SVKM’s NMIMS, Mukesh Patel School of Technology Management & Engineering



DOCUMENTATION

FOR

**Future Predictions of COVID-19**

**(Mini Project)**

**By: Soham Vasudeo & Ritika Shetty**

**Team No: L039 & L040**

**Case Title: Future Predictions of COVID-19**

**College: Mukesh Patel School of Technology Management & Engineering, NMIMS**

Contents

[**Goal / Problem Statement** 3](#_Toc85921519)

[**Understanding the Training Data at hand** 3](#_Toc85921520)

[**Data Cleaning** 4](#_Toc85921523)

[**Code** 5](#_Toc85921524)

[**Exploratory Data Analysis** 8](#_Toc85921525)

[**Methodology used for data analysis** 15](#_Toc85921527)

[**Model fitting** 15](#_Toc85921529)

[**Results** 16](#_Toc85921530)

[**Dashboards** 17](#_Toc85921531)

[**Covid Cases Dashboard** 17](#_Toc85921532)

[**Covid Vaccination Dashboard** 18](#_Toc85921533)

# **Goal / Problem Statement**

The goal of this assignment/ combined project is to use all three subjects namely: Python Programming, Business Visualization and Statistics, Analytics and Modelling to visualize the given data and make Future Predictions of COVID cases using the SARIMA model.

# **Understanding the Training Data at hand**

# The data set we chose was namely “Covid-19 India, State wise Testing, Covid-19 Vaccinations­­­­­­­­­­­”. The elements in the data set were as follows:

# COVID-19 India:

* States/ Union Territories
* Dates and Times
* Confirmed Cases- both Indians and Foreign Nationals
* Cured Cases
* Deaths

1. COVID Vaccine:

* States
* Total Doses Administered
* Sessions
* Sites
* First Dose Administered
* Second Dose Administered
* Male (Doses Administered)
* Female (Doses Administered)
* Transgender (Doses Administered)
* Covaxin (Doses Administered)
* CoviShield (Doses Administered)
* Sputnik V (Doses Administered)
* AEFI
* 18-44 Years (Doses Administered)
* 54-60 Years (Doses Administered)
* 60+ Years (Doses Administered)
* 18-44 Years(Individuals Vaccinated)
* 18-44 Years(Individuals Vaccinated)
* 45-60 Years(Individuals Vaccinated)
* 60+ Years(Individuals Vaccinated)
* Male(Individuals Vaccinated)
* Female(Individuals Vaccinated)
* Transgender(Individuals Vaccinated)
* Total Individuals Vaccinated

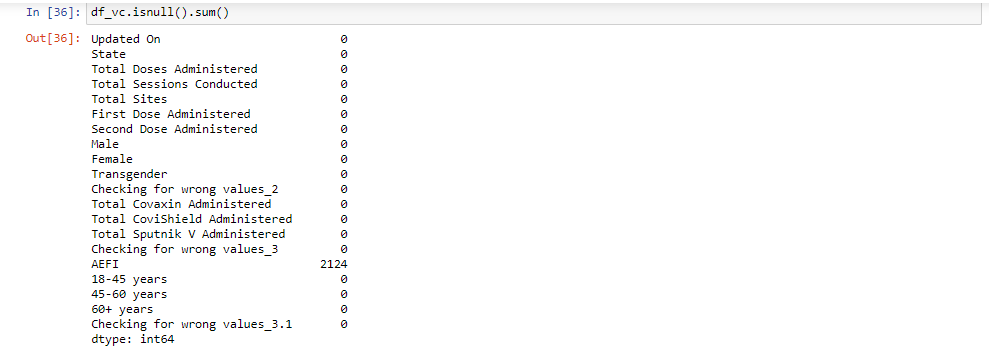
1. Statewise Testing Details:

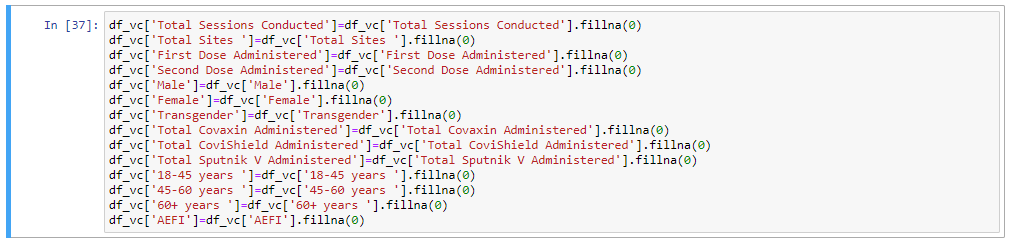
* Date
* State
* Total Samples
* Negative
* Positive

