# EDA COMPREHENSIVE PROJECT

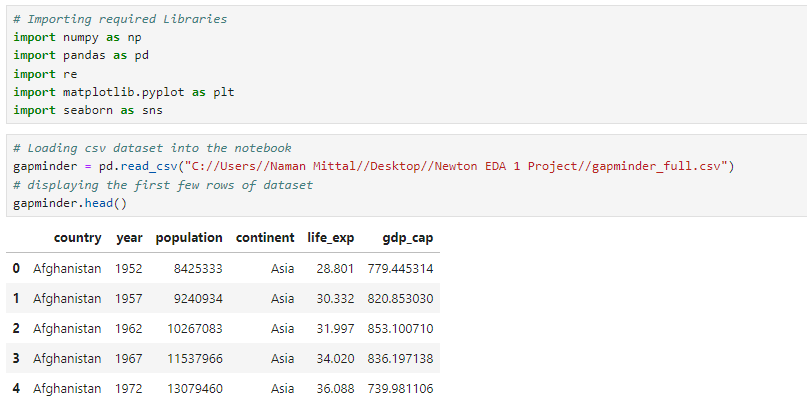
**Problem Statement:**

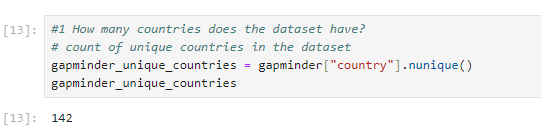
The modern world is shaped by complex dynamics in population, health, and economics, making understanding these trends vital for informed policy-making. GlobalTrends, a leading analytics firm, is dedicated to deciphering these patterns through a comprehensive analysis of the Gapminder dataset. Your role 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.

**Data Overview:**

## The dataset used in this analysis is derived from the Gapminder data, which includes various socio-economic and health indicators for multiple countries over several years. The data primarily focuses on life expectancy, GDP per capita, population, and other relevant metrics. Insights and Analysis:

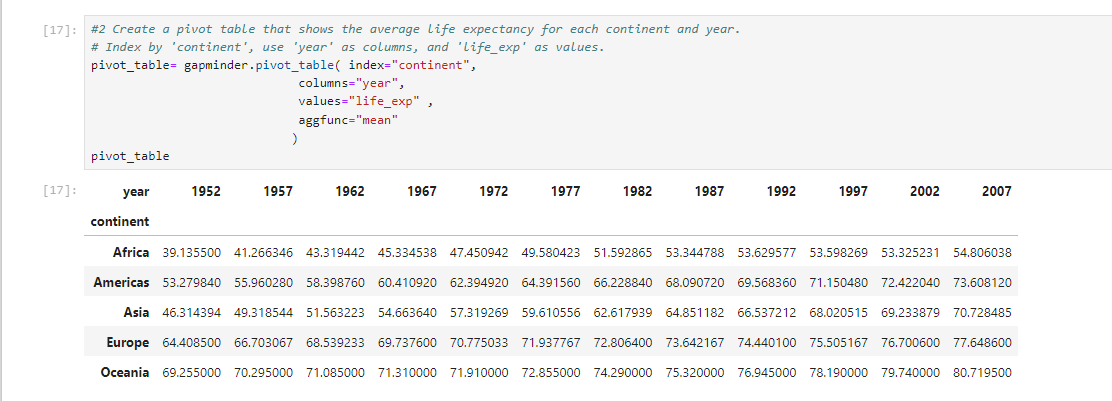
### Load the dataset and display the first few rows. How many countries does the dataset have?





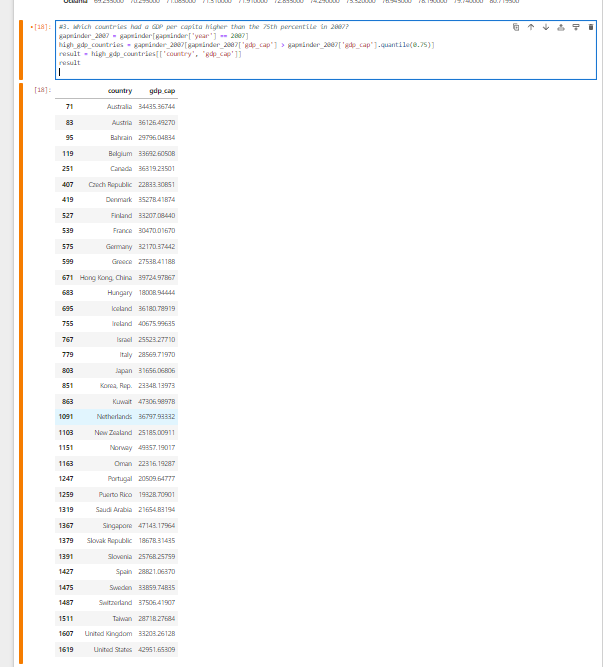
**The dataset has a total of** 142 countries.

