**Data Analysis — Case Study in Power BI — Covid Vaccination Process**

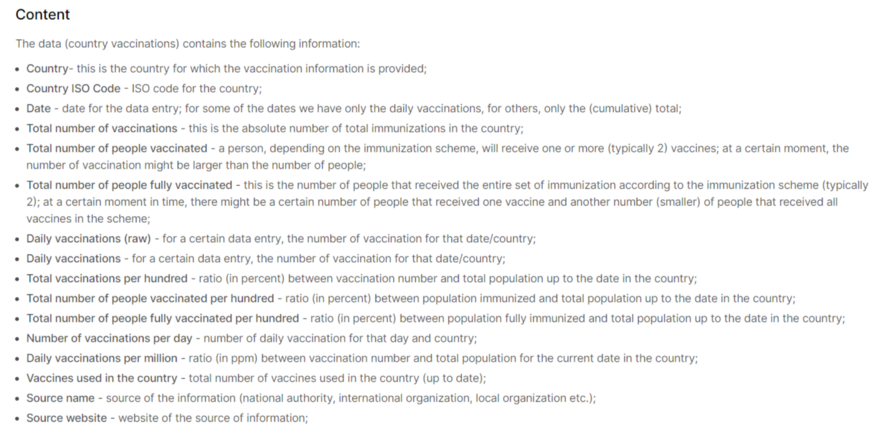
1. **Introduction**

We are going to explore COVID vaccine data and try to find out the inference of how the immunization process is going on worldwide.

1. **Source of Data**

For the case studies, Kaggle is a good source of data. So, we are considering below mentioned dataset.

* It has two files, but we will use one file named “country\_vaccinations.csv” for this case study.
* If you want to know more in detail about each field, you can have a look at the below picture. Though all the fields are self-explanatory.



