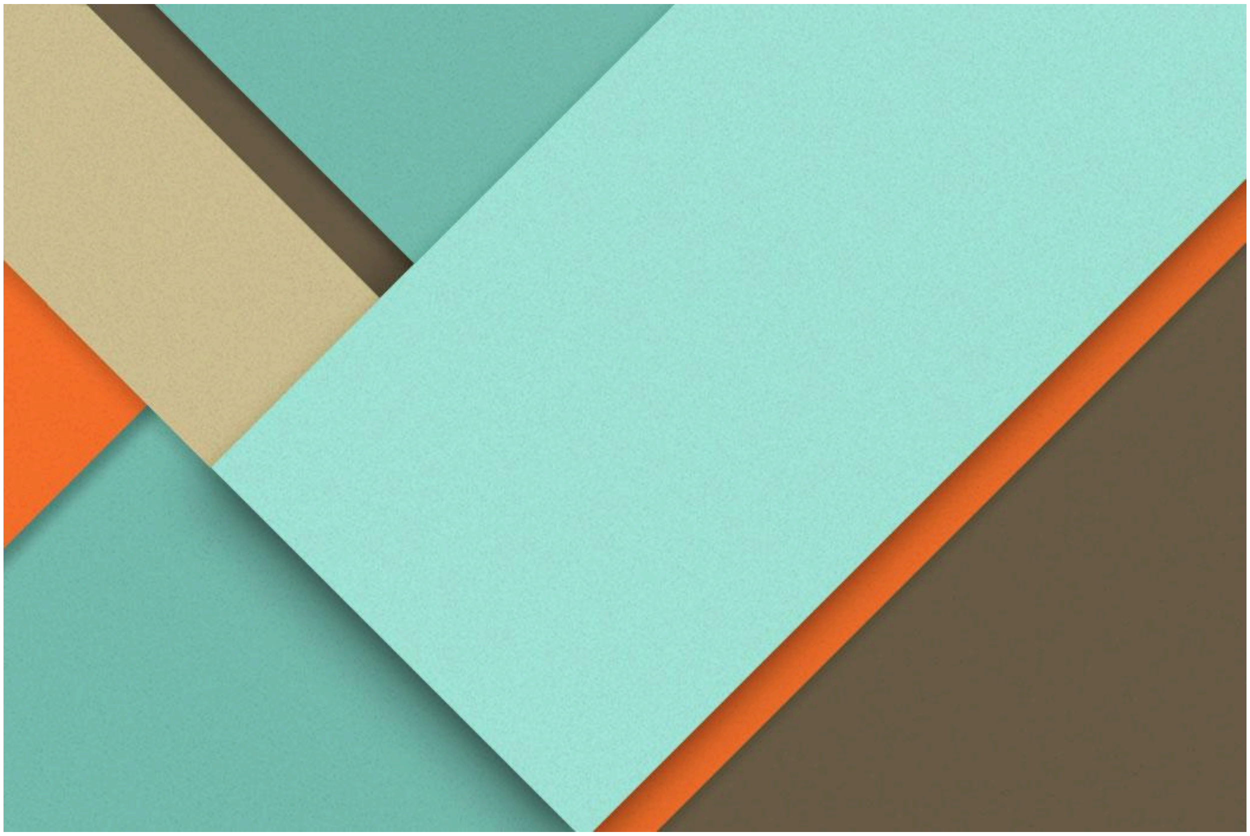


Assignment 1

Question 3

Abhinav Prakash
21CH10001
Course Code: CH620003

Problem Statement

Most viral outbreaks have a finite time period (short duration latency) during which the individual is infected but not yet infectious. Typically, this is modelled using the susceptible (S) – infected (I) – recovered (R) model. Of course there are many variations to the SIR model, for this problem, we stick to the basic version. The recovered fractions include the individual who are expired as well as healthy (after getting infected) too.

During the outbreak period, the total population (N) is assumed constant (through new births, migration, deaths, etc.), $\dot{N} = \dot{S} + \dot{I} + \dot{R} = 0$.

The dynamics of the disease transmission in this case is described as

$$\frac{dS}{dt} = -\frac{\beta}{N}SI; \ \frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I; \ \frac{dR}{dt} = \gamma I. \qquad (2)$$

Here, β represents the average contact rate of infection, signifying the transmission rate in the population. The average recovery rate is denoted by γ, and can be strongly correlated to the clinical observation for different virus variants. The reproduction number, $R_0 = \beta/\gamma$, which suggests when the transmission rate is more than the recovery, $R_0 > 1$, leads to an outbreak.

(a) Use the data available (from any online resource) for the COVID-19 spread dynamics in India and report the parameters R_0 and γ based on the above SIR model equation, for the first , second and possibly the third wave.

(b) Develop an optimised Neural network model to correlate the number of infected and recovered individuals with time from the initial period, starting from mid-March 2020 (you can ignore the few cases which were reported in Jan and Feb 2020). You may need to use the long short term memory (LSTM) time series network model or other recurrent network models.

Divide the total available dataset (available till today from March 2020) into training and testing data. The train dataset can be the first 'x' number of months (say 0-30 months, or the first wave as training and second wave as testing, or something similar) and remaining as test dataset. You should NOT randomly pick data points for the testing.

The mean squared error on the test data set should be more than 0.9. Note that adjusting the relative proportion of data available for training and testing, can drastically affect the model predictions over the test data.

(c) Find out the optimised number of nodes needed for the problem. You may consider only 1 hidden layer.

