**EDA COMPREHENSIVE PROJECT**

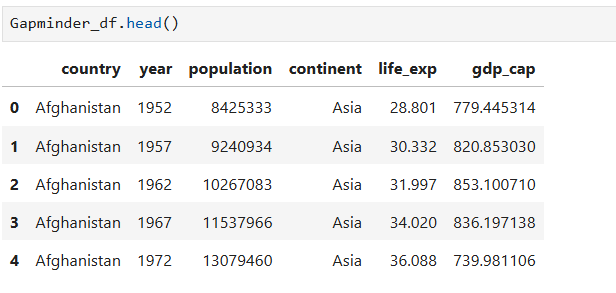
**Problem Statement:**

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.

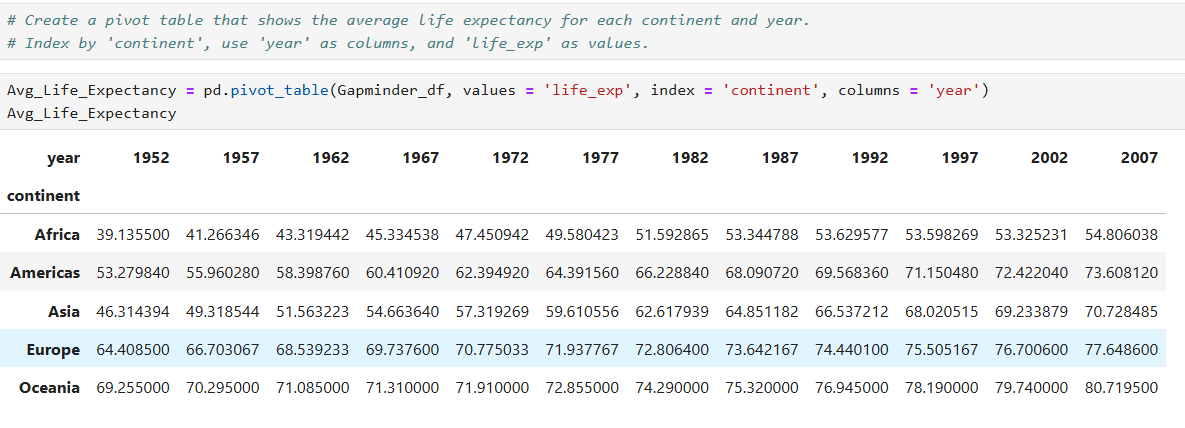
**Dataset:** [Gapminder World](https://drive.google.com/uc?id=1fDGZh86UPUkt2K6enlNQfB0mswU8pB_P)

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

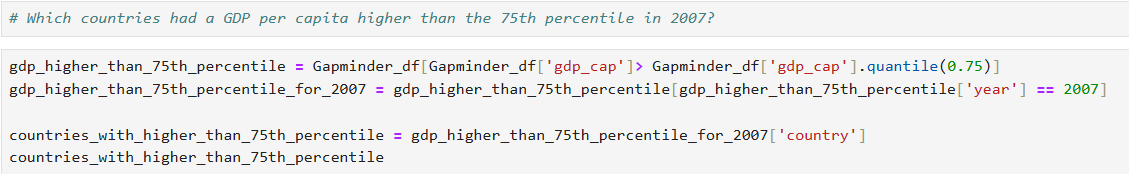


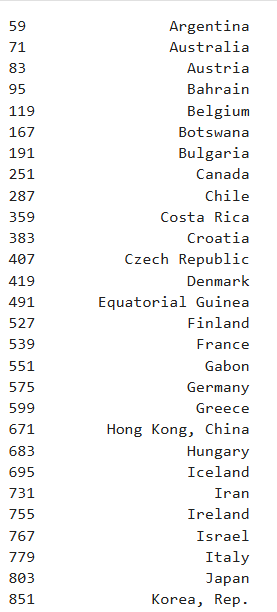


1. **Create a pivot table that shows the average life expectancy for each continent and year. Index by 'continent', use 'year' as columns, and 'life\_exp' as values.**

