**The Battle of Neighbourhoods**

# Introduction

## Background

Bengaluru is a capital city of Karnataka. The centre of India's high-tech industry, the city is also known for its parks and nightlife. As a resident of city I decided to use Bengaluru in my project. In recent decades, the city has witnessed rapid growth in population and urbanized area. Many businessmen arrives here with a dream of investment.

## Problem

When we think as an investor, we expect from them to prefer locality where the type of business they want to install is less intense. However, it is difficult for investors to get these information easily.

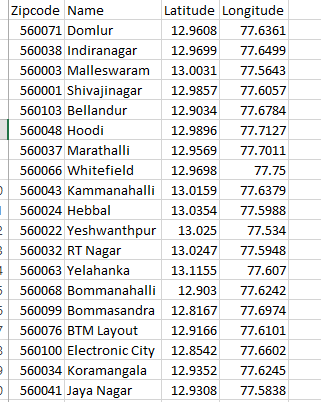
So, we can create a map and information where each neighbourhood of the city is clustered according to the venue density.

# Data

To solve the above problem, that is to create a map and information on all the neighbourhoods of Bengaluru city, we need data of the city.

There are many sources on internet from where we can get list of neighbourhoods of the Bengaluru City, e.g. Wikipedia.

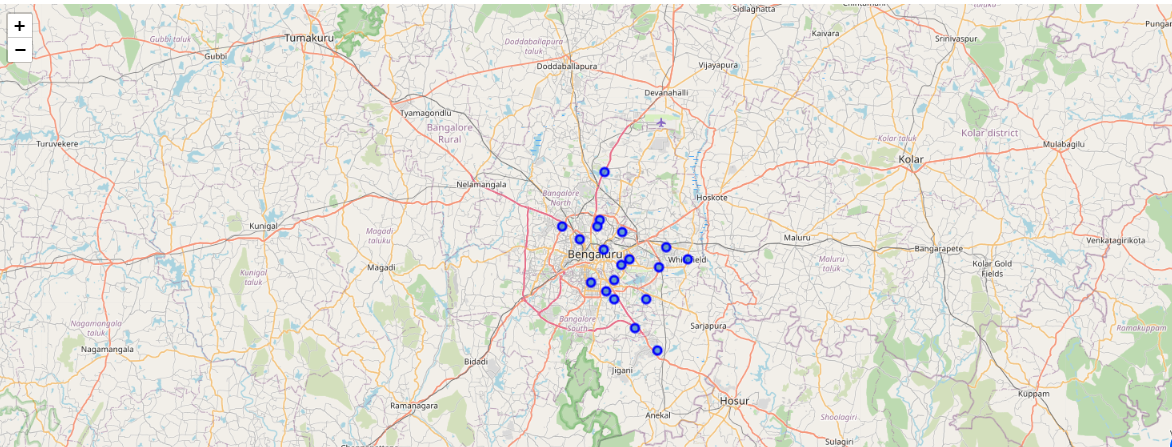
I have organized a data set in csv file with a list of all neighbourhoods along with their latitude and longitude.



# Methodology

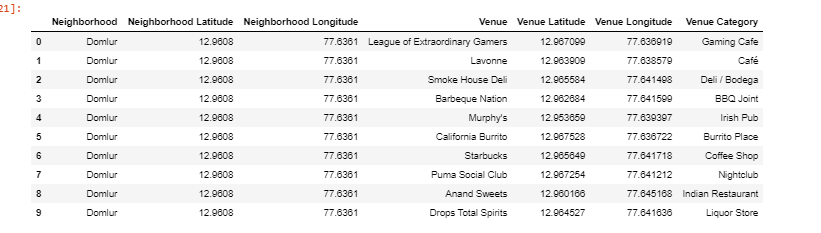
The data which I have organized is clean and can be used as it for the project.

I used the **folium** library to visualize the geographic details of the Bengaluru city and its neighbourhoods. I created a map of the city with neighbourhoods superimposed on top. To create the map I used **geopy** library to get the latitude and longitude values of the city.



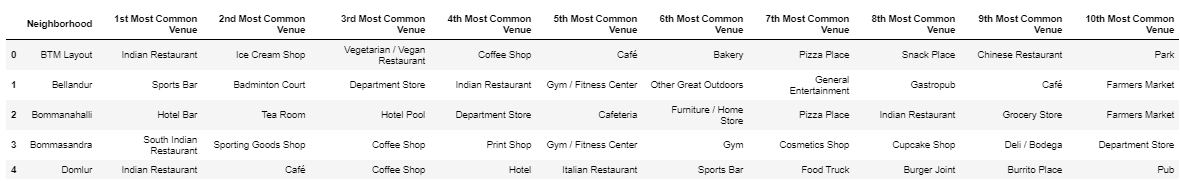
I utilized the **Foursquare API** to explore the venues of each neighbourhoods of the city.

