Project — The Battle of Neighbourhoods

Narasimha Rao

July 2020

Table of contents

1. Introduction: Business Problem

1.1 Background

Bangalore , officially known as Bengaluru, is the capital of the Indian state of Karnataka. It has a population of about 10 million and a metropolitan population of about 8.52 million, making it the third most populous city and fifth most populous urban agglomeration in India.Located in southern India on the Deccan Plateau, at a height of over 900 m (3,000 ft) above sea level, Bangalore is known for its pleasant climate throughout the year. Its elevation is the highest among the major cities of India.

Bengaluru is widely regarded as the "Silicon Valley of India" (or "IT capital of India") because of its role as the nation's leading information technology (IT) exporter.Indian technological organisations such as ISRO, Infosys, Wipro and HAL are headquartered in the city. A demographically diverse city, Bangalore is the second fastest-growing major metropolis in India.

Bangalore is also home to many vegan-friendly restaurants and vegan activism groups, and has been named as India's most vegan-friendly city by PETA India.

With it's diverse culture , comes diverse food items. There are many restaurants in Bangalore City, each belonging to different categories like South Indian, North Indian, Andhra, Chinese , Italian , French etc. So as part of this project , we will list and visualise all major parts of Bangalore City.

1.2 Problem

The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the aggregate rating of each restaurant, establishment of different types of restaurant at different places, Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry has'nt been saturated yet and the demand is increasing day by day. Inspite of increasing demand it however has become difficult for new restaurants to compete with established restaurants. Most of them serving the same food. Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality?. Which are locations has more restaurants? Do the entire locality loves Chinese or Andhra food. If yes then is that locality populated by a particular sect of people for eg. Andhra, Foreigners, Jain, Marwaris, Gujaratis,etc. These kind of analysis can be done using the data, by studying different factors.

1.3 Interest

The project aims to select the best location to stay in Bangalore if someone likes Chinese or Andhra cuisines and Best locations to start a new Chinese and Andhra Cuisines or the best location to stay for edible person based on the total restaurents, locality, riviews and neighborhoods.

➢ What places are having best restaurant in Bangalore?
➢ Which place are suitable for edible person in Bangalore city
➢ Which place are not suitable for edible person in Bangalore city
➢ Which is the best place to stay if I prefer Chinese Cuisine or Which areas have large number of Chinese Resturant Market ?
➢ Which is the best place to stay if I prefer Andhra Cuisine ?
➢ What is best location in Bangalore City for Chinese Cuisine ?
➢ What is best location in Bangalore City for Andhra Cuisine ?
➢ Which all areas have less number of resturant ?

2.Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project from the following data sources:

Bangalore Resturants data that contains list Locality, Resturant name,Rating along with their latitude and longitude.

➢ Data source : Zomato kaggel dataset
➢ Description : This data set contains the required information. And we will use this data set to explore various locality of Bangalore city.

Nearby places in each locality of Bangalore city.

➢ Data source : Fousquare API
➢ Description : By using this api we will get all the venues in each neighborhood.

Latitude and Longitude values from the geolocater.geocodes.

The Zomoto data set contains the following columns:

url: URL of the restaurant in the zomato website.

address: address of the restaurant in Bengaluru.

name: Name of the restaurant.

online_order: whether online ordering is available in the restaurant or not.

book_table: table book option available or not.

rate: contains the overall rating of the restaurant out of 5.

votes: total number of votes.

phone: phone number of the restaurant.

location: the neighborhood in which the restaurant is located.

rest_type: restaurant type.

dish_liked: dishes people liked in the restaurant.

cuisines: food styles, separated by comma.

approx_cost: approximate cost for meal for two people.

reviews_list: list of tuples containing reviews for the restaurant, each tuple consists of two values, rating and review by the customer.

menu_item: list of menus available in the restaurant.

listed_in(type): type of meal.

listed_in(city): contains the neighborhood in which the restaurant is listed.

2.2 Data Analysis and Cleansing

Data preparation for all the phases done separately.

Dataset is not clean and contains redundant as well as unnecessary data. So, done the cleaning process and deleted the unnecessary or redundant features.

For data analysis, we do not need the contact details of the restaurant so, deleting the following features :

- ➢ url
- ➢ address
- ➢ phone

The menu_item feature contains the names of the dishes available in the restaurant. This will hardly impact our analysis, as it is highly mathematics driven. Also, we have other features like rest_type, cuisines, listed_in(type) and dish_liked which give us a fair idea about what the restaurants have to offer, as we do not need to be very specific about what all dishes are available in the restaurant for the analysis. So, droping this feature as well.

The reviews_list feature also will hardly contribute for the analysis as it is pure text and we also have features like rate and votes that cover the necessary information. But, one can still extract this feature and play around like running an NLP algorithm. As far as the data analysis is considered deleting this feature as well.

Deleted some of the features and lets clean the remaining.

The rate feature is a string and it contains the character '/'.This is not required, and so I will remove the '/5' from each entry in the column, changing the datatype of the feature to float.

Renamed the some features, for simplicity sake.

The following features were renamed :

➢ approx_cost(for two people) : **avg_cost**
➢ listed_in(type) : **meal_type**
➢ listed_in(city) : **city_loc**

The feature description given on kaggle, makes one think about the two features : location and city. Both have the same description, so this is redundant information. Now, which one to keep?

In this approach I kept the city feature(0 null values) and deleting the location feature.

Now Data is ready, lets have a look at it.

	name	online_order	book_table	rate	votes	location	rest_type	dish_liked	cuisines	avg_cost	meal_type	city_loc
0	Jalsa	Yes	Yes	4.1	775	Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja...	North Indian, Mughlai, Chinese	800	Buffet	Banashankari
1	Spice Elephant	Yes	No	4.1	787	Banashankari	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G...	Chinese, North Indian, Thai	800	Buffet	Banashankari
2	San Churro Cafe	Yes	No	3.8	918	Banashankari	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc...	Cafe, Mexican, Italian	800	Buffet	Banashankari
3	Addhuri Udupi Bhojana	No	No	3.7	88	Banashankari	Quick Bites	Masala Dosa	South Indian, North Indian	300	Buffet	Banashankari
4	Grand Village	No	No	3.8	166	Basavanagudi	Casual Dining	Panipuri, Gol Gappe	North Indian, Rajasthani	600	Buffet	Banashankari

	name	online_order	book_table	rate	votes	rest_type	dish_liked	cuisines	avg_cost	meal_type	city_loc
count	51672	51672	51672	51672.000000	51672.000000	51466	23639	51672	51352	51672	51672
unique	8785	2	2	NaN	NaN	93	5271	2723	70	7	30
top	Cafe Coffee Day	Yes	No	NaN	NaN	Quick Bites	Biryani	North Indian	300	Delivery	BTM
freq	96	30428	45223	NaN	NaN	19129	182	2913	7576	25917	3278
mean	NaN	NaN	NaN	2.983062	283.920073	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	1.515475	804.138258	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	0.000000	0.000000	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	3.000000	7.000000	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	3.600000	41.000000	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	3.900000	198.000000	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	4.900000	16832.000000	NaN	NaN	NaN	NaN	NaN	NaN