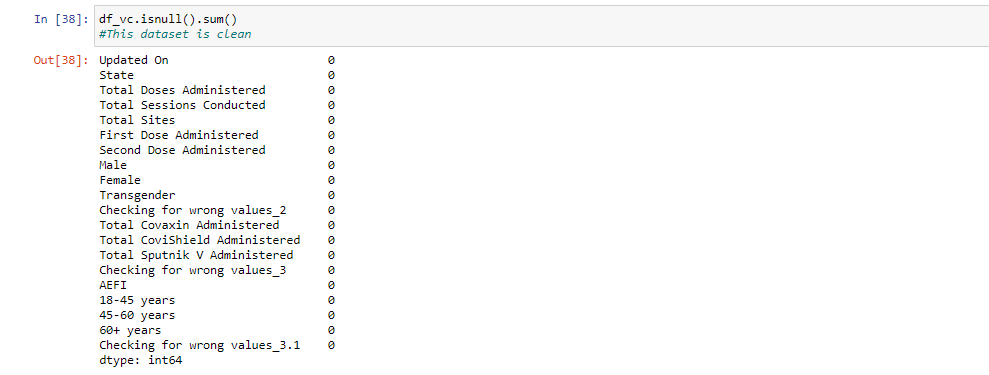
# **Data Cleaning**

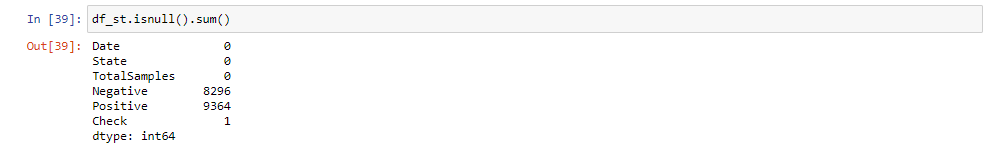


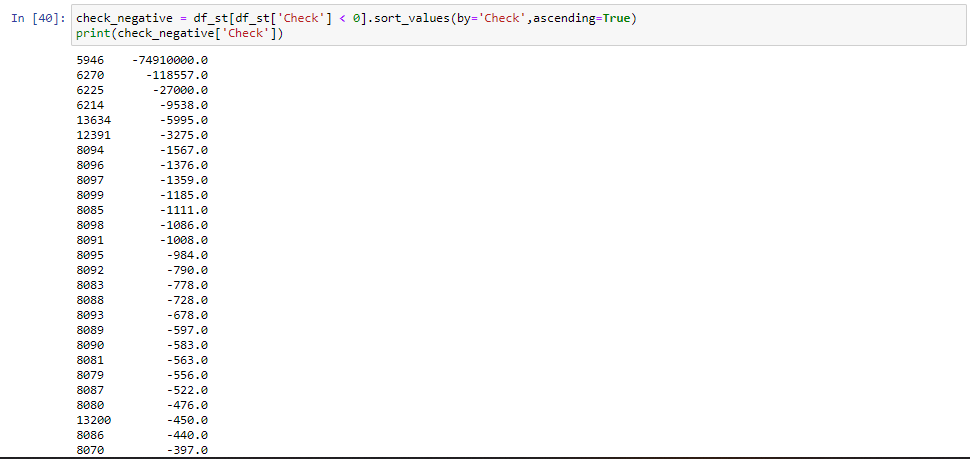




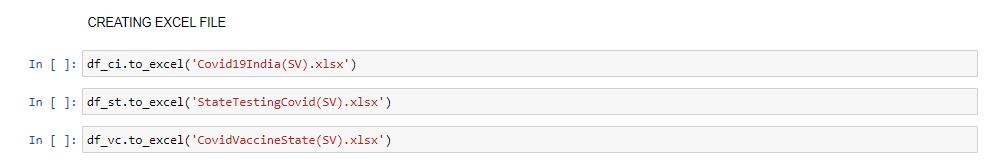




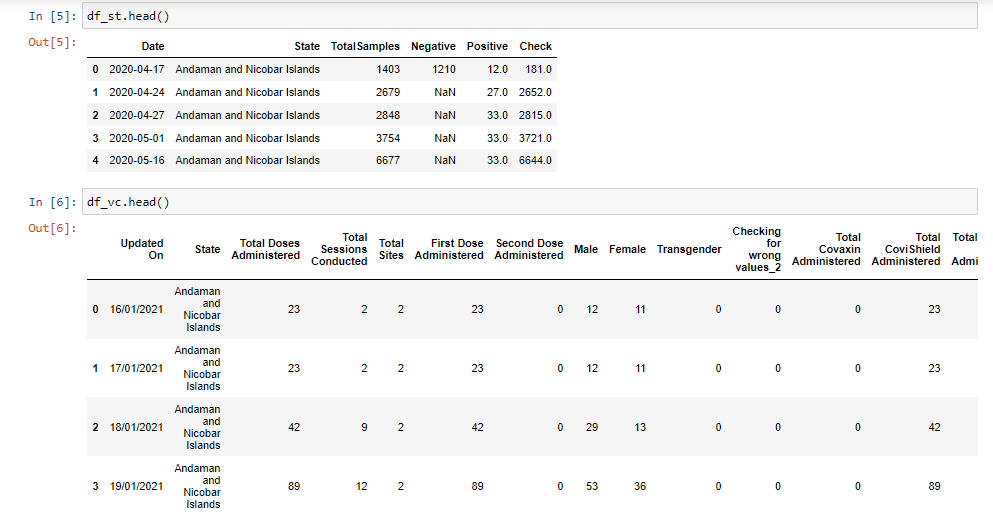
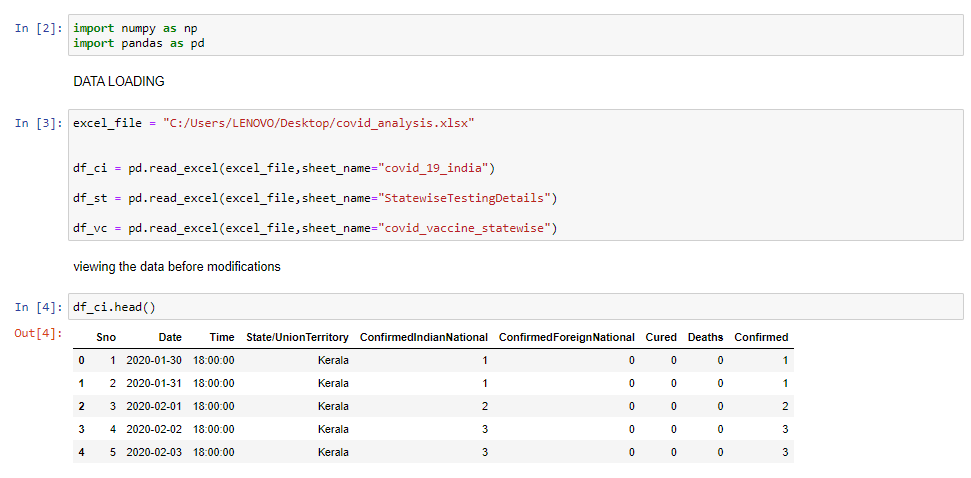


