



ASSIGNMENT 1 DATA VISUALISATION

Implemented in fulfillment of Bootcamp in Data Technician



APRIL 19, 2024

TAIWO OLUWAFEMI

TABLE OF CONTENT

1.0	Policies and Procedures.	1
2.0	Importance of Data Policies and Procedure adherence	1
3.0	Importance of Data Analyst awareness to Data Policies.....	2
4.0	Task Overview	2
5.0	About the Data	2
6.0	Data Observation	3
7.0	Measures of Central Tendency	6
8.0	Measure of dispersion.....	6
9.0	Correlation Analysis.....	7
10.0	Data Implication.....	8
11.0	Data Prediction	8
12.0	Data Limitation	9
13.0	Data Protection.....	9

1.0 Policies and Procedures.

It is crucial to ensure data integrity, security, privacy and legal compliance when data is being worked on. The following policies and procedure will ensure adherence to these standard;

- Data Privacy policy: It specify individual right with data and how personal identifiable information and sensitive data are collected, processed, stored and shared.
- Data Security policy: This policy establish procedures to prevent unauthorized access to data and restricts data access base on roles and permissions.
- Ethical Data use policy: This presents an acceptable way to share and collaborate with data while upholding intellectual properties and confidentiality agreement.
- Legal compliance policy: It defines the necessary processes required to obtain consent and permission before processing or analyzing data.
- Data Retention policy: The policy outline how data should be deleted or anonymized when no more needed.

2.0 Importance of Data Policies and Procedure adherence

In this work, it is important to adhere to the policies listed above to ensure compliance, accuracy and trustworthiness of the analysis activities. Accuracy means that, the data collected and used for this work represents the real-world realities it represents, these reduces the possibility of error and mis-interpretation of the results that can lead to unreliable insights and decision. Also, compliance to the data collection, usage, sharing, and collaboration policy while also respecting intellectual property rights will avoid any litigation issues against un-ethical use of data. Last, the integrity of the data will heighten the confidence level and trustworthiness of the conclusions and recommendations derived from the analysis.

3.0 Importance of Data Analyst awareness to Data Policies

Essentially, in order to meet stakeholder's expectation and ensure customer satisfaction entirely, data analysts must be aware of these rules, policies and procedure to enable them the work within the boundaries of established blueprints.

4.0 Task Overview

According to [finance and development](#), GDP (Gross domestic product) measures the monetary value of final goods and services—that is, those that are bought by the final user—produced in a country in a given period of time (say a quarter or a year). It counts all of the output generated within the borders of a country. GDP is composed of goods and services produced for sale in the market and also includes some nonmarket production, such as defense or education services provided by the government. The task is to analyze the GDP of some countries provided in the dataset collected and presents meaningful insights with data visualizations.

5.0 About the Data

The dataset used for the task as provided contains data between years 2002 to 2020, for 228 countries in the world. This means that, there are 228 rows (records) in the dataset. However, the task is limited only to data in the year 2019. The variables included in the dataset are;

- Rank
- Country
- GDP - per capita (PPP)
- Year of Information

Then, from the original dataset collected, the currency symbol was changed from dollar (\$) to Great Britain Pounds (£). After that, the entire dataset was filtered to 2019 data.

only, so as to limit the analysis to this year of information according to the requirement. On plotting the histogram figure 4, for the distribution Life expectancy of top 15 countries, the range of values in table 1 below was used for each bin.

S/N	Class Limit
1	80.0-80.5
2	80.6-81.0
3	81.1-81.5
4	81.6-82.0
5	82.1-82.5
6	82.6-83.0
7	83.1-83.5
8	83.5-84.0
9	84.1-84.5
10	84.6-85.0

Table 1. Histogram class limit

Furthermore, GDP and life expectancy tables were merged, and two new calculated normalized columns were also created. These enabled the calculation of correlation between GDP and life expectancy, plotting and visualizing these correlation or relationship using scatter plot and line chart.

