Capstone Project Presentation

By: Ritika Bhole

Analysing and Predicting the COVID-19 Trend in India

- Introduction
- Problem
- Interest
- Data sources
- Results
- Discussion

Introduction

Coronaviruses are a large family of viruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). The most recently discovered coronavirus causes coronavirus disease COVID-19. COVID-19 is an infectious disease caused by the most recently discovered coronavirus. This new virus and disease were unknown before the outbreak began in Wuhan, China, in December 2019. COVID-19 is now a pandemic affecting many countries globally.

Problem

The spread of COVID-19 has been rapid. The government is collecting plenty data from every state in an attempt to understand the spread and act accordingly. There is insufficient information available to analyse the current and predict the future scenarios. The spread of COVID-19 has to be analysed thoroughly to predict its expanse and when its growth will be curbed.

Interest

Currently, everyone is interested in knowing how and when this pandemic will stop. If we focus specifically on the specialists, then the analysts will be interested in this as to figure out ways to handle the situation. As the country is currently in lockdown and on the situation where it may be extended, this will help the strategists to estimate the duration of lockdown needed to curb the pandemic in India.

Data sources

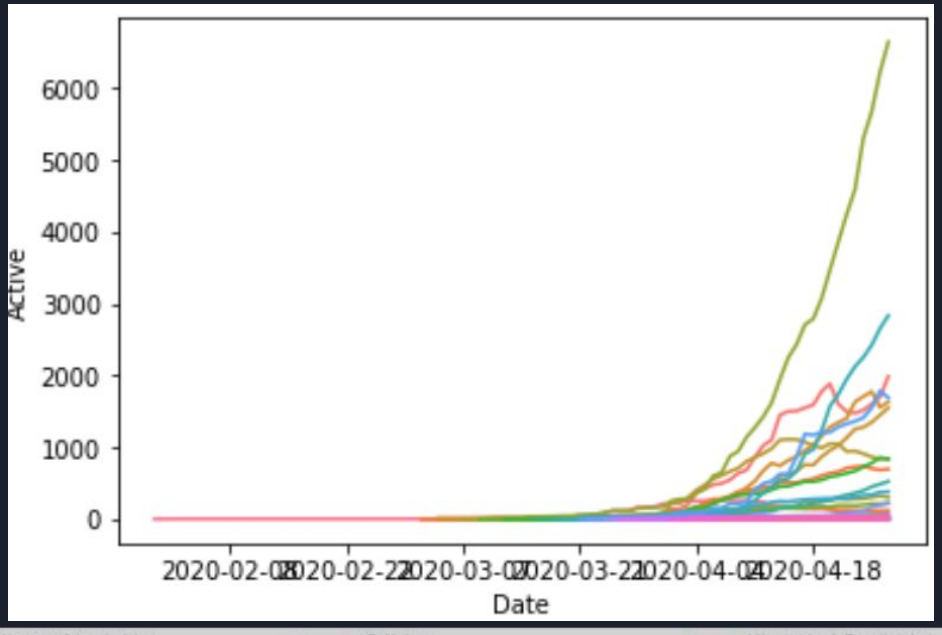
The data is scraped from the authentic data source of the Indian Government's COVID dashboard and website. The data has the following information:

- Date - Date of cumulative report
- Name of State / Union Territory / National Capital Region
- Total Confirmed cases (Indian National) - Cumulative count of Indian national confirmed with COVID-19
- Total Confirmed cases (Foreign National) - Cumulative count of foreign national confirmed with COVID-19
- Cured/Discharged/Migrated - Cumulative count of cured/ discharged cases
- Latitude - Latitude of the location
- Longitude - Longitude of the location
- Death - Cumulative count of deaths reported
- Total Confirmed cases

