

**IMPLEMENTACIÓN DE INTELIGENCIA DE NEGOCIO PARA EL ÁREA DE VENTAS Y
COMPRAS EN LA EMPRESA MATERIALES MADERAS SANTANDER SAS.**

**MACIAS REYES ANYELYN
ORTIZ REALES JUAN CARLOS**

**UNIVERSIDAD POPULAR DEL CESAR
FACULTAD DE INGENIERÍA Y TECNOLÓGICAS
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS
VALLEDUPAR- CESAR**

2024

**IMPLEMENTACIÓN DE INTELIGENCIA DE NEGOCIO PARA EL ÁREA DE VENTAS Y
COMPRAS EN LA EMPRESA MATERIALES MADERAS SANTANDER SAS.**

**MACIAS REYES ANYELYN
ORTIZ REALES JUAN CARLOS**

**DIRECTOR:
ING. AMILKAR SIERRA ROMANO E.**

**UNIVERSIDAD POPULAR DEL CESAR
FACULTAD DE INGENIERÍA Y TECNOLÓGICAS
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS
VALLEDUPAR- CESAR**

2024

CONTENIDO

Introducción	1
SECCIÓN I. DESCRIPCIÓN GENERAL	3
1.1 Título	3
1.2 Dirección de ejecución	3
1.3 Lapso de ejecución	3
1.4 Organismo o institución responsable	3
1.5 Información de contacto	3
1.6 Línea, Sublínea y grupo de investigación del proyecto.	3
1.6.1 Línea de investigación.	3
1.6.2 Sublínea de investigación.	3
1.6.3 Área de investigación.	4
1.6.4 Grupo de investigación.	4
SECCIÓN II. DESCRIPCIÓN SITUACIONAL	5
2.1 Identificación del problema	5
2.1.1 Formulación del problema	7
2.2 Justificación del proyecto	7
2.3 Objetivos del proyecto	8
2.3.1 Objetivo general	8
2.3.2 Objetivos específicos	8
2.4 Marco referencial	9
2.4.1 Marco conceptual	9
2.4.1.4 Modelo de madurez de inteligencia empresarial	10
2.4.2 Antecedentes	28
2.5 Marco metodológico	31

2.5.1 Tipo de estudio y diseño de investigación	31
2.5.2 Población y muestra	32
2.5.3 Instrumentos y técnicas de recolección de información	33
2.5.4 Metodología para el desarrollo del proyecto	34
2.6 Resultados esperados	37
2.7 Cronograma de actividades	38
2.8 Presupuesto	38
SECCIÓN III: DESARROLLO CIENTÍFICO – TECNOLÓGICO	41
3.1 Desarrollo de las fases de la metodología propuesta.	41
3.1.1 Planificación del proyecto	41
3.1.2 Análisis de requerimientos	42
3.1.3 Modelado dimensional	44
3.1.4 Diseño físico	53
3.1.5 Diseño e implementación del subsistema ETL	74
3.1.6 Implementación y desarrollo de los algoritmos seleccionados.	88
3.1.6 Implementación	104
3.2 ANÁLISIS DE RESULTADOS Y DISCUSIÓN	122
3.3 CONCLUSIONES	124
3.4 RECOMENDACIONES	126
3.5 BIBLIOGRAFÍA	

ÍNDICE DE TABLAS.

Tabla 1 Información de contacto.	3
Tabla 2 Top 10 mejores lenguajes de programación.	27
Tabla 3 Información base de datos fuentes.	33
Tabla 4 Descripción de las tablas relacionadas al área de ventas.	34
Tabla 5 Presupuesto talento humano.	39
Tabla 6 Presupuesto recursos tecnológicos.	39
Tabla 7 Presupuesto inversión de recursos por utilizar.	40
Tabla 8 Matriz de procesos y dimensiones.	44
Tabla 9 Definición de los atributos de la dimensión categoría del producto.	48
Tabla 10 Definición de los atributos de la dimensión cliente.	48
Tabla 11 Definición de los atributos de la dimensión producto.	49
Tabla 12 Definición de los atributos de la dimensión tiempo.	49
Tabla 13 Definición de los atributos de la dimensión vendedor.	49
Tabla 14 Definición de los atributos de la tabla hecho facturas.	50
Tabla 15 Definición de los atributos de la tabla hecho productos.	50
Tabla 16 Características de la base de datos fuente	53
Tabla 17 Volumetría de los datos fuente que serán procesados.	54
Tabla 18 Descripción de las características de los equipos de cómputo disponibles para el desarrollo del proyecto.	54
Tabla 19 Descripción y especificación del recurso software necesario para el desarrollo del proyecto.	55
Tabla 20 Estructura de la tabla sub grupo en la base de datos fuente.	56
Tabla 21 Estructura de la tabla grupo en la base de datos fuente.	57
Tabla 22 Definición de los atributos de la Dimensión Categoría del producto.	57
Tabla 23 Estructura de la tabla tercero en la base de datos fuente.	58
Tabla 24 Definición de los atributos de la dimensión cliente.	59
Tabla 25 Estructura de la tabla producto en la base de datos fuente.	60
Tabla 26 Estructura de la tabla unidad de medida en la base de datos fuente.	61
Tabla 27 Estructura de la tabla sub grupo en la base de datos fuente.	62
Tabla 28 Definición de los atributos de la dimensión producto.	63
Tabla 29 Estructura de la tabla producto en la base de datos fuente.	63
Tabla 30 Definición de los atributos de la dimensión vendedor.	64

Tabla 31 Estructura de la tabla devolución en la base de datos fuente.	64
Tabla 32 Estructura de la tabla devolución en la base de datos fuente.	66
Tabla 33 Definición de los atributos de la tabla hecho facturas.	68
Tabla 34 Estructura de la tabla devolución en la base de datos fuente.	68
Tabla 35 Estructura de la tabla bodega en la base de datos fuente.	69
Tabla 36 Estructura de la tabla dfactura en la base de datos fuente.	69
Tabla 37 Estructura de la tabla devolución en la base de datos fuente.	70
Tabla 38 Estructura de la tabla producto en la base de datos fuente.	72
Tabla 39 Definición de los atributos de la tabla hecho facturas.	74
Tabla 40 Información componentes utilizados.	75
Tabla 41 Muestreo de pruebas.	97
Tabla 42 Información librerías empleadas.	100

ÍNDICE DE ILUSTRACIONES.

Ilustración 1 Representación gráfica de un ejemplo de un esquema en estrella.	15
Ilustración 2 Representación gráfica de un ejemplo de un esquema copo de nieve.	16
Ilustración 3 Representación gráfica de una bodega de datos.	17
Ilustración 4 Diagrama de flujo ETL.	18
Ilustración 5 Cuadrante Mágico de Gartner las herramientas ETL.	20
Ilustración 6 Cuadrante Mágico de Gartner plataformas analíticas de inteligencia de negocio.	22
Ilustración 7 Ciclo de vida de la metodología de Kimball.	37
Ilustración 8 Cronograma de actividades.	38
Ilustración 9 Modelo conceptual.	47
Ilustración 10 Modelo lógico.	48
Ilustración 11 Mockup para el reporte de ventas totales.	51
Ilustración 12 Mockup para el reporte de vendedores.	51
Ilustración 13 Mockup para el reporte de productos.	52
Ilustración 14 Mockup para el reporte de clientes.	52
Ilustración 15 Arquitectura propuesta para el desarrollo del proyecto.	55
Ilustración 16 Diagrama relacional de las tablas origen para la creación de la dimensión categoría de producto.	57
Ilustración 17 Diagrama relacional de las tablas origen para la creación de la dimensión producto.	62
Ilustración 18 Diagrama relacional de las tablas origen para la creación de la tabla de hecho factura.	67
Ilustración 19 Diagrama relacional de las tablas origen para la creación de la tabla de hecho factura.	74
Ilustración 20 Creación del Job Dimensión Categoría de producto.	78
Ilustración 21 Creación del Job Dimensión Cliente.	79
Ilustración 22 Creación del Job Dimensión Producto.	80
Ilustración 23 Código fuente en java para la creación de funciones para la dimensión tiempo.	81
Ilustración 24 Código fuente en java para la creación de funciones para la dimensión tiempo.	81

Ilustración 25 Creación del Job dimensión tiempo.	82
Ilustración 26 ETL dimensión Tiempo	82
Ilustración 27 Creación del Job dimensión vendedor.	83
Ilustración 28 Creación del Job tabla hecho facturas.	84
Ilustración 29 Proceso ETL tabla hecho productos.	85
Ilustración 30 Proceso ETL tabla hecho productos.	85
Ilustración 31 Creación del Job tabla hecho productos.	86
Ilustración 32 Proceso ETL tabla hecho productos.	87
Ilustración 33 Proceso ETL tabla hecho productos.	87
Ilustración 34 Código java tabla hecho productos.	88
Ilustración 35 Código Algoritmo SES.	89
Ilustración 36 Código Algoritmo SES.	89
Ilustración 37 Código Algoritmo SES.	90
Ilustración 38 Código Algoritmo SES.	90
Ilustración 39 Código Algoritmo SES.	90
Ilustración 40 Código Algoritmo SES.	91
Ilustración 41 Código algoritmo Prophet.	91
Ilustración 42 Código algoritmo Prophet.	92
Ilustración 43 Código algoritmo Prophet.	92
Ilustración 44 Código algoritmo Prophet.	92
Ilustración 45 Código algoritmo Prophet.	93
Ilustración 46 Código algoritmo Prophet.	93
Ilustración 47 Código algoritmo Prophet.	93
Ilustración 48 Código algoritmo ARIMA.	94
Ilustración 49 Código algoritmo ARIMA	94
Ilustración 50 Código algoritmo ARIMA	95
Ilustración 51 Código algoritmo ARIMA	95
Ilustración 52 Código algoritmo ARIMA	95
Ilustración 53 Código algoritmo ARIMA	96
Ilustración 54 Código algoritmo ARIMA	96
Ilustración 55 Desarrollo código fuente ARIMA.	101
Ilustración 56 Desarrollo código fuente ARIMA.	101
Ilustración 57 Desarrollo código fuente ARIMA.	101

Ilustración 58 Desarrollo código fuente ARIMA.	102
Ilustración 59 Desarrollo código fuente ARIMA.	103
Ilustración 60 Desarrollo código fuente ARIMA.	104
Ilustración 61 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Ventas (Resumen financiero).	105
Ilustración 62 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S Ventas (Estadísticas facturas).	106
Ilustración 63 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Ventas (Análisis ventas anual).	106
Ilustración 64 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Clientes (Análisis clientes).	107
Ilustración 65 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Clientes (Análisis clientes).	107
Ilustración 66 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Vendedores (Análisis vendedores).	108
Ilustración 67 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Vendedores (Análisis vendedores).	108
Ilustración 68 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).	109
Ilustración 69 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).	109
Ilustración 70 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).	110

Ilustración 71 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).	110
Ilustración 72 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis categoría y productos destacados).	111
Ilustración 73 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).	111
Ilustración 74 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S (Proyección ventas).	112
Ilustración 75 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S (Crecimiento empresarial).	112
Ilustración 76 Comportamiento de las ventas por meses.	113
Ilustración 77 Proyección de ventas.	114
Ilustración 78 Clientes destacados por compras.	114
Ilustración 79 Vendedores destacados por ventas.	115
Ilustración 80 Categorías con mayor devolución.	116
Ilustración 81 Productos con mayor devolución.	116
Ilustración 82 Categorías de productos que representan un mayor valor en relación con las ventas totales.	117
Ilustración 83 Productos con mayor utilidad.	118
Ilustración 84 Productos más vendido por cantidad.	118
Ilustración 85 Crecimiento empresarial por año.	119
Ilustración 86 Proyección de compras por producto.	120
Ilustración 87 Exposición y capacitación por parte Juan Ortiz y Anyelyn Macias a los administrativos de la empresa Materiales Maderas Santander.	121
Ilustración 88 Capacitación por parte Juan Ortiz y Anyelyn Macias al gerente general de la empresa Materiales Maderas Santander.	121

Agradecimientos

El principal agradecimiento es a Dios por guiarnos, acompañarnos, apoyarnos y darnos la sabiduría a lo largo de este camino.

Gracias a nuestros padres, por ser incondicionales con nosotros, siempre brindándonos su amor, respaldo, confianza y alentarnos a cumplir nuestras metas.

De manera especial, a nuestro director de tesis, Amilkar Sierra, por aceptar guiarnos en este proyecto, por compartir sus conocimientos con nosotros, por su colaboración y confianza en nuestro trabajo.

Al cuerpo de docentes de ingeniería de sistemas, por su compromiso, así como también su dedicación para difundir sus conocimientos y enseñanzas necesarias para formarnos como profesionales.

A nuestros compañeros de estudio, por animarnos, apoyarnos, y brindarnos ayuda en cada una de las dificultades que se presentaron durante este proceso, y por siempre animarnos a cumplir nuestros objetivos.

A la universidad Popular del Cesar, por formarnos como profesionales, por su dedicación y compromiso con la excelencia académica.

Dedicatoria

Dedico este logro, en primer lugar, a Dios, quien ha sido extraordinariamente especial en mi vida, brindándome su constante compañía y apoyo incondicional. Agradezco profundamente por mantenerme en salud, por otorgarme la sabiduría, como también las fuerzas necesarias en todo momento. Reconozco que su presencia ha sido de vital importancia para alcanzar este logro. Expreso mis más sinceros agradecimientos, pues los planes que ha tenido conmigo y su voluntad han sido perfectos.

En memoria de mi padre, Laureano Macias, quien me educó con un amor infinito, me inculcó valiosos principios y valores. Admirable en todos los sentidos, me enseñó a valorar el estudio. Fue un hombre fuerte que enfrento las dificultades con determinación. Gracias a él, comprendí el verdadero significado de la perseverancia, aprendiendo a no rendirme y enfrentar los problemas. Por todo esto y mucho más, dedico este logro al legado de mi padre.

A mi madre, Miryam Reyes, quien personifica la fuerza y la valentía. Su dedicación incansable y sacrificio han sido fundamentales para brindarme la oportunidad de estudiar. Es una de las personas que más admiro en este mundo, su compañía incondicional me recuerda que puedo alcanzar cualquier objetivo. Estoy eternamente agradecida por tenerla en mi vida; valoro mucho su amor y apoyo, gracias, por ser el pilar de la fuerza que siempre necesito.

Para mis queridos hermanos, Yoleinis y Alex, quienes han sido mi apoyo incondicional en el estudio, brindándome su constante respaldo y amor. Agradezco por siempre animarme, así como también el orgullo que sienten por mí y su grata presencia en mi camino.

A mi amado sobrino, Santiago Arias, quien siempre ha sido mi fiel admirador y se llena de orgullo por todo lo que soy y he logrado. Es una de las personas que más me motiva a salir adelante. Este logro es para ti con todo mi amor.

A mi compañero Juan Carlos Ortiz, por su amabilidad, compañerismo y apoyo constante en todo este tiempo de estudio, siempre estando a mi lado motivándome. Su ayuda y cariño han sido fundamentales para poder salir adelante en este proyecto y en muchos otros aspectos de la vida. Sin él, nada de estos sería posible; gracias por ser una parte tan importante en mi camino.

Al ingeniero Amilkar Sierra, una persona amable quien noto mi talento y me motivo a superarme siempre, valorando el lado positivo de las cosas y animándome en mis estudios, así como en mi vida laboral y personal. Agradezco cada uno de sus consejos, su confianza en mí ha sido fundamental para mi desarrollo.

Anyelyn Macias Reyes.

Dedicatoria

Dedico este logro en primer lugar a Dios, por ser mi fuente de fortaleza y guía no solo en el camino académico, sino también en todas las facetas de mi vida. Reconozco su apoyo y ayuda en momentos difíciles, que han sido fundamentales para alcanzar este logro. Agradezco su amor incondicional y su dirección constante en cada paso.

A mi padre, Juan Ortiz, le agradezco sinceramente por haber sido mi guía y mi ejemplo en la vida. Su amor incondicional y su dedicación han sido importantes en mi camino. Su apoyo constante ha sido fundamental en mi formación. En cada momento difícil, han estado ahí para ayudarme, motivarme y brindarme su sabio consejo. Su presencia y su influencia han dejado una marca en mi vida. Aprecio enormemente todo lo que ha hecho por mí.

A mi madre, Erika Reales, le agradezco de todo corazón por ser mi roca, mi mayor fuente de amor y apoyo. Su ternura, su comprensión y su inquebrantable presencia han sido un regalo invaluable en mi vida. En cada dificultad, ha estado ahí para escucharme, consolarme y alentarme a seguir adelante. Su generosidad y su sacrificio me han enseñado mucho.

A mis queridas hermanas, Erika, Andrea y Sahuris, ustedes son mi apoyo incondicional y mi mayor orgullo. Siempre han sido mis cómplices y mis compañeras de vida. Su admiración y respaldo significan mucho para mí. Gracias por estar siempre a mi lado, por creer en mí y por ser un motor constante de motivación. Este logro también es de ustedes.

A Miriam Reyes, quien ha sido como una madre para mí a lo largo de mi trayectoria académica. Su bondad y generosidad no solo me brindaron consuelo, sino que también me dieron la fuerza y la determinación para seguir adelante. Agradezco profundamente su amor maternal y su guía, que han sido fundamentales en mi camino hacia el éxito académico.

A mi compañera, Anyelyn Macias, agradezco su constante apoyo y ánimo durante los momentos de desaliento que fueron un verdadero regalo para mí. A lo largo de este viaje, su motivación inquebrantable y su disposición para emprender este camino juntos han sido inspiradores. Agradezco su comprensión, su cariño y su aceptación incondicional.

A mi director de tesis, Amilkar Sierra, quiero expresar mi más sincero agradecimiento por su orientación experta y paciencia a lo largo de este proceso.

Juan Carlos Ortiz Rea

Introducción

Los sistemas informáticos desempeñan un papel crucial en la gestión de procesos empresariales en la actualidad. Es poco probable encontrar una empresa que no cuente con un sistema informático donde se almacenen los datos de sus operaciones. Sin embargo, en la mayoría de las empresas, estos datos simplemente se guardan en la base de datos sin ser aprovechados, ni transformados en información relevante o valiosa para la toma de decisiones.

La inteligencia de negocio es un proceso en el cual se recopilan y analizan datos históricos, así como también datos actuales, para luego representarlos de manera gráfica y respaldar la toma de decisiones estratégicas. En la actualidad, implementar una solución de inteligencia de negocio es una necesidad imperativa para las empresas. Los beneficios derivados de esto ayudan a éstas a posicionarse en el mercado mejorando el rendimiento de sus procesos, lo que implica un crecimiento empresarial generalizado.

De acuerdo a lo establecido previamente, la empresa Materiales Maderas Santander S.A.S. busca expandirse en el mercado mediante estrategias que respalden las decisiones con análisis de la información. Hasta ahora, los datos de la empresa han sido guardados sin ningún tratamiento, aunque se ha intentado realizar análisis mediante Excel, este proceso resulta tedioso igual que poco eficiente. Por lo tanto, el objetivo del proyecto es implementar una inteligencia de negocio respaldada por una bodega de datos para el área de ventas y compras que proporcionando información valiosa para la toma de decisiones en la empresa.

El presente trabajo consta de tres etapas, cada una de vital importancia para alcanzar el objetivo propuesto.

En la primera etapa, denominada Descripción General, se define el título del proyecto, el tiempo de ejecución y se proporciona información detallada sobre la línea y sublínea de investigación a la que pertenece el proyecto.

En la segunda etapa, denominada Descripción Situacional, se identifica y se formula el problema que enfrenta la empresa, se establece la justificación del problema y se definen los

objetivos. Además, se presenta el marco referencial, donde se definen conceptos importantes y se exponen los antecedentes del trabajo. Luego, se explica a detalle el marco metodológico empleado en el proyecto.

En la tercera etapa, Desarrollo Científico Tecnológico, se describe la planificación del proyecto y el desarrollo de las fases de la metodología propuesta. Se muestra el paso a paso del diseño, desarrollo e implementación del subsistema ETL, así como del algoritmo de autoaprendizaje para la proyección de compras. Se presenta el uso del módulo de visualización, que incluye informes, reportes y gráficas, permitiendo una representación más práctica de la información.

Finalmente, se exponen el análisis de los resultados, las conclusiones y las recomendaciones para la empresa.

SECCIÓN I. DESCRIPCIÓN GENERAL

1.1 Título

Implementación de inteligencia de negocio para el área de ventas y compras en la empresa Materiales Maderas Santander SAS.

1.2 Dirección de ejecución

Calle 31 # 4h/76 Villa del Rosario, Valledupar

1.3 Lapso de ejecución

Ocho (8) meses

1.4 Organismo o institución responsable

Materiales Maderas Santander SAS.

1.5 Información de contacto

Tabla 1 Información de contacto.

Nombres	Apellido	Identificación	Celular	Correo
Juan Carlos	Ortiz reales	1.064.121.751	320 8201482	juancortiz@unicesar.edu.co
Anyelyn	Macias reyes	1.003.244.070	301 3222303	amaciasr@unicesar.edu.co

Autor: Macias Anyelyn, Ortiz Juan.

1.6 Línea, Sublínea y grupo de investigación del proyecto.

1.6.1 Línea de investigación.

Transformación digital

Puntualmente, se quiere encaminar a la empresa Materiales Maderas Santander SAS al uso de las tecnologías emergentes, tratando de aprovechar los beneficios de la transformación digital en pro de esta misma y finalmente lograr que la empresa Materiales Maderas Santander SAS utilice estas nuevas tecnologías.

1.6.2 Sublínea de investigación.

Big Data y Analytics

Concretamente, se quiere desarrollar e implementar una herramienta de ayuda que permita transformar los datos en información que beneficie a la empresa Materiales Maderas Santander SAS ayudando con esto a la toma de decisiones. Para lograr esto se pretende

crear tableros de control, graficas dinámicas, reportes, entre otros que les facilite esta tarea a las personas interesadas en la empresa Materiales Maderas Santander SAS.

1.6.3 Área de investigación.

Business Intelligence

El proyecto se centra en la creación de una bodega de datos para el área de ventas y compras, por lo que es de suma importancia que los datos sean tratados y resulten confiables para la generación de informes, ayudando a la toma de decisiones en general y más específicamente en las ventas de la empresa y en la proyección de compras de esta misma.

1.6.4 Grupo de investigación.

GISICO

El proyecto se ajusta a este grupo de investigación ya que apoya el desarrollo de conocimientos en cuanto a la creación de herramientas que se encargan de transformar y analizar los datos convirtiéndolos en información relevante que ayudará a la empresa en la toma de decisiones de una manera más accesible y fácil.

SECCIÓN II. DESCRIPCIÓN SITUACIONAL

2.1 Identificación del problema

Frente el panorama actual de los negocios, es de suma importancia que las organizaciones presten atención a las preferencias en innovación y busquen adoptar los procedimientos de las diferentes áreas de acuerdo con el nacimiento de soluciones que puedan mejorar sus resultados. Para acompañar los nuevos movimientos del mercado, las empresas implementan soluciones tecnológicas capaces de concentrar la información, apresurar los procesos y garantizar resultados mucho más efectivos y alineadas con las expectativas de los clientes. [1]

Una empresa que maneja grandes volúmenes de información de cualquier naturaleza, que no se gestionan o procesan, esta empresa estaría perdiendo competitividad frente a otras que sí lo hacen. Por lo tanto, es crucial que las empresas dispongan de sistemas de información que posibiliten supervisar, visualizar, organizar, gestionar y conectar eficientemente el flujo de datos en su interior.[2]

A partir de lo expuesto, se puede observar que, sin importar las dimensiones o el alcance de las actividades, que resulta esencial implementar sistemas de información en las organizaciones. La revolución digital ha proporcionado mejores sistemas y herramientas que contribuyen a mejorar la eficiencia de los negocios, optimizar sus procesos y obtener una visión más completa de las operaciones.[2]

En conclusión, con lo anteriormente mencionado sobre la importancia de los sistemas de información en las empresas y los datos como recurso para la toma de decisiones, cabe mencionar que los datos son el recurso más relevante en el cual la empresa se puede apoyar, sin embargo, actualmente algunas empresas no les dan la atención que verdaderamente necesita. Ignorando así los beneficios que estos datos podrían traerles si se les da la correcta administración.

La empresa Materiales Maderas Santander SAS se dedica a la comercialización de materiales de construcción, artículos de ferretería, cortes de láminas, laminas RH con

variedad de formatos, entre otros productos. Se encuentra ubicada en la Calle 19 B # 13 – 33 en la ciudad de Valledupar, Cesar, cuenta con más de 5 años en el mercado.

Por lo tanto, según la entrevista realizada a la gerente administrativa de la empresa Materiales Maderas Santander SAS, se evidenció que el manejo de la empresa son las ventas de sus productos y que además carecen de una proyección de compras hacia los proveedores. Los reportes son de suma importancia para saber el estado de la empresa y para la ayuda a la toma de decisiones, siendo estas mismas las que ayuda a cumplir los objetivos de las empresas. La empresa cuenta con un sistema que le genera algunos informes, estos no muestran información estratégica para la toma de decisiones como estrategias de mercadeo, promociones de ciertos productos, proyección de compras, empleados que más vende, productos menos vendidos, crecimiento de la empresa entre otros.

Pero, por otro lado, la empresa ha logrado realizar un tratamiento y análisis de los datos para la toma de decisiones, esto lo hacen de manera manual por medio de la herramienta Excel, lo cual tiene un tiempo estimado de dos horas de trabajo porque son más de 1500 productos y se tienen más de 5 archivos de Excel para este proceso. Teniendo en cuenta esto, se puede evidenciar lo tedioso que se vuelve el proceso debido a que implica extraer los datos de la base de datos e importarlos a Excel, tratarlos y posteriormente crear aproximadamente por archivo más de 10 gráficas dinámicas que muestre la información que requiere la empresa.

Si la empresa Materiales Maderas Santander SAS no implementa un sistema de análisis de los datos y proyección de compras, se arriesga a perder crecimiento en el mercado y estaría en grandes desventajas frente a las empresas que, si poseen este sistema, desaprovechando el auge de esta era interconectada, y se perderían clientes potenciales. Teniendo en cuenta lo anterior, se puede decir que la empresa no podría subsistir sin un sistema que le permita tomar decisiones de manera más confiable y rápida.

Para el desarrollo de este proyecto, se propone la creación de un módulo funcional de inteligencia de negocio con un enfoque en el análisis de ventas y compras de la empresa. Este módulo contendrá una bodega de datos integral que servirá como base para

proporcionar a los directivos informes y reportes a través de gráficos dinámicos. Estas herramientas visuales permitirán un estudio más intuitivo y sencillo de la situación actual, pasada y futura de la empresa.

El propósito principal de esta iniciativa es facilitar el proceso de toma de decisiones para los líderes de la organización. Al aprovechar los datos recopilados en las áreas de ventas y compras, estos informes ayudarán a la empresa a alcanzar sus objetivos, fomentar su crecimiento, fortalecer su posición en el mercado y elevar su competitividad.

2.1.1 Formulación del problema

¿Cómo desarrollar una inteligencia de negocio para el área de ventas y compras en la empresa Materiales Maderas Santander SAS que les ayude en la toma de decisiones de manera estratégica?

2.2 Justificación del proyecto

Teóricamente, el propósito fundamental de este proyecto radica en demostrar la capacidad de la empresa Materiales Maderas Santander SAS para potenciar su toma de decisiones mediante la implementación de sistemas informáticos de inteligencia de negocio. Para la empresa Materiales Maderas Santander SAS, es fundamental contar con un acceso fluido a la información en el momento oportuno. Estos datos engloban una variedad de aspectos, tales como detalles sobre clientes, productos, vendedores, proveedores, precios, devoluciones y otros elementos clave. Mediante un análisis exhaustivo de estos elementos, la empresa estará en una posición sólida para tomar decisiones informadas, estratégicas, que influyan en su propio beneficio.

Por otro lado, desde una perspectiva práctica, la implementación de un módulo operativo de inteligencia de negocio destinado al ámbito de ventas y adquisiciones en la compañía constituirá una solución efectiva para abordar los desafíos identificados. Entre estos desafíos, destaca la problemática del tiempo perdido en el proceso de análisis de datos, la cual se verá mitigada significativamente. Además, este módulo no solo agilizará la toma de decisiones, sino que también optimizará los procesos internos de la empresa.

Con la puesta en marcha de este módulo funcional de inteligencia de negocio, la organización experimentará mejoras notables en sus procedimientos, evitando así la pérdida de tiempo que conlleva el procesamiento manual de datos para su posterior utilización. Una vez que este módulo esté desarrollado y en funcionamiento, se automatizarán tanto los procesos de análisis

de datos como las proyecciones de compras. La información resultante será presentada a través de paneles de control, informes dinámicos, gráficos interactivos, lo que simplificará la visualización de datos y los hará más accesibles e intuitivos para los stakeholders de la empresa.

Asimismo, la justificación metodológica del proyecto radica en la implementación de un módulo funcional de inteligencia empresarial respaldado por una bodega de datos multidimensional. Esta estructura permitirá analizar los movimientos de ventas, abarcando aspectos como productos, vendedores y clientes, entre otros. Además, se busca la incorporación de una proyección de compras mediante un algoritmo de aprendizaje automático. En este enfoque, los datos serán analizados en intervalos temporales específicos para presentar a la empresa una perspectiva de sus futuras compras.

Todo esto con el propósito de agilizar el análisis de datos para respaldar la toma de decisiones empresariales, fomentando un crecimiento más sólido en el mercado y fortaleciendo la fidelización de clientes, además de optimizar la proyección de compras y reducir pérdidas.

2.3 Objetivos del proyecto

2.3.1 Objetivo general

Implementar una inteligencia de negocio para el área de ventas y compras en la empresa Materiales Maderas Santander SAS.

2.3.2 Objetivos específicos

- Analizar los requerimientos del negocio referentes a los procesos que se realizan en el área de ventas y compras de la empresa.
- Diseñar la arquitectura lógica y física para el Data mart.

- Determinar el algoritmo de machine learning adecuado para la proyección de las compras en la empresa.
- Diseñar el módulo de visualización para representar la información a través de dashboard que permita el análisis de los datos.

2.4 Marco referencial

2.4.1 Marco conceptual

En este apartado se explicará los conceptos fundamentales de las variables de estudio del proyecto y la relación que existe entre estas, desde lo más simple a lo más complejo.

2.4.1.1 Inteligencia de negocio (BI)

La inteligencia de negocio se refiere a un conjunto de enfoques, recursos y técnicas empleadas con el fin de recopilar, examinar y convertir datos en información valiosa y pertinente para respaldar las decisiones estratégicas de una empresa.[3]

La definición mencionada hace hincapié en que la inteligencia de negocio no solo se trata de recopilar datos, sino de convertirlos en información de calidad y utilizar esa información para obtener conocimiento que pueda ser utilizado para tomar decisiones informadas y estratégicas. La meta final de la inteligencia de negocio es ayudar a las empresas a mejorar su competitividad mediante la toma de decisiones más acertadas y fundamentadas en datos.

Los sistemas de BI buscan dar respuestas a los siguientes interrogantes:

- ¿Qué pasó?
- ¿Qué pasa ahora?
- ¿Por qué pasó?
- ¿Qué pasará?

2.4.1.2 ¿Cuándo es necesaria la inteligencia de negocio?

Según los autores Conesa y Curto [4] la inteligencia de negocio es necesario cuando:

- La toma de decisiones se realiza de forma intuitiva en la organización.
- Identificación de problemas de calidad de información.
- Uso de Excel como repositorios de información corporativos o de usuario.
- Lo que se conoce como Excel caos.
- Necesidad de cruzar información de forma ágil entre departamentos.
- Evitar silos de información.
- Las campañas de marketing no son efectivas por la información base usada.

- Existe demasiada información en la organización para ser analizada de la forma habitual. Se ha alcanzado la masa crítica de datos.
- Es necesario automatizar los procesos de extracción y distribución de información.

2.4.1.3 Beneficios de la inteligencia de negocio

Según los autores Conesa y Curto [4] los beneficios de la inteligencia de negocio son los siguientes.

- Establecer un ciclo virtuoso de la información (donde los datos se convierten en información que genera conocimiento, lo cual permite tomar decisiones más acertadas, que a su vez se traducen en mejores resultados y generan nuevos datos).
- Facilitar una visión integral, consistente, histórica, duradera y de alta calidad de toda la información disponible.
- Desarrollar, administrar y mantener métricas, indicadores clave de rendimiento (KPI, Key Performance Indicators) e indicadores clave de metas (KGI, Key Goal Indicators) que sean fundamentales para la empresa.
- Proporcionar información actualizada tanto a nivel global como detallado.
- Reducir la brecha de entendimiento y enfoque entre el departamento de Tecnología de la Información (TI) y la organización en general.
- Mejorar la comprensión y documentación de los sistemas de información en el contexto de la organización.

2.4.1.4 Modelo de madurez de inteligencia empresarial

Cuando se proporciona una solución de inteligencia de negocio, es crucial identificar el nivel de madurez en el que se encuentra la organización donde se implementará la solución. Esto permitirá clasificar la empresa en función de su grado de madurez en la implementación de inteligencia de negocio y, por lo tanto, ofrecer una solución adecuada que satisfaga sus necesidades específicas.

Según los autores Conesa y Curto [4] las fases del modelo de madurez de inteligencia empresarial son las siguientes:

- **Etapa 1:** Ausencia de Business Intelligence. Los datos se encuentran dispersos en sistemas de procesamiento de transacciones en línea (OLTP), en diferentes soportes o simplemente en el conocimiento interno de la organización. Las decisiones se basan en la intuición y la experiencia, pero no en datos consistentes. No se ha reconocido la

utilización de datos corporativos en la toma de decisiones ni se ha identificado la necesidad de herramientas adecuadas.

- **Etapa 2:** Acceso a datos sin Business Intelligence. Aunque los datos son accesibles, no se realiza un procesamiento formal de los mismos para la toma de decisiones. Algunos usuarios tienen acceso a información de calidad y son capaces de justificar decisiones con dicha información. En muchos casos, este proceso se lleva a cabo mediante el uso de Excel u otras herramientas de reporting. Existe una intuición de que deben existir soluciones para mejorar este proceso, pero no se tiene conocimiento de la existencia del Business Intelligence.
- **Etapa 3:** Surgimiento de procesos formales de toma de decisiones basados en datos. Se establece un equipo encargado de gestionar los datos y generar informes que respalden la toma de decisiones fundamentadas. Los datos se extraen directamente de los sistemas transaccionales sin realizar procesos de limpieza o modelado, y no se cuenta con un data warehouse.
- **Etapa 4:** Implementación de un data warehouse. El impacto negativo en los sistemas OLTP lleva a reconocer la necesidad de un repositorio de datos para la organización. El data warehouse se percibe como una solución deseable, pero el reporting aún es personalizado.
- **Etapa 5:** Crecimiento de la data warehouse y formalización del reporting. El data warehouse funciona y se busca que todos se beneficien de él, por lo que se formaliza el proceso de generación de informes a nivel corporativo. Se menciona el concepto de OLAP, pero solo algunos comprenden plenamente sus beneficios.
- **Etapa 6:** Implementación de OLAP. Después de cierto tiempo, ni el reporting ni la forma de acceder a la data warehouse son satisfactorios para responder a preguntas sofisticadas. Se implementa OLAP para atender las necesidades de estos perfiles. Las decisiones comienzan a tener un impacto significativo en los procesos de negocio de toda la organización.

- **Etapa 7:** Formalización del Business Intelligence. Surge la necesidad de implementar otros procesos de inteligencia empresarial, como Data Mining, a medida que el Business Intelligence se consolida y se reconoce como una parte integral de la organización.

2.4.1.5 KPI

El acrónimo KPI, que representa Indicador Clave de Desempeño o Medidor de Desempeño, se refiere a un conjunto de medidas utilizadas para resumir la información sobre el rendimiento y la productividad de las acciones llevadas a cabo en una empresa. Estas medidas son útiles para tomar decisiones y determinar qué acciones han sido más eficaces para alcanzar los objetivos establecidos en un proceso o proyecto particular.[5]

2.4.1.6 Data warehouse

Un data warehouse es un sistema de gestión de datos especialmente diseñado para respaldar las actividades de inteligencia empresarial (BI), en particular, el análisis. Su principal objetivo es permitir consultas y análisis de grandes volúmenes de datos históricos provenientes de diversas fuentes, como registros de aplicaciones o sistemas de transacciones. Los data warehouses se destacan por su capacidad para centralizar y fusionar grandes cantidades de datos de múltiples fuentes. Esto brinda a las organizaciones información empresarial valiosa y mejora la toma de decisiones. Con el tiempo, se genera un valioso repositorio histórico para expertos en datos y analistas de negocios, convirtiéndose en una herramienta fundamental para la gestión empresarial. Debido a sus capacidades analíticas, un data warehouse se considera la "fuente única de datos" de una organización. Proporciona una vista completa y centralizada de todos los datos empresariales relevantes, lo que facilita la toma de decisiones basadas en datos y el análisis de tendencias y patrones comerciales.[6]

Características de un data warehouses

- Integración de datos: El data warehouse recopila y almacena datos de diferentes fuentes, incluyendo sistemas operativos, aplicaciones, bases de datos y otros almacenes de datos.
- Orientado a temas: El data warehouse está diseñado para almacenar datos específicos de un tema en particular, como ventas, finanzas o recursos humanos.

- No volátil: Los datos almacenados en el data warehouse son de sólo lectura y no pueden ser modificados. Esto garantiza que los datos sean coherentes y precisos.
- Histórico: El data warehouse almacena datos históricos, lo que permite a los usuarios realizar análisis de tendencias y patrones a lo largo del tiempo.
- Consultas complejas: El data warehouse permite realizar consultas complejas y análisis de datos, lo que permite a los usuarios obtener información valiosa de grandes conjuntos de datos.
- Estructurado para la toma de decisiones: El data warehouse está diseñado para apoyar la toma de decisiones empresariales mediante la generación de informes y análisis de datos que proporcionan información valiosa a los usuarios de negocio.