6.0 Data Observation

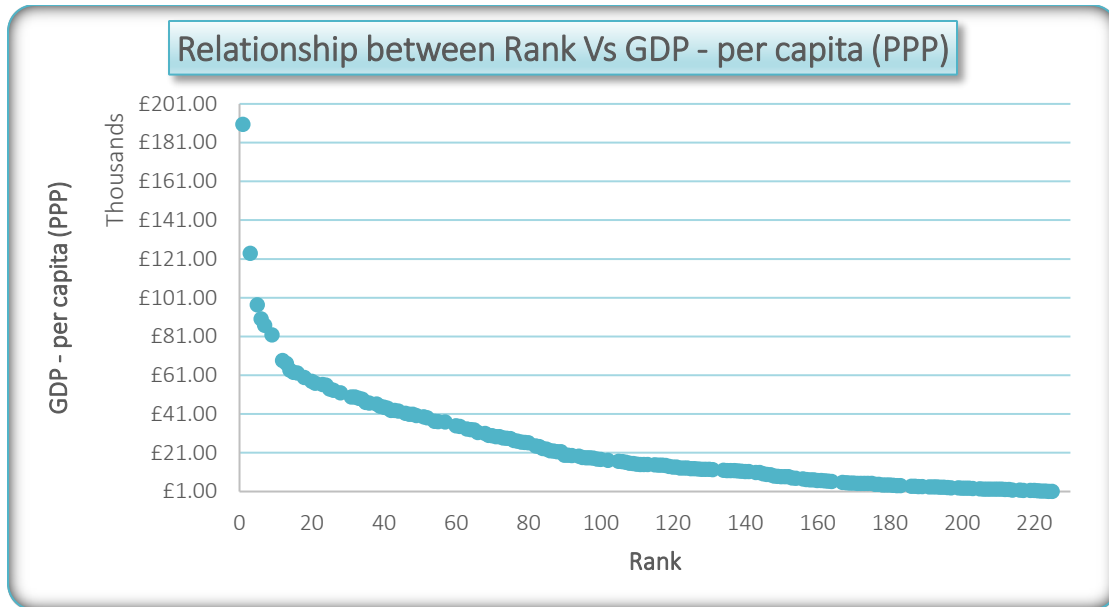


Figure 1. Relationship between Rank and GDP – per capita (PPP)

- It was observed in the chart in figure 1 that, as Rank value increases, there is a decrease in GDP - per capita (PPP).
- In 2019, Monaco has the highest GDP - per capita (PPP) with the value of £190,513.00 and a Rank of 1, while Burundi has the least GDP -per capita with a value of £752.00 with a Rank of 228.

Rank	Country	GDP - per capita (PPP)	Year of Information
1	Monaco	£190,513.00	2019
3	Macau	£123,965.00	2019
5	Singapore	£97,341.00	2019
6	Qatar	£90,044.00	2019
7	Ireland	£86,781.00	2019

9	Bermuda	£81,798.00	2019
12	Switzerland	£68,628.00	2019
13	United Arab Emirates	£67,119.00	2019
14	Norway	£63,633.00	2019
15	United States	£62,530.00	2019
16	Brunei	£62,100.00	2019
18	Hong Kong	£59,848.00	2019
20	Denmark	£57,804.00	2019
21	Netherlands	£56,935.00	2019
23	Austria	£56,188.00	2019
24	Iceland	£55,874.00	2019
25	Germany	£53,919.00	2019
26	Sweden	£53,240.00	2019
28	Belgium	£51,934.00	2019
31	Australia	£49,854.00	2019
32	Kuwait	£49,854.00	2019

