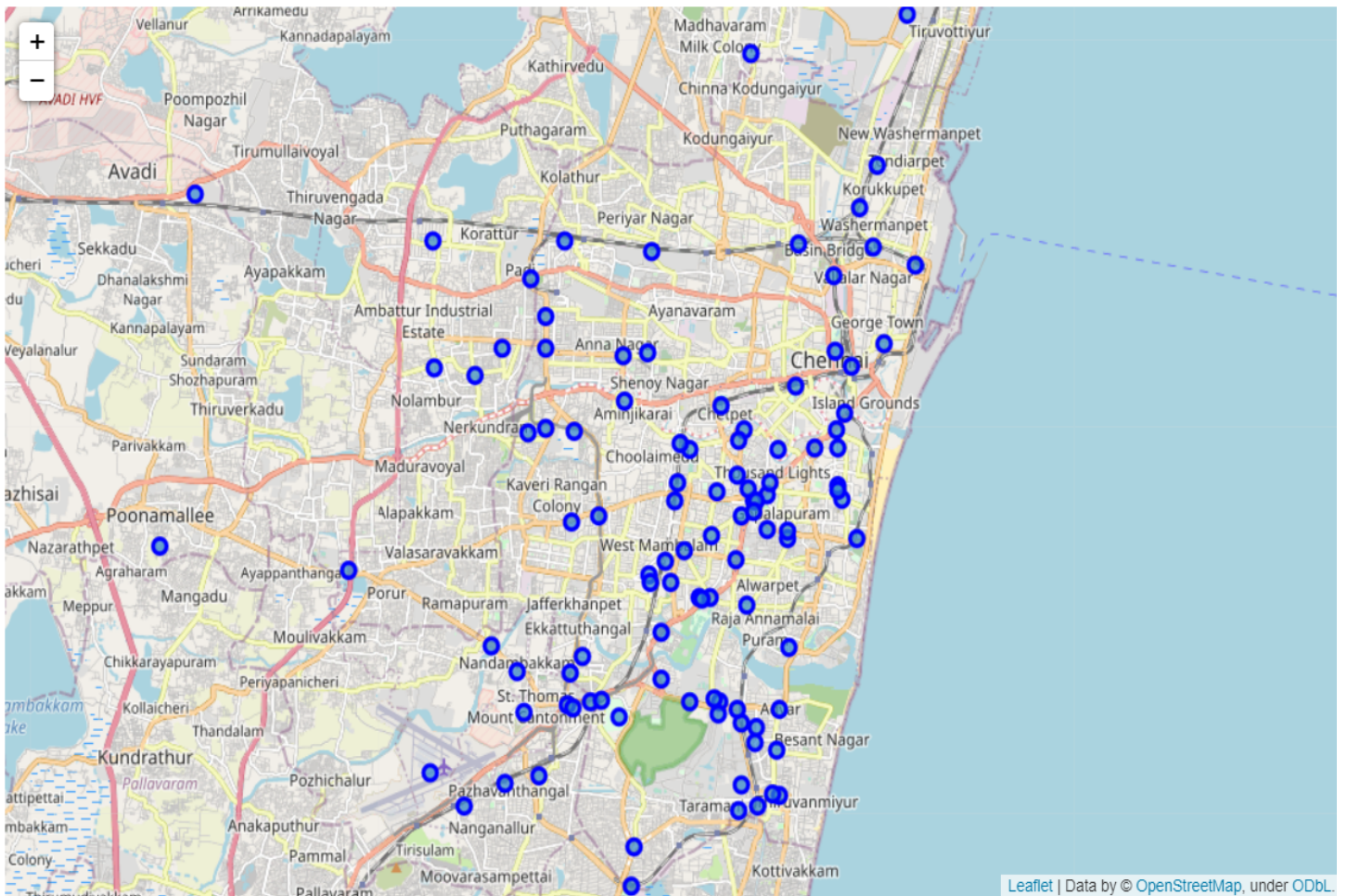


Capstone Project

THE BATTLE OF NEIGHBORHOODS

CHENNAI



Naidugari Manjunath

Business Problem Section

Background

Many people want to invest their money in the business to make profits. There are different types of business like Restaurants, Shopping Malls, and Departmental Stores etc. It is better to start their business in major cities in the country. In this project, I'd like to study the neighborhoods in Chennai, Tamil Nadu, India

Business Problem

While starting a business, we need to search for a neighborhood that is suitable to their preferences. In this project, we study neighborhoods and venues to know which locality is good to set up a business.