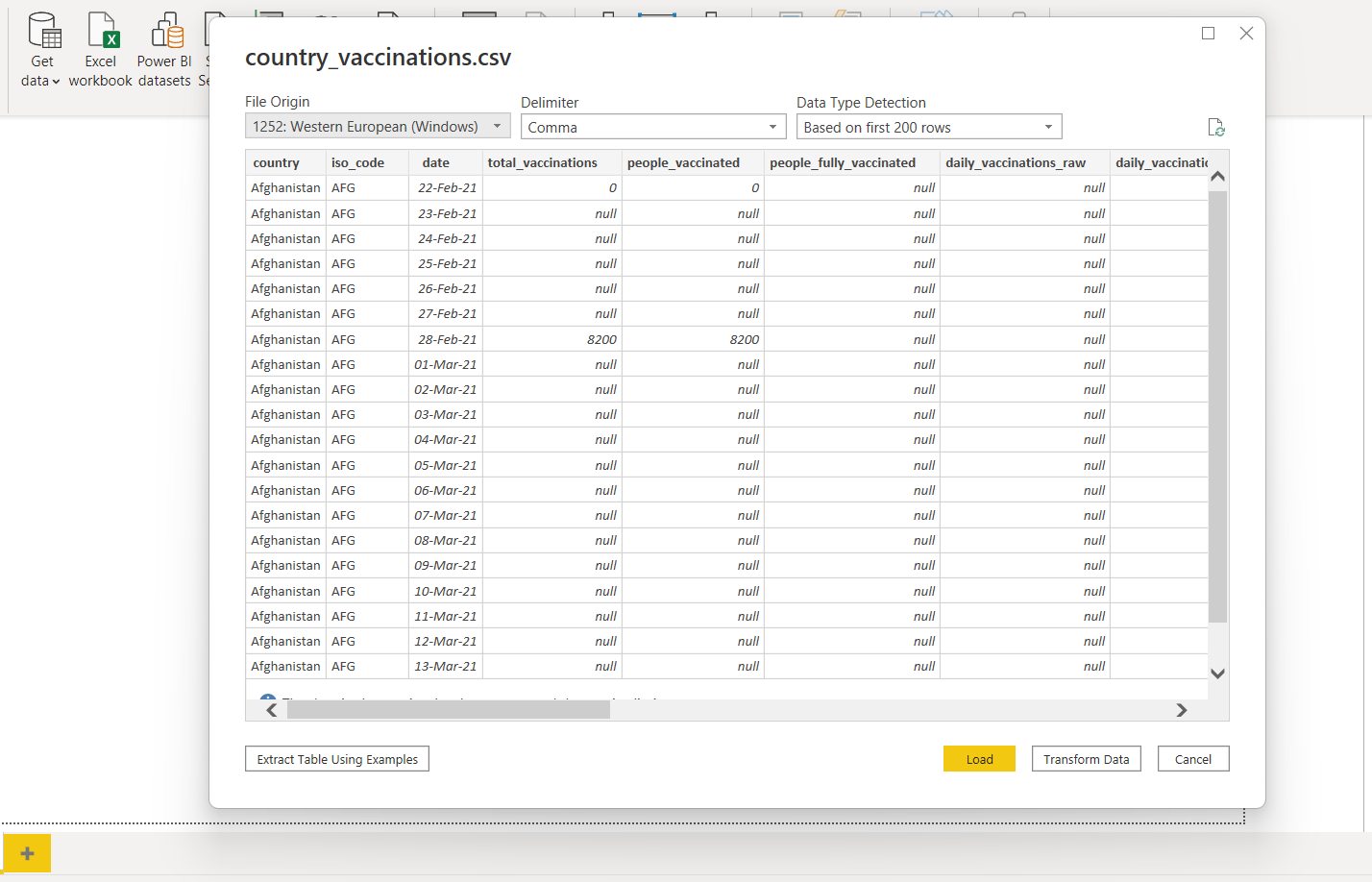
1. **Our Objective**

The objective of this case study is to follow the primary steps of a data analysis project.

* Data Understanding
* Data Import and Cleaning process
* Create Relationship
* Choose recommended graphs for visualization and know the purpose of those graphs
* Publish to Power BI Service

