iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 9 **Issue:** VII **Month of publication:** July 2021

**DOI:** https://doi.org/10.22214/ijraset.2021.36286

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 9 Issue VII July 2021- Available at www.ijraset.com*

# An Exploratory Data Analysis of Covid-19 Outbreak Situation in India

Arpit Verma[1], Kajal Rai[2], Muskan Goel[3], Shubhangi Sharma[4], Ujjawal Goel[5]
[1]*Department of Electronics & Communication Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India*
[2, 3, 4]*Department of Computer Science and Engineering,Raj Kumar Goel Institute of Technology & Management,Ghaziabad, India*
[5]*Department Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India*

***Abstract:*** ***The number of COVID-19 cases in India is increasing expeditiously. The National and native authorities are having a tough time to make a pattern, analyze and forecast the spread of COVID-19 in India. The main focus of this paper is to draw a statistical model for better understanding of COVID-19 spread in India by thoroughly studying the reported cases in the country till 14 March 2020.*** **An** ***Exploratory Data Analysis (EDA) technique is being implemented to review and analyze the reported COVID-19 cases in India. The results of the analysis divulge the impact of COVID-19 in India on daily and weekly manner, analysis of different states of India, analogize India with abutting countries also like the countries who are badly affected.***
***Keywords: COVID-19, Statistical Analysis, Exploratory Data Analysis, Social Distancing, India's Analysis, Abutting Countries Analysis, Trend Analysis***

## I. INTRODUCTION

COVID-19 is a contamination belongs to the "Nidovirus family", or "Nidovirales" which covers "Coronaviridae", "Artieviridae" and "Roiniviridae" family, responsible for respiratory illness in humans which may cause common cold to more harsh diseases such as MERS(Middle East Respiratory Syndrome) and SARS(Severe Acute Respiratory Syndrome). The most common manifestations or features of COVID-19 are fever, exhaustion, dry cough, aches and pain, nasal congestion, runny nose or sore throat. The main thing to note here is that some people get infected and don't get these symptoms or traits and doesn't feel unwell. All age group people who has a medical record of blood pressure, cardiovascular disease or diabetes are more prone to get infected and if anyone with fever, cough and breathing difficulties should immediately seek for medical attention. COVID-19 is a "communicable" disease, it means it spreads when a person comes in contacts with a COVID-suffered person and can be passes through the droplets from nose or mouth when an infected person coughs or exhales and this is the main reason to maintain 1metre (3 feet) distance from the unwell person. Studies till date designate that COVID-19 is mainly spread through contact rather than channeled through air. As many people only experienced mild symptoms so it is a high probability to catch COVID-19 from the person who has mild cough or doesn't feel ill.

Prevention of spreading COVID -19 can be minimized by including some of the simple and easy precautions that everyone should adopt in their daily habits which include thoroughly cleaning hands with sanitizer (containing alcohol) or washing them with soap and water. Avoid touching eyes, nose and mouth as hands touches several surfaces which might be contaminated and hands could act as a carrier for COVID- 19 and virus can enter in our body, so please stay home if you feel unwell and most importantly avoid travelling as much as possible. Stay updated to National and local authorities only as they will have the most up to date information about the situation.

On 30 January 2020, India announced its first coronavirus case in Kerala when a student returned from Wuhan (epicenter of coronavirus) and till then the number of cases has been increasing exponentially. In a short time ago, there is no vaccine or medicine obtainable particularly for treatment of COVID-19 and currently are under investigation. This paper examines the trend of COVID-19 based on certain criterion using "Exploratory Data Analysis". Exploratory Data Analysis (EDA) is the way to traverse the data with the aim of extracting convenient and actionable information from it. EDA is the explanatory step in any kind of analysis.

## II. LITERATURE SURVEY

In [1], the researchers analyzed the transmission trend of COVID-19 from China to other countries, confirmed cases on day to day , surveillance strategy of India, Italy, Korea and Wuhan from the first day of outbreak along with the effect of government policies of these countries in controlling the COVID-19 outbreak by finding the relation (linear relation) between outbreak condition and CFR(Case Fatality Rate) by taking global daily statistics such as confirmed, death and recovered cases and making prediction with respect to China using "Linear Regression".

