

Capstone Project - The Battle of Neighborhoods

Bangalore, Karnataka

Introduction

Business Problem & Discussion

I'm going to focus on problems one might face while opening a new restaurant. Well there are many factors to be considered to open a small or big level business. First and foremost, important point to decide is the location for her new restaurant. On what basis one can decide her restaurant's location? Does he want to rent place or can he afford to buy it? While selecting the place there are key points to consider like he needs to check out how many restaurants are there in that specific location? What all factors will help one to run her business above average?

Let's discuss the above-mentioned problem statements. Firstly, one needs to choose a suitable location for his/her restaurant. To open a new Restaurant in Bangalore. Bangalore is a famous tourist place, so assuming that one wants to rent a place for the new restaurant. And now he needs to figure out how many restaurants are there in say neighborhood A, B, C, etc. If there are more than two restaurants in a neighborhood then that would be a great risk to open new restaurant of same cuisine in that neighborhood. Selecting a place where there is less or no restaurant would be of great choice, considering the rent of neighborhood too. one needs to look for a place where many people frequently visit so that business is above average. Places like Downtown, Movie theatre, Malls & Gas stations would help her business running. Restaurants ratings, check-in of customers might help in deciding location crowd. I would also suggest that one should check for opening and closing timings of other restaurants.

Data Sets and APIs

Data Set:

The data acquired from Wikipedia pages and restructure to csv file for easier manipulation and reading. The dataset consist of Neighborhood, postal code and district name.

Foursquare API

This API has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source. Many organizations are

using to geo-tag their photos with detailed info about a destination, while also serving up contextually relevant locations for those who are searching for a place to eat, drink or explore. This API provides the ability to perform location search, location sharing and details about a business. Foursquare users can also use photos, tips and reviews in many productive ways to add value to the results.

Work Flow

HTTP requests would be made to this Foursquare API server using zip codes of the Bangalore city neighborhoods to pull the location information (Latitude and Longitude). Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.

Folium-

Python visualization library would be used to visualize the neighborhoods cluster distribution of Bangalore city over an interactive leaflet map. Extensive comparative analysis of two randomly picked neighborhoods would be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods. These clusters from each of those two chosen neighborhoods would be analyzed individually collectively and comparatively to derive the conclusions.

Python packages and Dependencies

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

Data wrangling

```
[3]: #read csv file contain Bangalore data
#df_mum = pd.read_csv('Bangalorepostalcodes1.csv')
#df_mum.head()
df_mum = pd.read_excel('Bangalorepostalcodes.xlsx')
df_mum.head()
#df_mum.shape
```

```
[3]:
```

	Location	Pincode	State	District
0	A F station yelahanka	560063	Karnataka	Bangalore
1	Adugodi	560030	Karnataka	Bangalore
2	Agara	560034	Karnataka	Bangalore
3	Agram	560007	Karnataka	Bangalore
4	Air Force hospital	560007	Karnataka	Bangalore

Adding Geographical coordinates of the neighborhoods

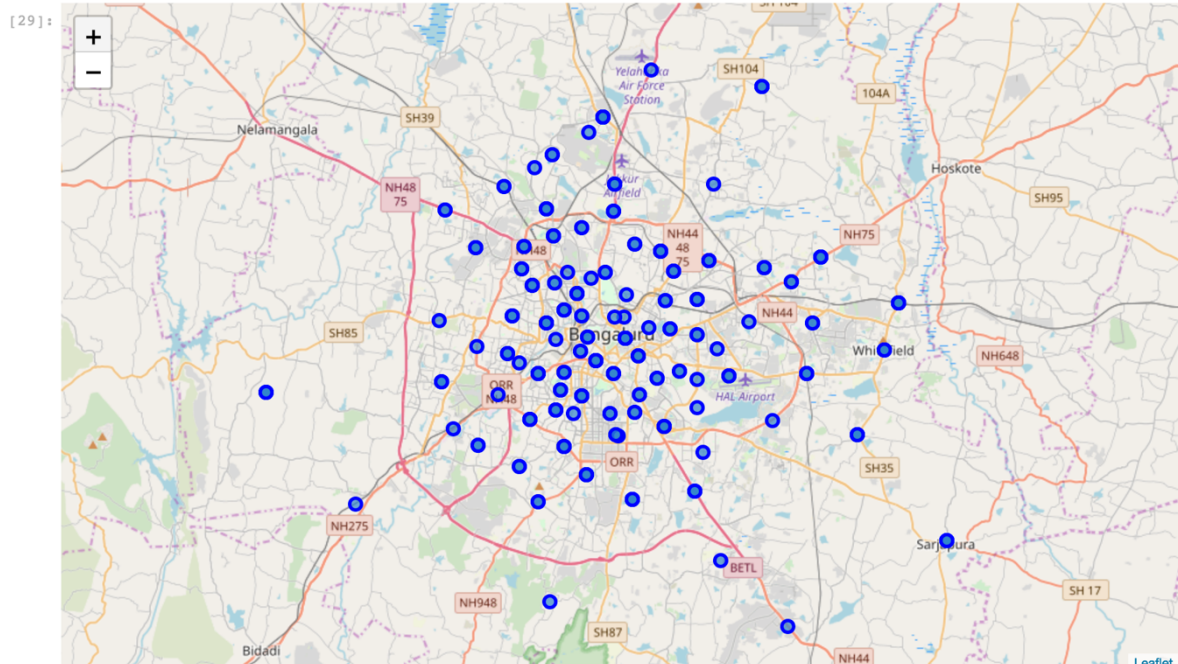
Geographical coordinates obtained using google and added to dataframe.

```
[6]:
```

