Introduction

A startup wants to open a new chain of coffee cafes in the city of Bengaluru, India. Mr. Murthi, who is the owner of the business recently moved from USA to India and wants to start a new business. He is not much aware about the city but is sure that people love coffee.

He wants to know the areas which lack in coffee shops and the areas having tech parks. He wants to target the office goers who would want to have quality coffee in office hours..

The foursquare api will be used to identify the office areas along with the areas which have coffee shops. Then we will cluster the areas and find out our target cluster.

Data

For this project the Foursquare API will be used. A list of neighborhoods in Bangalore, India is downloaded and their respective location in longitude and latitude coordinates is obtained. The sources are the following:

- Bangalore neighborhoods: https://data.gov.in

The data downloaded are the neighborhoods located in Bangalore. Moreover, their specific coordinates are merged. A Foursquare API GET request is sent in order to acquire the surrounding venues that are within a radius of 500m. The data is formatted using one hot encoding with the categories of each venue. Then, the venues are grouped by neighborhoods computing the mean of each feature.

The similarities will be determined based on the frequency of the categories found in the neighborhoods. These similarities found are a strong indicator for a customer and can help us decide whether to open a coffee shop area or not.

Methodology

Feature Extraction

For feature extraction One Hot Encoding is used in terms of categories. Therefore, each feature is a category that belongs to a venue. Each feature becomes binary, this means that 1 means this category is found in the venue and 0 means the opposite. Then, all the venues are grouped by the neighborhoods, computing at the same time the mean. This will give us a venue for each row and each column will contain the frequency of occurrence of that category.

	Neighbourhood	Andhra Restaurant	Antique Shop	Arts & Crafts Store	Asian Restaurant	Auto Garage	BBQ Joint	Bakery	Bar	Basketball Court	...
0	Arabic College	0	0	0	0	0	0	0	0	0	...
1	Arabic College	0	0	0	0	0	0	0	0	0	...
2	Arabic College	0	0	0	0	0	0	0	0	0	...
3	Bellandur	0	0	0	0	0	0	0	0	0	...
4	Bellandur	0	0	0	0	0	0	0	0	0	...

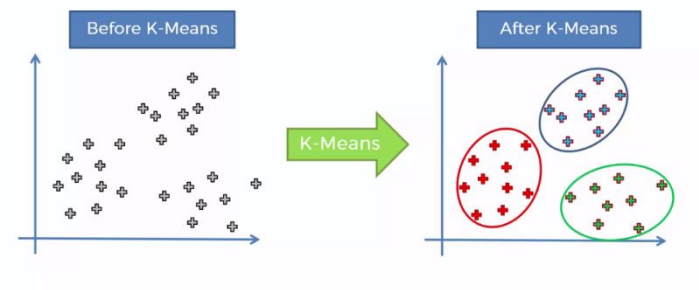
Unsupervised Learning

For the purpose of doing unsupervised learning to found similarities between neighborhoods, a clustering algorithm is implemented. In this case K-Means is used due to its simplicity and its similarity approach to found patterns.

- **K-Means:**

K-Means is a clustering algorithm. This algorithm search clusters within the data and the main objective function is to minimize the data dispersion for each cluster. Thus, each group found represents a set of data with a pattern inside the multidimensional features.

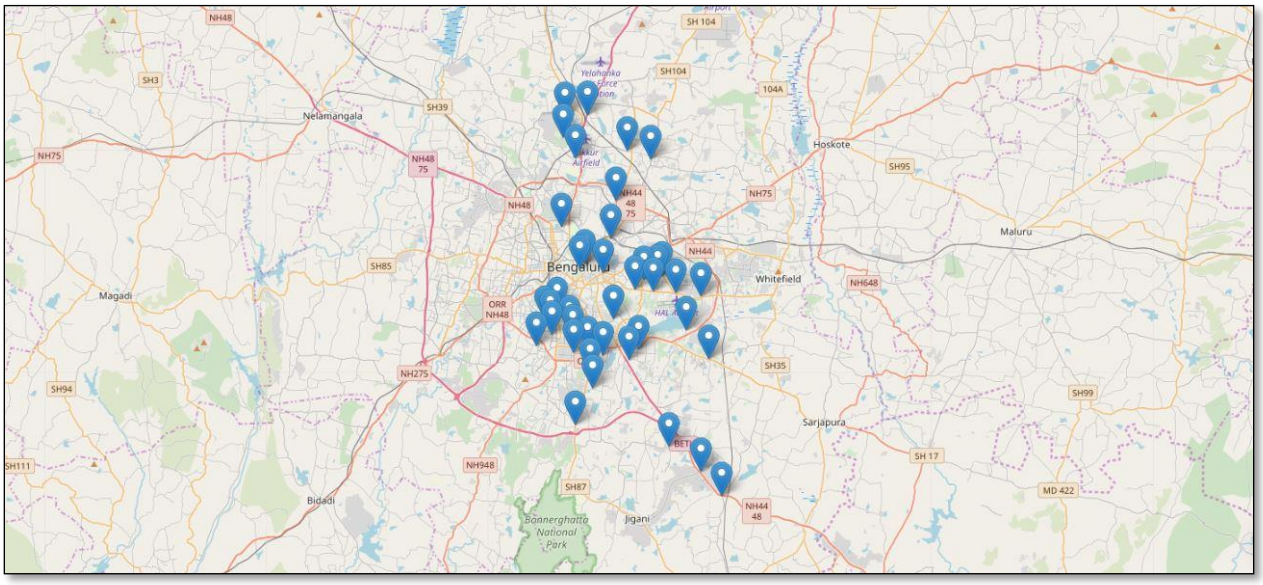
In the following figure there is a graphical example of how a K-Means algorithm works. As it is possible to see, dispersion is minimized by representing all clustered data into one group or cluster.