I designed the limit of 100 and radius of 1000mtrs for each neighbourhoods from their given latitude and longitude present in my data set. The result gave me list of all the venues,along with their respective geographical locations, for all the neighbourhoods. As the result was a json file, so I cleaned the file and put the data in a dataframe. Below is the head of the dataframe.



From the above data set I retrieved the unique categories of the venue. There were total of 127 unique categories.

I created a table which shows list of top 10 venue categories for each neighbourhood according to the frequency of each venue category.



We observed from the above table, that there are some common venue categories in neighbourhoods. So, I decided to segment and cluster the neighbourhood.

For, this reason I used unsupervised learning K-means algorithm of ML to cluster the neighbourhoods. K- Means algorithm is the most common cluster algorithm and fits apt for my purpose. I created 3 clusters.



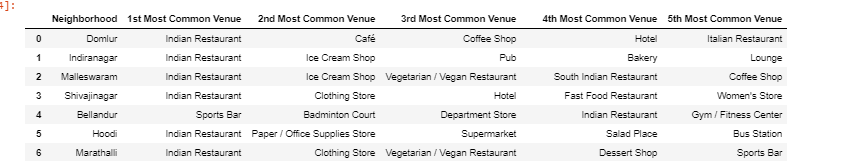
Now, I will label each cluster as below:

### Cluster 1: "Indian Restaurant" Cluster 2: "Sports/Gym Center" Cluster 3: "Market

# Results

Let’s create 3 tables for each cluster with 5 most common venue categories.

### Cluster 1: Indian Restaurant



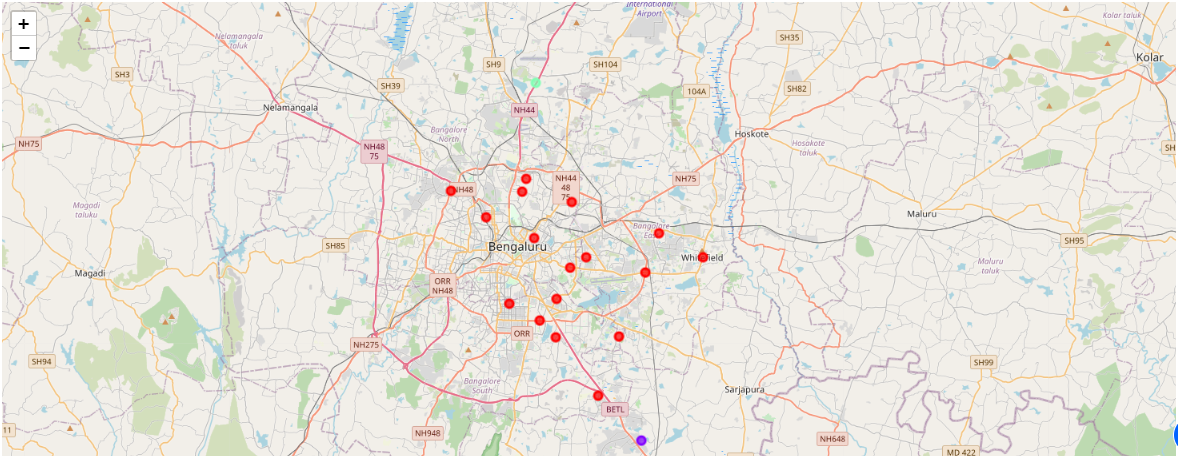
### Cluster 2: Sports/Gym Center



### Cluster 3: Market



Finally, we can see the map with the clusters



Looking at the clusters a person can make decision for investing on type or place for business.

For example if a person wants to invest on openening a restaurant then he can look on cluster 3 where there are very few restaurants.

# Discussion

As I mentioned before, Bengaluru is a major city of India and densely populated. There are far more neighbourhoods which can be included in the dataset for the data analysis which will give more detailed information for every business type. Using Foursquare API, we can get more information about every venue and we can find out which venue is trending in which area. I used K-means clustering for clustering the neighbourhoods which gave more clear idea for business owners.

At the end, I created a map using folium library which gives a better visualization of the data analysis performed.

# Conclusion

As more people are turning to big cities like Bengaluru for their business set up, they can achieve better outcomes using the above analysis.