



**UNIVERSIDAD TECNOLÓGICA
INDOAMÉRICA**

FACULTAD DE INGENIERÍAS

MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

TEMA:

SEGMENTACIÓN AVANZADA DE CLIENTES EN LA INDUSTRIA DE SUPLEMENTOS NATURALES PARA MEJORAR LAS ESTRATEGIAS DE MARKETING MEDIANTE HERRAMIENTAS DE BIG DATA.

Trabajo de Titulación previo a la obtención del título de Magíster en Big Data y Ciencia de Datos.

Autor(a)

Ing. Jorge Enrique Paz Ojeda

Tutor(a)

Ing. Ivanna Sánchez. MBA

AMBATO – ECUADOR

2025

**AUTORIZACIÓN POR PARTE DEL AUTOR PARA LA CONSULTA,
REPRODUCCIÓN PARCIAL O TOTAL, Y PUBLICACIÓN ELECTRÓNICA
DEL TRABAJO DE TITULACIÓN**

Yo, Jorge Enrique Paz Ojeda, declaro ser autor del Trabajo de Titulación con el nombre “SEGMENTACIÓN AVANZADA DE CLIENTES EN LA INDUSTRIA DE SUPLEMENTOS NATURALES PARA MEJORAR LAS ESTRATEGIAS DE MARKETING MEDIANTE HERRAMIENTAS DE BIG DATA”, como requisito para optar al grado de Magister en Big Data y Ciencia de Datos autorizo al Sistema de Bibliotecas de la Universidad Indoamérica, para que con fines netamente académicos divulgue esta obra a través del Repositorio Digital Institucional (RDI-UTI).

Los usuarios del RDI-UTI podrán consultar el contenido de este trabajo en las redes de información del país y del exterior, con las cuales la Universidad tenga convenios. La Universidad Indoamérica no se hace responsable por el plagio o copia del contenido parcial o total de este trabajo.

Del mismo modo, acepto que los Derechos de Autor, Morales y Patrimoniales, sobre esta obra, serán compartidos entre mi persona y la Universidad Indoamérica, y que no tramitaré la publicación de esta obra en ningún otro medio, sin autorización expresa de la misma. En caso de que exista el potencial de generación de beneficios económicos o patentes, producto de este trabajo, acepto que se deberán firmar convenios específicos adicionales, donde se acuerden los términos de adjudicación de dichos beneficios.

Para constancia de esta autorización, en la ciudad de Ambato a los 31 días del mes de Julio de 2025, firmo conforme:

Autor: Jorge Enrique Paz Ojeda

Firma:

Número de Cédula: 1711443711

Dirección: Pichincha, Quito, Pomasqui, Urb. La Pampa De Las Nueces Lote 78.

Correo Electrónico: jpaz8@indoamerica.edu.ec

Teléfono: 0997016845

APROBACIÓN DEL DIRECTOR

En mi calidad de Director del Trabajo de Titulación “**SEGMENTACIÓN AVANZADA DE CLIENTES EN LA INDUSTRIA DE SUPLEMENTOS NATURALES PARA MEJORAR LAS ESTRATEGIAS DE MARKETING MEDIANTE HERRAMIENTAS DE BIG DATA.**” presentado por Jorge Enrique Paz Ojeda, para optar por el Título de Magíster en Big Data y Ciencia de Datos.

CERTIFICO

Que dicho Trabajo de Titulación ha sido revisado en todas sus partes y considero que reúne los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte los Examinadores que se designe.

Ambato, 29 de Julio del 2025

.....

Ing. Ivanna Sánchez. MBA

DIRECTOR

DECLARACIÓN DE AUTENTICIDAD

Quien suscribe, declaro que los contenidos y los resultados obtenidos en el presente Trabajo de Titulación, como requerimiento previo para la obtención del Título de Magíster en Big Data y Ciencia de Datos, son absolutamente originales, auténticos y personales y de exclusiva responsabilidad legal y académica del autor.

Ambato, 31 de Julio del 2025

.....
Jorge Enrique Paz Ojeda

1711443711

APROBACIÓN DE EXAMINADORES

El Trabajo de Titulación ha sido revisado, aprobado y autorizada su impresión y empastado, sobre el Tema: **SEGMENTACIÓN AVANZADA DE CLIENTES EN LA INDUSTRIA DE SUPLEMENTOS NATURALES PARA MEJORAR LAS ESTRATEGIAS DE MARKETING MEDIANTE HERRAMIENTAS DE BIG DATA**, previo a la obtención del Título de Magíster en Big Data y Ciencia de Datos, reúne los requisitos de fondo y forma para que el estudiante pueda presentarse a la sustentación del Trabajo de Titulación.

Ambato, 31 de Julio del 2025

.....

Dr. José Varela Aldás, Ph.D.

EXAMINADOR

.....

Ing. Ignacio Ayala Chauvin, Ph.D.

EXAMINADOR

DEDICATORIA

Dedico este trabajo a mi familia, cuyo acompañamiento constante ha sido fundamental en mi formación personal y profesional, y especialmente a mis hijas, con el anhelo de dejarles un legado de disciplina y superación. Que este esfuerzo sea para ellas una muestra de que, con constancia y compromiso, todo objetivo es alcanzable.

A mis padres, por inculcarme con su ejemplo los valores del esfuerzo, la responsabilidad y la perseverancia, pilares esenciales en la consecución de este logro.

•

AGRADECIMIENTO

Agradezco a Dios y a mi familia por su apoyo en esta meta, a mis profesores y tutores de la Universidad Tecnológica Indoamérica por compartir sus conocimientos y a Natures Sunshine Latam por su apertura para esta investigación.

ÍNDICE DE CONTENIDOS

PORTADA	i
AUTORIZACIÓN PARA EL REPOSITORIO DIGITAL	ii
APROBACIÓN DEL DIRECTOR	iii
DECLARACION DE AUTENTICIDAD	iv
APROBACION DE EXAMINADORES	v
DEDICATORIA	vi
AGRADECIMIENTO	vii
INDICE DE CONTENIDOS	viii
INDICE DE TABLAS	x
INDICE DE GRÁFICOS	xi
INDICE DE IMÁGENES.....	xii
RESUMEN EJECUTIVO	xiii
ABSTRACT	xiv

CAPÍTULO I

INTRODUCCIÓN.....	1
ANTECEDENTES	2
JUSTIFICACIÓN.....	4
OBJETIVO GENERAL	4

CAPÍTULO II

METODOLOGÍA.....	5
ÁREA DE ESTUDIO	5
ENFOQUE METODOLÓGICO	5
DESCRIPCIÓN DE LA METODOLOGÍA.....	5
DISEÑO DEL TRABAJO.....	5
PROCEDIMIENTO PARA OBTENCION Y ANÁLISIS DE DATOS	6
POBLACIÓN Y MUESTRA	6

CAPÍTULO III

DESARROLLO.....	10
RECOPIACIÓN DE LOS DATOS.....	10

PROCESAMIENTO DE LOS DATOS	11
MODELADO	13
CAPÍTULO IV	
RESULTADOS Y DISCUSIÓN	14
CAPÍTULO V	
CONCLUSIONES Y RECOMENDACIONES	25
REFERENCIAS BIBLIOGRAFICAS	26
ANEXOS	28

ÍNDICE DE TABLAS

Tabla No. 1: DataSet 1	17
Tabla No. 2: DataSet 2	18
Tabla No. 3: Tabla de medias de los resultados del agrupamiento (K-means)	28
Tabla No. 4: Tabla de características de los segmentos de clientes	28

ÍNDICE DE GRÁFICOS

Gráfico No. 1: Método Elbow para K-means.....	21
Gráfico No. 2: Silhouette Score para K-means	22
Gráfico No. 3: Método Elbow para ϵ	24
Gráfico No. 4: Dendrograma AHC	25
Gráfico No. 5: Silhouette Score para AHC).....	25
Gráfico No. 6: Comportamiento de las variables Monetary vs Frecuency	27
Gráfico No. 7: Comportamiento de las variables Monetary vs Recency.....	27

ÍNDICE DE IMÁGENES

Imagen No. 1: (Metodología aplicada).....	14
Imagen No. 2: Función que elimina valores nulos, Phyton.....	19
Imagen No. 3: Función que elimina valores duplicados, Phyton.	19
Imagen No. 4: Función que elimina columnas no requeridas, Phyton.....	19
Imagen No. 5: Escalamiento de datos, Phyton	19
Imagen No. 6: Función K-Means, Phyton.	22
Imagen No. 7: Función NearestNeighbors, Phyton.....	23
Imagen No. 8: Aplicación de DBSCAN y silhouette-score, Phyton.....	23
Imagen No. 9: Aplicación Agglomerative Clustering y silhouette_score, Phyton.....	26
Imagen No. 10: Resultados Silhouette Score para los métodos aplicados, Phyton.....	26

UNIVERSIDAD TECNOLÓGICA INDOAMÉRICA
FACULTAD DE INGENIERÍAS
MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

TEMA: SEGMENTACIÓN AVANZADA DE CLIENTES EN LA INDUSTRIA DE SUPLEMENTOS NATURALES PARA MEJORAR LAS ESTRATEGIAS DE MARKETING MEDIANTE HERRAMIENTAS DE BIG DATA

AUTOR(A): Ing. Jorge Enrique Paz Ojeda

TUTOR (A): Ing. Ivanna Sanchez, MBA.