In [2], authors describes Coronavirology (the research performed in this field) with the overview of coronavirus facsimile and pathogenesis along with the evolution of coronavirus, the organ cultures and cells preparation, also as techniques for analyzing the virus function, commonly used reverse genetic techniques of coronavirus and virus cell fusion also as titration techniques, identification of cellular receptors and virus cell fusion along side visualization of virus replication complexes and covers the "coronavirus life cycle" in great detail.

In paper [3], researchers study and surveyed the COVID-19 virus spreading statistics from the cases of various countries using "Bailey's model". High correlation coefficients (91.4%) were resulted using Pearson correlation method and determinants (83.98) were also considered for the accurateness of the model. "World Health Organization's" daily report were considered were considered for analyses of 204 countries and also indicates the difficulties in correctly predicted the longer term spread-reduction variable of the pandemic.

In [4], authors review the "pathogen", "clinical features", "diagnosis" and "treatment" of COVID-19 and also explain the "epidemiology" and "pathology" supported current evidence and recommend that symptoms, exposure history and manifestation on chest CT imaging might be consider as a clinical diagnosis within the COVID-19 affected areas where there's shortage of "RT-PCR" testing kits. Further the crucial role of "S-protein" is additionally depicted for COVID-19 as S-protein mediates receptor binding and membrane fusion which is crucial for transmission is additionally explained along side the suggestion that transmission mode is human to human and majorly it gets transmitted through droplets and shut contact. The detailed analysis of 14 days "quarantine period" is additionally clearly explained as 95% of individuals experience symptoms within 12.5 days of contact.

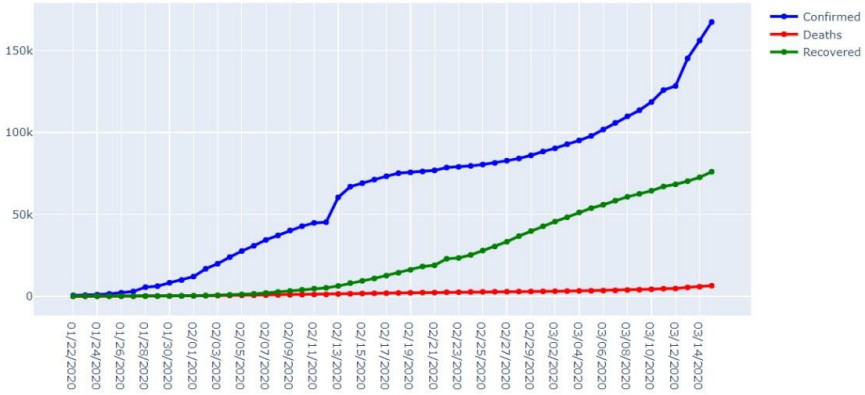
The paper [5] investigates the influence of air temperature and ratio on the transmission of COVID-19 by calculating the "effective reproductive number"(R) and under "Linear Regression" framework they found out that one-degree Celsius rise in temperature and one percent expand in the relative humidity lower R by 0.0225 and 0.0158 respectively and provides a signal that arrival of summer and rainy season in the northern hemisphere can effectively bring down the transmission of COVID-19.

## III. COVID-19 INDIA's DATA ANALYSIS

Current COVID-19 outbreak inspires to do an EDA on the datasets, scraped from different sources such as "Ministry of Health and family Welfare" [6], "COVID-19 India website" [7], "John Hopkins GitHub repository" [8], "Worldometer" [9], "Wikipedia" [10] using "Python" and consequently examining the spread and trend of the COVID-19 in India and comparison with the neighboring and other affected countries of the world. The dataset that uses EDA go through the process of normalization, selecting the necessary columns using filtering out, deriving new columns, and visualizing the data in the graphical format. This paper uses Python for Data Processing and Web Scrapping, Pandas, Matplotlib library to process and extract information from the available dataset. Appropriate graphs were created for better visualization are the results of "Plotly" library of the Python.

