International Journal of Innovations in Engineering and Technology (IJIET)
http://dx.doi.org/10.21172/ijiet.162.05

COVID-19 impact on the World using Machine Learning Libraries and Data Analysis

Saloni Shedge, Manish Shinde, Simran Sahetia
Department of Computer Engineering
VESIT, Mumbai University, Mumbai, Maharashtra, India

Abstract - At present COVID-19 is a major global challenge to humans. In March 2020, the World Health Organization (WHO) declared COVID-19 a pandemic, which is caused by the novel SARS-CoV-2 virus. To fight with COVID-19 virus in India, proper analysis of available data related to this disease is important. In this paper we are studying available information to provide reliable statistics as well as analysis on COVID-19 in India. Various models of machine learning are referred to. COVID spread data on the world was taken and graphical state wise information was created using data science libraries. Spread and Growth prediction in India is also done using various data science tools such as NumPy, Pandas, matPlotlib etc. More research has to be conducted in the areas that are less explored. This pandemic is now an enormous challenge for researchers, health-care workers and many more. We would like to contribute to this global endeavour with the help of our research and analysis.

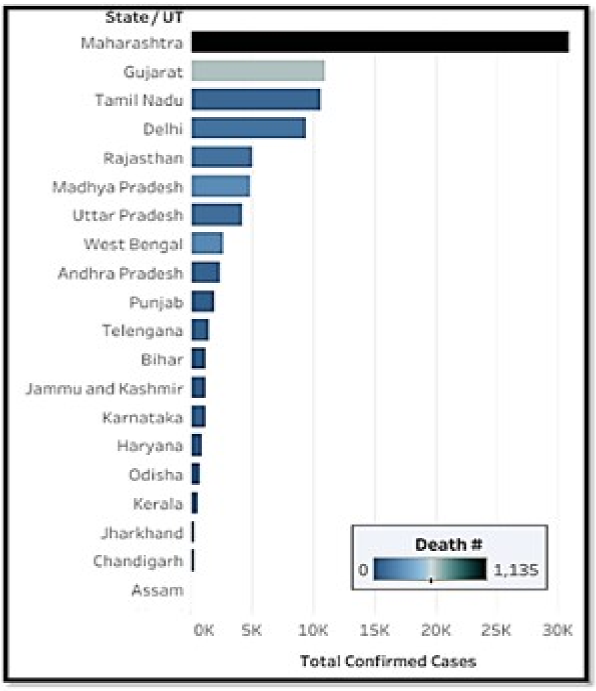
Keywords – COVID-19, Machine Learning, Geographic data analysis, Matpotlib, Pandas, Data Frames, Pandemic, Plotly.

I. INTRODUCTION

We are trying to analyze data on COVID hence, we are studying the past work done on the research of COVID. There are many papers to analyze and forecast covid spread on global level as well as its spread in India. We have referred few of these papers

II.LITERATURE SURVEY

Research paper [4] introduces Data Statistics and Forecasting of COVID, they use tables to analyze data and give information of each state in India as shown in figures below which is obtained using a dataset from kaggle.



