

Capstone Project by IBM for Data Science Course on  
Coursera (Submitted by Apoorva Bajaj)

# Opening a Restaurant in South Indian City of Chennai

## Section Index

1. Section 1: Introduction: Background and Problem Discussion
2. Section 2: Data Description and Requirements
3. Section 3: Methodology
4. Section 4: Analysis
5. Section 5: Results
6. Section 6: Conclusion

## 1. Introduction: Background and Problem Discussion

I am from India, one of my close friends stays in Chennai. He wants to open a restaurant in Chennai. He is not sure what parameters to consider while choosing a perfect location for his restaurant project in Chennai. I am helping him with data-driven decision-making using Data science. This Capstone project will explore the neighbourhoods of Chennai, a city in the Southern Part of India. I am of the opinion, that this project, its results and conclusions will be helpful for my friend who is planning to open a Restaurant in Chennai and would also be helpful to someone who is thinking of opening a similar food-outlet in Chennai.

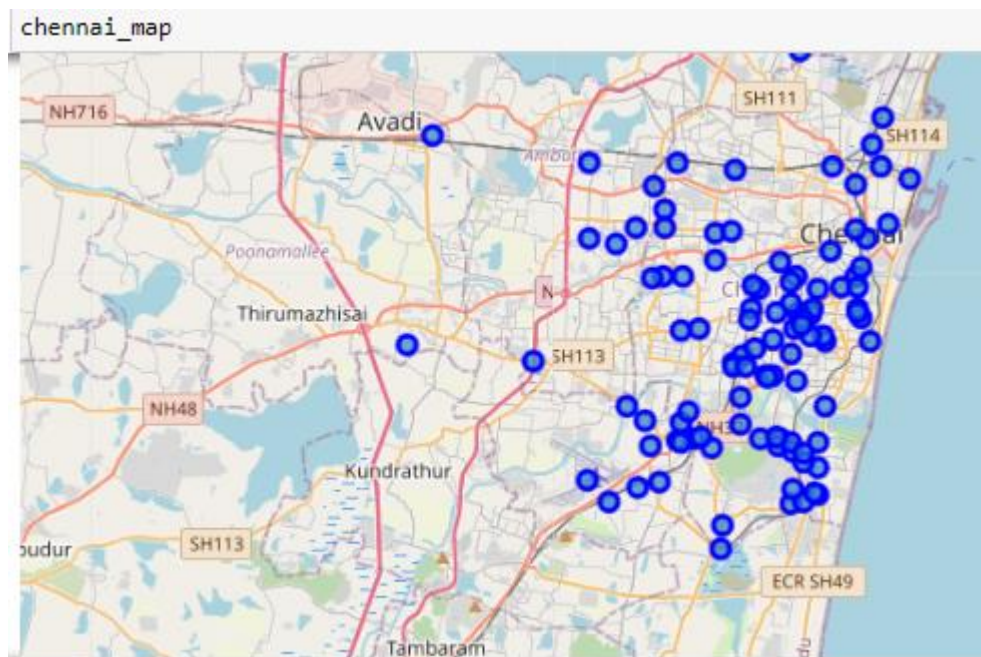
For the purpose of this project, I will be using “Foursquare API” to explore the neighbourhoods of Chennai. Specifically, I am interested in analysing the neighbourhood areas for which several venues can be obtained. Using Data Science methodologies, they are clustered based on the venues. I will be using “k-means clustering algorithm” to achieve the task. The optimal number of clusters will be gathered using silhouette score. I will be using “Folium visualization library\*\*” to visualize the clusters superimposed on the map of the city of Chennai. These clusters can be analysed to provide data driven and visual insights to my friend or other small-scale businesses to select an apt location for their requirement, such as Restaurants or Hotels.

## 2. Data Description and Requirements

Chennai has several neighbourhoods. I have found this dataset which includes the list of locations in Chennai city: [https://chennaiiq.com/chennai/latitude\\_longitude\\_areas.asp](https://chennaiiq.com/chennai/latitude_longitude_areas.asp) along with their Latitude and Longitude co-ordinates. One thing to note is that the Latitude and Longitude data provided in this link is in Degrees, Minutes and Seconds format. This has to be converted to Decimal Degrees before starting the analysis.