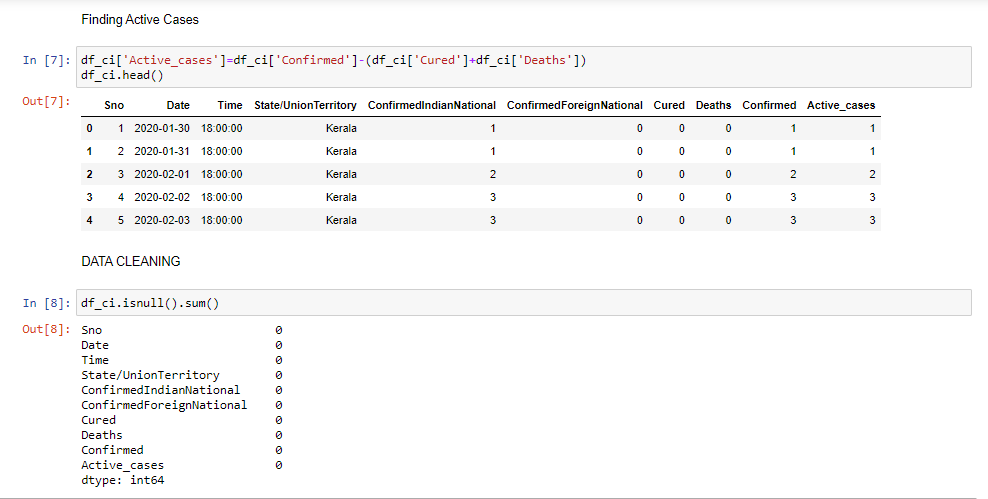






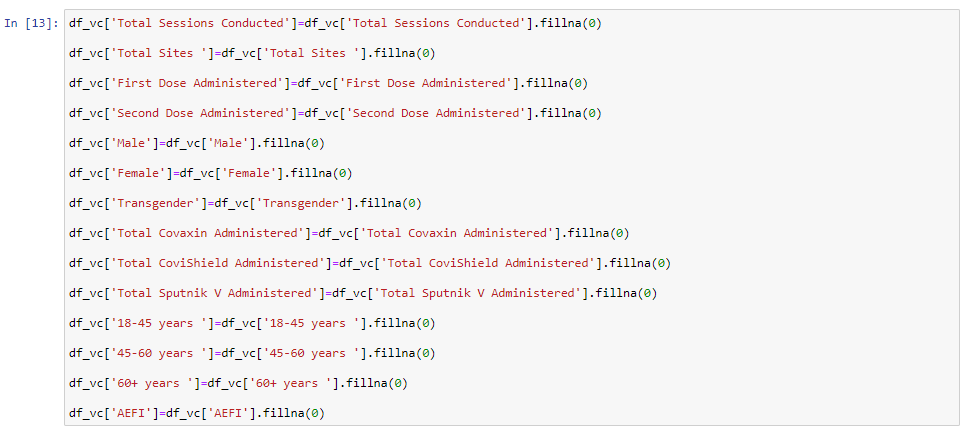
# **Code**

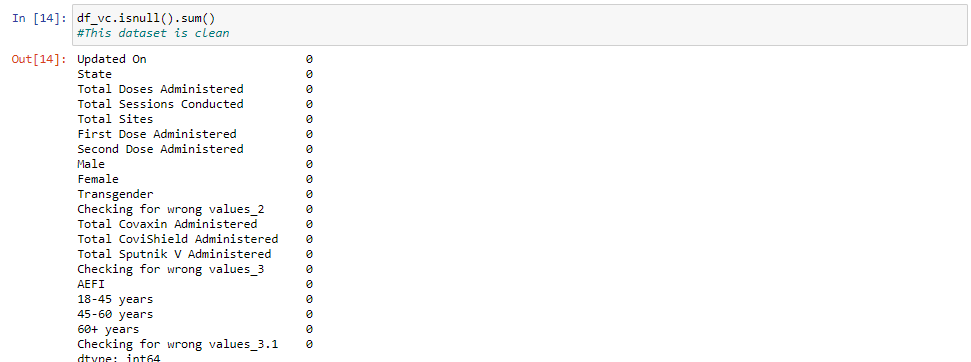
















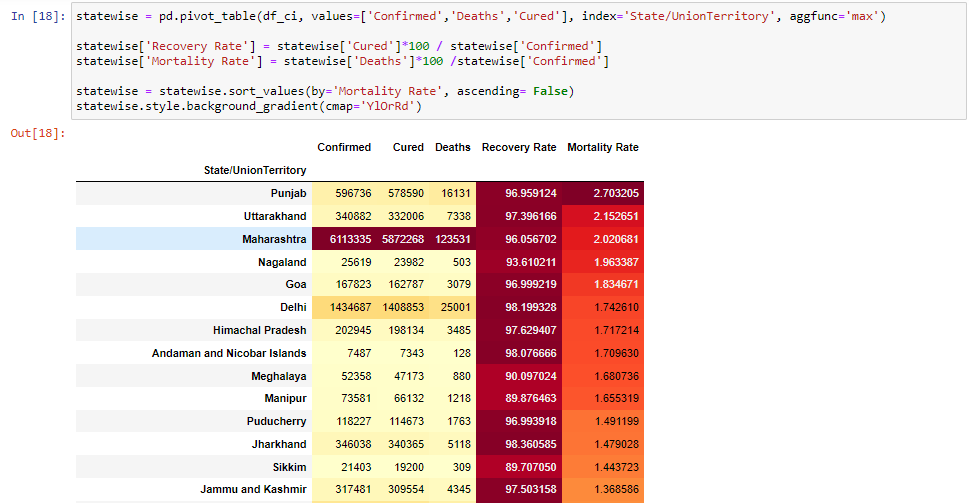


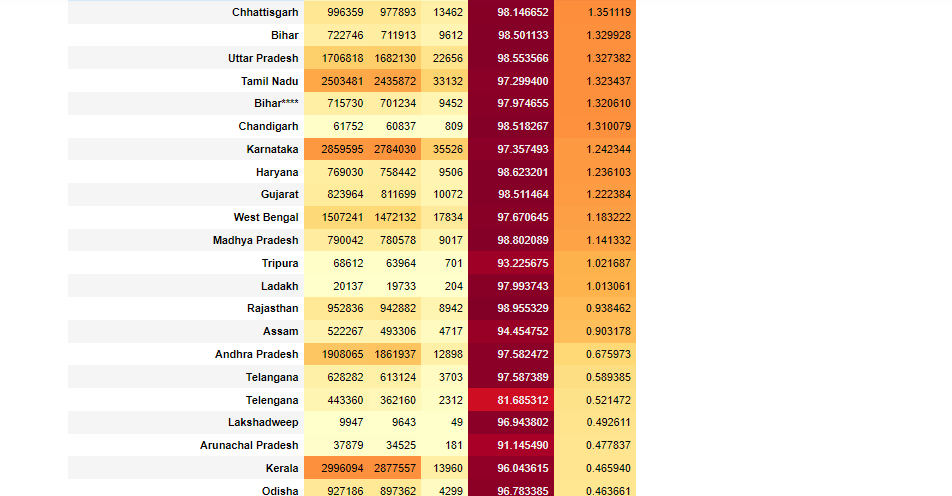
# **Exploratory Data Analysis**

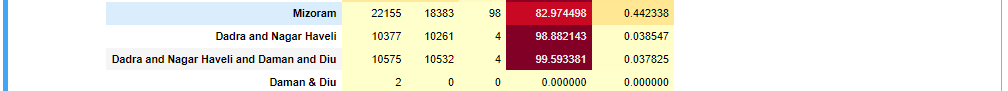
# We have done some exploratory analysis on python itself to understand the data better. The code snippets are given below:



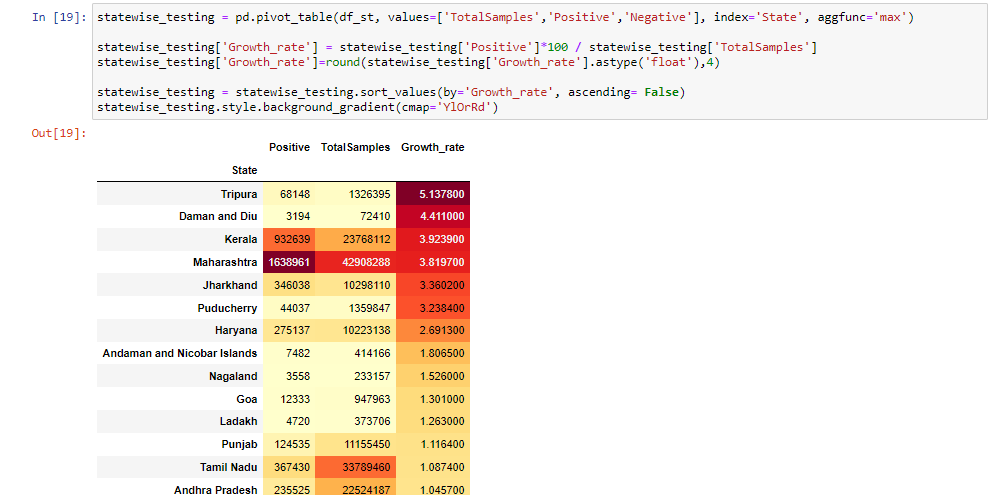
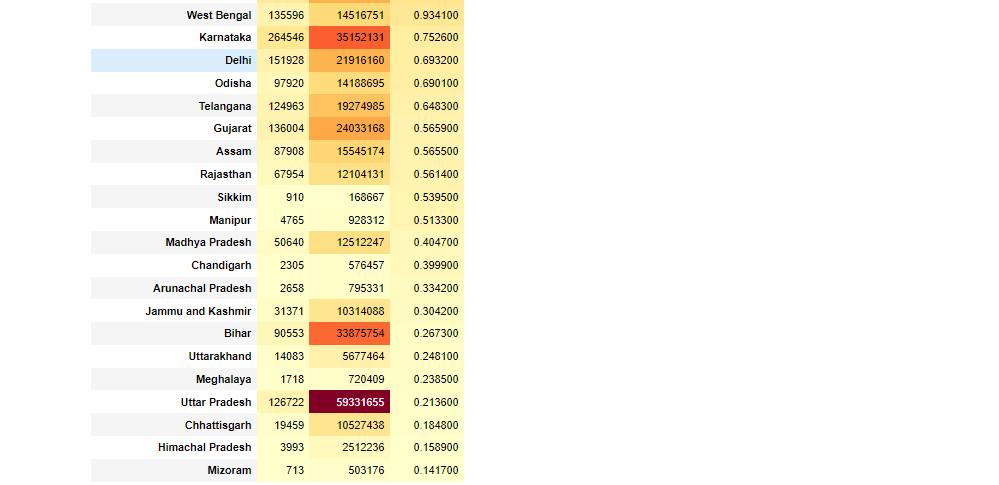
1. Firstly, we have made a crosstab in which we are showing numeric sum values for confirmed cases, cured cases, deaths and upon further calculation we have also given mortality and recovery rate for each state.