Extracted the Latitude and longitude values from Geolocator and added both the columns to the Data frame.

	name	online_order	book_table	rate	votes	rest_type	dish_liked	cuisines	avg_cost	meal_type	city_loc	Latitude	Longitude
0	Jalsa	Yes	Yes	4.1	775	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Laja...	North Indian, Mughlai, Chinese	800	Buffet	Banashankari	12.915221	77.573598
1	Spice Elephant	Yes	No	4.1	787	Casual Dining	Momos, Lunch Buffet, Chocolate Nirvana, Thai G...	Chinese, North Indian, Thai	800	Buffet	Banashankari	12.915221	77.573598
2	San Churro Cafe	Yes	No	3.8	918	Cafe, Casual Dining	Churros, Cannelloni, Minestrone Soup, Hot Choc...	Cafe, Mexican, Italian	800	Buffet	Banashankari	12.915221	77.573598
3	Addhuri Udupi Bhojana	No	No	3.7	88	Quick Bites	Masala Dosa	South Indian, North Indian	300	Buffet	Banashankari	12.915221	77.573598
4	Grand Village	No	No	3.8	166	Casual Dining	Panipuri, Gol Gappe	North Indian, Rajasthani	600	Buffet	Banashankari	12.915221	77.573598

With the help of city_loc, rate, name, Latitude and Longitude features we can predict the Best restaurants and locations in Bangalore city and we can make the clustering.

3.Methodology

3.1 Exploratory Data Analysis – EDA

Using this dataset, we begin by analyzing the top localities in Bangalore that have the most number of restaurants types. This will allow us to better understand the places where many venues co-exist and are worth visiting. I'll also explore the venues based on the ratings.
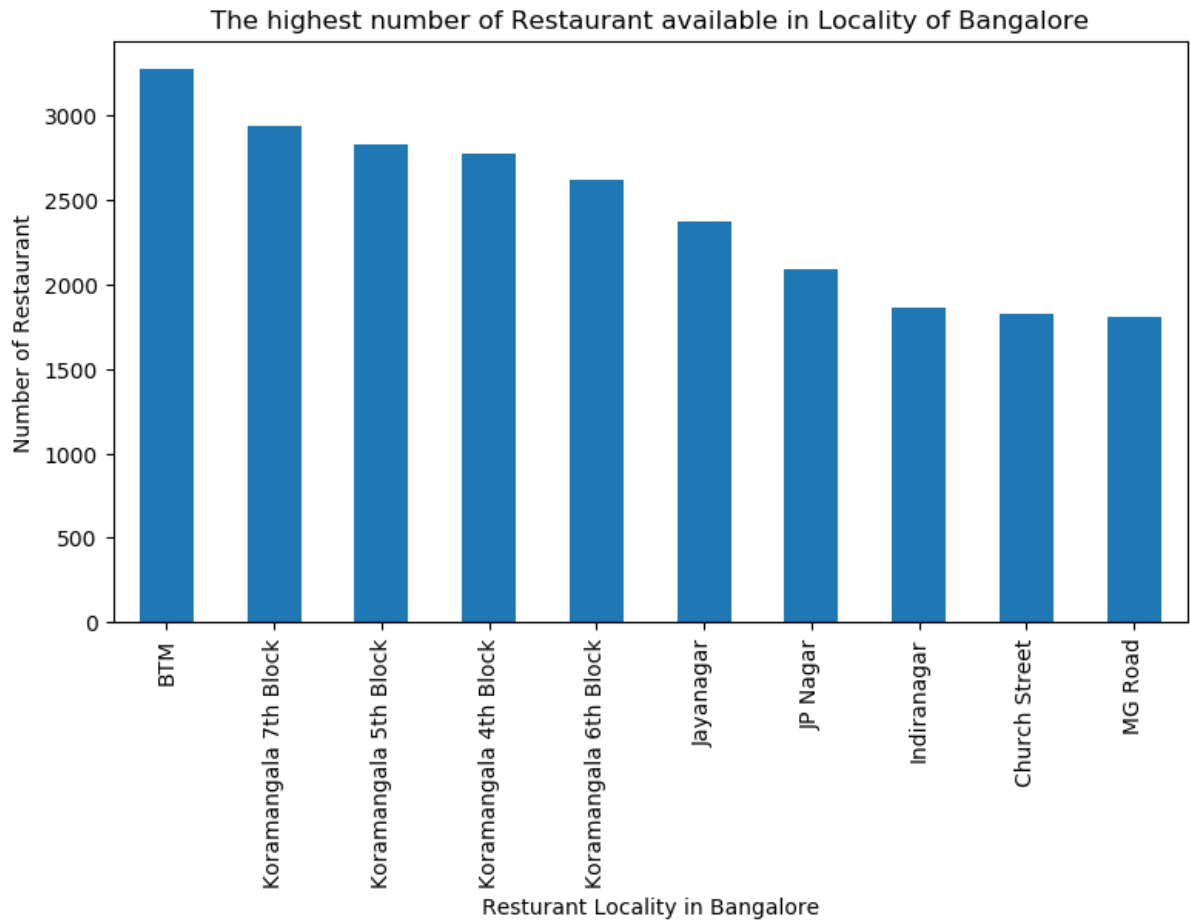
The aim to identify best localities based on the total restaurants, locality, reviews and neighbourhoods in Bangalore city for:

- ➢ More restaurants to stay for edible persons
- ➢ Best restaurants to visit
- ➢ Best Locality for Chinese cuisine
- ➢ Best Locality for Andhra cuisine

1. **Number of Restaurants in each Locality:**

From the below bar graph, we see that the locality with the greatest number of restaurants is BTM, followed by Koramangala and Jayanagar. Based on the above localities, an edible person can stay in BTM or Koramangala.