	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

*Exploring and Clustering neighborhoods in Chennai*



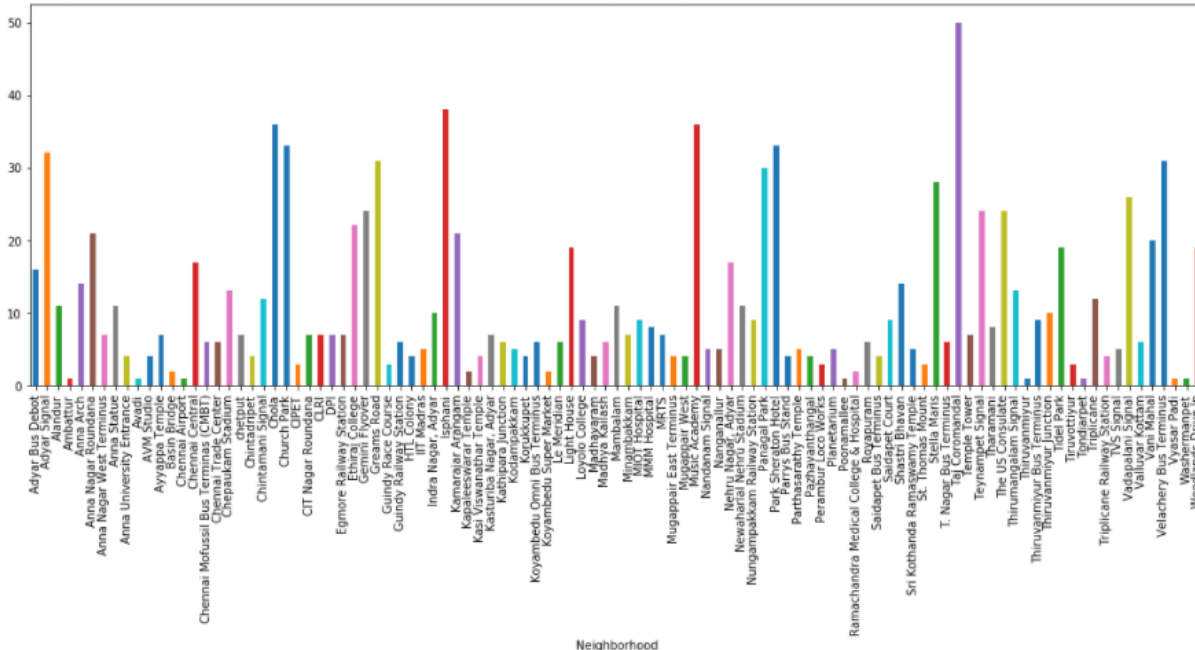
### 3. Methodology

Post the steps above, we have obtained the neighbourhoods data of Chennai (105). We also have the most popular venues in each neighbourhood obtained using Foursquare API. A total of 1130 venues have been obtained in the whole city and 145 unique categories. But as seen we have multiple neighbourhoods with less than 10 venues returned. In order to create a better analysis let us consider only the neighbourhoods with more than 10 venues.

We can perform one hot encoding on the obtained data set and use it find the 10 most common venue category in each neighbourhood. Then clustering can be performed on the dataset. Here K - Nearest Neighbour clustering technique have been used. To find the optimal number of clusters silhouette score metric technique is used. The clusters obtained can be analysed to find the major type of venue categories in each cluster.

## 4. Analysis

First, lets visualize the number of venues obtained in all the neighborhoods of Chennai.



As we can see from the above bar chart, there are many neighborhoods with less than 10 venues which can be remove 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.

