



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Implementación de business intelligence utilizando minería de datos
para la toma de decisiones en las ventas

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero de Sistemas

AUTOR:

Tantalean Salazar, Billy Brandon (orcid.org/0000-0002-7143-360X)

ASESOR:

Dr. Necochea Chamorro, Jorge Isaac (orcid.org/0000-0002-3290-8975)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2023

DEDICATORIA

Primeramente, a Dios que me ha llevado por el buen camino gracias a la fortaleza en cada etapa de mi vida hasta este momento. También se los dedico a mis padres que me han brindado su apoyo absoluto para ayudarme a llegar a donde estoy hoy.

AGRADECIMIENTO

Primeramente, agradecer a Dios por haberme guiado por el buen camino, y así poder llegar hasta este momento. También agradezco a mis padres que me han brindado su apoyo absoluto en cada momento, y así haber completado mis metas trazadas.

ÍNDICE DE CONTENIDO

CARÁTULA	i
DEDICATORIA	ii
AGRADECIMIENTO.....	iii
ÍNDICE DE CONTENIDO.....	iv
ÍNDICE DE TABLAS	v
ÍNDICE DE GRÁFICOS Y FIGURAS	vi
RESUMEN	vii
ABSTRACT	viii
I. INTRODUCCIÓN	09
II. MARCO TEÓRICO	12
III. METODOLOGÍA.....	22
3.1. Tipo y diseño de investigación	22
3.2. Variables y operacionalización	23
3.3. Población, muestra, muestreo, unidad de análisis	24
3.4. Técnicas e instrumentos de recolección de datos.....	25
3.5. Procedimientos	28
3.6. Método de análisis de datos	29
3.7. Aspectos éticos.....	31
IV. RESULTADOS	32
V. DISCUSIÓN	49
VI. CONCLUSIONES	51
VII.RECOMENDACIONES.....	52
REFERENCIAS.....	53
ANEXOS	58

Índice de Tablas

Tabla 1: Nivel de confiabilidad Kappa	27
Tabla 2: Confiabilidad indicador crecimiento de ventas	28
Tabla 3: Confiabilidad indicador crecimiento de clientes	28
Tabla 4: Confiabilidad indicador índice de rotación de inventario.....	29
Tabla 5: Confiabilidad indicador eficiencia de pedidos.....	29
Tabla 6: Indicador CV en Pre-Test y Post-Test.....	34
Tabla 7: Indicador EP en Pre y Post-Test	35
Tabla 8: Indicador CC en Pre y Post-Test.....	36
Tabla 9: Indicador IRI en Pre y Post-Test	37
Tabla 10: Prueba de Normalidad de CV en pre y post-test.....	39
Tabla 11: Prueba de Normalidad de EP en pre y post-test	40
Tabla 12: Prueba de Normalidad de CC en pre y post-test.....	42
Tabla 13: Prueba de Normalidad de IRI en pre y post-test	43
Tabla 14: Prueba T-Student Crecimiento de ventas.....	45
Tabla 15: Prueba T-Student Eficiencia de Pedidos	46
Tabla 16: Prueba T-Student Crecimiento de Clientes	48
Tabla 17: Prueba T-Student Índice de Rotación de Inventario.....	49

Índice de Figuras

Figura 1: Business Intelligence para reforzar la gestión de la organización	17
Figura 2: Arquitectura de Business Intelligence	18
Figura 3: Interfaz Weka	20
Figura 4: Fases de CRISP-DM.....	21
Figura 5: Diseño de Investigación	23
Figura 6: Crecimiento de ventas antes y después de la implementación.....	35
Figura 7: Eficiencia de Pedidos antes y después de la implementación	36
Figura 8: Crecimiento de clientes antes y después de la implementación	37
Figura 9: Índice de rotación de inventario antes y después de la implementación...	38
Figura 10: Prueba de normalidad de Crecimiento de ventas Pre-Test.....	39
Figura 11: Prueba de normalidad de Crecimiento de ventas Post-Test	40
Figura 12: Prueba de normalidad de eficiencia de pedidos Pre-Test.....	41
Figura 13: Prueba de normalidad de eficiencia de pedidos Post-Test	41
Figura 14: Prueba de normalidad de crecimiento de clientes PreTest	42
Figura 15: Prueba de normalidad de crecimiento de clientes PostTest.....	43
Figura 16: Prueba de normalidad de Índice de rotación de inventario PreTest....	44
Figura 17: Prueba de normalidad de Índice de rotación de inventario PostTest ..	44
Figura 18: Región de la Prueba T-Student de CV	46
Figura 19: Región de la Prueba T-Student de EP	47
Figura 20: Región de la Prueba T-Student de CC.....	48
Figura 21: Región de la Prueba T-Student de IRI	50

RESUMEN

La presente tesis, tuvo como objetivo principal evaluar cómo la implementación de Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas, debido a que la empresa no cuenta con herramientas necesarias que le ayuden a analizar sus registros de datos guardados tanto el programa Excel, como boletas o cuadernillos.

Se optó por utilizar la metodología CRISP-DM debido a su capacidad para ofrecer una perspectiva inicial sobre el negocio y su alineación con el Business Intelligence. Para llevar a cabo este estudio, se emplearon varios programas, incluyendo SQL Server para la gestión de la base de datos, WEKA para realizar la minería de datos y Power BI Desktop para la visualización general de los datos de la empresa.

El tipo de investigación es aplicada con un diseño Pre- Experimental y con un enfoque cuantitativo. La población se realizó de los reportes diarios de ventas del negocio por 15 días junto a 20 productos de catálogo de construcción, proporcionando así los datos necesarios para los indicadores ya mencionados.

Finalmente, los resultados indican que el Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas, llegando a satisfacer las necesidades de la empresa.

Palabras Clave: Business Intelligence, ventas, toma de decisiones, minería de datos

ABSTRACT

The main objective of this thesis was to evaluate how the implementation of Business Intelligence using data mining for decision-making improves sales, because the company does not have the necessary tools to help it analyze its data records stored in Excel programs such as reports or notebooks.

The CRISP-DM methodology was chosen because of its ability to provide an initial perspective on the business and its alignment with Business Intelligence. To carry out this study, several programs were used, including SQL Server for database management, WEKA to perform data mining and Power BI Desktop for general visualization of the company's data.

The type of research is applied with a Pre-Experimental design and with a quantitative approach. The population was made from the daily sales reports of the business for 15 days along with 20 construction catalog products, thus providing the necessary data for the aforementioned indicators.

Finally, the results indicate that Business Intelligence using data mining for decision-making improves sales, thus meeting the needs of the company.

Keywords: Business Intelligence, sales, decision making, data mining.

I. INTRODUCCIÓN

En la actualidad, la toma de decisiones estratégicas en ventas es muy esencial para lograr un éxito en el negocio. A medida que avanza la tecnología y aumenta la cantidad de datos en la era digital, existen grandes desafíos en cuanto a cómo las empresas pueden utilizar esta información para mejorar su competitividad.

En este contexto, la implementación de BI mediante minería de datos puede ser especialmente útil. Las organizaciones pueden hacer uso de la tecnología de minería de datos para analizar las ventas y descubrir patrones que les permitan tomar buenas decisiones informadas sobre los nuevos productos o los existentes y así generar una expansión de nuevos mercados.

Además, la minería de datos es significativa para las empresas al mejorar el rendimiento de sus ventas, detectar posibles oportunidades de ventas adicionales y analizar a la competencia. En definitiva, la implementación de Business Intelligence con técnicas de minería de datos puede contribuir a que las empresas tomen decisiones más acertadas, aumenten sus ingresos y mejoren su rentabilidad.

Un aspecto relevante es la realización de análisis en tiempo real proporcionada por la inteligencia empresarial (BI) y la minería de datos. Las compañías pueden supervisar de cerca las ventas y mitigar prontamente los cambios dados en el mercado, adaptando sus estrategias en consecuencia. Esto resulta especialmente valioso en entornos altamente dinámicos y competitivos, donde tomar decisiones ágiles puede determinar el éxito del negocio.

Este estudio se llevará a cabo en la Distribuidora San Cristóbal, una empresa que opera en el sector de la comercialización de materiales de construcción y ofrece servicios de alquiler de maquinaria pesada especializada en la limpieza de terrenos, eliminación de escombros y tareas relacionadas.

En la actualidad, la compañía carece de herramientas que pueda apoyar en el registro de sus operaciones de ventas, que suelen estar almacenadas en hojas de Excel, la cual, provoca un control y monitoreo inadecuado, ya que, al momento de consultar la información requerida, el tiempo no es el adecuado, provocando dificultades en la obtención de reportes o informes, asimismo no hay un control en

la gran cantidad de datos almacenados, lo que genera una deficiente toma de estrategias.

Estas estrategias empresariales se hacen de acuerdo a los reportes obtenidos, como mencionado anteriormente, su realización no es la adecuada, lo que puede llevar a tomar decisiones inexactas o incompletas. Asimismo, resulta crucial tener en cuenta que los informes pueden contener sesgos o errores, lo que puede provocar de manera negativa en la calidad de las decisiones adoptadas.

El proceso manual de ingresar los datos en hojas de Excel es propenso a errores y consume mucho tiempo, lo que impide que los trabajadores se concentren en otras actividades importantes. Otro desafío es la falta de integración de los datos de ventas con otros datos que se utilizan. Esto complica la capacidad de ver y analizar los datos de las ventas en conjunto con otros aspectos del negocio, como el inventario, los proveedores o los gastos. La falta de integración limita la capacidad de obtener un enfoque más general y preciso de la situación de la empresa.

Por lo tanto, debe ser implementada en la empresa, ya que, la Inteligencia de Negocio (BI) utiliza todos los datos generados y presenta informes informativos que logran ser de mayor importancia al momento de las decisiones empresariales [1]. Además, la minería de datos pretende captar patrones o modelos a partir de un modelo o conjunto de datos de forma masiva, utilizando algoritmos inteligentes que aprenden con ejemplos o experiencia y así extraer conocimientos valiosos para el negocio [2].

La investigación plantea el siguiente problema general: ¿En qué forma la implementación de Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas?, en este contexto, se identifican los específicos. Comenzando con el primero: ¿En qué forma la implementación de Business Intelligence utilizando minería de datos ayuda a aumentar las ventas?, segundo específico: ¿En qué forma la implementación de Business Intelligence utilizando minería de datos ayuda en la eficiencia de pedidos de las ventas?, tercer específico: ¿En qué forma la implementación de Business Intelligence utilizando minería de datos ayuda a aumentar los clientes en las ventas? y el cuarto específico: ¿En qué

forma la implementación de Business Intelligence utilizando minería de datos ayuda en el índice de rotación de inventario para las ventas?.

La implementación de business intelligence utilizando minería de datos se realiza para mejorar las ventas en la Distribuidora San Cristóbal, y así obtener una mejor toma de decisiones utilizando herramientas de monitoreo y visualización, la cual obtendrá un mejor rendimiento en sus procesos y reportes, agilizando la información de manera efectiva. Por ello, se busca que utilizando estas herramientas se logre mejorar la eficiencia de pedidos, el índice de rotación de inventario, crecimiento en las ventas y clientes.

De acuerdo a lo explicado anteriormente, se formuló el siguiente objetivo general: Evaluar cómo la implementación de Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas y cuatro objetivos específicos, el primero es: Aumentar las ventas mediante la implementación de Business Intelligence utilizando minería de datos, el segundo es: Mejorar la eficiencia de pedidos mediante la implementación de Business Intelligence utilizando minería de datos en ventas, el tercero es: Aumentar los clientes mediante la implementación de Business Intelligence utilizando minería de datos en ventas y el cuarto es: Mejorar el índice de rotación de inventario mediante la implementación de Business Intelligence utilizando minería de datos en ventas

Por lo tanto, se definió la siguiente hipótesis general: El Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas, junto con cuatro hipótesis específicas, la primera es: El Business Intelligence utilizando minería de datos mejora el crecimiento de ventas, la segunda es: El Business Intelligence utilizando minería de datos mejora la eficiencia de pedidos en las ventas, la tercera es: El Business Intelligence utilizando minería de datos mejora el crecimiento de clientes y la cuarta es: El Business Intelligence utilizando minería de datos mejora el índice de rotación de inventario para las ventas.

II. MARCO TEÓRICO

Se analizaron algunos estudios relacionados al tema que fueron considerados como antecedentes, las cuales comienzan con 5 internacionales y 5 nacionales.

En el siguiente artículo, se mencionó que las pequeñas empresas pueden enfrentarse a problemas si no invierten en la creación de nueva tecnología, el monitoreo de la satisfacción del cliente y la rentabilidad. Es necesario utilizar las tecnologías de la información empresarial para mejorar su rendimiento, aumentar sus posibilidades de supervivencia y crecimiento. Para obtener información relevante que les permita tomar decisiones fundamentadas a través del análisis de datos, es importante que estas empresas utilicen organizaciones de investigación externas especializadas para realizar encuestas periódicas de muestras representativas de clientes. Como conclusión, se vio que muchas pequeñas empresas no recopilan los datos que necesitan para administrar sus negocios con éxito porque no utilizan estas tecnologías. Por otro lado, se tuvieron como resultados que las empresas pequeñas, no utilizan tecnologías para recopilar los datos que necesitan para administrar con éxito su negocio, también es necesario ver la satisfacción del usuario, junto con la rentabilidad, ya que son dos aspectos fundamentales que las empresas deberían monitorear continuamente, pero que a menudo se descuidan [3].

En la investigación, se desarrolló e implementó un software de soporte a la decisión de marketing basado en tecnología de minería de datos, que tiene como objetivo solucionar las debilidades actuales de los sistemas. Este sistema se encarga de la integración de datos, el análisis histórico y las funciones de consulta, y además ofrece un modelo basado en redes neuronales para predecir ventas y mejorar las estrategias en las decisiones. El análisis realizado del estudio permitió identificar los límites válidos. En conclusión, el modelo basado en redes neuronales usado para el pronóstico de ventas demostró ser efectivo cuando se probó con datos reales y se comparó con otros métodos existentes. Por otro lado, se tuvieron como resultados que el sistema sugerido tiene el potencial de respaldar a las empresas en la toma de decisiones respaldadas por datos más precisos y basadas en la ciencia. Además, se destaca que la precisión de las predicciones de ventas y

beneficios generadas por este sistema supera a la de los enfoques tradicionales de pronóstico, lo que facilita la identificación de patrones y tendencias en los datos [4].

En el siguiente estudio, se investigó que para cualquier empresa es crucial contar con una gestión y crecimiento eficaz de los recursos humanos en su operación y gestión. De lo contrario, una gestión inadecuada puede resultar en ineficiencias, mala toma de decisiones y un rendimiento comercial disminuido. Por ello, esta medida posibilitará abordar de forma global los desafíos relacionados con la gestión de recursos humanos. Como conclusión, el sistema pudo mejorar significativamente la informatización comercial, la estandarización de datos, el análisis inteligente de decisiones, mejorar la eficiencia empresarial, proporcionar apoyo a las estrategias empresariales y fomentar el crecimiento empresarial. Por otro lado, se tuvieron como resultados que el sistema respaldado por la minería de datos puede potenciar la eficacia y eficiencia al momento de administrar los recursos humanos, ofreciendo un análisis inteligente y mejorando la optimización de los procesos empresariales. Asimismo, se destacó que la tecnología de minería de datos brinda la capacidad de analizar extensos conjuntos de datos de recursos humanos, extrayendo patrones y conocimientos valiosos [5].

En el artículo, se tuvo como objetivo analizar cómo la herramienta de Business Intelligence contribuye a las estrategias empresariales y como comprende sus procesos, seleccionaron a 73 personas de diferentes edades para completar una encuesta sobre las compras de alimentos de sus productos. La investigación se basa en cómo esta herramienta ayuda a las empresas comerciales a integrar mejor sus datos y mejorar su posición en el mercado, creando así nuevas estrategias de toma de decisiones. La encuesta publicada muestra que esta herramienta de negocios genera costos y beneficios para las empresas comercializadoras, ya que las industrias centradas en los alimentos son de mayor importancia a nivel local y mundial. Por otro lado, se tuvo como resultado que es importante con el propósito de optimizar las operaciones internas y asegurar la sostenibilidad en el sector de la alimentación en Europa [6].

Los autores en su artículo, tuvieron como objetivo general investigar el uso de la inteligencia comercial como una herramienta de gran importancia para que las PYMES usen sus datos como entrada para respaldar la toma de decisiones

comerciales. Utilizaron diferentes bases de datos para encontrar estudios sobre este tema y así compararon los resultados, haciendo diferentes aportes a la investigación. Se ha comprobado que esta herramienta ayuda y moldea la toma de decisiones inteligentes, ya que la información se convierte en el valor clave que impulsa los cambios en la empresa. Por otro lado, se tuvo como resultado que la implementación de inteligencia de negocios representa una opción valiosa y factible para las empresas, mostrando resultados muy favorables de acuerdo a los datos que se analizan, brindando así soluciones muy estratégicas [7].

El siguiente artículo científico, tuvo como propósito usar métodos de minería de datos es poder identificar patrones y relevancia en los datos del sistema SCADA, donde la generación de informes sobre el apagado del turbogenerador hidráulico tiene como objetivo incrementar la eficiencia y reducir el tiempo de inacción. La metodología empleada es el modelo CRISP-DM para desarrollar un modelo de decisión útil a partir de la información disponible. En conclusión, el uso de métodos de minería de datos siguiendo el modelo CRISP-DM ha brindado información altamente relevante para tomar decisiones relacionadas con el funcionamiento del turbogenerador hidráulico. Por otro lado, se tuvieron como resultados que el modelo de minería de datos empleado demostró ser eficaz en la evaluación del funcionamiento del turbogenerador y en la detección de patrones de comportamiento inusuales. Además, se subrayó la relevancia de poseer un conocimiento previo del proceso analizado para lograr una interpretación adecuada de los hallazgos [8].

La investigación de este autor, se centró en implementar un modelo adaptable de Business Intelligence (BI) para diagnosticar la situación actual de una organización y mejorar su eficacia y beneficios. También, se utilizó un método de investigación que utiliza encuestas para determinar los datos más relevantes. Como resultado, se vio que la inteligencia empresarial (BI) se reconoce como un elemento distintivo entre las organizaciones y una herramienta esencial para agilizar los procesos comerciales y optimizar la gestión de datos en los sistemas de TI. Por otro lado, se tuvo como resultado que después de llevar a cabo la encuesta en las empresas de la región central del Perú, se constató que las organizaciones del ámbito privado

disponen de empleados capacitados en el ámbito de inteligencia empresarial (BI), siendo el 66.67 % del personal capacitado concentrado en el sector privado [9].

En el siguiente artículo, se tuvo como objetivo implementar Business Intelligence en el campo del procesamiento de documentos, con el propósito de evaluar la duración y el número de informes relacionados con la configuración ideal, poniendo un énfasis particular en la mejora constante, el desarrollo y la eficiencia de la empresa en todas las fases del procedimiento legal. Además, se utilizó un nuevo enfoque EVOLUTION, que incluye investigaciones sobre los métodos RALPH KIMBALL, HEFESTO y BILL INMON, que ayudarán a establecer procesos de experiencia de usuario específicos. Como conclusión, se vio que el uso de Business Intelligence ha mejorado el seguimiento y control del tiempo de servicio en el área de procesamiento de documentos. Por otro lado, se tuvieron como resultados que el tiempo requerido para generar información de gestión se redujo en un 99,99%, el costo promedio de reproducción de información disminuyó en un 59%, y el nivel de satisfacción del usuario al utilizar las soluciones de inteligencia empresarial aumentó en un 62,2%. Además, se observaron mejoras notables en los indicadores relacionados con el tiempo promedio para la creación de informes de atención, la cantidad de informes generados y la satisfacción del personal responsable [10].

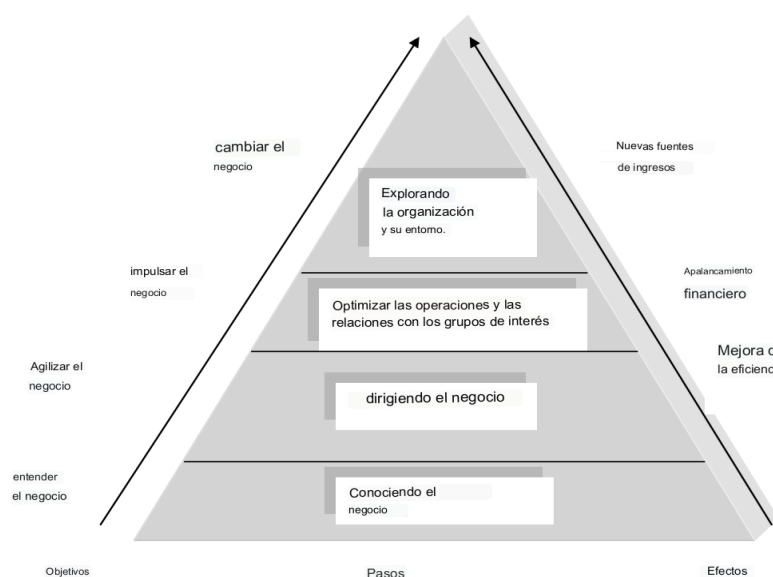
En el estudio, los autores tuvieron como objetivo general analizar el uso de la minería de datos en la gestión de clientes y la toma de decisiones empresariales, donde se consideraron diversos estudios de su aplicación. Concluyen, respaldados por la literatura, que la minería de datos proporciona una visión general, destacando la importancia de estas técnicas para comprender mejor a los clientes, mejorar la eficiencia y la eficacia de CRM y ejecutar el marketing de manera más efectiva. Por otro lado, se tuvieron como resultados que se identificaron conexiones entre las visitas a productos específicos en tiendas minoristas en línea y sus ventas, además, se notó que a medida que los clientes prolongan sus compras, adquieren un mayor número de productos. Además, se introdujeron diversos algoritmos y enfoques para extraer reglas de asociación y crear modelos que describen el comportamiento del consumidor [11].

En esta investigación, tuvieron como objetivo general crear un modelo de predicción que aplique técnicas de minería de datos con el fin de examinar el patrón de

comportamiento de los clientes y categorizar a aquellos que presentan un mayor riesgo de ser rechazados., ayudando así a la empresa a tomar decisiones informadas y crear una estrategia de mantenimiento. Se desarrolla utilizando el método CRISP-DM, que brinda un 83% de precisión en la identificación de clientes de alto riesgo, con un análisis de confianza del 80%. Concluyen que la técnica XG Boost puede ser una herramienta eficaz de detección y clasificación, además se resalta que la minería de datos mediante el reconocimiento de pautas y tendencias en conjuntos extensos de datos puede contribuir a que las empresas tomen decisiones más fundamentadas y exactas. Por otro lado, se tuvieron como resultados la implementación de la inteligencia empresarial y la administración del conocimiento puede ser efectiva en la resolución de desafíos comerciales. También se destacó el empleo de algoritmos de agrupación como K-means y análisis de componentes principales (PCA) para anticipar la demanda futura de productos o servicios [12].

Business Intelligence (BI) se refiere a un conjunto de tecnologías y enfoques utilizados por las empresas para analizar datos y organizar la inteligencia empresarial [13]. Además, pueden mejorar la capacidad de una empresa para administrar y procesar grandes cantidades de datos, al mismo tiempo que facilitan el acceso a la información para todos en la organización [14].

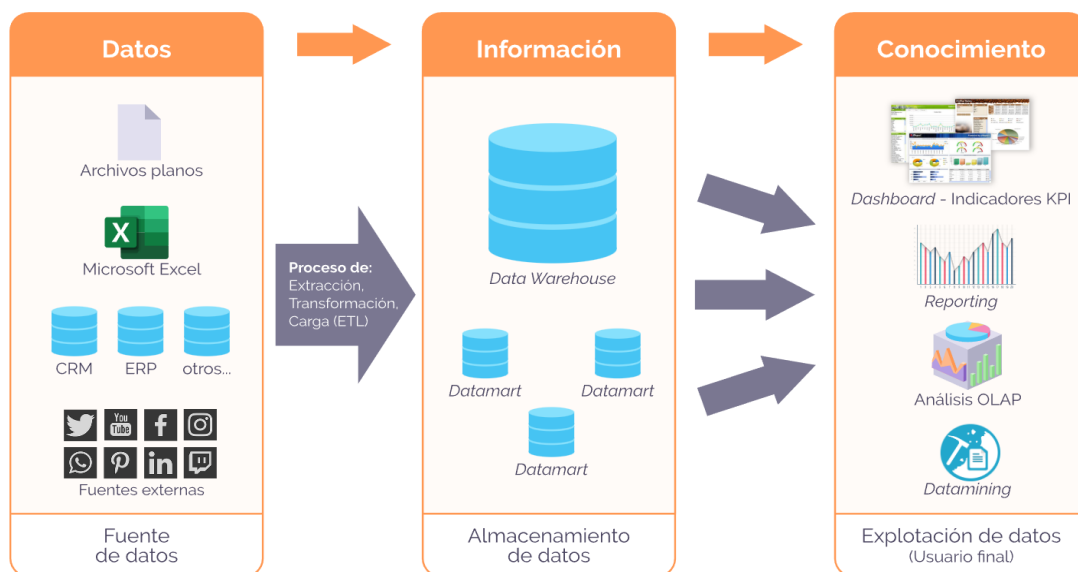
Figura 1: Business Intelligence para reforzar la gestión de la organización



Fuente: Extraído de [15]

Business Intelligence (BI) se define como un proceso que consiste en recopilar información interna y externa utilizando programas o sistemas informáticos que apoyan a las empresas a tomar mejores decisiones estratégicas y lograr una ventaja competitiva [15]. Esto puede conducir a una mejor productividad, menores costos y una asignación más eficiente de los recursos disponibles [16].

Figura 2: Arquitectura de Business Intelligence



Fuente: http://virtual.umng.edu.co/distancia/ecosistema/ovas/ingenieria_informatica/innovacion_tecnologica/unidad_3/medios/graficos/p7.svg

Recogida de datos: Las empresas pueden obtener información de diversas fuentes para su adquisición, que van desde sistemas CRM y ERP, hasta bases de datos, archivos y API. Las organizaciones pueden seleccionar las herramientas y técnicas adecuadas para recopilar datos tanto de procesos internos como de fuentes externas, en función de sus objetivos y recursos [17].

Integración de datos: El siguiente paso es la integración de datos, que incluye la utilización de información recopilada previamente y el proceso de Extracción, Transformación y Carga (ETL). La extracción de datos se realiza desde diversas fuentes externas y luego se transforman para cumplir con los estándares de calidad, lo que asegura que los datos estén limpios y libres de información innecesaria. Posteriormente, los datos se suben en el repertorio correspondiente [17].

Almacén de datos: La consolidación de datos se trata de la unificación de información proveniente de múltiples fuentes en una única base de datos centralizada, como un almacén de datos. Los datos se transforman a un formato estándar para facilitar su análisis y se realizan modificaciones para asegurar su calidad y coherencia [18].

Análisis de datos: Permite a los usuarios trabajar con los datos almacenados en la organización y utilizar técnicas como OLAP y minería de datos con el propósito de adquirir datos significativos que respalden la toma de decisiones. La información generada se pone a disposición de los gestores para su análisis y uso posterior [19].

Distribución de datos: La minería de datos (DM) es una herramienta crucial para tomar mejores estrategias empresariales y se focaliza en identificar patrones y reglas en los recursos de información. La minería de datos permite encontrar relaciones lógicas entre los datos y resumirlos de una manera novedosa, útil y fácilmente comprensible para la organización [20].

Power BI es un tipo de software que se ejecuta en Internet y está diseñado para ayudar a los usuarios a visualizar y analizar datos sin procesar. La visualización de datos tiene como objetivo simplificar el proceso de recuperación de información para los usuarios y acelerar la toma de decisiones [21].

SQL Server es un servidor de bases de datos relacionales que admite el uso del lenguaje de bases de datos estructurado SQL (Structured Query Language). Es ampliamente reconocido por su facilidad de instalación, alto rendimiento, bajo costo de propiedad y características de seguridad superiores [22].

