



# DS 250: Data Analysis and Visualization



## Covid 19 Analysis

Sai Rishitha (11840540)

Meghna Singh(11840710)

Prashant Kumar(11840820)

Shanu Gautam(11841040)

# Motivation

- The idea of the project is to find the rate of spread of the disease in the upcoming days with the help of regression analysis models and forecast of COVID-19 in India with the next days for better management for doctors and various government organisations.

# Objective

- Covid 19 case Trend in India
- Creating an ARIMA model to predict the rate of spread of the disease

# Data Collection and Cleaning

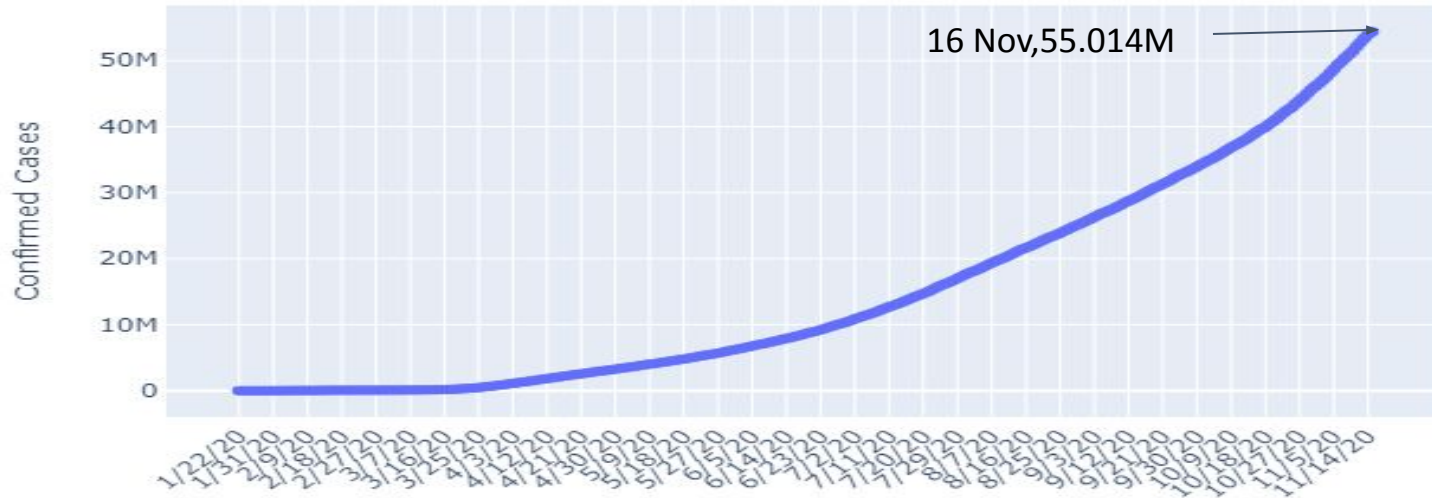
- We collected the links to raw files for COVID-19 dataset provided by JHU CSSE and downloaded them into respective dataframes for further analysis.
- Then , we cleaned the dataset, i.e. converted the numerical values to numbers , dropped insignificant columns such as ['Lat', 'Long', 'Country/Region' , 'Province/State'] .