RESUMEN EJECUTIVO

La compañía Nature's Sunshine Products, dedicada a la comercialización de suplementos naturales. Sus sucursales de Latinoamérica (NSP Latam) cada vez enfrentan mayores dificultades para alcanzar sus metas comerciales una de las causas es la falta de aplicación de herramientas tecnológicas que le permitan comprender a profundidad el comportamiento de sus clientes. Esta situación ha limitado el impacto de sus estrategias de marketing, afectando tanto la fidelización como el aprovechamiento de oportunidades de venta. Ante este problema, el objetivo del estudio fue aplicar una metodología de segmentación avanzada basada en el análisis RFM (Recencia, Frecuencia y Valor Monetario) y técnicas de agrupamiento como K-means, DBSCAN y clustering jerárquico, con el fin de identificar patrones de comportamiento que orienten decisiones comerciales más efectivas. La hipótesis planteada sostiene que, mediante el uso de algoritmos de segmentación y análisis de datos, es posible clasificar a los clientes en grupos con características similares, lo que permitirá implementar estrategias personalizadas y mejorar los resultados del negocio. Para esto, se trabajó con una base de datos de más de 79.000 registros de ventas, de los últimos dos años, la cual fue procesada mediante con herramientas de Big Data en entornos como Python y Google Colab. Los resultados demostraron que el algoritmo K-means fue el más efectivo, permitiendo identificar cuatro segmentos claros: Platinum, Gold, Silver y Bronze, cada uno con niveles distintos de compra, compra más reciente y frecuencia de compra. Esta clasificación ofrece a la empresa una base sólida para optimizar sus campañas, enfocar sus recursos y fortalecer la relación con sus clientes. En conclusión, la integración de tecnologías de análisis de datos en esta industria no solo mejora la comprensión del mercado, sino que también impulsa la competitividad y sostenibilidad del negocio en el tiempo.

DESCRIPTORES: BigData, Clustering, Estrategias Marketing, Segmentación de Clientes

UNIVERSIDAD TECNOLÓGICA INDOAMÉRICA

FACULTY OF ENGINEERING

MASTER'S IN BIG DATA AND DATA SCIENCE

AUTHOR: PAZ OJEDA JORGE ENRIQUE

TUTOR: MSc. SANCHEZ MONTERO IVANNA

ABSTRACT

ADVANCED CUSTOMER SEGMENTATION IN THE NATURAL SUPPLEMENTS INDUSTRY TO IMPROVE MARKETING STRATEGIES USING BIG DATA TOOLS

Nature's Sunshine Products, a company dedicated to the commercialization of natural supplements, is facing increasing challenges in its Latin American branches (NSP Latam) in achieving their commercial goals. One of the main causes is the lack of implementation of technological tools that enable a deep understanding of customer behavior. This situation has limited the impact of their marketing strategies, affecting both customer loyalty and the capitalization of sales opportunities. In response to this issue, the objective of the study was to apply an advanced segmentation methodology based on RFM analysis (Recency, Frequency, and Monetary value) combined with clustering techniques such as K-means, DBSCAN, and hierarchical clustering, with the aim of identifying behavioral patterns that can guide more effective commercial decisions. The proposed hypothesis is that by using segmentation algorithms and data analysis, it is possible to classify customers into groups with similar characteristics, thereby enabling the implementation of personalized strategies and improving business outcomes. To achieve this, a database of more than 79,000 sales records from the past two years was processed using Big Data tools in environments such as Python and Google Colab. The results showed that the K-means algorithm was the most effective, allowing the company to identify four distinct segments: Platinum, Gold, Silver, and Bronze, each differing in purchase volume, recency, and frequency. This classification provides the company with a solid foundation to optimize marketing campaigns, allocate resources more efficiently, and strengthen customer relationships. In conclusion, the integration of data

KEYWORDS: Keywords: bigdata, clustering, customer segmentation, marketing strategies.



CAPITULO I

INTRODUCCIÓN

Contexto macro:

En la era de la Industria 4.0, la digitalización de los procesos ha evolucionado hacia la utilización de Big Data (BD) e Inteligencia Artificial (IA) como herramientas esenciales para mejorar el rendimiento empresarial, especialmente en áreas como el marketing, la gestión de clientes y la toma de decisiones estratégicas (Shirazi & Mohammadi, 2019) (Gupta et al., 2021). Estas tecnologías permiten analizar grandes volúmenes de datos estructurados y no estructurados para generar conocimiento accionable y personalizado.

Contexto meso:

En América Latina, la industria de suplementos naturales ha crecido con rapidez, impulsada por el aumento del interés en productos saludables. No obstante, la mayoría de empresas del sector aún no han incorporado tecnologías de análisis avanzado para segmentar a sus clientes. Casos prácticos en el comercio electrónico y bienes de consumo han demostrado que una correcta implementación de Big Data puede transformar el enfoque comercial, optimizando la cadena de valor y la experiencia del cliente (Elía et al., 2020).

Contexto micro:

Nature's Sunshine Latam posee una base de datos con más de un millón de registros de clientes y transacciones. Sin embargo, la falta de aplicación de herramientas de análisis de datos ha limitado la efectividad de sus estrategias de marketing, impidiendo identificar segmentos clave de clientes y formular campañas personalizadas. Esta situación representa una oportunidad estratégica para integrar modelos como RFM y algoritmos de clustering que, según Mensouri et al. (2022), permiten clasificar de forma efectiva a los clientes y mejorar el retorno de la inversión comercial.

Problematización

A pesar del volumen y riqueza de los datos disponibles, Nature's Sunshine Latam no cuenta con un sistema de segmentación que traduzca esa información en conocimiento estratégico. El modelo actual de marketing es homogéneo, dirigido sin distinción a toda la base de clientes, lo que genera baja fidelización y un desperdicio de recursos. Además, la ausencia de una cultura basada en datos dificulta el diseño de estrategias personalizadas. Esta problemática ha sido evidenciada en diversos estudios, donde se plantea que la segmentación basada en variables RFM permite generar agrupamientos homogéneos y accionables (Antonius & Devi Fitriannah, 2024) (Zerbino et al., 2018).

Estudios recientes demuestran que el uso de modelos como RFM combinado con algoritmos de clustering permite generar segmentaciones más precisas y útiles para la toma de decisiones estratégicas (Mensouri et al., 2022). Sin embargo, Nature's Sunshine Latam continúa empleando métodos de marketing tradicionales que no responden a la heterogeneidad de su mercado. Esta situación se agrava por factores como la baja integración tecnológica, inconsistencias en la calidad de los datos y la resistencia interna a la adopción de herramientas de Big Data (Mukhopadhyay et al., 2024).

La falta de una estrategia basada en datos limita la eficacia de las campañas comerciales y reduce el retorno sobre la inversión en acciones de captación y fidelización. Por tanto, se requiere con urgencia la implementación de un sistema de segmentación avanzada que permita categorizar a los clientes en grupos significativos y formular estrategias diferenciadas para cada uno, alineadas con los objetivos de negocio.

ANTECEDENTES

La literatura respalda ampliamente el uso de modelos de segmentación basados en Big Data. Hughes (1994) introdujo el modelo RFM como herramienta para analizar el valor de los clientes en marketing relacional. Estudios más recientes como el de Aditya y Fitriannah (2021) comparan algoritmos de agrupamiento para encontrar el más adecuado en la identificación de clientes potenciales. Zerbino et al.(2018) desarrollaron un enfoque integral de gestión basada en datos para mejorar la personalización y fidelización de los consumidores.

Mensouri et al. (2022) propusieron una combinación de RFM y K-means que mostró efectividad en contextos reales de negocio. Asimismo, Antonius y Fitriyah (2024) integraron el modelo RFM con proporciones de descuento para enriquecer la segmentación. Mukhopadhyay et al. (2024) presentan un marco de Marketing 4.0 habilitado por Big Data que permite una mayor precisión en la segmentación. Elia et al. (2020) resaltan el valor que la creación multidimensional basada en datos genera para las empresas en industrias intensivas en información. Finalmente, Erevelles et al. (2016) señalan que la analítica de consumidores transforma la forma en que las empresas diseñan sus estrategias comerciales.

