WORLD BANK - LIFE EXPECTANCY PORTFOLIO PROJECT

May 22, 2024

[1]:
```python
#Loading the libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```python
#Loading the data
wb = pd.read_csv("C:\\Users\\Brian Motee\\Desktop\\Python Portfolio␣
↪projects\\Life-Expectancy-and-GDP-Starter\\all_data.csv")
print(wb.head())
```

```
  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
```

[3]:
```python
#Inspecting the data types
print(wb.info())
```

```
<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
```

[4]:
```python
#Getting numerical summaries
print(wb.describe())
```

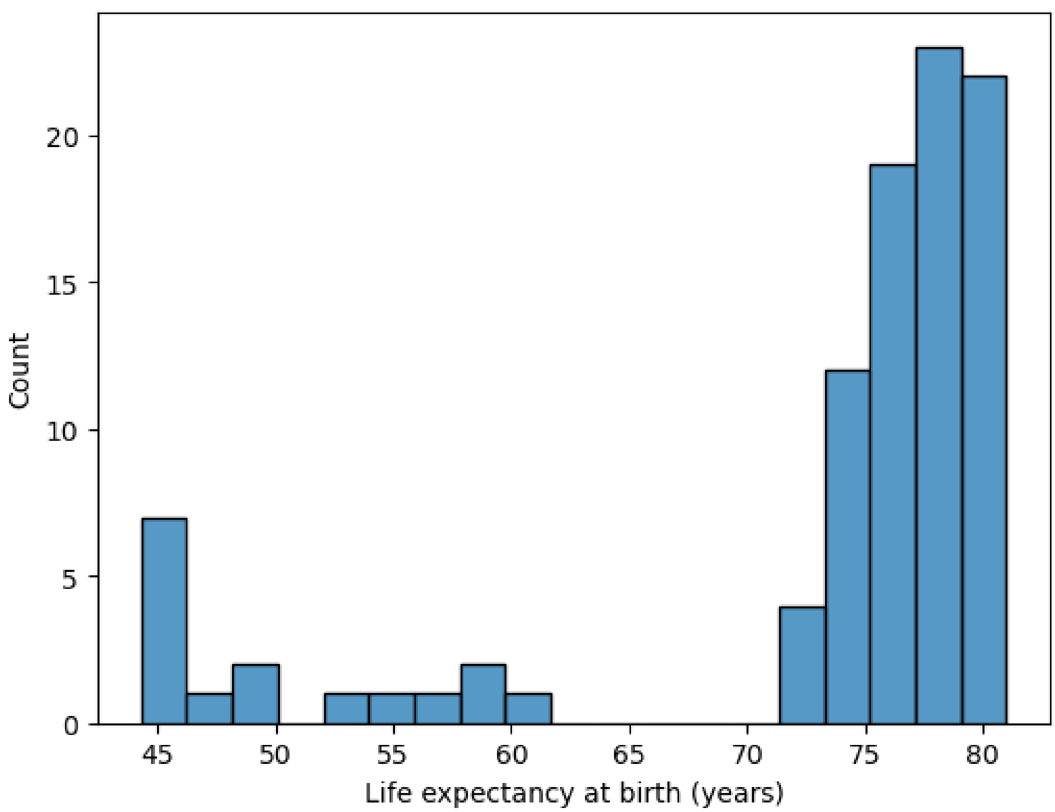
```
              Year  Life expectancy at birth (years)           GDP
count    96.000000                         96.000000  9.600000e+01
```

mean	2007.500000	72.789583	3.880499e+12
std	4.633971	10.672882	5.197561e+12
min	2000.000000	44.300000	4.415703e+09
25%	2003.750000	74.475000	1.733018e+11
50%	2007.500000	76.750000	1.280220e+12
75%	2011.250000	78.900000	4.067510e+12
max	2015.000000	81.000000	1.810000e+13

The average life expectancy is 73 yrs(Rounded Off) while the average GDP is 3.88. The is a big difference between the mode and the mean of GDP indicating that the distribution of GDP might be skewed