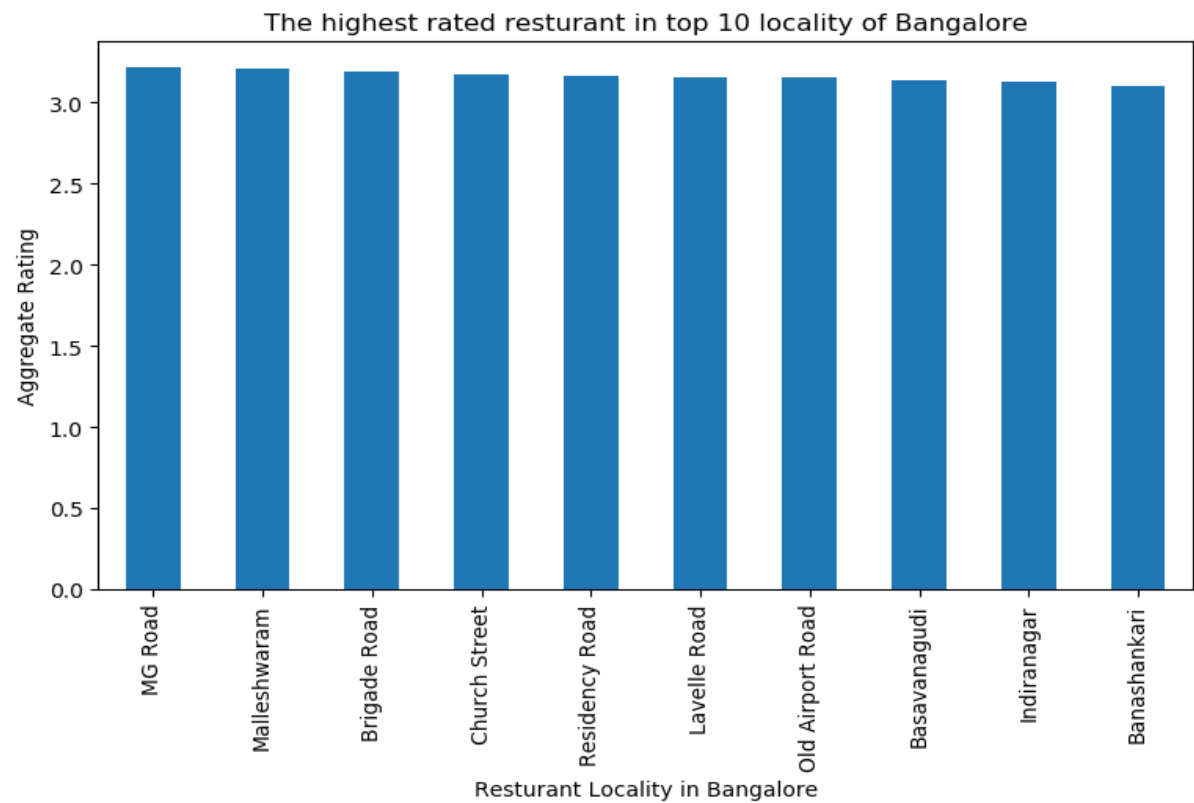Places are suitable for edible person in Bangalore city: BTM, Koramangala and Jayanagar

The highest number of Restaurant available in Locality of Bangalore

Places are not suitable for edible person in Bangalore city: New BEL Road & Banashankari

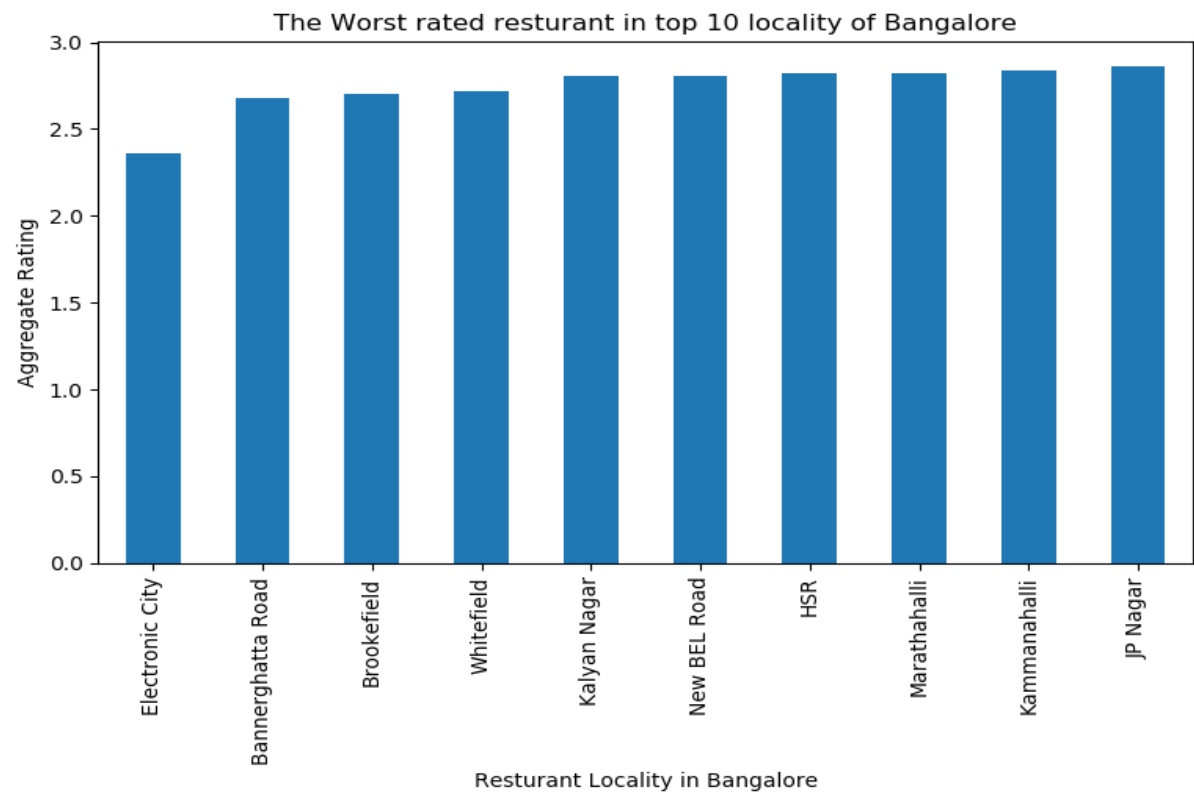The lowest number of Restaurant available in Locality of Bangalore

2. **Best and Worst restaurants in Bangalore:**

Best restaurant Localities in Bangalore City: MG Road, Malleshwaram and Brigade Road

