**EDA COMPREHENSIVE PROJECT**

**Problem Statement:**

The modern world is shaped by complex dynamics in population, health, and economics, making understanding these trends vital for informed policy-making. Global Trends, a leading analytics firm, is dedicated to deciphering these patterns through a comprehensive analysis of the Gapminder dataset. Your role in this project is to conduct an in-depth Exploratory Data Analysis (EDA), uncovering the intricate relationships between demographic changes, economic development, and health advancements over recent decades.

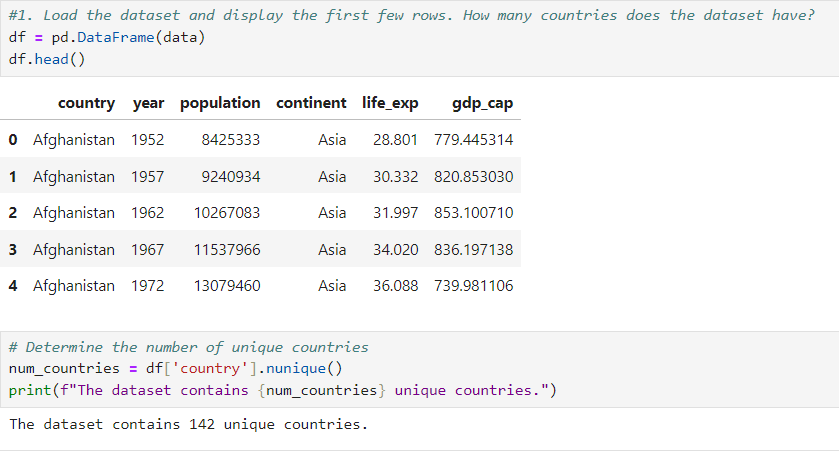
**NOTE:** For subjective questions, explain your answer in theory with clear examples and thoughtful analysis of the reasons behind the trends.

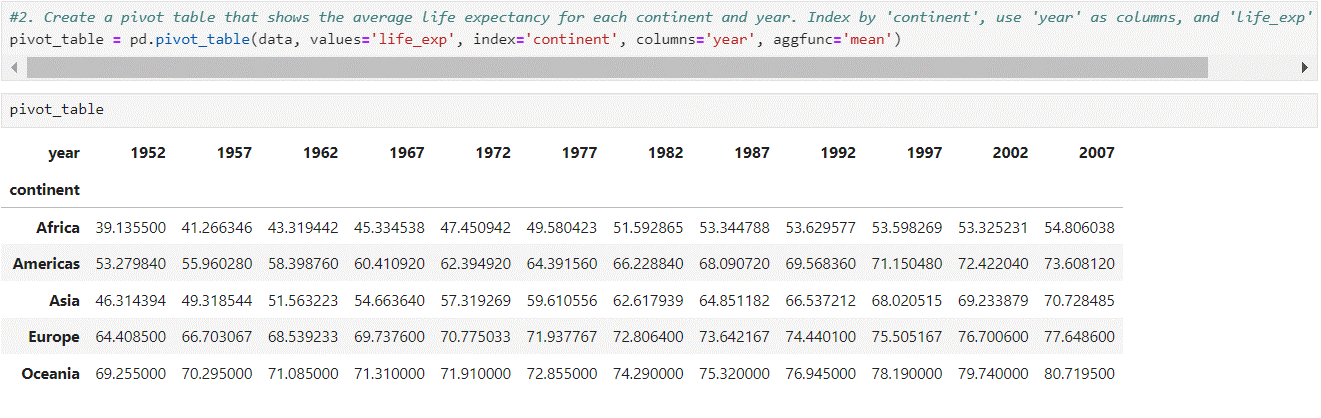
**Dataset:** [Gapminder World](https://drive.google.com/uc?id=1fDGZh86UPUkt2K6enlNQfB0mswU8pB_P)

**CodeLink:** [EDA-1\_Project.ipynb](https://colab.research.google.com/drive/1hA5Cq7J3TruiQTLV7HKw652kSLpDi37x?usp=sharing)

**Objectives:**

1. Load the dataset and display the first few rows. How many countries does the dataset have?



1. Create a pivot table that shows the average life expectancy for each continent and year. Index by 'continent', use 'year' as columns, and 'life\_exp' as values.
2. Which countries had a GDP per capita higher than the 75th percentile in 2007?



1. Categorize the 'life\_exp' into 4 equally ranged bins from 'Low' to 'Very High'. Use cut to create these categorical life expectancy groups and add them as a new column 'Life\_Exp\_Range'.