Source:Monika Mishra at BigDAI center,2020

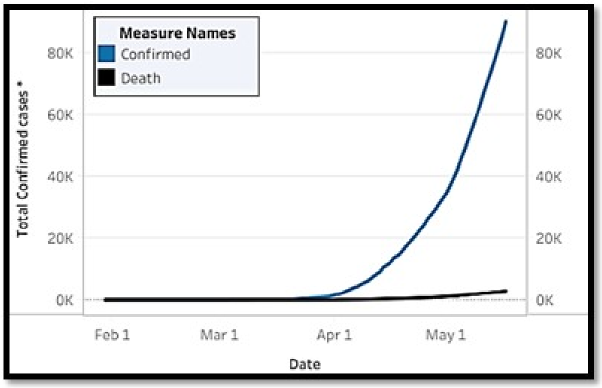


Fig. 2. . Confirmed and Death measure

Source:Monika Mishra at BigDAI center,2020

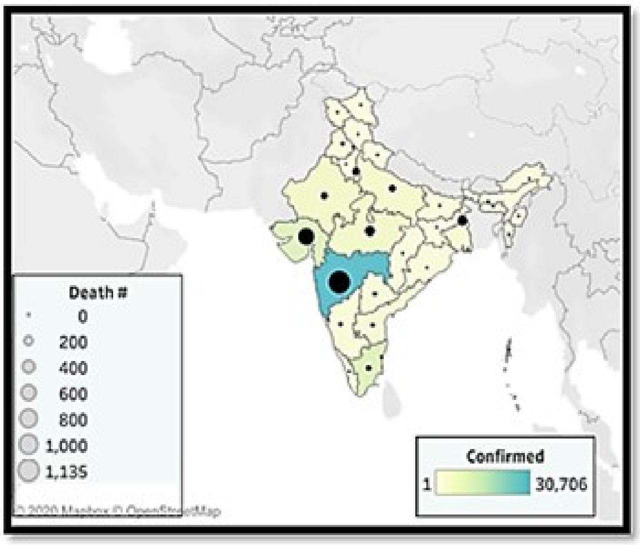


Fig. 3. Analysis of cases all over India
Source:Monika Mishra at BigDAI center,2020

Paper[5] introduces an objective approach to predict COVID. This paper has analysed different models from exponential smoothing families and then produced forecasting models..

Study introduced in [6] is based on a differential equation which is formed to find how the number of patients increases with time, they also provide a statistical model for the spread of COVID in clusters..This statistical model is specifically created for data on COVID in India.

Application of python to data on COVID is described in [7]. They create the predictor for the possible end number of COVID patients in Italy. Pandas and scikit libraries have been used by them. Model building for COVID with data science and challenges related to it are explained in [8].

Paper[9] introduces a mathematical model to analyze COVID.They have investigated practical data on COVID in India.Established model by them follows actual data of COVID spread. As shown in the figure below:

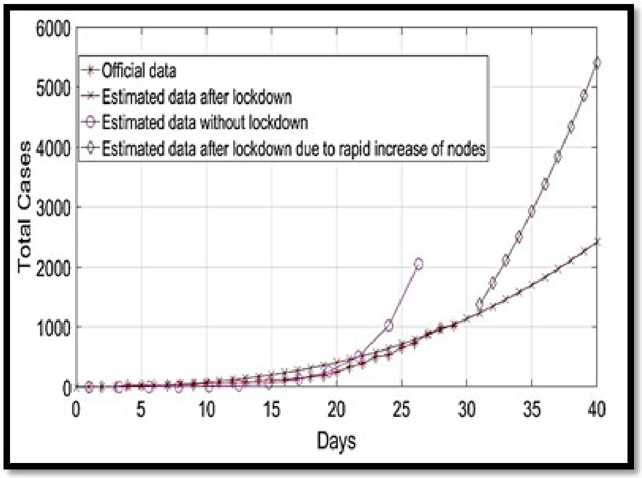


Fig. 4. Post Lockdown Estimation.

Source:M.K., Arti. (2020).

[10] also gives official information on COVID spread in India.

III.TOOLS USED DURING ANALYSIS

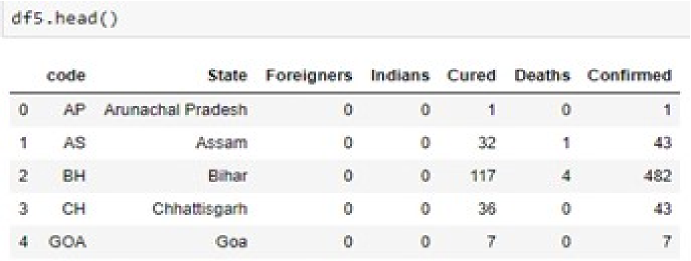
For Data Visualization:

- Numpy: Numpy is used to create matrices that help in plotting the data. We can create one dimensional as well as two dimensional matrices using this library. Usually it is abbreviated as np
- Pandas: This python library is used for cleaning and analysing the data. Pandas make the data look like a table. This has two forms: Series and Dataframes.
- Seaborn: The library helps to plot advanced graphs
- Jupyter Notebooks: It is an IDE in which the coding for data analysis is done
- Matplotlib: Matplotlib is used to plot graphs in the notebook. It analyzes the dataset and plots the appropriate graphs.
- Plotly: Plotly is a tool that develops online data analysis and it converts the data into visual form.
- DataFrames: They are basically used to transform the data into tabular form which in turn is passed to matplotlib library to plot graphs.
- Machine Learning: It is the ability of the machine to learn and find patterns in the dataset that is provided.
- Python:It is a high level programming language that is used in machine learning to code and train the models. It has many features.

