ANALYSIS REPORT
ON
"Covid-19 PANDEMIC INIDA"

2019-2020

SUBMITTED BY

AKASH KUNDU (11001)
AKSHAY KALE (11004)
SHUBHAM RAJPUT (11023)
SIDDHANT FULZELE (11024)
TEJAS AKADKAR (11027)

Submitted in partial fulfillment of the requirements for qualifying

M.Sc.(Data Science) – SEM II
Department of Computer Science

DECCAN EDUCATION SOCIETY'S

FERGUSSON COLLEGE (AUTONOMOUS),

PUNE:- 411 004

Deccan Education Society's
Fergusson College (Autonomous), Pune
Department Of Computer Science

CERTIFICATE

This is to certify that the project entitled
_____ **Covid-19 PANDEMIC INIDA** _____ submitted by

1. **_Aakash Kundu (11001)_____**
2. **_Akshay Kale (11004)_____**
3. **_Shubham Rajput (11023)_____**
4. **_Siddhant Fulzele (11024)_____**
5. **_Tejas Akadkar (11027)_____**

in partial fulfillment of the requirement of the completion of M.Sc.(Data Science)-I

[Semester-II], has been carried out by them under our guidance satisfactorily

during the academic year 2019-2020.

Place: Pune

Date: /   /2020

Head of Department

Department Of Computer Science

Fergusson College, Pune

Project Guide:
1.   _____

Examiners Name   **Sign**

1. _____   _____

2. _____   _____

ACKNOWLEDGEMENT

ACKNOWLEDGEMENT

Words fall short to express our deep sense of gratitude towards them all who have imparted their valuable time, energy and intellect towards the beautification of our Analysis project entitled as, **"Covid-19 Pandemic India".**

It gives us a great pleasure in presenting this report. Its justification will never sound good if we do not express our vote of thanks to our guide **Prof. Swati Satpute** without whose help our Analysis & its thesis would have neither began nicely nor would have reached a fine ending.

Never can we forget the hard labor & pain taken by our **H.O.D. & All Professors** who's hard – working nature, sophisticated teaching & guidance helped us framing & building this project & lastly we are also thankful to **www.covid19india.org** and their Research Team for being directly or indirectly helpful to us to build our project to be presented to the college.

By:-

AKASH JAGDISH KUNDU
AKSHAY PARSHURAM KALE
SHUBHAM RAJENDRA RAJPUT
SIDDHANT BALIRAM FULZELE
TEJAS PRADEEP AKADKAR

Content

LIST OF DIAGRAM

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION:

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, China, and has resulted in an ongoing pandemic. The first case may be traced back to 17 November 2019. As of 8 June 2020, more than 6.98 million cases have been reported across 188 countries and territories, resulting in more than 401,000 deaths. More than 3.13 million people have recovered.

The virus is primarily spread between people during close contact, most often via small droplets produced by coughing, sneezing, and talking. The droplets usually fall to the ground or onto surfaces rather than travelling through air over long distances. Less commonly, people may become infected by touching a contaminated surface and then touching their face. It is most contagious during the first three days after the onset of symptoms, although spread is possible before symptoms appear, and from people who do not show symptoms

The virus is primarily spread between people during close contact, most often via small droplets produced by coughing, sneezing, and talking. The droplets usually fall to the ground or onto surfaces rather than travelling through air over long distances. Less commonly, people may become infected by touching a contaminated surface and then touching their face. It is most contagious during the first three days after the onset of symptoms, although spread is possible before symptoms appear, and from people who do not show symptoms.

1.2 PANDEMIC :

The COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The outbreak was first identified in Wuhan, China, in December 2019. The World Health Organization declared the outbreak a Public Health Emergency of International Concern on 30 January, and a pandemic on 11 March.

A global coordinated effort is needed to stop the further spread of the virus. A pandemic is defined as "occurring over a wide geographic area and affecting an exceptionally high proportion of the population." The last pandemic reported in the world was the H1N1 flu pandemic in 2009.

Coronaviruses are important human and animal pathogens. At the end of 2019, a novel coronavirus was identified as the cause of a cluster of pneumonia cases in Wuhan, a city in the Hubei Province of China. It rapidly spread, resulting in an epidemic throughout China, followed by an increasing number of cases in other countries throughout the world. On 30th January 2020 India recorded its first COVID-19 case in state of Kerala. It was a student who had travel history to china. And till the start of June India has over 200 thousand confirmed cases.

1.3 PROBLEM STATEMENT:

In this project we dived deep into 'What does data say about Covid-19 situation in India?'. And with available data we came up with some observations and conclusions.

This analysis mainly focuses on:

✔ What is the current COVID-19 situation in India?

✔ State-wise comparison.

✔ What could be the reasons behind cases clusters found in India.

✔ Is lockdown in India successful or not?

.

CHAPTER 2

LITERATURE SURVEY

2.1 RESEARCH PAPER REFERRED

1) Covid-19 impact on Indian Economy

   Authers:
   - S. Mahendra Dev (Indira Gandhi Institute of Development Research)
   - Rajeshwari Sengupta (Indira Gandhi Institute of Development Research)

   https://ideas.repec.org/p/ind/igiwpp/2020-013.html

2) COVID-19 in India: Potential Impact of the Lockdown and Other Longer-Term Policies

   Authers:
   - Emily Schueller (CDDEP)
   - Eili Klein (Department of Emergency Medicine, Johns Hopkins School of Medicine; CDDEP)
   - Gary Lin (Department of Emergency Medicine, Johns Hopkins School of Medicine

   https://cddep.org/wp-content/uploads/2020/04/India-Shutdown-Modeling-Slides-Final-2.pdf

2.2 DATA SOURCES

For the COVID-19 data we have scrapped https://api.covid19india.org which is a volunteer-driven, crowdsourced database for COVID-19 stats & patient tracing in India .

For facts and information we have referred www.wikipedia.com and www.twitter.com .

CHAPTER 3

DATA &
PRE-PROCESSING

3.1 DATA DISCRIPTION

Data contains multiple files which includes

Raw_data : It has all the raw information about COVID-19 in India.

Having Columns ['agebracket', 'backupnotes', 'contractedfromwhichpatientsuspected', 'currentstatus', 'dateannounced', 'detectedcity', 'detecteddistrict', 'detectedstate', 'estimatedonsetdate', 'gender', 'nationality', 'notes', 'numcases', 'patientnumber', 'source1', 'source2', 'source3', 'statecode', 'statepatientnumber', 'statuschangedate', 'typeoftransmission', 'entryid']

Statewise_data : It contains state wise data of COVID-19.

Having Columns ['active', 'confirmed', 'deaths', 'deltaconfirmed', 'deltadeaths', 'deltarecovered', 'lastupdatedtime', 'migratedother', 'recovered', 'state', 'statecode', 'statenotes']

Daily_data : It contains number of patient recorded on daily basis.