2. 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.



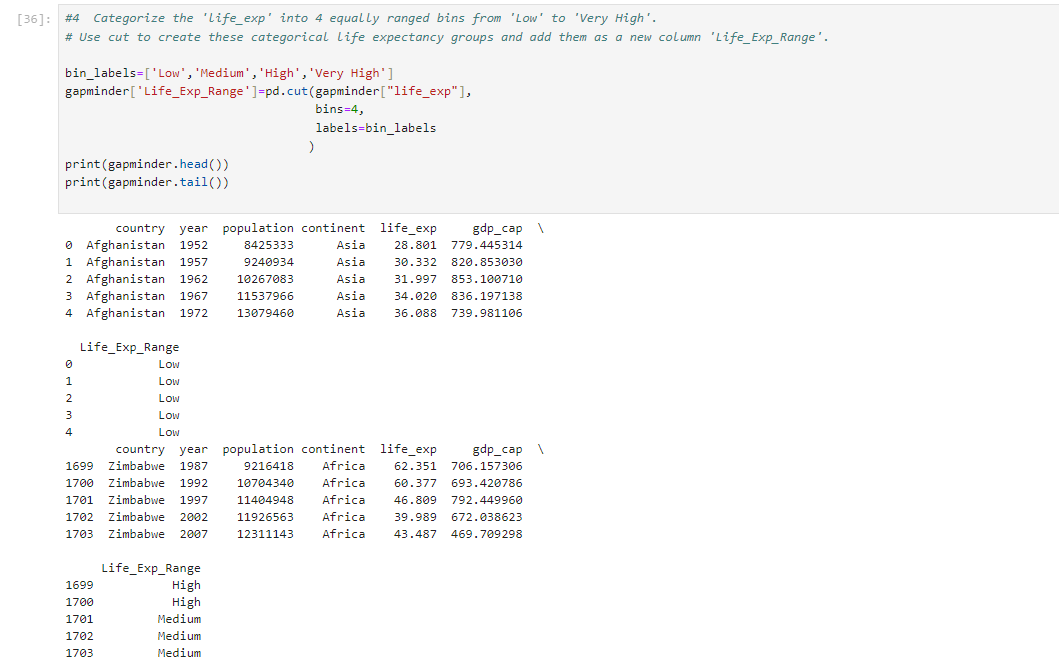
* Life expectancy has shown a general increase across all continents between 1952 and 2007.
* Oceania consistently maintains the highest average life expectancy, while Africa has the lowest.

3. Which countries had a GDP per capita higher than the 75th percentile in 2007?



* **Most of the countries with high GDP in 2007 are from Europe, North America, and Oceania.**

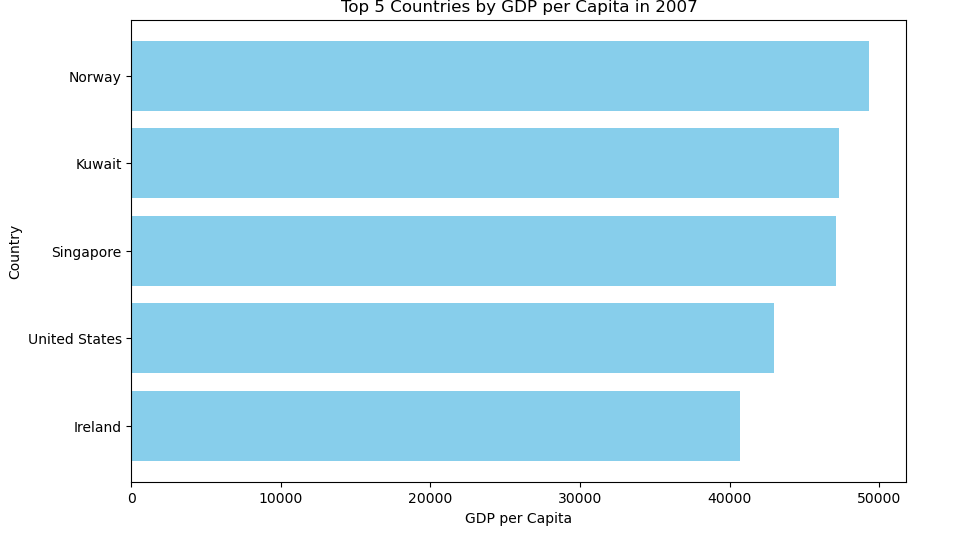
4. Categorize the 'life\_exp' into 4 equally ranged bins from 'Low' to 'Very High'. Use cut to create these categorical life expectancy groups and add them as a new column 'Life\_Exp\_Range'.