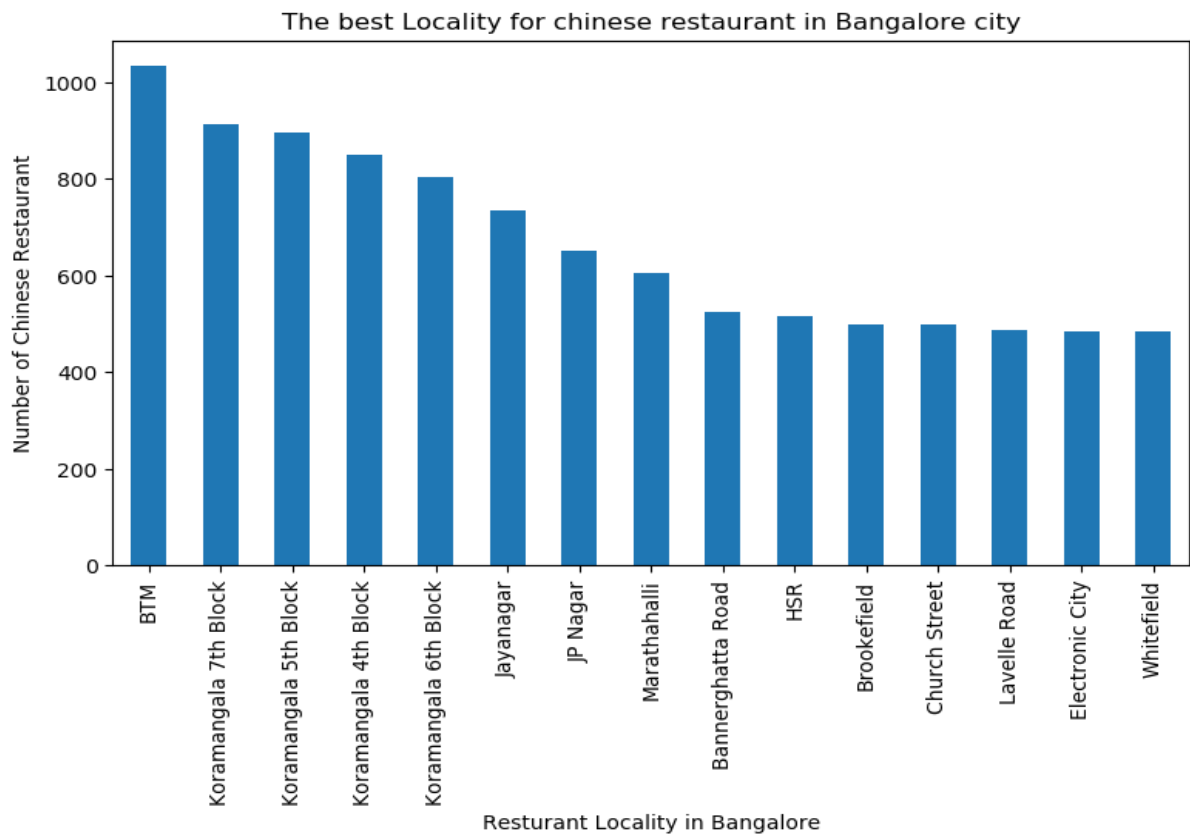The highest rated resturant in top 10 locality of Bangalore

Low rating restaurant Localities in Bangalore City: Electronic City, Bannerghatta Road

The Worst rated resturant in top 10 locality of Bangalore

3. **Best Locality for Chinese cuisine**

Best locations to stay if someone likes Chinese cuisine are BTM and Koramangala

The best places to start for Chinese restaurant in Bangalore city: Malleshwaram and Rajaji Nagar
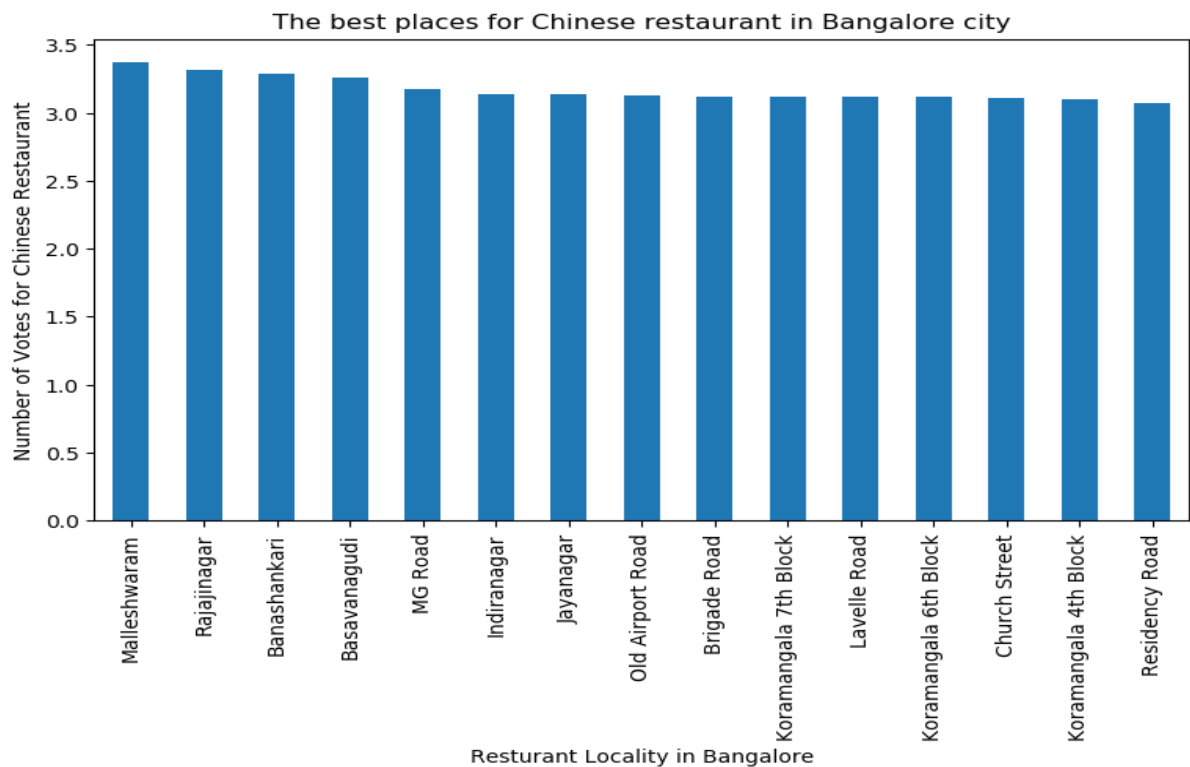
4. **Best Locality for Andhra cuisine**

