Covid-19 Data Analysis

1Ayush Khandelwal, 2Gaurav Kothari, 3Brijesh Kumar Singh
Assistant Professor*Information Technology, Jaipur Engineering College and Research Centre, Sitapura, Jaipur-302022, India

Abstract- The spread of coronavirus in India is increasing at a very fast speed. The Government of India is having a hard time on how to deal with or decrease the spread of covid-19 in India. The main objective of this research is to perform analysis on the covid-19 data and gain insights about the data by using exploratory Data Analysis(EDA). This research includes India’s analysis of covid-19 cases till October 2020. The result of the analysis shows the impact of COVID-19 in India on daily and weekly basis, and also the impact of covid-19 cases in different states of India, and what are the most common symptoms of covid-19, and also what are the measures that we can take to prevent being affected from Covid-19.

Keywords— COVID-19, exploratory data analysis , India’s analysis, symptoms, measures

I. INTRODUCTION
Coronaviruses (CoV) are an oversized cluster of viruses which can cause health problems starting from the cold to more severe diseases like Middle East Respiratory Syndrome and Severe Acute metabolic process. A coronavirus (CoV) is a new virus that has never been identified in humans. COVID-19 is the illness caused by a new coronavirus known as SARS-CoV-2. Several known coronaviruses have been found in animals that haven’t yet found in humans. The most common signs of covid-19 virus are fever, dry cough and tiredness. The less known symptoms of covid-19 are aches and pains, pharyngitis, diarrhea, conjunctivitis, headache, loss of taste and smell, a rash on skin and color transformation of fingers and toes. The serious symptoms of covid-19 include difficulty in breathing, chest pain, loss of speech or movement. We should minimize any close contact with anyone who is having respiratory problems like cold and cough. As many of us doesn’t know about the health of other people, so it is possible that we may catch covid-19 from them who are having mid cough or who does not feel illness. To get protection from and prevention from COVID-19 you can follow the number of straight-forward steps such as: (i) Continuously sanitizing your hands with an alcohol based sanitizer or washing your hands with soap, (ii) Do not touch your eyes and nose as it will cause the virus to enter your body, (iii) If you are having cold or cough and feeling unwell, then stay at home and take advice from the doctor for proper consultation, (iv) Do not go outside your house if it is not necessary and if your work can be done from home, (v) Do follow the guidelines of the Government as they have maximum information about the disease.

II. LITERATURE SURVEY
In paper [1], the researchers have implemented exploratory data analysis on Covid-19 data. The analysis focuses on the statistical information about covid-19 such as what are the number of confirmed cases, deaths, recovered cases and active cases in different states of India.

In paper [2], the authors review and analyze the covid-19 spreading statistics in different countries using the Bailey’s method. The World Health Organization reports were considered when performing the analysis of different countries.

In paper [3], the authors have performed data visualization techniques using different data visualization libraries in python to get the graphical representation of the covid-19 cases. This also involves visualizing the covid-19 cases in the lockdown period and how the cases had been decreased gradually in that period.

In paper [4], the researchers study and analyze the influen of air temperature on the transmission of covid-19 and they find out that the arrival of summer and rainy season can to some extent reduce the transmission of covid-19. They also observed the different symptoms that are caused due to covid-19.

In paper [5], the researchers study and analyze the origin of covid-19 as well as it covers the entire “coronavirus cycle” in great detail.

III. IMPORTANCE OF COVID-19 INDIA’S DATA ANALYSIS
The main objective of this analysis is to keep the track of the amount of covid 19 cases in India. This analysis tells the number of active cases, confirmed cases, deaths and fatality rate in India and what are the current trends in the rise of covid-19 cases and the prediction of the number of covid cases in India.
IV. COVID-19 INDIA’S DATA ANALYSIS

The Covid-19 outbreak in India has motivated me to perform EDA on covid-19 cases in India and also perform data visualization with different types of charts and graphs available through the use of matplotlib and seaborn library, and also other data visualization libraries. I have taken data from “Kaggle” and thus analyzed the spread and trend of the COVID-19 in India. The dataset consists of the covid-19 cases such as confirmed cases, active cases, recovered cases and more. This exploratory Data Analysis uses “Python” for data analysis and pandas library to extract the data from the covid-19 dataset. Appropriate graphs were created to better visualize the covid cases with the help of various data visualization libraries in python.