Diversas investigaciones han abordado la segmentación avanzada de clientes como una estrategia clave para optimizar la relación empresa-cliente y mejorar la eficiencia del marketing, especialmente en contextos apoyados por Big Data. En el sector retail, Maguiña Figueroa et al. (2018) demostraron que la aplicación de técnicas como el clustering y el análisis de componentes principales (PCA) permite identificar grupos de consumidores con mayor precisión, lo que resultó en una mejora significativa en la efectividad de las campañas promocionales.

En el mercado de suplementos orgánicos, Tarazona y Villanueva (2021) identificaron patrones de comportamiento en consumidores con base en valores y estilos de vida, permitiendo redefinir estrategias de posicionamiento de marca más efectivas. Este enfoque fue complementado por Colqui Palacin y Guevara (2023), quienes utilizaron análisis de clustering para identificar segmentos clave en el lanzamiento del producto MoradOx, proyectando un impacto positivo en las ventas. Desde una perspectiva más teórica, Kartajaya, Kotler y Setiawan (2018), en su obra *Marketing 4.0*, explican cómo las tecnologías digitales como Big Data han transformado la manera de atraer y fidelizar clientes, permitiendo una personalización más eficaz en entornos cada vez más conectados.

Por su parte, Palomo Fernández (2024) profundizó en el uso de algoritmos de inteligencia artificial y análisis de datos masivos para la identificación de microsegmentos en la industria del bienestar, subrayando la importancia de estos enfoques para maximizar el retorno sobre la inversión en campañas digitales. En el ámbito asegurador, Aguirre (2022) desarrolló una segmentación basada en clustering para personalizar los servicios

ofrecidos por las aseguradoras, mejorando la fidelización y la anticipación del comportamiento del cliente.

JUSTIFICACIÓN

Este proyecto es relevante tanto en el ámbito corporativo como en el académico. Para Nature's Sunshine Latam, la implementación de un sistema de segmentación avanzada basado en Big Data permitirá identificar patrones de consumo, personalizar estrategias de marketing y mejorar la rentabilidad. Los principales beneficiarios serán los departamentos de marketing, ventas y alta dirección, quienes contarán con herramientas analíticas para optimizar su toma de decisiones. A nivel teórico, el estudio amplía el conocimiento sobre aplicación de clustering y arquitecturas modulares en sistemas de analítica avanzada. La utilidad práctica radica en la posibilidad de replicar este modelo en otras industrias, y en la generación de valor mediante una mejor comprensión del comportamiento del consumidor. La factibilidad del proyecto está garantizada por la disponibilidad de datos reales, el compromiso institucional y la experiencia del investigador en ciencia de datos.

Objetivo general. - Implementar un modelo de segmentación avanzada de clientes mediante el análisis RFM y algoritmos de clustering basados en Big Data, con el propósito de optimizar las estrategias de marketing de Nature's Sunshine Latam y mejorar la efectividad comercial en su mercado regional.

Objetivos específicos

1. Diseñar un sistema de procesamiento y depuración de datos transaccionales que permita estructurar una base confiable para el análisis de clientes.
2. Aplicar técnicas de segmentación basadas en el modelo RFM y algoritmos de clustering (K-means, DBSCAN, AHC) e identificar el mejor de estos para este tipo de compañías.
3. Formular estrategias de marketing diferenciadas para cada segmento identificado, que incluyan responsables directos, presupuesto tentativo y duración en el tiempo de acuerdo con el comportamiento identificado de los clientes.

CAPITULO II

METODOLOGÍA

Área de estudio

El presente trabajo se ubica en el área de Ingeniería en Tecnologías de la Información, según la clasificación establecida por la Universidad Tecnológica Indoamérica y se enmarca en el área de Big Data y Ciencia de Datos, con énfasis en el análisis de datos masivos aplicados a problemas empresariales, específicamente la segmentación de clientes basada en técnicas de minería de datos y aprendizaje automático no supervisado.

Enfoque metodológico

El enfoque metodológico adoptado es mixto. Por un lado, el enfoque cuantitativo permite aplicar modelos matemáticos y estadísticos, tales como el análisis RFM y algoritmos de clustering, para categorizar a los clientes con base en datos transaccionales históricos. Esta aproximación facilita establecer relaciones entre variables numéricas y realizar inferencias válidas. Por otro lado, el enfoque cualitativo permite interpretar los resultados obtenidos, comprender el comportamiento de los segmentos y formular estrategias de marketing contextualizadas a las características observadas.

Descripción de la metodología:

El tipo de investigación es explicativo y asociativo, ya que busca entender cómo las variables Recency, Frequency y Monetary están relacionadas con la probabilidad de que un cliente pertenezca a un determinado segmento. El método de investigación es mixto. Se recurrió a la investigación bibliográfica-documental para la construcción del marco teórico y la revisión de antecedentes. Asimismo, se aplicó un enfoque de campo, mediante el uso de datos reales proporcionados por Nature's Sunshine Latam. Finalmente, se incorporó un componente experimental, utilizando modelos de segmentación en entornos de desarrollo de datos (Python, Google Colab).

Diseño del trabajo:

El diseño incluye la conceptualización y evaluación de las variables RFM, que se definieron como: Recency (cantidad de días desde la última compra), Frequency (número total de compras realizadas) y Monetary (valor total gastado por el cliente). Estas

variables fueron transformadas y escaladas para garantizar su compatibilidad con algoritmos de clustering. Se diseñó un experimento de comparación entre tres algoritmos de clustering: K-means, DBSCAN y AHC, evaluados con métricas como el Silhouette Score, seleccionando el modelo más eficiente para representar a los clientes.

Procedimiento para obtención y análisis de datos:

La obtención de datos se realizó a partir del sistema de gestión de ventas SQL Server de Nature's Sunshine Latam, seleccionando registros del periodo 2022-2024. Los datos fueron depurados y estructurados utilizando herramientas de Python, con librerías como Pandas y NumPy para la manipulación de datos, y Scikit-learn para el modelado. Se utilizó StandardScaler para la normalización, y los resultados fueron analizados con técnicas de visualización, evaluación estadística de métricas de clústeres y síntesis de perfiles. Posteriormente, se propusieron estrategias por segmento basadas en los hallazgos.

Población y muestra:

La población corresponde a todos los clientes registrados por Nature's Sunshine Latam durante el periodo de análisis. Esta población alcanza un total de 1.138.000 registros. La muestra estuvo compuesta por 79.916 clientes cuyos datos eran completos, limpios y aptos para el cálculo del modelo RFM. El muestreo fue no probabilístico por conveniencia, al trabajar con los datos disponibles en condiciones reales y útiles para el modelado. Esta muestra fue considerada suficiente para garantizar validez analítica y representatividad en la segmentación.

En la Imagen No. 1 se muestra el flujo de la metodología CRISP-DM que se empleó en este estudio:

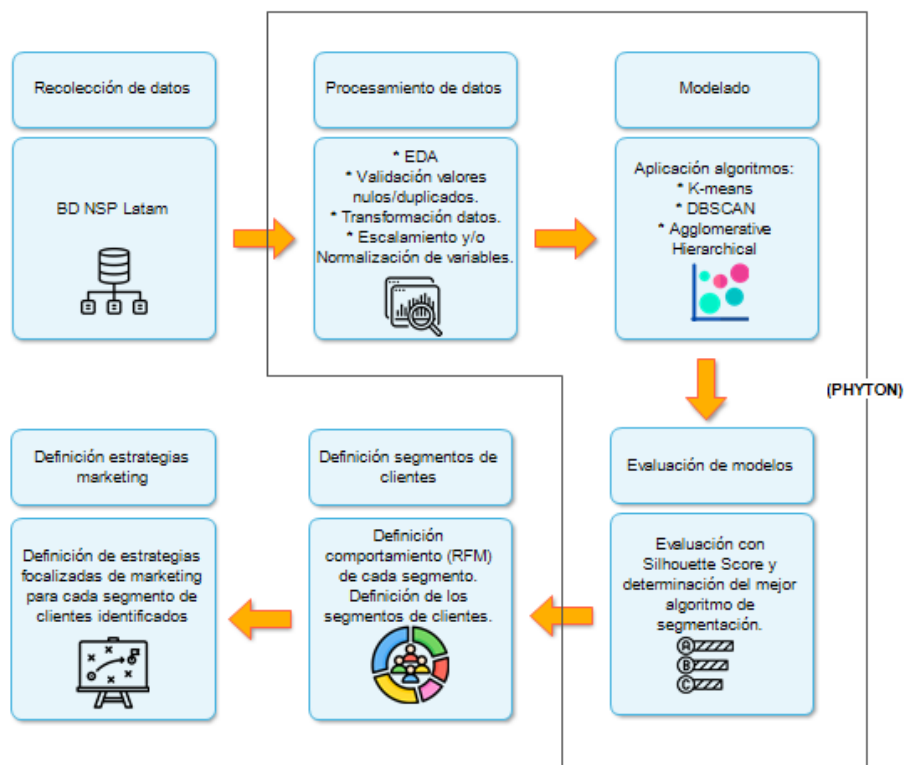


Imagen No. 1: Metodología Aplicada

Elaborado por: Paz, Jorge (2025).