# Data Collection and Cleaning

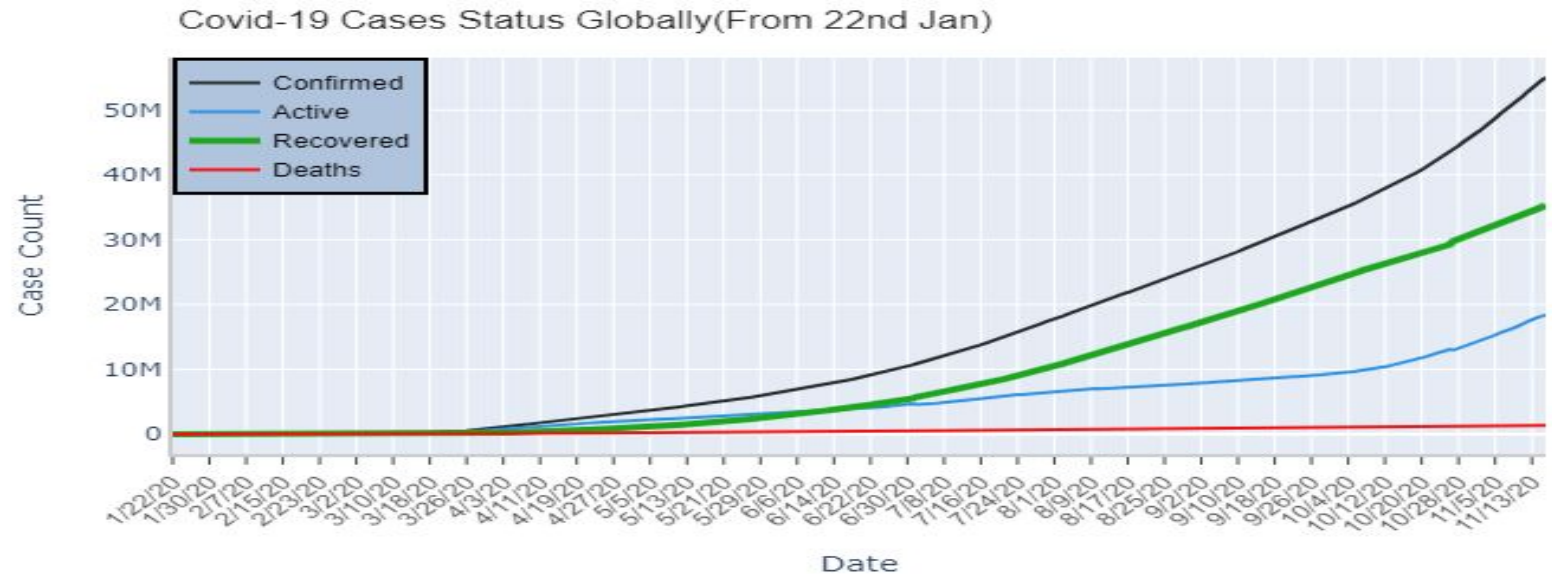
- Since the dataset for active cases was not available, we prepared the dataset of active cases by:  
active = confirmed -(recovered + death)
- Then, we extracted the data for India (as we mainly focus on India) to study the Covid 19 case Trend in India.

# Confirmed cases globally

Total Coronavirus Confirmed Cases (Globally)



