Numerical Analysis of Health Indicators with Pandas

1. Data Loading and Inspection

Task: Load the CSV data into a pandas DataFrame and inspect the structure.

Explanation: Use pd.read_csv() to load data. Utilize .info() and .describe() to understand column types and summarize statistics.

Key Functions: pd.read_csv(), .info(), .head(), .describe()

2. Handling Missing Values

Task: Identify missing values and fill them appropriately.

Explanation: Use .isnull() and .sum() to detect missing values. Fill missing numeric values with the mean using .fillna().

Key Functions: .isnull(), .sum(), .fillna()

3. Descriptive Statistics for Life Expectancy by Country

Task: Calculate the mean, standard deviation, minimum, and maximum for life expectancy values by country.

Explanation: Use groupby() to group data by country, and apply aggregate functions like mean(), std(), min(), and max().

Key Functions: .groupby(), .agg()

4. Life Expectancy Comparison by Gender

Task: Calculate the average life expectancy for each gender across all countries.

Explanation: Filter data by gender column and apply .mean() to the life expectancy column.

Key Functions: .loc[], .mean()

5. Correlation Analysis of Life Expectancy Range

Task: Calculate the correlation between low and high life expectancy ranges.

Explanation: Use .corr() to compute Pearson correlation between two columns.

Key Functions: .corr()

6. Regional Analysis of Life Expectancy

Task: Group countries by region and calculate the average life expectancy for each region.

Explanation: Use .groupby() to group data by region and calculate mean life expectancy.

Key Functions: .groupby(), .mean()

7. Identifying Highest and Lowest Life Expectancy

Task: Identify countries with the highest and lowest life expectancy for 'Male' and 'Both sexes.'

Explanation: Use .idxmax() and .idxmin() to locate the rows with maximum and minimum values.

Key Functions: .idxmax(), .idxmin()

8. Outlier Detection in Life Expectancy

Task: Identify outliers in life expectancy using the Interquartile Range (IQR) method.

Explanation: Compute Q1 and Q3 with .quantile(), calculate IQR, and filter outliers.

Key Functions: .quantile(), boolean indexing

9. Comparing Life Expectancy Against a Threshold

Task: Calculate the proportion of countries with life expectancy above a threshold (e.g., 50 years).

Explanation: Use boolean indexing to filter rows and .count() to determine proportions.

Key Functions: Boolean indexing, .count()

10. Trend Over Time (If Period Data Exists)

Task: Calculate life expectancy changes over time for each country.

Explanation: Use .diff() to compute changes across time periods, highlighting trends.

Key Functions: .diff()