

BIG DATA AND BUSINESS INTELLIGENCE MODULE

COVID-19 PANDEMIC ANALYSIS

BY: SALISU ALIYU ABDULLAHI

C2224671

Table of Contents

Executive Summary	3
Key Findings	3
Recommendations	4
Introduction	4
Data Source and Description	4
BI Requirements\Questions	6
Finding based on analysis and evaluation	7
Recommendations	13
Conclusions	13
Appendix: BI Design	14
Data Pre-Processing or Data Cleansing	14
Data Loading	14
Data Cleaning	17
BI Data Modelling via Star Schema - Facts and Dimensions.	24
Dimension Tables	26
Fact Table	35
DAX and M Language	37
M. Language.	42
Dashboard	46
Global Situation Dashboard	47
Continent Situation Dashboard	48
Country Situation Dashboard	49
Country-Specific Situation (Nigeria)	49
Summary Information	50
References	51

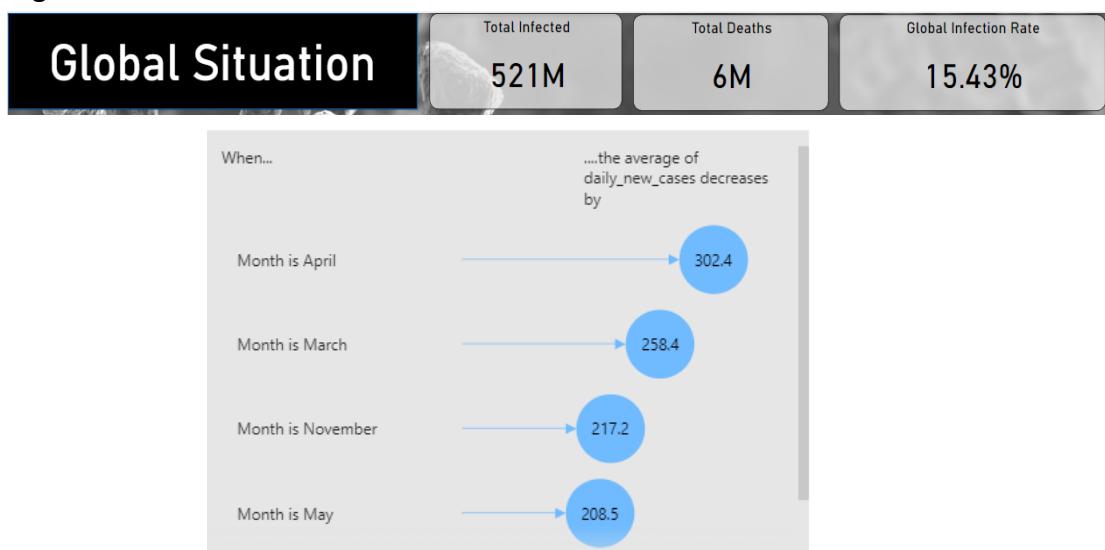
Executive Summary

This report examines the coronavirus (COVID-19) pandemic spread since its outbreak in the world. This report investigates confirmed infection and death cases in the world, across the continents, countries and specifically in Nigeria. The purpose of this report is to show how COVID has spread since its outbreak in the world and specifically in Nigeria; and to suggest possible influences that can reduce its spread and aid in the fight to reduce the spread of the pandemic and/or eradicate the disease totally.

Key Findings

The below are the findings from the business intelligence analysis of the COVID-19 pandemic (as of May 14, 2022,):

- The average COVID-19 infection rate currently stands at 15.43%
- The total number of confirmed infected cases since its outbreak was at five hundred and twenty-one million (521M) with the 2022 being the year with the highest number of infection cases totalling at two hundred and twenty-nine million (229M).
- Over 6 million (6M) people have died from the corona virus pandemic globally with the year 2021 accounting for the largest number of COVID deaths (3.53M) in a year since the pandemic started.
- Europe is the worst hit (194 million) continent by the corona virus with Australia/Oceania being the least hit (8 million).
- Only 19.5 percent of countries in the world have its infection rates below the global infection risk rate.
- Nigeria with a population of over two hundred and fifteen (215) million has an infection rate of 0.12 percent with about five million (5,114,703) covid tests done.
- Since the pandemic, the months of March, April, May, October, and November on average have seen a decline in the daily confirmed infection cases of COVID-19 in Nigeria.



Recommendations

- There is need for increased attention towards reducing the number of confirmed infected cases in the world as this has been increasing every year.
- Countries across Europe, Asia and North America need to take more stringent measures to reduce the spread of COVID in its regions.
- Lockdown and travel restrictions should be implemented in top countries with high infection risk rates to help reduce the spread of the virus.
- In Nigeria, more tests should be carried out in its population as analysis suggests a high number of the population's infection is not being tested, confirmed, and reflected.
- More research and analysis need to be done to understand why there is a decrease in daily infection in the months of March, April, May, October, and November and how this can be applied to other months of the year.

Introduction

Since its outbreak at the end of 2019, the corona virus pandemic has become a name on everyone's lips. This is not for a positive reason as the outbreak of the disease put the world on hold for at least month. This disease has affected everything in the world, from businesses to how we interact with each other and the environment. While there has been progress so far in the fight against the disease, there is still much to be done. This project seeks to investigate how the disease has fared so far across various demographics since its outbreak. This is examined based on its impact in the world, across the continents, countries and specifically in Nigeria. Analytical skills such as data pre-processing, modelling, in-depth analysis using AI and visualisations were employed towards this project.

Data Source and Description

The covid dataset was compiled on woldometers.info by Joseph Assaker and contains covid infection, deaths, and tests information about 225 countries. The organization main goal is to provide easily accessible data on the covid pandemic and assist in the fight to reduce the spread of the infection and possible eradication. Furthermore, the dataset was extracted from Kaggle via the link: <https://www.kaggle.com/datasets/josephassaker/covid19-global-dataset>.

The dataset is an file containing two sets of data; the first dataset(daily data) is a dataset that shows the daily infections and deaths due to covid across countries in the world while the second dataset (summary data) provides data on the total confirmed cases of covid, the total confirmed deaths due to covid across each country, the severity of these cases, and the number of cases, deaths and tests in relation to each country's population.

Information provided in this dataset was used to measure infection and death rates across various demographic, compare the average infection rate of the world against all the countries in this dataset and examine how the pandemic has fared in Nigeria.

In addition, this data only contains data from February 15th, 2020, to May 14th, 2022, for all countries except China, whose data started from January 22nd, 2020, and Palau whose data started from August 25th, 2021.

Both Datasets are in Comma Separated Values (CSV) format with the daily data set containing 7 columns and 184788 rows while the summary data contains 12 columns and 227 rows. Description of columns contains in the table in each data set (Daily Data and Summary Data) is outlined below.

Table 1: worldometer_coronavirus_daily_data

Index	Column Name	Description
1	Date	Date information was recorded
2	Country	Country Location of the information
3	Cummulative_Total_Cases	Total number of confirmed cases since first case
4	Daily_New_Cases	Total number of cases confirmed for a particular date
5	Active_Cases	Number of cases that have not recovered or died
6	Cummulative_total_deaths	Total number of confirmed deaths due to covid since first case
7	Daily_new_deaths	Total number of confirmed deaths for a particular date

Table 2: worldometer_coronavirus_summary_data

Index	Column Name	Description
1	Country	Country Location of the information
2	Continent	Continent in which the Country Location is in
3	Total_confirmed	Total number of confirmed cases since first case
4	Total_deaths	Total number of confirmed deaths due to covid since first case
5	Total_recovered	Total number of people who have recovered from covid since first case
6	Active_cases	Number of cases that have not recovered or died
7	Serious_critical	Number of people left in critical state due to covid infection
8	Total_cases_per_1m_popuation	Total number of confirmed cases since first case per 1 million of the country's population
9	Total_deaths_per_1m_popuation	Total number of confirmed deaths due to covid since first case per 1 million of the country's population
10	Total_tests	Total number of covid tests carried out
11	Total_tests_per_1m_popuation	Total number of covid tests carried out since first case per 1 million of the country's population
12	Population	Describes Population of the country

Below is also a screenshot of the datasets table in excel:

	A	B	C	D	E	F	G	H	I	J	K	L
1	country	continent	total_conf	total_deat	total_reco	active_cas	serious_o	total_case	total_deat	total_tests	total_tests	population
2	Afghanista	Asia	179267	7690	162202	9375	1124	4420	190	951337	23455	40560636
3	Albania	Europe	275574	3497	271826	251	2	95954	1218	1817530	632857	2871945
4	Algeria	Africa	265816	6875	178371	80570	6	5865	152	230861	5093	45325517
5	Andorra	Europe	42156	153	41021	982	14	543983	1974	249838	3223924	77495
6	Angola	Africa	99194	1900	97149	145		2853	55	1499795	43136	34769277
7	Anguilla	North Amer	2984	9	2916	59	4	195646	590	51382	3368870	15252
8	Antigua And	North Amer	7721	137	7511	73	1	77646	1378	18901	190076	99439
9	Argentina	South Amer	9101319	128729	8895999	76591	372	197992	2800	35716069	776974	45968174
10	Armenia	Asia	422896	8623	412048	2225		142219	2900	3068217	1031834	2973558
11	Aruba	North Amer	35693	213	35199	281		331689	1979	177885	1653053	107610
12	Australia	Australia/N	6593795	7794	6199822	386179	129	253112	299	70885598	2721042	26050899
13	Austria	Europe	4212492	18303	4135885	58304	58	462804	2011	1.85E+08	20328801	9102106
14	Azerbaijan	Asia	792638	9709	782869	60		76885	942	6838458	663324	10309383
15	Bahamas	North Amer	33871	801	32488	582		84626	2001	233473	583330	400242
16	Bahrain	Asia	576997	1479	569758	5760	4	318491	816	9775981	5396149	1811659
17	Bangladesh	Asia	1953012	29127	1899419	24466	1273	11643	174	14051455	83767	1.68E+08
18	Barbados	North Amer	76261	433	71497	4331		264768	1503	669421	2324136	288030
19	Belarus	Europe	982867	6978				104078	739	13220483	1399951	9443535
20	Belgium	Europe	4116397	31613	3941350	143434	117	352324	2706	33846023	2896893	11683561
21	Belize	North Amer	57896	676	56761	459	6	140843	1644	549217	1336070	411069

	A	B	C	D	E	F	G
1	date	country	cumulative_total_cases	daily_new_cases	active_cases	cumulative_total_deaths	daily_new_deaths
2	15/02/2020	Afghanistan	0		0	0	0
3	16/02/2020	Afghanistan	0		0	0	0
4	17/02/2020	Afghanistan	0		0	0	0
5	18/02/2020	Afghanistan	0		0	0	0
6	19/02/2020	Afghanistan	0		0	0	0
7	20/02/2020	Afghanistan	0		0	0	0
8	21/02/2020	Afghanistan	0		0	0	0
9	22/02/2020	Afghanistan	0		0	0	0
0	23/02/2020	Afghanistan	0		0	0	0
1	24/02/2020	Afghanistan	1		1	0	0
2	25/02/2020	Afghanistan	1	0	1	0	0
3	26/02/2020	Afghanistan	1	0	1	0	0
4	27/02/2020	Afghanistan	1	0	1	0	0
5	28/02/2020	Afghanistan	1	0	1	0	0
6	29/02/2020	Afghanistan	1	0	1	0	0
7	01/03/2020	Afghanistan	1	0	1	0	0
8	02/03/2020	Afghanistan	1	0	1	0	0
9	03/03/2020	Afghanistan	1	0	1	0	0
0	04/03/2020	Afghanistan	1	0	1	0	0
1	05/03/2020	Afghanistan	1	0	1	0	0

Figure 1: Screenshot of data in excel

BI Requirements\Questions

This project seeks to provide the World Health Organization (WHO) and the Nigerian Centre for Disease Control (NCDC) with information on how corona virus has fared across the world and in Nigeria; and to suggest possible influences that can reduce its spread and aid in the

fight to reduce the spread of the pandemic and/or eradicate the disease totally. The analysis is majorly concerned with the following questions:

- What is the daily infection and death count of the virus in the world across each year since its outbreak and what is the average infection rate globally?
- What is the virus infection spread across the continents, how many deaths has it led to and what are the worst hit and least hit continents by the virus?
- What percentage of countries are below and above the global infection rate average and what are the top 10 countries with high risk of infection?
- To show the daily count distribution of infection and deaths due to the virus in Nigeria since outbreak
- To identify the periods which there is a decline in the number of daily infections in Nigeria

Finding based on analysis and evaluation

The Global situation dashboard below visualises the result of analysis carried out to answer question one. The visualisation contains a drill down timeline zoom chart of Confirmed infected cases and confirmed deaths. These charts show the breakdown of covid infected cases by year (2020 to date). Furthermore, these zoom charts allow drill down into any year of choice should you need further analysis of covid cases based on the months of the year and much further into each day of any month. This chart was used because it is an interactive chart as it provides specific details into any selected period as shown below. The drill down timeline zoom chart shows increase in daily infected cases yearly with 2022 having the highest cases of infection. On the drill down timeline showing death cases, 2021 had the highest death cases when compared to the other years. The dashboard also contains visuals of three cards which show the total number of confirmed infected cases, the total number of deaths due to covid and the average infection rate which from the card stands at 15.43 percent globally.

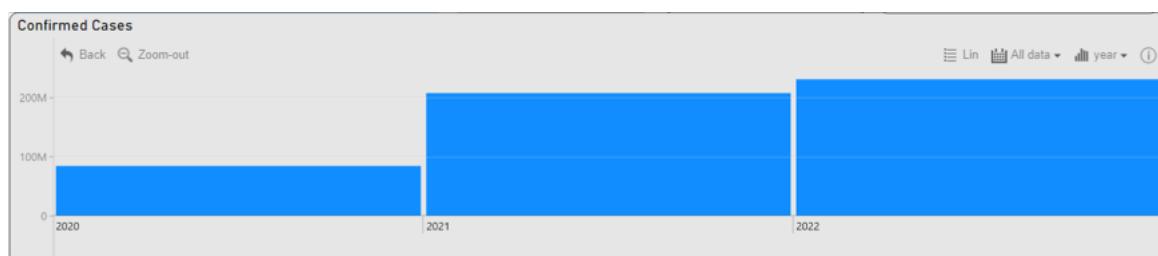


Figure 2: Distribution of confirmed infection cases by year

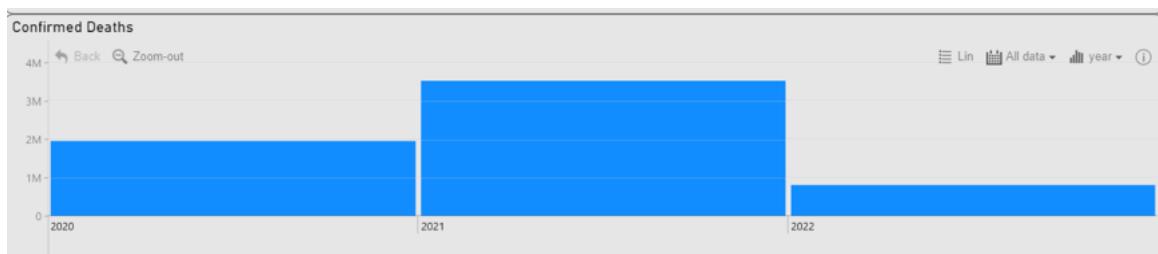


Figure 3: Distribution of confirmed covid deaths by year



Figure 4: Overview of Total Infection, deaths, and Infection Rate

The Continental Situation dashboard visualises the result of analysis carried out to answer question two. The visualisation contains an aster plot of infection cases distribution across continents in the world. From the plot, we can infer that Europe was the worst hit continent by the corona virus with over 194 million infection cases over the period and Australia/Oceania continent the least hit by the virus with about 8 million infection cases over the same period. Another visualisation in this dashboard (Clustered bar chart) shows the confirmed deaths due to covid across continents in the world. This chart shows that Europe witnessed the highest number of deaths due to the virus with 1.83 million people dying from the pandemic. Also, Australia/Oceania witnessed the lowest number of deaths due to covid with its number at a low of less than a million (about ten thousand). A slicer was added to the dashboard to aid easy navigation and highlight any particular year of choice in the period. A multi-row card was also visualised to aid in highlighting the list countries that fall under each continent under this report.

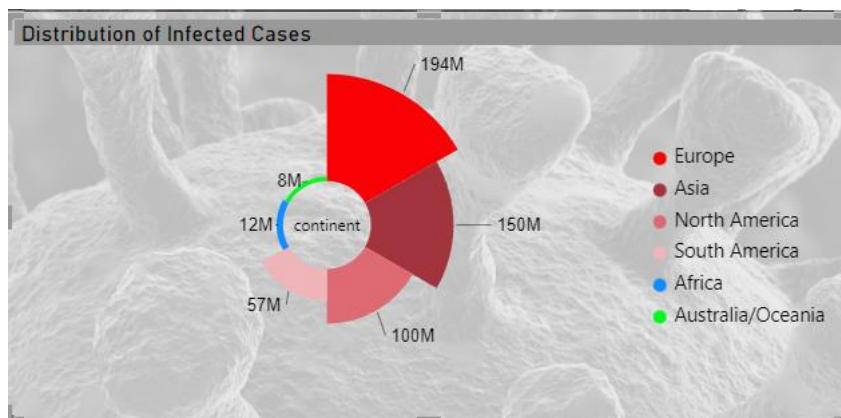


Figure 5: Distribution of infection cases across Continents

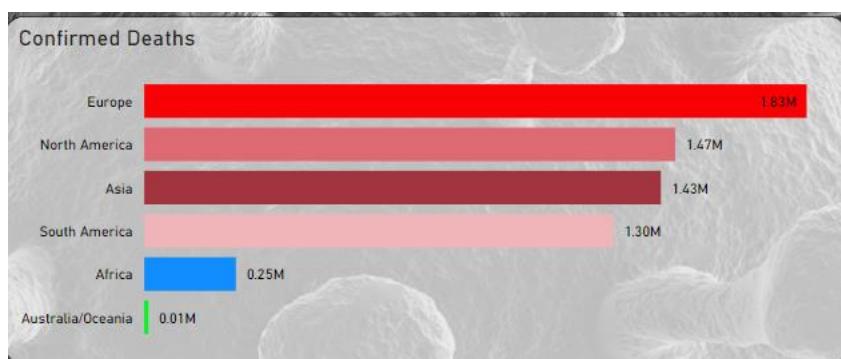


Figure 6: Distribution of covid deaths across Continents

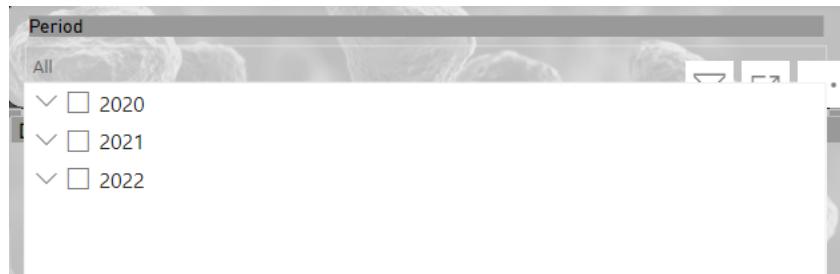


Figure 7: Period Slicer

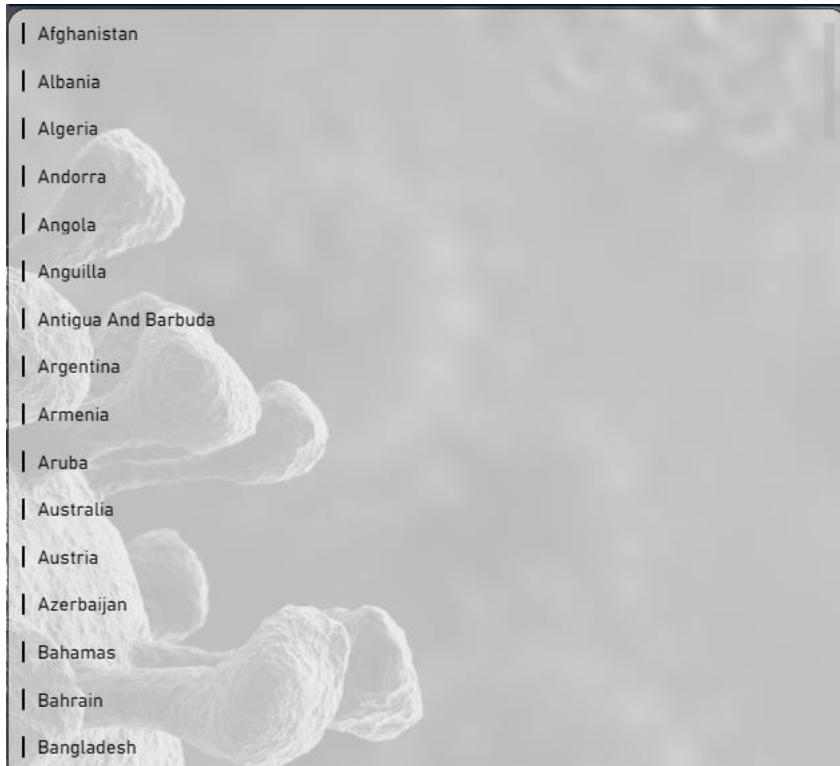


Figure 8: Card showing List of Countries

The Country Situation dashboard was created to visualise the result of analysis carried out on the data to answer question three. The visualisation contains a table, a donut chart, and a funnel chart. The table was used to show vital information for each country with regards to the corona virus. The table contains columns which houses information such as country, total infected people, total deaths, infection rate of the country, the mortality rate of covid in the country, the recovery rate of people infected from covid. The table also comes with an identification colour to show countries whose infection rate are above the global average infection rate (red) and countries who are below as well (green). The donut chart visualisation shows the percentage of countries spread according to their infection risk category. From the visualisation, we can infer that only about 19.5 percent of countries in the world are in the low-risk category (between 0-15%) with the largest risk category being countries that have an infection rate between 30 – 50 percent (46.59). The funnel chart visualises the top 10 countries with the highest infection rate of covid. From the chart, Faroe Island has the highest rate of infection with a 70.43% infection rate and Saint Barthelemy at the 10th position with an infection rate of 46.42%.

