



# COVID-19 DATA ANALYSIS AND PREDICTION OF INDIA

<sup>1</sup>Shubham Sharma, <sup>2</sup>Chetan Sharma, <sup>3</sup>Shivansh Dwivedi

<sup>1</sup>B.tech School of Computer science, Galgotias University, Greater Noida, India

<sup>2</sup>B.tech School of Computer science, Galgotias University, Greater Noida, India

<sup>3</sup>B.tech, School of Computer science Galgotias University, Greater Noida, India

Article DOI: <https://doi.org/10.36713/epra7237>

DOI No: 10.36713/epra7237

## ABSTRACT

Words frustrated to specific our deep sense of feeling towards all UN agency have imparted their valuable time, energy and intellect towards the modification of state of our Analysis project entitled as, "Covid-19 Pandemic India".

It offers u. s. a decent pleasure in presenting this report. Its justification will never sound sensible if we have a tendency to tend to do not categorical our vote of attributable to our guide academic. Kuldeep Kaswan whereas not whose facilitate our Analysis & its thesis would have neither began nicely nor would have reached a fine ending.

Never can we have a tendency to tend to forget the taxing labor & pain taken by our H.O.D. & All Professors who's taxing – operative nature, refined teaching & steering helped u. s. framing & building this project & finally we have a tendency to tend to area unit appreciative to web.covid19india.org and their analysis Team for being directly or indirectly helpful to u. s. to create our project to be conferred to the college.

## INTRODUCTION

COVID-19, or a lot of popularly called Novel Corona Virus, is related to the respiratory disorder in humans that has been declared as a worldwide epidemic and pandemic within the half-moon of the year 2020 by the globe Health Organization. As per the newest information (10 th Gregorian calendar month 2020) by John Hopkins University and alternative trailing websites, there area unit presently over seven.3 million individuals infected by the Novel Corona Virus all round the world and shut to 109 thousand deaths reportable from totally different elements of the globe. the highest ten countries with most range of infected cases area unit the us of America,

India, Italy, Germany, France, China, Iran, uk, Turkey and Swiss Confederation. the highest countries with most range of reportable deaths area unit European nation, India, us of America, France and uk. With relevancy the recovered patients list, China is at the highest of the list followed by Republic of India, Germany, Italy, Iran and also the us of America.

India was 1st placed well out the list of infected nations by immense margins, however recent events semiconductor diode to its rise to

second position that may be a purpose of concern. The rate is controlled at but three-dimensional immediately, that is best than the ~5.5% rate of world, however the model of unfold is slowly moving towards associate exponential trend which may cause huge loss of lives and infrastructure.

India is being looked upon by numerous nations currently as a World Leader and even WHO acknowledged that world is trying towards Indian methods to contain the natural event of this epidemic. India accounts for pretty much simple fraction of the world's population and is second leading country in terms of population within the world. Republic of India contributes heavily to the world's gross domestic product and is amongst the foremost distinguished developing country within the world with fairly sturdy economic process percentages. India's sensible sociableness with majority of the nations within the world and its useful nature makes it a perfectly for alternative countries. Therefore, the analysis of COVID-19 natural event in Indian region is closely watched and monitored by the globe and there's a desire of comprehensive analytical studies supported totally different methods taken by Indian directors from time to time. Republic of India has been following a nationwide



internment since 22-March-2020, that was a one-day internment, followed by a 21-day internment once 2 days. each activity in Republic of India since then has been happening with permission from numerous administration units and most the domestic and international travels are either prohibited or monitored closely. Republic of India is already got into the third section of COVID-19 natural event i.e. the community natural event as seen by numerous countries round the world, however the cases are rising unceasingly. Republic of India's internment amount has been compact by 2 major events within the recent days that were associated with the mass exodus of laborers and staff from one state to alternative states (especially from urban center/city/metropolis/urban center} to neighboring states) and physical phenomenon of a non secular event in Delhi that semiconductor diode to spike within the range of cases in numerous states of India. throughout this point, the Indian Prime Minister has been attempting to attach with Indian voters through innovative methods and turning out with numerous engagement activities that area unit impacting the total nation. With such a lot happening in Republic of India immediately, it becomes imperative that we tend to study this scenario and impact of varied such events in Republic of India through information analysis strategies and are available up with totally different plans for future which may be useful for the Indian directors and medical professionals.

