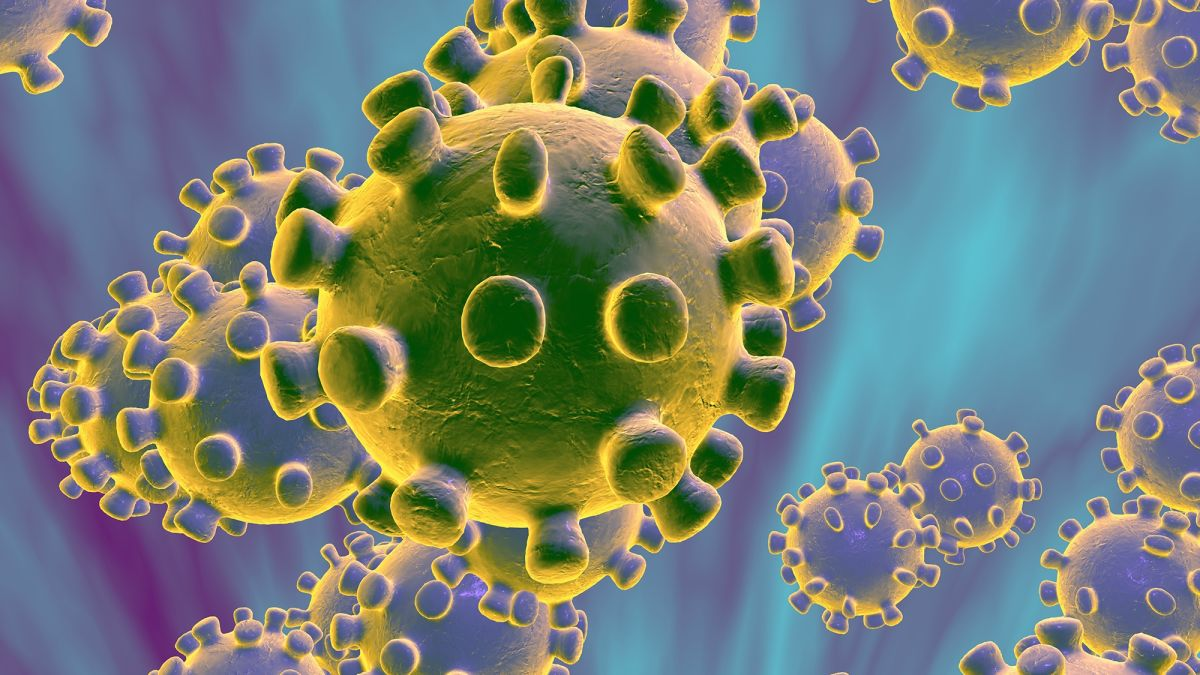
Data Warehousing

**FINAL PROJECT DOCUMENTATION**

A detail visual representation of COVID-19 patients in the United States

horizontal line



**Team 4:**

Ankit Arun Patil A20451742

Daphney Carol Valiatingara A20446937

Divya Rathnakar Poojari A20446668

Neha Desai A20402675

Tanmay Deepak Patil A20439531

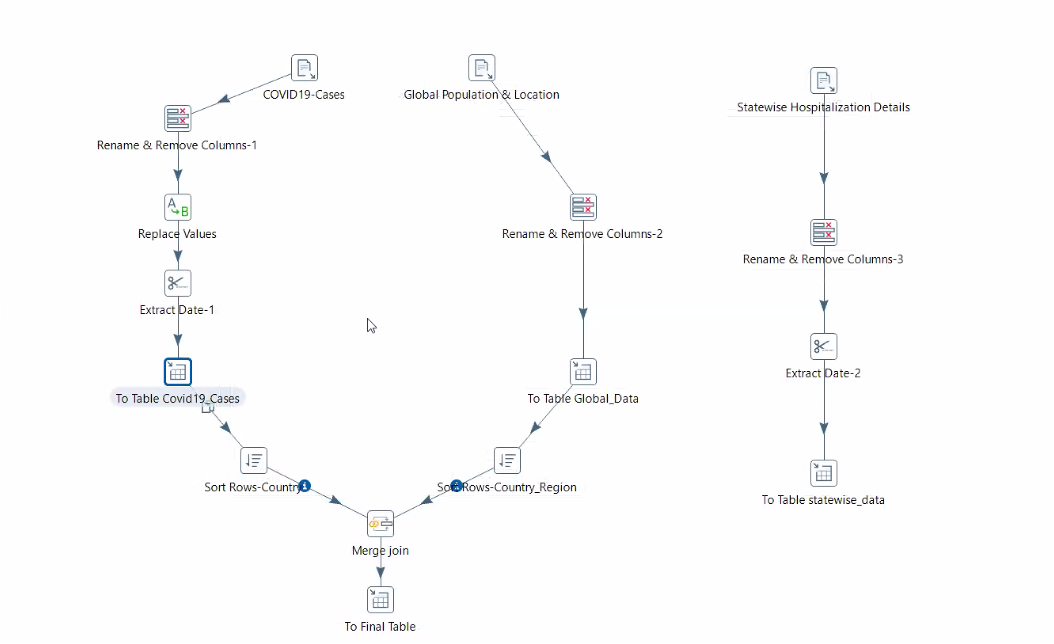
# Introduction

## ETL

Datasets used:

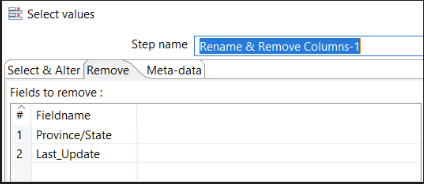
1. Datewise csv files : It comprises of deaths, confirmed, and recovered cases globally.
2. UID\_ISO\_FIPS\_LookUp\_Table.csv : It provides population and its latitude and longitude details of all the states affected by COVID19 globally.

