



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Sistema de pronóstico de ventas de herramientas hidráulicas y
electromecánicas

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero de Sistemas

AUTORES:

Hijar Saman, Juan Carlos (orcid.org/0000-0002-5542-2118)

Ortiz Salazar, Gianfranco Porfirio (orcid.org/0000-0002-2185-2283)

ASESOR:

Dr. Alfaro Paredes, Emigdio Antonio (orcid.org/0000-0002-0309-9195)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2019

Dedicatoria

Se lo dedicamos a Dios por habernos permitido llegar hasta las últimas instancias en las mejores condiciones. También dedicamos esta investigación a nuestras familias, por sus consejos, enseñanzas y tolerancia.

Agradecimiento

A mi padre Luis, quien siempre me brindó su apoyo en todos los aspectos, su confianza y sobre todo siempre tuvo una palabra para animarme y seguir en este largo camino de aprendizaje, realización personal y la verdadera misión en este paseo terrenal.

Gianfranco Porfirio Ortiz Salazar

A mis padres Juan y Frida quienes me enseñaron a valorar los resultados de un gran esfuerzo, a conocer el precio de tener una gota de sudor en la frente, por ser mis amigos y darme palabras de aliento, por ser ejemplos y por ocupar un lugar muy especial en mi corazón.

Juan Carlos Híjar Samán

Índice de contenidos

Carátula.....	i
Dedicatoria	ii
Agradecimiento	iii
Índice de contenidos	iv
Índice de tablas.....	v
Índice de figuras.....	vi
Índice de anexos.....	vii
Índice de abreviaturas.....	viii
Resumen.....	ix
Abstract	x
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO	14
III. METODOLOGÍA	21
3.1 Tipo y diseño de investigación	22
3.2 Variables y su operacionalización.....	23
3.3 Población, muestra y muestreo	23
3.3.1 Población.....	23
3.3.2 Muestra y Muestreo.....	23
3.4 Técnicas e instrumentos de recolección de datos	23
3.5 Procedimientos	23
3.6 Método de análisis de datos	24
3.7 Aspectos éticos.....	25
IV. RESULTADOS	26
4.1 Prueba de la hipótesis específica 1	29
4.2 Prueba de la hipótesis específica 2.....	29
4.4 Prueba de la hipótesis específica 3.....	30
4.4 Prueba de la hipótesis específica 4.....	30
4.5 Prueba de la hipótesis específica 5.....	31
4.6 Prueba de la hipótesis específica 6.....	32
4.7 Prueba de la hipótesis específica 7.....	32
4.8 Resumen de las hipótesis	33
V. DISCUSIÓN	35
VI. CONCLUSIONES.....	39
VII. RECOMENDACIONES.....	42
REFERENCIAS	44

Índice de tablas

Tabla 1: ANOVA del modelo de regresión lineal múltiple de las ventas	27
Tabla 2: Resumen del modelo de regresión lineal múltiple de las ventas	27
Tabla 3: Estadísticos descriptivos de las variables del modelo de regresión lineal múltiple de las ventas	28
Tabla 4: Coeficiente del modelo de regresión lineal múltiple de las ventas	28
Tabla 5: Resumen de los resultados de las comprobaciones de las hipótesis	33
Tabla 6: Matriz de consistencia	50
Tabla 7: Comprensión de los datos - Tabla transaccion_soles	54
Tabla 8: Comprensión de los datos - Tabla cantidad_almacen	54
Tabla 9: Comprensión de los datos - Tabla cantidad_proveedor	54
Tabla 10: Comprensión de los datos - Tabla monto_por_factura.....	55
Tabla 11: Comprensión de los datos - Tabla numero_asesores	55
Tabla 12: Comprensión de los datos - Tabla movimiento_por_articulo	55
Tabla 13: Comprensión de los datos - Tabla transaccion_por_cliente	56
Tabla 14: Comprensión del desarrollo – Inventario de recurso de hardware	63
Tabla 15: Comprensión del desarrollo – Inventario de recurso de software	64

Índice de figuras

Figura 1: Parámetros de conexión a la base de datos de las ventas.....	52
Figura 2: Conexión exitosa a las tablas de la base de datos de las ventas	52
Figura 3: Comprensión de los datos – Diagrama base de datos relacional	53
Figura 4: Modelo de tratamiento de datos del DWH.....	56
Figura 5: Inserción de información calculada	57
Figura 6: Muestreo de datos para inyección SQL.....	57
Figura 7: Conversión de tipos de datos de las tablas de la base de datos de las ventas	58
Figura 8: Orientación y creación de servidor de base de datos.....	58
Figura 9: DataStage aplicado a las ventas	59
Figura 10: Cubo OLAP del modelo de ventas post ejecución de DataStage	59
Figura 11: Definiciones de las dimensiones de las ventas.....	60
Figura 12: Diagrama de dispersión de dimensiones de las ventas	60
Figura 13: Consulta integradora en la base de datos de las dimensiones independientes	61
Figura 14: Consulta integradora en la base de datos de la dimensión dependiente	62
Figura 15: Comprensión del desarrollo – Plugins Analytics y Power BI	64
Figura 16: Comprensión del desarrollo – Arquitectura del modelo de proyección de ventas.....	66
Figura 17: Comprensión del desarrollo – Diagrama de pseudocódigo del algoritmo de proyección de ventas.....	67

Índice de anexos

Matriz de consistencia	50
Conexión y Desarrollo de la ETL	52
Desarrollo de la aplicación.....	63
Ficha de observación	68

Índice de abreviaturas

Abreviatura	Significado	Página
DM	Data Mining (Minería de datos)	4
MYPE	Micro y pequeña empresa	4
PYME	Pequeña y mediana empresa	4
PEN	Peruvian Sol (Sol peruano)	8
KPI	Key Performance Indicator (Indicadores claves de desempeño)	14
SQL	Structured Query Language (Lenguaje estructurado en base a consultas)	17
ERP	Enterprise Resources Planning (Planificación de recursos empresariales)	17
OLAP	On Line Analytical Processing (Procesamiento analítico en línea)	19
IA	Inteligencia Artificial	19
ANOVA	Analysis of the Variance (Análisis de varianza)	28

Resumen

El problema de la investigación fue: ¿Cuáles fueron los porcentajes de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes? El objetivo de la investigación fue determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes. El tipo de investigación fue explicativo con enfoque transversal y el tipo de diseño fue correlacional-causal ya que se analizaron los datos de las ventas totales durante el periodo de un año. La información fue obtenida de la empresa Probotix S.A.C. (empresa con nombre ficticio para reserva de información). Para el desarrollo de la investigación se usaron procesos de minería de datos y la técnica estadística usada para el análisis de datos fue la regresión lineal múltiple.

Los resultados obtenidos en la presente investigación demostraron que con la técnica de regresión lineal múltiple se obtuvo una explicación de la variabilidad de las ventas del 88.7%. Con los resultados obtenidos en la investigación se llegó a concluir que cuatro de las siete dimensiones del modelo de proyección de ventas fueron significativas para la estimación de las ventas, las cuales fueron: transacción por cliente con 45.1%, movimiento por artículo con 19.0%, cantidad de proveedores con 29.7% y cantidad de almacenes con 18.3%, mientras que las tres dimensiones restantes no fueron significativas para la estimación de las ventas, las cuales fueron: transacciones en soles, monto por facturas y número de asesores. Finalmente, se recomendó combinar la regresión lineal con alguna otra técnica estadística como por ejemplo análisis de series temporales con la finalidad que el modelo de predicción tenga mayor precisión.

Palabras clave: Minería de datos, porcentaje de variabilidad, regresión lineal múltiple, series temporales, modelo de predicción.

Abstract

The research problem was: What were the percentages of the variability of the sales of hydraulic and electromechanical tools with the variable's transaction per customer, movement per item, transactions in soles, amount per invoice, number of advisers, number of suppliers, and number of stores? The purpose of the study was to determine the percentage of the variability of the sales of hydraulic and electromechanical tools with the variables: transaction per customer, movement per item, transactions in soles, amount per invoice, number of advisers, number of suppliers, and number of warehouses. The type of research was explanatory with a cross-sectional approach and the type of design was correlational-causal, due to that the total sales data were analyzed during a period of one year. The information was obtained from the company Probotix S.A.C. (company with a fictitious name for confidentiality of information). For the development of the research, data mining processes were used, and the technique used for data analysis was multiple linear regression.

The results obtained in the present investigation demonstrated that an explanation of the variability of sales of 88.7% was obtained with the multiple linear regression technique. With the results obtained in the investigation, it was concluded that four of seven dimensions of the sales projection model were significant for the estimation of sales, which were: transactions per customer with 45.1%, movement per item with 19.0%, number of suppliers with 29.7%, and number of warehouses with 18.3%, while the three remaining dimensions were not significant for the estimation of sales, which were: transactions in soles, invoice amount, and number of advisors. Finally, it was recommended to combine linear regression with some other statistical technique, such as time series analysis, with the purpose that the prediction model has a higher precision.

Keywords: Data mining, percentage variability, multiple linear regression, time series, prediction model.

I. INTRODUCCIÓN

Es fundamental contar con un pronóstico preciso de las ventas en toda organización debido a que estos datos contribuyen a planificar, proyectar, programar, decidir y organizar los planes estratégicos para obtener ventas más exitosas a la dirección o gerencia (Montemayor, 2012, p. 205). Por consiguiente, el modelo de proyección debe basarse en cada dimensión para ser más preciso y exacto (Montemayor, 2012, p. 205). Se planteó como problema general: ¿Cuáles fueron los porcentajes de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes?

El objetivo general de la investigación fue determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes. Esta investigación fue justificada teórica y financieramente, debido a que ha ayudado a lograr un modelo de pronóstico de las ventas que ayudará a la toma de decisiones de inversiones, abastecimiento y producción de bienes o servicios. Como hipótesis general, se formuló que las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes estuvieron relacionadas con las ventas de herramientas hidráulicas y electromecánicas.

Las ventas engloban todas las actividades y ejercicios necesarios para suministrar o proporcionar a futuros consumidores u organizaciones de un bien o servicio después del acuerdo monetario (Aleman, 2011, p. 142). Por otro lado, el significado de esta palabra “ventas” para las organizaciones va mucho más allá. Al respecto, Guzman (2014) expresó: “La venta consta de un proceso y muchos autores coinciden en la importancia de cada una de sus fases las cuales se han determinado de acuerdo con distintos criterios basados en un esquema sencillo que recoge diferentes fases para concretar el proceso de ventas” (p. 11) y que “Los pronósticos de ventas se realizan para períodos de tiempo de

proyección y se usan varias fórmulas de cálculo para determinar su valor” (p. 30). Por otro lado, Elías y Zorrilla (2014) indicaron:

Si la cifra de ventas es positiva no importaban ni las ineficiencias ni la forma y modos en los que se conseguían estas mejoras en la facturación y que con la llegada de caídas de ventas a partir de la llegada de las crisis económicas muchos de los que realizaban este ejercicio se dieron cuenta de su equivocación. Para llegar al resultado debemos medir estadísticamente los montos consumidos por cada cliente fidelizado (básicamente el importe mensual). (p. 8)

Por lo antes mencionado, se planteó lo siguiente: ¿Por qué es importante conocer las transacciones que realizan los clientes, la rotación o el movimiento que existe por producto, servicio o artículo? Al respecto, los especialistas de PR Newswire (2013) señalaron:

Se puede lanzar ofertas, promociones, descuentos y rebajas, lo cual hace que las organizaciones de diferentes rubros busquen aprender la elección de los consumidores en base a la demanda y transacciones de los mismos, por ello es que para obtener esta información se planifican promociones, descuentos o rebajas agresivas en distintos rubros y aplicados a diferentes productos y servicios con la finalidad de aprender de la elección, error o acierto de esta estrategia y así volverlo un punto demasiado ventajoso, ya que uno puede generar información para orientarla a la predicción o pronóstico de ventas basado en los criterios mencionados. Finalmente, aquel será el punto de quiebre de una toma de decisión para poder realizar los movimientos de acuerdo con lo que apunta la entidad dentro de un plazo determinado. (p. 132)

Debido a ello es fundamental contar con nuevos métodos, modelos o herramientas que ayuden en el día a día a proyectar las ventas de un determinado producto. Al respecto, Sifones et al. (2016) explicaron:

Las organizaciones poseen todo tipo de procedimientos entre buenas y malas prácticas dentro de todas sus áreas. El no saber proyectar o determinar distintos tipos de procesos, malas aplicaciones de regresión para su proyección manual o empírica, con poca información, pocas dimensiones. Es por ello por lo que las MYPE's y PYME's necesitan de un acceso a este tipo de tecnología con el fin de incrementar su capacidad de demanda de un producto o servicio entre todos sus consumidores. (p. 201)

Este estudio cuenta con una justificación teórica, ya que a través de su aplicación fue posible lograr un modelo para pronosticar las ventas. Este estudio también fue justificado económicamente, porque a través de su aplicación se identificó las variables que pueden influir en las decisiones de inversiones financieras para mejorar las ventas, con un enfoque claro, preciso y conciso.

Actualmente las técnicas de Data Mining (DM) se están convirtiendo en métodos efectivos para extraer grandes conjuntos de información y transformarla en diferentes estructuras legibles para su posterior uso (Hernández et al., 2014, p. 351). Además, Pando (2017) indicó que un modelo estadístico orientado a la información con una herramienta de automatización ágil puede tener un impacto importante dependiendo de cómo se aplique sobre una organización; también, se debe contar con un plan estratégico y un estudio de mercado para seguir escalando a donde se desea. (p. 36)

Sobre la base de la realidad problemática presentada se planteó el problema general y los problemas específicos de la investigación. El problema general de la investigación fue: ¿Cuáles fueron los porcentajes de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes? Los problemas específicos de la investigación fueron los siguientes:

- **PE1:** ¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacción por cliente?
- **PE2:** ¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable movimiento por artículo?
- **PE3:** ¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacciones en soles?
- **PE4:** ¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable monto por facturas?
- **PE5:** ¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable número de asesores?
- **PE6:** ¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de proveedores?
- **PE7:** ¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de almacenes?

El objetivo general de la investigación fue determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes. Los objetivos específicos fueron los siguientes:

- **OE1:** Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacción por cliente.
- **OE2:** Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable movimiento por artículo.

- **OE3:** Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacciones en soles.
- **OE4:** Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable monto por facturas.
- **OE5:** Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable número de asesores.
- **OE6:** Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de proveedores.
- **OE7:** Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de almacenes.

La hipótesis general fue: “Las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes estuvieron relacionadas con las ventas de herramientas hidráulicas y electromecánicas”. Las hipótesis específicas fueron las siguientes:

- **HE1:** La variable transacción por cliente estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Esta hipótesis se planteó en base a varios estudios sobre el pronóstico de ventas con transacciones por cliente utilizando la técnica de regresión lineal múltiple. Respecto a ello, Storch de Gracia (2018) expresó:

Los datos que permiten evaluar el desempeño de la empresa y que son posteriormente extraídos de los datos maestros y de los datos de las transacciones son: ventas por productos, ventas por clientes, proyectos por clientes, compras por proveedor, costo de reposición o control de la producción, los que son clasificados y denominados como datos de

control y como parte de la gestión y ejecución de la planificación de las ventas dentro de una organización que tenga definido el modelo de información que será requerido por el sistema; además, los datos deben mantener una estructura lógica para que finalmente se arme una verdadera base de conocimiento. (p. 122).

Las transacciones por cliente suelen ser información clave para la proyección de ventas. En este sentido, Stuart (2010) mencionó:

La impresión general es que, en contraste con los mercados en los que una empresa será grande ante la elevada competitividad en los mercados y las lucrativas transacciones por clientes concentradas en un único mercado principal hace que estos mercados conduzcan a empresas pequeñas y emprendedores que posteriormente maximizan sus ventas y terminan siendo un elemento esencial en las futuras planificaciones, pronósticos de ventas y crecimiento de las empresas. (p. 136)

- **HE2:** La variable movimiento por artículo estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Esta hipótesis se planteó en base a varios estudios sobre el pronóstico de ventas con movimientos por artículos utilizando la técnica de regresión lineal múltiple. En este sentido, Keat et al. (2011) mencionaron:

Si un pronóstico muestra una tendencia desfavorable en el movimiento de los artículos en particular para el año siguiente, el plan podría recomendar acciones que contrarrestaran esta predicción. Entonces, dadas las nuevas tácticas o estrategias incorporadas dentro del plan, el pronóstico original se sustituye por un pronóstico nuevo y revisado, con condiciones y suposiciones alteradas basados en la información obtenida por el movimiento de

todos los productos provisionados de los diferentes abastecimientos. (p, 224).

