



# A Comprehensive analysis of covid-19

Sk Abul Fazal<sup>[1]</sup> Gagninder Kaur<sup>[2]</sup>

21MSM3148@cuchd.in, gagninder.e9655@cumail.in

Msc. (data science), department of mathematics,

Chandigarh University, mohali, punjab,

140413, india

---

**ABSTRACT:-** The last two years have been extremely volatile for the entire world. The covid-19, which was discovered in Wuhan, has spread throughout the world. The primary goal of this paper is to analyze the impact of the Corona virus over the last two years and visualize the data in order to gain a better understanding of the data collected from various sources. Also, we discussed some papers that we had gathered in order to review the works of various authors on the covid-19 and its effects. This paper puts an emphasis on the information of COVID-19 'phase 2' which was from February 2021 to December 2021. The outcome of the investigation demonstrates the effect of COVID-19 in India as well as all over the world. Graphs were created using Python libraries like "Matplotlib" and "Seaborn" which produces better results for visualization. Also, we have compared the situation of different countries who are suffering from Covid-19 pandemic with India. In this paper we have also discussed about the doses of vaccine administered in different countries.

**Keywords:-** COVID-19, Data analysis, Data Visualization, Vaccination

---

## 1.1 INTRODUCTION:-

The new coronavirus (COVID-19) was widely reported to have been discovered in December 2019 in Wuhan, China. After the initial pandemic, COVID-19 quickly spread over all provinces in China and then to other Asian and non-Asian countries [1]. Currently, more than 480 million cases of infection have also been finally confirmed in over 180 nations, including over 6.5 million deaths. Fever, tiredness, a dry cough, aches and pains, nasal congestion, etc. are the most prevalent signs or characteristic features of COVID-19. The remarkable thing here is sometimes the people get infected without exhibiting these symptoms or traits and do not even feel ill. COVID-19 is a "communicable" disease that is transmitted through droplets out from nose or mouth when an infected person coughs or outbreathes, that's why you should keep at least one metre (3 feet) away from the sick person. According to current research, COVID-19 is spread majorly through direct contact instead of airborne transmission [2]. COVID-19 protection and prevention can be mitigated by incorporating some simple and easy precautions into everyday lifestyle, like thoroughly sanitising hands with alcohol-based sanitizer or washing them with soap and water, and avoid touching the sense organs like mouth, nose or eyes as hands can act as a carrier for COVID-19 as hands touch multiple surfaces which might be infected and virus can enter our body [3]. The majority of medical research has prioritized research on acknowledging and characterizing the symptoms of COVID-19, and ultimately predicting its incubation periods. According to a review of the literature, many of the patients are asymptomatic while the incubation period scales from 4 to 14 days for this viral infection. Due to the high rate of infection of this disease, it is extremely crucial to forecast and assess the population contaminated with COVID-19. This persuades teams of researchers to investigate on some other perspective of their investigation. The forecasting of COVID-19 cases is the focus of this area. There are two primary ways for prediction used in the literature. The 1st is statistical modeling and the 2nd is analytics [4]. The COVID-19 was measured four phases.

