Predicting the suitable location to open a Restaurant in Chennai Neighborhood

Rameshkumar Shanmugam

January, 2020

## 1. Introduction

1.1. Background

The Indian Restaurants and Food Services Industry has continued to expand at a healthy pace, aided by year on year growth in the incomes and largely unaffected by the prevalent economic scenario that has slowed growth in sectors like manufacturing and infrastructure. The market size of the Indian restaurants and food services industry stands at Rs 3.7 trillion as of 2018 registering a year on year growth of about 10% and a CAGR growth of 8.4% between 2013 and 2018

Chart 1: Indian Restaurant & Food Service Industry
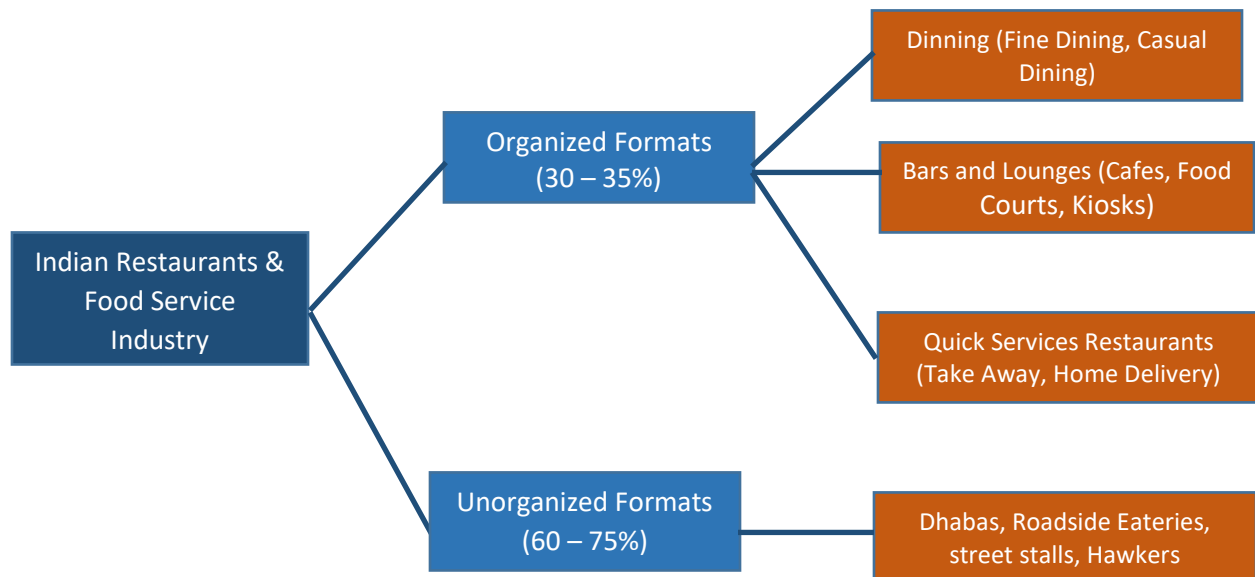
Source: Company presentations, Industry

Chart2: Indian Organized Food market segmentation – FY19 (%)
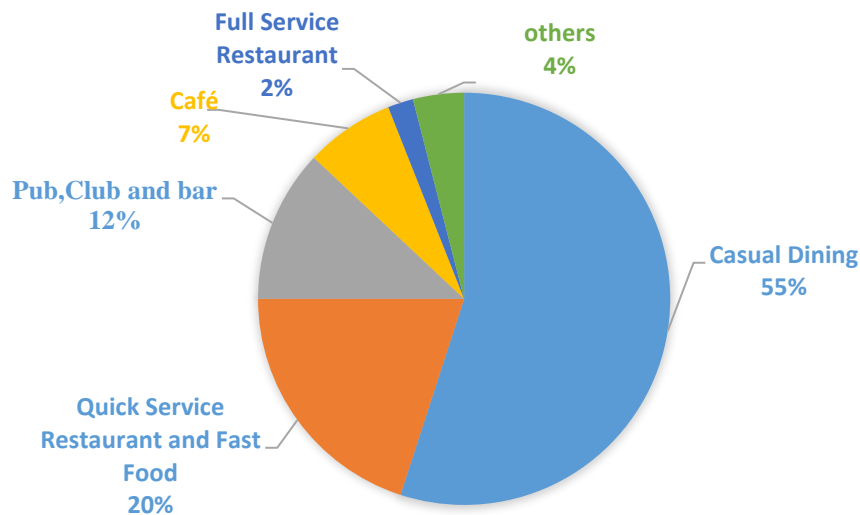
Source: FICCI, Company presentations

There are a lot of good restaurants already exist in Chennai city. Opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurant is one of the most important decision that will determine whether the investor will be able to make a profit out of it or not.