Methodology

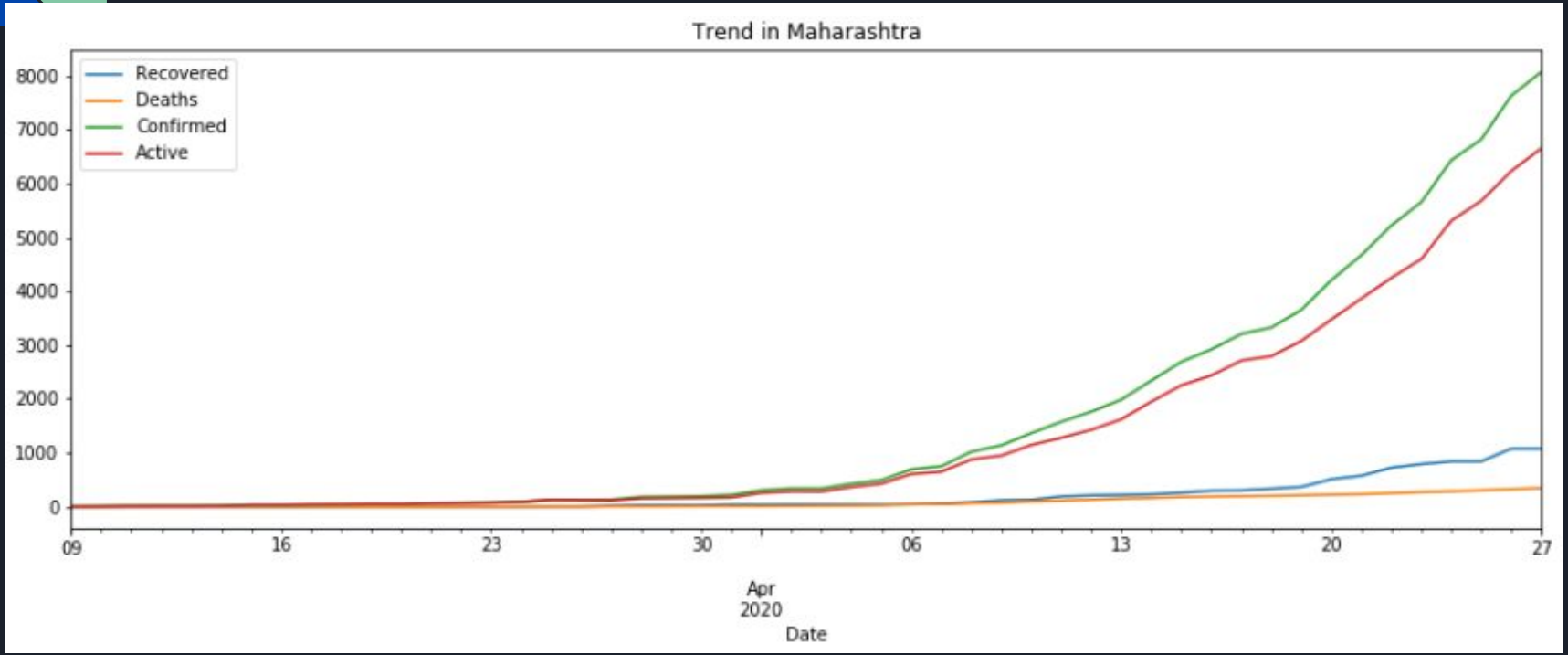
The date attribute in the data is of ‘object’ data type. Conversion of this to datetime datatype helps us to extract more features.
From the date attribute, the days from start can be enumerated, which helps in the plotting of the graph.
‘Active’ attribute is generated, which indicates the number of active cases. It is calculated simply as follows:
Active = Confirmed - Recovered - Deaths