Esta metodología es muy reconocida en proyectos de análisis de datos por su enfoque práctico e iterativo, esta se implementó en 6 etapas:

Generalmente esta metodología tiene una etapa inicial de entendimiento del negocio, sin embargo, al ser el investigador colaborador de la compañía con varios años de experiencia en la industria prescindió de esta etapa. La metodología para este estudio se definió con las siguientes etapas:

Etapa 1: Recopilación de datos

En esta fase se identificaron y extrajeron las fuentes de información relevantes, tanto internas como externas. Se recopiló un conjunto de datos histórico compuesto por variables demográficas, conductuales y transaccionales de los clientes de la empresa. La calidad y representatividad de los datos fueron criterios clave para asegurar la validez del análisis posterior.

Etapa 2: Procesamiento de datos

Esta etapa incluyó tareas de limpieza, transformación y estandarización de los datos. Se eliminaron registros duplicados, se gestionaron los valores faltantes y se normalizaron las variables numéricas para facilitar el análisis estadístico. Asimismo, se aplicaron técnicas de codificación para convertir variables categóricas en un formato adecuado para el modelado. Esta fase fue fundamental para asegurar la integridad y consistencia de los datos.

Etapa 3: Modelado (aplicación de los algoritmos de clustering)

Una vez preparados los datos, se procedió a la construcción de los modelos de segmentación mediante algoritmos de clustering. Se evaluaron diferentes métodos, como K-means y DBSCAN, seleccionando aquel que mejor se adaptó a las características de los datos y a los objetivos del negocio. El propósito de esta etapa fue identificar grupos homogéneos de clientes en función de sus patrones de comportamiento y características comunes.

Etapa 4: Evaluación de los modelos

Durante esta fase, se validaron los resultados obtenidos a través de métricas estadísticas y visualizaciones gráficas. Se analizaron indicadores como la inercia, la silueta y la cohesión interna de los clústeres para determinar la calidad del modelo. Además, se verificó la interpretabilidad de los segmentos generados y su utilidad práctica para la toma de decisiones estratégicas en la empresa.

Etapa 5: Definición de los segmentos

Con base en los resultados del modelado, se describieron detalladamente los perfiles de cada segmento identificado. Esta descripción incluyó variables como edad, frecuencia de compra, ticket promedio, tipo de producto adquirido, entre otros. Se establecieron nombres y características clave para facilitar la comprensión y comunicación de los segmentos a los equipos de marketing, ventas y desarrollo de producto.

Etapa 6: Definición de las estrategias de marketing

Finalmente, se diseñaron estrategias diferenciadas de marketing dirigidas a cada segmento, considerando sus necesidades, preferencias y valor potencial. Se plantearon acciones específicas en términos de comunicación, promociones, canales de distribución y fidelización, con el objetivo de optimizar los recursos de la empresa y maximizar el

retorno de inversión. Esta etapa permitió traducir los hallazgos del análisis de datos en decisiones estratégicas concretas orientadas al crecimiento del negocio.

Dado el alto requerimiento de procesamiento y manejo de grandes volúmenes de datos, se utilizó herramientas de computación en la nube, seleccionando Google Colab Pro, dada su integración con bibliotecas de Python como scikit-learn, pandas, matplotlib y PySpark. Esto no solo soporta las necesidades computacionales del proyecto, sino que también permite el uso de GPU y TPU para acelerar el procesamiento y reducir el tiempo de ejecución. Google Drive facilitará el almacenamiento y acceso a los datos, promoviendo un flujo de trabajo eficiente y la posibilidad de compartir resultados con colaboradores.

CAPITULO III DESARROLLO

Recopilación de los datos

En el presente estudio se utilizaron datos proporcionados por NSP Latam, compañía norteamericana de suplementos naturales con más de 50 años de presencia en los mercados, estos corresponden a su cartera de clientes y registros de ventas en sus mercados en Latinoamérica y corresponden a registros desde septiembre 2022 hasta agosto 2024 obtenidos directamente de sus bases de datos en SQL Server mediante Microsoft SQL Management Studio 18. Este primer conjunto de datos incluía cerca de 1138000 observaciones, en la Tabla No. 1 se muestra la estructura del DataSet 1, resultado de la recopilación inicial de datos.

Tabla No. 1: DataSet 1.

<i>Atributo</i>	<i>Descripción</i>
OrderID	ID de orden
Total_orderCost	Monto compra (USD)
OrderDate	Fecha de compra
OCountry	País de compra
AccountNumber	ID cliente
ACountry	País cliente
ChannelType	Canal venta
CenterID	ID Tienda/Centro
ItemID	ID Producto
Quantity	Cantidad

Elaborado por: Paz, Jorge (2025).

Los autores realizaron un análisis exploratorio para entender la calidad, distribución y correlaciones entre las variables. Adicionalmente los autores aplicaron el análisis RFM (Recency, Frecuency, Monetary), este fue desarrollado por primera vez por A.M. Hughes (A. M. Hughes, 1994), cómo un método de análisis del valor del cliente. De acuerdo con

Aditya D (Aditya & Fitriannah, 2021), ayuda a determinar los atributos utilizados para la segmentación de clientes. RFM hace referencia a sus siglas en inglés:

- Recency (R): representa el valor del período transcurrido entre la última compra de un cliente y el momento actual. Un valor de recencia más bajo implica que el cliente compra con frecuencia. De manera similar, un valor alto implica que el cliente no realizará una compra en el corto plazo.
- Frequency (F): representa el número de compras realizadas por el cliente. Es igual al número total de compras. Un valor alto de frecuencia implica un alto nivel de fidelidad del cliente.
- Monetary (M): que corresponde a la suma de las compras realizadas por cada cliente en el periodo analizado (USD).

Mediante el análisis RFM se pretende comprender el comportamiento y el valor de los clientes en función de sus características (Antonius & Devi Fitriannah, 2024). Con estas consideraciones se generó un nuevo dataset (dataset2) con 79916 observaciones, en la Tabla No. 2 se muestra la estructura del DataSet 2, resultado de la implementación de la implementación del análisis RFM.

Tabla No. 2: DataSet 2.

<i>Atributo</i>	<i>Descripción</i>
Country	País orden
Customer	ID Cliente
Recency	Días desde últ. compra
Frequency	Núm. compras
Monetary	Valor compra (USD)
Gender	Género Cliente
CanalMasUtilizado	Canal de compra

Elaborado por: Paz, Jorge (2025).

Procesamiento de los datos

Para la limpieza de datos, se eliminó valores duplicados y valores nulos utilizando la librería Pandas de Python, las funciones *dropna()* y *duplicated()* posteriormente se

eliminó las columnas que no serán consideradas para este análisis: Customer, Gender, Country y CanalMasUtilizado, mediante la función `drop(columns=[])`, en la Imagen No. 2, 3 y 4 se muestra el código Python utilizado para eliminar: valores nulos, valores duplicados y columnas no requeridas para este análisis respectivamente.

```
[6] dfx = df
    df = df.dropna()
    dfx = dfx.dropna()
    df.isnull().sum()
```

Imagen No. 2: Función que elimina valores nulos, Python.

Elaborado por: Paz, Jorge (2025).

```
[7] df[df.duplicated()]
```

Imagen No. 3: Función que elimina valores duplicados, Python.

Elaborado por: Paz, Jorge (2025).

```
[8] df=df.drop(columns=['country','Customer','gender','CanalMasUtilizado'])
    print(df.head())
```

Imagen No. 4: Función que elimina columnas no requeridas, Python.

Elaborado por: Paz, Jorge (2025).