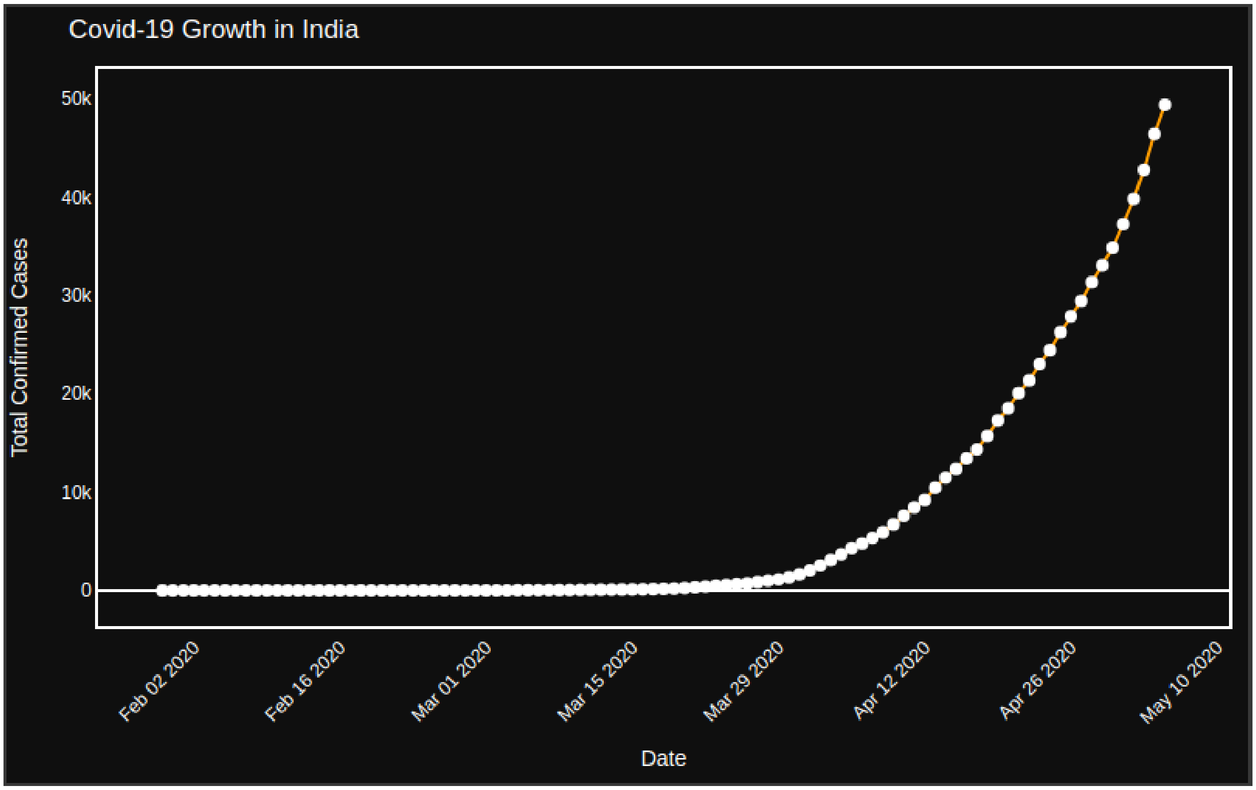
(d) Make a future forecast of the infection dynamics for the next 2 years, starting from today, using your trained Neural Network model.

(Hint: Please make some background reading about the LSTM model and use standard available codes for the same)

Google Colab Link

https://colab.research.google.com/drive/1MLokZlpQUagPBkgOUe-mV-uFqsLTWhgV?usp=sharing

Growth Plot



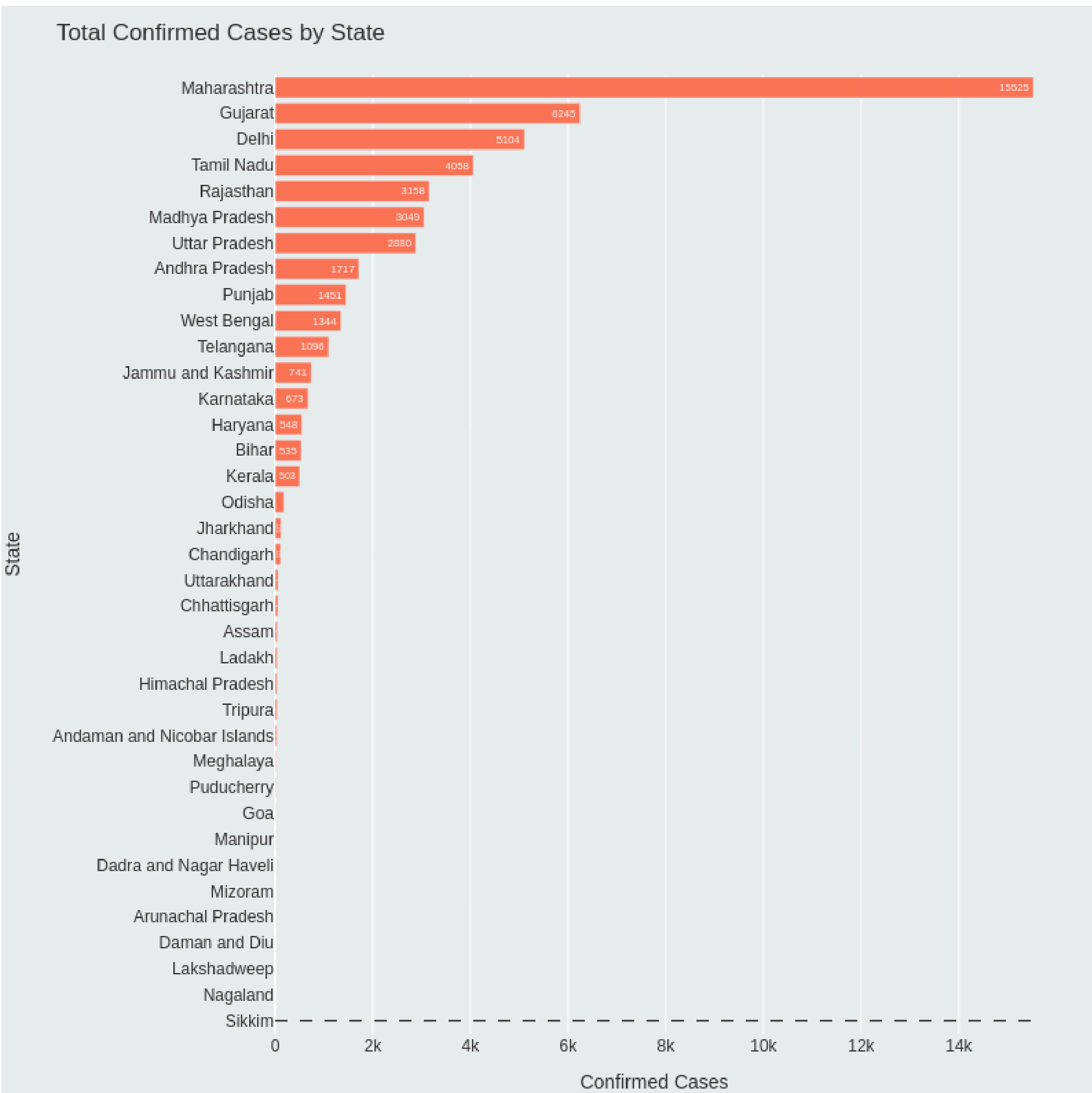
The curve exhibits an exponential upward trend, but the surge in the number of cases has not been as significant. Several plausible explanations would account for this:

