



**CHENNAI**

# **THE BATTLE OF NEIGHBORHOODS**

**BY: BALKRISHNA MISHRA**

**DATE: 13 MAR 2021**

# TABLE OF CONTENTS

1. [Introduction: Business Problem](#)
2. [Data Gathering](#)
3. [Problem Solving Methodology](#)
4. [Analysis & Results](#)
  - A. [Analysis](#)
  - B. [Results](#)
5. [Discussion](#)
6. [Conclusion](#)

## 1. INTRODUCTION: BUSINESS PROBLEM

This project deals with discussing the neighborhoods of Chennai, Tamilnadu in India. This project would specifically help Business community planning to start Restaurants, Hotels, Café etc. in Chennai

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. Folium visualization library can be used to visualize the clusters superimposed on the map of Chennai city. These clusters will be further analyzed to help business owners selecting a potential location to open-up Hotels, Shopping Malls, Restaurants or Coffee shops.

## 2. DATA GATHERING

Chennai has multiple potential neighborhoods. Let's use the following dataset with data wrangling and beautification techniques to produce meaningful information.

### *LOADING DATA FROM VARIOUS SOURCES*

---

Chennai Neighborhood data with their Latitude and Longitude from below path and then converting html data into a data frame (table).

[https://chennaiiq.com/chennai/latitude\\_longitude\\_areas.asp](https://chennaiiq.com/chennai/latitude_longitude_areas.asp)

But the Latitude and Longitude data provided here is in Degrees Minute Seconds format. Thus before starting the analysis, the location data has to be converted to Decimal Degrees. Let's first, review the html file.

A response value of **200** shows that the process was successful. Now we need to clean the file to access the Neighborhoods and their Locations. Let's use BeautifulSoup library to parse the html file in order to make it easier to access.

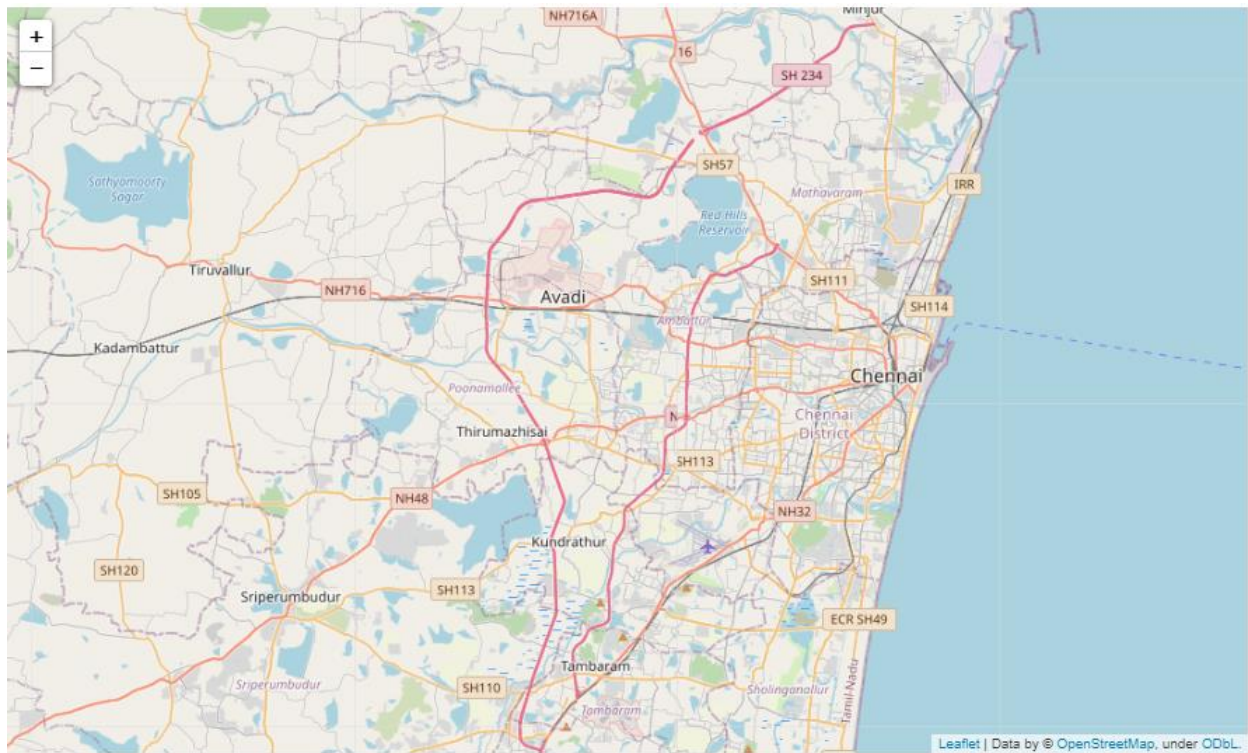
The html has been parsed and it can be used to build the dataset. Let us access the Location data and store it in a pandas data frame, `chennai_data`.