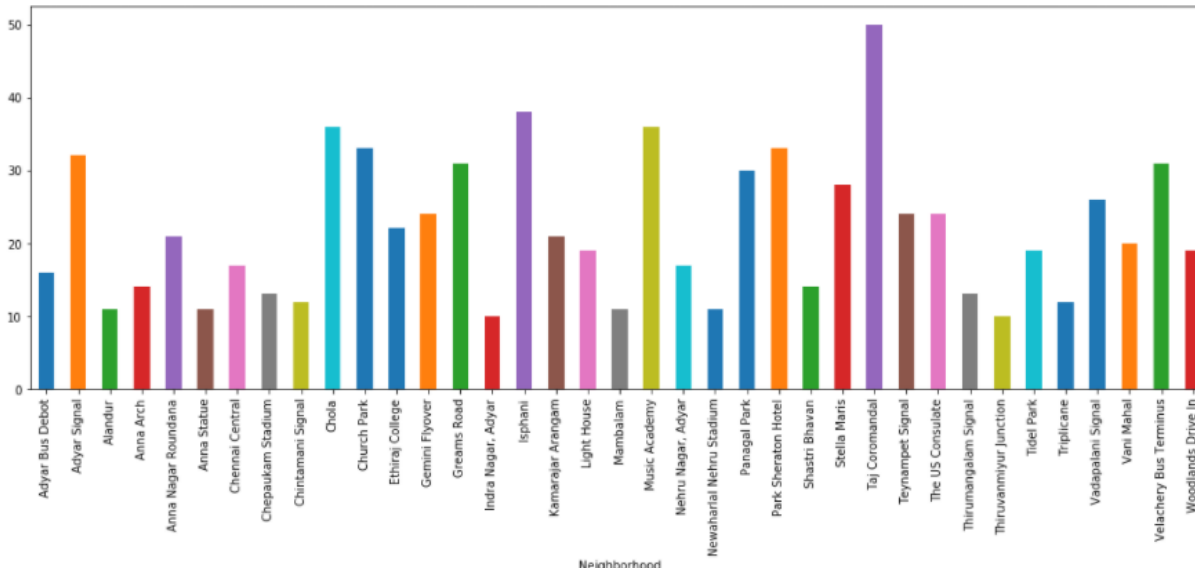
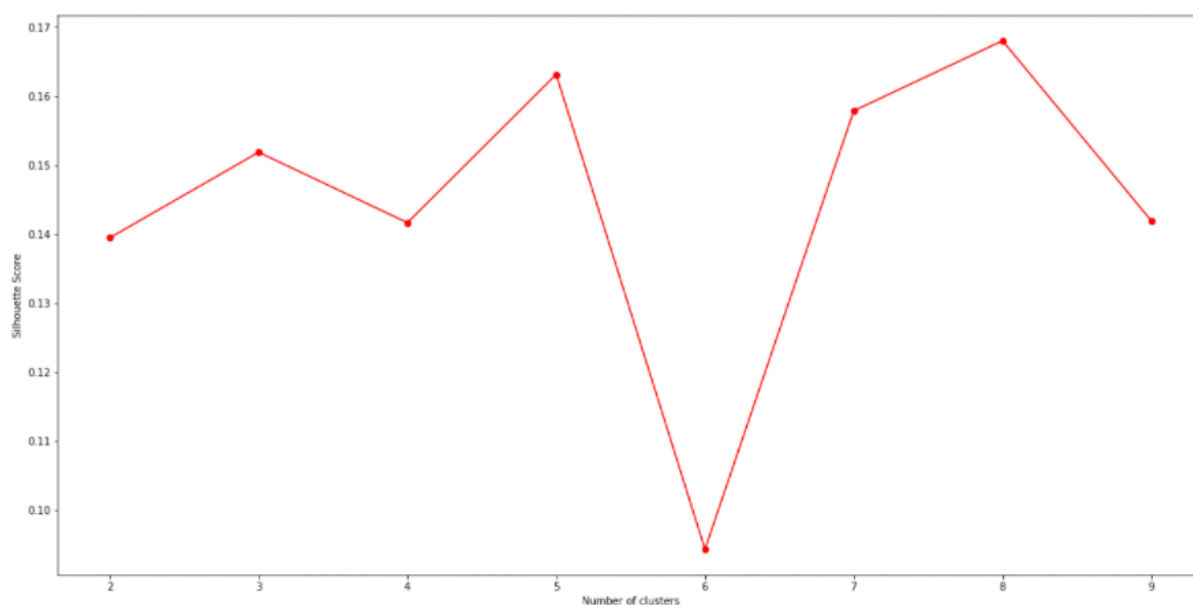


Fig 4.1 Filtered Neighborhood Dataset

Next, used a function to obtained the 10 most common venues in each neighborhood and store in the new pandas dataframe.

	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	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Pizza Place	Sandwich Place	Breakfast Spot	Fruit & Vegetable Store	Restaurant	BBQ Joint	Middle Eastern Restaurant
1	Adyar Signal	Indian Restaurant	Electronics Store	North Indian Restaurant	Coffee Shop	Rock Club	Dessert Shop	Bookstore	Lounge	Café	Shoe Store
2	Alandur	Indian Restaurant	South Indian Restaurant	Hotel	Bus Station	Bus Line	Bar	Metro Station	Airport	Gym	Grocery Store
3	Anna Arch	Fast Food Restaurant	Clothing Store	Electronics Store	Mediterranean Restaurant	Café	Multiplex	Pub	Bookstore	Scenic Lookout	Shopping Mall
4	Anna Nagar Roundana	Indian Restaurant	Chinese Restaurant	South Indian Restaurant	Clothing Store	Paper / Office Supplies Store	Café	Electronics Store	Fast Food Restaurant	Middle Eastern Restaurant	Bookstore