country	Total_Infected	total_deaths	Infection_rate	mortality_rate	Recovered	Active
Afghanistan	179290	7685	0.44%	0.02%	91000	81405
Albania	275559	3496	9.60%	0.12%	240000	41603
Algeria	265815	6875	0.59%	0.02%	240000	39110
Andorra	42155	153	54.40%	0.20%	35000	6652
Angola	99187	1898	0.29%	0.01%	90000	88411
Anguilla	2644	8	19.56%	0.06%	2500	2636
Antigua And Barbuda	7702	136	7.76%	0.14%	7000	6966
Argentina	9101310	128728	19.80%	0.28%	850000	802402
Armenia	422895	8623	14.22%	0.29%	350000	335672
Aruba	35600	212	33.17%	0.20%	30000	33178
Australia	6551609	7773	25.31%	0.03%	550000	602396
Austria	4209155	18303	46.28%	0.20%	350000	376952
Azerbaijan	792637	9709	7.69%	0.09%	700000	704928
Bahamas	33850	800	8.46%	0.20%	30000	33050
Bahrain	576992	1479	31.85%	0.08%	50000	56313
Bangladesh	1952965	29126	1.16%	0.02%	180000	172169
Barbados	76205	432	26.48%	0.15%	60000	71773
Belarus	982866	6978	10.41%	0.07%	80000	91308
Belgium	4116396	31613	35.23%	0.27%	350000	375583
Belize	57889	675	14.08%	0.16%	50000	52214
Benin	26926	162	0.21%	0.00%	25000	26764
Bermuda	14096	133	22.85%	0.22%	13000	12963
Brunei	50200	00	0.52%	0.00%	45000	5200
Total	520780007	6287453				

Figure 9: Table showing Countries with their Infection rates

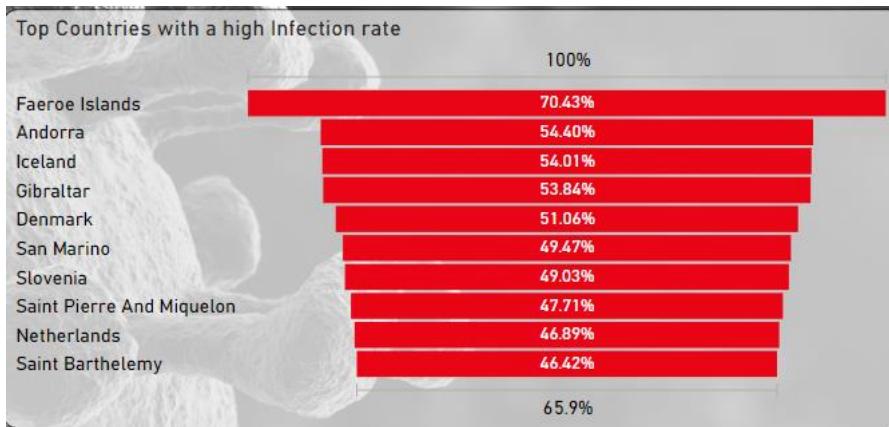


Figure 10: Top Countries with the highest infection rate

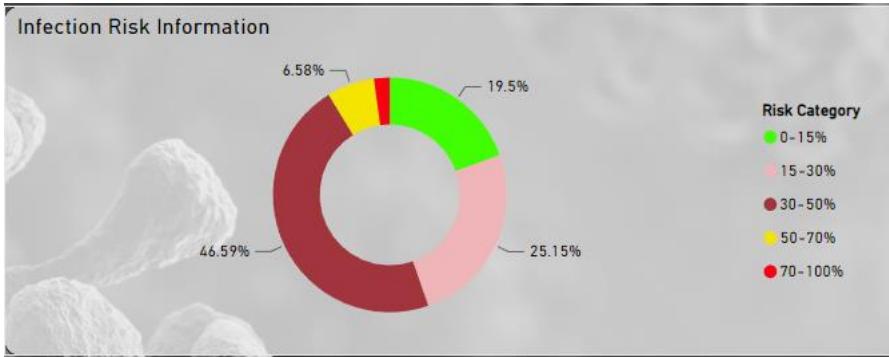


Figure 11: Distribution of Countries across Infection risk Categories

The Country-specific situation dashboard was created to visualise results of analysis carried out to daily count distribution of infection and deaths due to the virus in Nigeria since outbreak and identify the periods in which there were decline in the infection number in Nigeria. The visualisation contains a multi row card which gives shows the population of Nigeria (about 215 million), the total number of confirmed infection cases (over 250,000), the total number of deaths due to covid (3,143), the total covid tests done (about 5 million) and the infection rate of covid in Nigeria which stands at 0.12%. There are also two area charts which shows the daily reported infection cases and the daily confirmed covid deaths across the country respectively since the outbreak of corona virus. From the area charts, we can infer that the highest daily infection number in Nigeria was between December 2021 and January 2022. The dashboard also contains a key influencer chart which shows the periods which the average daily number of confirmed infection cases in Nigeria has reduced. From the visualisation, the daily infection cases are likely to reduce during the months of March, April, May, October, and November. Lastly, for easy navigation, there is a Timeline 2.4.0 slicer which is of advantage should you want information specific to a range of period since the outbreak of Covid in Nigeria.

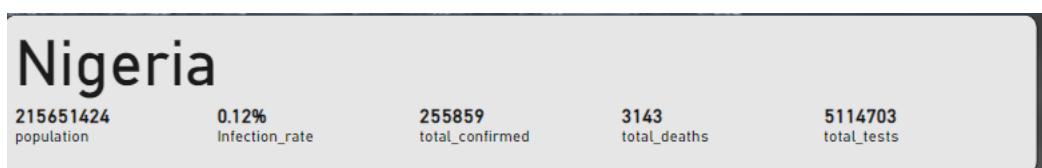


Figure 12: Overview of Covid infection in Nigeria

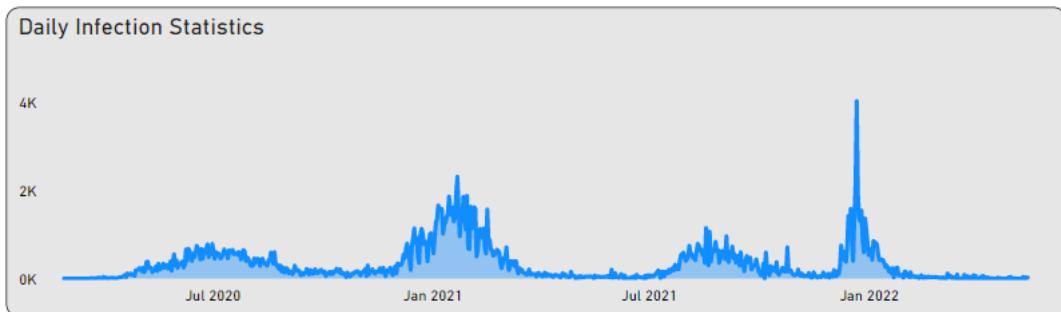


Figure 13: Daily Covid Infection in Nigeria since outbreak

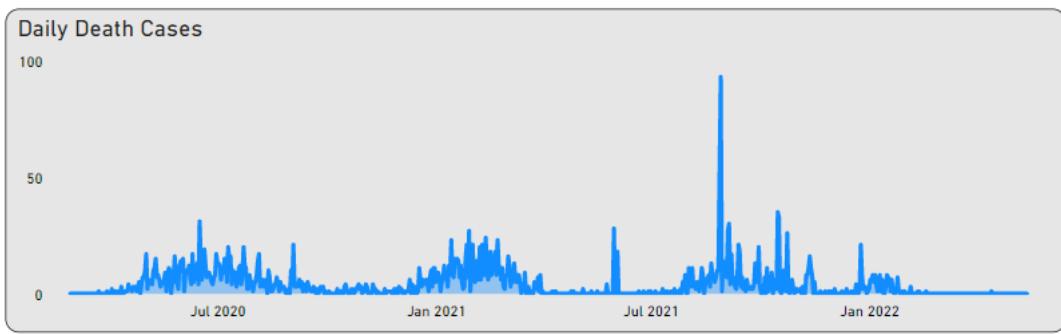


Figure 14: Daily Covid deaths in Nigeria since outbreak

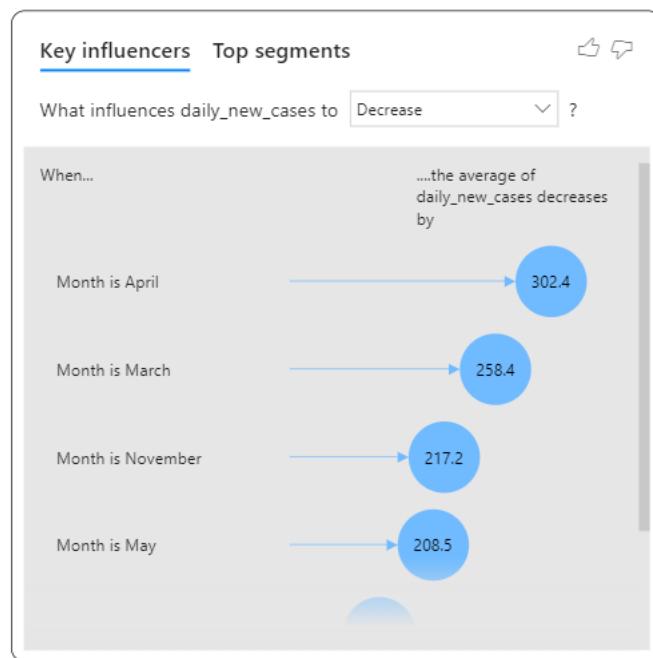


Figure 15: Period when there is Decline in Daily infection

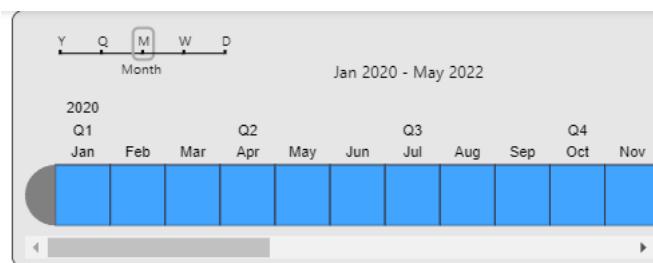


Figure 16: Timeline Slicer

The Information Summary dashboard was created to help users gain insights to the graphs and charts used in other dashboards. This dashboard was created using the smart narration visualisation to get information on every dashboard. From the visual, the longest period of rise in infection cases was from January 2020 to July 2020 with the longest drop in deaths being from August 2021 to May 2022. The longest period of drop in infection cases in Nigeria was from January 9, 2022, lasting 125 days.

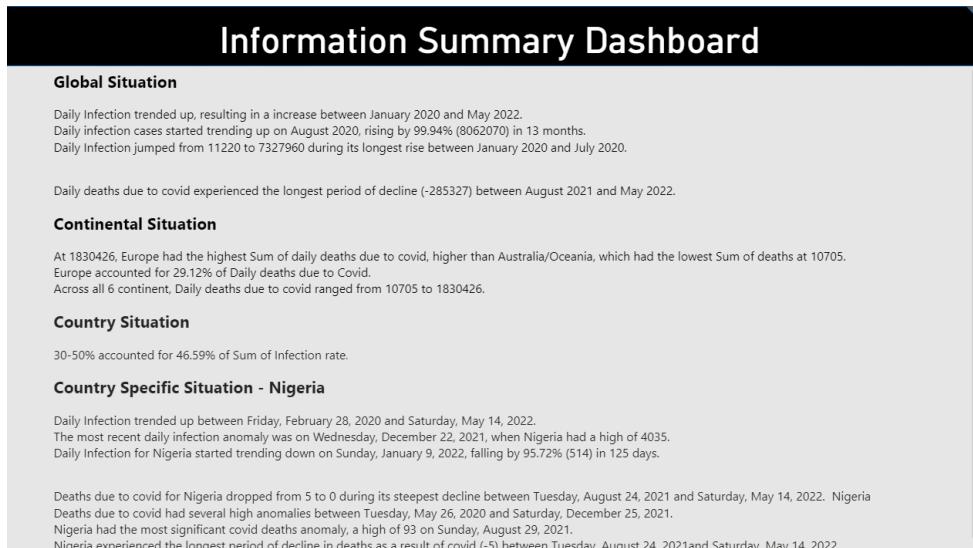


Figure 17: Smart Narration of Dashboards

Recommendations

Based on the findings from the visualisations created. Below are my recommendations.

- While good work has been done so far to reduce the number of deaths caused by covid, there is more work to be done to reduce the number of daily infections of covid as it has been increasing with every year. This would also have a positive effect on the global infection rate.
- In Europe, Asia and North America, there is need to intensify measures to reduce the number of covid infections and consequently deaths as a result as well as the statistics in these continents are higher compared to the remaining continents.
- More efforts need to be added towards improving the distribution percentage of countries below 15% infection risk category by reducing the spread across the other infection risk especially countries between 30-50 percent infection rate.
- Strict measures need to be put in place to effectively reduce the infection rate in the top 10 countries with the highest infection risk.
- In Nigeria, more tests should be carried out per million in its population as analysis suggests a high number of the population's infection is not being confirmed and reflected.
- More research and analysis need to be done to understand why there is a decrease in daily infection in the months of March, April, May, October, and November and how this can be applied to other months of the year.

Conclusions

I use Excel for analysis and reports before this module and only became aware of power BI through this module. However, I was fascinated by the extensive amount of analysis I can do with Power BI and was mostly impressed by features such as data pre-processing which makes it easier for me to streamline data to be precise and accurate; data modelling which allows me to group my data in smaller sets and still have them connected; the extensive library of visuals and the integration of AI and machine learning tools which are particularly

interesting to use. Finally, the issue I had creating this report mostly had to deal with the delay and lagging at random intervals by power BI when processing the data however I have not been able to determine if this is based on one's personal computer specs. Overall, power BI is a tool with many advantages especially with regards to data analysis.

Appendix: BI Design

Data Pre-Processing or Data Cleansing

- **Data Loading**

The pre-processing stage of the COVID dataset begins with importing of the dataset into the Microsoft power BI application. This is done using the 'Get Data' menu found on the '**Home**' tab of the application to select the file and import. The pictorial steps are shown below

1. Click on '**Get Data**'

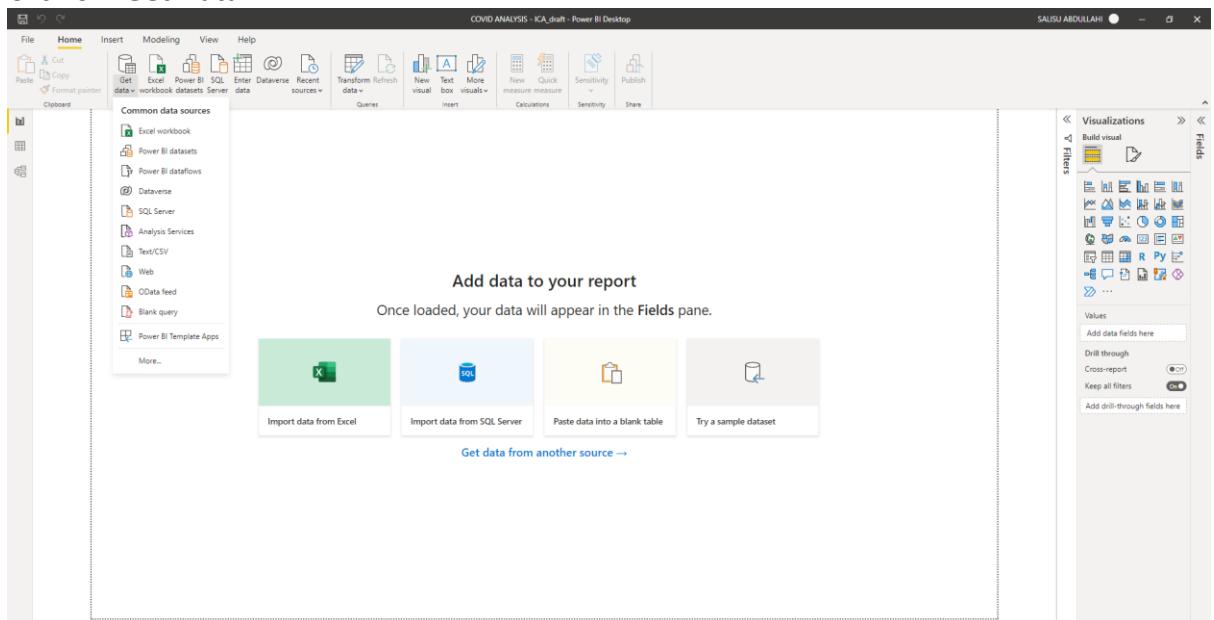


Figure 1.0: 'Get Data' process

2. Select appropriate data format. '**Text/CSV**' was selected as dataset is in '**Microsoft Excel Comma Separated values**' format.

