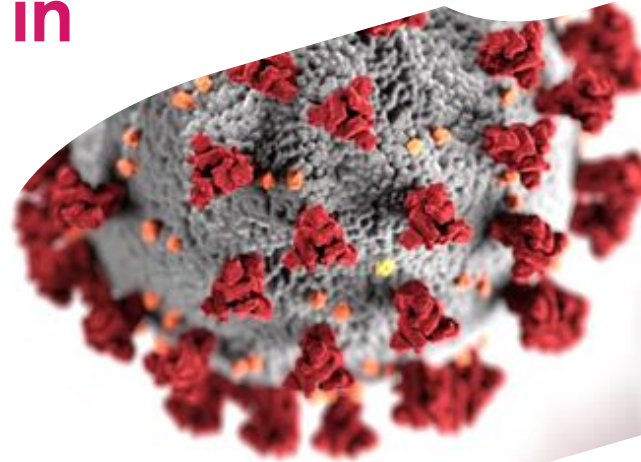# Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

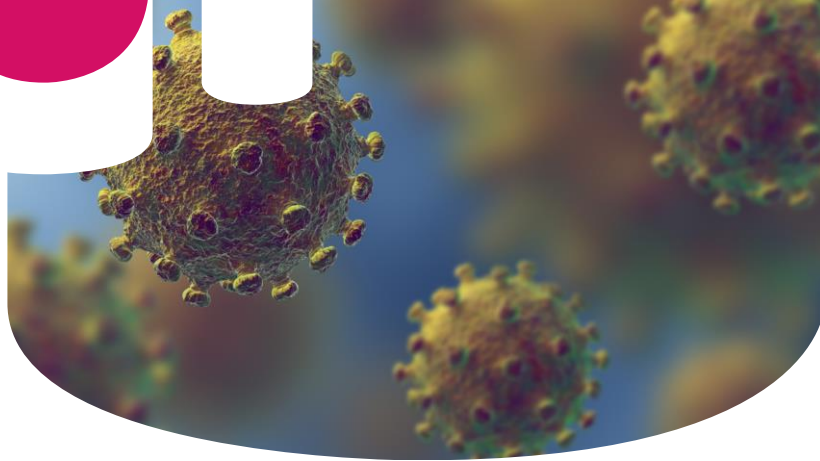**Saquib Mohiuddin Siddiqui**

Capstone Project

IBM Data Science Professional Certificate

July
2020

The Story Behind

Key Insights into COVID

- **COVID-19**) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

- First identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic

- As of 27 July 2020, more than 16.2 million cases have been reported across 188 countries and territories, resulting in more than 648,000 deaths. More than 9.4 million people have recovered.

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

2

The Problem at Hand

A Country with over a Billion People.

India is the second most populous country with a population of over 1.3 Billion, and a population density of 700 per sq. km and boasts an incredibly diverse culture

- it is at one of the highest risk of community transmission and uncontainable spread of the disease so it is compelling to have quality and adequate health care facilities in place to provide necessary health care sufficiently to all
- In this Project we will analyze and visualize the transmission and deaths caused by the virus in India using open source data.

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

3

DATA

Data Acquisition and Wrangling

First Data Set

- For the first data set, the Data has been acquired from open source website https://www.covid19india.org/ using their API's. https://api.covid19india.org/documentation/csv/

- District wise data have acquired from the above source

Second Data Set

- For the second data set, the data for all neighborhoods in Hyderabad have been acquired via Web Scraping using BeautifulSoup, I've referred to Wikipedia https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India

