

Universidad EAN

Facultad de Ingeniería

Especialización en Machine Learning

**SEGMENTACIÓN DE CLIENTES EN
EMPRESA COMERCIALIZADORA DE
EQUIPOS BIOMÉDICOS: ESTRATEGIAS
DE MARKETING BASADAS EN
CLUSTERIZACIÓN**

Autores:

**Diego Ojeda Espitia, Fabian Vargas Alvis, Lady Nuñez Vargas,
Victor Ríos Sánchez**

Directores:

Diego Armando Garcia Garcia

**Bogotá – Colombia
2024**

RESUMEN:

El presente documento aborda la problemática de segmentación de clientes en una empresa de venta de equipos biomédicos en Colombia cuyas estrategias de marketing y perfilamiento de clientes carece de un respaldo analítico robusto que le permita generar planes personalizados para aumentar la venta y la rentabilidad del negocio. Lo anterior, sumado al incremento de la competencia en el sector, dado que para 2022 se presentó un incremento de 8% de las importaciones de este tipo de productos.

Por lo cual, el objetivo será desarrollar un modelo de clusterización que soporte la toma de decisiones del negocio.

Palabras clave: Clustering, Segmentación de clientes, RFM (Recency, Frequency, Monetary), equipos biomédicos, Marketing.

ÍNDICE

CAPÍTULO 1	6
1. PLANTEAMIENTO DEL PROBLEMA.....	6
1.1 Antecedentes del Problema	6
1.2 Descripción del problema.....	7
2. PREGUNTA DE INVESTIGACIÓN	8
3. OBJETIVOS.....	8
Objetivo general	8
Objetivos específicos.....	8
4. CONVENIENCIA DE LA INVESTIGACIÓN	9
5. JUSTIFICACIÓN.....	9
6. MARCO TEORICO	10
Modelos de Clusterización y Segmentación para el Comportamiento de Compra de los Clientes.....	11
6.1. Método K-means.....	11
6.2. Método K-Medoids.....	12
6.3. Análisis ABC	13
6.4. Customer Lifetime Value (CLV).....	14
6.5. Clustering Jerárquico	16
6.6. DBSCAN	17
6.7. Modelo RFM.....	18
CAPÍTULO 2.....	19
DISCUSIÓN CRÍTICA DE LOS MODELOS Y JUSTIFICACIÓN DE LA ELECCIÓN DEL MODELO A IMPLEMENTAR.....	19
7. MARCO INSTITUCIONAL	21
CAPÍTULO 3.....	22
8. METODOLOGÍA	22
Primer nivel	22
Enfoque de la investigación.....	22
Alcance de la investigación	23
Diseño de la investigación	23

Pseudocódigo	35
CAPÍTULO 4.....	37
9. ANÁLISIS Y DISCUSIÓN DE RESULTADOS	37
Desarrollo de los Modelos de Segmentación	37
Filtrado y Análisis Exploratorio de Datos	37
Modelo RFM – Enfoque Cuantiles.....	44
Modelo RFM – Enfoque Clústeres	48
Normalización de datos	49
Número óptimo de clústeres	50
Desarrollo del Algoritmo K-means	51
Análisis de Resultados.....	54
Elección de Enfoque.....	56
Validación y Adaptabilidad del Modelo.....	57
Estrategias de Marketing Recomendadas	58
CAPITULO 5.....	59
10. CONCLUSIONES	59
REFERENCIAS.....	61

Índice de tablas y figuras

Tabla 1 Criterios de Evaluación de Modelos	20
Tabla 2 Presentación general de la Compañía	21
Tabla 3 Rangos de clasificación de las micro, pequeñas, medianas y grandes empresas en Colombia.....	21
Tabla 4 Técnica de Análisis de Datos.....	33
Tabla 5 Comparativo de Productos más rentables VS más vendidos.....	41
Tabla 6 Resultados pruebas de normalidad	50
Tabla 7 Cantidad de observaciones por clúster.....	51
Tabla 8 Descripción de Clústeres k-means.....	54
Tabla 9 Estrategias de Marketing	58
Figuras	
Ilustración 1 Instrumentos para Recolección de Información	30
Ilustración 2 Proceso para Efectuar Análisis Estadístico.....	31

Ilustración 3 Pseudocódigo Modelo RFM (dos enfoques)	36
Ilustración 4 Proceso de Transformación de Datos en Resultados	37
Ilustración 5 Cantidad de clientes por tipo	39
Ilustración 6 Cantidad de clientes Únicos por Zonas vs Top 20 Lugares con más Ventas	39
Ilustración 7 Top 20 productos más vendidos	40
Ilustración 8 Top 20 Productos más rentables económicamente	40
Ilustración 9 Top 20 Clientes más compradores.....	42
Ilustración 10 Top 20 Clientes Que más Ingresos Dejan a la Empresa	42
Ilustración 11 Análisis de tendencias temporales	43
Ilustración 12 Comparación de ventas y montos entre los periodos de tiempo establecidos	44
Ilustración 13 Número de Ventas Mensual 2019 (con Tendencia, Picos y Valles).....	44
Ilustración 14 Número de Ventas Mensual 2023-2024 (con Tendencia, Picos y Valles)	45
Ilustración 15 Distribución de Fechas de Compra	46
Ilustración 16 Distribución de Recencia de Visitas de los Clientes.....	46
Ilustración 17 Distribución Monetaria de los Clientes	47
Ilustración 18 Número de Clientes por Segmento	48
Ilustración 19 Histogramas de variables RFM.....	49
Ilustración 20 Método del codo y Método de Silueta.....	51
Ilustración 21 Observaciones antes y después de clasificación	52
Ilustración 22 Calificación promedio de variables por clúster	52

CAPÍTULO 1

1. PLANTEAMIENTO DEL PROBLEMA

1.1 Antecedentes del Problema

Frente al aumento de importaciones de equipos biomédicos presentado en el país en los últimos 5 años reportado por la ANDI donde expone que para el año 2022 hubo un total de 1.442.538 USD, lo que implica un aumento del más del 8% con respecto al año anterior, (ANDI, 2023) las empresas comercializadoras e importadoras de equipos biomédicos cuentan con una nueva tarea: sobresalir de la competencia, por lo que se hace necesario practicar la identificación de grupos de clientes con necesidades, características o comportamientos en común, permitiendo optimizar la atención y el servicio, buscando generar en los clientes la satisfacción y motivándolos a ser leales a la empresa con el fin de generar una relación de largo plazo con ellos. Para conseguir este propósito, se debe implementar un proceso denominado segmentación de clientes o de mercado, por consiguiente, un segmento de mercado es “un subgrupo de gente o empresas que comparten una o más características que provocan necesidades de productos similares” (Lamb, Mcdaniel, & Hair, 2008, pág. 244).

Para realizar segmentación de clientes o mercado existen diversas técnicas, dentro de las más populares se encuentra el análisis RFM (Recency, Frequency, Monetary), el cual propone que el rendimiento del marketing se puede agrupar en diferentes grupos (Chen, R, 2016, pág. 3).

El análisis RFM se usa por su fácil comprensión y aplicabilidad. Permite conocer el comportamiento de compra de los clientes según unos parámetros básicos, en busca de la identificación de aquellos clientes que aportan mayor valor a la empresa.

Adicionalmente, este análisis se puede complementar con otras técnicas como técnicas de manipulación de datos, métodos gráficos de visualización, métodos de agrupación, junto con la

introducción de nuevas variables, para obtener un análisis adaptado a las necesidades particulares de una empresa o industria ". (López et al., 2017).

En el entorno empresarial actual, contar con información clara, precisa y oportuna es vital para el éxito de las empresas. Poder realizar estudios que permitan identificar grupos de clientes con características de consumo similares es esencial para comprender el valor de cada cliente en la empresa. Este enfoque facilita la creación de campañas de fidelización basadas en las necesidades y diseño de productos enfocados en los segmentos específicos a los que se desea llegar (Kumar, V., & Reinartz, W, 2018).

La empresa comercializadora e importadora de equipos biomédicos en estudio, enfrenta una situación común en el sector donde, al no conocer cómo acercarse bien a su cliente final, es necesario requerir distribuidores o intermediarios reduciendo sus ganancias por el mayor porcentaje de descuento ofrecido a estas empresas. Actualmente la empresa gestiona la información de clientes y ventas utilizando hojas de Excel, aunque estos datos son útiles para analizar el inventario y planificar las compras a proveedores, no se aprovechan para estrategias de marketing lo cual es esencial tal como señala Lisa Arthur: “Los datos generan insights, y esos insights impulsan mejores interacciones, lo que te permite realmente entregar el mensaje correcto a través del canal adecuado en el momento oportuno al cliente indicado. A su vez, estas interacciones mejoradas resultan en un aumento de ingresos y una diferenciación competitiva”. (Arthur, L, 2013, pág. 43).

La empresa ha enfrentado cambios en su imagen y equipo de gestión, lo que ha generado inconsistencia en el mercado y menor fidelización de clientes. Aunque opera como distribuidor multimarca de equipos biomédicos, esta estrategia ha causado problemas de comercialización y pérdidas de clientes y mercadería.

1.2 Descripción del problema.

La empresa enfrenta desafíos críticos que afectan su desempeño y rentabilidad. La falta de una estrategia de marketing adecuada y una gestión deficiente de los datos de clientes han provocado baja retención y una oferta ineficaz de productos. La ausencia de un equipo de

marketing dedicado ha agravado estos problemas, limitando el desarrollo de estrategias para captar y retener clientes.

Además, la empresa carece de un modelo estructurado para analizar y utilizar la información de los clientes, lo que limita su capacidad para ajustar productos y servicios a las necesidades específicas de los clientes basándose en su historial de compras. Esto ha resultado en una disminución de ventas y en la pérdida de confianza y fidelidad de los clientes.

2. PREGUNTA DE INVESTIGACIÓN

¿Cómo la implementación de un modelo de clusterización ayuda a crear segmentos para la categoría de clientes finales en la empresa de equipos biomédicos y de qué manera pueden ser utilizados para diseñar estrategias de marketing que apalanquen el incremento de las ventas?

3. OBJETIVOS

Objetivo general

Desarrollar estrategias de marketing para la categoría de clientes finales, mediante la implementación de un modelo de clusterización que permita obtener segmentos de clientes, basándose en criterios prácticos definidos por el equipo de investigación.

Objetivos específicos

- Evaluar los diferentes modelos de clusterización que se pueden desarrollar y elegir el más adecuado de acuerdo con el sistema de evaluación que defina el equipo de investigación.
- Implementar el modelo de clusterización seleccionado y segmentar a la categoría de clientes finales, basándose en características relevantes del marketing.
- Analizar las características y necesidades de cada uno de los segmentos generados para personalizarlos con estrategias de marketing.

- Generar las estrategias de marketing para cada uno de los segmentos de clientes finales resultantes del modelo.
- Presentar a la empresa un informe con los resultados obtenidos en la segmentación de los clientes de interés, incluyendo una descripción de cada segmento, las diferentes estrategias de marketing propuestas para cada uno y el detalle por cliente del segmento al que pertenece.

4. CONVENIENCIA DE LA INVESTIGACIÓN

Dado que la organización no segmenta a sus clientes y utiliza estrategias generales, la implementación de un modelo de segmentación basado en el análisis de datos permitirá identificar distintos grupos de clientes y entender sus características, facilitando la creación de estrategias de marketing más efectivas.

Los beneficios clave para la empresa incluyen: la generación de segmentos de clientes, la caracterización detallada de cada segmento, y la obtención de insights para desarrollar estrategias de marketing basadas en las características de cada grupo.

Este proyecto es crucial ya que es el primero en detallar, mediante el uso de datos, las características de los clientes para entender mejor sus necesidades y mejorar el rendimiento de ventas.

5. JUSTIFICACIÓN

Teniendo en cuenta el contexto propuesto (un entorno empresarial altamente competitivo), el conocimiento de las necesidades y preferencias de los clientes es determinante para lograr el éxito en la implementación de estrategias de marketing.

La presente investigación busca implementar una estrategia de clusterización para segmentar a los clientes finales, lo cual es fundamental por varias razones: optimización de recursos, mejora de competitividad, aumento de la satisfacción del cliente, disminución de la tasa

de churn (abandono), evidencia de resultados y la definición de las estrategias comerciales de acuerdo con el grupo de clientes.

Respecto a la relevancia social, al personalizar las ofertas de los equipos biomédicos se asegura que cada uno de los grupos de clientes identificados, recibe dispositivos médicos eficientes y de calidad en pro del cuidado de la salud.

Dentro de las implicaciones prácticas se pueden reconocer el ajuste de las campañas de marketing y ventas, de acuerdo con los segmentos generados por el modelo implementado, mejoras en la captura y procesamiento de datos y la relación con otros departamentos: Atención al cliente, logística, soporte y capacitación personalizada.

La utilidad metodológica de la investigación se basará en la aplicación del modelo de segmentación seleccionado para los clientes de la empresa de interés en el sector del comercio al por mayor de venta de equipos biomédicos, utilizando datos reales y así poder generar estrategias específicas para este nicho de mercado.

Campo de investigación: Ciencia, tecnología e innovación. **Grupo de investigación:** Ciencias básicas. **Línea de investigación:** Estadística aplicada y ciencia de datos.

6. MARCO TEORICO

Las reglas del juego están cambiando para la industria de dispositivos médicos. Aunque la fuerte demanda, la demografía y el valor siguen generando interés e inversión, las expectativas cambiantes están imponiendo una carga mayor a las organizaciones (Liu et al., 2015) es decir, es importante abordar todo el motor comercial B2B, incluyendo nuevos enfoques para dirigir a clientes específicos, construir proposiciones de valor de productos y marcas, llegar a los clientes elegidos y maximizar el valor total capturado de cada uno (Liu et al., 2015).

Para llegar a estos es relevante contar con una correcta segmentación, para que las empresas aprovechen su base de datos, dividiéndola en grupos y diferenciándolos por sus características y necesidades propias.

Para lograr un posicionamiento eficiente y efectivo como empresa, se debe trabajar no sobre el perfil amplio de cliente potencial, sino en conjuntos específicos de clientes para poder realmente satisfacer sus expectativas: “La elaboración de perfiles eficaces es una piedra angular de la toma de decisiones estratégicas para las empresas emergentes digitales que buscan un crecimiento sostenible y la satisfacción del cliente” Kasem et al. (2024).

Desde el enfoque de la distribución y comercialización de equipos biomédicos, la segmentación cobra alta importancia por la variedad de clientes como hospitales, clínicas, consultorios médicos y centros de diagnóstico. Cada uno de ellos tienen distintas necesidades, lo que demanda desarrollar estrategias de marketing diferenciadas, como lo resalta un estudio reciente sobre métodos heurísticos de clusterización para la segmentación de clientes: “La segmentación precisa de los clientes es fundamental para que los tomadores de decisiones comprendan completamente los requisitos de los clientes (CR) en el mercado y luego diseñen actividades de mercado para satisfacerlos” Sun et al. (2021).

Modelos de Clusterización y Segmentación para el Comportamiento de Compra de los Clientes.

6.1. Método K-means

Este algoritmo de agrupamiento no supervisado “clasifica puntos dados en k grupos de tal manera que se minimice la distancia entre los miembros del mismo grupo” (Natingga y Toth, 2018).

Ventajas.

- Facilidad para comprender su funcionamiento e implementación.
- Versatilidad de uso con criterios como: distancia, selección de centroides y criterio de terminación en el algoritmo (Rendon y Abundez, 2016).

Caso de Uso.

