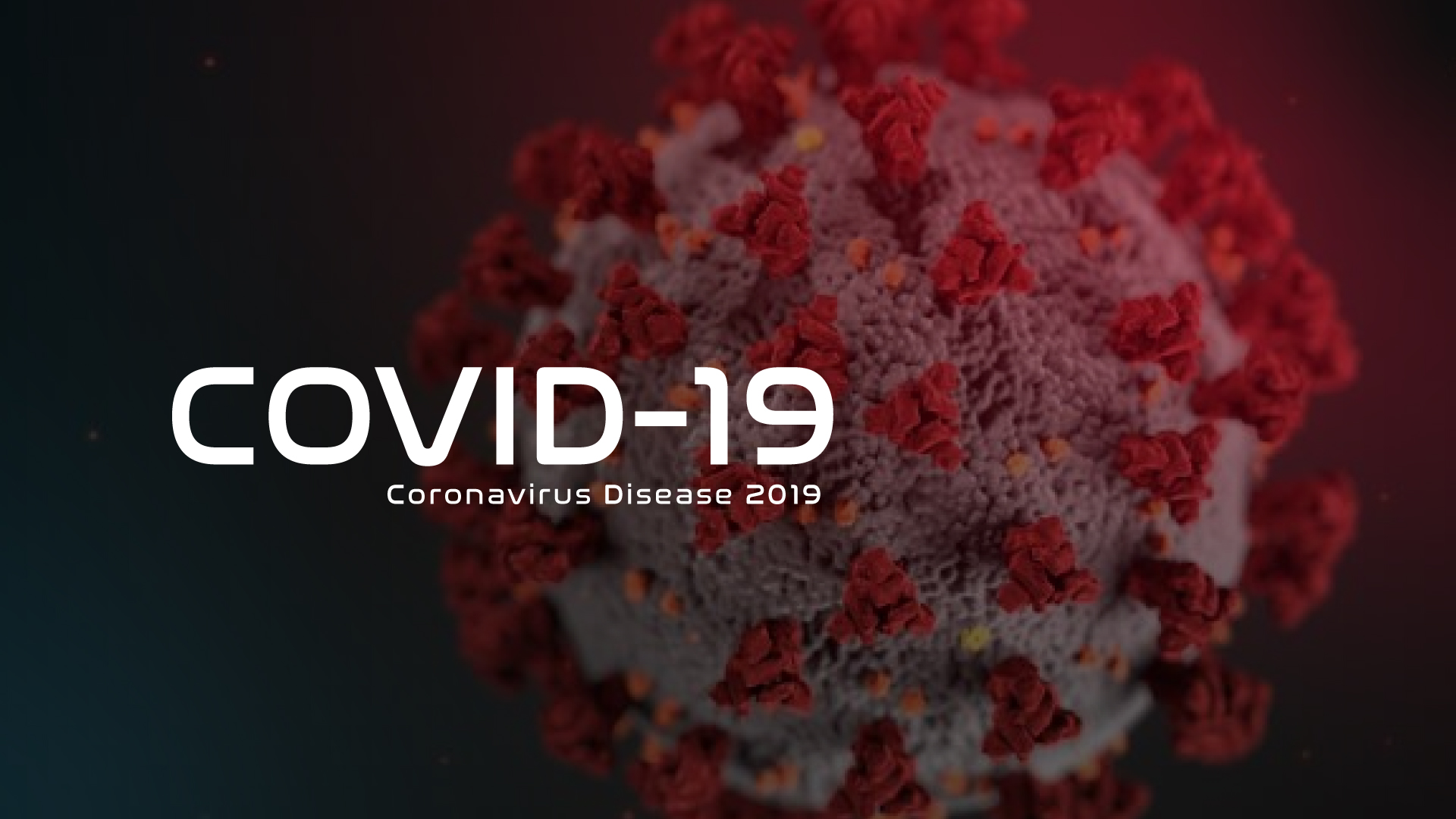
**Exploratory Data Analysis of COVID-19 pandemic (State level)**



**Jyotsna Pyarasani**

# 

# 

# **ABSTRACT**

**The current Covid-19 pandemic has forced a sudden and unimaginable turnaround on everyone, at all levels. Primarily our behavioral aspects and different aspects of our professional and personal lives have been strongly influenced and changed. The imperative and urgent COVID-19 research, which involves numerous areas, has changed the way of working, by requiring the investigation and response to be available as soon as possible. This study presents the current situation of COVID-19 spread in India along with the major factors that lead to the rapid spread with the inclusion of major impacts of the Outbreak on various sectors and Domains. The current Data is used as a source for Data-centric comparison of COVID-19 with the previous pandemics. An Exploratory Data analysis((EDA) technique is implemented to study and analyze the reports. The result of the analysis helps in predicting the future economic and social losses caused by the second/third wave.**

