

ASSIGNMENT 2: DATA VISUALIZATION

Malaika Akram

A. Load the data in the tool. Briefly explain the dataset

I used the covid_19_clean_complete.csv dataset from Kaggle. This dataset provides daily global information on COVID-19 cases from early 2020, tracking confirmed cases, deaths, recoveries, and active cases across various countries

Dataset Overview

Some key pieces of information in the dataset are:

Date: The day the data was recorded

Country: The country or region being reported on

Confirmed Cases: The total number of people who tested positive for COVID-19

Deaths: The total number of people who died due to the virus

Recovered: The number of people who got better after having the virus

Active Cases: The number of people still sick (calculated as confirmed cases minus deaths and recoveries)

New Cases & New Deaths: The number of new cases or deaths reported on a specific day

The screenshot shows the Power BI Data Load interface. On the left, the 'Navigator' pane displays a list of files: 'covid_19.xlsx [1]' and 'covid_19_clean_complete'. The 'covid_19_clean_complete' file is selected, indicated by a checked checkbox. The main area shows a preview of the 'covid_19_clean_complete' dataset with columns: Province/State, Country/Region, Lat, and Long. The data includes rows for various countries like Afghanistan, Albania, Algeria, Andorra, Angola, Antigua and Barbuda, Argentina, Armenia, Australia, and many others. At the bottom of the interface, there are buttons for 'Load', 'Transform Data', and 'Cancel'. A page navigation bar at the bottom left shows 'Page 1'.

The screenshot shows the Power BI Data Editor interface. The top navigation bar includes tabs for 'Structure' (selected), 'Relationships', 'Calculations', and 'Calendars'. Below the navigation bar is a toolbar with icons for 'New measure', 'Quick measure', 'New column', 'New table', and 'Mark as date table'. The main area displays a table with the following schema:

Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Active
	Lesotho	-29.61	28.2336	Wednesday, January 22, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Thursday, January 23, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Friday, January 24, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Saturday, January 25, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Sunday, January 26, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Monday, January 27, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Tuesday, January 28, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Wednesday, January 29, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Thursday, January 30, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Friday, January 31, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Saturday, February 1, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Sunday, February 2, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Monday, February 3, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Tuesday, February 4, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Wednesday, February 5, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Thursday, February 6, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Friday, February 7, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Saturday, February 8, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Sunday, February 9, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Monday, February 10, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Tuesday, February 11, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Wednesday, February 12, 2020	0	0	0	0
	Lesotho	-29.61	28.2336	Thursday, February 13, 2020	0	0	0	0

B. Transform the data (Conditional/Custom Columns etc.)

Conditional Column

I added a column that labels the situation in each country as “Low”, “Moderate”, or “High” based on the number of confirmed cases using a conditional rule:

If Confirmed Cases < 10,000 → Low

If between 10,000 and 100,000 → Moderate

If > 100,000 → High

Power Query Editor

Transform Add Column View Tools Help

General Conditional Column Index Column Duplicate Column Format ABC Extract Parse Statistics Standard Scientific Trigonometry From Number Date Time Duration From Date & Time

Add Conditional Column

Add a conditional column that is computed from the other columns or values.

New column name: severity level

Column Name	Operator	Value	Output
If Confirmed	equals	ABC 123 10000	Then ABC 123 Low
Else If Confirmed	equals	ABC 123 100000	Then ABC 123 Moderate
Add Clause			
Else	ABC 123	High	

OK Cancel

19_clean_complete

	Deaths	Recovered	Active	WHO Region	Severity level
1	0	0	0	Eastern Mediterranean	High
2	0	0	0	Europe	High
3	0	0	0	Africa	High
4	0	0	0	Europe	High
5	0	0	0	Africa	High
6	0	0	0	Americas	High
7	0	0	0	Americas	High
8	0	0	0	Europe	High
9	0	0	0	Western Pacific	High
10	0	0	0	Western Pacific	High
11	0	0	0	Western Pacific	High
12	0	0	0	Western Pacific	High
13	0	0	0	Western Pacific	High
14	0	0	0	Western Pacific	High
15	0	0	0	Western Pacific	High
16	0	0	0	Western Pacific	High
17	0	0	0	Europe	High
18	0	0	0	Europe	High
19	0	0	0	Americas	High
20	0	0	0	Eastern Mediterranean	High
21	0	0	0	South-East Asia	High
22	0	0	0	Arabian	High

Add Column View Tools Help

General Conditional Column Index Column Duplicate Column Format ABC Extract Parse Statistics Standard Scientific Trigonometry From Number Date Time Duration From Date & Time

