**COVID ANALYSIS USING POWER BI**

Contents

Acknowledge………………………………………………………………………………………………………………... 2

Data Source Description………………………………………………………………………………………………… 3

Business Intelligence Questions……………………………………………………………………………………… 5

Loading Data………………………………………………………………………………………….……………………… 7

BI Data Pre-Processing………………….………………………………………………………………………….……. 8

1. Create Date Table ……………………………………………………………………………………………… 11
2. Create Relationship …………………………………………………………………………………………... 11
3. Modify Column Name ………………………………………………………………………………………… 12
4. Select Theme ……………………………………………………………………………………………………… 13
5. Create Snapshot View ………………………………………………………………………………………… 13
6. Add Year, Month Slicer ………………………………………………………………………………………. 15

Section 2 Business Report ……………………………………………………………………………………………… 17

Key Findings

1. Creating Line Graph on “People Vaccinated by Date” ……………………………………….… 19
2. Creating Clustered Column Chart on “Top 10 Country by People Fully Vaccinated”. 20
3. Create Clustered Bar Chart “Top 10 Country by Daily Vaccinations”…………………….. 23
4. Create Map Visual “Total Vaccinations by Country”……………………………………………… 25

Insights ……………………………………………………………………………………………………………………………. 27

Dashboard ……………………………………………………………………………………………………….…………..…. 28

Publishing to Power BI Service ………………………………………………………………………………….… 29

Conclusion ………………………………………………………………………………….…………………………………... 30

Data Source Description

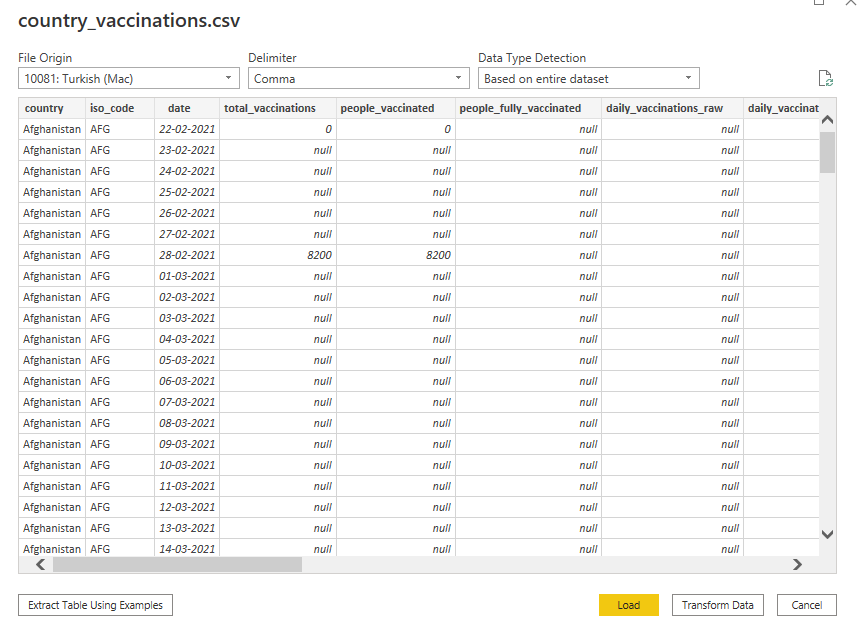
**Data Source**: <https://github.com/owid/covid-19-data>

Data is collected from [**Our World in Data**](https://ourworldindata.org/) GitHub repository for [covid-19](https://github.com/owid/covid-19-data). Country level vaccination data is gathered and assembled in one single file. Then, this data file is merged with locations data file to include vaccination sources information. A second file, with manufacturer’s information, is included.

The data (country vaccinations) contains the following information:

* **Country** - this is the country for which the vaccination information is provided
* **Country ISO Code**  - ISO code for the country
* **Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total
* **Total number of vaccinations** - this is the absolute number of total immunizations in the country
* **Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people
* **Total number of people fully vaccinated** - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme
* **Daily vaccinations (raw)** - for a certain data entry, the number of vaccination for that date/country
* **Daily vaccinations** - for a certain data entry, the number of vaccination for that date/country
* **Total vaccinations per hundred** - ratio (in percent) between vaccination number and total population up to the date in the country
* **Total number of people vaccinated per hundred** - ratio (in percent) between population immunized and total population up to the date in the country
* **Total number of people fully vaccinated per hundred** - ratio (in percent) between population fully immunized and total population up to the date in the country
* **Number of vaccinations per day** - number of daily vaccination for that day and country
* **Daily vaccinations per million** - ratio (in ppm) between vaccination number and total population for the current date in the country
* **Vaccines used in the country** - total number of vaccines used in the country (up to date)
* **Source name** - source of the information (national authority, international organization, local organization etc.)
* **Source website** - website of the source of information

**Data Overview**



Business Intelligence Questions

It has been over a full year since we started this difficult battle against Covid-19 and it has cost us dearly. These fast-spreading microscopic creatures that are too small to be even be seen with the naked eye have somehow shaken the entire world and changed life as we know it

People all over the world are racing against the clock to fight this pandemic. The main objective is to explore COVID vaccine data and try to find out the inference of how the immunization process is going on worldwide.

The selected dataset has detailed information of the Covid - 19 Vaccination details, thus the primary objective of this report is to analyse the vaccination count & progress of vaccination in different parts of the world using power BI tool

**The business problems that can be solved are**

1. Total Number of people Vaccinated sorting by Date
2. Top 10 Country by People Fully Vaccinated
3. Total Vaccinations by Country
4. Where are vaccinated more people per day?
5. Which country is using what vaccine?
6. In which country the vaccination programme is more advanced?
7. Top 10 Country by Daily Vaccinations
8. Comparison of countries based on Partial vaccinations & people fully vaccinated

Loading Data

The very first step of data analysis and creating Business Intelligence is loading raw data. Power BI supports data from various sources like web api, databases, flat files in excel, csv or space separated format, and many more.

On startup, Power BI pops up a dialog box to load existing project or Get data as shown in figure below.