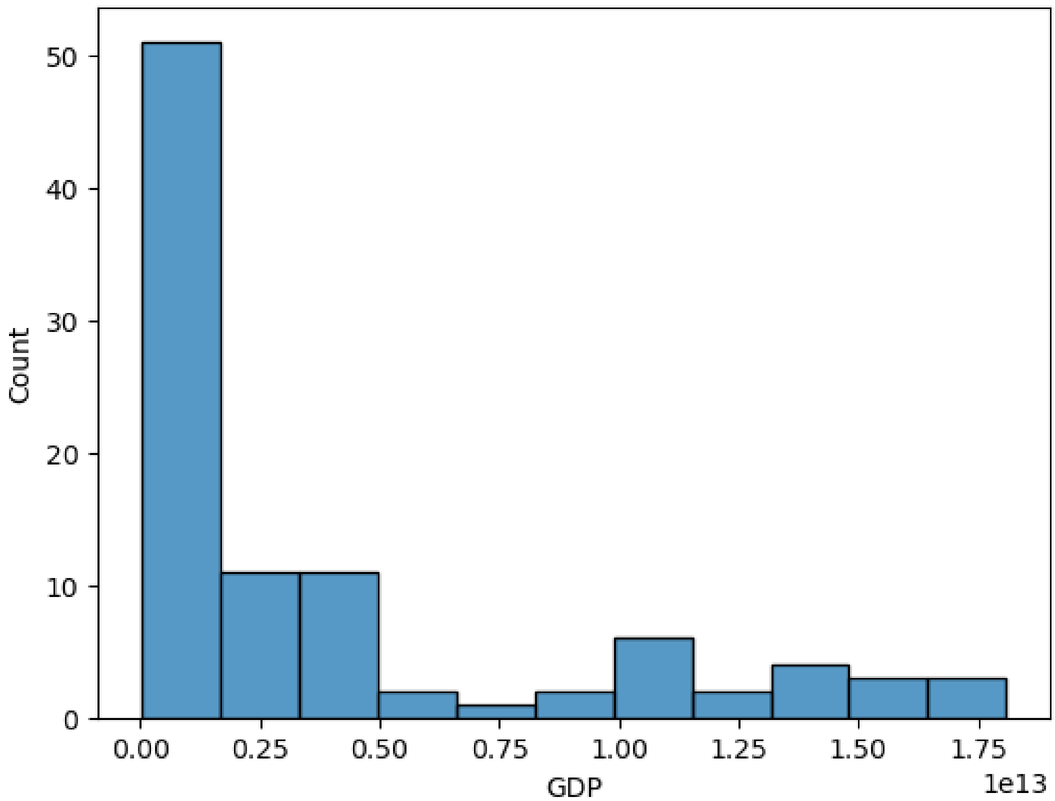
1 EXPLORATORY DATA ANALYSIS

[5]:
```python
#Assessing the distribution of life expectancy
sns.histplot(wb["Life expectancy at birth (years)"])
plt.show()
plt.clf()
```




```
<Figure size 640x480 with 0 Axes>
```

[6]:
```
#Assessing the distribution of GDP
sns.histplot(wb["GDP"])
plt.show()
plt.clf()
```




```
<Figure size 640x480 with 0 Axes>
```

The histogram is right skewed which means the data has outliers. This makes the mean Unreliable, hence we can use the mode to destribe the GDP rather than the mean. The mode of the GDP is 1.28

[7]:
```
#Getting the frequency of the countries
wb.Country.value_counts()
```

[7]:
```
Country
Chile                       16
China                       16
Germany                     16
Mexico                      16
United States of America    16
Zimbabwe                    16
Name: count, dtype: int64
```

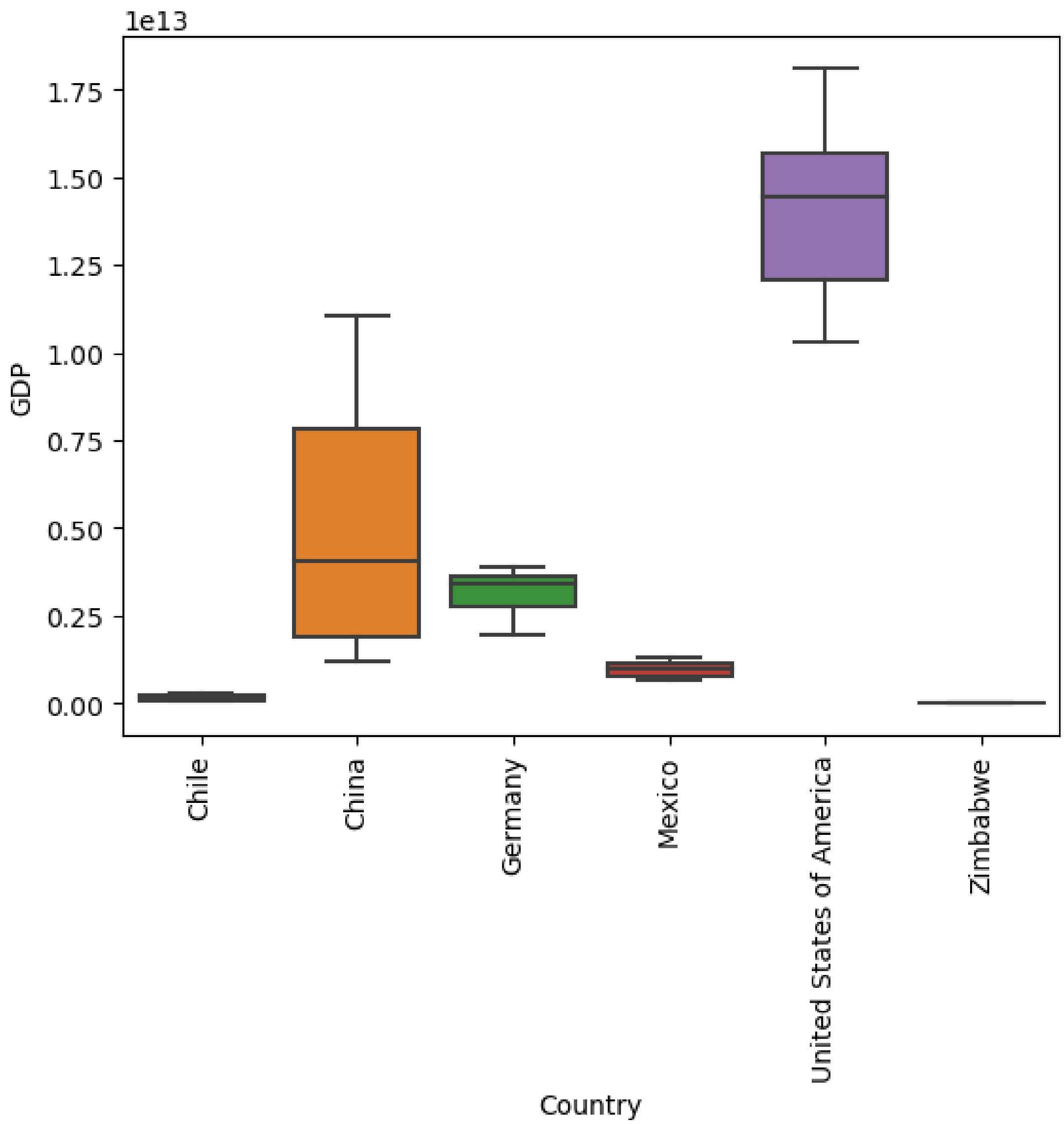
There are equal frequencies of the countries in this dataset