Add Column

= Table.AddColumn(#"Changed Type", "Severity level", each if [Confirmed] = 10000 then "Low" else if [Confirmed] = 100000 then "Moderate" else "High")

Deaths	Recovered	Active	WHO Region	Severity level
0	0	0	Eastern Mediterranean	High
0	0	0	Europe	High
0	0	0	Africa	High
0	0	0	Europe	High
0	0	0	Africa	High
0	0	0	Americas	High
0	0	0	Americas	High
0	0	0	Europe	High
0	0	0	Western Pacific	High
0	0	0	Western Pacific	High
0	0	0	Western Pacific	High
0	0	0	Western Pacific	High
0	0	0	Western Pacific	High
0	0	0	Western Pacific	High
0	0	0	Europe	High
0	0	0	Europe	High
0	0	0	Americas	High
0	0	0	Eastern Mediterranean	High
0	0	0	South-East Asia	High
0	0	0	Arabian	High

Query Settings

Properties

Name: covid_19_clean_complete

All Properties

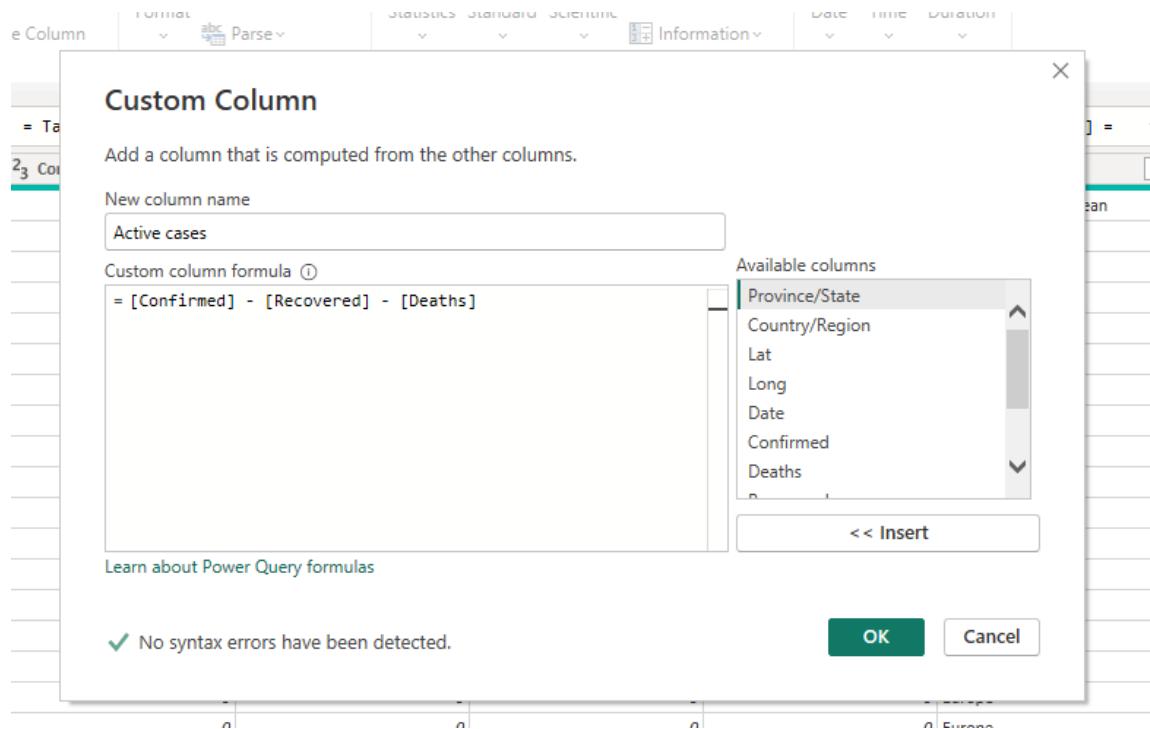
Applied Steps

- Source
- Navigation
- Promoted Headers
- Changed Type
- Added Conditional

Custom column

I added a new column called Active Cases to calculate the number of currently infected people. The formula used was:

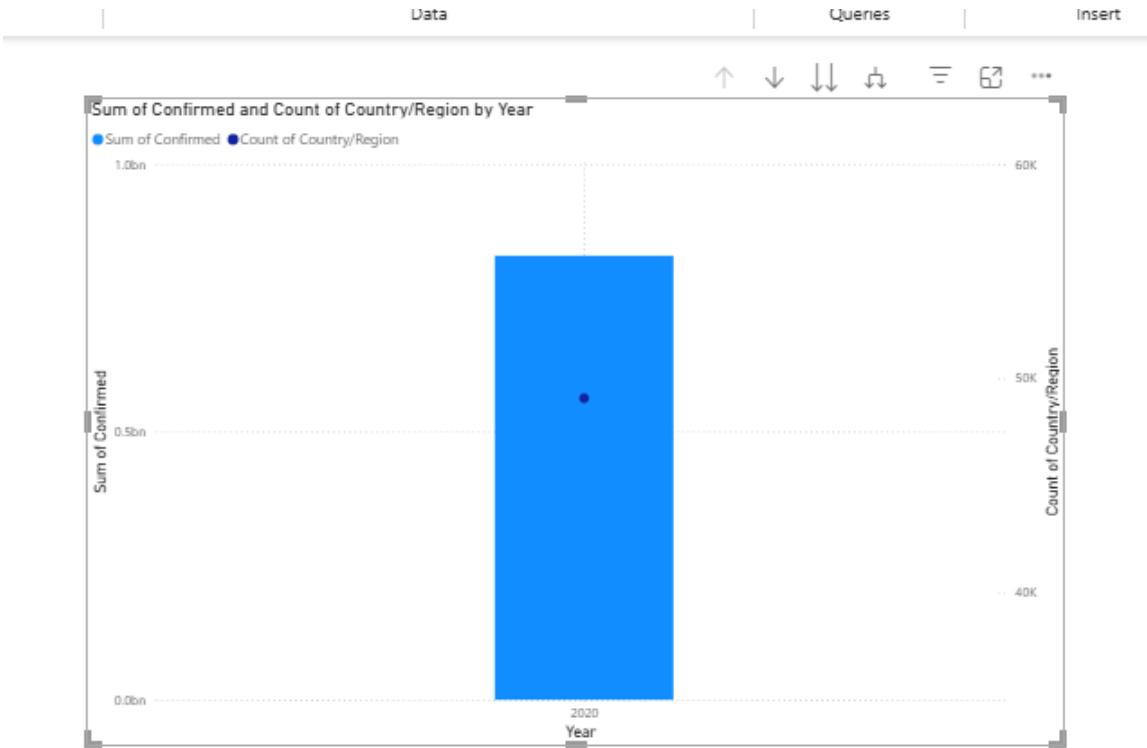
Active Cases = Confirmed Cases - Recovered - Deaths



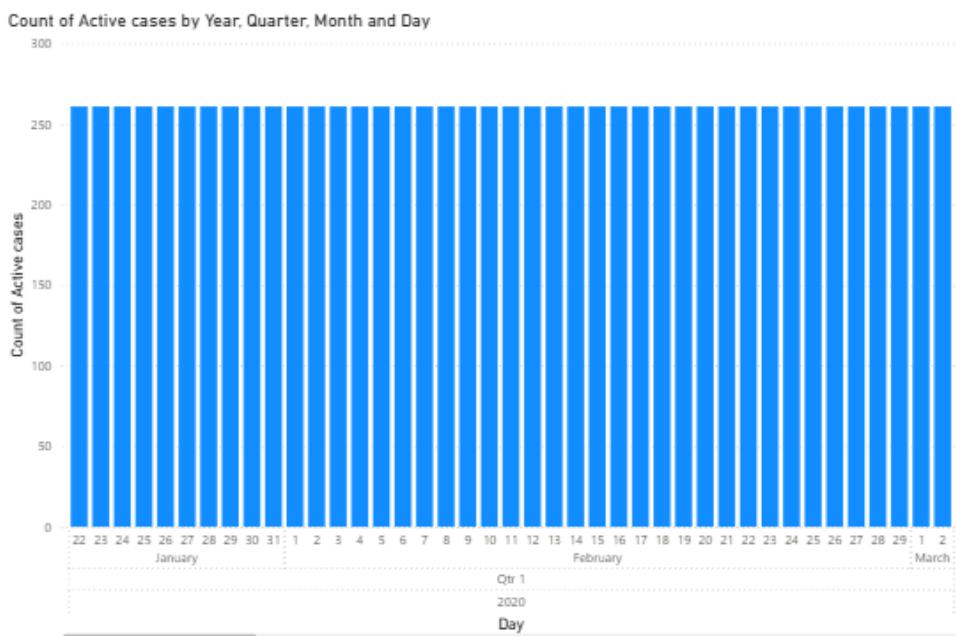
	Recovered	Active	WHO Region	Severity level	Active cases
44	0	0	Americas	High	0
45	0	0	Americas	High	0
46	0	0	Africa	High	0
47	0	0	Africa	High	0
48	0	0	Americas	High	0
49	0	0	Western Pacific	High	1
50	0	0	Western Pacific	High	14
51	0	0	Western Pacific	High	6
52	0	0	Western Pacific	High	1
53	0	0	Western Pacific	High	0
54	0	0	Western Pacific	High	26
55	0	0	Western Pacific	High	2
56	0	0	Western Pacific	High	1
57	0	0	Western Pacific	High	4
58	0	0	Western Pacific	High	1
59	0	0	Western Pacific	High	0
60	0	0	Western Pacific	High	5
61	0	0	Western Pacific	High	0
62	17	28	Western Pacific	High	399
63	0	0	Western Pacific	High	4
64	0	0	Western Pacific	High	0
65	0	0	Western Pacific	High	1

