



Escuela de Administración de Tecnologías de Información

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP-DM para la Mejora del Módulo de Inteligencia de Negocios de la Empresa PROCOM

Trabajo Final de Graduación para optar al grado de Licenciatura en
Administración de Tecnología de Información

Modalidad Proyecto de Graduación

Elaborado por: José Andrés Brenes Granados

Prof. Tutora: MSc Jacqueline Tatiana Solís Céspedes

Cartago, Costa Rica

Semestre II

Octubre, 2024



Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP-DM para la Mejora del Módulo de Inteligencia de Negocios de la Empresa PROCOM © 2024 por Jose Brenes Granados está licenciado bajo Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional

Hoja de Aprobación

INSTITUTO TECNOLÓGICO DE COSTA RICA
ESCUELA DE ADMINISTRACIÓN DE TECNOLOGÍAS DE
INFORMACIÓN GRADO ACADÉMICO: LICENCIATURA

Los miembros del Tribunal Examinador de la Escuela de Administración de Tecnologías de Información, recomendamos que el siguiente informe del Trabajo Final de Graduación del estudiante José Andrés Brenes Granados sea aceptado como requisito parcial para obtener el grado académico de Licenciatura de Tecnología de Información.



Profesora Tutora



Lectora externa

Lector académico

Coordinadora de
Trabajo Final de Graduación

Dedicatoria

A mis amados padres, Porfirio e Isabel, por ser los pilares fundamentales de mi vida y sembrar en mí la semilla de la perseverancia.

A mis queridos hermanos: Jacqueline, Gerardo, Melania y Fanny, por su apoyo incondicional y por creer siempre en mí.

Gracias por forjar mi carácter, por enseñarme que no hay meta inalcanzable cuando se persigue con determinación y esfuerzo.

La fortaleza que hoy me define es el fruto de su amor y sus enseñanzas. Esta dedicatoria es para ustedes, mi amada familia.

Resumen

El presente trabajo final de graduación se enfoca en la implementación de un proceso de análisis de datos fundamentado en la metodología *Cross-Industry Standard Process for Data Mining (CRISP-DM)* para mejorar el módulo de Inteligencia de Negocios en Procom, una empresa especializada en la creación de soluciones tecnológicas. La investigación se desarrolló utilizando un enfoque cualitativo, siguiendo un diseño de investigación-acción.

El estudio comprende el desarrollo y aplicación de cuatro modelos analíticos: (1) segmentación de clientes mediante *clustering* K-means, (2) análisis diagnóstico de tendencias de servicios, (3) diagnóstico de ingresos por habitaciones y (4) análisis comparativo de patrones de reservas. Estos se diseñaron para abordar objetivos específicos de minería de datos, incluyendo la identificación de segmentos de clientes, la evaluación de la rentabilidad de servicios y el análisis de patrones temporales en ingresos y reservas.

Los resultados revelaron tres segmentos de clientes claramente diferenciados, identificaron los diez servicios de mayor rentabilidad y detectaron patrones estacionales en ingresos y ocupación. Estos hallazgos se tradujeron en dashboards (tableros de control) interactivos desarrollados con DevExpress Report Designer, proporcionando una herramienta visual y analítica para la toma de decisiones estratégicas en el sector hotelero.

La implementación de este proceso no solo mejora las capacidades analíticas de Procom, sino que también establece una base sólida para fomentar una cultura de toma de decisiones basada en datos. El proyecto culmina con recomendaciones estratégicas para cada segmento de clientes y propuestas para mejorar la oferta de servicios y la gestión de habitaciones.

Este trabajo contribuye al campo de la inteligencia de negocios aplicada al sector hotelero, demostrando la eficacia de la metodología *CRISP-DM* en la transformación de datos en conocimientos accionables para la mejora de operaciones y la satisfacción del cliente.

Palabras clave: Inteligencia de Negocios; *CRISP-DM*; Análisis de Datos; Sector Hotelero; Segmentación de Clientes; Tableros de Control Interactivos.

Abstract

This graduation project focuses on implementing a data analysis process based on the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology to enhance the existing Business Intelligence (BI) module at Procom, a company specialized in creating technological solutions. The research was conducted using a qualitative approach, following an action-research design.

The study encompasses the development and application of four analytical models: (1) customer segmentation using K-means clustering, (2) diagnostic analysis of service trends, (3) room revenue diagnosis, and (4) comparative analysis of booking patterns. These models were designed to address specific data mining objectives, including customer segment identification, service profitability evaluation, and analysis of temporal patterns in revenue and reservations.

The results revealed three clearly differentiated customer segments, identified the top ten most profitable services, and detected seasonal patterns in revenue and occupancy. These findings were translated into interactive dashboards developed with DevExpress Report Designer, providing a visual and analytical tool for strategic decision-making in the hotel sector.

The implementation of this process not only enhances Procom's analytical capabilities but also establishes a solid foundation for fostering a data-driven decision-making culture. The project concludes with strategic recommendations for each customer segment and proposals to improve service offerings and room management.

This work contributes to the field of business intelligence applied to the hotel sector, demonstrating the effectiveness of the CRISP-DM methodology in transforming data into actionable insights for improving operations and customer satisfaction.

Keywords: Business Intelligence; CRISP-DM; Data Analysis; Hotel Sector; Customer Segmentation; Interactive Dashboards.

Tabla de Contenidos

1. Introducción.....	1
1.1 Descripción General.....	1
1.2 Antecedentes.....	2
1.2.1 Descripción de la organización	2
1.2.2 Trabajos similares realizados dentro y fuera de la organización.....	5
1.3 Planteamiento del problema.....	8
1.3.1 Situación problemática	8
1.3.2 Justificación del proyecto	10
1.3.3 Beneficios esperados o aportes del Trabajo Final de Graduación	11
1.4 Objetivos del Trabajo Final de Graduación	12
1.4.1 Objetivo General	12
1.4.2 Objetivos Específicos	12
1.5 Alcance.....	13
1.6 Exclusiones.....	13
1.7 Supuestos.....	14
1.8 Entregables	14
1.8.2. Gestión del proyecto.....	16
1.9 Limitaciones.....	17
2. Marco Conceptual.....	18
2.1 Mapa de Conceptos.....	18
2.2 Inteligencia de negocios	19
2.3 Sistemas de información	19
2.4 Recopilación y almacenamiento de datos.....	21
2.4.1 Gestor de bases de datos	21
2.4.2 <i>Data Warehouse</i>	21
2.4.2.1 <i>Data mart</i>	21
2.4.3 Proceso de Extracción, Carga y Transformación (ETL).....	22
2.4.4 Diccionario de datos	23

2.4.5 Modelado y Estructuración de Datos	24
2.4.5.1 Modelo relacional	25
2.5 Análisis de datos	28
2.5.1 Tipos de análisis de datos	28
2.5.2 Minería de datos	31
2.5.3 Visualización de datos.....	32
2.5.3.1 Mejores prácticas de visualización	32
2.5.3.2 Tipos de gráficos	33
2.5.3.3 <i>Dashboards</i>	36
2.6 Cultura <i>data-driven</i>	37
2.7 Modelos y Metodologías	37
2.7.1 Metodología <i>CRISP-DM</i>	38
2.7.2 Metodología Kimball.....	39
2.7.3 Metodología Inmon	40
2.7.4 Comparación entre Kimball e Inmon	41
2.7.5 Priorización de requerimientos <i>MoSCoW</i>	42
2.8 Indicadores Clave de Desempeño (KPIs) Hoteleros	42
2.9 Análisis de riesgos.....	45
2.9.1 Identificación de riesgos	45
2.9.2 Evaluación del riesgo	45
2.9.3 Análisis de riesgo inherente.....	46
2.9.5 Medición de los riesgos	47
3 Marco Metodológico	49
3.1 Tipo de Investigación	49
3.2 Enfoque de la Investigación.....	50
3.3 Diseño de la investigación	51
3.5 Sujetos de investigación	54
3.6 Variables o categorías de la investigación.....	55
3.7 Técnicas e instrumentos de recolección de datos.....	57
3.8 Procedimiento metodológico de la Investigación	58
3.8.1 Diagrama propuesto para las fases del procedimiento metodológico	58

3.8.2 Cronograma	59
3.9 Operacionalización de las variables o categorías.....	61
3.10 Resumen del procedimiento metodológico de la investigación	65
4 Análisis de Resultados	67
4.1 Comprensión del negocio	67
4.1.1 Determinación de objetivos empresariales	67
4.1.3 Determinación de <i>KPIs</i> , objetivos de minería de datos y criterios de éxito	81
4.1.4 Plan de implementación del proyecto.....	84
4.2 Entendimiento de los datos.....	85
4.2.1 Recolección inicial de los datos	86
4.2.2 Descripción de los datos	88
4.2.3 Exploración de los datos	94
4.2.4 Verificación de la calidad de los datos.....	120
5. Propuesta de Solución.....	124
5.1 Preparación de los datos	124
5.1.1 Seleccionar los datos.....	124
5.1.2 Limpieza de los datos.....	125
5.1.3 Integrar los datos	126
5.1.4 Construcción de datos	128
5.1.5 Formatear los datos	130
5.2 Modelado.....	131
5.2.1 Seleccionar las técnicas de modelado	131
5.2.2 Diseño de prueba	132
5.2.3 Evaluar el modelo	132
5.2.4 Ajustes de los parámetros del modelo y los análisis realizados	134
5.3 Evaluación.....	135
5.3.1 Evaluar resultados.....	135
5.3.2 Modelos aprobados	142
5.4 Despliegue	143
5.4.1 Objetivos del despliegue.....	143
5.4.2 Plan de despliegue	143

5.4.3 Integración de Datos.....	145
5.4.4 Dashboards realizado en el entorno.....	145
5.4.5 Recomendaciones.....	148
5.4.6 Matriz de cumplimiento de requerimientos	149
5.4.7 Matriz de cumplimiento de objetivos de negocio	150
5.4.8 Matriz de cumplimiento de objetivos de minería	151
5.4.10 Análisis de la viabilidad de la propuesta.....	152
6 Conclusiones.....	154
7 Recomendaciones	161
8 Referencias	168
9 Apéndices	172
Apéndice A. Plantilla de minutas para reunión.....	172
Apéndice B. Plantilla de gestión de cambios	173
Apéndice C. Plantilla de entrevista	173
Apéndice D. Plantilla Observaciones.....	173
Apéndice F. Plantilla Revisión Documental.....	174
Apéndice G. Entrevista con el gerente general 1.....	174
Apéndice H. Entrevista con el gerente general 2.....	175
Apéndice I. Entrevista con Gerente de Operaciones y Encargado de BI	176
Apéndice J. Entrevista con Gerente General y Gerente de Operaciones	178
Apéndice K. Entrevista con encargado de soporte de TI	179
Apéndice L. Reunión con profesora tutora 1	180
Apéndice M. Reunión con profesora tutora 2	181
Apéndice N. Reunión con profesora tutora 3	182
Apéndice O. Reunión con profesora tutora 4	183
Apéndice P. Reunión con profesora tutora 5.....	184
Apéndice Q. Reunión con profesora tutora 6	185
Apéndice R. Reunión con profesora tutora 7.....	186
Apéndice S. Reunión con profesora tutora 8.....	187
Apéndice T. Reunión con profesora tutora 9.....	188
Apéndice U. Reunión con profesora tutora 10	189

Apéndice V. Reunión con el encargado de BI	190
Apéndice W Reunión profesora tutora contraparte 1	191
Apéndice Y Reunión profesora tutora contraparte 2	192
Apéndice Z Reunión profesora tutora contraparte 3.....	193
Apéndice AA Primera entrevista con el gerente general.....	194
Apéndice AB. Segunda entrevista con el gerente general.....	195
Apéndice AC. Entrevista con la gerente de operaciones y el encargado de BI	196
Apéndice AD. Entrevista con el gerente general y la gerente de operaciones.....	198
Apéndice AE. Entrevista con encargado de Soporte de TI.....	199
Apéndice AF Entrevista con encargado de desarrollo	199
Apéndice AG. Revisión documental	199
Apéndice AH. Diccionario de datos a explorar.....	219
Apéndice AI. Datos seleccionados para el análisis.....	221
Apéndice AJ. Diccionario de datos de los modelos.....	231
Apéndice AK. Diccionario de datos del <i>Data Mart</i> y <i>reglas del ETL</i>	233
Apéndice AK. Firma de minutas del encargado de desarrollo	245
Apéndice AL. Firma de minutas del encargado de <i>BI</i>	245
Apéndice AM. Firma de minutas del encargado de soporte de TI.....	246
Apéndice AN. Firma de minutas del gerente general.	246
Apéndice AO. Firma de minutas de la gerente de operaciones.....	247
Apéndice AP. Acta de validación por parte del gerente general	248
Apéndice AQ. Acta de validación por parte de la gerente de operaciones	249
Apéndice AR. Firma de minutas de la profesora tutora	250
10 ANEXOS.....	251
11 Glosario	251

Índice de Figuras

Figura 1. Diagrama organizacional del equipo de proyecto.....	4
Figura 2. Árbol del problema	9
Figura 3. Mapa de Conceptos.....	18
Figura 4. <i>Ejemplo de diccionario de datos.</i>	24
Figura 5. <i>Ejemplo de un modelo entidad- relación</i>	25
Figura 6. <i>Ejemplo de esquema estrella</i>	27
Figura 7. <i>Ejemplo de esquema copo de nieve.</i>	28
Figura 8. <i>Ejemplo gráfico de barras</i>	33
Figura 9. <i>Ejemplo de gráfico de líneas</i>	34
Figura 10. <i>Ejemplo de gráfico circular</i>	34
Figura 11. Ejemplo de un gráfico de indicadores.....	35
Figura 12. <i>Ejemplo de Mapa de Calor</i>	35
Figura 13. <i>Guía de la metodología CRISP-DM</i>	38
Figura 14. <i>Ciclo de vida de Kimball</i>	39
Figura 15. <i>Ciclo de vida de Inmon</i>	41
Figura 16. Fases Metodológicas.....	58
Figura 17. Cronograma	59
Figura 18 <i>Estructura de tablas principales de la base de datos</i>	74
Figura 19. Diagrama de tablas unificadas	86
Figura 20. Análisis estadístico de las variables categóricas de la tabla Bill	88
Figura 21. Análisis estadístico de la fecha de la tabla Bill.....	89
Figura 22. Análisis estadístico de las variables numéricas de la tabla Bill.....	89
Figura 23. Análisis estadístico de las variables categóricas de la tabla ReservationBook.....	89
Figura 24. Análisis estadístico de las variables numéricas de la tabla ReservationBook	89
Figura 25. Análisis estadístico de las variables categóricas de la tabla	90
Figura 26. Análisis estadístico de las variables categóricas de la tabla Service	90
Figura 27. Análisis estadístico de las variables categóricas de Order.....	91
Figura 28. Análisis estadístico de las fechas de la tabla Order	91
Figura 29 Análisis estadístico de las variables categóricas de OrderCPL	91
Figura 30 Análisis estadístico de las fechas de la tabla Room.....	92
Figura 31. Análisis estadístico de las variables numéricas de OrderCPL.....	92
Figura 32. Análisis estadístico de las variables de ReservationService	92
Figura 33. Análisis estadístico de las variables de la tabla Log_ Stays	93
Figura 34 Prueba de normalidad para las variables numéricas de Bill	94
Figura 35 .Matriz de correlación para las variables categóricas Bill	94
Figura 36. Boxplot de las variables numéricas de la tabla Bill	95
Figura 37. Exploración de la variable Service en la tabla Bill.....	95
Figura 38 Exploración de la variable IDReservationDetail en la tabla Bill.....	96
Figura 39. Exploración de la variable Room en la tabla Bill	96
Figura 40. Exploración de la variable Client en la tabla Bill	97
Figura 41. Prueba de normalidad para las variables numéricas de la tabla ReservationBook.....	97
Figura 42. Matriz de correlación para las variables numéricas de la tabla ReservationBook.....	98

Figura 43. Matriz de correlación para las variables categóricas de la tabla ReservationBook	98
Figura 44. Exploración de la variable Client en la tabla ReservationBook	99
Figura 45. Exploración de la variable OrderID en ReservationBook	100
Figura 46. Exploración de la variable Confirmed en ReservationBook	100
Figura 47. Exploración de la variable ServiceID en ReservationBook.....	101
Figura 48. Exploración de las fechas en la tabla ReservationBook	101
Figura 49. Boxplots de las variables numéricas de ReservationBook	102
Figura 50. Exploración de la variable roomTypeName en la tabla Room	103
Figura 51. Exploración de las variable CostCenter y ServiceCategory en la tabla Service.....	103
Figura 52. <i>Matriz de correlación para las variables categóricas de la tabla Order</i>	104
Figura 53. Exploración de la variable Client en la tabla Order.....	105
Figura 54. Exploración de la variable Applied en la tabla Order.....	106
Figura 55. Exploración de la variable ReservationID en la tabla Order	106
Figura 56. Exploración de la variable InProcess en la tabla Order	107
Figura 57. Exploración de la variable RoomMain en la tabla Order	107
Figura 58. Exploración de las fechas en la tabla Order.....	108
Figura 59 Prueba de normalidad para las variables numéricas de la tabla OrderCPL	108
Figura 60. Matriz de correlación para las variables categóricas de la tabla OrderCPL	108
Figura 61. Exploración de la variable Client en la tabla OrderCPL	109
Figura 62. Exploración de la variable OrderCPL_Type en la tabla OrderCPL	110
Figura 63. Exploración de la variable Room en la tabla OrderCPL	110
Figura 64. Exploración de la variable Service en la tabla OrderCPL	111
Figura 65. Exploración de la variable OrderID en la tabla OrderCPL.....	111
Figura 66. Exploración de la variable ReservationID en OrderCPL.....	112
Figura 67. Exploración de las fechas en la tabla OrderCPL	112
Figura 68 Boxplot de la variable Total_Paid en la tabla OrderCPL	113
Figura 69. Prueba de normalidad para las variables numéricas de la tabla ReservationService..	113
Figura 70. Matriz de correlación para las variables numéricas de la tabla ReservationService. ..	113
Figura 71. Matriz de correlación para las variables categóricas de la tabla ReservationService.	114
Figura 72. <i>Exploración de la variable ClientID en ReservationService</i>	115
Figura 73. Exploración de la variable IsCPL en ReservationService	115
Figura 74. Exploración de las fechas en la tabla ReservationService.....	116
Figura 75. Boxplots de las variables numéricas de la tabla ReservationService	116
Figura 76. Matriz de correlación para las variables numéricas de la tabla Log_Stays.	117
Figura 77. Matriz de correlación para las variables categóricas de la tabla Log_Stays.....	118
Figura 78. Exploración de la variable Bill_ID en Log_Stays	118
Figura 79. Exploración de las fechas en la tabla Log_Stays.....	119
Figura 80. Boxplot de la variable Qty_Guests de la tabla Log_Stays.....	119
Figura 81. Porcentaje de valores nulos en la tabla Bill	120
Figura 82. Porcentaje de valores nulos en la tabla ReservationBook	121
Figura 83. Porcentaje de valores nulos en la tabla Order.....	122
Figura 84. Porcentaje de valores nulos en la tabla OrderCPL	122
Figura 85. Porcentaje de valores nulos después de la limpieza de datos.....	126
Figura 86. Diagrama dimensional	127
Figura 87. Diagramas de flujo del proceso ETL	128
Figura 88. Evaluación del Codo y Silueta.....	132

Figura 89. Modelo de segmentación de clientes	136
Figura 90. Tendencia de uso de los servicios.....	137
Figura 91. Modelo de tendencia de consumo de servicios.....	138
Figura 92. Diagnóstico de rentabilidad de habitaciones	139
Figura 93. Evolución mensual de ingresos por reserva de habitaciones	140
Figura 94. Análisis de los tipos de reservas	141
Figura 95. Tablero de análisis de segmentación de clientes	146
Figura 96. Tablero de análisis de servicios	146
Figura 97. Tablero de análisis de habitaciones	147
Figura 98. Tablero de análisis de ocupación y servicios.....	147

Índice de Tablas

Tabla 1. Proyectos similares internos.....	5
Tabla 2. Proyectos similares externos	6
Tabla 3. Entregables de la empresa.....	14
Tabla 4. Entregables académicos	16
Tabla 5. Sistemas de Información	19
Tabla 6. Mejores prácticas para la visualización de datos	32
Tabla 7. Comparación de metodologías Inmon y Kimball	42
Tabla 8. Indicadores Clave de Desempeño para el Sector Hotelero	43
Tabla 9. Cálculo de Riesgo Inherente	47
Tabla 10. Matriz de probabilidad-impacto.....	47
Tabla 11. Diseños de investigación cualitativa	51
Tabla 12. Fuentes primarias de la investigación	53
Tabla 13. Fuentes secundarias de la investigación.....	54
Tabla 14. Sujetos de investigación.....	55
Tabla 15. <i>Variables de Investigación</i>	55
Tabla 16. Instrumentos y plantillas de recolección de datos.....	57
Tabla 17. Operacionalización de las variables	61
Tabla 18. Resumen del procedimiento metodológico de la investigación	65
Tabla 19. Sistemas de Información de Procom.....	70
Tabla 20. Reportes encontrados en el módulo de BI.....	71
Tabla 21. Hardware utilizado	72
Tabla 22. Herramienta de visualización.....	73
Tabla 23. Gestor de base de datos	74
Tabla 24. Requisitos del proyecto	75
Tabla 25. Identificación de riesgos	78
Tabla 26. Resultado de inherencia	78
Tabla 27. Mapa de calor de riesgos.....	79
Tabla 28. Riesgos medidos.....	80
Tabla 29. Objetivos de minería y criterios de éxito	82
Tabla 30. Indicadores clave por objetivo de minería	82

Tabla 31. Herramientas seleccionadas por fase CRISP-DM.....	84
Tabla 32. Variables excluidas del análisis	124
Tabla 33. Atributos derivados del análisis	128
Tabla 34. Campos generados para el análisis.....	129
Tabla 35. Requisitos de formato para los modelos	130
Tabla 36. Técnicas de modelado implementadas.....	131
Tabla 37. Parámetros del modelo de segmentación	134
Tabla 38. Diagnóstico de Tendencias de Servicios.....	134
Tabla 39. Diagnóstico de ingresos por habitaciones	134
Tabla 40. Análisis de Duración de Estancias y Relación con Servicios	135
Tabla 41. Descripción de dashboards a implementar.....	143
Tabla 42. Matriz de cumplimiento de requerimientos	149
Tabla 43. Matriz de cumplimiento de objetivos de negocio	150
Tabla 44. Matriz de cumplimiento de objetivos de minería.....	151

1. Introducción

1.1 Descripción General

El enfoque basado en datos es fundamental para la estrategia empresarial actual. Procom enfrenta el desafío de aumentar el uso de los datos que genera para tomar decisiones y ayudar a sus clientes a hacerlo también.

Davenport y Harris (2017) presentan evidencia sobre cómo las organizaciones que adoptan un enfoque basado en datos tienen mayores probabilidades de éxito en diversos aspectos de su negocio. Sin embargo, a pesar de esta evidencia, cada empresa tiene un contexto único que requiere soluciones específicas.

Procom está generando una cantidad significativa de datos que no están siendo analizados, lo cual se refleja en reportes que no generan valor para sus clientes hoteleros en el módulo de Inteligencia de Negocios de su sistema de gestión hotelera (véase Apéndice H para la transcripción de la entrevista con el gerente general). Esta situación no es exclusiva de Procom; según DataAnalytics (2023), el 75 % de las empresas del sector tecnológico no aprovechan la información generada por sus sistemas.

En este proyecto se busca implementar un proceso de análisis de datos que permita extraer, transformar y cargar la información de un cliente hotelero de Procom para generar visualizaciones que faciliten la toma de decisiones. Se pretende que este proceso sirva como modelo escalable para aplicarlo a otros hoteles y segmentos de clientes.

A su vez, se espera que este proceso fomente el uso de los datos como eje para la toma de decisiones. Davenport y Harris (2017) argumentan que cuando los empleados y los líderes observan resultados tangibles derivados del análisis de datos, son propensos a creer en el valor de un enfoque basado en datos y a adoptarlo en sus prácticas.

El presente documento se estructura en siete secciones principales: (1) Introducción, (2) Antecedentes, (3) Planteamiento del problema, (4) Marco conceptual y metodológico, (5) Análisis de resultados, (6) Propuesta de solución y (7) Conclusiones y recomendaciones.

En la sección de antecedentes se proporciona una descripción de Procom, que incluye su misión, visión, valores y equipo de trabajo. También se examinan trabajos similares realizados dentro y fuera de la organización. El planteamiento del problema profundiza en la situación problemática, la justificación del proyecto y los beneficios esperados.

El objetivo general se enfoca en implementar un proceso de análisis de datos basado en la metodología CRISP-DM, y los objetivos específicos abordan el análisis de reportes, diseño de procesos de Extracción, Transformación y Carga (ETL), elaboración de cuadros de mando y validación del proceso implementado.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Posteriormente, se desarrolla un marco conceptual y metodológico, seguido por un análisis de resultados y una propuesta de solución. El documento finaliza con las conclusiones y recomendaciones derivadas del estudio.

1.2 Antecedentes

1.2.1 Descripción de la organización

Procom surge en 2002 con la misión de ofrecer soluciones innovadoras para automatizar operaciones, impulsar la productividad y fortalecer la competitividad de empresas en diversos sectores. En dos décadas, la empresa ha expandido su presencia a El Salvador y Nicaragua, consolidándose como un aliado para organizaciones que buscan sobresalir en el entorno tecnológico actual.

La organización se destaca por su enfoque centrado en brindar acompañamiento integral a sus clientes, escuchar las necesidades de las empresas y ofrecer herramientas personalizadas. Ha establecido relaciones con alrededor de mil clientes, basando su crecimiento en un modelo de innovación tecnológica.

Ofrece soluciones completas de facturación y punto de venta (Point of Sale [POS]) que permiten a las empresas gestionar sus operaciones comerciales. Estas soluciones abarcan desde la gestión de inventarios, pedidos, clientes y pagos, hasta la integración con otros sistemas empresariales como la administración de inventarios y recursos humanos.

Además, desarrolla sistemas de comercio electrónico que posibilitan a las empresas vender sus productos y servicios en línea. Estas soluciones incluyen la gestión de inventarios, pedidos, envíos y pagos, con la capacidad de integrarse con otros sistemas empresariales para una administración integral.

En el ámbito de los pagos y cobros automatizados, ha establecido una alianza estratégica con Ridivi, permitiendo a las empresas recibir pagos de manera segura, reduciendo el riesgo de pérdida de datos y mejorando la experiencia del cliente.

Adicionalmente, ha forjado alianzas estratégicas con Softland, una destacada empresa de software empresarial en Latinoamérica. Esta colaboración facilita a las empresas de sectores como alimentos y bebidas, hotelería, turismo y servicio automotriz el acceso a una plataforma integrada con los módulos contables del Enterprise Resource Planning (ERP) de Softland.

1.2.1.1 Misión

De acuerdo con la documentación institucional, la misión de Procom es:

"Transformar empresas por medio de la digitalización, utilizando herramientas tecnológicas innovadoras y a la vanguardia, para optimizar procesos, mejorar la eficiencia operativa, impulsar la productividad y elevar la competitividad en un entorno empresarial exigente y globalizado" (Procom, 2024).

1.2.1.2 Visión

Según la documentación institucional, la visión de Procom es:

"Ser la empresa de puntos de venta número uno en la región latinoamericana, ofreciendo soluciones tecnológicas innovadoras y a la vanguardia que transformen la forma en que las empresas gestionan sus operaciones comerciales y mejoran su competitividad" (Procom, 2024).

1.2.1.3 Valores

Los valores institucionales de Procom, según su documentación oficial, son:

- **Esfuerzo:** La organización se caracteriza por trabajar con pasión y entusiasmo, manteniéndose dispuesta a superarse y mejorar sus procesos y servicios. El esfuerzo y la perseverancia constituyen elementos fundamentales para alcanzar objetivos y brindar soluciones de calidad a los clientes.
- **Respeto:** Constituye la base de toda relación saludable y productiva en Procom. La empresa promueve un entorno de trabajo inclusivo y respetuoso, donde cada persona se siente valorada y apoyada.
- **Puntualidad:** La organización se compromete a mantener la precisión en sus plazos y cumplir sus compromisos, adaptándose constantemente a las necesidades de sus clientes.
- **Trabajo en Equipo:** La empresa fomenta la colaboración y la comunicación entre sus empleados para trabajar hacia objetivos comunes. Mantiene el compromiso de apoyo mutuo y el intercambio de conocimientos y habilidades para ofrecer soluciones de alta calidad.
- **Comunicación:** La organización prioriza la claridad y concisión en sus comunicaciones, manteniendo la disposición de escuchar y adaptarse a las necesidades de clientes y empleados.

1.2.1.4 Equipo de trabajo

A continuación, se detallan los roles y responsabilidades asignados al equipo encargado de ejecutar el proyecto:

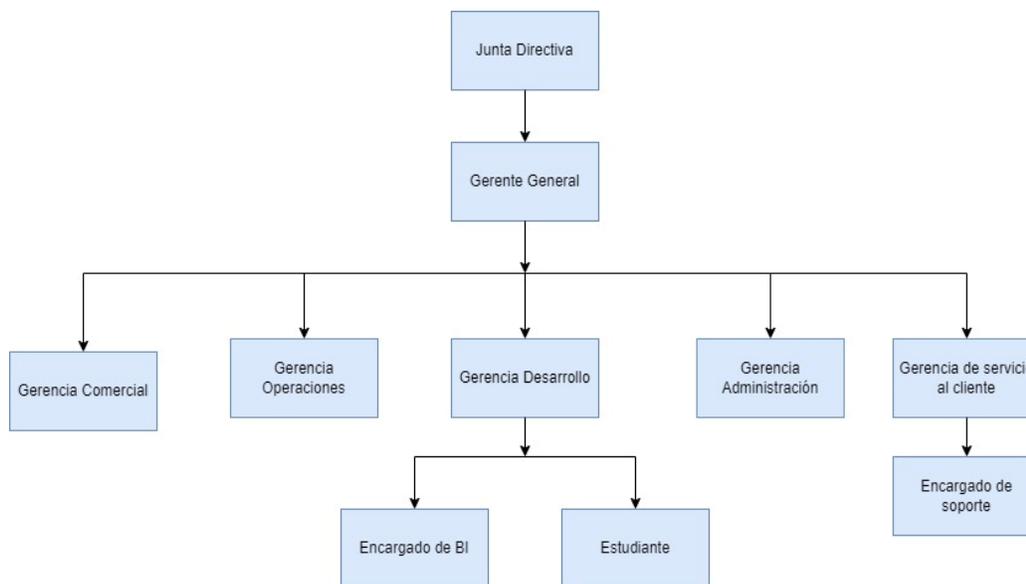
- **Estudiante:** Responsable del desarrollo de los elementos especificados en el alcance del proyecto, referentes a la implementación del proceso de análisis de datos para la mejora del módulo de Inteligencia de Negocios de Procom.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- Gerente General: Responsable de la gestión del proyecto y la asignación de recursos necesarios para su desarrollo.
- Encargado de desarrollo tecnológico: Responsable de orientar la comprensión de la estructura de las bases de datos.
- Encargado de soporte de los sistemas: Responsable de la instalación y configuración de las herramientas necesarias para el desarrollo del proyecto.
- Encargado de BI: Responsable de la gestión del módulo de Inteligencia de Negocios en los sistemas de Procom.
- Gerente de operaciones: Responsable de proporcionar información sobre los requerimientos del cliente.

En la Figura 1 se presenta el diagrama organizacional que ilustra la posición de este equipo dentro de la jerarquía de la compañía. Este organigrama facilita la comprensión de las líneas de reporte y las relaciones jerárquicas involucradas en la gestión del proyecto.

Figura 1. *Diagrama organizacional del equipo de proyecto*



Nota. Gerente general (comunicación personal, 2024).

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

1.2.2 Trabajos similares realizados dentro y fuera de la organización

A continuación, se describen proyectos similares desarrollados tanto dentro como fuera de la organización. La descripción incluye el nombre del proyecto, el objetivo general y su utilidad o aporte para el desarrollo del proyecto actual.

1.2.2.1 Proyectos similares internos

En la Tabla 1 se presentan los proyectos similares realizados dentro de la organización. Según la página web de Procom (2024), estos proyectos fueron desarrollados por la organización.

Tabla 1. *Proyectos similares internos*

Nombre del Proyecto	Objetivo General	Utilidad para el proyecto
Implementación de <i>PMS</i> y <i>POS</i> de alimentos y bebidas en Hotel Boutique.	Implementar una solución tecnológica integral que permita la centralización de la información, optimización de procesos, mejora de la eficiencia operativa y la experiencia del cliente en un hotel boutique ubicado en una zona turística de Costa Rica.	Proporciona una base de datos e información para alimentar el módulo de <i>BI</i> , permitiendo el análisis de datos para el hotel. También aporta lecciones aprendidas y buenas prácticas en la implementación de soluciones tecnológicas en el sector hotelero.
Digitalización de procesos de cobro a Cooperativa <i>Retail</i>	Implementar una solución tecnológica integral que transforme las operaciones de la cooperativa mediante la centralización y automatización de procesos clave, como compras, gestión de inventario, facturación, contabilidad, crédito y cobranza, para mejorar la eficiencia y reducir costos.	Aporta experiencia en la implementación de soluciones que optimizan la eficiencia y reducen costos.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Nombre del Proyecto	Objetivo General	Utilidad para el proyecto
Digitalización de procesos de restaurante con varias sucursales	Implementar una solución tecnológica integral que permita al restaurante la centralización y automatización de sus operaciones, mejora de la eficiencia, reducción de costos y toma de decisiones estratégicas informadas en tiempo real para mejorar la experiencia del cliente.	Sirve como ejemplo práctico de cómo una solución de datos centralizada mejora la eficiencia y la toma de decisiones en una empresa de servicios, similar al caso del hotel.

Nota. Adaptado de la página web de Procom (2024)

1.2.2.2 Proyectos similares externos

En la Tabla 2 se presentan los proyectos similares realizados fuera de la organización.

Tabla 2. *Proyectos similares externos*

Nombre del Proyecto	Objetivo General	Utilidad para el Proyecto
Implementación de BI para optimizar el rendimiento de las aplicaciones en la nube	Según García, J., y López, M. (2024), el objetivo fue mejorar el rendimiento y la escalabilidad de las aplicaciones en la nube mediante la implementación de un módulo de BI que proporcione información en tiempo real sobre el uso de recursos, rendimiento de las aplicaciones y experiencia del usuario.	Este proyecto proporciona conocimientos sobre la implementación de soluciones de inteligencia empresarial (BI) y el uso de datos en tiempo real para la toma de decisiones. Sirve como referencia para la implementación de herramientas y tecnologías de BI en Procom.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Nombre del Proyecto	Objetivo General	Utilidad para el Proyecto
Utilización de BI para análisis de datos de clientes en empresas de comercio electrónico	Según Rodríguez, E., y Fernández, D. (2022), el objetivo fue mejorar la comprensión del comportamiento del cliente, optimizar las estrategias de marketing y aumentar las ventas en el comercio electrónico mediante la implementación de un módulo de BI que proporciona información sobre las compras de los clientes.	Este proyecto aporta conocimientos sobre cómo implementar soluciones de BI para obtener información valiosa sobre los clientes y mejorar las estrategias de negocio.

Nota. Elaboración propia (2024)

1.3 Planteamiento del problema

En este apartado se expone la problemática identificada dentro del contexto organizacional que impulsa la realización del proyecto. Asimismo, se detallan los beneficios esperados de su implementación.

1.3.1 Situación problemática

Actualmente, Procom enfrenta un desafío crítico en la utilización efectiva de los datos generados por su sistema de gestión hotelera (*PMS*). A pesar de contar con una infraestructura tecnológica que captura información detallada de las operaciones hoteleras, la empresa no logra transformar estos datos en *insights* accionables para sus clientes, según lo documentado en entrevistas con el gerente general (véase **Apéndice H**).

Esta situación se manifiesta principalmente en el módulo de *BI* ofrecido por Procom a sus clientes, donde los reportes actuales se limitan a presentar datos sin procesar, careciendo de análisis que faciliten la toma de decisiones estratégicas. El gerente general señala que esta limitación afecta directamente la capacidad de los clientes hoteleros para optimizar sus operaciones y mejorar su desempeño financiero.

La problemática se agrava por la ausencia de personal especializado en análisis de datos y la falta de procesos estructurados para la generación de *insights*. Esta carencia organizacional, documentada en el **Apéndice H**, resulta en una subutilización sistemática de la información disponible en el sistema *PMS*.

Un estudio reciente de DataAnalytics (2023) contextualiza esta situación dentro de una tendencia más amplia en el sector tecnológico, donde el 75% de las empresas enfrentan desafíos similares en el aprovechamiento de sus datos. En el caso específico de Procom, esta tendencia se materializa en la generación de reportes que, aunque técnicamente funcionales, no satisfacen las necesidades analíticas de sus clientes hoteleros.

El gerente general (véase **Apéndice H**), señaló que la problemática de Procom se origina en una falta de cultura *data-driven* en la empresa, ya que los empleados no están familiarizados con herramientas avanzadas de análisis de datos y no existe una estrategia clara para su uso en la toma de decisiones.

Existe, además, una percepción errónea sobre el significado de una organización impulsada por datos, ya que el módulo de *BI* utiliza datos sin procesar que no generan información valiosa para la toma de decisiones. Esta confusión quedó evidenciada en un estudio de la Universidad de Tecnología (2022), el cual mostró que el 60% de las empresas del sector confunden los conceptos de *data-driven* e inteligencia empresarial.

Adicionalmente, Procom identificó la necesidad de mejorar los reportes de su módulo de *Business Intelligence (BI)*. Para lograrlo, la empresa planea analizar los datos

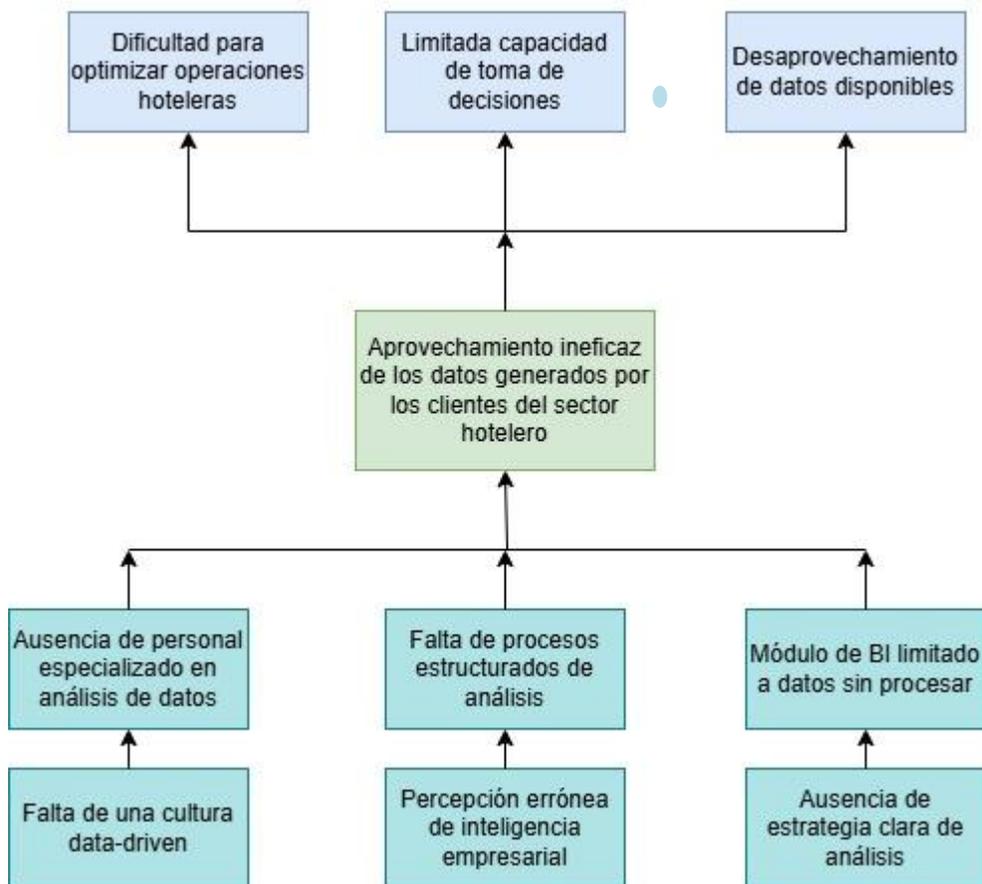
Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

generados por su sistema de *Property Management System (PMS)* para enriquecer y mejorar las visualizaciones actuales.

El proyecto comenzará con un cliente piloto del sector hotelero. A partir de los resultados obtenidos, se planea escalar esta solución a toda la cartera de clientes, permitiendo la estandarización del proceso.

A continuación, en la Figura 2, se ilustra esta problemática de manera gráfica.

Figura 2. *Árbol del problema*



Nota. Elaboración propia (2024)

1.3.2 Justificación del proyecto

La implementación del proceso de análisis de datos para mejorar el módulo de *Business Intelligence (BI)* existente en Procom representa una solución innovadora que aborda aspectos clave de la Administración de Tecnología de Información (ATI). Este proyecto se fundamenta en la metodología *CRISP-DM (Cross-Industry Standard Process for Data Mining)*, un enfoque estructurado para la gestión de proyectos de minería de datos (Chapman et al., 2000). La aplicación de *CRISP-DM* en este contexto asegura un abordaje sistemático y riguroso para el manejo de los datos.

El núcleo del proyecto se centra en la Inteligencia de Negocios, un campo que ha evolucionado en los últimos años. Según Larson y Chang (2016), la *BI* moderna se caracteriza por su agilidad y adaptabilidad, características que este proyecto busca implementar en Procom. La implementación del proceso de análisis de datos no solo busca mejorar el módulo de *BI* existente, sino que también fomenta una cultura de toma de decisiones basada en datos (Davenport & Harris, 2017).

En el ámbito de la Gestión de Datos e Información, el proyecto incorpora prácticas como la creación de un *data mart* y el modelado dimensional, siguiendo los principios establecidos por Kimball y Ross (2013).

La Estrategia de Tecnología de Información desempeña un papel fundamental en este proyecto, ya que la implementación del proceso de análisis de datos para mejorar el módulo de *BI* existente debe alinearse con los objetivos estratégicos de Procom. Ward y Peppard (2016) enfatizan la importancia de esta alineación para maximizar el valor de las inversiones en TI, un principio que guía el diseño y la implementación de la solución propuesta.

Es importante destacar que dicha implementación requiere no solo habilidades técnicas, sino también competencias de comunicación y liderazgo. Un profesional de ATI debe demostrar inteligencia emocional para afrontar los desafíos interpersonales que surgen en los proyectos.

1.3.3 Beneficios esperados o aportes del Trabajo Final de Graduación

En esta sección se mencionan los beneficios que se obtendrán al realizar el presente proyecto dentro de Procom.

1.3.3.1 Beneficios Directos

Dentro de los beneficios directos del proyecto se incluyen:

- Desarrollo de *data mart* centrado en el sector hotelero.
- Presentación de cuadros de mando interactivos con visualizaciones de indicadores clave de desempeño (*KPIs*).
- Definición de un proceso de análisis de datos basado en la metodología *CRISP-DM*.
- Identificación de oportunidades de mejora específicas para el sector hotelero.
- Estandarización del proceso de análisis para otros clientes del sector hotelero de Procom.

1.3.3.2 Beneficios Indirectos

En cuanto a los beneficios indirectos que la empresa obtendría mediante la implementación de esta propuesta, se incluyen:

- Primer paso hacia la promoción de una cultura de toma de decisiones basada en datos.
- Desarrollo de un proceso de análisis replicable en otros segmentos de clientes de Procom.

1.4 Objetivos del Trabajo Final de Graduación

En esta sección se describen los objetivos del proyecto según el problema descrito en la **Sección 1.3**, que abarca tanto el objetivo general como los específicos.

1.4.1 Objetivo General

Implementar un proceso de análisis de datos apoyado en la metodología CRISP-DM, con enfoque inicial en un cliente del sector hotelero para la mejora del módulo de BI existente, durante el segundo semestre de 2024.

1.4.2 Objetivos Específicos

Analizar los patrones de los reportes solicitados por los clientes hoteleros y los sistemas de información que sirven como fuente de datos para la definición de indicadores claves de rendimiento para los hoteles.

Construir un *data mart*, junto con los procesos de extracción, transformación y carga de los datos (*ETL*) generados, para la presentación de datos depurados y la creación de los modelos analíticos para la medición de los indicadores claves de rendimiento de los hoteles.

Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los *KPIs* en el módulo de *BI*, para la generación de recomendaciones sobre los resultados.

Validar el proceso de análisis de datos implementado con los gerentes de Procom que garantice la entrega de valor a la empresa.

1.5 Alcance

A continuación, se detalla el alcance del proyecto, que incluye los entregables derivados de su ejecución, así como las exclusiones, los supuestos y las limitaciones en función de los objetivos planteados en la **Sección 1.4**. Los entregables del proyecto comprenden los documentos utilizados para la gestión del proyecto y los resultados de su implementación.

1.6 Exclusiones

A continuación, se listan las exclusiones del proyecto:

- El proyecto excluye la implementación de los *dashboards* en el entorno de producción; los resultados se limitarán al entorno de desarrollo.
- No se incluye soporte técnico continuo ni mantenimiento posterior a la entrega de los *dashboards* en su versión no productiva.
- El proyecto no abarcará la integración de los nuevos *dashboards* con sistemas operativos del hotel que no estén directamente relacionados con el módulo de *BI* existente.
- Se excluirá un programa completo de capacitación para los usuarios finales del hotel.
- El proyecto se enfocará en el análisis descriptivo y diagnóstico; no se incluirán modelos predictivos o prescriptivos debido a la disponibilidad de datos limitada para estos análisis.
- El alcance se restringirá a la visualización y análisis de datos, excluyendo la automatización de procesos de negocio basados en los *insights* obtenidos.
- No se desarrollarán *APIs* para la integración de los *dashboards* con sistemas de terceros.
- Las visualizaciones y análisis estarán orientados a un cliente del sector hotelero, sin adaptaciones para otros clientes en esta versión.

1.7 Supuestos

Los siguientes supuestos se considerarán válidos y aplicables durante la ejecución del proyecto:

- La alta dirección de Procom respaldará y apoyará el proyecto, asignando los recursos necesarios para su desarrollo.
- Procom facilitará acceso a la documentación de sus sistemas, bases de datos y procesos de negocio, además de los recursos humanos necesarios para comprender el funcionamiento del módulo de *BI* actual.
- Procom cuenta con los recursos necesarios para llevar a cabo el proyecto, incluyendo personal, hardware y software adecuados.
- Los datos requeridos para mejorar el módulo de *BI* están disponibles, con acceso garantizado para los objetivos del proyecto.

1.8 Entregables

En esta sección, se detallan los entregables del proyecto, diseñados para cumplir con los objetivos establecidos en la **Sección 1.4** y abordar la problemática identificada. Los entregables se clasifican en dos categorías: entregables de producto y entregables de gestión del proyecto.

1.8.1. Entregables del producto

Los entregables de esta sección están asociados a cada objetivo del proyecto, con el fin de cumplir los objetivos establecidos y abordar la problemática identificada.

1.8.1.1 Entregables de la empresa

A continuación, se presenta la Tabla 3, que describe los entregables desarrollados para la empresa.

Tabla 3. *Entregables de la empresa*

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo	Entregable	Descripción
Analizar los patrones de los reportes solicitados por los clientes hoteleros y los sistemas de información que sirven como fuente de datos para la definición de indicadores claves de rendimiento para los hoteles.	Documento de Análisis de Reportes Solicitados por los Clientes, Sistemas de Información	Este informe capturará los reportes solicitados por los clientes hoteleros y mapeará los sistemas de información existentes. Incluirá un resumen de los reportes solicitados para identificar patrones de solicitud, y un inventario completo de los sistemas de información con características y contenidos de cada fuente de datos relevante para el proyecto.
	Catálogo de Indicadores Clave de Desempeño (<i>KPIs</i>)	Este documento contendrá una lista de los <i>KPIs</i> identificados. Cada <i>KPI</i> tendrá su definición, fórmula de cálculo, fuentes de datos requeridas y su relevancia para los objetivos de negocio de Procom.
Construir un <i>data mart</i> , junto con los procesos de extracción, transformación y carga de los datos (<i>ETL</i>) generados, para la presentación de datos depurados que permitan la medición de los indicadores claves de rendimiento de los hoteles.	Diseño del Proceso <i>ETL</i> , diagramas de flujo y diccionarios de datos	Este entregable incluirá documentos técnicos que detallan el diseño del proceso de Extracción, Transformación y Carga (<i>ETL</i>) de datos, con diagramas de flujo que ilustran el recorrido de los datos desde cada fuente hasta su destino en el módulo de BI. También se adjuntará el código aplicado en el proceso.
	Diseño del <i>Data Mart</i>	Este entregable detallará el modelo de datos dimensional (esquema estrella o copo de nieve) junto con diccionarios de datos que incluyen descripciones, justificación, tipo de dato y reglas de transformación.
Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los <i>KPIs</i> en el módulo de <i>BI</i> , para la generación de recomendaciones sobre los resultados.	Cuadros de Mando Interactivos	Entregable que incluye al menos dos cuadros de mando funcionales e interactivos implementados en el módulo de <i>BI</i> existente, con visualizaciones gráficas intuitivas respaldadas en los <i>KPIs</i> específicos del sector hotelero.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo	Entregable	Descripción
Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los <i>KPIs</i> en el módulo de <i>BI</i> , para la generación de recomendaciones sobre los resultados.	Documentación técnica de Cuadros de Mando	Documento con detalles técnicos sobre las visualizaciones y cálculos utilizados en cada cuadro de mando, facilitando futuras actualizaciones o personalizaciones.
Validar el proceso de análisis de datos implementado con los gerentes de Procom que garanticen la entrega de valor a la empresa.	Presentación de resultados	Documento de referencia para Procom, que presenta una visión clara del proceso implementado y su contribución al cliente y a la empresa.

Nota. Elaboración propia (2024)

1.8.1.2 Entregables académicos

Tabla 4. *Entregables académicos*

Entregable	Descripción
Informes de avance	Informes periódicos que detallan el progreso en el desarrollo del proyecto.
Informe Final	Documento final que incluye los elementos estipulados en el reglamento del proyecto de graduación.

Nota. Elaboración propia (2024)

1.8.2. Gestión del proyecto

Además de los entregables asociados a los objetivos del proyecto definidos en la **Sección 1.4**, se generarán una serie de entregables relacionados con la gestión y administración del proyecto, que servirán como evidencia y documentación del proceso de ejecución.

1.8.2.1 Minutas

Las minutas se realizarán al finalizar cada sesión y su objetivo será capturar y documentar de manera clara y concisa los aspectos más relevantes de cada reunión. Los campos se detallan en el **Apéndice A**.

1.8.2.2 Gestión del Cambio

La plantilla de gestión de cambios del proyecto se utilizará para recopilar y evaluar las solicitudes de cambio de manera estructurada y estandarizada. La plantilla se encuentra en el **Apéndice B**.

1.9 Limitaciones

A continuación, se detallan los factores y elementos que limitan el avance del proyecto:

- La base de datos proporcionada del cliente piloto contiene 7 meses de información.
- La base de datos recibe información de múltiples sistemas, lo que genera tablas de estructura compleja.
- Existen registros incorrectos en determinados campos de la base de datos.
- No se dispone de documentación sobre la estructura de la base de datos de los hoteles.
- Retrasos en el inicio estipulado del proyecto.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

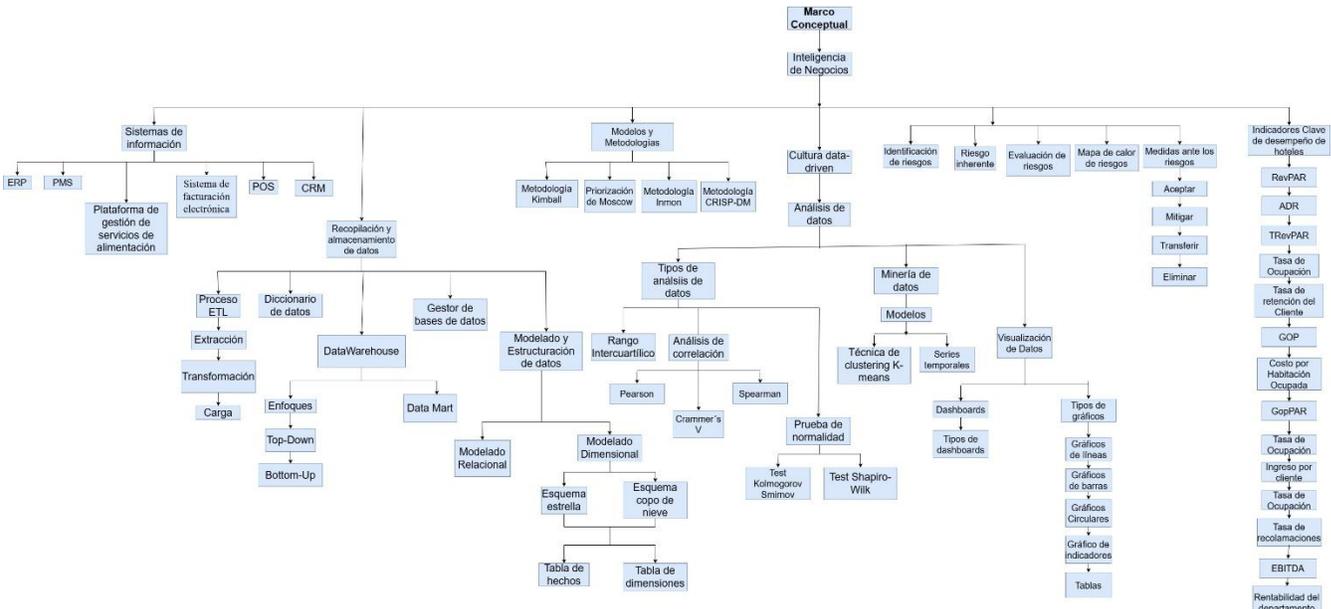
2. Marco Conceptual

En este capítulo, se desarrolla el marco conceptual que sustenta la investigación sobre la mejora del módulo de BI. Se exploran los conceptos fundamentales, metodologías y modelos que permiten comprender y analizar la intersección entre la gestión hotelera, los sistemas de información y la inteligencia de negocios.

2.1 Mapa de Conceptos

La Figura 3 muestra una visualización de los conceptos necesarios para la comprensión del trabajo.

Figura 3. Mapa de Conceptos



Nota. Elaboración Propia (2024)

2.2 Inteligencia de negocios

La inteligencia de negocios (*BI*, por sus siglas en inglés) se define como un conjunto integrado de estrategias, procesos, aplicaciones, datos, tecnologías y arquitecturas técnicas orientadas a la recopilación, análisis, presentación y difusión sistemática de información empresarial (Larson & Chang, 2016). En el contexto de la gestión hotelera, la *BI* desempeña un papel crucial al transformar datos brutos en conocimientos accionables, facilitando así la toma de decisiones tanto a nivel estratégico como operativo.

Davenport y Harris (2017) enfatizan que la *BI* no se limita a la implementación de herramientas tecnológicas, sino que representa un enfoque sistemático para utilizar los datos y el análisis como una ventaja competitiva sostenible. Su aplicación efectiva en el sector hotelero permite a las organizaciones anticipar tendencias del mercado, personalizar las experiencias de los huéspedes y tomar decisiones informadas basadas en evidencia empírica (Bilgihan et al., 2021).

La inteligencia de negocios incluye varios conceptos que se desarrollan a lo largo de esta sección.

2.3 Sistemas de información

Los sistemas de información son un componente esencial en la arquitectura tecnológica de las organizaciones modernas, especialmente en el sector hotelero. Según Silberschatz et al. (2020), un sistema de información es un conjunto de componentes interrelacionados que recopilan, procesan, almacenan y distribuyen información para apoyar la toma de decisiones, la coordinación y el control en una organización.

En la industria hotelera, estos desempeñan un papel clave en la gestión eficiente de las operaciones diarias y en la toma de decisiones estratégicas. Bilgihan et al. (2021) destacan que en el sector hotelero abarcan diversas aplicaciones, incluyendo sistemas de gestión de propiedades (*PMS*), de reservas, de gestión de ingresos y de gestión de relaciones con los clientes (*CRM*), entre otros. A continuación, en la Tabla 5, se describen algunos sistemas de información relevantes para este proyecto:

Tabla 5. Sistemas de Información

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Sistema	Descripción
<i>ERP (Enterprise Resource Planning)</i>	Bernal (2010) lo define como un sistema de planificación de recursos empresariales que integra varias funciones clave de negocio, como finanzas, gestión de inventarios, compras y recursos humanos, en una única plataforma de software. Este tipo de sistema ayuda a las organizaciones a automatizar y gestionar procesos empresariales, proporcionando una visión integrada y en tiempo real de sus operaciones.
<i>PMS (Property Management System)</i>	Según Bilgihan et al. (2021), es una solución integral de gestión empresarial diseñada para la industria de la hospitalidad. Está dirigida a diversos segmentos del sector, como hoteles, parques de aventura, establecimientos con restaurante y clubes sociales.
Plataforma de gestión para servicios de alimentación	La Cámara Costarricense de Hoteles (2021) la define como una plataforma para gestionar operaciones en restaurantes o servicios de alimentos, incluyendo el control de inventarios, pedidos, planificación de menús y seguimiento de costos. Facilita la administración eficiente de los recursos y mejora la toma de decisiones a partir de datos integrados.
Sistema de facturación electrónica	La Cámara Costarricense de Hoteles (2021) lo define como un sistema que permite emitir, recibir y almacenar facturas de manera digital, cumpliendo con las regulaciones locales y mejorando la eficiencia administrativa. Facilita el control financiero y contable, reduciendo la necesidad de procesos manuales.
Sistema de punto de venta	Según Bilgihan et al. (2021), es una herramienta tecnológica utilizada en establecimientos comerciales para realizar transacciones de ventas. Además de gestionar pagos, permite el control de inventarios, la generación de reportes de ventas y facilita la integración con otros sistemas, como la contabilidad o el CRM.
Sistema integral de Gestión de Relaciones con Clientes (<i>Customer Relationship Management</i>)	Davenport et al. (2017) lo definen como un sistema que ayuda a las empresas a gestionar la relación con sus clientes, centralizando la información de estos y permitiendo un seguimiento detallado de las interacciones. Facilita la personalización del servicio, mejora la retención de clientes y aumenta la eficacia en los procesos de ventas y marketing.