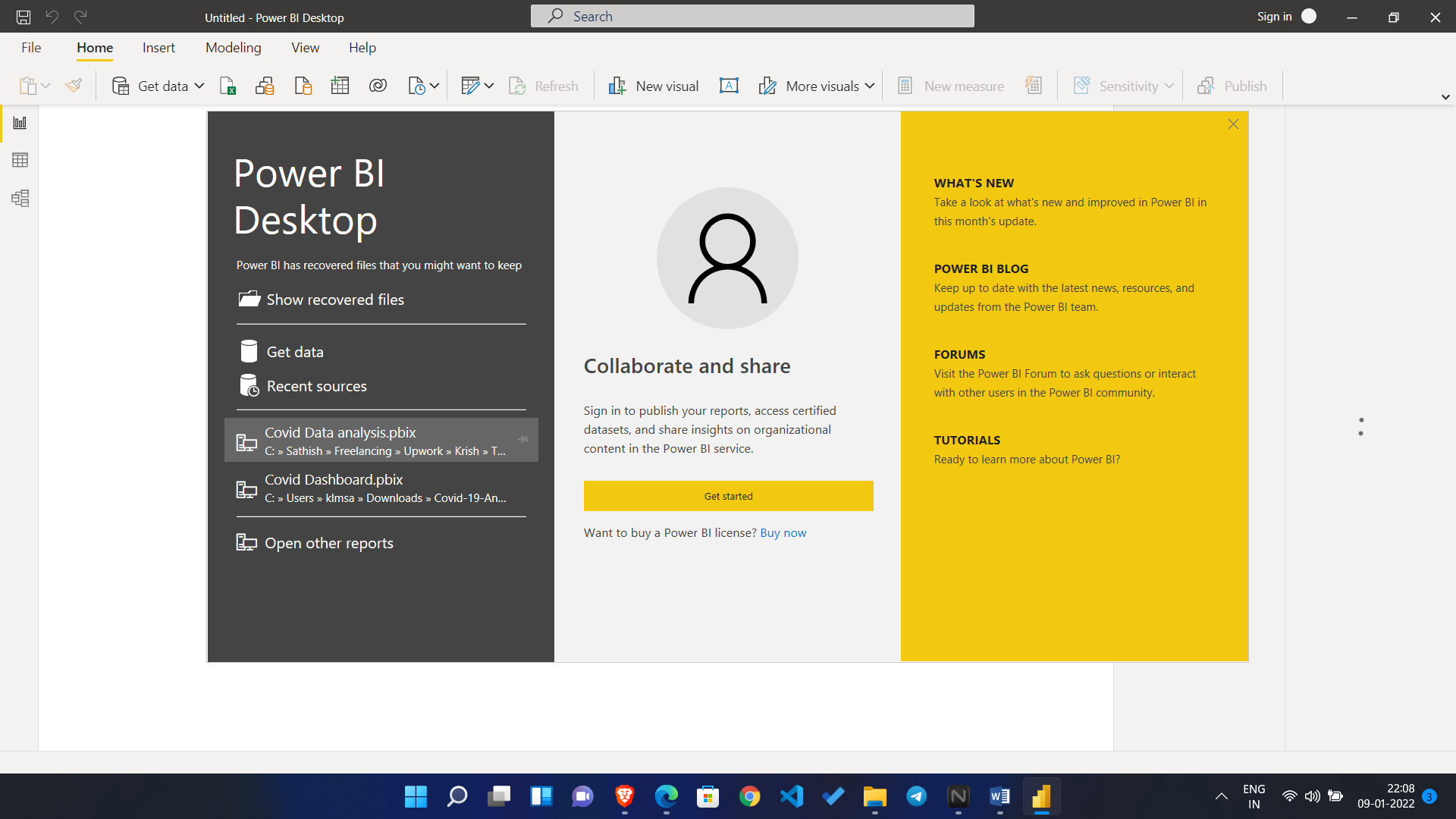


Fig: Power BI startup

For our project, we have to load dataset, dataset country\_vaccinations.csv which is in csv form by clicking “Get Data” menu and selecting “Text/CSV” option

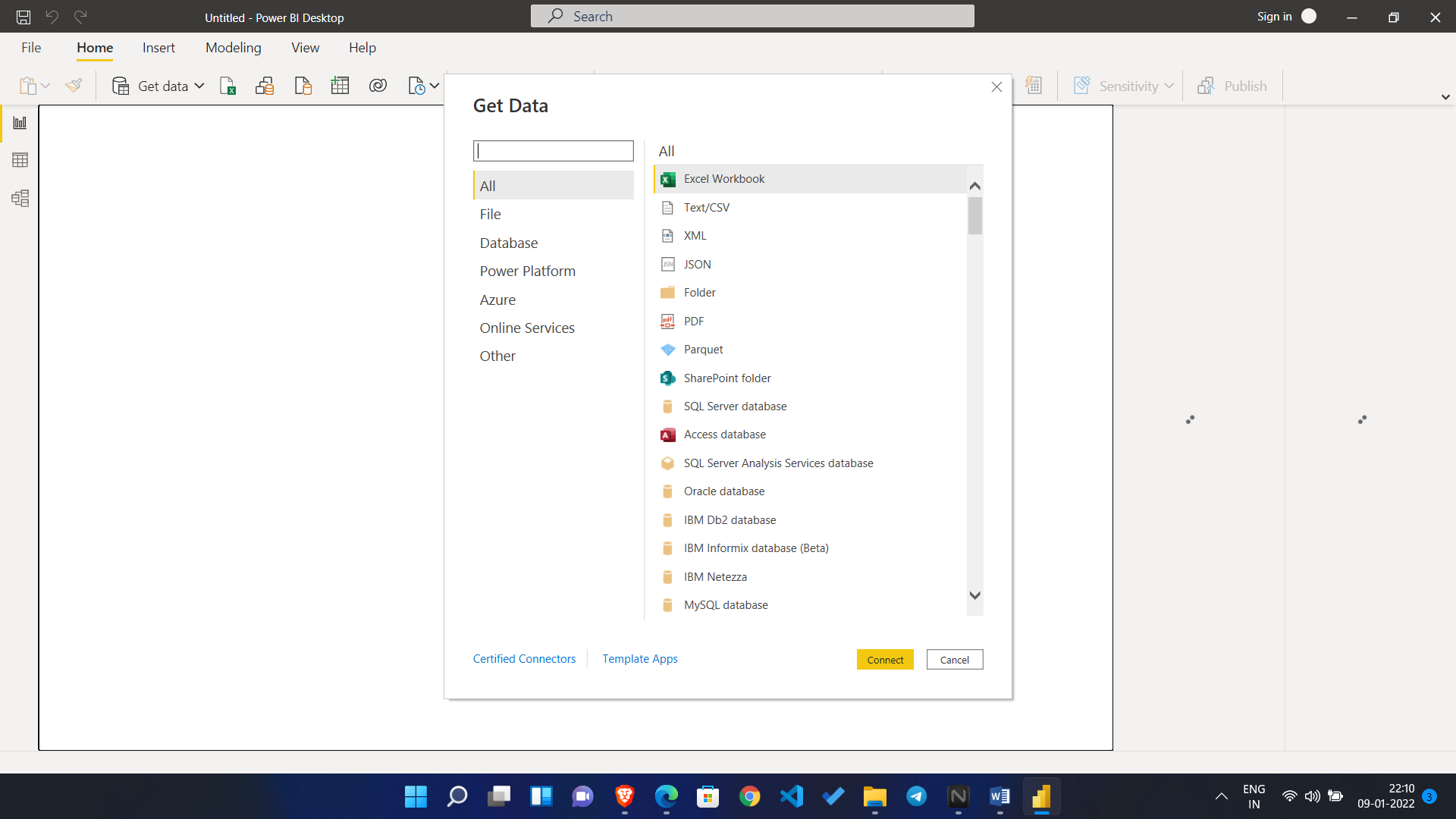


Fig: uploading Data Dialog box

Upon selecting the csv file, a dialog was popped out as shown in figure below giving an option to load or transform the data. The csv data was loaded by selecting ‘Load’ button without transforming the data at this moment. This step automatically determined the datatype of column of the table based upon the values of first 200 rows