### *DATA WRANGLING*

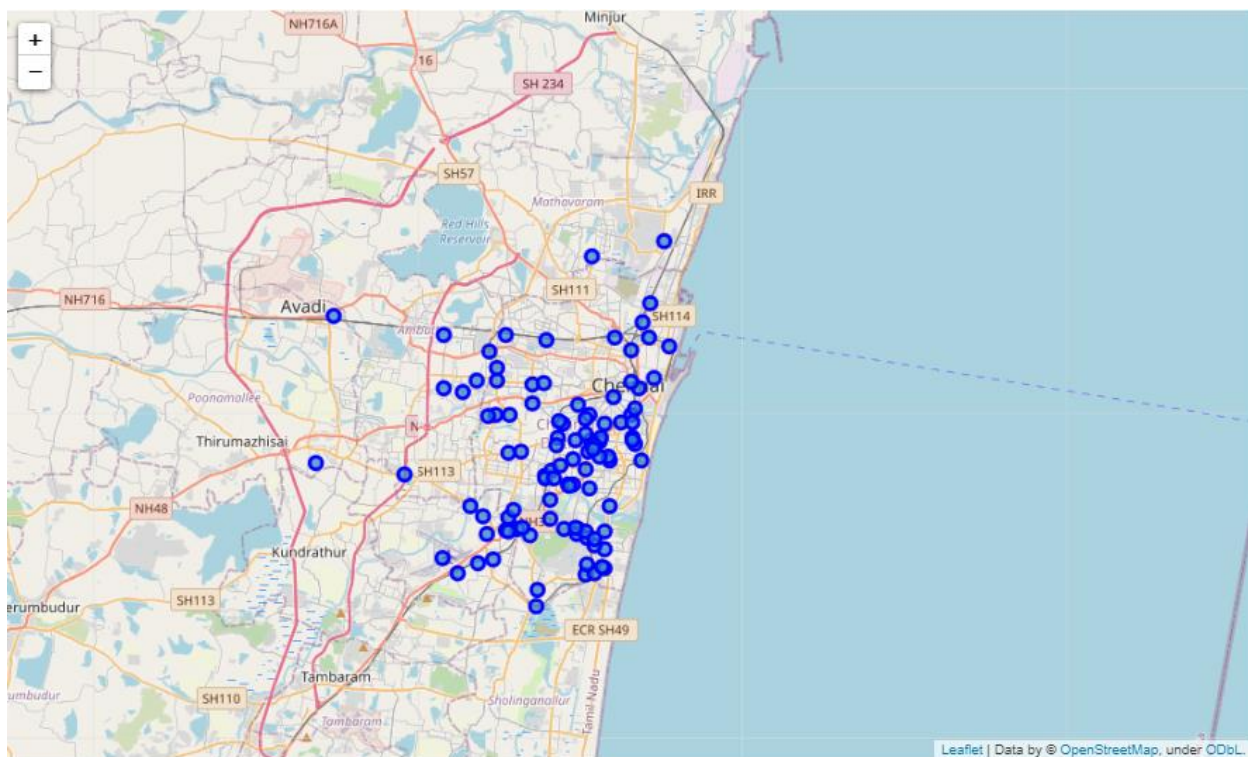
---

Convert Degrees Minute Seconds to Decimal Degrees using the `dms2dd` function defined below will solve this problem.