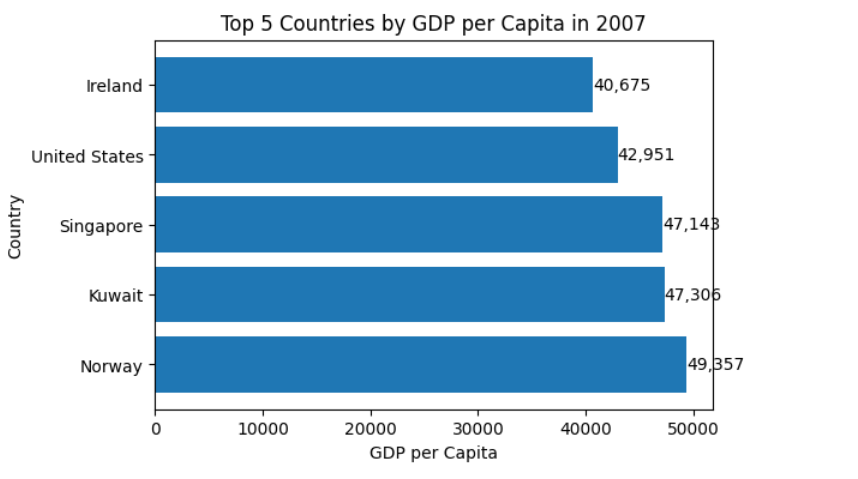
A screenshot of a computer

Description automatically generated

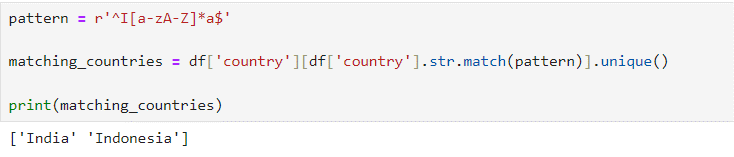
1. Identify the top 5 countries with the highest GDP per capita in 2007. Use a horizontal bar chart to display this data.

A screen shot of a computer code

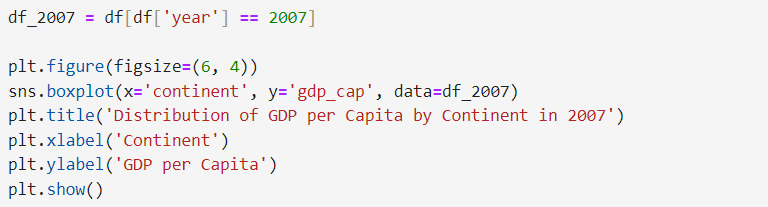
Description automatically generated



1. Find all country names that start with "I" and end with "a" using regex.



1. Create a boxplot using Seaborn to compare the distribution of GDP per capita for each continent in 2007.



A chart of a distribution of gdp per continent

Description automatically generated

1. Find all countries with a life expectancy of over 80 years in 2007. List these countries and their respective continents.

