TABLE OF CONTENTS

CERTIFICATE

This is to certify that **Sagar Agrawal** student of BCA Final Semester enrollment No.**PV-17210839**, January 2020 – June 2020 session of this institute has completed his final semester project entitled "**Data Analysis on Covid-19**".
He has submitted satisfactory project report for the award of degree of Bachelor Of Computer Applications (BCA) of GRAPHIC ERA HILL UNIVERSITY.

Project Guide: Mr. Aditya Harbola
School of Computing
(Assistant Professor, G.E.H.U.)

ACKNOWLEDGEMENT

“No task is a single man’s effort”

This project is a culmination of task undertaken by me. Acknowledgement is not mere formality or ritual but a genuine opportunity to express the indebtedness to all those without whose active support and encouragement this project wouldn’t have been possible. One of most pleasing aspects in collecting the necessary information and compiling it is the opportunity to thank those who have actively contributed to it.

I am extremely grateful to Mr. Aditya Harbolafor his reversed guidance and encouragement, which led to the completion of this project till now. Without her constant appraisal and efforts, this task would have been a mere dream. She was always there to help us throughout this project. She provided us with all the necessary resources and guidance during the project which helped us to complete the project successfully. My special thank to all faculty members for their active support, affectionate guidance and constant encouragement. I am highly indebted to Ms. SonikaKandari (HOD, Department of Computer Applications) for her continuous effort in building a good infrastructure and develop a professional.

Sagar Agrawal
BCA VI Sem
Enrollment No: PV-17210839

ABSTRACT

The outbreak of COVID-19 in different parts of the world is a major concern for all the administrative units of respective countries. India is also facing this very tough task for controlling the virus outbreak and has managed its growth rate through some strict measures.

This study presents the current situation of coronavirus spread in India along with the impact of various measures taken for it. With the help of data sources (till 7th-8th April 2020) from various state units of India and Ministry of Health and Family Welfare, Government of India, this study presents various trends and patterns. This study answers six different research ques-tions in a comprehensive manner. It has been reported that growth rate of infected cases has been controlled with the help of National Lockdown, however some uncontrolled mass level events had negatively impacted the infected cases. With the help of exponential and polyno-mial regression modelling, the predictions of up to 75000 cases have been done by the end of April 2020. It has also been seen that there are some prominent clusters and patient nodes in the network of patients which are the major influencers for COVID-19 spread. Also, death rate case predictions have been done through two-class classification models with an accuracy of 60%.

At the end, strategies for continuation for lockdown has been discussed and present-ed. It appears that only essential services should be open for the citizens of India and the na-tional lockdown should be carried on for next 2-4 weeks. This study will be useful for the Government of India and various states of India, Administrative Units of India, Frontline health workforce of India, researchers and scientists. This study will also be favorable for the administrative units of other countries to consider various aspects related to the control of COVID-19 outspread in their respective regions.

LIST OF FIGURES

LIST OF TABLES

LIST OF GRAPHS

ABBREVIATIONS

PANDAS	pd
MATPLOTLIB.PYPLOT	plt
PLOTLY.EXPRES	px

INTRODUCTION

COVID-19, or more popularly known as Novel Corona Virus, is associated with the respiratory disorder in humans which has been declared as a global epidemic and pandemic in the first quarter of the year 2020 by the World Health Organization. As per the latest data (6th April 2020) by John Hopkins University and other tracking websites, there are currently more than 1.3 million people infected by the Novel Corona Virus all around the world and close to 75 thousand deaths reported from different parts of the world. The top 10 countries with maximum number of infected cases are the United States of America, Spain, Italy, Ger-many, France, China, Iran, United Kingdom, Turkey and Switzerland. The top countries with maximum number of reported deaths are Italy, Spain, United States of America, France andUnited Kingdom. With respect to the recovered patients list, China is at the top of the list followed by Spain, Germany, Italy, Iran and the United States of America. India was placed comfortably out the list of infected nations by huge margins, but recent events led to its rise to 27th position which is a point of concern. The mortality rate is controlled at less than 3% right now, which is better than the ~5.5% mortality rate of world, but the model of spread is slowly moving towards an exponential trend which can lead to massive loss of lives and infrastructure.

India is being looked upon by various nations now as a World Leader and even WHO acknowledged that world is looking towards Indian strategies to contain the outbreak of this epidemic. India accounts for almost one-fifth of the world's population and is second leading country in terms of population in the world. India contributes heavily to the world's GDP and is amongst the most prominent developing country in the world with fairly strong economic growth percentages. India's good camaraderie with majority of the nations in the world and its helpful nature makes it a perfect ally for other countries. Therefore, the analysis of COVID-19 outbreak in Indian region is closely watched and monitored by the World and there is a need of comprehensive analytical studies based on different strategies taken by Indian administrators from time to time.

India has been following a nationwide lockdown since 22-March-2020, which was a one-day lockdown, followed by a 21-day lockdown after two days. Every activity in India since then has been happening with permission from various ad-ministration units and almost all the domestic and international travels have been either banned or monitored closely. India is yet to get into the

third phase of COVID-19 outbreak i.e. the community outbreak as seen by various countries around the world, but the cases have been rising continuously. India's lockdown period has been impacted by two major events in the recent days which were related to the mass exodus of labours and workers from one state to other states (especially from Delhi to neighbouring states) and conduction of a religious event in Delhi which led to spike in the number of cases in various states of India.

During this time, the Indian Prime Minister has been trying to connect with Indian citizens through innovative strategies and coming up with various engagement activities which are impacting the whole nation. With so much happening in India right now, it becomes imperative that we study the current situation and impact of various such events in India through data analysis methods and come up with different plans for future which can be helpful for the Indian administrators and medical professionals.

The current study explores various aspects associated with the COVID-19 outbreak in India and the various regions situated in India.

1.1 DATA ANALYSIS

FIG.1 DATA ANALYSIS

Today we are living in a world that is basically on digitalization. In order to that in daily life we use many digital products, gadgets etc. we can take an simple example of **YOUTUBE**. Every day a huge amount of data is being uploaded on this social platform in the form of videos that is also a kind of data. The ultimate matter is being surrounded by data as we do. Data can be any form like video, image, pdf, sound etc. The technology or process which is used to cleaning, transforming, and modeling data to discover useful information for business decision making. The ultimate purpose of data analysis is to retrieve useful information from the data and taking the appropriate decision on the basis of data analysis.

1.2 APPLICATION AREAS OF DATA ANALYSIS

Data analysis is commonly is used everywhere whenever It comes to data and it is used in policing and security, banks and finance, delivery logistics, customer relationship management, health care and medical field , online searching and surfing etc.Below are the various areas where data analytics applications have been employed:

1.2.1 POLICING/SECURITY

Several cities all over the world have employed predictive analysis in predicting areas that would likely witness a surge in crime with the use of geographical data and historical data. This has seemed to work in major cities such as Chicago, London, Los Angeles, etc. Although, it is not possible to make arrests for every crime committed but the availability of data has made it possible to have police officers within such areas at a certain time of the day which has led to a drop in crime rate.

This shows that this kind of data analytics application will make us have safer cities without police putting their lives at risk.

3. TRANSPORTATION

A few years back at the London Olympics, there was a need for handling over 18 million journeys made by fans in the city of London and fortunately, it were sorted out.