## VISUALISING CHENNAI MAP



## CHENNAI MAP WITH THE NEIGHBORHOODS MARKERS IN BLUE



## EXPLORE NEIGHBORHOODS IN CHENNAI WITH FORESQUARE

We will send a explore request for each neighborhood and return the 100 most popular places in the neighborhood around 500 meters or 1/2 KM.

	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	Kuttanadu Restaurant	12.997010	80.257799	Asian Restaurant
2	Adyar Bus Debot	12.997222	80.256944	Zha Cafe	12.999730	80.254806	Café
3	Adyar Bus Debot	12.997222	80.256944	Kovai Pazhamudir Nilayam	12.996522	80.259776	Fruit & Vegetable Store
4	Adyar Bus Debot	12.997222	80.256944	Barbeque N Biryani	12.995907	80.256011	BBQ Joint

A total of **1167** venues were obtained from the above data frame and there are 148 unique categories.

We can see that **Taj Coromandal** returned the highest number of venues i.e. 51. Now let's check the unique categories of all the venues returned.

## 3. PROBLEM SOLVING METHODOLOGY

Now, we have the neighborhoods data of Chennai (i.e. total 105 neighborhoods). We also have the most popular venues in each neighborhood obtained using Foursquare API. A total of **1167 VENUES** have been obtained in the whole city and **148 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.

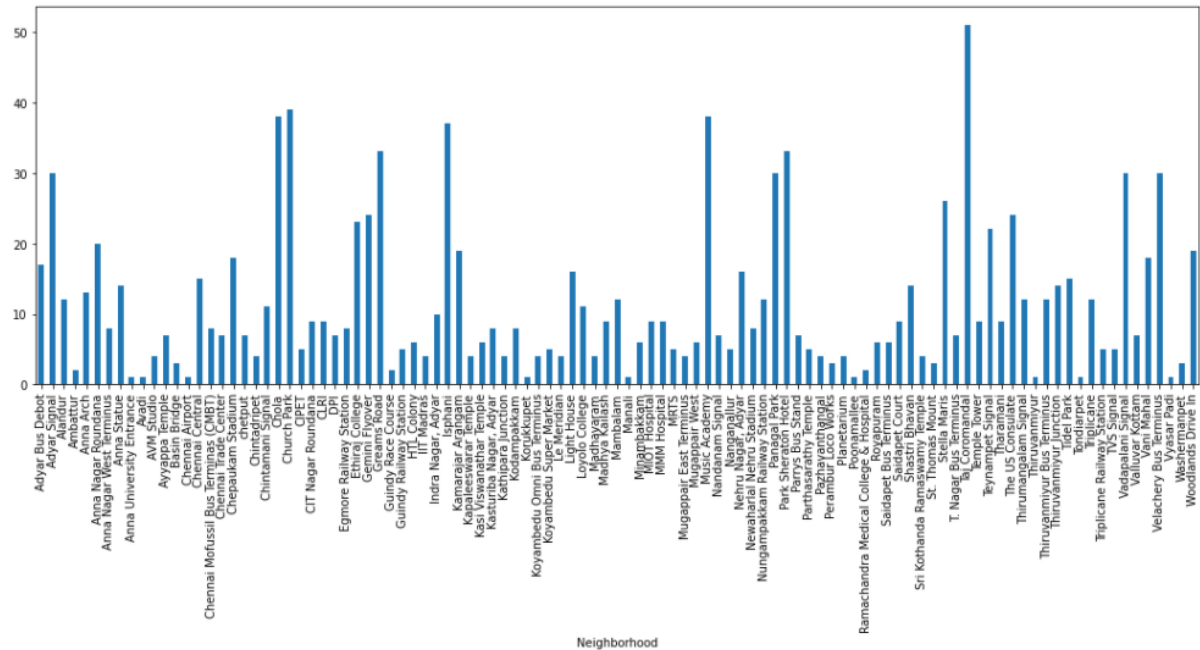
We can perform one hot encoding on the obtained data set and use it find the 10 most common venue category in each neighborhood. Then clustering can be performed on the dataset. K - Nearest Neighbor clustering technique have been used and to find the optimal number of clusters silhouette score metric technique is used. The clusters obtained can be analyzed to find the major type of venue categories in each cluster. The outcome then can be used to suggest business people, potential locations based on the category.



