

## Proyecto de Investigación

Diseño de sistema de segmentación de clientes a través de machine learning para aplicar estrategias de reactivación y fidelización de clientes.

Elaborado por:

Carlos Andrés Zambrano Betancourt

Programa:

Especialización Machine Learning

Universidad Ean

Escuela de Formación en Investigación

Seminario de Investigación de Pregrado

Bogotá

21/05/2024

## Resumen

La falta de una herramienta de segmentación de clientes para dirigir los esfuerzos de marketing adecuadamente genera como resultado desconocimiento de necesidades y preferencias de los clientes, falta de personalización en la comunicación, asignación imprecisa de recursos en las estrategias de marketing y pérdida de oportunidades de ventas cruzadas. Este proyecto tiene como objetivo diseñar un sistema de segmentación utilizando Machine Learning para optimizar el plan de fidelización y retención de los clientes de Diverplaza.

Palabras clave: segmentación de clientes, machine learning, estrategias de retención, personalización en la comunicación, ventas cruzadas.

## **Problema de Investigación**

### **Descripción del Problema:**

En el contexto de los centros comerciales en Colombia, la falta de un sistema de segmentación de clientes representa una oportunidad de mejora significativa frente a la competencia. Esta situación se ha identificado como un obstáculo para la optimización de la rentabilidad y la fidelización de los clientes de DIVERPLAZA. La competencia cada vez más intensa y las expectativas cambiantes de los consumidores han puesto de manifiesto la necesidad de estrategias más sofisticadas de gestión de la experiencia del cliente.

La ausencia de un sistema de segmentación de clientes afecta directamente la capacidad de DIVERPLAZA para adaptarse a las preferencias y necesidades individuales de sus clientes, lo que conduce a una menor satisfacción, lealtad del cliente y disminución de las oportunidades de venta cruzada.

### **Origen del Problema:**

La ausencia de una herramienta efectiva para dirigir los esfuerzos de marketing genera una falta de comprensión profunda de las necesidades y comportamientos de los clientes. El no aprovechamiento de las fuentes de datos disponibles (APP de fidelización, Portal Cautivo, Tráfico centro comercial, Tráfico parqueaderos) dificulta la personalización de las estrategias de marketing y la identificación de oportunidades de ventas cruzadas.

## **Síntomas o Situaciones Anómalas:**

Este problema se manifiesta en la comunicación genérica y no personalizada con los clientes, falta de estrategias personalizadas según el tipo de cliente, fallas en el sistema de retención de clientes de alto valor, ineficacia en la maximización de los ingresos y la falta de dirección estratégica en la gestión del tráfico hacia el centro comercial, todo esto conduce a la pérdida de oportunidades de venta y a una experiencia del cliente poco satisfactoria.

## **Pronóstico de la Situación:**

La falta de personalización en la comunicación con los clientes y la incapacidad para identificar las necesidades y tendencias individuales de estos generarán disminución en la lealtad de los clientes de alto valor, una menor rentabilidad y pérdida de competitividad en el mercado actual.

## **Control Pronóstico:**

El diseño de un sistema de segmentación de clientes basado en machine learning proporcionara una solución al problema al permitir una comprensión de los comportamientos y preferencias de los clientes como lo manifiesta America Retail “el uso de big data, análisis de datos y estadísticas permite pronosticar de mejor manera las decisiones de compra de las personas, ya que, por ejemplo, al brindar información a los consumidores basada en sus decisiones previas.” (Retail, 2023). Este enfoque permitirá una comunicación asertiva y personalizada, aplicación de estrategias de captación, fidelización y retención específicas, identificación de oportunidades de venta cruzada y una mejor gestión del tráfico hacia el centro comercial.

## **Pregunta de Investigación:**

¿Cuál es la mejor herramienta de machine learning para diseñar un sistema que permita segmentar los clientes del centro comercial de manera efectiva, con el fin de fortalecer las estrategias de retención y maximizar el valor del cliente?

## **Objetivos**

### **Objetivo general**

Diseñar un sistema de segmentación de clientes a través de machine learning para fortalecer las estrategias de retención de clientes y maximizar el valor del cliente.

### **Objetivos específicos**

- Evaluar los modelos de machine learning existentes para la segmentación de clientes.
- Seleccionar el modelo de segmentación que más se ajuste acuerdo con las necesidades del centro comercial.
- Identificar y recopilar los datos relevantes de los clientes, incluyendo características demográficas, comportamientos de compra y preferencias según el modelo seleccionado.
- Desarrollar el modelo seleccionado para segmentar de manera efectiva la base de clientes en grupos homogéneos según criterios predefinidos.
- Evaluar la precisión y efectividad de los modelos de segmentación de clientes utilizando métricas de desempeño adecuadas.

## Justificación

El estudio es conveniente debido a la creciente importancia del marketing personalizado teniendo en cuenta la constante evolución de los clientes, de sus necesidades y comportamientos. La falta de herramientas efectivas para dirigir los esfuerzos de marketing puede resultar en la pérdida de oportunidades de ventas cruzadas y la insatisfacción de los clientes por estrategias de comunicación genéricas que no generan valor. Tener un sistema efectivo de segmentación de clientes permite “ubicar al cliente en el centro del negocio, dando una oportunidad no sólo de perfilar sus gustos e intereses, sino también de identificar el rumbo al que se deben enfocar los negocios a través de un apoyo libre de sesgos” (Retail, 2023).

La investigación aborda una problemática relevante en el ámbito empresarial, ya que la retención de clientes y la maximización del valor de estos son aspectos cruciales para la sostenibilidad y el crecimiento de las organizaciones. Mejorar la comunicación con los clientes y ofrecerles experiencias personalizadas puede contribuir a fortalecer la relación empresa-cliente y mejorar la satisfacción del cliente cómo se evidencia en el estudio, denominado ‘Sales and Marketing Insights Report:

Se basó en encuestas cualitativas a 100 perfiles directivos de grandes empresas globales, en el cual afirman que, en términos de tecnología, la inteligencia artificial (IA) emerge como un factor clave para habilitar ventas y generar experiencias personalizadas para los clientes, según el 60% de los encuestados. Además, el análisis de datos y el big data son considerados instrumentos esenciales para mejorar la toma de decisiones por el 25% de los participantes. (González, 2024).

El estudio tiene importantes implicaciones prácticas para DIVERPLAZA, ya que proporcionará una herramienta de segmentación de clientes basada en machine learning que le permitirá obtener una comprensión más profunda de los comportamientos y necesidades de los clientes y dirigir de manera efectiva los esfuerzos del área de marketing. Esto puede conducir a la obtención de mejores resultados con los planes de captación, retención y fidelización de los clientes, el aumento de los ingresos y una mejor reputación de la marca a través de experiencias personalizadas marcando un factor diferencial frente a la competencia.

El estudio evaluará metodologías avanzadas de análisis de datos y machine learning para desarrollar el modelo de segmentación de clientes bajo la metodología que más se ajuste a las necesidades de DIVERPLAZA y también teniendo en cuenta los recursos existentes (APP de fidelización, Portal Cautivo, Tráfico centro comercial, Tráfico parqueaderos) para el desarrollo del proyecto. Esto proporcionará un enfoque innovador y riguroso para abordar el problema objeto de estudio.

En cuanto al campo, grupo y línea de investigación, el estudio se enmarca en el campo de Ciencia, Tecnología e Innovación, específicamente en el grupo de Ciencias Básicas y la línea de investigación en Estadística aplicada y Ciencia de Datos. Esto asegura que el estudio esté alineado con los objetivos institucionales y contribuya al desarrollo del conocimiento en el área de marketing estratégico, gestión empresarial y machine learning.

## **Marco Teórico**

### **Estado del arte**

#### **Introducción**

La segmentación de clientes es uno de los principales herramientas para generar campañas de marketing efectivas, lo cual permite a las empresas comprender las necesidades y preferencias de sus clientes con la finalidad de ofrecer productos y servicios adaptados a sus características específicas. En el contexto de los centros comerciales, la segmentación de clientes juega un papel crucial en la mejora de las estrategias de marketing para optimizar las estrategias de captación, fidelización, reactivación de clientes y la promover de la venta cruzada.

#### **Métodos de segmentación**

Existen muchos métodos de segmentación que han sido empleados a lo largo del tiempo, como lo son el análisis de cluster (Jain & Dubes, 1988), análisis discriminante (Johnson & Wichern, 2007), árboles de decisión (Breiman et al., 1984), técnicas de aprendizaje automático como regresión logística (Hosmer Jr., Lemeshow , & Rodney , 2013) y redes neuronales (Bishop, 2006). Estos métodos permiten identificar grupos homogéneos de clientes con características similares, con la finalidad de generar estrategias de marketing personalizadas.

