

Identify the location to increase COVID-19 Healthcare facilities

Tamil Nadu, India

May 29, 2021

1. Introduction

1.1 Background

The first case of the COVID-19 pandemic in the Indian state of Tamil Nadu was reported on 7 March 2020.

The largest single-day spike (30,987 cases) was reported on 13 May 2021 and Tamil Nadu now has the fourth-highest number of confirmed cases in India after Maharashtra, Kerala, and Karnataka. All 37 districts of the state are affected by the pandemic, with capital district Chennai being the worst affected.

1.2 Problem

Hospital facilities in Tamil Nadu are adequate, however, it is not distributed as per the density of the population in all the districts of Tamil Nadu. People who require medical care have to travel 100 to 200 km as there are no beds available nearby by healthcare facilities.

1.3 Interest

State Government of Tamil Nadu, taking intensive measures to curb the Covid-19 infections and providing temporary field hospitals for Covid-19 patients across the state. This project will help the State Government to focus on the districts with higher populations but limited health care facilities.

2. Data acquisition and cleaning

2.1 Data sources

- *Wikipedia*

As Wikipedia is a reliable source of information, following URL is used to extract list of districts in Tamil Nadu with population.

https://en.wikipedia.org/wiki/List_of_districts_of_Tamil_Nadu

- **Geopy**

Geopy is a Python client for several popular geocoding web services. This library makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

- **Foursquare**

Foursquare is the most trusted, independent location data platform for understanding location, explore venues and neighborhoods

2.2 Data cleaning

Data downloaded from Wikipedia is stored in the Pandas data frame and the following cleansing exercises were performed.

No.	District	Code	Capital	Uate of official formation	Split from	Area (km2)	Population[34]	Populationdensity(/km2)	Taluks[35]	Map	
0	1.0	Ariyalur	AR	Ariyalur	23 November 2007	Perambalur	1949.31	754894	390	Andimadam Ariyalur Udayarpalayam Sendurai	NaN
1	2.0	Chengalpattu	CGL	Chengalpattu	29 November 2019	Kanchipuram	2944.96	2556244	868	Chengalpattu Cheyyur Madurantakam Pallavaram Tambaram Thiruporur Thirukazhukundram Vandalur	NaN
2	3.0	Chennai	CH	Chennai	1 November 1956	One of the original 13 districts (under former name of "Madras District")	426	4646732	26076	Alandur Ambattur Aminjikarai Ayanavaram Egmore Guindy Madhavaram Maduravoyal Mambalam Mylapore Perambur Purasawalkam Sholinganallur Tiruvottiyur Tondiarpet Velachery	NaN
3	4.0	Coimbatore	CO	Coimbatore	1 November 1956	One of the original 13 districts	4,723[36]	3458045	732	Anaimalai Annur Coimbatore-North Coimbatore-South Kinathukadavu Madukkarai Mettupalayam Perur Pollachi Suler Valparai	NaN

Rename the following column header to make it more appropriate for the content

- 'Area(km2)':'Area',
- 'Population[34]':'Population',
- 'Populationdensity(/km2)':'Populationdensity_per_km2',
- 'Taluks[35]':'Taluks'

'Area' column contains some square brackets '[' and it was removed the following python code.

```
for ind in df.index:
    df['Area'][ind] = df['Area'][ind].replace('[', '')
    if df['Area'][ind].find("[") != -1:
        print(df['Area'][ind])
        string = df['Area'][ind]
        string = string[0: (df['Area'][ind].find("["))]
        print( string)
        df['Area'][ind] = string
```

After removing the square brackets, the data type of column 'Area' is changed from 'Object' to 'Float'. Moreover, we also remove the column 'Map' as it doesn't contain any values.

2.3 Adding Latitude and Longitude for districts

By using 'geopy.geocoders' libraries, every district of the data frame is appended with longitude and latitude using the following code.

```
def findGeocode(district):  
  
    # try and catch is used to overcome  
    # the exception thrown by geolocator  
    # using geocodertimedout  
    try:  
  
        # Specify the user_agent as your  
        # app name it should not be none  
        geolocator = Nominatim(user_agent="foursquare_agent")  
  
        return geolocator.geocode(district + " Tamil Nadu, India")  
  
    except GeocoderTimedOut:  
  
        return findGeocode(district)  
  
#geolocator = Nominatim(user_agent="foursquare_agent")  
#location = geolocator.geocode(address)  
#latitude = location.latitude  
#longitude = location.longitude  
#print(latitude, longitude)
```

Four square agents is used to extract the co-ordinates and it is merged to the existing table.

