Optimal Location and Rating for a new Restaurant in Bangalore

1. Introduction

1.1 Background

Bangalore is composed of a number of neighbourhoods spread across a total area of 709 sq Km. There are many venues (especially restaurants, hotels and cafes) which can be explored. This project explores various venues in Bangalore and attributes the data based on user ratings and average price. To explore this information, this project involves the juxtaposition of both the Foursquare API and the Zomato API to fetch complete information of various venues (including name, address, category, rating, and price). Further, a map of the neighbourhoods with specific color attributes will be plotted to highlight their position, and information about these neighbourhoods. Such plots imbibe bountiful information in the form of their colored representations and location on the map.

1.2 Interested Audience

The target audience for such a project is twofold. Firstly, any person who is visiting Bangalore, India can use the plots and maps from this project to quickly select places that suit their budget and rating preferences. Secondly, a stakeholder who is looking to open a new restaurant can use this information and get the best neighbourhood based on price and restaurant type they are interested in. We have directed the project targeting more to the stakeholders.

2. Data

2.1 Data Sources and Cleaning

Pincodes of neighborhoods in Bangalore city through a csv downloaded from **https://github.com/sanand0/pincode/blob/master/data/IN.csv** . Below are the first 5 entry of the database.

	Country	Pincode	Neighborhood	City	Latitude	Longitude
0	IN	744101	Marine Jetty	South Andaman	11.6667	92.7500
1	IN	744101	Port Blair	South Andaman	11.6667	92.7500
2	IN	744101	N.S.Building	South Andaman	11.6667	92.7500
3	IN	744102	Haddo	South Andaman	11.6833	92.7167
4	IN	744102	Chatham	South Andaman	11.7000	92.6667

Then we cleaned the data and filtered data only related to our project, that is Bangalore city.

	Neighborhood	City	Latitude	Longitude
0	Bangalore G.P.O.	Bangalore	12.9914	77.5944
1	Legislators Home	Bangalore	12.9914	77.5944
2	Vasanthanagar	Bangalore	12.9914	77.5944

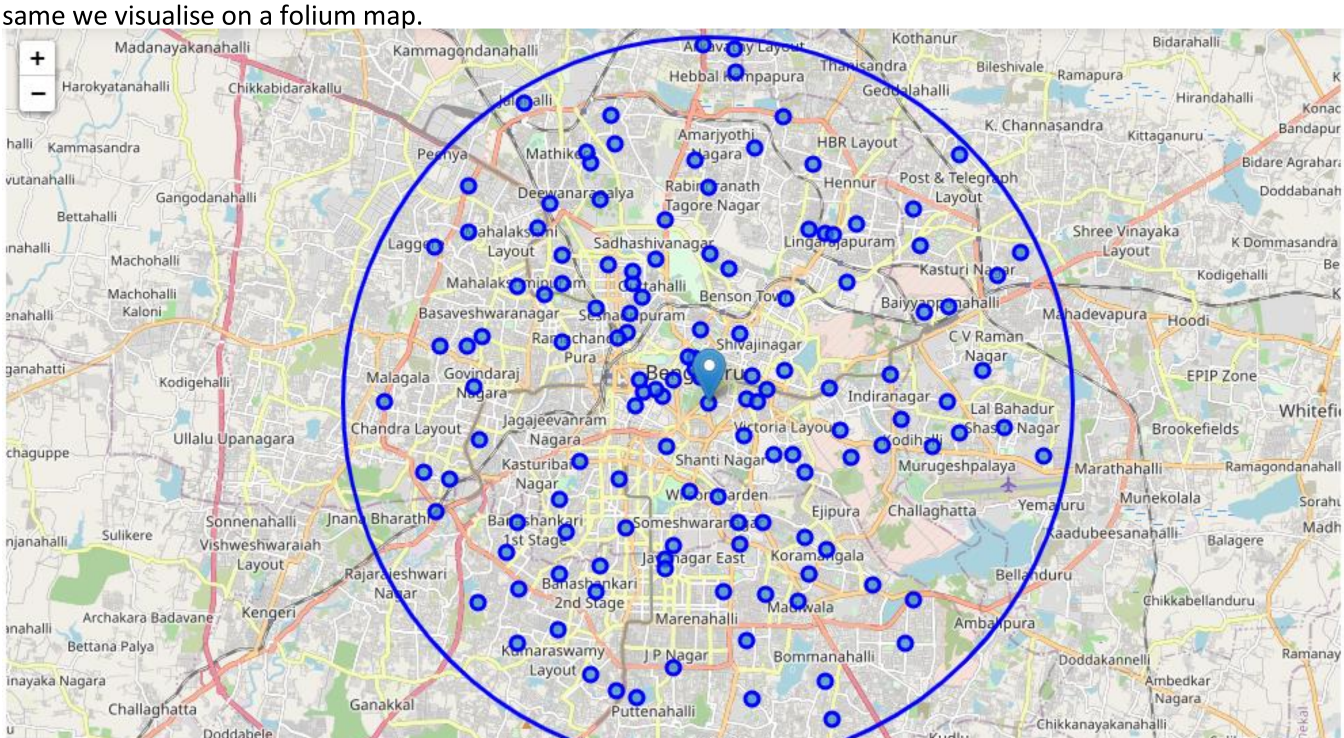
As the Latitude and longitude were against pin code and there were multiple pin codes in same neighbourhood, so we used Neighbourhood name to get exact centre of that neighbourhood from Google API.