La minería de datos implica la extracción de información valiosa de un conjunto extenso de datos observados, con el propósito de descubrir relaciones y patrones inesperados. Esta tecnología ofrece una herramienta innovadora, fácil de entender y útil para los usuarios, que resulta crucial para las organizaciones, ya que un análisis exhaustivo puede definir el valor de los patrones encontrados, convirtiéndolos en información útil y relevante para el mercado [23] [24].

Existen múltiples técnicas asociadas a la minería de datos que pueden aplicarse para mejorar el rendimiento empresarial en el mercado. Sin embargo, es crucial

comprender cómo utilizar estos grandes conjuntos de datos en beneficio de la organización. En general, estas han sido la principal ventaja competitiva que ha permitido a las empresas mantenerse competitivas [24].

Algunas técnicas relacionadas con la minería de datos: **Técnicas de reglas de asociación (algoritmo Apriori)**: El algoritmo apriori es uno de los algoritmos de análisis de reglas de asociación para determinar con precisión el número de índice apropiado, a menudo se usa para determinar el número de ocurrencias de los elementos, este algoritmo es simple y adecuado para encontrar relaciones entre relaciones y elementos. un determinado conjunto de datos [25].

Técnica de árboles de predicción de clasificación (algoritmo CHAID): Una técnica de análisis estadístico comúnmente utilizada es el algoritmo CHAID, el cual ayuda a identificar patrones entre una variable dependiente categórica y múltiples variables independientes, que también pueden ser categóricas. Los árboles de decisión, que se construyen mediante el uso de CHAID, son una herramienta para modelar eventos y comprender su impacto en los resultados [25].

WEKA: El programa WEKA es un software creado en Java. Este software ofrece diversos algoritmos de aprendizaje automático para su utilización y herramientas para la transformación y preprocesamiento de datos, como técnicas de muestreo y discretización [26].

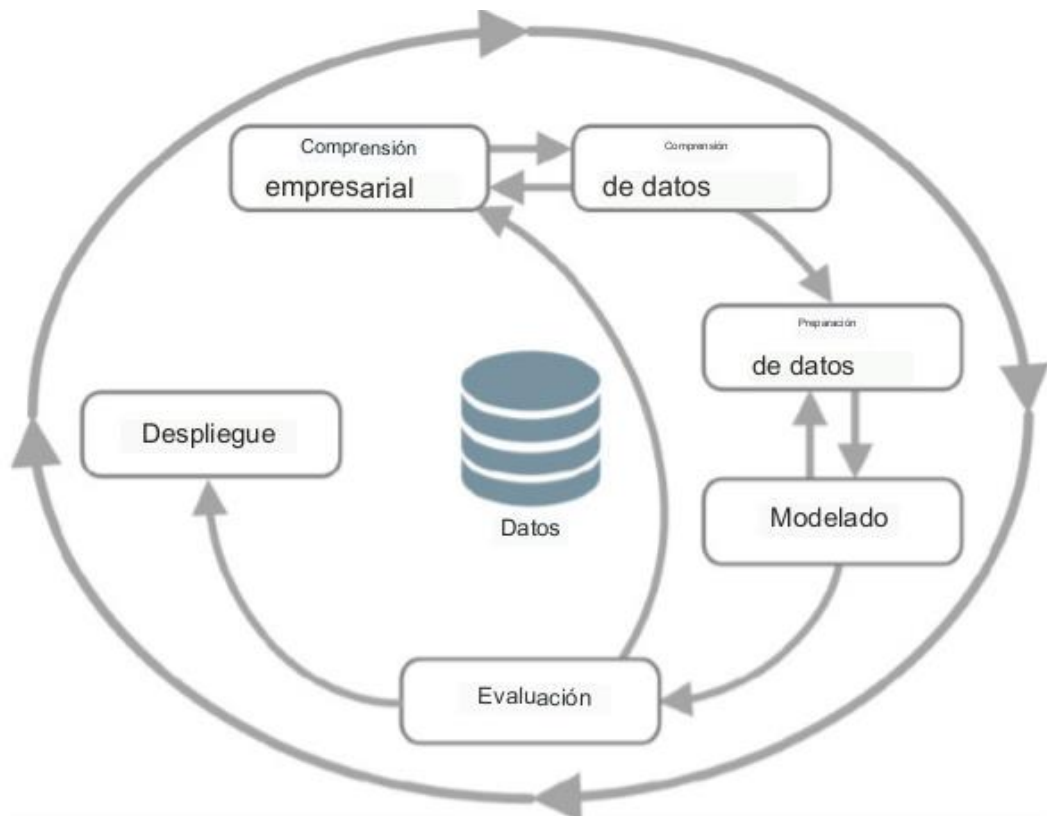
Figura 3: Interfaz Weka



Fuente: Extraído de [26]

La metodología CRISP-DM es un modelo de proceso independiente de la industria que consta de varias fases iterativas para realizar un proyecto orientado a la minería de datos. Normalmente, estas etapas abarcan el entendimiento del negocio, la comprensión de los datos, la estructuración de los datos, la creación del modelo, la evaluación y la implementación [27].

Figura 4: Fases de CRISP-DM



Fuente: Extraído de [24]

Comprensión del negocio: Es esencial hacer un análisis completo de la posición actual de la empresa para obtener una visión integral de los recursos disponibles y aquellos que son necesarios. Uno de los elementos cruciales de esta etapa es la definición del propósito de la minería de datos [27]

Comprensión de los datos: Aquí, concluye con la creación de la tarea y la explicación del procedimiento aproximado a seguir. El usuario y el analista colaboran para intercambiar ideas sobre las tareas y las expectativas. Se revisan y establecen los procedimientos apropiados para llevar a cabo la tarea. Asimismo, en

esta etapa se establecen los estándares que se emplearán para evaluar si se ha logrado el éxito [29].

Preparación de los datos: En esta fase, los datos son preparados para que sean utilizados en el modelo. Si se utilizan diversas fuentes de datos, se deben vincular durante la preparación de los mismos. También se abordan las cuestiones de calidad de los datos y se diseña una estrategia para tratar cualquier problema relacionado. Es crucial reservar datos de validación y prueba durante esta fase para poder evaluar el modelo en fases posteriores [30].

Modelado: En esta fase, se efectúa la selección de las técnicas de modelado más apropiadas para crear y evaluar el modelo. Las elecciones de las técnicas utilizadas se basan en las particularidades de los datos [31].

Evaluación: En la etapa de evaluación del modelo, se busca comprender la calidad del modelo creado y los resultados que produce, asegurando que aborda adecuadamente el problema de negocio. Esta fase es esencial como paso intermedio entre la creación del modelo y su implementación [32].

Despliegue: La fase de despliegue se encarga de implementar los resultados obtenidos del proyecto de minería de datos para asegurar su disponibilidad y cumplir con las necesidades finales de los clientes en la empresa [33].

III. METODOLOGÍA

3.1. Tipo y diseño de investigación

Para este proyecto se eligió la investigación aplicada, porque se buscará proporcionar soluciones prácticas a los problemas identificados, generando nuevas estrategias para el negocio. El objetivo de este tipo de investigación es plantear problemas o interrogantes específicos, enfocándose en el estudio y la consolidación del conocimiento para su aplicación en situaciones reales [34].

En el presente proyecto se emplea el diseño pre-experimental. Es un diseño sin grupo de control en el cual se realiza una medición inicial seguida de una medición posterior. Este diseño es de un solo grupo experimental, caracterizado por un estricto control en los niveles de intervención y variables [35].

Figura 5: Diseño de Investigación

$$Ge = O1 \quad X \quad O2$$

Fuente: extraído de [35]

Dónde:

Ge: grupo experimental

O1: Pre-Test: Antes de aplicar la implementación del business intelligence utilizando minería de datos.

X: La aplicación del business intelligence utilizando minería de datos.

O2: Post-Test: Después de aplicar la implementación del business intelligence utilizando minería de datos.

3.2. Variables y operacionalización

Variable Independiente:

Business intelligence

- Definición Conceptual: La inteligencia empresarial desempeña un papel fundamental como elemento estratégico en empresas y organizaciones al ofrecer una ventaja competitiva al proporcionar conocimientos exclusivos para abordar desafíos empresariales [36].

Variable Dependiente: ventas

- Definición Conceptual: El término "ventas" se utiliza para describir la transacción de bienes y servicios a cambio de una compensación monetaria. Estas transacciones se llevan a cabo debido a necesidades específicas y subjetivas de los clientes [37].
- Definición Operacional: Las buenas estrategias realizadas en las ventas mejoran el estado financiero del negocio, donde se verá reflejado en el crecimiento de ventas, así como también de los clientes.
- Dimensión:
 - Perspectiva Financiera [38].
 - Perspectiva Cliente [38].
 - Eficiencia [39]
 - Rotación de inventario [39]
- Indicadores:
 - Crecimiento de ventas
 - Fórmula: crecimiento en porcentaje
 - $CV = (venta\ total\ actual - venta\ total\ pasada / venta\ total\ pasada) \times 100$
 - Crecimiento de clientes
 - Fórmula: crecimiento en porcentaje
 - $CC = (total\ cliente\ actual - total\ cliente\ pasado / total\ cliente\ pasado) \times 100$
 - Eficiencia de Pedidos

Fórmula: porcentaje

$$\left(\frac{\text{Pedidos entregados conformes}}{\text{total de pedidos realizados}} \right) * 100$$

Índice de rotación de inventario

Fórmula:

$$\frac{\text{Costo de venta}}{\text{inventario promedio}}$$

- Escala de medición

Razón

3.3. Población, muestra y muestreo

3.3.1. Población

Es fundamental determinar el tamaño de la población para obtener resultados precisos, especialmente dado que la investigación evaluará las ventas. Se utilizará un diseño pre-experimental, comparando los datos periódicos de la empresa, antes y después de la implementación. La población hace referencia a todo el elemento que son objeto de estudio, especificados por el investigador de acuerdo con la definición incorporada en el estudio [40].

Por lo tanto, la población consta de reportes diarios, los cuales incluyen las ventas del día, pero solo de acuerdo a 20 productos de construcción, esto debido al tiempo de realización del proyecto, por ello los resultados obtenidos serán de acuerdo a esos datos específicos.

También, los productos seleccionados se basan en aquellos que tienen mayor demanda dentro de la empresa.

3.3.2. Muestra

La muestra será igual que la población, debido a que es pequeña. Por ello, una muestra es una porción más reducida de una población o conjunto que resulta de interés, del cual se obtendrán datos relevantes y es importante que refleje de manera adecuada a toda la población en cuestión [41].

3.3.3. Muestreo

En este proyecto no se aplicará el muestreo. Este método se emplea cuando la cantidad de elementos en la población es considerable, mientras que en casos de poblaciones pequeñas no se necesita utilizar la técnica de muestreo [40].

3.3.4. Unidad de análisis

La unidad de análisis del presente estudio, son las ventas generadas en los días mencionados en la población.

3.4. Técnicas e instrumentos de recolección de datos

Esta investigación optó por utilizar la técnica del fichaje, ya que se obtendrá información de los registros de ventas. El fichaje es una técnica utilizada en la investigación, la cual sirve para recolectar datos ya existentes [42]. Por ello, como instrumento se usará la ficha de registro, que es la compilación de datos e información de las diversas fuentes que están siendo consultadas, las fichas son creadas y diseñadas considerando la información deseada para el estudio, lo que significa que no hay un formato estándar predefinido [43].

Validez: La validez del instrumento a utilizar en el siguiente proyecto se apoya en la información de libros, lo cual proporciona una base sólida y confiable para la investigación. Los libros son fuentes de conocimiento que a menudo han sido revisadas por expertos en el campo y respaldadas por investigaciones previas. Esto asegura que el instrumento esté respaldado por una sólida fundamentación teórica y conceptual.

Confiabilidad

Este proyecto de investigación debe tener confiabilidad ya que, implica la aplicación de una escala o instrumento de medición a un grupo de individuos en dos ocasiones distintas. Esta aplicación puede llevarse a cabo de manera inmediata o con un lapso de tiempo entre la primera y la segunda vez [44].

Para evaluar la confiabilidad nos basaremos en la confiabilidad por test-retest donde habrá dos tomas o dos momentos (test y retest) donde la segunda toma será pasado un cierto tiempo definido respecto de la primera toma, se espera que los datos sean iguales o cercanos a uno, en las dos tomas. Para el desarrollo de la confiabilidad se usará el programa SPSS.

Además, se establece la técnica a utilizar, la cual servirá para obtener la escala, esto consistirá en el uso del coeficiente de correlación de Pearson. Este coeficiente representa una medida estadística inferencial que indica la fuerza de la relación lineal entre dos variables cuantitativas [43].

Tabla 1: Nivel de confiabilidad Kappa

VALORES	INTERPRETACIÓN
<0,01	No acuerdo
0,01 – 0,20	Ninguna a escaso
0,21 – 0,40	Regular o razonable
0,41 – 0,60	Moderado
0,61 – 0,80	Substancial
0,81 – 1,00	Casi perfecto

Fuente: Extraído de [39]

De acuerdo a las medidas realizadas en el programa SPSS para cada indicador se obtuvieron los siguientes resultados del test retest:

Tabla 2: Confiabilidad indicador crecimiento de ventas

Correlaciones			
		TEST	RETEST
TEST	Pearson	1	0,775
	Sig. (bilateral)		0,001
	N	15	15
RETEST	Pearson	0,775	1

	Sig. (bilateral)	0,001	
	N	15	15

Fuente: Elaboración propia

El indicador de crecimiento de ventas evaluado a través del coeficiente de Pearson, alcanza un valor de 0.775. Este resultado indica que el instrumento muestra una confiabilidad substancial, lo que implica que las mediciones obtenidas son confiables. Por lo tanto, se puede decir que el instrumento es fiable.

Tabla 3: Confiabilidad indicador crecimiento de clientes

Correlaciones			
		TEST	RETEST
TEST	Pearson	1	0,686
	Sig. (bilateral)		0,005
	N	15	15
RETEST	Pearson	0,686	1
	Sig. (bilateral)	0,005	
	N	15	15

Fuente: Elaboración propia

Este indicador evaluado a través del coeficiente de Pearson, alcanza un valor de 0.686. Este resultado indica que el instrumento muestra una confiabilidad substancial, lo que implica que las mediciones obtenidas son confiables. Por lo tanto, se puede decir que el instrumento es fiable.

Tabla 4: Confiabilidad indicador índice de rotación de inventario

Correlaciones			
		TEST	RETEST
TEST	Pearson	1	0,657
	Sig. (bilateral)		0,002
	N	20	20
RETEST	Pearson	0,657	1
	Sig. (bilateral)	0,002	
	N	20	20

Fuente: Elaboración propia

El indicador índice de rotación de inventario evaluado a través del coeficiente de Pearson, alcanza un valor de 0.657. Este resultado indica que el instrumento muestra una confiabilidad substancial, lo que implica que las mediciones obtenidas son confiables. Por lo tanto, se puede decir que el instrumento es fiable.

Tabla 5: Confiabilidad indicador eficiencia de pedidos

		TEST	RETEST
TEST	Pearson	1	0,61
	Sig. (bilateral)		0,016
	N	15	15
RETEST	Pearson	0,61	1
	Sig. (bilateral)	0,016	
	N	15	15

Fuente: Elaboración propia

El indicador eficiencia de pedidos evaluado a través del coeficiente de Pearson, alcanza un valor de 0.61. Este resultado indica que el instrumento muestra una confiabilidad substancial, lo que implica que las mediciones obtenidas son confiables. Por lo tanto, se puede decir que el instrumento es fiable.

3.5. Procedimientos

- Se solicita el permiso para recolección de los registros de ventas en la Distribuidora San Cristóbal, la cual será de apoyo para esta investigación.
- Se acceden a los documentos que contienen los datos a medir, ellos se registraron en el instrumento (anexos)
- Luego, para los procesos de pre-test se tendrá un registro de acuerdo a las estadísticas del negocio antes de realizar la implementación de acuerdo a sus ventas.

- Se desarrollará la Metodología CRISP-DM referente a la minería de datos, elegida en el proyecto.
- Luego, los datos recopilados se almacenarán en Sql Server y luego se transferirán a POWER BI para tener un vista general e intuitiva de los datos
- Se realizará un objetivo visual para que luego sea exportado a excel, seguidamente se llevará a WEKA para realizar el análisis de patrones y predicciones del registro de ventas.
- Posteriormente, se utilizará Power BI, lo que permitirá comprender mejor la información almacenada y tomar decisiones estratégicas más efectivas para el negocio.
- Seguidamente para el proceso de post-test, después del tiempo estipulado para su evaluación y ya una vez implementada la solución propuesta, se tendrá los resultados de acuerdo a los indicadores mencionados en el proyecto.

3.6. Método de análisis de datos

Este presente proyecto tendrá datos estadísticos que comprueben si la hipótesis es correcta o no. El análisis descriptivo, se da usando el SPSS, con su respectiva interpretación de acuerdo a los datos estadísticos.

Por otra parte, el análisis inferencial será la clasificación de las pruebas estadísticas, mediante la prueba de normalidad:

- si $p < 0.05$ se utilizará el método de Shapiro-Wilk
- si $p \geq 0.05$ se usará el método de Kolmogorof Smimov.

Por último, se va a desarrollar la prueba de las hipótesis, tanto la hipótesis nula (H_0) como la hipótesis alternativa (H_a):

- si $p < 0.05$ posee una distribución no normal (prueba no paramétrica) y se usará la prueba Wilcoxon
- si $p \geq 0.05$ posee una distribución normal (prueba paramétrica) y se usará la prueba T Student

Hipótesis General:

El Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas

Indicadores:

CV: Crecimiento de ventas

IRI: Índice de rotación de inventario

CC: Crecimiento de clientes

EP: Eficiencia de Pedidos

HE: Hipótesis Específico

a: antes

d: después

Indicador 1:

Hipótesis Específico (HE1): El Business Intelligence utilizando minería de datos mejora el crecimiento de ventas.

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora el crecimiento de ventas.

$$H0 = CVa \geq CVd$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos mejora el crecimiento de ventas.

$$Ha = CVa < CVd$$

Indicador 2:

Hipótesis Específico (HE2): El Business Intelligence utilizando minería de datos mejora la eficiencia de pedidos en las ventas

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora la eficiencia de pedidos en las ventas

$$H0 = EPa \geq EPd$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos mejora la eficiencia de pedidos en las ventas

$$Ha = EPa < EPd$$

Indicador 3:

Hipótesis Específico (HE3): El Business Intelligence utilizando minería de datos mejora el crecimiento de clientes.

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora el crecimiento de clientes.

$$H0 = CCa \geq CCd$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos aumenta los clientes.

$$Ha = CCa < CCd$$

Indicador 4:

Hipótesis Específico (HE4): El Business Intelligence utilizando minería de datos mejora el índice de rotación de inventario para las ventas.

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora la eficiencia de pedidos en las ventas

$$H0 = EPa \geq EPd$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos mejora la eficiencia de pedidos en las ventas

$$Ha = EPa < EPd$$

3.7. Aspectos éticos

Esta presente investigación se consolida a los elementos éticos establecidos por la Universidad César Vallejo, en línea con la Resolución de Consejo Universitario RCUN°0340-2021. Se respeta la autoría original de todas las fuentes de investigación, garantizando un compromiso firme contra el plagio, con el objetivo de asegurar que los autores reciban el reconocimiento adecuado sin ser desacreditados. Este aspecto puede ser verificado mediante el análisis del software TURNITIN. Además, se cumplen las políticas y principios éticos que rigen la recopilación de los datos de la empresa bajo investigación. Se cuenta con el permiso del responsable de la empresa para acceder a la información necesaria, siguiendo todos los requisitos y procedimientos establecidos. Asimismo, se ha seguido la norma ISO 690, que dicta las pautas para la citación y referenciación en tesis en el ámbito de la ingeniería, de acuerdo con las regulaciones de la universidad.

IV. RESULTADOS

4.1. Análisis descriptivo

Indicador 1: Crecimiento de ventas (CV)

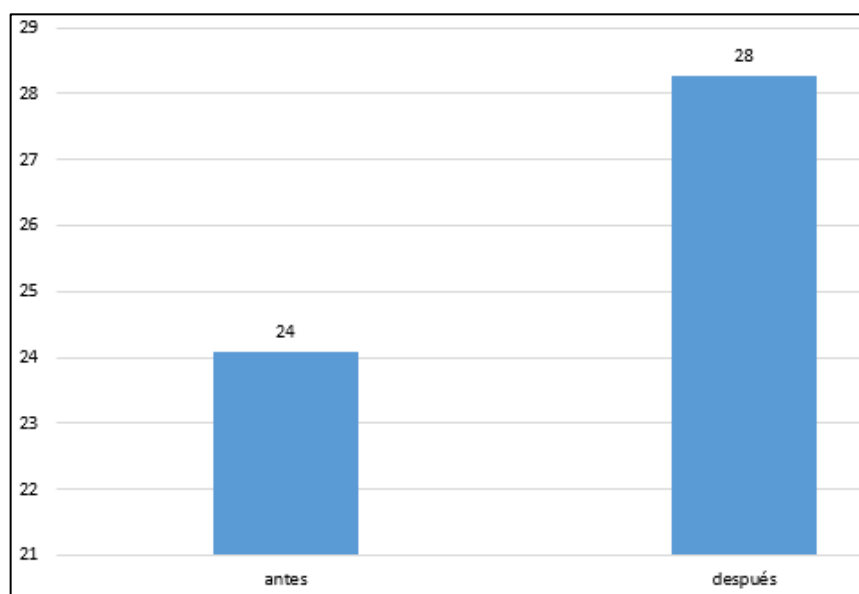
Tabla 6: Indicador CV en Pre-Test y Post-Test

Pruebas descriptivas					
	N	Mínimo	Máximo	Media	Desv. Desviación
CC_Pre-Test	15	13,83	34,83	24,0833	6,19077
CC_Post-Test	15	17,46	44,71	28,268	7,76653
N válido (por lista)	15				

Fuente: Elaboración propia

En este análisis estadístico descriptivo, se observó que el crecimiento de ventas antes de la implementación tenía una media de 24,08, mientras que después de la implementación, la media fue de 28,268. También se registró una desviación de 6.19077 para el pre-test y 7,7665 para el post-test. Seguidamente, para los valores mínimos y máximos en el PreTest fueron 13,83 y 34,83, respectivamente, y en el PostTest, estos valores fueron 17,46 y 44,71, destacando el cambio de mejora en el indicador. Este resultado se mostrará a continuación.

Figura 6: Crecimiento de ventas antes y después de la implementación



Fuente: Elaboración propia

Esto indica que la eficiencia de pedidos ha experimentado un incremento del 24% al 28% lo que refleja una diferencia de 4% que se alcanzó mediante esta implementación.

Indicador 2: Eficiencia de Pedidos (EP)

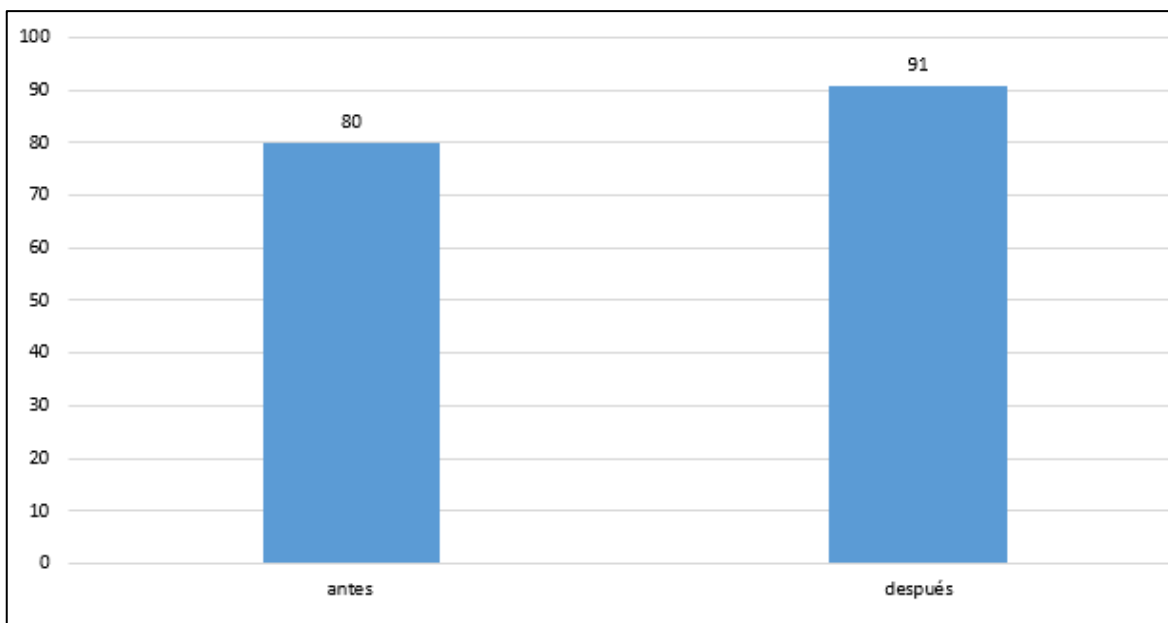
Tabla 7: Indicador EP en Pre y Post-Test

Pruebas descriptivas					
	N	Mínimo	Máximo	Media	Desv. Desviación
EP_Pre-Test	15	71,43	87,5	79,8067	6,12185
EP_Post-Test	15	81,82	100	90,7833	4,71013
N válido (por lista)	15				

Fuente: Elaboración propia

En este análisis estadístico descriptivo, se observó que la eficiencia de pedidos antes de la implementación tenía una media de 79.80, mientras que después de la implementación, la media fue de 90.78. También se registró una desviación de 6.1218 para el pre-test y 4.7101 para el post-test. Seguidamente, para los valores mínimos y máximos en el PreTest fueron 71.43 y 87.5, respectivamente, y en el PostTest, estos valores fueron 81.82 y 100, destacando el cambio de mejora en el indicador. Este resultado se mostrará a continuación.

Figura 7: Eficiencia de Pedidos antes y después de la implementación



Fuente: Elaboración propia

Esto indica que la eficiencia de pedidos ha experimentado un incremento del 80% al 91%, lo que refleja una diferencia de 11% que se alcanzó mediante esta implementación.

Indicador 3: Crecimiento de clientes (CC)

Tabla 8: Indicador CC en Pre y Post-Test

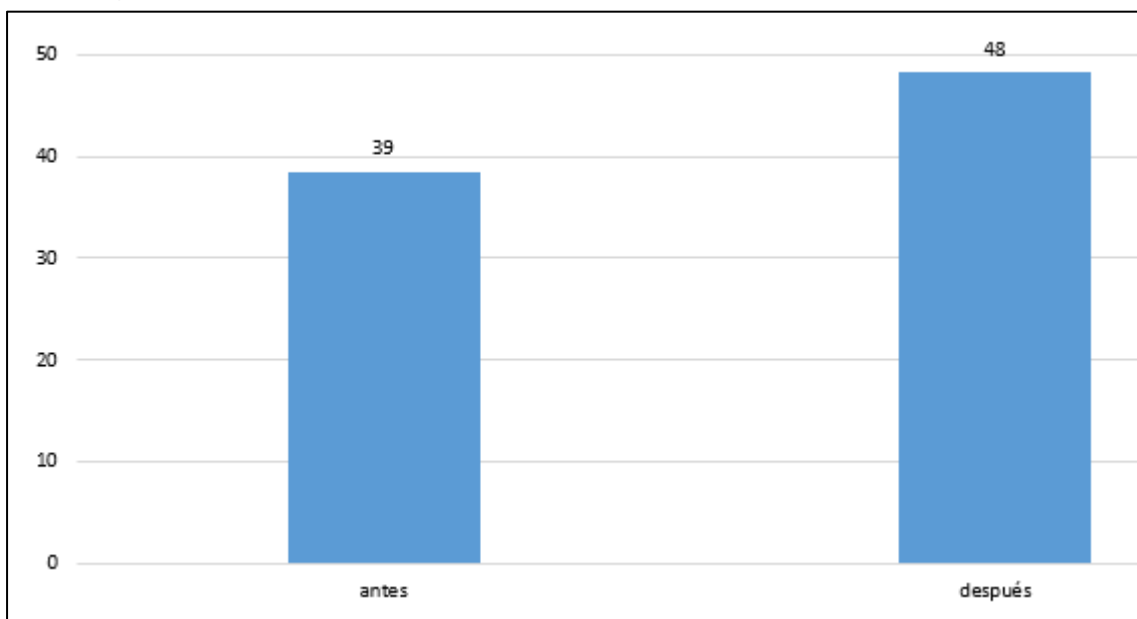
Pruebas descriptivas					
	N	Mínimo	Máximo	Media	Desv. Desviación
CC_Pre-Test	15	15,0	50,0	38,5387	11,66092
CC_Post-Test	15	21,05	66,67	48,282	13,97167
N válido (por lista)	15				

Fuente: Elaboración propia

En este análisis estadístico descriptivo, se observó que el indicador antes de la implementación tenía una media de 38,538, mientras que después de la implementación, la media fue de 48,282. También se registró una desviación de 11,6609 para el pre-test y 13,97167 para el post-test. Seguidamente, para los valores mínimos y máximos en el PreTest fueron 15 y 50, respectivamente, y en el

PostTest, estos valores fueron 21,05 y 66,67, destacando el cambio de mejora en el indicador. Este resultado se mostrará a continuación.