	Location	Pincode	State	District	Latitude	Longitude
0	A F station yelahanka	560063	Karnataka	Bangalore	13.129087	77.614226
1	Adugodi	560030	Karnataka	Bangalore	12.944150	77.607623
2	Agara	560034	Karnataka	Bangalore	12.926138	77.622109
3	Agram	560007	Karnataka	Bangalore	12.957917	77.630912
4	Air Force hospital	560007	Karnataka	Bangalore	12.957917	77.630912

Bangalore Neighborhoods in Folium Map

```
[29]: map_bangalore
```



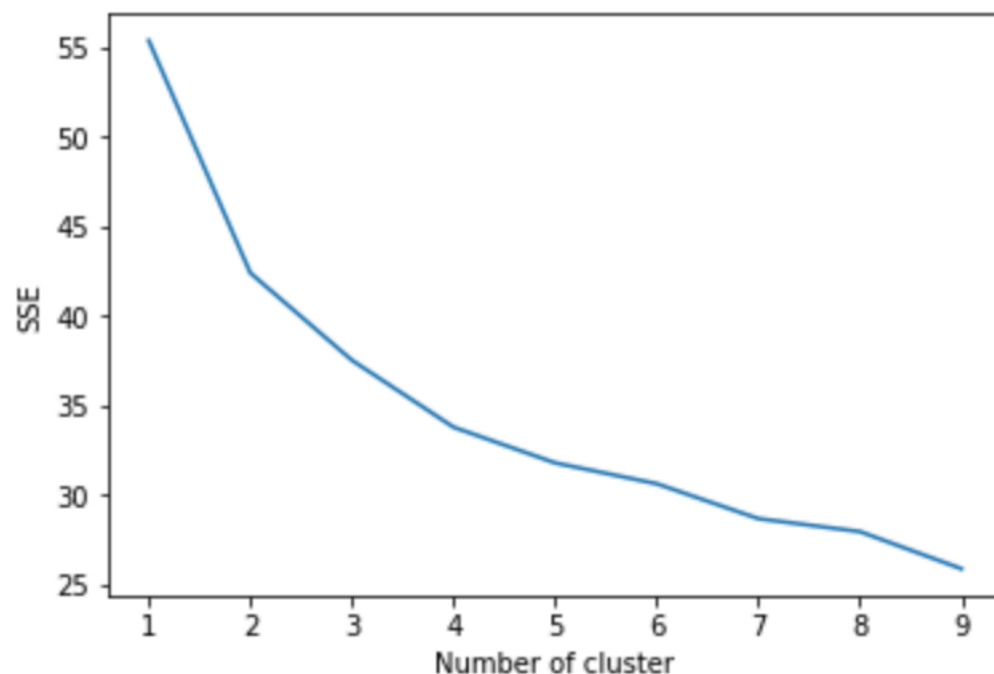
Common venues for each neighborhood in Bangalore

[47] :

