



**Estudio de algoritmos y técnicas de minería de datos para descubrir patrones que influyen en el comportamiento del abastecimiento de productos en la industria retail.**

Caiza Barrera, Mónica Gabriela

Vicerrectorado de investigación, innovación y transferencia de tecnología

Centro de postgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magíster en Gestión de Sistemas de Información e Inteligencia de Negocios

Ing. Vinuesa Trujillo, Jaime Augusto, MBA

28 de noviembre del 2022



**COPYLEAKS**

TESIS - CAIZA BARRERA MONICA GABRIELA v5 - revisión pl...

Scanned on: 2:57 December 19, 2022 UTC



Identical Words	252
Words with Minor Changes	154
Paraphrased Words	676
Omitted Words	206

Firma:

Ing. Vinueza Trujillo, Jaime Augusto, MBA  
Director



Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

### Certificación

Certifico que el trabajo de titulación “**Estudio de algoritmos y técnicas de minería de datos para descubrir patrones que influyen en el comportamiento del abastecimiento de productos en la industria retail**”, fue realizado por la Señora **Caiza Barrera Mónica Gabriela**, el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 28 de noviembre del 2022

Ing. Jaime Augusto Vinueza Trujillo, MBA

C.C.:1716028509

DIRECTOR



## Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

### Centro de Posgrados

#### Responsabilidad de Autoría

Yo, **Caiza Barrera Mónica Gabriela**, con cédula de identidad No. 1717535254, declaro que el contenido, ideas y criterios del trabajo de titulación: **"Estudio de algoritmos y técnicas de minería de datos para descubrir patrones que influyen en el comportamiento del abastecimiento de productos en la industria retail"** es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 28 de noviembre del 2022

Ing. Mónica Gabriela Caiza Barrera

C.C. 1717535254



Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

### Autorización de Publicación

Yo, **Caiza Barrera Mónica Gabriela**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Estudio de algoritmos y técnicas de minería de datos para descubrir patrones que influyen en el comportamiento del abastecimiento de productos en la industria retail”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 28 de noviembre del 2022

Ing. Mónica Gabriela Caiza Barrera

C.C. 1717535254



### **Dedicatoria**

En el camino de la vida, existen momentos difíciles e inolvidables en los cuales siempre conté con el amor incondicional de mi Padre Celestial, de quien nunca he soltado su mano y a quien dedico este trabajo.

A mi Auxiliadora, quien siempre ha sido mi auxilio, guiando mis pasos en el camino de la vida.

A mis padres Alfredo y Gladys, por ser mi apoyo incondicional, mi ejemplo de lucha constante ante cualquier adversidad.

A mi hermano Luis Alfredo, que con sus ocurrencias siempre me saca una sonrisa y hace de mis días más llevaderos.

A mi hermana Evelyn Carolina, mi ángel, que, aunque ya no está conmigo, siempre me apoyó y me acompañó en cada una de las etapas de mi vida

A mi esposo Diego, con su amor y paciencia me apoyó para culminar este trabajo con esfuerzo y dedicación.

A mi hijo Diego Julián, que con solo su presencia me motivó a culminar este trabajo.

Mónica Gabriela



## **Agradecimientos**

A mi Padre Celestial por sus bendiciones diarias para culminar con éxito esta formación académica.

A mis padres y hermanos, por su apoyo incondicional, por sus palabras de aliento y cariño para poder culminar con éxito la maestría.

A mi esposo y a mi hijo que han estado a mi lado, sacrificando tiempo para culminar este trabajo de titulación.

A mi director de tesis, Jaime Vinuesa, quien con su paciencia y enseñanzas ha guiado este trabajo de titulación con su aporte invaluable.

A mis compañeros de aula, con quienes disfrutamos de esta formación académica.

Mónica Gabriela



## Índice de Contenido

Dedicatoria.....	6
Agradecimientos .....	7
Resumen .....	15
Abstract.....	16
Capítulo 1 .....	17
Análisis del estado actual de la empresa.....	17
<i>Antecedentes</i> .....	17
<i>Planteamiento del problema</i> .....	18
<i>Objetivo del proyecto</i> .....	19
Objetivo General .....	19
Objetivos Especificos .....	19
<i>Justificación y Alcance del proyecto</i> .....	20
Preguntas de Investigación. ....	21
<i>Hipótesis</i> .....	22
Capítulo 2 .....	23
Marco referencial.....	23
<i>Categorización de variables</i> .....	23
Variables dependientes .....	23
Variables Independientes .....	23





<i>Marco Teórico</i> .....	24
Industrial Retail [6].....	25
Tasa de Descuento de un Producto .....	26
Gestión de Stock [13] .....	27
Stock Excesivo [7] .....	28
Pirámide de Valor [14].....	28
Minería de Datos [3] .....	31
Modelo de Minería de Datos [4].....	33
“Arquitectura del modelo de minería de datos” [4] .....	34
<i>Modelo de red neuronal [5]</i> .....	35
<i>Modelo de árbol de decisión [5]</i> .....	36
<i>Reglas de inducción [15]</i> .....	36
Metodologías para el proceso de Minería de Datos [16].....	37
<i>KDD</i> .....	37
<i>CRISP – DM (CRoss Industry Standard Process for Data Mining) [6]</i> .....	38
<i>SEMMA [16]</i> .....	41
<i>Revisión Preliminar de Literatura</i> .....	42
Criterios de Inclusión .....	42
Criterios de Exclusión .....	42
Grupo de Control.....	43
Palabras clave y coincidencias.....	47
Cadena de búsqueda y resultados obtenidos.....	47



Resultados obtenidos de estudio preliminar de literatura.....	53
<i>Metodología de investigación</i> .....	55
Definición de la situación actual .....	58
Estudio de la viabilidad de la solución .....	58
Construcción de la solución.....	58
Validación de la solución .....	59
Contribuciones .....	59
Capítulo 3 .....	60
Desarrollo de la solución .....	60
<i>Requerimientos del Negocio</i> .....	60
<i>Recolección y Exploración de Datos</i> .....	62
<i>Preparación de la data a utilizar</i> .....	63
<i>Métodos o técnicas de minería de datos</i> .....	83
Agrupación de 4 Clústers .....	84
Agrupación de 6 Clústeres .....	85
Agrupación de 8 Clústers .....	87
Capítulo 4 .....	93
Resultados obtenidos .....	93
<i>Patrones obtenidos</i> .....	93
<i>Análisis de los resultados</i> .....	94
Capítulo 5 .....	95



Conclusiones y recomendaciones .....	95
<i>Conclusiones</i> .....	95
<i>Recomendaciones</i> .....	96
Bibliografía .....	98



## Índice de Tablas

<b>Tabla 1</b> <i>Relación - Objetivos Específicos y Preguntas de Investigación</i> .....	21
<b>Tabla 2</b> <i>Grupo de Control [14]</i> .....	44
<b>Tabla 3</b> <i>Palabras Clave en Grupos de Control</i> .....	47
<b>Tabla 4</b> <i>Artículos Seleccionados</i> .....	48
<b>Tabla 5</b> <i>Top 10 de clientes</i> .....	67
<b>Tabla 6</b> <i>Top 10 de ID de producto de mayor y menor demanda</i> .....	72
<b>Tabla 7</b> <i>Top 10 de producto de mayor y menor demanda</i> .....	76
<b>Tabla 8</b> <i>Ganancia de producto de mayor y menor demanda</i> .....	92



## Índice de Figuras

<b>Figura 1</b> Espina de Pescado .....	19
<b>Figura 2</b> Categoría de variables Dependientes e Independientes.....	24
<b>Figura 3</b> Elementos de la pirámide de valor del consumidor [14].....	29
<b>Figura 4</b> “Modelo de minería de datos” [4].....	34
<b>Figura 5</b> Red Neuronal .....	35
<b>Figura 6</b> Árbol de decisiones [5] .....	36
<b>Figura 7</b> Metodología KDD [24].....	38
<b>Figura 8</b> Metodología CRISP – DM [6] .....	39
<b>Figura 9</b> Metodología SEMMA [16].....	41
<b>Figura 10</b> Diagrama de solución.....	62
<b>Figura 11</b> Información cargada a la Base de datos SQL .....	63
<b>Figura 12</b> Limpieza de datos .....	64
<b>Figura 13</b> Conteo de datos únicos.....	65
<b>Figura 14</b> Análisis de fechas de ventas .....	66
<b>Figura 15</b> Clientes únicos .....	67
<b>Figura 16</b> Concentración de clientes. ....	68
<b>Figura 17</b> Análisis del campo ciudad. ....	69
<b>Figura 18</b> Análisis del campo provincia. ....	70
<b>Figura 19</b> Análisis del campo sucursal. ....	71
<b>Figura 20</b> Productos Únicos .....	72
<b>Figura 21</b> Análisis del código del producto .....	73



<b>Figura 22</b> Análisis de la categoría del producto .....	74
<b>Figura 23</b> Análisis de la subcategoría del producto .....	75
<b>Figura 24</b> Análisis del nombre del producto.....	76
<b>Figura 25</b> Análisis del campo venta.....	78
<b>Figura 26</b> Análisis del campo cantidad .....	79
<b>Figura 27</b> Análisis del campo descuento .....	80
<b>Figura 28</b> Análisis del campo ganancia .....	81
<b>Figura 29</b> Análisis del campo COSTO_ENVIO.....	81
<b>Figura 30</b> Análisis del campo Prioridad .....	82
<b>Figura 31</b> Variables descartadas para el análisis .....	83
<b>Figura 32</b> 4 Clústers.....	84
<b>Figura 33</b> Tabla de Centroides de 4 Clústers .....	85
<b>Figura 34</b> 6 Clústers.....	85
<b>Figura 35</b> Tabla de Centroides de 6 Clústers .....	86
<b>Figura 36</b> 8 Clústers.....	87
<b>Figura 37</b> Tabla de Centroides de 8 Clústers .....	88



## Resumen

La empresa retail analizada se dedica a la venta de productos textiles en todo el territorio ecuatoriano, teniendo sus principales centros de atención en Guayaquil y Manta; la venta de los productos retail depende de la temporada, al no tener un correcto manejo del inventario de productos, podría ocasionar la pérdida de ventas si no existe suficiente, o en caso contrario, se deberá almacenar el inventario sobrante hasta la próxima temporada. La empresa presenta un alto índice de cobertura de stock considerando que el indicador para ventas al retail es de 2 a 3 meses, por lo que la empresa está generando un gasto excesivo en almacenamiento de dicho stock. La empresa tiene inconvenientes con el excesivo stock de la mercadería, lo que implica la existencia de productos retail que no rotan, altas cantidades de productos almacenados, mercadería obsoleta o fuera de temporada, esto lleva a crear promociones frecuentes para disminuir el excesivo stock. Este trabajo de titulación pretende dar recomendaciones al personal de la empresa para poder determinar patrones de comportamiento que influyen en el abastecimiento de la mercadería para disminuir el sobre stock, mediante herramientas de inteligencia de negocios como el rapidminer, motor de base de datos (SQL Server) y herramientas para realizar la extracción, transformación y carga (Talend Open Studio).

*Palabras Clave:* Retail, minería de datos, stock, análisis de datos.



### **Abstract**

The analyzed retail company is dedicated to the sale of textile products throughout the Ecuadorian territory, having its main service centers in Guayaquil and Manta; the sale of retail products depends on the season, by not having a correct management of the inventory of products, it could cause the loss of sales if there is not enough, or otherwise, the excess inventory must be stored until the next season. The retail company has a high stock coverage rate considering that the indicator for retail sales is 2 to 3 months, so the company is generating excessive spending on storage of said stock. The retail company has problems with the excessive stock of merchandise, which implies the existence of retail products that do not rotate, high amounts of stored products, obsolete or out-of-season merchandise, this leads to creating frequent promotions to reduce excessive stock. This titling work intends to give recommendations to the company's personnel in order to determine behavior patterns that influence the supply of merchandise to reduce overstock, through business intelligence tools such as rapidminer, a database engine (SQL Server ) and tools to perform extraction, transformation, and loading (Talend Open Studio).

*Keywords:* Retail, data mining, stock, data analysis.



## Capítulo 1

### **Análisis del estado actual de la empresa**

#### ***Antecedentes***

La empresa está en el país desde 1963, dedicada a la venta de productos textiles a nivel nacional, orientando sus esfuerzos a satisfacer las necesidades de sus clientes de acuerdo a su ubicación y estaciones climáticas principalmente.

Su oficina matriz se ubica en la ciudad de Quito, teniendo sus principales centros de atención en Guayaquil y Manta.

Al momento la empresa es el líder en las ventas al retail, obligándola a tener un stock variado y suficiente, orientado a satisfacer a los clientes con los más variados gustos por la moda, tomando en cuenta cantidad, calidad, diversidad y marcas de sus productos; generando un servicio de calidad y lealtad del cliente.

La mayor cantidad de venta de los productos retail, depende de la temporada (principalmente cambios de clima y festividades), al no tener un correcto manejo del inventario de productos, podría ocasionar la pérdida de ventas si no existe suficiente, o en caso contrario, se deberá almacenar el inventario sobrante hasta la próxima temporada.

Actualmente, la empresa presenta un alto índice de cobertura de stock considerando que el indicador para ventas al retail es de 2 a 3 meses, por lo que la empresa está generando un gasto excesivo en almacenamiento de dicho stock.

### ***Planteamiento del problema***

La empresa ecuatoriana de retail, ubicada su matriz en la ciudad de Quito, con sucursales en las principales ciudades del país, presenta problemas con el excesivo stock de la mercadería expuesta a sus clientes, lo que implica la existencia de productos retail que no rotan, altas cantidades de productos almacenados, mercadería obsoleta o fuera de temporada.

El problema del alto stock en la empresa lleva a crear promociones frecuentes para disminuir el excesivo stock, con rebajas en sus productos hasta del 70% del valor de venta, lo que implica mayor inversión de capital para la obtención de mercadería nueva o de temporada.

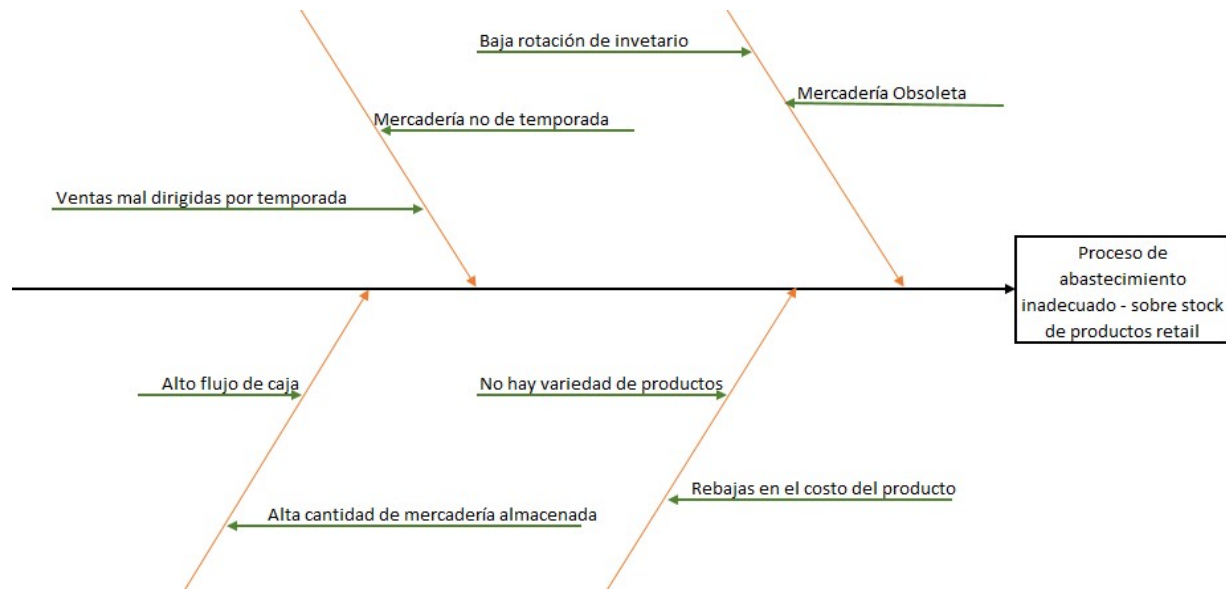
En la figura 1 se presenta la espina de pescado utilizada para la identificación del problema.

Los productos de retail presentan un gran problema con el proceso de abastecimiento debido, provocando el sobre stock de los productos retail.

Este problema ocasiona bajo nivel de rotación de inventario, provocando que su mercadería quede obsoleta y fuera de temporada, lo que provoca que sus ventas se encuentren mal dirigidas al público.

Con el sobre stock, se realizan promociones, lo que provoca rebaja en el costo del producto y una posible pérdida de ganancia.

Si los productos que se encuentran con sobre stock, no son vendidos con sus promociones, este es almacenado, lo cual induce a un alquiler de bodega ocasionando nuevos gastos y un alto flujo de caja.

**Figura 1****Espina de Pescado**

*Nota.* La figura detalla los problemas que se presentan en la empresa retail.

**Objetivo del proyecto**

En esta sección se presenta el objetivo general y los objetivos específicos que llevan a la solución del problema del alto stock en una industria retail ubicada en la ciudad de Quito.

**Objetivo General**

Diseñar un modelo de patrones de comportamiento que influyen en el abastecimiento de productos en la industria retail utilizando algoritmos y técnicas de minería de datos.

**Objetivos Específicos**

**OE1:** Realizar un estudio de la situación actual del proceso de abastecimiento de productos con el propósito de determinar las causas del elevado índice de cobertura de stock con respecto al índice normal de retail de moda.

**OE2:** Obtener información de estudios realizados acerca del manejo del stock, a través de la revisión preliminar de literatura para conocer técnicas de administración de la industria de productos retail

**OE3:** Recopilar y analizar la data histórica mediante técnicas de minería de datos para identificar los patrones de comportamiento para el abastecimiento de productos en la empresa.

**OE4:** Validar el modelo propuesto para mejorar el índice de cobertura del stock, realizando pruebas que permitan visualizar la confiabilidad de los resultados

### ***Justificación y Alcance del proyecto***

La empresa ecuatoriana, líder en la industria retail, realiza la adquisición de sus productos de forma masiva para la comercialización de los mismos en el país.

Esta empresa adquiere sus productos de 5 a 6 veces en el año para satisfacer las necesidades de sus clientes de acuerdo a la temporada.

La empresa ecuatoriana líder en retail, presenta un alto índice de cobertura, generando un gasto excesivo en el almacenamiento de los productos adquiridos. Estos productos que permanecen en bodega no tienen rotación de inventario, lo que implica pérdida de venta, por tanto, disminución del flujo de efectivo, reducción del precio del producto a través de descuentos o promociones.

Con la problemática planteada, es importante la realización de este proyecto con el que se pretende recomendar una solución al stock excesivo en la empresa, para lo cual es importante realizar el estudio del comportamiento de patrones que influyen en el abastecimiento de productos con mayor precisión y así mejorar el índice de cobertura que actualmente se presenta la empresa.

## Preguntas de Investigación.

De acuerdo a los objetivos específicos planteados, se tiene las preguntas de investigación:

**Tabla 1**

*Relación - Objetivos Específicos y Preguntas de Investigación*

Objetivo	Descripción del objetivo
OE1:	Realizar un estudio de la situación actual de la empresa del proceso de abastecimiento de productos con el propósito de determinar las causas del elevado índice de cobertura de stock con respecto al índice normal de retail de moda.
<b>OE1-RQ1</b>	¿Cuáles son las causas por las cuales las empresas de retail tiene un alto índice de cobertura de stock?
<b>OE1-RQ2</b>	¿Cuáles son los efectos del alto índice de cobertura de stock en la industrial retail?
OE2:	Obtener información de estudios realizados acerca del manejo del stock, a través de la revisión preliminar de literatura para conocer técnicas de administración de la industria de productos retail.
<b>OE2-RQ1</b>	¿Qué soluciones presentan los estudios consultados sobre el alto índice de cobertura de stock en la industria retail?
<b>OE2-RQ2</b>	¿Qué métodos utilizan para solucionar el alto índice de cobertura de stock en la industria retail?
OE3:	Recopilar y analizar la data histórica mediante técnicas de minería de datos para identificar los patrones de comportamiento para el abastecimiento de productos en la empresa.

Objetivo	Descripción del objetivo
<b>OE3-RQ1</b>	¿Qué modelo de minería de datos permite encontrar los patrones de comportamiento para reducir el índice de cobertura del stock en la empresa retail?
<b>OE3-RQ2</b>	¿Qué procesos de datos se utilizarán para el análisis de la data histórica proporcionada por la empresa?
OE4:	Validar el modelo propuesto de minería de datos para mejorar el índice de cobertura del stock, realizando pruebas que permitan visualizar la confiabilidad de los resultados
<b>OE4-RQ1</b>	¿La solución presentada mejora el índice de cobertura de stock?
<b>OE4-RQ2</b>	¿Con los patrones de comportamiento encontrados se puede mejorar la adquisición de los productos en stock?

Nota. Esta tabla detalla los objetivos específicos y las preguntas de investigación

### ***Hipótesis***

El abastecimiento de productos retail en la empresa, se puede mejorar utilizando patrones de comportamiento que permitan predecir eventos futuros a través de algoritmos y técnicas de minería de datos

## Capítulo 2

### Marco referencial

#### *Categorización de variables*

En el análisis de la hipótesis, se han determinado las variables dependientes e independientes, las cuales se detallan a continuación.

#### **Variables dependientes**

La industria retail, presenta problemas con el abastecimiento de productos debido al excesivo stock.

