|  |
| --- |
| Photo displaying partial image of two pie charts on a canvas-textured page |
| Wealth Of Nations Analysis  With MS Excel & Tableau |
| |  |  |  | | --- | --- | --- | | Deepak Assudani | 11 March 2024 | JustIT Data Bootcamp Assignment 1 | |

# **Table of Contents**

[Policies and Procedures 2](#_Toc161007938)

[Excel Progress 3](#_Toc161007939)

[Workbook password 3](#_Toc161007940)

[Currency format 3](#_Toc161007941)

[Turning sheet into a table 4](#_Toc161007942)

[Freeze panes for viewing ease 4](#_Toc161007943)

[Exploratory Data Analysis (EDA): Pivot Table 5](#_Toc161007944)

[EDA: Pivot Chart 6](#_Toc161007945)

[Ad hoc: Importing population information 7](#_Toc161007946)

[Filter table: 2019 data 10](#_Toc161007947)

[Chart: 2019 data 10](#_Toc161007948)

[Moving map chart to a newly labelled sheet 11](#_Toc161007949)

[Sort table (manual 1): Top 20 countries 12](#_Toc161007950)

[Sort table (manual 2): Top 20 countries 14](#_Toc161007951)

[Sort table (auto): Top N countries 14](#_Toc161007952)

[Bar charts: Top 20 countries 16](#_Toc161007953)

[Tableau Progress 17](#_Toc161007954)

[Import xlsx data 17](#_Toc161007955)

[Set table relationships 18](#_Toc161007956)

[Check data types 19](#_Toc161007957)

[Client requirements: Colour-blind friendly & only top 20 countries 19](#_Toc161007958)

[Formatting axis 19](#_Toc161007959)

[Change filter to top 20 20](#_Toc161007960)

[Ad hoc: Smartphone users as % of population 21](#_Toc161007961)

[Tableau chart1 21](#_Toc161007962)

[Tableau chart2 22](#_Toc161007963)

[Tableau chart3 22](#_Toc161007964)

[Tableau chart4 23](#_Toc161007965)

[Dashboard 1 23](#_Toc161007966)

[Project Reflection 24](#_Toc161007967)

# **Policies and Procedures**

As data analysts, when working with data within the UK, we must adhere to at least two main frameworks: the rules within the Data Protection Act 2018, and the guidance within the Data Ethics Framework.

The Data Protection Act 2018 (DPA) is the UK’s version of the EU’s General Data Protection Regulation (GDPR). They are largely similar and both provide legal frameworks when personal data is of concern (including but not limited to its collection, usage, and storage). The DPA also covers other general data, including law enforcement and national security data. As there is no personal data, law enforcement data or national security data involved in this current project, there is no risk of violation of the DPA (to the extent where such data is not used), therefore no specific measures will be taken on this front. Any further analysis or data collection performed outside of this project will need a separate risk assessment, and potentially separate measures in order to adhere to the DPA. For more information, please visit <https://www.gov.uk/data-protection>.

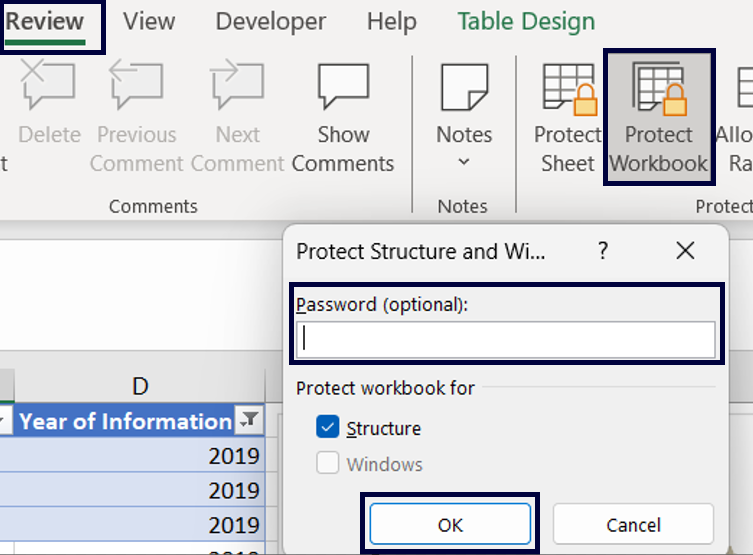
The Data Ethics Framework provides the data controller a set of ethical guidelines when handling data. This framework has overlaps with the DPA, and it is aimed primarily at public servants and policymakers that make use of data, it is also for those working directly or indirectly with data in the public sector (e.g. data analysts, statisticians data scientists). Its overarching principles are: Transparency, Accountability and Fairness, and it also provides guidance on how to ensure these principles are adhered to. This framework is highly relevant to this project, as our data involves data from different countries (and depending on the interpretation of the data, can potentially include misleading information relating to aspects such as spending behaviours, wealth and longevity). Although it is highly unlikely the analysis coming from this project will be used by policymakers, it is still a good habit to keep this framework in mind, as we provide interpretations and snippets of data, and it is important we process the data and present the results in a fair and unbiased manner. For more information on the framework, please visit <https://www.gov.uk/government/publications/data-ethics-framework>.

It should be noted that these regulations and frameworks exist not only to protect the identity and personal of individuals, but also data relating to any particular group. Any misuse or abuse of data or results could have detrimental effects, including but not limited to the misrepresentation or misinterpretation of any particular individual’s or group’s characterisation or behaviour, resulting in illegal or unfair discrimination, targeting or attack on any particular individual or group. Therefore it is vital that these frameworks and regulations are adhered to the highest standard where relevant, and if in doubt, data controllers and analysts should err on the side of caution and effect maximal implementation. Although the topic of ‘Artificial Intelligence’ (AI) is outside of the current project scope, it should be noted that these frameworks and regulations also govern AI safety and ethics, and as the field of AI continues to expand, it is even more crucial for data controllers to conduct thorough data related risk assessments before collecting and using any such data.

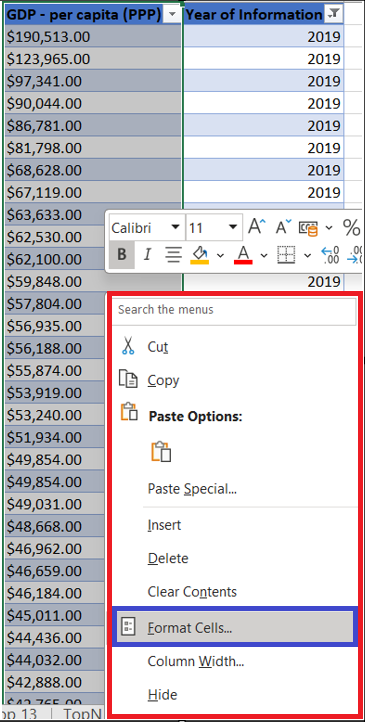
# **Excel Progress**

For the screenshots in this document, a red outline indicates a right click menu, and a blue outline indicates a left click selection.

