Analysis of COVID-19 in India

1st Rohit Vartak
Department of Electrical Engineering
190010058
190010058@iitb.ac.in

2nd Raghav Singhal
Department of Electrical Engineering
19D070049
19D070049@iitb.ac.in

3rd Kushal Kejriwal
Department of Electrical Engineering
190110040
190110040@iitb.ac.in

***Abstract*—Year 2020 has been all about COVID-19. A pandemic of the modern age, has, with all of its fury, left the world ravaged. Life as we knew it, was brought to a complete stand still.Thus, we were motivated to understand how India has fared in the battle against COVID-19, how different states in India have coped with testing and analyse the trends related to the COVID-19 outbreak.Thus, our project demonstrates the distribution of ICMR testing facilities in India, trends in the daily confirmed cases , daily cured cases, daily deaths and the doubling time for infections. Our project highlights the Rural-Urban divide, the testing capabilities of each state in India, the integral role that GDP plays in healthcare and the distribution of healthcare centres and hospital beds in India.We also analyse the age which is most at risk from COVID-19 and compare India with the rest of the world in her battle with COVID-19.**

***Index Terms*—COVID-19, formatting, style, styling, insert**

I. Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus(2019-nCoV). The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. Most people infected with the COVID-19 virus experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness [1]. The coronavirus outbreak came to light on December 31, 2019 when China informed the World Health Organisation of a cluster of cases of pneumonia of an unknown cause in Wuhan City in Hubei Province. Subsequently the disease spread to more Provinces in China, and to the rest of the world. The World Health Organization(WHO) soon declared it a pandemic. The virus has been named SARS-CoV-2 and the disease is now called COVID-19 [2]. COVID-19 has been researched intensively since its birth in Wuhan, China, about its symptoms, its after-effects, and how it spreads.There has been intensive research done on how India has fared too, however we wanted to understand how each state in India has fared, how GDP plays an integral role in the testing and general infrastructure related to healthcare a state has and the rural-urban divide. We also wanted to understand the trends associated with the COVID-19 outbreak namely- confirmed cases, daily cured cases, daily deaths and the doubling time for infections. We further wanted to analyse which age group is most at risk from COVID-19.

II. DATASETS

We used, in total, 11 datasets obtained from Kaggle(9) [3] and Wikipedia(2) [4] [5]. Some datasets(2 : - *popindia and GDP*) were cleaned using Microsoft Excel and the rest using data cleaning techniques which were implemented in Python. We had to take special care of the names of the States and the use of the ampersand (&) in their names while merging different datasets. Further, few datasets had 'Date' as a column and so this column had to be converted to the Python's *'time'* datatype for ease of plotting.

- *HospitalBedsIndia* : - Contains the number of Health Care Centres(Primary, Community, District, Sub-District and total public care centres), number of hospital beds and number of Hospitals in Rural and Urban areas in each State/UT(Union Territory) in India.
- *popindia* : - Contains Population(number and percentage of total population of India (%)), decadal growth(%), Rural and Urban population(number and percentage of the population of that state), Area(km^2), Density($/km^2$) and Sex ratio for each State/UT in India.
- *GDP* : - Nominal Gross Domestic Poduct (GDP in Billion Dollars)
- *TotalSamples* : - Contains Total Samples taken by the concerned authorities, the number of postives and the negatives for the tests taken for the mentioned dates for each State/UT in India.
- *ICMR Testing Facilities* :- Contains the City and State of each Testing facility throughout India along with the latitude and longitude
- *Time Series Confirmed Cases, Deaths, and Recoveries* :- These 3 datasets give the time series of the confirmed cases, deaths and recoveries of all the countries in the World from the period of January to November
- *Covid India Cleaned* :- Gives the time series of confirmed cases of all the states of India from the period of March to November.
- *Confirmed Cases India* :- Contains the cumulative nummber of cases in each state and Union Territory through each date.
- *AgeGroup* :- Contains the Total Number of Cases and Percentage of Cases per each Age Group.