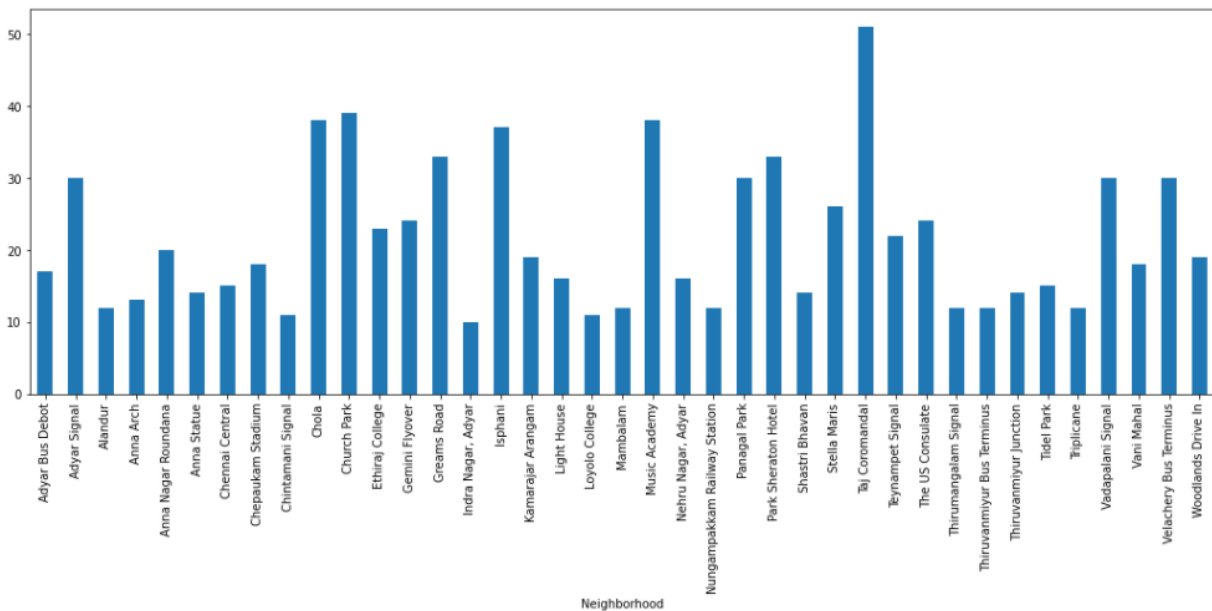
## 4. ANALYSIS & RESULTS

### 4.1 ANALYSIS

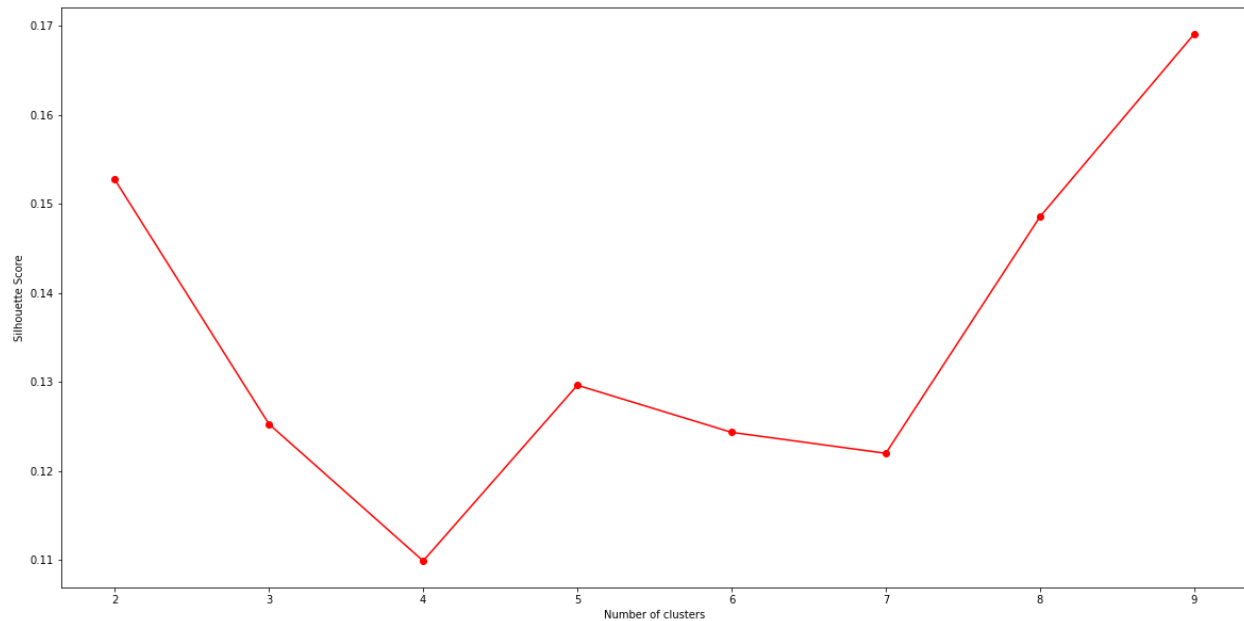
Let's visualize the number of venues obtained in all the neighborhoods of Chennai.



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

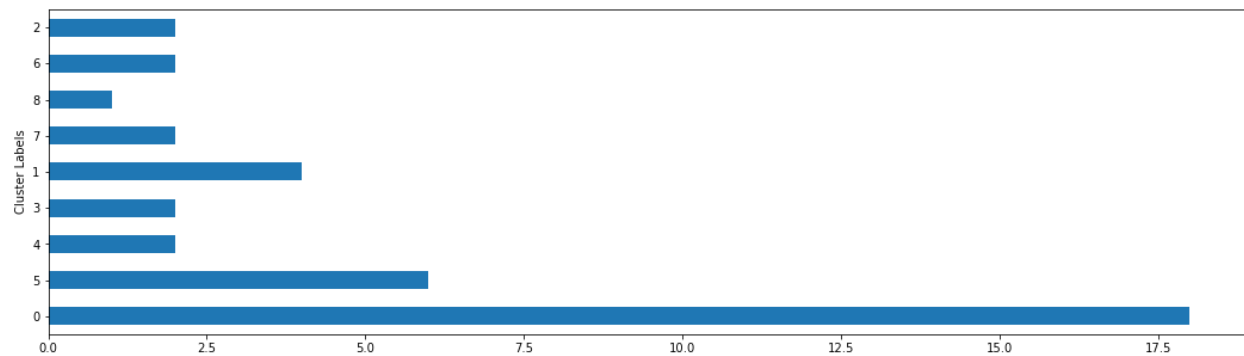


Cluster Neighborhoods Here k-Nearest Neighborhoods clustering technique is used. I will use the **silhouette\_score** to obtain the best value for the number of clusters.



As seen from the above line plot, the best number of clusters having the **highest silhouette score is 9**. So, let's consider the number of clusters as 8.

We can also visualize the density of Neighborhood by cluster label in a chart below.



