Introduction

This project will investigate if there is a strong correlation between the economic output of a country and the life expectancy of its citizens.

The goals are to prepare data, followed by analysis with plots, and seek to explain the findings from the study.

Here are a few questions that this project will seek to answer:

- Has life expectancy increased over time in the six nations?
- What is the average life expectancy in these nations?
- Has GDP increased over time in the six nations?
- Is there a correlation between GDP and life expectancy of a country?

Data sources

- GDP Source: World Bank national accounts data, and OECD National Accounts data files.
- Life expectancy Data Source: World Health Organization

Exploring the Data

In [ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('all_data.csv')
print(df.info())
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96 entries, 0 to 95
Data columns (total 4 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   Country                         96 non-null     object 
 1   Year                            96 non-null     int64  
 2   Life expectancy at birth (years)  96 non-null     float64
 3   GDP                             96 non-null     float64
dtypes: float64(2), int64(1), object(1)
memory usage: 3.1+ KB
None
```

Out[ ]:

	Country	Year	Life expectancy at birth (years)	GDP
0	Chile	2000	77.3	7.786093e+10
1	Chile	2001	77.3	7.097992e+10
2	Chile	2002	77.8	6.973681e+10
3	Chile	2003	77.9	7.564346e+10
4	Chile	2004	78.0	9.921039e+10

In [ ]:
```
countries = df.Country.unique()
print(countries)
```

```
['Chile' 'China' 'Germany' 'Mexico' 'United States of America' 'Zimbabwe']
```

In [ ]:
```
print(df.Year.unique())
```

