**Logotipo

Descripción generada automáticamente**

**Segmentación de Clientes Mediante Análisis de Patrones de Compra para la optimización de Estrategias Comerciales.**

Jose Antonio Berrio Lasprilla.

Orlando José Olea Gómez.

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos.

Asesor  
David Manuel Villanueva Valdes, M.Sc.

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

|  |  |
| --- | --- |
| **Cita** | ( Berrio Lasprilla.& Olea Gómez., 2024) |
| **Referencia**  **Estilo APA 7 (2020)** | Berrio Lasprilla,J., & Olea Gómez., O. J. (2024). Segmentación de Clientes Mediante Análisis de Patrones de Compra para la optimización de Estrategias Comerciales.]. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Dedicatoria**

"A nuestras familias,

Con cada paso dado en este viaje hacia la especialización en analítica y ciencia de datos, ustedes han sido nuestra inspiración y nuestro constante apoyo. Sus palabras de aliento, paciencia y comprensión han sido el motor que ha impulsado nuestra determinación y persistencia.

A través de las noches de estudio, los desafíos académicos y los momentos de incertidumbre, su amor incondicional ha sido nuestra mayor motivación. Cada logro alcanzado en este camino es también suyo, pues cada uno de ustedes ha sido parte fundamental de nuestra trayectoria.

Dedicamos este logro a ustedes, nuestra familia querida, por ser nuestra fuente de fuerza y ​​alegría. Que este logro sea también un tributo a su sacrificio y apoyo inquebrantable.

Con todo nuestro amor y gratitud,

Jose Antonio Berrio Lasprilla.

Orlando José Olea Gómez.

**Agradecimientos**

A Dios y a nuestras familias, quienes, de forma directa e indirecta, presencial y no presencial, siempre nos han estado mostrando el camino que debemos recorrer para alcanzar nuestros objetivos. Gracias.

**Tabla de contenido**

Contenido

[Resumen 5](#_Toc167744301)

[Abstract 6](#_Toc167744302)

[1. Descripción del problema. 7](#_Toc167744303)

[1.1. Problema de negocio 7](#_Toc167744304)

[1.2. Aproximación desde la analítica de datos 7](#_Toc167744305)

[1.3. Métricas de desempeño 8](#_Toc167744306)

[2. Objetivos 8](#_Toc167744307)

[2.1. Objetivo general 8](#_Toc167744308)

[2.2. Objetivos específicos 8](#_Toc167744309)

[3. Datos 9](#_Toc167744310)

[3.1. Datos originales 9](#_Toc167744311)

[3.2. Datasets 10](#_Toc167744312)

[3.3. Analítica descriptiva 10](#_Toc167744313)

**Lista de figuras**

[Ilustración 1: Relación de variables 11](#_Toc167745670)

[Ilustración 2: Encabezado de la tabla de datos 11](#_Toc167745671)

[Ilustración 3:Variable edad empresa cliente 11](#_Toc167745672)

[Ilustración 4: Variable Exhibidor 12](#_Toc167745673)

[Ilustración 5:Matriz de correlación entre variables 12](#_Toc167745674)

# Resumen

Con el objetivo de potenciar la productividad del equipo de ventas de Mercantil Zafiro, una empresa líder en el sector de consumo masivo de alimentos, se plantea una estrategia innovadora basada en análisis de datos. Esta estrategia se centra en el desarrollo de un modelo de clusterización utilizando el algoritmo K-means. Este modelo agrupará a los clientes en segmentos homogéneos, aprovechando datos históricos de compras. La comprensión detallada de las diferentes tipologías de usuarios dentro del público objetivo permitirá diseñar estrategias comerciales personalizadas para cada segmento, optimizando así las operaciones de ventas y asegurando un enfoque más eficiente y orientado a resultados.

[**https://github.com/JoseBelas/Momentos\_Evaluativos**](https://github.com/JoseBelas/Momentos_Evaluativos)

[**https://github.com/categat/Momentos\_Evaluativos**](https://github.com/categat/Momentos_Evaluativos)

*Palabras clave*: Modelo de clusterización, segmentación, ventas, Algoritmo K-means.

