DETECTING NOVEL CORONA VIRUS COVID-19 FOR MEDICAL EMERGENCIES

1. Data Collection

From the health ministry affairs, it is very important in finding out the severity of covid-19 cases over the districts of TAMIL NADU State. Many urban and rural districts of TAMIL NADU are affected by this disease and the count is highly alarming, it becomes necessary to group affected districts based on severity level which would go a long way in assisting health organizations to give importance for such places where the spreading of the disease is rapid. The following data sources are utilized to extract/generate the required information:

1. Addresses of Tamil Nadu Districts will be obtained using Google Maps API reverse geocoding (geopy.geocoders)

2. Number of neighborhood hospitals and their locations will be obtained using Foursquare API

3. Latitude and Longitude Dataset of Tamil Nadu state will be obtained from the URL: https://www.mapsofindia.com/lat_long/tamilnadu/ This dataset comprises of the following columns:

   Location: District Name(s) of TAMIL NADU

   Latitude: Latitude data of the District (DMS Format*)

   Longitude: Longitude data of the District (DMS Format*)

   *DMS Format: (° : degree, ' : minute, " : second)+Direction(N,E,W,S)

4. COVID-19 pandemic data of Tamil Nadu, India medical cases by district will be obtained from the URL:https://en.m.wikipedia.org/wiki/Template:COVID19_pandemic_data/India/Tamil_Nadu_medical_cases_by_district This dataset comprises of the following columns:

   District: TAMIL NADU District Name

Diagnosed cases: Total No. of Diagnosed cases for COVID-19

Deaths: Total No. of Deaths caused by COVID-19

Recovered cases: Total No. of Recovered cases from COVID-19

Active cases : Total No. of COVID-19 cases still active

Population: Total population of the District

Cases per M: Cases per Million

Last case reported on: Last reported date

2. Dataset creation and Cleaning

The COVID-19 pandemic data is extracted from Wikipedia webpages using Web Scraping with the help of BeautifulSoup library in Python. Taking a closer look at the dataset, the collected data contains the identifier comma (,) to represent thousands in the dataset.

	District	Diagnosed cases	Deaths	Recovered cases	Active cases	Population	Cases per M	Last case reported on
0	District	Diagnosed cases[a]	Deaths	Recovered cases	Active cases[b]	Population	Cases per M	Last case reported on
1	Tamil Nadu	102,721	1,385	58,378	42,955	75,695,000	1,357	3 July 2020
2	Ariyalur	463	0	425	38	752,481	615	1 July 2020
3	Chengalpattu	6,139	106	3,113	2,920	2,556,423	2,401	3 July 2020
4	Chennai	64,689	996	40,111	23,582	7,088,000	9,127	3 July 2020

The cleaning of the collected data from the web table is performed to create a uniform and usable dataset before they could be subjected to further analysis.

	Diagnosed cases	Deaths	Recovered cases	Active cases	Population	Cases per M
0	6139.0	106.0	3113.0	2920.0	2556423.0	2401.0
1	64689.0	996.0	40111.0	23582.0	7088000.0	9127.0
2	645.0	1.0	252.0	392.0	3172578.0	203.0
3	1143.0	5.0	813.0	325.0	2600880.0	439.0
4	107.0	0.0	41.0	66.0	1502900.0	71.0

Data Transformation

The Latitude and Longitude data had been extracted from the mapsofindia web page. The spherical coordinates (latitudes and longitudes) are in degreesminutes-seconds (DMS) format. DMS data is recorded in the format: (° : degree, ' : minute, " : second) with direction (N,E,W,S). The range of minutes and seconds values is from 0 to 60.

	Active	Location	Latitude	Longitude
0	1	Ariyalur	11° 8' N	79° 4' E
1	1	Chengalpattu	12° 42' N	80° 01' E
2	1	Chennai	13° 04' N	80° 17' E
3	1	Coimbatore	11° 00' N	77° 00' E
4	1	Cuddalore	11° 43' N	79° 49' E

The geographic coordinates can be converted to decimal degrees. It's just another way to represent the same location in a different format which can be used to plot markers over the map.

	Active	Location	Latitude	Longitude	DD_Lat	DD_Lng
0	1	Ariyalur	11° 8' N	79° 4' E	11.133333	79.066667
1	1	Chengalpattu	12° 42' N	80° 01' E	12.700000	80.016667
2	1	Chennai	13° 04' N	80° 17' E	13.066667	80.283333
3	1	Coimbatore	11° 00' N	77° 00' E	11.000000	77.000000
4	1	Cuddalore	11° 43' N	79° 49' E	11.716667	79.816667

3. Neighborhoods data returned by FourSquare API:

A small sample of generated neighborhood hospitals along with geographical coordinate addresses returned by FourSquare API under hospital category has been shown in the table below.

	name	categories	lat	lng
0	Apollo Hospitals	Hospital	13.062761	80.251790
1	Miot Hospital	Hospital	13.021841	80.185831
2	Hande Hospital	Hospital	13.076504	80.228414
3	Apollo Speciality Hospital	Hospital	13.033841	80.245289
4	Fortis Malar Hospital	Hospital	13.010271	80.258978
5	Balaji Dental & Craniofacial Hospital	Dentist's Office	13.035352	80.247048
6	Vijaya Group Hospital	Hospital	13.049794	80.208666
7	Apollo First Med Hospitals	Hospital	13.077976	80.246140
8	Rajan Eye Care Hospital	Hospital	13.050975	80.242846
9	Apollo hospitals	Hospital	13.062767	80.251862

4. Data Integration

The fields of the above two datasets containing district-wise covid-19 cases and district-wise geographical coordinates are merged based on district column to visualize and gain further insights from the obtained data.

	Active	Location	Latitude	Longitude	DD_Lat	DD_Lng	District	Diagnosed cases	Deaths	Recovered cases	Active cases	Population	Cases per M	Last case reported on
0	1	Chengalpattu	12° 42' N	80° 01' E	12.700000	80.016667	Chengalpattu	6,139	106	3,113	2,920	2,556,423	2,401	3 July 2020
1	1	Chennai	13° 04' N	80° 17' E	13.066667	80.283333	Chennai	64,689	996	40,111	23,582	7,088,000	9,127	3 July 2020
2	1	Coimbatore	11° 00' N	77° 00' E	11.000000	77.000000	Coimbatore	645	1	252	392	3,172,578	203	3 July 2020
3	1	Cuddalore	11° 43' N	79° 49' E	11.716667	79.816667	Cuddalore	1,143	5	813	325	2,600,880	439	3 July 2020
4	1	Dharmapuri	12° 08' N	78° 13' E	12.133333	78.216667	Dharmapuri	107	0	41	66	1,502,900	71	3 July 2020
5	1	Dindigul	10° 22' N	78° 00' E	10.366667	78.000000	Dindigul	618	7	310	301	2,161,367	286	3 July 2020