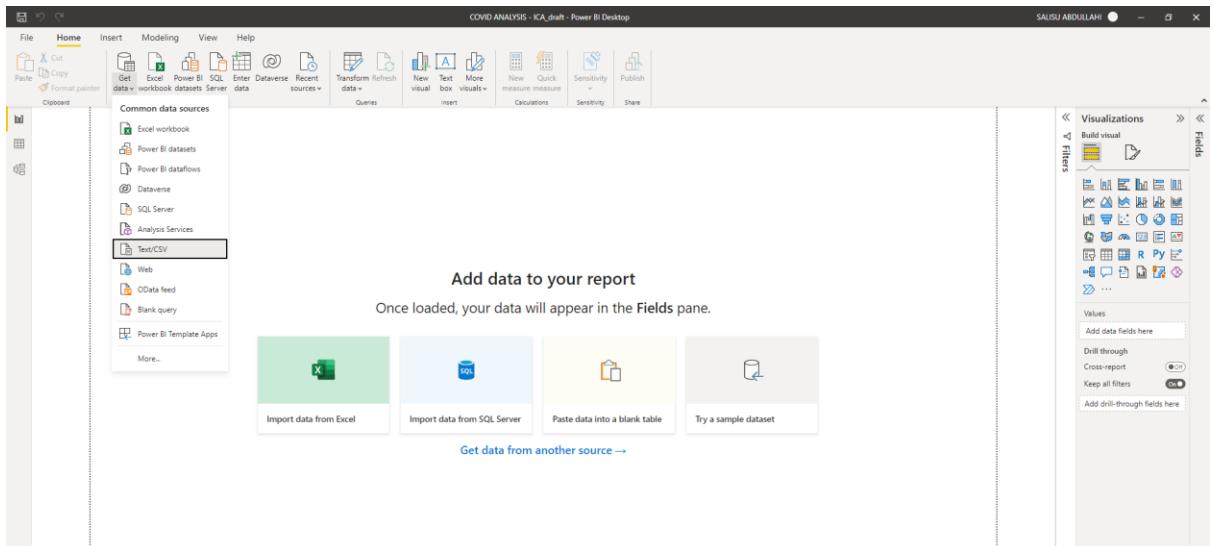


Figure 1.1: 'Text/CSV' selection

3. Select dataset files from the directory it was saved on the system and click on open.

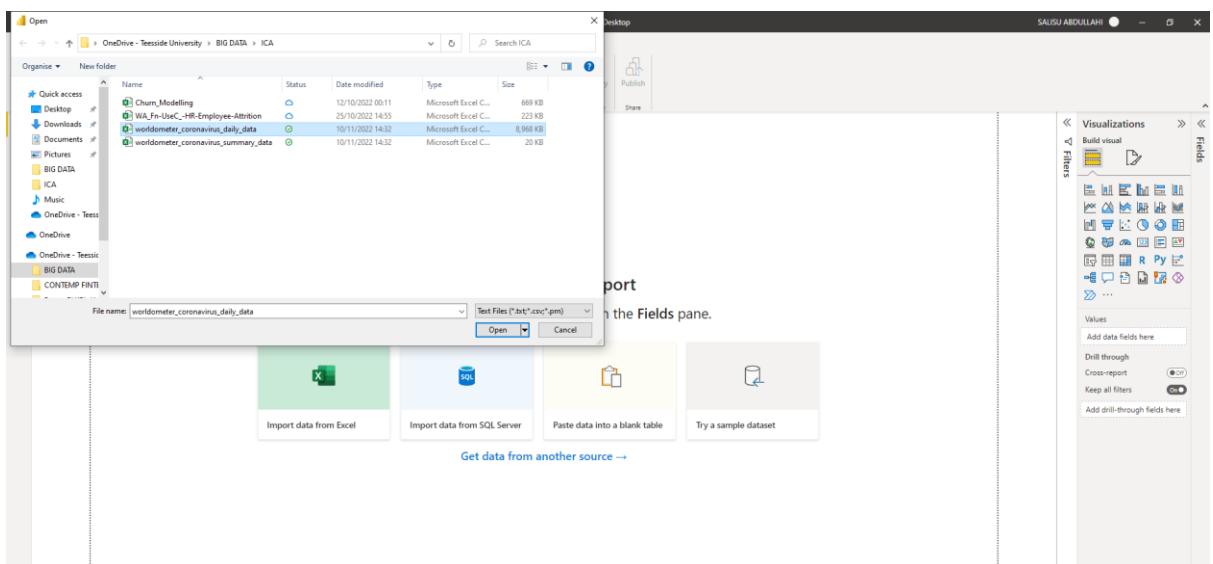


Figure 1.2: File selection

4. After selecting the file, a dialogue box opens showing a preview of the dataset selected. This also shows the file origin, the delimiter which the file is separated by (commas as file is a CSV file) and the basis on which the data type was detected.

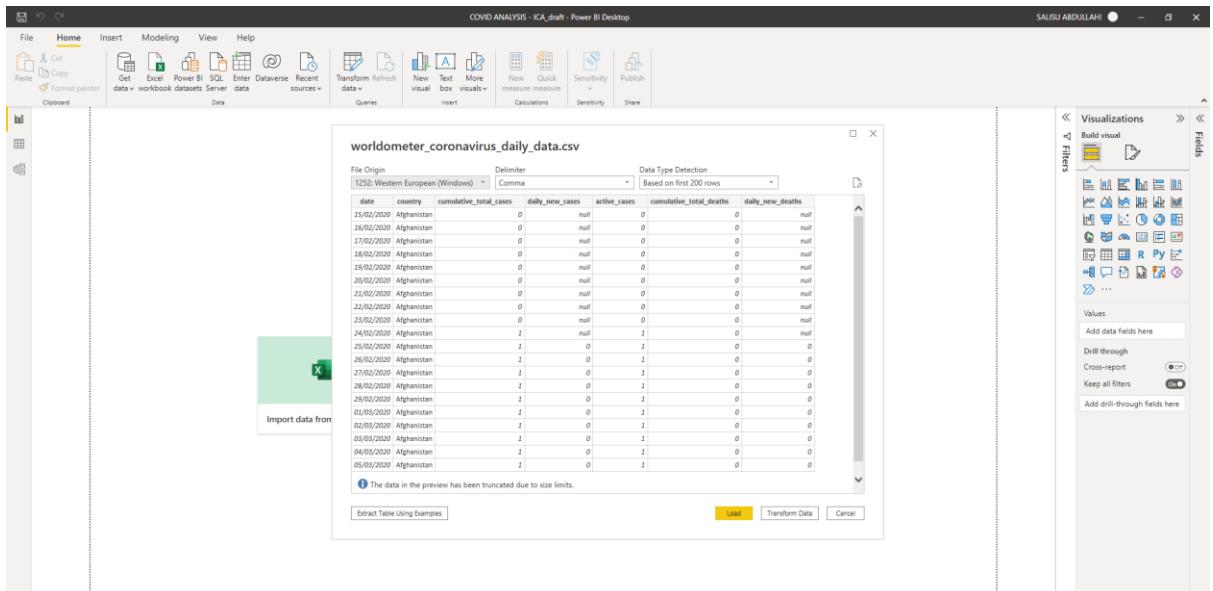


Figure 1.3.0: Dataset Preview

- Upon review of the dialogue box to further confirm that right dataset was selected, the load option was selected.

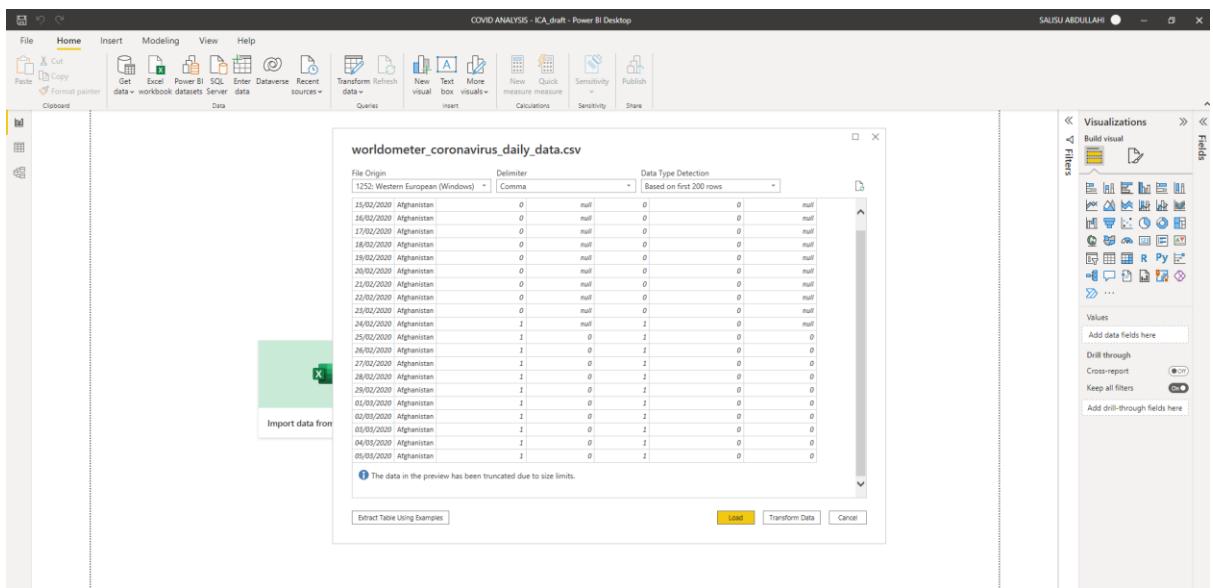


Figure 1.3.1: Dataset Preview

- Steps 2 to 5 was repeated to select the second dataset file to be used after which we have below page as shown. The imported datasets can be found under the fields menu as shown below.

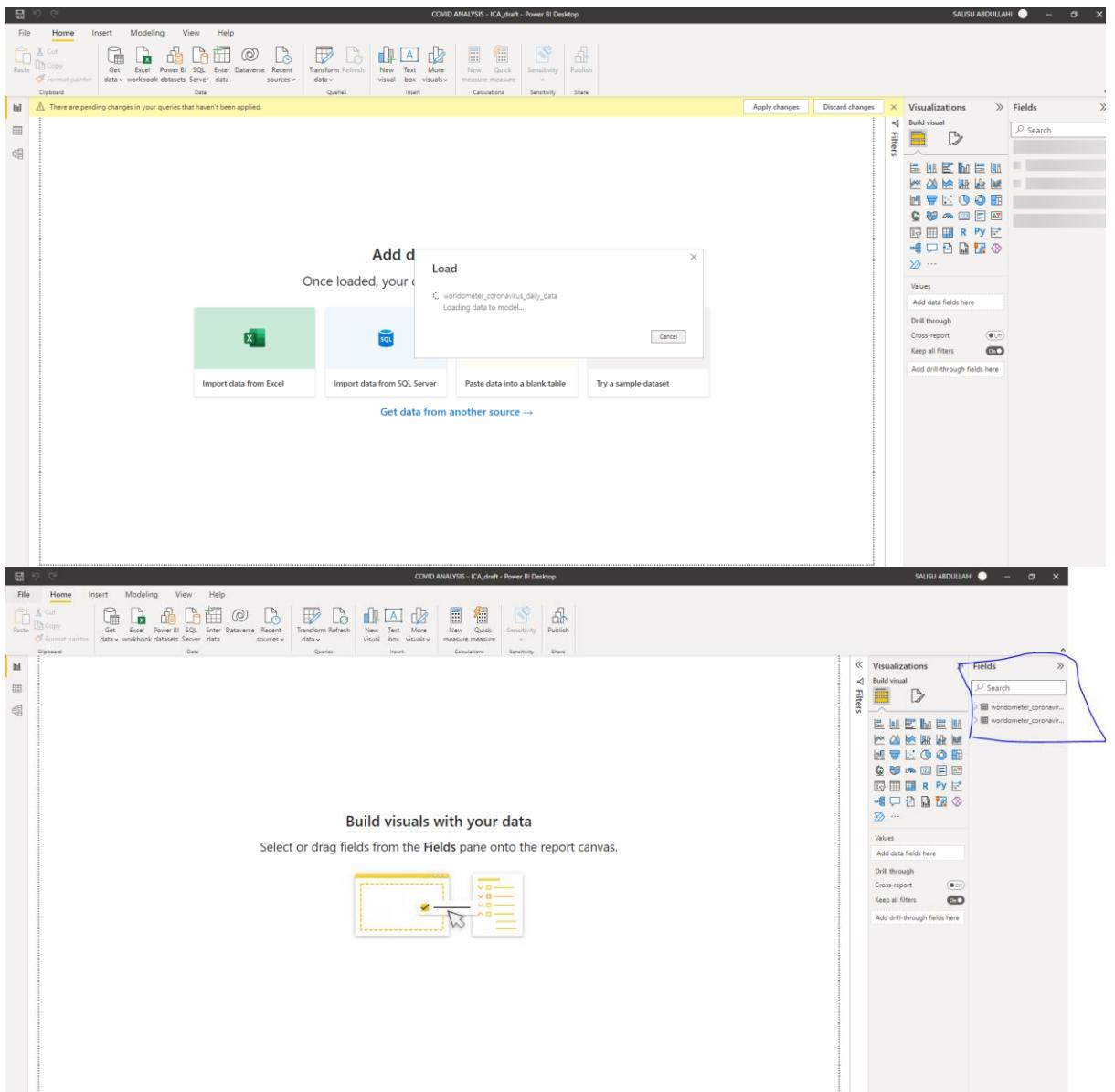


Figure 1.4: Data loaded into Power BI

- **Data Cleaning**

After importation of the dataset, the dataset was then cleaned. This process allows for me to identify errors and fix them as well as remove duplicated data and columns that are not important to the analysis to be done.

Data Cleaning underwent the following procedures.

1. The **Transform data** icon was selected from the query group to open the power query editor

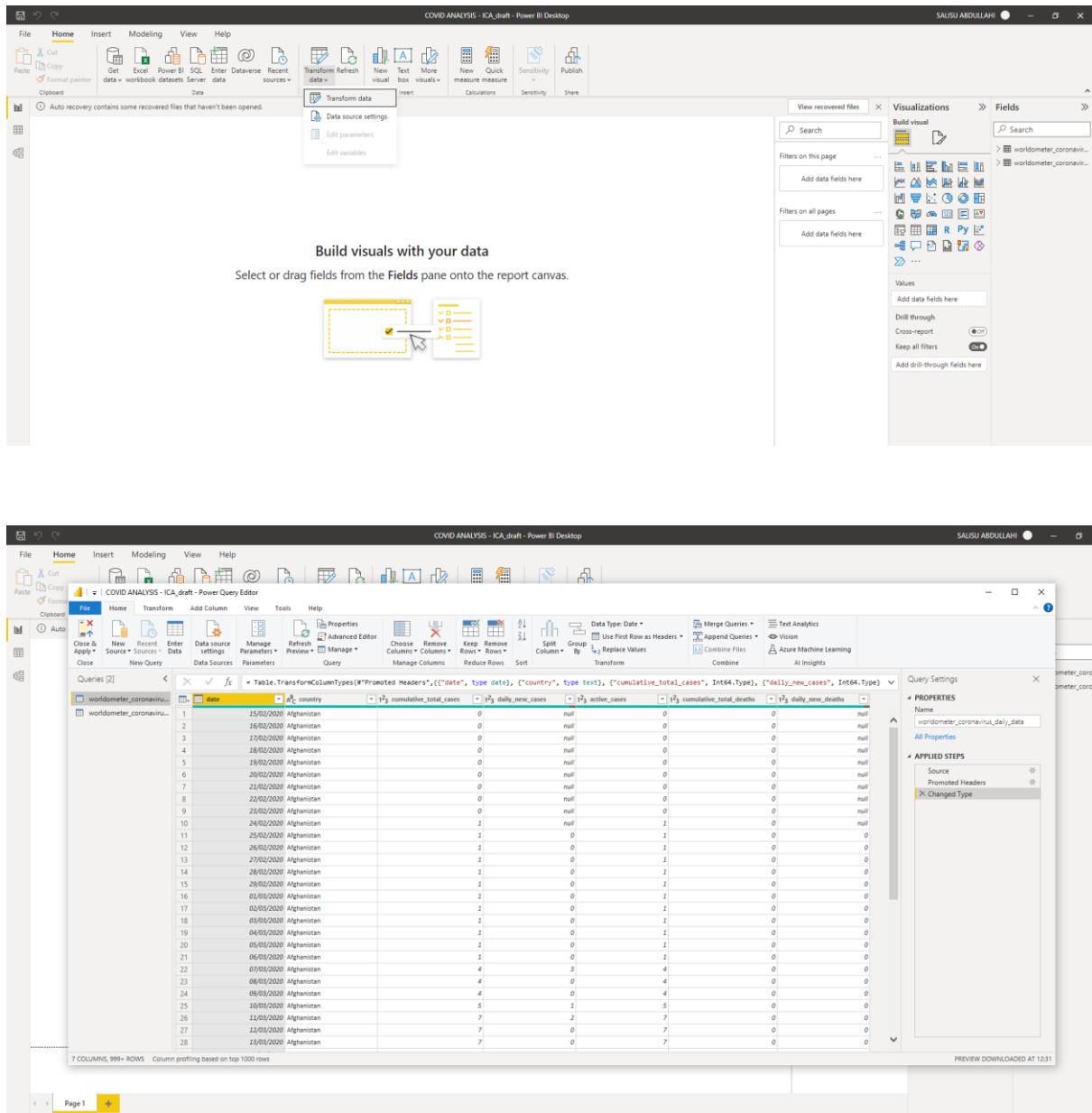


Figure 2.0: 'Transform Data' selection

2. The next step was to identify blank, null or error data from our columns. To do this, I first used the '**Column Quality**' under the View Menu bar to identify the percentage of data under each column which can be identified as blank, null or error. The output of this procedure is shown below. The aim of this procedure was to decide the best way to deal with blank, null, and erroneous data. If blank, null, or erroneous data in a column was less than 10 percent of the data, data will be dealt by elimination; if blank data is 10 percent or more than 10 percent, data will be dealt by identification.

Figure 2.1: 'Column Quality' selection

- Based on the column quality, I decided to eliminate the blank data in each column of my data. To do this, I first selected all columns, then clicked on the dropdown icon beside the **daily_new_cases** column and selected the '**remove empty**' filter option. This removed the null values in my data and gave resulting column quality as shown below. Negative figures in the column were filtered out too. This was done by repeating same step but instead of selecting '**remove empty**', I unticked the negative values off the list.

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Layout** ribbon tab selected.
- Query Settings** pane on the left shows the current query is named "daily_data".
- Data Preview** pane shows a table with columns: date_ID, country, daily_new_cases, and daily_new_deaths. A context menu is open over the "daily_new_cases" column, specifically targeting the first row which contains a null value.
- Applied Steps** pane on the right lists the steps taken so far, including "Removed Columns".
- Properties** pane on the right shows the "Name" is set to "daily_data".
- Preview** at the bottom right indicates the preview was downloaded at 15:07.

Figure 2.2: Removal of null values

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Layout** ribbon tab selected.
- Query Settings** pane on the left shows the current query is named "daily_data".
- Data Preview** pane shows a table with columns: date_ID, country, daily_new_cases, and daily_new_deaths. A context menu is open over the "daily_new_cases" column, specifically targeting the first row which contains a negative value (-39).
- Applied Steps** pane on the right lists the steps taken so far, including "Removed Columns".
- Properties** pane on the right shows the "Name" is set to "daily_data".
- Preview** at the bottom right indicates the preview was downloaded at 15:07.

