# An Investigatory Data Analysis of COVID 19 India Data: With Reference to First Three Phases of Lockdown in India

**E.Prabhakar[1], T.M.Jacob[2], Aman Khurana[3], N.Santhiya[4]**

[1]*Assistant Professor, Department of Computer Science and Engineering,*
*Nandha College of Technology, Erode, Tamil Nadu.*
[2]*Associate Professor, Department of Statistics, Nirmala College, Muvattupuzha, Kerala.*
[3]*Applied Data Scientist, Dunnhumby, Gurgoan, Haryana.*
[4]*UG Scholar, Department of Information Technology,*
*Nandha College of Technology, Erode, Tamil Nadu.*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract -** *COVID-19 is a transferrable disease and its origin was first reported in December 2019. The World Health Organisation (WHO) had declared it as a pandemic by the in March 2020. As of now 28th May 2020, more than 5.69 million people have been tested positive. The number of infected Corona Virus cases in the world is increasing rapidly. WHO and Governments putting tremendous efforts to fight the disease. Experts are finding it very tough to understand the pattern and predict the virus spread in the world. The primary objective of this research work is to analyze what happened in India during the first 3 phases of lockdown. We have drawn a statistical model to understand the virus spread pattern in India. The investigation results identify the effect of COVID-19 in India on weekly and lockdown fashion. Exploratory data analysis is performed on COVID 19 India data. This paper statistically analyzed statewide data. It has drawn statistical model for the percentage of variation between states on the reported numbers of the recovered, mortal, cured patients and samples tested. The analysis of this data will help us to identify the areas of improvement and making better decisions in upcoming days.*

*Key Words*: COVID19, Corona Virus, Exploratory Data Analysis, Statistical Model, WHO, COVID 19 India.

## 1. INTRODUCTION

Coronavirus disease 2019 (abbreviated as COVID-19) is a transferrable disease that put the entire world in the struggle. It is mainly caused due to the virus named severe respiratory syndrome coronavirus 2 (SARS-CoV-2). This virus was first reported in December 2019 in Wuhan, China. After that it witnessed a global spread, leading to an ongoing world pandemic [17], [18]. As of 19th May 2020, greater than 4.81 million cases have been identified across 188 nations and territories. It resulted in much higher than 318,000 deaths and greater than 1.78 million people have recovered from the disease.

According to the WHO, no vaccines/treatments are available for COVID-19. The WHO proclaimed the COVID 19 to be a Public Health Emergency of International Concern (PHEIC)on 30th January 2020. On 11th March 2020, WHO announced COVID19 as a pandemic.

India is the second-largest nation in the world in terms of population. So, it is important that India efficiently handles this pandemic. On 30th January 2020, the first COVID-19 case was identified in India China was the origin of this virus. As of 19th May 2020, the Ministry of Health and Family Welfare has reported an overall 101,139 confirmed cases. 39,174 people recovered from the disease, while 3,163 deaths happened in the country.

The COVID 19 Pandemic has been proclaimed as an epidemic in more than 10 states and union territories. Conditions of the Epidemic Diseases Act, 1897 have been evoked in these states and union territories. All academic institutions and usual companies have been locked down, except necessary shops like grocery and medical shops[19], [22].

On 22nd March 2020, a 14-hour voluntary public curfew was observed by entire India to handle the Corona Virus outbreak. On that lockdown happened in 75 districts where the COVID-19 cases had been reported along with other metro cities. Then on 24th March, the prime minister of India issued the order for a nationwide lockdown for 21 days. On 14th April, the central government extended the nationwide lockdown for another two weeks till 3 May. Again on 1st May, the lockdown was further extended till 17 May[22], [23].

## 2. LITERATURE SURVEY

Nadia AL-Rousan and Hazem AL-Najjar [1] studied the present development of the spread of CoVID-19. Their study identified a correlation between independent variables and dependent variable. The article provided a statistical analysis of the impact of sex, location, causes for infection, year of birth, and released or disease contracted date on the reported numbers on the cured and succumbed cases.

T.Dinesh Kumar et al., [2] proposed a new approach to handle big data. They performed exploratory data analysis on airline services big data. Authors in [3] introduced a new mechanism to identify public opinion mining. This paper analysed Government Scheme Advertisement.

Cleo Anastassopoulou et al., [4] analysed the current situation and stated that accurate reasons based on the prediction of the dynamics of the spread simply cannot be attained at this phase. Also, they pointed out the unpredictability of accessible official data will lead to false forecasts.