- The Missing Data has been appended using list function

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

4

Rendered Map of the city under Analysis

```
In [5]:  1  hyderabad_map = folium.Map(location = [float(HYD_Latitude), float(HYD_Longitude)], zoom_start = 12)
         2  hyderabad_map.save("hyderabad_map.html")
         3  hyderabad_map
Out[5]:
```
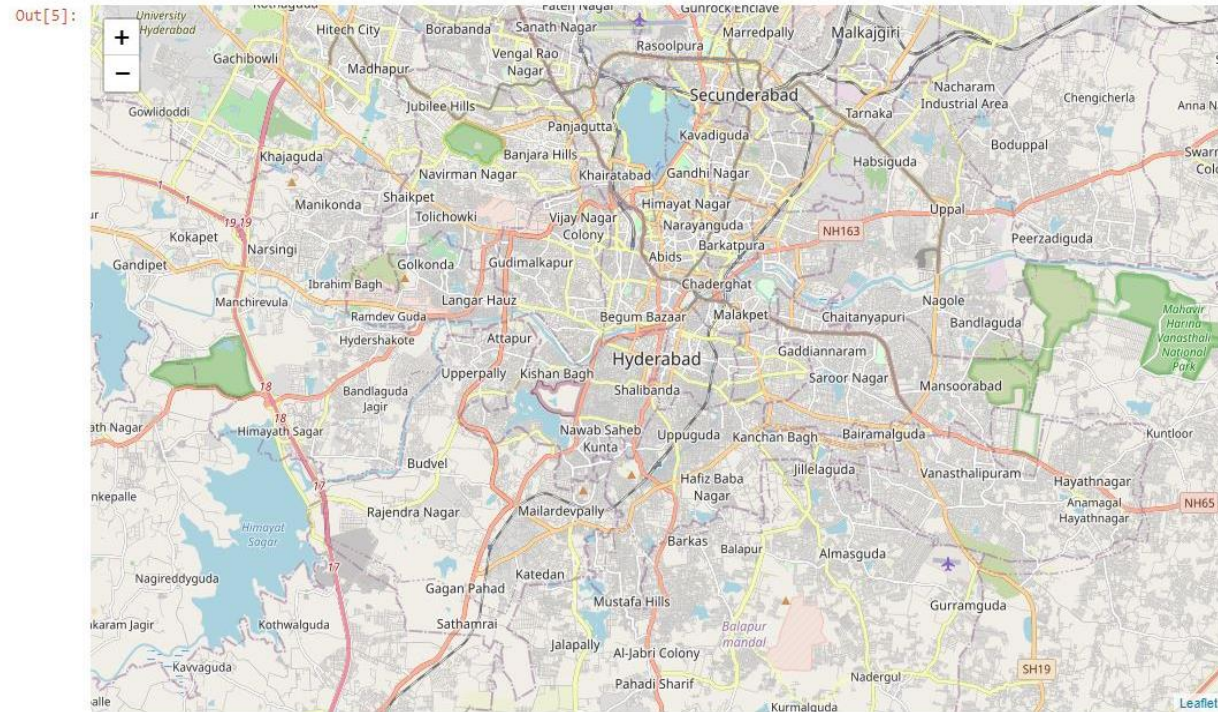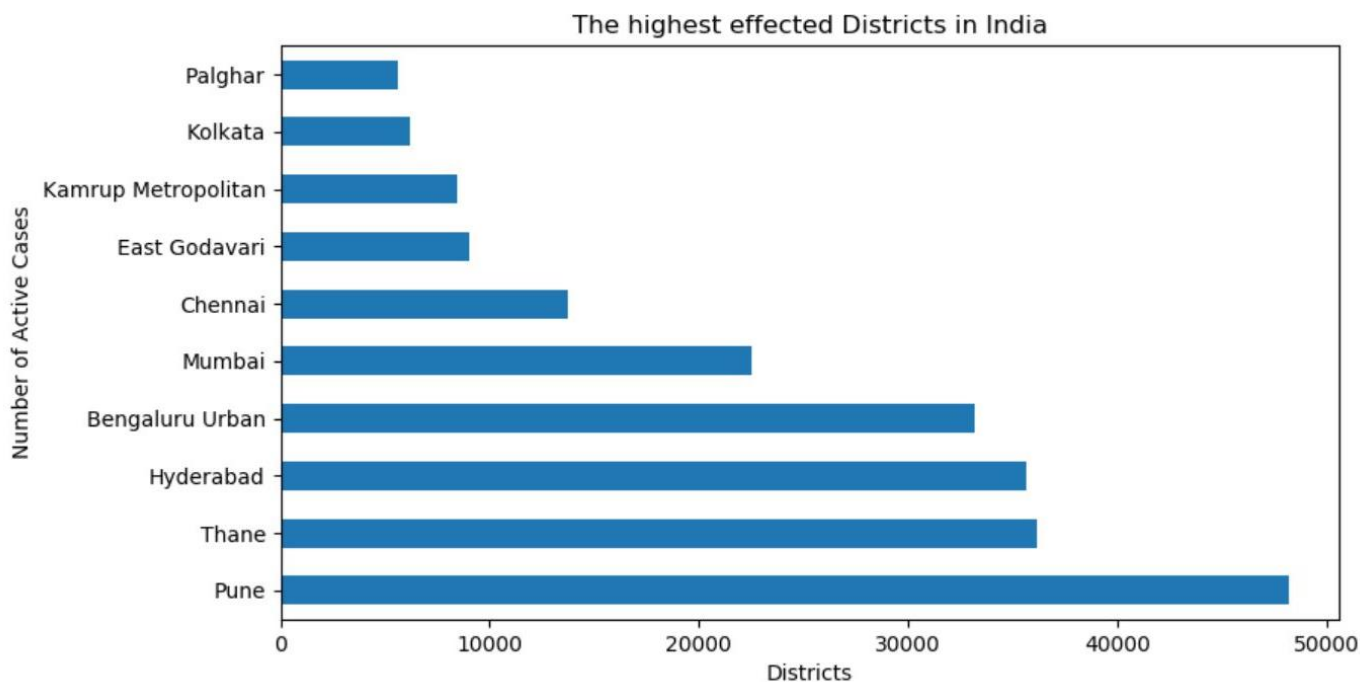