2 #Bivariate Analysis

[8]:
```
#Evaluating the association between Country and GDP
#Getting the DATA
Chile_data = wb.GDP[wb.Country=='Chile']
China_data = wb.GDP[wb.Country=='China']
Germany_data = wb.GDP[wb.Country=='Germany']
Mexico_data = wb.GDP[wb.Country=='mexico']
USA_data = wb.GDP[wb.Country=='United States of America']
Zimbabwe_data = wb.GDP[wb.Country=='Zimbabwe']
#Getting the means
mean_Chile = np.mean(Chile_data)
mean_China = np.mean(China_data)
mean_Germany = np.mean(Germany_data)
mean_Mexico = np.mean(Mexico_data)
mean_USA = np.mean(USA_data)
mean_Zim = np.mean(Zimbabwe_data)
#Printing the means
print("CHILE MEAN: " + str(mean_Chile))
print("CHINA MEAN: " + str(mean_China))
print("GERMANY MEAN: " + str(mean_Germany))
print("MEXICO MEAN: " + str(mean_Mexico))
print("USA MEAN: " + str(mean_USA))
print("ZIMBABWE MEAN: " + str(mean_Zim))
```

```
CHILE MEAN: 169788845015.3125
CHINA MEAN: 4957713750000.0
GERMANY MEAN: 3094775625000.0
MEXICO MEAN: nan
USA MEAN: 14075000000000.0
ZIMBABWE MEAN: 9062579595.0625
```

[9]:
```
#Visualizing the means
sns.boxplot(data=wb,x='Country',y='GDP')
plt.xticks(rotation=90)
plt.show()
plt.clf()
```



<Figure size 640x480 with 0 Axes>

From the boxplots it is easy to tell that GDP is associate with the country and that it varies from country to countr with USA having the highest GDP and Zimbabwe the lowest

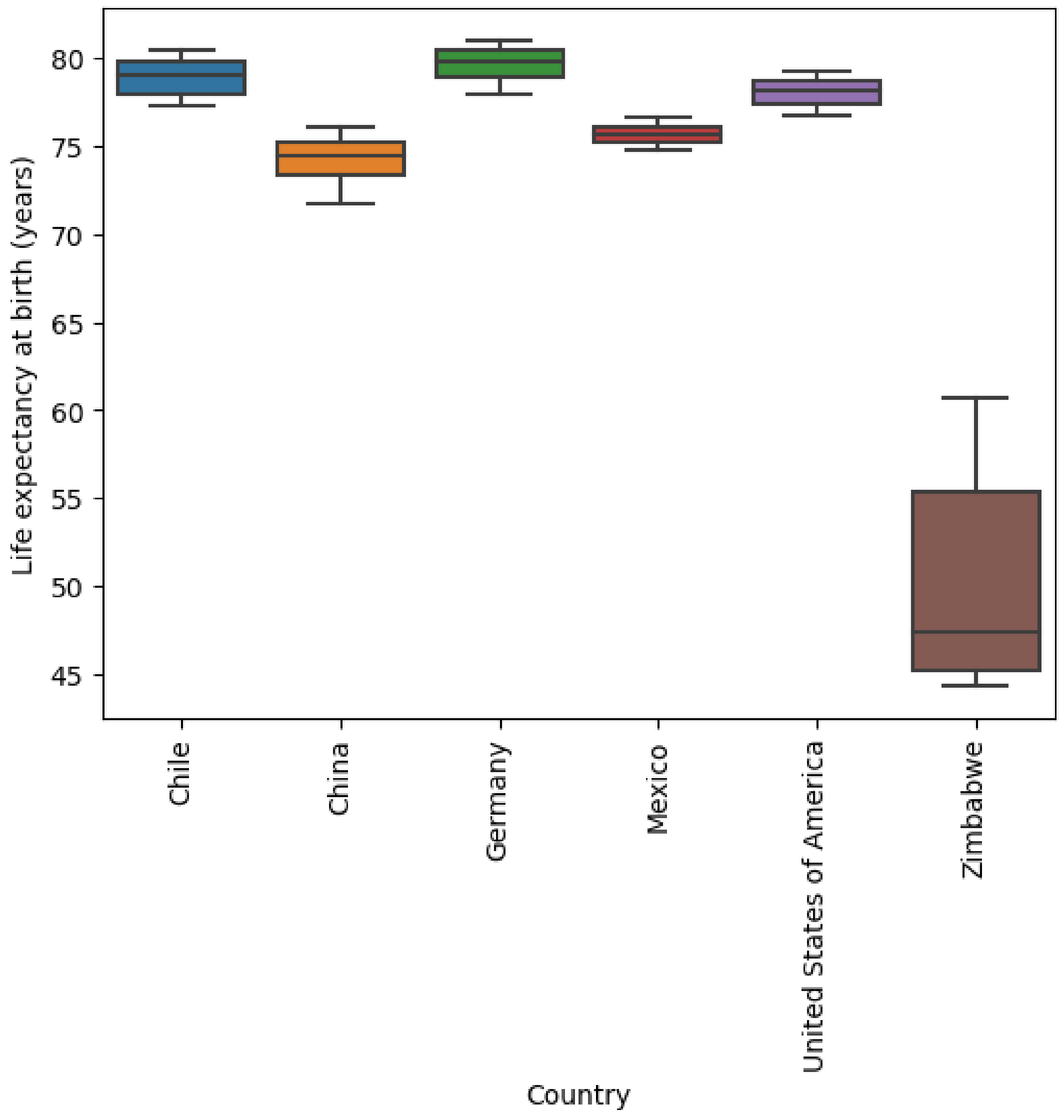
[10]:
```
#Evaluating the association between Country and Life Expectancy
#Getting the DATA
Chile_data2 = wb["Life expectancy at birth (years)"][wb.Country=='Chile']
China_data2 = wb["Life expectancy at birth (years)"] [wb.Country=='China']
Germany_data2 = wb["Life expectancy at birth (years)"] [wb.Country=='Germany']
Mexico_data2 =  wb["Life expectancy at birth (years)"][wb.Country=='mexico']
```

```
USA_data2 =  wb["Life expectancy at birth (years)"] [wb.Country=='United States␣
 ↪of America']
Zimbabwe_data2 =  wb["Life expectancy at birth (years)"] [wb.
 ↪Country=='Zimbabwe']
#Getting the means
mean_Chile2 = np.mean(Chile_data2)
mean_China2 = np.mean(China_data2)
mean_Germany2 = np.mean(Germany_data2)
mean_Mexico2 = np.mean(Mexico_data2)
mean_USA2 = np.mean(USA_data2)
mean_Zim2 = np.mean(Zimbabwe_data2)
#Printing the means
print("CHILE MEAN: " + str(mean_Chile2))
print("CHINA MEAN: " + str(mean_China2))
print("GERMANY MEAN: " + str(mean_Germany2))
print("MEXICO MEAN: " + str(mean_Mexico2))
print("USA MEAN: " + str(mean_USA2))
print("ZIMBABWE MEAN: " + str(mean_Zim2))
```

```
CHILE MEAN: 78.94375
CHINA MEAN: 74.26249999999999
GERMANY MEAN: 79.65625
MEXICO MEAN: nan
USA MEAN: 78.0625
ZIMBABWE MEAN: 50.09375
```