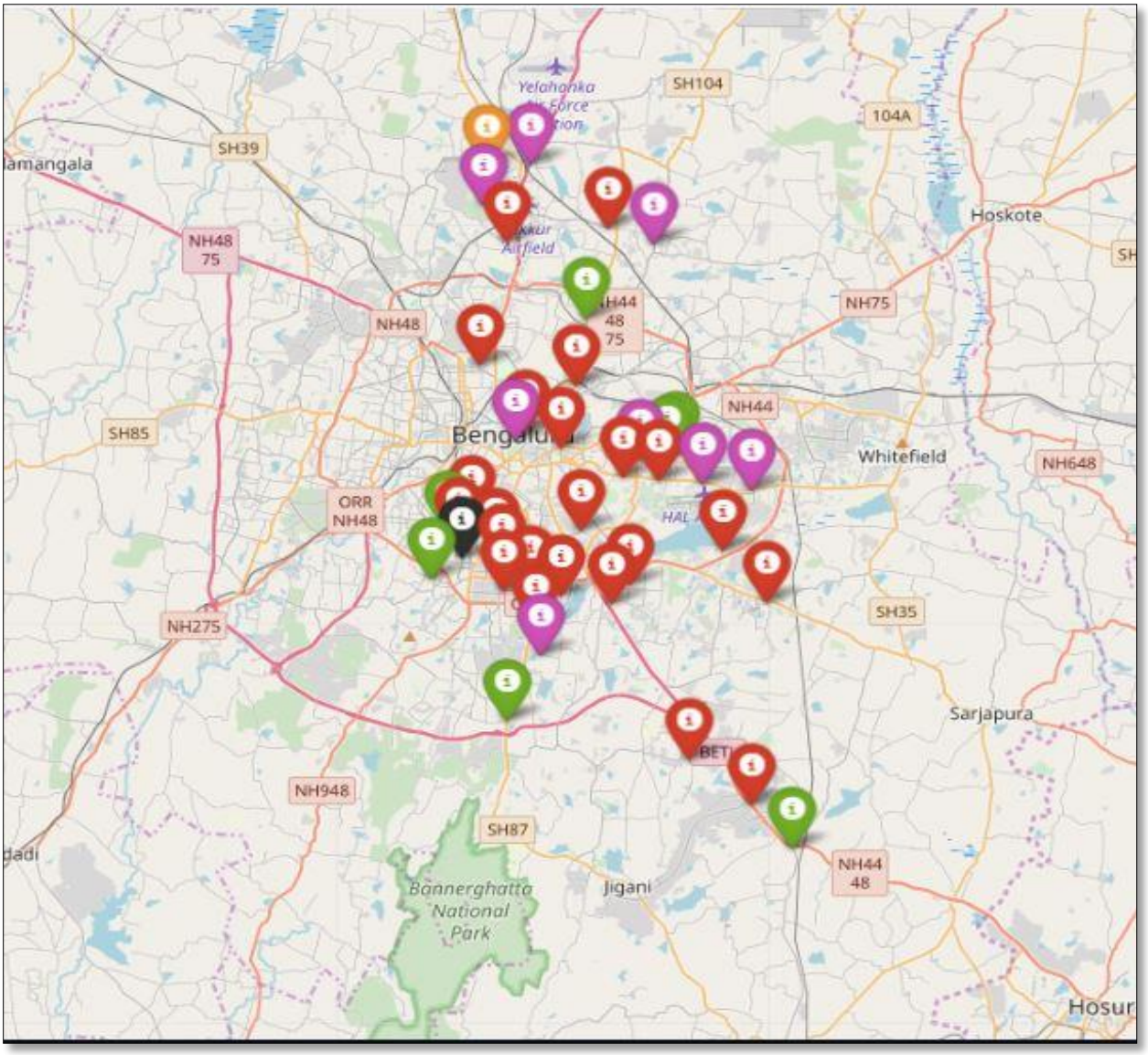
It is necessary for this algorithm to have a prior idea about the number of clusters since it is considered an input of this algorithm. For this reason, the elbow method is implemented. A chart that compares error vs number of clusters is done and the elbow is selected. Then, further analysis of each cluster is done.