Having Columns ['dailyconfirmed', 'dailydeceased', 'dailyrecovered', 'date', 'totalconfirmed', 'totaldeceased', 'totalrecovered']

Test_data : Tests conducted in every state.

Having Columns ['individualstestedperconfirmedcase', 'positivecasesfromsamplesreported', 'samplereportedtoday', 'source', 'testpositivityrate', 'testsconductedbyprivatelabs', 'testsperconfirmedcase', 'testspermillion', 'totalindividualstested', 'totalpositivecases', 'totalsamplestested', 'updatetimestamp']

State_wise_GDP : All States GDP values of last 8 years.

Having Columns ['state/ut', '2019-20']

State_wise_Population : All states' population figures according to 2011 Census

Having Columns ['state/ut', 'Population', 'Decadal Groth', 'Rural Population', 'Urban Population', 'Area', 'Density', 'Sex ratio'

3.2 TECHNOLOGIES   USED

Python for data scrapping , preprocessing , visualization etc.

Libraries Used :

1) Scrapy : Scraping data from Web

2) Pandas : Working with data files

3) Numpy : for Scientific Calculation

4) Matplotlib : Basic Visualization

5) Plotly : Advance Visualization

6) Dash : Creating Dashboard

Tableau for visualization.

Tableau is powerful visualization tool which allow us to plot geo-informatical Data.

3.3 DATA PRE-PROCESSING

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing.

Steps Used :

1.Scrapped data was in raw form all information in a single file. So we needed to make separate files based on their purpose.

2.Death rate in each state is calculated by the formula (number of deaths/total confirmed cases)*100.

3. Recovery rate in each state is calculated by the formula (number of   recoveries / total confirmed cases)*100.

4.Split the data according to the lockdown periods.

 5.Null values in state wise GDP data filled with average.

3.4 DATA EXPLORATION

3.4.1 TYPES OF VISUALIZATION USED

Pie chart: To show active ,recovered, death percentage due to COVID and gender distribution.

Bar Plot: To show state comparison on various factors, age-wise distribution etc.

Line Plot : To show confirmed , recovered , deceased trend on daily basis.

3D plot: To show state wise daily count of confirmed cases.

3.4.2 TECHNOLOGIES USED FOR VISUALIZATION

Python and Tableau are used to make all the visualizations which are displayed on the dashboard.

Plotly and matplotlib are the python libraries used for the visualization

CHAPTER 4

ANALYSIS

4.1 ANALYSIS (IMPLEMENTATION)

India recorded its first COVID-19 case on 30th January 2020 in kerala. The infected person was a student who had travelled to china for academic purpose. And since then cases in India is rising exponentially.

A complete Dashboard is released on **http://siddfulzele.pythonanywhere.com/covid_India/**

What is the current situation in India?
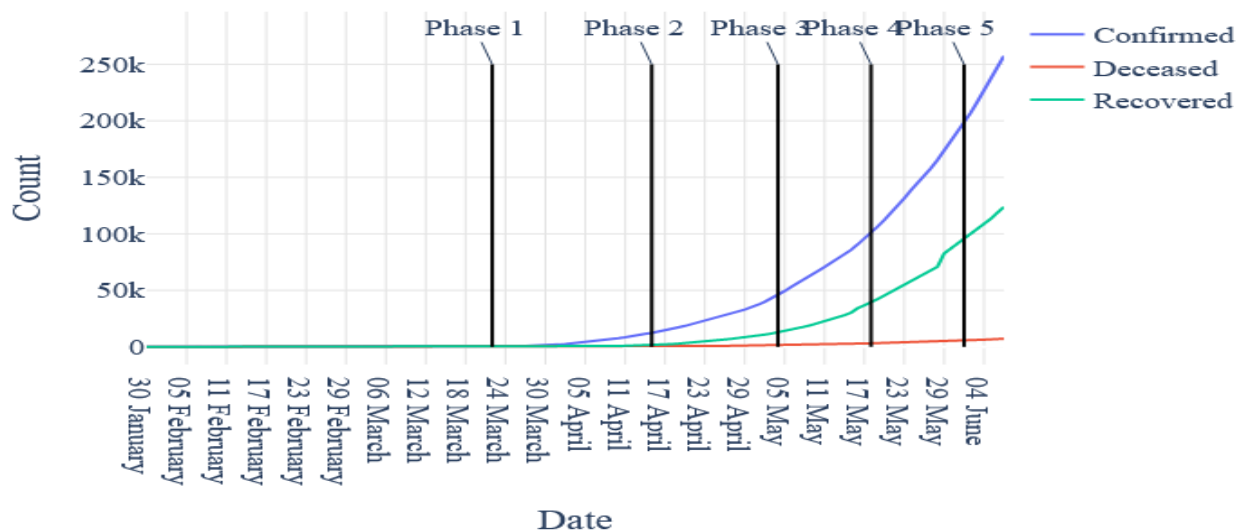
(Fig 4.1)

As shown above, India had recorded over 500 cases till 24th March. So government declared nation-wide lockdown from 25th march to 14th April also known as lockdown 1.0 and after this government has been extending nation-wide lockdown step by step.
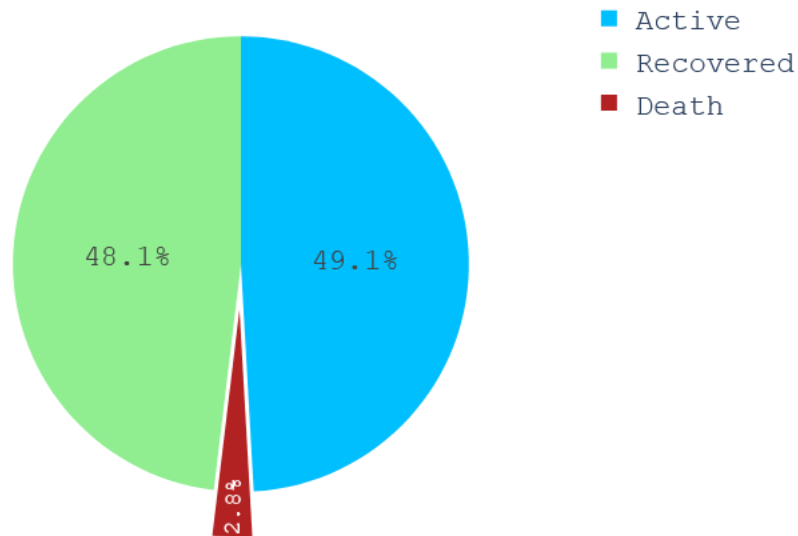
Situation in India till 7th June 2020.

