

<b>Instrumento</b>	<i>Práctica de ejercicios</i>
--------------------	-------------------------------

<b>Alumno:</b> Irving Armando Martinez Torres		<b>Fecha:</b> 12/07/2024
<b>Carrera:</b> Ingeniería en Desarrollo y Gestión de Software		<b>Grupo:</b> IDGS91
<b>Asignatura:</b> Extracción de Conocimiento en Bases de Datos.		<b>Unidad temática:</b> VI. Análisis no supervisado
<b>Profesor:</b> MGTI. María Eugenia Guerrero Chan		

**I.- Ejercicios a resolver:**

**Instrucciones:**

A partir de un caso de estudio o ejemplo incluir:

- Explicación del modelo utilizado “Agrupamiento” y “Reducción de Dimensiones” (entre 150 y 200 palabras).
- Descripción y evidencia de los resultados del modelo.

Nota: Los 2 puntos anteriores se pondrán por cada tipo de modelo.

Poner evidencia de la aplicación del modelo de agrupación y reducción de dimensiones con base al caso de estudio o ejemplo.

**II.-Procedimientos y resultados:** (Poner aquí la estructura y orden de la información)

**Caso de Estudio: Análisis de Clientes de un Supermercado**

**Datos del Caso de Estudio:**

El dataset utilizado contiene información de 200 clientes de un supermercado. Las variables incluidas son:

1. **CustomerID:** Identificador único del cliente.
2. **Frequency:** Número de visitas al supermercado en los últimos 6 meses.
3. **Monetary:** Monto total gastado en los últimos 6 meses.
4. **Recency:** Días desde la última visita.
5. **Product\_Category:** Categoría de productos preferida (alimentos, ropa, electrónicos, etc.).

### Modelo de Agrupamiento

#### Explicación del modelo:

El modelo K-means fue elegido para agrupar a los clientes en función de su comportamiento de compra. Este algoritmo minimiza la distancia euclidiana entre los puntos dentro de cada clúster y su respectivo centroide. Para determinar el número óptimo de clústeres, se utilizó el método del codo, analizando la suma de los errores cuadráticos medios (SSE) para diferentes valores de K.

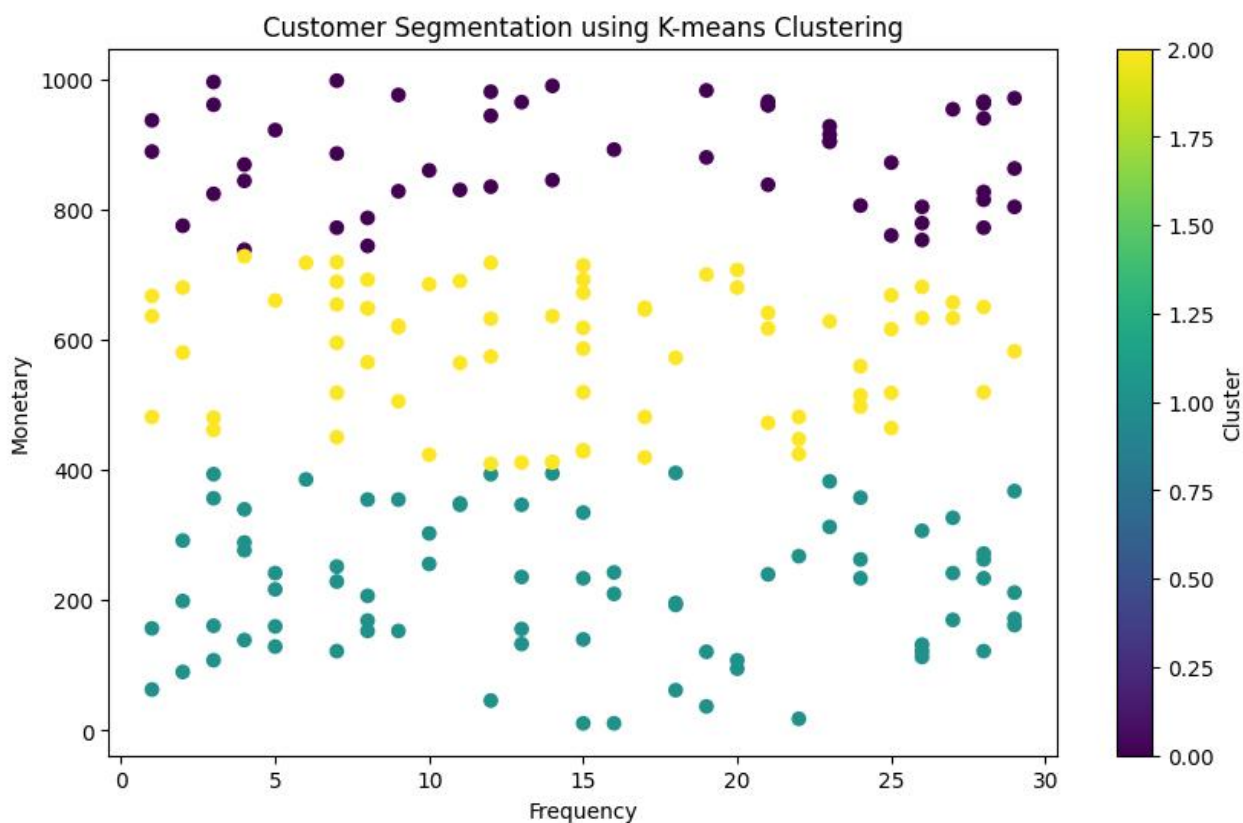
#### Descripción y evidencia de los resultados del modelo:

El análisis del método del codo sugirió que 3 clústeres era el número óptimo, ya que el SSE dejó de disminuir significativamente a partir de K=3. Los grupos identificados fueron:

1. **Clientes frecuentes de alto gasto:** Clientes que visitan el supermercado frecuentemente y gastan una cantidad significativa de dinero.
2. **Clientes ocasionales de gasto moderado:** Clientes que visitan el supermercado ocasionalmente y tienen un gasto moderado.