III. Analysis and Results

A. Abbreviations and Acronyms

We have used the following abbreviations/acronyms -

- ICMR : - Indian Council of Medical Research
- GDP : - Gross Domestic Product
- UT : - Union Territory

B. Units

- latitude, longitude : - °(Degree)
- area : - km^2 (square kilometer)
- Density : - /km^2 (per square kilometer)
- Nominal GDP($ B) : $ (Dollar)

C. Equations

Doubling Time for Infection

Number of cases of a^{th} day $= q1$

Number of cases of b^{th} day $= q2$

$$\text{Doubling time} = (b-a) \cdot \frac{\ln(2)}{\ln(\frac{q2}{q1})} \quad (1)$$

D. Analysing ICMR testing facilities and statewise distribution throughout the country

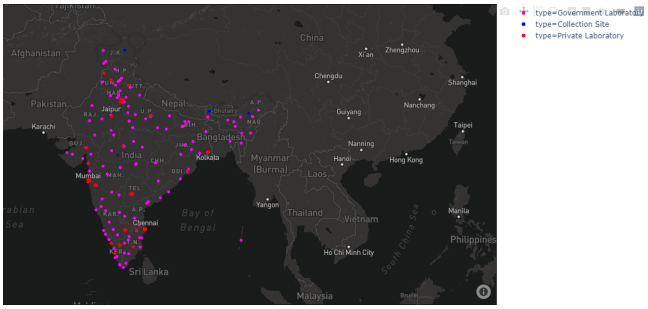

Fig. 1: Mapbox plot of ICMR testing facilities throughout India

1) ICMR facilities throughout India: In Figure 1 We see that majority of the testing facilities are government laboratories, while some are privately owned. Carefully looking along the Indo-China border, we observe some collection sites, which is owing to the high altitute and rough terrain.

2) Statewise Distribution of Testing Facilities: In Figure 2 We observe that states like Maharashtra, Tamil Nadu and Uttar Prades have higher number of testing facilities. This can be attributed to large land area and higher COVID cases. Many North-Eastern states have less labs and some only have collection sites.

E. Comparing India to the World in COVID 19 (Confirmed Cases, Recoveries and Deaths)

1) Comparing India to the World in COVID 19 confirmed cases: In Figure 3 We see that initially China led the total number of cases (in the months of Jan and Feb). Then in April, the virus exploded in the United States and it reported a surge in cases. In mid May, it exploded in Brazil and in June

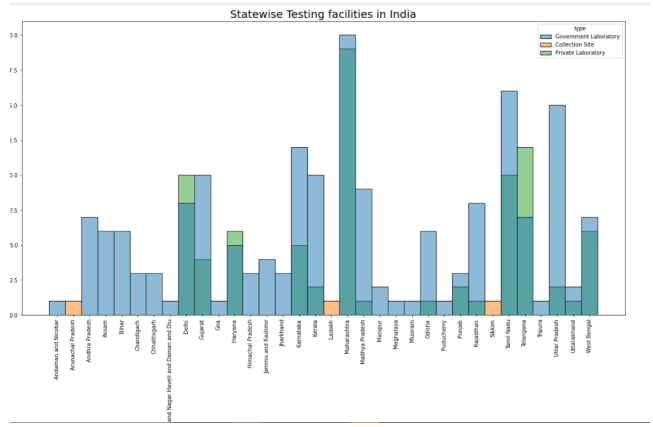

Fig. 2: Histogram of the number of statewise facilities filtered by their type

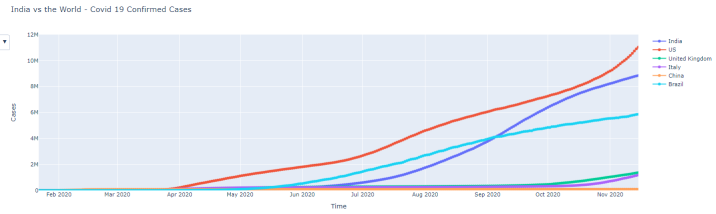

Fig. 3: India vs World: COVID 19 Confirmed Cases

it started in India before attaining its peak in September (we can clearly see the point of inflection, we also overtook Brazil around that point). India was quite close to overtaking US in October but US again reported a surge in cases and the cases dropped in India leading to an increasing gap.

We can also attribute the surge in cases in the United States to the US Presidential Elections 2020.

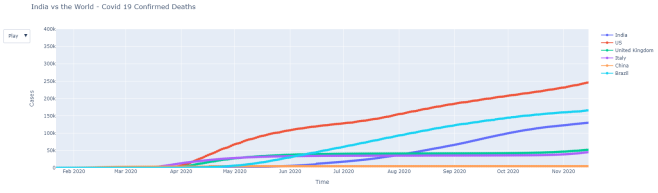

Fig. 4: India vs World: COVID 19 Confirmed Deaths

2) Comparing India to the World in COVID 19 confirmed deaths: In Figure, 4The US reported more deaths and had a poor recovery rate. Even though India has significantly higher cases then Brazil, the number of deaths are lower.

