



Wealth of Nations

Excel and Tableau Assignment

Tamima Tarofdar

Data Technician Skills Bootcamp - Just IT (April 2024)



Table of Contents

Table of Contents.....	2
Scenario.....	3
Policies and Procedures.....	4
Excel.....	6
Tableau.....	11
Conclusions.....	19
Evaluation.....	20

Scenario

- - - - X

Data visualisation has become an essential business capability to help transform information into insights that can drive meaningful business outcomes and improved experiences. Today, most organisations have accumulated a wealth of data from the different corners of their businesses they are then unable to see how this data can help them make better decisions, making actions, and results.

You have been asked to look at the data workbook and familiarise yourself with this data. You have also been asked to create a visual report that will show the data in the form of charts and maps using Tableau to the client's requirements. You will also need to consider data protection and computer misuse policies.

Task 1 - Data Policies and Procedures

- - - - X

In an increasingly digital world, data protection policies are fundamental for maintaining the integrity and security of data. It ensures information is used ethically and transparently. By adhering to laws like GDPR (EU) and the Data Protection Act (UK), such policies safeguard the rights and interests of individuals and organisations.

They ensure that various aspects of data protection are considered when handling data which include the following:

- 1) **Encryption Policies** dictate when and how data should be encrypted, whether at rest or in transit, to prevent unauthorised access.
- 2) **Acceptable Use Policies** outline permissible and prohibited actions concerning data usage by employees, preventing misuse and safeguarding organisational reputation and productivity.
- 3) **Password Policies** govern password management, enhancing data security through requirements like length, complexity, and update frequency.
- 4) **Data Retention Policies** establish protocols for determining data retention periods and proper disposal methods to manage the data lifecycle efficiently.
- 5) **Data Breach Policies** outline procedures and responsibilities in the event of a data breach, including notifying authorities and affected parties.

For data analysts, understanding and adhering to these policies are crucial because they:

- Ensure compliance with legal and ethical obligations.
- Maintain data reliability and validity in analysis.
- Safeguard data confidentiality, integrity, and availability.
- Foster trust, transparency, and accountability in data practices.
- Minimise potential harm from data misuse or breaches.

When analysing datasets like "The Wealth of Nations," which contains structured data in an Excel format without personal or sensitive information, adherence to relevant legislation such as the Computer Misuse Act 1990, GDPR, and Data Protection Act 2018 is still essential. Even though it's considered open data, legal frameworks must be respected to prevent unlawful access, misuse, or modification without consent.

For data analysts handling sensitive or personal data, ethical responsibilities are paramount. They must ensure data security, prevent breaches, and prioritise privacy and fairness in data handling. By upholding data protection policies, analysts contribute to building public trust in data usage and protection.

Task 2 - Excel

- - - - X

As I was using the online web version of Microsoft Excel, the password protection functionality for the entire workbook was unavailable. Therefore, I decided to password-protect each sheet individually to ensure the data was secure.

To do this I navigated to the ‘Review’ ribbon in Excel and selected ‘Manage Protection’. As shown in Figure 1, a menu appeared on the right side of the screen and I moved the toggle underneath ‘Protect Worksheet’ to On.

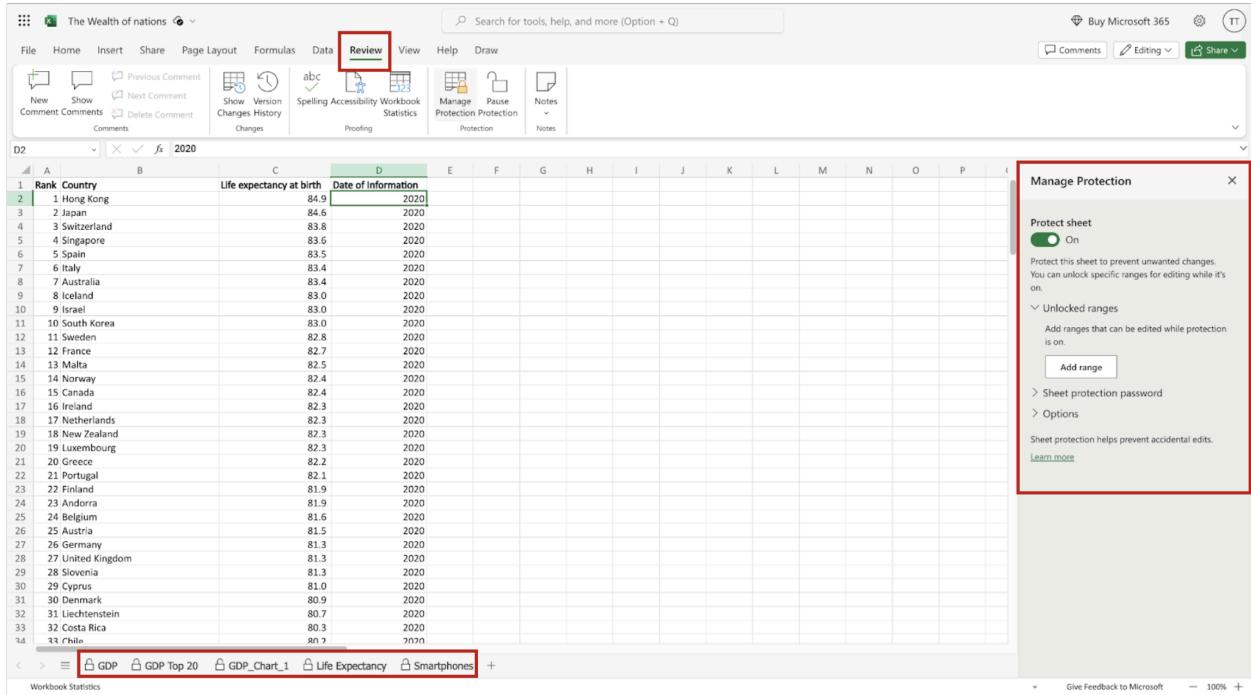


Figure 1

Each protected sheet was signified with a small lock symbol next to its name on the sheet tab at the bottom of the document window (also shown in Figure 1).