The major Target Audience would be small-scale business owners and stakeholders planning to start their business at a location in Chennai. This project would help them find the optimal location based on the category of their business such as,

- What is the best location to start a Restaurant in Chennai?
- Which area is best suitable for opening a Shopping Mall and a Departmental Store in Chennai?

Data Requirements

Chennai has multiple neighborhoods. The chennaiiq.com website has a dataset which has the list of neighborhoods in Chennai along with their Latitude and Longitude. Foursquare API is used to obtain the venue details in each neighborhood.

1. https://chennaiiq.com/chennai/latitude_longitude_areas.asp

(105, 3)

:

	Neighborhood	Latitude	Longitude
0	Adyar Bus Debot	12°59'50" N	80°15'25" E
1	Adyar Signal	13°00'23" N	80°15'27" E
2	Alandur	13°00'28" N	80°12'35" E
3	Ambattur	13°06'36" N	80°10'12" E
4	Anna Arch	13°04'28" N	80°13'06" E

But latitude and longitude are in Degree minute second format which is converted as Decimal format.

(105, 3)

	Neighborhood	Latitude	Longitude
0	Adyar Bus Debot	12.997222	80.256944
1	Adyar Signal	13.006389	80.257500
2	Alandur	13.007778	80.209722
3	Ambattur	13.110000	80.170000
4	Anna Arch	13.074444	80.218333

The details of venues in neighborhoods like **Venue**, **Venue Latitude**, **Venue Longitude**, **Venue Category** should be obtained with the help of Foursquare API.

2. <https://www.foursquare.com>

A total of 1009 venues are obtained from Foursquare, which are used in further analysis

(1009, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adyar Bus Debot	12.997222	80.256944	Zaitoon Restaurant	12.996861	80.256178	Middle Eastern Restaurant
1	Adyar Bus Debot	12.997222	80.256944	Zha Cafe	12.999730	80.254806	Café
2	Adyar Bus Debot	12.997222	80.256944	Kuttanadu Restaurant	12.997010	80.257799	Asian Restaurant
3	Adyar Bus Debot	12.997222	80.256944	Barbeque N Biryani	12.995907	80.256011	BBQ Joint
4	Adyar Bus Debot	12.997222	80.256944	Domino's Pizza	13.001470	80.256626	Pizza Place