Escalado de datos, para este tipo de ejercicios de clustering, que se basan en la distancia euclidiana, si los datos no están escalados, las variables con rangos más grandes como Monetary dominarían el proceso de agrupamiento, resultando en una segmentación distorsionada, esto se resuelve escalando los datos. En la Imagen No. 5 se muestra la función `fit_transform()` en aplicada en Python obteniendo el dataframe `df_scaled`.

```
[ ] from sklearn.preprocessing import StandardScaler
    scaler = StandardScaler()
    df_scaled = scaler.fit_transform(df)
    df_scaled = pd.DataFrame(df_scaled, columns = df.columns )
    df_scaled
```

Imagen No. 5: Escalamiento de datos, Python.

Elaborado por: Paz, Jorge (2025).

Modelado

En este estudio aplicamos 3 algoritmos conocidos por su versatilidad en segmentación de datos:

- **K-Means:** Eficiente para grandes volúmenes de datos, segmenta clientes en grupos homogéneos según características similares, para esto utilizaremos KMeans de scikit-learn.
- **DBSCAN:** Adecuado para detectar grupos de diferentes densidades y gestionar outliers, implementado con DBSCAN de scikit-learn.
- **Agglomerative Hierarchical:** Genera una jerarquía de segmentos para analizar relaciones entre grupos, aplicaremos AgglomerativeClustering de scikit-learn.

CAPITULO IV RESULTADOS Y DISCUSIÓN

K-means clustering:

Determinamos el número de clústeres óptimo mediante el método Elbow, este determinó que valor óptimo es 4, el Gráfico No.1 muestra el Método Elbow para K-means y su valor óptimo.

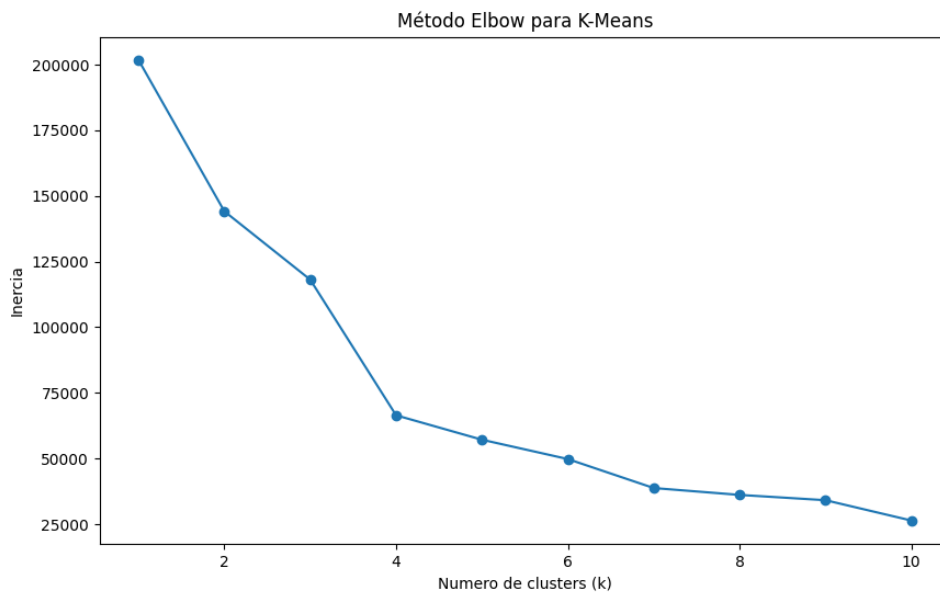


Gráfico No. 1: Método Elbow para K-means.

Elaborado por: Paz, Jorge (2025).

Mediante el método de Silhouette Score, obtuvimos un K óptimo de 4, en el Gráfico No. 2 se muestra esta métrica y su valor óptimo:



Gráfico No. 2: Silhouette Score para K-means.

Elaborado por: Paz, Jorge (2025).

Finalmente ejecutamos el algoritmo K-Means para K=4, la Imagen No. 6 presenta el código Phyton para aplicar K-means.

```
km = KMeans(n_clusters = 4, random_state=20)
y = km.fit_predict(df_scaled)
print(silhouette_score(df_scaled, y))
dfx['k_means'] = y
```

Imagen No. 6: , Función K-Means, Phyton.

Elaborado por: Paz, Jorge (2025).

DBSCAN Clustering

Calculamos las distancias al vecino más cercano, la Imagen No 7. muestra la aplicación de la función *NearestNeighbors()* en Phyton.

```
# Calcular las distancias al vecino más cercano
neighbors = NearestNeighbors(n_neighbors=2)
neighbors_fit = neighbors.fit(df_scaled)
distances, indices = neighbors_fit.kneighbors(df_scaled)
```

Imagen No. 7: Función NearestNeighbors, Python.

Elaborado por: Paz, Jorge (2025).

Ejecutamos el algoritmo DBScan para obtener el mejor Silhouette Score y los parámetros epsilon y número mínimo de puntos, La Imagen No. 8 muestra la ejecución de DBScan para obtener el mejor Silhouette Score, junto con los parámetros epsilon y el número mínimo de puntos.

```
for p in dbscan_params:
    try:
        dbscan = DBSCAN(eps=p[0], min_samples=p[1])
        y_dbscan = dbscan.fit_predict(df_scaled)

        # Calcular Silhouette Score solo si hay más de un cluster
        if len(np.unique(y_dbscan)) > 1:
            sil = silhouette_score(df_scaled, y_dbscan)
            sil_score.append(sil)
            no_of_clusters.append(len(np.unique(y_dbscan)))

            # Actualizar el mejor modelo si se encuentra una mejor puntuación
            if sil > best_sil_score:
                best_sil_score = sil
                best_params = p
                best_labels = y_dbscan
        else:
            sil_score.append(0)
            no_of_clusters.append(0)
    except Exception as e:
        sil_score.append(0)
        no_of_clusters.append(0)
```

Imagen No. 8: Aplicación de DBSCAN y Silhouette-score, Python.

Elaborado por: Paz, Jorge (2025).

Graficando los resultados obtuvimos, el Gráfico No. 3 muestra los resultados obtenidos al graficar la ejecución de DBScan, destacando el número de segmentos óptimos.

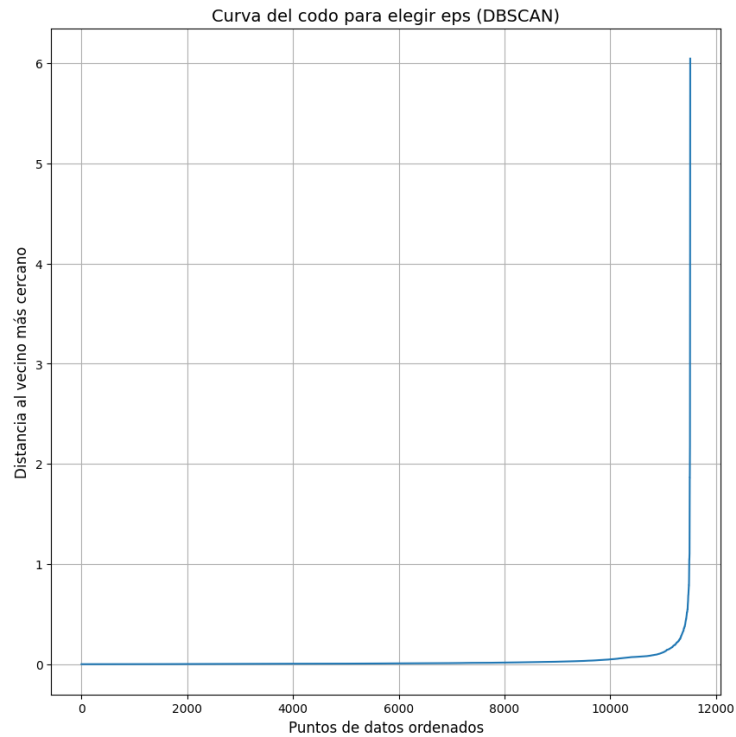


Gráfico No. 3: Método codo (Elbow) para ϵ .

Elaborado por: Paz, Jorge (2025).