3) Comparing India to the World in COVID 19 confirmed recoveries: Figure 5 again displays the excellent recovery rate of India. India overtook US in total recoveries as early as mid August (US had 110000 more cases than India), and is head and shoulders above any country. We have a remarkable recovery rate of 95%.

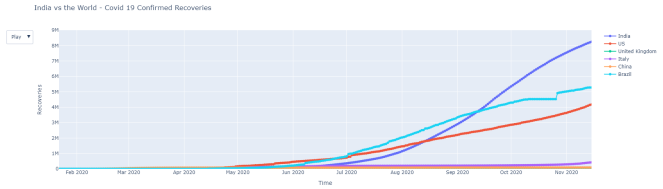

Fig. 5: India vs World: COVID 19 Confirmed Recoveries

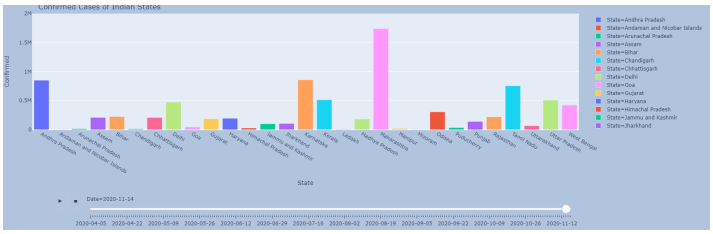

Fig. 6: Histogram of Statewise Confirmed Cases of India

F. Analysing Cumulative Cases of Various States of India

In Figure 6Maharashtra has the highest number of cases (almost double than the next highest state). Mumbai and Pune were covid hotspots. The south Indian states were also badly affected with AP, TN, Karnataka recording around 800k cases

G. Studying the Doubling Time for Infections in the Worst 5 Affected States in India

The worst 5 affected states in India, as per number of cases; are Maharashtra, Andhra Pradesh, Tamil Nadu, Karnataka and Uttar Pradesh. A higher doubling time implies a better

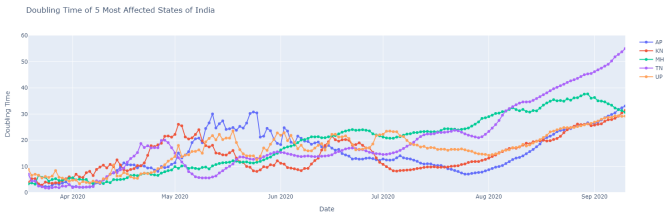

Fig. 7: Plot of Doubling Time of 5 Most Affected States of India.

condition and more control over the situation. With reference to Figure 7 we initially see that there is a lot of variation in the doubling time. This is followed by a number of troughs and peaks, but with gradual increase throughout. Towards the end of the graph, all the states showcase an overall increase in the doubling time, however only Tamil Nadu has a significant increase, displaying it has handled the situation in the best way among these 5 states.

H. Analysing the Trends in Daily Confirmed Cases, Daily Cured Cases and Daily Deaths

For each of the 3 graphs: Daily Confirmed Cases(Figure 8), Daily Cured Cases(Figure 9) and Daily Deaths(Figure 10); we

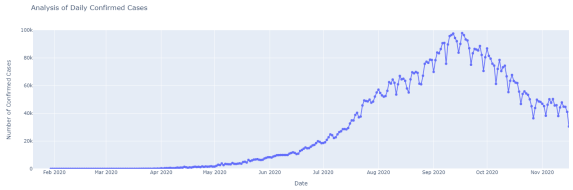

Fig. 8: Analysis of Daily Confirmed Cases.

Fig. 9: Analysis of Daily Cured Cases.

can see a similar pattern. All 3 graphs increased till a point (around Sep-Oct) and more or less decreased thereon. Thus, they hit a local maxima around that time. The sharp peak in deaths at Jun 17 can be attributed to the recataloging of previous deaths that were falsely identified as Non-Covid related, and were correctly catalogued on that day, leading to a sharp spike.

I. Analysis of Age Distribution of Covid-19 Cases in India

The pie graph Figure 11 clearly illustrates that the age group '20-29' has been most affected by Covid-19 in India (in terms of number of cases), followed by '30-39' and '40-49', which together form the bulk of the cases encountered in our country.