Best locations to stay if someone likes Andhra cuisine are Marathahalli and BTM
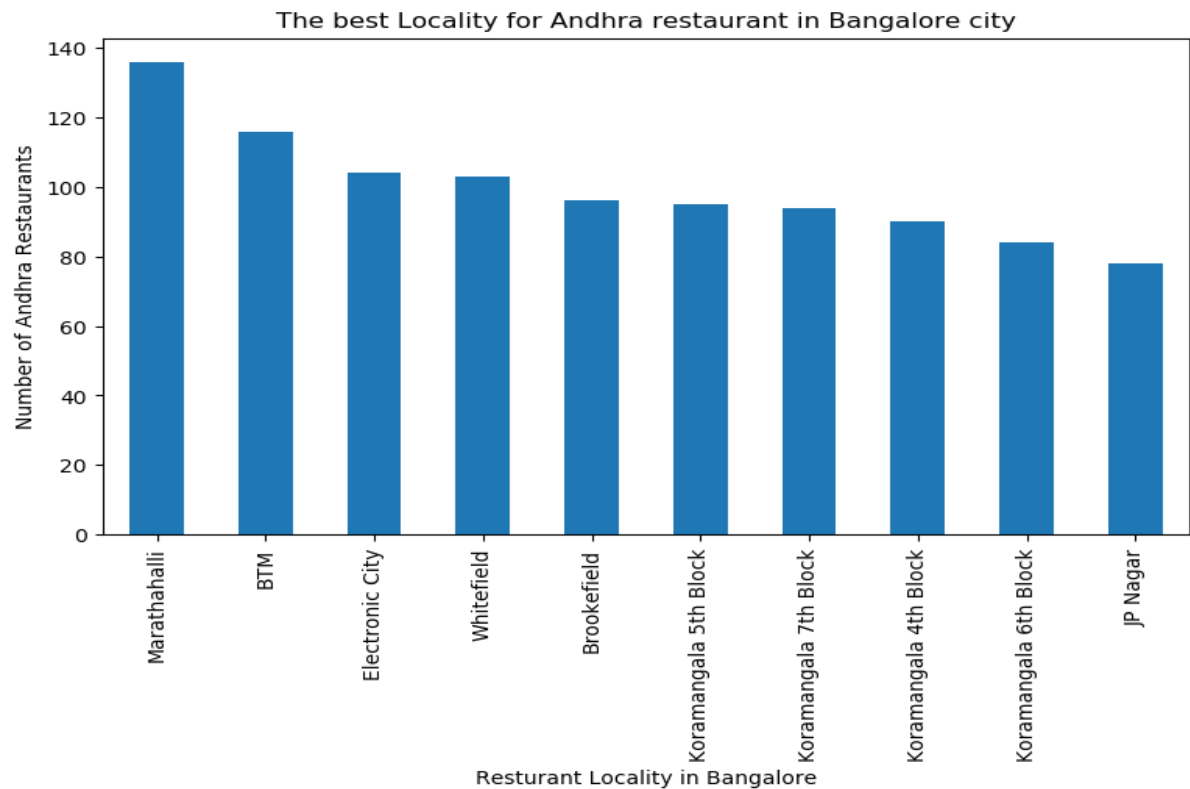
The best places to start for Andhra restaurant in Bangalore city: Lavelle Road and Residency Road
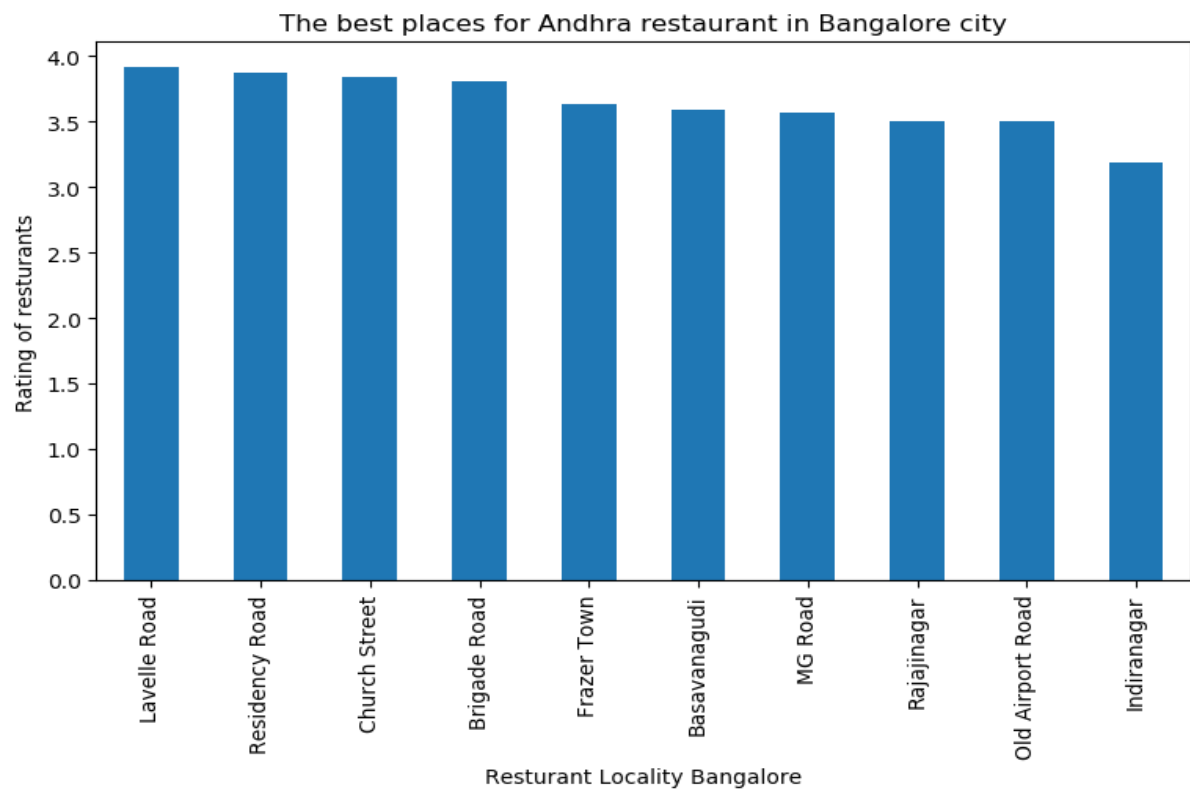
3.2 Modelling

Using the final data set with the latitude and longitude we can build a clustering.

We will cluster the various localities in Bangalore based on its restaurants rating and city location. We will use the K-mean Clustering algorithm to cluster the localities into K different clusters. Restaurants in a cluster will be similar to each other, while restaurants in other clusters will be dissimilar to restaurants in another cluster. To find out the best value of K in K-means we will use the elbow method. We'll plot:

· values for K on the horizontal axis

· the distortion on the Y axis

When K increases, the centroids are closer to the cluster's centroids. The improvements will decline, at some point rapidly, creating the elbow shape. That point is the optimal value for K. In the plot below, K=2 and K=4 are optimal. We will choose K=4 for this project.
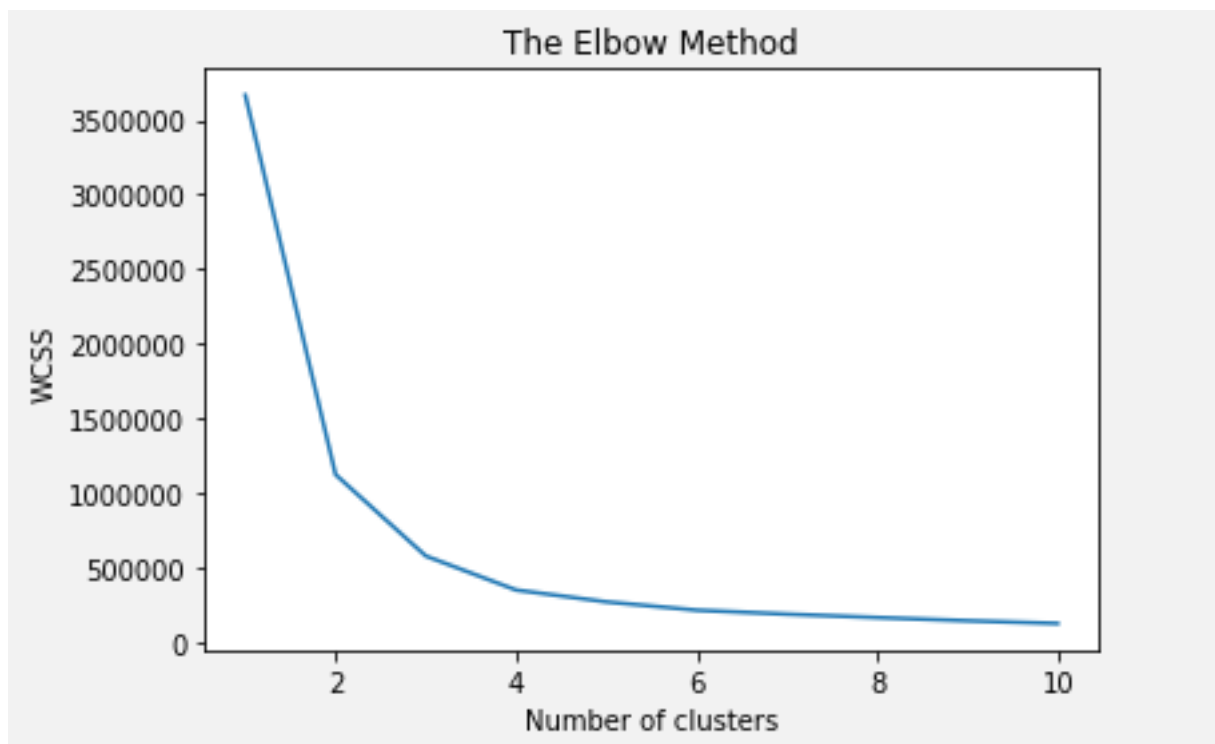