El siguiente estudio se centra en la segmentación de clientes en la industria minorista, haciendo uso del modelo RFM y el algoritmo K-means, cuyo objetivo es el análisis del comportamiento de compra histórico de los clientes y diseño de campañas promocionales más efectivas.

Para ello, definieron el número de clústeres que representan la cantidad de segmentos para dividir la base de clientes. Luego, utilizaron el algoritmo para agrupar a los clientes bajo sus características RFM y finalmente evaluaron los clústeres para identificar patrones y comportamientos comunes.

Los resultados cuantitativos de la investigación determinaron que haciendo uso de la métrica Silhouette Score que evalúa la calidad de la segmentación, con una separación de 3 clústeres obtuvieron una métrica de 0.362, en comparación de 5 clústeres con un puntaje de

0.349, siendo $K=5$ menos óptimo que $K=3$, ya que el valor de silueta más cercano a +1 representa el óptimo en comparación con otros clústeres.

Luego, realizaron un análisis de ventas que reveló que el clúster 1 tiene la mayor recencia de ventas, pero el clúster 2 tiene la mayor frecuencia de venta y el mayor monto de ventas que otros.

La segmentación con K-means en esta investigación, permitió a las empresas minoristas diseñar campañas de marketing más efectivas y acciones de mejora en la retención de clientes. (Anitha y Patil, 2022).

6.2. Método K-Medoids

Sembiring Berahmana, Mohammed, y Chairuang (2020) discuten que K-medoids es un algoritmo de agrupamiento (clustering) utilizado para dividir un conjunto de datos en clústeres, similar a K-Means. Sin embargo, a diferencia de K-Means, este método selecciona un objeto real del conjunto de datos como el "medoide" o centro del clúster. Esto hace que sea más robusto frente a outliers (valores atípicos) y ruido en los datos, ya que minimiza la suma de las distancias entre los objetos y el medoide dentro de cada clúster.

Caso de Uso.

Según Sembiring Berahmana, Mohammed y Chairuang (2020), el método K-Medoids se utiliza para agrupar datos de clientes en función del modelo RFM, asignando objetos a clústeres mediante la distancia euclidiana y recalculando los medoids hasta que los clústeres se estabilizan, el proceso consiste en lo siguiente:

1. Inicializar el centro del clúster con el número de clústeres (k).
2. Cada dato u objeto se asigna al clúster más cercano utilizando la distancia euclidiana.
3. Seleccionar aleatoriamente objetos en cada clúster como nuevos candidatos a medoid.
4. Calcular la distancia de cada objeto contenido en cada clúster con el nuevo candidato a medoid.
5. Calcular la desviación total (S) calculando el valor total de la nueva distancia menos el valor total de la distancia anterior. Si se obtiene $S < 0$, intercambiar el objeto con el clúster de datos para crear un nuevo conjunto de k objetos como medoid.
6. Repetir los pasos tres al cinco hasta que no haya cambios en el medoid para que se obtengan los clústeres y sus miembros."

Después de aplicar el K-Medoids, los resultados de la segmentación muestran que este método es efectivo para formar clústeres que pueden ser utilizados para identificar diferentes clases de clientes, como Dormant (inactivos) y Golden (valiosos), dependiendo de las características de recencia, frecuencia y valor monetario de las transacciones de los clientes.

6.3. Análisis ABC

El análisis ABC está basado en el principio de Pareto donde el 20% del esfuerzo genera el 80% del beneficio. La esencia del análisis ABC es dividir la base de clientes en categorías según el porcentaje de las compras del cliente en relación con las ventas totales:

- Categoría A – 80%;
- Categoría B – 15%;
- Categoría C – 5%.

Trabajar con los clientes de la categoría A es una prioridad para la empresa, los clientes de la categoría B están en un nivel intermedio, y no se debe perder tiempo trabajando con los clientes del grupo C."(Evdokimova, 2021).

Ventajas

- **Segmentación Efectiva del Inventario:** El análisis ABC permite a las empresas clasificar los productos en diferentes categorías según su relevancia, lo que facilita una mejor priorización y manejo del inventario.
- **Optimización de la Gestión de Inventarios:** Al centrarse en los productos más importantes, las empresas pueden mejorar la rotación de inventarios, lo que contribuye a una gestión más eficiente y a la reducción de costos asociados.
- **Enfoque Estratégico en la Promoción:** Este análisis ayuda a las empresas a dirigir sus esfuerzos de marketing y contenido hacia los productos más significativos, lo que aumenta la efectividad de las campañas de ventas.
- **Mejora en la Relación con el Cliente:** Al integrar el análisis ABC con actividades de seguimiento y servicio al cliente, las empresas pueden aumentar la satisfacción del cliente y mejorar la fidelización, lo que a su vez potencia las ventas y la lealtad del cliente.

Estas ventajas hacen que el análisis ABC sea una herramienta valiosa para optimizar tanto las operaciones internas como la relación con los clientes en las pequeñas y medianas empresas. (Galvez-Torres et al., 2020)

6.4. Customer Lifetime Value (CLV)

Se trata de un método (también considerada métrica) para determinar la potencial ganancia que representa un cliente a lo largo de todo su ciclo de vida y permite identificar segmentos de clientes, específicamente aquellos de alto valor que tienen una importancia mayor para la estrategia del negocio.

En una definición más técnica, se puede interpretar como el valor presente de los beneficios futuros obtenidos de un cliente durante su vida en relación con la empresa (Montero et al., 2020).

Las estimaciones realizadas por medio del CLV se basan en la frecuencia, recencia y valor promedio de cada compra, es decir, en otro modelo: el RFM, con el fin de tratar de predecir el comportamiento de los clientes en el futuro y con base en esa estimación, crear grupos de clientes que permitan generar estrategias de marketing más rentables (Montero et al., 2020).

Algunos puntos de vista señalan que el enfoque del CLV suele subestimarse por no ser una técnica demasiado compleja, sin embargo, algunos autores han combinado el concepto detrás de esta métrica junto a modelos predictivos de regresión, uno de ellos, el Robust M-estimation usado para optimizar el rendimiento de la estimación CLV (Elveny et al., 2024)

Ventajas.

La aplicación de modelos que determinen el valor de los clientes, contribuyen a tomar decisiones que optimicen el gasto en campañas publicitarias, enfocando mejor los recursos y validando la rentabilidad entre diferentes segmentos de mercado (Berger & Nasr, 1998).

La rápida proliferación del uso de tecnologías Big Data ha contribuido a la disponibilidad de información en masa, por lo que la aplicación de estos modelos ha sido viable en entornos donde antes era muy complejo manejarlos. Tal es el caso de los eCommerce, en donde la aplicación del CLV permite tomar decisiones en periodos cortos de tiempo y ajustar sus estrategias de marketing.

Se pueden inferir segmentos de ciclo de vida del cliente con base en patrones específicos para ciertos grupos de personas, esto ha sido probado extensivamente en la industria retail (Nie et al., 2024).

Caso de Uso.

El siguiente estudio tomó datos de churn de clientes de uno de los mayores proveedores de servicios de viajes del mundo (www.ctrip.com), en donde se contó con un set de datos etiquetado para saber si el cliente había abandonado o no.

A través de un exhaustivo proceso de preprocesamiento de datos (valores atípicos, faltantes, negativos, creación de nuevas variables, etc.) y posterior manejo para dividir la información y set de prueba y entrenamiento, se realizó la evaluación de performance y de precisión del modelo para saber si se había hecho una adecuada clasificación de las etiquetas churn/No churn de los diferentes registros.

Tal vez lo más notable de este estudio es que pasó de la aplicación de analítica descriptiva y predictiva para centrarse en el ámbito prescriptivo, centrándose en la comprensión de los hechos, en el soporte a la toma de decisiones y en mejorar su aplicabilidad. La conclusión del mismo, es que el uso de PAM (prescriptive analytics method) permite una rápida adaptación por parte de las organizaciones a los constantes cambios tanto del macroentorno como del entorno próximo de las organizaciones, específicamente en un ambiente dinámico como es el eCommerce en la industria de viajes y turismo (Feng et al., 2024).

6.5. Clustering Jerárquico

Se trata de una técnica de agrupamiento que organiza los datos en un árbol jerárquico para poder construir la estructura que permita definir la cantidad de cluster óptimo. A diferencia de algoritmos como el k-means, el clustering jerárquico no necesita definir el número de clústeres desde el inicio.

Se identifican dos enfoques en esta técnica: 1) Aglomerativo: Que toma cada uno de los puntos de datos como cluster único y progresivamente los va uniando con otros puntos de datos similares hasta que quede un único cluster global y 2) Divisivo: Que es el enfoque contrario, partiendo de un clúster único y dividiendo en cada etapa en 2 subgrupos hasta llegar a puntos individuales de datos (Correa et al., 2024)

El proceso de agrupamiento de puntos de datos en el enfoque aglomerativo es realmente muy intuitivo: Si hay N puntos de datos significa que se tienen N grupos (clústeres). Luego se identifican los dos puntos más cercanos y se combinan en un solo grupo, es decir que luego de la primera iteración, se tendrán N-1 grupos. Se actualizan las distancias con base en este nuevo punto de datos creado a partir del agrupamiento y se continúa el proceso de una manera iterativa

reduciendo en cada paso la cantidad de puntos de datos en 1 hasta quedar en el único grupo resultante (Yang & Trewn, 2004).

Ventajas.

Tal como se indicó antes, una de las principales ventajas de aplicar el método de clustering jerárquico es que no se necesita conocer de antemano el número de clústeres, lo que a menudo demanda procesos complejos de cálculo. A través del dendograma resultante, se ayuda al investigador a determinar el mejor número de agrupamientos o clústeres que se deben usar para posteriormente complementarlos con otro tipo de técnicas o modelos (Kopřivová & Matušínková, 2023)

En los conjuntos de datos donde existen (sin verse a simple vista) relaciones de clústeres y subclusters, el mejor escenario es contar con métodos de agrupamiento jerárquico para explicar mejor las relaciones y lograr una mayor eficacia al momento de procesar los datos y generar el resultado (Gopal, 2019).

Caso de Uso.

Desarrollo de un caso de asignación de créditos de consumo a agencias y clientes por parte de una cooperativa financiera en Ecuador: El caso tomó en cuenta varias variables financieras de los clientes (capacidad de pago, nivel de riesgo, características demográficas del cliente, etc.).

Una vez realizada la etapa de preprocesamiento y limpieza de datos, se ejecutó el modelo con el cual se estructuró el dendograma (diagrama de árbol) que permitió a su vez identificar los diferentes clústeres de clientes.

La aplicación de este modelo, dio como resultado la creación de 3 clústeres dependiendo principalmente del monto de ingresos de los clientes y del nivel de estabilidad laboral. A partir de allí, se propusieron diferentes estrategias comerciales y de marketing para cada grupo de clientes, entendiendo sus características (Correa et al., 2024).

6.6. DBSCAN

Es un algoritmo de agrupamiento que permite identificar clústeres de alta densidad en un grupo de datos, donde maneja el ruido es capaz al detectar clústeres de formas arbitrarias y tamaños variables, los parámetros de DBSCAN utiliza dos el radio de búsqueda (ϵ) y el número mínimo de puntos (MinPts) necesarios para considerar una región como un clúster denso. Estos parámetros permiten al algoritmo adaptarse a diferentes distribuciones de datos.

Manejo de Ruido Una de las ventajas significativas de DBSCAN es su capacidad para identificar y clasificar puntos de datos que no pertenecen a ningún clúster, denominados "ruido". Esto es útil en aplicaciones donde los datos pueden contener errores o anomalías.

Aplicaciones: DBSCAN se utiliza en diversas aplicaciones, desde la detección de patrones en datos geoespaciales hasta la segmentación de clientes en marketing, donde la identificación de grupos de comportamiento similar es crucial. (Discovery and Data Mining ,1996).

6.7. Modelo RFM

La segmentación de Clientes mediante el Modelo RFM es una técnica esencial en marketing que permite a las empresas clasificar a sus clientes en diferentes grupos para personalizar estrategias y mejorar la relación con ellos sus características principales:

Recencia (R): Mide el tiempo transcurrido desde la última compra del cliente, sin embargo, los clientes que han comprado recientemente son más propensos a responder a futuras campañas de marketing.

Frecuencia (F): Evalúa la cantidad de compras realizadas por el cliente en un período determinado. Los clientes que compran con frecuencia suelen ser más leales y valiosos para la empresa.

Valor Monetario (M): Calcula el gasto total de un cliente durante un período específico. Los clientes que generan mayores ingresos son prioritarios para las estrategias de marketing. (Discovery and Data Mining, 96.)

CAPÍTULO 2

DISCUSIÓN CRÍTICA DE LOS MODELOS Y JUSTIFICACIÓN DE LA ELECCIÓN DEL MODELO A IMPLEMENTAR.

Para dar inicio a la elección del modelo más acuerdo se tendrá en cuenta las ventajas y casos de uso de cada uno de los modelos anteriormente investigados, el equipo de investigación evaluó los diferentes modelos a través de la definición del sistema de calificación teniendo en cuenta las características específicas del mercado de equipos biomédicos en Colombia y las necesidades de segmentación de la empresa con ayuda de los siguientes criterios prácticos que se van a tener en cuenta:

1. Modelo que mejor resultados ha generado en mercadeo.
2. Facilidad de implementación, ajuste e interpretación de los resultados.
3. Experiencia de construcción por el equipo de investigación.
4. Capacidad del modelo para trabajar con el volumen de datos dado por la empresa.

Con las variables anteriormente definidas, se construye la Tabla 1 de ponderación por criterio frente a cada modelo investigado, estableciendo un sistema de puntuación en una escala del 1 al 5, donde:

- 1 = Muy deficiente
- 2 = Deficiente
- 3 = Aceptable
- 4 = Bueno
- 5 = Excelente

Tabla 1 Criterios de Evaluación de Modelos

Modelo / Criterio	Resultados (35%)	Implementación e Interpretación (10%)	Experiencia de Construcción (25%)	Capacidad (30%)	Puntuación Total
K-means	4	5	4	4	4,1
K-medoids	4	4	3	4	3,7
Análisis ABC	2	4	4	3	3,0
Customer Lifetime Value	5	5	4	4	4,4
Clustering Jerárquico	4	3	2	4	3,4
DBSCAN	3	3	2	4	3,0
Segmentación de Clientes RFM	5	5	5	4	4,7

Fuente: Elaboración propia.

Desde el inicio de la investigación y junto con la definición de los objetivos, se establecieron criterios claros para seleccionar el modelo de segmentación que se implementará con la información disponible. Este enfoque fue adoptado por los investigadores debido a la falta de información específica sobre modelos aplicados en la industria de equipos biomédicos. Además, diseñar y ejecutar de manera rigurosa cada uno de los modelos investigados requeriría más tiempo del que se tiene asignado para el desarrollo de la investigación.

En este contexto y avanzando de lo general a lo específico, se determina que el modelo RFM (Recencia, Frecuencia y Valor Monetario) es el más adecuado, ya que cumple con todas las condiciones necesarias para ser incluido en este enfoque de investigación. Además, se incorporarán métodos de clusterización y análisis de cuantiles en el marco del RFM, dado que estos enfoques permiten segmentar eficazmente a los clientes, proporcionando una comprensión más profunda de su comportamiento de compra y facilitando la identificación de patrones relevantes en el mercado.