	Neighborhood	City	Latitude	Longitude	address	latitude	longitude
0	Bangalore G.P.O.	Bangalore	12.9914	77.5944	Bangalore G.P.O.,Bangalore	12.981220	77.595069
1	Legislators Home	Bangalore	12.9914	77.5944	Legislators Home,Bangalore	12.982845	77.589609
2	Vasanthanagar	Bangalore	12.9914	77.5944	Vasanthanagar,Bangalore	12.989619	77.592795
3	Mahatma Gandhi Road	Bangalore	12.9914	77.5944	Mahatma Gandhi Road,Bangalore	12.974707	77.609409
4	Vidhana Soudha	Bangalore	12.9914	77.5944	Vidhana Soudha,Bangalore	12.979673	77.591244
5	Rajbhavan (Bangalore)	Bangalore	12.9914	77.5944	Rajbhavan (Bangalore),Bangalore	12.982467	77.591303
6	HighCourt	Bangalore	12.9914	77.5944	HighCourt,Bangalore	12.977874	77.592635
7	Cubban Road	Bangalore	12.9914	77.5944	Cubban Road,Bangalore	12.978070	77.605785
8	Bangalore Bazaar	Bangalore	12.9914	77.5944	Bangalore Bazaar,Bangalore	12.928798	77.676381
9	CMM Court Complex	Bangalore	12.9914	77.5944	CMM Court Complex,Bangalore	12.973127	77.583151
10	Dr. Ambedkar Veedhi	Bangalore	12.9914	77.5944	Dr. Ambedkar Veedhi,Bangalore	12.979702	77.592547

We also obtained Latitude and Longitude of Bangalore Centre from Google API. Then we calculated distance between neighbourhood and Bangalore centre, and removed all neighbourhoods which were more than 10kms away from Bangalore Centre. Database after this looked something like this.

Neighborhood	City	address	latitude	longitude	distance from center
Bangalore City	Bangalore	Bangalore City,Bangalore	12.971599	77.594563	0.0
HighCourt	Bangalore	HighCourt,Bangalore	12.977874	77.592635	729.0
Dr. Ambedkar Veedhi	Bangalore	Dr. Ambedkar Veedhi,Bangalore	12.979702	77.592547	928.0
Vidhana Soudha	Bangalore	Vidhana Soudha,Bangalore	12.979673	77.591244	968.0
Museum Road	Bangalore	Museum Road,Bangalore	12.972280	77.604245	1053.0
Bangalore G.P.O.	Bangalore	Bangalore G.P.O.,Bangalore	12.981220	77.595069	1072.0
Sri Jayachamarajendra Road	Bangalore	Sri Jayachamarajendra Road,Bangalore	12.977216	77.585946	1124.0
CMM Court Complex	Bangalore	CMM Court Complex,Bangalore	12.973127	77.583151	1249.0
Rajbhavan (Bangalore)	Bangalore	Rajbhavan (Bangalore),Bangalore	12.982467	77.591303	1260.0
Richmond Town	Bangalore	Richmond Town,Bangalore	12.963382	77.603489	1331.0
Ashoknagar (Bangalore)	Bangalore	Ashoknagar (Bangalore),Bangalore	12.971885	77.607009	1350.0

The same we visualise on a folium map.



We then used this Neighbourhood data to get list of restaurants from **Four Square API.** We kept the limit to 500 restaurant per neighbourhood within a radius of 1km. below is the data format what we extracted from four square.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Neighborhood Distance	Venue	Venue Latitude	Venue Longitude	Venue Distance from center	Venue Id	Venue Category
Bangalore City	12.971599	77.594563	0.0	Toscano	12.971980	77.596066	168	4bc1cd90b492d13a4e74a660	Italian Restaurant
Bangalore City	12.971599	77.594563	0.0	Shiro	12.971900	77.596236	184	4b895510f964a520442c32e3	Japanese Restaurant
Bangalore City	12.971599	77.594563	0.0	Café Noir	12.971995	77.596001	162	4baef172f964a5202ce33be3	French Restaurant
Bangalore City	12.971599	77.594563	0.0	J W Kitchen	12.972410	77.594592	90	523de40611d2996a150886fc	Asian Restaurant
Bangalore City	12.971599	77.594563	0.0	Bengaluru Baking Co.	12.971761	77.595128	63	51efe570498e01081549f692	Bakery

After removing duplicates venue and cleaning the data we used this dataframe to get details of each venue from **Zomato API.** From Zomato we obtained below details of each venue.

Venue_Id	Venue_Lat	Venue_Lng	Venue_Dist	Avg_Cost_For_2	Venue_Cuisines	Table_Booking	Online_Order	Venue_Rating	Venue_Res_Type
534d46b2498e81f5e91eddda	13.010413	77.648123	8.0	600.0	North Indian, Mughlai, Biryani, Chinese	0.0	1.0	3.8	Casual Dining
4c3c263b980320a1f9468ae4	12.935222	77.624375	11.0	1600.0	Continental, Mediterranean, North Indian, Chin...	1.0	1.0	3.9	Casual Dining
4d9cdc26c593a1cd66205319	12.996995	77.614308	12.0	200.0	Bakery, Fast Food, Beverages	0.0	0.0	3.0	Dessert Parlour
4be67bb32457a593aae7ac15	12.978357	77.640717	12.0	600.0	Pizza, Fast Food, Finger Food, American	0.0	1.0	3.3	Casual Dining
4bab3b34f964a520a99a3ae3	13.005642	77.569207	14.0	300.0	Fast Food	0.0	1.0	3.3	Quick Bites

3. Methodology and Data Analytics

In this project we have directed our efforts on detecting areas of Bangalore which will be best suitable for opening a new Specific restaurant.

