HSO201A

Research Paper

Predicting the number of covid cases in Indian States using Machine Learning

Instructor: S. K. Mathur

GROUP INFORMATION:

Name - Roll No.	**Name - Roll No.**
Akashdeep Bhateja - 190085 (Leader)	Ajit Meena - 190076
Prakarsh Agrawal - 190615	Pratyush Baiplawat - 190640
Chandra Shekhar - 190246	Akash Kar - 190080
Kuldeep Meena – 190442	Poluri Bhakti Reddy - 190604
Harsh Mishra - 190362	Dhananjay Gupta - 190280
Vijay Fagaria - 190961	

ABSTRACT:

It has been a year and a half now since the covid was first introduced to the world. In India, this covid pandemic has become worse now. Every household in India is now under the effect of this calamitous situation. Amidst all the havoc, many studies have been accomplished to track the disease, its growth prediction, and possible policies to address the severity faced by the people. Among these, most of the studies have deployed machine learning and data analysis to produce effective and accurate results. However, most of this prior research focused on the number of infected people in the entire country. In this study, we focus on the number of infected people in each Indian state (with sufficient data for prediction) and predict the number of infected people for that state in the next 60-90 days. We will try distributions like Weibull, Gaussian, Log-normal, Log-logistics to decide the best fit for our data. We have also used methods like Levenberg-Marquardt (LM) for curve fitting. Hopefully, such research would promote proactive and confident responses from both citizens and government in this catastrophic situation.

INTRODUCTION:

BACKGROUND:

The whole world is currently facing an unprecedented crisis due to novel coronavirus. People are witnessing the reality of the line "*These are hard times"* for more than a year now. According to the World Health Organisation (WHO), Coronavirus circulate in some wild animals and have the capability to transmit from animals to humans. Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are

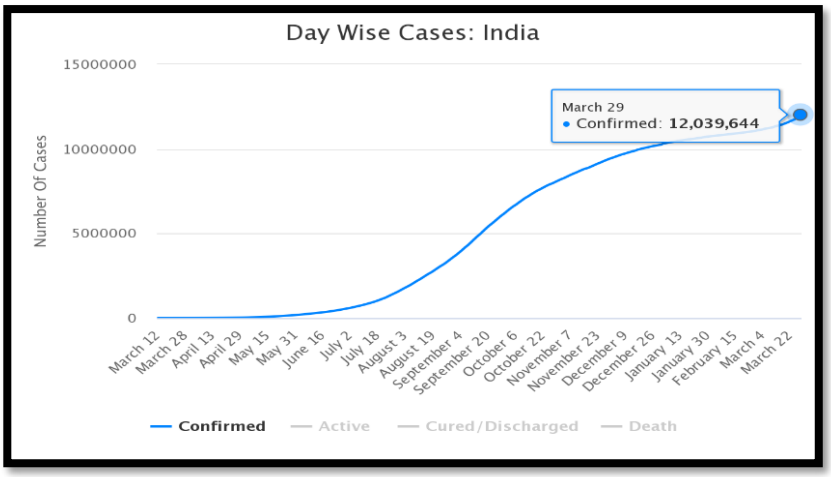
more likely to develop serious illness. Maybe this is appearing as the most dangerous thing about this virus, but the main concern lies in its exponential rate or way of spreading.
The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. It also spreads through social contact. It means that, if one person in a group gets affected, the whole group may get infected in no time. Till date, there are no specific treatments for coronavirus to date. However, one can avoid infection by maintaining basic personal hygiene and social distancing from infected persons. On 31 December 2019, the first reported case in the COVID-19 outbreak was reported in Wuhan, China. The first case outside of China was reported in Thailand on 13 January 2020. The cumulative incidence of the causative virus (SARS-CoV-2) is rapidly increasing and has affected 196 countries and territories with USA, Spain, Italy, U.K. and France being the most affected. WHO declares COVID-19 outbreak as a Public Health Emergency of International Concern (PHEIC) by WHO on 30 January 2020 and declared Coronavirus disease 2019 (COVID-19) as a global pandemic on 11 March 2020?
In India, the disease was first detected on 30 January 2020 in Kerala in a student who returned from Wuhan. The total (cumulative) number of confirmed infected people is 12.1 million till now (30 March 2021) across India out of which 11.4 million have been recovered and around 162K lives are lost.

IMPACTS OF COVID:

The COVID-19 pandemic has led to a dramatic loss of human life worldwide and presents an unprecedented challenge to public health, food systems and the world of work. Millions of enterprises face an existential threat. Nearly half of the world's 3.3 billion global workforce are at risk of losing their livelihoods. Informal economy workers are particularly vulnerable because the majority lack social protection and access to quality health care and have lost access to productive assets. Also, due to some long lockdown stages all over the world, the effect of price hike cannot be ignored. Also, the education system is fully revolutionised. Online education has become a necessity to secure the lives of students. At the same time, we must not forget its impact on people's minds. Such a pandemic creates a huge panic among masses. They fear to go anywhere, to meet someone, to celebrate some occasion and so on. Many rumours also spread at such times, that can mould people's opinion.

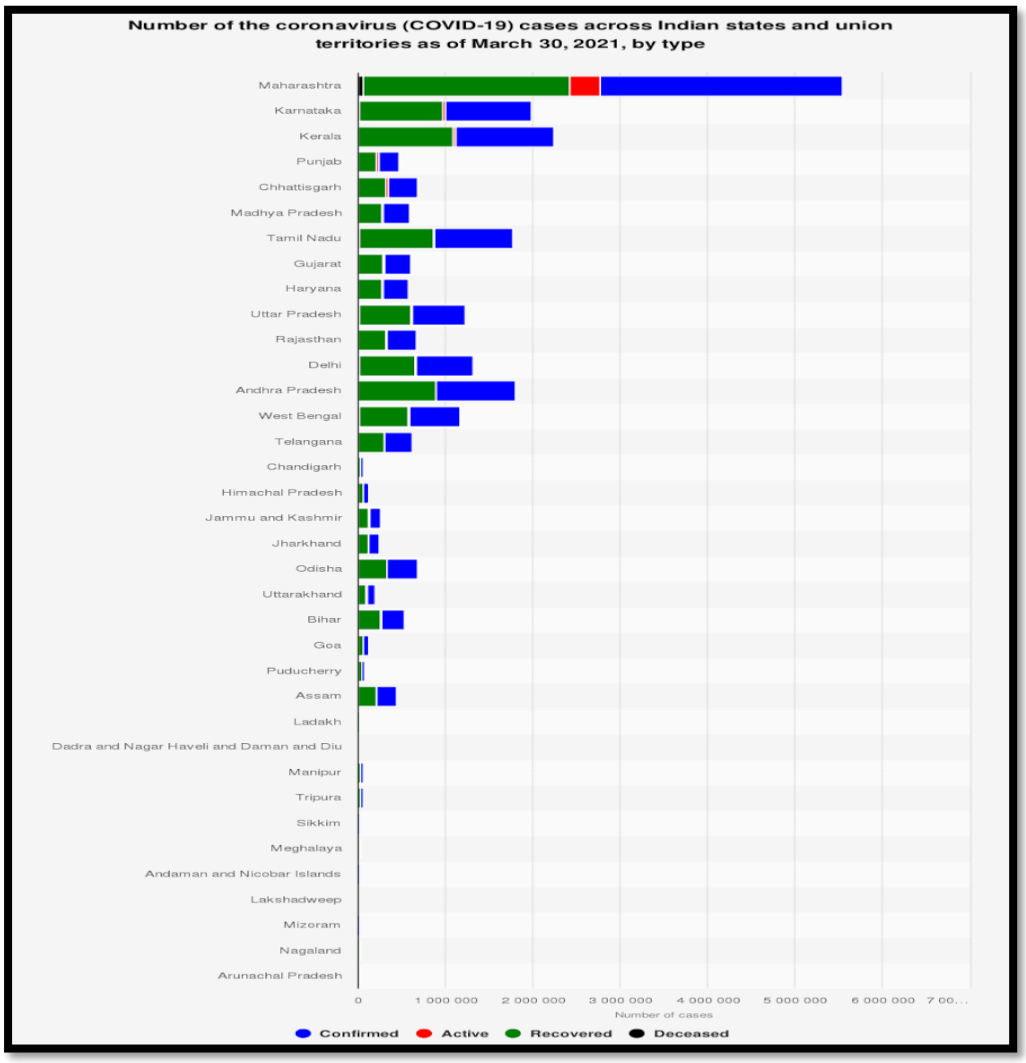
MOTIVATION:

The world requires efficient methodologies and research to confront this chaotic situation. Deep learning presents itself as one such option today. For example, ML and AI are used to augment the diagnosis and screening process of the identified patient with radio imaging technology akin to Computed Tomography (CT), X-Ray, and Clinical blood sample data. The list of such examples is endless. Wrapping covid datasets with Machine Learning (ML) and Artificial Intelligence (AI), researchers can forecast where and when the disease is likely to spread and notify those regions to match the required arrangements. Our main purpose is to use these tools in the study of covid pandemic situations in India.



This line graph depicts the increase of corona cases in India. It may appear here that cases were negligible for initial months of spread. But, if we shorten our scale, then the picture will become clearer. Cases began increasing more rapidly during July 2020 and since then, it has only gone up.

The second graph below shows the number of cases on March 30, 2021 for various states and union territories. Bar graph is not uniform. Large differences between various states can



be clearly observed. A factor that can be associated with such differences is population and area. Maharashtra is much larger as compared to Manipur.
If we make predictions on an all-India basis, the model may not prove appropriate for various regions because distribution has no uniformity. Also, covid growth in various regions depends upon the policies developed by local authorities and how strictly people in that region follow them. For example, Maharashtra is now adopting Britain's lockdown scheme to control covid pandemic in the state. Therefore, from the point of view of usefulness, a model based on state will give more productive results.

OBJECTIVES AND PROBLEM STATEMENT DESCRIPTION:
The main **objective** of our study is to explore the trend of covid 19 in Indian states and in some metro cities. Most of the prior research has focused on India as a single unit. However, extending the research to the state-level is important as the State governments have played a separate role in the control of Covid. Moreover, analysis at a smaller level will give more accuracy in implementing measures according to need. Through this study, we will try to answer the following questions:

1. Which states are under adverse conditions and which are under good control?
2. What is the expected number of cases after the end of November?
3. Which state will get rid of COVID-19 first (i.e., Number of cases < 1)?
4. When will COVID-19 vanish in India?
5. Which state has a chance to suffer from another wave and when?

Answering these can help authorities to know the scenario for lifting up the lockdown in a phased manner, especially in states which fall under the green zone.

BIASES OR ASSUMPTIONS:
The number of cases in any region or the pandemic situations depend upon factors like population density, area. More than these, it also depends upon the type of people living in that area. Do people strictly follow the protocols for safety from pandemic? At the same time, to combat such a situation, the mental health of people matters a lot which in turn depends upon various factors. Government policies and its executions play its own part. For example, the time of implementation of lockdown or other restrictions placed upon the activities of recreation(say). However, quantification of these indicators is quite difficult and require collection and interpretation of even larger dataset. For now, we are restricting our study based on the available dataset.

This research is limited to the first wave of corona in India due to the unavailability of a proper data set of current severity. However, it opens a new dimension of application of these tools with some more technical analysis for predicting results in case of future pandemics or some other havoc.

LITERATURE REVIEW:

INTRODUCTION:
COVID-19 is caused by SARS-Cov-2, a newly emergent coronavirus that was first recognized in Wuhan, China, in December 2019. SARS-CoV-2 is a new coronavirus closely related to SARS-CoV and genetically clusters within Beta coronavirus subgenus Sarbecovirus. The first

whole-genome sequence was published on January 5, 2020, and thousands of genomes have been sequenced since this date. Till now Over 57 000 genome sequences have been deposited in the GISAID EpiCoV database. A meta-analysis of different time estimates to the virus's last common ancestor indicates that the pandemic could have started between October 6 and December 11, 2019 [ii]. (SAGE, 2020) There is a need for innovative solutions to develop, manage, and analyse big data on infected subjects' growing network, patient details, their community movements and integrate with clinical trials and pharmaceutical, genomics, and public health data.
The Machine Learning (ML) and Data science community are striving hard to improve the forecasts of epidemiological models and analyse the information flowing over Twitter to develop management strategies and the assessment of the impact of policies to curb its spread. Various datasets in this regard have been openly released to the public. However, there is a need to capture, develop and analyse more data as the COVID-19 grows worldwide. [i]

BROAD:
The base paper analyses and predicts the growth of COVID – 19 worldwide; however, this paper focuses on using linear regression models for COVID -19 prediction on a state-level. These models have already been used to predict epidemics like COVID-19 worldwide, including China, Ebola outbreak in Bomi, Liberia (2014).
We have used Indian COVID-19 data available publicly. There are a few works that are based explicitly on Indian COVID-19 data. Das [iii] has used the epidemiological model to estimate the primary reproduction number at national and some state levels. Ray et al. [iv] used a predictive model for case-counts in India. They also discussed hypothetical interventions with various intensities and provided projections over a time horizon.
Both the articles have used the SIR (susceptible infected-removed) model for their analysis and prediction. As we discussed earlier, considering the great diversity in every aspect of India and its vast population, it would be a much better idea to look at each of the states individually. The purpose of the SIS model is to reflect the effect of the major preventive measures like the nationwide 21-day lockdown from March 25 to April 14, 2020. The SIS model is critically dependent on the infection-rate parameter (β). It is defined as the number of people infected per unit time from an infected person. Note that this parameter is subject to change due to the effect of lockdown and other preventive measures to ensure social distancing. When people are at home, the infection rate is expected to be on the lower side.
The study of each of the states individually would help decide further actions to contain the disease's spread, which can be crucial for the specific states only.

DETAILED:
The general framework observed from the study by Davies et al. is compartmental modelling. In this, individuals are categorised into different categories according to their infection or symptom status. The prototypical compartmental model is the Susceptible-Infectious-Removed (SIR) model, the parameter β sets the infection rate, and the average infectious period is 1/μ days. The basic reproduction number, $R0=\beta N/\mu$, represents the expected number of individuals that a single infectious host will infect if introduced into a population of N susceptible hosts

$$\frac{dS}{dt} = \beta SI; \qquad \frac{dI}{dt} = \beta SI - \mu I; \qquad \frac{dR}{dt} = \mu I;$$

In this article, the author mainly focused on the SIS model and the logistic and the exponential models at each state (restricting to only those states with enough data for prediction). The SIS model considers the possibility that an infected individual can return to the susceptible class on recovery because the disease confers no long-standing immunity against reinfection. WHO is aware of these reports of patients who were first tested negative for COVID-19 using PCR (polymerase chain reaction) testing and then after some days tested positive again?

CONCLUSION:

In case of Developing countries like India (which also possesses a great bio-diversity), there have been a very rapid and uncontrollable growth in terms of number of cases on daily basis. Due to rapid transmission, many states in India should increase their attention and approach toward handling of COVID-19 positive scenario. They need to improve their response operations including establishments of real time rapid response team.

METHODS & METHODOLOGIES:

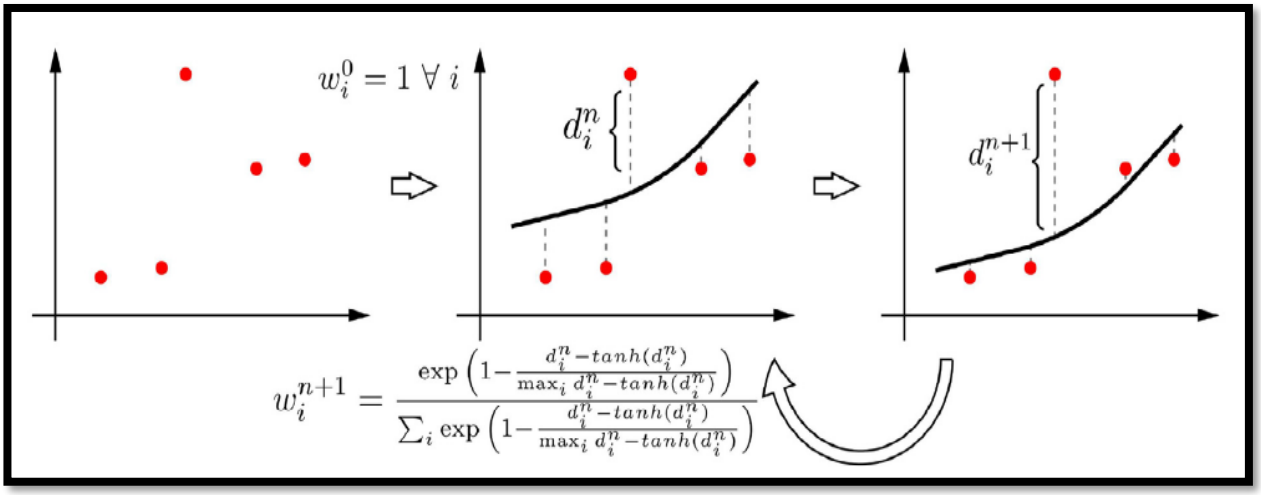
Here, we used mathematical modelling to predict the trend of patient diagnosis in Indian states. According to all diagnosis numbers from the WHO website and combining with the transmission mode of infectious diseases, the mathematical model was fitted to predict the future trends of outbreaks. We have also used methods like **Levenberg-Marquardt (LM)** for curve fitting.

The **Levenberg - Marquardt (LM)** principle offers great flexibility and a wide range of applications due to many reasons. This technique is particularly effective in solving systems of nonlinear equations, and this makes it useful for applications involving semiconductor devices that have both large sets of non-linear equations and large parameter sets. Levenberg-Marquardt Optimization is a virtual standard in nonlinear optimization which significantly outperforms gradient descent and conjugate gradient methods for medium sized problems. It is a pseudo-second order method which means that it works with only function evaluations and gradient information but it estimates the Hessian matrix using the sum of outer products of the gradients.

Levenberg - Marquardt(LM) can be extended to get an approximate fit of various distributions and finding the best-fit parameters corresponding to them. One of the density functions proposed are

$$f(x) = k.\gamma.\beta.\alpha^{\beta}.x^{-(\beta+1)}exp\left[-\gamma\left(\frac{\alpha}{x}\right)^{\beta}\right], \ x, \alpha, \beta, \gamma > 0$$

Here, this function $f(x)$ denotes the no of cases, with respect to the variable x, which represents the days from the first case reported. The parameters of the model are α, β, γ > 0, ∈ R. Now, we are going to use the weighting strategy because initial stage data of covid have various data points that are different from the other values. And We will be calling this fitting technique *"Robust Fitting".*



THE PROPOSED ALGORITHM FOR CURVE-FITTING:

The Following algorithm is proposed in the paper, which is used for curve fitting. We have written same logic for the code in python, which uses all the inbuilt functions from the **Numpy** package. The threshold value was chosen to be: 0.001

Requirements: $\boldsymbol{x}$ (# days), y (actual number of cases), $\boldsymbol{\epsilon}$ (Threshold)

Procedure:

W0 = Unit Vector [1] x Sizeof($\boldsymbol{x}$)

$for\ iterations\ (0\ to\ n)$:

$$f = LM\ (input = x,\ target = y,\ weights = w^n$$

$$d_i = |f(x_i - y_i| \; \forall \; i$$

$$w_i^{n+1} = \frac{\exp\left(1 - \frac{d_i^n - \tanh(d_i^n)}{max_i d_i^n - \tanh(d_i^n)}\right)}{\sum_i \exp\left(1 - \frac{d_i^n - \tanh(d_i^n)}{max_i d_i^n - \tanh(d_i^n)}\right)}$$

If (old_weight - new_weight) < $\boldsymbol{\epsilon}$: **break for**

end:

DATA SET REVIEW:

The dataset we are using consists of data from the whole nation and its respective states. The datasets which we are using can be found on the GitHub link here. The names of the columns, as well as their description, for the major datasets are as follows:

- **Date:** representing the time on which the data had recorded

- **State:** representing the data of that particular state only.
- **Confirmed:** representing the total number of covid cases on that specific date in that state.
- **Recovered:** representing the total recovered cases on that particular date in that state.
- **Deceased:** representing the total deaths on that specific date in that state.
- **Tested:** representing the total persons detected on that specific date in that state.

