1/12/2021

COVID-19 DATA ANALYSIS

CIS4008-N

BIG DATA AND BUSINESS INTELLIGENCE

IN COURSE ASSESSMENT

SECTION 1 : BUSINESS INTELLIGENCE DESIGN

OLUSESAN, CLEMENT
A0368816

Contents

INTRODUCTION

It is fair to say that the novel coronavirus (CoV) named 2019-nCoV by the WHO (World Health Organisation) or popularly known as COVID-19 has been the worst pandemic the world has seen for at least a decade if not more, it has created the biggest global crisis in generations. It has created an unprecedented challenge to public health, food systems, energy sector and basically every sector of the world's economy, it has led to a dramatic loss of life all around the world and according to WHO (World Health Organisation), nearly half of the world's 3.3 billion global workforce are at risk of losing their jobs. Its spread has devasted many economies and businesses and many Governments are struggling to bring the disease under control and revive their economies. As of 2021, COVID-19 has reached more than 150 nations and has already been declared a worldwide pandemic.

COVID-19 is a pathogenic virus caused by the SARS-CoV-2 virus that began at the beginning of December 2019 in Wuhan City, Hubei Province. Bats occur to have been the COVID-19 virus reservoir from the phylogenetic analysis carried out on it, but the intermediate host has not been detected till now. Coronaviruses mostly causes gastrointestinal and respiratory tract infections and is spread from an infected person's mouth or nose in small liquid particles when they cough, sneeze, speak, sing or breathe. These particles can range from large respiratory droplets to smaller aerosols. It has many common symptoms like fever. Cough, tiredness, loss of taste or smell, sore throat, headache, aches and pains, diarrhoea etc.

However, fortunately there has been many great strides made to curb the spread of the disease by first documenting the cases through vigorous testing which allowed it easier to identify and isolate those that have been infected. The fast and alarming infection rate of COVID-19 incentivized international alliances and government to urgently organize resources to create multiple vaccines as fast as possible. The first vaccine called CanSino was approved by the Chinese government for limited use in the military on June 2020, then Russia also approved Sputnik V vaccine for emergency use. On December 2020, the UK government gave temporary approval for the use of the Pfizer–BioNTech vaccine and since then many different vaccines have been approved and administered all around the world.

OBJECTIVES AND APPROACH

Due to the profound effect that COVID-19 has had on the entire world, I decided to visualize the effects of safety measures such as vaccinations and testing, and to check how effective they have been. Also, I wanted to find some of the correlation between some details about each country in relation to the effects of the disease.

By the end of this report, I expect to have answered the questions below

1. What are the Overall values for confirmed cases, vaccinations, testing and deaths all around the world?
2. What are the total cases, Deaths, Tests and Vaccinations for each continent?
3. What are the rates of vaccinations, cases, deaths and tests since the beginning of the year?
4. What are the rates of daily cases and daily deaths?
5. In which continents and countries are the vaccination programme more advanced?
6. Which country has achieved or is closest to herd immunity?
7. What is the relationship between the confirmed cases and the deaths caused by covid-19?
8. The correlation between a Country's GDP and the death rate.
9. The Effects of Covid-19 on the number of ICU patients admitted.
10. Does the median age of a country affect the death rate of Covid-19 in that country?
11. What are the key influencers that affects each continent?

DATASET

1. THE DATA SOURCE

The dataset used in this report was helpfully provided by a GitHub user, who collated the data from various sources such as the COVID-19 Data Repository by the centre for Systems science and Engineering (CSSE) at John's Hopkins University (JHU), the European Centre for Disease Prevention and Control (ECDC) and some other sources. The link to the dataset is below:

https://github.com/owid/covid-19-data/tree/master/public/data

2. DATASET DESCRIPTION AND TABLE

The downloaded dataset is a CSV file that has one single table which contains 67 columns and 147495 rows. However, I only need 22 columns for the data I want to visualize so I removed the unneeded columns as will be shown the data pre-processing step. The table below shows the used column names and a description for what each column data represents.

Column Name	**DESCRIPTION**
total_cases	Total confirmed cases of COVID-19
new_cases	New confirmed cases of COVID-19
Total_deaths	Total deaths attributed to COVID-19
new_deaths	New deaths attributed to COVID-19
icu_patients	Number of COVID-19 patients in intensive care units (ICUs) on a given day
total_tests	Total tests for COVID-19
new_tests	New tests for COVID-19 (only calculated for consecutive days)
total_vaccinations	Total number of COVID-19 vaccination doses administered
people_vaccinated	Total number of people who received at least one vaccine dose
people_fully_vaccinated	Total number of people who received all doses prescribed by the vaccination protocol
total_boosters	Total number of COVID-19 vaccination booster doses administered (doses administered beyond the number prescribed by the vaccination protocol)
new_vaccinations	New COVID-19 vaccination doses administered (only calculated for consecutive days)
iso_code	ISO 3166-1 alpha-3 – three-letter country codes
continent	Continent of the geographical location
location	Geographical location
date	Date of observation

population	Population of each geographical location
Median_age	It's is the age which divides a population into two numerically equally sized groups.
aged_65_older	Share of the population that is 65 years and older, most recent year available
gdp_per_capita	Gross domestic product at purchasing power parity (constant 2011 international dollars), most recent year available
female_smokers	Share of women who smoke, most recent year available
male_smokers	Share of men who smoke, most recent year available

LOADING THE DATA

The first step is to load the data into Power BI which can be done by using the **Get Data** button in the home ribbon, then there are two ways to load the data into Power BI, since Power BI supports a range of data from various sources such as excel, web API, text/CSV and more. We could just click on the Text/CSV file upload but here, the second method was chosen which is using Blank Query Editor and M language to upload the data.

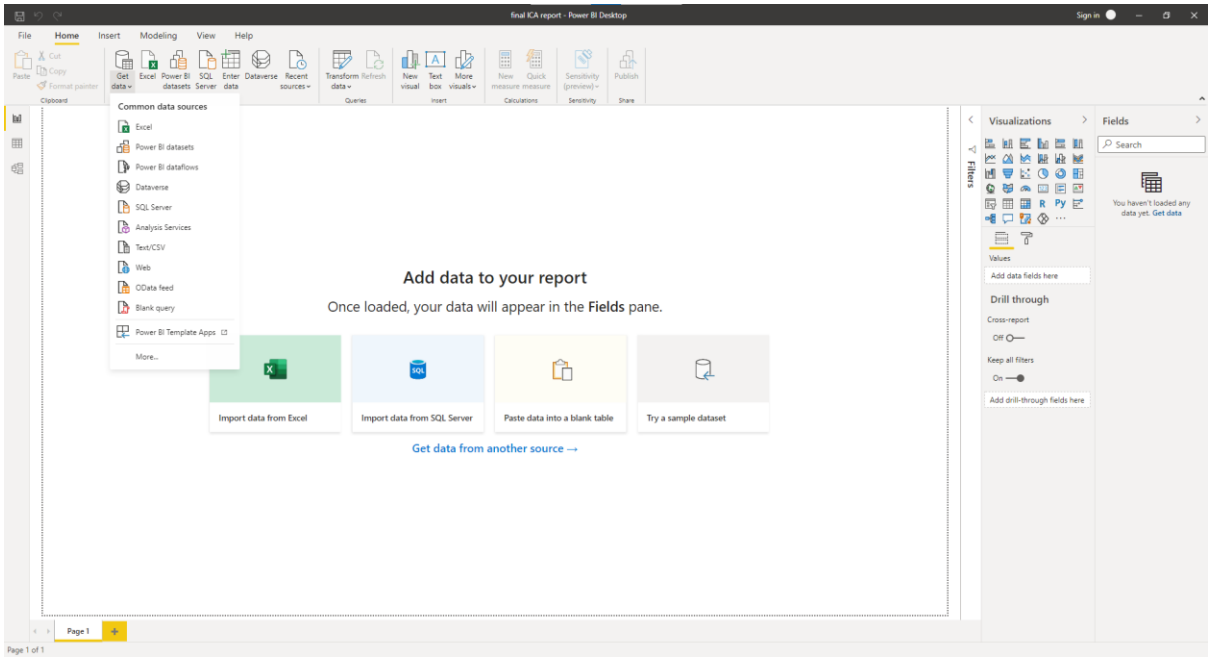


Figure 1 - loading data using blank Query

After Clicking on the Blank Query tab, the Power Query Editor window appears and then the Advanced Editor is chosen to create a new query. Using M language, the data is then loaded from

the file path using the code as shown in the figure below.

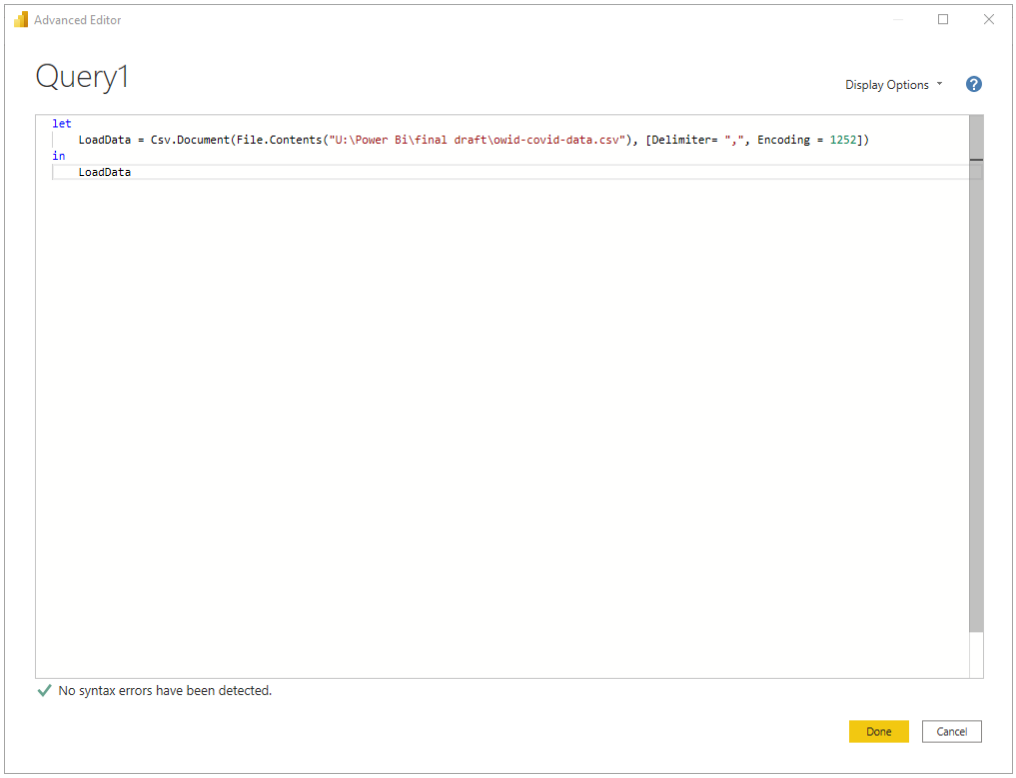


Figure 2 - M language code to load the data

The data was successfully loaded, and the Query name was changed to 'COVID-19 data', then the **Close and Apply** button was selected to the apply our changes and close the Power Query Editor window when it was done as shown below

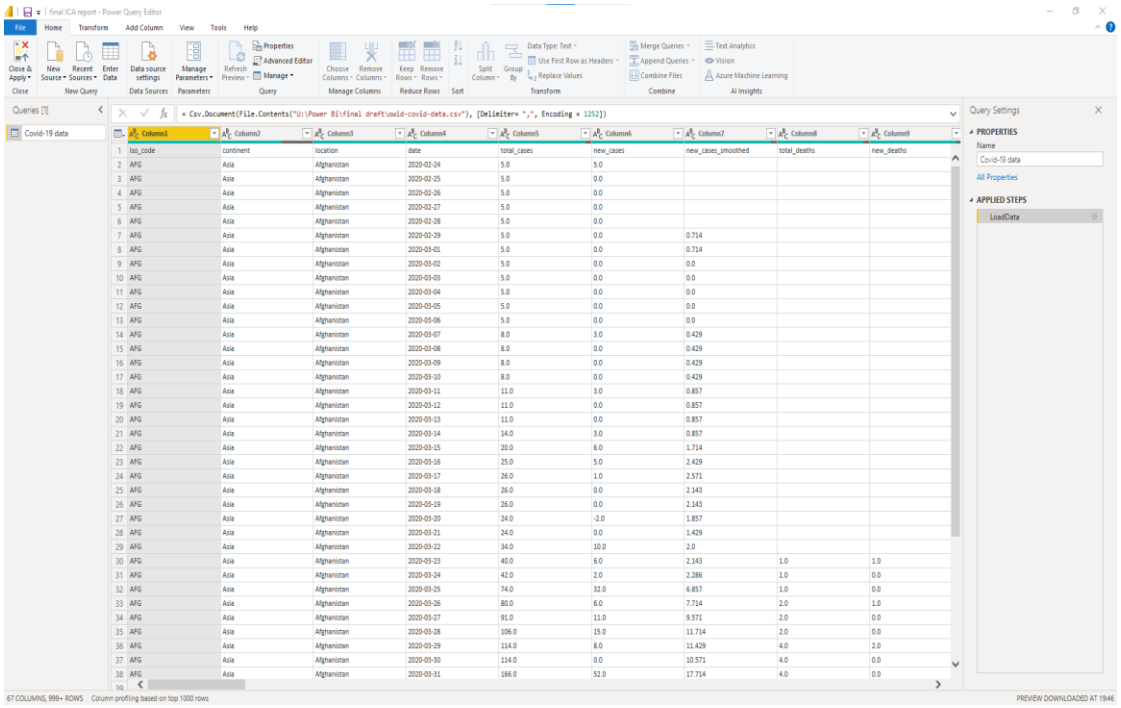


Figure 3 – Changing table name

DATA PRE-PROCESSING OR DATA - CLEANING.

After loading the data, the next step is the data pre-processing, cleaning and transforming the data to generate the fact and dimension tables. To start this process, we opened the Power Query Editor Window by selecting the Transform Data button as shown below

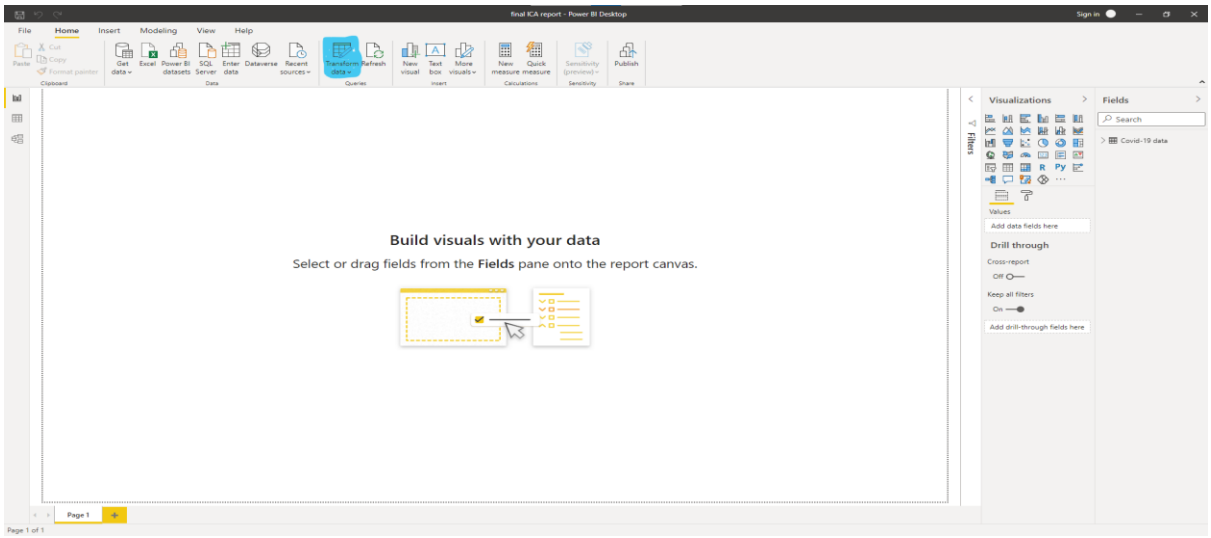


Figure 4 - The transform Data button

1. Removing Unused Columns.

My dataset has many unneeded columns, so the first step is removing those and leaving only the needed columns. To do that, we opened the Advanced editor and wrote the code to remove the columns as shown below

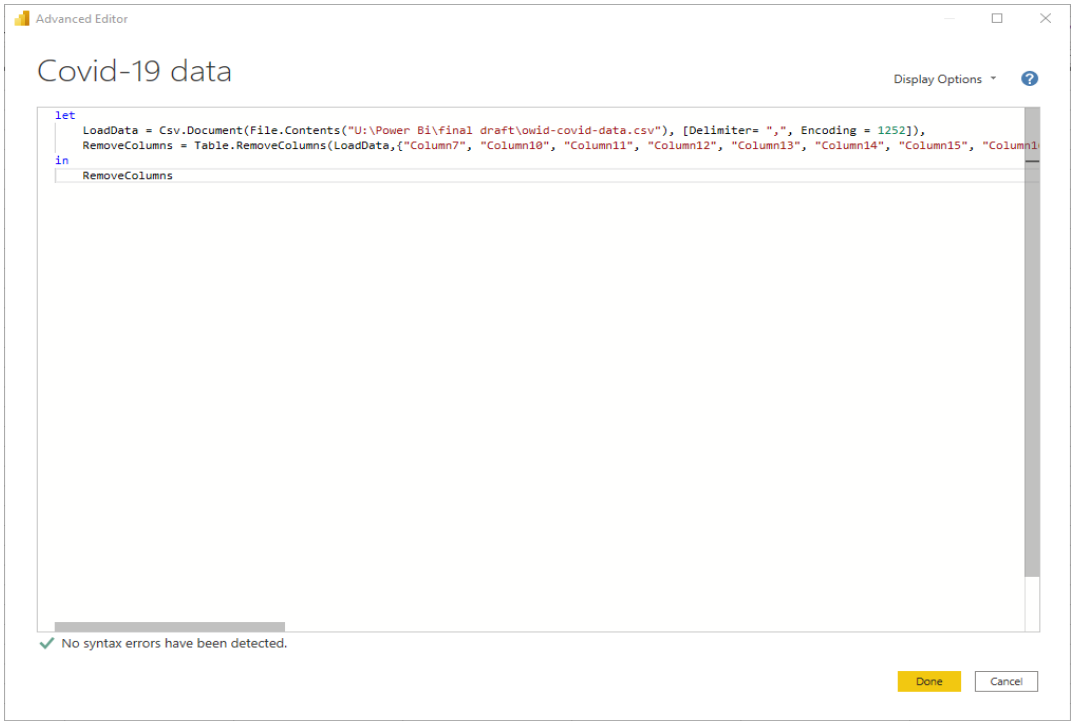


Figure 5 - M-Language code for removing columns

After removing the unneeded columns, there are only 22 columns left but the data is still very disorganised as shown below.

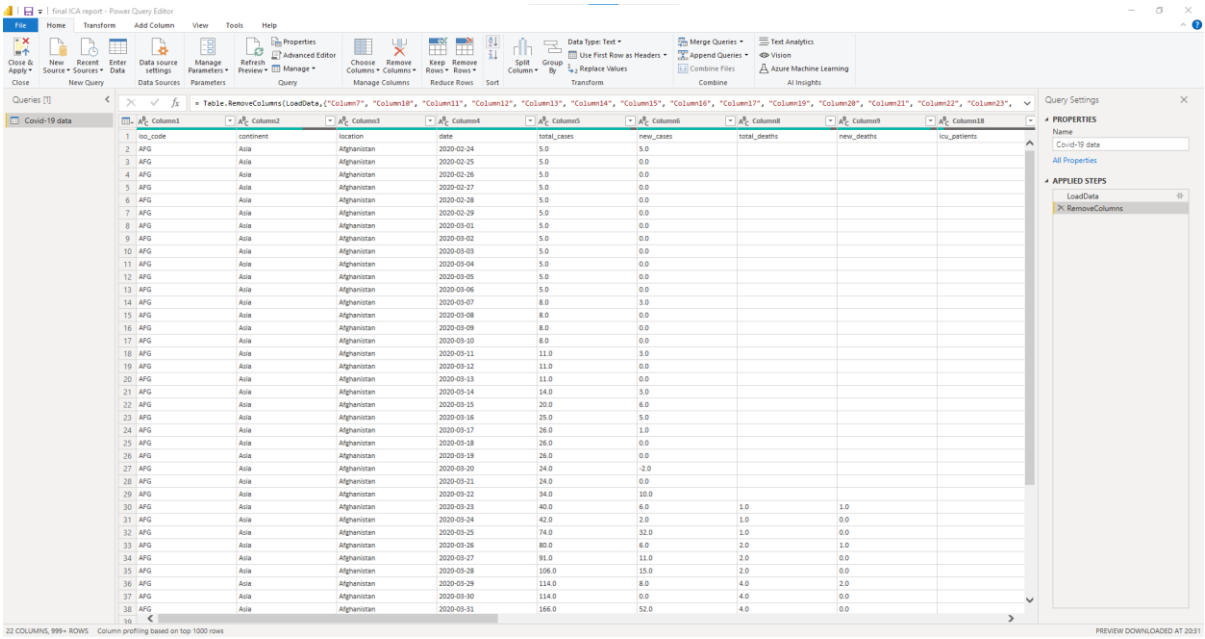


Figure 6 - Disorganised Data

2. Promoting Column names.

The next step is to promote the column names. To do this, we opened the advanced Editor using the steps explained above and wrote the code for promoting column names as shown below.

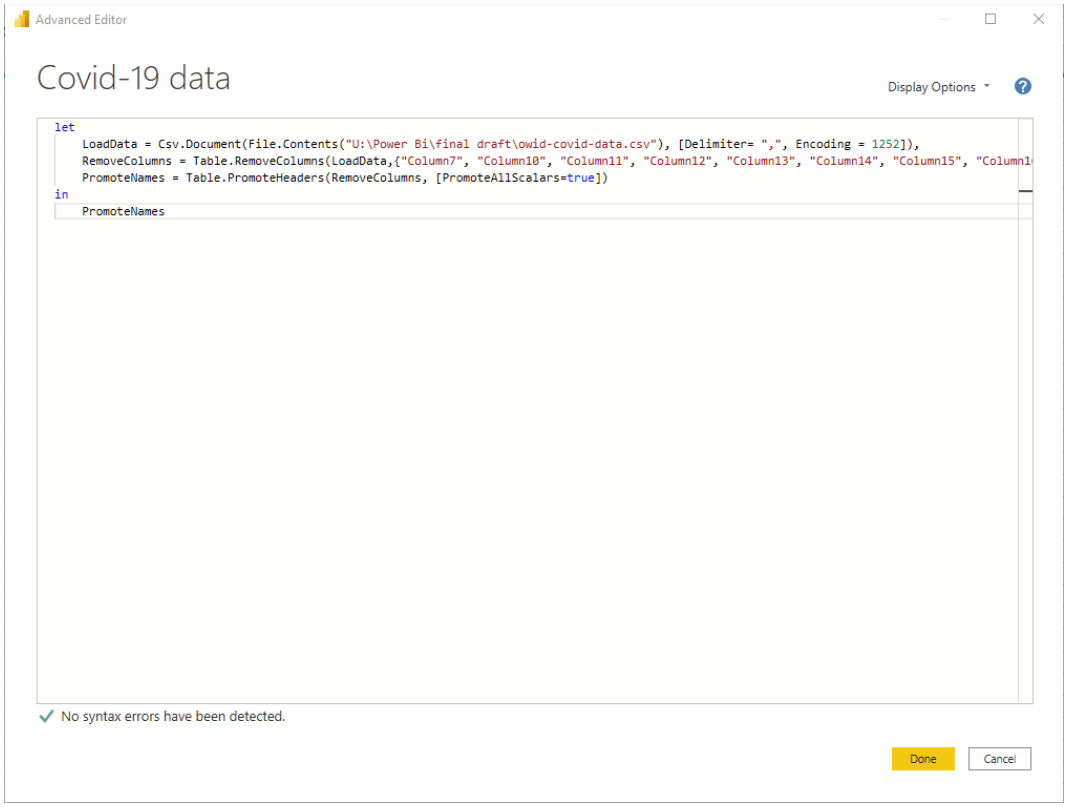


Figure 7 - M-language Code for promoting column Names

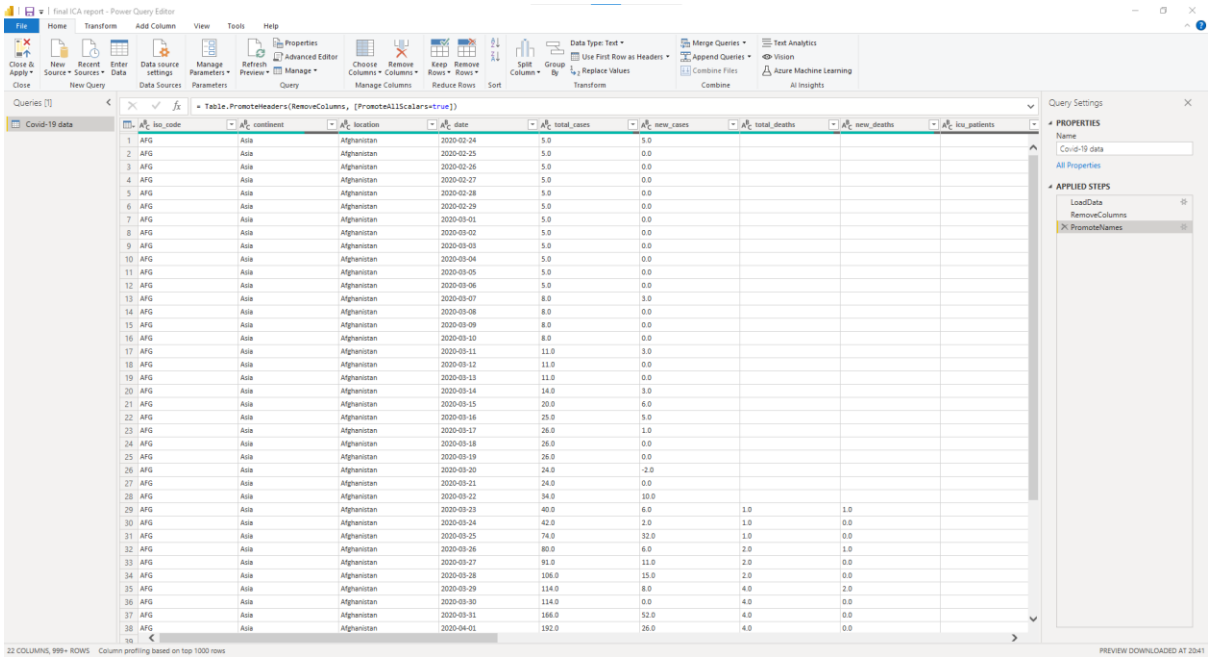


Figure 8 - Table showing after promoting the names

3. Changing Data types

The next step is assigning the correct data types to each column, since Power Bi has already assigned the Text Data type to all, some of the columns will not need their data type changed but to change the others, We again use M-language as shown below to change the data types of ‘total_cases’, ‘new_cases’, ‘total_deaths’, ‘new_deaths’, ‘icu_patients’, ‘new_tests’, ‘total_tests’, ‘total_vaccinations’, ‘people_vaccinated’, ‘people_fully_vaccinated’, ‘total_boosters’, ‘new_vaccinations’ and ‘population’ to Whole number data types and the date column into date data type and the gdp_per_capita column into fixed decimal number data type.

Figure 9 - M-Language code for changing data types

4. Renaming the location column

To make the data slightly easier to read and easily understandable, I decided to rename the location column to 'country' to easily show what the column represents and differentiate from the region column.

