Effects of Lockdown and Post Lockdown on Covid19 cases across India using Supervised Learning Techniques

Narayana Darapaneni
Director - AIML
Great Learning/Northwestern University
Illinois, USA
darapaneni@gmail.com

Amol Kobal
Student - AIML
Great Learning
Pune, India
kobal.amol@gmail.com

Rohit Chaoji
Student - AIML
Great Learning
Pune, India
rohitchaoji@gmail.com

Rajiv Tiwari
Student - AIML
Great Learning
Pune, India
rztiwari@gmail.com

Suman Saurav
Student - AIML
Great Learning
Pune, India
sumansaurav1987@gmail.com

Anwesh Reddy Paduri
Research Assistant - AIML
Great Learning
Pune, India
anwesh@greatlearning.in

***Abstract*— The Coronavirus (or Covid-19) outbreak is a major global pandemic, which has infected millions of people across the world. India has been one of the worst hit country by this pandemic with over 4.5 million people affected by the virus as of 10th Sept 2020. The governments in India took various methods to contain the outbreak including non-medical interventions like lockdown. This document provides detailed analysis of effect of lockdown in containing the spread of Covid-19 in India. This work comprises of applying different statistical techniques to comprehend the Virus progression during various phases of lockdown implemented across India.**

Keywords—Covid-19, lockdown, unlock, supervised learning, infection rate

I. INTRODUCTION

The Coronavirus Disease 2019 (Covid-19) outbreak was first detected in Wuhan district, China in December 2019 [1]. Over the course of eight months, it has spread to more than 200 countries [2]. In the early stages, countries without proper lockdown restrictions, such as Italy, Iran and the United States were badly affected by this virus [3]. Covid-19 spreads from person-to-person via aerosols such as respiratory droplets [4]. On March 11, Word Health Organization (WHO) declared Covid-19 as a pandemic and urged countries to take measures to contain its spread [5].

In an effort to contain the spread of Covid-19, India implemented a lockdown with strict social distancing norms [6]. Implementation of the lockdown, as well as adherence to social distancing showed positive results in containing the spread of the virus [7].

Existing studies on the subject show an overall reduction of transmission rate resulting from lockdown strategies implemented in various parts of the world [8,9,10,11]. India implemented 3 phases of strict lockdown, however, during the fourth phase of the lockdown implemented by the Government of India, restrictions were eased in certain regions. Beyond that India has started the unlock phases to bring the economy back on track. Available data can be used to analyse the effect of easing lockdown restrictions on the spread of Covid-19.

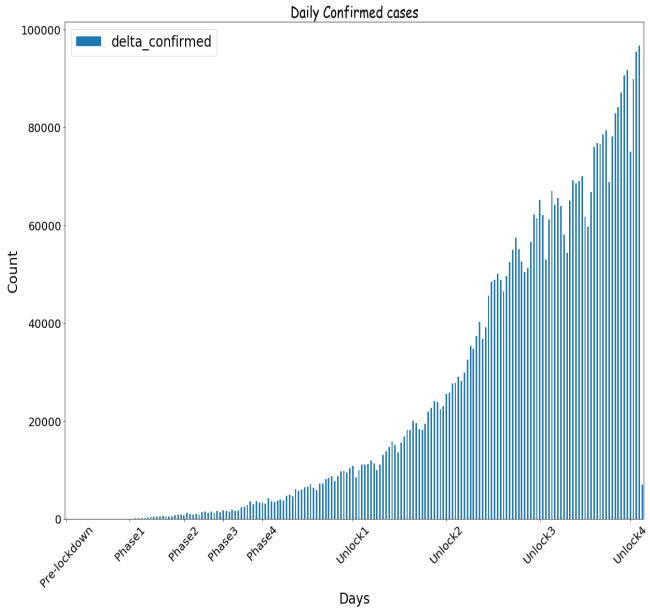

FIGURE 1. The number of Daily Confirmed Cases (DCC) across India over the period of last 190 days. X-axis is days since start 5th March 2019, Y-axis is daily confirmed cases [12].

In this study we try to use available data for daily confirmed cases with exploratory and feature engineering techniques to identify the infection rate (I_R), defined as - Percentage change of confirmed cases over daily testing done. The study will identify the impact if various phases of the lockdown on the rate of increase of infection rate in India. Our results can be used by governments across the world to understand and implement the lockdown measures to contain the spread of highly infectious diseases like Covid-19.