Transformation screenshot:

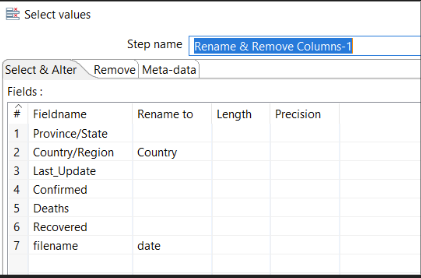


Data Pre-processing

1. Removal of fields using the ‘Remove’ tab of the ‘Select values’ component



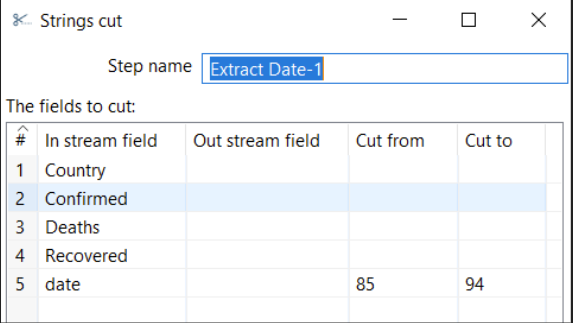
1. Renames fields using the ‘Select & Alter’ tab of the ‘Select values’ component



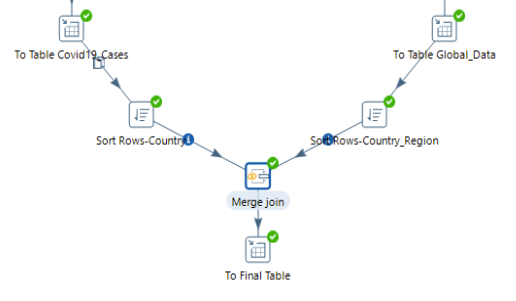
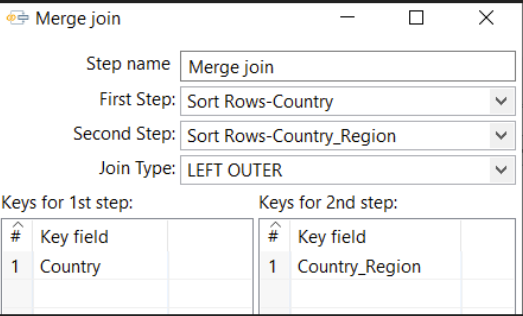
1. Using the ‘Replace in string’ component to replace certain values of the field ‘Country’



1. Using the ‘Strings cut’ component to extract date from the timestamp.

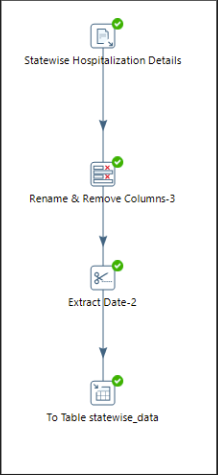


We have performed left outer join on the two datasets on the Country names and the final output table is Global\_Covid19\_Cases.