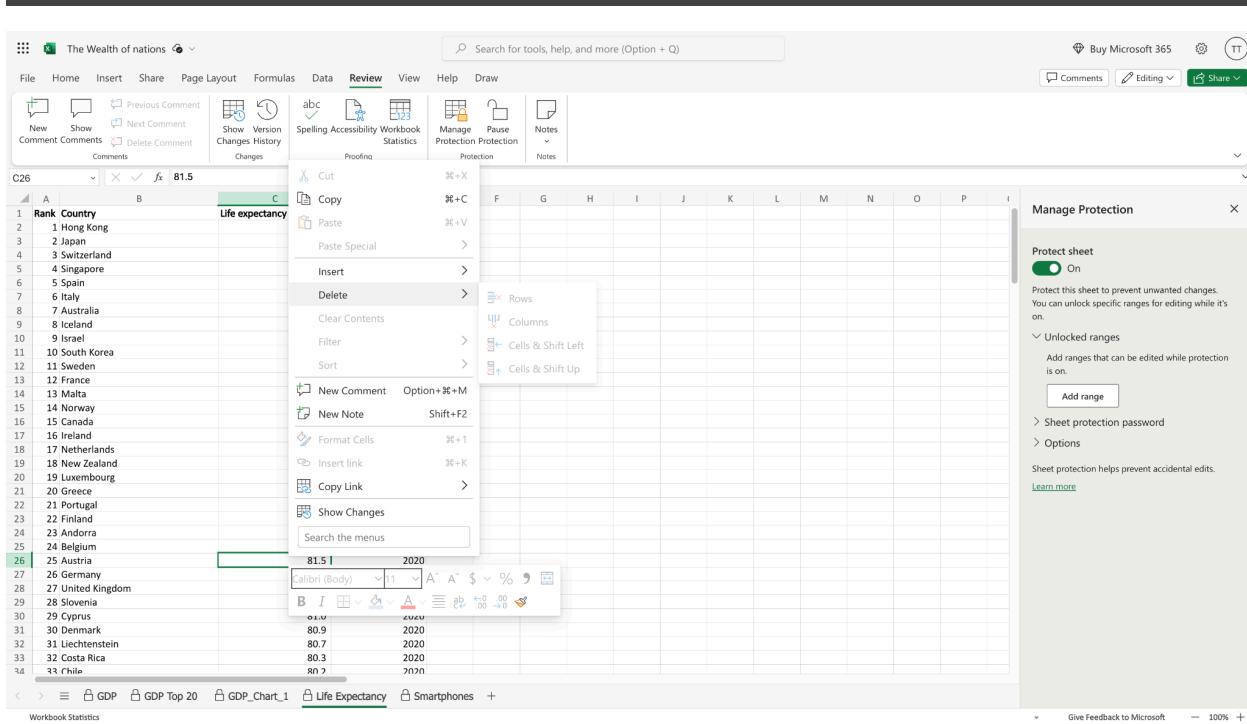


Figure 2

As shown in Figure 2, by protecting the worksheet a viewer cannot make changes such as deleting or inserting data. These options have been greyed out.

Rank	Country	GDP - per capita (PPP)	Year of Information
1	Monaco	\$190,513.00	2019
2	Liechtenstein	\$180,367.00	2018
3	Macau	\$123,965.00	2019
4	Luxembourg	\$115,874.00	2020
5	Singapore	\$97,341.00	2019
6	Qatar	\$90,044.00	2019
7	Ireland	\$86,781.00	2019
8	Isle of Man	\$84,600.00	2014
9	Bermuda	\$81,798.00	2019
10	Cayman Islands	\$71,549.00	2018
11	Falkland Islands	\$70,800.00	2015

Figure 3

Before analysis, I needed to make it clear that the data for GDP in column C was monetary and in the correct currency. Therefore, I highlighted the column and changed the data to display the British Pound symbol as shown in Figure 3.

To analyse the data efficiently I had to transform the data in the GDP sheet into a table. As shown in Figure 4, I navigated to the ‘Insert’ ribbon, highlighted the relevant data and formatted the data as a table to include headers.