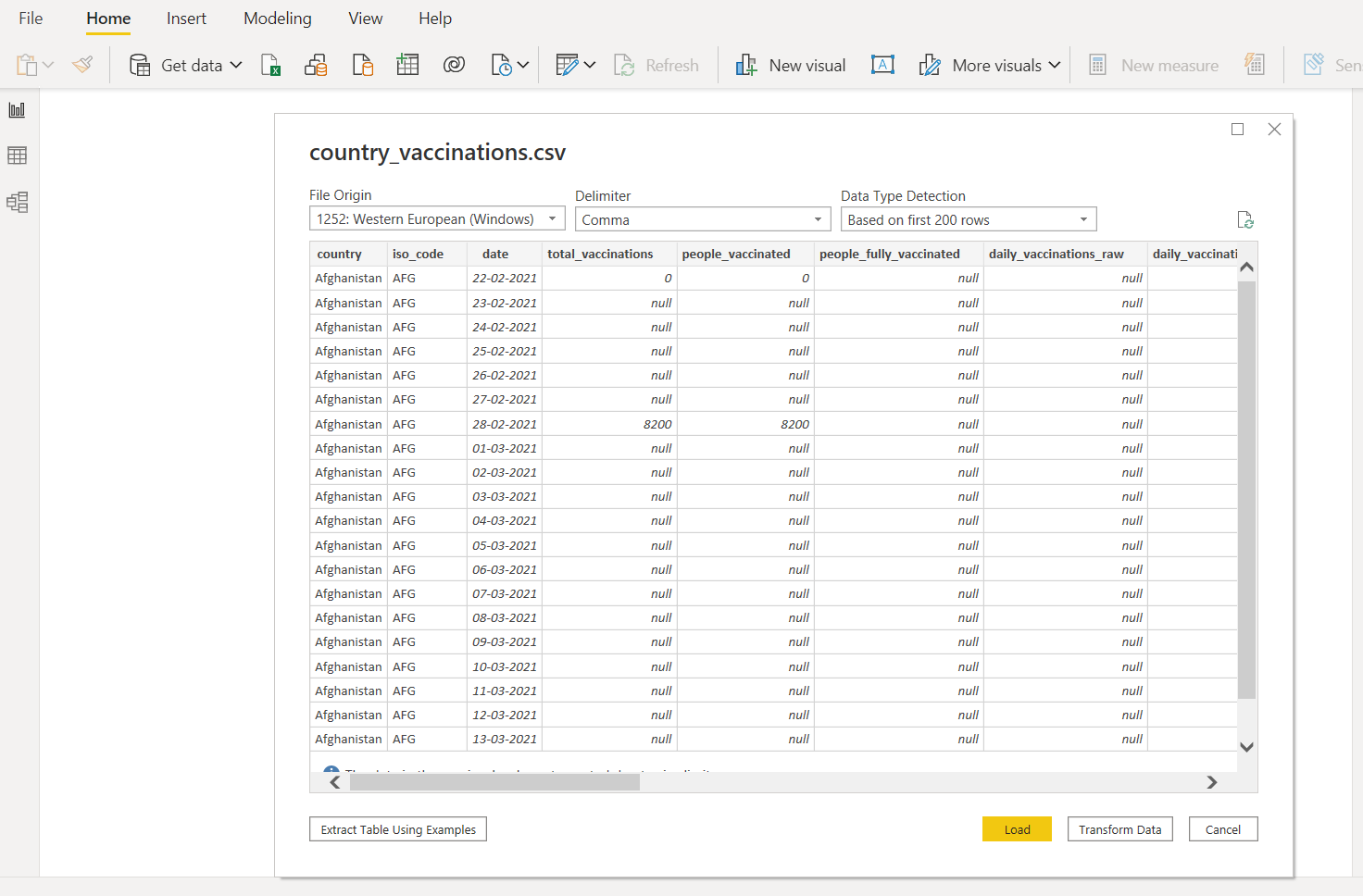


Fig: Loading csv file

# Data Cleaning

After importing, it is obvious to go for the data cleaning process.

Click on Transform Data under the Home tab and go to Power Query Editor.

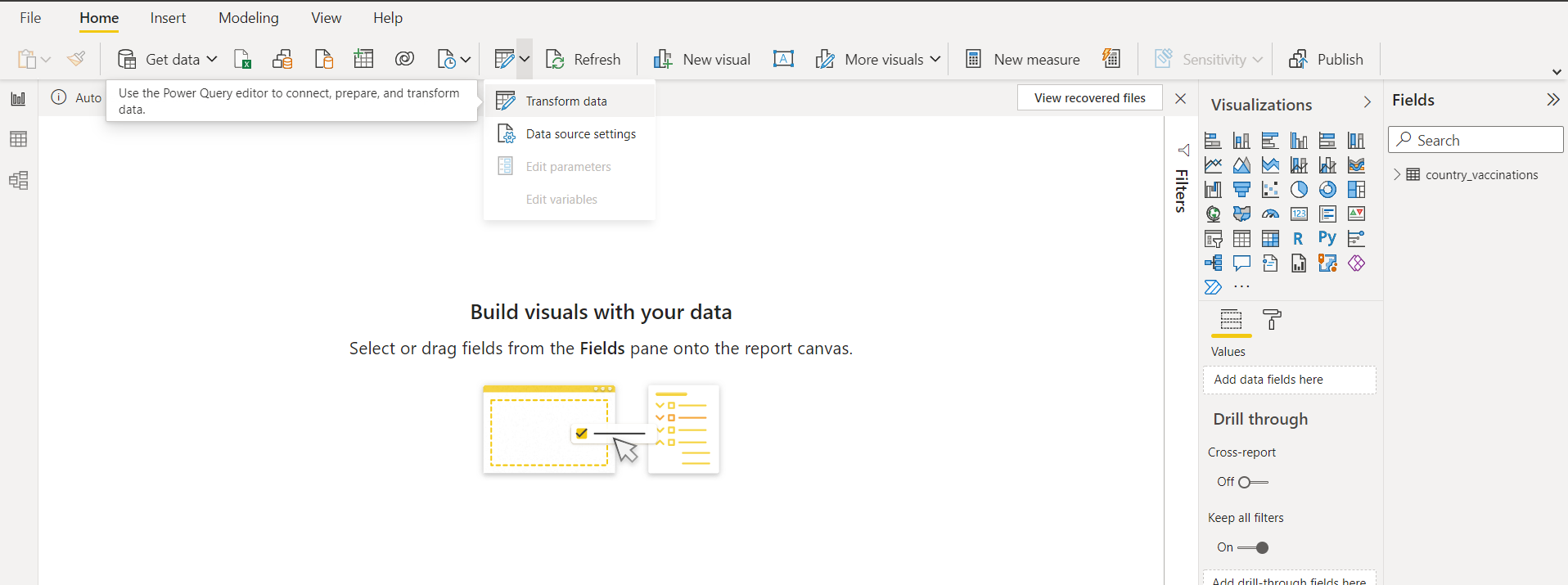


Fig: Entering Query Editor

In Power Query Editor, go to the View tab, enable Column Distribution, Column Quality and Column Profile.

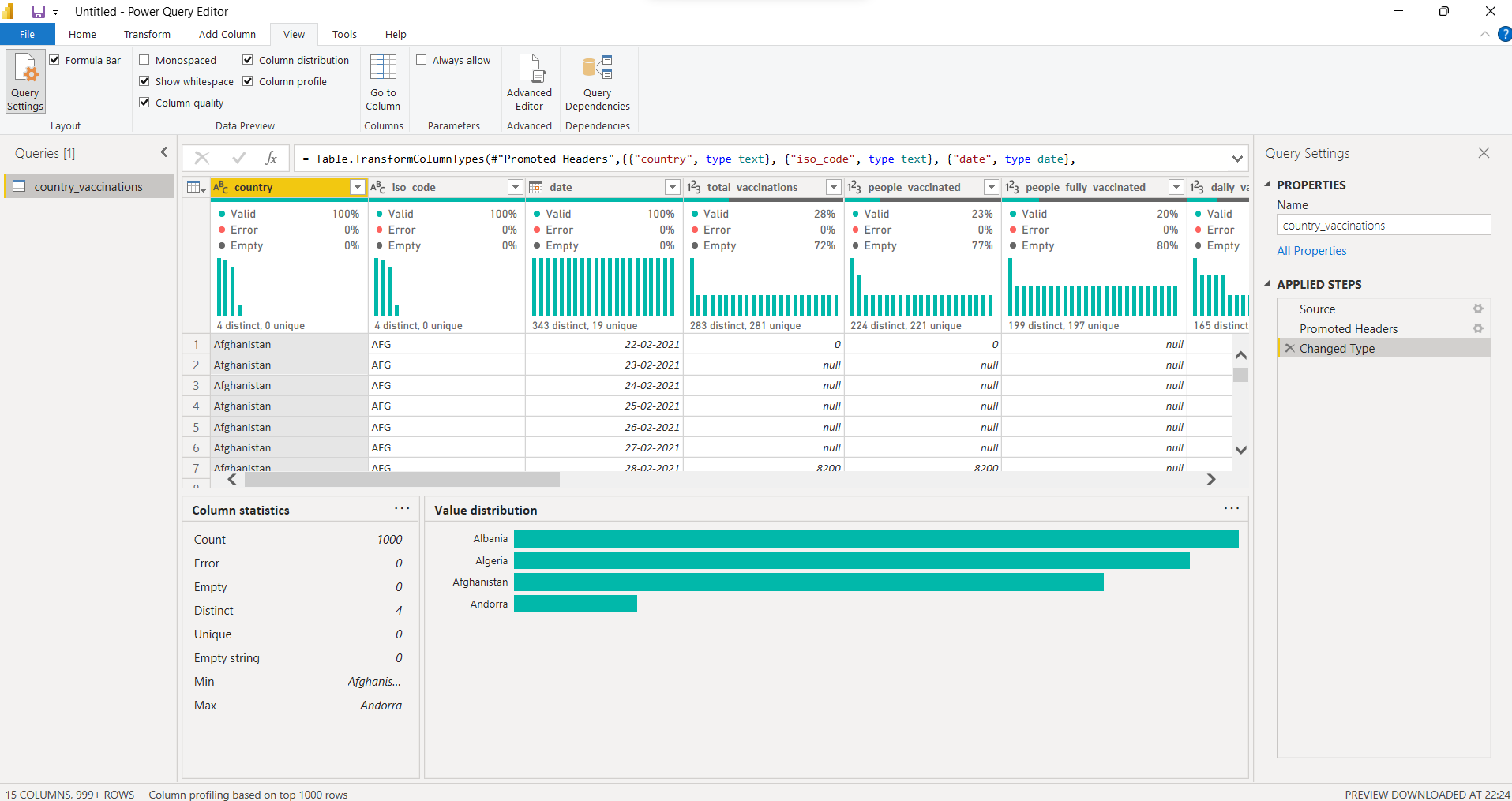
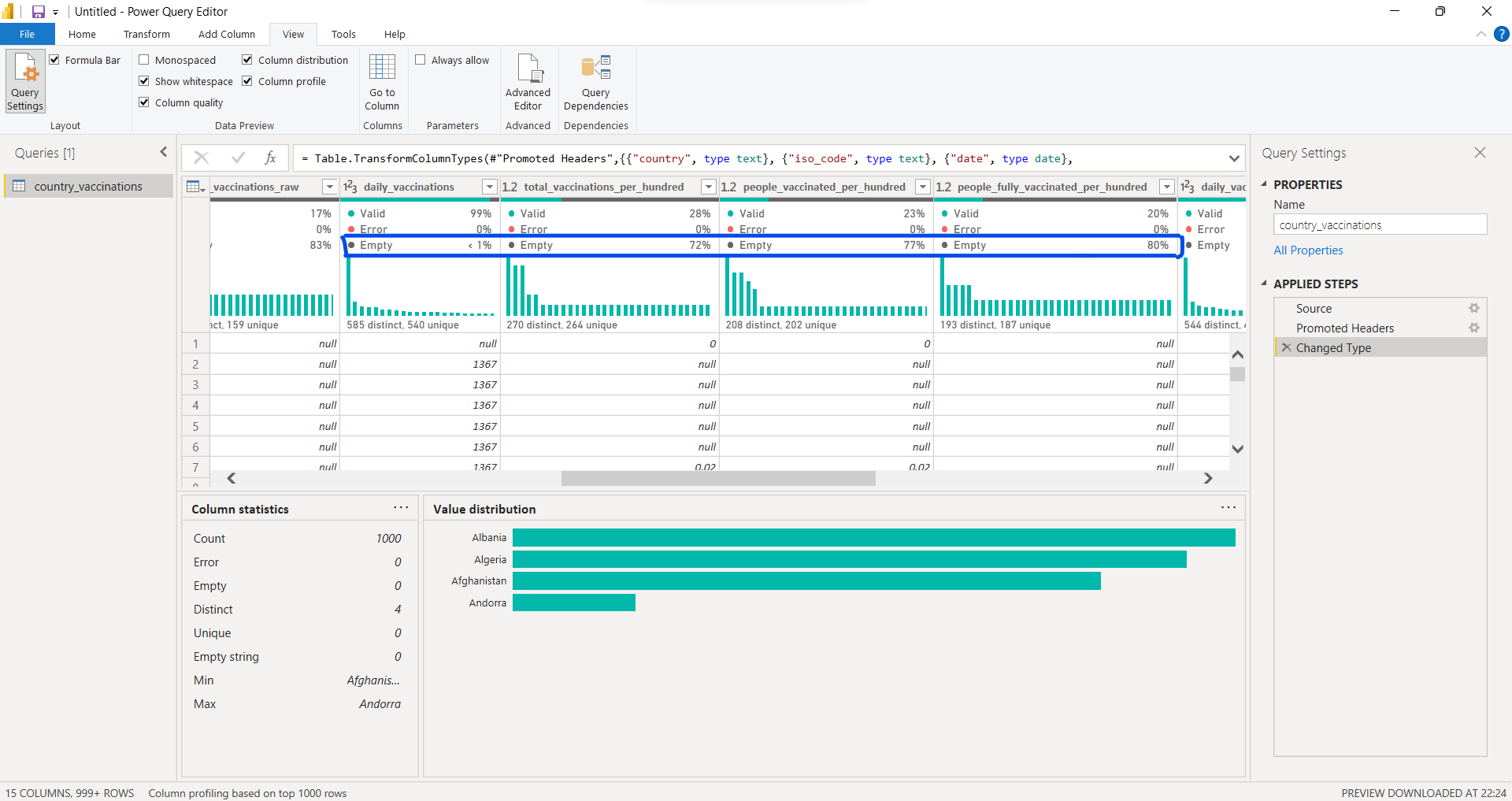


