**DETECTING NEARBY HOSPITAL VENUES FOR MEDICAL EMERGENCIES**

**DATA DESCRIPTION**

**1. Data Collection**

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 TN

**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)

1. 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:** TN 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

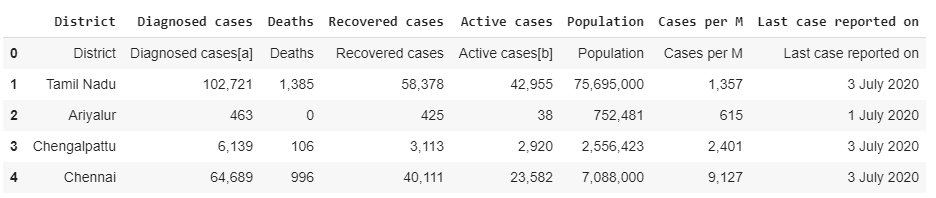
**Population:** Total population of the District

**Cases per M:** Cases per Million

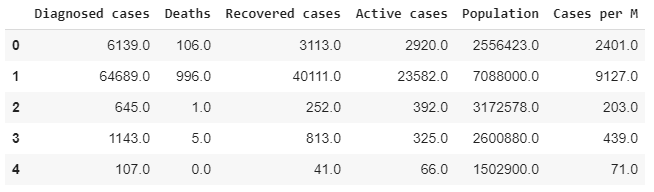
**Last case reported on:** Last reported date

**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.

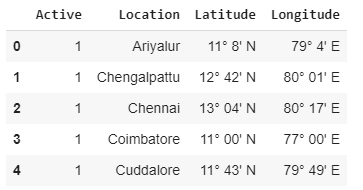


The cleaning of the collected data from the web table is performed to create a uniform and usable dataset before they could be subjected to further analysis.

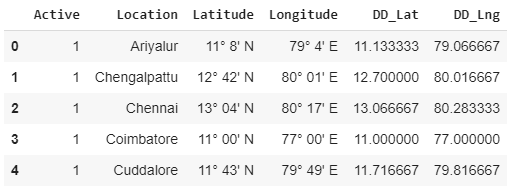


**Data Transformation**

The Latitude and Longitude data had been extracted from the mapsofindia web page. The spherical coordinates (latitudes and longitudes) are in degrees-minutes-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.



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.



**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.



**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.