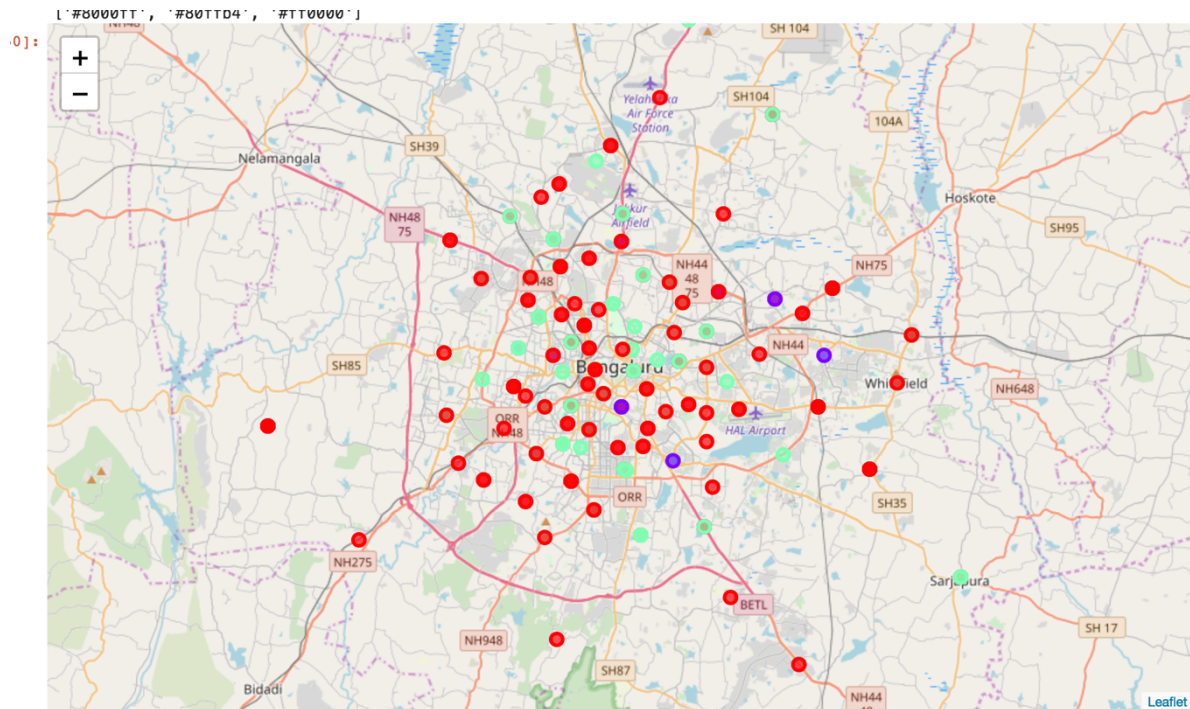
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	A F station yelahanka	Vegetarian / Vegan Restaurant	Lake	Dessert Shop	Diner	Farmers Market	Event Space	English Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant
1	Adugodi	Fast Food Restaurant	Ice Cream Shop	Restaurant	Bus Station	Design Studio	Furniture / Home Store	Cricket Ground	Cupcake Shop	English Restaurant	Electronics Store
2	Agara	Indian Restaurant	Fast Food Restaurant	Restaurant	Café	Halal Restaurant	Ice Cream Shop	Department Store	Bus Station	Furniture / Home Store	American Restaurant
3	Agram	Vegetarian / Vegan Restaurant	Snack Place	Breakfast Spot	Bakery	Mobile Phone Shop	Wine Shop	Dog Run	Event Space	English Restaurant	Electronics Store
4	Air Force hospital	Vegetarian / Vegan Restaurant	Snack Place	Breakfast Spot	Bakery	Mobile Phone Shop	Wine Shop	Dog Run	Event Space	English Restaurant	Electronics Store

Elbow Criterion Method

The idea behind elbow method is to run k-means clustering on a given dataset for a range of values of k and for each value of k, calculate sum of squared errors (SSE). After that, plot a line graph of the SSE for each value of k.



Neighborhood clusters using K-mean



Neighborhood venues

[67] :

	Location	Brigade Road	Lalbagh West
	Pincode	560001	560004
	State	Karnataka	Karnataka
	District	Bangalore	Bangalore
	Latitude	12.9766	12.9438
	Longitude	77.5993	77.5738
	Cluster Labels	2	0
	1st Most Common Venue	Lounge	Indian Restaurant
	2nd Most Common Venue	Indian Restaurant	Café
	3rd Most Common Venue	Italian Restaurant	Fast Food Restaurant
	4th Most Common Venue	Mexican Restaurant	Sandwich Place
	5th Most Common Venue	Ice Cream Shop	Ice Cream Shop
	6th Most Common Venue	Pub	Bakery
	7th Most Common Venue	Plaza	Restaurant
	8th Most Common Venue	Pizza Place	Smoke Shop
	9th Most Common Venue	Breakfast Spot	Farmers Market
	10th Most Common Venue	Hotel	Road

Results

Based on cluster for each neighborhood above, we believe that classification for each cluster can be done better with calculation of venues categories (most common) in neighborhoods. Referring to each cluster, we can't determine clearly what represent in each cluster by using Foursquare - Most Common Venue data.

However, for the purpose of this project we assumed each cluster as follow:

Cluster 1: Bangalore: restaurant and Tourism

Cluster 2: Bangalore: Residential and Shopping

Cluster 3: Bangalore: Residential

What is lacking at this point is a systematic, quantitative way to identify and distinguish different district and to describe the correlation most common venues as recorded in Foursquare. The reality is however more complex: similar neighborhoods might have or might not have similar common venues. A further step in this classification would be to find a method to extract these common venues and integrate the spatial correlations between different of areas or district.

Conclusion

Using Foursquare API, we can capture data of common places all around the world. Using it, we refer back to our main objectives, which is to determine

- the similarity or dissimilarity of cities
- classification of area located inside the city whether it is residential, tourism places, or others