1. **Import Data**

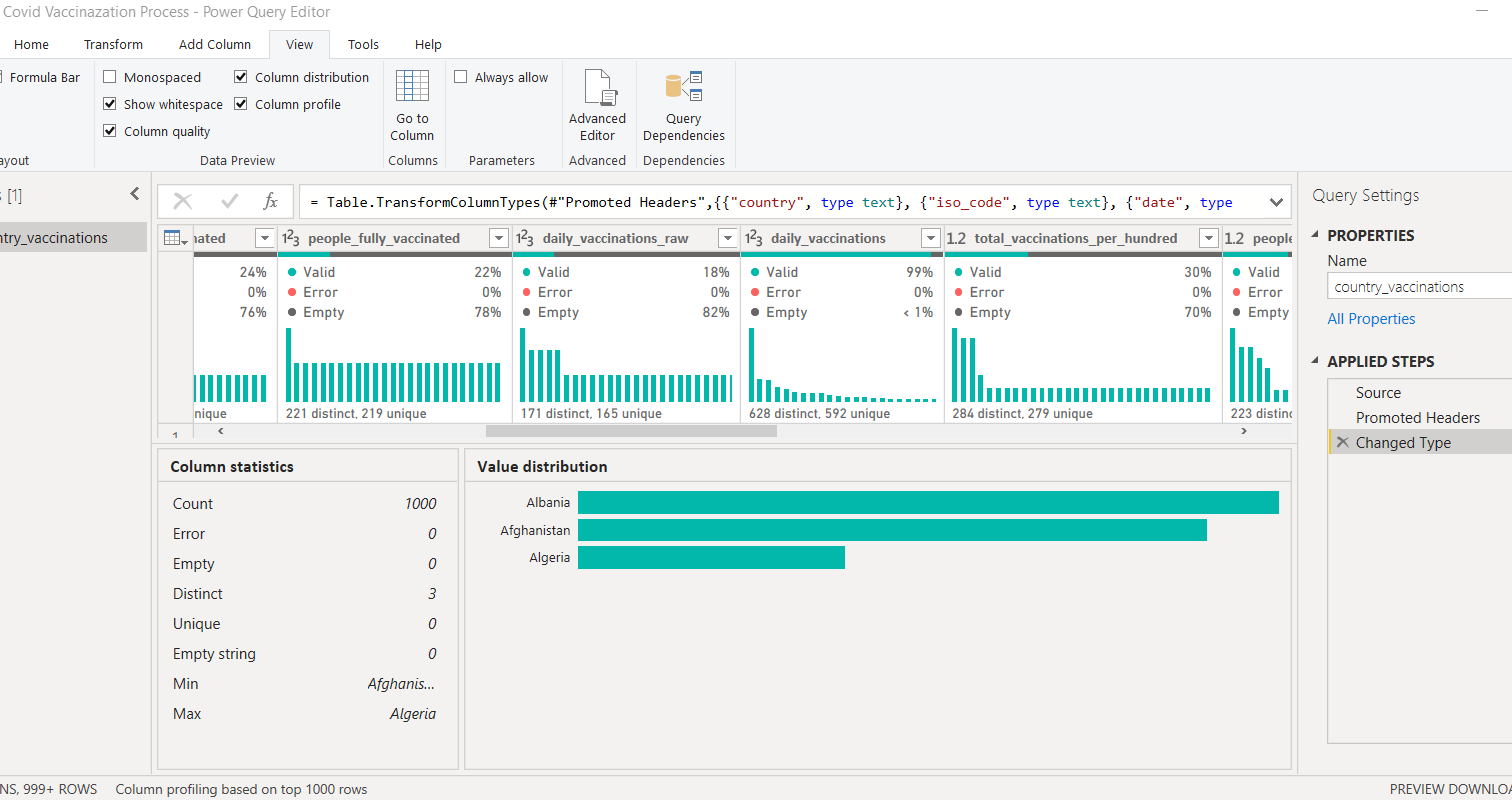
* Let’s start with the Get Data option under the Home tab. As this is a CSV file, select the Text/CSV option from the drop-down list
* Select the file named country\_vaccinations.csv
* After selecting the file, data will be displayed in the below format
* Click on Load and save data.