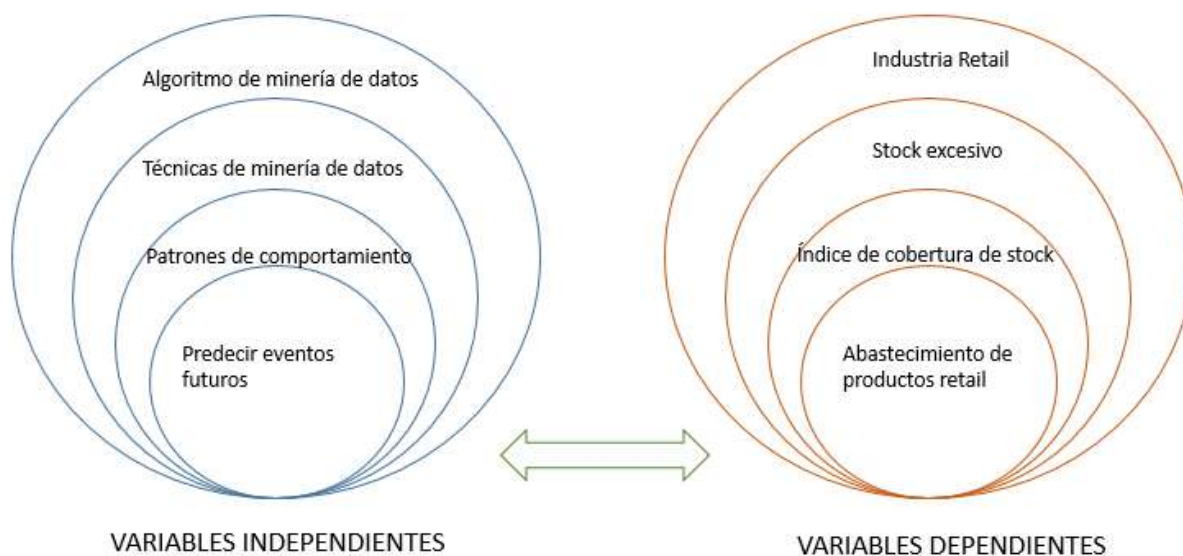
#### **Variables Independientes**

Implementar un modelo de predicción de eventos a través de algoritmos de minería de datos y descubrimientos de patrones de comportamiento que eviten el sobre stock de productos

En la figura 2, se puede visualizar la red de categorías de variables dependientes e independientes citadas anteriormente.

**Figura 2**

*Categoría de variables Dependientes e Independientes*



*Nota.* La figura detalla las variables dependientes e independientes del proyecto.

### **Marco Teórico**

Ecuador es uno de los mercados más atractivos para la implementación de la industrial Retail de moda, siendo así, varias marcas se encuentran en negociaciones de ingreso al país [1].

*“Hasta abril de este 2018, la actividad del sector comercial de Ecuador creció significativamente, pues solo en los dos primeros meses del año acumuló US\$20.703,5 millones en ventas, según el diario El Comercio de Ecuador” [1].*

La empresa reconocida como líder en Retail, se encuentra atendiendo en las principales ciudades del país como: Quito Guayaquil, Cuenca, Ibarra, Machala, Ambato, Manta, Loja, Esmeraldas [2].



Para la realización de este proyecto de titulación se presenta un plan de trabajo que permite identificar las actividades que se deben realizar en el tratamiento de los datos y poder identificar patrones de comportamiento para el abastecimiento de productos retail.

### **Industrial Retail [6]**

Retail es un término inglés utilizado para nombrar a la venta al por menor de productos al consumidor final; este sector económico comercializa de forma masiva productos y servicios a grandes cantidades de clientes. Las ventas de retail, pretenden satisfacer las necesidades de sus clientes ya sea por tiendas físicas o ventas online.

Para poder clasificar a la industria retail, se toma en cuenta los metros cuadrados, el surtido de sus productos y el grado de especialidad, precios, entre otras características, así tenemos [10]:

- Tiendas departamentales: Se caracteriza por ser una tienda de muchos metros cuadrados con diferentes categorías de productos separadas por departamentos.
- Especialistas: Se caracteriza por atender una categoría específica, dando mayor asesoramiento al cliente.
- Supermercados: Se caracterizan por la excelencia dentro del sector retail, practicando técnicas de marketing con excelentes resultados
- Tiendas de conveniencia o cercanía: Son aquellas tiendas que se encuentran cerca de sus clientes, siendo la comodidad y el tiempo las cosas más valoradas por los clientes.
- Tiendas de descuento: Se caracterizan por tener precios bajos y de calidad inferior.

- Hipermercados: Se caracterizan por su gran espacio de exhibición de productos, por lo general se encuentran alejados de la ciudad
- Category Killers: Se especializa en una categoría especial, pero con precios muy bajos, ocasionando que competidores queden fuera del mercado.

El sector de retail, está conformado por sectores y sub sectores sin distinción de tiendas físicas o ecommerce, citados a continuación [9]:

- Alimentación y Bebidas
- Moda
- Equipamiento del hogar
- Servicios (bancarios, telefonía, juegos, seguros, consultorías, transporte, turismo, salud, belleza, entre otros)
- Automotriz
- Farmacias

El retail en el Ecuador, es un sector económico estable que genera más de 16 800 millones anualmente, teniendo un crecimiento anual de 1.3% según datos de Euromonitor Internacional, firma de investigación de mercado con base en el Reino Unido. El segmento retail está evolucionando a un buen ritmo, permitiendo tener un crecimiento continuo, tomando en cuenta la evolución de la urbe en las distintas ciudades [11].

### **Tasa de Descuento de un Producto**

Los precios de un producto en un negocio de retail, es un factor determinante al momento de la compra, para ello se debe tener en cuenta la tasa de descuento que se va aplicar al producto sin perjudicar al comerciante [12].

La mayoría de las tiendas retail, utilizan la estrategia de precios de Keystone, el cual indica que se debe duplicar el precio del producto, generando una ganancia del 50%, aunque en diferentes casos, se deba vender a precios más altos o más bajos dependiendo del retail [12].

Los consumidores de la industrial retail, se interesan por los precios bajos utilizando cupones, descuentos, precios estacionales, promociones. Sin embargo, para la industria retail, es conveniente realizar descuentos que permitan aumentar la afluencia de personas a la tienda para reducir el inventario difícil de vender [12].

Para el cálculo de la tasa de descuento de un producto, se puede aplicar la siguiente fórmula:

$$\text{Precio de Venta} = [(\text{Costo del Producto}) \div (100 - \% \text{de recargo})] * 100$$

Las empresas de retail utilizan los descuentos para incrementar la venta de sus productos a bajos precios, pero en grandes cantidades. Esta estrategia, debe ser manejada cuidadosamente, ofreciendo descuentos ocasionales para la fidelidad de los clientes y evitando la disminución del margen de ganancia.

### **Gestión de Stock [13]**

La gestión de stock es una tarea importante en la industria retail, permitiendo dar seguimiento al inventario de variedad y volumen de productos. Este sistema permitirá:

- **Control de robo o pérdidas:** Facilita el control del inventario para detectar faltantes de productos,
- **Servicio al cliente:** Permite verificar productos en stock y localizarlos rápidamente.

- **Gestión financiera:** Permite dar seguimiento de funcionamiento del negocio sin esperar el estado de cuenta.
- **Seguimientos de productos:** Realiza el seguimiento de artículos y controla el stock para un mejor servicio al cliente.

### **Stock Excesivo [7]**

“En la actualidad las empresas ponen todos sus esfuerzos en ofrecer un mejor servicio a sus clientes, sin que esto repercuta necesariamente en un gasto excesivo. **La correcta gestión del stock almacén juega un papel decisivo** en la búsqueda de esta competitividad. Hacer un uso correcto de nuestro inventario nos permitirá fidelizar clientes, disponer de más tiempo para dedicarlo al desarrollo de nuestro negocio y evitar la pérdida de parte de nuestro presupuesto.”

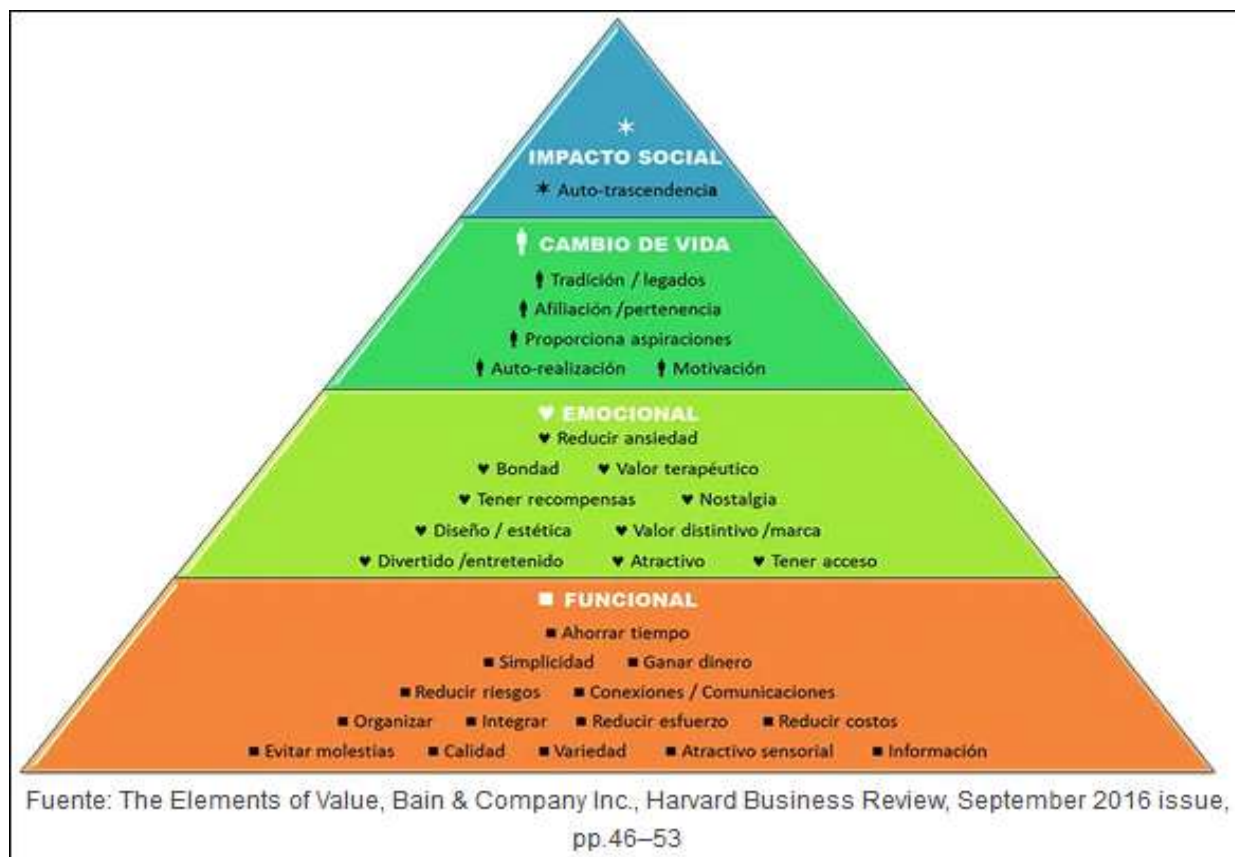
La mercadería retail debe ser controlada para poder brindar al usuario final, productos de calidad, para así mantener la fidelidad de sus clientes y que la misma se entregue en tiempo adecuado y en buenas condiciones.

### **Pirámide de Valor [14]**

En la figura 3, se identifica 30 valores que una empresa debe entregar al cliente y los clasifica por nivel según Maslow. En la parte superior se encuentran los valores más fuertes de la pirámide, sin embargo, una empresa para poder entregar valores fuertes necesita entregar algún valor débil de los niveles inferiores, estos valores varían según el tipo de producto o servicio.

Figura 3

Elementos de la pirámide de valor del consumidor



*Nota.* Esta figura detalla cada etapa de la pirámide de valor del consumidor. Tomado de <http://www.ebizlatam.com/la-creacion-valor-cliente-maslow-la-piramide-valores-del-consumidor/>

Los 30 valores principales de la pirámide, corresponde al resultado de la indagación de consumidores, clientes corporativos y de estudios analizados por los autores (*Eric Almquist, John Senior, Nicolas Bloch*). Esto permitió descubrir 3 patrones de creación de valor.

1. Los valores citados en la pirámide, tiene diferente valor según el negocio, sin embargo, el valor común que toda industria debe brindar al cliente es la calidad del producto, lo que es percibido por el usuario; los otros valores dependerán de la

industria. A continuación, se citan los 5 valores con más relevancia en la fidelidad de un cliente:

- a. "Calidad
- b. Tener acceso
- c. Tradición de familia
- d. Evitar molestias
- e. Reducir la ansiedad

En el retail de descuento, es más valorado por los clientes:

- a. Calidad
- b. Variedad
- c. Reducir costos
- d. Ahorrar tiempo
- e. Recibir recompensas

En el retail de vestimenta, los valores más apreciados son:

- a. Calidad
- b. Variedad
- c. Evitar molestias
- d. Diseño/estética
- e. Ahorro de tiempo" [14]

2. Los clientes exponen que las compañías digitales dan más valor, ya que evitan molestias y ahorran tiempo y la comunicación es más fácil y conveniente.
3. Las empresas físicas tienen valores que las empresas digitales no los tienen, ganando valor de marca, atractivo, afiliación y permanencia, obteniendo calificaciones más altas con respecto a las digitales

### **Minería de Datos [3]**

Este proceso trabaja con grandes volúmenes de datos y análisis matemáticos para descubrir patrones y tendencias de comportamiento.

La identificación de los diferentes patrones de comportamiento se la realiza a través de un proceso de minería de datos, ya que un análisis tradicional de datos, no permite detectar con certeza este comportamiento.

Mediante un análisis de datos, se descubren patrones y tendencias de comportamiento lo que define a un modelo de minería de datos, los cuales se pueden aplicar a escenarios como los siguientes:

- **Previsión:** Este escenario permite identificar ventas, cargas de servidor y tiempo de inactividad del mismo.
- **Riesgo y probabilidad:** Para estos escenarios, determina el punto de equilibrio, los mejores clientes y designa probabilidades de los resultados.
- **Recomendaciones:** Para generar recomendaciones, se debe identificar los productos que se pueden unificar
- **Buscar secuencias:** Predice las futuras compras, mediante la selección de productos realizados por el cliente.
- **Agrupación:** Predecir sus futuras compras mediante distribución de eventos y clientes comunes por clústers.

Para la realización del trabajo de minería de datos de la industria retail se tiene:

- Identificación el problema
- Indagación de los datos
- Corrección de los datos
- Discriminación de datos

- Minería de datos

**Identificación el problema:** identificar el problema y definir métodos de uso de los datos para proporcionar dar una respuesta.

Este ítem incluye:

- Identificar los requisitos del cliente
- Especificar el problema
- Proponer métricas con las que se calculará el modelo
- Puntualizar los objetivos del proyecto.

**Indagación de los Datos:** Permite unificar, consolidar y limpiar los datos

Para la limpieza de los datos se debe verificar información no válida y faltante, registros duplicados e identificar las correlaciones, origen de los datos y columnas que son necesarias para el análisis de la información.

**Exploración de datos:** Se debe verificar los datos limpios.

Para el modelo de minería de datos, se debe revisar los datos seleccionados para tomar decisiones acertadas.

**Modelos de Datos:** Permite crear el modelo de minería de datos.

El modelo de los datos utilizar columnas de datos para la entrada, atributos y parámetros que permiten implementar el algoritmo y procesar los datos.

El entrenamiento o procesamiento de un modelo hace referencia al algoritmo matemático aplicado a los datos para obtener patrones de comportamiento.

Los patrones de comportamiento dependen de los datos seleccionados, de los algoritmos y de su configuración.



**Validación y Exploración de modelos:** Se estudia los modelos de minería de datos que se ha creado y se verifica su eficacia.

Los datos de entrenamiento son utilizados para crear el modelo y los datos de prueba permiten comprobar la validez del modelo mediante consultas de predicción.

**Implementación de datos:** Consiste en ejecutar modelos en un entorno de producción.

Con los modelos certificados, se puede realizar diferentes tareas según la necesidad de la empresa:

- Los modelos de minería permiten crear predicciones y se usados en la toma de decisiones comerciales
- Utilizar la minería de datos en una aplicación para poder explotar su funcionalidad.

Una vez seleccionadas las tareas necesarias para la minería de datos, se procede al análisis de la data histórica obtenida para poder identificar los parámetros necesarios con los cuales se realizará el estudio de los patrones de comportamiento de esta industria. Las tendencias de comportamiento y los patrones de datos definen a un **modelo de minería de datos**.

#### **Modelo de Minería de Datos [4]**

El conjunto de estadísticas, datos y patrones, permiten identificar relaciones y generar predicciones.

El origen de los datos es almacenado en la estructura de minería de datos y un modelo de minería de datos guarda información resultante del procesamiento estadístico de los datos, como los patrones encontrados.

### “Arquitectura del modelo de minería de datos” [4]

La estructura de minería de datos, es procesada y analizada para obtener los metadatos, patrones y enlaces de la estructura de minería de datos con se india en la figura 4.

#### Figura 4

*Modelo de minería de datos [4]*



*Nota.* La figura detalla el modelo de minería de datos general. Tomada de <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/mining-models-analysis-services-data-mining?view=sql-server-2017>

Los metadatos permiten almacenar el nombre del modelo de minería de datos, el servidor donde se procesan, la estructura de minería de datos y algoritmos aplicados en el análisis de los datos.

Se debe tomar en cuenta los siguientes pasos para crear un modelo de minería de datos:

- Crear la estructura de minería de datos y las columnas de datos que se requieren.
- Identificar el algoritmo más acertado para el análisis.

- Seleccionar las columnas de la estructura de minería de datos para utilizarlas en el modelo, identificar el resultado que se quiere obtener.
- Establecer parámetros de procesamiento del algoritmo.
- Completar el modelo procesando los datos.

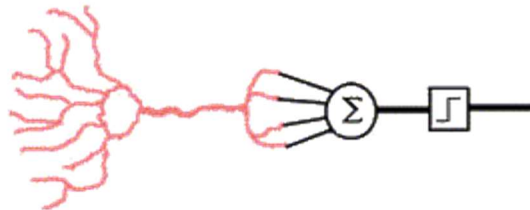
Los modelos de minería de datos más conocidos son:

### ***Modelo de red neuronal [5]***

En la red neuronal se presentan varios nodos que representan puntos de entrada de los datos, los cuales se los agrupa y se analizan mediante algoritmos para obtener resultados. En la figura 5 se ilustra una red neuronal.

### **Figura 5**

#### *Red Neuronal*



*Nota.* La figura representa una red con varios nodos y puntos de entrada de datos. Tomado de [https://www.aprenderaprogramar.com/index.php?option=com\\_content&view=article&id=258:mineria-de-datos-2o-parte-modelos-tecnicas-herramientas-dv00106a&catid=45&Itemid=164](https://www.aprenderaprogramar.com/index.php?option=com_content&view=article&id=258:mineria-de-datos-2o-parte-modelos-tecnicas-herramientas-dv00106a&catid=45&Itemid=164)

Las redes neuronales son conocidas como cajas negras, debido que el proceso para el tratamiento de datos no tiene pautas lógicas para el ser humano, pero son útiles para obtener predicciones, por lo que se utiliza en muchas aplicaciones.

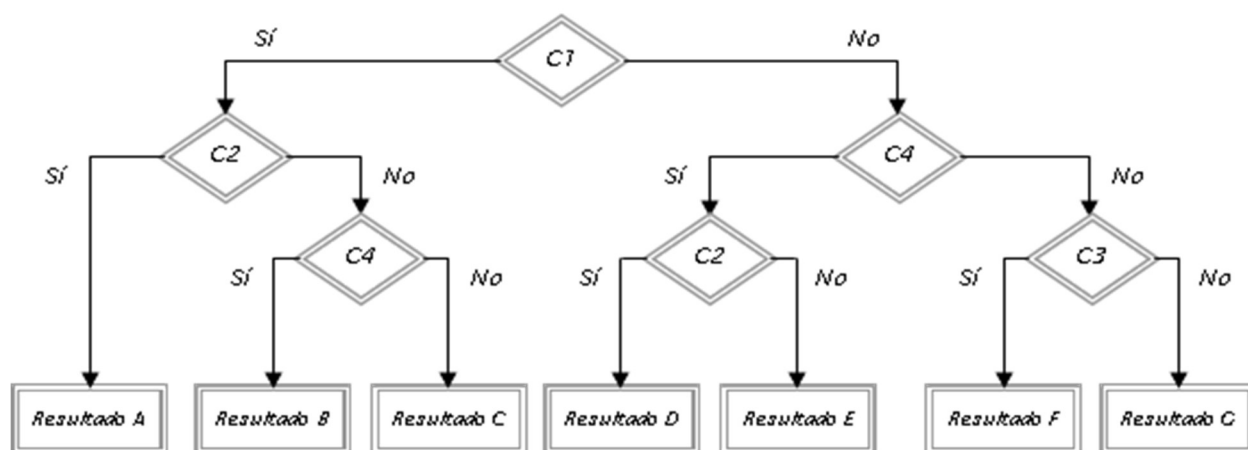
### **Modelo de árbol de decisión [5]**

Este modelo es conocido por el “divide y vencerás”, ya que agrupa datos según las características necesarias para el modelo mediante bifurcaciones hasta obtener el resultado. Este modelo tiene pautas lógicas y es conocido como “caja blanca”. Con las bifurcaciones de los datos se produce la “adivinación”.

En la figura 6 se ilustra un árbol de decisiones.

**Figura 6**

*Árbol de decisiones*



*Nota.* La figura detalla un árbol de decisiones agrupado según sus características. Tomado de [https://www.aprenderaprogramar.com/index.php?option=com\\_content&view=article&id=258:mineria-de-datos-2o-parte-modelos-tecnicas-herramientas-dv00106a&catid=45&Itemid=164](https://www.aprenderaprogramar.com/index.php?option=com_content&view=article&id=258:mineria-de-datos-2o-parte-modelos-tecnicas-herramientas-dv00106a&catid=45&Itemid=164)

### **Reglas de inducción [15]**

Mediante cálculos estadísticos, se genera reglas que siguen un patrón de si – entonces, con lo que pueden aplicarse a la data para extraer los resultados requeridos y hacer pronósticos.