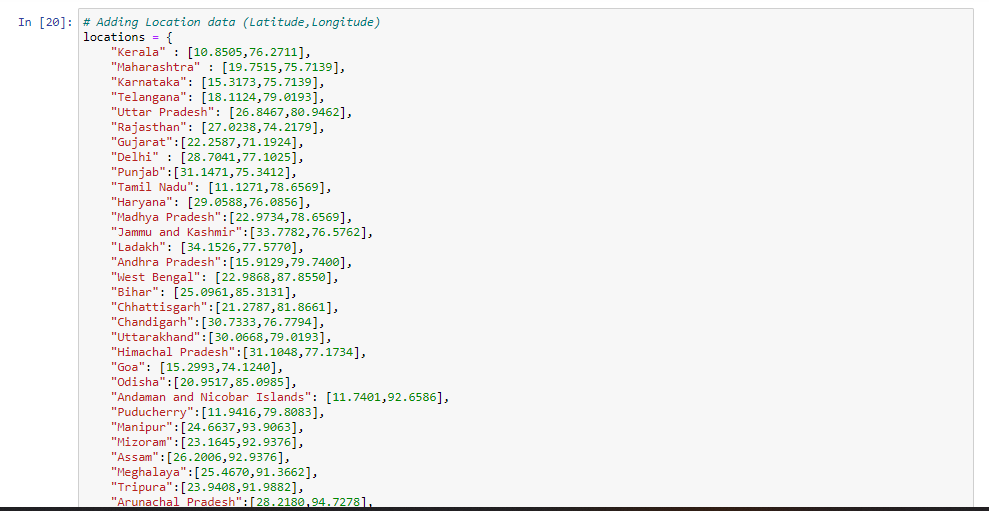




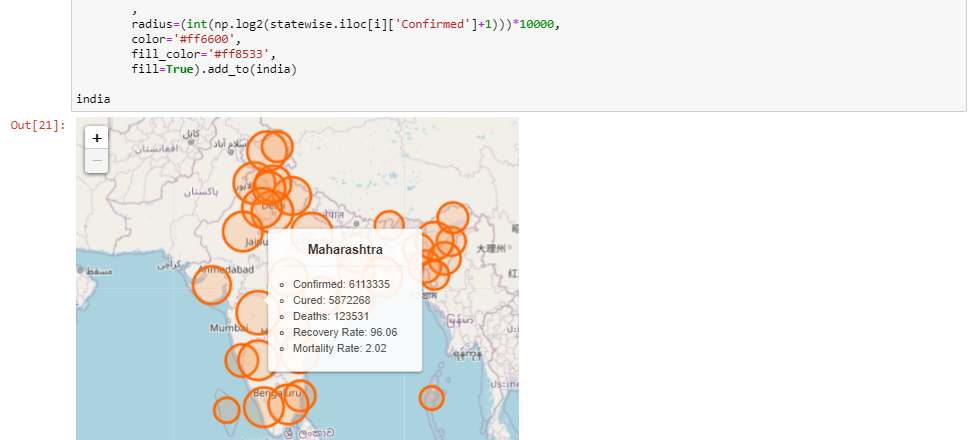
1. Next, we have calculated the growth rate of cases for each state.

1. For better understanding of data, we have plotted a geo map. On hovering the mouse tip on the state, we get data tip values for cured, deaths, confirmed, mortality and recovery rate.

