Introduction

**Background:** I live in Bengaluru, India, one of the cities famous for Software companies.

It is tough to get a right for you at the right budget when you are trying to relocate to a different part of the city.

**Problem:** Its always too much hassle if we plan to buy house especially in the right location and within our budget. And then we must look at different facilities that is available around you. We normally are comfortable being with specific set of features like restaurant, park, groceries stores., etc., around us. This project aims to consider all this to find a place in the city that is close to our preferences.

**Interest:** This project interest people who would like to find out neighborhood that are close to their existing place, when they decide move out /plan to buy a house in new locality.

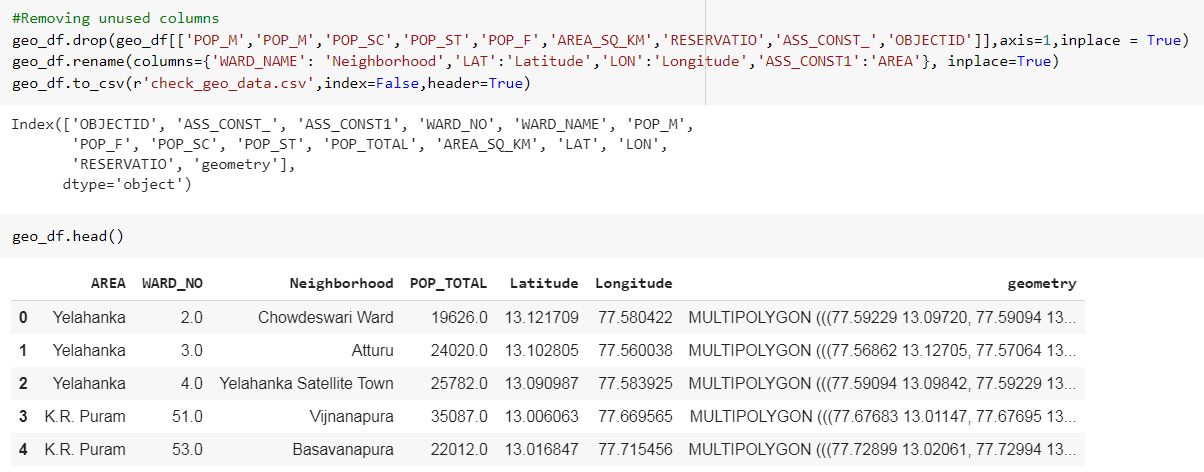
Data

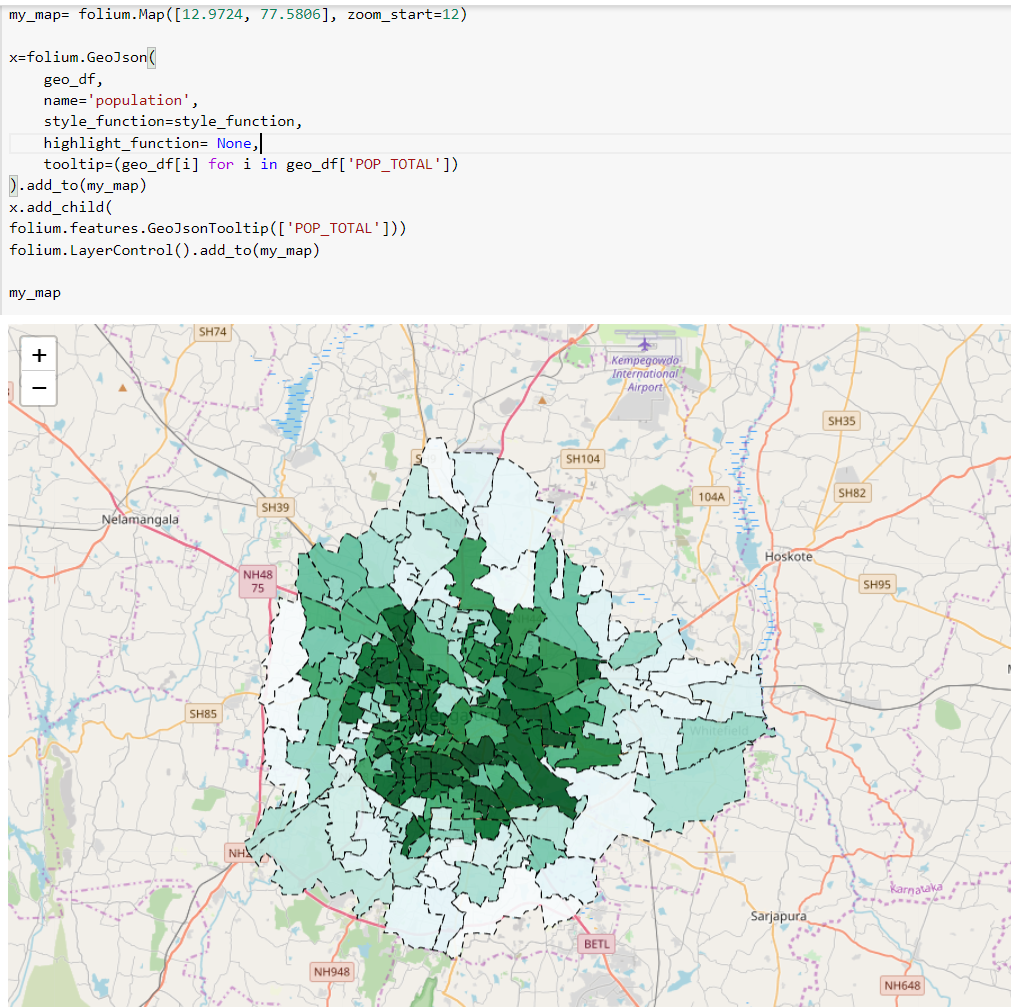