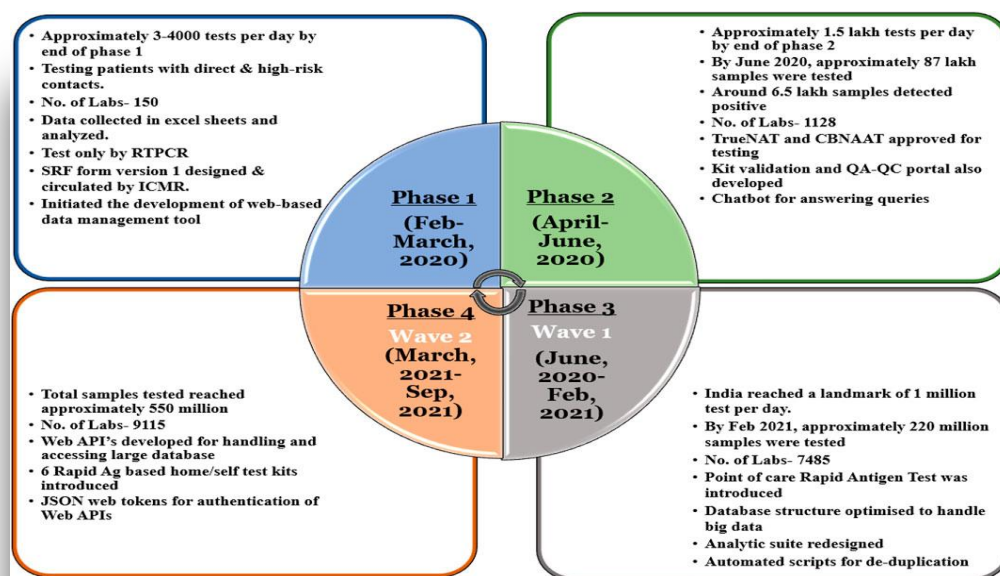


FIGURE. 1

The COVID-19 was at its peak from early 2020 to 2021 causing large amount of deaths for which the scientists work day and night and they found the vaccine against COVID-19 which has accuracy of approx. 91%. In this paper we aim to analyze the changes of situations of covid-19 such as no. of deaths, no. of infected person, etc. Also, how COVID-19 situation has changed from 2020 to 2022. We applied different data analytics model to see the vaccination status and its effect.

## 1.2 LITERATURE REVIEW:-

In [1]The researchers reported various facets of the latest corona-virus disease, visualized the transmission rate of the pathogen, and investigated potential role of predictive analytics to the spreading of this viral pathogen. First, COVID-19 is the subject of a study, which focuses on a range of things including its origin, resemblance to preceding coronaviruses, transmission rate, symptoms, and so on.

Second,to evaluate the propagation of the viral infection, predictive modelling is applied to a dataset from Johns Hopkins University. As of June 4, 2020, the United States had the highest confirmed cases, regardless of the fact that the disease first emerged in China in December 2019. Third, a second – order polynomial algorithm is then used to model the global surge in the prevalence of reported cases over time. Fourth, classification algorithms are used to diagnose COVID-19 using a dataset sample size of 5644 given by a Hospital in Brazil. It is demonstrated that the different libraries of Machine Learning and NLP can effectively categorise COVID-19 patients.

In [2] the researchers have done clinical test for corona virus in 2510 patients in China. They have accumulatedbacterial DNA and laboratory data from 2510 patientsfor retrogressive analysis. After clinical test it was analysed that the percentage rate of 3 diseases COVID-19, influenza A and B is 1.3 percent, 3%, and 3%. COVID-19 bacterial DNA was found in the feces but not in the tears of 8(+ve) patients. Among, the 32 infected patients of COVID-19, 47% and50% had decreased inflammatory cells count and inflammatory cells ratio, 66% and 75%, respectively, had significantly reduced erythrocytes count and erythrocytes ratio, and in 56% C-reactive protein levels had risen. 10 haematological measure were found to be significantly different in the blood of COVID-19 infected patients and those with influenza A and influenza B (P 0.05). 18 haematological measures differed importantly (P 0.05) between COVID-19 +ve and -ve infected persons.So, they madea conclusion that the +ve rate of Influenza-A and B was greater than the level of COVID-19 and Faecal samples

can be examined when subglottic swab collection could result in infection. The use of subglottic swabs and faecal samples can help to increase the effective rate of bacterial DNA detection. Some haematological indices may change as a result of the COVID-19.

In [3] the author has discussed about the trend of COVID-19 in India. The author uses the data till 22 April 2020 and had done an EDA to show how rapidly Corona is increasing. In this paper comparison of India with some other countries has also done to know where is the highest cases of corona in a weekly basis, where China is on the top as of 22<sup>nd</sup> April 2020. Also, the author researches the cases of Corona in different state of India where Maharashtra is on the top of the list. The author analyzes the symptoms of Covid-19 through visualization which helps us to understand the effect of COVID-19 in India.

