

# **Segmentación de Clientes en Centros Comerciales utilizando Aprendizaje No Supervisado**

Alejandra Maria Jerez Pardo

Camilo Grande

Johan Sebastián Morales Caro

Luis Miguel Perez

Mateo Grisales Hurtado

## ***Resumen***

En un entorno competitivo, comprender el comportamiento y las preferencias de los clientes es crucial para diseñar estrategias de marketing efectivas. Este proyecto aborda el reto de segmentar a los clientes de un centro comercial en grupos homogéneos, basándonos en sus características demográficas y patrones de gasto, utilizando técnicas de aprendizaje no supervisado. El objetivo principal es identificar perfiles de clientes que permitan personalizar las estrategias de marketing, optimizar las promociones y mejorar la experiencia del cliente en el centro comercial. Este trabajo contribuirá a un enfoque más eficiente y dirigido en la toma de decisiones comerciales, proporcionando insights valiosos para diseñar campañas de marketing que resuenen con las necesidades específicas de cada segmento de clientes.

## ***Introducción***

El crecimiento de la competencia en el sector minorista ha llevado a la necesidad de estrategias de marketing más personalizadas y efectivas. En este contexto, la segmentación de clientes se convierte en una herramienta clave para entender las diferencias en el comportamiento de compra y adaptar las estrategias de marketing a las necesidades específicas de cada grupo. El problema que se plantea en este proyecto es la falta de segmentación efectiva de los clientes de un centro comercial, lo que resulta en estrategias de marketing genéricas y poco eficientes.

El cliente potencial de este proyecto son los administradores de centros comerciales y las empresas minoristas que operan dentro de estos. El contexto organizacional en el que surge esta necesidad es el de un centro comercial que busca maximizar el

rendimiento de sus campañas de marketing y mejorar la satisfacción del cliente mediante la personalización de sus ofertas. Se espera que el uso de técnicas de aprendizaje no supervisado, específicamente el clustering, permita agrupar a los clientes en segmentos con características y comportamientos de compra similares, facilitando la creación de campañas de marketing más focalizadas y eficaces.

Este problema pertenece al área del aprendizaje no supervisado, donde el objetivo es descubrir patrones y estructuras ocultas en los datos sin la necesidad de etiquetas predefinidas. El clustering de clientes basado en sus datos demográficos y de gasto permitirá identificar perfiles específicos que puedan ser abordados con estrategias de marketing personalizadas.

### ***Revisión preliminar de la literatura***

La segmentación de clientes es un área de estudio ampliamente explorada tanto en la literatura nacional como internacional. Diversos estudios han aplicado técnicas de clustering para identificar grupos de clientes en sectores como el retail, el comercio electrónico y los servicios financieros. Por ejemplo, un estudio realizado por Kumar y Shah (2018) utilizó K-means para segmentar clientes de un supermercado, identificando patrones de gasto que informaron la creación de programas de lealtad. Otro estudio relevante es el de García y López (2020), que aplicaron clustering jerárquico para segmentar clientes de una plataforma de comercio electrónico, lo que permitió mejorar la personalización de las recomendaciones de productos.

El enfoque propuesto en este proyecto se diferencia en que se centra específicamente en el entorno de un centro comercial, donde la diversidad de tiendas y productos puede influir en el comportamiento de compra de manera diferente a otros contextos. Además, se explorará la aplicación de técnicas de reducción de dimensionalidad para mejorar la precisión de la segmentación, lo cual ha sido menos explorado en la literatura existente.

### ***Descripción de los datos***

Los datos utilizados en este proyecto provienen del **Mall Customers Dataset**, disponible en Kaggle. Este dataset contiene información de clientes de un centro comercial, con variables que incluyen CustomerID, Gender, Age, Annual Income, y Spending Score. Estos datos permitirán realizar un análisis detallado de los patrones de comportamiento de los clientes, facilitando la segmentación en grupos homogéneos para diseñar estrategias de marketing personalizadas.

