* LIBRERIAS

**##LIBRERIAS EDA**

**import numpy as np**

**import pandas as pd**

**import seaborn as sns**

**import matplotlib.pyplot as plt**

* CARGA DE DATOS

**artists\_billboard = pd.read\_csv("/Users/pere amat/CE/PIA/CLASSIFICACIÓ/ENTREGABLES/ejer1/artists\_billboard\_fix3.csv")**

* EXPLORACIÓN INICIAL

**VISUALIZAR BASE DE DATOS**

**df.head()**

**TIPO DE DATOS**

**df.info()**

**NOMBRE COLUMNAS**

**df.columns**

**VISUALIZACIÓN , MIN, MAX STD,**

**artists\_billboard.describe()**

**## VISUALIZACIÓN CAMPOS ÚNICOS**

**artists\_billboard['mood'].unique()**

**OBSERVAR VALORES VACIOS**

**df.isnull().sum()**

**VALORES ATÍPICOS DE UN CAMPO**

**Q1 = df['tempo'].quantile(0.25)**

**Q3 = df['tempo'].quantile(0.75)**

**IQR = Q3 - Q1**

**outliers = df[(df['tempo'] < Q1 - 1.5 \* IQR) | (df['tempo'] > Q3 + 1.5 \* IQR)]**

* VISUALIZACIÓN UNIVARIADA

HISOTGRAMA

**plt.hist(df\_imputacion['study\_hours'],**

**bins=15,**

**color='lightblue',**

**edgecolor='black')**

**plt.title('Frecuencia de horas estudiadas')**

**plt.xlabel('study hours')**

**plt.ylabel('Frecuencia')**

**# Show the plot**

**plt.show()**

GRAFICO DE BARRAS

**activity\_means = df[['Very\_Active\_Minutes', 'Fairly\_Active\_Minutes', 'Lightly\_Active\_Minutes']].mean()**

**plt.bar(activity\_means.index,**

**activity\_means,**

**color=['red', 'blue', 'green'],**

**edgecolor='black')**

**plt.title('Media de Minutos Activos por Nivel de Actividad')**

**plt.xlabel('Nivel de Actividad')**

**plt.ylabel('Minutos Promedio')**

**plt.xticks(rotation=0)**

**plt.show()**

BOXPLOT

**# Boxplot de Calories\_Burned**

**plt.boxplot(df['Calories\_Burned'],**

**patch\_artist=True,**

**boxprops=dict(facecolor='skyblue', color='black'),**

**medianprops=dict(color='black'))**

**plt.title("Boxplot de Calorías Quemadas")**

**plt.ylabel("Calorías Quemadas")**

**plt.show()**

* VISUALIZACION BIVAriada

GRAFICO DE DISPERSION

**plt.scatter(df['Fairly\_Active\_Minutes'], df['Calories\_Burned'],**

**color='blue',**

**s=60) # s is equivalent to cex in R for point size**

**plt.title("Relación entre Minutos Bastante Activos y Calorías Quemadas")**

**plt.xlabel("Minutos Bastante Activos")**

**plt.ylabel("Calorías Quemadas")**

**plt.show()**

* ANALISIS DE CORRELACIÓN

**# Matriz de correlación**

**correlaciones = df\_imputacion.corr()**

**print(correlaciones['study\_hours']) # Correlación de la variable faltante con otras**

**# Mapa de calor de correlación**

**sns.heatmap(correlaciones, annot=True, cmap='coolwarm')**

**plt.show()**

* GESTION DE VALORES PERDIDOS

**OBSERVAR VALORES VACIOS**

**df.isnull().sum()**