Results

Firstly, data is plotted in a geographical map to get a notion of the Bangalore location on map. In the following image is shown the neighborhoods in Bangalore.



For visualization purposes, the geographical data is again plotted but with different colors. Each color represents the cluster for which that neighborhood belongs. This image is shown below.



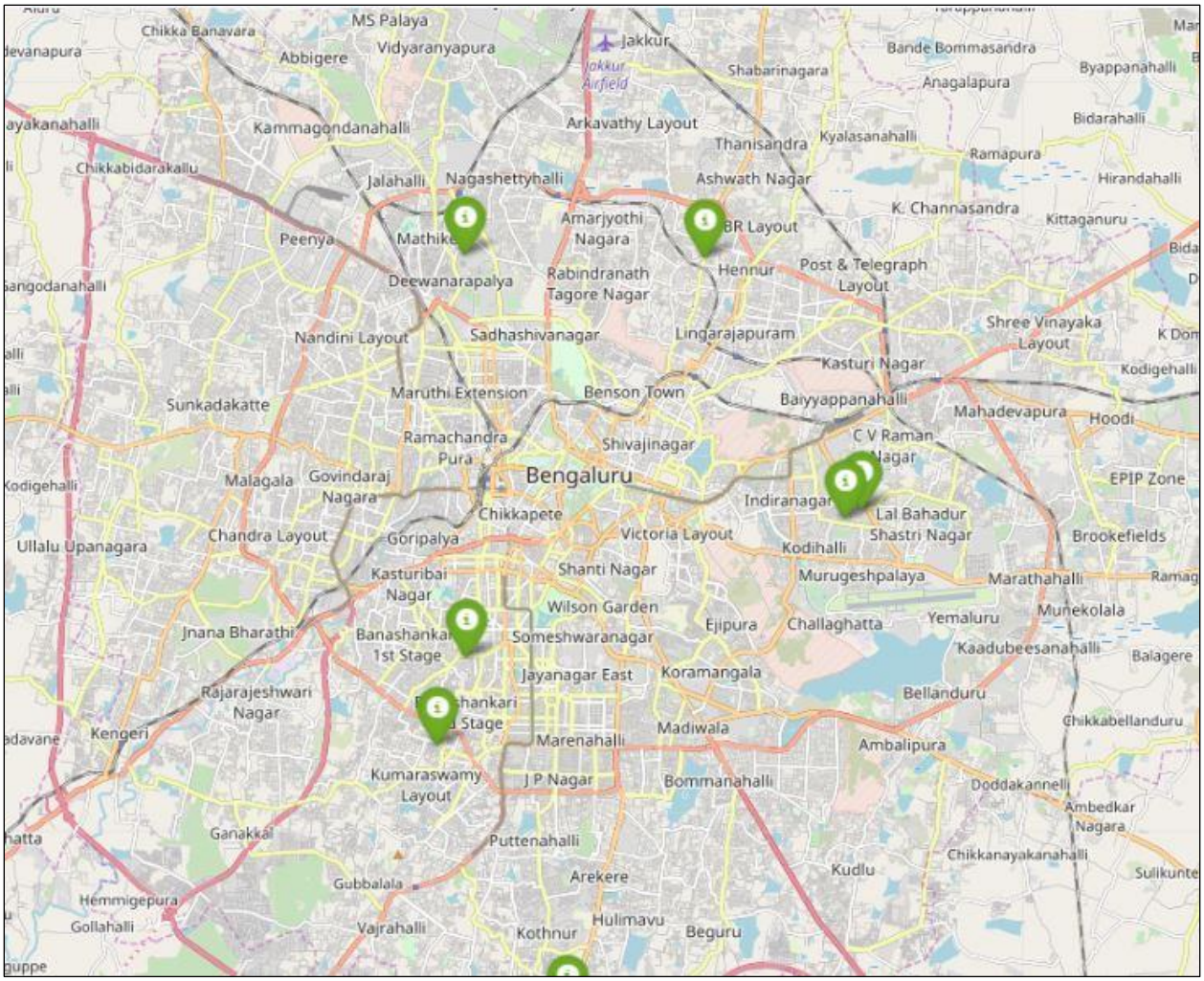
In this image it is evident that cluster algorithm is not segmenting the neighborhoods for location areas. This means that it is not true that geolocation of neighborhoods is correlated with the categories of the venues around each neighborhood. Yet, it is possible to see which neighborhoods within Bangalore, the people prefer to have coffee and within which areas do we have sufficient coffee shops.