Location address: Hyderabad, Telangana, India
Latitude and Longitude of the said address:
(17.3616079, 78.4746286)

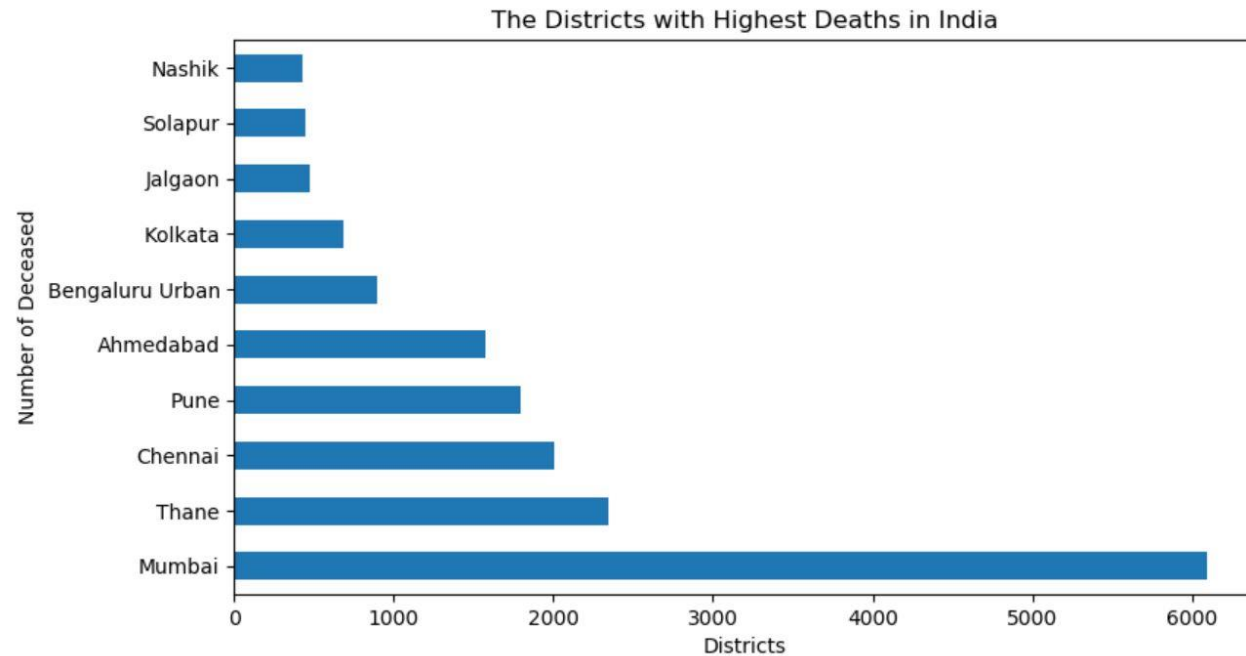The Coordinates of Hyderabad city are, Latitude: 17.3616079 and Longitude: 78.4746286