In first step we have collected the required data of location and type (category) of every restaurant within 10km from Bangalore centre. We have also obtained rating and other details of all restaurants.

Second step in our analysis will be calculation and exploration of Restaurants across different neighbourhoods of Bangalore. We will focus on Distance from Centre, Rating, Average Price, and Density of all restaurants and restaurant type selected by Stake Holder, to get the best Neighbourhood for Opening a new restaurant.

Third and Final Step will be involving Prediction of Rating (above average or below) of the restaurant given stake holder selects cuisines, res type and neighbourhood where he wants to open the restaurant.

3.1 Understanding the data extracted

We started with getting the data type of the data frame obtained till now and discussing all features.

```
Out[180]: Neighborhood              object
          Neighborhood_Distance    float64
          Neighborhood_Lat         float64
          Neighborhood_Lng         float64
          Venue                     object
          Venue_Category            object
          Venue_Id                  object
          Venue_Lat                float64
          Venue_Lng                float64
          Venue_Dist               float64
          Avg_Cost_For_2           float64
          Venue_Cuisines            object
          Table_Booking            float64
          Online_Order             float64
          Venue_Rating              object
          Venue_Res_Type            object
          dtype: object
```

Neighbourhood - There are 131 unique Neighbourhood in Bangalore. and a total of 2301 restaurants. With Kormangala VIBk, haveing the most restaurants , ie 89

Neighborhood Distance - Average distance of Neighborhood distance form Bangalore Centre is approx 5km, which should be the case as we have considered Neighborhood till 10kms range only.

Venue - there are a total of 2301 restaurants with 1713 unique ones. That means there are restaurants with their chain in Bangalore, CCD contributing the most 64 in it.

Venue Category - There are 89 types of restaurants in Bangalore. And "Indian Restaurant" sitting at top with 682 restaurants.

Venue ID - We see a few '0' in Venue id, suggesting we didn’t got details of few restaurant from Zomato APi. We will drop these restaurants as we won’t be able to use them in our project.

Average Cost for 2 - INR 662/- is the amount which is the average charges by restaurants in Bangalore for 2 people according to Zomato.

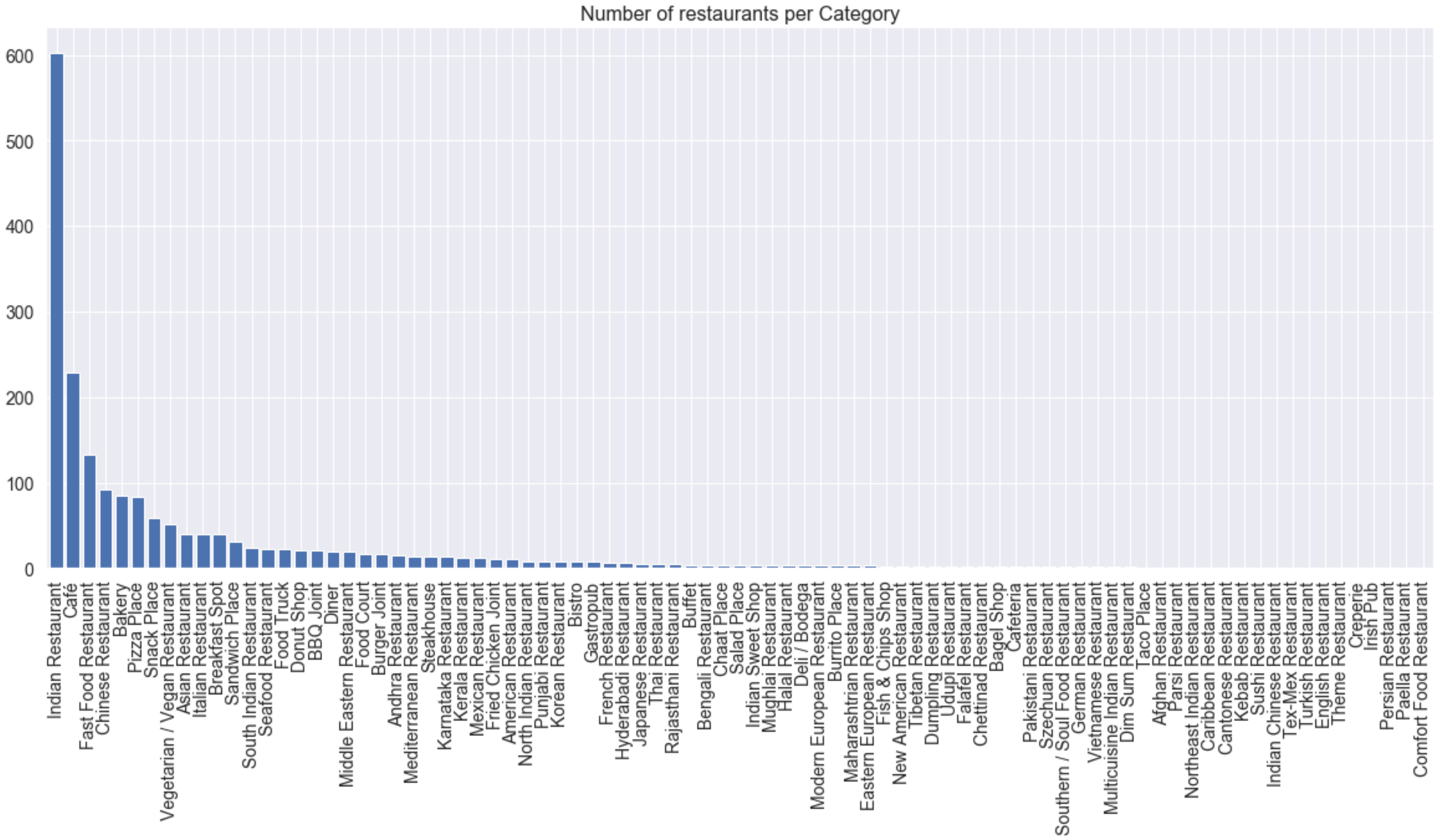
Venue-Cuisines - As cuisines is maintained as comma separated it is showing a very high value of unique cuisines. We will deal with this in our 3rd step of project while predicting the ratings.Same goes with **Venue_Res_Type**.

