

1-.¿Hay alguna variable que no aporta información?

CustomerID generalmente no aporta información útil para análisis, ya que solo es un identificador.

2-.Si tuvieras que eliminar variables, ¿cuáles quitarías y por qué?

CustomerID porque no aporta información sobre patrones de consumo o agrupamiento.

Gender depende del análisis: puede ser útil para segmentar, pero si solo quieres clusters basados en números, no es necesario convertirlo en numérico.

3-.¿Existen variables que tengan datos extraños?

Annual income tiene algunos outliers.

4-.Si comparas las variables, ¿todas están en rangos similares? ¿Crees que esto afecte?

Annual Income (en miles) y Spending Score (1-100) están en rangos distintos.

Sí afecta si quieres hacer clustering (k-means, por ejemplo), porque variables con rangos mayores dominarán la distancia. Solución: normalizar o escalar.

5-.¿Puedes encontrar grupos que se parezcan? ¿Qué grupos son estos?

Puedes usar k-means clustering para identificar grupos. Por ejemplo, segmentar clientes en:

- Clientes con alto ingreso y alto gasto.

- Clientes con bajo ingreso y alto gasto.

- Clientes con ingreso medio y gasto bajo, etc.