IV..DATASET USED

Two different datasets were used for data analysis.

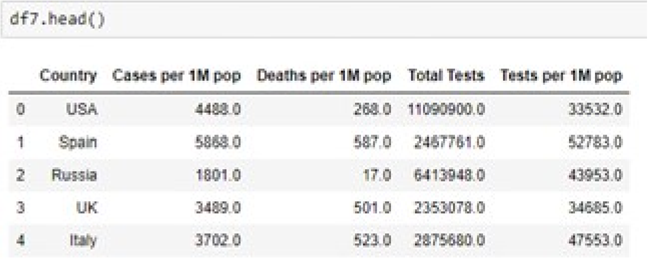
1)The first dataset was the impact of COVID-19 on India. It was the statewise impact and the number of people that were cured,death or confirmed as per the states. The dataset was found on Kaggle and data was cleaned.The dataset consists of data till 3rd May 2020.



```
df5.head()
```

	code	State	Foreigners	Indians	Cured	Deaths	Confirmed
0	AP	Arunachal Pradesh	0	0	1	0	1
1	AS	Assam	0	0	32	1	43
2	BH	Bihar	0	0	117	4	482
3	CH	Chhattisgarh	0	0	36	0	43
4	GOA	Goa	0	0	7	0	7

2)The second dataset is COVID's impact on the entire globe. The data was segregated in the form of different countries and consisted of cases per 1 million population,Deaths per 1 million population and total test. The dataset was found on Kaggle



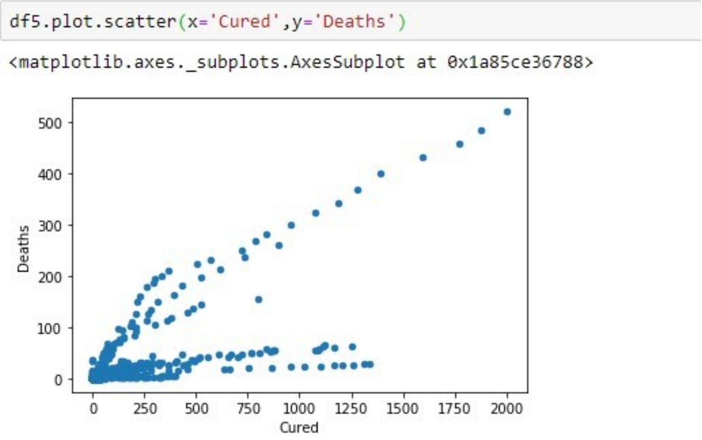
```
df7.head()
```

	Country	Cases per 1M pop	Deaths per 1M pop	Total Tests	Tests per 1M pop
0	USA	4488.0	268.0	11090900.0	33532.0
1	Spain	5868.0	587.0	2467761.0	52783.0
2	Russia	1801.0	17.0	6413948.0	43953.0
3	UK	3489.0	501.0	2353078.0	34685.0
4	Italy	3702.0	523.0	2875680.0	47553.0

V.IMPLEMENTATION

First dataset for global COVID spread was downloaded.From the given dataset information was filtered out and growth trajectories on cases across globe was created with respect to country by using python libraries pandas,NumPy and Seaborn. From the second dataset COVID spread data on India was taken and graphical state wise information was created using data science libraries.From those visualizations we were able to gain insights on the spread of COVID.We were able to find out which states have impacted largely and which states are not impacted till date. Hammer graphs were also created.Map View ,Heat Map were used to visualize information .Comparative visualization of state wise availability of Data on Laboratories and Testing kits availability was also analyzed .Information on that data was displayed graphically. Containment zones and their distribution was analyzed.