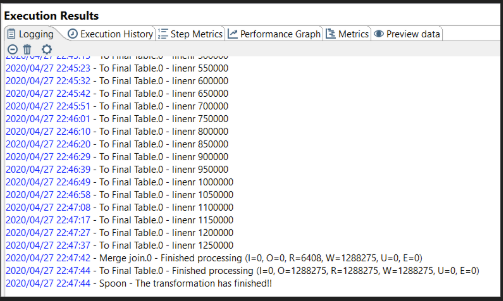
Second ETL transformation:

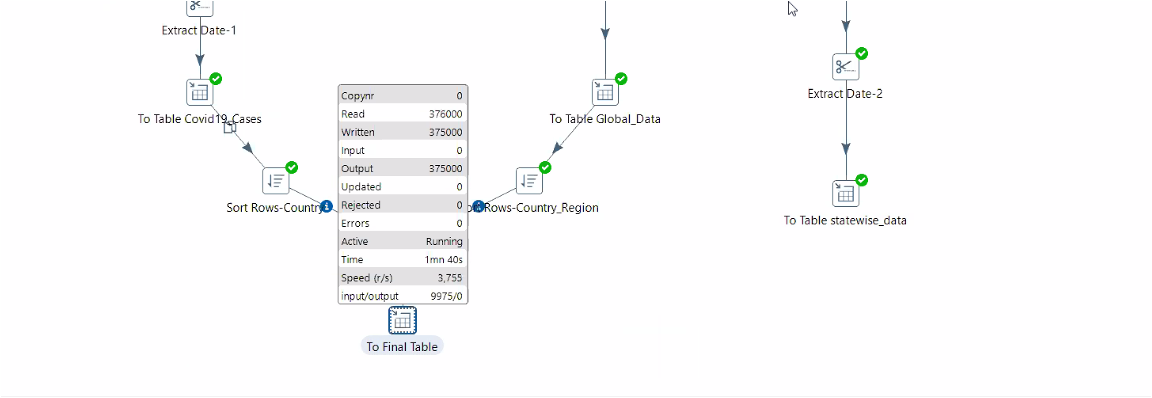
We have used the dataset states\_daily.csv which gives state-wise information on the hospitalizations, ICU, and ventilator usage, as well as the number of cases, tested positive and negative in the US. The data was loaded into the final output table named ‘statewise\_data’ after pre-processing.



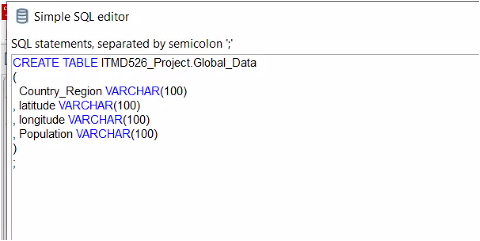
Execution of job

The job referencing the transformations was successfully executed and we have two output tables which will be used for analysis and visualizations.