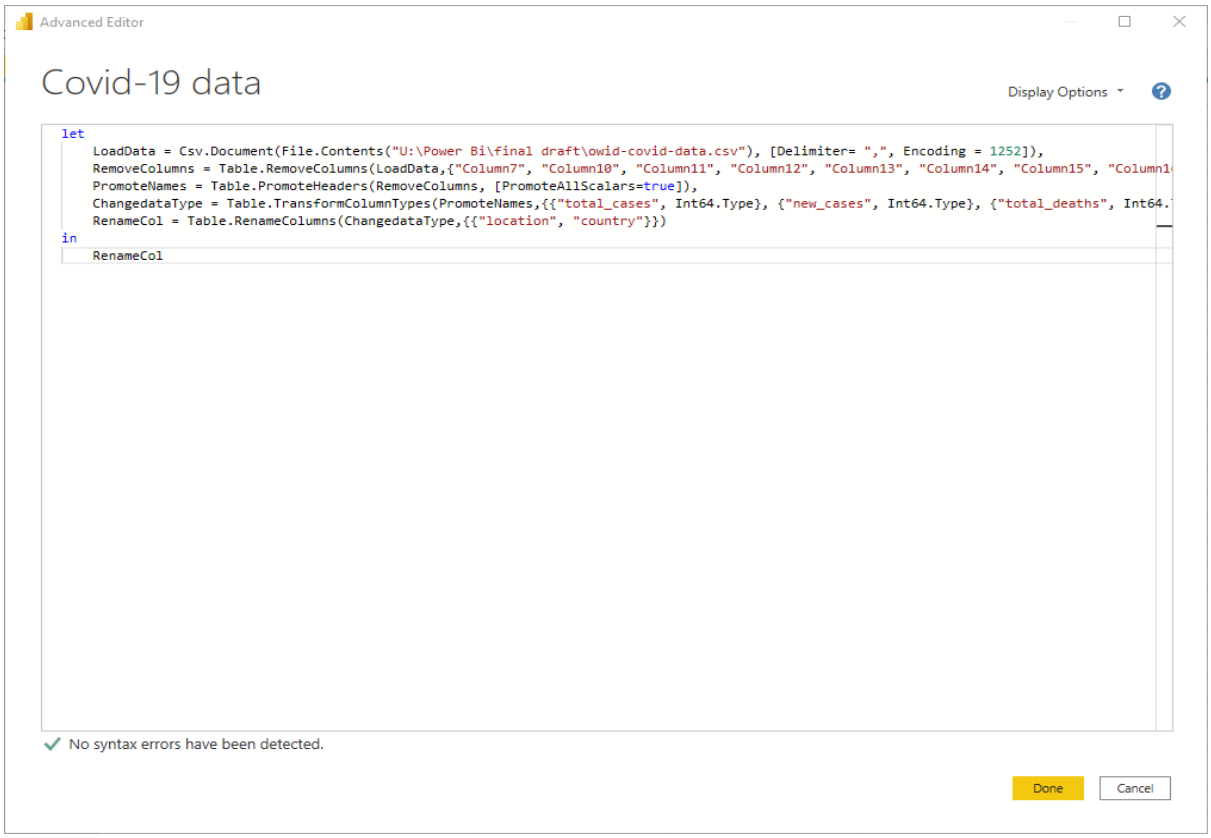
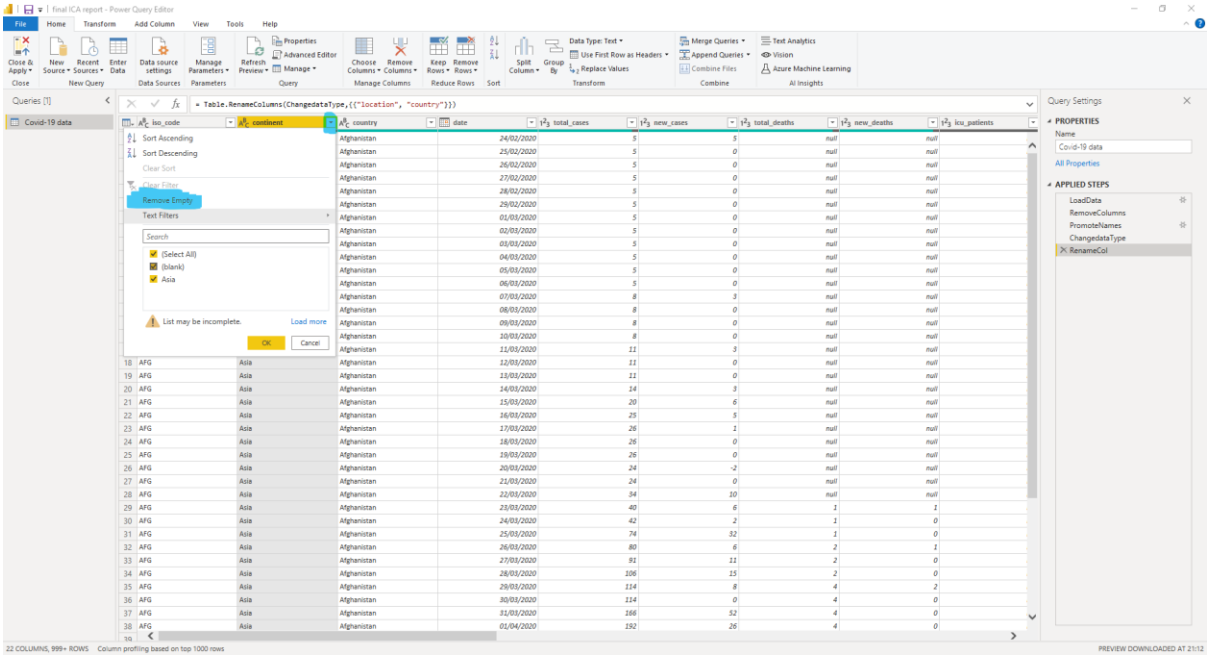


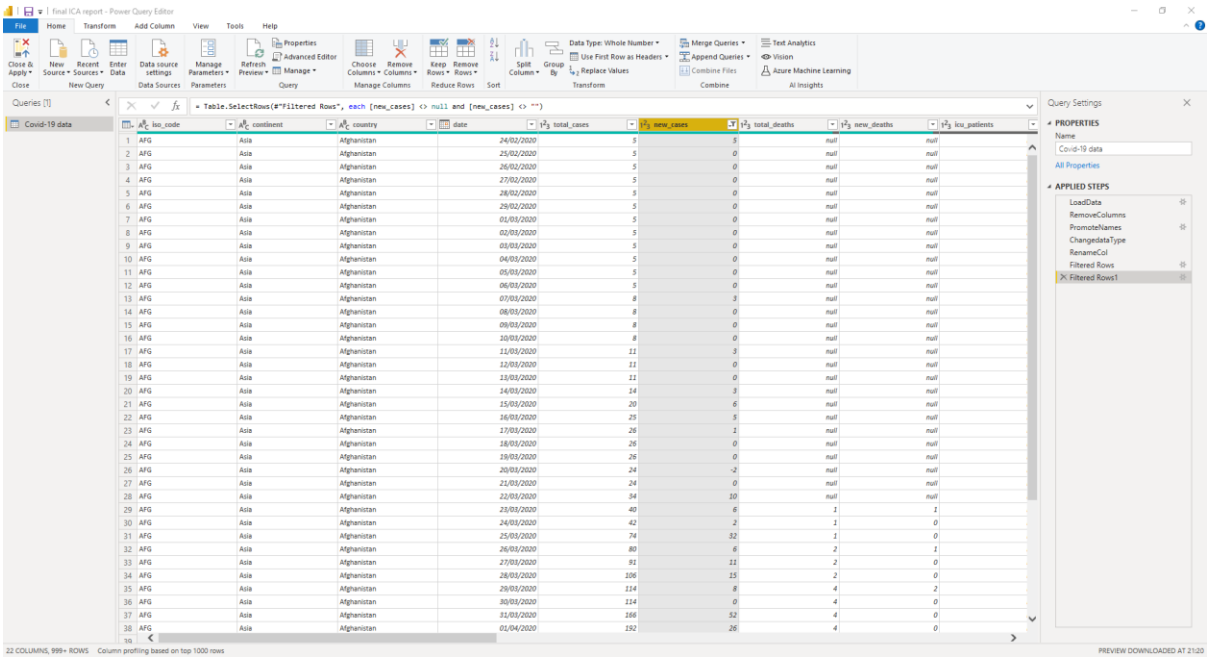
Figure 10 – M-language to Rename column

5. Removing empty values in the dataset

The next step is removing all the empty values or errors in the data. To do that, the first thing I did was look critically at my data and I quickly discovered that the empty rows in the continent column was affecting the country column so removing those empty columns would correct both columns. Instead of using M-language, I decided to let Power BI help me by just clicking on the icon in the continent column and selecting 'Remove Empty' as shown below



Then the next thing I did was repeat the same process to remove the empty rows in the ‘new-cases’ column. The figure below shows the table after the removal of those empty values.



Then instead of removing the empty values in the other columns with empty columns, I instead replaced the values with ‘0’. This was done to show how many deaths, vaccinations, testing were available in relation to the ‘cases’ column. Since the first data collated were the cases data, the other columns need to reflect that there were initially 0 records for them when cases were being collated.

Using M-language, I replaced the values in ‘total_deaths’, ‘new_deaths’, ‘total_tests’, ‘total_vaccinations’, ‘people_vaccinated’, ‘people_fully_vaccinated’, ‘total_boosters’, ‘new_vaccinations’ columns as shown below

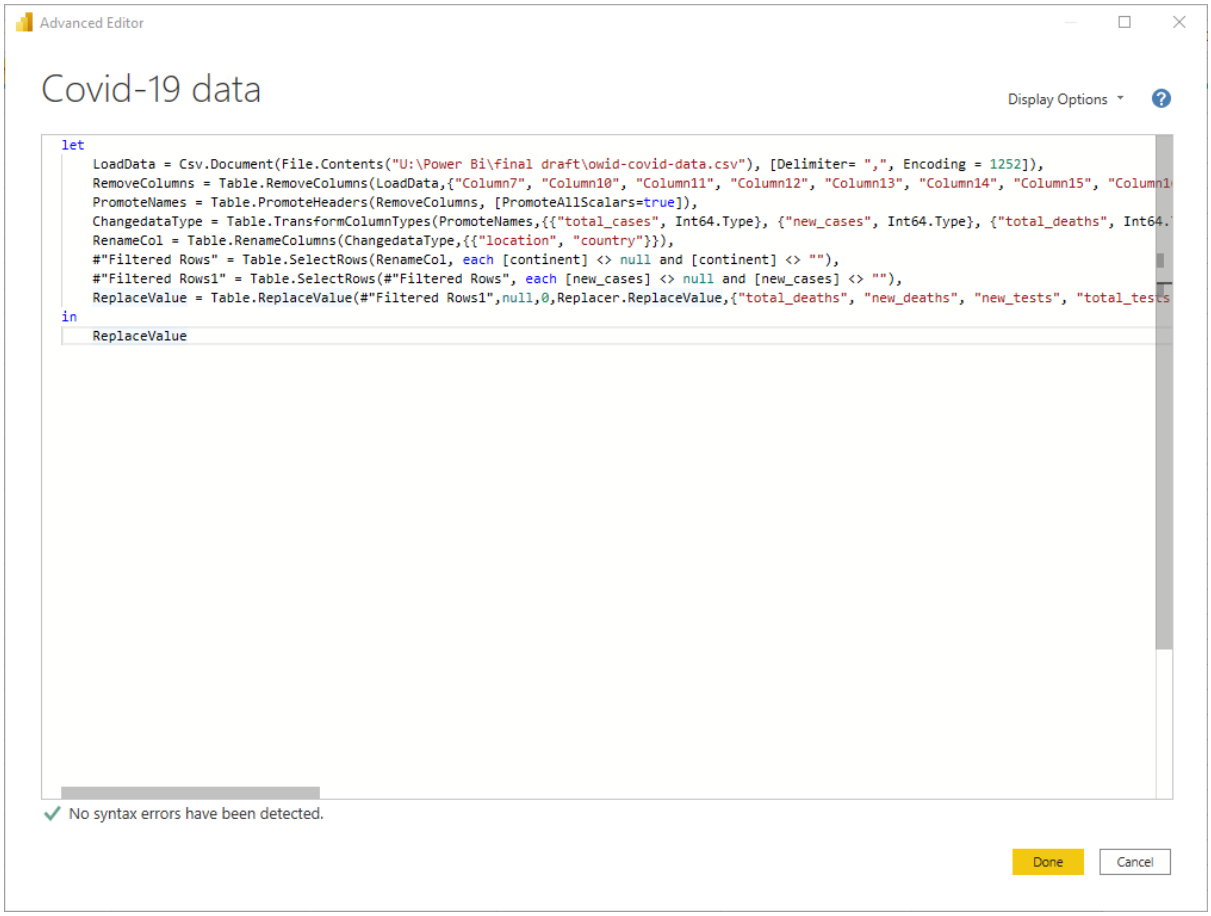


Figure 11 - M-language to replace values

Now the data looks much better as shown below

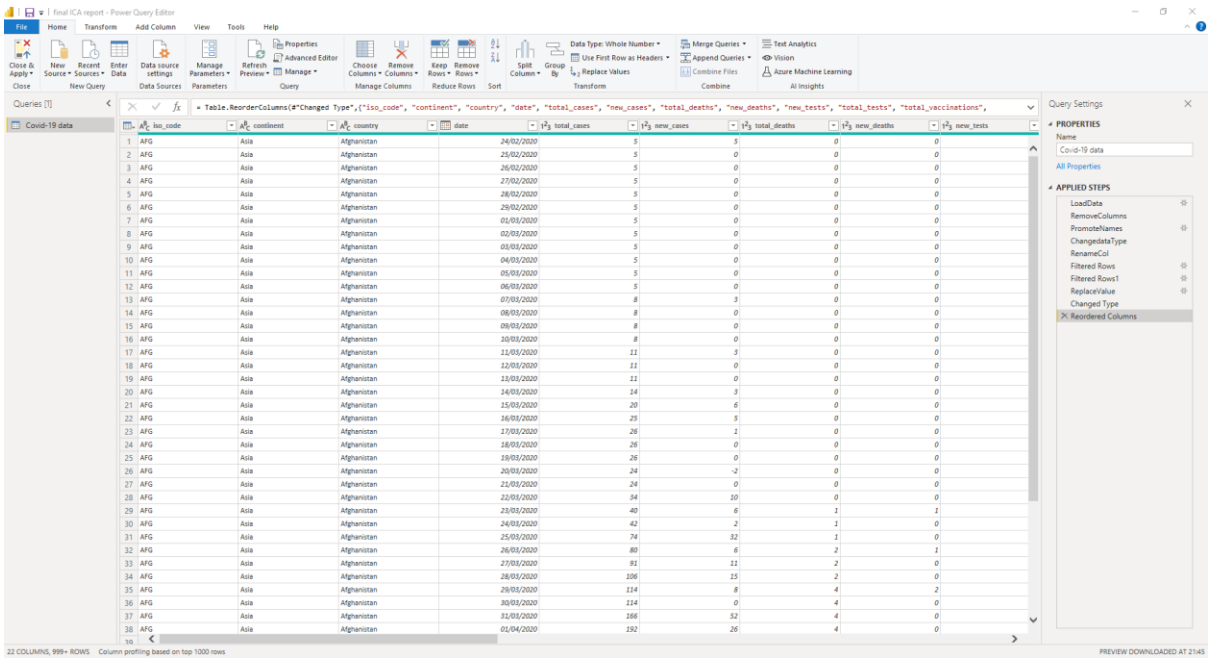


Figure 12 – Showing the organised data

6. Creating the Dimension Tables

After cleaning my data, I am left with a single dataset with 22 columns and 135649 rows. This table has many different types of record and as all the information is represented in this single flat table, it becomes difficult to handle or visualize the data well therefore this flat table was normalized into several dimension tables and one fact table to form a star schema.

- Creating Location Dimension table.

To create the Location Dimension table, a duplicate Covid-19 data table was created by right-clicking on the table and selecting 'duplicate' as shown below

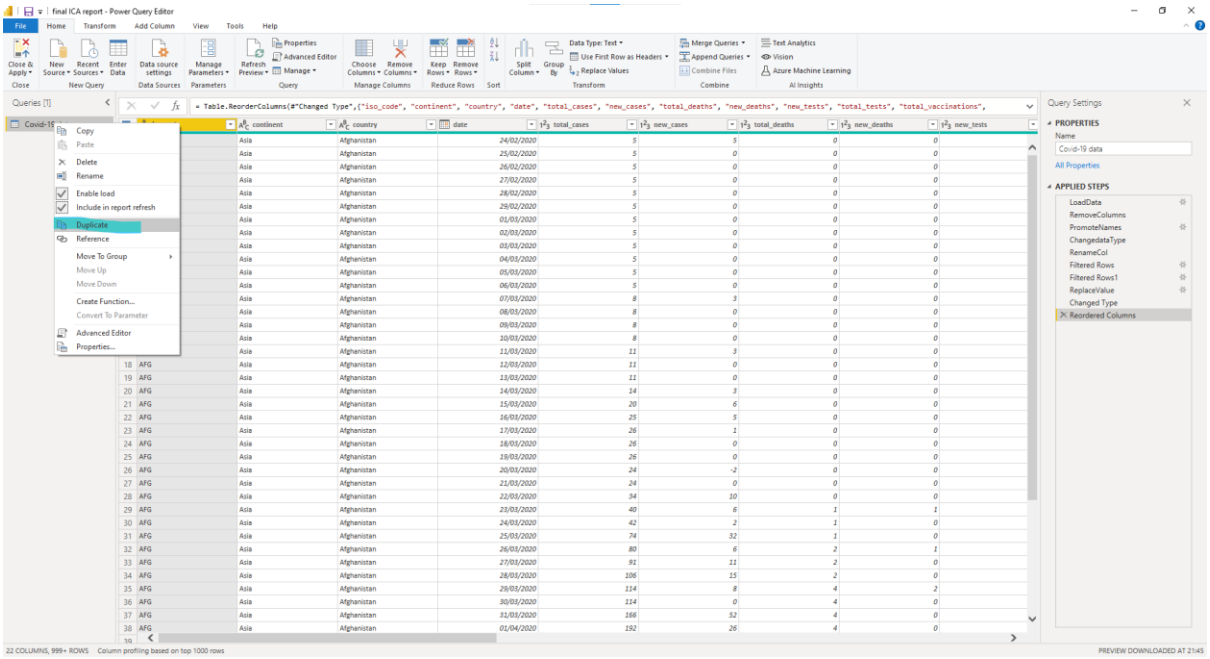


Figure 13 - Showing how to make duplicates of a table

Then the newly created Table was renamed Location and all the columns except 'Iso_code', 'Continent' and 'country' were removed. To do that, the three columns were selected with a right click and 'Remove other columns' option was selected as shown below

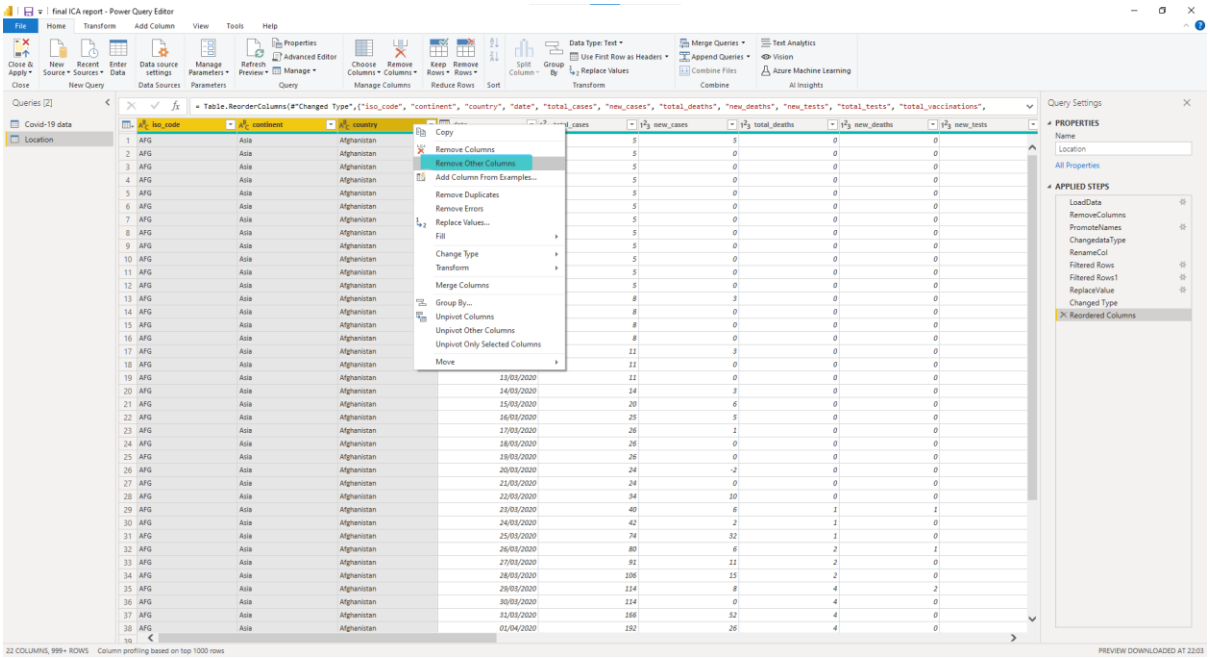


Figure 14 - Showing how to remove other columns

Dimension tables need to have unique values so to do that I needed to remove the duplicated rows. To do that I selected all the columns in the table and right-clicked on them, then I selected the 'Remove duplicates' tab as shown below

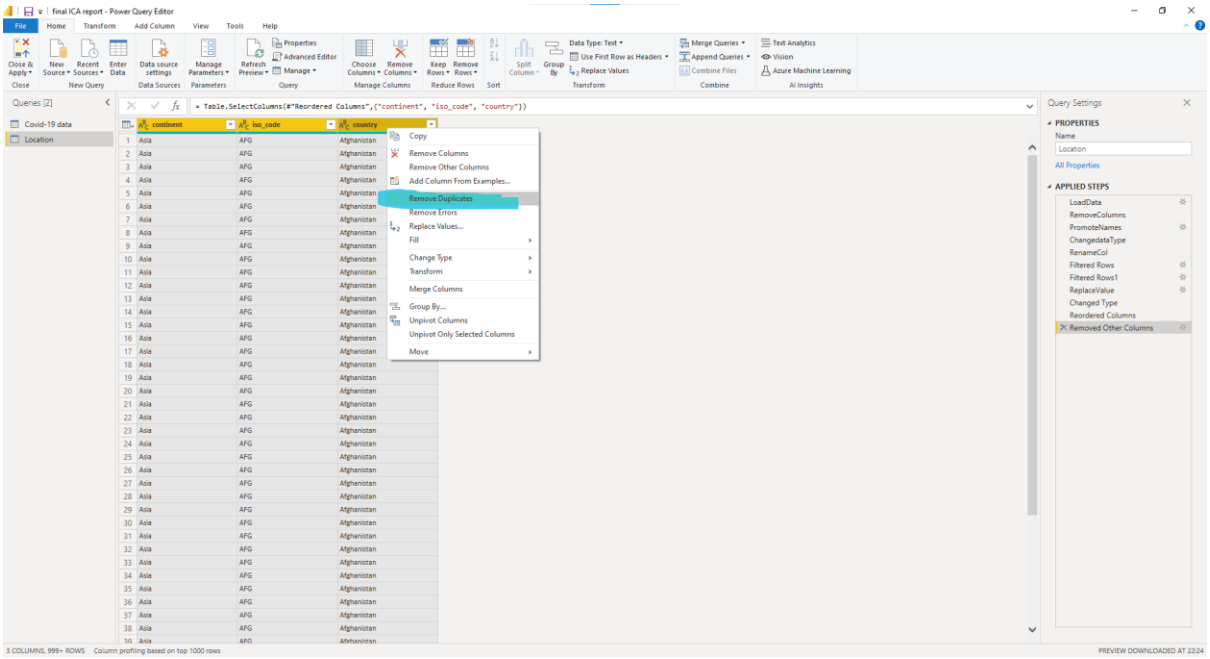


Figure 15 - Removing duplicates from columns

So, the data in the three columns now have distinct values as shown below

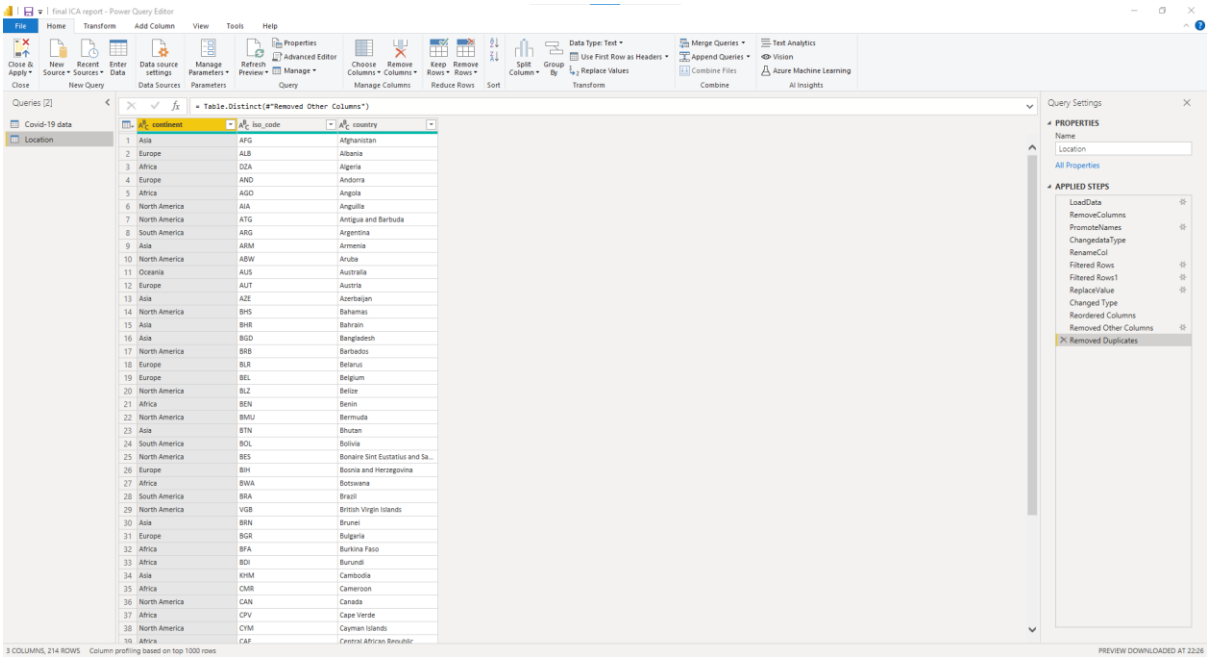


Figure 16 - Showing the new table

- Creating the cases dimension table

To create the cases table, I first needed to create a unique column that will link it to the fact table when we are creating our relationships. To do that I first duplicated the 'total_cases' and 'new_cases' columns by right- clicking on each column and choosing the 'Duplicate column' as shown in the figures below

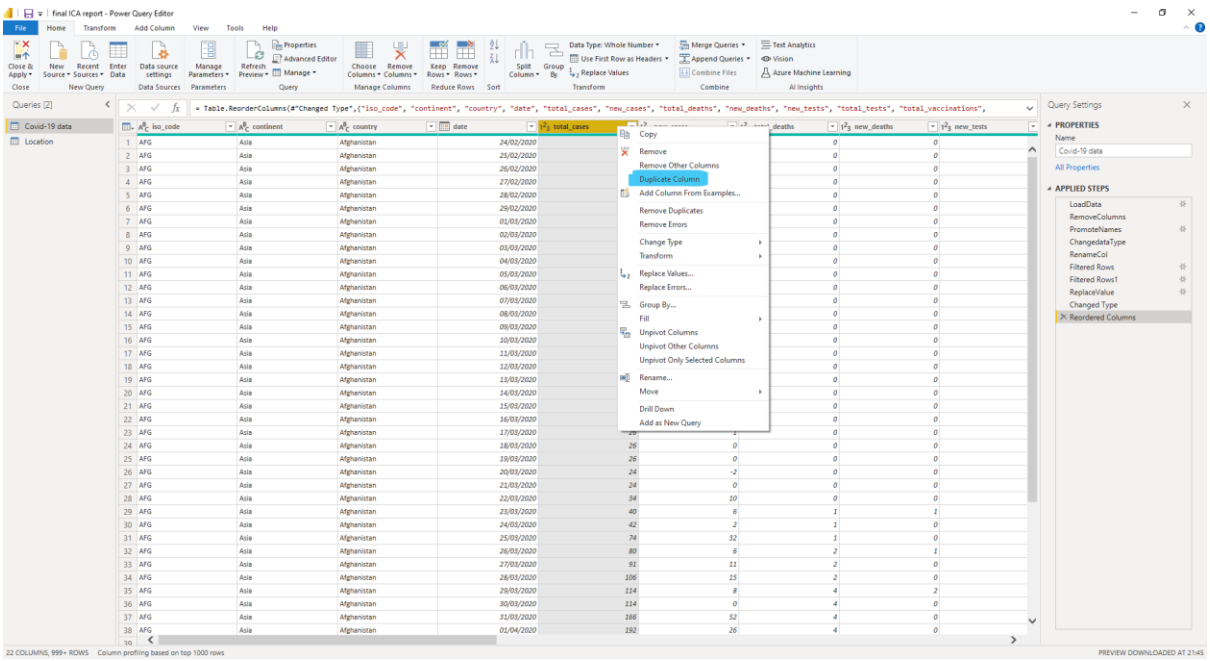


Figure 17 - Duplicating columns

Then I merged the two newly created columns and renamed it Cases_ID, to do that I needed to first select both new columns and then click on the transform tab, then I selected the merge columns button.

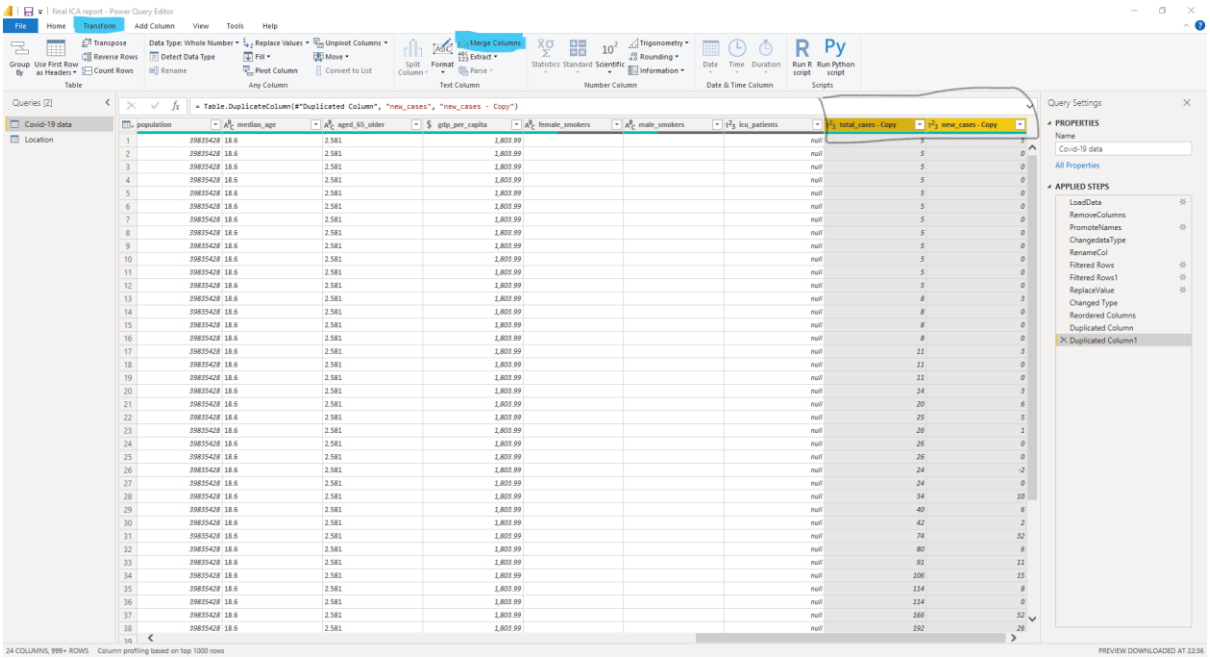


Figure 18 - Creating Case_ID column

A dialog Box appears and in it we change the column name to Cases_ID and then I clicked the ok button.