Here are few of the rows and columns of our dataset:

Date	State	Confirmed	Recovered	Deceased	Other	Tested
2021-01-16	India	10558715	10196223	152311	4370	186544868.0
2021-01-17	India	10572677	10210736	152456	4377	187093036.0
2021-01-18	India	10582664	10227863	152593	4387	187802827.0
2021-01-19	India	10596451	10245092	152755	4402	188566947.0
2021-01-20	India	10611730	10265163	152907	4413	189347782.0
2021-01-21	India	10626225	10282897	153068	4422	190148024.0
2021-01-22	India	10640548	10300063	153221	4435	190985119.0
2021-01-23	India	10655444	10316096	153377	4443	191766871.0
2021-01-24	India	10668676	10329244	153508	4449	192337117.0
2021-01-25	India	10677774	10345336	153624	4462	193062694.0
2021-01-26	India	10690507	10358586	153762	4469	193613120.0
2021-01-27	India	10702063	10372847	153885	4478	194338773.0
2021-01-28	India	10720975	10393162	154047	4484	195081079.0

The dataset which we have used for analysing the vaccine part is shown below. It comprises of the total number of vaccines (CoviShield or Covain) being administered to the people of the India.

- **Total Covaxin Administered** comprises of the total number of Covaxin being administered to the people of India.

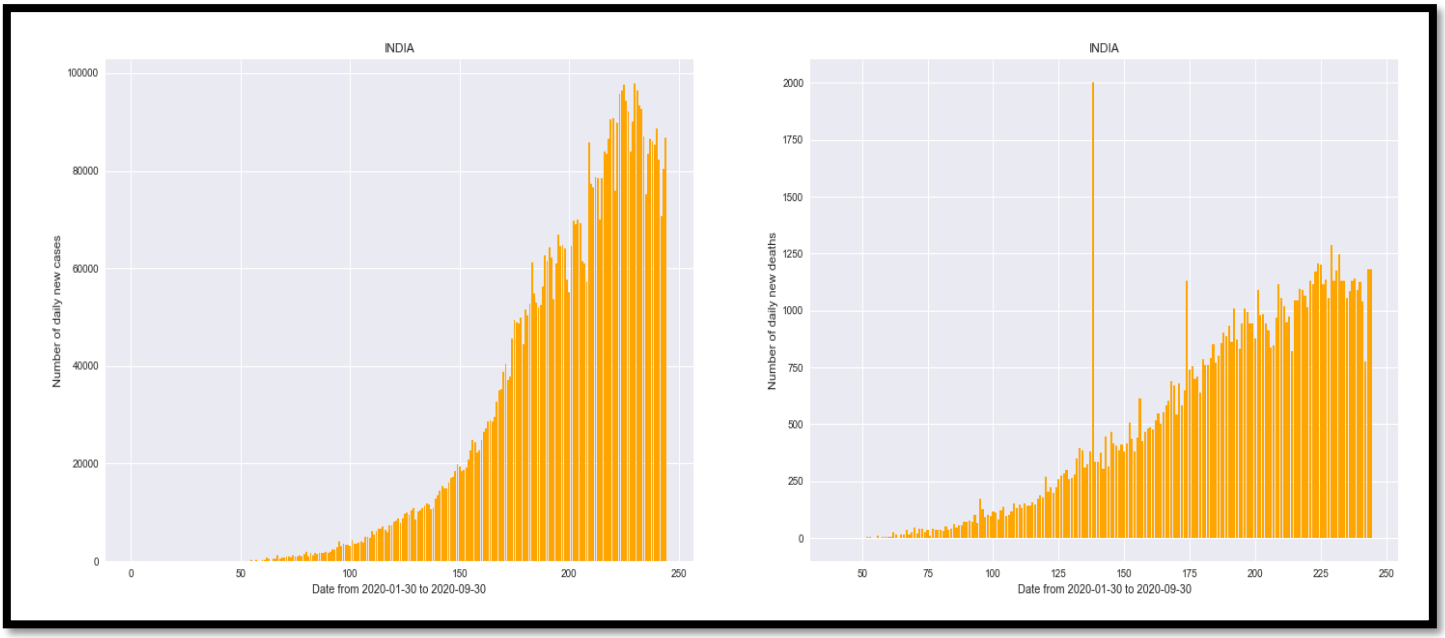
Total Doses Administered: comprises of both the CoviShield and Covaxin being administered to the people of India.

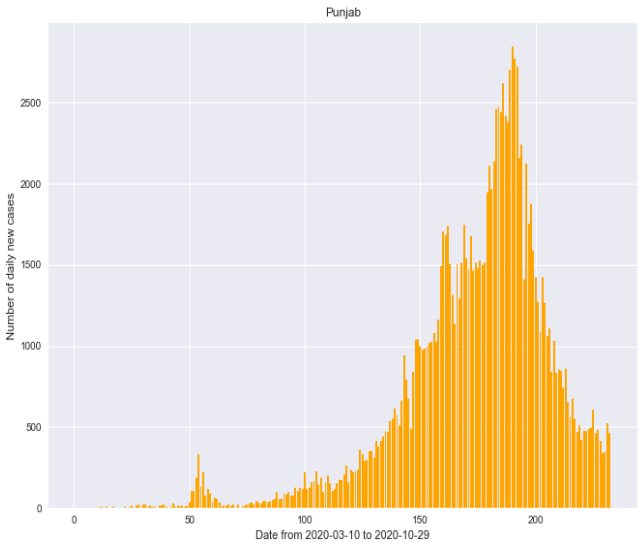
- **Total CoviShield Administered** comprises of the total number of CoviShield vaccine being administered to the people of India.

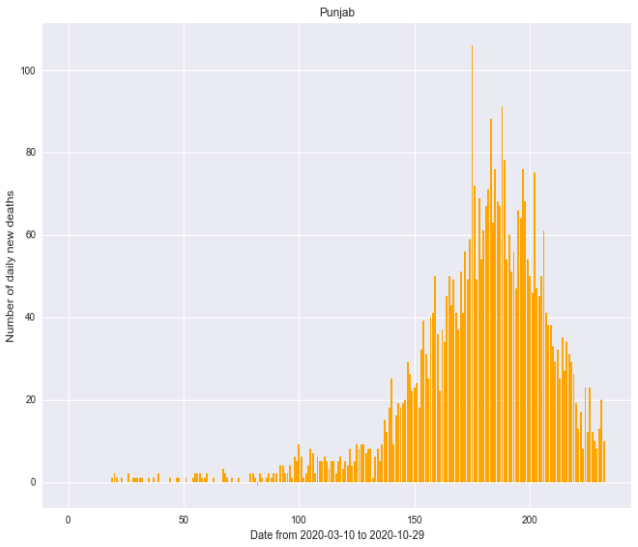
Updated On													
2021-01-16	Maharashtra	812649.0	179.0	174.0	5726.0	0.0	3668.0	2057.0	1.0	85.0	5641.0	5726.0	5726
2021-01-17	Maharashtra	814714.0	269.0	216.0	6521.0	0.0	3953.0	2566.0	2.0	94.0	6427.0	6521.0	6521
2021-01-18	Maharashtra	811257.0	772.0	320.0	6151.0	0.0	3569.0	2581.0	1.0	105.0	6046.0	6151.0	6151
2021-01-19	Maharashtra	850966.0	1196.0	340.0	13699.0	0.0	6328.0	7367.0	4.0	214.0	13485.0	13699.0	13699
2021-01-20	Maharashtra	870296.0	1547.0	347.0	23880.0	0.0	9658.0	14205.0	17.0	439.0	23441.0	23880.0	23880
2021-01-21	Maharashtra	880118.0	1820.0	357.0	24148.0	0.0	9784.0	14347.0	17.0	443.0	23705.0	24148.0	24148
2021-01-22	Maharashtra	902635.0	2012.0	374.0	44369.0	0.0	16546.0	27802.0	21.0	840.0	43529.0	44369.0	44369
2021-01-23	Maharashtra	921922.0	2435.0	483.0	69295.0	0.0	25333.0	43937.0	25.0	1156.0	68139.0	69295.0	69295
2021-01-24	Maharashtra	925794.0	2756.0	581.0	70196.0	0.0	25661.0	44510.0	25.0	1157.0	69039.0	70196.0	70196
2021-01-25	Maharashtra	950534.0	3170.0	617.0	107073.0	0.0	38512.0	68527.0	34.0	1433.0	105640.0	107073.0	107073

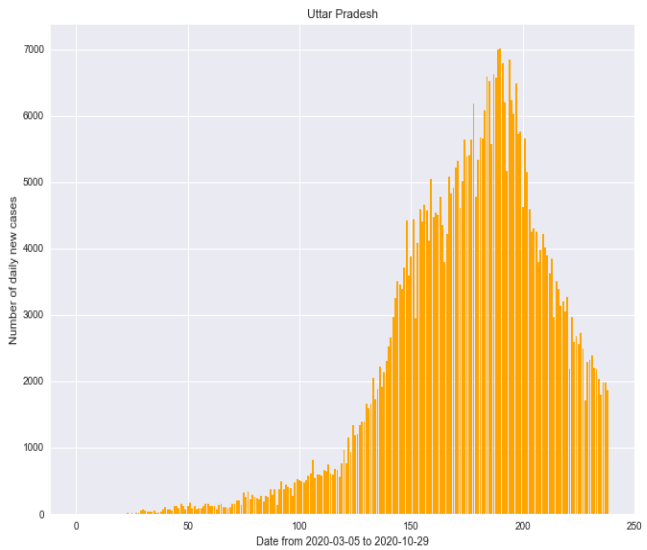
EXPLORATORY DATA ANALYSIS:

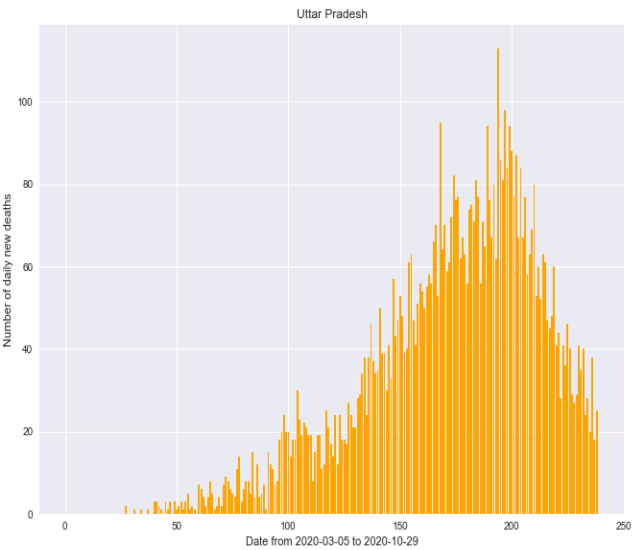
We have used python libraries such as **Seaborn**, **Matplotlib** for plotting and libraries such as **pandas** and **Numpy** were used for data manipulation. The Data is taken upto 30th October, 2020. The results for India and some states are shown below:

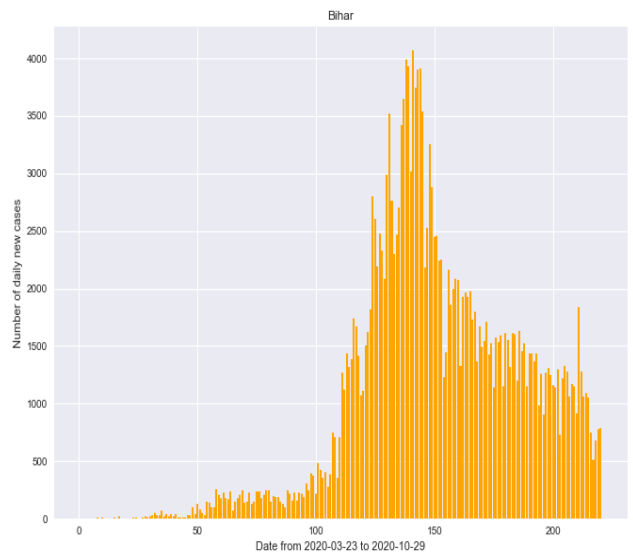


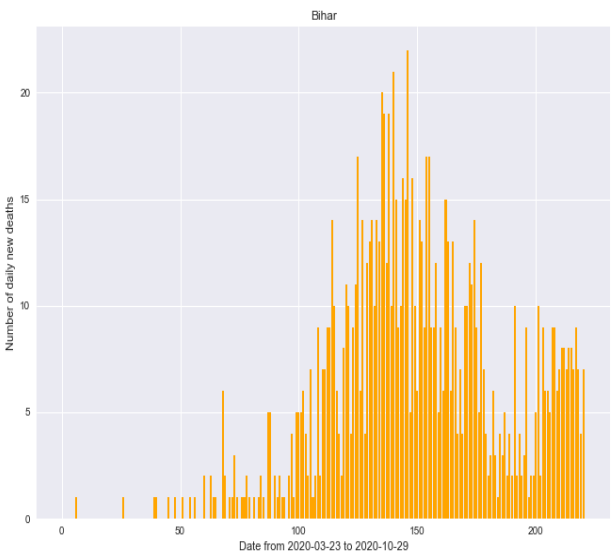

Punjab
Number of daily new cases
Date from 2020-03-10 to 2020-10-29

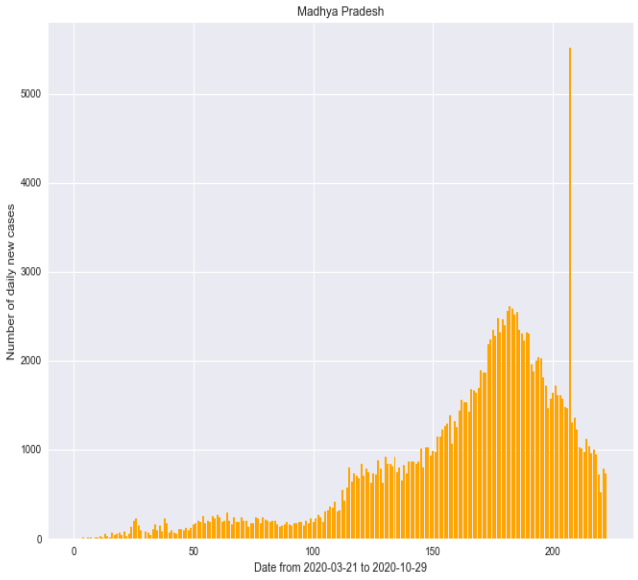

Punjab
Number of daily new deaths
Date from 2020-03-10 to 2020-10-29

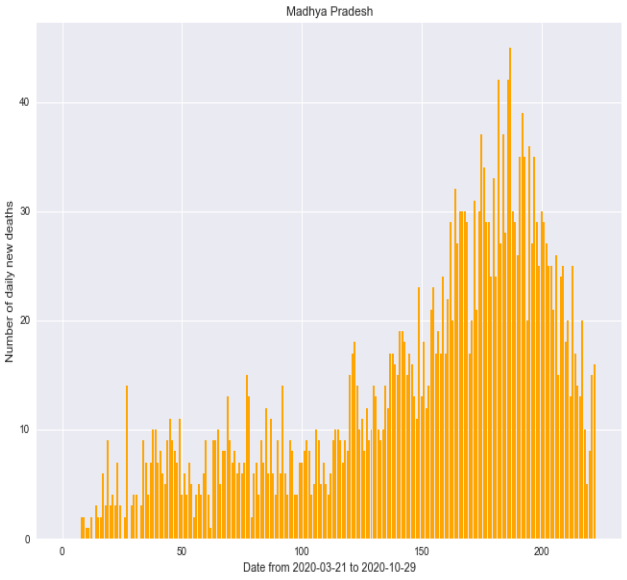

Uttar Pradesh
Number of daily new cases
Date from 2020-03-05 to 2020-10-29

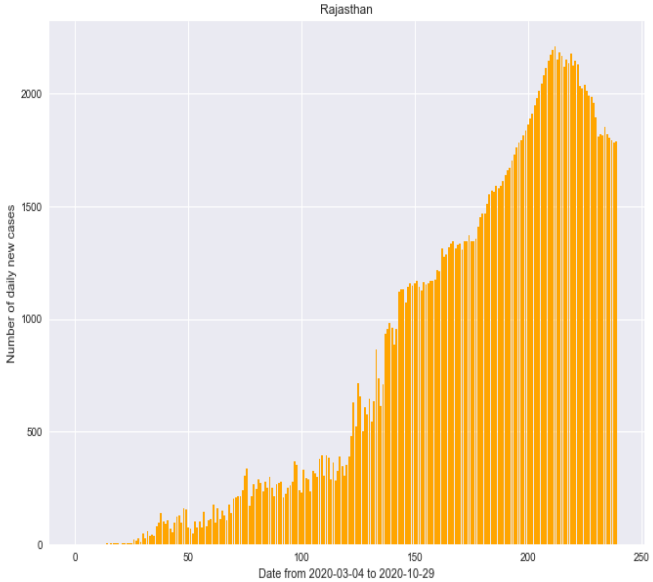

Uttar Pradesh
Number of daily new deaths
Date from 2020-03-05 to 2020-10-29

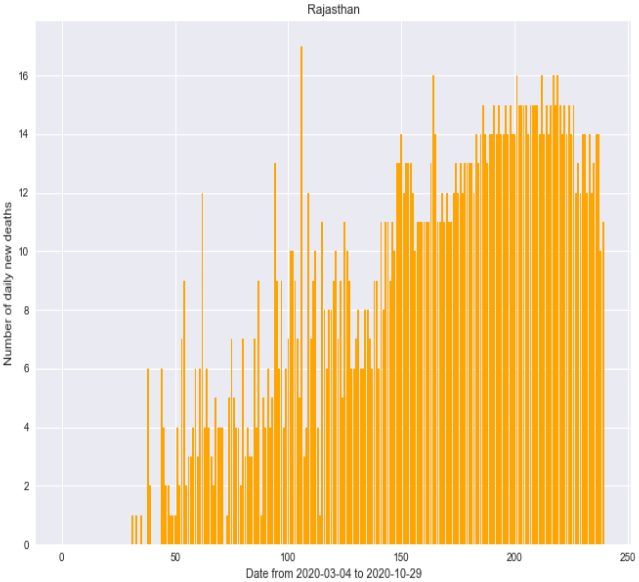

Bihar
Number of daily new cases
Date from 2020-03-23 to 2020-10-29

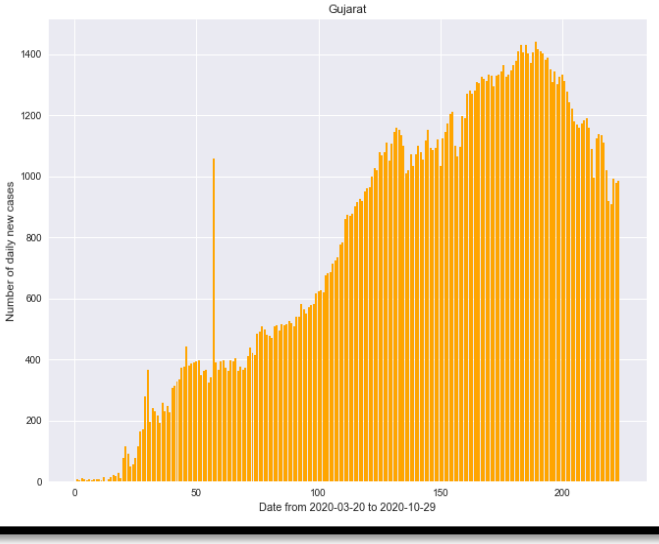

Bihar
Number of daily new deaths
Date from 2020-03-23 to 2020-10-29

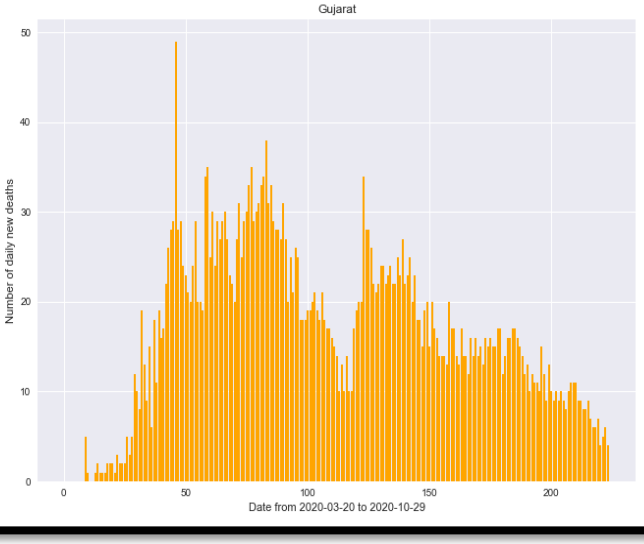

Madhya Pradesh
Number of daily new cases
Date from 2020-03-21 to 2020-10-29


Madhya Pradesh
Number of daily new deaths
Date from 2020-03-21 to 2020-10-29


Rajasthan
Number of daily new cases
Date from 2020-03-04 to 2020-10-29

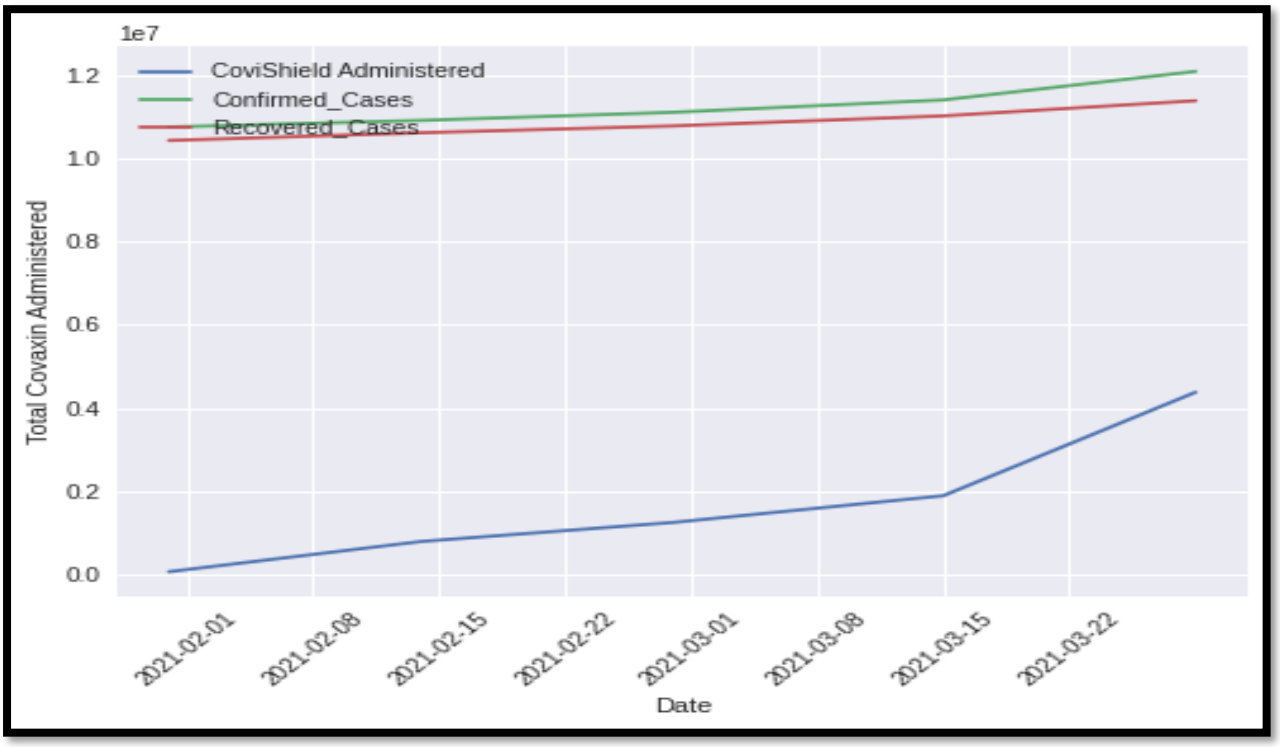

Rajasthan
Number of daily new deaths
Date from 2020-03-04 to 2020-10-29

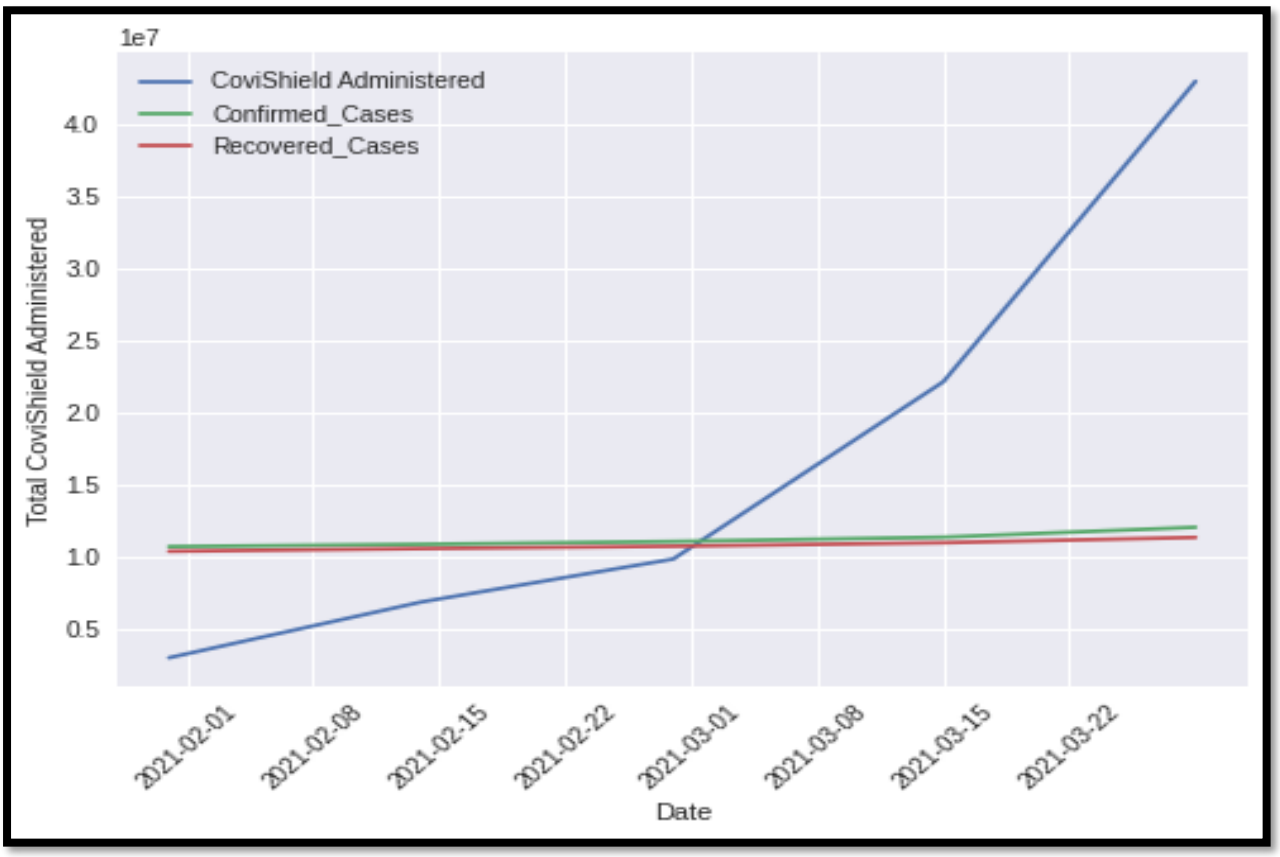

Gujarat
Number of daily new cases
Date from 2020-03-20 to 2020-10-29


Gujarat
Number of daily new deaths
Date from 2020-03-20 to 2020-10-29

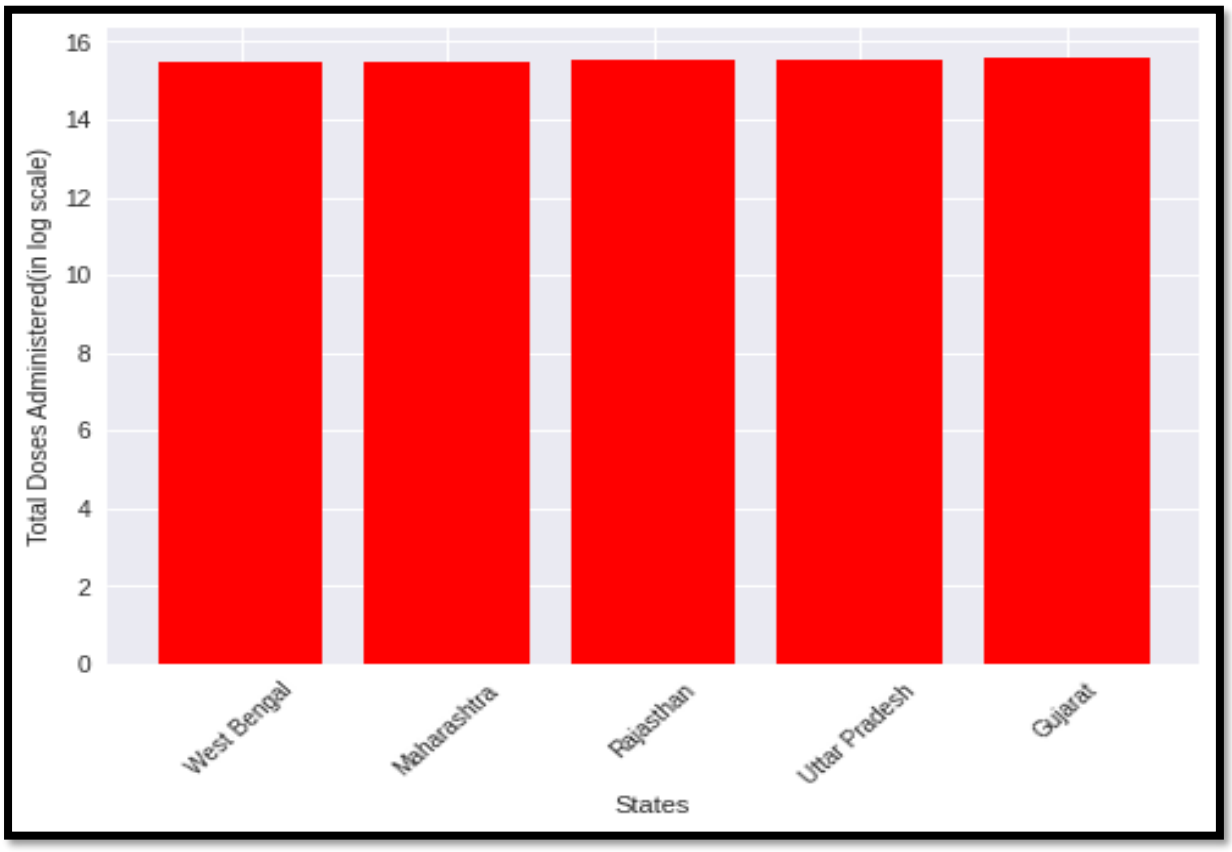
VACCINATION ANALYSIS:

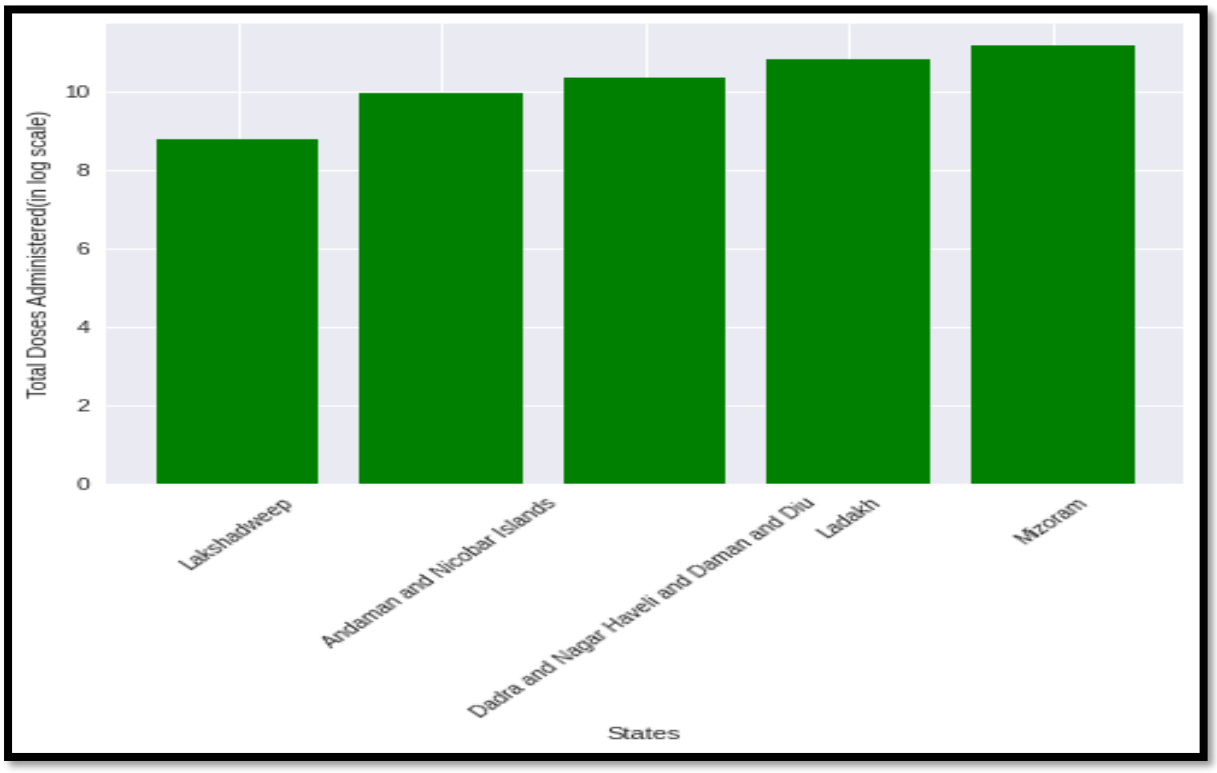
The following graphs show the trend of vaccinaton.



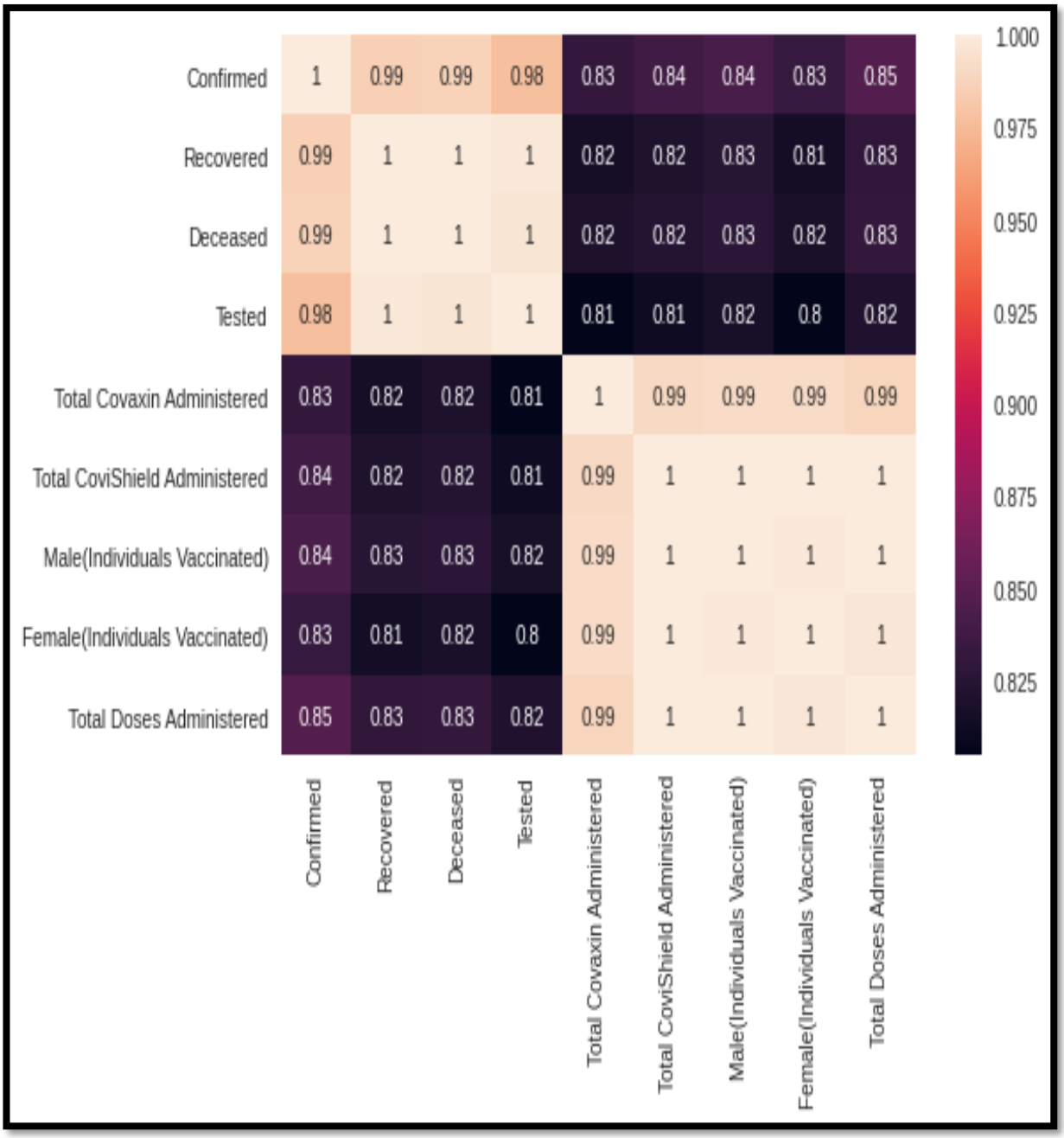


The following graphs shows the top and least vaccinated states:





The relation between the various parameters can be understood by calculating a correlation matrix that will explain how different variables are correlated among themselves and how the change in one variable will affect other variables.



A +1 value in a correlation matrix indicates that the other variable also grows up if one variable goes up. A -1 value in a correlation matrix shows that the other goes up if one variable goes down. 0 value indicates no correlation among the variables chosen.

DIFFERENT DISTRIBUTIONS:

In this paper, we are trying to plot the graph using these 4 Probability Distributions, and then decide the Which plot is fitting best to our data.

1. Gaussian Distribution:

The probability distribution function of a Gaussian distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{\frac{-(x - \mu)^2}{2\sigma^2}}$$

Where, μ = mean, σ = Standard Deviation

This distribution peak at its mean and It is also symmetric in nature. The shape of curve will be bell shaped, peak value depend on the values of the parameters such as scale (used in code).

2. Log-Logistic Distribution:

$$f(x) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^{\beta}\right)^2} \quad x,\ \alpha,\ \beta,\ \gamma > 0$$

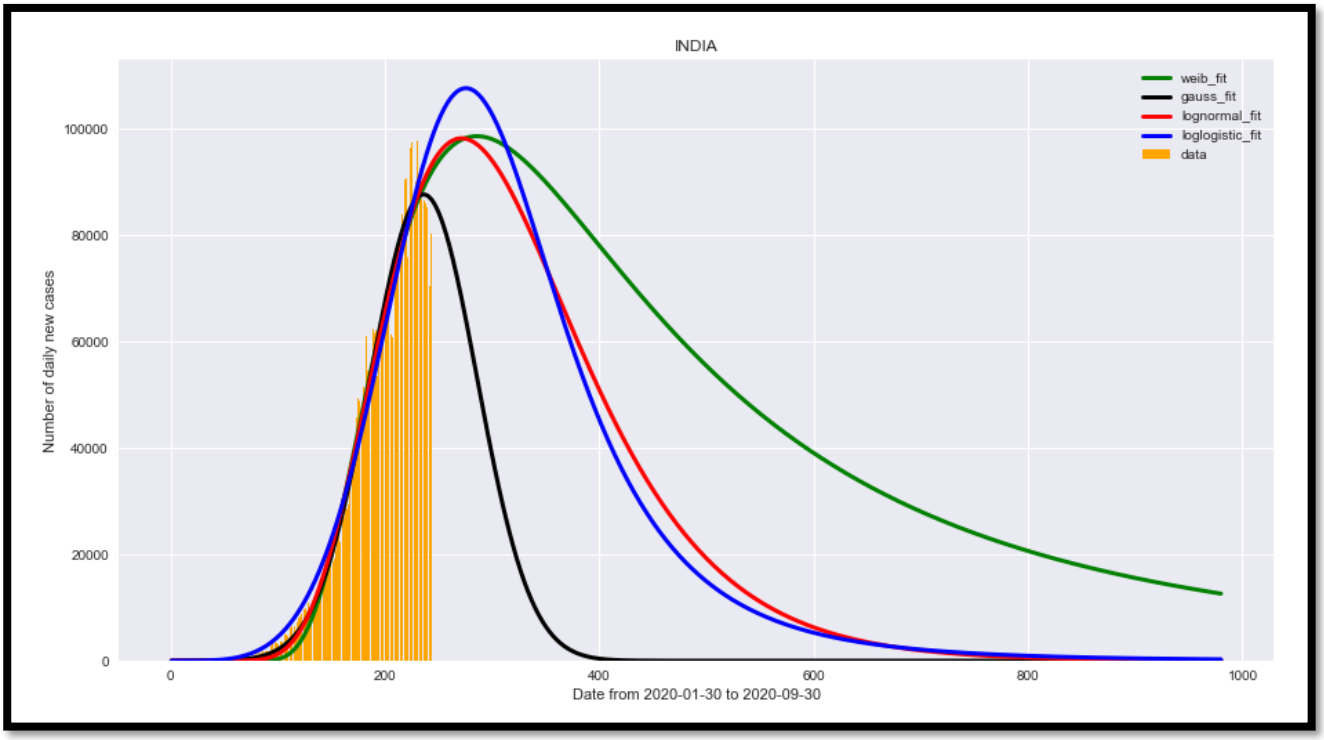
3. Log-Normal Distribution:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \; e^{\frac{-(ln(x) - \mu)^2}{2\sigma^2}}$$

4. Weibull Distribution:

$$f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} e^{(\frac{x}{\lambda})^k}$$

RESULTS AND CONCLUSIONS:



This graph has all the plots corresponding to all the distribution plots mentioned above. The characteristics of this curve are:

The data was taken till 30th October, 2020 for getting the optimum parameters corresponding to a distribution function.