### *A. COVID-19 Spread in India over time*

In the Figure shown below, the X axis represents the Dates on an interval of Days from 22 Jan 2020 till 14 March 2020 and Y axis represents the number of cases (in thousands). Blue line shows "Confirmed cases" (positive cases), Red line shows "Deaths" representing the number of cases who had lost their lives, Green line shows "Recovered" which depicts the count of people who has recovered and, the difference of Deaths and Recovered from Confirmed cases.



Inference from the Figure is as follows:

1) From last week of January, COVID-19 cases started to pop up in India, till January 28 the number of cases had been increasing on a constant level but after that it started increasing significantly and from February 11 it have risen exponentially.
2) Approximately 50 thousand "confirmed" cases has been registered till Feb 12 out of which 150 thousand are currently active.
3) Out of 50 thousand confirmed cases many cases have been recovered and some resulted in death.
4) As on 22 April 2020, India has registered around 21 thousand cases out of which 700 has been resulted in death and around 4300 has been recovered leaving around 16 thousand as active and these numbers are rising on regular basis.

*B. COVID-19 Spread in India vs Other Countries*

In the Figure shown below, the X axis represents the Dates and Y axis represents the number of cases. "Orange" line represents the cases in Italy, "Green" in Korea, "Purple" in Wuhan and "Blue" in India.

Inference from the Figure below is as follows:

1) All the countries started to detect COVID-19 cases from the month of January end except China which tells us that the epicenter of the global virus is China.

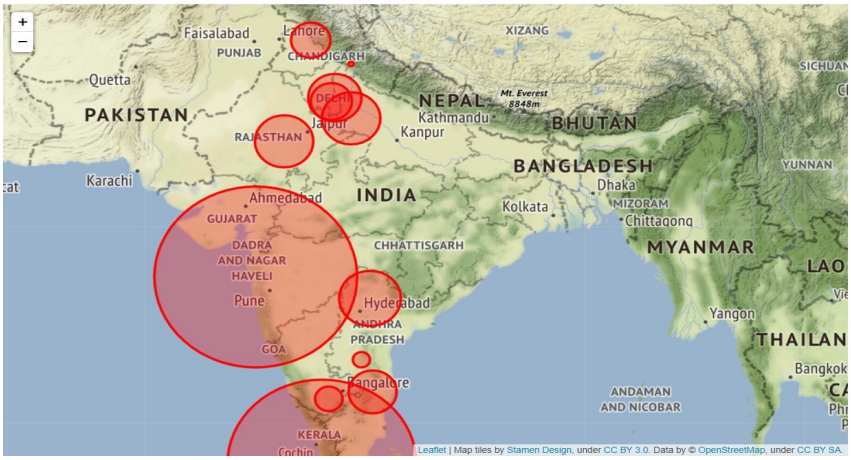

2) China(Wuhan) shows a growth in number of cases from January till mid Feb but then the amount of cases has been constant within the China.
3) Italy shows about no growth in number of cases from January till mid of February. On the other hand, the quantity of cases has touched there approximately 40 thousands till 15 March 2020.
4) Korea shows about no growth in number of cases from January till mid of February like Italy. On the other hand, the quantity of cases has touched there approximately 8 thousand (less than the cases of Italy) till 15 March 2020.
5) In India, the condition is best compared to other countries because the number of cases has touched only around 300 cases which is extremely less as compared to other countries till March 15 2020.