Figura 8: Crecimiento de clientes antes y después de la implementación



Fuente: Elaboración propia

Esto implica que el indicador ha experimentado un aumento de 39% a 48%, lo que representa una diferencia de 9% que se logró gracias a la implementación.

Indicador 4: Índice de rotación de inventario (IRI)

Tabla 9: Indicador IRI en Pre y Post-Test

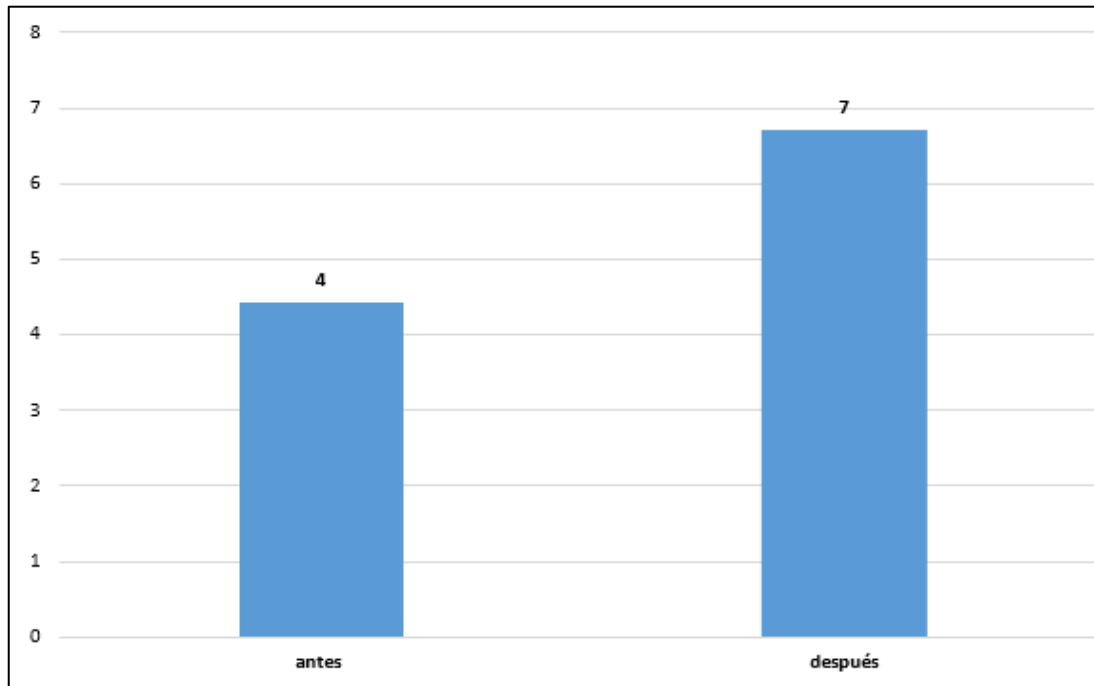
Pruebas descriptivas					
	N	Mínimo	Máximo	Media	Desv. Desviación
IRI_Pre-Test	20	2,67	7,71	4,438	1,29037
IRI_Post-Test	20	4	9,3	6,7215	1,47261
N válido (por lista)	20				

Fuente: Elaboración propia

En este análisis estadístico descriptivo, se encontró que el indicador antes de la implementación tenía una media de 4.43, mientras que después, la media fue de 6.72. También se registró una desviación de 1.2903 para el pre-test y 1.4726 para

el post-test. Seguidamente, para los valores mínimos y máximos en el PreTest fueron 2.67 y 7.71, respectivamente, y en el PostTest, estos valores fueron 4 y 9.3, destacando el cambio de mejora. Este resultado se mostrará a continuación.

Figura 9: Índice de rotación de inventario antes y después de la implementación



Fuente: Elaboración propia

Esto implica que el índice de rotación de inventario ha experimentado un aumento de 4 a 7, lo que representa una diferencia de 3 que se logró gracias a la implementación.

4.2. Análisis Inferencial

Indicador 1: Crecimiento de ventas (CV)

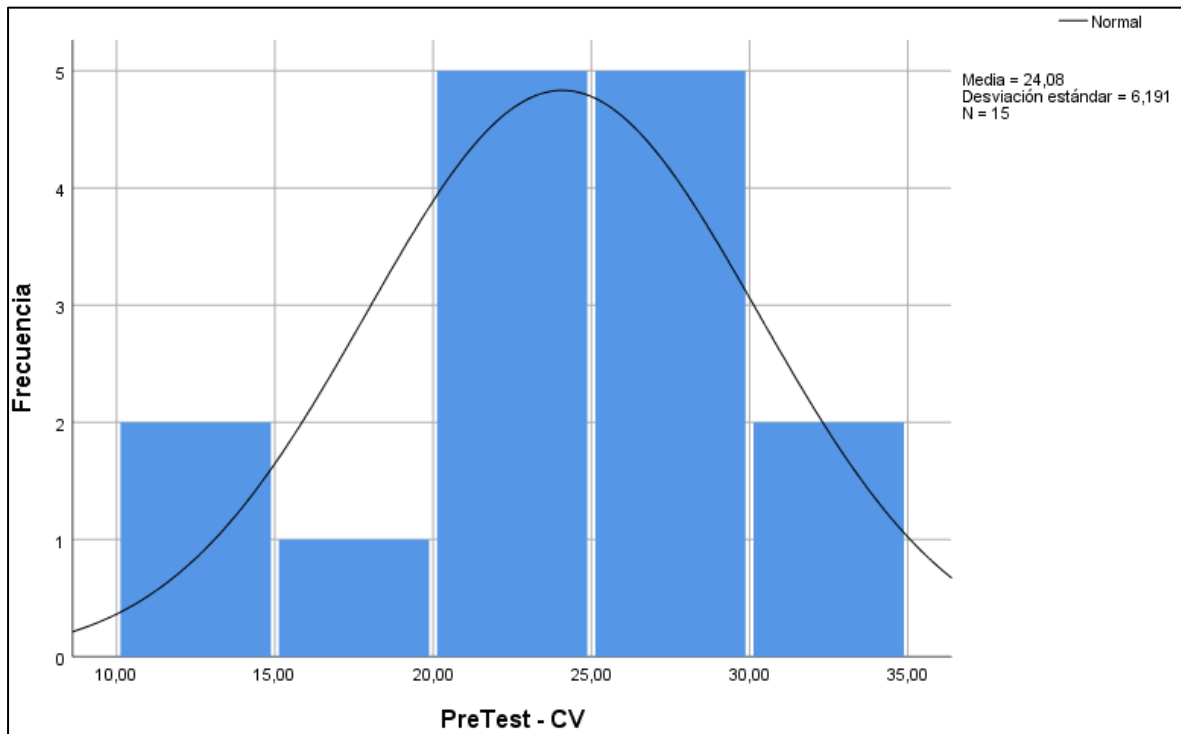
Tabla 10: Prueba de Normalidad de CV en pre y post-test

Prueba CV			
	Shapiro-Wilk		
	Estadístico	gl	Sig.
CV_PreTest	0,969	15	0,851
CV_PostTest	0,954	15	0,589

Fuente: Elaboración propia

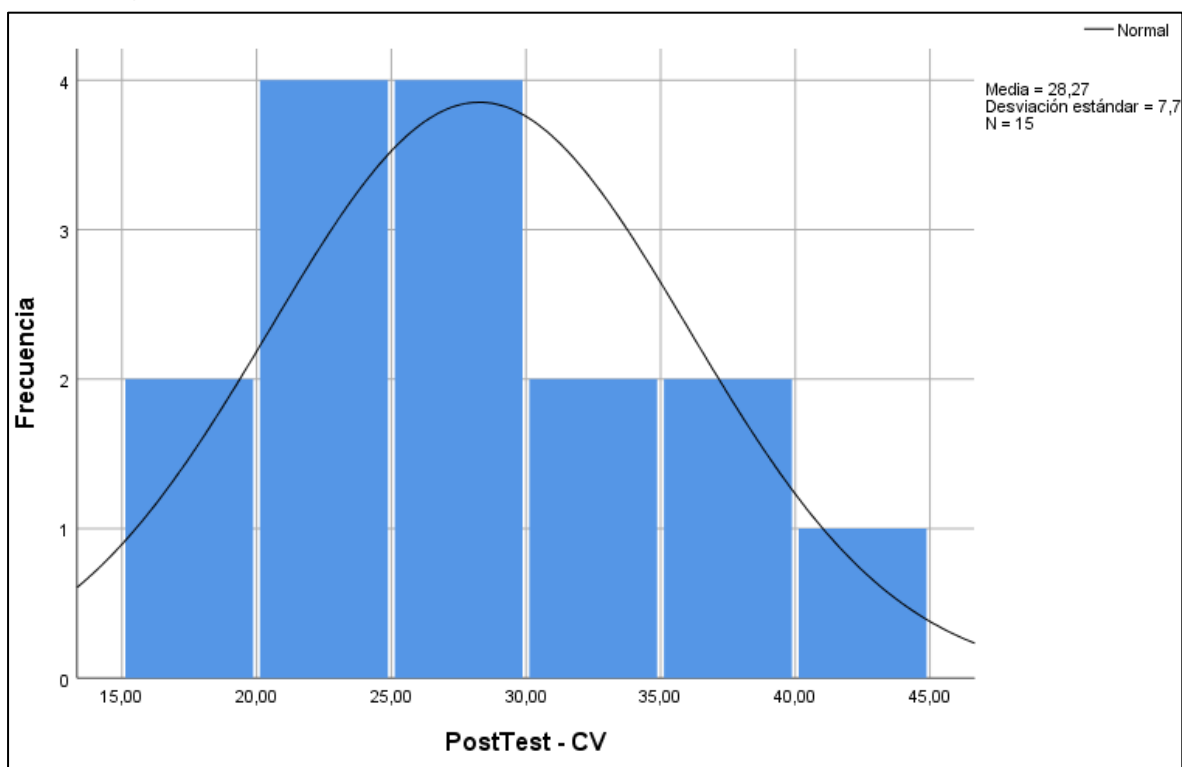
Basándonos en la información de la tabla generada de acuerdo al indicador, se puede inferir que los datos del Sig, tanto para el pre-test con un valor de 0,074 y el post-test con 0,184, muestran que son datos superiores al umbral de 0.05, lo que significa que los datos alcanzan una normalidad adecuada. Estos resultados se enseñan en las imágenes 10 y 11.

Figura 10: Prueba de normalidad de Crecimiento de ventas Pre-Test



Fuente: Elaboración propia

Figura 11: Prueba de normalidad de Crecimiento de ventas Post-Test



Fuente: Elaboración propia

Indicador 2: Eficiencia de Pedidos (EP)

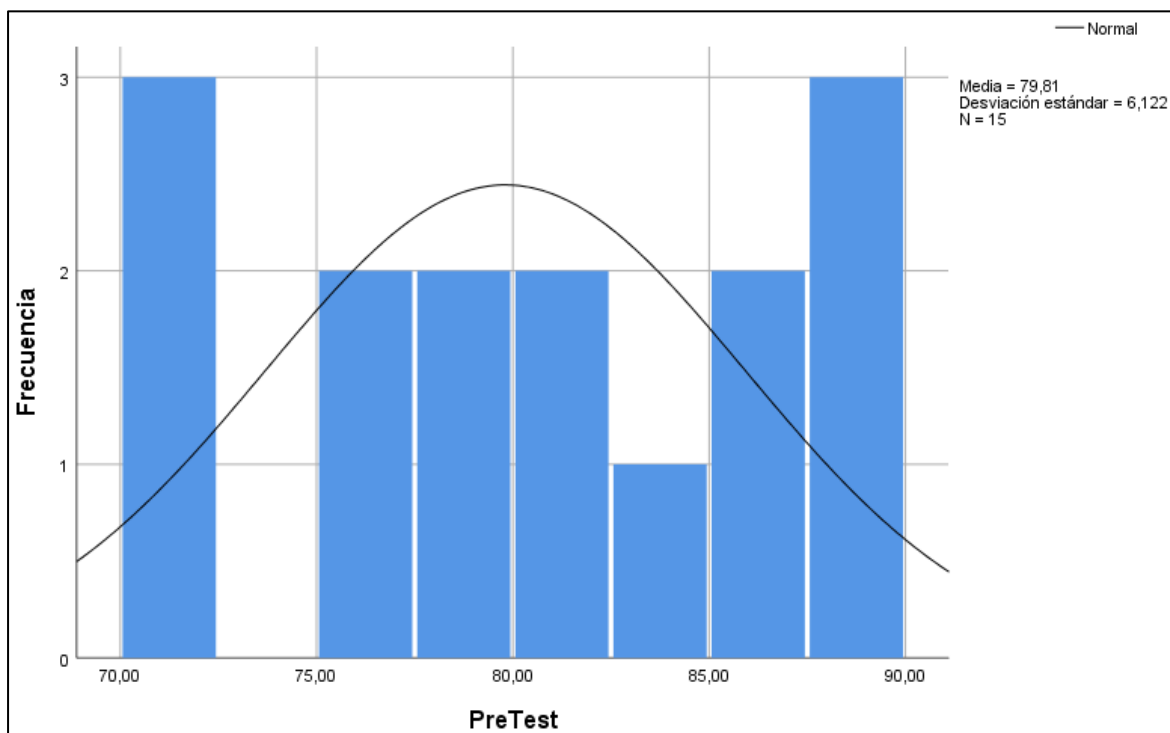
Tabla 11: Prueba de Normalidad de EP en pre y post-test

Prueba EP			
	Shapiro-Wilk		
	Estadístico	gl	Sig.
EP_PreTest	0,893	15	0,074
EP_PostTest	0,919	15	0,184

Fuente: Elaboración propia

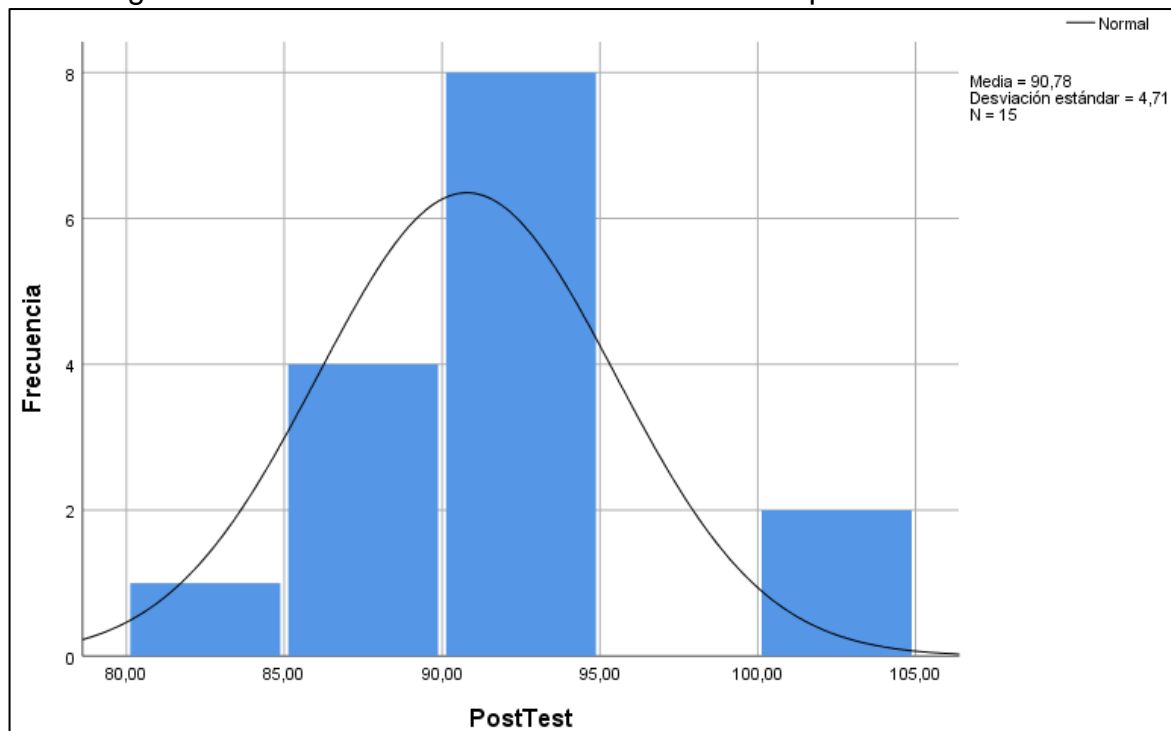
Basándonos en la información de la tabla generada de acuerdo al indicador, se puede inferir que los datos del Sig, tanto para el pre-test con un valor de 0,074 y el post-test con 0,184, muestran que son datos superiores al umbral de 0.05, lo que significa que los datos alcanzan una normalidad adecuada. Estos resultados se enseñan en las imágenes 12 y 13.

Figura 12: Prueba de normalidad de eficiencia de pedidos Pre-Test



Fuente: Elaboración propia

Figura 13: Prueba de normalidad de eficiencia de pedidos Post-Test



Fuente: Elaboración propia

Indicador 3: Crecimiento de clientes (CC)

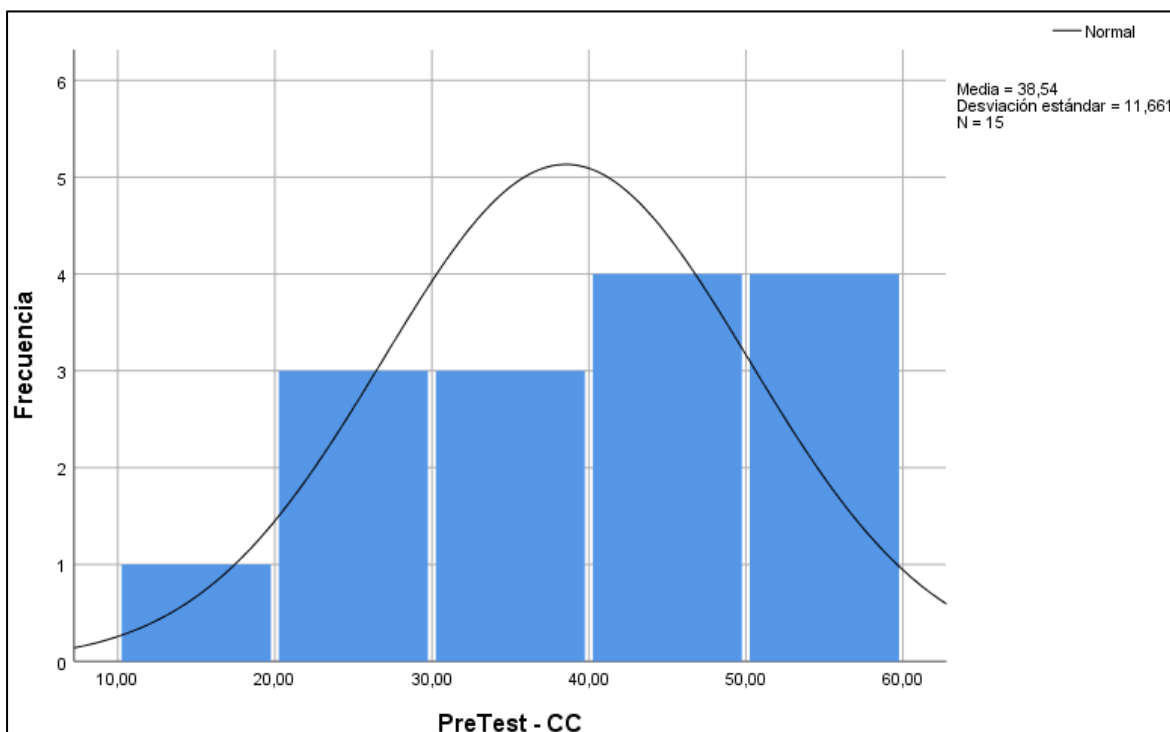
Tabla 12: Prueba de Normalidad de CC en pre y post-test

Prueba CC			
	Shapiro-Wilk		
	Estadístico	gl	Sig.
CC_PreTest	0,882	15	0,051
CC_PostTest	0,905	15	0,114

Fuente: Elaboración propia

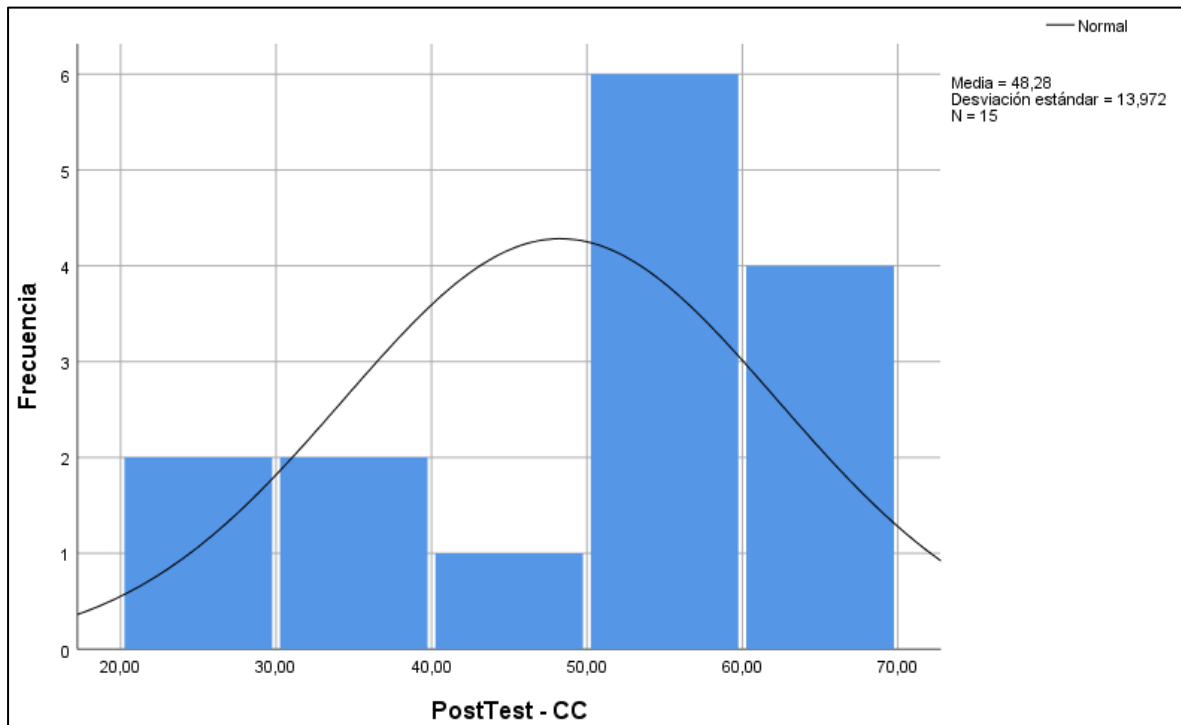
Basándonos en la información de la tabla generada de acuerdo al indicador, se puede inferir que los datos del Sig, tanto para el pre-test con un valor de 0,51 y el post-test con 0,114, muestran que son datos superiores al umbral de 0.05, lo que significa que los datos alcanzan una normalidad adecuada. Estos resultados se enseñan en las imágenes 14 y 15.

Figura 14: Prueba de normalidad de crecimiento de clientes PreTest



Fuente: Elaboración propia

Figura 15: Prueba de normalidad de crecimiento de clientes PostTest



Fuente: Elaboración propia

Indicador 4: Índice de rotación de inventario (IRI)

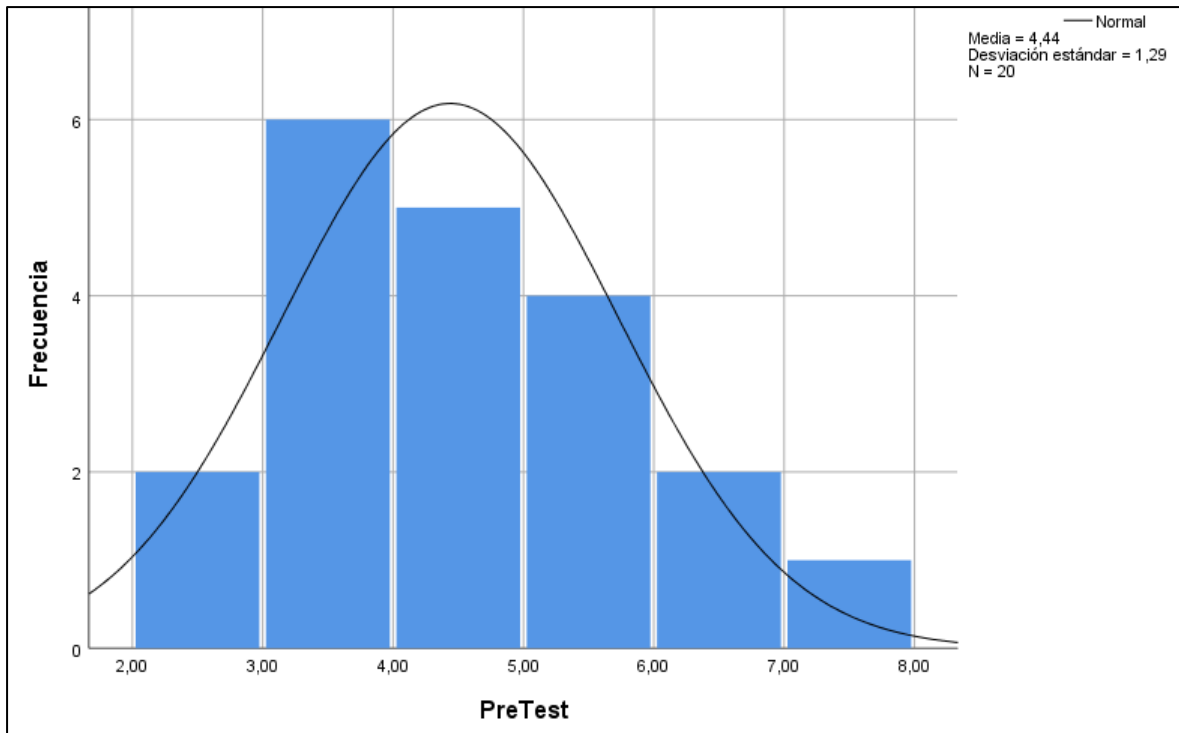
Tabla 13: Prueba de Normalidad de IRI en pre y post-test

Prueba IRI				
	Shapiro-Wilk			
	Estadístico		gl	Sig.
IRI_PreTest	0,931		20	0,161
IRI_PostTest	0,975		20	0,856

Fuente: Elaboración propia

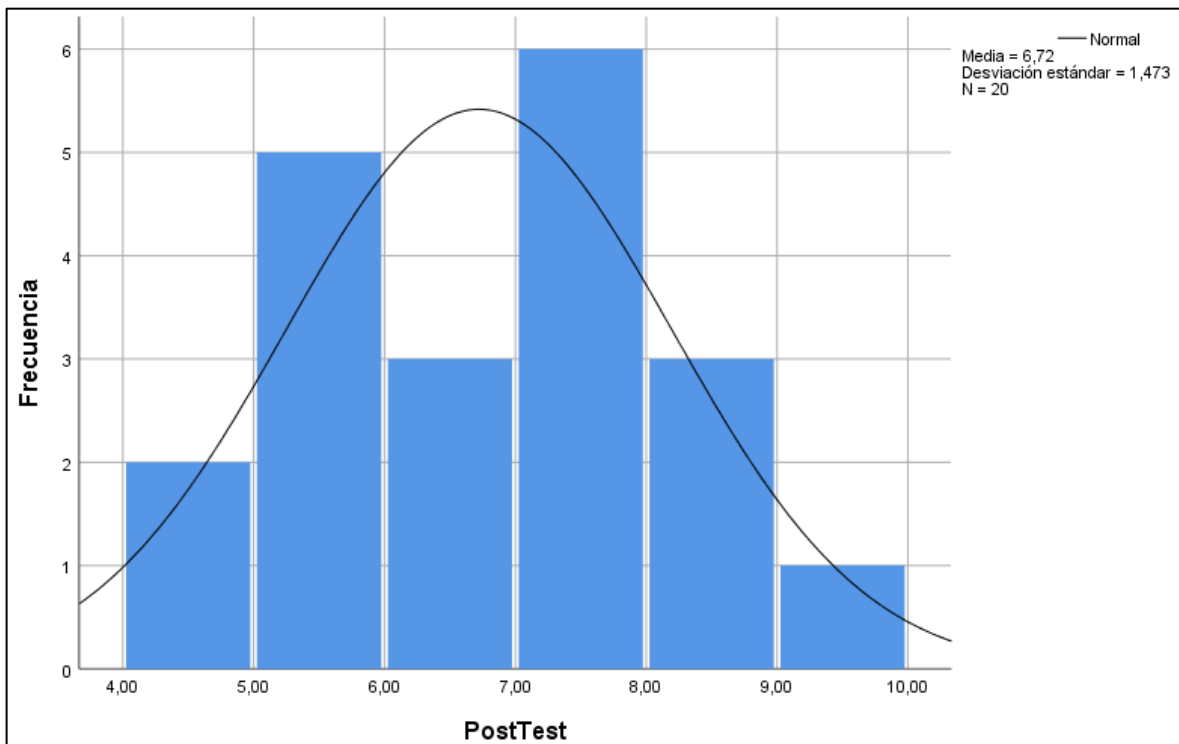
Basándonos en la información de la tabla generada de acuerdo al indicador, se puede inferir que los datos del Sig, tanto para el pre-test con un valor de 0,161 y el post-test con 0,856, muestran que son datos superiores al umbral de 0.05, lo que significa que los datos alcanzan una normalidad adecuada. Estos resultados se enseñan en las imágenes 16 y 17.

