

# ELEMENTOS DE APRENDIZAJE DE MÁQUINA Y BIG DATA

## Proyecto Final

### “Segmentación de Clientes en un Shopping”

#### 1. Introducción

**Contexto del Problema:** Un shopping busca implementar estrategias de marketing efectivas para aumentar las ventas. Se dispone de datos anónimos de clientes, incluyendo información sobre género, edad, ingresos anuales y un puntaje de comportamiento.

**Importancia y Relevancia:** El proyecto busca segmentar a los clientes en grupos homogéneos para diseñar estrategias de marketing personalizadas, maximizando la eficacia de las campañas y mejorando la retención de clientes.

#### Objetivos del Proyecto:

- Utilizar el algoritmo K-Means para segmentar a los clientes.
- Analizar las características distintivas de cada grupo.
- Proponer estrategias de marketing específicas para cada segmento.

#### 2. Metodología

**Datos Utilizados:** Se empleará un conjunto de datos proporcionado por un shopping[1], incluyendo género, edad, ingresos anuales y puntajes de comportamiento.

**Herramientas y Tecnologías:** Se utilizarán bibliotecas de Python como NumPy[2], Pandas[3] y Plotly[4], así como algoritmos de Machine Learning, específicamente K-Means[5].

#### Proceso de Análisis/Desarrollo:

1. Limpieza de Datos: Se establece 'CustomerID' como índice y se renombran las columnas.
2. Análisis Exploratorio: Se realiza un análisis descriptivo de las variables.
3. Preparación de Datos: Se eliminan variables no relevantes y se escalan las características.
4. Número Óptimo de Clústeres: Se utiliza el método del codo para determinar el número óptimo de clústeres.
5. Creación del Modelo: Se aplica K-Means con el número óptimo de clústeres.
6. Métricas del Modelo: Se evalúan las métricas Silhouette y Calinski-Harabasz para validar la calidad del agrupamiento.
7. Resultados: Se muestran los grupos obtenidos y se describen sus características distintivas.

### 3. Resultados

**Número Óptimo de Clústeres:** la curva del codo (Figura 1) sugiere que el número óptimo de clústeres es entre 2 y 3 (la disminución de la inercia se produce de forma brusca). Me parece más apropiado usar 5 clústeres. Esto permitiría identificar cinco grupos de clientes distintos en términos de edad, ingresos y puntajes de comportamiento.

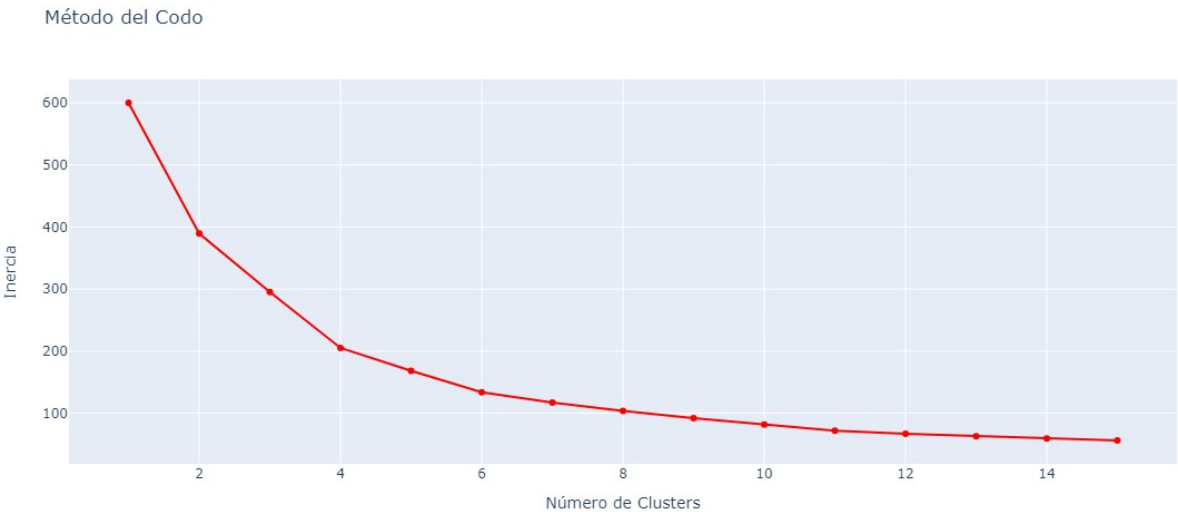


Figura 1. Curva del codo.