Agglomerative Hierarchical Clustering (AHC):

Calculamos la distancia entre los puntos mediante la métrica Ward, el Gráfico No. 4 muestra el dendrograma en función de la distancia obtenida entre los puntos:

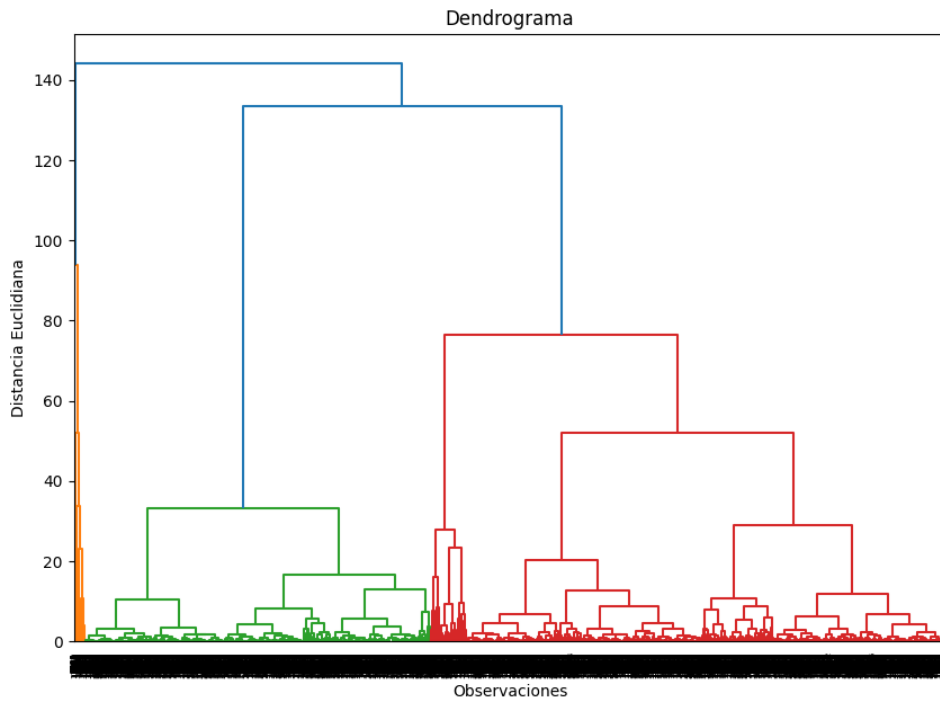


Gráfico No. 4: Dendrograma AHC.

Elaborado por: Paz, Jorge (2025).

Al calcular el número óptimo de clústeres mediante el Silhouette Score, obtuvimos 2 clústeres. El Gráfico No. 5 muestra la representación gráfica de este resultado.

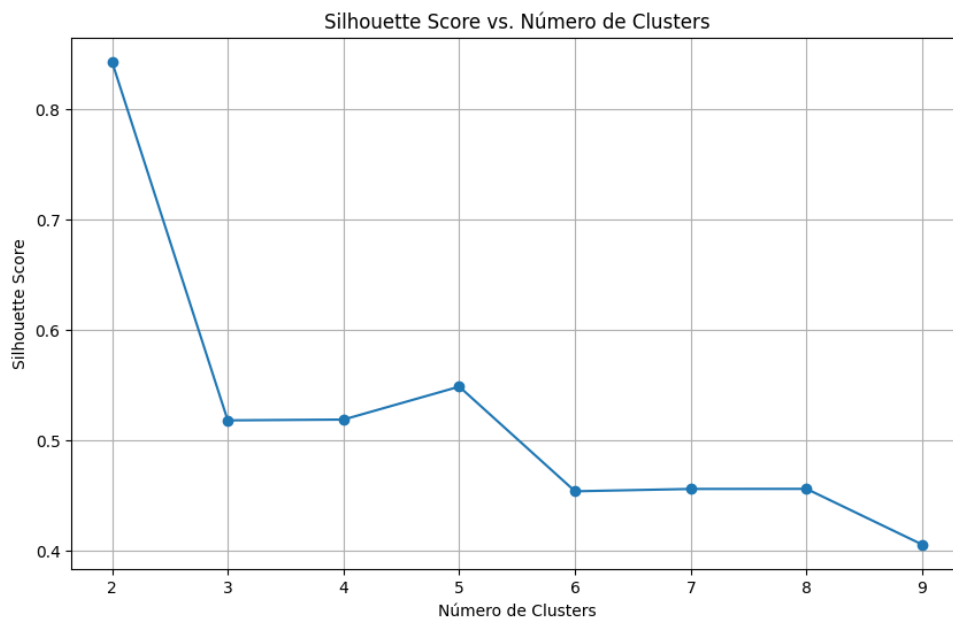


Gráfico No. 5: Silhouette Score para AHC.

Elaborado por: Paz, Jorge (2025).

Se ejecutó el algoritmo de agrupamiento jerárquico aglomerativo (AHC) con una configuración de cuatro clústeres, conforme a criterios estratégicos definidos por Nature's Sunshine Latam. El Gráfico No. 9 muestra su implementación y el resultado del cálculo del *Silhouette Score* correspondiente, utilizando código Python y la librería *scikit-learn* para validar la coherencia de la segmentación.

```
# Ajustar el modelo con 4 clusters
hc = AgglomerativeClustering(n_clusters=4, metric='euclidean', linkage='ward')
y_hc = hc.fit_predict(df_scaled)

# Evaluar el score de silueta con 4 clusters
silhouette_score_hc = silhouette_score(df_scaled, y_hc)
print(f'Silhouette Score para 4 clusters: {silhouette_score_hc}')

# Añadir la columna de clusters al dataframe original
dfx['hc'] = y_hc
```

Imagen No. 9: Aplicación Agglomerative Clustering y silhouette_score, Python.

Elaborado por: Paz, Jorge (2025).

Los resultados obtenidos de la métrica utilizada para evaluar cada uno de los algoritmos de segmentación se muestran en la Imagen No. 10.

	Método	Silhouette Score
0	KMeans	0.548764
1	Clustering Jerárquico	0.518465
2	DBSCAN	0.852849

Imagen No. 10: Resultados Silhouette Score para los métodos aplicados, Python.

Elaborado por: Paz, Jorge (2025).

Aunque DBSCAN obtuvo el mejor Silhouette Score solo fue capaz de identificar un clúster en la segmentación, los investigadores consideran que los resultados de este algoritmo no son útiles para este estudio en particular, por lo tanto los investigadores determinaron que K-means ofrecía una segmentación más adecuada para continuar con el análisis. El comportamiento de las variables RFM fue graficado utilizando los resultados obtenidos con K-means, como se muestra en el Gráfico No. 6. Adicionalmente, se incluye el Gráfico No. 7, que representa la relación entre las variables *Monetary* y *Recency*, permitiendo visualizar patrones de comportamiento entre los clústeres.

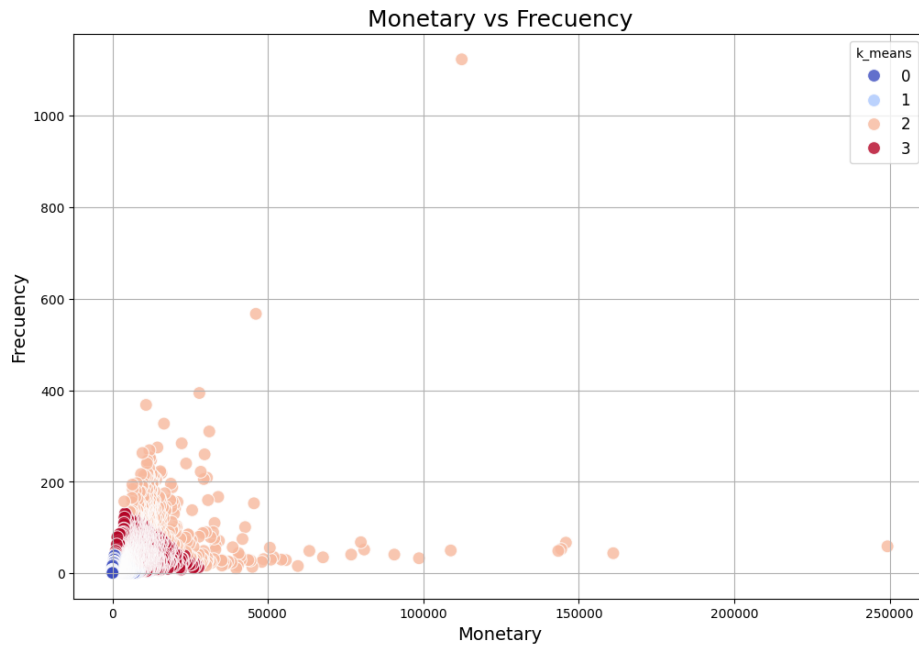


Gráfico No. 6: Comportamiento de las variables Monetary vs Frequency.

Elaborado por: Paz, Jorge (2025).