J. Analysing the Hospital and Population data

1) Analysing Total Public Health Care Centres per unit area: In Figure 12, we see that the distribution of health care centres across each State/UT is approximately the same. In other words, assuming uniform distribution of Health care centres in terms of area, we note that 'accesibility' of health care centres for the population is approximately the same across different states/UT in India.

2) Analysing Total Public Health Care Centres per unit population: In Figure 13, we see a great disparity in the 'availability' of Health Care centres for a person across different states/UT.Higher ratio implies better 'availability' of

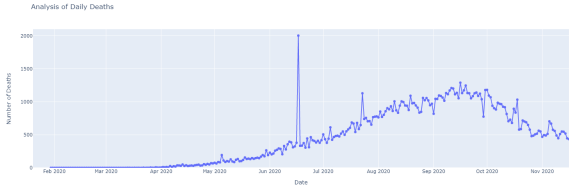

Fig. 10: Analysis of Daily Deaths.

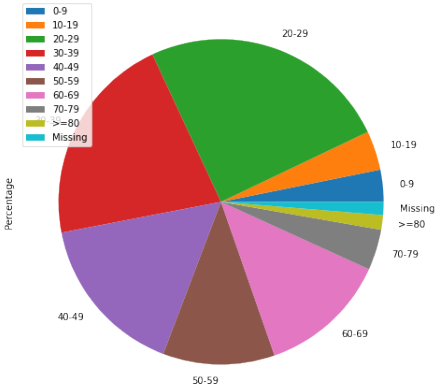

Fig. 11: Analysis of Age Distribution of Covid-19 Cases in India.

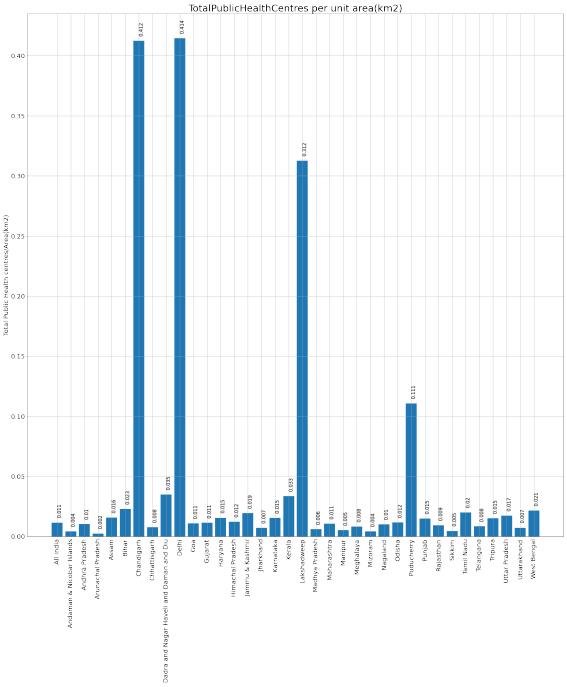

Fig. 12: Plot of Total Public Health Care Centres per unit area(km^2) for each state/UT in India.

Health Care centres for the person living in the state/UT. We disregard social, political, economic, geographical and other such reasons which may result in friction while admitting/ treating a patient.

3) Analysing Total number of Public beds per unit area: In Figure 14, we see approximately the same ratio for almost all the states/UT which in other words, means, approximately the same 'accesibility' to a hospital bed for a person across different states/UT across India.

4) Analysing Total number of Public beds per unit population: In Figure 15, we see a great divide here. Some states fare far better than the others, however, India still lacks far behind in infrastructure as compared to Europe which has on

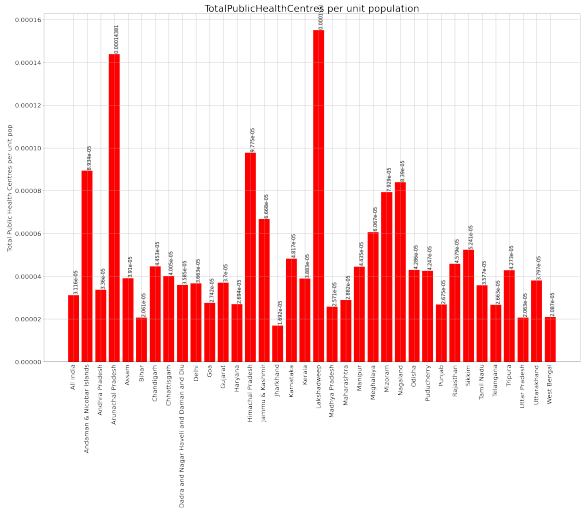

Fig. 13: Plot of Total Public Health Care Centres per unit population for each state/UT in India.