7. MARCO INSTITUCIONAL

Tabla 2 Presentación general de la Compañía

Objeto social:	<p>Actividad principal - CIIU 4659 - Comercio al por mayor de otros tipos de maquinaria y equipo.</p> <p>Actividad secundaria:</p> <ol style="list-style-type: none"> 1. CIIU 3313 - Mantenimiento y reparación especializado de maquinaria y equipo. 2. CIIU 4646 - Comercio al por mayor de otros artículos de uso doméstico 															
Número de colaboradores	40 colaboradores en Colombia.															
Tamaño de la empresa	<p>El decreto 957 del 5 de junio de 2019 en el capítulo 13, artículo 2.2.1 establece los rangos de clasificación de las micro, pequeñas, medianas y grandes empresas en Colombia, de acuerdo con esto y considerando que la actividad principal de la compañía es la importación y venta de equipos biomédicos, entre otros.</p> <p>Tabla 3 Rangos de clasificación de las micro, pequeñas, medianas y grandes empresas en Colombia</p> <table border="1"> <thead> <tr> <th><i>Tipo de Empresa</i></th> <th><i>Número de Empleados</i></th> <th><i>Activos Totales (en SMMLV)</i></th> </tr> </thead> <tbody> <tr> <td><i>Microempresa</i></td> <td><i>Hasta 10</i></td> <td><i>Hasta 500</i></td> </tr> <tr> <td><i>Pequeña Empresa</i></td> <td><i>11 a 50</i></td> <td><i>501 a 5.000</i></td> </tr> <tr> <td><i>Mediana Empresa</i></td> <td><i>51 a 200</i></td> <td><i>5.001 a 30.000</i></td> </tr> <tr> <td><i>Gran Empresa</i></td> <td><i>Más de 200</i></td> <td><i>Más de 30.000</i></td> </tr> </tbody> </table> <p>Fuente: Elaboración propia con información tomada de mipymes.gov.co</p> <p>Considerando los ingresos de actividades ordinarias, que equivalen a 1.100.000 COP millones, de acuerdo con lo anterior, esta empresa es considerada como una compañía pequeña al tener en cuenta el parágrafo 1 de dicho decreto que establece “Se considera pequeña empresa.</p>	<i>Tipo de Empresa</i>	<i>Número de Empleados</i>	<i>Activos Totales (en SMMLV)</i>	<i>Microempresa</i>	<i>Hasta 10</i>	<i>Hasta 500</i>	<i>Pequeña Empresa</i>	<i>11 a 50</i>	<i>501 a 5.000</i>	<i>Mediana Empresa</i>	<i>51 a 200</i>	<i>5.001 a 30.000</i>	<i>Gran Empresa</i>	<i>Más de 200</i>	<i>Más de 30.000</i>
<i>Tipo de Empresa</i>	<i>Número de Empleados</i>	<i>Activos Totales (en SMMLV)</i>														
<i>Microempresa</i>	<i>Hasta 10</i>	<i>Hasta 500</i>														
<i>Pequeña Empresa</i>	<i>11 a 50</i>	<i>501 a 5.000</i>														
<i>Mediana Empresa</i>	<i>51 a 200</i>	<i>5.001 a 30.000</i>														
<i>Gran Empresa</i>	<i>Más de 200</i>	<i>Más de 30.000</i>														
Sector al que pertenece	Comercio - Salud															
Principales competidores	Incolmedica, Tecnicaelectromedica, Las Electromedicina															

Fuente: Elaboración propia con diferentes consultas realizadas (indagaciones, mipymes.gov.co)

CAPÍTULO 3

8. METODOLOGÍA

A continuación, se detalla la metodología utilizada para la implementación del modelo de clusterización en la empresa comercializadora de equipos biomédicos y la descripción del nivel de investigación, metodología desarrollada, definición de variables, los métodos de recolección de información utilizados y técnicas de análisis de los datos.

Primer nivel

Enfoque de la investigación

El enfoque de la investigación es cuantitativo, teniendo en cuenta que “de las preguntas se establecen hipótesis y determinan variables; se traza un plan para probarlas (diseño); se miden las variables en un determinado contexto; se analizan las mediciones obtenidas utilizando métodos estadísticos, y se extrae una serie de conclusiones respecto de la o las hipótesis” (Hernández et al., 2010).

Con base en lo anterior y, teniendo en cuenta que el trabajo propuesto a desarrollar se enmarca en el desarrollo de un modelo de segmentación a partir de una técnica específica que requiere la recolección de mediciones históricas y su posterior tratamiento mediante técnica estadísticas y de análisis de datos para la verificación de la hipótesis inicial: análisis descriptivo de variables, limpieza y depuración de datos, modelo RFM (Recency, Frequency, Monetary Value). Por lo anterior, cumple con las condiciones para ser catalogado dentro de este enfoque de investigación.

Además, que la propuesta cumple con varias características que permiten clasificarlo en este enfoque según proponen Hernández et al., (2010):

- Los datos son productos de mediciones y se analizan con métodos estadísticos.
- Los objetos medidos no son afectados por los investigadores, promoviendo la objetividad en la investigación.
- El proceso sigue un patrón lógico, predecible y estructurado para llegar a cabo la investigación.

Alcance de la investigación

De acuerdo con la clasificación dada por Hernández et al., (2010) en la cual se pueden identificar cuatro tipos de alcance: Exploratorios, descriptivos, correlacionales y explicativos. Se determina que para este proyecto se aplica un alcance correlacional, argumentado por lo siguiente:

Se va a trabajar con variables históricas para utilizar una técnica estadística que permita crear segmentos dentro de la población a través del comportamiento de los valores que toman las variables analizadas. Esto supone una relación entre estas variables que permitan crear dichos grupos.

A lo largo de la investigación, se tenderá a encontrar relaciones entre las variables de la información histórica de los individuos y analizar cómo su comportamiento hace que sean asignados a un grupo u otro.

Diseño de la investigación

El diseño de la investigación se enmarca en una metodología de “construcción de modelos - estudio aplicado” dado que se va a desarrollar un modelo analítico por medio de la aplicación de técnicas estadísticas dentro del contexto de una organización con una problemática específica que se va a tratar.

Como resultado del trabajo, se obtendrá un modelo RFM (Recencia, Frecuencia y Valor Monetario) personalizado para la empresa de interés, desde la creación de dos métodos Cuantiles y Clusterización y posterior selección del modelo que mejor se adapte a las necesidades de la organización, con el objetivo de mejorar la toma de decisiones tanto comerciales como de marketing con una base analítica construida desde el análisis de información histórica. En caso de que quiera replicarse en otras industrias o empresas se deberá hacer un análisis exploratorio de la información que cambie el enfoque del desarrollo del modelo.

Población y Muestra

Descripción de la Población

La población objeto de estudio está compuesta por los clientes particulares de una empresa comercializadora de equipos biomédicos. Las características de la población son las siguientes:

Tipo de Clientes: Instituciones de salud (hospitales, clínicas) y profesionales independientes (médicos, terapeutas).

Ubicación Geográfica: Clientes ubicados en todas las zonas del país donde opera la empresa, abarcando tanto áreas urbanas como rurales.

Aval de Venta por Contabilidad: Clientes debe contar con Aval “Generado de Guia” lo que indica que el área de facturación generó visto bueno para realizar la venta.

El número total de individuos pertenecientes a la población es de 781 clientes en el rango de fechas entre el 2016 y 2024.

Tipo de Muestreo Elegido

Para este estudio, se optará por un muestreo no probabilístico debido a que la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación o los propósitos del investigador (Johnson, 2014, Hernández-Sampieri et al., 2013 y Battaglia, 2008b) en este caso las causas van ligadas al modelo de clusterización elegido para segmentar (RFM), a los ciclos de compra de la industria biomédica y al comportamiento global de la ventas de equipos biomédicos.

En Colombia, la normativa actual no exige la renovación obligatoria de tecnología al cumplirse el tiempo de vida útil establecido por los fabricantes, que varía entre 1 y 10 años para equipos de riesgo bajo. Para establecer un límite en la muestra y garantizar que el modelo RFM no se vea afectado por compras antiguas, se calculará un promedio entre estos extremos.

Adicionalmente, y a consideración del gerente comercial y general de la empresa analizada se excluye el período de tiempo transcurrido entre el 06/03/2020 y el 30/06/2022 caracterizado por un aumento relevante en las compras de la industria biomédica, debido a la alta demanda de equipos durante la pandemia mundial. Por lo tanto, las adquisiciones realizadas durante este período podrían afectar la muestra final, y necesitamos asegurar la validez y relevancia de los datos analizados.

Por ende, la muestra quedaría con la siguiente característica:

Clientes que realizaran compras entre los periodos del 01/01/2019 hasta el 31/08/2024 excluyendo las compras transcurridas en el periodo de pandemia de la población descrita en el ítem anterior.

Segundo nivel

Selección de métodos o instrumentos para recolección de información

Como el objetivo de esta investigación empieza por extraer y filtrar información directamente de la base de datos ofrecida por la empresa sin necesidad de realizar preguntas a terceros, realizar formularios o buscar otro tipo de dato, el enfoque para diseñar nuestro instrumento de medición estaría diseñado basado en los siguientes 6 aspectos:

1. Definición de Objetivos:

Según Aravind Hebbali (2023), para calcular la puntuación RFM es necesario contar con un ID de cliente único, el número de transacciones u órdenes, las compras totales realizadas por el cliente y la cantidad de días desde la última visita, basado en el período que se desea calcular el RFM. Por ende, es necesario que la muestra cumpla con estos criterios.

2. Diseño del Instrumento y Revisión de la Base de Datos:

Herramientas de Extracción: Se usará la herramienta de Excel para la primera exploración de la información y el lenguaje python para la extracción y el procesamiento de esta misma.

Especificación de Consultas: Estas consultas deben incluir los siguientes campos que ayudaran en esta investigación son:

CIA = día del mes de la venta

ANIO = Año que se realizó la venta

MES = Mes en el que se realizó la venta

FEC_EMI = fecha de emisión de la factura

NOM_CLIENTE = nombre del cliente

CODPROD = Código del producto.

PRODUCTO = Nombre del Producto

MODELO = Modelo del producto

MARCA = marca del producto

PROVEEDOR = Proveedor del producto

UMED = unidad de medida

CANTIDAD = cantidad de productos vendidos

PREC_ACTUAL = valor del producto actual

VVTA = valor de la venta

%DESCUENTO = porcentaje de descuento ofrecido

NOM_ZVTA = nombre de la zona de venta

COD_TIPCLIENTE = código asociado al tipo del cliente

NOM_TIPCLIENTE= nombre asociado al tipo de cliente

GLOSA: Estado de la venta.

UBICACIÓN=ubicación de la venta.

NOM_TRANSACCION= nombre del tipo de transacción

Con esta información logramos contar con cada uno de los criterios establecidos por el autor Aravind Hebbali (2023), y adicionalmente se cuenta con otros datos que pueden ser relevantes al momento de realizar recomendaciones al respecto de que productos ofrecerle a los clientes que se validen en el modelo.

Criterios de Filtrado:

Criterio 1: Fecha de venta entre: Clientes que realizaran compras entre los periodos del 01/01/2019 hasta el 31/08/2024. Excluyendo ventas realizadas entre 06/03/2020 y el 30/06/2022 (periodo correspondiente a la pandemia).

Criterio 2: Tipo de Cliente: Clientes Clasificados en la columna NOM_TIPCLIENTE como "Cliente Particular" y "Distribuidor".

Criterio 3: Clientes que cuenten en la columna de glosa con: Aval de Venta por Contabilidad: "Generado de Guia".

Criterio 4: Clientes que adquirieron algún producto o equipos y esta ha sido registrado en la columna NOM_TRANSACCION como "Venta de mercadería a terceros".

Agrupamiento y Ordenación: Si es necesario, para presentación de la información y para manejo de los clientes se agruparán por zonas.

3. Validación de Datos

Verificación de Resultados: Tras la extracción, se verifica que los datos sean correctos y cumplan con los criterios establecidos, mediante validación cruzada y se verificará que estos tengan coherencia, y mantengan los formatos requeridos.

Control de Calidad: se realizará análisis de calidad para identificar posibles inconsistencias o errores en los datos usando herramientas en Python.

4. Análisis de Datos

Métodos de Análisis: se definirá cómo se analizarán los datos extraídos, usando la respectiva tabla de análisis de modelos presentada anteriormente, adicional a esto se realizarán los respectivos análisis estadísticos, visualizaciones y técnicas de modelado.

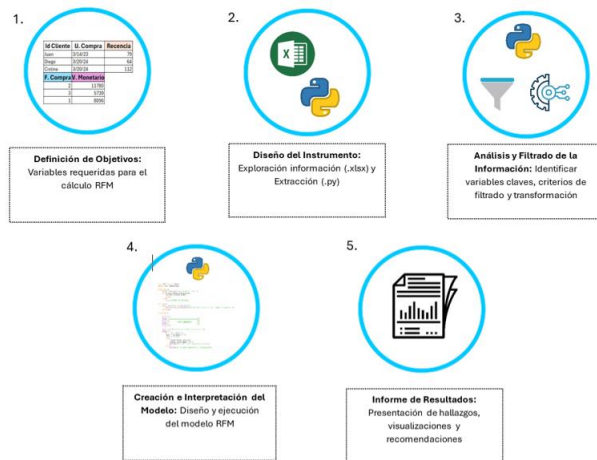
Interpretación de Resultados: se establecerá un plan para interpretar los resultados en función de los objetivos propuestos en la investigación.

5. Informe de Resultados y Documentación

Estructura del Informe: se diseñará un informe que presente claramente los hallazgos y las visualizaciones de datos usando el lenguaje Python en Google Colab.

Recomendaciones: Ofreceremos recomendaciones a la empresa basadas en los análisis realizados.

Ilustración 1 Instrumentos para Recolección de Información



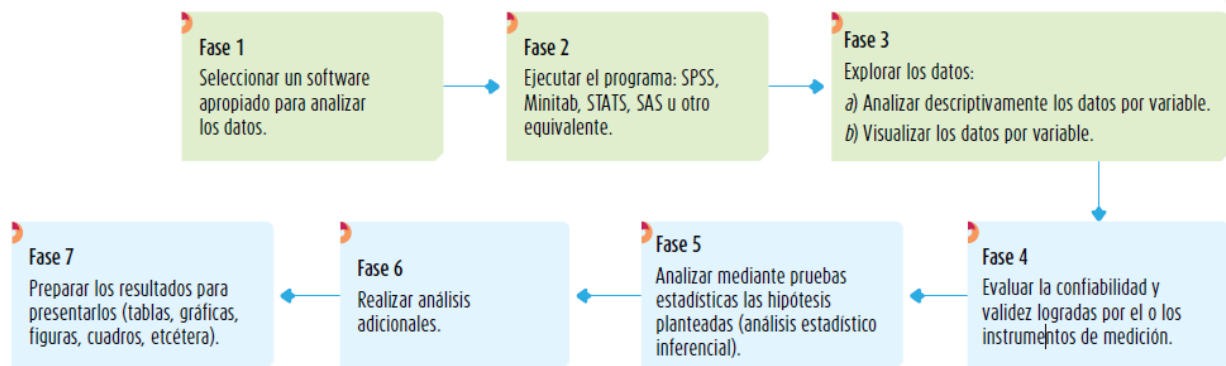
Fuente: Elaboración Propia.

Técnicas de análisis de datos

Una vez definido el instrumento para la recolección de información, el paso a seguir es realizar el análisis de datos de la información disponible.

Para esto, se seguirá el proceso propuesto por Hernández et al. (2010, p. 272) el cual consta de 7 fases las cuales dan una guía general acerca del proceso de análisis estadístico, tal como se observa en la ilustración 2.