Figure 2.3: Removal of negative values

The screenshot shows the Microsoft Power Query Editor interface. The main area displays a table with columns: date, #country, cumulative_total_cases, daily_new_cases, active_cases, cumulative_total_deaths, and daily_new_deaths. The 'Query Settings' pane on the right shows the step 'Filtered Rows' under 'APPLIED STEPS'. The 'Properties' pane shows the query name as 'worldometer_coronavirus_daily_data'. The status bar at the bottom right indicates 'PREVIEW DOWNLOADED AT 13:19'.

Figure 2.4: Result of negative and null value removal on daily_new_cases column

- Step 3 was repeated, this time filter was applied on the **daily_new_death** column. Result shown below.

The screenshot shows the Microsoft Power Query Editor interface. The main area displays a table with columns: date_ID, #country, daily_new_cases, and daily_new_deaths. A 'Number Filters' dialog box is open, showing a list of values from -1 to 10. The 'Query Settings' pane on the right shows the step 'Filtered Rows2' under 'APPLIED STEPS'. The 'Properties' pane shows the query name as 'daily_data'. The status bar at the bottom right indicates 'PREVIEW DOWNLOADED AT 19:00'.

Figure 2.5: Removal of negative values

Figure 2.6: Result of negative and null value removal on daily_new_deaths column

5. I am now done cleaning the first data table (worldometer_coronavirus_daily_data) which I then renamed to **daily_data** table. Steps 2 to 4 was also applied to the second table (worldometer_coronavirus_summary_data) table which I renamed to **summary_data**. Worthy to note that the second table contained a column (serious_or_critical) with 10% more than 10 percent (36) of empty values in that column blank. Thus, before performing steps 2 to 4, I dealt with the missing values by identification, replacing null values in the column with 0. To do this, I selected the serious_or_critical column and then right clicked and selected the ‘replace value’ option to open the replace value dialogue box. On the box, I inputted ‘null’ in the ‘value to find’ space and ‘0’ in the ‘replace with’ field. This process replaced all the null values with 0. Alternatively, the replace value option can be found on the **Home ribbon**. This was also applied to other columns with 10 percent empty column.

Figure 2.7.1: Dealing with missing values by identification

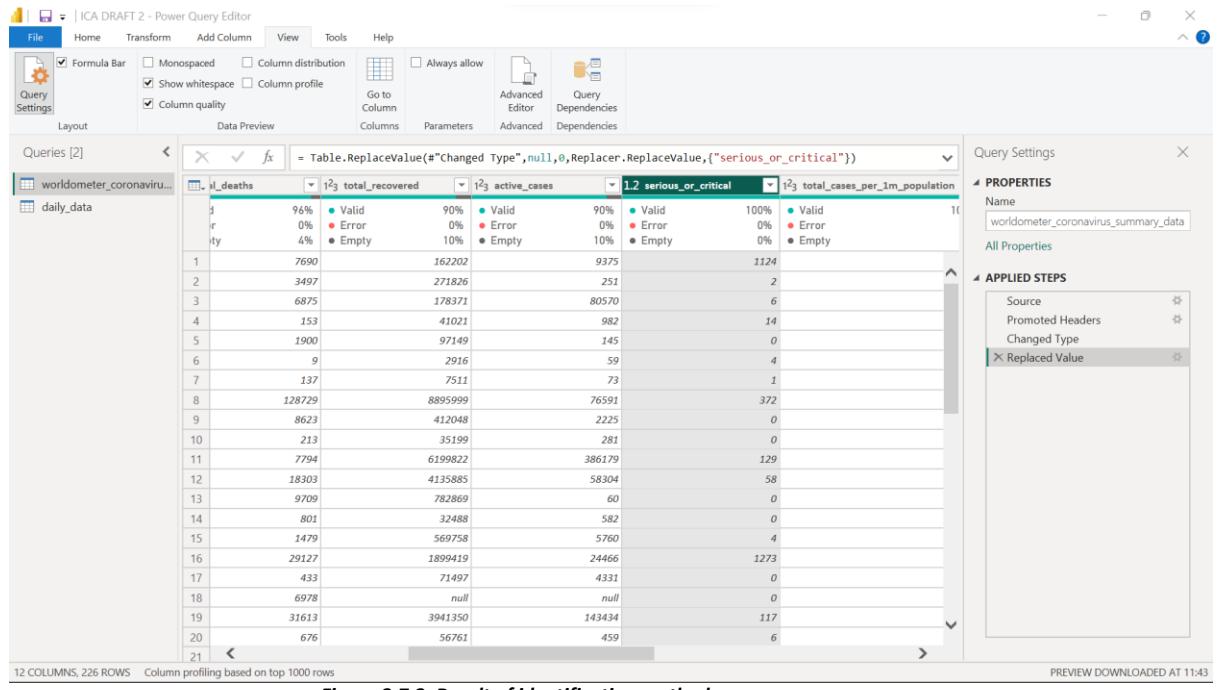


Figure 2.7.2: Result of identification method

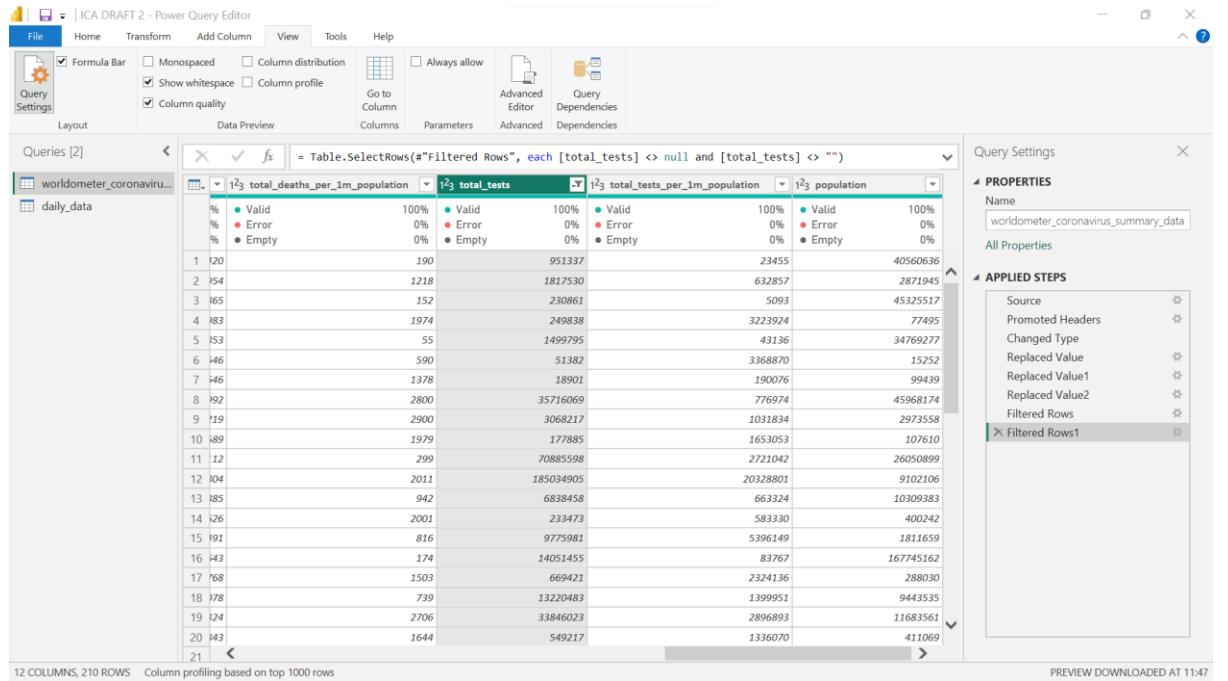


Figure 2.8.1: Renaming the datasets

Figure 2.8.2: screenshot of 'transform data' query showing renamed Datasets

- Now that I am done with cleansing for both data tables, we select the **close and apply** icon to apply all the changes made to the dataset table.

BI Data Modelling via Star Schema - Facts and Dimensions.

- Before dataset modelling begun, the data model state look was shown below. Existing relationship between the two tables was deleted as well. To achieve this, the relationship line was right-clicked and the 'delete' option was selected. This action can also be done by selecting the manage relationship icon from the home ribbon, ticking the relationship to be deleted from the dialogue box that opens and selecting delete.

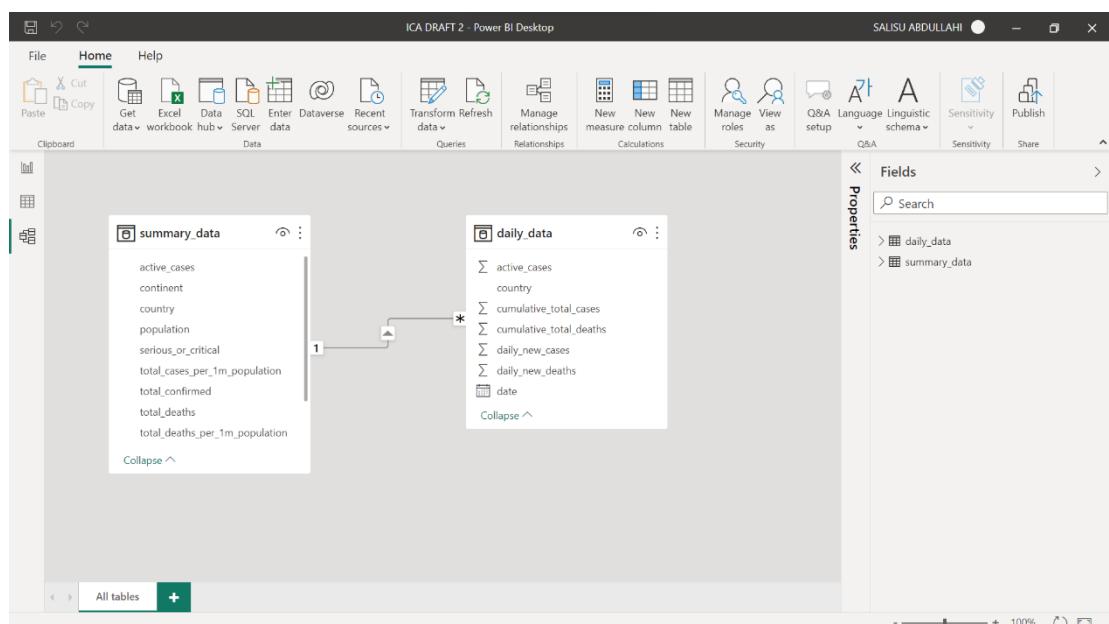
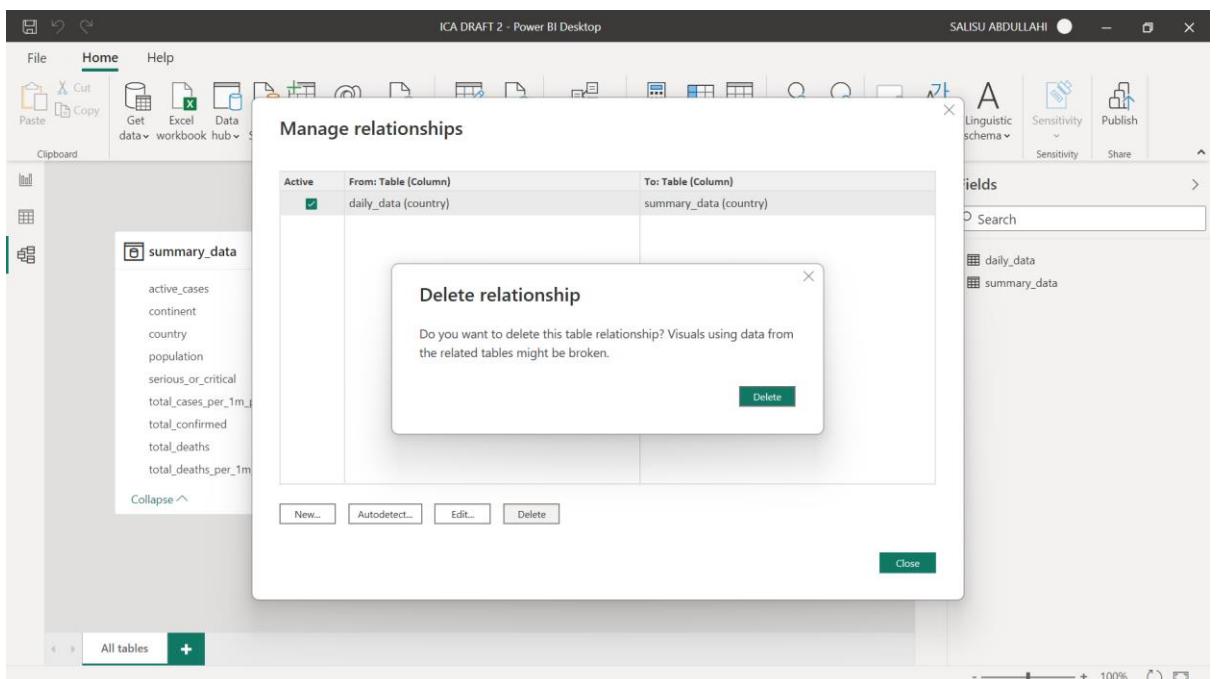
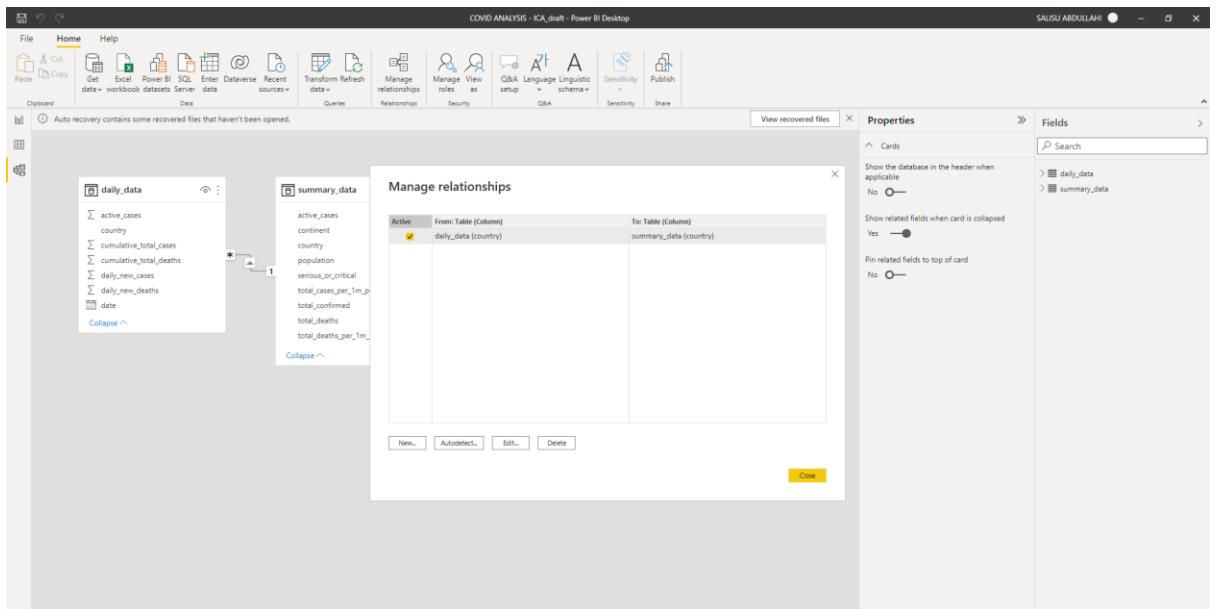


Figure 3.0: screenshot of existing data model and relationship



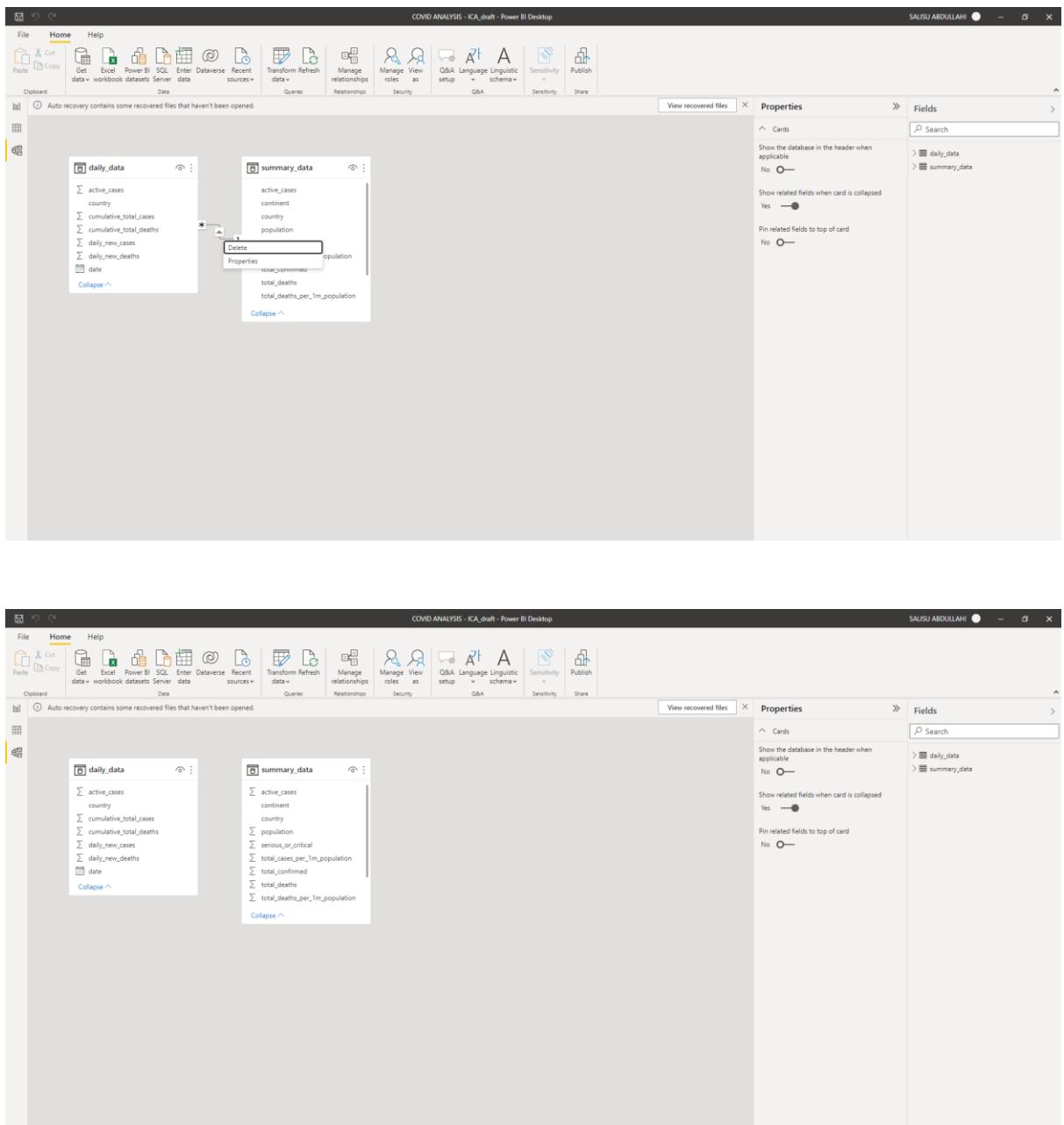


Figure 3.1: screenshots showing data model relationship removal process

- When creating my data model for this analysis, I had to create dimension tables and fact tables. The aim of this is to de-congest and simplify my data table and make processing of my analysis easier. Dimension tables to be created include data that are contextual to the fact table.