- **Limited Testing:** The relatively low increase might be attributed to the possibility that the number of conducted tests is significantly lower compared to other nations. A higher testing volume often correlates with the identification of more cases.
- **Absence of Community Transmission Trigger:** It's conceivable that community transmission has not been activated in India. This could be influenced by various factors, including climatic conditions or other yet-to-be-identified variables.
- **No Ongoing Community Transmission:** Another hypothesis is that community transmission is not occurring at present. Stringent measures, public awareness, or early detection may be contributing to the containment of community spread.

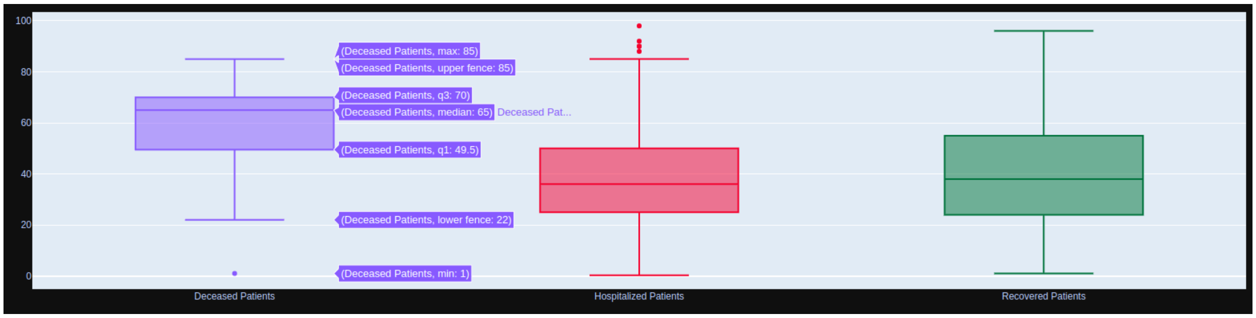
- **Higher Immunity Levels:** The population might exhibit a higher level of immunity, either due to previous exposure to similar pathogens or other immunization factors. This increased immunity could potentially contribute to a more controlled spread of the disease.

It's important to note that these are potential explanations, and a comprehensive analysis considering various factors is necessary to gain a clearer understanding of the observed trend.