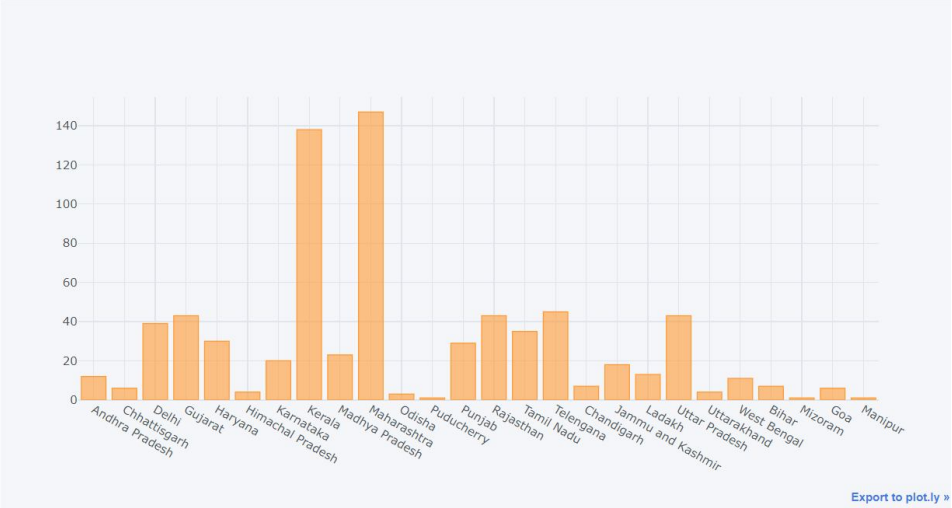
*C. State-wise Analysis of COVID-19*

Kerala reported the primary coronavirus case in India on January 30 when a student who had returned from Wuhan. Till the month of February , two more students were tested positive after their return from China. Table depicts the different states of India having their number of Total Cases. Inference from the figure are often drawn that Maharashtra tops the list by having 147 Total cases, then Kerala having 138 Total cases followed by Telengana, Rajasthan, Gujarat, Tamil Nadu, Uttar Pradesh, Delhi, Punjab, Madhya Pradesh, Haryana, Jammu and Kashmir, Karnataka, Ladakh, Andhra Pradesh, West Bengal, Chandigarh, Bihar, Chattisgarh, Goa, Uttarakhand, Himachal Pradesh, Odisha, Manipur, Mizoram and Puruchery.

Table 1 : States of India having the number of Total cases

| Name of State | Active Cases |
|---|---|
| Maharashtra | 147 |
| Kerala | 138 |
| Telengana | 45 |
| Rajasthan | 43 |
| Gujarat | 43 |
| Tamil Nadu | 35 |
| Uttar Pradesh | 43 |
| Delhi | 39 |
| Punjab | 29 |
| Madhya Pradesh | 23 |
| Haryana | 30 |
| Jammu and Kashmir | 18 |
| Karnataka | 20 |
| Ladakh | 13 |
| Andhra Pradesh | 12 |
| West Bengal | 11 |
| Chandigarh | 7 |
| Bihar | 7 |
| Chattisgarh | 6 |
| Goa | 6 |
| Uttarakhand | 4 |
| Himachal Pradesh | 4 |
| Odisha | 3 |
| Manipur | 1 |
| Mizoram | 1 |
| Puducherry | 1 |