In [4] the authors had discussed about the increasing case of COVID-19 in India for the year 2020. They have collected the data from different government sources and analyzes the trend to predict about the COVID-19 in the future. Their main focus is when the Corona cases will be on its peak and when it will diminish. To predict the cases, they have used Python and different Visualizations technique. Different Machine Learning Algorithms, Deep learning Algorithms, NLP techniques was used to research the likelihood of a rise in cases and in the nearish future. Data was converted to a logarithmic scale using feature engineering techniques, which leads to improved correlation by isolating any data appendages or outliers. Based on short-term interval predictions, their concept can be optimized to predict long time intervals.

In [5] the researchers have focused on two of the first affected countries of European nations by Corona virus i.e. *ITALY and SPAIN*. They have used two mathematical model 1) SIR (Susceptible(S), Infectious(I) and Removed(R)) model and 2) the Logistic Regression model. They simulate the daily and consecutive occurrences of COVID-19 in the 2 countries during the initial diagnosis. Basic reproduction number was between 2 and 3 for Italy and for Spain it is between 2.5 and 4. Of the two model the logistic regression model was found to be more appropriate for the analysis.

### 1.3 COVID-19 DATA ANALYSIS:-

Because of the rapid growth of COVID-19 in the world, it was necessary to conduct Data Analysis on some datasets collected from various sources such as "Kaggle", "Worldometer", "WHO", and many others. Python was used for visualization. In addition, Python libraries such as "Matplotlib", "Seaborn", "Numpy", "Pandas" and "Pyplot" were used to analyze and visualize the trend and track the outspread of COVID-19 in India and around the whole world. Dataset used for Data Analysis is normalized so that there is no data redundancy. In addition, some columns with missing value and unsolicited columns were removed to improve data visualization.

### 1.4 SPREADING OF CORONA VIRUS OVER TIME:-

In the figure. 2, the X-axis refers to the **Country/Region** those are affected by COVID-19 and the Y axis represents the total **Deaths (in thousand)** from August 2021 to July 2021.

Inferences made from figure 1 are as follows:

- 1) As the title depict, it has top 10 countries with maximum number of deaths and each Country/Region is represented by different color.
- 2) With 1,40,000 US has maximum number of deaths in the given period which is represented by pink color bar.
- 3) As the virus spread from china to the whole world but China is not even in list f top 10 deaths.
- 4) Also, Brazil has 2<sup>nd</sup> most number of deaths (90000) followed by U.K (45000) which is followed by Mexico (40000) and then Italy (38000).
- 5) It is depicted from the graph India which maximum Active cases but its death rate is lower as compared to Cured cases.