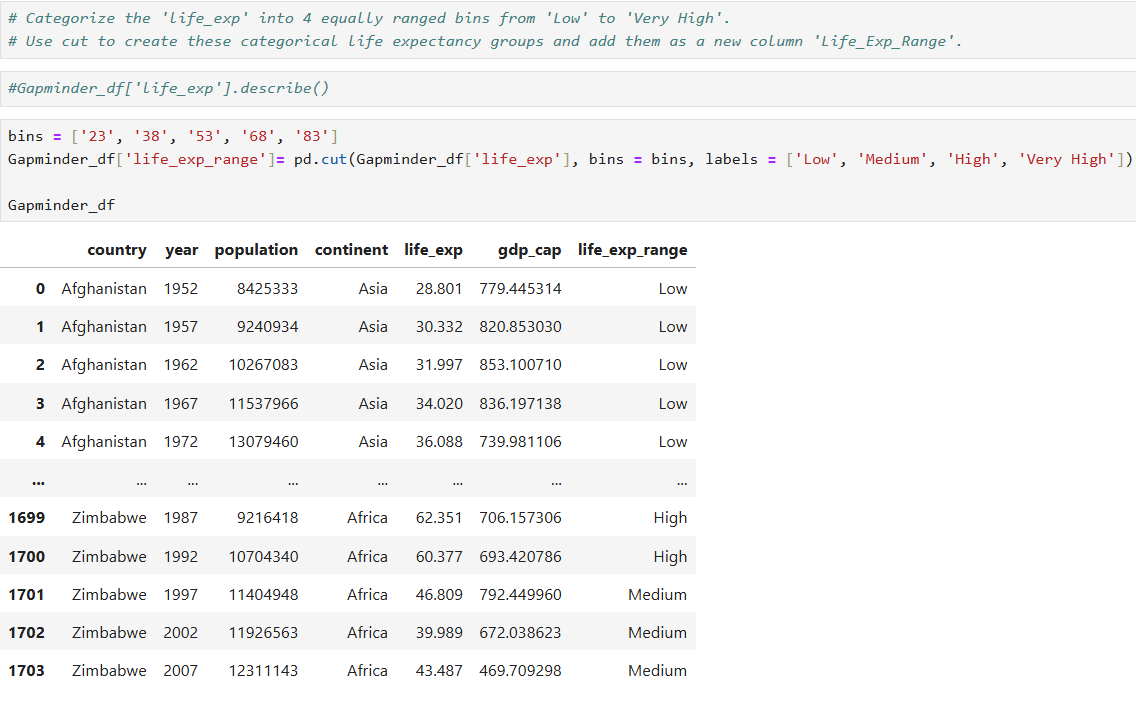


1. **Which countries had a GDP per capita higher than the 75th percentile in 2007?**