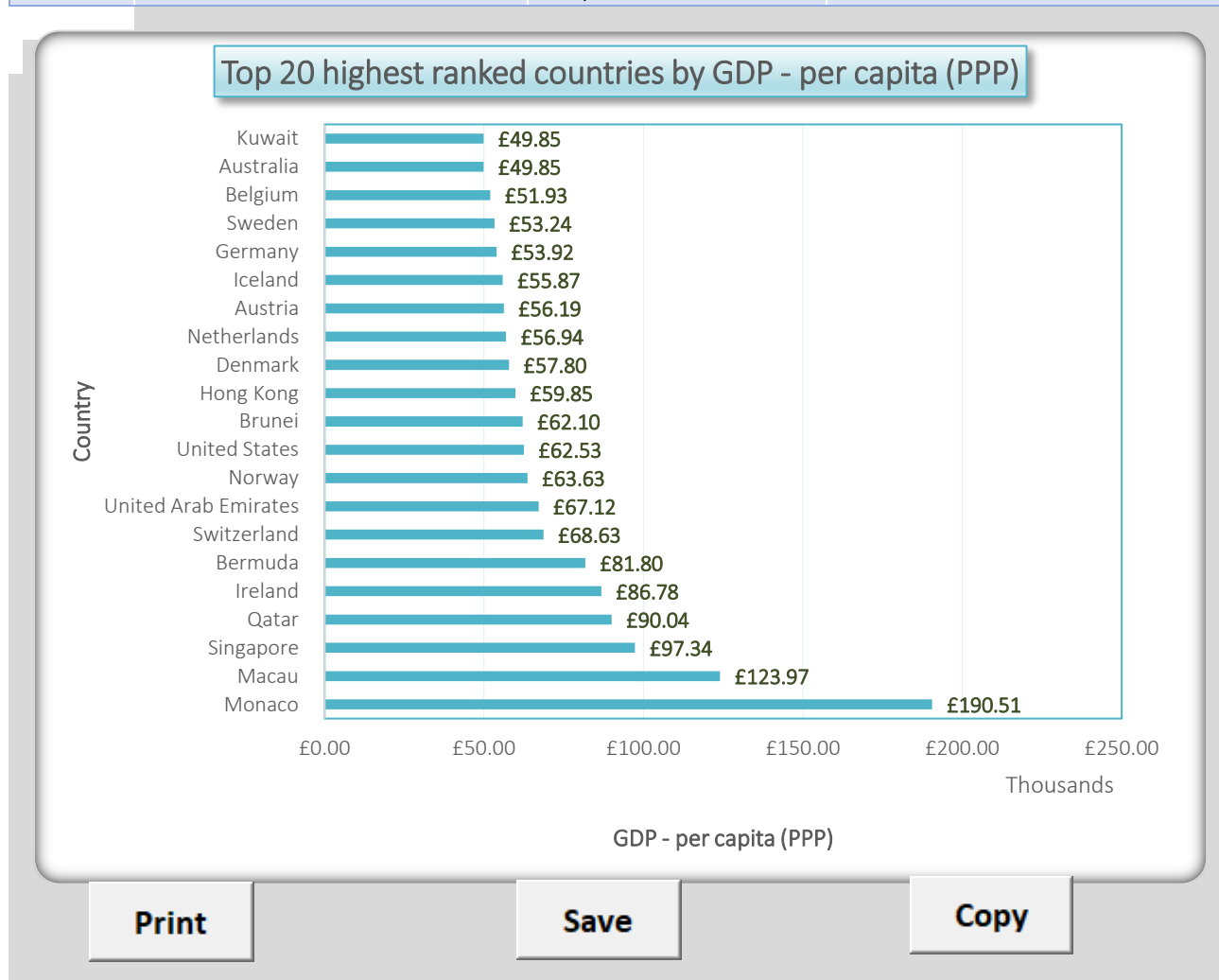


Figure 2. Relationship between Rank and GDP – per capita (PPP)

- According to the chart in figure 2. The country with the highest GDP - per capita in 2019 among the top 20 countries is Monaco.
- Kuwait and Australia has the least and same GDP - per capita in 2019 among the top 20 countries, with the value of £49,854.00. As a result of these, there are 21 countries in the list of top 20 countries.
- Also, so as to ease the repetitive copy function from the excel document, a macro was recorded on the copy operation and assigned to the copy button.

7.0 Measures of Central Tendency

- Average life expectancy across top 15 countries is 83.0
- The median life expectancy is 83.0
- The average and median life expectancy on the top 15 countries are same value.