Nota. Elaboración Propia (2024)

2.4 Recopilación y almacenamiento de datos

La recopilación y almacenamiento de datos son procesos fundamentales en la implementación de soluciones de inteligencia de negocios, especialmente en el contexto del sector hotelero. Estos procesos están intrínsecamente ligados a los objetivos del proyecto definidos en la **Sección 1.4** y a la infraestructura de sistemas de información descrita en la **Sección 2.3**.

La recopilación de datos implica la captura sistemática de información relevante de diversas fuentes dentro de la organización. Estas fuentes, exploradas en la **Sección 2.3**, incluyen sistemas de gestión de propiedades (*PMS*), de reservas, puntos de venta (*POS*) y de gestión de relaciones con los clientes (*CRM*), entre otros. Comprender estas fuentes es esencial para el primer objetivo específico del proyecto planteado en la **Sección 1.4**.

El almacenamiento de datos se refiere a la organización y mantenimiento de la información recopilada en una estructura definida por la empresa. Aunque en el contexto de la inteligencia de negocios es común encontrar soluciones como almacenes de datos (*data warehouses*) o *data marts*, existen múltiples enfoques y tecnologías para el almacenamiento de datos.

2.4.1 Gestor de bases de datos

Un sistema gestor de bases de datos (*SGBD*, por sus siglas en inglés) es un conjunto de programas que permiten la creación, mantenimiento y uso de bases de datos. Este software facilita la definición de estructuras de datos, el almacenamiento de la información, su manipulación y el acceso compartido por múltiples usuarios y aplicaciones (Silberschatz et al., 2020). Estos sistemas son esenciales para la implementación del segundo objetivo específico del proyecto planteado en la **Sección 1.4**.

2.4.2 Data Warehouse

Un *data warehouse* es un repositorio central de datos diseñado específicamente para el análisis y la toma de decisiones. Kimball y Ross (2013) lo definen como "una copia de los datos de transacción estructurada específicamente para consultas y análisis" (p. 3).

2.4.2.1 Data mart

Un data mart es un subconjunto de un *data warehouse*, generalmente enfocado en un área específica del negocio. Kimball y Ross (2013) describen los *data marts* como "subconjuntos lógicos del *data warehouse* completo" (p. 11). Este permitirá a Procom organizar y presentar los datos relevantes para los *KPIs* del sector hotelero de manera eficiente y accesible.

Primero, aborda el segundo objetivo planteado en la **Sección 1.4**, centrado en la construcción de un data mart junto con los procesos de Extracción, Transformación y Carga (*ETL*) asociados. Esta infraestructura de datos facilitará la consolidación y estructuración de la información para el análisis del rendimiento hotelero.

En segundo lugar, el data mart servirá como base para alcanzar el tercer objetivo específico planteado en la **Sección 1.4**, que contempla la elaboración de cuadros de mando interactivos. Estos cuadros de mando, alimentados por los datos organizados en el data mart, proporcionarán visualizaciones dinámicas de los *KPIs*, permitiendo así una interpretación intuitiva y accesible de la información crítica para la toma de decisiones en el sector hotelero.

2.4.2.2 Enfoques del *Data Warehouse*

Existen varios enfoques para la construcción de un *data warehouse*, los cuales también se asocian con la creación de *data marts*:

- Enfoque *Top-Down* (Inmon): Primero se construye un *data warehouse* empresarial centralizado, del cual luego se derivan *data marts* (Data Management Association International, 2017).
- Enfoque *Bottom-Up* (Kimball): Se construyen *data marts* de manera incremental, que juntos forman el *data warehouse* empresarial (Kimball & Ross, 2013).

Davenport y Harris (2017) destacan la importancia de los *data warehouses* y *data marts* en la analítica de negocios, subrayando que proporcionan una "única versión de la verdad" para la toma de decisiones basadas en datos.

2.4.3 Proceso de Extracción, Carga y Transformación (ETL)

El proceso de Extracción, Transformación y Carga (*ETL*, por sus siglas en inglés) constituye un componente esencial en la arquitectura de análisis de datos. Munappy et al. (2020) lo definen como un conjunto sistemático de procedimientos que implican la extracción de datos de diversas fuentes, su transformación para garantizar compatibilidad y coherencia, y su posterior carga en el sistema de almacenamiento destino. Este proceso es fundamental para la integración efectiva de datos en entornos de inteligencia de negocios.

Chapman et al. (2000), en su guía metodológica *CRISP-DM (Cross-Industry Standard Process for Data Mining)*, sitúan el proceso *ETL* dentro de la fase de "preparación de datos", descrita en la **Sección 2.7.1**. Los autores destacan que la inclusión de este es crucial para el éxito de los proyectos de análisis de datos y se descompone en tres fases principales:

2.4.3.1 Extracción:

La fase de extracción consiste en la obtención de datos de múltiples fuentes, tanto internas como externas a la organización (Munappy et al., 2020). En el contexto del proyecto de Procom, esta fase involucra la extracción de datos de diversos sistemas hoteleros, como los descritos en la **sección 2.3**.

2.4.3.2 Transformación:

Durante la fase de transformación, los datos extraídos se someten a diversas operaciones para cumplir con los requisitos del sistema destino. Munappy et al. (2020) señalan que existe la posibilidad de incluir limpieza de datos, eliminación de duplicados, conversión de formatos y aplicación de reglas de negocio. En el sector hotelero, es posible que incluya la estandarización de formatos de fechas o la agregación de datos para calcular los *KPIs*.

2.4.3.3 Carga

La fase final del proceso ETL implica la carga de los datos transformados en el sistema destino, generalmente un *data warehouse* o un *data mart* (Munappy et al., 2020). En el proyecto de Procom, esta fase se alinea con el segundo objetivo específico planteado en la **sección 1.4**, que abarca la construcción de un data mart para el análisis de *KPIs* del sector hotelero.

2.4.4 Diccionario de datos

Un diccionario de datos es una colección centralizada y estructurada de información sobre los datos almacenados en una base de datos. Su función principal es proporcionar información sobre la estructura, formato y uso de estos dentro de un sistema de información. El Data Management Association International (2017), en su guía DAMA-DMBOK, los define como "datos que describen otros datos" (p. 409). En este contexto, es esencial para organizar y presentar estos metadatos de manera sistemática y accesible.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

A continuación, en la Figura 5 se muestra un ejemplo de un diccionario de datos.

Figura 4. *Ejemplo de diccionario de datos.*

NOMBRE DE LA COLUMNA	PK	TIPO DE DATO	TAMANO	FK	DESCRIPCIÓN
COD_AGENCIA	1	INT			Código numérico secuencial autoincrementado (AUTO_INCREMENT).
NOMBRE_AGENCIA		VARCHAR	100		Nombre de la agencia.
COD_CIUADAD		INT		CIUDADES	Código de ciudad de la agencia, referencia a la tabla ciudades.
DIRECCION_AGENCIA		TEXT			Dirección de la agencia.
TELEFONO_AGENCIA		CHAR	9		Teléfono de la agencia.
RESPONSABLE_AGENCIA		CARCHAR	100		Responsable de la agencia.
ESTADO		CHAR	1		Estado de la agencia puede ser 'A' activa o 'I' inactiva.

Nota. Tomado de Vásquez, D. (2012).

En el proyecto de Procom, la implementación de un diccionario de datos es crucial para asegurar la consistencia y comprensión de los campos utilizados en el análisis de *KPIs* del sector hotelero. Este recurso facilitará la interpretación correcta por parte de los analistas y usuarios finales, contribuyendo a la precisión de los cuadros de mando interactivos propuestos en el tercer objetivo específico planteado en la **Sección 1.4**.

La integración del proceso *ETL* y la implementación de un diccionario son componentes fundamentales para el éxito del proyecto. Estos elementos no solo facilitarán la construcción del data mart, sino que también sentarán las bases para un análisis de datos preciso y confiable, alineándose con los objetivos de proporcionar *insights* a los clientes del sector hotelero.

2.4.5 Modelado y Estructuración de Datos

El modelado y estructuración de datos es fundamental en la arquitectura de soluciones de inteligencia de negocios, estableciendo un puente entre los sistemas de información operativos y las herramientas analíticas avanzadas. Este proceso implica el diseño e implementación de una arquitectura que soporte eficazmente las necesidades de análisis y toma de decisiones de una organización (Silberschatz et al., 2020).

En el proyecto de Procom, el modelado es un elemento clave para la consecución de los objetivos planteados en la **Sección 1.4**, en particular para la construcción del *data mart* y la generación de *KPIs*. Este enfoque abarca tanto el modelo relacional como el dimensional, reconociendo sus roles complementarios en el análisis de datos empresariales.

2.4.5.1 Modelo relacional

Introducido por E.F. Codd en 1970, el modelo relacional se ha consolidado como el paradigma dominante en el diseño de bases de datos. Kimball y Ross (2013) subrayan su importancia en el diseño de *data warehouses*, especialmente en el enfoque de Inmon, que utiliza un modelo normalizado. Esta perspectiva se alinea con el segundo objetivo del proyecto planteado en la **Sección 1.4**, que busca construir un *data mart* para el análisis de datos del sector hotelero.

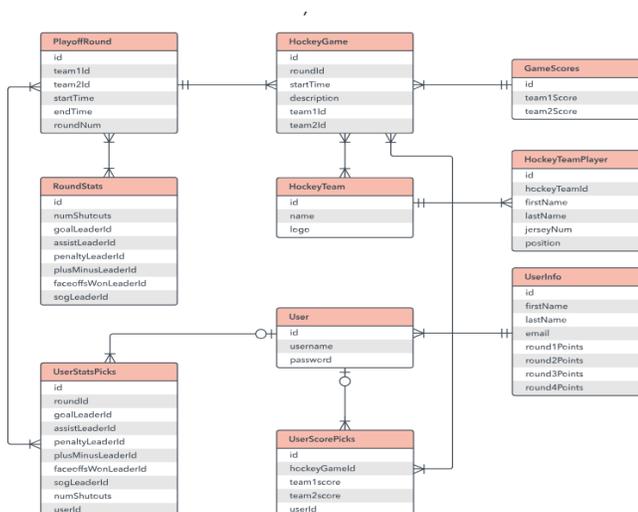
La aplicación del modelo relacional en el proyecto de Procom facilita la integración de datos provenientes de varios sistemas de información hoteleros, como se discute en la **Sección 2.3**. Esto proporciona una base sólida para el proceso *ETL*, abordado en la **Sección 2.4.3**, permitiendo una transformación y carga de datos eficientes en el *data mart* propuesto.

2.4.5.2 Modelo entidad-relación

El modelo entidad-relación, como lo define el Data Management Association International (2017), constituye una herramienta crucial en la fase de diseño conceptual de una base de datos. Este modelo facilita la visualización y el consenso sobre la estructura de datos entre diseñadores y *stakeholders*, aspectos esenciales para el éxito del proyecto de Procom.

A continuación, se muestra la Figura 5 con un ejemplo de cómo se visualiza el modelo entidad-relación.

Figura 5. *Ejemplo de un modelo entidad- relación*



Nota. Tomado de Lucid Software Inc. (2023).

Sus componentes principales incluyen:

- Entidades: Representan objetos o conceptos del mundo real
- Atributos: Características o propiedades de las entidades
- Relaciones: Conexiones o asociaciones entre entidades
- Cardinalidad: Indica cuántas instancias de una entidad están relacionadas con una instancia de otra.

La implementación efectiva del modelo entidad-relación en el proyecto contribuye a la creación de un diccionario de datos robusto, como se explicó en la **Sección 2.4.4**. Este facilita la comprensión y validación del análisis de datos por parte de los gerentes de Procom, alineándose con el cuarto objetivo del proyecto en la **Sección 1.4**.

Además, este modelo proporciona una base conceptual para el diseño de cuadros de mando interactivos, como se plantea en el tercer objetivo del proyecto de la **Sección 1.4**, permitiendo una visualización efectiva de los *KPIs* y otros indicadores relevantes para el sector hotelero.

2.4.5.2 Modelo dimensional

El modelo dimensional es una técnica de diseño lógico ampliamente utilizada en la construcción de *data warehouses* y *data marts*. Kimball y Ross (2013) destacan que este modelo facilita tanto la comprensión como el rendimiento de las consultas analíticas. Este enfoque se alinea con el segundo objetivo específico del proyecto en la **Sección 1.4**

2.4.5.2.1 Tabla de hechos

Kimball y Ross (2013) destacan la tabla de hechos como el componente central del modelo dimensional. Cada registro en esta tabla representa un evento de negocio específico, denominado "hecho" o "medida". En el contexto del proyecto de Procom, cabe la posibilidad que se incluyan eventos como reservas de habitaciones, transacciones de servicios hoteleros o registros de ocupación. Está relacionada con el proceso *ETL*, ya que es el destino principal de los datos transformados y cargados.

2.4.5.2.2 Tabla de dimensiones

Las tablas de dimensiones, según Kimball y Ross (2013), contienen atributos descriptivos y contextuales que permiten analizar y filtrar los datos numéricos almacenados en las tablas de hechos. En el proyecto de Procom, es posible que estas incluyan dimensiones como tiempo, cliente, habitación o servicio, proporcionando el contexto necesario para el análisis de los *KPIs* del sector hotelero, tal como se describe en el primer objetivo específico en la **Sección 1.4**.

2.4.5.2.3 Esquema estrella

El esquema estrella, descrito por Kimball y Ross (2013), consiste en una tabla de hechos central rodeada de dimensiones directamente relacionadas. Este esquema es relevante para el tercer objetivo del proyecto en la **Sección 1.4**, enfocado en la creación de cuadros de mando interactivos. La simplicidad y eficiencia de este facilita la creación de visualizaciones comprensibles de los *KPIs* del sector hotelero.

A continuación, se muestra la Figura 6 con un ejemplo de un esquema estrella.

Figura 6. *Ejemplo de esquema estrella*



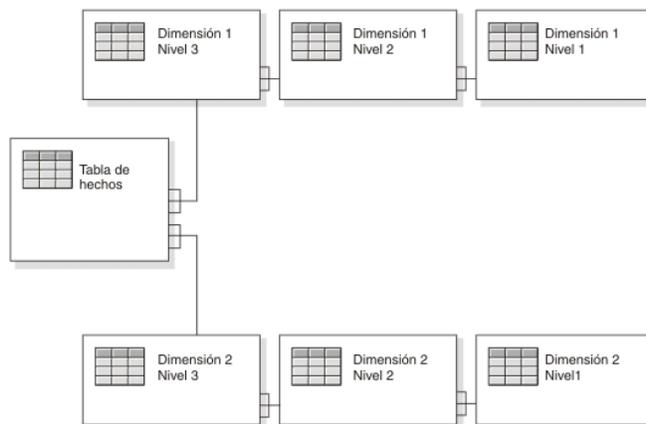
Nota. Tomado de IBM (2021).

2.4.5.2.4 Esquema copo de nieve

Según Kimball y Ross (2013), el esquema copo de nieve es una variante del esquema estrella en el modelado dimensional, en la que las tablas de dimensiones se normalizan en múltiples tablas relacionadas. La tabla de hechos central se mantiene igual que en el esquema estrella, pero las dimensiones se dividen en tablas adicionales, creando una jerarquía.

A continuación, se muestra la Figura 7 con un ejemplo del esquema copo de nieve.

Figura 7. Ejemplo de esquema copo de nieve.



Nota. Tomado de IBM (2021).

2.5 Análisis de datos

El análisis de datos constituye un componente esencial en el proceso de inteligencia de negocios, interconectando los conceptos previamente discutidos y contribuyendo directamente a los objetivos del proyecto planteados en la **Sección 1.4**. Davenport y Harris (2017) lo definen como "el uso extensivo de datos, análisis estadístico y cuantitativo, modelos explicativos y predictivos, y gestión basada en hechos para impulsar decisiones y acciones" (p. 7), lo cual se alinea con el propósito general del proyecto.

2.5.1 Tipos de análisis de datos

Algunos de los distintos tipos de análisis de datos descritos por Davenport y Harris (2017) se relacionan directamente con los objetivos del proyecto:

- **Análisis descriptivo:** Proporciona una base para comprender el estado actual de los datos y las necesidades de información del sector, es decir, resume lo que ha sucedido en el pasado.
- **Análisis diagnóstico:** Ayuda a comprender las razones detrás de los patrones observados en los datos hoteleros, es decir, explora por qué sucedió algún evento.
- **Análisis predictivo y prescriptivo:** Implica la elaboración de cuadros de mando interactivos, que permiten visualizar el estado actual, prever

tendencias futuras y recomendar acciones basadas en los datos.

2.5.1.1 Rango Intercuartílico

El *National Institute of Standards and Technology* (2012) define el rango intercuartílico (*IQR*) como una medida de dispersión que evalúa la variabilidad de la mitad central de los datos, minimizando la influencia de valores extremos. En el análisis exploratorio, el *IQR* es útil para identificar *outliers* y comprender la distribución de los datos, especialmente en gráficos como el *boxplot*, donde ayuda a visualizar la mediana, los cuartiles y los valores atípicos.

2.5.1.2 Prueba de normalidad

Ghasemi y Zahediasl (2012) definen la prueba de normalidad como un procedimiento estadístico utilizado para determinar si un conjunto de datos sigue una distribución normal. Esta prueba es esencial en métodos estadísticos como las pruebas t y el análisis de varianza (ANOVA), donde la normalidad es un supuesto crucial para la validez de los resultados. Entre las pruebas de normalidad más utilizadas están la de Shapiro-Wilk, Kolmogorov-Smirnov y Anderson-Darling.

2.5.1.2.1 Test de Shapiro-Wilk.

Es una prueba de normalidad que evalúa si una muestra de datos proviene de una población con una distribución normal. Calcula una estadística W que compara la varianza de los datos observados con la varianza de los datos esperados bajo una distribución normal.

Un valor de W cercano a 1 indica que los datos son consistentes con la normalidad, mientras que un valor menor sugiere que los datos no provienen de una población normal.

Asimismo, si el valor p es menor que el nivel de significancia (comúnmente 0.05), se rechaza la hipótesis nula, lo que indica que hay evidencia suficiente para concluir que los datos no se distribuyen normalmente.

2.5.1.2.2 Test de Kolomogorov Smirnov

Según Ghasemi et al. (2012), es una prueba estadística que evalúa la máxima diferencia absoluta entre la función de distribución empírica de un conjunto de datos y la función de distribución acumulativa de una distribución de referencia.

Esta prueba se utiliza para determinar si un conjunto de datos sigue una distribución específica, como la normal, y no requiere que los datos cumplan con supuestos estrictos de normalidad. Es sensible a discrepancias tanto en la localización como en la forma de la distribución.

2.5.1.3 Análisis de correlación

El análisis de correlación es una técnica estadística que permite evaluar la relación entre dos o más variables. Esta técnica proporciona información sobre la dirección (positiva o negativa) y la fuerza de la relación entre las variables, sin asumir causalidad (Field, 2013).

Es fundamental en el análisis exploratorio de datos, ya que ayuda a identificar patrones y tendencias en un conjunto de datos, resultando útil para la toma de decisiones. A continuación, se muestran algunos tipos de correlaciones de interés para este trabajo:

2.5.1.3.1 Correlación de Spearman

La correlación de Spearman es una medida no paramétrica que evalúa la relación entre dos variables ordinales o entre variables continuas que no siguen una distribución normal. Esta correlación se basa en el rango de los datos en lugar de los valores originales. El coeficiente de correlación de Spearman (ρ) también varía entre -1 y 1 (Field, 2013). Esta medida es útil en casos donde los datos no cumplen con los supuestos requeridos para la correlación de Pearson.

2.5.1.3.2 Correlación Cramer's V

Cramer's V es una medida de asociación que se utiliza para evaluar la relación entre dos variables categóricas. Esta medida se basa en tablas de contingencia y varía de 0 a 1, donde 0 indica que no hay asociación entre las variables y 1 indica una asociación perfecta (Vaske et al., 2017). Cramer's V es útil para complementar el análisis de correlación en contextos en los que se manejan datos categóricos.

2.5.1.3.3 Correlación de Pearson

La correlación de Pearson es una medida que cuantifica la relación lineal entre dos variables continuas. El coeficiente de correlación de Pearson (r) varía de -1 a 1, donde un valor cercano a 1 indica una fuerte correlación positiva, mientras que un valor cercano a -1 indica una fuerte correlación negativa. Un valor de r de 0 sugiere que no hay relación lineal entre las variables (Field, 2013).

2.5.2 Minería de datos

La minería de datos es una disciplina clave en la inteligencia de negocios, que permite transformar grandes volúmenes de datos en conocimientos accionables. Kotu y Deshpande (2015) la definen como "el proceso de descubrir patrones interesantes y conocimientos a partir de grandes cantidades de datos" (p. 2), lo cual se alinea estrechamente con los objetivos del proyecto establecidos en la **Sección 1.4**. Chapman et al. (2000) proponen la metodología *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) para realizar el proceso de minería de datos.

Dicha metodología proporciona un marco estructurado para la implementación de proyectos de minería de datos. Esta se relaciona directamente con el proceso *ETL* discutido en la **Sección 2.4.3** y con el ciclo de vida del análisis de datos presentado en la **Sección 2.5**.

2.5.2.1 Modelos

En el ámbito de la minería de datos, los modelos representan abstracciones de patrones, relaciones o estructuras subyacentes en los datos. Estos modelos se construyen a partir de algoritmos y técnicas estadísticas con el objetivo de descubrir patrones significativos y predecir tendencias futuras (Kotu & Deshpande, 2015). Algunas técnicas específicas de minería de datos, como el *clustering* y el análisis de series temporales, tienen aplicaciones concretas en este proyecto:

2.5.2.1.1 Técnica de clustering K-means

Este algoritmo de *clustering* permite segmentar clientes hoteleros basándose en patrones de reserva y comportamiento, agrupar hoteles según características similares de rendimiento y detectar patrones estacionales en la ocupación hotelera. Esta técnica se relaciona con el modelo dimensional descrito en la **Sección 2.4.5.2**, ya que los resultados se incorporarían como nuevas dimensiones o atributos en el *data mart*.

2.5.2.1.2 Series Temporales

El análisis de series temporales es relevante en el sector hotelero, donde los *KPIs* presentan una marcada componente temporal. Esta técnica se vincula con el análisis descriptivo, diagnóstico y predictivo discutido en la **Sección 2.5.1**, y con la estructura del *data mart*, particularmente en la dimensión temporal del esquema estrella. Además, contribuye a la creación de cuadros de mando interactivos que facilitan visualizaciones dinámicas de tendencias temporales en *KPIs* clave.

2.5.3 Visualización de datos

La visualización de datos es un componente esencial en la analítica y la toma de decisiones basada en datos. Davenport y Harris (2017) destacan que una visualización efectiva es fundamental para comunicar *insights* y facilitar la toma de decisiones.

2.5.3.1 Mejores prácticas de visualización

A continuación, se muestra la Tabla 6 con las mejores prácticas para la visualización de los datos, publicadas por Tableau en 2014.

Tabla 6. *Mejores prácticas para la visualización de datos*

Práctica recomendada	Descripción
Definir el Propósito	Hay que asegurar que la visualización tenga un propósito claro y selectivo, planteándose preguntas como "¿Qué intenta decir?" y "¿Quién es su audiencia?"
Título Claro	Incluir un título simple, informativo y atractivo que explique el propósito de la visualización.
Comprensión Rápida	La visualización debe ser comprensible en 30 segundos sin información adicional.
Organización Eficaz	Organizar los datos de forma que resalten los elementos relevantes en los ejes X e Y.
Evitar sobrecarga	Limitar la cantidad de información presentada para no abrumar a los espectadores.
Uso Limitado de Colores	Utilizar un número reducido de colores y formas en cada vista para facilitar la identificación de patrones.
Orientación Intuitiva	Alinear la orientación de las vistas de forma intuitiva y adecuada a la percepción de los espectadores.
Fuentes Legibles	Limitarse al uso de tres tipos distintos de fuentes.
Subtítulos Guiadores	Incluir subtítulos que guíen a los espectadores a través del contenido.
Revisión de Colores	Verificar que los colores empleados en el panel de control mantengan una paleta de

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Práctica recomendada	Descripción
	entre 7 y 10 colores.
Etiquetas Claras	Disponer de etiquetas claras y ubicarlas de forma que orienten al espectador.

Nota. Adaptado de Tableau (2014)

2.5.3.2 Tipos de gráficos

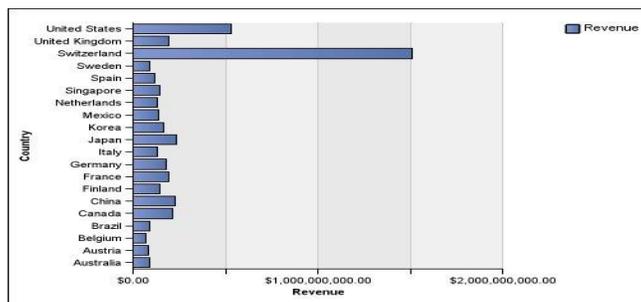
La guía de Tableau (2014) sugiere utilizar distintos tipos de gráficos según las características de los datos y el contexto en el que se presente la información.

2.5.3.2.1 Gráficos de Barras

Los gráficos de barras se utilizan para comparar diferentes categorías de datos, con barras que representan la magnitud de cada categoría. Según Malik (2005), los gráficos de barras son uno de los tipos de visualización de mayor adopción y efectividad.

A continuación se muestra la Figura 8 con un ejemplo de un gráfico de barras.

Figura 8. *Ejemplo gráfico de barras*



Nota. Tomado de IBM (2024).

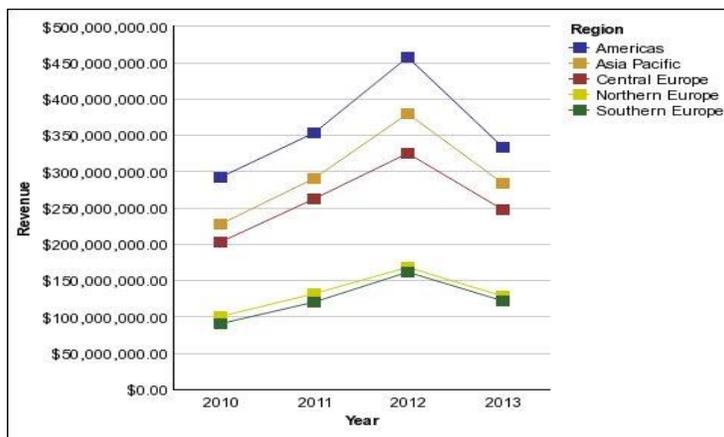
2.5.3.2.2 Gráficos de Líneas

Este tipo de gráfico es ideal para mostrar tendencias y patrones en datos históricos o series temporales. Se representan con una línea que conecta puntos de datos en un gráfico, lo que permite visualizar tendencias, picos o caídas en los datos. Malik (2005) sugiere los gráficos de líneas como una excelente manera de visualizar datos a lo largo del tiempo.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

A continuación se muestra la Figura 9 con un ejemplo de un gráfico de líneas.

Figura 9. Ejemplo de gráfico de líneas



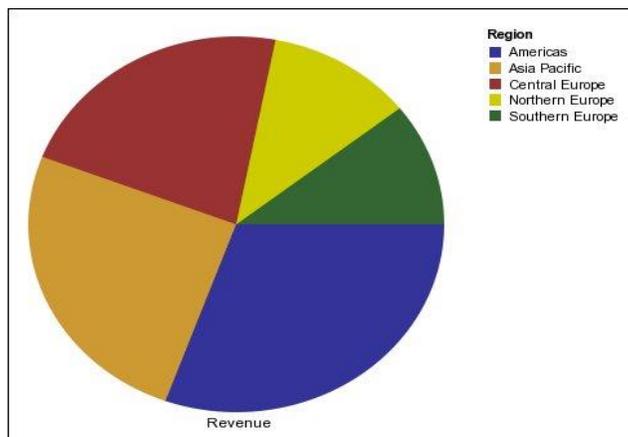
Nota. Tomado de IBM (2024).

2.5.3.2.3 Gráficos Circulares

Los gráficos circulares muestran la proporción de cada categoría dentro de un total, representando los datos como "rebanadas" de un círculo. Malik (2005) menciona los gráficos circulares como una forma de mostrar proporciones o partes de un todo.

A continuación se muestra la Figura 10 con un ejemplo de un gráfico circular.

Figura 10. Ejemplo de gráfico circular



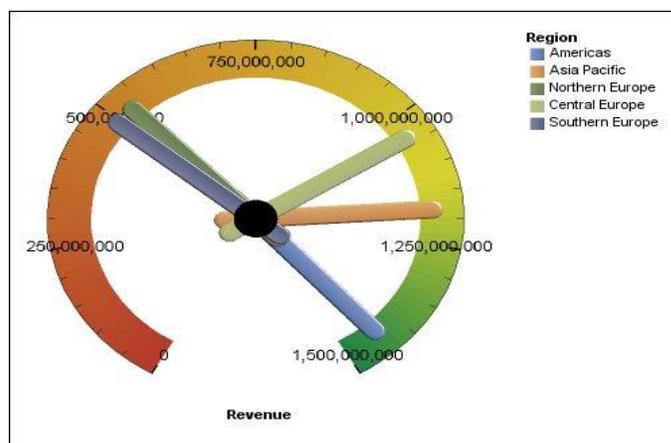
Nota. Tomado de IBM (2024).

2.5.3.2.4 Indicadores

Este tipo de gráfico se asemeja a los velocímetros de los autos y permite a los usuarios ver si un *KPI* está en el rango deseado o si necesita atención. Según Malik (2005), los indicadores visuales como medidores o diales son útiles para representar una sola métrica clave en relación con un valor objetivo.

A continuación se muestra la Figura 11 con un ejemplo de un gráfico de indicadores.

Figura 11. Ejemplo de un gráfico de indicadores



Nota. Tomado de IBM (2024).

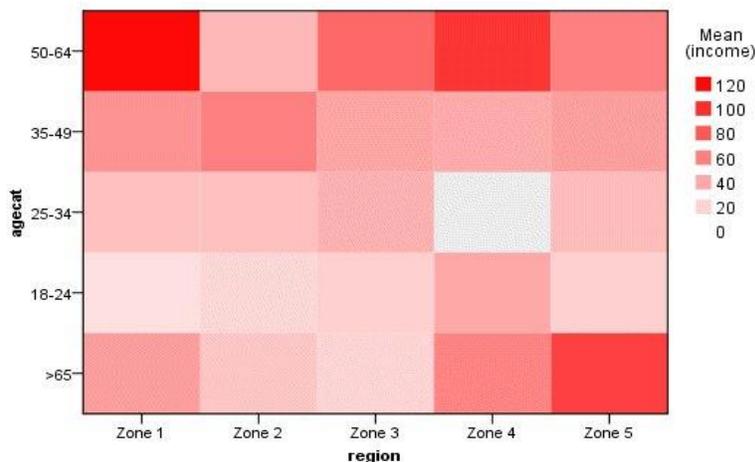
2.5.3.2.5 Mapas de Calor

Estos gráficos utilizan colores para representar valores de una métrica, facilitando la detección de áreas que requieren atención. Según Malik (2005), los mapas de calor resultan una herramienta poderosa para identificar patrones y anomalías en grandes conjuntos de datos.

A continuación se muestra la Figura 12 con un ejemplo de un mapa de calor.

Figura 12. Ejemplo de Mapa de Calor

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM



Nota. Tomado de IBM (2024).

2.5.3.2.6 Tablas

Aunque no es un gráfico en sí, las tablas de datos son importantes para ciertos usuarios que requieren ver detalles numéricos en un formato estructurado. Las tablas sirven como un complemento a los gráficos visuales, especialmente cuando se necesitan datos exactos o cuando es necesario mostrar grandes cantidades de información según Malik (2005).

2.5.3.3 Dashboards

Un *dashboard* (tablero de mando) es una herramienta visual que presenta información crítica de manera concisa y en tiempo real. Está diseñado para ofrecer una vista rápida y clara del estado actual de las operaciones, proyectos o métricas clave de una organización, facilitando la toma de decisiones. Según Malik (2005), estos son un instrumento que permite a las empresas mapear los indicadores de rendimiento clave (*KPIs*) que son cruciales para el éxito de la organización.

Los *dashboards* suelen incluir gráficos, tablas y visualizaciones personalizadas para mostrar la información de manera simple y directa. Según Malik (2005), estos deben tener una interfaz clara, visualmente intuitiva, que permita identificar rápidamente áreas de oportunidad o problemas.

Tipos de dashboards

Malik (2005) clasifica los *dashboards* en tres categorías principales, basadas en su uso y audiencia:

- *Dashboards* Operacionales: se enfocan en las operaciones diarias de la empresa y se actualizan en tiempo real o con alta frecuencia. Ejemplos incluyen indicadores de productividad, tiempos de respuesta, o estados de inventario.
- *Dashboards* Tácticos: estos se enfocan en la supervisión de métricas relacionadas con la ejecución de planes a corto o mediano plazo. Están diseñados para apoyar a los gerentes intermedios o jefes de departamento, quienes necesitan información sobre proyectos o iniciativas estratégicas en curso.
- *Dashboards* Estratégicos: son utilizados por los ejecutivos para monitorear el rendimiento general de la organización en relación con los objetivos estratégicos a largo plazo. Sirven para brindar una vista global del estado de la empresa y para apoyar la toma de decisiones estratégicas.

2.6 Cultura *data-driven*

Una cultura basada en datos es fundamental para que las organizaciones aprovechen el valor de la información generada por sus sistemas. Según Nguyen et al. (2020), crear una cultura *data-driven* implica desarrollar capacidades analíticas y promover el cambio organizacional.

Gartner (2021) enfatiza que los líderes de datos y análisis deben adoptar esta cultura para maximizar el valor de los datos, fomentando la toma de decisiones basada en evidencia y la integración del análisis en los procesos de negocio.

IBM (2020) destaca en su informe “The data-driven organization: Unlocking greater value from data” que las organizaciones *data-driven* están mejor posicionadas para explotar el valor de sus datos, lo cual implica no solo tecnología, sino también cambios en mentalidades y procesos.

2.7 Modelos y Metodologías

Para mejorar el módulo de Inteligencia de Negocios de Procom es necesario fundamentarse en modelos y metodologías que guíen el proceso de manejo de datos a implementar. Según Kotu y Deshpande (2015), un modelo en este contexto de análisis se define como "una representación matemática o lógica de un sistema o proceso que se utiliza para hacer predicciones o tomar decisiones" (p. 15).

La Data Management Association International (2017) describe una metodología como un enfoque para realizar una tarea o resolver un problema que incluye un conjunto

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

de técnicas, procedimientos y documentación.

La diferencia principal entre un modelo y una metodología radica en que un modelo constituye una representación o estructura que facilita la comprensión o predicción de un fenómeno, mientras que una metodología representa un enfoque sistemático para ejecutar un proceso o proyecto.

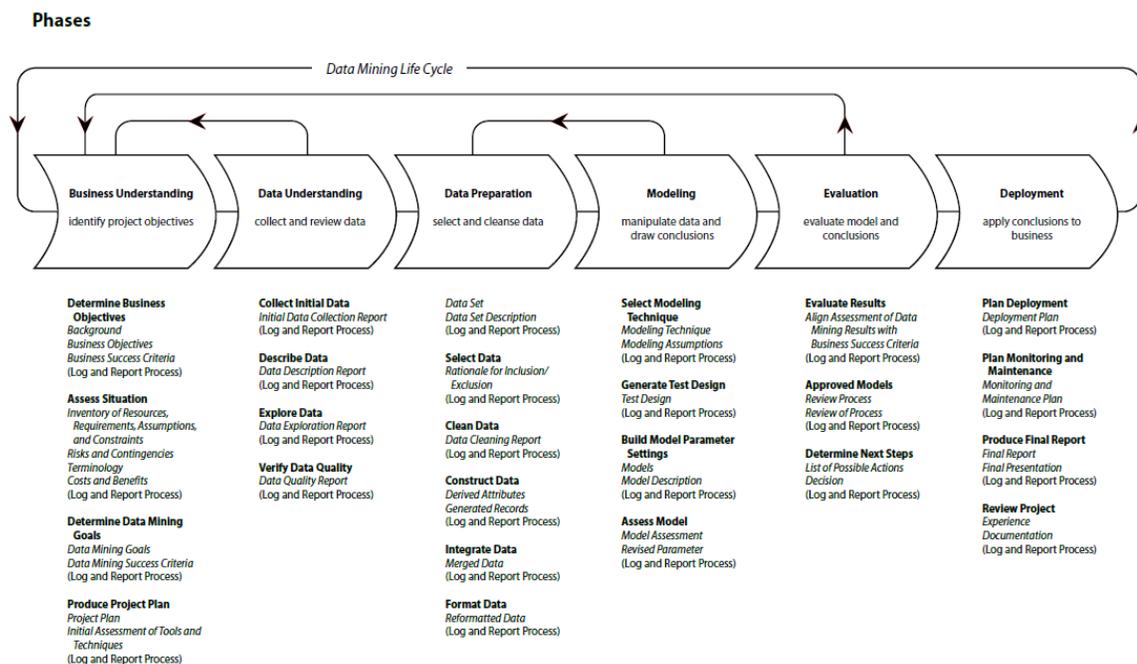
2.7.1 Metodología CRISP-DM

La metodología *Cross-Industry Standard Process for Data Mining (CRISP-DM)* constituye un enfoque estructurado y ampliamente utilizado para proyectos de minería de datos y análisis predictivo.

Chapman et al. (2000) la definen como "un modelo de proceso jerárquico, que consiste en un conjunto de tareas descritas en cuatro niveles de abstracción (de lo general a lo específico): fases, tareas genéricas, tareas especializadas e instancias de proceso" (p. 10).

En la Figura 13 se presentan las fases del proceso CRISP-DM.

Figura 13. *Guía de la metodología CRISP-DM*



Nota. De "CRISP-DM 1.0: Step by step data mining guide" por P. Chapman et al., 2000, p. 10.

2.7.1.1 Fases

La metodología *CRISP-DM* consta de seis fases principales:

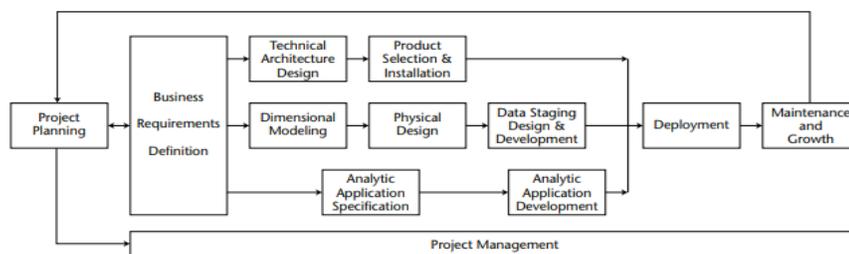
- **Comprensión del Negocio:** Identifica los objetivos empresariales y los transforma en un problema de minería de datos.
- **Comprensión de los Datos:** Inicia con la recolección inicial de datos y continúa con actividades para familiarizarse con estos, identificar problemas de calidad y descubrir hallazgos preliminares.
- **Preparación de los Datos:** Abarca las actividades necesarias para construir el conjunto de datos final a partir de los datos brutos iniciales.
- **Modelado:** Implica la selección y aplicación de diversas técnicas de modelado, así como la calibración de sus parámetros a valores óptimos.
- **Evaluación:** Comprende la evaluación del modelo y la revisión de los pasos ejecutados para su creación, asegurando el cumplimiento adecuado de los objetivos empresariales.
- **Despliegue:** Organiza y presenta el conocimiento obtenido de manera que la empresa genere valor para sus clientes.

2.7.2 Metodología Kimball

Kimball y Ross (2013) definen esta metodología como "un enfoque de abajo hacia arriba para el diseño de almacenes de datos que se centra en la identificación de las medidas de negocio clave y construye el almacén de datos para responder a estas medidas" (p. 3).

En la Figura 14 se presenta el ciclo de vida de la metodología de Kimball.

Figura 14. *Ciclo de vida de Kimball*



Nota. De "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling" por R. Kimball y M. Ross, 2013, p. 3.

2.7.2.1 Fases de Kimball

- Planificación del Proyecto: Define el alcance, justificación y recursos necesarios para el proyecto.
- Definición de Requerimientos del Negocio: Identifica las necesidades de información de los usuarios empresariales.
- Diseño de la Arquitectura Técnica: Establece el marco técnico para el ambiente del almacén de datos.
- Selección e Instalación del Producto: Evalúa y selecciona las herramientas de software y hardware requeridas.
- Modelado Dimensional: Crea los modelos dimensionales para representar los procesos de negocio.
- Diseño Físico: Traduce el modelo lógico a estructuras físicas de almacenamiento.
- Diseño y Desarrollo de la Presentación de Datos: Incluye procesos de Extracción, Transformación y Carga (*ETL*).
- Especificación de Aplicaciones de *BI*: Define las aplicaciones analíticas para uso de los usuarios.
- Desarrollo de Aplicaciones de *BI*: Construye las aplicaciones analíticas.
- Implementación: Integra los componentes técnicos, datos y aplicaciones de usuario final.
- Mantenimiento y Crecimiento: Proporciona soporte continuo y capacitación a los usuarios.

2.7.3 Metodología Inmon

Inmon (2005) define esta metodología como un enfoque descendente (*top-down*), que inicia con un modelo de datos empresarial completo y deriva posteriormente almacenes de datos departamentales (*data marts*) específicos de este almacén centralizado.

En la Figura 15 se ilustra el ciclo de vida de la metodología de Inmon.

Figura 15. Ciclo de vida de Inmon



Nota. De "Building the Data Warehouse" por W. H. Inmon, 2005.

2.7.3.1 Fases de Inmon

- Desarrollo del Modelo de Datos Empresarial: Implica la creación de un modelo de datos normalizado que representa la información relevante para la empresa.
- Análisis del Área de Interés: Identifica y analiza las áreas específicas del negocio que abarcará el almacén de datos (*data warehouse*).
- Diseño del Almacén de Datos: Desarrolla la estructura del almacén de datos central basándose en el modelo empresarial y el análisis de áreas de interés.
- Implementación del Almacén de Datos: Contempla la construcción del almacén de datos diseñado en la fase anterior.
- Creación de *Data Marts* Dependientes: Establece almacenes departamentales o funcionales que derivan sus datos del almacén central.

2.7.4 Comparación entre Kimball e Inmon

La Tabla 7 presenta una comparación entre las metodologías de Kimball e Inmon.

Tabla 7. Comparación de metodologías Inmon y Kimball

Metodología	Inmon	Kimball
Enfoque	<i>Top-down</i>	<i>Bottom-up</i>
Alcance	Toda la empresa	Procesos de negocio específicos
Uso principal	Organizaciones con necesidades de integración de datos corporativos	Empresas que requieren resultados inmediatos en áreas específicas

Nota. Elaboración Propia (2024)

2.7.5 Priorización de requerimientos MoSCoW

Según el Project Management Institute (2017), la técnica MoSCoW establece un método de priorización que categoriza los requisitos en:

- Must have (Debe tener): Requisitos fundamentales
- Should have (Debería tener): Requisitos importantes, pero no críticos
- Could have (Podría tener): Requisitos deseables
- Won't have this time (No tendrá esta vez): Requisitos pospuestos

2.8 Indicadores Clave de Desempeño (KPIs) Hoteleros

La Cámara Costarricense de Hoteles define los *KPI* como métricas que evalúan la efectividad de una organización en el logro de sus objetivos estratégicos y operacionales. En el contexto hotelero, estos indicadores resultan esenciales para medir aspectos críticos como la rentabilidad, la eficiencia operativa y la satisfacción del cliente. Los *KPI* facilitan a los gerentes hoteleros la toma de decisiones fundamentada en datos, la optimización de recursos y el mejoramiento de la experiencia del huésped.

La identificación de los indicadores clave de desempeño en la industria hotelera se fundamenta en diversas fuentes académicas y sectoriales:

- El estudio de Srivastava y Maitra (2016) sobre KPI en hoteles de cinco estrellas
- La investigación de Millán-García y Gómez-Díaz (2018) acerca de factores e indicadores de competitividad hotelera.
- La guía de la Cámara Costarricense de Hoteles (2021) sobre prácticas para la selección de KPI.
- El informe de Failte Ireland (2023) sobre indicadores de desempeño hotelero.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- El análisis de Rondi et al. (2019) sobre los principales indicadores de desempeño en la actividad hotelera.

En la Tabla 8 se presentan los indicadores clave investigados para el sector hotelero.

Tabla 8. *Indicadores Clave de Desempeño para el Sector Hotelero*

Indicador	Descripción	Fórmula o método
RevPAR (Revenue per Available)	Ingresos generados por habitación disponible.	$\text{RevPAR} = \frac{\text{Ingresos totales por habitaciones}}{\text{Número total de habitaciones disponibles}}$ $\text{RevPAR} = \text{ADR} * \text{Tasa de ocupación}$
Tasa de Ocupación	Porcentaje de habitaciones ocupadas con relación al total disponible.	$\text{Tasa de Ocupación} = \left(\frac{\text{Número de habitaciones ocupadas}}{\text{Número total de habitaciones disponibles}} \right) * 100$
ADR (Average Daily Rate)	Ingreso promedio por habitación ocupada.	$\text{ADR} = \frac{\text{Ingresos totales por habitaciones}}{\text{Número de habitaciones vendidas}}$
Costo por Habitación Ocupada	Costo asociado a cada habitación ocupada.	$\text{Costo por Habitación Ocupada} = \frac{\text{Costos totales de habitaciones}}{\text{Número de habitaciones ocupadas}}$
Ingreso por Cliente (RPC)	Total, de ingresos generados por cada cliente durante su estadía.	$\text{RPC} = \frac{\text{Ingresos totales}}{\text{Número total de clientes}}$
Índice de Satisfacción del Cliente	Mide la satisfacción general de los huéspedes.	Medida compuesta basada en encuestas.
Tasa de Retención de Clientes	Porcentaje de clientes que regresan.	$\text{Tasa de Retención} = \left(\frac{\text{Número de clientes que repiten}}{\text{Número total de clientes}} \right) * 100$
Ingreso por Persona	Total, de ingresos dividido por el número de huéspedes.	$\text{Ingreso por Persona} = \frac{\text{Ingresos totales}}{\text{Número total de huéspedes}}$
Tiempos de Respuesta de Servicio	Tiempo promedio en atender solicitudes de huéspedes.	Medida de tiempo
Costos de Personal	Gastos asociados a la nómina.	Suma de todos los gastos relacionados con el personal

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Indicador	Descripción	Fórmula o método
TRevPAR (Total Revenue Per Available Room)	Ingresos totales generados por habitación disponible.	$TRevPAR = \text{Ingresos totales} / \text{Número total de habitaciones disponibles}$
Número de Habitaciones Vendidas	Cantidad total de habitaciones alquiladas en un periodo.	Conteo directo
Costo de Adquisición de Clientes (CAC)	Costo total de adquirir un nuevo cliente.	$CAC = \text{Gastos totales en marketing y ventas} / \text{Número de nuevos clientes adquiridos}$
Tasa de Conversión	Porcentaje de consultas que se convierten en reservas.	$\text{Tasa de Conversión} = (\text{Número de reservas realizadas} / \text{Número total de consultas o visitas}) * 100$
GOP (Gross Operating Profit)	Resultado operativo bruto (ingresos totales menos gastos operativos).	$GOP = \text{Ingresos totales} - \text{Gastos operativos totales}$
GopPAR (Gross Operating Profit per Available Room)	Beneficio bruto por habitación disponible.	$GopPAR = GOP / \text{Número total de habitaciones disponibles}$
EBITDA	Ganancias antes de intereses, impuestos, depreciación y amortización.	$EBITDA = \text{Ingresos} - \text{Gastos} + \text{Intereses} + \text{Impuestos} + \text{Depreciación} + \text{Amortización}$
Margen de Ganancia Bruto	Ingresos totales menos costos directos.	$\text{Margen de Ganancia Bruto} = ((\text{Ingresos totales} - \text{Costo de bienes vendidos}) / \text{Ingresos totales}) * 100$
Rentabilidad del Departamento	Porcentaje de ganancias respecto al ingreso total del departamento.	$\text{Rentabilidad del Departamento} = (\text{Ingresos del departamento} - \text{Costos del departamento}) / \text{Ingresos del departamento}$
Costos Indirectos	Porcentaje de costos indirectos respecto a los ingresos totales.	Suma de todos los costos que no están directamente relacionados con la prestación del servicio.
Días de Ocupación	Total, de días en que las habitaciones están ocupadas en un periodo.	$\text{Días de Ocupación} = \text{Número total de noches vendidas} / \text{Número de habitaciones disponibles}$
Tasa de Reclamaciones	Porcentaje de huéspedes que presentan quejas durante su estancia.	$\text{Tasa de Reclamaciones} = (\text{Número de reclamaciones} / \text{Número total de estancias}) * 100$

Nota. Elaboración Propia (2024).

2.9 Análisis de riesgos

El análisis de riesgos constituye el proceso sistemático de identificación, evaluación y priorización de los riesgos que afectan el éxito de un proyecto. Según el Project Management Institute (2017), este proceso implica un enfoque metodológico para comprender la naturaleza de los riesgos y su posible impacto en los objetivos del proyecto.

El análisis de riesgos resulta fundamental en proyectos de minería de datos, dado que estos implican la extracción de patrones y conocimientos a partir de grandes volúmenes de datos, lo que conlleva incertidumbres y complejidades inherentes (Jones y Smith, 2023).

2.9.1 Identificación de riesgos

La identificación de riesgos comprende el proceso de reconocimiento y documentación de los riesgos potenciales que afectan el éxito de un proyecto u organización. Este proceso resulta fundamental en la gestión de riesgos, ya que permite anticipar amenazas y oportunidades antes de su materialización. Según el Project Management Institute (PMI, 2017), la identificación de riesgos implica un enfoque sistemático que abarca diversas fuentes, desde factores internos como limitaciones de recursos hasta externos como cambios en el entorno económico o regulatorio.

Las técnicas de identificación de riesgos incluyen la revisión de documentación, entrevistas con expertos, sesiones de lluvia de ideas y análisis FODA. El objetivo consiste en generar una lista de posibles riesgos para las fases posteriores del análisis y la gestión (Ñaupás et al., 2013).

En el contexto de la inteligencia de negocios definida en la **Sección 2.2**, resulta fundamental identificar riesgos que comprometan la calidad de los reportes, como la calidad de los datos disponibles. Los datos inexactos o incompletos afectan la fiabilidad de las recomendaciones resultantes.

2.9.2 Evaluación del riesgo

La evaluación del riesgo constituye el proceso de análisis e interpretación de los riesgos identificados en un proyecto, con el fin de determinar su impacto potencial y probabilidad de ocurrencia. Según el Project Management Institute (2021), este proceso implica el uso de métodos cualitativos y cuantitativos para evaluar y priorizar los riesgos según su severidad y la capacidad de respuesta. La evaluación permite a los gerentes de proyecto tomar decisiones fundamentadas sobre el abordaje de los riesgos, priorizando aquellos con mayor efecto potencial en el logro de los objetivos.

Según el Project Management Institute (2021), los niveles de probabilidad se clasifican en:

- Alta (3): Representa una probabilidad del 80 % al 100 % de ocurrencia del riesgo, lo que requiere acciones inmediatas de mitigación.
- Media (2): Indica una probabilidad del 30 % al 79 % de materialización del riesgo, lo que exige atención significativa.
- Baja (1): Señala una probabilidad del 0 % al 29 % de ocurrencia del riesgo, lo que requiere monitoreo continuo.

El impacto de un riesgo se refiere a las consecuencias negativas de su materialización. Estas generalmente son de naturaleza financiera, temporal, de alcance, de calidad o reputacional (PMI, 2021). Los niveles de impacto se definen como:

- Alto (3): Impacto significativo que ocasiona retrasos importantes o daños financieros considerables.
- Medio (2): Impacto moderado que genera retrasos menores o costos adicionales limitados.
- Bajo (1): Impacto mínimo que causa inconvenientes manejables en la organización.

2.9.3 Análisis de riesgo inherente

El Project Management Institute (2021) define el riesgo inherente como el nivel de riesgo existente en un proyecto antes de aplicar acciones de gestión o control.

El cálculo del riesgo inherente resulta de multiplicar la probabilidad de ocurrencia por el impacto del riesgo, lo que facilita la identificación y priorización de riesgos significativos (Project Management Institute, 2021).

A continuación, en la Tabla 9, se presentan ejemplos de cómo se calcula el riesgo inherente

Tabla 9. *Cálculo de Riesgo Inherente*

Probabilidad	Impacto	Riesgo Inherente
2	3	6
2	3	6
1	3	3
2	3	6
3	3	9
3	2	6
1	3	3
2	3	6
2	3	6
2	3	6
2	2	4

Nota. Adaptado de "Practice Standard for Project Risk Management" por Project Management Institute, 2021.

Para medir el riesgo inherente calculado en la **Tabla 9**, se utiliza un mapa de calor como el de la **Figura 15** para crear una matriz de probabilidad-impacto.

A continuación, se muestra la Tabla 10 con la distribución de valores para facilitar su comprensión.

Tabla 10. *Matriz de probabilidad-impacto*

Probabilidad	Impacto		
	Alto	Medio	Bajo
Alto	9	6	3
Medio	6	4	2
Bajo	3	2	1

Fuente. Elaboración Propia (2024)

2.9.5 Medición de los riesgos

La medición de riesgos constituye el proceso de cuantificación y evaluación de los riesgos identificados en un proyecto u organización. Este proceso utiliza técnicas cualitativas y cuantitativas para determinar la probabilidad de materialización de un riesgo y su impacto en los objetivos del proyecto (Project Management Institute, 2021).

Tras la identificación y evaluación de los riesgos, se procede a implementar medidas para su gestión. Esta etapa resulta crucial en la administración de proyectos, pues implica el desarrollo de estrategias específicas para abordar los riesgos identificados. El Project Management Institute (2017) propone las siguientes estrategias:

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- **Aceptar:** Se selecciona cuando la probabilidad de ocurrencia es baja o cuando el costo de mitigación supera los beneficios potenciales.
- **Mitigar:** Implica la implementación de acciones preventivas que reduzcan la probabilidad de materialización del riesgo o disminuyan su impacto.
- **Transferir:** Consiste en la asignación de la responsabilidad del riesgo a otra entidad.
- **Eliminar:** Se enfoca en la implementación de medidas que supriman la posibilidad de materialización del riesgo.

3 Marco Metodológico

La investigación, según Hernández Sampieri et al. (2014), constituye un proceso crítico, sistemático y empírico aplicado al estudio de un fenómeno o problema. Este marco proporciona la base para abordar interrogantes y desafíos de manera estructurada y rigurosa.

En esta sección se desarrollan estrategias y herramientas para la recolección y análisis de datos, asegurando la obtención de resultados válidos y confiables. La estructura define secciones específicas que respaldan la comprensión de la naturaleza y los objetivos del proyecto definidos en la **Sección 1.4**, abordando aspectos como: el tipo de investigación, enfoque de la investigación, fuentes de información, sujetos de investigación, variables consideradas y el procedimiento metodológico de las etapas del proyecto.

3.1 Tipo de Investigación

Según Pimienta (2012), existen dos tipos de investigación: básica y aplicada. A continuación, se describe cada una y se indica el tipo seleccionado para este proyecto.

La investigación básica persigue la generación de conocimiento, sin considerar directamente sus aplicaciones prácticas. Su propósito consiste en ampliar y profundizar los conocimientos sobre la realidad, centrándose en la construcción de generalizaciones, incluyendo hipótesis, leyes y teorías (Pimienta, 2012, p. 9).

Este tipo de investigación no responde a intereses económicos, sino que surge de la curiosidad y el interés por generar nuevo conocimiento. Se considera básica porque fundamenta la investigación aplicada o tecnológica, resultando esencial para el desarrollo científico (Ñaupas et al., 2013, p. 93).

La investigación aplicada, en cambio, busca la consolidación del conocimiento y su aplicación para incrementar el acervo cultural y científico, así como producir tecnología al servicio de la sociedad (Pimienta, 2012, p. 9). Esta se orienta a resolver problemas relacionados con procesos de producción, distribución, circulación y consumo de bienes y servicios en cualquier actividad humana, principalmente en áreas como industria, infraestructura, comercio, comunicaciones y servicios (Ñaupas, 2014, p. 93).

La implementación de un proceso de análisis de datos basado en CRISP-DM para mejorar el módulo de Inteligencia de Negocios de Procom corresponde a una investigación aplicada.

Pimienta (2017) señala que la investigación aplicada "tiene como finalidad resolver un problema concreto, práctico, de la realidad cotidiana de las empresas" (p. 72). Este proyecto aborda directamente un problema práctico de Procom: la subutilización de datos y la necesidad de mejorar su módulo de Inteligencia de Negocios.

Ñaupas et al. (2018) profundizan en esta definición, indicando que la investigación aplicada "se orienta a resolver problemas de los procesos de producción, distribución,

circulación y consumo de bienes y servicios, de cualquier actividad humana" (p. 136). El proyecto se alinea con esta descripción al abordar el procesamiento de datos para alimentar el módulo de Inteligencia de Negocios.

Adicionalmente, Ñaupas et al. (2018) indican que la investigación aplicada "se caracteriza por buscar la aplicación o utilización de los conocimientos adquiridos, mientras se obtienen otros, después de implementar y sistematizar la práctica basada en investigación" (p. 136). Esto se refleja en el proyecto, donde se aplican conocimientos sobre Inteligencia de Negocios y simultáneamente se generan nuevos conocimientos específicos del contexto empresarial.

Pimienta (2017) también destaca que la investigación aplicada "utiliza los conocimientos obtenidos en las investigaciones básicas, pero no se limita a utilizarlos, sino que busca nuevos conocimientos específicos para aplicaciones prácticas" (p. 72). El proyecto evidencia esto mediante la utilización de conocimientos fundamentales sobre gestión de datos e Inteligencia de Negocios, aplicándolos específicamente para mejorar los procesos de toma de decisiones empresariales.

3.2 Enfoque de la Investigación

Hernández Sampieri et al. (2014) identifican tres enfoques principales para la investigación:

El enfoque cuantitativo sigue un proceso secuencial y probatorio donde cada etapa precede a la siguiente. Parte de una idea que se delimita progresivamente, estableciendo objetivos, preguntas e hipótesis. Después de revisar antecedentes teóricos y definir variables, se traza un plan para probarlas mediante la recolección y análisis estadístico de datos numéricos, para extraer conclusiones sobre las hipótesis planteadas. Se caracteriza por su objetividad, predictibilidad y estructura sistemática.