Ilustración 2 Proceso para Efectuar Análisis Estadístico



Fuente: Hernández et al. (2010, p. 272)

A continuación, se describirán las primeras 3 Fases del proceso relacionado anteriormente, en dónde se encuentra el punto de partida para lograr el objetivo de análisis.

Fase 1

Inicialmente, se definen los softwares que se van a utilizar durante la etapa de análisis de datos:

Lenguaje Python: Python es un lenguaje de programación y análisis de datos muy utilizado para manipular, transformar y generar análisis en los datos. Este lenguaje cuenta con bibliotecas especializadas para diferentes tareas en el manejo de los datos. Una definición más formal de este lenguaje es “Python es un lenguaje de programación orientado a objetos claro y potente, comparable a Perl, Ruby, Scheme o Java.” (Python Software Foundation, 2022)

Google Collab: Tal como lo menciona Google en su apartado de preguntas frecuentes “Colab es un servicio alojado de Jupyter Notebook que no requiere configuración para su uso y brinda acceso gratuito a recursos informáticos, incluidas GPU y TPU. Colab es especialmente

adecuado para el aprendizaje automático, la ciencia de datos y la educación.” (Google Research, n.d.)

Microsoft Excel: Es una de las hojas de cálculo más utilizadas y conocidas en el mundo mediante la cual se pueden realizar múltiples operaciones con datos. De manera general, se puede decir que “Microsoft Excel es un programa de hojas de cálculo disponible en el paquete Microsoft Office. MS Excel se utiliza para crear hojas de cálculo para almacenar y organizar datos en formato de tabla.” (Excelx, n.d.)

Fase 2

Ejecutar el programa de Python haciendo uso de la información que se tiene en excel, para poder analizar los datos y construir el modelo RFM junto con la automatización de los procesos repetitivos. Todo lo anterior, en un ambiente colaborativo para el equipo de investigación como Google Colab.

Fase 3

Definir las técnicas de análisis de datos: datos descriptivo e hipótesis, que tipo de información se esperaba o el objetivo a conocer con los parámetros estadísticos y con qué fin, ¿para identificar que patrones? ¿Con que contexto?

Tabla 4 Técnica de Análisis de Datos

Objetivo Específico	Técnicas de Análisis	Actividades
<p>Limpieza de Datos</p> <p>Asegurar la integridad del conjunto de datos mediante la identificación y eliminación de valores nulos y la selección de columnas relevantes para el análisis, garantizando que solo se utilicen datos completos y consistentes.</p>	<p>Limpieza de Datos</p> <ul style="list-style-type: none"> -Verificación de valores nulos. -Eliminación de filas donde las columnas 'A' o 'B' tengan valores NaN. -Filtrado de columnas relevantes para el análisis. $df_{limpio} = df[df['A'].notna() \& df['B'].notna()]$	<p>Limpieza de Datos</p> <ul style="list-style-type: none"> -Cargar el dataset y verificar la integridad de los datos. -Realizar limpieza de datos según los criterios establecidos.
<p>Transformación de Datos</p> <p>Convertir las columnas de fechas y categóricas a tipos de datos adecuados, y aplicar filtros temporales para limitar el análisis a un rango de fechas significativo, facilitando así el análisis posterior.</p>	<p>Transformación de Datos</p> <ul style="list-style-type: none"> - Conversión de las columnas 'ANIO', 'MES' y 'CIA' a tipo entero. - Filtrado por fechas: Fecha de inicio: '2019-01-01' Fecha de fin: '2024-08-31' Exclusión de fechas: '2020-03-06' a '2022-06-30'. $X_{entero} = \text{int}(X)$ $df_{filtrado} = df[(Fecha_{ini} \leq FEC_{EMI}) \& (FEC_{EMI} \leq Fecha_{fin})]$	<p>Transformación de Datos</p> <ul style="list-style-type: none"> - Aplicar filtros de fecha y eliminar filas no deseadas. - Transformar columnas a tipos adecuado.
<p>Análisis Descriptivo</p> <p>Realizar un análisis descriptivo exhaustivo que incluya el conteo de valores únicos, la identificación de la cantidad total de clientes y la determinación de las fechas mínima y máxima, con el fin de obtener una visión general del comportamiento de los clientes.</p>	<p>Análisis Descriptivo</p> <ul style="list-style-type: none"> - Contar Valores Únicos: Esto se puede calcular utilizando la función de cardinalidad de un conjunto, que cuenta el número de elementos distintos. - Fechas Mínima y Máxima: Se utilizan funciones de agregación, como mínimo (min) y máximo (max), para identificar los extremos del conjunto de datos. <p>Contar Valores Únicos: $n_{valores_unicos} = \{x \in X\}$</p> <p>Fechas Mínima y Máxima: $fecha_min = \min(FEC_EMI)$</p> <p>$fecha_max = \max(FEC_EMI)$</p>	<p>Análisis Descriptivo</p> <ul style="list-style-type: none"> -Realizar análisis descriptivo y explorar los datos para identificar patrones y tendencias. -Visualizar datos mediante histogramas y gráficos de series temporales.

<p>Análisis de Series Temporales</p> <p>Calcular y visualizar el promedio de las fechas de última compra de los clientes, así como identificar tendencias en el comportamiento de compra a lo largo del tiempo, permitiendo una mejor comprensión de la lealtad del cliente.</p>	<p>Análisis de Series Temporales</p> <p>-Gráfico del promedio de fechas de última compra de los clientes. -Calcular la fecha promedio en la que los clientes dejaron de comprar.</p> $Fecha_{promedio} = \frac{1}{n} \sum_{i=1}^n FEC_EMI_i$ <p>Histogramas: $f(a, b) = \frac{n(a, b)}{N}$</p>	<p>Análisis de Series Temporales</p> <p>-Realizar análisis descriptivo y explorar los datos para identificar patrones y tendencias. -Visualizar datos mediante histogramas y gráficos de series temporales.</p>
<p>Análisis RFM</p> <p>Implementar el modelo RFM para segmentar a los clientes en función de su recencia, frecuencia y valor monetario, asignando puntuaciones que permitan identificar grupos de clientes y diseñar estrategias de marketing personalizadas.</p>	<p>Análisis RFM</p> <p>Formula Recencia: $R = Fecha_actual - Fecha_ultima_compra$</p> <p>Frecuencia: $F = \text{Número de compras en un período}$</p> <p>Histograma del Valor Monetario:</p> $M = \sum_{i=1}^n Monto_compra$ <p>Asignación de Puntuaciones:</p> $P_k = \frac{k}{N} \times 100$	<p>Implementación del Modelo RFM</p> <p>-Calcular las métricas RFM y asignar puntuaciones. -Agrupar clientes en segmentos basados en las puntuaciones RFM.</p>
<p>Pruebas de Normalidad</p> <p>Evaluar la normalidad de las distribuciones de las métricas RFM a través de pruebas estadísticas y gráficos, asegurando que los análisis posteriores sean válidos y que se puedan aplicar técnicas estadísticas adecuadas.</p>	<p>Pruebas de Normalidad</p> <p>-Gráfico Q-Q (Comparación de cuartiles)</p> $Q_i = Cuartil_i(\text{datos}) \text{ vs } Q_i = Cuartil_i(\text{distribución teórica})$ <p>-Prueba de Shapiro-Wilk para cada columna del RFM.</p> $\frac{(\sum_{i=1}^n a_i x_{(i)})^2}{(\sum_{i=1}^n x_i \bar{x})^2}$ <p>p-valor $< \alpha$ indica no normalidad.</p> <p>-Prueba Kolmogorov-Smirnov.</p> $D = \max Fn(x) - Gm(x) $ <p>p-valor $< \alpha$ indica que las distribuciones son diferentes.</p>	<p>Evaluación de Normalidad</p> <p>Realizar pruebas de normalidad y visualizar distribuciones.</p>

<p>Normalizar las métricas RFM para garantizar que todas las variables se encuentren en la misma escala, facilitando así comparaciones y análisis más precisos entre los diferentes segmentos de clientes.</p>	<p>Normalización Min-Max:</p> $X' = \frac{X - X_{min}}{X_{max} - X_{min}}$ <p>Estandarización (Z-score):</p> $Z = \frac{X - \mu}{\sigma}$ <p>Donde μ es la media y σ es la desviación estándar.</p>	<p>Documentación y Presentación Documentar los hallazgos y preparar una presentación que resuma el análisis y recomendaciones.</p>
---	--	---

Fuente: Elaboración Propia

Pseudocódigo

En la ilustración 3 se puede ver el detalle del pseudocódigo para la implementación del modelo RFM desde el tratamiento de los datos, cálculo de cada una de las tres variables del modelo y la aplicación de los dos enfoques: por cuantiles y por el algoritmo k-means.

Ilustración 3 Pseudocódigo Modelo RFM (dos enfoques)

```
# 1. Lectura de los datos
datos = leer_datos('datos.csv')

# 2. Limpieza, filtrado y transformacion
datos = limpiar_datos(datos)
datos = filtrar_datos(datos)
datos = transformar_datos(datos)

# 3. Analisis exploratorio
analisis_exploratorio_datos(datos)

# 4. Calculo de Recencia, Frecuencia y Valor Monetario (RFM)
recencia = calcular_recencia(datos)
frecuencia = calcular_frecuencia(datos)
valor_monetario = calcular_valor_monetario(datos)

# Escenario 1: Enfoque basado en cuantiles
def enfoque_cuantiles(rfm_data):
# 5. Calcular los cuantiles para cada variable de RFM
recencia_cuantiles = cuantiles(rfm_data['Recencia'])
frecuencia_cuantiles = cuantiles(rfm_data['Frecuencia'])
valor_monetario_cuantiles = cuantiles(rfm_data['ValorMonetario'])

# 6. Puntuar cada observacion en funcion de los cuantiles (1-5)
rfm_data['R_puntaje'] = puntuar(recencia_cuantil)
rfm_data['F_puntaje'] = puntuar(frecuencia_cuantil)
rfm_data['M_puntaje'] = puntuar(valor_monetario_cuantil)

# 7. Generar segmentos a partir de reglas de negocio
rfm_data['Segmento'] = asignar_segmento(rfm_data)

# 8. Analisis de resultados
analizar_segmentos(rfm_data)

# Escenario 2: Enfoque basado en K-means
def enfoque_kmeans(rfm_data):
# 5. Pruebas de normalidad para cada variable
verificar_normalidad(rfm_data['Recencia'])
verificar_normalidad(rfm_data['Frecuencia'])
verificar_normalidad(rfm_data['ValorMonetario'])

# 6. Normalizacion de los datos
rfm_data_normalizado = normalizar_datos(rfm_data)

# 7. Aplicacion de K-means con el numero optimo de clusters
numero_optimo_clusters = determinar_clusters_optimos(rfm_data_normalizado)
kmeans = KMeans(n_clusters=numero_optimo_clusters)
rfm_data['Cluster'] = kmeans.fit_predict(rfm_data_normalizado)

# 8. Analisis de resultados
analizar_clusters(rfm_data)

# Ejecucion de cada escenario
enfoque_cuantiles(rfm_data)
enfoque_kmeans(rfm_data)
```

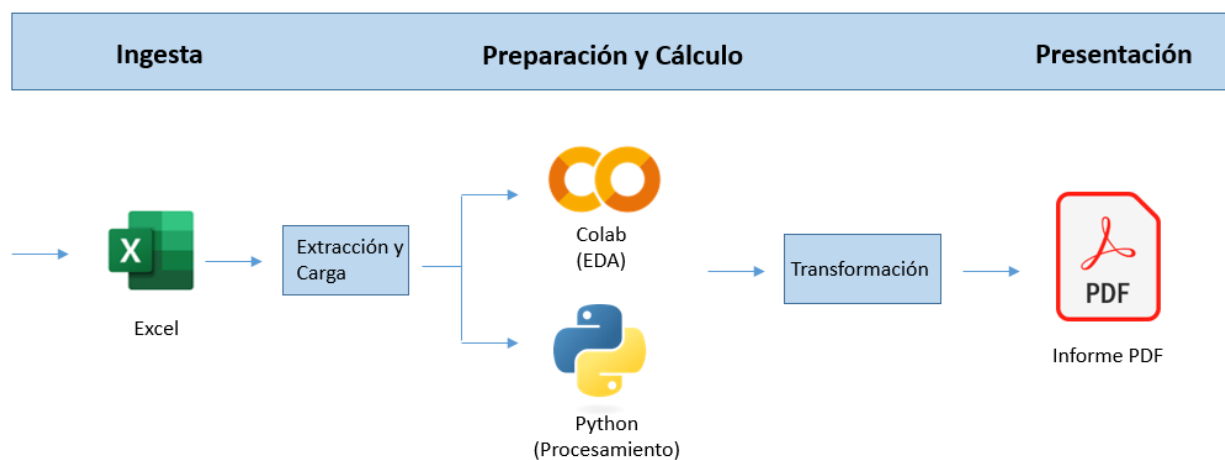
Fuente: Elaboración Propia

CAPÍTULO 4 9. ANÁLISIS Y DISCUSIÓN DE RESULTADOS

Desarrollo de los Modelos de Segmentación

El proceso de transformación de los datos para lograr obtener los resultados de la investigación se compone de unos pasos como se muestra en el siguiente diagrama.

Ilustración 4 Proceso de Transformación de Datos en Resultados



Fuente: Elaboración Propia

Filtrado y Análisis Exploratorio de Datos

Con el objetivo de optimizar la información y comprender mejor la estructura de la base de datos utilizada, se establece una etapa de filtrado y análisis. Este proceso se basa en el principio de organización, que permite asegurar que la información necesaria esté accesible para nuestros modelos analíticos.

La primera fase consiste en filtrar los datos según criterios específicos. Durante esta etapa, se eliminan las columnas duplicadas que puedan generar confusiones o redundancias en el análisis. También se unifican los datos de las columnas que, aunque representan la misma información, pueden contener variaciones como puntos o formatos distintos. Este proceso de unificación asegura que todos los registros sean consistentes y que la información se presente de manera uniforme. Además, se rellenan los campos vacíos con valores adecuados, lo que garantiza que el análisis posterior no se vea afectado por la falta de datos.

La etapa de filtrado no solo garantiza la relevancia de los datos seleccionados, sino que también mejora la eficiencia en el procesamiento de la información. Al limitar el conjunto de datos a aquellos registros que son significativos mediante la aplicación de los 4 filtros estipulados con anterioridad, se puede enfocar el análisis en las variables más pertinentes, permitiendo una mejor interpretación de los resultados, este paso nos permitió reducir nuestra base de datos de:

9302 Rows x 51 Columns.

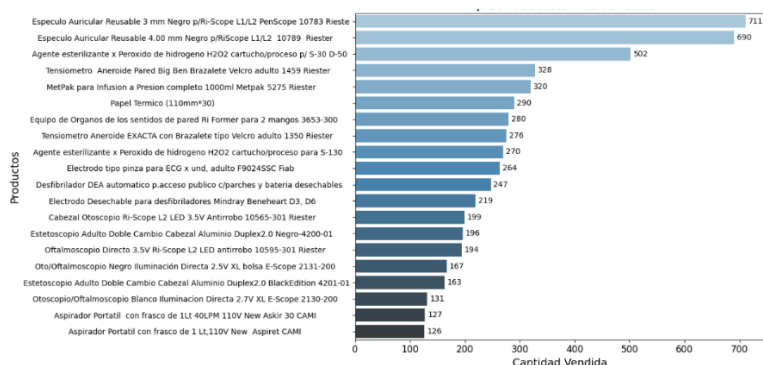
a

3323 Rows x 21 Columns.

De los cuales ninguna celda cuenta con valores nulos y se concluye que la cantidad de clientes a clasificar arroja un total de 425.