## Metodologías para el proceso de Minería de Datos [16]

Se citan 3 metodologías para implementar el proceso de minería de datos:

- KDD
- CRISP-DM
- SEMMA

### **KDD**

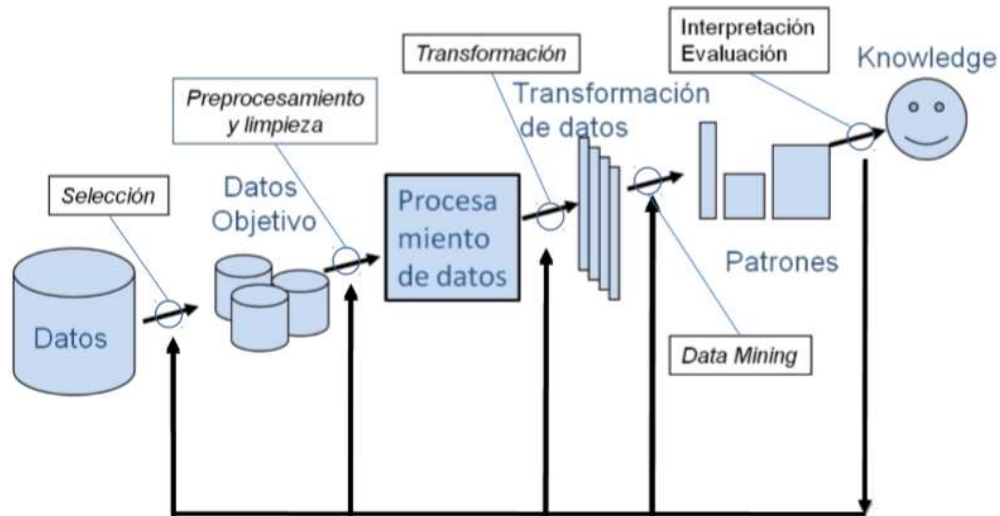
*“Es un proceso repetitivo y participativo. Es una metodología propuesta por Fayyad en 1996” [16], en donde propone 5 fases:*

- **Selección:** Determinar un conjunto de datos o variables para el análisis que se realizará
- **Pre procesamiento:** Limpieza y pre procesamiento de datos
- **Transformación:** reducción de datos con características útiles para el objetivo planteado
- **Minería de datos:** Se realiza el análisis de la hipótesis y modelo de datos, se selecciona el algoritmo de minería de datos para la obtención de patrones de interés.
- **Evaluación e implantación:** Interpretación de los patrones encontrados, uso del conocimiento obtenido con el análisis.

En la figura 7 se ilustra las fases de la metodología KDD

Figura 7

Metodología KDD



*Nota.* La figura representa los pasos de la metodología KDM. Tomado de León, E. P. (s.f).  
Minería de Datos.

### **CRISP – DM (CRoss Industry Standard Process for Data Mining) [6]**

El desarrollo de esta metodología fue iniciado en septiembre de 1996, financiado por red mundial (europeos) de 200 miembros.

El modelo presenta de 6 fases, algunas de ellas son bidireccionales, con lo cual se puede revisar de manera parcial o total las fases anteriores. En la figura 8 se ilustra las fases de la metodología.

- Conocimiento del negocio
- Conocimiento de los datos
- Proceso de los datos
- Modelado
- Evaluación

Figura 8

Metodología CRISP – DM

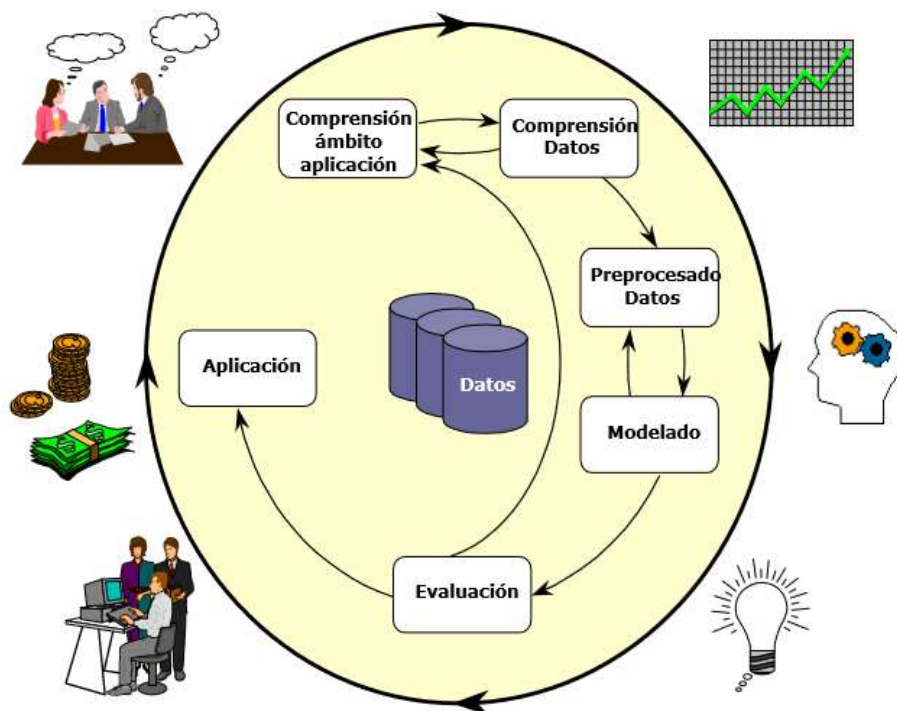


Imagen original: <http://www.crisp-dm.org/>

Nota. La figura representa la secuencia de la metodología Crisp – DM. Tomado de <http://crisp-dm.org>

### Conocimiento del negocio

- Presentación de los objetivos del negocio
- Valoración de la situación actual
- Establecer los objetivos de la minería de datos
- Creación del plan del proyecto

### Conocimiento de los datos

- Selección de datos iniciales
- Conocimiento de los datos
- Búsqueda de los datos

- Comprobación de calidad de datos

### **Preparación de los datos**

- Selección de los datos
- Limpieza de datos
- Construcción de datos
- Integración de datos
- Formateo de datos” [6]

### **Modelado**

- “Selección de la técnica de modelado
- Diseño de la evaluación
- Construcción del modelo
- Evaluación del modelo” [6]

### **Evaluación**

- “Evaluación de resultados
- Revisar el proceso
- Establecimiento de los siguientes pasos o acciones” [6]

### **Despliegue**

- “Planificación de despliegue
- Planificación de la monitorización y del mantenimiento
- Generación de informe final
- Revisión del proyecto” [6]



**SEMMA [16]**

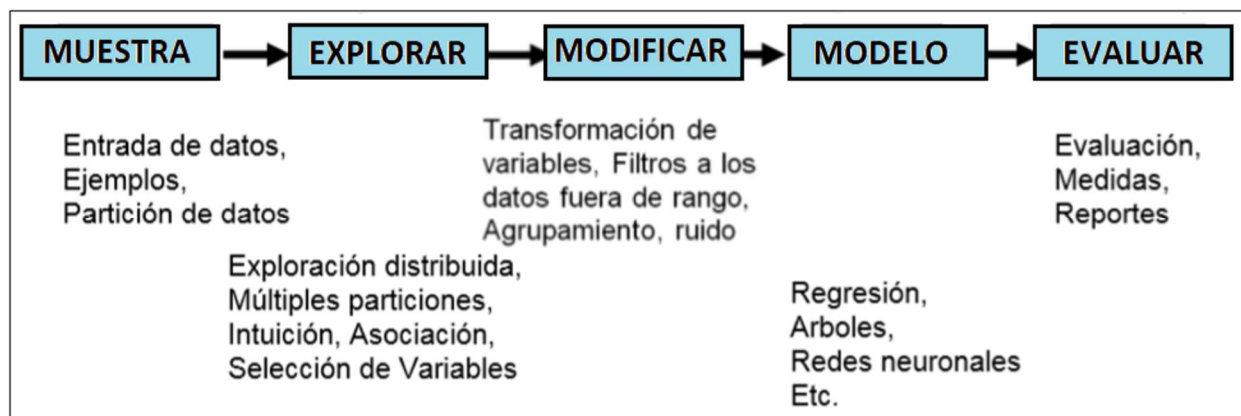
Esta metodología es una propuesta de SAS Institute Inc, el que indica que el proceso de sus fases con grandes cantidades de datos permite descubrir patrones de negocio desconocidos. Está conformado por 5 fases:

- Muestra
- Explorar
- Modificar
- Modelo
- Evaluar

En la figura 9 se ilustra las fases de la metodología y las acciones a realizarse en cada una de ellas.

**Figura 9**

*Metodología SEMMA [16]*



*Nota.* La figura representa los pasos de la metodología SEMMA. Tomado de León, E. P. (s.f.).

Minería de Datos.

Para el desarrollo de proyecto, se trabajará con la metodología CRISO – DM, debido que presenta las siguientes ventajas en relación a las otras metodologías:

- “No – propietario
- Independiente de la aplicación o la industria.
- Neutral con respecto a herramientas
- Enfocado en problemas de negocios.
- Plataforma guía
- Experiencia Base
- Plantillas para Análisis” [16]

### ***Revisión Preliminar de Literatura***

Para realizar el estado del arte, se utilizó la técnica de Revisión Preliminar de Literatura (PLR – Preliminary literatura reviews) con fuente digital de búsqueda de información IEEE. A continuación de detalla los criterios de Inclusión y Exclusión

#### **Criterios de Inclusión**

A continuación, se citan los criterios de inclusión:

- Industria retail
- Venta de productos retail
- Sobre stock de productos
- Índice de cobertura retail
- Análisis de ventas

#### **Criterios de Exclusión**

A continuación, se citan los criterios de exclusión:

- Artículos de Minería de datos que no estén relacionados a la industria retail
- Modelado de Datos

### **Grupo de Control**

A continuación, se presenta los grupos de control con los que se trabajará y los que se obtuvieron del análisis de los criterios de inclusión y exclusión



Tabla 2

*Grupo de Control*

GRUPO DE CONTROL	TITULO	PALABRAS CLAVE
EC01	<p><b>Título:</b> <i>“Big data, knowledge co-creation and decision making in fashion industry”</i></p> <p><b>Fecha:</b> octubre 2018</p> <p><b>URL:</b></p> <p><a href="https://www.sciencedirect.com/science/article/pii/S0268401218305024?via%3Dihub">https://www.sciencedirect.com/science/article/pii/S0268401218305024?via%3Dihub</a></p>	<p>Big data, Decision making, Organizational performance, Fashion industry, fashion retail, Business analytics</p>
EC02	<p><b>Título:</b> <i>“Tactical sales forecasting using a very large set of macroeconomic indicators”</i></p> <p><b>Fecha:</b> 16 de enero del 2018</p> <p><b>URL:</b></p> <p><a href="https://www.sciencedirect.com/science/article/abs/pii/S0377221717305957">https://www.sciencedirect.com/science/article/abs/pii/S0377221717305957</a></p>	<p>Forecasting, Tactical planning, Leading indicators, Variable selection, retail</p>

GRUPO DE CONTROL	TITULO	PALABRAS CLAVE
EC03	<p><b>Título:</b> <i>“Forecasting replenishment orders in retail: value of modelling low and intermittent consumer demand with distributions”</i></p> <p><b>Fecha:</b> 4 de febrero del 2018</p> <p><b>URL:</b>  <a href="https://www.tandfonline.com/doi/abs/10.1080/00207543.2018.1431413">https://www.tandfonline.com/doi/abs/10.1080/00207543.2018.1431413</a></p>	<p>forecasting, replenishment orders, planned orders, fashion retail supply chain, low and intermittent demand</p>
EC04	<p><b>Título:</b> <i>“Sales forecasting using extreme learning machine with applications in fashion retailing”</i></p> <p><b>Fecha:</b> diciembre 2008</p> <p><b>URL:</b>  <a href="https://www.sciencedirect.com/science/article/pii/S0167923608001371">https://www.sciencedirect.com/science/article/pii/S0167923608001371</a></p>	<p>Fashion sales forecasting, fashion retail, extreme learning machine, artificial neural network, backpropagation neural networks, decision support system</p>
EC05	<p><b>Título:</b> <i>“Fast fashion sales forecasting with limited data and time”</i></p> <p><b>Fecha:</b> marzo 2014</p>	<p>Fashion forecasting, fashion retail, fast fashion, quick forecasting, time series, intelligent forecasting</p>

GRUPO DE	TITULO	PALABRAS CLAVE
CONTROL	URL:	
	<a href="https://www.sciencedirect.com/science/article/pii/S0167923613002558">https://www.sciencedirect.com/science/article/pii/S01679</a>	
	<a href="https://www.sciencedirect.com/science/article/pii/S0167923613002558">23613002558</a>	

*Nota.* Esta tabla detalla los artículos de los grupos de control que se va analizar.

### Palabras clave y coincidencias

En la tabla siguiente, se enumeran las palabras claves y la repetición de las mismas en los diferentes artículos.

**Tabla 3**

*Palabras Clave en Grupos de Control*

Palabra clave	EC01	EC02	EC03	EC04	EC05	Repeticiones
Big data	1					1
Decision making	1					1
Fashion Retail	1	1	1	1	1	5
Fashion industry	1	1	1	1	1	5
Business analytics	1					1
Forecasting		1	1	1	1	4
Tactical planning		1	1			2
Leading indicators		1				1

*Nota.* La tabla detalla las palabras clave de los grupos de control.

### Cadena de búsqueda y resultados obtenidos

Con las palabras clave, se debe armar la cadena de búsqueda hasta obtener resultados óptimos, es decir, que la cadena devuelva los mejores artículos que hagan referencia a la problemática del stock en la empresa retail.

Los resultados de búsqueda se los realizó en la IEEE Xplore® Digital Library. La cadena de búsqueda ideal con la que se obtuvo 45 artículos es:

("RETAIL INDUSTRY" OR "RETAILING") AND ("FORECAST" OR "SALES FORECAST" OR "FORECASTING") AND "SUPPLY CHAIN"

De los resultados obtenidos se realizó el filtro de los artículos que cubren el tema de investigación, filtrado desde el año 2012 hasta el 2019, teniendo como resultado 11 artículos.

Adicional se descartan 6 artículos, debido que no tratan de la industria retail, dando como resultado 5 artículos que presentan información útil sobre la solución a la problemática del stock de la empresa retail.

A continuación, se presenta los artículos seleccionados.

**Tabla 4**

*Artículos Seleccionados*

TITULO	RESUMEN
<p><i>"Research on demand forecasting of retail supply chain emergency logistics based on NRS-GA-SVM"</i></p> <p><b>Autor:</b> Hong Xue; Chi Jiang; Boyu Cai; Yi Yuan</p> <p><b>Publicado en:</b> '42018 Chinese Control And</p>	<p><i>"Con el objetivo de las características de dimensión alta, condición dinámica y auto adaptación de parámetros para el pronóstico de la demanda, el modelo de pronóstico de demanda de la logística de emergencia de la cadena de suministro minorista se creó basándose en el algoritmo NRS-GA-SVM. El modelo de reducción del índice de atributos de muestra para el pronóstico de demanda de logística de emergencia se estableció en base al algoritmo NRS. Se adoptó el método de procesamiento continuo de datos. Los factores de influencia clave se extrajeron con mayor precisión. El modelo dinámico de pronóstico de demanda de logística de emergencia se estableció sobre la base de la teoría de regresión de la máquina de vectores de soporte no lineal y el algoritmo de aprendizaje automático"</i></p>



TITULO	RESUMEN
Decision Conference (CCDC)	<p><i>de optimización de parámetros para obtener el efecto de predicción óptimo. Los resultados numéricos de los experimentos muestran que el pre procesamiento de los índices con NRS y la optimización de los parámetros con GA no solo pueden mejorar la precisión de la logística de emergencia, sino que también pueden reducir el tiempo de ejecución del modelo de pronóstico, lo que promueve la capacidad de salvaguarda de emergencia del minorista. Emergencia de la cadena de suministro y verifica la viabilidad del modelo de pronóstico de demanda logística de emergencia para la cadena de suministro minorista.”</i></p>
<p><i>“A retail demand forecasting model based on data mining techniques”</i></p>	<p><i>“Este documento aborda el problema de pronosticar varias demandas de productos de los principales almacenes de distribución. El pronóstico de demanda es la actividad de construir modelos de pronóstico para estimar la cantidad de un producto que los clientes comprarán. Se ve afectado por numerosos factores diferentes, como el tamaño de la región del almacén, el recuento de clientes, el tipo de producto, etc. Cuando el número de almacenes y productos de distribución aumenta, se vuelve considerablemente difícil estimar la demanda de los clientes. En este estudio, proporcionamos una metodología adecuada para el pronóstico de la demanda que es capaz de superar las limitaciones mencionadas al tiempo que proporciona una alta precisión de estimación. La metodología propuesta agrupa almacenes similares de acuerdo con su</i></p>
<p><b>Autores:</b> İrem İşlek ; Şule Gündüz Öğüdücü</p> <p><b>Publicado en:</b> 2015</p> <p>IEEE 24th International Symposium on Industrial Electronics (ISIE)</p>	

TITULO	RESUMEN
	<p><i>comportamiento de venta utilizando agrupamiento de gráficos bipartitos. Después de eso, se aplica la fase de pronóstico híbrido que combina el modelo de promedio móvil y el algoritmo de aprendizaje automático de la red Bayesiana. Nuestros resultados experimentales en el conjunto de datos reales muestran que este enfoque mejora considerablemente el rendimiento del pronóstico”</i></p>
<p><i>“Inventory control systems with safety stock and reorder point approach”</i></p>	<p><i>“Los sistemas de control de inventario tienen un papel importante para monitorear la condición de los bienes en la etapa de la cadena de suministro de la distribución de bienes desde el almacén central hasta el comercio minorista de acuerdo con la demanda del consumidor. La investigación apunta a desarrollar sistemas de control de inventario en la cadena de suministro, distribución de bienes desde proveedores hasta el comercio minorista a través del almacén central con costos mínimos de cadena de suministro. Minimización del costo de la cadena de suministro utilizando el modelo de evaluación de pronóstico y el resultado como base para el cálculo del stock de seguridad y el punto de pedido. Los resultados de la investigación son el sistema que se puede utilizar para controlar el inventario y la renovación de los bienes en el almacén central al calcular la minimización del costo de la cadena de suministro en función de la demanda de los consumidores. Se realizaron estudios de caso sobre la distribución de la cadena de suministro de mercancías desde el almacén central hasta el comercio minorista. Los datos de</i></p>
<p><b>Autor:</b> Devi Ajeng Efrilianda; Mustafid; R. Rizal Isnanto</p>	
<p><b>Publicado en:</b> 2018 International Conference on Information and Communications Technology (ICOIACT)</p>	

TITULO	RESUMEN
	<p><i>observación derivados de la distribución de mercancías desde el almacén central hasta la venta al por menor de ciertos productos durante 4 períodos de tiempo. A través de dicho sistema de control de inventario, se puede minimizar el costo de la cadena de suministro, y el nivel de inventario del producto se puede mantener en el tiempo y la cantidad de inventario que se debe cumplir.”</i></p>
<p><i>“Fast fashion sales forecasting with limited data and time”</i></p> <p><b>Autor:</b> Tsan-MingChoiChi-LeungHuiNaLiuSau-FunNgYongYu</p> <p><b>Publicado en:</b> marzo 2014</p>	<p><i>“La moda rápida es una estrategia comúnmente adoptada en la venta de moda. En forma rápida, las decisiones operativas deben tomarse con un calendario ajustado y el método de pronóstico correspondiente debe completarse con datos muy limitados dentro de un período de tiempo limitado. Motivado por las rápidas prácticas comerciales de la moda, en este artículo se desarrolla un algoritmo de pronóstico inteligente que combina herramientas como la máquina de aprendizaje extrema y el modelo gris. Nuestro análisis de datos reales demuestra que este nuevo algoritmo derivado puede generar un pronóstico razonablemente bueno bajo el tiempo y las restricciones de datos dados. Un análisis adicional con un conjunto de datos artificiales muestra que el algoritmo propuesto se desempeña especialmente bien cuando la pendiente de la tendencia de la demanda es grande, o la variación del ciclo estacional es grande. Estas dos características se adaptan muy bien al patrón de demanda de moda rápida porque el factor de tendencia es significativo y el ciclo estacional suele ser muy variable de manera rápida. Los resultados</i></p>

TITULO	RESUMEN
<p data-bbox="203 527 509 827"> <i>“Tactical sales forecasting using a very large set of macroeconomic indicators”</i> </p> <p data-bbox="203 863 509 1024"> <b>Autor:</b> Yves R.Sagaert  <b>Publicado en:</b> enero 2018         </p>	<p data-bbox="537 321 1443 491"> <i>de este documento sientan las bases que pueden ayudar a lograr pronósticos de ventas en tiempo real para operaciones de moda rápidas en el futuro.”</i> </p> <p data-bbox="537 527 1443 1860"> <i>“El pronóstico táctico en la gestión de la cadena de suministro permite la planificación del inventario, la programación de la producción y la compra de materias primas, entre otras funciones. Por lo general, se refiere a las previsiones con 12 meses de antelación. Los modelos de previsión tradicionales tienen en cuenta la información extrapolada del pasado, pero no pueden anticipar eventos macroeconómicos, como aumentos abruptos o disminuciones en la actividad económica nacional. En la práctica, esto se contrarresta mediante el uso de juicios de expertos de gestión, que se sabe que sufren diversos sesgos, es costoso y no es escalable. Este documento evalúa múltiples enfoques para mejorar el pronóstico de ventas tácticas utilizando indicadores líderes macroeconómicos. El pronóstico estadístico propuesto selecciona automáticamente tanto el tipo de indicadores adelantados como el orden de los candidatos para cada uno de los indicadores seleccionados. Sin embargo, como los valores futuros de los indicadores adelantados son desconocidos, se introduce una incertidumbre adicional. Esta incertidumbre se controla en nuestra metodología al restringir las entradas a una configuración de pronóstico incondicional. Comparamos esto con la configuración condicional, donde se supone que se conocen los valores de los</i> </p>

TITULO	RESUMEN
	<p><i>indicadores futuros y evaluamos la pérdida teórica de la precisión del pronóstico. También evaluamos modelos puramente estadísticos a partir de modelos basados en juicios, en los que los potenciales indicadores adelantados son pre filtrados por expertos, cuantificando el equilibrio entre precisión y costo. El marco propuesto mejora la precisión de los pronósticos con respecto a los puntos de referencia establecidos de las series temporales, al tiempo que proporciona información útil sobre los principales indicadores clave. Evaluamos el enfoque propuesto en un estudio de caso real y encontramos un 18,8% de ganancias de precisión en el proceso de pronóstico actual.”</i></p>

*Nota.* La tabla presenta el resumen de los artículos seleccionados de los grupos de control

En base a los estudios analizados, se tiene varios artículos que tratan sobre sistemas de control de inventarios, modelos estadísticos, que permiten estimar la cantidad de productos que los clientes comprarán, con lo que se estima mejorar el stock que maneja cada industria.