# Abstract

With the aim of enhancing the productivity of the sales team of Mercantil Zafiro, a leading company in the mass food consumption sector, an innovative strategy is proposed based on data analysis. This strategy focuses on the development of a clustering model using the K-means algorithm. This model will group customers into homogeneous segments, leveraging historical purchase data. Detailed understanding of the different types of users within the target audience will allow the design of personalized commercial strategies for each segment, thus optimizing sales operations and ensuring a more efficient and results-oriented approach.

https://github.com/JoseBelas/Momentos\_Evaluativos

https://github.com/categat/Momentos\_Evaluativos

Keywords: Clustering model, segmentation, sales, K-means algorithm.

# Descripción del problema.

Mercantil zafiro, es una empresa comercial del sector de consumo masivo de alimentos, actualmente se encuentra ante el reto de incrementar la productividad de su equipo de ventas. Para lograrlo, se enfocará en la utilización eficiente y la gestión estratégica de la información recaudada en sus operaciones. Se analizarán los datos históricos comerciales del año 2022, así como la información detallada de los puntos físicos de venta y características específicas de los clientes. Este análisis permitirá desarrollar estrategias precisas, ajustadas a los distintos segmentos de usuarios. Para evaluar el éxito del modelo predictivo, se implementarán métricas de desempeño que incluirán la precisión y la evolución de cada cliente, asegurando así un enfoque orientado a resultados y una mejora continua en las operaciones comerciales.

## Problema de negocio

La empresa Mercantil zafiro desea aumentar la efectividad de sus estrategias comerciales por lo tanto busca aplicar modelos de ciencia de datos para segmentar su maestra de clientes con base en los históricos de ventas en el año 2022

## Aproximación desde la analítica de datos

Para abordar este desafío, se desarrollará un modelo de clusterización que agrupe a los clientes en segmentos homogéneos, se utilizará un enfoque basado en el algoritmo K-means entrenado con un conjunto de datos de compras previamente etiquetadas, clasificación que, al ser entregada, permitirá al equipo de ventas comprender mejor las diferentes tipologías de usuarios que existen dentro del público objetivo de la empresa, con el fin utilizarlo como insumo para el desarrollo de un portafolio de estrategias comerciales y de llegada para cada segmento. Una vez completada esta fase, el modelo clasificará automáticamente a todos los clientes nuevos de la compañía, facilitando así la implementación efectiva de las estrategias previamente establecidas.

## Métricas de desempeño

Las métricas de desempeño incluirán la inertia “Within-Cluster Sum of Squares” (WCSS) del modelo, la cual mide la suma de las distancias al cuadrado entre cada punto y el centroide de su cluster. Adicionalmente, se utilizará el método del codo para ajustar la elección de número de clusters al más óptimo posible. Graficando la suma de las distancias al cuadrado dentro del cluster (WCSS) en función del número de clusters.

# Objetivos

## Objetivo general

Desarrollar un modelo de segmentación de clientes basado en la técnica de machine learning denominada K-Means, utilizando patrones de compras del año 2022, el cual será un insumo para el desarrollo de metodologías comerciales ajustadas para cada grupo y permitirá clasificar automáticamente los clientes nuevos de la empresa con el fin de optimizar los tiempos de ejecución de las estrategias de la empresa.

## Objetivos específicos

* Diseñar algoritmos de segmentación que clasifiquen la base de clientes en grupos homogéneos basados en características compartidas, como tipología de cliente, preferencias de producto, volumen de compra.
* Validar la efectividad del modelo mediante pruebas y análisis a los segmentos identificados.
* Contribuir al desarrollo de estrategias específicas de marketing y ventas para cada segmento de clientes, centrándose en la personalización y la metodología de llegada.
* Reducir el tiempo y los recursos dedicados a la extracción, procesamiento y entrega manual de información comercial por parte del jefe de ventas, supervisores y asesores comerciales.
* Incrementar la rentabilidad de la compañía al dirigir de manera más eficiente los recursos de marketing y ventas hacia los segmentos de clientes más rentables y/o con mayor potencial de crecimiento.
* Establecer métricas claras para medir el rendimiento del sistema de segmentación, incluyendo la mejora en la eficiencia operativa, la retención de clientes y el impacto en las ventas y la rentabilidad.
* Realizar evaluaciones periódicas para identificar áreas de mejora y ajustar las estrategias según sea necesario con el fin de optimizar los resultados obtenidos.