```
[2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013
 2014 2015]
```

In [ ]:
```
df = df.rename({"Life expectancy at birth (years)":"LEABY"}, axis = "columns")
df.head()
```

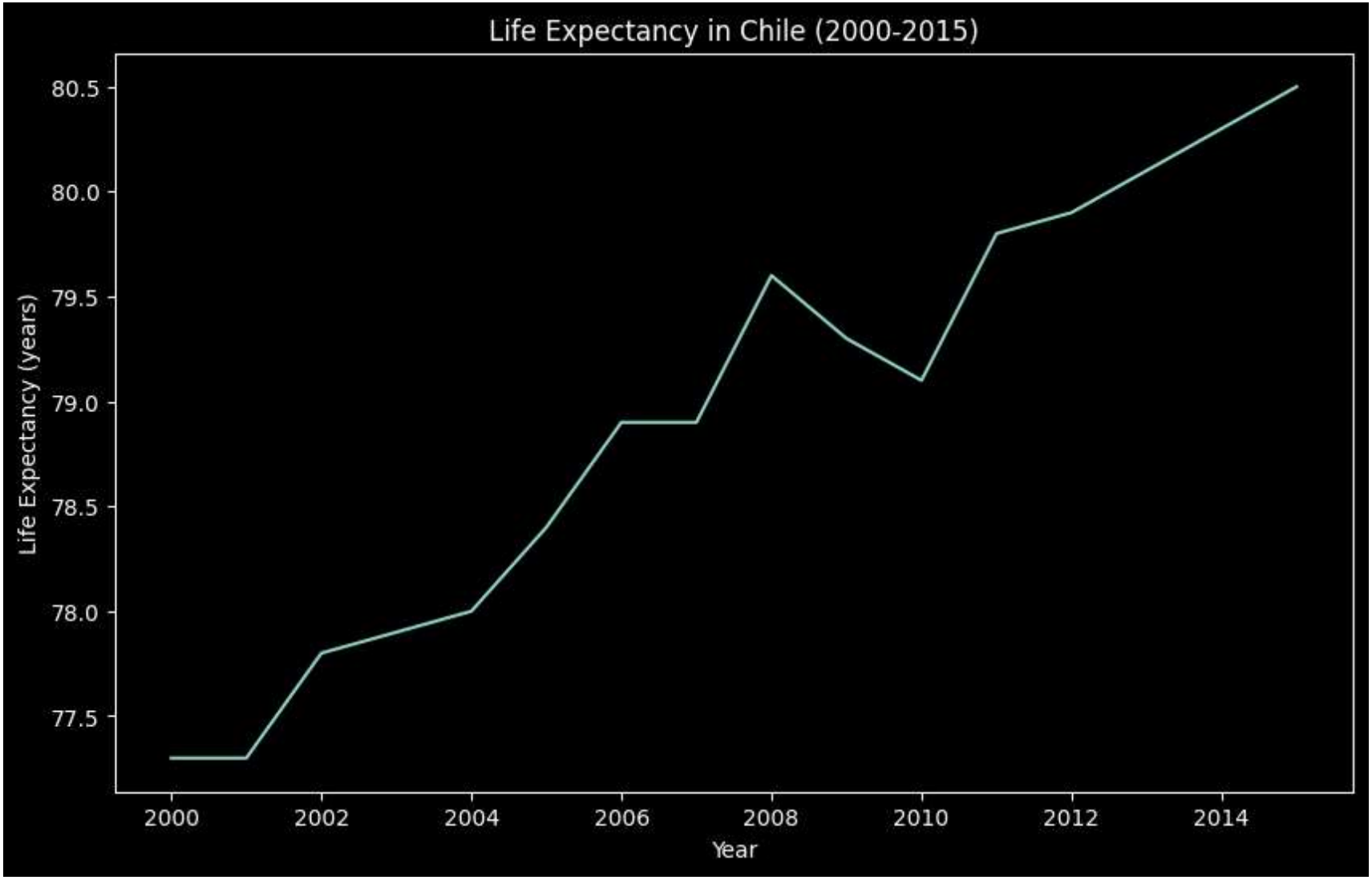
Out[ ]:

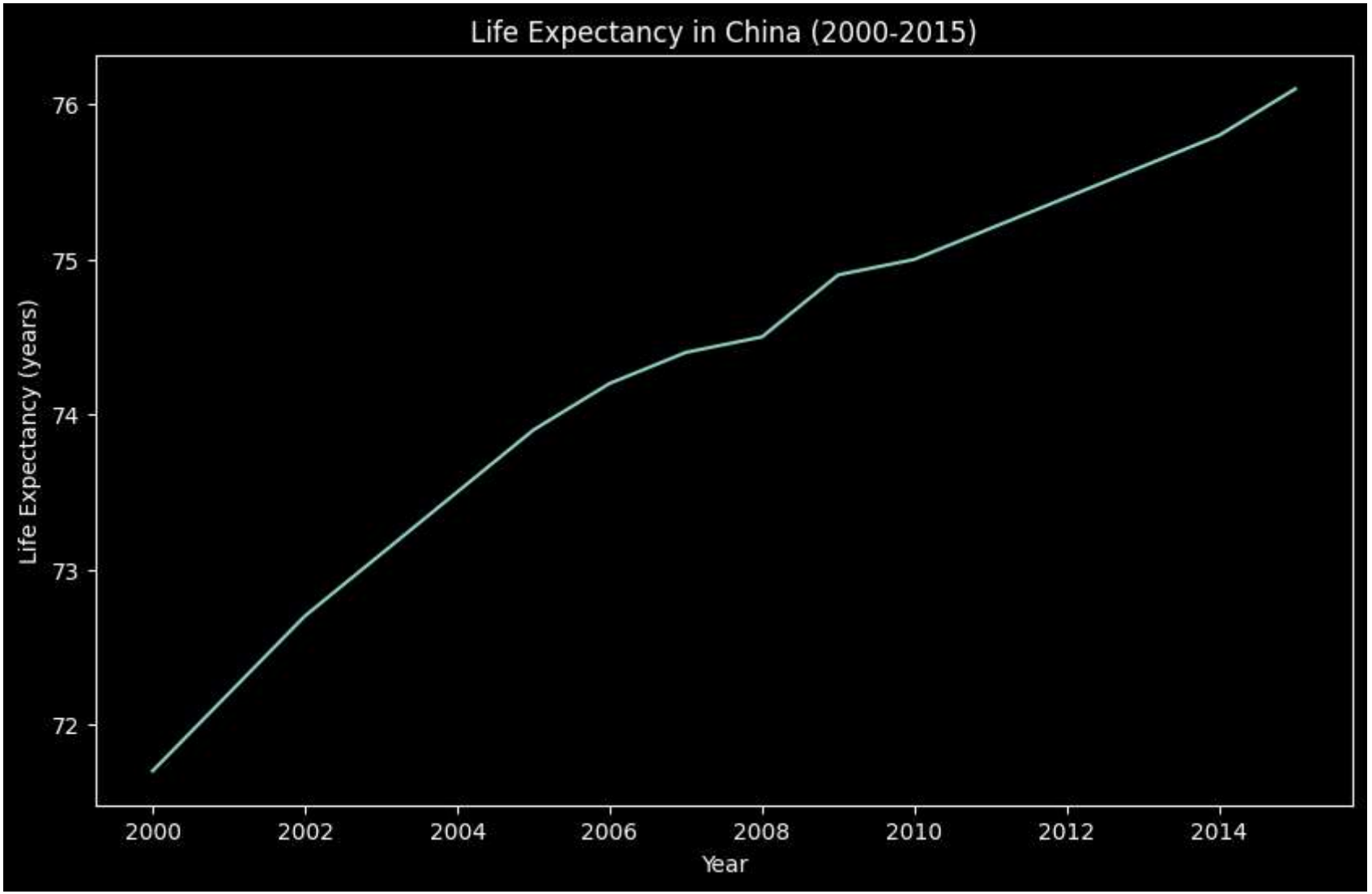
	Country	Year	LEABY	GDP
0	Chile	2000	77.3	7.786093e+10
1	Chile	2001	77.3	7.097992e+10
2	Chile	2002	77.8	6.973681e+10
3	Chile	2003	77.9	7.564346e+10
4	Chile	2004	78.0	9.921039e+10

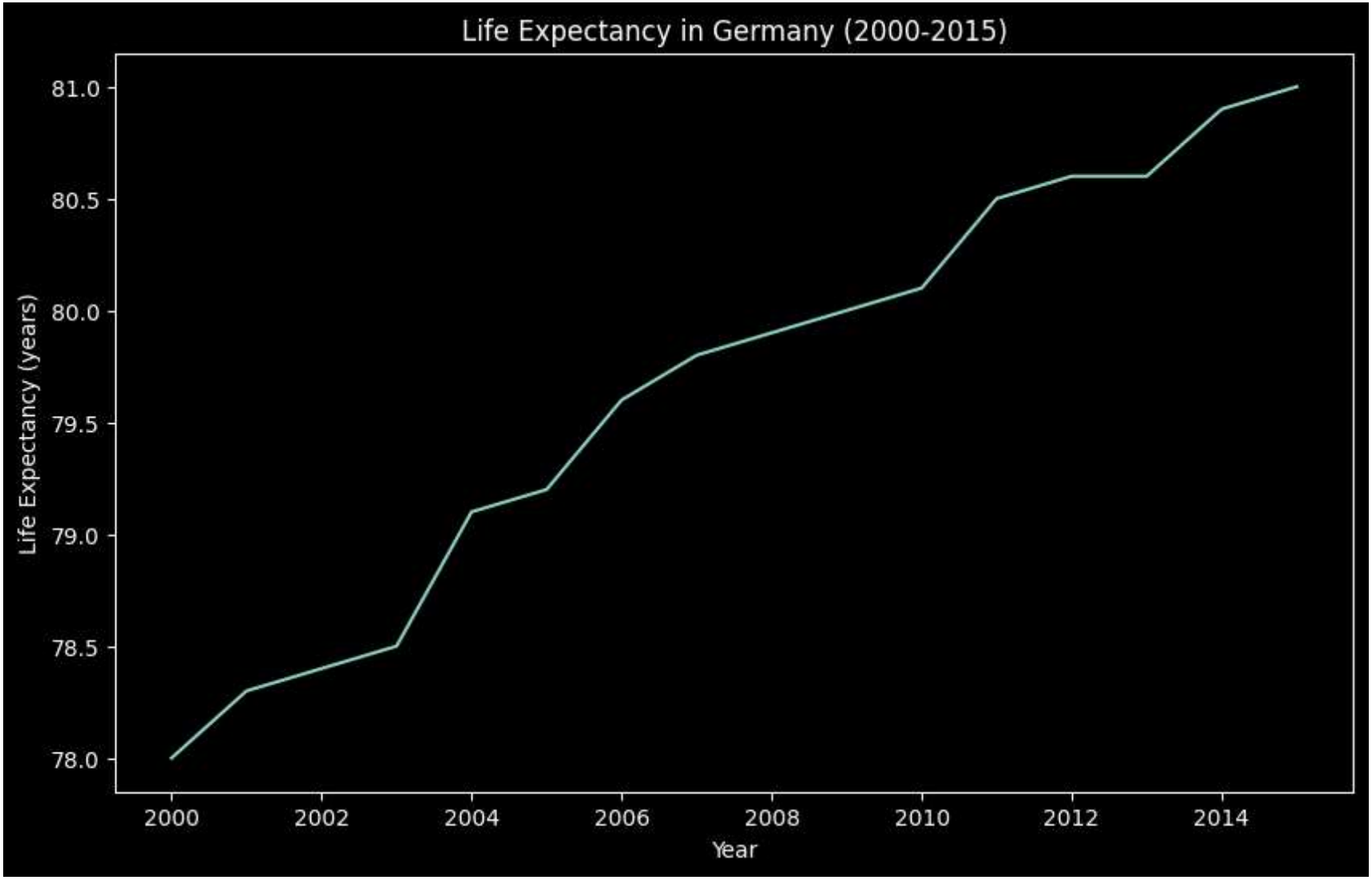
Exploring Life Expectancy in All Six Countries

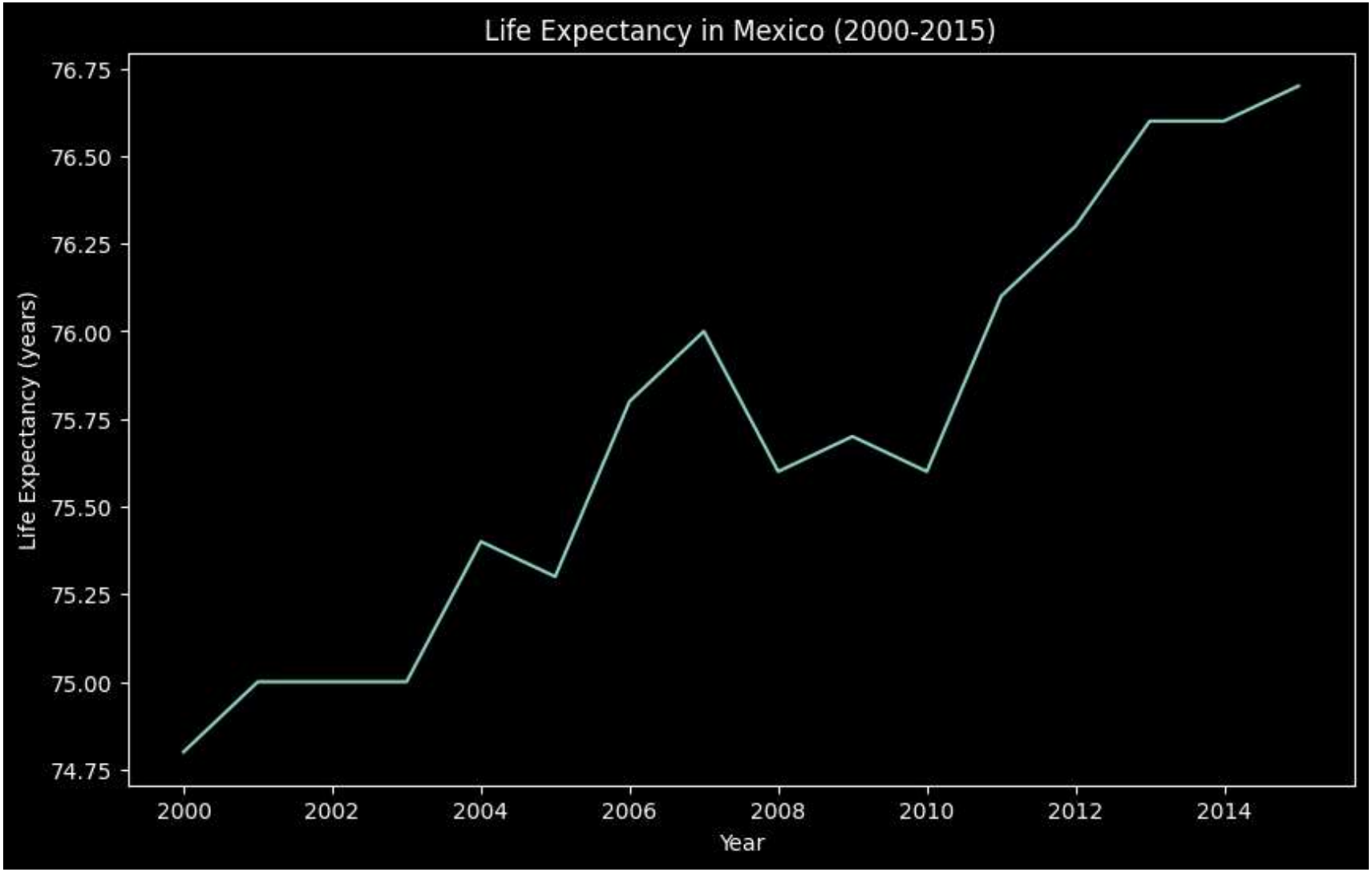
In [ ]:

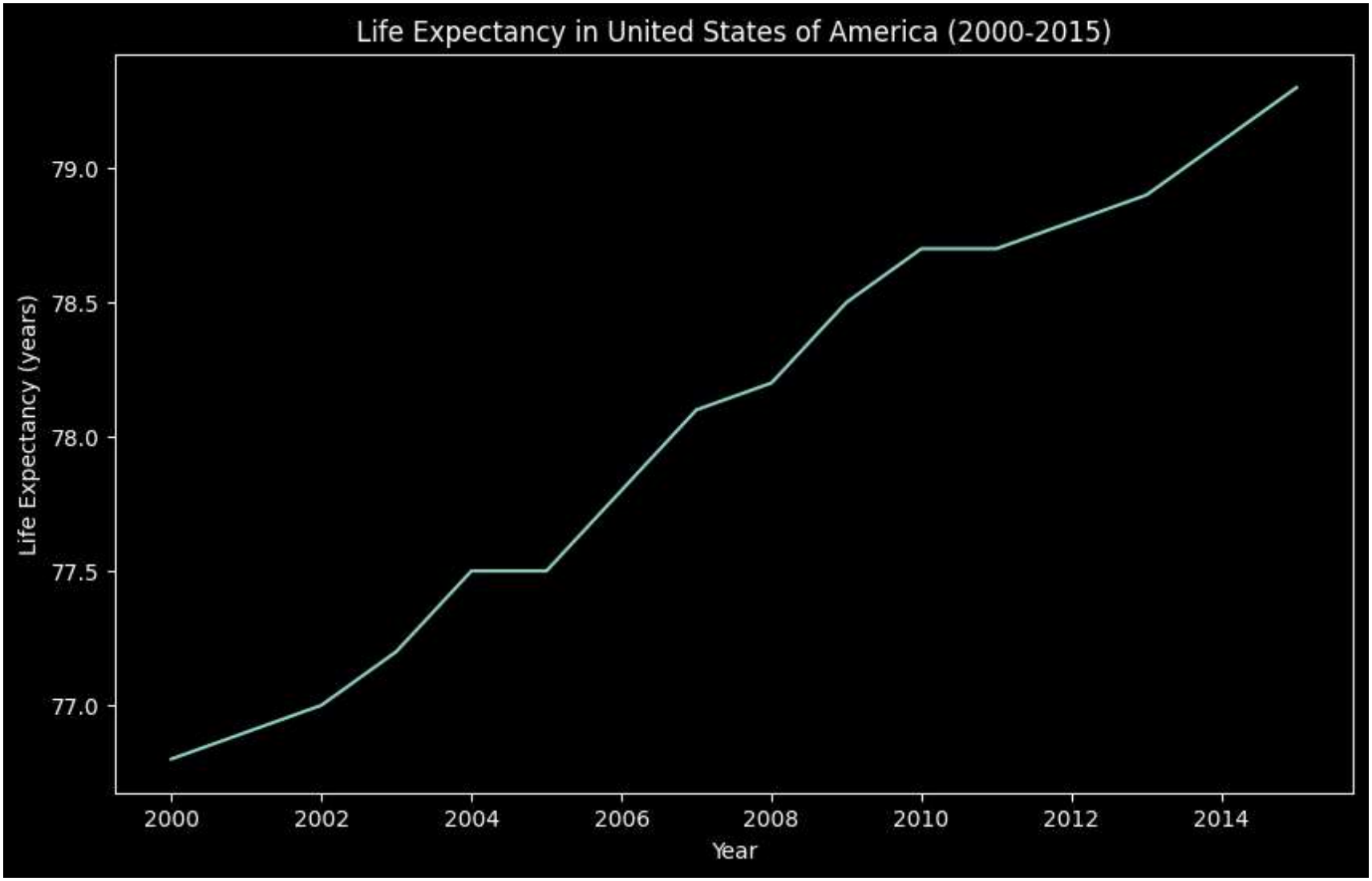
```
# Get unique country names
unique_countries = df['Country'].unique()

# Create a line plot for each country
for country in unique_countries:
    country_data = df[df['Country'] == country]

    plt.figure(figsize=(10, 6))
    sns.lineplot(data=country_data, x='Year', y='LEABY')
    plt.title(f'Life Expectancy in {country} (2000-2015)')
    plt.xlabel('Year')
    plt.ylabel('Life Expectancy (years)')