II. MATERIALS AND METHODS

A. Data collection

For the purpose of our research we wanted a continuous and reliable source of time series data and the covid19india.org website provided the perfect data source which has been collated as a crowdsourced initiative [13].

The data has been collected and organized by volunteers who gather data from various bulletins and other official handlers. Below is the exposed application protocol interface (API) which has been used for the research.

https://api.covid19india.org/v4/timeseries.json

The API exposes the JavaScript object notation (JSON) [14] data for each state/ union territories and the country wise total and daily confirmed, tested, recovered and deceased cases. Data is available for each date 30-Jan-2020 to till date. We see that cases are increasing every day.

For the purpose of our research we have taken the columns relevant for our analysis which includes Date, Total Confirmed, Daily Confirmed, Total Recovered, Total Tested, Daily Tested as shown in the table extract below:

TABLE1.Sample of the data showing the Total Confirmed, Recovered and Tested cases reported across India starting from March-04-2020.

Date	Total Confirmed	Total Recovered	Total Tested
04 Mar	28	3	2842
05 Mar	30	3	3248
06 Mar	31	3	3654
07 Mar	34	3	4060
08 Mar	39	3	4466
09 Mar	48	3	4872
0 Mar	63	3	5278
11 Mar	71	3	5684
12 Mar	81	3	6090
15 Mar	91	3	7194
15 Mar	102	3	8298

This data can be taken to build other features like Daily Confirmed and Daily Tested cases which can help in analyzing the infection rate over a period of time.

The rough analysis of data over the period of time suggests a continuous growth of cases, with the initial phase suggesting a high rate of growth, while later the growth became more consistent. For the purpose of statistical analysis we had to do some approximation of data for the initial phase when the data collection process was crude and was collated only after a few days. We averaged out the data for such phases to provide a better approximation of analysis. All these approximations are on the real data set and hence averaging out the data during the phase would not have any significant effect on the analysis.

Besides the data on the number of cases we also need to consider the lockdown phases in India for our statistical analysis. The various phases of lockdown in India as defined by the government of India are as follows [15]:

1. Phase 1 - 23 March to 14 April'20
2. Phase 2 - 15 April to 3 May'20
3. Phase 3 - 4 May to 17 May'20
4. Phase 4 - 18 May to 31 May'20
5. Unlock 1.0 - 1 June to 30 June'20
6. Unlock 2.0 - 1 July to 31 July'20
7. Unlock 3.0 – 1 August to 31 August'20
8. Unlock 4.0 – 1 September to 30 September'20

These phases are used as the timelines of analysis to perform statistical testing to conclude on effect of the lockdown in the process to curb the growth of infection.

B. Statistical Methods

With the formatted data-set we do statistical tests to perform hypothesis testing of samples over various phases to statistically conclude the effect of lockdown.

T-test: In this case, we are testing whether implementation of the lockdown led to any significant changes in mean growth rate of the spread, so we will apply t-tests on sample means over different phases of the lockdown. The test statistic in a two-sample t-test is given by following formula [16]:

$$T = \frac{Y1 - Y2}{\sqrt{\left(\frac{S1^2}{N1} + \frac{S2^2}{N2}\right)}} \quad (1)$$

Y1, Y2 = Means over corresponding phases

s1, s2 = Variances over corresponding phases

N1, N2 = Number of samples in corresponding phases

If the absolute value of test-statistic exceeds the critical value, we should reject the null hypothesis stating that there is a significant difference in the sample means.

C. Graphical Representation and Exploratory Data Analysis

When dealing with medium-to-large datasets, it is helpful to visualize it in order to quickly find trends and patterns in data at a glance. For this purpose, Python packages such as matplotlib, Seaborn and Plotly were used.

Bar Graph: A bar graph can be used to display categorical data over a period of time – display time series data [17]. We can also plot multiple categorical features on the graph to compare the features and changes over a period of time. Covid-19 cases in India can be plotted to gain more information and visualize the cases - tested, recovered and reported across India.

Line Graph: A line graph can be used to display categorical feature values over a period of time [18]. We can use a line graph to represent the categorical values like confirmed cases, testing done, percentage change in confirmed cases and testing done. The line plots can visually present a clean representation of these features.