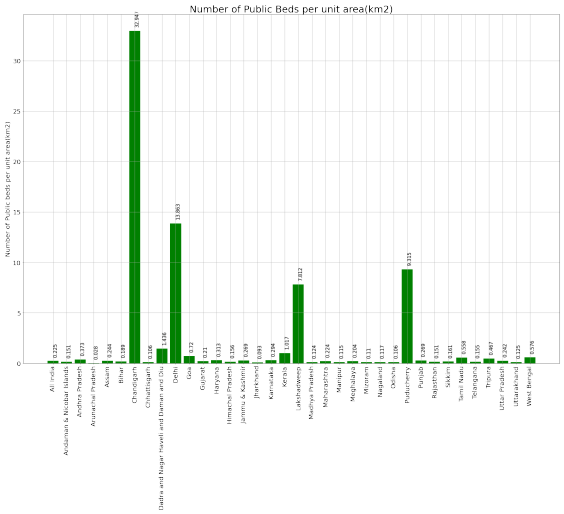

Fig. 14: Plot of Total number of Public beds per unit area for each state/UT in India.

average 538 hospital beds per 100,000 of their population.

5) Analysing the Rural-Urban divide - Comparing Total number of Public beds per unit population: In Figure 16, we see great disparity between the rural and urban health care infrastructure, where the ratio is usually higher for urban areas as compared to rural for most of the states/UT in India.(*Note*)The data does NOT take into account the social, political and economic inequalities which plague India and which are more prevalent in Urban areas where the economic gap is huge.

K. Analysing the Statewise Testing, hospital and Population data

1) Analysing Total number of Samples Taken per unit population: In Figure 17, we again see a huge disparity in the total testing carried by the states/UT across India.

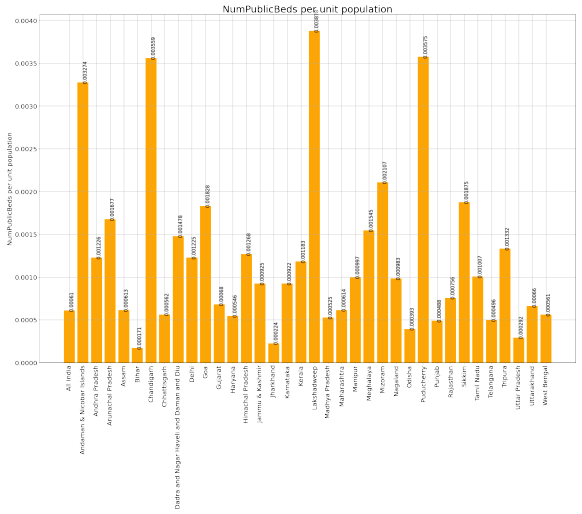

Fig. 15: Plot of Total number of Public beds per unit population for each state/UT in India.

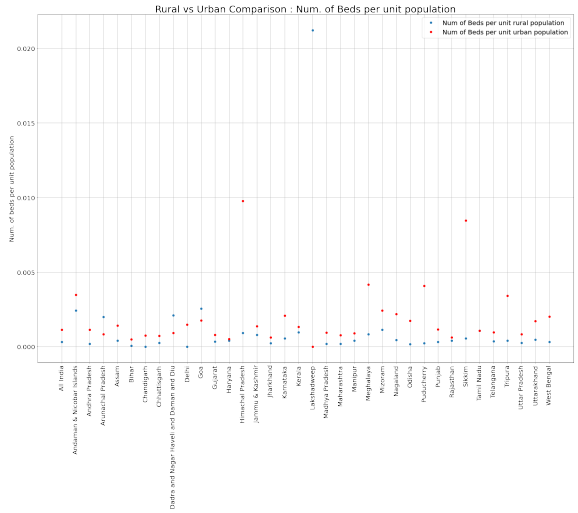

Fig. 16: Plot of Total number of Public beds per unit population for Rural and Urban areas of each state/UT in India.