Samrat K. Dey et al., [5] presented an attempt at the compilation and analysis of epidemiological outbreak information on COVID-19 using the multiple public datasets on 2019-nCoV obtained from the Johns Hopkins University, World Health Organization, Chinese Center for Disease Control and Prevention, National Health Commission, and DXY. An investigatory data study with visual images has been carried out to interpret the number of various cases identified(confirmed, deceased, and cured) in various districts of China and outside China.

In [6], [8], [11], [12], the authors analysed the available data with in-depth exploratory data analysis. And their results have shown the importance of exploratory data analysis for analyzing any data, before applying machine learning techniques.

Sarvam Mittal [7] drawn a statistical framework for getting to know the spread of COVID-19 in India better through a rigorous analysis of the cases reported in the country till 22 April 2020. An Exploratory Data Analysis (EDA) method is being realized for the research and analysis of the COVID-19 cases reported in India.

Davit Gondauri et al., in [9] studied the chain-binomial type of the framework which consists of brief phases of high infectivity and nearly consistent incubation periods. This research work studies and analyses the statistics on the COVID-19 Virus spread using the examples of the cases obtained from various nations.

Rajan Gupta et al. [10] studied six different aspects. They correspond to depicting the growth trends of the number of infected patients in India, predictions made for the number of infected patients for days to come, the effect of social distancing on the Indian citizens, the effect of mass events on the number of infected patients in India, network study and mining of patterns on the individuals suffering from the disease, and reviewing the mechanics for lifting the lockdown in India.

Ahmad Yame [13] gathered all kinds of Data Sources from the current researches and studied about Coronavirus Disease (COVID-19)around the world. Then he discussed data collection planning by analysing on the extent to which the available data can be utilized.

In papers [14], [15]. [16]authors studied and analyzed the various exploratory data analysis using various statistics for the given data.

## 3. EXPLORATORY DATA ANALYSIS

The present COVID-19 pandemic inspires to carry out an EDA on the datasets, acquired from various resources like "Ministry of Health and Family Welfare" [6], "COVID-19 India website" [7], "John Hopkins GitHub repository" [8], "Worldometer" [9] and "Wikipedia" [10] with the help of "Python" and thereby evaluating the spread and trend followed by COVID-19 in India and comparing it with the neighbouring and worst suffering world nations.

Dataset is taken from
- ✓ Kaggle repository[20], [21]
- ✓ https://www.mohfw.gov.in/ [24]
- ✓ https://www.covid19india.org/ [25]
- ✓ https://github.com/CSSEGISandData/COVID-19 [26]
- ✓ https://www.worldometers.info/coronavirus/country/india/ [27]

Dataset consists of various parameters like number of samples tested, confimed cases, cured, death count.
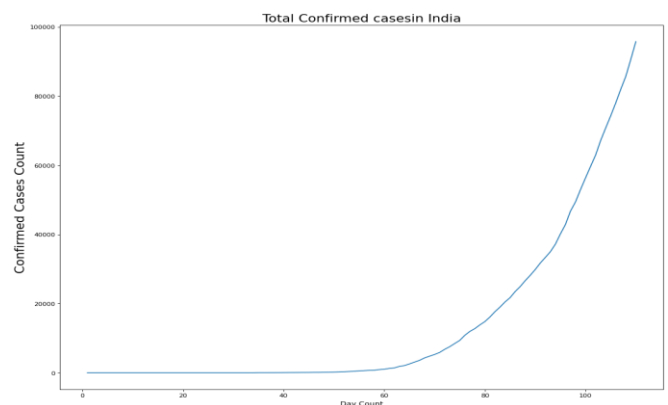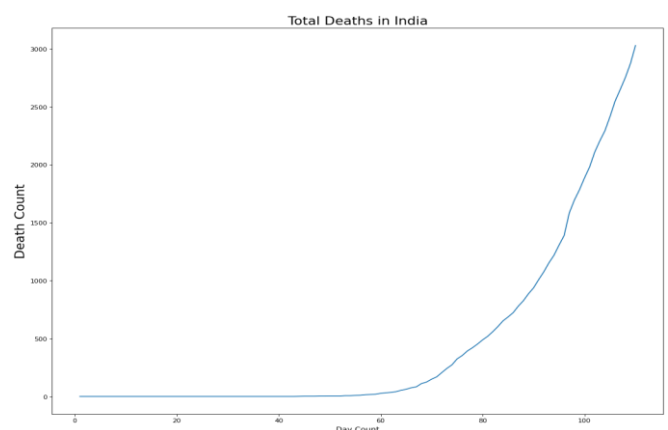


