

## **Coursera Capstone Project**

### **Utilizing Foursquare API to analyze Chennai's Neighborhoods**

#### **Introduction:**

Chennai is the capital of Tamilnadu and it is one of the metropolitan cities of India. Chennai is the 6th Largest city in India with the population of 9,714,000 as of 2015 with Land area of 971 km<sup>2</sup>. Diversity of cuisine available is reflective of social and economic diversity of Chennai.

The aim of project is to find neighborhoods with a high frequency of restaurants/food stalls/café's in Chennai, India. Chennai is one of the Metropolitan City in India. To find popular restaurants or food stalls or cafes, respective coordinates need to be retrieved, so that Foursquare can be used to find nearby venues.

#### **Problem Description:**

Chennai has many Ancient temples like Triplicane parthasarathy temple, Mylapore Sivan temple and Hertiage sites like Mahabalipuram, St.Geroge Fort etc. Due to these, many foreign tourists across globe and people from various parts of India are visiting Chennai daily. The main problem they are facing is good lodging and food.

In such situation, they mainly relied on Internet and searching various restaurants nearby their lodging and there are few questions to be addressed:

1. How many types of foods available in the restaurant?
2. Which is the most nearest me having good rating?
3. How many "similar" restaurants are available nearby?
4. Do the "similar" restaurants cost more? If so what is the specialty they will have?

Expectation from proposed recommender system is to address these questions and also to recommend new places based on their rankings.

#### **Target Audience:**

Target Audience is not particular to a tourist who are new to Chennai city but also to everyone who is living in Chennai city also since no one aware of popular restaurants in all the places. People who rarely visits restaurants would prefer to have the most rated restaurants nearby them and all this could be handled by our recommender system.

## Data

The Foursquare API is used to access the venues in the neighborhoods. Since, it returns less venues in the neighborhoods, we would be analyzing areas for which countable number of venues are obtained. • Then they are clustered based on their venues using Data Science Techniques. Here the k-means clustering algorithm is used to achieve the task. The optimal number of clusters can be obtained using silhouette score metrics. •

Folium visualization library can be used to visualize the clusters superimposed on the map of Chennai city. These clusters can be analyzed to help small scale business owners select a suitable location for their need such as Hotels, Shopping Malls, Restaurants or even specifically Indian restaurants or Coffee shops.

Chennai has multiple neighborhoods. The chennaiiq.com website has a dataset which has the list of locations in Chennai along with their Latitude and Longitude. In order to obtain the venue details in each neighborhood Foursquare API is used. • [https://chennaiiq.com/chennai/latitude\\_longitude\\_areas.asp](https://chennaiiq.com/chennai/latitude_longitude_areas.asp) • <https://foursquare.com/> There is a total of 105 neighborhoods. But the Latitude and Longitude data obtained are in Degrees Minute Seconds format which needs to be converted to Decimal Degrees Format. The details obtained from Foursquare API are Venue, Venue Latitude, Venue Longitude and Venue Category. A total of 1130 venues data have been obtained from Foursquare.