<div align="center">

Confirm Cases : 257487

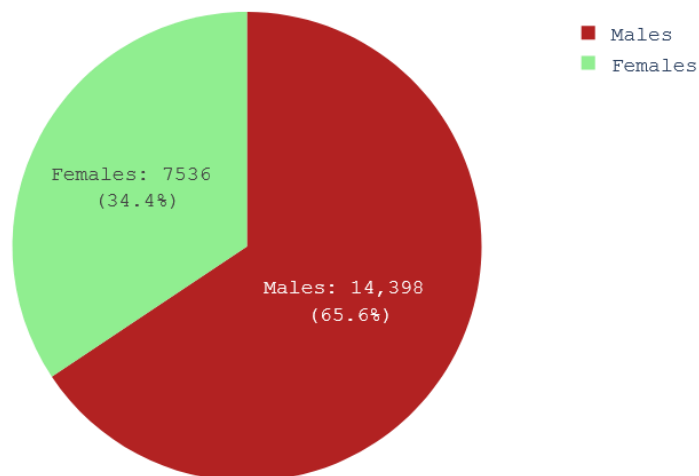## Recovered : 123848

Active : 126433

Deceased : 7206

</div>

(Fig 4.2)

India is showing good recovery rate day by day with low rate of deaths but on the other hand it is also reaching new peak of confirmed cases everyday.
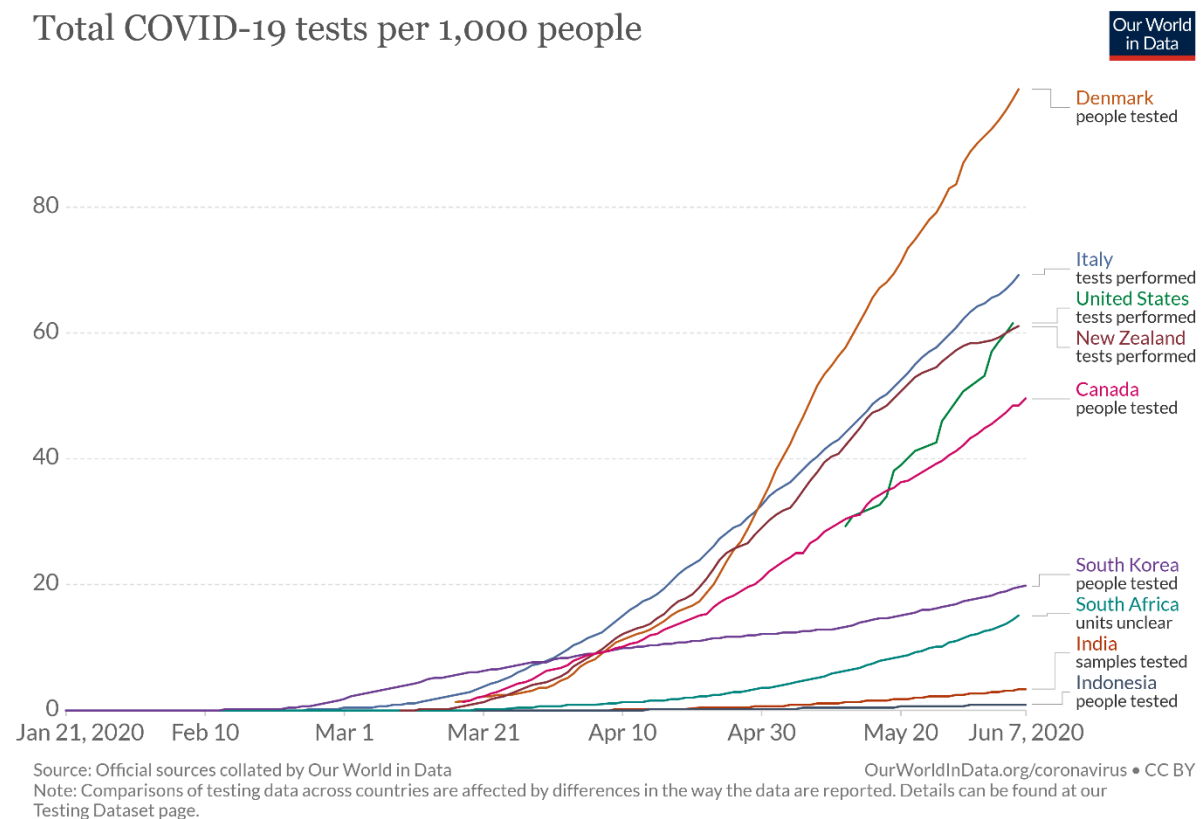
If we talk at the level of patients then as per sample size 21936 cases,

(Fig 4.3)

As shown above, number of infected males is greater than that of females. The more exposure of males in Indian family to outside work could be the reason behind this.

India is highly criticized by experts for the low number of tests being conducted.

Total COVID-19 tests per 1,000 people

Our World in Data

Denmark
people tested

Italy
tests performed
United States
tests performed
New Zealand
tests performed

Canada
people tested

South Korea
people tested
South Africa
units unclear
India
samples tested
Indonesia
people tested

80

60

40

20

0

Jan 21, 2020   Feb 10   Mar 1   Mar 21   Apr 10   Apr 30   May 20   Jun 7, 2020

Source: Official sources collated by Our World in Data   OurWorldInData.org/coronavirus • CC BY
Note: Comparisons of testing data across countries are affected by differences in the way the data are reported. Details can be found at our
Testing Dataset page.

(Fig 4.4)

(Image source: https://ourworldindata.org/grapher/full-list-cumulative-total-tests-per-housand?time=..)

As shown in the graph above India places itself at the tail-end of this comparison. Low number of tests conducted was the reason behind at the lower end in world wide tally of COVID-19 cases in early days. But as of now we are observing 8-9 thousand people being found positive every day.
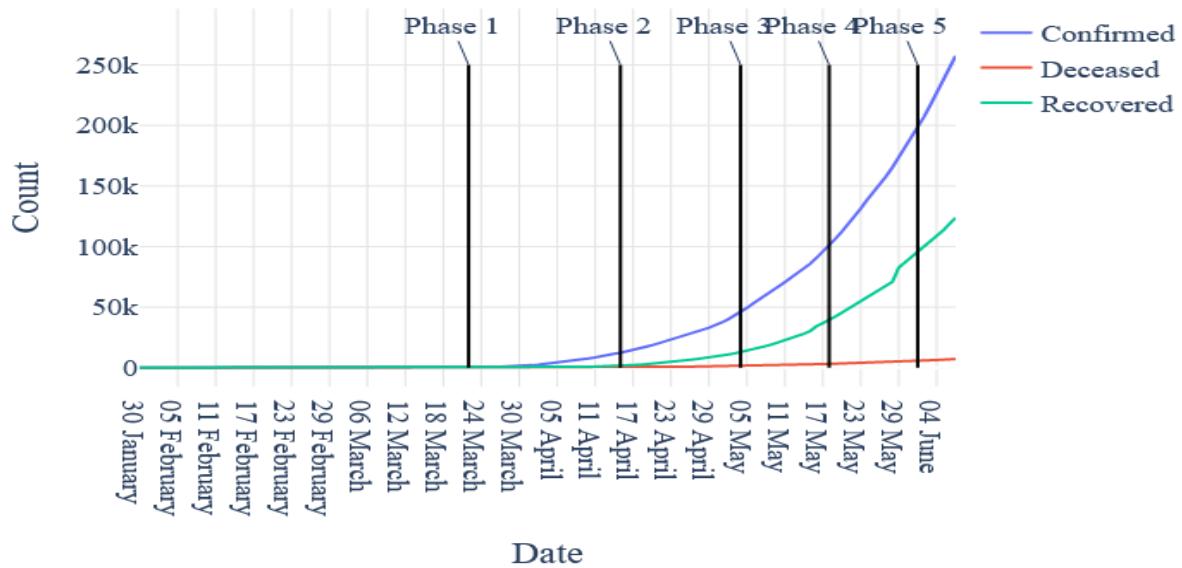
(Fig 4.5)