2) Analysing the effect of higher GDP on testing and hospital beds: In Figure 18, 19, 20 and Figure 21, we observe positive correlation for all. Therefore, Figure 18 implies that urban areas are more prone to be COVID-19 hotspots while Figures 19 and 20 imply that higher GDP implies better testing abilities and higher supply of hospital beds.In Figure 21 we see that a higher per capital GDP implies better availability of a hospital bed for a person.

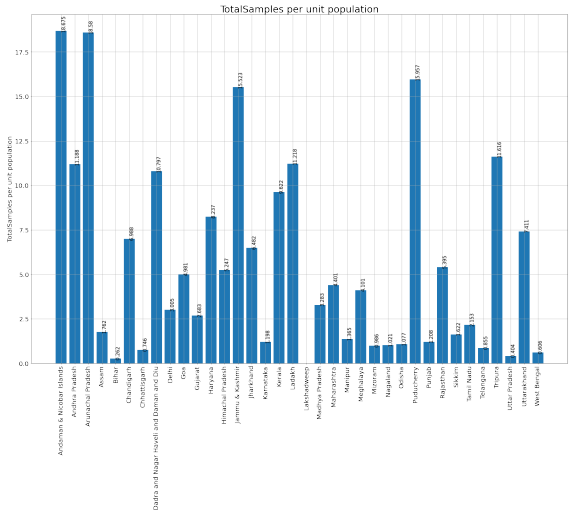

Fig. 17: Plot of Total number of Samples Taken per unit population for each state/UT in India.

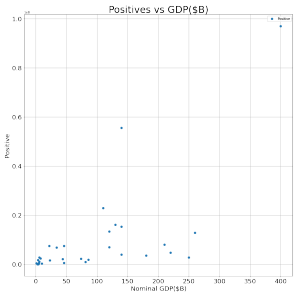

Fig. 18: Positives vs GDP.

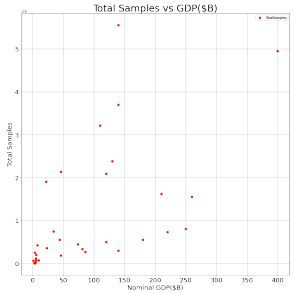

Fig. 19: Total samples taken vs GDP

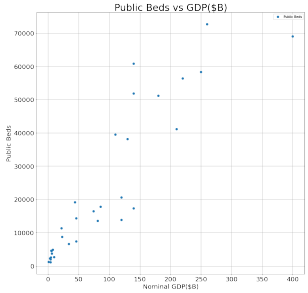

Fig. 20: Total Number of hospital beds vs GDP

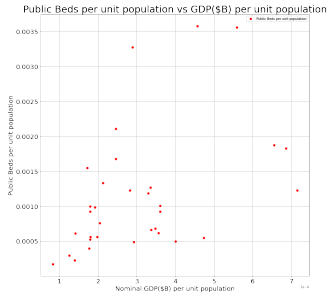

Fig. 21: Total hospital beds per unit population vs GDP per capita

IV. DISCUSSION

We believe we have performed a detailed and extensive analysis of Covid-19 in India. Along with the conventional and popular questions such as studying doubling rates, gender distribution of Covid-19 victims, daily analysis of cases, comparing India with other countries, comparing different states of India; we also studied slightly more unpopular questions such as ICMR distribution of testing facilities, comparing rural and urban healthcare facilities, the correlation of GDP with Covid-19 in different states and availability of hospital beds throughout India. This extensive study over different spheres is what we believe makes our analysis well-rounded and complete. We completely focussed on the data analysis and visualization part, not giving emphasis to the prediction/machine learning aspect of it. Our project could be insightful to anyone looking for a detailed analysis of Covid-19 in India through the majority of 2020, since we encompassed a wide range of ideas and topics, and have gained significant insights from the same. It could be a stepping stone for someone looking to study and analyse the unpopular questions regarding Covid-19 in India in closer detail.

ACKNOWLEDGMENT

We thank the Professors of DS203 course, who have given us this opportunity to learn something new. We would also like to thank the TAs for their invaluable feedback throughout this course and especially during the course of this Project.

REFERENCES

[1] https://www.who.int/health-topics/coronavirus#tab=tab_1
[2] https://www.thehindu.com/topic/coronavirus/
[3] https://www.kaggle.com/anshuls235/india-vs-covid19/data
[4] https://en.wikipedia.org/wiki/List_of_Indian_states_and_union_territories_by_GDP
[5] https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_population