

# Assignment 5 EDA

Caitlin Wei

2024-10-02

Load the dataset using your preferred programming language (R or Python).

```
import pandas as pd

# Load the dataset
df = pd.read_csv("wdi.csv")

# Show the first few rows of the dataset
df.head()
```

	country	inflation_rate	exports_gdp_share	gdp_growth_rate	gdp_per_capita	adult_literacy_rate
0	Afghanistan	NaN	18.380042	-6.240172	352.603733	NaN
1	Albania	6.725203	37.395422	4.856402	6810.114041	98.5
2	Algeria	9.265516	31.446856	3.600000	5023.252932	NaN
3	American Samoa	NaN	46.957520	1.735016	19673.390102	NaN
4	Andorra	NaN	NaN	9.563798	42350.697069	NaN

Conduct exploratory data analysis on at least three indicators of your choice. Summarise your findings in markdown sections. Show your code and results.

```
# Descriptive statistics for selected indicators
gdp_per_capita_stats = df['gdp_per_capita'].describe()
life_expectancy_stats = df['life_expectancy'].describe()
inflation_rate_stats = df['inflation_rate'].describe()

gdp_per_capita_stats, life_expectancy_stats, inflation_rate_stats
```

```

(count      203.000000
mean       20345.707649
std        31308.942225
min         259.025031
25%        2570.563284
50%        7587.588173
75%        25982.630050
max        240862.182448
Name: gdp_per_capita, dtype: float64,
count      209.000000
mean        72.416519
std         7.713322
min         52.997000
25%         66.782000
50%         73.514634
75%         78.475000
max         85.377000
Name: life_expectancy, dtype: float64,
count      169.000000
mean        12.493936
std         19.682433
min         -6.687321
25%          5.518129
50%          7.967574
75%         11.665567
max         171.205491
Name: inflation_rate, dtype: float64)