#### **Variables utilizadas**

Las variables utilizadas para segmentar clientes en diferentes industrias incluyen características demográficas (Kotler & Armstrong, 2008), comportamiento de compra (Lee, Hosanagar, & Harikesh, 2018) y segmentación de los clientes según las tres variables del comportamiento: última transacción, frecuencia y valor monetario (Hair Jr., Bush, & Ortinau,

2010), la combinación de las variables citadas anteriormente permite realizar una segmentación más precisa y efectiva.

## **Aplicaciones prácticas**

Los estudios han demostrado que la segmentación de clientes puede mejorar significativamente la efectividad de las estrategias de marketing y promoción. Al comprender mejor las necesidades y preferencias de los diferentes segmentos de clientes, los centros comerciales pueden ofrecer experiencias más personalizadas, aumentar la fidelidad del cliente y promover la venta cruzada entre diferentes áreas del centro comercial.

## **Desafíos y tendencias futuras**

A pesar de los beneficios de la segmentación de clientes, existen desafíos en su implementación, como la recopilación y gestión de datos, la privacidad del cliente y la interpretación de los resultados (Fader, Hardie, & Lee, 2004). Además, las tendencias en segmentación de clientes incluyen el uso de tecnologías como la inteligencia artificial y el big data para ofrecer experiencias aún más personalizadas.

## **Marco Teórico**

En este estudio, se abordan los siguientes ejes conceptuales: dirigir efectivamente los esfuerzos de marketing, personalización en la comunicación con los clientes, ventas cruzadas, retención de clientes y maximización del valor del cliente.

Para abordar la falta de personalización en la comunicación con los clientes, (Hughes, Hagander, & Pal, 2016) destacan la importancia de utilizar técnicas de segmentación de clientes basadas en el análisis de datos para adaptar los mensajes de marketing a las preferencias individuales de los clientes. Además, la investigación (Smith & Jones, 2019) subraya la necesidad de implementar estrategias de marketing relacional que fomenten la lealtad del cliente a través de interacciones personalizadas y relevantes.

En cuanto a las oportunidades de ventas cruzadas y ventas adicionales, estudios como el de (DOĞAN, AYÇİN, & BULUT, 2018) sugieren que la segmentación efectiva de clientes puede facilitar la identificación de oportunidades de ventas adicionales al revelar patrones de comportamiento de compra y preferencias. Además, la investigación (García & Martínez, 2017) destaca la importancia de la recomendación personalizada de productos como una estrategia efectiva para aumentar las ventas cruzadas.

Para abordar la retención de clientes y maximización del valor del cliente, (Kotler & Armstrong, 2018) proponen un enfoque centrado en el cliente que involucre la gestión proactiva de la satisfacción del cliente, la calidad del servicio y la experiencia del cliente.

## **Marco institucional**

### **Descripción de la Organización**

El Centro Comercial "DIVERPLAZA" se encuentra ubicado en la ciudad de Bogotá, Colombia y pertenece al sector del Comercio Minorista (CIU 4711).

DIVERPLAZA se destaca por su diversidad de tiendas y servicios que atienden a una amplia gama de clientes. Entre sus principales nichos de mercado se encuentran los segmentos de bienes y servicios de conveniencia, moda, tecnología, entretenimiento y gastronomía.

### **Principales Productos y Servicios**

El centro comercial ofrece una variedad de productos y servicios, incluyendo ropa, calzado, electrodomésticos, dispositivos electrónicos, alimentos y bebidas, servicios de belleza y entretenimiento para todas las edades.

### **Estructura Organizacional**

DIVERPLAZA opera bajo una estructura organizacional jerárquica, con un equipo directivo encabezado por un gerente de Activos, quien supervisa las operaciones diarias del centro comercial. El personal se distribuye en departamentos clave, como tecnología, mercadeo (Marketing, atención al cliente, fidelización), comercial, operaciones (Seguridad, Mantenimiento, SST, Parqueaderos).

## **Procesos**

Los procesos en DIVERPLAZA están diseñados para garantizar una experiencia óptima para los visitantes. Esto incluye variedad de tiendas y nichos, la planificación y ejecución de eventos y promociones, la seguridad de las instalaciones, la atención al cliente y la gestión de relaciones con los arrendatarios.

El Centro Comercial DIVERPLAZA se ha convertido en un destino emblemático para las compras y el entretenimiento en el sector donde se encuentra ubicado, atrayendo a miles de visitantes cada semana con su amplia oferta de productos, servicios y experiencias.

## **Metodología**

### **Primer nivel**

#### **Enfoque, alcance y diseño de la investigación**

El enfoque de investigación será mixto, combinando elementos cuantitativos y cualitativos para obtener una comprensión completa del problema de falta de segmentación de clientes en DIVERPLAZA. Desde una perspectiva cuantitativa, se recolectarán datos numéricos sobre la frecuencia de compra (qué tantas veces compran productos del centro comercial), el monto gastado (cuánto gastan en las compras realizadas en el centro comercial) y la recencia (Tiempo transcurrido desde la última compra hasta el momento de realizar el modelo de segmentación), adicionalmente desde una perspectiva cualitativa, se recolectarán datos del género.

El alcance de la investigación se limitará a DIVERPLAZA y sus clientes, centrándose en comprender las necesidades, preferencias y comportamientos de estos. No se incluirán otros centros comerciales ni se ampliará el estudio a otros sectores.

El diseño de investigación será experimental y transversal. Se realizará intervención y manipulación de variables, con la finalidad de garantizar la fiabilidad de la información para el desarrollo de la investigación.

## **Definición de Variables**

Las variables a definir son fundamentales para establecer el marco de la investigación y comprender los aspectos específicos que se analizarán en relación con la falta de segmentación de clientes en DIVERPLAZA. Se abordan conceptual y operacionalmente de la siguiente manera:

### **1. Características demográficas:**

**Definición conceptual:** Estas variables representan atributos sociodemográficos de los clientes, como edad y género. Sirven para comprender la composición demográfica de la base de clientes de DIVERPLAZA y cómo influye en sus comportamientos de compra y preferencias.

**Definición operacional:** La captación de datos se llevará a cabo mediante la APP de fidelización del centro comercial, la cual permite a los usuarios obtener puntos redimibles por bonos y premios del centro comercial a cambio de registrar las facturas de las compras realizadas, es importante resaltar que para hacer uso de la APP los clientes de DIVERPLAZA

deben registrarse en esta indicando nombres, apellidos, edad, género, fecha de nacimiento, celular, correo electrónico y documento de identificación.

## 2. Comportamiento de compra:

**Definición conceptual:** Estas variables reflejan los patrones y hábitos de compra de los clientes, basado en las facturas registradas en la APP de fidelización, mediante las cuales se puede identificar la frecuencia de compra, el monto gastado por visita y las tiendas en la que realizan compras con mayor frecuencia, entre otros. Son importantes para entender cómo los clientes interactúan con DIVERPLAZA y los productos o servicios de que tiendas prefieren.

**Definición operacional:** Se obtendrán a partir del análisis de datos de las facturas registradas en la APP por los clientes de DIVERPLAZA. Mediante estas facturas se identificará la frecuencia de compra de los clientes, el monto total gastado y las tiendas en las que más realizan compras.

## 3. Preferencias:

**Definición conceptual:** Estas variables representan las preferencias y gustos de los clientes relacionados con las tiendas. Son importantes para personalizar las estrategias de marketing y mejorar la experiencia del cliente.

**Definición operacional:** Se identificarán mediante las facturas registradas en la APP de fidelización.

## **Población y Muestra**

La población objetivo de este estudio comprende todos los clientes registrados en la aplicación de fidelización del centro comercial DIVERPLAZA durante el período de investigación. Dado que los datos provienen de la aplicación de fidelización y se utilizará la totalidad de ellos, no se realizará un proceso de muestreo, ya que se trabajará con la población completa disponible en la base de datos de la aplicación.

### **Características de la población:**

**Número de individuos:** La cantidad exacta de clientes que registran facturas en la APP de fidelización de DIVERPLAZA varía según el día, la temporada y otros factores. El número total de clientes registrados en la aplicación de fidelización de DIVERPLAZA puede variar con el tiempo, actualmente hay 11.897 usuarios activos.