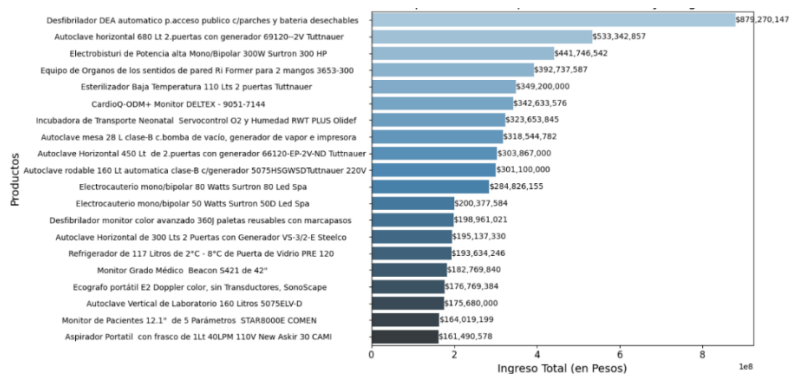
Una vez completada la etapa de filtrado, se procede a la exploración de los datos. Esta fase es crucial para profundizar en el conocimiento del comportamiento de las ventas durante el periodo seleccionado. Se examinan desde la cantidad y tipo de clientes que se van a clasificar en el modelo Ilustración 5, las variaciones en las ventas a lo largo del tiempo y su distribución geográfica en el país. A través de técnicas de visualización y análisis descriptivo, se puede identificar cómo diferentes factores, como la ubicación, el tipo de cliente y el producto, afectan las ventas.

Ilustración 7 Top 20 productos más vendidos



Fuente: Elaboración Propia.











Ilustración 8 Top 20 Productos más rentables económicamente



Fuente: Elaboración Propia.

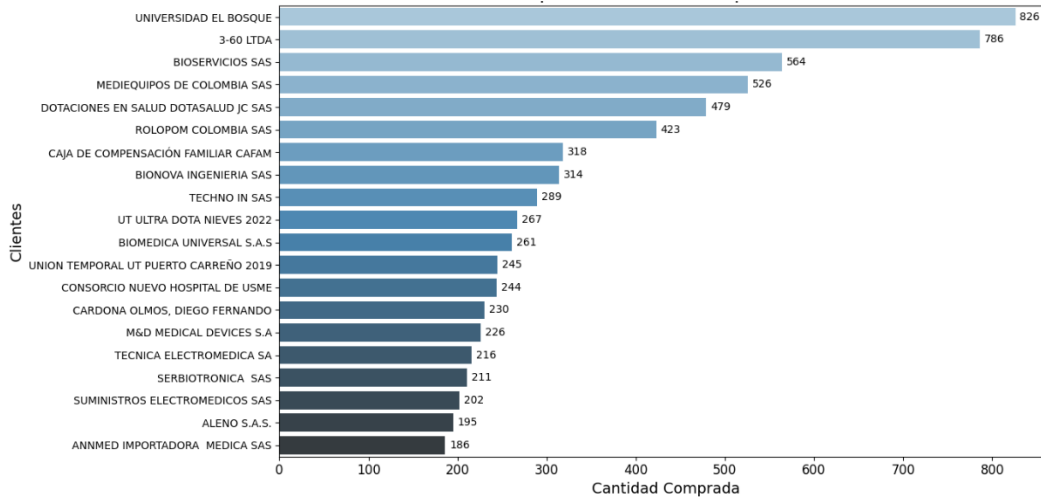
De la Ilustración 7 podemos ver que no todo lo que se vende en volumen es más rentable económicamente, por eso tenemos que diferenciar estos 2 aspectos permitiéndonos descubrir que los productos más vendidos son equipos pequeños y de bajo costo, y en algunas situaciones solo consumibles comparados con los equipos de alto costo pero que son vendidos en cantidades más limitadas ilustración 8, en la Tabla 5 podemos conocer estos equipos e insumos.

Tabla 5 Comparativo de Productos más rentables VS más vendidos.

Top 5 Productos más rentables económicamente				
				
Desfibrilador DEA automático p. acceso publico c/parches y batería desechables	Autoclave horizontal 680 Lt 2.puertas con generador 69120--2V Tuttnauer	Electrobisturi de Potencia alta Mono/Bipolar 300W Surtron 300 HP	Equipo de Organos de los sentidos de pared Ri Former para 2 mangos 3653-300	Esterilizador Baja Temperatura 110 Lts 2 puertas Tuttnauer
Top 5 Productos más vendidos				
				
Espejo Auricular Reusable 3 mm Negro p/Ri-Scope L1/L2 PenScope 10783 Rieste	Espejo Auricular Reusable 4.00 mm Negro p/RiScope L1/L2 10789 Riester	Agente esterilizante x Peroxido de hidrogeno H2O2 cartucho/proceso p/ S-30 D-50	Tensiómetro Aneroide Pared Big Ben Brazaletes Velcro adulto 1459	Riester MetPak para Infusion a Presion completo 1000ml Metpak 5275

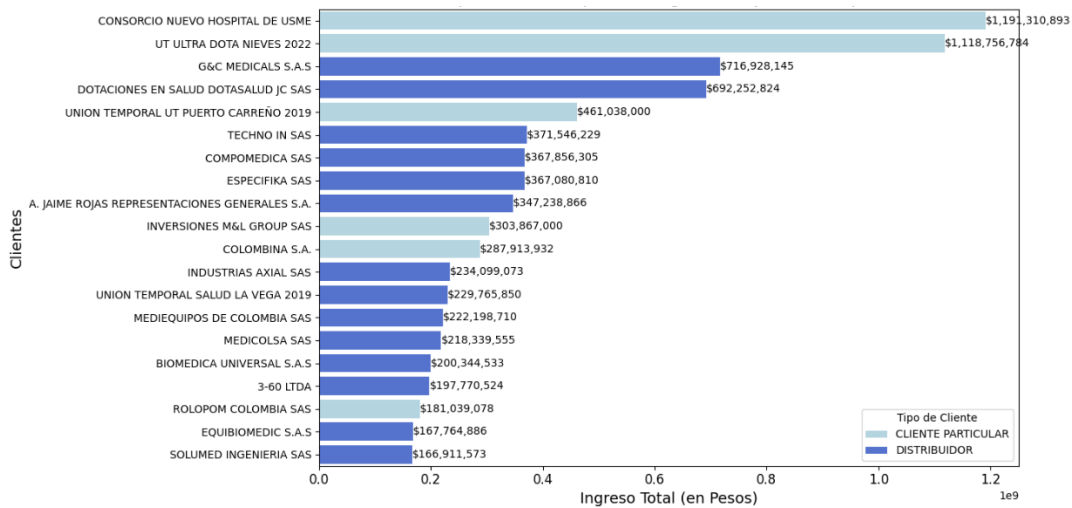
Fuente: Elaboración Propia.

Ilustración 9 Top 20 Clientes más compradores



Fuente: Elaboración Propia.

Ilustración 10 Top 20 Clientes Que más Ingresos Dejan a la Empresa

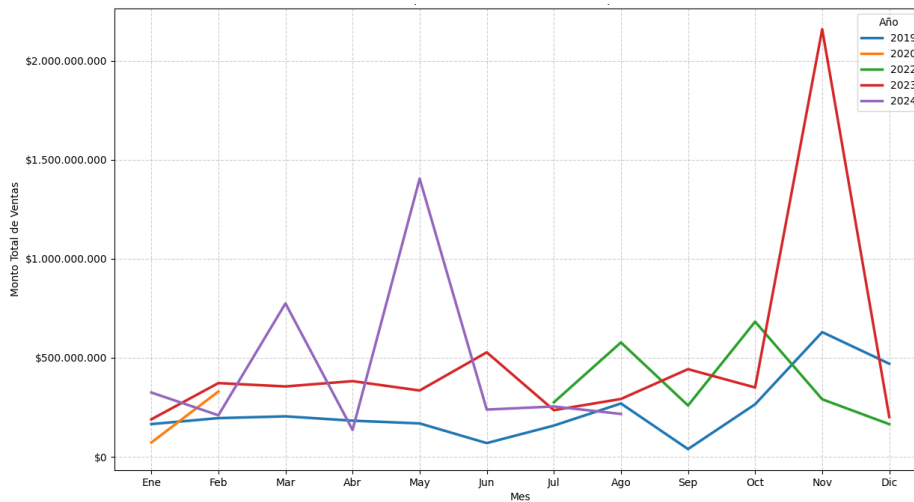


Fuente: Elaboración Propia.

Además, se implementa un análisis de tendencias temporales que permitirá entender las fluctuaciones estacionales y la evolución de las ventas a lo largo del año, permitiéndonos visualizar que el periodo de mayor venta puede llegar a ser en diciembre, pero el resto del año es

demasiado oscilante. Esta información es esencial para la toma de decisiones estratégicas y el desarrollo de acciones de marketing más efectivas.

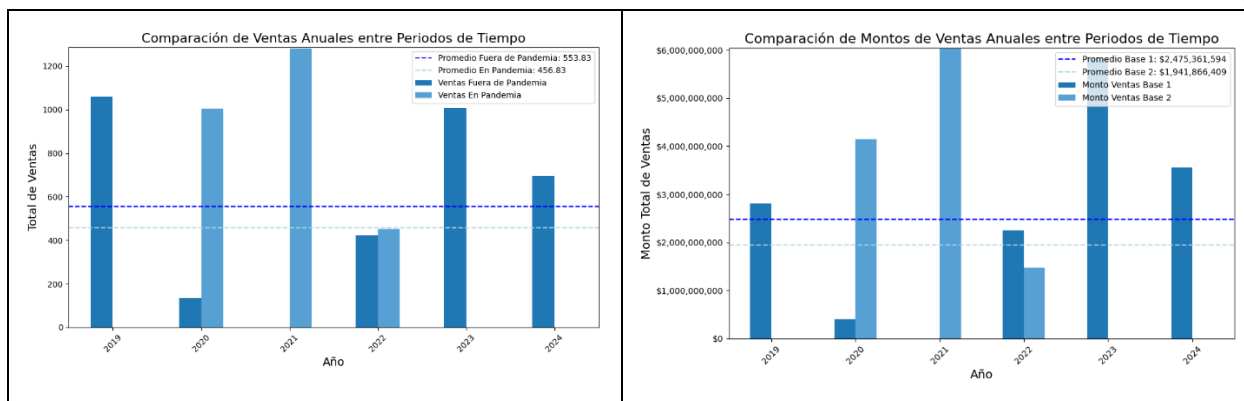
Ilustración 11 Análisis de tendencias temporales



Fuente: Elaboración Propia.

Para finalizar el análisis exploratorio de los datos se realizó una comparación de los promedios de montos de ventas y cantidades de equipos/insumos vendidos en los periodos externos a pandemia Vs Dentro de pandemia para corroborar si era necesario excluir estos datos, concluyendo que, aunque los promedios en pandemia si eran mayores no son considerables Ver ilustración 12, de igual forma al ser una petición del cliente se realizan los siguientes pasos excluyendo estos datos.

Ilustración 12 Comparación de ventas y montos entre los periodos de tiempo establecidos



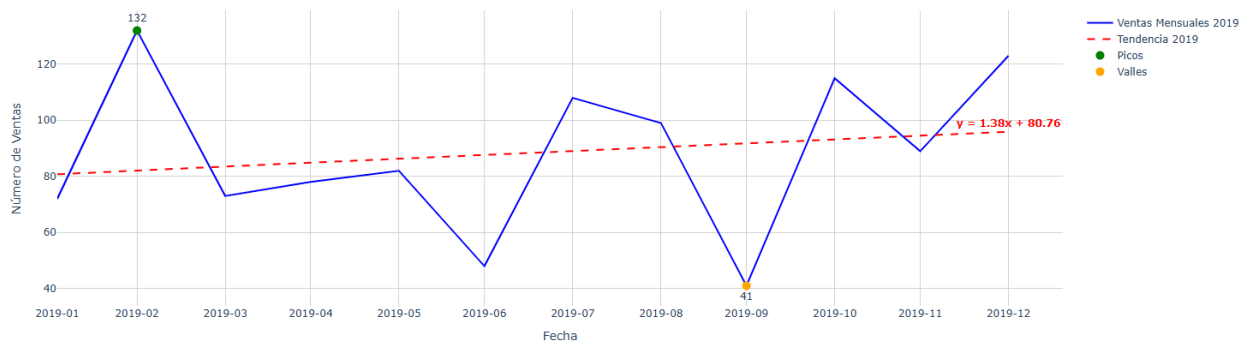
Fuente: Elaboración Propia.

Modelo RFM – Enfoque Cuantiles

En el contexto de RFM cuando se aplica el método por Cuantiles, se dividen las métricas de Recencia, Frecuencia y Monetario en cinco partes iguales (o quintiles) para asignar una puntuación en función de cada quintil. A cada cliente se le asigna una puntuación de 1 a 5 en función del quintil en el que se ubica en cada métrica: los clientes en el quintil superior obtienen la puntuación más alta, mientras que los del quintil inferior reciben la puntuación más baja.

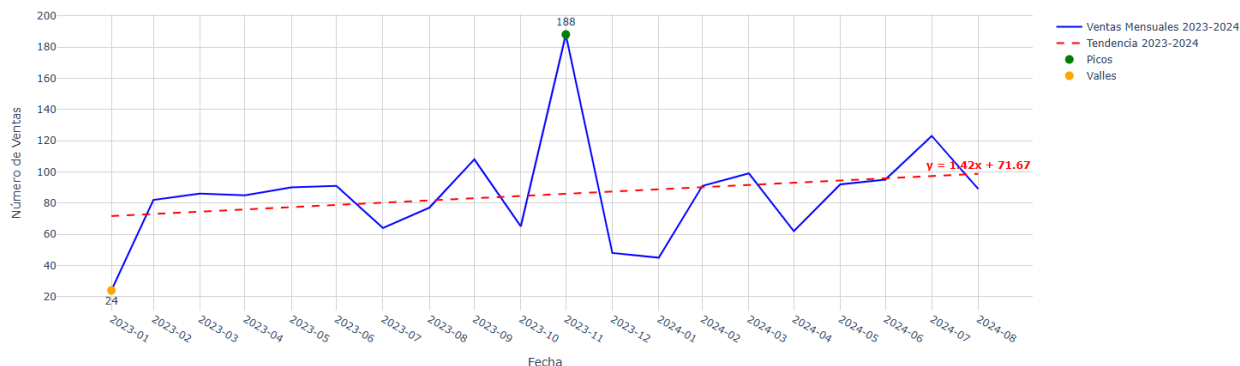
Para dar inicio a la construcción del modelo RFM, previamente se realizó un análisis de serie temporal para entender el comportamiento del número de ventas de los años completos sin tener en cuenta años de pandemia.

Ilustración 13 Número de Ventas Mensual 2019 (con Tendencia, Picos y Valles)



Fuente: Elaboración Propia.