Those neighborhoods that are similar among them belong to the same cluster. Hence, they have the same color in the image above.

In the image above it is found the proportion of the neighborhoods assigned to each cluster. There are two major clusters and three minor clusters. Moreover, there is a cluster that has only one neighborhood.

Analysis of all the clusters

Cluster 1

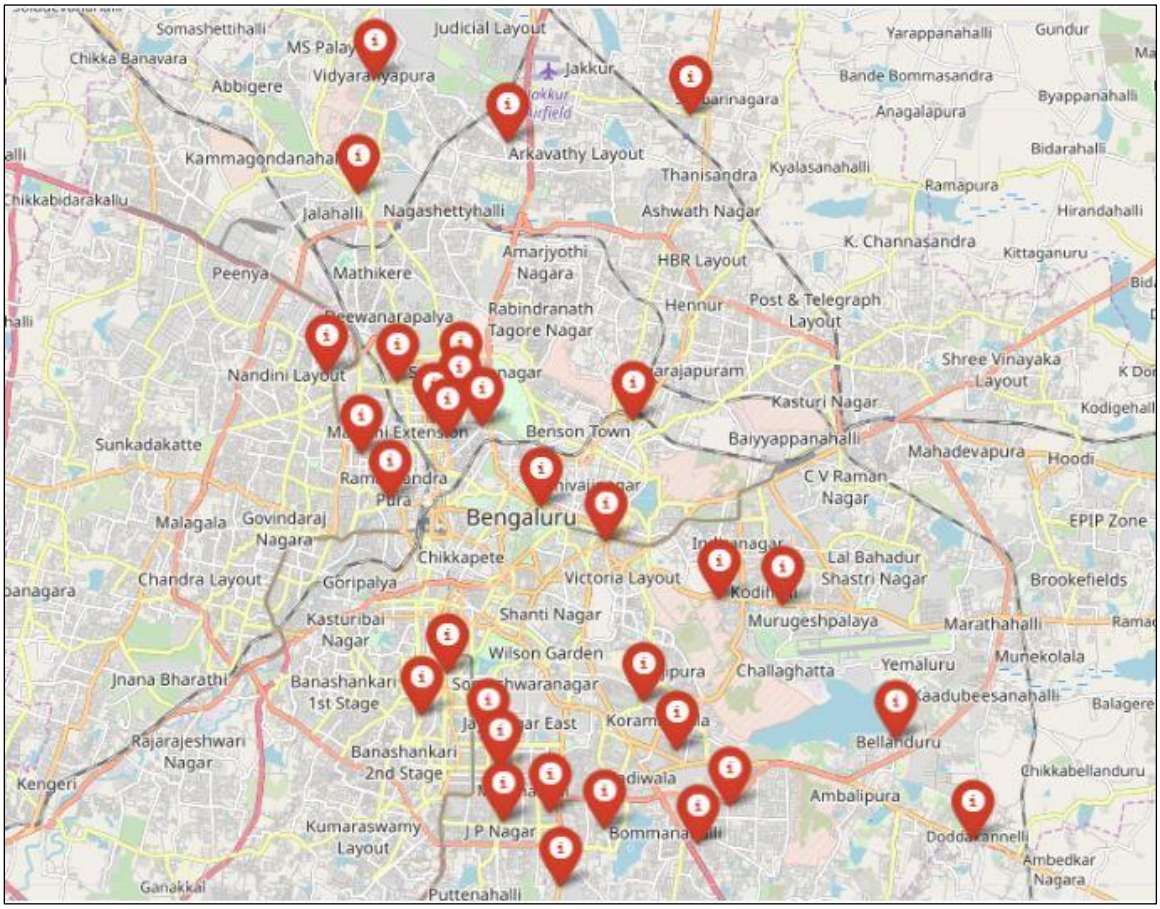


	Neighbourhood	Postal Code	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Arabic College	560045	77.6206	13.0291	1.0	Indian Restaurant	Coffee Shop	Electronics Store	Women's Store	Donut Shop
8	Jeevanbhimanagar	560075	77.6535	12.9695	1.0	Indian Restaurant	Park	Chinese Restaurant	Department Store	Kerala Restaurant
13	New Thippasandra	560075	77.6572	12.9716	1.0	Indian Restaurant	Park	Department Store	Food Truck	Bakery
20	Ashoknagar (Bangalore)	560050	77.5645	12.9377	1.0	Indian Restaurant	Breakfast Spot	Juice Bar	Szechuan Restaurant	Indie Movie Theater
27	Chandapura	560081	77.7041	12.8016	1.0	Indian Restaurant	Asian Restaurant	Women's Store	Donut Shop	Fish Market
30	Gottigere	560083	77.5883	12.8564	1.0	Indian Restaurant	Department Store	Women's Store	Food Court	Deli / Bodega
39	Padmanabhnagar	560070	77.5576	12.9176	1.0	Indian Restaurant	Park	Convenience Store	Coffee Shop	Snack Place
48	Msrit	560054	77.5642	13.0298	1.0	Indian Restaurant	Diner	Fast Food Restaurant	Women's Store	Food Court

Cluster 1 is widely spread on the outskirts of Bangalore. The Most common spot people visit is the Indian Restaurant followed by Park. We can see Coffee shop appearing in only Arabic College area.