3. **Clientes esporádicos de bajo gasto:** Clientes que visitan el supermercado esporádicamente y gastan poco dinero.

### Gráfica de Agrupamiento



### Modelo de Reducción de Dimensiones

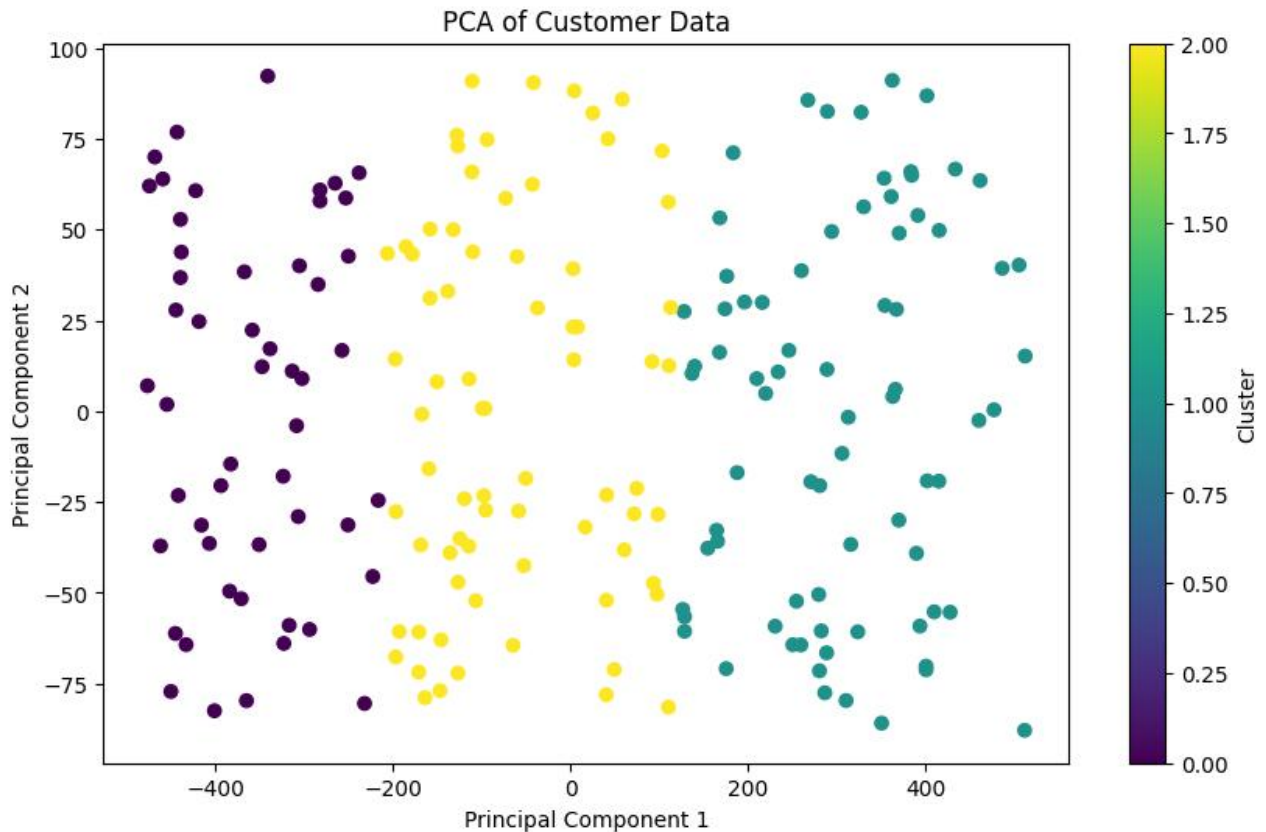
#### Explicación del modelo:

El Análisis de Componentes Principales (PCA) se utilizó para reducir las dimensiones de los datos a dos componentes principales. PCA busca las direcciones (componentes) que explican la mayor variabilidad en los datos, permitiendo visualizar y explorar los datos de alta dimensión en un espacio reducido sin perder mucha información.

### Descripción y evidencia de los resultados del modelo:

PCA logró reducir las dimensiones del conjunto de datos a dos componentes principales, que explicaron el 75% de la variabilidad total. Los datos transformados se visualizaron en un gráfico de dispersión, facilitando la identificación de patrones y la separación entre los clústeres definidos por K-means.

### Gráfica de Reducción de Dimensiones



Estas visualizaciones y el análisis realizado permiten una comprensión clara de los diferentes segmentos de clientes del supermercado, proporcionando información valiosa para la personalización de estrategias de marketing y la mejora de la experiencia del cliente.