Keat et al. (2011) indicaron que la principal tendencia está basada en los movimientos por artículos. Por otro lado, Torres (2014) detalló: “El pronóstico de ventas es una estimación de los movimientos futuros (ya sean físicos o monetarios) de uno o varios artículos para un periodo determinado. Llevar a cabo el pronóstico de ventas permite elaborar el presupuesto de ventas y a partir de este, crear los demás presupuestos; como el de producción, de compra de insumos o mercadería, requerimiento de personal, flujo de efectivo, entre otros” (p. 131). El pronóstico permitirá saber cuánto movimiento va a existir en uno o varios artículos, cuanto se va a necesitar producir, cuanto se necesita de insumos o mercadería y cuanta inversión se necesitará para poder contar con ventas exitosas.

- **HE3:** La variable transacciones en soles estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Esta hipótesis se planteó en base a diferentes estudios sobre el pronóstico de ventas con transacciones en soles utilizando la técnica de regresión lineal múltiple. Respecto a ello, Gutiérrez (2013) explicó:

Para tener una proyección que sirva de base para la toma de decisiones gerenciales, deben proyectarse tres estados financieros básicos de la contabilidad que posteriormente sirven de puente para lograr que el balance general cuadre según cierta ecuación contable, la empresa proyecta transacciones en soles por el valor de 500 PEN., distribuidos en 100 PEN. De contado y 400 PEN. A crédito. Esta transacción en soles generará varios movimientos en los diferentes estados financieros que finalmente servirá para realizar una proyección más acertada respecto al movimiento del dinero. (p. 27)

Por lo mencionado anteriormente, se puede observar que las transacciones en soles son esenciales para que los modelos de regresión lineal múltiple sean más acertados financieramente. En consecuencia, Pacheco (2016) indicó:

Normalmente, una empresa mantiene dinero en efectivo y en inversiones transitorias para cubrir las transacciones en soles que realiza respecto a las ventas. El mantenimiento de fondos excedentes para fines preventivos en realidad representa un inventario de seguridad que, dependiendo de su volumen, tiene asociado un costo de inmovilización que repercute en la rentabilidad de la empresa, aunque inversamente acota el riesgo de insolvencia para el modelo de pronóstico de ventas (p. 261).

- **HE4:** La variable monto por facturas estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Esta hipótesis se planteó en base a diversos estudios sobre el pronóstico de ventas con montos por facturas utilizando la técnica de regresión lineal múltiple. Respecto a ello, Montes et al. (2019) detallaron:

La experiencia se refiere a los datos históricos en el comportamiento de las facturas, se hacen en base al análisis de los montos facturados y de los montos emitidos, esta información se acumula como parte de la rutina contable normal; es poco común que se pueda obtener de otra forma o fuente, a menos que se realice un costoso estudio especial. Por eso es importante un procedimiento bien definido del análisis de ventas para que el control presupuestario sea la base principal para realizar el análisis de ventas y donde el departamento de ventas debe de conceder un interés fundamental al procedimiento de codificación de los montos facturados. (p. 87)

Montes et al. (2019) explicaron que el análisis de las ventas es fundamental debido a que se basa en los montos por facturas. Por otro lado, Inmon (2012) explicó: “Al tratarse de pagos asociados a operaciones en gran medida vendrán determinadas por las ventas y existirán montos por facturas, planes de trabajo, etc., que podrán ser transformados en flujos de caja utilizando la metodología de proyección de ventas con información decisiva que finalmente servirán para las fuentes de financiación a corto plazo analizados por alguna herramienta” (p. 86). Inmon (2012) explicó que producir para generar facturas teniendo en cuenta las variaciones de existencias en función de los inventarios que se quieran tener de materias primas y de productos terminados, independientemente del caso que sea siempre que se consiga facturar los pagos pueden ser tratados como flujos principales y que posteriormente ayudarán con información fundamental para el pronóstico de las ventas de los siguientes trimestres.

- **HE5:** La variable número de asesores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Esta hipótesis se planteó en base a distintos estudios sobre el pronóstico de ventas con número de asesores utilizando la técnica de regresión lineal múltiple. Referente a eso, Albisetti (2018) explicó:

El posicionamiento estratégico es la forma como nos diferenciamos de la competencia ante el segmento de mercado objetivo. Respecto a la planeación que tiene como objetivo proyectar las ventas en el mediano plazo, una de sus primeras tareas es definir el número de asesores a nuestro alcance que, con alta probabilidad, estarán en total disponibilidad de crear la demanda del producto-servicio que se desea comercializar. (p. 236)

Por lo mencionado anteriormente, se puede decir que el número de asesores provee con observación personal, entrevistas, encuestas, búsquedas y el análisis de fuentes de inteligencia de mercado que

finalmente servirán para la proyección de ventas. Por consiguiente, Pacheco (2016) indicó:

Las consultas son frecuentemente realizadas a los asesores para los presupuestos quienes presentan un plan y estrategia de ventas con la finalidad de impulsar las mismas, partiendo del método de regresión para visualizar cuáles son los factores que influyen (o que tienen algún tipo de correlación entre sí) en las preferencias de los productos que se deseen comercializar con la finalidad de realizar la proyección. En base al planteamiento de los asesores, se inician los procesos del plan de ventas de los productos en conjunto con los diferentes departamentos que forman parte del proceso [mercadotecnia, manufacturación, etc.] (p. 247).

- **HE6:** La variable cantidad de proveedores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Esta hipótesis se planteó en base a distintos estudios sobre el pronóstico de ventas con cantidad de proveedores utilizando la técnica de regresión lineal múltiple. En cuanto a ello, Garduño (2014) mencionó:

La cantidad de proveedores es una pieza fundamental y estratégica para las empresas porque de ellos dependen los recursos e insumos para la manufactura de los productos en el plazo correspondiente y con calidad, dentro de la cadena logística de producción y entrega de estos a las ventas finales. Aún más importante, para el área de marketing, cuyo vínculo es de suma significancia para la entrega de los productos o servicios a los clientes con los estándares de calidad y en el momento requerido es la disponibilidad de materias primas y el costo de estos (p. 67).

Garduño (2014) planteó que las alianzas con los proveedores se logran con mucha comunicación a fin de sintonizarlos con la filosofía de la empresa, enfocada en la satisfacción del cliente y posteriormente para que el margen de ventas sea muy superior continuamente. Por otro lado,

Rajan (2014) describió: “Siempre se requiere y necesita diferentes características dentro de la cantidad de proveedores con los que la empresa tenga alguna alianza estratégica para potenciar sus ventas, las principales razones son de coste. Normalmente se negocia con un proveedor que pueda proporcionar las suficientes piezas para la empresa básicamente que cumpla con la demanda que se le exige y de la que se cuenta con una certeza absoluta, sin embargo, la siguiente razón prioritaria es el tiempo ya que si minimizas costes, pero aumentas el tiempo de producción de los productos, entonces se pierde el valor de la proyección de las ventas y no sería exacto” (p. 149).

- **HE7:** La variable cantidad de almacenes estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Esta hipótesis se planteó en base a diferentes estudios sobre el pronóstico de ventas con cantidad de almacenes utilizando la técnica de regresión lineal múltiple. Referente a eso, De La Arada (2015) manifestó:

Una de las tareas fundamentales a realizar en los almacenes es la gestión del inventario, entendido este como el conjunto de productos, bienes y mercancías que están disponibles para su venta o manufacturación. La pérdida de oportunidades y posterior impacto en la proyección de las ventas se da cuando el inventario se puede dañar o deteriorar, puede quedar obsoleto o caduca y sobre todo incurre en gastos administrativos y de seguros. Un abastecimiento sobregirado origina irregularidades en la venta o incluso en manufacturar otros productos, tener un abastecimiento sin uso y ocupando un espacio en los almacenes generan un gasto extra y por ende una pérdida en la proyección de ventas. (p. 77).

Por lo mencionado anteriormente, se puede observar que mientras más almacenes se tenga, es mayor el riesgo de perder oportunidades. Por lo tanto, Inmon (2012) explicó:

Los principales abastecimientos (seguridad y efecto) hacen referencias al nivel de existencias que debe haber en un almacén para hacer frente a cualquier eventualidad derivada de un incremento en la demanda o de posibles problemas que no aseguren la entrega o manufacturación del producto final. Gracias a la permanencia de ellos de manera habitual en los almacenes sin intención real de utilizarse, ya que ayudará a hacer frente ante posibles contingencias y que, dado la complejidad de prever la demanda exacta de un artículo o tendencia en una categoría de productos, se necesita trabajar con herramientas y realizar proyecciones y previsiones fiables para asegurar la venta de mercancía en las tiendas. (p. 27).

II. MARCO TEÓRICO

En este segundo capítulo se detalla los trabajos relacionados con investigación donde algunos fueron: redes neuronales para realizar procedimiento de pronóstico de ventas, inteligencia de negocios para desarrollo de nuevas tecnologías ágiles, inteligencia de negocios y sus comportamientos en el desarrollo y un marco para la implementación de la detección de problemas a través de tableros de control (KPI's), etc.; pero, no se encontró investigaciones con el mismo objetivo. Asimismo, se explica las teorías relacionadas: minería de datos para detectar problemas, aprendizajes basados en redes neuronales que evalúan el rendimiento de los pronósticos del esquema a nivel futuro y metodologías de minería de datos, de las cuales se aplicó el modelo de regresión lineal múltiple para la proyección de las ventas.

Salini y Pérez (2019) investigaron una estrategia de ventas para la orientación de variables e indicadores para el panorama empresarial en Chile para mejorar el proceso de venta y conseguir la demanda de los productos en el ranking para realizar una proyección en base a la demanda del mercado y tener un constante abastecimiento para la demanda del consumidor por temporada. Son estrategias con tipo de diseño de investigación transversal explicativo que se emiten en base al departamento de mercadeo y logística, entre las cuales se resalta principalmente el producto junto a la oferta que se determinó en el mismo en el catálogo en base a los tiempos que representan los picos de venta al momento del establecimiento de ofertas (Salini y Pérez, 2019, p. 68).

Pando (2017) analizó un caso de redes neuronales aplicado en una empresa con sede en Cuba, la que brindó la información para realizar el pronóstico de ventas, con un diseño metodológico de tipo no experimental con la información en el periodo 2008-2016. Pando (2017) concluyó que para posicionar el aprendizaje continuo se debe crear los procedimientos automatizados en base a la lógica de negocio y recopilación de información, haciendo afinidad para poder tener un margen de precisión del 80% en los resultados de confiabilidad (Pando, 2017, p. 352)

Syntetos y Zied (2015) aplicaron la inteligencia de negocios para detectar los problemas a través de tableros de control diseñados mediante un árbol de

decisión en una sede empresarial en Ecuador, con un diseño no experimental para trazar los gastos por problemas de producción de la empresa y solucionando las incidencias transaccionales. Syntetos y Zied (2015) concluyeron que, para ahorrar recursos para crear un nuevo proceso para la empresa en el área financiera, se debe potenciar la idea por explotar según las necesidades de la empresa, a través de una área creativa en la cual pueden brindar ideas importantes y su recurso fundamental es la información y en este caso las incidencias de la empresa (Syntetos y Zied, 2015, p. 31).

Hard (2014) implementó la predicción basada en una metodología con diseño no experimental para un modelo de redes neuronales, un esquema horizontal que también cuenta con algorítmica y el factor suerte no funciona, sino al contrario un modelo matemático y variables constituyen la correcta vía de soluciones. En este caso el autor se traza un objetivo en base al fútbol y sus mediciones son realizadas en base a información histórica de parte de cada equipo de fútbol y realiza un seguimiento por año y estados por fecha, además de un correlativo por juego ganado y perdido. Hard (2014) logró un margen alto de pronóstico para el partido ganado y recomendó el uso de variables que se fundamentan en base al juego para ver los puntos y crear la adaptación constante al juego hasta llegar a finalizar el producto (p. 177).

Saleh (2014) definió un marco de trabajo para determinar el tiempo de paradas por geolocalización en una empresa de transportes, adecuando factores o indicadores atípicos con un componente de preprocesamiento. Saleh (2014) concluyó que es funcional en un entorno y se aplica a la PYME en cuestión, bajo estándares y medidas de la propia empresa, todo esto para poder ahorrar costos y tiempo a la empresa. Saleh (2014) recomendó que se evalúe la posibilidad de magnificar el proceso de desarrollo a nivel servidor para poder consumir servicios web (p. 187).

Ryan (2013) estudió aplicaciones de escritorio orientadas a las empresas del sector industrial, bajo un esquema colaborativo de inteligencia de negocios. El diseño de la investigación fue no experimental descriptivo. Ryan (2013) concluyó que la utilización de metodologías predictivas (árbol de decisión o

regresión múltiple) que pueden aplicar a PYME's utilizando un planeamiento de procesos para la fase despliegue comercial es rentable, ya que analiza los precios y lanza nuevas alternativas comerciales de acuerdo con las temporadas de mayor consumo del cliente (p. 215). Finalmente, Ryan (2013) recomendó que para analizar un entorno siempre es requerida una metodología que permita analizar datos cuantitativos y cualitativos (p. 215).

Salini y Pérez (2019) indicaron que mediante la regresión lineal desarrollaron una mejora continua en el diagnóstico médico en el área metropolitana de Toluca. En este caso los patrones e indicadores se orientan a un paciente y también de acuerdo a los síntomas y descartes médicos, una línea muy delicada de la salud prima en esta investigación ya que es el bienestar humano. El diseño de la investigación fue preexperimental convencional con la participación de determinado grupo de profesionales del área médica. Por tanto, la información que se destaca es del diagnóstico acertado y preciso con un margen de error de 0.19% y va acorde de síntomas y de la mano de un profesional en la rama para poder realizar un trabajo denso y mostrar el reporte de síntomas que se reflejarán en diagnósticos médicos (Salini y Pérez, 2019, p. 54)

Salazar et al. (2019) estudiaron el efecto de la inteligencia de negocios en el área de logística, basándose netamente en indicadores y puntos clave, para la detección o demanda en el mercado por variables como tiempo, costo o consumo, lo cual establece un enlace directo con el tema a tocar. El diseño de la investigación fue descriptivo y estaba enfocado en los casos de negocio de la empresa. Salazar et al. (2019) concluyeron al analizar y aplicar la inteligencia de negocios que el desarrollo para la implementación de modelos lineales es un puente de buenas prácticas para solventar la investigación (p. 172).

Salazar et al. (2019) concluyeron que crear un enlace para la administración de puntos de entrega y reducción de tiempo en el área de distribución. Aunque no especifica las fases, demuestra la gran escala de una elaboración continua de datos, con perspectiva de operativa, asimismo reducir la carga de operaciones repetitivas y mejorando la oferta demanda para el estilo

de la empresa. Salazar et al. (2019) recomendaron que para futuras investigaciones se utilice un motor de base de datos no SQL para usar patrones de baja latencia dentro de la información (Salazar et al. 2019, p. 173).

Navarro (2015) explicó el problema de inventarios denominado producto hueso a través del historial de venta de producto para poder diagnosticarlo en un ERP de una empresa ubicada en Colombia, para esta evaluación se determinó el comportamiento de rotación del producto mediante las metodologías de Croston y Winters que son las más adecuadas a las dimensiones del producto y problemática hallada en Colombia. Navarro (2015) recomendó que antes ver el historial de un producto, es preferible medir el alcance de rotación entre almacenes u obtener un historial de venta de más de cinco años (Navarro, 2015, p. 32)

Salini y Pérez (2019) determinaron un modelo relacional de base de datos para un proyecto en el área comercial en Argentina, se determinó el ciclo de comportamiento para luego determinar las variables y realizar las proyecciones con un modelo de regresión lineal, asimismo evaluaron varios modelos estadísticos para comparar los resultados. Salini y Pérez (2019) recomendaron evaluar de manera rigurosa el modelo que se aplicará para la compatibilidad de la información y orientarlo al objetivo de la investigación (p. 32).