Online Order and Table Booking - Boolean feature, with 1, and 0 as values. Showing 64% and 9.7% of restaurant have that feature respectively.

Venue-Rating - We see an average rating of 3.8 out of 5 in Bangalore.

3.2 Venue Category

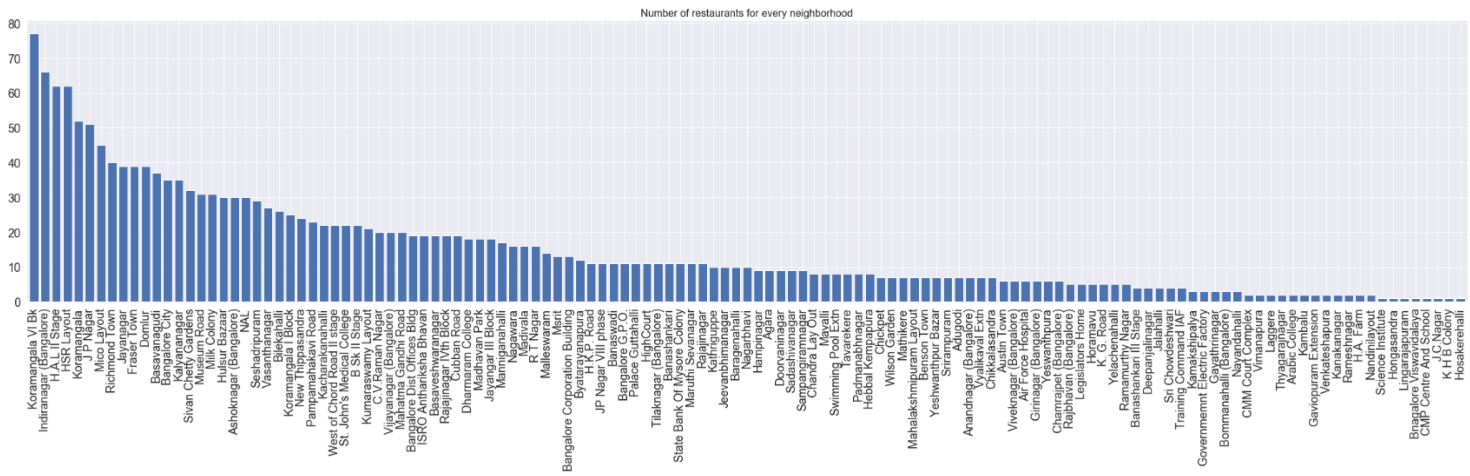
We begin our Analysis taking a look at the various categories of venues that exist in Bangalore



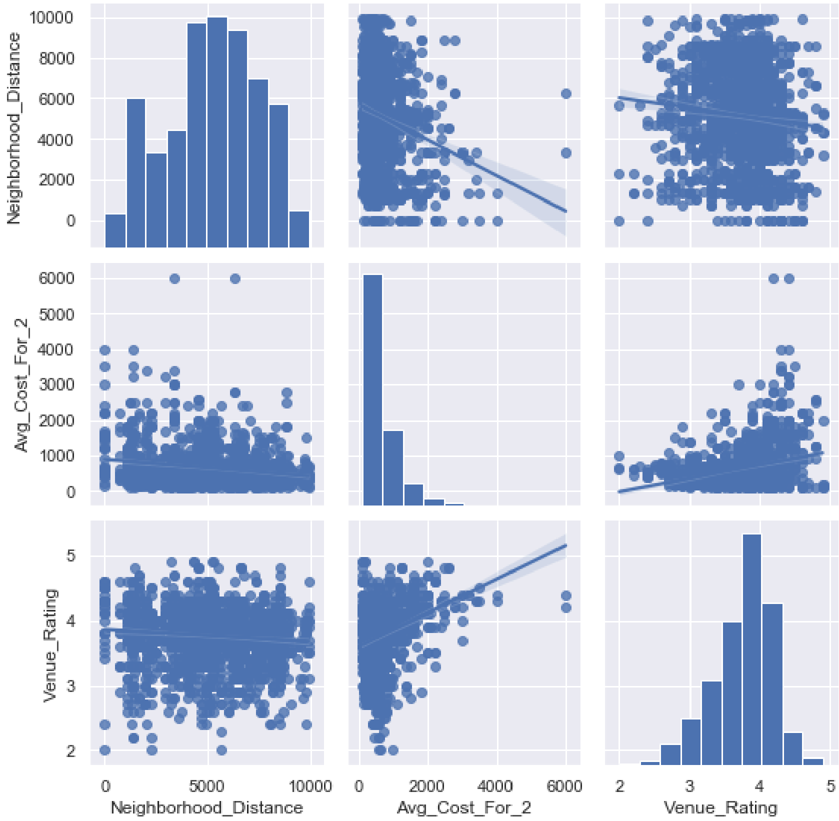
Indian restaurants is almost twice as the next Restaurant type, that is Cafe. While 50% of the type are having just 1 restaurant in Bangalore. So while selecting the neighborhood for **cost and rating**, we will have to consider all restaurants instead of 1 specific restaurant which we want to open.