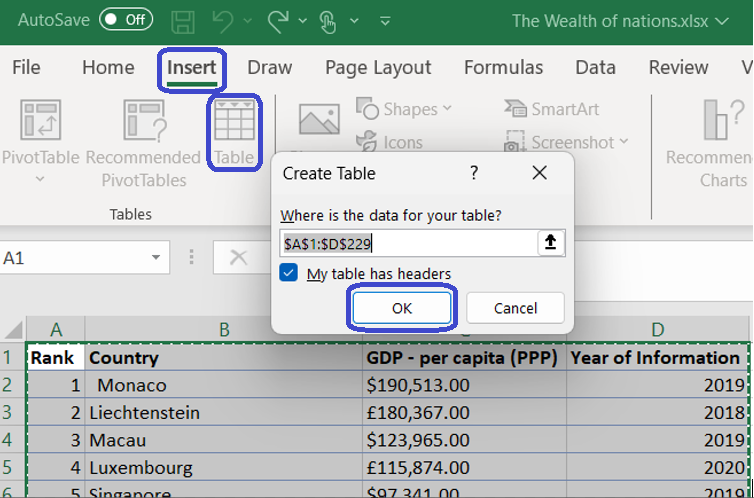
## Workbook password



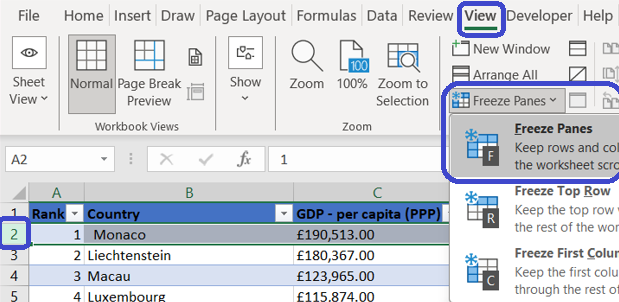
## Currency format



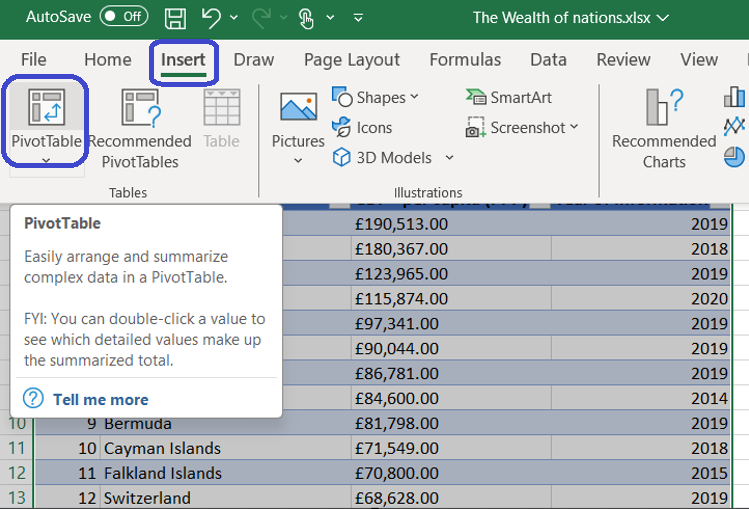
## Turning sheet into a table

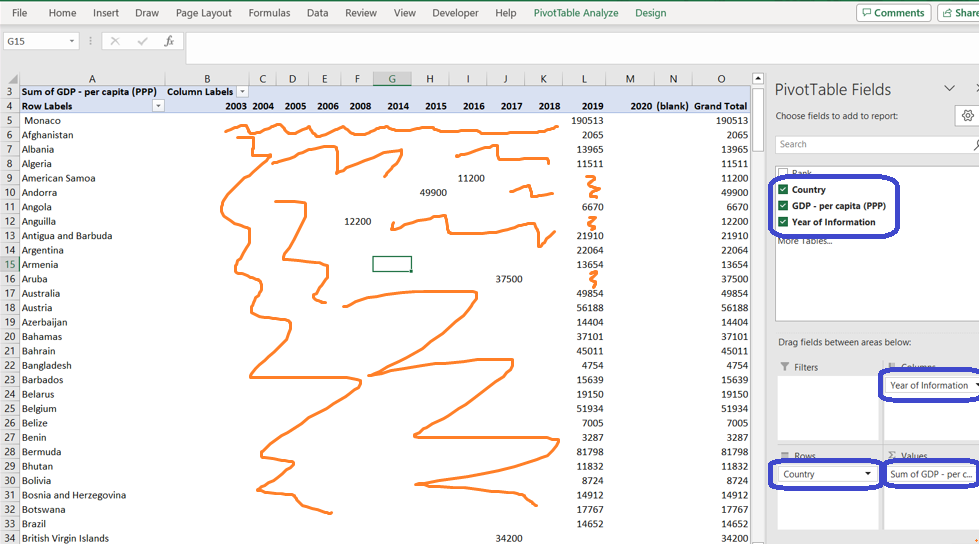


## Freeze panes for viewing ease



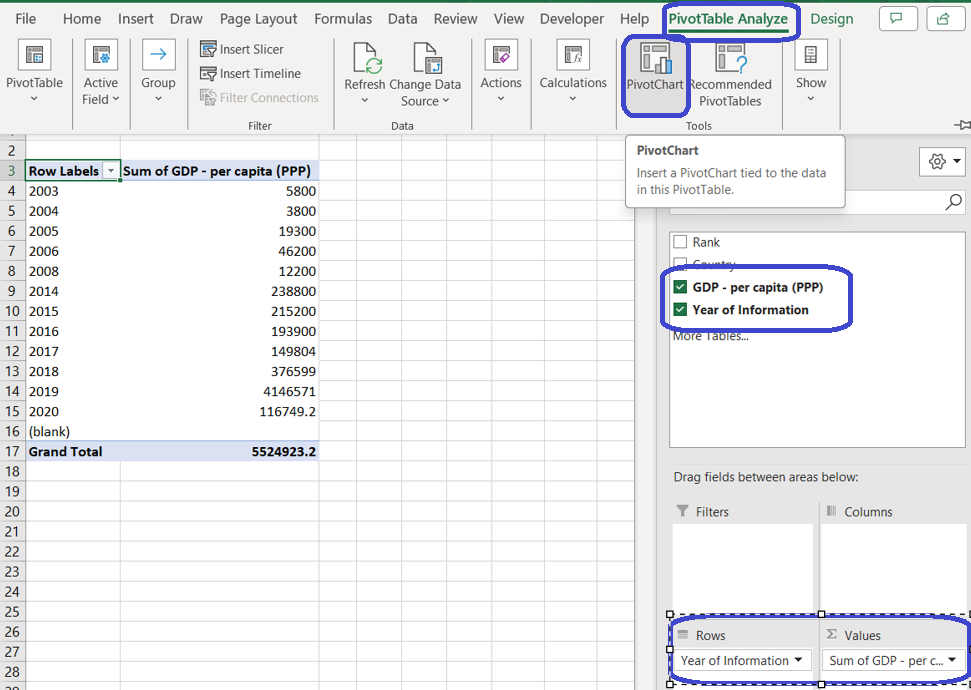
## Exploratory Data Analysis (EDA): Pivot Table

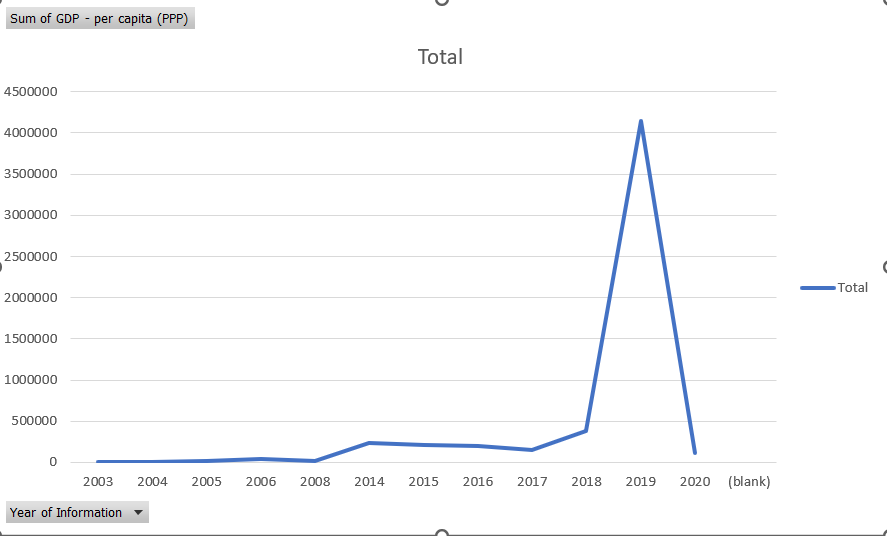




Note: Missing values for many years in many countries, indicating incomplete data

## EDA: Pivot Chart

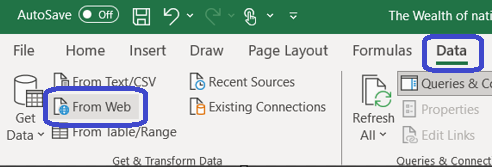




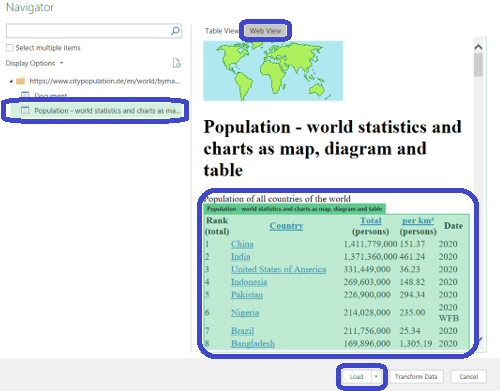
Notice massive spike in 2019 GDP PPP data, indicating missing data in other years (or less likely, huge increase in GDP in just this year)

## Ad hoc: Importing population information

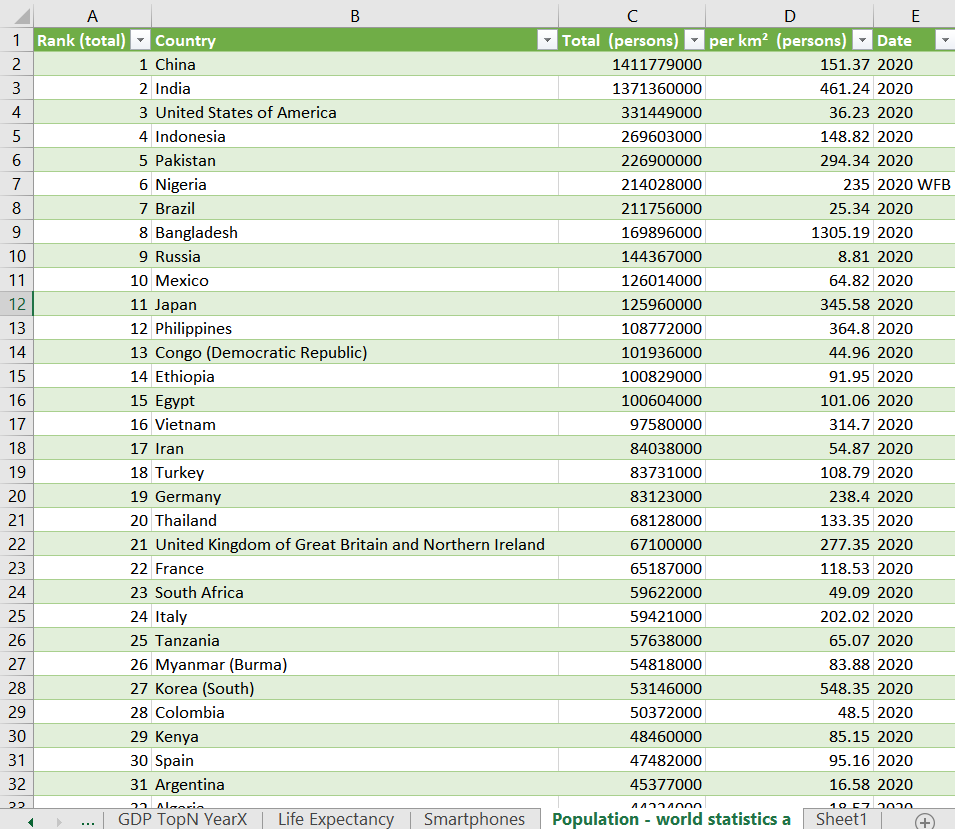
As the number of smartphone users may be skewed by a country’s population, this ad hoc of country population will be imported from an external source.