    # Save each plot with a unique filename (optional)
    plt.savefig(f'{country}_life_expectancy.png')

    plt.show()
```

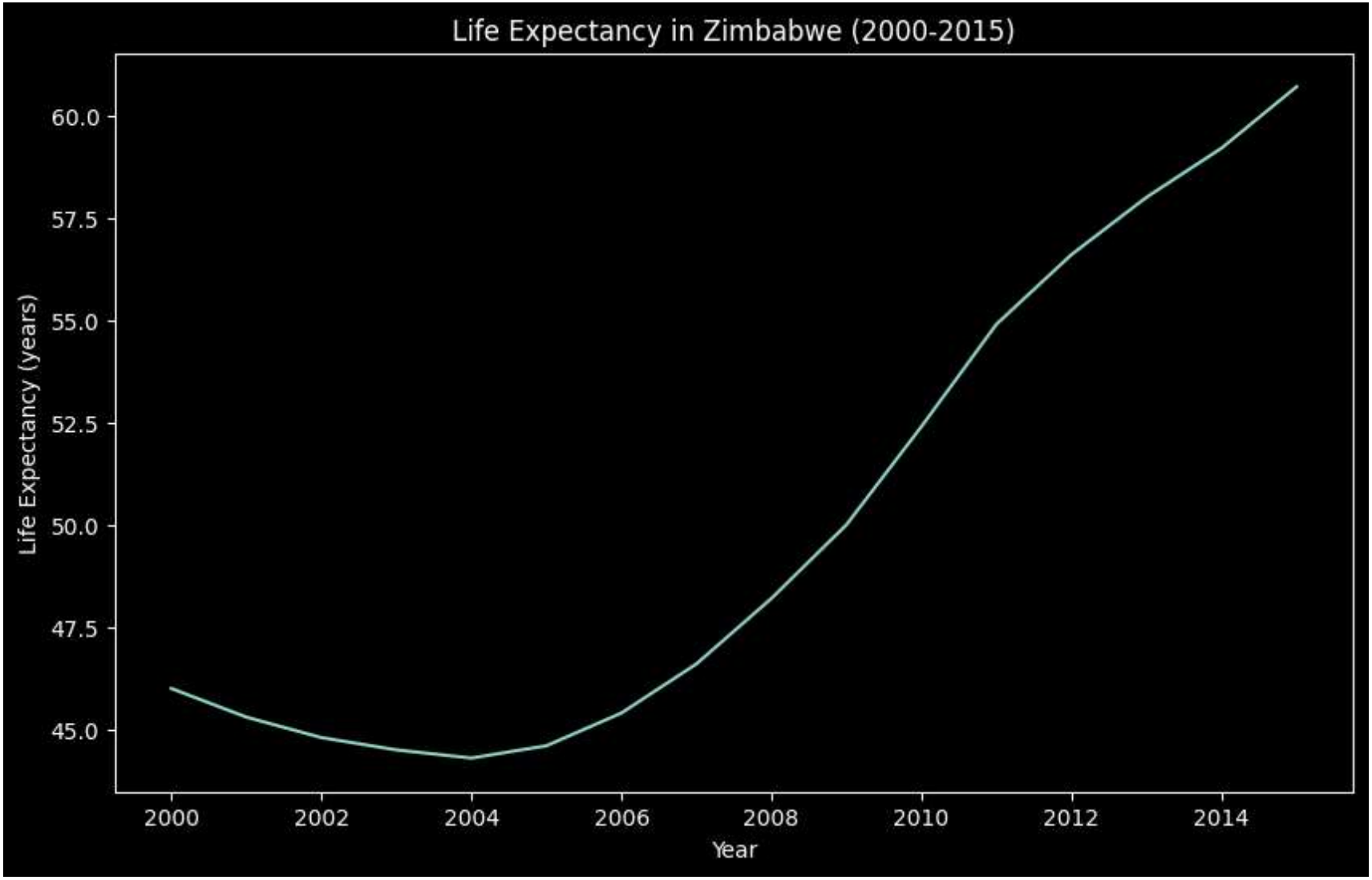

Life Expectancy in Chile (2000-2015)
Life Expectancy (years)
80.5
80.0
79.5
79.0
78.5
78.0
77.5
2000
2002
2004
2006
2008
2010
2012
2014
Year


Life Expectancy in China (2000-2015)
Life Expectancy (years)
76
75
74
73
72
2000
2002
2004
2006
2008
2010
2012
2014
Year


Life Expectancy in Germany (2000-2015)
Life Expectancy (years)
81.0
80.5
80.0
79.5
79.0
78.5
78.0
2000
2002
2004
2006
2008
2010
2012
2014
Year

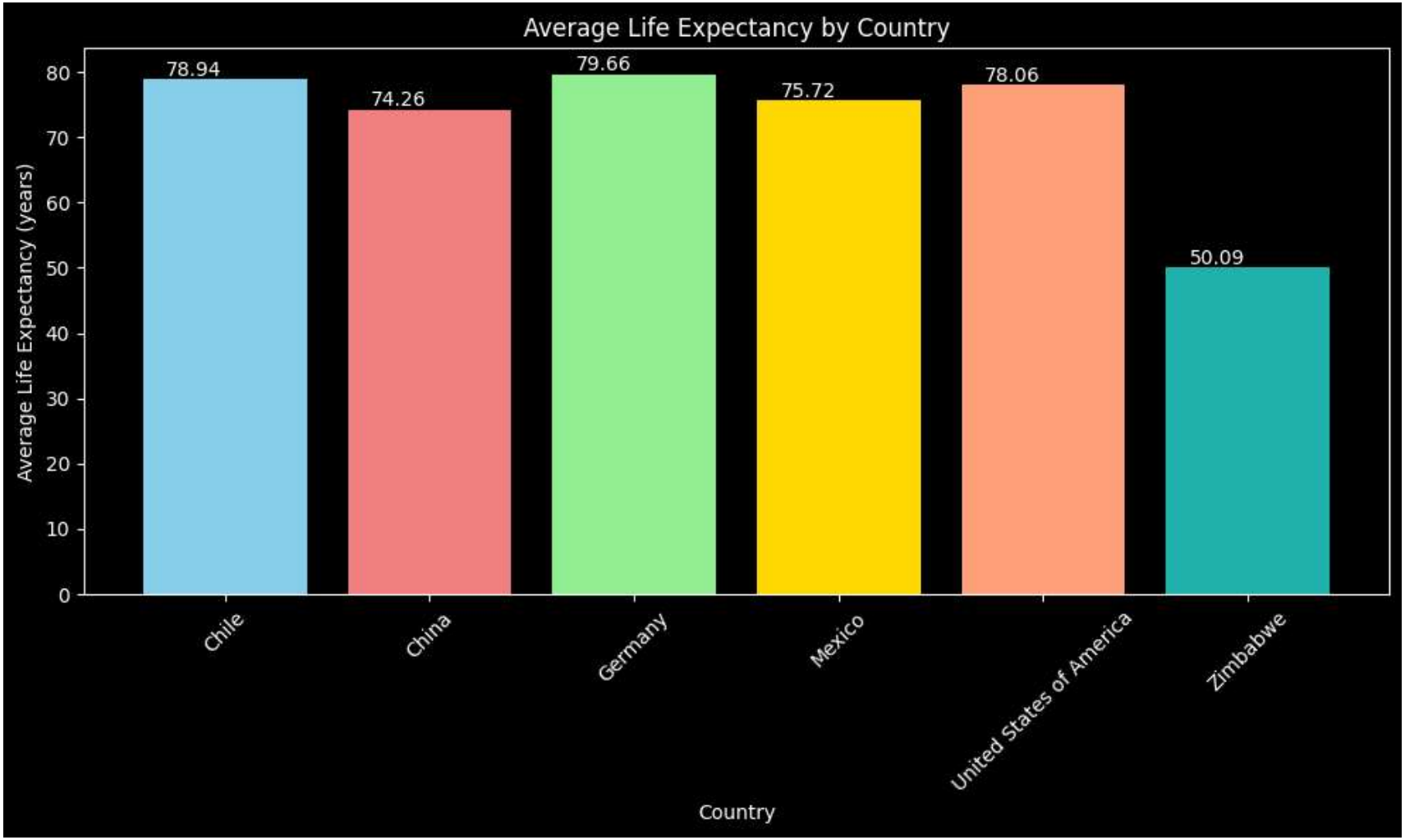

Life Expectancy in Mexico (2000-2015)
Life Expectancy (years)
76.75
76.50
76.25
76.00
75.75
75.50
75.25
75.00
74.75
2000
2002
2004
2006
2008
2010
2012
2014
Year


Life Expectancy in United States of America (2000-2015)
Life Expectancy (years)
79.0
78.5
78.0
77.5
77.0
2000
2002
2004
2006
2008
2010
2012
2014
Year



In [ ]:
```
# Calculate and print the average life expectancy for each country
avg_life_expectancies = []

for country in countries:
    country_data = df[df['Country'] == country]
    avg_life_expectancy = country_data['LEABY'].mean()
    avg_life_expectancies.append(avg_life_expectancy)

colors = ['skyblue', 'lightcoral', 'lightgreen', 'gold', 'lightsalmon', 'lightseagreen']
```

```
# Create a bar chart with custom colors
plt.figure(figsize=(10, 6))
bars = plt.bar(countries, avg_life_expectancies, color=colors)
plt.xlabel('Country')
plt.ylabel('Average Life Expectancy (years)')
plt.title('Average Life Expectancy by Country')

# Rotate the x-axis labels for better readability if needed
plt.xticks(rotation=45)

# Add data labels above each bar
for bar, avg in zip(bars, avg_life_expectancies):
    plt.text(bar.get_x() + bar.get_width() / 2 - 0.15, bar.get_height() + 0.5, f'{avg:.2f}', ha='center')