Del análisis de los gráficos obtenidos del modelo (Figura 2 y 3), podemos observar las características de cada grupo:

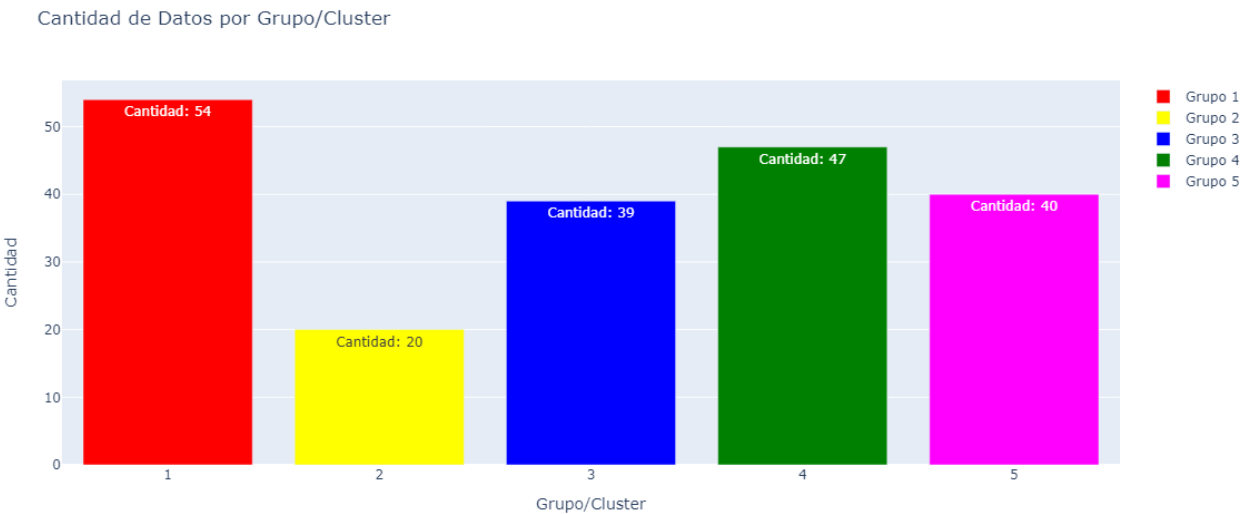


Figura 2. Cantidad de clientes por grupo.

## Grupos de clientes

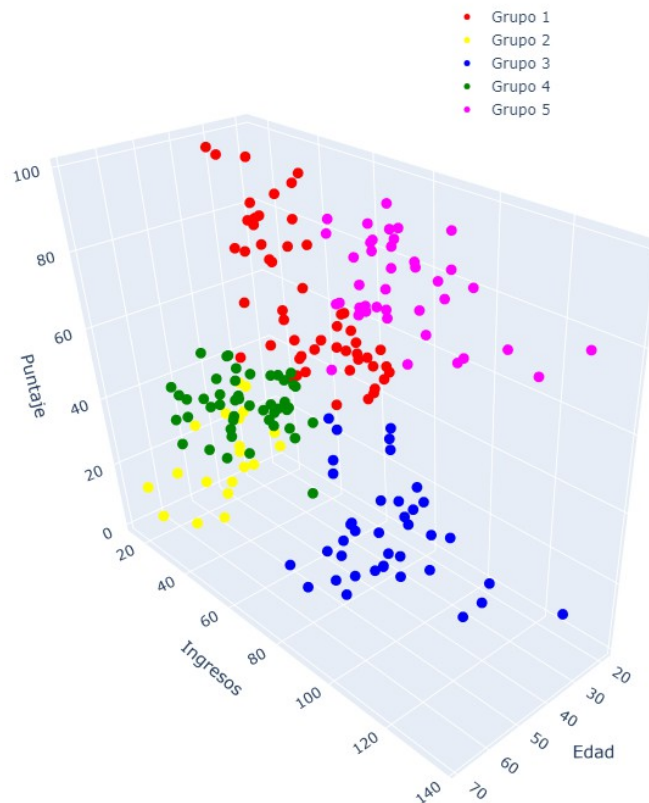


Figura 3. Grupos de clientes

### **Grupo 1 (rojo, cantidad 54):**

Características: Edades jóvenes. Ingresos anuales moderados. Puntuaciones de gasto variadas, algunas altas.