Figura 16: Prueba de normalidad de Índice de rotación de inventario PreTest



Fuente: Elaboración propia

Figura 17: Prueba de normalidad de Índice de rotación de inventario PostTest



Fuente: Elaboración propia

4.3 Prueba de Hipótesis

Indicador 1:

HE2: El Business Intelligence utilizando minería de datos mejora el crecimiento de ventas.

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora el crecimiento de ventas.

$$H_0 = CVa \geq CVd$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos mejora el crecimiento de ventas.

$$H_a = CVa < CVd$$

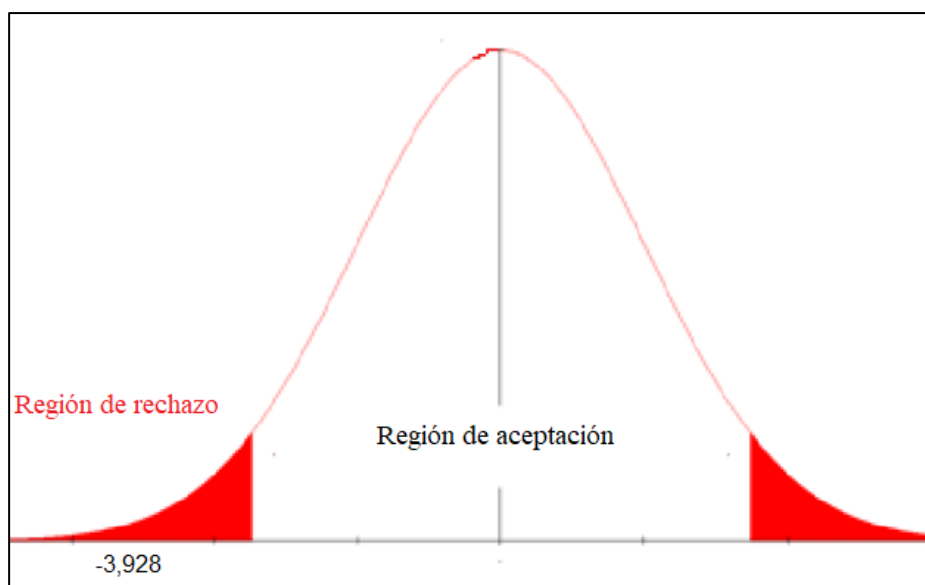
Tabla 14: Prueba T-Student Crecimiento de ventas

Prueba de T						
	Media	N	Desviación típ.	t	gl	Sig. (bilateral)
CC_Pre-Test	24,08	15	6,191	-3,928	14	,002
CC_Post-Test	28,27	15	7,767			

Fuente: Elaboración propia

Para la evaluación de la hipótesis de este indicador, se empleó T-Student, lo cual los datos utilizados son del pre-test y el post-test. El valor calculado para T resultó ser -3,928. Dado que el valor de Sig. es menor que 0.05, se opta por aceptar la hipótesis alternativa y descartar la nula. Por lo tanto, esto significa que la implementación del proyecto efectivamente mejora el crecimiento de ventas. Además, cabe recalcar que el valor de T se halla en la zona de rechazo, como se puede apreciar en la figura siguiente.

Figura 18: Región de la Prueba T-Student de CV



Fuente: Elaboración propia

Indicador 2:

HE2: El Business Intelligence utilizando minería de datos mejora la eficiencia de pedidos en las ventas

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora la eficiencia de pedidos en las ventas

$$H_0 = EP_a \geq EP_d$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos mejora la eficiencia de pedidos en las ventas

$$H_a = EP_a < EP_d$$

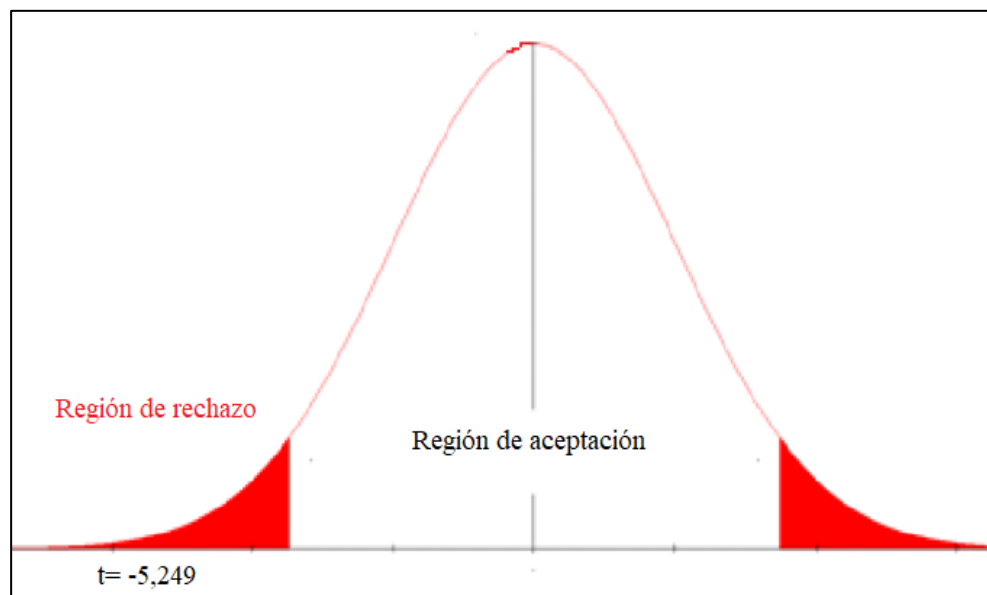
Tabla 15: Prueba T-Student Eficiencia de Pedidos

Prueba de T						
	Media	N	Desviación típ.	t	gl	Sig. (bilateral)
EP_Pre-Test	79,81	15	6,122	-5,249	14	0,000
EP_Post-Test	90,78	15	4,710			

Fuente: Elaboración propia

Para la evaluación de la hipótesis de este indicador, se empleó T-Student, lo cual los datos utilizados son del pre-test y el post-test. El valor calculado para T resultó ser -5,249. Dado que el valor de Sig. es igual a 0 y es menor que 0.05, se opta por aceptar la hipótesis alternativa y descartar la nula. Por lo tanto, esto significa que la implementación del proyecto efectivamente mejora la eficiencia de pedidos en las ventas. Además, cabe recalcar que el valor de T se halla en la zona de rechazo, como se puede apreciar en la figura siguiente.

Figura 19: Región de la Prueba T-Student de EP



Fuente: Elaboración propia

Indicador 3:

HE2: El Business Intelligence utilizando minería de datos mejora el crecimiento de clientes.

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora el crecimiento de clientes.

$$H_0 = CCa \geq CCd$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos mejora el crecimiento de clientes.

$$H_a = CCa < CCd$$

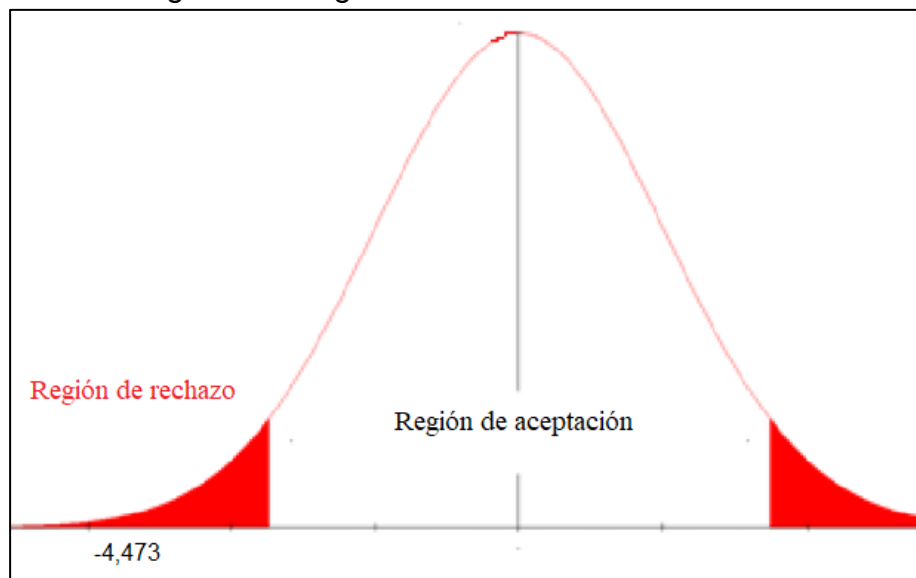
Tabla 16: Prueba T-Student Crecimiento de Clientes

Prueba de T						
	Media	N	Desviación típ.	t	gl	Sig. (bilateral)
CC_PreTest	38,54	15	11,661	-4,473	14	,001
CC_PostTest	48,28	15	13,972			

Fuente: Elaboración propia

Para la evaluación de la hipótesis de este indicador, se empleó T-Student, lo cual los datos utilizados son del pre-test y el post-test. El valor calculado para T resultó ser -4,473. Dado que el valor de Sig. es menor que 0.05, se opta por aceptar la hipótesis alternativa y descartar la nula. Por lo tanto, esto significa que la implementación del proyecto efectivamente mejora el crecimiento de clientes. Además, cabe recalcar que el valor de T se halla en la zona de rechazo, como se puede apreciar en la figura siguiente.

Figura 20: Región de la Prueba T-Student de CC



Fuente: Elaboración propia

Indicador 4:

HE2: El Business Intelligence utilizando minería de datos mejora el índice de rotación de inventario para las ventas

Hipótesis nula (H0): El Business Intelligence utilizando minería de datos no mejora el índice de rotación de inventario para las ventas

$$H0 = IRIa \geq IRIb$$

Hipótesis alternativa (Ha): El Business Intelligence utilizando minería de datos no mejora el índice de rotación de inventario para las ventas

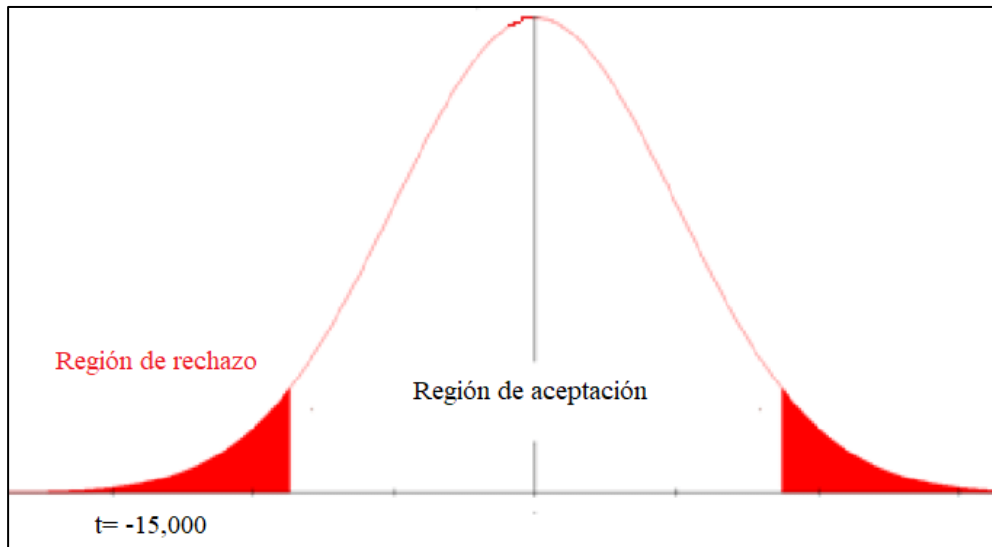
$$Ha = IRIa < IRIb$$

Tabla 17: Prueba T-Student Índice de Rotación de Inventario

Prueba de T						
	Media	N	Desviación típ.	t	gl	Sig. (bilateral)
IRI_PreTest	4,44	20	1,290	-15,000	19	0,000
IRI_PostTest	6,72	20	1,473			

Para la evaluación de la hipótesis de este indicador, se empleó T-Student, lo cual los datos utilizados son del pre-test y el post-test. El valor calculado para T resultó ser -15,000. Dado que el valor de Sig. es igual a 0 y es menor que 0.05, se opta por aceptar la hipótesis alternativa y descartar la nula. Por lo tanto, esto significa que la implementación del proyecto efectivamente mejora el índice de rotación de inventario para las ventas. Además, cabe recalcar que el valor de T se halla en la zona de rechazo, como se puede apreciar en la figura siguiente.

Figura 21: Región de la Prueba T-Student de IRI



Fuente: Elaboración propia

V. DISCUSIÓN

Conforme a los resultados obtenidos de los indicadores de la presente investigación, se realizará una comparativa con otros estudios relacionados al tema abordado, con el fin de ver si existe similitudes de mejora después de implementar dichos proyectos. Esta comparación permitirá obtener una vista de manera general sobre los resultados, viendo el impacto de diversas áreas aplicadas.

El crecimiento de ventas ha aumentado de 24% a 28%, evidenciando una diferencia de 4%, así mismo la eficiencia de pedidos ha aumentado de 80% a 91%, evidenciando una diferencia de 11%. Por otro lado, el índice de rotación de inventario ha aumentado de 4 a 7 rotaciones, evidenciando una diferencia de 3. Por último, el crecimiento de clientes ha aumentado de 39% a 48%, notando la diferencia de 9%. Estos resultados indican que la implementación del proyecto fue adecuada debido a las mejoras obtenidas.

De acuerdo a nuestros resultados positivos para cada indicador de estudio, se realiza un comparativo general de la tecnología implementada, en este caso se comenzará con la inteligencia de negocios. Por ello, el estudio [3], destaca la importancia de utilizar esta tecnología para recopilar y analizar datos en la gestión empresarial. Enfatiza que muchas pequeñas empresas no aprovechan la tecnología para recopilar los datos necesarios para gestionar con éxito sus operaciones. Por otro lado, el estudio [4], indica que la implementación de tecnologías de Business Intelligence beneficia significativamente a las pequeñas empresas, mejorando así su gestión empresarial.

También, este estudio [7], aborda sobre la inteligencia de negocios en empresas, donde sus resultados destacan que esta tecnología tiene muy buena capacidad para optimizar la obtención de información, mejorar el monitoreo de indicadores y convertirse en una ventaja competitiva. A su vez, el estudio [10], se enfoca en la Inteligencia empresarial en los Juzgados Civiles de Huaura, donde obtuvo como resultado, que el nivel de respuesta a la atención de los reclamos fue de un 87.50%, junto con el nivel de satisfacción del producto, que mejoró un 19.99%, el nivel del servicio en un 40.89% y por último se obtuvo un incremento del 30.11% en los niveles de venta.

Haciendo mención ahora a nuestra otra tecnología, que es la minería de datos, el cual se basó nuestra metodología de estudio y respectivos análisis, tanto como predicciones y patrones de los datos, se compara con otros estudios relacionados.

El estudio [5], se centra en aplicar minería de datos y la red neuronal BP y así optimar la exactitud de los pronósticos de ventas y beneficios. Los resultados demostraron que esta aplicación ayuda a las organizaciones a tomar estrategias más fundamentadas en cuanto a precios, inventarios y satisfacción del cliente. Destacando que esta tecnología mejora la efectividad y así poder reducir costos, además de incrementar la satisfacción del consumidor y la lealtad a la marca.

A su vez, el estudio [9], aplica minería de datos en una central hidroeléctrica y cómo puede generar información necesaria para una toma de decisiones oportuna. La cual utilizó la metodología CRISP-DM, teniendo como resultados que esta metodología debe ser usada si existen datos históricos y cuya finalidad sea prevenir o evitar algún problema en específico.

Por otro lado, en este presente proyecto, se usaron otras técnicas de análisis, lo cual generó algoritmos diferentes, pero dando los mismos resultados favorables, del estudio [11], el cual muestra una alta precisión en los resultados de la minería de datos, identificando tres clusters distintos utilizando el algoritmo k-means. Estos clusters pueden brindar información ventajosa sobre el comportamiento de los interesados, lo que puede mejorar la estrategia de marketing de una empresa.

En conjunto, estos estudios brindan una visión completa y coherente de cómo la tecnología, la minería de datos y la inteligencia de negocios pueden ser fundamentales en la toma de decisiones informadas en diversos contextos empresariales y organizacionales, destacando su importancia en la mejora de la eficiencia, la competitividad y la adaptabilidad en el entorno empresarial actual.

Por lo tanto, como anteriormente mencionado en algunos estudios más cercanos al tema abordado, se puede ver que estas tecnologías son esenciales e importantes al momento de implementarlas, dando resultados favorables a la empresa.

VI. CONCLUSIONES

- Respecto a los resultados alcanzados de cada indicador, se determina que la implementación del Business Intelligence utilizando minería de datos para la toma de decisiones mejoró las ventas.
- El indicador de crecimiento de ventas ha experimentado un aumento del 24% al 28%, lo que representa una diferencia del 4%. Este incremento en las ventas es un reflejo directo de la efectividad del proyecto en impulsar el crecimiento de la empresa.
- La eficiencia de pedidos ha mejorado sustancialmente, pasando de un 80% a un 91%, lo que representa una diferencia del 11%. Esto indica que la empresa ha logrado optimizar sus procesos de gestión de pedidos de manera significativa.
- El aumento en el índice de rotación de inventario, de 4 a 7 rotaciones, demuestra que la empresa está gestionando su inventario de manera más efectiva. Esto puede traducirse en una reducción de costos de almacenamiento y una mayor liquidez, ya que los productos se venden más rápidamente.
- El crecimiento de clientes ha aumentado del 39% al 48%, con una diferencia del 9%. Esto sugiere que el proyecto no solo ha ayudado a retener a los clientes existentes, sino que también ha atraído a nuevos clientes, lo que es fundamental para la empresa.
- En conjunto, estos resultados son indicativos de que la implementación del proyecto ha sido adecuada y ha generado mejoras significativas en varios aspectos clave del negocio. Estas mejoras contribuyen a fortalecer la posición competitiva de la empresa y a garantizar su crecimiento a largo plazo.

VII. RECOMENDACIONES

1. Dado que la implementación del proyecto ha demostrado ser efectiva en el aumento de las ventas y la eficiencia operativa, es fundamental mantener y fortalecer el uso de estas herramientas en la empresa. Esto implica continuar la recopilación y análisis de datos para tomar decisiones informadas y adaptar estrategias en tiempo real.
2. La mejora en los indicadores de crecimiento de ventas, eficiencia de pedidos, rotación de inventario y crecimiento de clientes son positivos, pero deben seguir siendo monitoreados y en todo caso tener un enfoque más profundo.
3. Considerar ampliar las fuentes de datos, esto puede incluir datos de redes sociales, datos de mercado y otros que puedan proporcionar información adicional para la toma de decisiones estratégicas.
4. Es importante tener información más relevante al tema implementado, ya que esto ayudará a tener más enfoques de estudio y poder brindar más soluciones utilizando otras herramientas o técnicas de datos.

REFERENCIAS

[1] THAKAR, S. A Pipeline for Business Intelligence and Data-Driven Root Cause Analysis on Categorical Data. *Lecture Notes in Networks and Systems* [online]. 2023, 587, 389–398. ISSN 2367-3370. https://doi.org/10.1007/978-981-19-7874-6_28

[2] REBELO, F.J.P. Use of data mining techniques to explain the primary factors influencing water sensitivity of asphalt mixtures. *Construction & building materials* [online]. 2022, 342. ISSN 0950-0618. <https://doi.org/10.1016/j.conbuildmat.2022.128039>

[3] D'ARCONTE, C. Business intelligence applied in small size for profit companies. *Procedia Computer Science* [online]. 2018, 131, 45–57. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.04.184>

[4] HOU, R. Marketing Decision Support System Based on Data Mining Technology. *Applied sciences*. [online]. 2023, 13(7). ISSN 2076-3417. <https://doi.org/10.3390/app13074315>

[5] JIAN, L. Design of enterprise human resources decision support system based on data mining. *Soft Computing* [online]. 2022, 26(20), 10571–10580. ISSN 1432-7643. <https://doi.org/10.1007/s00500-021-06659-4>

[6] GARCÍA-JIMÉNEZ, Alejandra De-Jesús et al. La Inteligencia De Negocios: Herramienta Clave Para El Uso De La Información Y La Toma De Decisiones Empresariales. *Revista de Investigaciones Universidad del Quindío* [en línea]. 2021, 33(1), 132–139 [consultado el 15 de julio de 2023]. ISSN 2500-5782.

Disponible en: <https://doi.org/10.33975/riuq.vol33n1.514>

[7] VITERI-CEVALLOS, Cristóbal Josué y Dayana Yasmín MURILLO-PÁRRAGA. Inteligencia de negocios para las organizaciones. *Revista Arbitrada Interdisciplinaria Koinonía* [en línea]. 2021, 6(12), 304 [consultado el 15 de julio de 2023]. ISSN 2542-3088. Disponible en: <https://doi.10.35381/r.k.v6i12.1291>

[8] RISCO, Redy H., PÉREZ, Daniel A., CASAVARDE, Luis A. and LA-ROSA, Gerson. Aplicación de técnicas de minería de datos para determinar un modelo de operación de un turbogenerador hidráulico [en línea]. *Información tecnológica*. 2022. Vol. 33, no. 6p. 31–44. <https://doi.org/10.4067/s0718-07642022000600031>

[9] TASA, M., LAGOS, R. y SUCARI, W. Efficient use of data and knowledge transfers in information systems. *Revista Innova Educación*. 2022. Vol. 4, ISSN 2664-1496. <https://dialnet.unirioja.es/servlet/articulo?codigo=8326425>

- [10] Rios, J. J. (2022). Business intelligence and its influence on decision making civil courts superior court of justice, Huaura [Inteligencia de negocios y su influencia en la toma de decisiones Juzgados Civiles Corte Superior de Justicia, Huaura]. 1st LACCEI International Multiconference on Entrepreneurship, Innovation and Regional Development - LEIRD 2021. <http://dx.doi.org/10.18687/LEIRD2021.1.1.23>
- [11] ALI, E. Role of Business Intelligence and Knowledge Management in Solving Business Problems. Technical journal: scientific professional journal of University North [online]. 2022, 16(3), 371–378. ISSN 1846-6168. <https://doi.org/doi:10.31803/tg-20220531145604>
- [12] HURTADO, C. L., & Álvarez, L. E. (2020). Tendencias del comportamiento de consumidores mediante herramientas de Data Mining, en supermercados del Perú: una revisión de la literatura científica. Universidad Privada del Norte. Disponible en: <https://hdl.handle.net/11537/25688>
- [13] VELA, Mirko Bruno. Application of a data mining model to predict customer defection. case of a telecommunications company in peru. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications [en línea]. 2023, 14(1), 144–158 [consultado el 15 de julio de 2023]. Disponible en: <https://doi.org/doi:10.58346/jowua.2023.i1.012>
- [14] KHATIBI, V. Deployment of a business intelligence model to evaluate Iranian national higher education. Social Sciences and Humanities Open [online]. 2020, 2(1). ISSN 2590-2911. <https://doi.org/10.1016/j.ssaho.2020.100056>
- [15] OLSZAK, C.M. Business Intelligence Systems for Innovative Development of Organizations. *Procedia Computer Science* [online]. 2022, 207, 1754–1762. ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2022.09.233>
- [16] GOTTFRIED, A. Mining open government data for business intelligence using data visualization: A two-industry case study. Journal of theoretical and applied electronic commerce research [online]. 2021, 16(4), 1042–1065. ISSN 0718-1876. <https://doi.org/10.3390/JTAER16040059>
- [17] IRIS, N. Implementation of Business Intelligence Solution for United Airlines. International journal of advanced computer science & applications. [online]. 2023, 14(1), 843–852. ISSN 2156-5570. <https://doi.org/10.14569/IJACSA.2023.0140192>
- [18] KONGTHANASUWAN, T. [et al.]. Market Analysis with Business Intelligence System for Marketing Planning. Information [online]. 2023, 14(2). ISSN 2078-2489. <https://doi.org/10.3390/info14020116>

- [19] DUQUE, J. Knowledge data extraction for business intelligence A design science research approach. *Procedia Computer Science* [online]. 2022, 204, 131–139. ISSN 1877-0509.
<https://doi.org/10.1016/j.procs.2022.08.016>
- [20] BANY, A. [et al.]. Business Intelligence and Analytics (BIA) Usage in the Banking Industry Sector: An Application of the TOE Framework. *Journal of Open Innovation: Technology, Market, and Complexity* [online]. 2022, 8, no. 4p. 189.
<https://doi.org/10.3390/joitmc8040189>
- [21] HALIM, K.K. Business intelligence for designing restaurant marketing strategy: A case study. *Procedia Computer Science* [online]. 2019, 161, 615–622. ISSN 1877-0509.
<https://doi.org/10.1016/j.procs.2019.11.164>
- [22] MADYATMADJA, E. HAKIM, A. and SEMBIRING, D. Performance testing on Transparent Data Encryption for SQL Server's reliability and efficiency. *Journal of Big Data* [online]. 2021, 8, no. 1.
<https://doi.org/10.1186/s40537-021-00520-z>
- [23] DUQUE, J. SILVA, F. and GODINHO, A. Data Mining applied to Knowledge Management. *Procedia Computer Science*. [online]. 2023, 219, 455-461. ISSN 1877-0509.
<https://doi.org/10.1016/j.procs.2023.01.312>
- [24] GAZZAWE, F. Data Mining and Soft Computing in Business Model for Decision Support System. *Scientific programming* [online]. 2022. ISSN 1058-9244.
<https://doi.org/10.1155/2022/9147444>
- [25] SHAABAN, A.G. Knowledge discovery in manufacturing datasets using data mining techniques to improve business performance. *Indonesian Journal of Electrical Engineering and Computer Science* [online]. 2022, 26(3), 1736–1746. ISSN 2502-4752.
<https://doi.org/10.11591/ijeecs.v26.i3.pp1736-1746>
- [26] MERLINI, Donatella y ROSSINI, Martina. Text categorization with WEKA: A survey. *Machine Learning with Applications*. [online]. 2021, 4, 100033. ISSN 2666-8270.
<https://doi.org/10.1016/j.mlwa.2021.100033>
- [27] SCHRÖER, C. A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science* [online]. 2021, 181, pp 526–534. ISSN 1877-0509.
<https://doi.org/10.1016/j.procs.2021.01.199>
- [28] HUBER, S. [et al.]. DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM mode. *Procedia CIRP* [online].