Cases Timeline

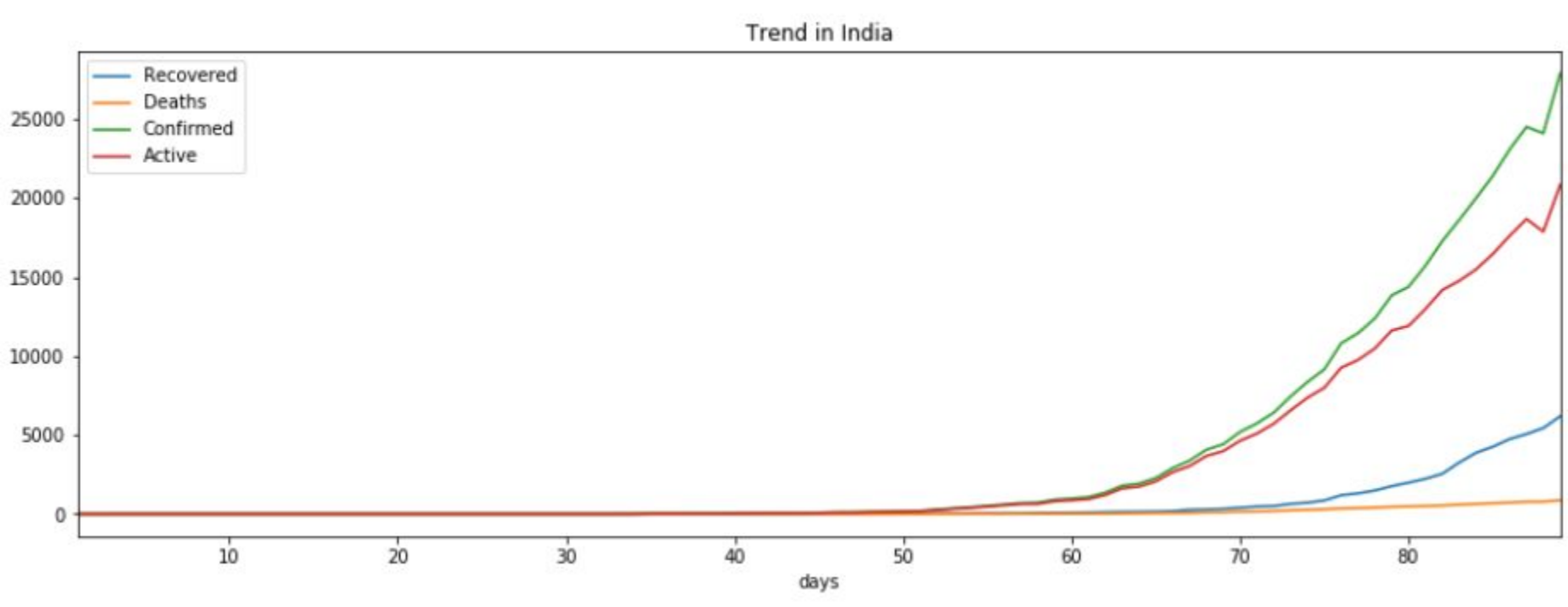


State
Kerala
Delhi
Telengana
Haryana
Rajasthan
Uttar Pradesh
Tamil Nadu
Union Territory of Ladakh
Karnataka
Maharashtra
Punjab
Union Territory of Jammu and Kashmir
Andhra Pradesh
Uttarakhand
Odisha
Puducherry
West Bengal
Chhattisgarh
Union Territory of Chandigarh
Gujarat
Chandigarh
Himachal Pradesh
Jammu and Kashmir
Ladakh
Madhya Pradesh
Bihar
Manipur
Mizoram
Andaman and Nicobar Islands
Goa
Assam
Jharkhand
Arunachal Pradesh
Tripura
Meghalaya