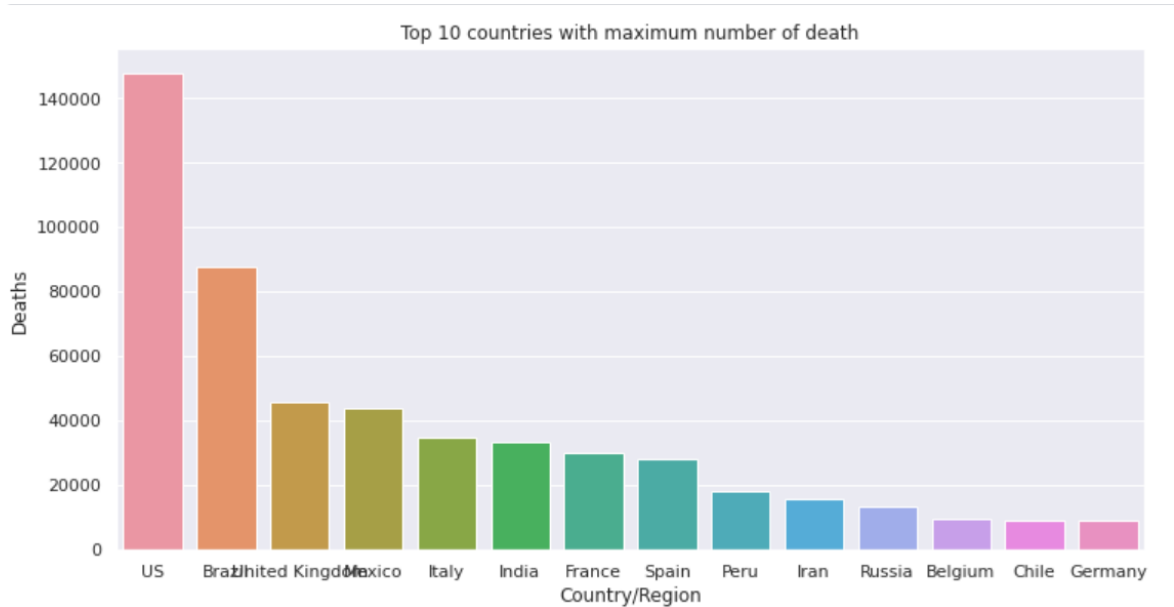


Fig.2

### Cumulative confirmed COVID-19 cases

Due to limited testing, the number of confirmed cases is lower than the true number of infections.

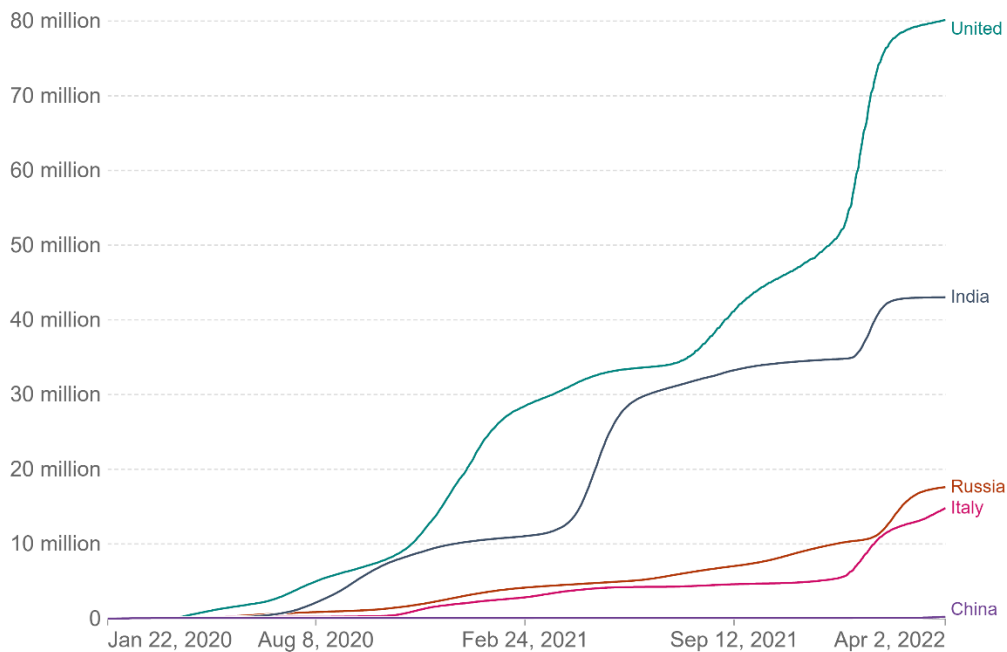


Fig. 3

In figure. 3, four different countries have been compared for the confirmed COVID-19 cases from Jan 2020 to Apr-2022. For this graph different dataset has been taken from **Kaggle**. The X-axis in the graph represents the dates, and the Y-axis stands for the confirmed COVID-19 cases. Green line denotes increase in the confirmed cases in USA, Blue line make up the set of confirmed cases over the period in India.

Brown line represents same for Russia, Pink line depicts the cases in Italy and Violet line is of China.

- 1) As COVID-19 cases first appeared in China but from the Graph it is clearly visible that by the time the number of cases in China were decreasing.

- 2) US turned out to be most affected country among the four others. From Jan 2020, there is a spike in the confirmed cases in US and in Apr 2022 the confirmed COVID-19 cases touched a number i.e. a whopping 80 million.
- 3) Russia and Italy have almost equal number of confirmed COVID-19 cases i.e. Russia has 18 million cases whereas Italy has 16 million cases as of April 2, 2022.
- 4) In that course of time India was just behind the top most in the list of most recorded cases of COVID-19.