2019, 79. 403-408. ISSN 2212-8271.
<https://doi.org/10.1016/j.procir.2019.02.106>

[29] MORLOCK, F. Concept for Enabling Customer-oriented Data Analytics via Integration of Production Process Improvement Methods and Data Science Methods. *Procedia CIRP* [online]. 2021, 104, 542–546. ISSN 2212-8271.
<https://doi.org/10.1016/j.procir.2021.11.091>

[30] VAN DER VOORT, H. Data science as knowledge creation a framework for synergies between data analysts and domain professionals. *Technological forecasting and social change* [online]. 2021, 173. ISSN 0040-1625.
<https://doi.org/10.1016/j.techfore.2021.121160>

[31] SÁNCHEZ, A. Sustainable e-Learning by Data Mining—Successful Results in a Chilean University. *Sustainability*. [online]. 2023, 15(2). ISSN 2071-1050.
<https://doi.org/10.3390/su15020895>

[32] SKARPATHIOTAKI, C.G. Cross-Industry Process Standardization for Text Analytics. *Big Data Research* [online]. 2022, 27. ISSN 2214-5796.
<https://doi.org/10.1016/j.bdr.2021.100274>

[33] PLOTNIKOVA, V. Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data and Knowledge Engineering* [online]. 2022, 139. ISSN 0169-023X.
<https://doi.org/10.1016/j.datak.2022.102013>

[34] ¿Qué es la investigación aplicada y cuáles son sus principales características? En: *IBERO tijuana* [en línea]. Disponible en: <https://blogposgrados.tijuana.iberomex.mx/investigacion-aplicada/> [fecha de consulta: 20 de mayo de 2023].

[35] ESCOBAR, P. y BILBAO, J. INVESTIGACION Y EDUCACION SUPERIOR [en línea]. 2.ª ed. EEUU: Lulu.com, 2020 [fecha de consulta: 22 de mayo del 2023]. Disponible en: https://www.google.com.pe/books/edition/INVESTIGACION_Y_EDUCACION_SUPERIOR/W67WDwAAQBAJ?hl=es&gbpv=0

[36] DAMASCENO, E. Data Mining, Business Intelligence, Grid and Utility Computing: A Bibliometric Review of the Literature from 2015 to 2020. *International Conference on Enterprise Information Systems, ICEIS - Proceedings*. 2021, 1, 367–373. ISSN 2184-4992.
<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85136825901&partnerID=40&md5=28dba0b7a88daf167c2edc10e7bda826>

[37] DA SILVA, D., 2022. 4 estrategias para tener éxito en 2023. En: *zendesk* [en línea]. Disponible en: <https://www.zendesk.com.mx/blog/ventas/> [fecha de consulta: 22 de mayo de 2023].

[38] MARÍN, J., 2019. Desarrollo de una versión electrónica del cuadro de mando integral (CMI) [en línea]. Cartagena: Universidad Politécnica de Cartagena [fecha de consulta: 27 de mayo de 2023]. Disponible en: <http://hdl.handle.net/10317/8079>

[39] PORTUGAL, B., MAEDA, A. 2022. SISTEMA DE GESTIÓN DE INVENTARIOS PARA MEJORAR LA PRODUCTIVIDAD EN EL PROCESO DE RECARGAS DE EXTINTORES DE LA EMPRESA COIMSER S.A.C. CALLAO, 2021. Tesis, Universidad Nacional del Callao. Disponible en: <https://hdl.handle.net/20.500.12952/7616>

[40] ARIAS, J y COVINOS, M. Diseño y metodología de la investigación [en línea]. Perú: Enfoques Consulting EIRL, 2021 [fecha de consulta: 22 de mayo de 2023]. Disponible en: <https://repositorio.concytec.gob.pe/handle/20.500.12390/2260>

[41] HERNÁNDEZ, R. y MENDOZA, C. Metodología de la investigación: las rutas: cuantitativa, cualitativa y mixta [en línea]. México: McGRAW-HILL INTERAMERICANA EDITORES, 2018 [fecha de consulta: 22 de mayo de 2023]. Disponible en: <http://repositorio.uasb.edu.bo:8080/handle/54000/1292>

[42] TAFUR, R. e IZAGUIRRE, M. Cómo hacer un proyecto de investigación [en línea]. Bogotá: Alpha Editorial, 2022 [fecha de consulta: 22 de mayo de 2023]. Disponible en: https://books.google.com.pe/books?id=3tavEAAAQBAJ&newbks=1&newbks_redir=0&hl=es&redir_esc=y

[43] ARIAS, J. Técnicas e instrumentos de investigación científica [en línea]. Arequipa: ENFOQUES CONSULTING EIRL, 2020 [fecha de consulta: 22 de mayo de 2023].

Disponible en: <http://hdl.handle.net/20.500.12390/2238>

[44] MANTEROLA, Carlos et al. Confiabilidad, precisión o reproducibilidad de las mediciones. Métodos de valoración, utilidad y aplicaciones en la práctica clínica. Rev. chil. infectol. [online]. 2018, vol.35, n.6 [citado 2023-06-17], pp.680-688. Disponible en: <http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0716-10182018000600680&lng=es&nrm=iso>. ISSN 0716-1018. <http://dx.doi.org/10.4067/S0716-10182018000600680>

ANEXOS

ANEXO 1: Matriz de Operacionalización de variables

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores	Fórmula	Instrumento	Escala de medición
INDEPENDIENTE Business intelligence	BI actúa como un factor estratégico para una empresa u organización, generando una ventaja competitiva que proporciona información privilegiada para responder a problemas empresariales DAMASCENO (2021).						
DEPENDIENTE ventas	El término "ventas" se utiliza para describir la transacción de bienes y servicios a cambio de una compensación monetaria. Estas transacciones se llevan a cabo debido a necesidades específicas y subjetivas de los clientes, Da Silva (2022).	Las buenas estrategias realizadas en las ventas mejoran eficientemente su crecimiento, así como también sus clientes	Perspectiva Financiera Eficiencia Perspectiva cliente Rotación de inventario	Crecimiento de ventas Eficiencia de Pedidos Crecimiento de clientes Índice de rotación de inventario	$CV = ((\text{venta total actual} - \text{venta total anterior}) / \text{venta total anterior}) \times 100$ $(\text{Pedidos entregados conformes} / \text{total de pedidos realizados}) \times 100$ $CC = ((\text{total cliente actual} - \text{total cliente anterior}) / \text{total cliente anterior}) \times 100$ Costo de venta / inventario promedio	Ficha de registro	RAZÓN

ANEXO 2: Matriz de consistencia

Problemas	Objetivos	Hipótesis	Variables	Dimensiones	Indicadores	Método
General	General	General	Independiente			Tipo de investigación: Aplicada
¿En qué forma la implementación de Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas?	Evaluar cómo la implementación de Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas	El Business Intelligence utilizando minería de datos para la toma de decisiones mejora las ventas	Business intelligence utilizando minería de datos			
Específicos	Específicos	Específicos	Dependiente			Diseño de investigación: Experimental: pre-experimental Población: Registro de 15 días junto con 20 productos Técnica de investigación: Fichaje Instrumento de investigación: Ficha de registro
¿En qué forma la implementación de Business Intelligence utilizando minería de datos ayuda a aumentar las ventas?	Aumentar las ventas mediante la implementación de Business Intelligence utilizando minería de datos	El Business Intelligence utilizando minería de datos mejora el crecimiento de ventas	ventas	Perspectiva Financiera	Crecimiento de ventas	
¿En qué forma la implementación de Business Intelligence utilizando minería de datos ayuda en la eficiencia de pedidos de las ventas?	Mejorar la eficiencia de pedidos mediante la implementación de Business Intelligence utilizando minería de datos en ventas	El Business Intelligence utilizando minería de datos mejora la eficiencia de pedidos en las ventas		Perspectiva cliente	Crecimiento de clientes	
¿En qué forma la implementación de Business Intelligence utilizando minería de datos ayuda a aumentar los clientes en las ventas?	Aumentar los clientes mediante la implementación de Business Intelligence utilizando minería de datos en ventas	El Business Intelligence utilizando minería de datos mejora el crecimiento de clientes		Eficiencia	Eficiencia de Pedidos	
¿En qué forma la implementación de Business Intelligence utilizando minería de datos ayuda en el índice de rotación de inventario para las ventas?	Mejorar el índice de rotación de inventario mediante la implementación de Business Intelligence utilizando minería de datos en ventas	El Business Intelligence utilizando minería de datos mejora el índice de rotación de inventario para las ventas		Rotación de inventario	Índice de rotación de inventario	

ANEXO 3: Ficha de registro de crecimiento de ventas – Pretest

FICHA DE REGISTRO					
Investigador	Tantalean Salazar Billy Brandon				
Organización	Distribuidora San Cristobal			RETEST	
Método de investigación	Pre-experimental				
Fecha inicio	01-ago-23	Fecha Final	31-ago-23		
Variable	Indicador	Medida	Fórmula		
Ventas	crecimiento de ventas	razón	$CV = ((VTA - VTP) / VTP) * 100$		
N°	FECHA (VTP)	FECHA (VTA)	Venta total actual (VTA)	Venta total pasada (VTP)	Crecimiento de Venta (CV)
1	01/08/2023	17/08/2023	S/ 11.779	S/ 10.277	14,62
2	02/08/2023	18/08/2023	S/ 12.105	S/ 10.245	18,16
3	03/08/2023	19/08/2023	S/ 10.403	S/ 8.615	20,75
4	04/08/2023	20/08/2023	S/ 9.502	S/ 7.467	27,25
5	05/08/2023	21/08/2023	S/ 8.925	S/ 7.310	22,09
6	06/08/2023	22/08/2023	S/ 9.085	S/ 7.298	24,49
7	07/08/2023	23/08/2023	S/ 7.054	S/ 5.434	29,81
8	08/08/2023	24/08/2023	S/ 10.879	S/ 8.690	25,19
9	09/08/2023	25/08/2023	S/ 6.558	S/ 5.434	20,68
10	10/08/2023	26/08/2023	S/ 9.800	S/ 8.036	21,95
11	11/08/2023	27/08/2023	S/ 9.466	S/ 7.505	26,13
12	12/08/2023	28/08/2023	S/ 8.735	S/ 6.860	27,33
13	13/08/2023	29/08/2023	S/ 5.845	S/ 4.335	34,83
14	14/08/2023	30/08/2023	S/ 7.800	S/ 5.815	34,14
15	15/08/2023	31/08/2023	S/ 10.380	S/ 9.119	13,83

ANEXO 4: Ficha de registro de crecimiento de clientes – Pretest

FICHA DE REGISTRO					
Investigador	Tantalean Salazar Billy Brandon				
Organización	Distribuidora San Cristobal			RETEST	
Método de investigación	Pre-experimental				
Fecha inicio	01-ago-23	Fecha Final	31-ago-23		
Variable	Indicador	Medida	Fórmula		
Ventas	crecimiento de clientes	razón	$CV = ((TCA - TCP) / TCP) * 100$		
N°	FECHA (TCP)	FECHA (TCA)	Total cliente actual (TCA)	Total cliente pasado (TCP)	Crecimiento de clientes (CC)
1	01/08/2023	17/08/2023	10	8	25,00
2	02/08/2023	18/08/2023	10	7	42,86
3	03/08/2023	19/08/2023	15	10	50,00
4	04/08/2023	20/08/2023	24	20	20,00
5	05/08/2023	21/08/2023	25	17	47,06
6	06/08/2023	22/08/2023	12	8	50,00
7	07/08/2023	23/08/2023	22	15	46,67
8	08/08/2023	24/08/2023	23	20	15,00
9	09/08/2023	25/08/2023	12	9	33,33
10	10/08/2023	26/08/2023	23	16	43,75
11	11/08/2023	27/08/2023	15	10	50,00
12	12/08/2023	28/08/2023	12	8	50,00
13	13/08/2023	29/08/2023	22	17	29,41
14	14/08/2023	30/08/2023	22	16	37,50
15	15/08/2023	31/08/2023	11	8	37,50

ANEXO 5: Ficha de registro de índice de rotación de inventario – Pretest

FICHA DE REGISTRO					
Investigador	Tantalean Salazar Billy Brandon				
Organización	Distribuidora San Cristobal				RETEST
Método de investigación	Pre-experimental				
Fecha inicio	17-jul-23	Fecha Final	31-jul-23		
Variable	Indicador	Medida	Fórmula		
Ventas	Índice de rotación de inventario	razón	$IRI = CVE / IP$		
N°	fecha	Nombre producto	Costo de venta (CVE)	Inventario promedio (IP)	Índice de rotación de inventario (IRI)
1	Julio	Ladrillo de pared	S/ 20.000	S/ 7.500	2,67
2	Julio	Ladrillo de techo	S/ 15.500	S/ 4.500	3,44
3	Julio	Cemento	S/ 27.000	S/ 3.500	7,71
4	Julio	Arena Fina	S/ 13.000	S/ 3.000	4,33
5	Julio	Hormigón	S/ 12.500	S/ 2.500	5,00
6	Julio	Piedra Chancada	S/ 10.000	S/ 2.500	4,00
7	Julio	yeso	S/ 8.000	S/ 1.500	5,33
8	Julio	Varilla de fierro de 3/8	S/ 13.500	S/ 4.200	3,21
9	Julio	Varilla de fierro de 1/4	S/ 5.500	S/ 1.300	4,23
10	Julio	Varilla de fierro de 1/2	S/ 16.000	S/ 2.500	6,40
11	Julio	Martillo	S/ 2.100	S/ 500	4,20
12	Julio	Clavos	S/ 2.100	S/ 350	6,00
13	Julio	Sierra circular	S/ 1.500	S/ 400	3,75
14	Julio	Interruptores de luz	S/ 2.500	S/ 450	5,56
15	Julio	Toma corrientes	S/ 2.000	S/ 600	3,33
16	Julio	cable eléctrico	S/ 4.200	S/ 1.200	3,50
17	Julio	Pintura	S/ 3.400	S/ 900	3,78
18	Julio	Focos de luz	S/ 2.000	S/ 500	4,00
19	Julio	Lijas	S/ 1.000	S/ 340	2,94
20	Julio	Tubo pvc	S/ 3.500	S/ 650	5,38

ANEXO 6: Ficha de registro de eficiencia de pedidos – Pretest

FICHA DE REGISTRO				
Investigador	Tantalean Salazar Billy Brandon			
Organización	Distribuidora San Cristobal			TEST
Método de investigación	Pre-experimental			
Fecha inicio	17-jul-23	Fecha Final	31-jul-23	
Variable	Indicador	Medida	Fórmula	
Ventas	Eficiencia de Pedidos	razón	EP = (PEC / TPR) * 100	
N°	fecha	Pedidos entregados conformes (PEC)	Total pedidos realizados (TPR)	Eficiencia de Pedidos (EP)
1	Julio	7	9	77,78
2	Julio	5	7	71,43
3	Julio	6	7	85,71
4	Julio	5	7	71,43
5	Julio	6	8	75,00
6	Julio	5	7	71,43
7	Julio	7	9	77,78
8	Julio	7	8	87,50
9	Julio	5	6	83,33
10	Julio	7	8	87,50
11	Julio	8	10	80,00
12	Julio	8	10	80,00
13	Julio	7	8	87,50
14	Julio	6	7	85,71
15	Julio	6	8	75,00

ANEXO 7: Ficha de registro de datos Crecimiento de ventas– PostTest

FICHA DE REGISTRO					
Investigador	Tantalean Salazar Billy Brandon				
Organización	Distribuidora San Cristobal			POST-TEST	
Método de investigación	Pre-experimental				
Fecha inicio	02-oct-23	Fecha Final	01-nov-23		
Variable	Indicador	Medida	Fórmula		
Ventas	crecimiento de ventas	razón	$CV = ((VTA - VTP) / VTP) * 100$		
N°	FECHA (VTP)	FECHA (VTA)	Venta total actual (VTA)	Venta total pasada (VTP)	Crecimiento de Venta (CV)
1	02/10/2023	18/10/2023	10568	8997	17,46
2	03/10/2023	19/10/2023	9500	7857	20,91
3	04/10/2023	20/10/2023	11224	9106	23,26
4	05/10/2023	21/10/2023	10872	8348	30,23
5	06/10/2023	22/10/2023	10674	8398	27,10
6	07/10/2023	23/10/2023	9340	7320	27,60
7	08/10/2023	24/10/2023	9635	7198	33,86
8	09/10/2023	25/10/2023	10095	7870	28,27
9	10/10/2023	26/10/2023	9869	8010	23,21
10	11/10/2023	27/10/2023	9595	7784	23,27
11	12/10/2023	28/10/2023	9833	6795	44,71
12	13/10/2023	29/10/2023	11287	8733	29,25
13	14/10/2023	30/10/2023	9115	6529	39,61
14	15/10/2023	31/10/2023	9684	7090	36,59
15	16/10/2023	01/11/2023	10852	9143	18,69

ANEXO 8: Ficha de registro de datos Crecimiento de clientes– PostTest

FICHA DE REGISTRO					
Investigador	Tantalean Salazar Billy Brandon				
Organización	Distribuidora San Cristobal				POST-TEST
Método de investigación	Pre-experimental				
Fecha inicio	02-oct-2023	Fecha Final	01-nov-2023		
Variable	Indicador	Medida	Fórmula		
Ventas	crecimiento de clientes	razón	$CV = ((TCA - TCP) / TCP) * 100$		
N°	FECHA (TCP)	FECHA (TCA)	Total cliente actual (TCA)	Total cliente pasado (TCP)	Crecimiento de clientes (CC)
1	02/10/2023	18/10/2023	15	11	36,36
2	03/10/2023	19/10/2023	20	13	53,85
3	04/10/2023	20/10/2023	16	10	60,00
4	05/10/2023	21/10/2023	22	18	22,22
5	06/10/2023	22/10/2023	25	15	66,67
6	07/10/2023	23/10/2023	16	10	60,00
7	08/10/2023	24/10/2023	21	14	50,00
8	09/10/2023	25/10/2023	23	19	21,05
9	10/10/2023	26/10/2023	20	13	53,85
10	11/10/2023	27/10/2023	27	18	50,00
11	12/10/2023	28/10/2023	23	15	53,33
12	13/10/2023	29/10/2023	20	14	42,86
13	14/10/2023	30/10/2023	23	17	35,29
14	15/10/2023	31/10/2023	25	16	56,25
15	16/10/2023	01/11/2023	26	16	62,50

ANEXO 9: Ficha de registro de datos Índice de rotación de inventario– PostTest

FICHA DE REGISTRO					
Investigador	Tantalean Salazar Billy Brandon				
Organización	Distribuidora San Cristobal			POST-TEST	
Método de investigación	Pre-experimental				
Fecha inicio	02-oct-2023	Fecha Final	16-oct-2023		
Variable	Indicador	Medida	Fórmula		
Ventas	Índice de rotación de inventario	razón	IRI = CVE / IP		
N°	fecha	Nombre producto	Costo de venta (CVE)	Inventario promedio (IP)	Índice de rotación de inventario (IRI)
1	Octubre	Ladrillo de pared	S/ 35.000	S/ 7.500	4,67
2	Octubre	Ladrillo de techo	S/ 25.000	S/ 4.500	5,56
3	Octubre	Cemento	S/ 40.000	S/ 4.300	9,30
4	Octubre	Arena Fina	S/ 15.000	S/ 2.000	7,50
5	Octubre	Hormigón	S/ 20.000	S/ 3.000	6,67
6	Octubre	Piedra Chancada	S/ 25.000	S/ 3.500	7,14
7	Octubre	yeso	S/ 10.000	S/ 1.350	7,41
8	Octubre	Varilla de fierro de 3/8	S/ 21.500	S/ 4.000	5,38
9	Octubre	Varilla de fierro de 1/4	S/ 10.200	S/ 1.300	7,85
10	Octubre	Varilla de fierro de 1/2	S/ 16.000	S/ 1.800	8,89
11	Octubre	Martillo	S/ 2.500	S/ 350	7,14
12	Octubre	Clavos	S/ 2.000	S/ 250	8,00
13	Octubre	Sierra circular	S/ 2.500	S/ 400	6,25
14	Octubre	Interruptores de luz	S/ 2.550	S/ 300	8,50
15	Octubre	Toma corrientes	S/ 2.150	S/ 400	5,38
16	Octubre	cable eléctrico	S/ 4.150	S/ 750	5,53
17	Octubre	Pintura	S/ 3.500	S/ 540	6,48
18	Octubre	Focos de luz	S/ 2.000	S/ 400	5,00
19	Octubre	Lijas	S/ 1.200	S/ 300	4,00
20	Octubre	Tube pvc	S/ 3.500	S/ 450	7,78

ANEXO 10: Ficha de registro de datos Eficiencia de pedidos – PostTest

FICHA DE REGISTRO				
Investigador	Tantalean Salazar Billy Brandon			
Organización	Distribuidora San Cristobal		POST-TEST	
Método de investigación	Pre-experimental			
Fecha inicio	02-oct-2023	Fecha Final	16-oct-2023	
Variable	Indicador	Medida		Fórmula
Ventas	Eficiencia de Pedidos	razón		$EP = (PEC / TPR) * 100$
N°	fecha	Pedidos entregados conformes (PEC)	Total pedidos realizados (TPR)	Eficiencia de Pedidos (EP)
1	Octubre	10	11	90,91
2	Octubre	7	8	87,50
3	Octubre	9	10	90,00
4	Octubre	8	9	88,89
5	Octubre	9	10	90,00
6	Octubre	10	11	90,91
7	Octubre	7	8	87,50
8	Octubre	6	7	85,71
9	Octubre	9	9	100,00
10	Octubre	11	12	91,67
11	Octubre	12	13	92,31
12	Octubre	13	14	92,86
13	Octubre	11	12	91,67
14	Octubre	9	11	81,82
15	Octubre	10	10	100,00

ANEXO 11: DESARROLLO DE LA METODOLOGÍA CRISP-DM

A. Comprensión del negocio

1. Establecer los objetivos del negocio

1.1. Problema

El problema del negocio se explicó dentro del proyecto, precisamente en la introducción, por ende, ya no es necesario completar este punto.

1.2. Objetivos del negocio

Los objetivos definidos están orientados más a la minería de datos a diferencia de los objetivos del proyecto, por lo que se consideró los siguientes objetivos:

- Realizar análisis predictivos de los productos
- Realizar análisis para descubrir patrones de los productos

1.3. Criterios de éxito del negocio

Desde el punto de vista del negocio se establece como criterio de éxito la posibilidad de realizar predicciones y patrones sobre los productos vendidos con un buen porcentaje de fiabilidad, de tal forma que se mejore la toma de decisiones para mejorar las ventas, ya que se tendrá un conocimiento sobre qué productos son de mayor relevancia para el negocio.

2. Evaluar el problema

2.1. Inventario de recursos

En cuanto a recursos de software se dispone de Power BI Desktop, Microsoft Sql Server (Developer) y WEKA. Los recursos de hardware de los que disponemos son un ordenador de sobremesa con 8GB de ram y un almacenamiento de 500GB.

La fuente de datos es el programa Excel donde se almacena la información de ventas y otras áreas del negocio.

2.2. Requisitos, supuestos y restricciones

Requisitos del proyecto:

Identificar los principales patrones y predicciones de los productos vendidos en un determinado tiempo.

Describir los supuestos:

Los datos se seleccionarán de acuerdo a los atributos necesarios para cada análisis como cantidad, producto, entre otros.

Comprobar las restricciones:

Para el procesamiento y valoración de los datos se han solicitado para poder insertarlo en nuestra base de datos, el cual nos servirá para seleccionar los datos necesarios de manera óptima y rápida.

3. Establecer los objetivos de la minería de datos

- Modelar los datos adecuadamente para su análisis
- Encontrar predicciones de los productos vendidos
- Encontrar patrones de los productos vendidos
- Realizar análisis con diferentes algoritmos

4. Generar el plan del proyecto

Tabla: Plan de Proyecto

PLAN DE PROYECTO		
FASE	FECHA INICIAL	FECHA FINAL
Comprensión del negocio	01/08/2023	07/08/2023
Comprensión de los datos	08/08/2023	22/08/2023
Preparación de los datos	23/08/2023	31/08/2023
Modelamiento	01/09/2023	13/09/2023
Evaluación	14/09/2023	21/09/2023
Despliegue	22/09/2023	30/09/2023

Fuente: Elaboración Propia

B. Comprensión de los datos

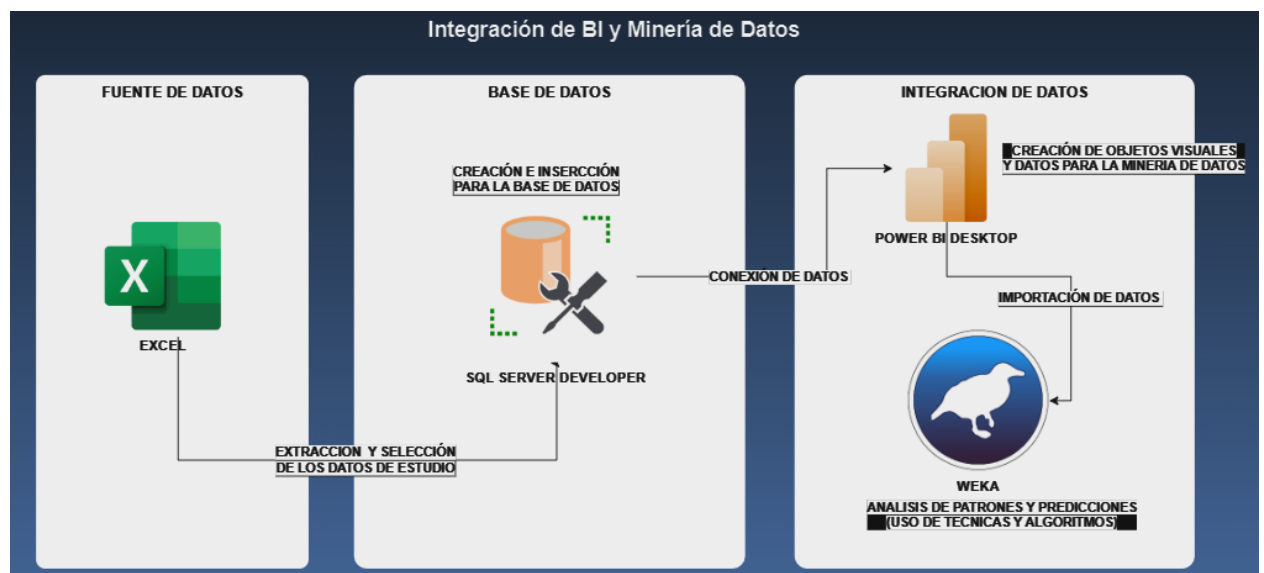
1. Recolectar datos iniciales

Los datos utilizados en este proyecto son datos referentes a las ventas del negocio, en este caso son los reportes diarios, el cual incluye todas las ventas del día, a su vez el

esquema general de la base de datos, como la tabla clientes incluyen información personal como puede ser, sus nombres y apellidos, correo, etc., por lo que no se ha podido utilizar estos datos reales, debido a impedimentos de la empresa. Por lo tanto, se ha tenido que crear y utilizar datos ficticios que hagan referencia a estos, pero se ha mantenido la relación de sus pedidos o ventas con los datos reales, los cuales se han referenciado con su ID_Cliente, al final los datos mencionados anteriormente no perjudican a los objetivos del proyecto. Así que, esta información estará exclusivamente en la visualización general que se presentará a través de Power BI. Esto permitirá observar de manera específica cómo se lleva a cabo un análisis más efectivo de los datos. Posteriormente, la empresa podrá comenzar a generar su propia información con sus datos reales referente a la tabla mencionada basándose en estos resultados.