Estos modelos estadísticos predictivos que mencionan los autores serán analizados para poder acoplarlos o mejorarlos en la ejecución de este proyecto para la industria retail.

### **Resultados obtenidos de estudio preliminar de literatura**

A continuación, se presenta los resultados obtenidos dando respuesta a las preguntas de investigación planteadas:

**OE2:** Obtener información de estudios realizados acerca del manejo del stock, a través de la revisión preliminar de literatura para conocer técnicas de administración de la industria de productos retail.

**OE2-RQ1:** ¿Qué soluciones presentan los estudios consultados sobre el alto índice de cobertura de stock en la industria retail?

Con los artículos obtenidos en la revisión de literatura, se puede evidenciar casos de estudio en donde se utiliza algoritmos matemáticos para solucionar la adquisición de productos en la industrial retail. Estos algoritmos utilizan información real de ventas, tendencia de la demanda, la variación del ciclo estacional.

Con los algoritmos matemáticos, se puede identificar los patrones de demanda en la industria retail, tomando en cuenta el ciclo estacional que es variable y con los cuales se puede pronosticar las ventas de los productos.

Estos resultados fueron obtenidos gracias a las herramientas de minería de datos, con los cuales se puede jugar con las diferentes variables para un mejor resultado.

**OE2-RQ1:** ¿Qué métodos utilizan para solucionar el alto índice de cobertura de stock en la industrial retail?

Los algoritmos utilizados para la reducción del índice de cobertura, los autores proponen el algoritmo NRS-GA-SVM que permiten mejorar la precisión del modelo y reducir el tiempo de ejecución, la red Bayesiana como algoritmo de aprendizaje automático, lo que permite lograr pronósticos de ventas en tiempo real para operaciones de moda rápidas en el futuro.

Estos algoritmos se encuentran en las herramientas de minería de datos, como el Rapidminer, con la que se trabajará en este proyecto de titulación.

A continuación, se presenta un resumen del estado del arte con el que se trabajó el presente trabajo de titulación.

Este proyecto de titulación se trabajó con una PLR (revisión preliminar de literatura) con ayuda de una búsqueda digital de información, utilizando el repositorio académico IEEE.

Se ha determinado criterios de inclusión como industrial retail, venta de productos, sobre stock, índice de cobertura, análisis de ventas y criterios de exclusión como minería de datos y modelado de datos.

Con los criterios de inclusión y exclusión se realiza grupos de control para determinar los artículos que proporcionen información para la realización de este proyecto.

Se realiza cinco grupos de control con las siguientes palabras claves:

- Big data
- Decision making
- Fashion Retail
- Fashion industry
- Business analytics
- Forecasting
- Tactical planning
- Leading indicators

Con estas palabras clave se realiza la cadena de búsqueda con la que se obtiene 45 artículos, los cuales fueron filtrados desde el año 2012 hasta el 2019, obteniendo 11 artículos, descartando 5 artículos que no pertenecen a la industria retail, quedando con 5 artículos de interés para este proyecto. En la tabla 4 se mencionan los artículos seleccionados.

### ***Metodología de investigación***

El enfoque investigativo que se utilizará en este proyecto de titulación será cualitativo debido que se va a utilizar la recolección y análisis de datos, concepciones, técnicas de estudio no cuantitativos para responder a las preguntas de investigación.

Este enfoque analiza realidades subjetivas de los datos extraídos para el análisis de los patrones de comportamiento de las ventas y compras en la industria retail, perfeccionando las hipótesis planteadas conforme se obtienen los datos.

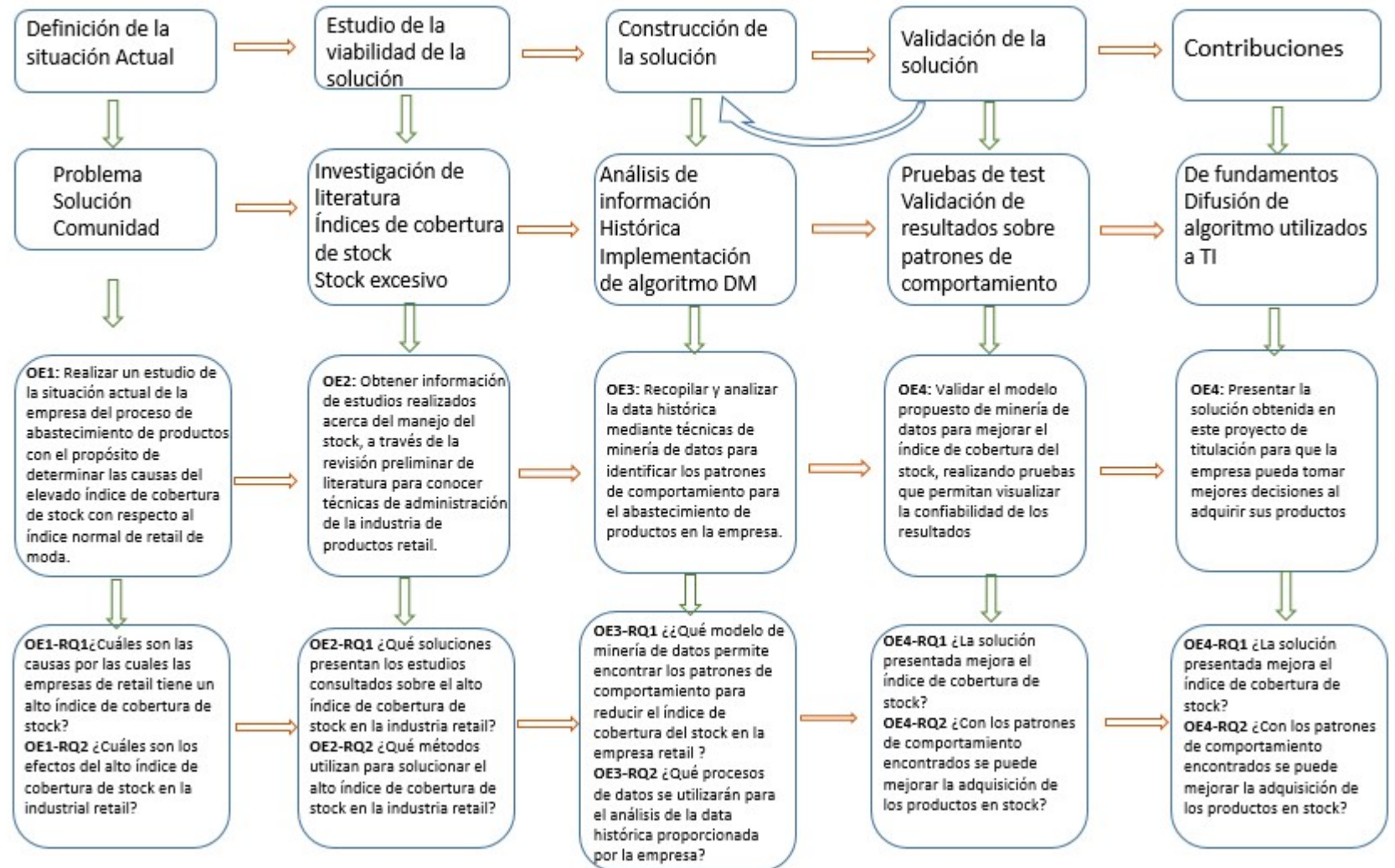
En este proyecto de titulación se utilizará una metodología de investigación híbrida, debido que con esta se podrá dar seguimiento ordenado en cada proceso de investigación.

La metodología utilizada se encuentra descrita en 5 pasos, uno de ellos pertenece a la metodología Design science research (Contribuciones) y a una metodología propia.

- Metodología Design Science Research
  - Contribuciones
- Metodología Propia
  - Definición de la situación actual
  - Estudio e la viabilidad de la solución
  - Construcción de la solución
  - Validación de la Solución

A continuación, se presenta la metodología que se utilizará en este proyecto de titulación y una descripción de cada fase.





Fases de la metodología de investigación:

### **Definición de la situación actual**

Con la información proporcionada por la empresa de retail, se identifica el problema que aqueja a la industria retail de este proyecto con respecto al sobre stock de sus productos. La empresa presenta un índice alto de stock con un indicador de ventas de 2 a 3 meses, lo cual genera un gasto adicional de almacenamiento de mercadería.

Esta primera fase, contempla el primer objetivo específico planteado y su respuesta a esta.

### **Estudio de la viabilidad de la solución**

En esta fase se trabaja con el estudio del arte, con lo cual se identifican las posibles soluciones a los problemas de sobre stock de retail, técnicas de administración de productos retail y propuestas para su solución.

En esta fase se trabaja con el segundo objetivo específico, dando respuestas a dos preguntas de investigación, de acuerdo a los resultados del estado del arte realizado en el literal 2.3.

### **Construcción de la solución**

Para la construcción de la solución se realiza la recopilación de la información requerida para el análisis de la misma, la cual es entregada por el departamento de tecnología. Con esta información se realiza su tratamiento (limpieza de datos), así como la identificación de los campos requeridos y que no afecten a la solución.

Se define las herramientas que se va a utilizar para el tratamiento de la información y obtener los patrones de comportamiento para evitar el sobre stock de los productos retail de la industria de la moda.

En esta fase se maneja con el tercer objetivo específico y las respuestas a sus preguntas.

### **Validación de la solución**

En esta etapa se valida los resultados obtenidos con respecto a los patrones de comportamiento encontrados para evitar el sobre stock de los productos retail de moda.

Con esta validación de resultados se recomendará una solución a este problema para mejorar el sobre stock de sus productos y cada vez ir disminuyendo este índice para evitar pérdidas económicas.

En esta fase se plantea el cuarto objetivo específico con sus respectivas respuestas.

### **Contribuciones**

En esta fase se capacitará al personal interesado sobre los algoritmos y herramientas utilizadas en la obtención de los resultados y aporte al giro del negocio.

En esta fase se utiliza el quinto objetivo específico en conjunto con sus respuestas obtenidas.

## Capítulo 3

### Desarrollo de la solución

#### *Requerimientos del Negocio*

La industria de retail ha marcado tendencias en la creación de experiencias y exclusividad, registrando cambios con las nuevas tendencias tecnológicas.

Las ventas de la industria ecuatoriana de moda también llegaron a la Unión Europea, en donde alcanzaron la cifra de US\$ 1,4 millones de dólares. La moda de Ecuador busca crecer fuera de sus fronteras.

*“En 2017, las exportaciones ecuatorianas de moda alcanzaron los US\$ 86,6 millones de dólares, un 15% menos que en 2016. Sin embargo, el sector se encuentra desarrollando un plan quinquenal que culminará en 2022, con el objetivo de triplicar el volumen de las ventas al exterior, hasta llegar a los trescientos millones de dólares.*

*El sector situó sus exportaciones en US\$ 40,9 millones de dólares durante los primeros cinco meses del 2018, según datos de la Asociación de Industriales Textiles de Ecuador (Aite). La industria textil del Ecuador presentó el 58% de ventas hacia Sudamérica, a países como Bolivia, Colombia y Perú” [17]*

Actualmente, la industria retail ha tenido un gran crecimiento dentro del país, por lo que debe encontrar un equilibrio en la adquisición de sus productos para su crecimiento.

La industria retail debe considerar nuevos métodos de análisis para la adquisición de sus productos lo cual se encuentra revolucionando esta industria, con esto se tendrá un mejor enfoque en las necesidades de los consumidores, teniendo una mejor participación del presupuesto de la industria.

La industria retail, requiere métodos eficaces para poder adquirir sus productos conforme a las necesidades identificadas de los consumidores; es decir, mediante métodos de minería de datos, poder analizar el abastecimiento que debe realizar la empresa con un mínimo de stock excesivo, lo cual genera gastos a la empresa.

Los grupos de retail se esfuerzan por liquidar su mercadería, lo cual resulta difícil cuando el producto no es innovador, provocando que la mercadería se almacene y se pretenda exhibir el siguiente período, generando gastos adicionales a la empresa.

En esta pandemia, muchas empresas de retail de la moda se han visto forzadas en aumentar su stock y almacenarlo para la siguiente temporada incluyendo descuento para disminuir su pérdida. Se calcula que el 50% del stock pasará a ser vendido en el año 2021 [17].

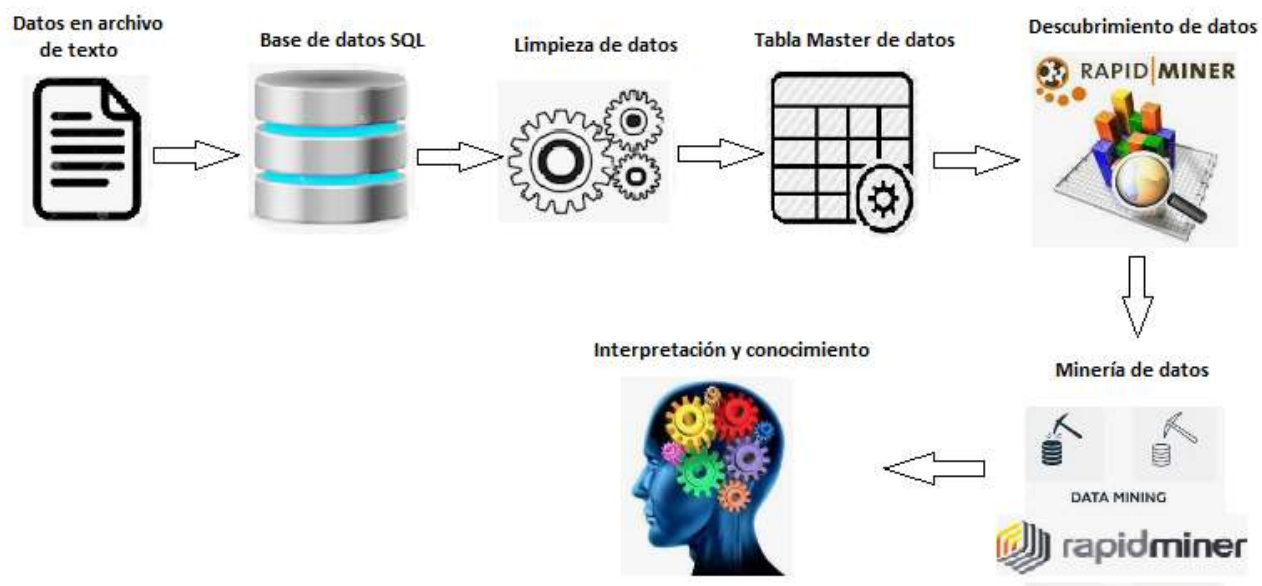
La planificación de abastecimiento de productos permite a las empresas reaccionar ante cambios en la industria y tener un mejor control sobre el stock adquirido. Sin esta planificación se pueden presentar los siguientes problemas [18]:

- “Altos costos de stocks y obsolescencia de productos.
- Frecuentes roturas de stock e inventario desequilibrado.
- Bajos niveles de servicio al cliente.
- Pérdida de ventas y resultados insatisfactorios.” [18]

En la figura 10, se presenta un esquema de la solución propuesta para este proyecto. La información proporcionada por la empresa se la obtuvo en un archivo plano, el mismo que se sube a un motor de base de datos para poder realizar la limpieza de los datos y así obtener una tabla Master de datos con la cual se trabaja para realizar el descubrimiento de los datos útiles para la minería de datos y así obtener los patrones y conocimiento en este análisis.

**Figura 10**

*Diagrama de solución*



*Nota.* La figura detalla la solución planteada al problema de la empresa.

**Recolección y Exploración de Datos**

Para el análisis del abastecimiento de productos de una industria retail, se tiene una muestra de datos de un período de cuatro años, lo cual permitirá identificar los productos de mayor afluencia o venta y cuáles son los productos que no son muy acogidos por el consumidor y así disminuir el stock excesivo de estos productos.

La data obtenida de la empresa de retail presenta los siguientes datos:

- Código de orden
- Fecha de venta
- Código del cliente
- Ciudad
- Provincia

- Sucursal
- Código del producto
- Categoría
- Sub Categoría
- Nombre del producto
- Precio venta
- Cantidad
- Descuento
- Ganancia
- Costo de envío
- Prioridad

### ***Preparación de la data a utilizar***

La data que se va a utilizar se encuentra en archivos planos, los cuales se subirán a un motor de base de datos SQL SERVER para la limpieza de la data que se va a utilizar para este análisis.

En la figura 11 se visualiza la data subida al motor de base de datos la misma que será limpiada para su análisis.

### **Figura 11**

#### ***Información cargada a la Base de datos SQL***

ID_ORDEN	FECHA_VENTA	ID_CLIENTE	CIUDAD	PROVINCIA	SUCURSAL	ID_PRODUCTO	CATEGORIA	SUB_CATEGORIA	NOMBRE_PRODUCTO	VENTA
MX-2014-137442	2018-10-23 00:00:00.000	TB-21280	Quito	Pichincha	Quicentro Sur	FUR-TA-10004937	Mujeres	Formal	Vestidos	928.256
IN-2014-16922	2018-10-24 00:00:00.000	DK-13150	Quito	Pichincha	Scala Shopping	OFF-ST-10004183	Hombres	Zapatos	Casual	648
IN-2014-11945	2018-10-27 00:00:00.000	SS-20140	Quito	Pichincha	Scala Shopping	TEC-CO-10000764	Niños	Niños/Niñas	Pantalones	1151.928
US-2014-146808	2018-10-23 00:00:00.000	HM-14980	Quito	Pichincha	Portal Shopping	OFF-AP-10003956	Mujeres	Deportivo	Buzos Capucha	133.824
ES-2014-1204900	2018-10-26 00:00:00.000	SN-20560	Quito	Pichincha	Quicentro Norte	FUR-CH-10002030	Mujeres	Basico	Camisetas Cortas	417.456
ES-2014-4325220	2018-10-25 00:00:00.000	DC-12850	Guayaquil	Guayas	Matriz	TEC-PH-10001937	Niños	Bebes	Zapatos	504.09

**Nota.** La figura detalla la información cargada a la base de datos.

La herramienta que se utiliza para la limpieza de datos es Talend Open Studio para Big Data, la cual es una herramienta de código abierto y permite la conexión con algunos motores de datos.

En el proceso de limpieza de datos se validará los siguiente:

- Campos blancos
- Campos nulos

En la figura 12 se puede visualizar el proceso de limpieza de datos que se realiza con cada campo de la tabla SuperStore.

**Figura 12**

*Limpieza de datos*



*Nota.* La figura detalla las tareas que se realizó para la limpieza de datos.



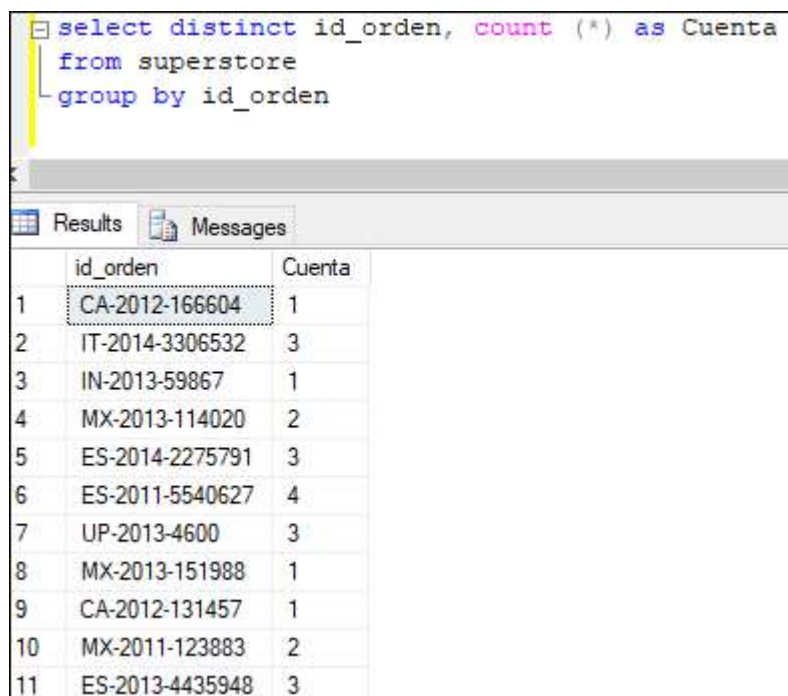
Una vez que se realiza la limpieza de los datos, la tabla SuperStore se convierte en una tabla Master, con la que se va a trabajar en el descubrimiento de los datos con la herramienta Rapidminer.