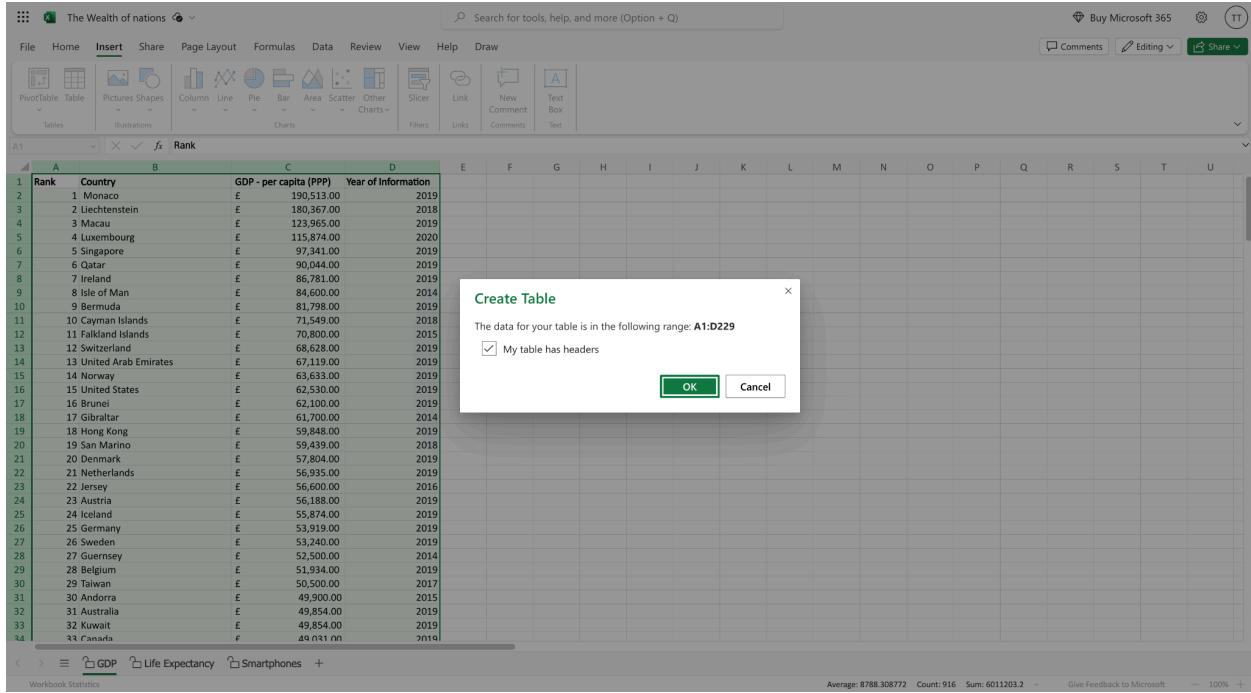


Figure 4

I needed the table to display data from 2019 only. To do this I used the filter button in the ‘Year of Information’ header and selected 2019 tickbox as illustrated in Figure 5.

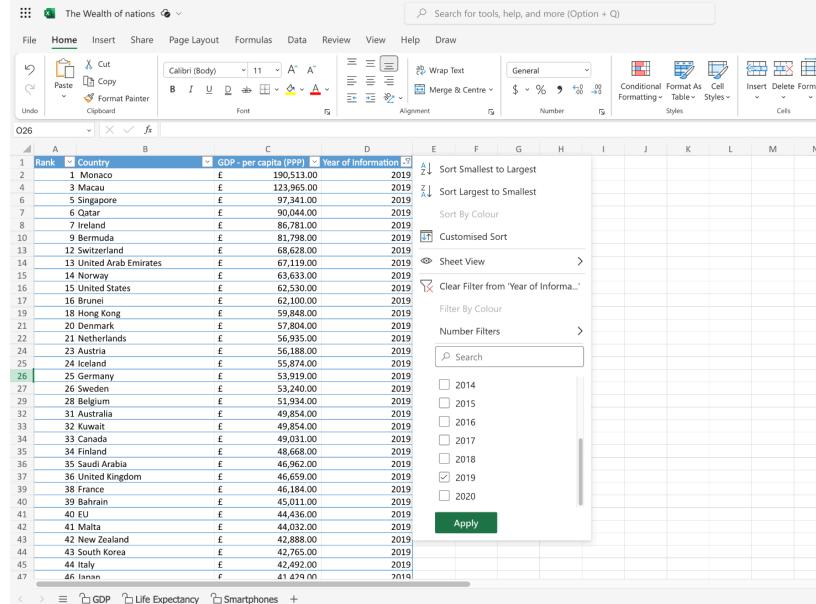


Figure 5

I wanted to create a chart that displayed information on Rank, Country, and GDP - per capita (PPP) in 2019. So I created the following bar chart shown in Figure 6, which I found to be the most effective way to display this information.

