

## TRABAJO AUTÓNOMO

### Objetivo:

Realizar un análisis completo del dataset proporcionado, aplicando el proceso **KDD (Knowledge Discovery in Databases)** y utilizando algoritmos de clustering no supervisado (**K-Means y Jerárquico**) para identificar patrones, agrupaciones naturales y extraer conclusiones relevantes.

### Tareas a Realizar:

#### 1. Preprocesamiento de Datos (KDD)

- **Limpieza de datos:** Manejar valores faltantes, outliers y ruido.
- **Transformación:** Normalización/Estandarización de variables (si es necesario).
- **Reducción de dimensionalidad:** Aplicar PCA o selección de características (opcional).

#### 2. Análisis Exploratorio de Datos (EDA)

- Estadísticas descriptivas (media, distribución, correlaciones).
- Visualizaciones (boxplots, histogramas, scatter plots) para entender la estructura de los datos.

#### 3. Aplicación de Algoritmos No Supervisados

- **K-Means:**
  - Determinar el número óptimo de clusters (método del codo, silueta).
  - Ejecutar el algoritmo e interpretar los centroides.
- **Clustering Jerárquico:**
  - Usar enfoque aglomerativo con diferentes enlaces (ward, single, complete).
  - Visualizar dendrograma y seleccionar clusters.

#### 4. Evaluación y Comparación

- Comparar resultados de ambos métodos (K-Means vs. Jerárquico).
- Métricas de evaluación (silueta, cohesión, separación).

#### 5. Conclusiones

- Interpretación de los clusters: ¿Qué características definen cada grupo?

- ¿Los clusters encontrados tienen sentido con el contexto del dataset?
- Limitaciones y posibles mejoras.

**Entregables:**

1. **Código (ipynb):** Script completo con comentarios del proceso realizado Paso a Paso.