In 2020 the total cases were less than Russia and Italy but due to the course of time the number of actual cases has accelerated and as reported in April 2022, India has nearly 45 million confirmed COVID-19 cases.

#### 1.4 COVID-19 CASES IN INDIA:-

In figure.4, we have top 10 states with high strength of active cases in India. The X-axis portrays the State/Union Territory and the Y-axis portrays the Active cases in 2021. Different states assigned different colors for better understanding of the graph.

Following inferences made from the graph:

- 1) The violet color bar represents Maharashtra which has maximum number of active cases as of 2021 i.e. it has 7 lakh active cases which is highest among all the states. Also, it has most number of deaths as compared to other states.
- 2) Karnataka has second highest active cases in India i.e. 6 lakh which is represented in orange color bar followed by Kerala which has 4.5 lakh active cases which is represented in green bar.
- 3) Tamil Nadu and Uttar Pradesh have equal number of active cases i.e. they have 3 lakh cases.
- 4) Among the ten states West Bengal and Chhattisgarh have 1 lakh active cases which is much lower than the others. So, in 2021 the condition of these two states is better than the others.

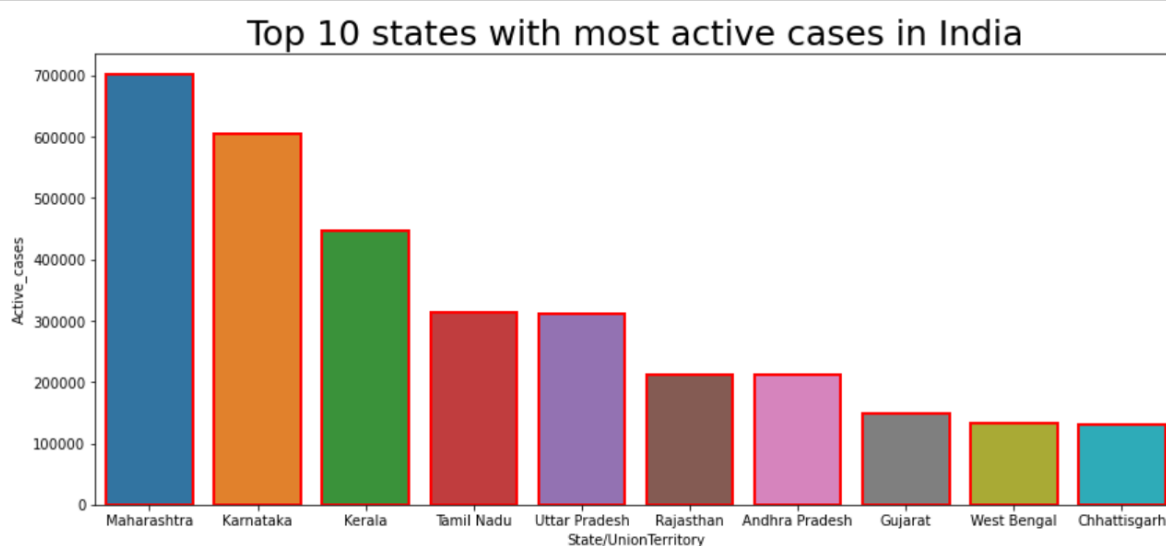


Fig.4 Covid-19 cases analysis In India:

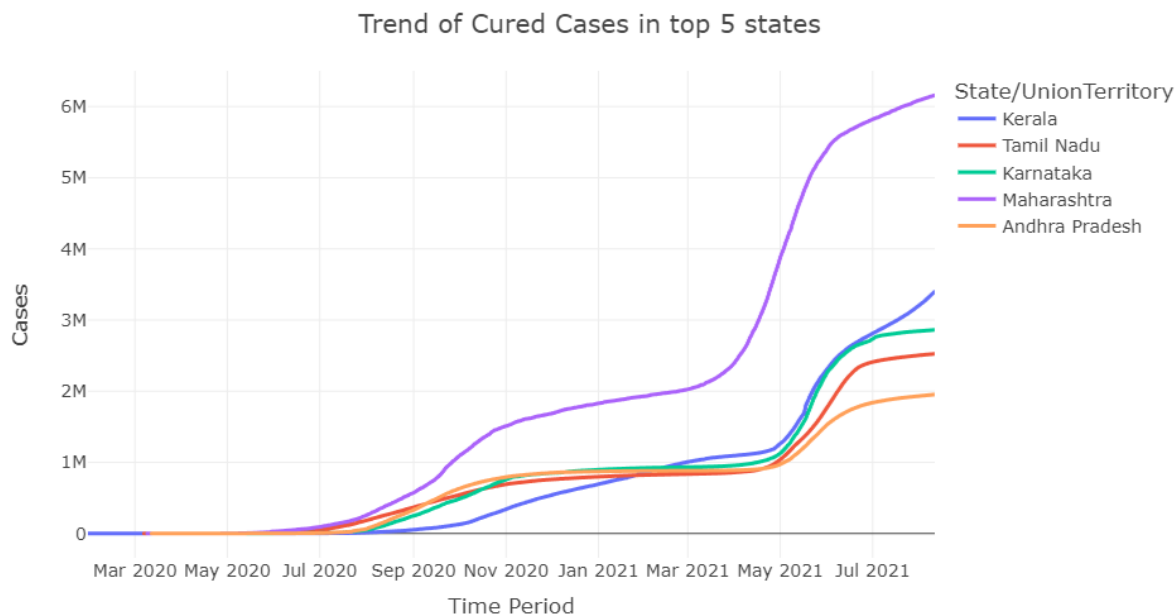


Fig.5

The above graph tells us about the cured cases in top 5 states over time. In Figure.5, the time period is signified by the X-axis, and the total number of cured cases is expressed by the Y-axis.

The violet line represents the cases in Maharashtra, Red line represents the cases in Tamil Nadu, Green line represents the Karnataka, Purple line portrays the cases in Kerala, Orange line describe increase in Andhra Pradesh. Some inferences made from the above graphs are as follows:

- 1) In March 2020, the first case was reported in Kerala and the cases of cured also started from Kerala.
- 2) With the course of time, the number of cured cases increases in Maharashtra which was below 1 million in July 2020 to 6 million in July 2021.
- 3) In the mid 2020 the number of cured cases decreases in Kerala which was about 2 million. But in July 2021 the number of cured cases in Kerala i.e. 3.5 million.
- 4) Tamil Nadu, Karnataka and Andhra Pradesh have nearly equal number of cured cases i.e. between 1.5 million to 2 million.

### 1.5 COVID-19 VACCINATION ANALYSIS:-

The below graph is made in the context to know the total numbers of individuals vaccinated in different states of India.

1) To plot the graph, we have taken top 12 States with maximum number of individuals vaccinated in India. In the graph, the marks were labelled by name of the states and sum of individuals vaccinated.

2) It is clearly visible from the graph that Maharashtra has highest rate of vaccinated individuals followed by Uttar Pradesh and Rajasthan. Also, it is inferred from the above graph that Odisha has least number of individuals vaccinated among the 12 States.

3) It is clear from the graph that many people in different parts of India are aware of the vaccination and they are getting vaccinated as soon as possible.

4) As we can see that the mentioned states have highest rate of vaccinations. It can also be inferred that these states have highest number of affected people by COVID-19 virus.

