Analysis of COVID-19 in India using Exploratory Method

Adarsh Satsangi

Computer Science & Engineering

SRM-IST

Delhi-NCR, India

Abstract— **The end of 2019 marked the outbreak of the new pandemic called "COVID-19" and since then number of infected cases are increasing globally, especially in India. India is now the 2nd most infected country and authorities are having a hard time in patterning and forecasting the spread of COVID-19. This paper aims at drawing the better statistical model with deep study of number of reported cases till September 15, 2020 with the implementation of Exploratory Data Analysis. The COVID-19 cases are shown on the basis of various trends for the better case study. The Python language is used and the case study is performed using the various Python Libraries and Future Prediction is done.**

Keywords— *Pandemic, COVID-19, Python, Exploratory Data Analysis.*

I.   INTRODUCTION

"The Novel Coronavirus appeared for the first time in the Wuhan city of China in December,2019 and its report was given to the World Health Organization (W.H.O) by the end of the month. The virus resulted in the state of emergency and also has created a large threat" [1]. When a healthy person comes in the contact of an infected person, the virus is transmitted through the respiratory tract while the roots of the virus are still unclear. The COVID-19 consists of viruses such as SARS and ARDS. Dry cough, fatigue, fever and shortness are breath are some the symptoms of this disease and they arise after 2-14days of contact. The key point to note down here is that most of the people get infected and still they don't show any symptoms of COVID-19 and this shows the abnormal and irregular nature of COVID-19. COVID-19 is easily transmissible disease and it starts with the symptoms of dry cough, cold, fever and may lead to the loss of life as well.

The disease can be minimized by taking some of the precautions such has using masks, using sanitizers, washing hands regularly and maintaining a proper social distance. Proper washing of hands is a key factor here as the hands could touch various contaminated surfaces and could act as the carrier of the virus and may result in the entry of virus in our body. If someone feels unwell then consulting a doctor immediately is highly recommended.

On January 30, 2020 the first case of COVID-19 was reported in India when a student returned to Kerala from the Wuhan City of China. There is no vaccine available till date to cure this disease, Although the Medical Department of many countries are working together for the formulation of vaccine but they have been unsuccessful till date and this motivates us

to perform the Exploratory Data Analysis and analyze the COVID-19 on the basis of the various trends.

II.   STAGES

The Hindu [2] published the reports of World Health Organization (W.H.O) and according to it the COVID-19 has four brief stages.

The first stage started with the reported cases of people with travel history to the affected areas. The second stage, when the family or friends came into the contact with the person with travel history and later the third stage came which made the situation more critical and resulted in the untraceable transmission source and the virus was transmitted to the people with no travel history. The worst of all was the stage four, when the transmission became endemic and uncontrollable. The Wuhan City in China was the first place with the COVID-19 transmission and it even affected the various other developed countries such as U.S.A, Spain and Italy as well. To reduce the impact of this pandemic, an immediate lockdown was implemented and social distancing was also practiced. The was use of masks was made mandatory to control the spread of the virus.[3]
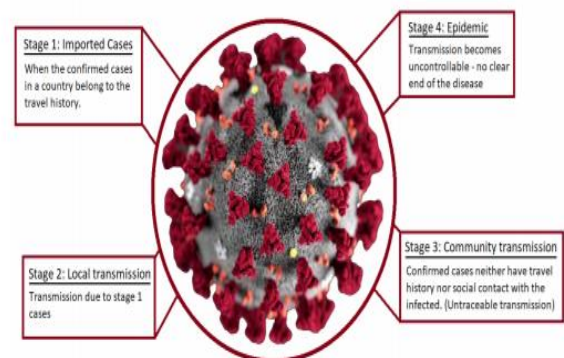

Fig. 1- Complete Transmission Stages of COVID-19

III.   DATA ANALYSIS OF COVID-19 IN INDIA

The outbreak and the rise in the number of C0VID-19 cases in India leads us to do an Exploratory Data Analysis for the current situation on the data obtained from the various sources such as "Ministry of Health and Welfare" [4], "World Health Organization", "Wikipedia" [5] using the trending programming language, Python and thus analyzing the data of India and comparing it with the other countries.