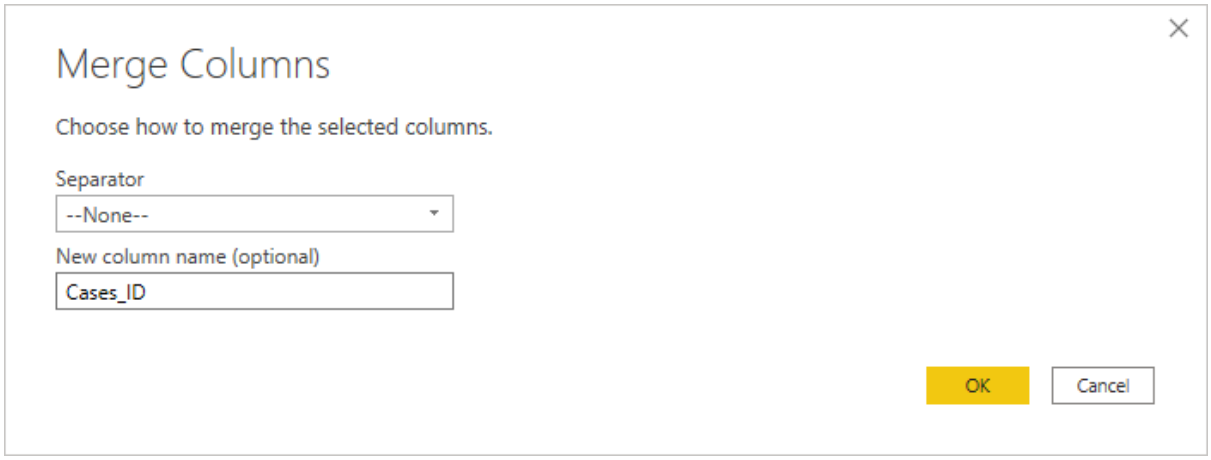


Figure 19 - merge Box

After that is done, I then needed to create a new table as we did for the location dimension table. After creating the new table, I renamed it ‘Cases’ and then removed every column except the ‘total_cases’, the ‘new_cases’ and the ‘Case_ID’ columns. Then Iremoved duplicates from all the columns as done in the location table and as shown below

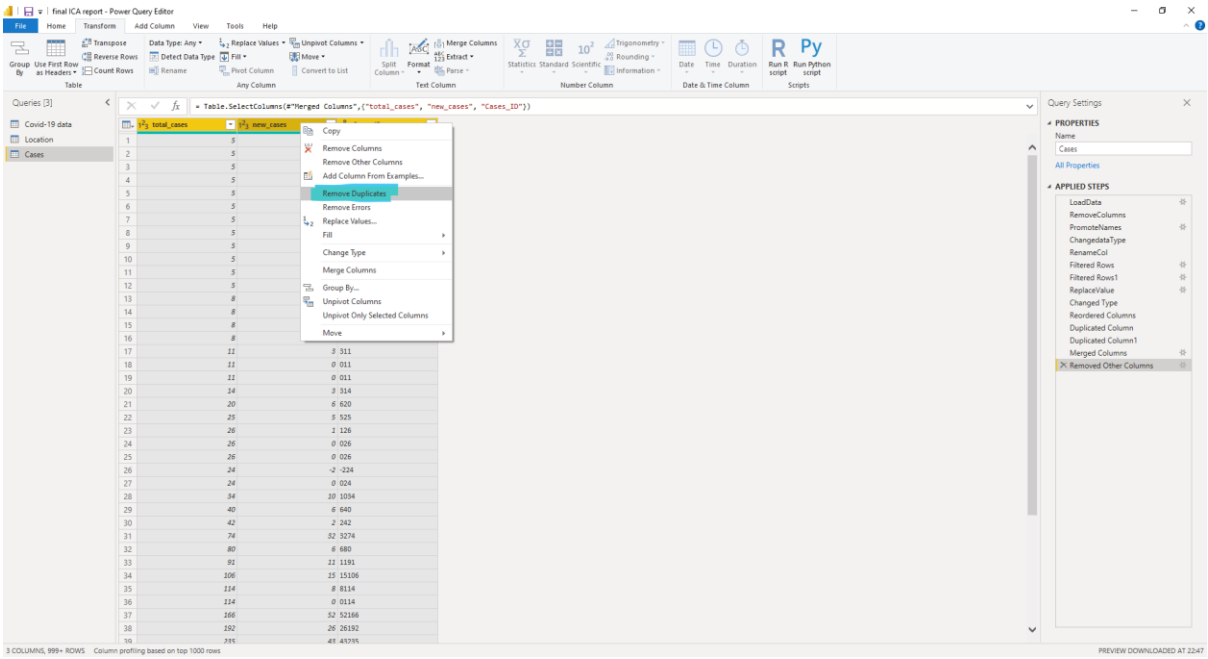


Figure 20 - removing duplicates from cases table

- Creating Deaths dimension table

I created the deaths dimension table using the same process I used to create the cases dimension table, I first duplicated the 'total_deaths' and 'new_deaths' columns, then merged the two columns and renamed the new merged column as Deaths_ID as shown below

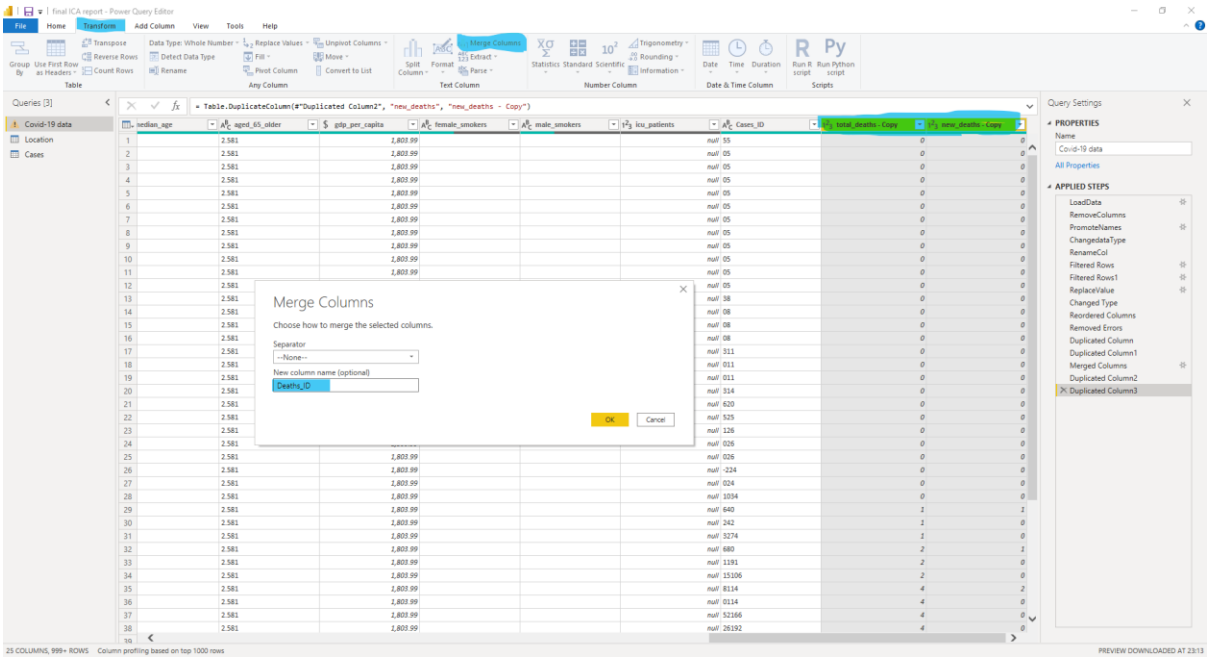


Figure 21 - duplicating and merging deaths column

Once that was done, I then duplicated the covid-19 data table and renamed it 'Deaths', then all the columns except the 'total_deaths', 'new_deaths' and 'deaths_ID' were removed and then I removed the duplicates from all the columns as shown in the figures below

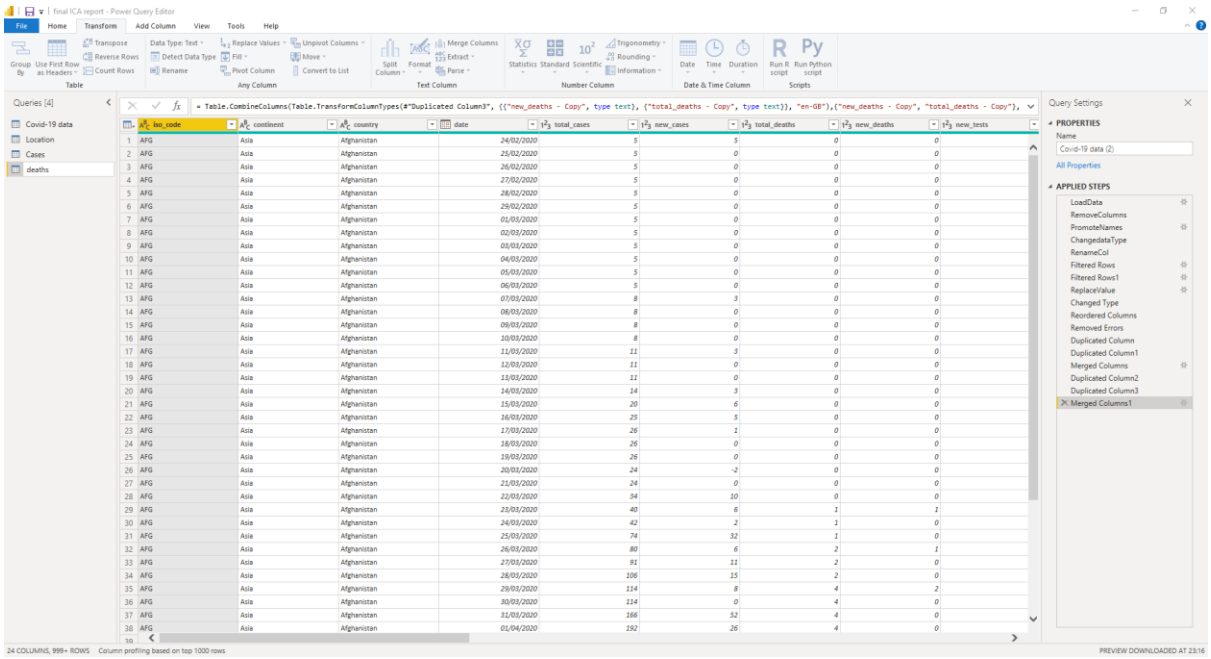


Figure 22 - creating and renaming deaths table

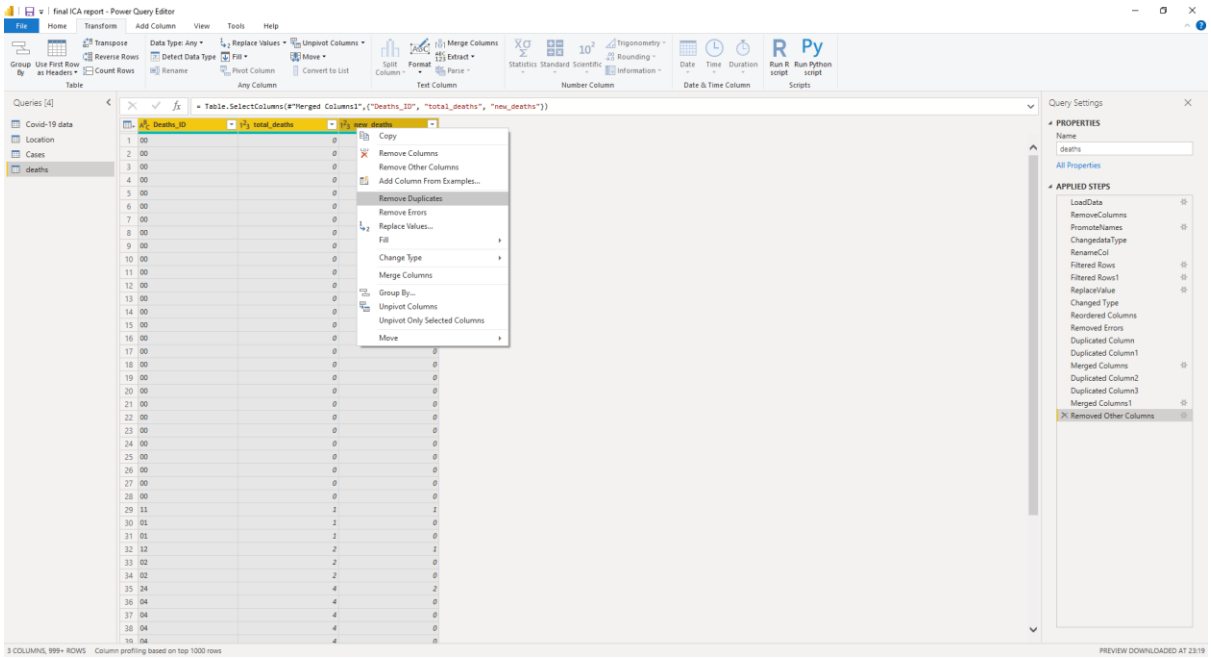


Figure 23 - removing duplicates from all columns

- Creating Tests Dimension table

Using the same procedure as before, I duplicated the 'new tests' and 'Total tests' columns in the Covid-19 data table, then merged the two tables and renamed it 'Tests ID', as shown in the diagram

below.

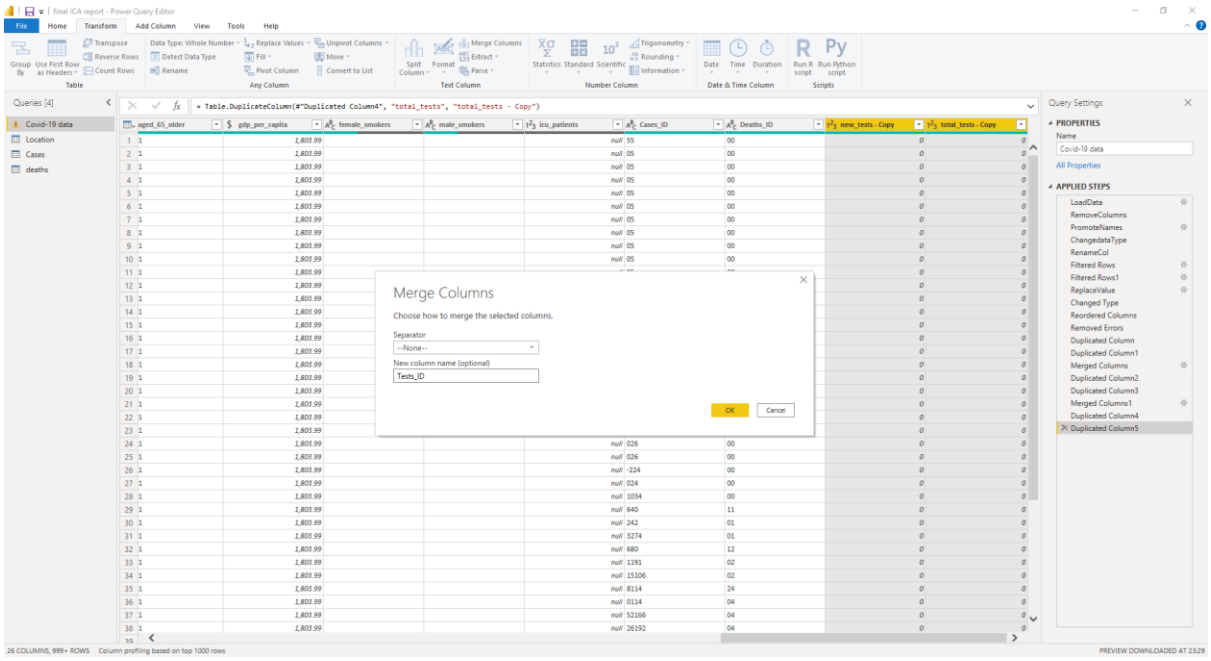


Figure 24 - Duplicating and merging Tests_ID column

Then I duplicated and renamed the Covid-19 data table to 'Tests.' I removed all columns except the 'new tests', 'total tests', and 'tests ID' columns from the newly created 'Tests' table. After that, I removed duplicates from all the test table's columns, as shown below.

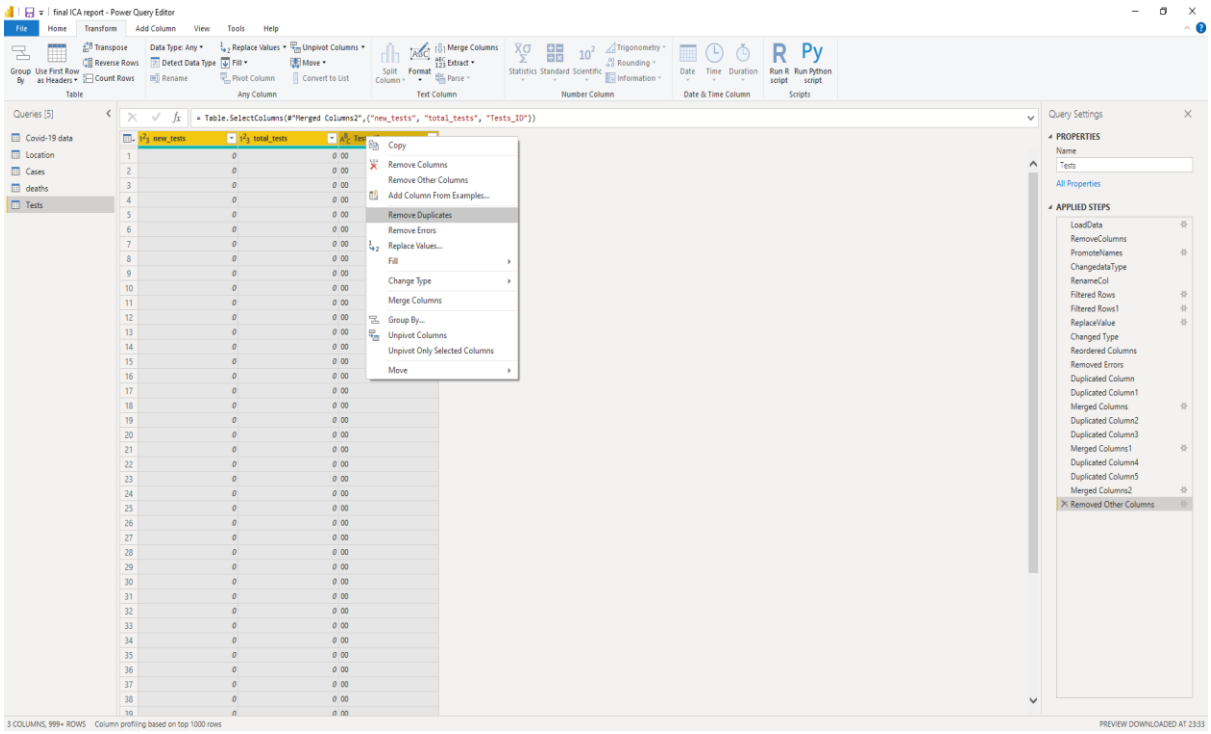


Figure 25 - removing duplicates from the Tests table

- Creating the Vaccinations Dimension table

As explained previously, I replicated all of the columns in the vaccination table that was required, including 'Total vaccinations,' 'people vaccinated,' 'people fully vaccinated,' 'total boosters,' and 'new vaccinations.' Then, as seen below, I merged the newly formed columns to produce the 'Vaccinations ID' column.

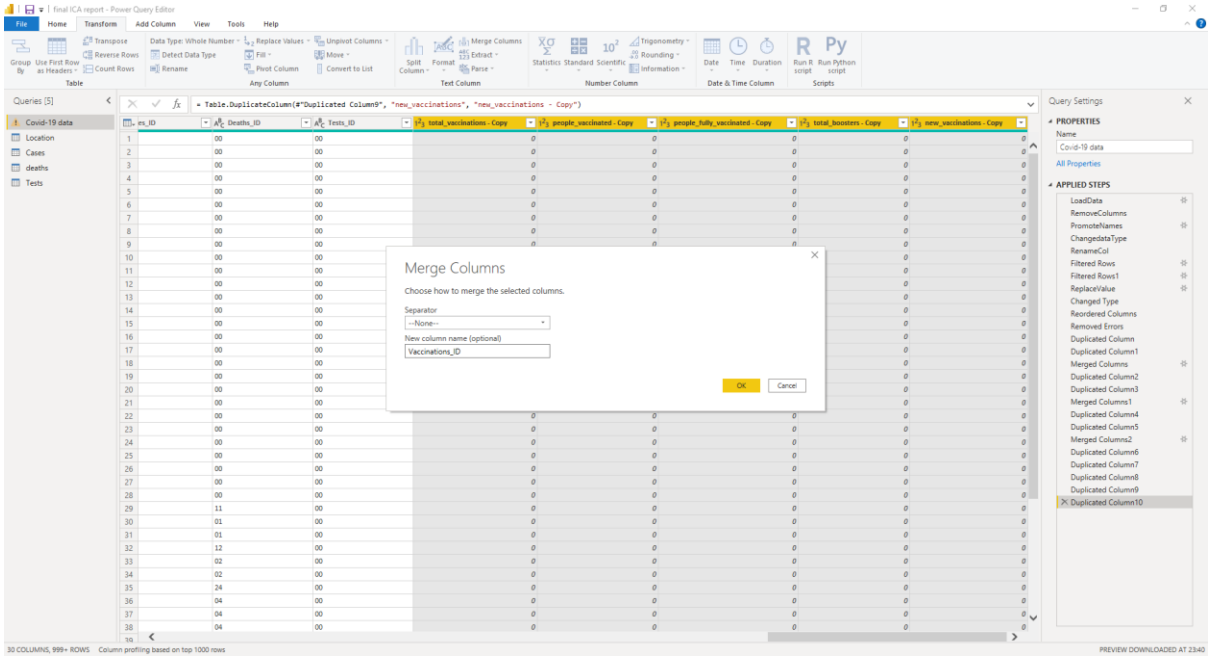


Figure 26 - Merging of columns to create Vaccinations_ID

Then I duplicated and renamed the covid-19 data table to 'Vaccinations.' I deleted all other columns in the new vaccinations table except the ones stated above, and then removed duplicates from all of the new columns in the Vaccination table.

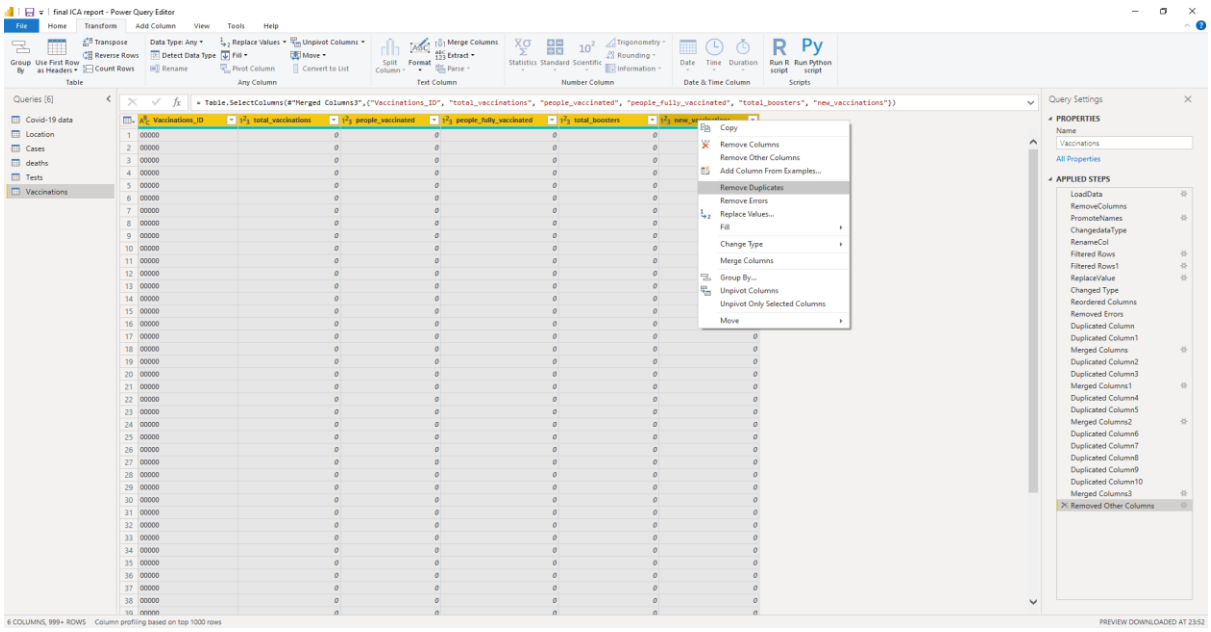


Figure 27 - Removing the duplicates in vaccinations

- Creating the Population Dimension table

To create this dimension table, I used the ISO_code column as a key so all I did was duplicate the Covid-19 data table and rename it as ‘Population’, then removed all the columns except the population column as shown below

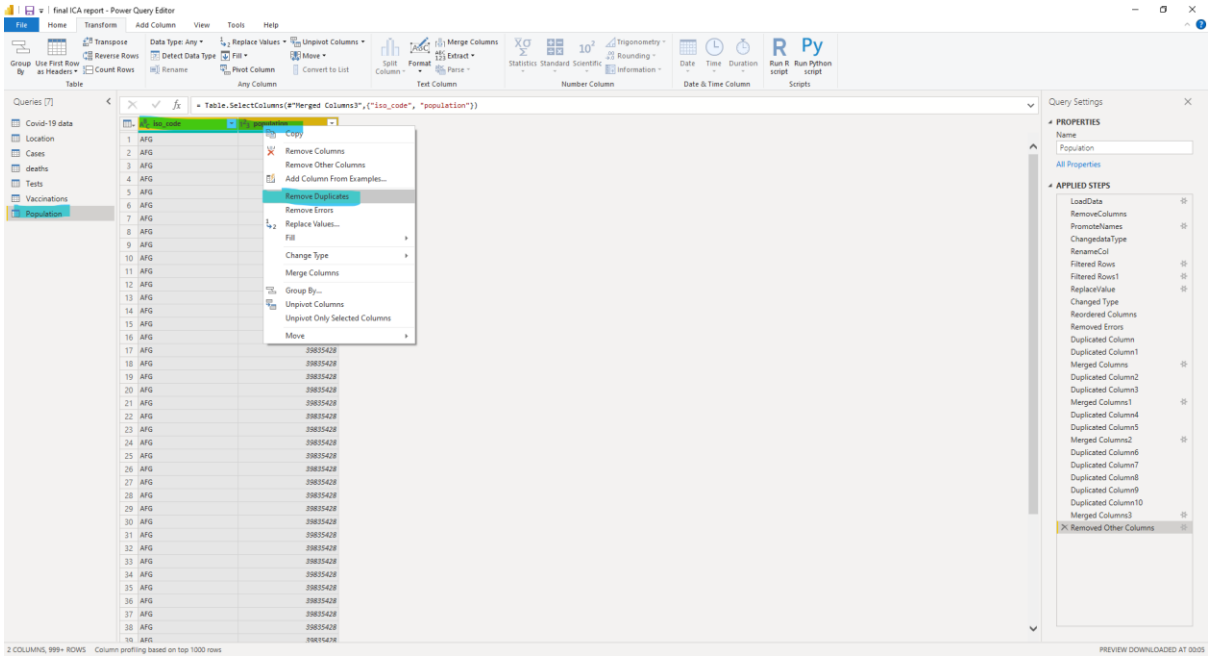


Figure 28 - creating Population details Dimension table

- Creating the population Details Dimension table

This dimension table was created to show the details of the population in the countries in the dataset, the data in these columns are not complete so I decided to group them together instead of grouping them with the population table. To create this table, I duplicated the covid-19 data table and renamed it population details, then removed all the columns in this new table except the ones we need which are the ‘median_age’, ‘aged_65_older’, ‘gdp_per_capita’ and the Iso_code columns. Then removed the duplicated data as shown below

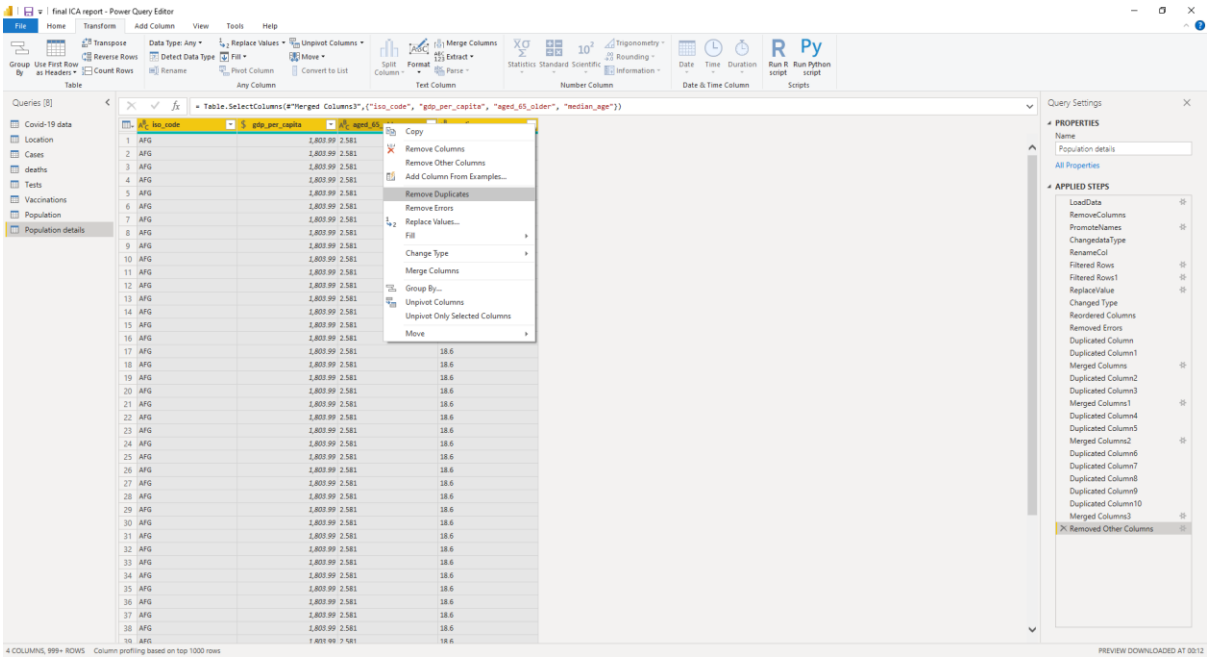


Figure 29 – creating the population details dimension table

After removing the duplicates, there are some empty values and null values in the data as shown below

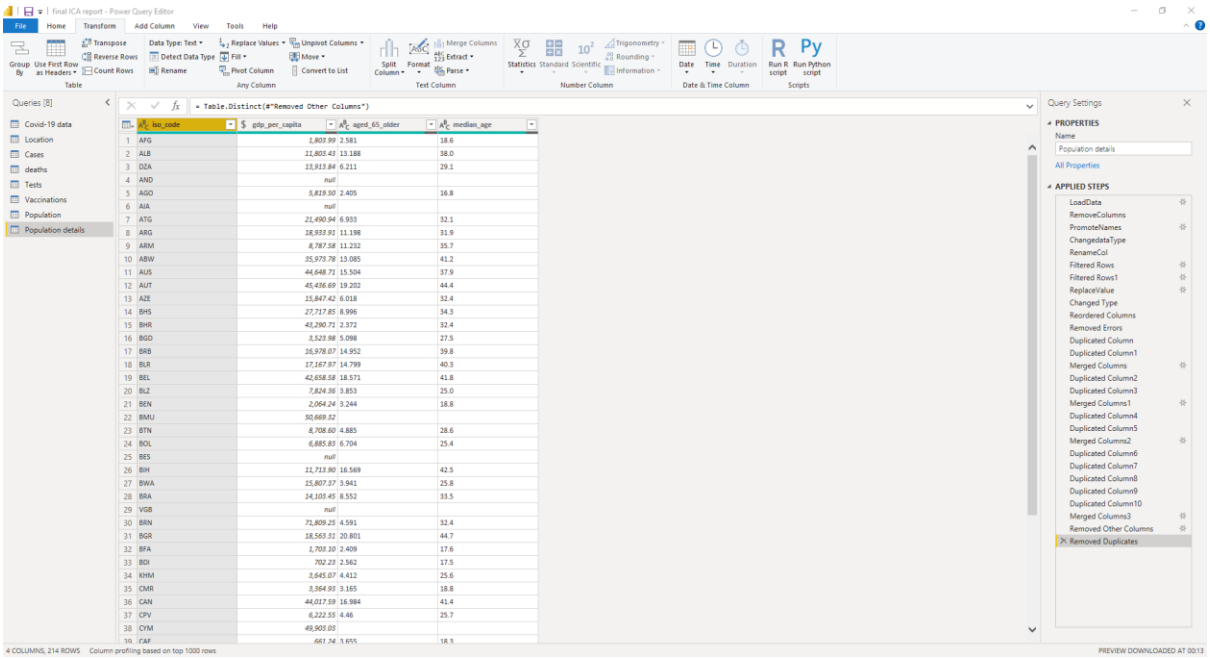


Figure 30 - Empty and Null values in table

To fix this, I removed the empty values in the ‘aged_65_older’ and the ‘gdp_per_capita’ columns as shown below