1.2. Business Problem

The objective of this project is to analyze and select the best locations in Chennai City, India to open a new Restaurant. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the Chennai City, India if an Investor is looking to open a new restaurant, where would be the best place/location to open it?

1.3. Targeted audience of this project

 Targeted audience would be investors who is looking for opening a restaurant in Chennai city. There are good restaurants already available in prime locations in Chennai, it is hard to find the right place to open a new one. Based on the current market, CARE Ratings expects the restaurant and food service industry to register a growth of about 10.4% CAGR for the next 5 years between 2018 and 2022 to reach Rs 5.5 trillion by 2022.

2. Data requirements

In order to solve the given problem, we have to collect the following data

1. List of neighborhoods in Chennai, Tamil Nadu
2. Latitude and longitude coordinates of those neighborhoods
3. Venue data, particularly data related to restaurants

2.1. Source of data

Data has been collected from following sources to fulfill our data requirements

1. This Wikipedia page (https://en.wikipedia.org/wiki/List_of_neighbourhoods_of_Chennai) contains a list of neighborhoods in Chennai, Tamil Nadu with a total of 180 neighborhoods.
2. Geographical coordinates of the neighborhoods were collected using google map API which will give us the latitude and longitude coordinates of the neighborhoods.
3. Foursquare API calls are used to get the venue data for those neighborhoods.

2.2. Data cleaning

Data collection process was performed through web scraping and API calls. Chennai neighborhood data is not well formatted in wiki page and there are few duplicate neighborhoods found due to region conflict which has been removed. Formatting and cleaning process completed after web scraping

Google API calls are used to get the Latitude and longitude coordinates of those neighborhoods then it is combined with Chennai Neighborhood data. Performed a quick check if there are no null values and no duplicates in the data set

Once base data collected, foursquare venue API calls are used to get the venue details like venue name, Category, Latitude and longitude coordinates of these venues located in the Chennai neighborhood. We have set the limit to 100 and radius to 800 while extracting venue details from Foursquare API. Finally, venues are related to Restaurants filtered for data analysis and modeling

2.3. Feature selection

After data cleaning, there are 860 venues and 7 features available in the final data set. After examining venue details feature, there are three new features need to be created based on the nature/type of Restaurants.

Features are

- Casual and fine dining restaurants categorized as **Dinning**
- Fast food cuisines as well as minimal table services are considered as Quick service Restaurants (**QSRs**) and
- Cafes, Food courts , Kiosks and bars are considered as **Cafes**

The below represents sample data before applying feature selection and data scaling

Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Royapuram	13.1137	80.29541	Ibaco	13.115612	80.294837	Dessert Shop
Royapuram	13.1137	80.29541	Hotel Pandia's	13.107839	80.291133	Indian Restaurant
Korukkupet	13.11855	80.278	Adayar Ananda Bhavan	13.11811	80.284481	Vegetarian / Vegan Restaurant
Korukkupet	13.11855	80.278	aws chats and sandwich	13.113717	80.282494	Sandwich Place
Korukkupet	13.11855	80.278	Idli Vilas	13.117964	80.285028	Vegetarian / Vegan Restaurant