2.4 Explore Health Care facilities in each district.

As we are interesting only for health care facilities, following attributes is added to the API to limit the output only for hospitals

Category Id=4bf58dd8d48988d196941735

```
# create the API request URL  
url = 'https://api.foursquare.com/v2/venues/explore?categoryId=4bf58dd8d48988d196941735&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}'.format(  
    CLIENT_ID,  
    CLIENT_SECRET,  
    VERSION,  
    lat, |  
    lng,  
    radius,  
    LIMIT)
```

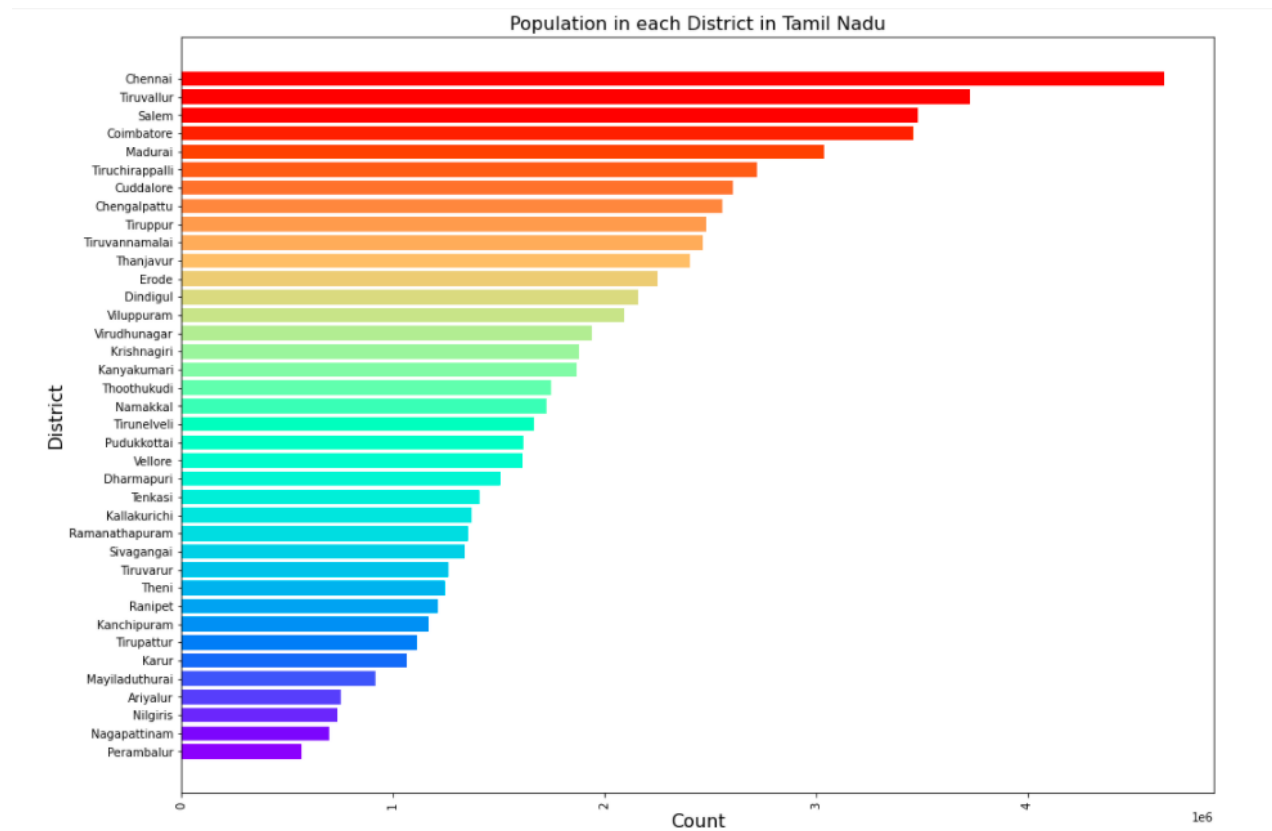
3. Exploratory Data Analysis

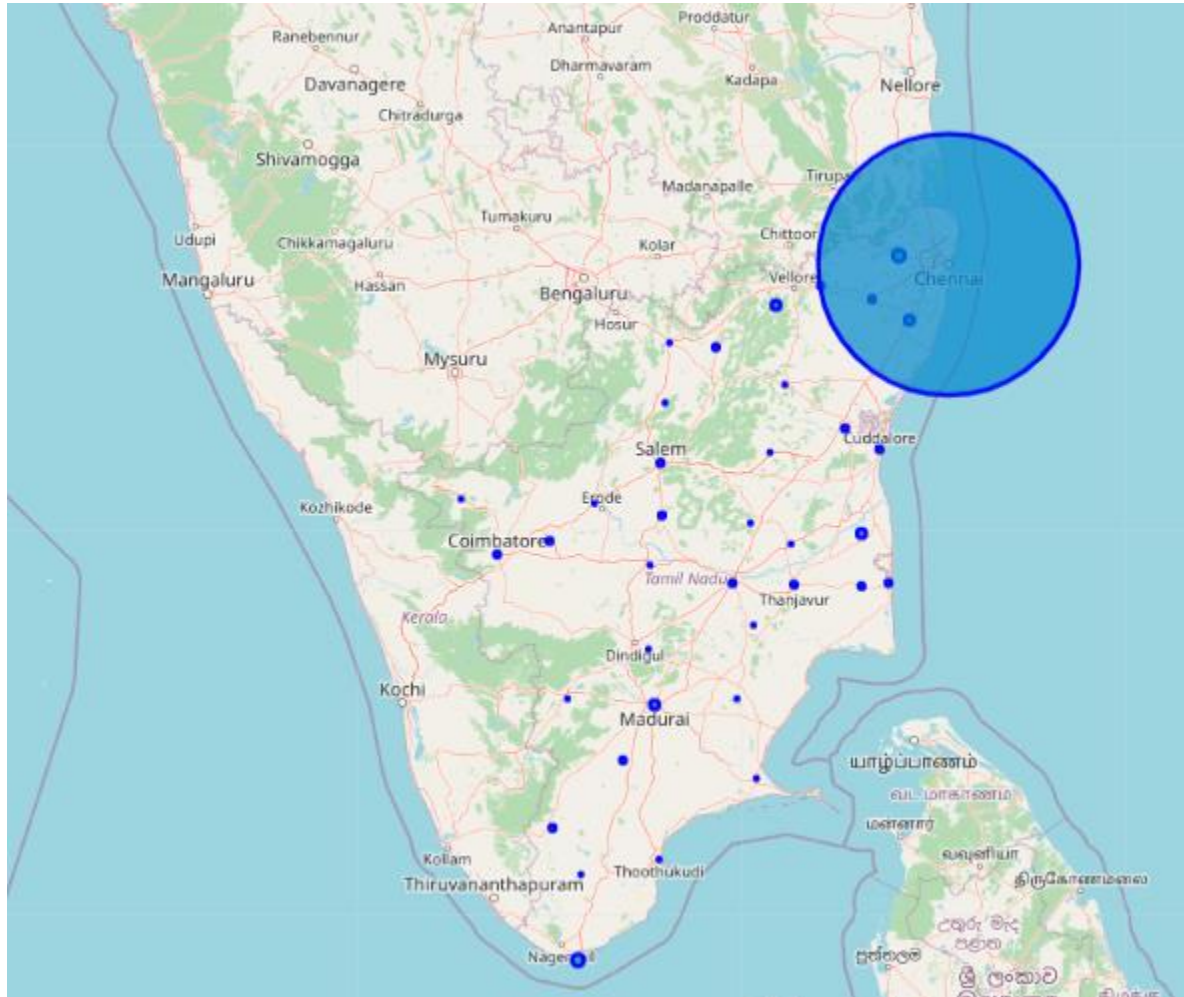
3.1 Analyze the population by Districts

With the help of horizontal bar charts, the population in each district is analyzed to determine the high and low population districts.

```
df.sort_values('Population',inplace=True)
population = df['Population']
district = df['District']

colors = cm.rainbow(np.linspace(0, 1, len(population.index)))
plt.figure(figsize = (16, 12))
plt.xticks(rotation = 90)
plt.xlabel("Count", fontsize = 16)
plt.ylabel("District", fontsize = 16)
plt.title("Population in each District in Tamil Nadu", fontsize = 16)
plt.barh(district, population.values, color = colors)
```