[11]:

```
#Visualizing the means
sns.boxplot(data=wb,x='Country',y="Life expectancy at birth (years)")
plt.xticks(rotation=90)
plt.show()
plt.clf()
```



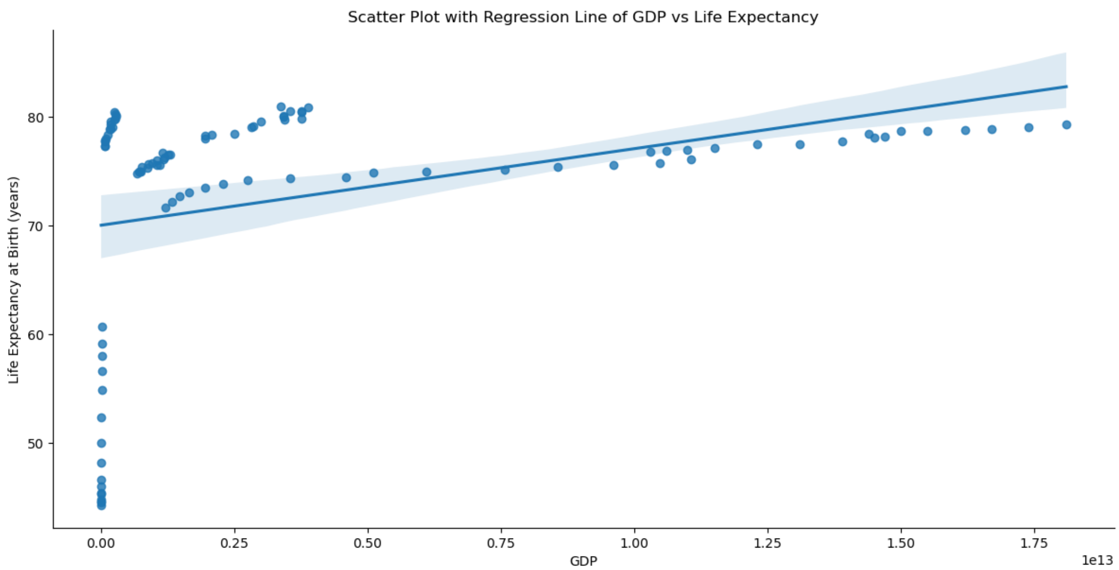

```
<Figure size 640x480 with 0 Axes>
```

The boxplot shows an association between life expectancy and country.Zimbambwe has the lowest life expenctancy and also the lowest GDP, is there a relationship between GDP and Life expectancy?

[25]:
```
#filtered_data = wb[wb["Life expectancy at birth (years)"]>70]
# Assuming 'wb' is your DataFrame with 'GDP' and 'Life expectancy at birth␣
↪(years)' columns
sns.lmplot(x='GDP', y='Life expectancy at birth (years)', data=wb, aspect=2,␣
↪height=6)

# Customize the plot
plt.title('Scatter Plot with Regression Line of GDP vs Life Expectancy')
```