Table 1- sample data format before feature selection

3. Exploratory Data Analysis

After introducing the three new features, Quick data analysis has been done to understand the distribution of data available in the data set and to understand the characteristics of the neighborhood as well
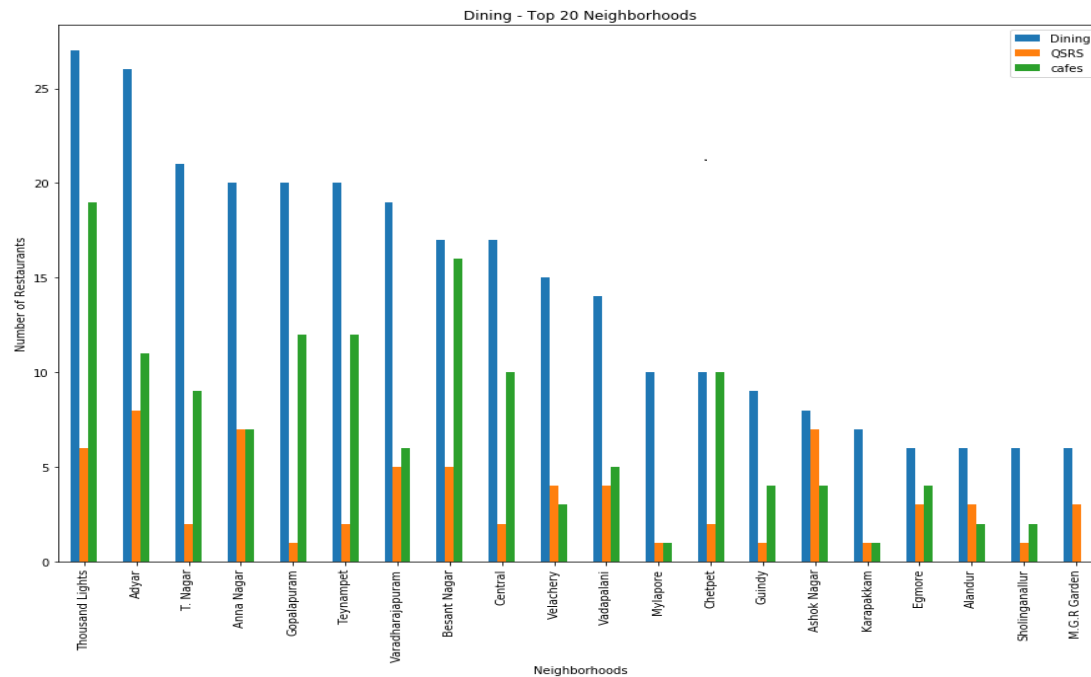

Figure 1 – Dinning top 20 neighborhoods

There are close to 6 neighborhoods which are having more than 20 dining/casual dining Restaurants. Those areas are **Thousand lights, Adyar, T.Nagar, Anna Nagar, Gopalapuram and Teynampet** . Cafes and lounges type of also bit populated in this Neighborhoods
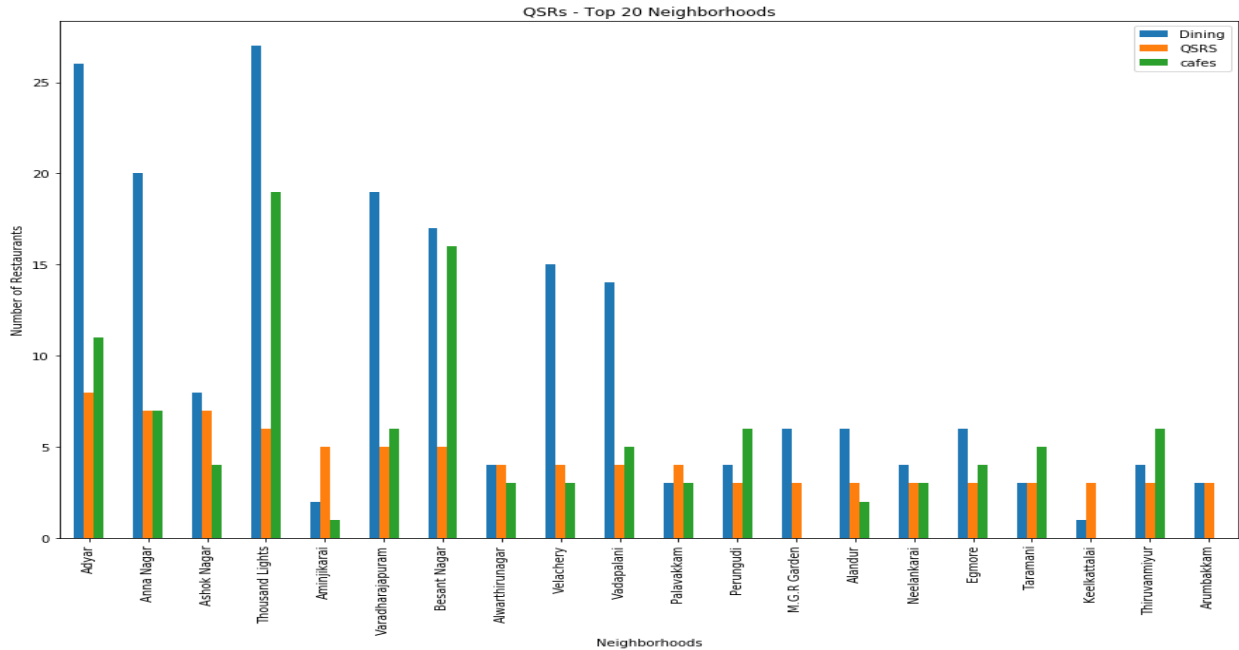
Figure 2 – QSRs top 20 neighborhoods