**Demografía:** La población de clientes de DIVERPLAZA es diversa en cuanto a edad, contando con clientes entre los 18 y 65 años, género (M/F), nivel socioeconómico (2, 3, 4), adicionalmente es importante resaltar que la mayoría de los clientes son residentes locales de Alamos norte y sus alrededores, pero también hay visitantes de otros sectores aledaños.

### **Procedimiento de selección de la muestra:**

Dado que se utilizará la totalidad de los datos capturados en la aplicación de fidelización, no se aplicará un procedimiento de selección de muestra. Se trabajará con todos los registros disponibles dentro del período de tiempo definido para el estudio.

Este enfoque garantiza la representatividad de la muestra y permite analizar de manera exhaustiva las características y comportamientos de los clientes registrados en la aplicación de fidelización de DIVERPLAZA.

## **Segundo nivel**

### **Selección de métodos o instrumentos para recolección de información**

Para la recolección de información, se utilizarán los datos recopilados de la aplicación de fidelización del centro comercial DIVERPLAZA. Estos datos incluirán información demográfica de los usuarios, comportamientos de compra, preferencias de tiendas y frecuencia de compras entre otros.

### **Métodos de recolección de datos:**

**Registro de la aplicación de fidelización:** Los datos de los usuarios se capturan automáticamente cuando utilizan la aplicación de fidelización de DIVERPLAZA. Estos datos incluyen información demográfica proporcionada por los usuarios al registrarse, así como registros de transacciones de compra, interacciones con promociones y redención de premios.

### **Instrumentos de recolección de datos:**

**Base de datos de la aplicación de fidelización:** La base de datos de la aplicación de fidelización del centro comercial DIVERPLAZA servirá como el principal instrumento para la recolección de datos. Esta base de datos contendrá registros estructurados de todas las interacciones de los usuarios con la aplicación, lo que proporcionará una amplia gama de información para el análisis.

## **Proceso de recolección de datos:**

Los datos se recopilarán de forma continua a medida que los usuarios interactúen con la aplicación de fidelización, es importante resaltar que el centro comercial cuenta con procedimientos para garantizar la integridad y seguridad de los datos recopilados, así como el cumplimiento de las regulaciones de protección de datos.

## **Validación de datos:**

Se realizará un proceso de validación de datos para verificar la calidad y consistencia de la información recopilada. Esto incluirá la detección y corrección de errores, la eliminación de duplicados y la identificación de valores atípicos.

Este enfoque garantiza la obtención de datos confiables y completos que permitirán realizar un análisis exhaustivo de las características y comportamientos de los clientes de DIVERPLAZA, lo que a su vez facilitará la segmentación y personalización de las estrategias de marketing.

## **Técnicas de análisis de datos**

Para obtener una comprensión profunda del comportamiento y las preferencias de los clientes, se seleccionan cuidadosamente las técnicas de análisis de datos más apropiadas lo que permitirá una segmentación efectiva y la creación de estrategias de marketing personalizadas.

Teniendo en cuenta lo anterior a continuación se relacionan las técnicas de análisis que se utilizarán para analizar la información recopilada de la APP de fidelización. Estas técnicas

abarcan desde el análisis estadístico descriptivo hasta modelos predictivos avanzados, con el objetivo de proporcionar una visión completa y detallada del perfil y comportamiento de los clientes del centro comercial.

## Técnicas de análisis de datos

Instrumento	Técnica de análisis	Descripción
Base de datos de la aplicación de fidelización	Análisis estadístico descriptivo	Se realizará un análisis estadístico descriptivo para examinar la distribución de variables, calcular medidas de tendencia central y dispersión, identificar valores atípicos y realizar análisis comparativos entre diferentes grupos de clientes. Esto proporcionará una comprensión inicial de las características de los clientes y sus comportamientos.
Base de datos de la aplicación de fidelización	Segmentación de clientes mediante técnicas de clustering	Se aplicarán técnicas de clustering, como: <ul style="list-style-type: none"> <li>- K-means</li> <li>- DBSCAN</li> <li>- Análisis RFM (Recencia, Frecuencia, Monto)</li> </ul> Todo esto con la finalidad de segmentar a los clientes en grupos homogéneos con características similares y elegir el mejor modelo, lo que permitirá identificar patrones de comportamiento y preferencias dentro de la base de clientes, lo que facilitará la personalización de las estrategias de marketing.
Base de datos de la aplicación de fidelización	Análisis de asociación (reglas de asociación)	Se emplearán técnicas de análisis de asociación, como el algoritmo Apriori, para descubrir relaciones entre diferentes tiendas que tienden a ser visitadas por los clientes. Esto ayudará a identificar oportunidades de ventas cruzadas y mejorar la oferta de productos en el centro comercial.

- **Evaluar los modelos de machine learning existentes para la segmentación de clientes.**

Para evaluar y seleccionar el modelo de machine learning más adecuado para la segmentación de clientes en Diverplaza, es crucial considerar tanto las características específicas del entorno de aplicación como las ventajas y desventajas de cada método. A continuación, se realiza un análisis de los modelos seleccionados:

### **K-means:**

K-means es uno de los algoritmos de clustering (agrupamiento) más conocidos y utilizados en el campo del machine learning y la minería de datos. Se clasifica como un método de clustering particional que divide un conjunto de datos en K grupos distintos basándose en las características de los datos. Aquí está una descripción más detallada de cómo funciona y sus características clave:

### **Funcionamiento del K-means**

- **Inicialización:** El proceso comienza seleccionando aleatoriamente K puntos del conjunto de datos como los centros iniciales de los clusters (centroides).
- **Asignación:** Cada punto en el conjunto de datos se asigna al centroide más cercano, basándose en la distancia euclidiana. Esto forma K clusters provisionales.
- **Actualización de Centroides:** Una vez que todos los puntos han sido asignados a un cluster, el centroide de cada uno se recalcula. Esto se hace tomando el promedio de todos los puntos que han sido asignados al cluster en el paso anterior.
- **Iteración:** Los pasos 2 y 3 se repiten iterativamente hasta que los centroides no cambian significativamente entre iteraciones consecutivas, lo que indica que el

algoritmo ha convergido, o hasta que se cumpla una condición de terminación predeterminada (como un número máximo de iteraciones).

## Características Clave

- **Simple y Eficiente:** K-means es fácil de entender e implementar. Es especialmente eficiente para grandes conjuntos de datos, aunque su complejidad computacional aumenta con el número de puntos y la dimensionalidad de los datos.
- **Sensibilidad a los Valores Iniciales:** El resultado del algoritmo K-means puede variar significativamente dependiendo de los centroides iniciales seleccionados. Para mitigar este problema, se puede ejecutar el algoritmo varias veces con diferentes centroides iniciales y seleccionar el resultado con la mejor medida de rendimiento (como la suma de errores cuadráticos dentro de los clusters).
- **Sensibilidad a Outliers:** Los valores atípicos pueden tener un gran impacto en la determinación de los centroides, ya que pueden distorsionar el cálculo del promedio.
- **Requiere Especificar K:** Uno de los desafíos clave del K-means es que requiere que el número de clusters K sea especificado de antemano, lo cual puede no ser trivial sin conocimiento previo de los datos. Métodos como el "método del codo" o el "análisis de la silueta" pueden ayudar a estimar el número adecuado de clusters.
- **Clusters de Formas Esféricas:** K-means tiende a funcionar mejor cuando los clusters son compactos y geoméricamente esféricos. No es ideal para identificar clusters de formas irregulares o alargadas.

K-means es ampliamente utilizado en una variedad de aplicaciones, como la segmentación de mercado, la organización de documentos, la compresión de imágenes, y más, debido a su simplicidad y eficacia en muchos escenarios prácticos

## **Ventajas:**

**Simplicidad y Facilidad de Implementación:** K-means es fácil de entender e implementar, lo que lo hace accesible para muchos usuarios y situaciones.

**Eficiencia en Grandes Conjuntos de Datos:** Es relativamente eficiente en términos computacionales, lo que lo hace adecuado para grandes volúmenes de datos.

**Fácil Interpretación de los Clusters:** Los clusters son fáciles de interpretar, y el método proporciona resultados claros y concisos.

## **Desventajas:**

**Necesidad de Especificar el Número de Clusters:** El número de clusters debe determinarse antes de la ejecución, lo cual puede ser un desafío si no se conoce la estructura subyacente de los datos.

**Sensibilidad a los Valores Atípicos:** Los outliers pueden distorsionar significativamente los resultados, llevando a una mala formación de los clusters.

**Limitación a Clusters de Formas Esféricas:** K-means tiende a funcionar mal si los clusters reales tienen formas no esféricas o tamaños muy variables.

## DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es un algoritmo de clustering que se destaca por su capacidad para identificar clusters basados en la densidad de los puntos en un conjunto de datos, así como su habilidad para manejar puntos atípicos o de ruido. Fue introducido en 1996 por Martin Ester, Hans-Peter Kriegel, Jörg Sander y Xiaowei Xu, y se ha convertido en una de las técnicas de clustering más populares y efectivas, especialmente útil en conjuntos de datos donde los clusters pueden tener formas irregulares.

### El algoritmo DBSCAN funciona basándose en dos parámetros principales:

- **Épsilon ( $\epsilon$ ):** Define el radio de la vecindad alrededor de un punto.
- **MinPts (Puntos Mínimos):** El número mínimo de puntos que debe tener una vecindad para que un punto sea considerado como un punto central (o core point).

### Pasos del Algoritmo:

- **Identificación de Puntos Centrales:** Para cada punto en el conjunto de datos, DBSCAN cuenta cuántos puntos se encuentran dentro del radio  $\epsilon$ . Si un punto tiene MinPts o más puntos dentro de su vecindad, se clasifica como un punto central.
- **Formación de Clusters:** Si un punto es un punto central y aún no es miembro de ningún cluster, se inicia un nuevo cluster. Luego, se exploran de manera recursiva todos los puntos dentro del radio  $\epsilon$  del punto central, añadiendo al cluster todos los puntos que son accesibles directa o indirectamente a través de otros puntos centrales.

- **Clasificación de Puntos de Frontera y Ruido:** Durante la exploración, los puntos que no son centrales, pero están dentro del radio  $\epsilon$  de un punto central se consideran puntos de frontera. Los puntos que no son ni centrales ni de frontera se clasifican como ruido.

## Características Clave de DBSCAN

- **No Necesita el Número de Clusters:** A diferencia de K-means, DBSCAN no requiere que se especifique el número de clusters de antemano, lo cual es una ventaja significativa en muchas aplicaciones prácticas.
- **Resistente a Outliers:** DBSCAN trata efectivamente los puntos atípicos como ruido, lo que ayuda a mantener la integridad de los clusters que se forman.
- **Capaz de Detectar Clusters de Formas Arbitrarias:** El algoritmo puede identificar clusters de cualquier forma y tamaño, lo que lo hace adecuado para conjuntos de datos con distribuciones de puntos complejas y variadas.
- **Sensibilidad a los Parámetros:** La elección de  $\epsilon$  y MinPts puede tener un impacto significativo en los resultados del clustering. Elegir estos parámetros puede ser desafiante y generalmente requiere un conocimiento previo del conjunto de datos o experimentación.
- **Escalabilidad:** Aunque DBSCAN es más complejo computacionalmente que K-means, su rendimiento puede ser limitado en conjuntos de datos de muy alta dimensionalidad o extremadamente grandes.

DBSCAN es particularmente útil en áreas como la identificación de conglomerados geográficos, la detección de anomalías, y la segmentación de imágenes, donde la flexibilidad en la forma y el tamaño de los clusters es crucial. Su capacidad para manejar el ruido y su independencia del número de clusters lo convierten en una herramienta robusta para el análisis de datos exploratorio.

### **Ventajas:**

**No Requiere Número Predefinido de Clusters:** DBSCAN no requiere que se especifique el número de clusters de antemano, lo que lo hace útil en escenarios donde la estructura de los datos es desconocida.

**Resistente a Outliers:** Este método es altamente resistente a los valores atípicos, ya que puede identificar y tratar outliers como un ruido separado.

**Capacidad para Identificar Clusters de Formas Arbitrarias:** Puede encontrar clusters de cualquier forma, lo que lo hace versátil para muchos tipos de datos.

### **Desventajas:**

**Sensibilidad a los Parámetros:** Los parámetros  $\epsilon$  y MinPts (que definen la densidad necesaria para formar un cluster) son cruciales y pueden ser difíciles de ajustar.

**No Escala Bien con Dimensionalidad Alta:** Su rendimiento disminuye significativamente con el aumento de la dimensionalidad de los datos.

## Análisis RFM

El análisis RFM (Recencia, Frecuencia, Monto) es una técnica de segmentación de clientes ampliamente utilizada en marketing directo y CRM (Customer Relationship Management). Se basa en tres parámetros clave que representan aspectos importantes del comportamiento de compra de un cliente. Estos son:

- **Recencia (R):** El tiempo desde la última compra del cliente. Un valor menor en "Recencia" indica que la compra más reciente fue hace poco tiempo, lo cual es un signo de mayor lealtad o interés continuo en los productos o servicios de la empresa.
- **Frecuencia (F):** El número total de compras realizadas por el cliente durante un período de tiempo definido. Un número mayor indica mayor compromiso con la marca.
- **Monto (M):** El total de dinero gastado por el cliente durante un período de tiempo. Esto puede indicar el valor del cliente para la empresa.

## Pasos del Análisis:

- **Recopilación de Datos:** Primero, se recopilan datos históricos de ventas que incluyen las fechas de compra, el número de transacciones y los montos gastados por cada cliente.
- **Cálculo de Valores RFM:** Para cada cliente, se calcula la recencia como el tiempo transcurrido desde su última compra, la frecuencia como el número de compras realizadas en un período específico, y el monto como el total gastado en ese período.

- **Segmentación de Clientes:** Los clientes se segmentan en diferentes grupos basados en sus valores RFM. Esto se puede hacer de diversas maneras, como, por ejemplo, asignando una puntuación a cada parámetro y sumando estas puntuaciones, o utilizando técnicas más complejas como clustering k-means sobre los valores RFM.
- **Estrategias de Marketing Personalizadas:** Basándose en la segmentación, se pueden diseñar estrategias de marketing específicas para cada grupo. Por ejemplo, los clientes con una alta frecuencia y monto, pero baja recencia podrían ser re-enganchados con ofertas especiales para incentivar una nueva compra.

## Ventajas:

**Específico para Comportamiento de Compras:** Diseñado específicamente para analizar y segmentar clientes según su comportamiento de compra, haciendo los resultados muy relevantes para aplicaciones de marketing.

**Directamente Vinculado a Estrategias de Marketing:** Facilita la implementación de estrategias de marketing dirigidas, como marketing por correo electrónico y promociones personalizadas.

**Fácil Integración de Datos:** Puede integrarse fácilmente con sistemas de CRM y bases de datos de ventas para obtener insights continuos.

## Desventajas:

**Limitado a Tres Dimensiones:** Se enfoca solo en recencia, frecuencia y monto, lo que puede omitir otras dimensiones importantes del comportamiento del cliente.

**No Captura la Variabilidad Completa:** Puede no capturar la complejidad completa del comportamiento del cliente, especialmente en contextos con múltiples influencias en las decisiones de compra.

## **Elección del Modelo para el Centro Comercial**

Para un centro comercial, donde es fundamental entender el comportamiento de compra y las necesidades de los clientes para personalizar las ofertas y aumentar la lealtad del cliente, el

- **Seleccionar el modelo de segmentación que más se ajuste acuerdo con las necesidades del centro comercial.**

Comparado con K-means y otras técnicas de clustering, RFM tiene algunas ventajas y particularidades que pueden hacerlo más adecuado en ciertos contextos, como en la segmentación de usuarios de DIVERPLAZA.

El modelo RFM es fácil de entender y de comunicar a los interesados. Cada dimensión (Recencia, Frecuencia, Monto) tiene un significado claro y directo relacionado con el comportamiento de compra del cliente, lo que lo hace altamente relevante para las áreas de marketing y la gestión de relaciones con clientes.

Los segmentos creados a partir de RFM son fácilmente interpretables y directamente accionables, como por ejemplo los clientes con alta recencia, alta frecuencia y alto monto pueden ser identificados como clientes VIP o clientes leales, mientras que aquellos con baja recencia, baja frecuencia y bajo monto son clientes perdidos o clientes en riesgo a los cuales se les puede aplicar campañas de reactivación.

## **Desventajas y Consideraciones de K-means y DBSCAN frente a RFM:**

### **Complejidad y Escalabilidad:**

Aunque K-means es un algoritmo de clustering sencillo y escalable, puede ser menos intuitivo en cuanto a la interpretación de los resultados. Los clusters formados por K-means no siempre tienen un significado claro en términos de comportamiento del cliente.

DBSCAN es bueno para detectar outliers y clusters de forma arbitraria, pero la selección de los parámetros eps y min\_samples puede ser complicada, y los resultados pueden ser difíciles de interpretar y de comunicar a un público no técnico.