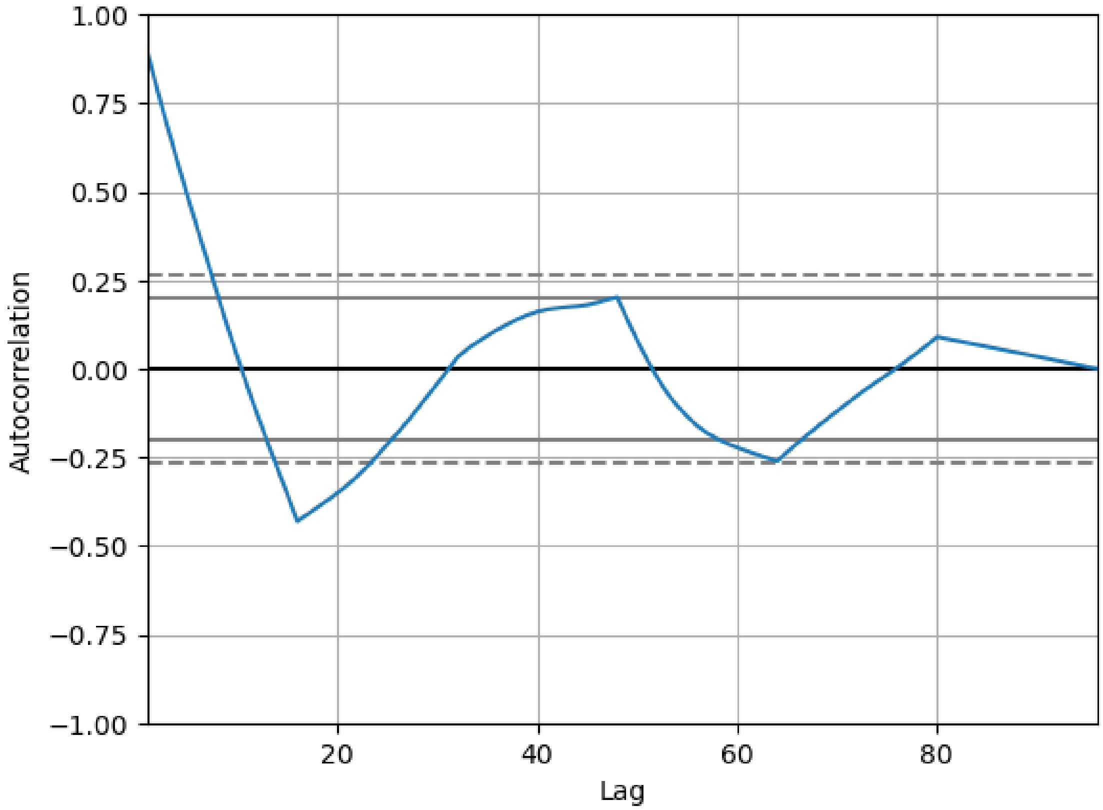
```
plt.xlabel('GDP')
plt.ylabel('Life Expectancy at Birth (years)')
plt.savefig("Scatterplot.png")
plt.show()
```

```
C:\Users\Brian Motee\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



The scatterplot shows that there is a positive relationship between GDP and life expectancy. This means the higher the GDP and the higher the life expectancy hence the life expectancy in USA is expected to be higher that the life Expectancy in Zimbabwe

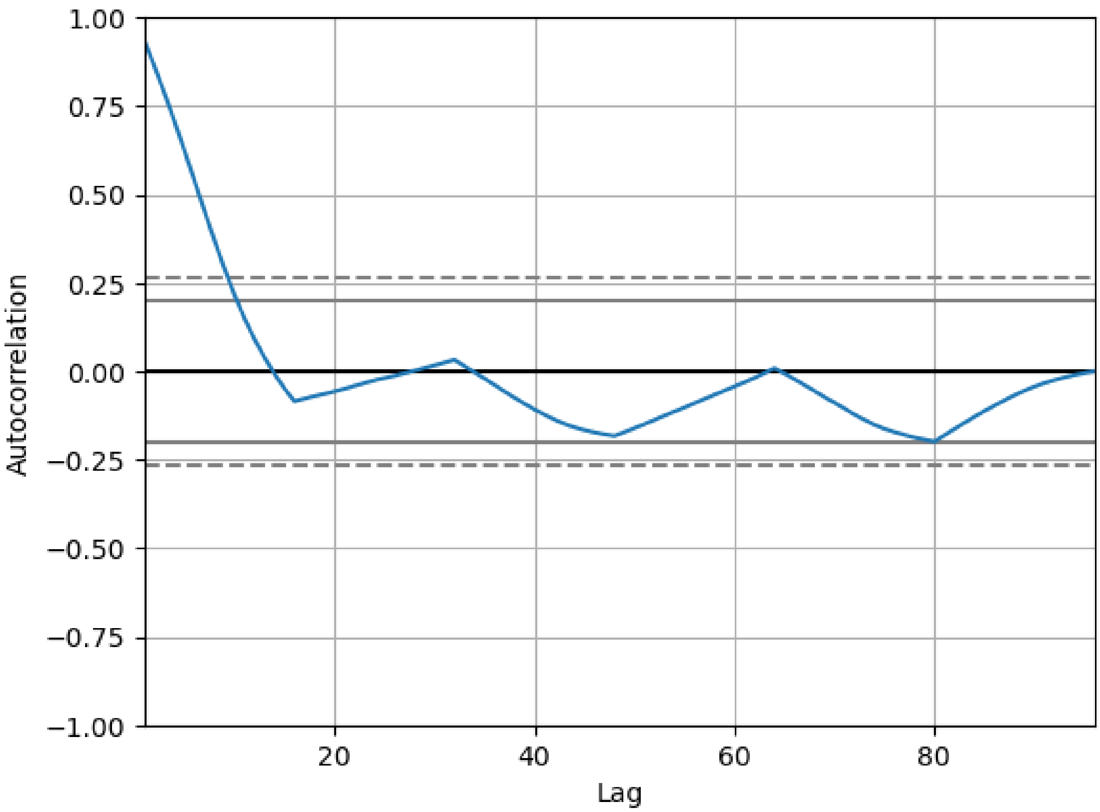
[26]:
```
#Investigating the correlation of GDP with previous years.
#Import autocorrelation function
from pandas.plotting import autocorrelation_plot
#Creating the autocorrelation plot
autocorrelation_plot(wb.GDP)
plt.savefig("Correlationplot.png")
plt.show()
plt.clf()
```




