

Capstone Project- The Battle of Neighborhoods

Vinodhini V.

10 – Aug – 2021

1. Introduction

For someone to start-up a chain of coffee shops or a single unit, enough market research should be carried out to identify the suitable location and time for start of the business. Market research is a wide area of analytics that requires one to collect, analyze and interpret the data based on customer demographics, competitors, overall industry trends.

Business problem here is to be able to find the best or most suitable locations in Bangalore's Neighborhood for start of a coffee-shop by researching on the competitors.

Business Objective:

This project will aid to narrow down the locations based on the coffee-shops in the neighborhood and their ratings.

- a. Identify the best location to setup a coffee shop in Bangalore based on location proximity and ratings of existing shops in the neighborhood.

2. Data Extraction and Cleaning

Multiple types of data are required for this analysis.

- a. List of Neighborhoods in Bangalore -
https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore
- b. Latitude and Longitude details of the Neighborhoods
- c. List of coffeeshops in each neighborhood (Extracted from Foursquare via API)
- d. Ratings of the coffee shops (Extracted from Foursquare via API)
- e. Neighborhood details and their latitude and longitude information helps to fetch the details of the coffee shops in each locality.

Data Extraction:

Neighborhood Data is extracted from Wikipedia using web-scraping techniques (beautifulsoup).

Latitude and Longitude information is prepared with data available on internet as a CSV.

Coffee shop information and ratings are extracted using FourSquare API. FourSquare API takes input for Latitude and Longitude of a location and fetches the neighboring places of interest.

Sample API Query to fetch the coffee shops around :

https://api.foursquare.com/v2/venues/search?client_id=XXX&client_secret=XXXX&ll=12.9784,77.6408&oauth_token=XXX&v=20180604&query=coffee&radius=500&limit=100

FourSquare API also provides information related to tips, ratings of a particular venue.

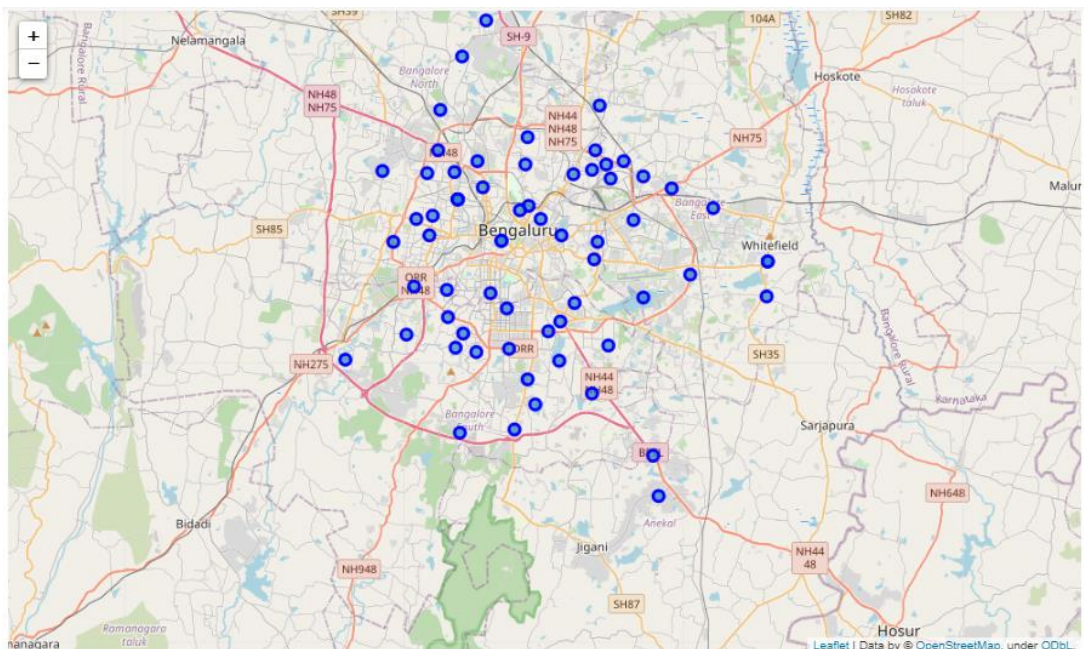
Data Clean-up/Preparation:

Any null values are treated by dropping the records.

Coffee shops with no rating are imputed suitably.