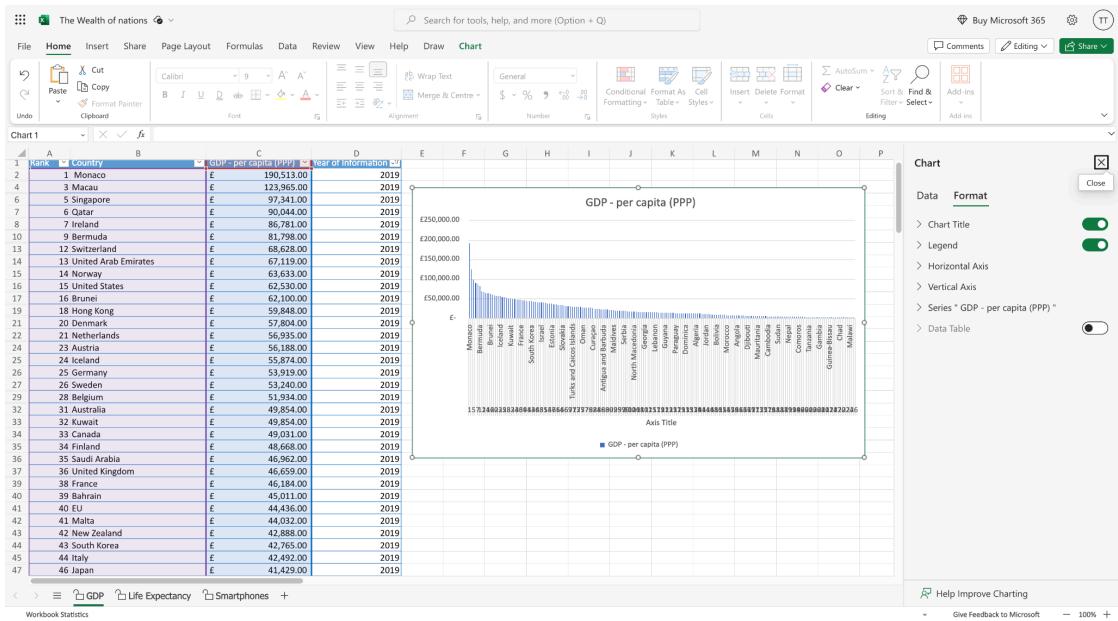


Figure 6

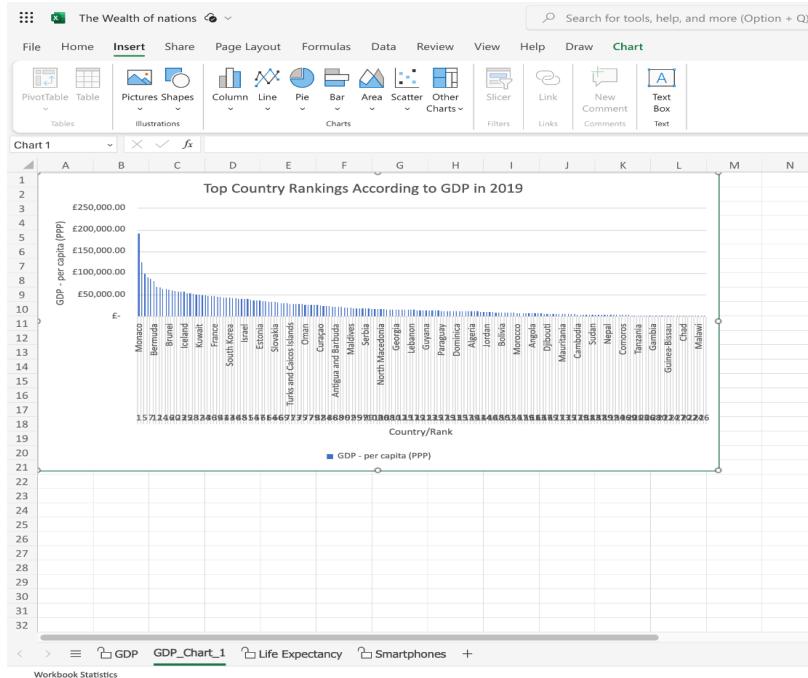


Figure 7

I then moved the chart to a separate sheet called 'GDP_Chart_1' to give me more room to add a title, label the x and y axes, and format the chart to be more visibly appealing (shown in Figure 7).

As per the client requirements I sorted and filtered the data in the table to ensure that information for only the top 20 highest ranking counties was displayed. I then created a bar chart to visually represent this data.

To ensure easy legibility I then placed the chart to be underneath the table, as shown in Figure 8. I also coloured the background by adding a fill colour to the cells around and behind the table and chart.

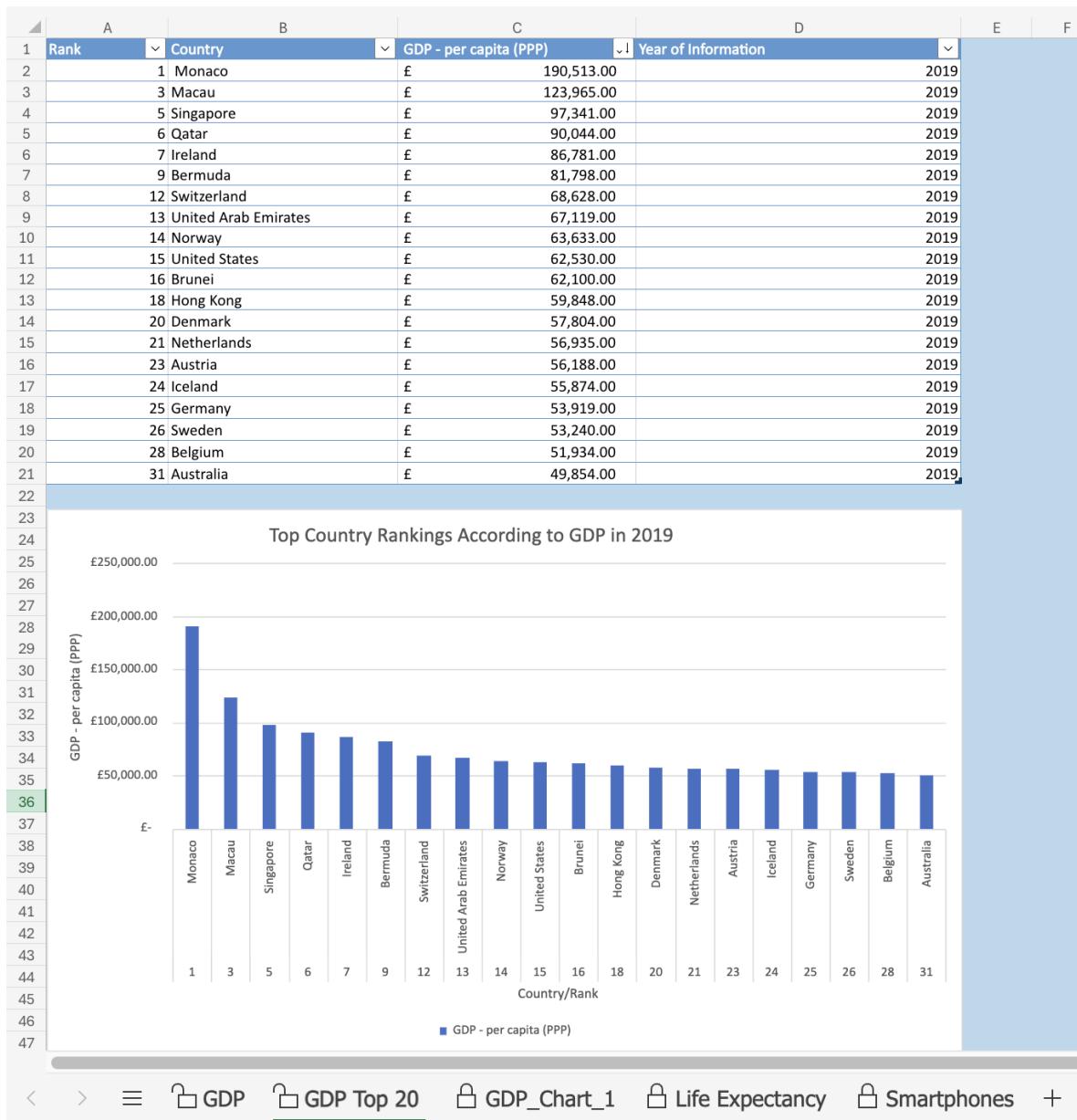


Figure 8

Third Task - Tableau

- - - - X

To carry out further analysis and create more sophisticated visualisations, I imported the ‘Wealth of Nations’ Excel file into Tableau.

As I had a chance to explore the data set in Excel I noticed that the common column in all the tables was the country column. As a result, once the data was imported, I visually arranged the sheets and made sure this commonality was reflected when I set the relationship between the various tables (shown in Figures 9 and 10).