```
&lt;Figure size 640x480 with 0 Axes&gt;
```

The correlation plot shows that high GDP is highly correlated with the previous year high GDP. therefore if the previous year the country had a high GDP it is expected that the current year the country will also have high GDP and Vice versa. The plot also shows that sometime back this relationship was very weak.

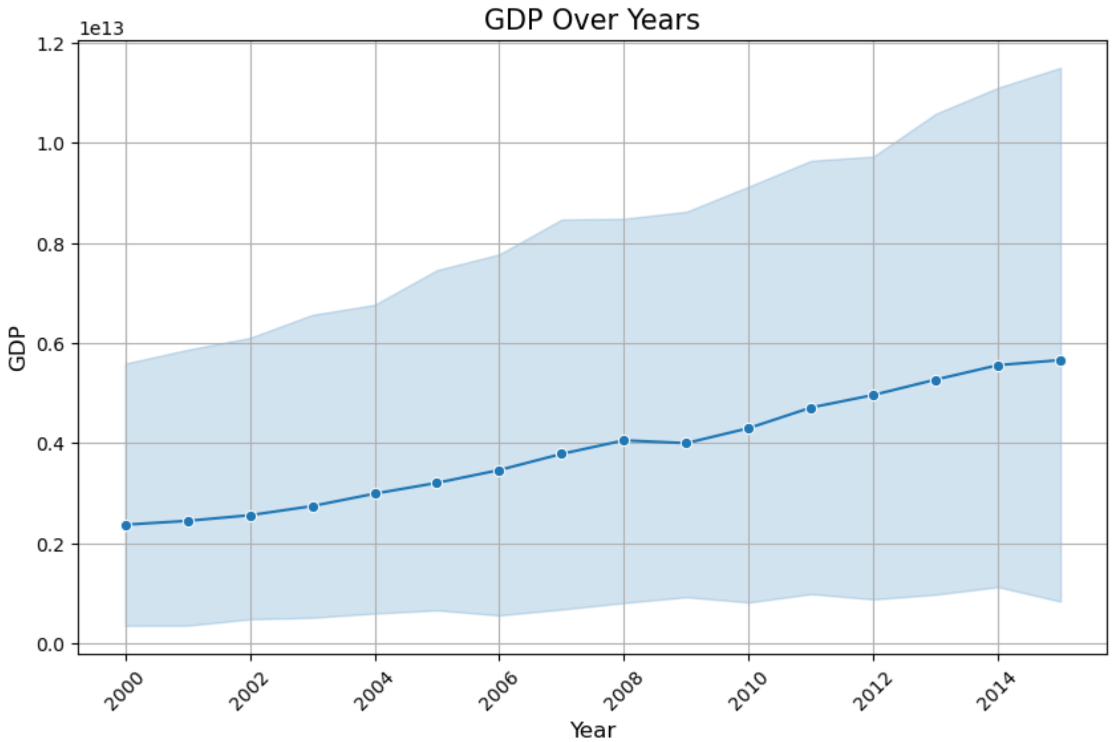
[27]:
```
#Investigating the correlation of life expectancy with previous years.
#Import autocorrelation function
from pandas.plotting import autocorrelation_plot
#Creating the autocorrelation plot
autocorrelation_plot(wb["Life expectancy at birth (years)"])
plt.savefig("Correlationyrs.png")
plt.show()
plt.clf()
```



<Figure size 640x480 with 0 Axes>

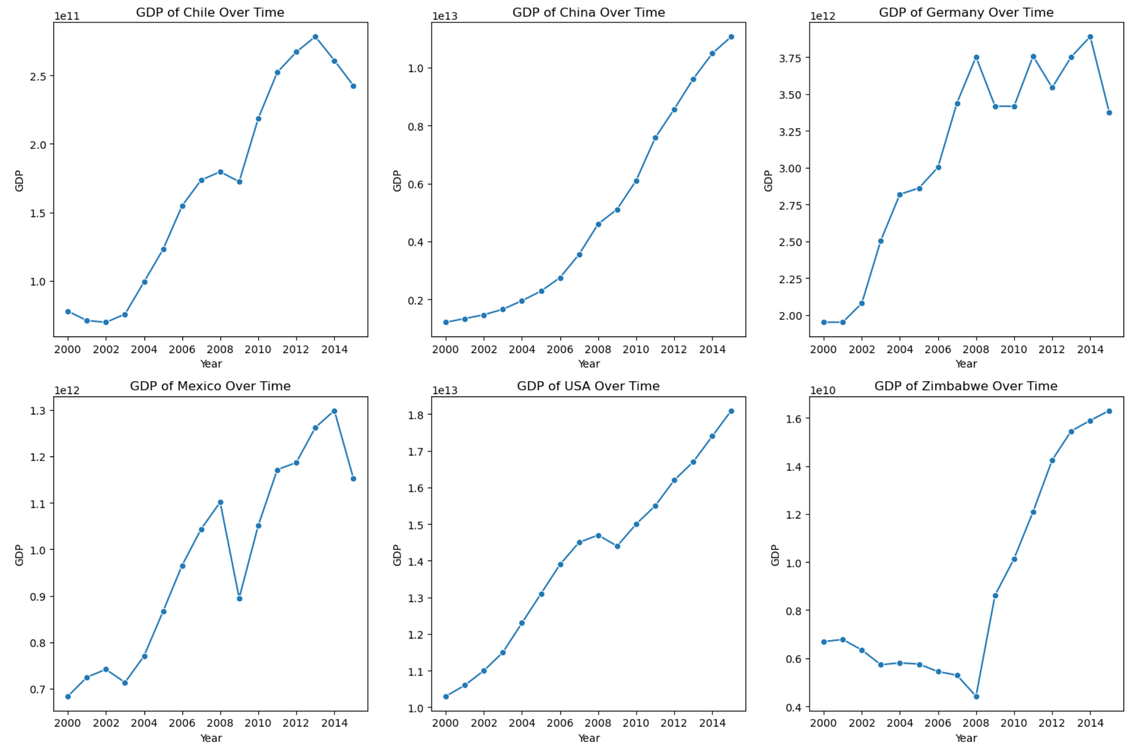
The graph shows that there is a very weak correlation between the previous years life expectancy and the current year life expectancy.

[16]:
```
#General GDP overtime;
plt.figure(figsize=(10, 6))
sns.lineplot(x='Year', y='GDP', data=wb, marker='o')
plt.title('GDP Over Years', fontsize=15)
plt.xlabel('Year', fontsize=12)
plt.ylabel('GDP', fontsize=12)
plt.xticks(rotation=45)  # Rotate x-axis labels if needed
plt.grid(True)
plt.show()
```



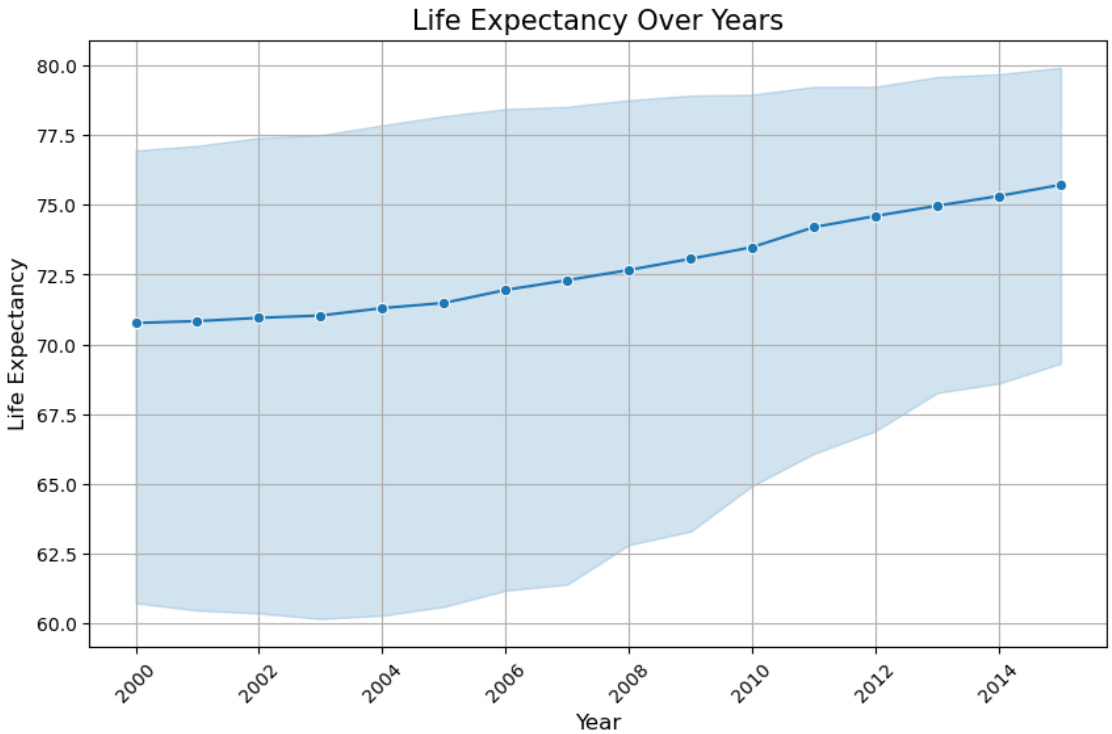
[28]:
```
# Filter data for each country
Chile_Data3 = wb[wb.Country == 'Chile']
China_Data3 = wb[wb.Country == 'China']
Germany_Data3 = wb[wb.Country == 'Germany']
Mexico_Data3 = wb[wb.Country == 'Mexico']
USA_Data3 = wb[wb.Country == 'United States of America']
Zim_Data3 = wb[wb.Country == 'Zimbabwe']

