Lab Task: Suicide Rate Dataset Analysis

In this lab, you will work with a dataset containing information about suicide rates across countries, years, genders, age groups, and socio-economic factors. The objective is to practice data cleaning, feature engineering, categorical encoding, and exploratory data analysis (EDA).

Part 1 – Data Understanding & Cleaning (Pandas library basic functions)

1. Load the dataset and explore its structure (rows, columns, data types, missing values).
2. Handle missing values using imputation if any (mean/median for numerical, mode for categorical)
3. Remove duplicates if any. (drop_duplicates() function) if any else 0

Part 2 – Feature Engineering

Create new features from existing columns:

1. GDP per capita (verify and convert to numeric). (or another column that should be converted to numeric)
2. Suicide-to-GDP ratio.
3. Drop irrelevant features

Part 3 – Exploratory Data Analysis (EDA)

Answer the following questions using plots and numerical summaries:

1. Is the suicide rate more prominent in some age categories than others?
2. Which countries have the most and the least number of suicides??
3. What is the effect of GDP and GDP per capita on suicide rates?
4. What is the trend of suicide rates across the years?
5. Is there a difference between male and female suicide rates?
6. Which generation shows the highest suicide rates?