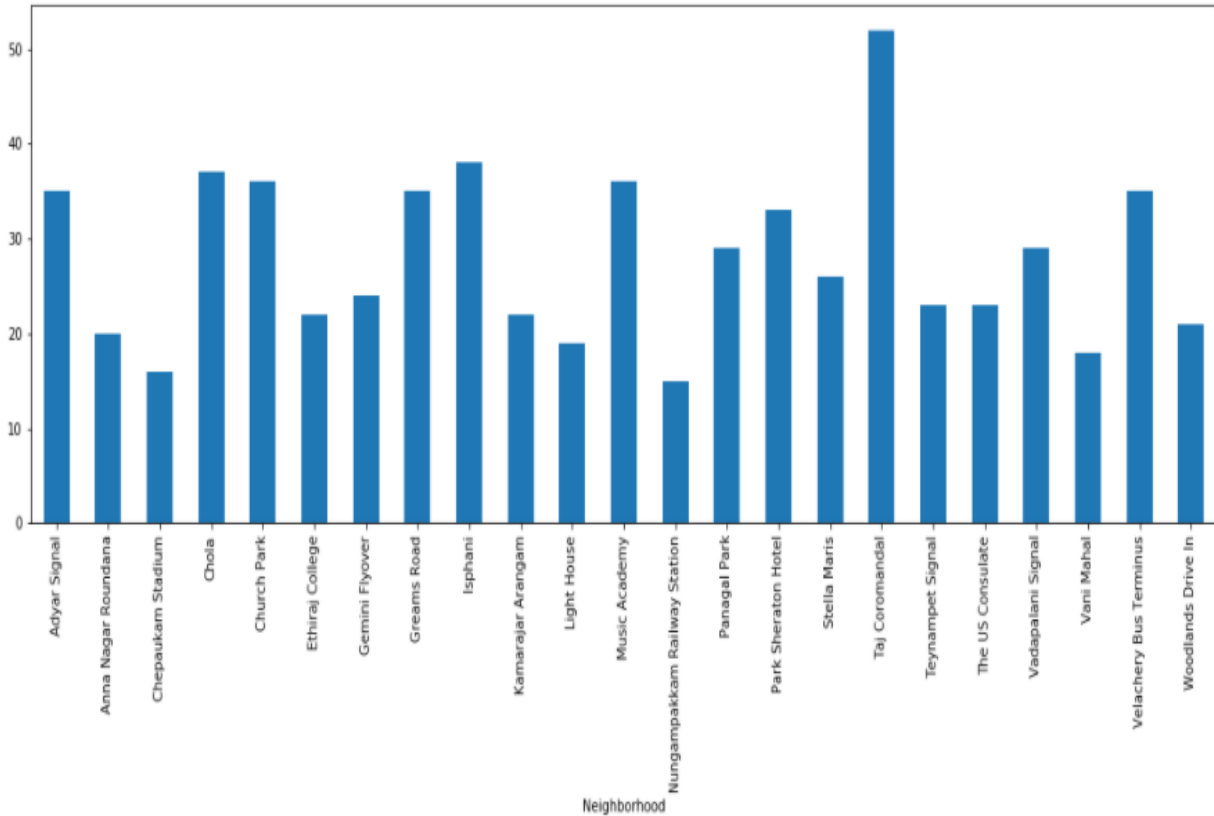
## Methodology

Now, we have the neighborhoods data of Chennai (105 neighborhoods). We also have the most popular venues in each neighborhood 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 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 find the 10 most common venue category in each neighborhood. Then clustering can be performed on the dataset. Here K - Nearest Neighbor clustering technique have 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, suitable locations based on the category.

## Analysis

Looking into the dataset we found that there were many neighborhoods with less than 15 venues which can be remove before performing the analysis to obtain better results. The following plot shows only the neighborhoods from which 15 or more than 15 venues were obtained.



One hot encoding is performed on the filtered data to obtain the venue categories in each neighborhood. Then group the data by neighborhood and take the mean value of the frequency of occurrence of each category.

