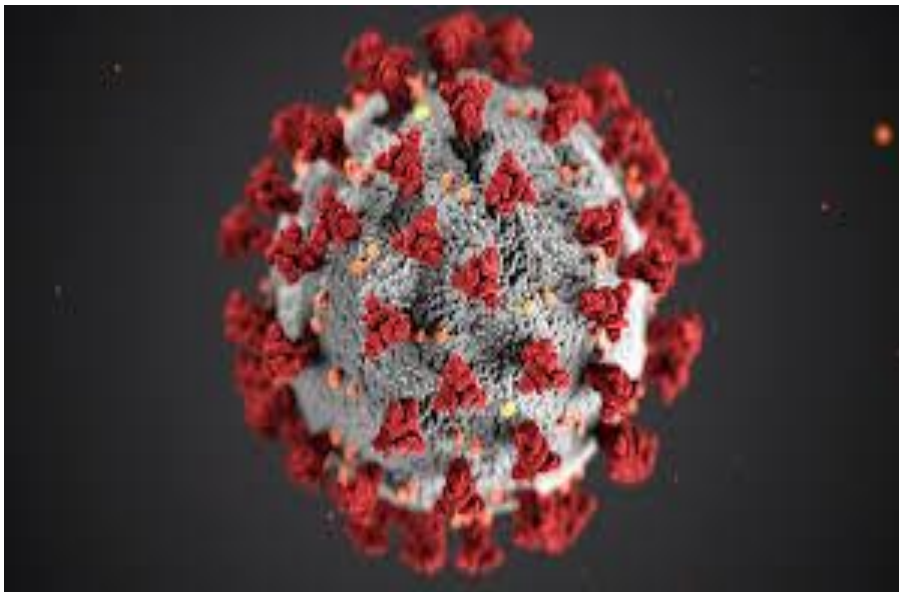


**Ravirajsinh Dabhi**

**Subject:** Applied Data Science Capstone  
(Full Report)

**Title:** Analysis on Covid-19 in India



# Table of Content

## Contents

<b>Introduction</b> .....	3
Steps: .....	3
Aim:- .....	3
<b>Data section</b> .....	4
<b>Methodology</b> .....	6
Bar Charts .....	7
Choropleth Map .....	8
I made choropleth map of India of Cases and death wise allocation.....	8
Using Foursquare API to find hospital in Mumbai .....	9
<b>Result</b> .....	10
<b>Conclusion</b> .....	10

## **Introduction**

corona virus is spread all over the world and in this situation I am doing IBM Data Science course so I could not find any other topic then Covid-19.

All over news channel, website and newspaper there are only one thing is CORONA. We have so much data and records about this disease so I decide to do some analysis on it. There are tons of data available like worldwide cases and deaths, country wise cases and deaths and many more. I decided to do analyse on India's all states cases and deaths from first case in the India to till 30<sup>th</sup> April, 2020.

This type of data is easily available on internet but not in direct way. We have first find the good source. After that we have to collect it and have to some data wrangling and manipulation so it can ready for analysis.

Steps:

1. Find data
2. Scrap from website
3. Process the data
4. Visualization

Aim:-

**Our main aim is to play with data about corona virus in the India. I will try to find some insights and visualization.**

**I will try to find some high rated hospitals using FOURSQUIRE API in state which have highest cases.**

## Data section

Data is most essential component of any data science project. After searching stuffs in internet I came up with one table in Wikipedia page [https://en.wikipedia.org/wiki/2020\\_coronavirus\\_pandemic\\_in\\_India](https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_India).

en.wikipedia.org/wiki/2020\_coronavirus\_pandemic\_in\_India

New COVID-19 cases in India by state and union territory (v·t·e) [hide]

Date (2020)	State/Union Territory																													Cases <sup>[a]</sup>		Deaths		Source(s)					
	Andaman and Nicobar Islands	Andhra Pradesh	Arunachal Pradesh	Assam	Bihar	Chandigarh	Chhattisgarh	Delhi	Goa	Gujarat	Haryana	Himachal Pradesh	Jammu and Kashmir	Jharkhand	Karnataka	Kerala	Ladakh	Madhya Pradesh	Maharashtra	Manipur	Meghalaya	Mizoram	Nagaland	Odisha	Puducherry	Punjab	Rajasthan	Tamil Nadu	Telangana	Tripura	Uttarakhand	Uttar Pradesh	West Bengal		New	Total	Diff.	New	Total
Jan-30																1																	1	1	—				
Feb-02																1																	1	2	+100%				
Feb-03																1																	1	3	+50%				
Mar-02								1																						1			2	5	+67%			[482]	
Mar-03																											1						1	6	+20%			[483]	
Mar-04								1			14																1					6	22	28	+356%			[484][485]	
Mar-05								1																							1		2	30	+7%			[485]	
Mar-06								1																									1	31	+3%			[485]	
Mar-07																	2												1				3	34	+10%			[485]	
Mar-08																5																	5	39	+15%			[486][487]	
Mar-09								1					1				1									1						1		5	44	+13%			[488][489]
Mar-10															4				2														6	50	+14%			[490]	
Mar-11								1								8											1						10	60	+20%			[491]	
Mar-12		1						2							0(1)		1		6								1				2		13	73	+22%	1	1	[492][493]	
Mar-13															2	2			3													1		8	81	+11%		1	[494][495]
Mar-14								1(1)					1						12								1					1		16	97	+20%	1	2	[496]
Mar-15																3			5											2			10	107	+10%			2	[497][498]
Mar-16													1		1	1	1		6					1									11	118	+10%		2		
Mar-17								1			1				3	3	4		3(1)										2		2		19	137	+16%	1	3		
Mar-18								2			2				1	1			3						1		1				2	1	14	151	+10%		3		

This table is what actually I want. It contains day to day record of corona cases in all states of the India from first case in record in India. It has all data which is necessary to do various kind of analysis. It contains each day cases and deaths of all states, Difference from previous day recorded data, states total cases and death and whole country's total deaths and cases.

I scrap the Wikipedia web page with help of BeautifulSoup python library and convert it in pandas dataframe. Finding particular table require some trial and error. Some code in below image.

```
[4]: res = requests.get("https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_India")
      soup = BeautifulSoup(res.content, 'lxml')
```

```
[5]: table = soup.find_all('table')[4]
      df= pd.read_html(str(table))
```

```
[6]: day_to_day = pd.DataFrame(df[0]).T
      day_to_day.head()
```

[6]:

		0	1	2	3	4	5	6	7	8	9	...	56	57	58	59	60	61	62		
	Date (2020)	Date (2020)	Jan-30	Feb-02	Feb-03	Mar-02	Mar-03	Mar-04	Mar-05	Mar-06	Mar-07	Mar-08	...	Apr-24	Apr-25	Apr-26	Apr-27	Apr-28	Apr-29	Total	De
State/Union Territory	Andaman and Nicobar Islands	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	4	5	6	NaN	NaN	NaN	33	
	Andhra Pradesh	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	60(2)	106(2)	36	80	82	73	1332	
	Arunachal Pradesh	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	1	
	Assam	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	1	NaN	NaN	NaN	2	NaN	38	

5 rows × 66 columns

There are many operation which you will find out in notebook. After this visualization of data I am using foursquare API to find top highest rated hospitals in state in which corona cases is highest.

I found out that Maharashtra state has highest cases in india and particular Mumbai city. So I use foursquare to find highest rated hospital in Mumbai.

I will use folium and geopy python library to visualization and find location of particular address. You can see step by step explanation and code in Notebook.

## Methodology

This raw data table has lot's of missing values. If there is no cases or death record for particular day that have null value. I have to convert it into zero(0). And most of data is 'object' Datatype and most of is 'string'. From string data type we can't do exploratory analysis of data.

Some of the data is in weird format like below image. Which I have to convert in appropriate format.

56(2)	82(3)	60(2)	106(2)	36	80	82	1259	31
0	0	0	0	0	0	0	1	–
0	0	1	0	0	0	2	38	1
12[d]	22	28	52	23	26	69	346	2
1	0	0	1	2	0	10	40	–
0	0	0	0	1	0	0	37	–
75	92(1)	128(2)	138(3)	111(1)	293	190	3108	54
0	0	0	0	0	0	0	7	–
206(18)	135(8)	217(9)	191(15)	256(6)	230(18)	247(11)	3548	162

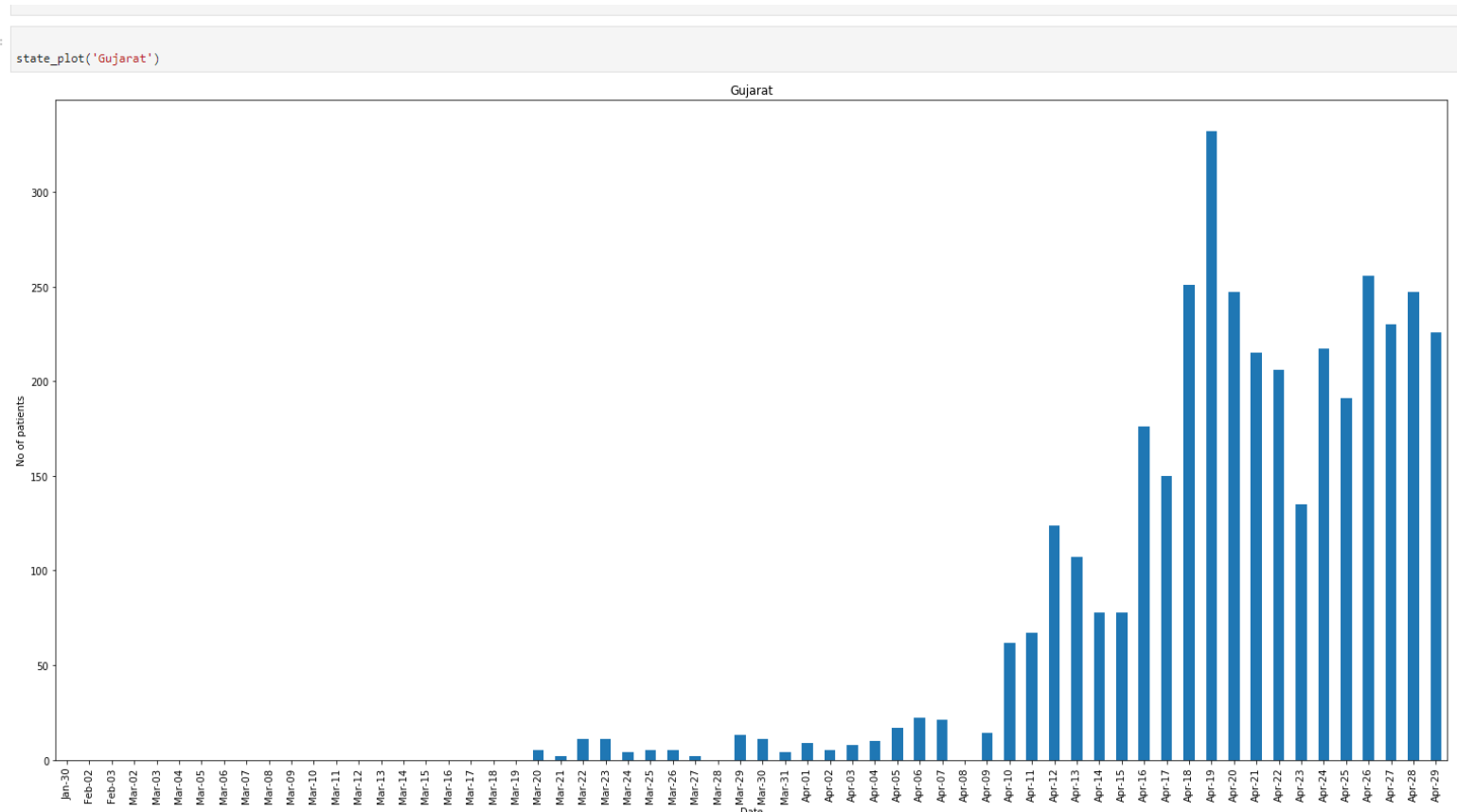
I converted into like below.

128	138	111	293	190	206	125	3439	56
0	0	0	0	0	0	0	7	0
217	191	256	230	247	226	308	4082	197
10	0	17	0	7	14	0	310	3
0	0	0	0	0	0	0	40	1
20	27	40	29	23	19	16	581	8
6	4	8	15	21	2	2	107	3
20	26	12	10	9	12	25	557	21
10	3	7	11	13	4	10	496	4

## Bar Charts

I end up with good Data Frame. From which I started to make some visualization. I made function in which you have to enter name of any state of india and you will get bar chart of cases on each day till 30<sup>th</sup> April, 2020.

Check function in notebook. Here I have example of Gujarat state.

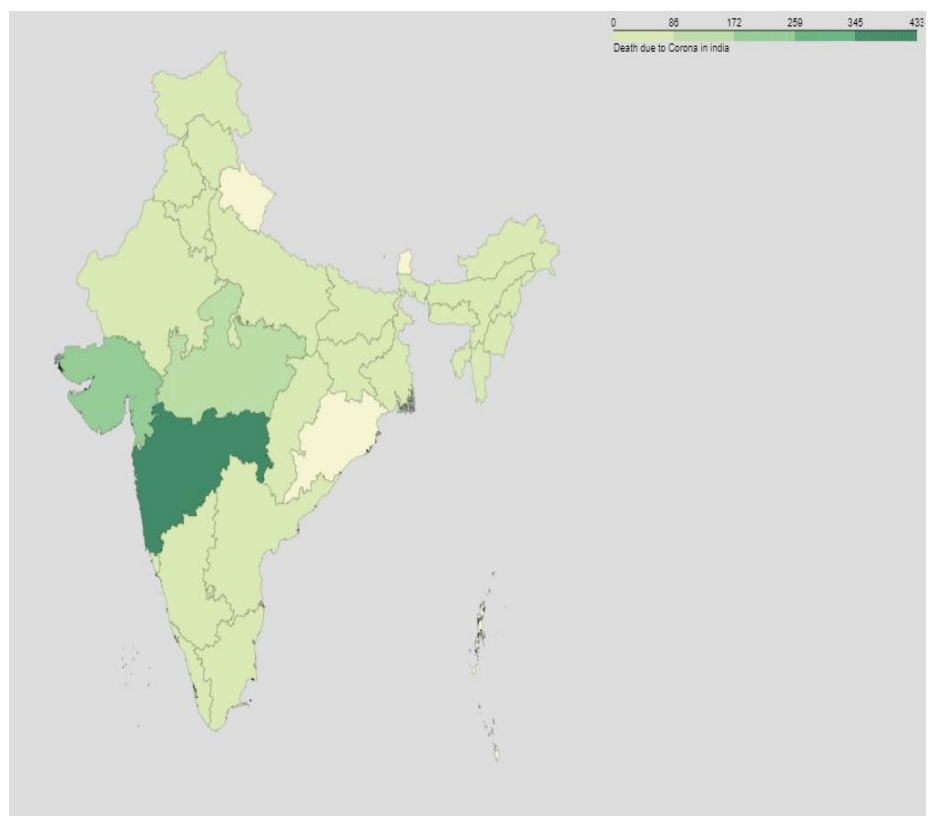
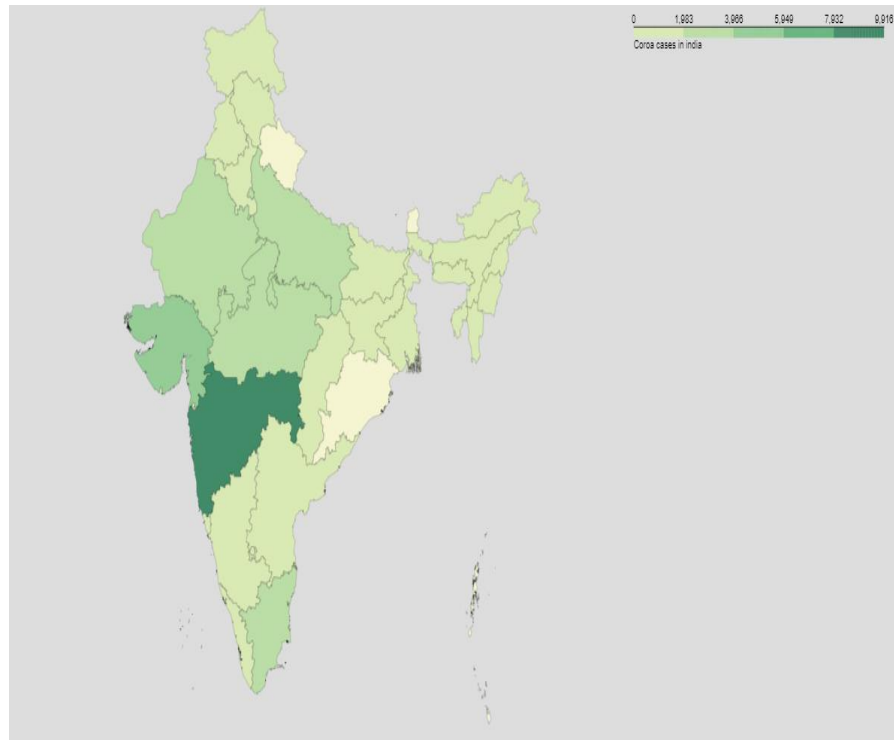


In Gujarat first case record on 20<sup>th</sup> March, 2020. Highest cases record on 19<sup>th</sup> April, 2020.

There are other state map in notebook.

## Choropleth Map

I made choropleth map of India of Cases and death wise allocation.

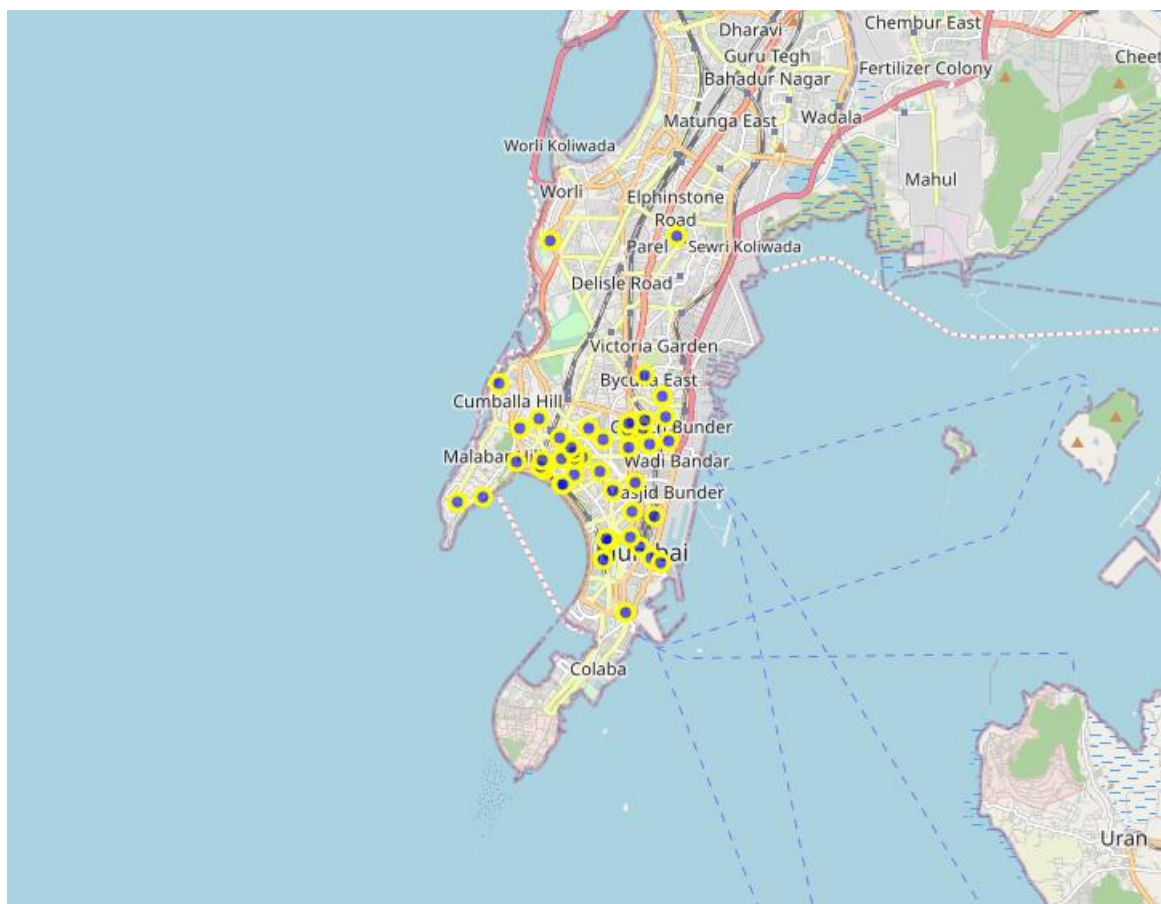




## Using Foursquare API to find hospital in Mumbai

I used Foursquare API to find out highest rated hospital in Mumbai city.

- First of all I used Geopy library to find out loction of Mumbai city.
- After that I used foursquare api to find out 50 hospitals in Mumbai city.
- Made dataframe of json file of hospital.
- Made Mumbai city map in folium.
- And last made location of each hospital in Mumbai map using folium.



## **Result**

- I end up with cleaned data of corona in each state in India.
- Find some comparison between various state on each Date recorded cases and death.
- Visualize data in many different ways.
- Used Foursquare API to find hospital name and location and mapped using folium.

## **Conclusion**

In this study, I learn about data scraping and use it for further analysis. Using different library I can find impact corona on each states, comparison between them, and visualized them. From this project I learn many things. At the end I would like to thank IBM for giving us lovely experience.

**Thank You**