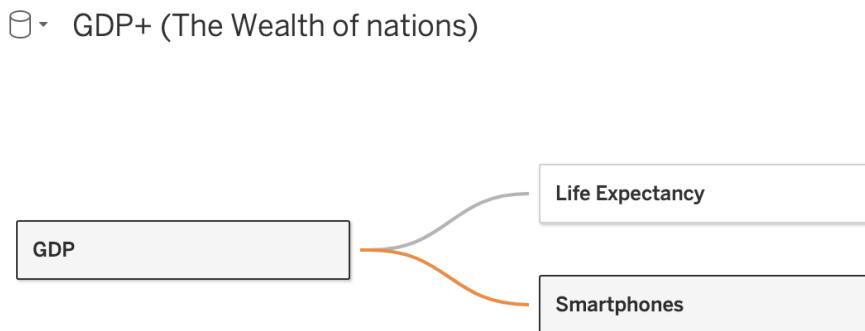


Figure 9

How do relationships differ from joins? [Learn more](#)

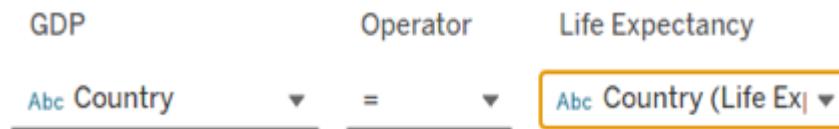


Figure 10

Upon checking the data I noticed the data type for the dates in the information column for each table was set to either Number(whole) or Number(decimal). I decided to change the data type to date manually. As there was no record of time in the data set, nor any relevance, I decided to select the date-only option. As shown in Figure 11 and Figure 12, I made this change for each table.

The screenshot shows the Tableau Data Editor interface. At the top, a table titled "Life Expectancy" is displayed with 4 fields and 196 rows. Below it, another table titled "Date of Information" is shown with 8 rows of data. A red box highlights the "Date of Information" table and its columns. To the left, a dropdown menu for "Date of Information" is open, showing various data types: Number (decimal), Number (whole), Date & Time, and Date. The "Date" option is selected and highlighted with a red box. An arrow points from the "Date" option in the dropdown to the "Date of Information" table.

#	Life Expectancy	Rank (Life Expectancy)	Country (Life Expectanc...	Life expectancy at birth	Date of Information
1	Hong Kong			84.9000	2.020
2	Japan			84.6000	2.020
3	Switzerland			83.8000	2.020
4	Singapore			83.6000	2.020
5	Spain			83.5000	2.020
6	Italy			83.4000	2.020
7	Australia			83.4000	2.020
8	Iceland			83.0000	2.020

#	Life Expectancy	Rank (Life Expectancy)	Country (Life Expectanc...	Life expectancy at birth	Date of Information
1	Hong Kong			84.9000	01/01/2020
2	Japan			84.6000	01/01/2020
3	Switzerland			83.8000	01/01/2020
4	Singapore			83.6000	01/01/2020
5	Spain			83.5000	01/01/2020
6	Italy			83.4000	01/01/2020
7	Australia			83.4000	01/01/2020
8	Iceland			83.0000	01/01/2020

Figure 11

Smartphones

4 fields 74 rows

Name
Smartphones

Fields

Type	Field Name	Phys...	Rem...
#	Rank (Smartphones)	Smart...	Rank (...)
@	Country (Smartphones)	Smart...	Count...
#	Smartphone Users	Smart...	Smart...
#	Date of Information (Smartph...	Smart...	Date o...

GDP

4 fields 228 rows

Name
GDP

Fields

Type	Field Name	Phys...	Rem...
#	Rank	GDP	Rank
@	Country	GDP	Country
#	GDP - per capita (PPP)	GDP	GDP - ...
#	Year of Information	GDP	Year of...

Smartphones

4 fields 74 rows

Name
Smartphones

Fields

Type	Field Name	Phys...	Rem...
#	Rank (Smartphones)	Smart...	Rank (...)
@	Country (Smartphones)	Smart...	Count...
#	Smartphone Users	Smart...	Smart...
#	Date of Information (Smartph...	Smart...	Date o...

GDP

4 fields 228 rows

Name
GDP

Fields

Type	Field Name	Phys...	Rem...
#	Rank	GDP	Rank
@	Country	GDP	Country
#	GDP - per capita (PPP)	GDP	GDP - ...
#	Year of Information	GDP	Year of...

Figure 12

After the data types were set I went on to create varied charts and visualisations to identify trends and patterns within the data. This was achieved by moving different dimensions and measures into the ‘columns’ and ‘rows’ sections, or by simply dragging them onto the main screen. The ‘Show me’ function provides options for a range of visualisations to be applied depending on the data.

It was imperative that while exploring the data I kept in mind the following client requirements:

- Visuals must be appropriate for colour-blind viewers
- The client is only interested in the top 20 highest-ranking countries
- There must be a minimum of four visuals

1) Visual 1 - Countries with the highest GDP rank (Bar chart)