Por otro lado, presenta flexibilidad en su estructura y no sigue un orden predeterminado. Las preguntas e hipótesis surgen en cualquier momento del proceso de recolección y análisis de datos. Este proceso dinámico se mueve entre los hechos y su interpretación, utilizando datos no numéricos para refinar las preguntas de investigación desde una perspectiva interpretativa.

Finalmente, el enfoque mixto integra procesos sistemáticos, empíricos y críticos que involucran la recolección y análisis de datos tanto cuantitativos como cualitativos. Su objetivo consiste en realizar inferencias a partir de la información recopilada para lograr una comprensión integral del fenómeno estudiado.

La presente investigación se alinea con el enfoque cualitativo descrito por Hernández Sampieri et al. (2014), aunque incorpora elementos cuantitativos en su desarrollo. No se clasifica como una investigación mixta debido a que los resultados esperados no buscan cuantificar el impacto del proyecto en términos numéricos para el cliente hotelero o la empresa.

El estudio se enfoca en comprender los procesos, interpretar contextualmente los

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

datos y generar conocimientos cualitativos que permitan mejorar el módulo de Inteligencia de Negocios existente y formular recomendaciones. Aunque se utilizan herramientas y técnicas asociadas con el análisis cuantitativo, como la construcción de un *data mart* y la elaboración de cuadros de mando, el núcleo de la investigación y sus objetivos principales mantienen una orientación cualitativa.

La interpretación y contextualización de los datos se manifiestan en el análisis de patrones de reportes y la definición de indicadores clave de rendimiento específicos para el ámbito hotelero. Esto demuestra un esfuerzo por comprender los fenómenos desde la perspectiva del sector, en lugar de aplicar métricas generalizadas.

Los objetivos revelan una perspectiva holística, característica del enfoque cualitativo, al abarcar diversos aspectos del negocio hotelero, desde los sistemas de información hasta los indicadores de rendimiento. Esta visión integral facilita una comprensión profunda del fenómeno estudiado.

Un aspecto particularmente alineado con la metodología cualitativa es el énfasis en la validación con los participantes, evidenciado en el objetivo de validar el proceso con los gerentes de Procom. Esta práctica subraya la importancia de incorporar las perspectivas de los actores involucrados, un principio fundamental en la investigación cualitativa.

La generación de recomendaciones basadas en la interpretación de resultados también refleja un proceso inductivo típico del enfoque cualitativo, donde se parte de datos específicos para llegar a conclusiones aplicables al contexto general.

3.3 Diseño de la investigación

Como se estableció en la **Sección 3.2**, el diseño de la investigación con enfoque cualitativo busca comprender fenómenos en su contexto real, explorar experiencias y percepciones, y analizar procesos sociales y organizacionales. En la Tabla 11 se presentan los principales diseños cualitativos.

Tabla 11. *Diseños de investigación cualitativa*

Pregunta de investigación	Diseño, marco o abordaje	Información que proporciona
Preguntas sobre procesos y relaciones entre conceptos que conforman un fenómeno.	Teoría fundamentada	Categorías del proceso o fenómeno y sus vínculos. Teoría que explica el proceso o fenómeno.
Preguntas sobre las características, estructura y funcionamiento de un sistema social (grupo, organización, comunidad, subcultura, cultura), desde una familia, hermandad o	Etnográfico	Descripción y explicación de elementos del sistema social: historia, evolución, estructura, interacciones, normas y patrones de conducta

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Pregunta de investigación	Diseño, marco o abordaje	Información que proporciona
hinchada hasta una megaciudad.		
Preguntas orientadas a comprender una sucesión de eventos, a través de las historias o narrativas de quienes la vivieron (experiencias de vida bajo una secuencia cronológica). Eventos como una catástrofe, una elección, la biografía de un individuo, etcétera.	Narrativo	Historias sobre procesos, hechos, eventos y experiencias, siguiendo una línea de tiempo, ensambladas en una narrativa general. Categorías relacionadas con tales historias y narrativa.
Preguntas sobre la esencia de las experiencias: lo que varias personas experimentan en común respecto a un fenómeno o proceso.	Fenomenológico	Experiencias comunes y distintas. Categorías que se presentan frecuentemente en las experiencias.
Preguntas sobre problemáticas o situaciones de un grupo o comunidad (incluyendo cambios).	Investigación-acción	Diagnóstico de problemáticas sociales, políticas, laborales, económicas, etc., de naturaleza colectiva. Categorías sobre las causas y consecuencias de las problemáticas y sus soluciones.

Nota. Adaptado de "Metodología de la investigación" por R. Hernández-Sampieri et al., 2014.

Para esta investigación se selecciona el diseño de investigación-acción, que según Hernández Sampieri et al. (2014), se caracteriza por su orientación práctica y capacidad para generar cambios en el entorno estudiado.

Este diseño resulta apropiado porque se orienta a resolver problemas cotidianos e inmediatos, mejorando prácticas concretas. El proyecto busca implementar un proceso de análisis de datos para optimizar el módulo de Inteligencia de Negocios de Procom, abordando una problemática específica de la organización.

Hernández Sampieri et al. (2014) señalan que la investigación-acción "se centra en aportar información que guíe la toma de decisiones para proyectos, procesos y reformas estructurales" (p. 496). Esta característica se refleja en los objetivos del proyecto, particularmente en la elaboración de cuadros de mando con indicadores de desempeño, que buscan generar recomendaciones fundamentadas y promover una cultura basada en datos.

La investigación-acción implica colaboración entre investigador y participantes en

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

la identificación de necesidades y la implementación de resultados. Este aspecto se evidencia en el primer objetivo específico, que busca analizar los requerimientos de los interesados, involucrando activamente a los miembros de la organización.

Otro elemento distintivo de la investigación-acción es su naturaleza cíclica y reflexiva. Hernández Sampieri et al. (2014) describen este proceso como una espiral de ciclos de planificación, acción, observación y reflexión. En este proyecto, dicho ciclo se manifiesta en la secuencia de objetivos: análisis, diseño e implementación, elaboración de cuadros de mando y validación.

Finalmente, la investigación-acción genera teoría y conocimiento a través de la práctica. El proyecto no solo busca implementar un proceso, sino también validarlo, contribuyendo al conocimiento sobre la aplicación efectiva de la metodología CRISP-DM en contextos empresariales específicos.

3.4 Fuentes de datos e información

Las fuentes de información constituyen los recursos utilizados para obtener datos y conocimientos relevantes para una investigación. Según Hernández Sampieri et al. (2014), estas abarcan todos los medios que permiten acceder a información valiosa y precisa.

Vargas y Ulate (2016) clasifican estas fuentes en dos categorías: primarias y secundarias. A continuación, se presentan las fuentes relevantes según esta clasificación.

3.4.1 Fuentes primarias

Según Hernández Sampieri et al. (2014), las fuentes primarias se refieren a datos originales sin procesar, obtenidos directamente de la realidad o de los sujetos de estudio. Estas fuentes proporcionan información de primera mano a través de la recolección directa de datos.

Tabla 12. *Fuentes primarias de la investigación*

Fuente	Justificación
Documentación del sistema PMS y módulo de BI	La documentación técnica brinda detalles sobre capacidades, funcionalidades, requisitos y limitaciones de las herramientas disponibles.
Libros de investigación sobre la metodología CRIPS-DM	Proporcionan fundamento teórico para implementar el proceso de gestión de datos.
Libros de investigación sobre creación de dashboards	Aportan marco teórico actualizado sobre conceptos, tecnologías, tendencias y metodologías para la implementación de tableros de control.
Artículos y libros de	Fundamentan la selección de indicadores

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Fuente	Justificación
investigación sobre los principales KPIs hoteleros	basándose en estudios empíricos y análisis de tendencias de la industria.
Entrevistas con el personal clave de Procom	Proveen percepciones, experiencias y conocimientos originales sobre el análisis de datos.
Talleres con el personal de Procom	Generan datos mediante interacción directa con participantes en actividades de análisis de datos.
Encuestas realizadas al personal de Procom	Recopilan opiniones directas sobre el análisis de datos mediante instrumentos estructurados.

Nota. Elaboración propia (2024)

3.4.2 Fuentes secundarias

Hernández Sampieri et al. (2014) definen las fuentes secundarias como recursos basados en la interpretación y análisis de datos previamente recolectados. Estas fuentes consisten en interpretaciones, análisis o resúmenes de fuentes primarias, proporcionando contexto o perspectivas sobre la información original.

Tabla 13. *Fuentes secundarias de la investigación*

Fuente	Justificación
Trabajos de Graduación con temática similar	Proporcionan información sobre proyectos previos, incluyendo enfoques metodológicos, desafíos encontrados y lecciones aprendidas
Artículos web sobre BI	Brindan información actualizada sobre tendencias, tecnologías y mejores prácticas en el campo
Material de los cursos TI6900 Inteligencia de Negocios y TI9905 Sistemas de Información Empresarial	Aportan fundamento teórico y práctico sobre implementación de soluciones de Inteligencia de Negocios e integración con sistemas empresariales académica.

Nota. Elaboración propia (2024)

3.5 Sujetos de investigación

De acuerdo con el enfoque seleccionado en la **sección 3.2**, resulta fundamental identificar y definir los sujetos de investigación, que constituyen los individuos directamente involucrados o que proporcionan datos primarios. Bernal (2010) destaca que el sujeto de investigación es quien proporciona información sobre los problemas y necesidades por resolver en un estudio, facilitando el desarrollo de soluciones efectivas. Estos sujetos desempeñan un papel crucial en la recolección de información para comprender la problemática.

En la Tabla 14 se presentan los sujetos seleccionados y su importancia relativa para el proyecto. Esta información resulta fundamental para entender la dinámica del sistema y desarrollar un proceso de gestión de datos que se integre con el módulo de Inteligencia de

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Negocios y satisfaga las necesidades específicas de los involucrados.

Tabla 14. Sujetos de investigación

Rol del sujeto	Caracterización del sujeto	Justificación de la importancia de este sujeto
Gerente General	Responsable de la visión global y la alineación del proyecto con los objetivos empresariales de Procom. Encargado de la toma de decisiones de alto nivel, asignación de recursos y aprobación final de iniciativas	Constituye el sujeto de mayor relevancia al asegurar que el proyecto cumpla las expectativas y genere valor para Procom.
Gerente de Operaciones	Coordina los equipos involucrados, asegurando el cumplimiento de objetivos dentro de plazos y presupuestos establecidos. Identifica y resuelve cuellos de botella operativos y mantiene informada a la gerencia general.	Facilita la integración entre las nuevas funcionalidades y el módulo de Inteligencia de Negocios existente.
Encargado de BI	Responsable de la gestión del módulo de Inteligencia de Negocios en los sistemas de Procom.	Proporciona información sobre la gestión actual del módulo de Inteligencia de Negocios.
Encargado de soporte de sistemas	Responsable del soporte ante fallos en los sistemas de Procom.	Garantiza la disponibilidad, seguridad y rendimiento de los sistemas.
Encargado de desarrollo tecnológico	Responsable del desarrollo y diseño de bases de datos de los sistemas de Procom.	Facilita la comprensión de herramientas e infraestructura tecnológica de los sistemas.

Nota. Elaboración Propia (2024)

3.6 Variables o categorías de la investigación

A partir de los objetivos específicos del estudio, se definen las variables asociadas a cada uno, lo que facilita la recolección y análisis de la información necesaria. Según Hernández Sampieri et al. (2014), una variable constituye una característica o atributo que tiene la posibilidad de variar o cambiar a lo largo del tiempo o en diferentes situaciones. Las variables permiten cuantificar y analizar los cambios que se producen.

A continuación, se muestra la Tabla 15 con las variables identificadas en los objetivos específicos.

Tabla 15. *Variables de Investigación*

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo de investigación	Nombre de la variable	Definición conceptual
Analizar los patrones de los reportes solicitados por los clientes hoteleros y los sistemas de información que sirven como fuente de datos para la definición de indicadores claves de rendimiento para los hoteles.	Patrones de reportes solicitados	Informes específicos requeridos por clientes del sector hotelero para su toma de decisiones.
	Sistemas de información como fuente de datos	Plataformas tecnológicas que almacenan y procesan datos operativos hoteleros.
	Indicadores clave de rendimiento hoteleros	Métricas que evalúan el desempeño en el sector hotelero.
Construir un <i>data mart</i> , junto con los procesos de extracción, transformación y carga de los datos (<i>ETL</i>) generados, para la presentación de datos depurados y la creación de los modelos analíticos para la medición de los indicadores claves de rendimiento de los hoteles.	Proceso <i>ETL</i>	Procedimientos de extracción, transformación y carga de datos.
	<i>Data mart</i>	Subconjunto especializado del almacén de datos enfocado en un área específica
	Datos depurados	Información procesada y estructurada lista para análisis
	Modelos analíticos	Representaciones matemáticas o computacionales para análisis de datos
Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los <i>KPIs</i> en el módulo de <i>BI</i> , para la generación de recomendaciones sobre los resultados.	Cuadros de mando con visualizaciones	Herramientas visuales que presentan indicadores clave de manera gráfica y concisa.
	Recomendaciones basadas en resultados	Conclusiones accionables derivadas del análisis de datos
Validar el proceso de análisis de datos implementado con los gerentes de Procom que garanticen la entrega de valor a la empresa.	Proceso de análisis implementado	Conjunto de procedimientos y herramientas para gestionar el ciclo de vida de los datos.
	Entrega de valor empresarial	Beneficios tangibles e intangibles generados por la implementación.

Nota. Elaboración propia (2024)

3.7 Técnicas e instrumentos de recolección de datos

A continuación, se describen los instrumentos para recopilar los datos que fundamentan las conclusiones del estudio. La selección de instrumentos depende del objetivo de la investigación y la población objetivo.

3.7.1 Entrevista

La entrevista constituye una técnica de recolección de datos basada en la interacción directa entre entrevistador y entrevistado, que busca obtener información detallada sobre el tema de estudio. Según Hernández et al. (2018), las entrevistas se clasifican en:

- Entrevista estructurada: Sigue un conjunto predeterminado de preguntas en orden específico, aplicadas sin variación a todos los entrevistados. Su estructura rigurosa facilita la comparación de información entre participantes.
- Entrevista semiestructurada: Utiliza un guion base con preguntas centrales, pero permite al entrevistador formular preguntas adicionales para profundizar en las respuestas. Este formato equilibra estructura y flexibilidad.
- Entrevista no estructurada: Prescinde de un guion establecido, permitiendo que la conversación fluya naturalmente. El entrevistador explora temas según las respuestas, obteniendo información relevante, aunque de menor estándar.

Para esta investigación se implementa la entrevista semiestructurada, considerando la necesidad de conocer en detalle aspectos relacionados con el módulo de Inteligencia de Negocios mediante preguntas predefinidas, mientras se mantiene la flexibilidad para profundizar en las respuestas obtenidas.

3.7.2 Revisión documental

La revisión documental constituye un método de investigación que comprende la búsqueda, recolección y análisis de documentos o fuentes de información existentes.

A continuación, se muestra la Tabla 16 con el instrumento y plantillas utilizadas para recopilar la información

Tabla 16. Instrumentos y plantillas de recolección de datos

Instrumento	Plantillas de referencia
Entrevista	Apéndice C
Revisión Documental	Apéndice F

Nota. Elaboración propia (2024)

3.8 Procedimiento metodológico de la Investigación

Esta sección presenta las fases metodológicas para abordar integralmente la estrategia de renovación cultural y mejora de la usabilidad de la plataforma de Inteligencia de Negocios en Procom. Estas fases se alinean con los objetivos específicos propuestos y comprenden un conjunto de tareas y entregables para su consecución.

3.8.1 Diagrama propuesto para las fases del procedimiento metodológico

A continuación, se muestra la Figura 16 con las fases metodológicas propuestas para este proyecto.

Figura 16. *Fases Metodológicas*



Nota. Adaptado de "CRISP-DM 1.0: Step-by-step data mining guide" por P. Chapman et al., 2000, p. 10.

3.8.1.1 Comprensión del negocio

En esta fase se inicia determinando los objetivos de negocio, lo que establece el marco para todo el análisis. Luego, se evalúa la situación actual, inventariando recursos, requisitos, suposiciones y restricciones, así como identificando riesgos y contingencias. Esta fase también incluye la definición de los objetivos específicos de minería de datos, criterios de éxito, patrones de reportes, *KPIs* y la creación de un plan de proyecto.

3.8.1.2 Entendimiento de los datos

En esta fase se identifican y analizan los sistemas de información que servirán como fuentes de datos. Se identifica la calidad y estructura de los datos necesarios para medir los *KPIs*. En esta etapa, se realiza la recopilación y exploración de los datos disponibles. Se inicia con la recolección de datos iniciales, seguido de una descripción de los conjuntos de los datos obtenidos. Se culmina con una verificación de la calidad de los datos, identificando posibles problemas como datos faltantes, inconsistencias o errores que afecten el análisis posterior.

3.8.1.3 Preparación de los datos

Esta fase implica la selección cuidadosa de los datos relevantes para el análisis, justificando la inclusión o exclusión de ciertos conjuntos. Los datos seleccionados se limpian, eliminando errores e inconsistencias. Se construyen nuevos atributos si es necesario, y se abarcan todas las actividades para construir el conjunto final de datos. Incluye el diseño y ejecución de los procesos de extracción, transformación y carga (*ETL*) de datos y la construcción del *data mart*.

3.8.1.4 Modelado

En esta fase se utilizan el *data mart* y los datos depurados para crear modelos analíticos. Incluye la selección de técnicas de modelado apropiadas para los *KPIs* definidos, la generación de modelos y su evaluación inicial. También, implica iteraciones, ajustando los modelos o volviendo a la fase de preparación de datos para mejorar los resultados.

3.8.1.5 Evaluación

Aquí se evalúa a fondo el modelo para asegurar que cumple con los objetivos de negocio. Se revisa el proceso para identificar cualquier tarea importante que haya sido omitida y se decide sobre los próximos pasos del proyecto. Esta fase es crítica para determinar si el modelo está listo para su implementación.

3.8.1.6 Despliegue

La fase final implica la implementación de los resultados en el entorno de negocio. Incluye la creación de cuadros de mando con vistas gráficas en el módulo de BI, la generación de recomendaciones basadas en los resultados, y la puesta en marcha del proceso de análisis de datos. También, los gerentes validarán que los resultados entreguen valor a la empresa.

3.8.2 Cronograma

A continuación, se muestra la Figura 17 con el cronograma planteado para la realización de este proyecto.

Figura 17. Cronograma

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Actividades/Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Analizar los requerimientos de los interesados, así como los sistemas de información que sirven como fuente de datos, para la definición de indicadores de desempeño para la toma de decisiones.															
Fase 1. Comprensión del negocio															
Iniciar entrevistas y encuestas con los stakeholders	X	X													
Documentación de requerimientos y definición de KPIs			X												
Entrega de I AVANCE (corresponde al capítulo I) al Prof. Tutor para revisión			X												
Redactar el Documento de Análisis de Requerimientos, Sistemas de Información.				X											
Redactar el "Catálogo de Indicadores Clave de Desempeño (KPIs)"				X											
Devolución a estudiante del I AVANCE revisado por parte del Prof. Tutor para las correcciones respectivas.				X											
Fase 2. Comprensión de los datos															
Análisis de los datos disponibles			X	X											
Identificación de problemas y soluciones para la calidad de datos					X										
Disenar un proceso de extracción, transformación y carga de los datos generados, para la medición de los indicadores de desempeño de la empresa.															
Fase 3. Preparación de los datos															
Entrega de II AVANCE al Prof. Tutor para su respectiva revisión					X										
Procesamiento y transformación de datos						X	X								
Devolución a estudiante del II AVANCE revisado, por parte del Prof. Tutor para las correcciones respectivas (capítulo III)						X									
Entrega III AVANCE (corresponde a capítulo II: Marco conceptual) al Prof. Tutor para revisión							X								
Segunda reunión con la organización por parte del Prof. Tutor							X								
Limpieza final y creación de nuevas variables								X							
Devolución a estudiante del III AVANCE revisado, por parte del Prof. Tutor para las correcciones respectivas (capítulo II)								X							
Fase 4. Modelado															
Diseño del proceso ETL									X						
Entrega de IV AVANCE (corresponde a Capítulo IV: Resultados, y Capítulo V: Propuesta de Solución) al Prof. Tutor para revisión									X						
Implementación de algoritmos de análisis										X					
Redactar el "Diseño del Proceso ETL con los diagramas de flujo, reglas de transformación y limpieza de datos y las especificaciones técnicas"										X					
Elaborar al menos dos cuadros de mando con vista gráfica de los indicadores de desempeño en el módulo de BI para el aprovechamiento de los datos para la toma de decisiones.															
Fase 5. Implementación															
Devolución a estudiante del IV AVANCE (Capítulo IV: Resultados, y Capítulo V: Propuesta de Solución) revisado, por parte del Prof. Tutor para las correcciones respectivas										X					
Implementación en el módulo BI y creación de cuadros de mando											X	X			
Entrega de V AVANCE (corresponde a Capítulo VI: Conclusiones y Recomendaciones) al Prof. Tutor para revisión											X				
Validar el proceso de manejo de datos implementado para el aseguramiento de la entrega de valor a la empresa.															
Fase 6. Evaluación															
Validación del proceso de manejo de datos														X	
Devolución a estudiante del V AVANCE revisado por parte del Prof. Tutor para las correcciones respectivas														X	
Finalizar el "Informe de Propuesta de Plan Integral"														X	X
Finalización del Trabajo Final de Graduación en la organización															X

Nota. Elaboración propia (2024)

3.9 Operacionalización de las variables o categorías.

Esta sección pretende conectar los elementos de la metodología de investigación. Esto permite validar que las actividades realizadas, los sujetos involucrados y el procedimiento metodológico utilizado son los adecuados para cumplir con los objetivos del proyecto.

En la Tabla 17, se presenta una descripción de las variables que se utilizarán en la investigación y las técnicas que se utilizarán para medirlas.

Tabla 17. *Operacionalización de las variables*

Objetivo específico	Capítulo	Fase de la investigación	Variables de investigación	Indicadores	Instrumentos utilizados
Analizar los patrones de los reportes solicitados por los clientes hoteleros y los sistemas de información que sirven como fuente de datos para la definición de indicadores claves de rendimiento para los hoteles.	Capítulo 4	Comprensión del negocio	Patrones de reportes solicitados	Número de tipos de reportes identificados y documentados	Revisión documental de los tipos de reportes generados.
			Indicadores clave de rendimiento hoteleros	Tipos de <i>KPIs</i> del sector hotelero identificados. Tipos de <i>KPIs</i> del sector hotelero identificados para cada objetivo de minería.	Revisión de literatura y <i>benchmarking</i> del sector hotelero.
		Entendimiento de los datos	Sistemas de información como fuente de datos	Número de fuentes de datos utilizadas para el sector hotelero	Ficha de inventario de sistemas

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo específico	Capítulo	Fase de la investigación	Variables de investigación	Indicadores	Instrumentos utilizados
Construir un data mart, junto con los procesos de extracción, transformación y carga de los datos (<i>ETL</i>) generados, para la presentación de datos depurados y la creación de los modelos analíticos para la medición de los indicadores claves de rendimiento de los hoteles.	Capítulo 5	Preparación de los datos	Proceso <i>ETL</i>	Número de procesos <i>ETL</i> diseñados y documentados	Documentación de procesos ETL a través de diccionarios
		Preparación de los Datos	<i>Data mart</i>	Número de tablas y columnas incluidas en el <i>data mart</i>	Diagrama del <i>data mart</i>
		Modelado	Datos depurados	Porcentaje de campos en el conjunto de datos que contienen valores válidos y utilizables para el análisis, después del proceso de depuración.	Análisis exploratorio y preparación de los datos con el lenguaje de programación R. Análisis estadístico
	Capítulo 5	Modelado Evaluación	Modelos analíticos	Número de modelos analíticos creados. Porcentaje de modelos que cumplieron con los criterios de éxito	Modelado de la información con el lenguaje de programación R.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo específico	Capítulo	Fase de la investigación	Variables de investigación	Indicadores	Instrumentos utilizados
	Capítulo 5			definidos	
Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los <i>KPIs</i> en el módulo de <i>BI</i> , para la generación de recomendaciones sobre los resultados.		Despliegue	Cuadros de mando con visualizaciones	Número de cuadros de mando implementados Número de <i>KPIs</i> incluidos por cuadro de mando	Herramienta de visualización <i>Devexpress Report Designer</i>
			Recomendaciones basadas en resultados	Número de recomendaciones específicas generadas a partir del análisis de datos.	Documentación de los principales hallazgos y su respectiva recomendación

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo específico	Capítulo	Fase de la investigación	Variables de investigación	Indicadores	Instrumentos utilizados
Validar el proceso de análisis de datos implementado con los gerentes de Procom que garanticen la entrega de valor a la empresa.		Despliegue	Proceso de análisis implementado	Número de fases de <i>CRISP-DM</i> implementadas. Porcentaje de pasos del proceso que están documentados. Duración promedio de cada fase del proceso.	Diagrama y documentación del flujo del proceso de análisis de datos implementado con la metodología <i>CRISP-DM</i> incluyendo la respuesta para los indicadores.
			Entrega de valor empresarial	Cantidad de gerentes que aprueban el valor entregado	Documento firmado que apruebe el valor entregado

Nota. Elaboración propia (2024)

3.10 Resumen del procedimiento metodológico de la investigación

A continuación, se muestra la Tabla 18 con el resumen del procedimiento metodológico de la investigación.

Tabla 18. Resumen del procedimiento metodológico de la investigación

Objetivo específico	Marco Conceptual	Metodología	Análisis y propuesta	Conclusiones	Recomendaciones
Analizar los reportes solicitados por los clientes hoteleros, así como los sistemas de información que sirven como fuente de datos, para la definición de indicadores de desempeño para la toma de decisiones estratégicas.	Capítulo 2 Secciones 2.2, 2.3, 2.4, 2.5, 2.7, 2.8, 2.9, 2.12	Capítulo 3 Secciones 3.8.1.1	Capítulo 4 Sección 4.1,4.2	Capítulo 6	Capítulo 7
Diseñar un proceso de extracción, transformación y carga de los datos (ETL) generados, para la medición de los indicadores de desempeño del sector hotelero.	Capítulo 2 Secciones 2.5 y 2.6	Capítulo 3 Secciones 3.8.1.2, 3.8.1.3, 3.8.1.4	Capítulo 5 Secciones 5.1,5.2	Capítulo 6	Capítulo 7

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los <i>KPIs</i> en el módulo de <i>BI</i> , para la generación de recomendaciones sobre los resultados.	Capítulo 2 Sección 2.11	Capítulo 3 Secciones 3.8.1.5	Capítulo 5 Sección 5.3,5.4	Capítulo 6	Capítulo 7
Validar el proceso de análisis de datos implementado con los gerentes de Procom para garantizar la entrega de valor a la empresa.	Capítulo 2 Sección 2.9	Capítulo 3 Secciones 3.8.1.6	Capítulo 5 Sección 5.3,5.4	Capítulo 6	Capítulo 7

Nota. Elaboración propia (2024).

4 Análisis de Resultados

Este capítulo presenta los resultados obtenidos siguiendo la metodología CRISP-DM para la mejora del módulo de Inteligencia de Negocios de Procom. El análisis se estructura en dos fases principales: comprensión del negocio y entendimiento de los datos.

En la fase de comprensión del negocio, se analizan los objetivos empresariales, la situación actual y los requisitos del proyecto. Se evalúan aspectos como recursos humanos, procesos existentes, cultura organizacional y sistemas de información. Además, se realiza un análisis de riesgos para identificar posibles amenazas al proyecto.

En la fase de entendimiento de los datos, se examina la estructura y calidad de la información disponible en las bases de datos del cliente piloto. Se realiza un análisis exploratorio de cada tabla relevante, incluyendo estudios de correlación, distribución y valores atípicos. También se identifican patrones en los datos y se evalúa la calidad de la información mediante el análisis de valores faltantes y consistencia.

Los resultados de este análisis proporcionan la base para las decisiones de diseño e implementación que se detallarán en el **Capítulo 5**.

4.1 Comprensión del negocio

Esta primera fase de CRISP-DM resulta fundamental para alinear los objetivos empresariales con las capacidades técnicas necesarias. Se analiza la situación actual de Procom desde múltiples perspectivas: objetivos estratégicos, recursos disponibles, procesos existentes y riesgos potenciales.

El análisis del contexto empresarial se desarrolló mediante un proceso iterativo y colaborativo con la gerencia de Procom. Las entrevistas con directivos y personal clave, documentadas en los **Apéndices G, H e I**, permitieron una comprensión de las necesidades organizacionales y los desafíos operativos.

4.1.1 Determinación de objetivos empresariales

Las sesiones de trabajo con la gerencia, particularmente las documentadas en el **Apéndice J**, revelaron una visión estratégica enfocada en la transformación de datos en valor tangible para los clientes hoteleros. Los objetivos identificados reflejan una comprensión madura de la necesidad de evolucionar desde la simple generación de reportes hacia un análisis integral que soporte la toma de decisiones.

La validación de estos objetivos con el equipo gerencial permitió establecer criterios de éxito realistas y medibles. Particularmente significativa resulta la priorización de la segmentación de clientes y el análisis de rentabilidad por servicio, reflejando un entendimiento de las necesidades del sector hotelero.

4.1.1.1 Descripción de la empresa

La descripción de la empresa se encuentra detallada en la **Sección 1.2.1**

4.1.1.2 Objetivos de negocio

El análisis de las entrevistas con la gerencia, documentadas en el **Apéndice J**, revela cinco objetivos estratégicos principales:

- Desarrollar recomendaciones fundamentadas en datos mediante tableros de control interactivos y visualizaciones
- Identificar indicadores clave de rendimiento específicos para la industria hotelera
- Identificar patrones y segmentos clave en la base de clientes hotelera
- Determinar tendencias estacionales de habitaciones y servicios
- Identificar servicios de mayor rentabilidad y sus tendencias de consumo

4.1.1.3 Criterios de éxito del negocio

Los criterios de éxito establecidos por Procom para los objetivos definidos en la **Sección 4.1.1.2** comprenden:

- Desarrollo de tableros de control que abarquen áreas clave del sector hotelero
- Generación de recomendaciones operativas para clientes hoteleros
- Incorporación de indicadores clave de rendimiento en cada tablero de control
- Implementación de un modelo de segmentación de clientes
- Identificación de servicios y habitaciones de mayor rentabilidad
- Desarrollo de análisis de rentabilidad por servicio

Los criterios de éxito establecidos reflejan un equilibrio entre ambición y practicidad, enfocándose en entregables concretos como tableros

de control y modelos de segmentación. Esto sugiere un entendimiento realista de las capacidades necesarias para la transformación deseada.

4.1.2 Situación actual

El diagnóstico de la situación actual reveló facetas críticas que impactan la capacidad de Procom para generar valor a partir de sus datos. Las entrevistas con el equipo gerencial, documentadas en los **Apéndices H, I y J**, permitieron una comprensión de sus capacidades analíticas, recursos humanos, procesos y cultura organizacional.

4.1.2.1 Recursos humanos para análisis de datos

La evaluación de recursos humanos, documentada en el **Apéndice H**, evidencia una brecha significativa en capacidades analíticas. La ausencia de personal especializado en análisis de datos, confirmada en múltiples entrevistas con la gerencia, representa una limitación fundamental para la transformación deseada. Esta situación refleja un desafío común en el sector tecnológico, donde según DataAnalytics (2023), el 75% de las empresas enfrentan dificultades similares en la gestión efectiva de datos.

4.1.2.2 Procesos actuales relacionados al análisis de datos

Según lo documentado en el **Apéndice H**, la empresa opera sin procesos formales establecidos para el análisis de datos. La gestión y generación de informes se realiza de manera específica según necesidad, sin una metodología estructurada o procesos estandarizados. Esta situación limita la capacidad organizacional para extraer valor consistente de los datos recopilados.

El módulo de Inteligencia de Negocios actual presenta limitaciones significativas en su funcionalidad. Los informes carecen de visualizaciones e información procesada previamente, evidenciando una brecha entre las capacidades actuales y las necesidades de los clientes.

El enfoque actual se centra en la generación de informes básicos, sin proporcionar recomendaciones para la toma de decisiones estratégicas en el sector hotelero. Esta situación enfatiza la necesidad de fortalecer las capacidades analíticas del módulo para ofrecer información accionable.

4.1.2.3 Cultura de datos en la organización

Según lo documentado en el **Apéndice H**, el personal de Procom no está familiarizado con herramientas avanzadas de análisis de datos. Esta carencia de habilidades analíticas, presente en los diferentes niveles organizacionales, dificulta la implementación de estrategias basadas en

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

datos.

La ausencia de especialistas en análisis y visualización de datos, particularmente para el sector hotelero, constituye una limitación que afecta la capacidad de generar informes útiles para los clientes.

4.1.2.4 Sistemas de información actuales

Los sistemas de información utilizados para la recopilación de datos se definieron en la **Sección 2.3**. La empresa se especializa en el desarrollo de estas soluciones para optimizar las operaciones organizacionales. A continuación, se presenta la Tabla 19, que detalla las soluciones que comercializa, según la documentación proporcionada por Procom (2024):

Tabla 19. *Sistemas de Información de Procom*

Sistema	Descripción
<i>Solaria X</i>	Sistema de gestión empresarial que integra módulos de ventas, facturación, inventario, gestión de relaciones con clientes y gestión financiera. Facilita la digitalización de procesos en diversos sectores.
<i>Solaria Hospitality (PMS)</i>	Solución integral de gestión diseñada para la industria hotelera. Adaptable para hoteles, parques de aventura, establecimientos con restaurante y clubes sociales.
<i>Solaria Food</i>	Plataforma especializada en gestión de servicios de alimentación, enfocada en comedores empresariales. Proporciona herramientas para la gestión operativa y mejora de experiencia del usuario.
<i>Solaria FE</i>	Solución web para gestión financiera y fiscal en Costa Rica. Integra operaciones financieras con otros sistemas, incluyendo módulos de caja, gastos, compras y facturación.
<i>Solaria POS</i>	Sistema de punto de venta basado en la nube. Incluye integración con Sinpe Móvil, comercio electrónico, control de inventario, costos y gestión de tiempo.
<i>SAO CRM</i>	Sistema integrado de gestión de relaciones con clientes y procesos de negocio. Optimiza la gestión de información de clientes, eficiencia operativa y toma de decisiones estratégicas.

Nota. *Adaptado de documentación institucional de Procom (2024).*

El análisis del ecosistema tecnológico de Procom revela una

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

infraestructura robusta de soluciones empresariales. La diversidad de sistemas identificados en la **Tabla 19** demuestra una base tecnológica sólida para la captura de datos, particularmente en el sector hotelero, aunque esta multiplicidad también representa un desafío para la integración y análisis coherente de la información.

4.1.2.5 Reportes generados por el PMS

Según lo documentado en el **Apéndice AE** mediante entrevista con el encargado de soporte de TI, la empresa ha estandarizado la generación de informes en el módulo del Sistema de Gestión Hotelera definido en la **Sección 4.1.2.4**.

En la Tabla 20, se muestran los reportes encontrados:

Tabla 20. *Reportes encontrados en el módulo de BI*

Reporte encontrado	Utilidad	Observaciones
Información de clientes	Presenta datos de clientes registrados, con filtros por origen, duración de estadía, tipo de reservación, frecuencia de visitas, requisitos alimentarios, servicios de interés, ubicación y preferencias gastronómicas.	Muestra información tabular con múltiples campos de análisis por cliente.
Ocupación	Presenta información sobre aforo diario y mensual.	Incluye dos visualizaciones: gráfico circular para aforo diario (Figura 13) y gráfico de barras para aforo mensual (Figura 10). El gráfico circular presenta exceso de etiquetas que dificultan su interpretación.
Ventas por Servicio	Muestra análisis de ventas por servicio con desgloses anual, mensual y diario.	Presenta visualización mediante gráfico circular (Figura 13).

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Reporte encontrado	Utilidad	Observaciones
Ventas por centro de costo	Detalla ventas por centro de costo.	Presenta información tabular limitada a un día específico, sin opciones de filtrado ni visualizaciones adicionales.

Nota. Elaboración Propia (2024).

La evaluación de los reportes existentes, documentada mediante entrevistas con el encargado de soporte de TI (**Apéndice AE**), revela limitaciones críticas en las capacidades analíticas actuales de Procom. Los cuatro tipos de reportes identificados evidencian un enfoque básicamente descriptivo, sin capacidad de generar *insights* accionables para la toma de decisiones.

La presentación visual de estos reportes muestra deficiencias significativas. Los gráficos circulares utilizados para el análisis de ocupación y ventas presentan un exceso de etiquetas que dificultan su interpretación, mientras que la información de clientes y ventas por centro de costo se limita a presentaciones tabulares sin elementos visuales o capacidades de filtrado avanzado.

4.1.2.6 Hardware

El análisis de la infraestructura técnica, documentado en las Tabla 21, revela una configuración que, aunque funcional, presenta limitaciones para el análisis avanzado de datos. Las sesiones de trabajo con el equipo técnico, registradas en el **Apéndice AE**, permitieron identificar que la utilización exclusiva de DevExpress Report Designer como herramienta de visualización impone restricciones significativas en las capacidades analíticas.

Tabla 21. *Hardware utilizado*

Tipo	Descripción
Servidores	<i>On premise</i>
Computadora	HP ZBook 15 G3
RAM	16GB
Capacidad de almacenamiento	256GB

Nota. Elaboración Propia (2024).

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

La infraestructura de hardware disponible para el proyecto, aunque básica, resulta funcional para los objetivos planteados. La configuración de servidores *on-premise* junto con equipos de desarrollo de capacidades moderadas sugiere la necesidad de considerar la eficiencia en el procesamiento de datos durante la implementación de las soluciones analíticas.

4.1.2.7 Herramienta de visualización de datos

En la Tabla 22 se muestra el software que utiliza la empresa para realizar los reportes en su módulo de *BI*.

Tabla 22. Herramienta de visualización

Software	Descripción
<i>Devexpress Report Designer</i>	Herramienta integrada en el módulo de Inteligencia de Negocios para la elaboración de informes. Se conecta directamente con la base de datos Microsoft SQL Server.

Nota. *Adaptado de documentación institucional de Procom (2024).*

La utilización exclusiva de DevExpress Report Designer como herramienta de visualización representa una limitación técnica. Si bien esta herramienta ofrece la ventaja de una integración nativa con la base de datos SQL Server existente, sus capacidades de visualización y análisis resultan restrictivas para los objetivos analíticos avanzados del proyecto.

4.1.2.8 Bases de datos

La empresa ha implementado una base de datos centralizada que almacena la información de los sistemas descritos en la **Sección 4.1.2.4**. Esta se caracteriza por su amplitud y adaptabilidad para atender las necesidades tanto del sector hotelero como de restauración.

El análisis de la infraestructura de datos revela una arquitectura compleja y adaptable, evidenciada en las 101 tablas que componen la base de datos del cliente piloto. La estructura flexible del esquema, aunque permite adaptabilidad a diferentes tipos de clientes hoteleros, presenta desafíos significativos para el análisis sistemático de datos. La descripción del volumen de datos y espacio de almacenamiento se encuentra en el **Apéndice AH**.

En la Tabla 23 se muestra el gestor de base de datos que utiliza la empresa para almacenar la información proveniente de sus sistemas.

La limitación temporal de siete meses en los datos disponibles impone restricciones importantes en la capacidad de identificar patrones estacionales y tendencias a largo plazo. Adicionalmente, la presencia de campos nulos y la variabilidad en el uso de tablas entre clientes sugiere la necesidad de un proceso robusto de limpieza y estandarización de datos. Esta restringe la capacidad de desarrollar modelos predictivos, según se especificó en las exclusiones de la **Sección 1.6**.

4.1.2.9 Requisitos del proyecto

El análisis de requisitos, desarrollado mediante sesiones iterativas con la gerencia y documentado en la Tabla 24, revela ocho requerimientos fundamentales para el éxito del proyecto. La priorización mediante la metodología MoSCoW, realizada en colaboración con el equipo gerencial, establece una clara jerarquía de necesidades donde la visualización efectiva de datos y la generación de recomendaciones accionables emergen como prioridades críticas.

Tabla 24. *Requisitos del proyecto*

RF-001	Visualización con DevExpress Report Designer
Descripción	Implementar DevExpress Report Designer para la creación de informes y visualizaciones
Objetivos de negocio asociados	Desarrollar recomendaciones basadas en datos mediante tableros de control interactivos
Prioridad	Must
RF-002	Desarrollo de tableros interactivos con indicadores clave
Descripción	Crear visualizaciones que muestren indicadores relevantes para la toma de decisiones.
Objetivos de negocio asociados	Desarrollar recomendaciones basadas en datos mediante visualizaciones Identificar indicadores clave de rendimiento específicos para la industria hotelera
Prioridad	Must
RF-003	Presentación clara y comprensible de los análisis en las visualizaciones.
Descripción	Implementar buenas prácticas de visualización de datos en el diseño de informes.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivos de negocio asociados	Identificar indicadores clave de rendimiento hoteleros Determinar tendencias estacionales Identificar servicios de mayor rentabilidad
Prioridad	Must
RF-004	Generación de recomendaciones en servicios, paquetes, reservas y habitaciones basadas en los datos recogidos.
Descripción	Desarrollar algoritmos para analizar datos y generar recomendaciones sobre servicios, paquetes, reservas y habitaciones
Objetivos de negocio asociados	Identificar patrones y segmentos de clientes Determinar tendencias estacionales Identificar servicios de mayor rentabilidad
Prioridad	Must
RF-005	Documentación de los nuevos modelos y análisis implementados
Descripción	Documentar todos los modelos estadísticos, análisis y tableros de control implementados
Objetivo de negocio asociado	Todos
Prioridad	Must
RF-006	Software de análisis estadístico compatible con el gestor de base de datos
Descripción	Seleccionar software de análisis estadístico compatible con SQL Server.
Objetivo de negocio asociado	Desarrollar recomendaciones basadas en datos mediante visualizaciones.
Prioridad	Must
RF-007	Diseño del proceso adaptable a otros segmentos
Descripción	Estructurar modelos y visualizaciones de forma modular para facilitar adaptación a otros sectores.
Objetivo de negocio asociado	Todos

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Prioridad	Should
RF-008	Cuadros de mando con filtros que permitan a los usuarios explorar los datos
Descripción	Implementar filtros avanzados para exploración de datos por diferentes criterios (fechas, tipos de habitación/servicio, clientes)
RF-008	Cuadros de mando con filtros que permitan a los usuarios explorar los datos
Objetivos de negocio asociados	Desarrollar visualizaciones interactivas Identificar patrones de clientes Analizar tendencias estacionales Identificar servicios rentables
Prioridad	Should

Nota. *Elaboración propia (2024).*

El análisis de requisitos mediante la metodología MoSCoW revela una clara priorización de necesidades técnicas y funcionales. Los requisitos clasificados como "Must" enfatizan aspectos críticos como la visualización efectiva de datos, la generación de recomendaciones basadas en evidencia y la documentación del proceso. Esta priorización refleja un entendimiento maduro de las necesidades del proyecto, equilibrando ambición con factibilidad.

4.1.2.10 Análisis de riesgos

El proceso de identificación y evaluación de riesgos se desarrolló mediante un esfuerzo colaborativo extensivo con la gerente de operaciones de Procom, documentado en múltiples sesiones de trabajo (**Apéndice J**). Este análisis se enriqueció significativamente con la revisión de literatura especializada en transformaciones digitales y proyectos de análisis de datos.

4.1.2.10.1 Identificación de riesgos

La identificación inicial de riesgos reveló diez categorías principales, fundamentadas tanto en la experiencia práctica de Procom como en investigaciones académicas. Bilgihan et al. (2021) proporcionaron un marco de referencia específico para el sector hotelero, mientras que Davenport y Harris (2017) contribuyeron con

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

perspectivas sobre transformaciones analíticas empresariales. Las observaciones de Munappy et al. (2020) sobre implementaciones ETL resultaron particularmente relevantes dado el enfoque técnico del proyecto.

Tabla 25. Identificación de riesgos

Identificador	Riesgo
R-001	Calidad de datos insuficiente
R-002	Resistencia al cambio
R-003	Limitaciones tecnológicas
R-004	Cambios en los requisitos
R-005	Problemas de integración de datos
R-006	Incumplimiento de plazos
R-007	Problemas de escalabilidad
R-008	Mal uso o interpretación errónea de los datos
R-009	Falta de adopción por parte de los clientes
R-010	Problemas de comunicación

Nota. Elaboración Propia (2024)

4.1.2.10.2 Evaluación del riesgo

En esta sección, se presenta una evaluación de riesgos basada en la probabilidad de ocurrencia y su posible impacto en la organización. Esta escala está basada en la guía del PMBOK séptima edición.

4.1.2.10.3 Análisis de riesgo inherente

La evaluación de riesgo inherente identifica la integración de datos como la amenaza crítica, con una puntuación máxima de 9. Este hallazgo, validado por la experiencia previa de Procom y respaldado por las observaciones de Hu et al. (2014) sobre integraciones en entornos hoteleros, sugiere la necesidad de una estrategia de mitigación particularmente robusta en esta área.

A continuación, en la Tabla 26, se presentan los riesgos asociados a su resultado de inherencia.

Tabla 26. Resultado de inherencia

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Identificador	Probabilidad	Impacto	Riesgo Inherente
R-001	2	3	6
R-002	2	3	6
R-003	1	3	3
R-004	2	3	6
R-005	3	3	9
R-006	3	2	6
R-007	1	3	3
R-008	2	3	6
R-009	2	3	6
R-010	2	2	4

Nota. Elaboración Propia (2024)

4.1.2.10.4 Mapa de calor de riesgos

Comprendiendo la **Tabla 26**, ahora se presenta la Tabla 27 con los riesgos clasificados según los resultados obtenidos en la sección 4.1.2.11.3

Tabla 27. Mapa de calor de riesgos

Probabilidad	Impacto		
	Alto	Medio	Bajo
Alto	R-005	R-006	
Medio	R-001 R-002 R-004 R-008 R-009 R-010	R-011	
Bajo	R-003 R-007		

Nota. Elaboración Propia (2024)

El mapa de calor resultante reveló una concentración significativa de riesgos en el cuadrante de probabilidad media e impacto alto. La calidad de datos y la resistencia al cambio emergen como amenazas secundarias pero significativas, un patrón que Nguyen et al. (2020) identifican como común en transformaciones basadas en datos. Esta distribución de riesgos sugiere la necesidad de un enfoque particularmente robusto en la gestión de cambio organizacional y la validación de datos.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

4.1.2.10.5 Medidas ante los riesgos

Después de identificar y evaluar los riesgos, se procede a la fase de implementación de medidas para gestionarlos.

A continuación, en la Tabla 28, se plantea la gestión de los riesgos identificados en la sección 4.1.2.11.1

Tabla 28. Riesgos medidos

Identificador	Riesgo inherente	Estrategia	Plan de acción
R-001	6	Mitigar	Implementar procesos de validación y limpieza de datos y establecer estándares de calidad
R-002	6	Mitigar	Desarrollar un plan de comunicación clara, capacitación y participación de los empleados en el proceso.
R-003	3	Aceptar	Documentar las limitaciones y buscar soluciones alternativas dentro de las capacidades actuales. Planificar actualizaciones futuras si es necesario.
R-004	6	Mitigar	Implementar una metodología ágil para manejar cambios y mantener una comunicación constante con los <i>stakeholders</i> .
R-005	9	Eliminar	Realizar un análisis de los sistemas y fuentes de datos, desarrollar un plan de integración.
R-006	6	Mitigar	Implementar una gestión de proyectos, establecer hitos claros y realizar un seguimiento continuo del progreso.
R-007	3	Aceptar	Monitorear el crecimiento y planificar mejoras de infraestructura para el

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Identificador	Riesgo inherente	Estrategia	Plan de acción
			futuro si es necesario.
R-008	6	Mitigar	Proporcionar capacitación sobre el uso e interpretación de datos.
R-009	6	Mitigar	Desarrollar una estrategia de adopción del cliente, incluyendo demostraciones.
R-019	4	Mitigar	Establecer canales de comunicación claros, programar reuniones regulares de equipo y <i>stakeholders</i> .

Nota. Elaboración Propia (2024)

Las estrategias de mitigación propuestas se desarrollaron considerando tanto el contexto específico de Procom como las mejores prácticas documentadas por el Project Management Institute (2017). La decisión de adoptar una estrategia de eliminación para el riesgo de integración de datos refleja su criticidad para el éxito del proyecto. Esta decisión fue respaldada por Brynjolfsson y McElheran (2019), quienes enfatizan la importancia fundamental de la integración de datos en transformaciones digitales exitosas.

La validación final de este análisis con la gerente de operaciones permitió refinar las estrategias de mitigación y establecer indicadores claros para el monitoreo continuo de riesgos. Se desarrollaron planes de contingencia específicos para cada categoría de riesgo, con especial énfasis en aquellos ubicados en los cuadrantes de alto impacto del mapa de calor. Este enfoque estructurado en la gestión de riesgos proporciona una base sólida para la implementación exitosa del proyecto.

4.1.3 Determinación de *KPIs*, objetivos de minería de datos y criterios de éxito

En esta sección se establecen los objetivos específicos de minería de datos y sus criterios de medición.

4.1.3.1 Objetivos de minería de datos y sus criterios de éxito

A continuación, en la Tabla 29 se presentan los objetivos de minería de datos identificados a partir de los requerimientos recopilados de la sección 4.1.2.10.

Tabla 29. *Objetivos de minería y criterios de éxito*

Objetivo	Criterio de éxito
Desarrollar modelo de segmentación de clientes basado en patrones de gasto	Identificar mínimo tres segmentos de clientes con perfiles de comportamiento de compra diferenciados
Analizar rentabilidad y tendencias de consumo en servicios	Identificar los diez servicios con mayor frecuencia de consumo
Identificar patrones significativos en ingresos por habitaciones	Detectar mínimo dos períodos de incremento significativo en ingresos
Analizar variaciones en reservas de servicios y habitaciones	Identificar al menos una tendencia mensual y un patrón semanal en la demanda

Nota. Elaboración Propia (2024)

Los objetivos de minería, detallados en la Tabla 29, reflejan una comprensión madura de las necesidades analíticas del sector hotelero. El enfoque en segmentación de clientes, análisis de rentabilidad y patrones temporales responde directamente a las necesidades identificadas en las entrevistas con la gerencia (**Apéndice J**). Los criterios de éxito establecidos proporcionan métricas cuantificables para evaluar el impacto de las soluciones implementadas.

4.1.3.2 Selección de *KPIs* para los objetivos de minería

Los indicadores clave de rendimiento definidos en la **Sección 2.10** se asocian con cada objetivo de minería de la siguiente manera:

Tabla 30. *Indicadores clave por objetivo de minería*

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo	KPI	Importancia
Desarrollar modelo de segmentación de clientes basado en patrones de gasto.	Ingreso por Cliente (RPC)	Facilita segmentación según valor económico
	Tasa de Retención de Clientes	Indica lealtad por segmento
	Días de Ocupación	Muestra preferencias de estancia por segmento
Analizar rentabilidad y tendencias de consumo en servicios	Número de Servicios Vendidos	Base para análisis de demanda
	Ingreso por Persona	Indica generación de valor
	Tasa de Ocupación	Contextualiza el consumo
Identificar patrones significativos en ingresos por habitaciones	RevPAR (Ingresos por Habitación Disponible)	Métrica central para identificar incrementos
	ADR (Tarifa Diaria Promedio)	Analiza impacto de tarifas
	Tasa de Ocupación	Evalúa impacto del volumen
	Ingreso por Cliente (RPC)	Incorpora ingresos adicionales
	TRevPAR (Ingresos Totales por Habitación Disponible)	Evalúa rendimiento por áreas
	Número de Habitaciones Vendidas	Relaciona ocupación e ingresos
Analizar variaciones en reservas de servicios y habitaciones	Número de Habitaciones y Servicios Vendidos	Métrica base para patrones
	Tasa de Ocupación	Relaciona reservas y capacidad

Nota. Elaboración Propia (2024)

La selección de indicadores clave, documentada en la Tabla 30, integra métricas estándar de la industria con indicadores específicos de comportamiento del cliente. Esta selección se fundamenta en las recomendaciones de la Cámara Costarricense de Hoteles (2021) y el análisis de Rondi et al. (2019) sobre indicadores de desempeño en la actividad hotelera. La validación con el equipo gerencial aseguró la relevancia práctica de los KPIs seleccionados.

4.1.4 Plan de implementación del proyecto

La estrategia de implementación refleja un enfoque pragmático hacia el desarrollo de capacidades analíticas. La selección de herramientas combina soluciones de código abierto para el análisis estadístico con la infraestructura existente para visualización, maximizando la eficiencia de recursos disponibles.

El plan evidencia una comprensión realista de las limitaciones organizacionales y técnicas identificadas en el análisis de la situación actual, proponiendo un desarrollo iterativo que permite la adaptación continua del proceso.

4.1.4.1 Plan de proyecto

El cronograma establecido refleja un equilibrio cuidadoso entre la urgencia de mejora y la necesidad de una implementación robusta. La estructura de fases, alineada con la metodología CRISP-DM, proporciona puntos de control claros para la validación de entregables. Las entrevistas con el equipo técnico (**Apéndice AF**) permitieron validar la viabilidad de los plazos propuestos, considerando las limitaciones de recursos y las dependencias técnicas identificadas. El cronograma detallado se encuentra en la **Sección 3.8.2**

4.1.4.2 Evaluación inicial de herramientas

La selección de herramientas, detallada en la Tabla 31, representa una decisión estratégica fundamentada en múltiples criterios. La elección de R Studio como plataforma principal de análisis surge de una evaluación exhaustiva de alternativas, considerando factores como:

- Capacidad de integración con la infraestructura existente
- Robustez en el manejo de análisis estadísticos complejos
- Disponibilidad de bibliotecas especializadas para el sector hotelero
- Curva de aprendizaje para el equipo técnico
- Costo-beneficio de la implementación

A continuación, se muestra la Tabla 31 con las herramientas consideradas para desarrollar el trabajo.

Tabla 31. *Herramientas seleccionadas por fase CRISP-DM*

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Fase CRISP-DM	Herramienta/Técnica	Descripción
Comprensión de los Datos	R Studio	Paquetes para exploración y resumen de datos
Preparación de los Datos		Paquetes para manipulación de datos
Modelado		Paquetes para modelado estadístico
Evaluación		Paquetes para evaluación de modelos
Despliegue	Devexpress Report Designer	Herramienta integrada para diseño de informes

Nota. Elaboración Propia (2024)

4.2 Entendimiento de los datos

Esta fase de CRISP-DM se desarrolló mediante un proceso iterativo en estrecha colaboración con la gerencia de Procom. A través de múltiples sesiones de trabajo, documentadas en los Apéndices I, J y K, se refinó progresivamente la comprensión de los datos disponibles y su relevancia para los objetivos del negocio.

El proceso iterativo permitió una evolución en el entendimiento de los datos, donde cada ciclo de análisis generaba nuevas preguntas que se discutían con la gerencia. Esta dinámica resultó particularmente valiosa en la identificación de patrones de negocio no evidentes en una primera exploración. Por ejemplo, la alta concentración de transacciones en un único cliente, inicialmente considerada como una anomalía en los datos, fue validada por la gerencia como un patrón de negocio real que requería atención.

La colaboración con el equipo gerencial fue instrumental en:

- La interpretación de patrones aparentemente anómalos
- La validación de relaciones identificadas entre variables
- La priorización de áreas de análisis según su relevancia comercial
- La identificación de limitaciones conocidas en los datos

Este enfoque colaborativo e iterativo permitió no solo una comprensión técnica de los datos, sino también una apreciación profunda de su contexto comercial y operativo. La retroalimentación constante de la gerencia enriqueció significativamente el análisis, asegurando que los hallazgos técnicos se alinearan con la realidad operativa del negocio.

4.2.1 Recolección inicial de los datos

El proceso de recolección de datos se realizó mediante la conexión entre R y Microsoft SQL Server utilizando el controlador RODBC. El análisis inicial de la estructura de datos reveló una complejidad significativa, con múltiples tablas interrelacionadas que requirieron un proceso de unificación para facilitar el análisis.

4.2.1.1 Tablas seleccionadas para el proyecto

Las tablas por utilizar se presentaron en la **Figura 16**. El **Apéndice AH** contiene la descripción de cada tabla y su relevancia para el proyecto.

