|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  | **29/12/2023** |  |
|  |
|  |
| Project Fortuna  Assignment 1  **Usmaan Dar** |
|  | **DTE Cohort 5** |  |
|  |  |  |
|  |  |  |

Table of contents

[Data protection policies 2](#_Toc2082011426)

[Why we use data protection policies 2](#_Toc238145174)

[Organisational data protection policy essentials 3](#_Toc739923783)

[How does this affect a data analyst? 3](#_Toc1131802657)

[Excel Tasks: Wealth of Nations 3](#_Toc259745486)

[Tableau dashboard project 6](#_Toc1095589694)

[Interesting points to note 10](#_Toc615272892)

[Conclusions 12](#_Toc682737862)

# Data protection policies

## Why we use data protection policies

These policies are essential for working with data to ensure information is used fairly, lawfully, transparently, and securely. Data protection policies protect the rights and interests of the people and companies whose data is being shared, so relevant laws and regulations (such as the GDPR, EU, Data Protection Act [UK]) must be followed.

## Organisational data protection policy essentials

* Encryption policies: These policies specify how and when to encrypt data, both at rest and in transit, to prevent unauthorised access or disclosure.
* Acceptable use policies: These policies define the permitted and prohibited activities for employees when using data i.e. to prevent misuse, abuse, or breach of data, as well as to protect the organisation’s reputation and productivity.
* Password policies: These policies set the rules for creating and managing passwords, such as the length, complexity, and frequency of change.
* Data retention and disposal policies: These policies determine how long data should be kept and how to dispose of it when it is no longer needed.
* Data breach policies: These policies outline the procedures and responsibilities for responding to a data breach, such as notifying the relevant authorities and the affected data subjects.

## How does this affect a data analyst?

As a data analyst, it is important to be aware of these rules because they help us to:

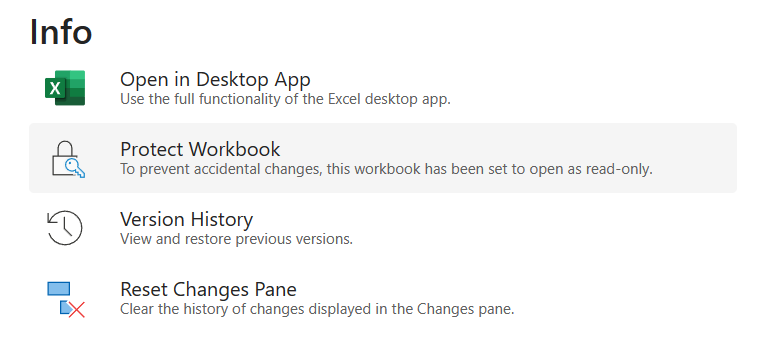
* Comply with the legal and ethical obligations and expectations of our organisation and our stakeholders.
* Ensure the quality, reliability, and validity of data and its analysis.
* Protect the confidentiality, integrity, and availability of data and systems.
* Enhance the trust, transparency, and accountability of data and work.
* Avoid or minimise the potential harm or damage that data misuse or breach can cause to oneself, the organisation, or others.