3. Exploratory Data Analysis

There are about 64 prominent neighborhoods in the city of Bangalore, spread on the north, south, east, and west of Bangalore. In these 64 locations, 128 coffee shops are located as found on Foursquare.



Only 37 out of 128 coffee shops has a rating listed on Foursquare. The mean of the rating is 6.3 with maximum being 7.7 and minimum as 5.1.

```
filter_df['rating'].describe()
```

```
count    37.000000
mean      6.337838
std       0.461490
min       5.100000
25%      6.300000
50%      6.300000
75%      6.400000
max       7.700000
Name: rating, dtype: float64
```

4. Methodology

The Neighborhood data is extracted from the Wikipedia page(https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore) using web-scraping technique. The required latitude and longitude data are read from the CSV.

Data clean-up is done as per steps mentioned in Section 2. Further data extraction on coffee shop and ratings is done using FourSquare APIs. Upto 100 available coffeeshops within the radius of 500 meters per location is considered. FourSquare API responds in JSON format. Pandas library is used to read CSV and JSON response and convert into the dataframe.

As the intention here is to find the segments that can show how the coffee shops are spread across, unsupervised Clustering using KMeans is the preferred approach. Clustering will be applied to understand the proximity and distribution of coffee-shops in the neighborhood. Clustering on the ratings will be done to understand how the coffee shops are distributed based on rating and locality.

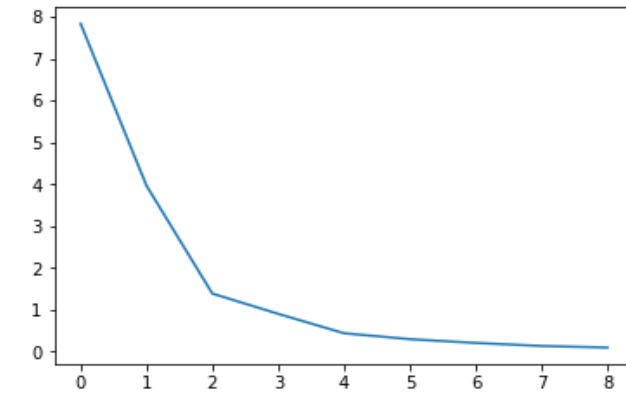
Elbow method will be applied to identify the suitable number of clusters. Validations done by plotting the output of cluster over maps using Folium and reviewed manually.

5. Results

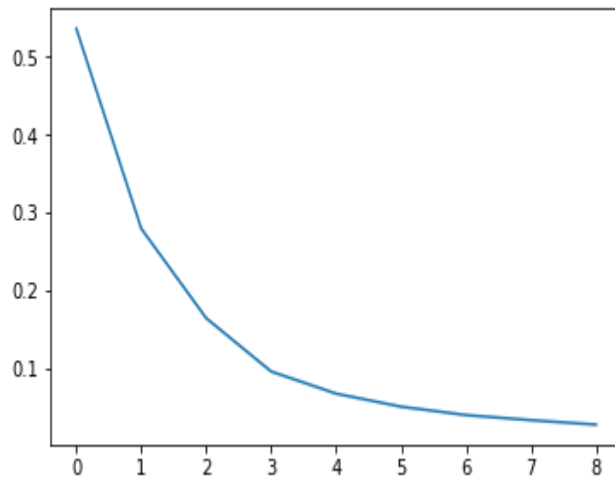
Elbow method is used to identify the optimum number of clusters. Clustering is done for two cases :

1. Cluster to show the coffee shop distributions across neighborhood.
2. Cluster to show the coffee shop distribution based on ratings

Optimum number of clusters for neighborhood proximity and rating of coffee shops was 2.