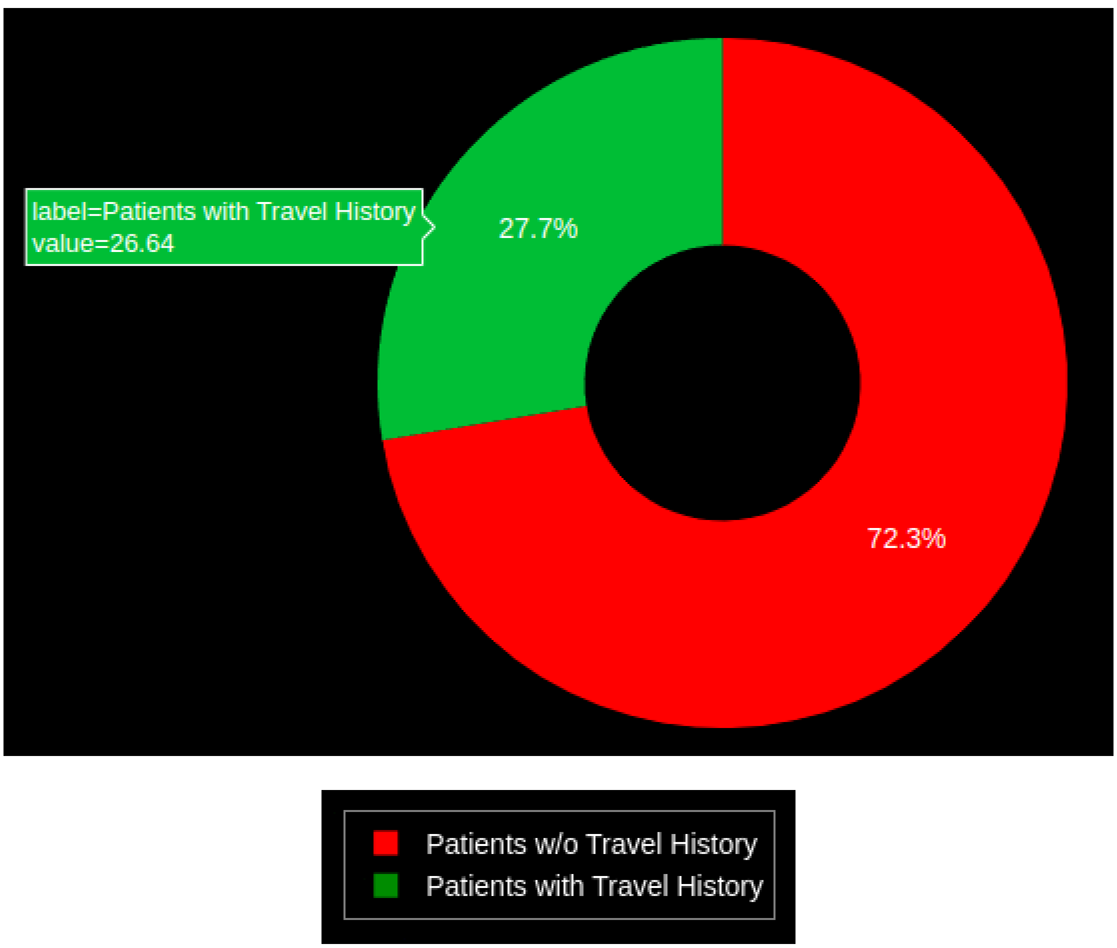
Number of cases in India <State-wise>



From Graph it is clearly evident that Maharashtra had highest no. of confirmed cases among all the states.

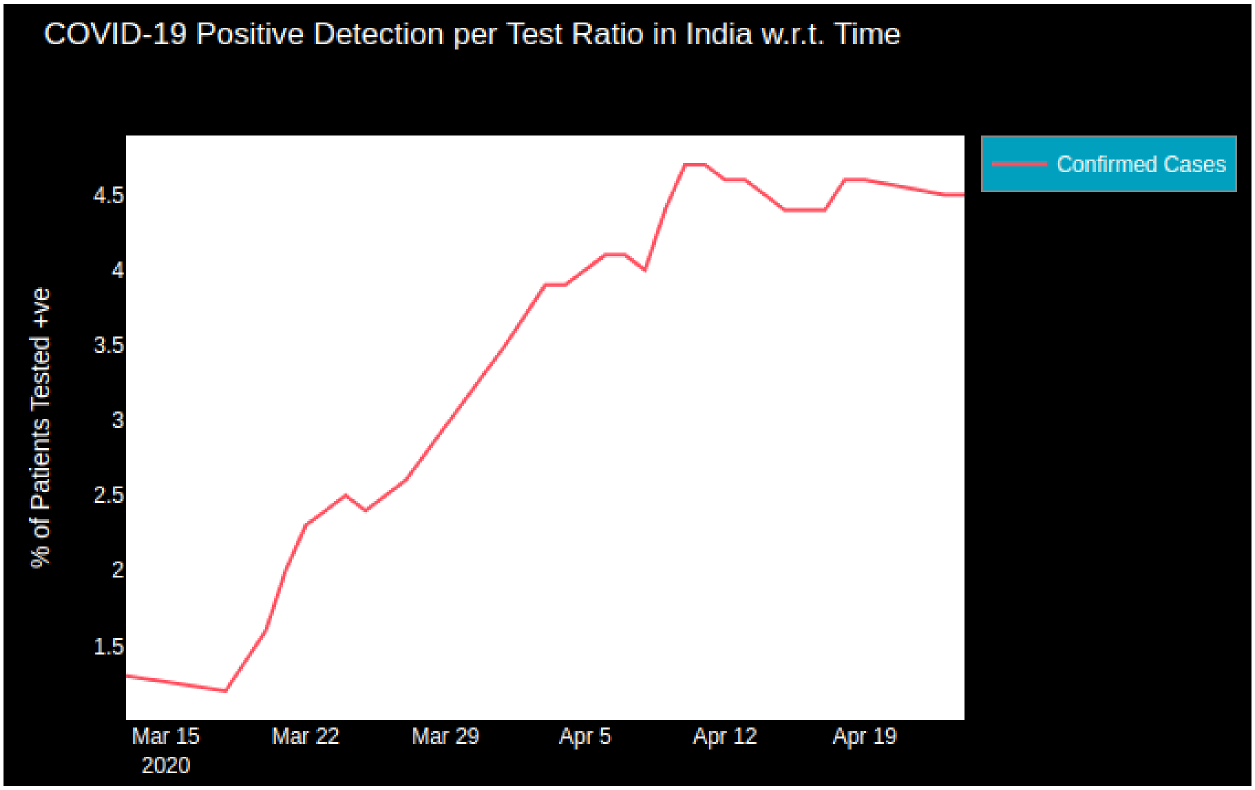


Based on the provided data, it is evident that older individuals exhibit higher vulnerability to diseases. The working-age population, typically falling within the 20-40 age range, appears to be more affected. However, their recovery rate is considerably high, likely attributed to their robust immune systems.



