1. **Carga tus datos**
2. **Si determinas que alguna variable no sirve basándose en la actividad pasada, elimínala y justifica por qué quitaste o no variables.**

Identificación y Eliminación de Variables

Variables para utilizar en K means

1. **Determina un valor de k**

Tomamos a k = 4 para poder detectar el codo

1. **Utilizando scikitlearn calcula los centros del algoritmo k-means**

**Basado en los centros responde las siguientes preguntas**

**¿Crees que estos centros puedan ser representativos de los datos? ¿Por qué?**

Sí, pero solamente cuando no se dividen en demasiados clusters, permitiendo entender el comportamiento y la similitud de nuestros datos.

**¿Cómo obtuviste el valor de k a usar?**

Mediante la observación de la gráfica, mientras se notara una mayor inflexión (en este caso), mejor

**¿Los centros serían más representativos si usaras un valor más alto? ¿Más bajo?**

Depende. La elección de K es el compromiso entre la representatividad de los centros y la utilidad del modelo. Un K bajo minimiza la representatividad, ya que fuerza la mezcla de patrones dispares en una media. En cambio, un K alto ofrece mayor precisión a nivel matemático, pero sacrifica la utilidad analítica. El punto del "codo" es el valor ideal que garantiza el balance entre la fidelidad de los datos y la simplicidad de la segmentación.

**¿Qué distancia tienen los centros entre sí? ¿Hay alguno que este muy cercano a otros?**

Euclidiana

**¿Qué pasaría con los centros si tuviéramos muchos outliers en el análisis de cajas y bigotes?**

Principalmente, moverían el centroide. Además, K means es muy sensible ante este tipo de circunstancias.

**¿Qué puedes decir de los datos basándose en los centros?**

Este grupo representa contenido que genera una respuesta emocional fuerte y es altamente compartido. Fomenta la conversación y el *engagement* interno (comentarios), típico de videos o preguntas. Son aquellos datos que nos son de gran utilidad.