Methodology

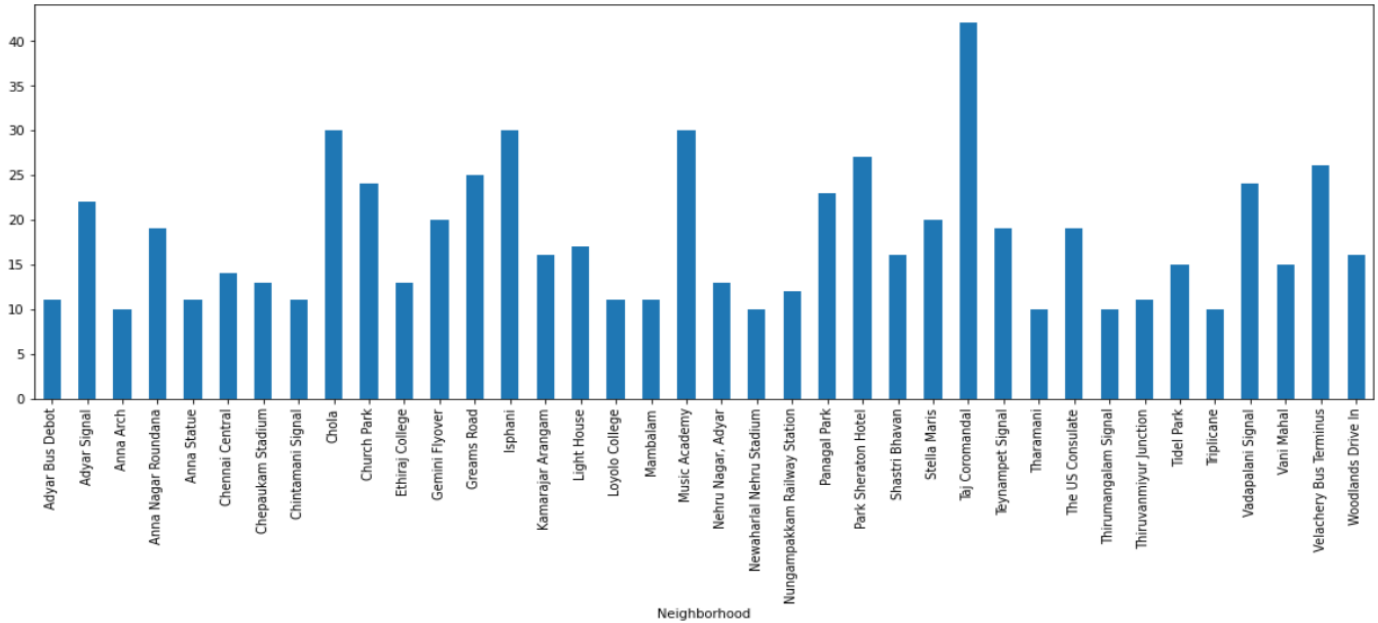
Now, we have the neighborhood data of chennai (**105 neighborhoods**). We also have the most popular venues in each neighborhood obtained using Foursquare API. A total of **1009 venues** have been obtained in the whole city and **136 unique categories**. But as seen we have multiple neighborhoods with less than 10 venues returned. In order to create a good analysis let's consider only the **neighborhoods with more than 10 venues**.

We can perform one hot encoding on the obtained data set and use it to find the 10 most common venue categories in each neighborhood. Then clustering can be performed on the dataset. Here, K - Nearest Neighbor clustering technique has been used. To find the optimal number of clusters silhouette score metric technique is used.

The clusters obtained can be analyzed to find the major type of venue categories in each cluster. This data can be used to suggest business people for suitable locations based on the category.

Analysis

From the dataset we get to know that there are many neighborhoods with less than 10 venues which can be removed before performing the analysis to obtain better results. The following plot shows only the neighborhoods from which 10 or more than 10 venues were obtained.



Now we will use One hot encoding to categorize the venues in each neighborhood. Then we group the data by neighborhood and take the mean value of frequency of occurrence of each category.

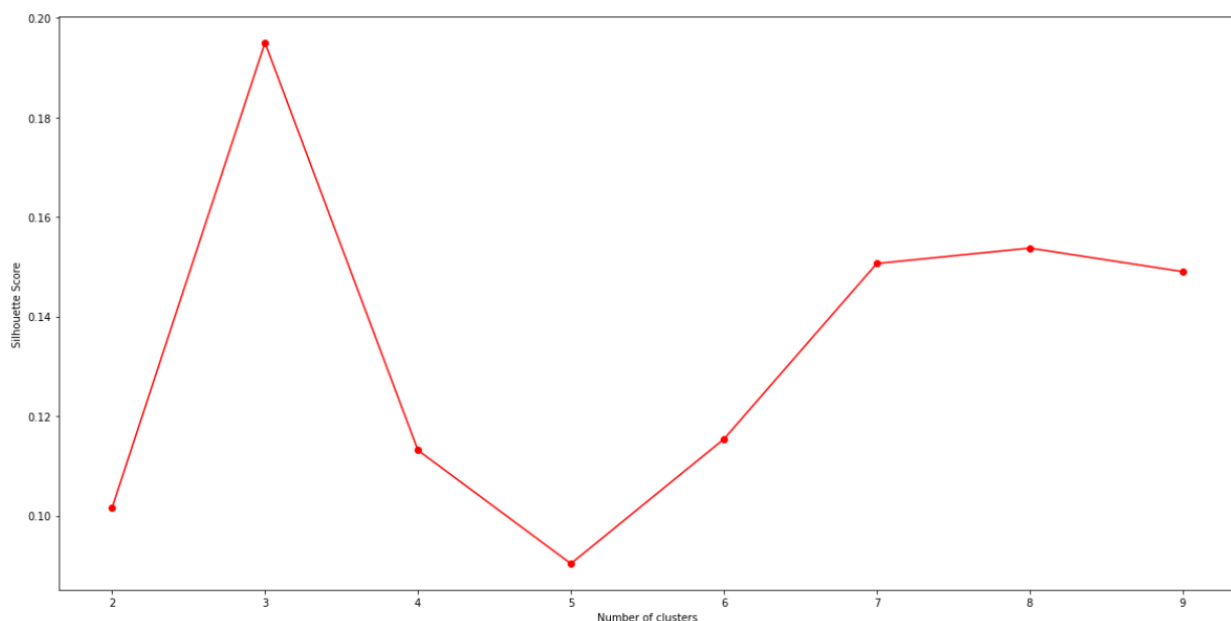
(676, 94)