# Set up the figure and subplots
plt.figure(figsize=(15, 10))

# Plot for Chile
plt.subplot(2, 3, 1)
sns.lineplot(x='Year', y='GDP', data=Chile_Data3, marker='o')
plt.title('GDP of Chile Over Time')
plt.xlabel('Year')
plt.ylabel('GDP')

# Plot for China
plt.subplot(2, 3, 2)
sns.lineplot(x='Year', y='GDP', data=China_Data3, marker='o')
plt.title('GDP of China Over Time')
plt.xlabel('Year')
```

```python
plt.ylabel('GDP')

# Plot for Germany
plt.subplot(2, 3, 3)
sns.lineplot(x='Year', y='GDP', data=Germany_Data3, marker='o')
plt.title('GDP of Germany Over Time')
plt.xlabel('Year')
plt.ylabel('GDP')

# Plot for Mexico
plt.subplot(2, 3, 4)
sns.lineplot(x='Year', y='GDP', data=Mexico_Data3, marker='o')
plt.title('GDP of Mexico Over Time')
plt.xlabel('Year')
plt.ylabel('GDP')

# Plot for USA
plt.subplot(2, 3, 5)
sns.lineplot(x='Year', y='GDP', data=USA_Data3, marker='o')
plt.title('GDP of USA Over Time')
plt.xlabel('Year')
plt.ylabel('GDP')

# Plot for Zimbabwe
plt.subplot(2, 3, 6)
sns.lineplot(x='Year', y='GDP', data=Zim_Data3, marker='o')
plt.title('GDP of Zimbabwe Over Time')
plt.xlabel('Year')
plt.ylabel('GDP')

# Adjust layout to prevent overlap
plt.savefig("GDP.png")
plt.tight_layout()
plt.show()
```



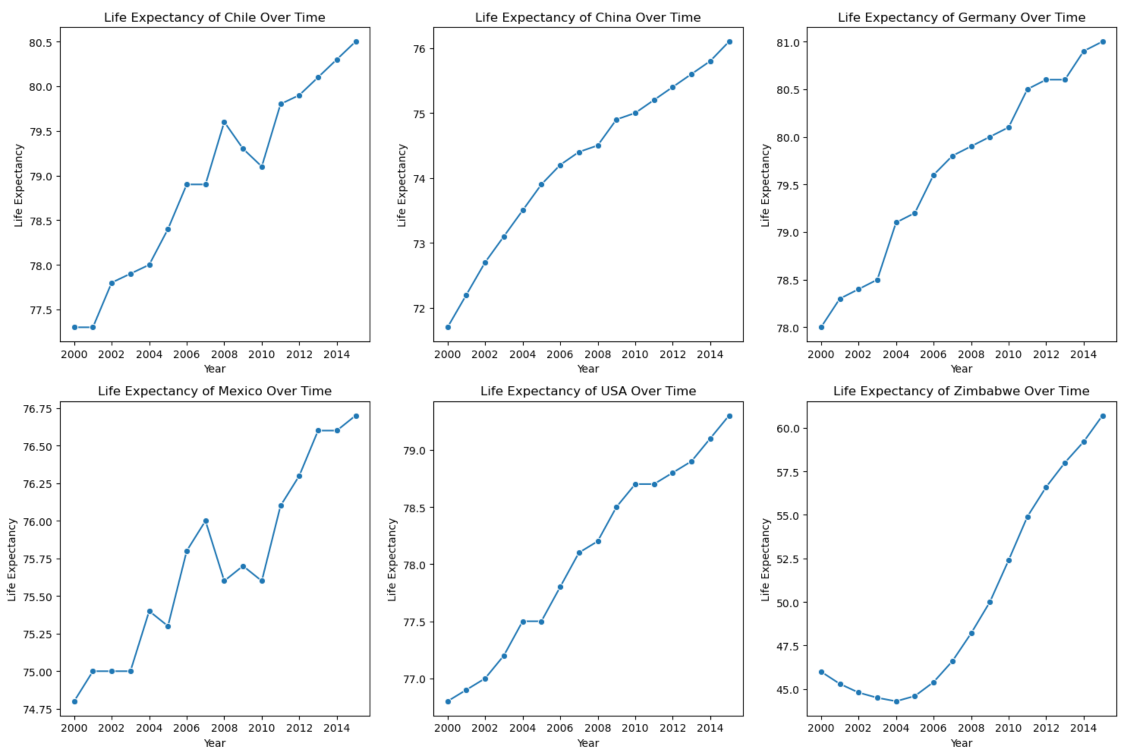
From the graphs,there seem to be a global event that happened between 2006 and 2008 making the GDP of most countries to drop and then oicking up in 2008

[49]:
```
#General Life Expectancy overtime;
plt.figure(figsize=(10, 6))
sns.lineplot(x='Year', y='Life expectancy at birth (years)', data=wb,
 ↪marker='o')
plt.title('Life Expectancy Over Years', fontsize=15)
plt.xlabel('Year', fontsize=12)
plt.ylabel('Life Expectancy', fontsize=12)
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