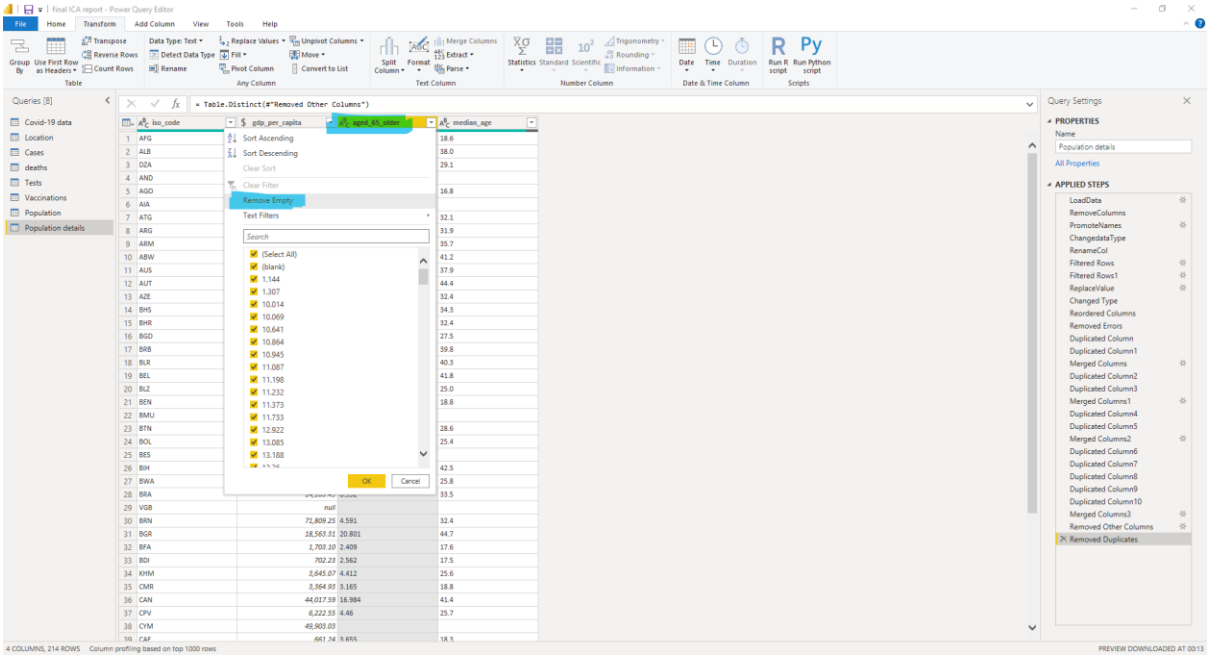


Figure 31 - Removing empty values

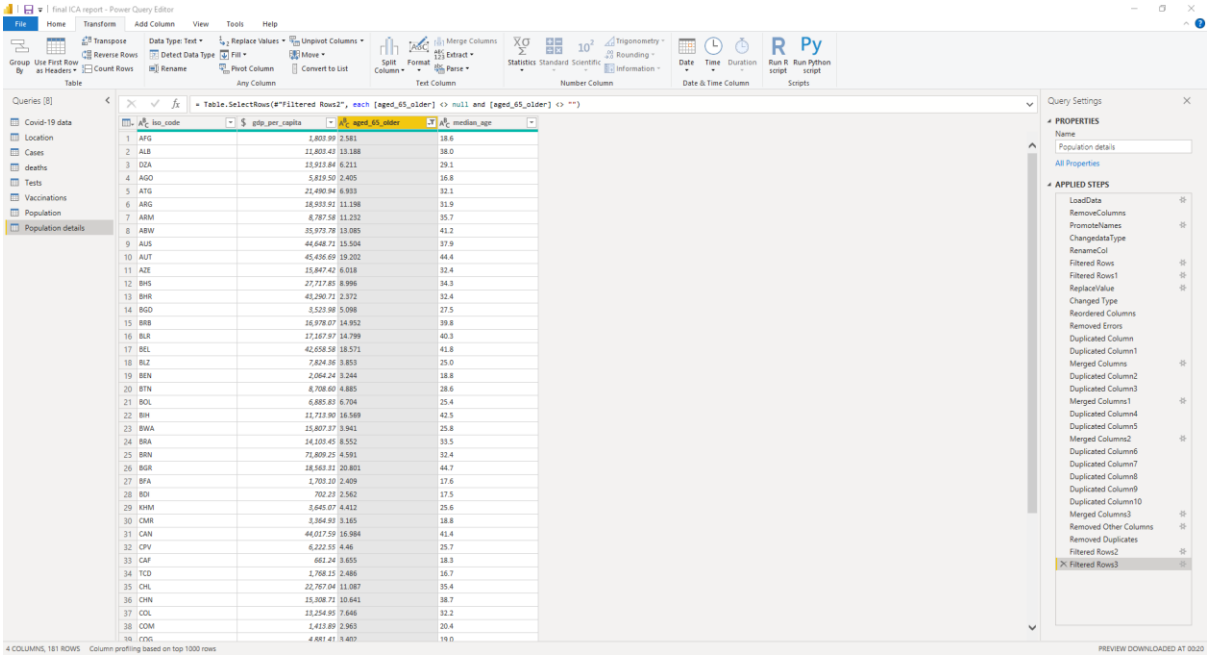


Figure 32 - Table after data cleaning

As seen in the figure above, the table is now free of errors and empty values.

- Creating Smoking detail dimension Table

This table was created to include the outlying data that doesn't have a lot of values. As a result, the visualisations that can be done with it will be limited. To make this table, I copied the Covid-19 data table and called it 'Other details', then deleted all but the 'iso code', 'male smokers', and 'female smokers' columns from the table, then removed duplicates from all the columns as shown below.

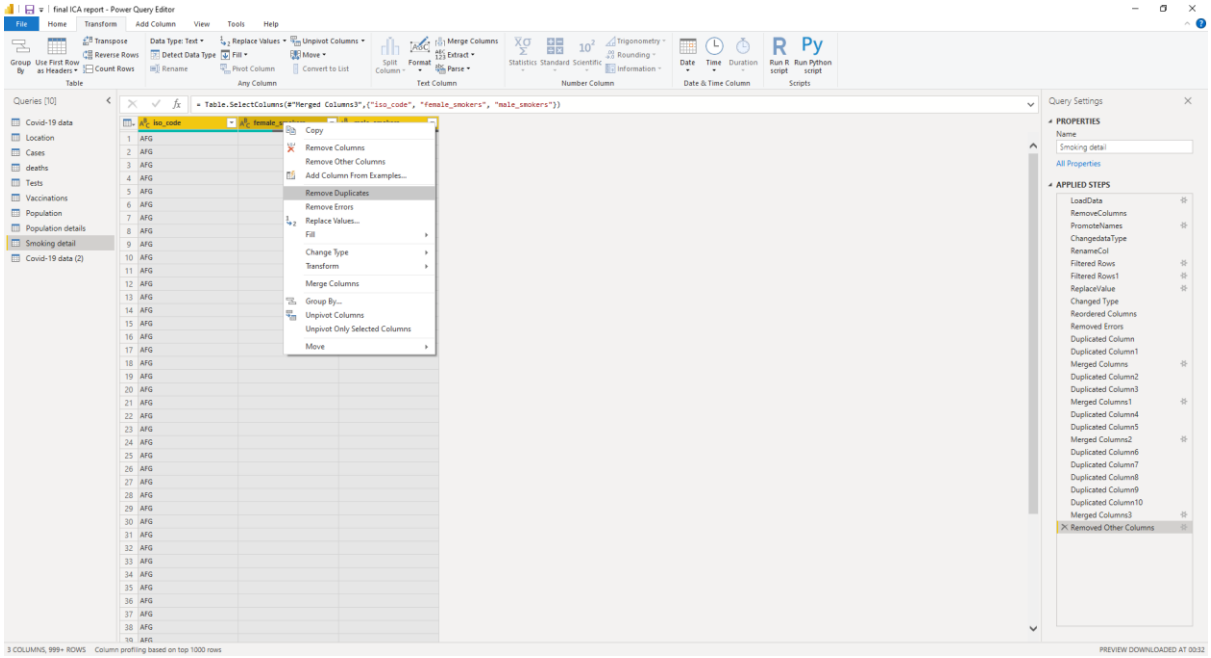


Figure 33 - Creating Smokers details table

After removing the duplicates, there were some empty rows so I removed the empty values from the female smokers column as shown below

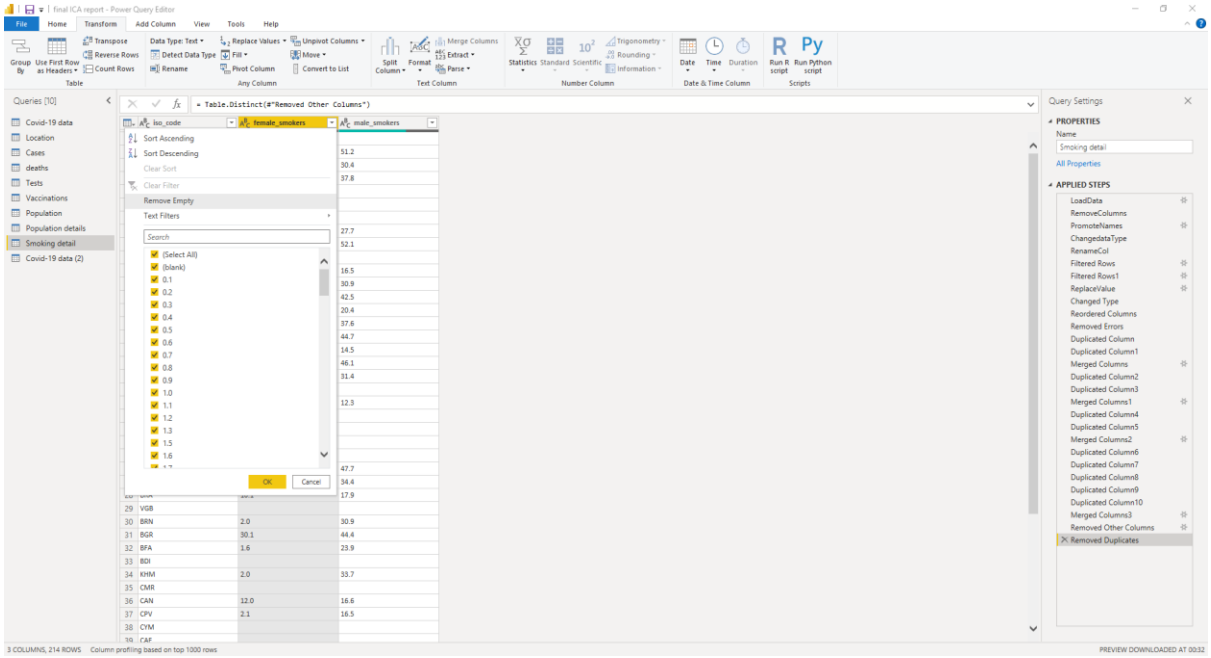


Figure 34 - removing empty rows

- Creating the ICU patients dimension table

To create this dimension table, I used the date column as the key column to link this table to the fact table, so the covid-19 data table was duplicated and renamed as ICU patients, then all the columns in the new table except the date and icu_patients table were removed. Then I removed duplicates from both tables as shown below.

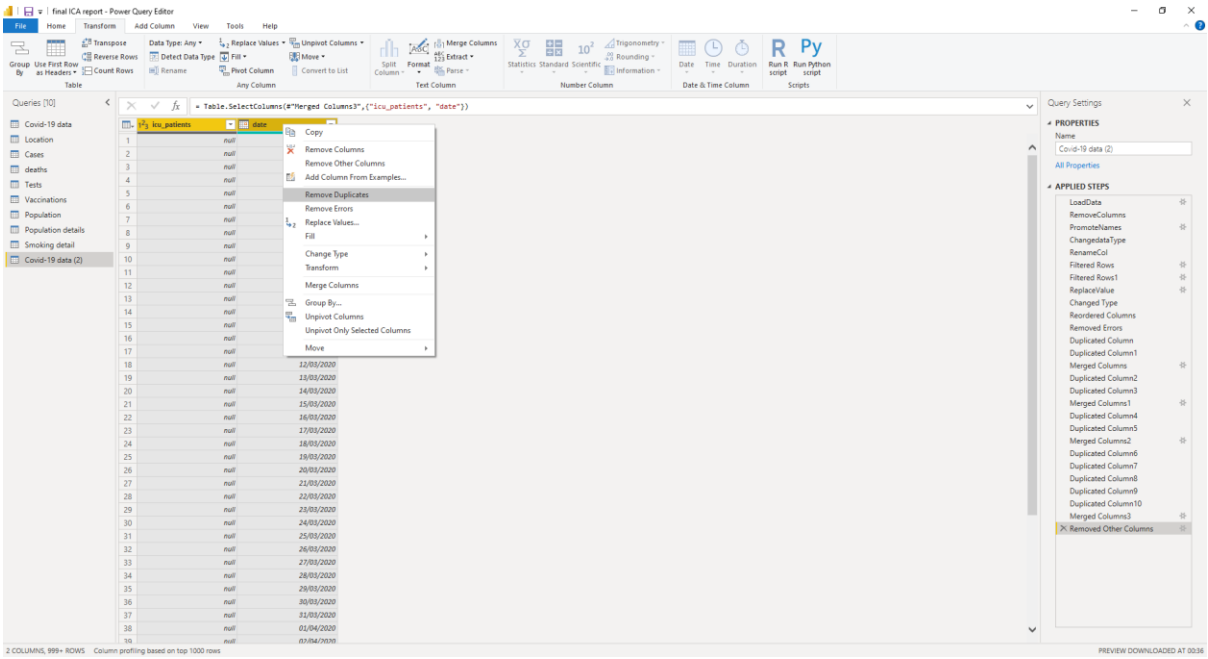


Figure 35 - ICU dimension table

After removing duplicates, there were some empty rows in the icu_patients column which were removed as shown below.

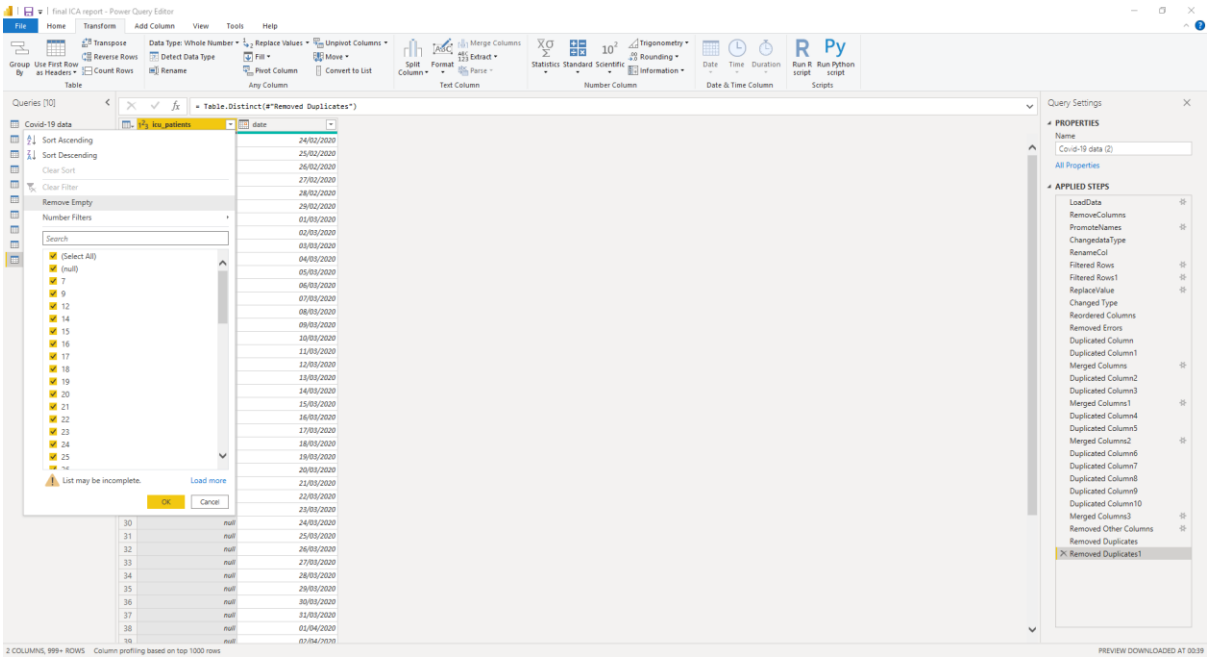


Figure 36 - removing empty rows

Now the table looks clean as shown below

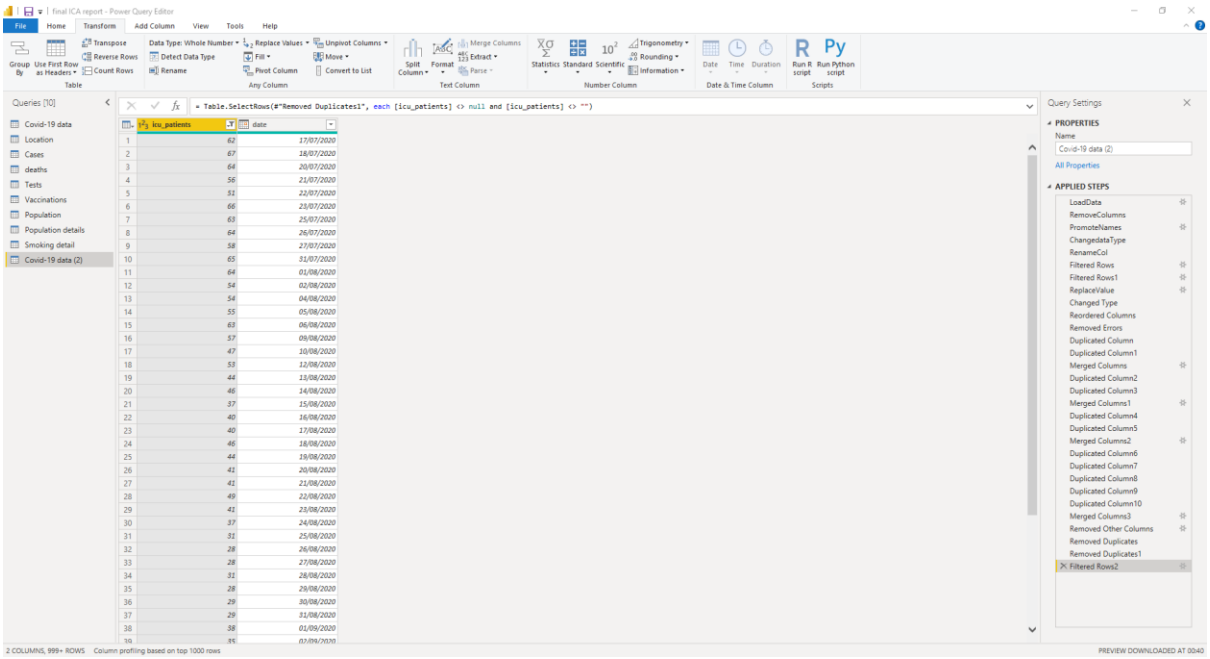


Figure 37 - ICU table after cleaning

7. Editing the fact table

Now that all the dimension tables have been created, I then needed to create the fact table which is now easy to do. In the covid-19 data, I just needed to leave only our primary key columns which are 'iso_code', 'date', 'cases_ID', 'Deaths_ID', 'Tests_ID' and 'Vaccinations_ID' and deleted all the other columns as shown below

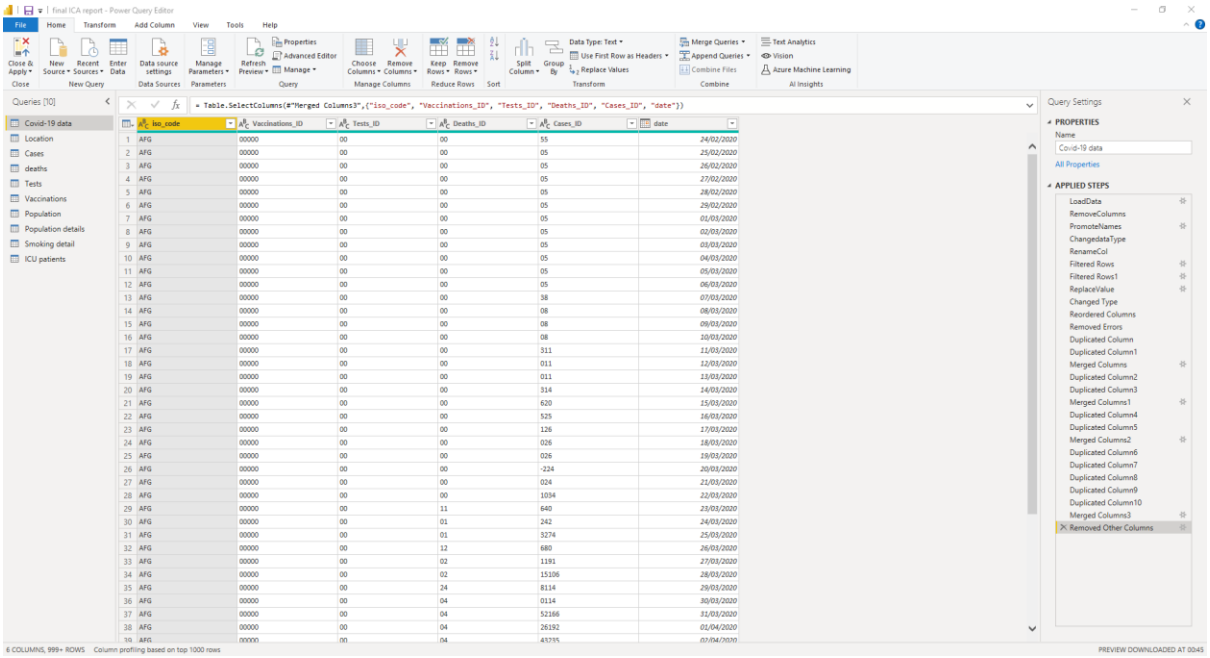


Figure 38 - Fact table after edit

So our fact table contains only the keys columns which means it is a factless fact table.

After this was done, then I clicked on the **Close & Apply** button to apply our changes and close the Power Query Editor as shown below

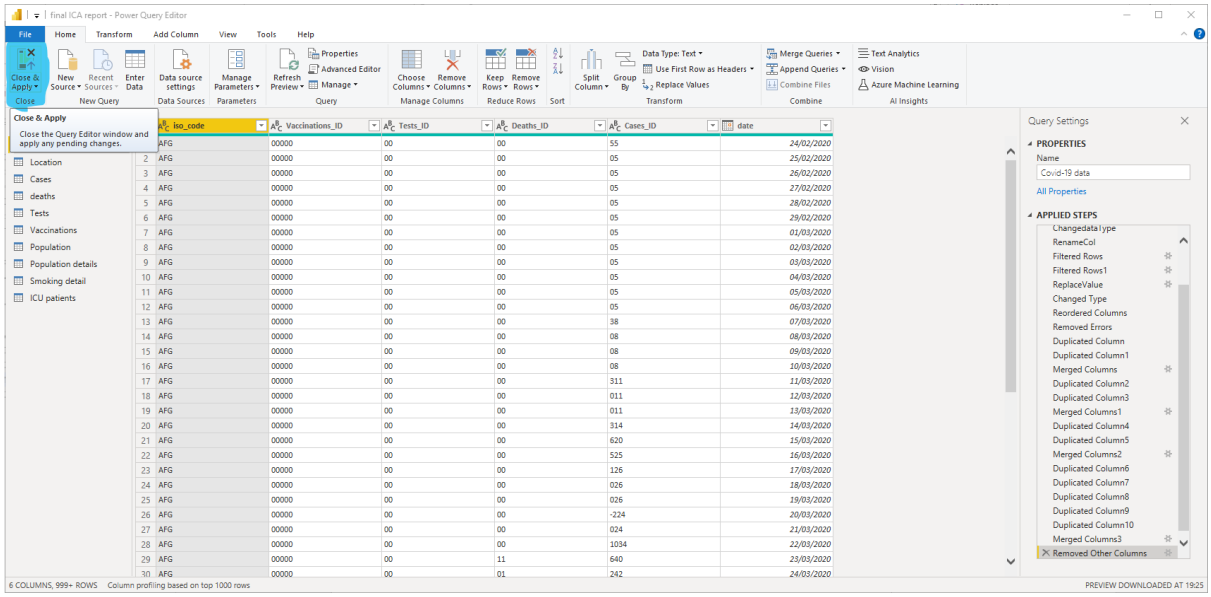


Figure 39 - close and Apply

8. Creating the Data models and relationships

After splitting the flat table into fact and dimension tables, the next step is connecting the data together to be able to visualize the data accurately which is called creating relationships. Power Bi provides a powerful feature called Power BI Data Modelling that can be used to perform this task. After loading and pre-processing the data, Power BI automatically generates relationships between the tables as shown below

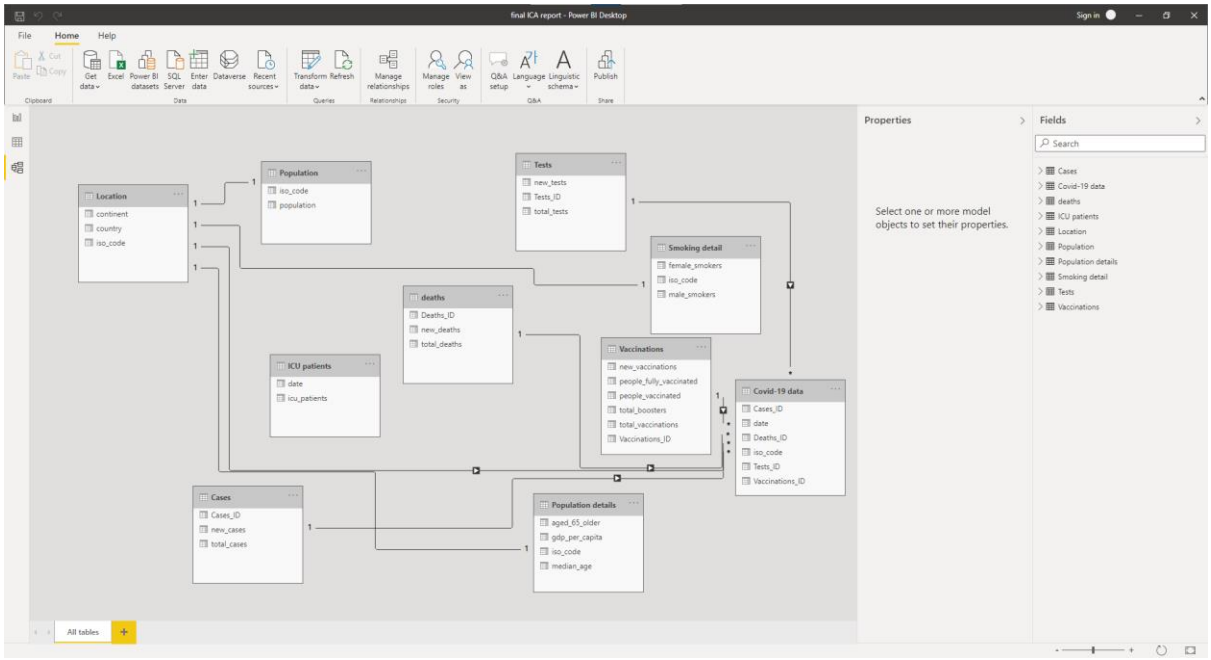


Figure 40 - Screenshot of Relationships generated by Power Bi

However, these are not the relationships we need, so these automatic relationships were deleted and new ones were created by first clicking on the 'Manage relationships' menu as shown below.

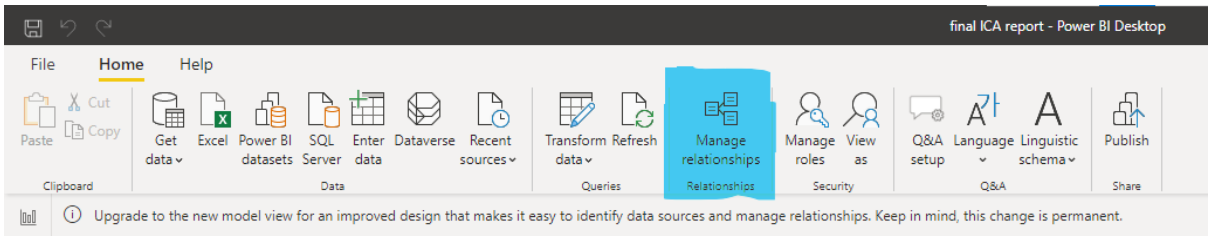


Figure 41 - Screenshot showing the Manage Relationships menu

This opened a dialog box which shows the already created relationships as shown below, then all the relationships were highlighted and the delete button shown below was clicked to delete all the relationships.