It might not be a good idea to target this cluster as the people are looking more for Restaurants and outside places. There is also a possibility that the Restaurants provide coffee to customers.

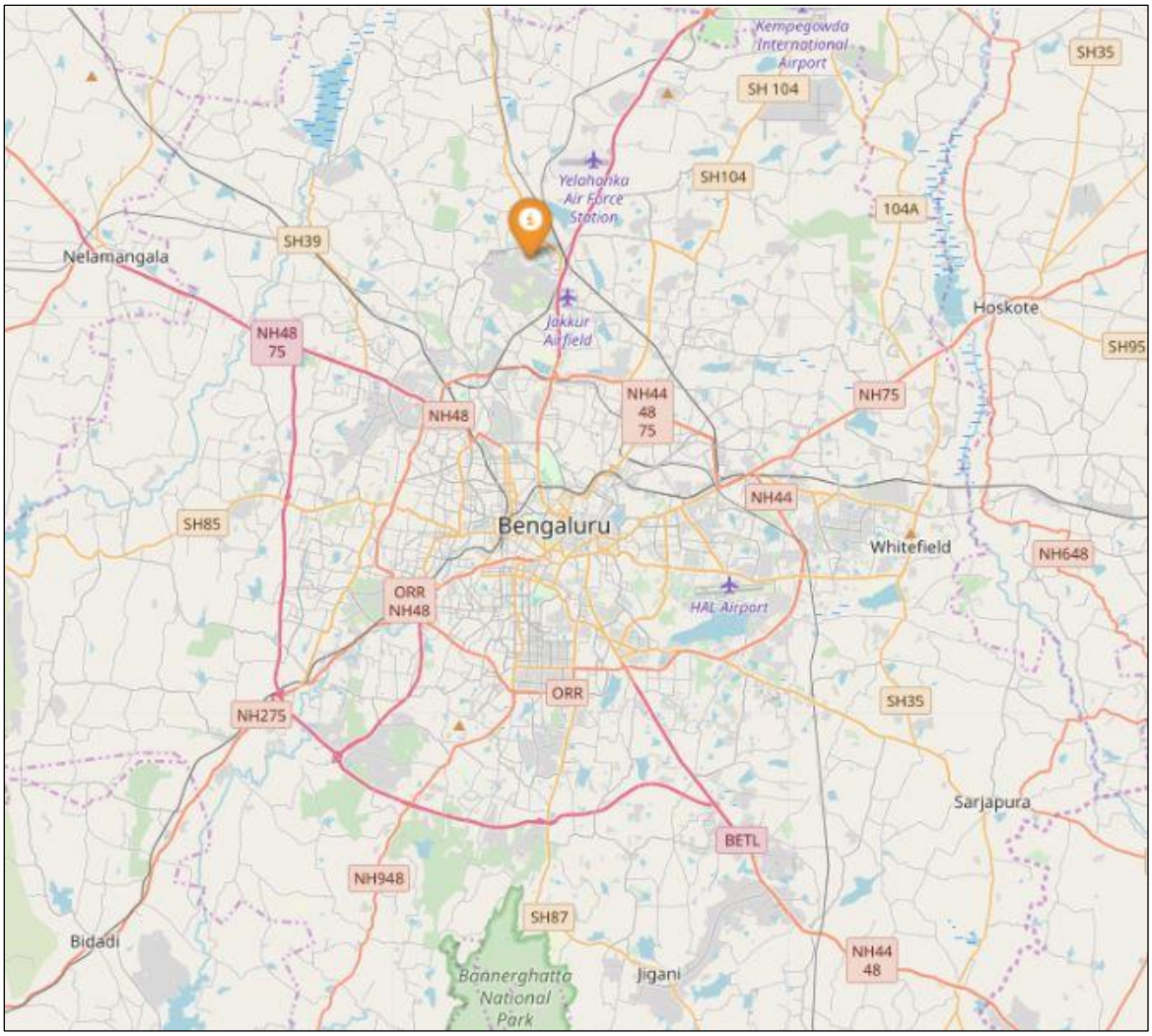
Cluster 2



	Neighbourhood	Postal Code	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Bellandur	560103	77.6760	12.9298	2.0	Fast Food Restaurant	Indian Restaurant	Kerala Restaurant	Pizza Place	Chinese Restaurant
2	Domlur	560071	77.6359	12.9611	2.0	Pizza Place	Indian Restaurant	Café	Garden	Women's Store
3	Dr. Shivarama Karanth Nagar	560077	77.6293	13.0681	2.0	Andhra Restaurant	Indian Restaurant	Asian Restaurant	Grocery Store	Deli / Bodega
4	Fraser Town	560005	77.6164	13.0005	2.0	Indian Restaurant	Ice Cream Shop	Juice Bar	Fast Food Restaurant	Bakery
10	Mahatma Gandhi Road	560001	77.6100	12.9739	2.0	Indian Restaurant	Café	Brewery	Donut Shop	Chinese Restaurant
12	NAL	560017	77.6502	12.9596	2.0	Café	Indian Restaurant	Coffee Shop	Nightclub	Middle Eastern Restaurant
14	Sadashivanagar	560080	77.5772	13.0094	2.0	Coffee Shop	Café	Fast Food Restaurant	Performing Arts Venue	Seafood Restaurant
15	Sahakaranagar P.O	560092	77.5880	13.0621	2.0	Indian Restaurant	Ice Cream Shop	Burger Joint	Italian Restaurant	Café
19	Bangalore G.P.O.	560001	77.5955	12.9814	2.0	Indian Restaurant	Café	Hotel	Bakery	Capitol Building
22	Bannerghatta Road	560076	77.5997	12.8976	2.0	Café	Food Court	Indian Restaurant	Sandwich Place	Office
23	Basavanagudi	560004	77.5741	12.9446	2.0	Indian Restaurant	Café	Fast Food Restaurant	Hookah Bar	Dessert Shop
24	Bommanahalli (Bangalore)	560068	77.6311	12.9068	2.0	Indian Restaurant	Clothing Store	Asian Restaurant	Auto Garage	Furniture / Home Store
25	Bommasandra Industrial Estate	560099	77.6878	12.8203	2.0	Department Store	Indian Restaurant	Fast Food Restaurant	Sporting Goods Shop	Chinese Restaurant
26	Carmelram	560035	77.6937	12.9078	2.0	North Indian Restaurant	Indian Restaurant	Sporting Goods Shop	Café	Women's Store
28	Dharmaram College	560029	77.5684	12.9352	2.0	Fast Food Restaurant	Indian Restaurant	Breakfast Spot	Sandwich Place	Plaza
29	Electronics City	560100	77.6626	12.8400	2.0	Hotel	Café	Coffee Shop	Fast Food Restaurant	Asian Restaurant

The Cluster 2 area is wide spread across Bangalore. Looking at the most preferred place, Restaurant, Café and Fast food joint tops the list. You can also see coffee shops as the 3rd Preferred Venue in most of the areas. Looks like the areas consists of majorly malls. This can be a good cluster to start with. We can target these areas. We might encounter some completion, but we may also get Coffee lovers from Bangalore.