## II. OVERVIEW

Coronavirus malady a pair of 019 (COVID-19) is associate communicable disease caused by severe acute metabolic process syndrome coronavirus 2 (SARS-CoV-2). it had been initial known in Dec 2019 in metropolis, China, associated has resulted in an current pandemic. the primary case is also derived back to Revolutionary Organization 17 November 2019. As of eight June 2020, quite half-dozen.98 million cases are rumored across 188 countries and territories, leading to quite 401,000 deaths. quite three.13 million folks have recovered.

The virus is primarily unfold between folks throughout shut contact, most frequently via little droplets created by coughing, sneezing, and talking. The droplets typically fall to the bottom or onto surfaces instead of travel through air over long distances. Less usually, folks could become infected by touching a contaminated surface then touching their face. it's most contagious throughout the primary 3 days when the onset of symptoms, though unfold is feasible before symptoms seem, and from those who don't show symptoms

The virus is primarily unfold between folks throughout shut contact, most frequently via little droplets created by coughing, sneezing, and talking. The droplets typically fall to the bottom or onto

surfaces instead of travel through air over long distances. Less usually, folks could become infected by touching a contaminated surface then touching their face. it's most contagious throughout the primary 3 days when the onset of symptoms, though unfold is feasible before symptoms seem, and from those who don't show symptoms.

## A. Pandemic

The COVID-19 pandemic, additionally called the coronavirus pandemic, is associate degree in progress pandemic of coronavirus illness 2019 (COVID-19), caused by severe acute metabolic process syndrome coronavirus a pair of (SARS-CoV-2). The happening was initial known in city, China, in Dec 2019. the globe Health Organization declared the happening a Public Health Emergency of International Concern on thirty January, and a scourge on eleven March.

A global coordinated effort is required to prevent the any unfold of the virus. a scourge is outlined as "occurring over a large region associate degreeed touching an exceptionally high proportion of the population." The last pandemic reported within the world was the H1N1 respiratory disorder pandemic in 2009.

Coronaviruses square measure vital human and animal pathogens. At the tip of 2019, a completely unique coronavirus was known because the explanation for a

cluster of respiratory disease cases in city, a town within the Hubei Province of China. It apace unfold, leading to a virus throughout China, followed by associate degree increasing variety of cases in alternative countries throughout the globe. On thirtieth January 2020 India recorded its initial COVID-19 case in state of Kerala. it absolutely was a student United Nations agency had travel history to china. And until the beginning of Gregorian calendar month India has over two hundred thousand confirmed cases.

## III. LITERATURE REVIEW

As per different papers available in literature, there are a few studies that focus on the trend analysis and forecasting for Indian region. The studies on Indian region presents long term and short term trend, respectively. These studies use time series data from John Hopkins University database and present forecasting using ARIMA model, Exponential Smoothing methods, SEIR model and Regression Model. However network modelling and pattern mining are not attempted in these versions of the studies and that too at the regional level, hence the current study attempts to do that. Also, the studies in Indian region from the past are more focused on presenting time series analysis based on



the overall data for Indian region rather than covering other sources of information apart from just considering the number of infected patients, so the need to analyze the patients background and information is required for the authorities to get better insight about the situation. Similarly, there are other mathematical models that were developed for analyzing the trends of COVID-19 outbreak in India. A model [7] for studying the impact of social distancing on age and gender of the patients in India was presented. It compared the country demographics amongst India, Italy and China and suggested the most vulnerable age categories and gender groups amongst all the nations. The study also predicted the rise of infected cases in India with different lockdown periods. Similarly, a network structure approach was used by one of the study [8] to see whether any specific node clusters were getting formed. But only travel data nodes were considered by the authors to check which the prominent regions are affecting Indian travelers coming back to the India. Also, the study presented the SIR model to see the rate of spread of the Corona Virus amongst patients in India. Analysis on the testing labs and infrastructure was also presented by earlier authors. Work of medical doctors and frontline health workers was also presented by some studies [9]. It was found that in India, the role of health workers was less stressed as the spread stage of corona virus was still in phase two or the phase of local transmission rather than the community transmission as compared to other nations like Italy, Spain and USA. However, it was also claimed that Indian healthcare infrastructure is not very strong as per the WHO guide lines and in case of community spread, the Indian government

may find it difficult to manage the spread. Some detailed discussion on the nature of the Corona Virus was also presented by some studies [10][11].