1.1. Arquitectura Para la integración de Business Intelligence y Minería de Datos

En la siguiente figura, se muestran los pasos de cómo se integra BI con minería de datos, haciendo uso de los softwares ya mencionados y mostrados en la imagen.



Ahora se detallará los pasos que se han seguido para la integración de estas tecnologías:

1.1.1. Capa de Extracción y Almacenamiento de Datos:

Los datos que se van a utilizar son los mencionados en la población. Los datos se extraen de un Excel generado por la empresa, el cual se le dará prioridad solo a los productos y relación de ventas como se presenta en la siguiente imagen:

PEDIDO	CLIENTE	FECHA	MONTO	PRODUCTOS	VENTA	DEPOSITO	FECHA DE VENTA	MONTO PAGADO	PRODUCTOS	FECHA	DIAL VENTA DEL DIA
PE01	C4	01-04-2023	\$1.400	Ladrillo de pared, Ladrillo de techo, Cemento, arena fina	V1	NALL	2023-07-01	\$170	Mantillo	01-04-2023	3700
PE02	C6	01-04-2023	\$1.150	Ladrillo de techo, arena fina, Hormigon, Piedra chancada, yeso	V2	NALL	2023-07-01	\$1.200	Interruptores de luz	02-04-2023	650
PE03	C9	01-04-2023	\$2.200	Ladrillo de pared, Ladrillo de techo, Cemento	V3	NALL	03/07/2023	\$138	Toma conectores	03-04-2023	6700
PE04	C50	02-04-2023	\$1.435	Ladrillo de techo, yeso, Varilla de hierro de 3/8, Varilla de hierro de 1/2	V4	NALL	03/07/2023	\$19	Llantas	04-04-2023	540
PE05	C17	02-04-2023	\$1.050	Ladrillo de pared, Cemento, arena fina	V5	NALL	03/07/2023	\$138	rocos de luz	05-04-2023	670
PE06	C17	02-04-2023	\$1.400	Ladrillo de pared, Ladrillo de techo, Cemento, yeso	V6	NALL	04/07/2023	\$170	Clavos	05-04-2023	640
PE07	C38	03-04-2023	\$1.350	Ladrillo de pared, Ladrillo de techo, arena fina, Hormigon, Piedra chancada	V7	NALL	04/07/2023	\$130	Clavos	07-04-2023	6570
PE08	C10	04-04-2023	\$1.850	Ladrillo de pared, Ladrillo de techo, Cemento	V8	NALL	04/07/2023	\$120	Sierra circular	08-04-2023	3340
PE09	C40	05-04-2023	\$1.250	Ladrillo de techo, Cemento, yeso, Varilla de hierro de 3/8	V9	NALL	04/07/2023	\$148	tubo pvc	08-04-2023	7637
PE10	C12	05-04-2023	\$1.200	Ladrillo de pared, Ladrillo de techo, Cemento, Varilla de hierro de 3/8, Varilla de hierro de 1/2	V10	NALL	05/07/2023	\$170	Mantillo, Sierra circular	10-04-2023	6340
PE11	C9	05-04-2023	\$1.000	Ladrillo de pared, Cemento	V11	NALL	05/07/2023	\$160	Clavos, Mantillo	10-04-2023	7660
PE12	C40	06-04-2023	\$1.350	yeso, Varilla de hierro de 3/8	V12	NALL	06/07/2023	\$148	tubo pvc	12-04-2023	6550
PE13	C39	06-04-2023	\$1.435	Varilla de hierro de 3/8, Varilla de hierro de 1/4, Varilla de hierro de 1/2	V13	NALL	06/07/2023	\$120	Pinura, llantas, tubo pvc	12-04-2023	7120
PE14	C9	07-04-2023	\$1.950	Ladrillo de techo, Cemento	V14	NALL	06/07/2023	\$150	cable electrico	14-04-2023	4704
PE15	C9	07-04-2023	\$1.700	Ladrillo de pared, Ladrillo de techo, Cemento	V15	NALL	07/07/2023	\$138	Clavos	15-04-2023	7300
PE16	C9	08-04-2023	\$1.300	Cemento, yeso, Varilla de hierro de 1/4	V16	NALL	07/07/2023	\$15	Llantas	16-04-2023	1285
PE17	C18	08-04-2023	\$1.950	Ladrillo de pared, Ladrillo de techo, Cemento	V17	NALL	07/07/2023	\$15	Clavos	17-04-2023	6540
PE18	C11	08-04-2023	\$1.610	Ladrillo de techo, Varilla de hierro de 3/8, Varilla de hierro de 1/4	V18	NALL	08/07/2023	\$138	Pinura, rocos de luz	18-04-2023	5400
PE19	C38	08-04-2023	\$1.240	Cemento, arena fina, Hormigon, Piedra chancada	V19	NALL	08/07/2023	\$135	Toma conectores, cable electrico	18-04-2023	7664
PE20	C4	08-04-2023	\$1.350	Ladrillo de pared, Ladrillo de techo, Cemento, Varilla de hierro de 3/8, Varilla de hierro de 1/2	V20	NALL	08/07/2023	\$148	Mantillo	20-04-2023	1663
PE21	C15	08-04-2023	\$1.415	Cemento, Varilla de hierro de 1/2	V21	NALL	08/07/2023	\$130	Sierra circular	21-04-2023	6860
PE22	C17	08-04-2023	\$1.950	yeso, Varilla de hierro de 1/4, Varilla de hierro de 1/2	V22	NALL	08/07/2023	\$135	Clavos, Sierra circular	22-04-2023	648
PE23	C38	08-04-2023	\$1.200	Ladrillo de pared, Cemento, arena fina, Hormigon, Piedra chancada	V23	NALL	08/07/2023	\$125	Mantillo	23-04-2023	6564
PE24	C41	08-04-2023	\$1.520	Cemento, yeso, Varilla de hierro de 1/4	V24	NALL	10/07/2023	\$130	Toma conectores	24-04-2023	9939
PE25	C4	10-04-2023	\$1.200	arena fina, Hormigon, Piedra chancada, yeso	V25	NALL	09/07/2023	\$160	Pinura	25-04-2023	6507
PE26	C16	10-04-2023	\$1.320	Ladrillo de pared, Ladrillo de techo, Cemento	V26	NALL	09/07/2023	\$144	tubo pvc	26-04-2023	6520
PE27	C18	10-04-2023	\$1.250	Ladrillo de pared, Ladrillo de techo, Cemento	V27	NALL	09/07/2023	\$135	Mantillo, Clavos	27-04-2023	6995
PE28	C46	10-04-2023	\$1.900	Ladrillo de pared, Ladrillo de techo, Cemento	V28	NALL	10/07/2023	\$130	Pinura, llantas	28-04-2023	7107
PE29	C18	12-04-2023	\$1.250	Ladrillo de techo, Cemento, arena fina	V29	NALL	10/07/2023	\$132	tubo pvc	29-04-2023	6745
PE30	C17	12-04-2023	\$1.235	yeso, Varilla de hierro de 3/8, Varilla de hierro de 1/4, Varilla de hierro de 1/2	V30	NALL	10/07/2023	\$130	cable electrico	30-04-2023	6140
PE31	C38	13-04-2023	\$1.200	Ladrillo de pared, Ladrillo de techo, Cemento	V31	NALL	10/07/2023	\$132	cable electrico	31-04-2023	6362
PE32	C18	13-04-2023	\$1.200	Ladrillo de pared, Ladrillo de techo, Cemento	V32	NALL	10/07/2023	\$125	Toma conectores, cable electrico	10-ago-2023	8077
PE33	C43	14-04-2023	\$1.430	Ladrillo de pared, Ladrillo de techo, Cemento, arena fina, Hormigon, Piedra chancada	V33	NALL	10/07/2023	\$138	rocos de luz	10-ago-2023	9245
PE34	C6	15-04-2023	\$1.225	Varilla de hierro de 3/8, Varilla de hierro de 1/4, Varilla de hierro de 1/2	V34	NALL	14/07/2023	\$130	Pinura, rocos de luz	03-ago-2023	839
PE35	C47	15-04-2023	\$1.625	Varilla de hierro de 3/8, Varilla de hierro de 1/2	V35	NALL	14/07/2023	\$138	Clavos	04-ago-2023	1467
PE36	C12	15-04-2023	\$1.100	Varilla de hierro de 3/8, Varilla de hierro de 1/4	V36	NALL	14/07/2023	\$135	Clavos	05-ago-2023	730
PE37	C21	15-04-2023	\$1.400	Varilla de hierro de 1/2	V37	NALL	14/07/2023	\$150	Mantillo	05-ago-2023	7280
PE38	C12	16-04-2023	\$1.800	Ladrillo de techo, Cemento	V38	NALL	14/07/2023	\$170	Mantillo, Interruptores de luz, Toma conectores	07-ago-2023	6414
PE39	C36	16-04-2023	\$1.225	Ladrillo de pared, Cemento, yeso	V39	NALL	15/07/2023	\$150	cable electrico	08-ago-2023	6930
PE40	C29	16-04-2023	\$1.225	Cemento, yeso, Piedra chancada	V40	NALL	15/07/2023	\$150	cable electrico	08-ago-2023	6414
PE41	C19	17-04-2023	\$1.650	Cemento	V41	NALL	16/07/2023	\$125	Mantillo, Sierra circular	10-ago-2023	6636
PE42	C4	17-04-2023	\$1.425	Yeso	V42	NALL	16/07/2023	\$138	Clavos	11-ago-2023	7600
PE43	C16	17-04-2023	\$1.430	Ladrillo de pared, Cemento	V43	NALL	16/07/2023	\$135	Clavos	11-ago-2023	6980
PE44	C10	17-04-2023	\$1.200	Ladrillo de pared	V44	NALL	17/07/2023	\$130	Clavos	11-ago-2023	6335
PE45	C40	18-04-2023	\$1.435	Cemento, yeso, Varilla de hierro de 3/8, Varilla de hierro de 1/4	V45	NALL	17/07/2023	\$135	Clavos	11-ago-2023	6306
PE46	C17	18-04-2023	\$1.150	Varilla de hierro de 3/8	V46	NALL	18/07/2023	\$145	Mantillo, Clavos	11-ago-2023	319
PE47	C21	18-04-2023	\$1.220	Varilla de hierro de 1/4, Varilla de hierro de 1/2	V47	NALL	18/07/2023	\$138	Clavos	11-ago-2023	6262
PE48	C9	18-04-2023	\$1.150	Cemento	V48	NALL	18/07/2023	\$125	Interruptores de luz	11-ago-2023	6775
PE49	C49	19-04-2023	\$1.150	Cemento	V49	NALL	19/07/2023	\$125	Mantillo	11-ago-2023	1205
PE50	C39	19-04-2023	\$1.450	Varilla de hierro de 3/8, Varilla de hierro de 1/4, Varilla de hierro de 1/2	V50	NALL	19/07/2023	\$125	Interruptores de luz, Toma conectores	11-ago-2023	6943
PE51	C2	19-04-2023	\$1.440	Ladrillo de techo	V51	NALL	20/07/2023	\$165	Toma conectores, cable electrico	10-ago-2023	9602
PE52	C48	20-04-2023	\$1.350	arena fina, Hormigon, Piedra chancada	V52	NALL	20/07/2023	\$144	tubo pvc	11-ago-2023	6525
PE53	C26	20-04-2023	\$1.150	Cemento	V53	NALL	22/07/2023	\$130	Ladrillo de pared, Cemento	10-ago-2023	6965
PE54	C48	20-04-2023	\$1.600	yeso	V54	NALL	22/07/2023	\$151	yeso	10-ago-2023	7054
PE55	C26	21-04-2023	\$1.800	yeso	V55	NALL	22/07/2023	\$130	Clavos	10-ago-2023	6979
PE56	C14	21-04-2023	\$1.250	Ladrillo de pared, Ladrillo de techo, Cemento	V56	NALL	23/07/2023	\$128	Varilla de hierro de 3/8, Varilla de hierro de 1/4, Varilla de hierro de 1/2	05-ago-2023	6558
PE57	C11	21-04-2023	\$1.650	Ladrillo de pared	V57	NALL	23/07/2023	\$138	Cemento	05-ago-2023	6988
PE58	C50	21-04-2023	\$1.150	Cemento	V58	NALL	24/07/2023	\$174	yeso, Mantillo, Clavos	07-ago-2023	3466
PE59	C17	21-04-2023	\$1.150	Ladrillo de pared, Ladrillo de techo	V59	NALL	24/07/2023	\$130	rocos de luz	08-ago-2023	6735
PE60	C19	21-04-2023	\$1.150	Varilla de hierro de 3/8	V60	NALL	24/07/2023	\$160	tubo pvc	08-ago-2023	6545
PE61	C48	22-04-2023	\$1.315	Varilla de hierro de 3/8, Varilla de hierro de 1/4, Varilla de hierro de 1/2	V61	NALL	24/07/2023	\$120	Ladrillo de techo, Cemento, Clavos	10-ago-2023	7800
PE62	C12	22-04-2023	\$1.430	Ladrillo de pared	V62	NALL	24/07/2023	\$138	Clavos	10-ago-2023	6380

De acuerdo a los datos y encontrar relación en estos, se procederá a crear la base de datos en sql server.

1.1.2. Capa de Análisis y Visualización:

Luego de la creación de la base de datos se conecta con Power bi Desktop, para la creación de objetos visuales y así tener un mejor entendimiento del comportamiento de los datos.

PREDICCIONES		
ID_Detalle	Suma de Cantidad	Suma de Subtotal
[-] DV1		
[-] V1		
Ladrillo de pared	300	900,00
[-] DV10		
[-] V3		
Ladrillo de pared	200	600,00
[-] DV100		
[-] V43		
yeso	26	442,00
[-] DV101		
[-] V43		
Varilla de fierro de 1/4	10	120,00
[-] DV102		
[-] V43		
Varilla de fierro de 1/2	40	1.400,00
[-] DV103		
[-] V44		
Ladrillo de techo	200	800,00
[-] DV104		
[-] V44		
Cemento	40	1.200,00
[-] DV105		
[-] V45		
Cemento	20	600,00
[-] DV106		
[-] V45		
yeso	10	170,00
[-] DV107		
[-] V45		

Estos archivos serán exportados y modificados debido a que los softwares no tienen una conexión directa, estos pasos son explicados más adelante. Cabe recalcar que para la minería de datos se necesitan cantidad de datos, los cuales los datos utilizados son de 2 meses que corresponden a julio y agosto, para ver cómo va la empresa antes de la implementación del proyecto.

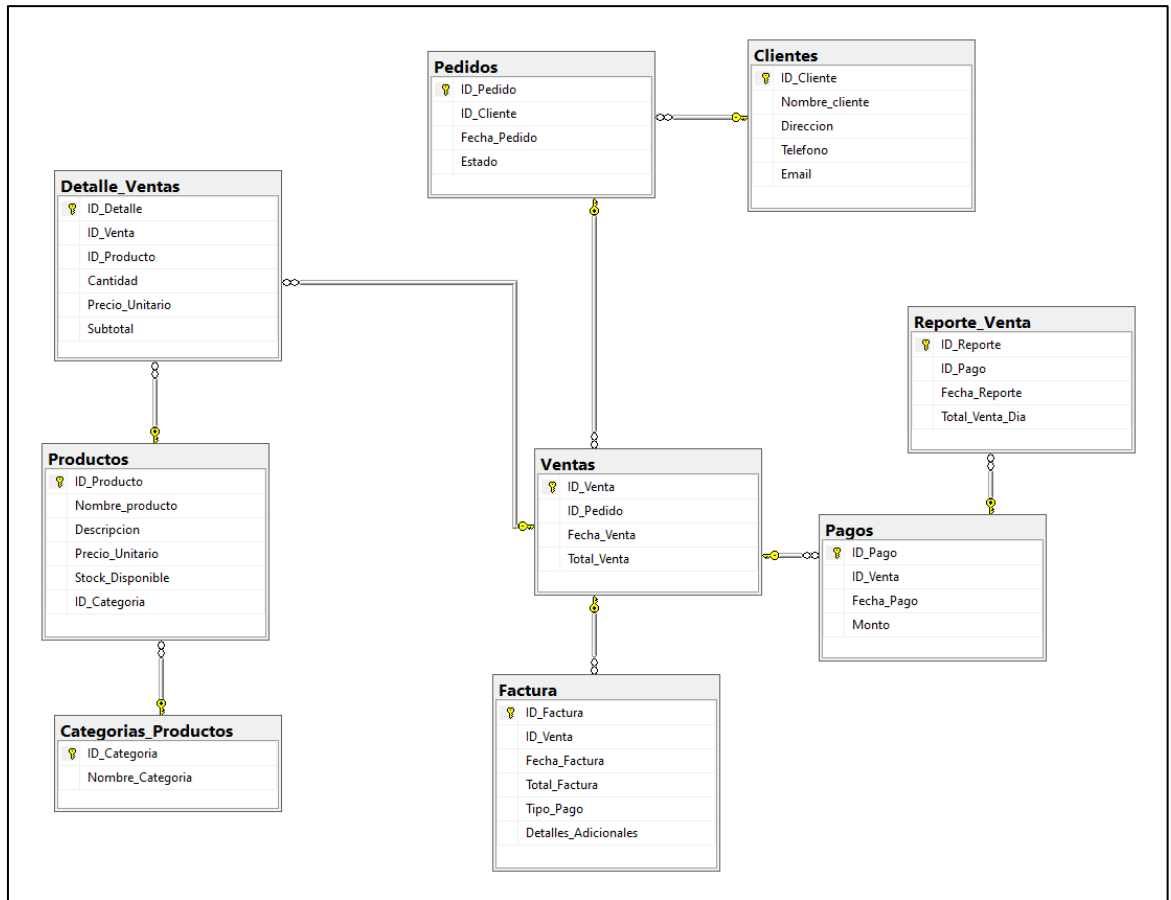
1.2. Herramientas y Tecnologías:

1.2.1. Extracción y Almacenamiento: Power BI, SQL Server.

1.2.2. Análisis y Visualización: Power BI Desktop.

1.2.3. Minería de Datos: Weka con técnicas de clasificación y asociación para descubrir predicciones y patrones dentro de los datos, con algoritmos como A priori, J48 y JCHAID.

A continuación, se muestra la base de datos general, la cual ayudará a extraer los datos necesarios para aplicar las predicciones y patrones en WEKA.



Fuente: Elaboración Propia

Los atributos específicos que serán útiles a la hora de hacer la minería de datos son:

Subtotal

Cantidad

Nombre_producto

ID_Venta

ID_Detalle

Las tablas en las que se recogen los datos necesarios para la minería de datos son:

Detalle_Ventas

Productos

Ventas

De acuerdo a los datos seleccionados desde power bi web, se exportará en archivo Excel, para después importarlo a WEKA, debido que se necesita un tipo de orden en los datos para ser procesados.

2. Describir los datos

Los datos extraídos son de la base de datos, por lo que se encuentran en sql server.

A continuación, se especifica los atributos para el análisis:

Tabla: Variable de Datos

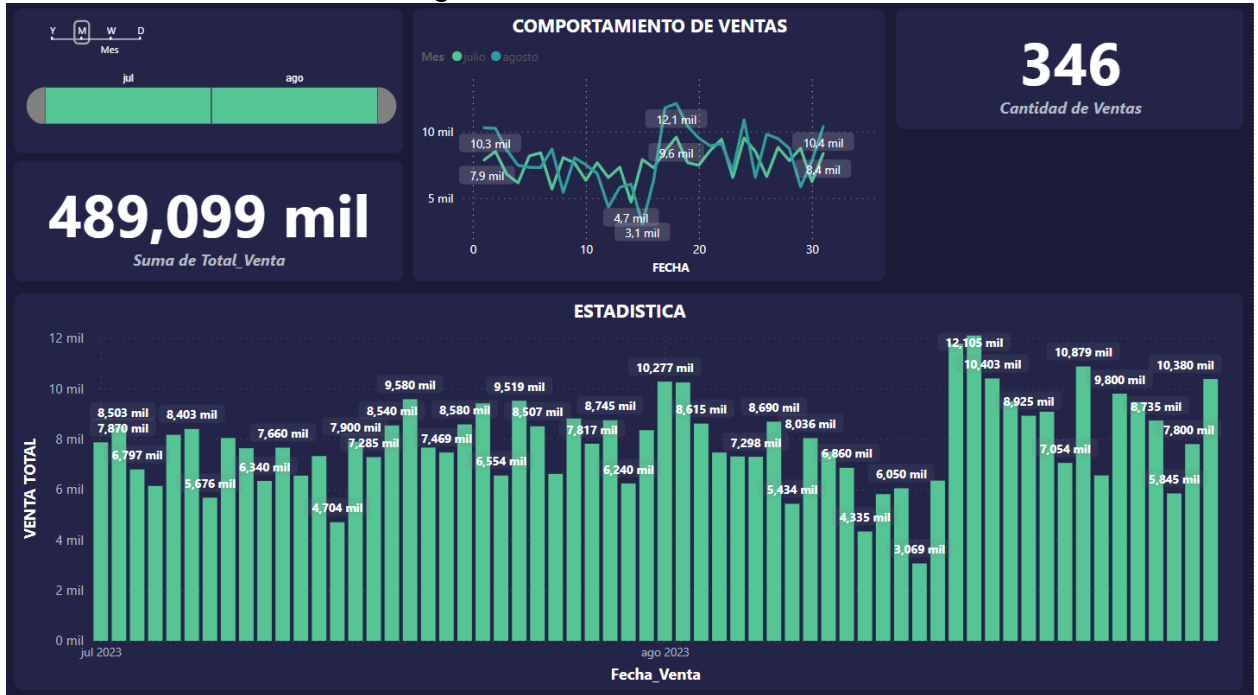
ID	Tabla referencial	Variable	Tipo de variable	Descripción de la variable
1	Detalle_Ventas	Subtotal	Decimal	El subtotal de los productos vendidos
2	Detalle_Ventas	Cantidad	Int	La cantidad de productos vendidos
3	Productos	Nombre_producto	Varchar	Indica el nombre del producto a vender
4	Categorias_Productos	Nombre_categoria	Varhcar	Categoría de los productos
5	Detalle_Ventas	ID_Venta	Varchar	Identificador único foránea de la venta
6	Detalle_Ventas	ID_Detalle	varchar	Identificador único del detalle de venta

Fuente: Elaboración Propia

3. Explorar los datos

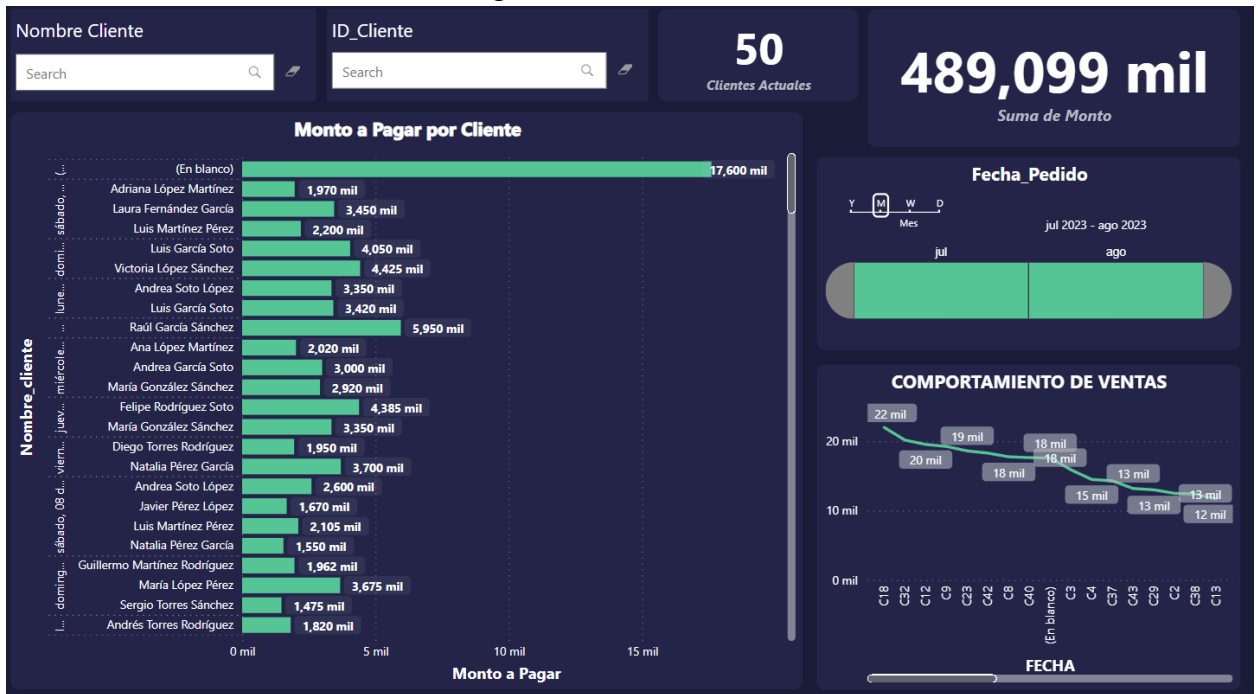
En el paso anterior se presenta una descripción de los datos, ahora se pasa a explorarlos mediante gráficos en Power bi que permitan tener una idea inicial del comportamiento y contenido de los datos, de sus características, esto probablemente ayude a obtener segmentaciones de datos que se podrán utilizar para los análisis correspondientes.

Figura: Grafico del Total de Ventas



Fuente: Elaboración Propia

Figura: Gráfico de Pedidos



Fuente: Elaboración Propia

Los datos para realizar el análisis de patrones y predicciones se mostraron en el punto 1.1 dentro de la comprensión de los datos.

4. Verificar la calidad de los datos

Después de hacer la exploración inicial de los datos se puede afirmar que estos son completos, ya que son extraídos de la base de datos y se dan un orden específico para cada análisis. Más adelante se mostrará los datos extraídos y generados para sus respectivos análisis.

C. Preparación de los datos

1. Seleccionar los datos

Como mencionado anteriormente los datos seleccionados para el análisis son los siguientes:

Datos para el análisis de predicciones:

TABLA Detalle_Ventas

- ID_Detalle
- ID_Venta
- Cantidad
- Subtotal

TABLA Productos

- Nombre_producto

Datos para el análisis de patrones:

TABLA Detalle_ventas

- Cantidad

TABLA Productos

- Nombre_producto

TABLA Ventas

- ID_Venta

2. Limpiar los datos

La base de datos con la que se cuenta para el proyecto contiene toda la información necesaria para poder cumplir los objetivos de la minería de datos, además, estos datos serán extraídos y modificados para poder ser analizados en el software WEKA, por lo que son datos limpios y por lo tanto no hay necesidad de hacer una limpieza más profunda sobre ellos.

3. Construir los datos

En la figura se muestra los datos necesarios para el análisis de patrones donde son extraídos desde power bi web.

A continuación, se mostrará los datos moldeados para importarlos a WEKA, el cual 1 significa que el producto está presente en la venta y 0 significa que no lo está, cabe recalcar que al final se guardará en formato csv.

Figura: Datos modificados para WEKA - Patrones

ID_Venta	Ladrillo de p	Arena Fina	cable eléctri	Cemento	Clavos	Focos de luz	Hormigón	Interruptore	Ladrillo de te	Lijas	Martillo	Piedra Chari	Pintura	Sierra circula	Toma corrier	Tubo pvc	Varilla de fie	Varilla de fie	Varilla de fie	yeso
1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
102	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
103	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
104	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
105	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
106	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
108	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
109	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
110	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
114	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
115	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
116	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
117	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
118	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
119	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
12	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0
120	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
122	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
125	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
126	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
127	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
128	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
129	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
13	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
130	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
131	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Fuente: Elaboración Propia

Para el tema de las predicciones también serán extraídos y armados en power bi, mostrado anteriormente.