Sugerencias: Estrategias de marketing enfocadas en productos de moda, tecnología y experiencias juveniles. Ofertas especiales o descuentos para incentivar el gasto de aquellos con puntuaciones de gasto altas. Programas de lealtad para fomentar la retención de clientes jóvenes.

### **Grupo 2 (amarillo, cantidad 20):**

Características: Edades variadas. Ingresos anuales moderados a altos. Puntuaciones de gasto más dispersas.

Sugerencias: Segmentación de marketing para abordar las necesidades específicas de diferentes grupos de ingresos. Promociones personalizadas basadas en las preferencias de gasto individuales. Programas de fidelización que ofrezcan recompensas acorde a la variabilidad en el gasto.

### **Grupo 3 (azul, cantidad 39):**

Características: Edades diversas, incluyendo adultos mayores. Ingresos anuales más altos. Puntuaciones de gasto variables.

Sugerencias: Servicio al cliente personalizado para clientes con ingresos más altos. Introducción de productos premium y experiencias exclusivas. Eventos o programas VIP para fomentar la lealtad.

#### **Grupo 4 (verde, cantidad 47):**

Características: Edades mayores con algunos adultos jóvenes. Ingresos anuales variados. Puntuaciones de gasto moderadas a altas.

Sugerencias: Estrategias de marketing centradas en la experiencia de compra y la comodidad. Programas de fidelidad que recompensen compras frecuentes. Promociones especiales para aumentar el gasto en categorías específicas.

#### **Grupo 5 (magenta, cantidad 40):**

Características: Edades variadas, con presencia significativa de adultos mayores. Ingresos anuales moderados. Puntuaciones de gasto variables.

Sugerencias: Ofertas especiales y descuentos para atraer a los adultos mayores. Programas de fidelización con beneficios adaptados a sus necesidades. Enfoque en productos y servicios que puedan ser de interés para este grupo demográfico.

Teniendo en cuenta los datos obtenidos, cómo obtener mejores beneficios depende de la imaginación. Como ejemplo, podría tener un día de descuento en productos que son objeto de interés del segmento objetivo, y evaluar posteriormente con Kmeans si hubo alguna variación favorable en los resultados de ese grupo, y si afectó el comportamiento de los otros por algún motivo.

## **4. Discusión**

Durante el desarrollo del proyecto, se identificaron desafíos relacionados con la selección del número óptimo de clústeres, así como posibles limitaciones en la interpretación de los resultados debido a la variabilidad en los datos y la naturaleza dinámica del comportamiento del cliente. Estos desafíos y limitaciones subrayan la importancia de abordar la incertidumbre y la complejidad asociadas con la segmentación de clientes y la aplicación de estrategias de marketing personalizadas.

En última instancia, los hallazgos principales destacan la relevancia de la segmentación de clientes para el shopping, así como la aplicación efectiva de conceptos de big data y machine learning en la resolución de problemas comerciales. Además, se sugiere que futuras investigaciones podrían incluir la exploración de nuevas variables (por ejemplo, tener en cuenta el género de los clientes) o la aplicación de otros algoritmos de segmentación para mejorar la precisión y la aplicabilidad de las estrategias propuestas. Estas reflexiones finales subrayan la importancia de continuar avanzando en la comprensión y aplicación de técnicas de segmentación de clientes para el beneficio de las estrategias comerciales y de marketing.

## 5. Conclusiones

**Segmentación de Clientes:** El proyecto ha logrado segmentar a los clientes en 5 grupos distintivos, proporcionando información valiosa para estrategias de marketing y se han dado algunas sugerencias para llevarlas a cabo.

## 6. Entrega de Código y Documentación

- Enlace a GitHub: <https://bit.ly/3vcA7Fy>
- El código está documentado y comentado para facilitar su comprensión y reutilización.

## 7. Referencias

- [1] Dataset: <https://bit.ly/3NDV96j>
- [2] Documentación oficial de Numpy: <https://numpy.org/doc/stable/>
- [3] Documentación oficial de Pandas: <https://pandas.pydata.org/pandas-docs/stable/>
- [4] Documentación oficial de Plotly: <https://plotly.com/python/>
- [5] Scikit-learn <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>