***Fig. 1.*** *Day by Day Confirmed Cases*



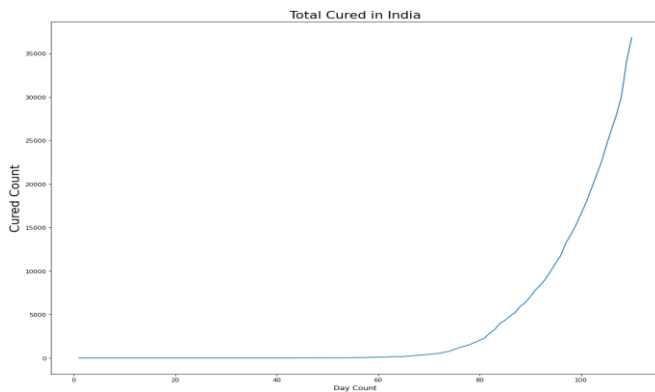***Fig. 2.*** *Day by Day Death Count*

**Fig. 3.** *Day by Day Cured Count*

Kerala is the first state in India, witnessed the first COVID 19 positive case on January 30. From that date to march last week, the curve is a little bit flatter. But after that, a number of confirmed cases increasing at a rapid pace. It continues to be steeper still this point of time. The cured count also increasing at a good pace. It is a positive side in many aspects.
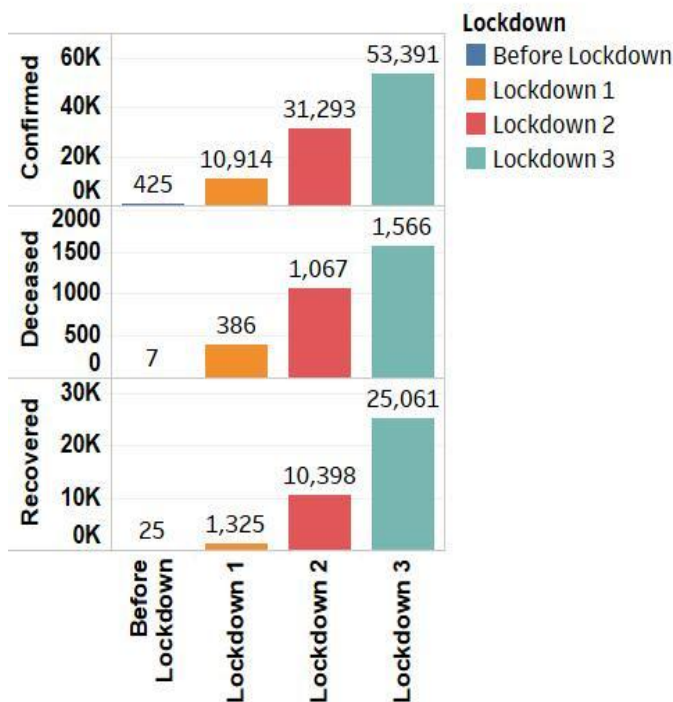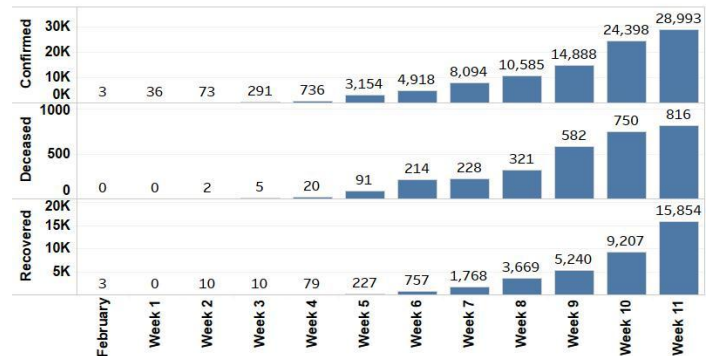