8.0 Measure of dispersion

- The range of life expectancies across top 15 countries is 2.5
- The standard deviation of life expectancies is 7.3
- The chart in figure 3 shows the distribution of top 15 countries.

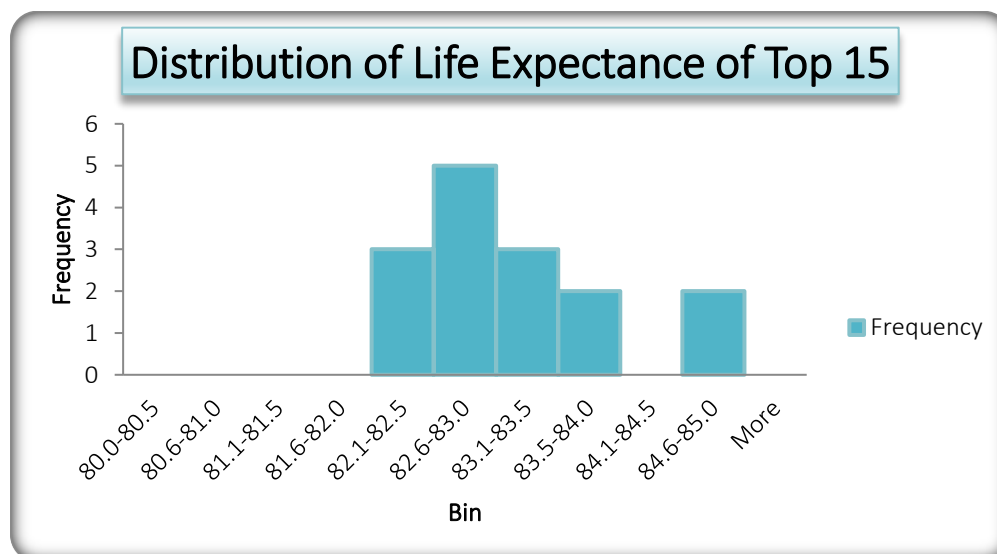


Figure 3. Distribution life expectancy of top 15 countries

9.0 Correlation Analysis

The calculated correlation of coefficients between Life expectancy at birth and GDP - per capita (PPP) is **0.66**. This means that there is a potential relationship between life expectancy and GDP of a country (i.e. life expectancy increase as the GDP grows). The image in figure 4 and figure 5 visualizes the relationship between these variables

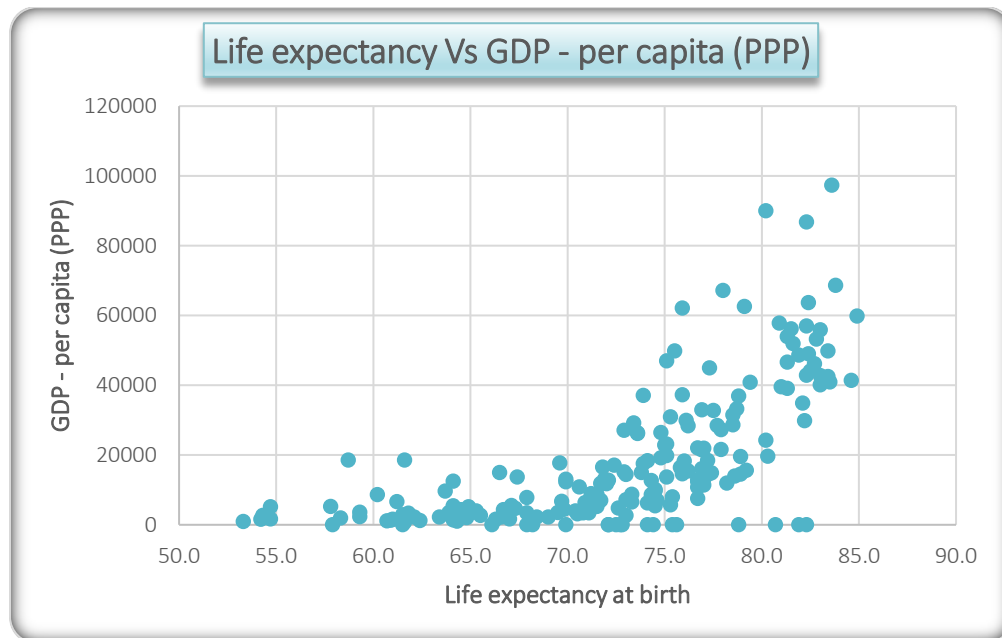


Figure 4. Relationship between Rank and GDP – per capita (PPP)