A continuación, se presenta el análisis de cada campo de la tabla master:

**ID\_ORDEN:** Este campo representa el código de la orden de la compra del artículo. En este análisis se visualiza 25.035 códigos únicos de una data de 51.290 filas. En la figura 13 se realiza un conteo de datos únicos para cada orden de compra. Este campo no discriminatorio para el modelamiento

### Figura 13

*Conteo de datos únicos*



```
select distinct id_orden, count (*) as Cuenta
from superstore
group by id_orden
```

	id_orden	Cuenta
1	CA-2012-166604	1
2	IT-2014-3306532	3
3	IN-2013-59867	1
4	MX-2013-114020	2
5	ES-2014-2275791	3
6	ES-2011-5540627	4
7	UP-2013-4600	3
8	MX-2013-151988	1
9	CA-2012-131457	1
10	MX-2011-123883	2
11	ES-2013-4435948	3

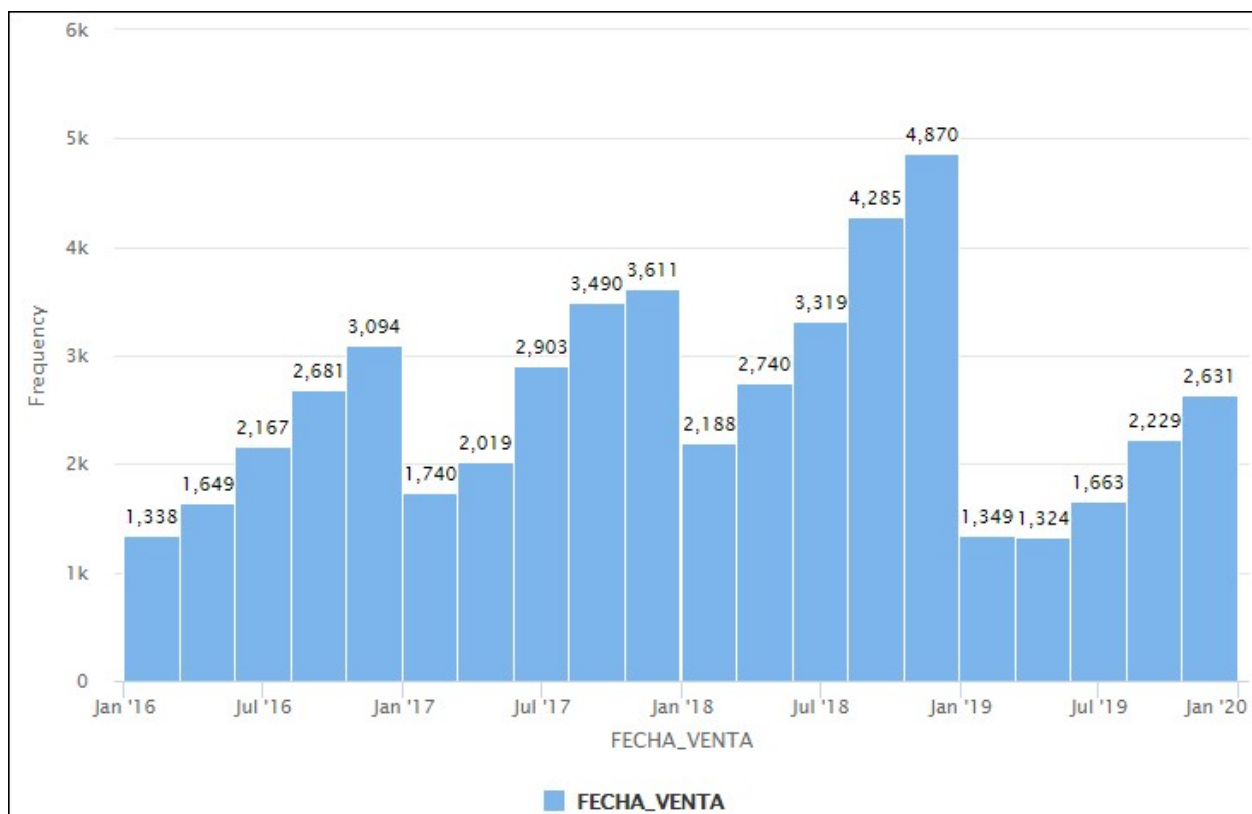
*Nota.* La figura detalla el conteo de los registros únicos de la data cargada.

**FECHA\_VENTA:** Este campo representa la fecha de venta del artículo, en la figura 14 se tiene un período de análisis de 4 años desde enero 2016 a diciembre 2019. Los meses con

mayores ventas son los meses de octubre a diciembre de cada año, meses de festividades en el país.

## Figura 14

### Análisis de fechas de ventas



*Nota.* La figura detalla la venta de los productos por fecha.

**ID\_CLIENTE:** Este campo representa el código del cliente que lo identifica de manera única. En la figura 15 se presenta 1.590 registros únicos, es decir, clientes que visitaron la tienda de retail en un período de 4 años.

**Figura 15***Clientes únicos*

```

select distinct ID_CLIENTE, COUNT (*) AS NUMERO
from superstore
group by ID_CLIENTE

```

	ID_CLIENTE	NUMERO
1	KC-6675	4
2	EH-3945	19
3	KL-6645	15
4	HA-4905	10
5	RB-19465	70
6	JE-16165	38
7	TS-21205	64
8	BW-11110	81
9	SJ-20125	58
10	DK-3150	19
11	IL-15100	48
12	GT-14755	48

*Nota.* la figura detalla un conteo de clientes de las ventas realizadas.

En la figura 16 se observa que la mayor concentración de ventas de artículos se encuentra en los siguientes clientes:

**Tabla 5***Top 10 de clientes*

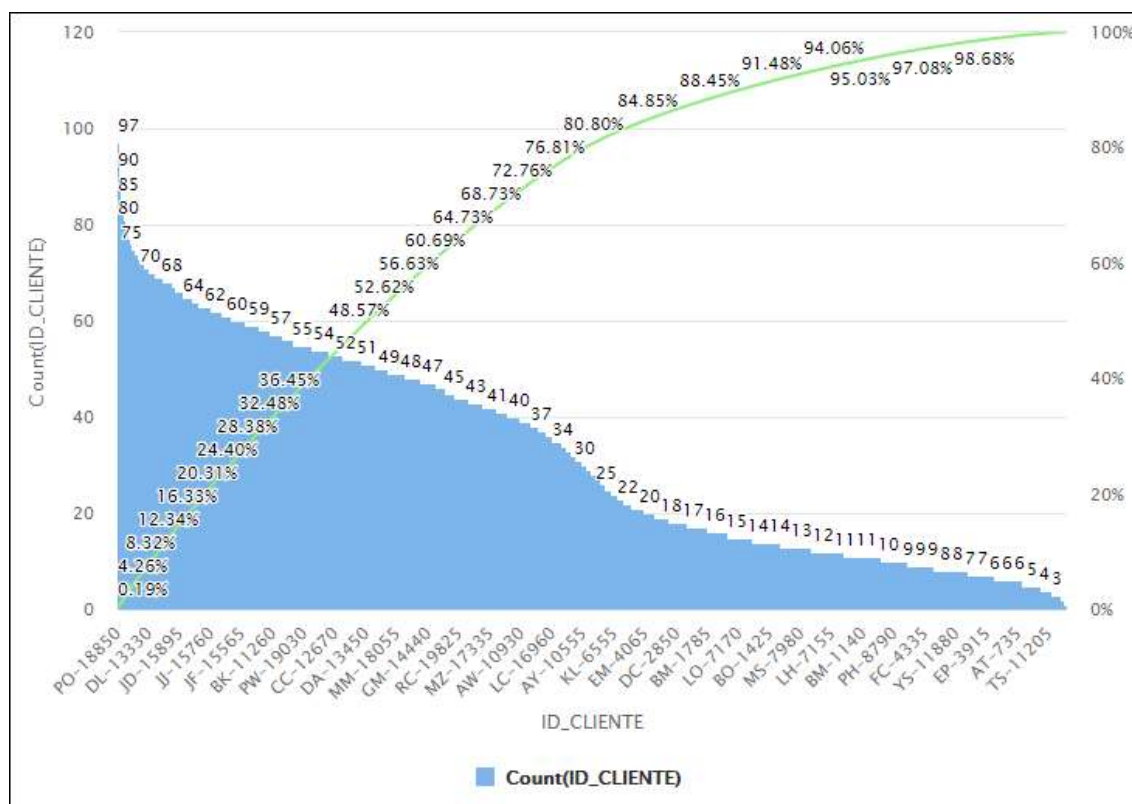
ID_CLIENTE	# DE REGISTROS
PO-18850	97
BE-11335	94
JG-15805	90
SW-20755	89

ID_CLIENTE	# DE REGISTROS
EM-13960	85
MY-18295	85
ZC-21910	84
MP-17965	84
CK-12205	83
BW-11110	81

*Nota.* La tabla detalla los clientes con mayores ventas

**Figura 16**

*Concentración de clientes.*

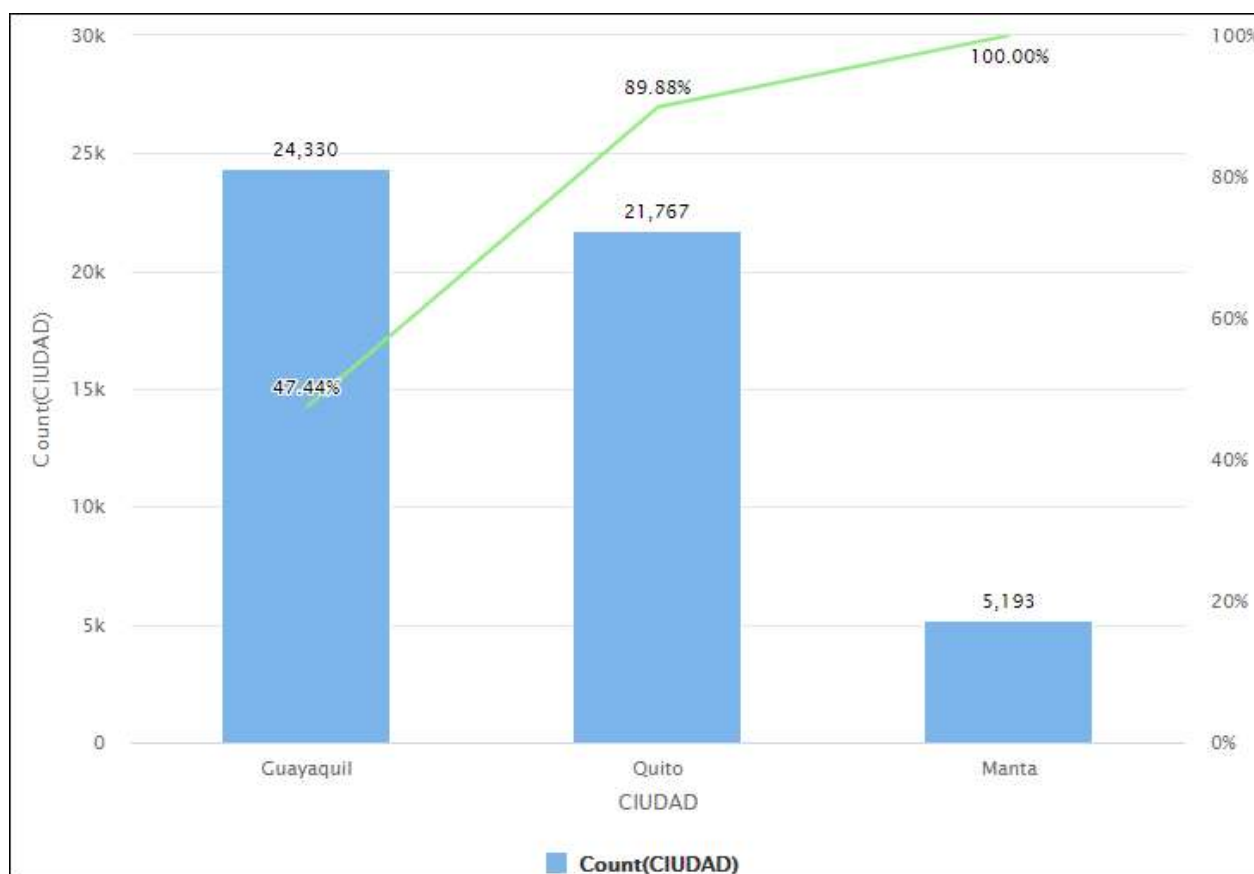


**CIUDAD:** Este campo representa a la ciudad en la que la tienda se encuentra ubicada.

En el gráfico 17 se puede observar que existe el 47.44% de ventas en Guayaquil, el 42.44% de ventas en Quito y el 10.12% en Manta, siendo Quito la ciudad donde se encuentra la matriz.

**Figura 17**

*Análisis del campo ciudad.*

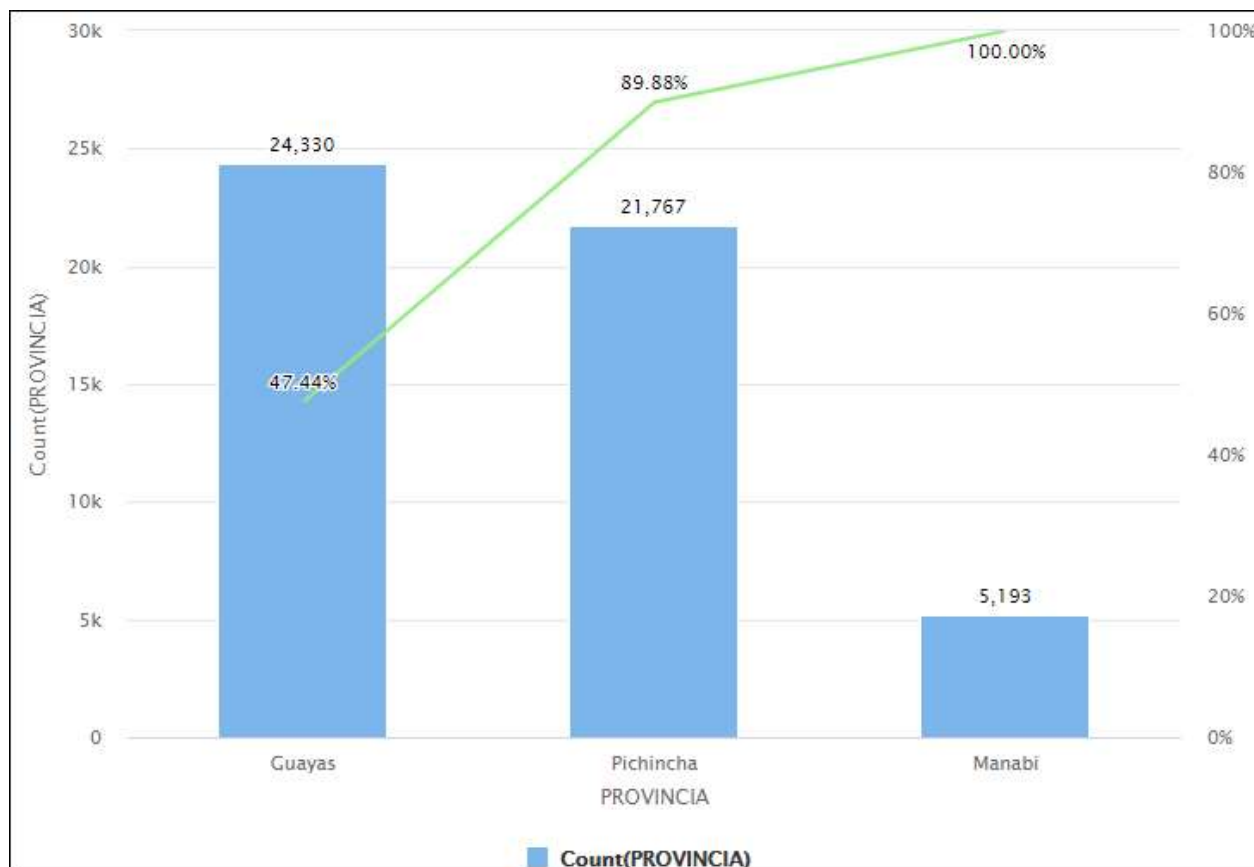


*Nota.* La figura detalla la concentración de ventas por ciudad.

**PROVINCIA:** Este campo representa a la provincia en la que la tienda se encuentra ubicada. En el gráfico 18 se puede observar que existe el 47.44% de ventas en Guayas, el 42.44% de ventas en Pichincha y el 10.12% en Manabí.

**Figura 18**

*Análisis del campo provincia.*

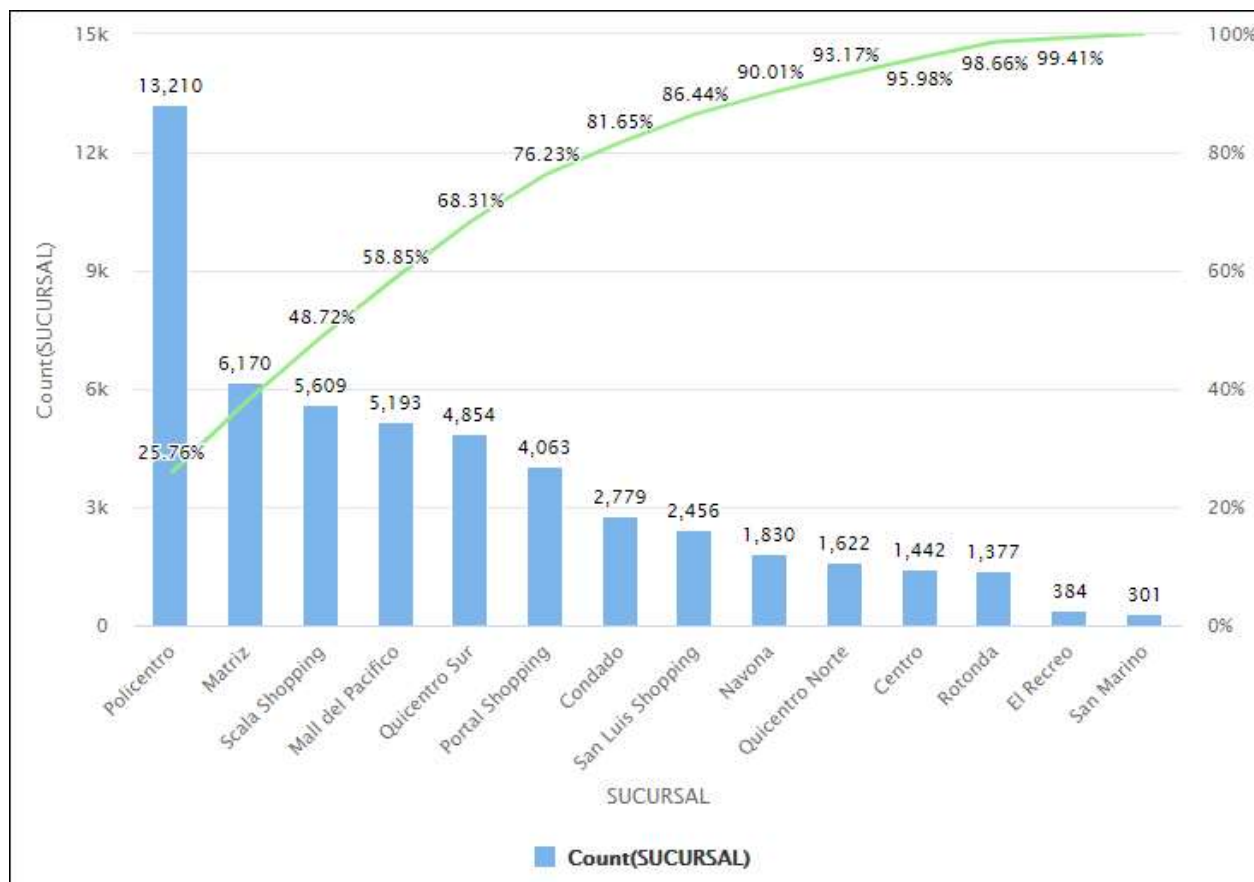


*Nota.* La figura detalla la concentración de ventas por provincia.

**SUCURSAL:** Este campo representa la sucursal en la que la tienda se encuentra ubicada. En el gráfico 19 se puede observar que existe el 25.76% de ventas registradas en el Policentro, 12.03% en la Matriz, 10.93% en el Scala Shopping, 10.13% en el Mall del Pacifico, 9.46% en el Quicentro del Sur, 7.92% en el Portal Shopping, 5.42% en el Condado, 4.79% en el San Luis Shopping, 3.57% en Navona, 3.16% en el Quicentro Norte, 2.81% en el centro, 2.68% en la Rotonda, 0.75% en el Recreo, 0.59% en San Marino

**Figura 19**

*Análisis del campo sucursal.*



*Nota.* La figura detalla la concentración de ventas por sucursal.

**ID\_PRODUCTO:** Este campo representa el código del producto que lo identifica de manera única. En la figura 20 se presenta 10.292 registros únicos.

**Figura 20***Productos Únicos*

```

select distinct ID_PRODUCTO, COUNT (*) AS NUMERO
from superstore
group by ID_PRODUCTO

```

	ID_PRODUCTO	NUMERO
1	OFF-AVE-10000608	13
2	OFF-AP-10003909	1
3	TEC-PH-10002070	8
4	TEC-CAN-10001376	1
5	TEC-MA-10002712	2
6	FUR-CH-10004676	7
7	OFF-ST-10004331	1
8	OFF-STA-10003956	12
9	FUR-CH-10001317	1
10	FUR-BO-10003668	1

*Nota.* La figura detalla un conteo de los productos por su ID.