# Case Trends Globally



# Ranking of Countries

## **Highest Confirmed cases:**

US, India, Brazil

## **Highest Deaths cases:**

US, Brazil, India

## **Highest Recovered cases:**

India, Brazil, US

## **Highest Active cases:**

US, France, UK

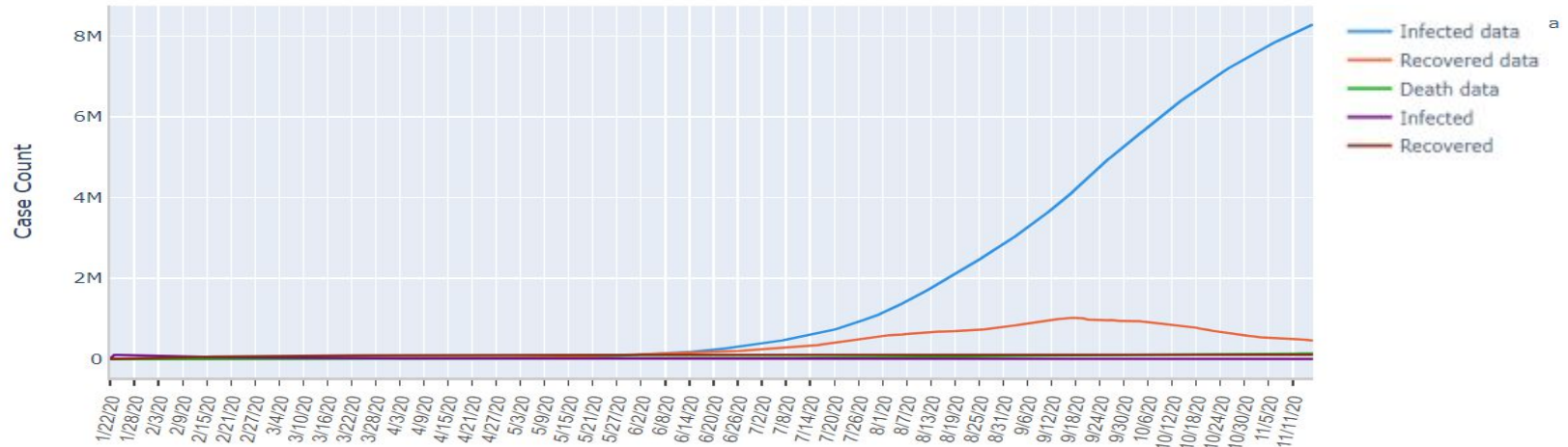


# Models

- SIR Model



SIR Model -- Covid-19 Transmission -- Prediction -- INDIA

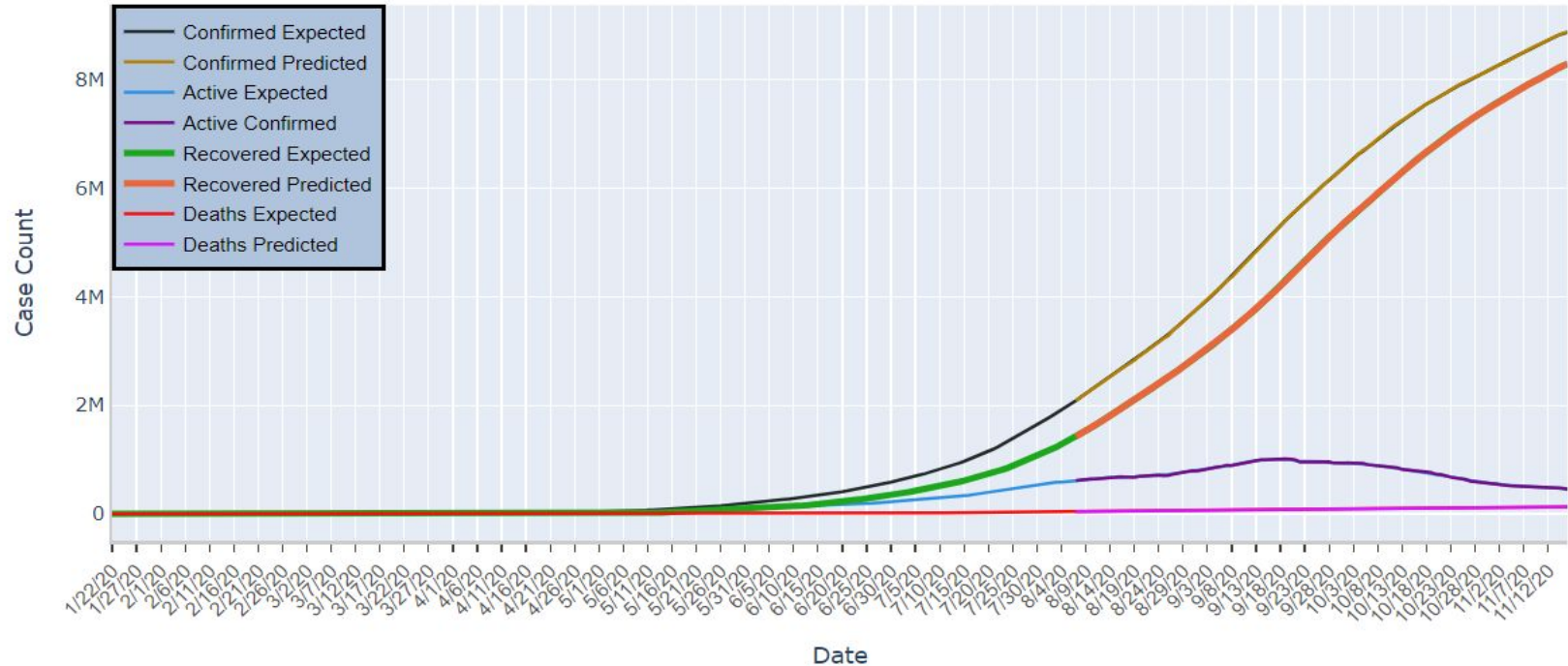


# ARIMA Model

- Rolling forecast ARIMA model
  - >Recreated the ARIMA model for each new observation and the received observation is again fed into the model to do the next forecast.
- How we trained and tested?