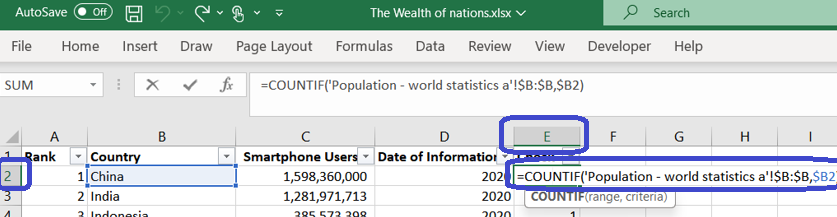




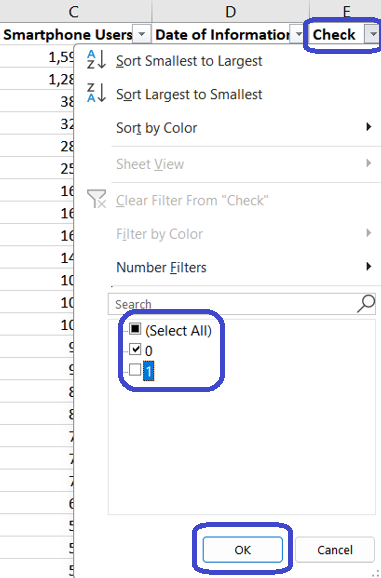
The below is the result



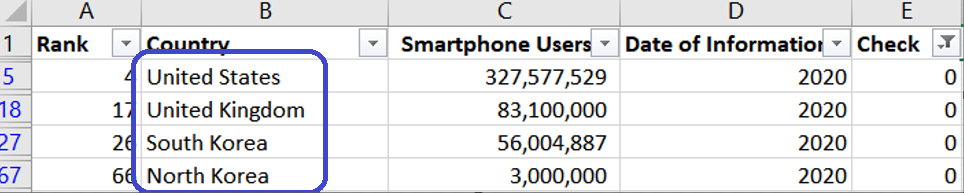
In preparation for a VLOOKUP, I used COUNTIF in a new column named “Check” to check the discrepancies in country names between the Smartphone sheet and the newly imported Population sheet

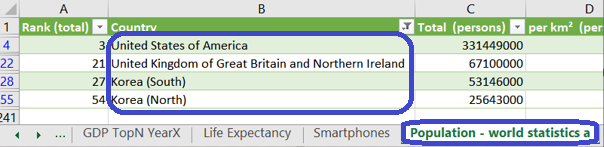


Filter in the 0’s to see name discrepancies

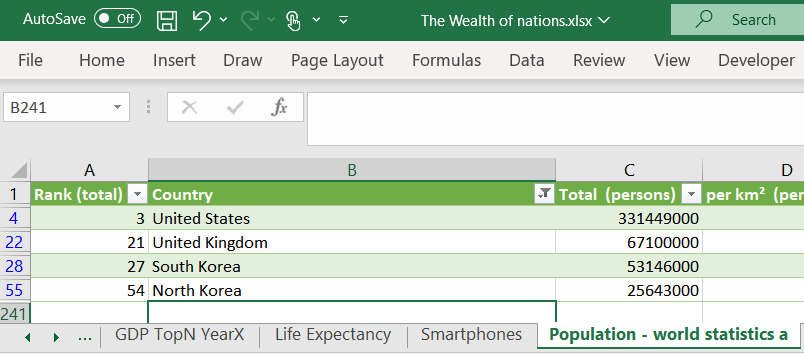


In this case, there are 4 discrepancies

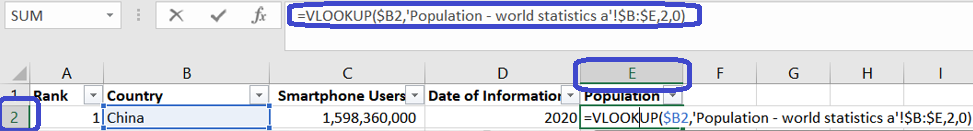




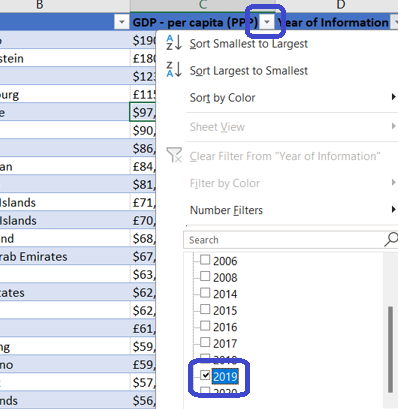
Alter the names in the Population tab



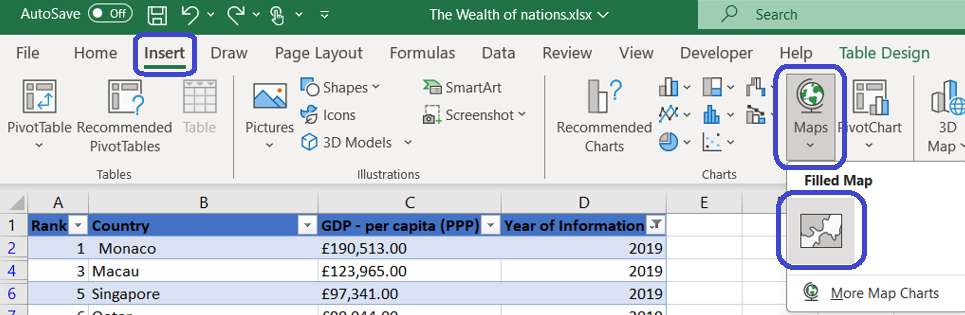
Using VLOOKUP to add a column of Population to the Smartphone tab