# **INTRODUCTION**

**The world is now facing an unprecedented crisis due to the novel coronavirus, first detected in Wuhan, China in December 2019. These viruses can cause respiratory symptoms in humans, along with other symptoms of the common cold and fever. The WHO declared the coronavirus disease (COVID-19) as a global pandemic on March 11, 2020. The disease has spread across 212 countries and territories around the world, with a total of more than 3 million confirmed cases.**

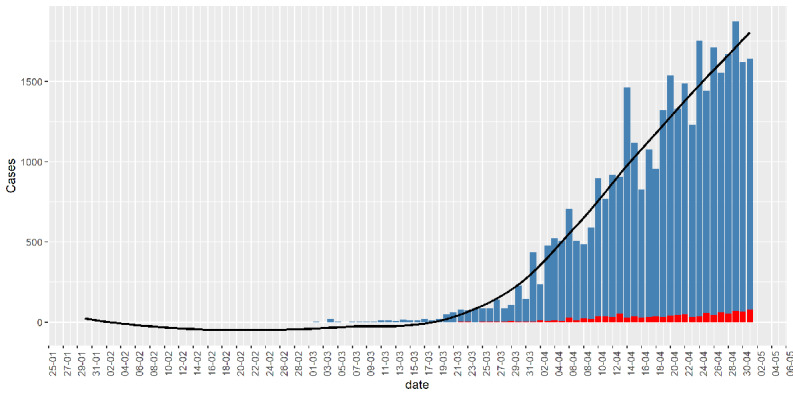


Figure 1:Bar chart of daily infected cases (blue) in India. The red bar denotes death. The black curve is a fitted smooth curve on the daily cases.

# **LIBRARIES USED**

* **Pandas is a useful library in data analysis. It can be used to perform data manipulation and analysis. Pandas provide powerful and easy-to-use data structures, as well as the means to quickly perform operations on these structures.**
* **NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, Fourier transform, and matrices.**
* **Seaborn is a library in Python predominantly used for making statistical graphics. Seaborn is a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python. Visualization is the central part of Seaborn which helps in the exploration and understanding of data.**
* **Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy.Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack.**
* **Folium is a Python library used for visualizing geospatial data. It is easy to use and yet a powerful library. Folium is a Python wrapper for Leaflet. js which is a leading open-source JavaScript library for plotting interactive maps.**

# **PROCEDURE**

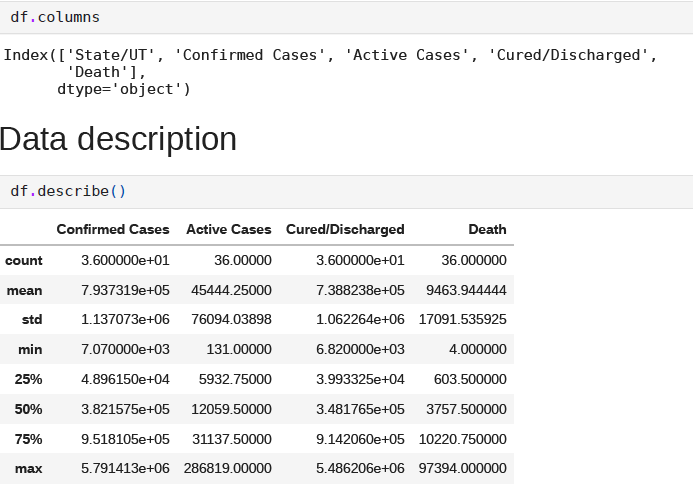
* **We acquired the required dataset from various websites.**
* **We have gone through the dataset and made a few modifications accordingly which include processing, removing NULL valued data.**
* **In the Jupyter notebook, we have interpreted the datasets using the above-mentioned library.**

# **DATA SET**

[Covid-19 India data set](https://github.com/Krishna1098/HCU-mini-project/blob/main/Datasets/india.csv)

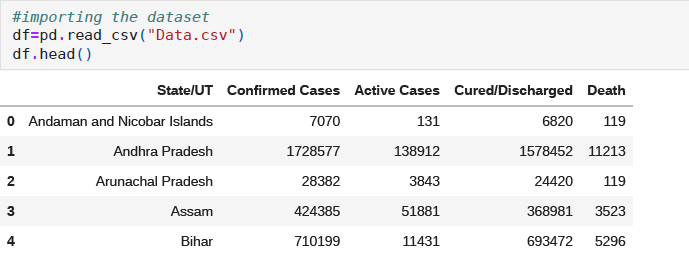
# **OBSERVATION**

* **Data Description**

****

**Fig 2: Description of dataset**

* **Pandas is a useful library in data analysis. It can be used to perform data manipulation and analysis.**