En la figura 21 se visualiza los porcentajes de los productos que son adquiridos por el usuario con mayor y menor demanda. A continuación, en la tabla 6 se detalla los id de los productos

**Tabla 6**

*Top 10 de ID de producto de mayor y menor demanda*

MAYOR DEMANDA		MENOR DEMANDA	
ID_PRODUCTO	# DE REGISTROS	ID_PRODUCTO	# DE REGISTROS
OFF-AR-10003651	35	FUR-BO-10001318	1
OFF-AR-10003829	31	OFF-PA-10000354	1
OFF-BI-10002799	30	OFF-SU-10001455	1

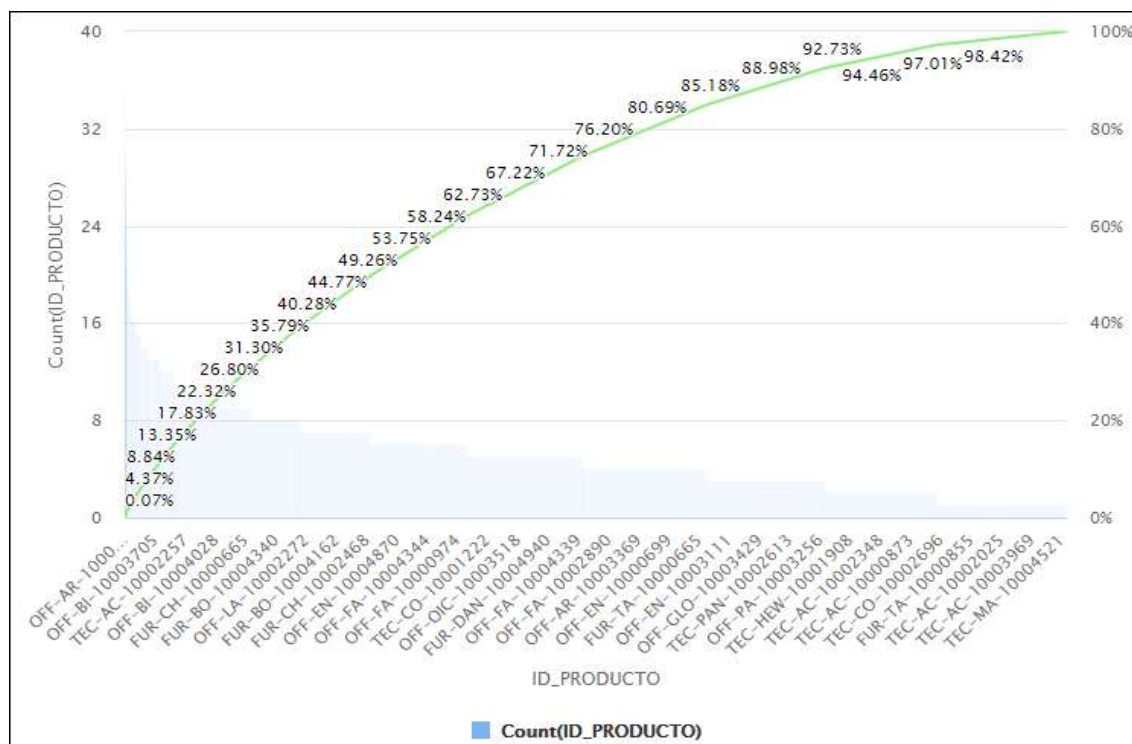


MAYOR DEMANDA		MENOR DEMANDA	
OFF-BI-10003708	30	OFF-AP-10000509	1
FUR-CH-10003354	28	OFF-AP-10000195	1
OFF-BI-10002570	27	FUR-FU-10000546	1
OFF-BI-10004140	25	OFF-BI-10003417	1
OFF-BI-10004195	24	OFF-PA-10000601	1
OFF-BI-10004632	24	OFF-EN-10000561	1
OFF-BI-10003650	24	OFF-LA-10001194	1

*Nota.* La tabla detalla los productos y el número de registros en la base de datos

**Figura 21**

*Análisis del código del producto*

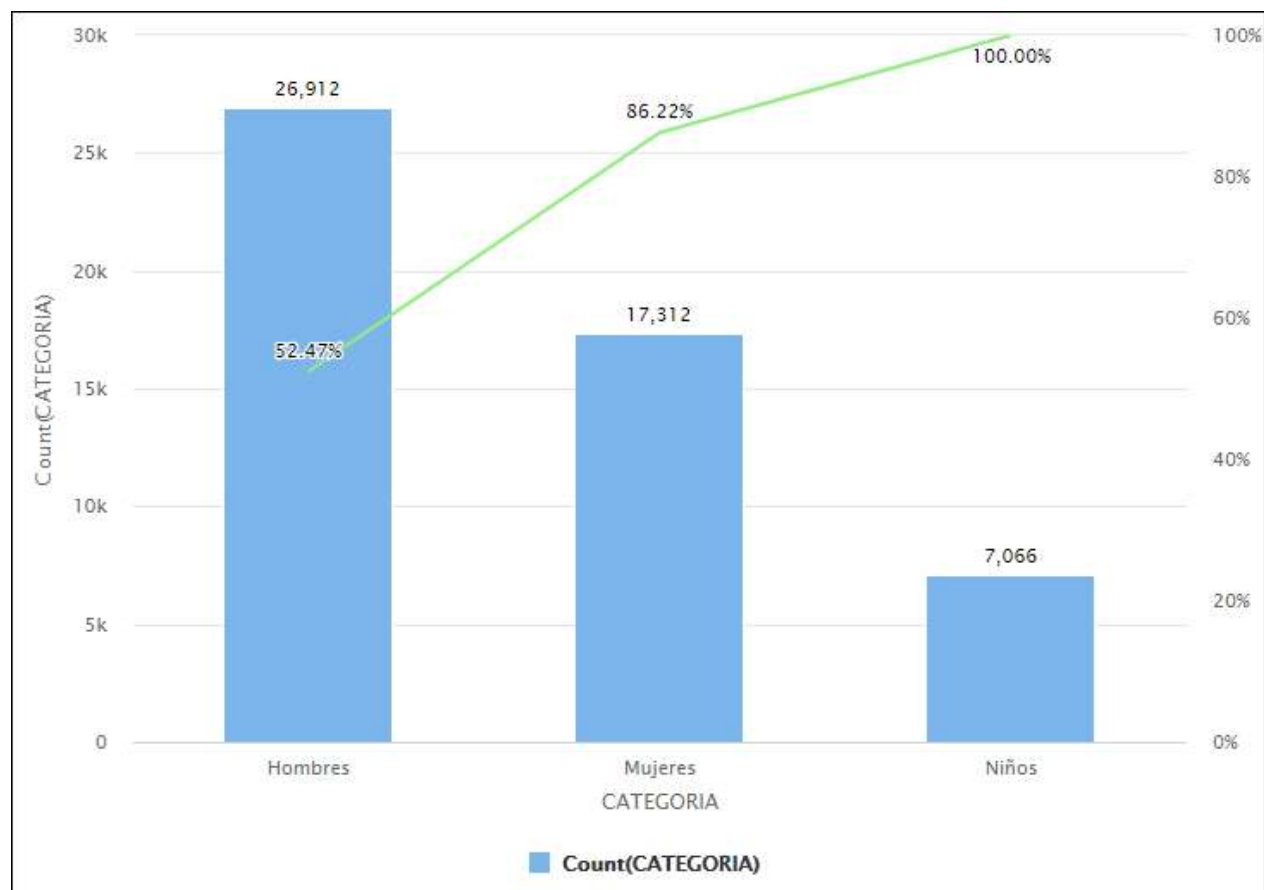


*Nota.* La figura detalla el código de los productos y su demanda en las ventas.

**CATEGORIA:** Este campo representa la categoría de los productos. En el gráfico 22 se puede observar que existe el 52.47% de productos para hombres, el 33.75% para mujeres y el 13.78% para niños.

**Figura 22**

*Análisis de la categoría del producto*



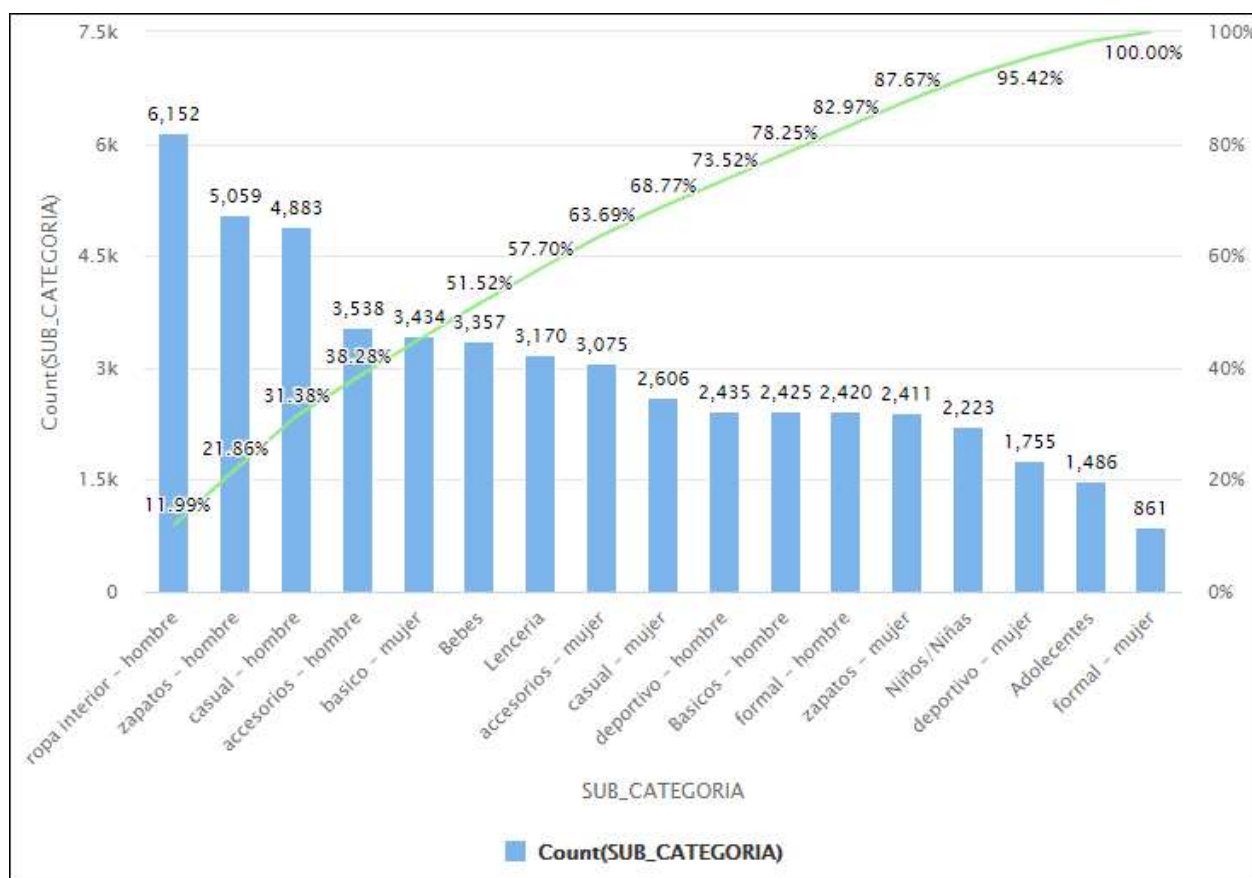
*Nota.* La figura detalla las ventas por categoría.

**SUBCATEGORIA:** Este campo representa la subcategoría de los productos. En el gráfico 23 se puede observar que existe el 11.99% de ropa interior - hombres, el 9.98% de zapatos hombre, el 9.52% de ropa casual hombre, el 6.9% de accesorios hombre, 6.69% de ropa básica mujer, 6.55% ropa de bebé, 6.18% lencería, 5.99% accesorios mujer, 5.08% de

ropa casual de mujer, 4.75% ropa deportiva hombre, 4.73% ropa básico hombre, 4.72% ropa formal hombre, 4.7% zapatos mujer, 4.33% ropa de niños, 3.42% ropa deportiva mujer, 2.9% ropa de adolescente, 1.68% ropa formal mujer.

**Figura 23**

*Análisis de la subcategoría del producto*

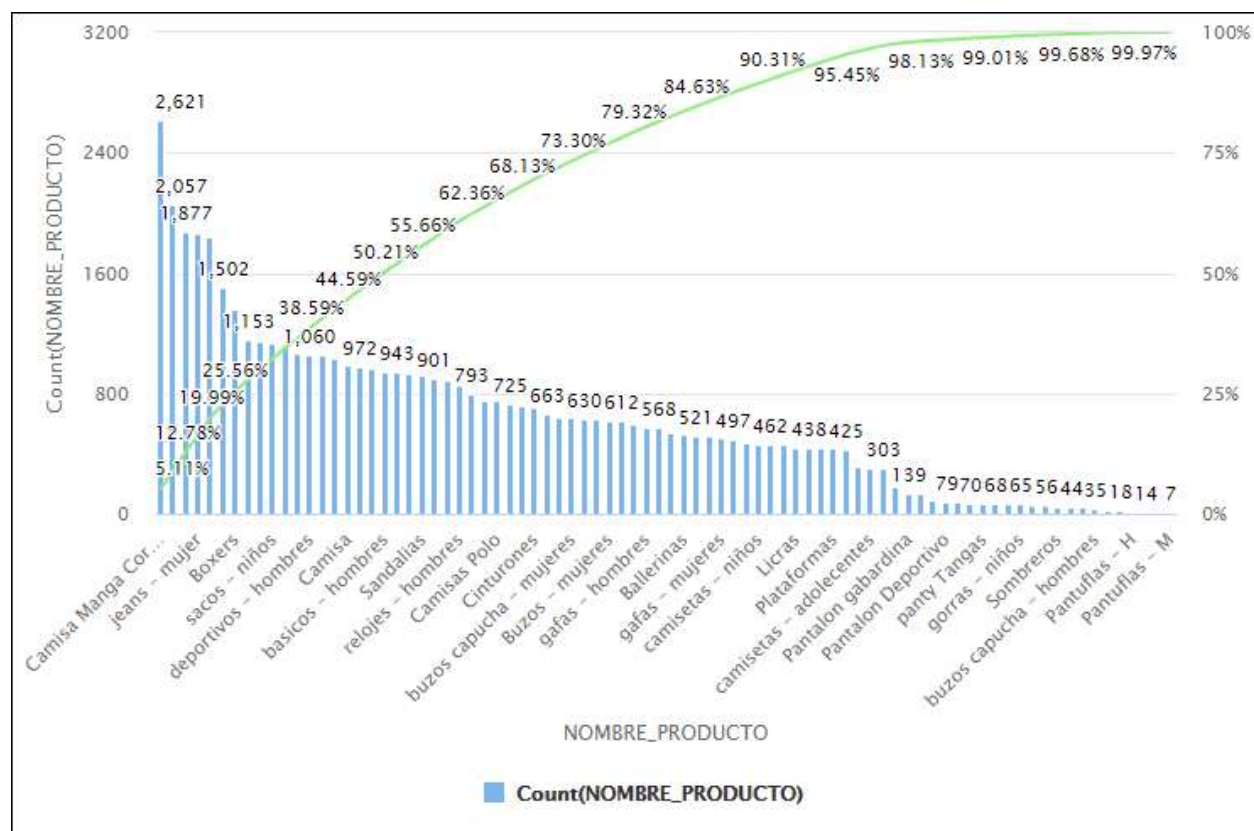


*Nota.* La figura detalla la venta por subcategoría.

**NOMBRE\_PRODUCTO:** Este campo representa el nombre de los productos. En el gráfico 24 se puede observar el porcentaje de mayor y menor venta de los productos retail.

Figura 24

## Análisis del nombre del producto



Nota. La figura detalla las ventas por producto.

A continuación, en la tabla 7 se detallan los nombres de los productos:

Tabla 7.

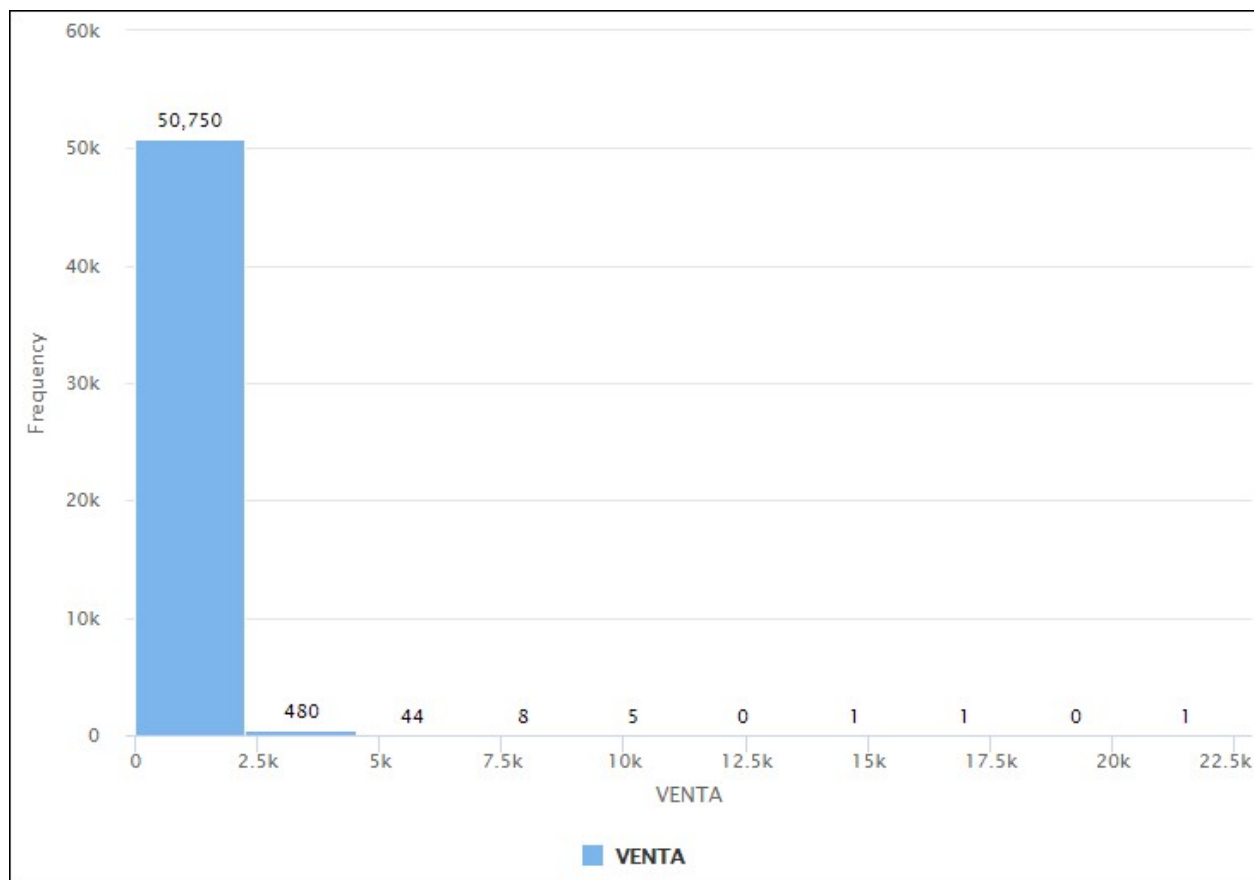
Top 10 de producto de mayor y menor demanda

MAYOR DEMANDA		MENOR DEMANDA	
NOMBRE_PRODUCTO	# DE REGISTROS	NOMBRE_PRODUCTO	# DE REGISTROS
Camisa Manga Corta	2621	Sombreros	50
medias - hombres	2057	Pañuelos	44

MAYOR DEMANDA		MENOR DEMANDA	
Botines	1877	Jeans – hombres	42
Jeans – mujer	1859	buzos capucha – hombres	35
Chaleco	1838	Corbatas	23
Correas	1502	deportivos – mujeres	18
Boxers	1358	Pantuflas – H	16
Panty	1153	Shorts – adolescentes	14
Calzoncillos	1145	Overol	8
sacos – niños	1129	Pantuflas – M	7

*Nota.* La tabla detalla los productos con mayor y menor demanda.

**VENTA:** Este campo representa el valor de la venta por artículo realizada. En la figura 25 se visualiza que la mayor concentración está en el primer intervalo de 0 a 2.2k

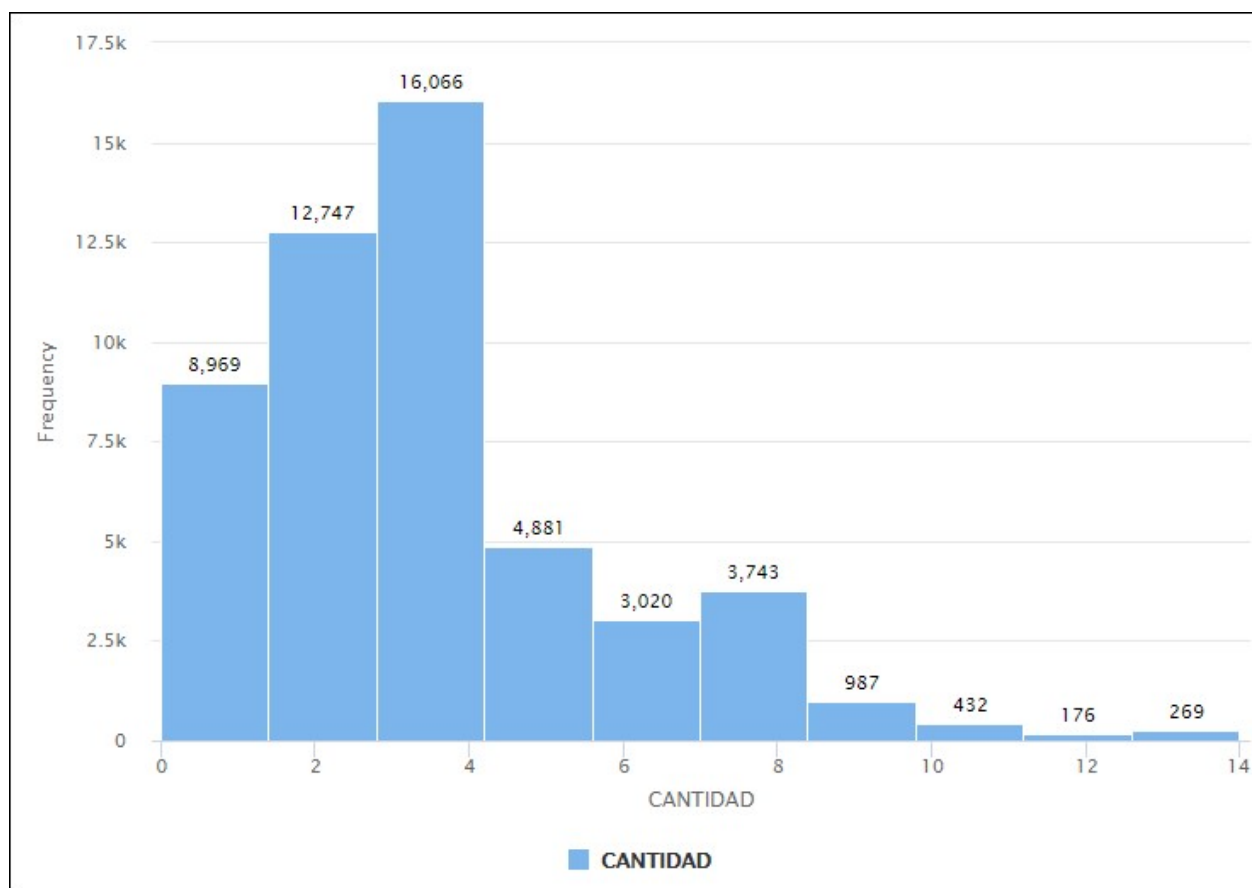
**Figura 25***Análisis del campo venta*

*Nota.* La figura detalla la concentración de ventas.