# Show the bar chart
plt.tight_layout()
plt.show()
```



The analysis unveils an encouraging trend in life expectancy across all six countries, implying significant progress in healthcare and living standards over the years.

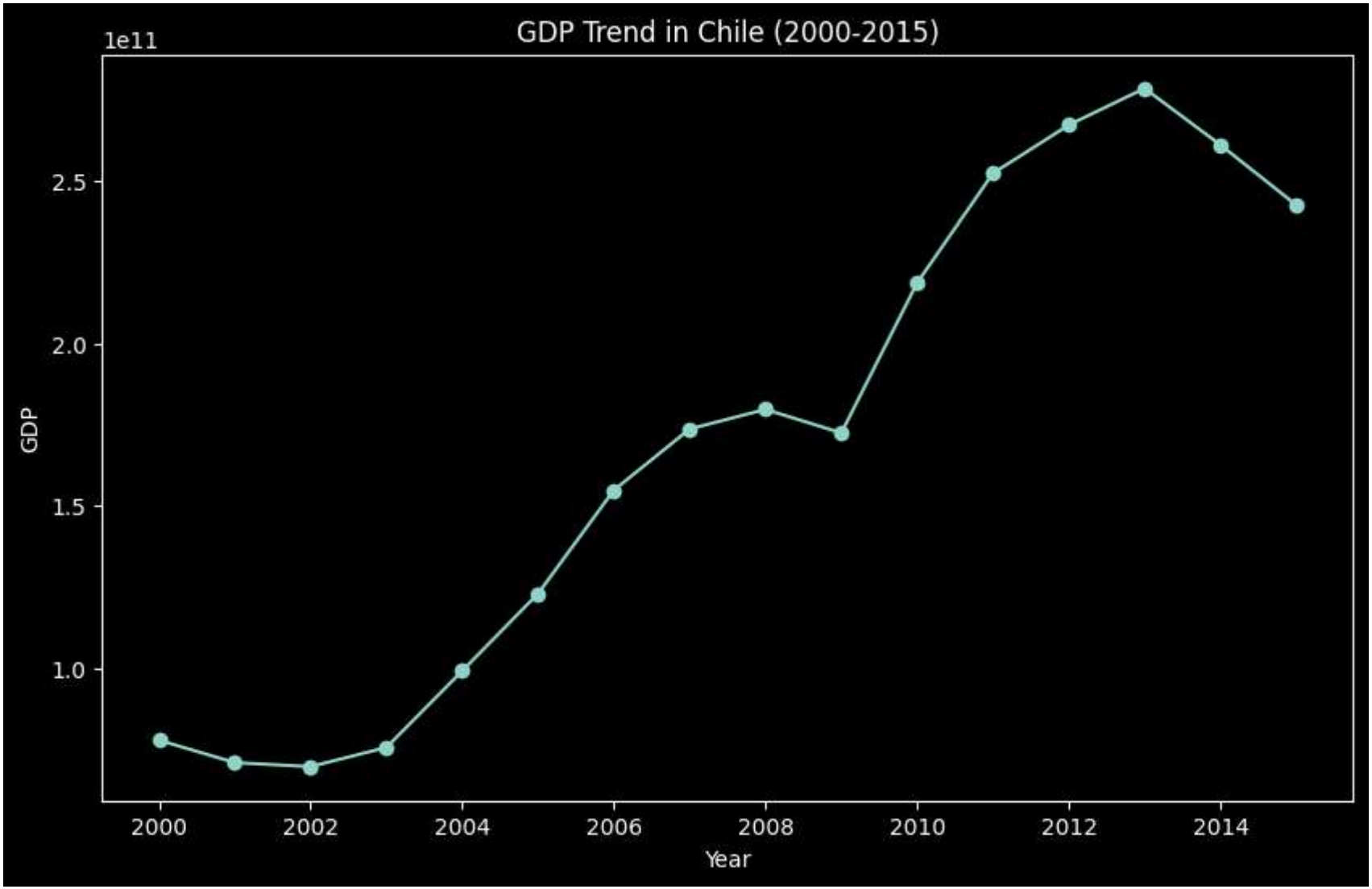
Country-Specific Insights

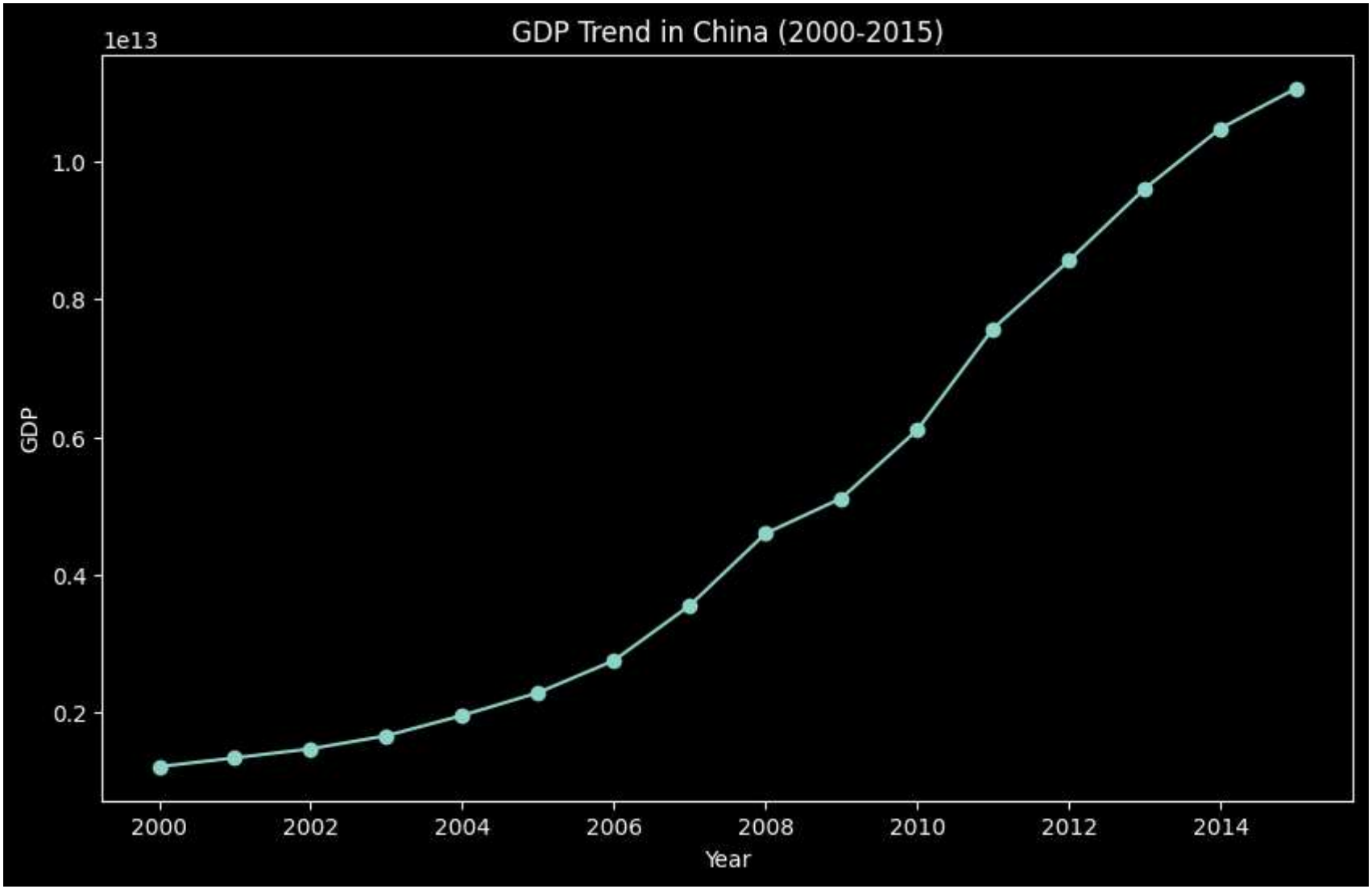
- **Chile:** Chile has consistently maintained a commendable life expectancy, holding steady at 78.94 years.
- **China:** China's life expectancy has seen a gradual rise, mirroring its ongoing economic development, with an average of 74.26 years.

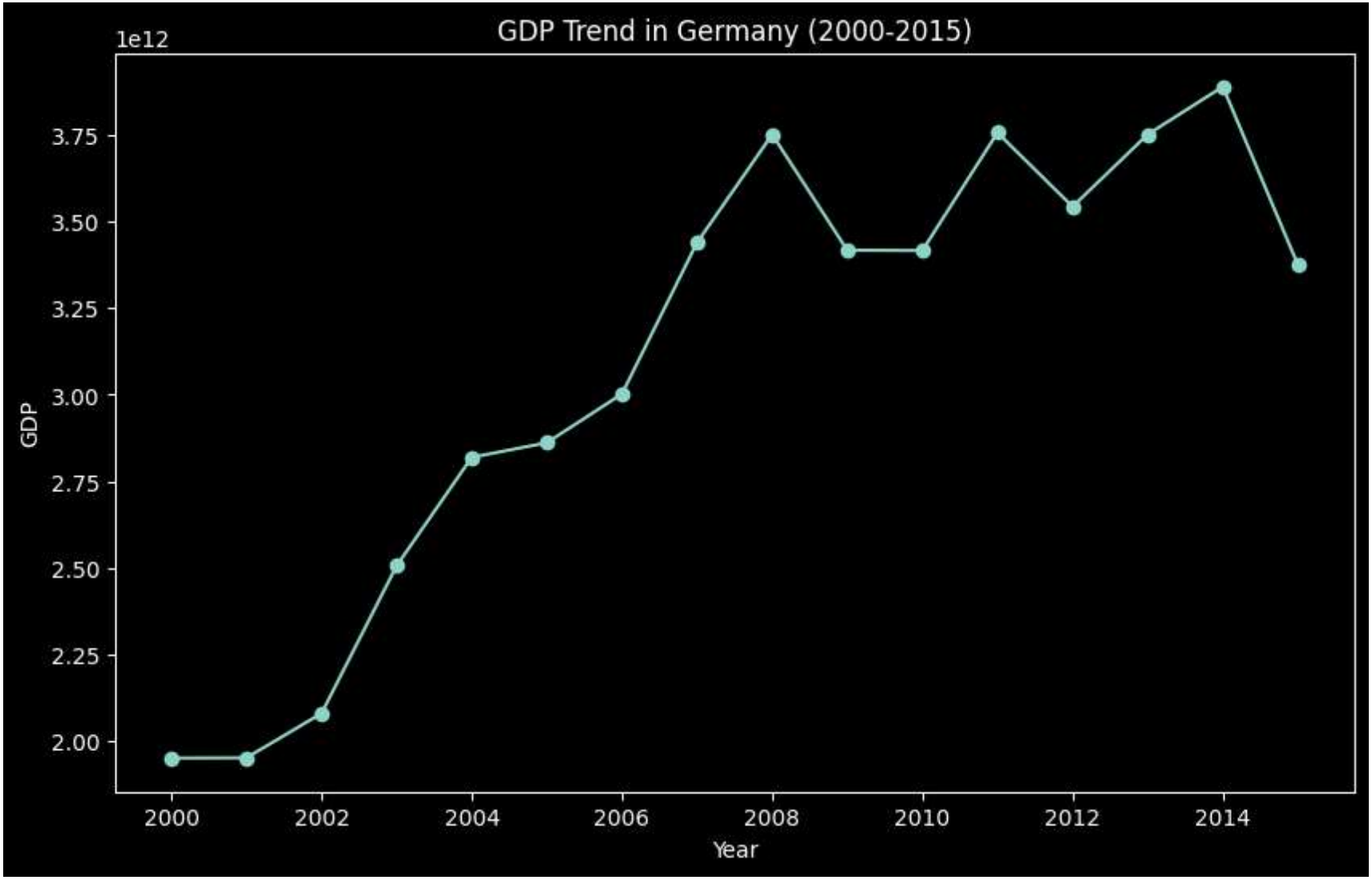
- **Germany:** Germany boasts a resiliently high life expectancy, averaging at 79.66 years.
- **Mexico:** Mexico exhibits a steady growth in life expectancy, averaging at 75.72 years.
- **United States of America:** The United States of America experienced a slower rate of increase in life expectancy compared to other countries, averaging at 78.06 years.
- **Zimbabwe:** Zimbabwe, despite having the lowest life expectancy initially, exhibits promising signs of improvement, with an average life expectancy of 50.09 years.

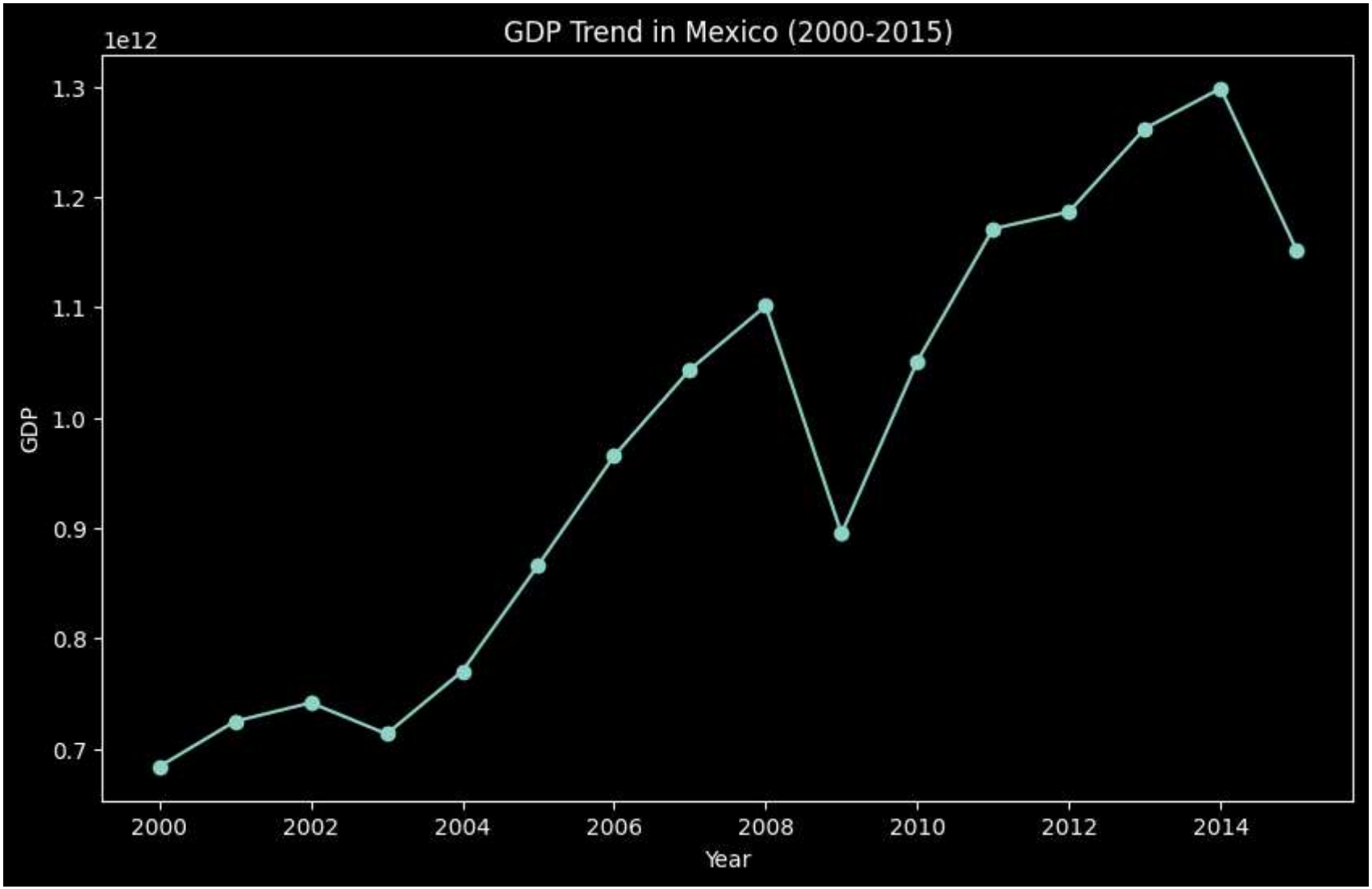
Exploring GDP Rates in All Six Countries

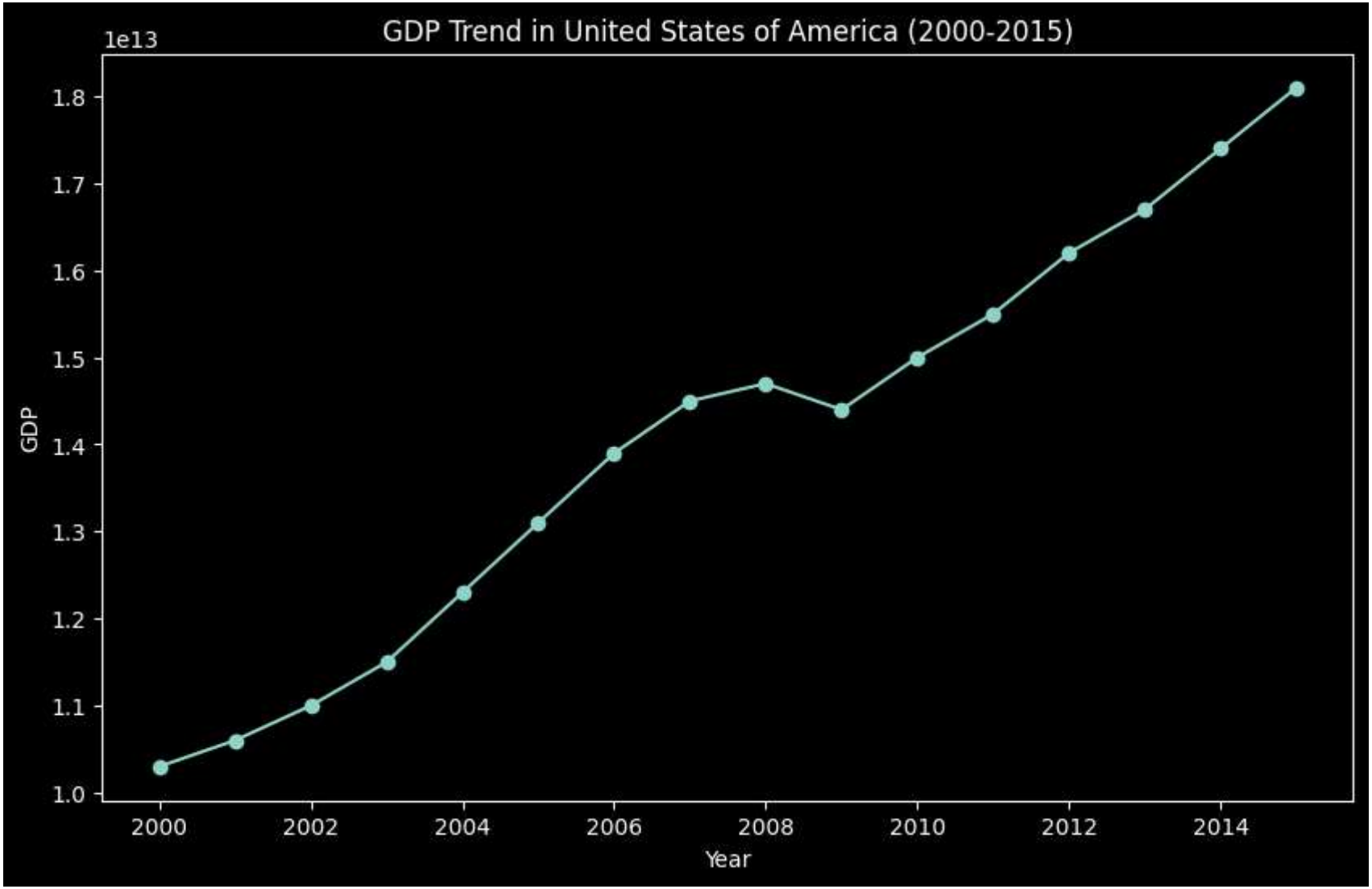
In [ ]:
```
# Get unique country names
unique_countries = df['Country'].unique()

