Capstone Project: Suicide Data Analysis

Introduction:

The dataset provided contains information on suicide rates across different countries and years. It includes demographic and economic factors that can be analyzed to identify patterns and trends in suicide occurrences. The dataset covers attributes such as country, year, gender, age group, suicide count, population, GDP, and other relevant metrics.

Data Dictionary:

- **Country:** The name of the country where the data is collected.
- **Year:** The year the data was recorded.
- **Sex:** Gender classification (Male or Female).
- **Age Group:** Age range of individuals in the dataset.
- **Suicides_no:** Total number of suicides reported.
- **Population:** Population of the respective age group and country.
- **Suicide_rate:** Suicide rate per 100,000 people.
- **GDP_per_capita:** Gross Domestic Product per capita.
- **HDI_for_year:** Human Development Index for the respective year.
- **Generation:** Generation classification (e.g., Silent, Boomers, Gen X, Millennials).

Instruction Guidelines:

Data Cleaning:

- Identify and handle missing values in the dataset.
- Ensure data types are consistent across all columns.
- Check for duplicate records and remove if necessary.

Exploratory Data Analysis (EDA):

- **Summary Statistics:**
  - Compute descriptive statistics (mean, median, min, max) for numerical columns.
  - Analyze the distribution of suicide rates across different years and countries.
- **Temporal Trends:**
  - Examine changes in suicide rates over the years.
  - Analyze gender-wise suicide trends.
- **Economic and Social Factors:**
  - Investigate the correlation between GDP per capita and suicide rates.
  - Compare suicide rates among different generations.

Visualization:

- Generate visualizations such as line charts, histograms, bar charts, and scatter plots.
- Represent suicide trends across years, gender distribution, and regional differences.
- Explore correlations between economic indicators and suicide rates.

Conclusion:

- Summarize the key findings from the analysis.
- Highlight trends and any significant observations.
- Provide insights into possible causes or patterns identified in the data.

Documentation & Submission:

Submit the following:

- GitHub Repository Link (Upload your Analysis to your Github).
- Jupyter Notebook Link (For easy access to your analysis and code and make sure you grant access).
- Word Document (Containing key insights and trends from your analysis).

Practice Questions(Just a guide)

1. Identify the top ten countries with the highest suicide rates./ Visualization
2. Identify the least ten countries with the lowest suicide rates./ Visualization
3. Analyze how suicide rates have changed over the years.
4. Compare the suicide rate between males and females.
5. Investigate if there is a correlation between GDP per capital and suicide rates.
6. Identify the most affected generation in terms of suicide rates.
7. Determine which age group has the highest suicide rate.
8. Find out if any country has shown a consistent decrease in suicide rates over the years.
9. Visualize the distribution of suicide rates across different continents.
10. Compare suicide rates for a specific year across multiple countries.
11. Analyze the impact of economic recessions on suicide rates.

This capstone project aims to help students apply data analysis skills to a real-world dataset, allowing them to derive meaningful insights and trends related to public health and socioeconomic factors.