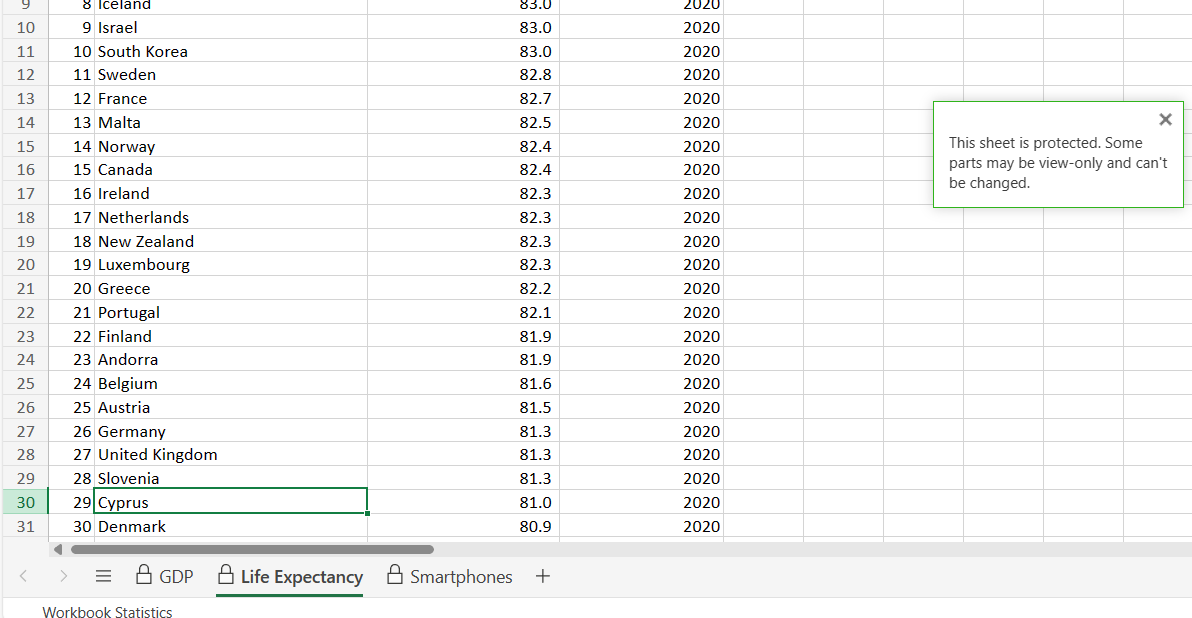
# Excel Tasks: Wealth of Nations

[Usmaan Dar DTE 5 Excel Assigment 1 - wealth of nations.xlsx](https://justit831-my.sharepoint.com/:x:/g/personal/usmaandar_bootcamp_justit_co_uk/EWJIpIPu1dJBtMKBKbVazOQBP1PvYCBcWtOJHMWQENky_A?e=optXhy)

[see Excel sheet in the above link]

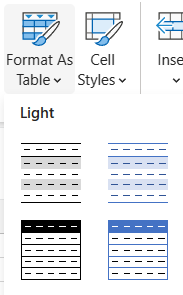
1. Password protection for workbooks is **unavailable** for Excel Online (it simply makes the doc ‘read only’, but there are alternate solutions:
   1. Make the excel sheet read-only
   2. Password-protect the worksheets individually e.g. see below (Password = wealthofnations):

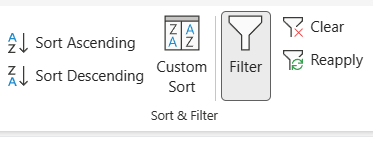


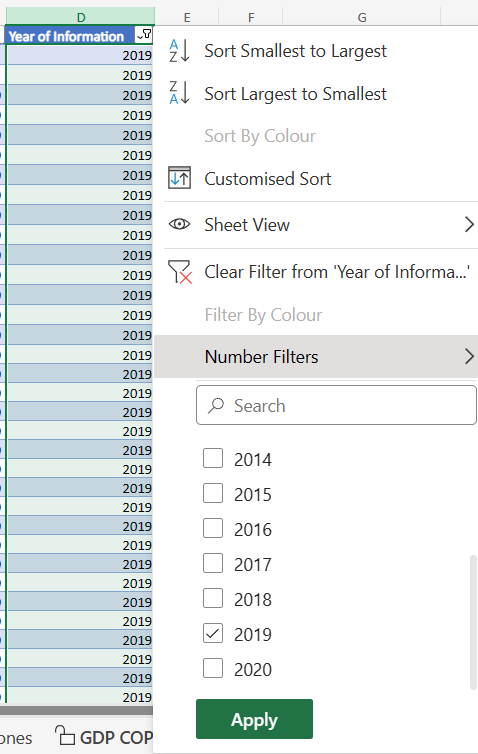


As you can see from the images above, I have password protected the original sheets (as the workbook cannot be password-protected on Excel Online). I have made separate copies of said sheets to display the changes from the original data to now.