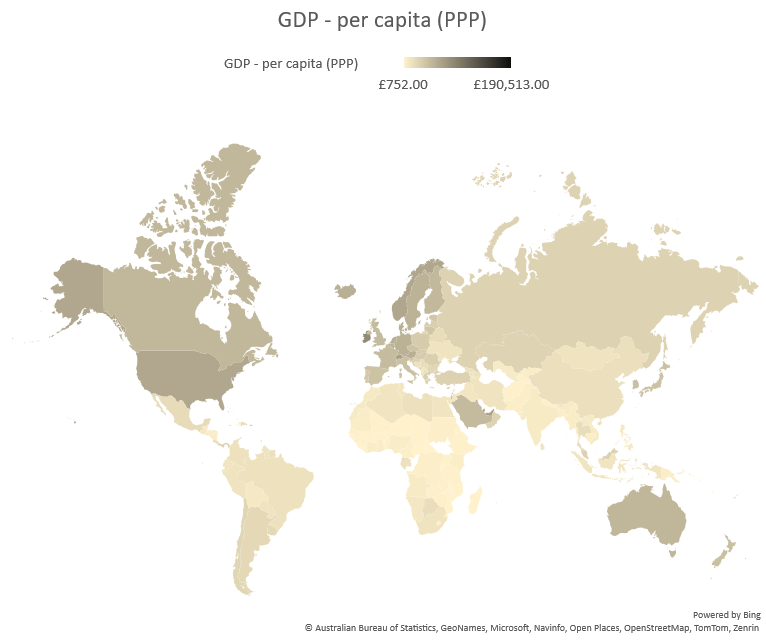


## Filter table: 2019 data

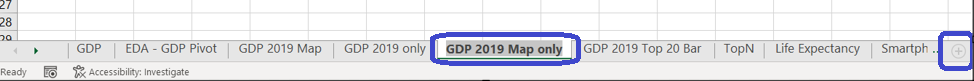


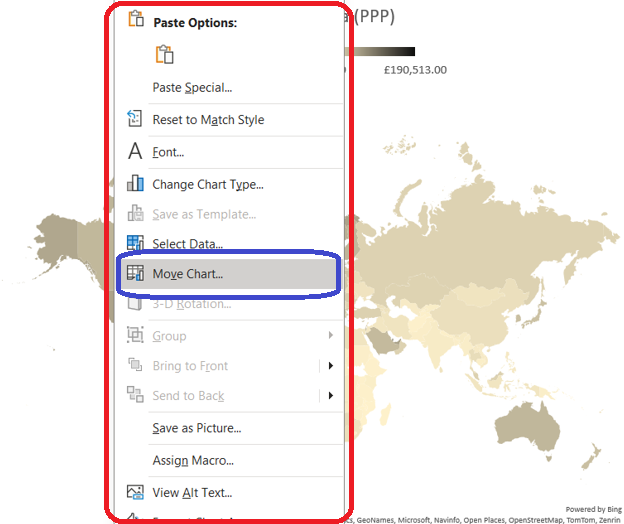
## Chart: 2019 data

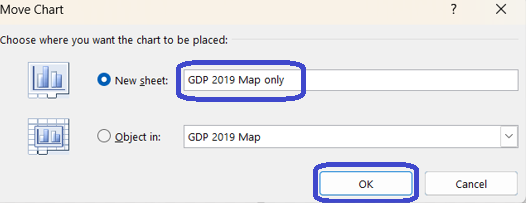




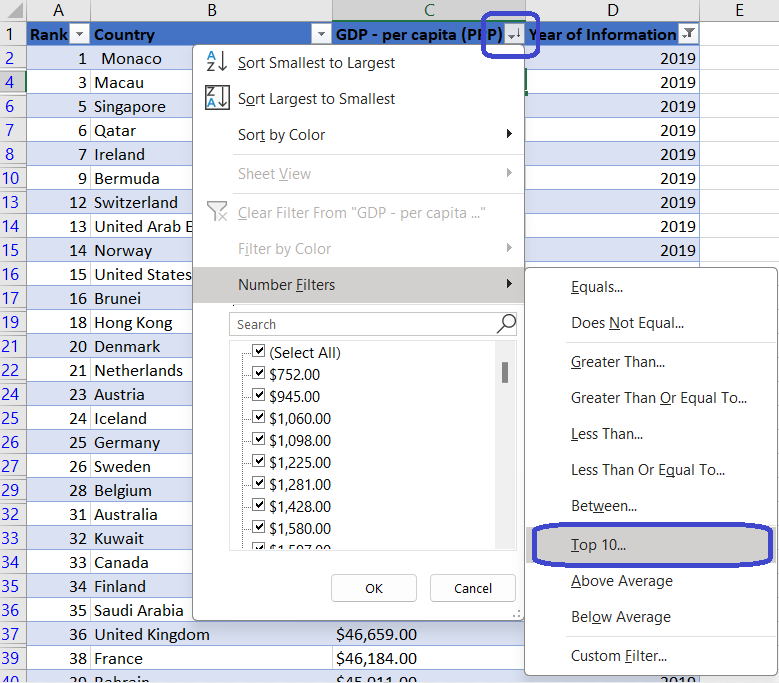
## Moving map chart to a newly labelled sheet

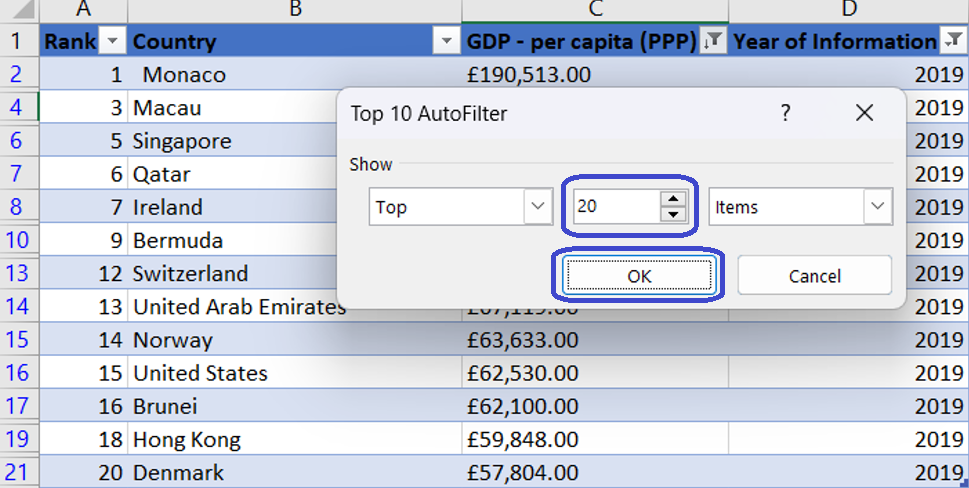




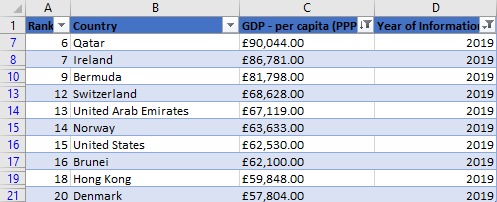


## Sort table (manual 1): Top 20 countries

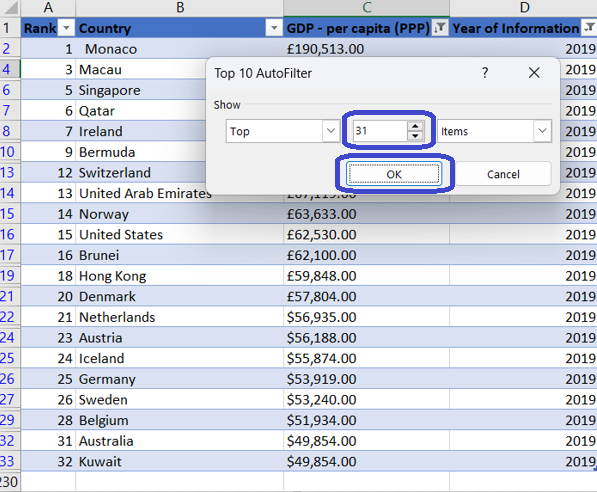




This will filter in the top 20 from the original data, and then filter in 2019 data (most likely due to internal priority of filtering). So the result is a list less than 20 countries. There are 2 methods to rectify this and achieve the top 20 as we originally intended to.

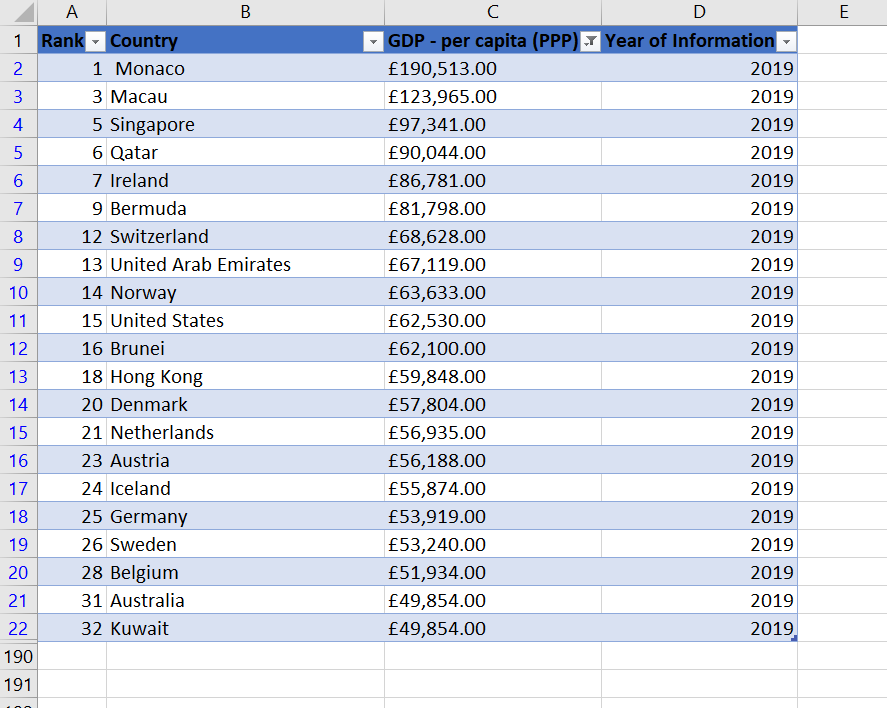


With some trial and error I found out that to get the top 20 (top 21 including a pair of countries that have the same GDP), I need to enter 31 into the sorting field.