The dataset uses the technique of normalization, filtration for selecting the essential data columns and it also visualizes the data in the proper graphical format. In this paper the "data-preprocessing" and "web-scrapping" is done by using widely used in-trend language "Python" and for the extraction for information from the given dataset and it processing the "pandas" library is used. The "Matplotlib" and "Seaborn" libraries of the Python language resulted in the formation of the accurate graphs.[6]

A. *Spread of COVID-19 in India over time*
The Figure 2 from [7] shows the Time VS Total Coronavirus cases. The time is plotted on the X-axis from February 15, 2020 to September 14, 2020 and the Total count of Coronavirus cases is plotted on the Y-axis (in millions) and the Blue line shows the variation of cases over the given time duration.
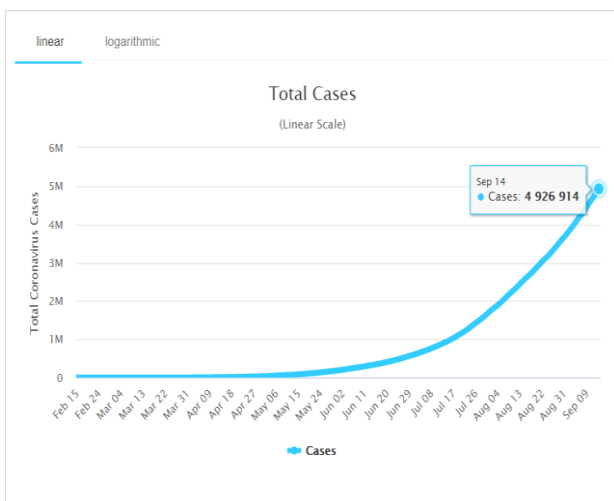

Fig. 2- Spread of COVID-19 in India over time

Inferences: -
1- The COVID-19 cases popped up in India from the January last week to March 15 on a constant scale. After March 15 to April 22 it rose significantly and after April 22 to the current date it is rising on an exponential level.
2- A total of 49,26,914 confirmed cases have been reported till now.

B. *Total COVID-19 deaths in India over time*
The Figure 3 shows the Time VS Total Coronavirus Deaths. The time is plotted on the X-axis from February 15, 2020 to September 14, 2020 and the total count of Coronavirus deaths is plotted on the Y-axis (in thousands) and the orange line shows the variation of deaths over the given time duration.
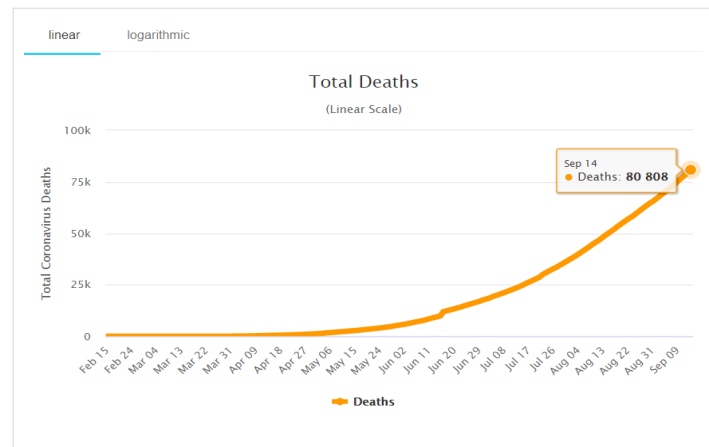

Fig. 3- Total COVID-19 deaths in India over Time

Inferences-
1- The COVID-19 deaths popped up in India from the April first week to May 06 on a constant scale. After May 06 to May 24 it rose significantly and after May 24 to the current date it is rising on an exponential level.
2- A total of 80,808 deaths have been reported till now.
3- The total deaths are increasing day by day in India.