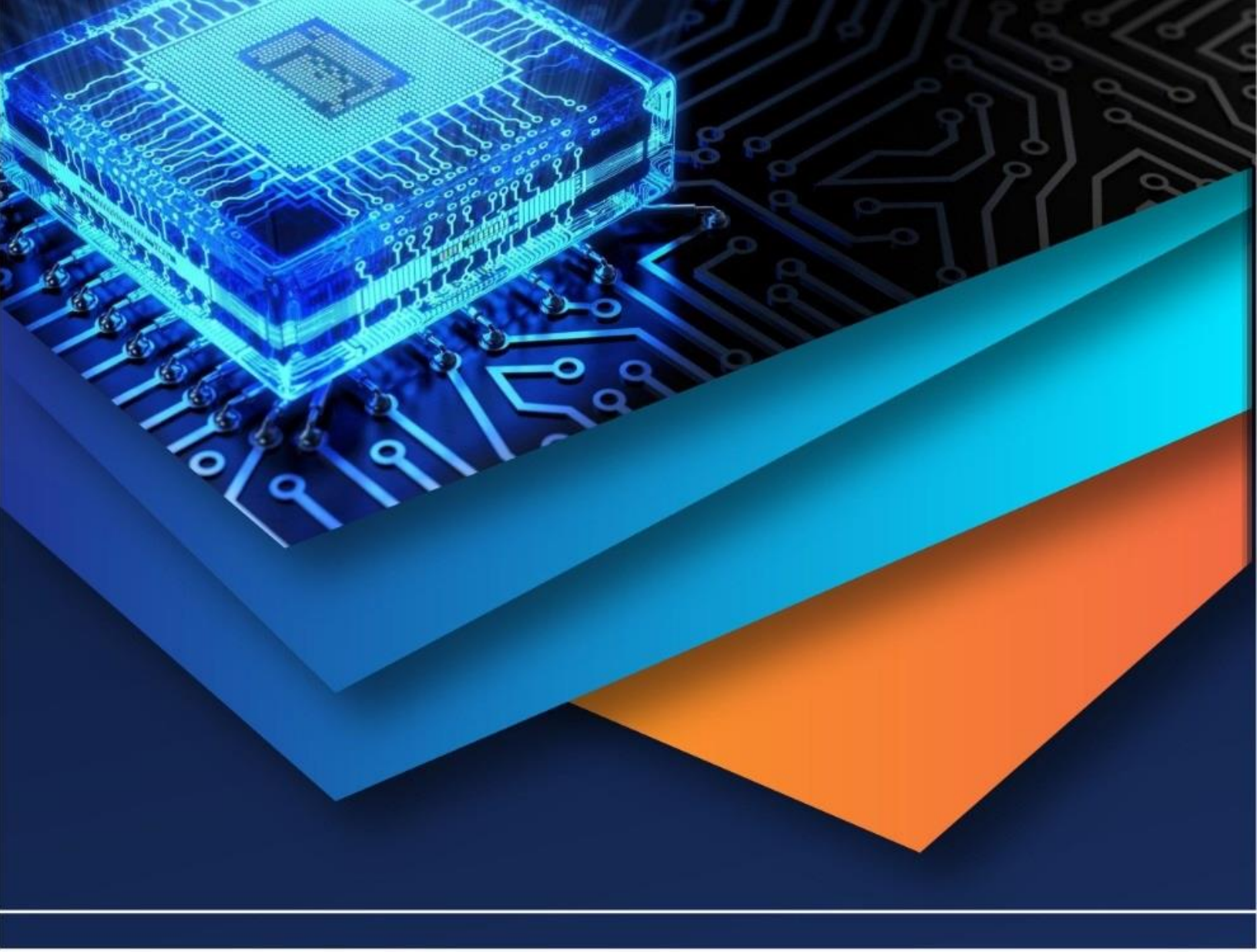

## IV. CONCLUSIONS

The main aim of the paper is to review and analyze the COVID-19 spread in India since the day of outbreak and pattern of spreading of virus in India and to know why National and native authorities are having a difficult time in handling the COVID19. Moreover to review about the common symptoms of COVID-19 that are observed till now, the spread of disease in India compared to other countries, the state wise trend of the epidemic to urge detail understanding of how this is often spreading and lastly to predict the longer term of epidemic in India.

This paper work are often extended to higher level in future, Predictive model for lasting of COVID-19 that uses Machine Learning algorithms, where the results from each graph of the paper are often taken as independent criteria for the machine learning algorithm. Moreover the longer term Prediction analysis are often extended ad resulted in additional accurate prediction on predict more accurate number of total cases in India.

## REFERENCES

[1] Hoseinpour Dehkordi, M. Alizadeh, P. Derakhshan, P. Babazadeh and A. Jahandideh,"Understanding epidemic data and statistics:A Case Study of COVID-19," in press.
[2] A. R. Fehr,S.Pearlman, " Coronavirues: An Overview of Their Replication and Pathogenesis," Springer, Methods Mol Biol., 2015;1282:1-23,doi: 10.1007/978-1-4939-2438-7_1
[3] D. Gondauri, E.Mikautadze and M. Batiashvili, "Research on covid-19 virus spreading statistics based on the examples of the cases from different countries," Electron J Gen Med.2020;17(4),em(209)
[4] Y. Deng, F. He, W. Li, "Coronavirus disease 2019: What we know?," J Med Virol,March 2020; doi:10.1002/jmv.25766
[5] W. Jingyuan, T. Wang, F.Kai and L. Weifeng, "High temperature and high humidity reduce the transmission of covid19," in press
[6] https://www.mohfw.gov.in/
[7] https://www.covid19india.org/
[8] https://github.com/CSSEGISandData/COVID-19
[9] https://www.worldometers.info/coronavirus/country/india/
[10] https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_population



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY