Analysing and Predicting an optimal location for an Italian Cuisine in Bangalore City

Harshith BS

March 09, 2021

1. Introduction

1.1 Background

Bangalore (officially known as Bengaluru) is the capital city and the largest city of the Indian State of Karnataka. It has a population of more than 8 million and a metropolitan population of around 11 million, making it the third most populous city and fifth most populous urban agglomeration in India. Located in the Southern India on the Deccan Plateau, at a height of over 900m (3000ft) above sea level, Bangalore is known for its pleasant climate throughout the year, its elevation is the highest among the major cities of India.

The diversity of the cuisine available is reflective of the social and economic diversity of Bengaluru. South Indian, North Indian, Arabic food, Italian, American, Chinese are all very popular in the city. Our focus will be on Italian restaurants and to predict a suitable location for an Italian Restaurant.

From being an unfamiliar food segment in Indian industry, Italian food has gone through many things which finally resulted in marking their presence in the market. Earlier people were just familiar about the native flavours and tastes to which their taste bud was adapted. But with globalization in the food industry, things have changed eventually for this industry.

Earlier chefs and recipes were imported from Italy as there was no one who could bring out that very taste in serving the people of our country. It eventually took a lot of time for our people in adapting to this taste and flavour which caused many outlets and restaurants to close down. But now as the trend has changed, Italian food products are one popular segment towards which people are really inclined and Italian has become the 2nd favourite International cuisine in the country.

1.2 Problem Definition

In this project, we will try to visualize all major parts of the Bangalore City and try to predict an optimal location for an Italian restaurant. Since there are lots of restaurants in Bangalore we try to detect locations that are not already crowded with restaurants in vicinity. We would also prefer locations as close to city center as possible.

Specifically, this report will target stakeholders who wants to set up an Italian restaurant in Bangalore, Karnataka, India.

We will use our data science powers to generate a few most promising neighbourhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2. Data Acquisition:

For the analysis we need the following data:

1. Bangalore Restaurants data that contains the details of locality, restaurant name, ratings along with their location details.

Data Source: ZOMATO_KAGGLE_DATASET

Data description: This data contains all the necessary information for our analysis. It consists of restaurants data from different countries including India, Sri Lanka, Brazil, Indonesia, New Zealand and USA. Each country has its own unique "Country Code". For example "Country Code for India is 1". Each country consists data of its cities popular restaurants. So for India there are total of 8652 restaurants from various cities of the country. Since our focus is on Bangalore city, we need to extract data particular to Bangalore city.

The above data consists of following features Restaurant ID, Restaurant Name, Country Code, City, Address, Locality, Locality Verbose, Longitude, Latitude, Cuisines, Average cost for two, Currency, Has table Booking, Has online delivery, Is delivering now, Switch to order menu, Price range, Aggregate rating, Rating colour, Rating text, votes.

Among these features we need to select a suitable features for our analysis. Our feature selection is limited to Restaurant Name, Locality, Longitude, Latitude, Cuisines, Aggregate Rating, Rating Text, Votes. A Sample of the data is shown below.

	url	address	name	online_order	book_table	rate	votes	phone	location	rest_type
0	https://www.zomato.com/bangalore/jalsa-banasha...	942, 21st Main Road, 2nd Stage, Banashankari, ...	Jalsa	Yes	Yes	4.1/5	775	080 42297555\r\n+91 9743772233	Banashankari	Casual Dining
1	https://www.zomato.com/bangalore/spice-elephan...	2nd Floor, 80 Feet Road, Near Big Bazaar, 6th ...	Spice Elephant	Yes	No	4.1/5	787	080 41714161	Banashankari	Casual Dining
2	https://www.zomato.com/SanchurroBangalore?cont...	1112, Next to KIMS Medical College, 17th Cross...	San Churro Cafe	Yes	No	3.8/5	918	+91 9663487993	Banashankari	Cafe, Casual Dining
3	https://www.zomato.com/bangalore/addhuri-udupi...	1st Floor, Annakuteera, 3rd Stage, Banashankar...	Addhuri Udupi Bhojana	No	No	3.7/5	88	+91 9620009302	Banashankari	Quick Bites
4	https://www.zomato.com/bangalore/grand-village...	10, 3rd Floor, Lakshmi Associates, Gandhi Baza...	Grand Village	No	No	3.8/5	166	+91 8026612447\r\n+91 9901210005	Basavanagudi	Casual Dining

2. To get nearby places in each locality of Bangalore city:

Data Source: FOURSQUARE.API

Data description: This API allows us to get information about all the venues in the neighbourhood of Bangalore City. For example to get the top venues near the neighbourhood of Bangalore like Kormangala, we can use the Foursquare Credentials to access the data.