As confirmed cases are increasing day by day the positive side is recovered number is also showing somewhat same behaviour with less number of people dying of COVID-19.India is yet to reach its peak of confirmed cases which is once reached then after that cases will start to decrease and recovered number will go higher to intersect confirmed cases at a point.


State-wise Comparison

India consist of 28 states and 8 Union Territories with varying features such as demography, geography, location, lifestyle etc. Which can decide the spread of contagious virus. Comparing COVID-19 situation in states gives us insight of which state/area is to be focused.
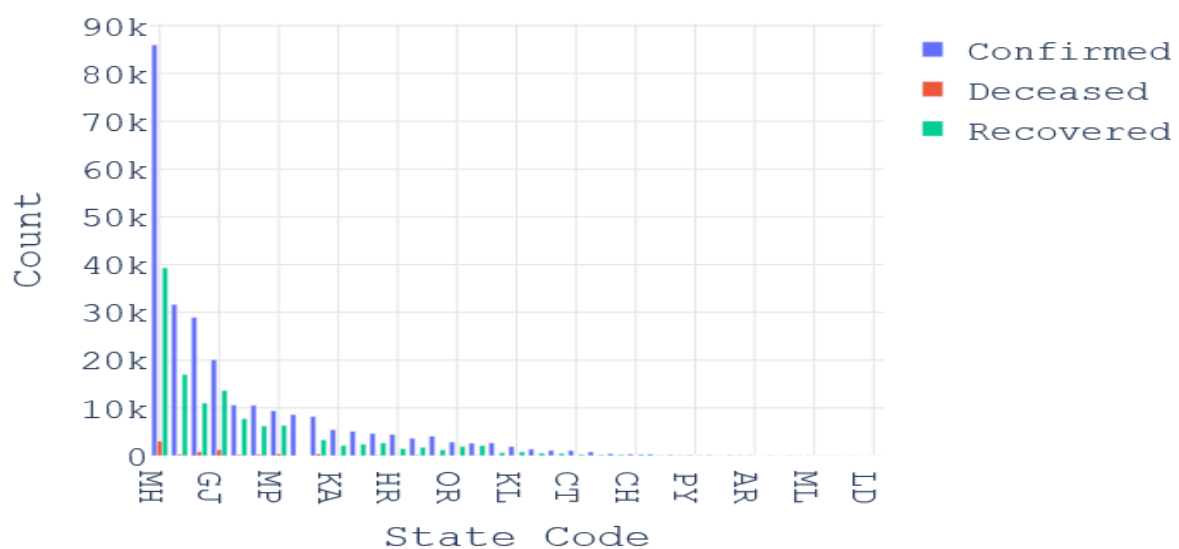
(Fig 4.6)

As shown above, Maharashtra, Tamil Nadu, Delhi, Gujrat, Uttar Pradesh, Madhya

Pradesh are top 6 states in confirmed cases in India. When the cases started to be detected in India since that time only Maharashtra has been leading this tally. As of 5th June 2020, over 80,000 cases are confirmed in Maharashtra.
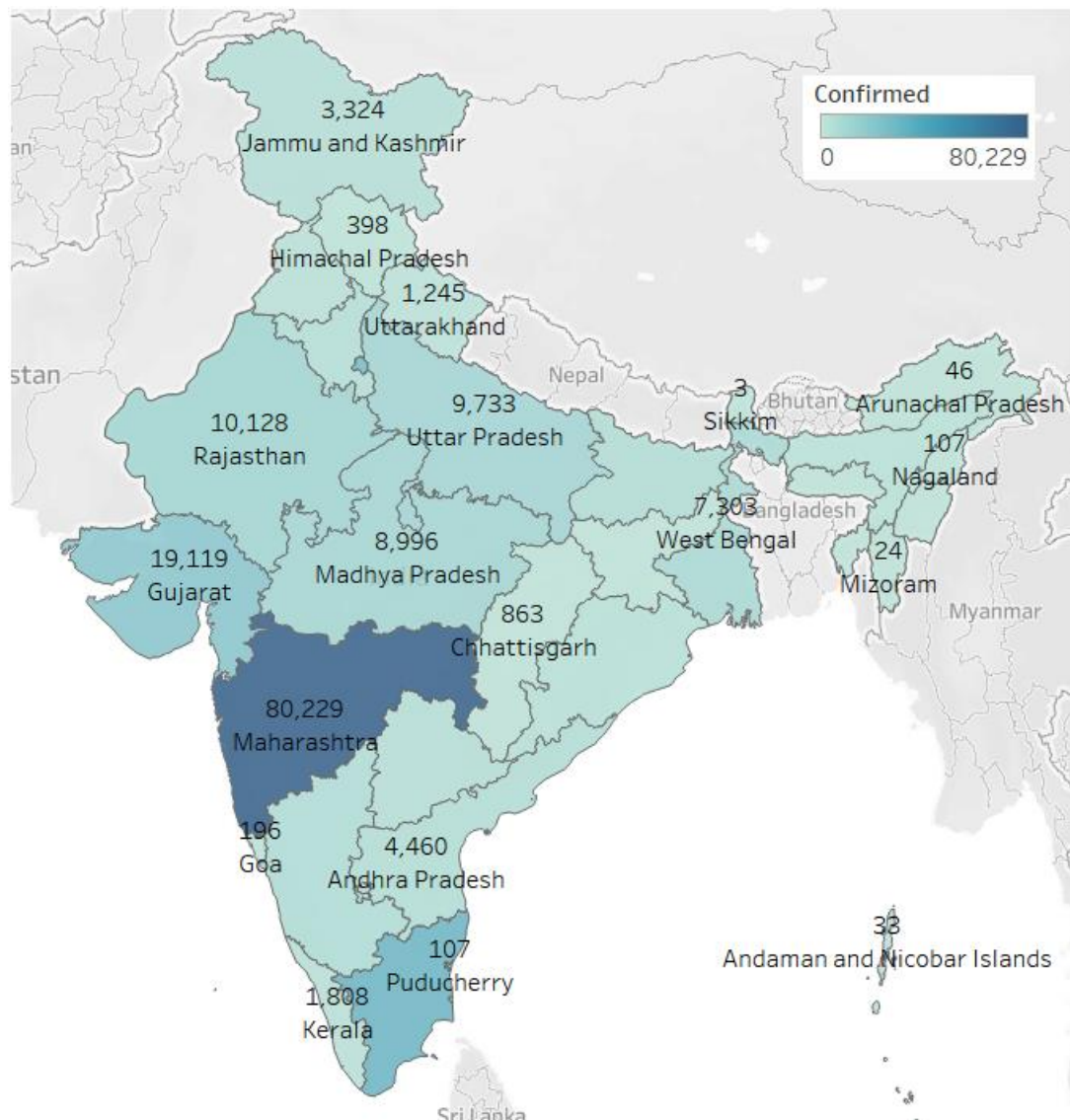
(Fig 4.7)