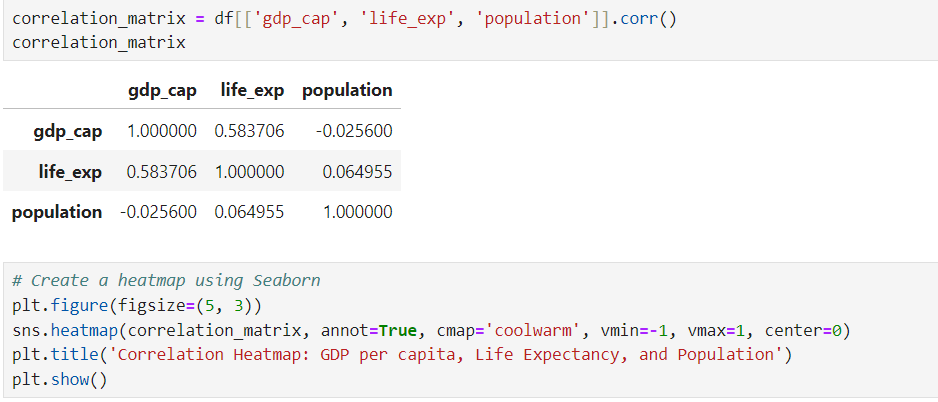


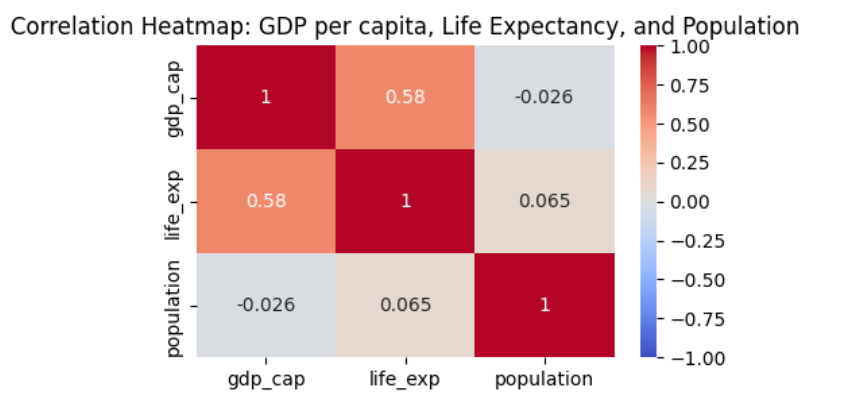
1. Convert the 'year' column to a datetime type and extract the decade. Create a new column 'Decade' that groups the years into decades (e.g., the 1950s, 1960s).

A screenshot of a computer

Description automatically generated

1. Compute the correlation matrix between GDP per capita, life expectancy, and population for the dataset. Then, use Seaborn to visualize this correlation matrix as a heatmap.





1. How has the global average life expectancy changed from 1952 to 2007? Plot a line graph to visualize this trend.

Subjective Question: Discuss the various reasons that could have contributed to the change.

The reasons that could have contributed to this change are numerous and complex. Some possible factors include:

1. **Improvements in healthcare:** Advances in medical technology, increased access to healthcare services, and the development of new treatments and vaccines have all contributed to a decline in mortality rates and an increase in life expectancy.
2. **Decline in infectious diseases:** The discovery of antibiotics and the development of vaccines have led to a significant decline in the incidence of infectious diseases, which were once a major cause of death.
3. **Improved sanitation and hygiene:** Advances in sanitation and hygiene have reduced the spread of diseases and improved overall health.
4. **Changes in lifestyle:** Changes in diet, exercise, and other lifestyle factors have contributed to a decline in the incidence of chronic diseases, such as heart disease and diabetes.
5. **Economic development:** Economic growth and development have led to improved living standards, better access to healthcare, and a decline in poverty, all of which have contributed to an increase in life expectancy.
6. **Education:** Increased access to education has led to better health literacy and healthier behaviors.
7. **Government policies:** Government policies and programs aimed at improving healthcare and reducing poverty have also contributed to the increase in life expectancy.

These are just a few possible factors that could have contributed to the change in global average life expectancy. The actual causes are likely to be complex and multifaceted and may vary depending on the specific region or population being studied.

A screenshot of a computer code

Description automatically generated

A graph with a line graph

Description automatically generated

1. For the year 2007, analyze the relationship between life expectancy and GDP per capita. Subjective Question: Is there a noticeable trend or correlation? Represent this using a scatter plot.

A screenshot of a computer

Description automatically generated

A graph showing a red line and blue dots

Description automatically generated

**Observation from Scatter Plot:**

* **Correlation/Trend**: There is a noticeable correlation or trend between life expectancy and GDP per capita.
* **General Trend**: As GDP per capita increases, life expectancy tends to increase as well.
* **Exceptions and Outliers**: There are exceptions like Afghanistan (low GDP per capita and life expectancy) and Japan (high GDP per capita and life expectancy).
* **R-squared**: The R-squared value is approximately 0.73, indicating a strong positive correlation between life expectancy and GDP per capita.

**Possible Reasons for the Correlation:**

* **Access to Resources**: Higher GDP per capita often reflects better access to healthcare and education, which can lead to improved health outcomes.
* **Standard of Living**: Economic development generally leads to a better standard of living, reduced poverty, and increased access to essential services like clean water and sanitation.
* **Investment in Healthcare**: Higher GDP per capita may lead to more investment in healthcare infrastructure, research, and development, which can translate into better health outcomes.
* **Complex Relationship**: The relationship between life expectancy and GDP per capita is complex and influenced by many factors.

1. Compare the average GDP per capita for each continent in the year 2007. Use a bar chart for this comparison.

Subjective Question: Why is the average GDP per capita for Oceania higher than the Americas even though the Americas have more countries?

A screenshot of a computer program

Description automatically generated

A graph of blue bars

Description automatically generated

From the bar chart, we can see that Oceania has the highest average GDP per capita, followed closely by Europe. The Americas have a lower average GDP per capita than Oceania, despite having more countries.