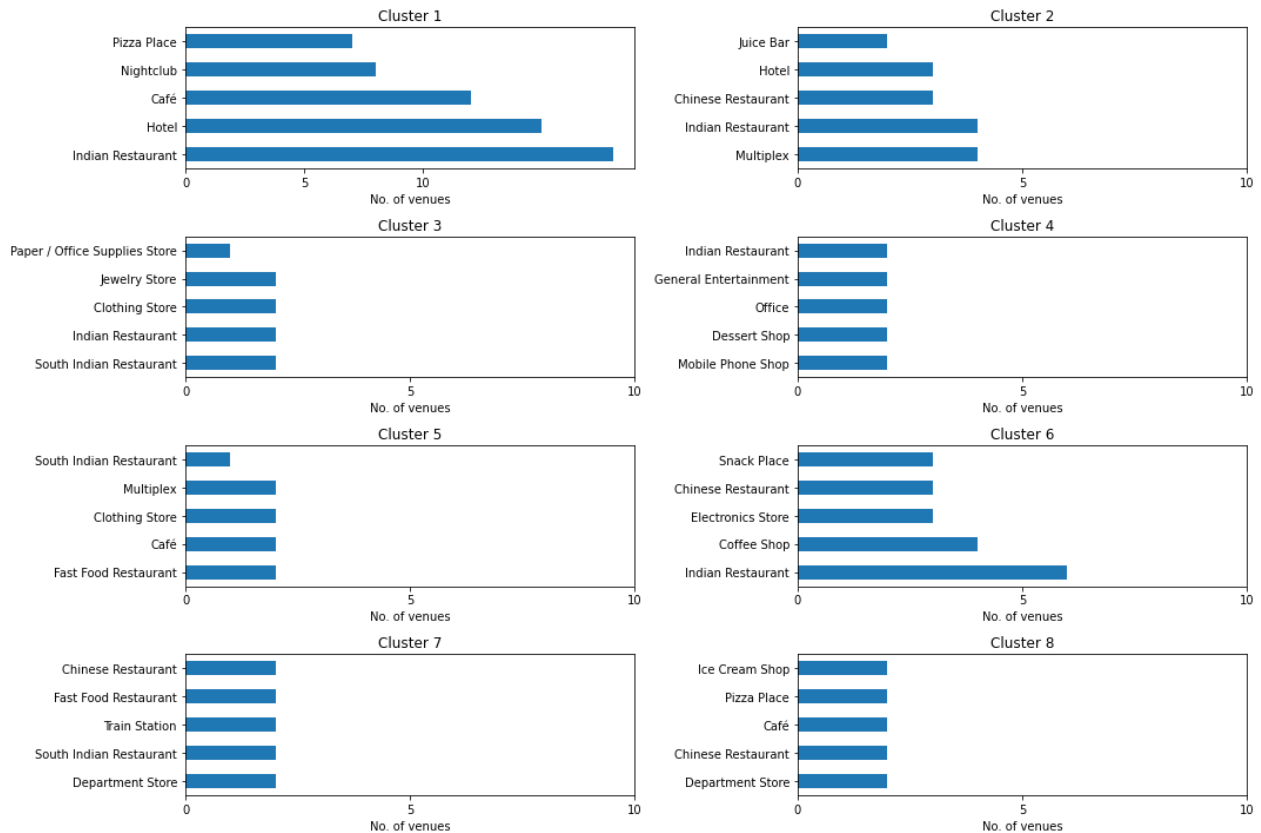
As we can see in above data frame **Cluster 0** Contains the highest cluster density. We need to find the geographic centroid for this cluster. This is the most suitable location for a new business set-up.

With the help of reverse lookup, we can find the best suitable Neighborhood to locate a New Business is in: ['Teynampet Signal'] while we know the geographical coordinate of Chennai are 13.0836939, 80.270186.

---

## 4.2 RESULTS

We need to dig more to find out complimentary business set-ups in each clusters in order to suggest a most suitable cluster for new business set-up. So let's examine the clusters and find the discriminating venue categories that distinguish each cluster. For this purpose let's also print the ten most common venue category in each cluster.



These above plots 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.

## 5. DISCUSSION

Let us discuss about appropriate business opportunity by Venue Category for certain targeted business set-ups.

### A. HOTELS

The neighborhoods in **cluster 1** has the greatest number of hotels, Thus, an optimal place would be one which has less hotels, but also have restaurants and other places to explore. Considering all these facts, the best choice would be Cluster 6 and Cluster 7 as it has some local cuisine restaurant whihc would attract tourists who would love to experience these.

### B. MULTIPLEX

The neighborhoods in **cluster 2 and 5** has notable number of Multiplexes. By using the same procedure as above, the suitable cluster would be the Cluster 1 and since it has not much shopping malls and also it has many Hotels and Restaurants which gives an advantage.