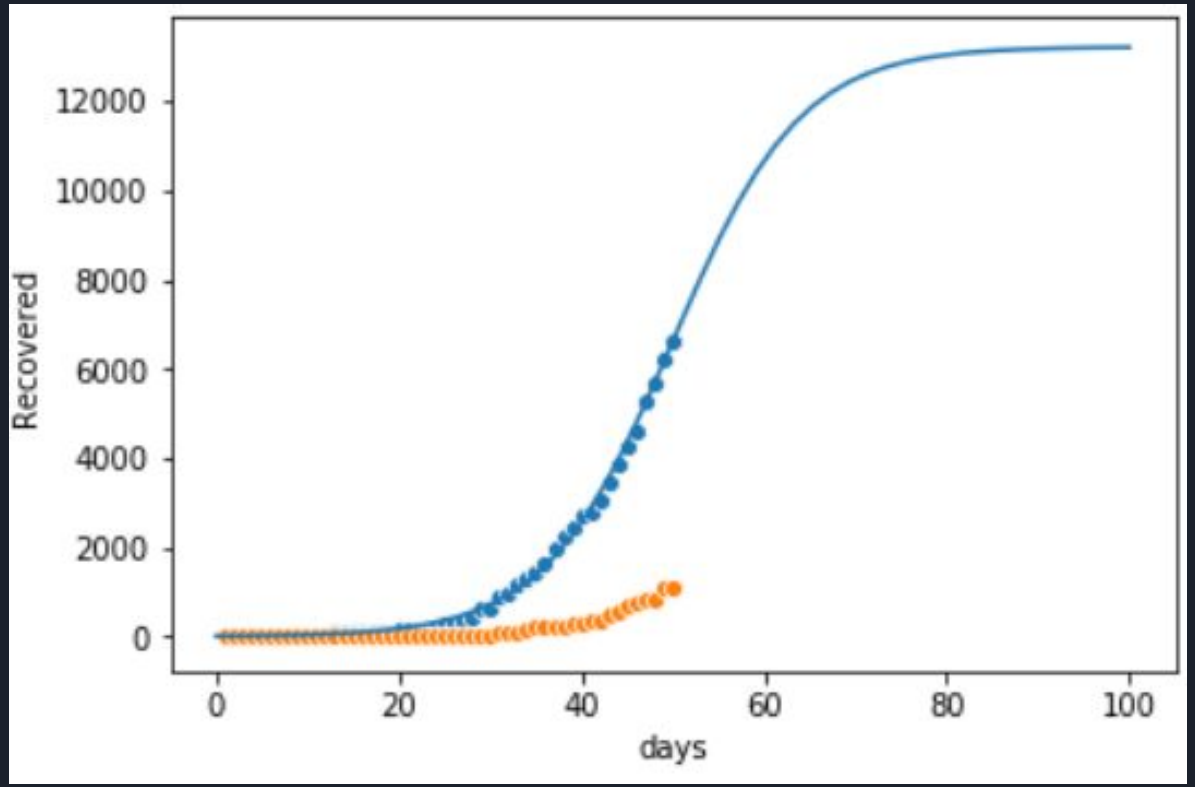
Trend in Maharashtra



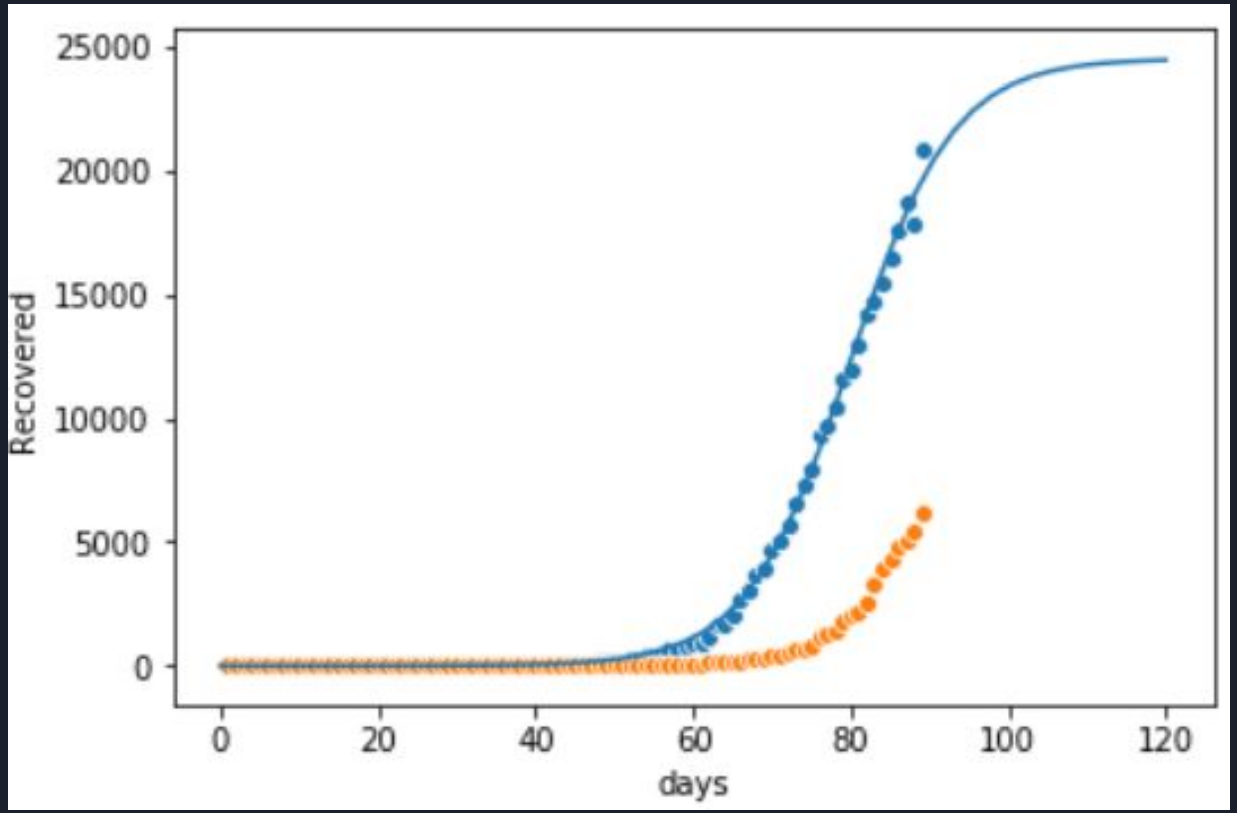
Trend in India



Sigmoidal Curve - Maharashtra



Sigmoidal Curve - India



Discussion

Although we see that the sigmoid curve fits the data well, we're still unsure of how the factors affect the spread.

Adding more features could be the future scope of the analysis

Conclusion

Curve flattens at around day 75 for Maharashtra and at around 115 for India.

As seen the duration of the dates is when lockdown had started and is continued. Hence, continuation of it till the recommended days will help to curb this pandemic.