Garduño (2014) señaló las diversas metodologías para poder aplicar a un pronóstico de ventas que utilizó para aplicarlo a un caso de estudio, utilizó la metodología Winter para realizar un hipercubo que abarca registros de mayor proporción, solo se basa en el historial de información. El objetivo fue aplicar la metodología propuesta para producir inteligencia artificial en base a la metodología winter. Determinó los valores predictivos y un cubo OLAP para almacenamiento de información. (Garduño, 2014, p. 84). Se recomendó un mayor alcance en las casuísticas, ya que el aprendizaje de la aplicación se determina mediante decisiones y actividades con un patrón en común (Garduño, 2014, p. 84).

Ionos (2019) impulsó un modelo de regresión lineal para una empresa petrolera que ubicar un patrón puntos petroleros en Venezuela, construyeron una investigación de diseño pre-experimental. El objetivo fue realizar pronóstico de puntos consumidos en función a algoritmos verticales para el diseño terrestre de los planos, con lo cual finalizaron con una integración para pronosticar un punto de quiebre petrolero usando la técnica de pronóstico tradicional como base de su investigación. Ionos (2019) recomendó no ejecutar los mismos diseños algorítmicos, la información geológica es muy variable, ya que geológicamente es irregular y se requiere actualizar de manera diaria la información (Ionos, 2019, p. 52)

Molina et al. (2016) realizaron Cubos OLAP'S para crear un modelo de predicción mediante el árbol de decisión para el registro de una librería privada en Chile, su estudio fue de diseño no experimental en tendencia longitudinal. lograron crear la aplicación de escritorio para facilitar el registro por cada libro con lector de barras, así la automatización les permite la agilización de permiso de productos y llegada al cliente para poder motivarlo a la compra y obtención de productos de la empresa proveedora. Concluyeron que mediante los cubos de información pudieron clasificar los libros para la biblioteca mediante la información de código de barras, recomiendan seguir las buenas prácticas del modelo de generación de cubos OLAP'S que usa la metodología de árbol de decisión (Molina et al., 2016, p. 64)

Garduño (2014) definió medir los costos de compra y precios de venta del producto, cantidades y movimiento de rotación del producto para armar el modelo de indicadores. El objetivo fue ejecutar el pronóstico de ventas en base a los indicadores, lo cual nos llevaría a la siguiente fase que sería construir un modelo de la distribución de ventas de cada producto para calcular el riesgo que se tendría en situaciones extremas de ventas (Garduño, 2014, p. 65). Se concluyó que usando los indicadores mencionados aumentan la confiabilidad para la investigación y recomiendan que para objetivos se necesitan mapear los procesos de la empresa y adaptar nuevos indicadores para futuras investigaciones (Garduño, 2014, p. 65).

Maheshwari (2015) realizó la minería de datos con tecnología ubiti para descubrir patrones útiles en la compra de medicina en boticas. Con una investigación de tipo longitudinal eligió un patrón para crear un modelo aplicable de manera simple. Maheshwari (2015) concluyó que los datos revelaron patrones de consumismo en el cliente, referenciando los productos medicinales más comprados, ya que los datos pasados pueden tener valor predictivo en casuísticas complejas (Maheshwari, 2015, p. 198).

Sifones et al. (2016) manifestaron que Python es un lenguaje de programación interpretado para generar dar una curva de aprendizaje. La sintaxis de Python es simple y es un lenguaje dinámico (Sifones et al., 2016, p. 218). Con una investigación de tipo explicativo indicó que cuenta con estructuras de datos eficientes de alto nivel y está enfocado a programación orientada a objetos. Sifones et al. (2016) concluyeron que las características mencionadas permiten un desarrollo extenso de aplicaciones en diversas áreas y la mayoría de las plataformas. Recomendó que es preferible implantarlo en un servicio web para poder recoger más información y así poder implementar más bucles del negocio (Sifones et al., 2016, p. 218)

Rosas et al. (2016) indicaron que los objetivos a corto plazo son una meta automatizada para aumentar el volumen de ventas en el rubro ferretero (Rosas et al., 2016, p. 69). Mediante una investigación de tipo explicativo se logró concluir que el volumen total de ventas necesario para alcanzar esta cantidad de facturación restando todos los costos y gastos anuales de la empresa, con la ayuda de los sectores de contabilidad y financiero. Rosas et al. (2016) recomendaron determinar siempre la fuente de la información y clasificarla en fases de un proceso (Rosas et al., 2016, p. 69)

Everitt (2014) indicó que la regresión múltiple se utiliza para construir una fórmula con dimensiones medibles de margen cuantitativo para poder detectar incidencias o relaciones dentro del campo matemático, utilizando una investigación de diseño explicativo. Everitt (2014) concluyó que el modelo de regresiones tuvo un margen alto de confiabilidad. Recomendó que existen límites estándares para medir su confiabilidad y es importante diferenciar las variables de las dimensiones para evitar errores de confiabilidad (Everitt, 2014, p. 73).

III. METODOLOGÍA

En este capítulo se explica el método que se utilizó para llevar a cabo la investigación del sistema de pronóstico de ventas de herramientas hidráulicas y electromecánicas, en donde el estudio fue de tipo aplicado, con enfoque cuantitativo y diseño no experimental, transversal y correlacional-causal porque se buscó describir el efecto y relación entre dos o más variables. La variable dependiente fue las ventas y las variables independientes fueron las siguientes: transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes. La población estuvo compuesta por los registros de ventas de 96 meses. El método que se utilizó para obtener una precisión fiable fue la regresión lineal múltiple, además se realizó una depuración de datos para contar con información más legible para el modelo. Asimismo, esta investigación cumplió con lo establecido en el Código de Ética en Investigación de la Universidad César Vallejo y el Código de Ética del Colegio de Ingenieros del Perú.

3.1 Tipo y diseño de investigación

La presente investigación fue de tipo aplicada, de enfoque cuantitativo y de diseño no experimental, transversal y correlacional-causal debido a que se analizó la información de las ventas totales en los registros de 12 meses. Hernández et al. (2014) mencionaron que una investigación es aplicada cuando se tiene como objetivo determinar un problema o planteamiento específico, enfocándose en la búsqueda y consolidación del conocimiento para su aplicación. Asimismo, Hernández et al. (2014) precisaron: “La investigación cuantitativa considera que el conocimiento debe ser objetivo, y que este se genera a partir de un proceso deductivo en el que, a través de la medición numérica y el análisis estadístico inferencial, se prueban hipótesis previamente formuladas” (p. 129).

Con respecto al diseño correlacional-causal, Hernández et al. (2014) indicaron: “Pueden limitarse a establecer relaciones entre variables sin precisar sentido de causalidad o pretender analizar relaciones causales. Cuando se limitan a relaciones no causales, se fundamentan en planteamientos e hipótesis

correlacionales; del mismo modo, cuando buscan evaluar vinculaciones causales, se basan en planteamientos e hipótesis causales” (p. 157).

3.2 Variables y su operacionalización

Las variables de estudio están en el **anexo “Matriz de consistencia”**.

3.3 Población, muestra y muestreo

3.3.1 Población

El equipo de seguimiento al análisis proporcionó los registros de ventas de 96 meses, los que constituyeron la población en estudio.

3.3.2 Muestra y Muestreo

Para la selección de la muestra se usó el muestreo no probabilístico por conveniencia con un total de registros de ventas de 12 meses.

3.4 Técnicas e instrumentos de recolección de datos

Para la elaboración de esta investigación se usó la técnica de la observación y como instrumento se utilizó la ficha de observación. Se usó la validez de contenido en base a citas y referencias de artículos científicos. Además, no se usó cuestionarios, sino fichas de observación, por lo que no se realizó cálculos de confiabilidad. Sin embargo, cabe resaltar que se usó un nivel de confianza de 95% en la aplicación de las pruebas estadísticas (Herrera, 2015, p. 206). El instrumento del estudio se encuentra en el **anexo “Desarrollo de la aplicación”**.

3.5 Procedimientos

Cuando se concluyó la implementación del modelo de pronóstico, se continuó con la evaluación de la muestra de los registros proporcionados. La muestra por conveniencia se sometió a una observación directa para analizar si el modelo implementado obtuvo un resultado significativo respecto al pronóstico realizado. Se siguieron los pasos detallados a continuación:

- a) Se definieron los objetivos que se buscan o persiguen, con la finalidad de definir su entidad de observación.
- b) Se realizó la planificación de los objetivos a observar, con el propósito de congregar la validez y confiabilidad.
- c) Se realizó la preparación de los investigadores que cumplen el rol de observadores para la evaluación de la información.
- d) Los observadores realizan la recolección de los datos que directamente implican con la tarea y/o actividad objeto de observación.
- e) Se realizó un proceso de conversión y estructuración de la base de datos de las ventas, para ello se necesitó crear un origen de información basado en los datos almacenados. Se recuperaron dichos datos y se aplicó un proceso de conversión basado en reglas de transformación para limpiar (resolver inconsistencia en la información, identificar y remover valores atípicos con un margen del 5% de los datos superiores e inferiores), filtrar (supresión de información irrelevante), disgregar (minimizar volúmenes de información para mayor eficiencia en la carga de trabajo) e integrar (identificación de redundancia) la información y un destino para almacenar la información ya procesada.
- f) Se creó una estrategia de análisis y almacenamiento de información mediante un data mart para la necesidad de las ventas con la finalidad de poder descomponer y descentralizar la información por dimensión para la eficiencia en la carga de trabajo de la siguiente actividad.

3.6 Método de análisis de datos

En este estudio se utilizó la regresión lineal múltiple como técnica estadística (Ramos et al., 2018, p. 88) previa depuración de los datos, la cual se ejecutó mediante un proceso de conversión y estructuración de la base de datos de las ventas. Para ello se necesitó crear un origen de información basado en los datos almacenados, se recuperaron dichos datos y se aplicó un proceso de conversión basado en reglas de transformación para limpiar (resolver inconsistencia en la información, identificar y remover valores atípicos con un margen del 5% de los datos superiores e inferiores). Asimismo, se procedió a filtrar (supresión de información irrelevante), disgregar (minimizar volúmenes de información para mayor eficiencia en la carga de trabajo) e integrar (identificación de redundancia) la información y un destino para almacenar la información ya procesada.

3.7 Aspectos éticos

Esta investigación se acogió a los lineamientos de la ética profesional expuestos por el Código de Ética del Colegio de Ingenieros del Perú que se mencionan en el artículo 15, disponiendo los principios y conductas fundamentales, así como: inclusión social, justicia, respeto, solidaridad, responsabilidad, honor profesional, honestidad y lealtad profesional (Colegio de Ingenieros del Perú, 2018, p. 3).

Esta investigación cumplió con los principios de ética en investigación propuestos en los capítulos I, II y III del Código de Ética en Investigación de la Universidad César Vallejo, siendo la honestidad, el respeto a la autoría de las fuentes y el rigor científico los que fueron aplicables a esta investigación. Asimismo, se cumplió con dar el consentimiento por escrito para la realización de la investigación y se mantuvo bajo anonimato el nombre de la institución donde se llevó a cabo el estudio de acuerdo con el artículo 14 (Universidad César Vallejo, 2017, p. 7).

IV. RESULTADOS

En este capítulo se detalla los resultados obtenidos dentro de la investigación con respecto al pronóstico de las ventas (variable dependiente), basándose en los indicadores beta de las variables independientes: transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes. Se describió el efecto del uso del sistema de pronóstico de ventas. A continuación, se detallará los resultados obtenidos con el modelo de regresión lineal múltiple.

En la tabla 1 se muestra el análisis ANOVA que indica que el modelo tiene un ajuste significativo porque presenta un valor inferior a 0.05.

Tabla 1: ANOVA del modelo de regresión lineal múltiple de las ventas

Modelo	Suma de cuadrados	Promedio de cuadrados	F	Sig.
Regression	5442.731	5442.731	759.928	.019
Residue	458.379	7.162		
Total	5901.110			

En la tabla 2 se puede ver que el modelo generado para las ventas presenta un R cuadrado ajustado de 0.887; es decir, la varianza de las ventas se explica al 88.7%. Por otro lado, en la prueba del estadístico de Durbin Watson, el valor estuvo dentro del rango de 1.5 a 2.5, lo que indica que los residuos no estuvieron auto correlacionados y el modelo es fiable.

Tabla 2: Resumen del modelo de regresión lineal múltiple de las ventas

Modelo	R Cuadrado	R Cuadrado Ajustado	Estimación del error estándar	Durbin – Watson
Ventas	0.887	0.880	1.970	1.879

En la tabla 3 se muestra el estadístico descriptivo correspondiente a las variables.

Tabla 3: *Estadísticos descriptivos de las variables del modelo de regresión lineal múltiple de las ventas*

Variable	Promedio	Desviación estándar
Ventas	86.154	6.4983
Trx_Cliente	5.5130	3.4105
Mov_Articulo	0.113	.1538
Trx_Soles	3.1230	2.2055
Monto_Facturas	0.3549	2.5718
Num_Asesores	0.624	.8233
Cant_Proveedores	0.1360	1.1578
Cant_Almacenes	0.1650	1.1875

En la tabla 4 se indica los coeficientes no estandarizados (B) que ayudarán a formar la ecuación para la predicción y los coeficientes estandarizados (β) indicarán la importancia de cada variable.

Tabla 4: *Coefficiente del modelo de regresión lineal múltiple de las ventas*

Variable	B	Desv. Error	β	Sig.	VIF
(Constante)	-.433	.573	-	.018	-
Trx_Cliente	.077	.009	.451	.044	1.257
Mov_Articulo	.080	.020	.190	.019	1.145
Trx_Soles	1.003	.997	.037	.104	1.569
Monto_Facturas	.810	.011	.475	.055	2.536
Num_Asesores	1.041	.959	.284	.193	3.256
Cant_Proveedores	.790	.210	.297	.032	1.002
Cant_Almacenes	.984	-.016	.183	.049	1.256

Ecuación del modelo de regresión lineal múltiple de las ventas

Al reemplazar los valores B donde su significancia sea menor a 0.05 en la ecuación, el modelo podría definirse de la siguiente manera:

$$\text{Ventas} = -0.433 + 0.077 \text{ Trx_Cliente} + 0.08 \text{ Mov_Artículo} + 0.79 \text{ Cant_Proveedores} + 0.984 \text{ Cant_Almacenes} + \epsilon$$

4.1 Prueba de la hipótesis específica 1

HE1: La variable transacción por cliente estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Interpretación de los valores B

El coeficiente no estandarizado B para la variable transacciones por cliente fue 0.77. Este valor indica que si la variable transacciones por cliente es incrementada en una unidad; entonces, Ventas aumenta en 0.77. Además, si la variable transacciones por cliente es reducida en una unidad; entonces Ventas disminuye en 0.77. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

Interpretación de los valores β

El coeficiente estandarizado beta (β) para la variable transacciones por cliente fue 0.451. Este valor indica que si la variable transacciones por cliente es incrementada o reducida en una desviación estándar (3.4105); entonces, Ventas es incrementada o reducida en 0.451 desviaciones estándares ($0.451 \times 6.4983 = 2.9307$), respectivamente. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

4.2 Prueba de la Hipótesis específica 2

HE2: La variable movimiento por artículo estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Interpretación de los valores B

El coeficiente no estandarizado B para la variable movimientos por artículo fue 0.080. Este valor indica que si la variable movimientos por artículo es incrementada en una unidad; entonces, Ventas aumenta en 0.080. Además, si la variable movimientos por artículo es reducida en una unidad; entonces Ventas disminuye en 0.080. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

Interpretación de los valores β

El coeficiente estandarizado beta (β) para la variable movimientos por artículo fue 0.190. Este valor indica que si la variable movimientos por artículo es incrementada o reducida en una desviación estándar (0.1538); entonces, Ventas es incrementada o reducida en 0.190 desviaciones estándares ($0.190 \times 6.4983 = 0.1234677$), respectivamente. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

4.3 Prueba de la hipótesis específica 3

HE3: La variable transacciones en soles estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Interpretación de los valores B

El coeficiente no estandarizado B para la variable transacciones en soles fue 1.003. Este valor indica que si la variable transacciones en soles es incrementada en una unidad; entonces, Ventas aumenta en 1.003. Además, si la variable transacciones en soles es reducida en una unidad; entonces Ventas disminuye en 1.003. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