The x-range of the plot is 979 days (which was a random choose), this day is signifying the date 4th October, 2022.

It can be observed that the curves such as Weibull, LogNormal and LogLogistic are giving more robust results here, which signifies the importance of the number of parameters and the characteristic they are associated with.

These fits are characterized by the optimized parameters and the best fit can be guessed by analysing the metrics such as MSE, MAE and R2 (R-Squared).

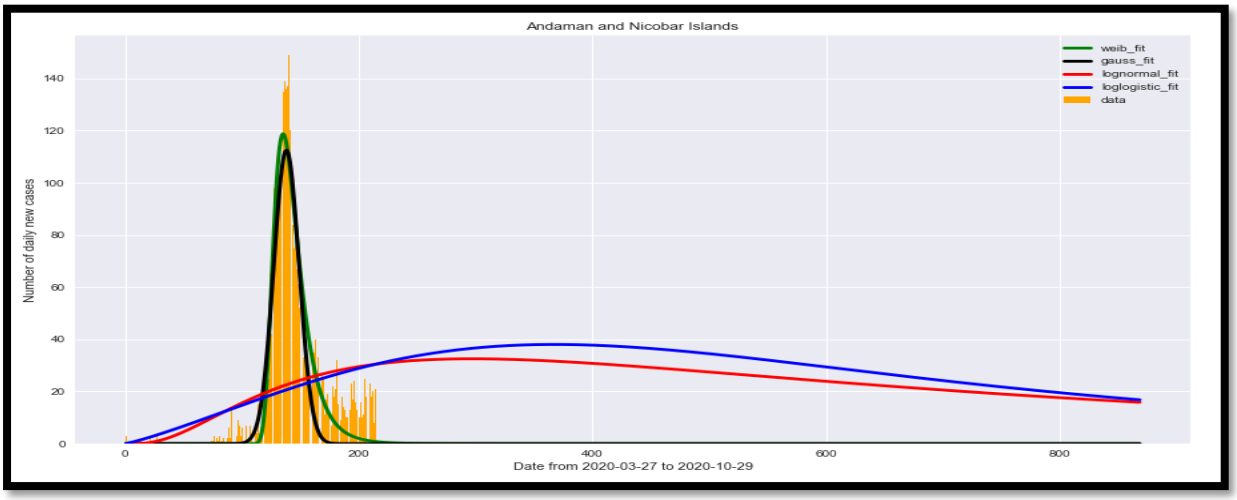
	Function	Mean Absolute Error	Mean Squared Error	R2-Score	Expected_Last_date	Expected_cases
0	Weibull	0.027411	0.001736	0.984982	2022-10-04	12627
1	Gauss	0.019964	0.001089	0.990098	2021-05-17	0
2	LogNormal	0.027270	0.001937	0.983180	2022-10-04	73
3	LogLogistic	0.035031	0.003485	0.969569	2022-10-04	271

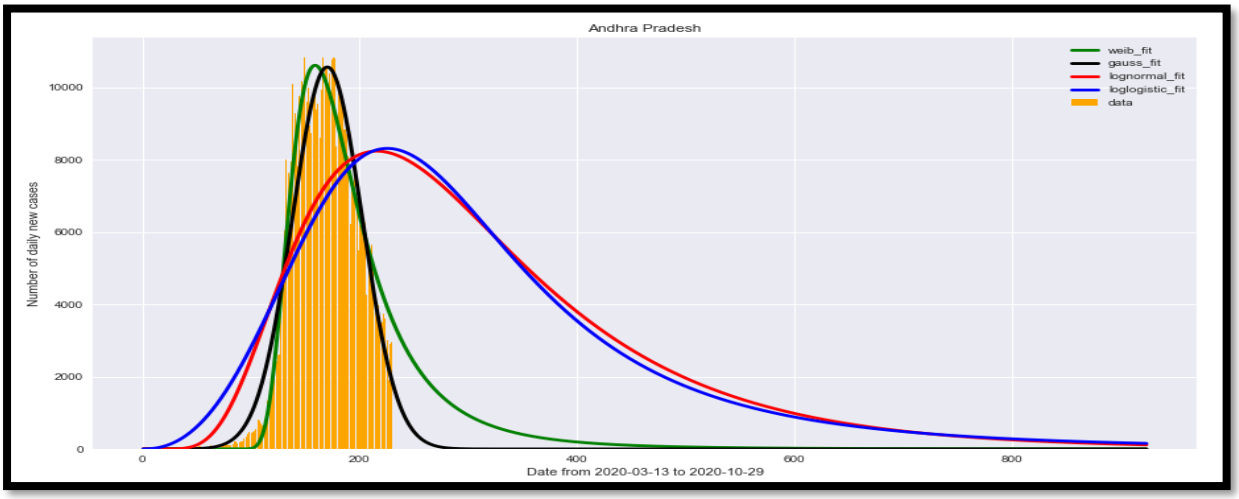
The Expected last date is calculated by creating a data-frame, which has all the values of predicted cases, i.e., corresponding to each function. We have values of x corresponding to them and thus querying a function named “expectedEnd,” we are successful in getting the expected end date and corresponding number of cases.

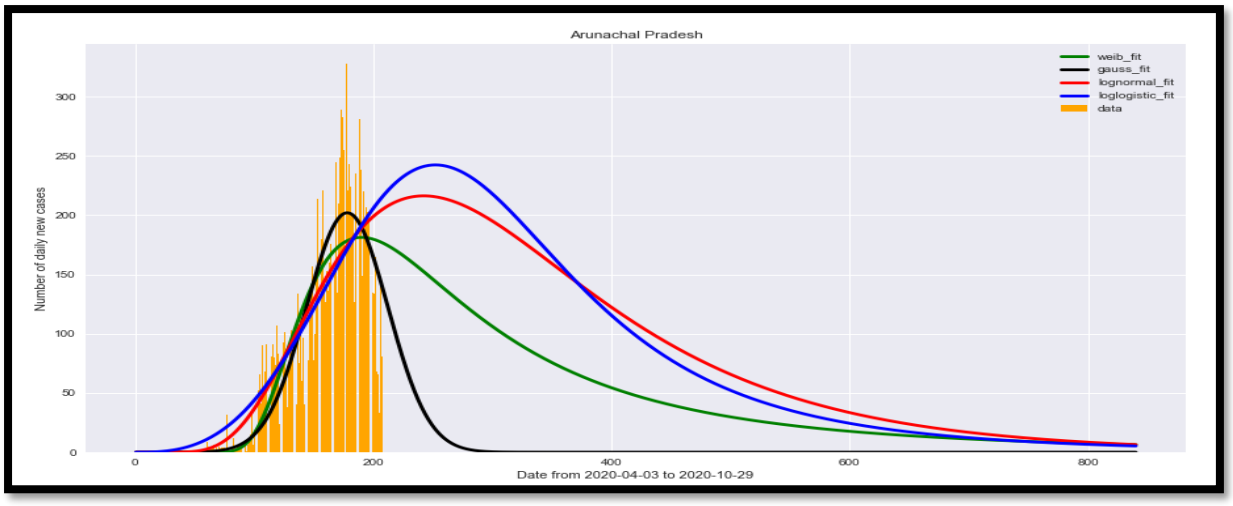
Please note that for the distributions where expected number of cases are not zero, that is due to the x-range in our plots. It can be easily calculated by changing the range of x (i.e., number of days).

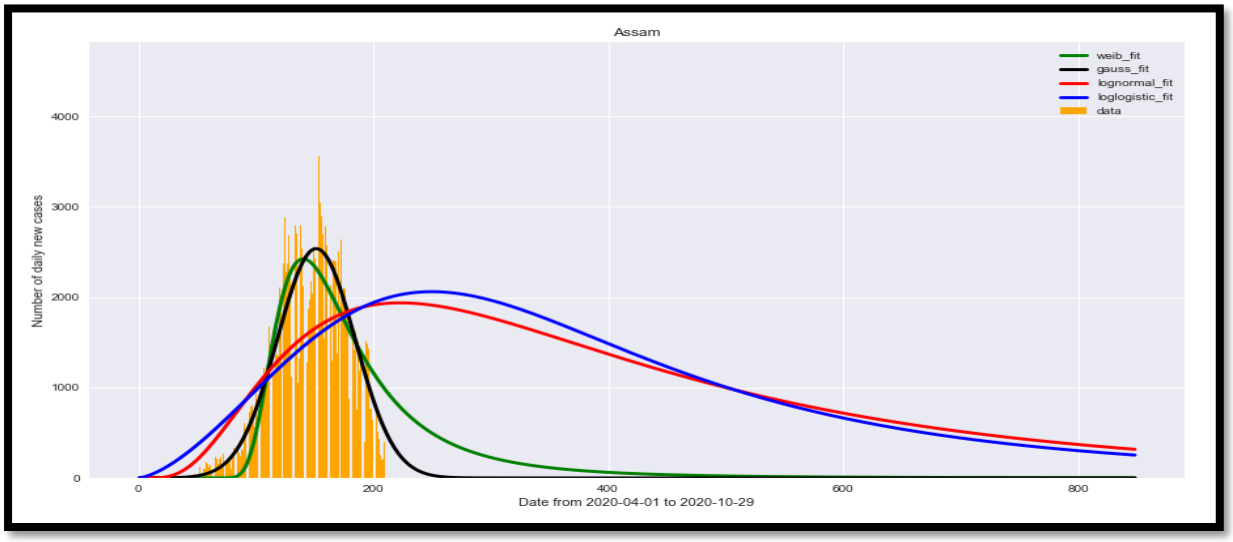
Hence looking on the metrics (MSE, MAE and R2-Score), It can be seen that **Gauss-Function** is the **best-fit** in the case of **INDIA.** But one should keep in mind that **overfitting and data limitations** (discussed in biases and assumptions section) can be the cases here, so a second better fit, here **LogNormal,** can also be considered as more general plot.

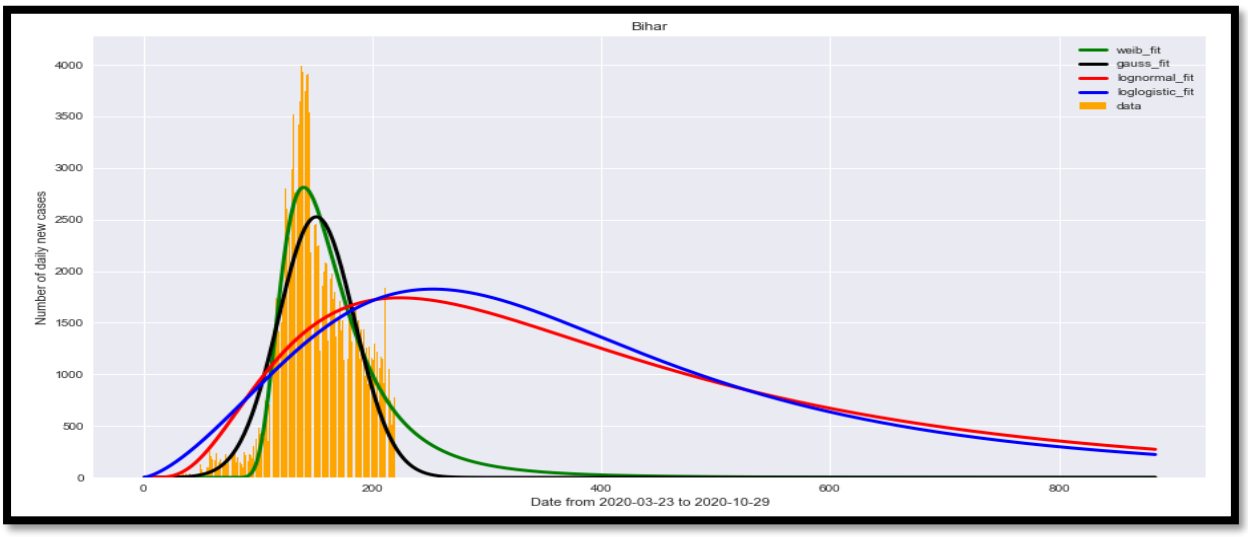
The Plots with respect to all the states are presented below:

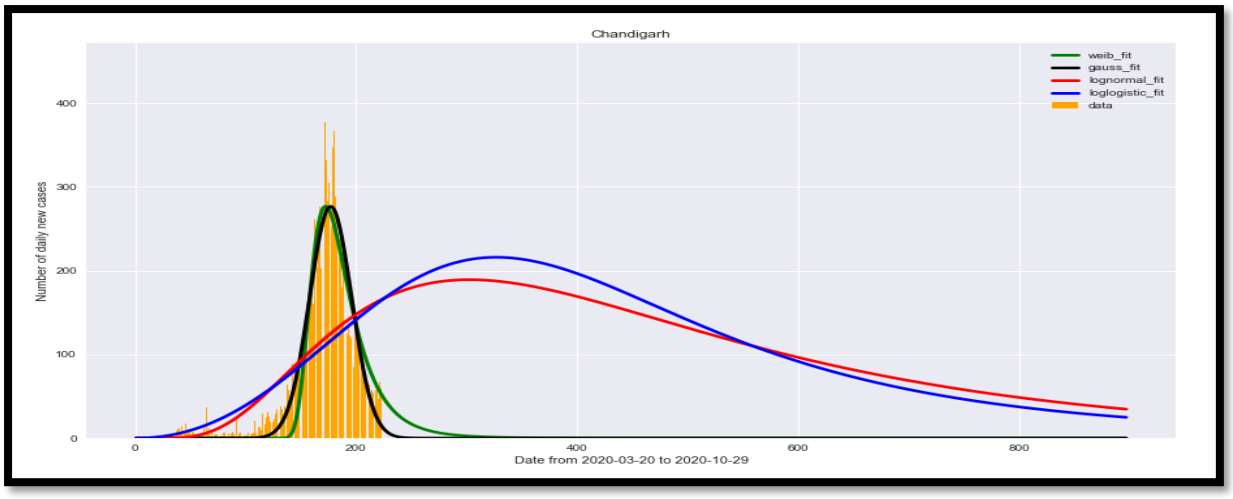


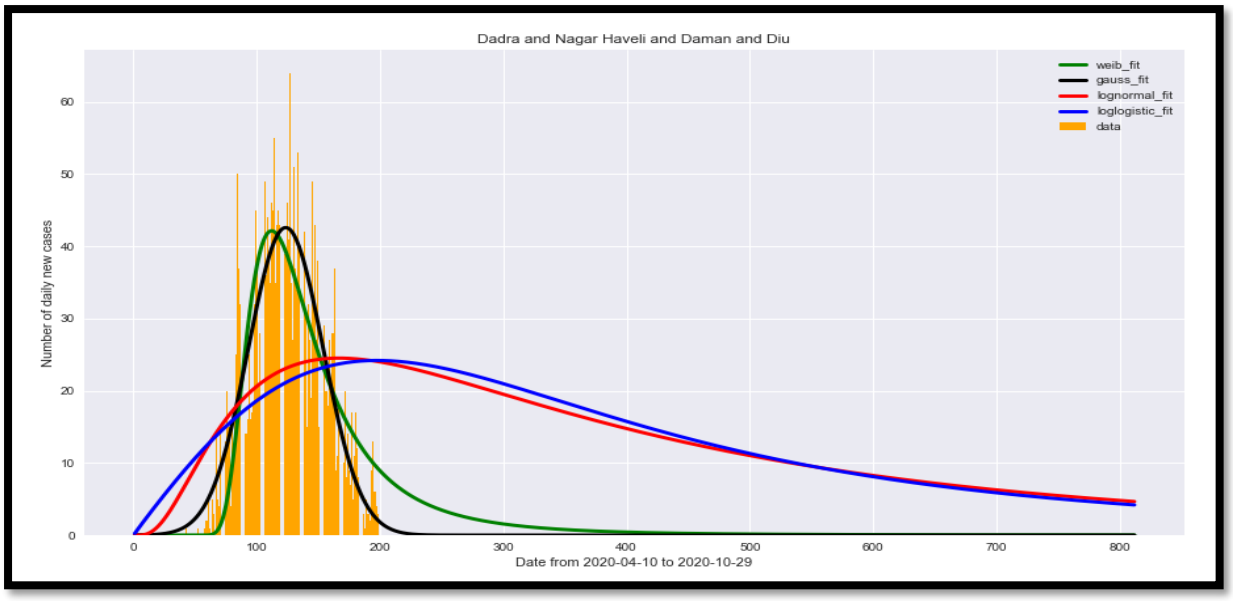


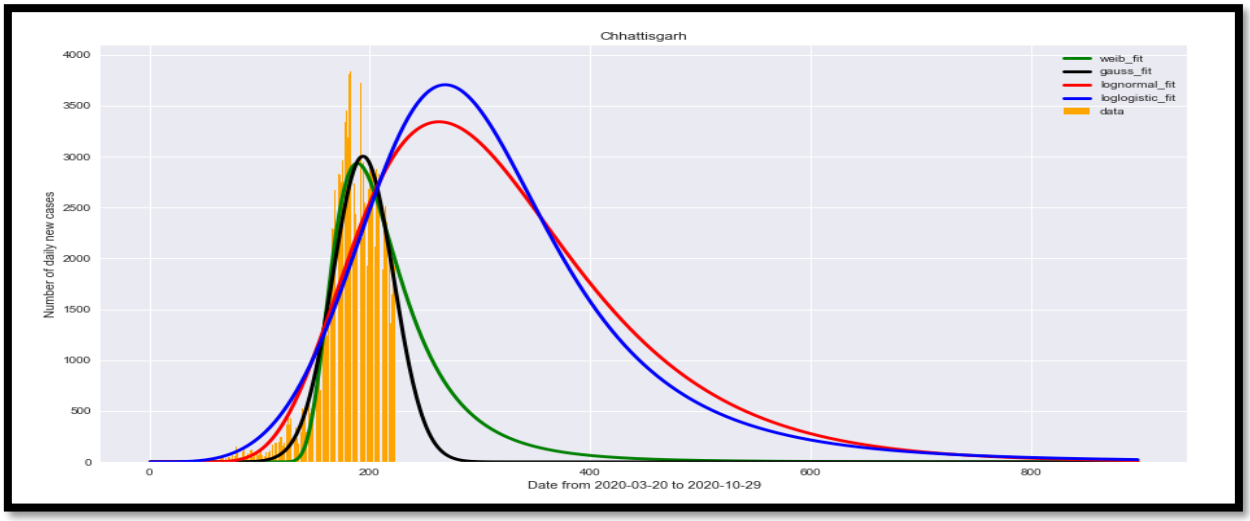

Arunachal Pradesh
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
50
100
150
200
250
300
0
200
400
600
800
Date from 2020-04-03 to 2020-10-29