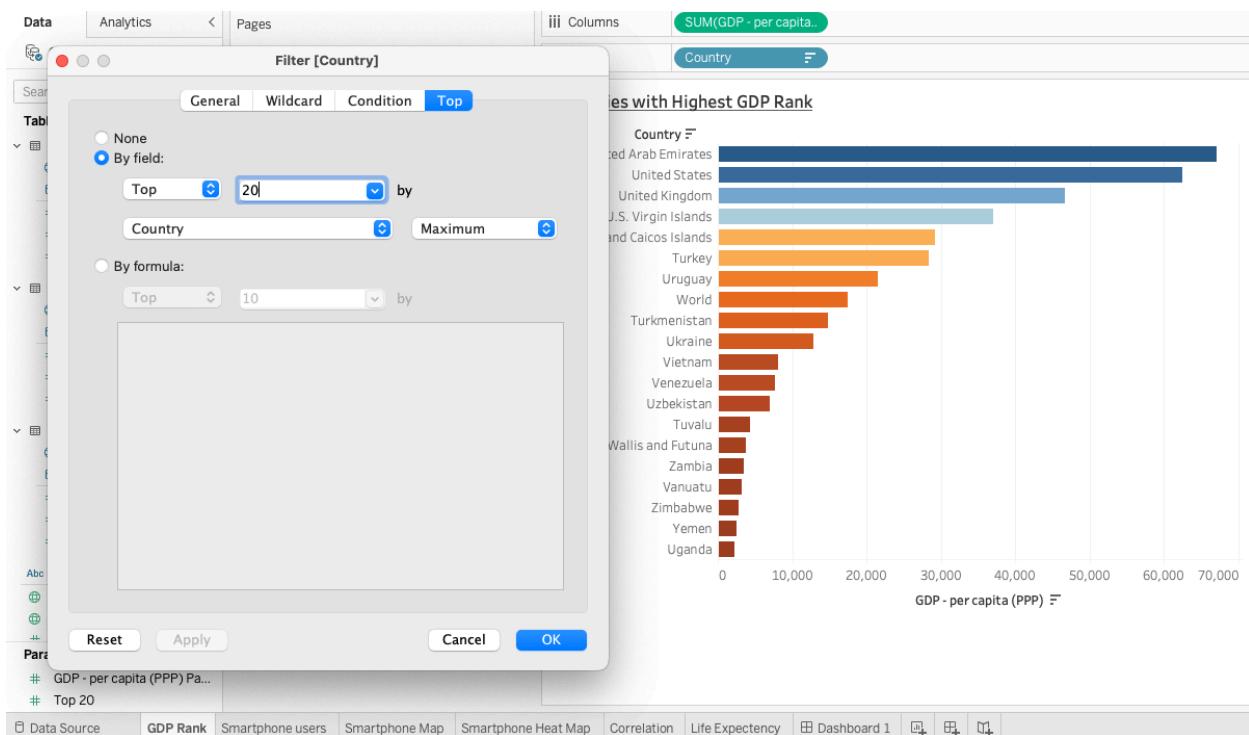


Figure 13

Visual 1 displays a bar graph with data on the highest-ranking countries according to GDP rank. I used the country data in the GDP sheet to filter the top 20 countries as

shown in Figure 13. I also used a colour gradient to visually highlight the differences in amounts for GDP per capita PPP for each nation.

2) Visual 2 - Countries with the highest no. of Smartphone users (Geographic heatmap)

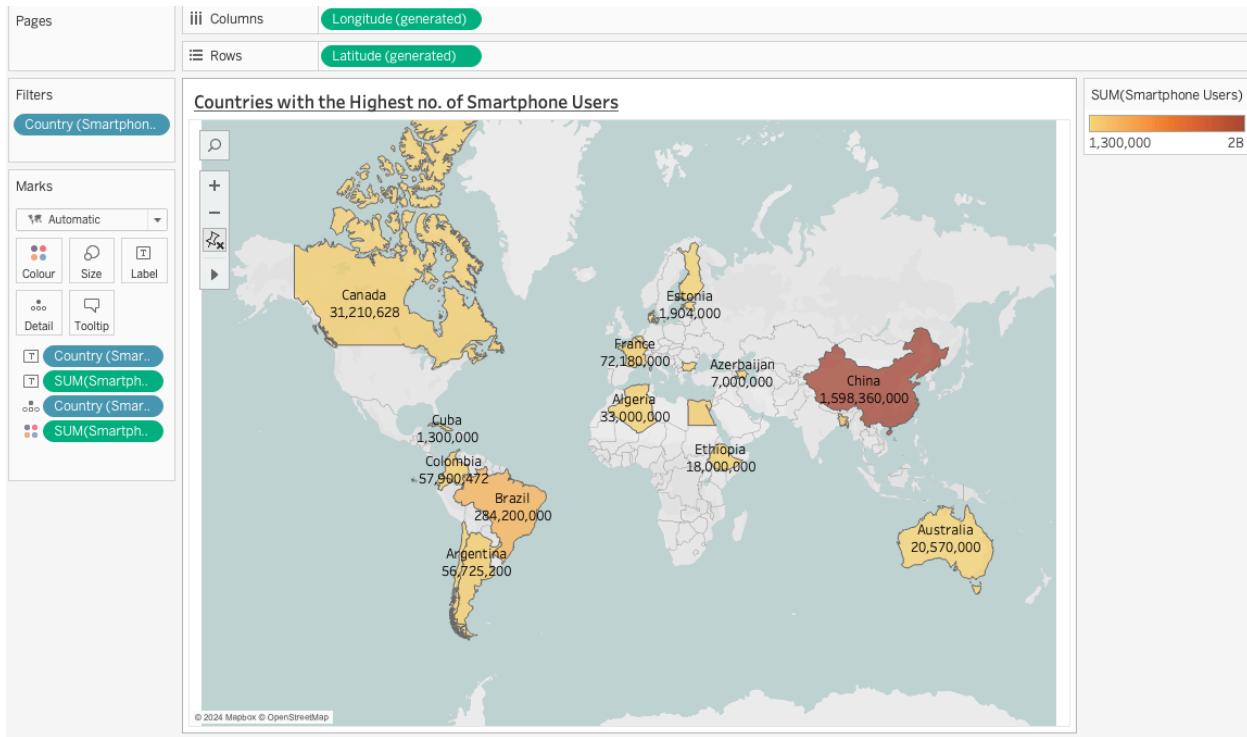


Figure 14

This map was created based on the number of smartphone users per country. The more intense the colour, the higher the number of smartphone users in that region. This visual was also filtered to show the top 20 countries in this regard as per the client's requirement. I also made the country name and no. of smartphone users visible to make the data easily digestible.

3) Visual 3 - GDP Per Capita vs Life Expectancy at Birth (Correlation)