### **Resultados Menos Directamente Accionables:**

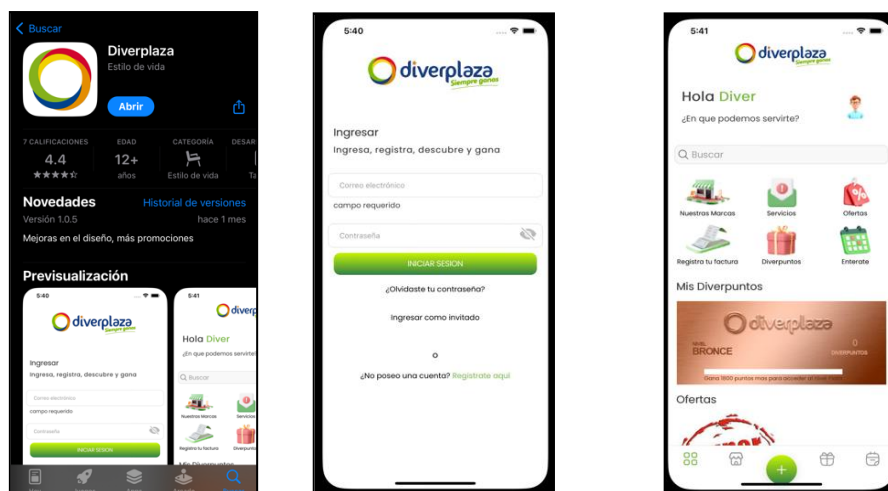
Aunque los clusters obtenidos mediante K-means y DBSCAN pueden ser útiles para ciertos análisis, pueden no ser tan directamente accionables como los segmentos RFM. Los segmentos RFM se pueden mapear directamente a estrategias de marketing, como campañas de retención o reactivación.

RFM es una técnica efectiva para la segmentación de usuarios en el contexto de DIVERPLAZA debido a su simplicidad, enfoque específico en el comportamiento del cliente (Recencia, Frecuencia y Monto), y su capacidad para crear segmentos que son fácilmente interpretables y accionables. Mientras que K-means y DBSCAN pueden ofrecer insights más complejos y detallados en ciertos contextos.

- **Identificar y recopilar los datos relevantes de los clientes, incluyendo características demográficas, comportamientos de compra y preferencias según el modelo seleccionado.**

Para la recopilación de los datos se utilizó la herramienta de fidelización de DIVERPLAZA (APP), la cual permite a los usuarios reclamar diferentes beneficios como boletas de cine, descuentos en diferentes tiendas dependiendo de los puntos acumulados en la APP, es importante resaltar que para hacer uso de la APP hay dos procesos que se deben llevar a cabo, los cuales detallo a continuación:

1. Ingresar a la APP store o Play Store dependiendo del sistema operativo del celular del cliente, aquí deberá buscar la APP llamada **Diverplaza**:



2. Descargarla e instalarla.
3. Una vez instalada, se deberá proceder con el registro en la APP, para lo cual el usuario deberá ingresar la siguiente información:

- a. Correo electrónico
- b. Primer Nombre

- c. Primer Apellido
  - d. Celular
  - e. Contraseña
4. Una vez registrado, para poder acumular puntos el usuario deberá registrar las facturas de las compras realizadas a través de la opción escanear factura, sin embargo, la aplicación solicitará completar datos adicionales:
- a. Dirección
  - b. Tipo de documento
  - c. Número de documento
  - d. Fecha de nacimiento
  - e. Género
5. Una vez completados los datos del paso anterior el usuario podrá carga las facturas de las compras realizadas en DIVERPLAZA.



**Nota:** Por cada \$300 COP registrados en las facturas el usuario acumulará 1 punto, estos puntos serán canjeables por todos los beneficios y premios disponibles en la aplicación.

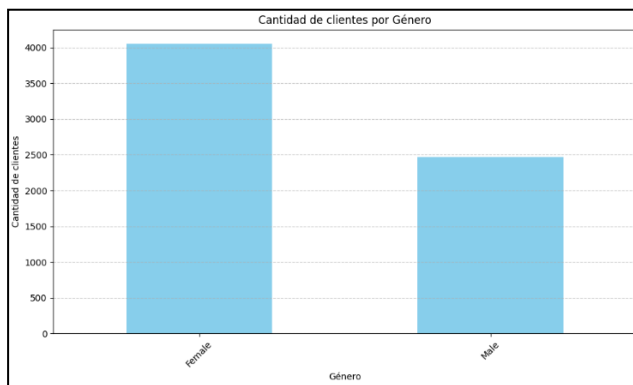
Teniendo en cuenta que la APP ya tiene más de 2 años operando se tienen un total de 11.897 descargas de la APP, sin embargo, del total de usuarios que han descargado la APP

solo 6.036 usuarios han cargado factura, sobre los cuales se llevó a cabo el desarrollo del modelo de segmentación de clientes.

## EDA - (Exploratory Data Analysis)

Previo al desarrollo del modelo se realizó el EDA (Exploratory Data Analysis) de la base de datos:

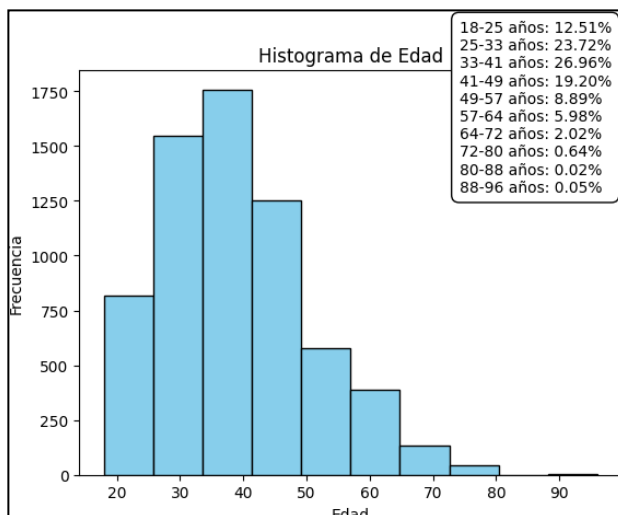
### Análisis variable categórica - Género



Genero	Conteo	Porcentaje (%)
Female	4.049	62%
Male	2.472	38%
<b>Total</b>	<b>6.521</b>	<b>100%</b>

Teniendo en cuenta los resultados anteriores es muy evidente que las mujeres hacen mayor uso de la aplicación

### Análisis variables cuantitativas - EDAD



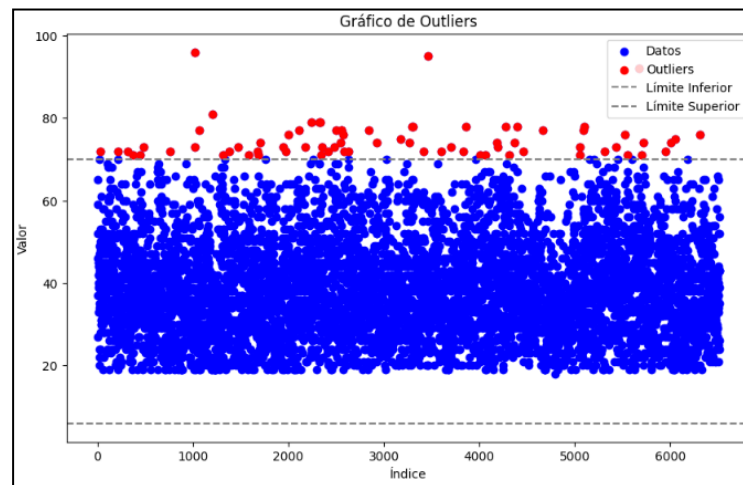
La variable edad tiene una distribución con sesgo positivo, adicionalmente, el mayor número de usuarios se encuentra entre los 25 y 40 años.

## Conclusiones Edad

- **Edad máxima:** La edad más alta es de 96 años.
- **Edad mínima:** La edad más baja es de 18.
- **Edad promedio:** La edad promedio es de 38 años. Esto indica que, en promedio, las personas que asisten a DIVERPLAZA tienen 38 años.
- **Moda:** 34 años. Esto sugiere que el grupo más numeroso dentro de la base de clientes tiene 34 años.
- **Mediana:** 38 años. Esto indica que la mitad de los clientes tienen 38 años o menos, lo cual es consistente con el promedio y refuerza la idea de una base de clientes en su adultez media.
- **Varianza:** La varianza es de aproximadamente 137.97. Esto indica la dispersión de las edades alrededor de su promedio, lo que sugiere una dispersión de los datos no tan alta.
- **Desviación estándar:** La desviación estándar de la edad es de aproximadamente 11 años. Esto indica la dispersión promedio de las edades con respecto a su promedio.
- **Curtosis:** Un valor positivo de curtosis (0.244896 en este caso) sugiere que la distribución de 'Edad' es aproximadamente mesocúrtica, similar a una distribución normal en términos de "picudez".