How was this feat achieved? The TFL and train operators made use of data analytics to ensure the large numbers of journeys went smoothly. They were able to input data from events that took place and forecasted a number of persons that were going to travel; transport was being run efficiently and effectively so that athletes and spectators can be transported to and from the respective stadiums.

3.) FRAUD AND RISK DETECTION

This has been known as one of the initial applications of data science which was extracted from the discipline of Finance. So many organizations had very bad experiences with debt and were so fed up with it. Since they already had data that was collected during the time their customers applied for loans, they applied data science which eventually rescued them from the losses they had incurred. This led to banks learning to divide and conquer data from their customers' profiles, recent expenditure and other significant information that were made available to them. This made it easy for them to analyze and infer if there was any probability of customers defaulting.

4.) MANAGE RISK

In the insurance industry, risk management is the major focus. What most people aren't aware of is that when insuring a person, the risk involved is not obtained based on mere information but data that has been analyzed statistically before a decision is made. Data analytics gives insurance companies information on claims data, actuarial data and risk data covering all important decision that the company needs to take. Evaluation is done by an underwriter before an individual insured then the appropriate insurance is set.

These days, analytical software is used for detecting the various forms of fraudulent claims. Risky claims are detected by red flag indicators which can be examined. It is very essential to bring such claims to the attention of administrators, due to the manner at which automation is improving claims processing efficiency.

5.) DELIVERY LOGISTICS

Well, data science and analytics have no limited applications. There are several logistic companies working all over the world such as UPS, DHL, FedEx, etc. that make use of data for improving their efficiency in operations. From data analytics applications, these companies have found the most suitable routes for shipping, the best delivery time, most suitable means of transport to select so as to gain cost efficiency and many others. Also, data generated by these companies through the use of GPS gives them enough opportunities to take advantage of data analytics and data science.

FIG.2 TOP BIG DATA USE CASES

6.) WEB PROVISION

There is this general belief that “Smart Cities” have fast internet speed provided either by their government or companies present there, therefore declaring them smart. Well, just because people can access Facebook or YouTube at the speed of lightning does not necessarily make a city smart.

Although there may be the presence of fast internet but this is just one thing; it needs to be present in the appropriate place and accessed by the right people as well. The key component of this is being able to shift bandwidth at the right time and location. This can only be achieved by the use of data.

The main assumption is that commercial and financial areas should have the highest bandwidth during weekdays while residential areas should get such on weekends. The real truth is that this situation is more complex than it looks and this can only be solved by data analytics application. For example, if a particular community wants to get the attention of web development companies and high-tech industries and make them establish there, a higher bandwidth would be required; only data analytics could get this done effectively.

7.) PROPER SPENDING

Another issue with Smart Cities is the large amount of money spent on little work. Small changes or landmark remodeling which one could dismiss as unnecessary projects consume so much money. Data analytics applications would target where taxpayers’ money would have a major impact on and the kind of work that would be adequate for it. The targeting of where this money should be spent would lead to the entire city’s infrastructure getting a facelift with a reduction of excess money spent.

8.) CUSTOMER INTERACTIONS

This is another one of the applications of data analytics in insurance. Insurers can determine a lot about their services by conducting regular customer surveys mainly after interacting with claim handlers. They could use this to know which of their services are good and the ones that would

need improvement. Various demographics may desire diverse methods of communication like in person interactions, websites, phone or just email. Taking the analysis of customer demographics with feedback can help insurers improve on customer experience depending on customer behavior and proven insights.

A study recently carried out showed that a lack of investment in technology was the cause customer dissatisfaction of the present generation of insurance customers because they prefer using mobile and online channels, social media and other recent mediums to interact with their agents. However, the older generation still prefers the use of the telephone. To improve the overall experience of customers, it is best for insurance companies to provide a wide range of communication methods for their customers.

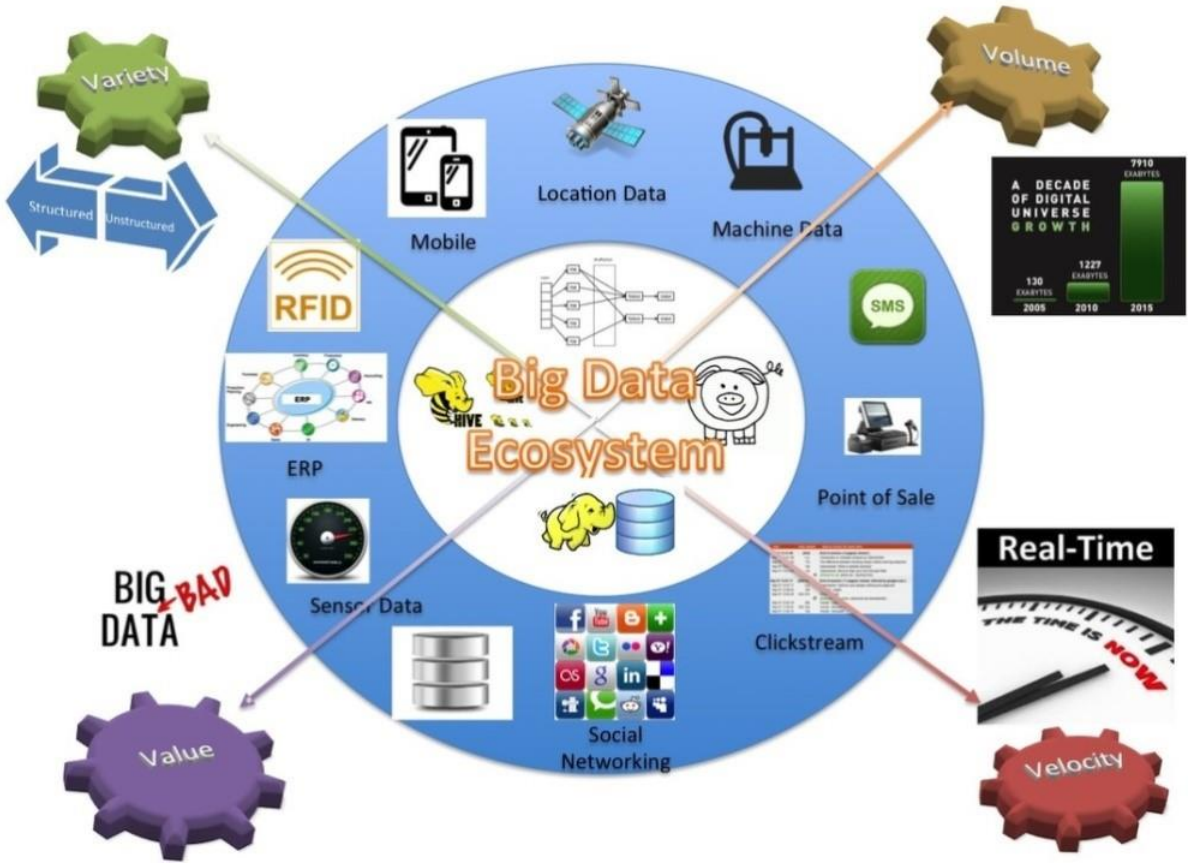


FIG.3 BIG DATA ECOSYSTEM

9.) CITY PLANNING

One big mistake being made in many places is that analytics is not considered when pursuing city planning. As a matter of fact, web traffic and marketing are still being used instead of the creation of spaces and buildings. This really causes a lot of issues to power over data due to its influence on things like building zoning and amenity creation. Models that are built will maximize the accessibility of specific areas or services while the risk of overloading significant elements of the infrastructure in the city is minimized. This implies that it creates efficiency.