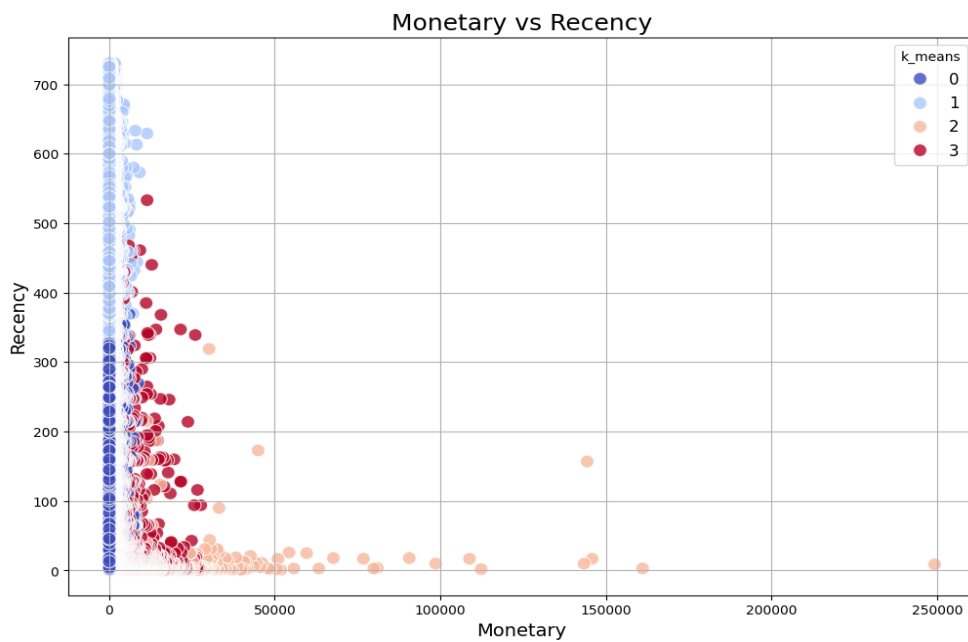


Gráfico No. 7: Comportamiento de las variables Monetary vs Recency.

Elaborado por: Paz, Jorge (2025).

A partir de los resultados obtenidos con el algoritmo K-means, se determinó la distribución de elementos en cada segmento, así como la media de las variables RFM

correspondientes a cada grupo. Esta información se resume en la Tabla No. 3, la cual permite identificar las características promedio de cada clúster generado:

Tabla No. 3: Tabla medias de los resultados del agrupamiento (K-means).

k_means	Recency_Mean	Frecuency_Mean	Monetary_Mean	Count
0	141.48	5.65	482.71	34024
1	533.03	2.21	230.44	30897
2	11.60	127.45	21614.34	344
3	39.11	48.40	7074.02	1988

Elaborado por: Paz, Jorge (2025).

Según la visión estratégica de la compañía, los clientes fueron clasificados en función del comportamiento de las variables RFM, como se detalla en la Tabla No. 4.

Tabla No. 4: Tabla de características de los segmentos de clientes.

	Recency	Frequency	Monetary
Platinum	Baja (↓)	Alta (↑)	Alta (↑)
Gold	Baja (↓)	Media (-)	Media (-)
Silver	Media (-)	Baja (↓)	Baja (↓)
Bronze	Alta (↑)	Baja (↓)	Baja (↓)

Elaborado por: Paz, Jorge (2025).

En el *cluster 2*, con 344 individuos representa al 1% de la muestra y corresponde a los clientes denominados **Platinum**, y son los clientes de más alto valor para la compañía, tienen compras recientes, realizan un alto número de compras y representan para la compañía un importante ingreso monetario, podríamos calificarlos como clientes leales.

En el *cluster 3*, con 1988 individuos representa el 3% de la muestra y corresponde a los clientes denominados *Gold*, a este segmento también podríamos identificarlos como clientes con potencial de crecimiento y también son importantes para la compañía en términos de lo que aportan monetariamente, tiene menos frecuencia de compra que los Platinum y su última compra no es tan reciente.

En el *cluster 1*, con 30897 individuos representa el 45% están representados los clientes denominados *Silver*, son aquellos que realizaron su última compra hace mucho tiempo atrás, hicieron pocas compras y su aporte monetario es relativamente bajo, sin embargo, llama la atención la gran cantidad de clientes en este segmento.

Y, finalmente en el *cluster 0*, con 34024 que representa el 51% los clientes denominados *Bronze*, están representados los clientes de menor valor para la compañía, casi podríamos calificarlos de clientes ocasionales, aquellos que realizaron 1 o 2 compras, su última compra fue hace más de 1 año y su aporte monetario a la compañía fue mínimo, podríamos concluir que es poco probable que realicen una nueva compra, clientes perdidos. Llama la atención a los investigadores su alto número.

Con la información obtenida luego de aplicar las herramientas de BD, podríamos sugerir la aplicación de las siguientes estrategias de marketing focalizadas hacia los segmentos identificados. (Ver Anexo 1.)

Este estudio abre la puerta para la investigación de nuevas y mejores aplicaciones de las herramientas de BD en el ámbito empresarial, por ejemplo: Ser capaz de identificar el mejor objetivo para una oferta específica le permite a la empresa ganar en varios niveles, ya sea optimizando su inversión, aumentando el nivel de satisfacción del cliente, sabiendo que prefieren recibir las ofertas adecuadas a sus preferencias.

Al igual que (Mensouri et al., 2022) aplicamos una combinación de K-means y RFM para identificar comportamientos de los clientes, en nuestro estudio además aplicamos otros dos algoritmos de segmentación DBSCAN y AHC, sin embargo, con los resultados obtenidos podemos concluir que los mejores resultados fueron con K-means. Además (Antonius & Devi Fitriannah, 2024) mediante la aplicación de estas herramientas pudo identificar características específicas de comportamiento en los segmentos de clientes al igual que en nuestro estudio lo que confirma la utilidad que pueden brindar a las

compañías de suplementos naturales.

Contraste con otras investigaciones Los resultados de esta investigación coinciden con hallazgos reportados por Mensouri et al. (2022), quienes comprobaron la efectividad del modelo RFM combinado con K-means para agrupar clientes de forma útil para el marketing. De igual manera, Antonius y Fitriannah (2024) integraron un enfoque similar con variables extendidas, confirmando que esta técnica permite extraer perfiles diferenciados con alto valor estratégico.

Zerbino et al. (2018) plantearon que una segmentación respaldada en datos masivos mejora la gestión de relaciones con los clientes (CRM), lo cual también fue evidenciado en el presente estudio mediante la formulación de campañas específicas por clúster. Además, los resultados empíricos refuerzan lo propuesto por Mukhopadhyay et al. (2024), quienes sostienen que el Marketing 4.0, basado en analítica avanzada, permite generar valor en sectores competitivos y saturados como el de suplementos naturales.

A diferencia de estudios como el de Aditya y Fitriannah (2021), que comparan múltiples algoritmos sin considerar la implementación operativa de estrategias, este trabajo no solo evalúa el rendimiento técnico del modelo, sino que propone aplicaciones prácticas alineadas al negocio. Asimismo, frente a investigaciones como las de Elia et al. (2020), que se enfocan en la infraestructura tecnológica para Big Data, el presente estudio aporta una visión aplicada directamente al análisis de comportamiento del consumidor.

Contraste con otras investigaciones La hipótesis planteada —que la implementación de un modelo de segmentación basado en RFM y clustering mejora la efectividad del marketing— fue verificada con base en métricas analíticas y criterios estratégicos. El Silhouette Score, como estadígrafo principal, alcanzó un valor óptimo para 4 clústeres, evidenciando que los datos presentaban patrones distinguibles y coherentes.

Además, la utilidad de los clústeres se confirmó mediante la formulación de estrategias comerciales diferenciadas y medibles. Este enfoque es congruente con lo planteado por Erevelles et al. (2016), quienes afirman que la segmentación basada en datos transforma no solo la táctica, sino la visión estratégica del marketing.

Por tanto, los resultados del modelado, el análisis descriptivo por segmento y las acciones derivadas confirman empíricamente la hipótesis planteada.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

1. Se logró implementar un modelo de segmentación avanzada de clientes mediante la aplicación del análisis RFM combinado con algoritmos de clustering no supervisado. El algoritmo **K-means** con 4 clústeres mostró el mejor desempeño según el **Silhouette Score (0.548764)**, lo cual permitió identificar grupos de clientes con características transaccionales distintivas y patrones de comportamiento útiles para la formulación de estrategias comerciales.
2. A través del diseño de un sistema de procesamiento y depuración de datos transaccionales, se estructuró una base limpia y confiable que sirvió de insumo para el modelado. Este proceso incluyó técnicas de normalización, eliminación de registros incompletos y codificación adecuada, garantizando la calidad de los datos analizados y cumpliendo el primer objetivo específico.
3. Se identificaron cuatro segmentos bien diferenciados de clientes (estratégicos, frecuentes, nuevos y en riesgo), para los cuales se diseñaron estrategias de marketing personalizadas. Estas incluyeron la asignación de responsables, un estimado de recursos y un cronograma tentativo, lo cual aporta directamente al cumplimiento del tercer objetivo específico.
4. Comparativamente, los resultados coinciden con estudios como el de Mensouri et al. (2022) y Zerbino et al. (2018), quienes demuestran que el análisis RFM y el clustering permiten optimizar la fidelización y rentabilidad del cliente. No obstante, este estudio va más allá al traducir la segmentación en acciones estratégicas concretas, alineadas a las metas comerciales de la empresa.
5. Una de las principales limitaciones del estudio fue la exclusión de variables demográficas o conductuales complementarias, que podrían enriquecer la segmentación. Asimismo, el análisis se realizó sobre datos de una sola región, lo cual restringe la generalización de los hallazgos a otras geografías o industrias.