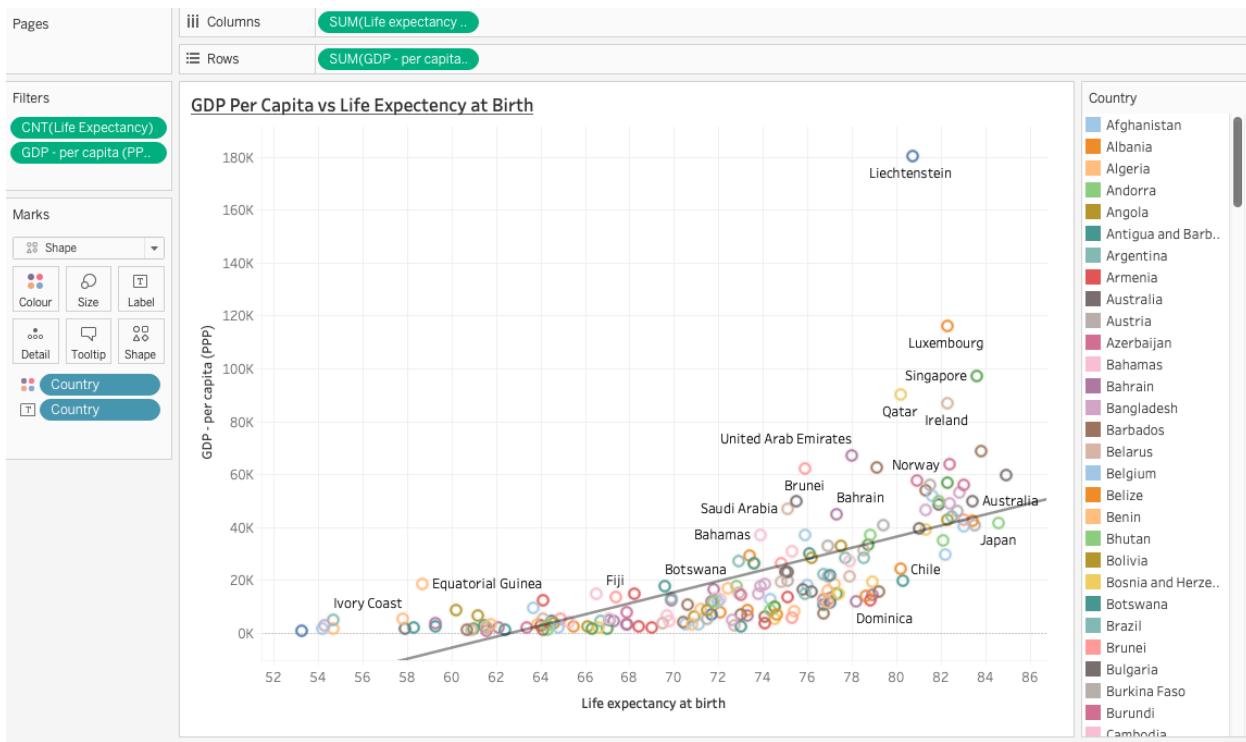


Figure 15

This chart illustrates the relationship between GDP per capita and life expectancy. I also added a trendline which calculated the p-value as <0.0001 and shows a slight positive correlation.

4) Visual 4 - Countries with Highest Life Expectancy After Birth (Treemap)

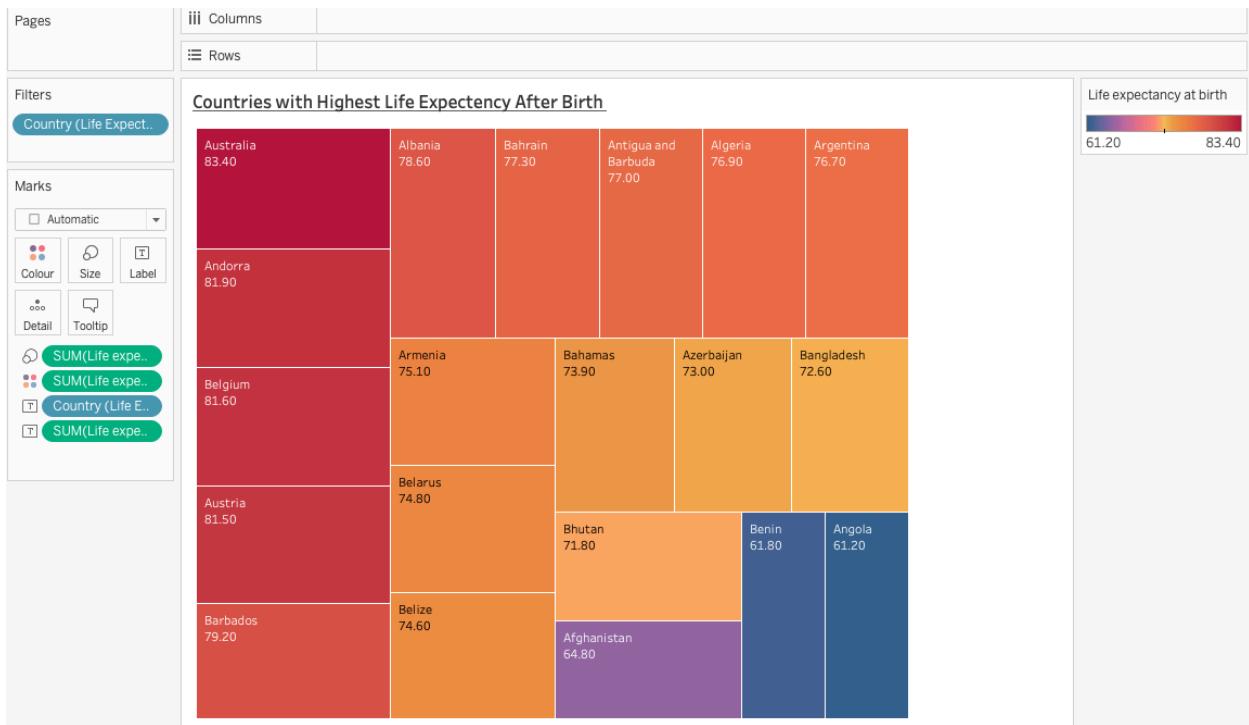


Figure 16

Visual 4 shows a treemap displaying data on life expectancy after birth for each country. The closer in colour to red the higher the average life expectancy of that nation. The closer in colour to blue the lower the average life expectancy of that nation. The treemap allows for quick perception of the countries with the highest life expectancy through the use of contrasting colour intensity highlighted in a colour-blind-friendly pallet. This visual was also filtered to show only the top 20 countries