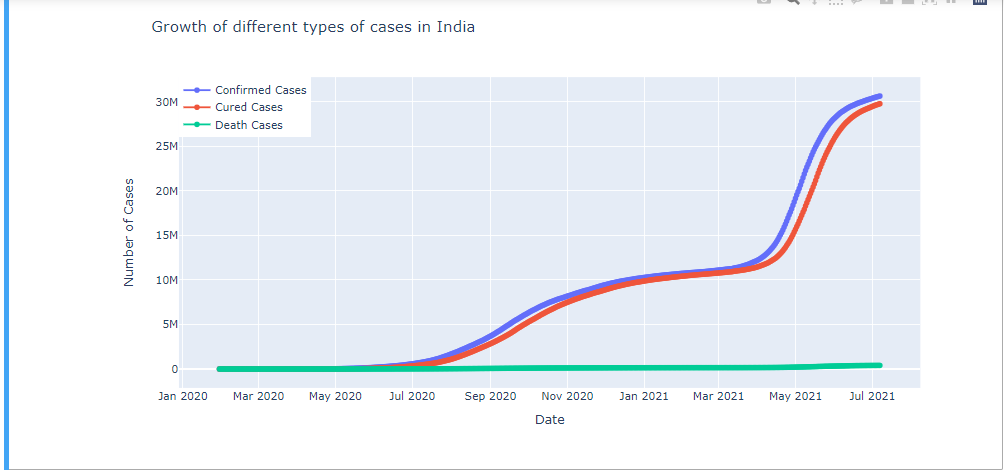




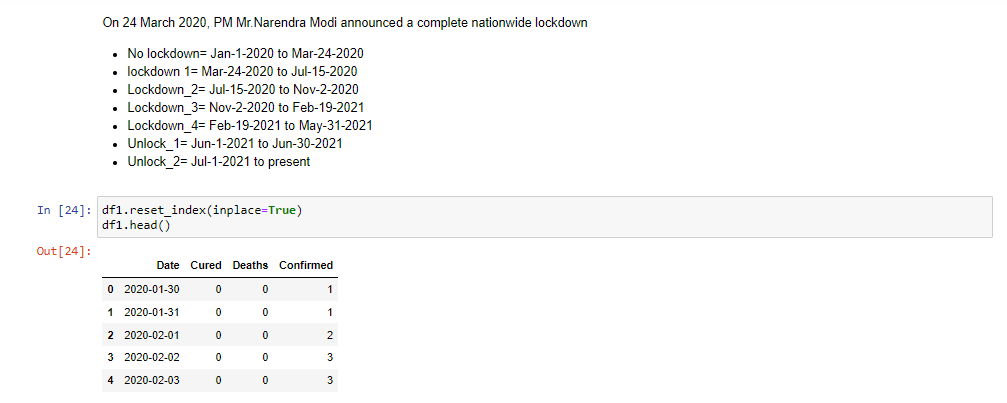


1. We have created a scatter plot with line and markers to show a growth chart for different cases.

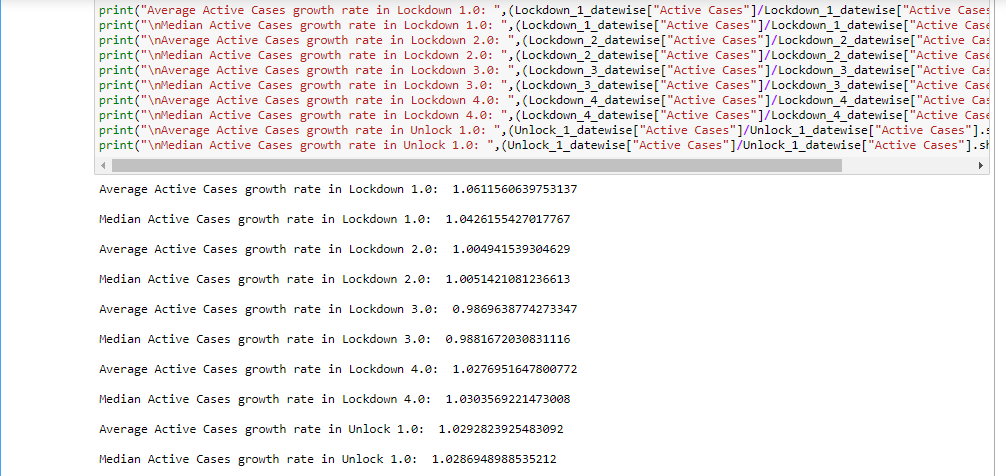


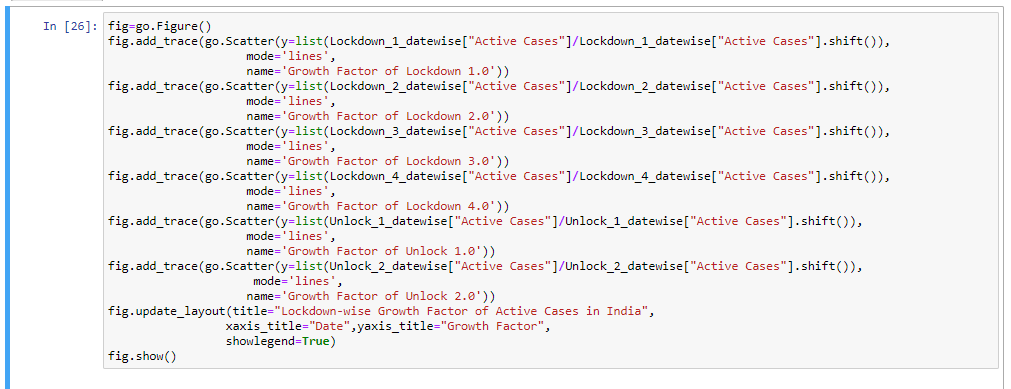


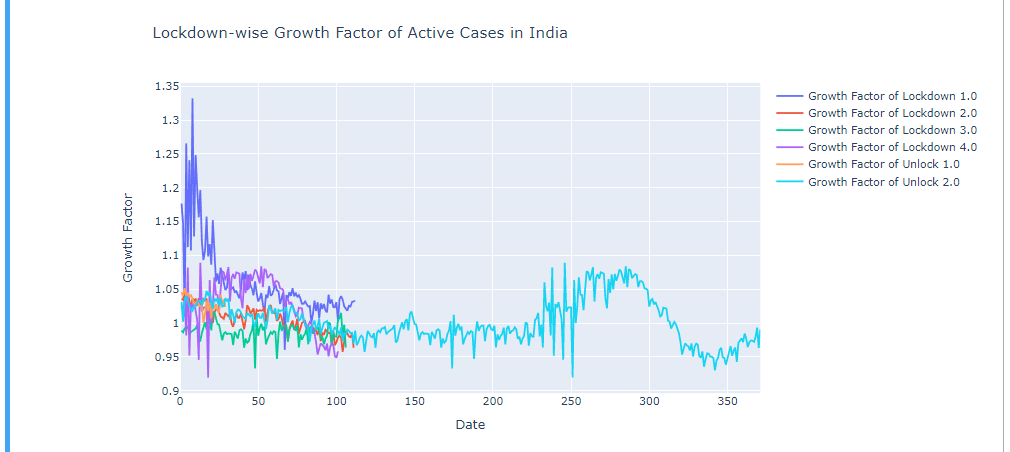
1. Government of India had announced a complete nationwide lockdown several times. Gathering all the time durations for the lockdown we have divided the dataset into groups. Then we have used these groups to show the growth of cases in India.