A continuación, se muestra los datos, el cual solo se ha eliminado la columna de is_detalle, ya que solo sirve para ordenar las ventas, también se guardará el archivo en formato csv.

Figura: Datos modificados para WEKA - Predicciones

A	B	C	D	E
ID_venta	Nombre_producto	Cantidad	Subtotal	
1	Ladrillo de pared	300	900	
3	Ladrillo de pared	200	600	
43	yeso	26	442	
43	Varilla de fierro de 1/4	10	120	
43	Varilla de fierro de 1/2	40	1400	
44	Ladrillo de techo	200	800	
44	Cemento	40	1200	
45	Cemento	20	600	
45	yeso	10	170	
45	Varilla de fierro de 1/2	30	1050	
46	Clavos	3	15	
46	Sierra circular	1	120	
3	Ladrillo de techo	100	400	
47	Martillo	1	25	
48	Arena Fina	1	450	
48	Hormigón	2	900	
48	Piedra Chancada	1	500	
48	yeso	30	510	
49	Ladrillo de pared	450	1350	
49	Ladrillo de techo	200	800	
49	Cemento	35	1050	
50	Ladrillo de pared	250	750	
50	Ladrillo de techo	100	400	
3	Cemento	40	1200	
50	Cemento	40	1200	
51	Ladrillo de pared	250	750	
51	Ladrillo de techo	100	400	
51	Cemento	25	750	
52	Toma corrientes	30	210	
53	Pintura	4	100	
54	Tubo pvc	4	64	
55	Ladrillo de techo	400	1600	

Fuente:

Elaboración Propia

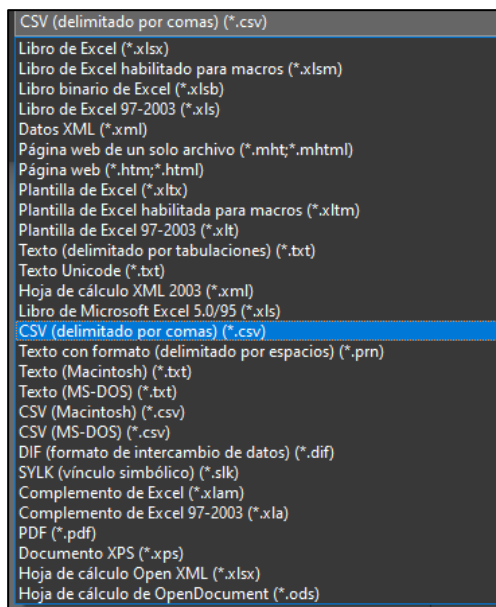
D. Modelamiento

Las técnicas utilizadas son de clasificación, junto con el algoritmo JCHAID y J48 que ayudarán al tema de las predicciones, a su vez se usará la técnica de asociación, junto con el algoritmo a priori para el análisis de patrones. La herramienta seleccionada es WEKA versión 3.8.6.

A continuación, se da una breve descripción del paso a paso para el cargue de los archivos en el software WEKA:

PASO 1: De acuerdo a las figuras anteriores, se guardarán en formato csv, como se muestra a continuación:

Figura: Formato de archivos

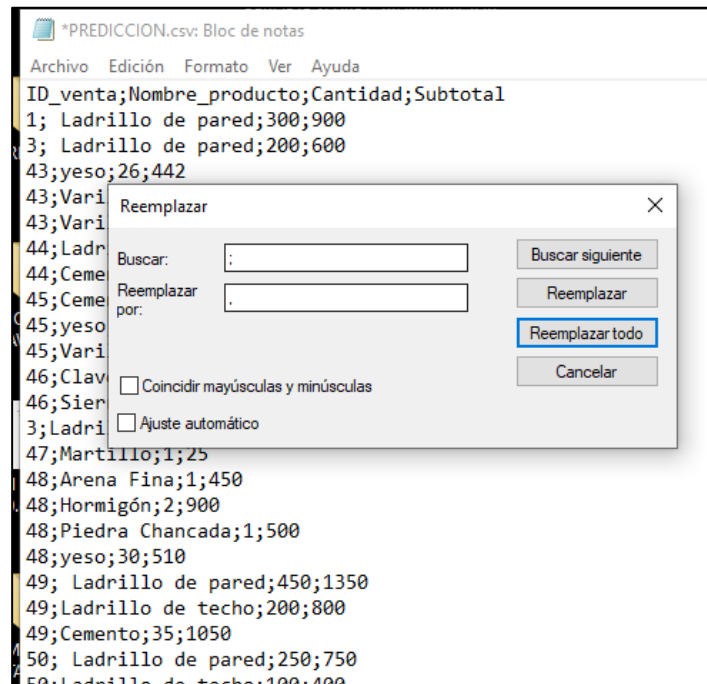


Fuente: Elaboración Propia

PASO 2: Una vez guardado lo abrimos con bloc de notas, donde se dará en el menú “Edición” y luego en “reemplazar”. En buscar se pondrá “;” y será reemplazado por “,”, ya que así permitirá una mejor lectura en el software, esto aplica para ambos archivos, tanto como para patrones y predicciones.

A continuación, se mostrará la figura del paso 2:

Figura: Reemplazar delimitador



Fuente: Elaboración Propia

PASO 3: Una vez se ingresa al aplicativo WEKA, aparece la ventana selectora de interfaces la cual permite seleccionar la interfaz con la cual se trabajará.

Figura: Interfaz WEKA

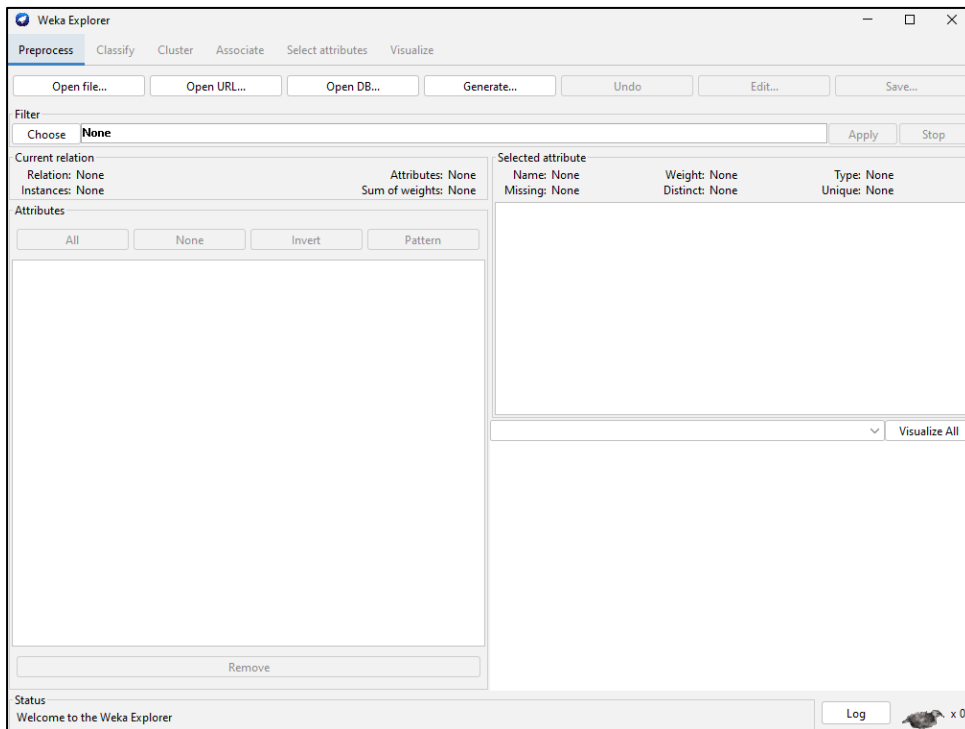


Fuente: Elaboración Propia

- Explorer: realiza operaciones sobre un solo conjunto de datos. Esta es la interfaz con la cual se trabajará.

La interfaz Explorer tiene seis pestañas: Preprocess, Classify, Cluster, Associate, Select atributes, Visualize. Primero se utilizará Preprocess.

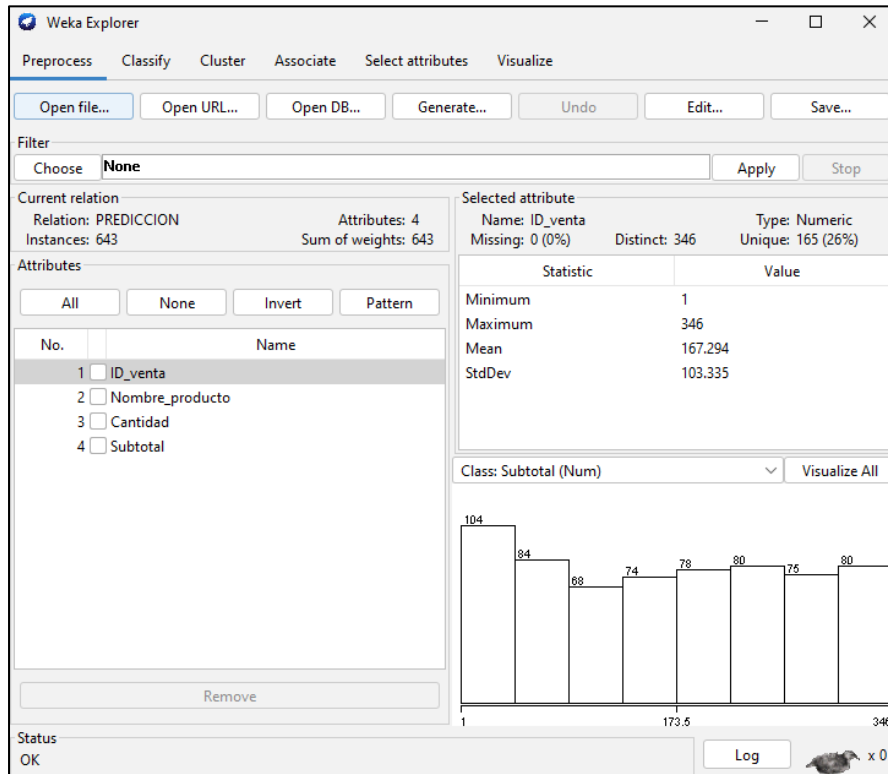
Figura: Interfaz de Explorer - WEKA



Fuente: Elaboración Propia

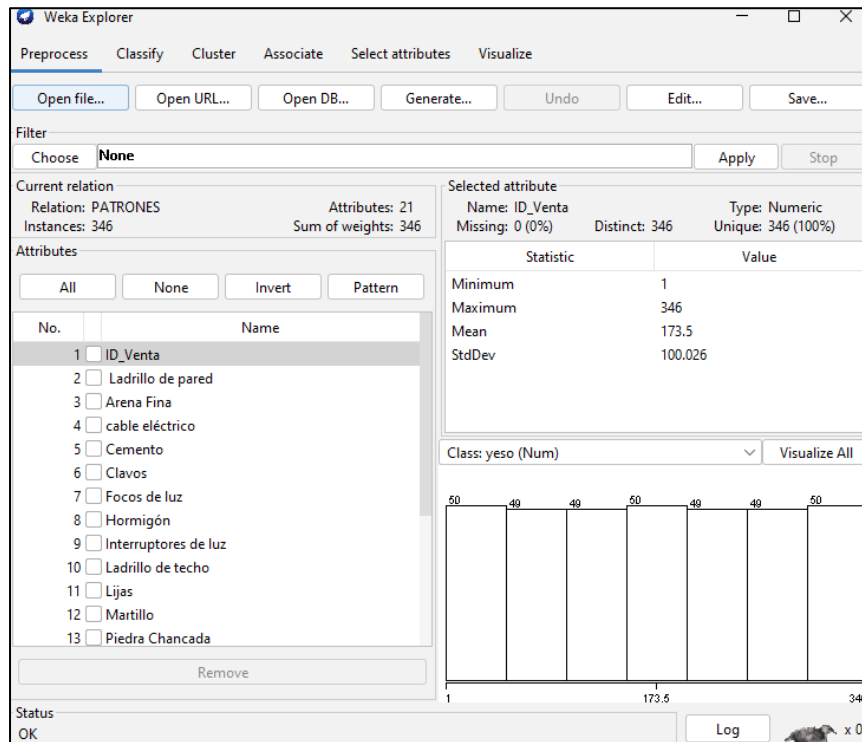
PASO 4: En la pestaña Preprocess se ingresa a la opción Open file para cargar el archivo a analizar, que en este caso sería el archivo csv.

Figura: Archivo para predicciones



Fuente: Elaboración Propia

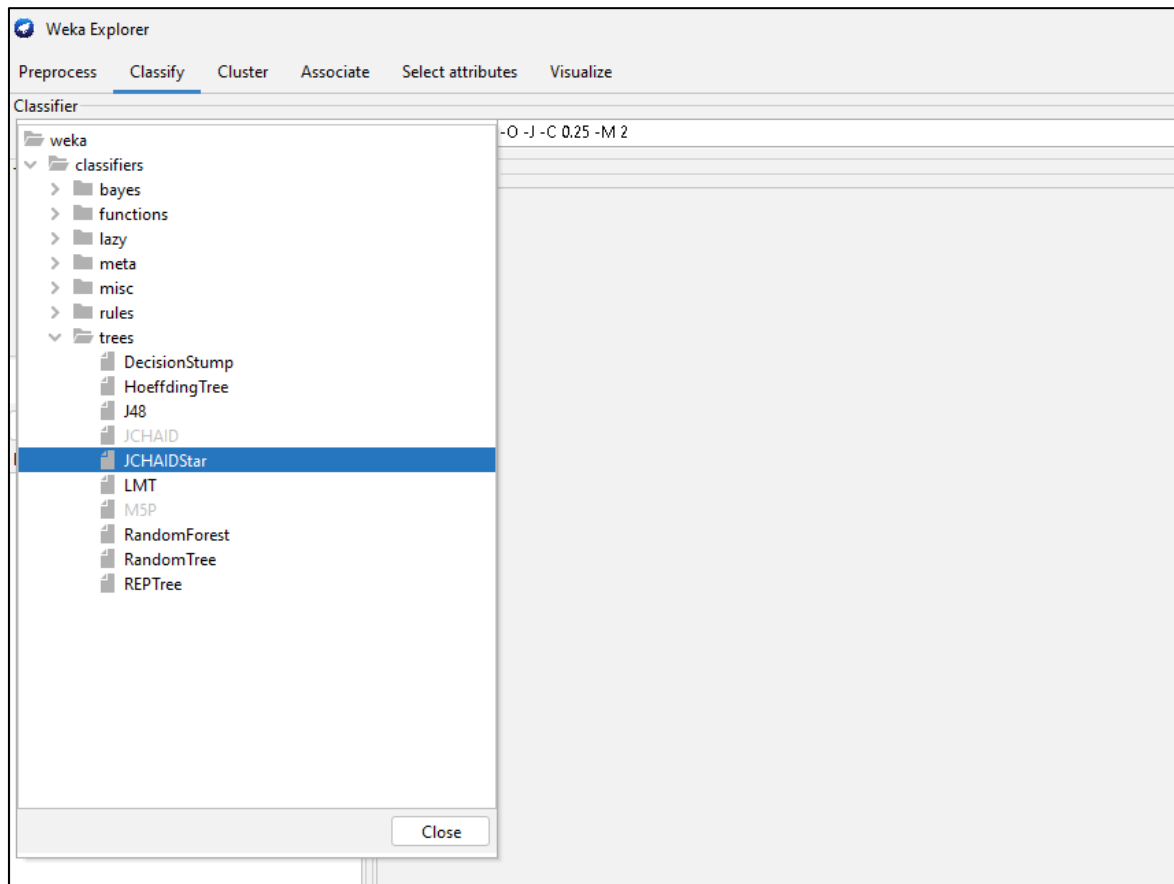
Figura: Archivo de patrones



Fuente: Elaboración Propia

PASO 5: Se comenzará con la predicción, en este caso se seleccionará la pestaña “classify”, donde se seleccionará “trees” que viene a ser el árbol de decisión, junto con el algoritmo para el análisis, comenzaremos con JCHAIDStar.

Figura: Selección de Algoritmo - Predicción

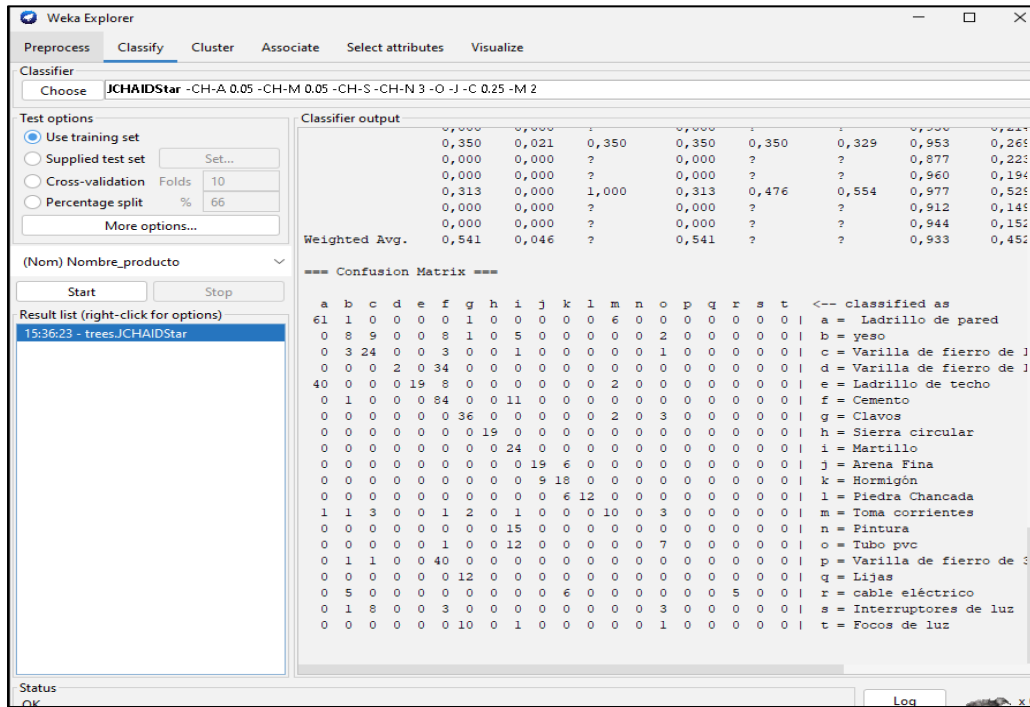


Fuente: Elaboración Propia

Una vez seleccionado la técnica y algoritmo, se dará en la opción “use training set”, seguidamente se seleccionará el nombre de la clase Nombre_producto, ya que queremos saber que productos son más probables a ser vendidos.

Se inicia dando en “start” y el programa comenzará a mostrar los resultados del análisis.

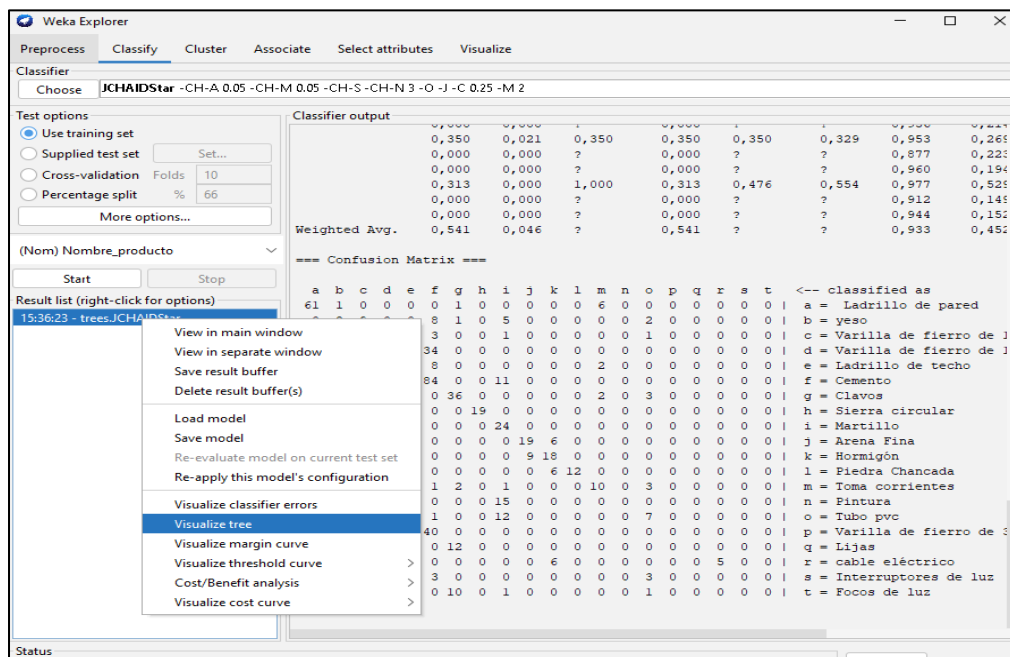
Figura: Análisis con JCHAIDStar



Fuente: Elaboración Propia

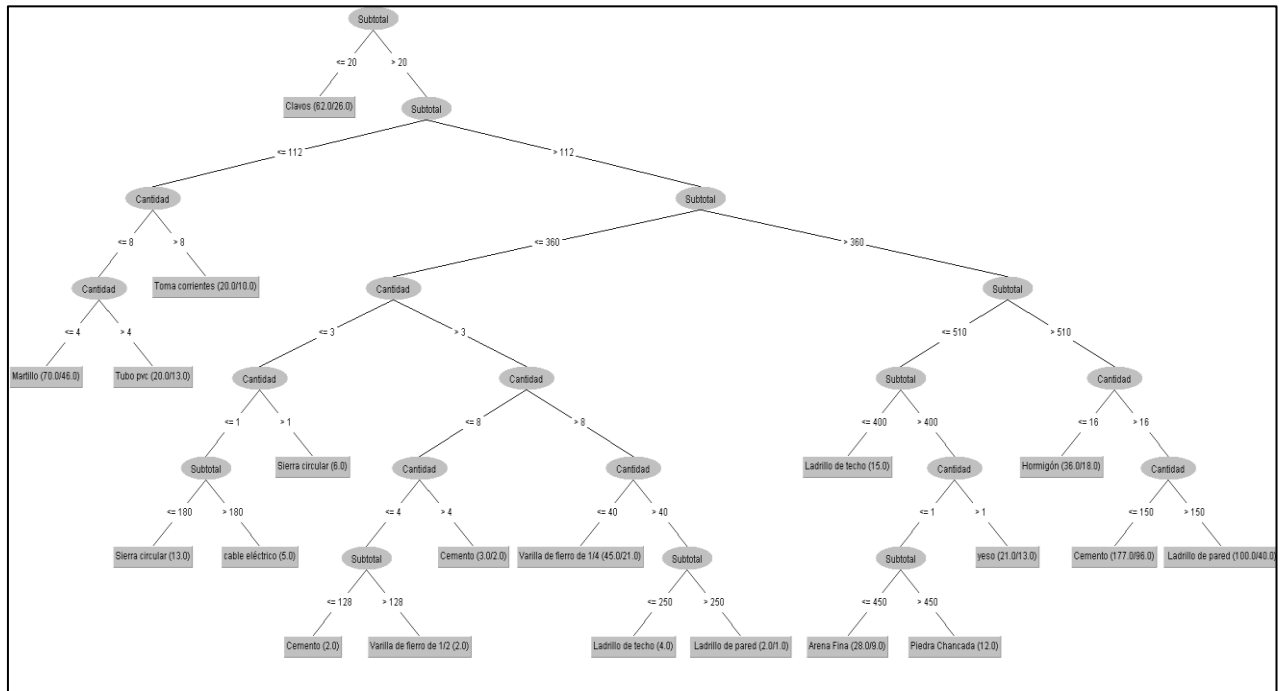
En el resultado obtenido dando clic derecho y luego en “visualize tree”, mostrará en árbol de decisión del análisis.

Figura: Opción “visualize tree”



Fuente: Elaboración Propia

Figura: Árbol de decisión JCHAID



Fuente: Elaboración Propia

Interpretación de los resultados

El árbol comienza con una condición en la parte superior, en este caso, "Subtotal ≤ 20 ", que significa que si el valor de "Subtotal" es menor o igual a 20, se clasifica como "Clavos" con una probabilidad de 62.0 y una confianza de 26.0.

Si el valor de "Subtotal" es mayor que 20, el árbol se ramifica en dos caminos:

1. Si "Subtotal" es menor o igual a 112 y "Cantidad" es menor o igual a 8, entonces se verifica una nueva condición: "Cantidad ≤ 4 ". Si esta condición es verdadera, se clasifica como "Martillo" con una probabilidad de 70.0 y una confianza de 46.0. Si "Cantidad > 4 ", se clasifica como "Tubo pvc" con una probabilidad de 20.0 y una confianza de 13.0.
2. Si "Subtotal" es mayor que 112, el árbol sigue ramificándose en subcategorías basadas en las condiciones "Cantidad" y "Subtotal". Por ejemplo, si "Cantidad" es menor o igual a 3 y "Subtotal" es menor o igual a 180, se clasifica como "Sierra circular" con una probabilidad de 13.0. El árbol continúa dividiéndose en categorías más específicas a medida que se cumplen ciertas condiciones.

Interpretación de los resultados

El árbol comienza con una raíz que establece una condición inicial: "Cantidad \leq 8". Esto significa que, si la cantidad es menor o igual a 8, se sigue el primer camino; de lo contrario, se sigue el segundo camino.

Si la cantidad es menor o igual a 8, se evalúan más condiciones. El árbol se divide en función de los valores de "Subtotal" y "Cantidad". Cada nodo representa una decisión basada en las condiciones que se evalúan en ese nodo.

Por ejemplo, si la cantidad es mayor que 8, se sigue el segundo conjunto de condiciones. El árbol se ramifica aún más en función de los valores de "Subtotal" y "Cantidad".

El proceso continúa, dividiendo el conjunto de datos en subconjuntos más pequeños a medida que se desciende por el árbol, y se asigna una categoría a cada subconjunto en función de las condiciones evaluadas en cada nodo.

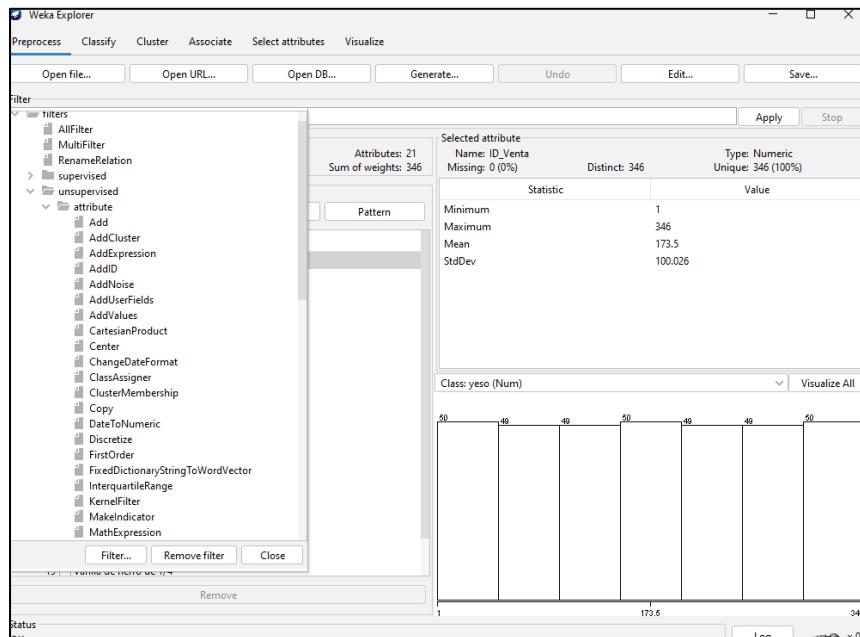
Las probabilidades y números que acompañan a las categorías (por ejemplo, "Clavos (30.0/3.0)") pueden indicar la probabilidad de que un elemento se clasifique en esa categoría y la confianza en esa predicción.

Siguiendo con el paso 5, ahora se analizará para encontrar patrones.