Scatter plot was plotted using plotly for the dataset of India as per the states. On the Y-axis we have the Deaths and on the X-axis we have the people cured. After analysis we can see that the highest amount of deaths India saw was approximately 500 people. Death data is clustered between 0-100 range. Also the total number of cured people maximizes at 1400 people. Cured people aggregate to the range of 0-300.

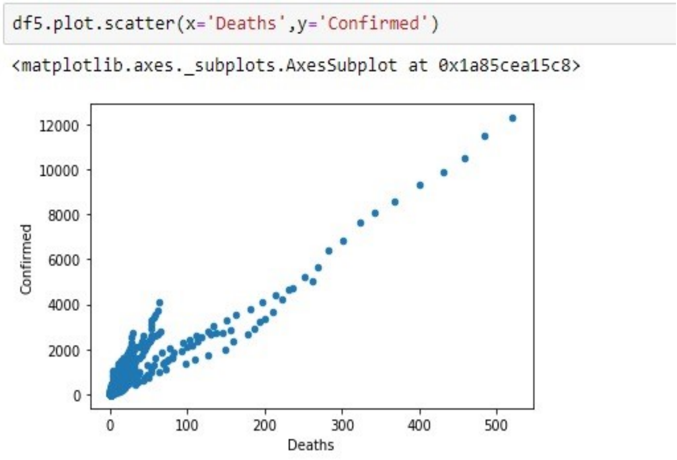


```
df5.plot.scatter(x='Cured',y='Deaths')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a85ce36788>
```

Fig. 5. Deaths v/s Cured cases in INDIA

Data analysis between the number of Death and Confirmed cases in India was also done. According to the analysis the maximum confirmed cases are around 5000 and the maximum deaths in India is around 500 people. We can say that the death rate is 1/10th of the confirmed cases.



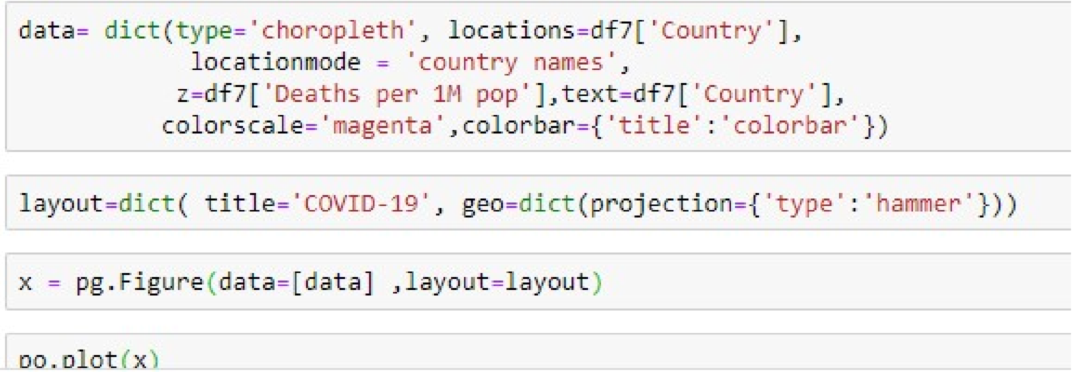
```
df5.plot.scatter(x='Deaths',y='Confirmed')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a85cea15c8>
```

Fig. 6.Confirmed v/s Death cases in INDIA

GEOGRAPHIC PLOTTING USING CHOROPLETH:

A data dictionary was created in which attributes such as type, location, locationmode, z, text, colorscale and colorbar were defined. The Map type used was Choropleth for its easy to understand nature.The locationmode was of country names as in the dataset data was divided as per countries After creating a dictionary for data, another dictionary for layout was created wherein attributes like title and geo were defined. We just hammer as a projection for plotting data in our globe map. Later x was defined and the plotly function was called. An HTML file was created as output which consisted of the World Map. The location in the output was mapped according to the countries and z defines the data that the map represented. Other than that the colorscale used was magenta and the title that was given to the colorbar which defines the range of the map was also named as colorbar.



```
data= dict(type='choropleth', locations=df7['Country'],
            locationmode = 'country names',
           z=df7['Deaths per 1M pop'],text=df7['Country'],
          colorscale='magenta',colorbar={'title':'colorbar'})
```

```
layout=dict( title='COVID-19', geo=dict(projection={'type':'hammer'}))
```

```
x = pg.Figure(data=[data] ,layout=layout)
```

```
po.plot(x)
```