# Create line charts for GDP trends in each country
for country in unique_countries:
    # Filter data for the current country
    country_data = df[df['Country'] == country]

    # Create a line chart
    plt.figure(figsize=(10, 6))
    plt.plot(country_data['Year'], country_data['GDP'], marker='o', linestyle='-')
    plt.title(f'GDP Trend in {country} (2000-2015)')
    plt.xlabel('Year')
    plt.ylabel('GDP')

    # Show the plot for each country
    plt.show()
```

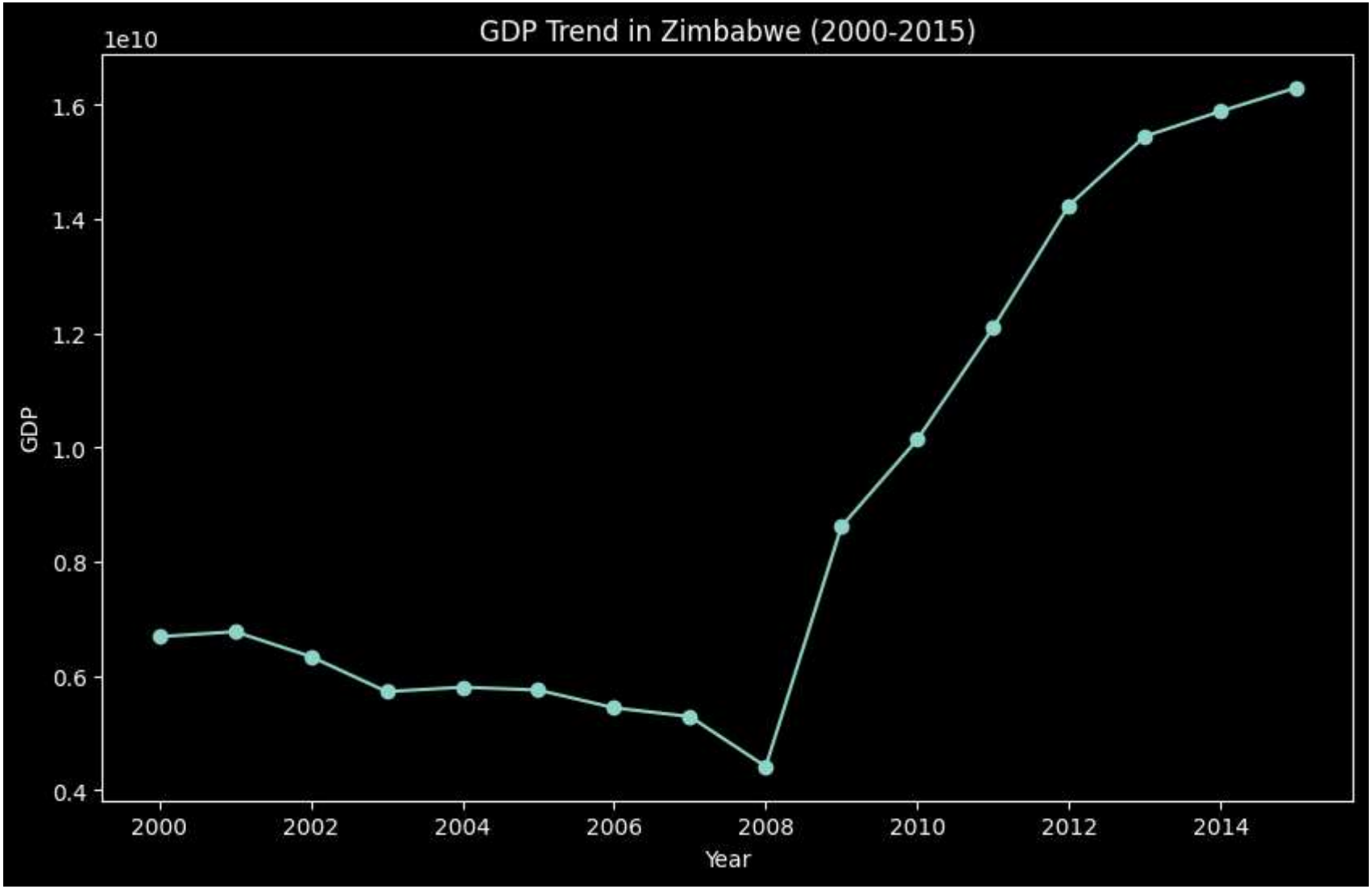

GDP Trend in Chile (2000-2015)
1e11
2.5
2.0
1.5
1.0
GDP
2000
2002
2004
2006
2008
2010
2012
2014
Year


GDP Trend in China (2000-2015)
1e13
1.0
0.8
0.6
0.4
0.2
GDP
2000
2002
2004
2006
2008
2010
2012
2014
Year


GDP Trend in Germany (2000-2015)
1e12
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
GDP
2000
2002
2004
2006
2008
2010
2012
2014
Year


GDP Trend in Mexico (2000-2015)
1e12
1.3
1.2
1.1
1.0
0.9
0.8
0.7
GDP
2000
2002
2004
2006
2008
2010
2012
2014
Year


GDP Trend in United States of America (2000-2015)
1e13
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
GDP
2000
2002
2004
2006
2008
2010
2012
2014
Year



In [ ]:
```
# Get unique country names
unique_countries = df['Country'].unique()