Interpretación de los valores β

El coeficiente estandarizado beta (β) para la variable transacciones en soles fue 0.037. Este valor indica que si la variable transacciones en soles es incrementada o reducida en una desviación estándar (2.2055); entonces, Ventas es incrementada o reducida en 0.037 desviaciones estándares ($0.037 \times 6.4983 = 0.24043$), respectivamente. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

4.4 Prueba de la hipótesis específica 4

HE4: La variable monto por facturas estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Interpretación de los valores B

El coeficiente no estandarizado B para la variable monto por facturas fue 0.810. Este valor indica que si la variable monto por facturas es incrementada en una unidad; entonces, Ventas aumenta en 0.810. Además, si la variable monto por facturas es reducida en una unidad; entonces Ventas disminuye en 0.810. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

Interpretación de los valores β

El coeficiente estandarizado beta (β) para la variable monto por facturas fue 0.475. Este valor indica que si la variable monto por facturas es incrementada o reducida en una desviación estándar (2.5718); entonces, Ventas es incrementada o reducida en 0.475 desviaciones estándares ($0.475 \times 6.4983 = 3.0866$), respectivamente. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

4.5 Prueba de la hipótesis específica 5

HE5: La variable número de asesores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Interpretación de los valores B

El coeficiente no estandarizado B para la variable número de asesores fue 1.041. Este valor indica que si la variable número de asesores es incrementada en una unidad; entonces, Ventas aumenta en 1.041. Además, si la variable número de asesores es reducida en una unidad; entonces Ventas disminuye en 1.041. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

Interpretación de los valores β

El coeficiente estandarizado beta (β) para la variable número de asesores fue 0.284. Este valor indica que si la variable número de asesores es incrementada o reducida en una desviación estándar (0.8233); entonces, Ventas es

incrementada o reducida en 0.284 desviaciones estándares ($0.284 \times 6.4983 = 1.8455$), respectivamente. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

4.6 Prueba de la hipótesis específica 6

HE6: La variable cantidad de proveedores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Interpretación de los valores B

El coeficiente no estandarizado B para la variable cantidad de proveedores fue 0.790. Este valor indica que si la variable cantidad de proveedores es incrementada en una unidad; entonces, Ventas aumenta en 0.790. Además, si la variable cantidad de proveedores es reducida en una unidad; entonces Ventas disminuye en 0.790. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

Interpretación de los valores β

El coeficiente estandarizado beta (β) para la variable cantidad de proveedores fue 0.297. Este valor indica que si la variable cantidad de proveedores es incrementada o reducida en una desviación estándar (1.1578); entonces, Ventas es incrementada o reducida en 0.297 desviaciones estándares ($0.297 \times 6.4983 = 1.92999$), respectivamente. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

4.7 Prueba de la hipótesis específica 7

HE7: La variable cantidad de almacenes estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.

Interpretación de los valores B

El coeficiente no estandarizado B para la variable cantidad de almacenes fue 0.984. Este valor indica que si la variable cantidad de almacenes es incrementada en una unidad; entonces, Ventas aumenta en 0.984. Además, si

la variable cantidad de almacenes es reducida en una unidad; entonces Ventas disminuye en 0.984. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

Interpretación de los valores β

El coeficiente estandarizado beta (β) para la variable cantidad de proveedores fue 0.183. Este valor indica que si la variable cantidad de proveedores es incrementada o reducida en una desviación estándar (1.1875); entonces, Ventas es incrementada o reducida en 0.183 desviaciones estándares (0.183 X 6.4983 = 1.18918), respectivamente. Esta interpretación es válida si los efectos de las otras variables permanecen constantes.

4.8 Resumen de las hipótesis

A continuación, se presenta la tabla resumen de las comprobaciones de las hipótesis de la investigación:

Tabla 5: *Resumen de los resultados de las comprobaciones de las hipótesis*

Código	Hipótesis	Resultado
HE1	La variable transacción por cliente estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Aceptado
HE2	La variable movimiento por artículo estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Aceptado
HE3	La variable transacción en soles estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Rechazado
HE4	La variable monto por facturas estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Rechazado

Código	Hipótesis	Resultado
HE5	La variable número de asesores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Rechazado
HE6	La variable cantidad de proveedores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Aceptado
HE7	La variable cantidad de almacenes estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Aceptado
HG	Las variables transacción por cliente, movimiento por artículo, transacciones en soles, monto por facturas, número de asesores, cantidad de proveedores y cantidad de almacenes estuvieron relacionadas con las ventas de herramientas hidráulicas y electromecánicas.	Rechazado

De acuerdo con la tabla 5, en base a los resultados obtenidos se demostró que cuatro de las hipótesis específicas planteadas fueron aceptadas, logrando cumplir cuatro objetivos específicos. Se generó un incremento de precisión de 88.7% sobre las ventas.

V. DISCUSIÓN

Los procesos de inteligencia de negocios y minería de datos fueron aplicados para la depuración de una gran cantidad de información en la que se buscó patrones y tendencias que ayudaron a explicar el comportamiento de la información de las ventas para las dimensiones determinadas. La regresión lineal múltiple tuvo un resultado mejorado con respecto a otras investigaciones. Se obtuvo una explicación de la variabilidad de las ventas del 88.7% con el modelo de pronóstico desarrollado.

En base a los resultados obtenidos en la presente investigación se realizaron comparaciones con otras investigaciones sobre pronósticos de ventas. La técnica de regresión lineal múltiple fue aplicada a una fuente de información que contiene las ventas de la empresa PROBOTIX S.A.C. (empresa con nombre ficticio para reserva de información), ya que el propósito fue pronosticar la cantidad de ventas. En esta investigación se establecieron 7 variables independientes. Las variables relacionadas fueron las siguientes: transacciones por cliente (Trx_Cliente), movimiento por artículo (Mov_Articulo), cantidad de proveedores (Cant_Proveedores) y cantidad de almacenes (Cant_Almacenes), con explicaciones de la variabilidad de las ventas del 45.1%, 19.0%, 29.7% y 18.3%, respectivamente. Por lo contrario, las variables no significativas al 95% de confianza fueron las siguientes: transacciones en soles (Trx_Soles), monto por facturas (Monto_Facturas) y número de asesores (Num_Asesores).

La variable transacción por cliente utilizada para la proyección de ventas de la presente investigación obtuvo un grado de contribución en las ventas del 45.1%, lo que fue mayor a los resultados de la investigación de Montemayor (2012), quien obtuvo una contribución de 18.32%. En el estudio de Montemayor (2012) se usó información genérica para negocios, lo cual no permitió que el modelo genere una precisión más certera, además de tener un modelo de fases, por lo cual los resultados fueron menores a los resultados de la presente investigación.

El resultado de la variable movimientos por artículo de la presente investigación fue 19.0%, el cual fue diferente al resultado de la investigación de Aleman (2011), ya que dicha investigación obtuvo en su variable de movimientos

por artículo de 5.17%. El estudio de Aleman (2011) estuvo basado en determinados artículos de fabricación genérica más no en todos los movimientos que existieron de artículos, además de tener objetivos diferentes a los de la presente investigación.

Los resultados de la variable transacción en soles fueron no significativos para explicar las ventas, a diferencia de los resultados de Elías et al. (2014), quienes obtuvieron 10.4% en la variable transacción en soles, debido a que las transacciones monetarias fueron convertidas a la divisa de soles en el proceso de depuración de la información, mientras que para la presente investigación en el proceso de la depuración y filtración de datos se utilizaron las transacciones que tenían como tipo a soles.

En este estudio se obtuvo un 47.5% de contribución en la variable de monto por facturas sobre la proyección de ventas, la cual obtuvo un resultado mayor al 14.7% del estudio de Guzman (2014). El estudio de Guzman (2014) se basó en la demanda de artículos y no en las ventas totales para realizar su modelo, además los objetivos fueron diferentes a los que se tuvo en la presente investigación.

En el presente estudio se obtuvo un 28.5% en la variable de número de asesores para el pronóstico de ventas, la cual fue un resultado mayor con respecto al estudio de la investigación de Aleman (2011), en que el resultado del estudio obtuvo 25.02%. En el estudio de Aleman (2011) se tuvo que la contribución de la variable número de asesores fue relativamente menor debido a que la información estuvo referida al producto del proceso de las ventas comerciales y factores externos (competencia y tendencias) y no de las ventas históricas como fue el caso de la presente investigación.

En este estudio se obtuvo un 29.7% de contribución en la variable cantidad de proveedores respecto a la proyección de las ventas y que obtuvo un mayor resultado a la de Inmon (2012), en la que obtuvo un 19.7% en su variable de cantidad de proveedores. En el estudio de Inmon (2012) se usó la inferencia de ventas a futuro, además los objetivos de su investigación fueron diferentes y

la información de la presente investigación, la que tuvo enfoque y priorización en los factores para las ventas históricas.

En cambio, el resultado obtenido de la variable cantidad de almacenes del presente estudio fueron diferentes a los resultados de Perier (2016), en los que obtuvo 23.8% en esa variable, debido a que se basó en optimización de la cadena logística de la venta. La investigación de Perier (2016) tuvo por objetivo determinar únicamente el almacenaje de los productos más no todo lo que abarca las ventas de estos, el ciclo de vida de la venta, su información histórica, y la adquisición de nuevo abastecimiento para el almacén, seguridad y preservación del abastecimiento.

VI. CONCLUSIONES

Las conclusiones de la presente investigación fueron las siguientes:

1. La variable de transacciones por cliente en función de las ventas tuvo un grado de contribución alto respecto a las investigaciones de otros autores, debido a que la información utilizada contempla el flujo de ventas de una determinada organización y no se basa en información genérica de ventas.
2. Es importante crear reglas que ayuden a poder contar una información más consistente y robusta para que cuando sea procesada toda la información la técnica sea capaz de calcular un mejor impacto en la contribución de la variable.
3. Las transacciones en soles con menor contribución no son muy útiles dentro del cálculo del modelo de proyección (relación del 3.7%), porque la información posterior a la depuración fue poca y porque la mayor cantidad de transacciones fueron registradas con una divisa diferente.
4. Con la técnica de regresión lineal múltiple en el modelo de regresión lineal se obtuvo un R cuadrado ajustado del 88.7% en las ventas, lo que puede mejorarse si el efecto de los montos por facturas son significativos para el modelo.
5. Con mayor impacto en las ventas se tuvo la cantidad de proveedores, debido a que en el proceso de venta los resultados dependen en mayor parte por la logística, la disposición de recursos o servicios, obteniendo un grado de contribución del 29.7% sobre las ventas y manteniendo las demás condiciones constantes.
6. La cantidad de almacenes tuvo un impacto del 18.3%, siendo la variable con una significancia mayor al promedio en el aumento en las ventas. Por ello, se debería enforzar más a la cantidad de almacenes en las ventas.

7. El uso de una herramienta como Power BI para la construcción del modelo de minería de datos es sumamente práctico, ya que tiene muchas funcionalidades de minería de datos, además de un módulo para la codificación en lenguaje de programación Python, el cual es el lenguaje más utilizado para realizar este tipo de modelos y que además estas herramientas son de uso fácil, rápido y sencillo.
8. En la depuración de información se debería considerar su mejor distribución y definición, ya que así genera que las variables cuenten con un mayor nivel de significancia y se genere un pronóstico más preciso.

VII. RECOMENDACIONES

Las recomendaciones para futuras investigaciones son las siguientes:

1. Elaborar modelos relacionados con los montos por facturas o el número de asesores, debido a que se pudo observar que el nivel de contribución fue alto y puede influenciar en las ventas.
2. Combinar la regresión lineal múltiple con otra técnica estadística como por ejemplo el procesamiento en stream, el cual permite trabajar el modelo con información continuamente para descentralizar la tasa de entrada de información.
3. Incrementar la muestra de información para evaluar. En el presente estudio se utilizó información con un período de 1 año, esto podría extenderse a más cantidad de información o registros para buscar y procesar datos más relevantes para el pronóstico de ventas.
4. Realizar planos de datos sin procesar y residuales que también pueden ser informativos y poder visualizar el pre y post de los datos utilizados mediante un proceso de conversión y depuración de información para poder revalorizar los datos que vienen filtrados por el conjunto de reglas de discriminación de información duplicada e inconsistente y se minimizaría el uso de recursos para la construcción del modelo.
5. Considerar que las variables con cierto grado de significancia en un modelo de regresión lineal múltiple solo dan a explicar que existe una correlación pero que más no implica una causalidad. Para el establecimiento de la causalidad se tiene que desarrollar y emplear un experimento diseñado con aleatorización de información.
6. Implementar gráficos de residuos, los cuales son una forma fácil y eficiente para validar si el modelo de regresión lineal múltiple desarrollado cuenta con algún grado de problema y también aporta valor para la mejora del modelo.

REFERENCIAS

- ALBISETTI ROMANOL, E. 2018. Técnicas para la estrategia empresarial. Naviery SA. Italia: Florencia [Consulta: 14 de Agosto de 2018] Disponible en: <https://search.proquest.com/docview/39252?accountid=842236>
- ALEMAN, Jorge. Definición de modelo de proyección de ventas comerciales. Argentina: Las Industrias REC E.I.R.L, 2011. ISBN: 828-622-9074-02-22
- DE LA ARADA, M. Análisis inferencial de Ventas a futuro. Brazil: REMORD S.A.C,2015. ISBN: 582-1458-4552-71-8.
- ELÍAS, J. & ZORRILLA, F. Desarrollo de un data mart para mejorar la toma de decisiones en el Área de Planificación Comercial del Segmento Premium de Telefónica del Perú. Universidad Mayor de San Marcos, Perú: Lima, 2014. [Consulta: 29 de abril de 2019] <http://ateneo.unmsm.edu.pe/ateneo/handle/123456789/2754?mode=full>
- EVERITT, Bances. A handbook of statistical analysis using R. Pro: University Champan Park. Estados Unidos: Cambridge,2014. ISBN: 800-192-491-8
- GARDUÑO GUILLEN, G. Metodología para calcular el pronóstico de ventas y una medición de su precisión en una empresa farmacéutica: caso de estudio. México: Monterey, 2014. ISBN: 987-42-5821-488-4
- GUZMAN, Fiorella. Demanda de artículos de mecánica pesada y electrónica orientada al mercado internacional. Colombia: Cali, 2014. ISBN: 928-62-7534-2-8.
- GUTIÉRREZ, Ulises. Estrategias, negocios y Mercados Empresariales. Colombia: Bogota ,2013. ISBN: 886-662-1255-84
- HARD SIDNEY, J. 2014. Road landslide information management and forecasting system based on GIS. Environ Monit Assessment. [Consulta: 21 de agosto de 2020]. Disponible en: <https://doi.org/10.1007/s10661-008-0492-2>.

- HERNÁNDEZ, R.J., FERNÁNDEZ C. y BAPTISTA M. Metodología de la Investigación (6ta Edición). México: Monterey, 2014. ISBN: 553-58-585-4
- INMON, B. Construyendo un Data Warehouse. 3ra ed. EEUU: Editorial Wiley, 2012. ISBN:958-58-2541-488-4
- IONOS SEGURA, S. 2019. Software de data mining. [Consulta: 29 de Setiembre de 2020] Disponible en: <https://www.ionos.es/digitalguide/online-marketing/analisis-web/softwarede-data-mining-las-mejorsherramientas=?>
- KEAT, M., RODRIGUEZ, A., OLIVER, H., PRADA, M. E., MESIAH ALBITES, M., & DÍAS ROSSI, M. E. Pasos de la metodología predictiva a través de la regresión Lineal. Universidad del Norte. Colombia: CSIC, 2011. [Consulta: 14 de enero de 2019]. Disponible en: https://books.google.com.pe/books?id=X73sjfjS7sjd&dq=libro+estad%C3%ADstico+Fjfsf+sds3&lr=&hl=es&source=gbs_navlinks_s
- MAHESHWARI SHIRA, A. Business Intelligence and Data Mining (1ra Edición), Estados Unidos: Cambrige, 2014. ISBN: 946-405-849-657-7
- MAHESHWARI SHIRA, A. Data Analytics Made Accessible. Iowa, Estados Unidos: Cambrige, 2015. ISBN: 996-255-582-455-4
- MOLINA, J. & VÁSQUEZ A, Arturo 2016. A visual guide to CRISP-DM methodology. [Consulta: 10 de marzo de 2019] Disponible en: <https://exde.wordpress.com/2009/03/13/a-visual-guide-to-crisp-dmmethodology/>
- MONTEMAYOR CASTILLO, E. Métodos de pronósticos para negocios. Editorial Digital, Instituto Tecnológico y de Estudios Superiores de Monterrey, México: Monterey, 2012.ISBN:9548-58-4215-644-4
- MONTES, M., MICHEL, Á., RAMÍREZ, H., & PASTOR, M. E. Analisis de vectores a través de regresión lineal. Colombia: CSIC, 2019. [Consulta: 02 de noviembre de 2019]. Disponible en:

https://books.google.com.pe/books?id=x3daDwAAQBAJ&dq=libro+estad%C3%ADstico+multivariado+pdf&lr=&hl=es&source=gbs_navlinks_s

NAVARRO, Alex. Definición de modelo de proyección de ventas comerciales. Ecuador: Los Ruiseñores, 2015. ISBN: 848-682-7474-12-5

PACHECO, Carlos. Presupuestos un enfoque general. México: México D.F., 2016. ISBN: 978-607-8463-57-2.