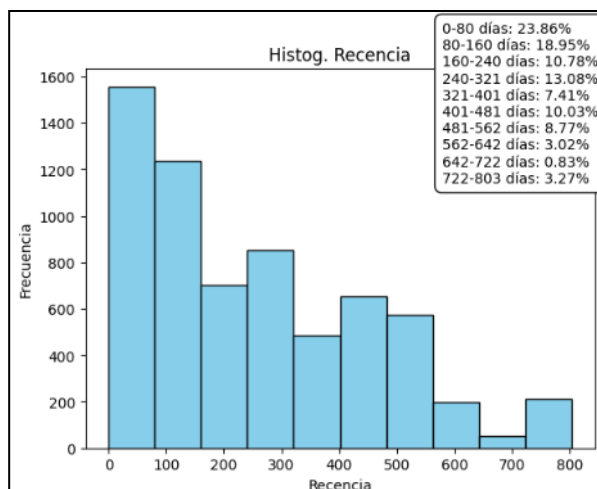
- **Asimetría:** Un valor positivo de asimetría (0.641999 en este caso), lo que significa que hay una cola más larga en el lado derecho de la distribución de edad. Esto puede implicar una proporción algo mayor de clientes más jóvenes que de mayores.

## Identificación de valores atípicos:



Como se evidencia en la gráfica anterior las personas mayores a 70 años, son valores atípicos que pueden afectar el desempeño del modelo, por esta razón son retirados, con el objetivo de optimizar el funcionamiento del modelo.

## Análisis variables cuantitativas - RECENCIA

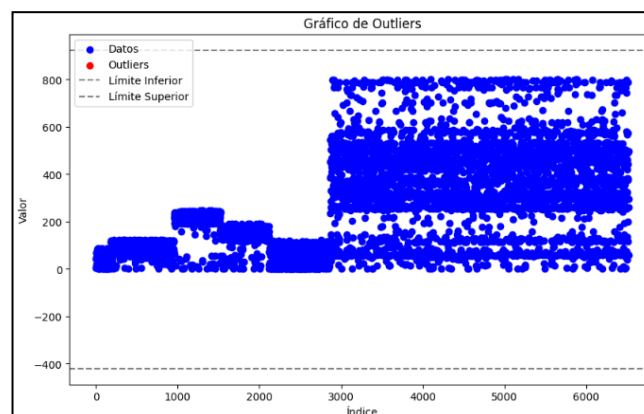


La variable Recencia tiene una distribución con sesgo a la derecha, es decir que tiene, adicionalmente, el mayor número de usuarios realizó su última compra entre 1 y 80 días.

## Conclusiones Recencia

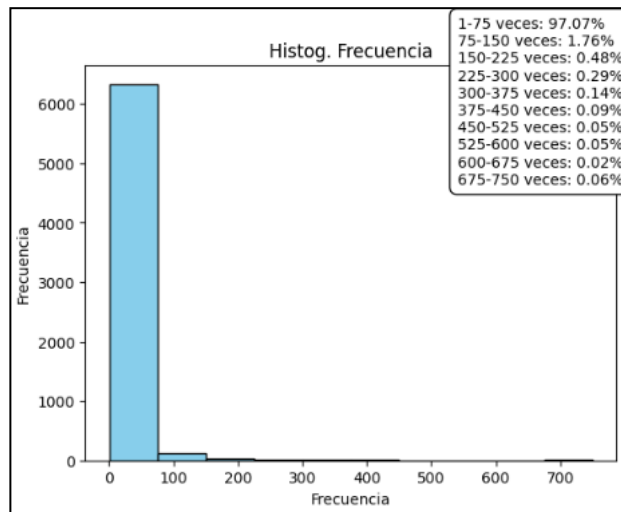
- **Recencia máxima:** La recencia más alta es de 803 días.
- **Recencia mínima:** La recencia más baja es de 0 días.
- **Recencia promedio:** La recencia promedio es de aproximadamente 257.87 días, indicando que, en promedio, los clientes hicieron su última compra hace unos 8 meses.
- **Moda:** 118 días. Esto sugiere que muchos clientes volvieron a comprar en torno a este tiempo después de su compra anterior.
- **Mediana:** 218 días. Esto indica que la mitad de los clientes ha realizado compras más recientemente que el promedio.
- **Varianza:** La varianza es de aproximadamente 39,304.45, indicando una alta variabilidad en los días desde la última compra de los clientes.
- **Desviación estándar:** La desviación estándar es de aproximadamente 198.25 días, mostrando una dispersión significativa en la recencia.
- **Curtosis:** Un valor negativo de curtosis (-0.224481) indica una distribución más aplanada que la normal.
- **Asimetría:** Un valor positivo de asimetría (0.749941), indicando una cola más larga hacia recencias más largas.

## Identificación de valores atípicos



Cómo se evidencia en la gráfica anterior no hay outliers en la variable recencia, razón por la cual no se le debe realizar ningún tipo de tratamiento.

## Análisis variables cuantitativas - FRECUENCIA:



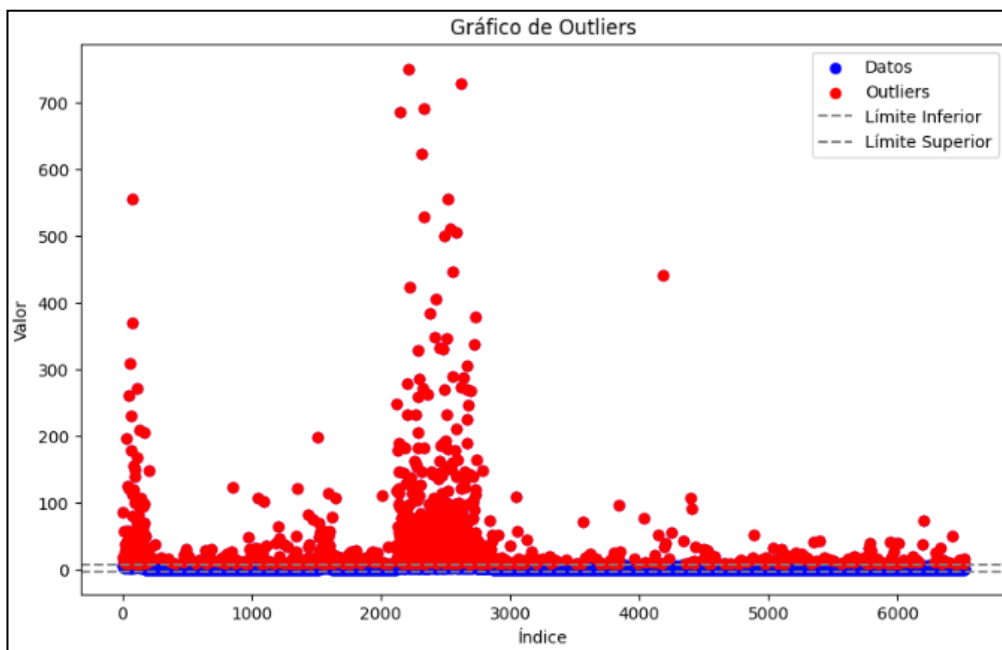
La variable frecuencia tiene una distribución sesgo a la derecha, adicionalmente, el mayor número de usuarios han realizado entre 1 y 75 compras.

## Conclusiones Frecuencia

- **Frecuencia máxima:** La frecuencia más alta es de 750 compras.
- **Frecuencia mínima:** La frecuencia más baja es de 1 compra.
- **Frecuencia promedio:** La frecuencia promedio es de aproximadamente 10.00 compras.
- **Moda:** 1 compra. Esto sugiere que la mayoría de los clientes compra sólo una vez.
- **Mediana:** 1 compra. Esto confirma que más de la mitad de los clientes han realizado sólo una compra.
- **Varianza:** La varianza es de aproximadamente 1,515.73, indicando una alta variabilidad en el número de compras por cliente.
- **Desviación estándar:** La desviación estándar es de aproximadamente 38.93 compras, mostrando una alta dispersión.