# Loop through each country and calculate summary statistics
for country in unique_countries:
    country_data = df[df['Country'] == country]

    # Calculate summary statistics for GDP
    summary_stats = country_data['GDP'].describe()
```

```
    print(f'Summary statistics for {country}:')
    print(summary_stats)
    print('\n')
```

```
Summary statistics for Chile:
count    1.600000e+01
mean     1.697888e+11
std      7.687884e+10
min      6.973681e+10
25%      9.387303e+10
50%      1.729975e+11
75%      2.449515e+11
max      2.783840e+11
Name: GDP, dtype: float64


Summary statistics for China:
count    1.600000e+01
mean     4.957714e+12
std      3.501096e+12
min      1.211350e+12
25%      1.881585e+12
50%      4.075195e+12
75%      7.819550e+12
max      1.106470e+13
Name: GDP, dtype: float64


Summary statistics for Germany:
count    1.600000e+01
mean     3.094776e+12
std      6.674862e+11
min      1.949950e+12
25%      2.740870e+12
50%      3.396350e+12
75%      3.596078e+12
max      3.890610e+12
Name: GDP, dtype: float64


Summary statistics for Mexico:
count    1.600000e+01
mean     9.766506e+11
std      2.095716e+11
min      6.836480e+11
25%      7.630910e+11
```

```
50%      1.004376e+12
75%      1.156992e+12
max      1.298460e+12
Name: GDP, dtype: float64


Summary statistics for United States of America:
count    1.600000e+01
mean     1.407500e+13
std      2.432694e+12
min      1.030000e+13
25%      1.210000e+13
50%      1.445000e+13
75%      1.567500e+13
max      1.810000e+13
Name: GDP, dtype: float64