There two sets of data that is used in this project,

1. BBMP Ward information : (<https://github.com/datameet/Municipal_Spatial_Data/tree/master/Bangalore>)
   1. Bengaluru city is managed by a body called BBMP. They have categorized the city in terms of wards (locations) politically. The data is published in their website.
   2. This data contains, different wards and their names along with the population in each of these wards.
2. Foursquare api for Bengaluru (latitude: 12.9791198, longitude: 77.5912997.). This has provided a limited set of information. This is not used much in Bengaluru, India

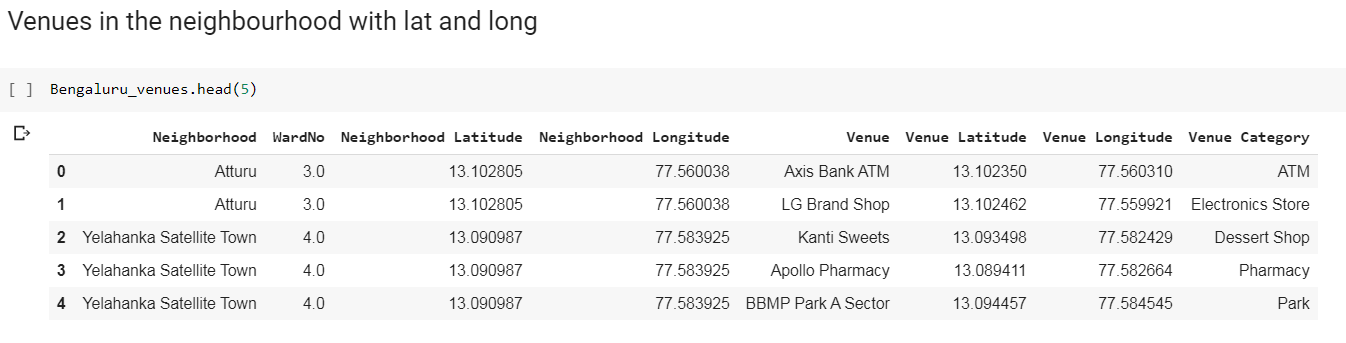
Methodology

We start by exploring the data that is used in this project. The data contains many fields that are not relevant for this project. So, we have cleaned such fields. Also, some of the columns are renamed to suit our project as shown below,