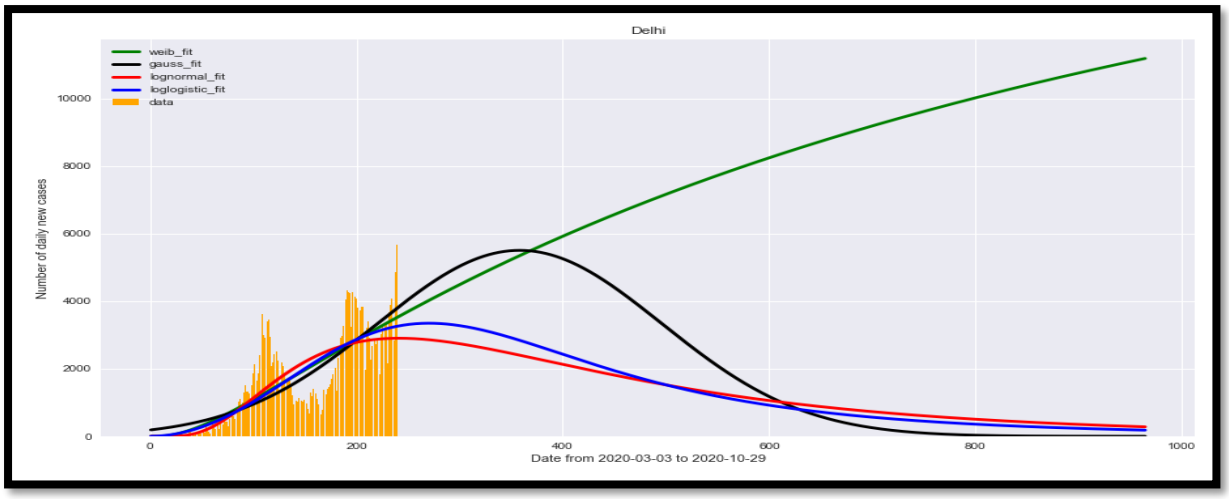

Assam
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
1000
2000
3000
4000
0
200
400
600
800
Date from 2020-04-01 to 2020-10-29

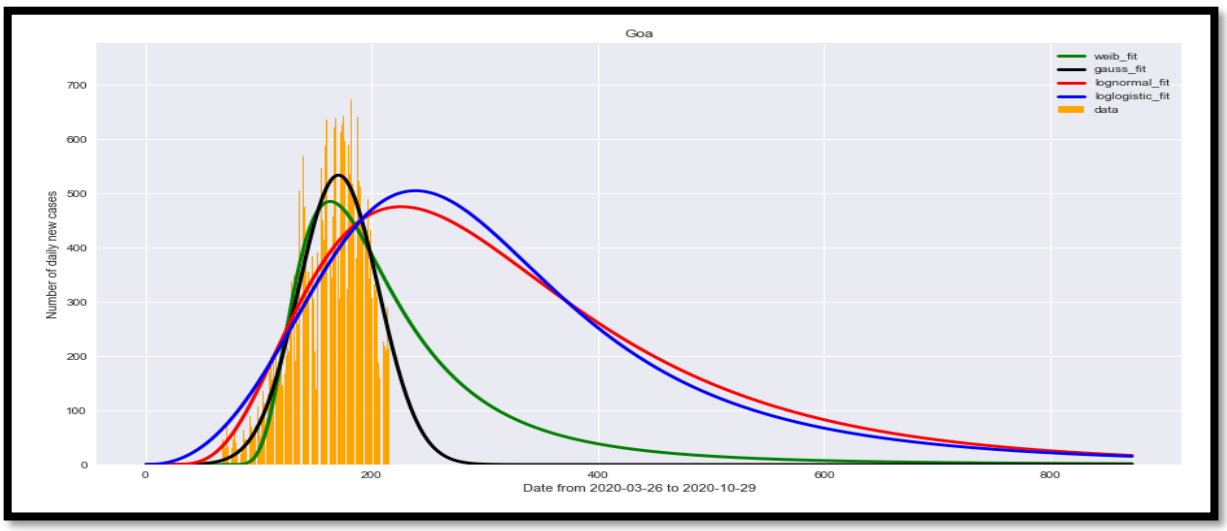

Bihar
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
500
1000
1500
2000
2500
3000
3500
4000
0
200
400
600
800
Date from 2020-03-23 to 2020-10-29

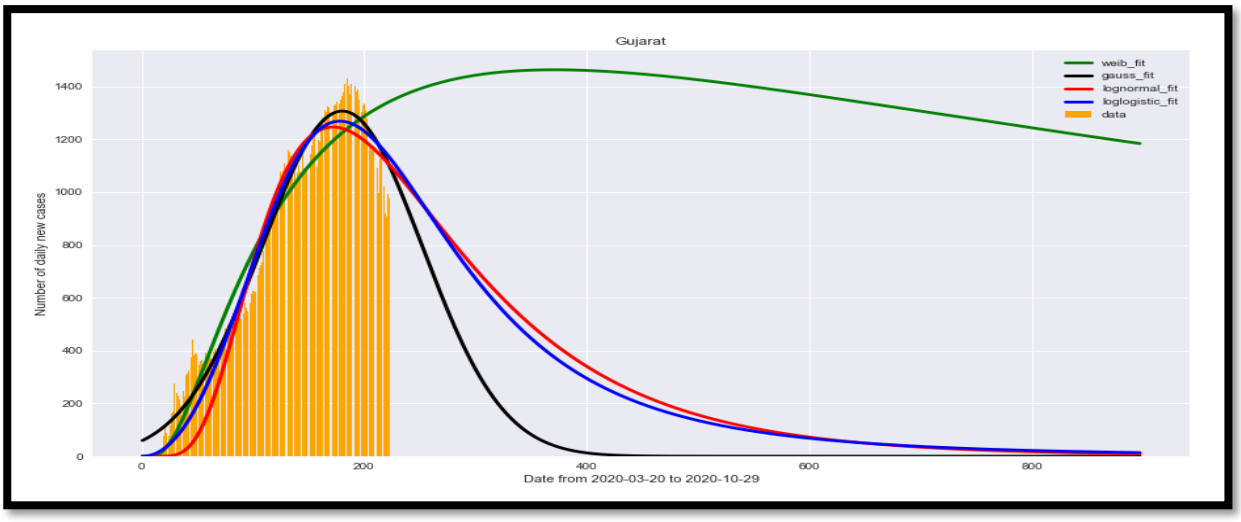

Chandigarh
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-20 to 2020-10-29

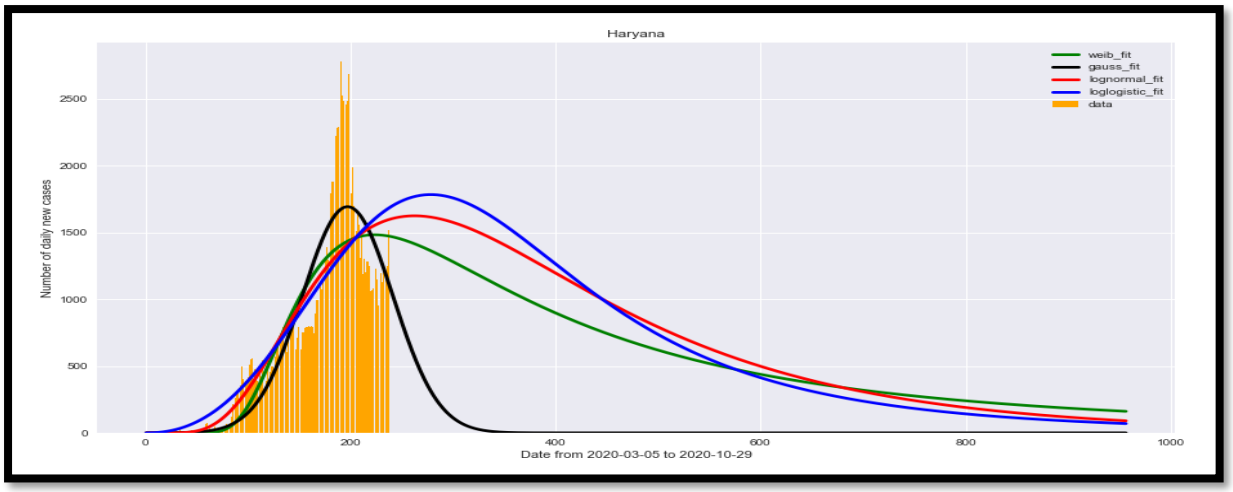

Dadra and Nagar Haveli and Daman and Diu
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-04-10 to 2020-10-29

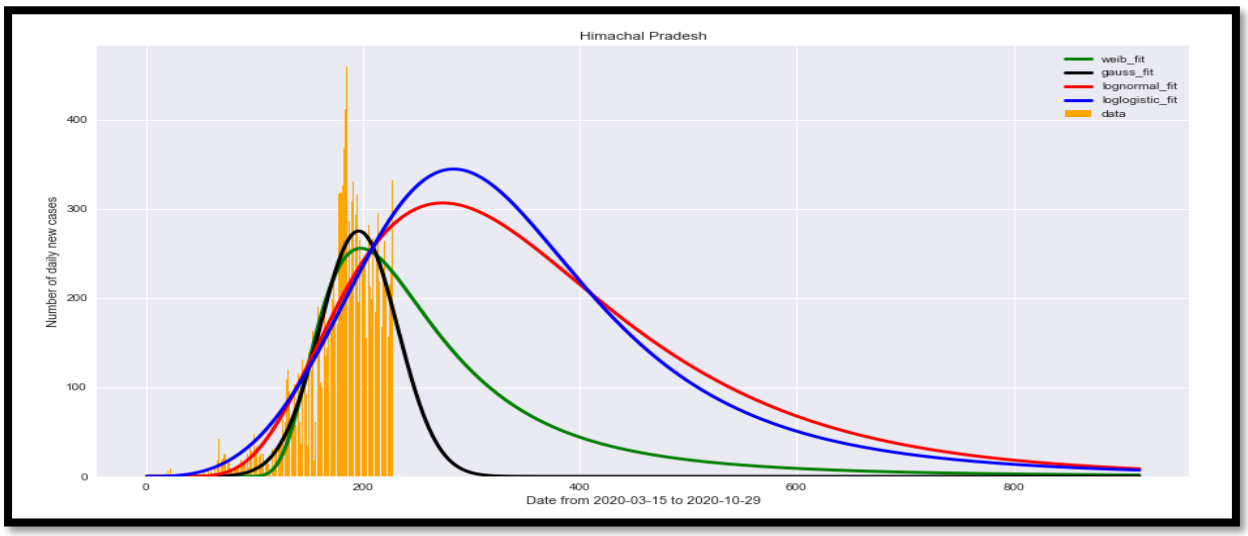

Chhattisgarh
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-20 to 2020-10-29

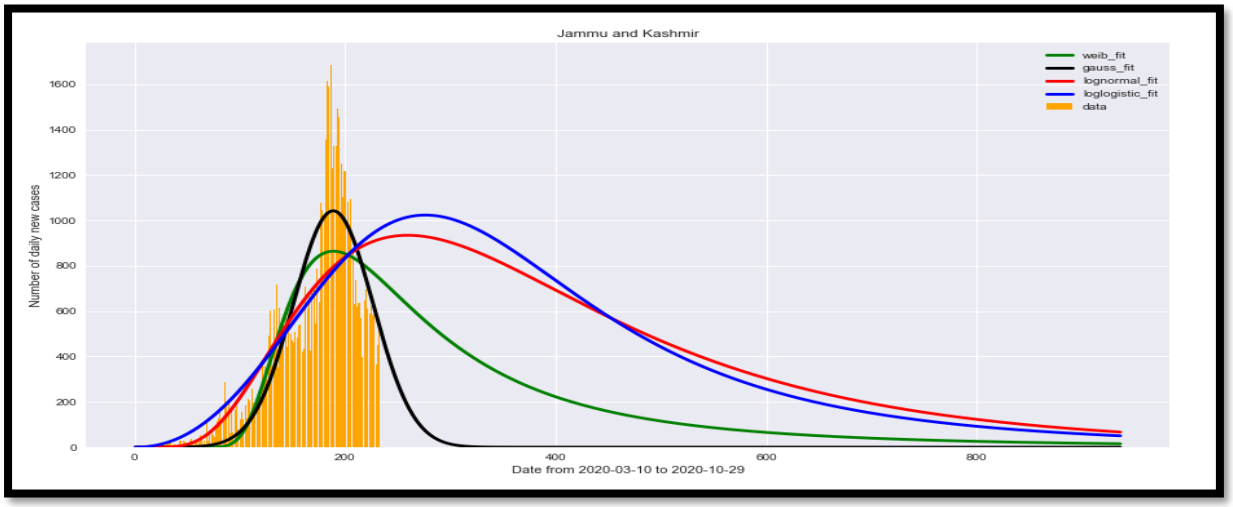

Delhi
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
2000
4000
6000
8000
10000
0
200
400
600
800
1000
Date from 2020-03-03 to 2020-10-29

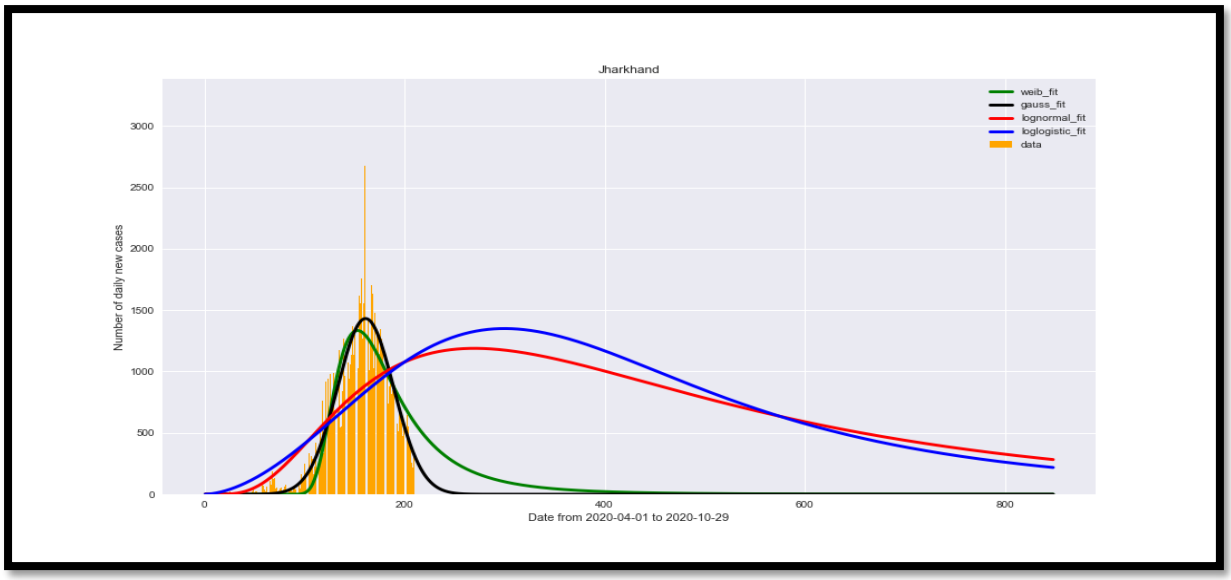

Goa
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
100
200
300
400
500
600
700
0
200
400
600
800
Date from 2020-03-26 to 2020-10-29

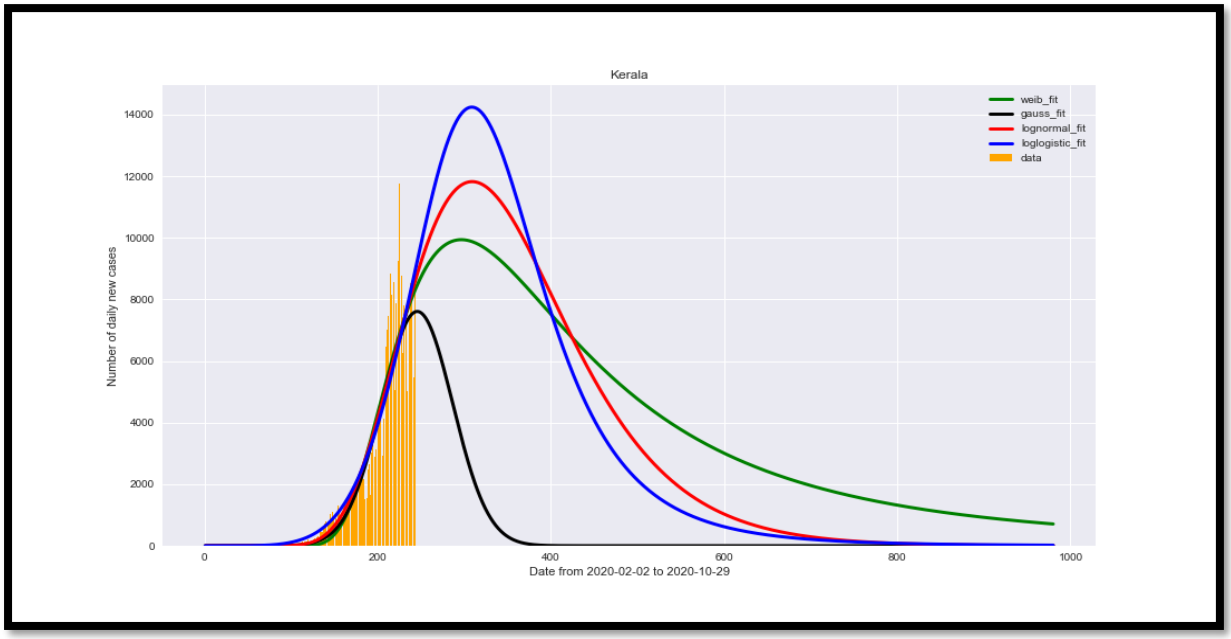

Gujarat
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
200
400
600
800
1000
1200
1400
0
200
400
600
800
Date from 2020-03-20 to 2020-10-29

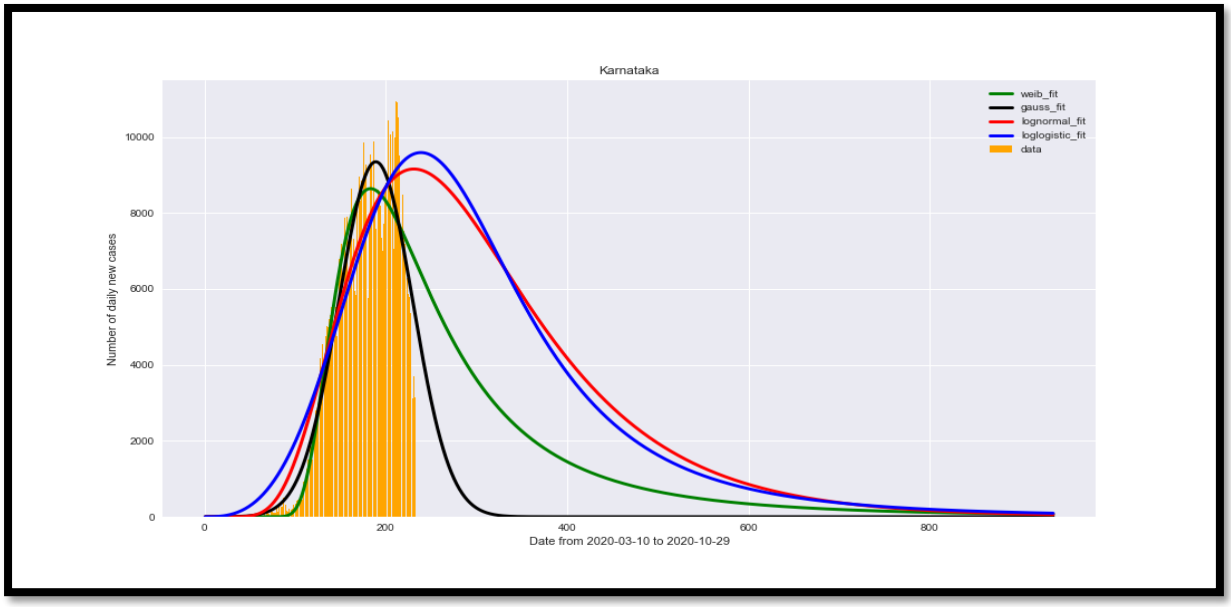

Haryana
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
500
1000
1500
2000
2500
0
200
400
600
800
1000
Date from 2020-03-05 to 2020-10-29

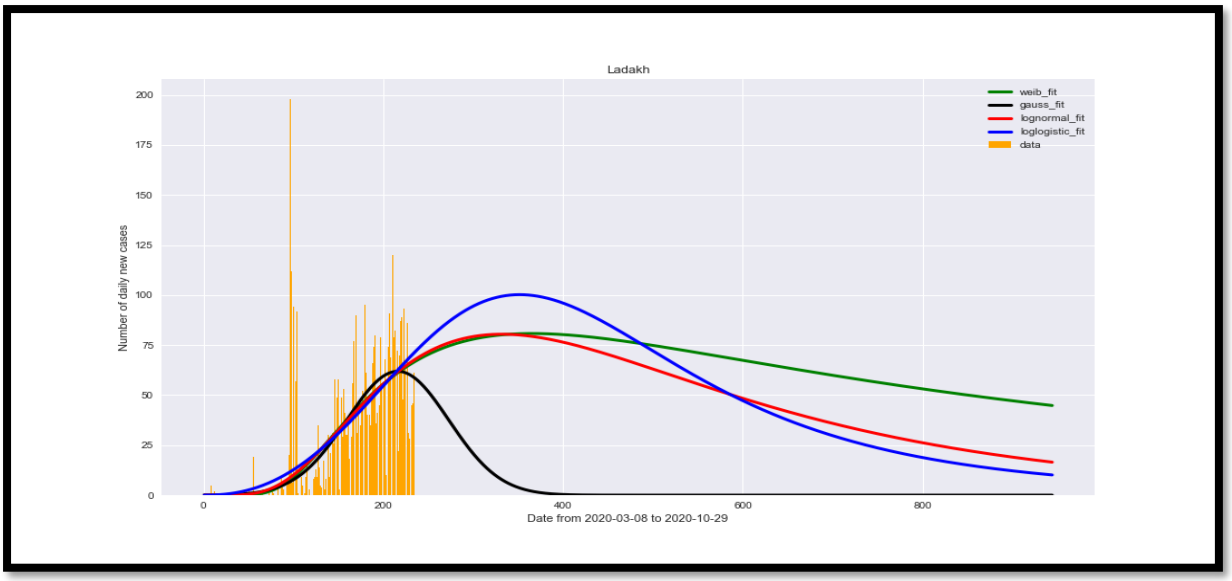

Himachal Pradesh
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
100
200
300
400
0
200
400
600
800
Date from 2020-03-15 to 2020-10-29

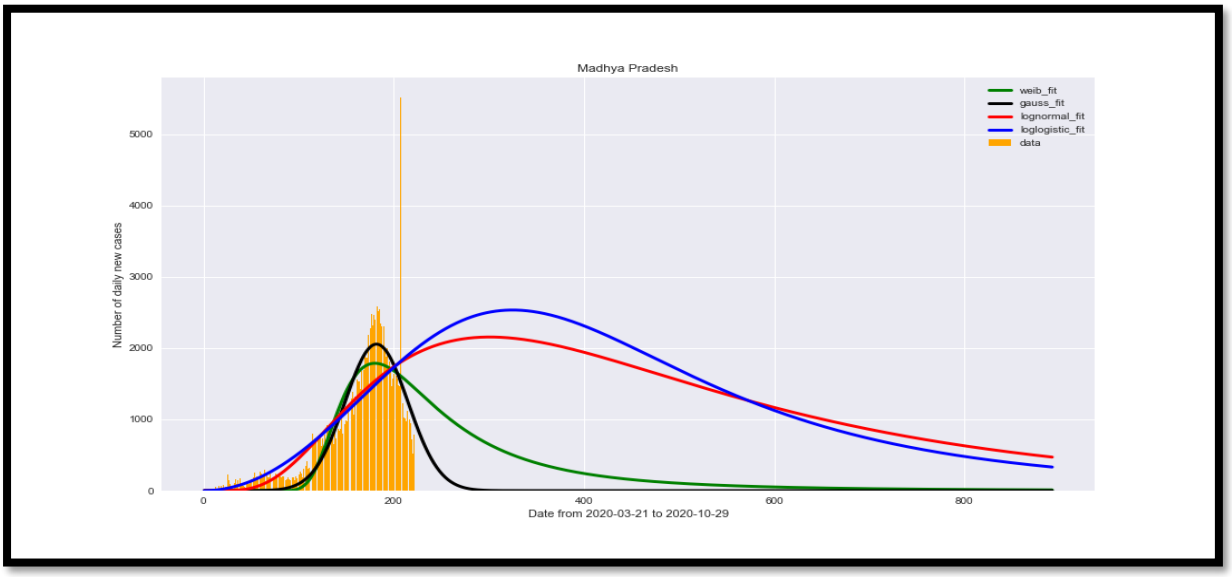

Jammu and Kashmir
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
200
400
600
800
1000
1200
1400
1600
0
200
400
600
800
Date from 2020-03-10 to 2020-10-29

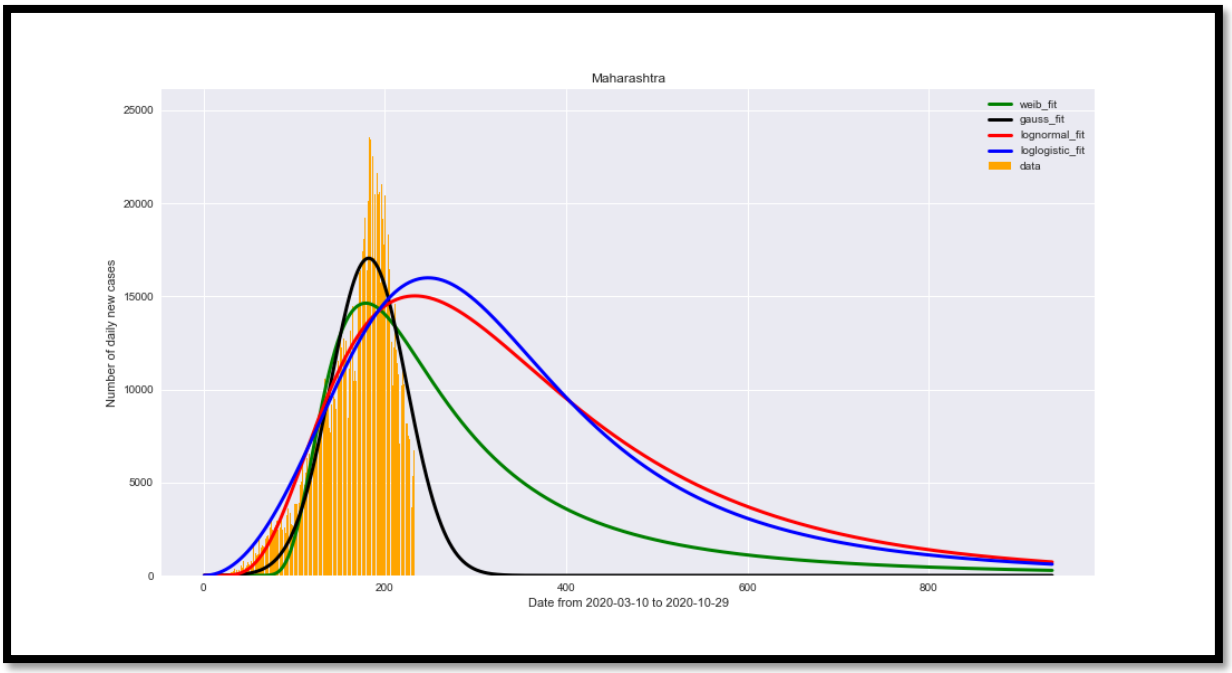

Jharkhand
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-04-01 to 2020-10-29

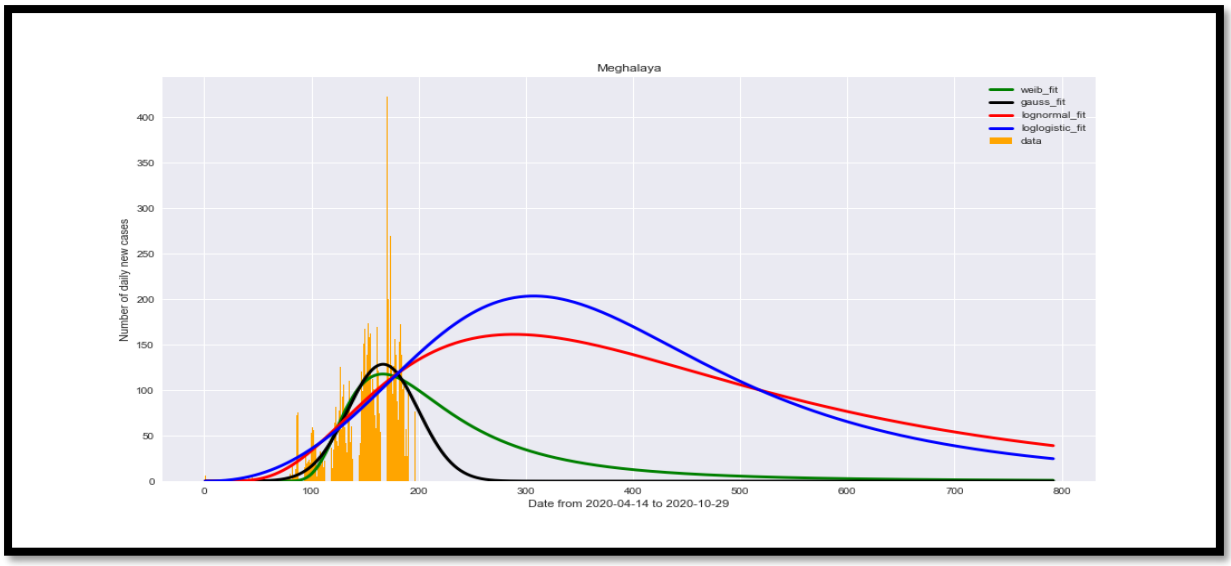

Kerala
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-02-02 to 2020-10-29

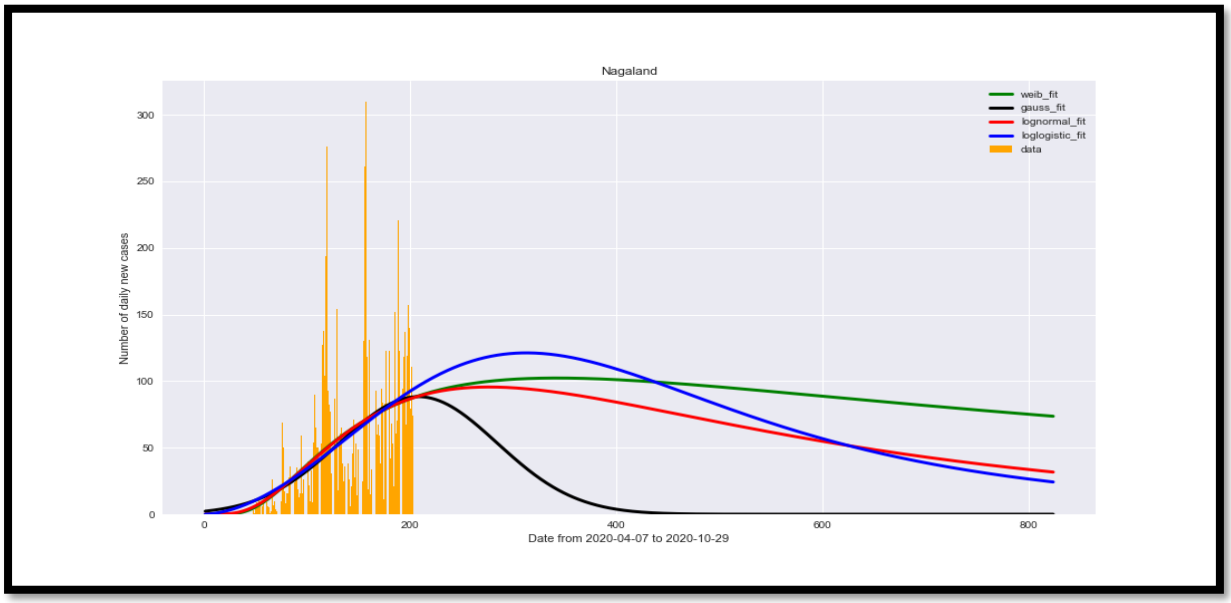

Karnataka
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-10 to 2020-10-29

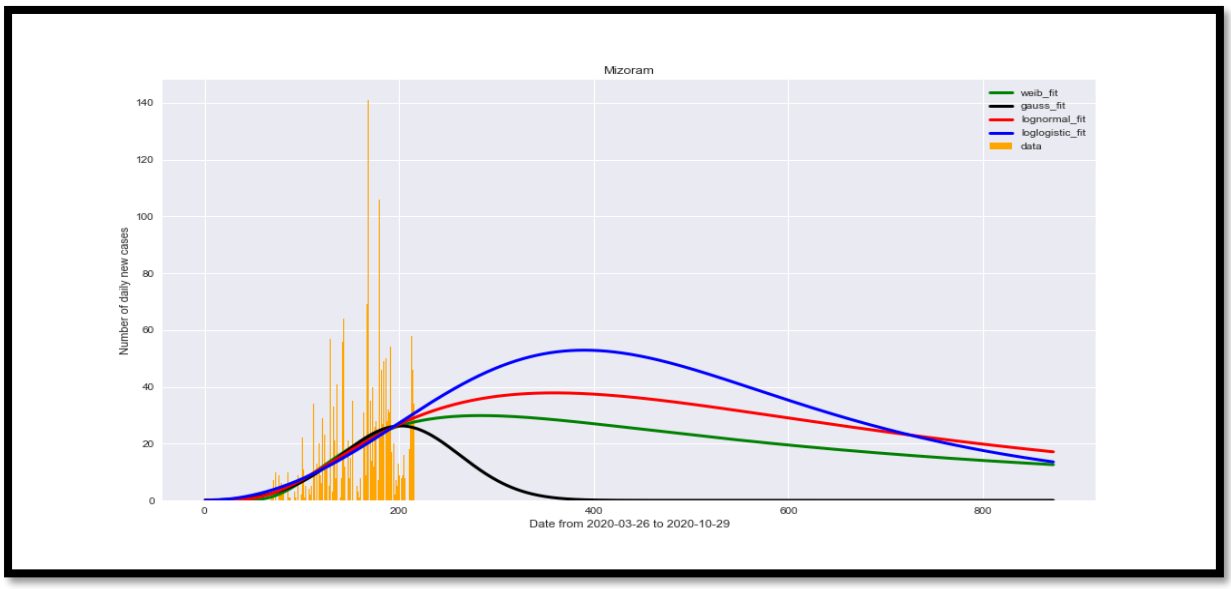

Ladakh
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-08 to 2020-10-29

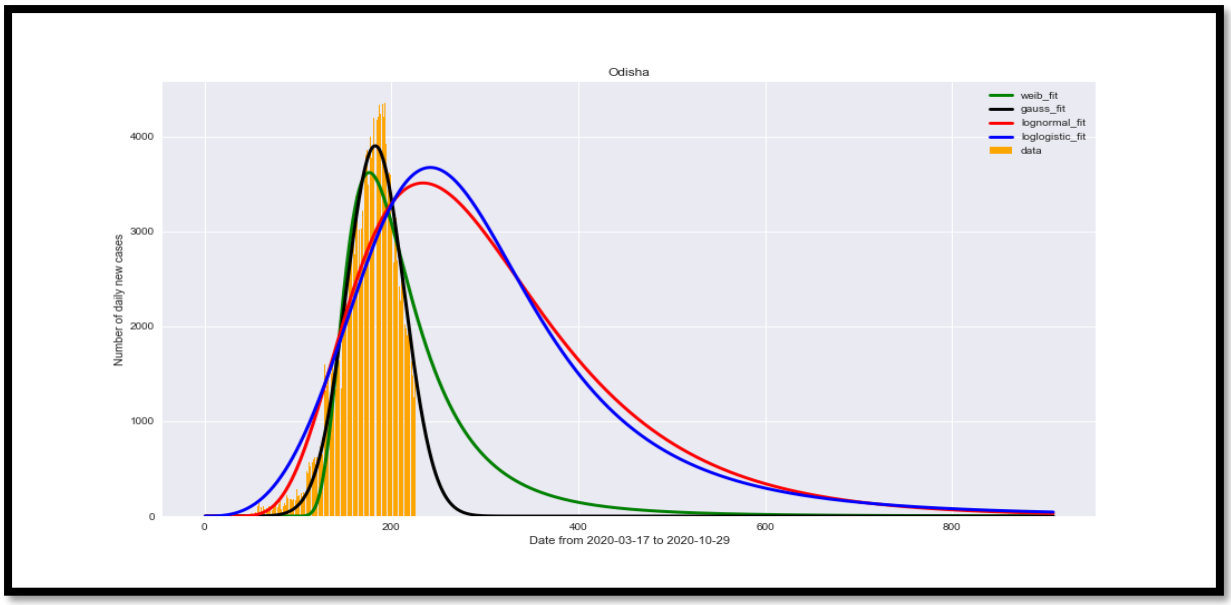

Madhya Pradesh
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-21 to 2020-10-29

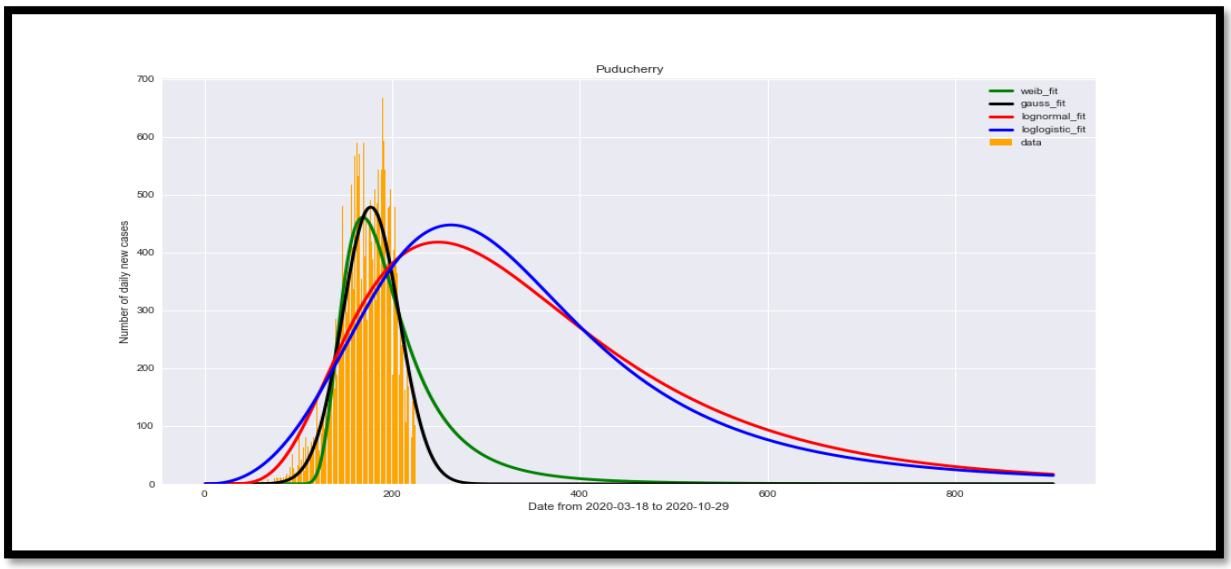

Maharashtra
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-10 to 2020-10-29

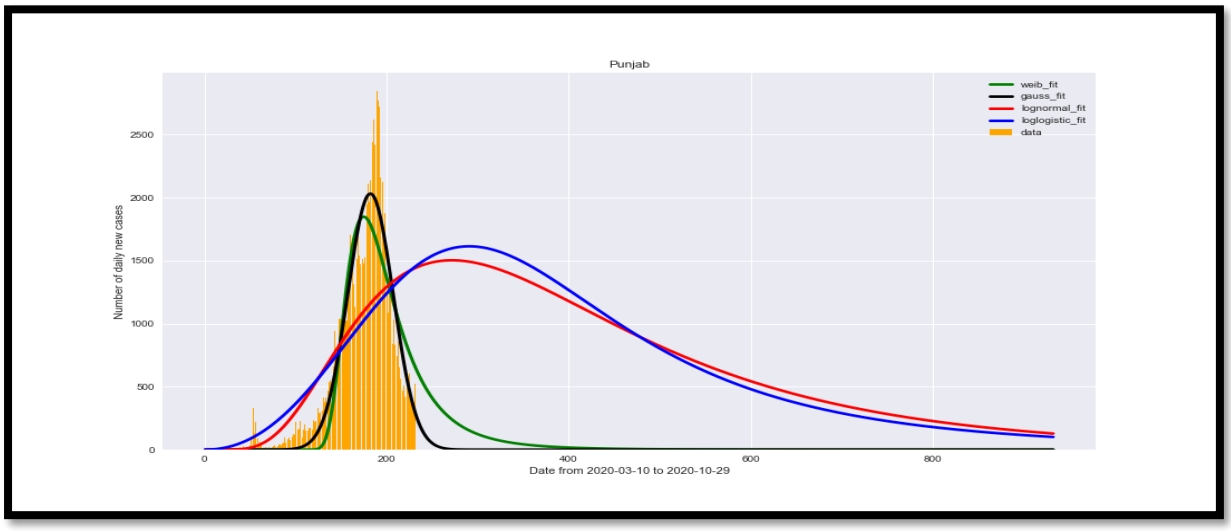

Meghalaya
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-04-14 to 2020-10-29

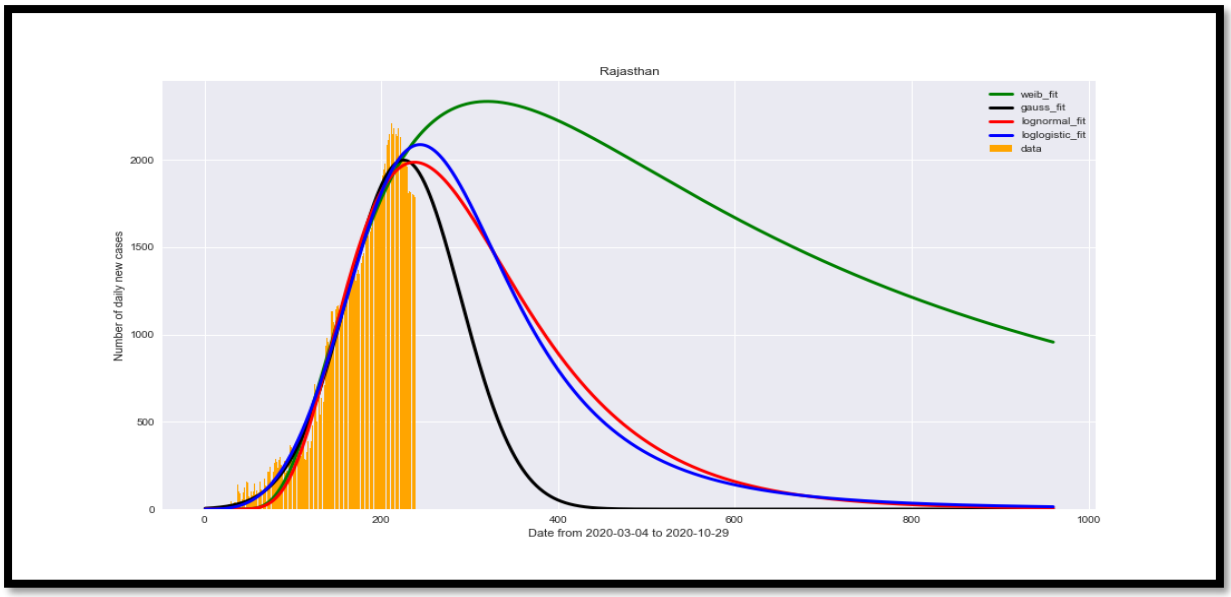

Nagaland
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-04-07 to 2020-10-29

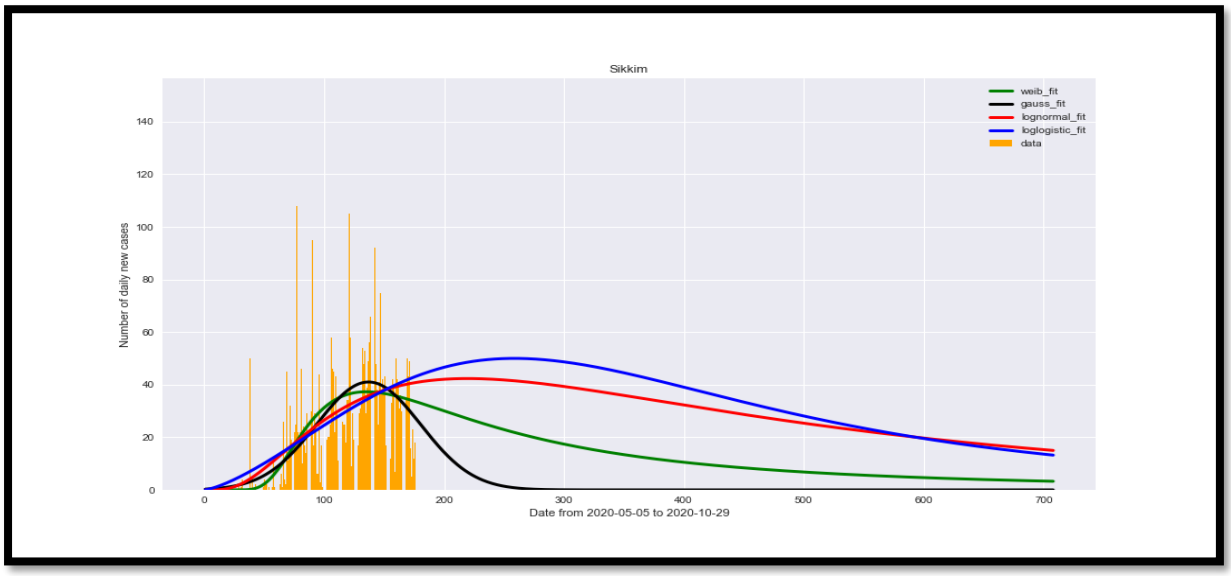

Mizoram
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-26 to 2020-10-29

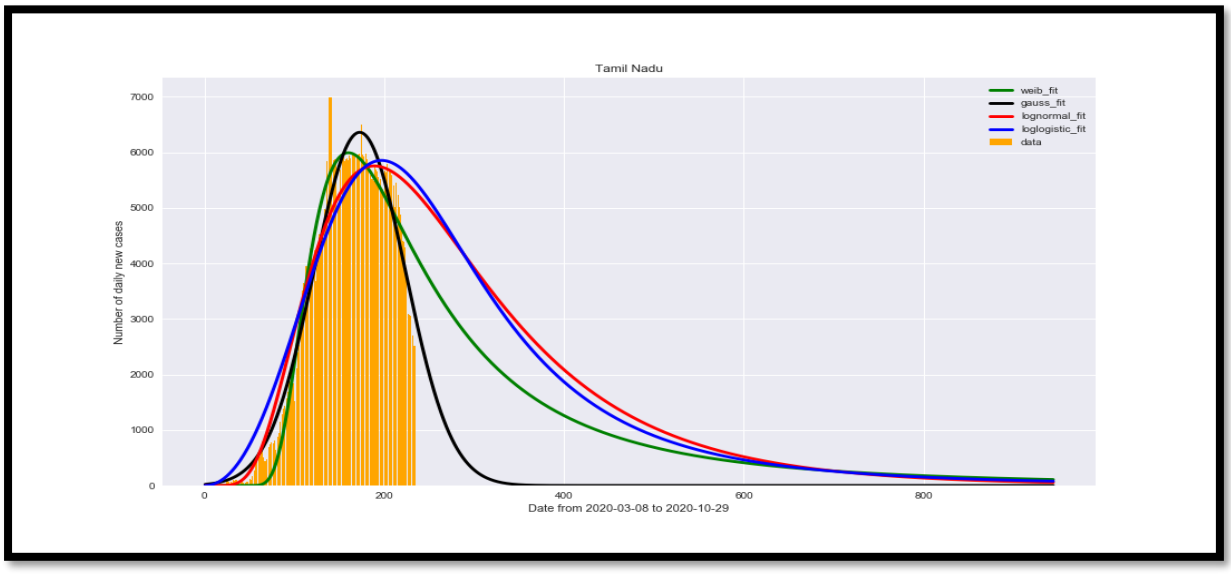

Odisha
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
1000
2000
3000
4000
0
200
400
600
800
Date from 2020-03-17 to 2020-10-29

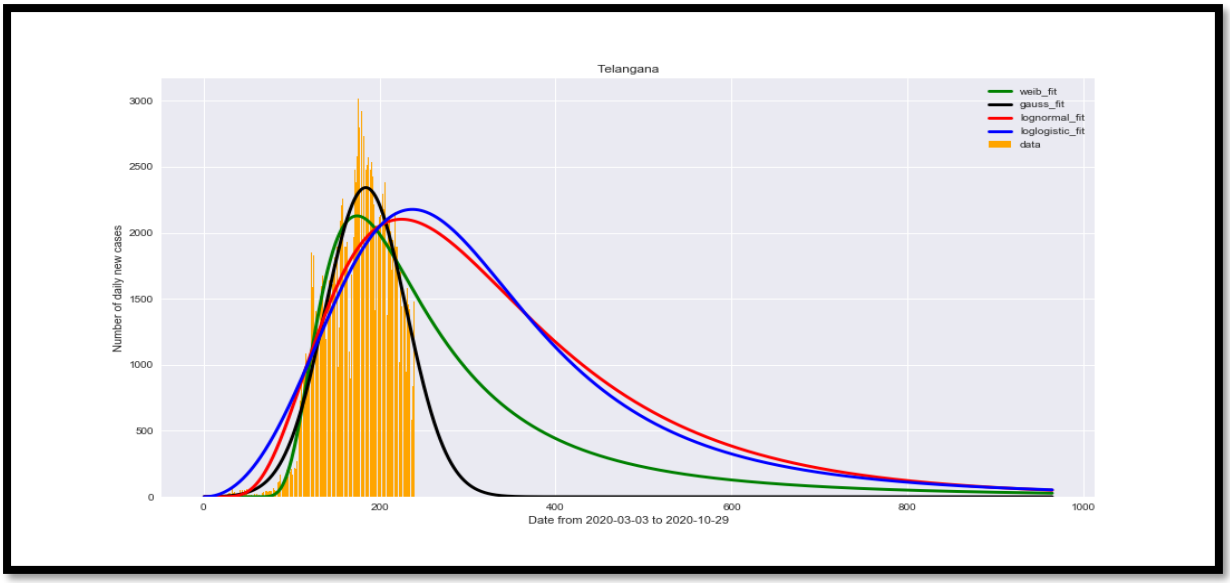

Puducherry
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
100
200
300
400
500
600
700
0
200
400
600
800
Date from 2020-03-18 to 2020-10-29

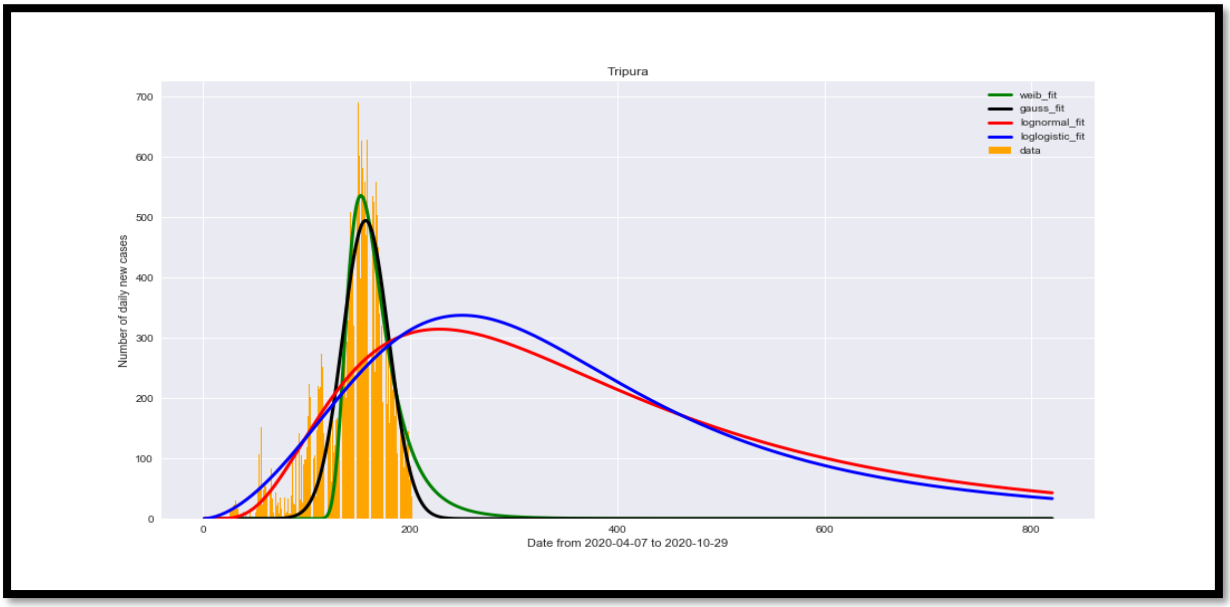

Punjab
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
500
1000
1500
2000
2500
0
200
400
600
800
Date from 2020-03-10 to 2020-10-29

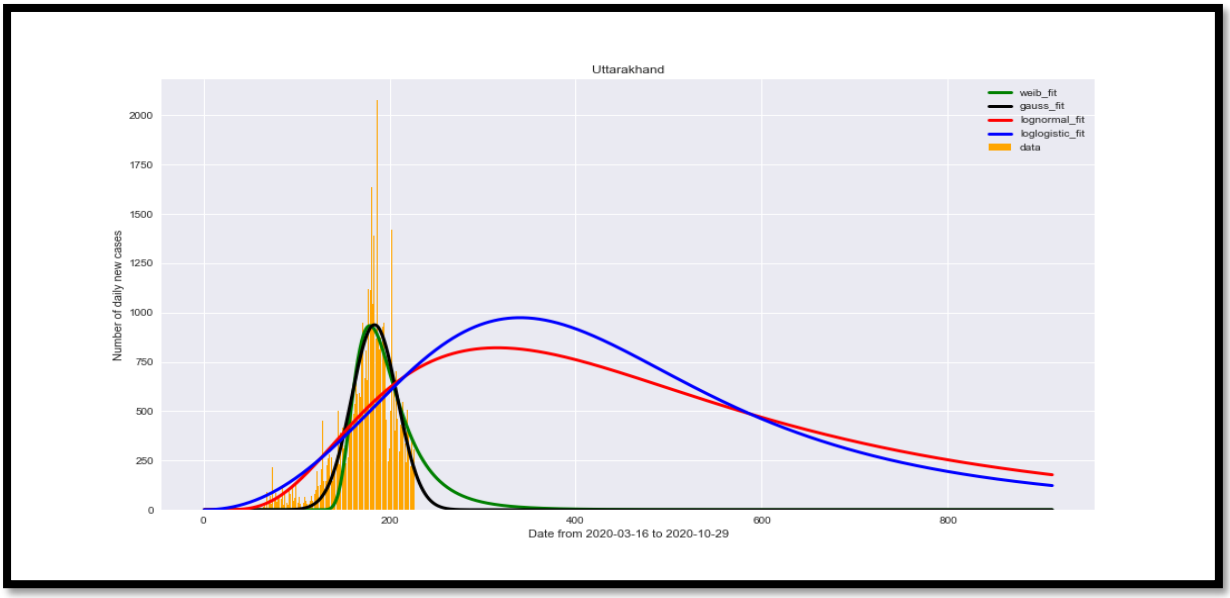

Rajasthan
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
500
1000
1500
2000
0
200
400
600
800
1000
Date from 2020-03-04 to 2020-10-29

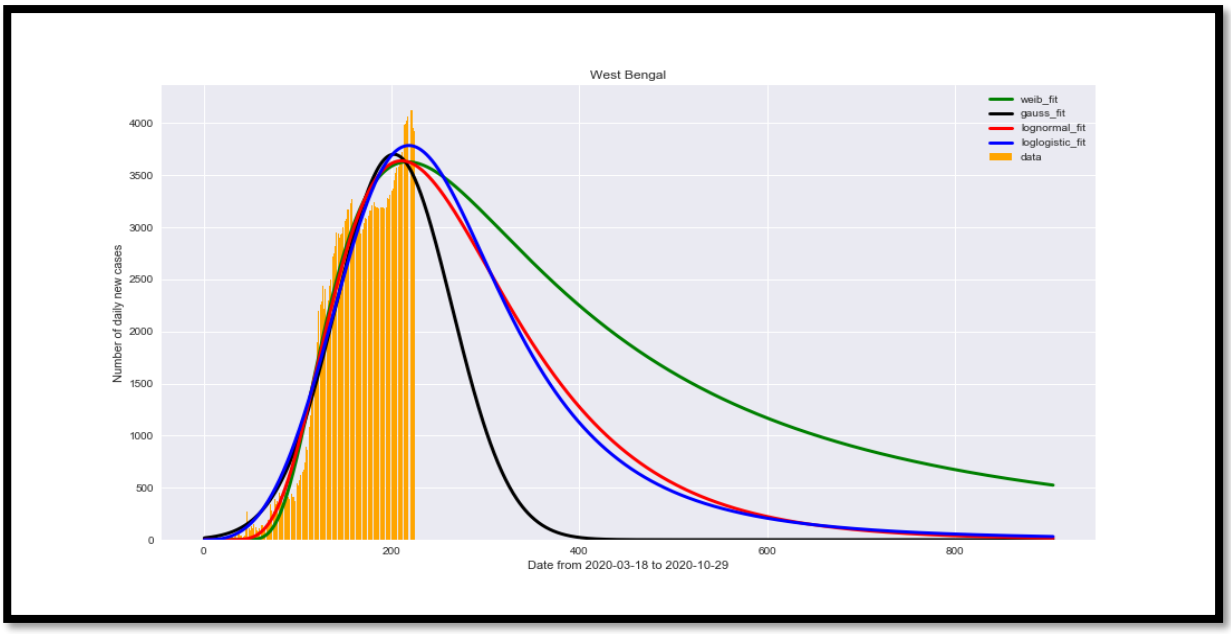

Sikkim
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
20
40
60
80
100
120
140
0
100
200
300
400
500
600
700
Date from 2020-05-05 to 2020-10-29

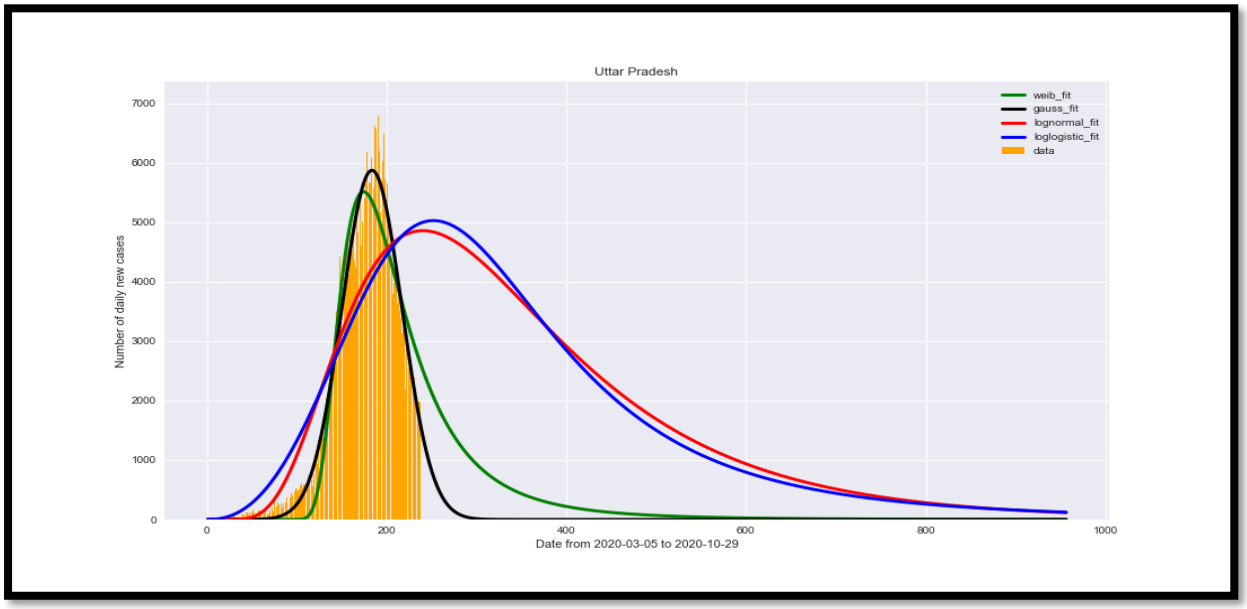

Tamil Nadu
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
0
1000
2000
3000
4000
5000
6000
7000
0
200
400
600
800
Date from 2020-03-08 to 2020-10-29


Telangana
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-03 to 2020-10-29


Tripura
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-04-07 to 2020-10-29


Uttarakhand
weib_fit
gauss_fit
lognormal_fit
loglogistic_fit
data
Number of daily new cases
Date from 2020-03-16 to 2020-10-29





Link to the code: [Google-Colab]

- These were some of the analyses done on all the states and thus based upon these fits we have calculated the value of all metrics. They are tabulated onto next page:
- There are two tables which describe the comparison between all the metrics of Weibull-Gauss and LogNormal-LogLogistic:
- The third table includes the Expected End Date (One and Two) and Expected Cases (one and Two) corresponding to all states.