3.3 Neighbourhoods

We will check for number of restaurants in each neighbourhood.

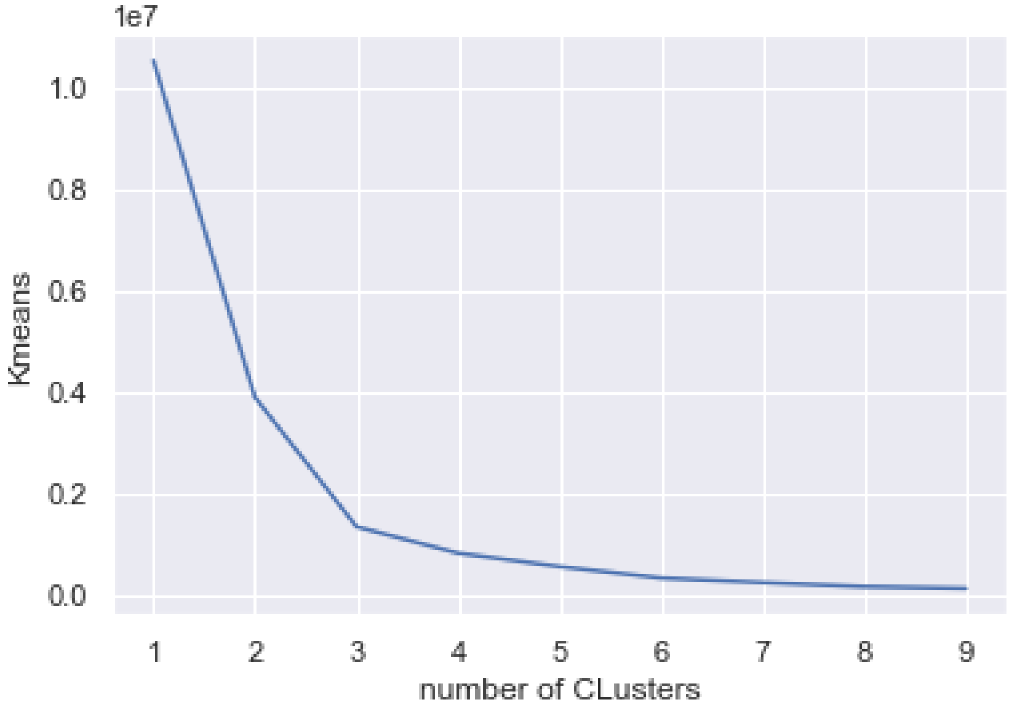


We see here that there are few locations with more than 50 restaurants and few neighbourhoods with just 1 restaurant. We have not only given priority to number of restaurants in a neighbourhood, but also to what is the average rating of the restaurants, average cost, distance of neighbourhood from centre. So let's take a look at the relationship between these features for all neighbourhood.

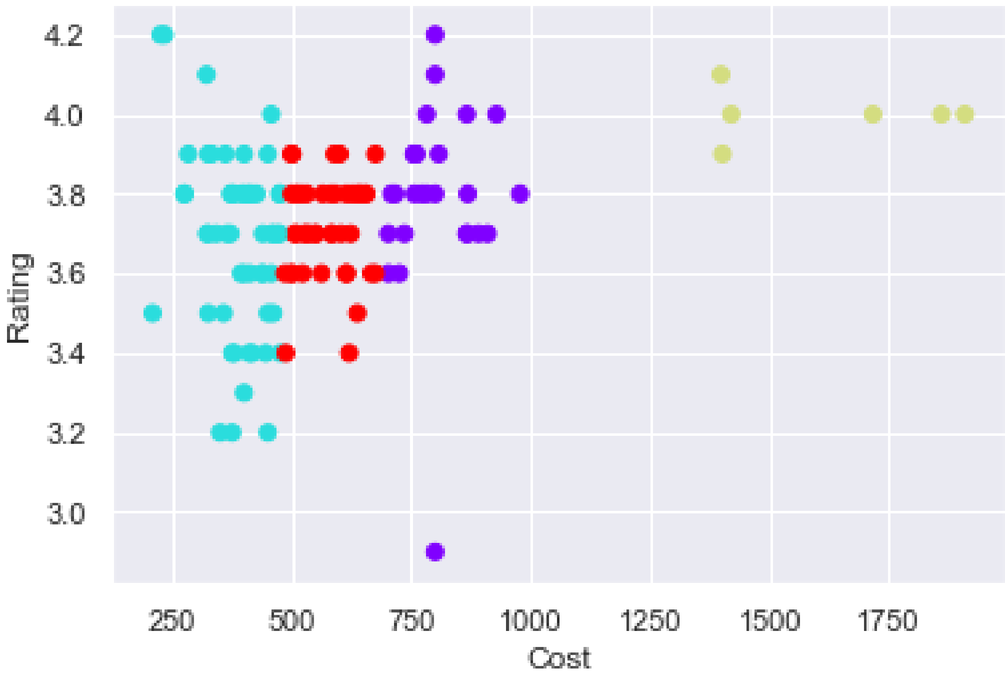


As the distance increases, average rating and average price both decreases but are not effected that much. So all the 3 features can be used for our analysis.