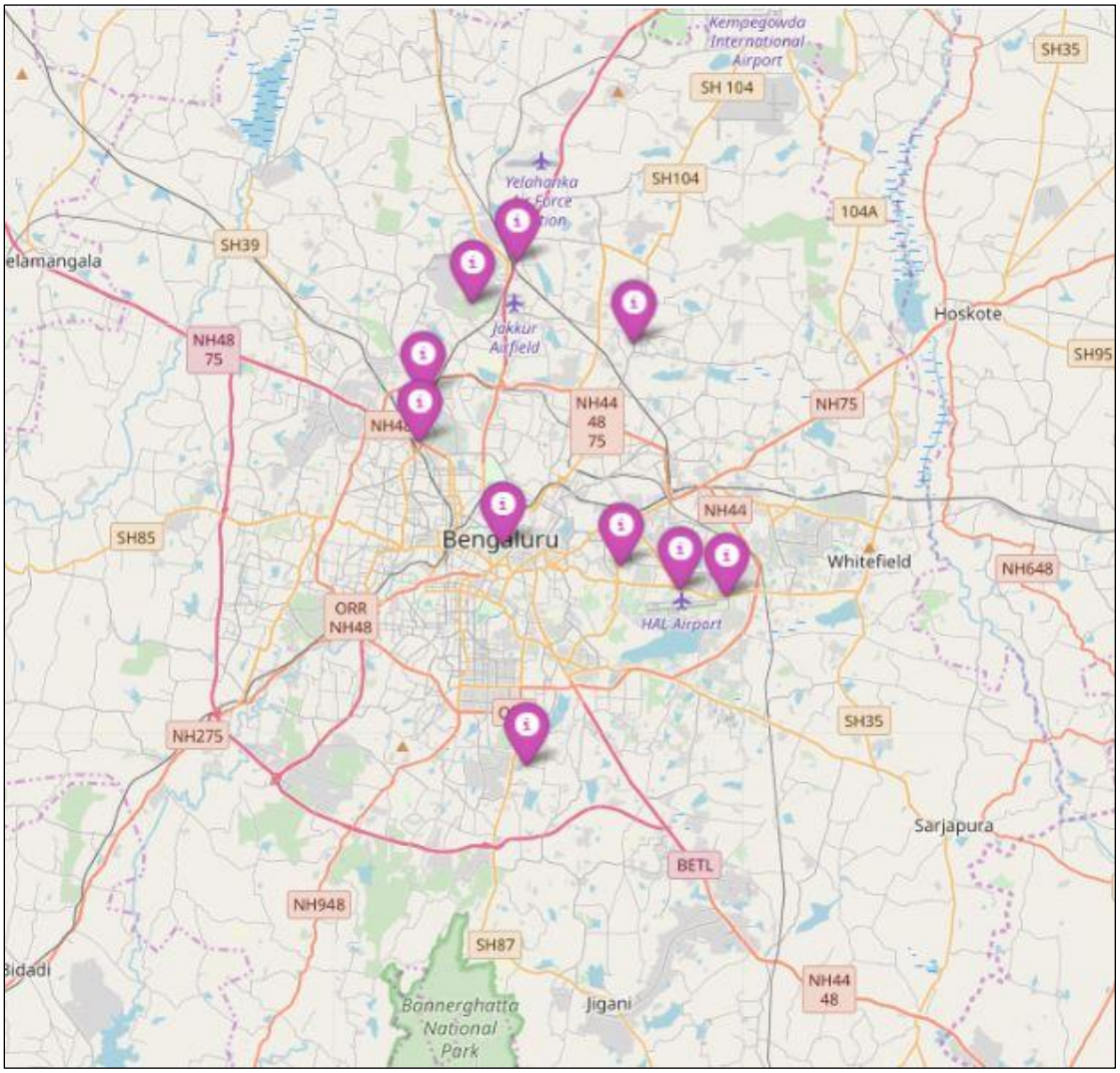
Cluster 3



	Neighbourhood	Postal Code	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
18	Yelahanka Satellite Town	560064	77.58	13.0946	3.0	Dessert Shop	Women's Store	Cupcake Shop	Deli / Bodega	Department Store

As this Cluster contains only 1 area, this might not be a very wise decision to put up a center here. Also, the most preferred destinations vary greatly with coffee shop.