**Fig. 4.** *Confirmed, Deceased and Recovered Count based on Lockdown*



**Fig. 5.** *Weekwise Confirmed, Deceased and Recovered Count*

The figure 4 shows the cases confirmed, deceased and recovered for each lockdown. Confirmed cases and recovered count increased in faster rate in each phase. The number of people deceased ranges from 7 to 1,566. Figure 5 represents confirmed cases, deceased and recovered for each week. Confirmed ranges from 3 to 28,993. Deceased ranges from 0 to 816. Recovered ranges from 0 to 15,854.
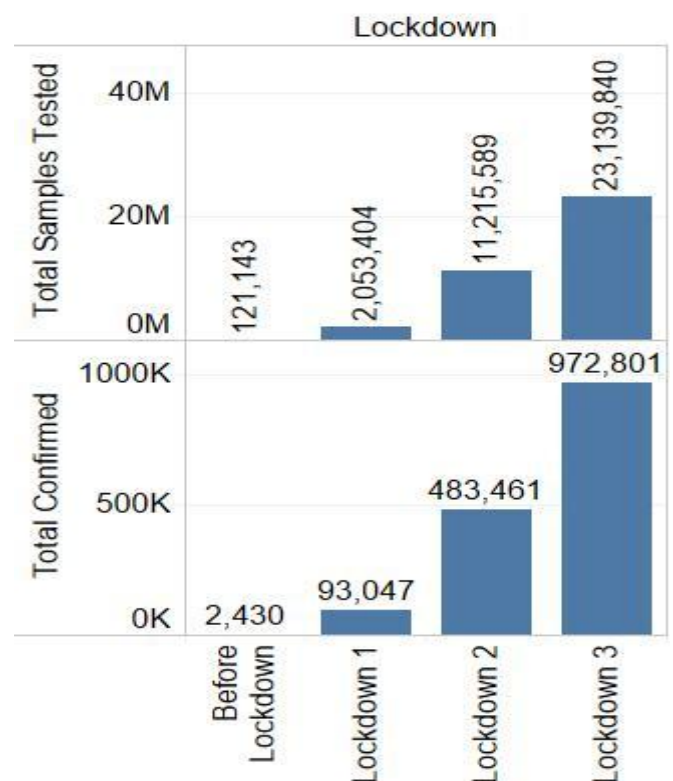


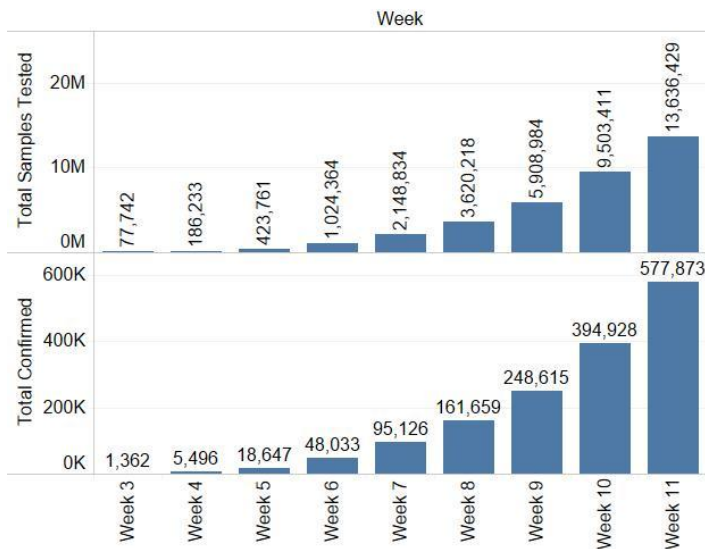**Fig. 6.** Confirmed, Tested Count based on Lockdown

**Fig. 7.** Weekwise Confirmed, Tested Count

Figure 6 shows the total samples tested and the total confirmed for each lockdown. Figure 7 shows the total samples tested and the total confirmed for each week.
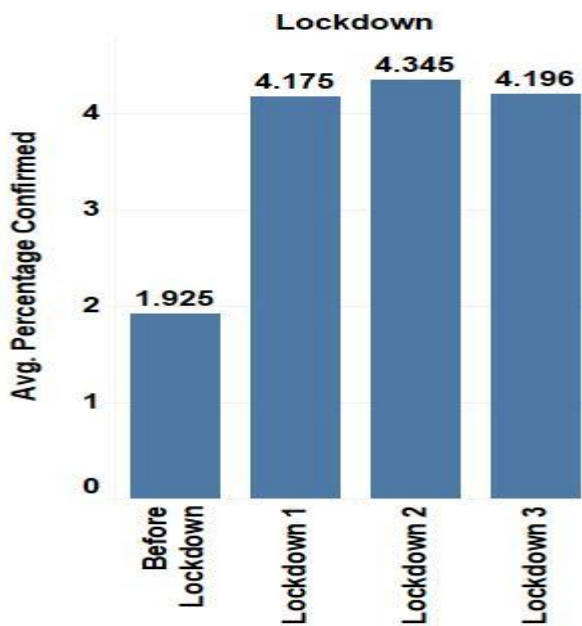