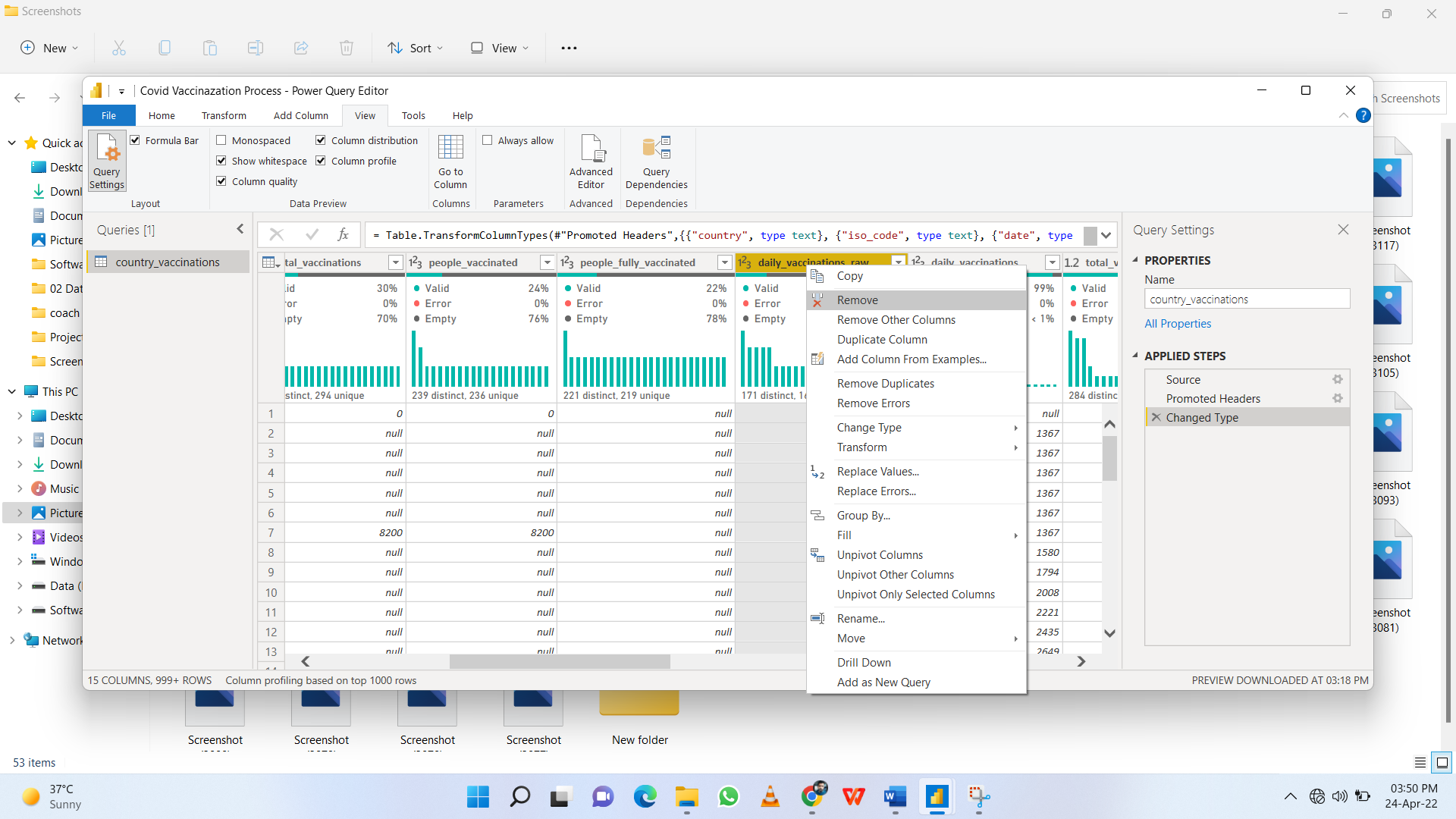
1. **Data Cleaning**

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

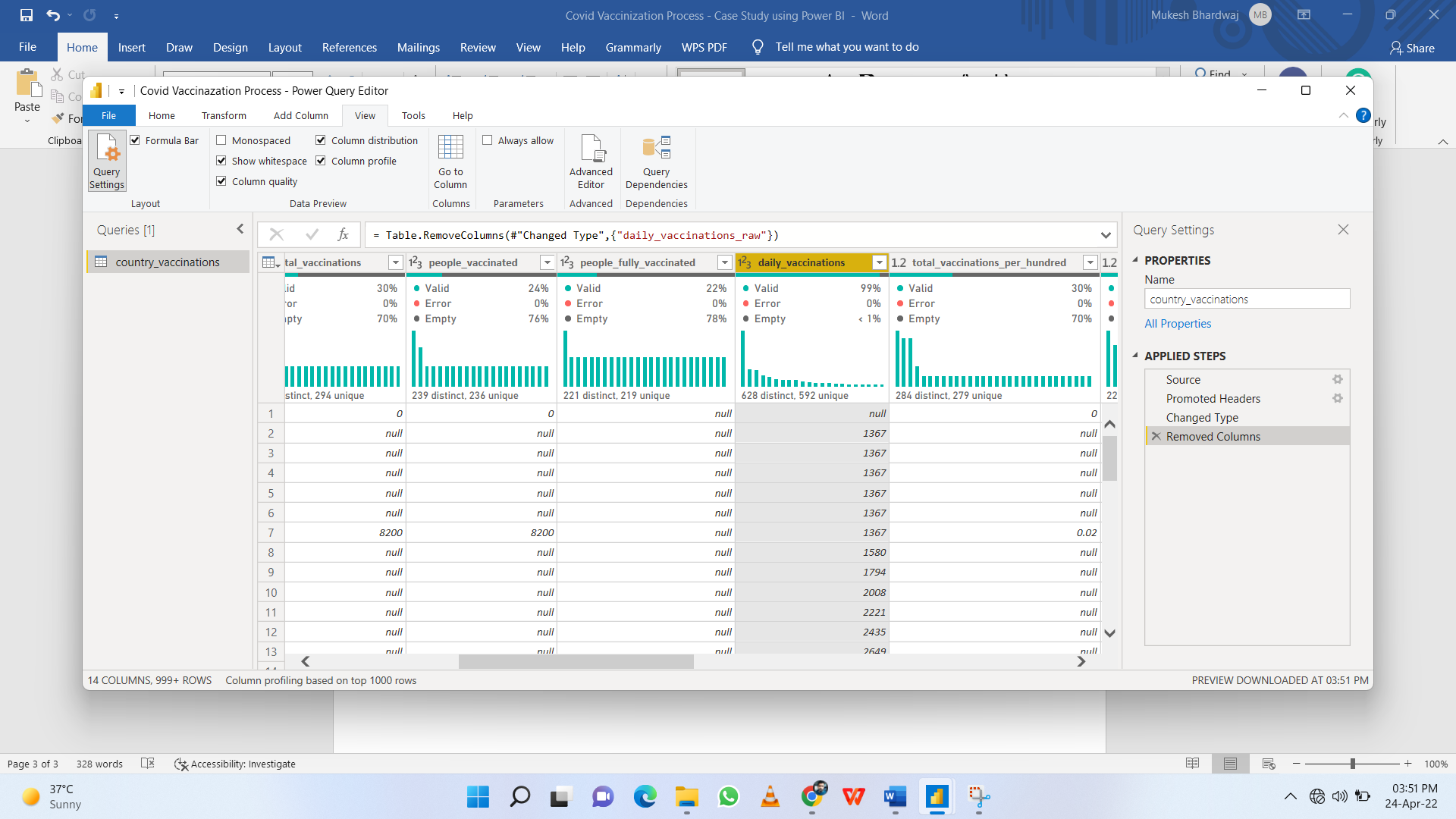
* Click on Transform Data under the Home tab and go to Power Query Editor.
* In Power Query Editor, go to the View tab, enable Column Distribution, Column Quality , and Column Profile.
* It will help you to find out missing values, any data errors, any data type mismatch, any outliers , etc.
* Based on the above findings, you can take appropriate actions.
* For example, in this data, we have daily\_vaccinations\_raw which has 82% empty rows which means it has missing values. Whereas daily\_vaccinations have less than 1% empty rows. Both columns have the same purpose. So, we can remove daily\_vaccinations\_raw.



