report

Kei Nie

2025-02-21

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

```
import pandas as pd
df = pd.read_csv("wdi.csv")
print(df.head())
```

```
          country  inflation_rate  exports_gdp_share  gdp_growth_rate  \
0     Afghanistan             NaN          18.380042        -6.240172
1         Albania        6.725203          37.395422         4.856402
2         Algeria        9.265516          31.446856         3.600000
3  American Samoa             NaN          46.957520         1.735016
4         Andorra             NaN                NaN         9.563798

   gdp_per_capita  adult_literacy_rate  primary_school_enrolment_rate  \
0      352.603733                  NaN                            NaN
1     6810.114041                 98.5                      95.606712
2     5023.252932                  NaN                     108.343933
3    19673.390102                  NaN                            NaN
4    42350.697069                  NaN                      90.147346

   education_expenditure_gdp_share  measles_immunisation_rate  \
0                              NaN                       68.0
1                          2.74931                       86.0
2                              NaN                       79.0
3                              NaN                        NaN
4                          2.66623                       98.0
```

```
   health_expenditure_gdp_share  income_inequality  unemployment_rate  \
0                           NaN                NaN             14.100
1                           NaN                NaN             11.588
2                           NaN                NaN             12.437
3                           NaN                NaN                NaN
4                           NaN                NaN                NaN

   life_expectancy  total_population
0           62.879        41128771.0
1           76.833         2777689.0
2           77.129        44903225.0
3              NaN           44273.0
4              NaN           79824.0
```

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

```
summary = df[['gdp_per_capita', 'life_expectancy', 'unemployment_rate']].describe()
print(summary)
```

```
       gdp_per_capita  life_expectancy  unemployment_rate
count      203.000000       209.000000         186.000000
mean     20345.707649        72.416519           7.268661
std      31308.942225         7.713322           5.827726
min        259.025031        52.997000           0.130000
25%       2570.563284        66.782000           3.500750
50%       7587.588173        73.514634           5.537500
75%      25982.630050        78.475000           9.455250
max     240862.182448        85.377000          37.852000
```

The analysis of GDP per capita, life expectancy, and unemployment rate from the dataset reveals significant economic and social disparities among countries. GDP per capita varies widely, with an average of $20,345 but a high standard deviation ($31,308), indicating strong inequality. Life expectancy is more stable, averaging 72.42 years with most countries falling between 66.78 and 78.48 years. Unemployment rates show substantial variation, with a mean of 7.27%, but extremes range from 0.13% to 37.85%, reflecting diverse labor market conditions globally. These findings highlight major global inequalities in economic prosperity, health, and employment opportunities.

4. 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.

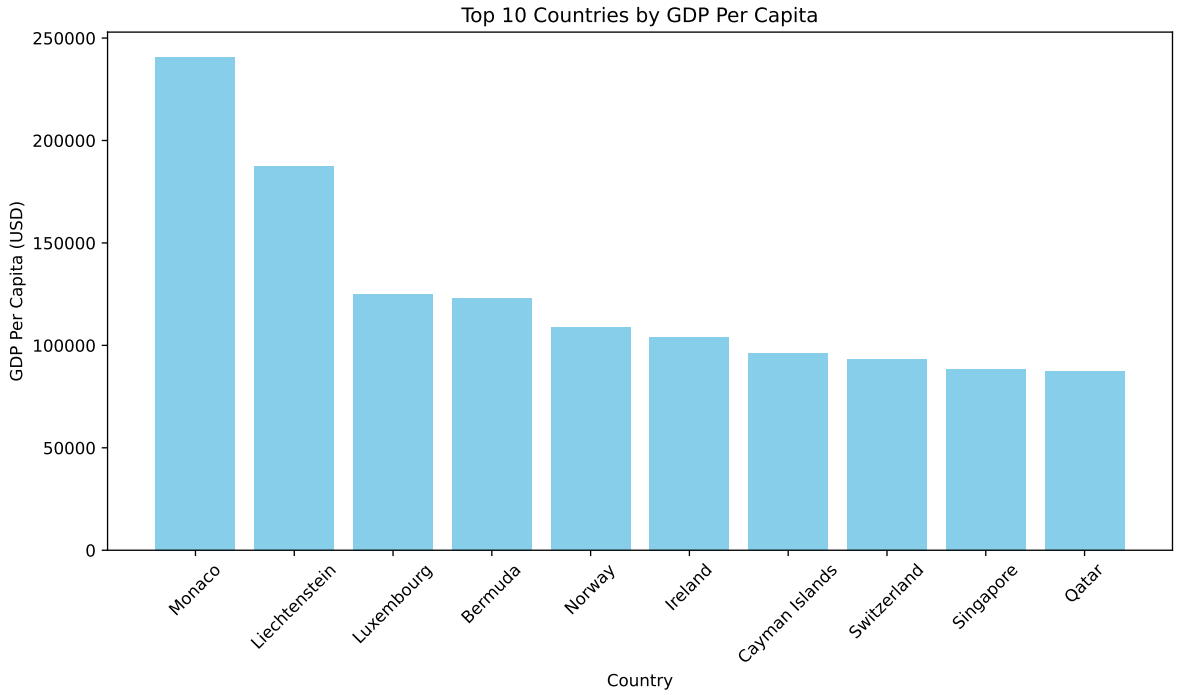

Figure 1: Bar chart showing the top 10 countries by GDP per capita (data from World Development Indicators).

Figure 2: Scatter plot illustrating the relationship between life expectancy and unemployment rate.

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

	index	gdp_per_capita	life_expectancy	unemployment_rate
0	count	203.000000	209.000000	186.000000
1	mean	20345.707649	72.416519	7.268661
2	std	31308.942225	7.713322	5.827726
3	min	259.025031	52.997000	0.130000
4	25%	2570.563284	66.782000	3.500750
5	50%	7587.588173	73.514634	5.537500

	index	gdp_per_capita	life_expectancy	unemployment_rate
6	75%	25982.630050	78.475000	9.455250
7	max	240862.182448	85.377000	37.852000

7. Add a bibliography using BibTeX (.bib). Cite at least two sources related to your analysis.

Several sources have analyzed economic trends (Bank 2020; Provost and Fawcett 2013).

Bank, World. 2020. "World Development Indicators." *World Bank Data*. https://data.worldbank.org/.

Provost, Foster, and Tom Fawcett. 2013. *Data Science for Business*. O'Reilly Media.