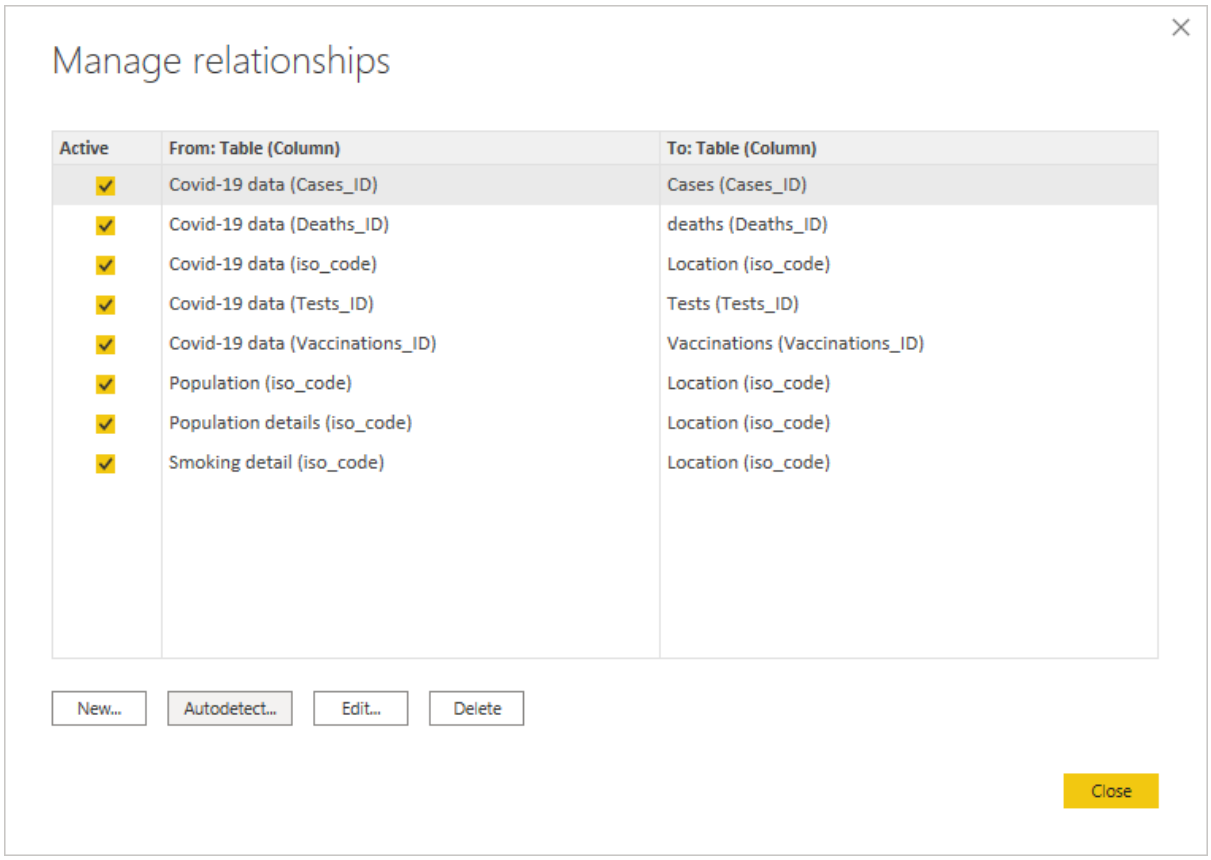


Figure 42 - Screenshot of the Manage relationships Dialog box

Then the 'New' button was selected to create new relationships, a 'Create relationship' dialog box popped up which allowed me to choose the specific columns in two different tables in which relationships are to be created and shows the cardinality to be used and also the cross filter direction which can be changed from single to both to make the relationship bi-directional. Then all the relationships needed were created and the close button was selected as shown below

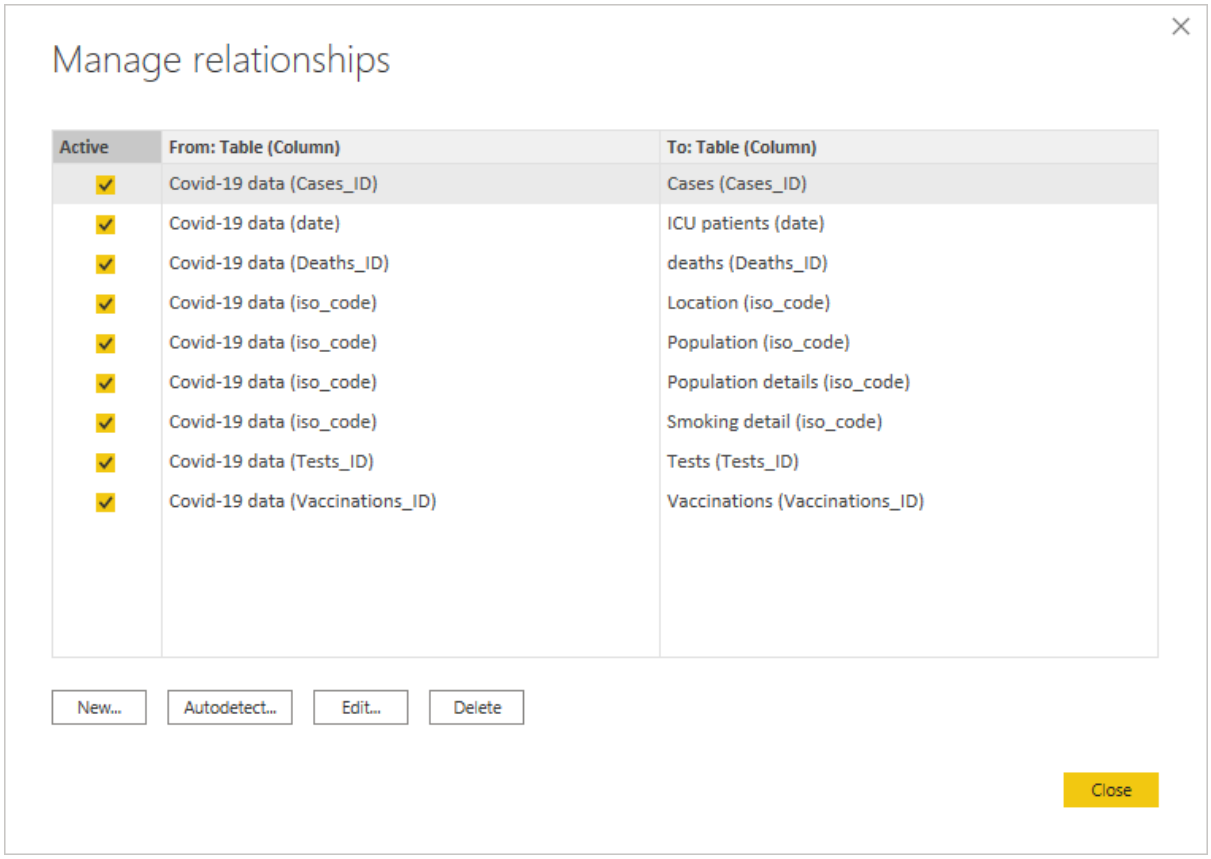


Figure 43 - Screenshot of the relationships

The tables were then arranged with the fact table (Covid-19 data) at the centre and all the dimension tables surrounding it as shown below

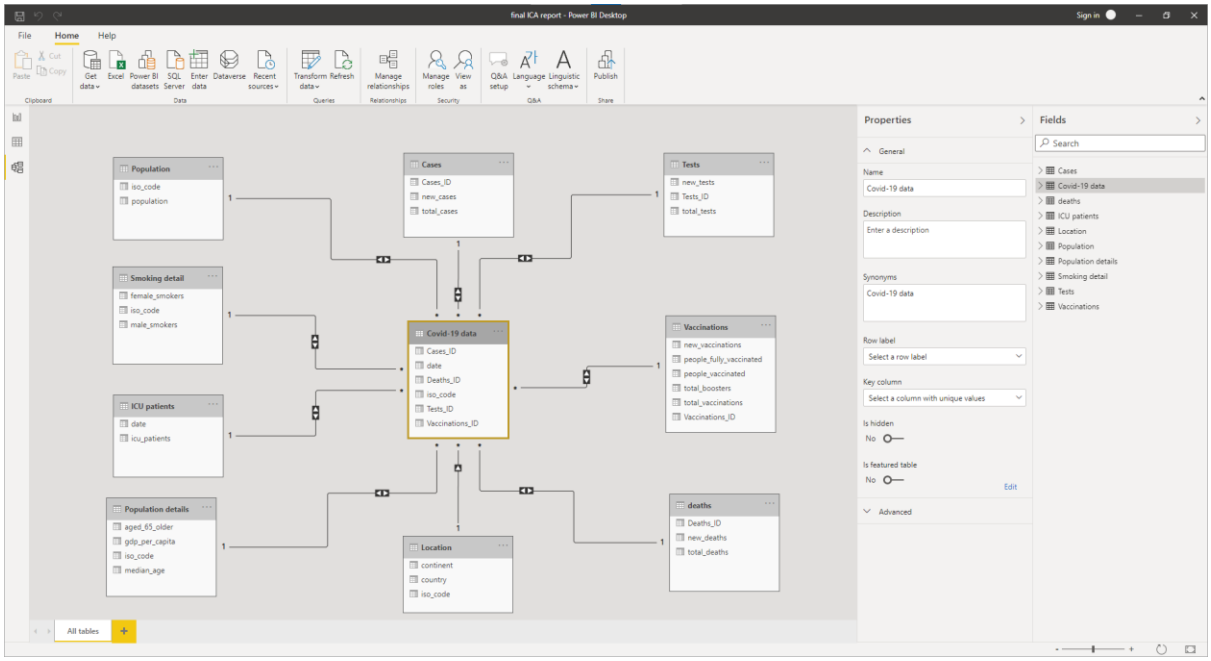


Figure 44 - Screenshot of the model

Most of the relationships had to be bi-directional so that the flow of data could be mutual between the related tables and to create data referencing across the connected forms.

EXECUTIVE SUMMARY

With the devasting impact that Covid-19 has had all over the world, it is Imperative that we try to understand the pandemic as much as possible and one of the best ways to do that is by analysing and visualising its data. This report helps to contextualize the effect of covid-19 in relation to deaths it has caused and to quantify the effects of the efforts to curb its spread, such as testing and vaccinations. The report shows the data gotten on the disease for all the continents and most of the countries in the world and it also

KEY FINDINGS

- Despite being a world Power, United states has not handled the virus well enough, They have had consistently high number of daily cases and deaths though their full vaccinations rate has been steadily rising.
- The rate of vaccination and testing in Africa and surprisingly Europe is quite low.
- The actual death rate compared the to the confirmed cases is low at 1.93% which shows that the advances in medicine has helped us reduce the fatality of the Virus.
- Oceania has one of the best vaccination rate as a continent and the lowest number of cases though that might be because it is a relatively small continent and it's a bit closed off to the world so the pandemic is easily controlled
- Gibraltar is the only country to have all its citizens fully vaccinated which made it the first country to achieve herd immunity.
- The median age of a country does not have much of an impact on the effects of Covid-19.
- The wealth of a country also doesn't offer much protection against the Virus, which means the Virus affected every country irrespective of its wealth or level of advancement.
- The Virus doesn't seem to lead to a rise in the number of ICU patients admitted in certain countries.

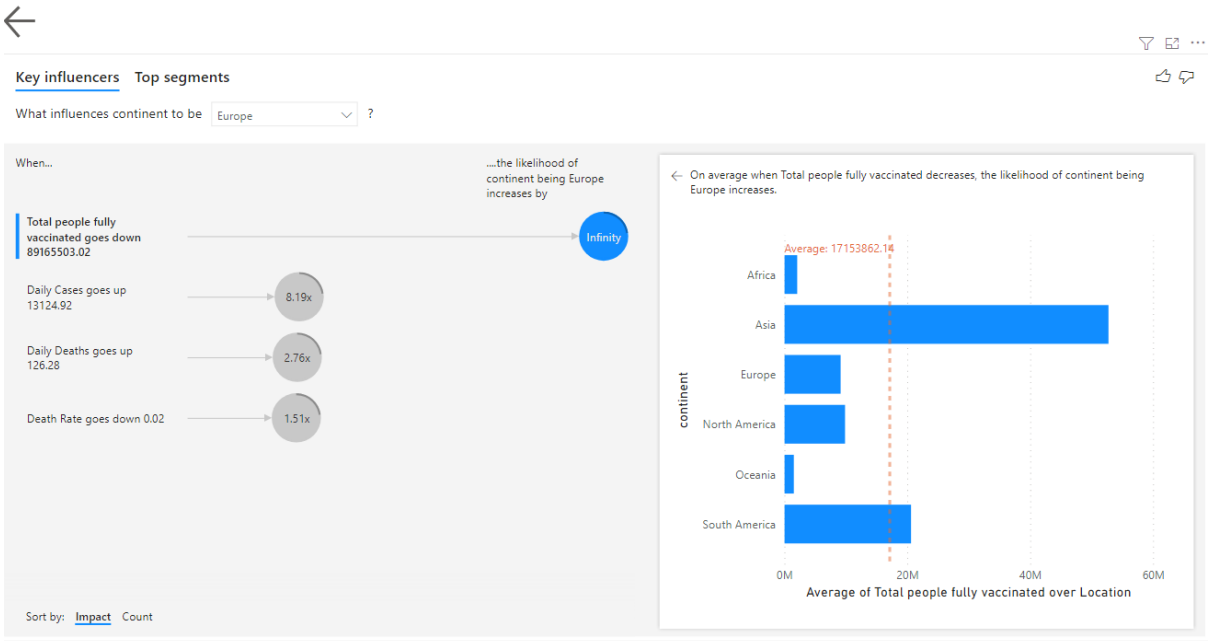


Figure 45 - showing key insights

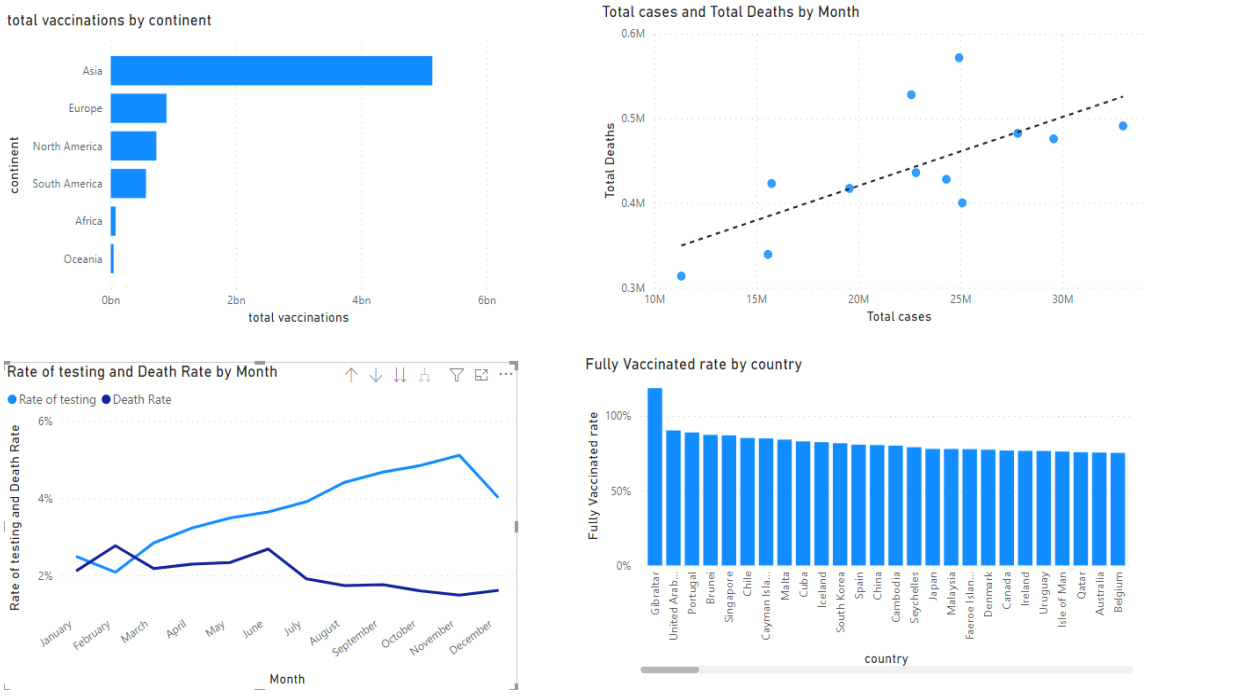


Figure 46 - key findings

RECOMMENDATIONS

- There needs to be an increase the rate of vaccinations provided in all the continents especially Africa and Asia as the number of people fully vaccinated is only 14.83% of the earth's population as at the time this data was collated.
- Testing and tracing has proven to an effective method of keeping track of the Virus but continents like Africa has very low testing rate, so this is something that needs to be improved.

INTRODUCTION

COVID-19 is a pathogenic virus caused by the SARS-CoV-2 virus that first appeared in Wuhan City, Hubei Province, in early December 2019. Based on phylogenetic analyses, bats appear to have been the COVID-19 viral reservoir, but the intermediate host has yet to be identified. Coronaviruses are mostly responsible for gastrointestinal and respiratory system illnesses, and they are disseminated through minute liquid particles from an infected person's mouth or nose when they cough, sneeze, speak, sing, or breathe. Large respiratory droplets to smaller aerosols are examples of these particles. Fever is one of the most common symptoms. Cough, fatigue, loss of taste or smell, sore throat, headache, aches and pains, diarrhoea, and other symptoms are common.

It is essential to use the data collated on the Virus to give important insights into how to stop its spread and reduce its effects on the world. This report serves to create visualizations for the some of the measures taken to combat the virus such as testing and vaccinations and to check how effective these measures have been.

This report is addressing the following questions.

1. What are the Overall values for confirmed cases, vaccinations, testing and deaths all around the world?
2. What are the total cases, Deaths, Tests and Vaccinations for each continent?
3. What are the rates of vaccinations, cases, deaths and tests since the beginning of the year?
4. What are the rates of daily cases and daily deaths?
5. In which continents and countries are the vaccination programme more advanced?
6. Which country has achieved or is closest to herd immunity?
7. What is the relationship between the confirmed cases and the deaths caused by covid-19?
8. The correlation between a Country's GDP and the death rate.
9. The Effects of Covid-19 on the number of ICU patients admitted.
10. Does the median age of a country affect the death rate of Covid-19 in that country?
11. What are the key influencers that affects each continent?

DATA MODEL

The dataset for this research was kindly provided by a GitHub user with the name 'baae386', who compiled it from multiple sources including the COVID-19 Data Repository at Johns Hopkins University (JHU), the European Centre for Disease Prevention and Control (ECDC), and others. The following is a link to the dataset:

https://github.com/owid/covid-19-data/tree/master/public/data

As discussed in section 1 above, the single flat file was normalised into Nine dimension tables and a single factless fact table. The data model and relationships were created using the tools available in power bi and explained in section.

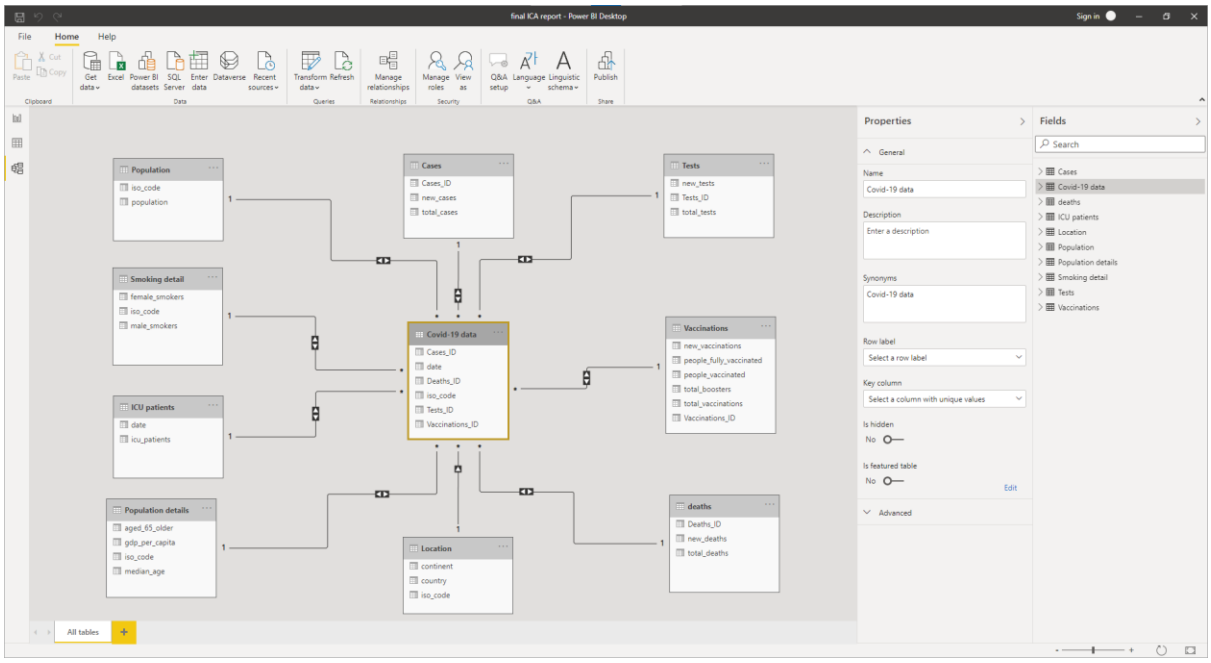


Figure 47 - THE DATA MODEL

FINDINGS BASED ON ANALYSIS AND EVALUATION

A Power BI report was produced to visualise and analyse various BI queries. Power BI visualisation tools were used to build a variety of graphs and tables.

1. What are the overall values for the confirmed cases, Vaccinations, Testing and deaths?

In order to visualize the total values for the testing, cases, deaths and vaccinations data so we can have a general idea of the scale of the pandemic, various Dax formulas were created to show those total values for the Confirmed cases, deaths, tests and the vaccinations.

To create these measures, the data page was clicked and then the new measure toolbar was selected to create different measures to aid in visualising our data as shown below

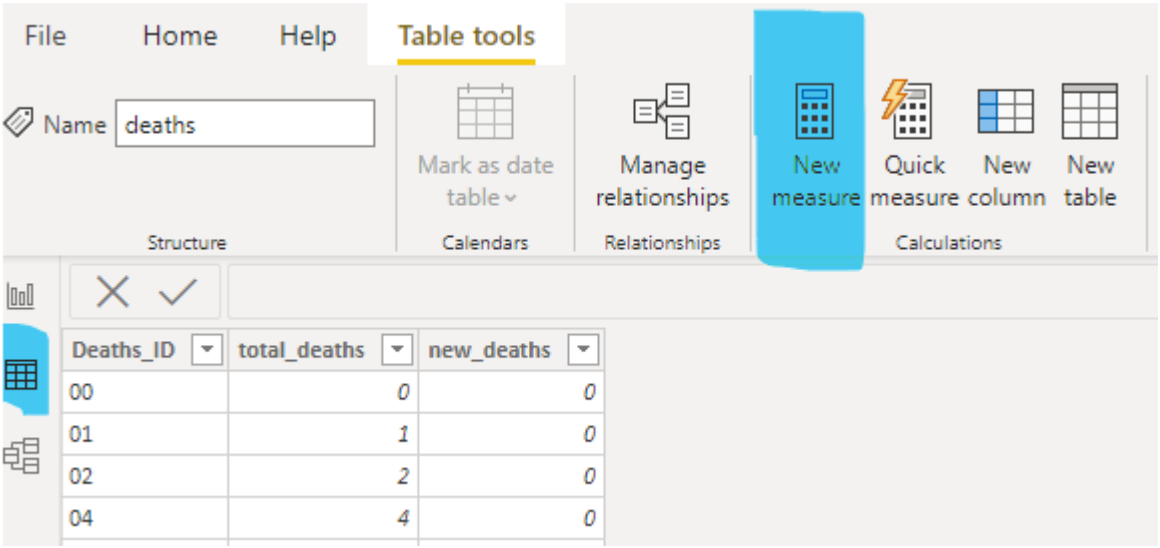


Figure 48 - Showing new measures creation

The formula to create the measure for total cases shown below is then inputted.

```
1 Total cases = SUM('Cases'[new_cases])
```

Figure 49 - DAX formula for total cases

The total vaccinations, deaths and tests were also created using this process as shown below.

NB – the data for the tests was not complete for all the countries due to the poor records kept by some countries so there are some slight discrepancies in its final figure.

```
1 Total Deaths = SUM(deaths[new_deaths])
```

Figure 50 - Total deaths

```
1 total vaccinations = sum(Vaccinations[new_vaccinations])
```

Figure 51 - total vaccinations

```
1 Total tests = SUM(Tests[new_tests])
```

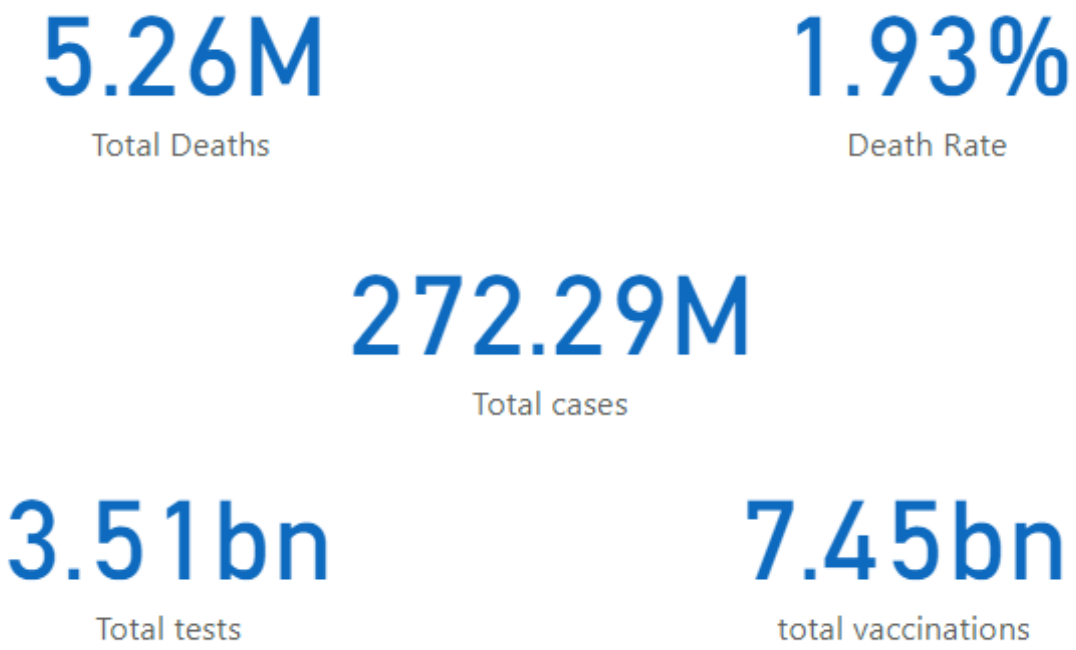
Figure 52 - total tests

Then I also wanted to show rate of deaths caused by covid based on the confirmed cases collected, which was created using the formula shown below

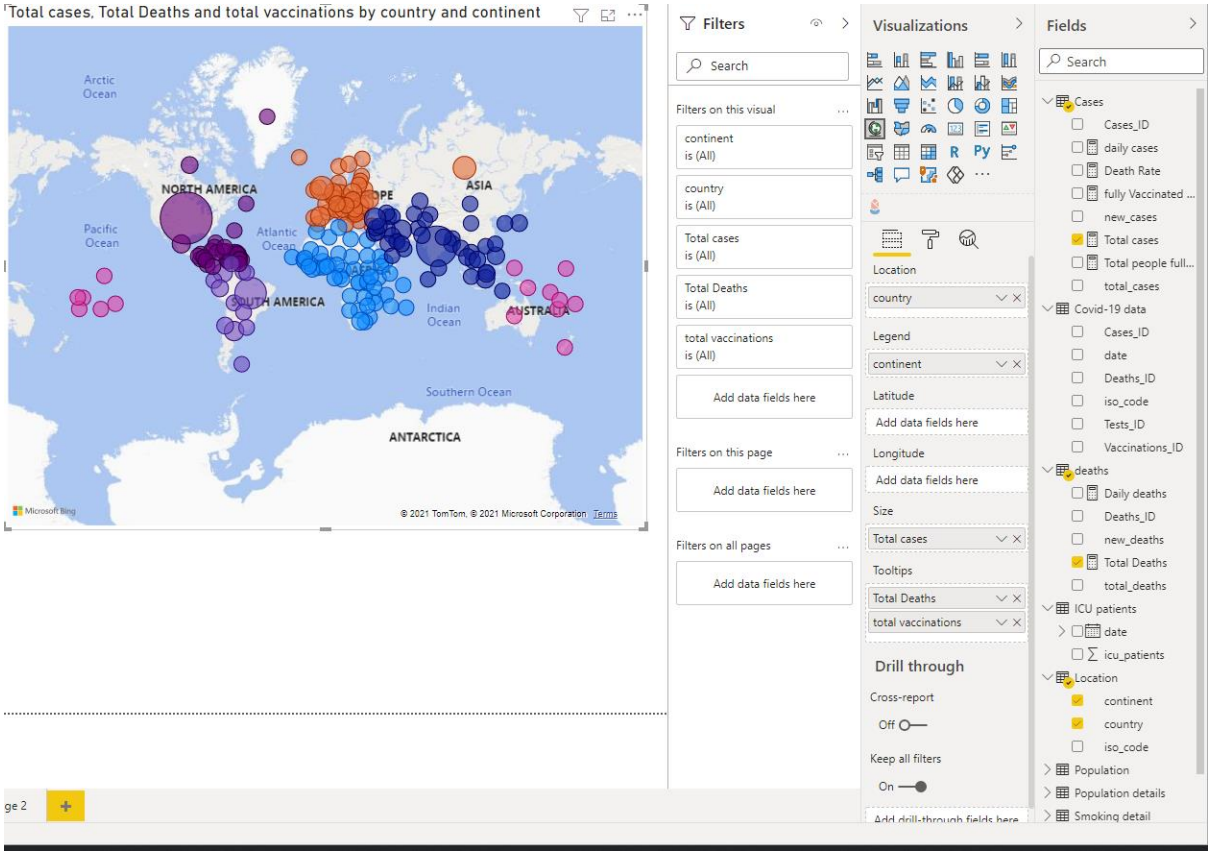
```
1 Death Rate = [Total Deaths] / [Total cases]
```

Figure 53 - death rate

Then using the Card visualization tool, the totals which shows the current figures for each metric as at 16th of December 2021 according to our dataset.

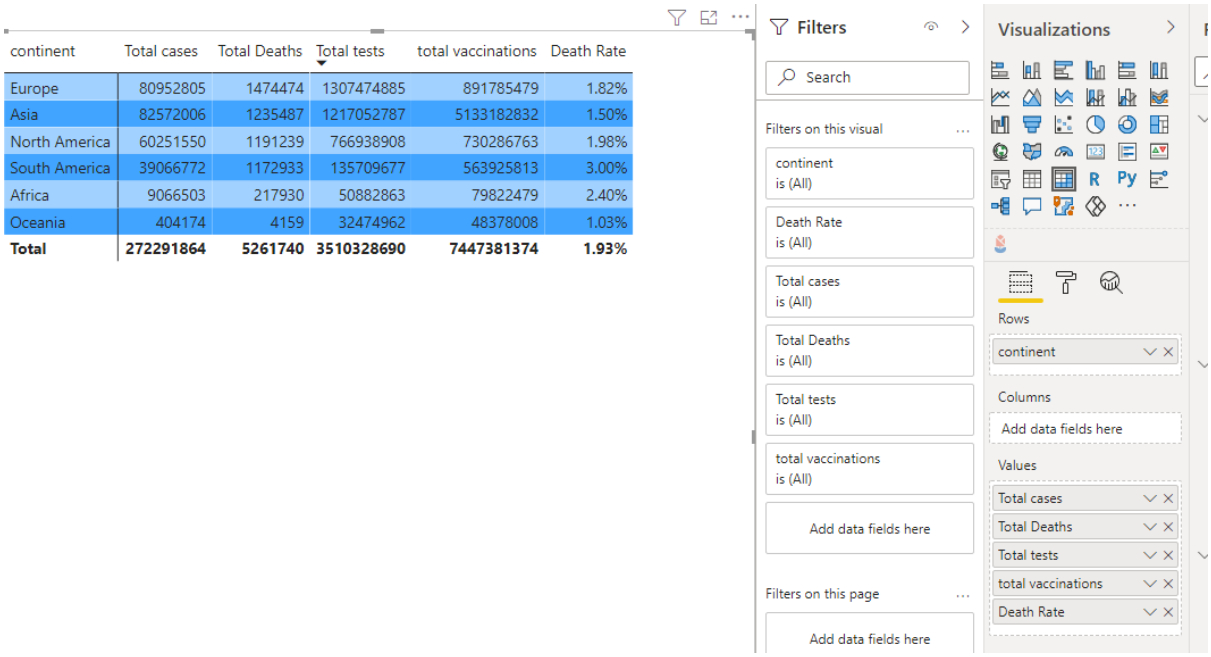


Map was then chosen to visualize the total Cases, total deaths and total vaccinations in each country. So, we can easily see the cases in each region based on the bubble sizes which after even a cursory glance, we can see that the United States has had the most number of confirmed cases which may be due to their advanced level of testing.