PANDO RISH, Y. Facultades del entorno SQL Server 2008. Perú: Macro E.I.R.L, 2017. ISBN: 978-612-4034-02-2.

PERIER, José. Aprovisionamiento y almacenaje en la venta. España: Vigo, 2016. ISBN: 978-84-9839-574-7.

PR NEWSWIRE. SageQuest Named NCTC Exclusive Provider for GPS Fleet Tracking and Business Intelligence. Serie de documentos exclusivos para el seguimiento de la Inteligencia de negocios [en línea]. USA: Boston [Consulta: 17 de septiembre del 2019] Disponible en: <https://www.prnewswire.com/news-releases/sagequest-named-nctc-exclusive-provider-for-gps-fleet-tracking-and-business-intelligence-191068081.html>

RAJAN, Suri. La producción es cuestión de tiempo: La ventaja competitiva de la Fabricación de Respuesta Rápida (QRM), España: Barcelona, 2014. ISBN: 978-849-414-067-9

ROSAS, J. & CORTES, E. Propuesta de una metodología de planeación de la demanda y de los inventarios de medicamentos y dispositivos médicos de uso en pacientes hospitalizados en una IPS de cuarto nivel. Facultad de Ingeniería, Universidad ICESI, Colombia: Cali, 2016. ISBN: 928-62-7534-2-8.

RYAN ADELL, E. 2013. Business intelligence changes the face of fleet management. Accountancy SA. Estados Unidos: Georgia [Consulta: 19 de

Julio de 2018] Disponible en:
<https://search.proquest.com/docview/1458788236?accountid=37408>

SALAZAR, Carlos A., VALLEJO, Cristian y MONTILLA, Omar. Presupuestos: Aplicación práctica en las organizaciones, México: CDMX, 2019. ISBN: 978-958-778-554-8

SALEH MOHAMS, N. 2014 Comprehensive frameworks for decision making support in medical equipment management. [en línea]. Facultad de Ingeniería de la Universidad de El Cairo Giza. Egipto: Giza [Consulta: 25 de octubre de 2018]. Disponible en:
<http://search.ebscohost.com/login.aspx?direct=true&db=b9h&AN=9096488&lang=es&site=ehost-live>

SALINI, G. & PÉREZ, P.D. Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales. Revista chilena de ingeniería. Chile: Santiago [Consulta: 15 de junio de 2019] Disponible en:
http://www.scielo.cl/scielo.php?pid=S071833052006000200012&script=sci_arttext&tlng=pt.

SYNTETOS, A. & ZIED, M. Forecasting of compound Erlang demand. Journal of the Operational Research Society, Vol. 66, pp. 2061-2074. 2015. Una Aplicación de Redes Neuronales. Turquía: Giza [Consulta: 25 de Febrero de 2019] Disponible en:
<https://search.proquest.com/docview/448408587?accountid=37408>

STORCH DE GRACIA, E. 2018. Lines temporals data mining. Alemania: Kreise [Consulta: 19 de Julio de 2018] Disponible en:
<https://search.proquest.com/docview/1458788236?accountid=37408>

SIFONES, M., LEÓN, Á., ALVIN, H., MORA, M. E., DÍAZ RODRÍGUEZ, M., & DÍAZ MORA, M. E. Introducción al análisis estadístico multivariado aplicado. Experiencia y casos en el Caribe colombiano. Universidad del Norte. Colombia: CSIC, 2016. [Consulta: 14 de enero de 2019]. Disponible en:

https://books.google.com.pe/books?id=x3daDwAAQBAJ&dq=libro+estad%C3%ADstico+multivariado+pdf&lr=&hl=es&source=gbs_navlinks_s

STUART, E. 2010. Administración de Ventas. Argentina: Buenos Aires [Consulta: 17 de Julio de 2019] Disponible en: <https://arequest.proquest.com/docview/14584233236?accountid=85243>

TORRES M, Virgilio. 2014. Gestión y proyección de empresas para Ventas, México: México D.F., 2014. ISBN: 978-607-438-858-9

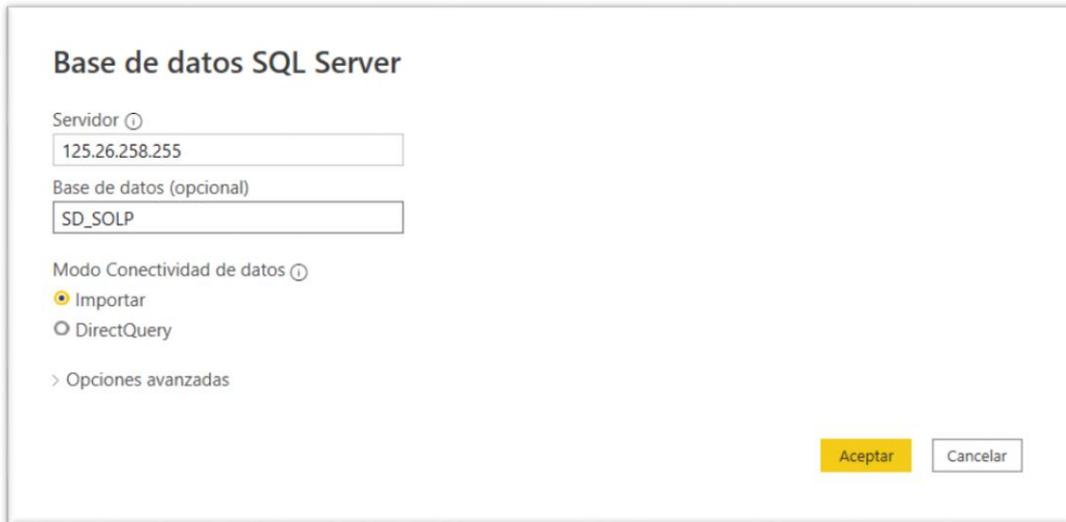
Matriz de consistencia

Tabla 6: Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE	DIMENSIONES	INDICADORES
General	General	General	Variable Dependiente		
¿De qué manera un sistema de pronóstico de ventas herramientas hidráulicas y electromecánicas con las variables Transacción por cliente, movimiento por artículo, transacciones en soles, Monto por facturas, Numero de asesores, ¿Cantidad de proveedores y cantidad de almacenes mejora la precisión del pronóstico de las ventas?	Determinar el efecto del uso de un sistema de pronóstico de ventas herramientas hidráulicas y electromecánicas con las variables Transacción por cliente, movimiento por artículo, transacciones en soles, Monto por facturas, Numero de asesores, Cantidad de proveedores y cantidad de almacenes en la precisión del pronóstico de las ventas	El uso de un sistema de pronóstico de ventas herramientas hidráulicas y electromecánicas con las variables Transacción por cliente, movimiento por artículo, transacciones en soles, Monto por facturas, Numero de asesores y Cantidad de proveedores y cantidad de almacenes mejora la precisión del pronóstico de las ventas	Ventas	Precisión del pronóstico	R2 ajustado (Precisión del modelo)
Específicos	Específicos	Específicos			
¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacción por cliente?	Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacción por cliente.	La variable transacción por cliente estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Transacción por cliente	Varianza explicada de las ventas con la variable transacción por cliente	Beta
¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable movimiento por artículo?	Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacciones en soles.	La variable movimiento por artículo estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Movimiento por artículo	Varianza explicada en las ventas con la variable movimiento por artículo	Beta

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE	DIMENSIONES	INDICADORES
Específicos	Específicos	Específicos			
¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacciones en soles?	Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable transacciones en soles.	La variable transacciones en soles estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Transacciones en soles	Varianza explicada en las ventas con la variable transacciones en soles	Beta
¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable monto por facturas?	Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable monto por facturas.	La variable monto por facturas estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Monto por facturas	Varianza explicada en las ventas con la variable monto por facturas	Beta
¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable número de asesores?	Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable número de asesores.	La variable número de asesores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Numero de asesores	Varianza explicada en las ventas con la variable número de asesores	Beta
¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de proveedores?	Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de proveedores.	La variable cantidad de proveedores estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Cantidad de proveedores	Varianza explicada en las ventas con la variable cantidad de proveedores	Beta
¿Cuál fue el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de almacenes?	Determinar el porcentaje de la variabilidad de las ventas de herramientas hidráulicas y electromecánicas con la variable cantidad de almacenes.	La variable cantidad de almacenes estuvo relacionada con las ventas de herramientas hidráulicas y electromecánicas.	Cantidad de almacenes	Varianza explicada en las ventas con la variable cantidad de almacenes	Beta

Conexión y Desarrollo de la ETL



Base de datos SQL Server

Servidor ⓘ
125.26.258.255

Base de datos (opcional)
SD_SOLP

Modo Conectividad de datos ⓘ
 Importar
 DirectQuery

> Opciones avanzadas

Aceptar Cancelar

Figura 1: Parámetros de conexión a la base de datos de las ventas

Nota: En los parámetros de conexión se muestran una ip pública y el nombre de la base de datos para la conexión directa al servidor e información que componen.

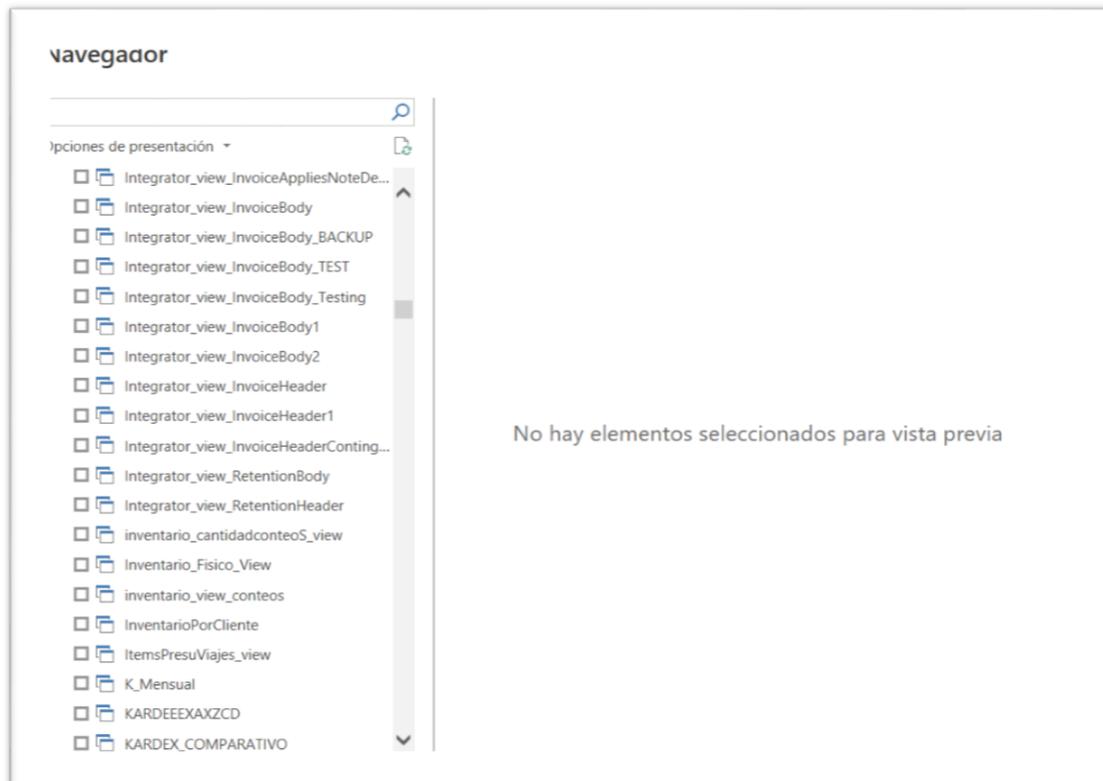


Figura 2: Conexión exitosa a las tablas de la base de datos de las ventas

Nota: Parámetros conformes y acceso a las tablas multidimensionales para realizar la carga de datos ETL

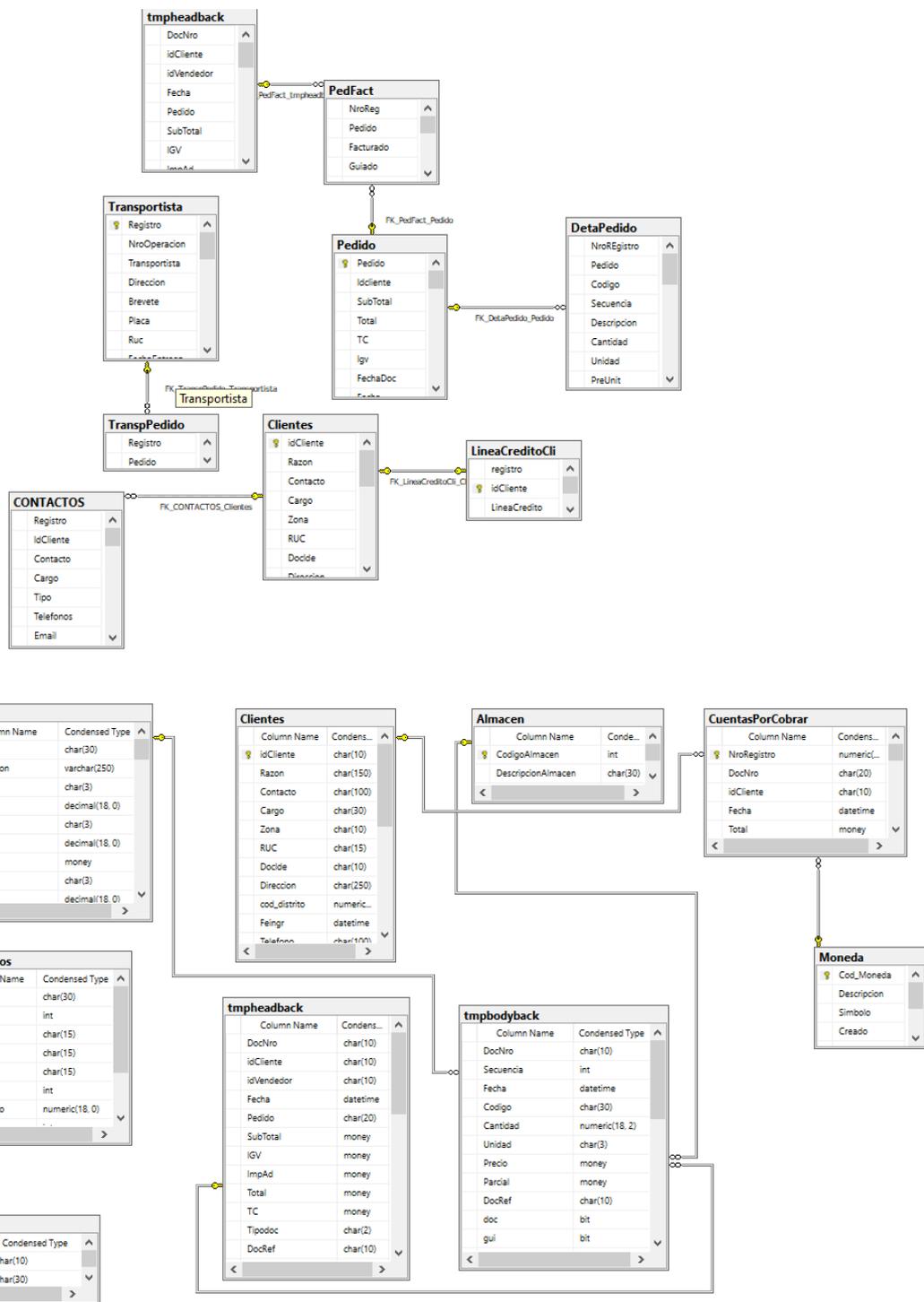


Figura 3: Comprensión de los datos – Diagrama base de datos relacional

Nota: Definición del diagrama físico de la empresa de consistencia SQL, estructura relacional proporcionada por la empresa.