A. COVID-19 Cases in India over Time

In Figure 1, the X-axis represents the Dates from March 2020 till 19th May 2021 and Y-axis represents the total number of cases. Red line depicts the number of Confirmed cases(total cases), Gray line depicts “Deaths” which represents the number of people who had lost their lives, Green line depicts the number of cured(recovered) cases which represents the total number of people who have recovered.

Figure 1:- Total Cases

B. Age-Wise Spread of COVID-19 Cases in India

The inferences from Figure 2 can be as shown:-

1. The bar chart analyzes the spread of covid-19 in India to understand that what are the age groups that are most affected with covid-19.
2. The person having age between 20 to 40 are most affected with covid-19.
3. The older age group is very less affected from this pandemic.

V. STATE-WISE ANALYSIS OF COVID-19

The first covid-19 case in India was reported in Kerela on January 30, where the student who returned from Wuhan, China tested positive. After that, 2 more students who had returned from Wuhan, China had tested positive. From there onwards, the spread of covid-19 in India is increasing at a significant speed. In figure 3, we can see the states of India with the statistical information about covid-19 such as confirmed, death, active, cured cases, cure rate(per 100) and death rate(per 100). From Figure 3, we may draw the inference that Maharashtra has the highest number of cases having 5433506 confirmed cases, 83777 deaths and 4927480 recovered cases having 1.54 death rate (per 100) and 90.69 recovery rate (per 100). Beyond this Karnataka has 2272374 confirmed cases followed by Kerela, Tamil Nadu, Uttar Pradesh, Andhra Pradesh, Delhi, West Bengal, Chhattisgarh and Rajasthan respectively.

Figure 2:- Age Wise Spread of COVID-19 Cases

Figure 3:- Statewise Statistics of COVID-19
VI. COVID-19 CASES IN LOCKDOWN PERIOD

The lockdown introduced by the government had resulted in the decrement of the covid-19 cases in most parts of the country. The lockdown period has put a hope of certainty that the country will definitely get win over this pandemic if we follow all the rules and guidelines issued by the government.

Figure 4: Daily New Confirmed Cases in Lockdown period

VII. SYSTEM SPECIFICATIONS

Software Implementation: -
PYTHON, DATA ANALYTICS and DATA VISUALIZATION are used in analyzing and visualizing the COVID-19 cases of the country.

Python: -
Python is a multi-purpose(web, GUI , Scripting, etc.) language. Python is an object oriented language and supports all the features of OOPS. Python is an interpreted language and is strongly typed and dynamically typed language. Python focuses on readability and productivity.

Data Analytics : -
Data analysis is outlined as a process of cleanup, remodeling, and extracting data to find out helpful information for businesses and various IT Companies. The aim of Data Analysis is to achieve insights from the information and from that answers the queries associated with the data that we’ve extracted.

Data Visualization :-
Data visualization is the process to show the data in a graphical format, chart format or other visual format. Data visualization is important because it allows us to see relationship and trends in the data in a more concise way.

VIII. CONCLUSION:

The main objective of this research is to review and analyze the spread of covid-19 in India since the day of arrival and also observe how the covid-19 is spreading in different states of India, and to understand why the Government of India and Healthcare sector are having a difficulty in controlling the spread of COVID-19. Moreover, to study the spread of covid-19 in various states by analyzing the total cases and the Age wise spread of covid-19, how the different age group people are affected from this outbreak, and finally the analysis of covid-19 in India during lockdown period and how the lockdown has helped seen a sudden decline in the number of covid-19 cases in India.

This research can often be extended to a better level by doing more analysis on covid-19 data and further applying machine learning algorithms to predict the amount of covid-19 cases in the near future.

ACKNOWLEDGEMENT

We would like to thank Mr. Arpit Agarwal and Dr. V.K. Chandna for giving us the facilities and providing us with a propitious environment for working in college. We would wish to express deep gratitude to Dr. Smita Agarwal, Head of Department (Information Technology), Jaipur Engineering College & Research Centre, Jaipur with whose support the research has been made possible. We are indebted towards our supervisors who have devoted their precious time and advice during the period, which is imminent to this research.

REFERENCES


AUTHOR INFORMATION

Ayush Khandelwal, Department of Information Technology, Jaipur Engineering College and Research Centre.

Gaurav Kothari, Department of Information Technology, Jaipur Engineering College and Research Centre.