Fig: Viewing column distribution

It will help you to find out missing values, any data errors, any data type mismatch, any outliers etc.



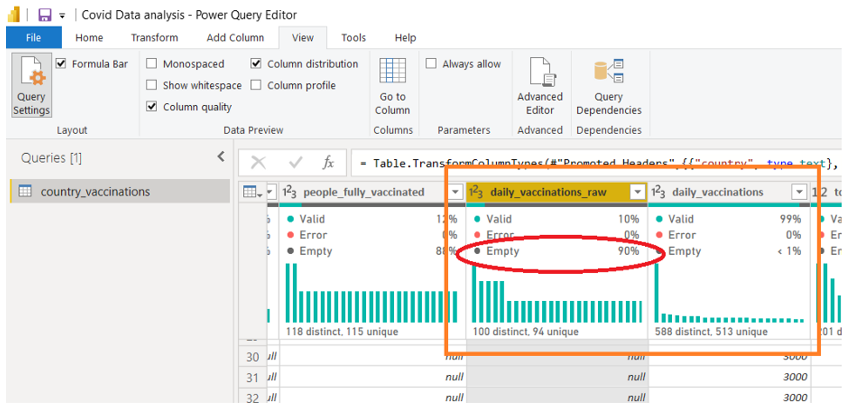
****

Fig: Missing value analysis

Based on the above findings, you can take appropriate actions. For example, in this data, we have total\_vaccinations\_per\_hundred which has 77% empty rows & daily\_vaccinations\_raw which has 90% empty rows that means it has missing values. Whereas daily\_vaccinations has less than 1% empty rows. So we can remove total\_vaccinations\_per\_hundred & Daily\_vaccinations\_raw as all columns have same purpose

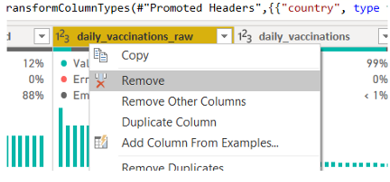


Fig: Removing unwanted columns

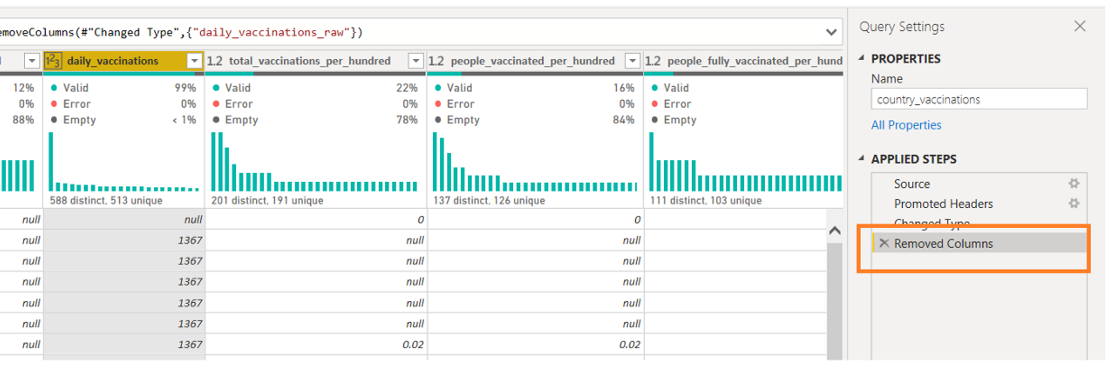
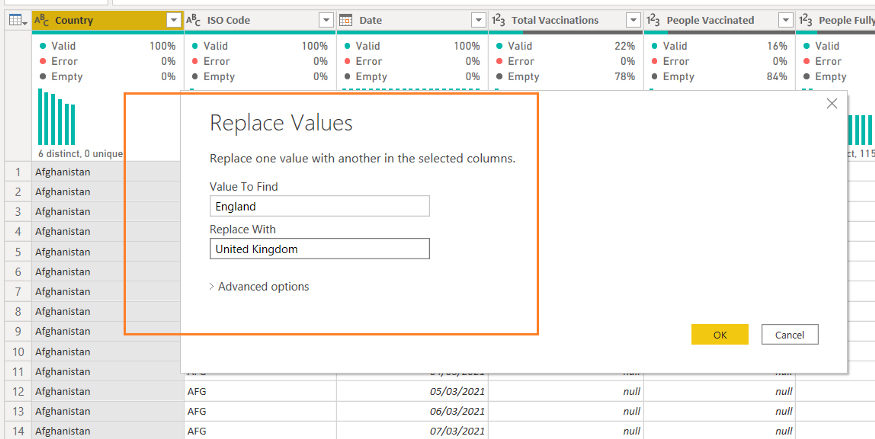


Fig: Viewing removed columns

Now Right Click on Country → Click on Replace Values and replace “England” with “United Kingdom”.