Now that we have clustered various localities based on their restaurants, we will plot these localities on the map of Bangalore with different colors for different clusters.

To plot these localities we will need their latitude and longitude values which we will obtain from the Geopy API. We will get a dataframe as follows:

Latitude and Longitude values of localities obtained from the GeoPy API

In the below map of Bangalore, we plot the different localities color coded according to their cluster groups.
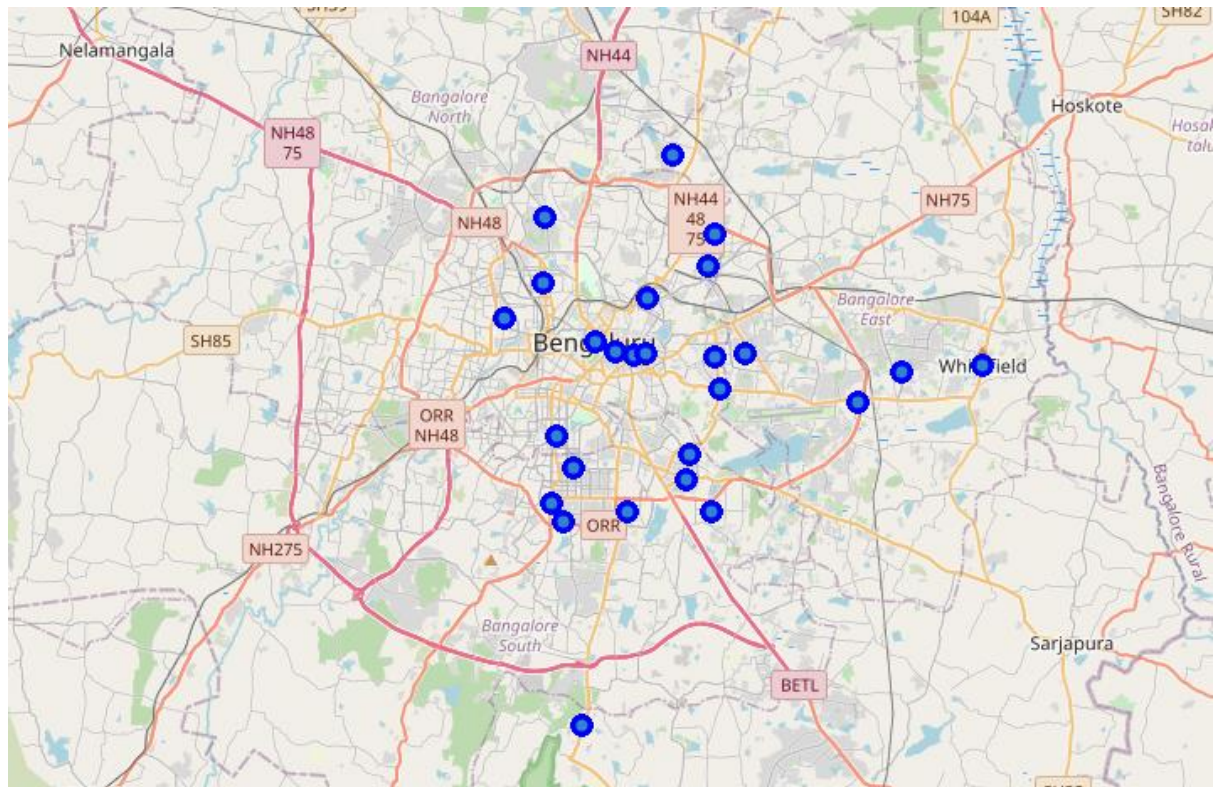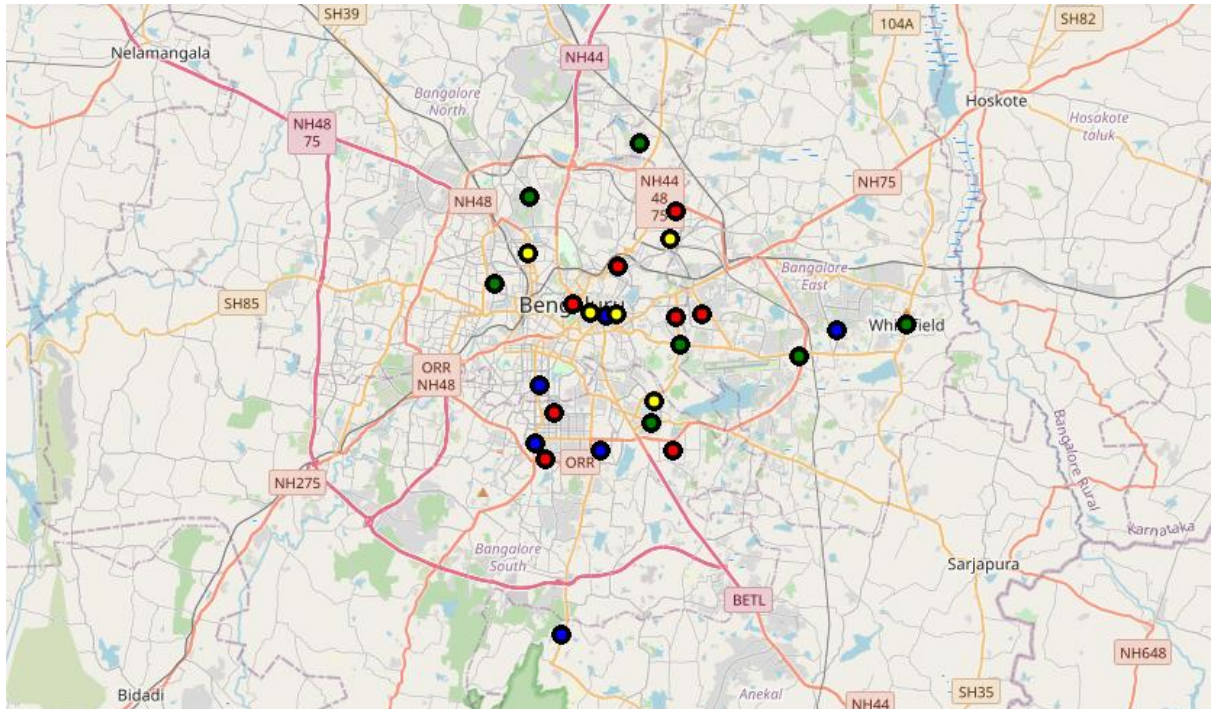
Map of Bangalore showing various localities clustered by Restaurants