La información en los data warehouses se representan de manera desnormalizada, esto con el fin de optimizar las consultas.

2.4.1.7 OLAP

La tecnología de software denominada Procesamiento Analítico en Línea (OLAP) posibilita el análisis de datos empresariales desde múltiples enfoques. Las organizaciones recopilan y almacenan información de diversas fuentes, como sitios web, aplicaciones, medidores inteligentes y sistemas internos. OLAP integra y organiza estos datos en categorías con el fin de proporcionar información relevante para la planificación estratégica de la empresa.[7]

2.4.1.8 Modelo multidimensional

Modelos de datos como conjuntos de medidas descritas por dimensiones. [8]

- **Hechos:** Los hechos son colecciones de datos relacionados que incluyen medidas y un contexto. El contexto es determinado por las dimensiones y cada hecho específico se encuentra vinculado a un miembro de cada dimensión. [8]

Existen varios tipos de tablas de hechos, a continuación, se mencionarán tres tipos.

1. **Transaccional:** es una tabla que contiene información detallada de transacciones o eventos individuales en un momento específico. En una tabla de hechos transaccional, cada registro representa una transacción individual, por lo que la tabla puede tener un gran número de filas. Esta tabla se utiliza como la base para las consultas y los informes analíticos, ya que permite un análisis detallado de los datos en una base de datos. Además, se puede utilizar para construir otras tablas de hechos, como las tablas de hechos agregados o resumidos.

2. **Vista periódica:** es una tabla en una base de datos de Business Intelligence que almacena datos agregados y precalculados para un lapso de tiempo específico, como un mes o un trimestre. Esta tabla de hechos contiene medidas agregadas para cada combinación de dimensiones y permite que las consultas se ejecuten más rápido en comparación con las tablas de hechos transaccionales detalladas.
3. **Acomulada:** es una tabla que almacena medidas que se acumulan a lo largo del tiempo en lugar de representar valores instantáneos. Esta tabla se utiliza para realizar análisis de tendencias y comparar el rendimiento de la organización a lo largo del tiempo. En una tabla de hechos acumulados, cada registro representa un período de tiempo específico, como un día, un mes o un trimestre, y las medidas acumuladas se calculan y almacenan en la tabla para cada período.

Tipos de dimensiones:

SCD Tipo 1: En este tipo de SCD, se sobrescribe el valor antiguo con el nuevo valor cuando se produce un cambio en la dimensión. Esto significa que no se mantiene un registro histórico de los cambios.

SCD Tipo 2: En este tipo de SCD, se agrega una nueva fila a la tabla de la dimensión cada vez que se produce un cambio, manteniendo el registro histórico de los cambios. Esto permite realizar un análisis histórico y comparar los cambios en la dimensión a lo largo del tiempo.

SCD Tipo 3: En este tipo de SCD, se agrega una columna adicional a la tabla de la dimensión para almacenar el valor anterior y el nuevo valor. Esto significa que se mantiene un registro histórico de los cambios, pero solo se pueden rastrear los cambios más recientes. Este enfoque es útil cuando los cambios en la dimensión son importantes, pero la organización solo necesita rastrear los cambios más recientes.

SCD Tipo 4: es un tipo de dimensión lentamente cambiante que combina elementos del SCD Tipo 1 y del SCD Tipo 2. En este enfoque, se mantiene un registro histórico de los cambios en la dimensión, pero en lugar de agregar nuevas filas a la tabla de la dimensión, se crea una tabla adicional para almacenar los cambios.

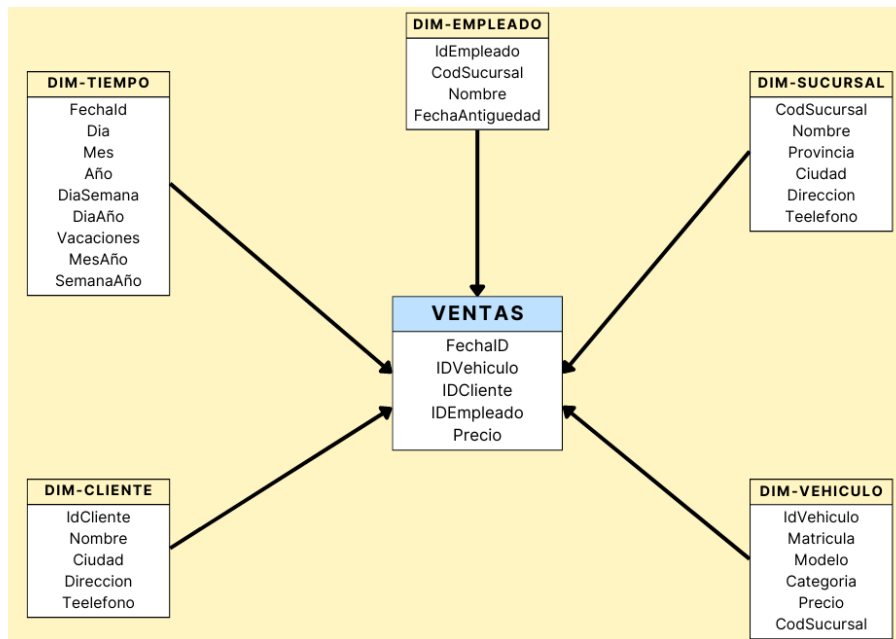
SCD Tipo 6: es otro tipo de dimensión lentamente cambiante que combina elementos del SCD Tipo 1 y del SCD Tipo 2, pero también incorpora la noción de versiones. En este enfoque, se mantiene un registro histórico de los cambios en la dimensión, al mismo tiempo que se crea una versión "actual" de la dimensión. La versión actual se relaciona con la tabla

principal de la dimensión, mientras que las versiones históricas se mantienen en una tabla adicional.

- **Dimensiones:** Las dimensiones se refieren a las perspectivas o entidades que una organización desea mantener para organizar sus datos, como el tiempo, la ubicación, los clientes, los proveedores, entre otros. [8]
- **Granularidad:** El nivel de detalle se refiere al grado de especificidad de los datos almacenados en una tabla. En otras palabras, se refiere al nivel de desagregación de los datos registrados. Un nivel de detalle alto implica que se están capturando datos detallados y específicos, mientras que un nivel de detalle bajo implica que los datos se están agregando y resumiendo a un nivel más general.
- **Medida:** Los indicadores son atributos numéricos asociados a los hechos que representan lo que se está midiendo en una base de datos. Por ejemplo, los indicadores pueden incluir el volumen de ventas, los costos de un producto, el número de transacciones realizadas o el porcentaje de beneficios. Los indicadores son importantes porque proporcionan información cuantitativa que se utiliza para analizar y tomar decisiones en una base de datos. [8]
- **Esquema en estrella**

Una tabla de hechos y una tabla adicional por cada dimensión. [8]

Ilustración 1 Representación gráfica de un ejemplo de un esquema en estrella.

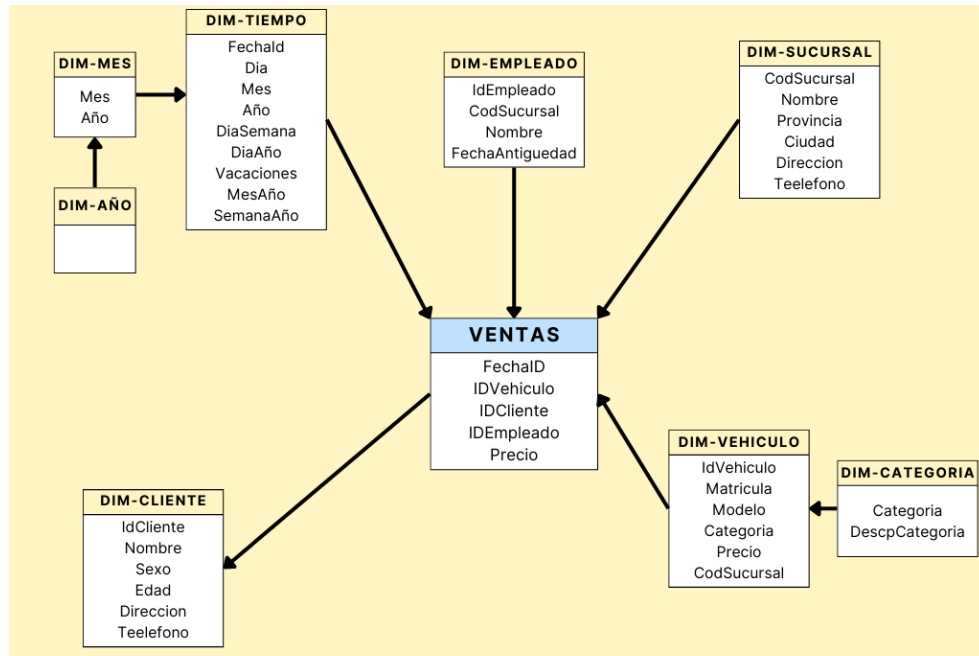


Autor: Macias Anyelyn, Ortiz Juan.

- **Esquema copo de nieve.**

Refleja la organización jerárquica de las dimensiones. [8]

Ilustración 2 Representación gráfica de un ejemplo de un esquema copo de nieve.

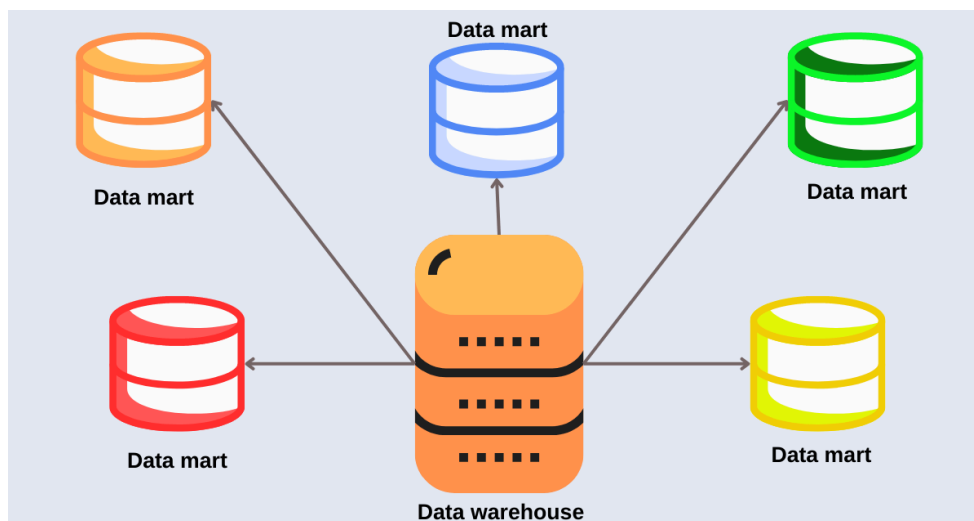


Autor: Macias Anyelyn, Ortiz Juan.

2.4.1.9 Data mart

Un data mart es un sistema de gestión de datos que almacena información específicamente dirigida a una unidad de negocio particular dentro de una organización. A diferencia del sistema de gestión de datos más amplio de la empresa, un data mart contiene solo una fracción de los datos totales. Su objetivo principal es proporcionar información analítica de manera eficiente y rápida al departamento o unidad de negocio correspondiente. Los data marts están diseñados para ofrecer datos resumidos y relevantes a las partes interesadas clave, facilitando la toma de decisiones informadas de manera ágil. Esto ayuda a las empresas a comprender mejor sus operaciones comerciales y a tomar decisiones basadas en hechos concretos, en lugar de suposiciones. En resumen, un data mart es una herramienta esencial para la gestión de datos empresariales y el análisis de información, permitiendo a las organizaciones tomar decisiones más precisas y fundamentadas.[9]

Ilustración 3 Representación gráfica de una bodega de datos.



Autor: Macias Anyelyn, Ortiz Juan

2.4.1.10 ETL

El proceso ETL (Extraer, Transformar, Cargar) se refiere a un conjunto de técnicas, herramientas y tecnologías utilizadas para obtener datos de diversas fuentes, modificarlos de manera coherente y valiosa, y cargarlos en otros sistemas para que sean accesibles por los diferentes niveles de la organización.

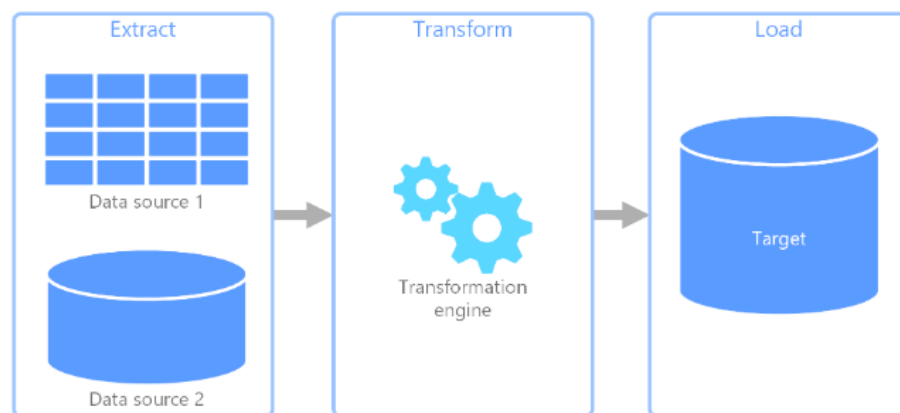
La etapa de extracción implica recopilar datos de múltiples fuentes, como bases de datos, archivos y aplicaciones. La etapa de transformación se encarga de convertir los datos en un formato uniforme y estándar, aplicando reglas de limpieza y validación para asegurar la precisión e integridad de los datos.

Por último, la etapa de carga implica transferir los datos transformados a otros sistemas, como un almacén de datos o un data mart, para que los usuarios finales puedan acceder y utilizar la información. El proceso ETL es fundamental para garantizar la disponibilidad y utilidad de los datos en la toma de decisiones empresariales.[10]

- **Fase 1 Extracción:** La fase de extracción en un proceso ETL tiene como objetivo recolectar datos de múltiples fuentes, tanto internas como externas. Estas fuentes pueden incluir sistemas internos de la organización, como un CRM o ERP, o fuentes externas como proveedores, clientes y datos públicos de la web. Una vez extraídos, los datos se almacenan en su formato original en almacenes de datos, que suelen estar alojados en sistemas cloud, para su posterior procesamiento. [10]

- **Fase 2 Transformación:** La fase de transformación en un proceso ETL es fundamental y consiste en procesar los datos para que sean coherentes con el modelo de negocio de la organización. Los datos extraídos en la fase anterior pueden ser estructurados o no estructurados, y deben ser transformados para poder obtener información útil y coherente con la empresa. Esta fase puede incluir la eliminación de duplicados, la corrección de errores, la conversión de formatos y la agregación de datos, entre otros procesos de limpieza y organización. Una vez transformados, los datos se preparan para ser cargados en los sistemas de destino. [10]
- **Fase 3 Carga:** La fase de carga en un proceso ETL es la última etapa donde los datos transformados se cargan en un sistema de destino. Este sistema de destino puede ser un Data Warehouse, que es una base de datos diseñada para el análisis de grandes volúmenes de datos. Los datos cargados en el Data Warehouse pueden ser utilizados por distintas herramientas de análisis, como herramientas de inteligencia empresarial y análisis de datos. A través de estas herramientas, las empresas pueden realizar análisis descriptivos, diagnósticos, predictivos y prescriptivos para obtener información valiosa y tomar decisiones informadas basadas en datos. [10]

Ilustración 4 Diagrama de flujo ETL.



Autor: Microsoft

Según el Cuadrante Mágico de Gartner las herramientas ETL más utilizadas son:

- **Talend:** Talend Open Studio es una herramienta de software de código abierto bajo licencia GPL que permite visualizar y modelar transformaciones de datos, generando código Java. Sus aplicaciones incluyen ETL, sincronización y replicación de bases de datos, migración y transformación de datos. Ofrece un entorno visual basado en Eclipse RCP, con más de 400 componentes predefinidos y la capacidad de crear nuevos. Su diseño gráfico permite mapear y transformar datos, además de funciones especializadas como filtrado y multiplexado de datos, así como soporte para RDBMS, formatos de archivo y directorios LDAP.
- **IBM:** Datastage es una herramienta ETL que facilita la extracción, transformación y carga de datos en el destino deseado, incluyendo entornos de Big Data. Proporciona datos de calidad para el análisis empresarial y cuenta con una interfaz gráfica que simplifica la gestión de procesos de integración de datos. Datastage es compatible con diversas fuentes de datos, como archivos secuenciales, indexados, bases de datos relacionales, mainframes, fuentes externas y aplicaciones empresariales.
- **Oracle:** ODI (Oracle Data Integrator) es una herramienta de integración de datos que permite extraer, transformar y cargar datos desde diversas fuentes hacia un destino deseado. ODI es una solución ETL eficiente y escalable, con características como integración en tiempo real, compatibilidad con múltiples fuentes y destinos de datos, soporte para Big Data, transformaciones complejas de datos y automatización de procesos de integración.
- **Informática:** PowerCenter es una plataforma de integración de datos de Informatica Corporation que realiza extracción, transformación y carga (ETL) de datos desde diversas fuentes hacia un destino deseado. PowerCenter se utiliza para integrar y consolidar datos de múltiples sistemas y aplicaciones en una única fuente, permitiendo la toma de decisiones en tiempo real. Sus características incluyen integración con diversas fuentes y destinos de datos, soporte para diferentes tipos de datos, transformaciones complejas, rendimiento escalable y alta disponibilidad.

Para el desarrollo de este proyecto de investigación como herramienta ETL se seleccionará talend Open Studio debido a sus características frente a otras herramientas como su versión gratis, la facilidad de uso, su amplia compatibilidad y variedad de componentes, teniendo en cuenta que esta herramienta cuenta con muy buen soporte y comunidad dispuestos ayudar cuando sea necesario.

Ilustración 5 Cuadrante Mágico de Gartner las herramientas ETL.

Figure 1: Magic Quadrant for Data Integration Tools



Source: Gartner (August 2022)

Autor: Gartner.

2.4.1.11 Dasboarh

Un panel de control es una herramienta que permite administrar y supervisar datos de manera visual, mostrando indicadores clave de rendimiento (KPI), métricas y datos relevantes para monitorear el estado de una empresa, departamento, campaña o proceso específico. Es una especie de resumen que recopila información de diversas fuentes en un solo lugar y la presenta de manera comprensible, resaltando lo más importante. Un panel de control debe adaptarse a las necesidades específicas del departamento, campaña o proceso en cuestión, incluyendo solo los KPI relevantes. Para lograrlo, es necesario identificar las principales preguntas que queremos responder. Además, la información se presenta visualmente, a través de gráficos y señales rápidas como colores, flechas hacia arriba o abajo y cifras destacadas.

Es importante que un panel de control sea práctico, es decir, que guíe las acciones de nuestro equipo y proporcione la información necesaria para determinar los siguientes pasos

a seguir con el fin de mejorar los resultados. Por último, la información debe actualizarse en tiempo real, ya que las estrategias de marketing digital evolucionan rápidamente y aprovechar el momento clave es esencial.[13]

Según el Cuadrante Mágico de Gartner las plataformas analíticas de inteligencia de negocio más utilizadas son:

- **Microsoft:** Microsoft Power BI es una herramienta de Business Intelligence y visualización de datos diseñada para empresas y usuarios individuales. Ofrece a los usuarios herramientas para agregar, analizar, visualizar y compartir datos. La interfaz de usuario de esta herramienta es bastante intuitiva, especialmente para aquellos que ya están familiarizados con Excel. Además, su integración con otros productos de Microsoft la convierte en una herramienta de autoservicio muy útil que requiere poco entrenamiento inicial.

Hay dos versiones de Microsoft Power BI disponibles: una versión gratuita destinada a pequeñas y medianas empresas y una versión de pago, Power BI Plus, que incluye más funcionalidades y está disponible mediante una suscripción mensual.

En resumen, Power BI es una herramienta que reúne, procesa y visualiza los datos de la empresa de forma clara e inteligible. Utilizando gráficos atractivos y fáciles de entender, permite a los usuarios generar y compartir instantáneas claras y útiles de lo que está sucediendo en la empresa. Además, se puede conectar a una gran variedad de fuentes de datos, desde hojas de cálculo básicas hasta bases de datos y aplicaciones en la nube.[15]

- **Salesforce tableau:** es una plataforma líder en el mercado de Business Intelligence que proporciona una solución completa para que los encargados de tomar decisiones tengan acceso en tiempo real a toda la información que necesitan. Una de las principales características de Tableau es su capacidad de visualización interactiva de datos. Los usuarios pueden interactuar con los datos y realizar comparaciones, filtrados y conexiones entre variables para obtener información valiosa de manera rápida y sencilla. Además, los informes y dashboards que se pueden crear son altamente visuales, lo que facilita la comprensión de los datos.

Tableau convierte a las empresas en organizaciones "data-driven" al hacer que los datos se conviertan en información útil. Esta herramienta ayuda a las

empresas a tomar decisiones informadas al proporcionar información clara y concisa. Además, Tableau se integra con una amplia variedad de fuentes de datos, lo que permite a los usuarios conectarse a múltiples fuentes de datos y combinarlos para obtener información más completa. [16]

- **Qlik:** es una herramienta de visualización y descubrimiento de datos de última generación que facilita el análisis de los datos y la toma de decisiones empresariales. Se trata de una solución basada en la nube, accesible desde cualquier dispositivo con conexión a Internet, que se enfoca en la visualización interactiva para proporcionar a los usuarios una experiencia de análisis de datos en tiempo real. La interfaz de Qlik Sense es muy intuitiva y ofrece una función de arrastrar y soltar que permite a los usuarios crear sus propios modelos analíticos personalizados sin la necesidad de tener conocimientos previos en programación. Además, permite la creación de visualizaciones, dashboards y aplicaciones que responden a las preguntas más relevantes del negocio.

En resumen, Qlik Sense ofrece una experiencia de análisis de datos fácil y efectiva que permite a los usuarios explorar y descubrir información útil en sus datos de forma rápida y eficiente, para tomar decisiones empresariales más informadas.[17]

Para el desarrollo de este proyecto de investigación se seleccionará Power Bi como plataforma analítica de inteligencia de negocio, debido a sus características frente a otra plataforma, como su facilidad de uso, integración con Microsoft, ofrece una versión gratuita, altamente personalizable ajustándose a las necesidades del proyecto.

Ilustración 6 Cuadrante Mágico de Gartner plataformas analíticas de inteligencia de negocio.

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Autor: Gartner.

2.4.1.12 Aprendizaje automático (Machine Learning)

El Machine Learning es una rama de la Inteligencia Artificial que se enfoca en desarrollar algoritmos capaces de aprender sin necesidad de ser programados explícitamente. En lugar de ello, se alimenta el algoritmo con grandes volúmenes de datos para que pueda aprender y actuar de forma adecuada en diferentes situaciones y escenarios. [18]

Existen dos tipos de aprendizaje en Machine Learning: supervisado y no supervisado:

En el aprendizaje supervisado, se proporcionan al algoritmo tanto las características como las etiquetas, para que pueda predecir y clasificar de forma precisa. El aprendizaje supervisado incluye dos tipos de algoritmos: clasificación y regresión. [18]

- **Algoritmo de clasificación:** este encuentra patrones en los datos proporcionados y clasifica un elemento en uno de los grupos establecidos. Con este método, esperamos que el algoritmo nos diga a qué grupo pertenece el elemento que estamos estudiando, y así puede predecir de qué se trata. [18]
- **Algoritmo de regresión:** este devuelve un valor específico en lugar de clasificar un elemento en un grupo. En este método, se espera obtener un número que permita

predecir el valor de una variable dependiente en función de una o varias variables independientes. [18]

En el aprendizaje no supervisado, solo se proporcionan características al algoritmo, sin etiquetas. El objetivo es que el algoritmo agrupe los datos según sus características, asumiendo que los datos que comparten características similares pertenecen al mismo grupo.[18]

Modelos de Machine Learning

Existen tres modelos de algoritmo de Machine Learning:

- **Modelos lineales:** como la regresión lineal o la logística, buscan encontrar una línea que se ajuste bien a los datos proporcionados. Sin embargo, estos modelos pueden tener el problema de "overfitting", ajustándose demasiado a los datos disponibles y no siendo efectivos para comportamientos más complejos. [18]
- **Modelos de árbol:** como los árboles de decisión y los random forest, son más precisos y estables, y construyen reglas de decisión que se representan como un árbol. Estos modelos pueden representar relaciones no lineales y resolver problemas más complejos. Sin embargo, pueden ser más complicados de interpretar y tener un rendimiento inferior en comparación con los modelos lineales. [18]
- **Las redes neuronales:** son un tipo de modelo de aprendizaje automático que intenta replicar el comportamiento del cerebro humano, donde millones de neuronas se interconectan para enviar mensajes. Este modelo es muy efectivo en el reconocimiento de patrones en datos, como imágenes o videos, gracias a su habilidad para aprender características complejas y abstraer información útil de grandes conjuntos de datos. Las redes neuronales pueden tener diferentes arquitecturas y se utilizan en diversas aplicaciones, como el reconocimiento de voz, el procesamiento del lenguaje natural, la detección de fraudes, entre otros. [18]

Fases de desarrollo

El proceso de desarrollo de un algoritmo de aprendizaje automático consta de dos fases principales: entrenamiento y prueba.

- **Fase de Entrenamiento:** Durante esta fase, se recopila una gran cantidad de datos y se divide una parte para entrenar el algoritmo. Se proporciona toda esta información

al algoritmo para que encuentre patrones importantes y pueda hacer predicciones precisas en el futuro. Durante el entrenamiento, los parámetros del modelo se ajustan para minimizar el error en las predicciones.

- **Fase de Prueba:** En la fase de prueba, se utiliza el resto de los datos para evaluar la eficacia del modelo. Se hacen preguntas al modelo y se comparan las respuestas con las respuestas esperadas. Si el modelo no tiene un rendimiento satisfactorio, se deben agregar más datos o ajustar el método utilizado. Si el modelo tiene una tasa de aciertos del 80% al 90%, se considera que ha aprendido lo suficiente y se puede utilizar en la aplicación correspondiente. [18]

2.4.1.13 Series temporales.

Una serie temporal es una colección secuencial de observaciones de una variable en el tiempo, en la que el orden de las observaciones es importante. Cada valor de la serie está asociado a un instante de tiempo, lo que significa que su análisis implica el manejo de dos variables: la variable en estudio y la variable tiempo. Las series temporales pueden tener diferentes periodicidades, como anual, semestral, trimestral, mensual, etc., según los periodos en los que se recolectan los datos. Ejemplos de series temporales son las ventas anuales de una empresa o el número de casos semanales de accidentes de trabajo en una empresa.

El análisis de las series temporales implica el uso de técnicas estadísticas que permiten estudiar y modelar el comportamiento de un fenómeno en el tiempo, así como hacer predicciones sobre los valores futuros de la serie.[27]

Las series temporales pueden ser descompuestas en diferentes componentes que permiten entender su comportamiento y realizar análisis más precisos.[28]

- **Nivel:** representa el valor medio en toda la serie temporal.
- **Tendencia:** indica la dirección general del comportamiento de la serie temporal a largo plazo, es decir, si la serie temporal aumenta, disminuye o se mantiene constante en el tiempo.
- **Componente cíclico:** refleja fluctuaciones periódicas que no están relacionadas con la estacionalidad y que suelen tener una duración mayor que esta. Estos ciclos no tienen por qué ser periódicos.

- **Ruido:** representa la variación aleatoria de la serie temporal que no puede ser explicada por los otros componentes.

La identificación y separación de estos componentes ayuda a comprender mejor el comportamiento de la serie temporal y a realizar pronósticos más precisos en el futuro mediante la aplicación de técnicas estadísticas. [28]

Principales algoritmos de Machine Learning que se utilizan para el análisis de series temporales.

- **Autoregressive Integrated Moving Average (ARIMA):** es un algoritmo de aprendizaje automático ampliamente utilizado en la modelización y predicción de series de tiempo. Este se compone de tres componentes principales: la parte autorregresiva (AR), la parte de media móvil (MA) y la parte de integración (I). La parte autorregresiva (AR) se utiliza para modelar la dependencia de una variable en sus valores pasados. La parte de media móvil (MA) modela la relación entre la variable y los errores de predicción pasados. La parte de integración (I) se utiliza para transformar la serie de tiempo en una serie estacionaria, lo que facilita el modelado y la predicción.
- **Algoritmo Suavización exponencial simple (SES):** Es una técnica de pronóstico utilizada en el análisis de series temporales. Este modelo busca predecir el próximo valor en una serie temporal basándose en los datos históricos observados y una tasa de suavización exponencial. Su funcionamiento consiste en asignar pesos decrecientes a los datos y observaciones anteriores, donde los pesos disminuyen exponencialmente con el tiempo. Esto muestra que las observaciones más recientes tienen un mayor impacto en la predicción que los datos u observaciones antiguas. El algoritmo determina un valor suavizado para cada intervalo de tiempo
- **Algoritmo Prophet:** Es un algoritmo de pronósticos de regresión aditiva desarrollado por Facebook con una tendencia de curva y crecimiento lineal o logístico por partes, que se utiliza para predecir series temporales con tendencias estacionales. Este algoritmo incluye tres componentes principales: La tendencia es la dirección general de los datos a lo largo del tiempo, la estacionalidad expresa patrones que se repiten

en intervalos regulares del tiempo, los días festivos los cuales permite incorporar información sobre eventos especiales que pueden llegar a afectar la tendencia de los datos, como días festivos.

2.4.1.14 Lenguaje de programación.

Un lenguaje de programación se refiere a un conjunto de normas y símbolos que se utilizan para crear programas de computadora. Está compuesto por un sistema formal diseñado para estructurar algoritmos y procesos lógicos que serán ejecutados por una computadora o sistema informático. Esto permite controlar su comportamiento físico, lógico y su interacción con el usuario humano.

Existen diferentes tipos de lenguajes de programación, incluyendo:

- **Lenguajes de bajo nivel:** Estos lenguajes están específicamente diseñados para un hardware particular y no se pueden transferir o utilizar en otras computadoras diferentes.
- **Lenguajes de alto nivel:** Estos lenguajes tienen como objetivo ser más universales y pueden ser utilizados independientemente de la arquitectura del hardware en diferentes tipos de sistemas. Pueden ser de propósito general, utilizados en una amplia gama de aplicaciones, o de propósito específico, diseñados para tareas especializadas.
- **Lenguajes de nivel medio:** Estos lenguajes se presentan como un punto intermedio entre los dos tipos anteriores. Permiten realizar operaciones de alto nivel al mismo tiempo que se tiene cierto control sobre la arquitectura del sistema en el que se ejecutan.

Cada tipo de lenguaje de programación tiene sus ventajas y desventajas, y la elección del mejor depende del propósito y las necesidades específicas del proyecto. [29]

Top 10 de los mejores lenguajes de programación para científicos de datos en el 2023.[30]

Tabla 2 Top 10 mejores lenguajes de programación.

1. Python	2.R	3.SQL	4.Java	5.Julia
6.Scala	7.C/C++	8.JavaScript	9.Swift	10.Go

- **Python:** Python es un lenguaje de programación de nivel alto que se caracteriza por ser interpretado, interactivo y orientado a objetos. Es ampliamente reconocido y utilizado en la industria del machine learning y la ciencia de datos. Su sintaxis clara y legible lo hace ideal para principiantes, mientras que su extensa colección de bibliotecas y frameworks lo convierten en una herramienta versátil y útil en diversas aplicaciones. Python es multiplataforma, lo que significa que puede ser ejecutado en diferentes sistemas operativos como Windows, Linux y Mac OS.
- **R:** R es un lenguaje de programación de código abierto e interpretado con un enfoque específico en estadísticas, visualización y análisis de datos. Es una herramienta potente para el análisis estadístico y modelado de datos. Además de ser multiplataforma y orientado a objetos, R destaca por su capacidad para generar gráficos de alta calidad y su interfaz de línea de comandos que permite una programación eficiente y flexible.
- **Java:** Java es un lenguaje de programación de alto nivel, orientado a objetos y de propósito general. Se destaca por ser seguro y portable, lo que significa que el código escrito en Java puede ejecutarse en diferentes plataformas sin necesidad de volver a compilarlo. Esta capacidad de portabilidad se debe en gran medida a la Máquina Virtual de Java (JVM), que actúa como una capa de abstracción entre el código Java y el sistema operativo subyacente. Java es ampliamente utilizado en el desarrollo de aplicaciones empresariales, juegos, aplicaciones móviles y de escritorio, entre otros.

Para llevar a cabo este proyecto de investigación, se ha elegido R como lenguaje de programación para desarrollar el algoritmo de aprendizaje automático. Se han considerado sus características, tales como que está diseñado especialmente para el análisis estadístico y la visualización de datos, es popular para desarrollar algoritmos de proyección, puesto que cuenta con una gran variedad de paquetes o librerías, su facilidad de aprendizaje, su amplia comunidad de usuarios dispuesta a ofrecer ayuda cuando se necesite y su gran uso en la industria del aprendizaje automático.

2.4.2 Antecedentes

Internacional

Zerpa R, Becerra G, Izquierdo O, Velásquez G (2020). Realizaron una investigación con el objetivo general de implementar un Data Warehouse para soporte de decisiones en empresas de comercialización de mercancía. La metodología utilizada fue HEFESTO, que se caracteriza por su facilidad de entendimiento y fases bien definidas.

La implementación del almacén de datos fue satisfactoria, utilizando herramientas SSDDT y un entorno gráfico de Visual Studio en conjunto con el servidor SQL. Se logró realizar el proceso de extracción, transformación y carga desde la base de datos transaccional hacia el Data Warehouse.

A través de la Inteligencia de Negocio, se analizaron las áreas de ventas y compras, lo que permitió generar informes dinámicos con el fin de obtener indicadores clave que apoyen al ejecutivo a la hora de tomar decisiones. [19]

Pereda M, Cabrera S (2019). Realizaron una investigación con el objetivo de mejorar el análisis de la información en los procesos de ventas de la empresa Ingenieros en Acción S.R.L. a través de la implementación de una Solución de Inteligencia de Negocios (BI).

La solución propuesta se basó en la arquitectura tecnológica de Inteligencia de Negocios de Pentaho y se apoyó en la metodología de Ralph Kimball. Se analizaron los requerimientos del área crítica del negocio y se implementó el proceso ETL, las dimensiones y cubos, para luego crear la vista de análisis y navegabilidad de información.

Como resultado, se obtuvieron reportes dinámicos para cada requerimiento, mejorando el análisis de la información del proceso de ventas de la empresa y reduciendo el tiempo para obtener dicha información. Los usuarios del BI se mostraron satisfechos con los resultados obtenidos. [20]

Nacional

Parra C, Rincón R, Romero S (2019). Realizaron una investigación con el objetivo de proponer la herramienta Microsoft Power BI y su aplicativo Power BI Desktop como solución de inteligencia de negocios para la Empresa ABC, que se dedica a la manufactura de productos plásticos. La propuesta tenía como finalidad mejorar el almacenamiento,

visualización y calidad de los reportes de información, para contribuir en la toma de decisiones del área comercial.

Para lograr este objetivo se utilizaron los métodos inductivo y deductivo. La investigación se enfocó en cómo la herramienta Power BI podría contribuir a mejorar la toma de decisiones en el área de ventas de la empresa.

Como resultado de la aplicación de la herramienta, se logró reducir los tiempos de concertación gracias a la eliminación de diferencias en los informes generados. Además, los datos dejaron de pertenecer a personas individuales y se convirtieron en materia prima para el área, mejorando así el proceso de toma de decisiones en el área comercial de la Empresa ABC.[21]

Álvarez R, Gómez S (2022). Realizaron una investigación con el objetivo de diseñar un modelo de inteligencia de negocios para la empresa Puntual Arquitectura Ingeniería. Se utilizó la metodología de las cinco fuerzas de Porter para analizar las fuerzas externas y variables de la industria y lograr la ventaja competitiva.

El modelo de inteligencia de negocios diseñado propuso el uso de una herramienta especializada en inteligencia empresarial y analítica que permita a la gerencia generar reportes de ventas, costos y margen por ciudad, producto, cliente y proveedor mediante gráficas de barras, líneas de tendencia, gráficas de torta, KPI y tablas en cualquier dispositivo tecnológico con acceso a Internet.

Como resultado de la aplicación de la herramienta, los gerentes y analistas de información ahorraron tiempo en el desarrollo y actualización de informes, lo que les permitió concentrarse en la generación de estrategias para acercarse al objetivo del direccionamiento estratégico de la empresa. El sistema también permitió identificar fácilmente los procesos que debían mejorarse dentro de la empresa para lograr las metas definidas.[22]

Guzmán A, Arenales R (2022) Realizaron una investigación con el objetivo de diseñar y analizar un modelo de inteligencia de negocios que permita mejorar y facilitar la toma de decisiones en las pequeñas y medianas empresas del sector de desarrollo de software en

Colombia. Para lograr este objetivo, se utilizó el método de investigación de análisis, el cual implicó la implementación de un modelo de inteligencia de negocios en las PYMES, considerando el carácter financiero que esto implica y la relación causa-efecto que puede tener en la toma de decisiones oportunas en las organizaciones.

La implementación del modelo de inteligencia de negocios permitió la creación de un modelo de datos consistente con la misión y visión de la organización desde su constitución inicial. Esto otorgó facilidad para la toma de decisiones estratégicas de manera eficaz y oportuna, lo cual mejorará la capacidad de las PYMES del sector de desarrollo de software en Colombia para enfrentar los desafíos y oportunidades del mercado.