C. Build multiple visualization charts

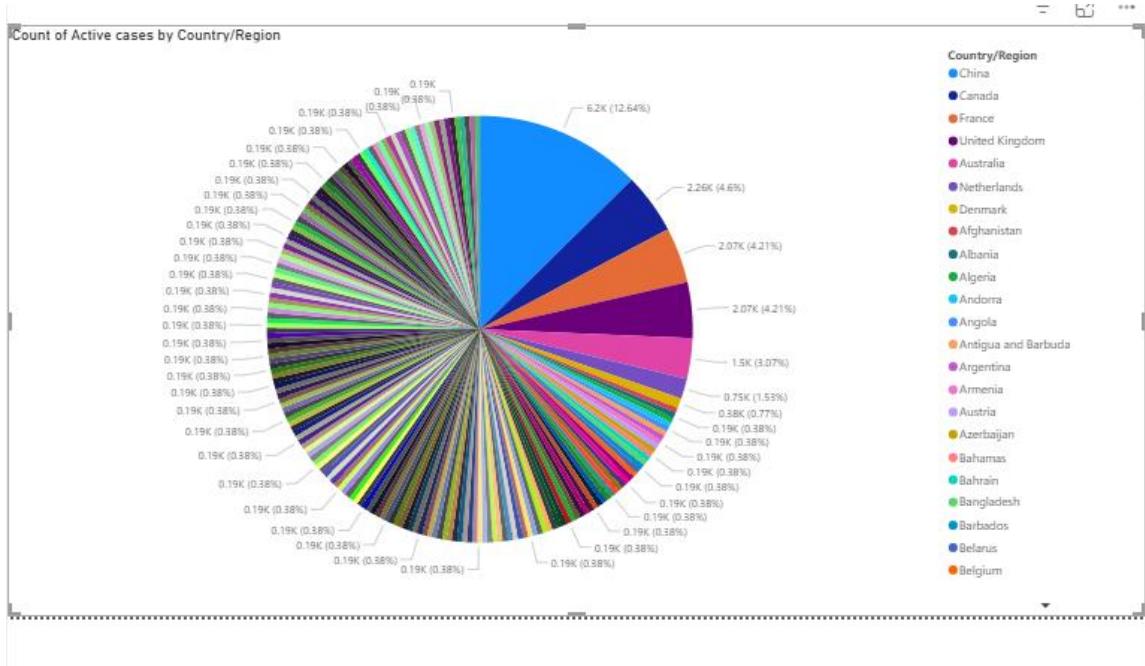
Line Chart – Confirmed Cases Over Time



Line Chart – Active Cases



Pie Chart



D. Build Measures and DAX functions

To perform deeper analysis on the COVID-19 dataset, I created several DAX measures in Power BI. These allow for real-time calculations based on the data shown in charts or filtered by slicers

Total confirmed

Home table covid_19_clean_co... \$ % Auto New Quick measure measure

Structure Formatting Properties Calculations

1 Total Confirmed = SUM('covid_19_clean_complete'[confirmed])

Build visuals with your data

Total Deaths

Home table covid_19_clean_co... \$ % Auto New Quick measure measure

Structure Formatting Properties Calculations

1 Total Deaths = SUM('covid_19_clean_complete'[deaths])

Total Recovered



Overall usefulness and interactive nature of the dashboard

Usefulness:

Summarizes key data like total confirmed cases, total deaths, recoveries, and active cases in one view.
Helps users identify trends through line charts (e.g., confirmed cases over time).
Allows comparison between countries, showing which were most affected.
Shows the death rate and recovery rate, which gives deeper insights into health system outcomes.
The map visual makes it easy to understand which regions were more impacted globally.

Interactivity:

I added slicers for country and date range so users can:
Focus on a specific country's data
Analyze data during a selected time period
All visuals automatically update based on the filters selected.
This allows the user to perform their own analysis without changing the report manually.