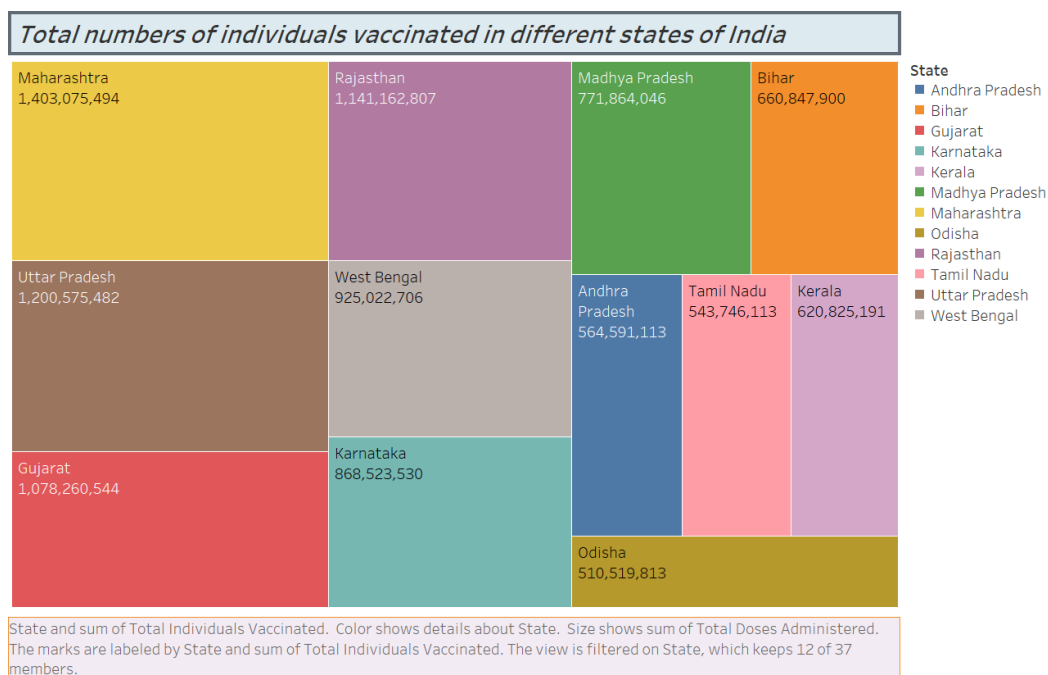


Fig. 6

## 1.6 CONCLUSION:-

The key intention of this paper is to do the analysis on the dataset taken and trace the spread and trends of novel corona virus around the world and in India and also, to visualize it in the graphs using different tools and techniques. Moreover, the study is conducted to analyze which countries are affected more over the recent periods, comparing different countries such as U.S.A, Russia, Italy and China with India that how the confirmed cases in these countries have been increasing in 2021. It also focuses on the number of active cases in India as well as the trend of COVID-19 cured cases in different affected states in India. At last, we have also seen the vaccination trends in different states of India. All these were visualized through graphs and charts using Python and Tableau so that a clear inference could be made for the readers and make this paper easy to understand by everyone.

## 1.7 FUTUREWORK:-

This paper work can be further extended higher level study and analyze more about COVID-19 and its variants in the future. Data Analytics can be used for predictive analysis and to create an advance model through which people can be alert it advance about COVID-19 and its variants affected areas so that this time they will become aware and stop the spread of the disease. Also, Machine learning models can be created to predict the effectiveness of the vaccinations on different people and also to track the people those are affected by disease and also this model can be used to make the world fully vaccinated against COVID-19 and its variants.

## REFERENCES:-

- [1] Mondal, M. R. H., Bharati, S., Podder, P., & Podder, P. (2020). Data analytics for novel coronavirus disease. *Informatics in Medicine Unlocked*, 20, 100374.
- [2] Yun, H., Sun, Z., Wu, J., Tang, A., Hu, M., & Xiang, Z. (2020). Laboratory data analysis of novel coronavirus (COVID-19) screening in 2510 patients. *Clinica Chimica Acta*, 507, 94-97.
- [3] Mittal, S. (2020). An exploratory data analysis of COVID-19 in India. *Int. J. Eng. Res. Technol. (IJERT)*.
- [4] Chintalapudi, N., Battineni, G., & Amenta, F. (2021). Sentimental analysis of COVID-19 tweets using deep learning models. *Infectious Disease Reports*, 13(2), 329-339.
- [5] Sengupta, S., Mugde, S., & Sharma, G. (2020). Covid-19 pandemic data analysis and forecasting using machine learning algorithms. *medRxiv*

- [6] Chu, J. (2021). A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain. *PloS one*, 16(3), e0249037.
- [7] [www.kaggle.com/dataset/covid-19-india](http://www.kaggle.com/dataset/covid-19-india).
- [8] [www.worldometerinfo.com](http://www.worldometerinfo.com).