**CANTIDAD:** Este campo representa la cantidad de artículos que se venden. En la figura 26 se visualiza que la mayor concentración o mayor número de unidad vendidas está en el rango de 0 a 4 unidades

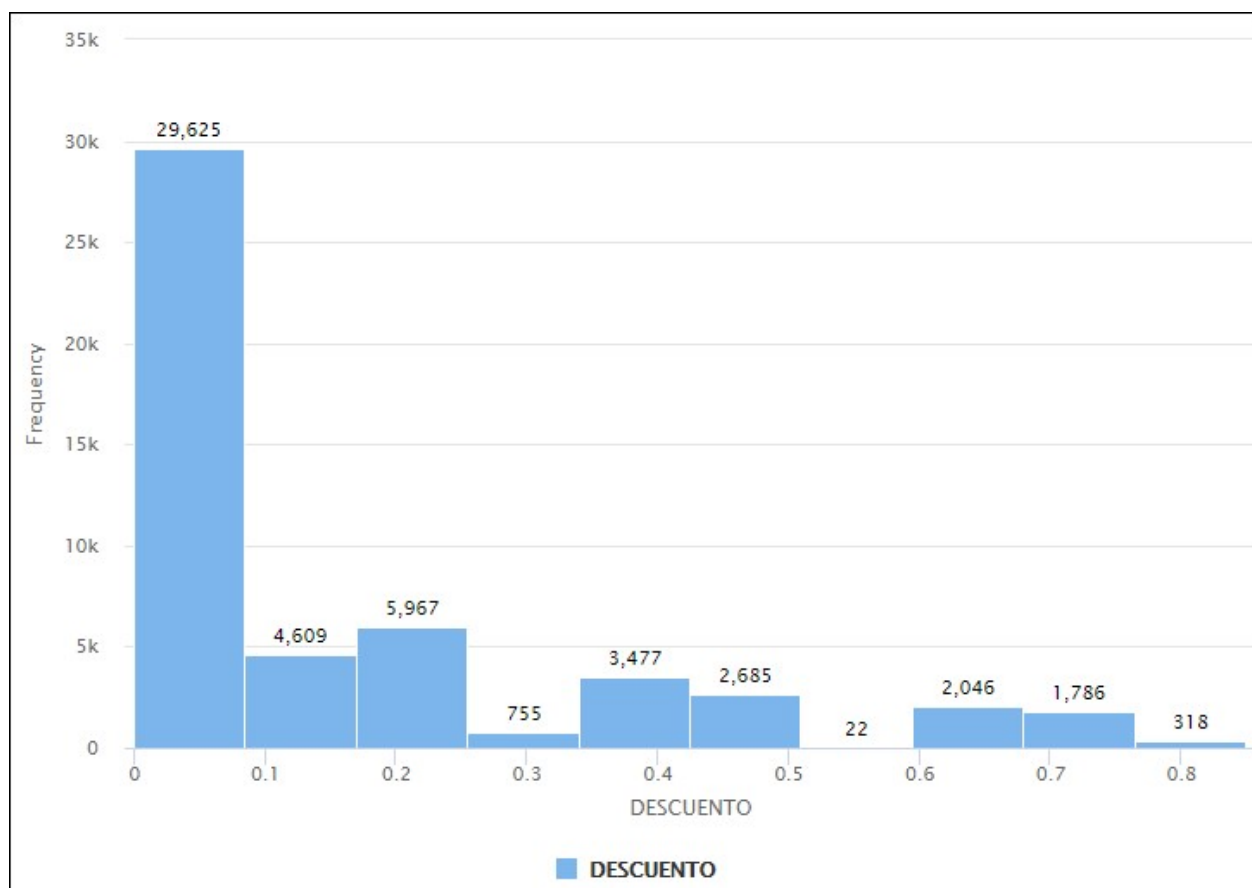
**Figura 26**

*Análisis del campo cantidad*



*Nota.* La figura detalla la concentración de la cantidad de los productos.

**DESCUENTO:** Este campo representa el descuento que tiene el cliente por su compra por artículo. En la figura 27 se visualiza que la mayor concentración se encuentra en el primer intervalo de 0 a 0.085

**Figura 27***Análisis del campo descuento*

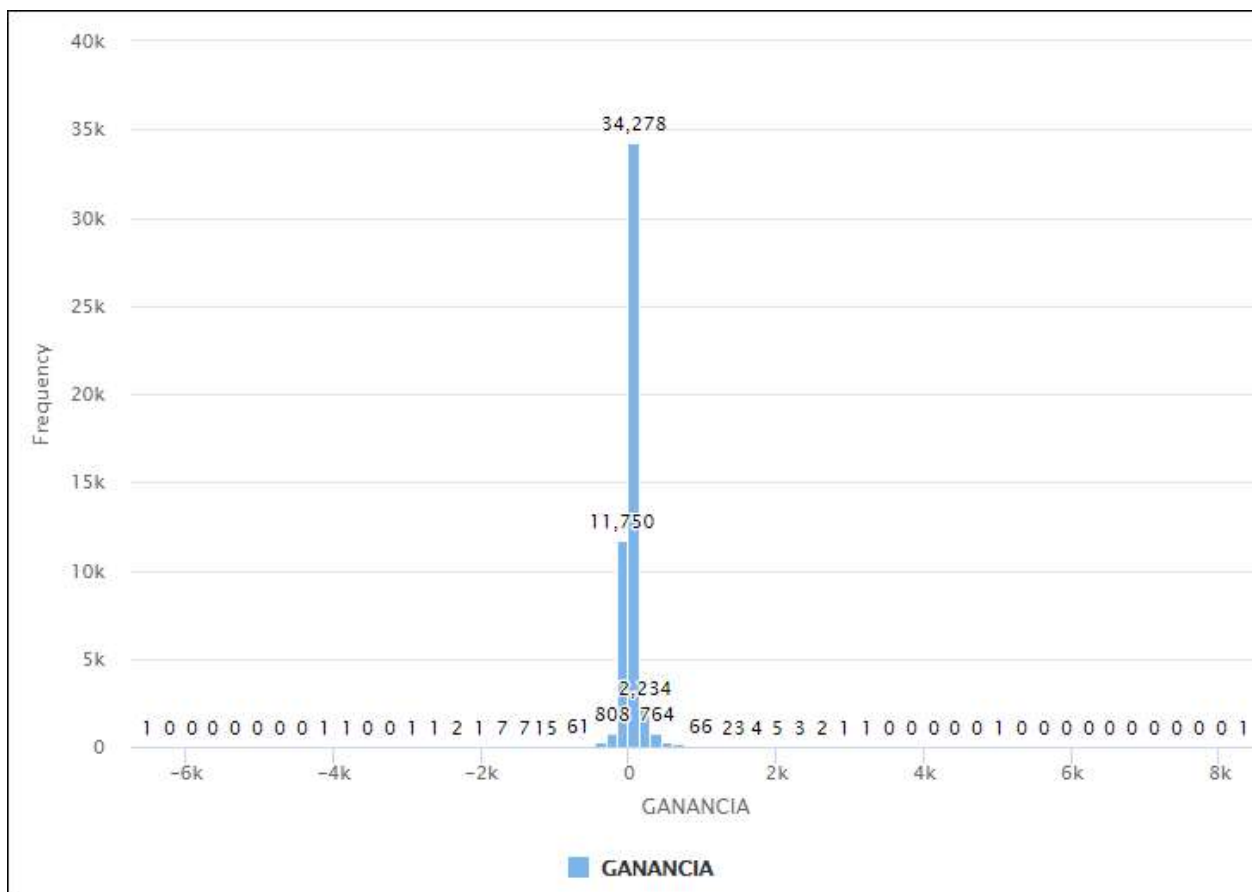
*Nota.* La figura detalla la concentración de los descuentos.

**GANANCIA:** Este campo representa la ganancia que se tiene por artículo vendido, en la figura 28 se observa que existe ganancia negativa, lo que indica que existe pérdida en la misma.



**Figura 28**

*Análisis del campo ganancia*

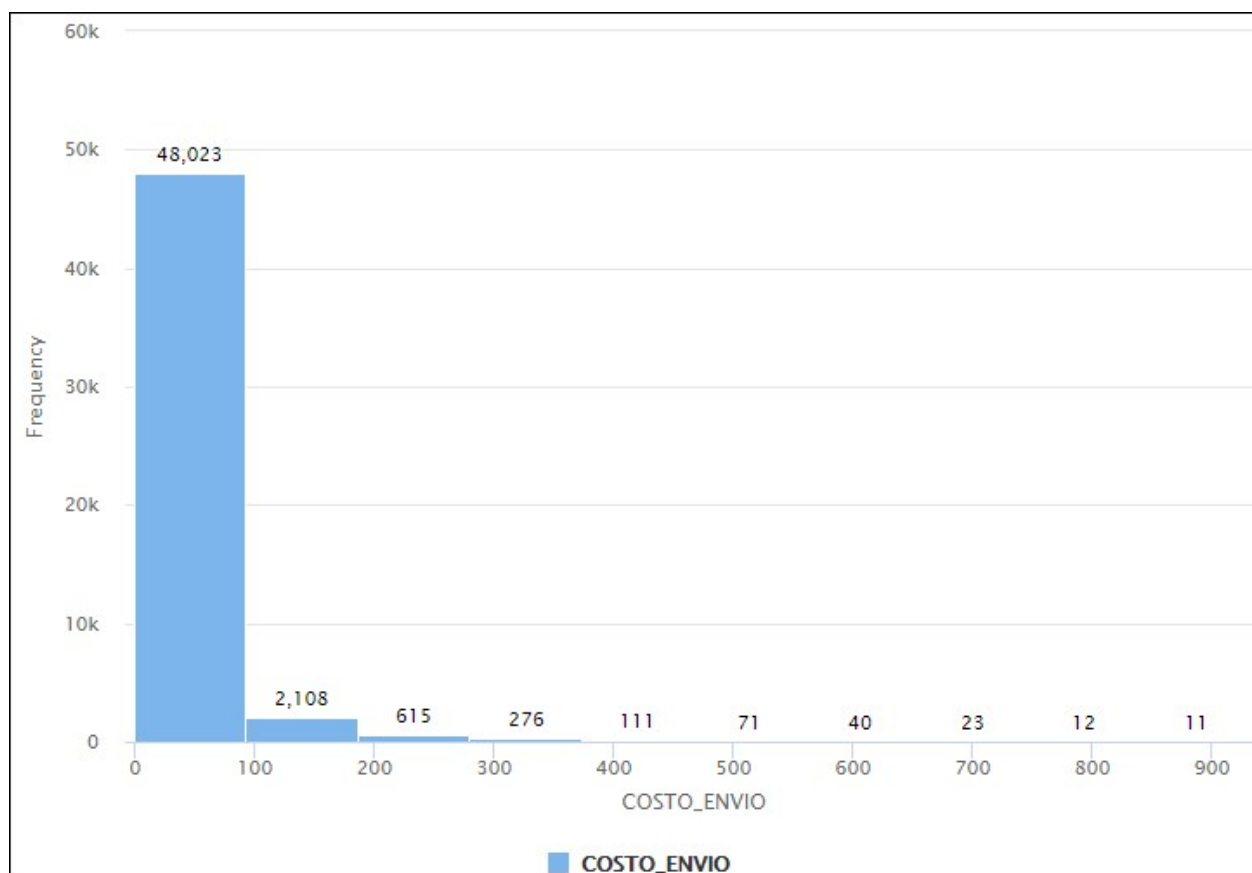


*Nota.* La figura detalla la concentración de la ganancia de los productos.

**COSTO\_ENVIO:** Este campo representa el valor del envío del artículo en una compra online. En la figura 29 se visualiza que la mayor concentración se encuentra en el primer intervalo de 0 a 100.

**Figura 29**

### Análisis del campo COSTO\_ENVIO

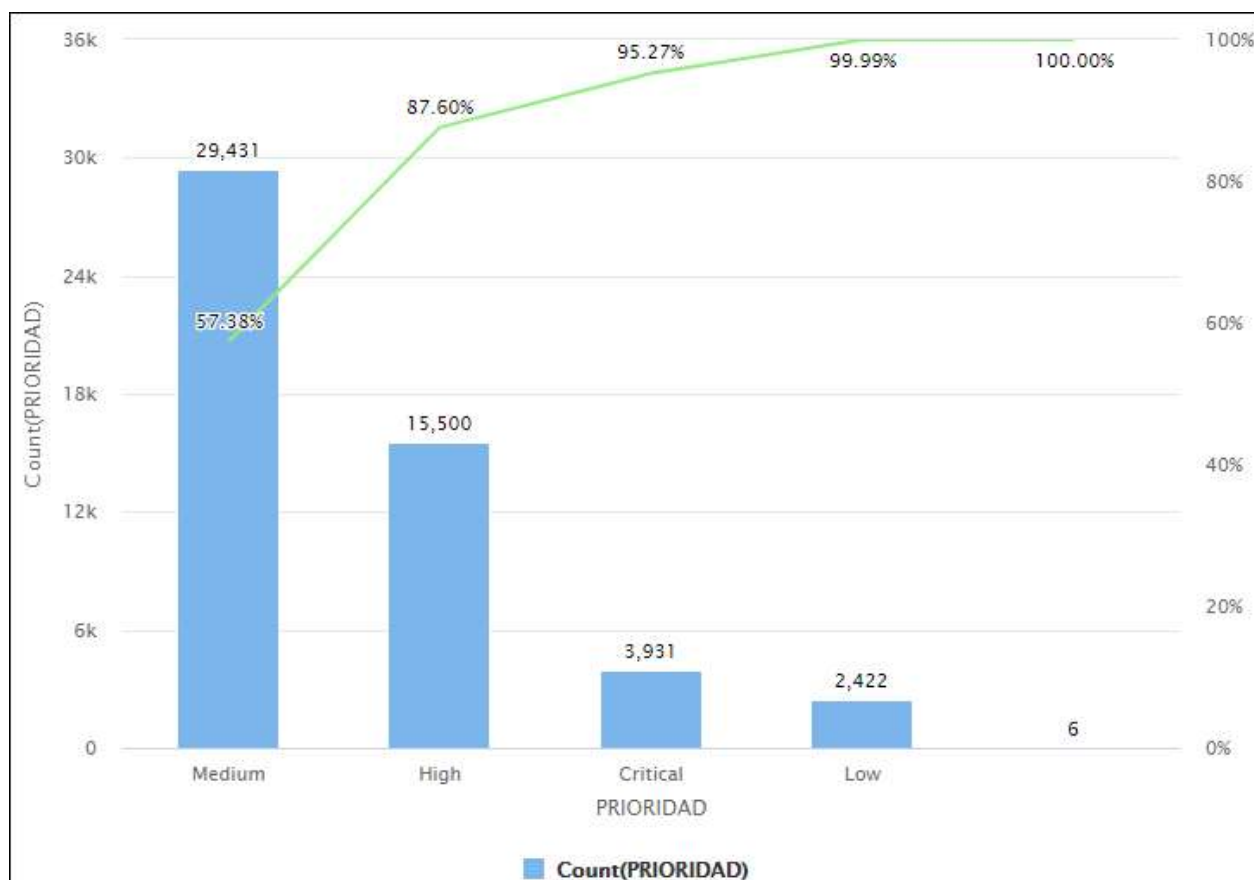


*Nota.* La figura detalla la concentración del costo de envío.

**PRIORIDAD:** Este campo representa la prioridad de la venta realizada por artículo. En la figura 30 se visualiza que el 57.38% de prioridad media, 30.22% prioridad alta, 7.67% de prioridad crítica, 4.72% prioridad baja, 0.01% no tiene prioridad.

**Figura 30**

### Análisis del campo Prioridad



*Nota.* La figura detalla la prioridad de venta de los productos.

### Métodos o técnicas de minería de datos

Una vez analizado cada campo de la tabla master se procede a segmentar por el algoritmo K-means, se escoge 4, 6 y 8 clústers que serán analizados.

En la figura 31 se visualiza las variables descartadas para la clasificación de los 4, 6 y 8 clústers que se van analizar.

- ID\_ORDEN
- FECHA\_VENTA

**Figura 31**

### Variables descartadas para el análisis

Selected	Status ↑	Quality	Name	Correlation	ID-ness	Stability	Missing	Text-ness
<input type="checkbox"/>	●		ID_ORDEN	?	48.81%	0.03%	0.00%	22.30%
<input type="checkbox"/>	●		FECHA_VENTA	?	?	0.25%	0.00%	0.00%

*Nota.* La figura detalla las variables descartadas.

Para el análisis de la data se ha realizado 3 agrupaciones, cada agrupación contiene 4, 6 y 8 clústers respectivamente.

### Agrupación de 4 Clústers

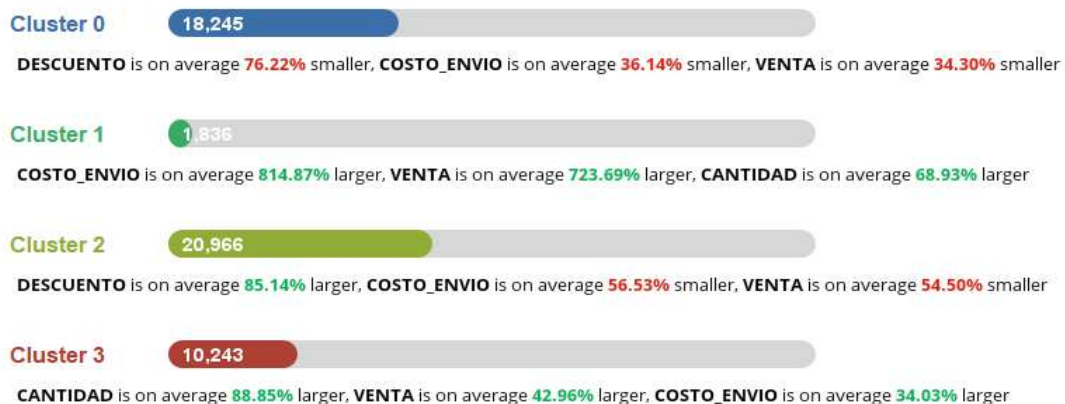
En la figura 32 se puede observar que la mayor concentración se encuentra en el clúster 2, seguido del clúster 0, clúster 3 y clúster 1

### Figura 32

#### 4 Clústers

#### k-Means - Summary

Number of Clusters: 4



*Nota.* La figura detalla los clústeres de la herramienta

En la figura 33 se presenta se presenta la tabla de los centroides para cada variable de la agrupación de 4 clústeres.

### Figura 33

*Tabla de Centroides de 4 Clústeres*

#### k-Means - Centroid Table

Cluster	CANTIDAD	CATEGORIA	CIUDAD	COSTO_E...	DESCUEN...	GANANCIA	PRIORIDAD	PROVINCIA	SUB_CAT...	SUCURSAL	VENTA
Cluster 0	2.626	0	2	16.841	0.034	33.225	4	2	14	13	161.908
Cluster 1	5.872	1	0	241.262	0.059	456.418	4	0	2	6	2030.018
Cluster 2	2.497	0	0	11.463	0.265	-16.537	4	0	8	6	112.127
Cluster 3	6.564	0	0	35.346	0.103	36.088	4	0	14	6	352.340

*Nota.* La figura detalla los centroides de los clústeres. Fuente: Figura propia

### Agrupación de 6 Clústeres

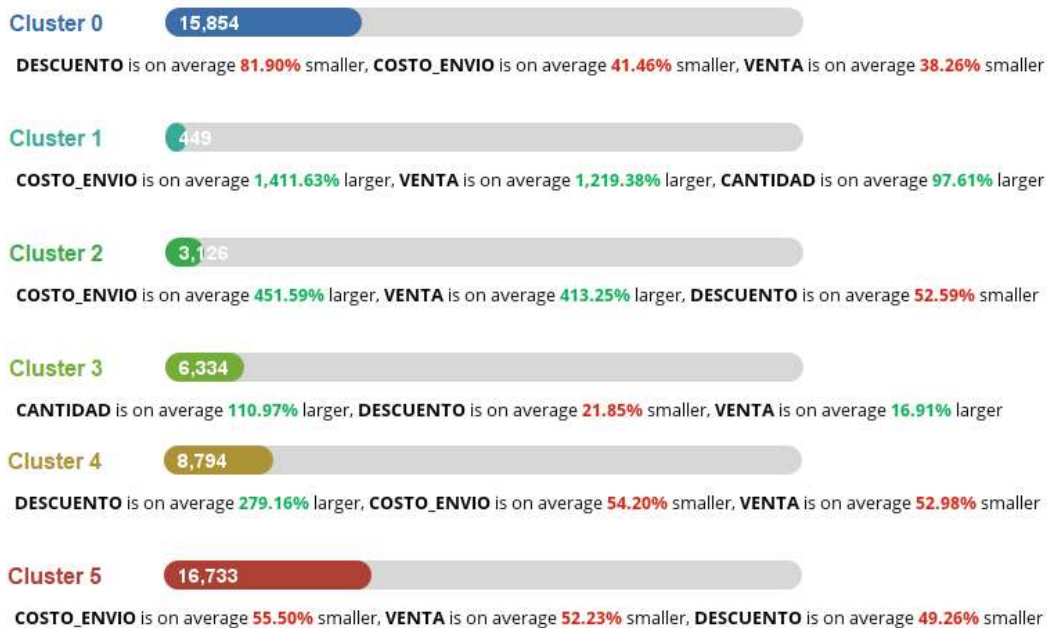
En la figura 34 se puede observar que la mayor concentración se encuentra en el clúster 5, seguido del clúster 0, clúster 4, clúster 3, clúster 2 y clúster 1

### Figura 34

## 6 Clústers

### k-Means - Summary

Number of Clusters: 6



*Nota.* La figura detalla los 6 clústeres.