Similarly, based on the requirement suggestions can be provided about the neighborhood that would be best suitable for the business.

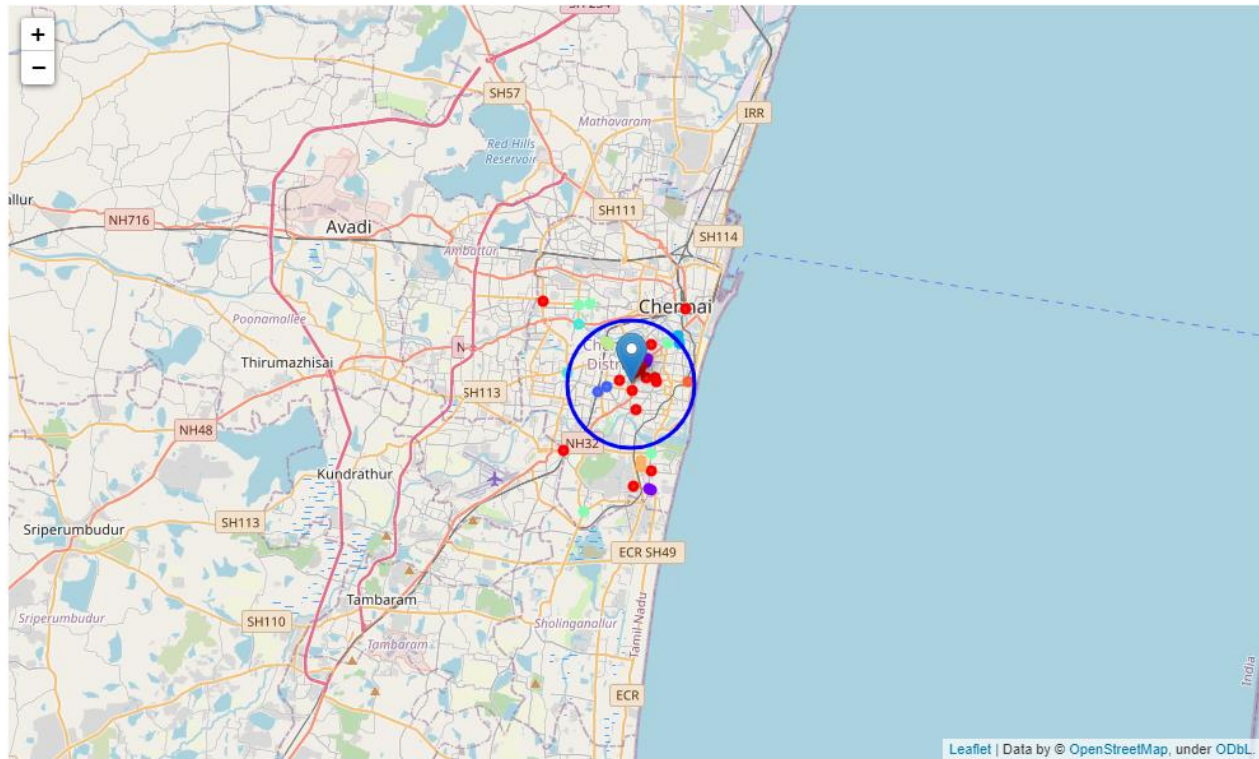
### C. BUISNESS STORES (Retail/Departmental/Wholesale)

The neighborhoods in **cluster 3** has the greatest number of stores, Thus, an optimal place would be one which has less stores, but also have other attractions and other places to explore. Considering all these facts, the best choice would be Cluster 1 and Cluster 2 as it has some local cuisine restaurant whihc would attract tourists who would love to experience these.

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 Venue category.

I have added a most suitable location on CHENNAI map based on above analytics for a new business set-up



This map can also be used to select a vast suggestion area for a particular type of business based on the category.

## 6. CONCLUSION

Objective of this project was to analyze the neighborhoods of Chennai and create a clustering model to suggest potential 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 an appealing recommendation 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 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.