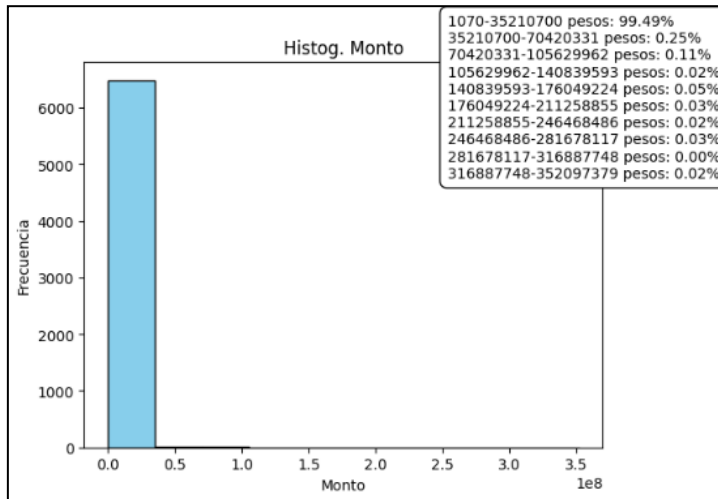
- **Curtosis:** Un valor extremadamente alto de curtosis (127.03937) indica una distribución muy leptocúrtica.
- **Asimetría:** Un valor extremadamente alto de asimetría (9.789148), indicando una cola muy larga hacia un número mayor de compras.

## Identificación de valores atípicos



Como se evidencia en la gráfica anterior hay gran presencia de valores atípicos que pueden sesgar el modelo, por esta razón se realiza el procesamiento de estos para generar una mayor confiabilidad del modelo de segmentación de clientes.

## Análisis variables cuantitativas - MONTO:



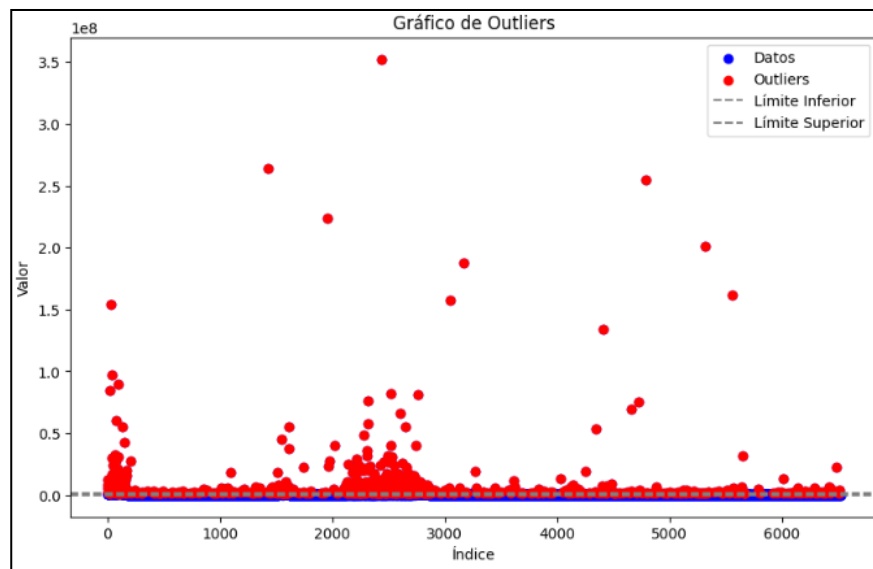
La variable edad tiene una una distribución con sesgo a la derecha, adicionalmente, el mayor número d usuarios se encuentra entre los \$1.070 y \$35.210.700 años.

## Conclusiones Monto

- **Monto máximo:** El monto más alto gastado es de \$352,097,379.
- **Monto mínimo:** El monto más bajo gastado es de \$1,070.
- **Monto promedio:** El monto promedio gastado es de aproximadamente 1,366,487.
- **Moda:** \$69,900. Esto sugiere que el monto más común gastado es relativamente bajo.
- **Mediana:** \$168,612. Esto indica que más de la mitad de los clientes gastan 168,612 o menos.
- **Varianza:** La varianza es extremadamente alta, aproximadamente 92,233,893,821,086.95, reflejando una enorme variabilidad en el monto gastado.
- **Desviación estándar:** La desviación estándar es extremadamente alta, aproximadamente 9,603,847.87.
- **Curtosis:** Un valor muy alto de curtosis (570.981538) indica una distribución muy leptocúrtica.

- **Asimetría:** Un valor extremadamente alto de asimetría (21.379082), indicando una cola muy larga hacia montos más altos gastados.

## Identificación de valores atípicos



Teniendo en cuenta los valores atípicos identificados se procede a realizar el tratamiento de los datos, con el fin de eliminar el sesgo del modelo de segmentación de clientes.

- **Desarrollar el modelo seleccionado para segmentar de manera efectiva la base de clientes en grupos homogéneos según criterios predefinidos.**

Una vez culminado el análisis exploratorio de los datos se procede a aplicar el modelo de clusterización elegido (Análisis RFM – Recencia – Frecuencia - Monto).

Para el desarrollo de este modelo solo se seleccionan las variables Recencia, Frecuencia y Monto. Para cada variable seleccionada se crea con indicador por percentiles de tal manera

que evalué con una calificación de 1 a 5 cada variable, según el percentil en el que se encuentre cada cliente evaluado.

Una vez evaluados todos los clientes por cada variable seleccionada, se procede a segmentar a los clientes según la siguiente codificación:

Cód	Segmento	Cód	Segmento	Cód	Segmento	Cód	Segmento
111	Perdidos	231	En riesgo	535	Prometedores	423	Cientes Potenciales
112	Perdidos	232	En riesgo	311	Prometedores	431	Cientes Potenciales
113	Perdidos	233	En riesgo	313	Prometedores	432	Cientes Potenciales
114	Perdidos	241	En riesgo	314	Prometedores	433	Cientes Potenciales
115	Perdidos	251	En riesgo	315	Prometedores	441	Cientes Potenciales
121	Perdidos	322	En riesgo	411	Prometedores	442	Cientes Potenciales
122	Perdidos	332	En riesgo	412	Prometedores	451	Cientes Potenciales
123	Perdidos	213	En riesgo	413	Prometedores	452	Cientes Potenciales
131	Perdidos	221	En riesgo	414	Prometedores	453	Cientes Potenciales
132	Perdidos	224	En riesgo	415	Prometedores	531	Cientes Potenciales
141	Perdidos	225	En riesgo	421	Prometedores	532	Cientes Potenciales
151	Perdidos	234	En riesgo	422	Prometedores	533	Cientes Potenciales
144	Perdidos	235	En riesgo	424	Prometedores	541	Cientes Potenciales
154	Perdidos	242	En riesgo	425	Prometedores	542	Cientes Potenciales
155	Perdidos	243	No podemos perderlos	511	Prometedores	551	Cientes Potenciales
124	Perdidos	244	No podemos perderlos	512	Prometedores	552	Cientes Potenciales
125	Perdidos	245	No podemos perderlos	513	Prometedores	553	Cientes Potenciales
133	Perdidos	252	No podemos perderlos	514	Prometedores	335	Cientes leales
134	Perdidos	253	No podemos perderlos	515	Prometedores	344	Cientes leales
135	Perdidos	254	No podemos perderlos	521	Prometedores	345	Cientes leales
142	Perdidos	255	No podemos perderlos	522	Prometedores	354	Cientes leales
143	Perdidos	312	Prometedores	523	Prometedores	355	Cientes leales
145	Perdidos	321	Prometedores	524	Prometedores	435	Cientes leales
152	Perdidos	324	Prometedores	525	Prometedores	444	Cientes leales
153	Perdidos	325	Prometedores	323	Prometedores	543	Cientes leales
214	En riesgo	331	Prometedores	333	Cientes Potenciales	445	VIP
215	En riesgo	334	Prometedores	341	Cientes Potenciales	454	VIP
211	En riesgo	343	Prometedores	342	Cientes Potenciales	455	VIP
212	En riesgo	434	Prometedores	351	Cientes Potenciales	544	VIP
222	En riesgo	443	Prometedores	352	Cientes Potenciales	545	VIP
223	En riesgo	534	Prometedores	353	Cientes Potenciales	554	VIP
						555	VIP

Una vez segmentados los clientes según la codificación citada previamente obtenemos los siguientes grupos:

Segmento	Recencia promedio	Frecuencia promedio	Monto promedio
Clientes Potenciales	119	2	\$ 115.167,89
Clientes leales	133	3	\$ 392.900,44
En riesgo	346	1	\$ 139.918,99
No podemos perderlos	368	2	\$ 337.179,33
Perdidos	562	1	\$ 162.366,60
Prometedores	145	1	\$ 153.076,20
VP	59	3	\$ 550.398,56

- **Evaluar la precisión y efectividad de los modelos de segmentación de clientes utilizando métricas de desempeño adecuadas.**

#### Interpretación de los Segmentos

- **Clientes Potenciales**
- **Recencia Promedio:** 119 días
- **Frecuencia Promedio:** 2 compras
- **Monto Promedio:** \$115,167.89
- **Interpretación:** Este grupo ha comprado recientemente y podría convertirse en clientes más leales si se les incentiva adecuadamente.

