

## Taller práctico EDA ( Análisis Exploratorio de Datos)

**Dataset:** gym\_members\_exercise.csv (Tracking de Ejercicios en el Gym)

**Librerías permitidas:** pandas, matplotlib, pyplot, seaborn

### Parte 1: Preguntas Teóricas

Analizar en detalle el dataset y dar respuesta a las siguientes preguntas:

1.1 Describirlo con sus propias palabras.

1.2 Definir 5 potenciales casos de uso que se puedan resolver con ML haciendo uso del dataset.

### Parte 2: Preguntas Prácticas

#### Instrucciones:

- ✓ Descarga el dataset.
- ✓ Responde cada pregunta en una celda de código diferente.
- ✓ Usa comentarios (#) para explicar tu proceso si es necesario.

#### 1. Exploración inicial

- ¿Cuántas filas y columnas tiene el dataset? Muestra las primeras 5 filas.

#### 2. Limpieza básica

- ¿Hay valores nulos en el dataset? Si es así, ¿cuántos por columna?

#### 3. Imputación de Datos

- Utilice una estrategia de las aprendidas hasta el momento en el bootcamp para imputar cada uno de los datos que ha encontrado nulos, hágalo por cada feature, explíqueme en cada caso por qué seleccionó cada estrategia.

#### 4. Análisis Univariado

- Realicé un análisis univariado de los features numéricos que presenta el dataset, por favor sacar cinco conclusiones de lo observado.

#### 5. Codificación de features

- Realicé un análisis y defina cuáles de los features categóricos que contiene el dataset deben ser codificados antes de empezar a realizar el análisis

bivariado. Para codificar considere los siguientes algoritmos: **LabelEncoder** ó **OneHotEncoding**.

## 6. Análisis Bivariado

- Realicé un análisis bivariado de los features numéricos que presenta el dataset, por favor sacar cinco conclusiones de lo observado específicamente en la matriz de correlación.

## 7. Escalamiento de features

- Realicé un escalamiento de features del dataset, observé el resultado y deduzca tres conclusiones de lo observado.

## 8. Correlación de features después de escalamiento

- Una vez ha realizado el escalamiento por favor volver a crear la matriz de correlación y observar cómo cambió con respecto el ejercicio anterior, sacar tres conclusiones principales.

## 9. Calcule los datos nulos del feature **Calories\_Burned** haciendo uso de regression

- Por favor vuelva a tomar el dataset original y enfóquese en el feature **Calories\_Burned**, a éste feature impute los datos nulos haciendo uso del algoritmo de regresión visto durante clase. Cuando lo haga observe qué tal el resultado y sugiera siguientes pasos de acuerdo a lo observado.

Los resultados serán revisados durante la sesión.

**LET'S CODING ; BE HAPPY**



**Cata L.**