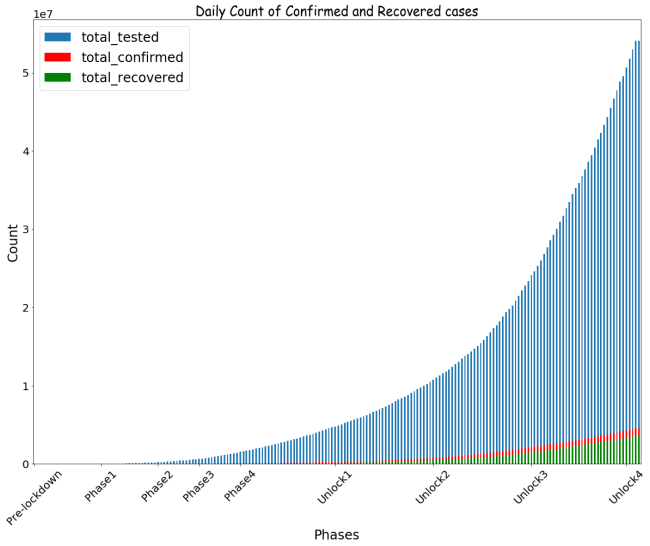

FIGURE2.Total Testing, Recovered and Confirmed cased over period of 190 days starting from 04-March including various phases.

As evident from the graph the testing across the country has been rapidly ramped up over the period of time however the number of confirmed cases have also seen a steep rise. Amidst this the recovery has also seen a great improvement over past months.

The data provides basic information about the number of reported cases across India but this does not help to achieve any conclusion about the effect of lockdown and unlock phases. The total number of cases is deemed to rise daily over a period of time. We need to extract more information out of this data to reach to any conclusion about the effects of lockdown.

D. Feature Selection

From the data downloaded from the API we can see the following structure of the data features:

1. Date of Record
2. Number of Confirmed Cases on Date
3. Number of Tests Conducted on Date
4. Total confirmed cases
5. Total testing done

These features give actual progression of cases, however to check the underlying facts we need to excavate more features via feature engineering. We have used simple mathematics to build new features for our analysis:

1. Daily Confirmed cases (DCC): Change in total confirmed cases daily.

$$DCC = TCC - TCC_p \quad (2)$$

2. Daily Tested cases (DTC): Change in total testing on daily basis

$$DTC = TTC - TTC_p \quad (3)$$

3. Percentage Rate of Increase of Testing (PrIT): Change in testing on daily basis

$$PrIT = \frac{DTC}{DTC_p} \quad (4)$$

4. Percentage Rate of Increase of Confirmed cases: Daily increase over Daily increase of previous day

$$PrIC = \frac{DCC}{DCC_p} \quad (5)$$

DCC = Daily confirmed cases

DCC_p = Daily confirmed cases till previous day

TCC = Total confirmed cases

TCC_p= Total confirmed cases till previous day

DTC = Daily testing conducted

TCC = Total testing conducted

TTC_p= Total testing conducted till previous day

PrIT = Percentage Rate of Increase of Testing

DTC_p= Daily testing conducted till previous day

PrIC = Percentage rate of increase of confirmed cases

The newly created features provide powerful mechanism for statistical analysis and tests.

TABLE2.Provides the Testing percentage change and confirm cases percentage change.

Date	Total Confirm	Total Tested	Confirm Percentage change	Test Percentage Change
04 Mar	28	2842	0.00	0.00
05 Mar	30	3248	7.14	14.29
06 Mar	31	3654	3.33	12.50
07 Mar	34	4060	9.64	11.11
08 Mar	39	4466	14.70	10.00
09 Mar	48	4872	23.07	9.09
0 Mar	63	5278	31.25	8.33
11 Mar	71	5684	12.70	7.69
12 Mar	81	6090	14.08	7.14
15 Mar	91	7194	12.35	18.12
15 Mar	102	8298	12.08	15.35

The sample data of percentage change in confirm cases and percentage change in testing has been captured in the above table. This data has been built across by the base data set by applying mathematical formulas as mentioned previously. This process of building new features for analysis using the existing data features and applying the domain knowledge is termed as “Feature Engineering” [19]. This enterprise feature provided us with new outcome for different phases of the lockdown. To understand the features it is worth to plot a graph of the new features “Rate of

change of confirmed cases over testing done daily" to get better insights.

