**World Bank Data**

Documentation: Reproducing Hans Rosling's Animated Graph

**1. Load Data:**

Life Expectancy at Birth (Dataset 1):

Load the dataset containing life expectancy at birth.

Check the structure and preview the data to understand column names and contents.

Fertility Rate (Dataset 2):

Load the dataset containing fertility rate information.

Inspect the dataset to understand its structure and check for any missing values.

Country Population (Dataset 3):

Load the dataset containing country population data.

Review the data to ensure proper loading and assess any potential issues.

**2. Data Overview:**

Explore Each Dataset:

Examine summary statistics, such as mean, median, and standard deviation, for each dataset.

Identify any outliers or unusual patterns in the data that may need attention.

Data Distribution:

Visualize the distribution of life expectancy, fertility rate, and population using appropriate charts (histograms, box plots, etc.)

**3. Handle Missing Values:**

Identify Missing Values:

Check each dataset for missing values and assess their impact on the analysis.

Imputation Strategy:

Decide on an imputation strategy for handling missing values, considering factors like dataset size and the importance of missing values in the analysis.

**4. Data Types:**

Data Type Verification:

Confirm that the data types of each column are appropriate for the analysis.

Convert data types if necessary (e.g., converting numeric values stored as strings to float).

Datetime Conversion:

If applicable, convert date columns to datetime objects for proper time-based analysis.

**5. Merge DataFrames:**

Utilize DataFrame.merge():

Combine the three datasets using the appropriate columns for merging (e.g., country code or name).

Create a unified dataset with information on life expectancy, fertility rate, and country population.

Verify Merged Data:

Confirm that the merged dataset accurately reflects the information from all three sources.

Check for any discrepancies or unexpected changes in the data after merging.

Additional Visualizations:

Population Trends (Years vs. Population):

Create a line graph illustrating population trends over the years.

Fertility Rate Distribution:

Explore the distribution of fertility rates across different regions.

Life Expectancy Variation:

Visualize the variation in life expectancy over time.

Correlation Analysis:

Conduct a correlation analysis between life expectancy, fertility rate, and population.

Regional Analysis:

Utilize the provided Metadata\_Country dataset to assign continents or regions to countries.

Perform a regional analysis to understand how life expectancy and fertility rate vary across different parts of the world.

Hint Implementation:

Create Continent Column:

Use the Metadata\_Country dataset to create a new column specifying the continent for each country using if-else conditions.

This ensures that the analysis can be stratified by continent for better insights.