Creation of tables



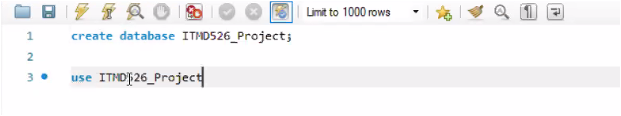
## 



## 

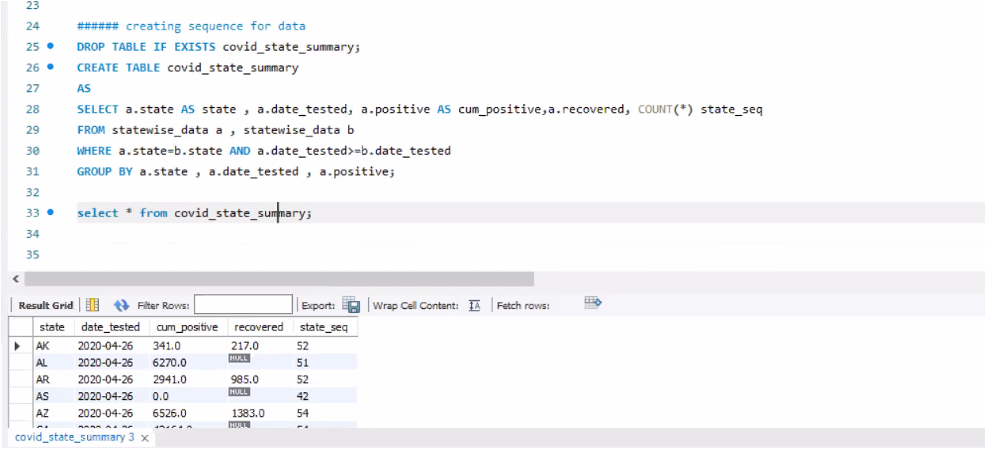
## SQL Queries

* Creation of database

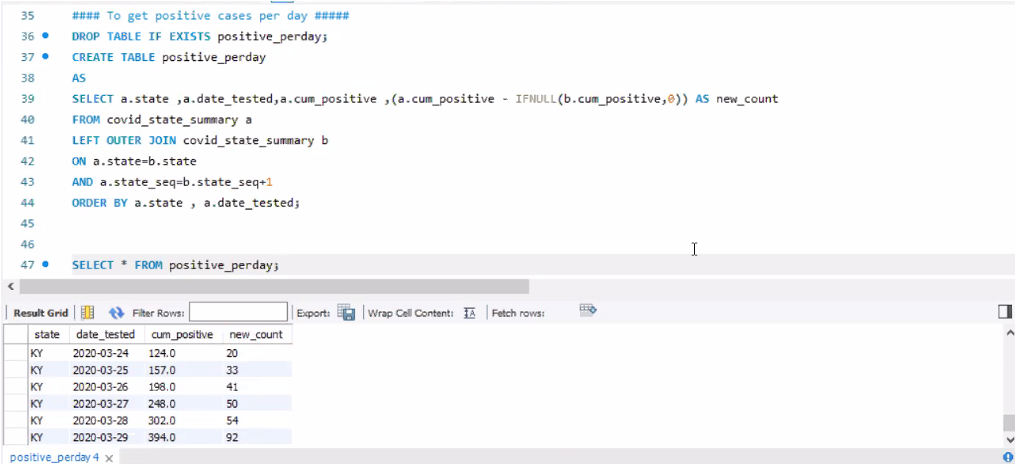


* SQL tables and queries created for simpler visualization

Query 1: Adds a sequence number to every state

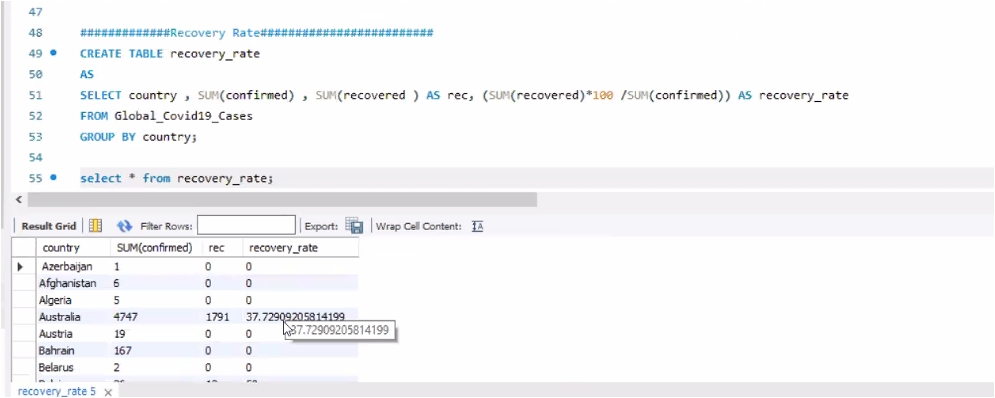


Quey 2: Fetches number of positive cases per day from cumulative positive cases as new\_count