2. What are the Total cases, Deaths, Tests, Vaccinations and Death Rate for each continent?

A Matrix visualization was used to show this in a clean, clear and concise manner as shown below



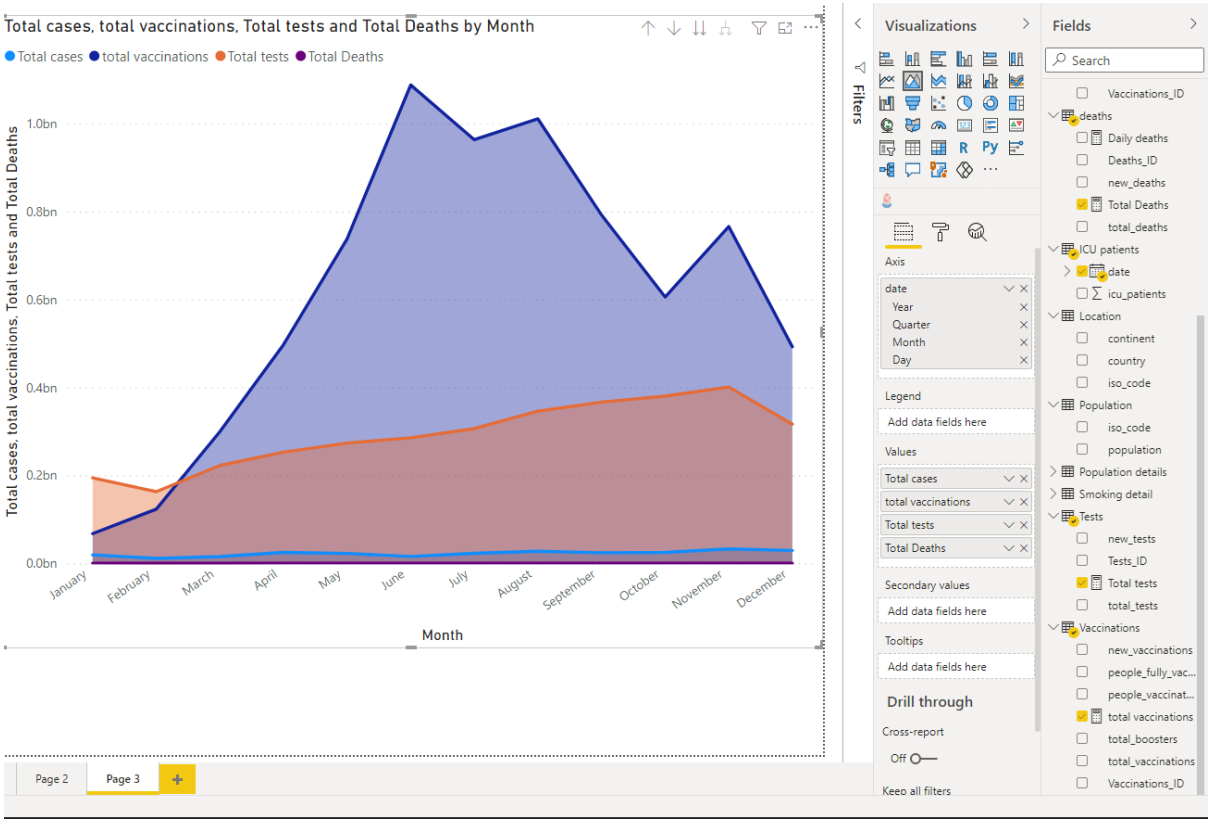
continent	Total cases	Total Deaths	Total tests	total vaccinations	Death Rate
Europe	80952805	1474474	1307474885	891785479	1.82%
Asia	82572006	1235487	1217052787	5133182832	1.50%
North America	60251550	1191239	766938908	730286763	1.98%
South America	39066772	1172933	135709677	563925813	3.00%
Africa	9066503	217930	50882863	79822479	2.40%
Oceania	404174	4159	32474962	48378008	1.03%
Total	**272291864**	**5261740**	**3510328690**	**7447381374**	**1.93%**

It is immediately obvious that Oceania has the lowest cases, deaths. Tests, vaccinations and death rate though that is probably due to its low population. The continent with the highest death rate was

south America followed by Africa which may be due to the general low level of medical equipments in both continents.

3. What are the rates of vaccinations, cases, deaths and tests since beginning of this year?

Area chart was chosen to visualize the rate of vaccinations, cases, deaths and tests since this year began and the visualization shows that that vaccinations given peaked in June before decreasing while the testing rate has been slowly but steadily increasing before slightly decreasing in November, however the death was too low to be shown properly on the chart and it has been relatively level.



4. What are the total cases in each continent and each country?

I wanted to visualize the total case and monthly cases of each country in each continent and the best visualization for that was the decomposition tree. Then I sorted the table by the total cases to make it easy to see the highest countries and months with the most cases as shown below.

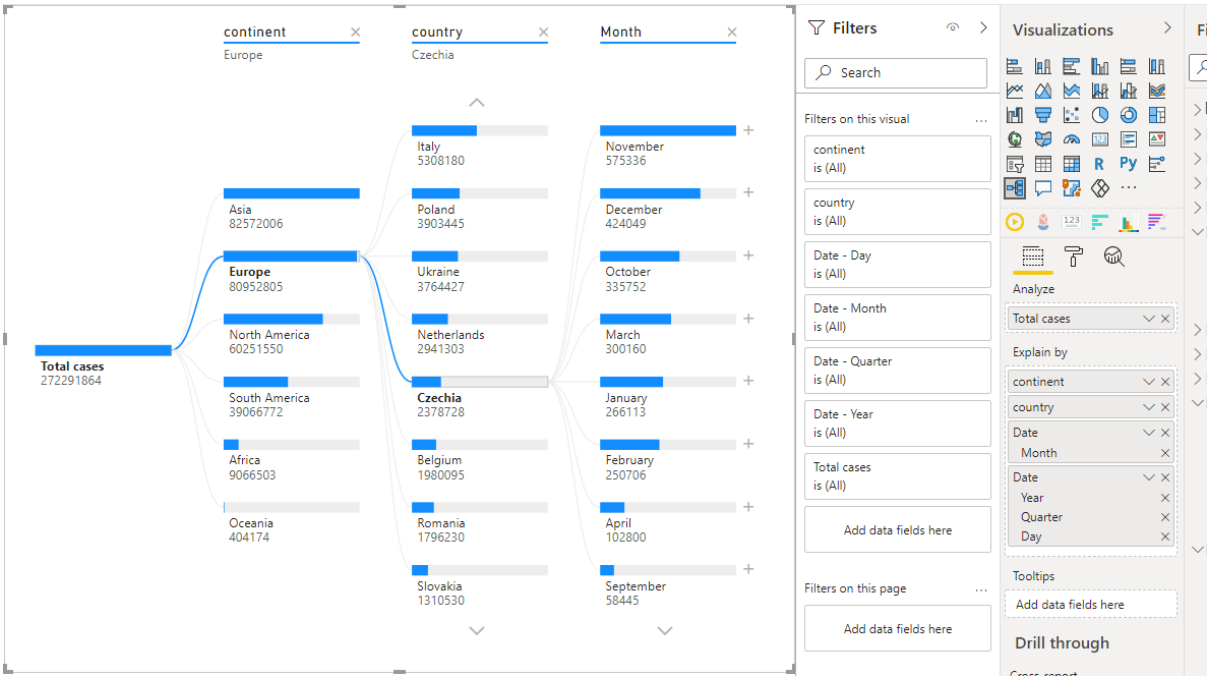


Figure 54 - Decomposition tree for Cases Analysis

5. What is the rate of cases and deaths daily?

To visualize the cases and deaths per day for each country, a new date table had to be created to create a contiguous date deselection, to do that the data model page was clicked and then the 'New table' toolbar. Then the DAX formula to create a calendar table was inputted as shown below

```
1 Calendar = CALENDAR ( MIN ( 'Covid-19 data'[date]), MAX ( 'Covid-19 data'[date] )
2 )
```

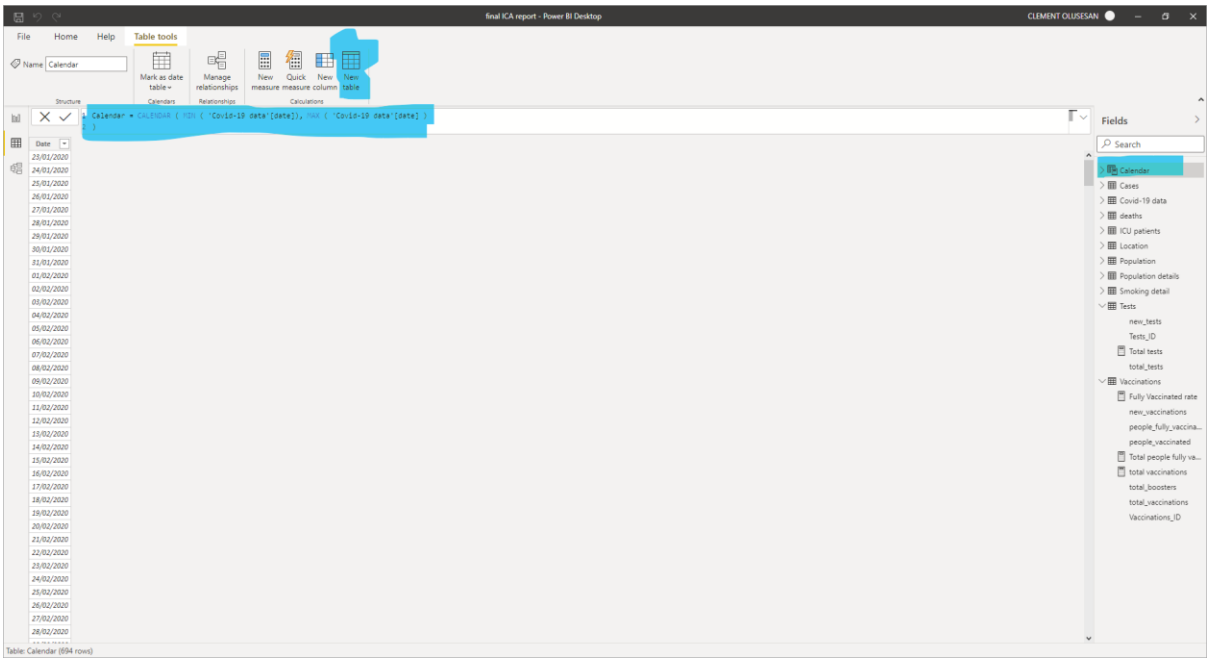


Figure 55 - Creating the Calendar table

After that was done, a relationship had to be created between the new calendar table and the fact table, therefore the 'Model' tab was clicked and then the 'manage relationship' toolbar was selected, then the 'New…' button was selected to create the relationship shown below

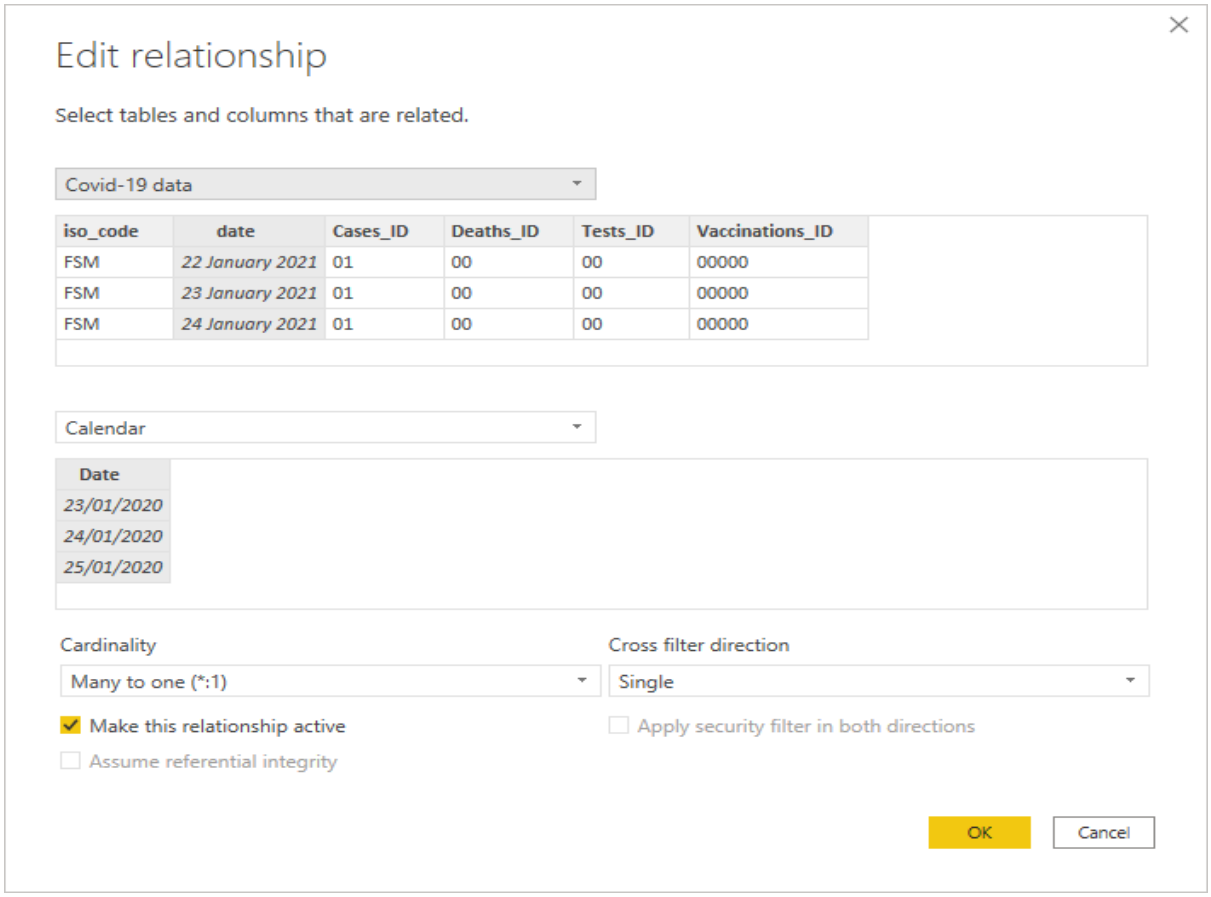


Figure 56 - Creating the relationship between the calendar and fact table.

Now we had the calendar table, the daily cases and deaths measures were created using the DAX formula shown below

```
1  Daily cases =
2  VAR current_day = [Total cases]
3  VAR prev_day =
4  CALCULATE (
5      [Total cases],
6      DATEADD( 'Calendar'[Date], -1, DAY )
7  )
8  RETURN
9  IF ( OR ( ISBLANK ( prev_day ), ISBLANK ( current_day ) ), BLANK(), current_day - prev_day )
```

Figure 57 - Daily Cases Dax formula

```
Daily deaths =
VAR current_day = [Total Deaths]
VAR prev_day =
CALCULATE (
    [Total Deaths],
    DATEADD( 'Calendar'[Date], -1, DAY )
)
RETURN
IF ( ISBLANK ( prev_day ), BLANK() , current_day - prev_day )
```

Figure 58 - Daily Deaths DAX formula

Gauge was chosen to visualize the daily cases and daily cases with a slicer for date and country created to toggle between the dates and the visualisation was formatted to show the data against a maximum of one million cases as shown below.

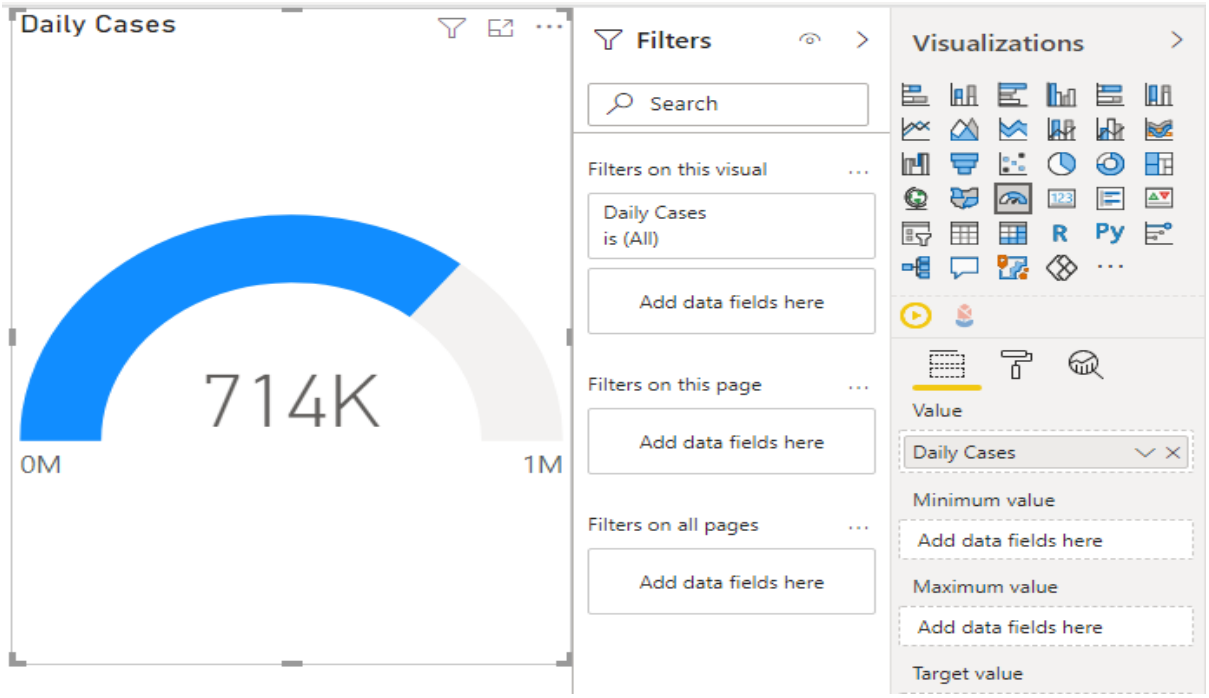


Figure 59 - Daily cases

The same was done to visualize the daily deaths for all the countries and continents and the Gauge was formatted to show the data against a maximum of Twenty thousand cases as shown below.

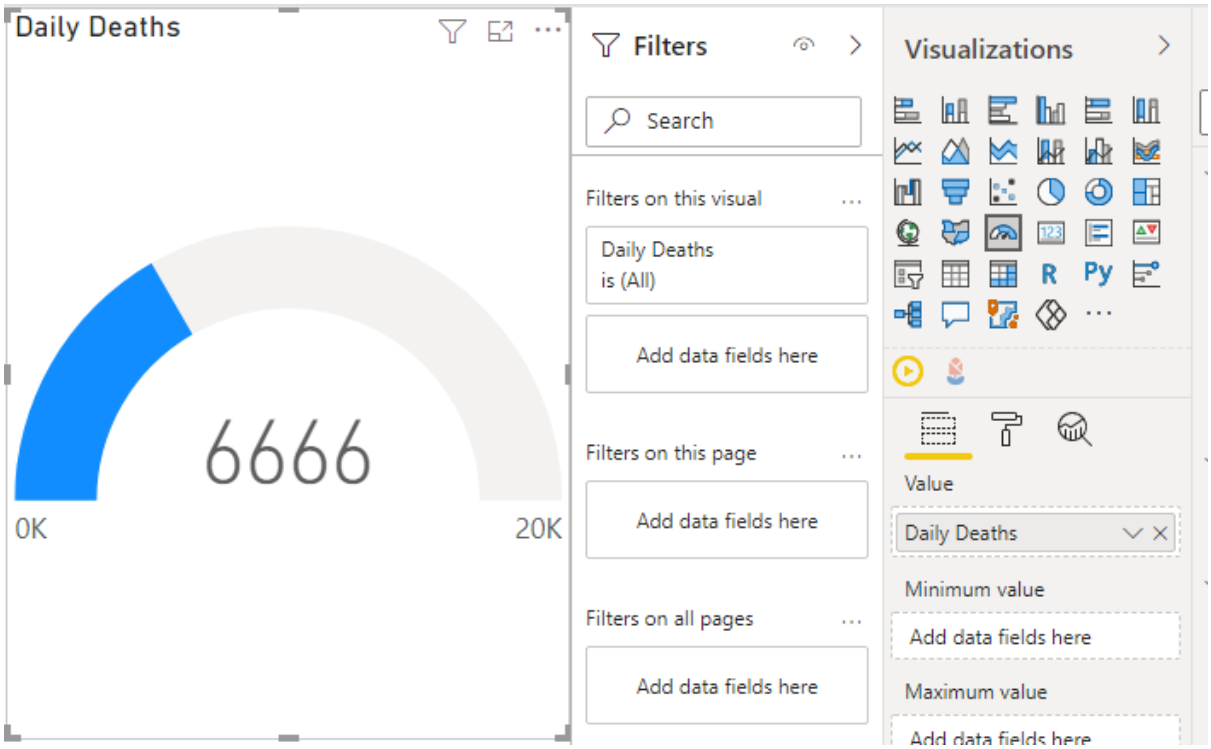
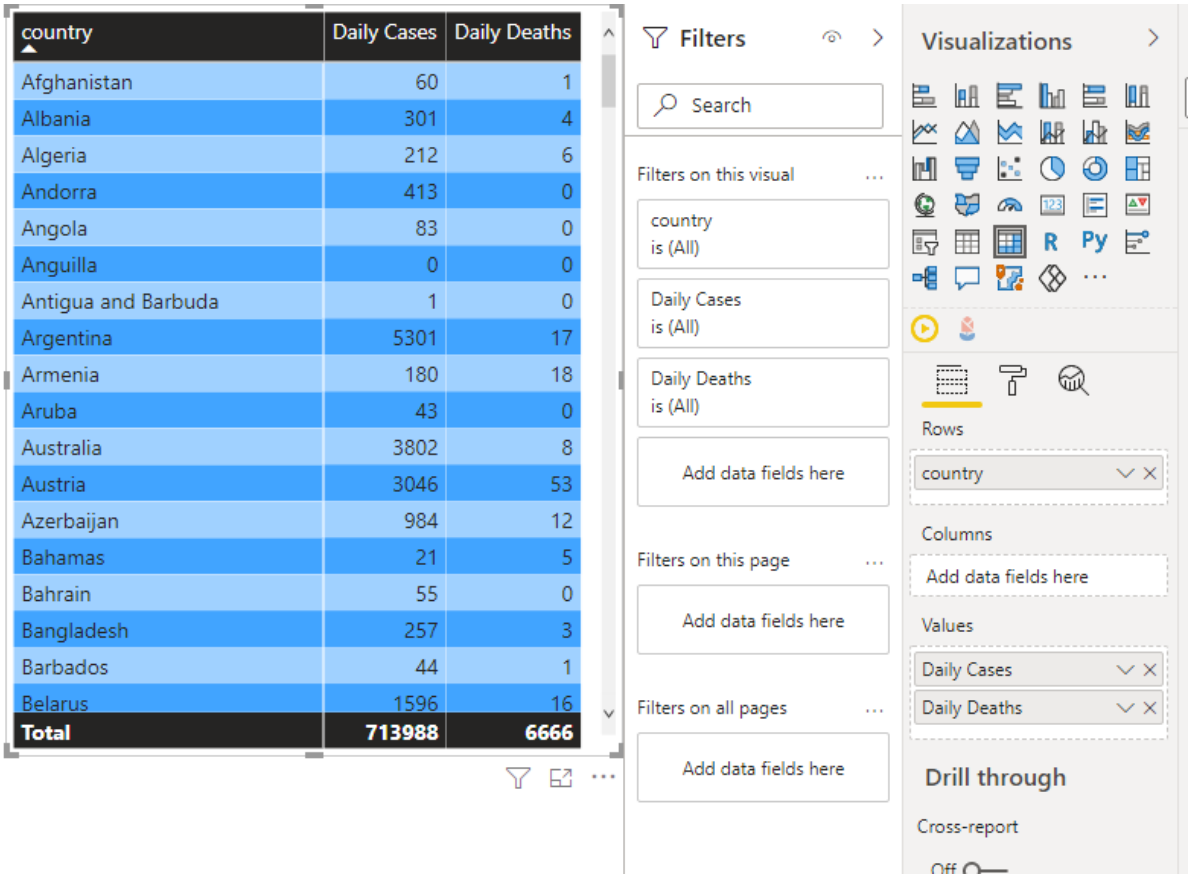


Figure 60 - Daily deaths

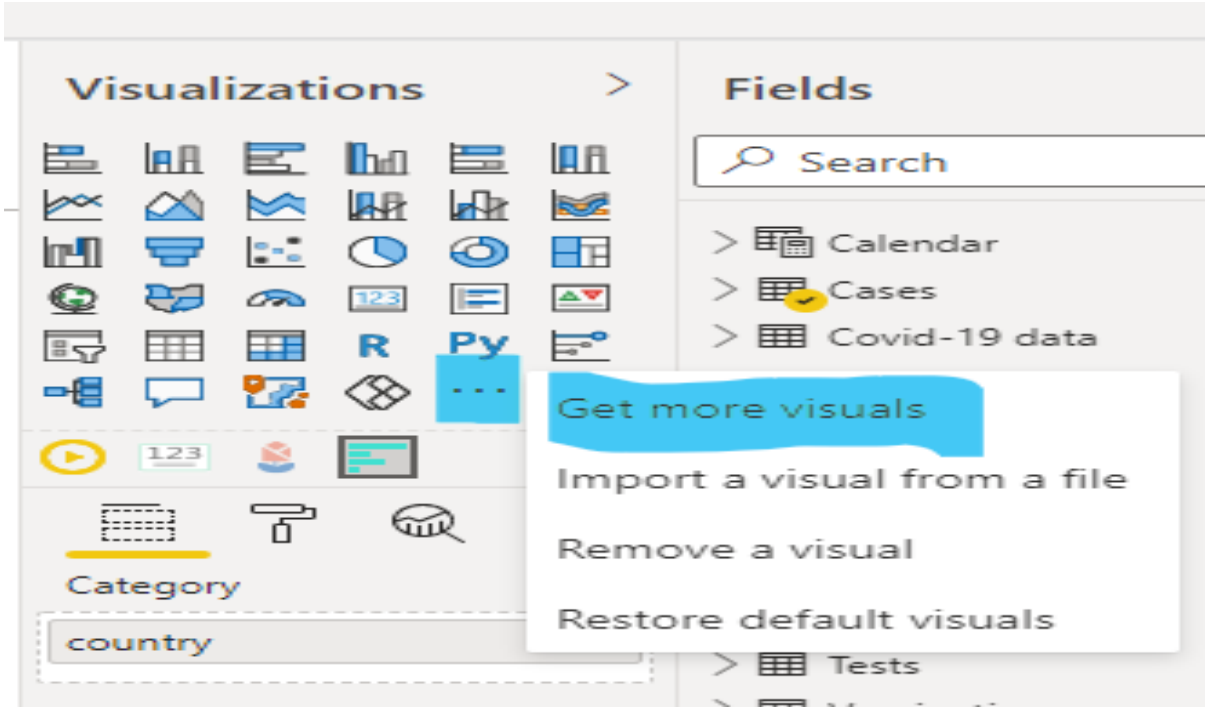
Then Matrix was chosen to visualize the Daily cases and daily deaths for each country in tabular form so it can be seen easily as shown below

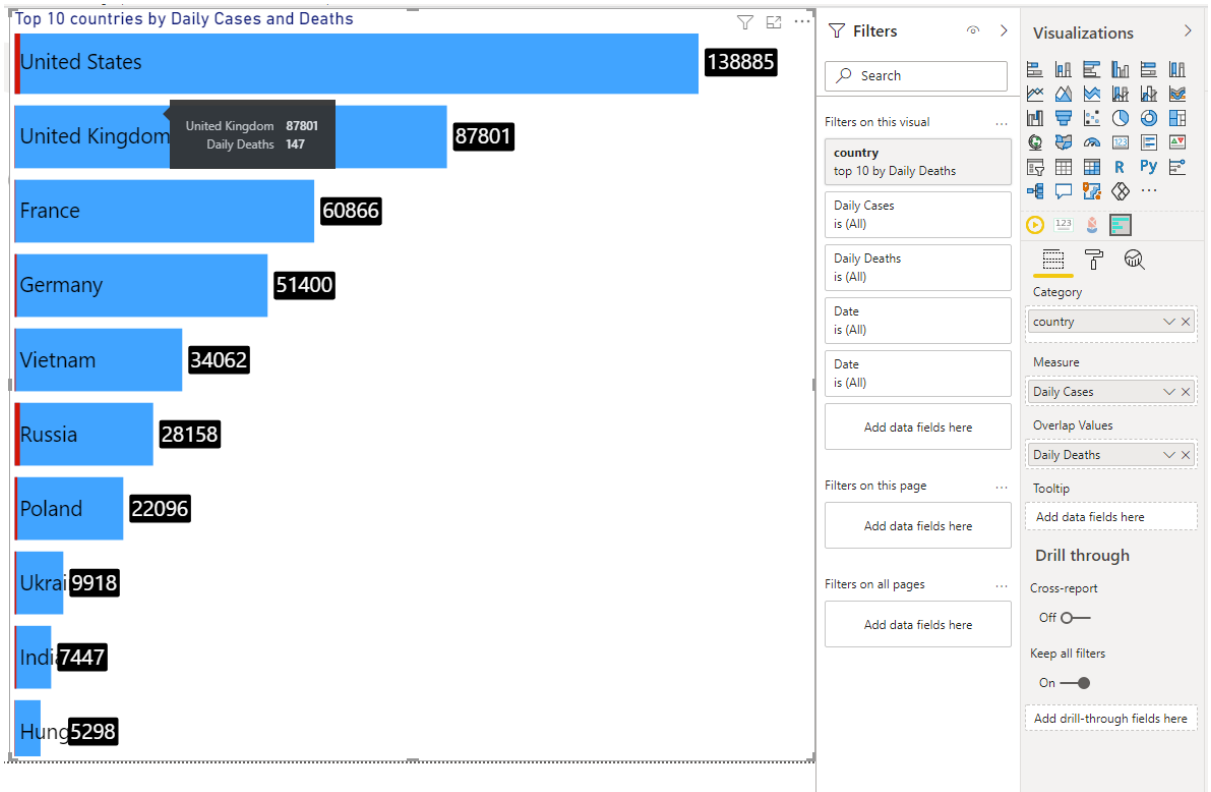


country	Daily Cases	Daily Deaths
Afghanistan	60	1
Albania	301	4
Algeria	212	6
Andorra	413	0
Angola	83	0
Anguilla	0	0
Antigua and Barbuda	1	0
Argentina	5301	17
Armenia	180	18
Aruba	43	0
Australia	3802	8
Austria	3046	53
Azerbaijan	984	12
Bahamas	21	5
Bahrain	55	0
Bangladesh	257	3
Barbados	44	1
Belarus	1596	16
Total	**713988**	**6666**