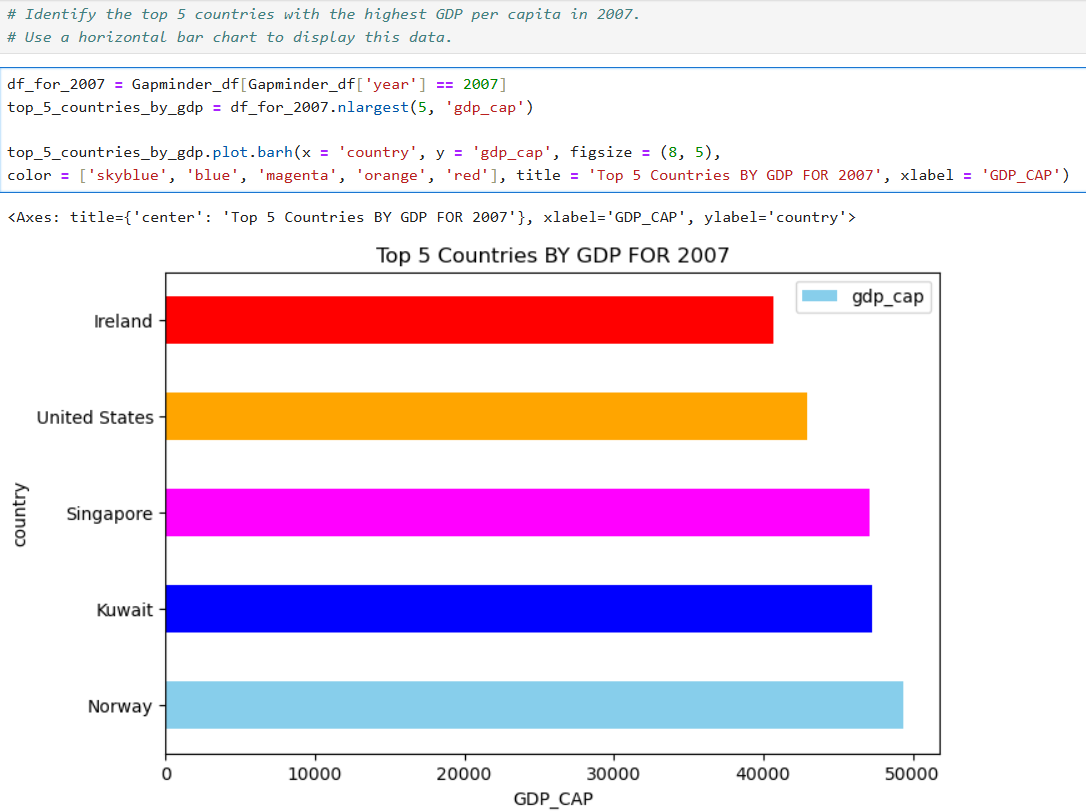


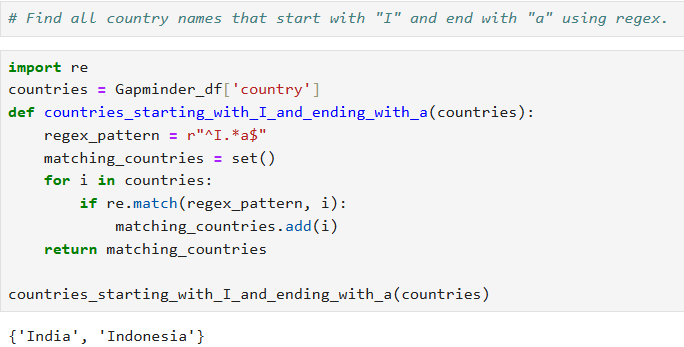
1. **Categorize the 'life\_exp' into 4 equally ranged bins from 'Low' to 'Very High'. Use cut to create these categorical life expectancy groups and add them as a new column 'Life\_Exp\_Range'.**



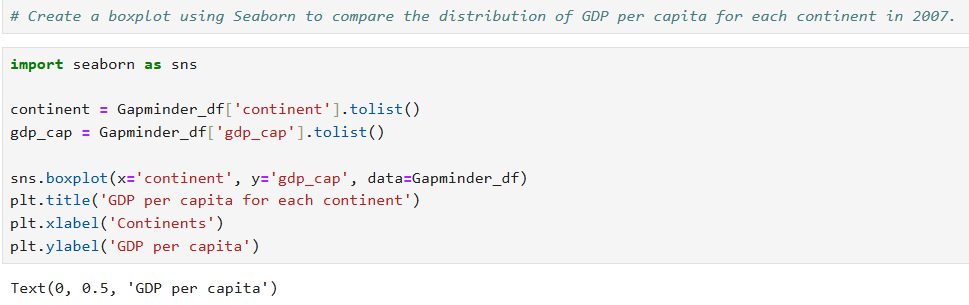
1. **Identify the top 5 countries with the highest GDP per capita in 2007. Use a horizontal bar chart to display this data.**

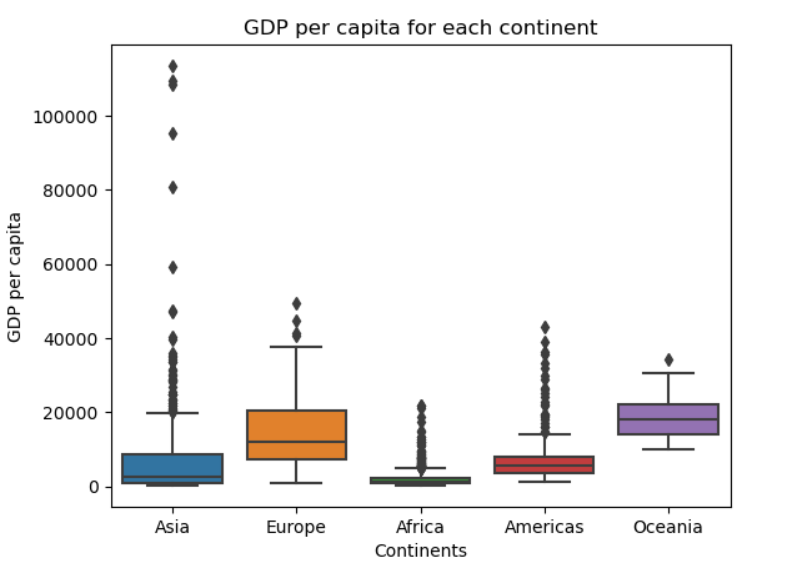


1. **Find all country names that start with "I" and end with "a" using regex.**



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

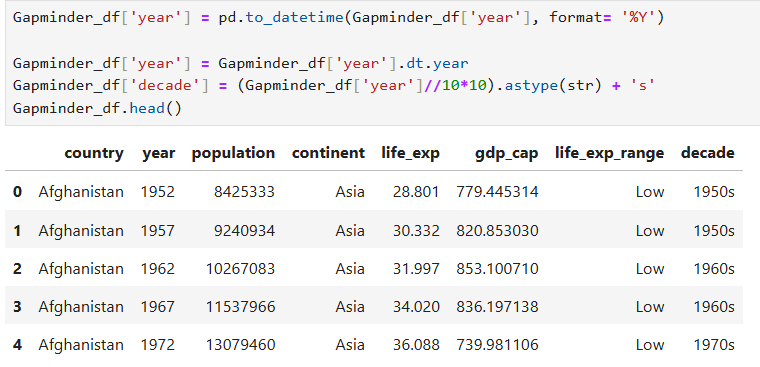




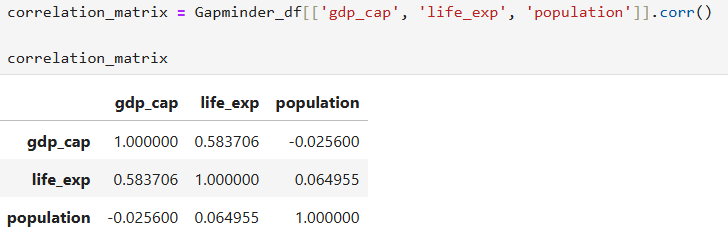
1. **Find all countries with a life expectancy of over 80 years in 2007. List these countries and their respective continents.**



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).**

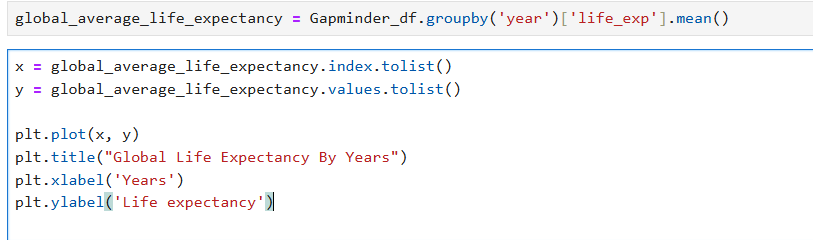


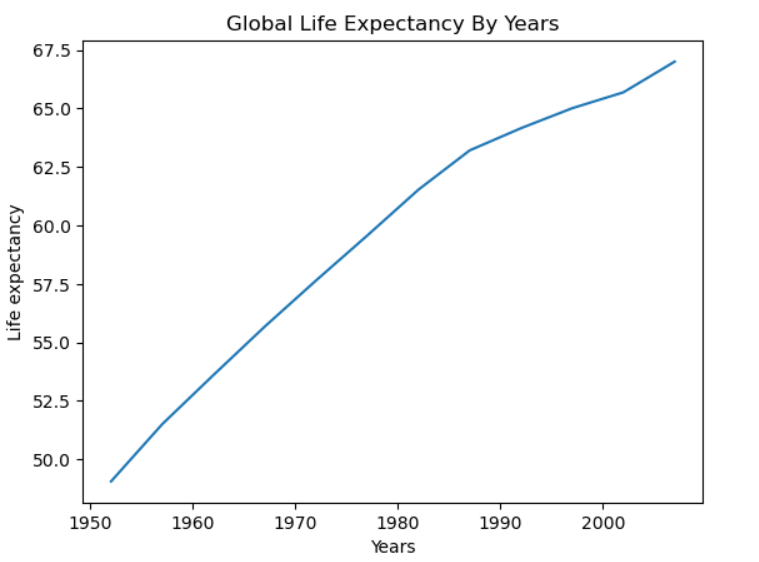
1. **Compute the correlation matrix between GDP per capita, life expectancy, and population for the dataset. Then, use Seaborn to visualize this correlation matrix as a heatmap.**





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





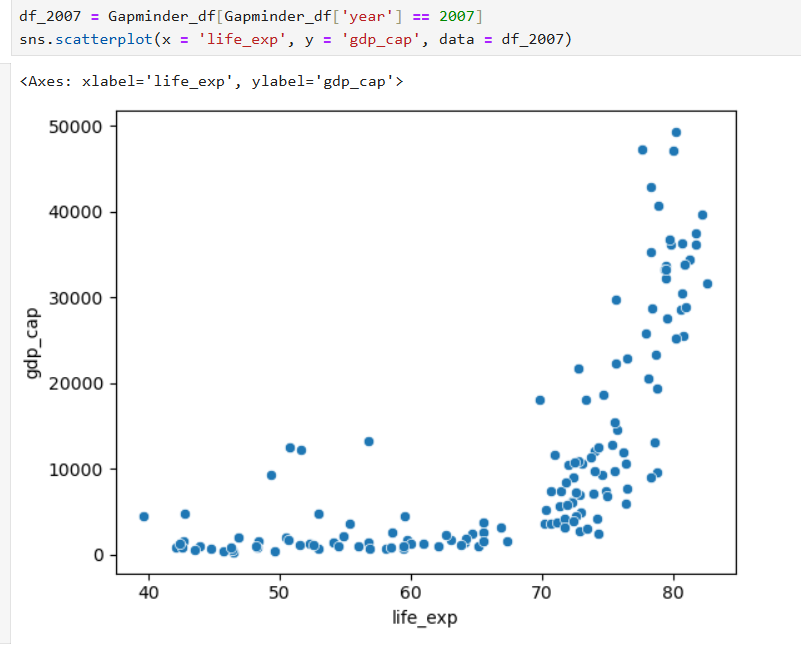
**Subjective Question: Discuss the various reasons that could have contributed to the change.**

**Several factors have contributed to the significant increase in global life expectancy from 1952 to 2007:**

1. **Medical Advancements:**
   * **Vaccinations and Disease Control:** The widespread availability of vaccines has significantly reduced the mortality rate from infectious diseases such as polio, measles, and smallpox.
   * **Antibiotics and Treatments:** The discovery and distribution of antibiotics like penicillin, as well as advances in medical treatments for various diseases, have decreased mortality rates and improved overall health.
2. **Improved Public Health Infrastructure:**
   * **Sanitation and Clean Water:** Improvements in public sanitation, access to clean water, and the development of sewage systems have reduced the spread of waterborne diseases.
   * **Healthcare Access:** Increased access to healthcare, including maternal and child health services, has contributed to the decline in infant and maternal mortality rates.
3. **Economic Growth and Improved Living Standards:**
   * **Nutrition:** Better economic conditions have led to improved nutrition and food security, which are crucial for enhancing health and life expectancy.
   * **Housing and Living Conditions:** Improved housing and living conditions have reduced exposure to environmental hazards and improved overall health.
4. **Education and Awareness:**
   * **Health Education:** Increased levels of education, particularly regarding health practices, have led to better hygiene, disease prevention, and the adoption of healthier lifestyles.
   * **Family Planning:** Education about and access to family planning services have contributed to smaller family sizes and better health outcomes for mothers and children.
5. **Global Initiatives and International Aid:**
   * **Global Health Programs:** International efforts, such as those led by the World Health Organization (WHO) and various NGOs, have focused on eradicating diseases, improving healthcare infrastructure, and addressing public health issues worldwide.
   * **Humanitarian Aid:** Humanitarian aid has played a significant role in improving health outcomes in regions affected by poverty, conflict, and natural disasters.
6. **Technological Innovations:**
   * **Medical Technology:** Advances in medical technology, such as imaging, diagnostics, and surgical techniques, have improved the ability to diagnose and treat diseases early and effectively.
   * **Communication and Information Technology:** Improved communication and information technology have facilitated the rapid dissemination of health information, best practices, and innovations across the globe.
7. **Political Stability and Peace:**
   * **Reduced Conflicts:** In regions where political stability has improved, life expectancy has increased due to the reduction in conflicts, leading to safer environments and better healthcare provision.
   * **Global Cooperation:** International cooperation on health issues has led to more effective responses to global health challenges, such as pandemics and chronic diseases.

**In summary, the combination of medical, economic, educational, technological, and political factors has driven the increase in global life expectancy from 1952 to 2007. These improvements reflect a collective effort to enhance the quality of life and health outcomes across the world.**

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.**

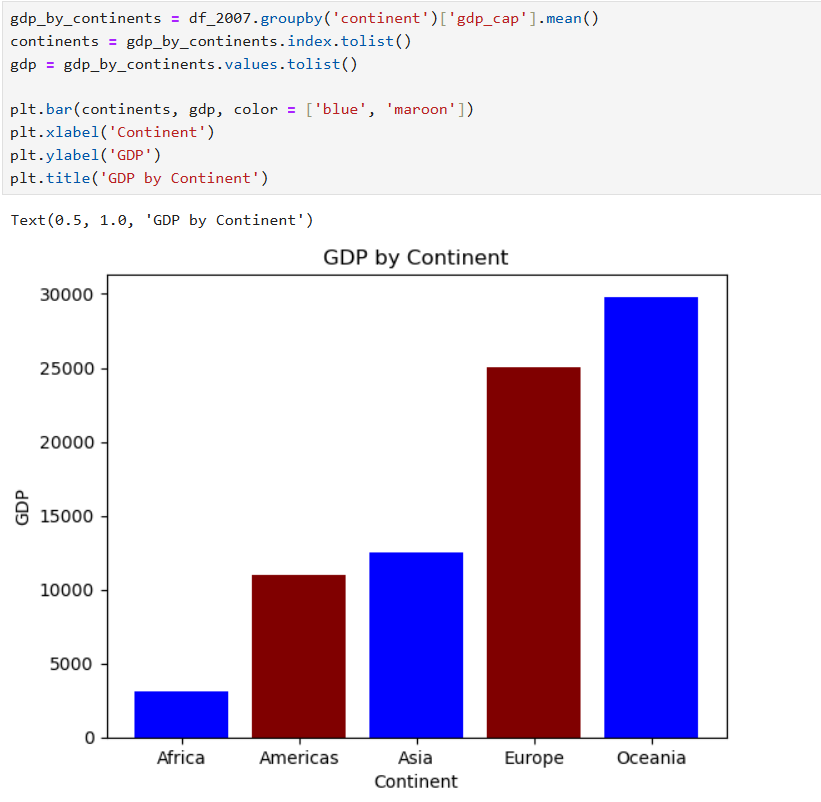
The scatter plot represents the relationship between life expectancy (life\_exp) and GDP per capita (gdp\_cap) for various countries in the year 2007. The following observations can be made:

1. **Positive Correlation:**
   * There is a noticeable positive correlation between life expectancy and GDP per capita. As GDP per capita increases, life expectancy tends to increase as well. This indicates that countries with higher economic prosperity generally have longer life expectancies.
2. **Non-Linear Relationship:**
   * The relationship between the two variables is not perfectly linear. At lower levels of GDP per capita, life expectancy varies more widely. As GDP per capita increases, the variation in life expectancy decreases, and most countries tend to cluster around higher life expectancies.
3. **Clusters of Countries:**
   * There appears to be a cluster of countries with low GDP per capita and lower life expectancy. These countries are likely developing nations where economic challenges and lower standards of living impact health and longevity.
   * Another cluster is observed at the higher end of GDP per capita, where countries typically have life expectancies above 70 years. These are likely developed nations with robust healthcare systems and higher standards of living.
4. **Outliers:**
   * A few outliers might be present in the scatter plot, where countries have a high GDP per capita but relatively lower life expectancy. These could be countries where specific factors (such as high levels of inequality, healthcare issues, or other socioeconomic factors) impact life expectancy despite economic prosperity.

**Conclusion:**

The scatter plot clearly shows that there is a positive correlation between life expectancy and GDP per capita. While the relationship is strong, it is not strictly linear, indicating that other factors may also influence life expectancy. Economic prosperity generally contributes to better healthcare, nutrition, and living conditions, which in turn, increase life expectancy. However, outliers and variations at lower GDP levels suggest that economic factors alone do not entirely determine life expectancy, and other social, political, and environmental factors may also play significant roles.

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?**

The average GDP per capita of a continent is calculated by taking the mean GDP per capita of all countries within that continent. Several factors could contribute to the higher average GDP per capita for Oceania compared to the Americas:

1. **Small Number of Countries in Oceania:**
   * Oceania consists of a small number of countries, with Australia and New Zealand being the most economically dominant. These countries have very high GDP per capita, which significantly raises the average for the continent.
2. **Economic Prosperity in Australia and New Zealand:**
   * Australia and New Zealand are highly developed nations with robust economies, high living standards, and advanced industries. Their GDP per capita is considerably higher than most countries in the Americas, particularly when compared to countries in Latin America and the Caribbean, which have lower GDP per capita.
3. **Diversity of Economies in the Americas:**
   * The Americas include a wide range of economies, from highly developed countries like the United States and Canada to developing nations in Central and South America. The inclusion of many countries with lower GDP per capita in Latin America dilutes the average GDP per capita for the continent.
4. **Population Size:**
   * Population size affects GDP per capita. Countries like the United States have a large population, and even though it has a high total GDP, the per capita value is lower than that of smaller, wealthier countries like Australia and New Zealand.
5. **Resource Distribution and Industry Focus:**
   * Australia and New Zealand have economies that benefit from vast natural resources, advanced technology, and significant investments in education and infrastructure, leading to higher productivity and, consequently, higher GDP per capita.