3. Methodology

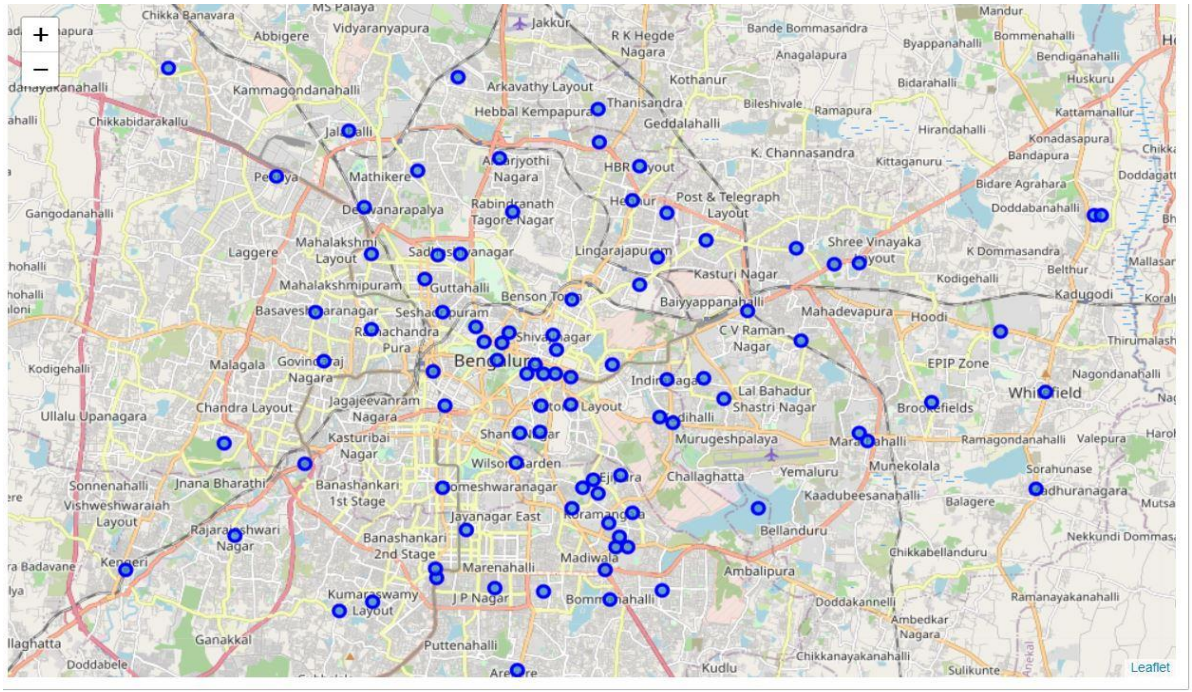
Firstly, we need to get the list of neighbourhoods in the city of Bangalore. This data is available from the Zomato Kaggle Dataset.

From this dataset we will acquire the list of names of the Neighbourhood's. We need to get the geographical coordinates in the form of latitude and longitude with the help of Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert the address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas as shown below.

	Neighborhood	Latitude	Longitude
0	Banashankari	12.915219	77.573621
1	Basavanagudi	12.941726	77.575502
2	Mysore Road	12.948657	77.535702
3	Jayanagar	12.929273	77.582423
4	Kumaraswamy Layout	12.908149	77.555318
...	...	...	...
87	Sahakara Nagar	13.062147	77.580061
88	Jalahalli	13.046453	77.548380
89	Nagarbhavi	12.954674	77.512172
90	Peenya	13.032942	77.527325
91	KR Puram	13.007516	77.695935

92 rows × 3 columns

DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinate's data returned by Geocoder are correctly plotted in the city of Bangalore. Next, we will use the Foursquare API to get the top 100 venues that are within a radius of 2000 meters as shown below.



We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering.

The venues obtained by Foursqaure API is shown below.

	Neighborhood	Latitude	Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	Banashankari	12.915219	77.573621	Shivaji Military Hotel	12.917919	77.573925	Indian Restaurant
1	Banashankari	12.915219	77.573621	Corner House	12.922647	77.573560	Ice Cream Shop
2	Banashankari	12.915219	77.573621	Stoned Monkey	12.923579	77.569689	Ice Cream Shop
3	Banashankari	12.915219	77.573621	Natural Ice Cream	12.923863	77.576513	Ice Cream Shop
4	Banashankari	12.915219	77.573621	Davanagere benne dosa	12.908932	77.572983	Breakfast Spot
...	...	...	...	...	...	...	...
6462	KR Puram	13.007516	77.695935	Tandoor Box	12.992469	77.702759	BBQ Joint
6463	KR Puram	13.007516	77.695935	ABB Cafeteria	12.993985	77.705566	Cafeteria
6464	KR Puram	13.007516	77.695935	City Kitchen	13.023871	77.690793	Indian Restaurant
6465	KR Puram	13.007516	77.695935	Namdhari Fresh	12.991481	77.702810	Grocery Store
6466	KR Puram	13.007516	77.695935	Reliance Fresh	13.022554	77.696372	Convenience Store

6467 rows × 7 columns

Since we are analysing the “Italian Restaurant” data, we will filter the “Italian Restaurant” as venue allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 4 clusters based on their frequency of occurrence for “Italian Restaurant”. The results will allow us to identify which neighbourhoods have a higher concentration of Italian Restaurant while which neighbourhoods have a fewer number of Italian Restaurant. Based on the occurrence of Italian Restaurant in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new restaurants.