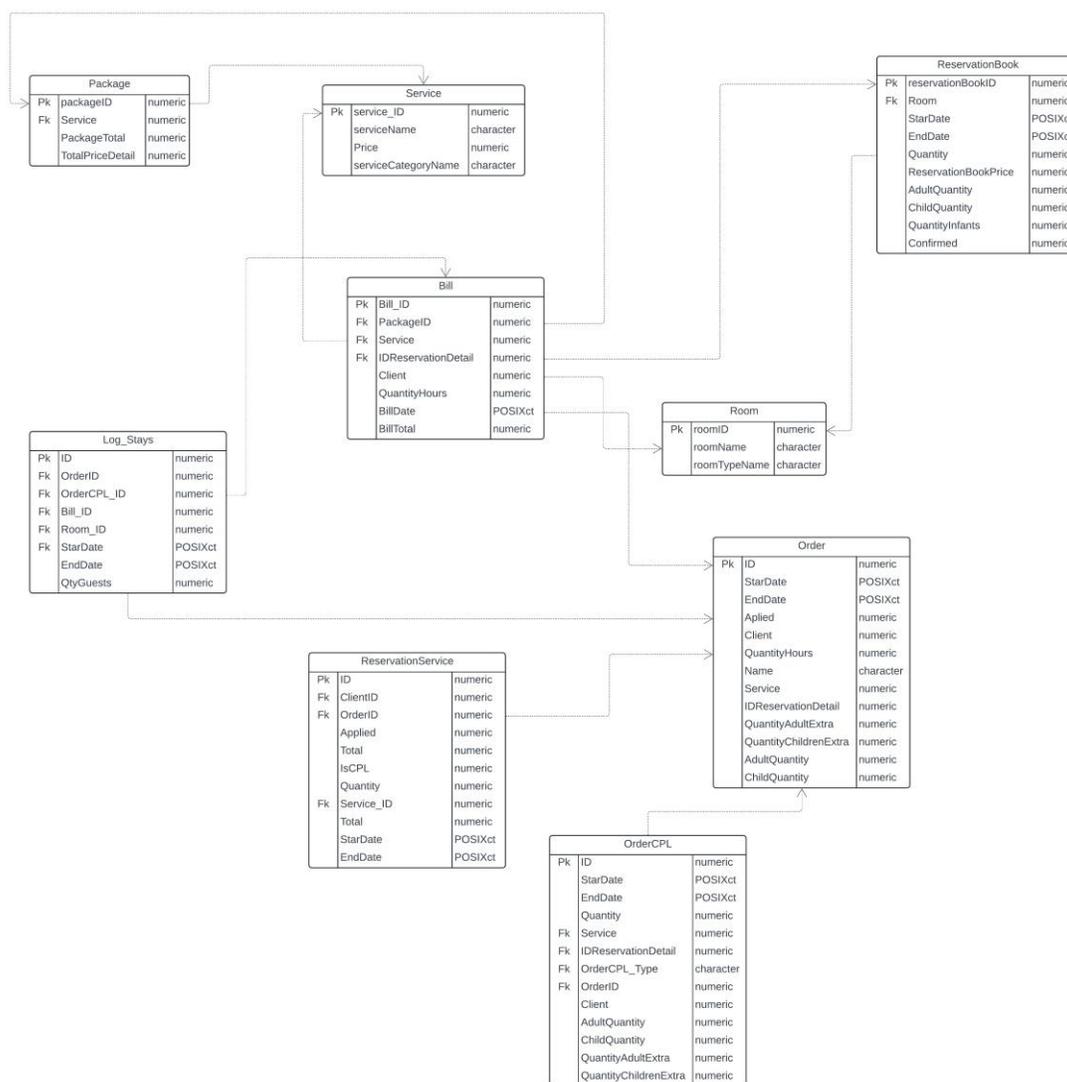
Para simplificar el análisis, se realizaron las siguientes unificaciones de tablas basadas en las relaciones presentadas en la **Figura 16**.

- Facturación (Bill) con Detalle de facturación (BillDetail)
- Detalle de reserva (ReservationBookDetail) con Tarifa (Rate)
- Libro de reservas (ReservationBook) con Detalle de reserva
- Categoría de servicio (ServiceCategory) con Centro de costo (CostCenter)
- Servicio (Service) con Categoría de servicio
- Habitación (Room) con Tipo de habitación (RoomType)
- Orden (Order) con Detalle de orden (OrderDetail)
- Reserva de servicio (ReservationService) con Detalle de reserva de servicio (ReservationServiceDetail)
- Orden complementaria (OrderCPL) con Detalle de orden CPL (OrderCPLDetail)

La Figura 19 presenta el diagrama entidad-relación resultante de las tablas unificadas.

Figura 19. Diagrama de tablas unificadas

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM



Nota. Elaboración Propia (2024).

La unificación de tablas, representada en la **Figura 19**, resultó en una estructura cohesiva y manejable. Este proceso de consolidación redujo la complejidad inicial de 101 tablas a un conjunto enfocado de relaciones, manteniendo la integridad de la información relevante para los objetivos del proyecto.

4.2.1.2 Variables consideradas para el análisis

Se realizó un análisis de las columnas de las tablas seleccionadas para examinar la varianza y el porcentaje de valores faltantes. Este proceso permitió identificar las variables relevantes para los objetivos del estudio.

Mediante la biblioteca "Data Explorer" de R, se ejecutó el siguiente proceso de selección:

- Se descartaron aquellas con porcentaje de valores nulos clasificado como no utilizable
- La presencia de estas columnas posibilita la aparición de sesgos en los resultados
- Se descartaron variables con valores constantes (igual a cero)
- Estas variables no aportan información significativa al análisis
- Se retuvieron las variables alineadas con los indicadores clave definidos en la **Sección 4.1.3.1**
- El diccionario de datos resultante se encuentra en el **Apéndice AI**, siguiendo la estructura de la **Figura 5**

4.2.2 Descripción de los datos

Esta etapa comprende el análisis de características de las columnas seleccionadas en la **Sección 4.2.1.2**. Según Chapman et al. (2000), esta fase busca familiarizarse con los datos y detectar posibles limitaciones que afecten el desarrollo del proyecto, permitiendo definir objetivos y tareas con mayor precisión.

4.2.2.1 Descripción de la Tabla de Facturación (Bill)

La completitud de la variable en servicios (68%) y detalles de reserva (32%) limita la capacidad de análisis de patrones de consumo, requiriendo estrategias de imputación o análisis parcial como se muestra en la Figura 20.

Figura 20. *Análisis estadístico de las variables categóricas de la tabla Bill*

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Bill_ID	0	1.00	1	4	0	3307	0
Service	2081	0.68	1	3	0	104	0
Room	5294	0.18	1	2	0	15	0
IDReservationDetail	4418	0.32	1	4	0	958	0
Client	0	1.00	1	4	0	1368	0
Order	0	1.00	1	4	0	3307	0

Nota. Elaboración propia (2024).

También la cobertura temporal de siete meses restringe la capacidad de análisis estacional, afectando directamente el objetivo de identificación

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

de patrones temporales como se muestra en la Figura 21.
 Figura 21. *Análisis estadístico de la fecha de la tabla Bill*

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
BillDate	0	1	2023-10-03	2024-05-22	2024-01-27	219

Nota. Elaboración propia (2024).

Asimismo, la distribución asimétrica en montos de facturación que se muestra en la Figura 22, sugiere la necesidad de técnicas robustas para el análisis financiero y la identificación de segmentos de valor.

Figura 22. Análisis estadístico de las variables numéricas de la tabla Bill

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
SubTotal	0	1	25256.90	25369.16	0	10000.00	17999.98	38500.00	566999.1
QuantityHours	0	1	2.29	3.50	0	1.00	1.00	3.00	114.0
BillTotal	0	1	79246.51	89162.57	0	23999.98	52499.98	98999.93	792315.0

Nota. Elaboración propia (2024).

4.2.2.2 Descripción de la Tabla de Reservaciones (ReservationBook)

La Figura 23 presenta una alta completitud general (>99% en mayoría de campos) que permite el análisis confiable de patrones de reserva y comportamiento de huéspedes. Sin embargo, se debe realizar un proceso de limpieza para la variable OrderID.

Figura 23. *Análisis estadístico de las variables categóricas de la tabla ReservationBook*

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
reservationBookID	0	1.00	1	4	0	1082	0
Room	0	1.00	1	2	0	16	0
OrderID	350	0.81	2	4	0	871	0
ServicelD	0	1.00	1	3	0	36	0

Nota. Elaboración propia (2024).

De acuerdo con la Figura 24, la variable Confirmed es una variable binaria (0 o 1). También se observa que QuantityAdultExtra y QuantityChildrenExtra tienen 13 valores faltantes (completitud 99%)

Figura 24. Análisis estadístico de las variables numéricas de la

tabla ReservationBook

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
AdultQuantity	0	1.00	1.71	1.55	0	0.0	2.00	2.00	8.0
ChildQuantity	0	1.00	0.14	0.46	0	0.0	0.00	0.00	4.0
Quantity	0	1.00	1.80	2.11	1	1.0	1.00	2.00	62.0
ReservationBookPrice	0	1.00	41453.33	30871.79	0	24336.3	41415.92	41415.93	476991.2
QuantityAdultExtra	13	0.99	0.42	0.99	0	0.0	0.00	0.00	6.0
QuantityChildrenExtra	13	0.99	0.13	0.45	0	0.0	0.00	0.00	4.0
QuantityInfants	0	1.00	0.03	0.18	0	0.0	0.00	0.00	2.0
Confirmed	0	1.00	0.81	0.39	0	1.0	1.00	1.00	1.0

Nota. Elaboración propia (2024).

4.2.2.3 Descripción de la Tabla de Habitaciones (Room)

La Figura 25 evidencia la completitud del 100% en todos los campos facilita el análisis de la capacidad instalada y su utilización. También la estructura clara de los tipos de habitación permite la segmentación precisa de ingresos y análisis de rentabilidad por tipo de unidad.

Figura 25. Análisis estadístico de las variables categóricas de la tabla

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
roomID	0	1	1	2	0	15	0
roomName	0	1	7	14	0	15	0
roomTypeName	0	1	9	10	0	2	0

Nota. Elaboración propia (2024).

4.2.2.4 Descripción de la Tabla de Servicios (Service)

La tabla Service presenta siete categorías y centros de costo definidos que proporcionan una base sólida para el análisis de rentabilidad por la línea de servicio tal como se evidencia en la Figura 26.

Figura 26. Análisis estadístico de las variables categóricas de la tabla Service

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
service_ID	0	1	1	3	0	173	0
serviceName	0	1	4	94	0	173	0
costCenterName	0	1	3	12	0	7	0
serviceCategoryName	0	1	3	12	0	7	0

Nota. Elaboración propia (2024).

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

4.2.2.5 Descripción de la Tabla de Órdenes (Order)

La Figura 27 evidencia un 97% de completitud en identificadores de reserva, lo cual permite la trazabilidad confiable entre órdenes y reservaciones.

Figura 27. Análisis estadístico de las variables categóricas de Order

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ID	0	1.00	1	4	0	4066	0
Client	0	1.00	1	4	0	1391	0
Applied	0	1.00	1	1	0	2	0
InProcess	0	1.00	1	1	0	2	0
RoomMain	0	1.00	1	2	0	16	0
ReservationID	195	0.97	1	4	0	2708	0

Nota. Elaboración propia (2024).

Asimismo, la Figura 28 muestra la presencia de campos temporales completos, lo cual facilita el análisis del ciclo de vida de las órdenes y patrones temporales de consumo.

Figura 28. Análisis estadístico de las fechas de la tabla Order

skim_variable	n_missing	complete_rate	min	max	median	n_unique
OpeningDate	0	1	2023-10-03	2024-05-22	2024-01-22	219
ClosingDate	12	1	2023-10-03	2024-05-22	2024-01-22	219

Nota. Elaboración propia (2024).

4.2.2.6 Descripción de la Tabla de Órdenes Complementarias (OrderCPL)

La Figura 29 evidencia valores faltantes en las variables Service ReservationID y Room lo cual requiere estrategias diferenciadas de análisis según el tipo de información, afectando la profundidad del análisis de las órdenes complementarias.

Figura 29 Análisis estadístico de las variables categóricas de OrderCPL

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ID	0	1.00	1	3	0	364	0
Client	0	1.00	1	4	0	34	0
ReservationID	177	0.61	1	4	0	208	0
OrderCPL_Type	0	1.00	7	35	0	10	0
Service	238	0.47	1	2	0	12	0
Room	428	0.06	1	2	0	12	0
OrderID	5	0.99	1	3	0	359	0

Nota. Elaboración propia (2024).

La Figura 30 evidencia valores faltantes para las variables StarDate y EndDate sugiriendo la aplicación de técnicas de limpieza para análisis de tiempo.

Figura 30 *Análisis estadístico de las fechas de la tabla Room*

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
StarDate	5	0.99	2023-10-06	2024-05-06	2023-12-29	164
EndDate	175	0.61	2023-10-06	2024-04-28	2023-11-29	89

Nota. Elaboración propia (2024).

La Figura 31 presenta las variables numéricas con una distribución asimétrica en pagos totales Impacto: Necesidad de técnicas robustas para el análisis financiero de servicios complementarios.

Figura 31. *Análisis estadístico de las variables numéricas de OrderCPL*

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75
Total_Paid	0	1	18250.91	29797.84	0	0	8849.56	21238.92

Nota. Elaboración propia (2024).

4.2.2.7 Descripción de la Tabla de Reserva de Servicios (ReservationService)

La Figura 32 presenta un análisis que da como resultado una completitud del 100% en todas las columnas. Asimismo, la diferencia de escalas entre variables Quantity y ReservationServiceTotal es un aspecto a considerarse para realizar los modelos analíticos planteados.

Figura 32. *Análisis estadístico de las variables de ReservationService*

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ID	0	1	1	4	0	2754	0
ClientID	0	1	1	4	0	619	0
Order_ID	0	1	1	4	0	2648	0
Applied	0	1	1	1	0	2	0
IsCPL	0	1	1	1	0	2	0
Service_ID	0	1	1	3	0	98	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
StartDate	0	1	2023-10-03	2024-12-07	2024-01-22	220
EndDate	0	1	2023-10-03	2024-05-07	2024-01-21	218

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
ReservationServiceTotal	0	1	43860.73	55074.84	750	17999.98	29999.97	53000.01	1222000.1
Quantity	0	1	3.51	4.28	1	2.00	2.00	4.00	114.0
ReservationServiceDetailTotal	0	1	23410.29	33680.51	750	7000.00	15000.00	26999.99	864499.9

Nota. Elaboración propia (2024).

4.2.2.8 Descripción de la Tabla de Reserva de Habitaciones (Log_Stays)

La Figura 33 evidencia una completitud del 100% en todos los campos de la tabla, lo cual resulta positivo para la realización de los modelos analíticos.

Figura 33. Análisis estadístico de las variables de la tabla Log_Stays

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
ID	0	1	1	3	0	988	0
Order_ID	0	1	2	4	0	866	0
OrderCPL_ID	0	1	1	3	0	21	0
Bill_ID	0	1	1	4	0	843	0
Room_ID	0	1	1	2	0	15	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
StartDate	0	1	2023-10-05	2024-05-22	2024-01-25	198
EndDate	0	1	2023-10-06	2024-05-24	2024-01-26	201

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75
Qty_Guests	0	1	3.54	2.66	0	2	2	4

Nota. Elaboración propia (2024).

4.2.3 Exploración de los datos

El análisis exploratorio de cada tabla revela patrones y características que impactan directamente en los objetivos de minería de datos establecidos. La exploración incluyó análisis de normalidad, correlaciones y distribuciones, proporcionando *insights* fundamentales para el diseño de los modelos analíticos. Es importante señalar que esta fase incluyó un proceso de iteración constante dada la naturaleza de los datos.

4.2.3.1 Exploración de la Tabla de Facturación (Bill)

En el análisis de facturación que se presenta en la Figura 34, las pruebas de normalidad revelan una distribución no gaussiana en las variables financieras ($p < 0.05$), característica típica en datos transaccionales que sugiere la necesidad de aproximaciones no paramétricas para el análisis posterior.

Figura 34 *Prueba de normalidad para las variables numéricas de Bill*

Columna	Estadistico_D	P_value
SubTotal	0.16914	1.593382e-161
BillTotal	0.18706	1.83865e-197

Nota. Elaboración propia (2024).

La matriz de correlación que se presenta en la Figura 35, identifica relaciones estructurales significativas, particularmente entre identificadores de factura y cliente, validando la integridad del sistema transaccional.

Figura 35 *.Matriz de correlación para las variables categóricas Bill*

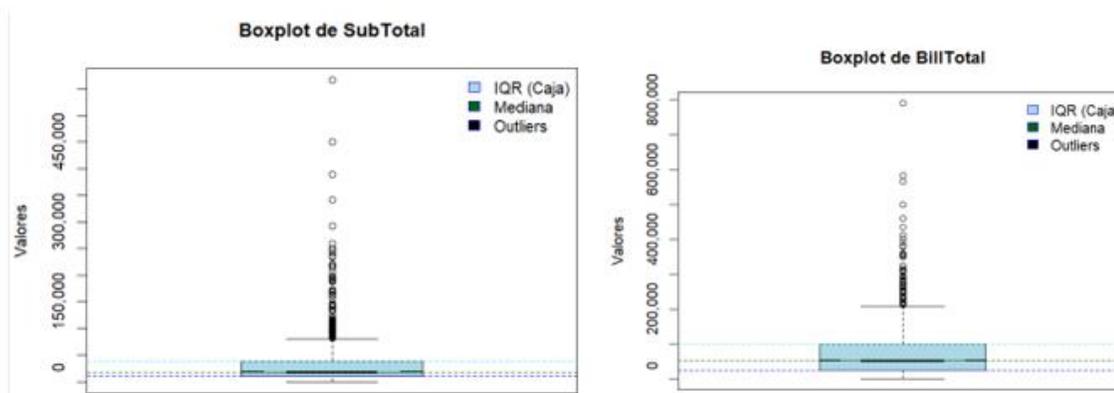
1.00	NaN	NaN	NaN	1.00	1.00	Bill_ID
NaN	1.00	NaN	NaN	NaN	NaN	PackageID
NaN	NaN	1.00	NaN	NaN	NaN	Service
NaN	NaN	NaN	1.00	NaN	NaN	IDReservationDetail
1.00	NaN	NaN	NaN	1.00	1.00	Client
1.00	NaN	NaN	NaN	1.00	1.00	Order
Bill_ID	PackageID	Service	IDReservationDetail	Client	Order	

Nota. Elaboración propia (2024).

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

El análisis de distribución de ingresos, visualizado en la Figura 36, revela una concentración de facturación en el rango de ₡10,000 a ₡35,000, con valores atípicos que alcanzan hasta ₡800,000. Esta distribución asimétrica sugiere la existencia de dos segmentos de mercado distintos: un mercado base de servicios regulares y un segmento premium de alto valor.

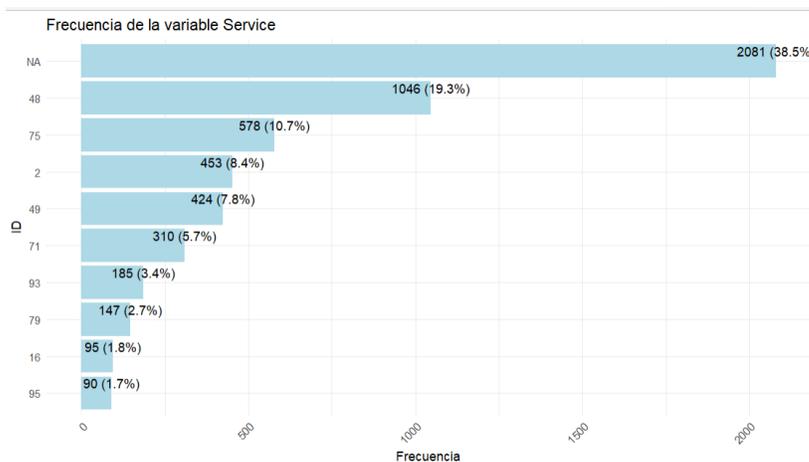
Figura 36. *Boxplot de las variables numéricas de la tabla Bill*



Nota. Elaboración propia (2024).

La exploración de servicios facturados que se presenta en la Figura 37 muestra un 38.5% de valores nulos. Esto implica su limpieza para pasos posteriores.

Figura 37. *Exploración de la variable Service en la tabla Bill*

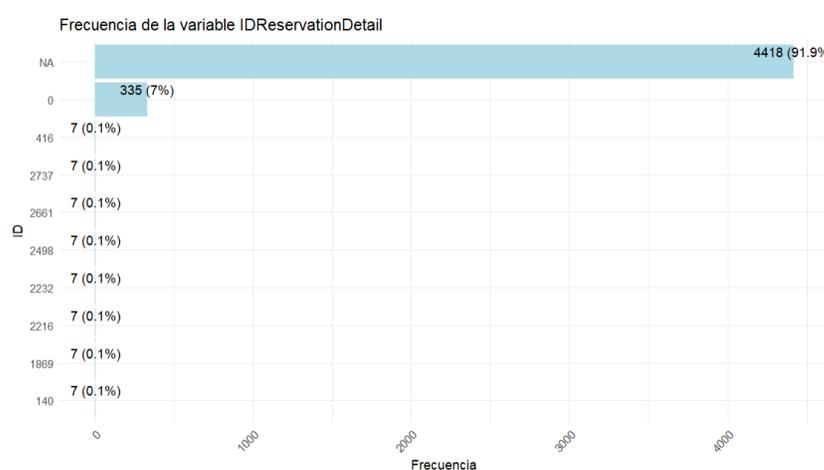


Nota. Elaboración propia (2024).

La exploración de las reservaciones facturadas que se presenta en la Figura 38 muestra un 91.9% de valores nulos. Esto implica su limpieza para pasos posteriores.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

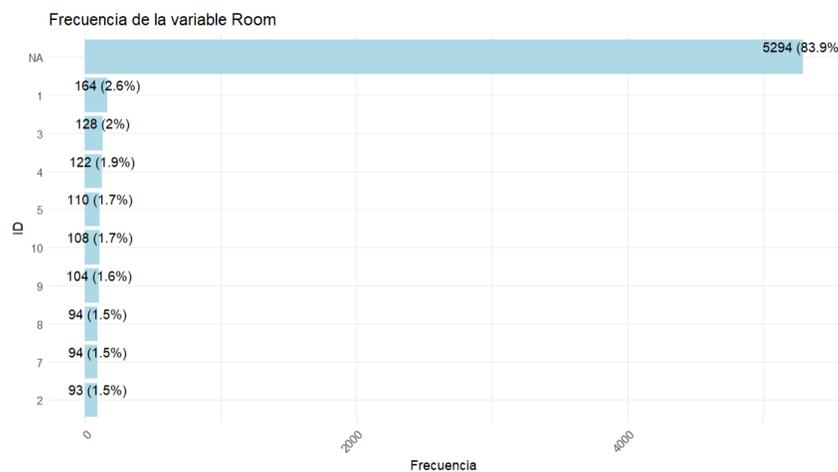
Figura 38 *Exploración de la variable IDReservationDetail en la tabla Bill*



Nota. Elaboración propia (2024).

La Figura 39 muestra la asignación de habitaciones con un 83.9% de valores faltantes. Esta situación impone restricciones significativas en la capacidad de análisis detallado por tipo habitación.

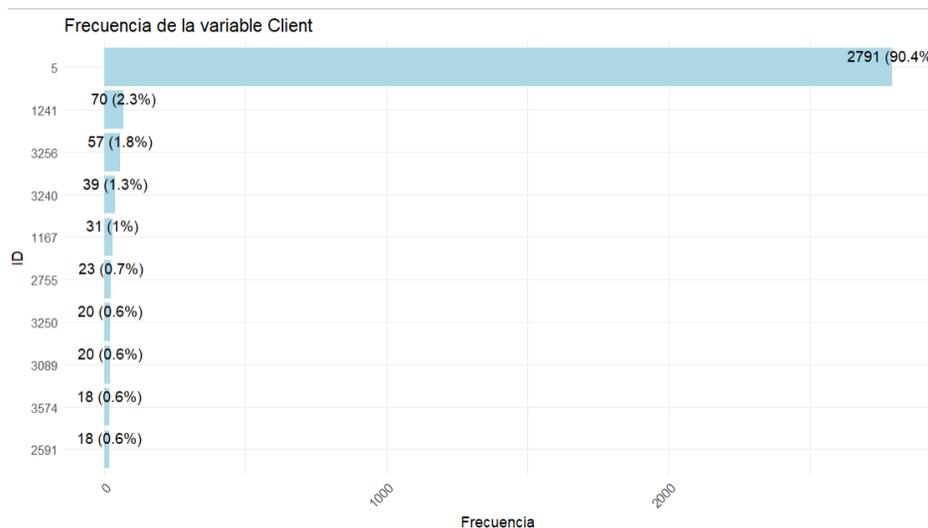
Figura 39. *Exploración de la variable Room en la tabla Bill*



Nota. Elaboración propia (2024).

El análisis de clientes que se presenta en la Figura 40 revela que un único cliente representa el 90.4% de las facturaciones, señalando una concentración de riesgo operativo significativa.

Figura 40. Exploración de la variable Client en la tabla Bill



Nota. Elaboración propia (2024).

4.2.3.2 Exploración de la Tabla de Reservas (ReservationBook)

En el análisis del libro de reservas que se presenta en la Figura 41, las pruebas de normalidad revelan una distribución no gaussiana en las variables financieras ($p < 0.05$), característica típica en datos transaccionales que sugiere la necesidad de aproximaciones no paramétricas para el análisis posterior.

Figura 41. Prueba de normalidad para las variables numéricas de la tabla ReservationBook

Columna	Estadistico_D	P_value
AdultQuantity	0.22690	1.235622e-81
ChildQuantity	0.51868	0e+00
Quantity	0.35300	5.55779e-197
ReservationBookPrice	0.27306	4.8989e-118
QuantityAdultExtra	0.46078	0e+00
QuantityChildrenExtra	0.51943	0e+00
QuantityInfants	0.53593	0e+00
Confirmed	0.49515	0e+00

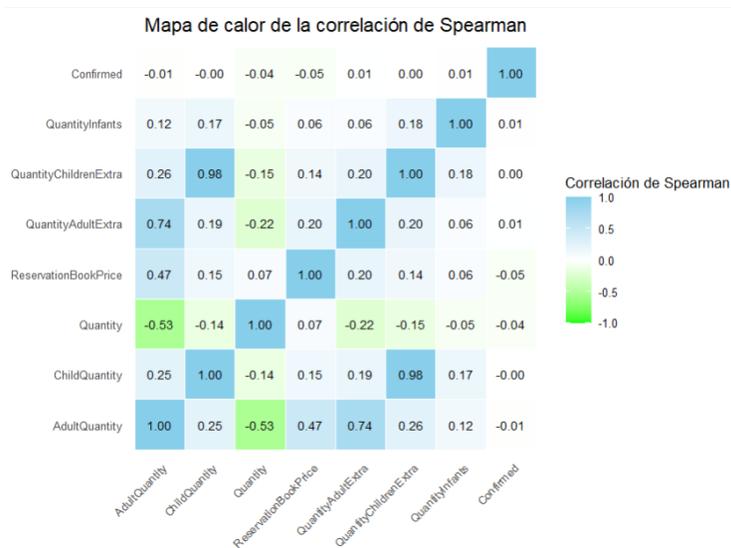
Nota. Elaboración propia (2024).

Las correlaciones entre variables de reservación, que se presentan en la Figura 42, revelan una estructura compleja de relaciones. Se observa una correlación positiva fuerte (0.98) entre la cantidad de niños adicionales y la cantidad total de niños, lo que valida la consistencia en el registro de huéspedes menores. Simultáneamente, la correlación positiva (0.74) entre la cantidad de adultos y adultos adicionales sugiere patrones predecibles en la

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

ocupación. Particularmente interesante resulta la correlación negativa moderada (-0.54) entre la cantidad total y el número de adultos, sugiriendo que las estadías de mayor duración tienden a involucrar grupos de menor tamaño.

Figura 42. Matriz de correlación para las variables numéricas de la tabla ReservationBook

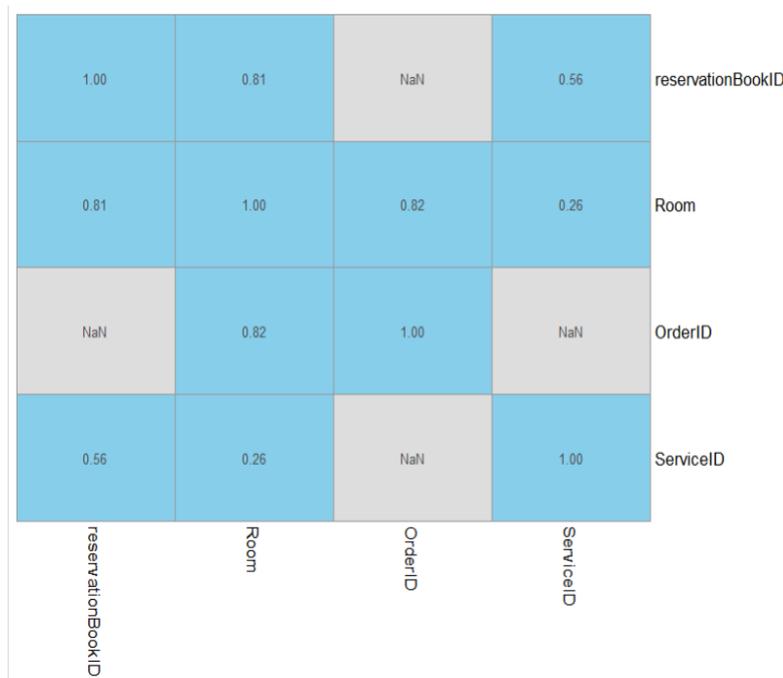


Nota. Elaboración propia (2024).

La Figura 43 expone las correlaciones entre variables categóricas de reservación, donde destaca una fuerte asociación (0.81) entre el identificador de reserva y la habitación asignada, así como entre el identificador de orden y habitación (0.82). Estas relaciones robustas validan la integridad del sistema de reservas y sugieren un proceso estructurado de asignación de habitaciones.

Figura 43. Matriz de correlación para las variables categóricas de la tabla

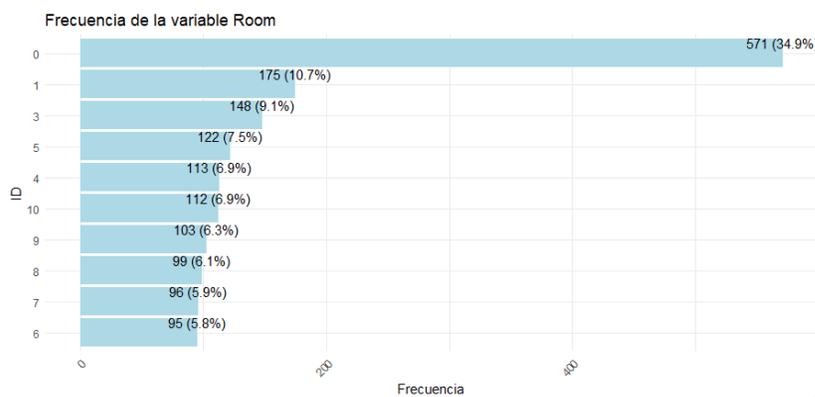
ReservationBook



Nota. Elaboración propia (2024).

El análisis del comportamiento de clientes en las reservaciones, que se ilustra en la Figura 44, revela que un 34.9% de las habitaciones registran valores nulos, sugiriendo una brecha significativa en la trazabilidad de la ocupación.

Figura 44. *Exploración de la variable Client en la tabla ReservationBook*

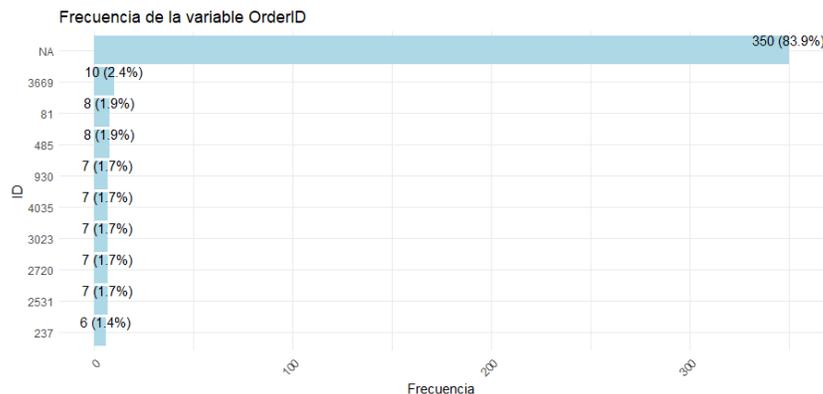


Nota. Elaboración propia (2024).

En la Figura 45, se observa que el 83.9% de las reservaciones carecen de identificador de orden asociado, lo cual compromete la capacidad de realizar un seguimiento completo del ciclo de reserva-estancia-facturación.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

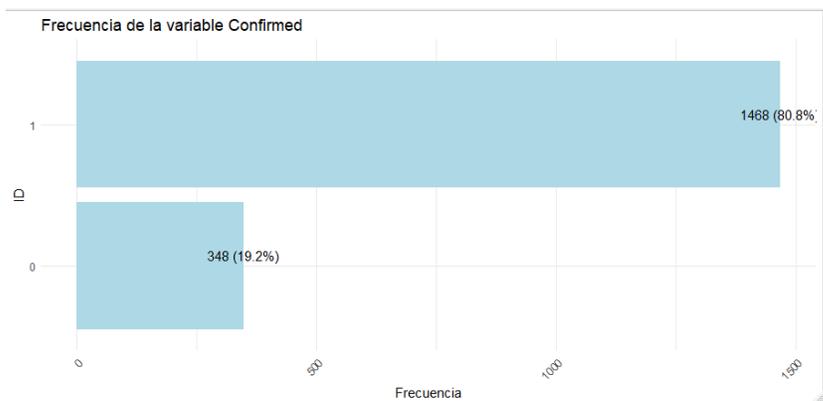
Figura 45. Exploración de la variable OrderID en ReservationBook



Nota. Elaboración propia (2024).

La tasa de confirmación de reservas, que se evidencia en la Figura 46, muestra un patrón operativo relevante: el 80.8% de las reservas alcanzan el estado de confirmación. Este indicador sugiere un proceso de conversión eficiente, aunque también señala un margen de mejora potencial en la gestión de reservas tentativas.

Figura 46. Exploración de la variable Confirmed en ReservationBook

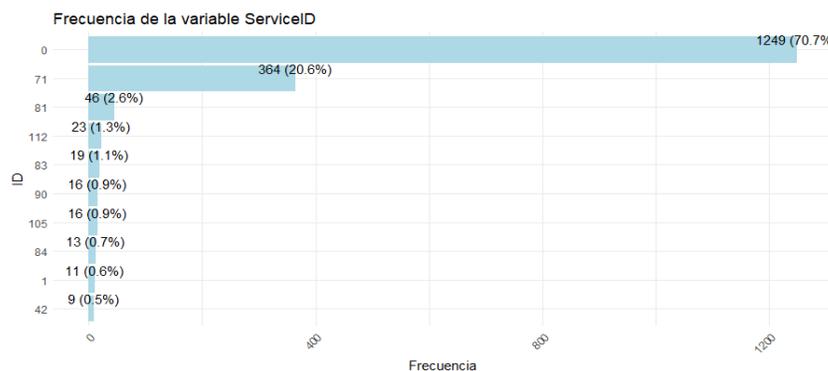


Nota. Elaboración propia (2024).

La Figura 47 revela que el 70.7% de las reservaciones no tienen servicios adicionales asociados, lo cual indica una oportunidad de pérdida en la venta cruzada de servicios.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

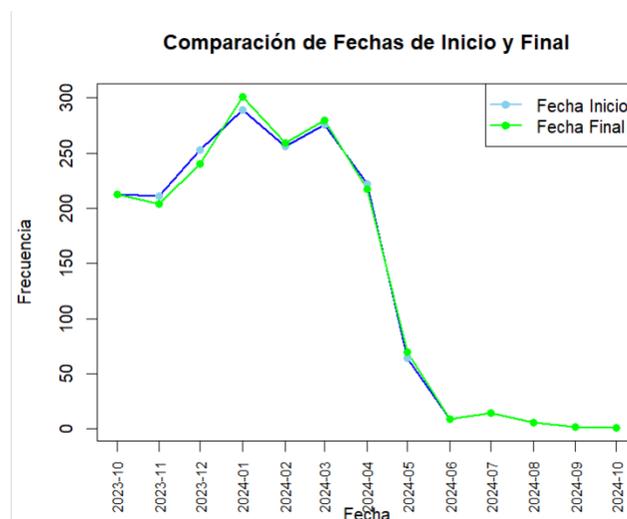
Figura 47. Exploración de la variable ServiceID en ReservationBook



Nota. Elaboración propia (2024).

Los patrones temporales de reservación, que se visualizan en la Figura 48, muestran una marcada estacionalidad. Los meses de enero y marzo de 2024 presentan los volúmenes altos de actividad, mientras que noviembre 2023 registra el punto de menor demanda. Esta variabilidad temporal tiene implicaciones directas para la planificación operativa y la estrategia de revenue management.

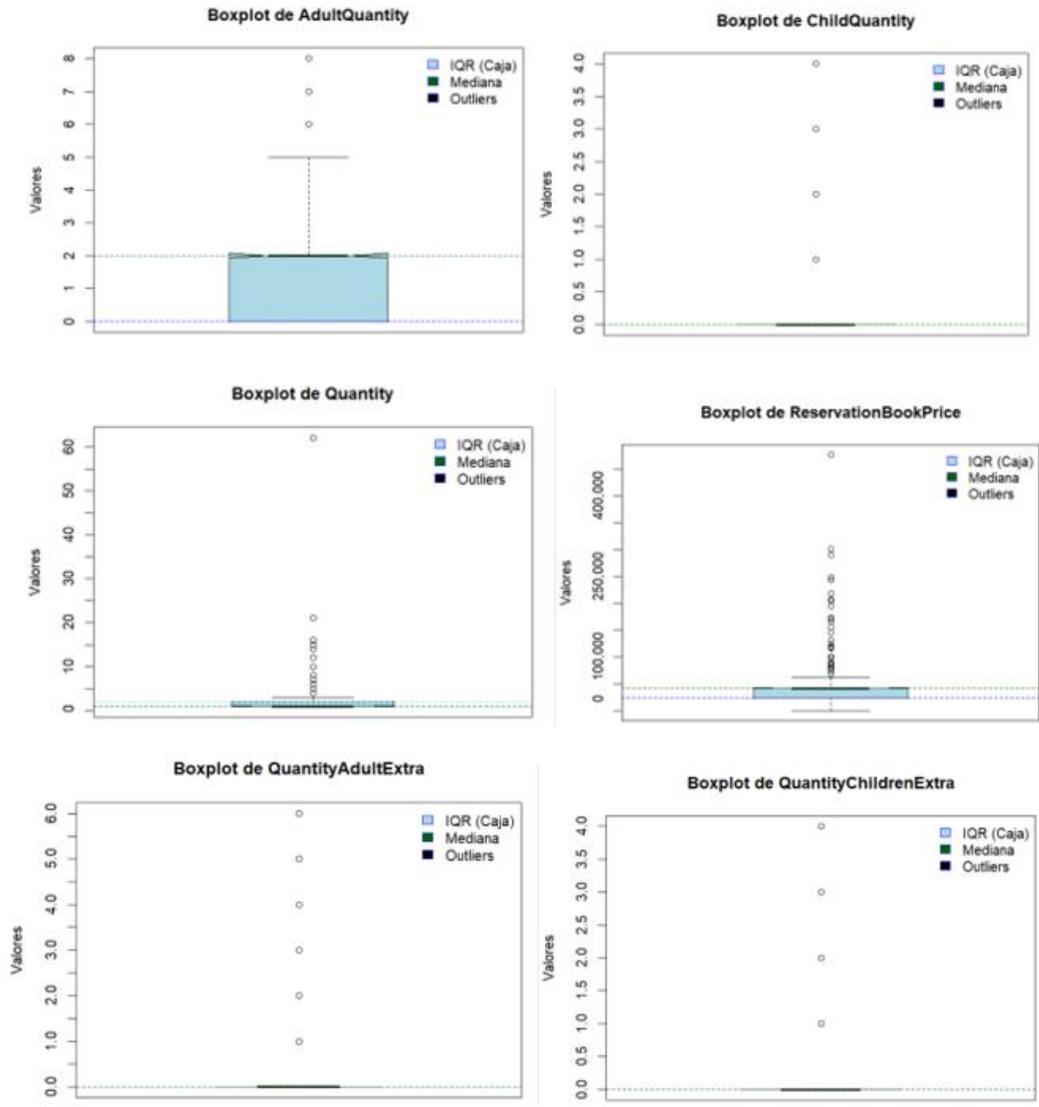
Figura 48. Exploración de las fechas en la tabla ReservationBook



Nota. Elaboración propia (2024).

El análisis de la distribución de huéspedes, que se presenta en los boxplots de la Figura 49, revela patrones característicos en la composición de las reservas. La mediana de dos adultos por reserva, con un rango intercuartílico de 1-2 adultos y valores atípicos que alcanzan hasta seis adultos, sugiere un perfil de cliente predominantemente familiar o de parejas. La presencia marginal de niños en las reservas, con una mediana cercana a cero y valores atípicos hasta cuatro niños, refuerza esta interpretación del segmento de mercado principal.

Figura 49. *Boxplots de las variables numéricas de ReservationBook*

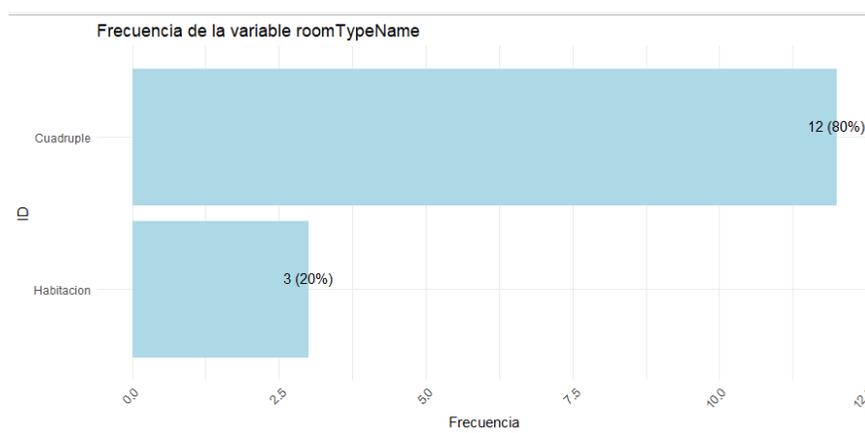


Nota. *Elaboración propia (2024).*

4.2.3.3 Exploración de la Tabla de Habitaciones (Room)

La distribución de tipos de habitación, visualizada en la Figura 50, muestra una predominancia de habitaciones cuádruples sobre sencillas, sugiriendo una orientación estratégica hacia grupos familiares o pequeños grupos. Esta configuración de inventario tiene implicaciones directas para la estrategia de revenue management y la optimización de la capacidad instalada.

Figura 50. Exploración de la variable `roomTypeName` en la tabla `Room`



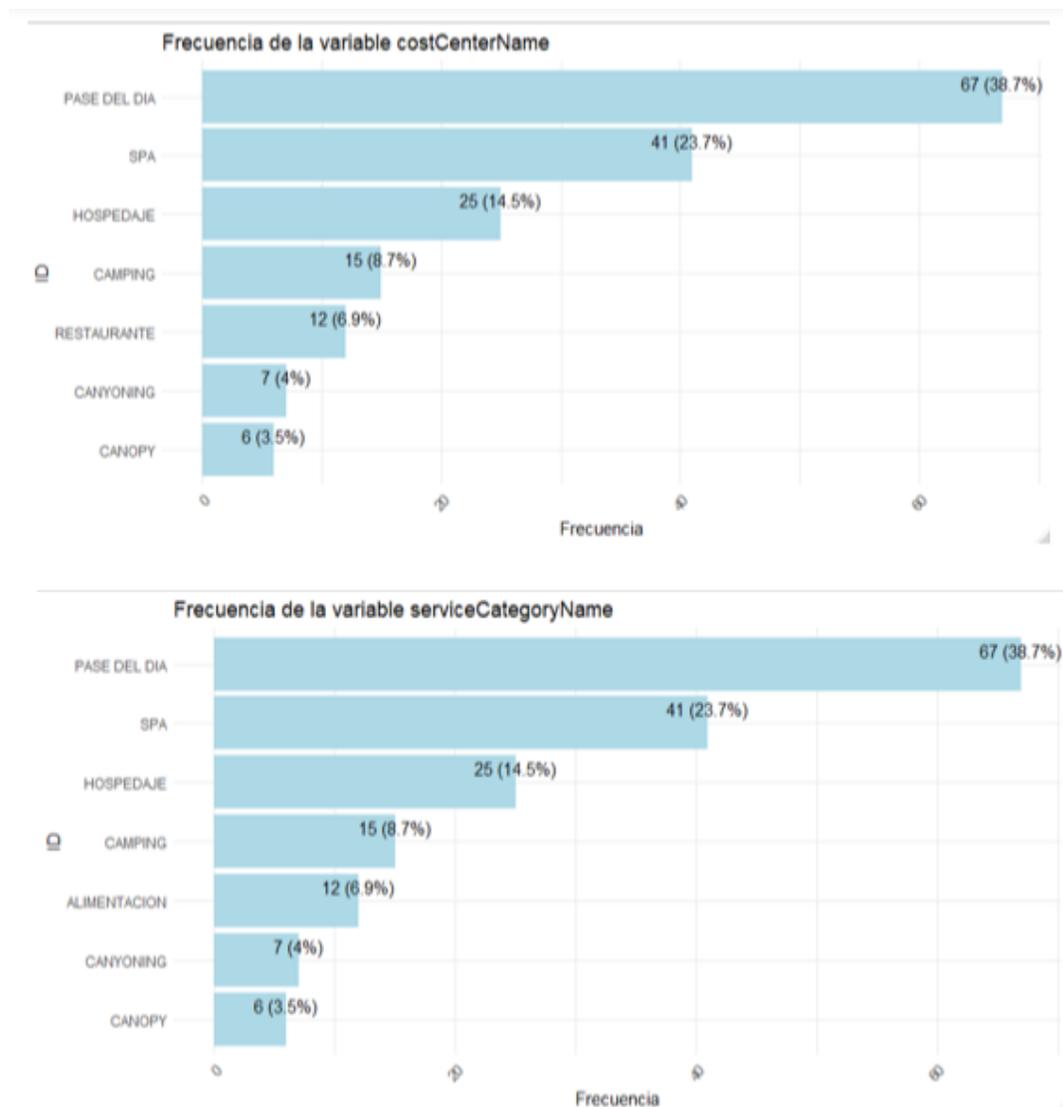
Nota. Elaboración propia (2024).

4.2.3.4 Exploración de la Tabla de Servicios (Service)

La organización de servicios, visualizada en la Figura 51, expone la estructura operativa del hotel a través de sus categorías y centros de costo. La correspondencia directa entre estas clasificaciones demuestra una organización lógica del negocio, aunque la duplicidad identificada entre las categorías "restaurante" y "alimentación" señala una inconsistencia que podría afectar la precisión del análisis financiero por la línea de servicio.

Figura 51. Exploración de las variables `CostCenter` y `ServiceCategory` en la tabla `Service`

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM



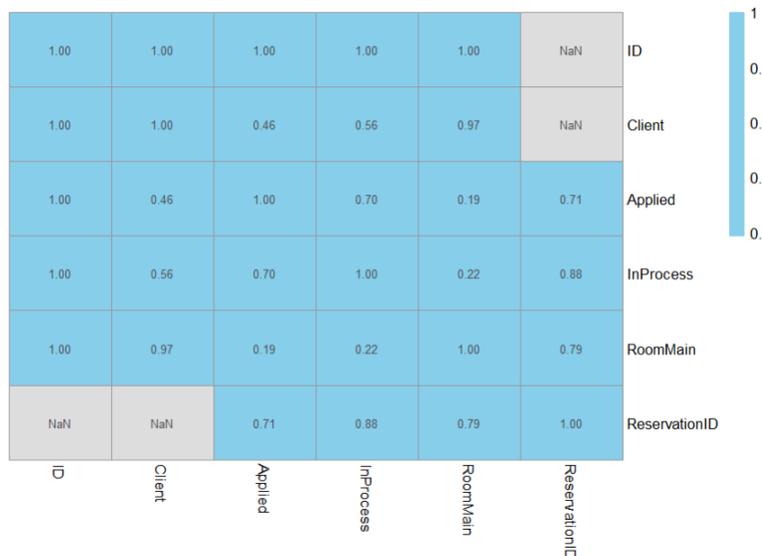
Nota. Elaboración propia (2024).

4.2.3.5 Exploración de la Tabla de Órdenes (Order)

El estudio de las relaciones entre variables categóricas en las órdenes, que se presenta en la Figura 52, revela una estructura de datos coherente con correlaciones significativas. La correlación perfecta (1.0) entre identificadores y clientes valida la integridad del sistema transaccional, mientras que las correlaciones fuertes entre habitación principal y estado de reserva (0.79-0.88) sugieren un proceso estructurado de gestión de órdenes.

Figura 52. *Matriz de correlación para las variables categóricas de la tabla Order*

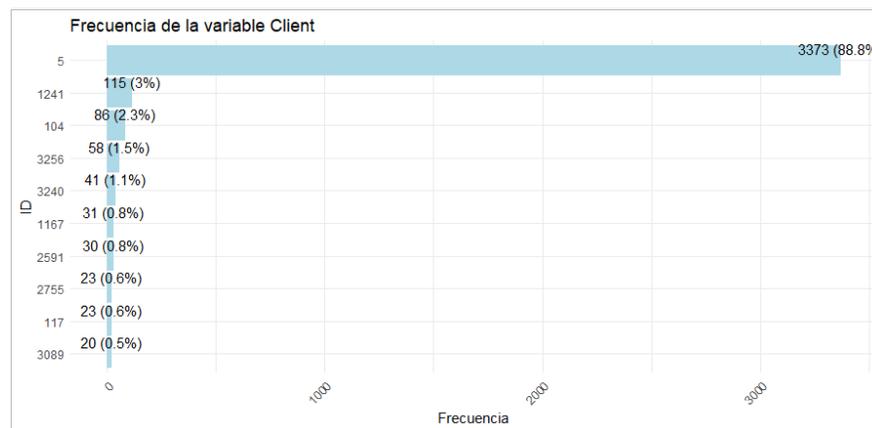
Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM



Nota. Elaboración propia (2024).

La concentración de transacciones por cliente, que se evidencia en la Figura 53, expone un patrón preocupante: el cliente identificado como "cliente 5" acumula el 88.8% de las órdenes totales. Esta concentración extrema, consistente con los hallazgos previos en facturación, confirma una dependencia crítica que representa un riesgo operativo significativo para el negocio.

Figura 53. Exploración de la variable Client en la tabla Order

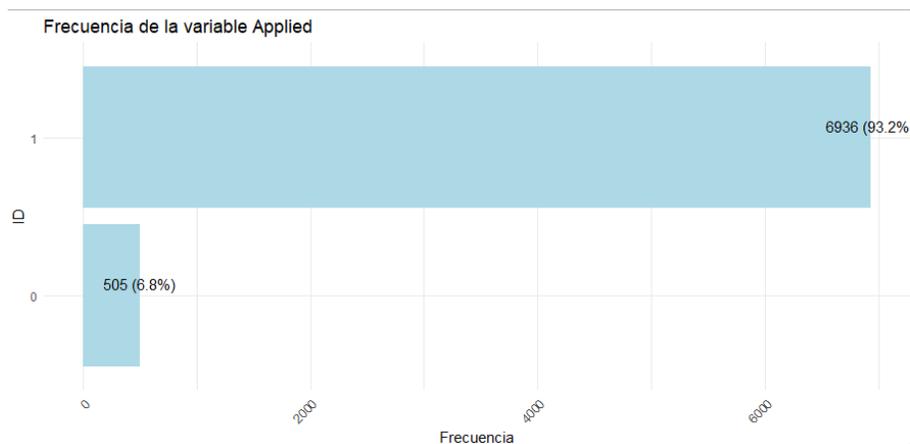


Nota. Elaboración propia (2024).

El análisis del estado de facturación, representado en la Figura 54, revela un proceso operativo eficiente con un 93.2% de órdenes efectivamente facturadas. Este alto porcentaje de conversión sugiere un sistema robusto de seguimiento y cierre de transacciones, aunque el 6.8% restante de órdenes no facturadas merece atención para optimizar la captura de ingresos.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

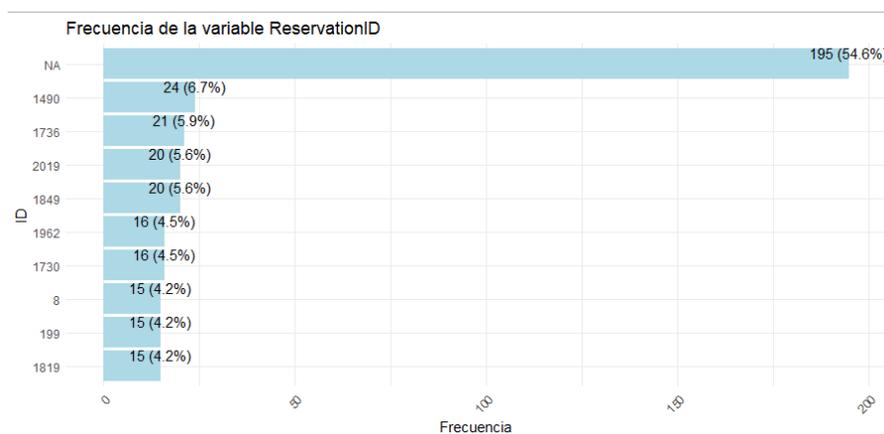
Figura 54. Exploración de la variable Applied en la tabla Order



Nota. Elaboración propia (2024).

La exploración de la vinculación con reservaciones, visualizada en la Figura 55, muestra que el 54.6% de las órdenes carecen de identificador de reserva asociado. Esta desconexión significativa entre órdenes y reservaciones sugiere un desafío importante para la trazabilidad completa del ciclo del cliente e impacta la capacidad de análisis integral del comportamiento de consumo.

Figura 55. Exploración de la variable ReservationID en la tabla Order

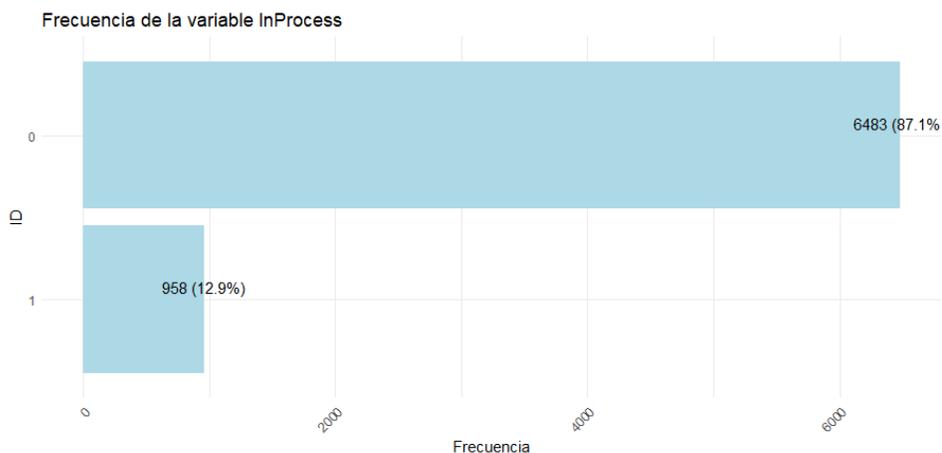


Nota. Elaboración propia (2024).

El estado de procesamiento de órdenes, que se presenta en la Figura 56, indica que el 87.19% de las transacciones han sido completamente procesadas.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

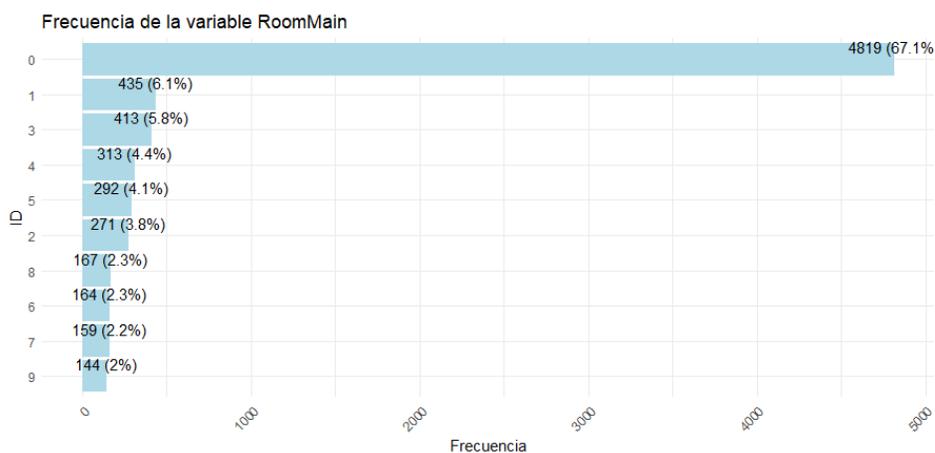
Figura 56. *Exploración de la variable InProcess en la tabla Order*



Nota. Elaboración propia (2024).

El análisis de la asignación de habitaciones principales que se presenta en la Figura 57, revela un 67.1% de registros en cero y sugiere una predominancia de servicios no vinculados directamente a hospedaje.

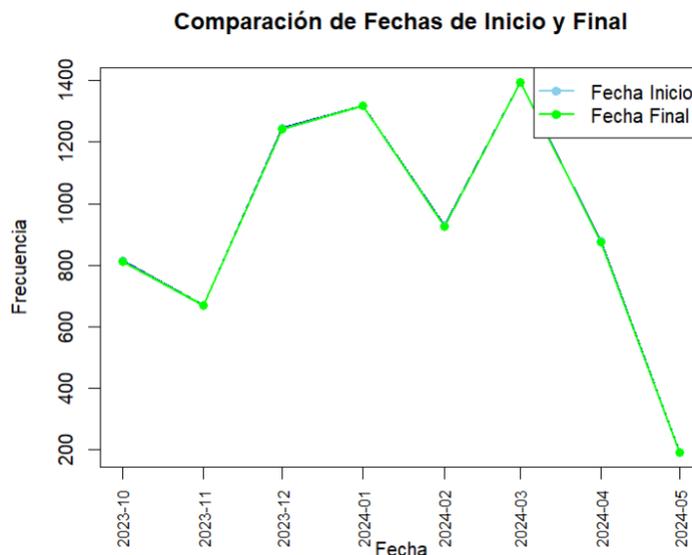
Figura 57. *Exploración de la variable RoomMain en la tabla Order*



Nota. Elaboración propia (2024).

Los patrones temporales de órdenes, representados en la Figura 58, confirman una marcada estacionalidad en la operación del hotel. Los meses de enero y marzo de 2024 emergen consistentemente como períodos de alta actividad, alineándose con los patrones previamente identificados en reservaciones. Esta consistencia en la estacionalidad proporciona una base sólida para la planificación operativa y la optimización de recursos.

Figura 58. Exploración de las fechas en la tabla Order



Nota. Elaboración propia (2024).

4.2.3.6 Exploración de la Tabla de Órdenes Complementarias (OrderCPL)

En el análisis de órdenes complementarias que se presenta en la Figura 59, las pruebas de normalidad revelan una distribución no gaussiana en las variables financieras ($p < 0.05$), característica típica en datos transaccionales que sugiere la necesidad de aproximaciones no paramétricas para el análisis posterior.

Figura 59 Prueba de normalidad para las variables numéricas de la tabla OrderCPL

Columna	Estadistico_D	P_value
Total_Paid	0.27011	3.930036e-29

Nota. Elaboración propia (2024).

La matriz de correlación para variables categóricas en órdenes complementarias, representada en la Figura 60, revela relaciones estructurales significativas. Las correlaciones perfectas (1.0) entre OrderID con OrderCPL_Type y Client, junto con la fuerte correlación (0.91) entre OrderID y Name, sugieren un sistema de registro robusto y coherente. Estas relaciones validan la integridad del sistema transaccional para servicios complementarios.