## Sort table (manual 2): Top 20 countries

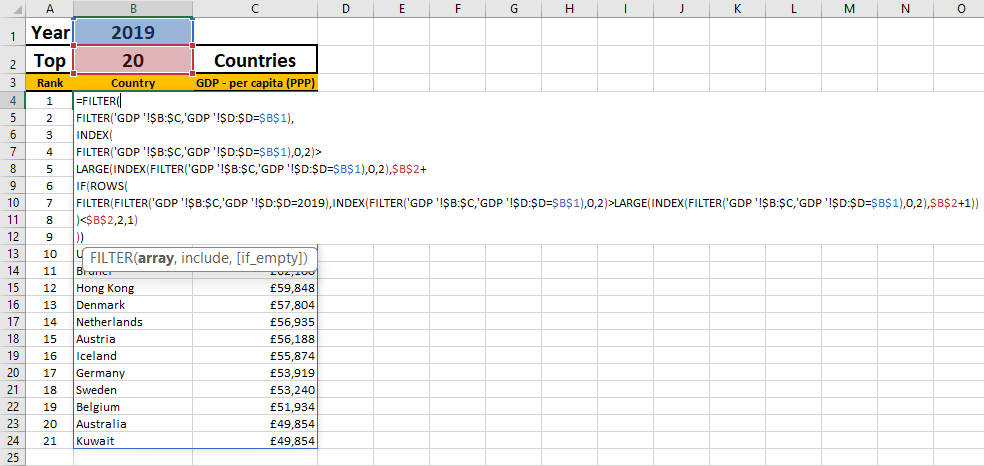
Another way is to copy the 2019 data and paste them As Values onto another sheet, and filter the top 20 countries from there.



## Sort table (auto): Top N countries

An alternative way to make a dynamic query to filter in the countries with top 20 GDP in 2019 is to use the FILTER function in Excel. This eliminates the need to create a new table just for 2019 data.

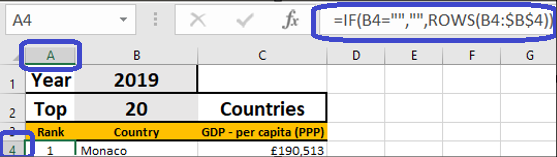
Though the main function is FILTER, other supplementary functions (IF, INDEX, ROWS and LARGE) will also be needed in order to minimise the need to create additional cells / columns / arrays.



The formula is as follows:

=FILTER(  
FILTER('GDP '!$B:$C,'GDP '!$D:$D=$B$1),  
INDEX(  
FILTER('GDP '!$B:$C,'GDP '!$D:$D=$B$1),0,2)>  
LARGE(INDEX(FILTER('GDP '!$B:$C,'GDP '!$D:$D=$B$1),0,2),$B$2+  
IF(ROWS(  
FILTER(FILTER('GDP '!$B:$C,'GDP '!$D:$D=2019),INDEX(FILTER('GDP '!$B:$C,'GDP '!$D:$D=$B$1),0,2)>LARGE(INDEX(FILTER('GDP '!$B:$C,'GDP '!$D:$D=$B$1),0,2),$B$2+1))  
)<$B$2,2,1)  
))

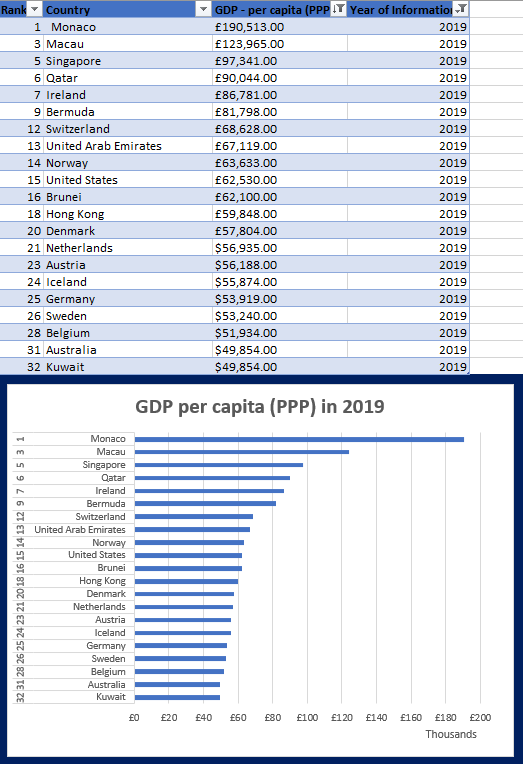
To create a rank for this dynamic table, a new column will be inserted, and a combination of IF and ROWS functions will be used

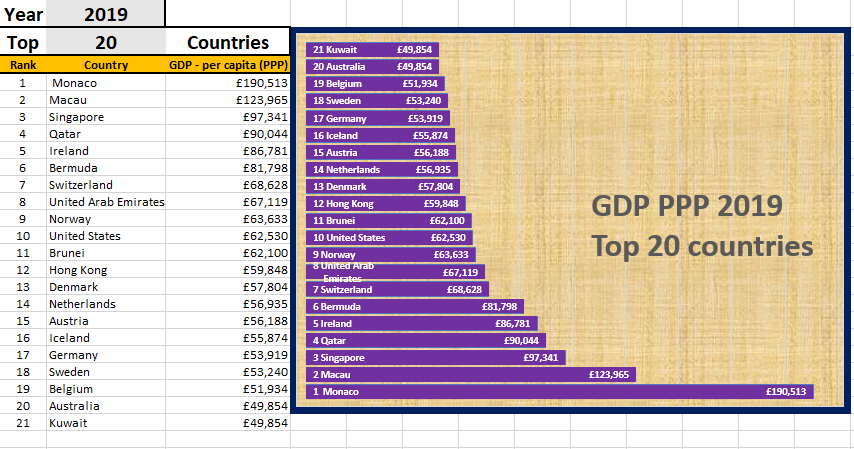


The formula is as follows:

=IF(B4="","",ROWS(B4:$B$4))