# Training and Testing

Covid-19 Case Trend in INDIA

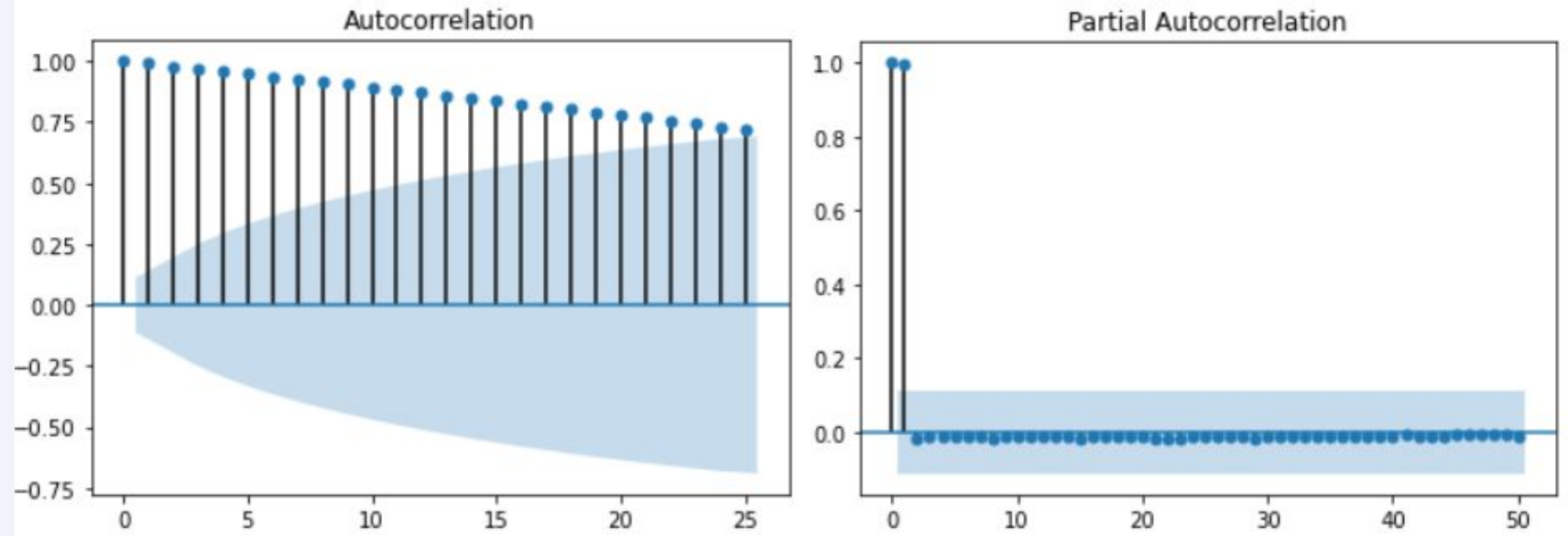


# Testing: Error

- We got the following value of root mean square error for our four datasets:

Dataset	RMSE
Confirmed Cases	6283.996
Recovered Cases	5313.729
Active Cases	7374.860
Deaths	104.470