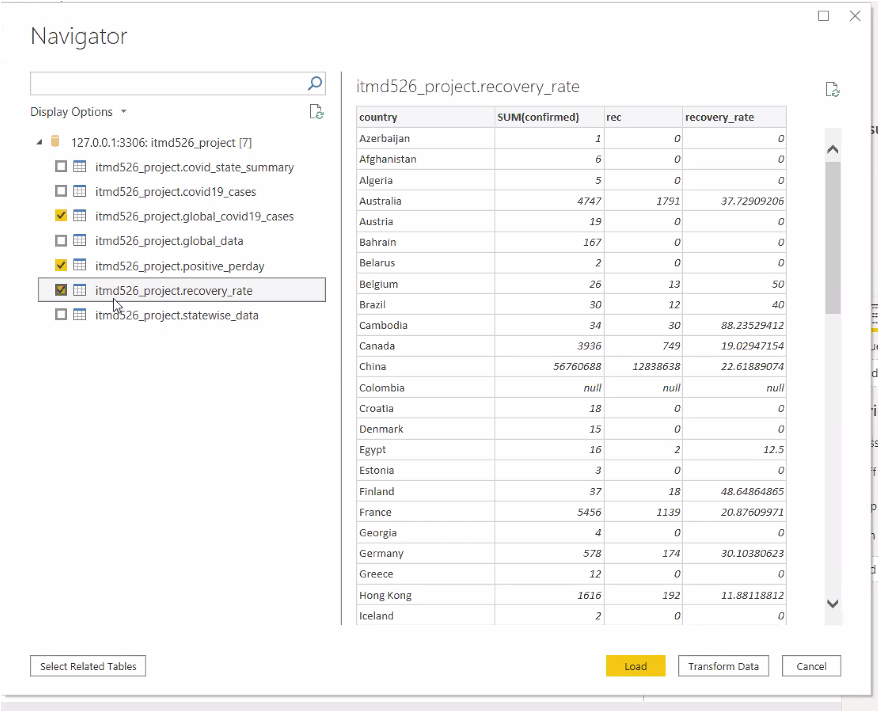
Query 3: This query calculates the recovery rate of every country to help visualize which country has a better covid recovery rate.

Recovery rate= confirmed cases for a country / recovered cases for a country



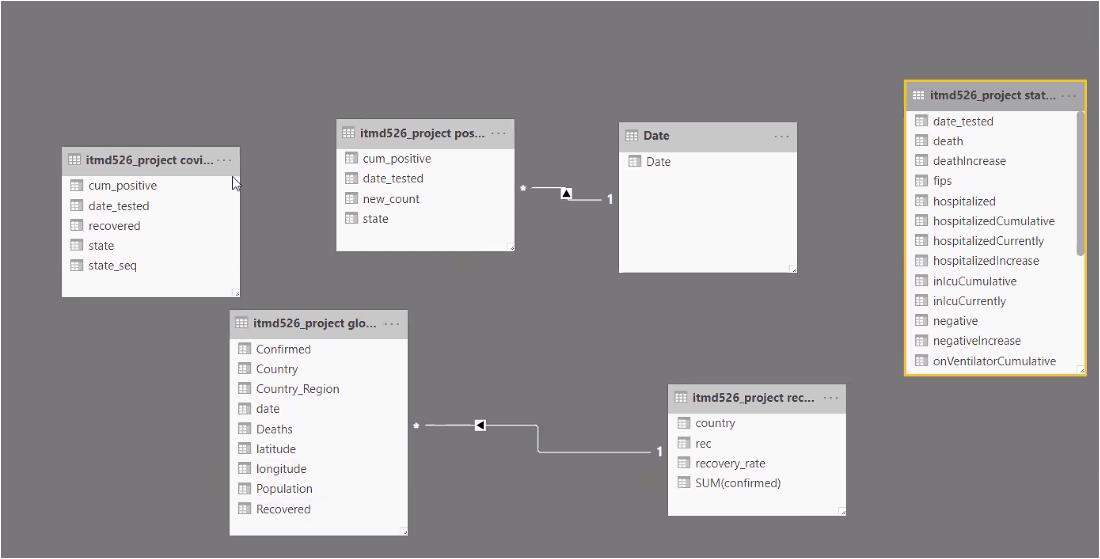
## POWER BI REPRESENTATION AND DASHBOARDS