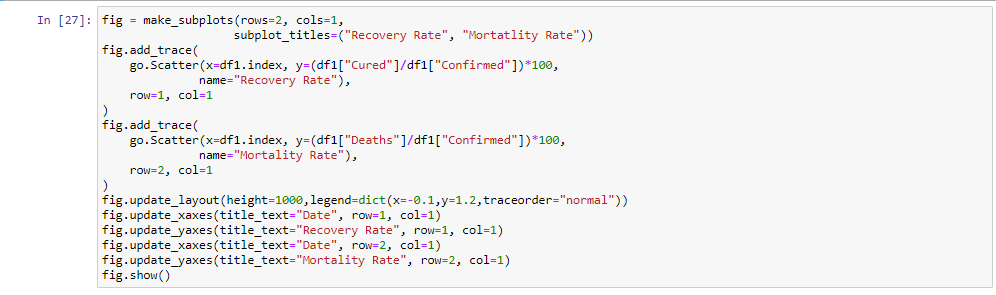


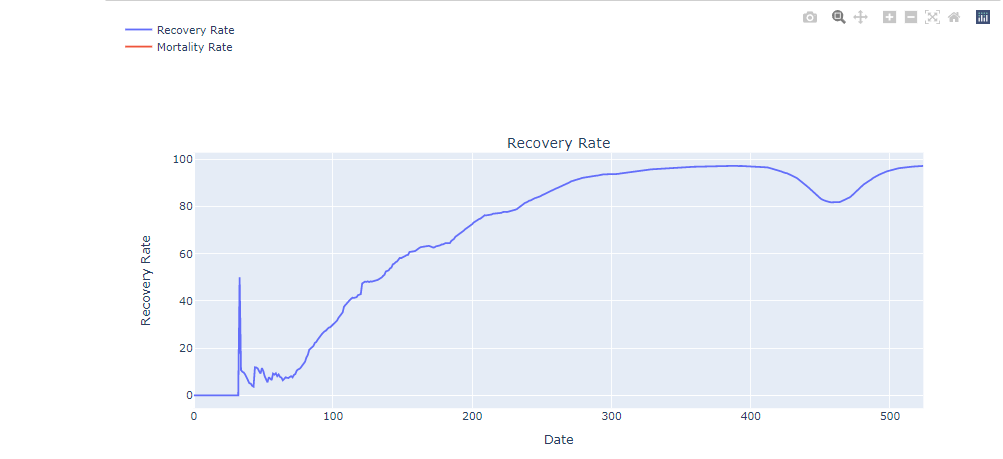


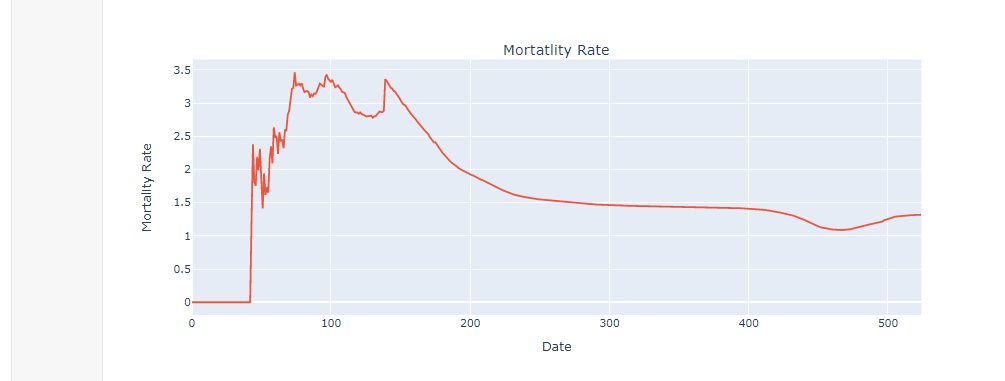




1. Using scatter plot with line and markers we have plotted a graph for Mortality and Recovery rate.







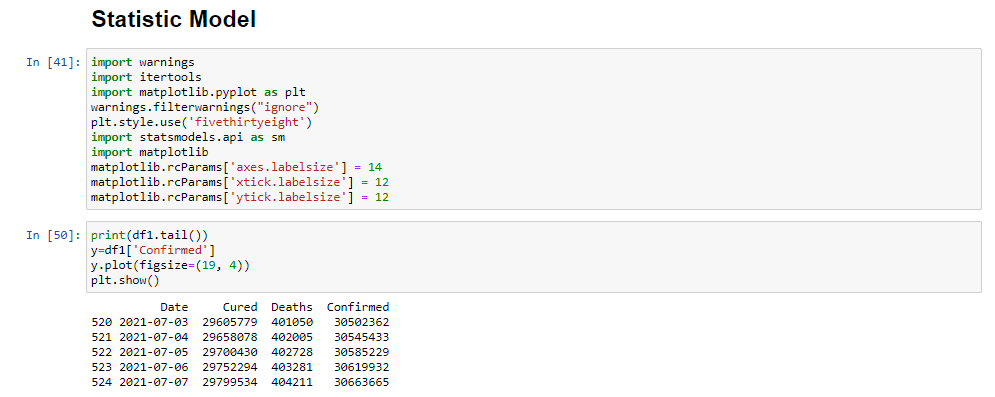
# **Methodology used for data analysis**

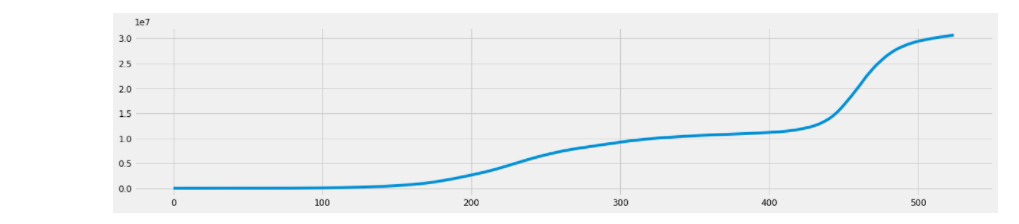
# We have used SARIMA model to forecast predictions of Covid-19 Cases.

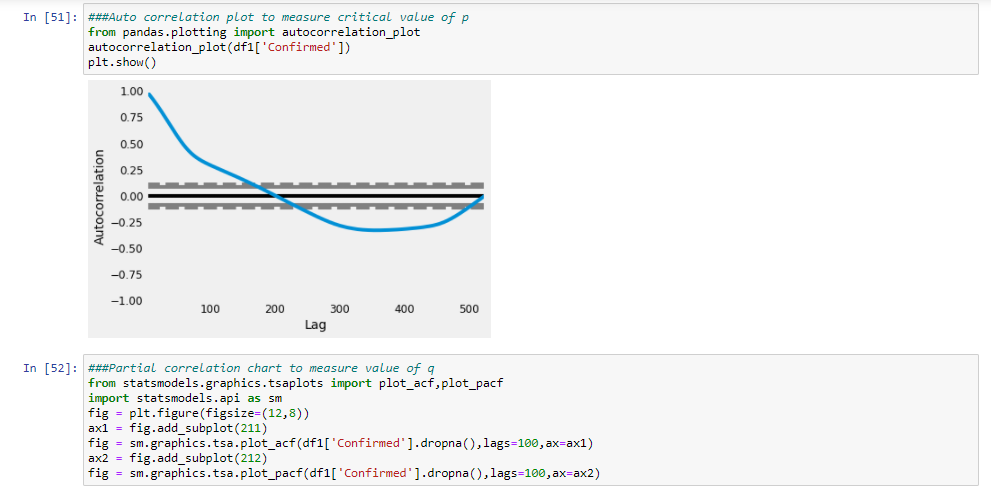
SARIMA stand for Seasonal Auto Regressive Integrated Moving Average.