```

Create at least two different types of plots (e.g., bar chart, scatter plot) to represent your analysis. Use Quarto code chunks to embed these visualisations. Add a title and axis labels to each plot. Use Quarto to include a caption and a reference to the source of the data. Hide your code in the final document.

```

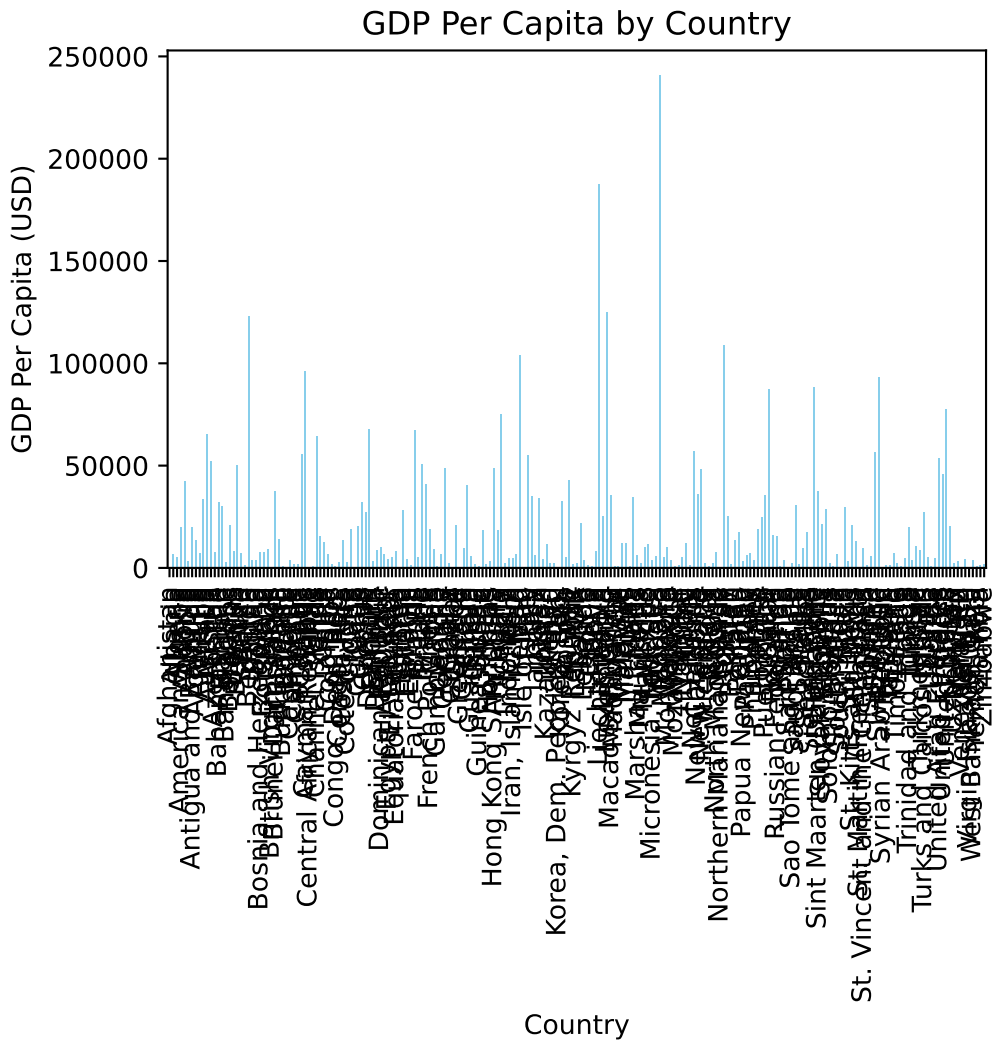
import matplotlib.pyplot as plt

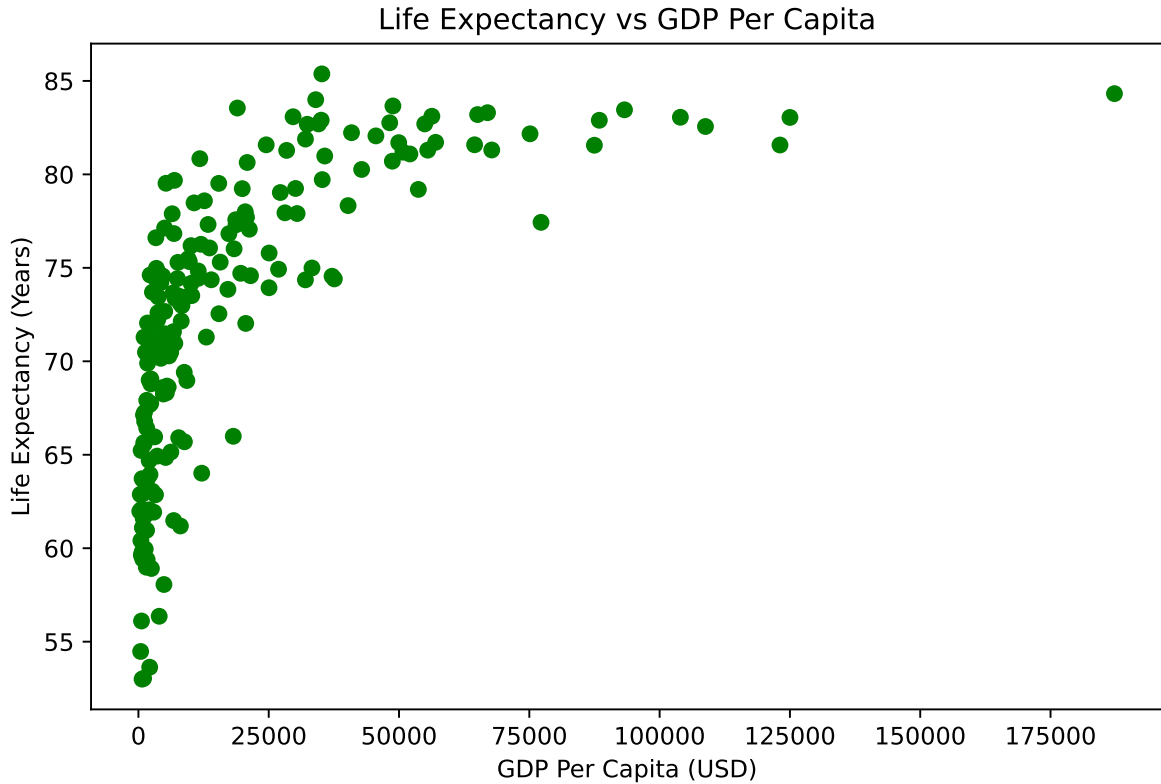
# Bar chart for GDP per capita
plt.figure(figsize=(8,5))
df.plot(kind='bar', x='country', y='gdp_per_capita', legend=False, color='skyblue')
plt.title('GDP Per Capita by Country')
plt.xlabel('Country')
plt.ylabel('GDP Per Capita (USD)')
plt.show()

```

```
# Scatter plot for Life Expectancy vs GDP per Capita
plt.figure(figsize=(8,5))
plt.scatter(df['gdp_per_capita'], df['life_expectancy'], color='green')
plt.title('Life Expectancy vs GDP Per Capita')
plt.xlabel('GDP Per Capita (USD)')
plt.ylabel('Life Expectancy (Years)')
plt.show()
```

<Figure size 2400x1500 with 0 Axes>





Construct a table that highlights some key statistics from your analysis. Ensure the table is well-formatted and included in the report.

```
# Table with key statistics
summary_table = df[['gdp_per_capita', 'life_expectancy', 'inflation_rate']].describe()
summary_table
```

	gdp_per_capita	life_expectancy	inflation_rate
count	203.000000	209.000000	169.000000
mean	20345.707649	72.416519	12.493936
std	31308.942225	7.713322	19.682433
min	259.025031	52.997000	-6.687321
25%	2570.563284	66.782000	5.518129
50%	7587.588173	73.514634	7.967574
75%	25982.630050	78.475000	11.665567
max	240862.182448	85.377000	171.205491

As shown in **Figure 1**, the GDP per capita varies widely across countries. **Table 1** summarizes the key statistics for GDP per capita, life expectancy, and inflation rate.

see references (Zaman et al. 2017)

Zaman, Sojib Bin, Naznin Hossain, Varshil Mehta, Shuchita Sharmin, and Shakeel Ahmed Ibne Mahmood. 2017. “An Association of Total Health Expenditure with GDP and Life Expectancy.” *Journal of Medical Research and Innovation* 1 (2): AU7–12.