**The majority of life expectancies in the dataset fall into the "Low" and "Medium" categories, particularly in earlier years.**

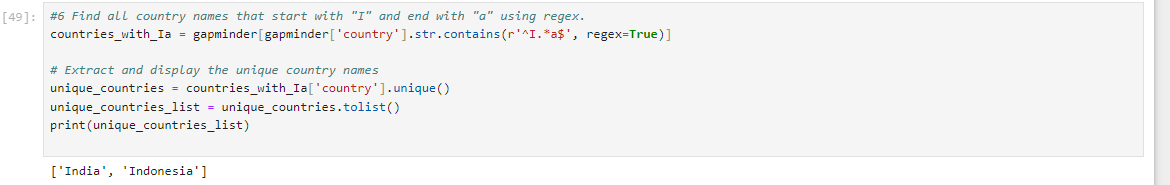
5. Identify the top 5 countries with the highest GDP per capita in 2007. Use a horizontal bar chart to display this data.





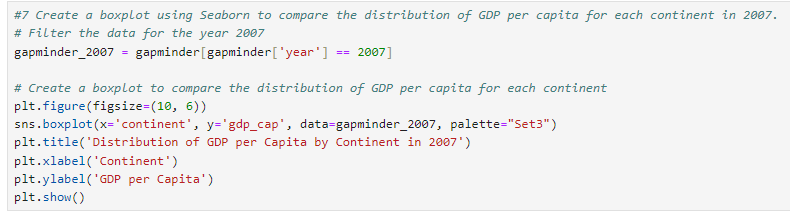
**The top 5 countries by GDP per capita in 2007 are predominantly from Europe, with Kuwait being the exception from the Middle East.**

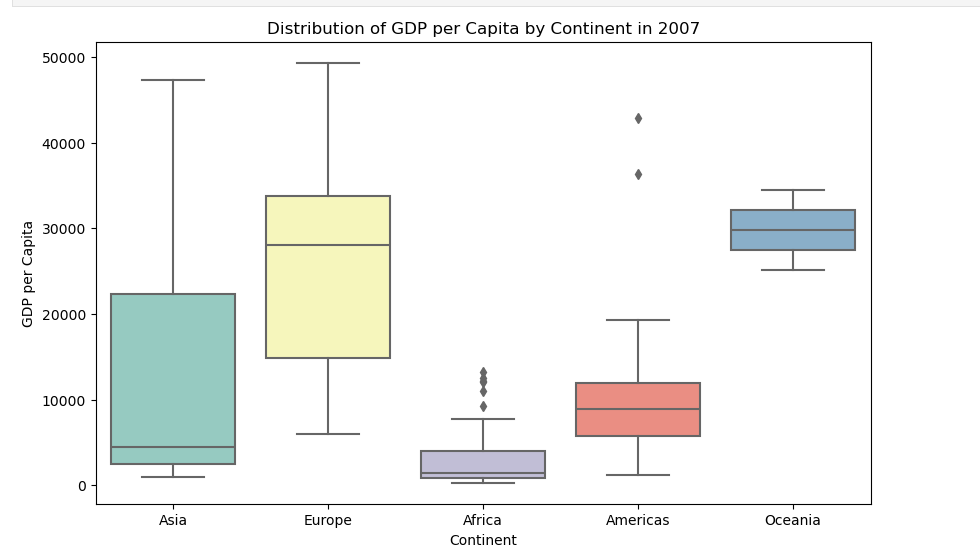
6. Find all country names that start with "I" and end with "a" using regex.