Apart from India, a few models are also available for other countries primarily for China, Italy and USA as the number of infected patients was high. Studied like [12-16] worked on various mathematical models to determine the spread of the disease, predict the number of infected patients, commenting on the preparedness for each country in tackling COVID-19 spread and finding the patterns of flattening curve in different conditions. A lot of researches are still in preprint stage for the world level and are yet to be peer reviewed. With respect to the research activities conducted in the Indian region, the studies are yet to work on the impact of different policies working towards containment of the corona virus. Even in the preprint databases, there are fewer evidences available which worked in the Indian region with more granularities and came up with analysis that can support the decision making of the various administrators in India to curb the lockdown and work on future strategies. Therefore, this study attempts to work on a comprehensive level to analyze the COVID 19 spread in India and impact of various strategies imposed by the Government at both state level and central level.

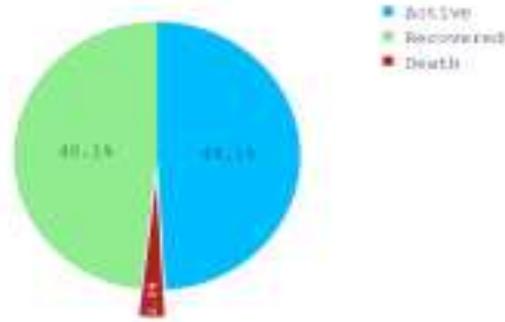
#### IV . FINDINGS AND DISCUSSION

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.



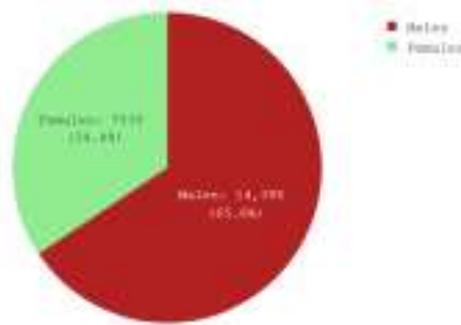
**What is the current situation in India?** As shown above, India had recorded over 500 cases till 24th March. So government declared nation-wide lockdown from 25th march to 14th April also known as lockdown 1.0 and after this government has been extending nation-wide lockdown step by

step. **Situation in India till 7th June 2020.**  
**Confirm Cases : 257487**  
**Recovered : 123848**  
**Active : 126433**  
**Deceased : 7206**



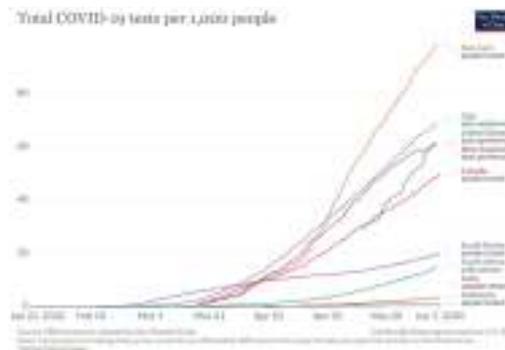
India is showing good recovery rate day by day with low rate of deaths but on the other hand it is also reaching new peak of confirmed cases everyday.

If we talk at the level of patients then as per sample size 21936 cases,



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



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

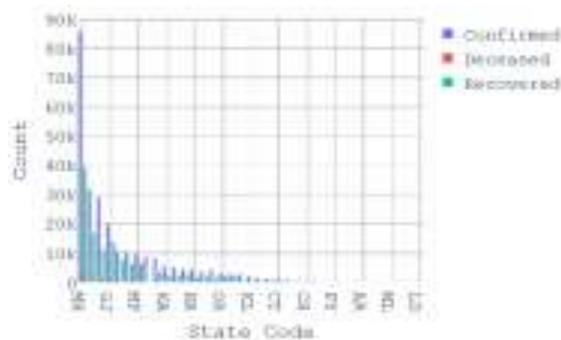
world wide tally of COVID-19 cases in early days. But as of now we are observing 8-9 thousand people being found positive every day.