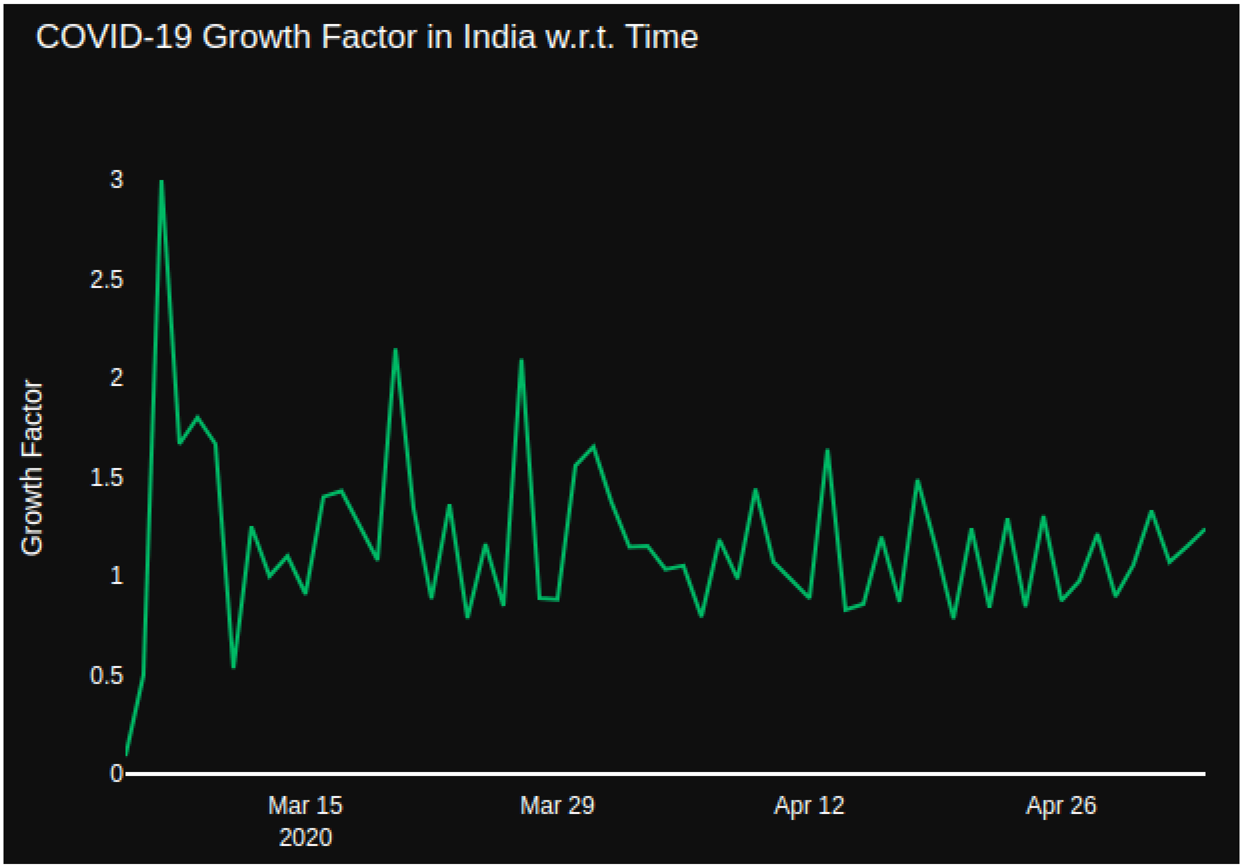
How did individuals without any travel history contract the infection? One possibility is that they might have been in close contact with individuals who did have a travel history. Various scenarios could contribute to this phenomenon.

Covid-19 Test results in India



Clearly after lockdown, there was a surge in the no. of confirmed cases i.e. they might have got infection during lockdown itself, may be infection spread via. Air or water?

Growth Factor in India

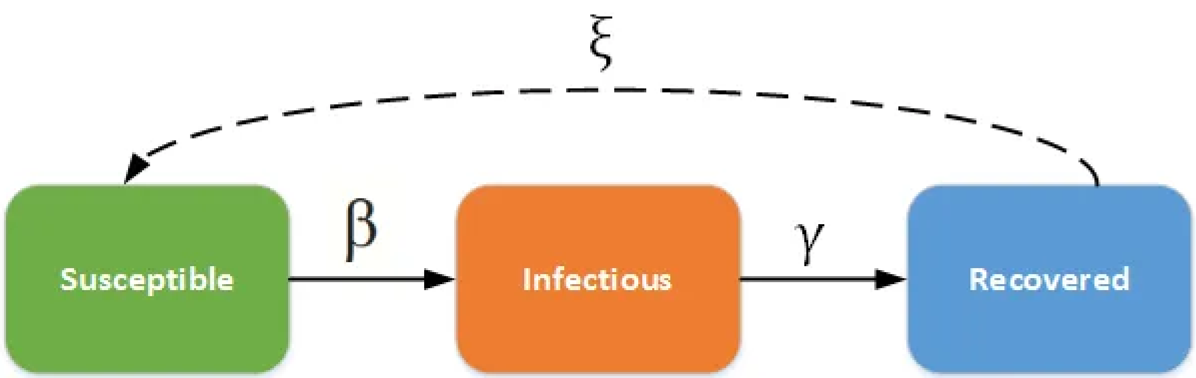


The growth factor is a measure obtained by dividing the number of new cases on a given day by the number of cases reported on the previous day. If this ratio stays consistently above 1, it signals exponential growth. In our case, the majority of the days show a ratio greater than 1, indicating the spread of the pandemic in our country. After analyzing the initial data, we need a model to quantify this growth. One straightforward model we can use is the SIR Model.

SIR Model:

The SIR model is a simple epidemiological model used to understand and predict the spread of infectious diseases within a population. The acronym stands for Susceptible (S), Infected (I), and Recovered (R). The model divides the population into these three compartments, tracking how individuals transition between them over time.

In the SIR model, individuals move from the Susceptible group to the Infected group, where they can infect others, and eventually to the Recovered group, gaining immunity. The model uses differential equations to describe these transitions based on parameters like the contact rate and recovery rate. While simplistic, the SIR model provides valuable insights into the dynamics of disease transmission and is widely used in epidemiology.