Dashboard

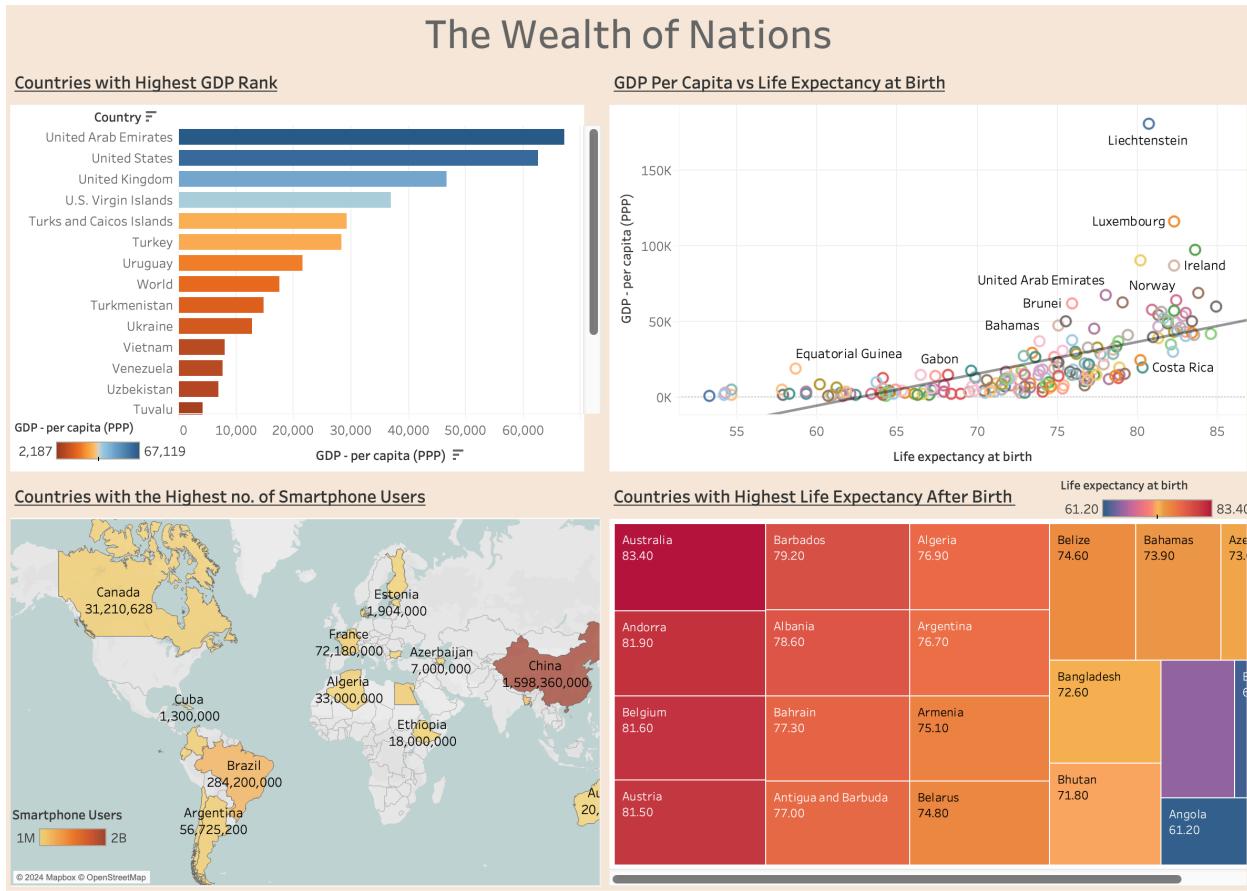


Figure 17

To satisfy the client's requirement, the dashboard, shown in Figure 17, was designed for colour-blind viewers. This meant a colour-blind pallet that avoided problematic colour combinations, like red/green, green/brown, green/blue, blue/grey, etc, was used. The background was also set to a light colour to improve legibility by reducing the stark contrast between white and black.

As the Wealth of Nations dataset is quite big, it would have been difficult to present data for each country. Therefore, as per the client's requirement, visuals 1, 2 and 4 display the top 20 highest-ranking countries only.

Link to dashboard:

<https://public.tableau.com/app/profile/tamima.tarofdar/viz/WealthofNations-TT/Dashboard1?publish=yes>

Conclusions

- - - - X

Upon analysing the data, one of the most significant findings to note is the low p-value of <0.0001 and the positive correlation between GDP PP and life expectancy. This suggests that the higher a nation's GDP, the higher the average life expectancy of its population. This is not an entirely surprising outcome due to the increased investment in policy measures, including healthcare, education and infrastructure, which typically follows when a country has a high GDP.

Another interesting observation is the discrepancy between Hong Kong, which ranked 1st concerning life expectancy, and Brunei, which ranked last at 73rd. Despite this Brunei ranked higher than Hong Kong in terms of GDP PPP. This contradicts the finding earlier mentioned as in some instances the positive relationship between high GDP PPP and high life expectancy does not align. It also highlights that other factors may influence life expectancy, and further investigation with a more comprehensive dataset may be required to identify these.

There were also other cases where the data did not align as expected. Upon closer inspection of the data, it was clear the reason for the misalignment was because there was missing data. For example, Brunei is in the top 20 highest-ranking countries for GDP, but has no recorded data on its smartphone users. This is just one example of the few instances where there was substantial data for one side of the relationship and little for another. Furthermore, Macau and Hong Kong are not countries and - despite serving as Special Administrative Regions - are considered part of mainland China. Therefore, the comparison between cities with countries can distort results and should be considered when interpreting the dataset.

Evaluation

- - - - X

Overall, the success of this project was dependent on my careful adherence to the assessment guidelines which ensured that the final outcome met the required standards. I enhanced clarity by including annotated screenshots and formatted the structure of the content in an organised manner so it was easy to follow.

Moving forward, I aim to utilise the desktop version of Microsoft Excel to fully leverage its advanced functionalities like macros and workbook protection. Additionally, I plan to refine my project planning to allocate more time for independent study, particularly focusing on developing skills such as coding for dashboard creation and implementing calculated fields in Tableau to streamline analysis.

Moreover, I intend to work with more comprehensive datasets and set more ambitious project objectives to challenge myself further. Through this experience, I was able to refine my recently acquired skills in Tableau and Excel and strengthen my understanding of GDPR implications and compliance policies for handling datasets.