Mapping of Neighbourhood with respect to their ratings and price for 2 using **KMeans clustering.**

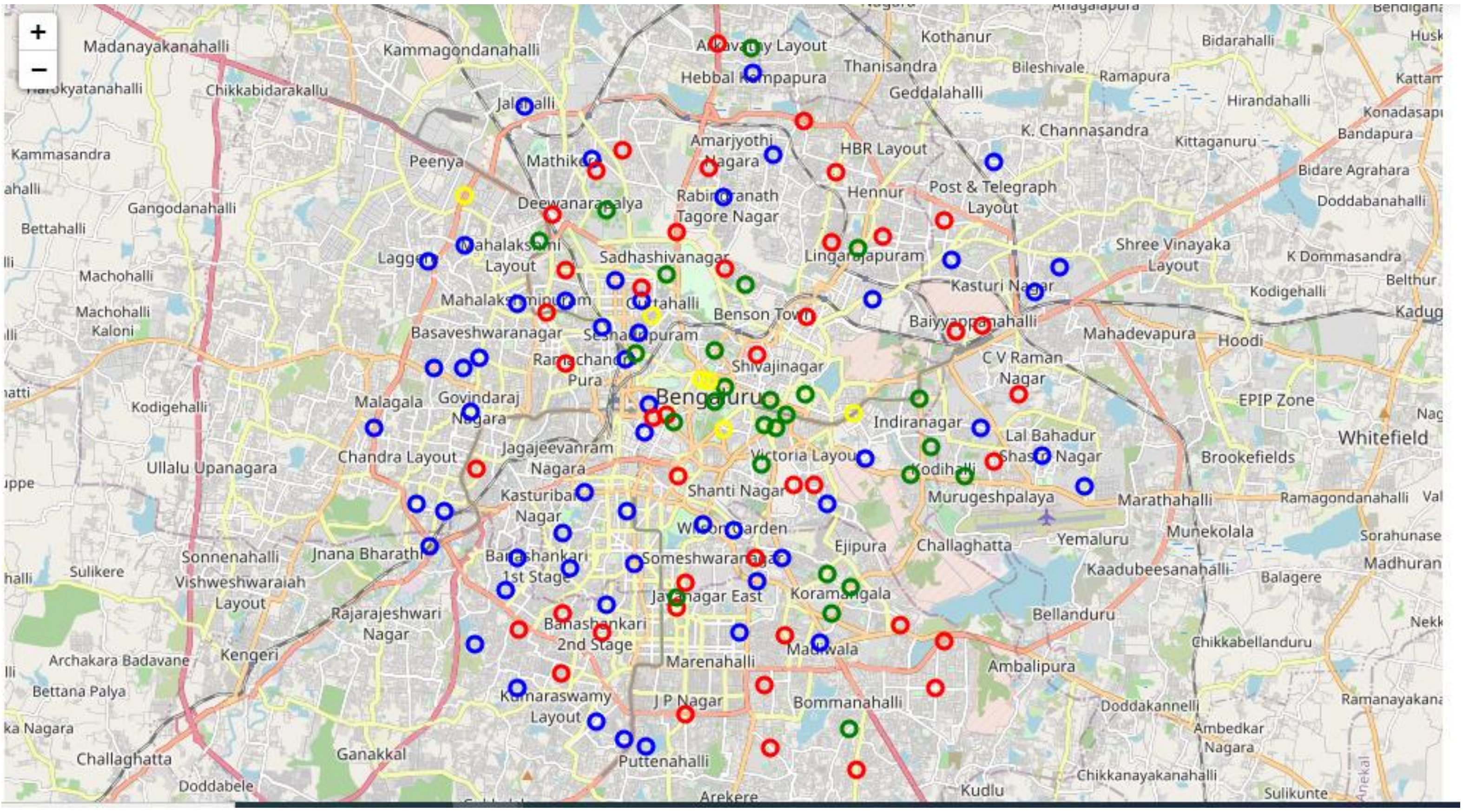


So according to **Elbow Method,** we can say that we can divide neighbourhoods in 4 different clusters. Below is the clustering result with k=4.



So if we see here, High Cost will get you better rating, but there is a mix review as we lower our price. So In our analysis, we will not give much weightage to Cost. Rather we will focus on distance and total number of restaurants more.

If we see the same on Bangalore map, Kmeans have not only divided the neighbourhood with cost and rate, but it can also be said it has divided the neighbourhood according to distance form centre as well.



So according to all graphs and observation, and discussion with stakeholders we are divided the weightage of different features to mark the best neighbourhoods.