En la figura 35 se presenta se presenta la tabla de los centroides para cada variable de la agrupación de 6 clústeres.

### Figura 35

*Tabla de Centroides de 6 Clústeres*

#### k-Means - Centroid Table

Cluster	CANTIDAD	CATEGORIA	CIUDAD	COSTO_E...	DESCUEN...	GANANCIA	PRIORIDAD	PROVINCIA	SUB_CAT...	SUCURSAL	VENTA
Cluster 0	2.781	0	2	15.439	0.026	31.755	4	2	14	13	152.170
Cluster 1	6.869	1	0	398.632	0.049	884.567	4	0	2	6	3251.661
Cluster 2	5.059	1	0	145.460	0.068	222.405	4	0	4	6	1264.919
Cluster 3	7.333	1	0	27.080	0.112	29.351	4	0	7	6	288.125
Cluster 4	3.000	0	2	12.077	0.542	-74.500	4	2	14	3	115.885
Cluster 5	2.537	0	0	11.736	0.073	20.346	4	0	14	6	117.723

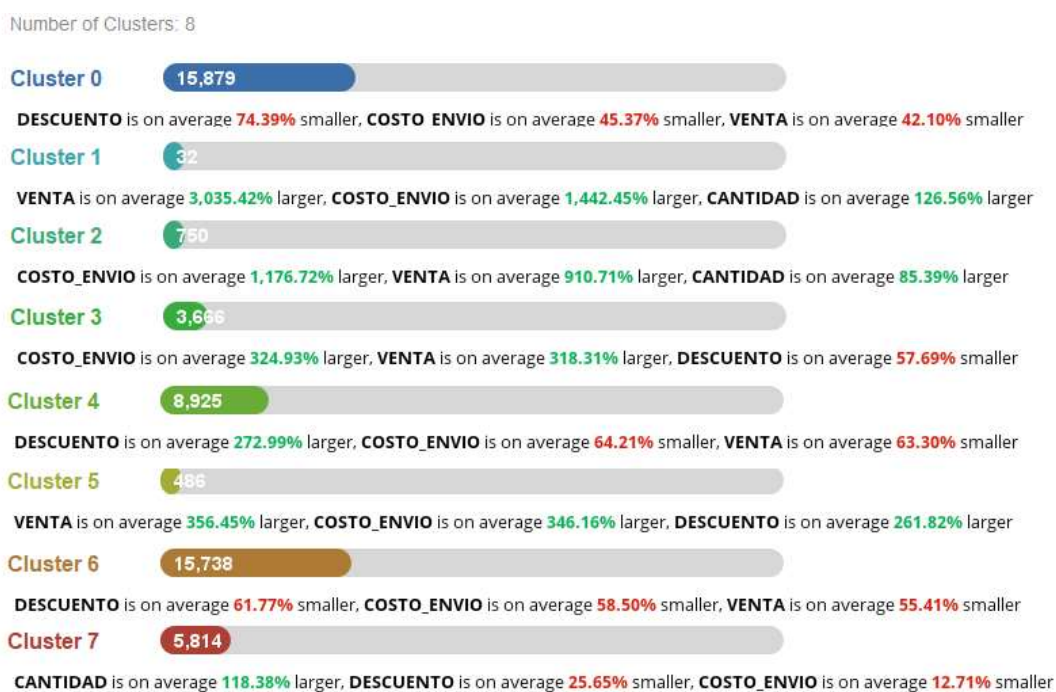
*Nota.* La figura detalla los centroides de 6 clústeres.

## Agrupación de 8 Clústers

En la figura 36 se puede observar que la mayor concentración se encuentra en el clúster 0, seguido del clúster 6, clúster 4, clúster 7, clúster 3, clúster 2, clúster 5 y clúster 1

### Figura 36

#### 8 Clústers



*Nota.* La figura detalla los 8 clústeres.

En la figura 37 se presenta se presenta la tabla de los centroides para cada variable de la agrupación de 8 clústeres.

## Figura 37

Tabla de Centroides de 8 Clústers

### k-Means - Centroid Table

Cluster	CANTIDAD	CATEGORIA	CIUDAD	COSTO_E...	DESCUEN...	GANANCIA	PRIORIDAD	PROVINCIA	SUB_CAT...	SUCURSAL	VENTA
Cluster 0	2.713	0	2	14.406	0.037	27.314	4	2	14	13	142.697
Cluster 1	7.875	2	0	406.759	0.050	2838.020	4	0	4	6	7727.333
Cluster 2	6.444	1	0	336.685	0.052	584.565	2	0	2	6	2490.919
Cluster 3	4.744	1	0	112.057	0.060	207.577	4	0	4	6	1030.943
Cluster 4	2.845	0	0	9.437	0.533	-47.272	4	0	14	6	90.443
Cluster 5	5.934	1	0	117.657	0.517	-718.404	4	0	13	6	1124.930
Cluster 6	2.561	0	0	10.944	0.055	20.958	4	0	8	6	109.896
Cluster 7	7.591	0	0	23.020	0.106	31.703	4	0	15	6	240.165

*Nota.* La figura detalla los centroides de los 8 clústeres

Una vez realizado la clusterización se analiza los resultados, obteniendo las 5 sub categorías más vendidos y las 5 menos vendidos.

Las sub categorías menos vendidos son:

- Ropa formal – mujer: Existen 3083 prendas vendidas con un número de 861 registros en la base distribuidos de la siguiente manera:
  - 82 Chaquetas vendidas
  - 524 Faldas vendidas
  - 647 Blusas vendidas
  - 1830 Vestidos vendidos.
- Ropa para adolescentes: Existen 4906 prendas vendidas con un número de 1486 registros en la base distribuidos de la siguiente manera:
  - 59 Shorts vendidos
  - 888 Ropa Interior vendidas
  - 890 sacos vendidos
  - 995 Chaquetas vendidas



- 1035 Jeans vendidos
- 1039 Camisetas vendidas
- Ropa deportiva – mujer: Existen 6078 prendas vendidas con un número de 1755 registros en la base distribuidos de la siguiente manera:
  - 336 Calentadores vendidos
  - 861 Camisetas vendidas
  - 1091 Shorts vendidos
  - 1604 Licras vendidas
  - 2196 Buzos capucha vendidos
- Ropa para Niños/Niñas: Existen 7454 prendas vendidas con un número de 2223 registros en la base distribuidos de la siguiente manera:
  - 1503 Ropa Interior vendida
  - 1538 pantalones vendidos
  - 1568 Camisetas vendida
  - 2845 Sacos vendidos
- Zapatos – mujer: Existen 8310 zapatos vendidos con un número de 2411 registros en la base distribuidos de la siguiente manera:
  - 28 Pantuflas vendidas
  - 91 Zapatos deportivos vendidos
  - 253 Botas vendidas
  - 1465 Zapatos Plataformas vendidas
  - 1536 Zapatos Tacones            vendidas
  - 1771 Zapatos Ballerinas vendidas
  - 3166 Zapatos tipo Sandalias vendidas

Las sub categorías más vendidos son:

- Ropa interior – hombre: Existen 21429 prendas vendidas con un número de 6152 registros en la base distribuidos de la siguiente manera:
  - 153 Pañuelos vendidos
  - 1692 Pijamas vendidas
  - 3775 Camisetas Interiores vendidas
  - 3998 Calzoncillos vendidos
  - 4777 Bóxeres vendidos
  - 7034 Medias vendidas
  
- Zapatos – hombre: Existen 16917 zapatos vendidos con un número de 5059 registros en la base distribuidos de la siguiente manera:
  - 48 Pantuflas vendidas
  - 216 Zapatos Urbanos vendidos
  - 248 Zapatillas vendidas
  - 3092 Zapatos básicos vendidos
  - 3399 Zapato casual vendidos
  - 3558 Zapatos deportivos vendidos
  - 6356 Botines vendidos
  
- Ropa Casual – hombre: Existen 16301 prendas vendidos con un número de 4883 registros en la base distribuidos de la siguiente manera:
  - 526 Pantalón gabardina vendidos
  - 1055 Chaquetas vendidas
  - 6081 Chalecos vendidos
  - 8639 Camisa Manga Corta vendidas
  
- Accesorios – hombre: Existen 12822 prendas vendidos con un número de 3538 registros en la base distribuidos de la siguiente manera:

- 259 Bolsos vendidos
  - 316 gorras vendidas
  - 1718 Billeteras vendidas
  - 2040 gafas vendidas
  - 3024 relojes vendidos
  - 5465 Correas vendidas
- Básico – mujer: Existen 12336 prendas vendidas con un número de 3434 registros en la base distribuidos de la siguiente manera:
- 30 overoles vendidos
  - 1102 camisetas vendidas
  - 2249 Buzos vendidos
  - 2342 Camisetas Cortas vendidas
  - 6613 Jeans vendidos

A continuación, se presenta los 5 más vendidos y los menos vendidos.

Los productos más vendidos se encuentran distribuidos en todas las sucursales, Navona, San Marino, Condado, Matriz, El Recreo, Mall del Pacifico, Portal Shopping, Scala Shopping, Quicentro Norte, Rotonda, San Luis Shopping, Quicentro Sur, Centro Policentro y son:

- Camisa Manga Corta: Existen 8639 prendas vendidas con un número de 2621 registros en la base.
- Medias – hombres: Existen 7034 prendas vendidas con un número de 2057 registros en la base
- Botines: Existen 6356 zapatos vendidos con un número de 1877 registros en la base
- Jeans – mujer: Existen 6613 prendas vendidas con un número de 1859 registros en la base

- Chaleco: Existen 6081 prendas vendidas con un número de 1838 registros en la base.
  - Los productos menos vendidos se encuentran en la sucursal Policentro y son:
- Pantuflas – M: Existen 28 zapatos vendidos con un número de 7 registros en la base.
- Overol: Existe 30 prendas vendidas con un número de 8 registros en la base
- Shorts – adolescentes: Existen 59 prendas vendidas con un número de 14 registros en la base.
- Pantuflas – H: Existen 48 zapatos vendidos con un número de 16 registros en la base
- Deportivos – mujeres: Existen 91 zapatos vendidos con un número de 18 registros en la base

Se realiza el análisis de la ganancia de los 5 productos con mayor y menor venta. En la tabla 8 se puede visualizar que los productos de menor venta generan menor ganancia y hasta presenta ganancia negativa es decir pérdida a la empresa, y los productos de mayor venta generan mayor ganancia a la empresa.

**Tabla 8**

*Ganancia de producto de mayor y menor demanda*

PRODUCTOS DE MENOR VENTA		PRODUCTOS DE MAYOR VENTA	
NOMBRE_PRODUCTO	GANANCIA	NOMBRE_PRODUCTO	GANANCIA
Pantuflas – M	-740.376	Camisa Manga Corta	30197.0449
Overol	80.98	medias – hombres	15222.1872
Shorts – adolescentes	2776.2637	Botines	44073.3079
Pantuflas – H	-236.5292	Jeans – mujer	87344.5795
deportivos – mujeres	580.719	Chaleco	25061.5286

*Nota.* La tabla detalla un resumen de los productos con mayor y menor ganancia

## Capítulo 4

### Resultados obtenidos

#### *Patrones obtenidos*

En el capítulo anterior, se realizó el análisis de las variables en donde se discriminó las que no aportaban en el análisis.

Los patrones encontrados con la data proporcionada para bajar el sobre stock en la empresa retail son:

- Los productos de menor venta se encuentran en la sucursal Policentro de la ciudad de Guayaquil, en esta sucursal se almacena la mercadería con mayor stock, lo que provoca que exista un mayor gasto en su almacenamiento. Con un nuevo análisis se debe verificar si los productos que se venden en esta sucursal pueden sustentar la permanencia de la tienda.
- Los productos de menor venta generan ganancia negativa, esto es una consecuencia de su sobre stock, debido que para su venta este producto pudo estar de promoción o descuento y no fue vendido. Los productos que se deben verificar su venta son: pantuflas, overoles, shorts para adolescentes, ropa deportiva para mujeres.
- Los productos se encuentran vendidos en fechas fuera de temporada, lo que implica que para el abastecimiento de estos productos se debe tomar en cuenta el clima o temporada en la que se encuentra la ciudad de la sucursal.
- Los productos menos vendidos se encuentran en la subcategoría adolescentes (shorts), básico – mujer (overol), zapatos – hombre (pantuflas), zapatos – mujer (deportivos, pantuflas).

Estas características se deben tomar en cuenta para el abastecimiento de los productos en especial en la sucursal Policentro.

Estos patrones de comportamiento se pudieron encontrar en un análisis de 4 años con variables básicas. Para un mejor resultado, más detallado, se debe incluir, las tallas, colores, modelo, entre otras, con lo que el resultado será más certero y evitar las pérdidas o gastos extras por el sobre stock de su mercadería.

### ***Análisis de los resultados***

Con la data cargada y la aplicación Rapidminer, se logró identificar los productos menos vendidos los mismos que generan pérdidas a la empresa, estos productos se encuentran la sucursal Policentro de la ciudad de Guayaquil y corresponden a pantuflas de mujer, overoles, shorts para adolescentes, pantuflas para hombre, zapatos deportivos para mujeres.

Las pantuflas de hombre y mujer son las que generan pérdidas negativas, es decir, estos productos pudieron ser vendidos con descuento, fuera de temporada y con alto sobre stock.

Los productos menos vendidos, son los que presentan sobre stock en la empresa retail, generando un alto gasto en el almacenamiento de dichos productos y baja rotación de inventario.

La cantidad de productos adquiridos en una venta está en un rango de 1 a 4 items.

Al no tener mayor información sobre los productos vendidos como talla, color, temporada, número de prendas en stock, el análisis realizado es de manera global con características generales como sucursal de venta, mes de venta y ganancia que generó en su venta.

## Capítulo 5

### Conclusiones y recomendaciones

#### *Conclusiones*

- Se identificó las causas que tenía la empresa sobre el proceso de abastecimiento de productos, provocando el sobre stock de algunos de sus productos, entre ellas están los productos fuera de temporada y en locales con menos afluencia de clientes, entre otras.
- Para un mejor análisis predictivo, se realizó el estudio del arte, para poder determinar los métodos y técnicas para encontrar los patrones idóneos para la solución al problema, así como número de stock en prendas, tallas, color, marca, temporada, sin embargo, con la información brindada se pudo identificar los patrones macros que permiten evitar el sobre stock.
- La herramienta Rapidminer, permite minar datos, archivos de texto para un análisis predictivo, inteligencia de negocios o segmentación de información. La versión utilizada en este proyecto, es académica, la cual tiene una vigencia de un año, la misma que tiene todos modelos activos y no hubo restricción en la data procesada como sucede con la versión libre que limita el número de registros a procesar.
- La data cargada fue validada por la herramienta Talend Open Studio, la cual permite la corrección de errores con los valores nulos o en blanco cargados a la base de datos. La versión utilizada en este proyecto es libre, la cual permite al usuario explorar la aplicación y conocer sus funciones para la valoración de la data.
- En la entrega de la data por parte de la empresa, fue limitado en el número de variables, por lo que el análisis no es más detallado en el abastecimiento de productos y así evitar

el sobre stock, sin embargo, se tiene un patrón macro para el abastecimiento de sus productos retail de moda.

- Los resultados obtenidos, fueron de validez para la empresa y se demostró que un sistema de minería de datos ayuda a la toma de decisiones, no solo en el sobre stock de los productos, sino también se puede incluir otras variables para identificar nuevos patrones y aumentar el valor en su análisis.

### **Recomendaciones**

- Se recomienda implementar procesos automáticos que verifiquen la integridad de los datos, así, al realizar la minería de los datos no se volverá a realizar nuevamente su validación
- Se recomienda realizar este análisis incluyendo mayor cantidad de variables para determinar con más exactitud los productos con sobre stock y a su vez generen nuevos conocimientos y una mejor predicción sobre el stock de los productos.
- Se recomienda tener personal que trabaje con la data y genere conocimiento a través de la minería de datos, con lo cual se puede realizar predicciones para aumentar los ingresos a la empresa o disminuir los gastos innecesarios como el almacenamiento de productos que no se venden.
- Supervisar periódicamente que las variables sean cargadas por parte del personal de facturación o de inventario para poder tener una data útil para el análisis.
- Si no se dispone de mayor información sobre los productos (variables de estudio), se debería realizar una re ingeniería de datos para incluir información necesaria para el análisis predictivo.
- Es indispensable que se realice este procedimiento cada año con nueva información cargada y actualizada para tener una mejor predicción sobre el comportamiento de



las ventas de los productos de moda y así disminuir paulatinamente el sobre stock de sus productos.

## Bibliografía

- Analítica de Retail [15] . (18 de diciembre de 2018). Obtenido de <http://analiticaderetail.com/tasa-de-descuento-de-un-producto/>
- Badoc, I. [. (19 de Abril de 2018). *Generixgroup*. Obtenido de <https://www.generixgroup.com/es/blog/indicadores-gestion-inventarios>
- Córdova, G. [. (10 de febrero de 2011). *Única 360*. Obtenido de <https://www.unica360.com/analisis-rfm-en-retail-empezando-a-segmentar-clientes-i>
- Datanet [8] . (s.f.). Obtenido de <https://datanetconsultores.es/stock-almacen-por-que-es-importante-su-optima-gestion/>
- ETATEX, C. (Abril de 2019). *Etafashion.com* [2]. Obtenido de <https://www.etafashion.com/empresa>
- IEEE [11] . (2019). Obtenido de <https://ieeexplore.ieee.org/Xplore/home.jsp>
- Krall, C. [. (2019). *Aprender a Programar*. Obtenido de [https://www.aprenderaprogramar.com/index.php?option=com\\_content&view=article&id=258:mineria-de-datos-2o-parte-modelos-tecnicas-herramientas-dv00106a&catid=45&Itemid=164](https://www.aprenderaprogramar.com/index.php?option=com_content&view=article&id=258:mineria-de-datos-2o-parte-modelos-tecnicas-herramientas-dv00106a&catid=45&Itemid=164)
- León, E. P. (s.f.). Minería de Datos.
- Logicalis [23] . (12 de enero de 2015). Obtenido de <https://blog.es.logicalis.com/analytics/modelos-de-data-mining-y-las-herramientas-mas-usadas>
- Maturana, J. [. (26 de mayo de 2017). *EbizLatam*. Obtenido de <http://www.ebizlatam.com/la-creacion-valor-cliente-maslow-la-piramide-valores-del-consumidor/>
- Microsoft [3]. (8 de enero de 2019). Obtenido de <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts?view=sql-server-2017>
- Microsoft [4] . (7 de mayo de 2019). Obtenido de <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/mining-models-analysis-services-data-mining?view=sql-server-2017>

- Nardi, G. [. (2015). *Desnudando el Marketing*. Obtenido de <https://desnudandoelmarketing.com/tipos-de-retail/>
- Pérez, J., & María, M. [. (2016). *Definicion*. Obtenido de <https://definicion.de/retail/>
- Perú Retail [1] . (9 de Julio de 2018). Obtenido de <https://www.peru-retail.com/ecuador-ingreso-retailers-de-moda/>
- Rams, M. [. (13 de junio de 2016). Obtenido de <https://blog.elogia.net/rfm-recency-frecuency-y-money-qu%C3%A9-valor-tiene-nuestro-cliente>
- Revista Enfasis - Logistica [27]. (12 de Agosto de 2019). Obtenido de <http://www.logisticasud.enfasis.com/notas/84162-la-cadena-del-fashion-retail>
- Revista Lideres [14] . (noviembre de 2014). Obtenido de 11: <https://www.revistalideres.ec/lideres/retail-ecuador-ventas-centros-comerciales.html>
- Riaño, P. [. (15 de Junio de 2020). *Klarna for business*. Obtenido de <https://www.modaes.es/back-stage/el-stock-condena-a-la-moda-al-aburrimiento-un-ano-sin-innovacion-para-salvar-las-finanzas.html>
- Sage, H. (19 de diciembre de 2017). *ControlGroup [10]*. Obtenido de <https://blog.controlgroup.es/cobertura-de-stock/>
- Soto, M. [. (13 de diciembre de 2016). *WIVO*. Obtenido de <https://blog.wivoanalytics.com/consejos-para-una-gesti%C3%B3n-de-stocks-efectiva-en-retail>
- Soto, M. [. (25 de abril de 2017). *WIVO*. Obtenido de <https://blog.wivoanalytics.com/cu%C3%A1l-es-el-mejor-indicador-de-rendimiento-en-retail>
- Turienzo, L. [. (14 de Octubre de 2017). *RetailNewsTrends*. Obtenido de <https://retailnewstrends.me/que-es-el-retail-2/>
- Urdaneta, A. [. (21 de enero de 2016). *WIVO*. Obtenido de <https://blog.wivoanalytics.com/conoce-tus-tiendas-top-11-kpis-para-exito-en-retail>

Turienzo, L. (14 de Octubre de 2017). *RetailNewsTrends*. Obtenido de <https://retailnewstrends.me/que-es-el-retail-2/> [9]