Now click on the Close & Apply button and return to the main Power BI Desktop pane

Create Date Table

First, create one Date table before proceeding to any calculation.

Here you will use the DAX function and this date table will help you to do Time Intelligence Analysis.

Go to the Modelling tab → Click on the New table

Write “Date Table = CALENDARAUTO()” and the automatic date table is now in place.

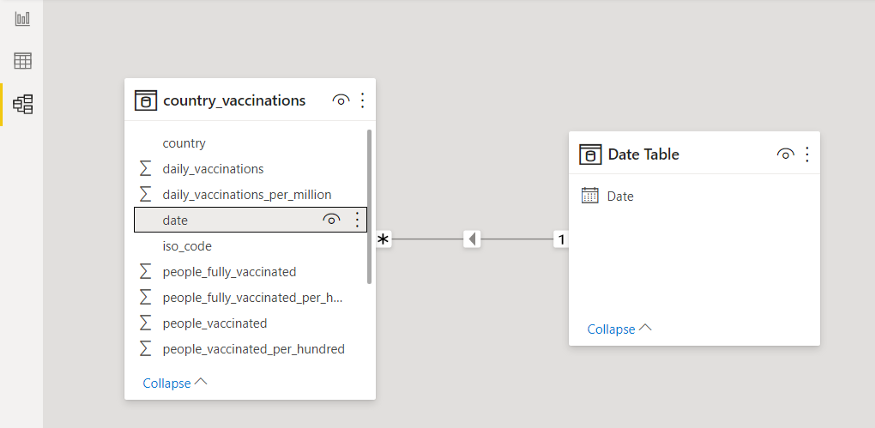
Create Relationship

Now you have two tables and it’s time to create a relationship between them.

Click on “Model” from the left side navigation bar.

Click on the date column of the country\_vaccinations table, then drag & drop to the date column of the Date Table.

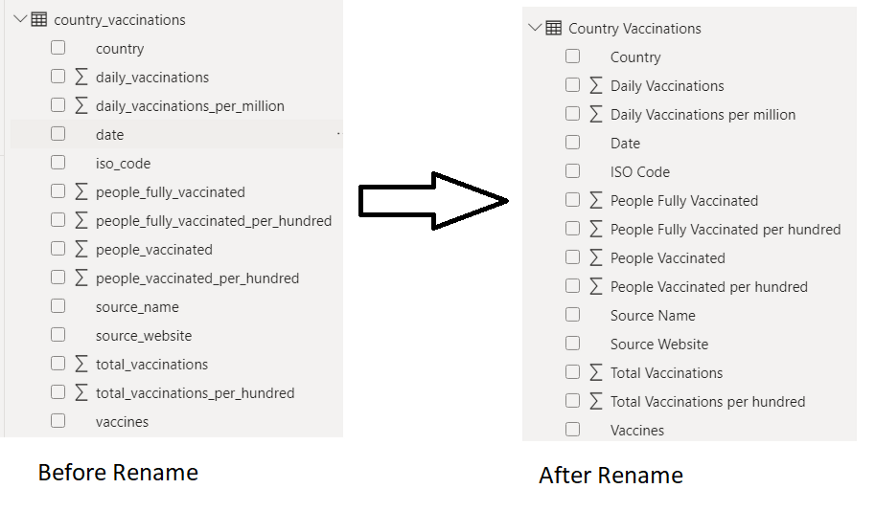
A many to one relationship is created.



Now you are ready to create visualizations. But one thing to remember, based on our visualization requirements, we will create different calculated measures or calculated columns.

Modify Column Name

Provide some proper names to all fields and tables. For example, remove underscore, words start with capital letters etc.



# **Select Theme**

Before proceeding to report, you can select one theme for your project. It will help you choose the proper colour combination. For each theme, there have some suggested colours, however, you can very well select any other colour also.

Go to View tab -> Under Themes Select Executive ( you can select based on your choice)

**Create Snapshot View**

For any summary report or dashboard, it is a good idea to have some snapshot views.

At a glance, the user will get some idea about the current scenario of the business/data.

1. Click on Card visual → Added to the canvas area.
2. Select field P**eople Fully Vaccinated**
3. Click on Format → Go to Data label and Category Label. Change Color, Font family and Text size. Add some background color to it.

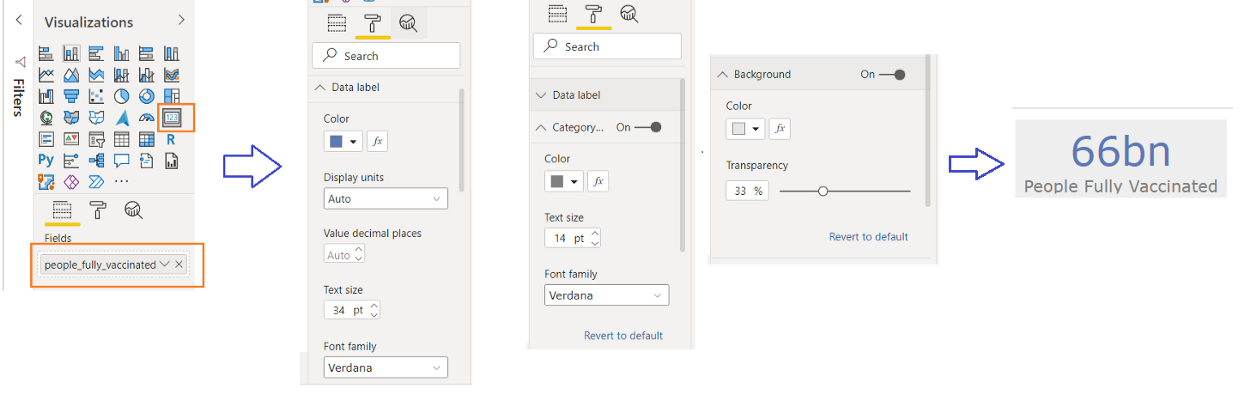


Fig: Creating Snapshots



Fig: Snapshots

4. Follow the same process for Total Vaccinations, People Vaccinated and Total Country.

5. To derive Total Country → Select the Country column and change to Count (Distinct) from the drop-down.

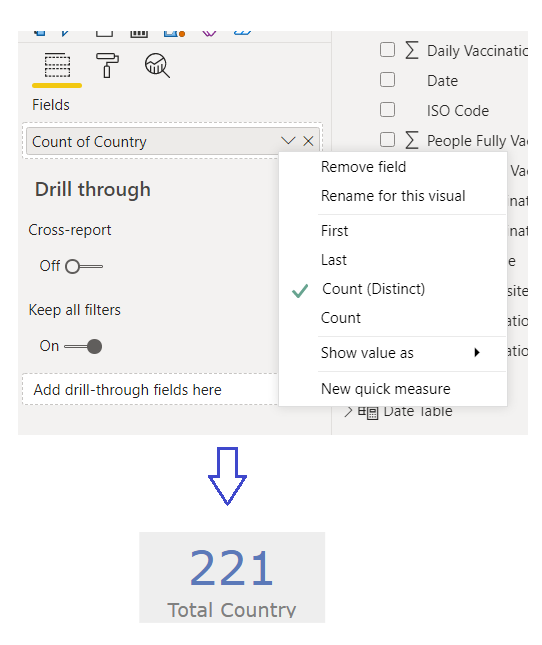


Fig: Creating Snapshots for other columns

Using Format painter, copy the same format for all the Card visuals.



Fig: Snapshots for all card visuals

# Add Year, Month Slicer

1. Add Slicer visual beside the card visuals.
2. Add Date Hierarchy → Keep only Year and Month.
3. Normally people are interested to know how the vaccine process is going month on month.