[illegible]

The above dataset is used to obtain the top 10 most common venues in each neighborhood i.e 10 values with highest mean of frequency of occurrence. A sample of five neighborhoods is shown in figure below.

	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	Adyar Bus Debot	Asian Restaurant	Indian Restaurant	Pizza Place	Sandwich Place	BBQ Joint	Fast Food Restaurant	Middle Eastern Restaurant	Café	Breakfast Spot	Department Store
1	Adyar Signal	Indian Restaurant	North Indian Restaurant	Bakery	Coffee Shop	Snack Place	Fast Food Restaurant	Electronics Store	Dessert Shop	Rock Club	Café
2	Anna Arch	Clothing Store	Fast Food Restaurant	Bookstore	Food Court	Electronics Store	Café	Multiplex	Pizza Place	Gym / Fitness Center	Indian Restaurant
3	Anna Nagar Roundana	Indian Restaurant	Clothing Store	Bookstore	Hotel Bar	Coffee Shop	Electronics Store	Fast Food Restaurant	Café	Middle Eastern Restaurant	Paper / Office Supplies Store
4	Anna Statue	Indian Restaurant	Multiplex	Movie Theater	Dessert Shop	Flea Market	General Entertainment	Department Store	Donut Shop	Eastern European Restaurant	Electronics Store

The resultant dataset after encoding is used for clustering. Here the K-Nearest Neighbor(KNN) clustering algorithm is used. It is an unsupervised Machine Learning that clusters the given data into n-different clusters. For optimal result we need to select the best value of k. Here, we can use silhouette score to find the best value of k. A range of clusters 2 to 10 considered in KNN line plot. From the plot we can see that a k value of 2 provides the best score. Now, this k value is used for K-Means Clustering Technique and we get data like in the figure.



Results

Let's examine the 8 clusters and find the discriminating venue categories which distinguish each cluster. For this purpose let's look at the most common venue category in each cluster.

Cluster 1

Top venue categories in cluster 1 are Restaurant, Pizza Place, Book Store, Coffee Shop, Sandwich Place.

	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	Adyar Bus Debot	Asian Restaurant	Indian Restaurant	Pizza Place	Sandwich Place	BBQ Joint	Fast Food Restaurant	Middle Eastern Restaurant	Café	Breakfast Spot	Department Store
1	Adyar Signal	Indian Restaurant	North Indian Restaurant	Bakery	Coffee Shop	Snack Place	Fast Food Restaurant	Electronics Store	Dessert Shop	Rock Club	Café
3	Anna Nagar Roundana	Indian Restaurant	Clothing Store	Bookstore	Hotel Bar	Coffee Shop	Electronics Store	Fast Food Restaurant	Café	Middle Eastern Restaurant	Paper / Office Supplies Store
6	Chepaukam Stadium	Indian Restaurant	Electronics Store	Bookstore	Multiplex	Hotel	Bar	General Entertainment	Café	Music Store	Dessert Shop
7	Chintamani Signal	Indian Restaurant	Restaurant	Coffee Shop	Dessert Shop	Bakery	Middle Eastern Restaurant	Smoke Shop	Shoe Store	Flea Market	Convenience Store
20	Newaharlal Nehru Stadium	Indian Restaurant	Bookstore	Soccer Stadium	Café	Sandwich Place	Platform	Vegetarian / Vegan Restaurant	Hotel Bar	Hotel	Convenience Store
24	Shastri Bhavan	Indian Restaurant	Chinese Restaurant	Japanese Restaurant	Hotel	Coffee Shop	Fast Food Restaurant	Multicuisine Indian Restaurant	Pizza Place	Convenience Store	Asian Restaurant
26	Taj Coromandal	Indian Restaurant	Sandwich Place	Chinese Restaurant	Café	Ice Cream Shop	Asian Restaurant	Dessert Shop	Italian Restaurant	Clothing Store	Park
36	Velachery Bus Terminus	Indian Restaurant	Fast Food Restaurant	Pizza Place	Clothing Store	Coffee Shop	Department Store	Chinese Restaurant	Ramen Restaurant	Restaurant	Multiplex