We usually see buildings that are built on spots that look suitable but actually have a negative effect on other places. This is because such issues were not considered during the period of planning. Data analytics applications, as well as modeling, would make it easy to mark the outcome of erecting a structure on any spot.

10.) HEALTHCARE

One challenge most hospitals face is coping with cost pressures in treating as many patients as possible, considering the quality of healthcare's improvement. Machine and instrument data use has risen drastically so as to optimize and track treatment, patient flow as well as the use of equipment in hospitals. There is an estimation that a 1% efficiency gain will be achieved and

would result to over $63 billion in worldwide health care services.

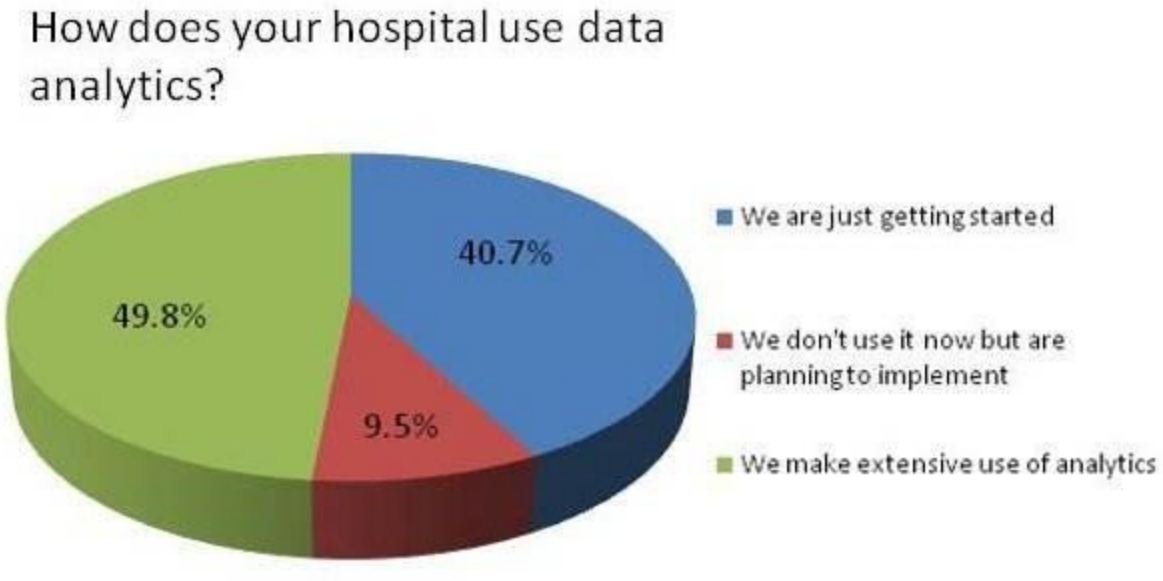


FIG.4USE OF DATA ANALYSIS IN HOSPITAL

11.) TRAVEL

Data analytics applications help in the optimization of traveler's buying experience via social media and mobile/weblog data analysis. This is because customers' preferences and desires can be obtained from this, therefore, making companies sell products from the correlation of the current sales to recent browse-to-buy conversion through customized offers and packages. Data analytics applications can also deliver personalized travel recommendations depending on the outcome from social media data.

12.) ENERGY MANAGEMENT

We are in an era where firms make apply data analytics to energy management and cover areas like energy optimization, smart-grid management, distribution of energy and building automation for utility companies. Data analytics application here focuses mainly on monitoring and

controlling of dispatch crew, network devices and make sure service outages are properly managed. Utilities get the ability to integrate as much as millions of data points within the performance of the network which allows the engineers make use of the analytics in monitoring the network.

13.) INTERNET/WEB SEARCH

When one mentions the word ‘search’, the first thing that comes to the mind is ‘Google’. In fact, Google to some point can be used in place of ‘search on the internet’ by saying ‘Google it’. Well, apart from Google, there are several other search engines such as Bing, Yahoo, Duckduckgo, AOL, Ask, etc. Each of these search engines is as a result of data science applications because they use algorithms to deliver the best results for any search query directed at them in just a split second. In respect to this, Google is known to process over 20 petabytes of data daily. Of course,

without analytics and data science, this feat wouldn’t have been possible.

FIG.5BUSINESS ANALYTICS

14.) DIGITAL ADVERTISEMENT

Apart from web search, there is another area where data analytics and data science serves a very important purpose – digital advertisements. From the banners displayed on several websites to the digital billboards seen in the big cities; all are controlled by data algorithms.

This shows why digital adverts get more CTR than the conventional way of advertisements. Targets depend solely on the past behavior of users.

The importance of data analytics applications cannot be overemphasized because it is used in almost all areas of life today. We can see that having data is very important before making certain decisions so as to avoid unnecessary issues.

Also, handling valuable data inefficiently could lead to several problems like different departments in an organization not understanding how to make use of it which would lead to data

CONCLUSION

However, data has become more available and accessible to more people therefore no longer at the disposal of data scientists and analysts. Almost everybody within an organization can make use of data for the increase of productivity and make very important decisions. Of course, proper use of data would have a positive impact on business and even the society in general.

1.3 ABOUT PROJECT

With the rapid spread in the novel corona-virus across countries, the World Health Organisation (WHO)and several countries have published latest results on the impact of COVID-19 over the past few months.

I have been going through many sources and articles to understand the fatality trend and I was excited to come across this data source and decided to see some visualization on the same. The aim here is to understand how visualization helps to derive informative insights from data sources.

For the visualization part, I am using Plotly. Plotly is a visualization tool available in python, and R which supports a number of interactive, high-quality graphs and is a great tool for data science beginners.

So in this project, we will preprocess and merge datasets to calculate needed measures and prepare them for an Analysis. In this project, we are going to work with the COVID19 dataset, provided by Kaggle, which consists of the data related to the cumulative number of confirmed cases, per day, in each Country. Also, we have another dataset consist of various life factors, scored by the people living in each country around the globe. We are going to merge these two datasets to see if there is any relationship between the spread of the virus in a country and how happy people are, living in that country.

1.4 BACKGROUND

COVID-19 outbreak was first reported in Wuhan, China and has spread to more than 50 countries.WHO declared COVID-19 as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020. Naturally, a rising infectious disease involves fast spreading, endangering the health of large numbers of people, and thus requires immediate actions to prevent the disease at the community level. Therefore, CoronaTracker was born as the online platform that provides latest and reliable news development, as well as statistics and analysis on COVID-19. This paper is done by the research team in the CoronaTracker community and aims to predict and forecast COVID-19 cases, deaths, and recoveries through predictive modelling. The model helps to interpret patterns of public sentiment on disseminating related health information, and assess political and economicinfluence of the spread of the virus.

2. COMPONENTS

2.1 PYTHON

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for rapid application development, as well as for use as a scripting or glue language to connect existing components together. Python's simple,easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

2.2 JUPYTER NOTEBOOK

The jupyter notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter notebook is maintained by the people at project jupyter.

Jupyter notebook are a spin-off projectr from the IPyhton project, which used to have an IPyhton notebook project itself. The name, jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the Ipython kernel, which allows you to write your programs in python, but there are currently over 100 other kernels that you can also use.

2.3 LIBRARIES

2.3.1 PANDAS

It is a popular python based data analysis toolkit which can be imported using "import pandas as pd". It presents a diverse range of utilities, ranging from parsing multiple file formats to converting an entire data table into a "Numpy" matrix array. This makes pandas a trusted ally in data science and machine learning.

```
In [6]: df=pd.read_excel(r"F:\COVID-19-Data-Analyisis-master\Covid cases in India.xlsx")
```

```
In [7]: df
```

Out[7]:

	S. No.	Name of State / UT	Total Confirmed cases (Indian National)	Total Confirmed cases (Foreign National)	Cured	Death
0	1	Andhra Pradesh	12	0	1	0
1	2	Chhattisgarh	6	0	0	0
2	3	Delhi	38	1	6	1
3	4	Gujarat	43	0	0	3
4	5	Haryana	16	14	11	0
5	6	Himachal Pradesh	4	0	0	1
6	7	Karnataka	20	0	3	2
7	8	Kerala	131	7	11	0
8	9	Madhya Pradesh	23	0	0	1
9	10	Maharashtra	144	3	15	4
10	11	Odisha	3	0	0	0
11	12	Puducherry	1	0	0	0

TABLE1. USE OF PANDAS

- In PANDAS we use read_excel() function to read the excel file data into dataframe object.
- Here we can see that dfis used to represent the two dimensional tabular data structure.
- Here is the tabular data of all the cases.

2.3.2 MATPLOTLIB

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002.

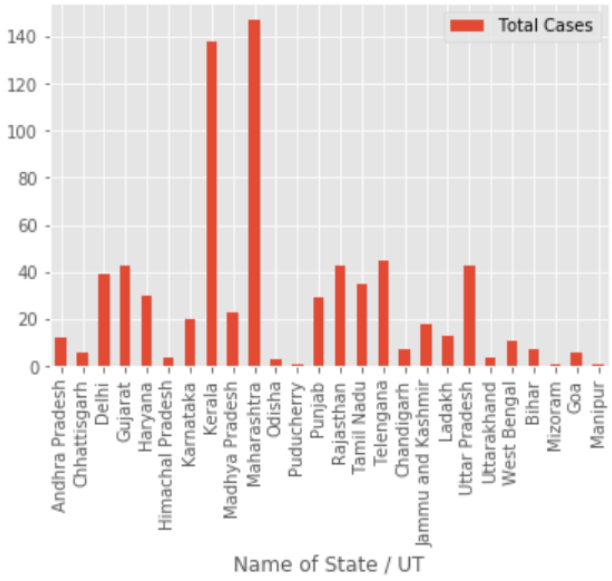
One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.

Funtions such as show(), bar(), scatter(), figure() of the matplotLib are used to present graph bars and other graphical representation in a more user interactive way .

Example given below :

```
In [161]: ################## Graphical Represantaion

In [162]: #Pandas vis
          df.plot(kind='bar',x='Name of State / UT',y='Total Cases')
          plt.show()
          #Plotly
          df.iplot(kind='bar',x='Name of State / UT',y='Total Cases')
```



GRAPH1. TOTAL CASES WITH STATE NAMES

2.3.3 PLOTLY

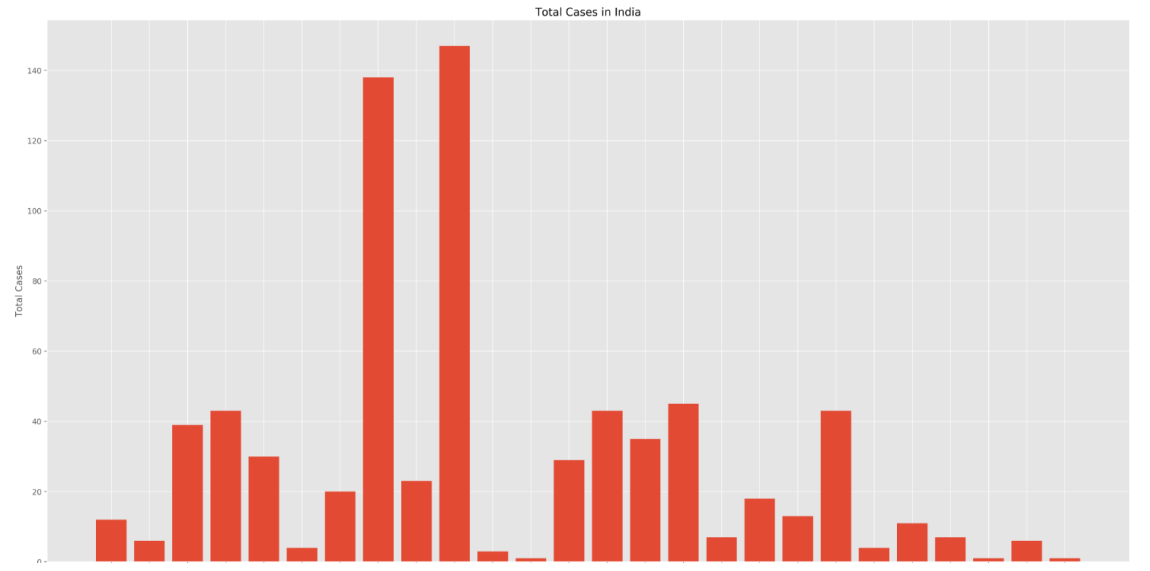
The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

Built on top of the Plotly JavaScript library, plotly.py enables Python users to create beautiful interactive web-based visualizations that can be displayed in Jupyter notebooks, saved to standalone HTML files, or served as part of pure Python-built web applications using Dash.

Funtions such as show(), bar(), scatter(), figure() of the plotlyare used to present graph bars and other graphical representation in a more user interactive way .

Example given below:

```
#plotly
fig=go.Figure()
fig.add_trace(go.Bar(x=df['Name of State / UT'],y=df['Total Cases']))
fig.update_layout(title='Total Cases in India',xaxis=dict(title='Name of State / UT'),yaxis=dict(title='Total Cases'))
```



GRAPH2. TOTAL CASES

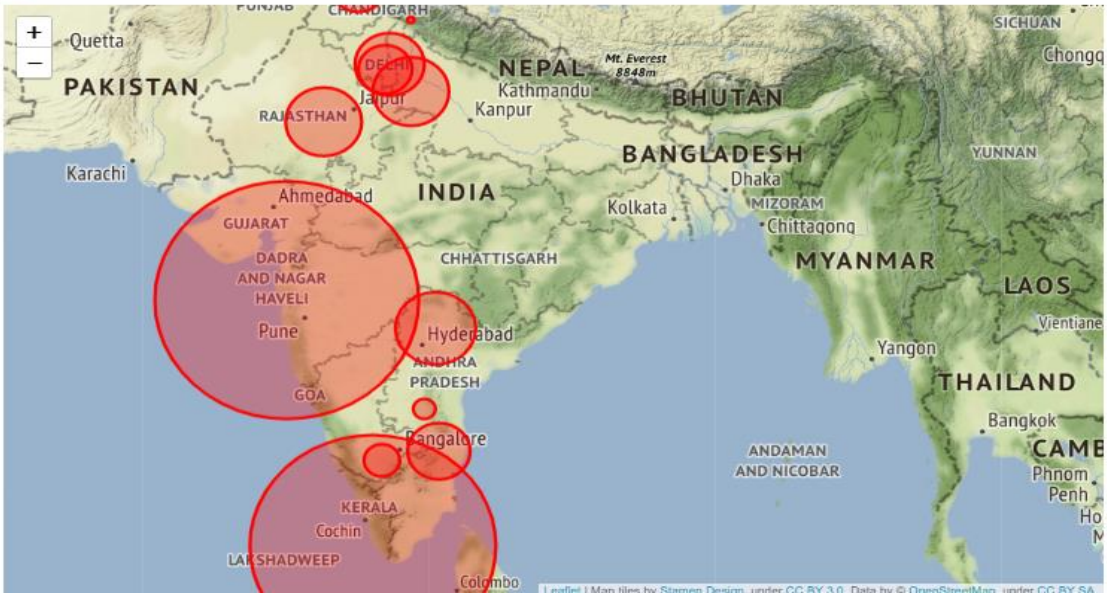
2.3.4 FOLIUM

Folium is a powerful Python library that helps you create several types of Leaflet maps. The fact that the Folium results are interactive makes this library very useful for dashboard building. To get an idea, just zoom/click
around on the next map to get an impression.

```
In [173]: map=folium.Map(location=[20,70],zoom_start=4,tiles='Stamenterrain')

for lat,long,value, name in zip(df_full['Latitude'],df_full['Longitude'],df_full['Total Cases'],df_full['Name of State / UT']):
    folium.CircleMarker([lat,long],radius=value*0.8,popup=('<strong>State</strong>: '+str(name).capitalize()+'<br>''<strong>Tota
l Cases</strong>: ' + str(value)+ '<br>'),color='red',fill_color='red',fill_opacity=0.3).add_to(map)

In [174]: map
Out[174]:
```



GRAPH3. CASES REPRESNTED IN MAP

By default, Folium creates a map in a separate HTML file. In case you use Jupyter (like myself), you might prefer to get inline maps. This Jupyter example shows how to display maps inline.

- **Folium()= Make beautiful, interactive maps with Python and Leaflet.js**
- **The map representation plays a major role in presenting a project, because in this a lot of effort is not wasted and is also easy to understand. Anyone can judge the project by just having a look at it.**
- **In this map we can clearly see that Maharashtra and Kerela and Delhi are the major paces that have the maximum number f the cases of the pandemic.**

2.3.5 CUFFLINKS

There also exists an independent third-party wrapper library around Plotly called Cufflinks, which provides similar functionality (with an API closer to that of Pandas' default matplotlib backend) by adding a .iplot() method to Pandas dataframes, as it was developed before Pandas supported configurable backends. Issues and questions regarding Cufflinks should be raised in cufflinks repository.

Example in the project :

```
In [147]: pyo.init_notebook_mode(connected=True)
          cf.go_offline()
```

3. METHODOLOGY

To answer the different research questions, specific methodologies have been used which covers up different data set, data sources, modeling techniques and outcome variables. The overall variables covered up in the study are shown in figure below.

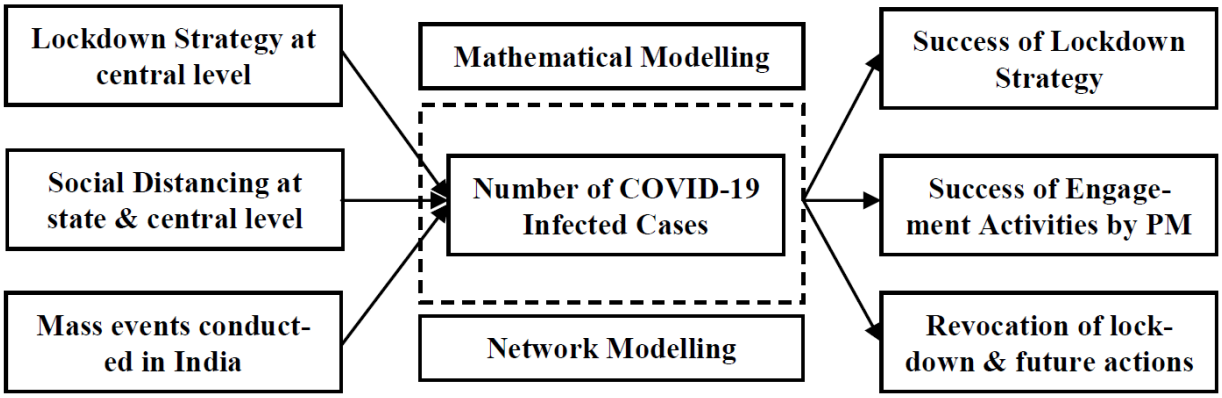


Figure 1. Research Model for the current study covering various variables and outcome measures.

For the first research question, the time series data from the time period of 30th January 2020 till 7th April 2020 has been covered from the Indian database of COVID-19 . This dataset was divided into three parts of 30th January 2020 to 4th March 2020, 5th March to 22nd March 2020 and 23rd March 2020 to 7th April 2020. Trend analysis and average number of infected cases were compared at both the national level. On the same dataset, the second re-search question has been answered using the Exponential Modelling to predict the short term trends for the next three weeks.

Also, the number of operational places and events conducted during lockdown data has been analyzed through news reports and channels. Based on the news reports, the major events were traced and people getting infected due to such events have been covered from the Indian data repository created using crowd sourcing channels. The patients were categorized as infected from the event as compared to other patients who were not related to the event. The trends and patterns were analyzed after conduction of this event i.e. in the first week of April 2020. The data was also confirmed from various news reports in India.

4. FINDING AND DISCUSSION

TABLE2. DAILY REPORTS DATA

This csv file contains information on the affected countries [in blue] which helps to identify the virus spread, information on infected cases, number of deaths and recoveries across countries. The country co-ordinates are also provided for analysis.

	Name of State / UT	Latitude	Longitude
0	Andaman And Nicobar	11.667026	92.735983
1	Andhra Pradesh	14.750429	78.570026
2	Arunachal Pradesh	27.100399	93.616601
3	Assam	26.749981	94.216667
4	Bihar	25.785414	87.479973
5	Chandigarh	30.719997	76.780006
6	Chhattisgarh	22.090420	82.159987
7	Dadra And Nagar Haveli	20.266578	73.016618
8	Delhi	28.669993	77.230004
9	Goa	15.491997	73.818001
10	Haryana	28.450006	77.019991
11	Himachal Pradesh	31.100025	77.166597
12	Union Territory of Jammu and Kashmir	33.450000	76.240000
13	Jharkhand	23.800393	86.419986
14	Karnataka	12.570381	76.919997
15	Kerala	8.900373	76.569993
16	Lakshadweep	10.562573	72.636867
17	Madhya Pradesh	21.300391	76.130019
18	Maharashtra	19.250232	73.160175
19	Manipur	24.799971	93.950017
20	Meghalaya	25.570492	91.880014
21	Mizoram	23.710399	92.720015
22	Nagaland	25.666998	94.116570
23	Orissa	19.820430	85.900017
24	Puducherry	11.934994	79.830000
25	Punjab	31.519974	75.980003
26	Rajasthan	26.449999	74.639981
27	Sikkim	27.333330	88.616647
28	Telengana	18.112400	79.019300
29	Tamil Nadu	12.920386	79.150042
30	Tripura	23.835404	91.279999
31	Uttar Pradesh	27.599981	78.050006

TABLE3. TIME-SERIES DATA:

A time series data which contains the counts on infected cases, deaths and recoveries across countries is also given. The time series data has individual files for each case and needs to be processed before visualization.

	SNo	Date	Province/State	Country	Last Update	Confirmed	Deaths	Recovered	Lat	Long	...	3/5/20	3/6/20	3/7/20	3/8/20	3/9/20	3/10/
0	32	01/22/2020	Washington	US	2020-01-22 17:00:00	1.0	0.0	0.0	47.4009	-121.4905	...	0	0	0	0	0	2
1	70	01/23/2020	Washington	US	2020-01-23 17:00:00	1.0	0.0	0.0	47.4009	-121.4905	...	0	0	0	0	0	2
2	118	01/24/2020	Washington	US	2020-01-24 17:00:00	1.0	0.0	0.0	47.4009	-121.4905	...	0	0	0	0	0	2
3	159	01/25/2020	Washington	US	2020-01-25 17:00:00	1.0	0.0	0.0	47.4009	-121.4905	...	0	0	0	0	0	2
4	203	01/26/2020	Washington	US	2020-01-26 16:00:00	1.0	0.0	0.0	47.4009	-121.4905	...	0	0	0	0	0	2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
3285	5879	03/15/2020	NaN	Saint Vincent and the Grenadines	2020-03-14 16:33:03	1.0	0.0	0.0	12.9843	-61.2872	...	0	0	0	0	0	
3286	5625	03/14/2020	NaN	Suriname	2020-03-14 16:33:03	1.0	0.0	0.0	3.9193	-56.0278	...	0	0	0	0	0	
3287	5881	03/15/2020	NaN	Suriname	2020-03-14 16:33:03	1.0	0.0	0.0	3.9193	-56.0278	...	0	0	0	0	0	
3288	5628	03/14/2020	Virgin Islands, U.S.	US	2020-03-14 16:15:18	1.0	0.0	0.0	18.3358	-64.8963	...	0	0	0	0	0	
3289	5885	03/15/2020	Virgin Islands, U.S.	US	2020-03-14 16:15:18	1.0	0.0	0.0	18.3358	-64.8963	...	0	0	0	0	0	

Analysis:

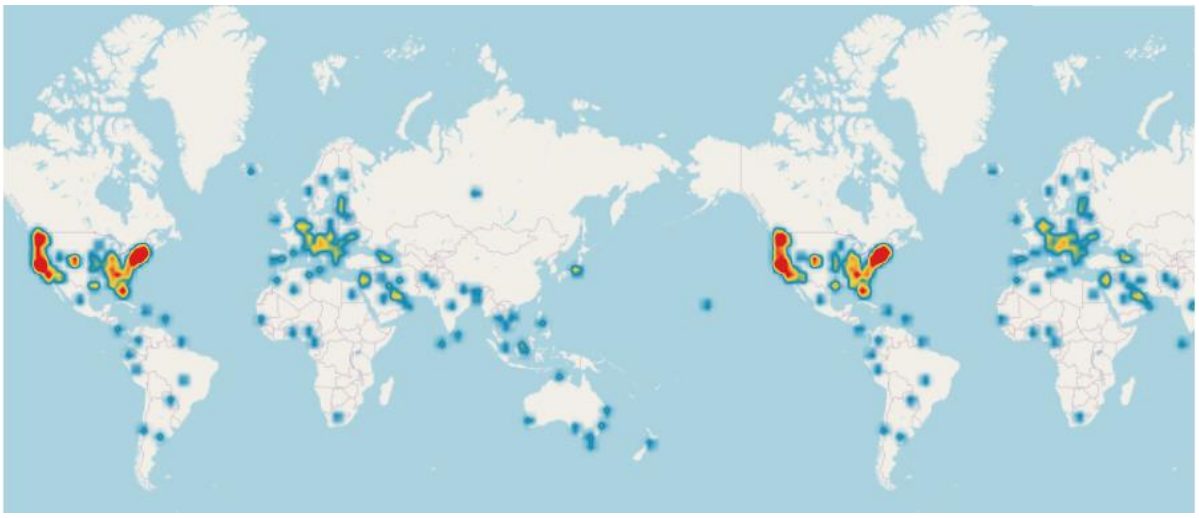
1. The global impact of COVID-19

2. Descriptive analysis on infected, mortality and recovery rates

3. Timeline analysis on spread of COVID-19 between [Jan — Mar]

THE GLOBAL IMPACT OF COVID-19

To understand the impact of the virus on a geographical landscape, I used the geographical scatter plot from Plotly. The code for this interactive plot is found in the shared link which will give a more clear interactive visualization.



Global impact of COVID-19

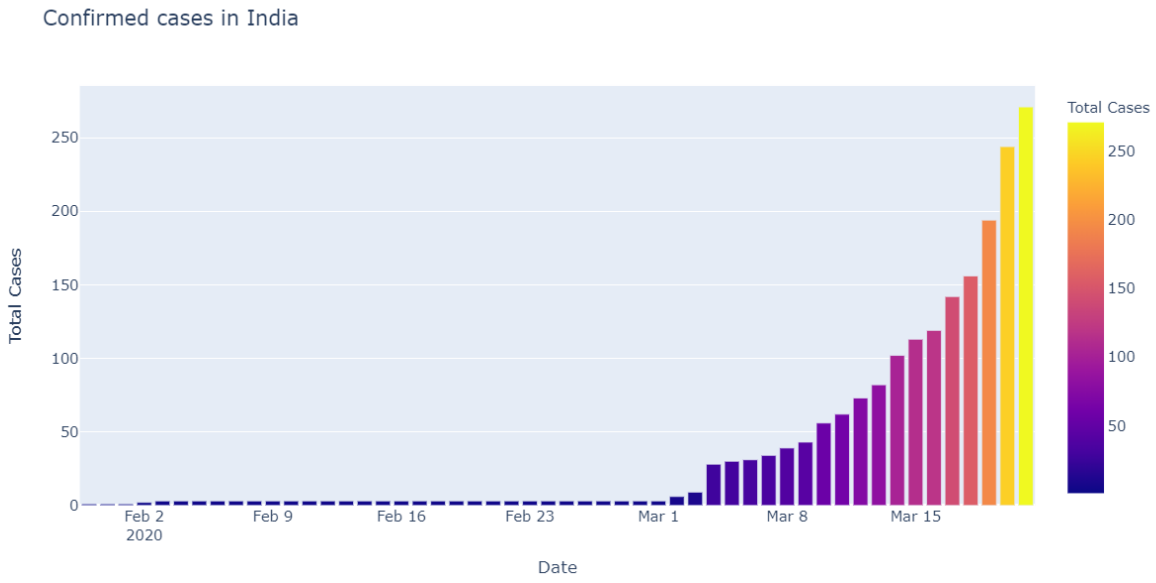
To understand the impact of the virus on a geographical landscape, I used the geographical scatter plot from Plotly. The code for this interactive plot is found in the shared link which will give a more clear interactive visualization.

DESCRIPTIVE ANALYSIS ON INFECTED, MORTALITY AND RECOVERY RATES

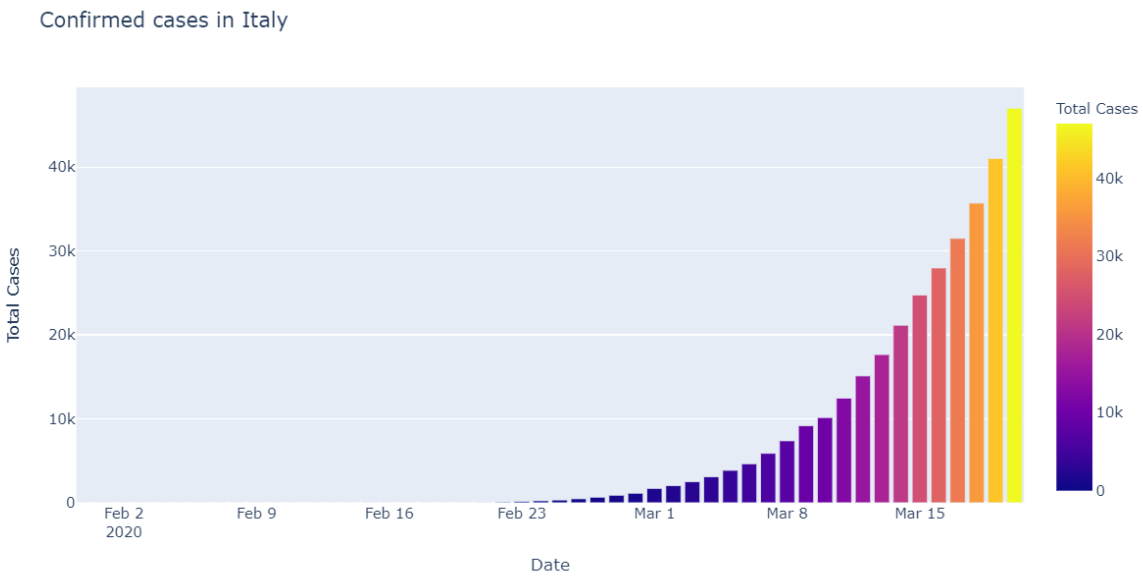
Here I have used various charts to show how information can be mined from data sources.

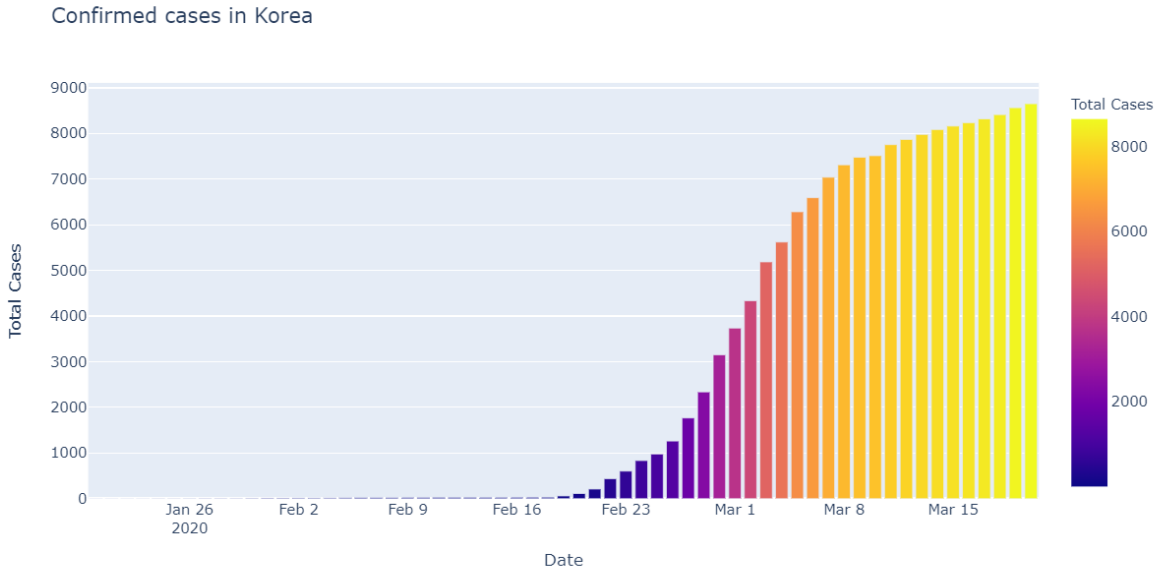
Note: Since China has a higher rate of infected cases which is greater than 85% compared to other countries, it would be good to exclude China and see the numbers on other countries. This will be applied to all the following charts.

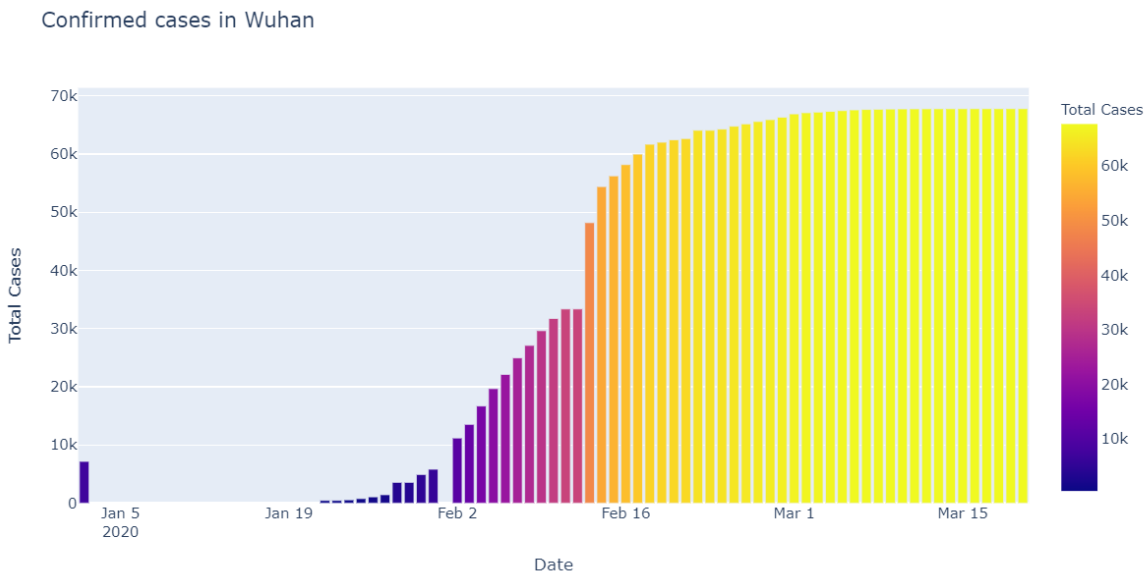
1. **Confirmed cases:** using bar representation



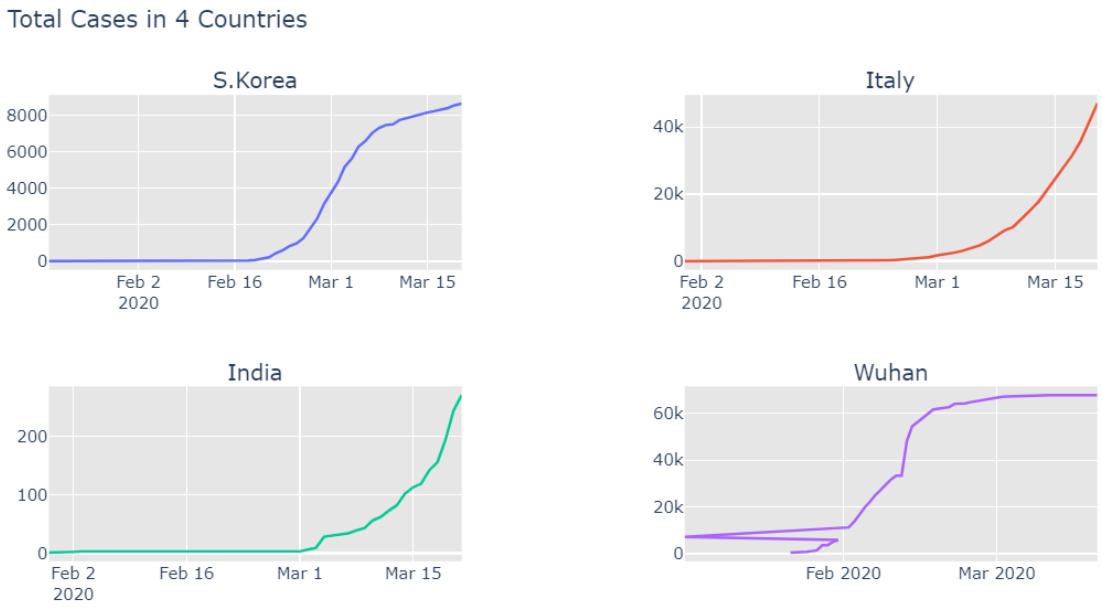
GRAPH4. CONFIRMED CASES USING BAR


Confirmed cases in Italy
Total Cases
40k
30k
20k
10k
0
Feb 2
2020
Feb 9
Feb 16
Feb 23
Mar 1
Mar 8
Mar 15
Date
Total Cases
40k
30k
20k
10k
0


Confirmed cases in Korea
Total Cases
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Jan 26
2020
Feb 2
Feb 9
Feb 16
Feb 23
Mar 1
Mar 8
Mar 15
Date
Total Cases
8000
6000
4000
2000



2. **Confirmed cases:** using line graph representation



GRAPH5. CONFIRMED CASES USING GRAPH REPRESENTATION

2. OVERALL ANALYSIS TABLE: USING GRADIENT STYLE

	Name of State / UT	Total Confirmed cases (Indian National)	Total Confirmed cases (Foreign National)	Cured	Death	Total Cases	Active Cases
0	Andhra Pradesh	12	0	1	0	12	11
1	Chhattisgarh	6	0	0	0	6	6
2	Delhi	38	1	6	1	39	32
3	Gujarat	43	0	0	3	43	40
4	Haryana	16	14	11	0	30	19
5	Himachal Pradesh	4	0	0	1	4	3
6	Karnataka	20	0	3	2	20	15
7	Kerala	131	7	11	0	138	127
8	Madhya Pradesh	23	0	0	1	23	22
9	Maharashtra	144	3	15	4	147	128
10	Odisha	3	0	0	0	3	3
11	Puducherry	1	0	0	0	1	1
12	Punjab	29	0	0	1	29	28
13	Rajasthan	41	2	3	0	43	40
14	Tamil Nadu	32	3	1	1	35	33
15	Telengana	34	11	1	0	45	44
16	Chandigarh	7	0	0	0	7	7
17	Jammu and Kashmir	18	0	1	1	18	16
18	Ladakh	13	0	0	0	13	13
19	Uttar Pradesh	42	1	11	0	43	32
20	Uttarakhand	4	0	0	0	4	4
21	West Bengal	11	0	0	1	11	10
22	Bihar	7	0	0	1	7	6
23	Mizoram	1	0	0	0	1	1
24	Goa	6	0	0	0	6	6
25	Manipur	1	0	0	0	1	1

TABLE4. SUMMARY OF CASES WITH COLOR GRADIENTS

5.BIBLOGRAPHY

1. Rajan Gupta1, Saibal K. Pal2 and Gaurav Pandey," A Comprehensive Analysis of COVID-19 Outbreak situation in India"medRxiv - the preprint server for Health Sciences, Volume:04, Issue:03, March 2020 (e-ISSN:2395:0056, o-ISSN:2395-0076).

3. Santhiya , "The impact of COVID-19 — data visualization using Plotly and comparative analysis with SARS" Issue 02, March,2020.

4. Nadia AL-RousanHazem AL-Najjar , “Data Analysis of Coronavirus CoVID-19 Epidemic in South Korea Based on Recovered and Death Cases” *Journal of Medical Virology* ;Issue:08, April 2020.

5. Wen Li ,Qinge Zhang, "Timely research papers about COVID-19 in China",Published:February 17, 2020.

6. Huang, Canping; Qi, Jianxun; Lu, Guangwen; Wang, Qihui; Yuan, Yuan; Wu, Ying; Zhang, Yanfang; Yan, Jinghua; Gao, George F; (2016). "Putative Receptor Binding Domain of Bat-Derived Coronavirus HKU9 Spike Protein: Evolution of Betacoronavirus Receptor Binding Motifs", Biochemistry, 55, 5977-5988.

7. World Health Organization (2020). Coronavirus disease (COVID-19) Pandemic, WHO. Accessed from https://www.who.int/emergencies/diseases/novel-coronavirus-2019 on 31st March 2020.

8. John Hopkins University (2020). Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE. Accessed from https://github.com/CSSEGISandData/COVID-19 on 6th April 2020.

9. Sharma, N. (2020). India’s swiftness in dealing with Covid-19 will decide the world’s future, says WHO, Quartz India. Accessed from https://qz.com/india/1824041/who-saysindias-action-on-coronavirus-critical-for-the-world/ on 25th March 2020.

10. Gupta, R., Pandey, G., Chaudhary, P., & Pal, S. K. (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. medRxiv. Accessed from https://www.medrxiv.org/content/10.1101/2020.04.01.20049825v1 on 5th April 2020

11. Tanne, J. H., Hayasaki, E., Zastrow, M., Pulla, P., Smith, P., & Rada, A. G. (2020). Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. Bmj, 368.

12. Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., ...& Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery.

13. Grasselli, G., Pesenti, A., &Cecconi, M. (2020). Critical care utilization for the COVID19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. JAMA.

14. Covid19India.org (2020a). India Covid-19 Tracker 2020. Accessed from https://www.covid19india.org/ on 7th April 2020

15. BBC News (2020). Coronavirus: Search for hundreds of people after Delhi prayer meeting, BBC News. Accessed from https://www.bbc.com/news/world-asia-india-52104753 on 2nd April 2020

16. Grasselli, G., Pesenti, A., &Cecconi, M. (2020). Critical care utilization for the COVID19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. JAMA.