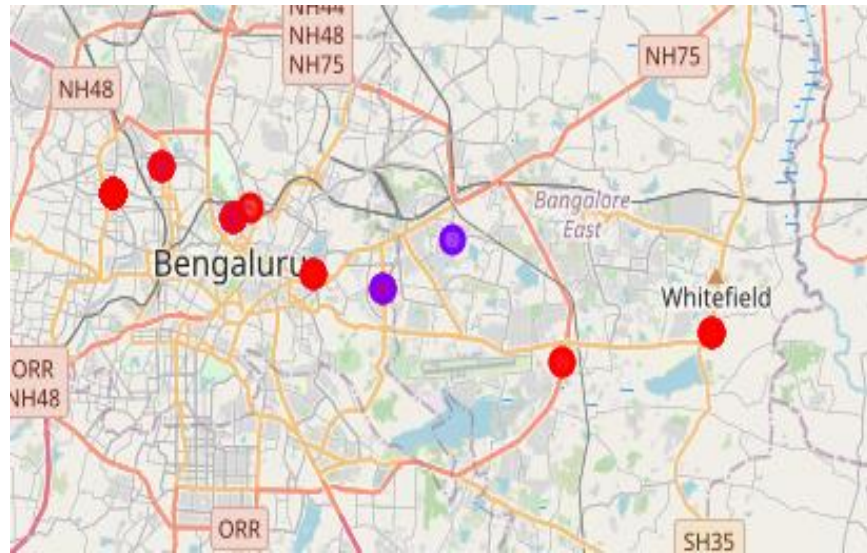
Optimum number of clusters for neighborhood proximity of coffee shops was 3.



Out of the many Coffee shops listed, only 37 have been rated. Most of the ratings hover around 5-7. Ratings of 6-7 are mostly seen around Bangalore North-West.

Cluster distribution – Coffeeshops with ratings

- Cluster 0 – 32 coffee shops [rating avg – 6.47]
- Cluster 1 – 5 coffee shops [rating avg – 5.52]

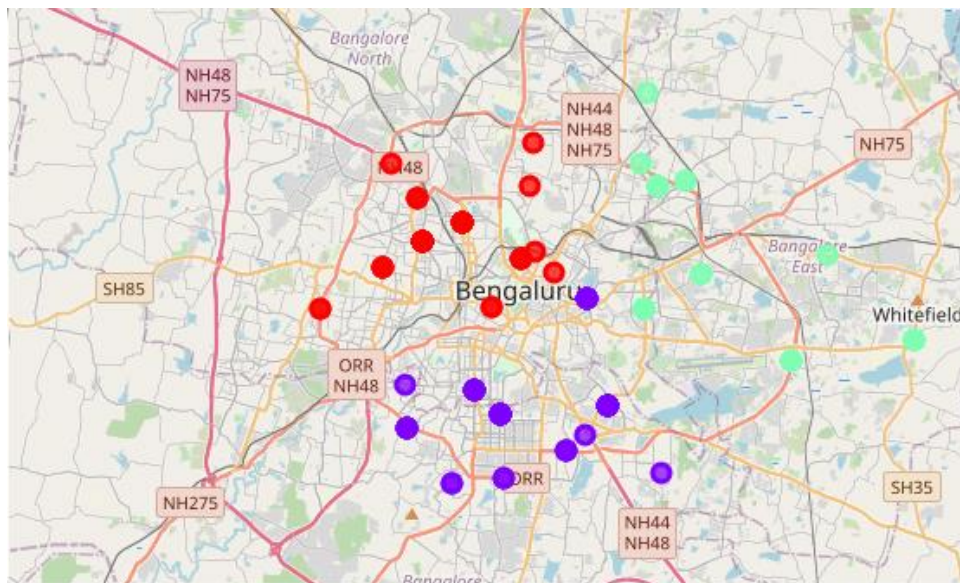


Cluster distribution – Coffeeshops in Neighborhood

Cluster 0 – 52 neighborhoods.

Cluster 1 – 51 neighborhoods

Cluster 2 – 33 neighborhoods



6. Limitations and Conclusion

Bangalore East seems to be a suitable location for new coffee shop if the business objective is to look for a less competitive environment as there is lack of sufficient and good rated coffee shops. If the business strategy is to go for more crowded place, then Bangalore Central seems to be a suitable option.

This project is restricted in analysis only on competitors and also the ratings available on a single platform (FourSquare). The market research can be further expanded to include the below aspects that can help finalize the optimal location for setting up a Coffee shop.

- population and surroundings
- ratings on other platforms for Coffee shops
- Coffee variants and other beverages offered

7. References

List of neighbourhoods in Bangalore , accessed 10 August 2021, https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Bangalore

Places API, accessed on 10 August 2021, <https://developer.foursquare.com/docs/>

QuickStart-Folium 0.12.1 documentation, accessed on 10 August 2021, <https://python-visualization.github.io/folium/quickstart.html>