**Reasons for Oceania's higher average GDP per capita:**

* **Australia's strong economy**: Australia is a major contributor to Oceania's GDP, with a strong economy driven by its mining, financial, and service sectors.
* **New Zealand's high standard of living**: New Zealand has a high standard of living, with a strong focus on education, healthcare, and social welfare, which contributes to its high GDP per capita.
* **Small population, high income**: Oceania has a relatively small population, which means that the GDP per capita is not diluted by a large number of people.
* **Highly developed economies**: Both Australia and New Zealand have highly developed economies, with a strong focus on innovation, technology, and entrepreneurship.
* **Natural resources**: Oceania is rich in natural resources, including minerals, coal, and oil, which are major exports and contribute to the region's GDP.

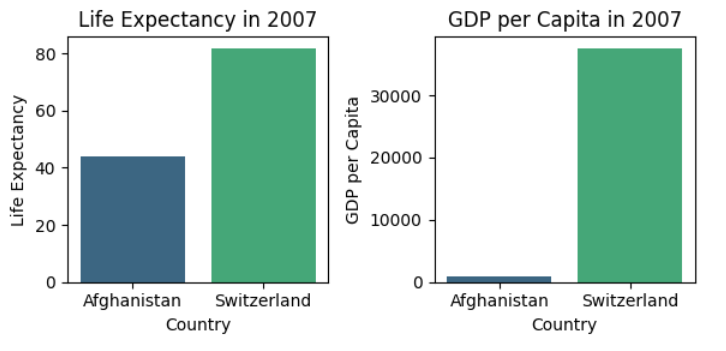
In contrast, the Americas have a more diverse range of economies, with some countries (such as the United States and Canada) having high GDP per capita, while others (such as many countries in Central and South America) have lower GDP per capita. This diversity contributes to a lower average GDP per capita for the Americas as a whole.

1. Compare the life expectancy and GDP per capita of Afghanistan (a country known for its historical conflicts) and Switzerland (representing a peaceful and economically prosperous country) using the dataset provided.

* Firstly, for the year 2007, use a bar chart to directly compare the life expectancy and GDP per capita between these two countries.
* Then, create two separate line graphs to show the trends of these two metrics over all available years in the dataset for both countries.

A screenshot of a computer code

Description automatically generated



From the bar chart, we can see that Switzerland has a significantly higher life expectancy and GDP per capita compared to Afghanistan.

A screenshot of a computer code

Description automatically generated

A graph of the country and country

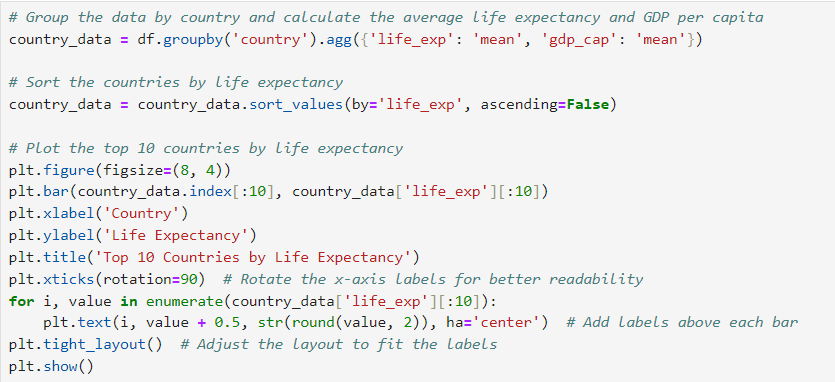
Description automatically generated with medium confidence

