

## Assignment 2

**Objective:** Use Pandas for data manipulation and analysis on a social science dataset, focusing on data cleaning, organization, and basic exploration.

**Dataset:** Use the 'Gapminder World' dataset, which includes information on countries' GDP, life expectancy, population, and more over several years. This dataset is appropriate for social science and is accessible from Gapminder.org.

### Tasks:

1. Load the Dataset:
  - Import the Gapminder dataset into a Pandas DataFrame.
2. Data Merging:
  - Convert from wide to long, if necessary.
  - Merge different parameters such as gdp, population, life expectancy, and so on.
3. Data Cleaning and Preparation:
  - Identify and handle missing values in the dataset.
  - Rename columns if necessary to make them more descriptive (e.g., rename 'pop' to 'Population').
4. Data Exploration:
  - Display the first and last 5 rows of the dataset.
  - Identify the number of unique countries represented in the dataset.
  - Extract data for a specific year (e.g., 2000) and create a new DataFrame.
5. Data Filtering:
  - Select all data related to a continent of your choice (e.g., Africa).
  - Filter out countries with a population less than 1 million.
6. Data Organization:
  - Sort the dataset based on life expectancy in ascending order.
  - Create a new column that categorizes countries into 'High', 'Medium', or 'Low' population based on defined thresholds.
7. Basic Data Analysis:
  - Calculate the average GDP per capita for each continent.
  - Identify the country with the highest life expectancy in a given year.
8. Data Visualization:
  - Use basic plotting functions in Pandas (e.g., plot, bar) to visualize the trends in life expectancy over the years for a selected country.

**Submission:** A Jupyter notebook containing the code, outputs, and brief comments explaining each step of the analysis. For your comments and headers, please use the *markdown* option.