In the below map of Bangalore, we plot the different localities color coded according to their cluster groups based on ratings.
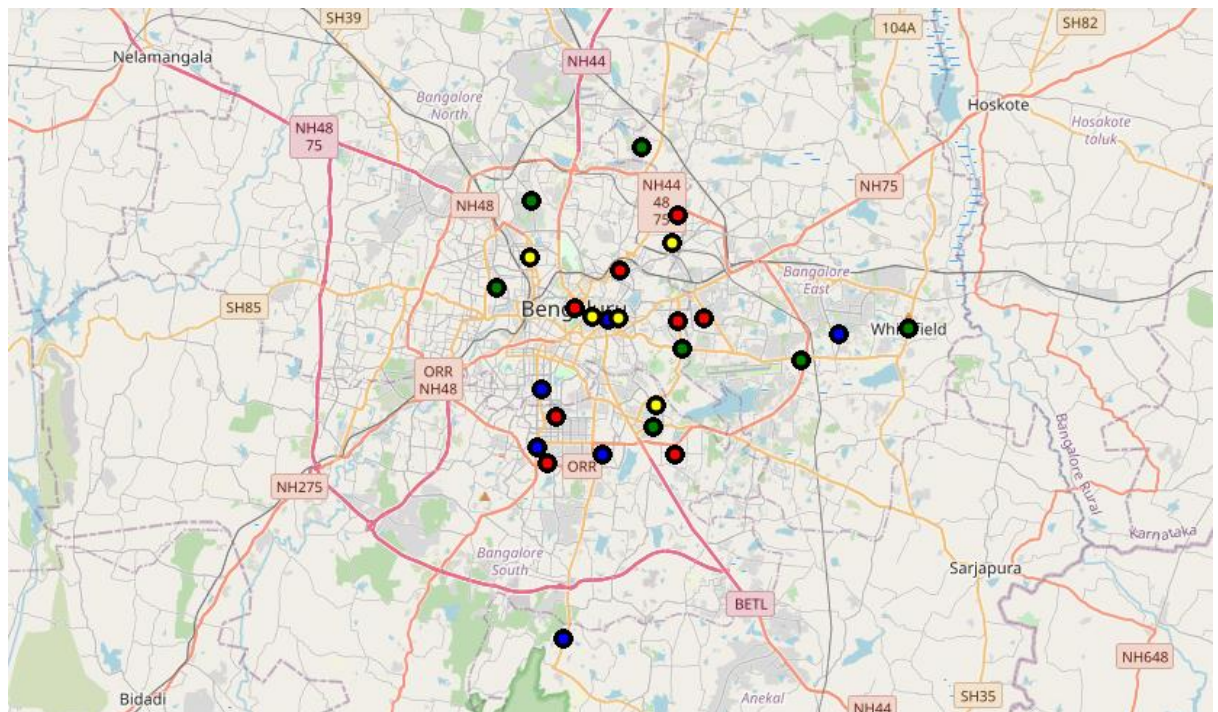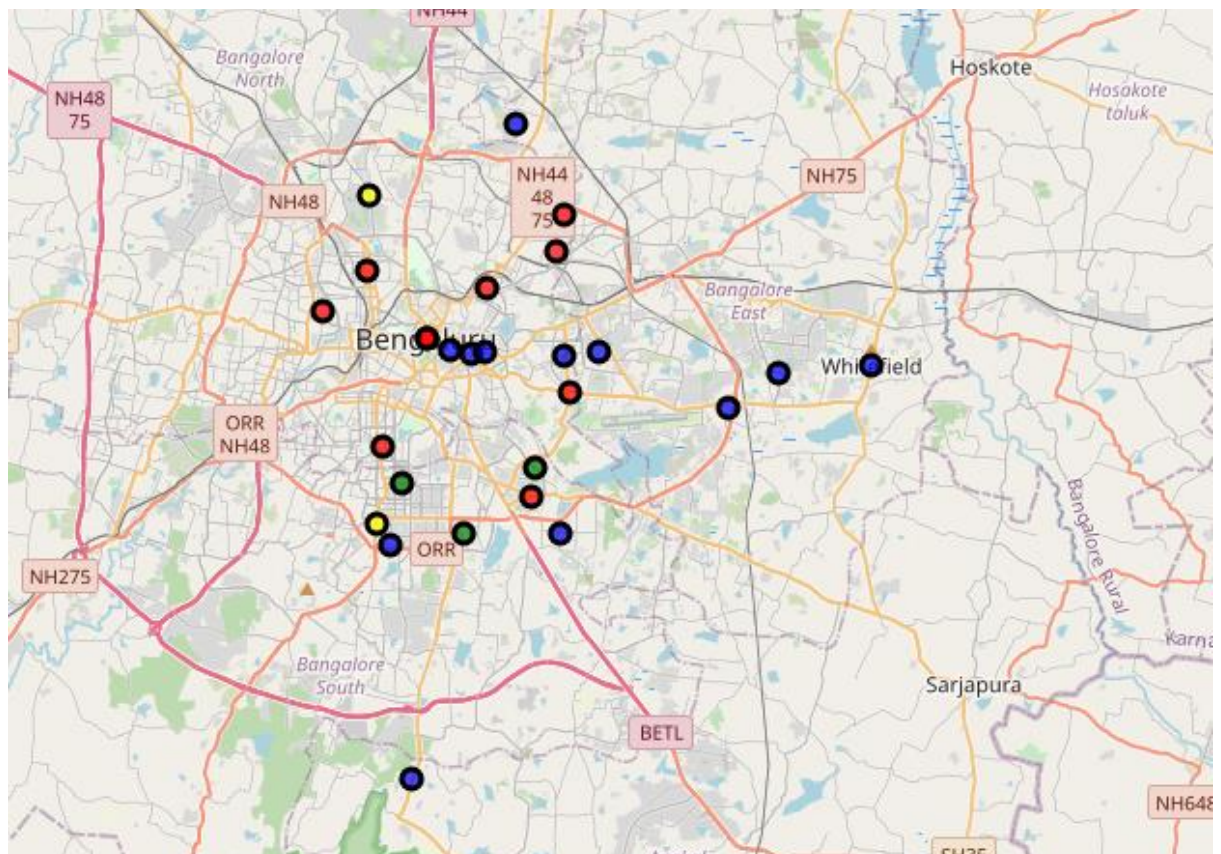
4.Results