Summary statistics for Zimbabwe:
count    1.600000e+01
mean     9.062580e+09
std      4.298310e+09
min      4.415703e+09
25%      5.748309e+09
50%      6.733671e+09
75%      1.263446e+10
max      1.630467e+10
Name: GDP, dtype: float64
```

1. **Chile:**
    - In Chile, the mean GDP over the observed years was approximately $169.79 billion, with a standard deviation of$76.88 billion. The minimum recorded GDP was around $69.74 billion, while the maximum reached$278.38 billion.
2. **China:**
    - China exhibited significant economic growth, with a mean GDP of approximately $4.96 trillion and a standard deviation of$3.50 trillion. The country's GDP ranged from a minimum of $1.21 trillion to a maximum of$11.06 trillion.

3. **Germany:**

   - Germany maintained economic stability, with a mean GDP of about $3.09 trillion and a standard deviation of $0.67 trillion. The country's GDP fluctuated between a minimum of $1.94 trillion and a maximum of $3.89 trillion.

4. **Mexico:**

   - Mexico experienced steady GDP growth, with an average of around $0.98 trillion and a standard deviation of $0.21 trillion. The lowest recorded GDP was $0.68 trillion, while the highest was $1.30 trillion.

5. **United States of America:**

   - The United States of America had the highest mean GDP among the six countries, approximately $14.07 trillion, with a standard deviation of $2.43 trillion. The country's GDP ranged from a minimum of $10.30 trillion to a maximum of $18.10 trillion.

6. **Zimbabwe:**

   - Zimbabwe, with the lowest GDP among the six countries, had an average GDP of about $9.06 billion and a standard deviation of $4.29 billion. The country's GDP ranged from a minimum of $4.42 billion to a maximum of $16.30 billion.

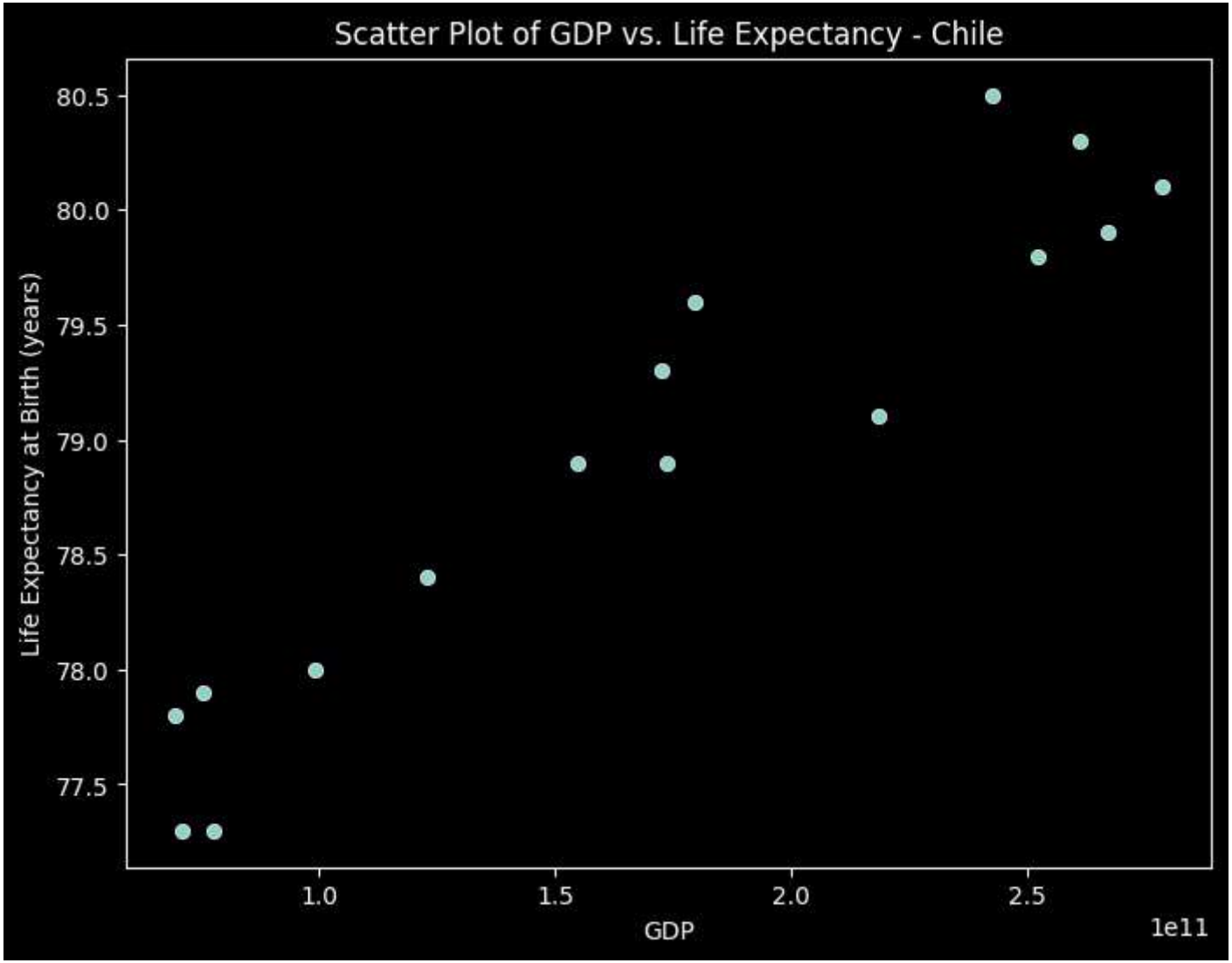
Correlation Between GDP and Life Expectancy of All Six Countries

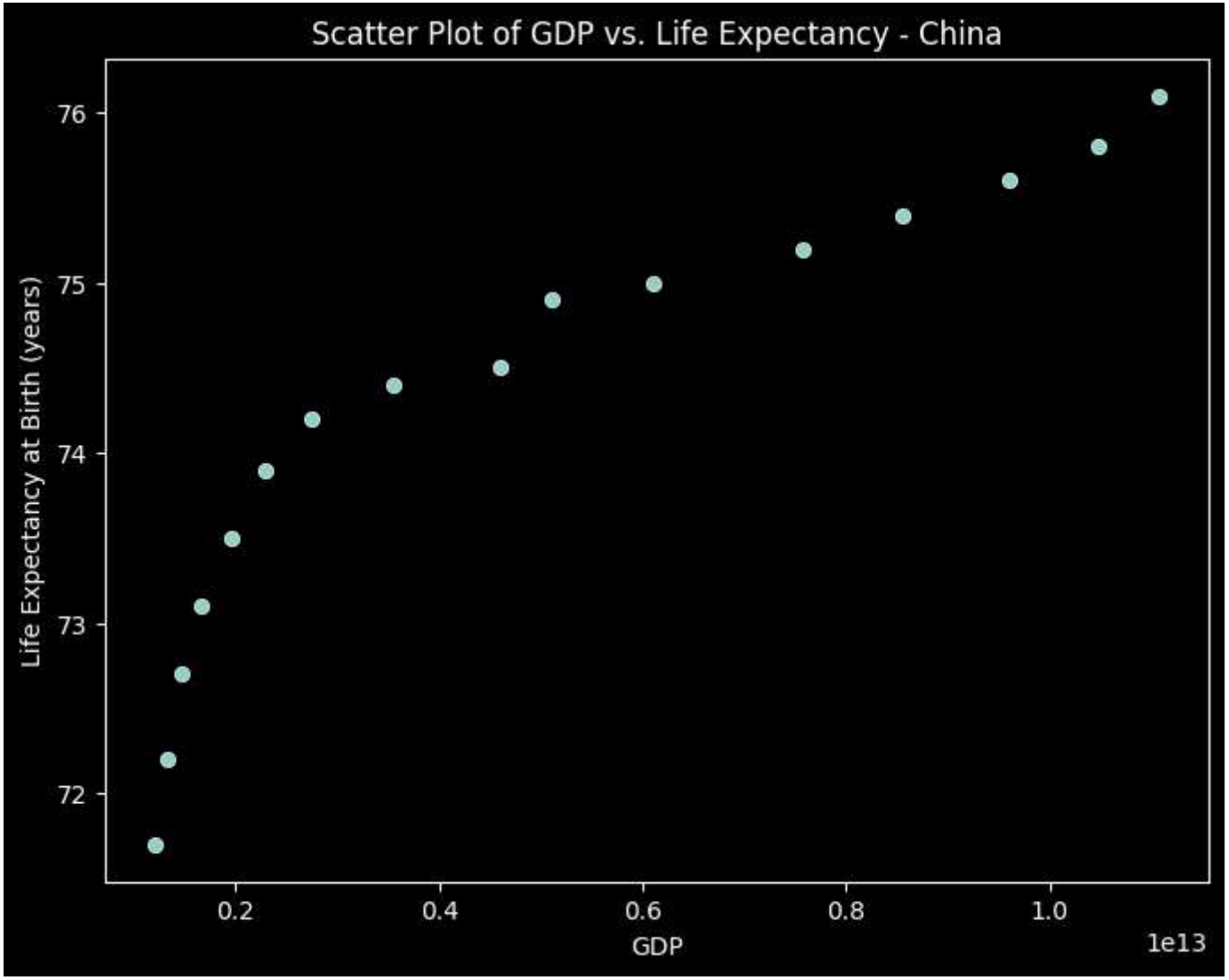
In [ ]:
```
# Calculate the correlation coefficient
correlation_coefficient = df['GDP'].corr(df['LEABY'])
print(f"Correlation Coefficient between GDP and Life Expectancy: {correlation_coefficient}")
```

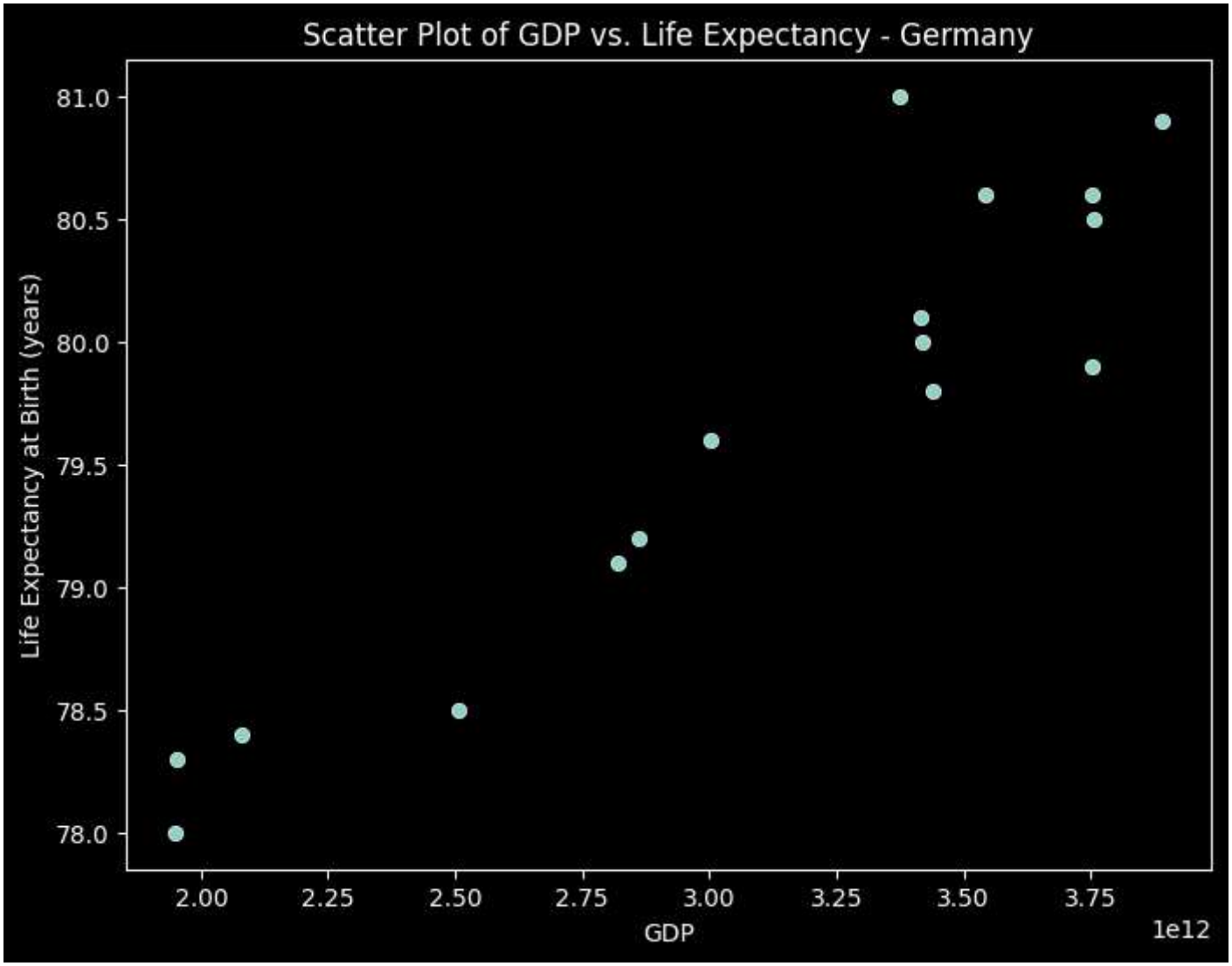
```
Correlation Coefficient between GDP and Life Expectancy: 0.3432067484491559
```

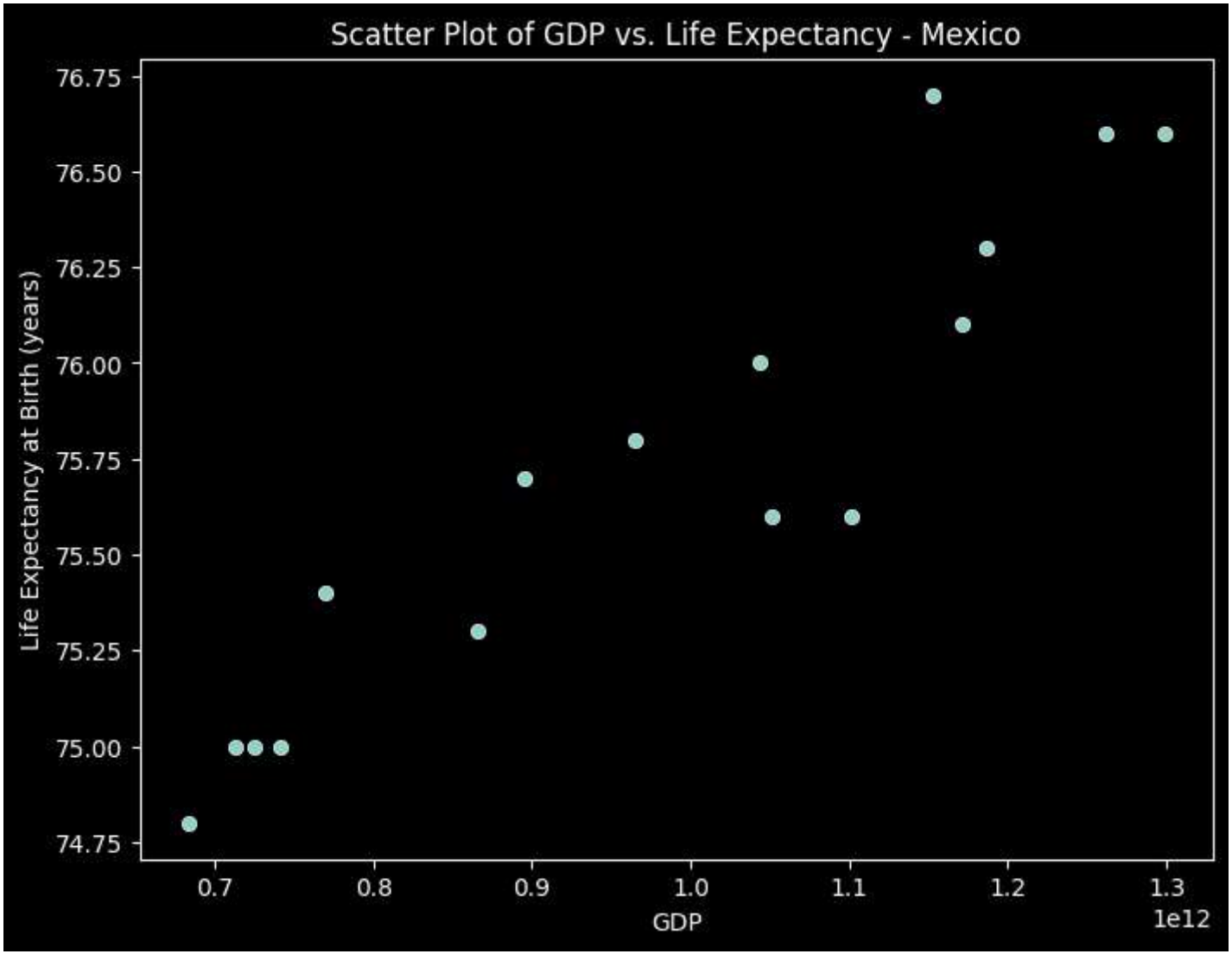
In [ ]:
```
# Get unique country names
unique_countries = df['Country'].unique()

# Create a scatter plot for each country
for country in unique_countries:
    # Filter data for the current country
    country_data = df[df['Country'] == country]