Figura 60. Matriz de correlación para las variables categóricas de la tabla

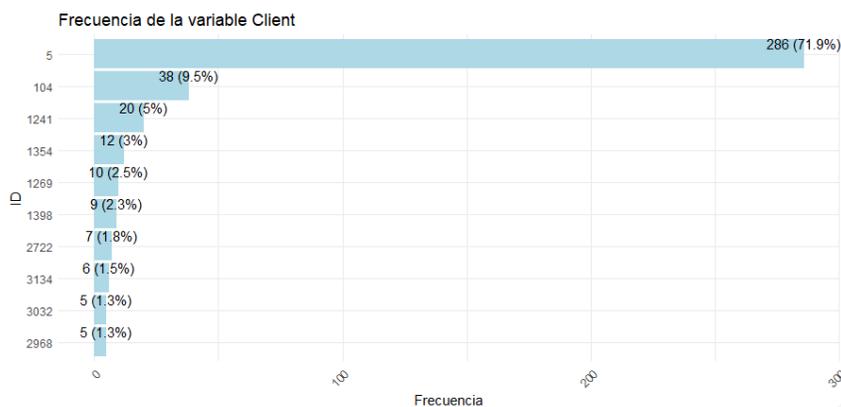
OrderCPL

1.00	1.00	NaN	1.00	NaN	NaN	NaN	ID
1.00	1.00	NaN	0.71	NaN	NaN	1.00	Client
NaN	NaN	1.00	1.00	NaN	NaN	NaN	ReservationID
1.00	0.71	1.00	1.00	NaN	NaN	1.00	OrderCPL_Type
NaN	NaN	NaN	NaN	1.00	NaN	NaN	Service
NaN	NaN	NaN	NaN	NaN	1.00	NaN	Room
NaN	1.00	NaN	1.00	NaN	NaN	1.00	OrderID
ID	Client	ReservationID	OrderCPL_Type	Service	Room	OrderID	

Nota. Elaboración propia (2024).

El patrón de distribución de clientes en servicios complementarios, ilustrado en la Figura 61, mantiene la tendencia de concentración observada en análisis previos. El cliente identificado como "cliente 5" representa el 71.9% de las órdenes de cortesía, aunque esta concentración es relativamente menor que la observada en transacciones regulares, sugiriendo una distribución ligeramente equilibrada en servicios complementarios.

Figura 61. *Exploración de la variable Client en la tabla OrderCPL*



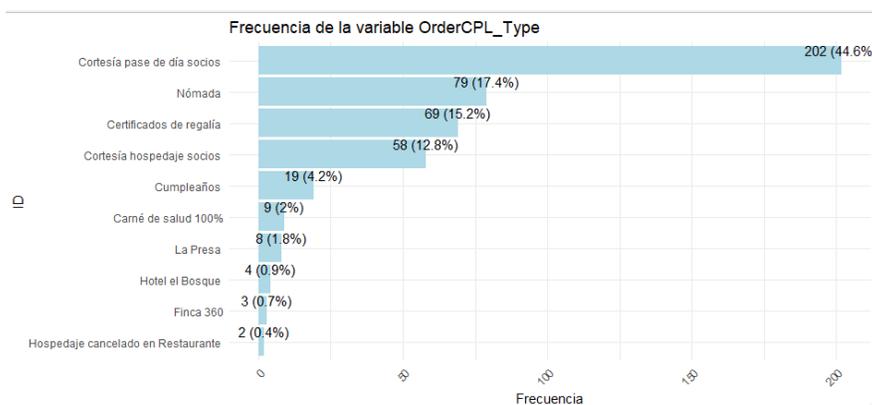
Nota. Elaboración propia (2024).

La exploración de tipos de órdenes complementarias, visualizada en

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

la Figura 62, revela una predominancia de cortesías tipo "pase de día" para socios. Este hallazgo tiene implicaciones significativas para la estrategia de fidelización y el análisis de rentabilidad, ya que sugiere un enfoque deliberado en la generación de valor para clientes frecuentes.

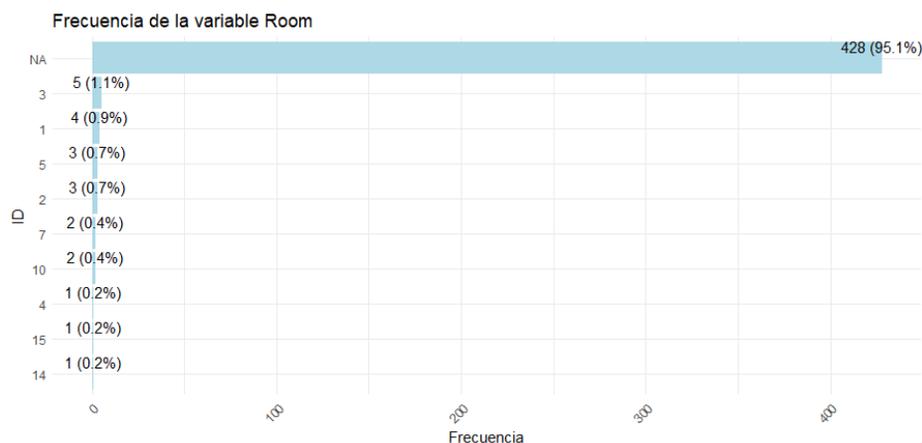
Figura 62. Exploración de la variable OrderCPL_Type en la tabla OrderCPL



Nota. Elaboración propia (2024).

La asignación de habitaciones en órdenes complementarias, representada en la Figura 63, muestra un patrón preocupante con un 95.1% de registros faltantes. Esta ausencia masiva de vinculación con habitaciones sugiere que los servicios complementarios operan de manera prácticamente independiente del sistema de hospedaje, limitando la capacidad de análisis integral de la experiencia del cliente.

Figura 63. Exploración de la variable Room en la tabla OrderCPL



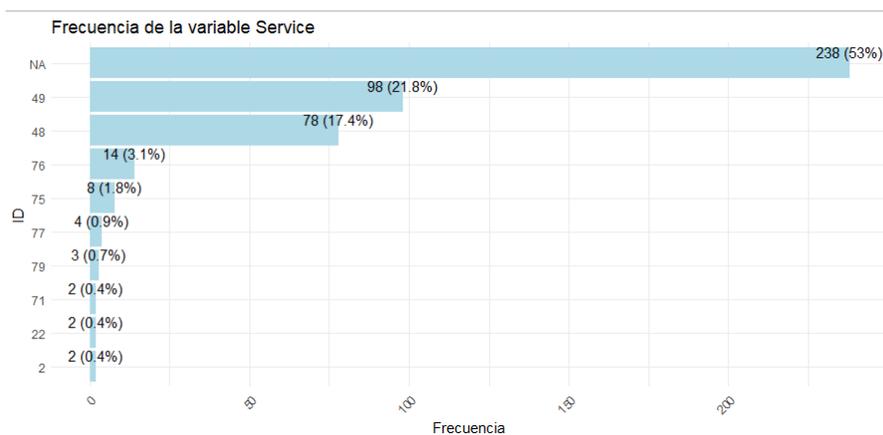
Nota. Elaboración propia (2024).

La distribución de servicios en órdenes complementarias, evidenciada en la Figura 64, revela que el 53% de las transacciones carecen

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

de identificación específica del servicio utilizado. Esta falta de granularidad en el registro de servicios complementarios compromete significativamente la capacidad de analizar patrones de consumo y preferencias en este segmento.

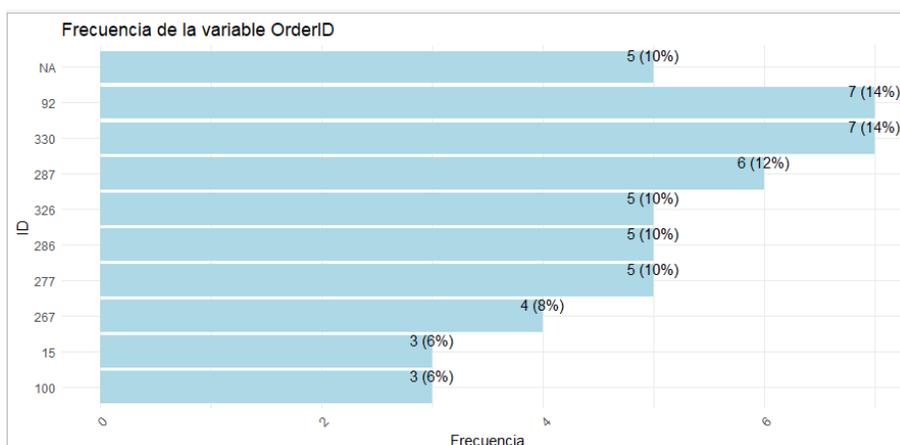
Figura 64. Exploración de la variable Service en la tabla OrderCPL



Nota. Elaboración propia (2024).

El análisis de la vinculación con órdenes principales, visualizado en la Figura 65, muestra una mejora notable en la integridad de datos con solo un 10% de registros faltantes.

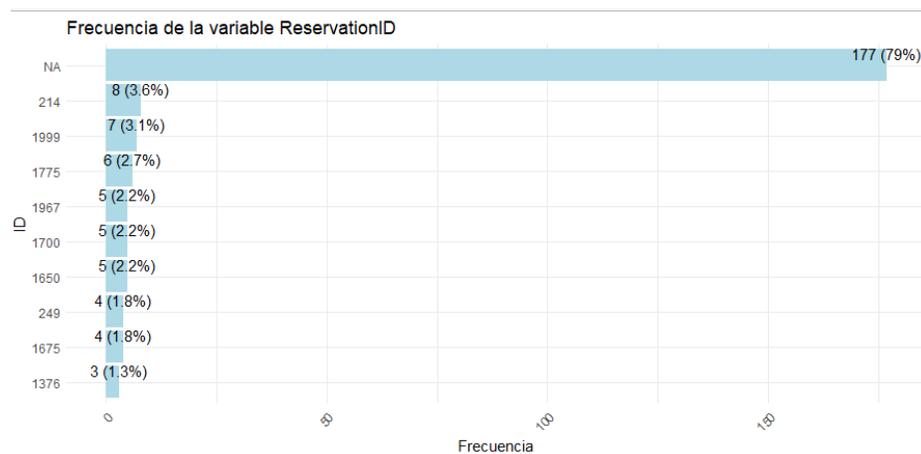
Figura 65. Exploración de la variable OrderID en la tabla OrderCPL



Nota. Elaboración propia (2024).

La exploración de la conexión con reservaciones de la Figura 66, revela que el 79% de las órdenes complementarias carecen de identificador de reserva, sugiriendo una desconexión significativa entre los servicios complementarios y el ciclo principal de hospedaje.

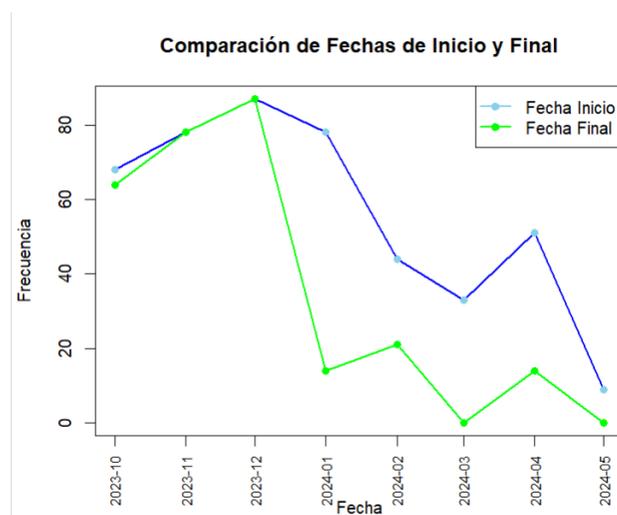
Figura 66. Exploración de la variable ReservationID en OrderCPL



Nota. Elaboración propia (2024).

Los patrones temporales de servicios complementarios, representados en la Figura 67, muestran una estacionalidad distintiva. Diciembre 2023 y enero 2024 emergen como los períodos de mayor actividad, mientras que marzo y mayo 2024 registran los volúmenes de menor demanda. Esta variación temporal, que difiere parcialmente de los patrones observados en servicios regulares, sugiere dinámicas específicas en el consumo de servicios complementarios.

Figura 67. Exploración de las fechas en la tabla OrderCPL

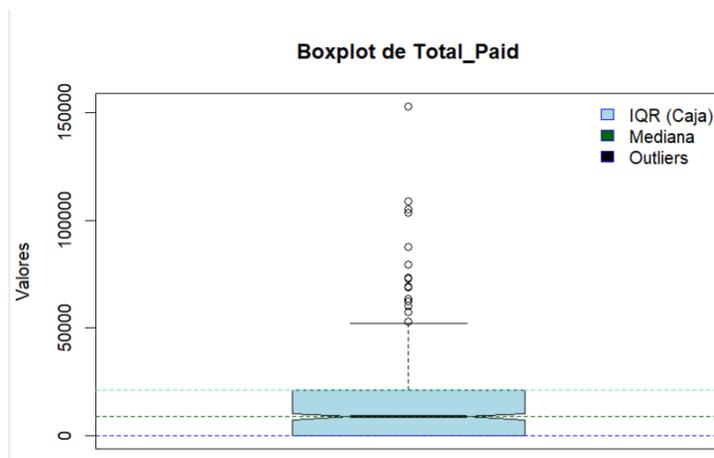


Nota. Elaboración propia (2024).

La distribución de montos pagados en servicios complementarios, visualizada en la Figura 68, revela una estructura de precios característica. La mayoría de las transacciones se concentran por debajo de ₡25,000, con valores atípicos que alcanzan hasta ₡150,000. Esta distribución asimétrica

sugiere una estrategia de precios escalonada, donde los servicios complementarios funcionan principalmente como generadores de valor agregado y no solo como fuentes primarias de ingreso.

Figura 68 *Boxplot de la variable Total_Paid en la tabla OrderCPL*



Nota. Elaboración propia (2024).

4.2.3.7 Exploración de la Tabla de Reserva de Servicios (ReservationService)

En el análisis de las reservaciones de servicios que se presenta en la Figura 69, las pruebas de normalidad revelan una distribución no gaussiana en las variables financieras ($p < 0.05$), característica típica en datos transaccionales que sugiere la necesidad de aproximaciones no paramétricas para el análisis posterior.

Figura 69. *Prueba de normalidad para las variables numéricas de la tabla ReservationService*

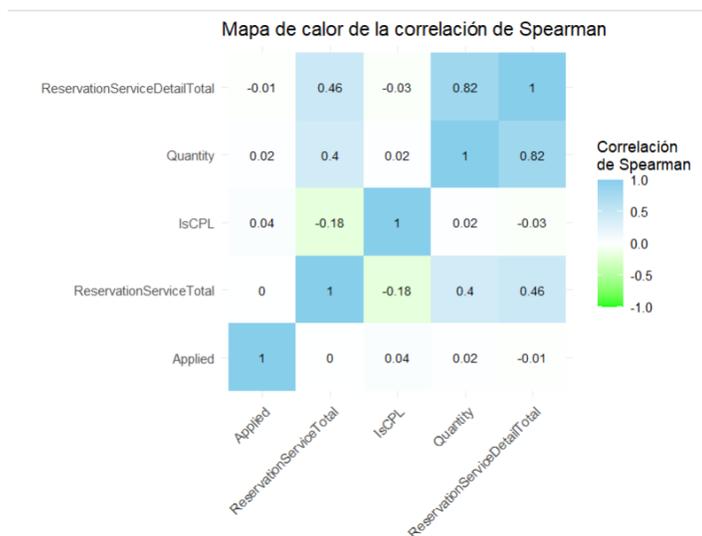
Columna	Estadístico_D	P_value
ReservationServiceTotal	0.22808	4.783198e-194
Quantity	0.27897	4.380111e-290

Nota. Elaboración propia (2024).

La matriz de correlación para variables numéricas que se presenta en la Figura 71 revela relaciones estructurales significativas, destacando una correlación fuerte (0.82) entre cantidad y total detallado del servicio, validando la consistencia en la política de precios.

Figura 70. *Matriz de correlación para las variables numéricas de la tabla*

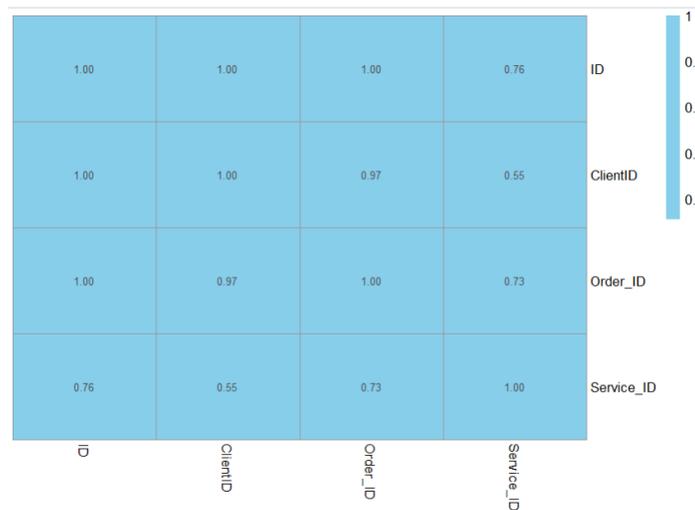
ReservationService.



Nota. Elaboración propia (2024).

La exploración de las relaciones categóricas en reservaciones de servicios, representada en la Figura 71, muestra correlaciones robustas entre identificadores de cliente y orden (0.97), así como entre servicio e identificador de orden (0.73). Estas relaciones sugieren un sistema estructurado de registro que mantiene la trazabilidad de las transacciones, aunque con algunas áreas de mejora en la vinculación de servicios específicos.

Figura 71. Matriz de correlación para las variables categóricas de la tabla ReservationService

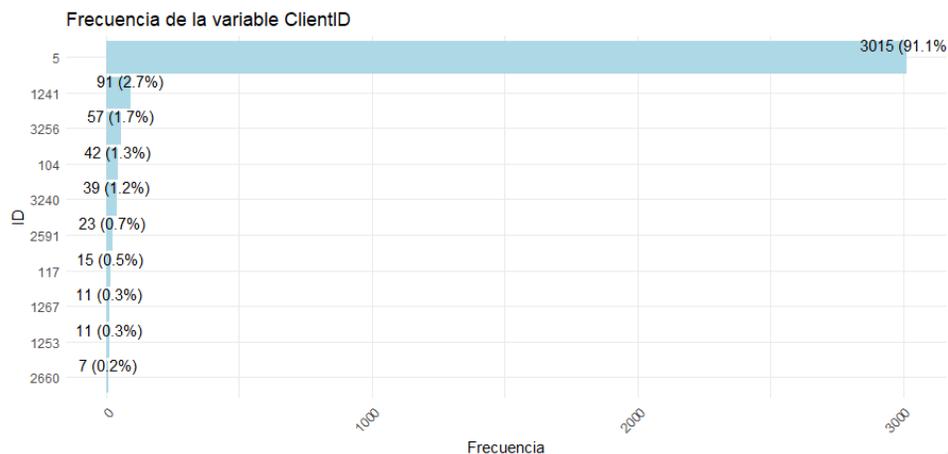


Nota. Elaboración propia (2024).

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

La concentración de clientes en reservaciones de servicios, evidenciada en la Figura 72, mantiene el patrón previamente identificado, con el cliente 5 representando el 91.1% de las reservaciones. Esta persistente concentración a través de diferentes tipos de transacciones confirma la dependencia crítica de un cliente principal en múltiples aspectos del negocio.

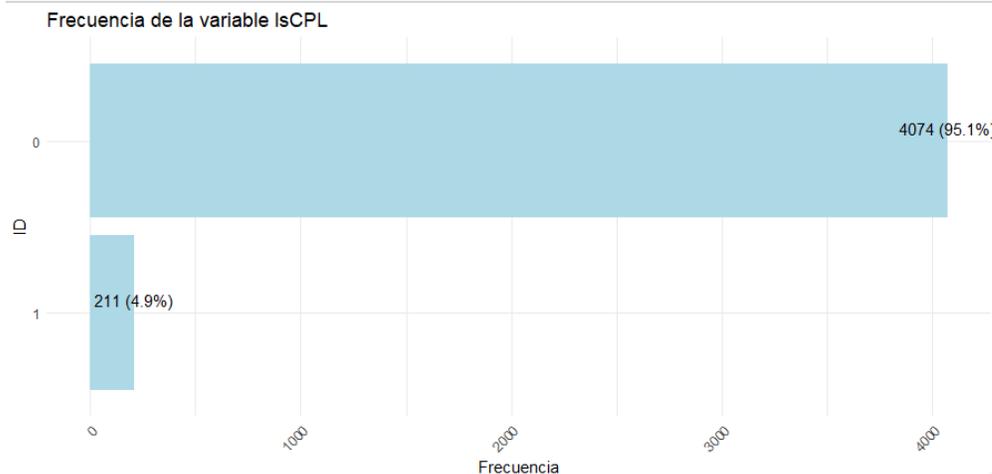
Figura 72. *Exploración de la variable ClientID en ReservationService*



Nota. Elaboración propia (2024).

El análisis de cortesías en reservaciones de servicios, representado en la Figura 73, muestra que solo el 4.9% de las reservaciones incluyen beneficios complementarios. Este porcentaje relativamente bajo sugiere una política selectiva en la asignación de cortesías, aunque también podría indicar una oportunidad perdida en la estrategia de fidelización de clientes.

Figura 73. *Exploración de la variable IsCPL en ReservationService*

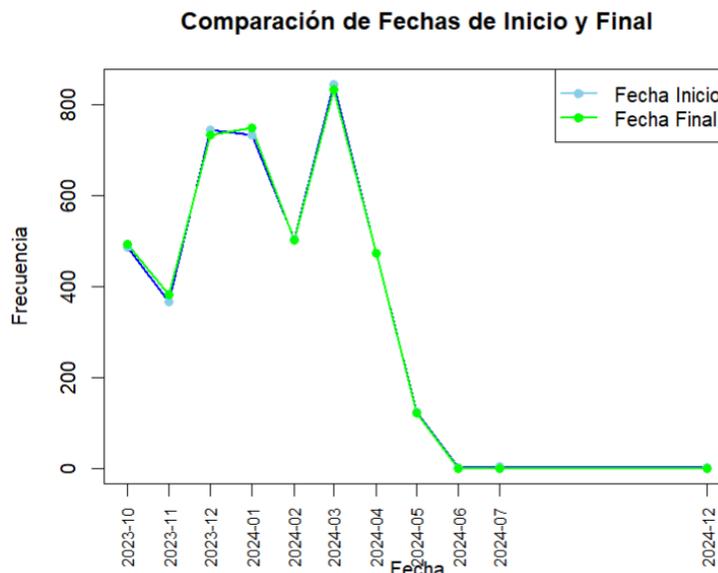


Nota. Elaboración propia (2024).

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Los patrones temporales en reservaciones de servicios, visualizados en la Figura 74, revelan picos de actividad en abril 2024 y diciembre 2023. Esta distribución temporal, que difiere parcialmente de los patrones identificados en reservaciones regulares, sugiere una dinámica estacional específica para servicios adicionales que merece consideración en la planificación operativa.

Figura 74. *Exploración de las fechas en la tabla ReservationService*

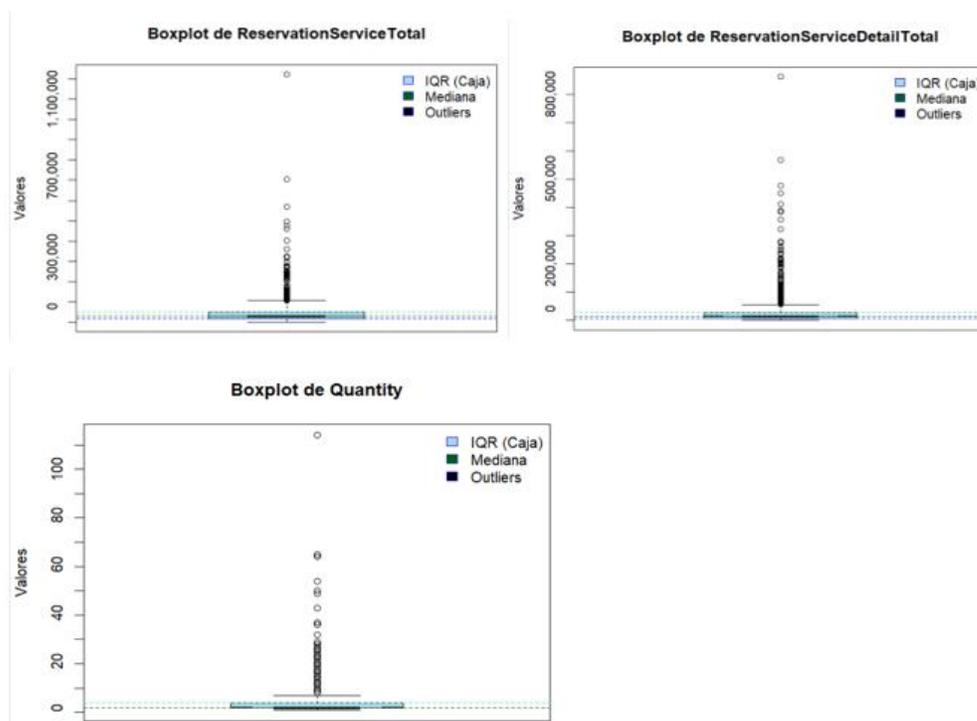


Nota. Elaboración propia (2024).

La distribución de variables numéricas en reservaciones de servicios, ilustrada mediante boxplots en la Figura 75, revela características significativas del comportamiento transaccional. Los montos totales y detallados muestran una distribución similar, con valores atípicos que alcanzan hasta ₡1,000,000, sugiriendo la existencia de servicios premium o paquetes especiales. La variable de cantidad, con una mediana de una hora y valores atípicos hasta 100 horas, indica una predominancia de servicios de corta duración con casos excepcionales de uso extensivo.

Figura 75. *Boxplots de las variables numéricas de la tabla ReservationService*

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM



Nota. Elaboración propia (2024).

4.2.3.8 Exploración de la Tabla de Reserva de Habitaciones (Log_Stays)

En el análisis de reserva de habitaciones que se presenta en la Figura 76, las pruebas de normalidad revelan una distribución no gaussiana en las variables financieras ($p < 0.05$), característica típica en datos transaccionales que sugiere la necesidad de aproximaciones no paramétricas para el análisis posterior.

Figura 76. Matriz de correlación para las variables numéricas de la tabla Log_Stays.

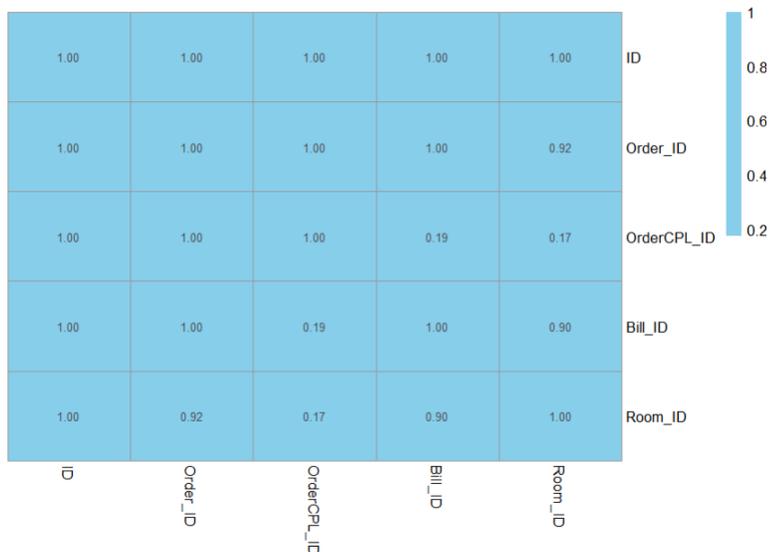
Columna	Estadistico_D	P_value
Qty_Guests	0.34397	5.829984e-102

Nota. Elaboración propia (2024).

El análisis de correlaciones numéricas en el registro de estadias de la Figura 77, revela una estructura de datos robusta. Las correlaciones perfectas (1.0) entre identificadores de factura, orden y órdenes complementarias, junto con correlaciones fuertes (0.90-0.92) con identificadores de habitación, sugieren un sistema integrado de registro que mantiene la trazabilidad de las transacciones a través de diferentes puntos de servicio.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

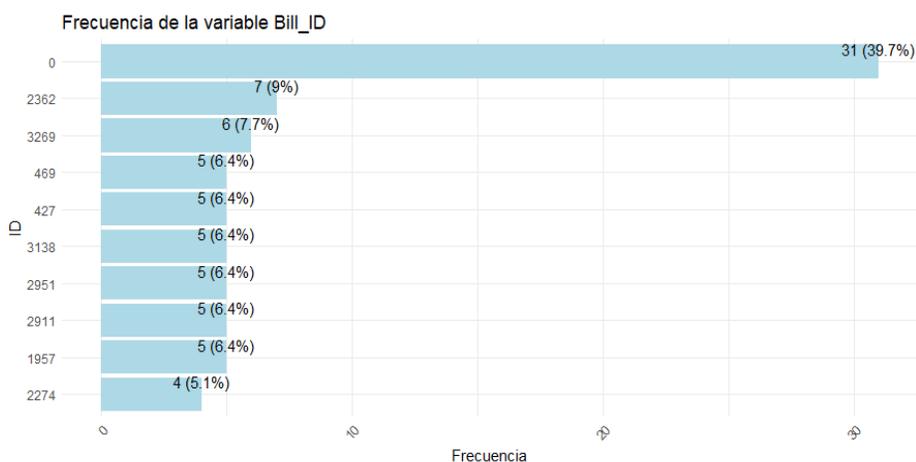
Figura 77. Matriz de correlación para las variables categóricas de la tabla Log_Stays



Nota. Elaboración propia (2024).

La exploración de la facturación en el registro de estadias, visualizada en la Figura 78, revela un hallazgo significativo: el 39.7% de las estadias carecen de registro de facturación asociado. Esta brecha en la vinculación entre estadias y facturación sugiere potenciales ineficiencias en el proceso de captura de ingresos.

Figura 78. Exploración de la variable Bill_ID en Log_Stays



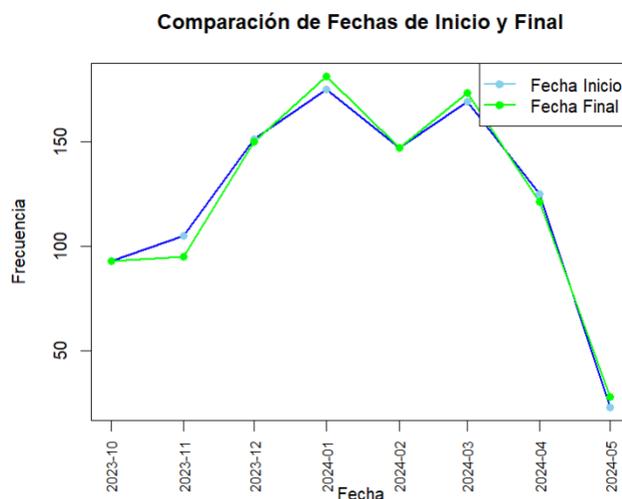
Nota. Elaboración propia (2024).

El análisis temporal de las estadias, representado en la Figura 79, confirma los patrones estacionales previamente identificados, con enero y

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

marzo de 2024 emergiendo consistentemente como los períodos de mayor actividad. Esta consistencia a través de diferentes métricas fortalece la validez de los patrones estacionales identificados y proporciona una base sólida para la planificación operativa y la gestión de recursos.

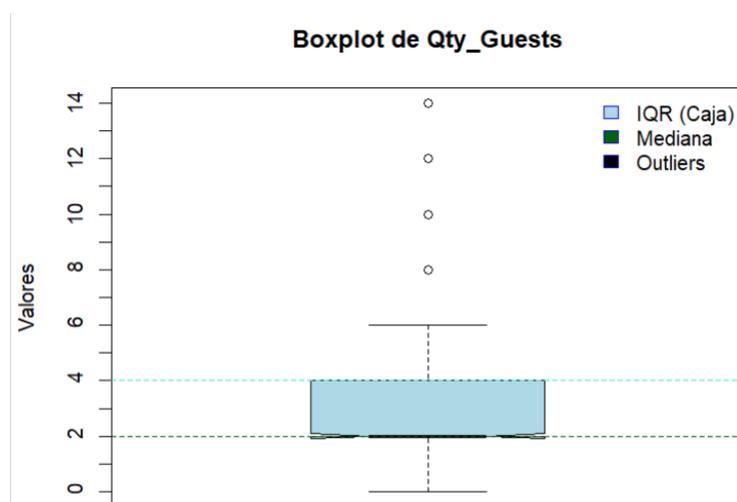
Figura 79. Exploración de las fechas en la tabla *Log_Stays*



Nota. Elaboración propia (2024).

La distribución de huéspedes por estadia, ilustrada mediante el boxplot en la Figura 80, revela un patrón característico del perfil de ocupación. La mediana de dos huéspedes por estadia, con un rango intercuartílico entre 2 y 4 huéspedes, sugiere una orientación predominante hacia parejas y grupos familiares pequeños. La presencia de valores atípicos que alcanzan hasta 14 huéspedes indica la capacidad del hotel para acomodar grupos más grandes, aunque estos casos representan excepciones al patrón general de ocupación.

Figura 80. Boxplot de la variable *Qty_Guests* de la tabla *Log_Stays*



Nota. Elaboración propia (2024).

4.2.4 Verificación de la calidad de los datos

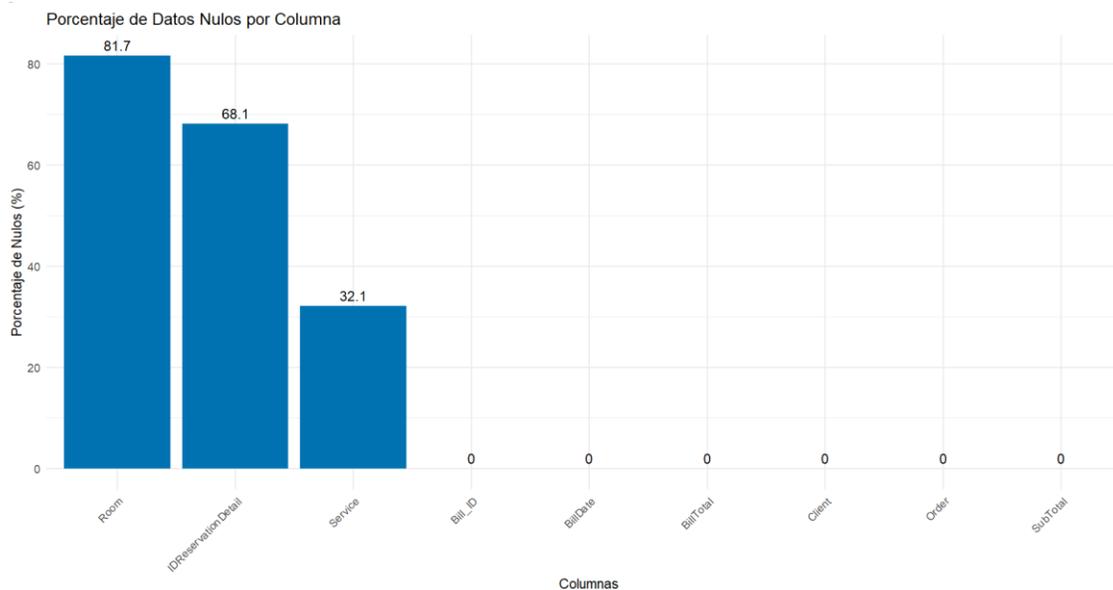
El análisis sistemático de la calidad de datos, realizado mediante la biblioteca Data Explorer de R, revela patrones significativos en la integridad y consistencia de la información a través de las diferentes tablas seleccionadas en la **Sección 4.2.1.1**.

4.2.4.1 Tabla Bill

La exploración de la tabla Bill, visualizada en la Figura 81, expone desafíos significativos en la completitud de datos críticos para el análisis. La variable Room presenta un alarmante 81.7% de datos faltantes, mientras que Service muestra un 68.1% de ausencias. Este patrón de valores faltantes en campos fundamentales sugiere una brecha significativa en el registro de la relación entre facturación y servicios específicos, comprometiendo la capacidad de realizar análisis de rentabilidad por unidad de servicio.

Figura 81. Porcentaje de valores nulos en la tabla Bill

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

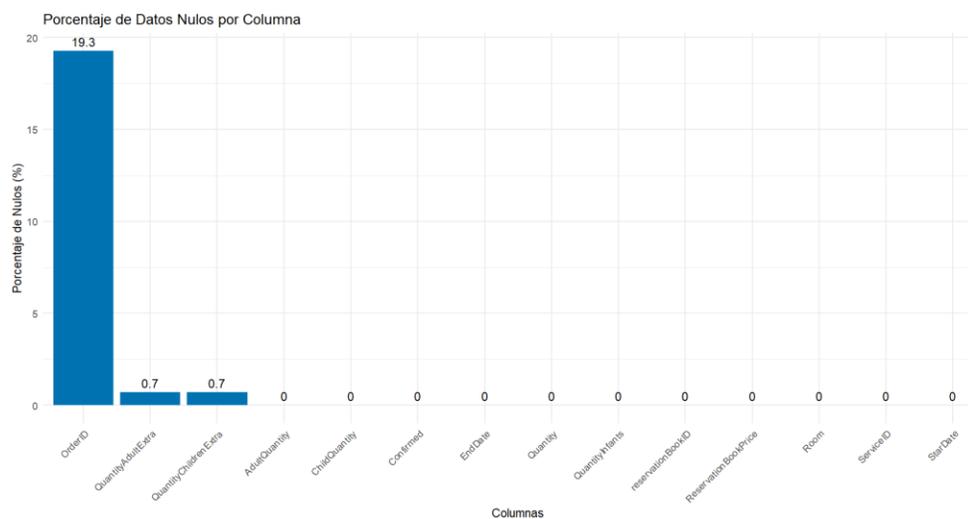


Nota. Elaboración propia (2024).

4.2.4.2 Tabla ReservationBook

El análisis de calidad en la tabla ReservationBook, representado en la Figura 82, revela una situación más favorable. Con solo un 19.3% de valores faltantes en OrderID y un marginal 0.7% en datos de huéspedes adicionales, esta tabla proporciona una base más sólida para el análisis de patrones de reserva. La alta completitud en los campos principales sugiere un proceso robusto en el registro de reservaciones, aunque la vinculación con órdenes presenta oportunidades de mejora.

Figura 82. Porcentaje de valores nulos en la tabla ReservationBook

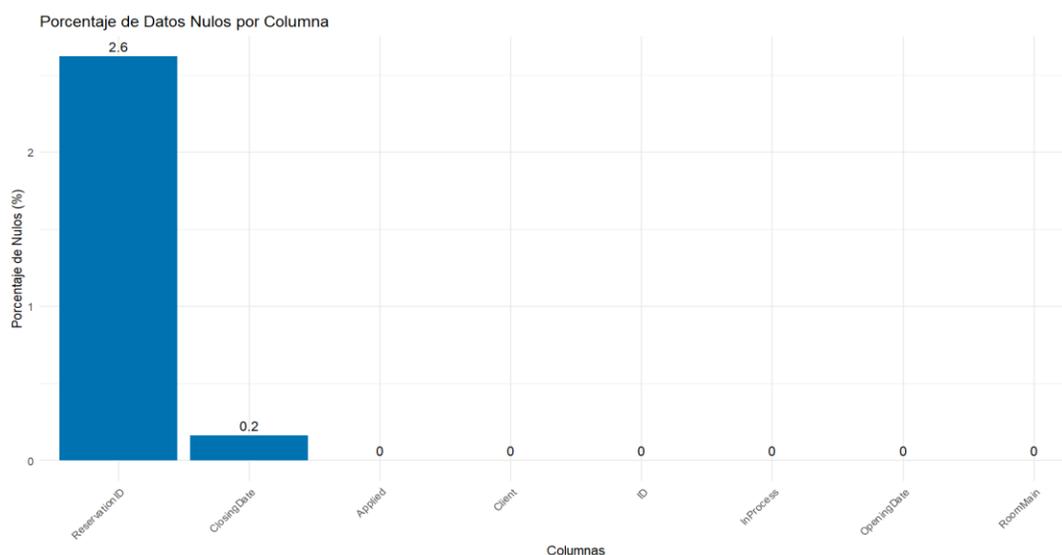


Nota. Elaboración propia (2024).

4.2.4.3 Tabla Order

La evaluación de la tabla Order, ilustrada en la Figura 83, muestra una integridad de datos robusta con apenas un 0.2% de fechas finales faltantes y un 2.6% de ausencias en reservaciones. Esta alta completitud valida la confiabilidad del sistema de gestión de órdenes como fuente primaria para el análisis de patrones transaccionales.

Figura 83. Porcentaje de valores nulos en la tabla Order



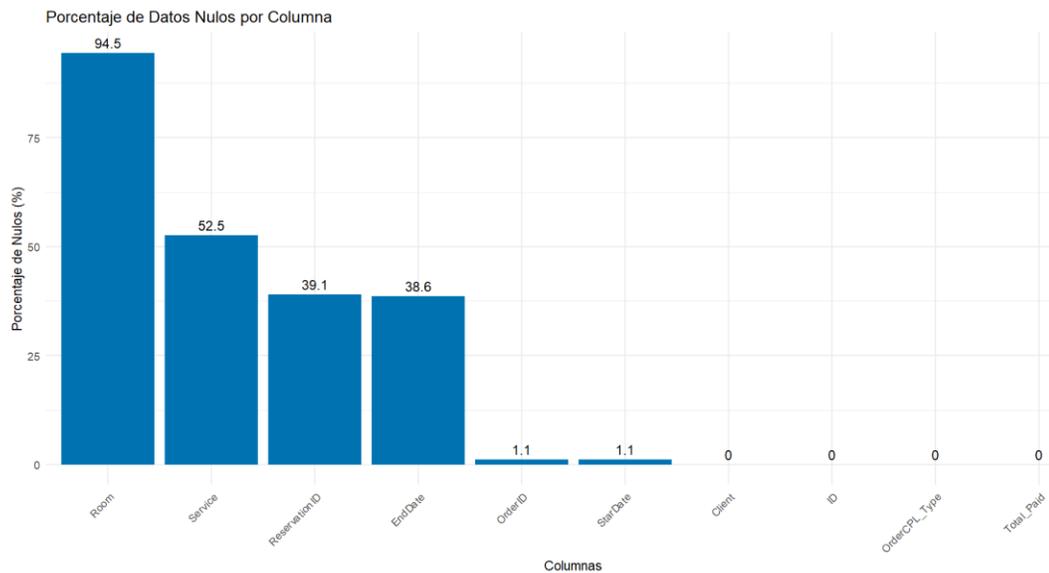
Nota. Elaboración propia (2024).

4.2.4.4 Tabla OrderCPL

El análisis de calidad de la tabla OrderCPL, visualizado en la Figura 84, revela las deficiencias significativas en la integridad de datos del sistema. La asignación de habitaciones muestra un crítico 94.5% de valores nulos, mientras que la identificación de servicios presenta un 52.5% de ausencias. Esta situación se agrava con un 38.6% de fechas finales faltantes y un 39.1% de vinculaciones ausentes con reservaciones. La magnitud de estos vacíos en los datos sugiere una desconexión sistemática en el registro de servicios complementarios, limitando severamente la capacidad de análisis integral de la experiencia del cliente.

Figura 84. Porcentaje de valores nulos en la tabla OrderCPL

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM



Nota. Elaboración propia (2024).

5. Propuesta de Solución

5.1 Preparación de los datos

Esta sección aborda la transformación del conjunto de datos en una estructura refinada y analizable, incluyendo selección, limpieza, formateo, construcción del almacén de datos departamental e integración.

5.1.1 Seleccionar los datos

5.1.1.1 Variables descartadas

Basado en el análisis de la **Sección 4.2.3** y los objetivos de minería definidos en la **Sección 4.1.3.1**, se identificaron las siguientes variables para exclusión:

Tabla 32. *Variables excluidas del análisis*

Tabla	Variable	Justificación de descarte
Bill	IDReservationDetail	Redundante con identificador de orden
Order	ClosingDate	Duplica fecha de inicio
	RoomMain	Redundante con ReservationBook y problemas de calidad
	ReservationID	Redundante con identificadores existentes y problemas de calidad
OrderCPL	StarDate	Redundante con fecha en Order
	EndDate	Duplica fecha de inicio
ReservationService	Quantity	<i>KPIs</i> enfocados en días, no horas
	EndDate	Duplica fecha de inicio
Log_Stays	Qty_Guests	<i>KPIs</i> enfocados en clientes facturados, no en detalle de invitados
ReservationBook	Room	Redundante con Log_Stays
	StarDate	Redundante con fechas de orden
	EndDate	No refleja salida real del cliente (disponible en Log_Stays)
	AdultQuantity	<i>KPIs</i> enfocados en clientes facturados
	ChildQuantity	<i>KPIs</i> enfocados en clientes facturados
	Quantity	<i>KPIs</i> enfocados en días, no horas
	ReservationBookPrice	Redundante con SubTotal en facturas

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla	Variable	Justificación de descarte
ReservationBook	QuantityAdultExtra	KPIs enfocados en clientes facturados
	QuantityChildrenExtra	KPIs enfocados en clientes facturados
	QuantityInfants	KPIs enfocados en clientes facturados
	ServiceID	Redundante con tabla Bill
Service	Price	Redundante con ReservationService y Bill
	serviceCategoryName	Redundante con centro de costo
Room	roomTypeName	Análisis basado en nombre de habitación

Nota. Elaboración Propia (2024)

5.1.2 Limpieza de los datos

Tras el descarte de variables de la **Sección 5.1.1**, que eliminó problemas de valores faltantes en Order, se aplicó un proceso de limpieza según lo acordado en el **Apéndice AF** para las tablas:

- ReservationBook
- Bill
- OrderCPL

Se realizó imputación de variables categóricas con el valor "No Registrado" en todas las tablas.

Figura 85. Porcentaje de valores nulos después de la limpieza de datos



Nota. Elaboración propia (2024).

5.1.3 Integrar los datos

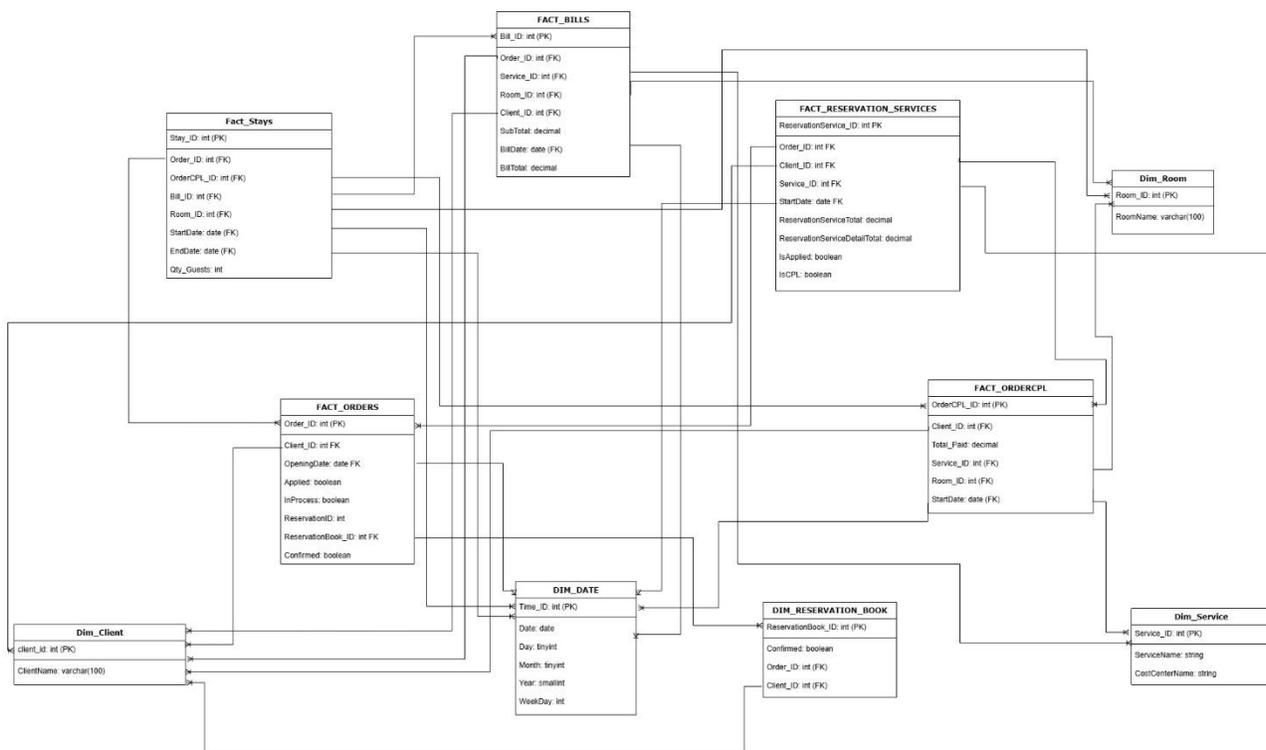
El proceso de integración de datos sigue los principios del ciclo de vida de Kimball, particularmente en su enfoque bottom-up para la construcción del data mart. Este enfoque resultó especialmente apropiado dado el alcance específico del proyecto en el sector hotelero y la necesidad de resultados tangibles en el corto plazo.

Siguiendo la metodología Kimball, se diseñó un esquema dimensional que captura las principales dimensiones del negocio hotelero, representado en la Figura 86. Este modelo dimensional facilita el análisis desde múltiples perspectivas analíticas mientras mantiene la simplicidad necesaria para consultas eficientes.

Se realizó imputación de variables categóricas con el valor "No Registrado" en todas las tablas en la **Sección 2.4.4.2**. El diccionario de datos completo con

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

campos seleccionados y reglas ETL se encuentra en el **Apéndice AK**.
 Figura 86. *Diagrama dimensional*



Nota. Elaboración Propia (2024).

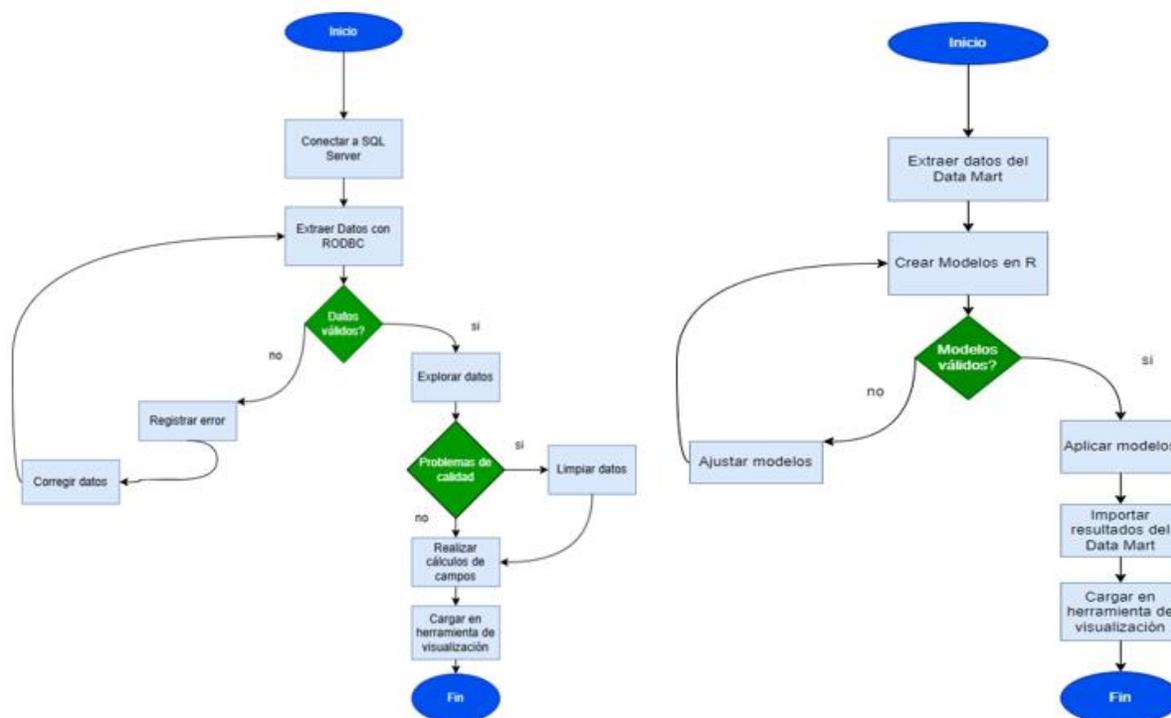
El proceso ETL desarrollado, ilustrado en la Figura 87, implementa las tres fases fundamentales descritas en el marco teórico:

- **Extracción:** El proceso inicia con la extracción de datos desde las tablas fuente identificadas en la fase de análisis. Esta etapa fue particularmente desafiante debido a la estructura compleja de la base de datos y la presencia de relaciones múltiples entre tablas. La extracción se diseñó para mantener la integridad referencial mientras se aislaban los datos relevantes para el análisis hotelero.
- **Transformación:** La fase de transformación aborda las inconsistencias y problemas de calidad identificados durante el análisis exploratorio. Las transformaciones implementadas incluyen la estandarización de formatos, la resolución de inconsistencias en nomenclatura de servicios, y la derivación de métricas agregadas necesarias para los KPIs hoteleros.
- **Carga:** La etapa final del proceso ETL popula el data mart diseñado según el modelo dimensional. Esta fase implementa validaciones para asegurar la integridad de los datos cargados y mantiene la trazabilidad entre las fuentes originales y las estructuras dimensionales.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

El proceso ETL resultante no solo aborda las necesidades inmediatas del análisis, sino que establece una base sólida para la evolución futura del sistema analítico, alineándose con los principios de escalabilidad y mantenibilidad propuestos por Kimball.

Figura 87. Diagramas de flujo del proceso ETL



Nota. Elaboración propia (2024).

5.1.4 Construcción de datos

5.1.4.1 Atributos derivados

A continuación, se muestra la Tabla 33 con el detalle de los atributos derivados de las variables seleccionadas en la **Sección 5.1.2**.

Tabla 33. Atributos derivados del análisis

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Nombre del Campo	Descripción	Fórmula/Método de Cálculo	Tipo de Dato	Propósito en el Análisis
RPC (Revenue Per Client)	Ingreso total generado por cada cliente	sum (BillTotal) agrupado por Cliente	Numérico	Medir valor monetario de cada cliente
Client_Visits_Frequency	Número de visitas únicas por cada cliente	n_distinct (ID) agrupado por Cliente	Numérico	Evaluar lealtad basada en frecuencia
AverageDuration	Duración promedio de estancia	difftime (EndDate, StartDate)	Numérico (días)	Analizar patrones de estancia
Growth	Porcentaje de crecimiento mensual	(BillTotal / lag (BillTotal) - 1) * 100	Numérico (%)	Analizar tendencias
DayOfWeek	Día de la semana	weekdays (Registered_Date)	Categorico	Identificar patrones semanales

Nota. Elaboración Propia (2024)

5.1.4.2 Campos generados

A continuación, se muestra la Tabla 34 con el detalle de los atributos generados a partir de las variables derivadas la **sección 5.3.3.1**.

Tabla 34. Campos generados para el análisis

Nombre del Campo	Descripción	Método de Cálculo	Tipo de Dato	Propósito en el Análisis
Type	Categorización de reserva (Habitación/Servicio)	Asignado manualmente en la preparación de datos	Categorico	Diferenciar entre reservas de habitaciones y servicios
Cluster_ID	Identificador de cluster para segmentación de clientes	Generado por algoritmo de clustering	Categorico	Agrupar clientes similares
Unique_Reservation_ID	Identificador único para cada reserva	Generado secuencialmente	Categorico	Identificación unívoca

Nota. Elaboración Propia (2024)

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

5.1.5 Formatear los datos

Esta etapa final de preparación verifica los requisitos específicos de formato para las técnicas seleccionadas. Aunque varios algoritmos realizan ordenaciones internas, el preprocesamiento ayuda a mejorar tiempos de ejecución.

En la Tabla 35 se detallan los requisitos de formato y las acciones de preparación para cada uno de los modelos propuestos en este estudio:

Tabla 35. *Requisitos de formato para los modelos*

Operación de Formateo	Campo(s) Afectado(s)	Descripción	Método/Función	Propósito en el Análisis
Unión de Tablas	FACT_BILLS DIM_DATE DIM_SERVICE DIM_ROOM FACT_STAYS	Combinación de datos de facturas con servicios, habitaciones y fechas	inner_join () left_join ()	Crear conjunto integral para análisis
Transformación de Fechas	Registered_Date	Formateo de fechas	floor_date (), format () weekdays ()	Facilitar análisis temporal
Escalar datos	RPC Client_Visits_Frequency AverageDuration	Estandarización	Scale ()	Optimizar K-means
Ordenamiento de Datos	Registered_Date	Orden cronológico	arrange()	Preparar series temporales
Manejo de Valores Faltantes	Varios campos numéricos provenientes de las uniones	Exclusión de nulos	na.rm = TRUE na.omit ()	Asegurar precisión

Nota. Elaboración Propia (2024)

5.2 Modelado

Este capítulo presenta seis modelos de minería de datos aplicados al análisis hotelero, alineados con los objetivos definidos en la **Sección 4.1.3.1**, buscando proporcionar una visión integral y generar recomendaciones estratégicas.

5.2.1 Seleccionar las técnicas de modelado

Las técnicas de modelado utilizadas para satisfacer los objetivos de minería definidos en la **Sección 4.1.3.1** son las siguientes:

Tabla 36. *Técnicas de modelado implementadas*

Técnica	Descripción	Supuestos
K-means (3 clústeres)	Segmentación de clientes basada en <i>RPC</i> , frecuencia, valor y duración.	Variables en misma escala. Las variables seleccionadas son relevantes para la segmentación de clientes. El número óptimo de clústeres determinado visualmente o mediante métodos heurísticos.
Análisis Diagnóstico de Datos	Análisis temporal de evolución de servicios.	Datos continuos y ordenados. Muestra representativa. La agregación mensual captura adecuadamente las tendencias.
Análisis Diagnóstico de Datos	Análisis de causas subyacentes en patrones y tendencias de habitaciones.	Categorías y períodos relevantes para el análisis.
Análisis Diagnóstico Comparativo de Reservas	Examina patrones de reservas mediante visualizaciones de tendencias temporales y semanales.	Los datos son representativos del periodo analizado. La agregación por mes y día de la semana es suficiente para revelar patrones relevantes.

Nota. Elaboración Propia (2024)

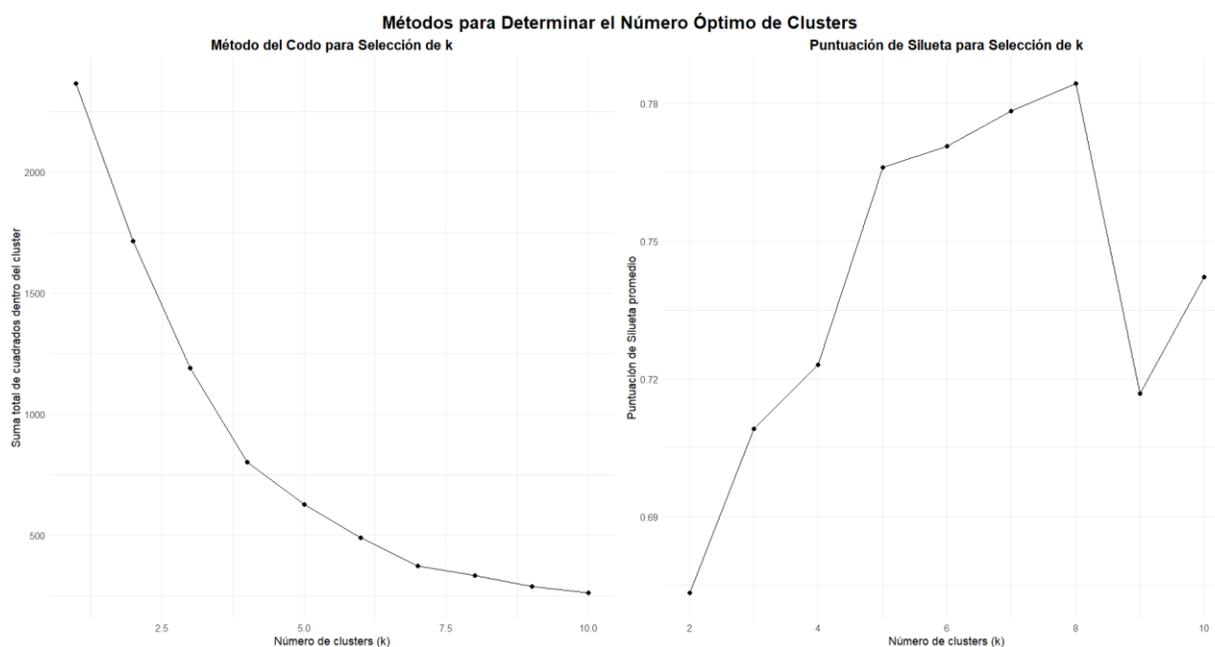
5.2.2 Diseño de prueba

La evaluación de los criterios de éxito planteados en la **Sección 4.1.3.1**, se centra en la identificación de patrones y obtención de *insights* sobre clientes y servicios, priorizando el diagnóstico del estado actual sobre métricas predictivas. Se utilizó el conjunto completo de datos segmentado por análisis, permitiendo un diagnóstico comprensivo.

5.2.3 Evaluar el modelo

5.2.3.1 Modelo 1: Diagnóstico de Segmentación de Clientes

Figura 88. *Evaluación del Codo y Silueta*



Nota. Elaboración propia (2024).

La determinación del número óptimo de clústeres se realizó mediante dos métodos complementarios: el Método del Codo y la Puntuación de Silueta. Los resultados de ambos métodos se analizaron para tomar una decisión informada sobre la estructura apropiada para los datos.

El Método del Codo reveló una disminución significativa en la suma de cuadrados *intra-clúster* hasta alcanzar tres clústeres. A partir de este punto, la tasa de disminución se atenuó considerablemente, formando el característico "codo" en la gráfica. Esta inflexión sugiere que la adición de mayor cantidad de clústeres proporciona rendimientos decrecientes en términos de la varianza explicada por el modelo.

Por otro lado, la Puntuación de Silueta alcanzó su valor máximo con ocho clústeres. Este resultado indica que, desde la perspectiva de la cohesión interna de los clústeres y la separación entre ellos, una solución de ocho grupos proporciona la mejor estructuración de los datos.

Ante la discrepancia entre los resultados de ambos métodos, se optó por una solución de compromiso. Se seleccionó una configuración de tres clústeres basándose en las siguientes consideraciones:

- Tres clústeres ofrecen una segmentación interpretable y manejable de los datos, respetando la inflexión significativa observada en el análisis del codo.
- Aunque no es óptima según la Puntuación de Silueta, la solución de tres clústeres mantiene un rendimiento satisfactorio en ambas métricas, equilibrando la necesidad de segmentar al menos tres comportamientos del cliente

5.2.3.2 Modelos restantes

Los tres modelos restantes presentados en este análisis se basan en una serie temporal de 7 meses. Esta limitación temporal restringe la aplicabilidad de algunos métodos estadísticos tradicionales que típicamente requieren series de mayor extensión para ser plenamente efectivos. No obstante, este enfoque proporciona valiosos *insights* sobre las tendencias recientes y patrones emergentes en diversos aspectos del negocio hotelero.

Es importante señalar que, debido a la cantidad limitada de datos, las métricas de evaluación de los modelos (como el MAPE - Error Porcentual Absoluto Medio) no alcanzan los niveles óptimos. Sin embargo, el análisis descriptivo y las visualizaciones generadas permiten identificar claramente patrones significativos y variaciones mes a mes en cada área estudiada.

5.2.4 Ajustes de los parámetros del modelo y los análisis realizados

Tabla 37. *Parámetros del modelo de segmentación*

Parámetro	Método	Proceso de Ajuste	Criterio de selección
Número de Clúster (k)	Método del codo y silueta	Evaluación de k entre 2 y 10	Equilibrio entre reducción de varianza e interpretabilidad
Selección de Variables	Selección iterativa	Prueba de combinaciones de RPC, frecuencia y duración	Priorización de segmentos diferenciados y relevantes
Técnicas de Escalado	Comparación de normalizaciones	Evaluación con y sin escalado	Mejor separación e interpretación de centroides

Nota. Elaboración Propia (2024)

Tabla 38. *Diagnóstico de Tendencias de Servicios*

Parámetro	Método	Proceso de Ajuste	Criterio de selección
Técnicas de Descomposición Estacional	Múltiples métodos de descomposición temporal.	Prueba de descomposiciones aditivas y multiplicativas con diferentes períodos	Selección del método que mejor captura patrones estacionales y proporciona <i>insights</i> accionables
Granularidad Temporal	Experimentación con niveles de agregación.	Comparación de análisis diarios, semanales y mensuales	Equilibrio entre detección de patrones a corto plazo y tendencias generales

Nota. Elaboración Propia (2024)

Tabla 39. *Diagnóstico de ingresos por habitaciones*

Parámetro	Método	Proceso de Ajuste	Criterio de selección
Selección de <i>KPIs</i>	Evaluación comprensiva	Prueba de diferentes conjuntos de indicadores	Priorización de KPI con mayor visibilidad sobre rendimiento operativo y financiero

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Nota. Elaboración Propia (2024)

Tabla 40. *Análisis de Duración de Estancias y Relación con Servicios*

Variable	Método	Proceso de Ajuste	Criterio de selección
Métricas de Comparación	Evaluación de múltiples métricas	Comparaciones basadas en reservas, ingresos y ocupación	Selección de métricas que proporcionan visión integral del rendimiento.
Granularidad del Análisis	Ajuste de nivel de detalle	Experimentación con diferentes niveles de agrupación	Equilibrio entre detalle informativo y claridad de presentación.

Nota. Elaboración Propia (2024)

5.3 Evaluación

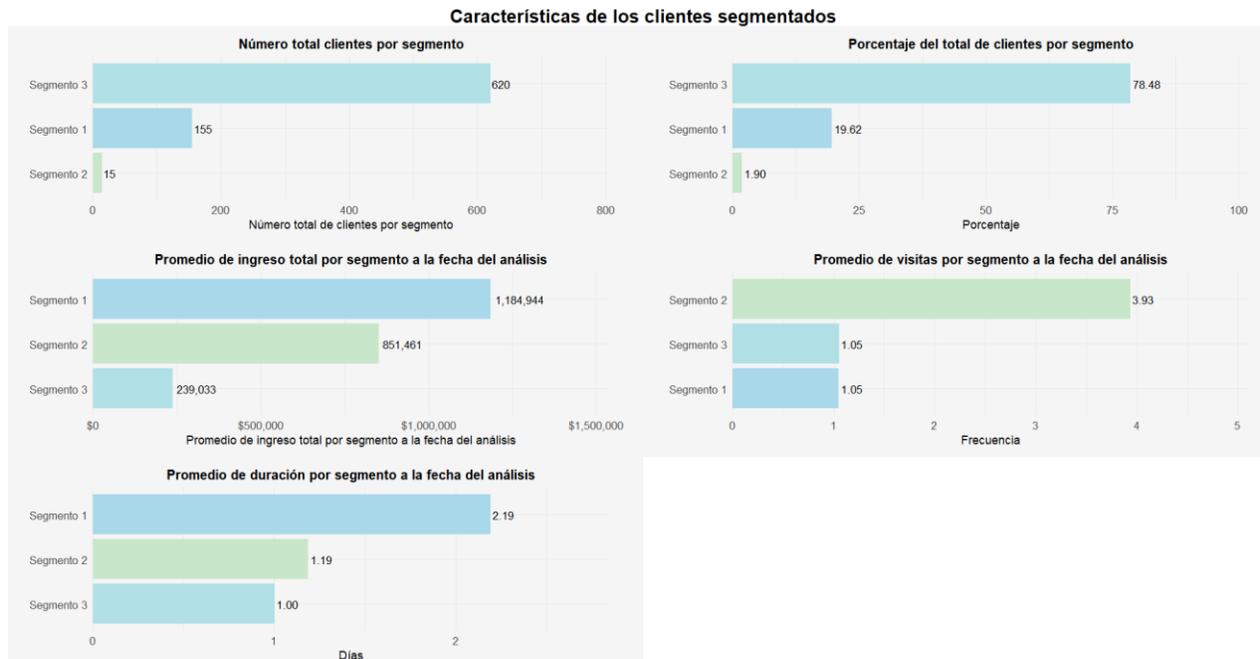
5.3.1 Evaluar resultados

La evaluación contempla tanto la precisión técnica como la relevancia para los objetivos comerciales establecidos. Los modelos han demostrado efectividad en proporcionar *insights* sobre los objetivos planteados.

Modelo 1: Diagnóstico de Segmentación de Clientes

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Figura 89. *Modelo de segmentación de clientes*



Nota. Elaboración propia (2024).

El modelo utiliza K-means para segmentar clientes según tres variables clave:

- RPC (Revenue Per Client): gasto total del cliente
- Client_Visits_Frequency: frecuencia de visitas
- AverageDuration: duración promedio de estancia

Se identificaron con éxito tres segmentos de clientes:

Segmento 1 (Naranja):

- 155 clientes (19.62% del total)
- Máximo RPC Medio: \$1,184,544
- Frecuencia de visitas media: 1.05
- Máxima duración promedio de estancia: 2.19 días

Segmento 2 (Azul):

- 15 clientes (1.90% del total)

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- RPC Medio: \$651,461
- Máxima frecuencia de visitas: 3.93
- Duración promedio de estancia: 1.19 días

Segmento 3 (Verde):

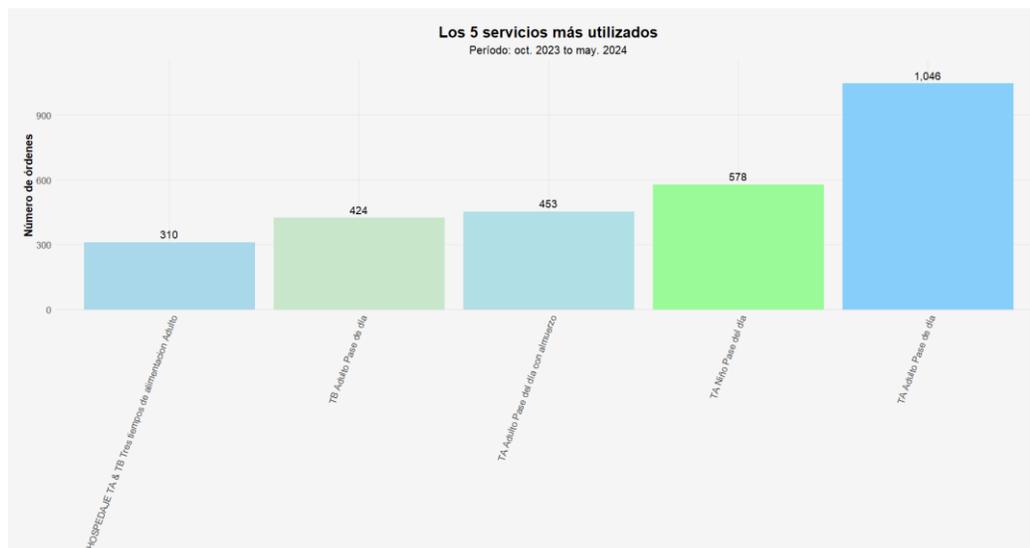
- 620 clientes (78.48% del total)
- Mínimo RPC Medio: \$239,033
- Frecuencia de visitas baja: 1.05
- Menor duración promedio de estancia: 1.00 día

Cada segmento muestra características distintivas en términos de número de clientes, gasto promedio (RPC), frecuencia de visitas y duración de estancia, cumpliendo así con el objetivo de identificar al menos 3 segmentos con perfiles de comportamiento definidos y diferenciados.

Se concluye que el modelo cumplió con los objetivos planteados, proporcionando una visión clara y accionable de los diferentes segmentos de clientes basados en sus patrones de gasto. Esta segmentación no solo identifica grupos distintos, sino que también proporciona *insights*.

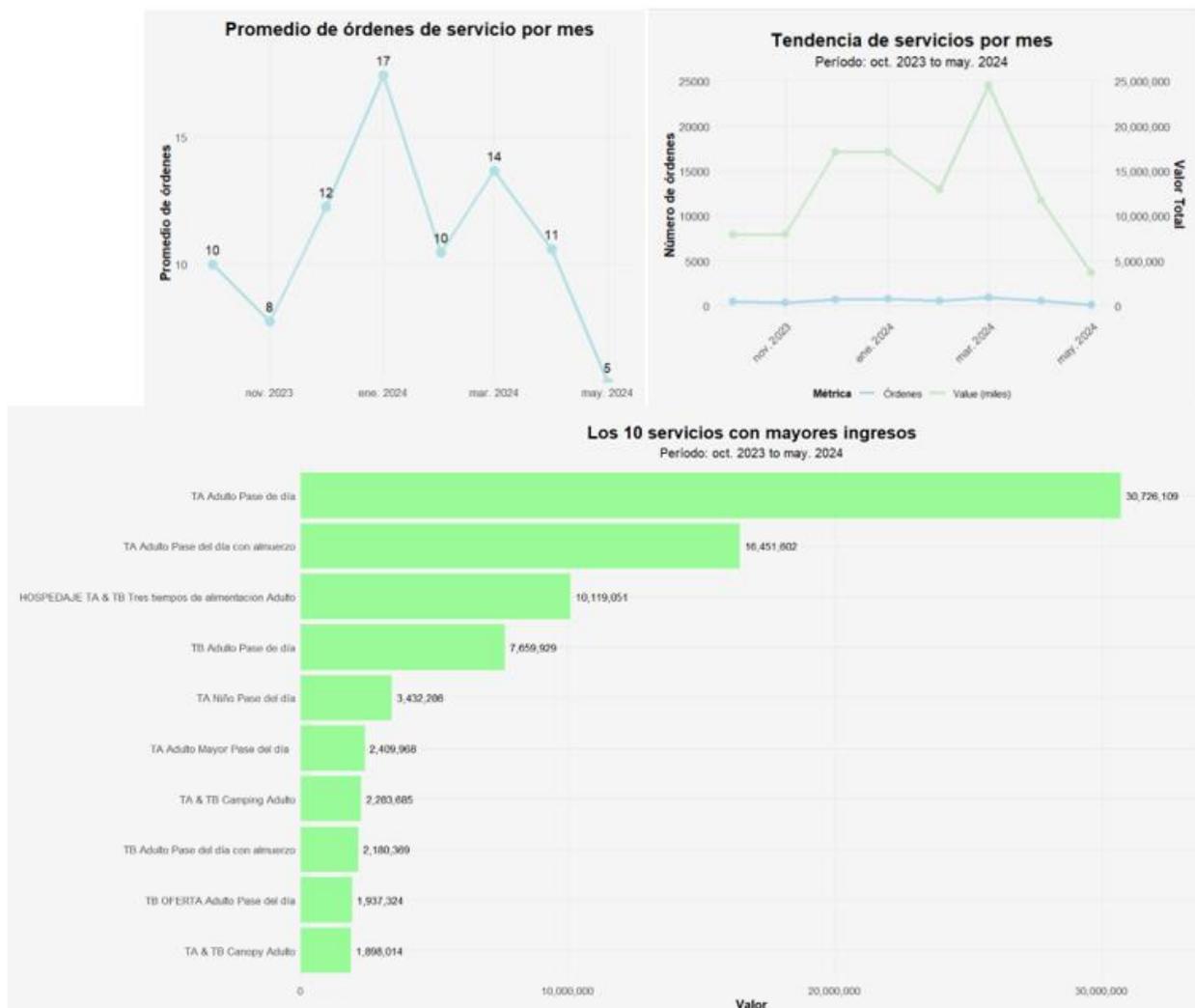
Modelo 2: Diagnóstico de Tendencias de Servicios

Figura 90. *Tendencia de uso de los servicios*



Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Nota. Elaboración propia (2024).
 Figura 91. Modelo de tendencia de consumo de servicios



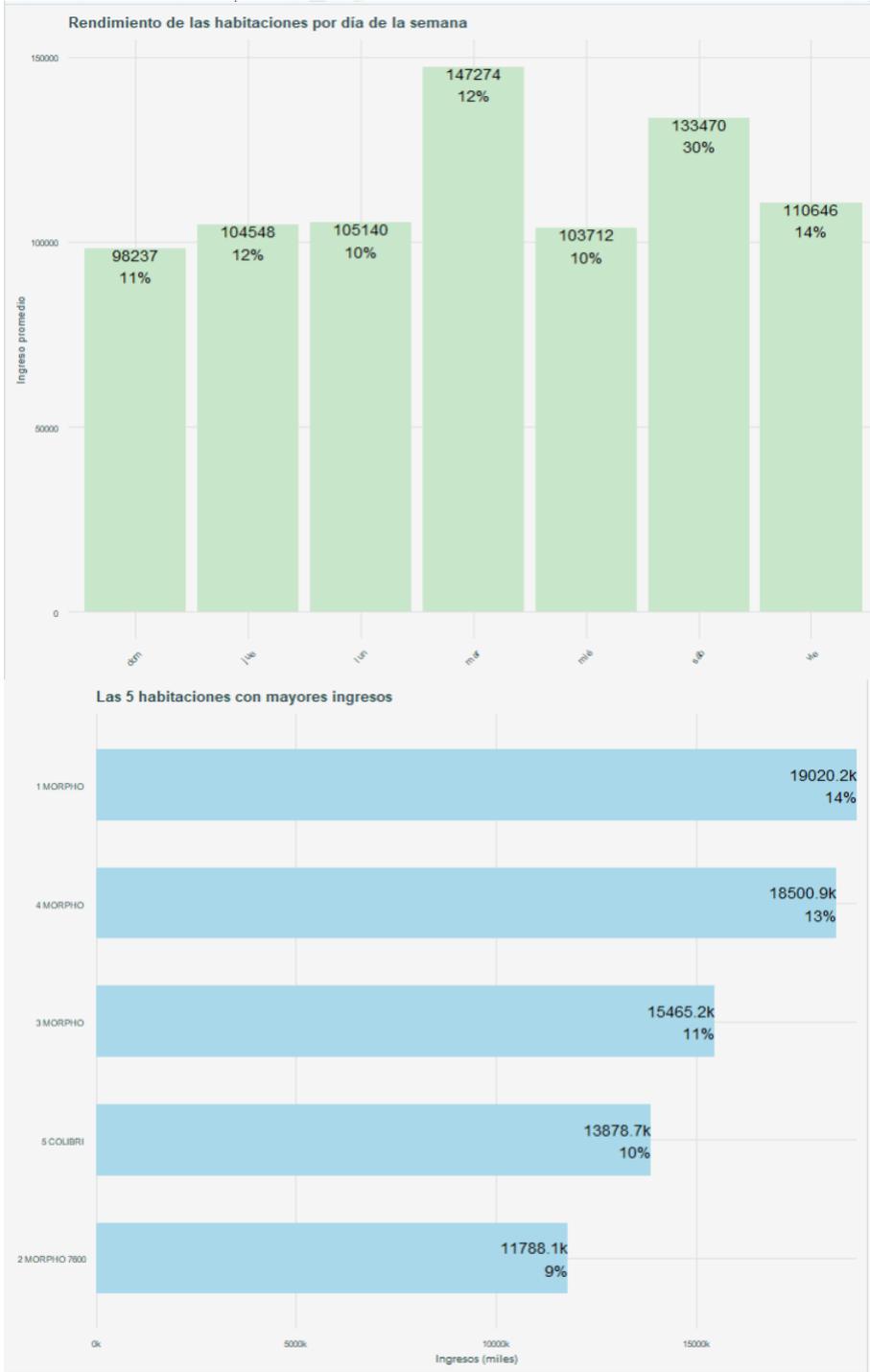
Nota. Elaboración propia (2024).

El modelo presentado logra identificar los 10 servicios con mayor frecuencia en las órdenes de los clientes. Este análisis se sustenta en una serie de visualizaciones gráficas que ofrecen perspectivas complementarias y detalladas sobre el desempeño de los servicios. Se encuentra que el servicio "TA Adulto Pase de día" órdenes tiene la mayor cantidad con 1,046 y es el que genera mayores ingresos con \$30,726,109.

También, un pico notable en marzo 2024. En cuanto al análisis de rentabilidad, se proporciona una visión clara de los 10 servicios con mayores ingresos. También, se identifican las tendencias de consumo de octubre 2023 a mayo 2024, ofreciendo una perspectiva global de la evolución tanto de la cantidad de órdenes como del valor total generado mes a mes.

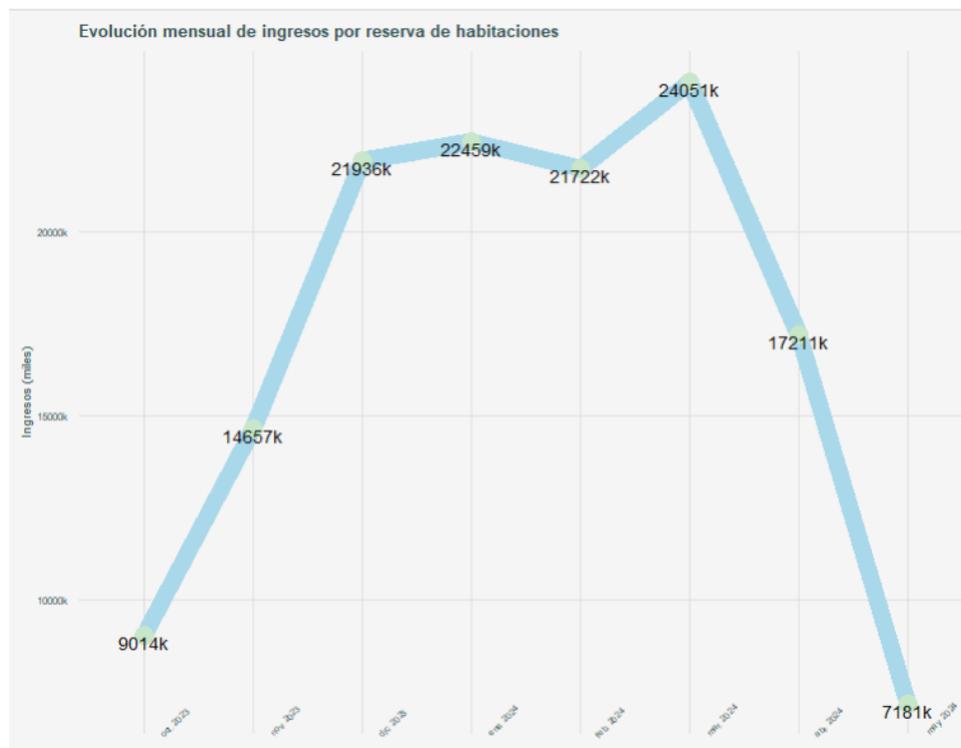
Modelo 3: Diagnóstico de Rentabilidad de Habitaciones

Figura 92. Diagnóstico de rentabilidad de habitaciones



Nota. Elaboración propia (2024).

Figura 93. Evolución mensual de ingresos por reserva de habitaciones



Nota. Elaboración propia (2024).

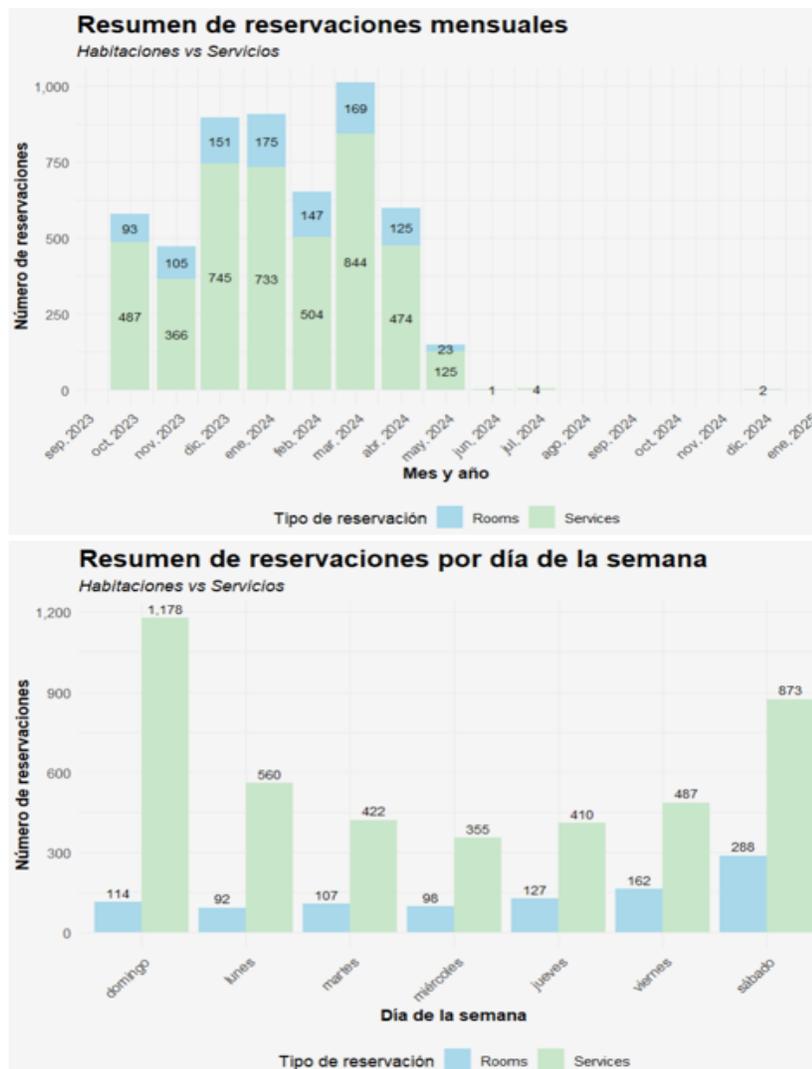
Los gráficos proporcionados analizan los puntos de alza significativos en los ingresos de habitaciones y detectar períodos de aumento significativo. Se muestran los ingresos totales por tipo de habitación. Aunque no proporciona una perspectiva temporal directa, esta visualización es crucial para identificar qué tipos de habitaciones generan mayores ingresos.

Se utiliza un gráfico de líneas donde se visualizan los puntos de alza significativos en los ingresos de habitaciones a lo largo del período. Este último, permite detectar períodos de aumento significativo en los ingresos por habitaciones. Se tiene que las habitaciones de mayor rentabilidad son la "1 MORPHO" y "4 MORPHO", generando el 14% y 13% de los ingresos respectivamente.

También, existe una marcada estacionalidad en los ingresos, con un pico notable en marzo. Los viernes y sábado muestran un rendimiento mayor que el resto de la semana.

Modelo 4: Análisis de Duración de Estancias y Relación con Servicios

Figura 94. Análisis de los tipos de reservas



Nota. Elaboración propia (2024).

Los gráficos proporcionados ofrecen un análisis que permite examinar las variaciones en las reservas de servicios y habitaciones, tanto a nivel mensual como semanal. Este análisis cumple efectivamente con el objetivo de identificar tendencias mensuales y patrones semanales en las reservas. Por ejemplo:

- Pico de reservas en enero 2024 con 1,113 reservas totales (944 servicios, 169 habitaciones).

- Disminución significativa en las reservas a partir de febrero 2024.
- Mayor número de reservas los domingos (1,287 total: 1,173 servicios, 114 habitaciones).
- Sábado es el segundo día de mayor ocupación (1,131 total: 873 servicios, 258 habitaciones).
- Menor actividad de reservas de martes a jueves.

5.3.2 Modelos aprobados

En la siguiente sección se muestran los modelos que cumplen con las condiciones establecidas en la **Sección 5.2.3**.

5.3.2.1 Revisión del proceso

Los modelos desarrollados para el análisis del hotel son principalmente descriptivos y de visualización. Estos incluyen gráficos de barras para mostrar los servicios de mayor rentabilidad, gráficos de líneas para las tendencias mensuales de ingresos y ocupación, y gráficos comparativos para analizar las reservas de servicios y habitaciones por día de la semana.

La revisión del proceso, según la guía de modelado de CRIPS-DM de IBM (s.f.), debe examinar cómo cada fase del *CRISP-DM* contribuyó al desarrollo y la calidad de los modelos finales. En este caso, la fase de comprensión del negocio guio la selección de variables y métricas relevantes para el análisis hotelero. La preparación de los datos, que incluyó la agregación de información por períodos y la creación de nuevas variables como las temporadas, fue crucial para permitir los análisis posteriores.

El proceso de modelado se centró en técnicas de visualización, lo cual es apropiado para los objetivos descriptivos del proyecto. Sin embargo, siguiendo las mejores prácticas de la guía de modelado de CRIPS-DM de IBM (s.f.) se considera la iteración del proceso para incluir técnicas de modelado avanzadas en futuras fases.

La evaluación de los modelos se ha realizado principalmente en términos de su capacidad para proporcionar *insights* útiles para el negocio. Esto está en línea con el enfoque de la guía de modelado CRIPS-DM de IBM (s.f.) que busca asegurar que los resultados del modelado sean accionables y relevantes para los *stakeholders* del negocio.

5.4 Despliegue

El despliegue representa la fase final del proyecto de análisis de datos para el cliente del sector hotelero. En esta etapa, se transformarán los *insights* obtenidos de los modelos analíticos en herramientas visuales y accionables. Se utilizará *DevExpress Report Designer* para crear *dashboards* interactivos que permitirán a los usuarios finales explorar y comprender los patrones de comportamiento de los clientes, las tendencias de servicios y la ocupación de habitaciones.

5.4.1 Objetivos del despliegue

- Visualizar de manera efectiva los segmentos de clientes identificados y sus características.
- Presentar las tendencias de servicios y habitaciones de forma clara y accionable.
- Facilitar el análisis de patrones temporales en reservas y ocupación.
- Proporcionar una herramienta interactiva para la toma de decisiones estratégicas.

5.4.2 Plan de despliegue

La Tabla 41 presenta una visión integral del plan de despliegue para el proyecto de análisis de datos en el sector hotelero. Esta consolida los elementos clave, incluyendo los resultados principales de los modelos realizados, los *KPIs* y métricas seleccionadas, y el diseño propuesto para los *dashboards* en *DevExpress Report Designer*.

Tabla 41. Descripción de *dashboards* a implementar

<i>Dashboard</i>	Resultados Clave	<i>KPIs/ Métricas clave</i>	Elementos visuales	Interactividad
Segmentación de Clientes	Tres segmentos claramente diferenciados basados en RPC, frecuencia de visitas y duración de	Distribución porcentual de clientes por segmento RPC medio por segmento Frecuencia de	Gráfico circular para la distribución de clientes por segmento Gráficos de barras para comparar RPC, frecuencia de visitas y duración de estancia entre segmentos	Filtros para seleccionar segmentos específicos <i>Tooltips</i> al pasar el cursor sobre los gráficos

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Dashboard	Resultados Clave	KPIs/ Métricas clave	Elementos visuales	Interactividad
	estancia.	visitas media por segmento Duración promedio de estancia por segmento	Tabla con características de cada segmento	
Análisis de Servicios	Identificación de los 10 servicios de mayor frecuencia y rentabilidad. Patrones estacionales en la demanda de servicios.	Top 10 servicios por frecuencia de orden Top 10 servicios por ingresos generados Tendencia mensual de órdenes y valor total de servicio	Gráfico de barras horizontal para los 10 servicios por frecuencia y valor Gráfico de líneas para la tendencia mensual de servicios Gráfico de área apilada para la comparación mensual entre cantidad de órdenes y valor total	Selector de rango de fechas Filtros por categoría de servicio <i>Tooltips</i> al pasar el cursor sobre los gráficos
Rentabilidad de Habitaciones	Identificación de las habitaciones de mayor rentabilidad. Patrones temporales en los ingresos por habitaciones.	Ingresos totales por tipo de habitación Tendencia temporal de ingresos por habitaciones Distribución de ingresos por día de la semana	Gráfico de anillos para la distribución de ingresos por tipo de habitación Gráfico de líneas para la tendencia temporal de ingresos Gráfico de barras para la distribución de ingresos por día de la semana	Filtros por tipo de habitación y rango de fechas <i>Tooltips</i> al pasar el cursor sobre los gráficos
Patrones de Reservas	Tendencias mensuales y semanales en reservas de servicios	Número de reservas mensuales (servicios vs. habitaciones)	Gráfico de barras apiladas para reservas mensuales (servicios vs. habitaciones)	Selector de rango de fechas Selector de tipos de reserva

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

<i>Dashboard</i>	Resultados Clave	KPIs/ Métricas clave	Elementos visuales	Interactividad
	y habitaciones .	Distribución semanal de reservas	Gráfico de calor para la distribución semanal de reservas	<i>Tooltips</i> de reservas al pasar el cursor

Nota. Elaboración Propia (2024).

5.4.3 Integración de Datos

En la fase de despliegue, se implementó la creación de tablas físicas en el almacén de datos departamental, en lugar de vistas o estructuras temporales. Esta decisión estratégica se fundamenta en que las tablas físicas proporcionan un rendimiento superior para consultas frecuentes, especialmente al manejar grandes volúmenes de datos característicos del sector hotelero. Adicionalmente, el almacenamiento físico permite mantener un historial consistente, facilitando análisis temporales y comparativos.

La capacidad de optimización constituyó otro factor determinante. Es posible la indexación de tablas para mejorar la velocidad de las consultas analíticas, aspecto vital para la generación eficiente de informes y tableros de control. Asimismo, garantizan una vista consistente de los datos en puntos específicos del tiempo, esencial para la generación de informes precisos y la toma de decisiones fundamentada.

Se estableció un proceso de actualización diario para todas las tablas, programado para horarios de baja actividad, minimizando el impacto en las operaciones hoteleras. Este proceso se fundamenta en un sistema *ETL* automatizado, desarrollado específicamente para cada tabla. El diccionario de datos completo de las diferentes tablas creadas para cada modelo se encuentra en el **Apéndice AJ**

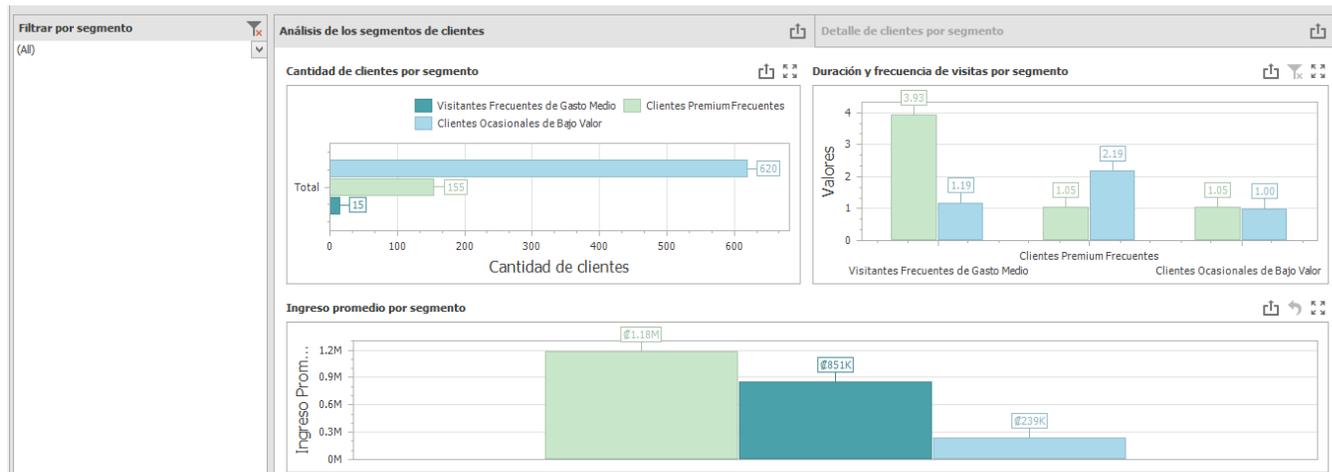
5.4.4 Dashboards realizado en el entorno

La implementación de los tableros siguió rigurosamente el plan de despliegue establecido en la **Tabla 41**, utilizando DevExpress Report Designer como herramienta de visualización. A continuación, se presenta el análisis de cada tablero implementado:

El primer tablero implementa el análisis de segmentación de clientes, integrando una evaluación bidimensional que relaciona la frecuencia de órdenes y la duración de estancias por segmento. Este tablero incorpora análisis comparativos entre segmentos, cuantificación de representatividad y evaluación de contribución económica, junto con la implementación de métricas específicas de rentabilidad.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

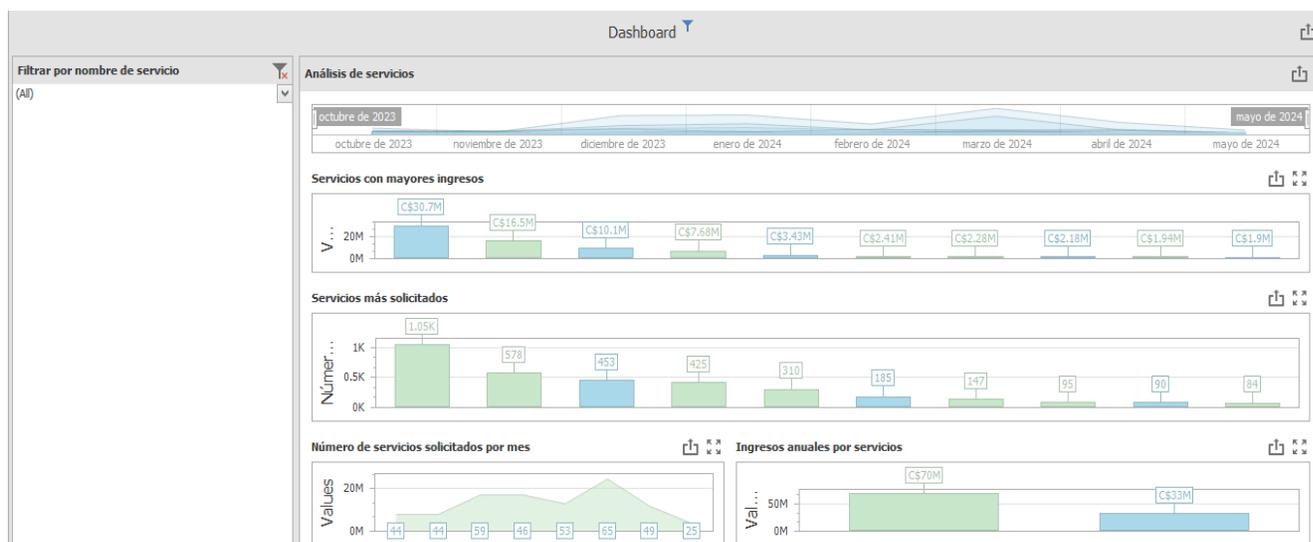
Figura 95. Tablero de análisis de segmentación de clientes



Nota. Elaboración propia (2024).

El segundo tablero desarrolla un sistema de análisis multidimensional de servicios, fundamentado en una clasificación jerárquica por ingresos. Este incorpora la implementación de métricas comparativas, evaluación de frecuencia de utilización, métricas de popularidad y análisis de patrones de consumo mediante series temporales.

Figura 96. Tablero de análisis de servicios

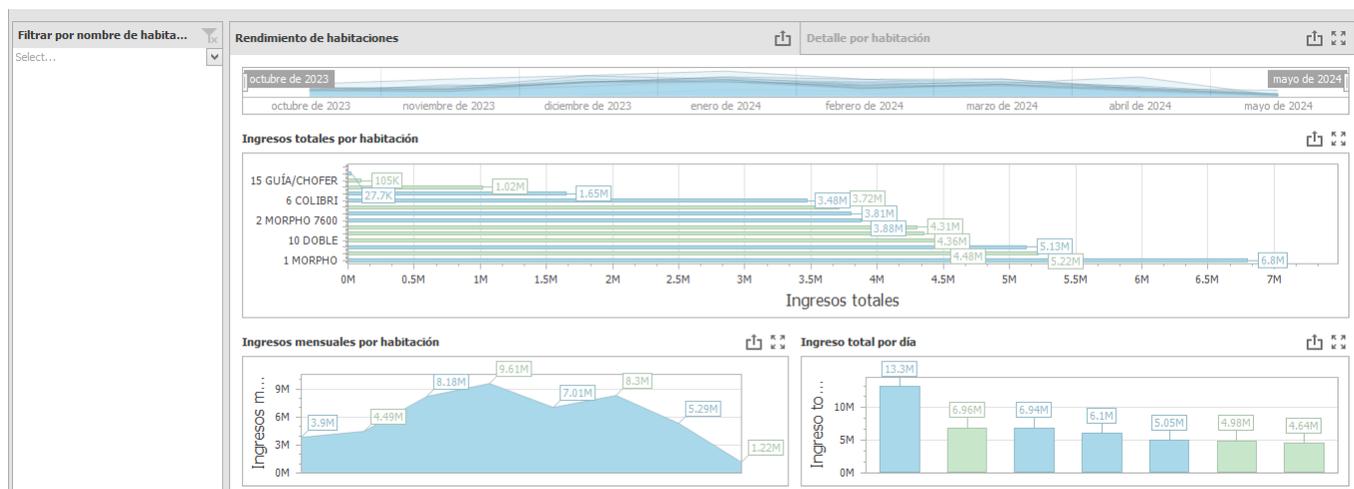


Nota. Elaboración propia (2024).

El tercer tablero se fundamenta en el análisis financiero de las habitaciones vendidas por el hotel. Este contempla la implementación de series temporales, identificación de puntos de inflexión en ingresos, cuantificación porcentual de contribución por habitación, segmentación de ingresos por períodos diarios y visualización de variaciones periódicas.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

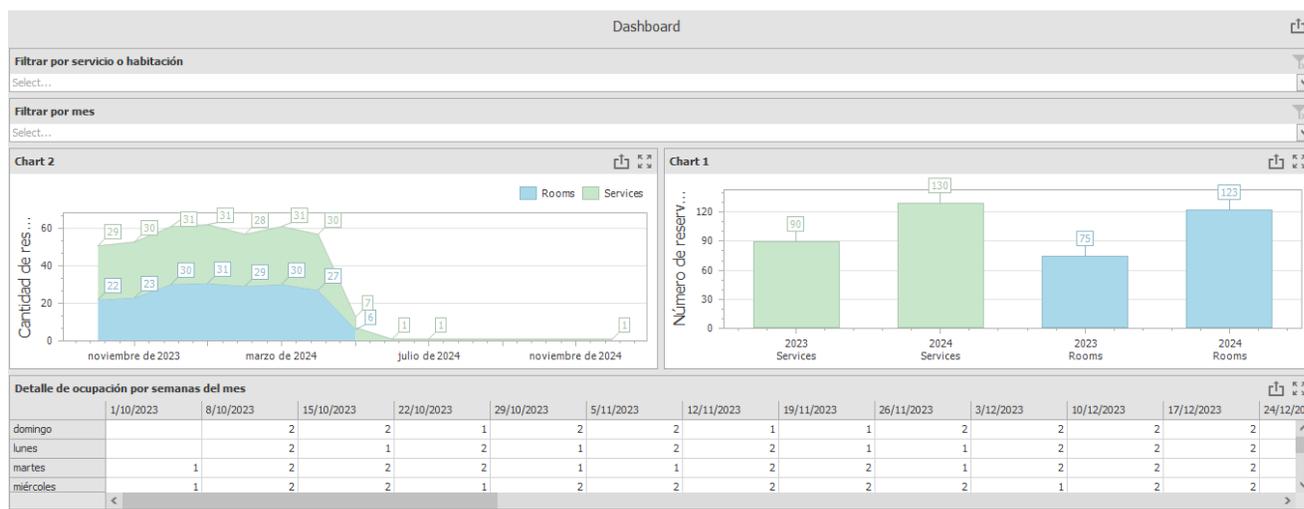
Figura 97. Tablero de análisis de habitaciones



Nota. Elaboración propia (2024).

El cuarto tablero implementa un análisis integrado de ocupación y servicios, respondiendo a la necesidad de monitorizar simultáneamente estas variables críticas en la operación hotelera. Este panel integra índices de utilización de servicios, series temporales longitudinales para el período noviembre 2023 - noviembre 2024, distribución semanal segmentada y patrones de comportamiento diario.

Figura 98. Tablero de análisis de ocupación y servicios



Nota. Elaboración propia (2024).

5.4.5 Recomendaciones

El análisis integral de los cuatro modelos ha generado una serie de recomendaciones estratégicas para optimizar las operaciones y aumentar la rentabilidad del hotel. Estas recomendaciones abordan las necesidades específicas de cada segmento de clientes y capitalizan las tendencias observadas en los patrones de reserva y uso de servicios.

La segmentación de clientes sugiere estrategias diferenciadas por grupo. Para el segmento premium, que representa el 19.62% de clientes de alto valor, se propone la implementación de un programa VIP que incluya acceso prioritario a servicios, mejoras de habitación y atención personalizada. Este programa busca maximizar la retención y el valor generado por este segmento estratégico.

Para los clientes frecuentes, que constituye el 1.90% de la base, requiere un programa de recompensas que incentive y premie la regularidad de visitas. Se recomienda implementar un sistema de descuentos progresivos o acumulación de puntos canjeables por servicios premium, fortaleciendo así la lealtad de este grupo.

Para el segmento de clientes ocasionales, que representa el 78.48% de la base, se recomienda desarrollar campañas de marketing focalizadas en incrementar la frecuencia de visitas. Estas campañas deben incluir ofertas y paquetes diseñados específicamente para convertir visitas esporádicas en hábitos regulares.

El análisis de servicios revela oportunidades centradas en el producto "TA Adulto Pase de día", que lidera tanto en demanda (1,046 órdenes) como en generación de ingresos (€30,726,109). Se recomienda crear un paquete premium para este servicio, incorporando beneficios adicionales que justifiquen un precio superior y aumenten el valor percibido por el cliente.

El análisis temporal indica la pertinencia de desarrollar una campaña promocional específica para marzo, período que muestra un pico significativo en la demanda. Esta campaña debería enfocarse en los cinco servicios de mayor demanda. Para contrarrestar los períodos de baja demanda identificados en mayo y noviembre, se sugiere implementar descuentos estratégicos en servicios como "HOSPEDAJE TA & TB Tres tiempos de alimentación Adulto".

Respecto al rendimiento de habitaciones, el análisis revela que las habitaciones "1 MORPHO" y "4 MORPHO" generan el 27% de los ingresos totales (14% y 13% respectivamente). Para las estas se recomienda:

- Implementar precios dinámicos durante los fines de semana, período de mayor demanda

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- Desarrollar paquetes especiales para marzo, aprovechando el pico estacional identificado
- Crear ofertas específicas para días entre semana, buscando incrementar la ocupación de lunes a jueves

Finalmente, el Modelo 4, que examina las variaciones en las reservas, sugiere implementar una estrategia de precios dinámicos para enero, mes que muestra un pico en las reservas. Para los meses de menor demanda, abril y mayo, se recomienda crear paquetes especiales que incentiven las reservas en estos períodos. Dado que el domingo se identifica como el día con mayor número de reservas, se propone lanzar una promoción específica para este día, posiblemente ofreciendo beneficios adicionales o descuentos para estancias que incluyan la noche del domingo.

5.4.6 Matriz de cumplimiento de requerimientos

Tabla 42. *Matriz de cumplimiento de requerimientos*

ID Requerimiento	Descripción	Evidencia	¿Se cumplió?
RF-001	Visualización de datos con Devexpress Report Designer	Sección 5.4.4	Cumplido
RF-002	Desarrollo de cuadros de mando interactivos que muestren KPIs relevantes	Sección 5.4.4	Cumplido
RF-003	Presentación comprensible de los análisis en las visualizaciones.	Sección 5.4.4	Cumplido
RF-004	Generación de recomendaciones en servicios, paquetes, reservas y habitaciones basadas en los datos recogidos.	Sección 5.4.5	Cumplido
RF-005	Documentación de los nuevos modelos y	Sección 5.2	Cumplido

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

ID Requerimiento	Descripción	Evidencia	¿Se cumplió?
	análisis implementados		
RF-006	Software de análisis estadístico compatible con el gestor de base de datos	Sección 4.1.4	Cumplido
RF-007	Diseño del proceso adaptable a otros segmentos	Sección 5	Cumplido
RF-008	Cuadros de mando con filtros que permitan a los usuarios explorar los datos	Sección 5.4.4	Cumplido

Nota. Elaboración propia (2024).

5.4.7 Matriz de cumplimiento de objetivos de negocio

Tabla 43. *Matriz de cumplimiento de objetivos de negocio*

Objetivo de negocio	Criterio de éxito	Evidencia	¿Se cumplió?
Desarrollar recomendaciones basadas en datos mediante cuadros de mando interactivos y visualizaciones.	Desarrollar cuadros de mando que cubran las áreas clave para los hoteles.	Sección 5.4.5	Cumplido
Identificar indicadores clave de rendimiento (<i>KPIs</i>) específicos para la industria hotelera.	Generar recomendaciones sobre la operación del cliente hotelero.	Sección 4.1.3.2	Cumplido
Identificar patrones y segmentos clave en la base de clientes del hotel.	Desarrollo de un modelo de segmentación de clientes.	Sección 5.3.1	Cumplido
Determinar las tendencias	Cada cuadro de mando contemple indicadores	Sección 5.3.1	Cumplido

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo de negocio	Criterio de éxito	Evidencia	¿Se cumplió?
estacionales de habitaciones y servicios.	claves de rendimiento (<i>KPIs</i>).		
Identificar los servicios de mayor rentabilidad, así como las tendencias de consumo.	Identificación de los servicios y habitaciones que han generado mayor rentabilidad para el hotel.	Sección 5.3.1	Cumplido
	Desarrollo de un análisis de rentabilidad por servicio	Sección 5.3.1	Cumplido

Nota. Elaboración propia (2024).

5.4.8 Matriz de cumplimiento de objetivos de minería

Tabla 44. *Matriz de cumplimiento de objetivos de minería*

Objetivo de minería	Criterio de éxito	Evidencia	¿Se cumplió?
Desarrollar modelo de segmentación de clientes basado en patrones de gasto.	Identificar al menos 3 segmentos de clientes con un perfil de comportamiento de compra definido y diferenciado.	Sección 5.3.1	Cumplido
Desarrollar un análisis de la rentabilidad y tendencias de consumo de los diferentes servicios ofrecidos.	Identificar los 10 servicios con mayor frecuencia en las órdenes de los clientes.	Sección 5.3.1	Cumplido
Analizar los puntos de alza significativos en los ingresos de habitaciones a lo largo del período.	Detectar al menos 2 periodos de aumento significativo en los ingresos por habitaciones.	Sección 5.3.1	Cumplido

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Objetivo de minería	Criterio de éxito	Evidencia	¿Se cumplió?
Analizar rentabilidad y tendencias de consumo en servicios	Identificar al menos una tendencia mensual y un patrón semanal sobre como varían las reservas de servicios y habitaciones	Sección 5.3.1	Cumplido

Nota. Elaboración propia (2024).

5.4.9 Validación por parte de los gerentes de Procom

La implementación del proceso de análisis de datos y sus resultados fueron validados formalmente por la gerencia de Procom mediante dos documentos oficiales. El **Apéndice AP** y el **Apéndice AQ** contienen las actas de aceptación firmadas por el gerente general y la gerente de operaciones, donde se certifica que los tableros de control desarrollados cumplen con los objetivos establecidos inicialmente y proporcionan valor para la empresa.

5.4.10 Análisis de la viabilidad de la propuesta.

El análisis de viabilidad para la implementación de un proceso de análisis de datos basado en metodología CRISP-DM para mejorar el módulo de Inteligencia de Negocios de Procom considera cuatro dimensiones fundamentales:

5.4.10.1 Viabilidad Técnica

La propuesta se fundamenta en herramientas tecnológicas consolidadas, incluyendo la metodología CRISP-DM, el desarrollo de un almacén de datos departamental y la implementación de procesos ETL estandarizados. Procom cuenta con una infraestructura tecnológica robusta, demostrada en la gestión efectiva de sistemas de información para su amplia cartera de clientes en el sector hotelero.

Sin embargo, se identifican riesgos técnicos relacionados con la complejidad de los datos, dado que estos provienen de múltiples sistemas de información, generando desafíos en los procesos de integración y limpieza. Adicionalmente, la base de datos del cliente piloto abarca menos de un año de información, limitando la capacidad del análisis para generar conclusiones basadas en tendencias a largo plazo.

5.4.10.2 Viabilidad Organizativa

El análisis organizativo revela una estructura definida en Procom, con roles designados para la ejecución del proyecto, incluyendo al gerente general y el

encargado de Inteligencia de Negocios. El compromiso de la alta dirección constituye un factor crítico que sustenta la viabilidad organizativa del proyecto, garantizando la asignación de recursos humanos y tecnológicos necesarios.

No obstante, la empresa enfrenta el desafío de carecer de personal dedicado exclusivamente al análisis de datos. Adicionalmente, aunque actualmente no existe una cultura consolidada de decisiones basadas en datos, la gerencia demuestra determinación para abordar esta limitación mediante un proceso de transformación cultural.

5.4.10.3 Viabilidad Económica

Desde la perspectiva económica, el proyecto presenta un potencial significativo de generación de valor tanto para Procom como para sus clientes. La implementación de un proceso de análisis de datos escalable y la creación de tableros de control interactivos basados en indicadores clave del sector hotelero no solo mejorarán la eficiencia operativa, sino que también fortalecerán la capacidad de toma de decisiones estratégicas fundamentada en datos.

La estrategia de implementación adoptada por Procom minimiza los riesgos económicos al iniciar con un cliente piloto antes de escalar la solución. Este enfoque resulta económicamente viable al permitir la validación del proyecto antes de realizar inversiones significativas en su expansión. Adicionalmente, la empresa podrá explorar nuevos modelos de negocio utilizando los conocimientos generados por los informes analíticos.

5.4.10.4 Viabilidad de Resultados

El éxito del proyecto depende fundamentalmente de su capacidad para generar valor tanto para Procom como para sus clientes hoteleros. Los beneficios directos esperados incluyen la creación de un almacén de datos departamental específico para el sector hotelero, la visualización efectiva de indicadores clave de rendimiento y la implementación de tableros de control interactivos que faciliten la toma de decisiones fundamentada en datos.

La escalabilidad del proyecto constituye un factor crítico de viabilidad, dado que los procesos desarrollados son aplicables a otros clientes hoteleros de Procom, generando un impacto positivo sostenible. Sin embargo, es importante señalar que el alcance actual excluye el desarrollo de modelos predictivos o prescriptivos, limitándose al análisis descriptivo y diagnóstico, lo que circunscribe el alcance de las recomendaciones generadas a partir de los datos disponibles.

6 Conclusiones

Este proyecto se centró en la implementación de un proceso de análisis de datos fundamentado en la metodología *CRISP-DM* para mejorar el módulo de Inteligencia de Negocios de Procom, focalizando inicialmente en un cliente del sector hotelero. El objetivo principal consistió en transformar los datos operativos en conocimientos accionables.

El proyecto abordó cuatro objetivos específicos: el análisis de patrones de reportes y la definición de indicadores clave, la construcción de un almacén de datos departamental con procesos ETL, la elaboración de tableros de control interactivos y la validación del proceso con los gerentes de Procom.

Las conclusiones presentadas a continuación reflejan los resultados obtenidos para cada objetivo, destacando los logros alcanzados, los desafíos enfrentados y las lecciones aprendidas durante el desarrollo e implementación del proyecto. Estas conclusiones derivan del análisis de resultados de la **Sección 4** y la propuesta de solución de la **Sección 5**.

Objetivo General: Implementar un proceso de análisis de datos apoyado en la metodología CRISP-DM, con enfoque inicial en un cliente del sector hotelero para la mejora del módulo de BI existente.

- Se implementó exitosamente un proceso de análisis de datos basado en *CRISP-DM*, mejorando el módulo de BI de Procom para el sector hotelero.
- La aplicación de la metodología *CRISP-DM* permitió abordar de manera estructurada y eficiente los desafíos de análisis de datos del cliente hotelero piloto.
- El proceso implementado demostró ser efectivo para transformar datos brutos en *insights* accionables, facilitando la toma de decisiones informadas en el sector hotelero.
- El enfoque sistemático de *CRISP-DM* permitió contener los ajustes imprevistos derivados de los hallazgos relacionados con la calidad de los datos y las particularidades del negocio.

Objetivo Específico 1: Analizar los patrones de los reportes solicitados por los clientes hoteleros y los sistemas de información que sirven como fuente de datos para la definición de indicadores claves de rendimiento para los hoteles.

- Los datos disponibles se centran en características financieras y de ocupación.
- La exploración de los sistemas de información existentes, como el *PMS*, permitió mapear las fuentes de datos relevantes para los *KPIs* identificados.