****

**Fig 2: Covid-19 state-wise Data**

* **Matplotlib is a multi-platform for data visualization**

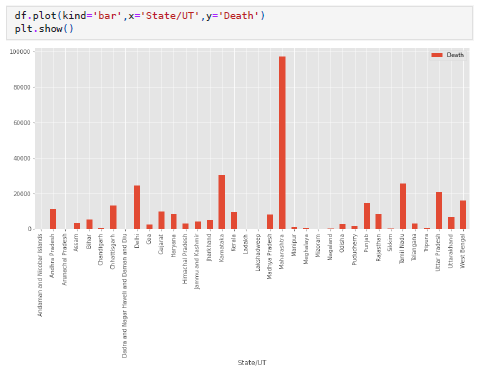


Fig 3: Representation of the total deaths recorded in each state

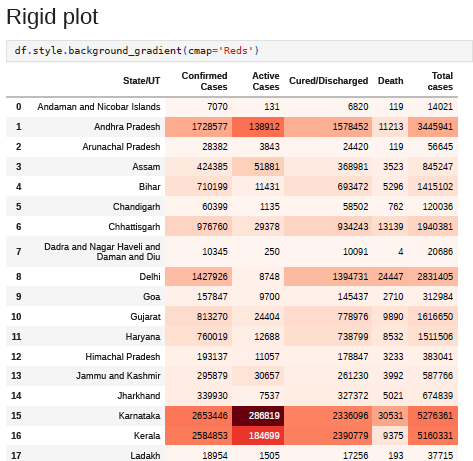


Fig 4: Gradient depth according to cases

* **Seaborn is a library in Python predominantly used for making statistical graphic representations.**

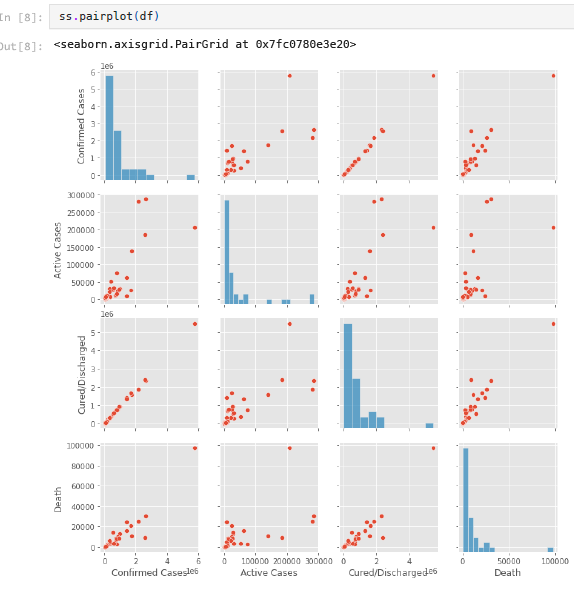


Fig 5: Scatter plots and bar graph

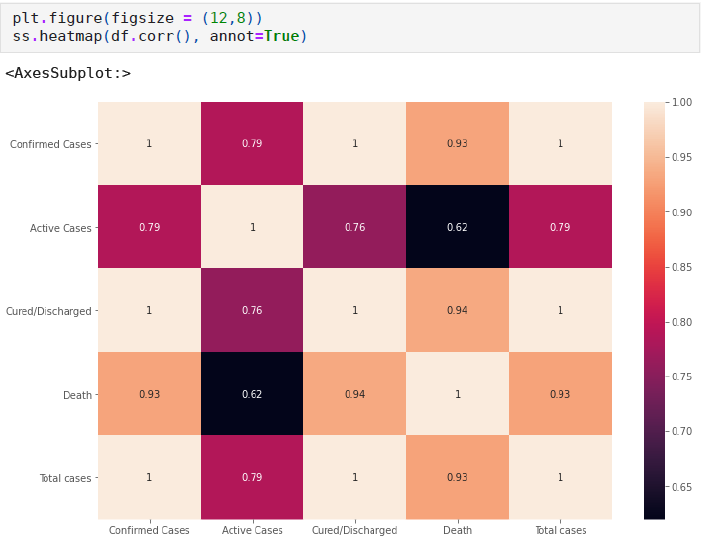


Fig 6: Heatmap

* **Folium is a Python library used for visualizing geospatial data.**



Fig 7: Indian map with Covid 19 situations using Folium

# **CONCLUSION**

**Using the above methods from libraries and plots, we obtained clear analysis and visualizations of the data. Wherein everyone can understand the range, Spread and Mortality rate of the cases by observing the bar, scatter plots, pie chart, Matrix table and the Map representation.**

# **REFERENCES**

* <https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_19_india.csv>
* <https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India>