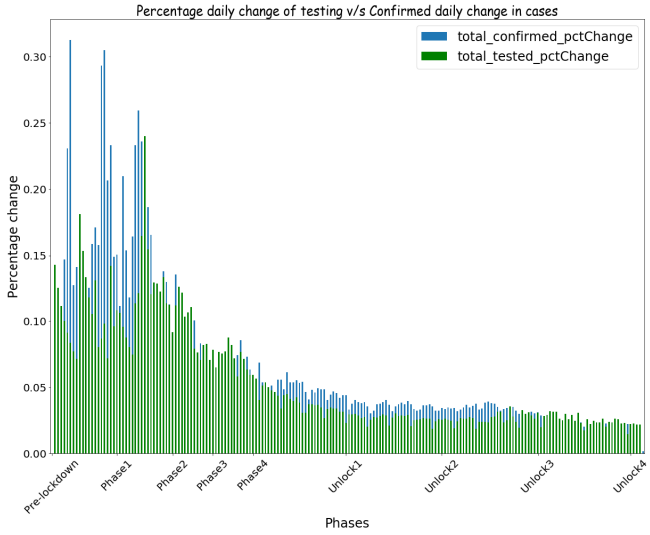

FIGURE3. Plotting the percentage change of testing and the confirmed cases.

The graph in the FIGURE3 above gives interesting insights about the rate of increase of cases over the testing. The rate of change increased substantially during the pre-lockdown and the first stage of lockdown. However, the virus has been found to have an incubation period of 12 days and any change will need about 2-3 weeks to show effect [20]. Analyzing the above graph we can see that the rate of increase of infection was pity high and was in an upward trend during the first 40 days and later on decreasing gradually to near constant rate of growth, thus providing an insight that the strict lockdown did have an effect on rate of growth of spread of the virus.

The percentage change though seems to be decreasing at a slow pace but the positive percentage change of graphs does not conclusively say anything about the trend of growth rate of the virus. The data above does not conclude to the results evidently, however applying a further mathematical computation on above data can provide us a conclusive pattern

E. Feature Engineering

The process of applying certain mathematical equations on existing dataset can produce new features in the data which is termed as Feature Engineering [19]. The above dataset has provided the percentage change of confirmed cases and testing done over a period of time. We can try creating a new using exiting features - percentage change of cases and testing done on daily basis.

Infection Rate (I_R): Percentage increase of Confirmed cases over Percentage increase of Testing. The infection rate could be a powerful feature to comprehend the growth of infection amongst the mass as both testing and number of cases are increasing daily. A positive trend in the infection rate would mean that the growth rate is increasing and a negative trend would conclude that the growth rate is decreasing.

$$I_R = \frac{PrIC}{PrIT}$$

I_R = Infection Rate

PrIC = Percentage rate of increase of confirmed cases

PrIT = Percentage Rate of Increase of Testing

TABLE3. Provides the sample of infection rate of the virus.

Date	Confirm Percentage change	Tested Percentage Change	Infection Rate (I_R)
04 Mar	0.00	0.00	0.00
05 Mar	7.14	14.29	0.49
06 Mar	3.33	12.50	0.24
07 Mar	9.64	11.11	0.73
08 Mar	14.70	10.00	1.23
09 Mar	23.07	9.09	2.22
0 Mar	31.25	8.33	3.69
11 Mar	12.70	7.69	1.97
12 Mar	14.08	7.14	.091
15 Mar	12.35	18.12	0.91
15 Mar	12.08	15.35	1.00

We can see that during the initial phase the infection rate of the virus was very high. The above data in the table provides the infection rate during the pre-lockdown phase. Plotting the infection rate across the various phases of lockdown gives some interesting results.

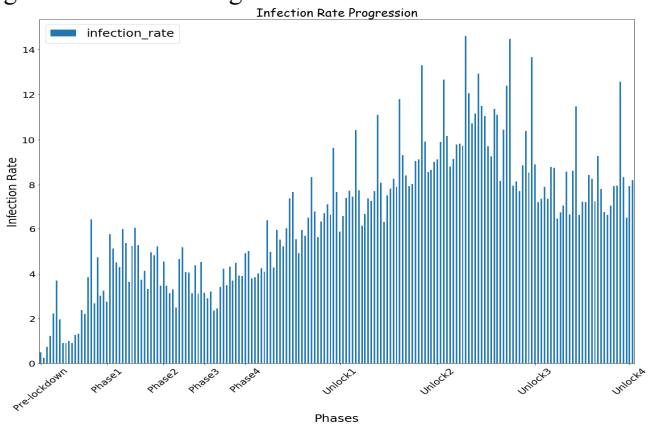

FIGURE4. Bar graph of daily infection rate.

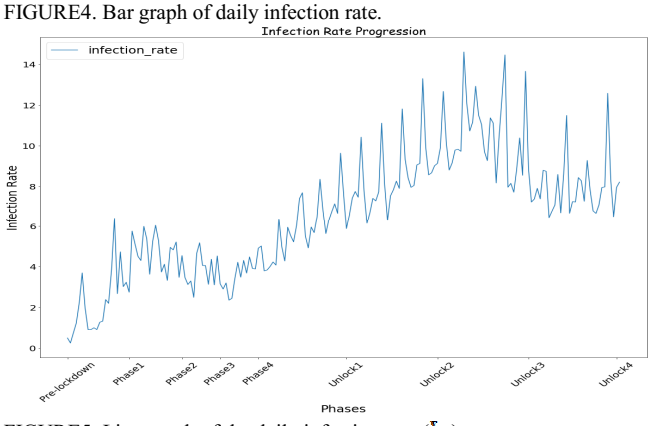

FIGURE5. Line graph of the daily infection rate (I_R)

The graphs above FIGURE4 and FIGURE5 provides a great deal of details on the growth rate of the virus. Analysing the infection rate change we can distinctly see that the graph had a steep rise during pre-lockdown with a decline during the strict lockdown phase followed by a consistent rise of infection rate.

We can can also see that during the unlock phase 3 and 4 there has been a slight decline in the infection rate.

III. Results

Performing structured approach of data analysis with inclusion of feature engineering techniques provided some deep insights on progression of infection rate across the country. From FIGURE4 and FIGURE5 we can conclude the following:

- The infection rate increased steeply during the pre-lockdown and the 1st phase of lockdown.
- During the 2nd and 3rd phase we can see that there was a steep decline in the infection rate and reached the lowest level around 1st week of May. We can observe that the strict lockdown phase was in place by around up to Phase 3 of the lockdown.
- Starting from Phase 3 of lockdown and continuing into Phase 4 and unlock phases we can see that there is a steep rise in the infection rate
- We can also observe a decline in later unlock phases. During the phase 3 and 4 of unlock we can see a decline in the infection rate

These observations needs to be backed up with proper statistical tests. Plotting the average infection rate change between different phases of lockdown to the corresponding phase will give the changes in p-value and t-value of the tests which can confirm the magnitude of change and the direction of change [21].

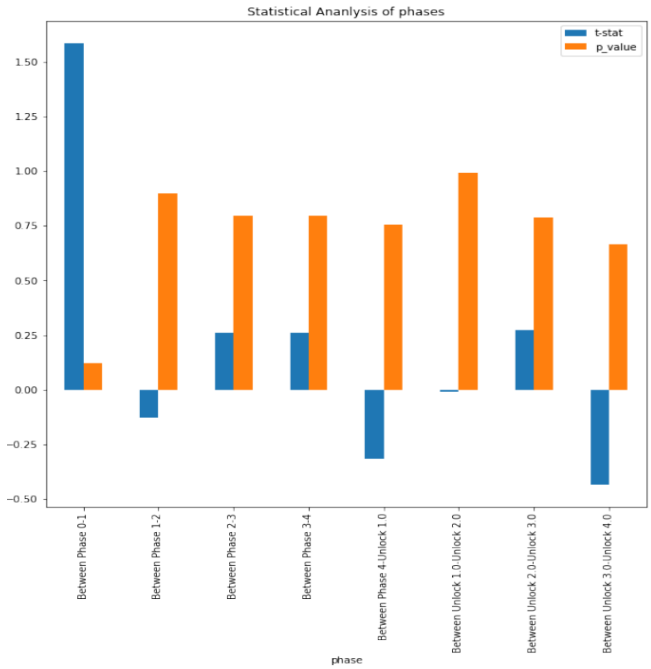

FIGURE6. Bar graph showing the p-value and t-value over different phase changes.

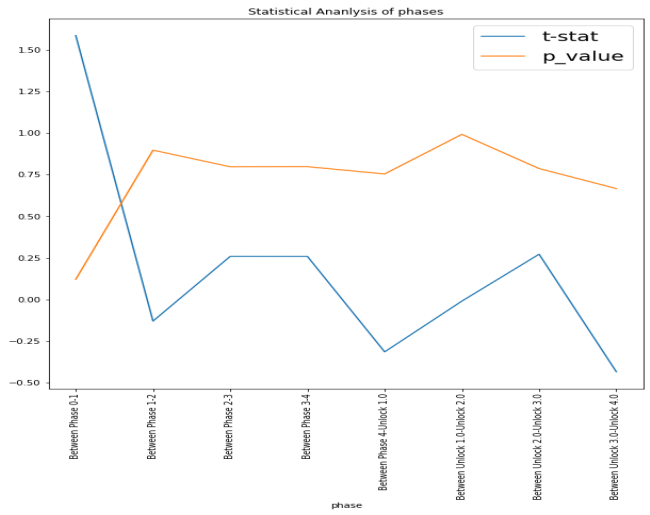

FIGURE7. Line graph showing the p-value and t-value over different phase changes.

Checking the graphs we see that p-value is significantly high between each phase signifying that the changes in either direction is significant during each phase. We can observe that t-stat value dips to negative values during phases 1 of lockdown and the last unlock phase (Unlock 3.0) and remains consistently positive during the other phases.

We can also observe that there is a significant dip in t-value below 0 during Phase4 and Unlock 1.0. This needs a closer observation in relation to percentage of testing and confirm case changes plot as per FIGURE3. We can see that during that phase there was significant jump in the testing rate change which might have caused the infection rate to lower down as the ratio will be effected if one parameter suddenly changes significantly.

IV. Discussion and Conclusion

A. *Discussion on the results*

The lockdown phases in India started with strict restriction in movement of people and resources during the first two phases and was increasingly relaxed during the next phases. The strict lockdown phase puts the country into a complete lockdown with all the activities and services including religious places, restaurants, gyms cinema halls, work places, industries etc, put into complete locked state. India implemented strict lockdown for phase 1 and 2 with minor relaxation for essential services like grocery and medicines. The unlock phase saw more ease provided and various activities including markets, worship places etc. being opened with light restrictions. The restrictions were lifted in a phased manner.

The initial phase where infection rate was very high can be attributed to lack of awareness within general mass about the practice of social distancing norms. We can also see that immediately after the strict lockdown implementation, the effect took about a couple of weeks to show due to the incubation period of the virus.

The steep decline in the infection rate during phase 2 and phase 3 of lockdown signifies that the strict lockdown did help in containing the spread of the virus. Also as the study

infers that the opening of restrictions during later phases increases the spread of the virus.

We can also see that during the Unlock 3.0 phase there has been a significant decline in the infection rate which could have been due to increase in general public awareness like following social distancing and compulsory wearing masks in the public places. Some local governments have also implemented strict laws like restricting the timing and days to open the markets which may have also contributed to the factor

B. Limitations in the study

Though the data collected and analyzed are verified data but the impact of the results could also be attributed to other factors performed by local governments. The study strictly does the analysis of infection rate growth based on lockdown mechanisms at the PAN India level and does not considers these local factors in place.

Besides the local governments have been taking many measures to curb the spread. In vast country like India the various local factors could affect the results. Our study is strictly based on PAN India level and specific to the effects of lockdown phases.

C. Conclusion

In the current times when the global economies are diminishing under the impact of pandemic, the only effective strategy to curb the growth rate of infection is performing social distancing and lockdown mechanism. We can see that strict lockdown (phase 1 and 2) did have significant impact in curbing the growth but removing lockdown saw the infection rate bounce back. The strict lockdown saw complete halt in movement of people across the country except the essential services like grocery and medical services. Since the movement stopped completely during this phase it coincides with the steep drop in the infection rate as evident in the FIGURE5. The period between Phase-1and Phase-2 was the period of strict lockdown in India during which we can see the infection rate fall down to the lowest levels.

However when the strict lockdown was relaxed we can observe that with start of movement of people, the infection rate jumped. Between Phase-3 to Unlock-2 the 3 months of relaxed lockdown and unlock phases the infection rate jumped on daily basis as can be validated in FIGURE5.

We can see that post 3 months of steady rise, the infection rate fell down a bit and then remained steady at a specific rate. This could be attributed to few factors like people acquiring immunity due to previous exposure to the virus; also factors like general improved awareness amongst the masses on the possible cause of spread of virus and precautions could have played significant role. The governments in India also implemented certain mandatory rules like compulsorily wearing face masks in public and limiting the number of people in gatherings which could have been a factor in curbing the infection rate.

We see from the outcomes of analysis that measures like public awareness of social distancing and mandatory face mask can be an effective measure. As analyzed from the Unlock 3.0 phase if the governments perform strict lockdown or general public awareness about social distancing, it can restrict the spread of the virus. The governments can thus open up the economic activities in phased manner performing and practicing general social distancing rules which can help bring economies on track and also limit the spread of Covid-19 virus.

REFERENCES

[1] Wikipedia contributors, "COVID-19 pandemic," *Wikipedia, The Free Encyclopedia*, 29-Sep-2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=COVID-19_pandemic&oldid=980878236. [Accessed: 29-Sep-2020].

[2] "Coronavirus update (live): 33,574,347 cases and 1,006,868 deaths from COVID-19 virus pandemic - worldometer," *Worldometers.info*. [Online]. Available: https://www.worldometers.info/coronavirus/. [Accessed: 29-Sep-2020].

[3] *Who.int*. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200327-sitrep-67-covid-19.pdf?sfvrsn=b65f68eb_4. [Accessed: 29-Sep-2020].

[4] "How does Coronavirus spread?," *Webmd.com*. [Online]. Available: https://www.webmd.com/lung/coronavirus-transmission-overview. [Accessed: 29-Sep-2020].

[5] "WHO announces COVID-19 outbreak a pandemic," 2020.

[6] BBC News, "Coronavirus: India enters 'total lockdown' after spike in cases," *BBC*, 25-Mar-2020.

[7] V. Moorthy, A. M. Henao Restrepo, M.-P. Preziosi, and S. Swaminathan, "Data sharing for novel coronavirus (COVID-19)," *Bull. World Health Organ.*, vol. 98, no. 3, p. 150, 2020.

[8] "Potential Long-Term Intervention Strategies for COVID-19," *Stanford.edu*. [Online]. Available: http://covid-measures.stanford.edu/. [Accessed: 29-Sep-2020].

[9] *Who.int*. [Online]. Available: https://www.who.int/emergencies/diseases/managing-epidemics-interactive.pdf. [Accessed: 29-Sep-2020].

[10] *Researchgate.net*. [Online]. Available: https://www.researchgate.net/publication/340598567_Impact_of_intervention_on_the_spread_of_COVID-19_in_India_A_model_based_study. [Accessed: 29-Sep-2020].

[11] CDC, "Coronavirus Disease 2019 (COVID-19) - Transmission," *Cdc.gov*, 21-Sep-2020. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprepare%2Ftransmission.html. [Accessed: 29-Sep-2020].

[12] "Coronavirus in India: Latest map and case count," *Covid19india.org*. [Online]. Available: https://www.covid19india.org/. [Accessed: 29-Sep-2020].

[13] *Covid19india.org*. [Online]. Available: https://www.covid19india.org/about. [Accessed: 29-Sep-2020].

[14] "JSON," *Json.org*. [Online]. Available: https://www.json.org/json-en.html. [Accessed: 29-Sep-2020].

[15] Wikipedia contributors, "COVID-19 pandemic lockdown in India," *Wikipedia, The Free Encyclopedia*, 27-Sep-2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=COVID-19_pandemic_lockdown_in_India&oldid=980645019. [Accessed: 29-Sep-2020].

[16] "Unpaired (Two Sample) t Test," *Statsdirect.co.uk*. [Online]. Available:https://www.statsdirect.co.uk/help/parametric_methods/utt.htm. [Accessed: 29-Sep-2020].

[17] Wikipedia contributors, "Bar chart," *Wikipedia, The Free Encyclopedia*, 14-Sep-2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Bar_chart&oldid=978334079. [Accessed: 29-Sep-2020].

[18] Wikipedia contributors, "Line graph," *Wikipedia, The Free Encyclopedia*, 21-Sep-2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Line_graph&oldid=979503485. [Accessed: 29-Sep-2020].

[19] Wikipedia contributors, "Feature engineering," *Wikipedia, The Free Encyclopedia*, 21-Sep-2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Feature_engineering&oldid=979583744. [Accessed: 29-Sep-2020].

[20] J. A. Backer, D. Klinkenberg, and J. Wallinga, “Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020,” *Euro Surveill.*, vol. 25, no. 5, 2020.

[21] “P Values,” Statsdirect.com. [Online]. Available: https://www.statsdirect.com/help/basics/p_values.htm. [Accessed: 29-Sep-2020].