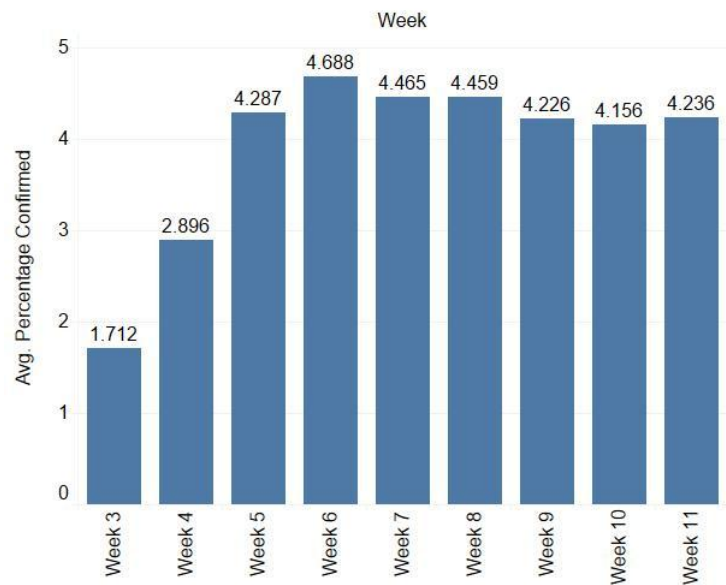**Fig. 8.** Percentage Confirmed - Lockdown



**Fig. 9.** Weekwise Percentage Confirmed

Figure 8 shows the average percentage confirmed for each lockdown. Average of Percentage Confirmed ranges from 1.925 to 4.345 in Figure 8. Figure 9 represents an average of Percentage Confirmed for each Week. Average of Percentage Confirmed ranges from 1.712 to 4.688.
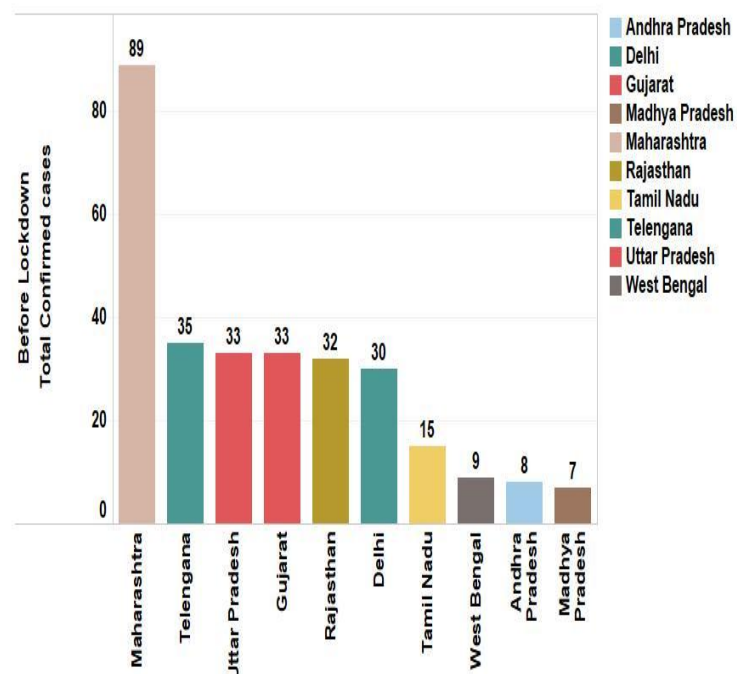


**Fig. 10.** Top 10 States having Higher Number of Cases before Lockdown

Before lockdown itself, Maharastra having the highest number of positive cases. It may be due to the density of population and more International movements.
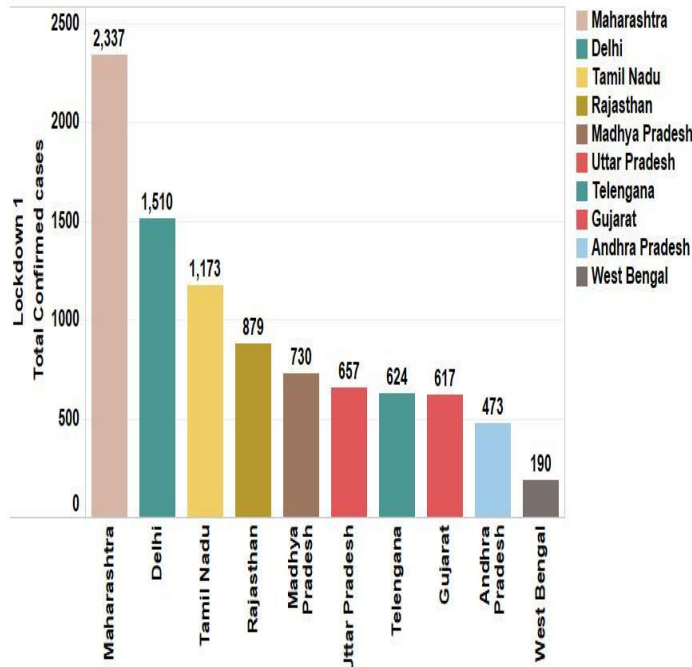


