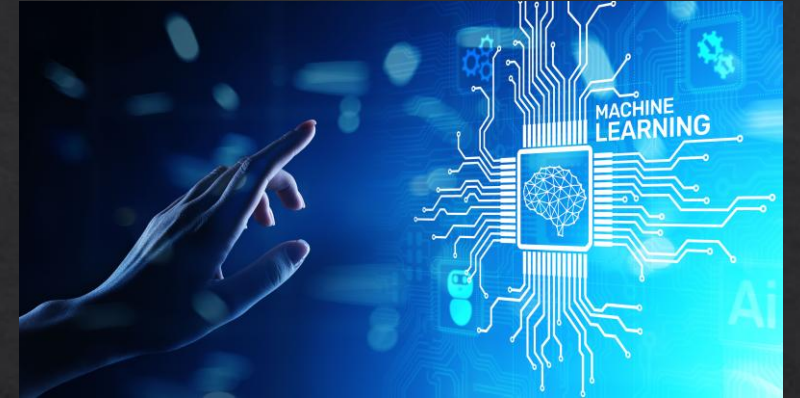


# Segmentación de Clientes utilizando Técnicas de Clustering



Aplicación de Machine Learning para la segmentación de clientes

- Objetivo: Identificar patrones en el comportamiento de clientes.
- Dataset utilizado: *Mall Customers* (Kaggle).



# Datos y Preprocesamiento

- **Descripción del dataset:**
  - Variables: CustomerID, Gender, Age, Annual Income, Spending Score.
- **Pasos de preparación:**
  - Carga y exploración de datos.
  - Limpieza: Eliminación de duplicados y verificación de valores nulos.
  - Conversión de variables categóricas a numéricas (ej. Gender: Male=0, Female=1).
  - Escalado de variables con StandardScaler.



# Exploración y Visualización Inicial

- **Análisis exploratorio:**
  - Visualización de las primeras filas y estructura del dataset.
  - Matriz de correlación para entender relaciones entre variables.
- **Importancia:**
  - Conocer la distribución de datos para detectar patrones y posibles problemas (outliers).

Primeras filas del dataset:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Información del dataset:

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 200 entries, 0 to 199
```

```
Data columns (total 5 columns):
```

#	Column	Non-Null Co
0	CustomerID	200 no
1	Genre	200 n
2	Age	200 n
3	Annual Income (k\$)	200 n
4	Spending Score (1-100)	200 n

```
dtypes: int64(4), object(1)
```

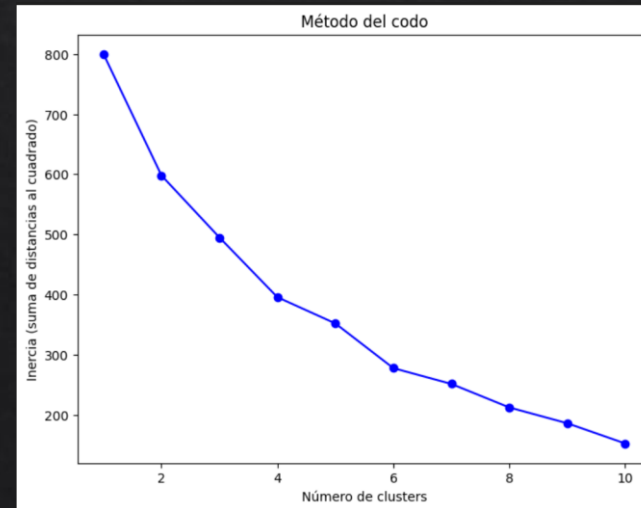
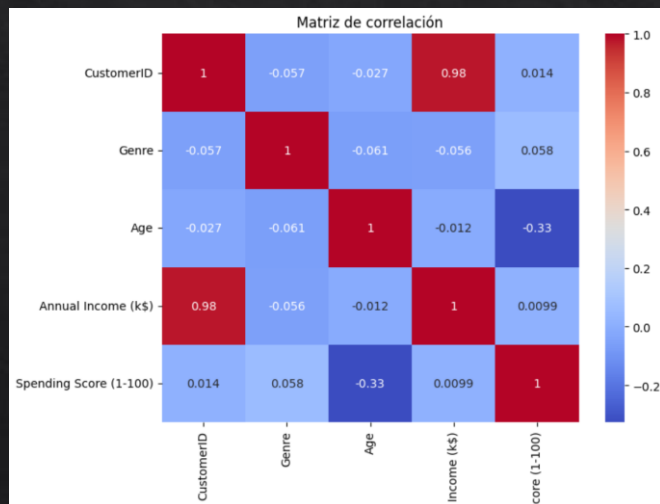
```
memory usage: 7.9+ KB
```





# Modelo de Clustering con K-Means

- **Técnicas utilizadas:**
  - Método del codo para determinar el número óptimo de clusters.
  - Evaluación con Silhouette Score.
- **Visualización:**
  - Gráficos de inercia y Silhouette Score.
  - Scatter plot de "Annual Income" vs "Spending Score" para observar la segmentación.



# Mejoras al Modelo

- **Eliminación de Outliers:**
  - Uso del método IQR para detectar y eliminar valores atípicos.
- **Reducción de Dimensionalidad:**
  - Aplicación de PCA para visualizar y mejorar la calidad del clustering.
- **Experimentación con otros algoritmos:**
  - Clustering jerárquico (Dendrograma).
  - DBSCAN.



# Resultados y Conclusiones

- **Resultados obtenidos:**
  - Distribución de clientes por cluster.
  - Visualización clara de segmentos de clientes.
- **Conclusiones:**
  - La segmentación permite orientar estrategias de marketing.
  - Las mejoras implementadas aumentan la robustez del modelo.