Fig. 7.Snapshot of the code

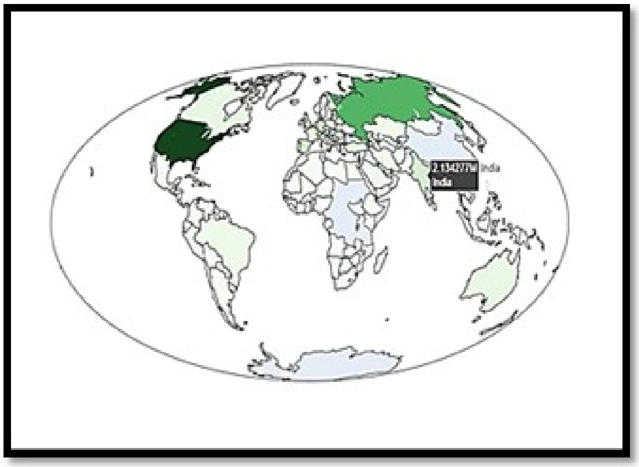


Fig. 8.Test Cases across the WORLD done for COVID-19

Data was plotted according to the total test cases performed in the entire World. The color scale is green and it is a hammer projection. The darker shade represents a higher number of tests performed while the light green color indicates that the number of tests were less. As we can see that the USA conducted the highest number of total tests, Russia takes the second lead in performing the total tests for Corona.

COVID-19 Analysis acc to Cases per 1M population:

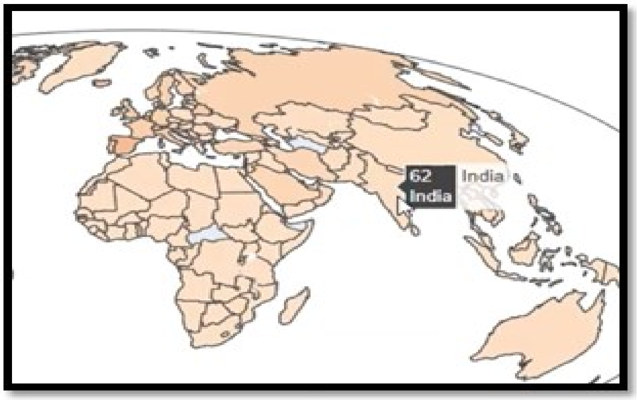


Fig. 9 . Cases per 1M population INDIA

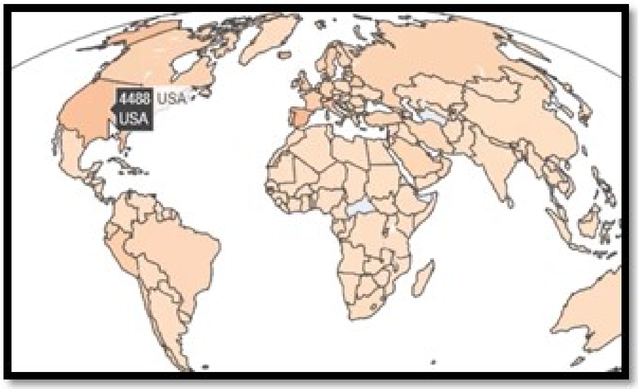


Fig. 10 . Cases per 1M population USA

The above graph shows the cases as per 1 million of population. We can see that Indian has 62 cases per 1M and USA has 4488 cases per 1M. From data analysis we can say that USA has been subjected to a greater impact of COVID-19.