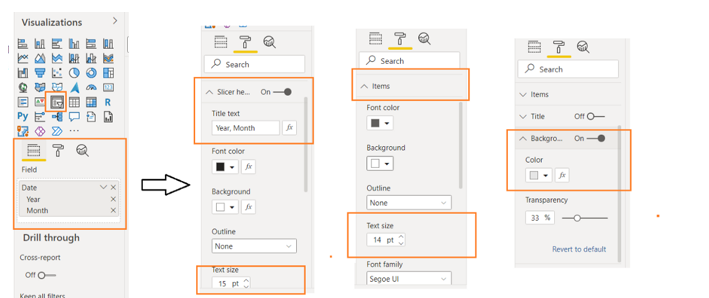


Fig: Formatting Year, Month

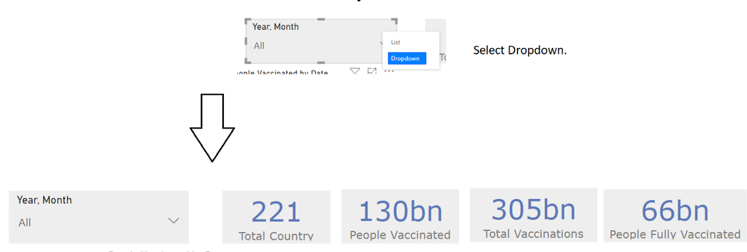


Fig: Adding customizations on sorting based on year, month



0

AUTHOR NAME  AUTHOR NAME

CIS4008-N Big Data and Business Intelligence

SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL TECHNOLOGIES

Business Intelligence Solution and Report

Section 2

Creating Line Graph on “People Vaccinated by Date”

1. Add **Line**Chart to the canvas area.
2. Add **Date**in Axis and **People Vaccinated**in Values. As it is a **trend**analysis (based on date, that’s why it is called trend), it is preferable to use a line graph to show how data varies over time.
3. In the Format section, you can do the following changes

a) Choose one Data Colors,

b) Enable Data labels, update Display units based on your choice so that values can be visible properly and easy to follow the data variations. But across all reports, try to keep the same display units, it will help any user to understand the data variations.

c) If you want, you can modify the title, or font size or different text style.

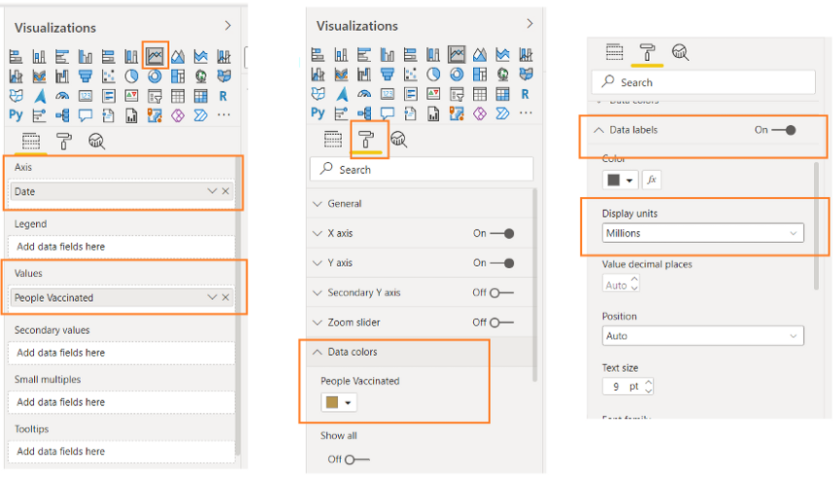


Fig: Customizations in Line graph

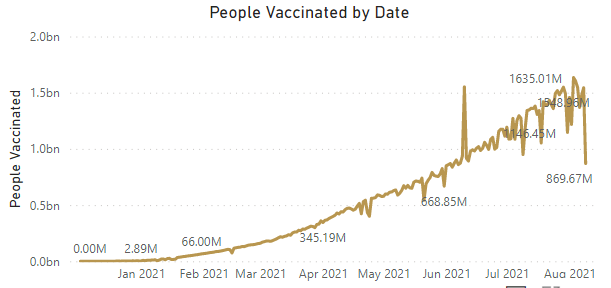
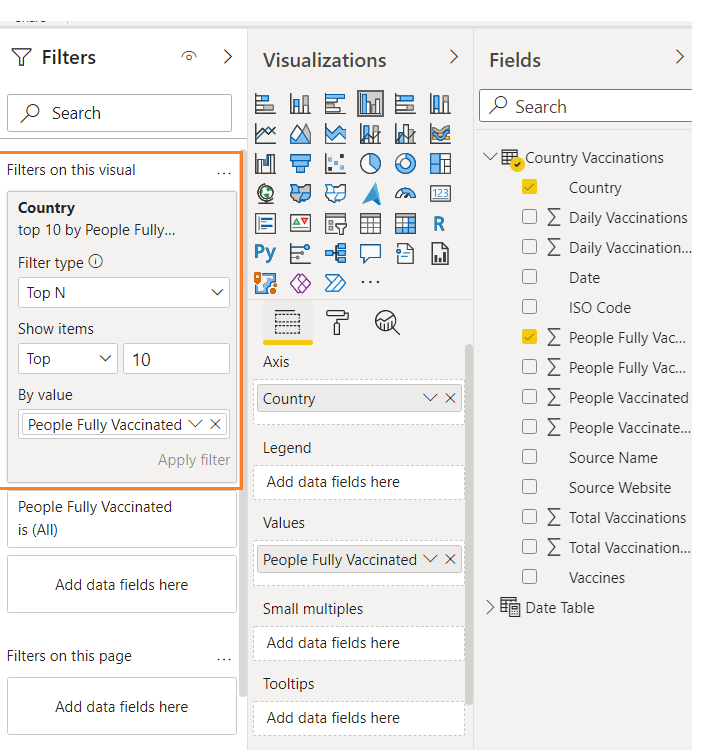


Fig: Final Output

Creating Clustered Column Chart on “Top 10 Country by People Fully Vaccinated”

1. Add **Bar**Chart to the canvas area.
2. Add **Country**in Axis and **People Fully Vaccinated** in Values.
3. To display any **comparison analysis**, it is preferable to use a **clustered column chart.**
4. In the **Filters**section, select **Top N**filter type from **Country**, add show items **10** and by value **People Fully Vaccinated.**



5. In the Format section, you can do the following changes

a) Choose one Data Colors,

b) Enable Data labels, update Display units based on your choice so that values can be visible properly and easy to follow the data variations.

c) If you want, you can modify the title, or font size or different text style.