Dimension Tables

The first dimension table to be created was the **Continent** table. To do this, the first step was to click on transform data, then right click on the **summary_data** table and select duplicate. Then rename the duplicated table to **Continent**.

COVID ANALYSIS - ICA_draft - Power Query Editor

Queries [3]

Table.Distinct(#'Filtered Rows2')

country continent total_confirmed total_deaths total_recovered active_cases serious_or_critical total_death_rate

1 Afghanistan Asia 179267 7690 162202 9375 1124

2 Albania Europe 275574 3497 271826 251 2

3 Algeria Africa 265816 6875 178871 80570 6

4 Andorra Europe 42156 153 41021 982 14

5 Angola Africa 99194 1900 97149 145 -1

6 Anguilla North America 2984 9 2916 59 4

7 Antigua And Barbuda North America 7721 137 7511 73 1

8 Argentina South America 910319 128729 8895999 76591 372

9 Armenia Asia 422896 8623 412048 2225 -1

10 Aruba North America 35693 213 35199 281 -1

11 Australia Australia/Oceania 6593795 7794 6199822 386179 129

12 Austria Europe 4212492 18303 4135885 58304 58

13 Azerbaijan Asia 792638 9709 782869 60 -1

14 Bahamas North America 33871 801 32488 582 -1

15 Bahrain Asia 579997 1479 569758 5760 4

16 Bangladesh Asia 1953012 29127 1899419 24466 1273

17 Barbados North America 76261 433 71497 4331 -1

18 Belgium Europe 4116397 31613 3941550 142434 117

19 Belize North America 57896 676 56761 459 6

20 Benin Africa 26952 163 25506 1283 5

21 Bermuda North America 14155 135 13664 336 -1

22 Bhutan Asia 59570 21 59491 58 -1

23 Bolivia South America 906146 21943 863665 20538 220

24 Botswana Africa 306324 2690 303026 608 1

25 Brazil South America 30682094 664920 29718402 298772 8318

26 Brunei Darussalam Asia 146659 220 142788 1651 -1

27 Bulgaria Europe 1161304 37045 1014302 110157 54

28 Burundi Africa 59570 21 59491 58 -1

12 COLUMNS, 191 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 14:40

Figure 3.2: screenshot showing duplicated data table

COVID ANALYSIS - ICA_draft - Power Query Editor

Queries [3]

Table.Distinct(#'Filtered Rows2')

Continent total_confirmed total_deaths total_recovered active_cases serious_or_critical total_death_rate

1 Afghanistan Asia 179267 7690 162202 9375 1124

2 Albania Europe 275574 3497 271826 251 2

3 Algeria Africa 265816 6875 178871 80570 6

4 Andorra Europe 42156 153 41021 982 14

5 Angola Africa 99194 1900 97149 145 -1

6 Anguilla North America 2984 9 2916 59 4

7 Antigua And Barbuda North America 7721 137 7511 73 1

8 Argentina South America 910319 128729 8895999 76591 372

9 Armenia Asia 422896 8623 412048 2225 -1

10 Aruba North America 35693 213 35199 281 -1

11 Australia Australia/Oceania 6593795 7794 6199822 386179 129

12 Austria Europe 4212492 18303 4135885 58304 58

13 Azerbaijan Asia 792638 9709 782869 60 -1

14 Bahamas North America 33871 801 32488 582 -1

15 Bahrain Asia 579997 1479 569758 5760 4

16 Bangladesh Asia 1953012 29127 1899419 24466 1273

17 Barbados North America 76261 433 71497 4331 -1

18 Belgium Europe 4116397 31613 3941550 142434 117

19 Belize North America 57896 676 56761 459 6

20 Benin Africa 26952 163 25506 1283 5

21 Bermuda North America 14155 135 13664 336 -1

22 Bhutan Asia 59570 21 59491 58 -1

23 Bolivia South America 906146 21943 863665 20538 220

24 Botswana Africa 306324 2690 303026 608 1

25 Brazil South America 30682094 664920 29718402 298772 8318

26 Brunei Darussalam Asia 146659 220 142788 1651 -1

27 Bulgaria Europe 1161304 37045 1014302 110157 54

28 Burundi Africa 59570 21 59491 58 -1

12 COLUMNS, 191 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 14:38

Figure 3.3: screenshot showing renamed data table

After renaming, the country and continent columns were selected. This is done by holding the control key and selecting both columns. Then other columns in the table were removed by right clicking on any of the selected columns and selecting the remove other columns option. I then remove duplicates from the data on the table by selecting the country column, right clicking on it and selecting remove duplicate.

The screenshot shows the Power Query Editor interface with the following details:

- File**: COVID ANALYSIS - ICA_draft - Power Query Editor
- Queries [3]** pane: daily_data, summary_data, Continent
- Table View**: A table titled "Table.Distinct(#"Filtered Rows2")" with columns: country, continent, total_confirmed, total_deaths, total_recovered, active_cases, serious_or_critical, total_death_rate. The data includes rows for countries like Afghanistan, Albania, Algeria, Andorra, Angola, Anguilla, Antigua And Barbuda, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belgium, Belize, Benin, Bermuda, Bhutan, Bolivia, Botswana, Brazil, Brunei Darussalam, Bulgaria, Burkina Faso, and so on.
- Applied Steps** pane: Shows the step "Removed Duplicates".
- Query Settings** pane: Shows the step "Continent".
- Bottom Status Bar**: 12 COLUMNS, 191 ROWS, Column profiling based on top 1000 rows, PREVIEW DOWNLOADED AT 1440.

Figure 3.4: screenshot showing data duplicate removal

The screenshot shows the Power Query Editor interface with the following details:

- File**: COVID ANALYSIS - ICA_draft - Power Query Editor
- Queries [3]** pane: daily_data, summary_data, Continent
- Table View**: A table titled "Table.SelectColumns(#"Removed Duplicates",{"country", "continent"})" with columns: country, continent. The data includes rows for countries like Afghanistan, Albania, Algeria, Andorra, Angola, Anguilla, Antigua And Barbuda, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belgium, Belize, Benin, Bermuda, Bhutan, Bolivia, Botswana, Brazil, Brunei Darussalam, Bulgaria, Burkina Faso, and so on.
- Applied Steps** pane: Shows the steps "Removed Duplicates" and "Removed Other Columns".
- Query Settings** pane: Shows the step "Continent".
- Bottom Status Bar**: 2 COLUMNS, 191 ROWS, Column profiling based on top 1000 rows, PREVIEW DOWNLOADED AT 1445.

Figure 3.5: screenshot showing other column removal

We have successfully created the continent dimension table. We delete the continent column from the summary_data table. To do this, the continent column was selected, right clicked and the option 'remove column' was selected.

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Queries [3]** pane: `daily_data`, `summary_data`, `Continent`.
- Table** view: A list of countries with columns: `#1 continent`, `#2 total_confirmed`, `#3 total_deaths`, `#4 total_recovered`, `#5 active_cases`, `#6 serious_or_critical`, and `#7 total_cases_per_1m_population`. The `continent` column is highlighted in yellow.
- Applied Steps** pane: Shows the step `Removed Duplicates`.
- Query Settings** pane: Shows the step `summary_data`.
- PREVIEW DOWNLOADED AT 14:18**

The screenshot shows the Power Query Editor interface after the 'Continent' column has been removed:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Queries [3]** pane: `daily_data`, `summary_data`, `Continent`.
- Table** view: A list of countries with columns: `#1 total_confirmed`, `#2 total_deaths`, `#3 total_recovered`, `#4 active_cases`, `#5 serious_or_critical`, and `#7 total_cases_per_1m_population`.
- Applied Steps** pane: Shows the steps `Removed Duplicates` and `Removed Columns`.
- Query Settings** pane: Shows the step `summary_data`.
- PREVIEW DOWNLOADED AT 14:49**

Figure 3.6: screenshots showing continent column removal

Next, I created a population_data dimension table. To do this, the previous dimension table steps were repeated but this time the columns selected were ‘Population’, ‘total_cases_per_1m_population’, ‘total_deaths_per_1m_population’, ‘total_tests_per_1m_population’ and ‘Country’.

COVID ANALYSIS - ICA_draft - Power Query Editor

File Home Transform Add Column View Tools Help

Queries [4]

summary_data (2)

Table.RemoveDuplicates("Removed Duplicates", "continent")

country	total_confirmed	total_deaths	total_recovered	active_cases	serious_or_critical	total_cases_per_1m_population
Afghanistan	179267	7690	162302	8075	1124	44
Albania	275574	3497	271826	251	2	98
Algeria	265816	6875	178371	80570	6	54
Andorra	42156	153	419221	982	14	5435
Angola	99194	1900	97249	145	4	28
Anguilla	2984	9	2916	59	4	1952
Anguilla And Barbuda	7721	137	7511	73	1	77
Argentina	9101319	128729	8895999	76591	372	1975
Armenia	422896	8623	412048	2225	-1	142
Aruba	35693	213	35199	281	-1	3311
Australia	6595795	7794	6199822	586179	129	2531
Austria	4212492	18301	4135885	58104	58	4628
Azerbaijan	792638	9709	782869	60	-1	768
Bahamas	33871	801	32488	582	-1	846
Bahrain	576997	1479	569758	5760	4	3184
Bangladesh	1958012	29127	1896419	24466	1273	114
Barbados	76061	433	71497	4331	-1	2645
Belgium	4116397	31613	3941350	143434	117	3521
Belize	57896	676	56791	459	6	1408
Benin	26952	163	25506	1283	5	21
Bermuda	14115	135	13664	336	-1	2385
Bhutan	59670	21	59491	58	-1	756
Bolivia	906146	21943	863665	20538	220	751
Botswana	306324	2690	303026	608	1	1255
Brazil	30582094	664920	29718402	298772	8318	1424
Brunel Darussalam	144659	220	142788	1651	-1	3246
Bulgaria	1161504	37045	1014902	110157	54	1691
Burkina Faso	99447	9511	95001	91	-1	1111

11 COLUMNS, 191 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 14:55

COVID ANALYSIS - ICA_draft - Power Query Editor

File Home Transform Add Column View Tools Help

Queries [4]

summary_data (2)

Table.RemoveDuplicates("Removed Duplicates", "continent")

country	total_cases_per_1m_population	total_deaths_per_1m_population	total_tests	total_tests_per_1m_population	population
9375	1224	4426	196	951337	29455
251	2	95954	1218	1817530	632857
80570	6	5865	152	250861	5093
982	14	549883	1974	249888	3223924
145	-1	2853	55	149795	40562656
59	4	195466	590	51382	3368870
73	1	77646	1378	18901	15252
876591	372	197992	2800	35716069	776974
2225	-1	142219	2900	3068217	49439
281	-1	331689	1979	177885	45968174
386179	129	253112	299	70885598	2973558
58304	58	462804	2011	185054905	2973558
60	-1	76885	942	6838458	45968174
582	-1	84626	2001	233473	400242
5760	4	318491	816	9775981	1811659
24466	1273	11643	174	14051455	83767
4331	-1	264768	1503	669421	167745162
143434	117	353234	2706	33846023	2324136
459	6	140843	1644	549217	288030
1283	5	2120	13	604310	11683561
336	-1	228537	2183	891980	153615070
58	-1	75651	27	2303690	411069
20538	220	75704	1833	205542	61850
608	1	125541	1102	2036808	202025
206772	8318	142460	3087	63776166	11969554
1651	-1	324871	494	717784	25373503
110157	54	189536	5407	9924259	26050899
8	-1	74441	49	1448573	851060

11 COLUMNS, 191 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 14:55

COVID ANALYSIS - ICA_draft - Power Query Editor

File Home Transform Add Column View Tools Help

Queries [4]

Population

Table.SelectColumns("Removed Columns", "country", "population", "total_tests_per_1m_population", "total_deaths_per_1m_population", "total_cases_per_1m_population")

country	population	total_tests_per_1m_population	total_deaths_per_1m_population	total_cases_per_1m_population
Afghanistan	40560636	234655	180	4420
Albania	2871445	632857	1218	35654
Algeria	4532517	5069	152	5665
Andorra	77495	3223924	1974	54383
Angola	34769577	43136	55	2853
Anguilla	15252	3368870	590	195646
Anguilla And Barbuda	99439	190076	1378	77646
Argentina	4568174	776974	2800	19792
Armenia	2975558	1031834	2900	44219
Aruba	107610	1653053	1979	331689
Australia	26050899	2721042	299	253112
Austria	9102106	20328801	2011	462804
Azerbaijan	10309383	665324	942	76885
Bahamas	400242	583330	2001	84626
Bahrain	1811659	5396149	816	218491
Bangladesh	16745162	87867	174	11643
Barbados	288030	232436	1503	264768
Belgium	11683561	2896893	2706	352324
Belize	41069	1356070	1644	240843
Benin	12714841	47530	13	2120
Bermuda	61850	1445402	2183	228537
Bhutan	787432	2925573	27	75651
Bolivia	11969554	2260325	1833	228537
Botswana	2440037	830883	1102	225541
Brazil	215375303	296119	3087	342460
Brunel Darussalam	445281	1611980	494	242471
Bulgaria	6851060	1448573	5407	289536
Burkina Faso	21970296	11533	17	949

5 COLUMNS, 191 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 15:00

Figure 3.7: screenshots showing Population table creation

Upon creation of the population dimension table and deletion of 'total_cases_per_1m_population', 'total_deaths_per_1m_population' and 'total_tests_per_1m_population' columns from the summary_data table, the summary_data table looks like below.

The screenshot shows the Power Query Editor interface with the 'summary_data' query selected. The table contains 191 rows and 8 columns. The columns are: country, total_confirmed, total_deaths, total_recovered, active_cases, serious_or_critical, total_tests, and popus. The 'APPLIED STEPS' pane on the right shows the step 'Removed Columns' has been applied. The preview at the bottom indicates 1459 rows.

country	total_confirmed	total_deaths	total_recovered	active_cases	serious_or_critical	total_tests	popus
Afghanistan	179367	7690	162202	9375	1224	951357	
Albania	275574	3497	271826	251	2	1817530	
Algeria	265816	6875	178371	80570	6	230861	
Andorra	42159	153	41021	982	14	248638	
Angola	99194	1900	97149	145	-1	1499795	
Anguilla	2984	9	2916	59	4	51582	
Antigua And Barbuda	7721	137	7511	73	1	18801	
Argentina	8101319	128729	8895999	76591	372	35716069	
Armenia	422896	8623	412048	2225	-1	177885	
Aruba	35693	213	35199	281	-1	177885	
Australia	6593795	7794	6199822	3865179	129	70885598	
Austria	4212492	18303	4135885	58304	58	285034605	
Azerbaijan	792638	9709	782869	60	-1	6838458	
Bahamas	33871	801	32488	582	-1	238473	
Bahrain	576997	1479	569758	5760	4	9775981	
Bangladesh	1955012	29127	1899419	24468	1273	14051455	
Barbados	76261	433	71497	4331	-1	669421	
Belgium	4116397	31613	3941550	143434	117	33846023	
Belize	57896	676	56761	459	6	549217	
Benin	26952	163	25506	1283	5	604510	
Bermuda	14135	135	13664	336	-1	893980	
Bhutan	59570	21	59491	58	-1	2303690	
Bolivia	906146	21943	885665	20538	220	2705422	
Botswana	306324	2690	303026	608	1	2026988	
Bosnia	3058294	664620	26718402	268777	8118	63779566	
Brunel Darussalam	144659	220	142788	1651	-1	717784	
Bulgaria	1515104	37045	1014302	110157	54	9924259	
Burkina Faso	20853	382	20439	32	-1	246995	
Cabo Verde	56105	401	55613	91	23	400982	
Cambodia	186262	8056	133194	12	-1	2961594	
Cameroun	119780	1927	117791	62	13	1751774	
Canada	3822346	40228	3499564	281554	423	81436458	
Caribbean Netherlands	9915	35	9829	51	-1	30126	

Figure 3.8: screenshot of summary_data table after removal of population columns

Next, I created a date dimension table. This was done by duplicating the daily_data table, selecting the date column and right click. Select 'remove other column' to remove the other columns.

The screenshot shows the Power Query Editor interface with the 'daily_data' query selected. The table contains 999+ rows and 1 column, labeled 'date'. The 'APPLIED STEPS' pane on the right shows the step 'Removed Other Columns' has been applied. The preview at the bottom indicates 1623 rows.

date
25/02/2020
26/02/2020
27/02/2020
28/02/2020
29/02/2020
01/03/2020
02/03/2020
03/03/2020
04/03/2020
05/03/2020
06/03/2020
07/03/2020
08/03/2020
09/03/2020
10/03/2020
11/03/2020
12/03/2020
13/03/2020
14/03/2020
15/03/2020
16/03/2020
17/03/2020
18/03/2020
19/03/2020
20/03/2020
21/03/2020
22/03/2020
23/03/2020
24/03/2020
25/03/2020
26/03/2020
27/03/2020
28/03/2020

Figure 3.9.0: screenshot of date dimension table creation

Next step is to create an index for the date. This can be done by first removing duplicates of the date in your table. Right click on the date column and select remove duplicates. Next step is to sort the date in order of ascending. Click on the dropdown icon beside the date column and select ‘sort ascending’. I then clicked the index column dropdown and selected ‘from 1’ from the ‘add column ribbon’. Rename the table to ‘Date’.

The screenshot shows the Power Query Editor interface with the 'date' table selected. The 'APPLIED STEPS' pane on the right lists the 'Removed Duplicates1' step. The table itself contains 28 rows of dates from 25/02/2020 to 25/03/2020. The ribbon at the top has tabs for File, Home, Transform, Add Column, View, Tools, and Help. The 'Transform' tab is active, showing various data manipulation tools like Conditional Column, Index Column, Merge Columns, Extract, Parse, Statistics, Standard, Scientific, Rounding, Information, Date, Time, Duration, Text Analytics, Vision, Azure Machine Learning, and AI Insights.

Figure 3.9.1: screenshot showing date table after duplicate removal

This screenshot shows the Power Query Editor after sorting the 'date' table. The 'APPLIED STEPS' pane now includes a 'Sorted Rows' step. The table displays 28 rows of dates from 23/01/2020 to 19/02/2020. The ribbon at the top remains the same as in Figure 3.9.1. The 'Transform' tab is still active, and the 'Sort' tool is visible in the ribbon's 'Text Analytics' section.

Figure 3.9.2: screenshot showing date table after sorting rows

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- General** ribbon tab selected.
- Applied Steps** pane on the right shows steps like "Promoted Headers", "Changed Type", "Filtered Rows", etc., with "Added Index" highlighted.
- Properties** pane on the right shows "Name: daily_data (2)" and "All Properties".
- Query Settings** pane on the right shows "Source" and "Applied Steps".
- Queries** list: daily_data, summary_data, Continent, Population, daily_data (2).
- Table View**: A table with columns "date" and "Index". The data shows dates from 25/01/2020 to 19/02/2020, with the "Index" column ranging from 1 to 28.
- Bottom Status Bar**: 2 COLUMNS, 843 ROWS, Column profiling based on top 1000 rows, PREVIEW DOWNLOADED AT 1631.

Figure 3.9.1: screenshot showing date table after 'index column' addition

Next, rename the index column on the date table to 'Date_ID' then merged the date table to the daily_data table. This was done by clicking on the merge queries icon from the 'home ribbon'. Selecting the date column from the daily_data and the date column from the date table and then click ok on the dialogue box. A new column was created. Click on the icon beside the new column and select 'Date_ID' only. Rename the new column to 'Date_ID'.

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Merge Queries** ribbon tab selected.
- Applied Steps** pane on the right shows "Removed Duplicates" highlighted.
- Properties** pane on the right shows "Name: daily_data" and "All Properties".
- Query Settings** pane on the right shows "Source" and "Applied Steps".
- Queries** list: daily_data, summary_data, Continent, Population, Date.
- Table View**: A table with columns "date", "country", "cumulative_total_cases", "daily_new_cases", "active_cases", "cumulative_total_deaths". Below it, a preview table shows data for Afghanistan from March 2020.
- Bottom Status Bar**: 7 COLUMNS, 999+ ROWS, Column profiling based on top 1000 rows, PREVIEW DOWNLOADED AT 1632.

The screenshot shows the Power Query Editor interface with the following details:

- File**, **Home**, **Transform**, **Add Column**, **View**, **Tools**, **Help** menu items.
- Merge Queries** ribbon tab selected.
- Applied Steps** pane on the right shows "Renamed Columns" highlighted.
- Properties** pane on the right shows "Name: daily_data" and "All Properties".
- Query Settings** pane on the right shows "Source" and "Applied Steps".
- Queries** list: summary_data, daily_data, Continent, Population_data, Date.
- Table View**: A table with columns "date", "date_ID". Below it, a preview table shows data for Afghanistan from January 2020 to February 2020.
- Bottom Status Bar**: 7 COLUMNS, 999+ ROWS, Column profiling based on top 1000 rows, PREVIEW DOWNLOADED AT 1432.

Figure 3.9.2: screenshots showing table merge process

ICA DRAFT 2 - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Data Data Sources Parameters Refresh Advanced Editor Properties Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Table Use First Row as Headers Merge Queries Append Queries Text Analytics Vision Combine Files Azure Machine Learning Combine AI Insights

Queries [5]

summary_data daily_data Continent Population_data Date

`= Table.NestedJoin(#"Removed Columns", {"date"}, Date, {"date"}, "Date.1", JoinKind.LeftOuter)`

	_new_cases	active_cases	cumulative_total_deaths	daily_new_deaths	Date.1
1	0	1	0	0	34
2	0	1	0	0	35
3	0	1	0	0	36
4	0	1	0	0	37
5	0	1	0	0	38
6	0	1	0	0	39
7	0	1	0	0	40
8	0	1	0	0	41
9	0	1	0	0	42
10	0	1	0	0	43
11	0	1	0	0	44
12	3	4	0	0	45
13	0	4	0	0	46
14	0	4	0	0	47
15	1	5	0	0	48
16	2	7	0	0	49
17	0	7	0	0	50
18	8	22	1	0	50
19	0	7	0	0	51
20	10	32	1	0	51
21	4	11	0	0	52
22	5	37	1	0	52
23	4	14	0	0	53

8 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 14:36

ICA DRAFT 2 - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Data Data Sources Parameters Refresh Advanced Editor Properties Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Whole Number Use First Row as Headers Merge Queries Append Queries Text Analytics Vision Combine Files Azure Machine Learning Combine AI Insights

Queries [5]

summary_data daily_data Continent Population_data Date

`= Table.ExpandTableColumn(#"Merged Queries", "Date.1", {"date_ID"}, {"Date.1.date_ID"})`

	_new_cases	active_cases	cumulative_total_deaths	daily_new_deaths	Date.1	Date.1.date_ID
1	0	1	0	0	34	34
2	0	1	0	0	35	35
3	0	1	0	0	36	36
4	0	1	0	0	37	37
5	0	1	0	0	38	38
6	0	1	0	0	39	39
7	0	1	0	0	40	40
8	0	1	0	0	41	41
9	0	1	0	0	42	42
10	0	1	0	0	43	43
11	0	1	0	0	44	44
12	3	4	0	0	45	45
13	0	4	0	0	46	46
14	0	4	0	0	47	47
15	1	5	0	0	48	48
16	2	7	0	0	49	49
17	0	7	0	0	50	50
18	8	22	1	0	50	50
19	0	7	0	0	51	51
20	10	32	1	0	51	51
21	4	11	0	0	52	52
22	5	37	1	0	52	52
23	4	14	0	0	53	53

8 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 14:41

Figure 3.9.3: screenshots of filtering table merged by date ID

Queries [5]

daily_data summary_data Continent Population_Data Date

`= Table.RenameColumns(#"Expanded Date.1", {"Date.1.Date_ID", "Date_ID"})`

	_new_cases	active_cases	cumulative_total_deaths	daily_new_deaths	Date_ID	Date.1.Date_ID
1	0	1	0	0	34	34
2	0	1	0	0	35	35
3	0	1	0	0	36	36
4	0	1	0	0	37	37
5	0	1	0	0	38	38
6	0	1	0	0	39	39
7	0	1	0	0	40	40
8	0	1	0	0	41	41
9	0	1	0	0	42	42
10	0	1	0	0	43	43
11	0	1	0	0	44	44
12	3	4	0	0	45	45
13	0	4	0	0	46	46
14	0	4	0	0	47	47
15	1	5	0	0	48	48
16	2	7	0	0	49	49
17	0	7	0	0	50	50
18	8	22	1	0	50	50
19	0	7	0	0	51	51
20	10	32	1	0	51	51
21	4	11	0	0	52	52
22	5	37	1	0	52	52
23	4	14	0	0	53	53

8 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows PREVIEW DOWNLOADED AT 14:41

Figure 3.9.4: screenshot of renamed date ID column

The Date column on the daily_data column was then removed and the date_ID column was reordered to the first column. Sort the table by country.

Query Settings

- APPLIED STEPS**
 - Source
 - Promoted Headers
 - Changed Type
 - Filtered Rows1
 - Removed Duplicates
 - Merged Queries
 - Expanded Date.1
 - Renamed Columns
 - Removed Columns
 - Reordered Columns

Figure 3.9.5: screenshot showing daily_data table after date column removal and column reordering

Query Settings

- APPLIED STEPS**
 - Source
 - Promoted Headers
 - Changed Type
 - Filtered Rows
 - Filtered Rows1
 - Removed Duplicates
 - Merged Queries
 - Expanded Date.1
 - Renamed Columns
 - Removed Columns
 - Reordered Columns
 - Filtered Rows2
 - Sorted Rows

Figure 3.9.6: screenshot showing daily_data table after sorting columns by Country

We are now done creating dimension tables, we click on the ‘close and apply’ icon to apply changes.

Fact Table

I used the daily_data table as my Fact Table. The reason behind my decision was because the daily_data table contains at least one column which are contextual or can be found on each of my dimension tables as well. My fact table included the following:

- **Date_ID:** representation of all the dates covering the period which this data was collected.
- **Country:** representation of all countries whose data was collected during this

exercise.

- **Daily_new_cases:** representation of confirmed cases of COVID-19 daily during the period of this study.
- **Daily_new_deaths:** representation of confirmed deaths due to COVID-19 daily during the period of this study.

All other columns in the daily_data table were removed as there were not relevant to the report. Click 'Close & Apply'.

The screenshot shows the Power Query Editor interface with the 'daily_data' query selected in the Queries list. The main pane displays a table with columns: cumulative_total_cases, daily_new_cases, active_cases, cumulative_total_deaths, and daily_new_deaths. The 'Transform' tab is active, showing various tools like 'Choose Columns', 'Remove Columns', 'Keep Rows', 'Remove Rows', 'Sort', and 'Split Column'. The 'APPLIED STEPS' pane on the right lists steps such as 'Promoted Headers', 'Changed Type', 'Filtered Rows', and 'Reordered Columns'. The status bar at the bottom indicates '8 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'.

This screenshot shows the same Power Query Editor interface after applying changes. The 'daily_data' query now has only four columns: date_ID, country, daily_new_cases, and daily_new_deaths. The 'Transform' tab is still active, and the 'APPLIED STEPS' pane now includes 'Removed Columns'. The status bar at the bottom indicates '4 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'.

Figure 4: screenshots showing removal of unwanted columns from the fact table

The resulting data model was then created as a star schema as shown below. The date table is a dimension table with its keys(date_ID) connected to the daily_data table in a many to one relationship. The Continent table is dimension table connected to the daily_data table through the country column in a many-to-one relationship. The Population table is connected to the fact table through the country column in a many to one relationship. Lastly, the summary_data table is connected to the fact table through the country column in a many to one relationship.

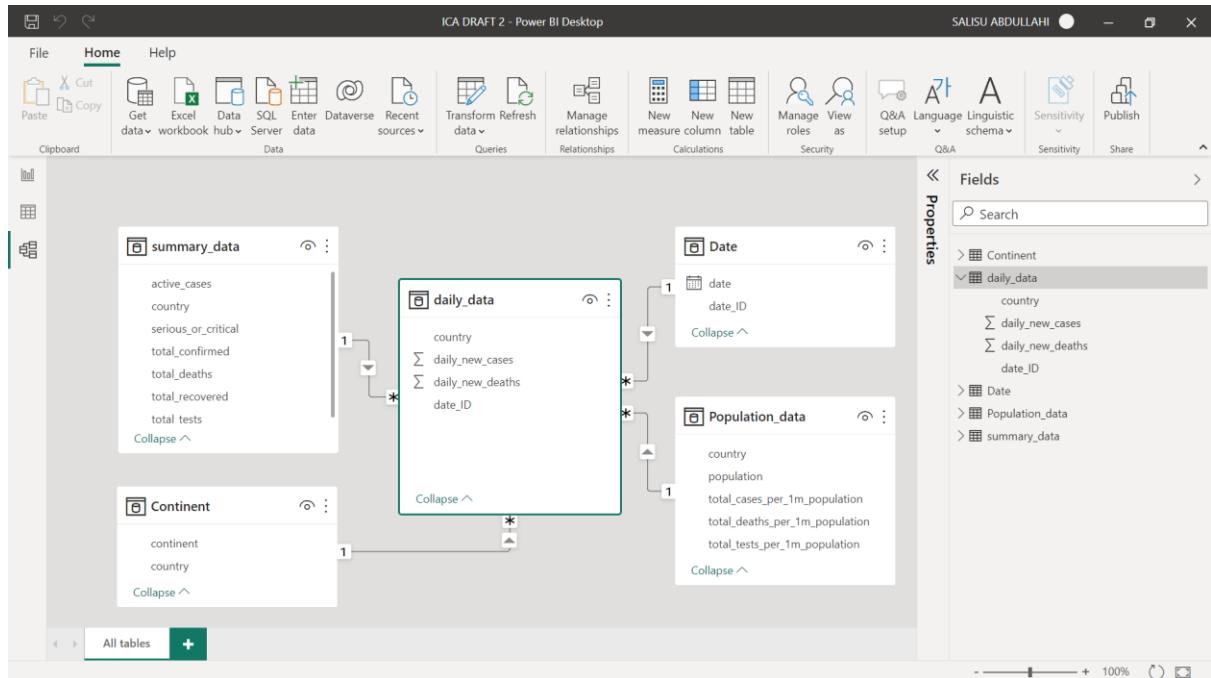


Figure 5: screenshot showing Star Schema data model for Analysis

DAX and M Language

During this analysis, I needed to use DAX to create a calculated column which determines the mortality rate of COVID in each country. To do this, I selected ‘new column’ under calculations in the table tools ribbon after clicking on the Population_data Table under the Data tab. Next step was to name the column ‘mortality_rate’ and write the following formulas and press enter key to create a new column.

```
mortality_rate = Population_data[total_deaths_per_1m_population] / 1000000
```

where mortality_rate is the total deaths per 1 million of the population divided by 1 million. Now that a new column has been created. I formatted the mortality_rate data to be presented in percentages. This was done by clicking on the percentage symbol on under the ‘column tools’ ribbon.

Figure 6.1 displays three screenshots of Power BI Desktop showing the creation of a mortality_rate column using DAX. The screenshots illustrate the process of defining the formula, applying it to the population data, and then summarizing the results.

Screenshot 1: Defining the DAX Formula

The screenshot shows the 'Table tools' ribbon selected. A new column named 'mortality_rate' is being created. The formula is defined as $\text{population} / (\text{total_cases_per_1m_population} + \text{total_deaths_per_1m_population})$. The formula bar also shows the condition $\text{total_deaths_per_1m_population} > 0$.

country	total_cases_per_1m_population	total_deaths_per_1m_population	total_tests_per_1m_population	population	mortality_rate
Afghanistan	4430	190	23455	40500000	0.000519
Albania	59554	1218	612857	5177045	0.02357
Algeria	5865	152	5093	4525517	0.00032
Andorra	543983	1974	222824	77495	0.00257
Angola	2853	55	43136	3978677	0.00161
Anguilla	295676	590	1098870	15252	0.00021
Antigua And Barbuda	77645	1378	109576	99439	0.00140
Argentina	19990	2800	779074	4596774	0.00410
Armenia	142210	2900	1021844	2075158	0.00285
Aruba	221689	1970	1631058	107510	0.00200
Australia	253112	299	272192	26050099	0.00010
Austria	462094	2011	2074092	9102109	0.00220
Azerbaijan	76085	942	661324	10308181	0.00074
Bahamas	81647	81	35964	131359	0.00060
Bahrain	11643	174	81749	187745102	0.00120
Bangladesh	204708	2101	224214	288104	0.00730
Barbados	204078	739	2899952	944205	0.00025
Belgium	357329	2706	2899893	1368581	0.00200
Belize	210813	2991	133670	412089	0.00060
Benin	2120	13	47530	12714341	0.00010
Bermuda	228537	2183	14454002	61850	0.00025
Bhutan	75651	27	2625573	787432	0.00030

Table: Population_data (210 rows)

Screenshot 2: Applying the DAX Formula

The screenshot shows the 'Column tools' ribbon selected. The formula is now applied to the 'population' column. The formula bar shows the full expression: $\text{population} / (\text{total_cases_per_1m_population} + \text{total_deaths_per_1m_population})$.

country	total_cases_per_1m_population	total_deaths_per_1m_population	total_tests_per_1m_population	population	mortality_rate
Afghanistan	4430	190	23455	40500000	0.000519
Albania	59554	1218	612857	5177045	0.02357
Algeria	5865	152	5093	4525517	0.00032
Andorra	543983	1974	222824	77495	0.00257
Angola	2853	55	43136	3978677	0.00161
Anguilla	295676	590	1098870	15252	0.00021
Antigua And Barbuda	77645	1378	109576	99439	0.00140
Argentina	19990	2800	779074	4596774	0.00410
Armenia	142210	2900	1021844	2075158	0.00285
Aruba	221689	1970	1631058	107510	0.00200
Australia	253112	299	272192	26050099	0.00010
Austria	462094	2011	2074092	9102109	0.00220
Azerbaijan	76085	942	661324	10308181	0.00074
Bahamas	81647	81	35964	131359	0.00060
Bahrain	11643	174	81749	187745102	0.00120
Bangladesh	204708	2101	224214	288104	0.00730
Barbados	204078	739	2899952	944205	0.00025
Belarus	104098	199	1899911	3648545	0.00050
Belgium	352321	2706	2899893	11635592	0.00200
Belize	210893	2644	133670	412089	0.00060
Benin	2120	13	47530	12714341	0.00010
Bermuda	228537	2183	14454002	61850	0.00025
Bhutan	75651	27	2625573	787432	0.00030

Table: Population_data (210 rows) Column: mortality_rate (1 distinct values)

Screenshot 3: Summarizing the Mortality Rate

The screenshot shows the 'Table tools' ribbon selected. The formula is now applied to the 'population' column. The formula bar shows the full expression: $\text{population} / (\text{total_cases_per_1m_population} + \text{total_deaths_per_1m_population})$.

country	total_cases_per_1m_population	total_deaths_per_1m_population	total_tests_per_1m_population	population	mortality_rate
Afghanistan	4420	190	23455	40500000	0.000519
Albania	59554	1218	612857	5177045	0.02357
Algeria	5865	152	5093	4525517	0.00032
Andorra	543983	1974	222824	77495	0.00257
Angola	2853	55	43136	3978677	0.00161
Anguilla	295676	590	1098870	15252	0.00021
Antigua And Barbuda	77645	1378	109576	99439	0.00140
Argentina	19990	2800	779074	4596774	0.00410
Armenia	142210	2900	1021844	2075158	0.00285
Aruba	221689	1970	1631058	107510	0.00200
Australia	253112	299	272192	26050099	0.00010
Austria	462094	2011	2074092	9102109	0.00220
Azerbaijan	76085	942	661324	10308181	0.00074
Bahamas	81647	81	35964	131359	0.00060
Bahrain	11643	174	81749	187745102	0.00120
Bangladesh	204708	2101	224214	288104	0.00730
Barbados	204078	739	2899952	944205	0.00025
Belarus	104098	199	1899911	3648545	0.00050
Belgium	352321	2706	2899893	11635592	0.00200
Belize	210893	2644	133670	412089	0.00060
Benin	2120	13	47530	12714341	0.00010
Bermuda	228537	2183	14454002	61850	0.00025
Bhutan	75651	27	2625573	787432	0.00030

Table: Population_data (210 rows) Column: mortality_rate (199 distinct values)

Figure 6.1: screenshots of creating mortality_rate column using DAX

Now that I have gotten the mortality, I decided to also get the infection rate in each country. Calculated columns were also created to determine the recovery rate in each country. To do this, same steps were applied to create a new column but below formula was used instead and the corresponding result is shown below.

Infection_rate = Population_data[total_cases_per_1m_population] / 1000000

Recovery_rate = summary_data[total_recovered] / summary_data[total_confirmed]

Where infection rate is total cases per 1 million of the population divided by 1 million and

Recovery rate is the total recovered cases divided by the total confirmed cases (calculated under summary table)

ICA DRAFT 2 - Power BI Desktop

SALISU ABDULLAHI

Column tools

Name: Infection_rate
Data type: Decimal number

Format: Percentage
Summarization: Don't summarize
Data category: Uncategorized

Sort by column
Sort groups
Data groups
Manage relationships
New column

Fields

Search

country, total_cases_per_1m_population, total_deaths_per_1m_population, total_tests_per_1m_population, population, mortality_rate, Infection_rate

Continent, daily_data, Date, Population_data, Infection_rate, mortality_rate, population, total_cases_per_1m..., total_deaths_per_1m..., total_tests_per_1m_p...

summary_data

Table: Population_data (210 rows) Column: Infection_rate (210 distinct values)

Figure 6.2: screenshot of creating infection_rate column using DAX

ICA DRAFT 2 - Power BI Desktop

SALISU ABDULLAHI

Column tools

Name: Recovery_rate
Data type: Decimal number

Format: Percentage
Summarization: Don't summarize
Data category: Uncategorized

Sort by column
Sort groups
Data groups
Manage relationships
New column

Fields

Search

country, total_confirmed, total_deaths, total_recovered, active_cases, serious_or_critical, total_tests, Recovery_rate

Continent, daily_data, Date, Population_data, summary_data, active_cases, country, Recovery_rate, serious_or_critical, total_confirmed, total_deaths, total_recovered, total_tests

Table: summary_data (210 rows) Column: Recovery_rate (192 distinct values)

Figure 6.3: screenshot of creating recovery_rate column using DAX

Now that I have gotten the infection rate for each country, the next step was to find the average infection rate globally. To do this, I created a calculated measure which took the total of the infection rates of all the countries under analysis and find the average. I clicked on the population_data table and selected on the new measure icon from the ‘table tools’ ribbon and wrote the following in the formula box.

Global_Average_Infection = AVERAGE(Population_data[Infection_rate])

Where the global average infection rate is the average of the infection rate total.

The resultant value was also shown in percentages using same steps I used on the calculated column (clicking on the percentage symbol on under the ‘column tools’ ribbon).

ICA DRAFT 2 - Power BI Desktop

File Home Help Table tools Column tools

Name: Population_data

Mark as date table Calendars Manage relationships Relationships New measure column New table Calculations

Fields

Search

country total_cases_per_1m_population total_deaths_per_1m_population total_tests_per_1m_population population mortality_rate Infection_rate

Afghanistan	4420	190	23455	4056036	0.02%	0.44%
Albania	95954	1218	632857	2871945	0.12%	9.60%
Algeria	5865	152	5093	4532517	0.02%	0.59%
Andorra	543983	1974	3223924	77495	0.20%	54.40%
Angola	2853	55	43136	34769277	0.01%	0.29%
Anguilla	195646	590	3366870	15252	0.06%	19.56%
Antigua And Barbuda	77646	1378	190076	99439	0.14%	7.76%
Argentina	197992	2800	776974	45968174	0.28%	19.80%
Armenia	142219	2900	1031834	2973538	0.20%	14.22%
Aruba	331689	1979	1653053	107610	0.20%	33.17%
Australia	253112	299	2721042	26050899	0.03%	25.31%
Austria	462804	2011	20328801	9102106	0.20%	46.28%
Azerbaijan	76885	942	663324	10309383	0.09%	7.69%
Bahamas	84626	2001	583330	400242	0.20%	8.46%
Bahrain	318491	816	5396149	1811659	0.08%	31.85%
Bangladesh	11643	174	83767	16745162	0.02%	1.16%
Barbados	264768	1503	2324136	288030	0.15%	26.48%
Belarus	104078	739	1399951	9443535	0.07%	10.41%
Belgium	352324	2706	2896893	11683561	0.27%	35.23%
Belize	140843	1644	1363070	411069	0.16%	14.08%
Benin	2120	13	47530	12714341	0.00%	0.21%
Bermuda	228537	2183	14454002	61850	0.22%	22.85%

Table: Population_data (210 rows) Column: Infection_rate (210 distinct values)

ICA DRAFT 2 - Power BI Desktop

File Home Help Table tools Measure tools

Name: Global_Average_Infection

Format: Percentage

Home table: Population_data

Data category: Uncategorized

New Quick measure measure Calculations

Fields

Search

country total_cases_per_1m_population total_deaths_per_1m_population total_tests_per_1m_population population mortality_rate Infection_rate

Afghanistan	4420	190	23455	4056036	0.02%	0.44%
Albania	95954	1218	632857	2871945	0.12%	9.60%
Algeria	5865	152	5093	4532517	0.02%	0.59%
Andorra	543983	1974	3223924	77495	0.20%	54.40%
Angola	2853	55	43136	34769277	0.01%	0.29%
Anguilla	195646	590	3366870	15252	0.06%	19.56%
Antigua And Barbuda	77646	1378	190076	99439	0.14%	7.76%
Argentina	197992	2800	776974	45968174	0.28%	19.80%
Armenia	142219	2900	1031834	2973538	0.20%	14.22%
Aruba	331689	1979	1653053	107610	0.20%	33.17%
Australia	253112	299	2721042	26050899	0.03%	25.31%
Austria	462804	2011	20328801	9102106	0.20%	46.28%
Azerbaijan	76885	942	663324	10309383	0.09%	7.69%
Bahamas	84626	2001	583330	400242	0.20%	8.46%
Bahrain	318491	816	5396149	1811659	0.08%	31.85%
Bangladesh	11643	174	83767	16745162	0.02%	1.16%
Barbados	264768	1503	2324136	288030	0.15%	26.48%
Belarus	104078	739	1399951	9443535	0.07%	10.41%
Belgium	352324	2706	2896893	11683561	0.27%	35.23%
Belize	140843	1644	1363070	411069	0.16%	14.08%
Benin	2120	13	47530	12714341	0.00%	0.21%
Bermuda	228537	2183	14454002	61850	0.22%	22.85%

Table: Population_data (210 rows) Column: Global_Average_Infection (0 distinct values)

Figure 6.4: screenshots of creating global average infection rate using DAX calculated measures

Calculated measures were also created under the daily_data table to determine the number of covid confirmed infection cases, number of covid deaths globally, and global number of recovered cases. Same steps as above were applied, and the corresponding formulas used are shown below.

`Total_Infected = SUM(daily_data[daily_new_cases])`
`total_deaths = SUM(daily_data[daily_new_deaths])`
`total_recovered = SUM(summary_data[total_recovered])`

where

- Total infected is the sum of daily new cases of covid
- Total Deaths is the sum of daily new deaths due to covid
- Total recovered is the sum of the total recovered cases of covid

The screenshot shows the Power BI Desktop interface with the 'Measure tools' tab selected. A new measure named 'Total_Infected' is being defined, with the formula `1 Total_Infected = SUM(daily_data[daily_new_cases])`. The formula bar shows a checkmark indicating the formula is valid. The 'Fields' pane on the right lists various data fields and measures, including 'country', 'daily_new_cases', 'daily_new_deaths', 'date_ID', 'total_Infected', and others. The main data grid displays rows for Tanzania with daily new cases values ranging from 0 to 599.

Figure 6.5: screenshot showing total infected measure created using DAX calculated measures

The screenshot shows the Power BI Desktop interface with the 'Measure tools' tab selected. A new measure named 'total_deaths' is being defined, with the formula `1 total_deaths = SUM(daily_data[daily_new_deaths])`. The formula bar shows a checkmark indicating the formula is valid. The 'Fields' pane on the right lists various data fields and measures, including 'country', 'daily_new_cases', 'daily_new_deaths', 'date_ID', 'total_deaths', and others. The main data grid displays rows for Tanzania with daily new deaths values ranging from 0 to 599.

Figure 6.5: screenshot showing total deaths measure created using DAX calculated measures

The screenshot shows the Power BI Desktop interface with the 'Measure tools' ribbon tab selected. A calculated measure named 'total_recovered' is being defined using the DAX formula `= SUM(summary_data[total_recovered])`. The Fields pane on the right displays a hierarchy of data sources and fields, including 'Continent', 'daily_data', 'country', 'daily_new_cases', 'daily_new_deaths', 'date_ID', 'total_deaths', 'Total_Infected', and 'total_recovered'. The main area shows a table with columns: country, daily_new_cases, daily_new_deaths, date_ID, and total_recovered, all containing the value 0.

Figure 6.5: screenshot showing total recovered measure created using DAX calculated measures

M. Language.

Now that I have created the respective calculated columns and measures, my next step is to outline countries based on their infection rates in my report to provide insight. To do this, I used M. Language to create a risk category based on infection rate.

This was done by first selecting 'Get data' and clicking on 'Blank Query' from the menu.

The screenshot shows the Power BI Desktop interface with the 'Home' ribbon tab selected. The 'Get Data' button is highlighted. The 'Common data sources' pane on the left lists various options such as 'Get data', 'Clipboard', 'Common data sources', 'Enter data', 'Transform data', 'Relationships', 'New measure', 'New column', 'New table', 'Manage roles', 'View as', 'Sensitivity', and 'Publish'. The 'Fields' pane on the right shows the same hierarchy of data sources and fields as in Figure 6.5. The main area shows a table with columns: date_ID, total_recovered, and total_deaths, all containing the value 0.

Figure 7.0: screenshot showing 'Get Data' and 'Blank Query' selection

The next step was to select 'Advanced Editor' from the dialogue box that opened.

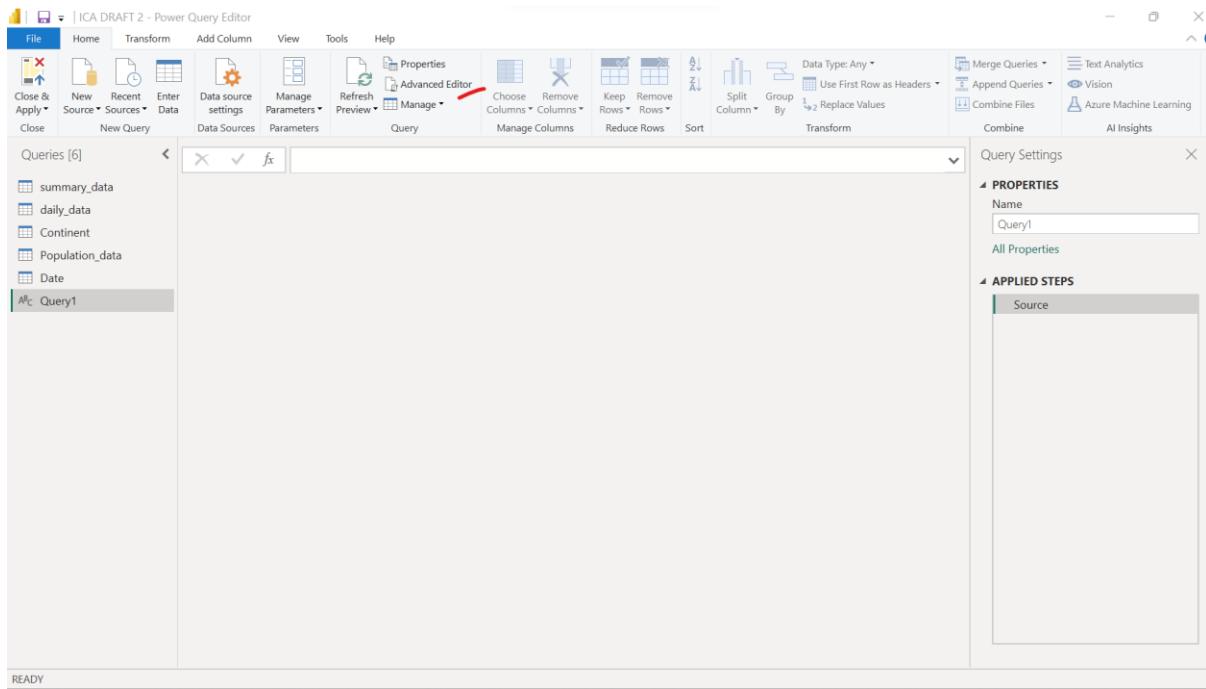


Figure 7.1: screenshot showing 'advanced editor' selection

The Table.fromrecord function was then called and the following below instructions were inputted and done was clicked. Rename the table to Risk Category and Click on close and apply.

```
Source = Table.FromRecords ({

    [Category = "0-15%", Lower Level = 0, Upper Level = 0.15], 

    [Category = "15-30%", Lower Level = 0.15, Upper Level = 0.30], 

    [Category = "30-50%", Lower Level = 0.30, Upper Level = 0.50], 

    [Category = "50-70%", Lower Level = 0.50, Upper Level = 0.70], 

    [Category = "70-100%", Lower Level = 0.70, Upper Level = 1]

})
```

Where

- Table.FromRecords instructs the programme to create a table.
- Category, Lower Level and Upper Level are columns created in the table.

Next step, I changed the data type on lower and upper level to decimals. This was done by selecting transform data and then clicking on the icon beside the lower level and upper level icon and selecting decimal number. Click close and apply.

The screenshot shows two windows side-by-side. The left window is the 'Advanced Editor' showing the following M Language code:

```

let
    Source = Table.FromRecords ({
        [Category = "0-15%", Lower Level = 0, Upper Level = 0.15],
        [Category = "15-30%", Lower Level = 0.15, Upper Level = 0.30],
        [Category = "30-50%", Lower Level = 0.30, Upper Level = 0.50],
        [Category = "50-70%", Lower Level = 0.50, Upper Level = 0.70],
        [Category = "70-100%", Lower Level = 0.70, Upper Level = 1]
    }),
    #"Changed Type" = Table.TransformColumnTypes(Source,{{"Lower Level", type number}, {"Upper Level", type number}})
in
    #"Changed Type"

```

The right window is the 'FINAL ICA DRAFT - Power Query Editor' showing the 'Transform' ribbon selected. The 'Queries [6]' pane lists 'summary_data', 'daily_data', 'Continent', 'Population_data', 'Date', and 'RiskCategory'. The 'RiskCategory' query is selected, showing a preview table with columns 'Category', 'Lower Level', and 'Upper Level'. The 'Applied Steps' pane shows the steps: 'Source' and 'Changed Type'.

Category	Lower Level	Upper Level
0-15%	0	0.15
15-30%	0.15	0.3
30-50%	0.3	0.5
50-70%	0.5	0.7
70-100%	0.7	1

Figure 7.2: screenshot showing creation of infection risk categories using M Language

I then converted both columns into percentages. This was done by clicking on the percentage symbol on the **Column Tools** Ribbon.

Next step was to map the risk category table to a risk category column (bin). To do this, I created a new calculated column and called it RiskCategory and used the CALCULATE function to map the infection rate to the risk category. See formula below

```

RiskCategory =
VAR CatVar=[Infection_rate]
RETURN
CALCULATE (VALUES ( RiskCategory[Category]), CatVar>RiskCategory[Lower
Level],CatVar<=RiskCategory[Upper Level])

```

Where

- VAR = variable which stores the result of the expression which can now be passed as an argument in another measure expression
- CatVar = name of the variable
- [infection_rate] is the expression returning a scalar or table value
- CALCULATE (VALUES (RiskCategory[Category]), CatVar>RiskCategory[Lower Level],CatVar<=RiskCategory[Upper Level]) calculates the values of the infection rate for each country with respect to their risk category.

The screenshot shows two separate Power BI Desktop windows side-by-side.

Top Window:

- Column Tools:**
 - Name: Upper Level
 - Data type: Decimal number
 - Format: Percentage
 - Summarization: Sum
 - Data category: Uncategorized
- Data View:**

Category	Lower Level	Upper Level
0-15%	0.00%	15.00%
15-30%	15.00%	30.00%
30-50%	30.00%	50.00%
50-70%	50.00%	70.00%
70-100%	70.00%	100.00%
- Power BI Data View:**
 - Search bar: Search
 - Navigation pane: Continent, daily_data, Date, Population_Data, Risk Category (selected), summary_data
 - Table structure: Category, Lower Level, Upper Level

Bottom Window:

- Column Tools:**
 - Name: RiskCategory
 - Data type: Text
 - Format: Text
 - Summarization: Don't summarize
 - Data category: Uncategorized
- Data View:**

```

1 RiskCategory =
2 VAR CatVar=[Infection_rate]
3 RETURN
4 CALCULATE (VALUES (RiskCategory[Category]), CatVar>RiskCategory[Lower Level],CatVar<=RiskCategory[Upper Level])
    
```

RiskCategory	total_cases_per_1m_population	total_deaths_per_1m_population	total_tests_per_1m_population	population	mortality_rate	Infection_rate	RiskCategory
0-15%	4420	190	23455	40560636	0.02%	0.44%	0-15%
0-15%	95954	1218	632857	2871945	0.12%	9.60%	0-15%
0-15%	5865	152	5093	45325517	0.02%	0.59%	0-15%
0-15%	543983	1974	3223924	77495	0.20%	54.40%	50-70%
0-15%	2853	55	43136	34769277	0.01%	0.29%	0-15%
0-15%	195646	590	3368870	15252	0.06%	19.56%	15-30%
0-15%	77646	1378	190076	99439	0.14%	7.76%	0-15%
0-15%	197992	2800	776974	45968174	0.28%	19.80%	15-30%
0-15%	142219	2900	1031834	2973558	0.29%	14.22%	0-15%
0-15%	331689	1979	1653053	107610	0.20%	33.17%	30-50%
0-15%	253112	299	2721042	26050899	0.03%	25.31%	15-30%
0-15%	462804	2011	20328801	912016	0.20%	46.28%	30-50%
0-15%	76885	942	663324	10309383	0.09%	7.69%	0-15%
0-15%	84626	2001	583330	400242	0.20%	8.46%	0-15%
0-15%	318491	816	5396149	1811659	0.08%	31.85%	30-50%
0-15%	11643	174	83767	167745162	0.02%	1.16%	0-15%
0-15%	264768	1503	2324136	288030	0.15%	26.48%	15-30%
0-15%	104078	739	1399951	9443535	0.07%	10.41%	0-15%
0-15%	352324	2706	2896893	11683561	0.27%	35.23%	30-50%
0-15%	140843	1644	1336070	411069	0.16%	14.08%	0-15%
- Power BI Data View:**
 - Search bar: Search
 - Navigation pane: Continent, daily_data, Date, Population_data (selected), Global_Average_Infe..., Infection_rate, mortality_rate, population, RiskCategory (selected), summary_data
 - Table structure: total_cases_per_1m_population, total_deaths_per_1m_population, total_tests_per_1m_population, population, mortality_rate, Infection_rate, RiskCategory

Figure 7.3: screenshot showing calculation of infection rate into risk categories using M Language

M. Language was also used to create category for recovery rate as well using the same steps as above but with formula below instead.

RecoveryCategory =

```

VAR CatVar=[Recovery_rate]
RETURN
CALCULATE (VALUES (RiskCategory[Category]), CatVar>RiskCategory[Lower
Level],CatVar<=RiskCategory[Upper Level])

```

The screenshot shows the Power BI Desktop interface with the 'Column tools' tab selected. In the code editor, the following M Language code is written:

```

1 RecoveryCategory =
2 VAR CatVar=[Recovery_rate]
3 RETURN
4 CALCULATE (VALUES (RiskCategory[Category]), CatVar>RiskCategory[Lower Level],CatVar<=RiskCategory[Upper Level])

```

Below the code is a table titled 'summary_data' with 210 rows. The columns include country, total_confirmed, total_deaths, total_recovered, active_cases, serious_or_critical, total_tests, Recovery_rate, and RecoveryCategory. The table shows data for various countries, such as Angola, Armenia, Aruba, Azerbaijan, Bahamas, Barbados, Belarus, Bermuda, Bhutan, Bosnia And Herzegovina, Brunei Darussalam, Burkina Faso, Burundi, Cambodia, Caribbean Netherlands, Chad, Channel Islands, Congo, Cook Islands, and Cote D'Ivoire.

The 'Fields' pane on the right lists several data sources and their fields, including 'Continent', 'daily_data' (with 'country', 'daily_new_cases', 'daily_new_deaths', 'date_ID', 'total_deaths', 'Total_Infected', and 'total_recovered'), 'Date', 'Population_data', 'RiskCategory', 'summary_data' (with 'active_cases', 'country', 'Recovery_rate', 'serious_or_critical', and 'total_confirm'), and 'RecoveryCategory'.

Figure 7.3: screenshot showing calculation of recovery rate into risk categories using M Language

Dashboard

My dashboard is made up of 5 pages which analyses and answers the business questions and provide other information. The Pages were titled in chronological order as listed below:

- Global Situation
- Continental Situation
- Country Situation
- Country Specific Situation
- Summary Information

First page provides visual information on covid globally, the second provides a breakdown of covid pandemic across the continents. The third page shows the pandemic effect across the countries under the analysis and the fourth provide information on how COVID has fared in Nigeria and last page provides a summary on the previous dashboards.

Global Situation Dashboard

This dashboard is made up of 3 cards showing the total infected, total deaths and the current global infection rate respectively. It also contains two drill down timeline zoom charts which visualises confirmed infected cases and confirmed deaths. These charts show the breakdown of covid infected cases by year (2020 to date). Furthermore, these zoom charts allow drill down into any year of choice should you need further analysis of covid cases based on the months of the year and much further into each day of any month. This dashboard is an interactive dashboard as it provides specific details into any selected period as shown below.