Ilustración 14 Número de Ventas Mensual 2023-2024 (con Tendencia, Picos y Valles)

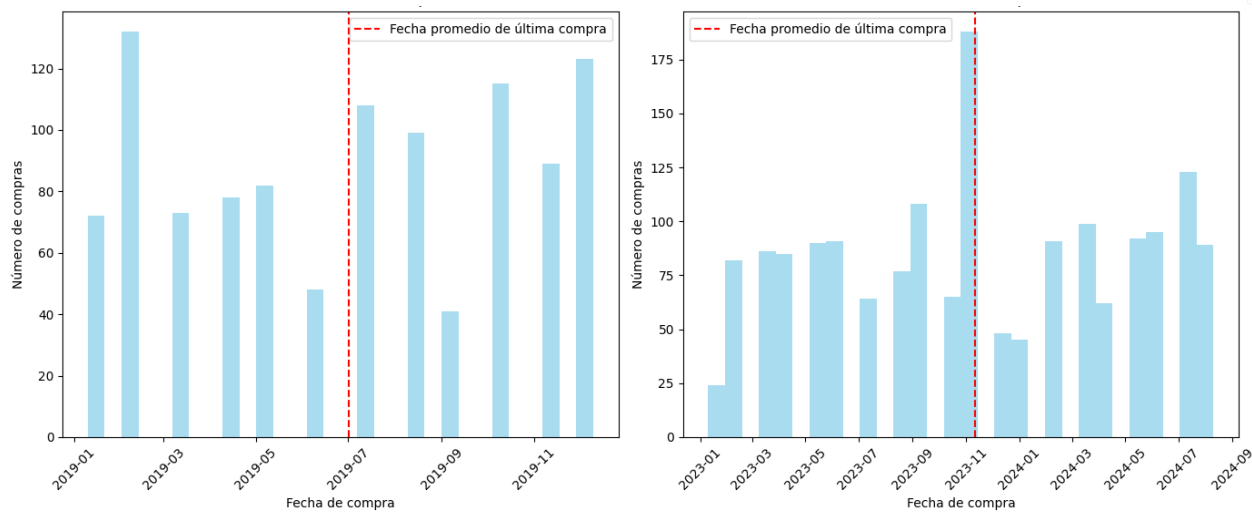


Fuente: Elaboración Propia.

Del comportamiento anterior en las gráficas, se concluye que la tendencia de las ventas año a año muestra un ligero incremento con variaciones estacionales. Antes de la época de pandemia el mayor número de ventas se presentó en febrero mientras que la zona valle de venta fue en el mes de septiembre, después de pandemia el comportamiento empezó a cambiar ya que el mayor volumen de ventas fue en noviembre 2023 y existe una marcada tendencia de poca venta al principio del mes de enero.

En cuanto a la fecha de la última compra de los clientes, en los siguientes gráficos se evidencia que, comparando la posición de la línea punteada, en promedio los clientes están comprando con menos frecuencia en el periodo 2023-2024 en comparación con 2019.

Ilustración 15 Distribución de Fechas de Compra

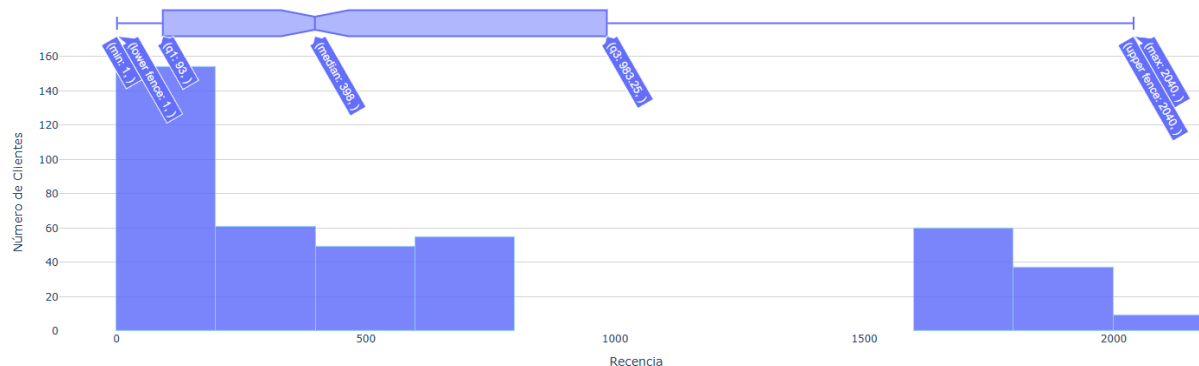


Fuente: Elaboración Propia.

Llevando a cabo el histograma de la Recencia presentado en la gráfica a continuación, en promedio son 398 días, tiempo en el que los clientes dejan de comprar. Sin embargo, la mayor parte de los clientes han realizado compras recientemente de acuerdo con la alta concentración en la parte izquierda del gráfico.

Por otro lado, a medida que se analiza el gráfico hacia la parte derecha, la frecuencia de compras disminuye, lo que indica un grupo considerable de clientes que ha realizado compras hace más tiempo debido a factores como la vida útil de los equipos, cambios en las necesidades del mercado o la misma satisfacción de los usuarios.

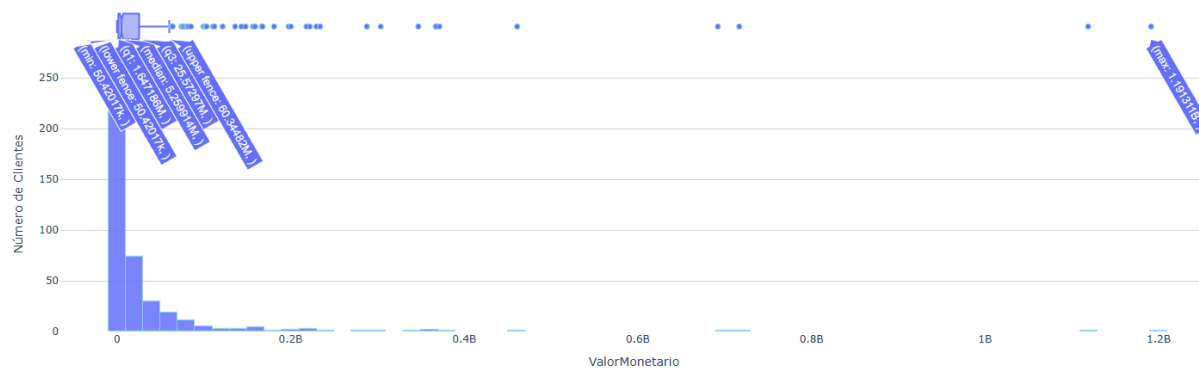
Ilustración 16 Distribución de Recencia de Visitas de los Clientes



Fuente: Elaboración Propia.

Por último, antes de iniciar con la asignación de la puntuación RFM por cliente, se realizó también el histograma de la distribución monetaria, arrojando este que la mayoría de los clientes han generado un valor monetario relativamente bajo, lo que puede indicar una tendencia hacia compras de menor ticket promedio.

Ilustración 17 Distribución Monetaria de los Clientes



Fuente: Elaboración Propia.

Para la asignación de puntuaciones RFM a cada cliente, se ejecuta el modelo en Python, aclarando que las puntuaciones se definen de la siguiente manera:

Recencia_score: 5 - más reciente, 1 - menos reciente

Frecuencia_score: 1 - menos frecuente, 5 - más frecuente

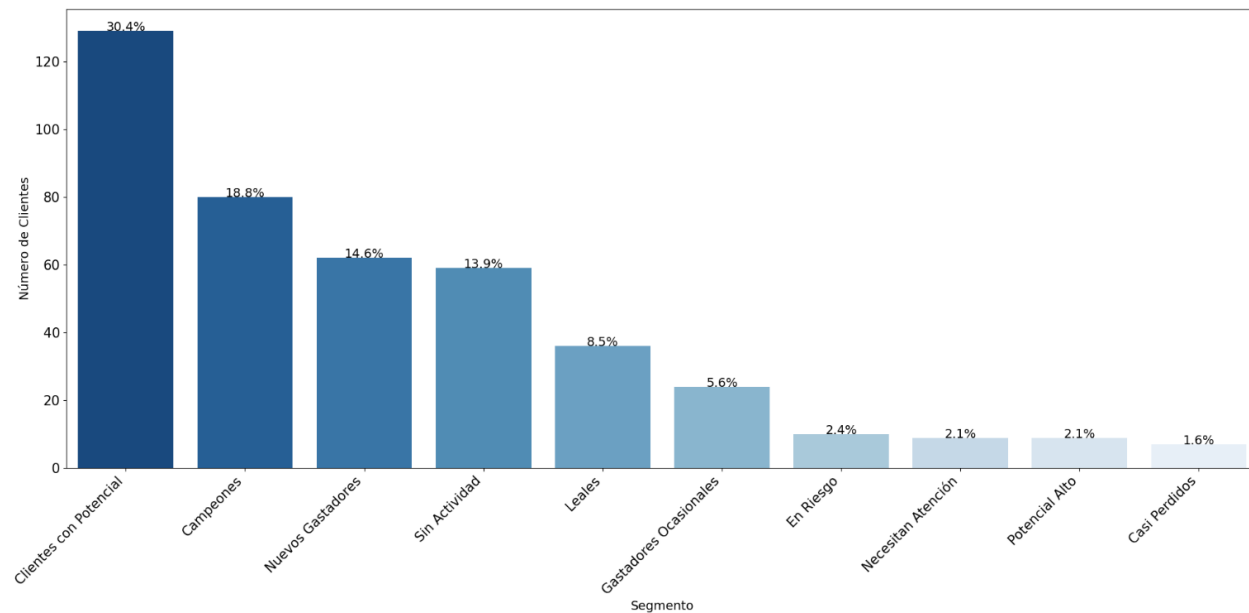
ValorMonetario_score: 1 - menos valioso, 5 - más valioso

RFM_SCORE: Suma de Recencia_score, Frecuencia_score y ValorMonetario_score

Bajo los criterios anteriores se definen las categorías de segmentación para cada grupo de clientes. Según Nair, proveedor en línea especializado en servicios analíticos: “Se han identificado 11 segmentos distintos, cada uno de los cuales corresponde a combinaciones únicas de las métricas RFM, lo que proporciona información sobre los patrones de comportamiento de los clientes” (Nair, A, 2023).

Para la presente investigación se personalizan los segmentos para considerar todas las combinaciones RFM que no están consideradas en las expresiones regulares definidas por el autor anterior.

Ilustración 18 Número de Clientes por Segmento



Fuente: Elaboración Propia.

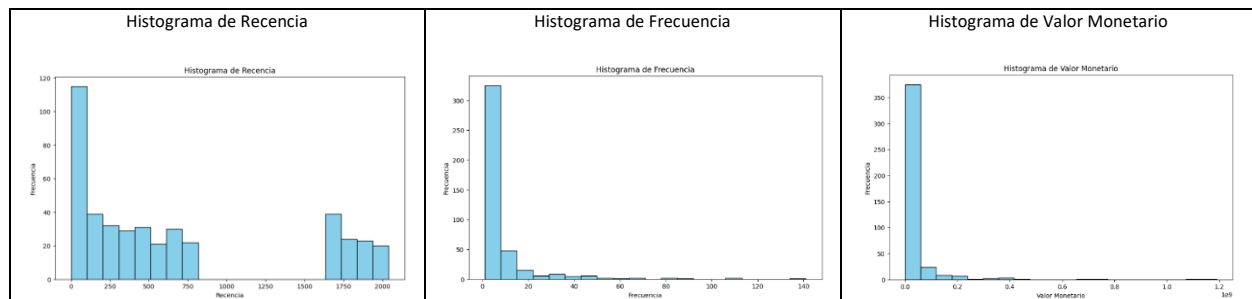
Modelo RFM – Enfoque Clústeres

Teniendo en cuenta el set de datos resultante después de aplicar los métodos de limpieza y transformación de datos, se generaron histogramas para cada una de las 3 variables del modelo (recencia, frecuencia y valor monetario) con el fin de entender cómo están distribuidos.

En la ilustración 19, se puede observar que una gran cantidad de observaciones se agrupan en valores bajos para todas las variables. En el histograma de la recencia se ve un rango de valores sin observaciones debido a la exclusión de fechas que se hizo inicialmente para el periodo comprendido entre 06 de marzo de 2020 y el 30 de junio de 2022 como se explicó en la etapa del manejo inicial de los datos.

Lo anterior no significa que haya un gran número de observaciones con valores bajos para las 3 variables, ya que puede haber una gran variedad de combinaciones posibles entre las mismas.

Ilustración 19 Histogramas de variables RFM



Fuente: Elaboración Propia.

Normalización de datos

La información de los histogramas, permite ver que hay una alta variabilidad en las magnitudes de cada una de las tres variables del modelo (eje X de cada una de las gráficas de la ilustración 19), por lo cual se hace necesario someter a los datos a un proceso de normalización de manera que los valores con una magnitud muy alta o muy baja no agreguen ruido a la creación de los clústeres.

Para ello, se evalúa la implementación de dos de las técnicas más conocidas: Estandarización (Z-score) y Escalado Min-Max. La decisión de cuál de las dos aplicar, depende de la distribución de los datos: Si los datos siguen una distribución normal, se usa Z-score, de lo contrario, será el escalado Min-Max.

Aunque los histogramas anteriores dan una idea inicial, es necesario aplicar técnicas estadísticas como la Prueba de Shapiro-Wilk y la Prueba Kolmogorov-Smirnov. La aplicación de estas técnicas confirmaron que los datos no siguen una distribución normal, como lo demuestran los resultados de la Tabla 6.

Tabla 6 Resultados pruebas de normalidad

Prueba de Shapiro-Wilk	Prueba de Kolmogorov-Smirnov
<p>Recencia: Estadístico: 0.796114106785001 Valor p: 1.0163460272497617e-22 <i>La distribución no es probablemente normal.</i></p> <p>Frecuencia: Estadístico: 0.4697179714902051 Valor p: 1.3793839077710775e-33 <i>La distribución no es probablemente normal.</i></p> <p>Valor Monetario: Estadístico: 0.31677356263287837 Valor p: 7.627872595956369e-37 <i>La distribución no es probablemente normal.</i></p>	<p>Recencia: Estadístico: 0.9105882352941177 Valor p: 0.0 <i>La distribución no es probablemente normal.</i></p> <p>Frecuencia: Estadístico: 0.8413447460685429 Valor p: 0.0 <i>La distribución no es probablemente normal.</i></p> <p>Valor Monetario: Estadístico: 1.0 Valor p: 0.0 <i>La distribución no es probablemente normal.</i></p>

Fuente: Elaboración Propia.

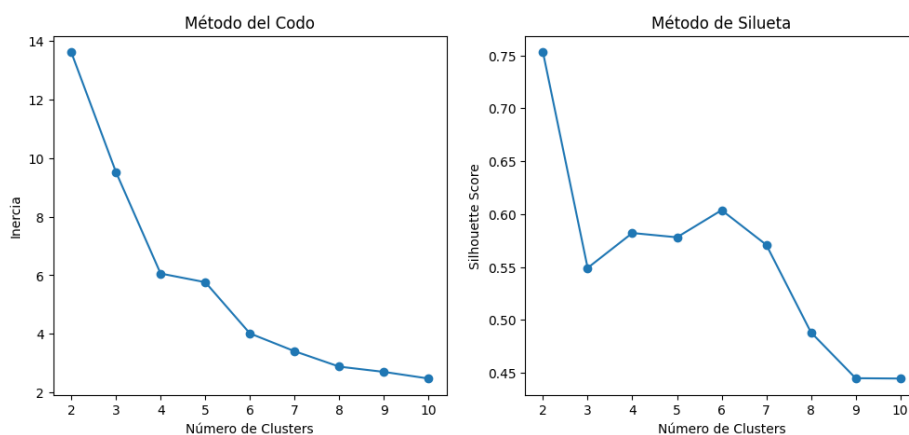
Con base en esta confirmación, se aplicó la técnica de escalado Min-Max a los datos.