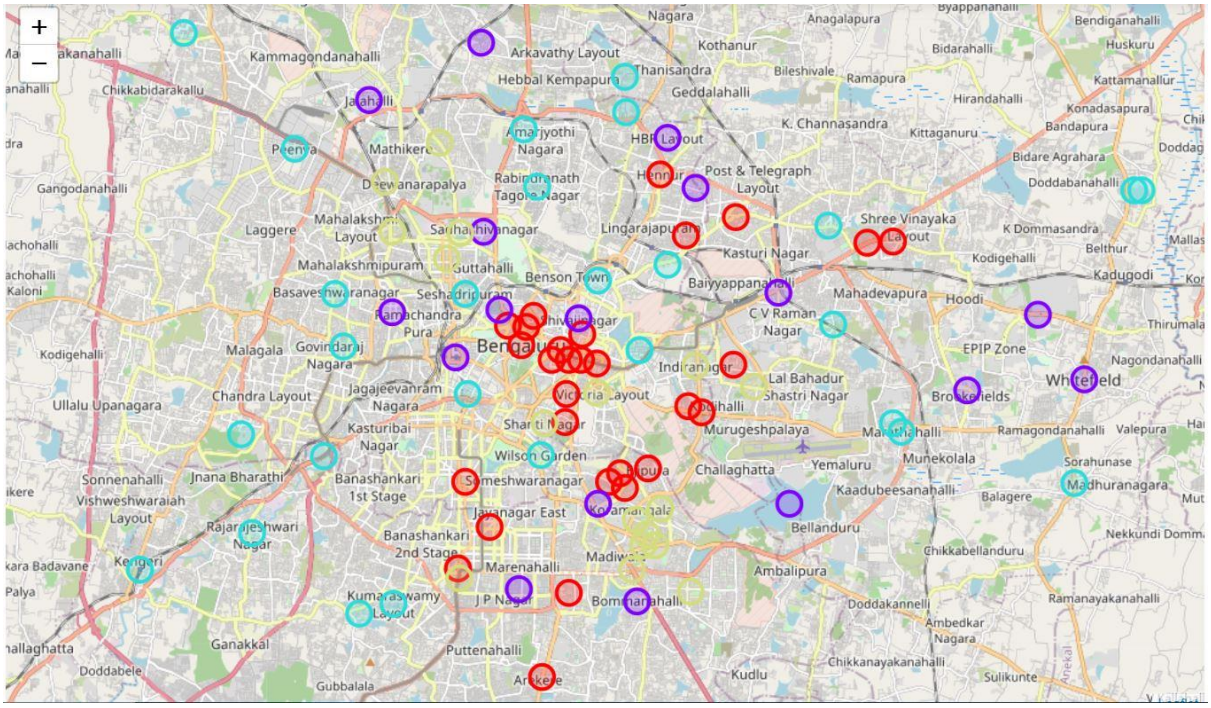
5. Analysis

Now we need to cluster all the neighbourhoods into different clusters. The results will allow us to identify which neighbourhoods have a higher concentration of Italian Restaurants while which neighbourhoods have a fewer number of Italian Restaurants. Based on the occurrence of Italian Restaurants in different neighbourhoods, it will help us answer the question as to which neighbourhoods are most suitable to open new Italian Restaurants. We set the number of clusters to 4 and run the algorithm. After applying the K-Means clustering algorithm, all the neighbourhoods get segregated and form different clusters. The resulted output is as shown below.

	Neighborhood	Italian Restaurant	Cluster Labels
0	BTM	2	0
1	Banashankari	3	3
2	Banaswadi	2	0
3	Bannerghatta Road	2	0
4	Basavanagudi	2	0
...	...	...	...
86	West Bangalore	0	2
87	Whitefield	1	1
88	Wilson Garden	0	2
89	Yelahanka	0	2
90	Yeshwantpur	4	3

91 rows × 3 columns

The result obtained from the K-Means clustering algorithm is used to create map using different markers for each cluster. Four clusters are obtained based on the algorithm. The red circles indicates cluster-1, the blue circle indicates cluster-2, the mint green colour indicates cluster-3 and the yellow colour indicates cluster-3. The map created using folium is shown below.