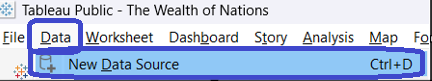
## Bar charts: Top 20 countries

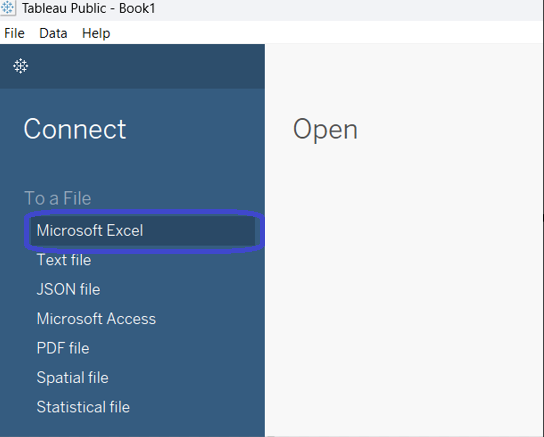




# **Tableau Progress**

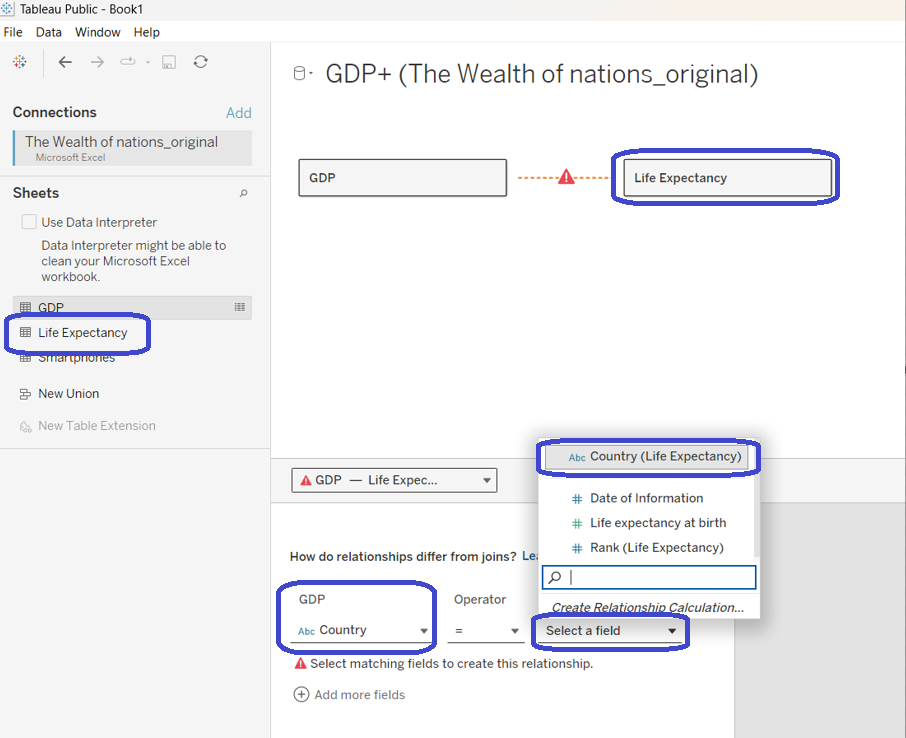
## Import xlsx data

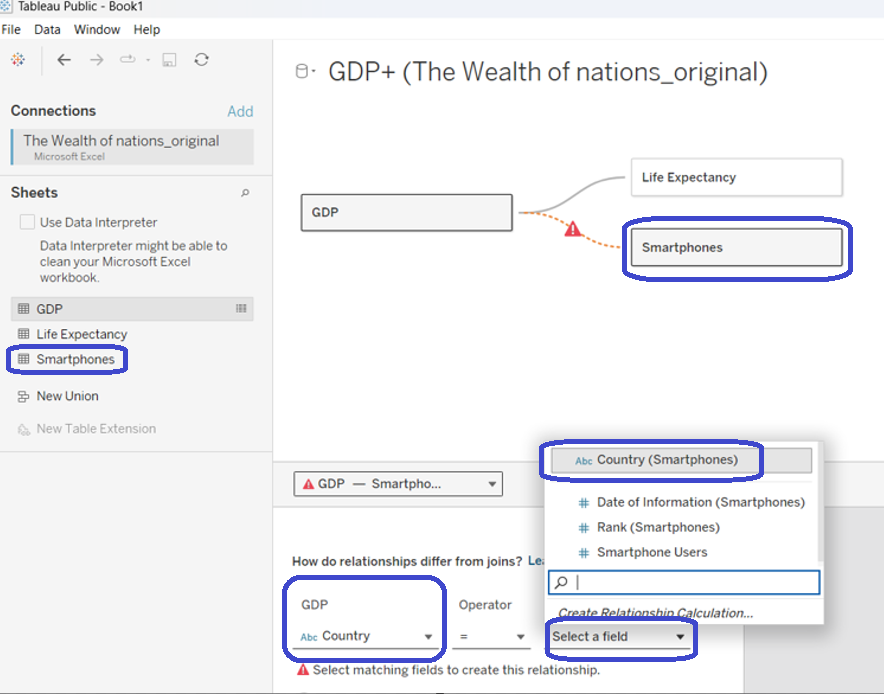






## Set table relationships





## Check data types

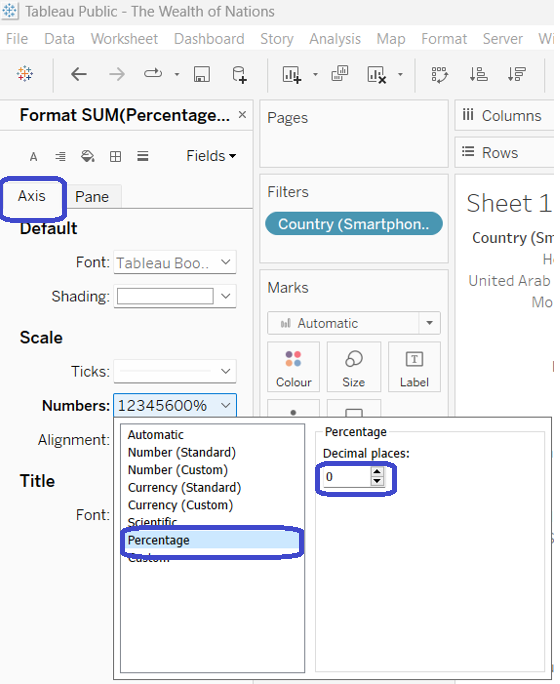


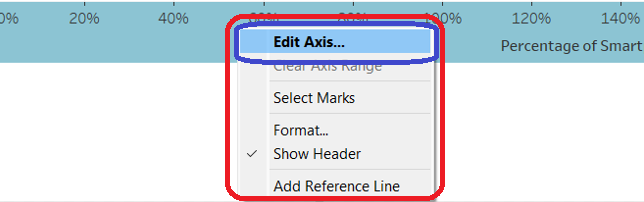
## Client requirements: Colour-blind friendly & only top 20 countries

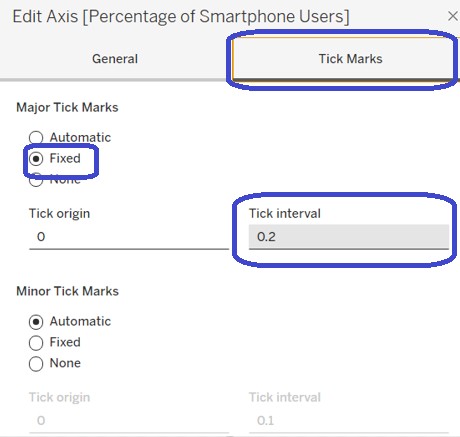
Our client has requested that our visualisations to be colour-blind friendly. As it is unclear whether these visualisations are intended for someone with anomalous trichromacy, dichromacy or the monochromacy, I will opt to mainly use different shapes and sizes within my data visualisation. In cases where colours are used, I will opt for contrasting hues within any one chart.

Another request from the client is that only the top 20 countries are to be presented. This will also be adhered to in the following visualisations.

## Formatting axis

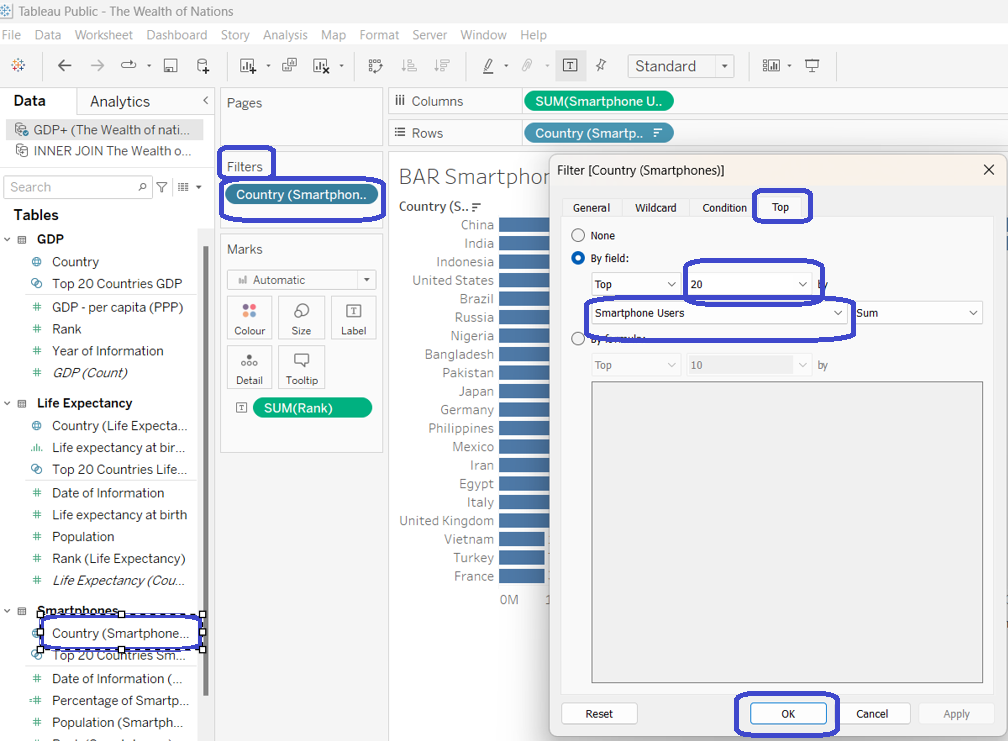






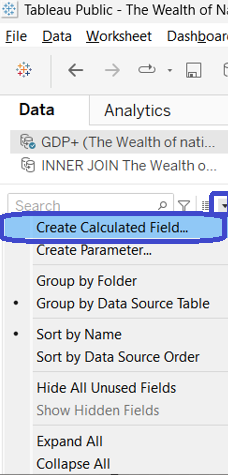
## Change filter to top 20

This method of filtering in the top 20 countries will be used in different scenarios using different Country pills (in this case – countries in the smartphone table.