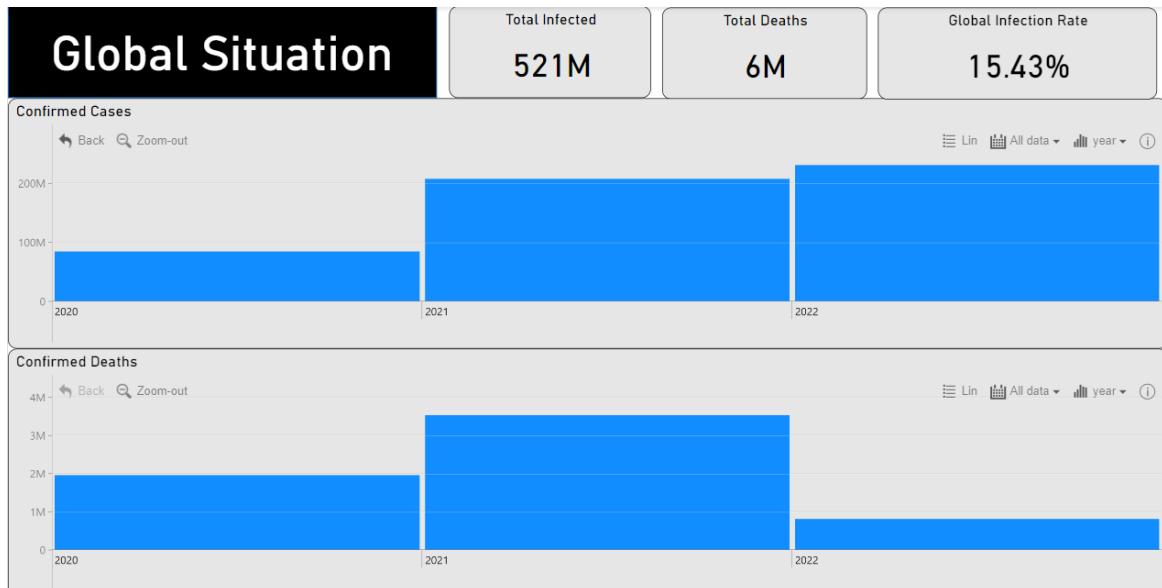


Figure 8.0: screenshot showing Global situation dashboard of covid by year

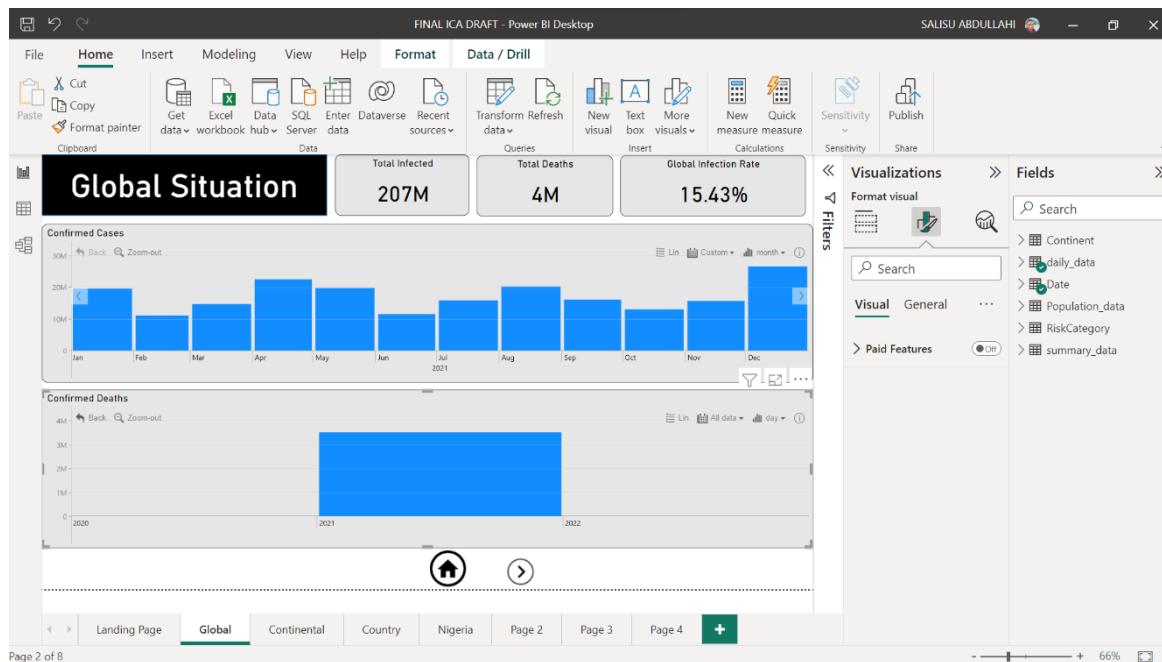


Figure 8.1: screenshot showing Global situation dashboard of covid by months in a specific year

Continent Situation Dashboard

In this dashboard, we have a combination of visuals that show the distribution of infected cases and deaths across the continents of the world while also showing the countries under these continents.

The Visual is made up of a slicer which allows you specify the year and period you want to visualise; an aster plot showing the infected case distribution; a clustered bar chart that shows the distribution of confirmed deaths due to COVID and a multi-row card that shows the countries under any selected continent.

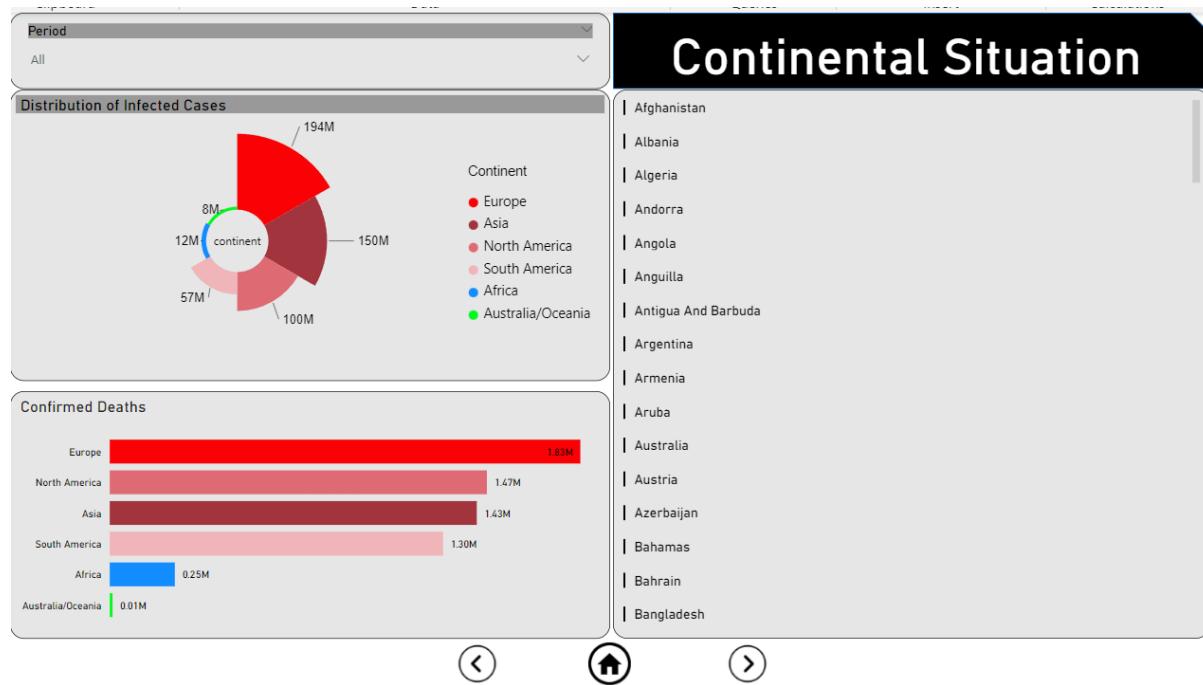


Figure 9.0: screenshot showing Continental situation dashboard of covid

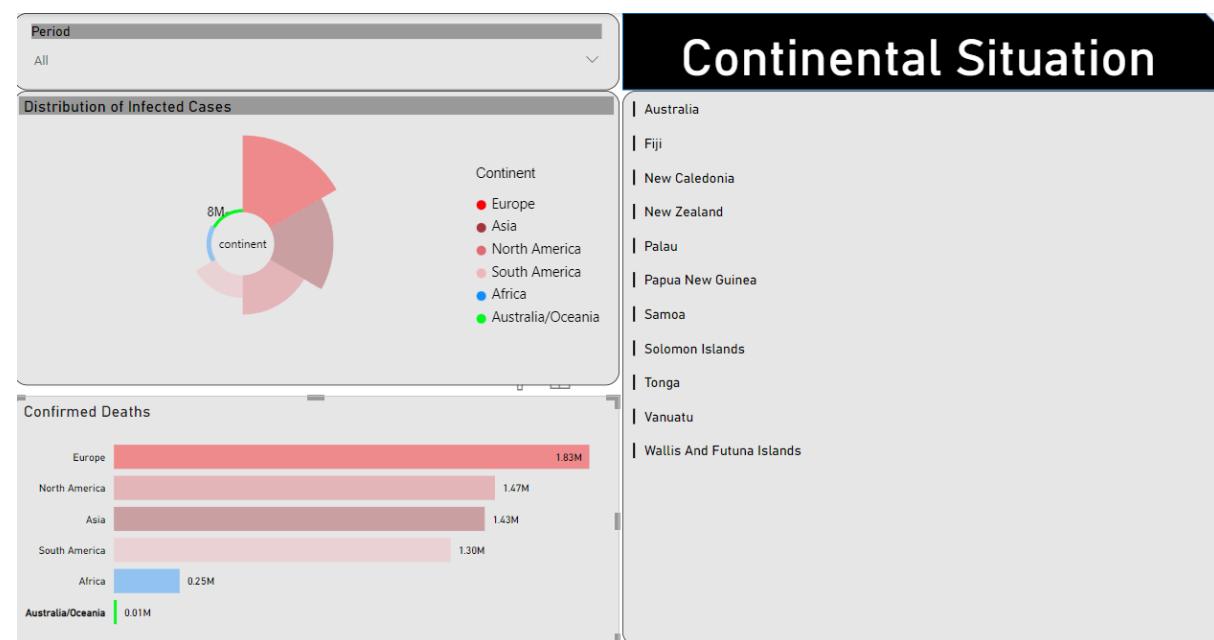


Figure 9.1: screenshot showing Continental situation dashboard of covid specifically highlighting Australia/Oceania

Country Situation Dashboard

Here, I visualised the impact of covid-19 across all the countries under analysis while also providing an infection risk probability for users to advise themselves.

The dashboard contains a table which shows the countries of the world, the total number of infected people, the number of deaths from covid, the infection rate, the covid mortality rate and the covid recovery rate in these countries. The table is also colour-padded to show countries whose infection rate are above the global infection rate average (red) and countries with infection rate lower than the global average(green).

The dashboard also contains a donut chart which visualises the distribution of infection risk across these countries. Countries in each infection risk category are shown under the table when the category is selected.

The last visualisation in this dashboard is a funnel chart which shows the top countries with a high infection rate. The aim of the visualisation is to advise travellers who intend to visit any of the listed countries.

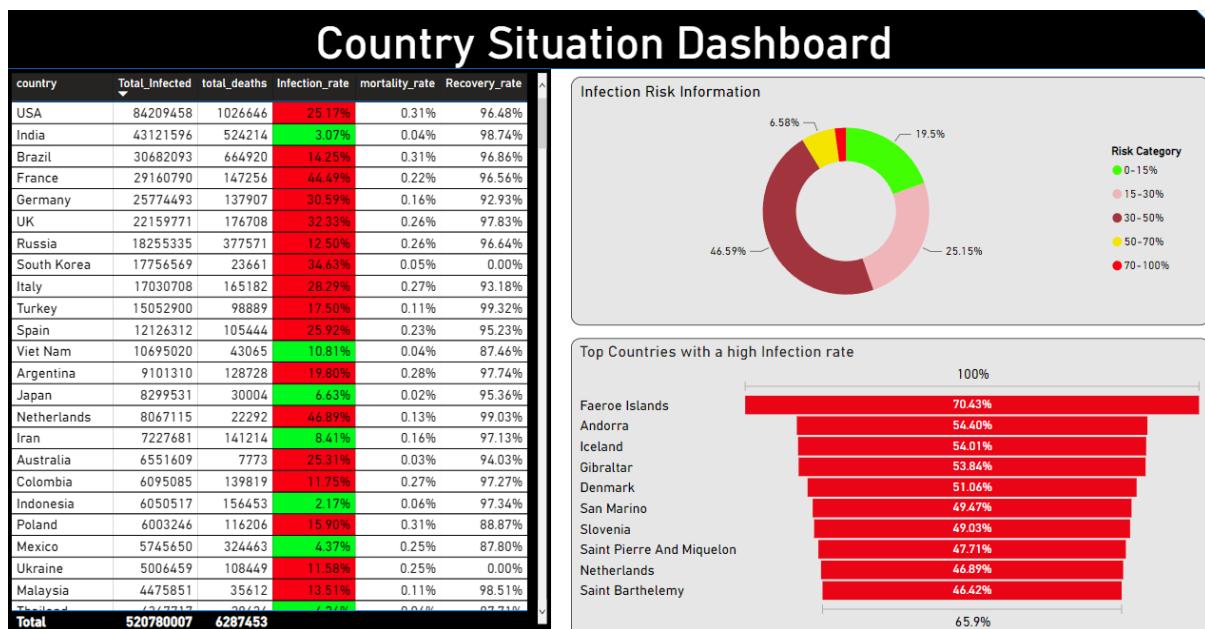


Figure 10.: screenshot showing Country situation dashboard of covid

Country-Specific Situation (Nigeria)

In the country-specific dashboard, I visualised COVID information for Nigeria. This visualisation showed the daily infection statistics and daily death cases since the COVID-19 pandemic started and the influence behind the tendency for the COVID infection to decrease.

This dashboard contains a multi-row card, which provides information regarding the population of Nigeria, the number of covid infection, deaths due to covid and tests for Coronavirus that have been done and the current infection rate in the country

There are also two area charts which shows the daily reported infection cases and the daily confirmed covid deaths across the country respectively since the outbreak of corona virus.

The dashboard also contains a key influencer chart which shows the periods which the average daily number of confirmed infection cases in Nigeria has reduced.

Lastly, for easy navigation, there is a Timeline 2.4.0 slicer which is of advantage should you want information specific to a range of period.

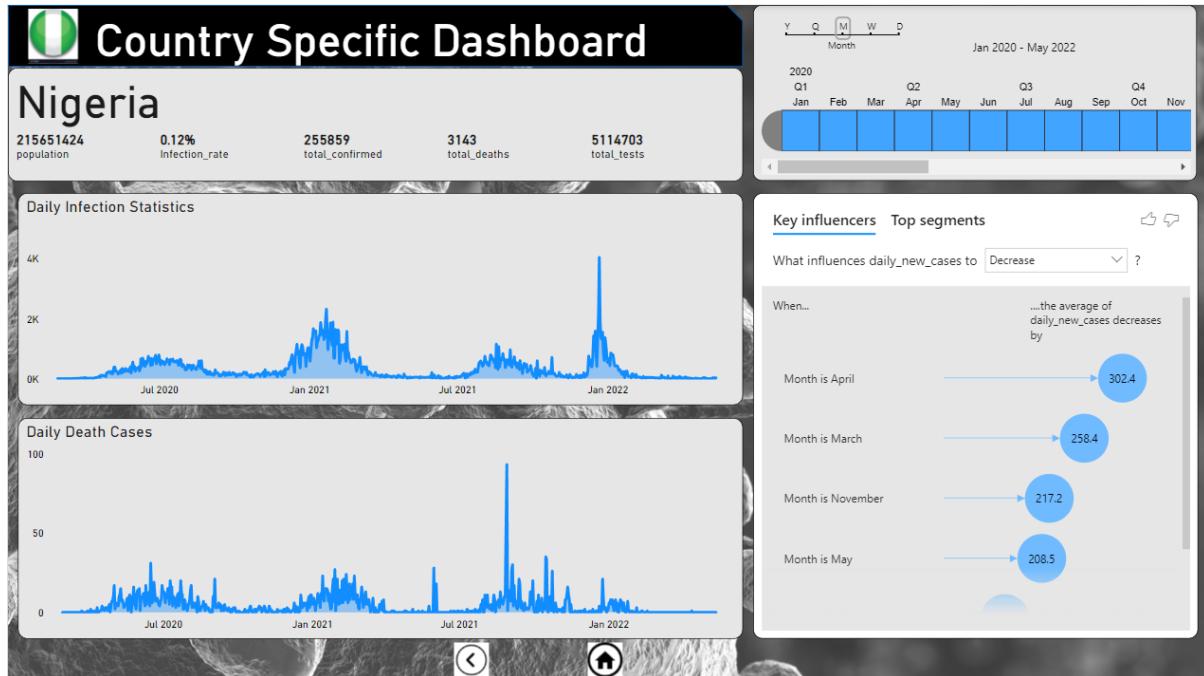


Figure 11: screenshot showing Nigeria situation dashboard of covid

Summary Information

In this dashboard, I visualised the detailed analysis of each of the previous dashboards. This was done using the smart narration visuals which uses artificial intelligence to analyse visualisations and provide insights.

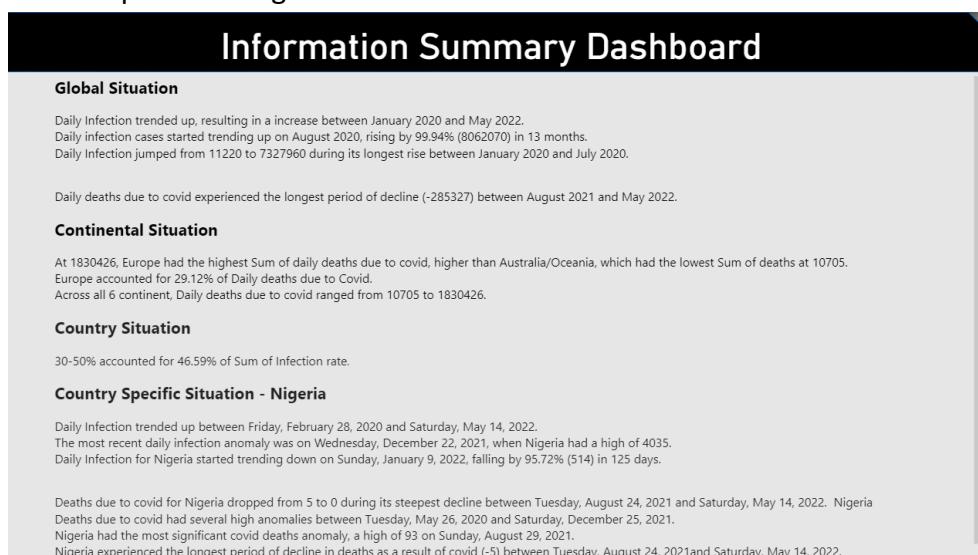


Figure 12: screenshot showing Summary Information dashboard of covid

References

1. Assaker, J. (2022) 'Covid-19 Global Dataset'. Available at:
<https://www.kaggle.com/datasets/josephassaker/covid19-global-dataset> (Accessed 02 January 2023).

Report Section	Description	Grade your work from 0 to 100
Report Structure	The report is well-written, and it contains all the relevant sections	85
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	85
Dax and M language	Both DAX and M Language have been extensively used in the report	80
Dashboard Design	The dashboard contains a variety of charts, including advanced ones not covered in the module.	90
Average		Add below the average of the four cells above: 85