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

5

Bar Chart Representing Nation Wide Cases

The highest effected Districts in India

Top 5 Districts:
- Pune with over 45,000 Active Cases
- Thane with over 35,000 Active Cases
- Hyderabad with over 30,000 Active Cases
- Bangalore Urban with over 30,000 active cases
- Mumbai with over 25,000 active cases

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

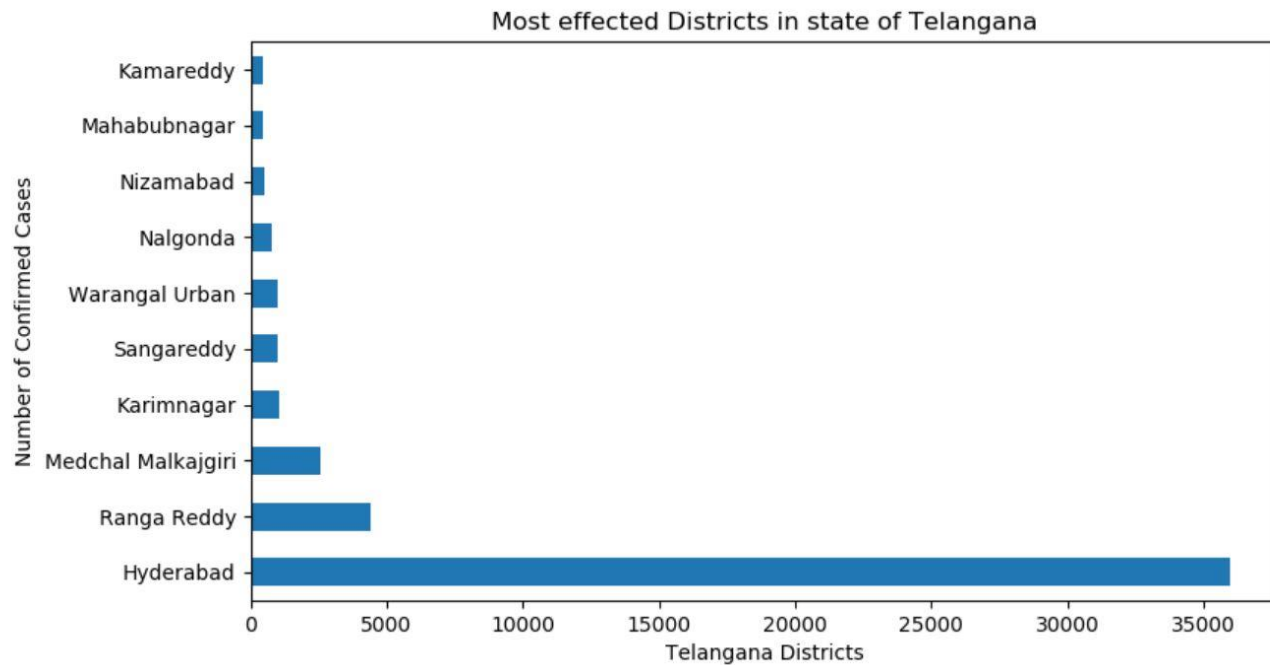6

Highest Mortality Rate

The Districts with Highest Deaths in India

Top 5 Districts:
- Mumbai with over 6,000 Deaths
- Thane with over 2500 Deaths
- Chennai with over 2000 Deaths
- Pune Urban with over 1800 Deaths
- Ahmedabad with over 1500 Deaths