COVID-19 Analysis acc to Deaths per 1M population:

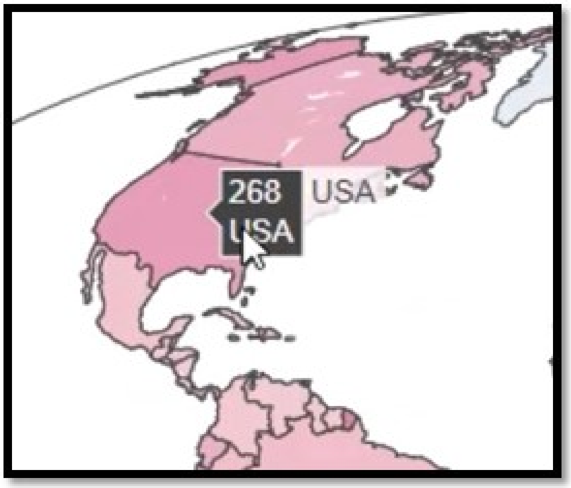


Fig. 11 .Death Cases per 1M population USA

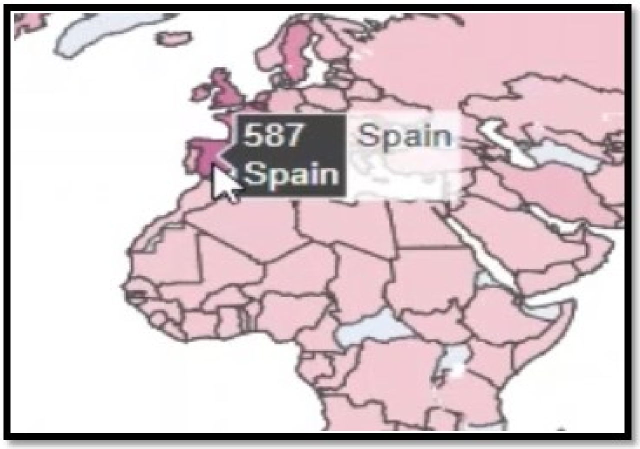


Fig. 12 .Death Cases per 1M population SPAIN

Based on the total death cases per 1M population, graphs have been plotted and Spain shows the highest cases of deaths with a range of 587 per 1M while USA is the second highest with the value of 268 per 1M of population.

VI.CONCLUSION

We have studied and investigated the problem of Covid19 spread in India as well as globally in practical scenarios. After data analysis we have shown the current status of India by representation of the number of deaths, cured and confirmed cases through scatter plot . We have also calculated the impact on various states in India.We have also analysed and plotted the situation of various countries affected by COVID19 throughout the world.

Hence,various visualizations made by using data science tools were beneficial to understand COVID spread.These visualizations enable us to find out which states are lagging to fight COVID. Government can take necessary actions to support weak areas found out by this visualization. Thus,comparative study helps us to analyse our current situation and enables us to fight with COVID by taking all the necessary preventive measures needed.

REFERENCES

[1] Mishra, Monika & Dauletbak, Dalyapraz & Woo, Jongwook. (2020). COVID-19 Data Statistics and Forecasting.

[2] "One COVID-19 positive infects 1.7 in India, lower than in hot zones". The Indian Express. 19 March 2020.

[3] “India becomes fifth country to isolate Covid-19 virus strain”,The .livemint.com,13 Mar 2020.

[4] COVID-19 Statistics and Data Analysis using Tableau 3/15/2020 By Monika Mishra at BigDAI center.

[5] Forecasting the novel coronavirus COVID-19 Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. PLOS ONE 15(3): e0231236. https://doi.org/10.1371/journal.pone.0231236

[6] Bhatnagar, Manav. (2020). A Statistical Model for the Spread of COVID19 in Clusters. 10.13140/RG.2.2.18583.52644.

[7] Italian COVID-19 Analysis with Python https://towardsdatascience.com/italian-covid-19-analysis-with-python-1bdb0e64d5ac

[8] Four Basic Data Science Lessons Illustrated by COVID-19 Data
https://medium.com/@ageitgey/four-basic-data-science-lessons-illustrated-by-covid-19-data-7d94134a5b0e

[9] M.K., Arti. (2020). Modeling and Predictions for COVID 19 Spread in India. 10.13140/RG.2.2.11427.81444.

[10] "Home | Ministry of Health and Family Welfare | GOI". www.mohfw.gov.in. Retrieved 5 May 2020.

.