Como resultado del proyecto, se definió un modelo de datos que funciona de manera general para las organizaciones orientadas al desarrollo de software. Este modelo se convierte en una guía inicial de implementación para las empresas emergentes, permitiéndoles tomar decisiones más informadas y eficaces en su gestión empresarial. En conclusión, la implementación del modelo de inteligencia de negocios es una herramienta valiosa para mejorar la toma de decisiones en las PYMES del sector de desarrollo de software en Colombia.[23]

Local

Araque F, Perdomo F, Reyes M, Vera C (2021) Realizaron una investigación con el objetivo de la construcción de un modelo artificial utilizando el software Rstudio para permitir al director de una empresa de servicios tener una visión intuitiva sobre el funcionamiento de su compañía y tomar decisiones acertadas. La investigación fue de carácter descriptivo, ya que se midieron variables o conceptos para integrarlos y determinar cómo se manifiesta un proceso y generar pronósticos a partir de este.

Se construyó una base de datos a partir de la información mensual suministrada por la empresa E.S.P EMSERPUCAR del municipio de Cartagena del Chairá, departamento del

Caquetá. Utilizando las series de tiempo se construyó un modelo matemático de regresión lineal múltiple (RLM) que se optimizó y adaptó para obtener un modelo de regresión lineal simple (RLS), ya que las variables no suministraban información suficiente para un modelo más efectivo.

Además, se construyó un modelo artificial utilizando el método de predicción ARIMA basado en las series de tiempo de los datos del consumo mensual del servicio de acueducto del municipio de Cartagena del Chairá. Se espera que este modelo sea capaz de predecir el número de usuarios que utilizarán el servicio de acueducto en el municipio para el año 2021 con una confiabilidad del 86,36%. El segundo modelo indica que se espera que el consumo de agua en el municipio sea estable para los primeros 10 meses del año 2021, con un promedio de aproximadamente 25.348 metros cúbicos de agua por mes.[24]

2.5 Marco metodológico

En esta sección se presentará una descripción detallada de los métodos y técnicas que se utilizarán para analizar y resolver el problema de investigación planteado. Se especificará el tipo de investigación, la población y muestra, los instrumentos y técnicas que se emplearán para la recolección de datos. Además, se expondrá la metodología que se utilizará para cumplir con los objetivos del proyecto.

2.5.1 Tipo de estudio y diseño de investigación

En este proyecto de investigación se empleará el enfoque cuantitativo, que consiste en la recolección y análisis de datos numéricos. Este tipo de estudio es fundamental, ya que permite la observación empírica y la conexión con la dimensión conceptual de la investigación. Asimismo, los procesos utilizados bajo este enfoque son objetivos, deductivos y probatorios, lo que permite obtener resultados precisos y medibles. Este enfoque resulta especialmente adecuado en este proyecto, puesto que se busca medir diversos datos para su análisis y toma de decisiones, y también se pretende hacer predicciones o proyecciones a partir de los datos obtenidos.

En este trabajo se utilizará un diseño de investigación descriptivo, el cual tiene como objetivo principal la descripción detallada y sistemática de los fenómenos, situaciones o características

de la población o muestra en estudio. Este tipo de diseño permite obtener información detallada sobre lo que se está investigando, pero no busca establecer relaciones causales o explicar las razones detrás de los fenómenos observados.

Basándonos en el enfoque cuantitativo, se llevará a cabo un análisis estadístico preciso y probatorio en la empresa del caso de estudio, con el fin de identificar patrones de comportamiento en las ventas y hacer predicciones sobre compras futuras. Estos resultados serán útiles para mejorar la toma de decisiones y resolver la problemática planteada en este proyecto de investigación.

2.5.2 Población y muestra

Población

La población objetiva del proyecto de investigación está delimitada por la empresa Materiales Maderas Santander SAS, ubicada en la ciudad de Valledupar, Cesar. El grupo que será encuestado está conformado por la gerencia de la empresa. Se ha llevado a cabo un análisis exhaustivo, considerando los roles fundamentales en la toma de decisiones estratégicas dentro de la empresa, quienes serán los principales receptores de los hallazgos de esta investigación. El propósito del proyecto es establecer una solución de inteligencia empresarial dirigida al departamento de ventas y compras de la organización, lo que proporcionará una herramienta eficiente para tomar decisiones en esta área.

Muestra

Dado que las ventas son la actividad principal de Materiales Maderas Santander, la empresa debe asegurarse de tener suficiente suministro de sus productos para llevar a cabo dichas ventas. Con el fin de cumplir con el objetivo de este proyecto de investigación, se tomó una muestra no probabilística de las áreas relacionadas en la empresa, representada por 00 tablas de la base de datos de la empresa.

El estudio se llevará a cabo durante el período que abarca desde enero de 2018 hasta diciembre de 2023.

2.5.3 Instrumentos y técnicas de recolección de información

Para llevar a cabo este proyecto de investigación, se obtuvo información a través de varias entrevistas con preguntas abiertas al Gerente Administrativo de la empresa Materiales Maderas Santander (ver Anexo I). Durante estas entrevistas, se identificó la problemática de la empresa y cómo se está abordando esta situación. Además, se recopilaron los requisitos

del negocio en el área de compras y ventas. También se obtuvo la fuente de datos, que se describe en la tabla 3.

Tabla 3 Información base de datos fuentes.

Sistema gestor de base de datos	Oracle 11 XE
Numero de tablas	124
Peso (MB)	100,8125
# tablas en el área de ventas	15
Peso tablas área de ventas (MB)	36,992

Autor: Macias Anyelyn, Ortiz Juan.

Según la base de datos proporcionada por el Gerente Administrativo de la empresa Materiales Maderas Santander, se seleccionó como muestra la información correspondiente al área de ventas y compras del periodo de enero a diciembre de 2022. En la Tabla 4 se muestra el volumen de datos seleccionados para el desarrollo de este proyecto de investigación.

Tabla 4 Descripción de las tablas relacionadas al área de ventas.

#	Nombre de la tabla	Descripción	# Registro	Peso (MG)
1	Tercero	Contiene información general de terceros.	6.404	2
2	Producto	Contiene información de los productos que oferta la empresa.	1.664	0,375
3	Bodega	Contiene información de las bodegas existentes.	6	0,062
4	Tercero_Fcredito	Contiene información relevante para realizar créditos a terceros.	8.677	0,125
5	Porcentaje_Iva	Contiene el porcentaje de IVA que se le aplica a los productos.	5	0,062
6	Mfactura	Contiene información de la factura.	68.290	13
7	Dfactura	Contiene información del detalle de la factura.	158.638	12
8	DDevolucion	Contiene información del detalle de las devoluciones.	2.602	0,187

9	Mdevolucion	Contiene información de la devolución.	1.966	0,312
10	Tipo_Comprobante	Contiene información de los tipos de comprobantes de pago que tiene la empresa.	32	0,062
11	Dfactura_Impuesto	Contiene información del impuesto que se le aplica a la factura.	197.844	7
12	Medio_Pago	Contiene los medios de pago que tiene la empresa.	8	0,62
13	Producto_Bodega	Contiene información de los productos registrados en cada bodega.	10.044	0,375
14	Clase_Impuesto	Contiene información de los impuestos que tiene la empresa.	13	0.062
15	Historico_Producto	Historial de los productos.	13.611	0,812
TOTAL			469.804	36,992

Autor: Macias Anyelyn, Ortiz Juan.

2.5.4 Metodología para el desarrollo del proyecto

En esta sección se describirá de manera detallada la metodología de inteligencia de negocio elegida para el desarrollo de este proyecto de investigación, incluyendo los pasos, técnicas y herramientas utilizadas.

2.5.4.1 Metodología de Kimball

Para el desarrollo de este proyecto de investigación, se utilizará la metodología de Kimball o también conocida como enfoque dimensional de Kimball, la cual se enfoca en la creación de modelos dimensionales simples y fáciles de entender, que permitan a los usuarios finales analizar la información de manera efectiva. Está compuesta por cuatro fases principales.

El diagrama representa de manera clara el ciclo de vida de la metodología, mostrando de forma secuencial los pasos y tareas necesarias para lograr un buen diseño, desarrollo e implementación de un Data Warehouse.

Fases del ciclo de vida de la metodología de Kimball.

2.5.4.1.1 Planificación del proyecto

Etapa inicial de la metodología, en esta se define el propósito del proyecto de Data Waterhouse Business Intelligence, los objetivos específicos y el alcance de este mismo, además de identificar los riesgos principales y aproximarse a las necesidades de información.

Para lograrlo, se llevan a cabo acciones características de un plan de proyecto, como la definición del alcance, la identificación y programación de tareas, la planificación de recursos y la asignación de cargas de trabajo. Todo esto se documenta en un plan de proyecto final.

Además, se establece la administración o gestión de esta subfase, que en sí misma es un proyecto completo, se realizan actividades como el monitoreo del estado de los procesos y actividades, la exploración de problemas, la elaboración de un plan de comunicación integral que involucre a la empresa y las áreas de tecnología de la información. [25]

2.5.4.1.2 Administración del proyecto DW/BI

La etapa de gestión de proyecto se centra en asegurar la continuidad junto al logro de los objetivos planteados. En esta etapa se desarrollan diversas acciones, incluyendo el seguimiento constante del progreso del proyecto para detectar desviaciones y aplicar correcciones, la identificación, así mismo la resolución oportuna de los problemas que surgen, el control de los cambios en los requisitos del proyecto, la elaboración de un plan de comunicación integral que involucre a todas las partes interesadas, y por último la administración de expectativas con el propósito de asegurar que los objetivos del proyecto se alcancen de manera efectiva, dentro de los límites presupuestarios teniendo en cuenta los plazos establecidos.

2.5.4.1.3 Definición de Requerimientos del Negocio

En esta etapa, se realiza el proceso de definición de requerimientos mediante entrevistas con el personal de ámbito empresarial y técnico. Este procedimiento implica una previa preparación, también una comprensión profunda del negocio, la competencia, la industria, la clientela, entre otros. Además, es necesario examinar informes, documentos de estrategia interna, así como analizar la cobertura mediática acerca de la organización, la competencia y la industria. Paralelamente, resulta crucial adentrarse en los términos y la terminología propios del negocio.

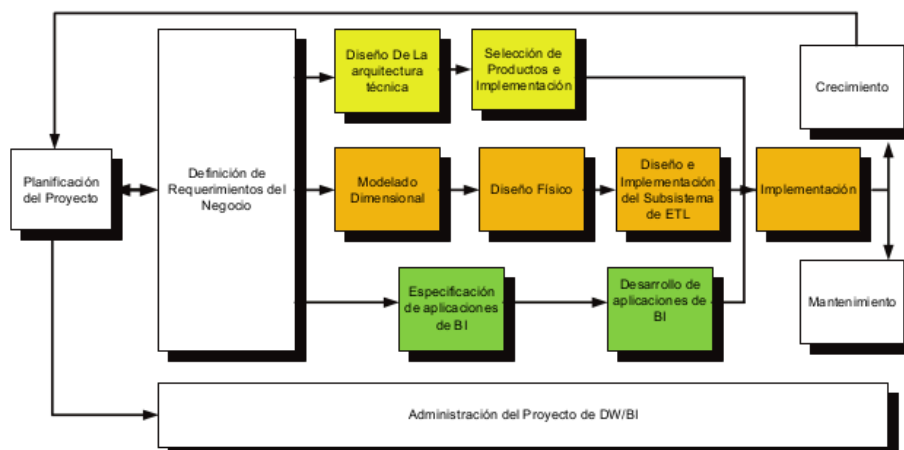
Se recomienda entrevistar a cuatro grupos de personas: el directivo responsable de tomar decisiones estratégicas, los administradores intermedios y de negocio responsables de explorar alternativas estratégicas y aplicar decisiones, el personal de sistemas (si existe), y el personal entrevistado por razones políticas. Los requerimientos del negocio son el soporte inicial de las tareas subsiguientes y también influyen en el plan de proyecto.[25]

Resulta esencial resaltar el valor primordial de la labor de definición de requerimientos, ya que estos constituyen el cimiento sobre el cual se erigen las actividades subsiguientes. En la

figura 7 se puede observar que esta fase se divide en tres caminos o subfases simultáneas que se enfocan en tres áreas: tecnología, datos y aplicaciones del BI.

- **Línea tecnológica (Camino superior):** La línea tecnológica comienza con el diseño de la arquitectura del sistema, que establece una lista de las capacidades necesarias. Luego, se procede a la selección e instalación de productos que satisfacen esas necesidades de la arquitectura. [26]
- **Línea de datos (Camino medio):** La estrategia de datos comienza con el diseño de un modelo dimensional que satisfaga las necesidades del negocio y considere las realidades de los datos subyacentes. En el mundo de Kimball, el modelo dimensional es la clave y los datos se organizan en medidas y dimensiones descriptivas. Estos modelos pueden implementarse en bases de datos relacionales, utilizando esquemas de estrella, o en bases de datos multidimensionales, utilizando cubos OLAP. En ambos casos, el objetivo principal es proporcionar facilidad de uso para los usuarios y velocidad en la ejecución de consultas. [26]
- **Línea de aplicación del BI (Camino inferior):** En el camino de implementación, se dedican a descubrir y crear una amplia gama de aplicaciones de inteligencia empresarial, que incluyen informes estándar, consultas con parámetros, paneles de control, tableros de datos, modelos analíticos, herramientas de extracción de datos y las interfaces de navegación correspondientes.[26]

Ilustración 7 Ciclo de vida de la metodología de Kimball.



Autor: Rivadera Gustavo R.

2.6 Resultados esperados

Mediante la ejecución de este proyecto de investigación, se busca implementar mejoras en la gestión de datos y la toma de decisiones en el ámbito de ventas y compras de la empresa Materiales Maderas Santander.

Principalmente, se plantea diseñar y desarrollar un data mart que permita almacenar y disponer de los datos cuando sea necesario de una forma más rápida y efectiva en comparación con los métodos actuales de la empresa materiales maderas Santander. Se busca disminuir, también optimizar el tiempo de consulta de los datos, facilitando el acceso a la información.

Por otro lado, se busca realizar una proyección de ventas por medio de un algoritmo de machine learning que permita dar respuesta a preguntas complejas sobre hechos pasados y futuros en el área de ventas y compras en la empresa materiales madera Santander.

Por último, no menos importante un módulo donde se pueda visualizar información mediante de dashboard iterativo con gráficas, tablas y filtros que faciliten el análisis de la información, para posteriormente tomar decisiones basándose en los datos. Estas decisiones beneficiarán a la empresa, puesto que mejorarán la eficiencia, efectividad en las estrategias de ventas y compras, todo esto con el fin de que la empresa logre tener un crecimiento significativo, se posicione en el mercado siendo una empresa competitiva, fidelizando clientes y disminuyendo pérdidas.

Teniendo en cuenta lo anterior, lo que se espera lograr es la integración del módulo de visualización con el data mart y los resultados del algoritmo de machine learning, para que los directivos de la empresa puedan explorar, entender los datos de una manera efectiva, logrando con esto una mejora en la toma de decisiones que ayudaran a la empresa a crecer y posicionarse en el mercado.

2.7 Cronograma de actividades

Autor: Macias Anyelyn, Ortiz Juan.

Recursos tecnológicos

Tabla 6 Presupuesto recursos tecnológicos.

Hardware	Estado	Cantidad
Computador procesador RYZEN 5, memoria RAM 8 G.	Buen estado	1
Computador procesador I5 12 th, memoria RAM 16 G.	Buen estado	1
Software	Estado	Cantidad
Microsoft power BI	Licencia por usuario gratis	2
Windows 10	vigente	2
Microsoft Office	vigente	2
Talend	Versión de prueba gratis	2
Visual code	Vigente	2
Servicio de internet	Vigente proveedor claro	1

Autor: Macias Anyelyn, Ortiz Juan.

Inversión de recursos por utilizar

Tabla 7 Presupuesto inversión de recursos por utilizar.

Recurso	Detalle	Valor (COP)
Personal	2 personas	\$10.000.000
hardware	Cumple con los requerimientos ya que se cuenta con estos equipos disponibles.	\$ 0
Software	Servicio de internet	Valor requerido por 8 meses \$800.000
Presupuesto para riesgos.		\$3.000.000
Presupuesto estimado valor total.		\$ 13.800.000

SECCIÓN III: DESARROLLO CIENTÍFICO – TECNOLÓGICO

Esta sección se enfocará en el proceso de desarrollo de la metodología de kimball la cual fue elegida cuidadosamente en el contexto de este proyecto. Teniendo en cuenta sus características, estas mismas que han demostrado ser las adecuadas para satisfacer de manera precisa los requerimientos del problema que se enfrenta.

3.1 Desarrollo de las fases de la metodología propuesta.

3.1.1 Planificación del proyecto

En esta etapa inicial, se define el propósito de la solución de Data Warehouse y Business Intelligence (DW/BI) que se pretende desarrollar, junto con sus objetivos, requerimientos y la extensión de dicho proyecto. También, se identifican los riesgos clave y se realiza un primer análisis de las condiciones y necesidades de los datos. Seguidamente, se presentan los aspectos fundamentales que deben abordarse en la fase de planificación de este proyecto.

3.1.1.1 Objetivo del proyecto

Ofrecer a la empresa Materiales Maderas Santander SAS una solución de Business Intelligence (BI) a través de la construcción de un Data Mart para el área de ventas y compras, que se ajuste a los requisitos previamente establecidos, cumpliendo con las mejores prácticas y principios metodológicos de Kimball para el desarrollo de un data warehouse. El propósito fundamental de esta iniciativa es facilitar la toma de decisiones por parte de los directivos de la empresa, apoyándose en el análisis de datos efectuado mediante la herramienta Power BI.

3.1.1.2 Alcance del Proyecto

Este proyecto tiene como objetivo diseñar, desarrollar e implementar una solución de inteligencia de negocio orientada al área de ventas y compras de la empresa Materiales Maderas Santander SAS. La implementación de esta solución tiene el propósito de mejorar la toma de decisiones empresariales y optimizar la proyección de compras.

Esta solución permitirá a los líderes de la empresa tener una visión global de la situación actual, así como comprender las implicaciones y beneficios potenciales de las decisiones que tomen.

Dentro del módulo de visualización, se presentará la información a través de paneles interactivos que incluyen una variedad de gráficos para facilitar la interpretación de los datos. Estos paneles permitirán el análisis de datos históricos de la empresa, lo que ayudará a responder preguntas como: ¿qué sucedió en el pasado? ¿Por qué ocurrió? ¿Qué está sucediendo en este momento? Además, el algoritmo de proyección de compras proporcionará respuestas a la pregunta "¿qué ocurrirá en el futuro?"

Para lograr esto, se creará una bodega de datos que recopilará y almacenará información relacionada con las áreas de ventas y compras, así como todos los procesos asociados a ellas. La creación de esta bodega de datos implicará un análisis exhaustivo del negocio, la extracción de datos, su procesamiento, transformación y su posterior carga (ETL) desde la base de datos existente proporcionada por la empresa hacia una nueva base de datos. Esto permitirá un acceso más rápido, eficiente y efectivo a los datos de la bodega, sin afectar los procesos de la fuente de datos original.

Además, se desarrollará un algoritmo de proyección de compras que se alimentará y entrenará con datos tanto de la fuente de datos proporcionada por la empresa como de la bodega de datos. Una vez entrenado, este algoritmo proporcionará proyecciones precisas de las compras futuras de la empresa.

3.1.2 Análisis de requerimientos

En esta etapa, se establecerán los requerimientos del negocio mediante la evaluación de los procedimientos operativos relacionados con las áreas de ventas y compras. Estos requisitos se derivarán de las conversaciones mantenidas con la directora administrativa de la empresa Materiales Maderas Santander SAS durante las entrevistas.

3.1.2.1 Identificación de temas analíticos y procesos del negocio

Durante la entrevista con la gerente administrativa, quedó claro que los procesos operativos fundamentales de la empresa son las ventas y las compras, es decir, la comercialización de productos de ferretería. Por lo tanto, se decidió que el enfoque principal de este proyecto se centrará en los procesos en el área de ventas y compras.

Además, se discutió que la toma de decisiones en la empresa se basa en la experiencia de los administradores y en los análisis de datos realizados por la gerente administrativa a través de los informes generados por el sistema actualmente en uso. Por lo tanto, la solución de inteligencia de negocios se concentrará en analizar datos relacionados con ventas y compras durante períodos específicos. De esta manera, los administrativos de la empresa contarán con información sólida para respaldar sus decisiones.

3.1.2.2 Definición de requerimientos

La gerente administrativa expresó durante la entrevista que sería altamente beneficiosa que la solución de inteligencia de negocio pudiera abordar las siguientes interrogantes:

- ¿Cómo se ha comportado el volumen de ventas en diferentes meses a lo largo del año?
- ¿Cuál es la proyección de ventas para el segundo trimestre del 2023?
- ¿Quiénes son los clientes más destacados en función de sus compras?

- ¿Quiénes son los vendedores más destacados en función de la utilidad de sus ventas?
- ¿Cuáles son las categorías de productos que presentan el mayor porcentaje de devolución?
- ¿Cuáles son los productos con el mayor porcentaje de devoluciones?
- ¿Cuáles son las categorías de productos que representan un mayor valor en relación con las ventas totales?
- ¿Cuáles son los productos que representan un mayor valor en relación con la utilidad total de la empresa?
- ¿Cuáles son los productos más vendidos por cantidad?
- ¿Cómo ha sido el crecimiento de la empresa a través de los años?
- ¿Cuál es la proyección de compras para el próximo mes?

3.1.2.3 Matriz de procesos/dimensiones

Una vez que hemos comprendido y evaluado los requisitos del negocio, avanzamos utilizando una herramienta clave, conocida como la matriz de procesos/dimensiones, recomendada por la metodología de Kimball. Esta herramienta desempeña un papel esencial en la creación de un diseño eficiente y estructurado para la gestión de datos empresariales. La matriz de

procesos/dimensiones facilita la conexión entre los procesos operativos de la empresa y las dimensiones que se emplearán para analizar estos procesos de manera más profunda y significativa. En la siguiente tabla se puede observar la matriz resultante.

Tabla 8 Matriz de procesos y dimensiones.

PROCESO DE NEGOCIO / DIMENSIONES	DIM TIEMPO	DIM VENDEDOR	DIM CLIENTE	DIM PRODUCTO	DIM CATEGORÍA PRODUCTO
GESTIÓN CLIENTES	x		x		
GESTIÓN VENEDORES	x	x			
PROCESO DE VENTAS	x			x	x
CONTROL DE UTILIDAD	x			x	x

CONTROL DE DEVOLUCIONES	x	x	x	x	x
-------------------------	---	---	---	---	---

Autor: Macias Anyelyn, Ortiz Juan.

3.1.3 Modelado dimensional

En esta etapa, desarrollamos los modelos dimensionales que organizan y estructuran los datos de manera que sea más sencillo analizarlos y consultarlos. Estos modelos se construyen tomando como base los procesos empresariales previamente identificados en la matriz mencionada anteriormente. Para crear el modelo, seguimos un proceso iterativo de cuatro.

3.1.3.1 Elegir el proceso de negocio

El primer paso implica la selección del área a modelizar, que en este contexto se centrará en las áreas de ventas y compras dentro de la empresa Materiales Maderas Santander S.A.S. Esta elección se basa en un análisis exhaustivo de los requisitos y los temas analíticos que fueron identificados en la etapa anterior. Los aspectos clave que serán abordados incluyen:

- Venta de productos en la empresa: Se analizarán las transacciones de venta de productos realizadas en la empresa.
- Compra de productos en la empresa: Se examinarán las adquisiciones de productos efectuadas por la empresa.
- Utilidad generada por las ventas: Se calculará y se evaluará la utilidad que se genera a partir de las ventas realizadas.
- Devoluciones por las ventas: Se analizarán las devoluciones que puedan haber surgido como resultado de las ventas.
- Clientes de la empresa: Se profundizará en la información relacionada con los clientes de la empresa, lo que permitirá un análisis detallado de su comportamiento.
- Vendedores de la empresa: Se examinarán las actividades de los vendedores dentro de la empresa, brindando una visión completa de su desempeño.

3.1.3.2 Establecer el nivel de granularidad

La granularidad se refiere a la especificación del nivel de detalle con el que se manejan los datos. La elección de la granularidad está influenciada por las necesidades del negocio y las

limitaciones de los datos proporcionados por la empresa. En este contexto específico, se optó por un nivel de granularidad que se centra en la **factura** y el **producto**. Esta elección se hizo con el propósito de desglosar y comprender en detalle los procesos de negocio relacionados con las áreas de ventas y compras.

3.1.3.3 Elegir las dimensiones

Las dimensiones propuestas para la bodega de datos son las siguientes:

- Dimensión producto: Describe cada uno de los productos que hacen parte del inventario de la empresa.
- Dimensión cliente: Describe la información de los clientes que comprar en la empresa.
- Dimensión vendedor: Describe la información de los vendedores con los que cuenta la empresa.
- Dimensión tiempo: Describe de manera detallada la fecha en la que se realizó un hecho.
- Dimensión categoría de producto: Describe la categoría a la que pertenece cada producto que se encuentra en el inventario de la empresa.

3.1.3.4 Identificación de las tablas de hecho y medidas

Las tablas de hecho y medidas seleccionadas para a la creación de la bodega de datos son:

Tabla hecho facturas, almacenara las siguientes métricas aditivas y no aditivas:

- Costo Total: Valor total de los productos vendidos en una factura, incluyendo todos los costos asociados.
- Total, devuelto: Valor total de la factura devuelta por el cliente.
- Estado de la Factura: Este indicador muestra el estado actual de la factura, que puede ser "Pagada" o "Anulada".
- Venta real: Valor resultante después de restar las devoluciones del total facturado. Representa los ingresos reales generados por las ventas.
- Costo Real: Valor resultante después de restar las devoluciones del costo total de la factura. Representa los costos reales incurridos en la venta de productos.

- Utilidad Total: Valor total de la utilidad obtenida en una venta, calculado restando los costos a la venta real.
- Porcentaje de Utilidad: Representa la utilidad obtenida en una venta como un porcentaje de los ingresos netos.

Tabla hecho productos, almacenara las siguientes métricas aditivas y no aditivas:

- Cantidad Vendida: La cantidad de productos vendidos por día.
- Valor Total: El valor total de ventas por producto en el día.
- Venta Real: El valor neto resultante después de restar las devoluciones del total de ventas.
- Valor Devuelto: El valor total de productos devueltos por producto.
- Cantidad Devuelta: La cantidad de productos devueltos por día.
- Porcentaje de Devolución: El porcentaje de devoluciones por producto con respecto a la cantidad vendida por día.
- Utilidad Total: El valor total de utilidad obtenido por producto, calculado restando los costos de la Venta Real.
- Costo Real: El costo neto resultante después de restar las devoluciones del costo total por producto.
- Porcentaje de Utilidad: El porcentaje de utilidad obtenida por producto en relación con los ingresos netos por día.

3.1.3.5 Modelado

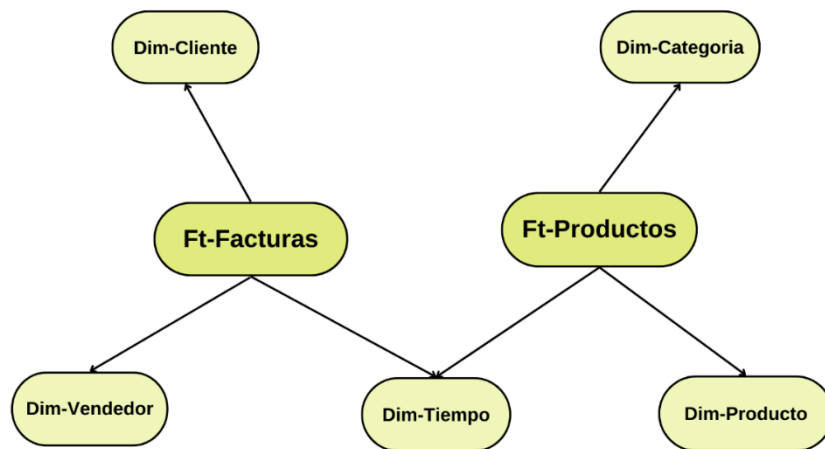
En esta parte se presentarán los modelos gráficos de nuestra bodega de datos que abarcan sus dimensiones lógica y conceptual. Estos modelos proporcionarán una representación visual detallada de la estructura y organización de la bodega de datos, lo que permitirá comprender mejor el diseño y el funcionamiento en cada uno de estos niveles. A través de estos modelos, podremos visualizar de manera clara cómo se relacionan y gestionan los datos dentro del entorno de inteligencia de negocio.

3.1.3.5.1 Modelo conceptual

En la ilustración 9, se presenta el modelo conceptual que se empleará como esquema para la organización de la bodega de datos. En este contexto, se adopta un enfoque de tipo

estrella, donde las tablas de hecho ocupan una posición central, representando así dos hechos distintos que son objeto de análisis. Se incluyen un total de cinco tablas de dimensión, cada una de ellas relacionada con perspectivas particulares que contribuyen a la descripción de estos hechos.

Ilustración 9 Modelo conceptual.

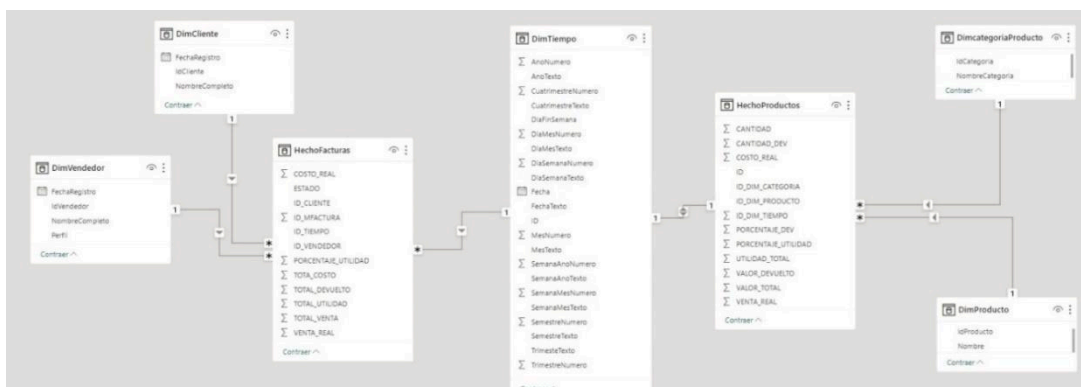


Autor: Macías Anyelyn, Ortiz Juan.

3.1.3.5.2 Modelo lógico

En la etapa de diseño del modelo lógico para el almacén de datos, se lleva a cabo un proceso meticuloso para identificar los atributos que formarán tanto las tablas de hechos como las dimensiones. Este proceso tiene como objetivo principal la creación de un Data Mart dirigido al departamento de ventas y compras de la empresa Materiales Maderas Santander S.A.S.

Ilustración 10 Modelo lógico.



- **Identificación de atributos de dimensiones y tablas de hechos**

Las tablas de la 9 a la 15 proporcionan un desglose de la estructura que conformará el Data Mart. Esto incluye el nombre de cada tabla, los nombres de sus atributos junto con sus respectivos tipos de datos, así como la especificación de claves primarias y foráneas para definir las relaciones entre ellas.

Tabla 9 Definición de los atributos de la dimensión categoría del producto.

DimCategoríaProducto		
PK	IdCategoría	DECIMAL(22,0)
	NombreCategoría	VARCHAR(50)

Autor: Macias Anyelyn, Ortiz Juan.

Tabla 10 Definición de los atributos de la dimensión cliente.

DimCliente		
PK	IdCliente	DECIMAL(22,0)
	NombreCompleto	VARCHAR(80)
	FechaRegistro	DATETIME

Autor: Macias Anyelyn, Ortiz Juan.

Tabla 11 Definición de los atributos de la dimensión producto.

DimProducto		
PK	IdProducto	DECIMAL(22,0)
	NombreCompleto	VARCHAR(80)
	FechaRegistro	DATETIME

Autor: Macias Anyelyn, Ortiz Juan.

Tabla 12 Definición de los atributos de la dimensión tiempo.

DimTiempo		
PK	ID	INT
	Fecha	DATE
	FechaTexto	VARCHAR(15)
	DíaSemanaNumero	INT
	DíaSemanaTexto	VARCHAR(15)
	DíaMesNumero	INT
	DíaMesTexto	VARCHAR(15)
	SemanaMesNumero	INT
	SemanaMesTexto	VARCHAR(20)
	SemanaAñoNumero	INT
	SemanaAñoTexto	VARCHAR(15)
	MesNumero	INT
	MesTexto	VARCHAR(10)
	AñoNumero	INT
	AñoTexto	VARCHAR(20)
	SemestreNumero	INT
	SemestreTexto	VARCHAR(15)
	TrimestreNumero	INT
	Trimeste Texto	VARCHAR(15)

	CuatrimestreNumero	INT
	CuatrimestreTexto	VARCHAR(15)
	DiaFinSemana	VARCHAR(15)

Autor: Macias Anyelyn, Ortiz Juan.

Tabla 13 Definición de los atributos de la dimensión vendedor.

DimVendedor		
PK	IdVendedor	VARCHAR(10)
	NombreCompleto	VARCHAR(80)
	Perfil	VARCHAR(30)
	FechaRegistro	DATETIME

Autor: Macias Anyelyn, Ortiz Juan.

Tabla 14 Definición de los atributos de la tabla hecho facturas.

HechoFacturas		
PK	ID MFACTURA	DECIMAL(22,0)
FK	ID CLIENTE	DECIMAL(22,0)
FK	ID VENDEDOR	VARCHAR(30)
FK	ID TIEMPO	INT
	TOTAL_VENTA	FLOAT
	TOTA_COSTO	FLOAT
	TOTAL_DEVUELTO	FLOAT
	ESTADO	VARCHAR(2)
	VENTA_REAL	FLOAT
	COSTO_REAL	FLOAT
	TOTAL_UTILIDAD	FLOAT
	PORCENTAJE_UTILIDAD	DOUBLE

Autor: Macias Anyelyn, Ortiz Juan.

Tabla 15 Definición de los atributos de la tabla hecho productos.

HechoProductos		
PK	ID	INT
FK	ID DIM TIEMPO	INT
FK	ID DIM PRODUCTO	VARCHAR(25)
FK	ID DIM CATEGORIA	DECIMAL(22,0)
	CANTIDAD	FLOAT
	VALOR_TOTAL	FLOAT
	VENTA_REAL	DOUBLE
	VALOR_DEVUELTO	DOUBLE
	CANTIDAD_DEV	FLOAT
	PORCENTAJE_DEV	DOUBLE
	UTILIDAD_TOTAL	DOUBLE
	COSTO_REAL	DOUBLE
	PORCENTAJE_UTILIDAD	DOUBLE

Autor: Macias Anyelyn, Ortiz Juan.

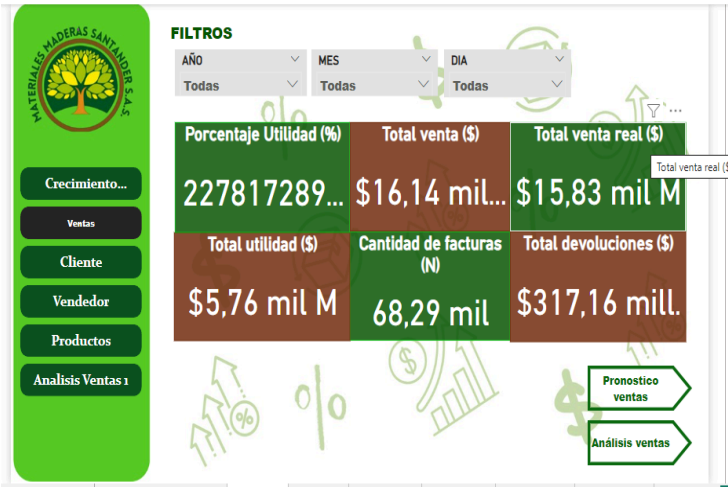
3.1.3.6 Pruebas modulo visualización

Durante el proceso de evaluación del módulo de visualización en Power BI, se realizaron una serie de pruebas para evaluar la integridad, precisión y consistencia de los datos en comparación con la fuente de datos proporcionada por la empresa. Se llevó a cabo una

verificación para asegurarse de que la información presentada en los dashboards se alinee de manera precisa con cada uno de los requisitos iniciales establecidos por la gerente

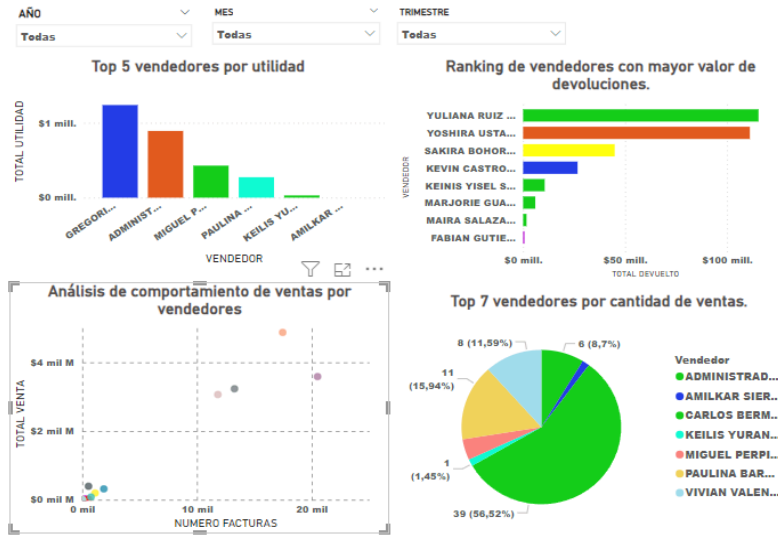
administrativa. Además, se efectuaron ajustes y modificaciones en el diseño de los dashboards con el fin de garantizar que cumplieran con los estándares establecidos, proporcionando así una representación efectiva de los datos clave. Estas pruebas se llevaron a cabo con el objetivo de asegurar que los dashboards fueran una herramienta confiable y precisa para la toma de decisiones en la empresa Materiales Madera Santander.

Ilustración 11 Mockup para el reporte de ventas totales.



Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 12 Mockup para el reporte de vendedores.



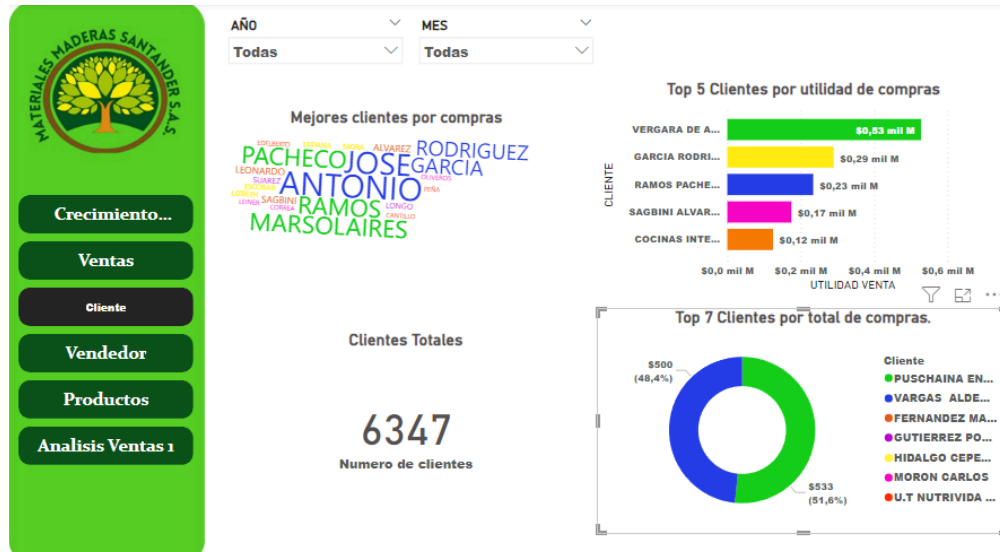
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 13 Mockup para el reporte de productos.



Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 14 Mockup para el reporte de clientes.



Autor: Macias Anyelyn, Ortiz Juan.

3.1.4 Diseño físico

En esta etapa, se lleva a cabo la concreción de la estructura, arquitectura y configuración del sistema, transformándolo desde un modelo de datos lógico en la implementación real de una base de datos relacional.

3.1.4.1 Descripción de los datos fuente

Tabla 16 Características de la base de datos fuente

Sistema gestor de base de datos	Oracle 11 XE
Numero de tablas	124
Peso (MB)	100,8125
# tablas en el área de ventas	15
Peso tablas área de ventas (MB)	36,992

Autor: Macias Anyelyn, Ortiz Juan.

- **Volumetría**

La volumetría de datos se emplea para calcular y determinar el espacio requerido en el motor de una base de datos o en el almacenamiento de un Data Warehouse con el fin de alojar y administrar los datos.

En el contexto de este proyecto, la volumetría se compone de las tablas asociadas a los procesos de ventas y compras de la empresa. La volumetría se describe en términos del tamaño actual, medido en megabytes, que representa la suma de los datos contenidos en las tablas utilizadas. En la siguiente tabla se puede observar la volumetría de los datos.

Tabla 17 Volumetría de los datos fuente que serán procesados.

#	Nombre de la tabla	Descripción	Numero registros	Peso (MG)
1	SubGrupo	Contiene información referente a la categoría a la que pertenece cada producto.		0.0625
2	Producto	Contiene información de los productos que oferta la empresa.	1.664	0.375
3	Usuario	Contiene información de los vendedores que generan las facturas.		0.06
4	Dfactura	Contiene información del detalle de la factura.	158.638	12
5	Tercero	Contiene información general de terceros.	6.404	2
6	Mfactura	Contiene información de la factura.	68.290	13
	TOTAL			27.4975

Autor: Macias Anyelyn, Ortiz Juan.

3.1.4.2 Configuración de la infraestructura para el desarrollo del proyecto

Dada la naturaleza de los datos que serán procesados, se requieren recursos computacionales adecuados para llevar a cabo este proceso. Las tablas 18 y 19 proporcionan una descripción detallada de las características de los recursos tecnológicos






disponibles en el equipo del proyecto, que serán utilizados en el desarrollo de este. En la ilustración 16 se realiza un diseño de la arquitectura propuesta para el proyecto.

Tabla 18 Descripción de las características de los equipos de cómputo disponibles para el desarrollo del proyecto.

Computador portátil	Asus vivobook HP Pavilion Laptop 15
Procesador	Intel i5 12th AMD Ryzen 5 2500U
Tipo de sistema	X64 X64
Memoria RAM	Memoria ram 16gb 8,00 GB
Disco duro	SSD de 500gb HDD de 1t, SSD de 124
Sistema operativo	Windows 11 Home Single Language Windows 11 Home Single Language

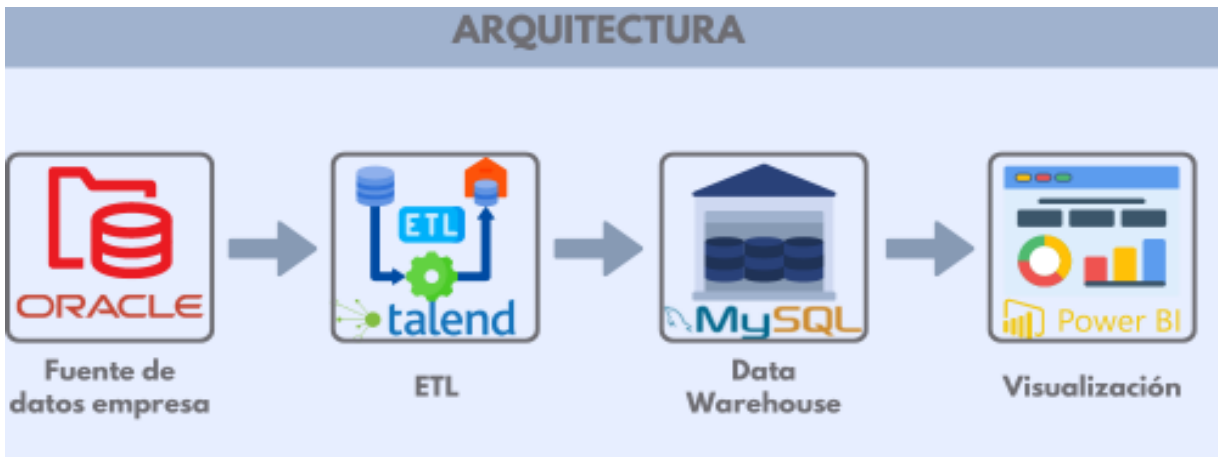
Autor: Macias Anyelyn, Ortiz Juan.

Tabla 19 Descripción y especificación del recurso software necesario para el desarrollo del proyecto.

	Software	Versión
Talend Open Studio for data integration		8.0.1
Power BI		2.121.942.0
Oracle		11
MySql		8.0
R		4.3.2

Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 15 Arquitectura propuesta para el desarrollo del proyecto.



Autor: Macias Anyelyn, Ortiz Juan.

La ilustración 15 representa la arquitectura propuesta para la solución de inteligencia de negocio de la empresa Materiales Maderas Santander. En el primer paso, se emplea Oracle, una base de datos relacional, como el almacén principal de los datos fuente de la empresa. El segundo paso implica el uso de Talend, una poderosa herramienta de ETL, que facilita la extracción, transformación y carga de los datos desde Oracle hacia la bodega de datos. La tercera etapa consiste en la implementación de MySQL, una base de datos que aloja la bodega de datos, donde se organizan y estructuran los datos para su posterior análisis. Por último, en el cuarto paso, se utiliza Power BI, una herramienta de visualización que despliega gráficos y tableros, permitiendo a los directivos de la empresa acceder a información valiosa extraída de la bodega de datos.

3.1.4.3 Modelo de datos físico

En la siguiente sección, se aborda la descripción y representación de las relaciones entre las tablas presentes en la base de datos de origen, las cuales se incorporan para conformar la bodega de datos. Adicionalmente, se ofrecen detalles acerca de los atributos más relevantes que se incorporarán en este procedimiento.

DimCategoriaProducto

Para establecer la dimensión de categoría de producto, se incorporaron las siguientes tablas procedentes de la base de datos de origen.

Tabla 20 Estructura de la tabla sub grupo en la base de datos fuente.

SUBGRUPO		
PK	ID_SUBGRUPO	NUMBER
FK	ID_GRUPO	NUMBER
	DESCRIPCION	VARCHAR2

Autor: Macias Anyelyn, Ortiz Juan.

La tabla SUBGRUPO contiene información de los sub grupos de cada producto de la empresa.

- La llave primaria (PK) es el código para identificar cada SUBGRUPO en la base de datos.
- La llave foránea (FK) indica la relación con la tabla GRUPO el cual contiene información general de los grupos a los que pertenece cada producto de la empresa.
- El atributo DESCRIPCION indica el nombre y detalle del SUGRUPO.

Tabla 21 Estructura de la tabla grupo en la base de datos fuente.

GRUPO		
PK	ID_GRUPO	NUMBER
	DESCRIPCION	VARCHAR2

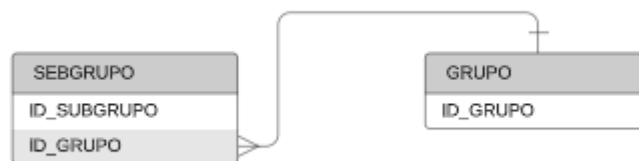
Autor: Macias Anyelyn, Ortiz Juan.

La tabla GRUPO contiene información de los grupos de cada producto de la empresa.

- La llave primaria (PK) es el código para identificar cada GRUPO en la base de datos.
- El atributo DESCRIPCION indica el nombre y detalle del GRUPO.

En la figura se representan las relaciones entre las tablas que conformarán la dimensión categoría de producto.

Ilustración 16 Diagrama relacional de las tablas origen para la creación de la dimensión categoría de producto.



Autor: Macias Anyelyn, Ortiz Juan.

El resultado arroja la tabla, que proporciona información sobre la estructura de la dimensión de categoría de producto "DimCategoriaProducto". Esta dimensión almacena los detalles de las categorías existentes en la empresa, incluyendo su identificador y su nombre.

Tabla 22 Definición de los atributos de la Dimensión Categoría del producto.

DimCategoriaProducto		
PK	IdCategoria	DECIMAL(22,0)
	NombreCategoria	VARCHAR(50)

Autor: Macias Anyelyn, Ortiz Juan.

DimCliente

Para establecer la dimensión cliente, se incorporaron las siguientes tablas procedentes de la base de datos de origen.

Tabla 23 Estructura de la tabla tercero en la base de datos fuente.

TERCERO		
PK	ID TERCERO	NUMBER
	PRIMER_APELLIDO	VARCHAR2
	SEGUNDO_APELLIDO	VARCHAR2
	PRIMER_NOMBRE	VARCHAR2
	SEGUNDO_NOMBRE	VARCHAR2
	RAZON SOCIAL	VARCHAR2
	ES CLIENTE	CHAR
	ES PROVEEDOR	CHAR
	ES EMPLEADO	CHAR
	DIRECCION	VARCHAR2
	TELEFONO	VARCHAR2
	CELULAR	VARCHAR2
	FAX	VARCHAR2
	EMAIL	VARCHAR2
	FECHA_NACIMIENTO	DATE
	FECHA_REGISTRO	DATE

Autor: Macias Anyelyn, Ortiz Juan.

La tabla TERCERO contiene información de los terceros de la empresa (cliente, vendedor y proveedor) en este caso solo se utilizarán los atributos necesarios para cliente.

- La llave primaria (PK) es el código para identificar cada tercero en la base de datos.

- El atributo PRIMER_APELLIDO indica el SEGUNDO apellido del tercero.
- El atributo SEGUNDO_APELLIDO indica el segundo apellido del tercero.
- El atributo PRIMER_NOMBRE indica el primer nombre del tercero.
- El atributo SEGUNDO_NOMBRE indica el segundo nombre del tercero.
- El atributo RAZON_SOCIAL indica el nombre legal de una empresa, sino es el caso indica el nombre completo del tercero.
- El atributo ES_CLIENTE indica si el tercero es cliente.
- El atributo ES_EMPLEADO indica si el tercero es empleado.
- El atributo ES_PROVEEDOR indica si el tercero es proveedor.
- El atributo DIRECCION describe la dirección del domicilio del tercero.
- El atributo TELEFONO describe el número telefónico del tercero.
- El atributo CELULAR describe el número celular del tercero.
- El atributo FAX especifica el fax del del tercero.
- El atributo EMAIL especifica la dirección de correo electrónico del tercero.
- El atributo FECHA_NACIMIENTO describe la fecha de nacimiento del tercero.
- El atributo FECHA_REGISTRO describe la fecha de registro a la base de datos del tercero.

El resultado arroja la tabla, que proporciona información sobre la estructura de la dimensión cliente "DimCliente". Esta dimensión almacena los datos de los clientes en la empresa.

Tabla 24 Definición de los atributos de la dimensión cliente.

DimCliente		
PK	IdCliente	DECIMAL(22,0)
	NombreCompleto	VARCHAR(80)
	FechaRegistro	DATETIME

Autor: Macias Anyelyn, Ortiz Juan.

DimProducto

Para establecer la dimensión producto, se incorporaron las siguientes tablas procedentes de la base de datos de origen.

Tabla 25 Estructura de la tabla producto en la base de datos fuente.

PRODUCTO		
PK	COD_PRODUCTO	NUMBER
FK	ID_UNIDAD_MEDIDA	NUMBER
FK	ID_SUBGRUPO	NUMBER
	DESCRIPCION	VARCHAR2
	REFERENCIA	VARCHAR2
	EXISTENCIA	NUMBER
	PESO	NUMBER
	STOCK_MIN	NUMBER
	STOCK_MAX	NUMBER
	COSTO_UNIT	NUMBER
	PRECIO	NUMBER
	EXIST	NUMBER
	PENDIENTE	NUMBER
	ESTANTE	NUMBER
	BD_REFERENCIA	VARCHAR2
	FECHA_VENCIMIENTO	DATE
	PORC_UTIL	NUMBER
	FECHA_REGISTRO	DATE
	PORC_IVA	NUMBER
	GARANTIA	NUMBER
	PRECIO_NEG	CHAR
	DISPONIBLE	CHAR

Autor: Macias Anyelyn, Ortiz Juan.

La tabla PRODUCTO contiene información de los productos de la empresa.

- La llave primaria (PK) es el código para identificar cada producto en la base de datos.
- Las laves foráneas (FK) indican la relación con las otras tablas: UNIDAD_MEDIDA, SUB_GRUPO.
- El atributo DESCRIPCION indica el nombre y detalle del PRODUCTO.
- El atributo REFERENCIA es un identificador único que se utiliza para distinguir un producto.
- El atributo EXISTENCIA indica la existencia disponible del producto.
- El atributo PESO indica el peso del producto si es necesario.
- El atributo STOCK_MIN indica la cantidad mínima del producto

- El atributo STOCK_MAX indica la cantidad máxima del producto
- El atributo COSTO_UNIT indica gasto total que la empresa incurre para producir o adquirir el producto.
- El atributo PRECIO indica el monto que la empresa cobra a sus clientes por el producto
- El atributo EXIST indica la cantidad existente del producto.
- El atributo PENDIENTE indica la cantidad que se encuentra pendiente del producto.
- El atributo ESTANTE indica la cantidad que se encuentra en estantería del producto.
- El atributo BD_REFERENCIA indica la cantidad que se encuentra pendiente del producto.
- El atributo FECHA_VENCIMIENTO indica la fecha de caducidad del producto.
- El atributo PORC_UTIL indica el porcentaje de utilidad que le queda a a la empresa por la venta del producto.
- El atributo FECHA_REGISTRO indica la fecha de registro en la base de datos del producto.
- El atributo PORC_IVA indica el porcentaje de IVA que se le aplica a ese producto del producto.
- El atributo GARANTIA indica el tiempo de garantía del producto.
- El atributo PRECIO_NEG indica si se negoció otro precio para el producto.
- El atributo DISPONIBLE indica la cantidad disponible del producto.

La tabla UNIDAD_MEDIDA contiene información de la unidad de medida y medida aplicada a los productos de la empresa.

Tabla 26 Estructura de la tabla unidad de medida en la base de datos fuente.

UNIDAD MEDIDA		
PK	ID UNIDAD MEDIDA	VARCHAR2
	DESCRIPCION	VARCHAR2
	FRACCION	CHAR
	COD DIAN	VARCHAR2

Autor: Macias Anyelyn, Ortiz Juan.

- La llave primaria (PK) identificador de la unidad de medida utilizada (ejemplo: bolsa, bulto, botella, etc)
- El atributo DESCRIPCION indica el nombre de la unidad de medida del producto.

- El atributo DESCRIPCION indica el nombre de la unidad de medida del producto.
- El atributo FRACCION indica si la medida del producto lleva una fracción.
- El atributo COD_DIAN indica si la medida del producto lleva una fracción.

La tabla SUBGRUPO contiene información de los sub grupos de cada producto de la empresa.

Tabla 27 Estructura de la tabla sub grupo en la base de datos fuente.

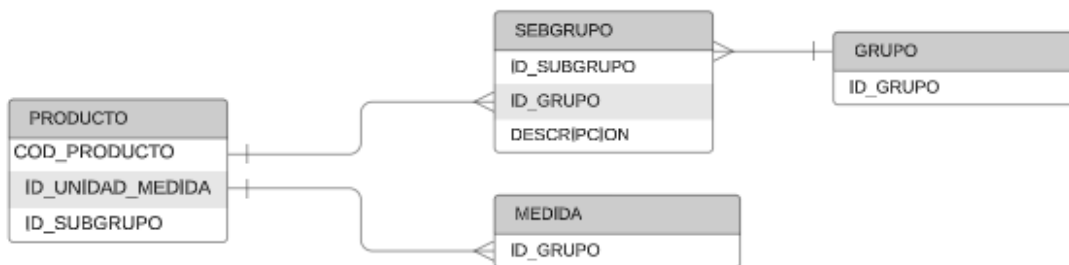
SUBGRUPO		
PK	ID_SUBGRUPO	NUMBER
FK	ID_GRUPO	NUMBER
	DESCRIPCION	VARCHAR2

Autor: Macias Anyelyn, Ortiz Juan.

- La llave primaria (PK) es el código para identificar cada SUBGRUPO en la base de datos.
- La llave foránea (FK) indica la relación con la tabla GRUPO el cual contiene información general de los grupos a los que pertenece cada producto de la empresa.
- El atributo DESCRIPCION indica el nombre y detalle del SUGRUPO.

En la figura 17 se representan las relaciones entre las tablas que conformarán la dimensión producto.

Ilustración 17 Diagrama relacional de las tablas origen para la creación de la dimensión producto.



Autor: Macias Anyelyn, Ortiz Juan.

El resultado arroja la tabla, que proporciona información sobre la estructura de la dimensión producto "DimProducto". Esta dimensión almacena los datos de los productos de la empresa.

Tabla 28 Definición de los atributos de la dimensión producto.

DimProducto		
PK	IdProducto	DECIMAL(22,0)
	NombreCompleto	VARCHAR(80)
	FechaRegistro	DATETIME

Autor: Macias Anyelyn, Ortiz Juan.

DimVendedor

Para establecer la dimensión vendedora, se incorporaron las siguientes tablas procedentes de la base de datos de origen.

Tabla 29 Estructura de la tabla producto en la base de datos fuente.

USUARIO		
PK	COD USUARIO	VARCHAR2
	NOM USUARIO	VARCHAR2
	PERFIL	VARCHAR2
	FECHA REGISTRO	DATE
	EMAIL	VARCHAR2

Autor: Macias Anyelyn, Ortiz Juan.

La tabla USUARIO contiene información sobre los usuarios que realizan las facturas en la empresa.

- La llave primaria (PK) es el código para identificar cada usuario en la base de datos.
- El atributo NOM_USUARIO indica el nombre completo del usuario.
- El atributo PERFIL indica el perfil del usuario.
- El atributo EMAIL especifica la dirección de correo electrónico del usuario.
- El atributo FECHA_REGISTRO describe la fecha de registro a la base de datos del usuario.

El resultado arroja la tabla, que proporciona información sobre la estructura de la dimensión vendedor "DimVendedor". Esta dimensión almacena los datos de los vendedores de la empresa.

Tabla 30 Definición de los atributos de la dimensión vendedor.

DimVendedor		
PK	IdVendedor	VARCHAR(10)
	NombreCompleto	VARCHAR(100)
	Perfil	VARCHAR(50)
	FechaRegistro	DATETIME

Autor: Macias Anyelyn, Ortiz Juan.

Hechos_facturas

Para crear la tabla de hecho facturas, se combinan las tablas de la base de datos de origen junto con las dimensiones que han sido previamente construidas en el data warehouse.

Tabla 31 Estructura de la tabla devolución en la base de datos fuente.

MDEVOLUCION		
PK	ID_MDEVOLUCION	NUMBER
FK	ID_MFACTURA	NUMBER
	COD_USUARIO	VARCHAR2
	NUM_DEVOLUCION	VARCHAR2
	FECHA	DATE
	PARCIAL	NUMBER
	SUB_TOTAL	NUMBER
	DESCUENTO	NUMBER
	GRAVADO	NUMBER
	EXENTO	NUMBER
	IVA	NUMBER
	TOTAL	NUMBER
	COSTO	NUMBER
	OBSERVACIÓN	VARCHAR2
	FECHAREGISTRO	DATE

Autor: Macias Anyelyn, Ortiz Juan.

La tabla MDEVOLUCION contiene información sobre las devoluciones realizadas a las facturas en la empresa.

- La llave primaria (PK) es el código para identificar la devolución realizada.
- La llave foránea (FK) indica la relación con la tabla MFACTURA el cual contiene información general de las facturas realizadas en la empresa.
- El atributo COD_USUARIO indica el código del vendedor que recibe la devolución.
- El atributo FECHA indica la fecha en la que se realiza la devolución.
- El atributo PARCIAL indica suma parcial de los precios de los productos
- El atributo SUB_TOTAL indica la suma parcial de los precios de los productos.
- El atributo DESCUENTO indica el valor del descuento realizado en la devolución si es el caso.
- El atributo GRAVADO indica el valor del impuesto.

- El atributo EXENTO indica el valor sino está exento de impuesto.
- El atributo IVA indica el valor del IVA aplicado.
- El atributo COSTO indica la suma del costo de la devolución.

- El atributo OBSERVACIÓN indica el comentario de la devolución.
- El atributo FECHA_REGISTRO describe la fecha de registro la devolución.

La tabla MFACTURA contiene información general sobre las facturas realizadas en la empresa.

Tabla 32 Estructura de la tabla devolución en la base de datos fuente.

MFACTURA		
PK	ID_MFACTURA	NUMBER
FK	ID_TERCERO	NUMBER
	ID_USUARIO	VARCHAR2
	TIPO_FACTURA	VARCHAR2
	NUM_FACTURA	VARCHAR2
	FECHA_FACTURA	DATE
	FECHA_PAGO	DATE
	PARCIAL	NUMBER
	SUBTOTAL	NUMBER
	DESCUENTO	NUMBER
	GRAVADO	NUMBER
	EXENTO	NUMBER
	IVA	NUMBER
	TOTAL	NUMBER
	DEVOLUCION	NUMBER
	PEDOENTE	NUMBER
	COSTO	NUMBER
	ABONO	NUMBER

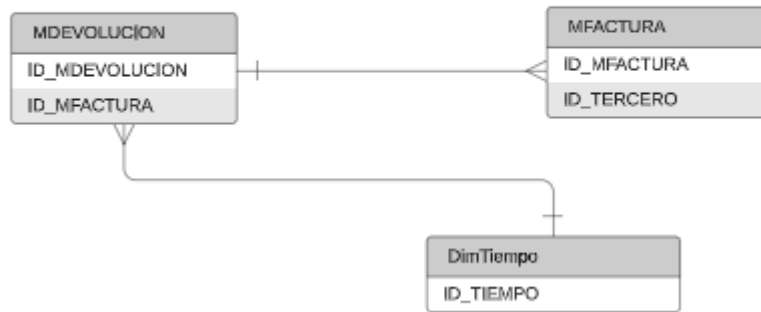
	INTERES	NUMBER
	ESTADO	CHAR
	DEVOLUCION_IVA	NUMBER

Autor: Macias Anyelyn, Ortiz Juan.

- La llave primaria (PK) es el código para identificar la factura realizada.
- La llave foránea (FK) indica la relación con la tabla TERCERO la cual contiene información general sobre los terceros.
- El atributo ID_USUARIO indica el código del vendedor que recibe la devolución.
- El atributo TIPO_FACTURA indica la categorización clasificación específica de la factura.
- El atributo NUM_FACTURA identificador de la factura
- El atributo FECHA_FACTURA indica la fecha en la que se realizó la factura.
- El atributo FECHA_PAGO indica la fecha en la que se realizó el pago de la factura.
- El atributo PARCIAL indica suma parcial de los precios de los productos.
- El atributo SUB_TOTAL indica la suma parcial de los precios de los productos.
- El atributo DESCUENTO indica el valor del descuento realizado en la factura si es el caso.
- El atributo GRAVADO indica el valor del impuesto.
- El atributo EXENTO indica el valor sino está exento de impuesto.
- El atributo IVA indica el valor del IVA aplicado.
- El atributo TOTAL indica el valor total de la factura.
- El atributo DEVOLUCIÓN indica el valor de la devolución.
- El atributo PEDIENTE indica la cantidad pendiente de la factura.
- El atributo COSTO indica la suma del costo de la devolución.
- El atributo ABONO indica el abono realizado a la factura.
- El atributo INTERES indica el valor del interés de la factura.
- El atributo ESTADO indica el estado en que se encuentra la factura.
- El atributo DEVOLUCION_IVA indica el valor devuelto del IVA.

Además de las tablas mencionadas anteriormente, se incorporó la Dimensión Tiempo para la creación de la tabla de hechos facturas. En la siguiente figura se pueden identificar las relaciones entre las tablas de las cuales se extraerán los atributos para la tabla de hechos facturas.

Ilustración 18 Diagrama relacional de las tablas origen para la creación de la tabla de hecho factura.



Autor: Macias Anyelyn, Ortiz Juan.

El resultado arroja la tabla, que proporciona información sobre la estructura de la tabla de hecho "Hecho_facturas". Esta tabla almacena los KPI de las facturas.

Tabla 33 Definición de los atributos de la tabla hecho facturas.

HechoFacturas		
PK	ID MFACTURA	DECIMAL(22,0)
FK	ID CLIENTE	DECIMAL(65,0)
FK	ID VENDEDOR	VARCHAR(100)
FK	ID TIEMPO	INT
	TOTAL VENTA	FLOAT
	TOTA COSTO	FLOAT
	TOTAL DEVUELTO	FLOAT
	ESTADO	VARCHAR(2)
	VENTA_REAL	FLOAT
	COSTO_REAL	FLOAT
	TOTAL UTILIDAD	FLOAT
	PORCENTAJE UTILIDAD	DOUBLE

Autor: Macias Anyelyn, Ortiz Juan.

Hechos_productos

Para crear la tabla de hecho productos, se combinan las tablas de la base de datos de origen junto con las dimensiones que han sido previamente construidas en el data warehouse.

Tabla 34 Estructura de la tabla devolución en la base de datos fuente.

PRODUCTO BODEGA		
PK	COD BD PROD	VARCHAR2
FK	COD BODEGA	VARCHAR2
FK	COD PRODUCTO	VARCHAR2
	EXISTENCIA	NUMBER

Autor: Macias Anyelyn, Ortiz Juan.

La tabla PRODUCTO_BODEGA contiene información sobre los productos que se encuentran en las bodegas de la empresa.

- La llave primaria (PK) es el código para identificar el producto y en que bodega se encuentra.
- La llave foránea (FK) indica la relación con la tabla BODEGA el cual contiene información general sobre las bodegas de la empresa y la tabla PRODUCTO la cual contiene información de los productos que oferta la empresa.
- El atributo EXISTENCIA indica la cantidad existente de un producto determinado en bodega.

La tabla BODEGA contiene información sobre las bodegas con las que cuenta la empresa.

Tabla 35 Estructura de la tabla bodega en la base de datos fuente.

BODEGA		
PK	COD BODEGA	VARCHAR2
	DESCRIPCION	VARCHAR2

Autor: Macias Anyelyn, Ortiz Juan.

- La llave primaria (PK) es el código para identificar la bodega.
- El atributo DESCRIPCION indica el nombre de la bodega.

La tabla DFACTURA contiene información del detalle de las facturas realizadas en la empresa.

Tabla 36 Estructura de la tabla dfactura en la base de datos fuente.

DFACTURA		
PK	ID DFACTURA	NUMBER
FK	COD BD PROD	VARCHAR2
FK	ID MFACTURA	NUMBER
	PORC IVA	NUMBER
	CANTIDAD	NUMBER
	VALOR UNIT	NUMBER
	PORC DCTO	NUMBER
	VALOR TOTAL	NUMBER
	COSTO_TOTAL	NUMBER
	COSTO_UNITARIO	NUMBER
	SERIAL	VARCHAR2

	CANTIDAD PEND	NUMBER
	CANTIDAD DEV	NUMBER
	SUBTOTAL	NUMBER

Autor: Macias Anyelyn, Ortiz Juan.

- La llave primaria (PK) es el código para identificar el detalle de factura.
- La llave foránea (FK) indica la relación con la tabla BODEGA PRODUCTO la cual contiene información sobre los productos que se encuentran en las bodegas de la empresa y la tabla MFACTURA la cual contiene información general sobre las facturas realizadas en la empresa.

Tabla 37 Estructura de la tabla devolución en la base de datos fuente.

MFACTURA		
PK	ID MFACTURA	NUMBER
FK	ID_TERCERO	NUMBER
	ID_USUARIO	VARCHAR2
	TIPO_FACTURA	VARCHAR2
	NUM_FACTURA	VARCHAR2
	FECHA_FACTURA	DATE
	FECHA_PAGO	DATE
	PARCIAL	NUMBER
	SUBTOTAL	NUMBER
	DESCUENTO	NUMBER
	GRAVADO	NUMBER
	EXENTO	NUMBER
	IVA	NUMBER
	TOTAL	NUMBER
	DEVOLUCION	NUMBER
	PEDOENTE	NUMBER
	COSTO	NUMBER
	ABONO	NUMBER
	INTERES	NUMBER
	ESTADO	CHAR
	DEVOLUCION IVA	NUMBER

Autor: Macias Anyelyn, Ortiz Juan.

La tabla MFACTURA contiene información general sobre las facturas realizadas en la empresa.

- La llave primaria (PK) es el código para identificar la factura realizada.
- La llave foránea (FK) indica la relación con la tabla TERCERO la cual contiene información general sobre los terceros.
- El atributo ID_USUARIO indica el código del vendedor que recibe la devolución.

- El atributo TIPO_FACTURA indica la categorización clasificación específica de la factura.
- El atributo NUM_FACTURA identificador de la factura
- El atributo FECHA_FACTURA indica la fecha en la que se realizó la factura.
- El atributo FECHA_PAGO indica la fecha en la que se realizó el pago de la factura.
- El atributo PARCIAL indica suma parcial de los precios de los productos.
- El atributo SUB_TOTAL indica la suma parcia de los precios de los productos.
- El atributo DESCUENTO indica el valor del descuento realizado en la factura si es el caso.
- El atributo GRAVADO indica el valor del impuesto.
- El atributo EXENTO indica el valor sino está exento de impuesto.
- El atributo IVA indica el valor del IVA aplicado.
- El atributo TOTAL indica el valor total de la factura.
- El atributo DEVOLUCIÓN indica el valor de la devolución.
- El atributo PEDIENTE indica la cantidad pendiente de la factura.
- El atributo COSTO indica la suma del costo de la devolución.
- El atributo ABONO indica el abono realizado a la factura.
- El atributo INTERES indica el valor del interés de la factura.
- El atributo ESTADO indica el estado en que se encuentra la factura.
- El atributo DEVOLUCION_IVA indica el valor devuelto del IVA.

Tabla 38 Estructura de la tabla producto en la base de datos fuente.

PRODUCTO		
PK	COD PRODUCTO	NUMBER

FK	ID UNIDAD MEDIDA	NUMBER
FK	ID_SUBGRUPO	NUMBER
	DESCRIPCION	VARCHAR2
	REFERENCIA	VARCHAR2
	EXISTENCIA	NUMBER
	PESO	NUMBER
	STOCK_MIN	NUMBER
	STOCK_MAX	NUMBER
	COSTO_UNIT	NUMBER
	PRECIO	NUMBER
	EXIST	NUMBER
	PENDIENTE	NUMBER
	ESTANTE	NUMBER
	BD_REFERENCIA	VARCHAR2
	FECHA_VENCIMIENTO	DATE
	PORC_UTIL	NUMBER
	FECHA_REGISTRO	DATE
	PORC_IVA	NUMBER
	GARANTIA	NUMBER
	PRECIO_NEG	CHAR
	DISPONIBLE	CHAR

Autor: Macias Anyelyn, Ortiz Juan.

La tabla PRODUCTO contiene información de los productos de la empresa.

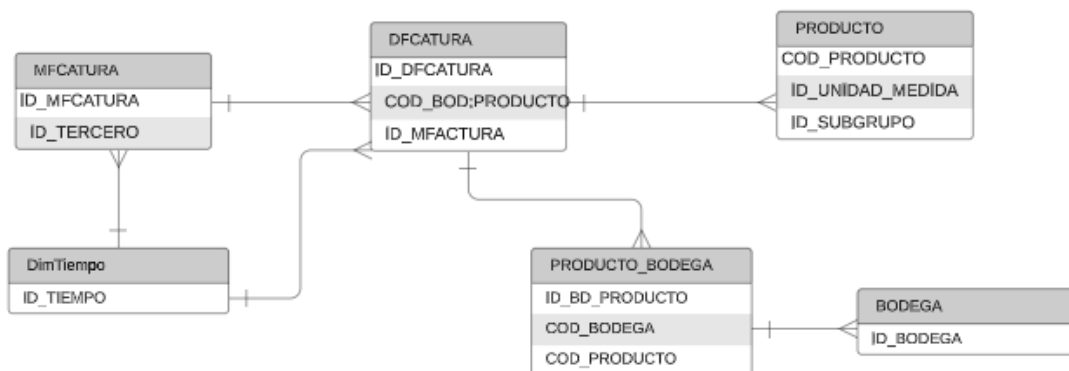
- La llave primaria (PK) es el código para identificar cada producto en la base de datos.
- Las laves foráneas (FK) indican la relación con las otras tablas: UNIDAD_MEDIDA, SUB_GRUPO.
- El atributo DESCRIPCION indica el nombre y detalle del PRODUCTO.
- El atributo REFERENCIA es un identificador único que se utiliza para distinguir un producto.
- El atributo EXISTENCIA indica la existencia disponible del producto.
- El atributo PESO indica el peso del producto si es necesario.
- El atributo STOCK_MIN indica la cantidad mínima del producto
- El atributo STOCK_MAX indica la cantidad máxima del producto
- El atributo COSTO_UNIT indica gasto total que la empresa incurre para producir o adquirir el producto.
- El atributo PRECIO indica el monto que la empresa cobra a sus clientes por el producto
- El atributo EXIST indica la cantidad existente del producto.
- El atributo PENDIENTE indica la cantidad que se encuentra pendiente del producto.
- El atributo ESTANTE indica la cantidad que se encuentra en estantería del producto.

- El atributo BD_REFERENCIA indica la cantidad que se encuentra pendiente del producto.
- El atributo FECHA_VENCIMIENTO indica la fecha de caducidad del producto.
- El atributo PORC_UTIL indica el porcentaje de utilidad que le queda a a la empresa por la venta del producto.
- El atributo FECHA_REGISTRO indica la fecha de registro en la base de datos del producto.
- El atributo PORC_IVA indica el porcentaje de IVA que se le aplica a ese producto del producto.

- El atributo GARANTIA indica el tiempo de garantía del producto.
- El atributo PRECIO_NEG indica si se negoció otro precio para el producto.
- El atributo DISPONIBLE indica la cantidad disponible del producto.

Además de las tablas mencionadas anteriormente, se incorporó la Dimensión Tiempo para la creación de la tabla de hechos productos. En la siguiente figura se pueden identificar las relaciones entre las tablas de las cuales se extraerán los atributos para la tabla de hechos productos.

Ilustración 19 Diagrama relacional de las tablas origen para la creación de la tabla de hecho factura.



Autor: Macias Anyelyn, Ortiz Juan.

El resultado arroja la tabla, que proporciona información sobre la estructura de la tabla de hecho "Hecho_productos ". Esta tabla almacena los KPI de los productos.

Tabla 39 Definición de los atributos de la tabla hecho facturas.

HechoProductos		
PK	ID	INT
FK	ID DIM TIEMPO	INT
FK	ID DIM PRODUCTO	VARCHAR(25)
FK	ID DIM CATEGORIA	DECIMAL(22,0)
	CANTIDAD	FLOAT
	VALOR TOTAL	FLOAT
	VENTA REAL	DOUBLE
	VALOR DEVUELTO	DOUBLE
	CANTIDAD DEV	FLOAT
	PORCENTAJE DEV	DOUBLE
	UTILIDAD TOTAL	DOUBLE
	COSTO REAL	DOUBLE
	PORCENTAJE UTILIDAD	DOUBLE

Autor: Macias Anyelyn, Ortiz Juan.

3.1.5 Diseño e implementación del subsistema ETL

Una vez que se ha completado la fase de diseño de la estructura de la bodega de datos y se han definido las tablas de la base de datos fuente junto con sus atributos correspondientes para la creación de dimensiones y tablas de hecho, el siguiente paso implica avanzar en el desarrollo de los procesos de extracción, transformación y carga (ETL). Para llevar a cabo esta tarea, se utiliza la herramienta de integración de datos, en este caso, Talend Open Studio en su versión 8.0.1, que incluye una serie de capacidades y funcionalidades específicas para esta finalidad.

En la tabla 40 se detallan todos los componentes empleados junto con sus respectivas definiciones y su aplicación en el proceso de desarrollo del ETL.

Tabla 40 Información componentes utilizados.

Componente	Información	Uso en el proyecto
Base de datos		
tDBInput	Permite extraer datos de una base de datos relacional y cargar estos datos en un flujo de trabajo de integración. Básicamente, se encarga de la lectura de datos desde una base de datos.	Este componente permitió extraer de la base de datos las siguientes tablas, para su posterior transformación. <ul style="list-style-type: none"> ● SUBGRUPO ● TERCEO ● PRODUCTO ● USUARIO ● MDEVOLUCION ● MFACTURA ● Dimtiempo ● PRODUCTO_BODEGA ● DFACTURA
tDBOutput	Permite escribir, cargar, actualizar y eliminar datos en una base de datos desde un flujo de datos dentro de un proceso de integración de datos. Su	Este componente permitió rellenar las siguientes tablas de hechos y dimensiones en el almacén de datos. <ul style="list-style-type: none"> ● Dimcategoriaproducto

	principal propósito es cargar o insertar datos en una tabla de una base de datos específica.	<ul style="list-style-type: none"> • Dimcliente • Dimproducto • Dimtiempo • Dimvendedor • Heho_facturas • Hecho_productos
tDBConnection	Permite definir y configurar una conexión de base de datos reutilizable que se puede utilizar en múltiples componentes dentro de un proyecto.	Este componente se empleó en todos los trabajos del proyecto para establecer conexiones con bases de datos MySQL y Oracle, tanto para los datos de origen de Oracle como para el almacén de datos de MySQL.
tDBCclose	Permite cerrar una conexión de base de datos de forma explícita, lo que significa que se fuerza el cierre de la conexión antes de que el flujo de trabajo termine su ejecución.	Este componente se empleó en todos los trabajos del proyecto para cerrar las conexiones con bases de datos MySQL y Oracle, tanto para los datos de origen de Oracle como para el almacén de datos de MySQL.
tDBCommit	Permite confirmar, asegurar que los cambios realizados en la base de datos dentro de una transacción sean exitosamente guardados de manera permanente.	Este componente posibilitó la ejecución de confirmaciones (commit) en la base de datos para cada uno de los procesos en todos los trabajos, siempre y cuando fueran exitosos.
tDBRollback	Permite revertir o deshacer una transacción de base de datos en caso de que ocurra un error o una condición que requiera la anulación de los cambios realizados dentro de la transacción.	Este componente permitió la cancelación de los procesos de los trabajos en caso de que se produjera un error en la transacción de la base de datos.
Procesamiento		
tFilterRow	Permitirte definir reglas que determinen qué registros deben pasar a través del flujo de trabajo y cuáles deben ser descartados o rechazados.	Este componente permitió la filtración de ciertos datos antes de su procesamiento y su posterior incorporación a la dimensión de clientes.
tAggregateRow	Permite realizar operaciones de agregación en los datos, lo que significa que puedes calcular sumas, promedios, conteos y otras funciones de agregación en un conjunto de datos. Este componente es conveniente cuando necesitas resumir o agrupar datos de acuerdo con ciertas condiciones.	Este componente posibilitó la agrupación por factura en la tabla "hecho_facturas" y por productos en la tabla "hecho_productos". Además, permitió la realización de sumas y conteos en los datos.
tSortRow	Permite organizar los datos de entrada en un orden específico, ya sea ascendente o descendente, según los criterios de clasificación que se definan.	Este componente posibilitó la organización de los datos en orden descendente antes de enviarlos a la tabla hechos_productos.
tJavaRow	Permite ejecutar código Java personalizado a nivel de fila en un flujo de trabajo. Su función principal realizar operaciones personalizadas en los datos.	Este componente posibilitó la generación del identificador (ID) para la tabla "hecho_productos".
tMap	Permite transformar datos dentro de un proceso de integración de datos. Su función principal es mapear y transformar los datos de una fuente a una estructura de destino, lo que permite realizar cambios en los datos, filtrar información, combinar datos de	Este componente fue de gran relevancia, ya que posibilitó la transformación de los datos de entrada y salida, creando así un flujo de extracción, transformación y carga (ETL) en cada uno de los trabajos del proyecto.

	múltiples fuentes y realizar otras transformaciones necesarias antes de cargarlos en una base de datos.	
Miselánea		
tRowGenerator	Permite generar la cantidad de filas de datos indicada.	Este componente permitió la generación de datos de días dentro de un intervalo de tiempo específico con el propósito de llenar la dimensión tiempo.
Orquestación		
tPrejob	Permite definir y ejecutar tareas específicas que deben llevarse a cabo antes de que comience la ejecución del trabajo.	Este componente permitió la apertura de la conexión a la base de datos fuente, Oracle y MySQL, para la bodega de datos de destino, antes de iniciar el trabajo.
tPosjob	Permite definir y ejecutar tareas después de que el trabajo principal haya finalizado su ejecución.	Este componente permitió cerrar la conexión a la base de datos fuente, Oracle y MySQL, para la bodega de datos de destino, después de terminar el trabajo.

Autor: Macias Anyelyn, Ortiz Juan.

A continuación, se presentan los resultados de los trabajos realizados en la herramienta de integración, los cuales se diseñaron para ejecutar los procesos ETL.

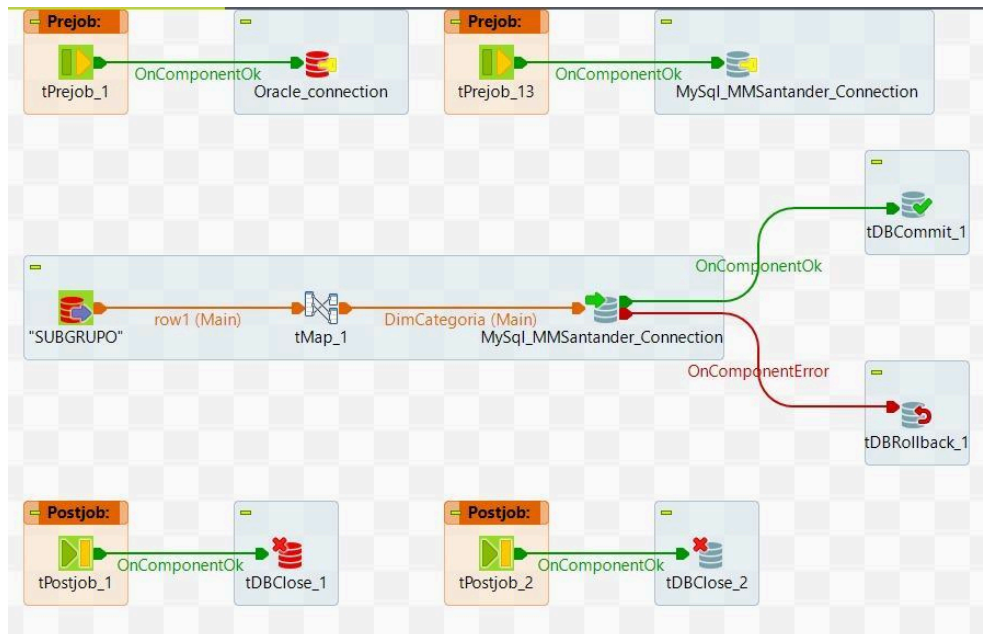
La primera fase comprendió el establecimiento de una conexión con la base de datos fuente alojada en el motor de base de datos Oracle 11. Luego, se extrajeron las tablas necesarias para la construcción del almacén de datos.

Posteriormente, se procedió a crear cada uno de los trabajos en los que se llevaron a cabo los procesos ETL. Se comenzó con las dimensiones, como se describe a continuación.

Job Dimension_Categoria

En la ilustración 20, se puede observar en detalle el job creado en Talend, en el cual se realiza un proceso ETL para la construcción de la dimensión categoría de producto.

Ilustración 20 Creación del Job Dimensión Categoría de producto.



Autor: Macias Anyelyn, Ortiz Juan.

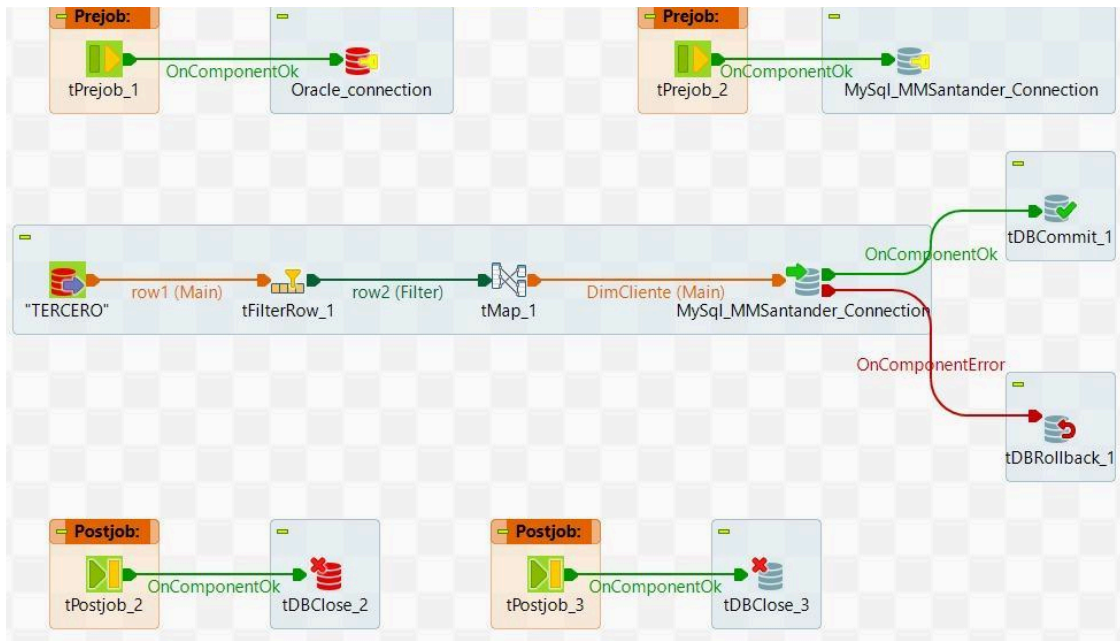
Inicialmente, se abre la conexión con la base de datos fuente y la base de datos de destino. A continuación, se inicia la extracción de tablas desde la base de datos fuente, lo que da paso a un proceso ETL. En este proceso, se extraen datos relacionados con los grupos y familias de productos, los cuales son transformados en la categoría correspondiente. Posteriormente, estos datos procesados se cargan en la dimensión de categoría de producto. Luego de completar este proceso, se lleva a cabo la confirmación (commit) en la base de datos si todos los procesos resultaron exitosos. En caso contrario, se procede a la cancelación de los procesos mediante un rollback.

Finalmente, se procede al cierre de la conexión tanto con la base de datos fuente como con la base de datos de destino.

Job Dimensión_Cliente

En la ilustración 21, se puede observar en detalle el job creado en Talend, en el cual se realiza un proceso ETL para la construcción de la dimensión cliente.

Ilustración 21 Creación del Job Dimensión Cliente.



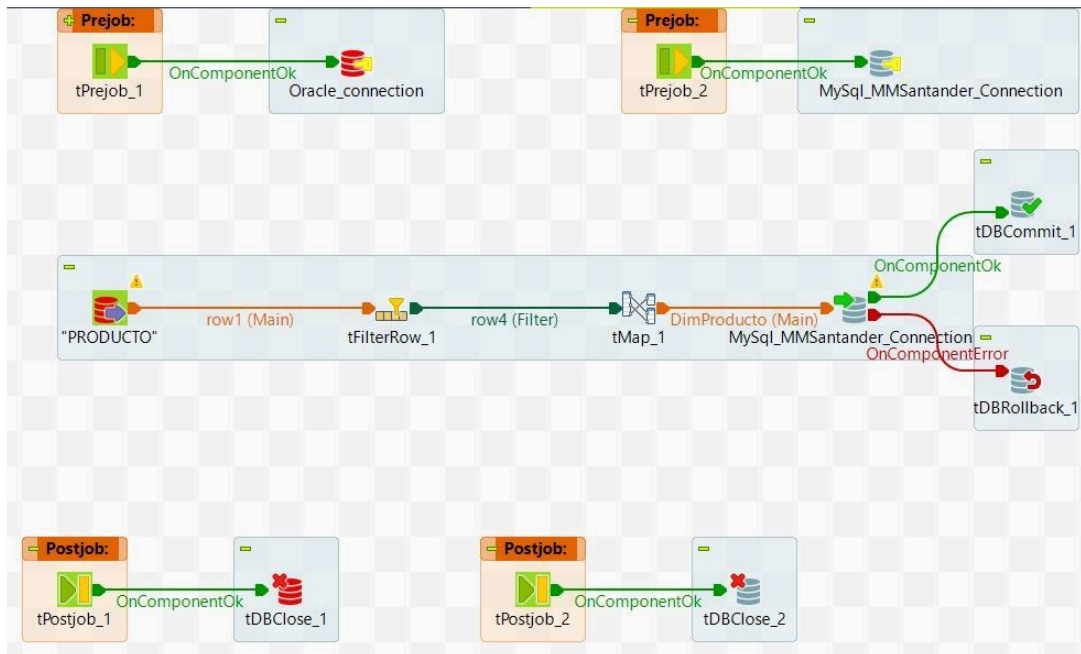
Autor: Macias Anyelyn, Ortiz Juan.

Inicialmente, se abre la conexión con la base de datos fuente y la base de datos de destino. A continuación, se inicia con la extracción de las tablas de la base de datos fuente, seguida de una fase de filtrado de datos. Luego, se ejecuta un proceso ETL para la extracción de datos correspondientes a terceros que se relacionan con los clientes, lo que permite crear la dimensión. Finalmente, se lleva a cabo la carga de los datos ya procesados en la dimensión de cliente. Después de completar este proceso, se efectúa la confirmación (commit) en la base de datos si todos los procesos resultaron exitosos. En caso de algún fallo, se procede a la cancelación de los procesos mediante el rollback. Finalmente, se procede al cierre de la conexión tanto con la base de datos fuente como con la base de datos de destino.

Job Dimension_Producto

En la ilustración 22, se puede observar en detalle el job creado en Talend, en el cual se realiza un proceso ETL para la construcción de la dimensión producto.

Ilustración 22 Creación del Job Dimensión Producto.



Autor: Macias Anyelyn, Ortiz Juan.

Inicialmente, se abre la conexión con la base de datos fuente y la base de datos de destino. A continuación, se inicia con la extracción de las tablas de la base de datos fuente, seguida de una fase de filtrado de datos. Posteriormente, se ejecuta un proceso ETL para la obtención de los datos necesarios con el fin de crear la dimensión de producto. Finalmente, se procede a la carga de los datos ya procesados en la mencionada dimensión. Luego de completar este proceso, se realiza la confirmación (commit) en la base de datos si todos los procesos fueron exitosos. En caso de algún fallo, se procede a la cancelación de los procesos mediante el rollback. Finalmente, se procede al cierre de la conexión tanto con la base de datos fuente como con la base de datos de destino.

Job Dimension_Tiempo

A continuación, se desarrollan diversas funciones en el lenguaje Java, los cuales serán empleados para manipular fechas en la construcción de esta dimensión. Las ilustraciones 23 y 24 muestran el código de las rutinas creadas.

Ilustración 23 Código fuente en java para la creación de funciones para la dimensión tiempo.

```
public static Date dateFromDays(Integer numDays) {  
  
    Calendar cal = Calendar.getInstance();  
    cal.set(2018, 06, 06);  
    cal.add(Calendar.DATE, numDays-1);  
    return (Date) cal.getTime();  
  
}  
  
public static String convertirSemanaALetras(int numeroSemana) {  
    String[] unidades = {"Cero", "Uno", "Dos", "Tres", "Cuatro", "Cinco", "Seis", "Siete", "Ocho", "Nueve"};  
    String[] decenas = {"Diez", "Veinte", "Treinta", "Cuarenta", "Cincuenta"};  
    String resultado = "";  
  
    if (numeroSemana < 10) {  
        resultado = unidades[numeroSemana];  
    } else if (numeroSemana >= 10 && numeroSemana < 60) {  
        if (numeroSemana % 10 == 0) {  
            resultado = decenas[numeroSemana / 10 - 1];  
        } else {  
            resultado = decenas[numeroSemana / 10 - 1] + " y " + unidades[numeroSemana % 10];  
        }  
    }  
  
    return resultado;  
}
```

Autor: Macias Anyelyn, Ortiz Juan.

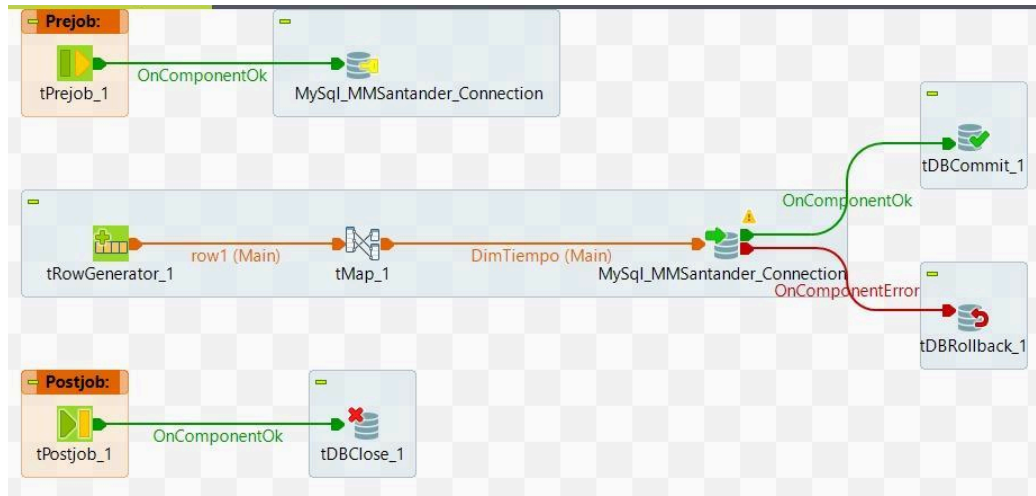
Ilustración 24 Código fuente en java para la creación de funciones para la dimensión tiempo.

```
public static String convertirNumero(int numero) {  
    if (numero < 2017 || numero > 2024) {  
        return "Número fuera de rango";  
    }  
  
    String[] unidades = {  
        "", "uno", "dos", "tres", "cuatro", "cinco",  
        "seis", "siete", "ocho", "nueve"  
    };  
  
    String[] decenas = {  
        "", "diez", "veinte", "treinta", "cuarenta", "cincuenta",  
        "sesenta", "setenta", "ochenta", "noventa"  
    };  
  
    String[] especiales = {  
        "diez", "once", "doce", "trece", "catorce", "quince",  
        "dieciséis", "diecisiete", "dieciocho", "diecinueve"  
    };  
  
    String[] miles = {  
        "", "dos mil", "tres mil"  
    };  
  
    int millares = (numero / 1000) - 1;  
    int centenas = (numero % 1000) / 100;  
    int decenaUnidad = numero % 100;  
  
    String resultado = miles[millares];  
  
    if (centenas > 0) {  
        resultado += " " + unidades[centenas] + " cientos";  
    }  
  
    if (decenaUnidad >= 10 && decenaUnidad <= 19) {  
        resultado += " " + especiales[decenaUnidad - 10];  
    } else {  
        int decena = decenaUnidad / 10;  
        int unidad = decenaUnidad % 10;  
  
        if (decena > 0) {  
            resultado += " " + decenas[decena];  
        }  
  
        if (unidad > 0) {  
            resultado += " y " + unidades[unidad];  
        }  
    }  
  
    return resultado;  
}
```

Autor: Macias Anyelyn, Ortiz Juan.

En la ilustración 25, se presenta en detalle el job creado en Talend, donde se ejecuta un proceso ETL para la generación de la dimensión tiempo.

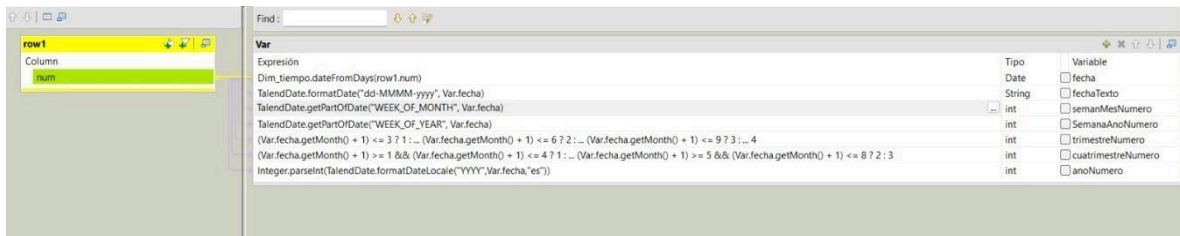
Ilustración 25 Creación del Job dimensión tiempo.



Autor: Macias Anyelyn, Ortiz Juan.

Inicialmente, se abre la conexión con la base de datos fuente y la base de datos de destino. Seguidamente, se generan días desde la fecha inicial hasta la fecha final. A continuación, se inicia un proceso ETL para la creación de las fechas, aprovechando tanto las funciones proporcionadas por Talend como las desarrolladas internamente.

Ilustración 26 ETL dimensión Tiempo



Autor: Macias Anyelyn, Ortiz Juan.

Inicialmente, se abre la conexión con la base de datos fuente y la base de datos de destino.

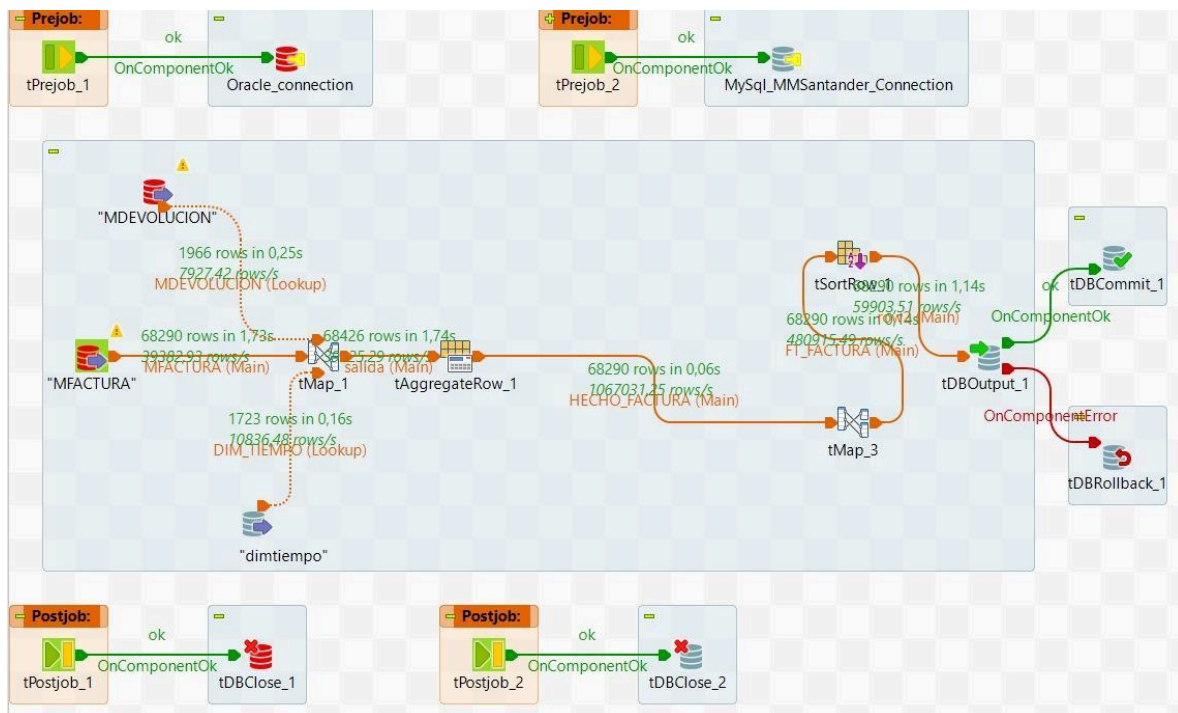
A continuación, se inicia con la extracción de las tablas de la base de datos fuente. Posteriormente, se ejecuta un proceso ETL para obtener los datos necesarios con el fin de crear la dimensión de vendedor. Finalmente, se procede a cargar los datos ya procesados en dicha dimensión. Luego de completar este proceso, se realiza la confirmación (commit) en la base de datos si todos los procesos resultaron exitosos. En caso de algún fallo, se procede a cancelar los procesos mediante el rollback.

Finalmente, se procede al cierre de la conexión tanto con la base de datos fuente como con la base de datos de destino.

Job Hecho_facturas

En la ilustración 28, se aprecia detalladamente el trabajo creado en Talend, donde se llevan a cabo varios procesos y agrupaciones con el propósito de construir la tabla de hechos de facturas.

Ilustración 28 Creación del Job tabla hecho facturas.



Autor: Macias Anyelyn, Ortiz Juan.

Inicialmente, se abre la conexión con la base de datos fuente y la base de datos de destino.

A continuación, se inicia el proceso con la extracción de las tablas desde la base de datos fuente. Posteriormente, se ejecuta un proceso ETL en el que se combinan las tablas extraídas. Durante esta etapa, se realizan transformaciones que implican el uso de funciones específicas de Talend para las métricas, y se establecen los atributos de la tabla de hechos, asignándoles los valores correspondientes a las métricas.

En la ilustración 29 se puede observar el proceso de transformación de los datos.

Ilustración 29 Proceso ETL tabla hecho productos.

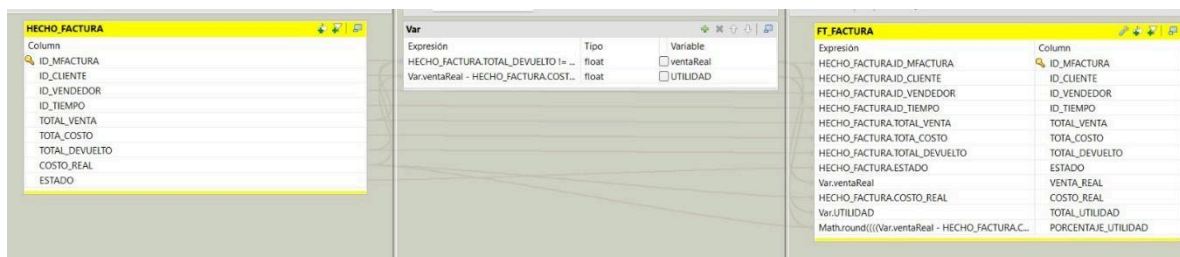


Autor: Macias Anyelyn, Ortiz Juan.

Después de esta etapa, los datos se transmiten a un componente que facilita operaciones de agregación, como la suma del costo real, costo total de los productos, entre otros. Una vez que se han calculado estos valores, los datos se asignan a los atributos correspondientes.

En la ilustración 30 se puede observar el proceso descrito anteriormente.

Ilustración 30 Proceso ETL tabla hecho productos.



Autor: Macias Anyelyn, Ortiz Juan.

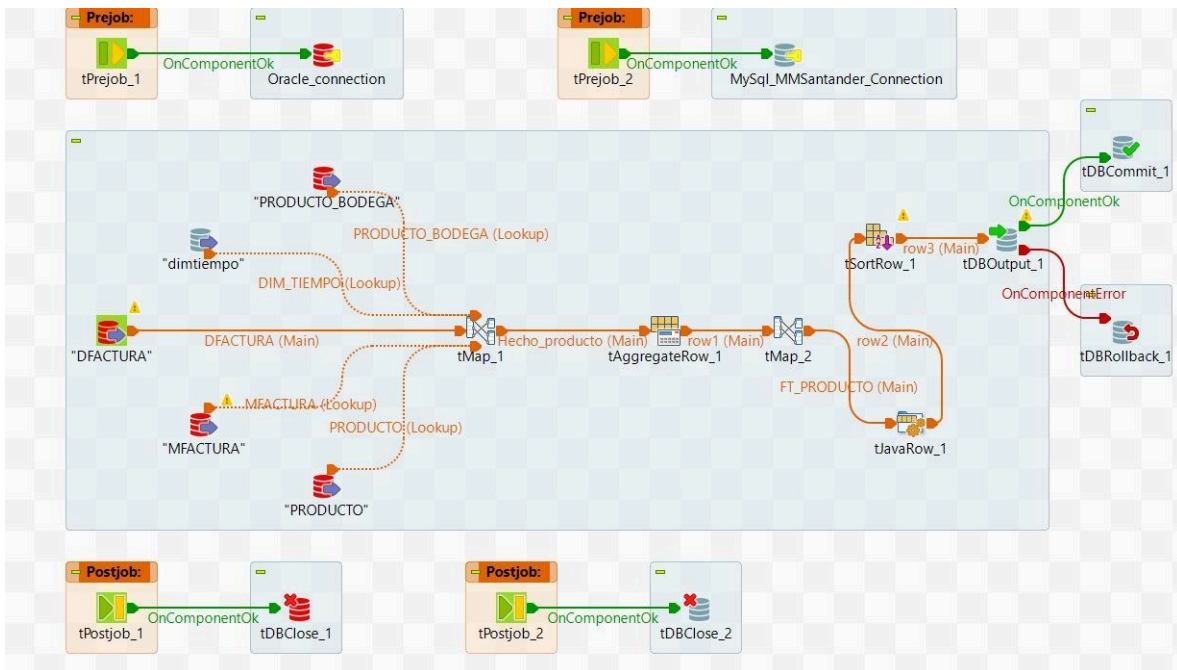
Para concluir, se empleó un componente que permitió ordenar los datos en orden descendente. Posteriormente, se realiza la confirmación (commit) en la base de datos si todos los procesos han tenido éxito. Si alguno de los procesos falla, se procede a la cancelación mediante el rollback.

Finalmente, se procede al cierre de la conexión, tanto con la base de datos fuente como con la base de datos de destino.

Job Hecho_productos

En la ilustración 31, se aprecia detalladamente el trabajo creado en Talend, donde se llevan a cabo varios procesos y agrupaciones con el propósito de construir la tabla de hechos de productos.

Ilustración 31 Creación del Job tabla hecho productos.



Autor: Macias Anyelyn, Ortiz Juan.

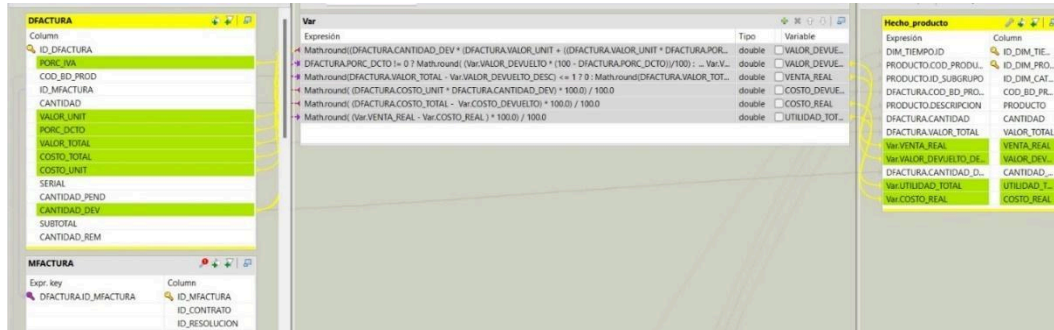
Inicialmente, se abre la conexión con la base de datos fuente y la base de datos de destino. A continuación, se inicia el proceso con la extracción de las tablas desde la base de datos fuente. Posteriormente, se ejecuta un proceso ETL en el que se combinan las tablas

extraídas. Durante esta etapa, se realizan transformaciones que implican el uso de funciones específicas

de Talend para las métricas, y se establecen los atributos de la tabla de hechos, asignándoles los valores correspondientes a las métricas.

En la ilustración 32 se puede observar el proceso de transformación de los datos.

Ilustración 32 Proceso ETL tabla hecho productos.

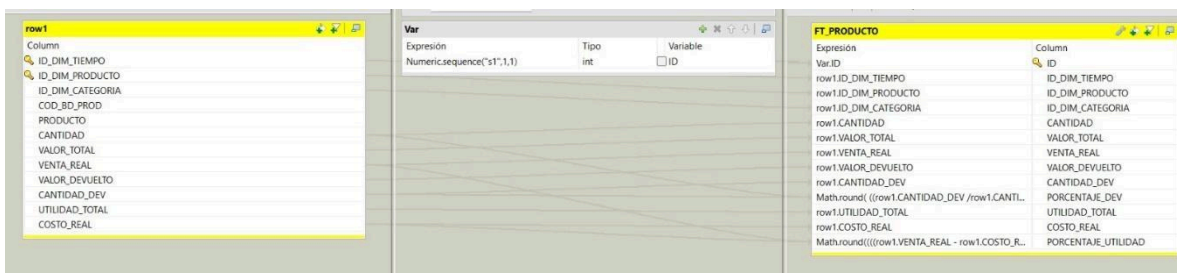


Autor: Macias Anyelyn, Ortiz Juan.

Después de esta etapa, los datos se transmiten a un componente que facilita operaciones de agregación, como la suma de las facturas, entre otras. Una vez que se han calculado estos valores, los datos se asignan a los atributos correspondientes.

En la ilustración 33 se puede observar el proceso descrito anteriormente.

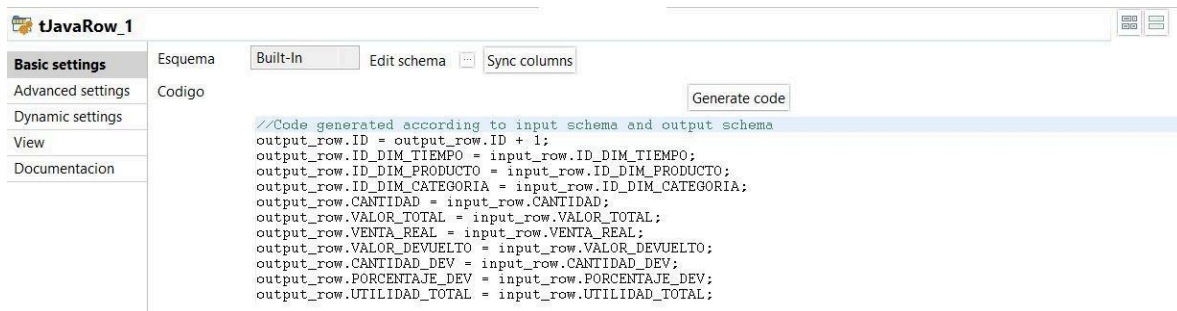
Ilustración 33 Proceso ETL tabla hecho productos.



Autor: Macias Anyelyn, Ortiz Juan.

Para concluir, se utilizó un componente que permitió la ejecución de código Java personalizado. Durante este proceso, se generó un identificador único y se asignaron valores a los atributos de la tabla de hechos de facturas.

Ilustración 34 Código java tabla hecho productos.



The screenshot shows the tJavaRow_1 application interface. On the left, there is a sidebar with menu items: Basic settings, Advanced settings, Dynamic settings, View, and Documentacion. The main area is titled 'Esquema' and contains a 'Built-In' dropdown, 'Edit schema', and 'Sync columns' buttons. Below this, the 'Codigo' section displays the following Java code:

```
//Code generated according to input schema and output schema
output_row.ID = output_row.ID + 1;
output_row.ID_DIM_TIEMPO = input_row.ID_DIM_TIEMPO;
output_row.ID_DIM_PRODUCTO = input_row.ID_DIM_PRODUCTO;
output_row.ID_DIM_CATEGORIA = input_row.ID_DIM_CATEGORIA;
output_row.CANTIDAD = input_row.CANTIDAD;
output_row.VALOR_TOTAL = input_row.VALOR_TOTAL;
output_row.VENTA_REAL = input_row.VENTA_REAL;
output_row.VALOR_DEVUELTO = input_row.VALOR_DEVUELTO;
output_row.CANTIDAD_DEV = input_row.CANTIDAD_DEV;
output_row.PORCENTAJE_DEV = input_row.PORCENTAJE_DEV;
output_row.UTILIDAD_TOTAL = input_row.UTILIDAD_TOTAL;
```

A 'Generate code' button is located to the right of the code block.

Autor: Macias Anyelyn, Ortiz Juan.

Una vez concluido el proceso anterior, se empleó un componente que permitió ordenar los datos en orden descendente. Posteriormente, se realiza la confirmación (commit) en la base de datos si todos los procesos han tenido éxito. Si alguno de los procesos falla, se procede a la cancelación mediante el rollback.

Finalmente, se procede al cierre de la conexión, tanto con la base de datos fuente como con la base de datos de destino.

3.1.6 Implementación y desarrollo de los algoritmos seleccionados.

En esta sección se llevará a cabo el desarrollo, la implementación y las pruebas de los algoritmos seleccionados: suavización exponencial simple, Prophet y ARIMA. El objetivo principal es determinar cuál de estos algoritmos ofrece un rendimiento óptimo al analizar los datos históricos y prever futuras compras. Los algoritmos fueron implementados en R. Las pruebas realizadas permitirán no solo verificar el correcto funcionamiento de los algoritmos, sino también evaluar la precisión del algoritmo de autoaprendizaje en la proyección de las ventas de la empresa, Para finalizar lo que se busca es identificar cuál de los tres algoritmos es el más adecuado para el caso específico que estamos abordando.

3.1.6.1 Implementación y desarrollo.

Durante esta fase, se procedió a recopilar los requerimientos de datos históricos proporcionados por la empresa, adaptando los algoritmos suministrados por R a las necesidades específicas del negocio. El proceso involucró la identificación de los datos esenciales y la modificación de los algoritmos para que se ajustará de manera óptima a

dichos datos. De esta forma, se implementaron las librerías apropiadas en R para garantizar una

integración y ejecución eficiente de las modificaciones realizadas en los algoritmos asegurando una adecuada alineación entre las características de los datos históricos y la funcionalidad del algoritmo.

Algoritmo Suavización exponencial simple (SES)

Inicialmente se realiza la instalación de las bibliotecas necesarios y luego se importan.

Ilustración 35 Código Algoritmo SES.

```
2 #Bibliotecas necesarias,previamente instaladas.  
3 library(dplyr)  
4 library(forecast)  
5 library(RMySQL)  
6 library(Metrics)
```

Autor: Macias Anyelyn, Ortiz Juan.

Segundo se realiza la conexión con la base de datos databasemmsantander, la conexión se establece con los siguientes parámetros, dirección del servidor de la base de datos, el nombre del usuario, la contraseña y el nombre de la base de datos. Una vez creada la conexión se procede a realizar la consulta SQL para obtener los datos de ventas del producto, la consulta se basa en el nombre del producto, cantidad vendida y rango de fechas históricas.

Ilustración 36 Código Algoritmo SES.

```
8 #Cineccion de la base de datos.
9 mysqlconnection <- dbConnect(
10   drv = MySQL(),
11   host = "localhost",
12   user = "root",
13   password = "301310",
14   dbname = "databasemmsantander"
15 )
16
17 producto <- "Producto"
18
19 # Crear la consulta SQL
20 consulta <- sprintf("
21 SELECT t.fecha as fecha, p.CANTIDAD as ventas
22 FROM hecho_productos p join dimtiempo t on t.ID = p.ID_DIM_TIEMPO
23 join dimproducto dm on dm.IdProducto = p.ID_DIM_PRODUCTO
24 where dm.Nombre = '%s'
25 and t.fecha between '2019-01-01' and '2022-12-31'
26 order by t.fecha asc limit 100000
27 ", producto)
28
```

Autor: Macias Anyelyn, Ortiz Juan.

Tercera ejecución de la consulta y manipulación de datos, la consulta se ejecuta utilizando **dbGetQuery**, y luego los datos se manipulan para convertir las fechas a formato mensual y calcular las ventas.

Ilustración 37 Código Algoritmo SES.

```
29 # Ejecutar la consulta
30 datos <- dbGetQuery(mysqlconnection, consulta)
31
32 # Convertir las fechas a formato mensual
33 datos <- mutate(datos, fecha = as.Date(fecha))
34
35 fechas <- datos$fecha
36
37 ventas_mensuales <- datos %>%
38   mutate(month = format(fecha, "%m"), year = format(fecha, "%Y")) %>%
39   group_by(month, year) %>%
40   summarise(ventas = sum(ventas))
41
42 ventas_mensuales <- ventas_mensuales[with(ventas_mensuales, order(ventas_mensuales$year))]
43
44
```

Autor: Macias Anyelyn, Ortiz Juan.

Cuarto se crea una serie temporal con las ventas mensuales especificando el inicio y la frecuencia.

Ilustración 38 Código Algoritmo SES.

```
45 serie <- ts(ventas_mensuales$ventas, start=min(ventas_mensuales$year), frequency=12)
46
```

Autor: Macias Anyelyn, Ortiz Juan.

Quinto entrenamiento del modelo SES, se entrena el modelo utilizando la función **forecast**, se especifican los parámetros.

Ilustración 39 Código Algoritmo SES.

```
46  
47 ses <- ses(serie, alpha=0.2)  
48
```

Autor: Macias Anyelyn, Ortiz Juan.

Sexto se realiza la proyección de un periodo (1 mes) hacia el futuro utilizando el modelo entrenado.

Finalmente, se imprime la proyección.

Ilustración 40 Código Algoritmo SES.

```
49 proyeccion <- forecast(ses, h=1)  
50  
51 print(proyeccion)  
--
```

Autor: Macias Anyelyn, Ortiz Juan.

Algoritmo Prophet

Inicialmente se realiza la instalación de las bibliotecas necesarios y luego se importan.

Ilustración 41 Código algoritmo Prophet.

```
1 # Instalar y cargar paquetes necesarios  
2 install.packages("prophet")  
3 install.packages("RMySQL")  
4 library(prophet)  
5 library(RMySQL)  
6
```

Autor: Macias Anyelyn, Ortiz Juan.

Segundo se realiza la conexión con la base de datos databasemmsantander, la conexión se establece con los siguientes parámetros, dirección del servidor se la base de datos, el nombre del usuario de la base de datos, la contraseña y el nombre de la base de datos. Una vez creada la conexión se procede a realizar la consulta SQL para obtener los datos de

ventas del producto, la consulta se basa en el nombre del producto, cantidad vendida y rango de fechas históricas.

Ilustración 42 Código algoritmo Prophet.

```
7 # Establecer la conexión a MySQL
8 mysqlconnection <- dbConnect(
9   drv = RMySQL::MySQL(),|
10  host = "localhost",
11  user = "root",
12  password = "301310",
13  dbname = "databasemmsantander"
14 )
15
16 producto <- "Producto"
17
18 # Crear la consulta SQL
19 consulta <- sprintf("
20 SELECT t.fecha as ds, p.CANTIDAD as y
21 FROM hecho_productos p join dimtiempo t on t.ID = p.ID_DIM_TIEMPO
22 join dimproducto dm on dm.IdProducto = p.ID_DIM_PRODUCTO
23 where dm.Nombre = '%s'
24 and t.fecha between '2020-05-01' and '2022-12-31'
25 order by t.fecha asc
26 ", producto)
27
```

Autor: Macias Anyelyn, Ortiz Juan.

Tercera ejecución de la consulta esta se ejecuta utilizando dbGetQuery

Ilustración 43 Código algoritmo Prophet.

```
27
28 # Ejecutar la consulta
29 df <- dbGetQuery(mysqlconnection, consulta)
30 |
```

Autor: Macias Anyelyn, Ortiz Juan.

Cuarto Ajustar la frecuencia de los datos, se agrega una columna para la frecuencia, se convierte la columna ds a un objeto de fecha y hora, por último, agrega la columna mes.

Ilustración 44 Código algoritmo Prophet.

```
31 # Ajustar la frecuencia de los datos históricos
32 df$freq <- "day"|
33 # Convertir la columna ds a un objeto de fecha y hora
34 df$ds <- as.Date(df$ds, format = "%Y-%m-%d")
35 # Agregar la columna ds_month
36 df$ds_month <- lubridate::floor_date(df$ds, "month")
37
```

Autor: Macias Anyelyn, Ortiz Juan.

Quinto Ajuste y entrenamiento del modelo, se crea una instancia del modelo, se configura el crecimiento de manera lineal y se especifica el intervalo de confianza a 99%, luego se añade la estacionalidad mensual. Entrenamiento del modelo en este internamente se buscan los parámetros que minimizan el error entre las observaciones reales y las predicciones del modelo en los datos históricos

Ilustración 45 Código algoritmo Prophet.

```
# Ajustar el modelo de crecimiento logístico
m <- prophet(growth = "linear", interval.width = 0.85)
m <- add_seasonality(m, name = "monthly", period = 30.44, fourier.order = 5)

m <- fit.prophet(m, df)
```

Autor: Macias Anyelyn, Ortiz Juan.

Quinto generación de datos futuros y predicciones se crea un dataframe co las fechas futuras para las cuales se desea realizar la predicción 1 mes. Luego se utiliza el modelo ajustado para predecir las ventas futuras.

Ilustración 46 Código algoritmo Prophet.

```
45 # Configurar la fecha inicial y predecir ventas
46 future <- make_future_dataframe(m, periods = 1, freq = "month")
47 fcst <- predict(m, future)
48
```

Autor: Macias Anyelyn, Ortiz Juan.

Finalmente, se imprime la proyección.

Ilustración 47 Código algoritmo Prophet.

```
52 # Imprimir la predicción
53 tail(dplyr::select(fcst, ds, yhat, yhat_lower, yhat_upper), 1)
54
```

Autor: Macias Anyelyn, Ortiz Juan.

Algoritmo Modelo autorregresivo integrado de media móvil (ARIMA)

Inicialmente se realiza la instalación de las bibliotecas necesarios y luego se importan.

Ilustración 48 Código algoritmo ARIMA.

```
1 # Cargamos los paquetes necesarios
2 library(fpp2)
3 library(readxl)
4 library(dplyr)
5 library(lubridate)
6 library(tidyr)
7 library(stringr)
8 library(RMySQL)
9
```

Autor: Macias Anyelyn, Ortiz Juan.

Segundo se realiza la conexión con la base de datos `databaseemmsantander`, la conexión se establece con los siguientes parámetros, dirección del servidor se la base de datos, el nombre del usuario de la base de datos, la contraseña y el nombre de la base de datos. Una vez creada la conexión se procede a realizar la consulta SQL para obtener los datos de ventas del producto, la consulta se basa en el nombre del producto, cantidad vendida y rango de fechas históricas.

Ilustración 49 Código algoritmo ARIMA

```
10 mysqlconnection <- dbConnect(
11   drv = MySQL(),
12   host = "localhost",
13   user = "root",
14   password = "301310",
15   dbname = "databaseemmsantander"
16 )
17
18 producto <- "Producto"
19
20 # Crear la consulta SQL
21 consulta <- sprintf("
22 SELECT t.fecha as fecha, p.CANTIDAD as ventas
23 FROM hecho_productos p join dimtiempo t on t.ID = p.ID_DIM_TIEMPO
24 join dimproducto dm on dm.IdProducto = p.ID_DIM_PRODUCTO
25 where dm.Nombre = '%s'
26 and t.fecha between '2020-05-01' and '2022-12-31'
27 order by t.fecha asc limit 100000
28 ", producto)
29
```

Autor: Macias Anyelyn, Ortiz Juan.

Tercero ejecuta la consulta y manipula los datos para obtener las ventas mensuales, se convierte la columna fecha a formato date, se agrupan los datos por año y mes, se calcula la suma de las ventas por mes y años, finalmente se ordenan por año.

Ilustración 50 Código algoritmo ARIMA

```
30 # Ejecutar la consulta
31 data <- dbGetQuery(mysqlconnection, consulta)
32
33 data <- mutate(data, fecha = as.Date(fecha))
34
35 fechas <- data$fecha
36
37 ventas_mensuales <- data %>%
38   mutate(month = format(fecha, "%m"), year = format(fecha, "%Y")) %>%
39   group_by(month, year) %>%
40   summarise(ventas = sum(ventas))
41
42 ventas_mensuales <- ventas_mensuales[with(ventas_mensuales, order(ventas_mensuales$year))]
43
```

Autor: Macías Anyelyn, Ortiz Juan.

Cuarto se preparan los datos de la serie temporales, se define el rango de datos a utilizar para el pronóstico y se crea la serie temporal a partir de los datos de las ventas utilizando las fechas definidas.

Ilustración 51 Código algoritmo ARIMA

```
44 fecha_mas_reciente <- max(fechas)
45 fecha_mas_antigua <- min(fechas)
46
47 startYear <- year(as.Date(fecha_mas_antigua))
48 startMonth <- month(fecha_mas_antigua)
49 endYear <- year(fecha_mas_reciente)
50 endMonth <- month(fecha_mas_reciente)
51
52 Y <- ts(ventas_mensuales[,3], start = c(startYear, startMonth), end = c(endYear, endMonth), frequency = 12)
53
```

Autor: Macías Anyelyn, Ortiz Juan.

Quinto se ajusta automáticamente el modelo a la serie de tiempo proporcionada y se entrena el modelo.

Ilustración 52 Código algoritmo ARIMA

```
69 modelo_arima <- auto.arima(Y, d=1, D=1, stepwise = FALSE, approximation = FALSE, trace = TRUE)
70
```

Autor: Macías Anyelyn, Ortiz Juan.

Sexto se imprime el modelo y se realiza un diagnóstico de los residuos del modelo, para comprobar si se cumple con las suposiciones de un proceso estacionario.

Ilustración 53 Código algoritmo ARIMA

```
70  
71 print(modelo_arima)  
72  
73 checkresiduals(modelo_arima)  
74
```

Autor: Macias Anyelyn, Ortiz Juan.

Séptimo se genera el pronóstico del siguiente mes, se visualiza el pronóstico.

Ilustración 54 Código algoritmo ARIMA

```
fcst <- forecast(modelo_arima, h=1, level = c(85))  
76 autoplot(fcst) +  
77   ggtitle("pronostico de ventas para el siguiente meses") +  
78   ylab("Cantidad")  
79
```

Autor: Macias Anyelyn, Ortiz Juan.

Octavo se imprime el resumen del pronóstico, el cual incluye estadísticas de los intervalos de confianza y por último se almacena el pronóstico en un dataframe.

Ilustración 55 Código algoritmo ARIMA

```
80 print(summary(fcst))  
81  
82 pronostico <- as.data.frame(fcst)  
83
```

Autor: Macias Anyelyn, Ortiz Juan.

3.1.6.2 Pruebas de los modelos

En esta sección, se llevará a cabo la evaluación de cada uno de los modelos con el objetivo de terminar cuál se adapta de manera más efectiva a este caso en específico. El método de prueba que será empleado se detallará a continuación: primer lugar, se suministrarán a los modelos datos históricos comprendidos entre el 01/01/2019 y el 31/12/2022, abarcando aproximadamente un periodo de 4 años. Posteriormente, se procederá a entrenar los modelos con estos datos, culminando con la realización del pronóstico para el mes de enero del 2023 con una confiabilidad del 85%. Los resultados obtenidos se compararán con los resultados derivados del método tradicional empleado por la empresa para gestionar los pedidos.

La metodología de la empresa consiste en calcular el promedio de las ventas de los últimos tres meses para, luego hacer el pedido o la compra de dicho producto en el mes siguiente. Este proceso se verificará mediante la confrontación con las ventas reales de dicho producto. Finalmente, el modelo seleccionado para este caso será que el que más se acerque al valor de las ventas reales.

A continuación, se presenta una muestra de las pruebas realizadas en una tabla comparativa con los resultados de cada modelo.

Tabla 41 Muestreo de pruebas.

Modelo	Producto	Intervalos de confianza 85%		Proyección	Metodología empresa	Venta real	Diferencias con respecto a la venta real	
		Límite inferior	Límite superior				Proyección	Metodología empresa
Prophet	Servicio perforación	-86	148	31	886	939	908	53
SES	Servicio perforación	178	2.139	1.159	886	939	220	53
ARIMA	Servicio perforación	-207	2.173	982	886	939	43	53
Prophet	Servicio de enchape 19 y 22 Flex	-636	2.056	740	12.268	10.070	9.330	2.198
SES	Servicio de enchape 19 y 22 Flex	5.070	24.290	14.680	12.268	10.070	4.610	2.198
ARIMA	Servicio de enchape 19 y 22 Flex	-442	18.204	8.847	12.268	10.070	1.223	2.198
Prophet	Tornillo drywall zincado 6x2" suma	-347	2.580	1.043	5.199	7.222	6.179	2.023
SES	Tornillo drywall zincado 6x2" suma	849	15.194	8.021	5.199	7.222	799	2.023

ARIMA	Tornillo drywall zincado 6x2" suma	-1.131	14.548	6.700	5.199	7.222	522	2.023
Prophet	Tornillo drywall 6x2" suma	-824	1.085	123	2.466	1.450	1.327	1.016
SES	Tornillo drywall 6x2" suma	-664	7.673	3.504	2.466	1.450	2.054	1.016
ARIMA	Tornillo drywall 6x2" suma	-3.113	6.740	1.813	2.466	1.450	363	1.016
Prophet	Canto blanco 19 flex	-313	561	124	2.024	3.042	2.918	1.018
SES	Canto blanco 19 flex	323	4.940	2.632	2.024	3.042	410	1.018
ARIMA	Canto blanco 19 flex	-26	6.146	3.060	2.024	3.042	18	1.018

Autor: Macias Anyelyn, Ortiz Juan.

Después de llevar a cabo un análisis y evaluación de los diferentes algoritmos de autoaprendizaje o modelos, se ha llegado a la conclusión que el modelo arima es la elección más apropiada para implementar en este proyecto teniendo en cuenta el comportamiento de los datos. Esta decisión se basa en diversos factores que respaldan la idoneidad de arima para este contexto en específico.

Al analizar los resultados obtenidos de cada modelo, se notó que ARIMA se aproximó significativamente a la cantidad real de ventas. La diferencia entre la proyección de ARIMA y las ventas reales fue la menos significativa en comparación con la metodología tradicional, lo que indica una aproximación más precisa a las ventas reales.

Teniendo en cuenta la naturaleza de los datos históricos de la empresa, los cuales no presentan estacionalidad. Se destaca por la capacidad de arima para abordar este desafío mediante la diferenciación, lo que mejora la efectividad del modelado y la precisión de las predicciones.

La flexibilidad de arima frente a un conjunto de datos relativamente pequeño también fue un factor determinante en su elección. Dado que contamos con un banco de datos no tan extenso, la capacidad del modelo para proporcionar predicciones precisas incluso en estas condiciones resulta valiosas.

La necesidad de realizar pronósticos a corto plazo, en la empresa para planificación mensual o trimestral, encontró respaldo en la capacidad de arima para generar predicciones precisas en estos horizontes temporales más inmediatos.

La simplicidad y la facilidad de interpretación de arima fueron aspectos destacados. Esta característica resulta valiosa, ya que se debe comunicar y explicar los resultados a los directivos de la empresa, estos puede que no tengan conocimientos profundos en el área, lo cual facilitara la comprensión y la toma de decisiones.

Para finalizar se tuvo en cuenta la consideración de arima como una herramienta estándar y de referencia en el análisis de series temporales, respaldada por investigaciones previas, lo cual facilitó su implementación y permitió una comparación rigurosa con los otros modelos.

En conclusión, la elección de arima se fundamenta en la capacidad para alinearse con las cantidades reales vendida, su adaptabilidad a datos no estacionarios, eficacia en conjunto de datos pequeños, idoneidad para pronosticar a corto plazo, enfoque univariante, facilidad de interpretación y su reconocimiento como una herramienta estándar en el análisis de series temporales. Todas estas características demuestran que el modelo arima es el más adecuado para este proyecto.

3.1.6.3 Diseño y desarrollo algoritmo seleccionado (ARIMA)

Una vez se ha completado la fase de selección del algoritmo por medio de pruebas y se haya definido este, el siguiente paso implica avanzar en el desarrollo, adecuación de acuerdo a los datos del algoritmo y el diseño de la interfaz de usuario. Para llevar a cabo esta tarea, se utiliza el lenguaje de programación R, junto con la herramienta RStudio un entorno de desarrollo para este lenguaje en su versión 4.3.2, que incluye una serie de funcionalidades específicas para esta finalidad.

En la tabla 41 se detallan todas las librerías empleadas junto con sus respectivas definiciones y su aplicación en el proceso de desarrollo del algoritmo de auto aprendizaje.

Tabla 42 Información librerías empleadas.

Librería	Información	Uso en ARIMA
shiny	Permite crear aplicaciones web iterativas.	Esta librería permitió la creación de la aplicación, la interfaz gráfica

		y la visualización de las proyecciones
fpp2	Brinda funciones para el análisis de series temporales y pronósticos.	Esta librería ayudo a crear, analizar y organizar la serie temporal.
dplyr	Permite realizar operaciones de manipulación de datos.	Esta librería ayudo a realizar el tratamiento de los datos históricos.
lubridate	Brinda funcionalidades para trabajar con fechas y tiempo.	Esta librería permitió el cálculo y extracción de componentes de las fechas de los datos.
tidyr	Brinda funcionalidades para realizar operaciones de manipulación de datos.	Esta librería nos permitió filtrar los datos, seccionar columnas y reordenar las filas de los datos.
stringr	Brinda funciones para realizar operaciones comunes en cadena de caracteres	Esta librería nos permitió construir la consulta SQL que permitió la extracción de los datos históricos de la base de dato.
RODBC	Permite la conexión a base de datos relacionales.	Esta librería permitió la conexión a la base de datos fuente Oracle.

Autor: Macias Anyelyn, Ortiz Juan.

3.1.6.4 Desarrollo código fuente

En esta fase se implementa el código fuente del algoritmo ARIMA, mencionado en las pruebas en el inciso 00, con algunas modificaciones y se le agrega código para el desarrollo de la interfaz gráfica, para mejorar la experiencia de usuario.

Inicialmente se cargan las librerías necesarias para la aplicación y se establece el directorio de trabajo con Setwd().

Ilustración 56 Desarrollo código fuente ARIMA.

```

1 # Cargamos los paquetes necesarios
2 library(shiny)
3 library(fpp2)
4 library(readxl)
5 library(dplyr)
6 library(lubridate)
7 library(tidyr)
8 library(stringr)
9 library(RODBC)
10
11
12 setwd("C:\\Users\\juank\\Documents\\Proyecto de grado\\Proyeccion de ventas\\R\\arima")
13

```

Autor: Macias Anyelyn, Ortiz Juan.

Segundo se establece la conexión a la base de datos Oracle donde se encuentra alojado los datos históricos, por medio de **odbcConnect()**.

Ilustración 57 Desarrollo código fuente ARIMA.

```
13  
14 conn <- odbcConnect(  
15   "conexionOracleStudio", uid = "ZUI_2018", pwd = "ZAFIROKEY"  
16 )  
17
```

Autor: Macias Anyelyn, Ortiz Juan.

Tercero se realiza la consulta a la base de datos para obtener una lista de productos, posteriormente se crea un menú desplegable en la interfaz Shiny con los productos obtenidos.

Ilustración 58 Desarrollo código fuente ARIMA.

```
18 # Ejecutar la consulta y obtener los productos  
19 productos <- sqlQuery(conn, "SELECT DESCRIPCION as nombre FROM PRODUCTO")  
20  
21 productos <- lapply(productos, iconv, from = "latin1", to = "utf8")  
22  
23 # Crear el menú desplegable  
24 productos_select <- selectInput(  
25   "producto",  
26   "Seleccione un producto",  
27   choices = productos$NOMBRE  
28 )  
29
```

Autor: Macias Anyelyn, Ortiz Juan.

Cuarto se define una función para ejecutar la consulta en la base de datos y le agrega un TryCatch(), para el control de excepciones.

Se define la función ejecutar consulta, encargada de ejecutar la consulta específica para obtener los datos de ventas del producto seleccionado. Aparte se realizan los siguientes procesos:

- Se modelan los datos para creación de la serie temporal.
- Se ajustan los parámetros para la creación de la serie temporal.
- Se configura el modelo arima.
- Se ejecuta la proyección de la compra.
- Se establecen los datos de salida de la proyección.

Ilustración 59 Desarrollo código fuente ARIMA.

```

30 # Crear la función para ejecutar la consulta
31- ejecutar_consulta_aux <- function(consulta) {
32-   tryCatch({
33     data <- sqQuery(conn, consulta)
34     return(data)
35   }, error = function(e) {
36     return(NULL)
37   })
38- }
39- }

40 # Crear la función para ejecutar la consulta
41- ejecutar_consulta <- function(producto_se) {
42- }
43 # Obtener el producto seleccionado
44 producto <- producto_se
45
46 print(producto_se)
47
48 # Crear la consulta SQL
49 consulta <- paste0("SELECT M.FECHA_FACTURA AS Fecha, ROUND(SUM(D.CANTIDAD)) AS Ventas FROM DFACTURA D join
50                    MFACTURA M on D.ID_MFACTURA = M.ID_MFACTURA join PRODUCTO P on D.COD_PROD LIKE
51                    CONCAT('%', P.COD_PRODUCTO) WHERE P.DESCRIPCION = '", producto, "' GROUP BY M.FECHA_FACTURA
52                    order by Fecha asc")
53
54 print(consulta)
55
56 # Ejecutar la consulta
57 data <- ejecutar_consulta_aux(consulta)
58
59- if (is.null(data)) {
60   # Mostrar un mensaje de error en la interfaz de Shiny
61   showNotification("Error al ejecutar la consulta. Asegúrate de que hay al menos 2 periodos en
62                    la serie temporal.", type = "error")
63   return(NULL)
64- }
65
66 data <- mutate(data, FECHA = as.Date(FECHA))
67
68 print(data)
69
70 data <- as.data.frame(data)
71 print(class(data$VENTAS))
72
73 fechas <- data$FECHA
74
75
76 ventas_mensuales <- data %>%
77   mutate(month = format(FECHA, "%m"), year = format(FECHA, "%Y")) %>%
78   group_by(month, year) %>%
79   summarise(VENTAS = sum(VENTAS), .groups = "drop")
80
81 print(ventas_mensuales)
82
83 fecha_mas_reciente <- max(fechas)
84 fecha_mas_antigua <- min(fechas)
85
86 startYear <- year(as.Date(fecha_mas_antigua))
87 startMonth <- month(fecha_mas_antigua)
88 endYear <- year(fecha_mas_reciente)
89 endMonth <- month(fecha_mas_reciente)
90
91- if (length(ventas_mensuales$VENTAS) < 2) {
92   # Mostrar un mensaje de error en la interfaz de Shiny
93   showNotification("Error: La serie temporal tiene menos de 2 periodos. No se puede descomponer.", type = "error")
94   return(NULL)
95- }
96
97 Y <- ts(ventas_mensuales[,3], start = c(startYear, startMonth), end = c(endYear, endMonth), frequency = 12)
98
99- tryCatch({
100   descom-decompose(Y)
101   autoplot(descom)
102   acf(Y)
103   pacf(Y)
104- }, error = function(e) {
105   return(NULL)
106- })
107
108 modelo_arima <- auto.arima(Y, d=1, D=1, stepwise = FALSE, approximation = FALSE, trace = TRUE)
109 fcst <- forecast(modelo_arima, h=1, level = c(99))
110 pronostico <- as.data.frame(fcst)
111 colnames(pronostico) <- c("Proyeccion", "Limite inferior", "Limite superior")
112- }

```

Autor: Macias Anyelyn, Ortiz Juan.

Quinto se define la interfaz de usuario donde se crea lo siguiente:

- Encabezado con el logo y el título del aplicativo. También se le agrega los diseños personalizados como: color, tipo de letra, tamaño, etc.
- Sidebar panel con el menú desplegable de productos con los diseños personalizados como: color, tipo de letra, tamaño, etc.

- Nota informativa para el uso de las proyecciones con los diseños personalizados como: color, tipo de letra, tamaño, etc.
- Botón para ejecutar la consulta con los diseños personalizados como: color, tipo de letra, tamaño, etc.
- Tabla para mostrar las proyecciones, con los diseños personalizados como: color de la tabla, color de la letra, tamaño, etc.

Ilustración 60 Desarrollo código fuente ARIMA.

```

116 ui = FluidPage(
117   tags/head(
118     tags/style(
119       HTML(
120         "
121         .header {
122           display: flex; /* Establecer el modelo de caja flexible */
123           align-items: center; /* Centrar verticalmente los elementos */
124           background-color: #56CD23; /* Color de Fondo */
125           color: white; /* Color del texto */
126           padding: 20px; /* Espaciado interno */
127           text-align: center; /* Alineación del texto */
128         }
129
130         .header img {
131           max-height: 90px; /* Altura máxima de la imagen */
132           margin-right: 20px; /* Margen derecho para separar la imagen del texto */
133         }
134         h1 {
135           font-weight: bold; /* Hacer que el texto sea negrita */
136           font-family: 'open sans', sans-serif;
137           font-size: 27px;
138         }
139
140         .content-text {
141           font-family: 'Open Sans', sans-serif;
142           font-size: 16px; /* Puedes ajustar el tamaño según tus preferencias */
143         }
144         .green-button {
145           background-color: #56CD23; /* Color verde por defecto */
146           color: white;
147           font-family: 'Open Sans', sans-serif;
148           margin-top: 20px;
149           font-size: 14px; /* Puedes ajustar el tamaño según tus preferencias */
150         }
151
152         #tabla th {
153           background-color: #56CD23;
154           color: white;
155           border: solid 1px;
156         }
157         #tabla td {
158           background-color: #E3E3E3;
159         }
160
161         #mainpanel {
162           width: 100%;
163           display: flex;
164           justify-content: center;
165           align-items: center;
166         }
167
168         "
169       )
170     )
171   ),
172   "
173   "
174   "
175   "
176   "
177   "
178   "
179   "
180   div(class = "header",
181     tags/img(src = "Logo.png", alt = "Logo"),
182     div(h1("PROYECCION DE COMPRAS MATERIALES MADERA SANTANDER"))
183   ),
184   "
185   sidebarlayout(
186     sidebarpanel(
187       fluidrow(
188         productos_select
189       ),
190       fluidrow(
191         helptext(style = "text-align: justify;", "Nota: Esta proyección de ventas es una estimación con un 87% de
192         confiabilidad, basada en datos históricos y tendencias actuales. Sin embargo, no garantizamos su
193         exactitud al 100%. Recomendamos utilizarla como guía, estando preparados para ajustes según sea
194         necesario.",
195         ),
196         actionbutton("ejecutar", "Ejecutar proyección", icon = icon("play"), class = "green-button", align = "center"
197         ), width = 12, style = "background-color: white; border: none; margin-top: 10px"
198       ),
199     )
200   ),
201   "
202   "
203   "
204   "
205   "
206   "
207   "
208   "
209   "
210   "
211   "
212   "
213   "
214   "
215   "
216   "
217   "
218   "
219   "
220   "
221   "
222   "
223   "
224   "
225   "
226   "
227   "
228   "
229   "
230   "
231   "
232   "
233   "
234   "
235   "
236   "
237   "
238   "
239   "
240   "
241   "
242   "
243   "
244   "
245   "
246   "
247   "
248   "
249   "
250   "
251   "
252   "
253   "
254   "
255   "
256   "
257   "
258   "
259   "
260   "
261   "
262   "
263   "
264   "
265   "
266   "
267   "
268   "
269   "
270   "
271   "
272   "
273   "
274   "
275   "
276   "
277   "
278   "
279   "
280   "
281   "
282   "
283   "
284   "
285   "
286   "
287   "
288   "
289   "
290   "
291   "
292   "
293   "
294   "
295   "
296   "
297   "
298   "
299   "
300   "
301   "
302   "
303   "
304   "
305   "
306   "
307   "
308   "
309   "
310   "
311   "
312   "
313   "
314   "
315   "
316   "
317   "
318   "
319   "
320   "
321   "
322   "
323   "
324   "
325   "
326   "
327   "
328   "
329   "
330   "
331   "
332   "
333   "
334   "
335   "
336   "
337   "
338   "
339   "
340   "
341   "
342   "
343   "
344   "
345   "
346   "
347   "
348   "
349   "
350   "
351   "
352   "
353   "
354   "
355   "
356   "
357   "
358   "
359   "
360   "
361   "
362   "
363   "
364   "
365   "
366   "
367   "
368   "
369   "
370   "
371   "
372   "
373   "
374   "
375   "
376   "
377   "
378   "
379   "
380   "
381   "
382   "
383   "
384   "
385   "
386   "
387   "
388   "
389   "
390   "
391   "
392   "
393   "
394   "
395   "
396   "
397   "
398   "
399   "
400   "
401   "
402   "
403   "
404   "
405   "
406   "
407   "
408   "
409   "
410   "
411   "
412   "
413   "
414   "
415   "
416   "
417   "
418   "
419   "
420   "
421   "
422   "
423   "
424   "
425   "
426   "
427   "
428   "
429   "
430   "
431   "
432   "
433   "
434   "
435   "
436   "
437   "
438   "
439   "
440   "
441   "
442   "
443   "
444   "
445   "
446   "
447   "
448   "
449   "
450   "
451   "
452   "
453   "
454   "
455   "
456   "
457   "
458   "
459   "
460   "
461   "
462   "
463   "
464   "
465   "
466   "
467   "
468   "
469   "
470   "
471   "
472   "
473   "
474   "
475   "
476   "
477   "
478   "
479   "
480   "
481   "
482   "
483   "
484   "
485   "
486   "
487   "
488   "
489   "
490   "
491   "
492   "
493   "
494   "
495   "
496   "
497   "
498   "
499   "
500   "
501   "
502   "
503   "
504   "
505   "
506   "
507   "
508   "
509   "
510   "
511   "
512   "
513   "
514   "
515   "
516   "
517   "
518   "
519   "
520   "
521   "
522   "
523   "
524   "
525   "
526   "
527   "
528   "
529   "
530   "
531   "
532   "
533   "
534   "
535   "
536   "
537   "
538   "
539   "
540   "
541   "
542   "
543   "
544   "
545   "
546   "
547   "
548   "
549   "
550   "
551   "
552   "
553   "
554   "
555   "
556   "
557   "
558   "
559   "
560   "
561   "
562   "
563   "
564   "
565   "
566   "
567   "
568   "
569   "
570   "
571   "
572   "
573   "
574   "
575   "
576   "
577   "
578   "
579   "
580   "
581   "
582   "
583   "
584   "
585   "
586   "
587   "
588   "
589   "
590   "
591   "
592   "
593   "
594   "
595   "
596   "
597   "
598   "
599   "
600   "
601   "
602   "
603   "
604   "
605   "
606   "
607   "
608   "
609   "
610   "
611   "
612   "
613   "
614   "
615   "
616   "
617   "
618   "
619   "
620   "
621   "
622   "
623   "
624   "
625   "
626   "
627   "
628   "
629   "
630   "
631   "
632   "
633   "
634   "
635   "
636   "
637   "
638   "
639   "
640   "
641   "
642   "
643   "
644   "
645   "
646   "
647   "
648   "
649   "
650   "
651   "
652   "
653   "
654   "
655   "
656   "
657   "
658   "
659   "
660   "
661   "
662   "
663   "
664   "
665   "
666   "
667   "
668   "
669   "
670   "
671   "
672   "
673   "
674   "
675   "
676   "
677   "
678   "
679   "
680   "
681   "
682   "
683   "
684   "
685   "
686   "
687   "
688   "
689   "
690   "
691   "
692   "
693   "
694   "
695   "
696   "
697   "
698   "
699   "
700   "
701   "
702   "
703   "
704   "
705   "
706   "
707   "
708   "
709   "
710   "
711   "
712   "
713   "
714   "
715   "
716   "
717   "
718   "
719   "
720   "
721   "
722   "
723   "
724   "
725   "
726   "
727   "
728   "
729   "
730   "
731   "
732   "
733   "
734   "
735   "
736   "
737   "
738   "
739   "
740   "
741   "
742   "
743   "
744   "
745   "
746   "
747   "
748   "
749   "
750   "
751   "
752   "
753   "
754   "
755   "
756   "
757   "
758   "
759   "
760   "
761   "
762   "
763   "
764   "
765   "
766   "
767   "
768   "
769   "
770   "
771   "
772   "
773   "
774   "
775   "
776   "
777   "
778   "
779   "
780   "
781   "
782   "
783   "
784   "
785   "
786   "
787   "
788   "
789   "
790   "
791   "
792   "
793   "
794   "
795   "
796   "
797   "
798   "
799   "
800   "
801   "
802   "
803   "
804   "
805   "
806   "
807   "
808   "
809   "
810   "
811   "
812   "
813   "
814   "
815   "
816   "
817   "
818   "
819   "
820   "
821   "
822   "
823   "
824   "
825   "
826   "
827   "
828   "
829   "
830   "
831   "
832   "
833   "
834   "
835   "
836   "
837   "
838   "
839   "
840   "
841   "
842   "
843   "
844   "
845   "
846   "
847   "
848   "
849   "
850   "
851   "
852   "
853   "
854   "
855   "
856   "
857   "
858   "
859   "
860   "
861   "
862   "
863   "
864   "
865   "
866   "
867   "
868   "
869   "
870   "
871   "
872   "
873   "
874   "
875   "
876   "
877   "
878   "
879   "
880   "
881   "
882   "
883   "
884   "
885   "
886   "
887   "
888   "
889   "
890   "
891   "
892   "
893   "
894   "
895   "
896   "
897   "
898   "
899   "
900   "
901   "
902   "
903   "
904   "
905   "
906   "
907   "
908   "
909   "
910   "
911   "
912   "
913   "
914   "
915   "
916   "
917   "
918   "
919   "
920   "
921   "
922   "
923   "
924   "
925   "
926   "
927   "
928   "
929   "
930   "
931   "
932   "
933   "
934   "
935   "
936   "
937   "
938   "
939   "
940   "
941   "
942   "
943   "
944   "
945   "
946   "
947   "
948   "
949   "
950   "
951   "
952   "
953   "
954   "
955   "
956   "
957   "
958   "
959   "
960   "
961   "
962   "
963   "
964   "
965   "
966   "
967   "
968   "
969   "
970   "
971   "
972   "
973   "
974   "
975   "
976   "
977   "
978   "
979   "
980   "
981   "
982   "
983   "
984   "
985   "
986   "
987   "
988   "
989   "
990   "
991   "
992   "
993   "
994   "
995   "
996   "
997   "
998   "
999   "
1000  "
1001  "
1002  "
1003  "
1004  "
1005  "
1006  "
1007  "
1008  "
1009  "
1010  "
1011  "
1012  "
1013  "
1014  "
1015  "
1016  "
1017  "
1018  "
1019  "
1020  "
1021  "
1022  "
1023  "
1024  "
1025  "
1026  "
1027  "
1028  "
1029  "
1030  "
1031  "
1032  "
1033  "
1034  "
1035  "
1036  "
1037  "
1038  "
1039  "
1040  "
1041  "
1042  "
1043  "
1044  "
1045  "
1046  "
1047  "
1048  "
1049  "
1050  "
1051  "
1052  "
1053  "
1054  "
1055  "
1056  "
1057  "
1058  "
1059  "
1060  "
1061  "
1062  "
1063  "
1064  "
1065  "
1066  "
1067  "
1068  "
1069  "
1070  "
1071  "
1072  "
1073  "
1074  "
1075  "
1076  "
1077  "
1078  "
1079  "
1080  "
1081  "
1082  "
1083  "
1084  "
1085  "
1086  "
1087  "
1088  "
1089  "
1090  "
1091  "
1092  "
1093  "
1094  "
1095  "
1096  "
1097  "
1098  "
1099  "
1100  "
1101  "
1102  "
1103  "
1104  "
1105  "
1106  "
1107  "
1108  "
1109  "
1110  "
1111  "
1112  "
1113  "
1114  "
1115  "
1116  "
1117  "
1118  "
1119  "
1120  "
1121  "
1122  "
1123  "
1124  "
1125  "
1126  "
1127  "
1128  "
1129  "
1130  "
1131  "
1132  "
1133  "
1134  "
1135  "
1136  "
1137  "
1138  "
1139  "
1140  "
1141  "
1142  "
1143  "
1144  "
1145  "
1146  "
1147  "
1148  "
1149  "
1150  "
1151  "
1152  "
1153  "
1154  "
1155  "
1156  "
1157  "
1158  "
1159  "
1160  "
1161  "
1162  "
1163  "
1164  "
1165  "
1166  "
1167  "
1168  "
1169  "
1170  "
1171  "
1172  "
1173  "
1174  "
1175  "
1176  "
1177  "
1178  "
1179  "
1180  "
1181  "
1182  "
1183  "
1184  "
1185  "
1186  "
1187  "
1188  "
1189  "
1190  "
1191  "
1192  "
1193  "
1194  "
1195  "
1196  "
1197  "
1198  "
1199  "
1200  "
1201  "
1202  "
1203  "
1204  "
1205  "
1206  "
1207  "
1208  "
1209  "
1210  "
1211  "
1212  "
1213  "
1214  "
1215  "
1216  "
1217  "
1218  "
1219  "
1220  "
1221  "
1222  "
1223  "
1224  "
1225  "
1226  "
1227  "
1228  "
1229  "
1230  "
1231  "
1232  "
1233  "
1234  "
1235  "
1236  "
1237  "
1238  "
1239  "
1240  "
1241  "
1242  "
1243  "
1244  "
1245  "
1246  "
1247  "
1248  "
1249  "
1250  "
1251  "
1252  "
1253  "
1254  "
1255  "
1256  "
1257  "
1258  "
1259  "
1260  "
1261  "
1262  "
1263  "
1264  "
1265  "
1266  "
1267  "
1268  "
1269  "
1270  "
1271  "
1272  "
1273  "
1274  "
1275  "
1276  "
1277  "
1278  "
1279  "
1280  "
1281  "
1282  "
1283  "
1284  "
1285  "
1286  "
1287  "
1288  "
1289  "
1290  "
1291  "
1292  "
1293  "
1294  "
1295  "
1296  "
1297  "
1298  "
1299  "
1300  "
1301  "
1302  "
1303  "
1304  "
1305  "
1306  "
1307  "
1308  "
1309  "
1310  "
1311  "
1312  "
1313  "
1314  "
1315  "
1316  "
1317  "
1318  "
1319  "
1320  "
1321  "
1322  "
1323  "
1324  "
1325  "
1326  "
1327  "
1328  "
1329  "
1330  "
1331  "
1332  "
1333  "
1334  "
1335  "
1336  "
1337  "
1338  "
1339  "
1340  "
1341  "
1342  "
1343  "
1344  "
1345  "
1346  "
1347  "
1348  "
1349  "
1350  "
1351  "
1352  "
1353  "
1354  "
1355  "
1356  "
1357  "
1358  "
1359  "
1360  "
1361  "
1362  "
1363  "
1364  "
1365  "
1366  "
1367  "
1368  "
1369  "
1370  "
1371  "
1372  "
1373  "
1374  "
1375  "
1376  "
1377  "
1378  "
1379  "
1380  "
1381  "
1382  "
1383  "
1384  "
1385  "
1386  "
1387  "
1388  "
1389  "
1390  "
1391  "
1392  "
1393  "
1394  "
1395  "
1396  "
1397  "
1398  "
1399  "
1400  "
1401  "
1402  "
1403  "
1404  "
1405  "
1406  "
1407  "
1408  "
1409  "
1410  "
1411  "
1412  "
1413  "
1414  "
1415  "
1416  "
1417  "
1418  "
1419  "
1420  "
1421  "
1422  "
1423  "
1424  "
1425  "
1426  "
1427  "
1428  "
1429  "
1430  "
1431  "
1432  "
1433  "
1434  "
1435  "
1436  "
1437  "
1438  "
1439  "
1440  "
1441  "
1442  "
1443  "
1444  "
1445  "
1446  "
1447  "
1448  "
1449  "
1450  "
1451  "
1452  "
1453  "
1454  "
1455  "
1456  "
1457  "
1458  "
1459  "
1460  "
1461  "
1462  "
1463  "
1464  "
1465  "
1466  "
1467  "
1468  "
1469  "
1470  "
1471  "
1472  "
1473  "
1474  "
1475  "
1476  "
1477  "
1478  "
1479  "
1480  "
1481  "
1482  "
1483  "
1484  "
1485  "
1486  "
1487  "
1488  "
1489  "
1490  "
1491  "
1492  "
1493  "
1494  "
1495  "
1496  "
1497  "
1498  "
1499  "
1500  "
1501  "
1502  "
1503  "
1504  "
1505  "
1506  "
1507  "
1508  "
1509  "
1510  "
1511  "
1512  "
1513  "
1514  "
1515  "
1516  "
1517  "
1518  "
1519  "
1520  "
1521  "
1522  "
1523  "
1524  "
1525  "
1526  "
1527  "
1528  "
1529  "
1530  "
1531  "
1532  "
1533  "
1534  "
1535  "
1536  "
1537  "
1538  "
1539  "
1540  "
1541  "
1542  "
1543  "
1544  "
1545  "
1546  "
1547  "
1548  "
1549  "
1550  "
1551  "
1552  "
1553  "
1554  "
1555  "
1556  "
1557  "
1558  "
1559  "
1560  "
1561  "
1562  "
1563  "
1564  "
1565  "
1566  "
1567  "
1568  "
1569  "
1570  "
1571  "
1572  "
1573  "
1574  "
1575  "
1576  "
1577  "
1578  "
1579  "
1580  "
1581  "
1582  "
1583  "
1584  "
1585  "
1586  "
1587  "
1588  "
1589  "
1590  "
1591  "
1592  "
1593  "
1594  "
1595  "
1596  "
1597  "
1598  "
1599  "
1600  "
1601  "
1602  "
1603  "
1604  "
1605  "
1606  "
1607  "
1608  "
1609  "
1610  "
1611  "
1612  "
1613  "
1614  "
1615  "
1616  "
1617  "
1618  "
1619  "
1620  "
1621  "
1622  "
1623  "
1624  "
1625  "
1626  "
1627  "
1628  "
1629  "
1630  "
1631  "
1632  "
1633  "
1634  "
1635  "
1636  "
1637  "
1638  "
1639  "
1640  "
1641  "
1642  "
1643  "
1644  "
1645  "
1646  "
1647  "
1648  "
1649  "
1650  "
1651  "
1652  "
1653  "
1654  "
1655  "
1656  "
1657  "
1658  "
1659  "
1660  "
1661  "
1662  "
1663  "
1664  "
1665  "
1666  "
1667  "
1668  "
1669  "
1670  "
1671  "
1672  "
1673  "
1674  "
1675  "
1676  "
1677  "
1678  "
1679  "
1680  "
1681  "
1682  "
1683  "
1684  "
1685  "
1686  "
1687  "
1688  "
1689  "
1690  "
1691  "
1692  "
1693  "
1694  "
1695  "
1696  "
1697  "
1698  "
1699  "
1700  "
1701  "
1702  "
1703  "
1704  "
1705  "
1706  "
1707  "
1708  "
1709  "
1710  "
1711  "
1712  "
1713  "
1714  "
1715  "
1716  "
1717  "
1718  "
1719  "
1720  "
1721  "
1722  "
1723  "
1724  "
1725  "
1726  "
1727  "
1728  "
1729  "
1730  "
1731  "
1732  "
1733  "
1734  "
1735  "
1736  "
1737  "
1738  "
1739  "
1740  "
1741  "
1742  "
1743  "
1744  "
1745  "
1746  "
1747  "
1748  "
1749  "
1750  "
1751  "
1752  "
1753  "
1754  "
1755  "
1756  "
1757  "
1758  "
1759  "
1760  "
1761  "
1762  "
1763  "
1764  "
1765  "
1766  "
1767  "
1768  "
1769  "
1770  "
1771  "
1772  "
1773  "
1774  "
1775  "
1776  "
1777  "
1778  "
1779  "
1780  "
1781  "
1782  "
1783  "
1784  "
1785  "
1786  "
1787  "
1788  "
1789  "
1790  "
1791  "
1792  "
1793  "
1794  "
1795  "
1796  "
1797  "
1798  "
1799  "
1800  "
1801  "
1802  "
1803  "
1804  "
1805  "
1806  "
1807  "
1808  "
1809  "
1810  "
1811  "
1812  "
1813  "
1814  "
1815  "
1816  "
1817  "
1818  "
1819  "
1820  "
1821  "
1822  "
1823  "
1824  "
1825  "
1826  "
1827  "
1828  "
1829  "
1830  "
1831  "
1832  "
1833  "
1834  "
1835  "
1836  "
1837  "
1838  "
1839  "
1840  "
1841  "
1842  "
1843  "
1844  "
1845  "
1846  "
1847  "
1848  "
1849  "
1850  "
1851  "
1852  "
1853  "
1854  "
1855  "
1856  "
1857  "
1858  "
1859  "
1860  "
1861  "
1862  "
1863  "
1864  "
1865  "
1866  "
1867  "
1868  "
1869  "
1870  "
1871  "
1872  "
1873  "
1874  "
1875  "
1876  "
1877  "
1878  "
1879  "
1880  "
1881  "
1882  "
1883  "
1884  "
1885  "
1886  "
1887  "
1888  "
1889  "
1890  "
1891  "
1892  "
1893  "
1894  "
1895  "
1896  "
1897  "
1898  "
1899  "
1900  "
1901  "
1902  "
1903  "
1904  "
1905  "
1906  "
1907  "
1908  "
1909  "
1910  "
1911  "
1912  "
1913  "
1914  "
1915  "
1916
```

Ilustración 61 Desarrollo código fuente ARIMA.

```
209 - server = function(input, output) {
210
211   library(shiny)
212   library(shinyjs)
213
214   # Generamos un texto vacío al inicio de la aplicación
215   output$tabla <- renderText("")
216
217   # Creamos una variable reactiva para almacenar la proyección
218 -  proyeccion <- reactive({
219     # Llamamos a la función de R
220     proyeccion <- ejecutar_consulta(input$producto)
221     return(proyeccion)
222 -  })
223
224   # Actualizamos la tabla con los datos de la función 'ejecutar_consulta()'
225 -  observeEvent(input$ejecutar, {
226     output$tabla <- renderTable({
227       proyeccion()
228     }, rownames = FALSE, digits = 2)
229 -  })
230 - }
231
232 shinyApp(ui, server)
233
```

Autor: Macias Anyelyn, Ortiz Juan.

3.1.6 Implementación

Después de llevar a cabo las etapas de análisis, diseño y desarrollo de la bodega de datos para el área de ventas y compras en la empresa Materiales Maderas Santander, se ha alcanzado exitosamente el cumplimiento de los objetivos del proyecto. Esto se ha logrado mediante la implementación de una solución integral que combina inteligencia empresarial acompañada de una proyección de compras. Por un lado, se creó un módulo de visualización utilizando la herramienta Power BI, que posibilita la representación visual de datos valiosos obtenidos después de completar la extracción, transformación y carga de información de gran relevancia en el proceso de negocio. Este módulo ha permitido satisfacer los requisitos planteados al inicio del proyecto, brindando una visión clara y concisa de los indicadores clave de rendimientos.

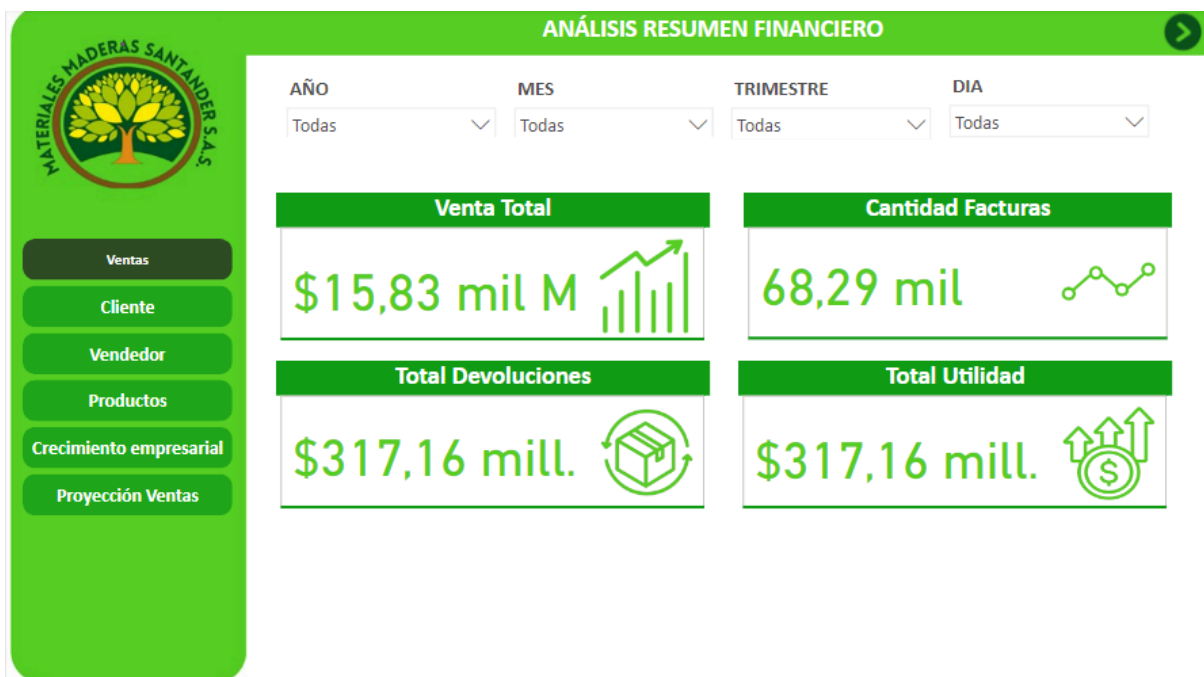
Por otro lado, se implementó un algoritmo de autoaprendizaje, más específicamente ARIMA, para la proyección de compras, utilizando la herramienta RStudio. Esta proyección, que se basa en datos históricos y tendencias, facilita la planificación de compras en la empresa, optimizando el costo de los productos al anticipar la demanda de cada artículo. Además,

reduce los riesgos de escasez de inventario en momentos críticos y evita la acumulación excesiva de existencias, lo que mejora la satisfacción y fidelización del cliente al garantizar la disponibilidad adecuada de productos en todo momento.

Es importante tener en cuenta que la proyección se debe utilizar como una guía o soporte al tomar decisiones, ya que no es 100% precisa debido a la presencia constante de incertidumbre y variables imprevistas que pueden afectar los resultados.

A continuación, presentamos los resultados del módulo de visualización y de la proyección de compras en respuesta a los requerimientos establecidos.

Ilustración 62 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Ventas (Resumen financiero).



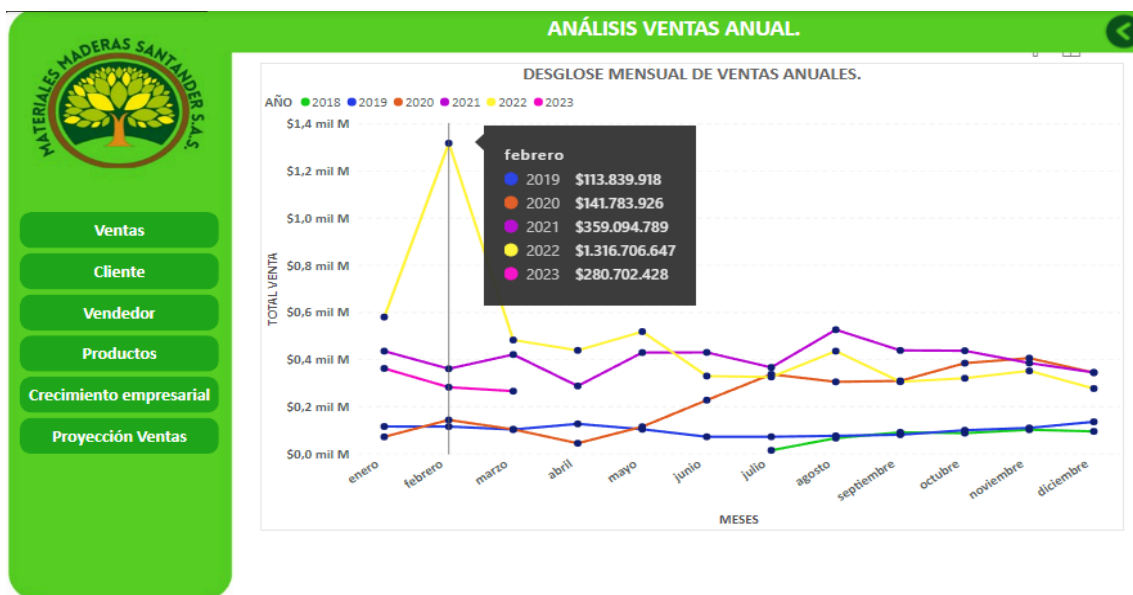
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 63 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S Ventas (Estadísticas facturas).



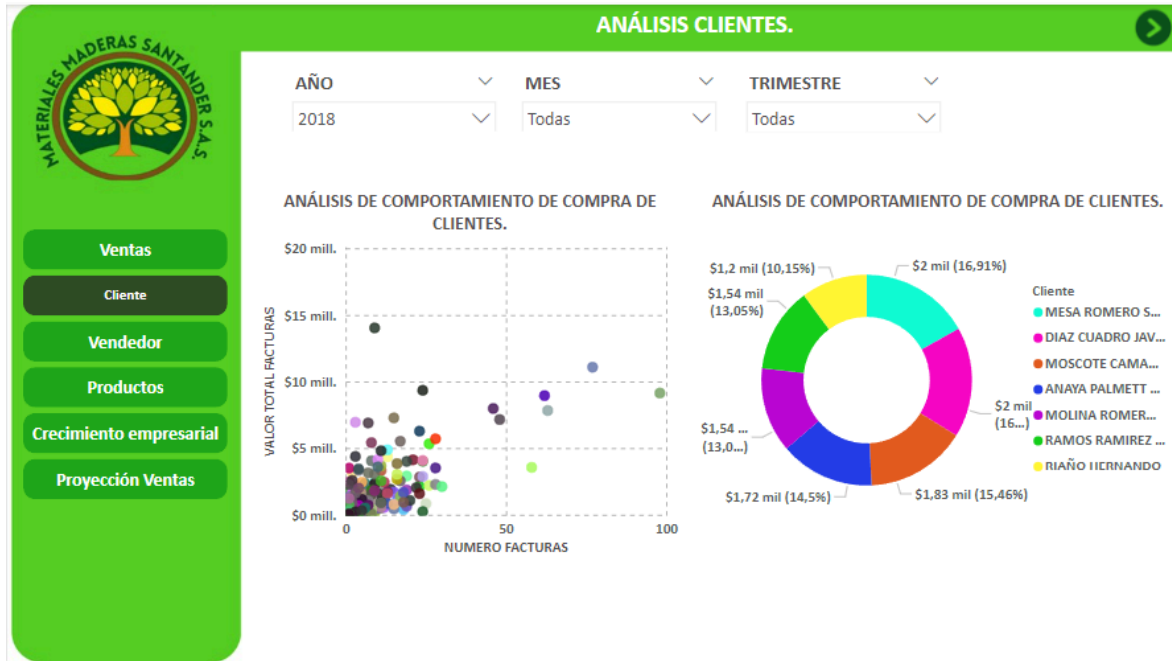
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 64 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S Ventas (Análisis ventas anual).



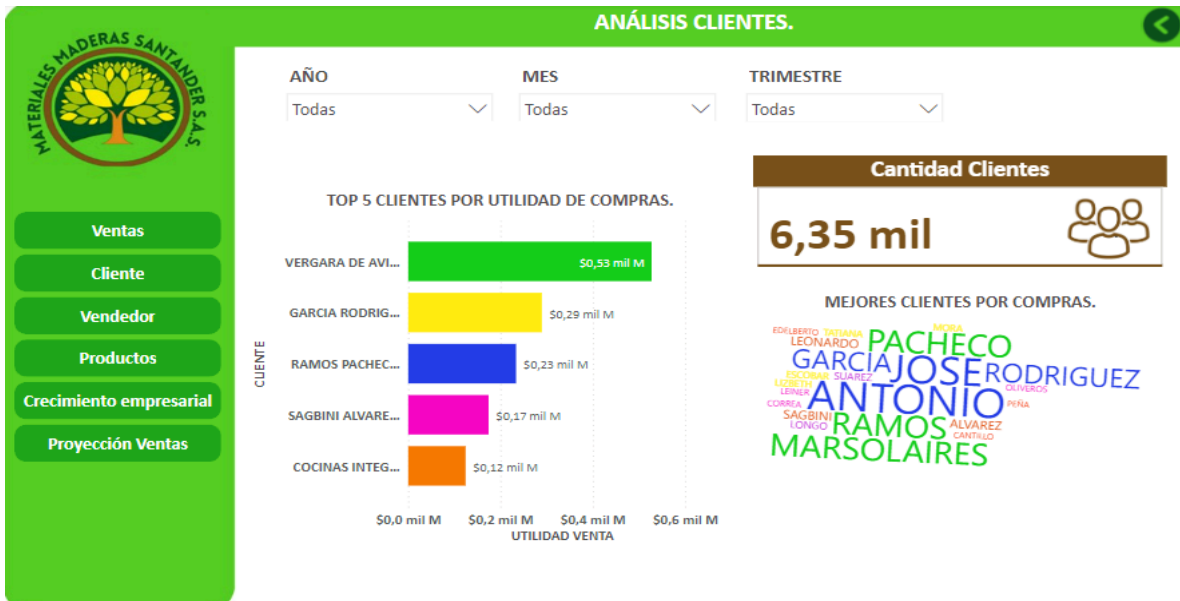
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 65 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Clientes (Análisis clientes).



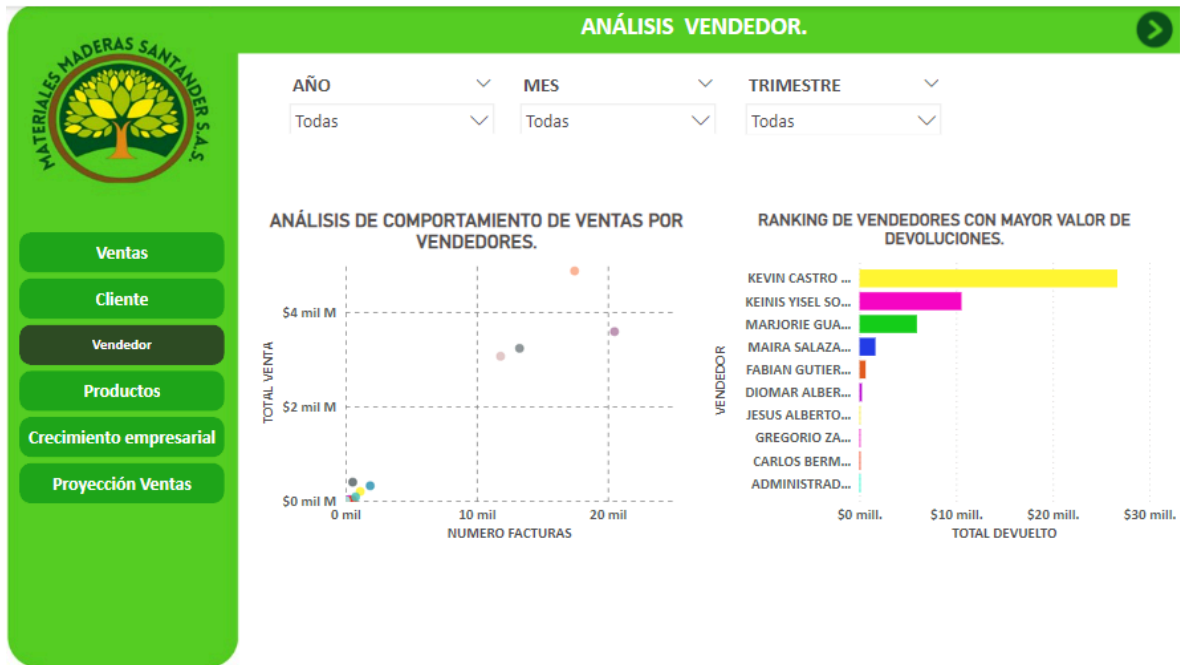
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 66 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Clientes (Análisis clientes).



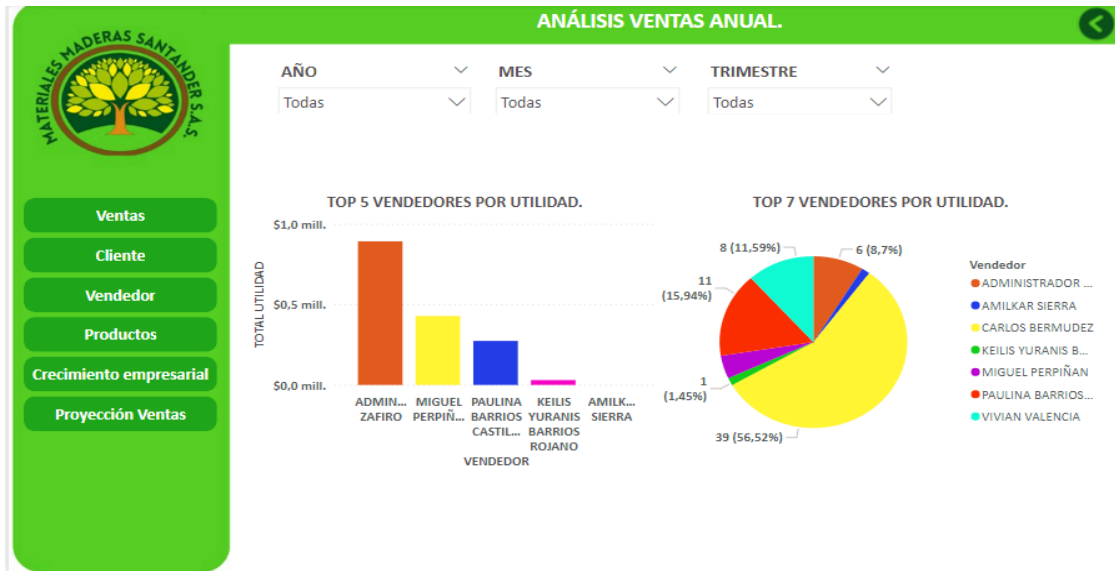
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 67 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Vendedores (Análisis vendedores).



Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 68 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Vendedores (Análisis vendedores).



Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 69 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).



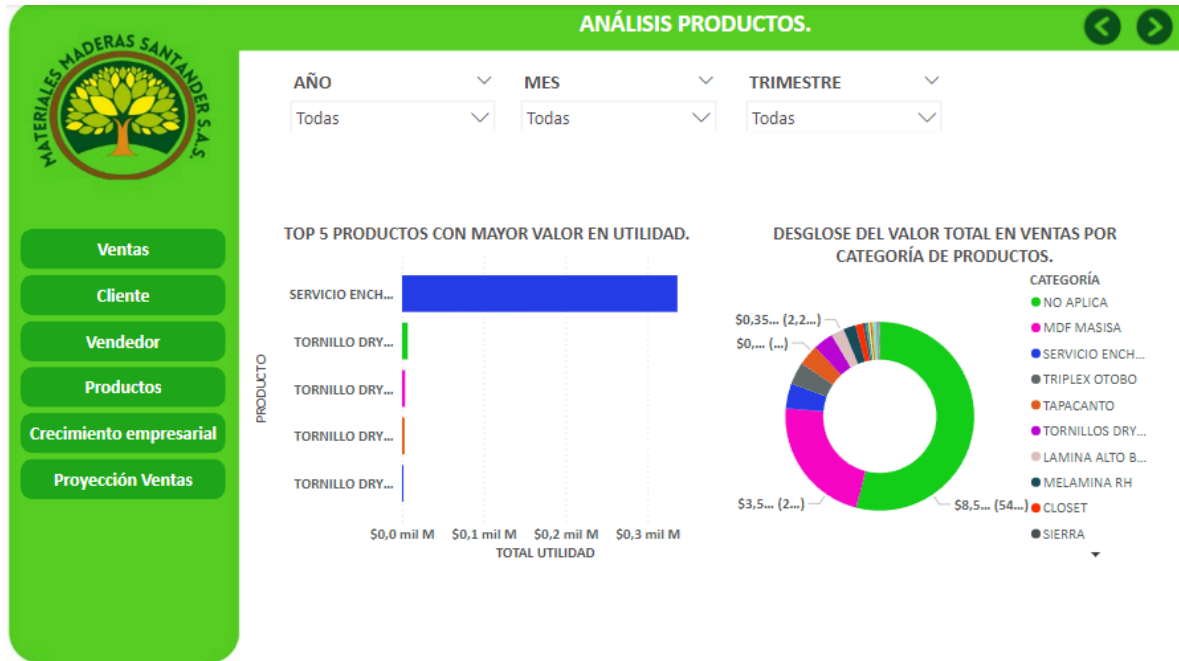
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 70 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).



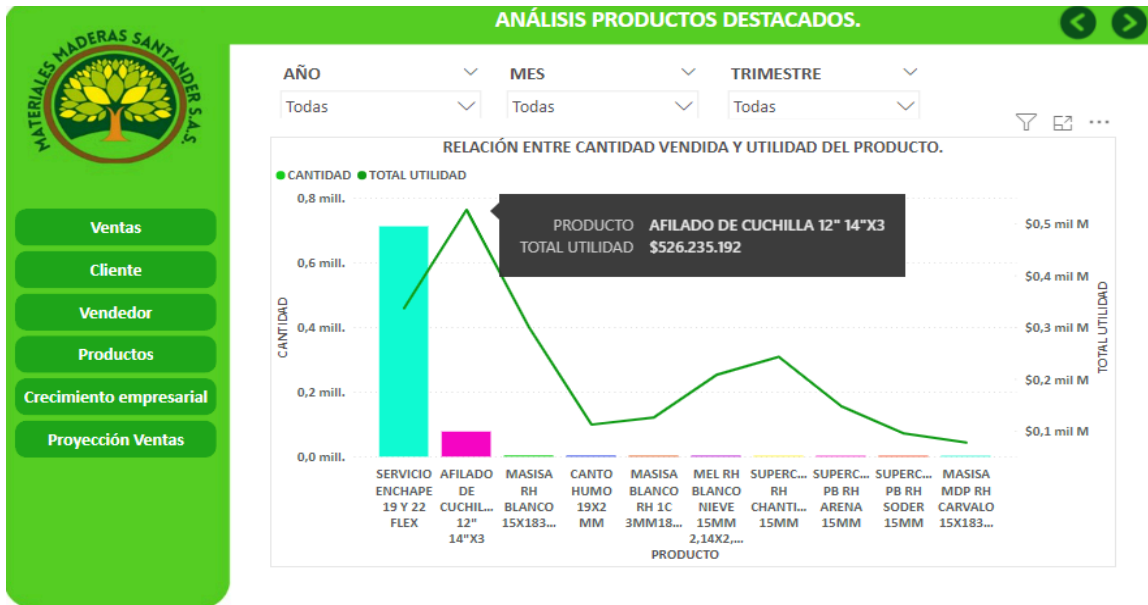
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 71 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).



Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 72 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).



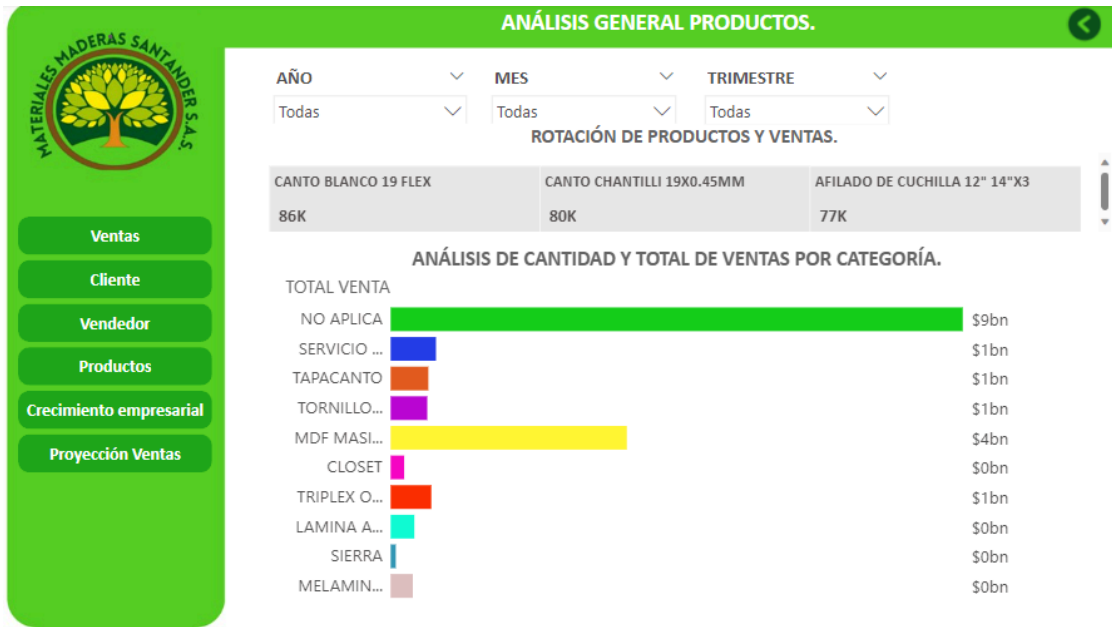
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 73 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis categoría y productos destacados).



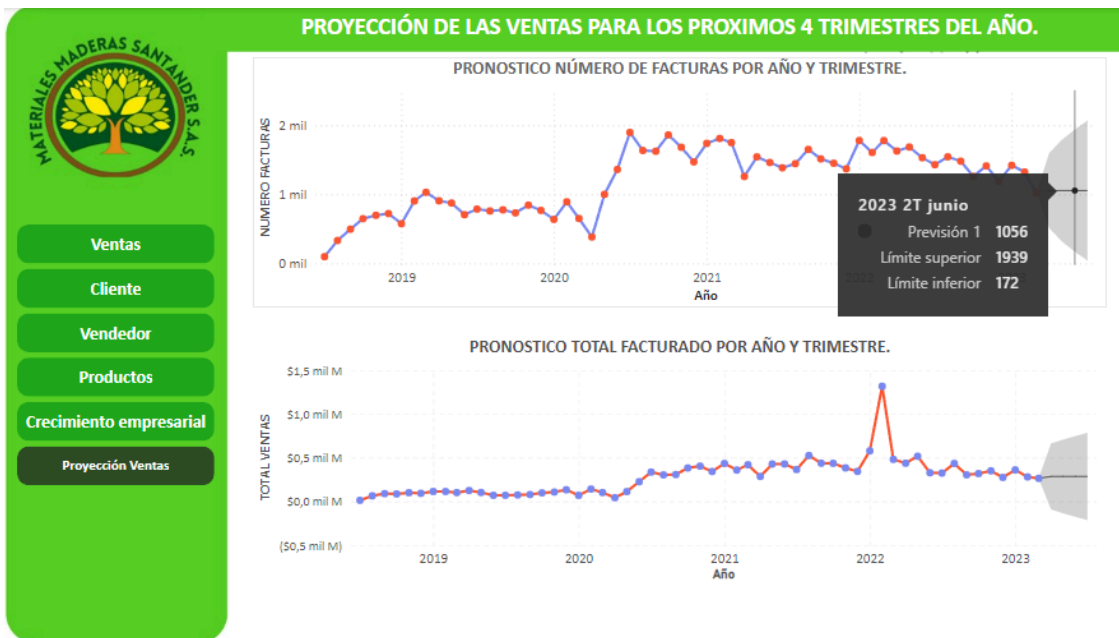
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 74 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S. Productos (Análisis productos).



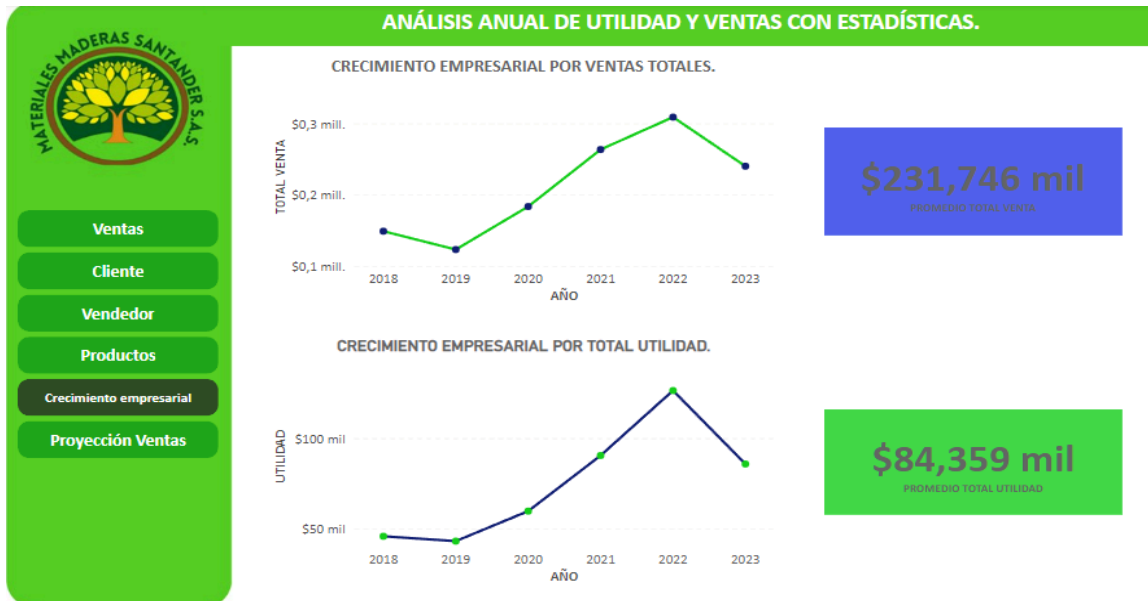
Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 75 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S (Proyección ventas).



Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 76 Visualización de los informes que muestran los resultados de la solución BI implementada en la empresa Materiales Maderas Santander S.A.S (Crecimiento empresarial).



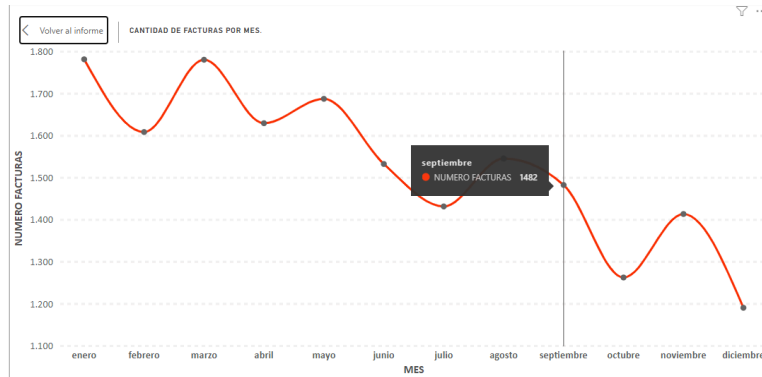
Autor: Macias Anyelyn, Ortiz Juan.

Como se puede apreciar en las ilustraciones de la 76 a la 86, se ha logrado la implementación de una solución de inteligencia de negocios en el área de ventas y compras de la empresa Materiales Maderas Santander. Esto proporciona a los administrativos una visualización general, clara e intuitiva de los procesos en estas áreas. Se han creado gráficos intuitivos, dinámicos y de fácil comprensión, acompañados de filtros respectivos que permiten a los usuarios interactuar y visualizar la información de manera conveniente. Este enfoque facilita la formulación de estrategias empresariales y contribuye a la toma de decisiones.

A continuación, se procede a dar respuestas a las preguntas formuladas por el gerente administrativo de la empresa.

¿Cómo se ha comportado el volumen de ventas en diferentes meses a lo largo del año?

Ilustración 77 Comportamiento de las ventas por meses.

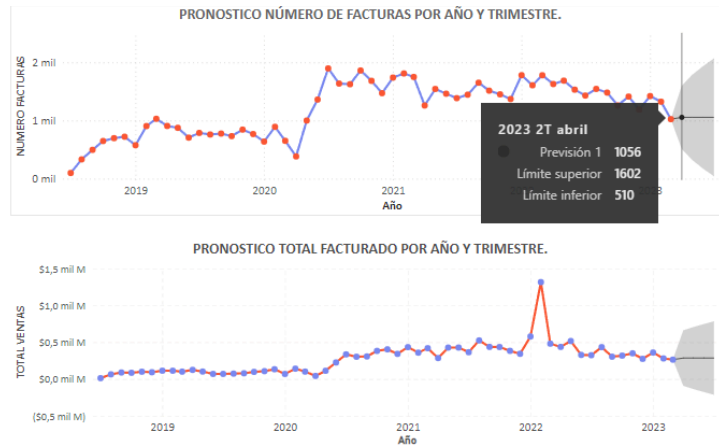


Autor: Macias Anyelyn, Ortiz Juan.

El gráfico presenta el volumen de ventas en relación con el número de facturas emitidas durante los meses del año 2022. Se puede ver que el mes con el mayor volumen de vetas fue enero, con un total de 1.190 facturas, mientras que diciembre registro el menor volumen de ventas, con un total de 1.190 facturas. El volumen de ventas experimento una tendencia decreciente desde septiembre hasta diciembre, mostrando una disminución gradual en la cantidad de facturas. Por otra parte, desde enero hasta agosto, se observa una tendencia creciente en el volumen de las ventas, indicando un aumento gradual en la cantidad de facturas. Por lo anterior, se puede decir que el volumen de ventas a lo largo de los meses del 2022 mostró variaciones, alcanzando su punto máximo en enero y disminuyendo en diciembre.

¿Cuál es la proyección de ventas para el segundo trimestre del 2023?

Ilustración 78 Proyección de ventas.

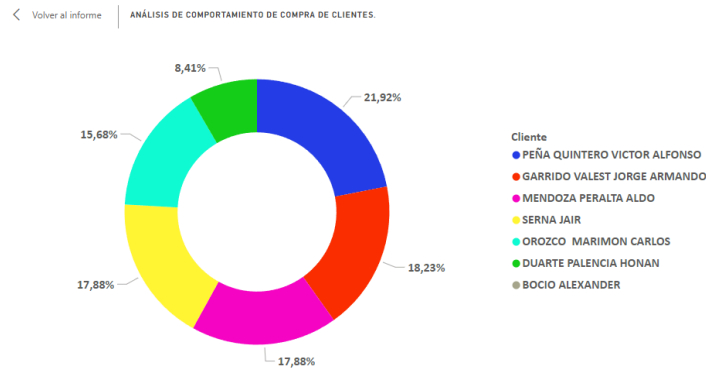


Autor: Macias Anyelyn, Ortiz Juan.

El gráfico muestra un pronóstico de facturación por año y trimestre basándose en el análisis de los datos históricos del trimestre anteriores, según este pronóstico se estima que el segundo trimestre del 2023 se generará alrededor de 1056 facturas con un margen de variación con un límite superior de 1602 facturas y un límite inferior de 510 facturas.

¿Quiénes son los clientes más destacados en función de sus compras?

Ilustración 79 Clientes destacados por compras.



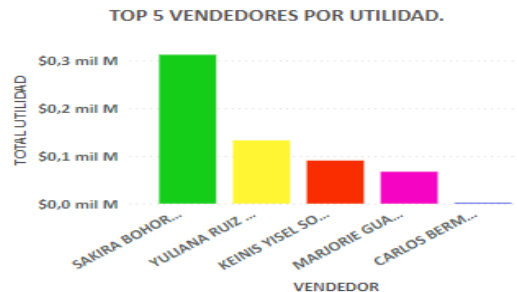
Autor: Macias Anyelyn, Ortiz Juan.

El gráfico muestra a los clientes con un mayor valor en sus compras. El cliente principal es Victor Peña Quintero, cuyo valor total en facturas asciende a 1,188, lo que representa un porcentaje del 21.92% del total de ventas. En segundo lugar, se encuentra Jorge Garrido

Valest, con un valor total de facturas de 988, lo que equivale al 18.23% del total de las ventas y así se puede ir analizando cada uno de los clientes que se muestran en la gráfica.

¿Quiénes son los vendedores más destacados en función de la utilidad de sus ventas?

Ilustración 80 Vendedores destacados por ventas.

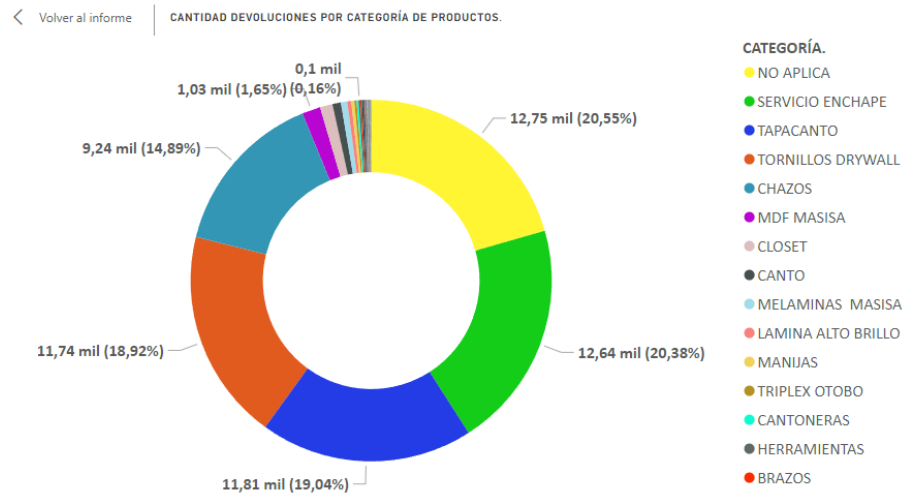


Autor: Macias Anyelyn, Ortiz Juan.

El gráfico de barras presenta el ranking de los cinco mejores vendedores en función de la utilidad que aportan a la empresa a través de las ventas que realizan. En el primer puesto se encuentra Sakira Bohorguez, quien ha generado una utilidad total de 312.421.155 millones. En el segundo lugar está Yuliana Ruiz, con un total de 132.760.340 millones. En el tercer lugar, se ubica Kenny Socara, con una contribución de 90.396.914 millones. En el cuarto puesto encontramos a Marjorie Guardela, con una utilidad de 66.939.157 millones, y finalmente, en el quinto lugar se sitúa Carlos Bermúdez, quien ha aportado un total de 1.367.409 millones.

¿Cuáles son las categorías de productos que presentan el mayor porcentaje de devolución?

Ilustración 81 Categorías con mayor devolución.

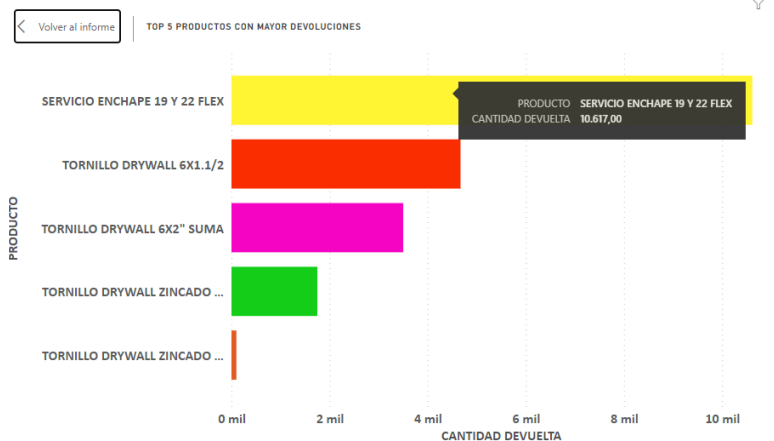


Autor: Macias Anyelyn, Ortiz Juan.

El análisis de gráfico revela que las categorías de productos con más devoluciones son las siguientes, primer lugar categoría no aplica que representa el 20.55% del total de las devoluciones, segundo lugar categoría servicios de enchape con un 20.38% del total de las devoluciones y en tercer lugar la categoría tapa caño con un 19.4% del total de las devoluciones.

¿Cuáles son los productos con el mayor porcentaje de devoluciones?

Ilustración 82 Productos con mayor devolución.

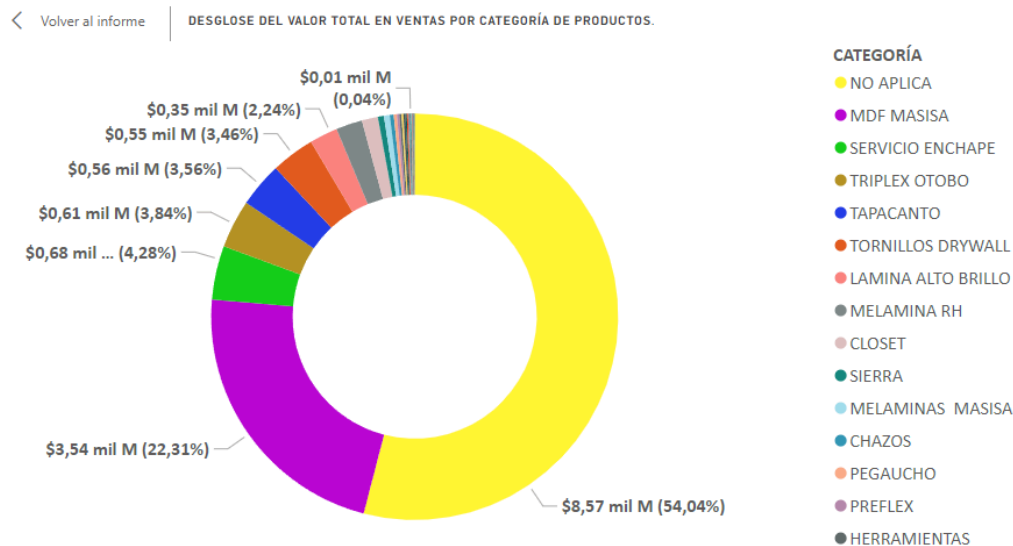


Autor: Macias Anyelyn, Ortiz Juan.

El análisis del gráfico muestra los productos con más devoluciones de la siguiente manera: en primer lugar, el producto "Servicio de enchape 19 y 22 flex" con un total de 10.617 devoluciones. En segundo lugar, se encuentra el producto "Tornillo drywall 6x1.1/2" con un total de 4.668 devoluciones. En tercer lugar, el producto "Tornillo drywall 6x2 suma" con un total de 3.500 devoluciones. En el cuarto lugar, está el producto "Tornillo drywall zincado 6x2 suma" con un total de 1.750 devoluciones, por último "Tornillo drywall zincado 6x1 suma" con un total de 100 devoluciones.

¿Cuáles son las categorías de productos que representan un mayor valor en relación con las ventas totales?

Ilustración 83 Categorías de productos que representan un mayor valor en relación con las ventas totales.

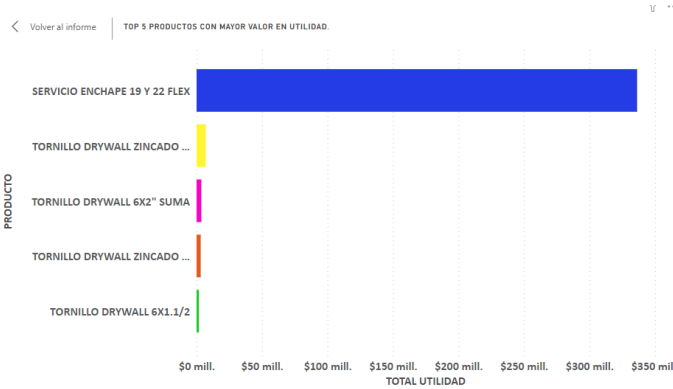


Autor: Macias Anyelyn, Ortiz Juan.

El gráfico destaca las categorías de productos que contribuyen significativamente a las ventas en todos los períodos analizados, es decir, aquellas categorías que aportan la mayor cantidad de ingresos a la empresa. En primer lugar, se encuentra la categoría "No Aplica," que representa el 54.04 % del total recaudado, equivalente a 8.567.201.504 millones. En segundo lugar, está la categoría "MDF masisa," con un porcentaje del 22.31 % y un total recaudado de 3.536.453.736 millones. El tercer puesto lo ocupa la categoría "Servicio de enchape", con un porcentaje del 4.28 % y un total recaudado de 678.683.182 millones.

¿Cuáles son los productos que representan un mayor valor en relación con la utilidad total de la empresa?

Ilustración 84 Productos con mayor utilidad.

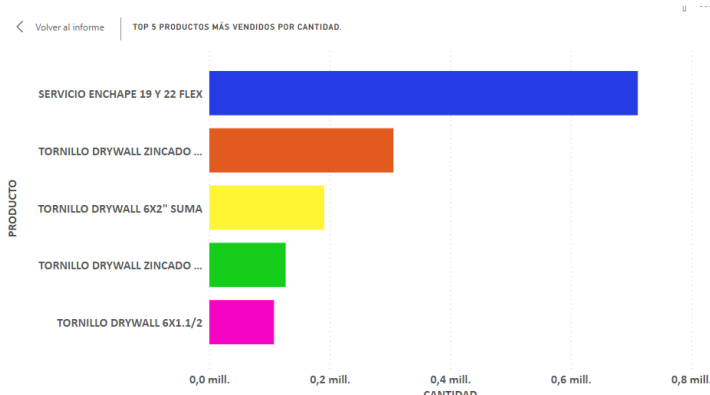


Autor: Macías Anyelyn, Ortiz Juan.

El gráfico de barras presenta los cinco productos más rentables para la empresa. En primer lugar, destaca el producto "Servicio de enchape 19 y 22 flex," con una utilidad total de 33.340.885 millones. En segundo lugar, se encuentra el producto "Tornillo drywall zincado 6x2 suma," con una utilidad de 6.958.775 millones. El tercer lugar lo ocupa el producto "Tornillo drywall 6x2 suma," con una utilidad total de 3.662.734 millones. En el cuarto lugar, se ubica el producto "Tornillo drywall zincado 6x1 suma" con una utilidad de 3.132.501 millones. Por último, el producto "Tornillo drywall 6 x 1.1/2" reporta una utilidad total de 1.713.153 millones.

¿Cuáles son los productos más vendidos por cantidad?

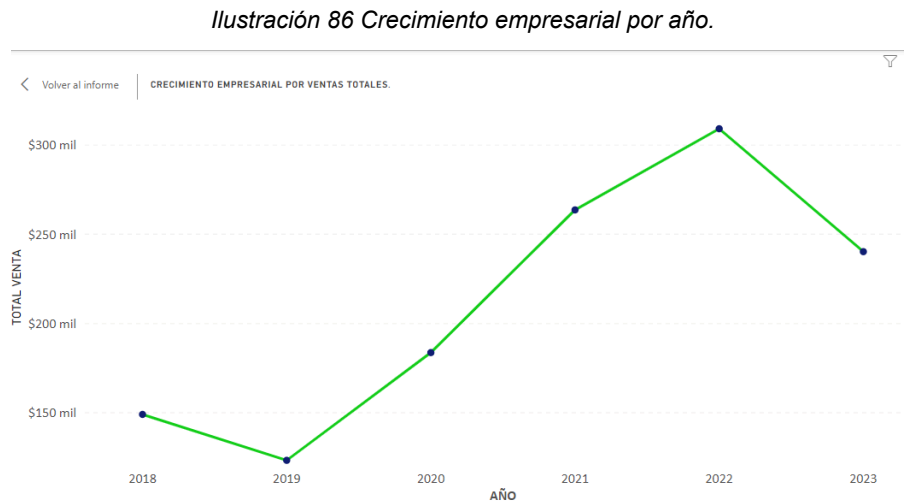
Ilustración 85 Productos más vendido por cantidad.



Autor: Macias Anyelyn, Ortiz Juan.

El gráfico ilustra los productos más destacados según su volumen de ventas. En la primera posición se encuentra el "Servicio de enchape 19 y 22 Flex," con un total de 711,315,000 unidades vendidas. En el segundo puesto, destacamos el producto "Tornillo drywall zincado 6x2 suma," con un total de 305,651,000 unidades vendidas. En el tercer lugar, encontramos el "Tornillo drywall 6x2 suma," que registra un total de 190,810 unidades vendidas. En el cuarto lugar, se ubica el producto "Tornillo drywall zincado 6x1 suma" con 126.658 unidades vendidas, y en la quinta posición se ubica el "Tornillo drywall 6 x 1.1/2", con un total de 107.350 unidades vendidas. Estos datos proporcionan una visión clara de los productos más demandados en función de su cantidad vendida.

¿Cómo ha sido el crecimiento de la empresa a través de los años?



Autor: Macias Anyelyn, Ortiz Juan.

El crecimiento de la empresa a lo largo de los años ha sido variable en relación con las ventas totales. En la gráfica, se aprecian los siguientes datos: en 2018, la empresa registró un rendimiento destacado con un volumen de ventas de 148,803 millones. Sin embargo, en 2019, experimentó una disminución del 17.3% en sus ventas. En el 2020 se logró una recuperación significativa con un crecimiento del 48.7% en su volumen de ventas, superando los 183.000 millones. En 2021, mantuvo su tendencia ascendente, alcanzando un nuevo

récord de 263,441 millones en ventas, lo que representa un aumento del 43.9% en comparación con 2020. En 2022, continuó su crecimiento y alcanzó el volumen de ventas más alto de toda la serie

histórica, con 308,956 millones. Sin embargo, en 2023, la empresa experimentó una disminución del 22.3% en su volumen de ventas, quedando en 240,038 millones.

En resumen, el crecimiento de la empresa a lo largo de los años ha sido en su mayoría positivo. Pero con algunos altibajos.

¿Cuál es la proyección de compras para el próximo mes?

Ilustración 87 Proyección de compras por producto.

The screenshot shows a web interface with a green header containing a tree logo and the text 'PROYECCIÓN DE COMPRAS MATERIALES MADERA SANTANDER'. Below the header is a dropdown menu labeled 'Seleccione un producto' with 'LAMINA TRIPLEX 4 MM ESPECIAL 122 X 244' selected. A note below the dropdown states: 'Nota: Esta proyección de ventas es una estimación con un 85% de confiabilidad, basada en datos históricos y tendencias actuales. Sin embargo, no garantizamos su exactitud al 100%. Recomendamos utilizarla como guía, estando preparados para ajustes según sea necesario.' Below the note is a green button labeled '▶ Ejecutar proyección'. At the bottom, there is a table with two columns: 'Mes' and 'Proyección'. The table contains one row with 'abr' in the 'Mes' column and '143.00' in the 'Proyección' column.

Mes	Proyección
abr	143.00

Autor: Macias Anyelyn, Ortiz Juan.

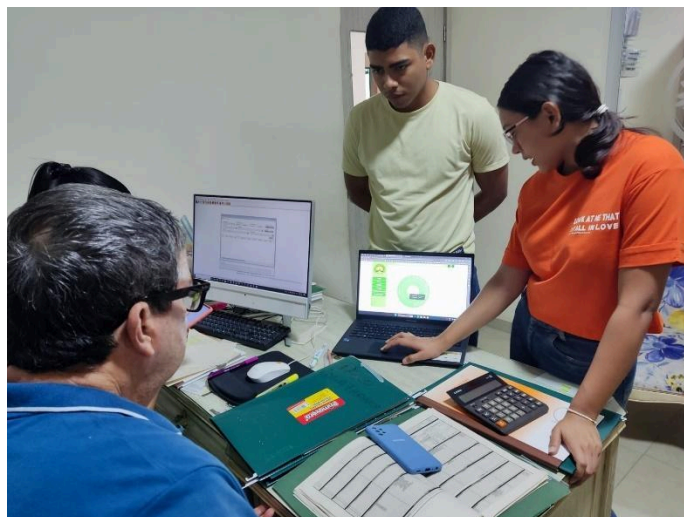
Al final se logró implementar la solución de inteligencia de negocio en la empresa Materiales Maderas Santander S.A.S., donde se realizó una exposición y capacitación sobre la solución al personal que se encarga de visualizar, analizar la información referente al área de ventas y compras como se observa en las ilustraciones 87 y 88.

Ilustración 88 Exposición y capacitación por parte Juan Ortiz y Anyelyn Macias a los administrativos de la empresa Materiales Maderas Santander.



Autor: Macias Anyelyn, Ortiz Juan.

Ilustración 89 Capacitación por parte Juan Ortiz y Anyelyn Macias al gerente general de la empresa Materiales Maderas Santander.



Autor: Macias Anyelyn, Ortiz Juan.

3.2 ANÁLISIS DE RESULTADOS Y DISCUSIÓN

Al inicio del desarrollo del proyecto, se identificó que la empresa materiales madera Santander se encontraba en la fase 3 del modelo de madurez de inteligencia de negocio. Esto se debe a que cuentan con un sistema de facturación que genera algunos informes, aparte se extraen datos directamente desde su base de datos fuente, que se analiza por medio de tablas y algunos gráficos no iterativos en la herramienta Excel. Todo esto se realiza para apoyarse en la toma de decisiones.

Durante el desarrollo del proyecto, se logró implementar una solución de inteligencia de negocios apoyada por una bodega de datos para el área de ventas y compras de la empresa. Este sistema permite la generación de informes y análisis detallados, así como la proyección de compras utilizando un algoritmo predictivo.

Los informes generados por medio del módulo de visualización revelan una serie de tendencias y patrones claves en las ventas y compras de la empresa. Además, la proyección de compra demostró ser una herramienta valiosa para optimizar el inventario, reduciendo los costos de almacenamiento. Con todo esto, la empresa puede visualizar de forma intuitiva, rápida y sencilla información relevante que antes no se podía identificar.

Por otro lado, la empresa enfrentaba dificultades para responder interrogantes complejos sobre hechos representativos. Esto se debía a la complejidad de extraer los datos de la base de datos fuente y pasarlos a Excel con el fin de realizar consultas y generar las tablas o gráficos. Este proceso resultaba complicado, demandando mucho tiempo, sin garantías de responder a todos los interrogantes planteados.

En consecuencia, al implementar una bodega de datos donde se puede ver la información a través de un módulo de visualización que genera informes y gráficas iterativas visualmente atractivas e intuitivas, le facilita a la empresa la interpretación de los datos. Además, al implementar el algoritmo de proyección para las compras se demostró que es de suma

importancia, dado que permite optimizar el inventario y reduce los costos de almacenamiento.

Todo esto hace posible que la empresa comprenda más a fondo la dinámica del mercado y permite a los directivos tomar decisiones informadas. También, da respuestas a preguntas complejas como: ¿Qué pasó?, ¿Qué pasa ahora?, ¿Por qué pasó? y ¿Qué pasará?

3.3 CONCLUSIONES

Actualmente, la inteligencia de negocio es esencial para las empresas porque les permite identificar oportunidades de negocio, tomar decisiones informadas basadas en datos reales y mejorar el rendimiento empresarial. Es crucial su implementación en la empresa Materiales Maderas Santander, porque les permitirá anticipar los cambios en el mercado, reducir los riesgos y mantenerse competitivos en su entorno empresarial que se vuelve cada vez más cambiante.

Al inicio del proyecto, se logró una comprensión profunda de las necesidades esenciales de la empresa. Esto permitió establecer los requerimientos del negocio y desarrollar un plan estratégico que abordó las necesidades identificadas de una manera efectiva y eficiente, con esto se dio cumplimiento al primer objetivo específico, analizar los requerimientos del negocio.

Seguidamente, se inició con el diseño de la arquitectura lógica y física de la bodega de datos, en conformidad con el segundo objetivo específico.

Posteriormente, se llevó a cabo definición del algoritmo de autoaprendizaje mediante pruebas, en línea con el tercer objetivo específico de determinar algoritmo de machine learning para la proyección de las compras. Todo esto con el fin de optimizar la eficiencia operativa, mejorar la calidad del análisis de los datos y garantizar una toma de decisiones más precisa fundamentada en la información.

Con el diseño de la arquitectura ya establecido, se procedió al desarrollo del proceso ETL, que tuvo una importancia significativa. Este proceso abarca la extracción de los datos de su base de datos fuente en un modelo transaccional, seguido por la transformación de los datos y, finalmente, su carga en la bodega de datos. Por otro lado, se procedió al desarrollo del algoritmo seleccionado, ARIMA. Teniendo en cuenta la necesidad de los datos históricos de la empresa para lograr el resultado deseado. Esto se basó en la comprensión previa de la importancia de la precisión en la predicción junto con la confiabilidad de los datos, teniendo

en cuenta la presencia constante de incertidumbre y variables imprevistas que pueden afectar los resultados en estos casos. También se consideró en el desarrollo de la interfaz que esta fuera sencilla, fácil de usar y amigable para los administrativos de la empresa.

Luego, se llevó a cabo el diseño y desarrolló del módulo de visualización en la herramienta Power Bi. Este módulo fue concebido para integrar una variedad de gráficos e información importante para la empresa, con el propósito de apoyar la toma de decisiones mediante una representación visual, clara y concisa de los datos. Cuenta con una interfaz intuitiva y fácil de usar, facilitando el entendimiento, el análisis de los datos, permitiendo así a los usuarios identificar tendencias y patrones de manera eficiente. Además, ofrece la flexibilidad necesaria para explotar los datos de manera detallada realizando comparaciones. Con esto, se cumplió el cuarto objetivo específico, diseñar el módulo de visualización.

Para finalizar, se logró alcanzar el objetivo general de implementar una solución de inteligencia de negocios para el área de ventas y compras en la empresa Materiales Maderas Santander. El proceso de desarrollo e implementación fue gratificante, también altamente beneficioso, puesto que se logró transformar significativamente la forma en que la empresa maneja sus datos y toma decisiones. Esta solución ha optimizado la comprensión de la información, mejorando la gestión del inventario con la implementación de la proyección de compras. Cada fase de este proyecto ha sido fruto del esfuerzo y la dedicación de todo el equipo, lo que ha finalizado en un resultado satisfactorio que no solo beneficiará a la empresa, sino que también ha fortalecido los conocimientos de los integrantes del proyecto.

3.4 RECOMENDACIONES

Teniendo en cuenta que el éxito de todo proyecto está vinculado a la implementación de buenas prácticas y recomendaciones que garanticen un buen funcionamiento, es fundamental considerar las siguientes recomendaciones en la implementación de la solución de inteligencia de negocios. Estas recomendaciones no solo contribuirán a un funcionamiento correcto, sino también a la capacidad de adaptarse al cambio, la posibilidad de crecer a futuro y completar implementación de las fases del modelo de inteligencia de negocio en la empresa.

- Se debe establecer un plan de mantenimiento para asegurar que la infraestructura de la solución de inteligencia de negocio este actualizada y funcionando correctamente.
- Se debe capacitar al personal nuevo que ingrese a la empresa relacionado con estas áreas para facilitar la comprensión de los datos, asegurando que todos los empleados involucrados en estos procesos estén familiarizados con la solución.
- Reunir la mayor información existente de los procesos de negocio para la posibilidad a futuro de expandir y completar el modelo de inteligencia de negocio en la empresa.
- Seguir buenas prácticas al realizar los procesos ETL a futuro para que se pueda procesar la información de manera más eficiente mientras se mantiene la calidad de los datos

3.5 BIBLIOGRAFÍA

- [1] Ikusi, “Tecnologías de la información en las empresas: ¿cómo impacta el negocio?”, Ikusi, 04-feb-2021. [En línea]. Disponible en: <https://www.ikusi.com/mx/blog/tecnologias-de-la-informacion-en-las-empresas-como-impacta-el-negocio/>. [Consultado: 25-may-2023].
- [2] D. Terreros, “Sistemas de información en las empresas: importancia, tipos y tips para tu negocio”, Hubspot.es, 26-may-2021. [En línea]. Disponible en: <https://blog.hubspot.es/marketing/sistemas-de-informacion-empresas>. [Consultado: 25-may-2023].
- [3] “¿Qué Es Business Intelligence? Introducción A Los Sistemas De BI (01)”, Visionarios, 02-jun-2017.
- [4] Jordi Conesa Caralt (coord.) Josep Curto Díaz, Introducción al Business Intelligence.
- [5] Colaboración, “KPI’s ¿Qué son, para qué sirven y por qué y cómo utilizarlos?”, Logicalis.com. [En línea]. Disponible en: <https://blog.es.logicalis.com/analytics/kpis-qu%C3%A9-son-para-qu%C3%A9-sirven-y-por-qu%C3%A9-y-c%C3%B3mo-utilizarlos>. [Consultado: 25-may-2023].
- [6] “¿Qué es un almacén de datos?”, Oracle.com. [En línea]. Disponible en: <https://www.oracle.com/co/database/what-is-a-data-warehouse/>. [Consultado: 25-may-2023].
- [7] “¿Qué es el OLAP? - Explicación del procesamiento analítico en línea - AWS”, Amazon.com. [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/olap/>. [Consultado: 25-may-2023].
- [8] O. Vs y D. Oltp, “El modelo multidimensional El modelo multidimensional Data Warehousing Warehousing Warehousing”, Ugr.es. [En línea]. Disponible en: <http://elvex.ugr.es/idbis/db/docs/intro/f%20modelo%20multidimensional.pdf>. [Consultado: 25-may-2023].
- [9] “¿Qué es un data mart? - Explicación de los data marts - AWS”, Amazon.com. [En línea]. Disponible en: <https://aws.amazon.com/es/what-is/data-mart/>. [Consultado: 25-may-2023].
- [10] E. Bello, “Guía de Procesos ETL: Qué son, cómo usarlos y herramientas clave”, Thinking for Innovation, 2022.
- [11] “Extracción, transformación y carga de datos (ETL)”, Microsoft.com. [En línea]. Disponible en:

<https://learn.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl>.

[Consultado: 25-may-2023].

[12] “2022 Gartner® Magic Quadrant™ for Data Integration Tools | Qlik®,” Qlik, 2022. <https://www.qlik.com/es-es/gartner-magic-quadrant-for-data-integration-tools> (accessed May 25, 2023).

[13] D. Ortiz y Cyberclick, “¿Qué es un dashboard y para qué se usa? (2023),” Cyberclick.es. [En línea]. Disponible en: <https://www.cyberclick.es/numerical-blog/que-es-un-dashboard>. [Consultado: 25-may-2023].

[14] “Cuadrante mágico de Gartner 2023 para BI y análisis | Qlik,” Qlik, 2023. <https://www.qlik.com/es-es/gartner-magic-quadrant-business-intelligence> (accessed May 25, 2023).

[14] E. Bello, “¿Qué es Microsoft Power BI? Todo lo que tienes que saber,” Thinking for Innovation, 2022.

[15] “Tableau Software - Herramienta de Visualización de Datos,” Neteris, 01-Jul-2019. [Online]. Available: <https://neteris.com/software/tableau-software-visualizacion-datos/>. [Accessed: 25-May-2023].

[16] S. A. Lempert, “¿Qué es Qlik Sense y para qué sirve?,” Lempert S.A. - Soluciones Tecnológicas, 18-Oct-2022.

[17] Org.sv. [Online]. Available: http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf. [Accessed: 25-May-2023].

[18] J. Lilian and Sandoval, “ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS MACHINE LEARNING ALGORITHMS FOR DATA ANALYSIS AND PREDICTION A A Palabras clave Keyword.” Available: http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf

[19] petrogli, “SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA SOPORTE DE DECISIONES EN EMPRESAS DE COMERCIALIZACIÓN DE MERCANCÍA,” Petroglifos Revista Crítica Transdisciplinaria, Dec. 18, 2020. <https://petroglifosrevistacritica.org.ve/revista/solucion-de-inteligencia-de-negocios-para-soporte-de-decisiones-en-empresas-de-comercializacion-de-mercancia/> (accessed May 25, 2023).

[20] Pereda Medina Cesar Antonio Cabrera Sanchez Martin Wilfredo, “Solución de inteligencia de negocios (bi) para mejorar el análisis de la información en los procesos de

ventas de la empresa Ingenieros en accion S.R.L. utilizando la arquitectura de PENTAHO BI” Universidad Privada Antenor Orrego, Perú, 2019.

[22] Magali Yilena Álvarez Rodríguez Paola Andrea Gómez Solórzano, “Modelo de Inteligencia de Negocios para la empresa Puntual Arquitectura Ingeniería,” Universidad EAN, Bogotá, Colombia, 2022.

[23] Jhon Edison Guzman Arias Carlos Enrique Arenales Rincón, “Diseño de un modelo de inteligencia de negocios para empresas del sector de desarrollo software en Colombia que fortalezca la toma de decisiones a nivel organizacional,” Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, 2022.

[24] W. Araque, R. Mora, A. Victoria, and Jasmidt Vera Cuenca, “INTELIGENCIA DE NEGOCIOS ADAPTATIVOS APLICADA A LA EMPRESA DE SERVICIOS PÚBLICOS E.S.P EMSERPUCAR DEL MUNICIPIO DE CARTAGENA DEL CHAIRÁ” ResearchGate, Jan. 29, 2021.

https://www.researchgate.net/publication/354255065_INTELIGENCIA_DE_NEGOCIOS_ADAPTATIVOS_APLICADA_A_LA_EMPRESA_DE_SERVICIOS_PUBLICOS_ESP_EMSEMPUCAR_DEL_MUNICIPIO_DE_CARTAGENA_DEL_CHAIRA (accessed May 25, 2023).

[25] WEB, “Metodología de Kimball,” Blogspot.com, May 25, 2023. <http://inteligenciadenegociosval.blogspot.com/2014/01/metodologia-de-kimball.html> (accessed May 25, 2023).

[26] M. Ross, “Data warehouse fácil: Consejo de diseño #115: El ciclo de vida Kimball resumido,” Datawarehouse.es. [Online]. Available: <https://datawarehouse.es/2009/ciclo-de-vida-kimball.html>. [Accessed: 25-May-2023].

[27] C. Catalán, “SERIES TEMPORALES.” Available: http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/SeriesTemporales.pdf

[28] F. Sanz, “Series Temporales - Introducción a esta técnica de predicción,” The Machine Learners, Sep. 29, 2020. <https://www.themachinelearners.com/series-temporales-intro/> (accessed May 25, 2023).

[29] “Lenguaje de Programación - Concepto, tipos y ejemplos,” Concepto, 2013. <https://concepto.de/lenguaje-de-programacion/> (accessed May 25, 2023).

[30] Javier Canales Luna, “Top programming languages for data scientists in 2023,” Datacamp.com, Jan. 21, 2022. <https://www.datacamp.com/blog/top-programming-languages-for-data-scientists-in-2022> (accessed May 25, 2023).

[31] G. Rivadera, "La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)." Available: http://bibliotecas.ucasal.edu.ar/opac_css/60536/678/67860536.pdf