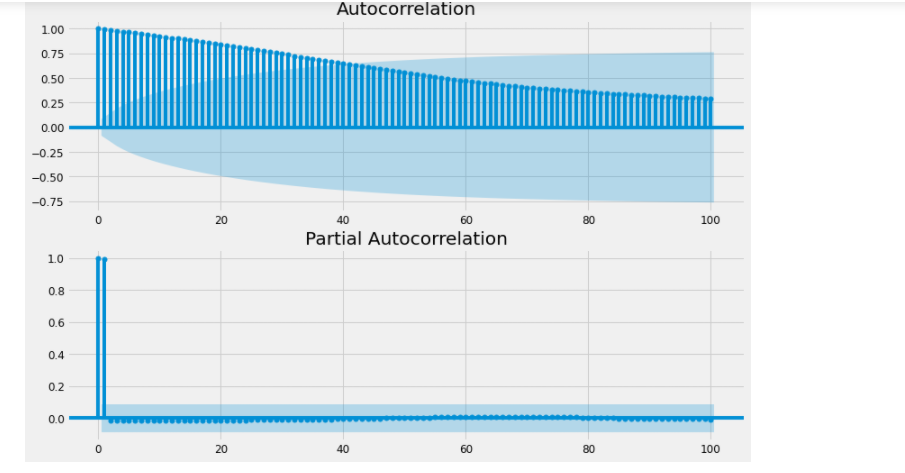
Firstly, we have loaded the data and imported necessary libraries in python and then we go for data preprocessing. In data preprocessing we first ensure that dates are used as index values and are understood by python as true “date” object

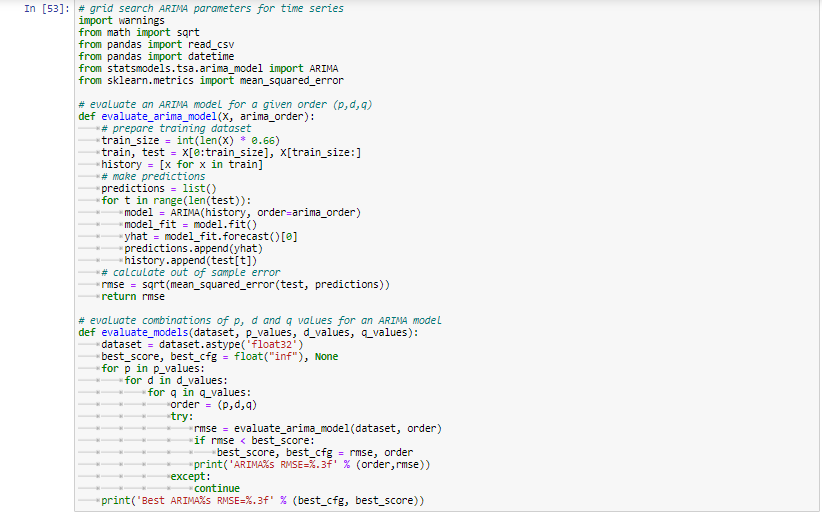
We make an Auto correlation plot to measure critical value of p and a Partial correlation chart to measure value of q. Then we wrote a code for grid search algorithm for arima parameters