Quick service Restaurants are slowly picking up in Chennai, it is not much populated like casual dining Restaurants. **Adyar,Anna Nagar,Ashok Nagar, Thousand lights and Aminjikarai** are having more than 5 QSRs. Based on the initial analysis, it is observed that Adyar neighborhood is quite populated with Casual dining and QSRs as well . It is clearly indicates that this is not suitable place to start the restaurant
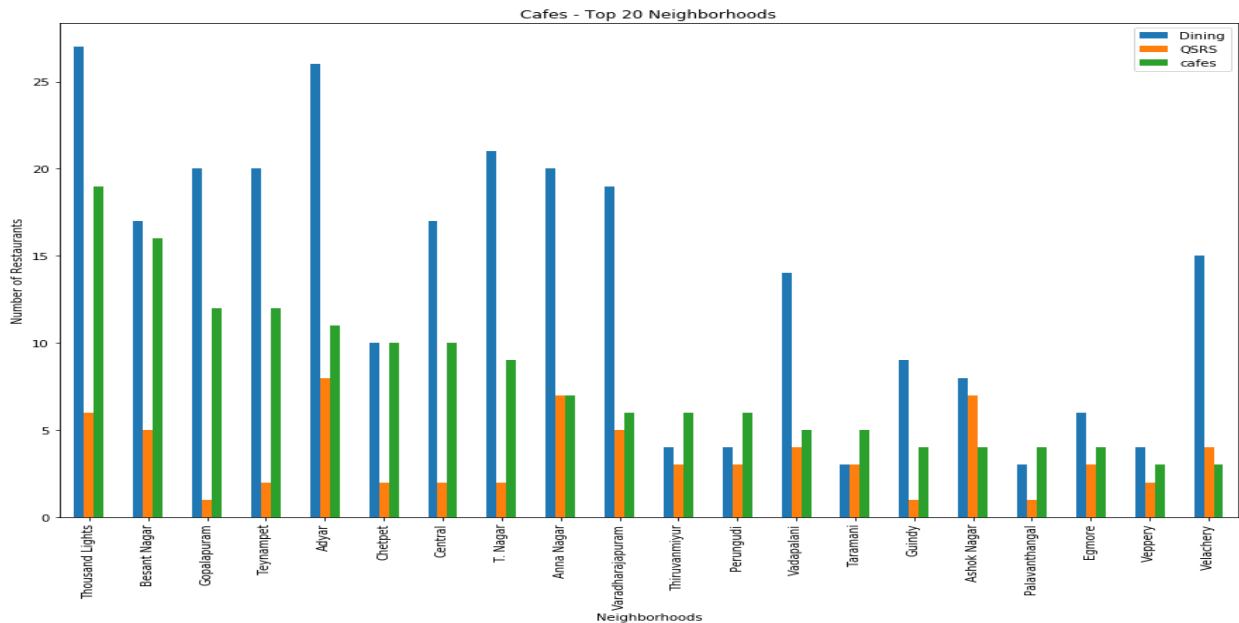