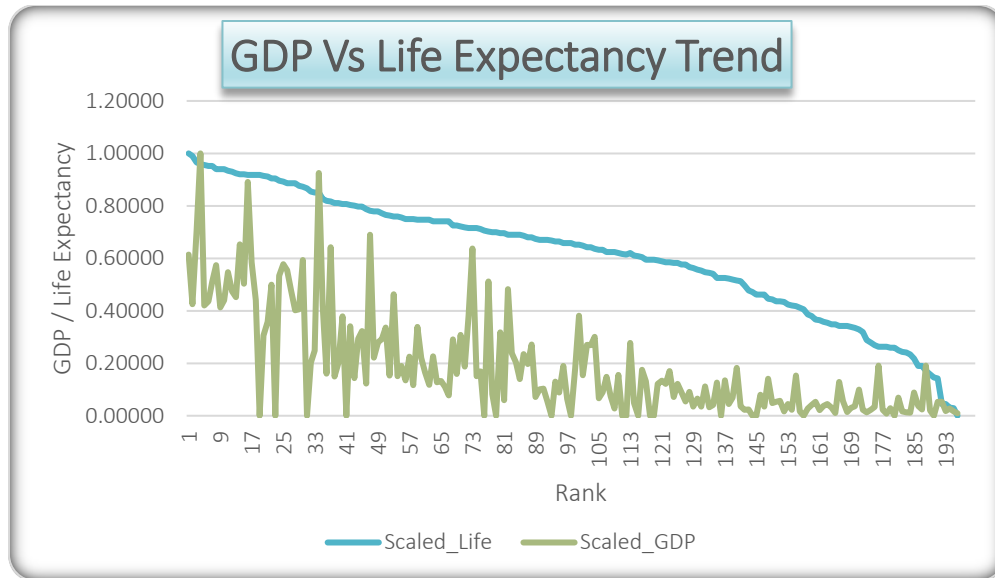


Figure 5. Trend analysis between Life expectancy and GDP – per capita (PPP)

10.0 Data Implication

The potential difference of life expectancy shows the dependences or relationship it has with GDP. This means that, according to trend analysis in figure 5 above, a country with a fall in GDP in a year, may possibly also experience a decrease of life expectancy at birth of children born in that same year.

As a result of these insight, in a country where the GDP is predicted to experience a fall or a decrease, the public health agencies will be well informed with facts that can influence policy formation or adjustment of existing policies on population or birth control in that particular year in other reduce the impact on birth in the year.

11.0 Data Prediction

Base on trend uncovered in the data, we not predict the future changes in life expectancy base on the changes in the GDP. However, such prediction cannot be establish on the countries in the dataset because of insufficient data. Predicting future changes in life expectancy will require a considerable large amount of yearly life

expectancy sample data of each countries, to uncover trend or pattern in such data, after that, use such pattern to make prediction on the future value for each countries.

12.0 Data Limitation

The major limitation on relying solely on life expectancy data as a measure of overall health is the insufficient volume of the data, because, there are countries that the GDP was not provided (e.g. Venezuela, Saint Vincent and the Grenadines etc.) which can have a significant effect on the correlation analysis and also the conclusion derived from the analysis.

For a more comprehensive understanding of public health, more organic data with more indicators or variables needs to be collected from the appropriate agencies or authorities in the each countries, or leverage on available public related dataset with the right authorization.

13.0 Data Protection

To ensure the compliance of the security policy, the update entire workbook is passworded to guarantee access to only authorized users. The password used to protect the updated workbook will be sent directly to the required users.