We can see that more number of COVID-19 clusters are in left half part of the India. In the North-East region of India ,despite being close to China , there are very less number of confirmed cases. The possible reason of this variation is discussed later part of this analysis.

4.2 R0 FACTOR

The R0 factor also known as Basic Reproduction Number. The R0 factor of an area denotes how many more persons can be infected due to a person who is having the disease. If R0 is 1.2 then it means 1 infected person can infect 1.2 person by the same disease. So that 5 people can infect total 6 new persons (1.2*5=6.0).

| Estimated R Value For States		
State	Period of Estimation	Estimated R
India	March 4 - April 11	1.83
	April 13-May 14	1.29
	May 16 - May 25	1.23
Maharashtra	April 13-April 26	1.49
	April 23-May 15	1.34
	May 4-May 25	1.27
Gujarat	April 30-May 9	1.23
	May 7-May 10	1.26
Tamil Nadu	April 29-May 4	1.83
	April 30-May 7	2.01
	May 7-May 10	1.31
Delhi	May 7-May 10	1.31
	May 20-May 25	1.2
Punjab	April 28-May 4	1.48
	May 5-May 8	1.32
Rajasthan	April 19-April 22	1.34
	May 10-May 25	1.27
Madhya Pradesh	April 16-May 1	1.23
	May 10-May 23	1.23
Uttar Pradesh	May 1-May 4	1.28
	May 19-May24	1.33
West Bengal	April 15-April 28	1.51
	April 28-May 1	1.14
	May 4-May 10	1.34
	May 15-May 25	1.22
Andhra Pradesh	April 6-May 1	1.27
Karnataka	May 16-May 24	1.62

(Fig 4.8)

(Image source: https://scroll.in/article/963252/in-india-how-many-people-does-a-covid-19-patient-infect-on-average)

Its is believed that India is having steady R0=1.23 since 25th May which is still higher than normal. While imposing nation-wide lockdown R0 is considered as one of the factor . If the R0 values is less than 1 i.e. one infected person can not infect another person then lockdown is to be lifted. But Lockdown in India is being lifted in steps despite having R0 factor greater than

1. It could lead to drastic results in future.

4.3 DOUBLING RATE

Doubling rate is number of days taken to double the number of confirmed cases in a particular area. This number should be as large as possible. Large number of doubling rate will increase the hopes of situation under control.
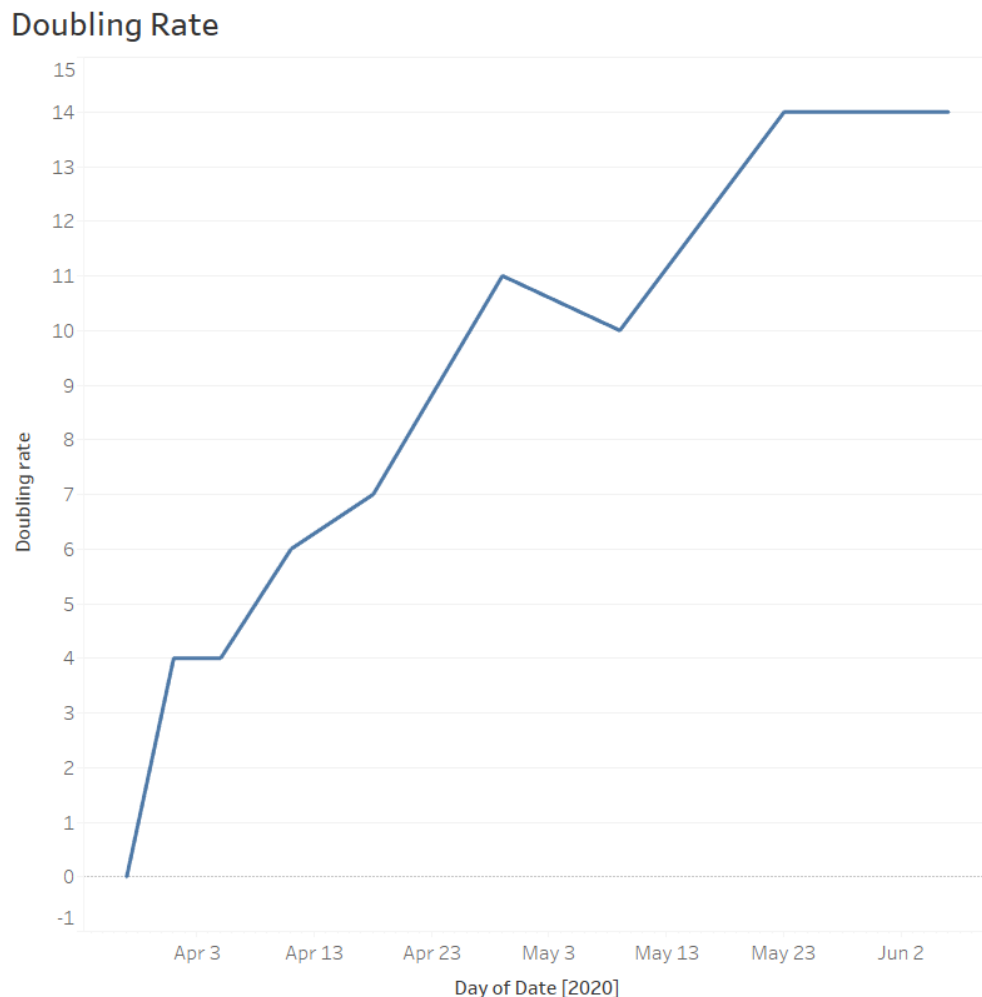

(Fig 4.9)

India is showing positive trend in doubling rate but the situation is still a not under the control of government and administration. This is a thing to worry about.