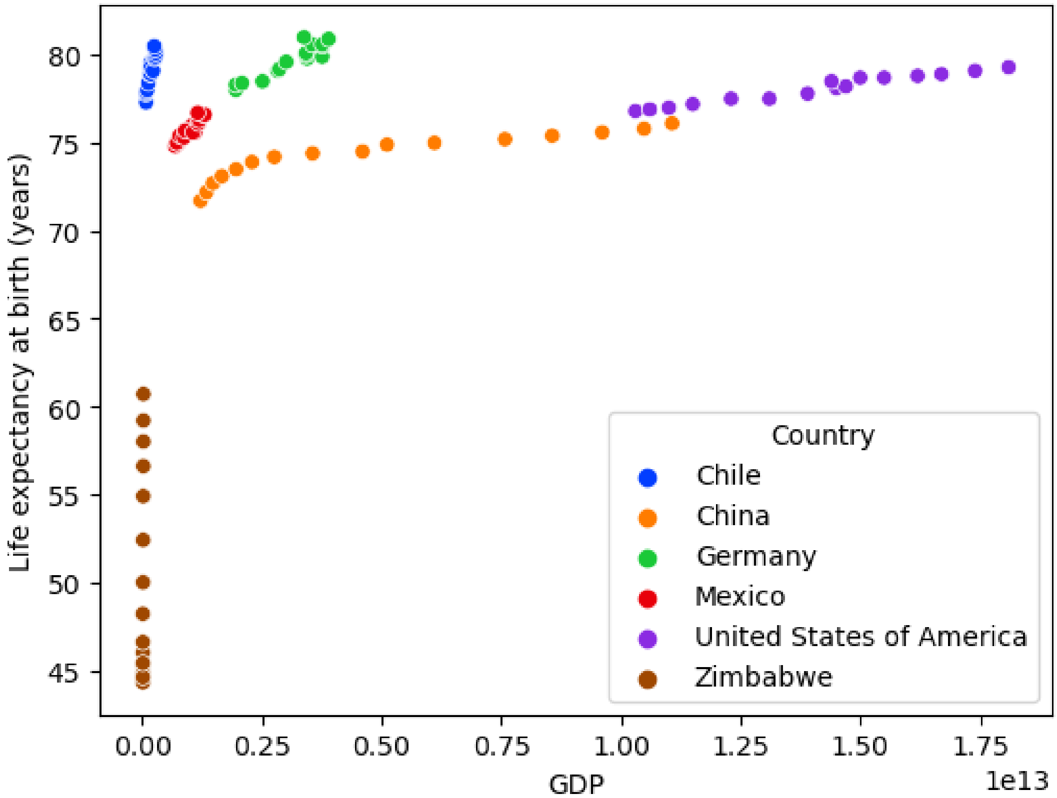
[29]:
```
#Investigating life expectancy of individual countries
# Set up the figure and subplots
plt.figure(figsize=(15, 10))

# Plot for Chile
plt.subplot(2, 3, 1)
sns.lineplot(x='Year', y='Life expectancy at birth (years)', data=Chile_Data3,␣
↪marker='o')
plt.title('Life Expectancy of Chile Over Time')
plt.xlabel('Year')
plt.ylabel('Life Expectancy')

# Plot for China
plt.subplot(2, 3, 2)
sns.lineplot(x='Year', y='Life expectancy at birth (years)', data=China_Data3,␣
↪marker='o')
plt.title('Life Expectancy of China Over Time')
plt.xlabel('Year')
plt.ylabel('Life Expectancy')

# Plot for Germany
plt.subplot(2, 3, 3)
```

```python
sns.lineplot(x='Year', y='Life expectancy at birth (years)',␣
 ↪data=Germany_Data3, marker='o')
plt.title('Life Expectancy of Germany Over Time')
plt.xlabel('Year')
plt.ylabel('Life Expectancy')

# Plot for Mexico
plt.subplot(2, 3, 4)
sns.lineplot(x='Year', y='Life expectancy at birth (years)', data=Mexico_Data3,␣
 ↪marker='o')
plt.title('Life Expectancy of Mexico Over Time')
plt.xlabel('Year')
plt.ylabel('Life Expectancy')

# Plot for USA
plt.subplot(2, 3, 5)
sns.lineplot(x='Year', y='Life expectancy at birth (years)', data=USA_Data3,␣
 ↪marker='o')
plt.title('Life Expectancy of USA Over Time')
plt.xlabel('Year')
plt.ylabel('Life Expectancy')

# Plot for Zimbabwe
plt.subplot(2, 3, 6)
sns.lineplot(x='Year', y='Life expectancy at birth (years)', data=Zim_Data3,␣
 ↪marker='o')
plt.title('Life Expectancy of Zimbabwe Over Time')
plt.xlabel('Year')
plt.ylabel('Life Expectancy')

# Adjust layout to prevent overlap
plt.savefig("Life Expectancy.png")
plt.tight_layout()
plt.show()
```



The GDP and Life Expectancy has been on constant rise overtime

[24]:
```
#plt.figure(figsize=(10,8))
#filtered_data = wb[wb["Life expectancy at birth (years)"]>70]
#Multivariate analysis
sns.scatterplot(x = 'GDP',y= "Life expectancy at birth (years)", hue='Country',data = wb, palette = "bright")
plt.show()
```



The scatter plot shows a positive correlation between GDP and life Expectancy with United States of America having the highest GDP and life expectancy

3 CONCLUSION

In conclusion,GDP and Life Expectancy has been increasing overtime in the 6 nations.

```
[ ]:
```