A continuación, se explicará cada tabla del modelo relacional, además de sus atributos, la descripción de cada atributo y el tipo de dato.

Tabla 7: *Comprensión de los datos - Tabla transaccion_soles*

Columna	Tipo de dato	Descripción
ID_moneda	INT	Código de moneda
descripcion	VARCHAR	Nombre de la moneda

Nota: Esta tabla tendrá las diferentes monedas que la organización utiliza, su clave principal es ID_moneda. Esta tabla tiene un total de 3 registros.

Tabla 8: *Comprensión de los datos - Tabla cantidad_almacen*

Columna	Tipo de dato	Descripción
ID_almacen	INT	Código de almacén
direccion	VARCHAR	Dirección del almacén
estado	INT	Estado del almacén
descripcion	VARCHAR	Descripción del almacén

Nota: La clave principal es ID_almacen. Esta tabla tiene un total de 36 registros.

Tabla 9: *Comprensión de los datos - Tabla cantidad_proveedor*

Columna	Tipo de dato	Descripción
ID_proveedor	INT	Código del proveedor
descripcion	VARCHAR	Descripción proveedor
razon_soc	VARCHAR	Razón social
direccion	VARCHAR	Dirección del proveedor
contacto	VARCHAR	Contacto del proveedor
correo	VARCHAR	Correo del proveedor
telefono	VARCHAR	Teléfono del proveedor
tipo_actividad	INT	Tipo de actividad
tipo_proveedor	INT	Tipo de proveedor
ruc	INT	RUC del proveedor

Nota: La clave principal es ID_proveedor. Esta tabla tiene un total de 1487 registros.

Tabla 10: *Comprensión de los datos - Tabla monto_por_factura*

Columna	Tipo de dato	Descripción
ID_factra	INT	Código de factura
importe	DECIMAL	Importe facturado
descripcion	VARCHAR	Serie de la factura
ID_articulo	INT	Código del artículo

Nota: La clave principal es ID_factra. Esta tabla tiene un total de 1254876 registros.

Tabla 11: *Comprensión de los datos - Tabla numero_asesores*

Columna	Tipo de dato	Descripción
ID_asesor	INT	Código del asesor
nombre	VARCHAR	Nombre del asesor
direccion	VARCHAR	Dirección del asesor
fecha_ingreso	DATE	Fecha de ingreso
fecha_cese	DATE	Fecha cese
tipo_doc	INT	Tipo de documento
num_doc	INT	Numero de documento
telefono	VARCHAR	Teléfono de contacto
tipo_asesor	INT	Tipo de asesor

Nota: La clave principal es ID_asesor. Esta tabla tiene un total de 1487 registros.

Tabla 12: *Comprensión de los datos - Tabla movimiento_por_articulo*

Columna	Tipo de dato	Descripción
ID_producto	INT	Código del producto
categoria	INT	Categoría del producto
descripcion	VARCHAR	Serie de la factura
fecha_mov	DATE	Fecha de rotación
tipo_mov	INT	Tipo de movimiento
ID_almacen	INT	Código de almacén

Nota: La clave principal es ID_producto y su clave foránea es ID_almacen. Esta tabla tiene un total de 6354125 registros.

Tabla 13: *Comprensión de los datos - Tabla transaccion_por_cliente*

Columna	Tipo de dato	Descripción
ID_cliente	INT	Código del cliente
razon	VARCHAR	Razón social del cliente
direccion	VARCHAR	Dirección del cliente
contacto	VARCHAR	Contacto del cliente
correo	VARCHAR	Correo del cliente
ruc	INT	RUC del cliente
telefono	VARCHAR	Teléfono del cliente
actividad	VARCHAR	Actividad del cliente

Nota: La clave principal es ID_cliente. Esta tabla tiene un total de 4125635 registros.

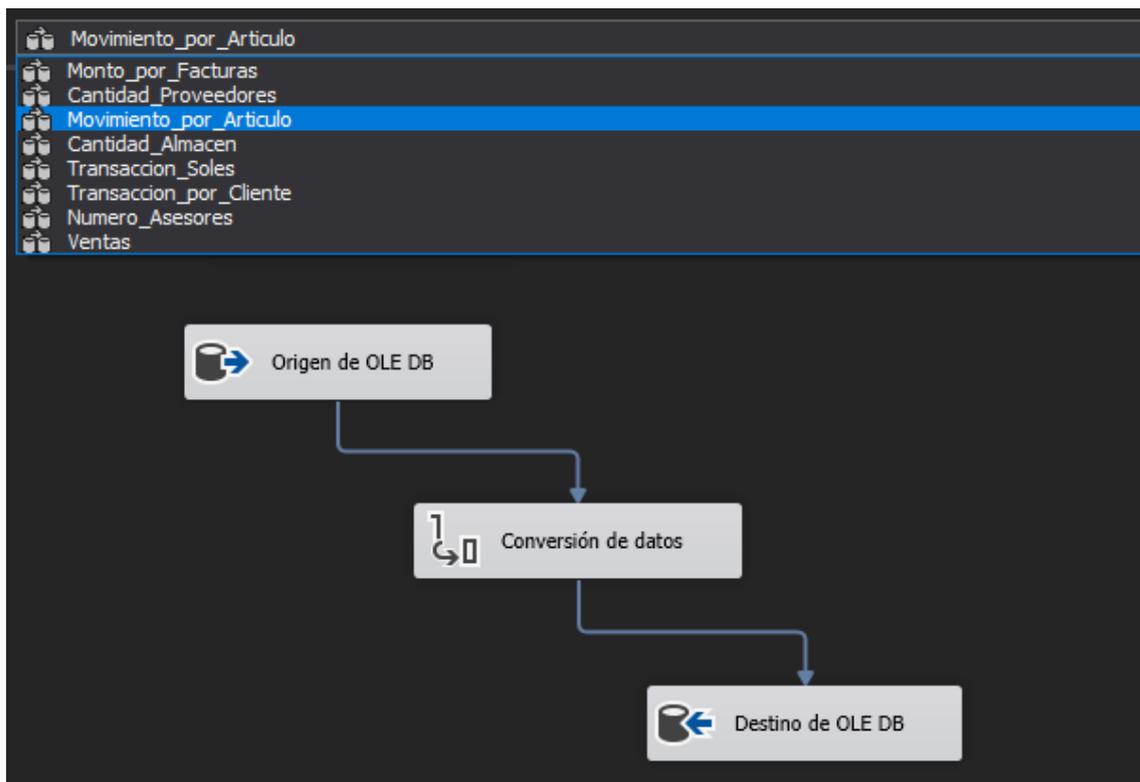


Figura 4: Modelo de tratamiento de datos del DWH

Nota: Proceso de conversión y estructuración del modelo estrella de la base de datos de las ventas, se necesitó crear un origen, un proceso de conversión y un destino. Dentro de los cuales se debe generar para cada dimensión.

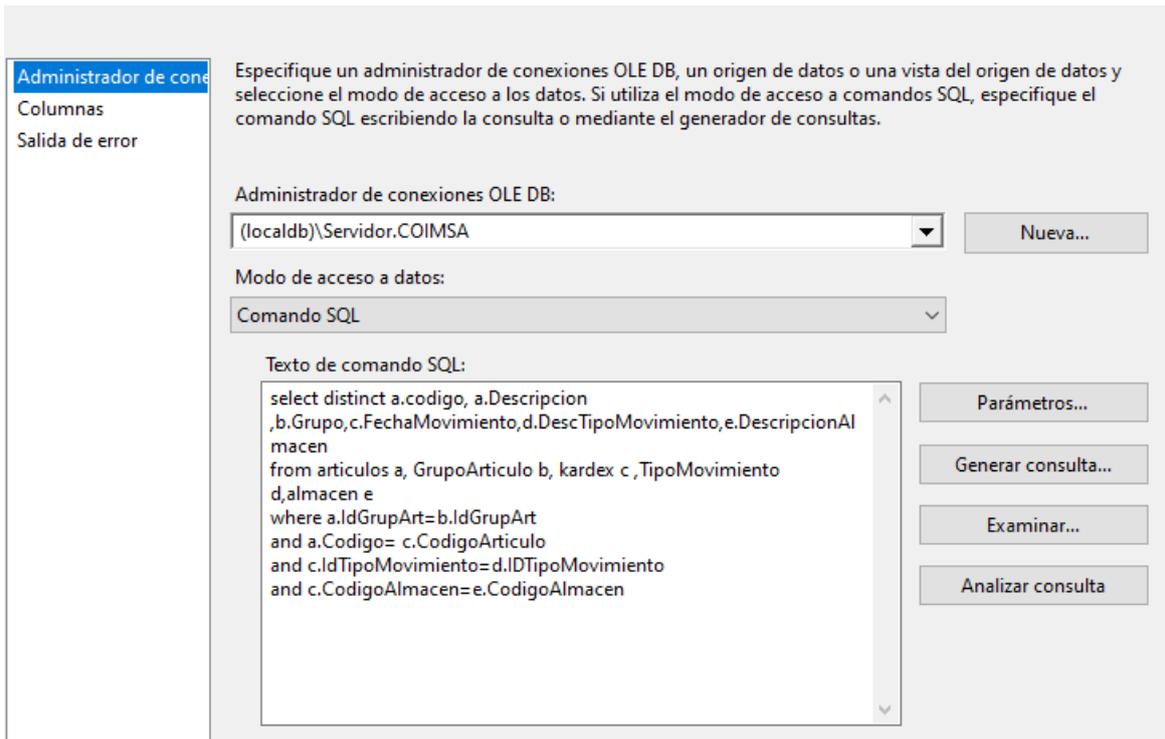


Figura 5: Inserción de información calculada

Nota: Se localiza el origen, en el cual se define la base de datos relacional, se genera una consulta alineada por dimensión y se procede a la sincronización.

Vista previa de los resultados de la consulta

Resultados de la consulta (hasta las primeras 200 filas):

codigo	Descripcion	Grupo	FechaMovi...	DescTipoM...	Desc
20922222...	22222 EK ...	RIPD ...	21/04/20...	RECEPCIÓN	ALM...
20922222...	22222 EK ...	RIPD ...	29/04/20...	VENTA	ALM...
20930214...	30214 J2/...	RIPD ...	21/04/20...	RECEPCIÓN	ALM...
20930214...	30214 J2/...	RIPD ...	29/04/20...	VENTA	ALM...
2093308 ...	3308 ATN...	RIPD ...	21/04/20...	RECEPCIÓN	ALM...
2093308 ...	3308 ATN...	RIPD ...	30/04/20...	VENTA	ALM...
2093310 ...	3310 A (R...	RIPD ...	05/05/20...	RECEPCIÓN	ALM...
2093310 ...	3310 A (R...	RIPD ...	15/07/20...	VENTA	ALM...
20951111...	51111 (RI...	RIPD ...	21/04/20...	RECEPCIÓN	ALM...
20951111...	51111 (RI...	RIPD ...	29/04/20...	VENTA	ALM...
20951207...	51207 (RI...	RIPD ...	21/04/20...	RECEPCIÓN	ALM...
20951207...	51207 (RI...	RIPD ...	29/04/20...	VENTA	ALM...
2096000...	6000-2RS	RIPD	21/04/20	RECEPCIÓN	ALM...

Cerrar

Figura 6: Muestreo de datos para inyección SQL

Nota: Se realiza la conversión de datos, introduciéndolos en la tabla de la base de datos relacional, para posteriormente generar una copia por campo de cada dimensión, la cual se convertirá en un tipo de dato beneficioso para el proceso de transformación.

Columnas de entrada disponibles

- Nombre
- codigo
- Descripcion
- DescripcionAlmacen
- DescTipoMovimiento
- FechaMovimiento

Columna de entrada	Alias de salida	Tipo de datos	Longitud	Precisión	Escala	Página
codigo	Copia de codigo	cadena [DT_STR]	250			1252
Descripcion	Copia de Descripcion	cadena Unicode [DT_WS...]	50			
Grupo	Copia de Grupo	cadena Unicode [DT_WS...]	50			
FechaMovimiento	Copia de FechaMovi...	marca de tiempo de bas...				
DescTipoMovimiento	Copia de DescTipoMo...	cadena Unicode [DT_WS...]	50			
DescripcionAlmacen	Copia de Descripcion...	cadena Unicode [DT_WS...]	50			

Figura 7: Conversión de tipos de datos de las tablas de la base de datos de las ventas

Configure las propiedades para insertar datos en una base de datos relacional mediante un proveedor OLE DB.

Administrador de conexiones OLE DB: Especifique un administrador de conexiones OLE DB, un origen de datos o una vista del origen de datos y seleccione el modo de acceso a los datos. Si utiliza el modo de acceso a comandos SQL, especifique el comando SQL escribiendo la consulta o mediante el generador de consultas. Para obtener acceso a datos de carga rápida, configure las opciones de actualización de tablas.

Administrador de conexiones OLE DB: (localdb)\Servidor.Hechos Nueva...

Modo de acceso a datos: Carga rápida de tabla o vista

Nombre de la tabla o la vista: [dbo].[Producto] Nueva...

Mantener valores de identidad Bloqueo de tabla

Mantener valores NULL Comprobar restricciones

Filas por lote:

Tamaño máximo de confirmación de inserción:

Figura 8: Orientación y creación de servidor de base de datos

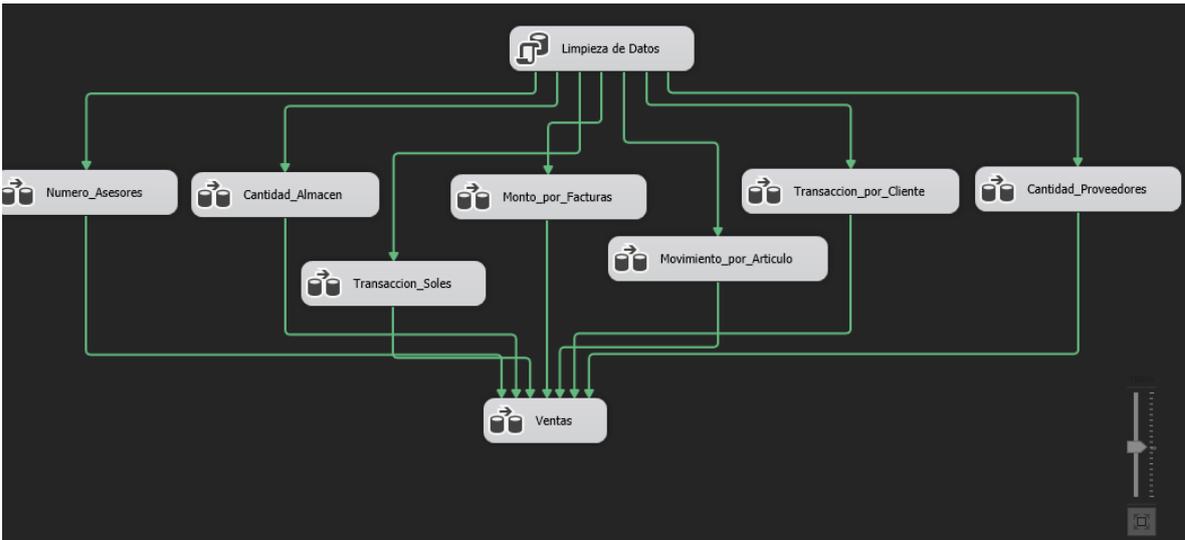


Figura 9: DataStage aplicado a las ventas

Nota: Proceso por dimensionamiento que tendrá un modelo enlazado de conversión de atributos e inserción de la información, sincronizado a la tabla de hechos, la cual proyectará la nueva y unificada información de las ventas. Se añadió un proceso de reinicio de información que se implementa con un borrado de datos cada vez se inicie el proceso el cual se puede observar en la parte superior de la figura.