# ARIMA Model



# ARIMA Model

```

=====
                        ARIMA Model Results
=====
Dep. Variable:          D2.y      No. Observations:          297
Model:                  ARIMA(3, 2, 0)  Log Likelihood          -2868.995
Method:                 css-mle    S.D. of innovations      3791.975
Date:                   Tue, 17 Nov 2020  AIC                    5747.989
Time:                   01:58:34    BIC                     5766.458
Sample:                 01-24-2020    HQIC                    5755.383
                        - 11-15-2020

=====
                        ARIMA Model Results
=====
                        coef      std err          z      P>|z|      [0.025      0.975]
-----
const          119.3312      143.612         0.831      0.406     -162.144      400.806
ar.L1.D2.y     -0.1815         0.059        -3.088      0.002       -0.297       -0.066
ar.L2.D2.y     -0.2962         0.057        -5.189      0.000       -0.408       -0.184
ar.L3.D2.y     -0.0577         0.059        -0.985      0.325       -0.173         0.057
Roots
=====

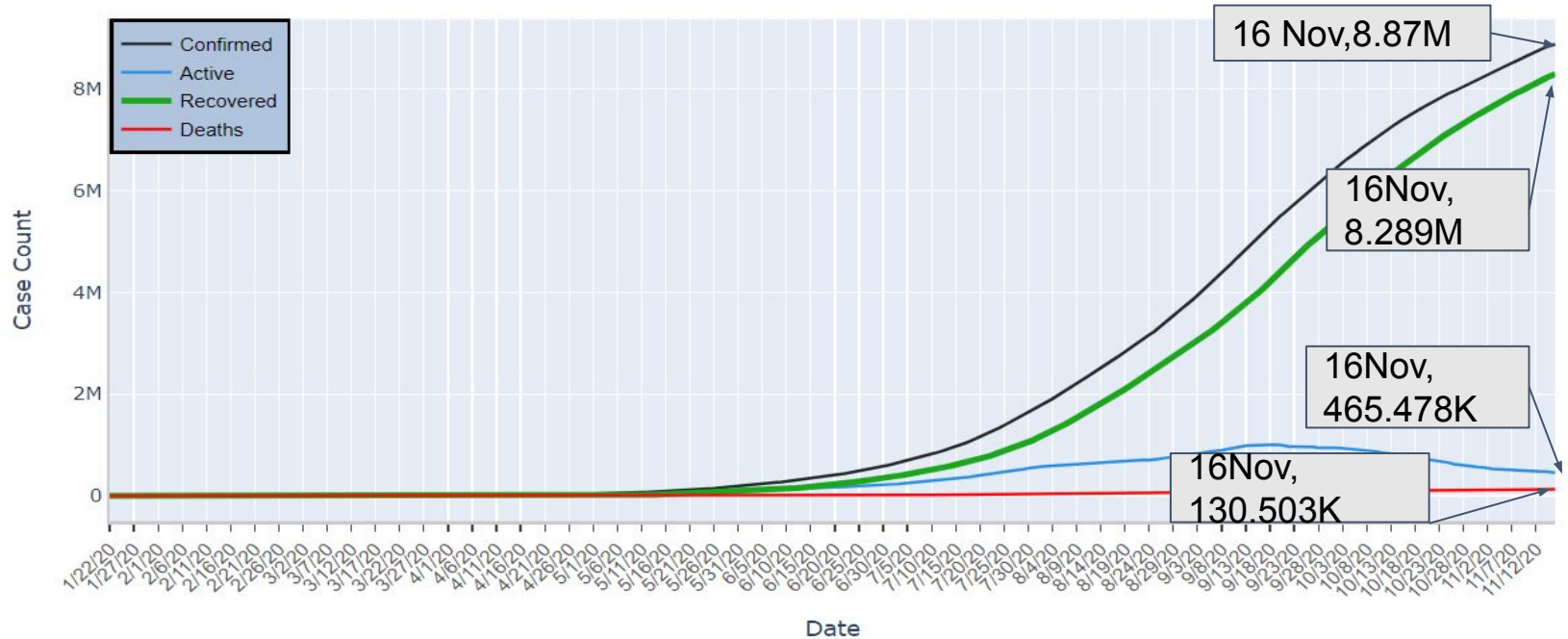
=====
                        ARIMA Model Results
=====
Dep. Variable:          D2.y      No. Observations:          297
Model:                  ARIMA(2, 2, 0)  Log Likelihood          -2869.479
Method:                 css-mle    S.D. of innovations      3798.229
Date:                   Tue, 17 Nov 2020  AIC                    5746.957
Time:                   01:59:55    BIC                     5761.732
Sample:                 01-24-2020    HQIC                    5752.872
                        - 11-15-2020

=====
                        ARIMA Model Results
=====
                        coef      std err          z      P>|z|      [0.025      0.975]
-----
const          117.0142      152.180         0.769      0.442     -181.253      415.281
ar.L1.D2.y     -0.1647         0.056        -2.923      0.003       -0.275       -0.054
ar.L2.D2.y     -0.2861         0.056        -5.086      0.000       -0.396       -0.176
Roots
=====

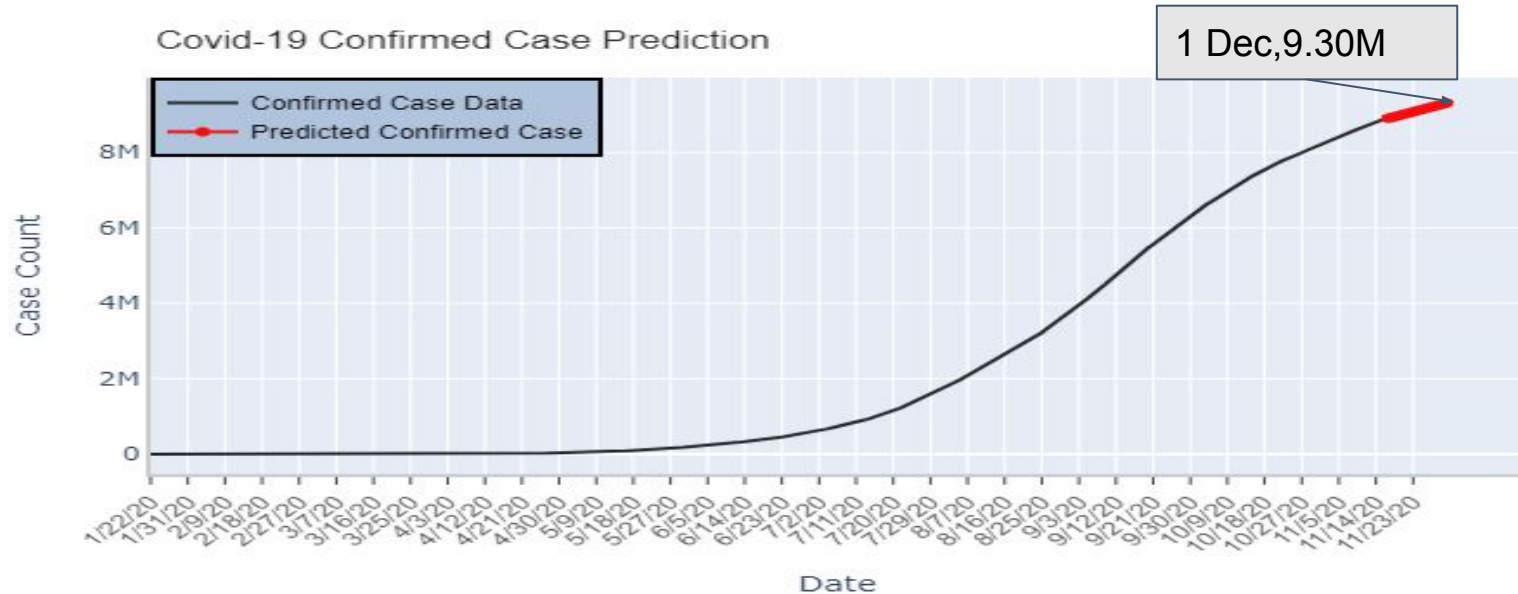
```