C. *Total Active cases in India over Time*
The Figure 4 shows the Time VS Total Coronavirus Active Cases. The time is plotted on the X-axis from February 15, 2020 to September 14, 2020 and the Total count of Coronavirus active cases is plotted on the Y-axis (in thousands) and the aqua line shows the variation of active cases over the given time duration.
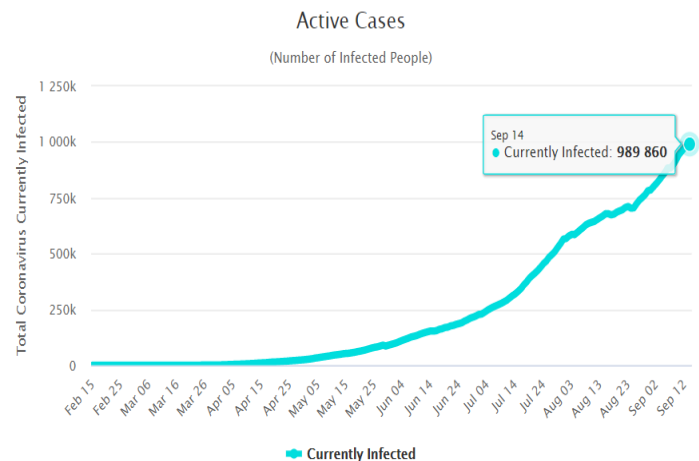

Fig. 4- Total COVID-19 active cases in India over time

Inferences-
1- From first week of March, COVID-19 cases started to pop up in India, till May 05 they increased on a constant level and then they increased significantly and on May 15 it rose exponentially and till current date, September 15, 2020 it is still rising exponentially.
2- A total of 9,89,860 cases are currently active.

D. *Spread of COVID-19 on the basis of Age*
We use a Histogram in Figure 5 to display the number of COVID-19 cases in India on the basis of age groups [8]. The

horizontal line shows the grouping on the basis of age groups whereas the vertical line shows the frequency of number of cases in percentage.
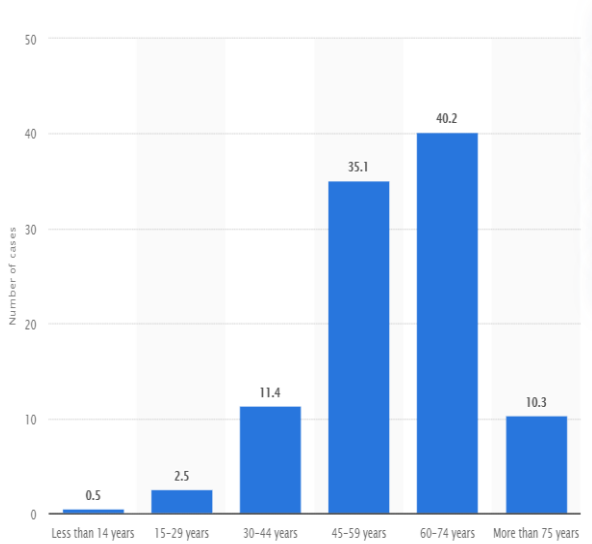

Fig. 5- Spread of COVID-19 in India on the basis of Age

Inferences-

1- The people of age group 60-74 years are affected the most and contribute 40.2% to the total.
2- The people of age group 45-59 years are affected the 2nd most and own 35.1% to the total.
3- The people of age group 30-44 years add about 11.4% to the total.
4- The people of age group above 75 years are also infected and make it up to 10.3%.
5- The number of infected people between age group 15-29 years are very low and the count is only 2.5%.
6- Only 0.5% percent of people are infected between the age group 0-4 years.
7- "In comparison to a healthy person, the people suffering from Blood Pressure, Diabetes or Cancer are at a higher risk of getting infected." [9]

E. State-Wise spread of COVID-19 in India

Figure 6 shows the Map of India along with the number of COVID-19 cases reported by each state.