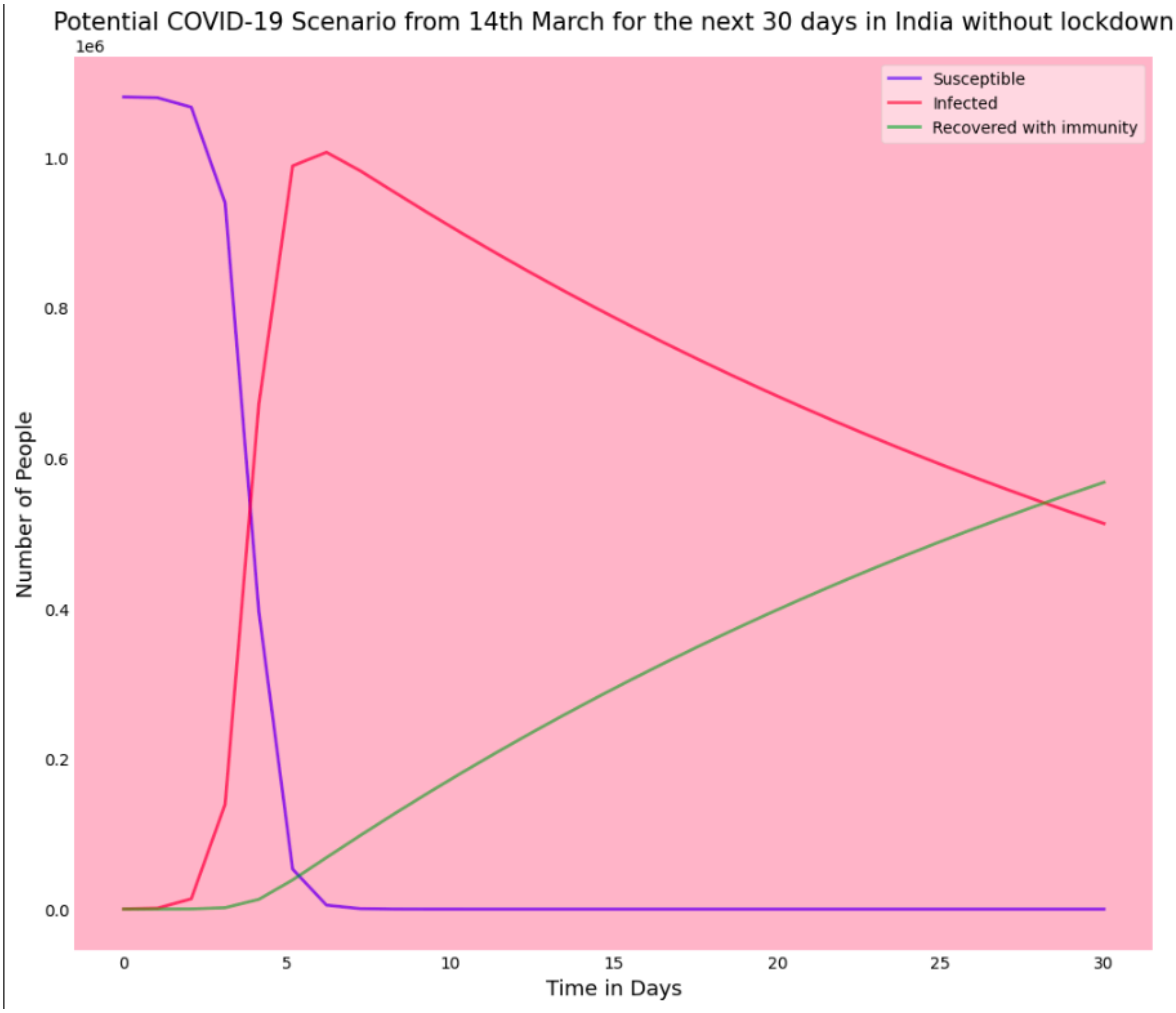


Code:

```
#Total population, N

N=1080000 #Considering a rough estimate of 10 lakhs as
population of India who might have been exposed out of 135 crore

#Initial number of infected and recovered individuals, I0 and
R0.

I0, R0 = 102, 19 #till india cross 100 cases
```

```
#Everyone else, S0, is susceptible to infection initially
S0 = N - I0 - R0

#Contact rate, beta and mean recovery rate gamma, (in 1/days).

beta, gamma = 2.4, 1./35 # Considering Beta and Gamma value
based on China's and Europe situation

#A grid of time points(in days)

t=np.linspace(0,30,30) # 30 values between 0 and 30

#The SIR model differential equations
def deriv(y, t, N, beta, gamma):
 S, I, R = y
 dsdt = -beta*I*S/N
 didt = beta*S*I/N-gamma*I
 drdt = gamma*I
 return dsdt, didt, drdt

#Initial conditions vector
y0 = S0, I0, R0

#Integrate the SIR equations over the time grid, t.
from scipy.integrate import odeint
ret = odeint(deriv, y0, t, args=(N, beta, gamma))
```

```
S, I, R = ret.T


# Plot the data on three separate curves for S(t), I(t) and R(t)
import matplotlib.pyplot as plt
fig = plt.figure(facecolor='w', figsize=(12, 10))
ax = fig.add_subplot(111, axisbelow=True)
ax.plot(t, S, 'b', alpha=0.5, lw=2, label='Susceptible')
ax.plot(t, I, 'r', alpha=0.5, lw=2, label='Infected')
ax.plot(t, R, 'g', alpha=0.5, lw=2, label='Recovered with
immunity')
ax.set_xlabel('Time in Days', size=13)
ax.set_ylabel('Number of People', size=13)
ax.yaxis.set_tick_params(length=0)
ax.xaxis.set_tick_params(length=0)
legend = ax.legend()
ax.set_facecolor('pink')
legend.get_frame().set_alpha(0.5)
for spine in ('top', 'right', 'bottom', 'left'):
   ax.spines[spine].set_visible(False)
ax.set_title('Potential COVID-19 Scenario from 14th March for
the next 30 days in India without lockdown', size=15)
plt.show()
```



Estimating Beta and Gamma for India for SIR Modelling and predicting for the next 6 months:

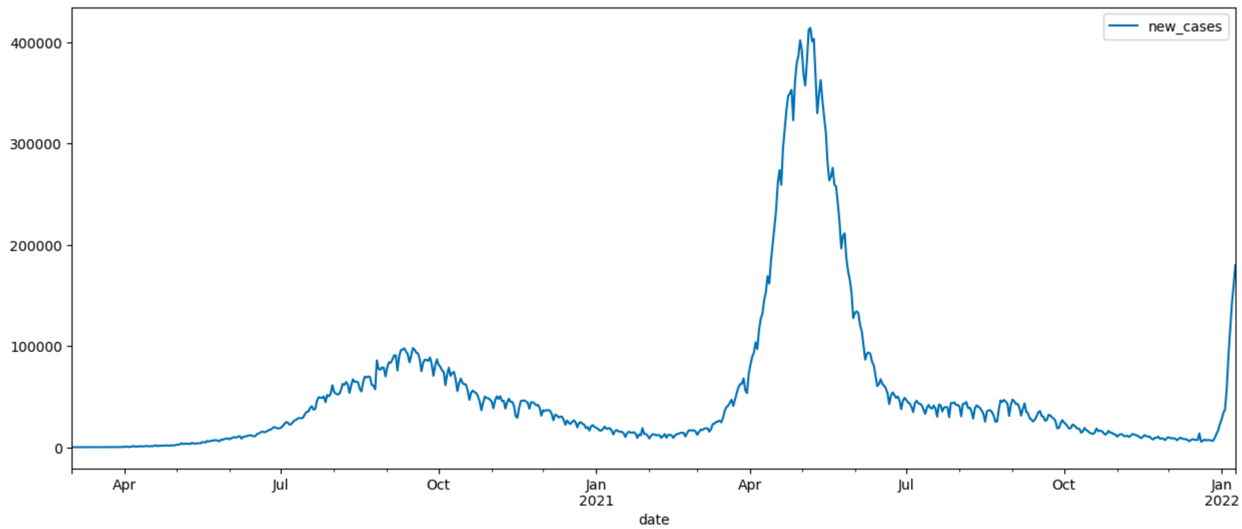
We can see for the first wave gave the following values

- beta(average contact rate of infection)=0.22795005
- gamma(average recovery rate)=0.01424044
- R0 (which suggests when the transmission rate is more than the recovery, R0 > 1, leads to an outbreak.):16.00723503

Lockdown Period (25th March-4th May)

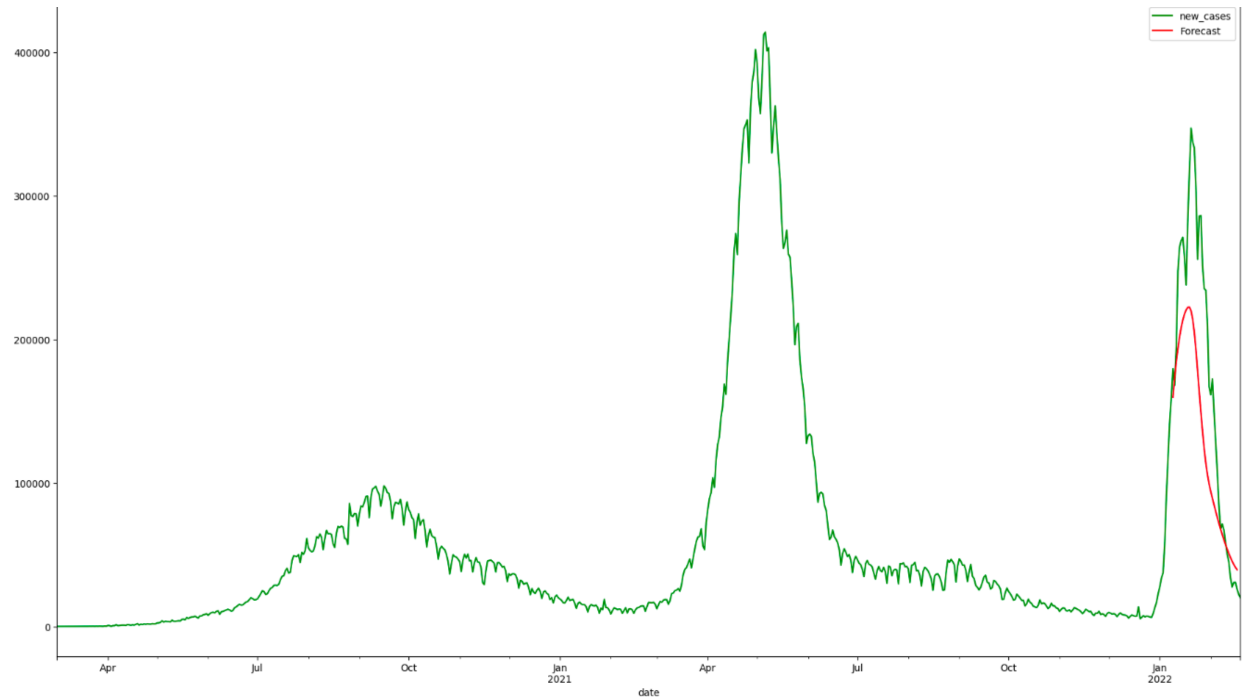
Predicting COVID Waves using LSTM artificial recurrent neural network architecture:

Plot of new cases <2021-2022>:



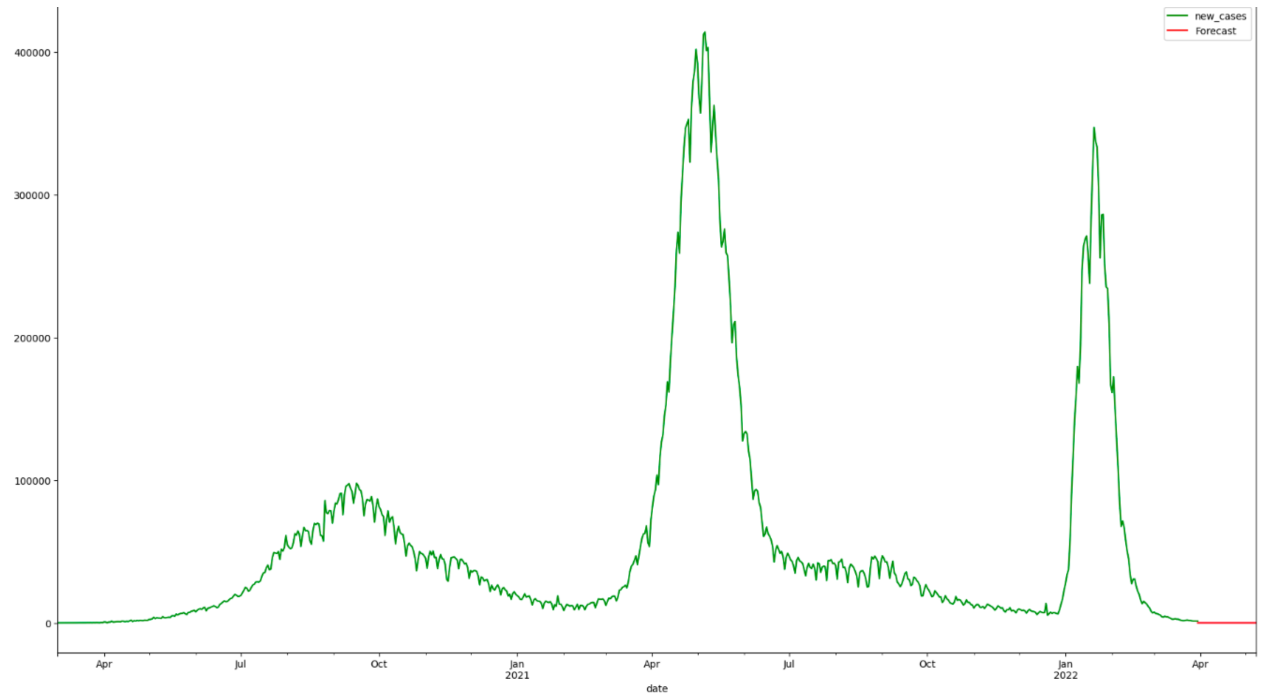
I am utilizing the most recent 50 days of scaled training data, spanning from 22-11-2021 to 10-1-2022, to forecast the number of new cases for the following day. The prediction involves using the information from the last 50 days to estimate the number of cases on 11-1-2022. Subsequently, I will exclude the data from 22-11-2021 and incorporate the predicted data from 11-1-2022 into the dataset.

The subsequent section displays the new cases forecasted by my LSTM model:

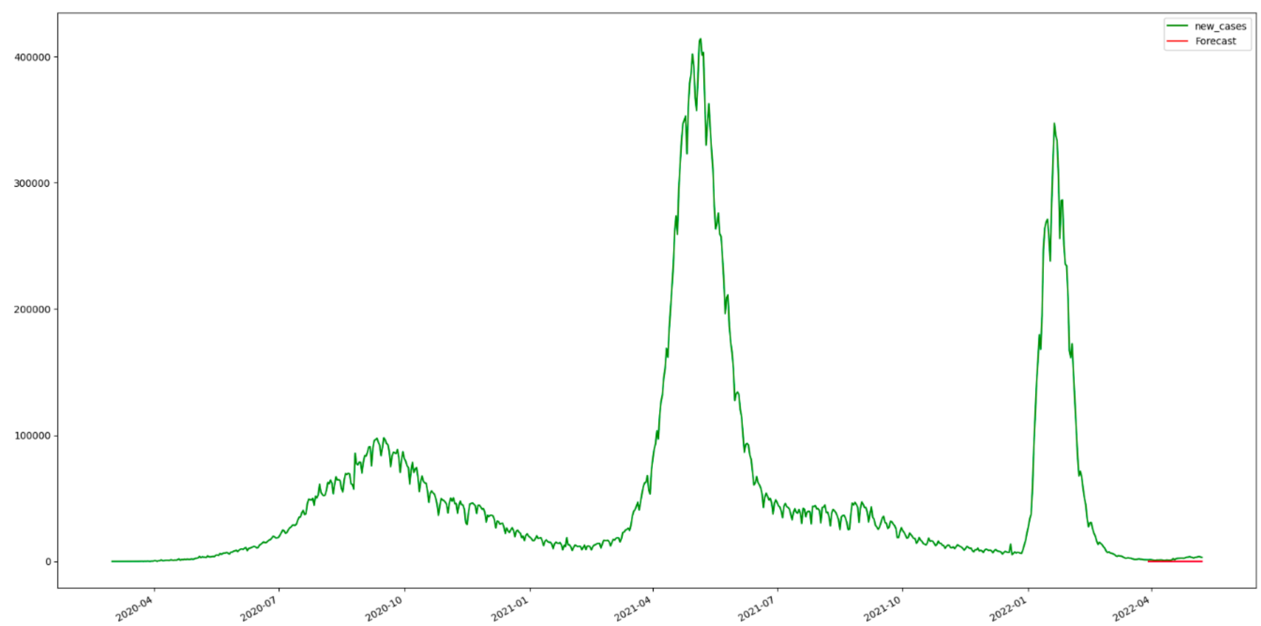


While my LSTM model effectively captures the overall trend (as depicted by the red line), it deviates significantly from the actual values. Despite experimenting with various parameters and methodologies for optimization, I believe the root issue lies in the dataset itself. The limited data available may hinder the model's ability to accurately predict values.

Now we will take cases till 30th March and we will predict the cases in the Month of April. The dataset which we use to train the model previous was upto 10th jan.



actual data:



As you can see my model is perfectly predicting the trend but the error lies in predicting the maxmina, minima and intermediate values.