# Data visualisations

Covid-19 Cases Trend in INDIA

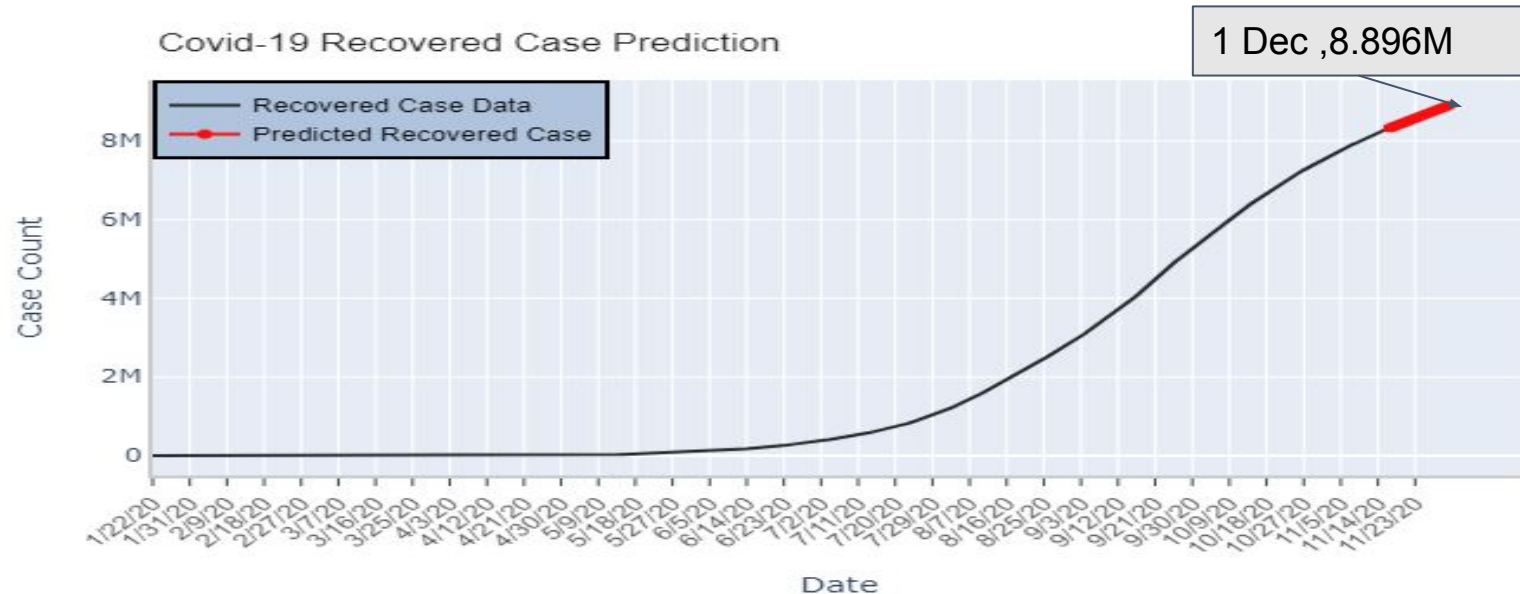


# Predicted confirmed cases for next 15 days.