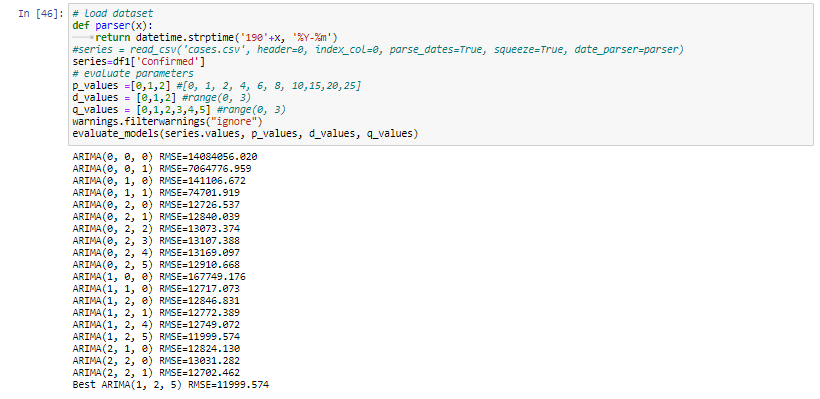




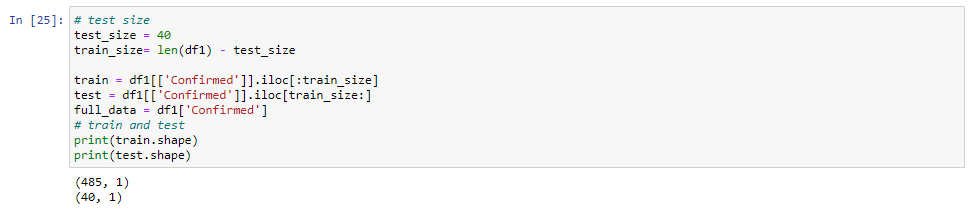


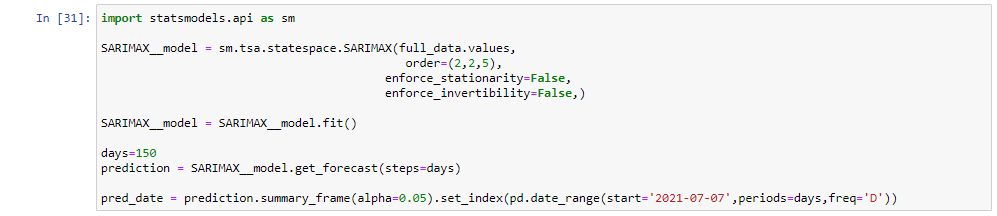






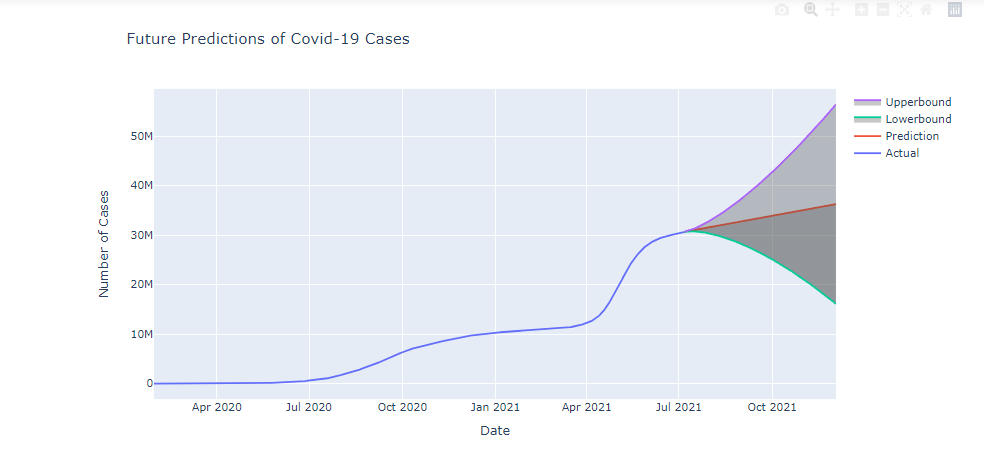
# **Model fitting**





# **Results**

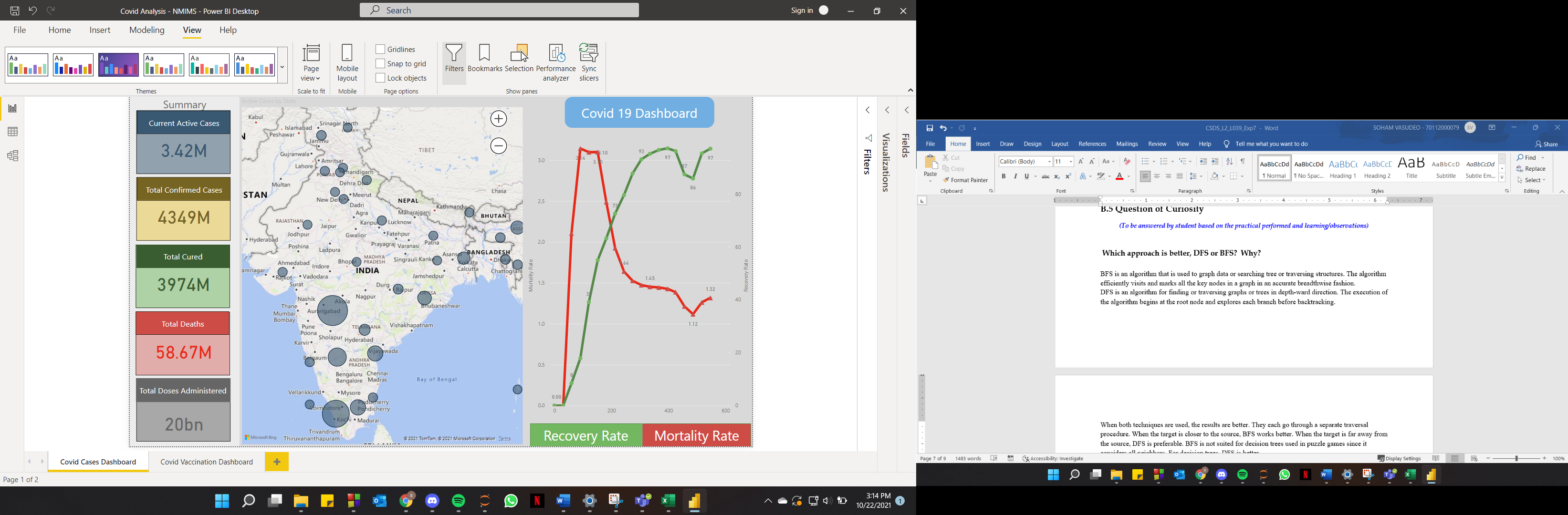




# **Dashboards**

We created 2 Dashboards using Microsoft Power BI Desktop version.

## **Covid Cases Dashboard**



The Covid Cases Dashboard is divided into 3 areas

1. Summary

The Infographics gives a bird’s eye view of the overall India wide current status of active covid cases, confirmed cases, citizens cured and total deaths as well as total vaccination doses administered till date.

1. Map based Analytics

The bubbles in the India map show the active covid cases across states of India.

The location of the bubble is based on latitude and longitude of the state. The size of the bubble denotes the no. of active cases in that state.

When the user hovers over a bubble, it shows a tooltip indicating active cases in the month of July 2021 (which is the current month as per the available data).

When the user clicks on a particular bubble, the details on the summary section also change showcasing the summary of active cases, confirmed cases, citizens cured and total deaths

1. Recovery and Mortality Rate

The Green line in the chart denotes the Recovery rate and the Red line denotes the Mortality rate.

The X axis indicates the number of days from covid -19 start.

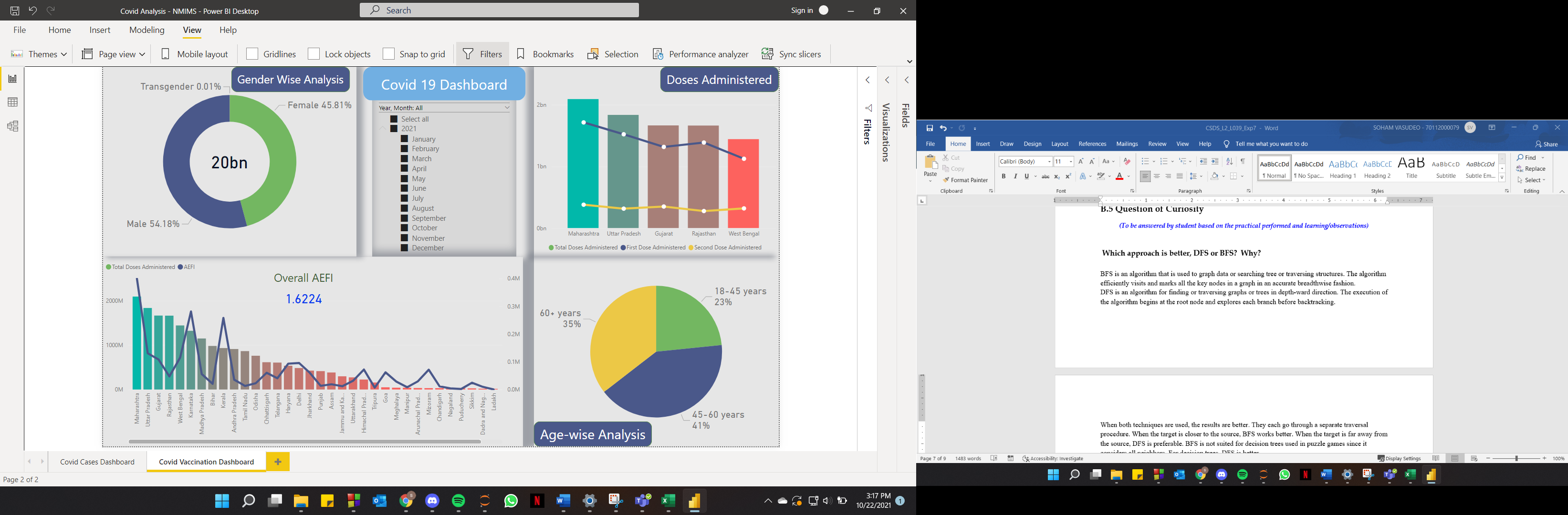
Mortality rate was highest at 91 days (around 4/1/2020) and was at 3.14%

The Mortality and Recovery Rate converged around 200 days (post vaccination took momentum)

Recovery Rate peaked to 97.22% at 397 days (around 2/1/2021) but again dropped to 85.79% when the 2nd wave of covid broke lose.

By July 2021 it was back to 97.08% and the mortality rate dropped to 1.32%

## **Covid Vaccination Dashboard**



The Covid Vaccination Dashboard is divided into 4 areas.

1. Gender Wise Analysis

The donut chart denotes the percentage of dose administered across all 3 genders. We have also provided the number of total doses administered till date in the middle of the donut.

1. Total Dose administered v/s AEFI

The Bar graph denotes the total dose administered across the states. The chart is also sorted in a descending order to showcase the best state on the left followed by others.

The Line graph on the Bar denotes AEFI (Adverse Event Following Immunization) which indicates how many vaccinated people got adverse reactions.

The overall AEFI is also denoted in the top-middle of the graph which indicate the current AEFI percentage across states.

The user can click a particular state or do a multi select with CTRL + States in the bar chart the other visualizations within the dashboard will also change to indicate the values for those states.

1. Doses Administered (1st and 2nd dose Analysis)

This bar graph shows the TOP 5 best states by doses administered and the 2-line graphs show how many 1st and 2nd doses have been administered within that particular state.

1. Age Wise Analysis

The pie chart shows the percentage of doses administered across all 3 age groups

1. Year – Month Filter

All the visualizations in the dashboard are linked by the multi-select Year-Month filter based on Dose administered. Using this filter, the user can do an analysis across all visualizations for a particular or multiple months.