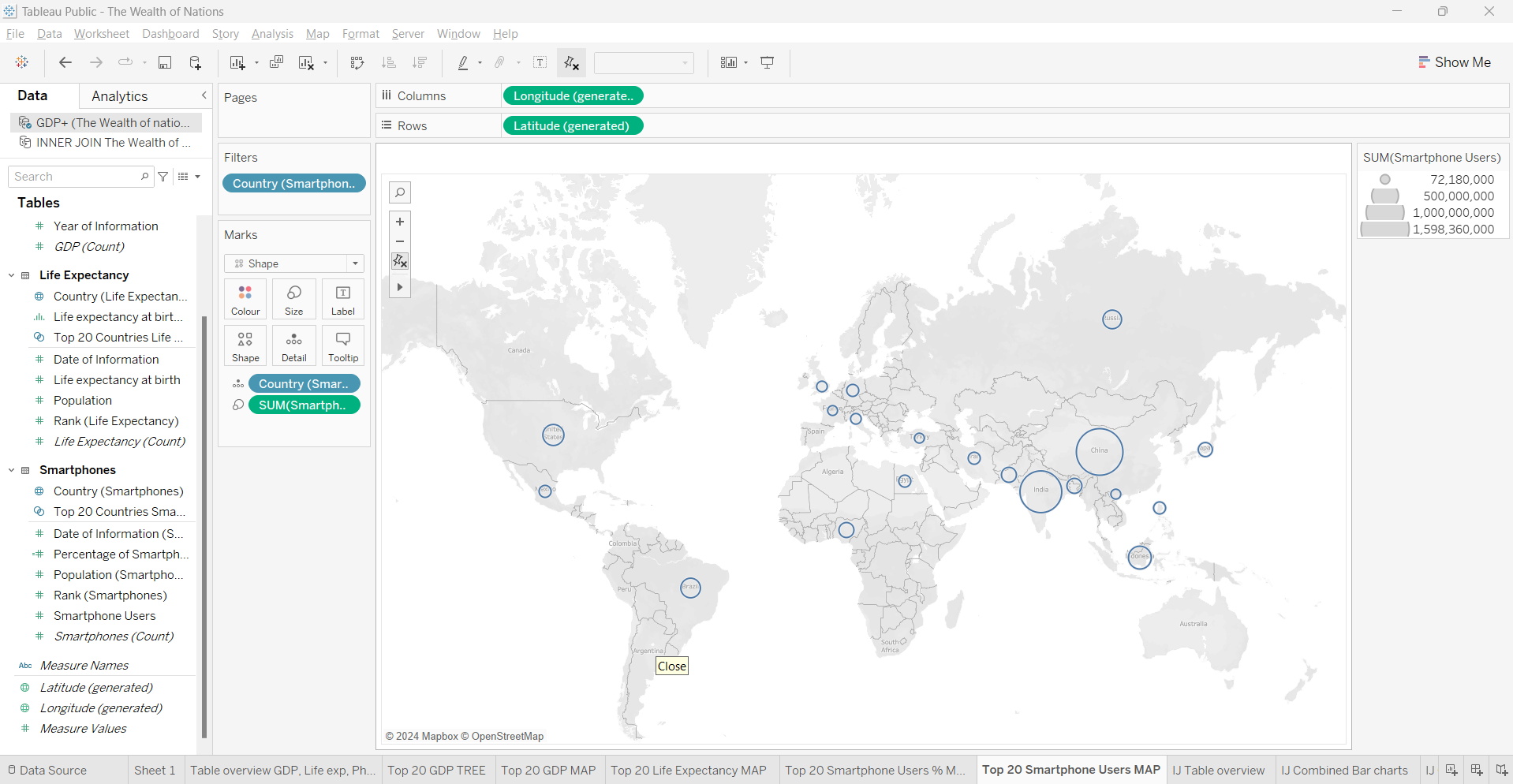
## Ad hoc: Smartphone users as % of population

A new field will be calculated to show smartphone users as a % of the population.





## Tableau chart1



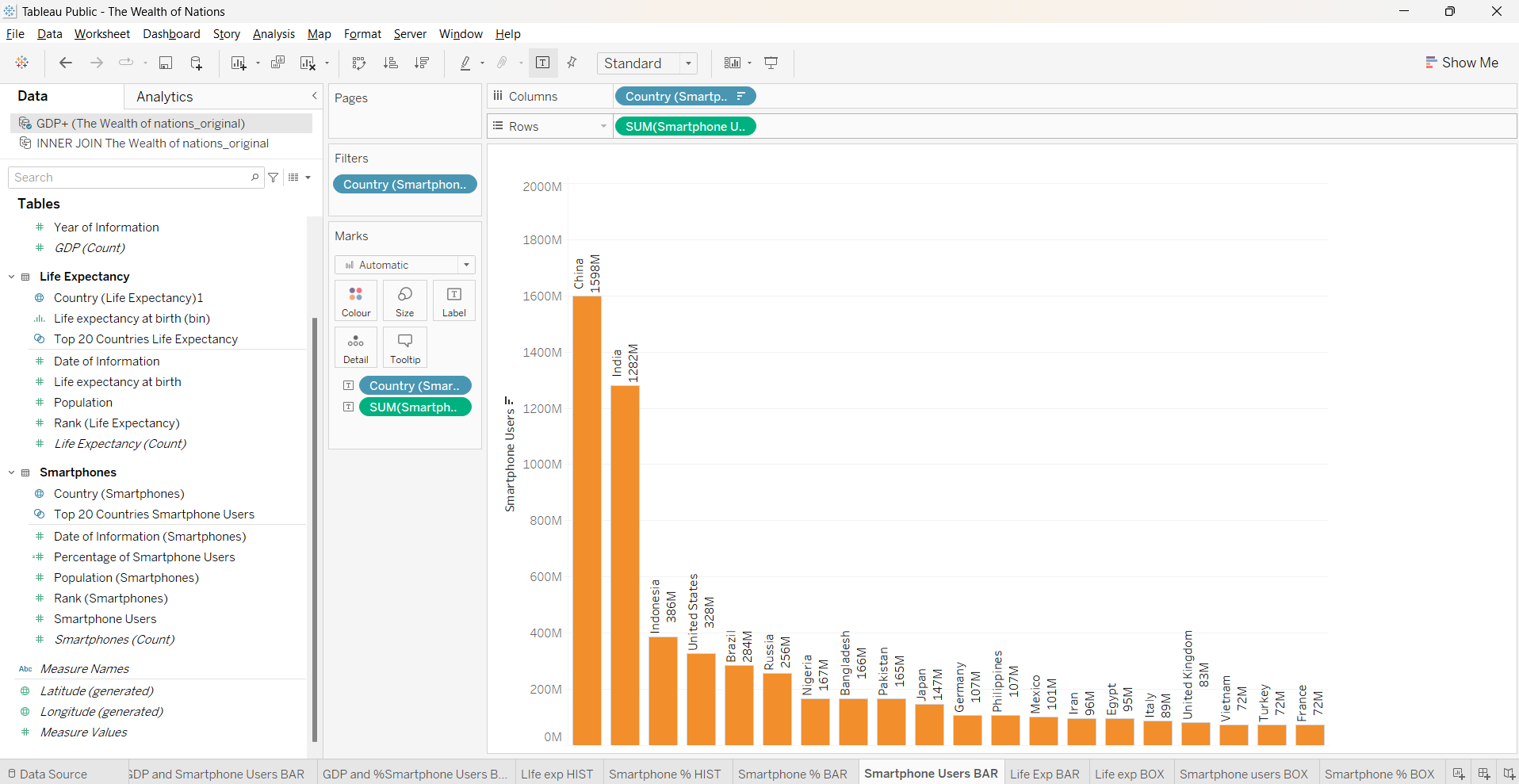
<https://public.tableau.com/app/profile/deepak.assudani/viz/TheWealthofNations_17094238619950/Top20SmartphoneUsersMAP_1?publish=yes>

## Tableau chart2



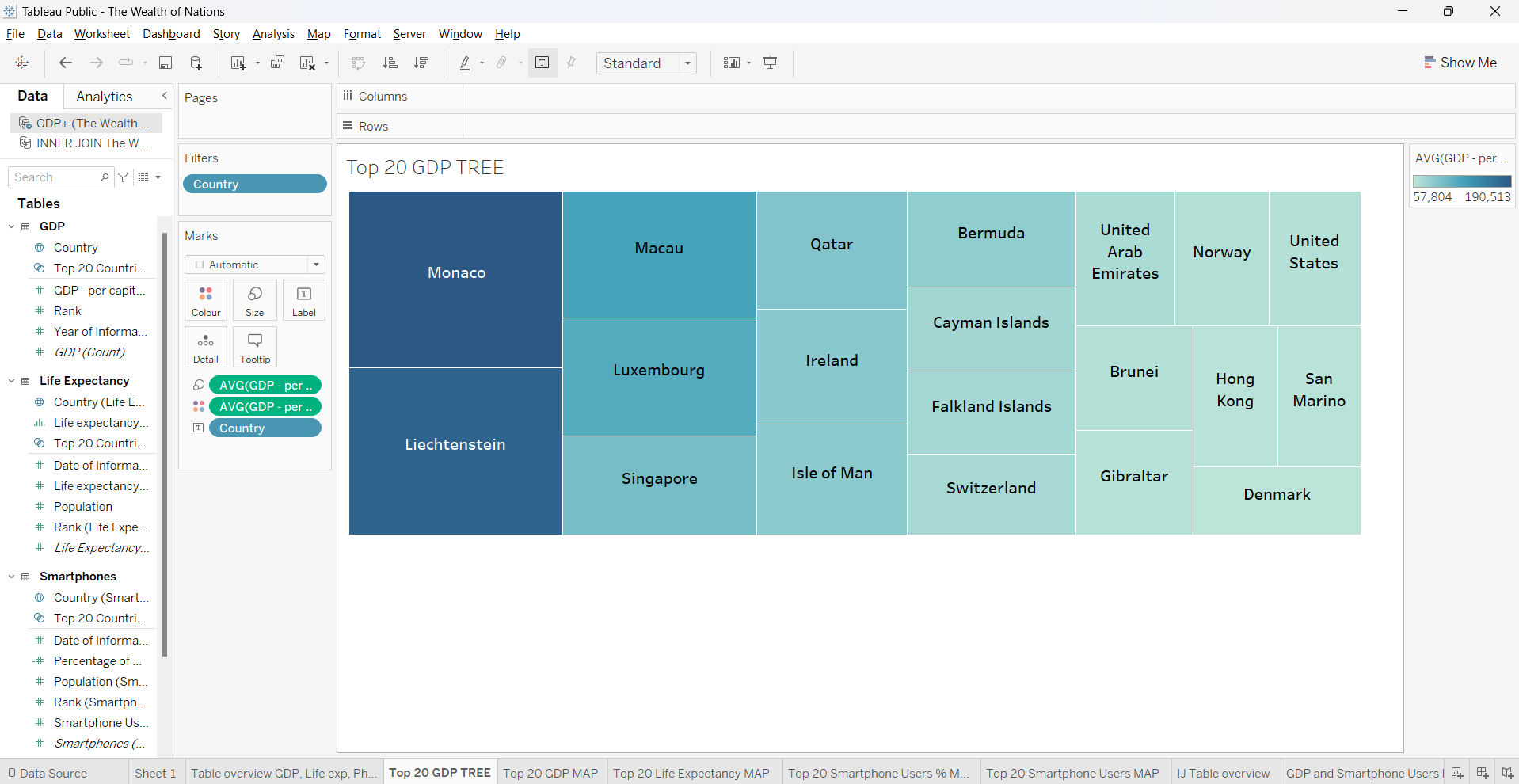
<https://public.tableau.com/app/profile/deepak.assudani/viz/TheWealthofNations_17094238619950/GDPandSmartphoneUsersBAR?publish=yes>