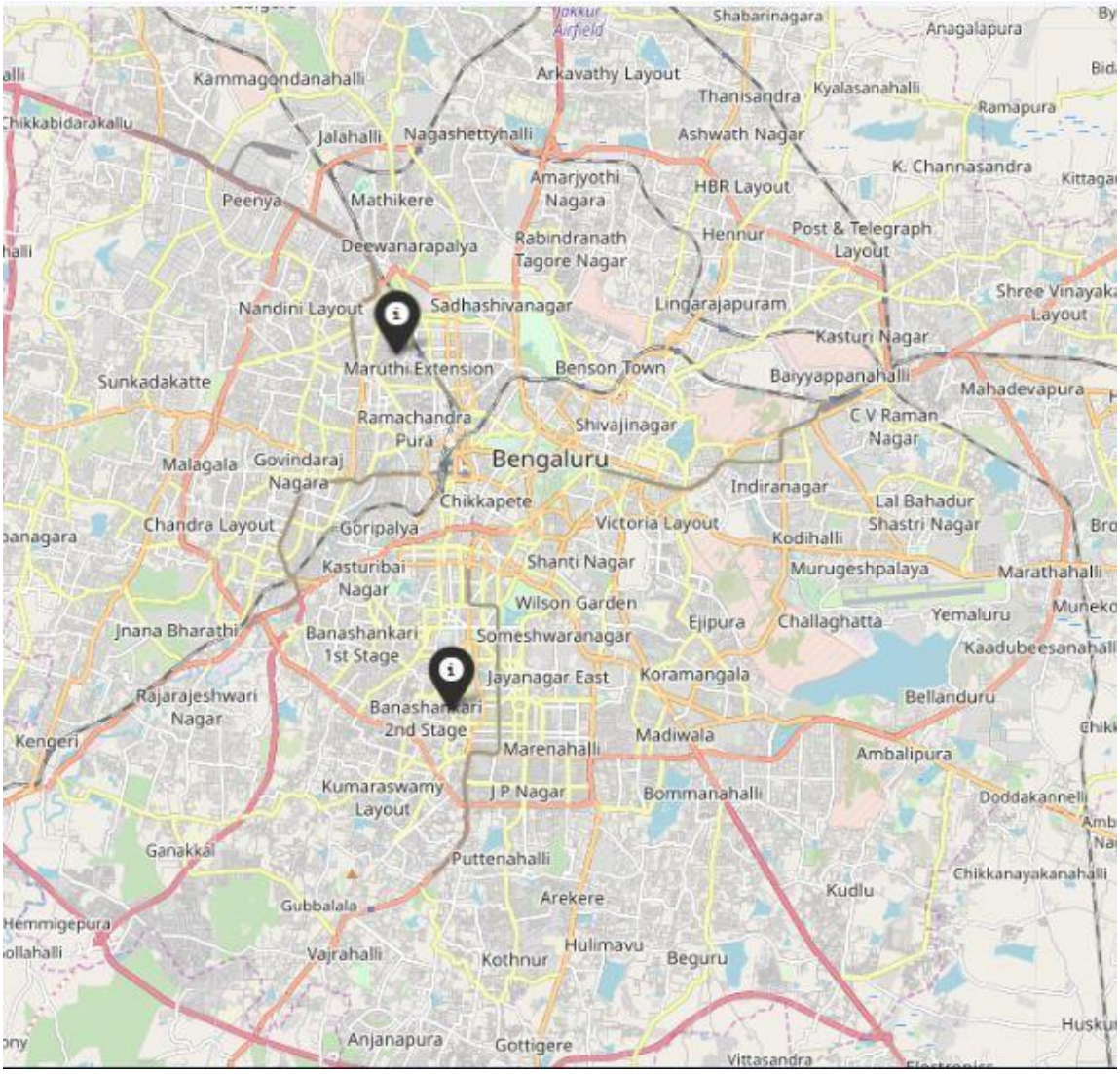
Cluster 4



	Neighbourhood	Postal Code	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	G.K.V.K.	560065	77.5784	13.0781	4.0	Basketball Court	Garden	Women's Store	Fish Market	Fast Food Restaurant
6	H.A.L II Stage	560008	77.6427	12.9686	4.0	Restaurant	Lounge	Asian Restaurant	Italian Restaurant	Ice Cream Shop
7	HighCourt	560001	77.5915	12.9773	4.0	Dog Run	Tennis Stadium	Metro Station	Capitol Building	Park
9	Kothanur	560077	77.6478	13.0614	4.0	Italian Restaurant	Coffee Shop	Pizza Place	Grocery Store	Bakery
11	Marathahalli Colony	560037	77.6880	12.9559	4.0	Shoe Store	Dessert Shop	Fried Chicken Joint	Road	Women's Store
16	Vimanapura	560017	77.6679	12.9585	4.0	Antique Shop	Bus Stop	Farmers Market	Women's Store	Dumpling Restaurant
17	Yelahanka	560064	77.5975	13.0958	4.0	Farmers Market	Clothing Store	Pub	Café	Women's Store
32	Hulimavu	560076	77.6022	12.8850	4.0	Pizza Place	Halal Restaurant	Department Store	Dumpling Restaurant	Fish Market
47	Mathikere	560054	77.5576	13.0403	4.0	Park	Italian Restaurant	Spa	Restaurant	Intersection
55	Yeswanthpura	560022	77.5562	13.0197	4.0	Vegetarian / Vegan Restaurant	Train Station	Boarding House	Women's Store	Donut Shop

The most preferred destinations for the areas in cluster 4 is mostly shops/Market or play areas outside home. This looks like residential areas and people may prefer having coffee at their home rather than in shop. Hence, we will not target this area.

Cluster 5



Neighbourhood	Postal Code	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
B Sk II Stage	560070	77.5695	12.9265	0.0	Ice Cream Shop	Performing Arts Venue	Pizza Place	Fast Food Restaurant	Dog Run