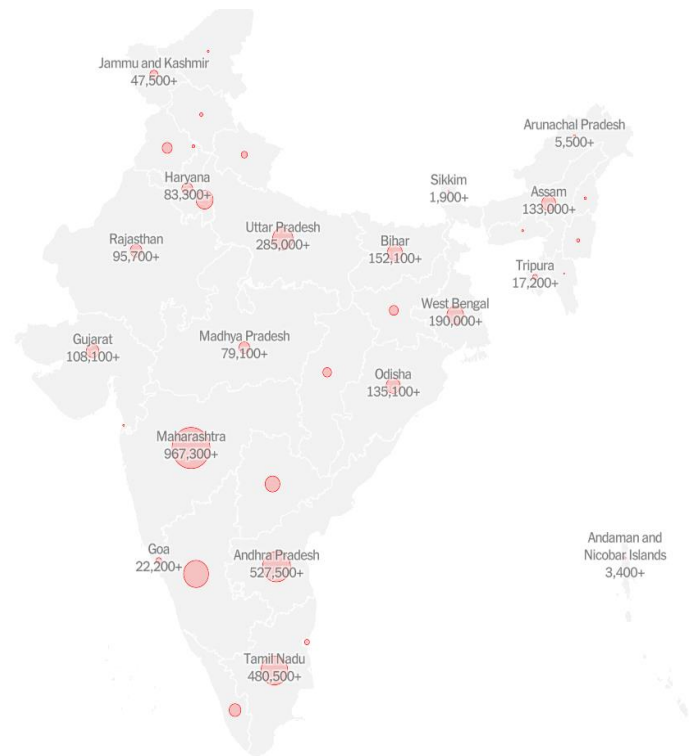The bigger the bubble shown, the higher the number of COVID-19 cases reported by that state.

Fig. 6- State-Wise spread of COVID-19 in India

Inferences-

1- Maharashtra is the most affected state in India and has a total around of around 10lacs cases.
2- Andhra Pradesh is the 2nd most affected state with around 5.3 lacs COVID-19 cases.
3- Tamil Nadu is 3rd most affected state with around 5 lacs cases.

F. Gender-Wise spread of COVID-19 in India

Figure 7 uses a Histogram to display the number of COVID-19 cases in India as per the gender. The horizontal line classifies the gender as Male and Female whereas vertical line shows the frequency of cases on a scale of 1 lac.
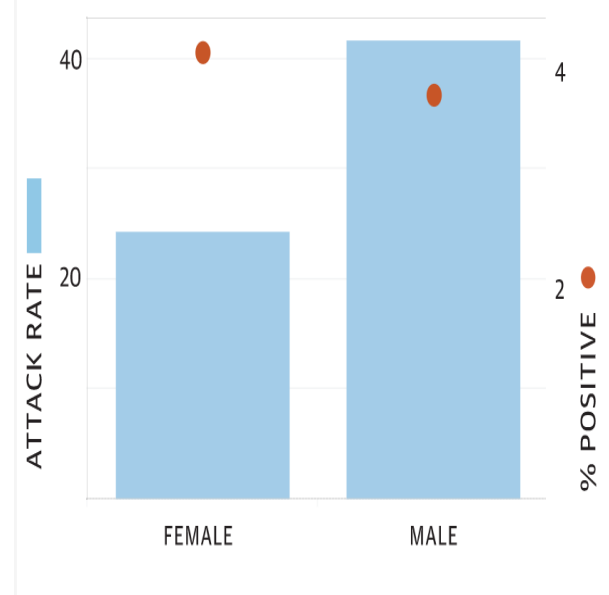

Fig. 7- Gender-Wise spread of COVID-19 in India

Inferences-
1- The attack rate is higher for males i.e. around 41.6 cases per 1 lac males.
2- The attack rate is lower for females and is around 24.3 cases per one lac females.[10]

G. Effect of COVID-19 on India's GDP

The outbreak of the COVID-19 has marked a great negative impact on India's GDP and its representation has been shown in the Figure-8. The horizontal line classifies the group of various months and the vertical line shows it growth.
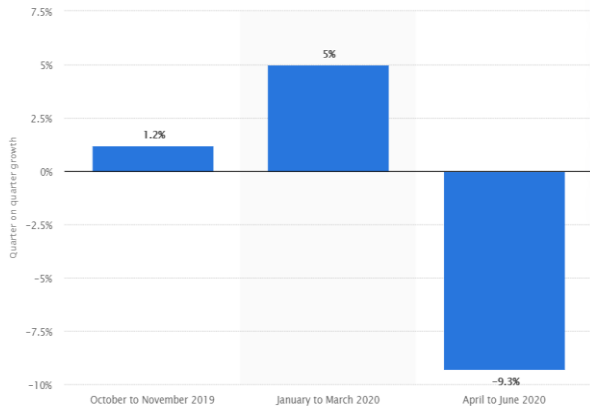