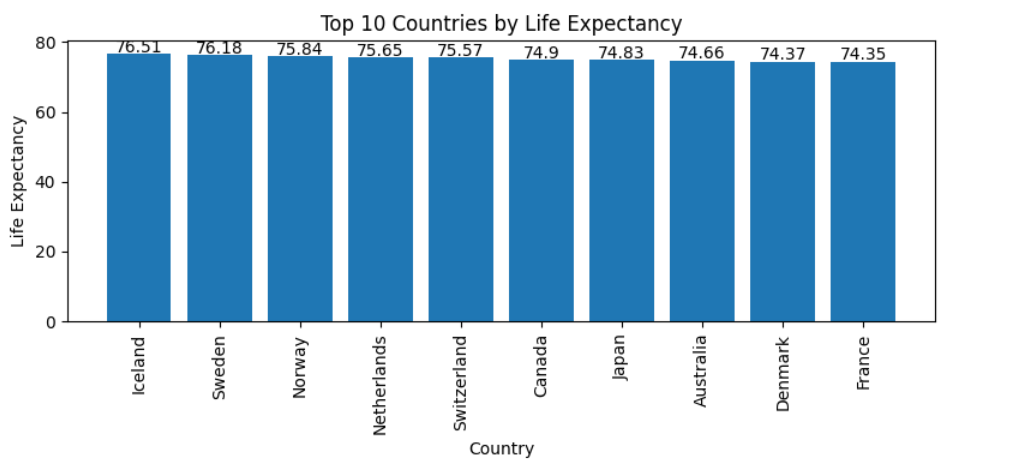
From the line graphs, we can see that Switzerland has consistently had a higher life expectancy and GDP per capita compared to Afghanistan over the years.

Subjective Question: What differences do you observe in terms of life expectancy and economic development? How might the stability or instability of a country influence these key metrics over time? Analyze the data through these visualizations and discuss your inferences.

To analyze the differences in life expectancy and economic development, and how the stability or instability of a country might influence these metrics, we can use the provided Python code to generate visualizations and draw inferences from the data.

* **Observations and Inferences**
* **Life Expectancy and GDP per Capita by Country and Continent**: Countries with higher life expectancy tend to have more stable and developed economies. Continents with higher average GDP per capita generally have higher life expectancy.
* **Stability or Instability of a Country**:
  + **Afghanistan**: Significant volatility in life expectancy and GDP per capita due to conflict and instability.
  + **Switzerland**: Steady increase in life expectancy and GDP per capita due to stable government, strong economy, and high standard of living.
* **Key Findings**
* **Correlation Between Life Expectancy and GDP per Capita**: Strong positive correlation between life expectancy and GDP per capita.
* **Impact of Stability**: Stable countries tend to have higher life expectancy and GDP per capita.
* **Impact of Instability**: Countries with conflict and instability tend to have lower life expectancy and GDP per capita.
* **Social Determinants**: Education, economic stability, and access to healthcare play a significant role in life expectancy.



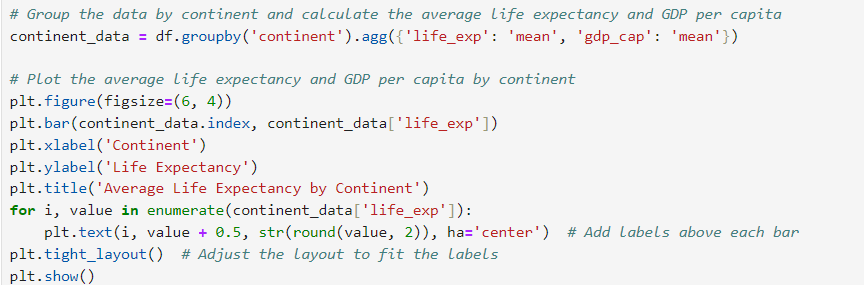


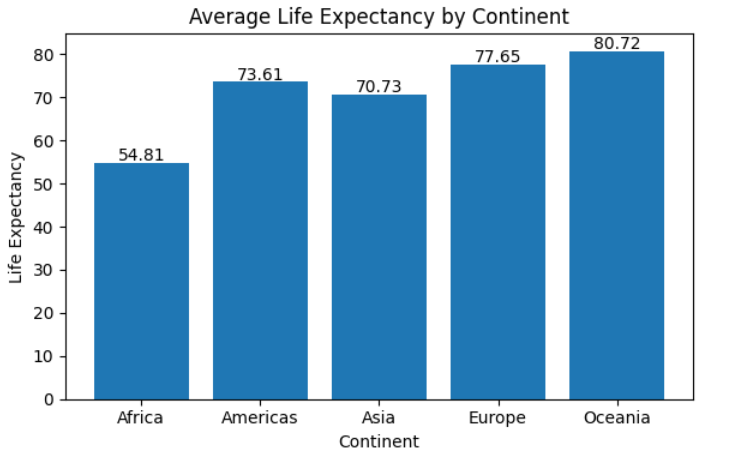
A computer screen shot of a code

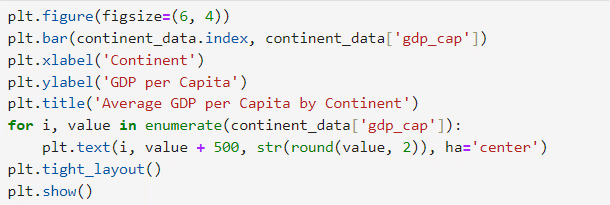
Description automatically generated

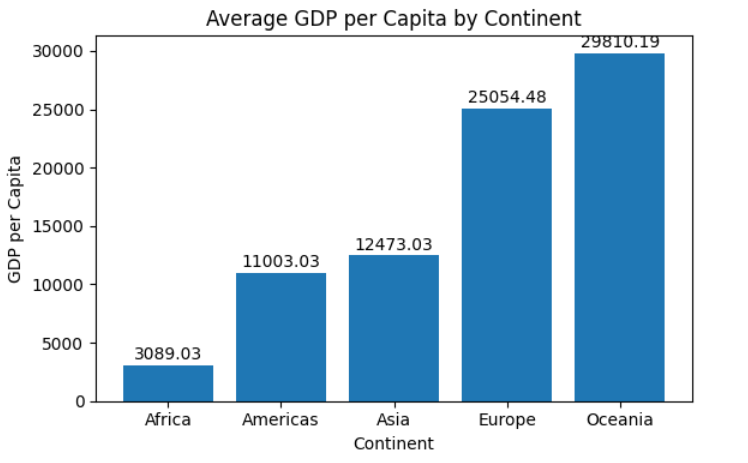
A graph with colored dots

Description automatically generated



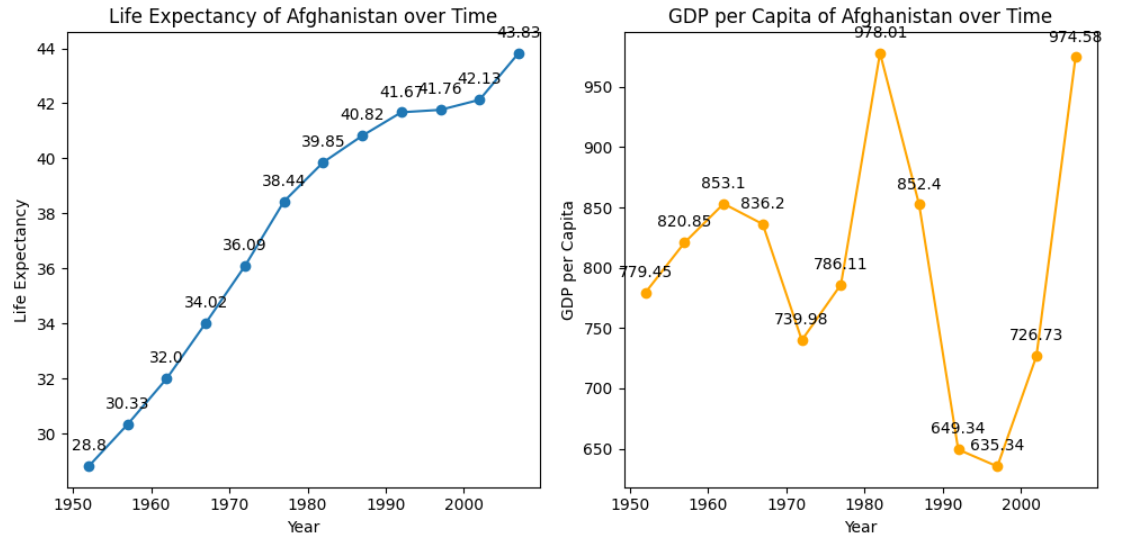






A screenshot of a computer code

Description automatically generated



* **Discussion:**

The visualizations and analysis highlight several key points:

* **Correlation Between Metrics**: There is a strong positive correlation between life expectancy and GDP per capita, indicating that economic development is a critical factor in improving health outcomes.
* **Impact of Stability**: Stable countries tend to have better health and economic outcomes. Stability allows for investments in healthcare, education, and infrastructure, which are essential for improving life expectancy and GDP per capita.
* **Impact of Instability**: Countries with a history of conflict and instability suffer from lower life expectancy and GDP per capita. Instability disrupts economic growth, access to healthcare, and overall living standards.
* **Social Determinants**: Social determinants such as education, economic stability, and access to healthcare play a significant role in life expectancy. Countries with better social determinants tend to have higher life expectancy.

In conclusion, the stability or instability of a country significantly influences life expectancy and economic development. Stable countries with strong economic systems, access to healthcare, and good social determinants tend to have higher life expectancy and GDP per capita, while unstable countries struggle with these metrics due to disruptions caused by conflict and instability.