Cluster 2

Top venue categories in cluster 2 are Restaurant, Bookstore, Hotel, Coffee Shop, Juice Bar.

	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
5	Chennai Central	Indian Restaurant	Bookstore	Train Station	Sandwich Place	Vegetarian / Vegan Restaurant	Platform	Hotel	Metro Station	Nightclub	Bus Station
8	Chola	Indian Restaurant	Hotel	Restaurant	African Restaurant	Russian Restaurant	Italian Restaurant	Juice Bar	Vegetarian / Vegan Restaurant	Kerala Restaurant	Lounge
9	Church Park	Chinese Restaurant	Indian Restaurant	Movie Theater	Multiplex	Bakery	Café	Middle Eastern Restaurant	Pub	Gaming Cafe	Pizza Place
12	Greams Road	Indian Restaurant	Café	Juice Bar	Middle Eastern Restaurant	Movie Theater	Multiplex	Chinese Restaurant	Vegetarian / Vegan Restaurant	Gaming Cafe	Mexican Restaurant
18	Music Academy	Indian Restaurant	Hotel	Restaurant	Café	Women's Store	Russian Restaurant	Bank	Bar	Chinese Restaurant	Comfort Food Restaurant
25	Stella Maris	Indian Restaurant	Hotel	Women's Store	Sandwich Place	Bar	Camera Store	Chinese Restaurant	Ice Cream Shop	Italian Restaurant	Jewelry Store
27	Teynampet Signal	Indian Restaurant	Hotel	Italian Restaurant	Lounge	Pub	Pizza Place	Café	Juice Bar	Mediterranean Restaurant	Gym / Fitness Center
31	Thiruvanniyur Junction	Juice Bar	Indian Restaurant	Pizza Place	Multiplex	Vegetarian / Vegan Restaurant	Indian Sweet Shop	Pharmacy	Hotel	Chinese Restaurant	Department Store
34	Vadapalani Signal	Indian Restaurant	Hotel	Fast Food Restaurant	Asian Restaurant	Multiplex	South Indian Restaurant	Market	Shopping Mall	Music Store	Clothing Store
35	Vani Mahal	Indian Restaurant	Hotel	Coffee Shop	Vegetarian / Vegan Restaurant	BBQ Joint	Department Store	Hotel Bar	South American Restaurant	Pub	Boutique

Cluster 3

Top venue categories in cluster 3 are Restaurant, Multiplex, Dessert Shop, General Entertainment.

	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
4	Anna Statue	Indian Restaurant	Multiplex	Movie Theater	Dessert Shop	Flea Market	General Entertainment	Department Store	Donut Shop	Eastern European Restaurant	Electronics Store
33	Triplicane	Indian Restaurant	Multiplex	General Entertainment	Dessert Shop	Gaming Cafe	Department Store	Donut Shop	Eastern European Restaurant	Electronics Store	Fast Food Restaurant

Cluster 4

Top venue categories in cluster 4 are Fast Food Restaurant, Cafe, Food Court, Bus Station, Clothing Store.