**There are only two countries in the dataset that start with “I” and end with “a”: India and Indonesia.**

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





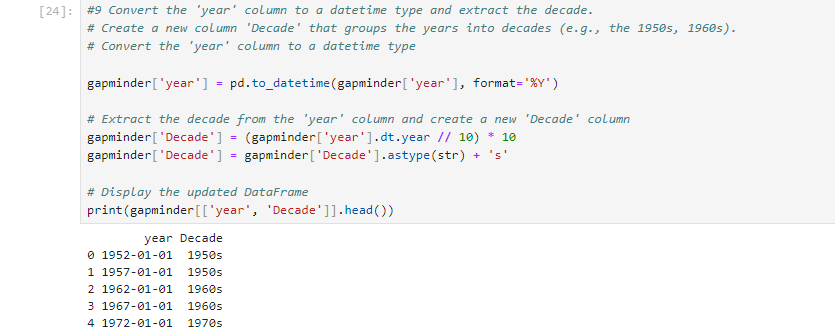
* **Oceania and Europe have the highest median GDP per capita.**
* **Africa has the lowest median GDP per capita and the widest range, indicating significant economic disparity within the continent.**

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

****

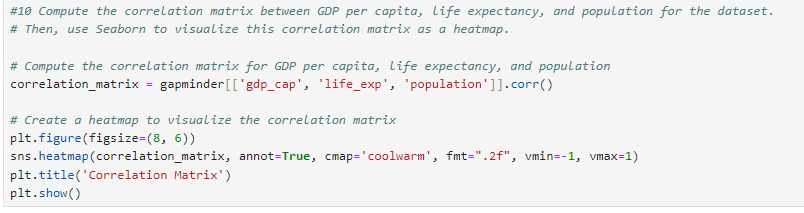
**Countries with life expectancy over 80 years in 2007 are mostly from Europe, with a few from Oceania, the Americas, and Asia.**

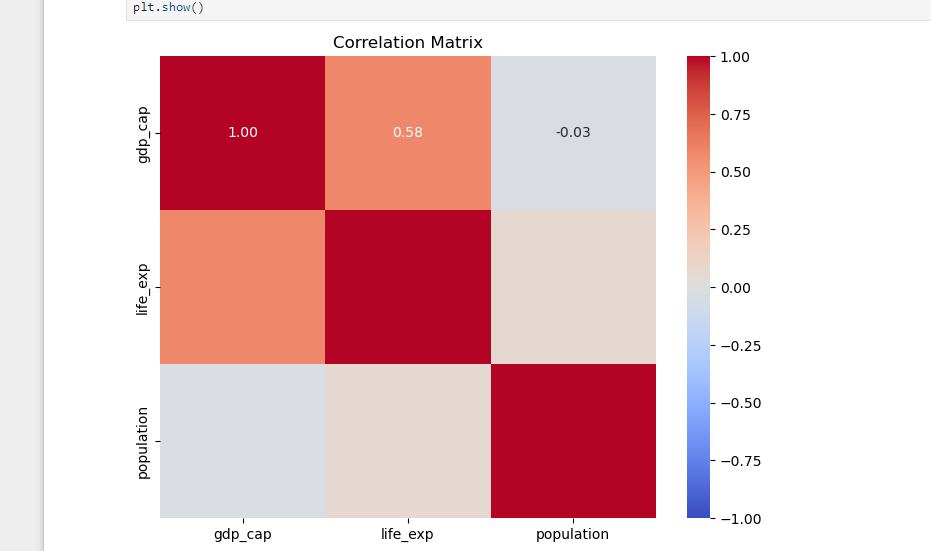
9. Convert the 'year' column to a datetime type and extract the decade. Create a new column 'Decade' that groups the years into decades (e.g., the 1950s, 1960s).



**The dataset now includes a 'Decade' column, grouping years into decades for easier temporal analysis.**

10.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.