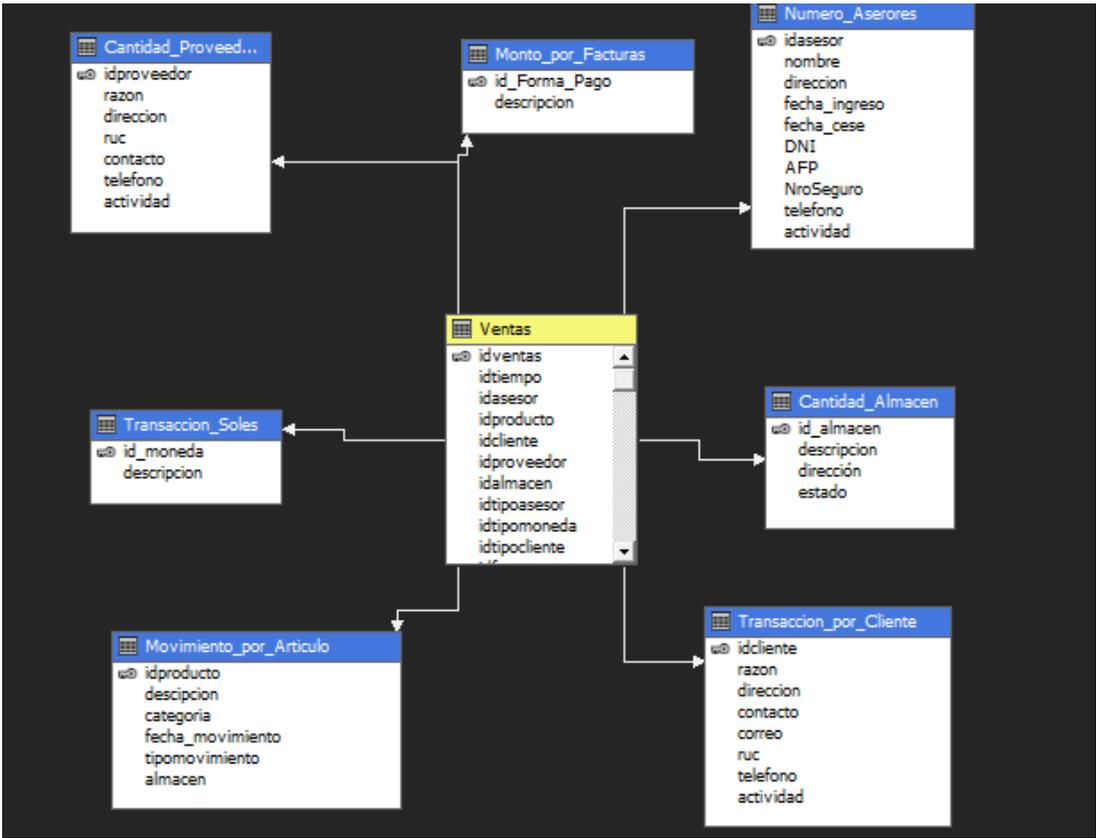


Figura 10: Cubo OLAP del modelo de ventas post ejecución de DataStage

Nota: Proceso de generación el cubo utilizando otra instancia en minería de datos, el cual parte de la tabla de hechos y genera las dimensiones dentro del cubo, que se observa dentro de la figura y sirve para la predicción continua.

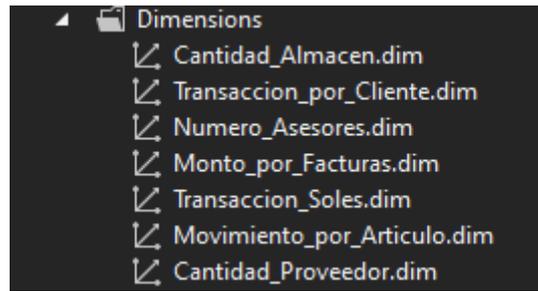


Figura 11: Definiciones de las dimensiones de las ventas

Nota: Proceso de generación de dimensiones de las ventas para integrar a la proyección mediante sus datos.

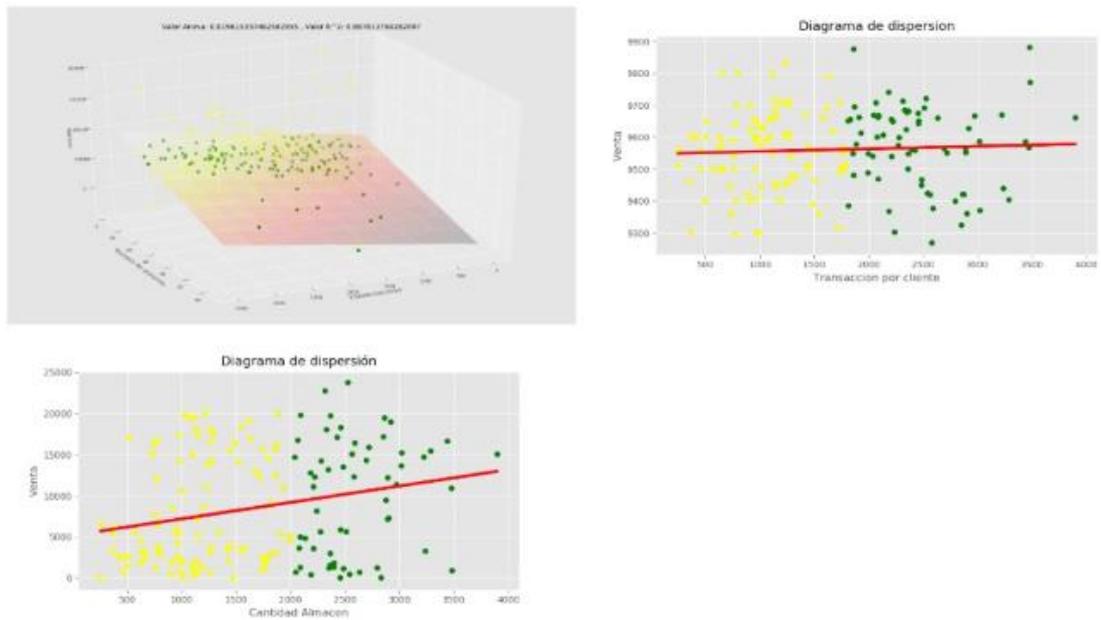


Figura 12: Diagrama de dispersión de dimensiones de las ventas

Nota: Resultados de las dimensiones en diagrama de dispersión de las ventas.

```

SELECT numerooperacion as Co_NumberOperation,
e.Co_TypeDocumentSUNAT as Co_TypeDocument,
--CASE e.Co_TypeDocumentSUNAT WHEN '01' THEN 'FJ' WHEN '03' THEN 'BJ' WHEN '08' THEN 'FD' ELSE 'FC' END
--+ SUBSTRING(a.DOCNRO, 2, 2) AS Tx_Serie,
SUBSTRING(a.DOCNRO, 1, 4) AS Tx_Serie,
SUBSTRING(a.DocNro, 5,10) AS Tx_DocNumber,
a.Fecha as Fe_Invoice,
CASE WHEN SUBSTRING(RTRIM(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(a.Pedido, ' ', ''), '- ', ''), '.', ''), '/', ''), '*', '')),1,20) = ''
THEN '000000' ELSE
SUBSTRING(RTRIM(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(REPLACE(a.Pedido, ' ', ''), '- ', ''), '.', ''), '/', ''), '*', ''), '_ ', '')),1,20) END as Tx_PurchaseOrder,
b.IdCliente as Co_Customer,
CASE e.Co_TypeDocumentSUNAT
WHEN '03' THEN '1'
WHEN '07' THEN
CASE WHEN b.RUC IS NULL THEN '1' ELSE
CASE WHEN LEFT(b.RUC, 2) = '10' OR LEFT(b.RUC, 2) = '20' OR LEFT(b.RUC, 2) = '15' THEN '6' ELSE '4' END
END
ELSE
CASE WHEN LEFT(b.RUC, 2) = '10' OR LEFT(b.RUC, 2) = '20' OR LEFT(b.RUC, 2) = '15' THEN '6' ELSE '4'
END
END AS Co_TypeDocumentIdentity,
b.RUC AS Tx_RUC, B.DocIde AS Tx_DNI, B.Razon AS Tx_BusinessName, A.Direccion AS Tx_InvoiceAddress,
a.Observaciones + ' - ' + c.Detalle as Tx_Observation,
ROUND(A.SubTotal,4) as Ss_SubTotal,
ROUND(a.IGV,4) as Ss_IGV,
ROUND(A.Total,4) as Ss_TotalBill,
f.Co_TypeOfCurrencySUNAT as Co_TypeOfCurrency,
a.TC as Ss_ExchangeRate,
g.Co_WayToPaySUNAT as Co_WayToPay,
replace(replace(replace(replace(c.Detalle, 'CHEQUE', 'CREDITO'), 'FACTURA', 'CREDITO'), 'FACTORING', 'CREDITO'), 'LETRA', 'CREDITO') as Tx_WayToPay,
a.VENCIMIENTO as Fe_Expiration,
ELSE
CASE WHEN LEFT(b.RUC, 2) = '10' OR LEFT(b.RUC, 2) = '20' OR LEFT(b.RUC, 2) = '15' THEN '6' ELSE '4'
END
END AS Co_TypeDocumentIdentity,
b.RUC AS Tx_RUC, B.DocIde AS Tx_DNI, B.Razon AS Tx_BusinessName, A.Direccion AS Tx_InvoiceAddress,
a.Observaciones + ' - ' + c.Detalle as Tx_Observation,
ROUND(A.SubTotal,4) as Ss_SubTotal,
ROUND(a.IGV,4) as Ss_IGV,
ROUND(A.Total,4) as Ss_TotalBill,
f.Co_TypeOfCurrencySUNAT as Co_TypeOfCurrency,
a.TC as Ss_ExchangeRate,
g.Co_WayToPaySUNAT as Co_WayToPay,
replace(replace(replace(replace(c.Detalle, 'CHEQUE', 'CREDITO'), 'FACTURA', 'CREDITO'), 'FACTORING', 'CREDITO'), 'LETRA', 'CREDITO') as Tx_WayToPay,
a.VENCIMIENTO as Fe_Expiration,
'0' + SUBSTRING(a.DocRef, 1, 3) AS Tx_GuideSerie, SUBSTRING(a.DocRef, 4,10) AS Tx_GuideNumber,
RTRIM(d.Per_appt) + ' ' + RTRIM(d.Per_apmt) + ' ' + RTRIM(d.Per_nomb) as No_Seller,
f.Tx_DescriptionForTextAmount as Tx_DescriptionOfCurrencyForNumberInText,
b.Email as Tx_EmailOfClient,
dbo.Integrator_function_isTTG(numerooperacion) as Fl_IsTTG,
dbo.Integrator_function_getSubTotalTTG(numerooperacion) as Ss_SubTotalTTG,
dbo.Integrator_function_IsBusinessForeing(b.IdCliente) as Fl_IsBusinessForeing,
a.MtoInafecto as Ss_UnaffectedAmount,
CASE Almacen WHEN 18 THEN 1 ELSE 0 END AS Fl_IsAffectoToDetraction,
h.CodEstableimiento as Co_BranchOffice,
convert(varchar, fechadoc, 108) as Tx_Hour
, CASE WHEN c.IdTipo = 39 THEN '1' ELSE '2' END Co_WayToPaySUNAT,
k.agenteretencion AS Fl_RetentionAgent
,
CASE WHEN a.Observaciones like '000%' THEN 1 ELSE 0 END Fl_WithoutDocument
FROM tmpheadback a JOIN Clientes b ON a.idCliente = b.idCliente
JOIN FormaDePago c ON a.IdTipo = c.IdTipo
JOIN CliOtroDat k ON k.idCliente = a.idCliente
JOIN Personal d ON a.IdVendedor = d.Per_Codi
JOIN Integrator_TypeDocument e ON a.Tipodoc = e.Co_TypeDocumentSEDISA
JOIN Integrator_TypeOfCurrency f ON a.Cod_Moneda = f.Co_TypeOfCurrencySEDISA
JOIN Integrator_relationWayToPaySEDISAAndSUNAT g ON c.IdTipo = g.Co_WayToPaySEDISA

```

Figura 13: Consulta integradora en la base de datos de las dimensiones independientes

Nota: Consulta para la transferencia de información en la tabla de hechos.

```

SELECT a.NumeroOperacion as Co_NumberOperation ,MIN(a.Secuencia) as Nu_Sequence, SUM(a.Cantidad) as Qt_InvoiceBody,
c.Co_UnitOfMeasurementSUNAT as Co_UnitOfMeasurement,
a.Precio as Ss_Price,
RTRIM(a.Codigo) + ' - ' + RTRIM(b.Descripcion) as Tx_DescriptProduct,
b.Unidad1 as Tx_Unity,
--a.Precio * 0.18 as Ss_IGV,
CASE WHEN b.Afecto = 1 THEN a.Precio * 0.18 ELSE 0.00 END as Ss_IGV,
CASE WHEN b.Afecto = 1 THEN 1 ELSE 0 END as Fl_Affected,
'01' as Co_TypeOfAffectationIGV,
'02' as Co_TypeOfAffectationIGVNotOnerous,
a.PorcDesc as Nu_DiscountRate,

CASE WHEN a.Precio = 0
THEN
--dbo.Integrator_getPriceTTG(a.DocRef, a.Codigo)
CASE WHEN dbo.Integrator_getPriceTTG(a.DocRef, a.Codigo) < 5
THEN 10 ELSE dbo.Integrator_getPriceTTG(a.DocRef, a.Codigo) END
ELSE 0
END as Ss_PriceTTG ,

CASE WHEN d.CODIGO_ANT IS NULL
THEN '31201501'
ELSE
CASE RTRIM(d.CODIGO_ANT) WHEN '' THEN '31201501' ELSE d.CODIGO_ANT END
END AS Co_GS1,

'GS1-129' AS Co_GTIN
FROM tmpbodyback a JOIN Articulos b ON a.Codigo = b.Codigo
LEFT JOIN Integrator_UnitOfMeasurement c ON b.Unidad1 = c.Co_UnitOfMeasurementSEDISA
LEFT JOIN DatAlterArt d ON b.Codigo = d.Codigo
GROUP BY a.NumeroOperacion,
a.DocRef,
a.codigo,
c.Co_UnitOfMeasurementSUNAT, a.Precio,
RTRIM(a.Codigo) + ' - ' + RTRIM(b.Descripcion),
b.Unidad1,

CASE WHEN a.Precio = 0
THEN
--dbo.Integrator_getPriceTTG(a.DocRef, a.Codigo)
CASE WHEN dbo.Integrator_getPriceTTG(a.DocRef, a.Codigo) < 5
THEN 10 ELSE dbo.Integrator_getPriceTTG(a.DocRef, a.Codigo) END
ELSE 0
END as Ss_PriceTTG ,

CASE WHEN d.CODIGO_ANT IS NULL
THEN '31201501'
ELSE
CASE RTRIM(d.CODIGO_ANT) WHEN '' THEN '31201501' ELSE d.CODIGO_ANT END
END AS Co_GS1,

'GS1-129' AS Co_GTIN
FROM tmpbodyback a JOIN Articulos b ON a.Codigo = b.Codigo
LEFT JOIN Integrator_UnitOfMeasurement c ON b.Unidad1 = c.Co_UnitOfMeasurementSEDISA
LEFT JOIN DatAlterArt d ON b.Codigo = d.Codigo
GROUP BY a.NumeroOperacion,
a.DocRef,
a.codigo,
c.Co_UnitOfMeasurementSUNAT, a.Precio,
RTRIM(a.Codigo) + ' - ' + RTRIM(b.Descripcion),
b.Unidad1,
a.Precio * 0.18,
a.PorcDesc,
B.Afecto,
CASE WHEN a.Precio = 0 THEN dbo.Integrator_getPriceTTG(a.DocRef, a.Codigo) ELSE 0 END,
CASE WHEN d.CODIGO_ANT IS NULL
THEN '31201501'
ELSE
CASE RTRIM(d.CODIGO_ANT) WHEN '' THEN '31201501' ELSE d.CODIGO_ANT END
END

```

Figura 14: Consulta integradora en la base de datos de la dimensión dependiente

Nota: Consulta para la transferencia de información en la tabla de hechos.