4.4 DEATH RATE

Death rate is calculated by,

100*(number of deaths in a state / number of confirmed cases in a state)

The number obtain after this calculation is the number of deaths behind every 100 confirmed cases in that state. Higher the number more the area is to be focused.
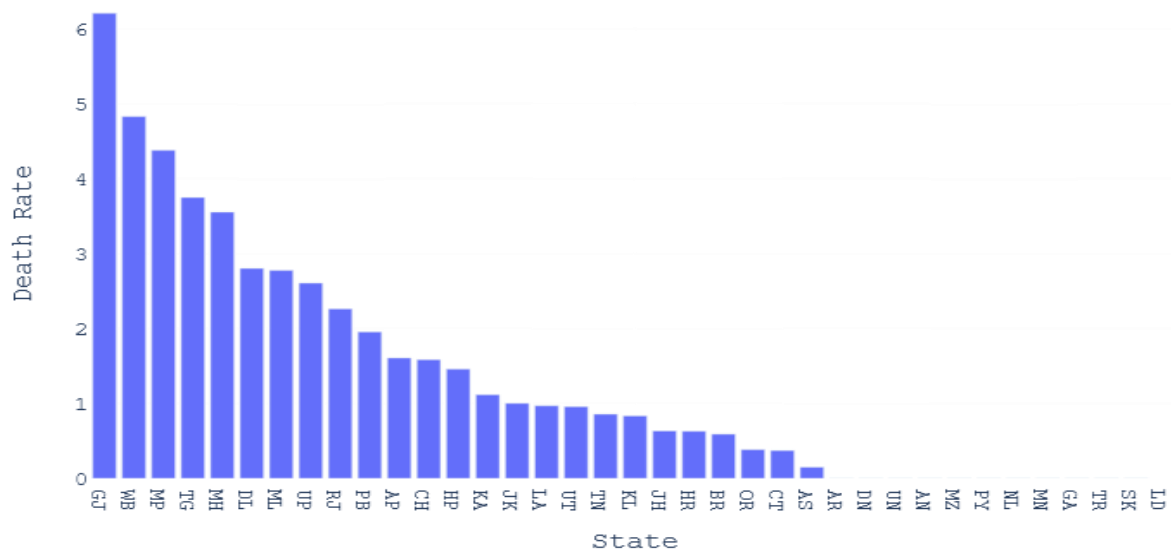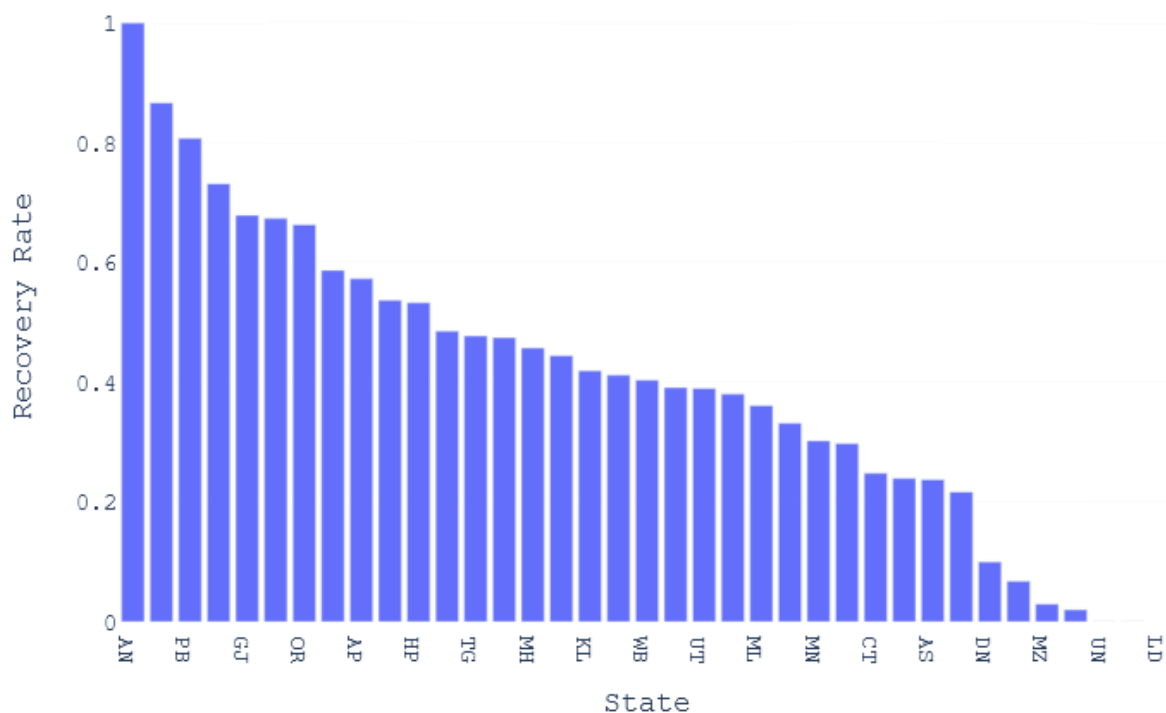

(Fig 4.10)

Gujrat is having death rate over 6 followed by West Bengal , Madhya Pradesh , Maharashtra , Telangana with death rates 5 , 4.2 , 3.5 , 3.4 respectively. Where as whole country's death rate on 5th June 2020 is 2.78 .All the states/UTs with death rate more than 2.78 are to be taken care of more intensively.

4.5 RECOVERY RATE

Recovery rate is calculated by,

100*(number of recoveries in a state / number of confirmed cases in a state)

The number obtain after this calculation is the number of recovered patients behind every 100 confirmed cases in that state.

(Fig 4.11)

Chandigarh is having recovery rate close to 90 followed by Punjab, Rajasthan , Gujrat ,Madhya Pradesh with recovery rates 84, 72, 68, 65 respectively. Where as whole country's death rate on 5th June 2020 is 47.48.

4.6 CLUSTER ANALYSIS

What could be the reasons behind cases clusters found in India.

4.6.1 URBAN POPULATION

As discussed earlier, there are more cases found in western part than the eastern part of India. We tried to find out some of the reasons that may have caused this situation in India.