**Fig. 11.** Top 10 States having Higher Number of Cases during Lockdown 1

During lockdown 1, Tamilnadu moved from 7 to 3 in terms of the highest number of confirmed cases.
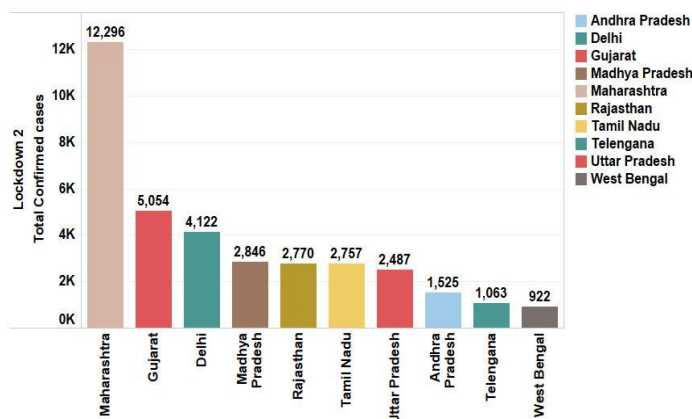


**Fig. 12.** Top 10 States having Higher Number of Cases during Lockdown 2

During lockdown 2, Tamilnadu started to reduce the number of confirmed cases, while Gujarat, Madhya Pradesh and Rajasthan has seen increasing numbers.
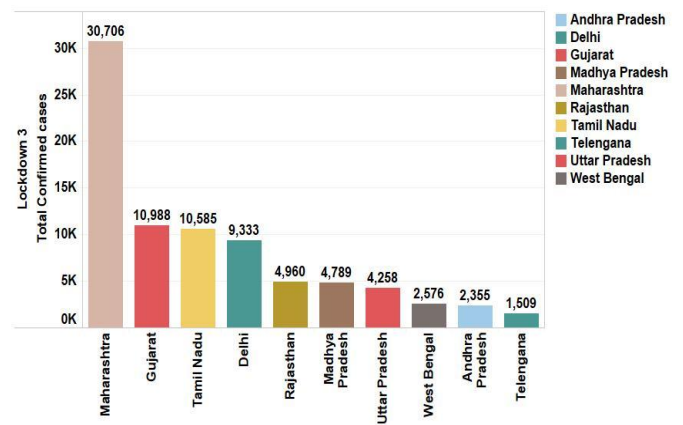


**Fig. 13.** Top 10 States having Higher Number of Cases during Lockdown 3

In each phase of lockdown Maharsatra not able to control the confirmed cases. It becomes epicenter of confirmed cases in India.
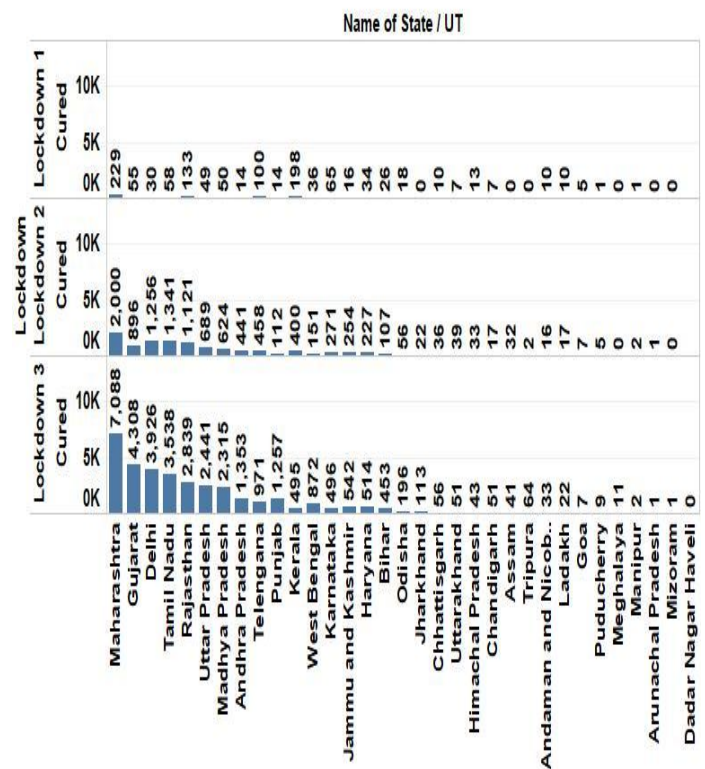