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India
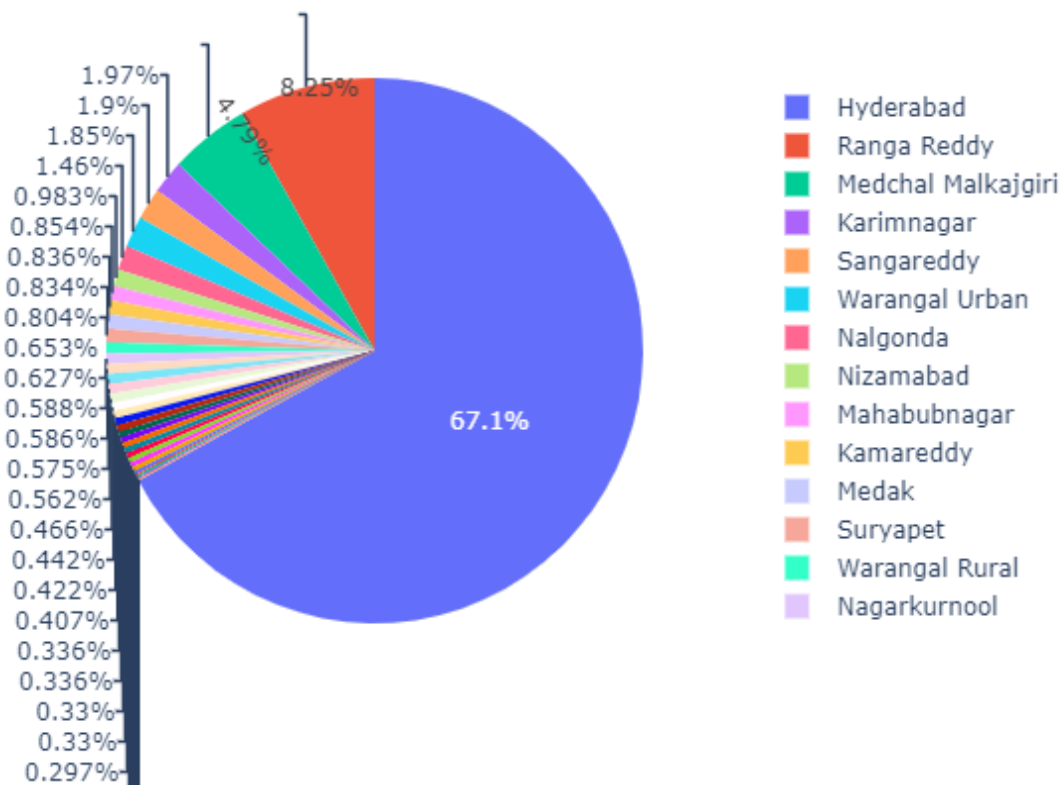
7

District Wise Cases at State Level

Most effected Districts in state of Telangana

Hyderabad in the Lead with at least 7 times higher number of cases then second most hit district, lets look at the distribution in the next slide

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

8

Pie Chart Representing the COVID distribution within state

District Wise Share for Telangana COVID19 Cases

Legend:
- Hyderabad
- Ranga Reddy
- Medchal Malkajgiri
- Karimnagar
- Sangareddy
- Warangal Urban
- Nalgonda
- Nizamabad
- Mahabubnagar
- Kamareddy
- Medak
- Suryapet
- Warangal Rural
- Nagarkurnool

Hyderabad has more than half of the cases (67.1%) precisely which indicates; most of the cases are concentrated within one district.