	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
2	Anna Arch	Clothing Store	Fast Food Restaurant	Bookstore	Food Court	Electronics Store	Café	Multiplex	Pizza Place	Gym / Fitness Center	Indian Restaurant
28	Tharamani	Fast Food Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Office	Hotel	Café	Sandwich Place	Food Court	Women's Store	Department Store
32	Tidel Park	Food Court	Fast Food Restaurant	Café	Sandwich Place	Platform	Bus Station	Vegetarian / Vegan Restaurant	Indian Restaurant	Office	Convenience Store

Cluster 5

Top venue categories in cluster 5 are Clothing Store, Jewelry Store, Indian Restaurant, Women's Store, Boutique.

	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
17	Mambalam	Clothing Store	Jewelry Store	Indian Restaurant	Boutique	Asian Restaurant	South Indian Restaurant	Snack Place	Department Store	Dessert Shop	Donut Shop
22	Panagal Park	Indian Restaurant	Clothing Store	South Indian Restaurant	Jewelry Store	Women's Store	Pizza Place	Concert Hall	Coffee Shop	Boutique	Dessert Shop

Cluster 6

Top venue categories in cluster 6 are Train Station, Departmental Store, Indian Restaurant, Departmental Store, Women's Store.

	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
16	Loyolo College	Women's Store	Train Station	Vegetarian / Vegan Restaurant	Department Store	Indian Restaurant	Ice Cream Shop	Clothing Store	South Indian Restaurant	Chinese Restaurant	Mexican Restaurant
21	Nungampakkam Railway Station	Department Store	Women's Store	Train Station	Vegetarian / Vegan Restaurant	Indian Restaurant	Ice Cream Shop	Clothing Store	South Indian Restaurant	Chinese Restaurant	Mexican Restaurant

Cluster 7

Top venue categories in cluster 7 are Hotel, Cafe, Indian Restaurant, Night Club, Concert Hall.

	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
10	Ethiraj College	Hotel	Kebab Restaurant	Restaurant	Indian Restaurant	Middle Eastern Restaurant	Fast Food Restaurant	Donut Shop	Pizza Place	Juice Bar	Asian Restaurant
11	Gemini Flyover	Hotel	Sandwich Place	Nightclub	Coffee Shop	Concert Hall	Bar	Café	Chinese Restaurant	Pizza Place	South Indian Restaurant
13	Isphani	Indian Restaurant	Hotel	Nightclub	Sandwich Place	Chinese Restaurant	Coffee Shop	South Indian Restaurant	Pizza Place	Bar	Pub
14	Kamarajar Arangam	Hotel	Café	Indian Restaurant	Pizza Place	Concert Hall	Amphitheater	Ice Cream Shop	Lounge	Park	South Indian Restaurant
23	Park Sheraton Hotel	Café	Restaurant	Italian Restaurant	Indian Restaurant	Thai Restaurant	Hotel	Coffee Shop	Sandwich Place	Japanese Restaurant	Dessert Shop
29	The US Consulate	Hotel	Sandwich Place	Nightclub	Coffee Shop	Electronics Store	Café	Chinese Restaurant	Bar	Park	South Indian Restaurant
37	Woodlands Drive In	Hotel	Indian Restaurant	Nightclub	Coffee Shop	Electronics Store	Park	Thai Restaurant	Ice Cream Shop	South Indian Restaurant	Bar