Figure 3 – Cafes top 20 neighborhoods

There are 5 neighborhoods having more than 10 cafes and lounges type restaurants. Those neighborhoods are **Thousand Lights, Besant Nagar, Gopalapuram, Teynampet and Adyar** there are 2 of them having

close 10 , those are **Chetpet and Chennai Central.** Based on this analysis, it is observed that **Thousand Lights and Besant Nagar** locality are having more number of casual dining and cafes as well.
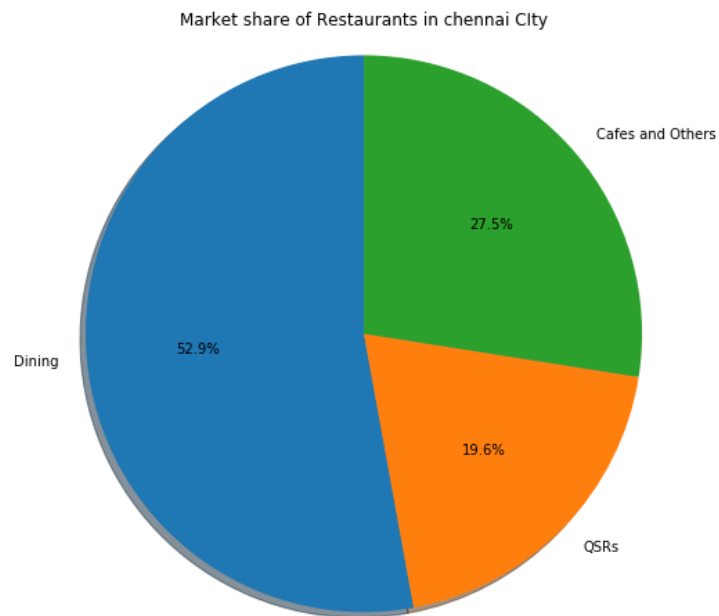

Figure 4 – Distribution/market share of restaurants in Chennai

It is observed that percentage of restaurants available in data set is similar to data represented in CAGR reporting. This data is sufficient to continue for further analysis and model building

4. Predictive modeling

After Data cleaning and EDA, Applied k-means clustering algorithm on our data set. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence for "Restaurants". The results will allow us to identify which neighborhoods have higher concentration of Restaurants while which neighborhoods have fewer number of Restaurants.

## 5. Results

The results from the k-means clustering show that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for "Restaurants":

- Cluster 0: Neighborhoods with high number of **Quick Service Restaurants(QSRs)** and less number of Causal dining and Cafes

- Cluster 1: Neighborhoods with high number of **Casual Dining** and less number of QSRs and Cafes

- Cluster 2: Neighborhoods with high numbers of **cafes** and there are Casual dining and QSRs available in this location.

The results of the clustering are visualized in the map below with cluster 0 in **Orange**, cluster 1 in **Green**, and cluster 2 in **Red**.
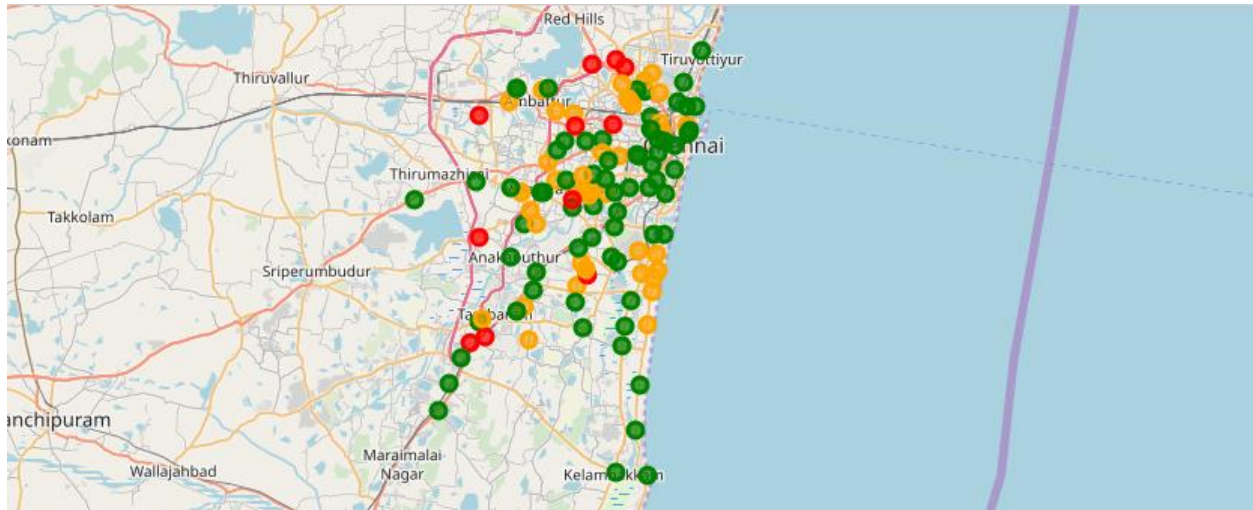
6. Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant investors who are looking for opening a restaurants in Chennai city. The answer to the proposed questions is: The neighborhoods listed in cluster 2(Refer Appendix) are the most preferred locations to open a restaurant like casual dining and QSRs. We have provided the recommendations where investors specifically open a casual dining type restaurants can refer cluster 0 neighborhoods. These locations do not have any casual dining type restaurants.