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 ademography, 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.



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.



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.

### 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).

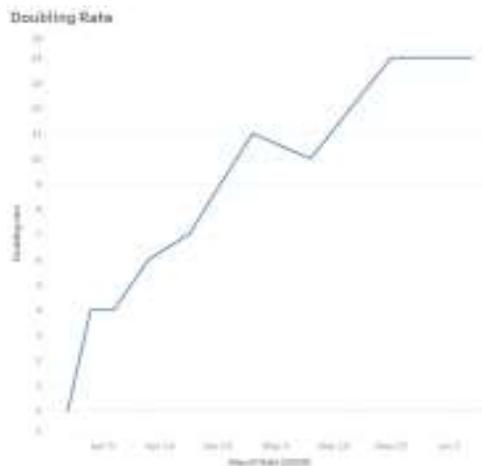


State	Period of Estimation	Estimation R
India	March 4 - April 11	1.83
	April 12-May 16	1.38
	May 17 - May 25	1.33
Andhra Pradesh	April 12-May 16	1.38
	May 17-May 25	1.37
	April 26-May 31	1.35
Bihar	April 26-May 31	1.35
	May 1-May 5	1.35
	May 7-May 11	1.35
Chhattisgarh	April 26-May 31	1.35
	April 26-May 31	1.35
	May 1-May 5	1.35
Goa	May 1-May 5	1.35
	May 7-May 11	1.35
	May 13-May 15	1.35
Haryana	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Jharkhand	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Kerala	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Madhya Pradesh	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Odisha	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Punjab	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Rajasthan	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Tamil Nadu	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Uttar Pradesh	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
West Bengal	May 13-May 15	1.35
	May 13-May 15	1.35
	May 13-May 15	1.35
Andhra Pradesh	April 4-May 1	1.37
Goa	May 13-May 15	1.35

It is believed that India is having steady  $R_0=1.23$  since 25th May which is still higher than normal. While imposing nation-wide lockdown  $R_0$  is considered as one of the factor. If the  $R_0$  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  $R_0$  factor greater

### DOUBLING RATE

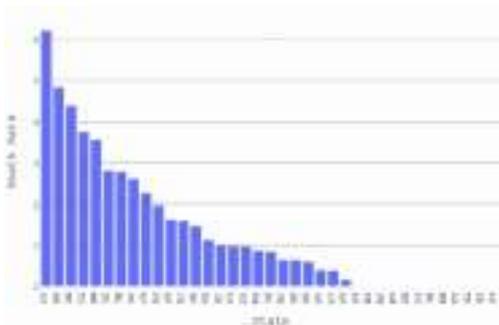
Doubling rate is number of days taken to double the number of confirmed cases in a particular area. This number should be as large as possible. Large number of doubling rate will increase the hopes of situation under control.



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.

### DEATH RATE

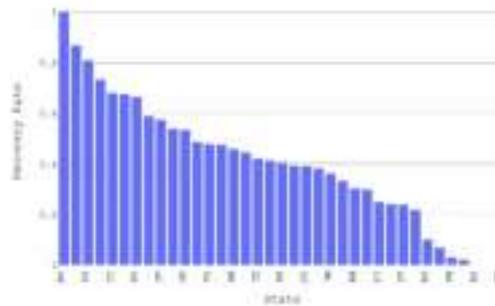
Death rate is calculated by,  $100 * (\text{number of deaths in a state} / \text{number of confirmed cases in a state})$ . The number obtain after this calculation is the number of deaths behind every 100 confirmed cases in that state. Higher the number more the area is to be focused.