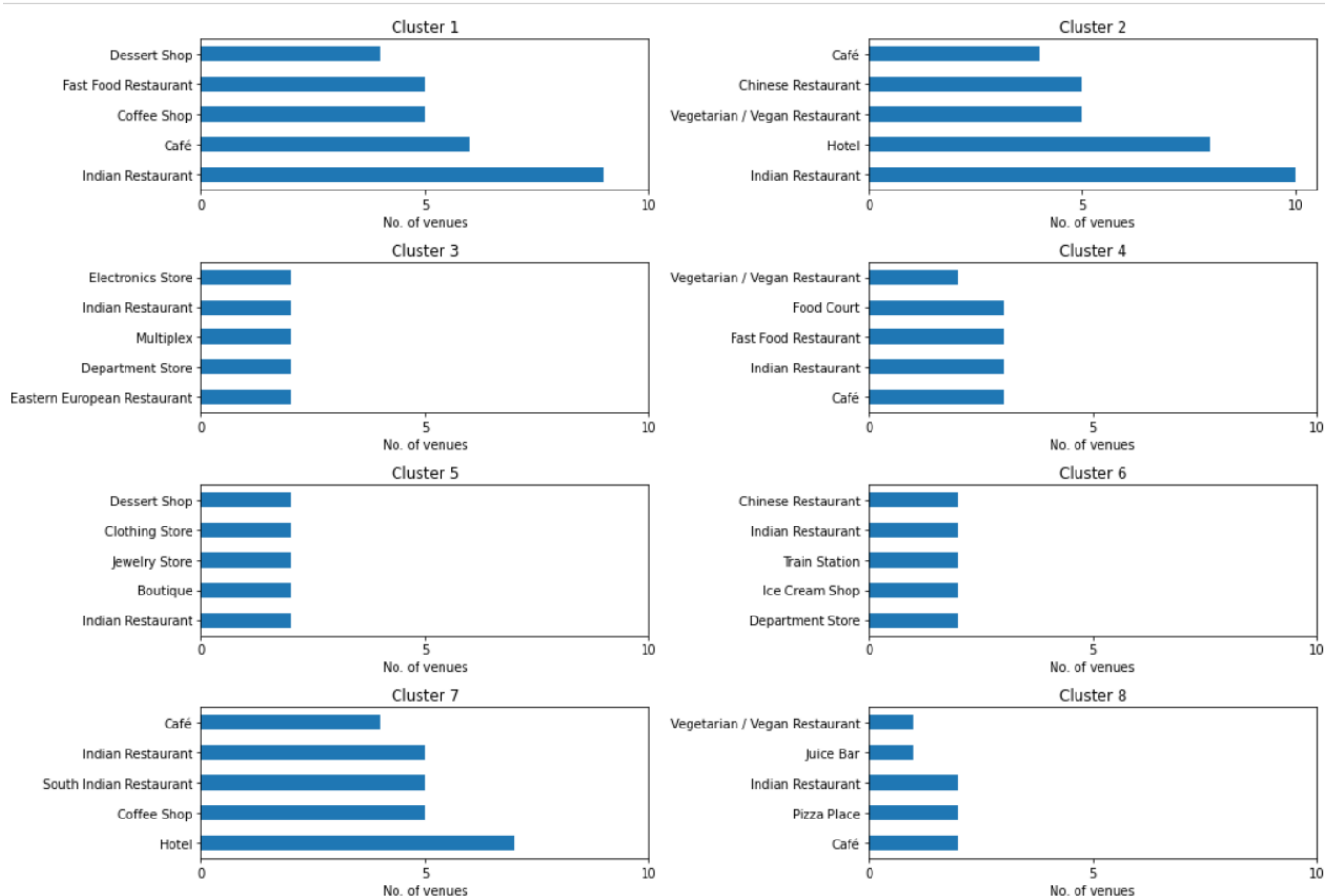
Cluster 8

Top venue categories in cluster 8 are Indian Restaurant, Pizza Place, Vegan Restaurant, Juice Shop.

	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
19	Nehru Nagar, Adyar	Café	Indian Restaurant	Chinese Restaurant	Pizza Place	Juice Bar	Ice Cream Shop	Sandwich Place	Flea Market	Department Store	Dessert Shop
30	Thirumangalam Signal	Indian Restaurant	Pizza Place	Jewelry Store	Vegetarian / Vegan Restaurant	Print Shop	Metro Station	Smoke Shop	Mobile Phone Shop	Café	Fast Food Restaurant

Discussion

Now we have clusters and top 5 venues in the neighborhood. Let's visualize the top 5 venue categories in each cluster.



This plot can be used to suggest valuable information to Business persons. Let's discuss a few examples considering they would like to start the following category of Business.

1. Hotel

The neighborhoods in cluster 2 has the greatest number of hotels, hence opening one here is not the best choice. So, is it best to open one in the neighborhoods in cluster 6 or 8? Not likely, since the place has a smaller number of food restaurants. Thus, an optimal place would be one which has less hotels, but also has restaurants and other places to explore.

Considering all these facts, the best choice would be Cluster 1 and Cluster 7. such as the Adyar Bus Depot, Gemini Flyover neighborhoods.