6. Results

The results from the K-means clustering show that we can categorize the neighbourhoods into 4 clusters based on the frequency of occurrence for “Italian Restaurant”. The four clusters along with their neighbourhoods are given below:

1. Cluster-1: In this cluster it contains neighbourhoods with 2 Italian Restaurants. The neighbourhoods present in this cluster are, BTM, Infantry Road, Jayanagar, KR Puram, Kammanahalli, Thippasandra, Koramangala 5th Block, Koramangala 6th Block, Hennur, Koramangala 8th Block, Lavelle Road, MG Road, St. Marks Road, Old Airport Road, Old Madras Road, Race Course Road, Richmond Road, Langford Town, Electronic City, Kanakapura Road, Commercial Street, Church Street, Brigade Road, Banaswadi, Domlur, Bannerghatta Road, Cunningham Road, Basavanagudi, Ejipura. Hence this cluster consists total of 29 places.

2. Cluster-2: In this cluster it contains neighbourhoods with 1 Italian Restaurant. The Neighbourhoods present in this cluster are, CV Raman Nagar, Whitefield, Koramangala 1st Block, Bommanahalli, Bellandur, Kalyan Nagar, Rajajinagar, Jalahalli, JP Nagar, Vasanth Nagar, “ITPL Main Road, Whitefield”, Sadashiv Nagar, Sahakara Nagar, Shivajinagar, HBR Layout, Majestic, Brookefield. Hence there are total of 17 places in this cluster.

3. Cluster-3: In this cluster it contains neighbourhoods with 0 restaurants. The Neighbourhoods present in this cluster are, North Bangalore, Frazer Town, RT Nagar, South Bangalore, Rajarajeshwari Nagar, Ramamurthy Nagar, Residency Road, Wilson Garden, Sanjay Nagar, Peenya, Nagawara, Kumaraswamy Layout, Mysore Road, West Bangalore, Hebbal, East Bangalore, “Varthur Main Road, Whitefield”, Kaggadasapura, City Market, Nagarbhavi, Seshadripuram, Uttarahalli, Ulsoor, Yelahanka, Magadi Road, Basaveshwara Nagar, Marathahalli, Kengeri. Hence there are total of 28 places in this cluster.

4. Cluster-4: In this cluster it contains neighbourhoods with 3 or more restaurants. The Neighbourhoods present in this cluster are, Vijay Nagar, Shanti Nagar, Koramangala 7th Block, Sankey Road, New BEL Road, Malleshwaram, Koramangala 4th Block, Koramangala 3rd Block, Koramangala 2nd Block, Jeevan Bhima Nagar, Indiranagar, Hosur Road, HSR, Central Bangalore, Banashankari, Sarjapur Road,Yeshwantpur. Hence there are total of 17 places in this cluster.

These clusters helps us to understand the neighbourhoods based on the frequency of Italian Restaurants. It also gives us the idea of which places is crowded with Italian restaurants with more competition and also helps us to identity places with moderate and least competiton.

7. Conclusions

Italian Restaurants are spread out throughout the city as we can visualize from the map. The moderate number of restaurants are concentrated at the central part of the city. As we can see the neighbourhoods present in the cluster-0, which consists of 2 Italian Restaurants are located at the central part of the city. The neighbourhoods which consists of maximum number of Italian restaurants are present in cluster-3, where the competition is very high. The neighbourhoods from the cluster-1 consisting of exactly 1 restaurant with moderate competition. Also there are neighbourhoods with zero Italian Restaurants, as we can see from cluster-2, also some places from this cluster are at the central part of the city.

Therefore, this project recommends restaurants owners to capitalize on these findings to open Italian Restaurants in neighbourhoods in cluster 2 where there is no competition. Restaurants owners who can really stand out and provide good quality Italian foods can also consider neighbourhoods in cluster-1 as there is moderate competition.

So we can apply the same approach for large datasets and can easily distinguish the venues based on the category. Suppose if there are 400 Chinese Restaurants in a city then we can easily segregate them into different clusters. We can apply this method not only for restaurants but shopping malls, movie theatre and much more. In this project, we only consider one factor i.e. frequency of occurrence of Italian Restaurant, there are other factors such as population and income of residents that could influence the location decision of a new Italian Restaurants.

But for setting up a restaurant we need to consider other factors such as the cost of rent, the surroundings around the restaurants, the kind of people in the locality, if it's a luxurious area many people prefer going out, their lifestyle will be different from others and therefore spend a lot. If we decide a place where the competition is less, then we need to consider the people living in that locality as well. If the people in that area spend a lot and love going out for food, then it’ll be a success. If the people staying near the restaurants don't prefer going out, then it's better to consider some other place with less competition and a good crowd.