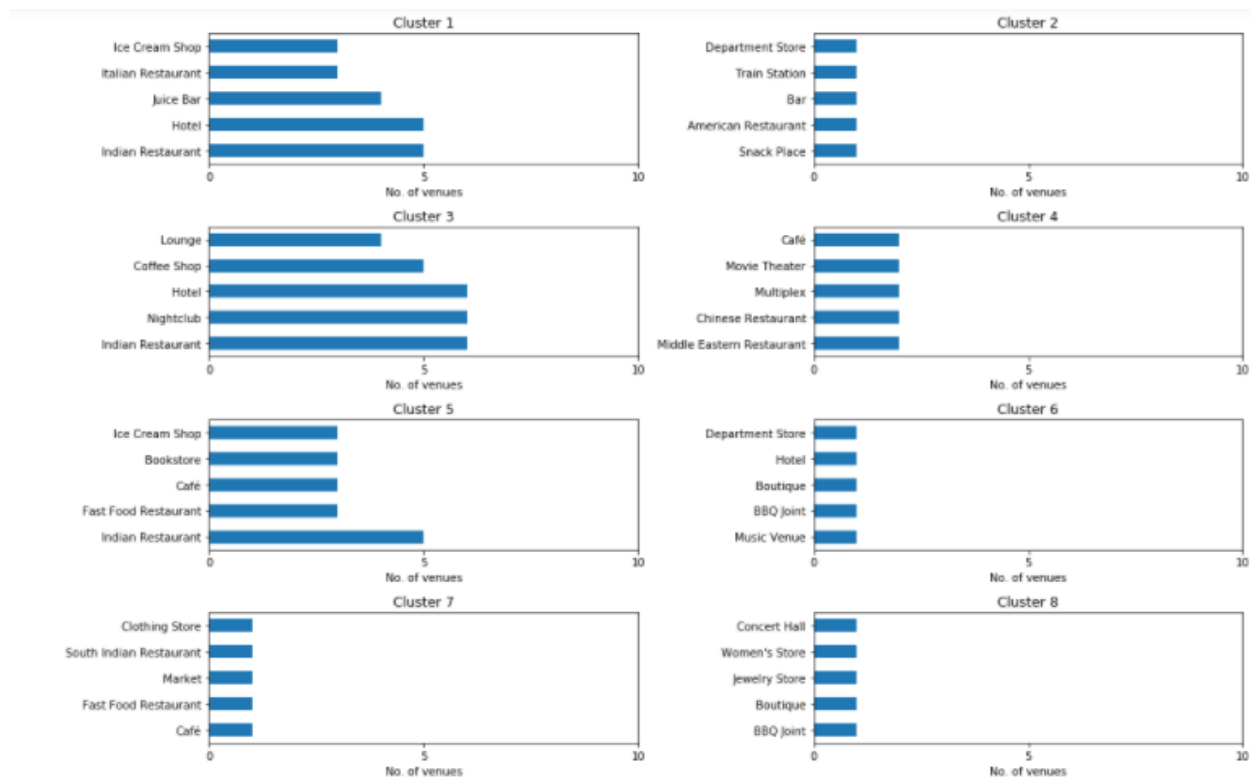
This is used to obtain the top 15 most common venues in each neighborhood i.e. the 15 venues with the highest mean of frequency of occurrence.

The resultant dataset can be used for the clustering algorithm. Here, the K-Nearest Neighbor (KNN) clustering algorithm is used. It is an unsupervised machine learning technique that clusters the given data into K number of clusters.

For optimal result we need to select the best value for K. Here, the silhouette score is used to find the best value for K.

Below final plot can be used to find Restaurants and malls available in particular area. Cluster 1 - Areas : Alwarpet around, Teynampet Signal around , Cluster 3 - Areas : Gemini Flyover, around Nungambakkam and Cluster 5 - Areas : Around Adyar Signal, Anna Nagar Roundtana, Velachery bus terminus, Chepauk Stadium are having variety of food outlets. So, One can visit these areas if they want to taste different varieties of foods.

Cluster 2, Cluster 7 and Cluster 8 are having very less variety of food outlets.



## Result:

With Chennai developing very fast and becomes IT hub, new restaurants are emerging quickly. We need a system that could help us access vast number of food varieties. It's impossible for a person to ask each and every one about their visit to a particular place and also not every one remembers everything.

On the other hand computers remembers everything and machine learning evolves to its peak, its high time technology will be our personal guidance and help us personally based on our likes and dislikes. So people would care about the project as their personal assistance and success rate could increase with time.

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. So one can easily find out the food outlets using this project but still we can see that foursquare is not upto date and in future this can be resolved using better data source with latest update.