**Based on Ratings: Localities having the best restaurants**

Based on Number of restaurants: Localities having more Restaurants

FourSquare Venue details:

	Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BTM	12.911276	77.604565	Ginger Tea House	12.912498	77.606373	Bakery
1	BTM	12.911276	77.604565	Ashirwad Departmental Store	12.915237	77.606057	Department Store
2	BTM	12.911276	77.604565	Cafe Coffee Day	12.912062	77.606156	Café
3	BTM	12.911276	77.604565	Flavours of China	12.913776	77.607662	Chinese Restaurant
4	BTM	12.911276	77.604565	Baskin-Robbins	12.913697	77.607022	Ice Cream Shop

Neighbourhoods:
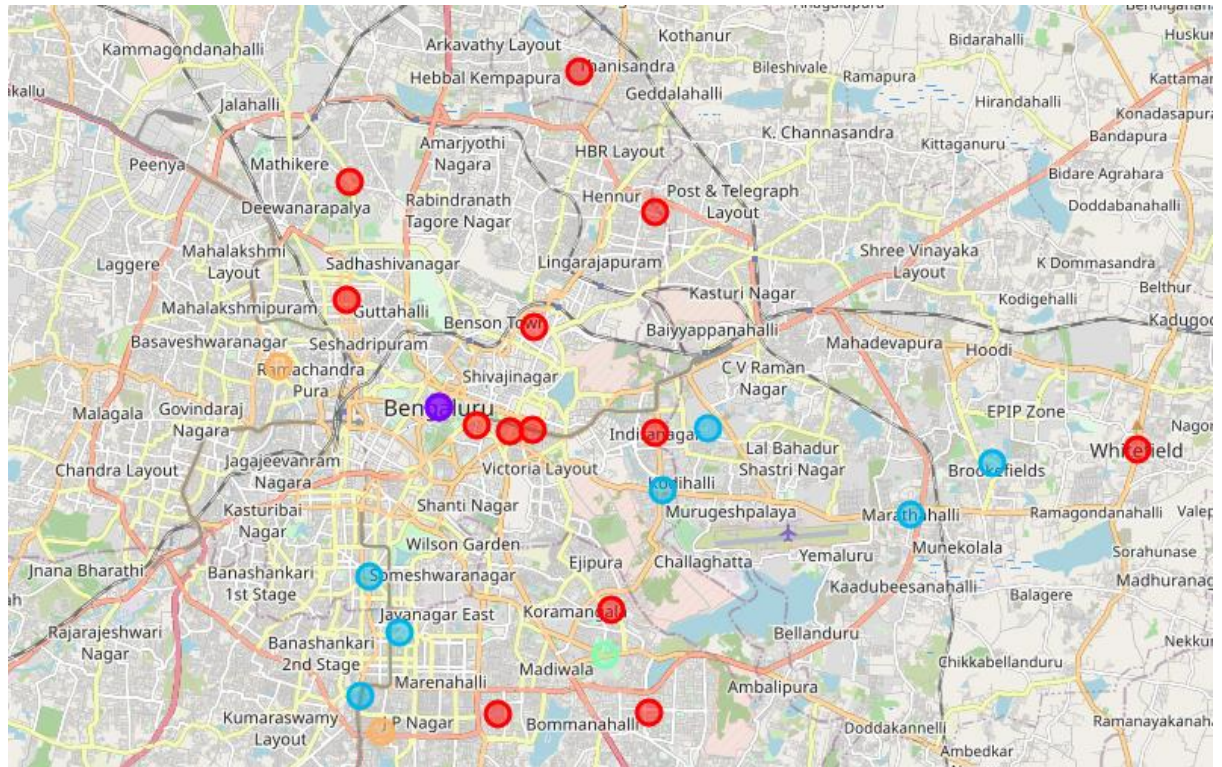
5.Discussion

The aim of this project is to help people who want to relocate to the Bangalore City or who wants to start a new Restaurant business with a specific cuisine to identify the best localities based on the total restaurants, locality, reviews and neighbourhoods. For example if a person is looking for a neighborhood with good restaurants then we can see that Clusters 1 and 3 have Indian Restaurants and Chinese Restaurants as the most common venues. If a person is looking for a neighborhood with Parks in a close proximity then the neighborhoods in the 2 and 4th clusters is suitable. For a person who wants to start new restaurant business then I feel that the neighborhoods in Cluster 1 and 3 are more suitable dues to the more common venues in that clusters. From the Analysis and Clustering, Best Places are suitable for edible person in Bangalore city are BTM, Koramangala and Jayanagar; Best restaurant Localities in Bangalore City are MG Road, Malleshwaram and Brigade Road; Best locations to stay if someone likes Chinese

cuisine are BTM and Koramangala; The best places to start for Andhra restaurant in Bangalore city are Lavelle Road and Residency Road.

6.Conclusion

This project helps a person get a better understanding of the neighborhoods with respect to the most common venues in that neighborhood. It is always helpful to make use of technology to stay one step ahead i.e. finding out more about places before moving into a neighborhood or starting a new business in new place. We have just taken Restaurants as a primary concern to shortlist the borough of Bangalore. The future of this project includes taking other factors such as cost of food and rent of the area in the areas into consideration to shortlist the borough based on selected cuisine and a predefined budget.

- Places are suitable for edible person in Bangalore city: BTM, Koramangala and Jayanagar

- Places are not suitable for edible person in Bangalore city: New BEL Road & Banashankari

- Best restaurant Localities in Bangalore City: MG Road, Malleshwaram and Brigade Road

- Low rating restaurant Localities in Bangalore City: Electronic City, Bannerghatta Road

- Best locations to stay if someone likes Chinese cuisine are BTM and Koramangala

- The best places to start for Andhra restaurant in Bangalore city: Lavelle Road and Residency Road

- The best places to start for Chinese restaurant in Bangalore city: Malleshwaram and Rajaji Nagar

- Best locations to stay if someone likes Andhra cuisine are Marathahalli and BTM

- The cluster one is the biggest cluster with 13 of the 25 neighborhoods in the Bangalore city. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Indian Restaurants, Chinese Restaurant, Cafe, and Ice Cream Shops.

- The second cluster has two neighborhoods which consists of Venues such as Park, Vineyard, Sports Bar, Capital Building and Dessert Shops.

- The third cluster has Seven neighborhoods which consists of Venues such as Indian Restaurant, Chinese Restaurants, and Fast Food shops.

- The fourth cluster has one neighborhood which consists of Venues such as Park, Bookstore, Spa and Stores.

- The fifth cluster has two neighborhoods which consists of Venues such as Department Store, Bakery, Busstation and Restaurants.