## Tableau chart3



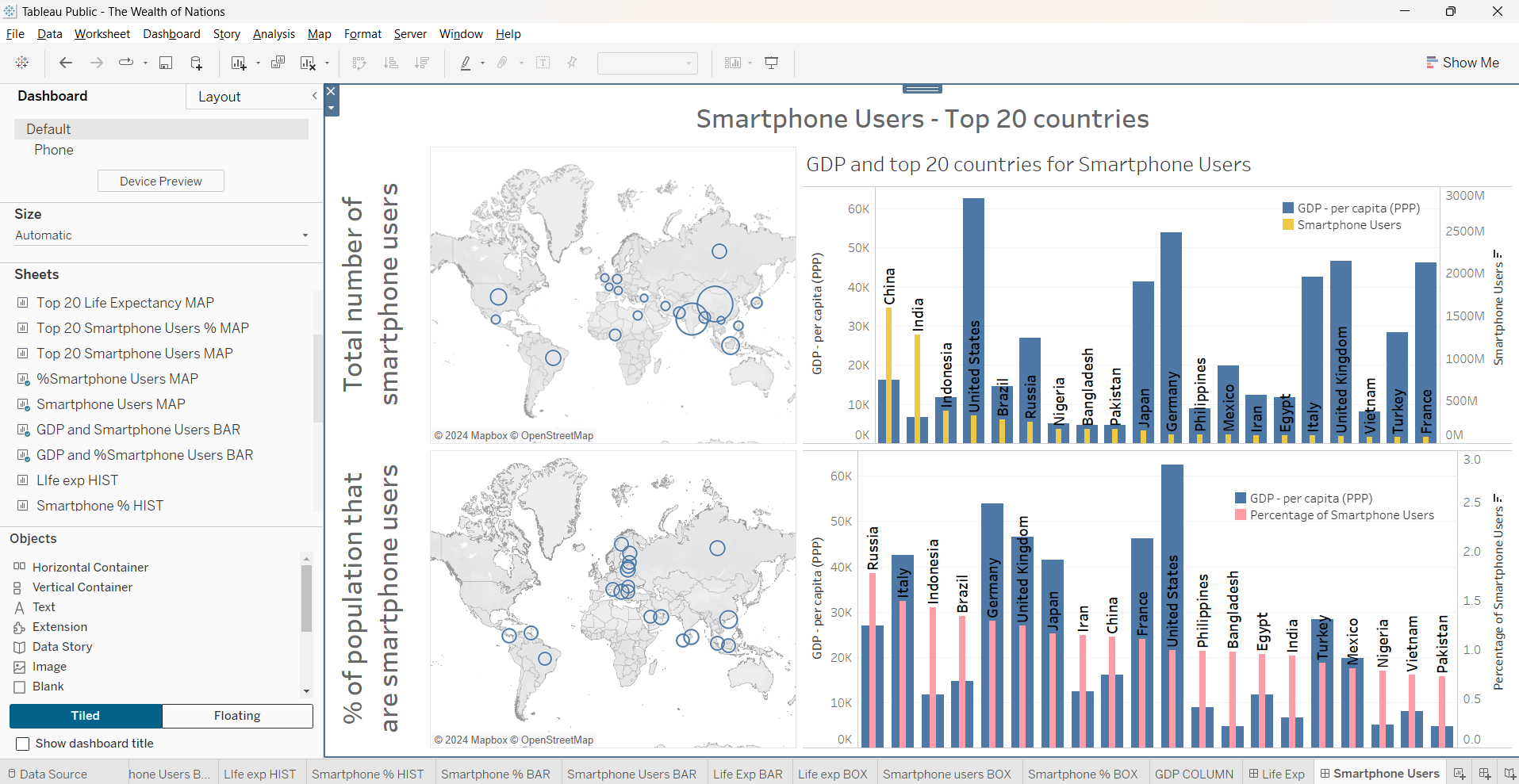
<https://public.tableau.com/app/profile/deepak.assudani/viz/TheWealthofNations_17094238619950/SmartphoneUsersBAR?publish=yes>

## Tableau chart4



<https://public.tableau.com/app/profile/deepak.assudani/viz/TheWealthofNations_17094238619950/Top20GDPTREE?publish=yes>

## Dashboard 1



<https://public.tableau.com/app/profile/deepak.assudani/viz/TheWealthofNations_17094238619950/SmartphoneUsers?publish=yes>

# **Project Reflection**

This project provided a lot of insight into my abilities and areas to improve when it comes to MS Excel and Tableau.

**MS Excel**

I consider myself to be fairly advanced when it comes to MS Excel, especially when it comes to writing query type formulas. In one part of the project where I had to filter in GDP data pertaining to the top 20 countries and in the year 2019, I went outside the scope of the project requirement and created a dynamic formula using a nested FILTER function in order to retrieve the requested data.

I performed a simple exploratory data analysis (EDA) which highlighted the following to me:

* The vast majority of GDP data is only available for 2019
* There is a lot of missing data in GDP for other years
* All of the Smartphone Users and Life Expectancy data are only available for 2020
* When combined, all 3 sets of data were not available for every single country. Some countries had data on all three measures: GDP, Smartphone Users and Life expectancy, but most countries only had partial data.

I also saw one part of the dataset that may lead to a skewed view of the data, specifically the Smartphone Users data, I opted to introduce a new column of population data. Though this was again outside the scope of the project, I felt it necessary, as the number of smartphone users is likely correlated to a country’s population, however, other measures such as GDP and life expectancy were likely to not have shared the same characteristic. Therefore, comparing smartphone users on its own will have introduced some bias. It could be argued that the client only wanted to know the total number of smartphone users in some of these countries, which is why I still kept this information in my visualisations.

However, this project also made me aware of my shortcomings in MS Excel, one of them being my manipulation of charts, which is still quite rudimentary and basic. I ended up doing a bit of research into chart making which also helped facilitate my MS Excel learning. I also realised that MS Excel is quite limited when it comes to data preparation. The FILTER function I used would have likely required less computing power if I did something similar using another programme, such as Python. This has made me realise that, although I feel very confident with MS Excel and I enjoy working with it, for my future prospects I should focus on other data software (whether it be other data preparation software like Python and SQL, or other data visualisation software such as PowerBI or Tableau).

**Tableau**

As for Tableau, I am a complete novice. However, I believe I made some advanced from being a complete beginner. Through some online tutorials and experimentations, I was able to create data relationships, charts with different bar widths, map charts, understand dual axes, and ultimately create a dashboard, which was quite visually appealing to me. I even created a new table using an inner join to combine all three tables, however, as previously discussed, the data was very insufficient, and I believed that any information gleaned from that inner join’ed table will not have provided any informative or representational data, therefore I did not proceed with creating visualisations with it.

However, I believe the skills I currently possess is far from sufficient if I am to look for a job that requires proof of data visualisation skills. There are still many fundamental aspects of Tableau that I still struggle with, including formatting charts, performing inclusion / exclusion criteria, and overall navigation of Tableau, though I believe this will come with time if I put in the effort on watching more tutorials online and experimenting with the software.

There were some parts of the project requirements that I was not able to adhere to (e.g. using Gestalt, using card to show aggregates, showing negative values), as I could not find relevance within the dataset. One example being the Gestalt design principle. As it was already a bit challenging for me to produce basic charts, it was difficult to even think about how to implement much of the Gestalt design principles. Another example being my lack of use of cards to showcase aggregate figures. Within the dataset, each country only had one figure per measure, and it was difficult to find appropriate use of an aggregate display. The only exception being a few countries that had GDP figures for more than 1 year – hence me using average instead of sum when creating my GDP visualisations, but I still did not see what I can achieve by showing aggregate figure here. Another example is negative values, which I was not able to find or calculate the data in any meaningful way to obtain negative values. One area of improvement is that perhaps I could have explored the data further to try and implement these aspects.

Finally, the dataset provided seems to have been quite incomplete, at times the different measures seemed non-relevant to each other. If this was a real life project, I would have most likely queried the source of the data and asked for further information regarding the method of data collection and the intended use of the results. However, I did not do this, and this was another aspect I could have improved on.

**Summary**

In summary, this project has provided me with further insight on my current abilities and a reminder on what to focus on in my data learning journey.