India has 2nd most population in the world and places itself at 7th position in the tally of surface area. So somewhere population is going to be major factor in spread of such pandemic.
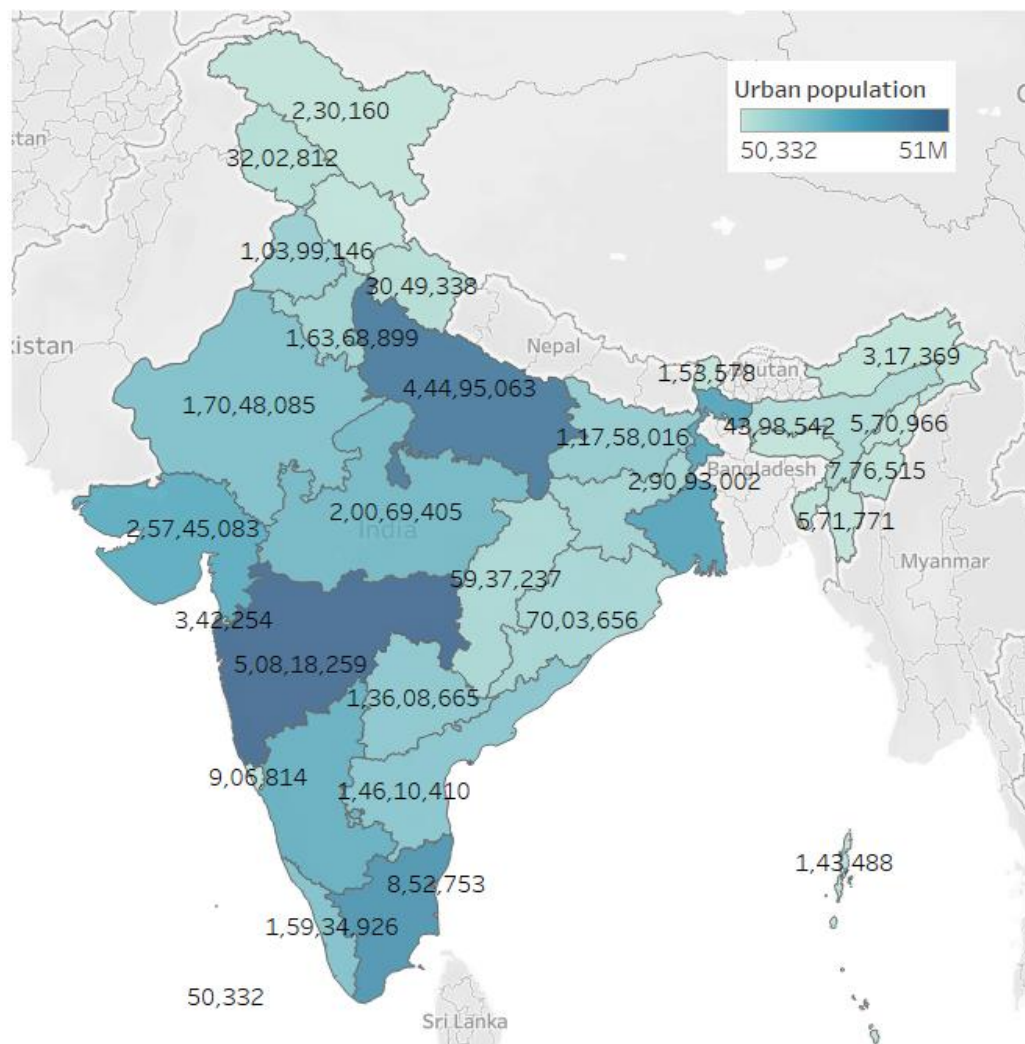
(Fig 4.12)

Above is the map of India showing urban population in each state.

And Fig 4.7 is the graph showing confirmed covid-19 cases in each state in India . They both look somewhat similar indicating 'More urban population causes quick spread of corona virus'.

Top 8 States with highest urban population:

Maharashtra ,Uttar Pradesh, Tamil Nadu , West Bengal ,Gujarat, Karnataka ,Madhya Pradesh , Rajasthan.
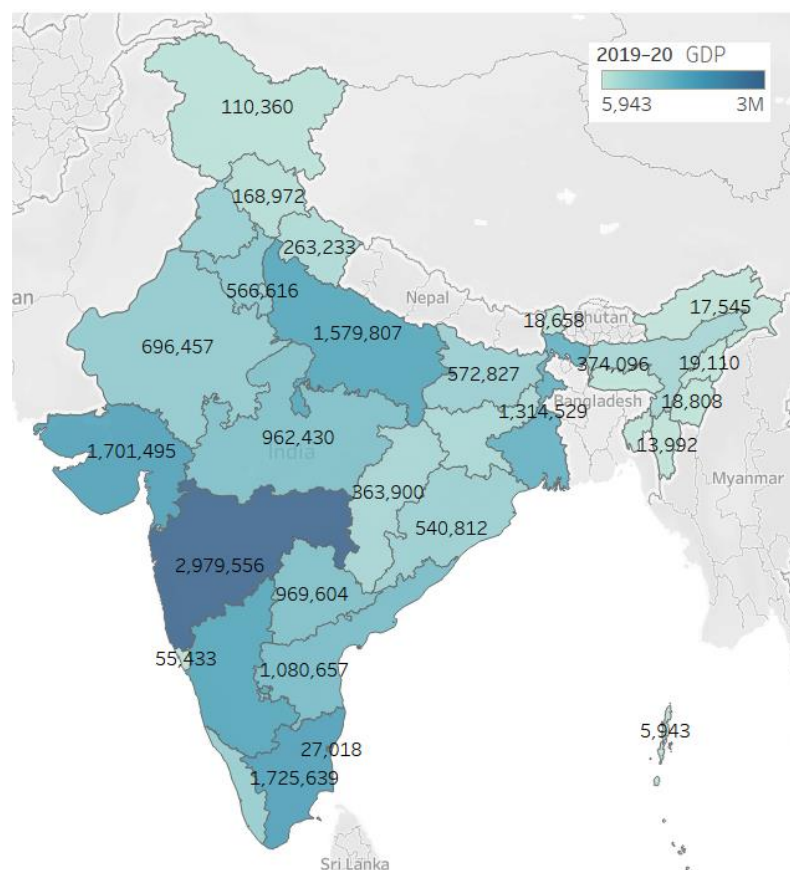
Top 8 States with highest number of COVID-19 cases:

Maharashtra , Tamil Nadu, Delhi ,Gujarat ,Rajasthan , Uttar Pradesh ,Madhya Pradesh ,West Bengal.

Out of top 8 states 7 states are common in both category. This indicates more urban population in these states causing higher number of cases.

4.6.2 GDP CONTRIBUTION

India is the world's 5th largest economy by nominal GDP. All states contributes in GDP of India.

(Fig 4.13)

These are the figures of state and its GDP for the financial year 2019-20 in Crores.

Fig 4.7 graph is state and its confirmed COVID-19 cases. These two graphs too look somewhat

27

identical indicating that 'Public movement , interaction ,contact with each other in activities which contribute in GDP, causes a favourable environment for the spread of such viruses.'

Top 8 States with highest GDP:

Maharashtra , Tamil Nadu, Gujarat, Karnataka, Uttar Pradesh ,West Bengal , Andhra Pradesh ,Telangana

Top 8 States with highest number of COVID-19 cases:

Maharashtra , Tamil Nadu, Delhi ,Gujarat ,Rajasthan , Uttar Pradesh ,Madhya Pradesh ,West Bengal.