Objetivo de negocio asociado: Identificar indicadores clave de rendimiento (KPIs) específicos para la industria hotelera.

- Se identificaron *KPIs* para el sector hotelero, incluyendo *RevPAR*, *ADR*, tasa de ocupación, e ingresos por cliente, proporcionando una base sólida para la medición del rendimiento hotelero.

RF-005: Documentación de modelos y análisis

- Se documentaron los *KPIs* identificados.
- Se crearon diccionarios de datos completos.
- Se establecieron definiciones claras para cada indicador.

Objetivo Específico 2: Construir un *data mart*, junto con los procesos de extracción, transformación y carga de los datos (ETL) generados, para la presentación de datos depurados y la creación de los modelos analíticos para la medición de los indicadores claves de rendimiento de los hoteles.

- Se diseñó e implementó con éxito un *data mart* utilizando una combinación de esquemas estrella y copo de nieve, para el análisis en el sector hotelero.
- Los procesos *ETL* desarrollados lograron integrar eficazmente datos de múltiples fuentes, incluyendo el sistema *PMS* y otras plataformas operativas del hotel.
- La construcción del *data mart* facilitó la creación de modelos analíticos robustos, permitiendo un análisis de los *KPIs* hoteleros identificados.
- El proceso de *ETL* mejoró la calidad de los datos (véase **Sección 5.1.2**), reduciendo inconsistencias y errores presentes en las fuentes de datos originales.
- La estructura del *data mart* se diseñó para acomodar futuras expansiones y la incorporación de nuevas fuentes de datos.

Objetivo de negocio asociado: Identificar patrones y segmentos clave

- Se identificaron tres segmentos claramente diferenciados basados en valor y frecuencia
- Los patrones de comportamiento por segmento son accionables y útiles para

estrategias

RF-004: Generación de recomendaciones

- Se generaron recomendaciones para cada segmento.
- Se desarrollaron propuestas de valor personalizadas para el cliente hotelero.

RF-006: Software de análisis estadístico

- Se realizaron análisis de segmentación exitosos usando R.
- Se implementó la técnica de clustering K-Means para identificar las características de cada segmento generado mediante el análisis de sus centroides.
- Se validaron estadísticamente los segmentos identificados.

RF-008: Cuadros de mando con filtros

- Se implementaron filtros para análisis por segmento.
- Se facilitó la exploración de características por segmento.
- Se permite el análisis comparativo entre segmentos.

Objetivo Específico 3: Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los *KPIs* en el módulo de BI, para la generación de recomendaciones sobre los resultados.

- Se crearon cuatro cuadros de mando interactivos utilizando *DevExpress Report Designer*, que visualizan los *KPIs* clave del sector hotelero.
- Los *dashboards* desarrollados incluyen análisis de segmentación de clientes, tendencias de servicios, rentabilidad de habitaciones y patrones de reserva, proporcionando una visión integral del rendimiento hotelero.
- Las visualizaciones interactivas permiten a los usuarios explorar los datos de manera intuitiva, facilitando la identificación de patrones y la generación de *insights* accionables.
- Los cuadros de mando han mejorado la capacidad de Procom para generar recomendaciones estratégicas basadas en datos para sus clientes hoteleros.

- La implementación de filtros en los *dashboards* ha aumentado la flexibilidad y profundidad del análisis disponible para los usuarios finales.

Objetivo de negocio asociado: Desarrollar recomendaciones basadas en datos mediante cuadros de mando interactivos y visualizaciones.

RF-001: Visualización de datos con Devexpress Report Designer

- Se implementó exitosamente DevExpress Report Designer como herramienta principal de visualización.
- Se crearon cuatro *dashboards* interactivos funcionales.
- Se logró una integración efectiva con la base de datos SQL Server de Procom.

RF-002: Desarrollo de cuadros de mando interactivos que muestren *KPIs* relevantes

- Se desarrollaron cuatro *dashboards* interactivos, superando el criterio mínimo de dos.
- Cada *dashboard* incorporó más de tres *KPIs* críticos.
- Se implementaron visualizaciones efectivas de *KPIs* como *RevPAR*, *ADR* y tasa de ocupación.

RF-006: Software de análisis estadístico compatible con el gestor de base de datos

- Se implementó exitosamente R como herramienta de análisis estadístico.
- Se logró una integración efectiva con SQL Server mediante RODBC.
- Se realizaron análisis estadísticos que respaldan las visualizaciones.

RF-008: Cuadros de mando con filtros

- Se implementaron filtros interactivos que permiten exploración multidimensional.
- Los usuarios tienen la posibilidad de analizar datos por múltiples criterios.
- La funcionalidad de filtrado permitió analizar períodos de tiempo, segmentos de clientes, tipos de habitaciones y servicios específicos según la

necesidad del hotel.

Objetivo de negocio asociado: Determinar tendencias estacionales de habitaciones y servicios.

- Se identificaron picos de demanda en enero y marzo 2024
- Se establecieron patrones claros de ocupación por día de la semana
- Se detectaron variaciones significativas en el uso de servicios por temporada

RF-003: Presentación clara y comprensible

- Se visualizaron efectivamente patrones estacionales.
- Se identificaron tendencias claras en ocupación e ingresos.
- Se presentaron variaciones temporales de forma comprensible.

RF-004: Generación de recomendaciones

- Se generaron recomendaciones basadas en patrones estacionales.
- Se identificaron oportunidades por temporada.
- Se propusieron estrategias específicas por período.

RF-008: Cuadros de mando con filtros

- Se implementaron filtros temporales efectivos.
- Se permite el análisis por diferentes períodos.
- Se facilita la identificación de patrones estacionales.

Objetivo de negocio asociado: Identificar los servicios de mayor valor y las tendencias de consumo:

- Se identificó "TA Adulto Pase de día" como el servicio de mayor rentabilidad
- Se establecieron patrones claros de consumo por temporada y día de la semana
- Se cuantificó el impacto financiero de cada servicio

RF-003: Presentación clara y comprensible

- Se visualizaron efectivamente los servicios de mayor rentabilidad.
- Se presentaron tendencias de consumo de forma clara.
- Se facilitó la identificación de servicios clave.

RF-004: Generación de recomendaciones

- Se identificaron los 10 servicios de mayor rentabilidad
- Se generaron recomendaciones para optimizar servicios.
- Se propusieron estrategias de mejora específicas.

RF-008: Cuadros de mando con filtros

- Se implementaron filtros por tipo de servicio.
- Se permite el análisis detallado de rentabilidad.
- Se facilita la comparación entre servicios.

Objetivo Específico 4: Validar el proceso de análisis de datos implementado con los gerentes de Procom que garanticen la entrega de valor a la empresa.

- El proceso de análisis de datos fue validado exitosamente con los gerentes de Procom, confirmando su valor para la empresa y sus clientes.
- La retroalimentación de los gerentes indicó que el nuevo enfoque proporciona insights relevantes sobre el desempeño del hotel.
- Los gerentes reconocieron el potencial del nuevo proceso para escalar a otros clientes del sector hotelero, indicando un posible aumento en el valor ofrecido por Procom.
- La validación reveló áreas de mejora y oportunidades para futuras iteraciones del proceso de análisis de datos, garantizando su evolución continua.
- El proceso implementado ha sentado las bases para fomentar una cultura de toma de decisiones basada en datos dentro de Procom, aunque se reconoce que este es un proceso en desarrollo que requerirá esfuerzos continuos.

RF-005: Documentación de modelos y análisis

- Se creó documentación completa de todos los modelos.
- Se desarrollaron diccionarios de datos detallados.
- Se documentaron procesos *ETL* y reglas de transformación.

RF-007: Diseño adaptable a otros segmentos

- Se desarrolló un proceso de análisis escalable.
- La estructura del *data mart* permite adaptación.
- Se estableció una base metodológica replicable.

7 Recomendaciones

En este capítulo se presentan recomendaciones basadas en los resultados obtenidos y la experiencia adquirida durante la implementación del proceso de análisis de datos en Procom. Estas sugerencias están diseñadas para maximizar el valor de la solución propuesta, optimizar el proceso de implementación y aprovechar las oportunidades identificadas durante el estudio. Además, se abordan aspectos que, aunque no se exploraron en profundidad en este proyecto, se consideran relevantes para el futuro desarrollo y expansión de las capacidades analíticas de Procom en el sector hotelero.

Objetivo General: Implementar un proceso de análisis de datos apoyado en la metodología CRISP-DM, con enfoque inicial en un cliente del sector hotelero para la mejora del módulo de BI existente.

- Estandarizar el proceso de análisis de datos basado en *CRISP-DM* como metodología central para todos los proyectos de *BI* en Procom.
- Desarrollar un plan de capacitación integral para el personal de Procom en metodología *CRISP-DM* y mejores prácticas de *BI* en el sector hotelero.
- Desarrollar una guía detallada de mejores prácticas para la implementación de *CRISP-DM* en el contexto del sector hotelero.
- Establecer un programa de capacitación continua en *CRISP-DM* y análisis de datos para el personal de Procom.
- Implementar un sistema de seguimiento y evaluación del impacto de la metodología

Objetivo Específico 1: Analizar los patrones de los reportes solicitados por los clientes hoteleros y los sistemas de información que sirven como fuente de datos para la definición de indicadores claves de rendimiento para los hoteles.

- Implementar un sistema de *feedback* automatizado para que los clientes hoteleros sugieran nuevos *KPIs* o modificaciones a los existentes.
- Implementar un sistema de seguimiento y evaluación del impacto de la metodología
- Desarrollar una biblioteca de *KPIs* específicos para diferentes tipos de hoteles (urbanos, resorts, entre otros), permitiendo una personalización efectiva de los análisis.
- Establecer un proceso de revisión trimestral de los *KPIs* con los clientes hoteleros para asegurar su relevancia continua.

Objetivo de negocio asociado: Identificar indicadores clave de rendimiento (KPIs) específicos para la industria hotelera.

- Implementar *benchmarking* automático contra promedios del sector hotelero.
- Crear un sistema de *scorecards* personalizados por nivel jerárquico.
- Implementar análisis de correlación entre *KPIs* para identificar relaciones causales.
- Establecer un programa de validación regular de la calidad de datos que alimentan los *KPIs*.

RF-002: Cuadros de mando interactivos

- Desarrollar un sistema de alertas para *KPIs* que se desvíen de rangos esperados.
- Implementar *benchmarking* automático contra promedios del sector.
- Crear visualizaciones predictivas de *KPIs* basadas en tendencias históricas.

RF-003: Presentación clara y comprensible

- Implementar *tooltips* informativos para cada visualización.
- Crear dashboards adaptables a diferentes niveles de experiencia del usuario.

RF-005: Documentación

- Establecer un proceso de revisión trimestral de *KPIs* relevantes.
- Desarrollar un repositorio centralizado de documentación.
- Implementar un sistema de versionado para la documentación.

Objetivo Específico 2: Construir un *data mart*, junto con los procesos de extracción, transformación y carga de los datos (ETL) generados, para la presentación de datos depurados y la creación de los modelos analíticos para la medición de los indicadores claves de rendimiento de los hoteles.

- Implementar un sistema de gestión de metadatos para mejorar la trazabilidad y el gobierno de los datos en el *data mart*.

- Desarrollar un módulo de calidad de datos automatizado que aplique reglas de negocio y validaciones personalizadas durante el proceso *ETL*.
- Establecer un proceso de revisión y optimización trimestral del rendimiento del *data mart*, incluyendo la reevaluación de índices y particiones.

Objetivo de negocio asociado: Identificar patrones y segmentos clave

- Desarrollar programas de fidelización personalizados para cada segmento identificado.
- Implementar análisis predictivo de comportamiento por segmento.
- Crear estrategias de marketing diferenciadas basadas en las características de cada segmento.
- Establecer un sistema de monitoreo continuo de la evolución de los segmentos.
- Desarrollar modelos de propensión a la compra por segmento.
- Implementar análisis de ciclo de vida del cliente por segmento.

RF-004: Generación de recomendaciones

- Desarrollar un sistema de recomendaciones automatizado basado en segmentos.
- Implementar análisis predictivo de comportamiento por segmento.
- Crear programas de fidelización personalizados por segmento.

RF-006: Software de análisis estadístico

- Implementar modelos avanzados de segmentación (como RFM).
- Desarrollar capacidades de análisis en tiempo real.
- Establecer procesos de validación automática de segmentos.

RF-008: Cuadros de mando con filtros

- Implementar análisis comparativo entre segmentos.
- Crear reportes automáticos por segmento.

Objetivo Específico 3: Elaborar al menos dos cuadros de mando interactivos con visualizaciones de los KPIs en el módulo de BI, para la generación de recomendaciones sobre los resultados.

- Crear una biblioteca de plantillas de *dashboards* personalizables para diferentes roles y niveles jerárquicos en los hoteles.
- Implementar un programa de capacitación para los usuarios finales sobre cómo interpretar y actuar sobre los *insights* proporcionados por los *dashboards*.
- Establecer un proceso de revisión y actualización regular de los *dashboards* basado en el *feedback* de los usuarios y las cambiantes necesidades del negocio.

Objetivo de negocio asociado: Desarrollar recomendaciones basadas en datos mediante cuadros de mando interactivos y visualizaciones.

- Implementar un programa de capacitación estructurado para usuarios de los *dashboards*, incluyendo sesiones básicas y avanzadas.
- Establecer un ciclo de retroalimentación regular con usuarios finales para mejorar continuamente las visualizaciones.
- Implementar un sistema de alertas personalizables basado en umbrales definidos por el usuario.
- Desarrollar capacidades de exportación de datos y reportes en múltiples formatos.
- Crear guías interactivas dentro de los *dashboards* para nuevos usuarios.

RF-001: Visualización de datos con DevExpress Report Designer

- Implementar un programa regular de actualización de DevExpress para aprovechar nuevas funcionalidades.
- Desarrollar una biblioteca de plantillas de visualización personalizadas para diferentes tipos de análisis.
- Establecer estándares de diseño corporativo para mantener consistencia visual.

RF-002: Desarrollo de cuadros de mando interactivos

- Implementar un sistema de *feedback* automatizado para evaluar la utilidad de los *dashboards*.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- Desarrollar versiones móviles de los dashboards para acceso en diferentes dispositivos.
- Crear tutoriales interactivos para nuevos usuarios de los dashboards.

RF-006: Software de análisis estadístico

- Establecer un pipeline automatizado para la actualización de datos en R.
- Desarrollar scripts reutilizables para análisis comunes.
- Implementar control de versiones para los scripts de análisis.

RF-008: Cuadros de mando con filtros

- Desarrollar funcionalidades de guardado de configuraciones de filtros personalizados.
- Añadir capacidades de exportación de datos filtrados.

Objetivo de negocio asociado: Determinar tendencias estacionales de habitaciones y servicios.

- Implementar un sistema de precios dinámicos basado en patrones estacionales.
- Desarrollar estrategias de marketing específicas para temporadas altas y bajas.
- Crear planes de gestión de capacidad basados en predicciones estacionales.
- Establecer políticas de inventario dinámicas según la temporada.
- Implementar promociones anticipadas basadas en tendencias históricas.

RF-003: Presentación clara

- Implementar visualizaciones predictivas de tendencias estacionales.
- Desarrollar alertas automáticas para cambios en patrones estacionales.
- Crear comparativas año contra año automatizadas.

RF-004: Generación de recomendaciones

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- Desarrollar un sistema de precios dinámicos basado en patrones estacionales.
- Implementar recomendaciones automáticas de gestión de inventario.
- Crear estrategias de marketing automatizadas por temporada.

RF-008: Cuadros de mando con filtros

- Implementar análisis predictivo de ocupación por temporada.
- Desarrollar comparativas multianuales de patrones estacionales.
- Crear reportes automáticos de rendimiento por temporada.

Objetivo de negocio asociado: Identificar los servicios de mayor valor y las tendencias de consumo:

- Implementar un sistema de optimización de precios por servicio basado en la demanda.
- Desarrollar estrategias de *cross-selling* y *up-selling* basadas en patrones de consumo.
- Crear paquetes personalizados combinando servicios de alta rentabilidad.
- Establecer un programa de desarrollo de nuevos servicios basado en tendencias identificadas.
- Implementar un sistema de monitoreo de la satisfacción del cliente por servicio.
- Desarrollar estrategias de promoción específicas para servicios con potencial de crecimiento.

RF-003: Presentación clara

- Implementar *scorecards* dinámicos de rendimiento de servicios.
- Desarrollar análisis de rentabilidad por segmento de cliente.
- Crear visualizaciones de análisis de margen por servicio.

RF-004: Generación de recomendaciones

- Desarrollar un sistema de optimización de precios por servicio.

- Implementar recomendaciones de cross-selling basadas en patrones de consumo.
- Crear estrategias de promoción personalizadas por servicio.

RF-008: Cuadros de mando con filtros

- Implementar análisis de rentabilidad por canal de venta.
- Desarrollar seguimiento de tendencias de consumo en tiempo real.
- Crear reportes de impacto de promociones por servicio.

Objetivo Específico 4: Validar el proceso de análisis de datos implementado con los gerentes de Procom que garanticen la entrega de valor a la empresa.

- Desarrollar un marco de evaluación del valor del negocio para cuantificar el impacto del sistema de BI en los resultados de los clientes hoteleros.
- Crear un programa dentro de Procom para promover la adopción y el uso efectivo del sistema.
- Establecer reuniones trimestrales de revisión con los gerentes de Procom para evaluar el impacto del sistema de *BI* y planificar mejoras futuras.
- Implementar un sistema de seguimiento de métricas clave de adopción y satisfacción del usuario para medir el éxito continuo del sistema de *BI*.

8 Referencias

- Bernal, C. A. (2010). Metodología de la investigación: administración, economía, humanidades y ciencias sociales (3ra edición.). Pearson Educación.
- Bilgihan, A., Cobanoglu, C., & Miller, B. L. (2021). Hospitality information technology: Learning how to use it (9th ed.). Kendall Hunt Publishing.
- Brynjolfsson, E., & McElheran, K. (2019). Data in action: Data-driven decision making in US manufacturing (Working Paper No. 25547). National Bureau of Economic Research. <https://www.nber.org/papers/w25547>
- Cámara Costarricense de Hoteles. (2021). Las 6 mejores prácticas para elegir KPIs para tu negocio. Recuperado de https://www.ecobusiness.fund/fileadmin/user_upload/Sustainability_Academy/Recursos/herramienta_turismo_5.1.pdf.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.
- Data Governance Institute. (2020). The data governance framework. <https://www.datagovernance.com/the-dgi-data-governance-framework/>
- Data Management Association International. (2017). The DAMA guide to the data management body of knowledge (DAMA-DMBOK) (2nd ed.). Technics Publications.
- DataAnalytics. (2023). Informe sobre el aprovechamiento de datos en empresas del sector de soluciones empresariales. DataAnalytics.
- Davenport, T. H., & Harris, J. G. (2017). Competing on analytics: The new science of winning (Updated, with a new introduction). Harvard Business Review Press.
- Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of business process management (2nd ed.). Springer.
- Failte Ireland. (2023). Hotel Key Performance Indicators. Recuperado de <https://supports.failteireland.ie/wp-content/uploads/2021/07/Hotel-Key-Performance-Indicators.pdf>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications Ltd.
- García, J., & López, M. (2024). Optimización del rendimiento de aplicaciones en la nube mediante análisis de datos: Un caso de estudio en una empresa de software como servicio

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- (SaaS). Conferencia Internacional de Computación en la Nube y Servicios.
- Gartner. (2021). Data and analytics leaders must embrace a data-driven culture to drive value. <https://www.gartner.com/en/documents/4004063>
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486-489. <https://doi.org/10.5812/ijem.3505>
- Hernández, R., Baptista, P. y Fernández, C. (2014). Metodología de la investigación. McGraw-Hill. <http://www.ebooks7-24.com.ezproxy.itcr.ac.cr/?il=285>
- Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, 2, 652-687. <https://doi.org/10.1109/ACCESS.2014.2332453>
- IBM Cognos Analytics 11.1.x. (s. f.). <https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=types-line-charts>
- IBM. (2021). *Schemas de estrella*. IBM Knowledge Center. Recuperado de <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-star>
- IBM. (2021). *Schemas de copo de nieve*. IBM Knowledge Center. Recuperado de <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-snowflake>
- IBM. (2020). The data-driven organization: Unlocking greater value from data. IBM Corporation. <https://www.ibm.com/downloads/cas/XKYL87ZJ>
- IBM. (s.f.). IBM SPSS Modeler CRISP-DM Guide. International Business Machines Corporation.
- Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). Wiley.
- Kalakota, R., & Robinson, M. (2001). *e-Business 2.0: Roadmap to success*. Addison Wesley.
- Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Wiley.
- Kotu, V., & Deshpande, B. (2015). *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Elsevier.
- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700-710.
- Lucid Software Inc. (2023). *Qué es un diagrama entidad-relación*. Lucidchart.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

<https://www.lucidchart.com/pages/es/que-es-un-diagrama-entidad-relacion>

Jones, A., & Smith, B. (2023). *Data Mining Project Management: Risks and Strategies*. Publisher Name.

Malik, S. (2005). *Enterprise dashboards: Design and best practices for IT*. Wiley.

Millán-García, C. H., & Gómez-Díaz, M. R. (2018). Factores e indicadores de competitividad hotelera. *Compendium*, 21(40), 1-18. Recuperado de <https://www.redalyc.org/articulo.oa?id=88055200021>

Munappy, A. R., Mattos, D. M. F., Boer, J. M. A., Moreira, V. P., Fonseca, M. J., & Ferreira, P. R. (2020). An overview of data extraction transformation and loading. *IEEE Access*, 8, 192095-192112. <https://doi.org/10.1109/ACCESS.2020.3031971>

Nguyen, T., Malecki, R., & Kanat, I. E. (2020). Building a data-driven culture: Creating change through analytic capability. *The Journal of Product Innovation Management*, 37(2), 142-156. <https://doi.org/10.1111/jpim.12527>

NIST/SEMATECH. (2012). *Handbook of statistical methods* (NIST handbook). National Institute of Standards and Technology. <https://doi.org/10.18434/M32189>

Ñaupas, H., Mejía, E., Novoa, E., & Villagómez, A. (2013). *Metodología de la investigación Cuantitativa - Cualitativa y Redacción de la Tesis*. Bogotá Colombia: Ediciones de la U.

Olszak, C. M. (2016). Toward better understanding and use of business intelligence in organizations. *Information Systems Management*, 33(2), 105-123. <https://doi.org/10.1080/10580530.2016.1155946>

Parmenter, D. (2015). *Key performance indicators: Developing, implementing, and using winning KPIs* (3rd ed.). John Wiley & Sons.

Pimienta, J. y Orden, A. (2012). *Metodología de la investigación*. Pearson.

Ponniah, P. (2000). *Data Warehousing fundamentals*. New York: John Wiley & Sons, Inc.

Procom. (2024, 27 de mayo). Algunos Casos de Estudio. Recuperado de <https://www.procom.cr/casos-de-exito/>

Procom. (2024, 27 de mayo). Sobre nosotros. <https://www.procom.cr/sobreprocom/>

Project Management Institute. (2021). *A guide to the project management body of knowledge (PMBOK guide)* (7th ed.). Project Management Institute.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM
para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- Rodríguez, E., & Fernández, D. (2022). La implementación de un módulo de BI para el análisis de datos de clientes en una tienda online de ropa. *Revista Internacional de Comercio Electrónico y Marketing*.
- Rondi, G. R., Casal, M. del C., & Giai Levra, V. (2019). Análisis de los principales indicadores de desempeño utilizados en la actividad hotelera. *Contabilidad y Auditoría*, 50, 142-181. <https://nulan.mdp.edu.ar/id/eprint/3285/1/rondi-et-al-2019.pdf>
- Silberschatz, A., Korth, H. F., & Sudarshan, S. (2020). *Database system concepts* (7th ed.). McGraw-Hill Education.
- Srivastava, N., & Maitra, R. (2016). Key Performance Indicators (KPI) in Hospitality Industry: An Emphasis on Accommodation Business of 5 Star Hotels of National Capital Region. *International Journal of Research in Tourism and Hospitality*, (1), 34-40. <http://dx.doi.org/10.20431/2455-0043.0201004>
- Tableau. (2014). *Tableau visual guidebook*. Tableau. <https://www.tableau.com/es-es/learn/whitepapers/tableau-visual-guidebook>
- The Data Warehousing Institute. (2019). *TDWI best practices reports* (4th ed.). TDWI.
- The Open Group. (2018). *TOGAF standard, version 9.2*. Van Haren Publishing.
- Universidad de Tecnología. (2022). *Estudio sobre la percepción de las empresas acerca de ser organizaciones data-driven*. Universidad de Tecnología.
- Vargas Morúa, E. y Ulate Soto, I. (2016). *Metodología para elaborar una tesis*. MONTES DE OCA, Costa Rica: Editorial Universidad Estatal a Distancia
- Vaske, J. J., Beaman, J., & Ring, A. (2017). *Cramer's V: A measure of association for nominal variables*. *Journal of Statistical Education*, 25(2), 86-95. <https://doi.org/10.1080/10691898.2017.1362402>
- Vásquez Núñez, J. D. (2012). *DICCIONARIO DE DATOS Sistema de Control de Producción SICPROD V1.0*. Universidad Técnica del Norte. <https://repositorio.utn.edu.ec/bitstream/123456789/1019/7/04%20ISC%20206%20DICCIONARIO%20DE%20DATOS.pdf>

9 Apéndices

Apéndice A. Plantilla de minutas para reunión

Minuta de Reunión			
Reunión No.		Fecha: dd/mm/aaaa	
Lugar:		Hora: Inicio/Finalización	
Objetivo de la reunión:			
Participantes:	Presentes:		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
Próxima reunión			
Temas para tratar	Fecha	Convocados	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice B. Plantilla de gestión de cambios

Solicitud de Cambio		
Número de solicitud		
Fecha	dd/mm/aaaa	
Solicitante		
Descripción del cambio		
Justificación del cambio		
Involucrados		
Prioridad	Bajo/Medio/Alto	
Impacto	Bajo/Medio/Alto	
Estado	Aceptado/Rechazado	Motivo

Apéndice C. Plantilla de entrevista

Guía de entrevista	
Fecha	
Objetivo	
Entrevistador	
Entrevistado(s)	
Nombre	Rol
Preguntas/Respuestas	
Pregunta	Respuesta
Observaciones	

Apéndice D. Plantilla Observaciones

Observación
Participantes
Rol
Temas por observar
Observaciones

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice F. Plantilla Revisión Documental

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
001					
00X					

Apéndice G. Entrevista con el gerente general 1

Minuta de Reunión			
Código de minuta	M-01	Fecha:	05/05/2024
Lugar:	Llamada telefónica	Hora: Inicio/Finalización	2:15pm / 3:00pm
Objetivo de la reunión:	Introducción de la empresa		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> Christian Vargas José Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Descripción de la organización	Se da una descripción de las funciones de la organización. Se expresa el interés de adoptar una cultura data-driven y de mejorar la funcionalidad de su módulo de BI	Proporcionar documentación Proporcionar los proyectos internos realizados
2	Descripción de misión, visión, valores de la empresa y proyectos similares	Se expresa la descripción de misión, visión, valores de la empresa y proyectos similares de	Proporcionar documentación

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

		forma clara	
3	Roles de la empresa	Se habla brevemente sobre la estructura de la empresa y los departamentos que esta tiene.	Proporcionar documentación
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Entender la problemática	20/05/2024	José Andrés Brenes Granados Christian Vargas	

Apéndice H. Entrevista con el gerente general 2

Minuta de Reunión			
Código de minuta	M-02	Fecha:	20/05/2024
Lugar:	Llamada Telefónica	Hora: Inicio/Finalización	3:00pm / 4:00pm
Objetivo de la reunión:	Entender la problemática de la empresa		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> Christian Vargas José Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Entendimiento a alto nivel de los procesos de Procom	Se identificó la carencia de procesos y tecnologías relacionadas a esto.	Proporcionar la documentación de los procesos
2	Problemática	Se identificaron	Crear una

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

	enfrentada por la empresa	problemáticas entorno al conocimiento general de los empleados sobre el análisis de datos. Se identificó la falta de decisiones estratégicas fundamentadas en datos, que les está generando deficiencias a la hora de aprovechar oportunidades en el mercado.	propuesta para enfrentar la problemática
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Presentación con el encargado de BI	25/06/2024	<ul style="list-style-type: none"> • Paola Ovares • Daniel Quesada 	

Apéndice I. Entrevista con Gerente de Operaciones y Encargado de BI

Minuta de Reunión			
Código de minuta	M-03	Fecha:	25/06/2024
Lugar:	Teams	Hora: Inicio/Finalización	11:10am / 11:40am
Objetivo de la reunión:	Presentación con el encargado de BI e introducción a los sistemas de la empresa		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Paola Ovares • Daniel Quesada • José Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No.	Asunto	Comentarios	Acuerdos
1	Explicación de la propuesta de trabajo	Se explicó al encargado de BI la propuesta de trabajo para abordar la problemática de Procom.	Se acordó trabajar en conjunto con el encargado de BI para realizar reportes que generen valor a los clientes. Los reportes aportarían valor para que las empresas estén preparadas para afrontar las demandas y gestionen mejor sus recursos.
	Introducción a los sistemas de Procom	El encargado de BI introdujo los diferentes sistemas que utiliza Procom para satisfacer las necesidades de los clientes.	Coordinación para realizar la instalación de los sistemas necesarios para comenzar a trabajar
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Definir el tipo de clientes para realizar el trabajo.	23/07/2024	<ul style="list-style-type: none"> • Christian Vargas • Paola Ovares 	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice J. Entrevista con Gerente General y Gerente de Operaciones

Minuta de Reunión			
Código de minuta	M-04	Fecha:	23/07/2024
Lugar:	Teams	Hora: Inicio/Finalización	10:00am / 10:30am
Objetivo de la reunión:	Definir el tipo de clientes para realizar el trabajo.		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Christian Vargas • Paola Ovares • José Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Tipos de clientes	Se brinda información sobre los diferentes clientes a los cuales la compañía vende sus soluciones de ERP y CRM	
2	Selección de clientes	La empresa seleccionó el sector de hotelería para la elaboración de los reportes	Se acordó pasar una copia de seguridad de la base de datos de los hoteles. Se empezaría con un cliente específico

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

			de inicio.
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Instalación de los sistemas de PROCOM y la transferencia de la copia de seguridad de la base de datos de los hoteles.	08/08/2024	<ul style="list-style-type: none"> • Esteban Córdoba • Jose Andrés Brenes Granados 	

Apéndice K. Entrevista con encargado de soporte de TI

Minuta de Reunión			
Código de minuta	M-05	Fecha:	08/08/2024
Lugar:	Teams	Hora: Inicio/Finalización	10:00am / 11:00am
Objetivo de la reunión:	Instalación de los sistemas de PROCOM y la transferencia de la copia de seguridad de la base de datos de los hoteles.		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Esteban Córdoba • José Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Instalación y configuración remota de los sistemas	Se realizó la instalación del Solaria PMS	
2	Accesos a los sistemas y copia de base de datos	Se transfirieron las credenciales de acceso al Solaria PMS y a la base de datos de Procom.	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Próxima reunión		
Temas para tratar	Fecha	Convocados
Reunión con el encargado de BI	15/08/2024	<ul style="list-style-type: none"> Daniel Quesada José Andrés Brenes Granados

Apéndice L. Reunión con profesora tutora 1

Minuta de Reunión			
Código de minuta	M-06	Fecha:	01/08/2024
Lugar:	Teams	Hora:	8:00pm / 9pm
		Inicio/Finalización	
Objetivo de la reunión:	Reunión de introducción con la profesora tutora		
Participantes:	Presentes: <ul style="list-style-type: none"> Jacqueline Solís Céspedes Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Coordinación para el desarrollo del TFG	Se estableció el jueves de cada semana como el día de reuniones semanales. Se revisó la idea de anteproyecto y se plantearon ideas para iniciar con el trabajo.	Obtener los objetivos de negocio enfocados en minería, ver años de información, ver los reportes y dashboards que la empresa genera para los clientes.
Próxima reunión			
Temas para tratar	Fecha	Convocados	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Avance de tercera	07/08/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados
-------------------	------------	--

Apéndice M. Reunión con profesora tutora 2

Minuta de Reunión			
Código de minuta	M-07	Fecha:	07/08/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:10pm / 8:00pm
Objetivo de la reunión:	Avance de tercera semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Avance de segunda semana	Se expusieron a la profesora tutora los hallazgos de los temas planteados en la semana. Se indagaron formas de abordar la problemática, pero aún quedan temas pendientes de investigación.	Script de R para la exploración de los datos. Exploración de los datos de las tablas de la base de datos del cliente que la empresa asignó
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 4 y 5	22/08/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice N. Reunión con profesora tutora 3

Minuta de Reunión			
Código de minuta	M-08	Fecha:	22/08/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:00pm / 8:00pm
Objetivo de la reunión:	Avance de cuarta y quinta semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Avance de tercera y cuarta semana	<p>Se expusieron a la profesora tutora los hallazgos de los temas planteados en las dos semanas principalmente sobre la gran cantidad de tablas que tenía la base de datos con cientos de columnas.</p> <p>Se realizó la entrega del primer avance la tercera semana y se dio la retroalimentación respectiva.</p> <p>Se definió Google</p>	La reducción dimensional de las tablas y columnas utilizando la desviación estándar e identificación de relaciones entre tablas

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

		Drive como el canal de transmisión de archivos. La profesora adjuntó materiales de referencia.	
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 6	29/08/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 	

Apéndice O. Reunión con profesora tutora 4

Minuta de Reunión			
Código de minuta	M-09	Fecha:	29/08/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:00pm / 8:00pm
Objetivo de la reunión:	Avance de la quinta semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Avance de la sexta semana	Se expusieron a la profesora tutora los hallazgos sobre la reducción de tablas y columnas. No se encontraron relaciones definidas en la base de datos de	Exploración de los datos de las tablas de la base de datos del cliente que la empresa asignó

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

		<p>SQL Server.</p> <p>Se compartieron archivos acerca de la construcción de los <i>data mart</i>.</p> <p>Se hace la entrega del segundo avance del proyecto.</p> <p>Se exploraron algunos gráficos relevantes para la exploración de datos</p>	
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 7	5/09/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 	

Apéndice P. Reunión con profesora tutora 5

Minuta de Reunión			
Código de minuta	M-010	Fecha:	05/09/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:00pm / 8:00pm
Objetivo de la reunión:	Avance de sexta semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

1	Avance de séptima semana	<p>Se expusieron a la profesora tutora los hallazgos sobre la exploración de datos realizada donde se encontró información “sucia” para el análisis.</p> <p>Se exploraron técnicas de limpieza entre ellas técnicas de imputación.</p> <p>La forma en que se abordaron los gráficos no fue la idónea y se debe replantear.</p>	<p>Mejorar la presentación de la exploración de los datos</p> <p>Explorar con los gerentes de cual forma les gustaría imputar los datos.</p>
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 8	12/09/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 	

Apéndice Q. Reunión con profesora tutora 6

Minuta de Reunión			
Código de minuta	M-011	Fecha:	12/09/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:10pm / 8:00pm
Objetivo de la reunión:	Avance de séptima semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No.	Asunto	Comentarios	Acuerdos
1	Avance de octava semana	<p>Se exponen los cambios realizados a la parte de exploración de datos.</p> <p>Se comparte las ideas de imputación dadas por los gerentes de Procom.</p> <p>Se habla acerca de análisis de correlación, series de tiempo y clustering.</p> <p>Se revisa el segundo avance del proyecto</p>	Realizar análisis de correlación, explorar series de tiempo y clustering.
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 9 y 10	15/08/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 	

Apéndice R. Reunión con profesora tutora 7

Minuta de Reunión			
Código de minuta	M-012	Fecha:	25/09/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:10pm / 8:00pm
Objetivo de la reunión:	Avance de octava y novena		
Participantes:	Presentes: <ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 		

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Avance de novena y décima semana	<p>Se expusieron a la profesora tutora los hallazgos sobre el análisis de correlación, explorar series de tiempo y clustering.</p> <p>Se realiza la entrega del tercer avance.</p> <p>Se da retroalimentación acerca de la reunión con la contraparte.</p>	<p>Realizar el diseño del data mart con la información preparada.</p> <p>Realizar cambios al avance 3.</p>
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 11	03/10/2024	<ul style="list-style-type: none"> Jacqueline Solís Céspedes Jose Andrés Brenes Granados 	

Apéndice S. Reunión con profesora tutora 8

Minuta de Reunión			
Código de minuta	M-013	Fecha:	03/10/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:10pm / 8:00pm
Objetivo de la reunión:	Avance de décima semana		
Participantes:	Presentes: <ul style="list-style-type: none"> Jacqueline Solís Céspedes Jose Andrés Brenes Granados 		

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

		Ausentes:	
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Avance de undécima semana	Se revisó de nuevo el avance. Se propusieron mejoras al avance 3 y 4.	Data mart implementado Mejoras al avance 3 Avance 4
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 11	10/10/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 	

Apéndice T. Reunión con profesora tutora 9

Minuta de Reunión			
Código de minuta	M-014	Fecha:	10/10/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:10pm / 8:00pm
Objetivo de la reunión:	Avance de undécima semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Avance de undécima semana	Se revisó de nuevo el avance.	Data mart implementado

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

		Se propusieron mejoras al avance 3 y 4.	Mejoras al avance 3 Avance 4
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 12	17/10/2024	<ul style="list-style-type: none"> Jacqueline Solís Céspedes Jose Andrés Brenes Granados 	

Apéndice U. Reunión con profesora tutora 10

Minuta de Reunión			
Código de minuta	M-015	Fecha:	17/10/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:10pm / 8:00pm
Objetivo de la reunión:	Avance de duodécima semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> Jacqueline Solís Céspedes Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Avance de duodécima semana	Se revisó de nuevo el avance. Se propusieron mejoras al avance 5	Data mart implementado Mejoras al avance 5 Avance final
Próxima reunión			
Temas para tratar	Fecha	Convocados	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Avance semana 3	24/10/2024	<ul style="list-style-type: none"> Jacqueline Solís Céspedes Jose Andrés Brenes Granados
-----------------	------------	--

Apéndice V. Reunión con el encargado de BI

Minuta de Reunión			
Código de minuta	M-016	Fecha:	27/08/2024
Lugar:	Teams	Hora: Inicio/Finalización	11:00am / 12:00pm
Objetivo de la reunión:	Entender el módulo de BI y cómo se conecta con la base de datos		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> Carlos Daniel Quesada Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Entender el módulo de BI y cómo se conecta con la base de datos	Se enseñaron todas las funcionalidades del módulo de BI y cómo estas se conectan con la base de datos.	Realizar todo el proceso y una vez lista la información ver los reportes
Próxima reunión			
Temas para tratar	Fecha	Convocados	
Avance semana 3	Sin fecha	<ul style="list-style-type: none"> Carlos Daniel Quesada Jose Andrés Brenes Granados 	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice W Reunión profesora tutora contraparte 1

Minuta de Reunión			
Código de minuta	M-017	Fecha:	16/08/2024
Lugar:	Teams	Hora: Inicio/Finalización	2:00pm / 2:30pm
Objetivo de la reunión:	Reunión de explicación sobre el proceso del TFG a la contraparte		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Christian Vargas • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Explicación del proceso de TFG de la profesora tutora a la contraparte	Se expusieron los puntos sobre cómo se desarrolla el TFG	Evaluación 1 por parte de la contraparte Segunda reunión en semana 8
Próxima reunión			
Temas para tratar	Fecha	Convocados	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Avance del proyecto	25/09/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Christian Vargas • Jose Andrés Brenes Granados
---------------------	------------	--

Apéndice Y Reunión profesora tutora contraparte 2

Minuta de Reunión			
Código de minuta	M-018	Fecha:	25/09/2024
Lugar:	Teams	Hora: Inicio/Finalización	11:00am / 12pm
Objetivo de la reunión:	Reunión para presentarle el avance del proyecto a la contraparte		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Christian Vargas • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Explicación del avance realizado a la fecha del TFG	Se expusieron los puntos sobre los desafíos presentados, así como el trabajo por realizar.	Evaluación 2 por parte de la contraparte Reunión en semana 14 para la entrega del trabajo final.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Próxima reunión		
Temas para tratar	Fecha	Convocados
Entrega final del proyecto	28/10/2024	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Christian Vargas • Paola Ovares • José Andrés Brenes Granados

Apéndice Z Reunión profesora tutora contraparte 3

Minuta de Reunión			
Código de minuta	M-019	Fecha:	28/10/2024
Lugar:	Teams	Hora: Inicio/Finalización	7:10pm / 8:00pm
Objetivo de la reunión:	Avance de décima semana		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • Jacqueline Solís Céspedes • Christian Vargas • Paola Ovares • Jose Andrés Brenes Granados 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Entrega final	Se revisaron los dashboards creados junto al proceso de análisis de datos realizado	Evaluar y firmar los resultados

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

--	--	--	--

Apéndice AA Primera entrevista con el gerente general

Guía de entrevista	
Código de entrevista	E-01
Fecha	05/05/2024
Objetivo	Introducción a la empresa PROCOM
Entrevistador	Jose Brenes
Entrevistado(s)	
Nombre	Rol
Christian Vargas	Gerente General
Preguntas/Respuestas	
Pregunta	Respuesta
¿Describa las principales funciones y servicios que ofrece PROCOM en el ámbito de dato, BI, comercio y marketing?	<ul style="list-style-type: none"> • Soluciones de Inteligencia Empresarial (BI) y análisis de datos. • Soluciones completas de facturación y punto de ventas (POS) para gestión de inventarios, pedidos, clientes y pagos. • Soluciones de pagos y cobros automatizados a través de una alianza estratégica con Ridivi. • Alianza con Softland para proporcionar plataformas integradas especialmente diseñadas para sectores como alimentos y bebidas, hotelería, turismo y servicio automotriz. • Un módulo específico de BI que, aunque actualmente presenta limitaciones en funcionalidad y visualizaciones, está en proceso de mejora.
¿Cuál es la misión, visión y valores de PROCOM?	<p>Misión: Transformar empresas por medio de la digitalización, por medio del uso de herramientas tecnológicas innovadoras para optimizar procesos, mejorar la eficiencia operativa, impulsar la productividad y aumentar la competitividad en entornos empresariales.</p> <p>Visión: Ser la empresa de puntos de venta número uno en Costa Rica y Región Latinoamericana, se busca ofrecer soluciones tecnológicas innovadoras que transformen la forma en la que las empresas gestionen sus operaciones comerciales y mejorar su competitividad, mediante decisiones en base a datos.</p> <p>Valores: Esfuerzo, respeto, puntualidad, trabajo en equipo, comunicación.</p>
¿Qué proyectos similares ha realizado PROCOM y cuáles han sido sus resultados?	<ul style="list-style-type: none"> ○ PROCOM ha implementado soluciones de BI para diversas industrias, incluyendo cobro a Cooperativa Retail. ○ Implementación de PMS y POS de alimentos y bebidas en Hotel Boutique, se logró una base sólida de datos e información para alimentar el módulo de BI permitiendo analizar los datos. Se logró proporcionar buenas prácticas en la implementación de soluciones tecnológicas en el sector de

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

	<p>hoteles.</p> <ul style="list-style-type: none"> ○ Digitalización de procesos de restaurante con varias sucursales: mejoró la eficiencia y toma de decisiones en tiempo real.
¿Cuál es la estructura organizacional de PROCOM y sus principales departamentos?	<p>Está organizada en puestos como: gerente general, gerente de operaciones, encargado de BI, encargado de soporte de sistemas, encargado de desarrollo tecnológico.</p> <p>Departamentos: consultoría de BI, desarrollo de software, integración de datos, soporte técnico y administración.</p>
¿Cuál es el enfoque de PROCOM para fomentar una cultura data-driven en las empresas cliente?	<ul style="list-style-type: none"> • La empresa reconoce la necesidad de adoptar una cultura basada en datos. • Actualmente carece de procesos formales establecidos para el análisis de datos. • No existe una metodología estructurada o procesos estandarizados. • La empresa está en proceso de reconocer la importancia de los datos para la toma de decisiones. • El módulo de BI existente muestra limitaciones importantes en su funcionalidad.
Observaciones	

Apéndice AB. Segunda entrevista con el gerente general

Guía de entrevista	
Código de entrevista	E-02
Fecha	20/05/2024
Objetivo	Entender la problemática de la empresa
Entrevistador	Jose Brenes
Entrevistado(s)	
Nombre	Rol
Christian Vargas	Gerente General
Preguntas/Respuestas	
Pregunta	Respuesta
¿Cuáles son las principales carencias en los procesos y tecnologías relacionadas con BI que ha identificado en PROCOM?	<ul style="list-style-type: none"> • Falta de procesos formales establecidos para el análisis de datos. • Ausencia de una metodología estructurada para el análisis. • El módulo de BI actual presenta limitaciones en visualizaciones y procesamiento de datos. • Los reportes no tienen información previamente procesada. • Existe una brecha considerable entre las capacidades actuales del módulo y las necesidades reales de los clientes.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

¿Qué nivel de conocimiento tienen los empleados sobre análisis de datos avanzados y cómo afecta esto a la empresa?	<ul style="list-style-type: none"> • Los empleados no están familiarizados con herramientas avanzadas de análisis de datos. • La carencia de habilidades analíticas se extiende a través de diferentes niveles de la organización. • Existe una falta de personal especializado en análisis y visualización de datos, especialmente para el sector hotelero. • Esta brecha de habilidades limita la capacidad de Procom para crear informes útiles.
¿Cómo la falta de decisiones estratégicas fundamentadas en datos está afectando el crecimiento de PROCOM?	<p>Se están viendo afectado en los siguientes ámbitos:</p> <ul style="list-style-type: none"> • Pérdida de oportunidades en el mercado. • Incapacidad para anticipar tendencias de la industria. • Limitación en la capacidad para proporcionar insights valiosos a los clientes. • Dificultad para crear reportes que generen valor real. • Restricción en la capacidad de ofrecer recomendaciones estratégicas. • Subutilización de los datos generados por los sistemas.
¿Qué expectativas tiene la empresa sobre la implementación de nuevas soluciones para abordar estas problemáticas?	<p>La empresa espera que la implementación de nuevas soluciones permita estandarizar nuestros procesos, mejorar la calidad de los análisis de BI y demás, y, aumentar la competitividad en el mercado. Por otra parte:</p> <ul style="list-style-type: none"> • Desarrollo de capacidades analíticas en el personal. • Mejora en la toma de decisiones estratégicas. • Implementación de un proceso de análisis de datos escalable. • Creación de cuadros de mando interactivos con visualizaciones de KPIs.
Observaciones	

Apéndice AC. Entrevista con la gerente de operaciones y el encargado de BI

Guía de entrevista	
Código de entrevista	E-03
Fecha	25/06/ 2024
Objetivo	Presentación con el encargado de BI e introducción a los sistemas de la empresa.
Entrevistador	Jose Andrés Brenes Granados
Entrevistado(s)	
Nombre	Rol
Paola Ovares	Gerente de Operaciones
Daniel Quesada	Encargado de BI
Preguntas/Respuestas	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Pregunta	Respuesta
¿Puede explicar de una forma clara la propuesta de trabajo para abordar la problemática de PROCOM en cuanto a BI?	Se busca una propuesta se centra en implementar una plataforma de BI integrada que automatice la recolección y procesamiento de datos de diversas fuentes. Se pretende que este incluya la creación de dashboards personalizados y la implementación de herramientas de análisis predictivo más avanzadas.
¿Qué tipos específicos de reportes se planean generar para aportar valor a los clientes de PROCOM?	Se desean los siguientes: <ul style="list-style-type: none"> ○ Segmentación de clientes: Distribución de clientes por segmento, RPC (Revenue Per Client) medio por segmento. ○ Análisis de servicios: tendencias mensuales por servicio, Patrones estacionales en la demanda de servicios. ○ Rentabilidad de habitaciones: Ingresos por tipo de habitación, Tendencias temporales de ingresos ○ Patrones de reservas: Análisis comparativo de reservas de habitaciones y servicios y distribución semanal de reservas.
¿Puede describir los diferentes sistemas que utiliza PROCOM actualmente para satisfacer las necesidades de BI de los clientes?	<p>PROCOM utiliza varios sistemas integrados:</p> <ul style="list-style-type: none"> ● DevExpress Report Designer: Principal herramienta para elaborar reportes en el módulo de BI, conectada con Microsoft SQL Server ● Microsoft SQL Server: Sistema gestor de base de datos para almacenar y gestionar la información ● Solaria Hospitality (PMS): Sistema integral de gestión hotelera ● Módulo de BI existente: Aunque con limitaciones actuales en visualizaciones y procesamiento de datos <p>Sistemas complementarios:</p> <ul style="list-style-type: none"> ● Solaria X (ERP) ● Solaria Food ● Solaria FE ● SAO CRM
¿Cuál es el plan detallado para la instalación e integración de los nuevos sistemas de BI necesarios?	<p>Preparación:</p> <ul style="list-style-type: none"> ● Construcción de un data mart específico para el sector hotelero. ● Diseño e implementación de procesos ETL. ● Desarrollo de modelos analíticos. <p>Implementación:</p> <ul style="list-style-type: none"> ● Proceso de actualización diario programado en horarios de baja actividad. ● Sistema ETL automatizado para cada tabla. ● Desarrollo de dashboards en DevExpress Report Designer.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

	<ul style="list-style-type: none"> Integración con las fuentes de datos existentes. <p>Validación y Despliegue:</p> <ul style="list-style-type: none"> Pruebas de los modelos analíticos. Validación con gerentes de Procom. Implementación en entorno de desarrollo.
Observaciones	

Apéndice AD. Entrevista con el gerente general y la gerente de operaciones

Guía de entrevista	
Código de entrevista	E-04
Fecha	23/07/2024
Objetivo	Definir el tipo de clientes para realizar el trabajo
Entrevistador	José Andrés Brenes Granados
Entrevistado(s)	
Nombre	Rol
Christian Vargas	Gerente General
Paola Ovaes	Gerente de Operaciones
Preguntas/Respuestas	
Pregunta	Respuesta
¿Tienen información sobre los diferentes tipos de clientes a los que PROCOM vende sus soluciones de ERP y CRM?	<p>PROCOM atiende principalmente a medianas y grandes empresas en Costa Rica y algunos países de Centroamérica. Sus clientes abarcan sectores como:</p> <ul style="list-style-type: none"> Sector hotelero Restaurantes Servicios de alimentación <p>Cada sector tiene necesidades específicas en cuanto a ERP y CRM, que adaptamos según sus requerimientos.</p>
¿Por qué se ha seleccionado específicamente el sector de hotelería para	<p>Se seleccionó el sector hotelero porque:</p> <ul style="list-style-type: none"> Se ha identificado un crecimiento significativo en el sector turístico. Existe una gran necesidad de soluciones de BI para optimizar

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

la elaboración de los nuevos reportes de BI?	<p>operaciones.</p> <ul style="list-style-type: none"> • Representa una oportunidad para mejorar la gestión de reservas y experiencia del cliente. • Será utilizado como caso piloto para escalar la solución a otros sectores.
¿Qué características específicas de los hoteles hacen que sean un buen candidato para este proyecto de BI?	<ul style="list-style-type: none"> • Los hoteles manejan grandes volúmenes de datos en tiempo real, desde reservas hasta preferencias de clientes y gestión de inventario. • Necesitan análisis predictivos para optimizar ocupación y precios. • La industria es altamente competitiva. • Requieren toma de decisiones basada en datos. • Generan datos de múltiples fuentes (reservas, servicios, clientes).
Observaciones	

Apéndice AE. Entrevista con encargado de Soporte de TI

Guía de entrevista	
Código de entrevista	E-05
Fecha	08/08/2024
Objetivo	Instalación de los sistemas de PROCOM y la transferencia de la copia de seguridad de la base de datos de los hoteles
Entrevistador	José Andrés Brenes Granados
Entrevistado(s)	
Nombre	Rol
Esteban Córdoba	Encargado de Soporte de TI
Preguntas/Respuestas	
Pregunta	Respuesta
¿Describa el proceso de instalación remota del sistema Solaria PMS y cómo se integra con las soluciones de BI de PROCOM?	La instalación remota de Solaria PMS se realiza utilizando Anydesk. Primero, se hace configuración de la base de datos en el servidor local, luego se instala el software Solaria PMS y finalmente se conecta con el sistema de BI.
Observaciones	

Apéndice AF Entrevista con encargado de desarrollo

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Guía de entrevista	
Código de entrevista	E-06
Fecha	07/10/2024
Objetivo	Aspectos técnicos del manejo de los datos
Entrevistador	José Andrés Brenes
Entrevistado(s)	
Nombre	Rol
Dennis Corrales	Encargado de desarrollo
Preguntas/Respuestas	
Pregunta	Respuesta
¿Cuáles son las principales tablas y relaciones en la base de datos que son cruciales para el análisis de datos del sector hotelero?	Las tablas principales para el análisis del sector hotelero incluyen Bill, ReservationBook, Room, Service, Order, OrderCPL, ReservationService y Log_Stays. Estas tablas están interrelacionadas y contienen información crucial sobre facturación, reservas, servicios, y estancias de los clientes. Las relaciones clave incluyen la conexión entre Bill y ReservationBook a través de OrderID, y la relación entre Service y las tablas de pedidos y reservas.
¿Cuál es la estructura actual de la base de datos para el análisis de datos?	La base de datos actual es una plantilla universal diseñada para adaptarse a diversos clientes hoteleros, lo que resulta en campos no utilizados por todos los clientes. Además, debido a la integración de múltiples sistemas de información, hay presencia de campos en blanco en numerosas tablas. Esto complica el análisis y requiere una cuidadosa limpieza y preparación de datos.
¿Cómo se manejan los datos históricos en la base de datos actual?	Actualmente, la base de datos proporcionada para el análisis no llega al año de información.
¿Cómo debería manejar los datos nulos?	Los principales procesos de limpieza incluyen el manejo de valores nulos, especialmente en tablas como Bill, OrderCPL y ReservationBook. Se realiza la imputación de variables categóricas con el valor 0 o “No registrado”.
¿Cómo se manejan las inconsistencias en los datos entre diferentes sistemas de información?	Dada la naturaleza integradora de la base de datos, que recibe información de diversos sistemas, se requiere un cuidadoso proceso de mapeo y estandarización. Esto incluye la identificación de campos comunes entre sistemas y la creación de reglas de transformación para asegurar la consistencia de los datos.