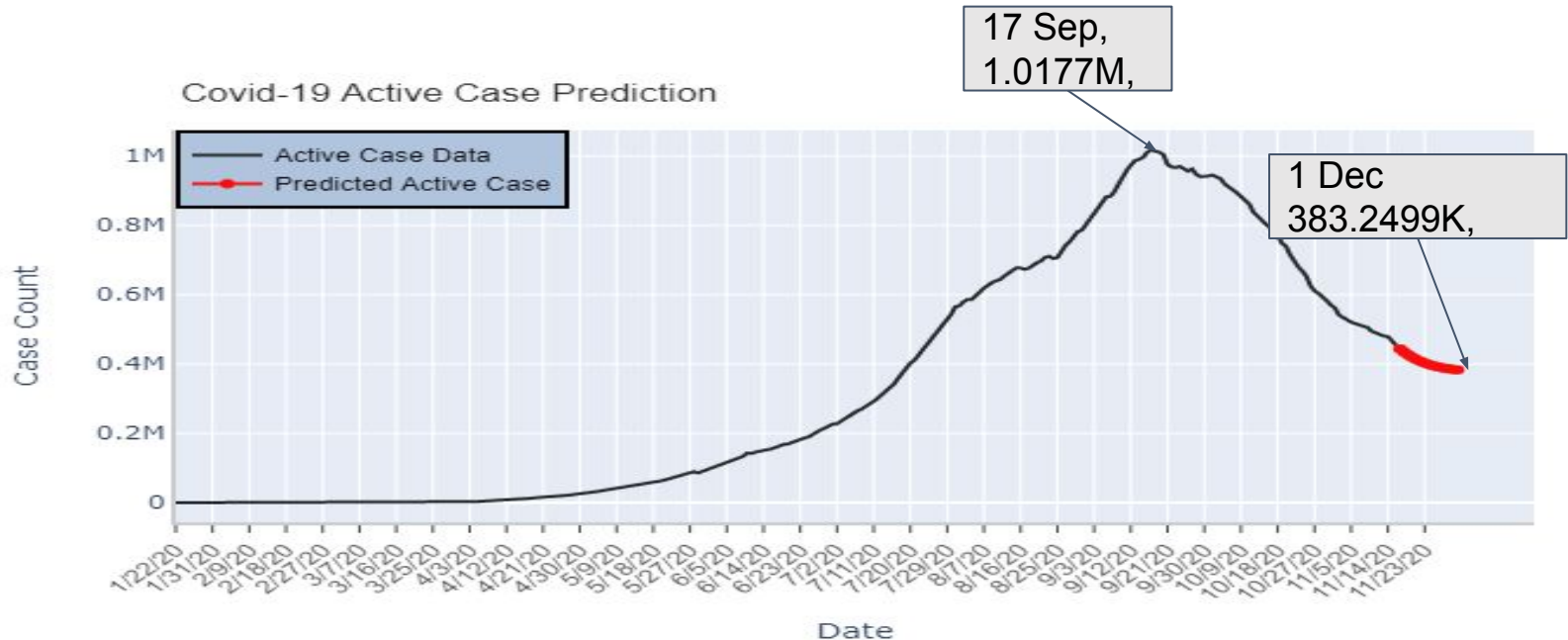




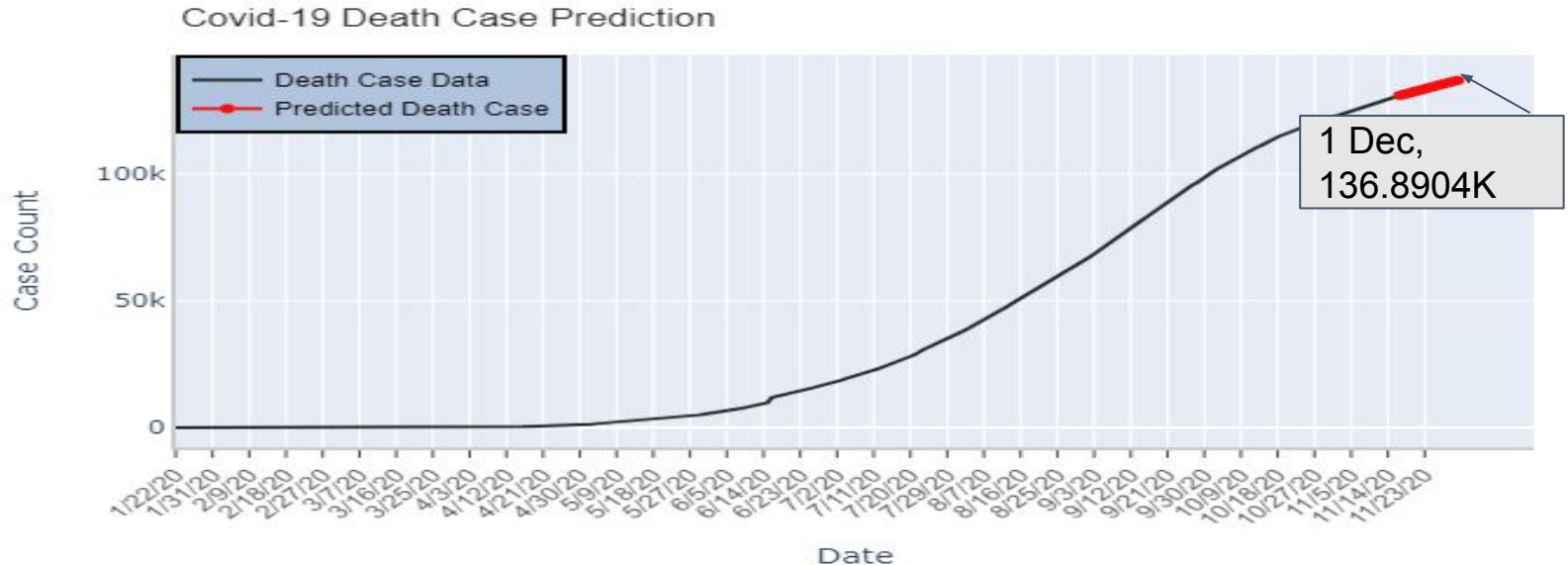
# Predicted Recovered cases for next 15 days.