The comparison above showing out of top 8 states 5 states falls in both of the category, Which indicates that if more number of businesses running in a state the more will such viruses spread in that state. Pandemics like COVID-19 hits the economy building factors the most.

4.7 LOCKDOWN

Is lockdown in India a successful or not?

Indian government implemented lock down in 4 phases. In comparison with other countries ,
India declared lockdown in its early days of COVID-19 outbreak.

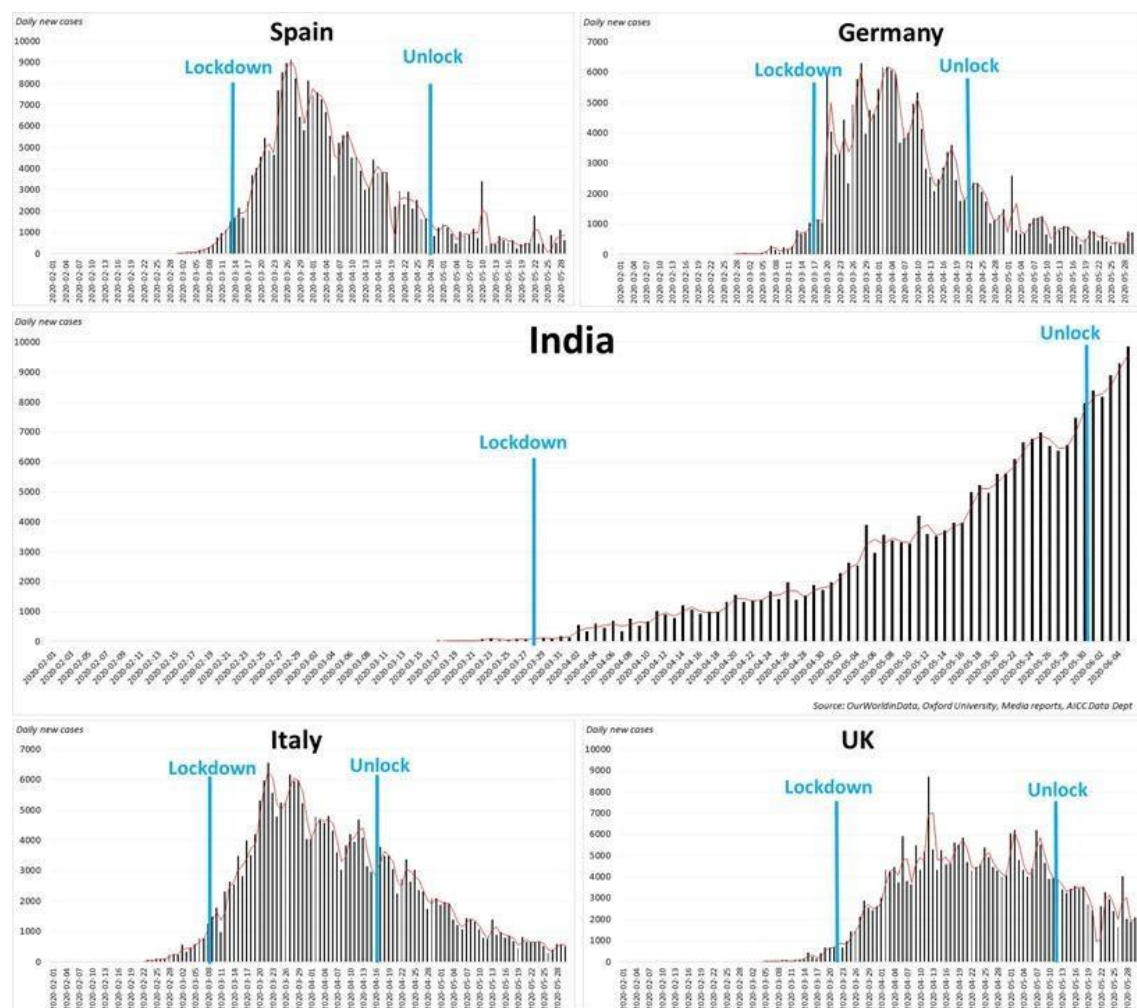   Phase 1: 25th March – 14th April

      Phase 2: 15th April – 3rd May

      Phase 3: 4th May – 17th May

      Phase 4: 18th May – 31st May

Among which phase 1 and 2 of lockdown were very strict whereas phase 3 and phase 4 were
comparatively lenient.

In all 4 phases of lockdown, not a single lockdown showed any down falling of the curve and yet
India has not reached the peak yet.

(Fig 4.14)

In all the phases of lockdown in India , cases graph is showing increasing trend only. Lockdown was meant to find a peak of cases in India but it failed to do so. When we compare India's lockdown phase with other country's lockdown phases , we see India loosing with great margin.

As per experts' analysis we are still very far from a peak which is quite scary. Lockdown in India did not serve its purpose and caused economic harm as well as non-decreasing cases in India.

CHAPTER 5

CONCLUSION & FUTURE ENHANCEMENT

5.1 CONCLUSION

- India is now 6th largest confirmed cases of COVID-19 in the world.
- India has not reached the peak yet so as of now we cannot predict the approximate number of confirmed cases in India.
- Recovery rate of India is also showing exponential behavior same as confirmed cases. But if cases increase beyond certain point then thing can go out of control which will affect the recovery rate.
- Gender information of most of the patients is nor released by the government but whatever data is available shows number of infected males is more than that of females. This is may be due to more exposure/contact of males with outdoor world.
- Same as gender, age information is unavailable for most of the patients , but whatever data is available shows age and cases are normally distributed and 21-40 is the age bin which has been infected more.
- This analysis showed that pandemic like this affects economy the most. Whichever the country's GDP source is , it gets targeted the most. Slowing down the economy then unemployment, job losses and then this chain reaction continues.
- India was in lockdown for more than 2 moths but still situation did not get any better. This may be due to weak administration or the violation of lockdown by citizens.

5.2 FUTURE ENHANCEMENT

1.Prediction Model

India has not reached the peak yet , once it reaches the peak the prediction model can be built to show that how much time it will take to get things back to the normal.

2.Sentiment Analysis

India has never experience such pandemic in last 100 years so what do people think about this pandemic, lockdown , government approach/policies etc can be studied to have sentiment insight of this pandemic.

CHAPTER 6

BIBLIOGRAPHY

6.1 REFERENCE BOOKS

- **Mastering Python For Data Science : Samir Madhavan**
- **Learning With Python**

6.2 WEBSITE

- https://api.covid19india.org

- **www.wikipedia.com**
- **www.twitter.com .**
- **https://www.w3schools.com/**
- **https://stackoverflow.com/**
- **https://plotly.com/**
- **https://www.pythonanywhere.com/**