Apéndice AG. Revisión documental

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
001	Julio	Pimienta, J. y Orden, A. (2012). Metodología de la investigación. Pearson. Libro: Bases de datos suscritas (SIBITEC).	Metodología de la investigación.	Delimita los tipos principales de investigación, sus características principales y el propósito.	En base a lo descrito por este libro, se determinó que tipo de investigación es la que se adapta más a este proyecto de investigación y que contempla el tipo de investigación a desarrollar.
002	Julio	Procom. (2024, 27 de mayo). Sobre nosotros. https://www.procom.cr/sobreprocom/ Entrevista a la empresa	Sobre nosotros.	Describe todos lo que requiere y realiza la empresa para optimizar procesos, mejorar la eficiencia operativa, impulsar la productividad y elevar la competitividad en un entorno empresarial.	Fue utilizada para la obtención de datos de la empresa como: misión, visión, valores, equipo de trabajo, diagrama organizacional
003	Julio	Rodríguez et al., (2023). Evaluación de la calidad de los reportes proporcionados por Procom a sus clientes hoteleros. <i>Revista de Gestión Hotelera</i> , 15(2), 55-72. Página Web, PROCOM.	Evaluación de la calidad de los reportes proporcionados por Procom a sus clientes hoteleros.	Estudio en el cual se destaca que los reportes proporcionados a los clientes hoteleros por parte de PROCOM no están cumpliendo con su objetivo de facilitar la toma de decisiones informadas.	Detalla información de suma relevancia sobre la problemática que enfrenta PROCOM tras no lograr que sus clientes no tomen decisiones informadas.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
004	Julio	Vargas Morúa, E. y Ulate Soto, I. (2016). Metodología para elaborar una tesis. MONTES DE OCA, Costa Rica: Editorial Universidad Estatal a Distancia. Sistema de Bibliotecas (SIBITEC)	Metodología para elaborar una tesis.	Este documento es una guía diseñada para ayudar a los estudiantes en la elaboración de tesis a nivel de grado. Su propósito es proporcionar orientaciones prácticas y herramientas necesarias para llevar a cabo investigaciones de manera efectiva.	Dio conocimiento sobre cómo se clasifican las fuentes de información.
005	Julio-Agosto	Bernal, C. A. (2010). Metodología de la investigación: administración, economía, humanidades y ciencias sociales (3ra edición.). Pearson Educación. Libro: Sistema de Bibliotecas (SIBITEC)	Metodología de la investigación: administración, economía, humanidades y ciencias sociales.	El texto se estructura para facilitar la comprensión de los métodos de investigación y su aplicación en diversas disciplinas, incluyendo administración y ciencias sociales. Proporciona herramientas y estrategias para el diseño, desarrollo y presentación de proyectos de investigación, así como para la recolección de datos y el uso de instrumentos de medición	Fue utilizado para determinar cuáles son los beneficios de este sistema de información y si se adapta a PROCOM.
006	Julio-Agosto	DataAnalytics. (2023). Informe sobre el aprovechamiento de datos en empresas del sector de soluciones empresariales. DataAnalytics	Informe sobre el aprovechamiento de datos en empresas del sector de soluciones	Estudio realizado, en el cual se determina que el 75% del sector tecnológico no aprovechan la información generada por sus sistemas, así como sucede en PROCOM.	Se usó para demostrar la importancia de dar buen uso a los datos para soluciones empresariales.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
		Google Académico	empresariales.		
007	Julio-Agosto	Davenport, T. H., & Harris, J. G. (2017). Competing on analytics: The new science of winning (Updated, with a new introduction). Harvard Business Review Press. Sistema de Bibliotecas (SIBITEC)	Competing on analytics: The new science of winning (Updated, with a new introduction).	Este documento presenta información acerca de cómo las organizaciones que adoptan un enfoque basado en datos tienen mayores probabilidades de éxito en diversos aspectos de su negocio, además abarca temas de evolución de la analítica, nivel de madurez analítica, desafíos y barreras.	Fue usado para demostrar que los cuando las empresas ven resultados tangibles derivados del análisis de datos, son propensos a creer en el valor de un enfoque basado en datos y a adoptarlo en sus propias prácticas.
008	Julio-Agosto	Hernández, R., Baptista, P. y Fernández, C. (2014). Metodología de la investigación. McGraw-Hill. http://www.ebooks7-24.com.ezproxy.itcr.ac.cr/?il=285 Libro: Bases suscritas (SIBITEC)	Metodología de la investigación.	Aborda los distintos enfoques y procesos de la investigación científica.	Fue usado para el desarrollo de toda la parte metodológica del desarrollo de este proyecto.
009	Julio-Agosto	Procom. (2024, 27 de Julio). Algunos Casos de Estudio. Recuperado de https://www.procom.cr/casos-de-exito/ Entrevista	Casos de Estudio.	Describe proyectos o casos realizados similares de manera interna y externa, mediante implementación de BI y otros.	Se utilizó para ver el tipo de proyectos realizados por PROCOM de manera interna y externa y si existía relación con este proyecto en proceso.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
010	Agosto	García, J., & López, M. (2024). Optimización del rendimiento de aplicaciones en la nube mediante análisis de datos: Un caso de estudio en una empresa de software como servicio (SaaS). Conferencia Internacional de Computación en la Nube y Servicios.	Optimización del rendimiento de aplicaciones en la nube mediante análisis de datos.	Describe proyectos que implementaron el uso de BI para optimizar el rendimiento de las aplicaciones en la nube y, así mismo, como lograron mejorar el rendimiento de las aplicaciones en la nube	Fue utilizado para la obtención de conocimientos sobre BI y como referencia para la implementación de herramientas y tecnologías de BI en PROCOM.
011	Agosto	Gartner. (2021). Data and analytics leaders must embrace a data-driven culture to drive value. https://www.gartner.com/en/documents/4004063 Página Web: Gartner, empresa de TI.	Data and analytics leaders must embrace a data-driven culture to drive value.	Detalla la importancia de adoptar una cultura data-driven para aquellas empresas líderes de datos y del análisis de estos mismos.	Se usó para determinar la importancia de la toma de decisiones basada en datos, y la urgencia de adopción de una cultura data-driven.
012	Agosto	Goleman, D. (1998). Emotional intelligence: Why it can matter more than IQ. Bantam Books. Sistemas de Bibliotecas del TEC (SIBITEC)	Emotional intelligence: Why it can matter more than IQ.	Destaca la importancia de la inteligencia emocional en el ámbito profesional, particularmente en roles de tecnología de la información (TI).	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
013	Agosto	Kalakota, R., & Robinson, M. (2001). e-Business 2.0: Roadmap for success. Addison Wesley. Sistemas de Bibliotecas del TEC (SIBITEC)	e-Business 2.0: Roadmap for success.	Documento en el cual se define a los sistemas de información como el componente eficaz dentro de la inteligencia de negocios.	Se extrajo que, los sistemas de información desempeñan un papel crucial en la gestión eficiente de las operaciones y en la mejora de la experiencia del cliente y como adaptar estos según objetivos de diferentes tipos de empresas.
014	Agosto	Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36(5), 700-710. Sistemas de Bibliotecas del TEC (SIBITEC)	A review and future direction of agile, business intelligence, analytics and data science.	Detalla la funcionalidad, características y demás información relevante sobre BI.	Menciona como la BI moderna fomenta una cultura de toma de decisiones basada en datos. Y considerando características la aplicación de BI calza a a la perfección este caso para ser utilizado.
015	Agosto	Jones, A., & Smith, B. (2023). Data Mining Project Management: Risks and Strategies. Publisher Name. Google académico	Data Mining Project Management: Risks and Strategies.	Detalla y define que el análisis de riesgos en proyectos de minería de datos y su importancia.	Uso en la comprensión y análisis de posibles riesgos a suceder en minería de datos y en el transcurso del desarrollo de cambios de la empresa.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
016	Agosto	Millán-García, C. H., & Gómez-Díaz, M. R. (2018). Factores e indicadores de competitividad hotelera. <i>Compendium</i> , 21(40), 1-18. Recuperado de https://www.redalyc.org/articulo.oa?id=88055200021 Sistema de Bibliotecas del Tec(SIBITEC)	Factores e indicadores de competitividad hotelera.	El documento aborda cómo la correcta identificación y utilización de estos indicadores ayudan a las empresas a mejorar su posición en el mercado y a tomar decisiones informadas.	Análisis de factores e indicadores de competitividad populares y que tan eficaces son.
017	Agosto	Munappy, A. R., Mattos, D. M. F., Boer, J. M. A., Moreira, V. P., Fonseca, M. J., & Ferreira, P. R. (2020). An overview of data extraction transformation and loading. <i>IEEE Access</i> , 8, 192095-192112. https://doi.org/10.1109/ACCESS.2020.3031971 Google Académico	An overview of data extraction transformation and loading.	Artículo usado para la definición y comprensión de El proceso ETL (Extracción, Transformación y Carga) en el análisis de datos.	Explica de manera concreta el proceso ETL y como se aplican en el análisis de datos en diversos ámbitos de estudio.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
018	Agosto	Ñaupas, H., Mejía, E., Novoa, E., & Villagómez, A. (2013). Metodología de la investigación Cuantitativa - Cualitativa y Redacción de la Tesis. Bogotá Colombia: Ediciones de la U. Sistema de Bibliotecas del Tec(SIBITEC)	Metodología de la investigación Cuantitativa - Cualitativa y Redacción de la Tesis	Menciona técnicas utilizadas para la identificación de riesgos y como varían dependiendo de la documentación y demás.	Aportó en la comprensión de la comprensión de riesgos según el tipo de documentación.
019	Agosto	Project Management Institute. (2021). A guide to the project management body of knowledge (PMBOK guide) (7th ed.). Project Management Institute. Sistema de Bibliotecas del Tec(SIBITEC)	A guide to the project management body of knowledge (PMBOK guide)	Esta guía se encarga de interpretar los riesgos identificados en un proyecto y el proceso implica el uso de diferentes métodos cualitativos y cuantitativos para evaluar los riesgos y priorizarlos en función de su severidad.	Se utilizó para la formación y evitar futuras problemáticas en el desarrollo del proyecto en cuanto al uso de métodos cualitativos y cuantitativos.
020	Agosto	Revista Internacional de Comercio Electrónico y Marketing.	La implementación de un módulo de BI para el análisis de datos de clientes en una tienda online de ropa.	Se centra en la implementación de un módulo de Inteligencia Empresarial (BI) en una tienda online de ropa. Su objetivo es mejorar el análisis de datos de clientes, lo que permite a la tienda tomar decisiones informadas y estratégicas basadas en datos.	Usado para la adquisición de conocimientos cómo: implementar soluciones de BI para obtener información valiosa sobre los clientes y mejorar las estrategias de negocio.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
021	Agosto	Ramamohanarao, Kotagiri, et al. Advances in Databases: Concepts, Systems, and Applications: 12th International Conference on Database Systems for Advanced Applications, DASFAA 2007, Bangkok, Thailand, April 9-12, 2007: Proceedings. 2007th ed. Vol. 4443. Sistema de Bibliotecas del Tec(SIBITEC)	Database system concepts	Define qué es un sistema gestor de bases de datos y sus aplicaciones.	Se estudio este sistema gestor de bases de datos para ver su adaptabilidad a diferentes bases de datos, cuáles son sus características y como funcionaría para PROCOM.
022	Agosto-septiembre	Bilgihan, A., Cobanoglu, C., & Miller, B. L. (2021). Hospitality information technology: Learning how to use it (9th ed.). Kendall Hunt Publishing. Biblioteca Dr. Jorge Villalobos Padilla/ Semantic Scholar.	Hospitality information technology: Learning how to use it.	Aborda la importancia de la tecnología de la información en la industria hotelera, centrándose en cómo estas herramientas mejoran la eficiencia, el servicio al cliente y, la rentabilidad. Los capítulos están diseñados para ofrecer respuestas a preguntas clave sobre el uso de sistemas de información para mejorar la toma de decisiones y optimizar las operaciones en la hospitalidad	Se encontró información detallada sobre la importancia de la anticipación de tendencias dentro del mercado, personalizar las experiencias de los huéspedes en base a datos y sus beneficios.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
023	Agosto-septiembre	Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc. Google Académico	CRISP-DM 1.0: Guía paso a paso de minería de datos.	Detalla el uso de CRISP-DM en el manejo de datos.	De este documento se extrajo información sobre el método Cross-Industry Standard Process for Data Mining y cómo aplicarlo para PROCOM.
024	Agosto-septiembre	Inmon, W. H. (2005). Building the Data Warehouse (4th ed.). Wiley. Sistema de Bibliotecas del Tec(SIBITEC)	Building the Data Warehouse	Este autor, mediante este artículo, describe la metodología Inmon, cuál es su enfoque, detalla sus fases y el orden por seguir para lograr un data marts o Data Warehouse.	Comprensión, uso y desarrollo de la metodología Inmon para Data Warehouse.
025	Agosto-septiembre	Kimball, R., & Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling (3rd ed.). Wiley. Sistema de Bibliotecas del Tec(SIBITEC)	The data warehouse toolkit: The definitive guide to dimensional modeling	Describe los principios fundamentales, fases de Kimball y demás en el ámbito de la Gestión de Datos e Información, también establece prácticas como la creación de un data mart y modelado dimensional.	Fue usada para extraer técnicas fundamentales para crear una arquitectura de datos robusta y escalable que soporte las necesidades analíticas de PROCOM.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
026	Agosto-septiembre	Olszak, C. M. (2016). Toward better understanding and use of business intelligence in organizations. Information Systems Management, 33(2), 105-123. https://doi.org/10.1080/10580530.2016.1155946 Sistema de Bibliotecas del Tec(SIBITEC)	Toward better understanding and use of business intelligence in organizations.	Explora cómo las organizaciones mejoran su comprensión y uso de la inteligencia empresarial (BI), se enfatiza la importancia de la BI en el proceso de toma de decisiones, así como en la planificación estratégica.	Para conocer acerca de los riesgos comunes en este tipo de proyectos o similares.
027	Agosto-septiembre	Project Management Institute. (2017). MoSCoW prioritization. En A guide to the project management body of knowledge (PMBOK® guide) (6th ed.) (pp. 250-252). Project Management Institute. Sistema de Bibliotecas del Tec(SIBITEC)	MoSCoW prioritization. En A guide to the project management body of knowledge	Define la técnica MoSCoW y cuáles son sus categorías.	Fue usado para el análisis y comprensión de la técnica MoSCow.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
028	Agosto-septiembre	Rondi, G. R., Casal, M. del C., & Giai Levra, V. (2019). Análisis de los principales indicadores de desempeño utilizados en la actividad hotelera. Contabilidad y Auditoría, 50, 142-181. https://nulan.mdp.edu.ar/id/eprint/3285/1/rondi-et-al-2019.pdf Sistema de Bibliotecas del Tec(SIBITEC)	Análisis de los principales indicadores de desempeño utilizados en la actividad hotelera. Contabilidad y Auditoría.	Se centra en la identificación y evaluación de indicadores clave de desempeño en el sector hotelero. Los autores realizan un análisis crítico de diversos indicadores, tales como RevPAR (Revenue Per Available Room), ARR/ADR (Average Room Rate/Average Daily Rate) y GopPAR (Gross Operating Profit Per Available Room), entre otros.	Fue usado para realizar una matriz comparativa entre indicadores en hoteles de interés y descripción.
029	Agosto-septiembre	Universidad de Tecnología. (2022). Estudio sobre la percepción de las empresas acerca de ser organizaciones data-driven. Universidad de Tecnología. Biblioteca electronica: SciELO	Estudio sobre la percepción de las empresas acerca de ser organizaciones data-driven.	Se analiza la percepción de las empresas sobre su madurez como organizaciones data-driven, es decir, aquellas que basan sus decisiones estratégicas y operativas en el análisis de datos. El estudio encontró que el 60% de las empresas del sector confunden los conceptos de data-driven e inteligencia empresarial.	Fue usada para demostrar el % de empresas del sector que confunden los conceptos de data-driven e inteligencia empresarial. Y definir la brecha de conocimientos en la industria.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
030	Septiembre	Brynjolfsson, Erik, and Kristina McElheran. "The Rapid Adoption of Data-Driven Decision-Making." <i>The American economic review</i> 106.5 (2016): 133–139. Print. Sistema de Bibliotecas del Tec(SIBITEC)	Data in action: Data-driven decision making in US manufacturing	Este documento investiga la adopción y el impacto del uso de decisiones basadas en datos en la industria manufacturera estadounidense. El documento enfatiza la importancia de complementariedades organizacionales, donde las decisiones basadas en datos y la alta inversión en capital tecnológico se refuerzan mutuamente.	Se usó para la identificación de riesgos en este tipo de proyectos y entablarlos según la probabilidad de ocurrencia y su posible impacto en la organización.
031	Septiembre	Cámara Costarricense de Hoteles. (2021). Las 6 mejores prácticas para elegir KPIs para tu negocio. Recuperado de https://www.ecobusiness.fund/fileadmin/user_upload/Sustainability_Academy/Recursos/herramienta_turismo_5.1.pdf . Página Web	Las 6 mejores prácticas para elegir KPIs para tu negocio.	Define y menciona las características relevantes del sistema de información (Plataforma de gestión para servicios de alimentación) y su aporte a restaurantes o servicios de alimentos.	Se usó para identificar <i>KPIs</i> relevantes para el análisis.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
032	Septiembre	Data Management Association International. (2017). The DAMA guide to the data management body of knowledge (DAMA-DMBOK) (2nd ed.). Technics Publications. Librería Digital ACM	The DAMA guide to the data management body of knowledge (DAMA-DMBOK)	Detalla los enfoques utilizados para la construcción de un data warehouse, en específico el siguiente: Enfoque Top-Down (Inmon).	Fue requerido para comprender los enfoques del data warehouse posible a aplicar.
033	Septiembre	Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of business process management (2nd ed.). Springer. Sistema de Bibliotecas (SIBITEC)	Fundamentals of business process management	Aborda el ciclo de vida de la gestión de procesos de negocio (BPM), que incluye la identificación, modelado, análisis, automatización y monitoreo de procesos. También explora conceptos, métodos y herramientas de gestión.	Para conocer acerca de los riesgos comunes en este tipo de proyectos o similares. Asimismo, se usó como ayuda para la fase de implementación de medidas para gestionar los riesgos entablados del proyecto.
034	Septiembre	Failte Ireland. (2023). Hotel Key Performance Indicators. Recuperado de https://supports.failteireland.ie/wp-content/uploads/2021/07/Hotel-Key-Performance-Indicators.pdf Página Web Failte Ireland: Publican sus artículos de investigación.	Hotel Key Performance Indicators.	Informe en el cual el tema central es sobre indicadores de desempeño hotelero.	Se analizaron los indicadores de desempeño hotelero estudiados en el informe y se relacionaron o sacó información parecida a algunos de los clientes de hoteles de PROCOM. Se analizaron los indicadores de este documento.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
035	Septiembre	Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE Publications Ltd. Base datos (SIBITEC)	Discovering statistics using IBM SPSS statistics	Está diseñado para ayudar a estudiantes e investigadores a aprender cómo aplicar técnicas estadísticas usando el software SPSS (Statistical Package for the Social Sciences).	Se usó para ver ejemplos prácticos y casos reales y aplicar las técnicas en diferentes situaciones de la investigación. El tema importante que se extrajo de este documento fue la información de correlaciones Pearson y Spearman.
036	Septiembre	Hu, H., Wen, Y., Chua, T. S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. IEEE Access, 2, 652-687. https://doi.org/10.1109/ACCESS.2014.2332453 Base de datos suscritas (SIBITEC)	Toward scalable systems for big data analytics: A technology tutorial.	Se centra en los sistemas escalables para el análisis de grandes volúmenes de datos. Y la necesidad de crear nuevas arquitecturas de sistemas para la adquisición, transmisión, almacenamiento y procesamiento a gran escala de estos datos.	Se usó para determinar posibles riesgos a futuro, los cuales son comunes o similares. en este tipo de proyectos.
037	Septiembre	IBM. (2020). The data-driven organization: Unlocking greater value from data. IBM Corporation. https://www.ibm.com/downloads/cas/XKYL87ZJ Página Web: IBM	The data-driven organization: Unlocking greater value from data.	El informe describe estrategias para cultivar una cultura de datos, incluido hacer que los datos sean accesibles en el formato correcto, con una gobernanza y las herramientas adecuadas para capacitar a los empleados para tomar decisiones informadas.	Para definir que es una organización Data-driven.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
038	Septiembre	Kotu, V., & Deshpande, B. (2015). Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner. Elsevier. Base de datos (SIBITEC)	Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner	En este artículo, se abarca la definición, características y factores importantes sobre la minería de datos, así como las técnicas que esta incluye.	Comprensión de la minería de datos, cuáles técnicas incluye y cómo se utilizarían estas en PROCOM.
039	Septiembre	Lucid Software Inc. (2023). Qué es un diagrama entidad-relación. Lucidchart. https://www.lucidchart.com/pages/es/que-es-un-diagrama-entidad-relacion Página Web: Lucidchart	Qué es un diagrama entidad-relación.	Conceptualiza que es un modelo entidad- relación mediante ejemplos gráficos y delimita cuáles son sus componentes claves.	Se usó para hacer un diagrama sobre entidad-relación de los datos de las tablas relevantes del análisis de datos de PROCOM.
040	Septiembre	Malik, S. (2005). Enterprise dashboards: Design and best practices for IT. Wiley. Bases de datos (SIBITEC)	Enterprise dashboards: Design and best practices for IT.	Definición de los diferentes gráficos, en qué casos de análisis de datos son usados y su visualización, entre otros. Por otra parte, abarca los tipos de dashboards en práctica de IT.	Comprensión de los gráficos de barra, mapas de color y demás. Reconocimiento y diferenciación de dashboards y cuáles se adecuan a PROCOM.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
041	Septiembre	Nguyen, T., Malecki, R., & Kanat, I. E. (2020). Building a data-driven culture: Creating change through analytic capability. The Journal of Product Innovation Management, 37(2), 142-156. https://doi.org/10.1111/jpim.12527 Base de datos (SIBITEC): ScienceDirect.	Building a data-driven culture: Creating change through analytic capability.	Se encuentra información sobre que cultura data-driven y lo que esta implica en una empresa para lograrla.	Conocimiento sobre cultura data-driven.
042	Septiembre	Parmenter, D. (2015). Key performance indicators: Developing, implementing, and using winning KPIs (3rd ed.). John Wiley & Sons. Sistema de Bibliotecas del TEC (SIBITEC).	Key performance indicators: Developing, implementing, and using winning KPIs.	Es una guía integral sobre cómo seleccionar, implementar y utilizar indicadores clave de rendimiento (KPIs) de manera efectiva en las organizaciones.	Fue usada para obtener información de (KPIs) relacionada a la industria hotelera.
043	Septiembre	Ponniah, P. (2000). Data Warehousing fundamentals. New York: John Wiley & Sons, Inc. Sistema de Bibliotecas del TEC (SIBITEC).	Data Warehousing fundamentals.	Explica mediante que es un esquema de estrella mediante ejemplos gráficos de otras empresas, delimita que factores abarca.	Solo fue explorado para ser comparado con el esquema de copo de nieve y cual se adaptaba al estudio.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
044	Septiembre	Srivastava, N., & Maitra, R. (2016). Key Performance Indicators (KPI) in Hospitality Industry: An Emphasis on Accommodation Business of 5 Star Hotels of National Capital Region. International Journal of Research in Tourism and Hospitality, (1), 34-40. Plataforma digital y base de datos SAGE Journals.	Performance Indicators (KPI) in Hospitality Industry: An Emphasis on Accommodation Business of 5 Star Hotels of National Capital Region.	En este artículo de revista, se detalla los indicadores clave de desempeño (KPIs) en hoteles de cinco estrellas.	KPIs usados en hoteles y ver qué relación o que información es importante adoptar para los clientes de industria hotelera de PROCOM.
045	Septiembre	Vaske, J. J., Beaman, J., & Ring, A. (2017). Cramer's V: A measure of association for nominal variables. Journal of Statistical Education, 25(2), 86-95. https://doi.org/10.1080/10691898.2017.1362402 Revista Académica de acceso abierto: Universidad Estatal Carolina del Norte.	Cramer's V: A measure of association for nominal variables.	Describe el uso de la medida de asociación Cramér's V, que se utiliza para evaluar la relación entre dos variables nominales. Este coeficiente cuantifica la fuerza de la asociación en tablas de contingencia.	Fue usado para comprender crammers v y realizar correlaciones categóricas.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
046	Septiembre	Vásquez Núñez, J. D. (2012). DICCIONARIO DE DATOS Sistema de Control de Producción SICPROD V1.0. Universidad Técnica del Norte. Sistema bibliotecas del TEC (SIBITEC). Página web del Repositorio Digital Universidad Técnica del Norte.	DICCIONARIO DE DATOS Sistema de Control de Producción SICPROD V1.0.	Explica de manera detallada mediante ejemplos como comprender una base de datos, según columnas y demás.	Describir y comprender la estructura de los datos seleccionados para el análisis.
047	Septiembre	Ward, J., & Peppard, J. (2016). Aligning information systems with business strategy. Journal of Information Systems, 30(2), 115-135. Sistema bibliotecas del TEC (SIBITEC).	Aligning information systems with business strategy	Menciona la relevancia de maximizar el valor de las inversiones en TI, un principio que guía el diseño y la implementación de soluciones o mejoras.	Fue usado para comprender La Estrategia de Tecnología de Información y como implementar los procesos de análisis de datos para mejorar el módulo de BI existente en la empresa en estudio.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
048	Septiembre-octubre	Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. International Journal of Endocrinology and Metabolism, 10(2), 486-489. Base de datos de libre acceso: PubMed	Normality tests for statistical analysis: A guide for non-statisticians.	Contiene información sobre pruebas de normalidad para datos no paramétricos.	Determina pruebas de normalidad, según la información dada por este artículo se seleccionó para saber que test de normalidad es el adecuado usar teniendo una muestra de $n > 5000$.
049	Septiembre-octubre	IBM. (2021). Schemas de copo de nieve. IBM Knowledge Center. Recuperado de https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-snowflake Página web: IBM	Schemas de copo de nieve. IBM Knowledge Center.	Este documento explica una técnica de modelado de datos utilizada en almacenes de datos.	Se utilizó como guía para el diseño del DataMart.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

No. Revisión	Fecha	Fuente del documento	Nombre del documento	Descripción	Observaciones
050	Septiembre-octubre	NIST/SEMATECH. (2012). Handbook of statistical methods (NIST handbook). National Institute of Standards and Technology. https://doi.org/10.18434/M32189 Sistema de Bibliotecas del TEC (SIBITEC).	Handbook of statistical methods (NIST handbook).	Es una guía práctica sobre una amplia gama de técnicas estadísticas. Abarca fundamentos de estadística, distribuciones de probabilidad, pruebas de hipótesis y estimación, entre otros.	Para definir los tipos de técnicas estadísticas.
051	Septiembre-octubre	Tableau. (2014). Tableau visual guidebook. Tableau. https://www.tableau.com/es-es/learn/whitepapers/tableau-visual-guidebook Página web: Tableau.	Tableau visual guidebook.	Muestra las mejores prácticas para la visualización de los datos y, asimismo, cuáles son los gráficos más recomendados para análisis de resultados de datos.	Se usó para las mejores prácticas de creación de dashboards.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice AH. Diccionario de datos a explorar

Nombre de la Tabla	Descripción	Importancia para el proyecto
Bill	Información general de facturas	Fundamental para el seguimiento financiero y análisis de ingresos. Permite un análisis de los patrones de gasto de los clientes.
BillDetail	Detalle de productos/servicios que componen la factura.	Proporciona una visión de los productos/servicios consumidos. Además, facilita la identificación de servicios populares.
CostCenter	Actividades que generan valor para el hotel	Crucial para visualizar los nombres de las actividades.
Package	Paquetes que ofrece el hotel	Esencial para entender los paquetes que se ofrecen en el hotel
PackageDetail	Detalles de paquetes ofrecidos a clientes	Ayuda a analizar la efectividad de los paquetes y su impacto en el RevPAR.
Rate	Tarifas de los servicios que ofrecen	Se desglosan las tarifas que de los servicios que se utilizan para ADR, RevPAR, y TRevPAR
ReservationBook	Libro de reservas	Contiene información que ayuda a observar patrones de demanda y la disponibilidad de habitaciones.
ReservationBookDetail	Libro de reservas detallado	Contiene información de reservas que permite analizar las preferencias de los clientes.
Room	Información de habitaciones	Contiene información sobre el nombre de las habitaciones que sirven para la tasa de ocupación y RevPAR
RoomType	Tipo de habitación	Contiene los tipos de habitaciones disponibles que ayuda a el análisis de rentabilidad por tipo de habitación, ADR y RevPAR.
Service	Servicios que ofrece el hotel	Contiene información sobre los servicios que se ofrecen en los hoteles que permite observar cómo estos impactan en las ingresos
ServiceCategory	Categoría de servicio	Ayuda a organizar y analizar los servicios ofrecidos.
ReservationService	Reservaciones de servicios	Permite entender qué servicios están siendo reservados por los huéspedes. Ayuda a identificar los servicios de mayor popularidad y rentabilidad.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Nombre de la Tabla	Descripción	Importancia para el proyecto
ReservationServiceDetail	Detalle de cada reserva dentro de ReservationService	Complementa a ReservationService proporcionando detalles de cada servicio reservado. Su importancia está en los datos granulares que ofrece, permitiendo un análisis de cómo se utilizan los servicios.
Order	Órdenes creadas a partir de la entrada de un cliente al hotel	Esta tabla es de suma importancia para el proyecto. Registra las órdenes o pedidos realizados por los huéspedes, lo que es fundamental para crear un modelo que segmente y analice los perfiles de clientes basados en sus gastos. Permite identificar patrones de consumo generales y es clave para entender el comportamiento de compra de los clientes durante su estancia.
OrderDetail	Detalle de las órdenes en la tabla Order	Proporciona información sobre cada orden, lo que la hace valiosa para el análisis del comportamiento del cliente. Su importancia radica en que permite entender exactamente qué están comprando los clientes, en qué cantidades y a qué precios.
OrderCPL	Órdenes de cortesía que no generan facturación	OrderCPL es importante para entender cómo las ofertas de cortesía impactan en las ventas.
OrderCPLDetail	Detalle de las órdenes en la tabla OrderCPL	Complementa a OrderCPL ofreciendo detalles de las cortesías otorgadas. Su importancia radica en permitir un análisis de cortesías.
Log_Stays	Registra información sobre la estancia de cada huésped.	Proporciona un registro detallado de las estancias de los huéspedes, lo que es esencial para identificar patrones estacionales y factores que influyen en la ocupación. Es fundamental para analizar la duración de las estancias.

Apéndice AI. Datos seleccionados para el análisis

Tabla	Columnas	Llave Primaria	Llave foránea de tabla	Tipo de dato	Descripción	Justificación para el proyecto
Bill	Bill_ID	1		Numeric	Identificador único de la factura	Permite identificar cada factura, crucial para todos los análisis de ingresos y gastos.
	PackageID		Package	Numeric	Paquete facturado	Permite evaluar la eficacia de los paquetes y ofertas especiales. Es crucial para analizar qué paquetes de mayor popularidad y rentabilidad.
	Service		Service	Numeric	Servicio facturado	Esencial para desglosar los ingresos por tipo de servicio. Permite calcular la rentabilidad por departamento y analizar qué servicios contribuyen al TRevPAR.
	IDReservationDetail		ReservationBook	Numeric	Reservación facturada	Facilita el análisis de la relación entre reservas y facturas, permitiendo evaluar patrones de gasto asociados a diferentes tipos de reservas.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Bill	SubTotal			Numeric	Monto total de un servicio, paquete o reservación	Facilita el análisis de ingresos netos por en el detalle de la factura de cada servicio, paquete o reserva
	QuantityHours			Numeric	Cantidad de horas que se usará el servicio, reservación o paquete	Útil para servicios facturados por hora, permitiendo analizar la relación entre duración del servicio e ingresos generados.
	Client			Numeric	Cliente facturado	Permite segmentar el análisis de gastos por cliente, crucial para calcular el Ingreso por Cliente (RPC).
	BillTotal			Numeric	Monto total de la factura sumando los subtotales del detalle	Componente clave para calcular RevPAR y TRevPAR.
	BillDate			Numeric	Fecha de emisión de la factura	Permite análisis temporales de ingresos para calcular ADR y ocupación por períodos.
ReservationBook	reservationBookID	1		Numeric	Identificador único de la reserva	Identifica cada reserva para todos los análisis relacionados.
	Confirmed			Numeric	Indica si la reserva está confirmada	Ayuda a distinguir entre reservas confirmadas y no confirmadas

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

ReservationBook	Client			Numeric	Cliente en el libro de reservas	Vincula reservas con clientes específicos para análisis de comportamiento
	Room		Room	Numeric	Identificador de la habitación reservada.	Permite analizar la ocupación y demanda por tipo de habitación.
	StarDate			POSIXct	Fecha de inicio de la estancia	Crucial para calcular la Tasa de Ocupación y Días de Ocupación.
	EndDate			POSIXct	Fecha de fin de la estancia	Crucial para calcular la Tasa de Ocupación y Días de Ocupación.
	AdultQuantity			Numeric	Cantidad de adultos	Permiten segmentar análisis por tipo de huésped y tamaño de grupo.
	ChildQuantity			Numeric	Cantidad de niños	Permiten segmentar análisis por tipo de huésped y tamaño de grupo.
	Quantity			Numeric	Número de horas de la reserva	Esencial para calcular la ocupación total y analizar patrones de reserva. Ayuda a determinar la demanda por tipo de habitación.
	ReservationBookPrice			Numeric	Precio total de la reserva	Fundamental para calcular el RevPAR y el ADR. Permite analizar los ingresos generados

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

ReservationBook						por cada reserva y compararlos entre diferentes tipos de habitaciones o temporadas.
	QuantityAdultExtra			Numeric	Número de adultos adicionales a la ocupación estándar.	Importante para analizar el impacto de huéspedes adicionales en los ingresos y costos.
	QuantityChildrenExtra			Numeric	Número de niños adicionales a la ocupación estándar.	Permite evaluar el impacto de los niños adicionales en los ingresos y costos
	QuantityInfants			Numeric	Número de infantes en la reserva.	Permite evaluar el impacto de los infantes adicionales en los ingresos y costos
Room	roomID	1		Numeric	Identificador único de la habitación	Identifica cada habitación para todos los análisis relacionados.
	roomTypeName		RoomType	Numeric	Nombre del tipo de habitación	Permite un análisis detallado y comprensible de los tipos de habitaciones.
	roomName			Character	Nombre de la habitación	Permite segmentar análisis por tipo de habitación, crucial para evaluar rendimiento y rentabilidad.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Service	Service_ID	1		Numeric	Identificador único del servicio	Identifica cada servicio para todos los análisis relacionados.
	serviceCategoryName			Character	Categoría del servicio	Permite agrupar servicios para análisis de rentabilidad y demanda.
	serviceName			Character	Nombre del servicio	Permite un análisis comprensible de los servicios ofrecidos.
	Price			Numeric	Identificador único de la categoría de servicio	Identifica cada categoría de servicio para análisis agrupados.
	CostCenterName			Character	Centro de costo asociado al servicio	Vincula servicios con centros de costo, crucial para calcular la Rentabilidad del Departamento.
Order	ID	1		Numeric	Identificador único de la orden	Esencial para rastrear y relacionar cada orden con otros datos del sistema.
	StarDate			POSIXct	Fecha de inicio de la orden	Analizar patrones temporales en las ventas y ocupación.
	EndDate			POSIXct	Fecha de finalización de la orden	Permite calcular la duración de las estancias y analizar patrones de reserva.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Order	Applied			Numeric	Indica si la orden se ha facturado	Útil para distinguir entre órdenes facturadas y las que no
	Client			Numeric	Identificador del cliente.	Fundamental para la segmentación de clientes y análisis de comportamiento individual.
	QuantityHours			Numeric	Cantidad de horas de la reserva.	Importante para analizar la duración de las estancias, especialmente en reservas por horas.
	Name			Character	Nombre de la orden o servicio.	Ayuda a categorizar y analizar tipos específicos de órdenes o servicios.
	Service		Service	Numeric	Identificador del servicio ordenado	Clave para analizar la popularidad y rentabilidad de diferentes servicios.
	IDReservationDetail		ReservationBook	Numeric	Enlace a los detalles de la reserva.	Permite relacionar la orden con información detallada de la reserva.
	QuantityAdultExtra			Numeric	Cantidades de adultos extra	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
	QuantityChildrenExtra			Numeric	Cantidades de Niños extra	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
Order	AdultQuantity			Numeric	Cantidades de adultos	Analizar la composición de los grupos de

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

						huéspedes y su impacto en las ventas y servicios.
	ChildrenQuantity			Numeric	Cantidades de niños	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
OrderCPL	ID	1		Numeric	Identificador único de la orden complementaria	Esencial para rastrear y relacionar cada orden con otros datos del sistema.
	StarDate			POSIXct	Fecha de inicio de la orden de cortesía	Analizar patrones temporales en las ventas y ocupación.
	EndDate			POSIXct	Fecha de finalización de la orden de cortesía	Permite calcular la duración de las estancias y analizar patrones de reserva.
	Quantity			Numeric	Cantidad de horas de la reserva.	Importante para analizar la duración de las estancias, especialmente en reservas por horas.
	Service		Service	Numeric	Servicio asociado a la orden complementaria	Clave para analizar la popularidad y rentabilidad de diferentes servicios.
	IDReservationDetail		ReservationBook	Numeric	Enlace a los detalles de la reserva.	Permite relacionar la orden con información detallada de la reserva.
	OrderCPLType			Numeric	Tipo de orden de cortesía.	Permite categorizar y analizar diferentes tipos de cortesías y su impacto.
OrderCPL						

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

	OrderID		Order	Numeric	Referencia a la orden principal.	Crucial para relacionar las cortesías con las órdenes regulares y analizar su efecto.
	Client			Numeric	Identificador del cliente.	Fundamental para la segmentación de clientes y análisis de comportamiento individual.
	QuantityAdultExtra			Numeric	Cantidades de adultos extra	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
	QuantityChildrenExtra			Numeric	Cantidades de niños extra	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
	AdultQuantity			Numeric	Cantidades de adultos	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
	ChildrenQuantity			Numeric	Cantidades de niños	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
ReservationService	ID	1		Numeric	Identificador único de la reserva de servicios	Esencial para rastrear y relacionar cada reserva de servicios con otros datos del sistema.
	ClientID			Numeric	Identificador del cliente.	Fundamental para la segmentación de clientes y análisis de

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

ReservationService						comportamiento individual.
	OrderID		Order	Numeric	Enlace a la orden relacionada.	Permite conectar servicios reservados con órdenes específicas
	Applied			Numeric	Indica si la orden se ha facturado	Útil para distinguir entre órdenes facturadas y las que no
	Total			Numeric	Monto total del servicio.	Crucial para análisis de ingresos y rentabilidad de servicios.
	IsCPL			Numeric	Indica si es un servicio de cortesía.	Importante para distinguir entre servicios pagados y cortesías en los análisis.
	Quantity			Numeric	Cantidad de horas del servicio reservado.	Útil para analizar el volumen de servicios solicitados y su popularidad.
	Service_ID		Service	Numeric	Servicio asociado a la reserva	Clave para analizar la popularidad y rentabilidad de diferentes servicios.
	Total			Numeric	Monto total por cada servicio.	Crucial para análisis de ingresos y rentabilidad de servicios.
	StarDate			POSIXct	Fecha de inicio de la orden complementaria	Analizar patrones temporales en las ventas y ocupación.
	EndDate			POSIXct	Fecha de finalización de la orden	Permite calcular la duración de las estancias y analizar patrones de

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

					complementaria	reserva.
Log_Stays	ID	1		Numeric	Identificador único del registro de estancia.	Esencial para rastrear cada estancia individualmente.
	OrderID		Order	Numeric	Enlaces a órdenes relacionadas.	Permiten relacionar las estancias con órdenes regulares y de cortesía para un análisis completo.
	OrderCPL_ID		OrderCPL	Numeric	Enlaces a órdenes de cortesía relacionadas.	Permiten relacionar las estancias con órdenes regulares y de cortesía para un análisis completo.
	Bill_ID		Bill	Numeric	Enlace a la factura.	Crucial para análisis financieros y de rentabilidad por estancia.
Log_Stays	Room_ID		Room	Numeric	Identificador de la habitación	Permite analizar preferencias de habitaciones y optimización de asignación de cuartos.
	StarDate			POSIXct	Fecha de inicio de la reserva de servicios	Analizar patrones temporales en las ventas y ocupación.
	EndDate			POSIXct	Fecha de finalización de la reserva de servicios	Permite calcular la duración de las estancias y analizar patrones de reserva.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

	QtyGuests			Numeric	Cantidad de invitados en la reservación	Analizar la composición de los grupos de huéspedes y su impacto en las ventas y servicios.
--	-----------	--	--	---------	---	--

Apéndice AJ. Diccionario de datos de los modelos

Tabla	Columna	Tipo de dato	Llave	Descripción	Justificación	Reglas ETL
model 1_cust omer_ segme ntatio n	id	integer	Primaria	Identificador único autoincremental	Garantiza un identificador único para cada registro	Autoincremental
	Client	integer		Identificador único del cliente	Permite enlazar con otras tablas y realizar análisis a nivel de cliente individual	Convertir a integer
	RPC	DECIMAL (18,2)		Revenue Per Client (Ingresos por Cliente)	Métrica clave para evaluar el valor monetario de cada cliente	Convertir a integer
	Client_Visits_Freque	integer		Frecuencia de visitas del cliente	Indica la lealtad del cliente y su potencial de generación de ingresos	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla	Columna	Tipo de dato	Llave	Descripción	Justificación	Reglas ETL
	ncy					
	AverageDuration	DECIMAL (18,2)		Duración promedio de estancia en días	Proporciona insights sobre los patrones de estancia de los clientes	Convertir a integer
	cluster	Integer		Segmento al que pertenece el cliente	Permite agrupar clientes con características similares para estrategias de marketing y servicio	Convertir a integer
model2service_analysis	id	Integer	Primaria	Identificador único autoincremental	Garantiza un identificador único para cada registro	Autoincremental
	Date	Date		Fecha del servicio	Identifica el día específico del servicio	Entra directo
	ServiceName	Varchar (100)		Nombre del servicio	Identifica el tipo de servicio ofrecido	Entra directo
	YearMonth	Nvarchar (100)		Año y mes en formato AAAA-MM	Permite agrupar y analizar datos por mes	Entra directo
	OrderCount	integer		Número de pedidos del servicio	Mide la popularidad de cada servicio	Convertir a integer
	TotalValue	DECIMAL (18,2)		Valor total de los pedidos del servicio	Mide el rendimiento financiero de cada servicio	Convertir a decimal
model2service_analysis	Season	Varchar (100)		Temporada (Invierno, Primavera, Verano, Otoño)	Permite analizar tendencias estacionales	Entra directo
model3room_revenue	id	integer		Identificador único autoincremental	Garantiza un identificador único para cada registro	Autoincremental
	Date	Date		Fecha de los ingresos	Identifica el día específico	Entra directo
	RoomName	Nvarchar(100)		Nombre de la habitación	Permite analizar el rendimiento por tipo de habitación	Entra directo
	BillTotal	DECIMAL (18,2)		Monto total de la factura	Mide los ingresos generados	Convertir a decimal
	Month_Year	Nvarchar(Mes y año en formato	Facilita el análisis mensual	Entra directo

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla	Columna	Tipo de dato	Llave	Descripción	Justificación	Reglas ETL
	ar	100)		AAAA-MM		
	MonthYear_Label	Nvarchar(100)		Etiqueta de mes y año (ej. "Ene 2023")	Proporciona una etiqueta legible para visualizaciones	Entra directo
	DayOfWeek	Nvarchar(100)		Día de la semana	Permite analizar patrones semanales	Entra directo
	Quarter	Nvarchar(100)		Trimestre y año (ej. "Q1-2023")	Facilita el análisis trimestral	Entra directo
model4reservationcomparison	id	integer	Primaria	Identificador único autoincremental	Garantiza un identificador único para cada registro	Autoincremental
	Date	Date		Fecha de la reserva	Identifica el día específico	Entra directo
	Type	Varchar(100)		Tipo de reserva (Habitaciones o Servicios)	Permite comparar entre reservas de habitaciones y servicios	Entra directo
	DayOfWeek	Varchar(100)		Día de la semana	Ayuda a identificar patrones semanales en las reservas	Entra directo
	Reservations	integer		Número de reservas	Cuantifica la demanda diaria	Convertir a integer

Apéndice AK. Diccionario de datos del *Data Mart* y reglas del ETL

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
DIM DATE	Time_ID	Identificador único para cada fecha	integer	Primaria	Permite una referencia eficiente a fechas específicas en otras tablas	PMS	Derivada de múltiples fuentes	Time_ID	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
	Registered Date	Fecha en formato estándar	Date		Facilita la visualización y el filtrado por fechas completas	PMS		RegisteredDate	
	Registered Year	Año de la fecha	integer		Permite análisis y agrupaciones por año	PMS		RegisteredYear	Convertir a integer
	Registered Month	Mes de la fecha	integer		Permite análisis y agrupaciones por mes	PMS		RegisteredMonth	Convertir a integer
	Registered Day	Día del mes	integer		Permite análisis detallados por día específico	PMS		RegisteredDay	Convertir a integer
DIM_CLIENT	Client_ID	Identificador único del cliente	integer	Primaria	Permite relacionar todas las actividades y transacciones de un cliente específico	PMS	Order OrderCP L Reservati onServic e Bill	Client_ID	Convertir a integer
DIM_ROOM	Room_ID	Identificador único de la habitación	integer	Primaria	Permite el seguimiento y análisis de actividades por habitación	PMS	Room RoomTy pe	Room_ID	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
					específica				
	RoomName	Nombre o número de la habitación	nVarchar(100)		Facilita la identificación humana de las habitaciones en reportes y análisis	PMS		RoomName	
DIM_SERVICO	ServiceID	Identificador único del servicio	integer	Primaria	Permite el seguimiento y análisis de servicios específicos	PMS		ServiceID	Convertir a integer
	ServiceName	Nombre del servicio	Nvarchar(100)		Proporciona una descripción clara del servicio para reportes y análisis	PMS	ServiceCategory CostCenter	ServiceName	
DIM_SERVICO	CostCenterName	Nombre del centro de costos asociado	Nvarchar(100)		Permite el análisis financiero y la asignación de costos por departamento	PMS		CostCenterName	
DIM_ORDER	Order_ID	Identificador único de la orden	integer	Primaria	Permite el seguimiento de órdenes específicas a lo largo del	PMS	Order	Order_ID	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
					sistema				
	Client_ID	ID del cliente asociado	integer	DIM CLIENT	Vincula la orden con un cliente específico para análisis de comportamiento del cliente	PMS		Client_ID	Convertir a integer
	OpeningDate	Fecha de apertura de la orden	integer	DIM DATE	Permite análisis temporales de las órdenes	PMS	OrderCPL	OpeningDate	Convertir a integer
	Applied	Indica si la orden ha sido aplicada	boolean		Facilita el seguimiento del estado de las órdenes	PMS	ReservationBook	Applied	Convertir a boolean
DIM_ORDER	InProcess	Indica si la orden está en proceso	boolean		Permite analizar la eficiencia en el procesamiento de órdenes	PMS		InProcess	Convertir a boolean
DIM_ORDERCPL	OrderCPLID	Identificador único de la orden complementaria	integer	Primaria	Permite el seguimiento de órdenes complementarias específicas	PMS	OrderCPL	OrderCPLID	Convertir a integer
	ClientID	ID del cliente asociado	integer	DIM CLIENT	Vincula la orden complementari	PMS		ClientID	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
DIM_ORDE RCPL					a con un cliente específico				
	ServiceID	ID del servicio asociado	integer	DIM SERVICE	Vincula la orden complementaria con un servicio específico	PMS		ServiceID	
	RoomID	ID de la habitación asociada	integer	DIM ROOM	Vincula la orden complementaria con una habitación específica	PMS		RoomID	Convertir a integer
	StartDate	Fecha de inicio de la orden complementaria	integer	DIM DATE	Permite análisis temporales de las órdenes complementarias	PMS		StartDate	Convertir a integer
	OrderCPL Name	Nombre o descripción de la orden complementaria	varchar (50)		Proporciona contexto adicional sobre el tipo de orden complementaria	PMS		OrderCPL Name	Directo
	TotalPaid	Total, pagado por la orden	Decimal (18,2)		Permite el análisis financiero de	PMS		TotalPaid	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
		complementaria			las órdenes complementarias				
DIM_RESERVATION BOOK	ReservationBook_ID	Identificador único del libro de reservas	integer	Primaria	Permite el seguimiento de reservas específicas	PMS	ReservationBook	ReservationBook_ID	Convertir a integer
	Confirmed	Indica si la reserva está confirmada	integer		Facilita el análisis de tasas de confirmación y cancelación	PMS		Confirmed	Convertir a integer
	Order_ID	ID de la orden asociada	integer	DIM ORDER	Vincula la reserva con una orden específica	PMS		Order_ID	Convertir a integer
	Client_ID	ID del cliente asociado	integer	DIM CLIENT	Permite el análisis de patrones de reserva por cliente	PMS		Client_ID	Convertir a integer
FACT_BILLS	Bill_ID	Identificador único de la factura	integer	Primaria	Permite el seguimiento y análisis de facturas específicas	PMS	Bill	Bill_ID	Convertir a integer
	Order_ID	ID de la orden asociada	integer	DIM ORDER	Vincula la factura con una orden	PMS		Order_ID	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
					específica				
FACT_BILLS	Client_ID	ID del cliente asociado	integer	DIM CLIENT	Permite el análisis de facturación por cliente	PMS		Client_ID	Convertir a integer
	Service_ID	ID del servicio facturado	integer	DIM SERVICE	Facilita el análisis de ingresos por tipo de servicio	PMS		Service_ID	Convertir a integer
	Room_ID	ID de la habitación facturada	integer	DIM ROOM	Permite el análisis de ingresos por habitación	PMS		Room_ID	Convertir a integer
	BillDate	Fecha de la factura	integer	DIM DATE	Permite análisis temporales de facturación	PMS		BillDate	Directo
	SubTotal	Subtotal de la factura	DECIMAL (18,2)		Facilita el análisis financiero detallado	PMS		SubTotal	Convertir a integer
	BillTotal	Total, de la factura	DECIMAL (18,2)		Permite el análisis de ingresos totales	PMS		BillTotal	Convertir a integer
	FACT_STAYS	StayID	Identificador único de la estancia	integer	Primaria	Permite el seguimiento y análisis de estancias específicas	PMS	Log_Stays	StayID

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
FACT_STAYS	Order_ID	ID de la orden asociada	integer	DIM ORDER	Vincula la estancia con una orden específica	PMS		Order_ID	Convertir a integer
	OrderCPL_ID	ID de la orden complementaria asociada	integer	DIM ORDER CPL	Permite relacionar la estancia con órdenes complementarias	PMS		OrderCPL_ID	Convertir a integer
	Bill_ID	ID de la factura asociada	integer	DIM BILL	Vincula la estancia con su facturación correspondiente	PMS		Bill_ID	Convertir a integer
	Room_ID	ID de la habitación	integer	DIM ROOM	Permite el análisis de ocupación por habitación	PMS		Room_ID	Convertir a integer
	StartDate	Fecha de inicio de la estancia	integer	DIM DATE	Facilita el análisis de duración de estancias y patrones temporales	PMS		StartDate	Convertir a integer
	EndDate	Fecha de fin de la estancia	integer	DIM DATE	Permite calcular la duración de las estancias	PMS		EndDate	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
	Qty_Guests	Cantidad de huéspedes	integer		Facilita el análisis de ocupación y capacidad	PMS		Qty_Guests	Convertir a integer
FACT_RESERVATION_SERVICES	ReservationService_ID	Identificador único del servicio de reserva	integer	Primaria	Permite el seguimiento de servicios de reserva específicos	PMS	ReservationService	ReservationService_ID	Convertir a integer
	Order_ID	ID de la orden asociada	integer	DIM ORDER	Vincula el servicio de reserva con una orden específica	PMS		Order_ID	Convertir a integer
	Client_ID	ID del cliente asociado	integer	DIM CLIENT	Permite el análisis de uso de servicios por cliente	PMS		Client_ID	Convertir a integer
	Service_ID	ID del servicio	integer	Numeric	Facilita el análisis de popularidad y rendimiento de servicios específicos	PMS		Service_ID	Convertir a integer
	StartDate	Fecha de inicio del	integer	POSIXct	Permite análisis	PMS		StartDate	Convertir a integer

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Tabla de destino	Columna de destino	Descripción	Tipo de dato	Llave	Justificación para el proyecto	Sistema de Origen	Tablas de Origen	Nombre del Campo	Reglas ETL
FACT_RESERVATION_SERVICES		servicio			temporales de uso de servicios				
	ReservationServiceTotal	Total, del servicio de reserva	DECIMAL (18,2)	POSIXct	Facilita el análisis financiero de los servicios de reserva	PMS		ReservationServiceTotal	Convertir a integer
	ReservationServiceDetailTotal	Total, detallado del servicio de reserva	DECIMAL (18,2)	Numeric	Permite un análisis financiero granular	PMS		ReservationServiceDetailTotal	Convertir a integer
	IsApplied	Indica si el servicio ha sido aplicado	boolean	Numeric	Facilita el seguimiento del estado de los servicios de reserva	PMS		IsApplied	Convertir a boolean
	IsCPL	Indica si es un servicio complementario	boolean	Numeric	Permite distinguir entre servicios estándar y complementarios	PMS		IsCPL	Convertir a boolean

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice AK. Firma de minutas del encargado de desarrollo

Constancia de participación	
Por medio de la presente, hago constar mi participación en las minutas y/o entrevistas siguientes:	
Nombre y Apellido:	Denis José Corrales Zepeda
Cargo/Rol:	PROCOM - Senior developer.
Código de Minutas o Entrevistas:	E-06
Firma:	
Con mi firma confirmo mi participación en esta sesión y valido el contenido tratado.	

Apéndice AL. Firma de minutas del encargado de BI

Constancia de participación	
Por medio de la presente, hago constar mi participación en las minutas y/o entrevistas siguientes:	
Nombre y Apellido:	Carlos Daniel Quesada Méndez
Cargo/Rol:	Encargado de BI
Código de Minutas o Entrevistas:	M-03,M-016, E-03
Firma:	
Con mi firma confirmo mi participación en esta sesión y valido el contenido tratado.	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice AM. Firma de minutas del encargado de soporte de TI

Constancia de participación	
Por medio de la presente, hago constar mi participación en las minutas y/o entrevistas siguientes:	
Nombre y Apellido:	Esteban Córdoba
Cargo/Rol:	Encargado de Soporte de TI
Código de Minutas o Entrevistas:	M-05, E-05
Firma:	
Con mi firma confirmo mi participación en esta sesión y valido el contenido tratado.	

Apéndice AN. Firma de minutas del gerente general.

Constancia de participación	
Por medio de la presente, hago constar mi participación en las minutas y/o entrevistas siguientes:	
Nombre y Apellido:	
Cargo/Rol:	
Código de Minutas o Entrevistas:	M-01,M-02,M-04, M-017,M-018,M-019,E-01,E-02,E-04
Firma:	<p>CHRISTIAN GERARDO VARGAS POVEDA (FIRMA)</p> <p>Firmado digitalmente por CHRISTIAN GERARDO VARGAS POVEDA (FIRMA) Fecha: 2024.10.29 11:50:10 -06'00'</p>
Con mi firma confirmo mi participación en esta sesión y valido el contenido tratado.	

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice AO. Firma de minutas de la gerente de operaciones

Constancia de participación	
Por medio de la presente, hago constar mi participación en las minutas y/o entrevistas siguientes:	
Nombre y Apellido:	
Cargo/Rol:	
Código de Minutas o Entrevistas:	M-03, M-04, M-019, E-03, E-04
Firma:	
Con mi firma confirmo mi participación en esta sesión y valido el contenido tratado.	



Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice AP. Acta de validación por parte del gerente general

ACTA DE VALIDACIÓN

Proceso de Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP-DM para la Mejora del Módulo de Inteligencia de Negocios de la Empresa PROCOM

Por medio de la presente, en mi rol como Gerente, valido el proceso de análisis de datos implementado, confirmando que:

1. He revisado detalladamente la metodología del análisis de datos
2. Los resultados obtenidos son coherentes y confiables para la empresa
3. El proceso genera valor tangible para las operaciones bajo mi supervisión
4. La información presentada cumple con los estándares de calidad requeridos

DATOS DEL GERENTE VALIDADOR

Nombre Completo: _____

Cargo: _____

Área/Departamento: _____

Fecha de Presentación: 28 de Octubre de 2024

Firma: **CHRISTIAN GERARDO VARGAS POVEDA (FIRMA)**
Firmado digitalmente por CHRISTIAN GERARDO VARGAS POVEDA (FIRMA)
Fecha: 2024.10.29 11:50:45 -06'00'

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice AQ. Acta de validación por parte de la gerente de operaciones

ACTA DE VALIDACIÓN

Proceso de Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP-DM para la Mejora del Módulo de Inteligencia de Negocios de la Empresa PROCOM

Por medio de la presente, en mi rol como Gerente, valido el proceso de análisis de datos implementado, confirmando que:

1. He revisado detalladamente la metodología del análisis de datos
2. Los resultados obtenidos son coherentes y confiables para la empresa
3. El proceso genera valor tangible para las operaciones bajo mi supervisión
4. La información presentada cumple con los estándares de calidad requeridos

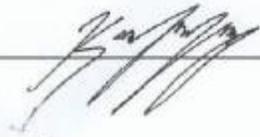
DATOS DEL GERENTE VALIDADOR

Nombre Completo: Paola Ovares Jara

Cargo: Gerente de Servicio

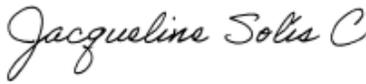
Área/Departamento: Customer Success.

Fecha de Presentación: 28 de Octubre de 2024

Firma: 

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

Apéndice AR. Firma de minutas de la profesora tutora

Constancia de participación	
Por medio de la presente, hago constar mi participación en las minutas y/o entrevistas siguientes:	
Nombre y Apellido:	Jacqueline Tatiana Solís Céspedes
Cargo/Rol:	Profesora Tutora
Código de Minutas o Entrevistas:	M-06, M-07, M-08, M-09, M-010, M-011, M-012, M-013, M-014, M-015, M-017, M-018, M-019
Firma:	
Con mi firma confirmo mi participación en esta sesión y valido el contenido tratado.	

10 ANEXOS

No existen anexos asociados para este documento.

11 Glosario

- **ADR** (*Average Daily Rate*): Tarifa diaria promedio que genera cada habitación ocupada.
- **Almacén de datos departamental** (*Data mart*): Subconjunto de un almacén de datos enfocado en un área específica del negocio.
- **Análisis de datos**: Proceso de inspección, limpieza, transformación y modelado de datos para descubrir información útil.
- **Análisis diagnóstico**: Tipo de análisis que examina por qué sucedió algo y sus causas.
- **Análisis predictivo**: Uso de datos históricos para predecir eventos futuros.
- **BI** (*Business Intelligence*): Inteligencia de Negocios, conjunto de estrategias y herramientas para analizar datos empresariales.
- **Benchmarking**: Proceso de comparar productos, servicios o procesos con los mejores del sector.
- **CRISP-DM**: Metodología estándar de minería de datos que define un ciclo de vida de proyectos de análisis.
- **Clustering**: Técnica de agrupamiento de datos según similitudes.
- **Cuadro de mando** (*Dashboard*): Panel visual que muestra métricas y KPIs clave del negocio.
- **DevExpress Report Designer**: Herramienta de software para crear reportes y visualizaciones interactivas.
- **Datos operativos**: Información generada por las operaciones diarias del negocio.
- **Datos depurados**: Información limpia y validada, lista para análisis.
- **ETL** (*Extract, Transform, Load*): Proceso de Extracción, Transformación y Carga de datos.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- **Esquema estrella:** Modelo de base de datos con una tabla central de hechos conectada a dimensiones.
- **Esquema copo de nieve:** Variante del esquema estrella donde las dimensiones están normalizadas.
Frecuencia de visitas: Número de veces que un cliente utiliza los servicios en un período.
- **Insights:** Conocimientos o comprensiones profundas derivadas del análisis de datos.
- **Indicadores clave de rendimiento (KPI):** Métricas que miden el desempeño de procesos críticos.
- **Inteligencia de Negocios:** Sistema para transformar datos en información útil para decisiones empresariales.
- **K-means:** Algoritmo de agrupamiento que divide datos en k grupos según similitud.
- **KPI (Key Performance Indicator):** Ver Indicadores clave de rendimiento.
- **Metodología:** Conjunto sistemático de métodos y procedimientos para alcanzar un objetivo.
- **Minería de datos:** Proceso de descubrir patrones en grandes conjuntos de datos.
- **Modelos analíticos:** Representaciones matemáticas o computacionales para analizar datos.
- **PMS (Property Management System):** Sistema de Gestión Hotelera.
- **Proceso ETL:** Ver ETL.
- **Punto de venta (POS):** Sistema para gestionar transacciones comerciales.
- **RevPAR (Revenue Per Available Room):** Ingresos por habitación disponible.
- **RPC (Revenue Per Client):** Ingresos por cliente.
- **R Studio:** Entorno de desarrollo integrado para el lenguaje de programación R.
- **Segmentación de clientes:** Proceso de dividir clientes en grupos según características similares.
- **SQL Server:** Sistema de gestión de bases de datos desarrollado por Microsoft.

Implementación de un Proceso de Análisis de Datos Basado en la Metodología CRISP- DM para la Mejora del Módulo de Inteligencia de Negocios de la empresa PROCOM

- **Stakeholders:** Partes interesadas o afectadas por un proyecto.
- **Tablero de control:** Ver Cuadro de mando.
- **Tasa de ocupación:** Porcentaje de habitaciones ocupadas respecto al total disponible.
- **TRevPAR** (*Total Revenue Per Available Room*): Ingresos totales por habitación disponible.
- **Up-selling:** Técnica de venta para ofrecer productos de mayor valor.
- **Usuario final:** Persona que interactúa directamente con el sistema o aplicación.
- **Valores atípicos** (*Outliers*): Datos que se desvían significativamente del comportamiento normal.
- **Variables derivadas:** Nuevas variables creadas a partir de datos existentes.
- **Visualización de datos:** Representación gráfica de información y datos.