We also provided the neighborhoods which are potential to open a QSR type restaurants, these locations are listed in the cluster 1.

7. Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of Restaurants (Causal Dining, QSRs, Cafes), there are other factors such as population, income of residents in their neighborhood and lifestyle that could influence the location decision of opening a Restaurant. However, to the best knowledge of this researcher such data are not available to the neighborhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to opening a Restaurant. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

8. References

1. http://www.careratings.com/upload/NewsFiles/Studies/Restaurants%20%20QSRs%20May%202019.pdf
2. https://www.mbaskool.com/business-concepts/marketing-and-strategy-terms/17868-quick-service-restaurant-qsr.html

9. Appendix

Cluster – 0 (the following Neighborhoods are potential to start casual Dining restaurants)

Neighborhood	Latitude	Longitude
Alapakkam	13.049031	80.167278
Choolai	13.091873	80.264191
Iyyapanthangal	13.038062	80.135355
Iyyappanthangal	13.038062	80.135355
Kodungaiyur	13.137455	80.24785
Kolappakkam	13.010112	80.149232
M.G.R Nagar	13.035225	80.19727
M.K.B Nagar	13.125841	80.262221
Maduravoyal	13.064142	80.160788
Mowlivakkam	13.021493	80.144323
Nesapakkam	13.037938	80.191968
Pattalam	13.100144	80.261511
Pattravakkam	13.110261	80.167278
Perambur	13.12103	80.232578
Puzhuthivakkam	12.969784	80.194282
Selavoyal	13.144191	80.255632
Sembiam	13.115401	80.236662
T.V.K Nagar	13.119938	80.234223
Ullagaram	12.969784	80.194282

Cluster – 1 (The following Neighborhoods are potential to start QSRs type restaurants)

Neighborhood	Latitude	Longitude
Anakaputhur	12.98019	80.12646
Chembarambakkam	13.03186	80.03823
Defence Colony	13.02616	80.20215
Erukanchery	13.12773	80.24847
Gerugambakkam	13.01009	80.13908
Guduvancherry	12.84388	80.05974
Jamalia	13.10483	80.25332

Kanathur	12.8663	80.24468
Kelambakkam	12.78967	80.22158
Kosapet	13.09224	80.25513
Kovalam	12.787	80.25044
Kovilambakkam	12.94086	80.18506
MMDA Colony	13.0665	80.2147
Mogappair	13.08372	80.17496
Moolakadai	13.12957	80.24163
Mudichur	12.92337	80.09731
Muttukadu	12.82702	80.24032
New Washermenpet	13.13535	80.2837
Old Washermenpet	13.11482	80.28718
Pallavaram	12.96751	80.1491
Ponneri	13.334	80.19438
Poonamallee	13.04726	80.09453
Purasawalkam	13.09414	80.25311
Royapuram	13.1137	80.29541
Saidapet	13.02128	80.2231
Thirumazhisai	12.93243	80.13215
Thirumullaivayal	13.13074	80.13144
Thirumullaivoyal north	13.13074	80.13144
Tiruvottiyur	13.16426	80.30014
Triplicane	13.05871	80.27571
Urapakkam	12.86737	80.06986
Vandalur	12.89126	80.081

Cluster – 2(The following Neighborhoods are potential to start QSRs and Casual Dining type restaurants)

Neighborhood	Latitude	Longitude
Avadi	13.106745	80.096951
ICF Colony	13.098074	80.219502
Kathirvedu	13.152116	80.20006
Kundrathur	12.997668	80.0972
Madhavaram	13.14879	80.230559
Madipakkam	12.964746	80.196083
Manjambakkam	13.155141	80.222419
Padi	13.096454	80.184523
Peerkankaranai	12.909278	80.102376
Perungalathur	12.904759	80.089084
Ramapuram	13.031744	80.181672