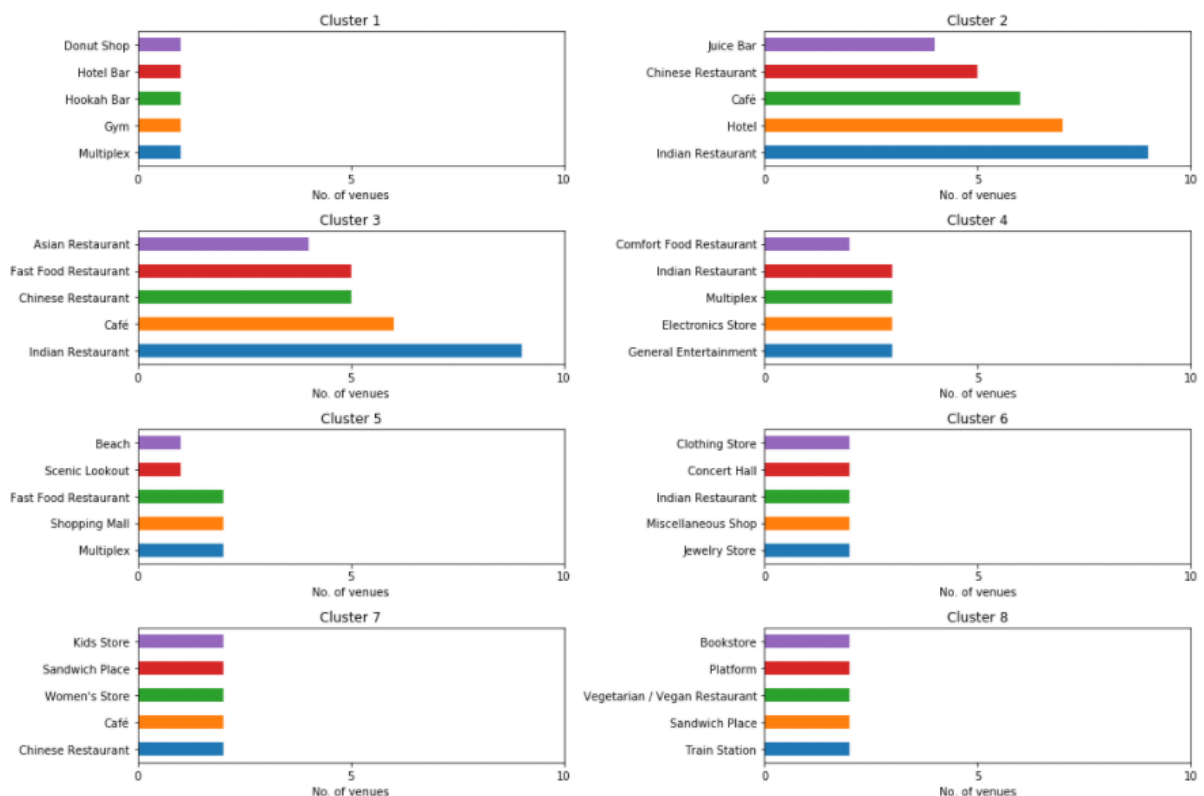
**Cluster Neighborhoods:** Here k-Nearest Neighborhoods clustering technique is used. Lets use the silhouette\_score to obtain the best value for the number of clusters.



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Cluster Labels	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
0	Adyar Bus Debot	12.997222	80.256944	2	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Pizza Place	Sandwich Place	Breakfast Spot	Fruit & Vegetable Store	Restaurant	BBQ Joint
1	Adyar Signal	13.006389	80.257500	2	Indian Restaurant	Electronics Store	North Indian Restaurant	Coffee Shop	Rock Club	Dessert Shop	Bookstore	Lounge	Café
2	Alandur	13.007778	80.209722	1	Indian Restaurant	South Indian Restaurant	Hotel	Bus Station	Bus Line	Bar	Metro Station	Airport	Gym
3	Anna Arch	13.074444	80.218333	4	Fast Food Restaurant	Clothing Store	Electronics Store	Mediterranean Restaurant	Café	Multiplex	Pub	Bookstore	Scenic Lookout
4	Anna Nagar Roundana	13.084444	80.218056	2	Indian Restaurant	Chinese Restaurant	South Indian Restaurant	Clothing Store	Paper / Office Supplies Store	Café	Electronics Store	Fast Food Restaurant	Middle Eastern Restaurant