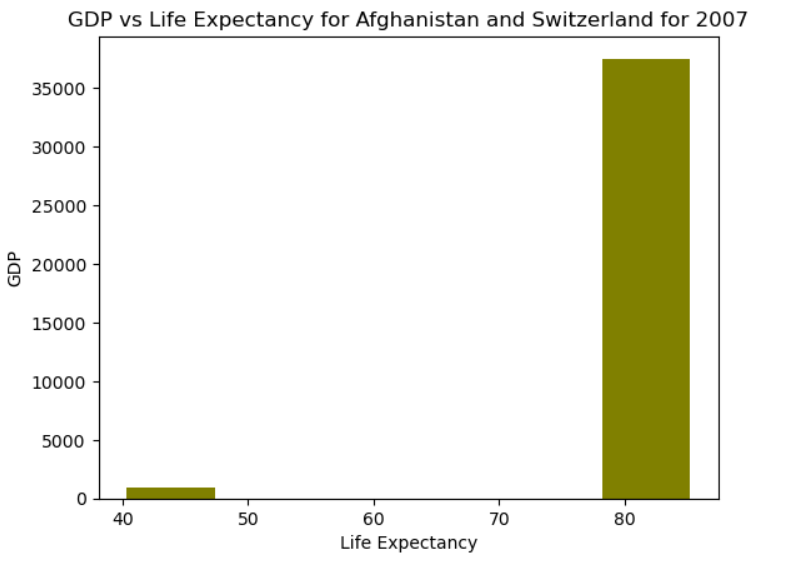
**Conclusion:**

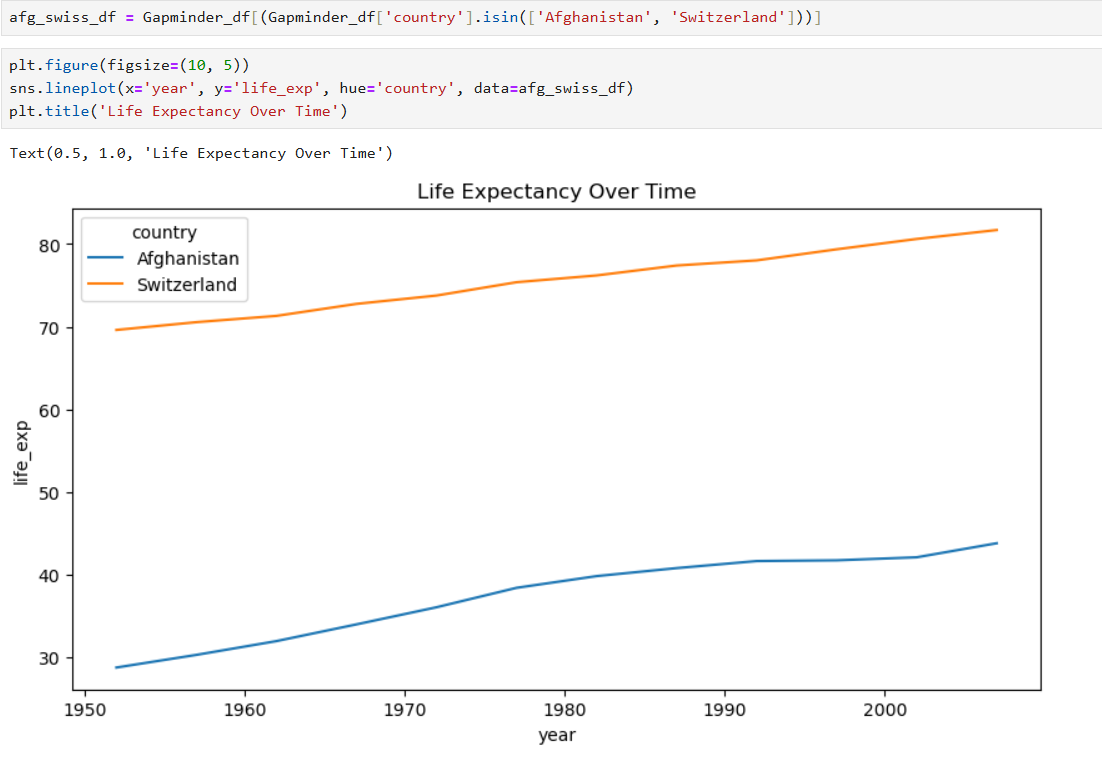
The higher average GDP per capita for Oceania is primarily driven by the strong economic performance of Australia and New Zealand, which are the key contributors to Oceania's average. In contrast, the Americas have a more diverse range of economies, including many countries with lower GDP per capita, which brings down the overall average for the continent. The economic prosperity and smaller number of countries in Oceania result in a higher average GDP per capita despite the Americas having more countries.