Número óptimo de clústeres

Durante la construcción de un modelo de clusterización por k-means, es necesario definir el número de clústeres a generar y éste se agrega como parámetro al modelo.

Para determinar el número óptimo de clústeres se cuenta con algunas herramientas visuales como el método del codo y el método de la silueta. Con base en el set de datos disponible y después del proceso de normalización de los datos, se ejecutaron estos dos métodos dando como resultado que seis clústeres era lo indicado. Los resultados se ven en la ilustración 20.

Ilustración 20 Método del codo y Método de Silueta



Fuente: Elaboración Propia.

Desarrollo del Algoritmo K-means

Una vez definido el número óptimo de clústeres, se ejecuta el modelo en Python. Mediante la ejecución del algoritmo k-means, cada observación se asigna a uno de los seis clústeres definidos previamente, de acuerdo con la dinámica del algoritmo. Este proceso agrupa observaciones similares, buscando maximizar la homogeneidad dentro de cada grupo y la heterogeneidad entre los diferentes grupos.

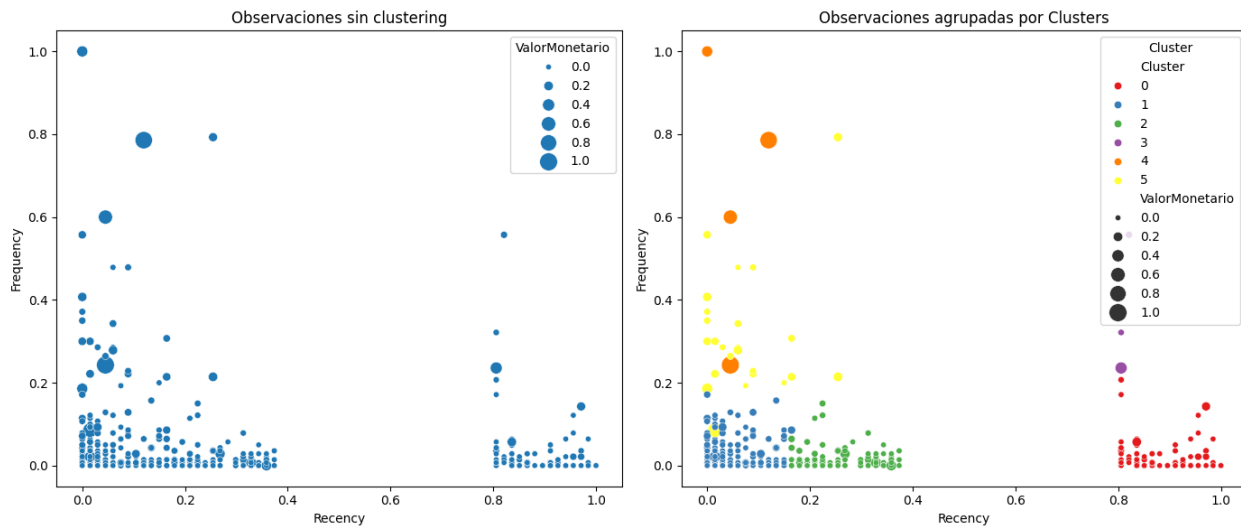
El resumen de los resultados tanto de cantidad de observaciones por cada cluster, como la gráfica de la clasificación de las observaciones, se pueden ver en la Tabla 7 y la ilustración 21, respectivamente:

Tabla 7 Cantidad de observaciones por clúster

Clúster	Cantidad Observaciones
1	161
2	128
0	103
5	26
4	4
3	3

Fuente: Elaboración Propia.

Ilustración 21 Observaciones antes y después de clasificación



Fuente: Elaboración Propia.

Con base en la clasificación de las observaciones en los seis clústeres diferentes, se genera el cálculo de la media para cada una de las variables del modelo RFM en cada cluster, manteniendo la normalización de las métricas, tal como se ve en la ilustración 22:

Ilustración 22 Calificación promedio de variables por clúster

Cluster	Recencia	Frecuencia	ValorMonetario
0	0.885009	0.017198	0.009696
1	0.050204	0.032786	0.019004
2	0.269698	0.013951	0.016281
3	0.810855	0.371429	0.160062
4	0.052477	0.657143	0.708003
5	0.066737	0.312363	0.115197

Fuente: Elaboración Propia.

Los resultados de la ilustración anterior se pueden interpretar de la siguiente manera:

- **Recencia:** Los clústeres con mayor valor (cercano a 1), corresponden a aquellas observaciones que tiene una recencia más alta, es decir, que su última compra es de

las más lejanas en el tiempo de todo el grupo de observaciones. En este caso, los clústeres 0 y 3 son los de mayor valor de recencia, esto se puede comprobar en la ilustración 16. De manera contraria, los clústeres 1, 4 y 5 son aquellos cuya última compra se presentó hace no mucho tiempo. En términos de negocio, un valor de recencia menor es mejor, ya que corresponde a los clientes que tuvieron una transacción recientemente.

- **Frecuencia:** En esta variable el sentido es contrario, ya que un valor mayor de frecuencia significa un cliente que ha comprado varias veces en el periodo de tiempo analizado. En este sentido, los clústeres 4, 3 y 5 representan las observaciones con mayor número de compras. En términos de negocio, un puntaje de frecuencia mayor es mejor.
- **Valor Monetario:** Finalmente, el valor monetario sigue la misma lógica de la frecuencia: un valor mayor es positivo para la organización y refiere a los clientes que han gastado una cantidad de dinero mayor en promedio. En la ilustración 21, los clústeres con mayor puntaje se pueden ver según el tamaño de la burbuja.

El promedio de cada variable del modelo RFM en cada uno de los seis clústeres generados, permite identificar la caracterización de las observaciones que componen cada uno de ellos, dando lugar a configurar los *perfiles* de clientes, similares a los generados en el enfoque anterior (cuantiles). En la tabla 8 se ve el detalle de la información de cada uno de los clústeres en cuanto a cantidad de observaciones clasificadas en cada uno de ellos y una definición de características según el puntaje de la recencia, frecuencia y valor monetario.

Tabla 8 Descripción de Clústeres k-means

ID Clúster	Cantidad Observaciones	Definición
0	103	Clientes Inactivos Baja frecuencia y bajo gasto, poco involucrados
1	161	Nuevos Clientes Clientes recientes con bajo valor, podrían estar en su primera compra.
2	128	Clientes Esporádicos Baja frecuencia y bajo valor, pero ligeramente más activos que el Clúster 0.
3	3	Clientes Ocasionales Alta recencia, con frecuencia y valor moderados; compran ocasionalmente.
4	4	Clientes Valiosos y Frecuentes Alta frecuencia y valor monetario, probablemente los mejores clientes.
5	26	Clientes en Crecimiento Recientes con frecuencia y valor moderado; pueden convertirse en clientes leales

Fuente: Elaboración Propia.

De esta manera, se consigue el resultado según el enfoque de generación de clústeres para el modelo de RFM. De acuerdo con las puntuaciones generadas en cada uno de los clústeres para cada una de las variables, se pueden identificar los clientes según el valor que le aportan a la compañía. Este resultado se puede potencializar al combinarlo con otros modelos como el CLV (*Customer Lifetime Value*) para tener la rentabilidad de cada grupo (Ahmad M. A. Zamil & T. G. Vasista, 2021).

Análisis de Resultados

De acuerdo con el desarrollo del modelo RFM a través de los dos enfoques presentados: Cuantiles y clústeres por k-means, se obtuvieron resultados interesantes para la segmentación de clientes en el contexto del estudio. Por un lado, el primer enfoque generó 10 segmentos de clientes con base en el resultado dado por las reglas de negocio, el conocimiento del mercado y la industria, mientras que para el algoritmo k-means se definió que el número óptimo de clústeres fuera de 6.

Por medio del enfoque de cuantiles, usando reglas de negocio para la configuración de los segmentos a generar se tiene que, de los 10 grupos generados, 6 de ellos representan el 91.8% de todas las observaciones disponibles. La definición dada para estos segmentos es: Clientes con potencial, campeones, nuevos gastadores, sin actividad, leales y gastadores ocasionales.

El grupo más representativo es el de Clientes con Potencial, lo cual es una noticia positiva para la empresa al tener un terreno por explorar y proponer estrategias de marketing que exploten el potencial de este segmento llevándolo a convertirse en clientes recurrentes o leales y aumenten tanto la frecuencia como el valor monetario.

Seguido están los campeones, quienes tienen una calificación por encima del promedio en cada una de las tres variables y son clientes para los que la empresa debe propender en su cuidado y mantenerlos, dado que son los del segmento preferencial que generan mejores resultados.

En tercer lugar, están los nuevos gastadores y con ellos una oportunidad de generar una relación perdurable en el tiempo y de confianza para convertirlos en clientes permanentes. Las actividades de consolidación de imagen y recordación son importantes para que no sean clientes de una sola vez.

Respecto al enfoque utilizando el algoritmo k-means, de los seis clústeres generados, los clústeres 1, 2 y 0 recogen el 92.2% de las observaciones. Las características de estos grupos se pueden ver en la ilustración 21 y una descripción más alineada a definición de negocio en la tabla 8.

Al analizar la gráfica de la segmentación de las observaciones, se ve claramente (ilustración 21) cómo se agrupan en 4 grupos definidos en el espacio y 2 restantes recogen observaciones atípicas con valores superiores al promedio, lo que hace que se agrupen de manera diferente a los demás.

En general, las observaciones que se agruparon en los 3 clústeres más representativos recogen una variación alta de valores en la recencia, desde las observaciones con mayor recencia hasta la más baja de ellas. Respecto a la frecuencia y al valor monetario, se agrupan las

observaciones con valores inferiores a la media normalizada de los datos, esto quiere decir que hay más observaciones con una frecuencia y valor monetario bajos y que solo algunas observaciones ostentan un alto valor que sube el promedio, razón por la cual se agrupan en clústeres diferentes, tal como se mencionó en el párrafo anterior.

Elección de Enfoque

Teniendo en cuenta el proceso llevado a cabo para la generación de cada uno de los dos modelos y el análisis de resultados posterior, se debe elegir cuál de los dos enfoques se presentará a la organización para su implementación y apoyo en las decisiones de marketing y comercial.

A pesar de que el modelo RFM ha sido ampliamente estudiado e implementado en las organizaciones, no es tan extensa la bibliografía respecto a la comparación de los dos enfoques elegidos en el presente estudio en la industria de equipos biomédicos. De manera general, algunos autores que han abarcado esta comparación como (Doğan et al., n.d.) refieren a las virtudes de cada uno de los dos enfoques, sin embargo, resaltan la importancia para el negocio en entender las características no solo de compra de los clientes sino también integrarlas a un conocimiento sociodemográfico que permita construir estrategias adecuadas tanto de marketing como de un enfoque comercial. Este enfoque de conocimiento al cliente desde diferentes contextos es soportado también por (Ahammed et al., 2023) al proponer variables demográficas específicas como género, edad, educación y nivel de ingresos para potencializar la segmentación y una optimización del CRM (Customer Relationship Management).

Así, se resalta la importancia de conocer las reglas del negocio que se adaptarán mejor a la realidad y que son más susceptibles a tener buenos resultados.

De acuerdo con lo anterior, el enfoque recomendado para la empresa es el modelo RFM basado en cuantiles teniendo en cuenta que la organización no ha tenido antes una estrategia soportada en datos, pero tienen un amplio conocimiento del negocio y de la caracterización de sus clientes. Si bien la ejecución de un algoritmo como k-means disminuye la subjetividad al

momento de crear los segmentos, también puede omitir detalles específicos del negocio y la relación entre variables que son importantes en la estrategia de marketing y comercial.

Otro argumento a favor de este enfoque es que la configuración de los segmentos es flexible ante cambios en las decisiones estratégicas de la empresa y permite ajustarse a la visión y objetivos de negocio que tengan las diferentes áreas. De esta manera, un cambio de enfoque de producto o estrategia comercial permite ajustar los clústeres, personalizar la segmentación e identificar rápidamente los clientes que clasifiquen según los intereses de la organización, delimitando el alcance de las iniciativas, aumentando así la efectividad de las mismas. Un ejemplo de esto puede ser un cambio en políticas de venta cruzada que aumente el potencial del segmento Gastadores Ocasionales y permita que migren al segmento de Leales.

Finalmente, se reitera que este será la primera iniciativa de la organización respecto al agrupamiento de sus clientes con base en un enfoque desde los datos, por lo cual se recomienda utilizar un enfoque que ofrezca simplicidad y facilidad de interpretación en la creación de los clústeres: El enfoque por cuantiles explica claramente cómo se hace la división y aunque el algoritmo k-means no es precisamente una caja negra, requiere mayor esfuerzo para explicar la configuración de los grupos y la separación de éstos.

Validación y Adaptabilidad del Modelo

El enfoque de Cuantiles permitirá a la empresa ajustar dinámicamente los parámetros de segmentación a medida que el mercado evolucione y la organización crezca, brindando flexibilidad para recalibrar el modelo. Esto permite, por ejemplo, actualizar los límites para cada cuantil a medida que cambie el comportamiento de compra o los objetivos comerciales sin tener que reconstruir el modelo desde cero.

Además, se pueden aplicar medidas periódicas de verificación interna. En primer lugar, esto implica analizar el desempeño de cada segmento en términos de ventas y participación, para verificar que los clientes en segmentos clave como “Leales” o “Campeones” realmente estén haciendo la contribución esperada a los ingresos y la frecuencia. En segundo lugar, se pueden realizar encuestas de satisfacción o análisis de comentarios de los clientes en estos segmentos para confirmar que las características de cada grupo coinciden con la experiencia del cliente.

Este proceso de ajuste y revisión le permitirá a la empresa modificar sus políticas de marketing, conforme el negocio crece. El modelo se alinea con los cambios estratégicos y responde a las necesidades cambiantes del mercado.

Estrategias de Marketing Recomendadas

Teniendo en cuenta que el enfoque recomendado a la organización es el basado en cuantiles, en el cual los segmentos se constituyen a partir de reglas definidas por el negocio y que como resultado la generación de 10 segmentos, las estrategias de marketing se centrarán atender los seis grupos de clientes más representativos por dos razones: i) Estos segmentos recogen el 92% de las observaciones y ii) la creación de estrategias, planeación y operatividad necesarias para atender a los restantes 4 grupos, no compensa la potencial generación de ingresos para la organización en una primera iteración del ejercicio.

La tabla 9 consolida todas las estrategias para cada uno de los segmentos generados a partir de sus características y el contexto de la empresa ya que no se trata de un retail común y los productos tienen un periodo de recompra o actualización alto.

Tabla 9 Estrategias de Marketing

Clúster	Estrategias
Clientes con Potencial	<ol style="list-style-type: none"> 1. Implementar estrategias de venta cruzada con productos complementarios para aumentar el valor de compra. 2. Ofrecer beneficios por fidelidad, motivando compras más frecuentes. 3. Diseñar campañas personalizadas enfocadas en las necesidades de los clientes y presentando posibilidades de ahorro.
Campeones	<ol style="list-style-type: none"> 1. Crear categorías de productos premium con descuentos especiales para reforzar su lealtad. 2. Generar programas de recompensas con descuentos exclusivos o acceso anticipado a lanzamientos para reforzar la percepción de exclusividad. 3. Involucrarlos en pruebas de producto o eventos exclusivos.
Nuevos Gastadores	<ol style="list-style-type: none"> 1. Implementar ofertas de bienvenida para aumentar la frecuencia de compra. 2. Campaña de comunicación de los beneficios únicos de la empresa en calidad y servicio. 3. Ofrecer soporte técnico personalizado para consolidar la relación con el cliente.
Sin Actividad	<ol style="list-style-type: none"> 1. Crear campaña de reactivación con descuentos o beneficios llamativos en productos. 2. Enviar correos informativos de uso y actualizaciones sobre nuevos productos. 3. Evaluar estrategia de fidelización temprana y su impacto en reducción de inactividad futura.