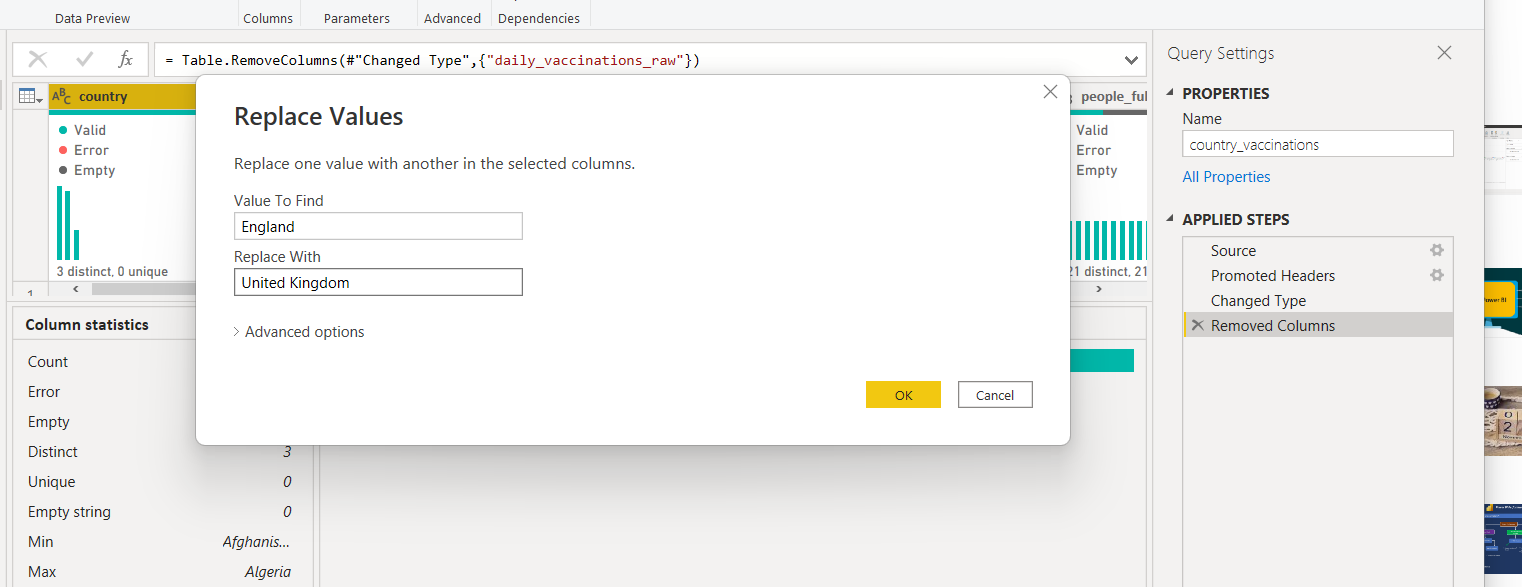
**Right Click on the Column**



**Click on Remove & added one step under APPLIED STEPS**

****

* 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.

1. **Create Date Table**

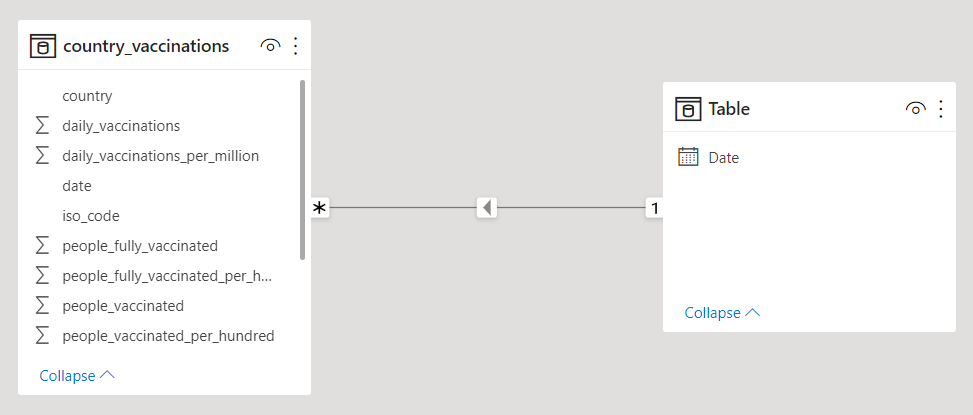
First, create one Date table before proceeding to any calculation.

* Here you will use the [DAX function](https://medium.com/analytics-vidhya/automatic-date-table-with-additional-columns-in-dax-power-bi-chapter-6-defcb162de7e) and this date table will help you to do [Time Intelligence](https://medium.com/analytics-vidhya/dealing-with-time-intelligence-functions-in-dax-power-bi-chapter-7-3f7915e6f4e1) Analysis.
* Go to the Modelling tab → Click on the New table
* Write “Table = CALENDARAUTO()” and the automatic date table is now in place.

1. **Create Relationship**

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

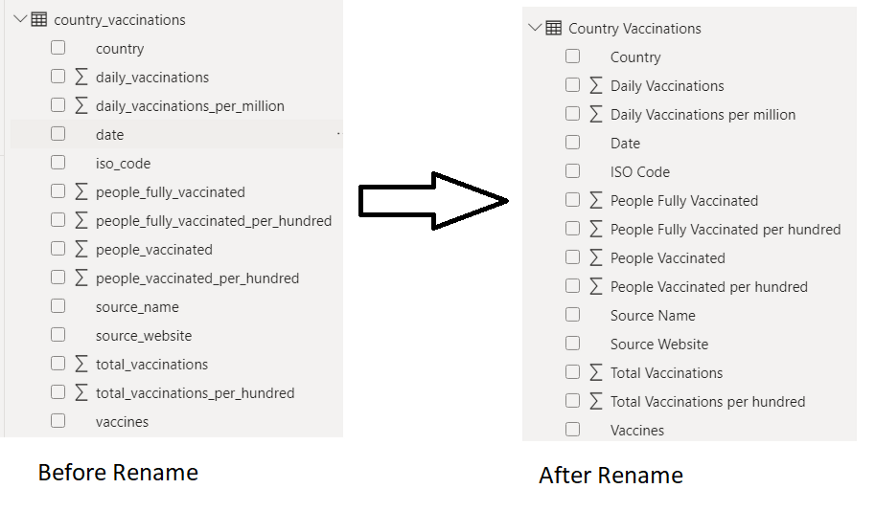
* Click on “Model” from the left side navigation bar.
* Click on the date column of the country\_vaccinations table, then drag & drop to the date column of the Date Table.
* A many-to-one relationship is created.