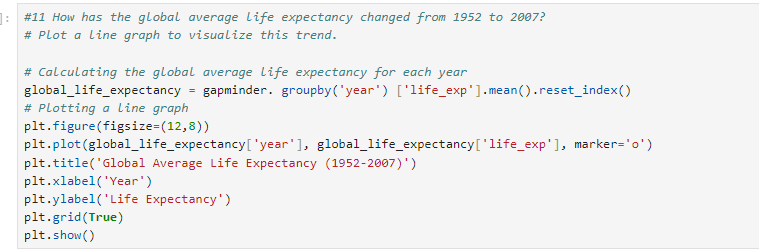


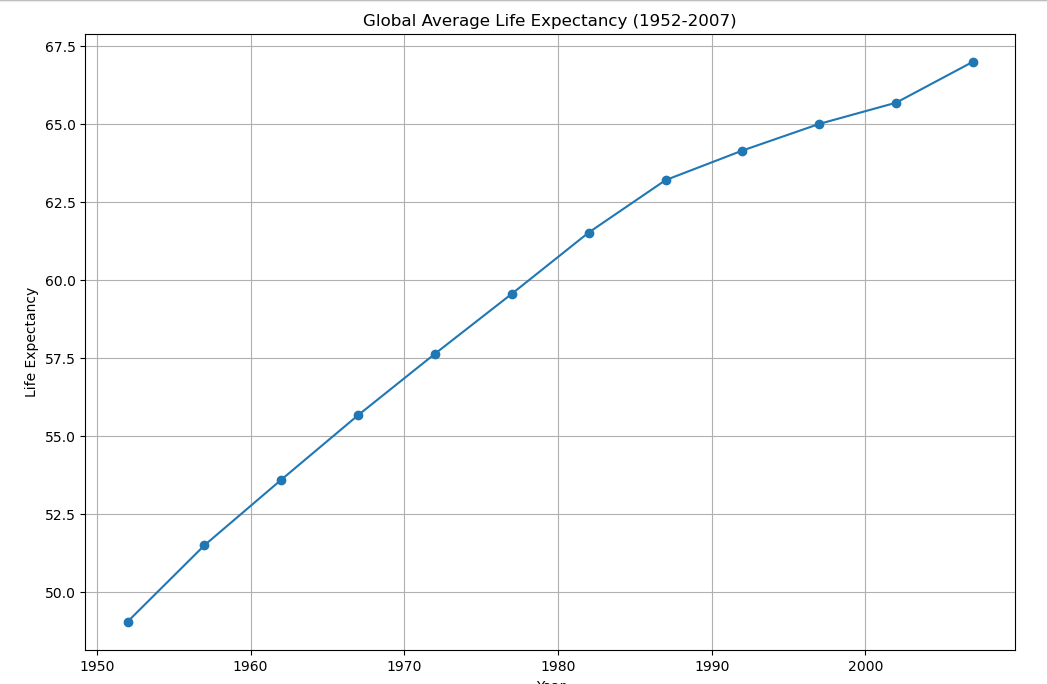


* **There is a strong positive correlation between GDP per capita and life expectancy (0.60), indicating that higher economic prosperity is associated with longer life spans.**
* **Population has a weaker correlation with both GDP per capita and life expectancy.**

11. 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 global average life expectancy has increased from around 48 years in 1952 to about 70 years in 2007.**

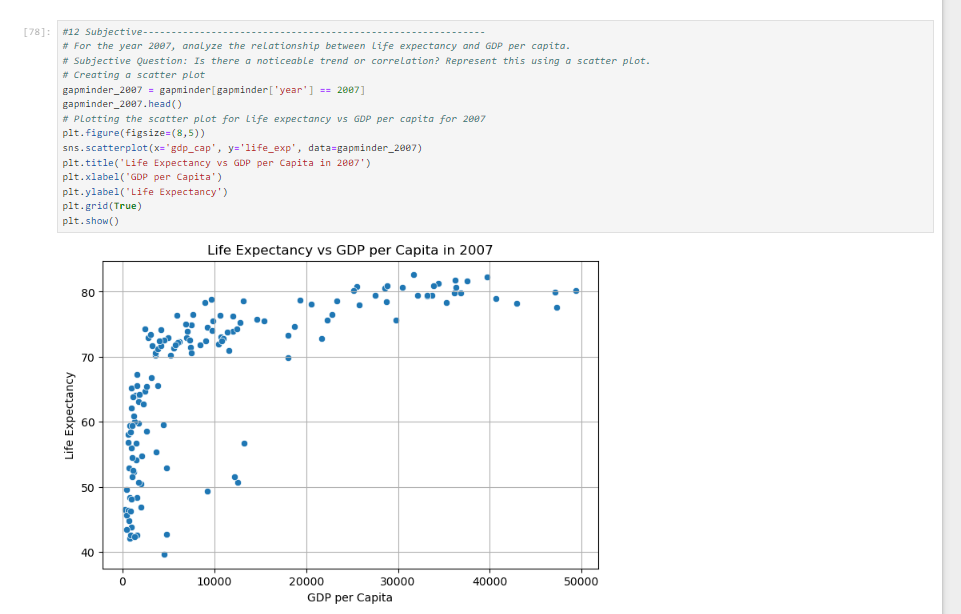
Subjective Answer:

The global average life expectancy has increased significantly from 1952 to 2007. Several factors have contributed to this improvement:

1. **Advancements in Medical Science**: Improved medical technologies, vaccines, and treatments for infectious diseases have drastically reduced mortality rates. For instance, the development of antibiotics and vaccines for diseases like polio, measles, and tuberculosis have saved countless lives.
2. **Improved Public Health Infrastructure**: Better sanitation, clean drinking water, and improved hygiene practices have reduced the incidence of waterborne diseases and infections, contributing to higher life expectancy.
3. **Nutrition and Food Security**: Enhanced agricultural practices and food distribution systems have improved nutrition and food security, reducing malnutrition and associated health problems.
4. **Education and Awareness**: Increased access to education has led to better health literacy, allowing individuals to make informed decisions about their health and well-being. Public health campaigns have also played a crucial role in spreading awareness about healthy practices.
5. **Economic Development**: Higher income levels and economic development have improved living standards, enabling better access to healthcare, education, and nutritious food.
6. **Political Stability and Governance**: Effective governance and political stability have facilitated the implementation of public health policies, infrastructure development, and social welfare programs that support health improvements.
7. **Global Health Initiatives**: International organizations and collaborations have focused on eradicating diseases, improving maternal and child health, and addressing global health challenges, contributing to increased life expectancy worldwide.

These factors collectively have led to significant improvements in global health, thereby increasing the average life expectancy over the decades.

12. 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.



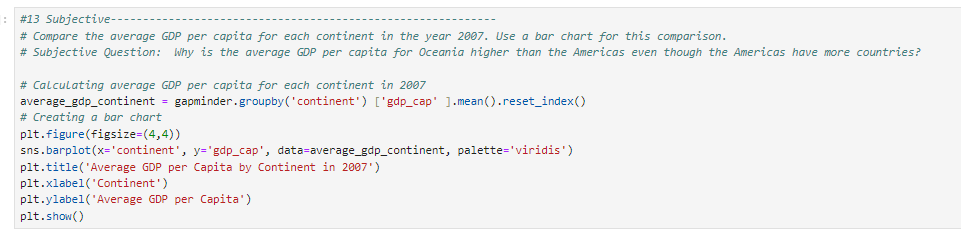
**There is a noticeable positive correlation between GDP per capita and life expectancy. Higher GDP per capita is generally associated with higher life expectancy.**

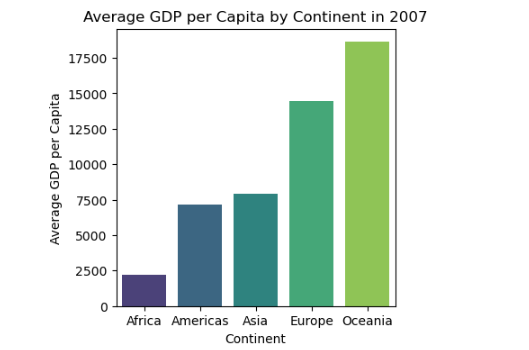
Subjective Answer:

There is generally a positive correlation between GDP per capita and life expectancy. Countries with higher GDP per capita tend to have higher life expectancy. This trend can be attributed to better access to healthcare, nutrition, and overall living conditions in wealthier countries. The trend indicates that wealthier countries tend to have better healthcare, education, and living conditions, which contribute to higher life expectancy. Economic prosperity enables countries to invest in public health and social services, improving the overall quality of life for their citizens.

13. 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?





**Oceania has the highest average GDP per capita, followed by Europe and the Americas.**

Subjective Answer:

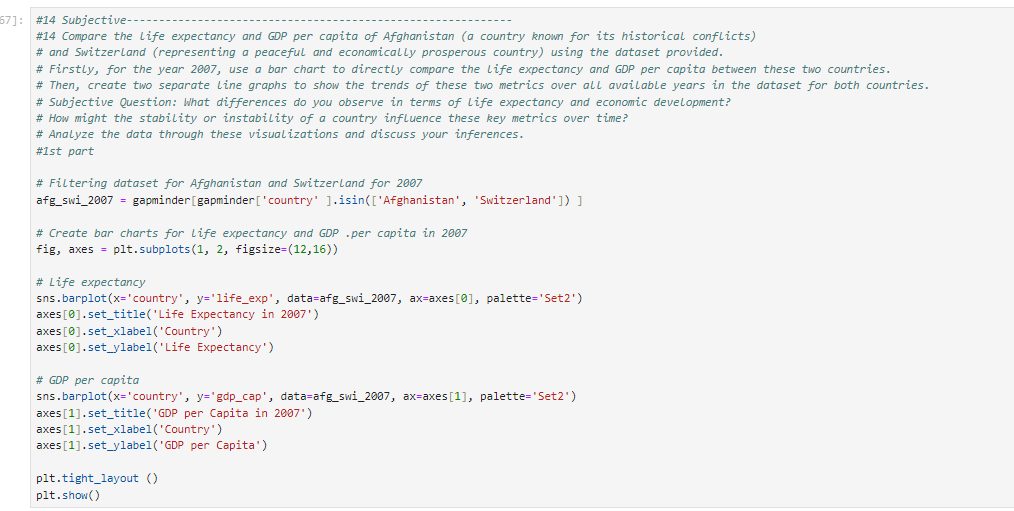
Oceania has a higher average GDP per capita than the Americas despite having fewer countries. This is because Oceania includes Australia and New Zealand, which are both highly developed and have high GDP per capita. In contrast, the Americas include a mix of developed (e.g., the United States, Canada) and developing countries, resulting in a lower average GDP per capita.