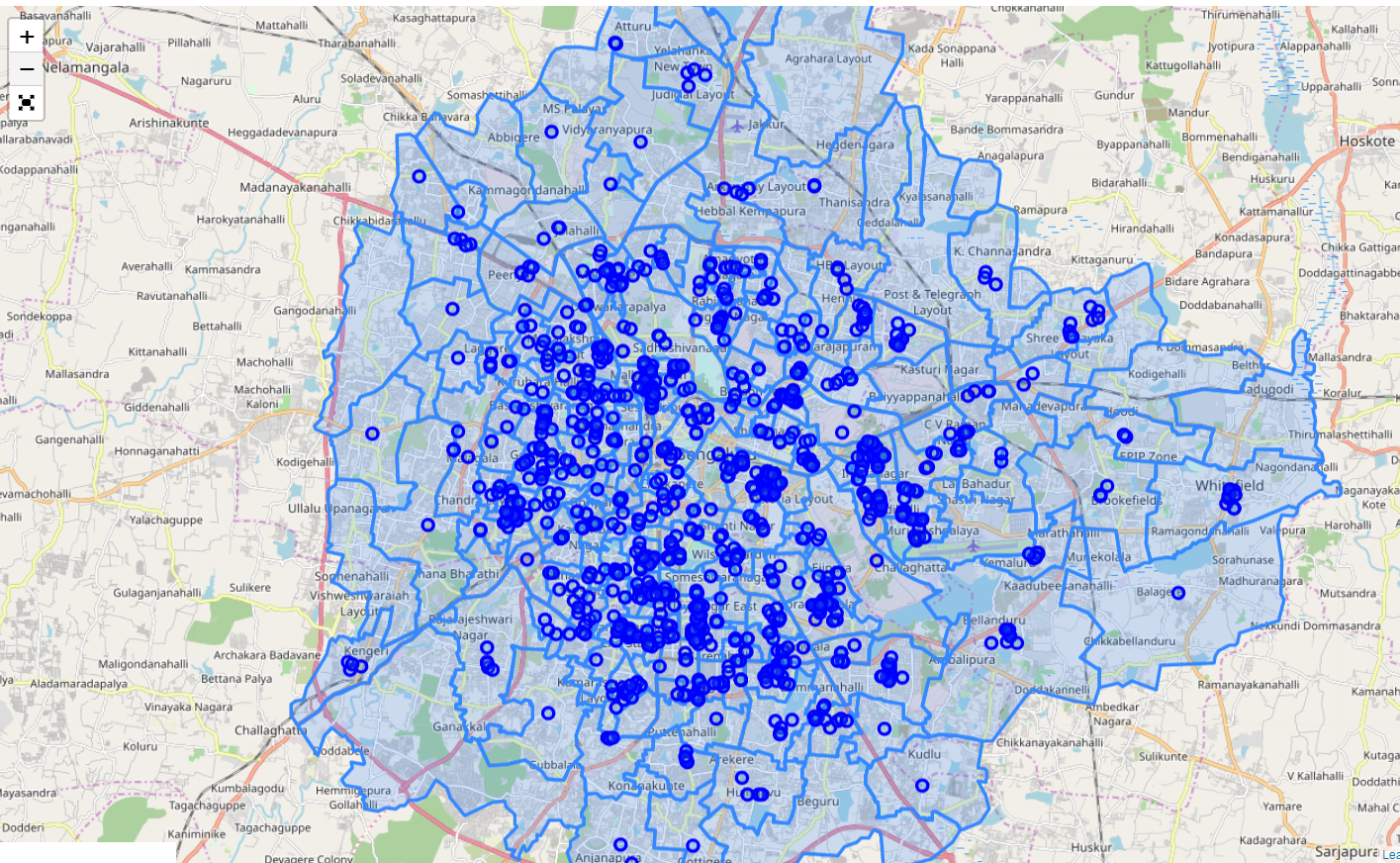


We can visualize the Bengaluru city in a folium map, based on the population,

We then use the Foursquare api with our credentials for all the neighborhood that we got from the dataset. Foursquare fetches all the venues along with the latitude and longitude for those venues