# Datos

## Datos originales

Los datos utilizados para este estudio se extrajeron del informe comercial llamado “Ventas x Material x Cliente” en formato .csv del sistema ERP de Mercantil Zafiro (ECOM). El conjunto de datos abarca variables como el Mes, Identificación del Cliente, Nombre del negocio, Antigüedad del Cliente, Exhibidor, Número de Referencias compradas, Cantidades ingresadas en Devolución, Valor Neto de Mercancía Recibida y Ventas Totales sin IVA, donde este último indica el monto total de compras por cliente antes de impuestos. Esta información refleja la actividad comercial de la empresa durante el año 2022, y al momento de su descarga tiene un tamaño de 3.16 MB

Descripción de las variables:

* Mes: Mes del año en el que se generaron las compras.
* Cliente: Código identificativo del cliente.
* Cliente2: Nombre del cliente.
* Edad Empresa Cliente: Años transcurridos desde el inicio de la relación entre la empresa y el cliente.
* Exhibidor: Activo comercial utilizado para fortalecer la imagen de marca dentro del punto de venta.
* Recuento de Referencia: Número total de distintas referencias que el cliente adquirió en sus compras.
* Suma de Cant. Dev: Cantidad de unidades de producto que el cliente devolvió al momento de las entregas.
* Suma de Cant. Neta: Cantidad de unidades de producto recibidas a satisfacción por el cliente.
* Suma de Vta. - IVA: Valor total de las compras realizadas por el cliente.

## Datasets

El informe comercial extraído del ERP de la empresa inicialmente contenía 17 columnas y 869.927 filas, las cuales fueron depuradas hasta alcanzar un tamaño de 9 columnas y 47.735 filas conservando solo las variables relevantes para el análisis.

Además, se realizó la eliminación de datos atípicos que afectaban la precisión del modelo, y aplicamos transformaciones en nuestra variable mes, reduciendo la dimensionalidad de la variable.

## Analítica descriptiva

Se realizará un análisis descriptivo preliminar de los datos para comprender mejor su distribución y características principales. Este análisis incluirá:

* Distribución de la edad del cliente en la empresa
* Matriz de correlación
* distribución de la variable exhibidor

Texto

Descripción generada automáticamente

Ilustración 1: Relación de variables

De esta manera podemos visualizar que los datos se distribuyen en 6 variables numéricas y 3 categóricas.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Ilustración 2: Encabezado de la tabla de datos

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamenteLa variable edad empresa cliente nos permite conocer los años de relación entre ambas partes, gracias a la gráfica es posible identificar que nuestro público objetivo se concentra entre 3 y 4 años, además queda la incógnita del por qué hay niveles tan bajos entre los 5 y 9 años, donde posiblemente podrán desarrollarse estrategias de negocio para evitar este descenso, a la vez que se mitigan las perdidas por el costo de oportunidad generado.

Ilustración 3:Variable edad empresa cliente

La variable Exhibidor hace referencia a que clientes cuentan con este activo comercial en su punto de venta el cual tiene el objetivo de fortalecer la imagen de marca, mejorar la exhibición de los productos e incrementar ventas. A partir de esta variable, es posible analizar qué tan efectivas han sido las asignaciones realizadas en función de las ventas por clientes, llegando a la conclusión de que el conjunto de clientes que no tienen activo comercial supera en ventas en 66,2% a los que, si tienen asignado activos comerciales, esto se debe a que los mayores compradores son mayoristas que no hacen uso de estos activos en su punto de ventas.

Imagen que contiene Gráfico de rectángulos

Descripción generada automáticamente

Ilustración 4: Variable Exhibidor

Teniendo en cuenta el grafico de correlación la suma de cantidad neta es la que mayor incidencia tiene en la variable relacionada a las ventas, seguida de la variable recuento de referencia. Cabe anotar que este grafico no sirve para evaluar las relaciones de las variables categóricas.

# 

Ilustración 5:Matriz de correlación entre variables