* Loading data from SQL Server

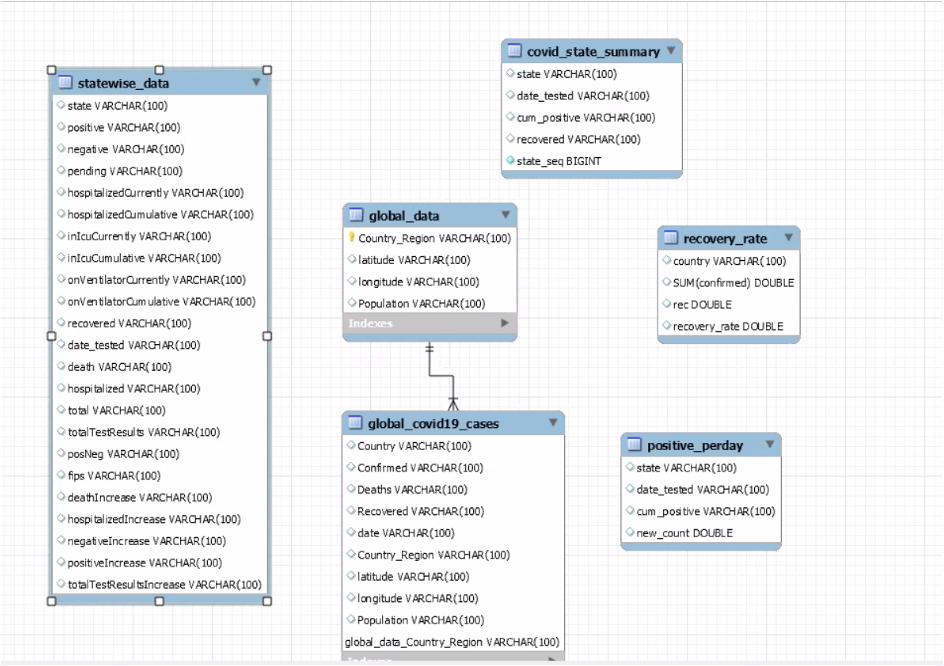


* Tables loaded in PowerBI





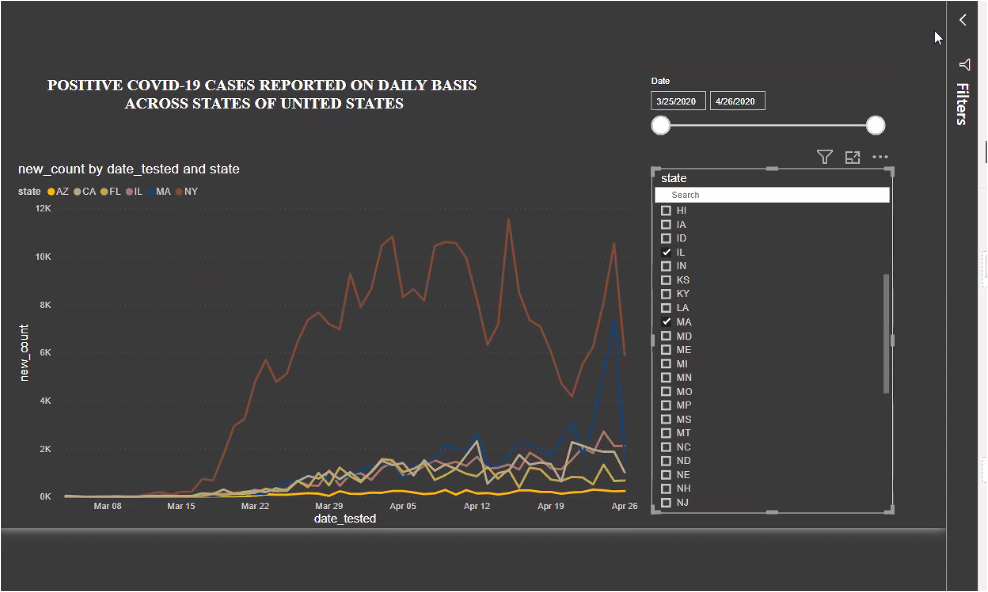
Data model:



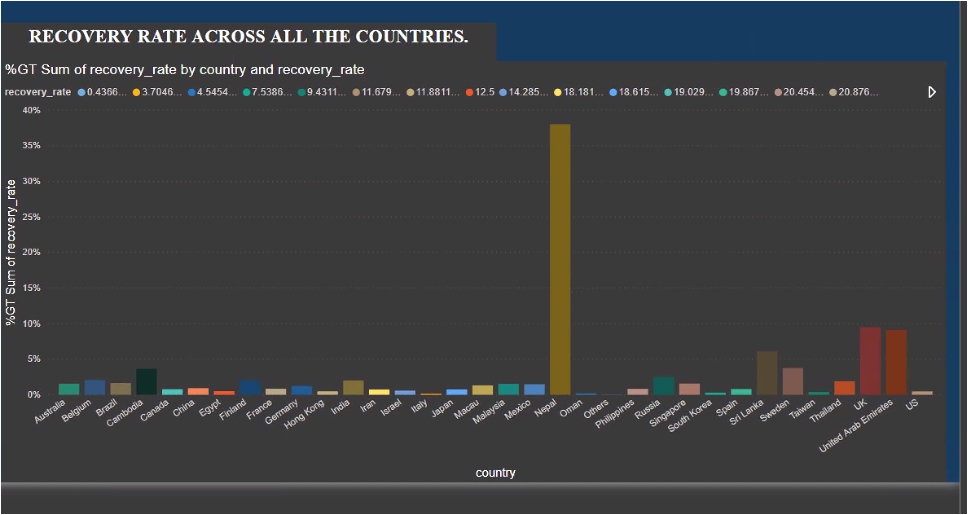
**DASHBOARDS IN POWER BI**

Below are the dashboards that were prepared.

* Dashboard that shows the positive cases across the different states in the US.