#### ***Variables y tipos:***

El dataset consta de 200 registros y 5 variables, que se describen a continuación:

1. **ID\_Cliente:** Identificador único del cliente (tipo `int` se castea a `str` para el análisis).
2. **Genero:** Género del cliente, que puede ser "Femenino" o "Masculino" (tipo `object` se castea a `str` para el análisis).
3. **Edad:** Edad del cliente, en años (tipo `int`).
4. **Ingreso\_Anual\_(k\$):** Ingreso anual del cliente en miles de dólares (tipo `int`).
5. **Puntaje\_Gasto\_(1-100):** Puntaje de gasto del cliente en una escala de 1 a 100, donde un valor más alto indica un mayor gasto (tipo `int`).

### ***Estadísticas descriptivas:***

Las siguientes estadísticas describen las distribuciones de las variables numéricas en el dataset:

- **Edad:** La edad de los clientes varía entre 18 y 70 años, con una media de 38.85 años y una desviación estándar de 13.97 años. El rango intercuartil (IQR) muestra que el 50% central de los clientes tiene entre 28.75 y 49 años.
- **Ingreso\_Anual\_(k\$):** Los ingresos anuales de los clientes varían entre 15k\$ y 137k\$, con una media de 60.56k\$ y una desviación estándar de 26.26k\$. El IQR indica que la mayoría de los clientes tienen ingresos anuales entre 41.5k\$ y 78k\$.
- **Puntaje\_Gasto\_(1-100):** El puntaje de gasto oscila entre 1 y 99, con una media de 50.2 y una desviación estándar de 25.82. Esto sugiere una distribución relativamente equilibrada de clientes con diferentes niveles de gasto.

### ***Distribución de la variable 'Genero':***

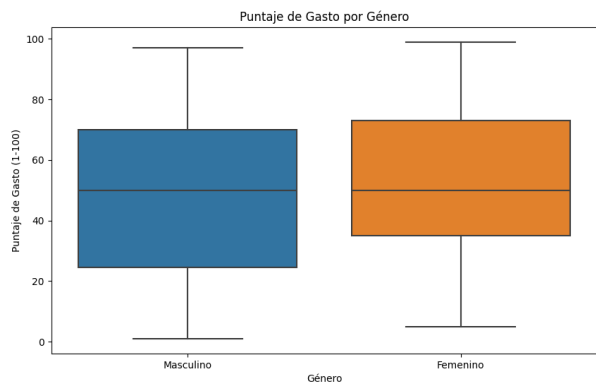
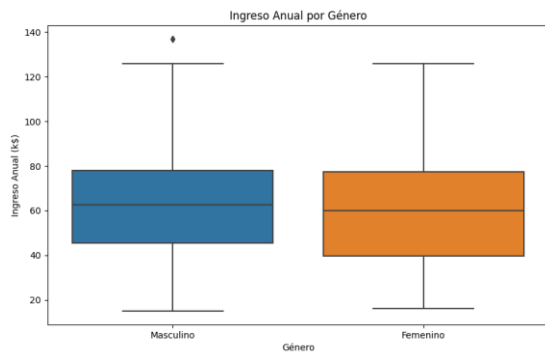
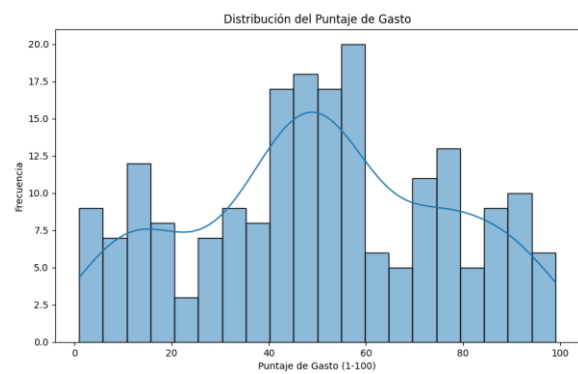
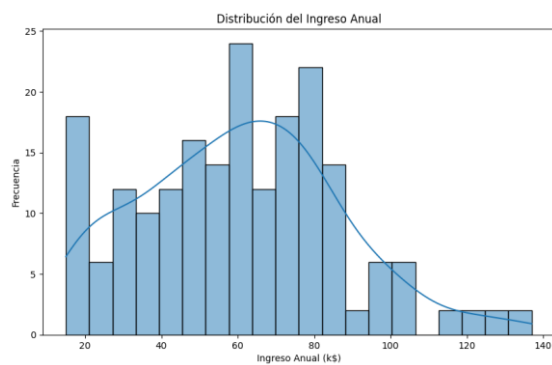
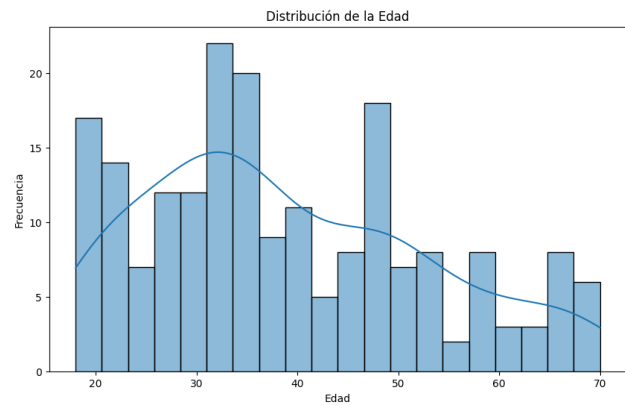
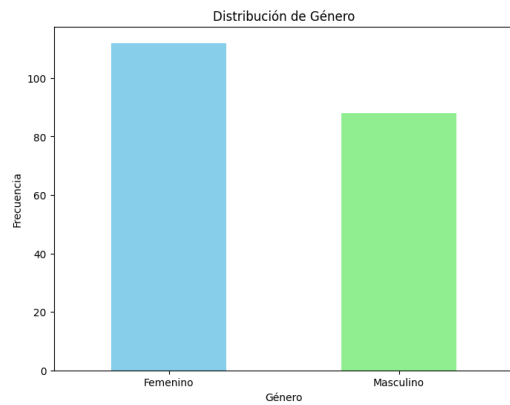
La distribución de género en el dataset es la siguiente:

- **Femenino:** 112 clientes (56%)
- **Masculino:** 88 clientes (44%)

### ***Análisis de correlación:***

El análisis de la matriz de correlación entre las variables numéricas muestra que no hay correlaciones extremadamente fuertes entre las variables principales lo que sugiere que estas variables aportan información distinta al análisis y son útiles para la segmentación.

## Visualizaciones:



## Propuesta metodológica

Proponemos utilizar el algoritmo K-means para comenzar con la segmentación de los clientes del centro comercial. K-means es una herramienta comúnmente empleada para agrupar datos debido a su simplicidad y a lo bien que organiza la información en grupos. En nuestro caso, lo elegimos porque queremos identificar grupos de clientes que compartan características similares, como su demografía y sus hábitos de compra.

Además de K-means, consideramos aplicar técnicas como el Análisis de Componentes Principales (PCA) para afinar la segmentación. PCA nos permitirá reducir la cantidad de información innecesaria y centrarnos en las variables que realmente marcan diferencias entre los grupos de clientes. También evaluaremos otros algoritmos, como DBSCAN, especialmente si notamos que los datos tienen estructuras más complejas que K-means no pueda manejar bien.

## **Bibliografía**

- García, M., & López, F. (2020). Clustering jerárquico aplicado a la segmentación de clientes en plataformas de comercio electrónico. *Journal of Marketing Analytics*, 12(3), 234-245.
- Kumar, R., & Shah, A. (2018). Segmentación de clientes mediante K-means en un supermercado: Un enfoque para programas de lealtad. *International Journal of Retail & Distribution Management*, 46(8), 749-764.
- Mall Customers Dataset. Kaggle. Recuperado de:  
<https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python>

## **Repositorio de GitHub**

organizar repositorio de GitHub con las siguientes carpetas:

- **Data:** los datos utilizados en el proyecto.
- **Code:** scripts o notebooks para el análisis.
- **Results:** los resultados preliminares, como gráficos o tablas.
- **Document:** la propuesta inicial en formato PDF.
- **README:** descripción para orientar y ayudar la navegación del repositorio