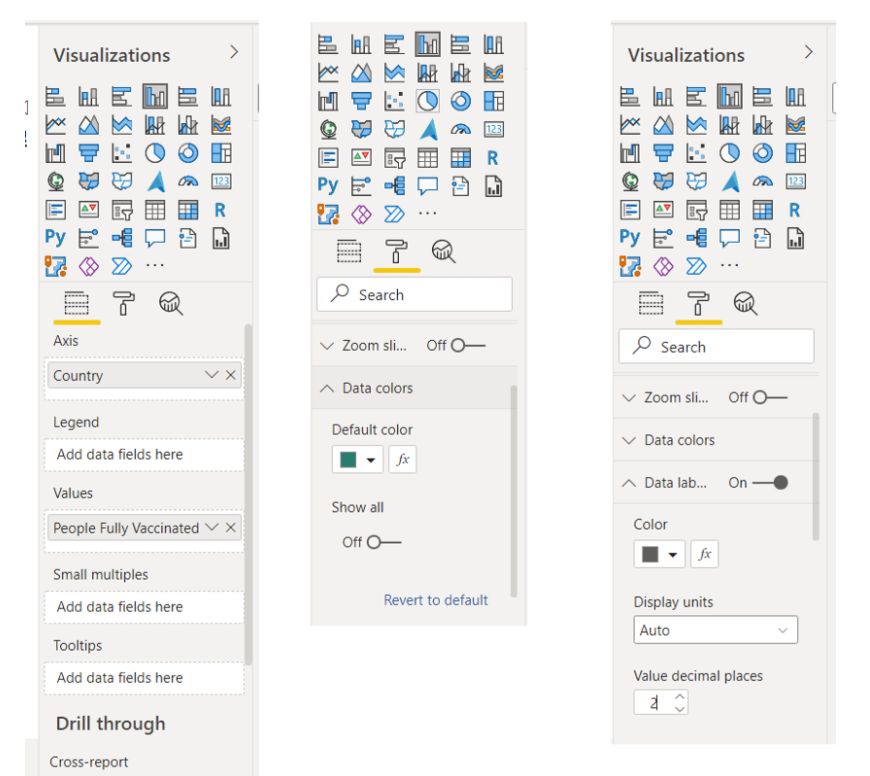


Fig: Customizations in Line graph

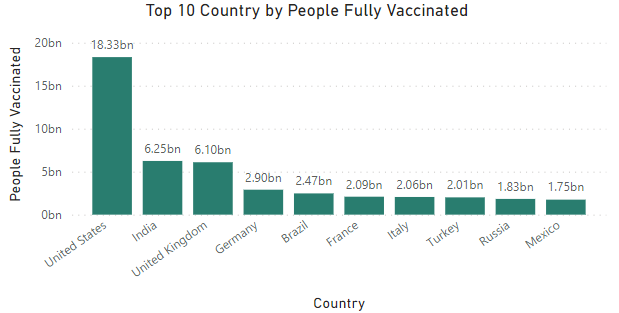
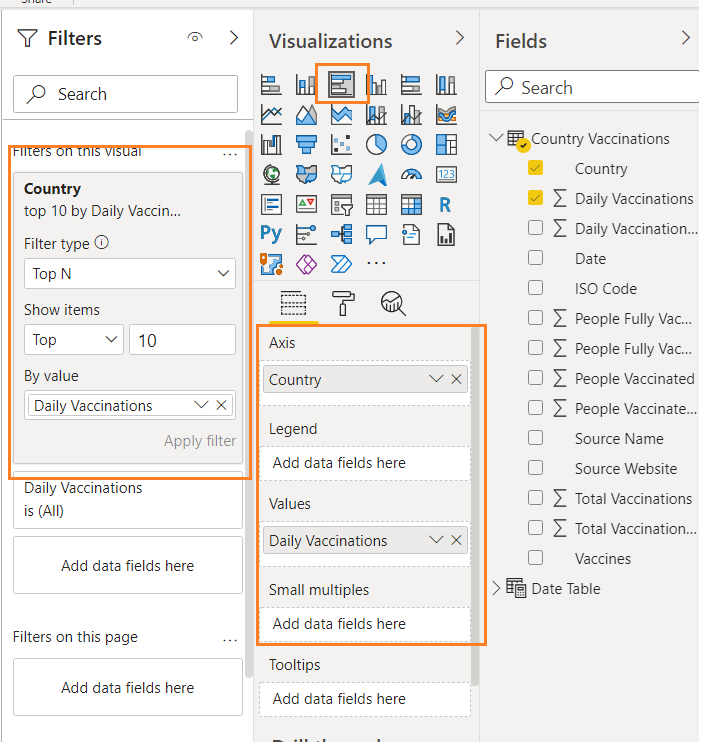


Fig: Final Output

Creating Clustered Bar Chart on “Top 10 Country by Daily Vaccinations”

1. Add **Bar**Chart to the canvas area.
2. Add **Country**in Axis and **Daily Vaccinations**in Values.
3. As it is also a **comparison analysis**, so you can use a **bar**graph. This time I am using Clustered Bar Chart.
4. In the **Filters**section, select **Top N**filter type from **Country**, add show items **10** and by value **Daily Vaccinations.**



5. In the Format section, you can do the following changes

a) Choose one Data Colors,

b) Enable Data labels, update Display units based on your choice so that values can be visible properly and easy to follow the data variations.

c) If you want, you can modify the title, or font size or different text style.

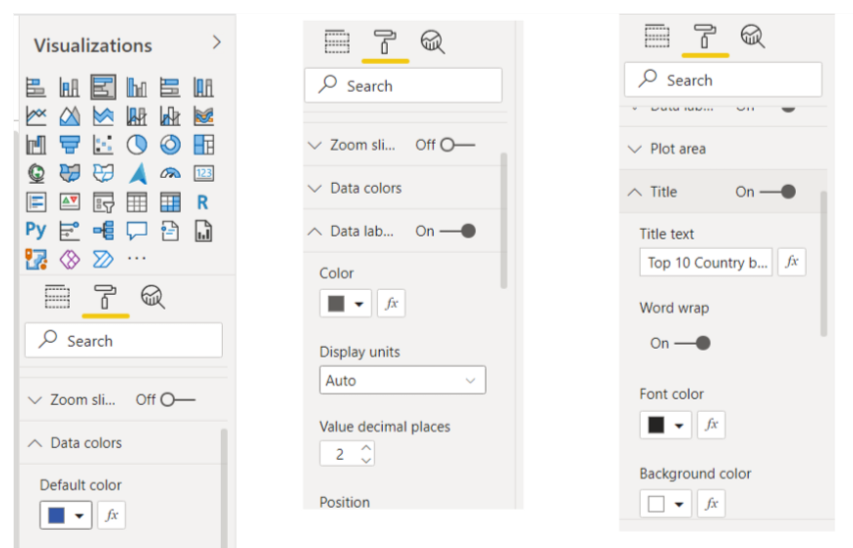


Fig: Customizations in Line graph

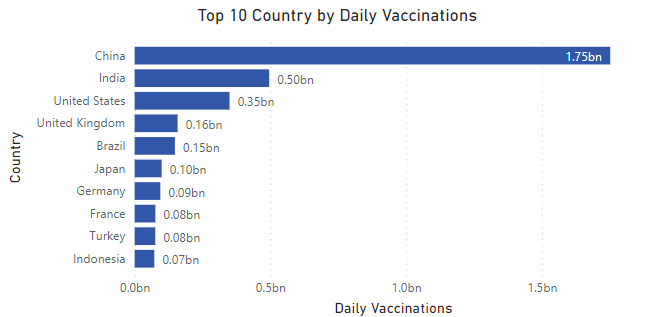


Fig: Final Output

Creating Map Visual on “Total Vaccinations by Country”

1. Add **Map visual**to the canvas area.
2. Add **Country**in **Location**and**Total Vaccinations** in **Size**.
3. In this analysis, you want to get some idea overall vaccinations process across all countries. For this, map visual is ideal.
4. In the Format section, you can do the following changes

a) Choose one **Data Colors**

b) Change the**Map styles** to Grayscale, update **Bubbles**size

c) If you want, you can modify the title, or font size or different text style.