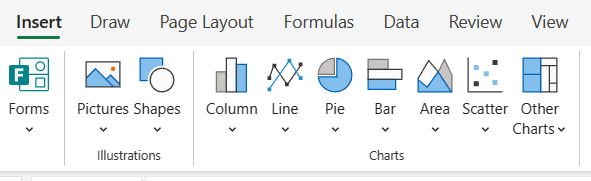
1. The data would first be made into a table, by highlighting the data and – on the ‘Home’ tab, and then ‘formatting as table’ - selecting the desired table. The table was then filtered to display only 2019 information by visiting the ‘Data’ tab, ‘sort and filter’ and highlighting only ‘2019’.

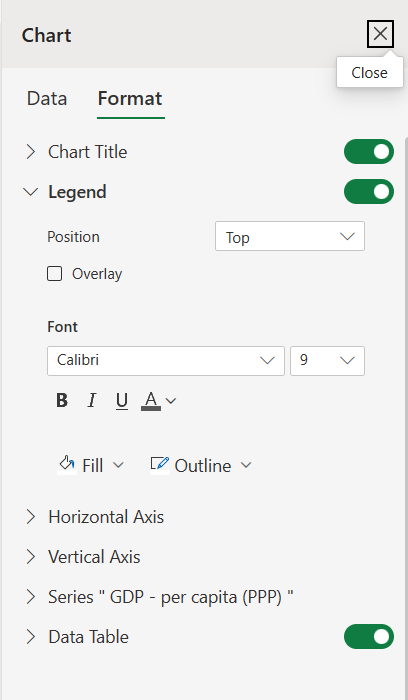




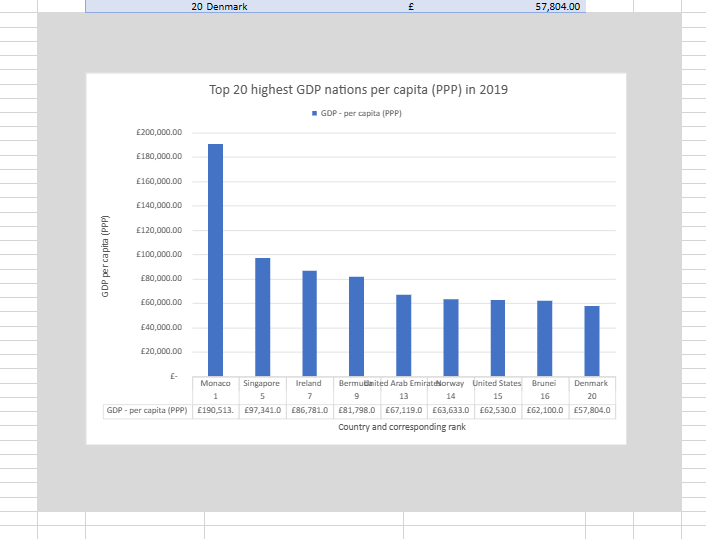


1. Next, a chart was created, to display only GDP PPP and moved to a separate sheet. The chart was made larger to accommodate all nations and their ranks. The chart was made larger by highlighting the data and selecting from the appropriate chart options. The ‘Format’ tab allowed me to edit the chart to make it more visually pleasing.





1. I would use my learnt skills to create a new chart and filter the data to display only the top 20 highest ranking GDP PPP nations. A grey background was also inserted behind the chart; by highlighting the desired cells, visiting the ‘insert’ tab, and using the ‘fill’ function.



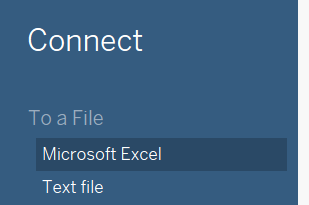
# Tableau dashboard project

<https://public.tableau.com/shared/CKRYZZ6YW?:display_count=n&:origin=viz_share_link>

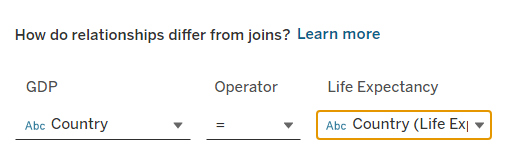
[see Tableau public profile in the above link]

