World Bank Data

In 2006, Hans Rosling gave a TED talk titled The best stats you've ever seen.(Must watch this video before working on the project). At the beginning of the talk, he showed an animation he made to debunk some misconceptions about today's world.

I enjoyed seeing this visualisation and I want you to reproduce it with the tools you know (i.e. Python, Pandas, Numpy, Seaborn and Matplotlib).

Dataset Information

- **Life expectancy at birth**: The number of years a newborn would live if the patterns of mortality at the time of birth remain the same throughout his life.

- **Fertility rate**: Number of children a woman would give birth to during her childbearing years.

- **Country population**: Total number of residents regardless of legal status or citizenship (midyear estimates)

Hans Rosling built this animation, after testing his students on global health, he realised that they still thought that the world was divided in two:

- The Western world: low fertility rate and high life expectancy
- The third world: high fertility rate and low life expectancy

There will be some differences between the original visualisation and the one you are going to build:

- **More data**: The talk was made in 2006 with data from 1962 to 2003. We will use data from 1960 to 2016.

- **Regions:** The original visualisation has five regions. We will keep the regions from the source data (i.e. seven regions).

- **Colours**. We can't get the exact colours of the regions. Feel free to use your colour mapping.

In this task, you have to build the same animated graph as you watched in the video. Before you start with the visualization you have to perform the following steps:

1. Load Data
2. Data Overview
3. Handle Missing Values
4. Data Types
5. Merge DataFrames (If required for any visualization)

So far we have performed the following steps on a few datasets so we are not mentioning minute details for the above task.

In case you want to perform basic visualizations apart from the animated one, consider the below pointers:

1. Population Trends (Years vs Population)(Line Graph)
2. Fertility rate distribution
3. Life expectancy variation
4. Correlation Analysis
5. Regional Analysis

This project might be challenging so I would recommend you seek feedback from your peers and be active on the discord channel to help each other out and build this.

Hint:
1. Load all three datasets

2. Data preprocessing on each dataset

3. Use the Pandas dataset merge method (DataFrame.merge) to combine the 3 datasets

4. Create a list of continents with their counties and then use if else to create a new column for each continent.

I have uploaded one more dataset `Metadata_Country` In this dataset you will find the column Region along with the Country code. You don't have to do the manual work to identify the Region.