State	MAE (W)	MAE(G)	MSE(W)	MSE(G)	R2-Score (W)	R2-Score (G)
Kerala	0.03190954	0.030080813	**0.004844447**	0.004934368	**0.901181177**	0.887784527
Delhi	**0.088914714**	0.09310033	0.01643328	**0.015898067**	0.588253366	**0.589107827**
Telangana	0.057876572	**0.052954191**	0.009905252	**0.008656675**	0.888775921	**0.899991855**
Rajasthan	0.036458312	0.031214747	0.002513531	**0.001826032**	**0.97752392**	0.983386331
Haryana	0.104155724	**0.087089344**	0.025297387	**0.015750034**	0.472823249	**0.728287807**
Uttar Pradesh	0.038149424	**0.03068761**	0.003034059	**0.001806599**	0.960579133	**0.975200803**
Ladakh	0.069444296	**0.067258942**	0.024986468	**0.024589822**	-0.621217232	-0.597899445
Tamil Nadu	**0.048898113**	0.048915736	0.005271666	0.004761384	0.955960686	**0.957897223**
Jammu and Kashmir	0.074893527	0.06674574	0.012998421	**0.008749017**	0.706822707	**0.829263638**
Karnataka	**0.041186034**	0.054442739	**0.003456665**	0.006415206	**0.963911632**	0.930319885
Maharashtra	0.105559875	**0.082556131**	0.021632254	**0.01322901**	0.685606161	**0.825314333**
Punjab	0.059135334	**0.040884005**	0.006733697	**0.003267583**	0.866054196	**0.939811172**
Andhra Pradesh	**0.032064379**	0.04359085	**0.002150114**	0.004601186	**0.982753117**	0.962152108
Himachal Pradesh	0.051609796	0.052056954	**0.007090735**	0.007892874	**0.85508684**	0.834692401
Uttarakhand	0.044245673	**0.040902736**	0.004501315	**0.003602487**	0.814722358	**0.852554422**
Odisha	0.046085107	0.026351647	0.004648816	**0.001676148**	0.953712764	**0.982282779**
Puducherry	0.056269057	**0.048214028**	0.007842193	**0.006807954**	0.889986517	**0.902855702**
West Bengal	0.047076266	0.059318875	**0.003204217**	0.005148223	**0.971397351**	0.948452695
Chandigarh	0.050802622	**0.037845088**	0.008199751	**0.005897934**	0.755745695	**0.829864656**
Chhattisgarh	0.04022414	**0.034895816**	0.003430631	0.003603575	**0.952043142**	0.946727879
Gujarat	0.076474867	**0.051765843**	0.011265894	**0.008829504**	0.872206379	**0.900203923**
Madhya Pradesh	0.03630597	0.030523616	0.002174918	**0.001471266**	0.852454917	**0.900302405**
Bihar	**0.04353353**	0.054259066	**0.003752768**	0.007617684	**0.93424381**	0.847521169
Manipur	**0.06716252**	0.068203885	0.01004391	**0.009978711**	0.748348096	0.746136965
Goa	0.072393419	**0.064642286**	**0.012192819**	0.010661254	0.845386346	**0.873576559**
Mizoram	**0.039453116**	0.041528944	**0.003732194**	0.003827153	**0.347086576**	0.303595741
Andaman and Nicobar Islands	**0.048294462**	0.064470773	**0.004551117**	0.00911203	**0.755891726**	0.551396215
Assam	0.073094592	**0.059408387**	0.012289926	**0.010099282**	0.675852135	**0.733051003**
Jharkhand	0.045157392	**0.033261258**	0.004914475	**0.00325189**	0.793027518	0.858774927
Arunachal Pradesh	0.056619686	**0.053829665**	0.009427053	**0.007582712**	0.820302109	**0.873069621**
Nagaland	**0.073092653**	0.076323751	**0.014206638**	0.014955008	-0.523313481	-0.665535324
Tripura	0.065095169	0.072261484	**0.011744021**	0.012464533	**0.82613588**	0.813180147
Dadra and Nagar Haveli and Daman and Diu	0.074052808	**0.063093843**	0.010657929	**0.007444338**	0.75004935	**0.84109568**
Meghalaya	0.038328778	**0.03486457**	0.004250495	**0.003733295**	0.615474253	**0.676228488**
Sikkim	**0.083389536**	0.086446578	**0.034207717**	0.035236517	**-1.639890192**	-1.919470017

State	MAE (LN)	MAE(LL)	MSE(LN)	MSE(LL)	R2-Score(LN)	R2-Score(LL)
Kerala	0.027435606	**0.027138037**	0.004880965	0.005301431	0.898911529	0.887090177
Delhi	0.099798758	0.09332905	0.020555844	0.01794588	0.453148938	0.539178515
Telangana	0.085571886	0.10452221	0.016836276	0.021154834	0.79564989	0.720379614
Rajasthan	0.041285423	**0.029540218**	0.002854928	0.001835251	0.975135748	0.983070771
Haryana	0.104022794	0.108677907	0.026886343	0.028859151	0.390701852	0.295168872
Uttar Pradesh	0.092740862	0.11041748	0.019733395	0.023476371	0.701157423	0.599308137
Ladakh	0.070772862	0.075797409	0.024789673	0.024639065	-0.63272753	-0.702939839
Tamil Nadu	0.060013531	0.081804748	**0.011075079**	0.014860077	0.906307601	0.863173388
Jammu and Kashmir	**0.065724201**	0.072059463	0.013699001	0.014712285	0.594377025	0.505294253
Karnataka	0.068533384	0.089846348	0.00965582	0.013511342	0.896670555	0.843564767
Maharashtra	0.087812884	0.09295138	0.022515737	0.023871494	0.601908681	0.536127623
Punjab	0.08867599	0.096328673	0.018359009	0.020135805	0.449641741	0.303312668
Andhra Pradesh	0.1535082	0.171667529	0.045750223	0.051821731	0.54459261	0.421494822
Himachal Pradesh	**0.050873411**	0.056747831	0.00817634	0.009140935	0.817364388	0.784156928
Uttarakhand	0.055499835	0.060256826	0.007611679	0.00833171	0.423730885	0.275446828
Odisha	0.087973377	**0.106910626**	0.019039444	0.023143529	0.800446652	0.741810648
Puducherry	0.106026306	0.123636245	0.026036543	0.030403716	0.473734721	0.301172179
West Bengal	**0.046900533**	0.05716299	0.003538179	0.00508107	0.967272359	0.947572294
Chandigarh	0.09453422	0.098538728	0.030283076	0.031791419	-0.714954071	-1.130144557
Chhattisgarh	0.072340364	0.083311476	**0.015705249**	0.01841513	0.734117008	0.666313936
Gujarat	0.086492055	0.066040601	0.019056178	0.011759232	0.824662889	0.877262725
Madhya Pradesh	0.028198046	**0.027949874**	0.001924938	0.002058466	0.843338334	0.812370439
Bihar	0.093732777	0.112741444	0.020759924	0.024141776	0.217799114	-0.195835626
Manipur	0.069199959	0.066769787	0.010271307	0.009923414	0.749733065	**0.751882865**
Goa	0.09235984	0.106066451	0.019223721	0.022177643	0.683238561	0.5866501
Mizoram	0.040314645	0.041594078	0.003846087	0.003970475	0.314899613	0.270042915
Andaman and Nicobar Islands	0.071733185	0.074919647	0.016533793	0.01715005	-2.52663426	-3.751849587
Assam	0.099634063	0.110905776	0.020324825	0.022553401	0.167882397	-0.09156167
Jharkhand	0.075632696	0.083774104	0.012837535	0.014328565	0.167874709	-0.057844704
Arunachal Pradesh	0.060814274	0.069254023	0.009177169	0.010241703	0.799291312	0.752808806
Nagaland	0.073499864	0.076109711	0.014287725	0.014923383	-0.542743332	-0.598110736
Tripura	**0.114428025**	0.124362853	0.028991124	0.03178542	0.043589359	-0.222230492
Dadra and Nagar Haveli and Daman and Diu	0.17344685	0.187337703	0.042633425	0.046823371	-1.503441872	-2.793076777
Meghalaya	0.035679223	0.040075058	0.004399375	0.004777324	0.47273794	0.356551965
Sikkim	0.086958611	0.092773373	0.035363661	0.036339322	-2.191250715	-2.786744063

State	Expected last date 1	Expected Cases-1	Expected last date2	Expected Cases-2
Kerala	2021-03-26(G)	0	2022-10-07(LN)	7
Delhi	2022-09-17(G)	0	2022-10-21(LL)	182
Telangana	2021-03-23(G)	0	2022-10-21(W)	27
Rajasthan	2021-06-23(G)	0	2022-10-18(LN)	6
Haryana	2021-06-23(G)	0	2022-10-15(LL)	69
Uttar Pradesh	2021-01-21(G)	0	2022-10-15(W)	2
Ladakh	2021-03-21(G)	0	2022-10-06(LL)	10
Tamil Nadu	2021-03-28(G)	0	2022-10-06(LN)	53
Jammu and Kashmir	2021-02-02(G)	0	2022-09-30(W)	15
Karnataka	2021-03-08(G)	0	2022-09-30(LN)	55
Maharashtra	2021-03-11(G)	0	2022-09-30(W)	275
Punjab	2020-12-14(G)	0	2021-10-24(W)	0
Andhra Pradesh	2021-01-14(G)	0	2022-09-21(W)	1
Himachal Pradesh	2021-01-25(G)	0	2022-09-15(W)	1
Uttarakhand	2020-12-15(G)	0	2021-06-25(W)	0
Odisha	2021-01-19(G)	0	2022-09-09(W)	1
Puducherry	2020-12-26(G)	0	2021-10-19(W)	0
West Bengal	2021-06-13(G)	0	2022-09-06(LN)	16
Chandigarh	2020-11-15(G)	0	2021-02-11(W)	0
Chhattisgarh	2021-01-17(G)	0	2022-04-05(W)	0
Gujarat	2021-06-15(G)	0	2022-08-31(LN)	8
Madhya Pradesh	2021-01-25(G)	0	2022-08-28(W)	10
Bihar	2020-12-30(G)	0	2022-02-24(W)	0
Manipur	2021-09-08(G)	0	2022-08-16(LL)	6
Goa	2021-01-15(G)	0	2022-08-13(W)	1
Mizoram	2021-03-18(G)	0	2022-08-13(W)	13
Andaman and Nicobar Islands	2020-10-31(G)	0	2020-10-31(W)	0
Assam	2021-01-07(G)	0	2022-07-26(W)	1
Jharkhand	2020-12-24(G)	0	2022-02-26(W)	0
Arunachal Pradesh	2021-01-17(G)	0	2022-07-20(W)	6
Nagaland	2021-06-17(G)	0	2022-07-08(LL)	24
Tripura	2020-11-28(G)	0	2021-03-14(W)	0
Dadra and Nagar Haveli and Daman and Diu	2020-11-02(G)	0	2021-03-06(W)	0
Meghalaya	2021-01-08(G)	0	2022-05-28(W)	0
Sikkim	2021-01-07(G)	0	2022-04-11(W)	3

OBJECTIVE RESULTS:

1.Which states are under adverse conditions and which are under good control?
Ans- Delhi, Maharashtra, Rajasthan and Gujarat are some of the states which are in adverse condition and Andhra Pradesh, Himachal Pradesh, Telangana and Assam are in good condition.

2. What is the expected number of cases after the end of November (2022)?
Ans - The expected number of cases after end of November is 12627 (as per Weibull's Model), 73 (as per Log-Normal), 231 (as per Log-Logistic)

3. Which state will get rid of COVID-19 first (i.e., Number of cases < 1)?
Ans - Andaman and Nicobar Islands is to get rid of Covid -19 first.

4. When will COVID-19 vanish in India?
Ans- The Covid -19 will vanish in India by 17th May 2021 (According to best fit). The second-best fit is Log-Normal with 73cases on 2022-10-04.

5. Which state has a chance to suffer from another wave?
Ans - Maharashtra, Delhi, West Bengal and Rajasthan are some of the states that are likely to have second wave.

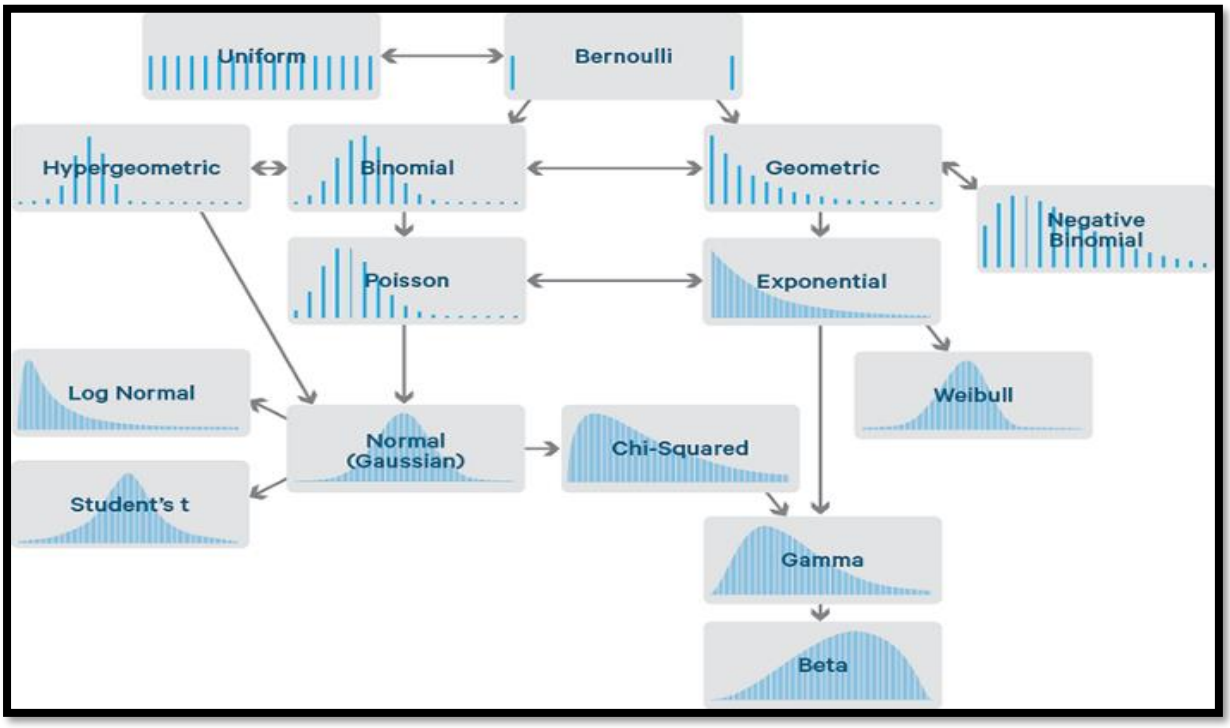
MISCELLANEOUS:

OTHER PACKAGES USED:

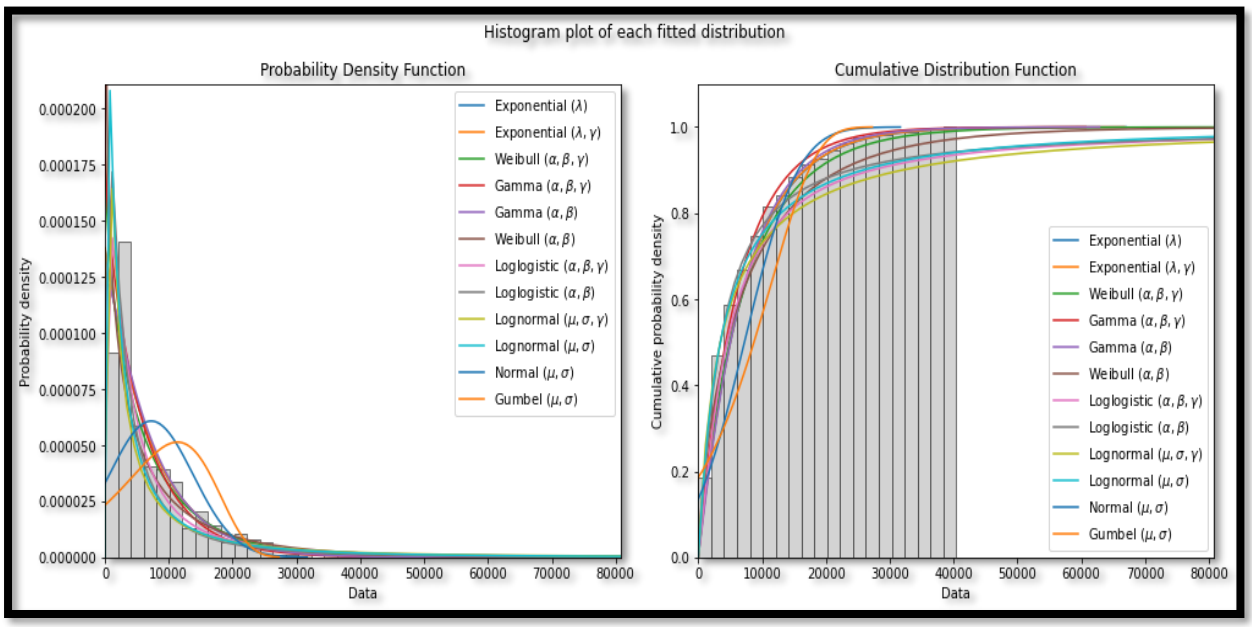
Reliability is a Python Library used for reliability engineering and survival analysis. We have plotted several plots with the use of this package. Using the function **Fit Everything,** we have plotted different types of plots and have found out the best distribution available for certain analysis.
Some of the plots which we have used for our analysis are described below:

- Histogram plot: which shows the PDF and CDF of the fitted distributions
- Semiparametric Probability-Probability (PP) plot: Provides a comparison of parametric vs non-parametric fit using a semiparametric Probability-Probability (PP) plot for each fitted distribution.
- Probability plot: shows a probability plot of each of the fitted distributions.

The method we have used for this analysis is Least Squares (LS) Method. All types of distributions including the extreme value distribution functions are being used. The functions we have used are:



Some of the graphs from the analysis of Maharashtra can be found below. It consists of finding out the probability plot of the best distribution as the pdf and CDF of the total number of confirmed cases per day found there.



FUTURE SCOPE:

The COVID-19 pandemic has opened several new directions of research for the current and future pandemics.

- The best fits can be found by involving some of two or three functions, which will be more robust to 2nd and 3rd Wave.

- Considering factors like climate change, population density, healthcare facilities, people's response to pandemic, strain type, distribution of age and individual and community movements will nourish the prediction accuracy.

- ARIMA with Weibull function can be used for time-based predictions.

- Thoughts can be put on AI based data collector from open sources, AI based Robots for contact-less delivery and risk assessment for variable age group can be predicted using AI.

- Following lockdown improvement in AQI has been observed, more extensive studies considering age distribution and demographics with other characteristic van be studied as part future work.

REFERENCES:

1. [Base Paper] - Predicting the growth and trend of COVID-19 pandemic using Machine learning and Cloud Computing.
2. Background Paper on COVID-19 WHO.
3. Prediction of COVID-19 Disease progression in India - under the effect of National Lockdown.
4. Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms.
5. Ghosh, Palash & Ghosh, Rik & Chakra borty, Bibhas. (2020). COVID-19 in India: State-wise Analysis and Prediction.
6. The Levenberg-Marquardt Algorithm - Theory and Implementation.
7. The Generalized Inverse Weibull Distribution