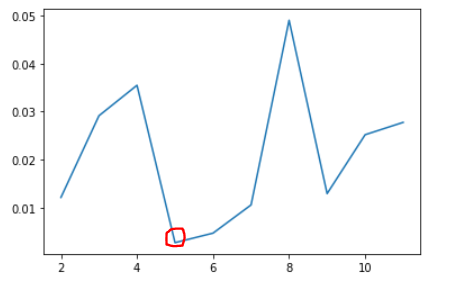
All these venues are plotted on a map and let looks like this



**Neighborhood clusters**

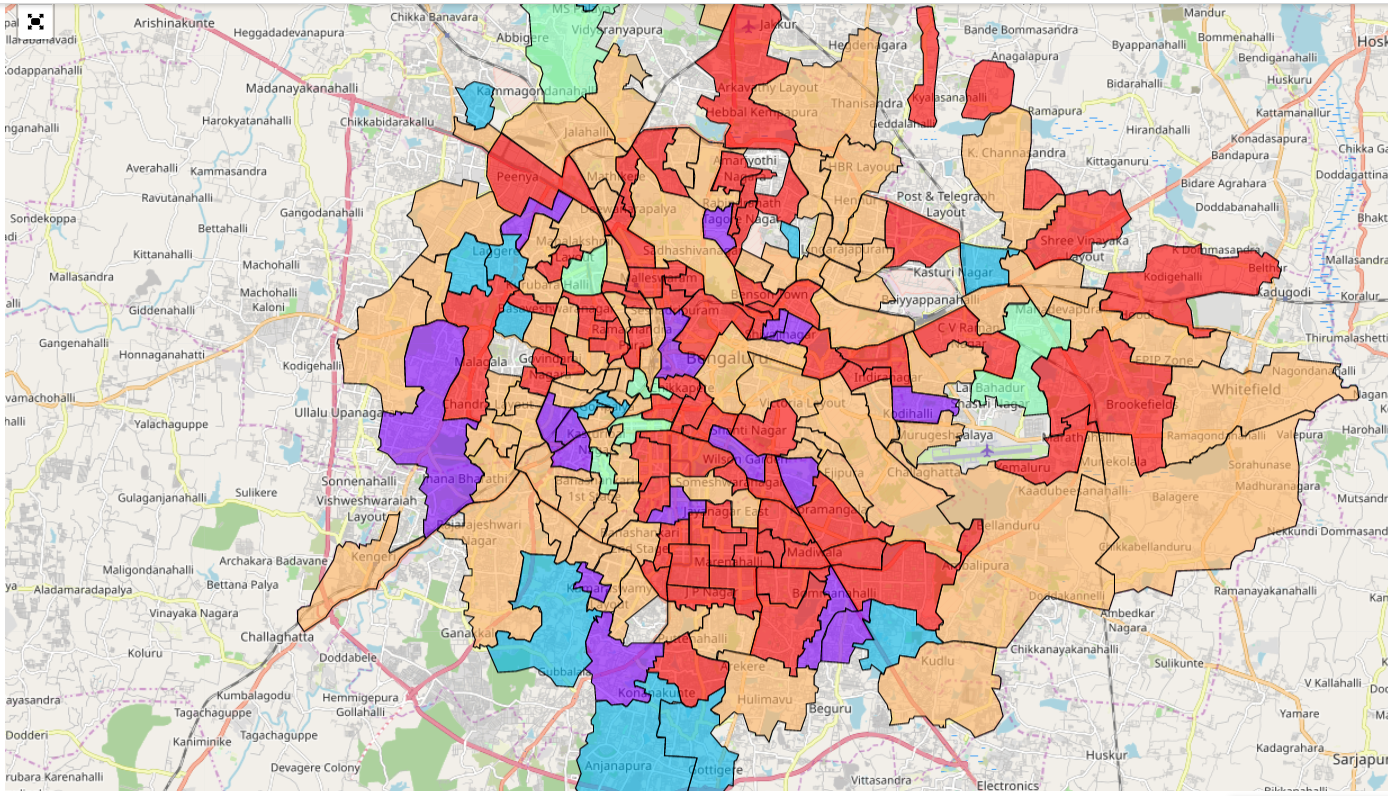
We form clusters of the venue based on the Foursquare and the geo Json data from the BBMP dataset. We will be using a K-means clustering algorithm.