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

9

	Neighbourhood	Latitude	Longitude	Hospital	Hospital_Latitude	Hospital_Longitude	Hospital_category
0	A. S. Rao Nagar	17.41120	78.50824	Sowmya Hospital	17.408888	78.506439	Hospital
1	A. S. Rao Nagar	17.41120	78.50824	Andhra Mahila Sabha Hospital	17.402475	78.510062	Hospital
2	A. S. Rao Nagar	17.41120	78.50824	Dr. Kirans' Dental Hospital	17.403120	78.507499	Dentist's Office
3	A. S. Rao Nagar	17.41120	78.50824	Abhaya BBC New Born Children's hospital	17.401986	78.509492	Hospital
4	A. S. Rao Nagar	17.41120	78.50824	Sagarlal Memorial Hospital	17.415921	78.498460	Hospital
...	...	...	...	...	...	...	... ...
1412	Yousufguda	17.43835	78.42855	FehmiCare Hospital	17.436987	78.427228	Hospital
1413	Yousufguda	17.43835	78.42855	Susheela hospital	17.434595	78.435801	Hospital
1414	Yousufguda	17.43835	78.42855	Tanvir Hospital	17.429536	78.433841	Hospital
1415	Yousufguda	17.43835	78.42855	Lavanya Group Of Dental Hospitals	17.434936	78.436330	Dentist's Office
1416	Yousufguda	17.43835	78.42855	J J hospitals	17.444909	78.432509	Not Mentioned

```
<matplotlib.axes._subplots.AxesSubplot at 0x253948eb188>
```
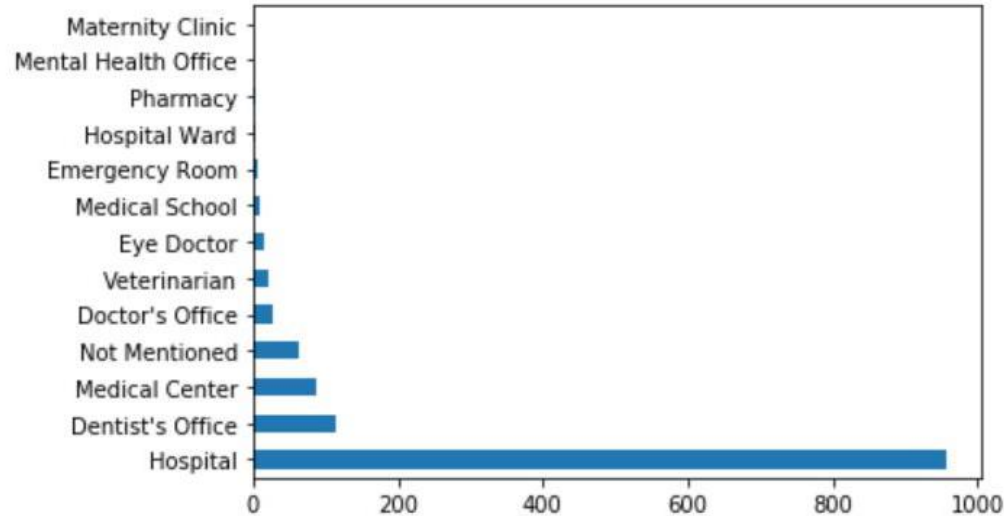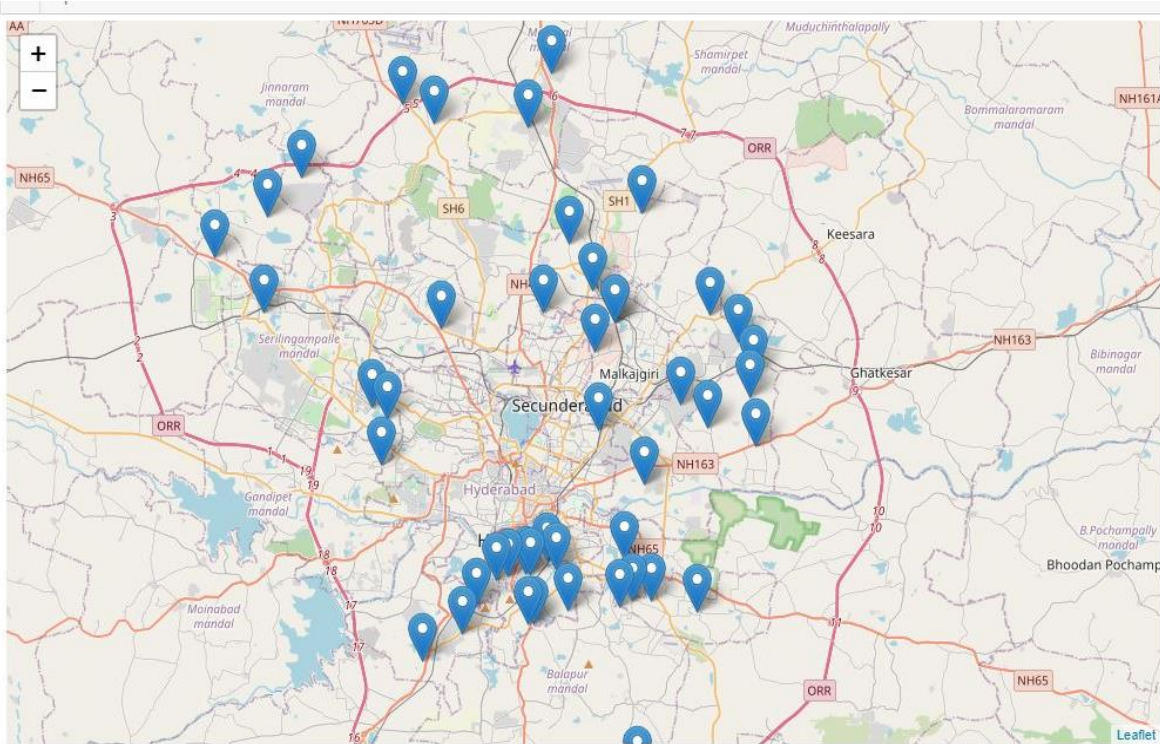