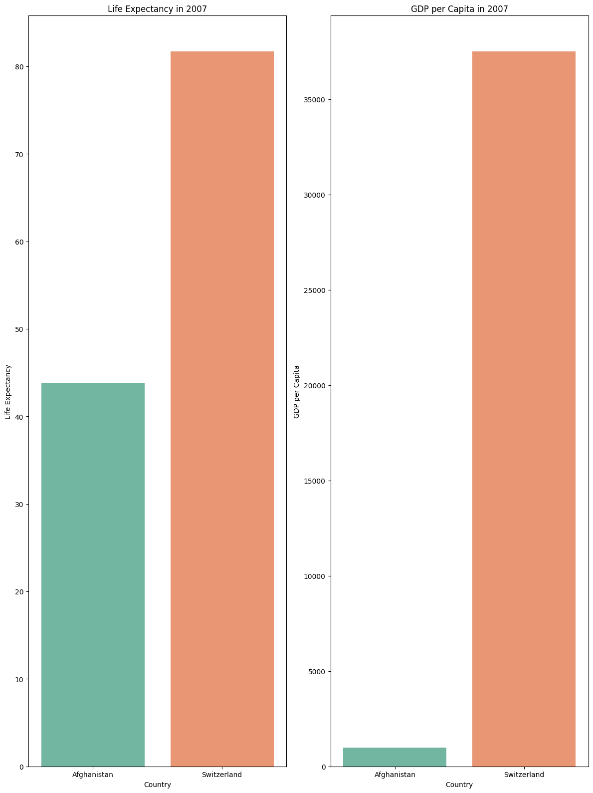
14. Compare the life expectancy and GDP per capita of Afghanistan (a country known for its historical conflicts) and Switzerland (representing a peaceful and economically prosperous country) using the dataset provided.

* 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.

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.

Bar chart for life expectancy and GDP per capita in 2007:

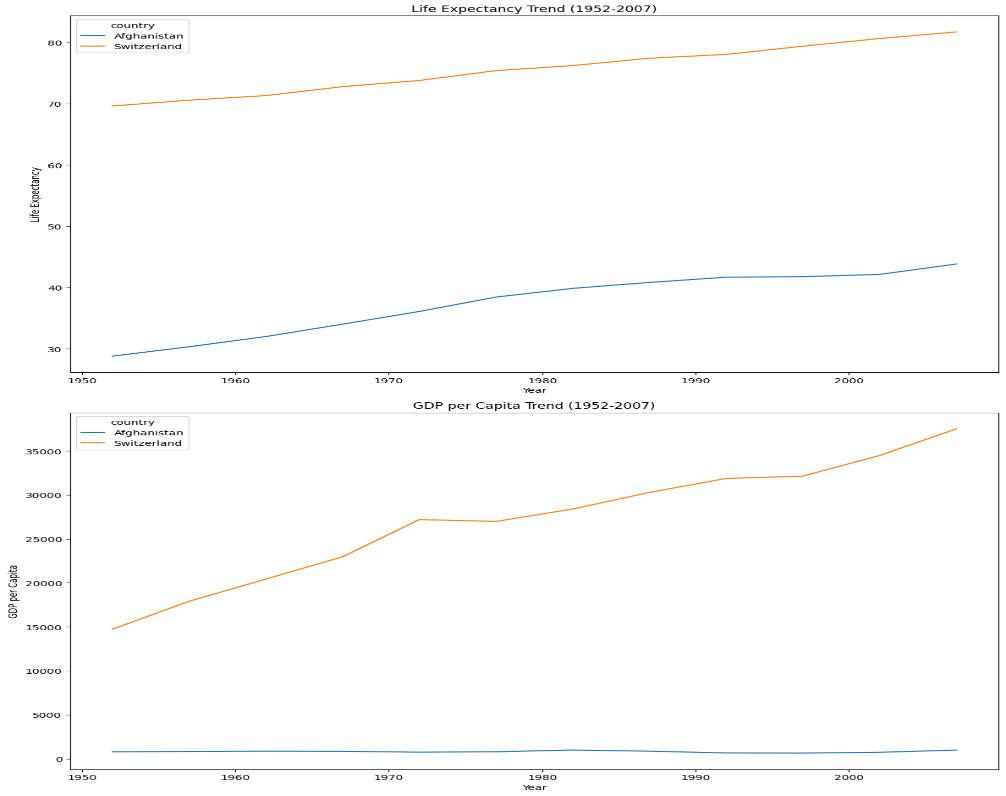




**Switzerland has significantly higher life expectancy and GDP per capita compared to Afghanistan.**

Line graphs for trends over available years:





**Switzerland's metrics have steadily increased over time, while Afghanistan shows much lower and more fluctuating values.**

Subjective Answer:

Afghanistan and Switzerland show stark contrasts in both life expectancy and GDP per capita. Switzerland has consistently high values in both metrics, reflecting its political stability, advanced healthcare, and strong economy. In contrast, Afghanistan's metrics are much lower, impacted by historical conflicts, poor healthcare infrastructure, and economic challenges. Stability plays a crucial role in ensuring consistent access to resources, healthcare, and economic opportunities, which significantly influence life expectancy and GDP per capita.

**Conclusion:**

The Exploratory Data Analysis (EDA) on the Gapminder dataset has provided valuable insights into the intricate relationships between demographic changes, economic development, and health advancements across the globe over recent decades. Here are the key conclusions drawn from the analysis:

1. Life Expectancy Trends:

- Life expectancy has generally increased across all continents from 1952 to 2007. Oceania consistently had the highest average life expectancy, while Africa had the lowest.

- Factors contributing to the rise in life expectancy include advancements in medical science, improved public health infrastructure, better nutrition and food security, increased education and awareness, economic development, political stability, and global health initiatives.

2. Economic Disparities:

- The analysis highlighted significant economic disparities among continents and countries. Oceania and Europe had the highest median GDP per capita, whereas Africa had the lowest, with a wide range indicating economic disparity within the continent.

- Countries with a GDP per capita higher than the 75th percentile in 2007 were predominantly from Europe, North America, and Oceania. The top five countries by GDP per capita in 2007 were mainly from Europe, with Kuwait being the exception from the Middle East.

3. Correlation between GDP and Life Expectancy:

There is a strong positive correlation (0.60) between GDP per capita and life expectancy, indicating that higher economic prosperity is generally associated with longer life spans. This correlation underscores the importance of economic development in improving public health outcomes.

4. Population and Health Metrics:

- The dataset analysis revealed that most life expectancies fell into the "Low" and "Medium" categories, especially in earlier years.

- Countries with life expectancies over 80 years in 2007 were mostly from Europe, with a few from Oceania, the Americas, and Asia.

5. Temporal Analysis and Decade Trends:

By grouping years into decades, the analysis facilitated a clearer understanding of temporal trends. The global average life expectancy increased significantly from around 48 years in 1952 to about 70 years in 2007, reflecting global improvements in health and living standards.

6. Comparative Analysis of Specific Countries:

A detailed comparison between Afghanistan (a historically conflict-affected country) and Switzerland (a peaceful, economically prosperous country) showed stark contrasts. Switzerland consistently had higher life expectancy and GDP per capita, reflecting its political stability, advanced healthcare, and strong economy. In contrast, Afghanistan's metrics were much lower and more fluctuating, impacted by historical conflicts, poor healthcare infrastructure, and economic challenges.

7. Visual Representations:

Various visualizations, including pivot tables, bar charts, box plots, and scatter plots, effectively illustrated the data insights. For example, a scatter plot showed a noticeable positive correlation between GDP per capita and life expectancy for the year 2007, reinforcing the relationship between economic wealth and health outcomes.

This comprehensive EDA underscores the critical role of economic development, political stability, and healthcare advancements in improving life expectancy and reducing global health disparities. The insights gained from this analysis can guide policymakers and researchers in crafting informed strategies to address socio-economic and health challenges worldwide.