Desarrollo de la aplicación

DESARROLLO DE LA APLICACIÓN

En la presente investigación se realizó el proceso de minería de datos, procesos ETL (DataStage) y subconjuntos de datos correctamente almacenados (Data Mart)

1. Inventario de recursos:

En esta sección, se mostrará la situación actual de los recursos con lo cual se cuenta para la realización del proceso de minería de datos.

- **Información:** Se cuenta con la base de datos de ventas de la empresa PROBOTIX S.A.C. (empresa con nombre ficticio para reserva de información), el cual contiene información detallada desde el año 2012 a 2019.
- **Hardware:** El recurso de hardware disponible es 1 computadora portátil con las siguientes especificaciones:

Tabla 14: *Comprensión del desarrollo – Inventario de recurso de hardware*

Especificación	Equipo Portátil
Marca	HP
Modelo	PAVILION
Procesador	AMD Ryzen 7 @ 2.0GHz
RAM	16GB
Almacenamiento	512GB SSD
Tarjeta Gráfica	AMD Radeon™ Graphics
Sistema Operativo	Windows 10 Pro (x64)

- **Software:** Respecto a los recursos de software, herramienta de minería de datos y gestión de base de datos (RDBMS) disponibles debido a que el origen de la fuente de información fue provisto por la empresa por lo que tuvo que importarse la información para poder realizar una mejor manipulación. Además de usar un lenguaje de programación para definir

los algoritmos del modelo estadístico, una herramienta para la ejecución de los algoritmos en formato visual (diagramas) y finalmente una herramienta para la depuración de la información. Estos programas tienen las siguientes características:

Tabla 15: *Comprensión del desarrollo – Inventario de recurso de software*

Especificación	RDBMS	DM	LP
Nombre	SQL Server	Power BI - Analytics	Python
Versión	2019 Express	V7.0.2	3.6.8
Desarrollado en	C++	C#	-
Desarrollado por	Microsoft	Microsoft	Python Software Foundation
Sistema Operativo	Windows 10 Pro (x64)	Windows 10 Pro (x64)	Windows 10 Pro (x64)

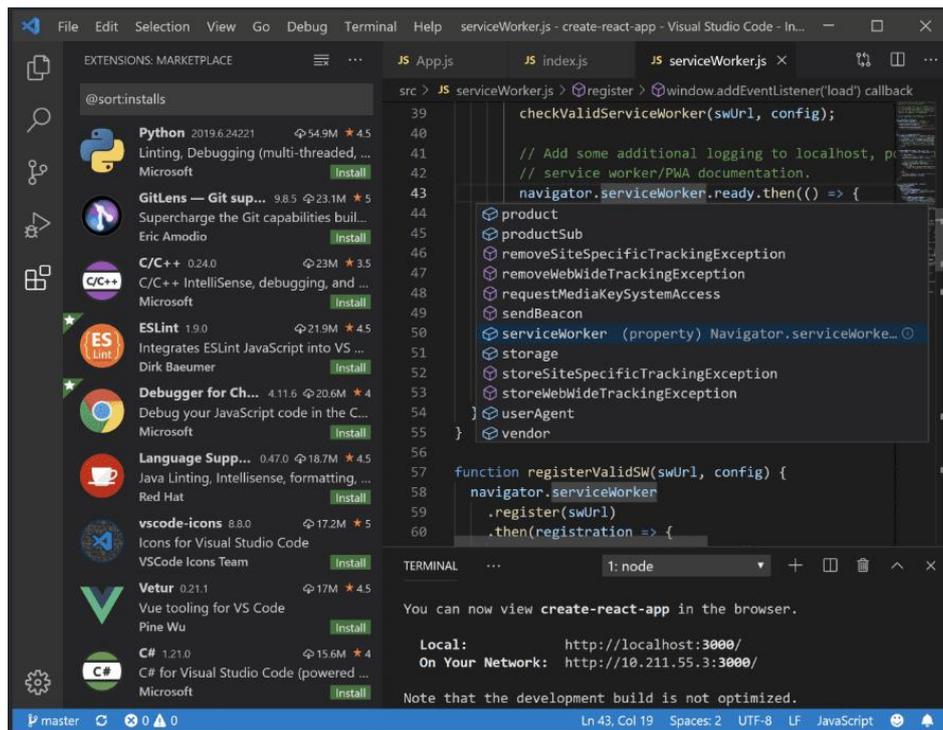


Figura 15: *Comprensión del desarrollo – Plugins Analytics y Power BI*

2. Construyendo la aplicación:

En esta tarea, debe de codificarse en el lenguaje de programación el conjunto de algoritmos para el cálculo de los valores de las dimensiones y los diagramas

(conjunto necesario para la construcción de un modelo), de los cuales se utilizaron las siguientes librerías:

- **Pandas:** Proporciona herramientas de análisis y manipulación de datos de alto rendimiento utilizando sus potentes estructuras de datos.
- **DataFrame:** Son estructuras de datos similares a las tablas de bases de datos relacionales como SQL.
- **Series:** Arrays unidimensionales con indexación (arrays con índice o etiquetados), similar a los diccionarios. Pueden generarse a partir de diccionarios o de listas
- **Panel, Panel4D y PanelND:** Estas estructuras de datos permiten trabajar con más de dos dimensiones, dado que es algo complejo y poco utilizado trabajar con arrays de más de dos dimensiones.
- **Numpy:** Utilizada para el cálculo de vectores y matrices con funciones matemáticas de alto nivel.
- **Shape:** Devuelve la dimensión del array, es decir, una tupla de enteros indicando el tamaño del array en cada dimensión. Para una matriz de **N** filas y **M** columnas obtendremos (n, m).
- **Data:** Buffer de información que contiene los elementos actuales del array.
- **Linspace:** Crea un array con valor inicial **start**, valor final **stop** y diferentes elementos.
- **Meshgrid:** Genera un plano a partir de dos los arrays x, y.
- **Matplotlib:** Utilizada generar gráficas a partir de datos contenidos en listas, vectores o extensiones matemáticas.
- **YLabel:** Utilizado para colocar un nombre a la recta Y.
- **Show:** Utilizado para mostrar los gráficos de acuerdo a la data en memoria.
- **XLabel:** Utilizado para colocar un nombre a la recta X.
- **Plot:** Utilizado para generar el grafico del plano.
- **Figure:** Utilizado para seleccionar tipo de grafico a mostrar.

- **Seaborn:** Permite la visualización de datos además de una interfaz para hacer gráficos estadísticos atractivos y explicativos cuyo objetivo es visualizar datos complejos de forma sencilla y extraer conclusiones.

El modelo estadístico se sincroniza con la base de datos y el modelo estrella depurado, para poder recuperar la información procesada para posteriormente realizar lo siguiente:

- Generación de tabla de coeficientes.
- Generación de tabla ANOVA.
- Generación de diagrama de dispersión con información sin discriminar los datos atípicos.
- Generación de diagrama de dispersión con la data discriminando los datos atípicos.
- Generación de influenciadores y segmentos clave.
- Generación de diagrama contemplando vectores **XYZ**.

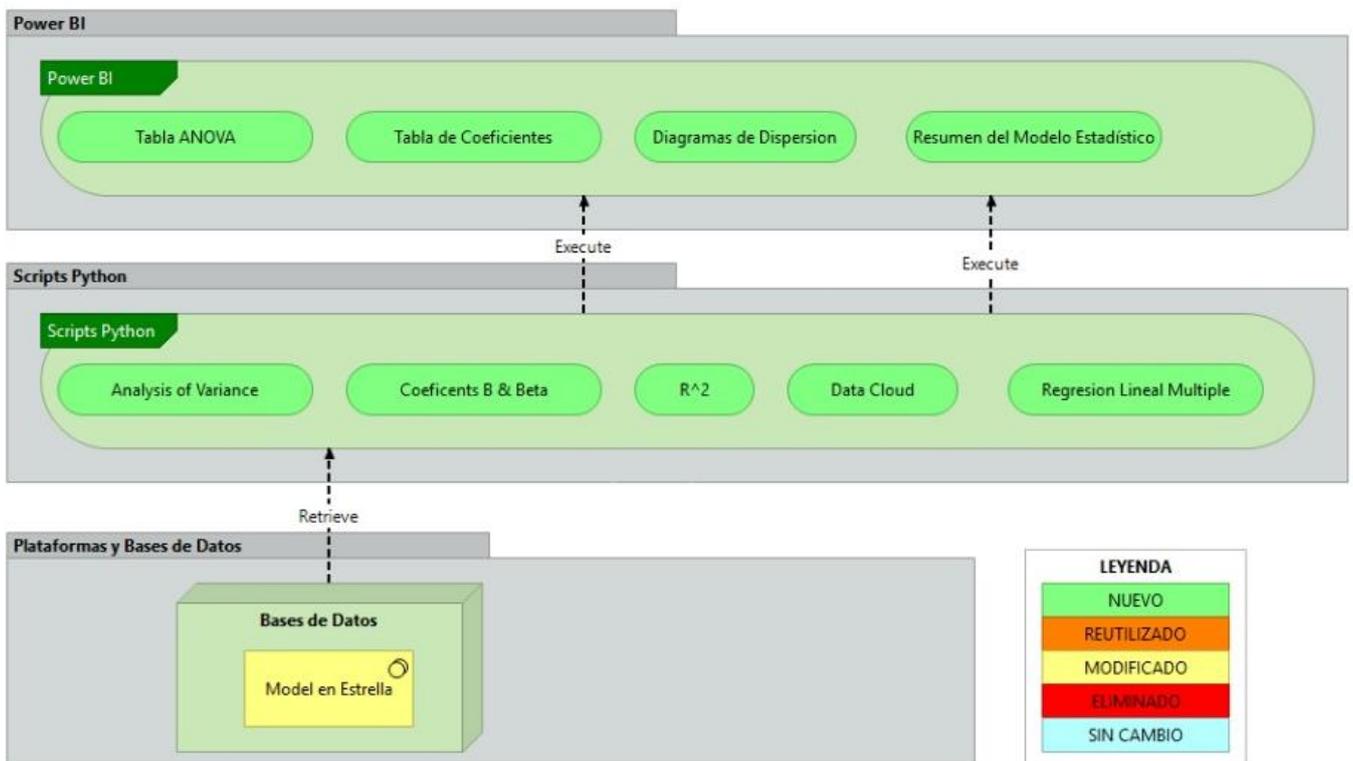


Figura 16: Comprensión del desarrollo – Arquitectura del modelo de proyección de ventas

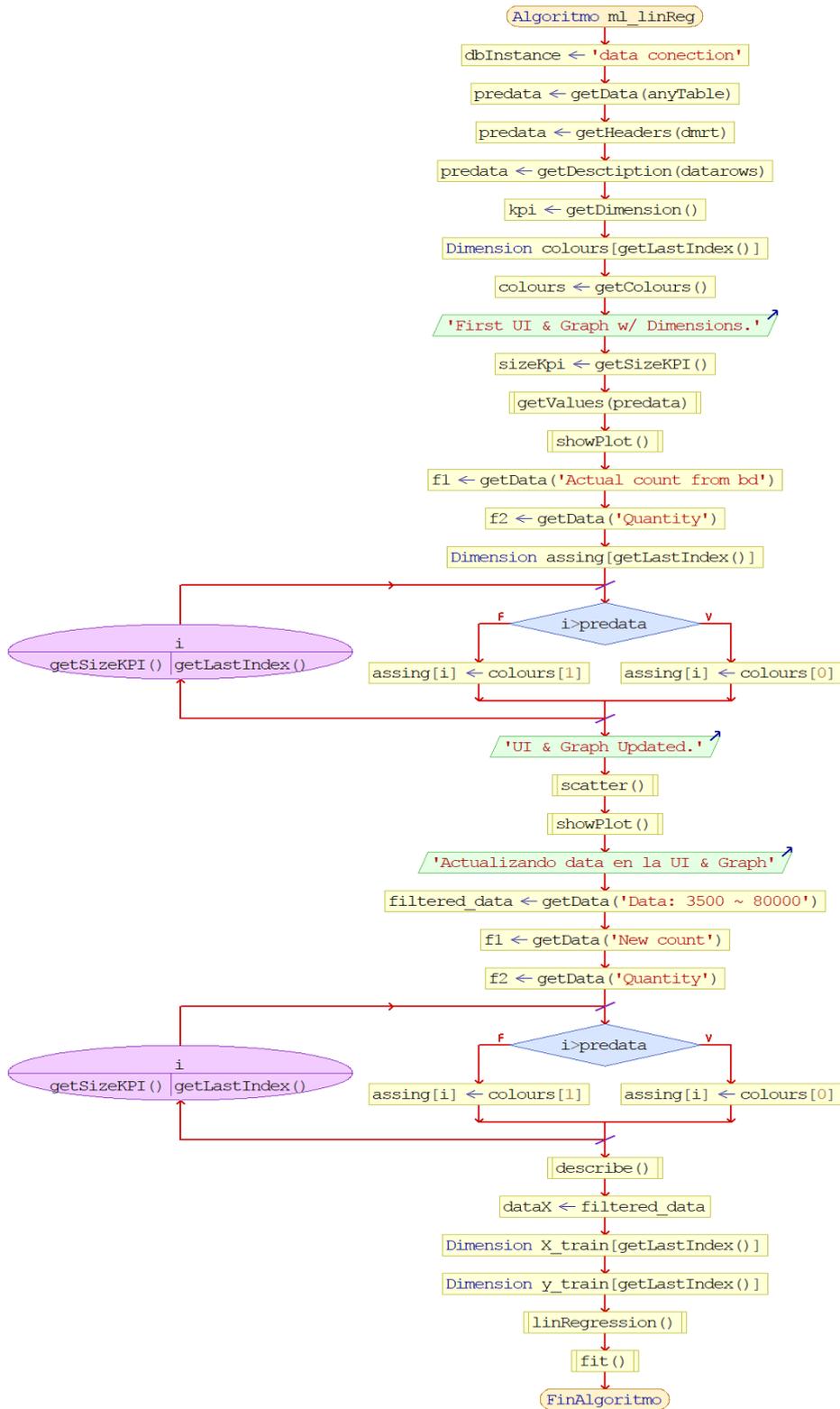


Figura 17: Comprensión del desarrollo – Diagrama de pseudocódigo del algoritmo de proyección de ventas

Ficha de observación

FICHA DE OBSERVACIÓN

Nº DE FICHA DE OBSERVACIÓN:	1
Observador:	Juan Carlos Híjar Samán
Institución donde se investiga:	Probotix S.A.C
Ubicación de la institución:	
Indicador observado:	Porcentaje del aumento de la precisión del modelo predictivo
Periodo de la observación:	2019

Variable	Indicador	Descripción	Técnica	Medida	Instrumento	Fórmula
Predicción de la gestión de Ventas	Porcentaje del aumento de la precisión del modelo predictivo	Calcula el porcentaje de aumento en la precisión del modelo predictivo	Observación	Porcentaje	Ficha de Observación	$PP = RP - \frac{VA}{100} \%$

Nº Registros	Fecha	Margen Real de precisión	Valor absoluto	Porcentaje de precisión del modelo predictivo de ventas
1	Enero	3%	100%	97%
2	Febrero	2%	100%	98%
3	Marzo	4%	100%	96%
4	Abril	2%	100%	98%
5	Mayo	1%	100%	99%
6	Junio	3%	100%	93%
7	Julio	5%	100%	95%
8	Agosto	4%	100%	96%
9	Septiembre	2%	100%	98%
10	Octubre	3%	100%	97%
11	Noviembre	1%	100%	99%
12	Diciembre	1%	100%	99%

DECLARATORIA DE AUTENTICIDAD DEL ASESOR

Yo, EMIGDIO ANTONIO ALFARO PAREDES, docente de la Facultad de Ingeniería y Arquitectura y Escuela Profesional de Ingeniería de Sistemas de la Universidad César Vallejo Campus Lima Este, revisor de la tesis titulada: "Sistema de pronóstico de ventas de herramientas hidráulicas y electromecánicas" de los estudiantes Híjar Samán Juan Carlos y Ortiz Salazar Gianfranco Porfirio, constato que la investigación tiene un índice de similitud de 16% verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y he concluido que cada una de las coincidencias detectadas no constituyen plagio. En tal sentido asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

24 de enero de 2023



.....
DR. EMIGDIO ANTONIO ALFARO PAREDES

DNI: 10288238