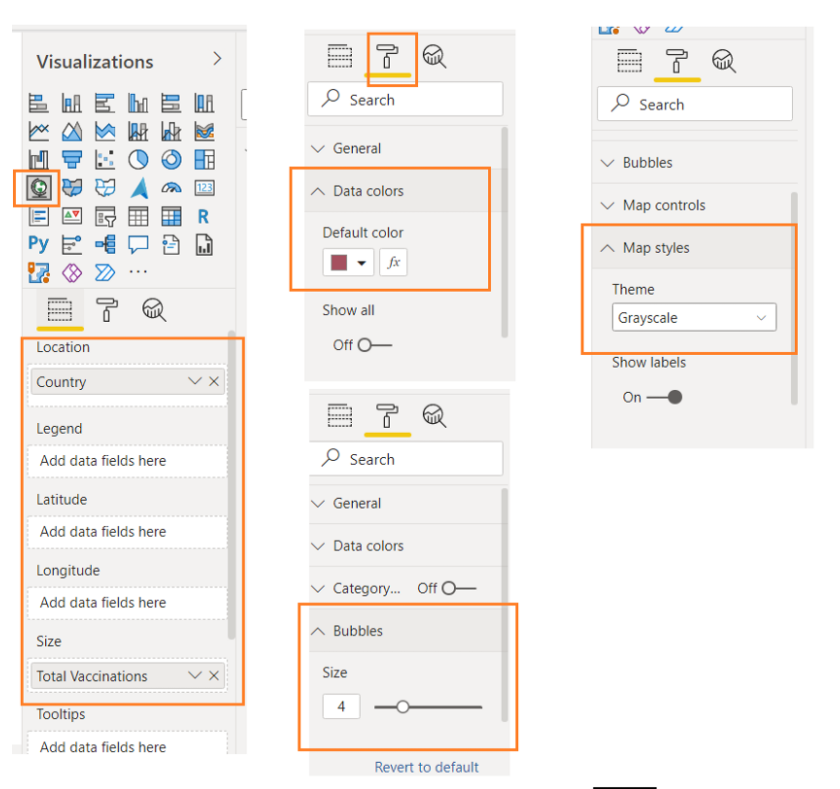


Fig: Customizations in Line graph

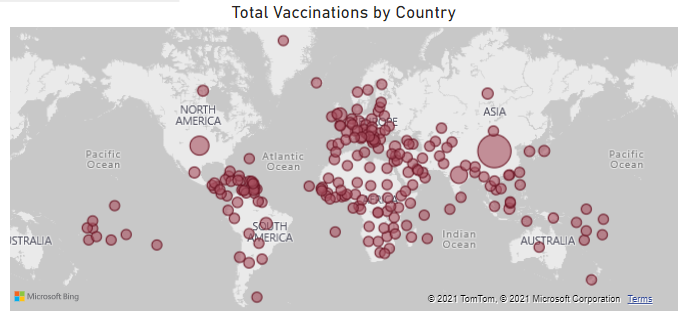
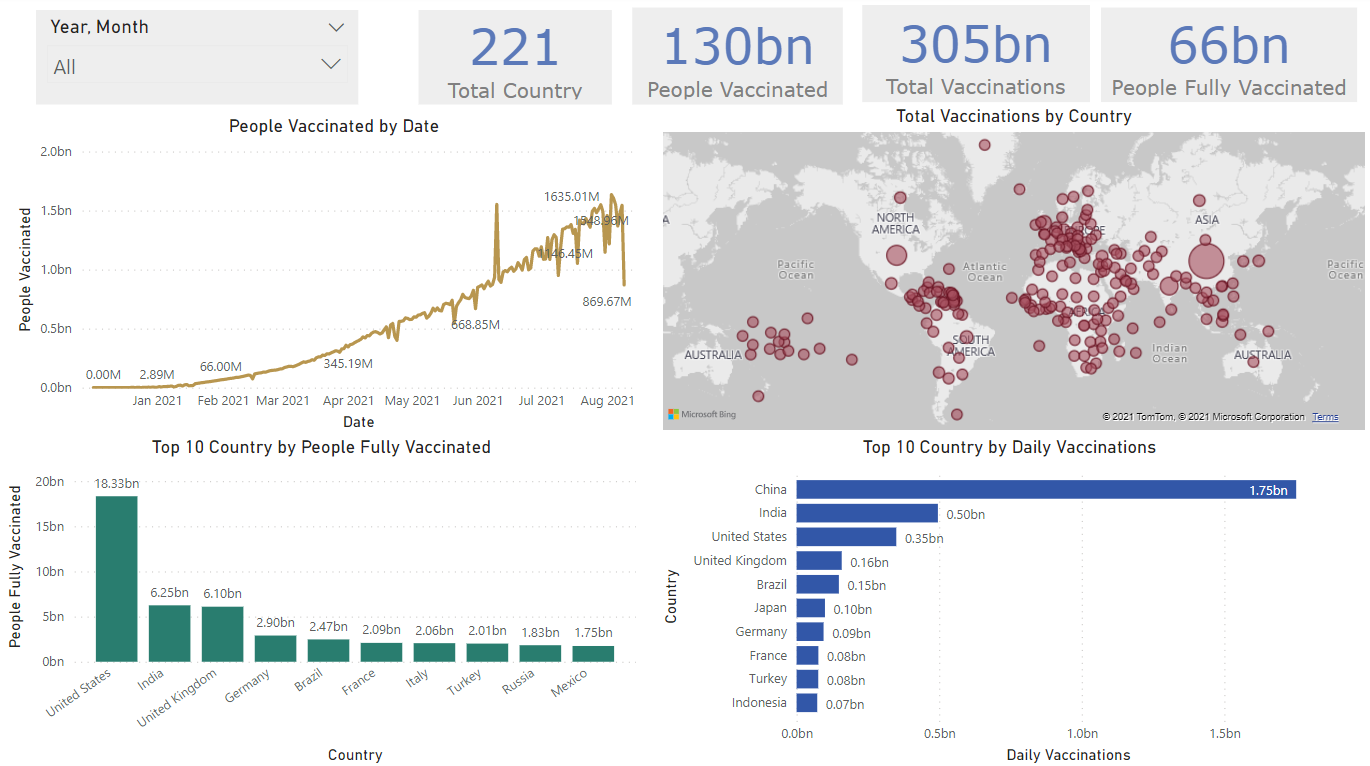


Fig: Final Output

Insights:

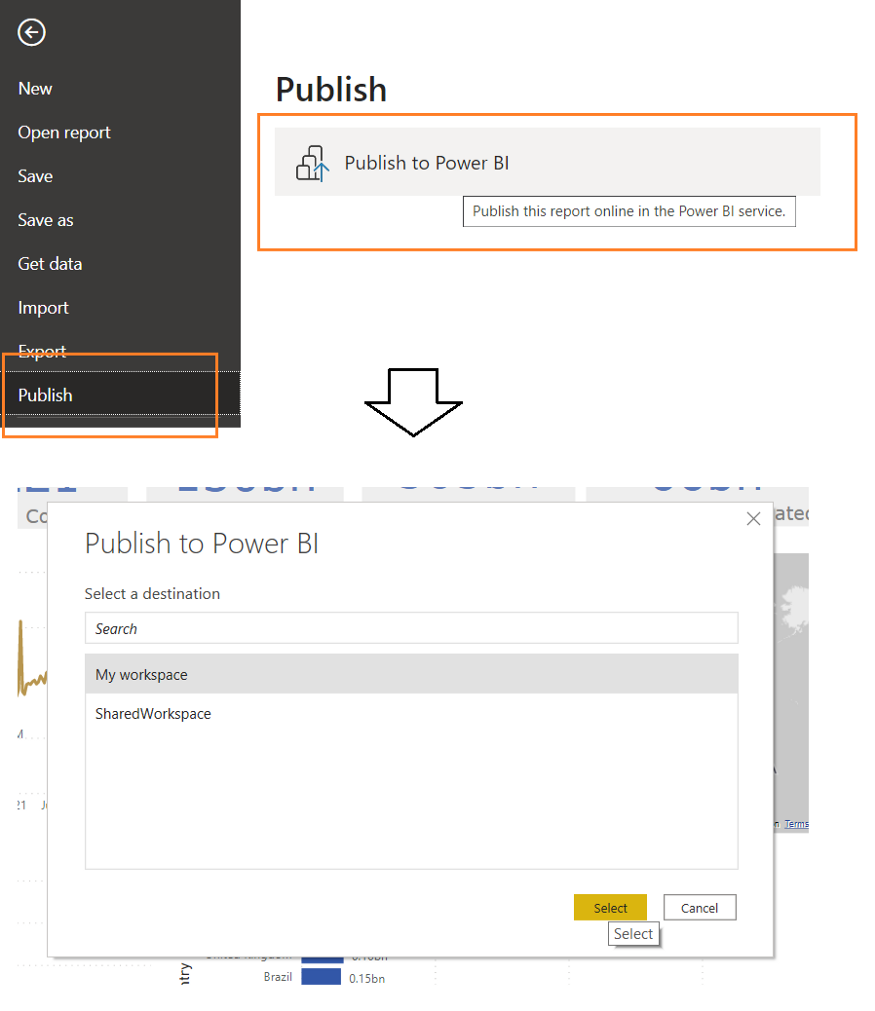
1. From this line graph, we can analyse that around June 10th “people vaccinated all over the world count” is higher than any time and the count gets higher exponentially in upcoming months and hits a saturation level around august.
2. Based on Clustered Column Chart on “Top 10 Country by People Fully Vaccinated”, we can see that United States stands top in the list of high vaccination rate and India, United Kingdom, Germany followed it.
3. From the Clustered Bar Chart on “Top 10 Country by Daily Vaccinations” we can analyse that china has daily vaccination count around three times as of India and most of the European countries are in top 10 Countries.
4. To be more elaborate in vaccinations by country all over the world, we can see that African countries have same amount in number and all the European countries have good amount in number of vaccinations. In Asia, China & India contribute more in number. Also some countries in South America like Brazil have very less count in vaccinations compared to their population and size.

Dashboard

****

# Publishing to Power BI Service

1. Go to the File tab → Click on Publish
2. Click on Publish to Power BI
3. Select a destination workspace and click on the Select button
4. Now your report is published to Power BI Service.



# **Open Report from Power BI Service**

1. Type in browser → <https://app.powerbi.com/>
2. Enter your credentials.
3. Go to the workspace where you have published.

Conclusion

COVID-19, a new and sometimes deadly respiratory illness has spread rapidly throughout that country and the world. The new coronavirus was first detected in Wuhan, China in December 2019. Tens of thousands of people were infected in China, with the virus spreading easily from person-to-person in many parts of that country. The novel coronavirus has now established itself in 177 countries and territories around the world in a rapidly expanding pandemic. Health officials in the United States and around the world are working to contain the spread of the virus through public health measures such as social distancing, contact tracing, testing, quarantines and travel restrictions. So these kind of statistics would be helpful to prevent corona virus as it gives us the whole picture of how many people gets vaccinated in each country and how people feel about the reliability of vaccines.