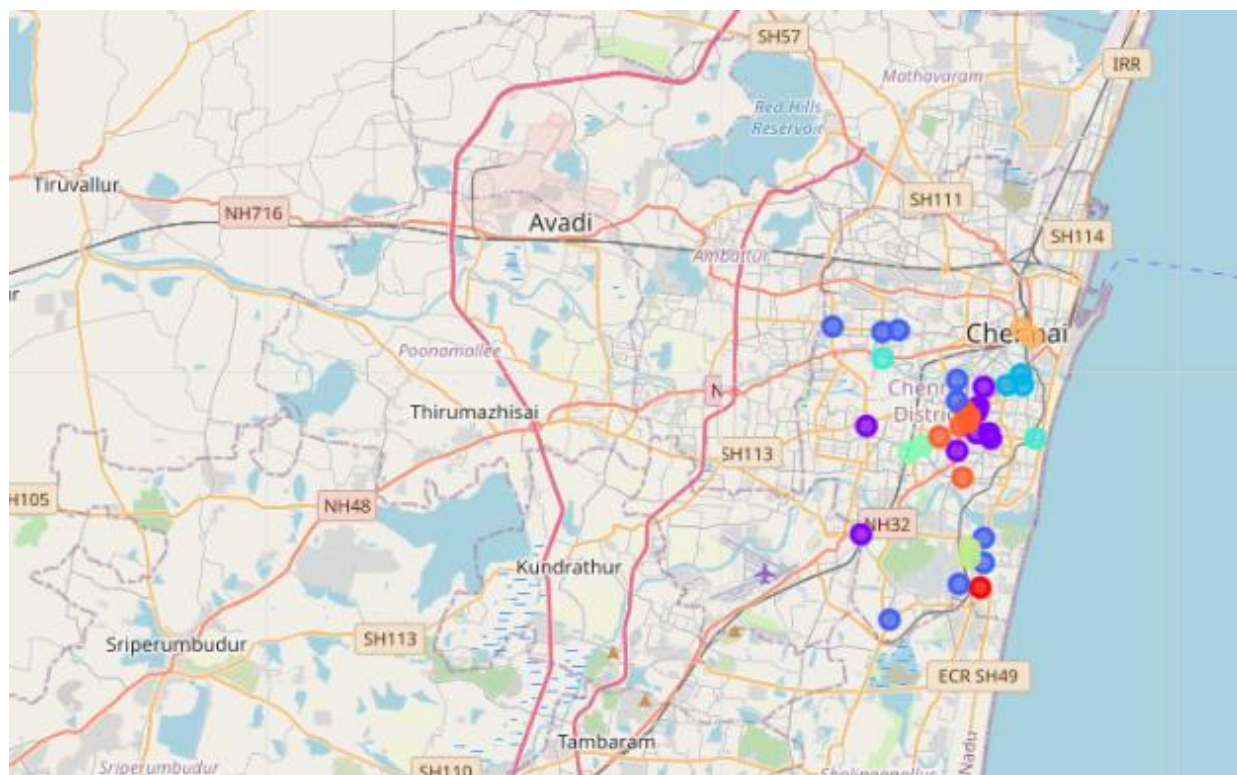
## 5. Results

I examined the clusters and find the discriminating venue categories that distinguish each cluster. For this purpose lets also print the ten most common venue category in each of the 8 clusters.



The plots post cluster analysis can be used to suggest valuable information to my friend. 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 at the neighborhoods in cluster 7 or 8. Not likely, since the place has a smaller number of food restaurants. Thus, an optimal place would be one which has relatively less restaurants. Considering all these facts, the best choice would be Cluster 3 and Cluster 4. Places such as the Adyar Bus Depot, Triplicane neighborhoods are the ones suggested by me to my friend for his use case of opening a restaurant in Chennai.

### Map of Chennai with the clusters



## 6. Conclusion

The main aim of this project was to analyse the neighbourhoods of Chennai and create a clustering model to suggest the most appropriate place to start a new restaurant business to my friend. The neighbourhoods data was obtained from an online source and the Foursquare API was used to find the major venues in each neighbourhood.

We found that many neighbourhoods 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 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 (hotels, restaurants or shopping malls).

Finally, given all the analysis above: places such as the Adyar Bus Depot, Triplicane neighbourhoods are the ones suggested by me to my friend for his use case of opening a restaurant in Chennai.