**Fig. 14. Statewise Cured Data**

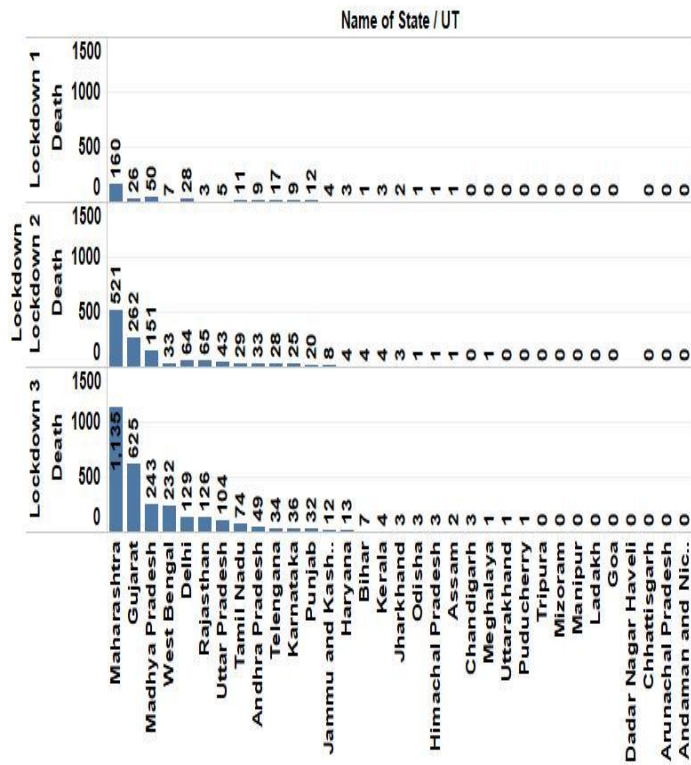Both Tamil Nadu, Punjab and Kerala done well in terms of cured count.
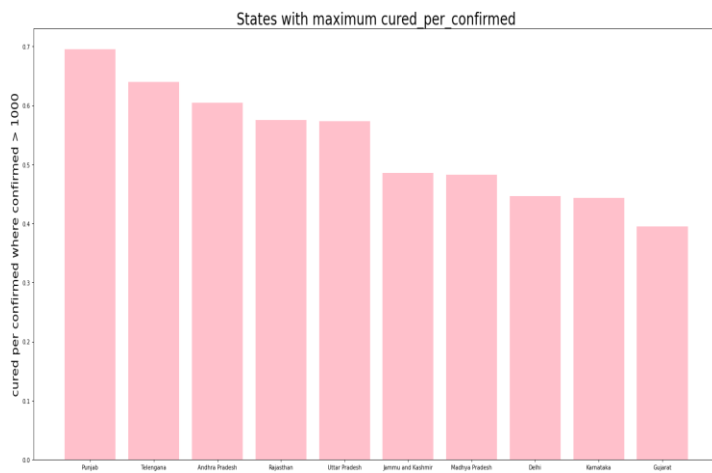
**Fig. 15. Statewise Death Data**



**Fig. 16. States with Maximum Cured Per Confirmed**

## 4. CONCLUSIONS

The important objective of the article is to explore and examine the statewiseCOVID-19 spread pattern in India. Few papers have been published in this regard. But none of the research focused on lockdown. So this paper focused on lockdown and taken many additional parameters to explore more information. Parameters like samples tested, confirmed cases, the death count and cured count are taken into consideration. Results indicate that what happened in each phase of lockdown. This will help to make better decisions in the upcoming days.

## REFERENCES

[1]    Nadia AL-Rousan, Hazem AL-Najjar, "Data Analysis of Coronavirus CoVID-19 Epidemic in South Korea Based on Recovered and Death Cases", Journal of Medical Virology, April 2020, DOI: 10.1002/jmv.25850.

[2]    T.Dinesh Kumar, E.Prabhakar, K.Nandhagopal, "An Enhanced Ensemble Classification Algorithm (EECA) for Airline Services Big Data Sentiment Analysis", Journal of Advanced Research in Dynamical & Control Systems, Volume 11, Issue 7, ISSN 1943-023X, Page No.1461-1467, August 2019.