#### Clientes Leales

- **Recencia Promedio:** 133 días
- **Frecuencia Promedio:** 3 compras
- **Monto Promedio:** \$392,900.44

- **Interpretación:** Compran con frecuencia y gastan significativamente. Son fundamentales para el negocio.

## En Riesgo

- **Recencia Promedio:** 346 días
- **Frecuencia Promedio:** 1 compra
- **Monto Promedio:** \$139,918.99
- **Interpretación:** Hace mucho que no compran y su compromiso es bajo. Es crucial reactivarlos.

## No Podemos Perderlos

- **Recencia Promedio:** 368 días
- **Frecuencia Promedio:** 2 compras
- **Monto Promedio:** \$337,179.33
- **Interpretación:** Aunque compran con buenos montos, hace mucho que no lo hacen. Es vital recuperarlos rápidamente.

## Perdidos

- **Recencia Promedio:** 562 días
- **Frecuencia Promedio:** 1 compra
- **Monto Promedio:** \$162,366.60
- **Interpretación:** Prácticamente inactivos, con baja frecuencia y recencia muy alta.

## Prometedores

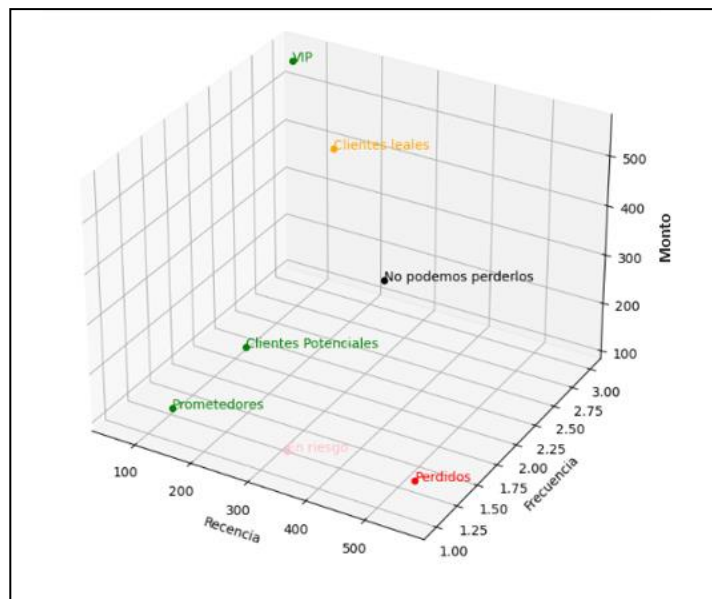
- **Recencia Promedio:** 145 días
- **Frecuencia Promedio:** 1 compra

- **Monto Promedio:** \$153,076.20
- **Interpretación:** Este grupo de clientes tienen compras recientes, pero con baja frecuencia. Tienen el potencial de convertirse en regulares con el estímulo adecuado.

## VIP

- **Recencia Promedio:** 59 días
- **Frecuencia Promedio:** 3 compras
- **Monto Promedio:** \$550,398.56
- **Interpretación:** Los más valiosos, compran frecuentemente, recientemente hicieron su última compra y gastan mucho en las tiendas de DIVEPLAZA.

A continuación, se detalla el comportamiento promedio de cada grupo con base en las variables evaluadas.



## Acciones de Marketing Sugeridas

- **Clientes Potenciales:** Ofrecer promociones o descuentos exclusivos para fomentar una tercera compra.  
  
Enviar comunicaciones personalizadas que destaquen nuevos productos o servicios que podrían interesarles.
- **Clientes Leales:** Programas de fidelización robustos que recompensen la frecuencia y el monto de sus compras.  
  
Ofertas personalizadas basadas en su historial de compra.
- **En Riesgo:** Campañas de correo electrónico para reconectar, ofreciendo incentivos para volver.  
  
Encuestas para entender su baja actividad y ajustar las ofertas según sus respuestas.
- **No Podemos Perderlos:** Contacto directo a través de llamadas o correos personalizados para entender su inactividad y ofrecer soluciones personalizadas.  
  
Invitaciones a eventos exclusivos o acceso anticipado a productos nuevos.
- **Perdidos:** Reintroducción de la marca a través de campañas de remarketing.  
  
Ofertas muy atractivas para intentar una reconexión.
- **Prometedores:** Programas de introducción que incentiven la segunda compra a corto plazo.  
  
Comunicación enfocada en construir la relación y reconocer su valor potencial.

- **VIP:** Servicios exclusivos de atención al cliente y ofertas personalizadas de alto nivel.  
Invitaciones a programas de embajadores de marca o eventos VIP.

Implementando estas acciones específicas, el centro comercial puede mejorar significativamente la eficacia de sus esfuerzos de marketing, aumentando así la retención y el valor de vida de los clientes.

## Referencias bibliográficas

- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Breiman, L. (1984). Breiman et al. En L. Breiman, *Breiman et al.*
- DOĞAN, O., AYÇİN, E., & BULUT, Z. A. (2018). CUSTOMER SEGMENTATION BY USING RFM. *International Journal of Contemporary Economics and*, 19.
- Fader, P. S., Hardie, B. G., & Lee, K. L. (Abril de 2004). *ResearchGate*. Obtenido de ResearchGate:  
[https://www.researchgate.net/publication/227442378\\_Counting\\_Your\\_Customers\\_the\\_Easy\\_Way\\_An\\_Alternative\\_to\\_the\\_ParetoNBD\\_Model/link/0912f510faa5eb675c000000/download?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19](https://www.researchgate.net/publication/227442378_Counting_Your_Customers_the_Easy_Way_An_Alternative_to_the_ParetoNBD_Model/link/0912f510faa5eb675c000000/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19)
- García, V., & Martínez, F. J. (2017). Personalized Product Recommendation System Based on Customer Segmentation and Collaborative Filtering. *ResearchGate*.
- González, V. (14 de Marzo de 2024). <https://americaretail-malls.com>. Obtenido de <https://americaretail-malls.com>: <https://americaretail-malls.com/secciones/estudios/perspectivas-optimistas-para-las-empresas-en-2024/>
- Hair Jr., J. F., Bush, R. P., & Ortinau, D. J. (2010). Investigación de mercados en un ambiente de información digital. En J. F. Hair Jr., R. P. Bush, & D. J. Ortinau, *Hair Jr., Joseph F. ; Bush, Robert P. ; Ortinau, David J. ;* Mc Graw Hill.
- Hosmer Jr., D. W., Lemeshow, S., & Rodney, X. S. (2013). *Applied Logistic Regression*. Wiley Series.
- Hughes, A. M., Hagander, P., & Pal, A. (14 de Diciembre de 2016). *McKinsey&Company*. Obtenido de McKinsey&Company: Unlocking the power of data in sales
- Jain, A. K., & Dubes, R. (1988). Algorithms for clustering data (Vol. 14). En A. K. Jain, & R. Dubes, *Algorithms for clustering data (Vol. 14)* (pág. 320). Prentice Hall.
- Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. En R. A. Johnson, & D. W. Wichern, *Applied Multivariate Statistical Analysis*. Pearson.
- Kotler, P., & Armstrong, G. (2008). *Fundamentos de marketing 8va edición*.
- Kotler, P., & Armstrong, G. (2018). *Principios de Marketing*. Pearson.
- Lee, D., Hosanagar, K., & Harikesh, S. N. (2018). *Advertising Content and Consumer Engagement on Social Media*.
- Retail, A. (29 de Mayo de 2023). [www-emis-com](http://www-emis-com). Obtenido de [www-emis-com](http://www-emis-com).: <https://www-emis-com.bdbiblioteca.universidadean.edu.co/php/search/searchv2>
- Smith, J., & Jones, M. (2019). The Impact of Personalized Marketing Communication on Customer Relationship Outcomes. *Journal of Marketing*.