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. 2425

### RECOVERY RATE

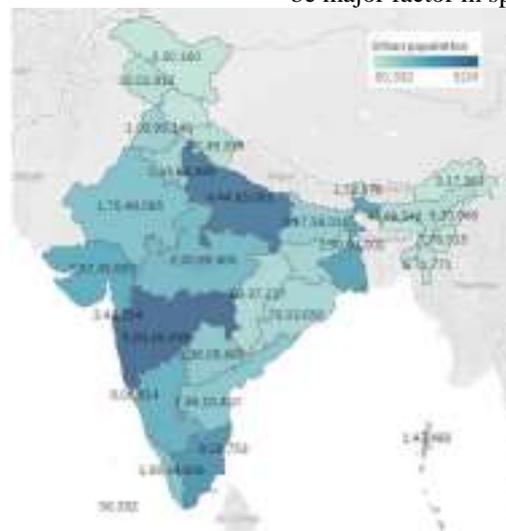
Recovery rate is calculated by,  $100 * (\text{number of recoveries in a state} / \text{number of confirmed cases in a state})$  The number obtain after this calculation is the number of recovered patients behind every 100 confirmed cases in that state.



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 47.48.26

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

- **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.



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

Bengal ,Gujarat, Karnataka ,Madhya Pradesh , Rajasthan.

**Top 8 States with highest urban population:** Maharashtra ,Uttar Pradesh, Tamil Nadu , West

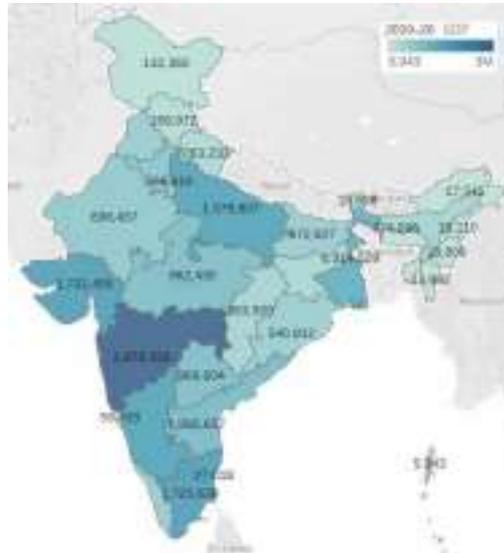
**Top 8 States with highest number of COVID-19 cases:** Maharashtra , Tamil Nadu, Delhi ,Gujarat ,Rajasthan , Uttar Pradesh ,Madhya Pradesh ,West Bengal.

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.

### GDP CONTRIBUTION

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



These are the figures of state and its GDP for the financial year 2019-20 in Crores. Fig graph is state and its confirmed COVID-19 cases. These two graphs too look somewhat 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.

### LOCKDOWN

Is lockdown in India a successful or not? Indian government implemented lockdown in 4 phases. In comparison with other countries, India declared lockdown in its early days of COVID-19 outbreak.

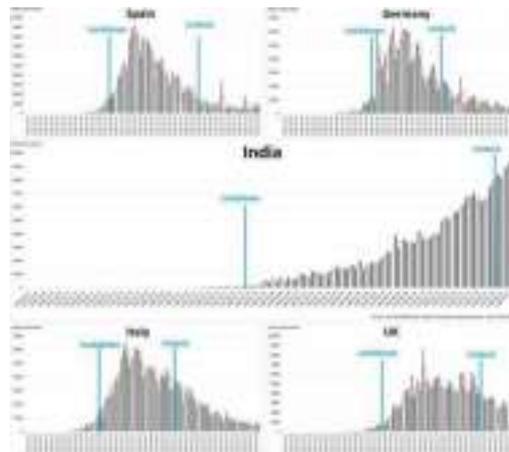
Phase 1: 25th March – 14th April Phase 2: 15th April – 3rd May

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 downward falling of the curve and yet



India has not reached the peak yet. In all the phases of lockdown in India, the cases graph is showing an 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 losing 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.

- India was in lockdown for more than 2 months but still the situation did not get any better. This may be due to weak administration or the violation of lockdown by citizens.

## 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 things can go out of control which will affect the recovery rate.
- Gender information of most of the patients is not released by the government but whatever data is available shows the number of infected males is more than that of females. This may be due to more exposure/contact of males with the 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 a pandemic like this affects the 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.