Steps taken to achieve desired results:

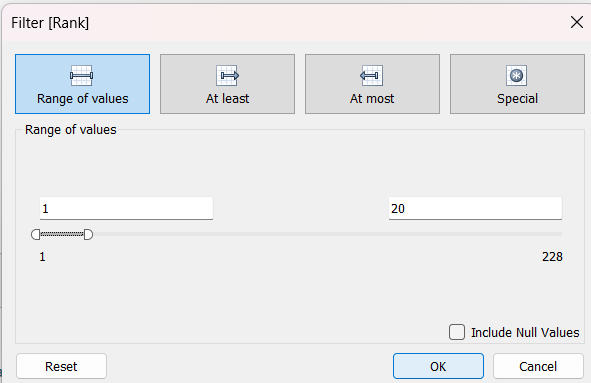
1. First, I enter the Tableau App and import the data from Excel.

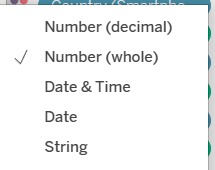


1. I set the relationship between the various data columns with 'country' being the common factor.

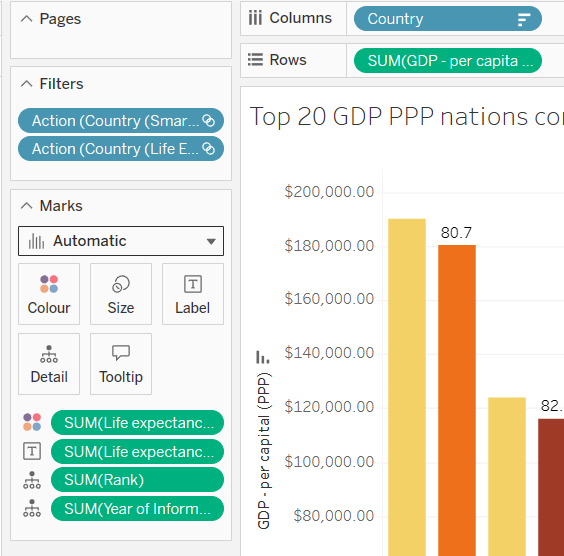


1. The ‘top 20 highest ranking countries’ were gathered by filtering the data source before moving to the worksheet (top right corner of the data source screen). Data types included; strings, date and time, and integers.





1. I created varied charts and visualisations to identify trends and patterns within the data. This was achieved by moving the various dimensions and measures into the ‘columns’ and ‘rows’ sections, or by simply dragging them on to the main screen. The ‘Show me’ function allows for different visualisations to be used.



1. A colour-blind palette was used to ensure accessibility for all users including those with Colour Vision Deficiency (CVD). There are several CVD types, so the two most common CVD types were covered in this Viz. The palette used, therefore, covered those with dichromacy and anomalous trichromacy (this means that ‘red’ and ‘green’ dual-colour combinations were avoided). Tableau Classic Medium is considered a CVD-friendly palette.

This also means that the colour palettes presented on the dashboard are such that a CVD person may distinguish between light and dark colours with ease. A minimalist approach was taken, given that a CVD person would struggle to identify differences in colour shades, so 3-stepped colours (max) were used.