Fig. 8- Effect of COVID-19 on India's GDP

H. Future Prediction on the basis of Data Analysis

We use the "Auto Regressive Integrated Moving Average" model on the various time series data set that we scrapped and we obtained the Figure-9. It is expected that India will follow an exponential curve as shown and there would be a constant increase in the number of COVID-19 cases. India will cross the 55 lacs cases mark by the September 21,2020. By the September 25, 2020 there would be around 59 lacs cases in India. The Figure-9 shows the Dates on the X-axis with an interval of two days whereas the Y-axis shows the total number of COVID-19 cases. The black colored line shows the variation of total cases with date.
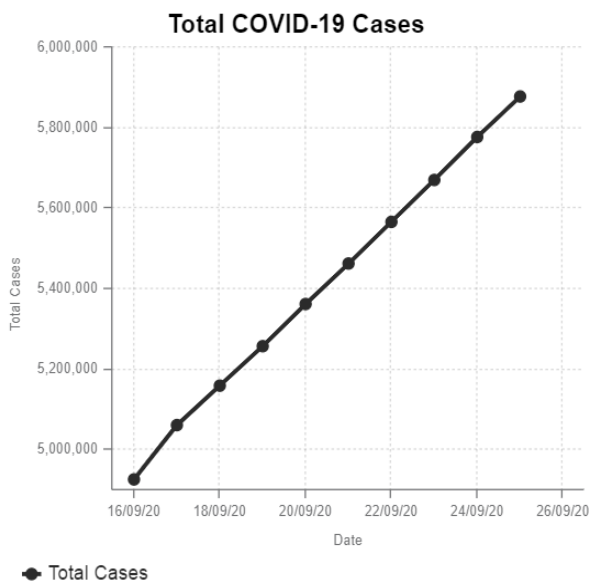

Fig. 9- Prediction of Total COVID-19 cases

I. Tools and Techniques Used

The dataset uses the technique of normalization, filtration for selecting the essential data columns and it also visualizes the data in the proper graphical format. In this paper the "data-preprocessing" and "web-scrapping" is done by using widely used in-trend language "Python" and for the extraction for information from the given dataset and it processing the "pandas" library is used. The "Matplotlib" and "Seaborn" libraries of the Python language resulted in the formation of the accurate graphs. The graph shown in Figure-9 has been framed by using the Python language in the Jupyter Notebook.[11]

IV. CONCLUSION AND FUTURE IMPLEMENTATION

The various patterns of COVID-19 are analyzed using this paper followed by a complete case study. It also helps in studying about the total COVID-19 cases in India, the active cases, the death ratio, the state-wise spread, the gender-wise spread and also the effect of COVID-19 on the India's GDP. At last its analysis the data and helps in predicting the total COVID-19 cases in India over the next 10 days.

Moreover, various Machine Learning Algorithms can also be applied to each individual graph for a better predictive model and also a better predictive model can also be done with more amount of data.

V. REFERENCES

[1] https://www.who.int/
[2] The Hindu https://thehindu.com
[3] Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal, "COVID-19 Epidemic Analysis Using Machine Learning and Deep Learning Algorithms" in medRxiv;
[4] https://www.mohfw.gov.in/
[5] https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India
[6] https://bedford-computing.co.uk/learning/wp-content/uploads/2015/10/Python-for-Data-Analysis.pdf
[7] https://www.worldometers.info/coronavirus/country/india/
[8] https://www.nytimes.com/india
[9] https://www.marketresearch.com/
[10] Y. Deng, F. He, W. Li, "Coronavirus disease 2019: What we know?," J Med Virol,March 2020; https://onlinelibrary.wiley.com/doi/10.1002/jmv.25766
[11] Vinay Chamola, Vikas Hassija, Vatsal Gupta, Mohsen Gulzani, "A Comprehensive Review of the COVID-19 Pandemic and the Role of the IoT, Drones, AI, Blockchain, and 5G in Managing its Impact" in IEE; https://www.researchgate.net/publication/341566740_A_Comprehensive_Review_of_the_COVID-19_Pandemic_and_the_Role_of_IoT_Drones_AI_Blockchain_and_5G_in_Managing_its_Impact