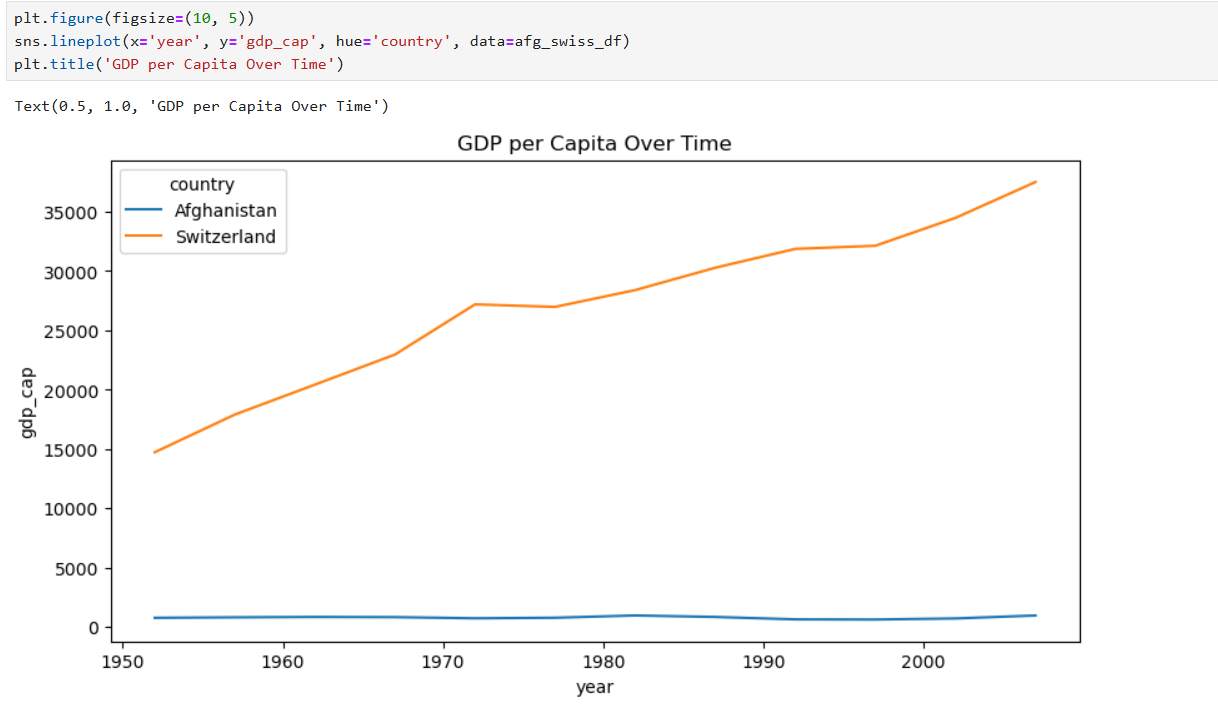
1. **Compare the life expectancy and GDP per capita of Afghanistan (a country known for its historical conflicts) and Switzerland (representing a peaceful and economically prosperous country) using the dataset provided.**

* **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.**









**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.**

**Key Observations:**

**Significant Life Expectancy Disparity:** The bar chart clearly illustrates a dramatic difference in life expectancy between Afghanistan and Switzerland for 2007. Switzerland's life expectancy is substantially higher, indicating a more advanced healthcare system and better living conditions.

**Contrasting Economic Development:** The line charts reveal a stark contrast in GDP per capita over time. Switzerland has experienced consistent economic growth and development, leading to a significantly higher GDP per capita. Afghanistan, on the other hand, has faced challenges in economic stability, resulting in a lower GDP per capita and fluctuations over the years.

**Correlation between Life Expectancy and Economic Development:** The visualizations suggest a strong correlation between life expectancy and economic development. Countries with higher GDP per capita tend to have better healthcare systems, education, and living standards, which contribute to increased life expectancy.

**Impact of Stability and Instability:**

**Switzerland:** Switzerland's relative political stability and economic prosperity have likely played a crucial role in its high life expectancy and economic development. A stable environment fosters investment, education, and healthcare, leading to improved living conditions.

**Afghanistan:** Afghanistan has faced significant political instability and conflict over the decades. These factors have hindered economic development, disrupted healthcare systems, and compromised the safety and well-being of its population, impacting life expectancy.

**Factors:**

**Economic Development:** The disparity in economic development between Afghanistan and Switzerland is evident. Switzerland's consistent growth and higher GDP per capita indicate a more robust economy and better infrastructure.

**Healthcare and Living Standards:** The significant difference in life expectancy suggests that Switzerland has a more developed healthcare system, better access to medical care, and improved living conditions compared to Afghanistan.

**Political Stability:** The political instability experienced by Afghanistan has likely contributed to its economic challenges and lower life expectancy. A more stable environment is essential for fostering development and improving the well-being of the population.

**Conclusion:**

The visualizations highlight the stark contrast between Afghanistan and Switzerland in terms of life expectancy and economic development. Switzerland's stability, economic prosperity, and advanced healthcare system have contributed to a significantly higher quality of life. In contrast, Afghanistan's challenges in these areas have impacted its population's well-being and overall development.