Figure 61 - daily cases and deaths in Matrix

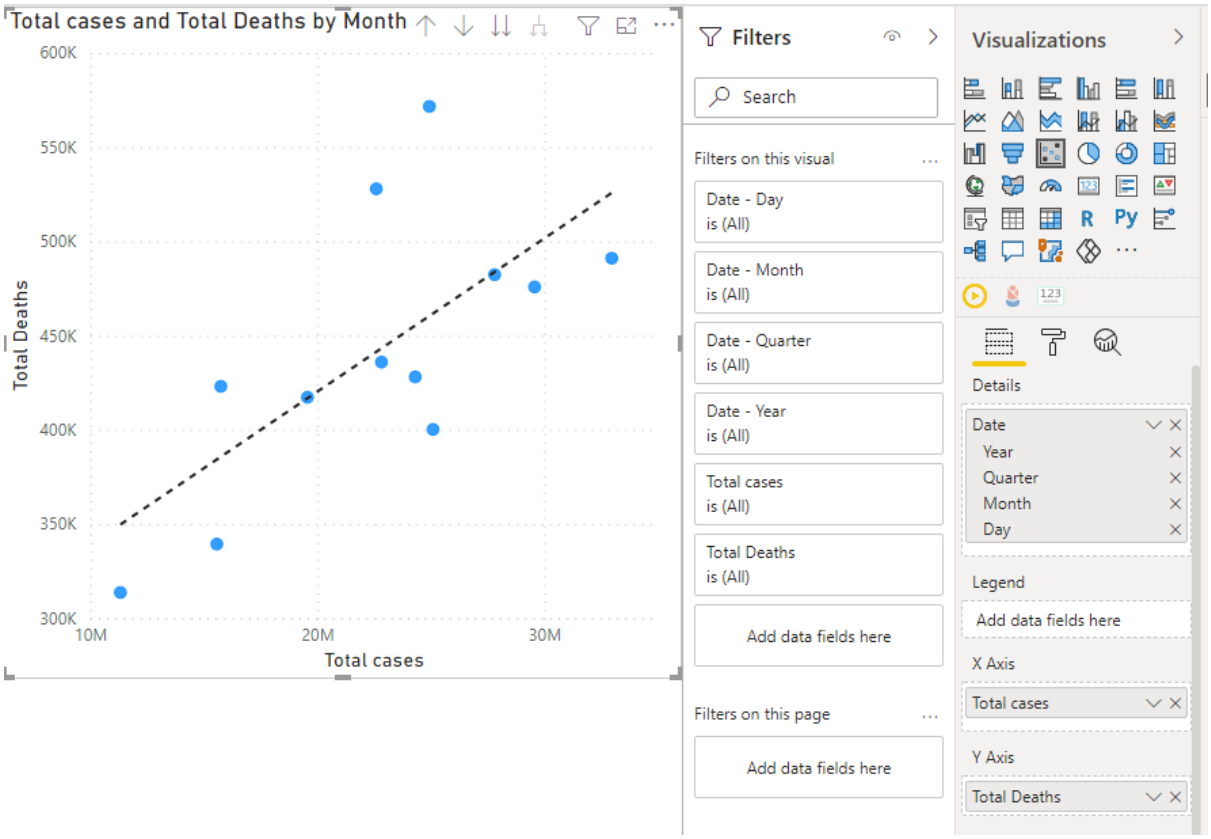
To visualize the Countries with the most Cases and Deaths on a daily basis, the Horizontal Bar chart needed to be imported by clicking on the '…' bar in the visualization tab and selecting 'Get more visuals'. After importing the chart, it was filtered to show only the top ten countries by deaths per day as shown below





This Visualization shows that on the 16th of December 2021 which is the last available date for our data, United states had the highest daily cases with a low death rate and United kingdom had the second highest cases with a very low death rate.

6. What is the relationship between Confirmed cases and deaths caused by Covid-19?



Scatter chart was chosen to visualize the relationship between the Total cases and the total death for each month.

Then to analyse the relationship further, a trend line was needed which was accessed by clicking on the analytics tab and clicking on the add button as shown below

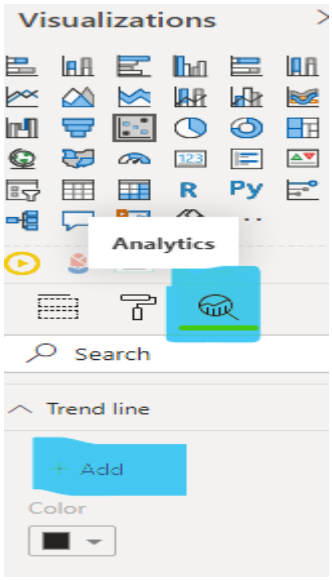


Figure 62 - Trend line

The trend line shows that there is a linear progression between the two data, which means as the total cases increase, the total deaths also increase.

7. What is the correlation between the death rate and the rate of testing in each month and which countries have the most tests?

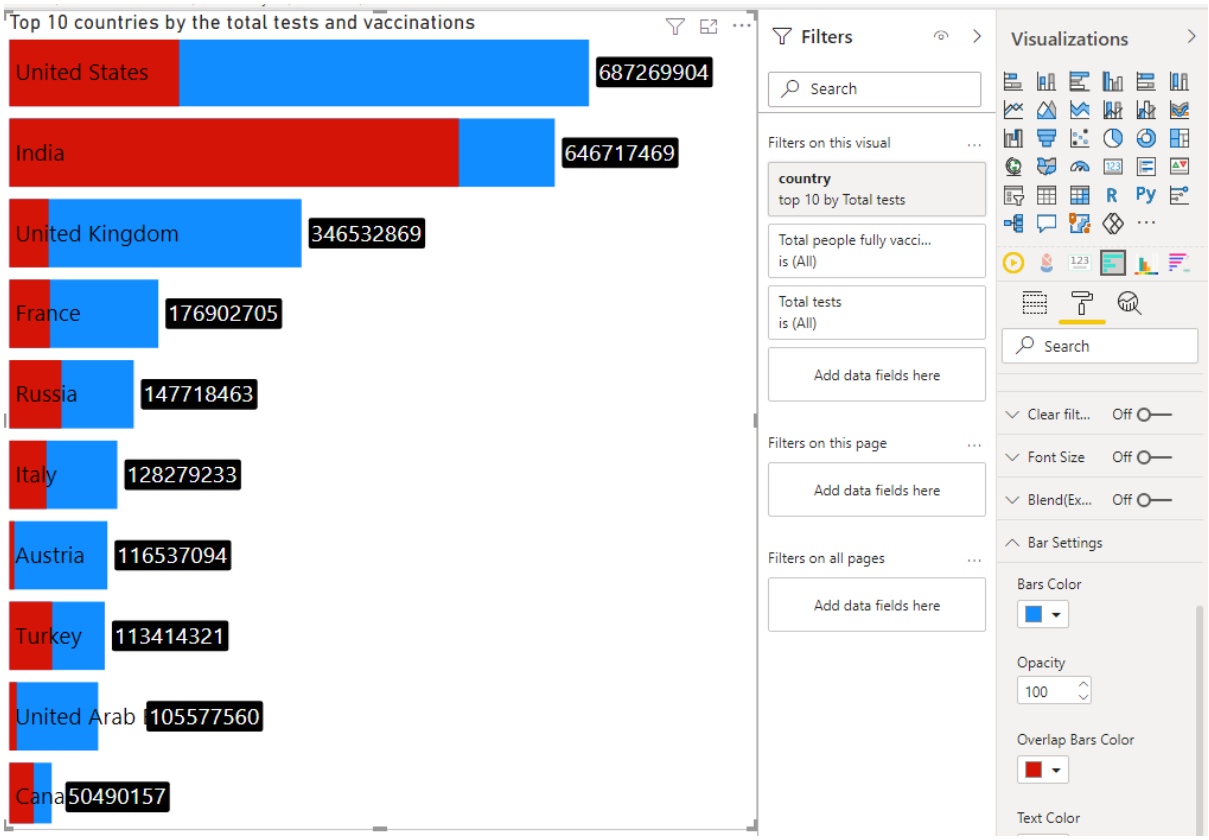


Figure 63 - Total tests and people fully vaccinated in each country

The horizontal bar chart which was imported earlier was chosen to visualize the top 10 countries by the total people fully vaccinated and testing that has been done in each country and it shows us that United states, India and United Kingdom are the countries that have done the most testing while India has a far higher rate of people fully vaccinated, it is slightly surprising that China is not in the top 10 because the virus started there and it is the most populous country in the world.

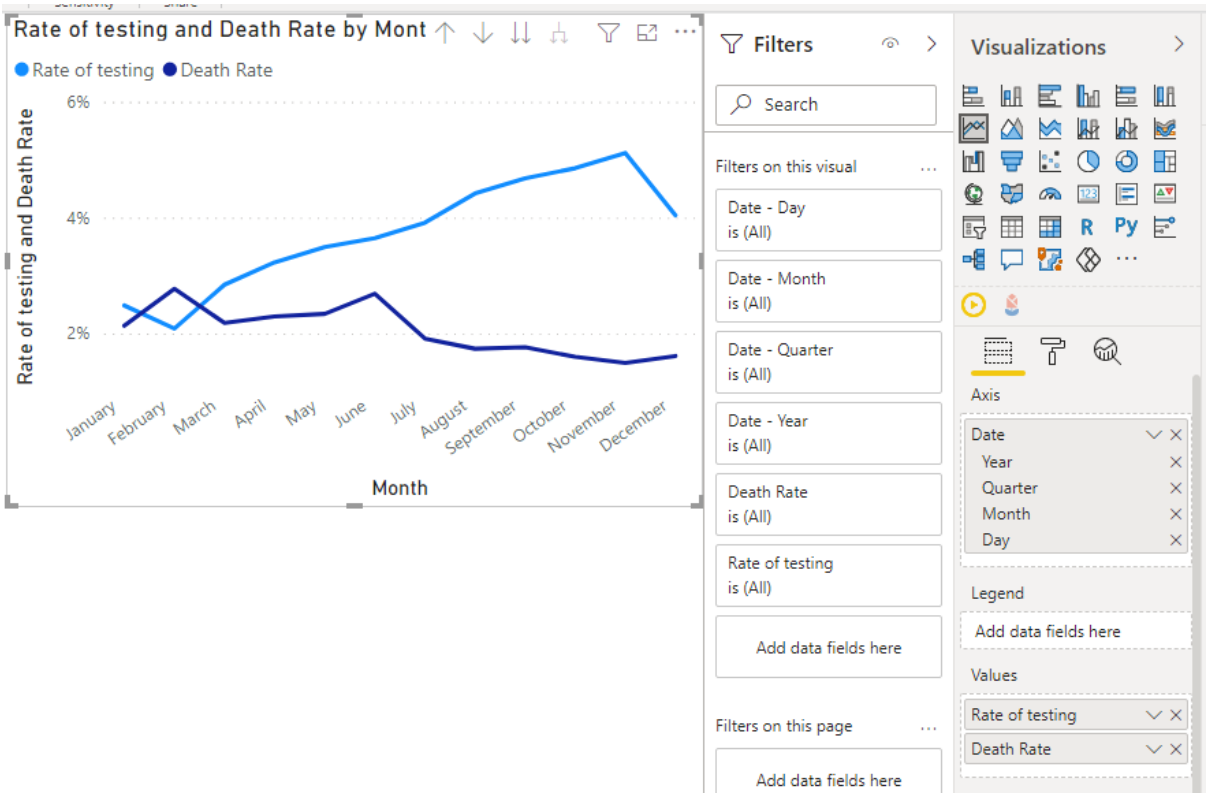


Figure 64 - correlation between rate of testing and deaths

Then to check the correlation between the rate of testing and the rate of death, a DAX measure for the Rate of Testing was created as shown below

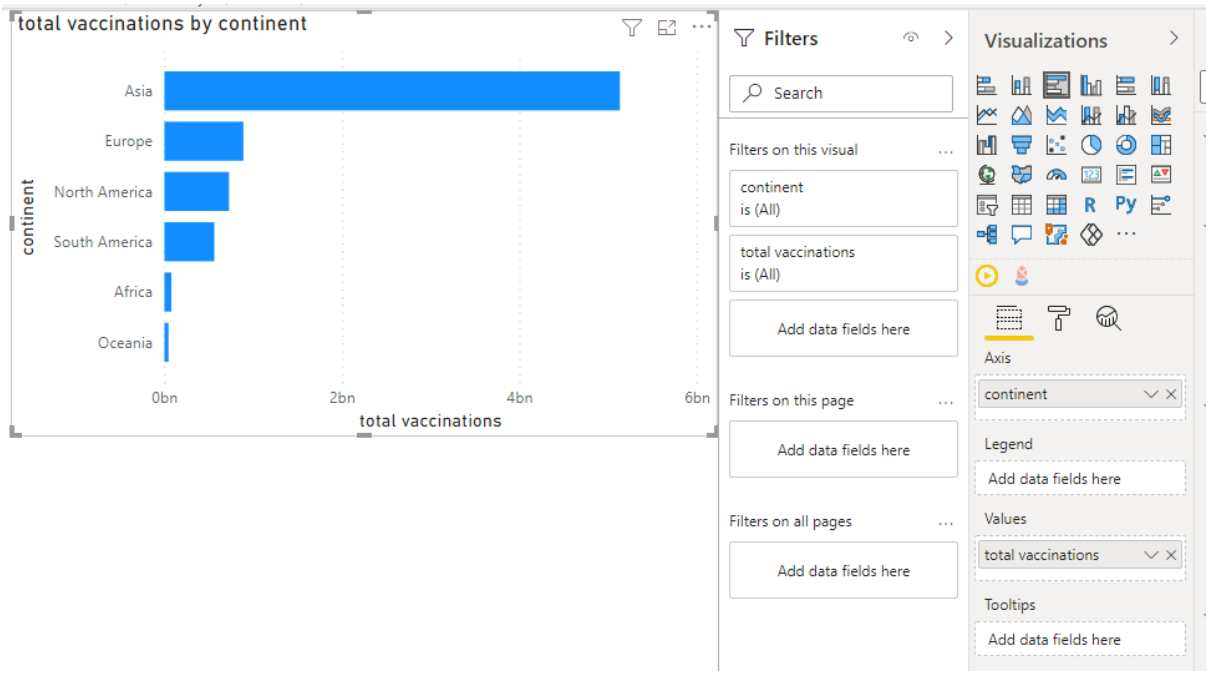
```
Rate of testing = [Total tests]/ SUM(Population[population])
```

Figure 65 - Dax formula for rate of testing

The line Chart was chosen to visualize the relationship between the rate of testing and deaths and as can be seen below, the increased testing rate had a very positive impact in reducing the rate of deaths over time, by November when there was a slight dip in the testing rate, the Death rate rose a little to show the importance of regular testing to contain Covid-19.

8. Which continents have the best vaccination process

Clustered bar chart was chosen to visualize the total vaccinations by continent, it was shown that Asia has had the most vaccinations followed by Europe while the likes of Africa and Oceania have had lower vaccinations done in their respective continents as shown below



To find the rate at which the population of each country has been fully vaccinated, a Dax formula was created as shown below

```
1  Fully Vaccinated rate = [Total people fully vaccinated]/sum(Population[population])
```

Then a slicer was created to toggle between the country whose fully vaccination rate is needed to be shown, Card was then chosen to visualize the vaccination rate as shown below

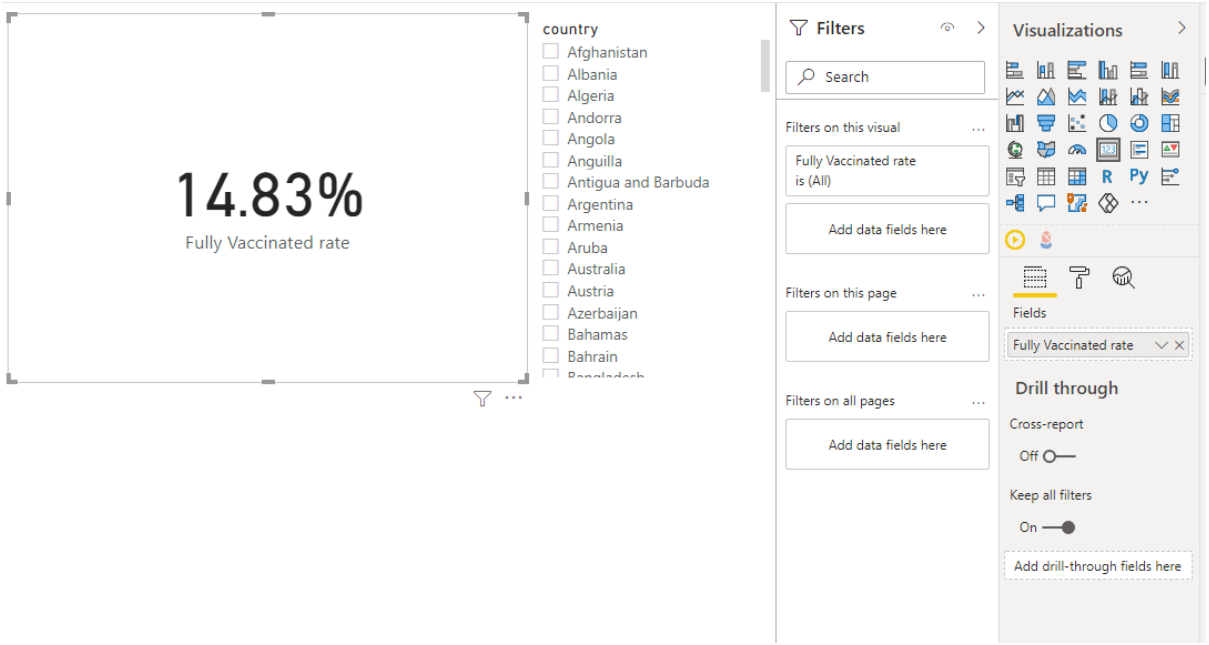
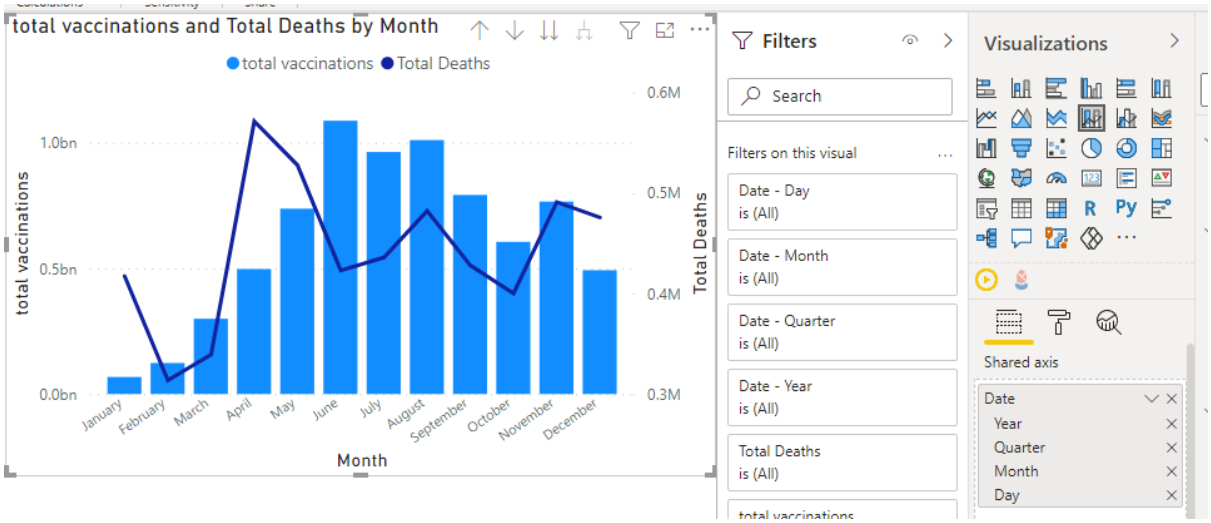


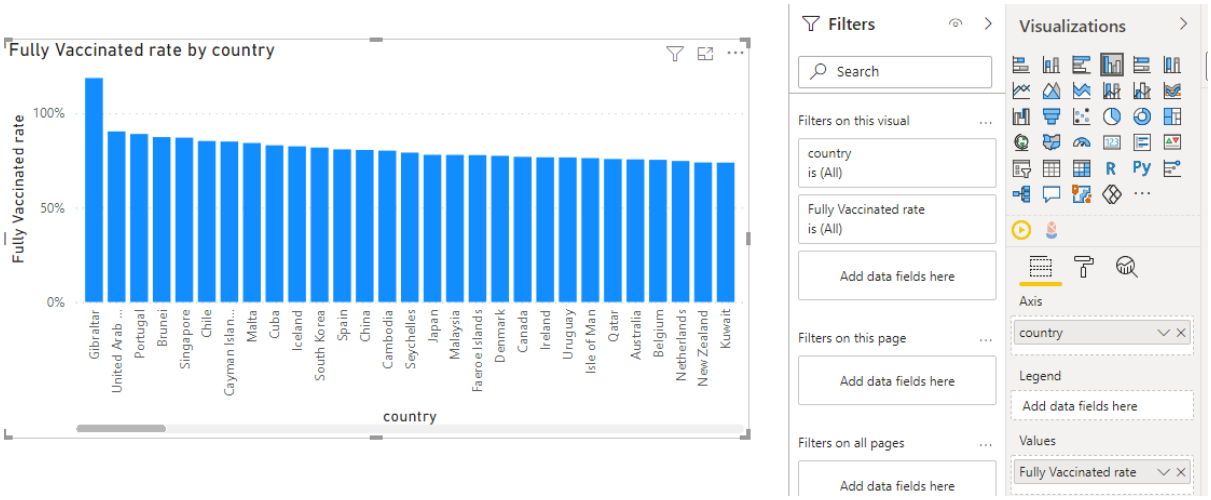
Figure 66 - Fully vaccinated rate

To visualize the total Vaccinations and Total Deaths by Month, the line and stacked column chart was chosen. It shows that when the Vaccination rate increases, the deaths tend to reduce which means vaccinations has been an effective measure against the Virus as shown below.

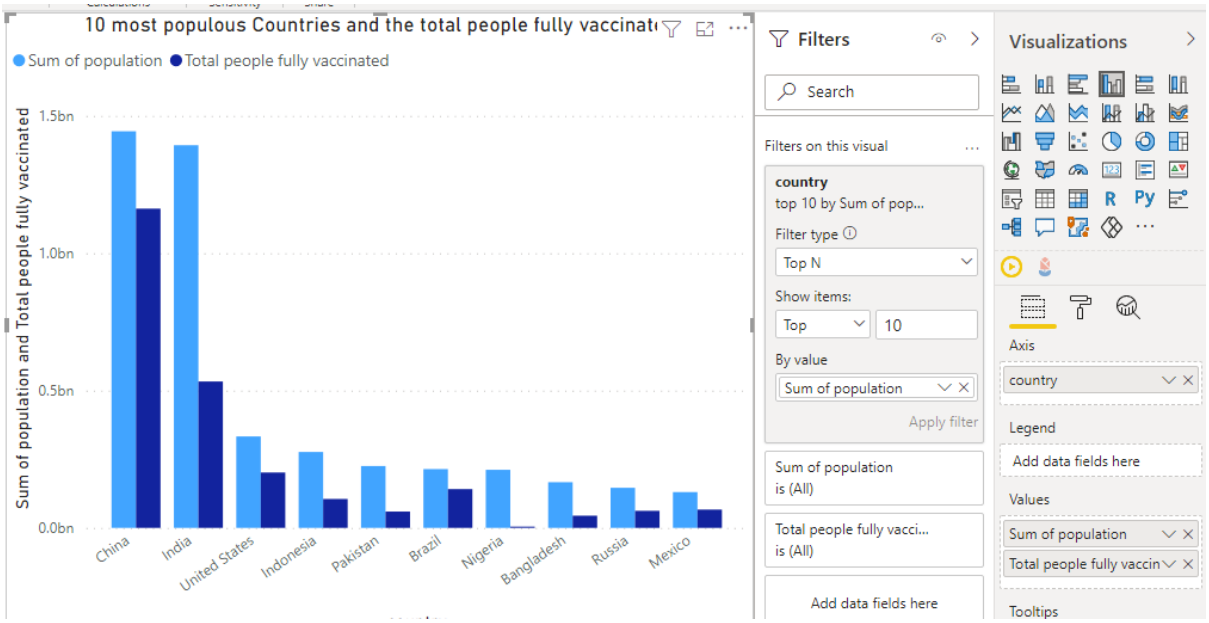


9. Which Countries has achieved Herd immunity?

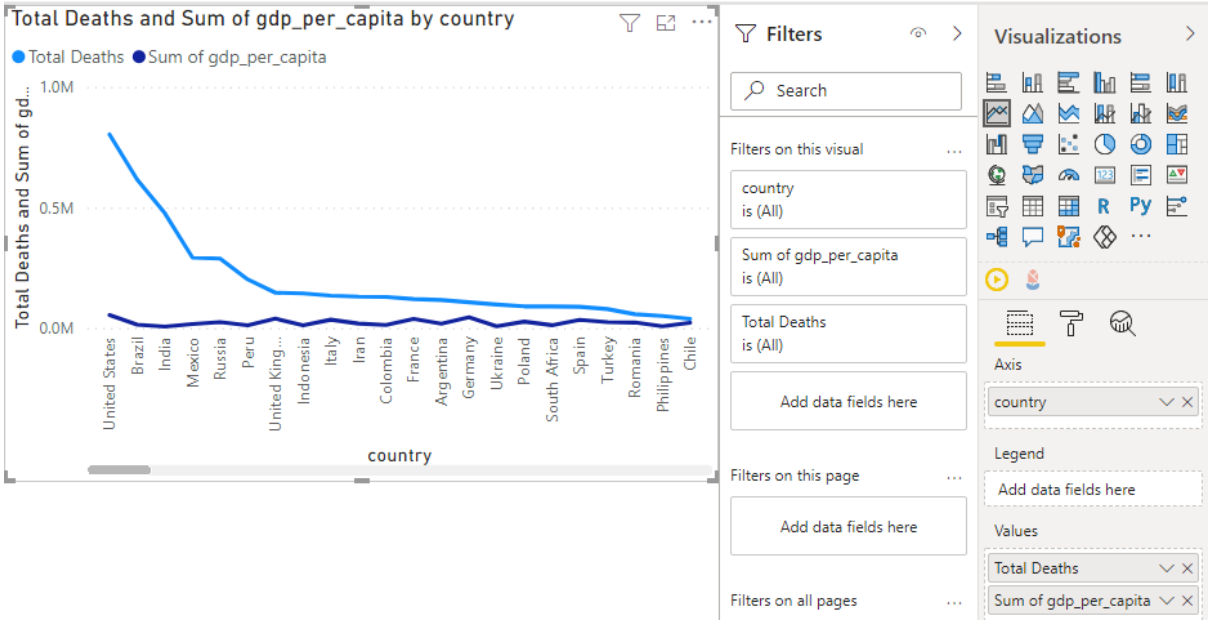
Herd immunity occurs when a large portion of a community (the herd) becomes immune to a disease, making the spread of disease from person to person unlikely. As a result, the whole community becomes protected — not just those who are immune. For Covid-19, full Vaccinations are the main way to be immune or at least have a high level of protection against the Virus, so to achieve that, the total number of people vaccinated must be very high against the overall population of the country. In order to visualize this, Clustered column chart was chosen and from the visualization. Because numerous cross-border workers from Spain were also given the vaccine in Gibraltar, Gibraltar has administered the full immunisation dose to more than its official population. UAE, Portugal, Brunei, Singapore, Chile, and other countries have the highest complete immunisation rates.



We also wanted to Visualize the total people fully vaccinated by the 10 most populous countries and clustered column was chosen to achieve that as shown below

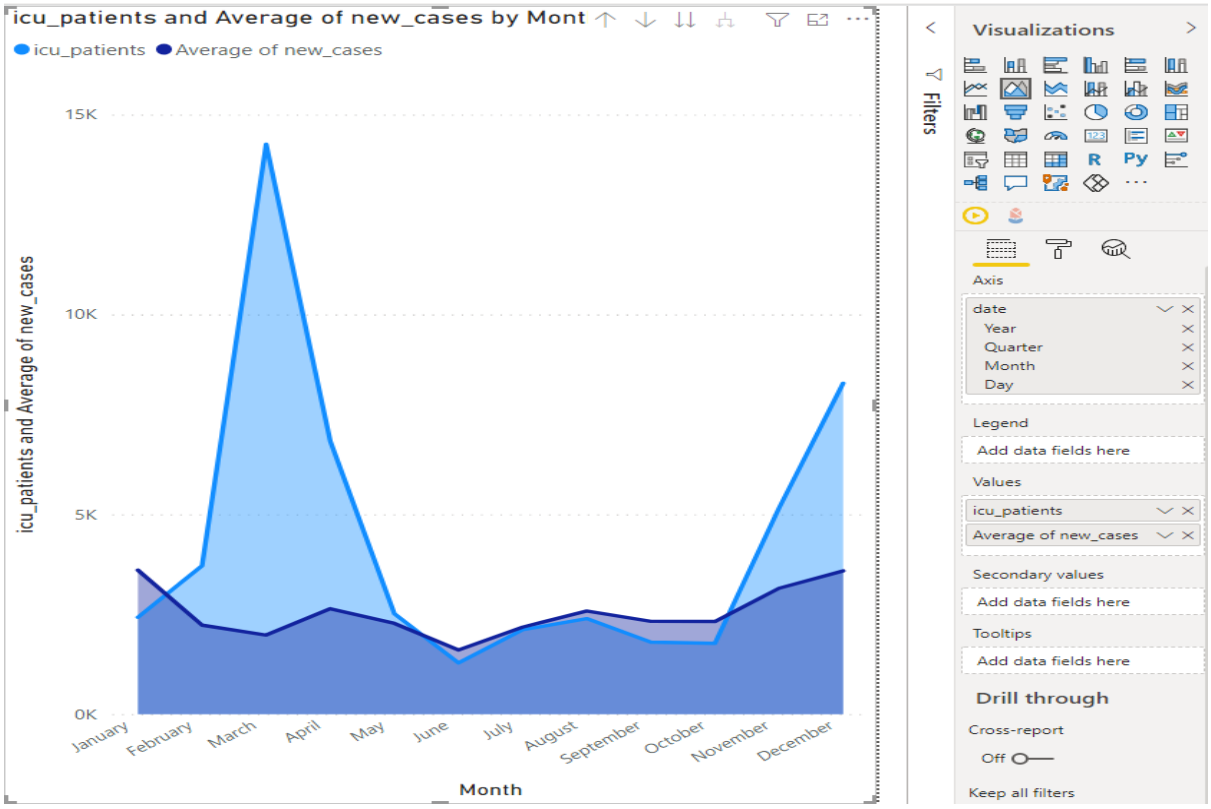


10. Does the GDP of a country affect the Death rate caused by Covid-19?



Line chart was used to visualize if there is any correlation between the wealth of a country and the deaths caused by Covid-19 and using the GDP-per-Capital. However, based on this chart there doesn't seem to be much of a correlation between the two which means Covid-19 affected every country irrespective of its financial strength.