Using Foursquare API and Geocoder, the list of hospitals and the corresponding coordinates have been obtained as listed above

On left, we could see the categorized Medical Facilities available

Rendered Map for localities at Risk

The above map represent neighbourhoods within Hyderabad which have low health care facilities available

List of Localities with no medical facilities within 1 Km radius have been obtained using the code below

```
latitudes=[]
longitudes=[]
count=0
# loop until you get the coordinates
for index,row in diff_df.iterrows():
    lat_lng_coords = None
    neighbourhood=row['Neighbourhood']
    while(lat_lng_coords is None):
        g =
geocoder.arcgis(f'{neighbourhood},
Hyderabad, India')
        lat_lng_coords = g.latlng

latitudes.append(lat_lng_coords[0])

longitudes.append(lat_lng_coords[1])
diff_df['Latitude']=latitudes
diff_df['Longitude']=longitudes
diff_df.head()
```

The Obtained localities and corresponding coordinates are then plotted on the rendered using folium.

11

Latest Stats

Data as for 28/July/2020

	Confirmed	Active	Recovered	Deceased
Maharashtra	3,83,723	1,47,592	2,21,944	13,883
Tamil Nadu	2,20,716	54,896	1,62,249	3571
Delhi	1,31,219	10994	1,16,372	3853
Andhra Pradesh	1,02,349	51,701	49,558	1090
Karnataka	1,01,465	61,818	37,685	1953
Uttar Pradesh	70,493	26,204	42,833	1456
West Bengal	60,830	19,502	39,917	1411
Telangana	57,142	13,753	42,909	480
Gujarat	56,874	13,048	41,482	2344
Bihar	41,111	13,011	27,844	255

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India

12

Discussion, Conclusion and Future Scopes

- With population density as high as India, it becomes pivotal that all neighborhoods have access to health care facilities without having the hassle.

- The Primary objective of the project was to analyze the COVID19 spread in India and to present a visual data of localities in my hometown which requires the attention of authorities to plan and implement measures to avoid the community transmission and facilitate easier access to health care

- The rendered map shows the southern area (Old City) and the eastern area requires immediate attention as large clusters of neighborhood's are at risk.

- After analyzing the different data sets, the project helps in better understanding of hotspots and neighborhood's which are at the relatively higher risk due to lack of facilities, thus, if implemented in real life could save millions living in the clusters.

- The future of the project include further improvement if data be provided, by adding additional parameters such as number of hospital beds, ventilators available to analyze the data efficiently and enabling the authorities to plan and act accordingly

- An ML model can also be developed to predict the capacities beforehand and build the necessary additional facilities required.

Analysis and Visualization of COVID19 in India at District Level and Availability of Health Care at regional level in Hyderabad, India