distance=0.1

average_rating=0.05

average_cost=0.05

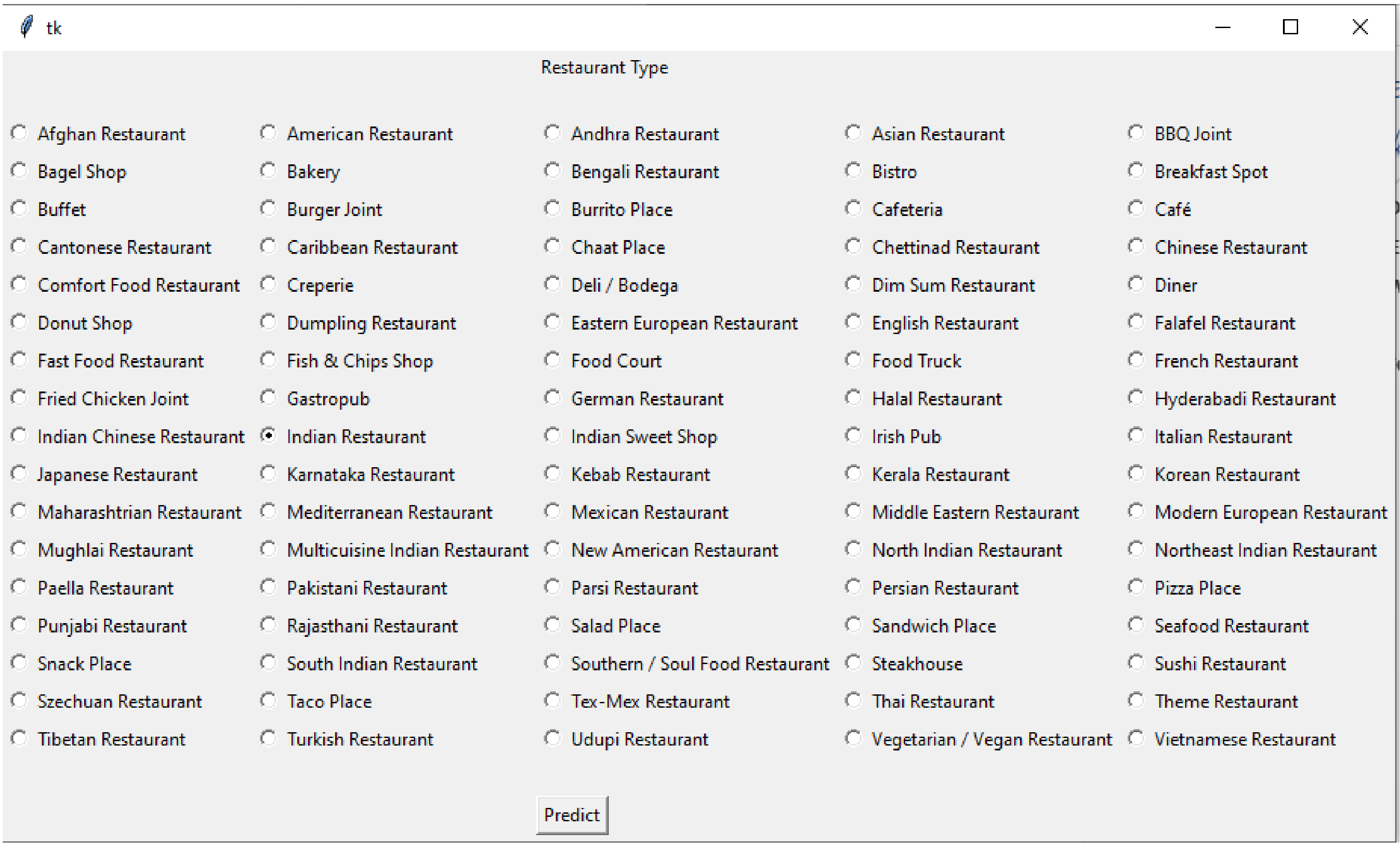
total_res=0.15

user_selected_average_rating=0.25

user_selected_average_cost=0.1

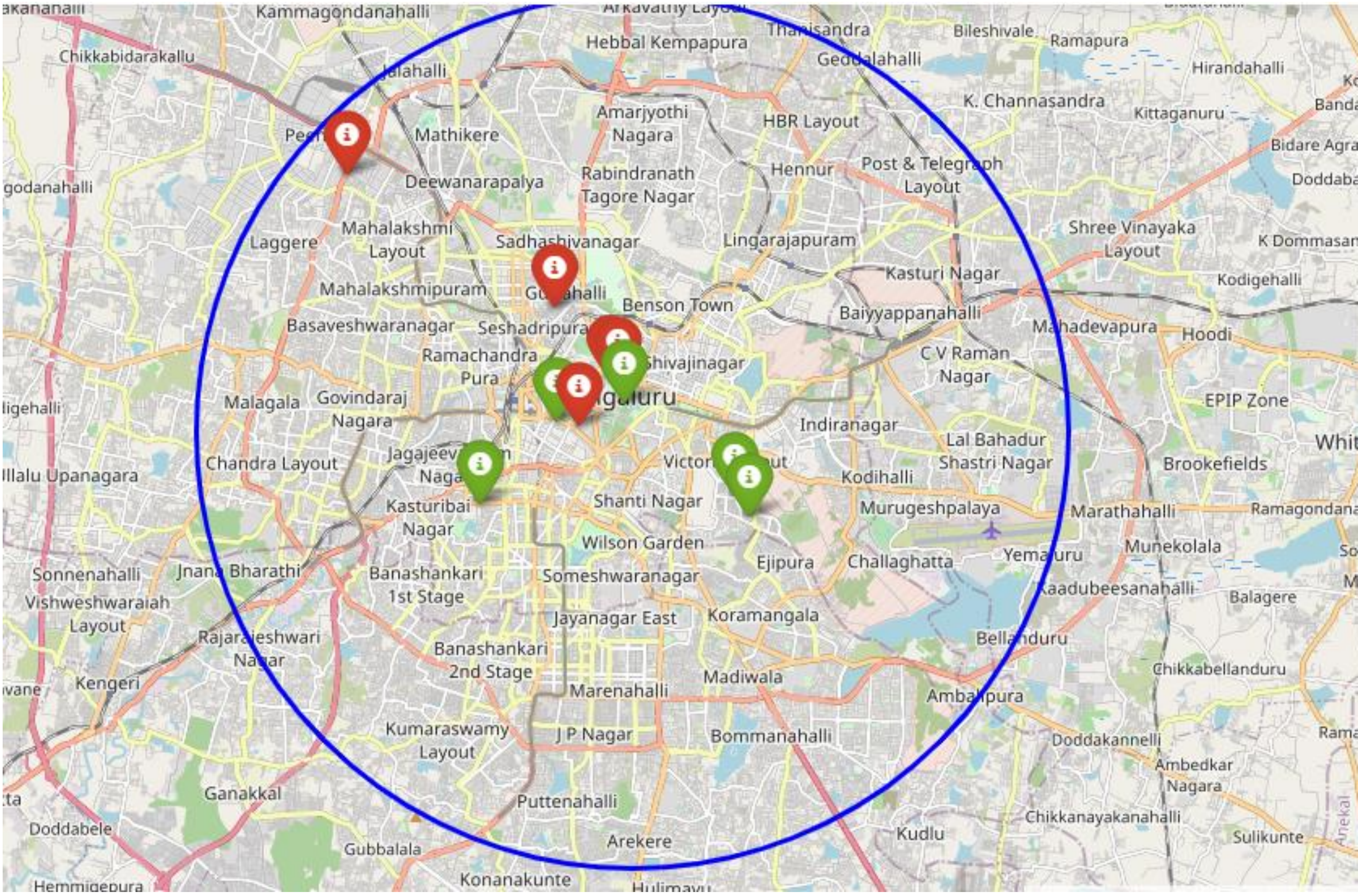
user_selected_total_res=0.3

Now, when we have divided weightage of each feature, we asked the stakeholder to select 1 Venue Category which they wanted to know the neighbourhood of. And then according to the weightage we sorted the top 10 preferred neighborhood.



So after all the calculation we made a dataframe with top 10 neighbourhood, where Stakeholder should target his restaurant. Please note this dataframe will completely depends on type of restaurant Stakeholder selected.

Neighborhood	Neighborhood Distance	Avg_Cost_For_2	Venue_Rating	Total Restaurants	User Selected-avg cost	User Selected-venue Rating	User Selected-Total Restaurants	Neighborhood_Lat	Neighborhood_Lng	Weightage
Legislators Home	1362	1860	4.0	4	1733	4.2	3	12.982845	77.589609	0.365716
Palace Guttahalli	3386	1909	4.0	11	2230	3.9	5	12.997414	77.578042	0.320346
Yeswanthpura	8859	1716	4.0	6	1800	4.1	1	13.025030	77.534024	0.300818
CMM Court Complex	1249	800	4.2	2	200	4.2	1	12.973127	77.583151	0.297917
Rajbhavan (Bangalore)	1259	1420	4.0	4	100	3.9	1	12.982467	77.591303	0.283893
HighCourt	729	890	3.7	11	1150	3.4	2	12.977874	77.592635	0.268478
Viveknagar (Bangalore)	3270	416	3.8	6	575	4.2	2	12.954487	77.619072	0.262048
K. G. Road	1797	510	3.7	4	483	3.9	3	12.974276	77.578219	0.248898
Austin Town	2728	641	3.8	6	600	3.7	2	12.958768	77.615995	0.245467
Chamrajpet (Bangalore)	3875	341	3.7	6	150	4.2	1	12.956987	77.562140	0.243401



Here red markers shows top 5 and green shows next 5 preferred neighbourhoods

3.4 Rating Prediction

Once we know the neighbrhood preference, we made a model which could predict rating of the new restaurant in each neighbourhood. We kept it simple for stakeholder and divided the rating into 2 categories, **below average rating** in Bangalore and **above average rating** in Bangalore.

For the same we did all the data cleaning similar to above step, just we included res_type and cuisines from Zomato API for this prediction.
As there were multiple cuisines for same restaurant we converted cuisine to features with one hot encoding and tried 4 models for doing the classification.

Algorithm	Jaccard	F1-score	LogLoss
KNN	0.70	0.69	NA
Decision Tree	0.75	0.75	NA
SVM	0.69	0.69	NA
LogisticRegression	0.63	0.57	0.63

4. Results and Discussion

Our analysis shows that although there is a great number of restaurants in Bangalore (~2000 in our initial area of interest which was 10x10km around centre), there are neighborhoods which are still rising up in terms of Venues around it.

Not only we gave stakeholder top 10 neighborhood for any type of Restaurant but also, a prediction of rating in those area depending on their average cost per 2 people. While selecting the neighborhood we took in consideration, type of restaurant Stakeholder wants to start, neighborhood distance from centre, average rating of all restaurant in neighborhood against average rating of particular restaurant and finally average cost at all restaurant vs average cost for that particular type which stakeholder wants to start.

Then we used classification methods to classify restaurants according to their type, cuisines, location, cost into 2 categories.

- Below Average Rating
- Above Average Rating

Best model was "Decision Tree" as it gave best Jaccard and f1 score. So we can use this model to predict rating as well.

So for a final comment - Recommended neighbourhood should be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

5. Conclusion

Purpose of this project was to identify Bangalore Neighborhood, suitable for any kind of restaurant preferred by stakeholder. By using Google , Foursquare and Zomato Api we were able to get restaurant details of all Neighborhood in Bangalore , which further helped us in finding the best Neighborhood. Also we were able to classify restaurants using a decision tree model and label them as below average and above average Rating.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every Neighborhood etc.