To find an optimal value for k, we run the clustering algorithm for various values and use the silhouette score. There was no elbow I could make but preferred to go with the first elbow at k=5.



**Plotting the clusters**

After some cleaning of neighborhood with no venue details from foursquare data and plotting the resulting clusters from k-mean algorithm in folium. The map looks like this,



**Decision Tree:**

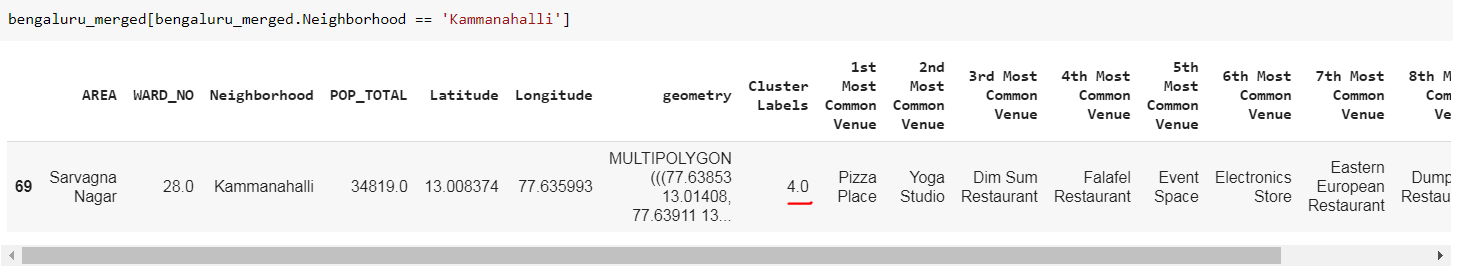
We build a decision tree in order to figure in which type of neighborhood I should select.

To do this we train a decision tree using certain venues as training data in order to predict the cluster labels, following result:

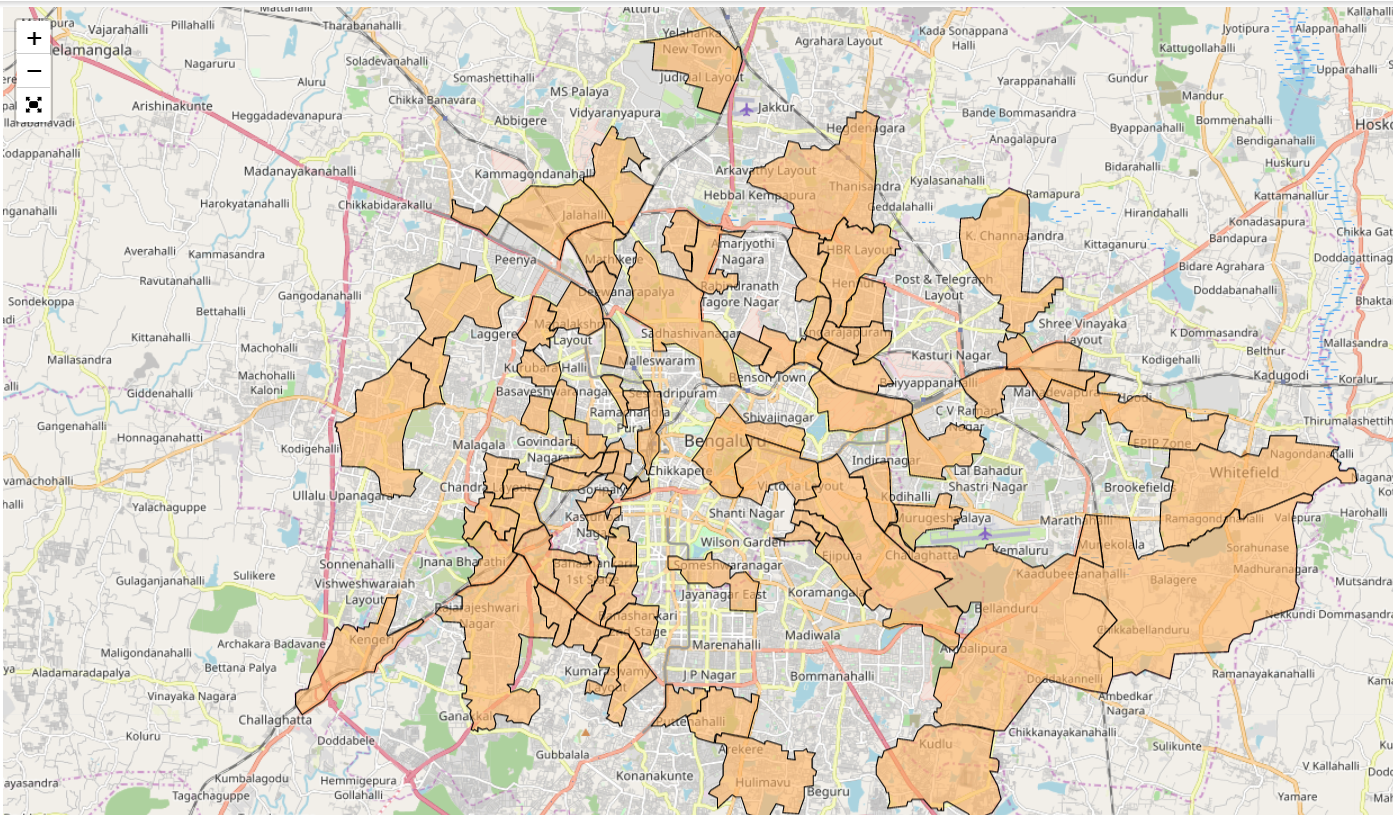


Result

I could figure out that my place belongs to cluster 4 and all we must do is to find the best places from this cluster that is suitable.



When we plot the cluster 4 in a map, it looks like this,



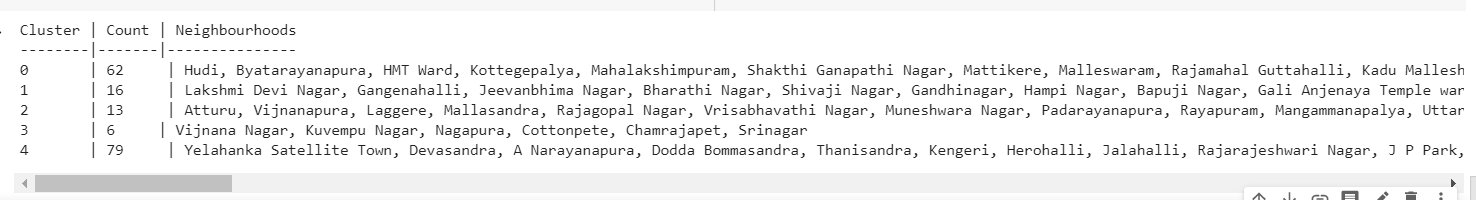
Then I try to get the top 5 places that is like my place from the cluster,



Discussion

Few observations, the Foursquare doesn’t cover too many areas in our city.

Some of the clusters have shown neighborhood that are similar while other clusters have less neighborhoods. This could be attributed to k value of 5 or dataset that is used for the analysis.



Folium is an interesting tool to visualize. The geoJson file have geometry data that helped to plot the area.

Conclusion

So, based on our analysis using k-mean clustering the following places were among top5 that are like my place Kammanahalli, Kodigehalli is my choice.