PASO 6: Una vez cargado el archivo los datos deberán ser transformados a datos nominales, ya que de esta manera el algoritmo lo acepta para el análisis.

PASO 7: Se selecciona "choose", luego en "unsupervised", que son datos no supervisados, lo cual es especial para encontrar patrones, seguidamente se dará en "apply" para aplicar el filtro.

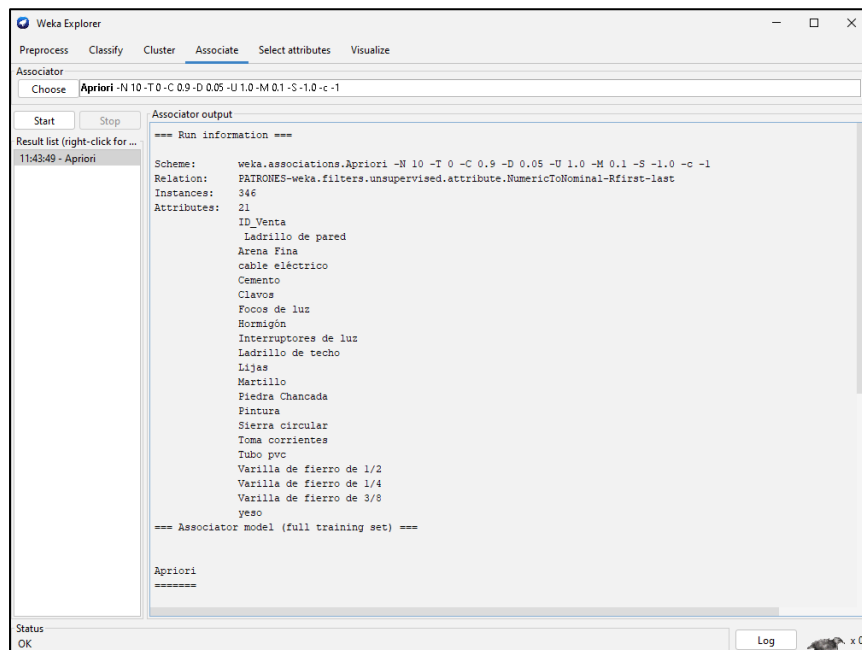
Figura: Cambio de datos a nominal



Fuente: Elaboración Propia

PASO 8: Se selecciona "Associate" en las pestañas, que sirve para el análisis de patrones, por defecto ya estará el algoritmo A priori y le daremos en start.

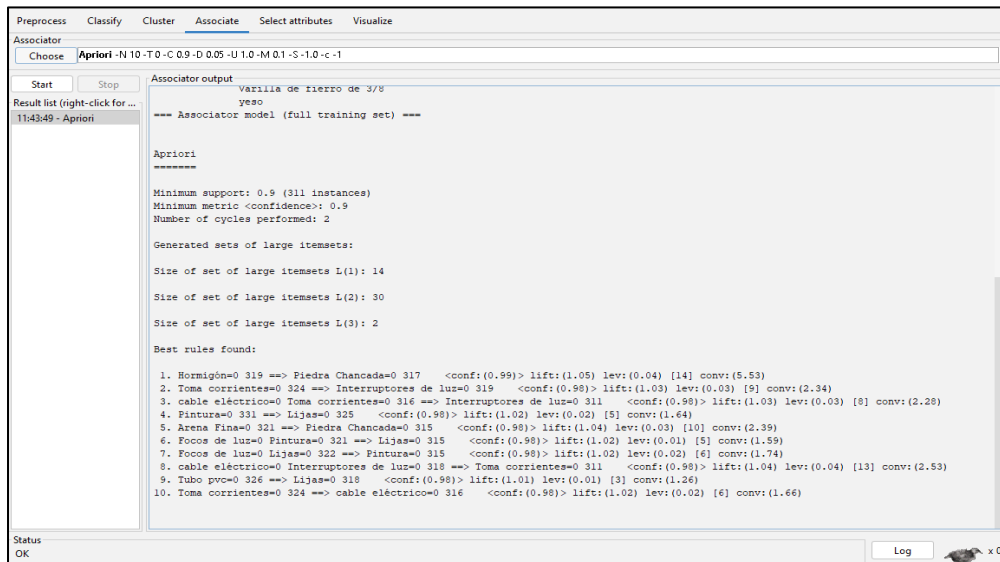
Figura: Análisis con algoritmo Apriori



Fuente: Elaboración Propia

Los resultados obtenidos se verán ahí mismo, lo cual nos genera los 10 mejores patrones que vienen por defecto.

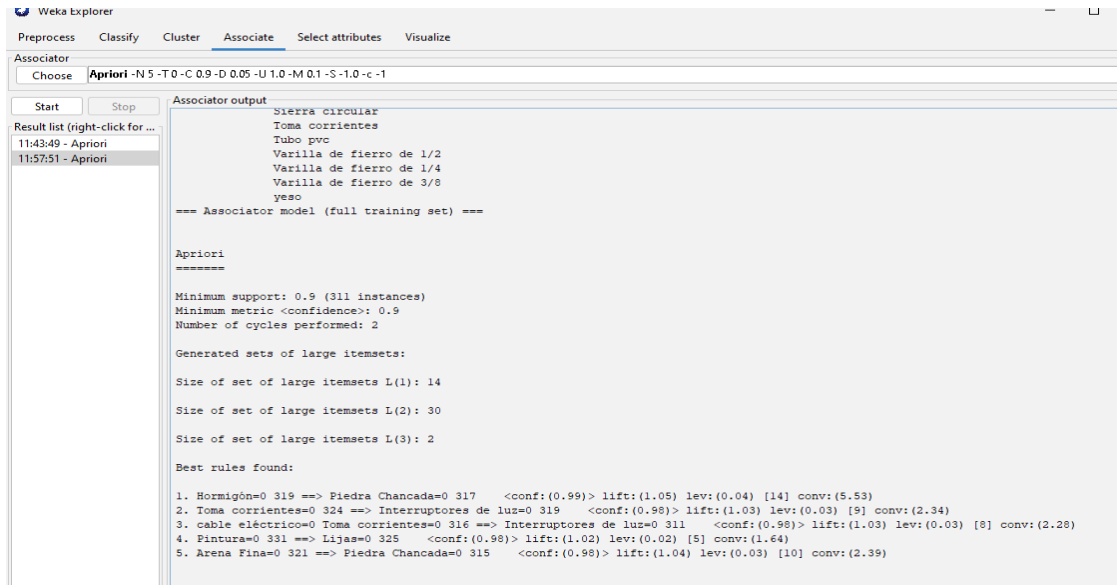
Figura: Resultados del análisis de patrones



Fuente: Elaboración Propia

Se puede cambiar para que evalúe y muestre una cierta cantidad de reglas, en este caso se seleccionó para que mostrara las 5 mejores reglas, la cual para cambiarlo se seleccionó en el nombre “Apriori” y luego en “numRules” se cambia el número.

Figura: Aplicando para 5 reglas de patrones



Fuente: Elaboración Propia

Interpretación de los resultados

Para entender mejor los resultados, se sigue el concepto de la figura anterior.

1. **Hormigón=0 319 ==> Piedra Chancada=0 317:** Esta regla indica que cuando el producto "Hormigón" no está presente (0), hay una alta probabilidad (confianza del 99%) de que "Piedra Chancada" tampoco esté presente (0). El valor de "lift" es mayor que 1 (1.05), lo que sugiere una asociación positiva entre la ausencia de ambos productos.
2. **Toma corrientes=0 324 ==> Interruptores de luz=0 319:** Cuando "Toma corrientes" no está presente, hay una alta probabilidad (confianza del 98%) de que "Interruptores de luz" tampoco esté presente. El valor de "lift" es mayor que 1 (1.03), lo que indica una asociación positiva entre la ausencia de ambos productos.

Así para cada regla generada, donde en este por los datos se ve mas la ausencia, lo que también indica que estos productos tienen mucha relación si no están presentes, ya que es muy probable que si se compra también puede estar presente el otro.

E. Evaluación

1. Evaluar los resultados

La aplicación de técnicas de minería de datos, algoritmos y la realización de tareas como la clasificación y asociación proporcionan un punto de partida inicial para entender cómo se aplica la minería de datos en situaciones del mundo real. Esto es

especialmente relevante para abordar problemas actuales que requieren este tipo de enfoque para una comprensión más profunda de los datos.

En la fase de Modelamiento, se encuentran disponibles tanto técnicas supervisadas, que se utilizan para el análisis predictivo, como sus respectivas tareas de clasificación. Además, se encuentran técnicas no supervisadas que se emplean en el análisis de patrones y la tarea de asociación. Dentro de estas técnicas supervisadas y no supervisadas, existen una variedad de tareas y métodos adicionales que amplían las posibilidades de investigación y análisis. Esto ofrece a los estudiantes futuros un amplio conjunto de enfoques que pueden explorar al adentrarse en el campo de la minería de datos.

2. Revisar el proceso

Para ejecutar técnicas en WEKA es importante ver el tipo de datos de los archivos, ya que a veces al querer aplicar cierto algoritmo no permitirá realizarlo ya que solo aceptan ciertos tipos de datos, lo que se tendrá que transformar en el mismo WEKA.

3. Planear los pasos siguientes

Refinar o sustituir los modelos: La técnica de minería de datos aplicados se considera apropiada para el tipo de datos que se produce en la predicción y patrones de los productos. Sin embargo, a estos datos se le pueden aplicar otros tipos de técnicas o algoritmos.

F. Despliegue

1. Desarrollar plan de implementación: Para desarrollar un plan de implementación, es importante primero resumir los resultados. También es conveniente que por cada descubrimiento nuevo que se haya obtenido, se cree un plan para dar a conocer esos resultados.

Utilizando los algoritmos para predicción se ve que, en el primer resultado (algoritmo JCHAID), se presenta una estructura jerárquica que categoriza productos en función de dos criterios principales: "Subtotal" y "Cantidad".

En el segundo resultado (algoritmo J48), se sigue una estructura similar, pero esta vez se enfoca principalmente en la "Cantidad" y luego en el "Subtotal".

Ambos resultados utilizan una estructura jerárquica para categorizar productos en función de múltiples criterios, como "Subtotal" y "Cantidad."

Proporcionan puntuaciones de soporte/confianza para cada categoría, lo que indica la probabilidad de que un producto pertenezca a una categoría específica dadas las condiciones establecidas.

En resumen, ambos resultados utilizan enfoques jerárquicos para clasificar productos en función de múltiples criterios, pero difieren en la prioridad dada a estos criterios y en las categorías específicas creadas.

Para los patrones ya no es necesario resumir, ya que sus resultados son breves y entendibles, también fueron explicados anteriormente, donde se ven que ciertos productos siguen patrones debido a su ausencia, lo que genera una idea para planear nuevas estrategias derivadas a estas.

2. Producir el informe final:

Para este caso, el informe constituye el presente documento que servirá como apoyo para futuros trabajos académicos de minería de datos.

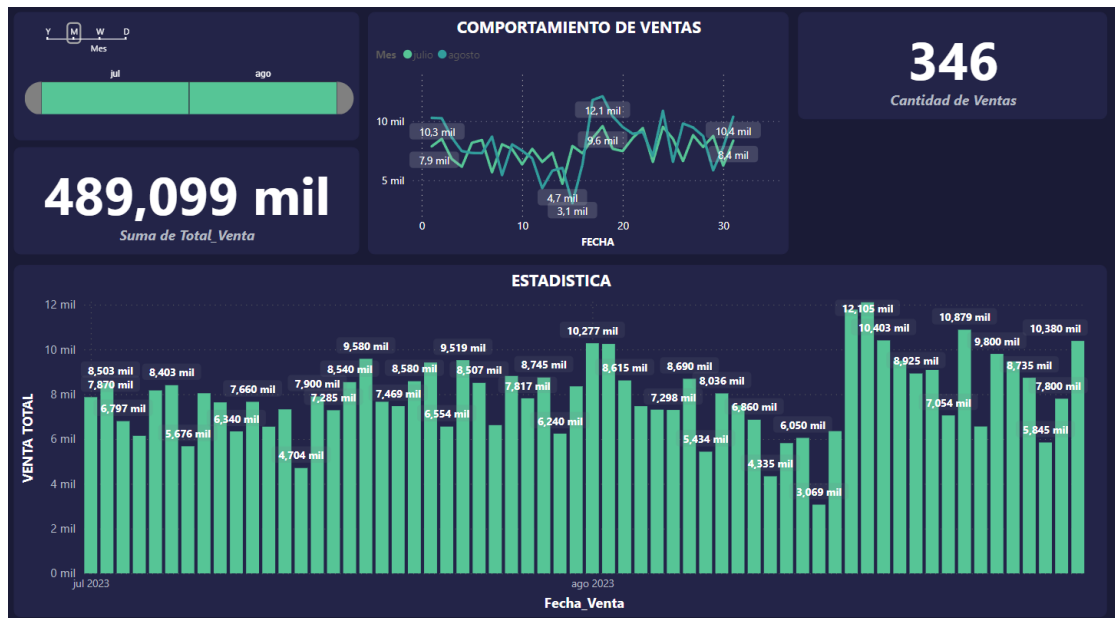
3. Revisar el proyecto

Las principales dificultades o demoras de tiempo se presentaron en la integración los archivos csv y la conversión de los datos para integrarlos a WEKA. La selección de las técnicas de minería de datos para aplicar deberá revisarse de forma que se tenga en cuenta el tipo de datos a los cuales se aplicarán (numéricos, categóricos), pues para cada uno de ellos hay técnicas.

ANEXO 12: COMPARACIÓN DE RESULTADOS DENTRO DE POWER BI

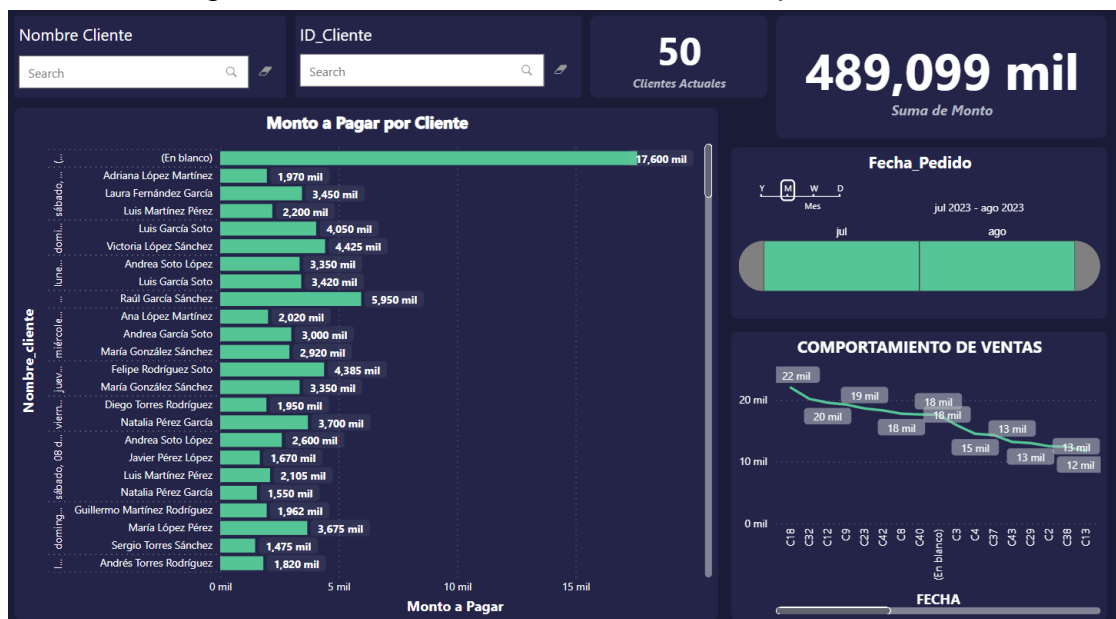
Los datos usados durante el desarrollo del proyecto constan de 2 meses. Luego para ver los resultados después de la implementación, fueron los datos del post test, de acuerdo al indicador de crecimiento de ventas, que consta de un mes.

Figura: Ventas Totales antes de la implementación



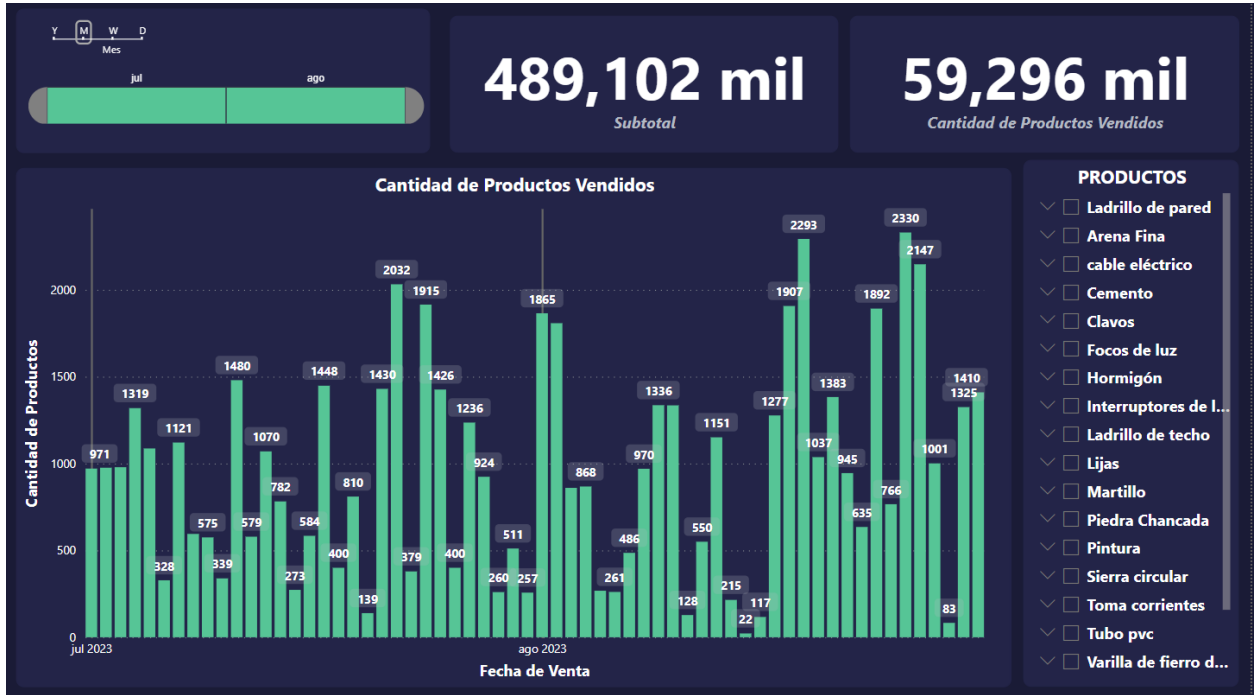
Fuente: Elaboración Propia

Figura: Pedidos Realizados antes de la implementación



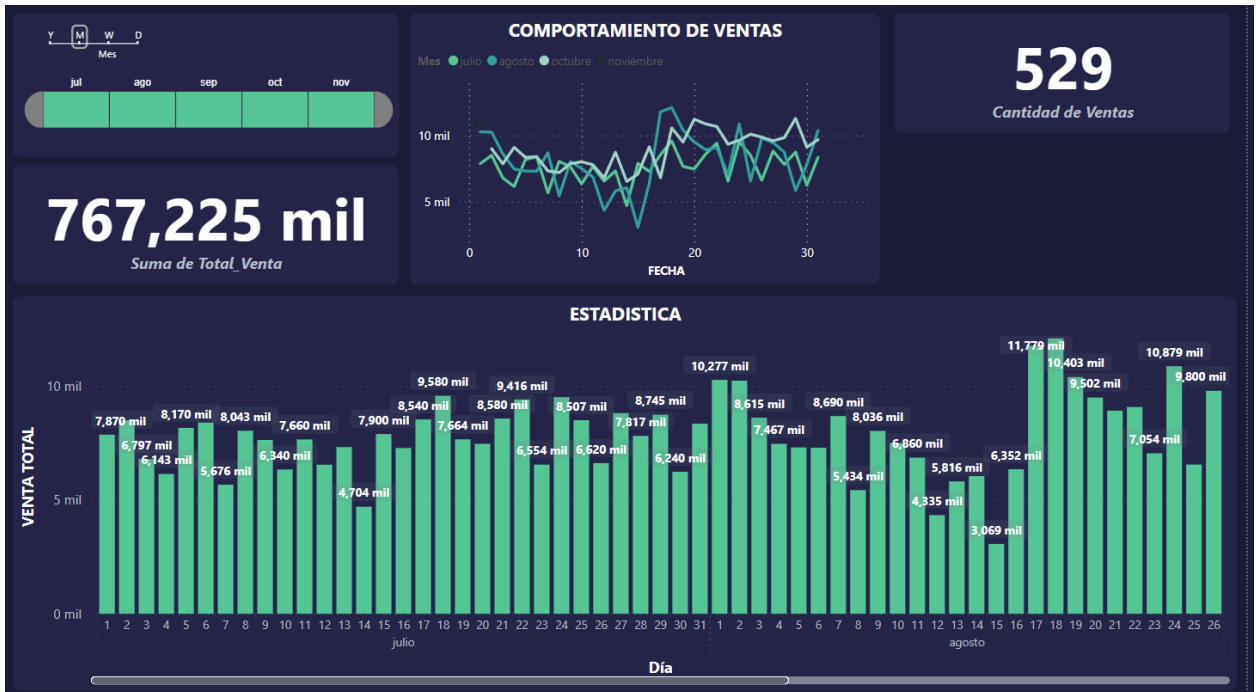
Fuente: Elaboración Propia

Figura: Detalle de productos vendidos antes de la implementación



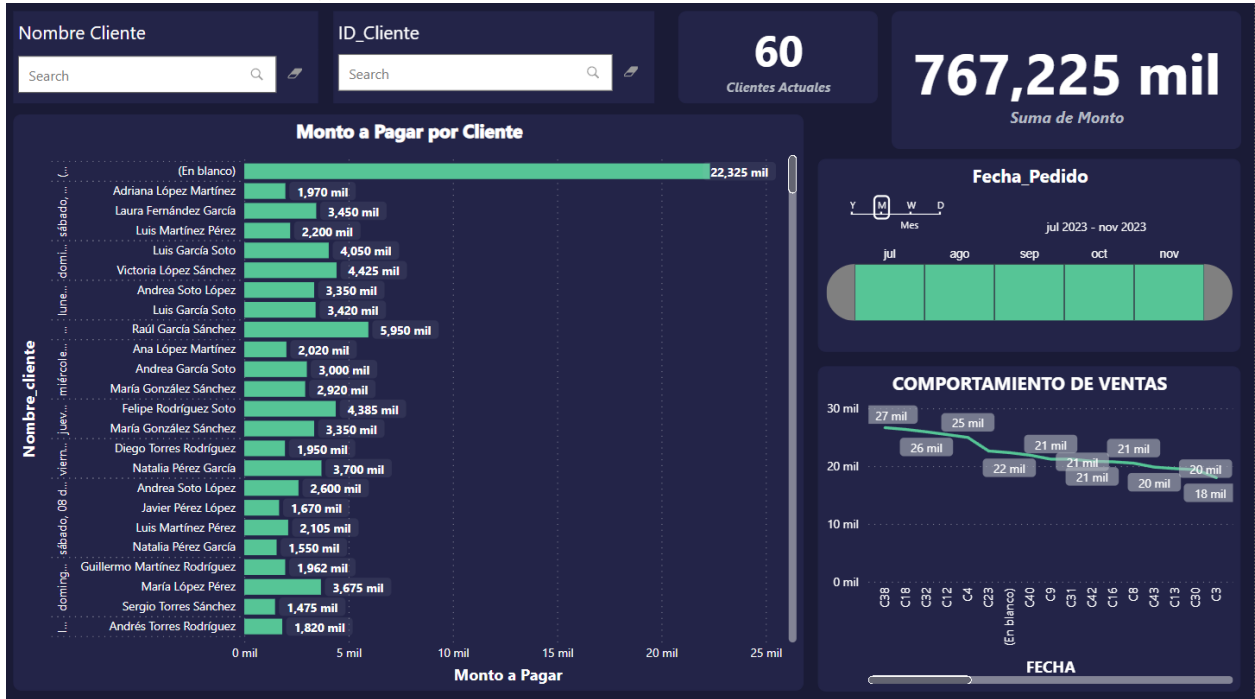
Fuente: Elaboración Propia

Figura: Ventas Totales después de la implementación



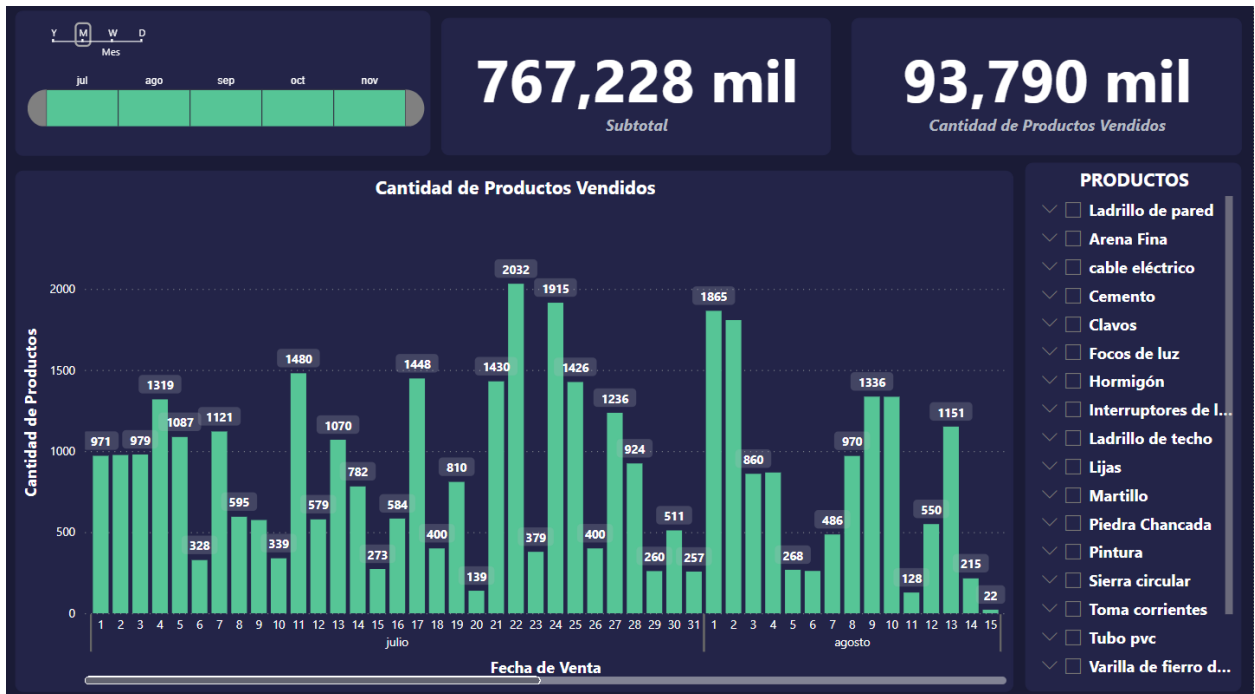
Fuente: Elaboración Propia

Figura: Pedidos Realizados antes de la implementación



Fuente: Elaboración Propia

Figura: Detalle de productos vendidos antes de la implementación



Fuente: Elaboración Propia



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, NECOCHEA CHAMORRO JORGE ISAAC, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "IMPLEMENTACIÓN DE BUSINESS INTELLIGENCE UTILIZANDO MINERÍA DE DATOS PARA LA TOMA DE DECISIONES EN LAS VENTAS", cuyo autor es TANTALEAN SALAZAR BILLY BRANDON, constato que la investigación tiene un índice de similitud de 15.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 27 de Noviembre del 2023

Apellidos y Nombres del Asesor:	Firma
NECOCHEA CHAMORRO JORGE ISAAC DNI: 18167347 ORCID: 0000-0002-3290-8975	Firmado electrónicamente por: JNECOCHEA el 28- 11-2023 09:21:43

Código documento Trilce: TRI - 0667055