11. Does Covid-19 cases cause an increase in the number of ICU patients?



Area Chart was chosen to visualize the effects of Covid-19 on the number of intensive Care patients. This graph shows that there was a spike in the number of ICU patients in the month of march but this was not caused by Covid-19 infections as while the confirmed cases have remained relatively constant, the number of ICU patients fluctuated quite a bit.

12. Does a country's Median age affect the Deaths caused by Covid-19 in different countries

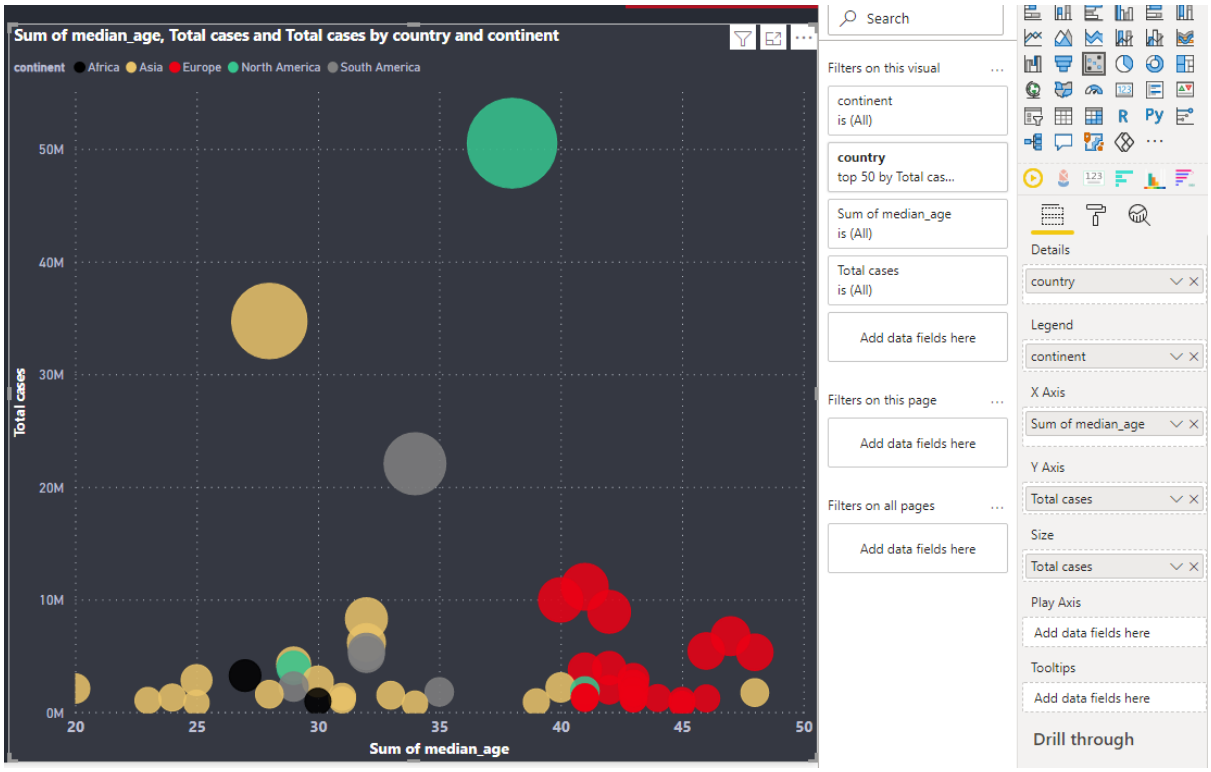


Figure 67 - median age

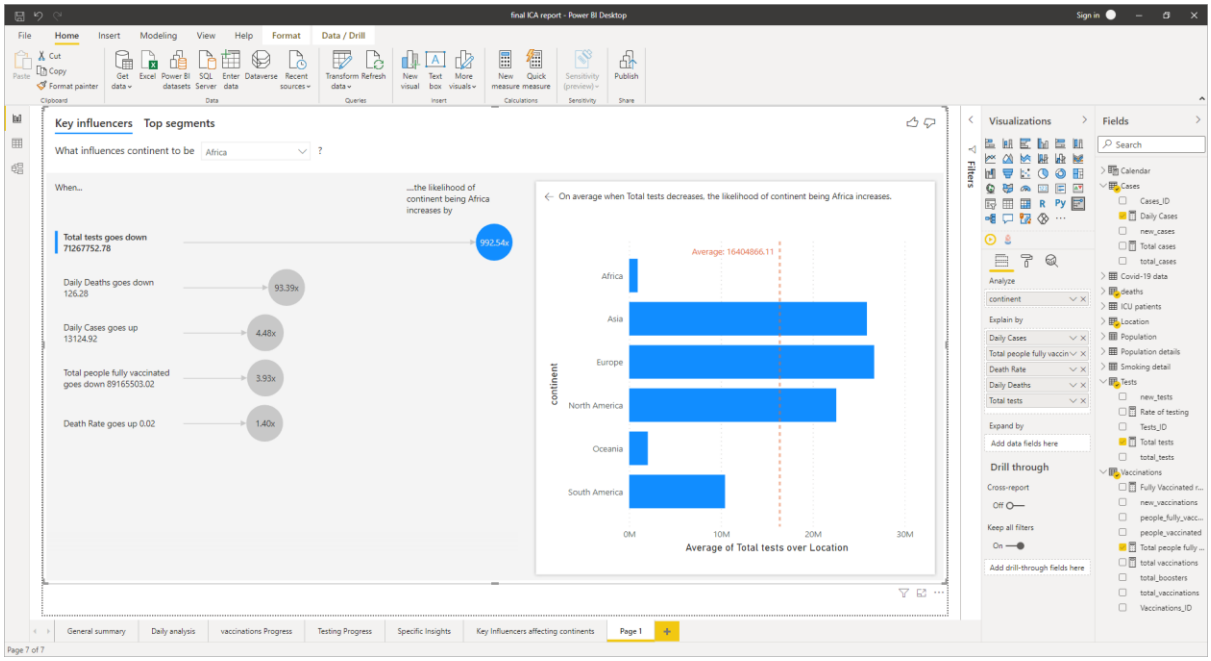
A country's median age is the age that divides the population into two numerical groups, which means half of the population are younger than that age and half are older. So, the lower the country's median age, the lower its populations average age. Since it was initially concluded that the Covid-19 was more deadly to older people, we decided to visualize the effect using scatter chart. We can see that countries in Europe generally have the most median age while Asian countries have the least. However, the median age does not appear to significantly affect the deaths caused by Covid-19.

KEY INFLUENCERS

1. What are the key Influencers that affects each Continent?

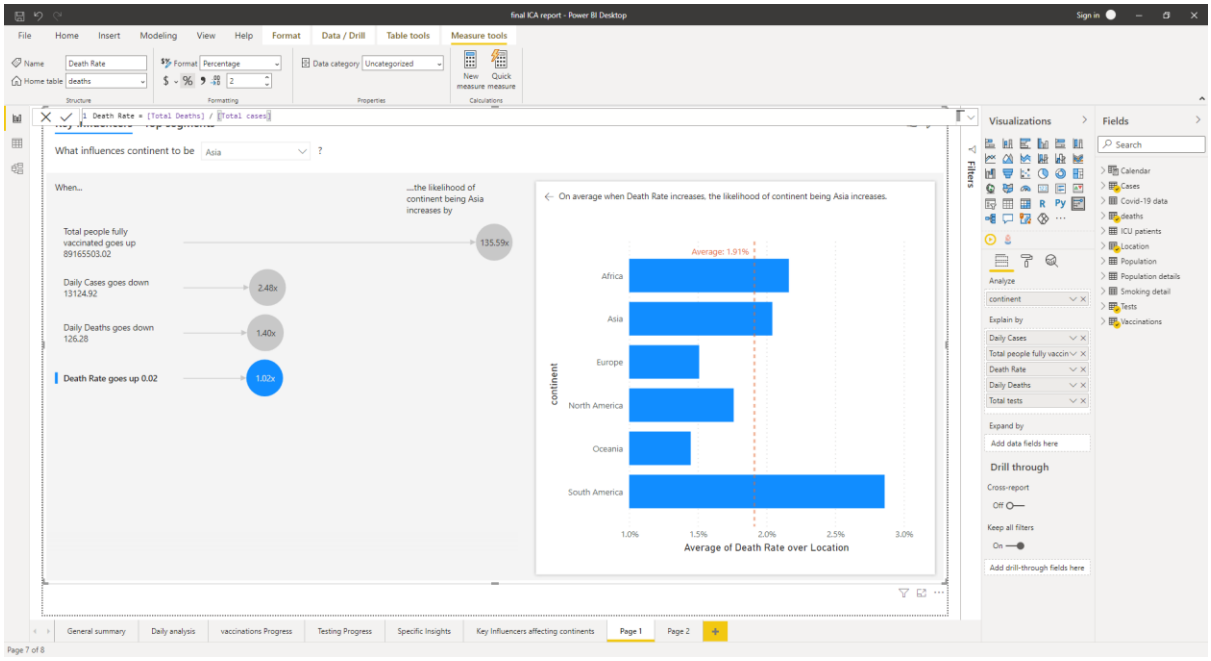
It is important to visualize key parameters that affects the rate of testing, vaccinations, confirmed cases and deaths in each continent so as to gain insights into the overall effects of COVID-19 in each continent. Fortunately, Power BI has a powerful tool to visualize the key influencers for each continent. Key influencers graph was then used to visualize these charts as shown below

- Africa



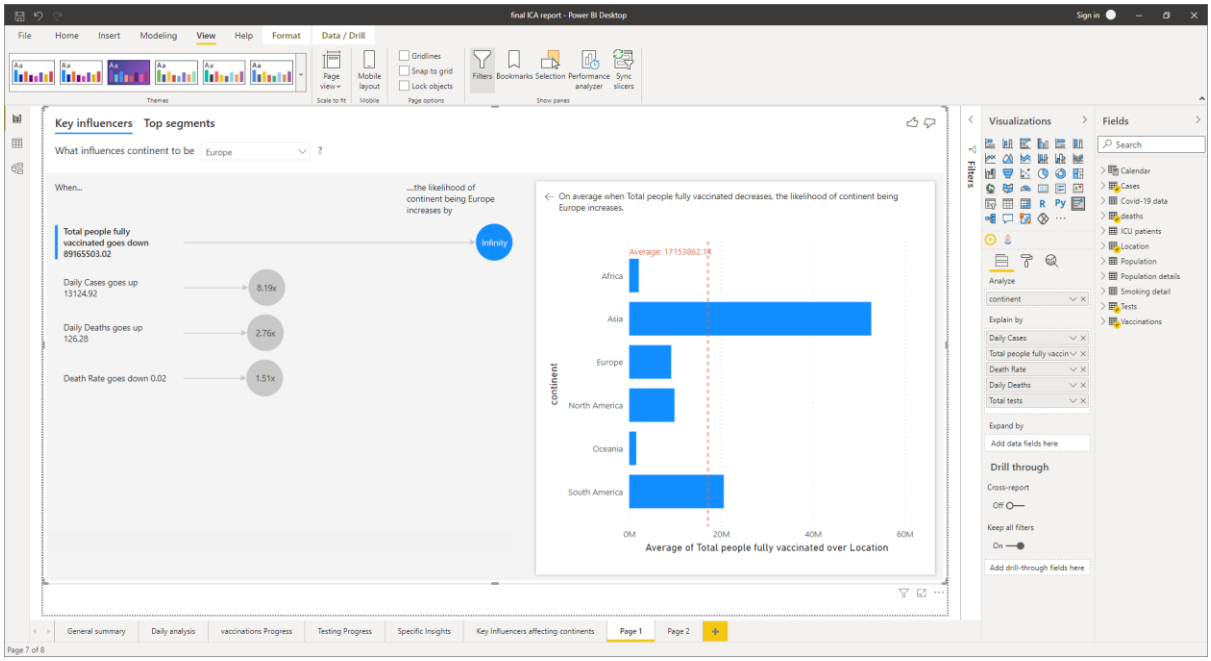
From this graph, Africa have carried out the lowest tests of the six continents so it also has the second lowest confirmed cases and the second lowest of fully vaccinated people. However, all these did not increase the number of people the virus kills on a daily basis, with the daily deaths the second lowest. The death rate is the second highest though which implies that on average more people diagnosed with Covid are more likely to die than in most of the other continents.

- Asia



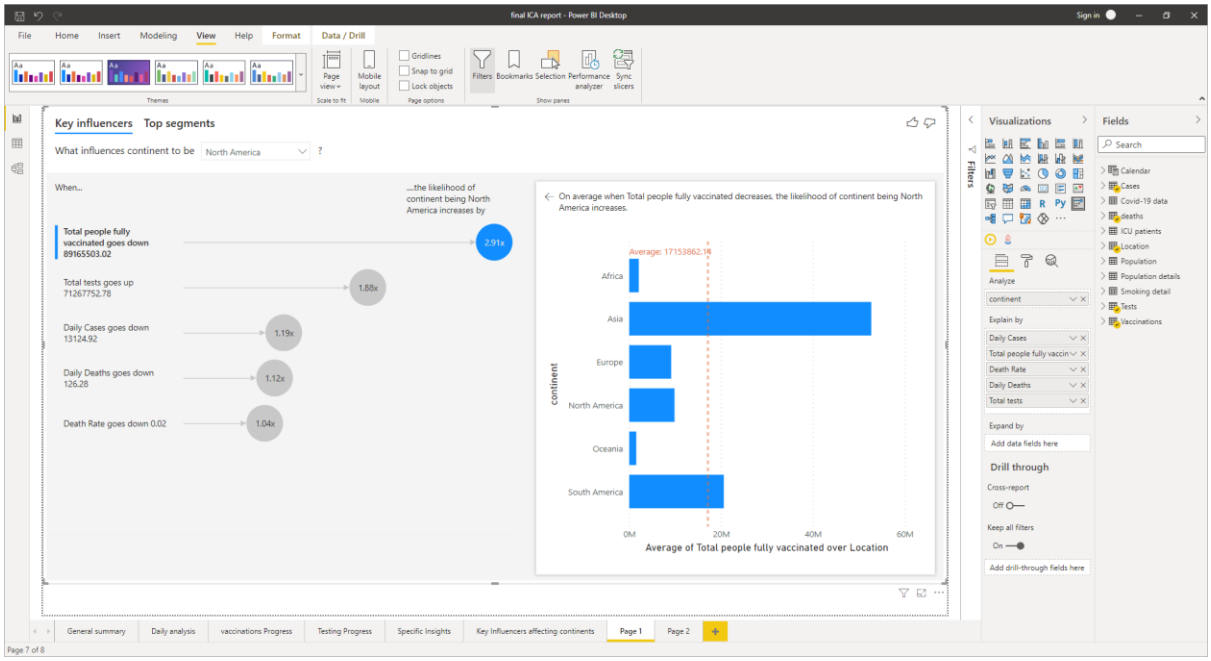
From this graph, Asia has the highest number of people fully vaccinated, one of the lowest confirmed cases, it also has on average one of the lower daily deaths. However, it has a higher death rate than most of the other continents on average.

- Europe



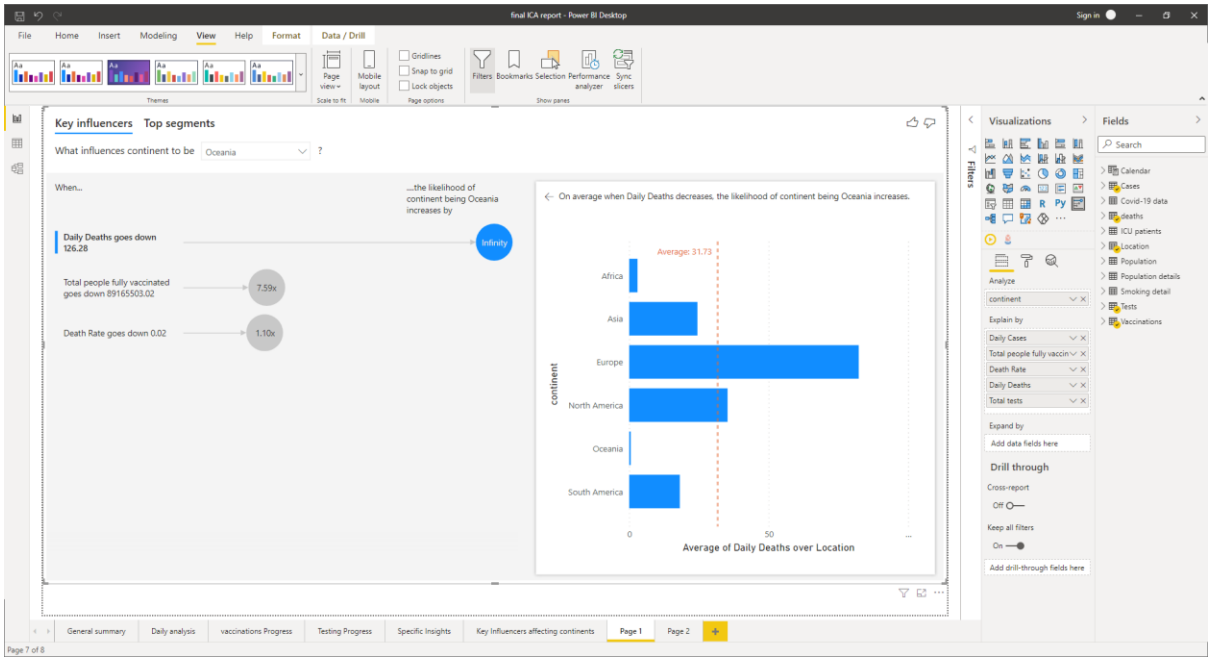
From this graph, Europe has a low rate of fully vaccinated people on average, the highest number of confirmed cases and the highest number of daily deaths. However, it has one of the lowest death rates on average.

- North America



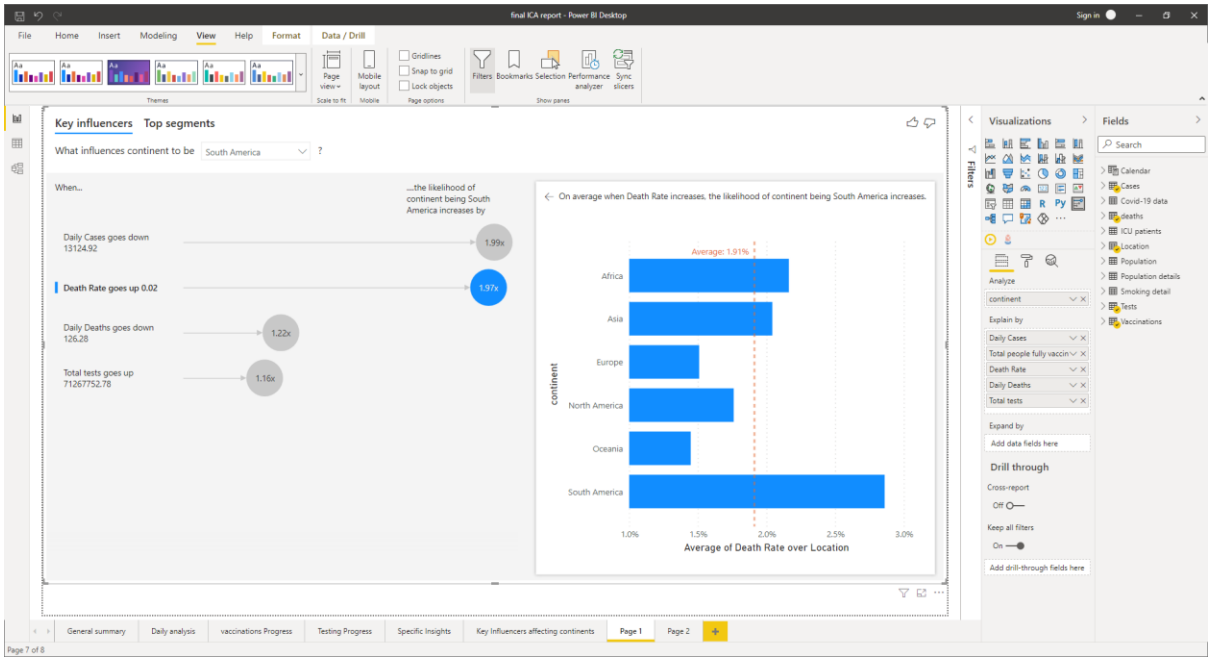
From this graph, North America has a lower average rate of fully vaccinated people, higher average rate of testing and general lower level of cases. It also has lower death rate and average number of daily deaths.

- Oceania



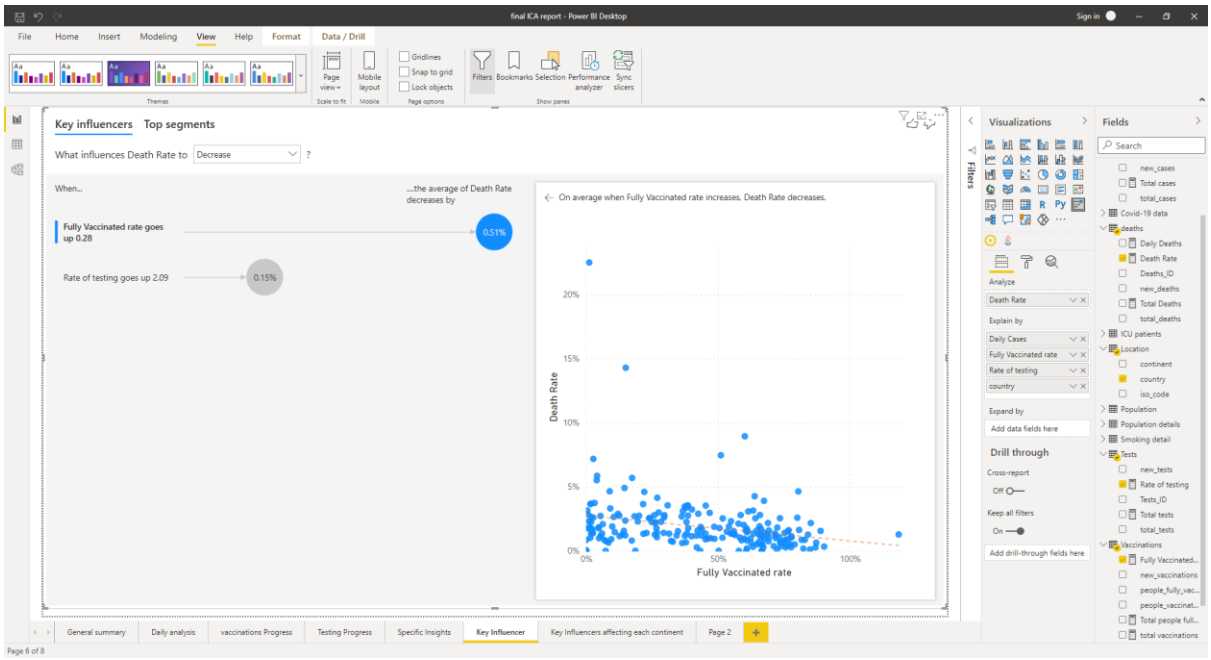
Oceania has a good rate of fully vaccinated people in relation to its population and the lowest number of daily deaths on average and the lowest death rate of all the continents. This shows that Oceania has done a good job of battling the Virus relative to the other continents.

- South America



South America has a low rate of daily confirmed cases but the worst death rate of all the continents. Which highlights how much more fatal Covid-19 has been in South America in relation to the other continents.

2. What are the key influencers that may increase or decrease the death Rate



The death rate decreases when the rate at which the population is fully vaccinated and when the rate of testing increases. This shows that the measures put in place to fight against Covid-19 has had a positive effect.

ARRANGING THE VISUALIZATIONS

The visualizations were then properly arranged, buttons were added to the report pages to make navigations between the pages easier. The details of the buttons are as follows

- PREVIOUS PAGE

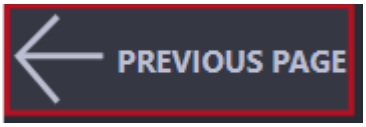


Figure 68 - Previous page navigation button

This button helps to navigate to the page in the Power BI report

- NEXT PAGE

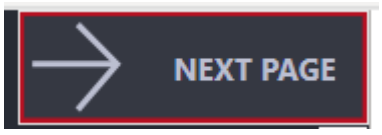


Figure 69 - Next page navigation button

After this was added, this is how one of the report pages looks

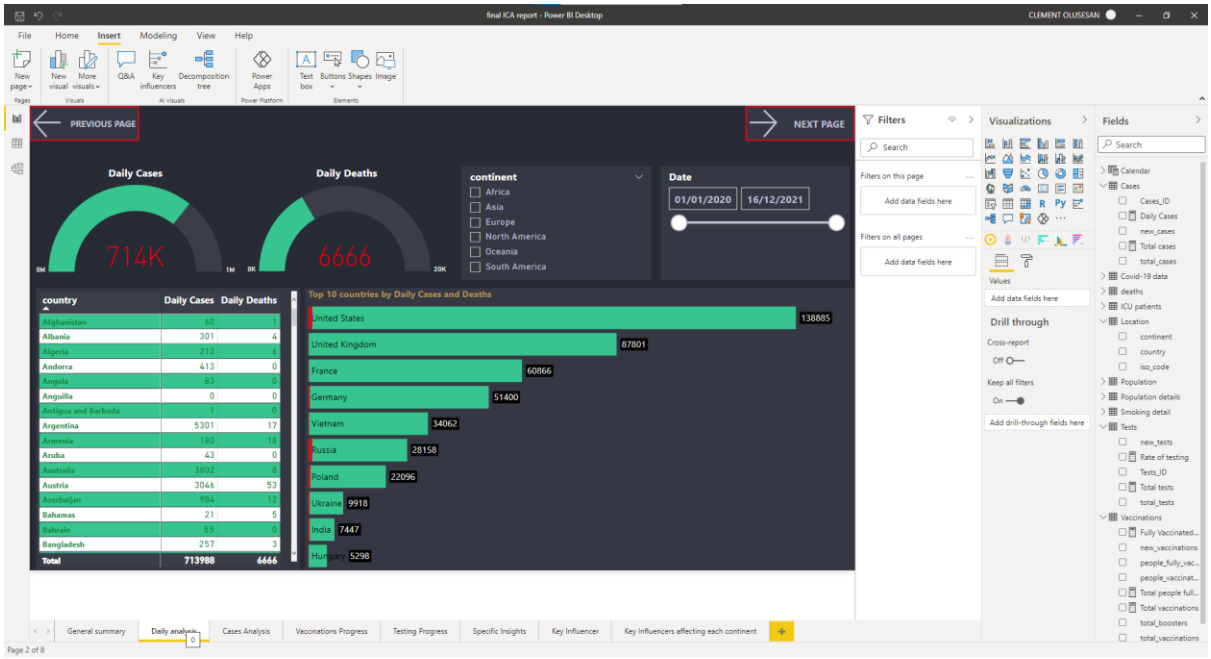
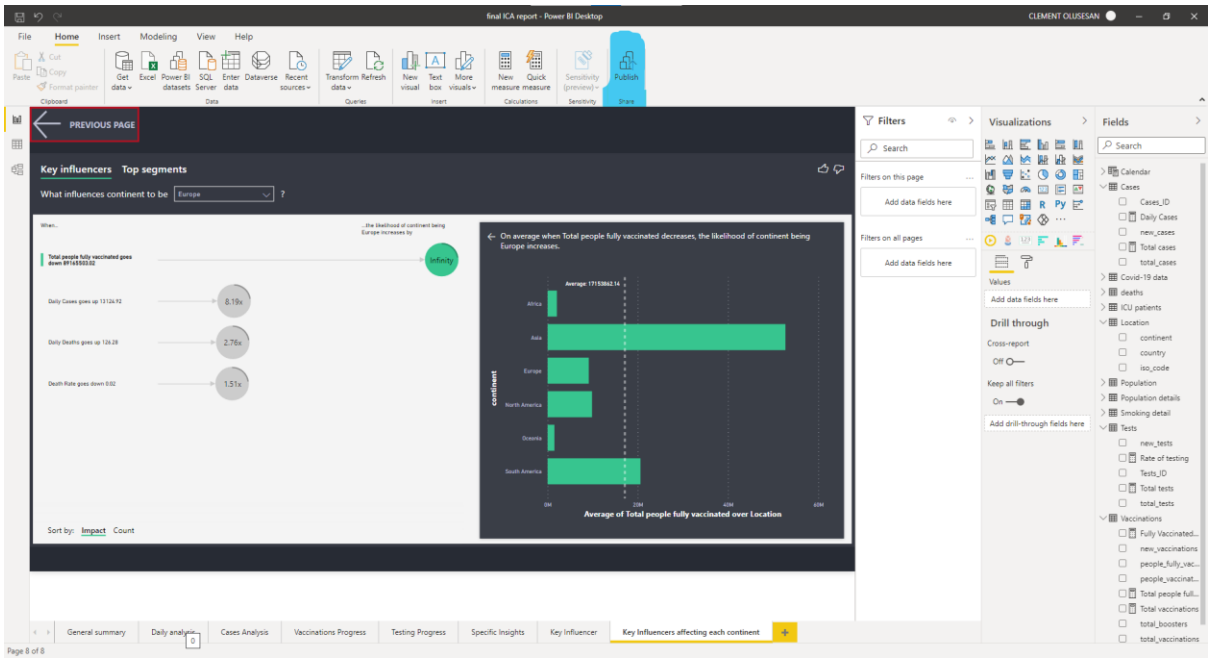


Figure 70 - report page after navigations added

Finally, the generated Business intelligence Report was published as shown in the figure below



CONCLUSIONS

RECOMMENDATIONS