6. Las implicaciones del estudio abarcan el fortalecimiento de la cultura organizacional basada en datos, la mejora del relacionamiento con los clientes y la generación de modelos replicables para otras unidades de negocio. Además, se promueve el uso de herramientas accesibles como Python y Google Colab para ejecutar proyectos de alto impacto con bajo costo.

Recomendaciones

1. Se recomienda implementar de forma progresiva las estrategias diseñadas para cada segmento, priorizando los clústeres más rentables (clientes estratégicos y frecuentes) con acciones de fidelización y campañas de valor agregado.
2. Establecer un sistema de monitoreo de KPIs por segmento, que permita evaluar la efectividad de las campañas y retroalimentar el modelo RFM de forma continua, garantizando así la mejora sostenida del proceso comercial.
3. Incluir en futuros modelos variables complementarias como frecuencia de reclamos, satisfacción, canal de compra o tipo de producto adquirido, lo que permitiría una segmentación aún más precisa y accionable.
4. Replicar la metodología utilizada en esta investigación en otras filiales o regiones donde opera Nature's Sunshine Latam, con el fin de homogeneizar la estrategia de marketing basada en datos en toda la organización.
5. Fomentar la capacitación interna en ciencia de datos y visualización para los equipos de marketing y comercial, de modo que se promueva una toma de decisiones fundamentada en evidencia y se fortalezca la autonomía técnica del personal.

REFERENCIAS BIBLIOGRÁFICAS

- A. Aguirre. (2022). *Modelos de predicción avanzados para el cálculo de reservas en la industria aseguradora*.
- A. M. Hughes. (1994). *Strategic database marketing: the masterplan for starting and managing a profitable, customer-based marketing program*. Chicago, Ill: Probus Pub. Co., 1994. (Probus Pub. Co., Ed.; III).
- Aditya, D. L., & Fitriana, D. (2021). COMPARATIVE STUDY OF FUZZY C-MEANS AND K-MEANS ALGORITHM FOR GROUPING CUSTOMER POTENTIAL IN BRAND LIMBACK. *Jurnal Riset Informatika*, 3(4).
<https://doi.org/10.34288/jri.v3i4.241>
- Antonius, V. H., & Devi Fitriana. (2024). Enhancing Customer Segmentation Insights by using RFM + Discount Proportion Model with Clustering Algorithms. (*IJACSA International Journal of Advanced Computer Science and Applications*, 15, 902–911).
- Elia, G., Polimeno, G., Solazzo, G., & Passiante, G. (2020). A multi-dimension framework for value creation through Big Data. *Industrial Marketing Management*, 90, 617–632. <https://doi.org/10.1016/J.INDMARMAN.2020.03.015>
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904.
<https://doi.org/10.1016/J.JBUSRES.2015.07.001>
- Gupta, S., Justy, T., Kamboj, S., Kumar, A., & Kristoffersen, E. (2021). Big data and firm marketing performance: Findings from knowledge-based view. *Technological Forecasting and Social Change*, 171, 120986.
<https://doi.org/10.1016/J.TECHFORE.2021.120986>
- H. Kartajaya, P. K. and I. S. (2018). *Marketing 4. 0: Transforma Tu Estrategia para Atraer Al Consumidor Digital*. Editorial Almuzara.
- H. M. Colqui Palacin and R. B. Guevara Amoretti. (2023). *Plan estratégico para la alianza Pillsens-CVS y la introducción del producto MoradOx en el periodo 2022-2026*.
- Mensouri, D., Azmani, A., & Azmani, M. (2022). K-Means Customers Clustering by their RFMT and Score Satisfaction Analysis. *International Journal of Advanced*

Computer Science and Applications, 13(6).
<https://doi.org/10.14569/IJACSA.2022.0130658>

- Mukhopadhyay, S., Singh, R. K., & Jain, T. (2024). Developing big data enabled Marketing 4.0 framework. *International Journal of Information Management Data Insights*, 4(1), 100214. <https://doi.org/10.1016/J.JJIMEI.2024.100214>
- R. Palomo Fernandez. (2024). *Evolución y tendencias emergentes en marketing digital: una exploración de IA, video marketing, personalización de contenidos y marketing de influencers*,.
- S. I. Maguiña Figueroa, J. C. M. V. M. A. B. Y. and H. J. A. G. (2018). *Innovación del modelo de negocio para mejorar la experiencia de compra de los clientes de un supermercado*.
- S. V. Tarazona Figueroa and E. del R. Villanueva Espíritu. (2021). *Estrategias de branding como medio de posicionamiento de la marca Nativo Perú Products de junio del 2021 a junio del 2022, Perú*,.
- Shirazi, F., & Mohammadi, M. (2019). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*, 48, 238–253. <https://doi.org/10.1016/J.IJINFOMGT.2018.10.005>
- Zerbino, P., Aloini, D., Dulmin, R., & Mininno, V. (2018). Big Data-enabled Customer Relationship Management: A holistic approach. *Information Processing & Management*, 54(5), 818–846. <https://doi.org/10.1016/J.IPM.2017.10.005>

ANEXO 1

Estrategias de marketing sugeridas para cada segmento de clientes identificados.

SEGMENTO CLIENTE	Estrategia marketing	Por qué?	Responsable	Temporalidad	Presupuesto	KPI
PLATINUM	1. Programas de fidelización exclusivos (ofrecer puntos, descuentos o acceso a productos exclusivos)	Los clientes Platinum tienen un alto valor para la empresa por lo que mantener su lealtad es fundamental para asegurar ingresos recurrentes.	Marketing / equipo de CRM	A largo plazo con revisión mensual de resultados	\$ 5000 a \$ 10000 anuales/Latam	Tasa de retención de clientes (Retention Rate), CLTV (Customer Lifetime Value)
	2. Ofertas personalizadas de alto valor (descuentos especiales, productos premium).	Aumentar su gasto medio y fomentar la compra recurrente	Marketing / equipo de ventas	Trimestral	\$ 1000 a \$ 3000 / campaña	Tasa de conversión de ofertas personalizadas
GOLD	1. Campañas de reactivación (email marketing, anuncios segmentados)	Reactivar clientes con compras recientes pero con posible riesgo de desconexión.	Marketing / equipo de CRM	A corto plazo / cada 1- 2 meses	\$ 1000 a \$ 3000 / campaña	Tasa de reactivación de clientes (reactivation rate)
	Upselling y Cross-selling (productos relacionados o de mayor valor)	Aumentar el valor de cada compra de los clientes Gold mediante ofertas adicionales	Marketing / equipo de ventas	Continuo	\$ 1,500 a 4,000 / campaña	Tasa de aceptación de upselling/Cross-selling
SILVER	1. Campañas de recordatorio y reactivación (recordatorios por email, SMS, anuncios)	Mejorar la frecuencia de compra mediante recordatorios específicos	Marketing / equipo CRM	A medio plazo / cada 2 meses	\$ 1,500 a 3,000 / campaña	Tasa de conversión de campañas de recordatorio
	2. Ofertas de descuento por tiempo limitado (Descuentos o promociones para generar urgencia)	Incentivar compras rápidas y aumentar la frecuencia de compra	Marketing / equipo de ventas	A corto plazo / cada 1- 2 meses	\$ 1,000 a 1,500 / campaña	Tasa de conversión de promociones
BRONZE	1. Campañas de activación masiva (email marketing, anuncios de facebook, retargeting)	Revivar el interés de clientes inactivos mediante campañas masivas.	Marketing / equipo CRM	A corto plazo / mensual	\$ 500 a \$ 2,000 / campaña	Tasa de apertura de emails, CTR (Click-Through Rate) .
	2. Descuentos por primera compra o incentivos (descuento por una compra o envío gratuito)	Incentivar compras iniciales para reactivar a los clientes	Marketing / equipo de ventas	A corto plazo, 1 vez al mes	\$ 1,000 a 2,000 / campaña	Tasa de conversión de incentivos, tasa de retorno de clientes

Elaborado por: Paz, Jorge (2025)