[3]    E.Prabhakar, R.Parkavi, N.Sandhiya, M.Ambika, "Public Opinion Mining For Government Scheme Advertisement", International Journal of Information Research and Review, Volume 3, Issue 4, Page No.2112-2114, April 2016.

[4]    Cleo Anastassopoulou, Lucia Russo, AthanasiosTsakris, ConstantinosSiettos, "Data-based analysis, modelling and forecasting of the COVID-19 outbreak", PLOS ONE, https://doi.org/10.1371/journal.pone.0230405, March, 2020.

[5]    Samrat K. Dey, Md. Mahbubur Rahman, Umme R. Siddiqi, ArpitaHowlader, "Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach", Journal of Medical Virology, March 2020,DOI: 10.1002/jmv.25743.

[6]    E.Prabhakar, G.Pavithra, R.Sangeetha, G.Revathy, "Mining Better Advertisement Tool for Government Schemes", International Journal for Technological Research in Engineering, Volume 3, Issue 5, ISSN (Online): 2347 – 4718, Page No.1023-1026, January 2016.

[7]    Sarvam Mittal, "An Exploratory Data Analysis of COVID-19 in India", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 9, Issue 04, April-2020.

[8]    E.Prabhakar, "Enhanced AdaBoost Algorithm with Modified Weighting Scheme for Imbalanced Problems", The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Volume 6, Issue 4, ISSN 2321-2381, Page No.22-26, July 2018.

[9]    Davit Gondauri, EkaterineMikautadze, MikheilBatiashvili, "Research on COVID-19 Virus Spreading Statistics based on the Examples of the Cases from Different Countries", Electronic Journal of General Medicine, 17(4), e-ISSN: 2516-3507, 2020.

[10] Rajan Gupta, Saibal K. Pal, Gaurav Pandey, "A Comprehensive Analysis of COVID-19 Outbreak situation in India", medRxiv, DOI: https://doi.org/10.1101/2020.04.08.20058347, 2020.

[11] E.Prabhakar, K. Sugashini, "New Ensemble Approach to Analyze User Sentiments from Social Media Twitter Data", The SIJ Transactions on Industrial, Financial & Business Management (IFBM), Volume 6, Issue 1, ISSN 2321-242X, Page No.7-11, June 2018.

[12] E.Prabhakar, V.S.Suresh Kumar, S.Nandagopal, C.R.Dhivyaa, "Mining Better Advertisement Tool for Government Schemes Using Machine Learning", International Journal of Psychosocial Rehabilitation, Volume 23, Issue 4, ISSN 1475-7192, Page No. 1122-1135, 2019.

[13] Ahmad Yame, "Overview: Data Collection and Analysis Methods in Impact Coronavirus Disease (COVID-19)", Experimental Findings, Research Gate, March 2020.

[14] E.Prabhakar, M.Santhosh, A.Hari Krishnan, T.Kumar, R.Sudhakar, "Sentiment Analysis of US Airline Twitter Data using New Adaboost Approach", International Journal of Engineering Research & Technology (IJERT), Volume 7, Issue 1, ISSN 2278-0181, Page No.1-3, April 2019.

[15] M.Vijayakumar, E.Prabhakar, "A Hybrid Combined Under-Over Sampling Method for Class Imbalanced Datasets", International Journal of Research and Advanced Development (IJRAD), Volume 2, Issue 5, ISSN 2581-4451, Page No. 27-33, December, 2018.

[16] BalGovind Shukla, Hariom Gupta, "An Exploratory Study Of Business Students Perspectives On Demonetization InIndia: With Special Reference To Allahabad City", Research Journal of Social Science & Management, ISSN: 2251-1571, January 2018.

[17] https://en.wikipedia.org/wiki/Coronavirus_disease_2019

[18] https://en.wikipedia.org/wiki/COVID-19_pandemic

[19] https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India

[20] https://www.kaggle.com/sudalairajkumar/covid19-in-india

[21] https://www.kaggle.com/imdevskp/covid19-corona-virus-india-dataset

[22] https://en.wikipedia.org/wiki/COVID-19_pandemic_lockdown_in_India

[23] https://en.wikipedia.org/wiki/Timeline_of_the_COVID-19_pandemic_in_India

[24] https://www.mohfw.gov.in/

[25] https://www.covid19india.org/

[26] https://github.com/CSSEGISandData/COVID-19

[27] https://www.worldometers.info/coronavirus/country/india/