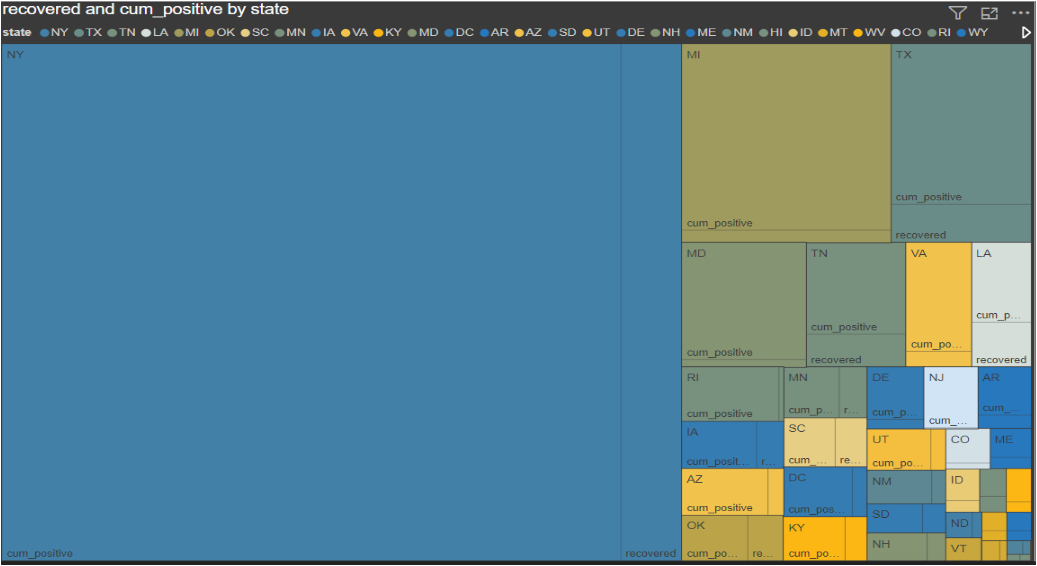


* Dashboard that shows recovery rate across all countries

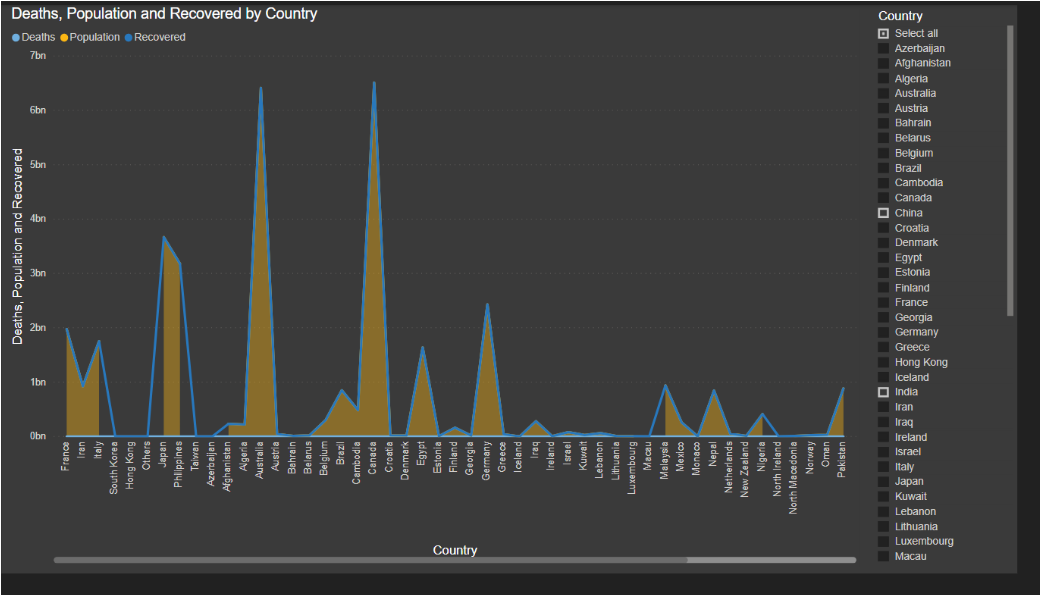


As seen above, the recovery rate for Nepal is the highest.

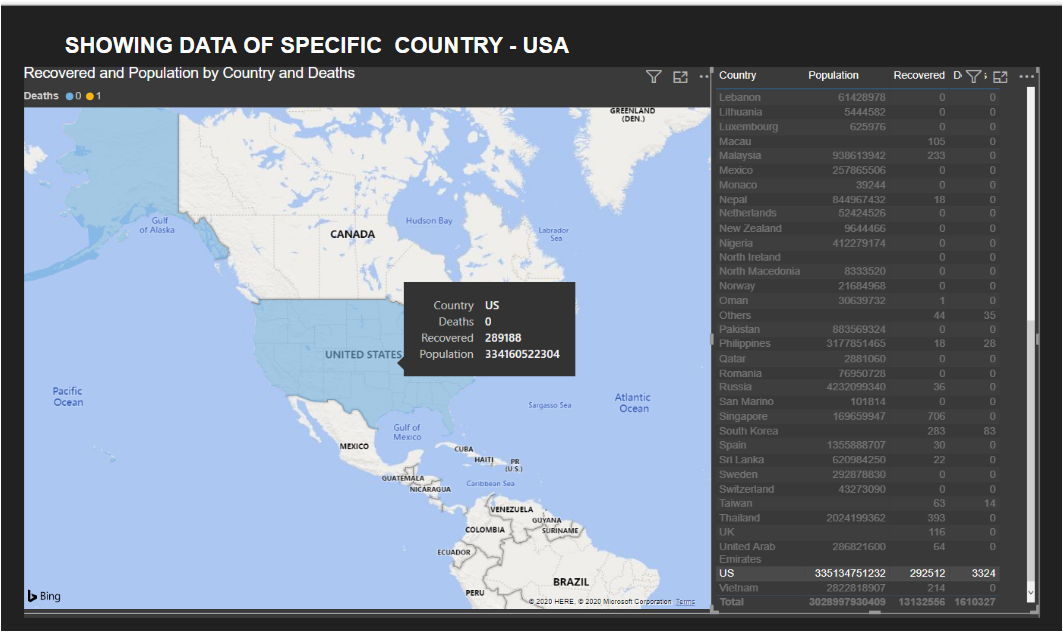
* Dashboard that shows recovered and cumulated positives by state.



* Stacked Area showing death, population and Recovered by Country.



* Map that shows the recovered and population country wise.



* Dashboard that shows an assortment of graphs.

**PENDING TESTING, NEGATIVE, POSITIVE CASES ACROSS THE STATES OF USA**