Gayathrinagar	560021	77.5581	13.0010	0.0	Ice Cream Shop	Light Rail Station	Women's Store	Donut Shop	Fish Market

Contains only 2 areas. Not a very profitable option, hence we will not target these areas.

Discussion

It is worth to note that this work is useful only for those who live in Bangalore or the people who have been to Bangalore in the past. The reason is because there is a limited amount of data, we can request using de Foursquare API. Consequently, it will have a greater cost than the Lite version.

Moreover, there is a cluster with one neighborhood. In the results we found out that this cluster has a Most preferred Venue as a Dessert Shop. Hence, we can say the algorithm is doing great since there is no other cluster with similar venues around.

Conclusion

In this work a segmentation of different areas of Bangalore is done. The data is downloaded and the venues around the neighborhoods is acquired using the Foursquare API. One Hot Encoding is used for converting the categories of the venues into a feature matrix. Then, all venues are grouped by neighborhoods and at the same time the mean is calculated. Hence, the resulting features used are the frequency of occurrence from each category in a neighborhood.

The K-Means clustering algorithm is used for finding similarities between all the neighborhoods listed in the feature matrix. Results show that there are 2 major groups and 3 minor groups. In addition, there is one group that contains only one neighborhood that is isolated from others.

We Will be targeting Cluster 2 as there is high probability that we may find many coffee lovers in that area.