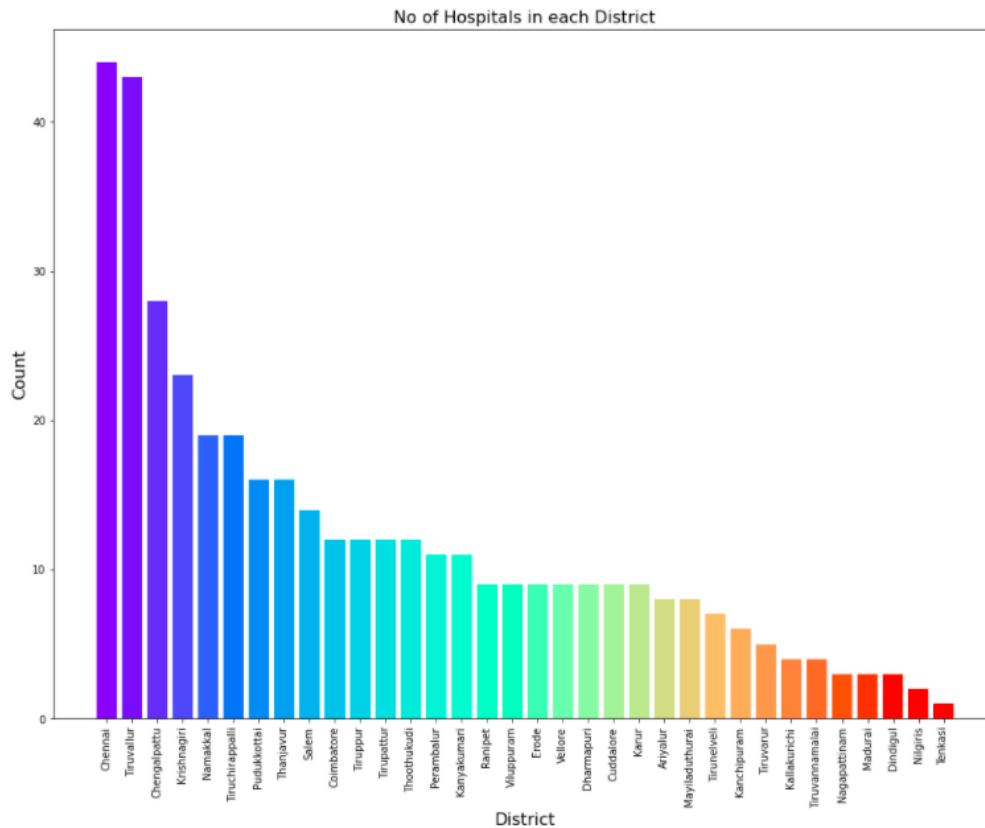


Based on the above analysis, we determined that the highest population district in Chennai and the lowest population district is Permabalar.

Like population, we also analyzed the hospitals in each district. With the help of Foursquare venue explore API, we extracted all the hospitals in each district as shown below.

```
no_hospitals_in_district.sort_values('No of Hospitals', ascending=False, inplace=True)
venues = no_hospitals_in_district['No of Hospitals']
district = no_hospitals_in_district['District']

colors = cm.rainbow(np.linspace(0, 1, len(venues.index)))
plt.figure(figsize = (16, 12))
plt.xticks(rotation = 90)
plt.xlabel("District", fontsize = 16)
plt.ylabel("Count", fontsize = 16)
plt.title("No of Hospitals in each District", fontsize = 16)
plt.bar(district, venues.values, color = colors)
```



3.2 Relationship between district population and the total number of hospitals

Data sets of population and number of hospitals by district were merged to determine the people per hospital for each district.

By using a folium map, people per hospital are analyzed and plotted.