    # Create a scatter plot
    plt.figure(figsize=(8, 6))
```

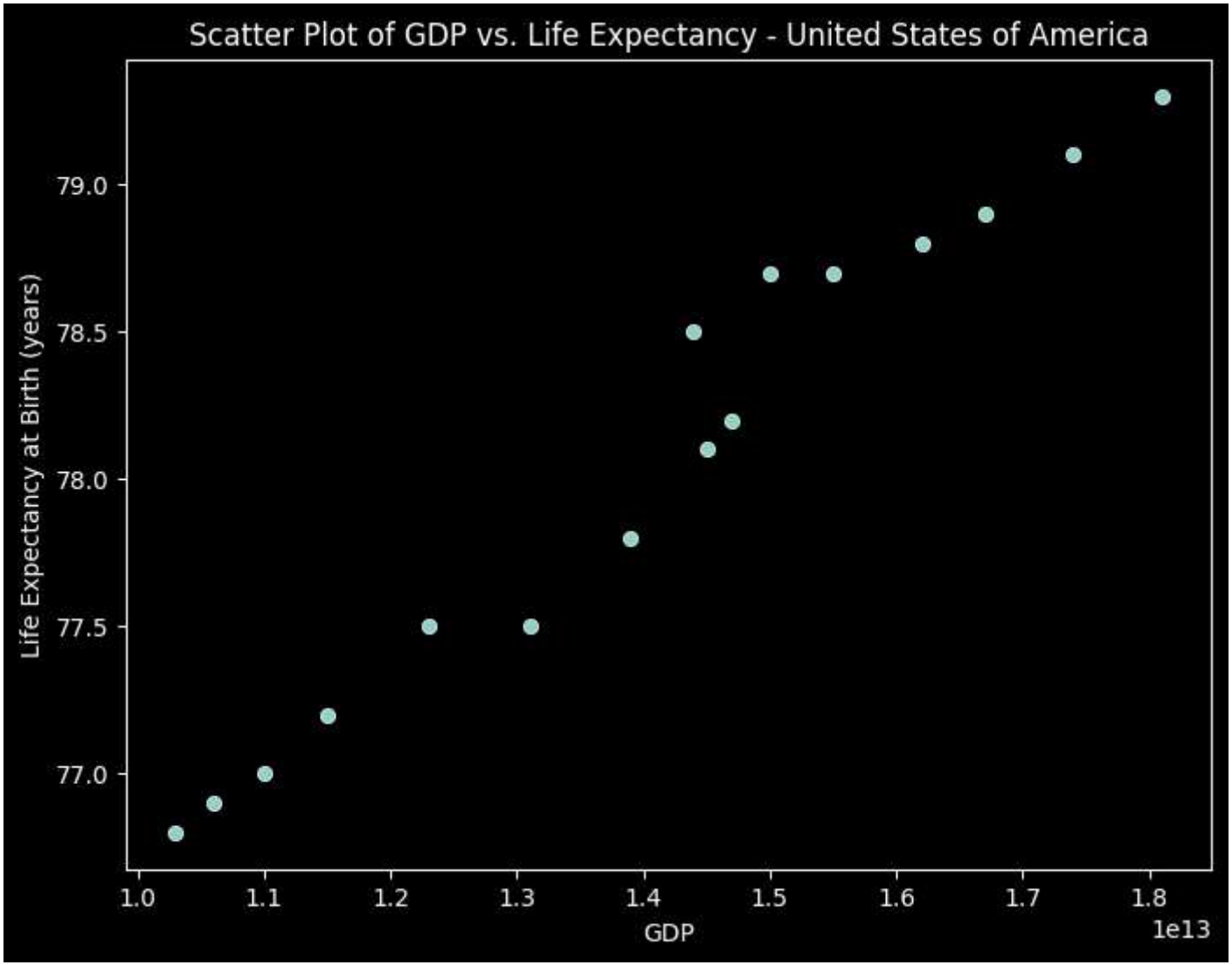
```
    sns.scatterplot(data=country_data, x='GDP', y='LEABY')

    # Customize plot labels and title
    plt.xlabel('GDP')
    plt.ylabel('Life Expectancy at Birth (years)')
    plt.title(f'Scatter Plot of GDP vs. Life Expectancy - {country}')

    # Show the plot for the current country
    plt.show()
```

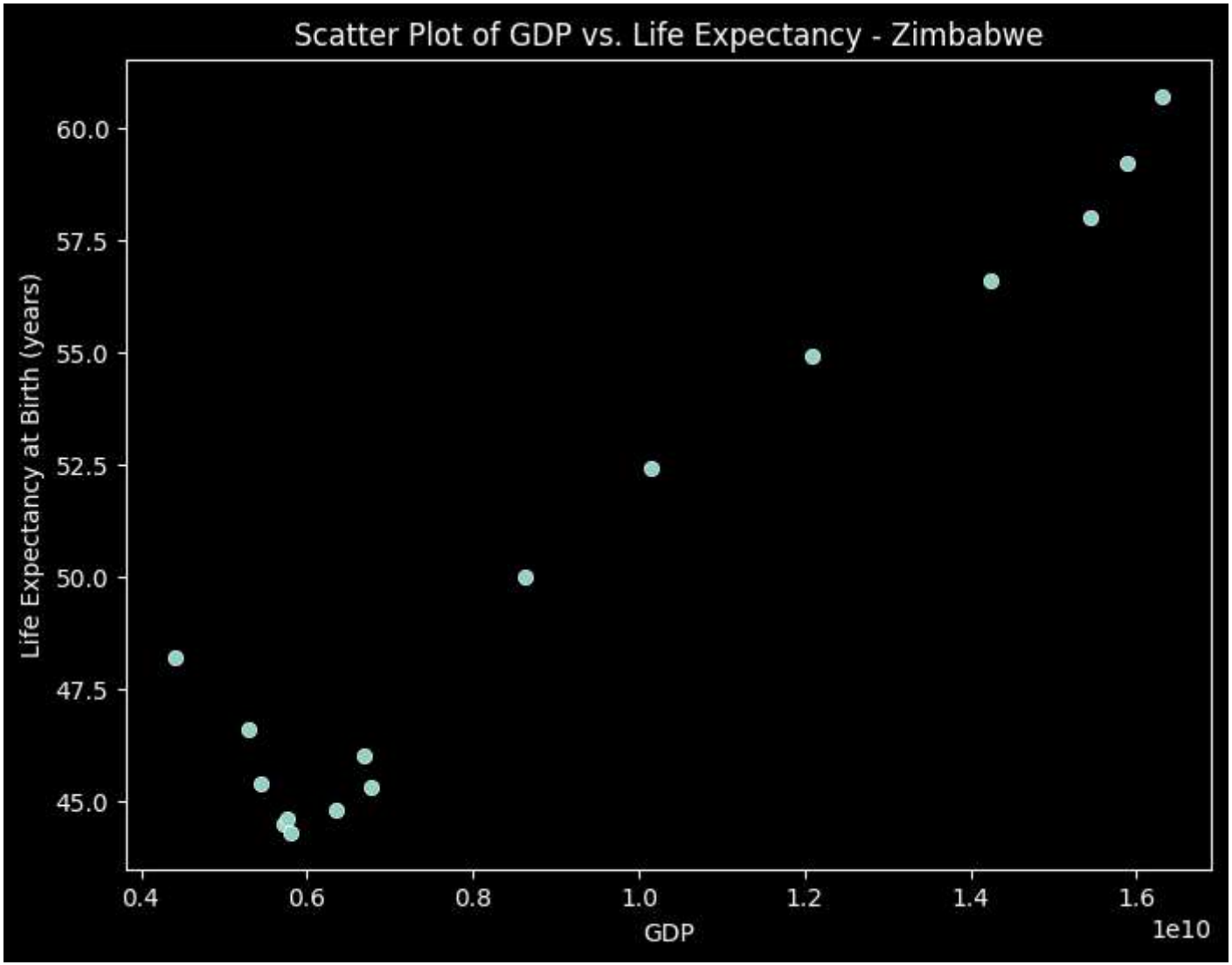



Scatter Plot of GDP vs. Life Expectancy - China
Life Expectancy at Birth (years)
76
75
74
73
72
0.2
0.4
0.6
0.8
1.0
GDP
1e13


Scatter Plot of GDP vs. Life Expectancy - Germany
Life Expectancy at Birth (years)
81.0
80.5
80.0
79.5
79.0
78.5
78.0
2.00
2.25
2.50
2.75
3.00
3.25
3.50
3.75
GDP
1e12


Scatter Plot of GDP vs. Life Expectancy - Mexico
Life Expectancy at Birth (years)
76.75
76.50
76.25
76.00
75.75
75.50
75.25
75.00
74.75
0.7
0.8
0.9
1.0
1.1
1.2
1.3
GDP
1e12


Scatter Plot of GDP vs. Life Expectancy - United States of America
Life Expectancy at Birth (years)
79.0
78.5
78.0
77.5
77.0
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
GDP
1e13



The correlation coefficient between GDP and life expectancy for the analyzed data is approximately 0.3432. This positive correlation coefficient indicates a positive linear relationship between a country's GDP and its life expectancy at birth (LEABY). In other words, as a country's GDP increases, its life expectancy tends to increase as well.

However, it's important to note that correlation does not imply causation. While there is a correlation between GDP and life expectancy, various other factors may also influence life expectancy, and this correlation does not necessarily indicate a direct causal relationship.

This information can be valuable for understanding the general trends in the dataset and exploring potential associations between economic prosperity (GDP) and public health (life expectancy) across different countries.

Conclusion

In this analysis, we delved into data on GDP and life expectancy from six diverse countries: Chile, China, Germany, Mexico, the United States of America, and Zimbabwe. Our primary goal was to explore the intricate relationship between GDP and life expectancy and derive meaningful insights from the dataset.

Life Expectancy Trends

- The analysis of life expectancy trends in six countries reveals a positive trajectory, reflecting substantial advancements in healthcare and living conditions. Each country demonstrates unique characteristics in its life expectancy journey. Chile maintains a commendable and stable life expectancy, while China showcases gradual growth in parallel with economic development. Germany stands out with consistently high life expectancy rates, while Mexico exhibits steady improvements. The United States experiences a comparatively slower rise, and Zimbabwe, despite starting from a lower point, shows promising signs of enhancement over time. These insights collectively illustrate the diverse patterns of life expectancy within this global context.

GDP and Life Expectancy Correlation

- We computed a correlation coefficient of approximately 0.3432 between GDP and life expectancy. This positive correlation suggests that as a country's GDP increases, so does its life expectancy. However, it is essential to remember that correlation does not imply causation as other factors may come into affect to increase life expectancy.

In summary, our analysis underscores the positive connection between economic prosperity, measured by GDP, and life expectancy across the examined countries. While this correlation provides valuable insights, it is vital to acknowledge that various other factors influence life expectancy. Our findings stress the importance of adopting a multifaceted approach to public health and development policies.

Comprehending the intricate interplay between GDP and life expectancy empowers policymakers, researchers, and healthcare professionals to make informed decisions aimed at enhancing public health and economic well-being. This analysis contributes to the ongoing global discourse on health and development.

By scrutinizing the data and trends presented here, we gain deeper insights into the complex relationship between economic determinants and public health outcomes, ultimately contributing to more effective strategies for enhancing the quality of life for individuals worldwide.