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.

1. **Modify Column Name**

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



1. **Select Theme**

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

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

1. **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.

* Click on Card visual → Added to the canvas area.
* Select field People Fully Vaccinated
* Click on Format your visual → Go to Data Visual and Callout Value. Change Color, Font family and Text size. Add some background color to it.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |

* Follow the same process for Total Vaccinations, People Vaccinated, and Total Country.
* To derive Total Country → Select the Country column and change to Count (Distinct) from the drop-down.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

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

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

1. **Add Year, Month Slicer**

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

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **After all changes** |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Change it to Dropdown** |  |  |  |  |

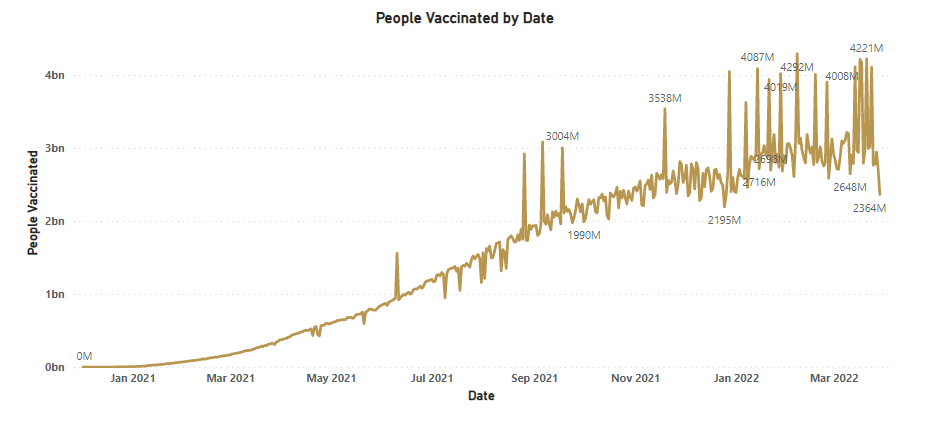
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |

1. **Create Line Graph “People Vaccinated by Date”**

* Add a Line Chart to the canvas area.
* 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.
* In the Format section, you can do the following changes
  + choose one Data Color,
  + 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.
  + if you want, you can modify the title, font size, or different text styles.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Final Output** |  |  |  |  |



1. **Create Clustered Column Chart “Top 10 Country by People Fully Vaccinated”**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

* In the Format section, you can do the following changes
  + choose one Data Color,
  + enable Data labels, update Display units based on your choice so that values can be visible properly and easy to follow the data variations.
  + if you want, you can modify the title, font size, or different text styles.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Final Output** |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |

1. **Create Clustered Bar Chart “Top 10 Country by Daily Vaccinations”**

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

|  |  |  |
| --- | --- | --- |
|  |  |  |

1. **Create Map Visual “Total Vaccinations by Country”**

* Add Map visual to the canvas area.
* Add Country in Location and Total Vaccinations in Size.
* In this analysis, you want to get some idea overall vaccinations process across all countries. For this, a map visual is ideal.
* In the Format section, you can do the following changes
  + choose one Data Colors
  + change the Map styles to Grayscale, update Bubbles size
  + if you want, you can modify the title, font size, or different text styles.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Final Output** |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |

1. **Publish to Power BI Service**

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |

1. **Open Report from Power BI Service**

* Type in browser → <https://app.powerbi.com/>
* Enter your credentials.
* Go to the workspace where you have published.
* Now your report is ready for presentation.