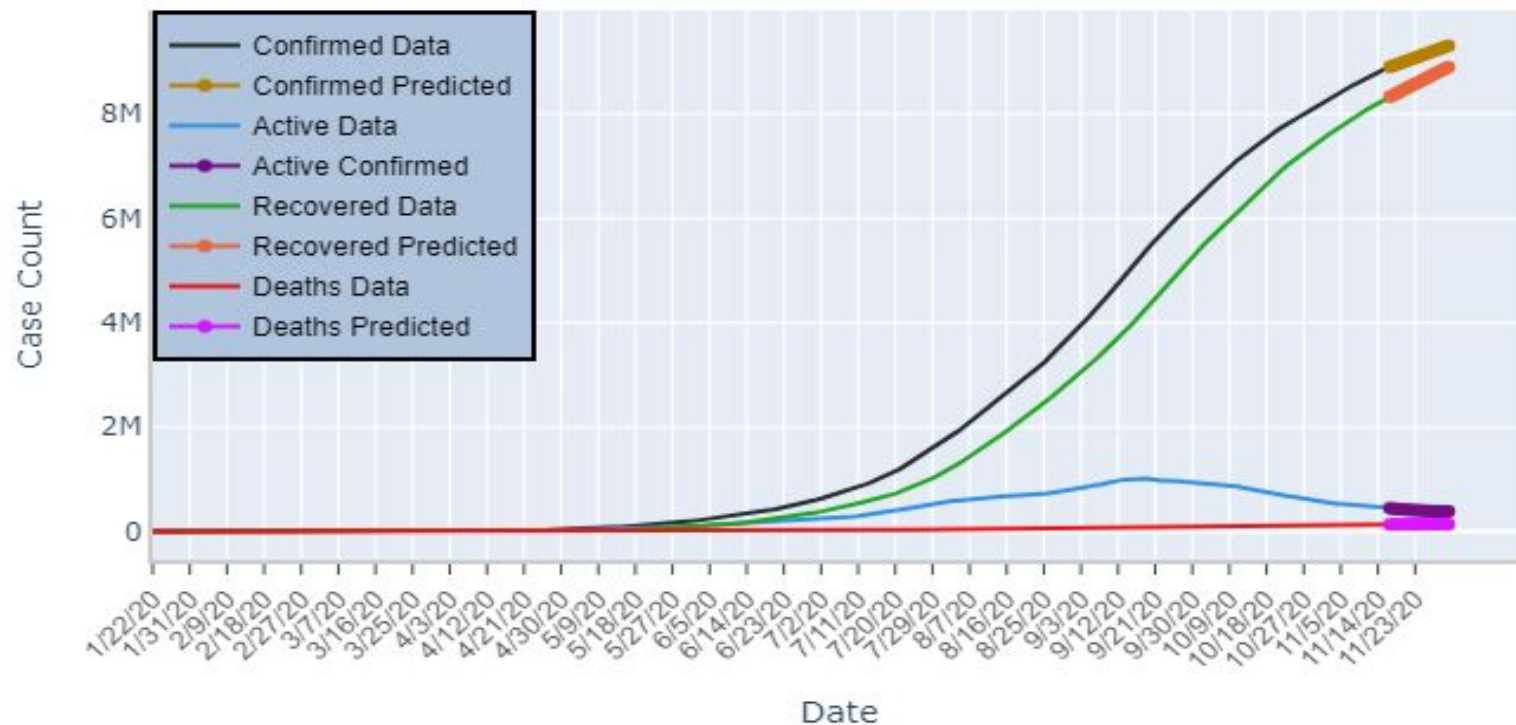
# Predicted Active cases for next 15 days.



# Predicted death cases for next 15 days.



## Covid-19 Case Prediction in INDIA



As people are going through a very challenging global health crisis, data analysis and visualisation helps us to understand what is happening and how to deal with the situation and how to take smart decisions to handle the situation. By doing this project we learnt how to use data science to predict future circumstances based on the current scenarios and datas.

**THANK YOU!**