Leales	<ol style="list-style-type: none">1. Reforzar la relación con beneficios de largo plazo como descuentos acumulativos según montos de compra.2. Crear una experiencia VIP en el servicio técnico para aumentar la satisfacción.3. Ofrecer charlas de valor agregado personalizadas sobre el mantenimiento y mejoras en el uso de sus equipos.
Gastadores Ocasionales	<ol style="list-style-type: none">1. Implementar campañas de email personalizados de recompra basados en el ciclo de vida de sus equipos.2. Ofrecer descuentos en combos o paquetes (venta cruzada) para incentivar compras más frecuentes.3. Promover el uso de repuestos o consumibles a precios preferenciales.

CAPITULO 5 10. CONCLUSIONES

En este proyecto se abordó la necesidad de desarrollar un modelo basado en datos que permitiera la segmentación de los clientes de una empresa de venta de equipos biomédicos en Colombia, la cual cuenta con información histórica de las transacciones de los clientes pero que había mantenido un enfoque subjetivo en sus estrategias de relacionamiento con el cliente tanto desde el área de marketing como comercial.

Para ello, el proyecto se dividió en dos etapas: Inicialmente se compararon varios modelos de segmentación usados regularmente, identificando sus ventajas y desventajas aplicadas al contexto de la empresa y la industria que se abordó. Cada uno de los modelos estudiados fue evaluado frente a cuatro criterios: Resultados, implementación e interpretación, experiencia de construcción y capacidad, dando como modelo seleccionado el RFM (Recencia, frecuencia y valor monetario).

Una vez seleccionado el modelo RFM, se propuso comparar los dos enfoques utilizados para la generación de los clústeres: Por un lado, está el enfoque tradicional basado en cuantiles definidos por reglas de negocio propias del conocimiento del contexto y, por otro lado, un enfoque basado en el algoritmo k-means que crea los segmentos de clientes por proximidad asegurando homogeneidad entre las observaciones de un mismo grupo y heterogeneidad entre los grupos.

El desarrollo de estos dos enfoques dio como resultado que las reglas de negocio definieron 10 clústeres, de los cuales 6 agrupaban cerca del 91% de las observaciones, mientras

que para el algoritmo k-means se definió un número óptimo de clústeres en 6, de los cuales 3 recogieron el 92% de las observaciones. Cada grupo generado cuenta con la caracterización según el puntaje obtenido para cada variable (recencia, frecuencia y valor monetario), ayudando a entender las características de cada uno y dando una idea más clara de abordarlos según las necesidades del negocio.

La comparación entre los dos enfoques y un análisis basado en facilidad de implementación e interpretación, transparencia y entendimiento de la creación de los segmentos, inclusión de criterios de negocio y facilidad para su posterior personalización en caso de ser necesario, hicieron que la decisión del modelo a recomendar a la organización fuera el primer enfoque, el basado en reglas de negocio o cuantiles.

Con base en esta decisión, se crearon estrategias de marketing para los 6 clústeres representativos que agrupa el 91% de las observaciones y para las cuales, las iniciativas que se definan crearán mayor valor para el negocio. Las estrategias están encaminadas en reconocer los clientes de mayor valor y los que tienen un gran potencial, enfocándose en atender esas necesidades que puedan hacer que los clientes valiosos sigan siéndolo y los que tienen un gran potencial, se pueda explotar según sus necesidades.

Iniciativas de lealtad, venta cruzada y de email marketing de recordación y activación son propuestas para los segmentos que mejor aplican y que tienen mayor potencial de aceptarlas y responder a ellas.

En cuanto a las recomendaciones para la organización, se dividen en tres sentidos: El primero, respecto a crear un proceso de mejora continua que permita la implementación del modelo recomendado, su evaluación y monitoreo para revisar el desempeño y posteriormente un ajuste de los parámetros según necesidades del negocio. Esto permite que el modelo siempre esté vigente y aportando valor al negocio y soporta la decisión de optar por una herramienta que sea fácilmente personalizable a lo largo del tiempo. " Estos mecanismos de puntuación permiten diferenciar los comportamientos de los clientes en las dimensiones del valor RFM, ofreciendo información crítica esencial para la segmentación efectiva y la formulación de estrategias de marketing específicas." (Wong et al., 2024).

En segundo lugar, el enfoque investigativo también se ha centrado en asignar una categoría de segmentación a cada cliente de manera individual a través del notebook Colab que contiene el código de desarrollo del modelo. Por ello, dicho insumo es relevante para que la organización pueda disponer de este, con el fin de aplicar las estrategias de marketing recomendadas de forma personalizada para cada usuario. Este enfoque de segmentación es una estrategia ampliamente utilizada en la industria del retail, ya que facilita afinar las iniciativas de marketing al atender a grupos específicos de clientes, logrando así satisfacer de manera más efectiva sus necesidades y preferencias individuales (Idowu & Kattukottai, 2019).

En tercer lugar, se recomienda unir los insights entregados por el modelo con otros complementarios como el CLV (Customer Lifetime Value) para incluir más variables como la rentabilidad, creando así una visión más robusta del valor de los clientes y generar una estrategia integral.

REFERENCIAS

Ahammed, K., Azad, R. U., Al Jaber, M., Tareq Rony, M. A., & Ahammad, M. (2023). A Comparative Study of RFM-Based Clustering Methods in Customer Segmentation. 2023 26th International Conference on Computer and Information Technology, ICCIT 2023. <https://doi.org/10.1109/ICCIT60459.2023.10441081>

Ahmad M. A. Zamil, & T. G. Vasista. (2021). Customer Segmentation Using RFM Analysis: Realizing Through Python Implementation. www.pbr.co.in

ANDI - Cámaras sectoriales. (s. f.). <https://www.andi.com.co/Home/Camara/8-dispositivos-medicos-e-insumos-para-la-salu>.

Anitha, P., Patil, M. (2022). *RFM model for customer purchase behavior using K-Means algorithm* [Archivo PDF]. <https://doi.org/10.1016/j.jksuci.2019.12.011>

Arthur, Lisa. Big Data Marketing : *Engage Your Customers More Effectively and Drive Value*, John Wiley & Sons, Incorporated, 2013. ProQuest Ebook Central,

<http://ebookcentral.proquest.com/lib/bibliotecaean-ebooks/detail.action?docID=1426518>

Berger, P. D., & Nasr, N. I. (1998). *Marketing Models and Applications. Journal of interactive marketing*.

Bult, J. R., & Wansbeek, T. J. (1995). Optimal selection for direct mail. *Marketing Science*, 14(4), 378-394. <https://doi.org/10.1287/mksc.14.4.378>

Correa Peralta , M. A., Almeida Salazar , B. A., Espinoza Bravo , M. G., Cabezas Reyes , A. T., & Castillo Villegas , K. G. (2024). *Segmentación Efectiva de Clientes utilizando R: Técnicas para Administración y Marketing Avanzado: Effective Customer Segmentation using R: Techniques for Advanced Marketing and Management. Editorial SciELa*, 1(1).

<https://doi.org/10.62131/978-9942-7173-2-0>

Chen, R. (2016). Feedback-Based Eco-Design for Integrating the Recency, Frequency, and Monetary Value of Eco-Efficiency into Sustainability Management. *Systems (Basel)*, 4(3), 30-. <https://doi.org/10.3390/systems4030030>

Doğan, M., & Can, M. (2018). *Customer Segmentation by Using RFM Model and Clustering Methods: A Case Study in Retail Industry. International Journal of Contemporary Economics and Administrative Sciences*, 8(1).

Elveny, M., Syah, R. B. Y., & Nasution, M. K. M. (2024). *An boosting business intelligent to customer lifetime value with robust M-estimation. IAES International Journal of Artificial Intelligence*, 13(2), 1632–1639. <https://doi.org/10.11591/ijai.v13.i2.pp1632-1639>

Evdokimova, S. A. (2021). *Segmentation of store customers to increase sales using ABC-XYZ-analysis and clustering methods*. Journal of Physics: Conference Series, 2032(1), 012117. <https://doi.org/10.1088/1742-6596/2032/1/012117>

Excelx. (n.d.). What is Excel? Meaning, definition and uses. <https://excelx.com/what-is-excel/>

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-54.
<https://doi.org/10.1609/aimag.v17i3.1230>

Feng, Y., Yin, Y., Wang, D., Ignatius, J., Cheng, T. C. E., Marra, M., & Guo, Y. (2024). *Enhancing e-commerce customer churn management with a profit- and AUC-focused prescriptive analytics approach*. *Journal of Business Research*, 184.
<https://doi.org/10.1016/j.ibusres.2024.114872>

Galvez-Torres, E., Cruz-Alfaro, M., Cespedes-Blanco, C., Raymundo, C., Mamani-Macedo, N., & Dominguez, F. (2020). *B2B marketing method adapted to sales improvement through the implementation of ABC classification tool and inbound marketing in SMEs*. En J. I. Kantola, S.

Google Research. (n.d.). Colaboratory frequently asked questions.
<https://research.google.com/colaboratory/faq.html>

Gopal, M. (2019). *OVERVIEW OF BASIC CLUSTERING METHODS*. In *Applied Machine Learning* (1st Edition). McGraw-Hill Education.
<https://www.accessengineeringlibrary.com/content/book/9781260456844/toc-chapter/chapter7/section/section20>

Hernández, R., Fernández, C., & Baptista, P. (2010). Metodología de la investigación. 6.

Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2), 85-126. <https://doi.org/10.1023/B:AIRE.0000045509.59082.9a>

Idowu, S., & Kattukottai, S. (2019). *Customer segmentation based on RFM model using kmeans, hierarchical and fuzzy c-means clustering algorithms*. ResearchGate. <https://doi.org/10.13140/RG.2.2.15379.71201>

Kasem, M., Hamada, M., Taj-Eddin, I. (2024). *Customer profiling, segmentation, and sales prediction using AI in direct marketing*. <https://doi.org/10.1007/s00521-023-09339-6>

Kopřivová, V., & Matušinská, K. (2023). UNLOCKING GENERATION Y: MARKET SEGMENTATION VIA LIFESTYLE INSIGHTS. *Communication Today*, 14(2), 122–139. <https://doi.org/https://doi.org/10.34135/communicationtoday.2023.Vol.14.No.2.9>

Kumar, V., & Reinartz, W. (2018). *Customer Relationship Management : Concept, Strategy, and Tools* (3rd ed. 2018.). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-55381-7>

Liu, A., McLellan, K., & Reinholt, B. (2015). *How medical-device manufacturers can transform marketing and sales capabilities*. McKinsey & Company. <https://www.mckinsey.com/industries/life-sciences/our-insights/how-medical-device-manufacturers-can-transform-marketing-and-sales-capabilities>

López, Á. J. C., Caicedo, C. G., & Oviedo, P. C. J. (2017). *Análisis multivariado para segmentación de clientes basada en RFM*. *Tecnura*, 21(54), 41-51. <https://doi.org/10.14483/22487638.12957>

Montero, M., Suarez, D., Aguirre, M. I. P., & Aguirre, R. E. C. (2020). *Estimating customer lifetime value: A metric alternative for customer segmentation applied to the retail sector. Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology*. <https://doi.org/10.18687/LACCEI2020.1.1.327>

Nair, A. (2023). *RFM analysis for successful customer segmentation*. Putler. <https://www.putler.com/rfm-analysis/>

Natingga, D. (2018). *Data science algorithms in a week : top 7 algorithms for scientific computing, data analysis, and machine learning* (Second edition, p. 107). Packt.

Nazir, & V. Salminen (Eds.), *Advances in Human Factors, Business Management and Leadership*. AHFE 2020. *Advances in Intelligent Systems and Computing* (Vol. 1209, pp. 167-173). Springer. https://doi.org/10.1007/978-3-030-50791-6_21

Nie, D., Scriney, M., Liang, X., & Roantree, M. (2024). *From data acquisition to validation: a complete workflow for predicting individual customer lifetime value. Journal of Marketing Analytics*, 12(2), 321–341. <https://doi.org/10.1057/s41270-022-00197-0>

Python Software Foundation. (2022, noviembre 22). *BeginnersGuide/Overview*. Python Wiki. <https://wiki.python.org/moin/BeginnersGuide/Overview>

Rendon, E., Abundez, I. (2016). *RENTOL: Un algoritmo de agrupamiento basado en K-means* [Archivo PDF]. https://rcs.cic.ipn.mx/2016_128/RENTOL_%20Un%20algoritmo%20de%20agrupamiento%20basado%20en%20K-means.pdf

Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (1998). *Density-based clustering in spatial databases: A summary*. In *Advances in Database Technology - EDBT 1998* (pp. 206-222).

Springer. https://doi.org/10.1007/3-540-49140-4_15

Sembiring Berahmana, R. W. B., Mohammed, F. A., & Chairuang, K. (2020). *Customer segmentation based on RFM model using K-Means, K-Medoids, and DBSCAN methods*.

Lontar Komputer, 11(1), 32-43. <https://doi.org/10.24843/LKJITI.2020.v11.i01.p04>

Sun, Z.-H., Zuo, T.-Y., Liang, D., Ming, X.-G., Chen, Z.-H., Qiu, S.-Q. (2021). *GPHC: A heuristic clustering method to customer segmentation* [Archivo PDF].

<https://doi.org/10.1016/j.asoc.2021.107677>

Wang, Y., & Wang, H. (2018). *A survey on density-based clustering algorithms*. *Journal of Computer Science and Technology*, 33(3), 453-469. [https://doi.org/10.1007/s11390-018-](https://doi.org/10.1007/s11390-018-1832-5)

[1832-5](https://doi.org/10.1007/s11390-018-1832-5)

Wong et al., (2024). *Exploring Customer Segmentation in E-Commerce using RFM Analysis with Clustering Techniques*. *Journal of Telecommunications and the Digital Economy*, 12, 105. [https://web-p-ebSCOhost-](https://web-p-ebSCOhost-com.bdbiblioteca.universidadean.edu.co/ehost/pdfviewer/pdfviewer?vid=3&sid=deeebf80-471f-47eb-8b57-8255d9132180%40redis)

[com.bdbiblioteca.universidadean.edu.co/ehost/pdfviewer/pdfviewer?vid=3&sid=deeebf80-](https://web-p-ebSCOhost-com.bdbiblioteca.universidadean.edu.co/ehost/pdfviewer/pdfviewer?vid=3&sid=deeebf80-471f-47eb-8b57-8255d9132180%40redis)

[471f-47eb-8b57-8255d9132180%40redis](https://web-p-ebSCOhost-com.bdbiblioteca.universidadean.edu.co/ehost/pdfviewer/pdfviewer?vid=3&sid=deeebf80-471f-47eb-8b57-8255d9132180%40redis)

Yang, K., & Trewn, J. (2004). *Cluster Analysis*. In *Multivariate Statistical Methods in Quality Management* (1st Edition). McGraw-Hill Education.

<https://www.accessengineeringlibrary.com/content/book/9780071432085/chapter/chapter7>