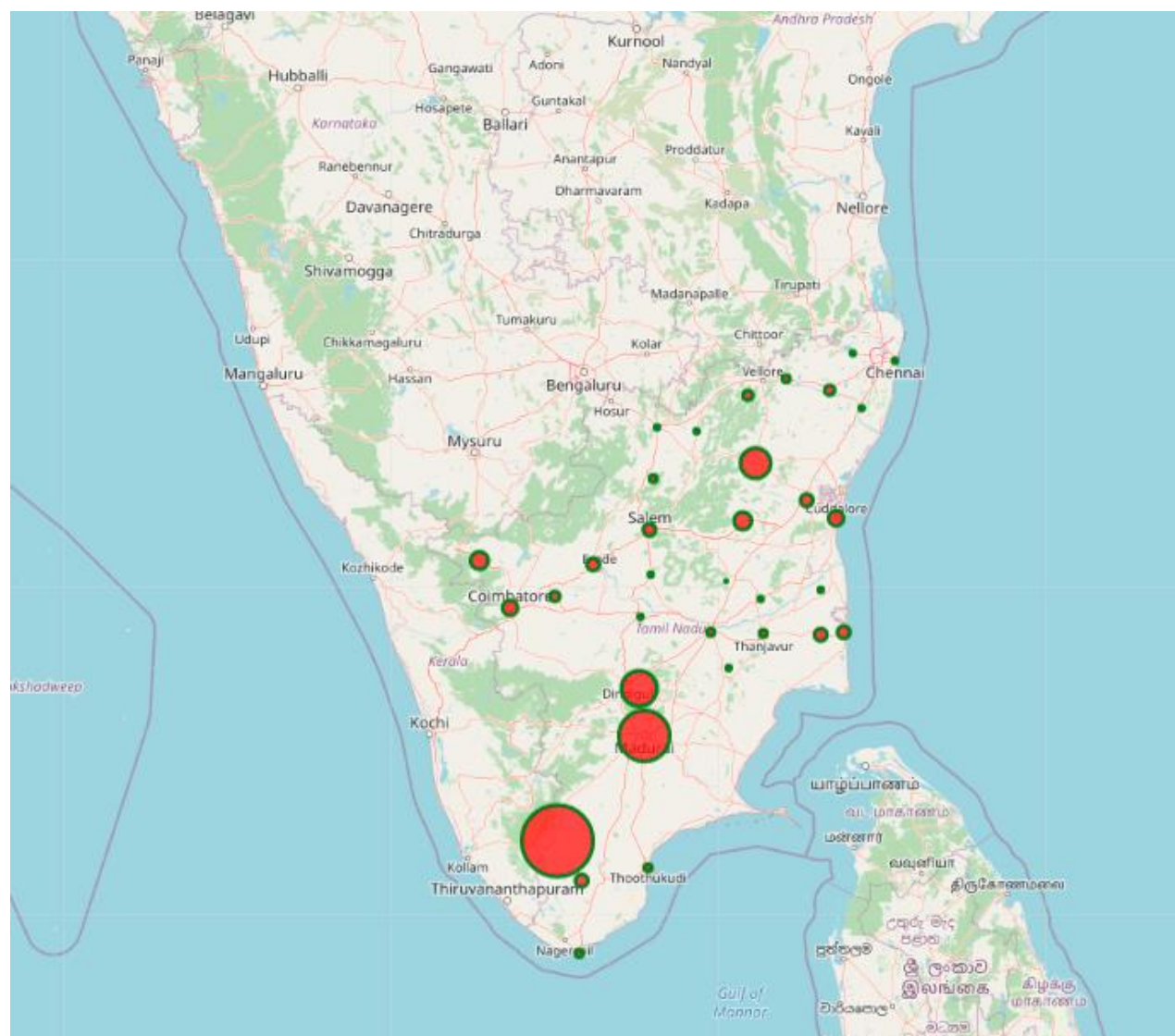
```
#df_hospitals.sort_values('Density Per Hospital',inplace=True)

#colors = cm.rainbow(np.linspace(0, 1, len(density_population.index)))

#map of State using Latitude and Longitude values
map_state = folium.Map(location=[latitude, longitude], zoom_start=7)

# add markers to map
for lat, lng, district, density in zip(df_hospitals['Latitude'], df_hospitals['Longitude'], df_hospitals['District'], df_hospitals['Density Per Hospital']):
    label = '{}', '{}'.format(district, density)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=density/50000,
        popup=label,
        color='green',
        fill=True,
        fill_color = '#FF0000',
        fill_opacity=0.7,
        parse_html=False).add_to(map_state)

map_state
```



4. Conclusions

In this study, we analyzed the population and hospitals in each district to determine which districts require more attention for health care facilities.

The analysis identified that the following districts do not have any medical facilities and they must travel nearby districts for hospitals. These districts require immediate attention by the State Government of Tamil Nadu to open temporary field hospitals.

District	Code	Capital	Date of official formation	Split from	Area	Population	Populationdensity_per_km2
Theni	TH	Theni	25 July 1996	Madurai	3242.30	1245899	406
Sivagangai	SI	Sivagangai	15 March 1985	Ramanathapuram	4189.00	1339101	328
Ramanathapuram	RA	Ramanathapuram	1 November 1956	One of the original 13 districts	4068.31	1353445	331
Virudhunagar	VR	Virudhunagar	15 March 1985	Ramanathapuram	4241.00	1942288	453

Also, we noted, following districts require immediate attention as the people per hospital in these districts are more than 300k to 1 million.

District	Density Per Hospital
Kallakurichi	342570.0
Nilgiris	367697.0
Tiruvannamalai	616219.0
Dindigul	719925.0
Madurai	1012751.0
Tenkasi	1407627.0