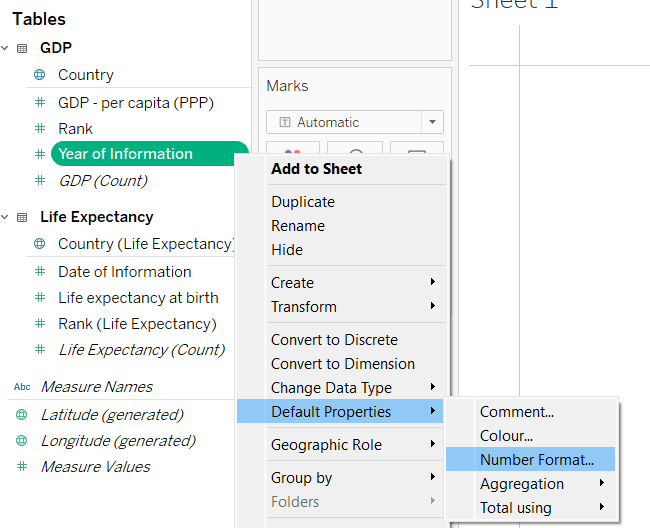


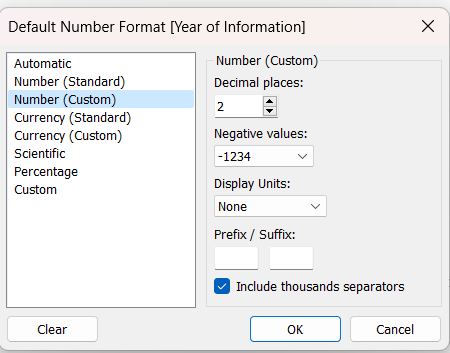
1. The dashboard would prove the trickiest to make, as all 4 visual diagrams had to fit onto a single page, including their legends. I would need to edit the container layout to ‘desktop’ and format the visuals from there.



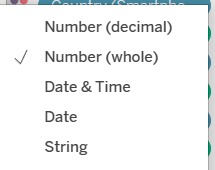
### Interesting points to note

1. It was difficult to understand at first why the data - when ranked according to countries - would not align in some cases. It turns out there is missing data i.e. where there is lots of data for one side of the relationship, there is little for another e.g. Brunei is in the top 20 highest ranking countries (for GDP), but is absent entirely from the data source under smartphone users.
2. I tried switching the ‘year’ parameter from integer to a string to remove the comma to allow for better formatting (from 2,020 to 2020). This did not work, so I used number formatting (custom) to untick the 'thousands separator' and removed decimals to allow the ‘year’ parameters to appear as whole numbers. This was duplicated across all 3 parameters.





1. Data types included the following: Strings, numbers, and dates and times



1. I added the ‘$’ symbol as a prefix to the GDP PPP parameter by right clicking on the parameter, visiting number formatting, and adding the prefix symbol (see image under ‘B’ above).

### Conclusions

The higher the GDP PPP, the more likely the nation will have a higher life expectancy. This is understandable given the greater funding opportunities for public policy measures surrounding housing, infrastructure, healthcare, educations, etc, among many other factors.

An interesting point to note is that, in the given list of values, Hong Kong is ranked 1st (in life expectancy) and Brunei is ranked last at 73rd. However, Brunei is ranked higher than Hong Kong in terms of GDP PPP. Therefore, the higher the GDP PPP of a nation, it is not necessarily a clear and definitive rule that a nation will have a higher life expectancy. It is arguable that the reverse trend is true when considering the data, yet, more data would be required to identify key factors at play here. For instance, from one chart alone, 40% of 'life expectancy' data is missing.

Further, Macau and Hong Kong are not countries and - despite serving as Special Administrative Regions - are considered as part of mainland China. Therefore, the data compares cities with countries which may skew results entirely.

Further, the USA has a far greater number of smartphone users, however, the evidence is far from conclusive owing to a very small sample size from the dataset. The data may be invalid, or incorrect, as the data states there are 327.58 million smartphone users, against the US population of 331,45 million (2020, US Census Bureau). This suggests that 98.8% of the US population uses a smartphone (slightly under 9 in 10 people). It is possible that 98.8% of the US population uses a phone, but the demarcation between cell phones and smart phones must be made clear. For example, both are different and maintain different functionalities and purposes. Although, this data may be accurate as the source doesn’t suggest that the US population must ‘own’ a smartphone to ‘use’ it. Even so, given the inaccuracies in the source we have seen so far, the veracity of this data is questionable at best.

Overall, for any conclusions to be drawn, more data is required. Life expectancy data and smartphone usage data show missing or inaccurate values for the top 20 highest ranking GDP PPP nations. In the end, the top 20 highest ranking nations (by GDP PPP) sees Monaco ranked 1st and Denmark ranked 20th.