2. Shopping Mall

By using the same procedure as above, the suitable clusters would be the Cluster 1, Cluster 2 and Cluster 7, since it does not have shopping malls in any of the clusters and also it has many Hotels and Restaurants which gives an advantage.

3. Departmental Stores

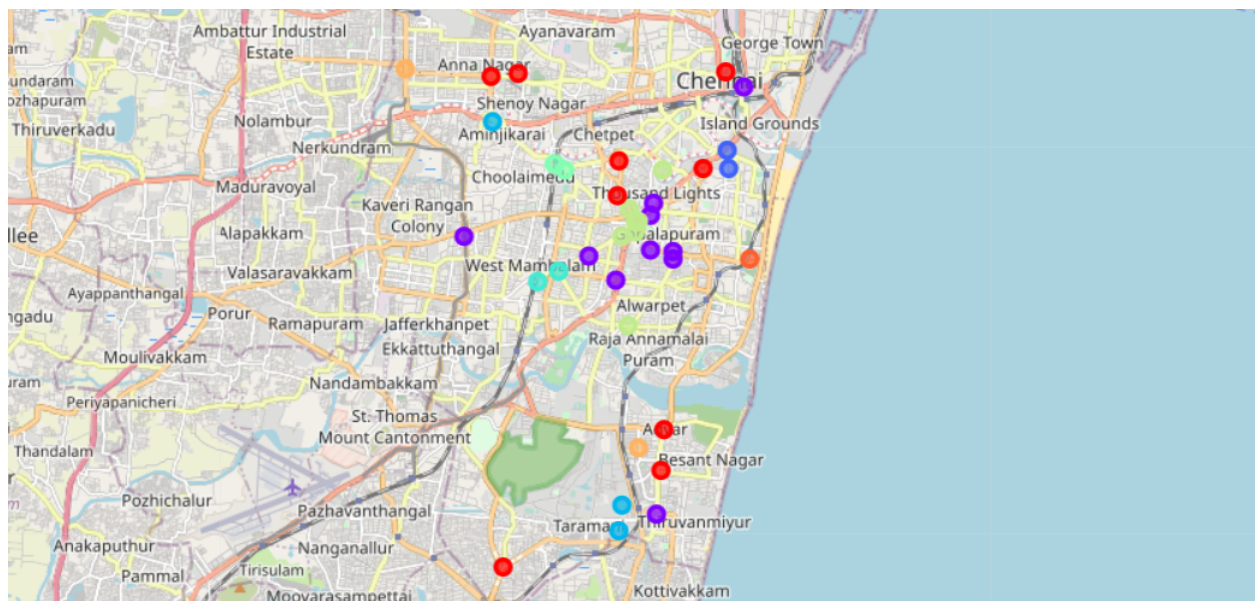
Repeat the procedure used in hotels. Although cluster 3 have departmental stores, cluster 4 and cluster 7 are also suitable neighborhoods for business. Since we don't have enough information from foursquare.

Map of Chennai with the clusters superimposed on top

This map can be used to find a suitable location to start a new business based on the category.

For example: Red = Cluster 1 , Violet = Cluster 2, Sky Blue = Cluster 4, Light Green = Cluster 7.

These clusters are suitable to start a business in Chennai.



Conclusion

Purpose of this project was to analyze the neighborhoods of Chennai and create a clustering model to suggest personal places to start a new business based on the category. The neighborhoods data was obtained from an online source and the Foursquare API was used to find the major venues in each neighborhood. But we found that many neighborhoods had less than 10 venues returned. In order to build a good Data Science model, we filtered out these locations. The remaining locations were used to create a clustering model. The best number of clusters i.e. 8 was obtained using the silhouette score. Each cluster was examined to find the most venue categories present, that defines the characteristics for that particular cluster.

A few examples for the applications that the clusters can be used for have also been discussed. A map showing the clusters have been provided. Both these can be used by stakeholders to decide the location for the particular type of business. A major drawback of this project was that the Foursquare API returned only a few venues in each neighborhood. As a future improvement, better data sources can be used to obtain more venues in each neighborhood. This way the neighborhoods that were filtered out can be included in the clustering analysis to create a better decision model.