## Introduction

Chennai, the capital city of South Indian state of Tamilnadu in India, is one of the largest cultural, economic, and educational centers of South India. The city is also 36th-largest urban area by population in the world (1). The traditional gateway of South India, Chennai is among the most-visited Indian cities by foreign tourists and ranked the 43rd-most visited city in the world for the year 2015 (2) and was ranked the 36th-most visited city in the world for the year 2019 (3). The Quality of Living Survey rated Chennai as the safest city in India (4). In such a big city, food is an important constituent of every household. To get good organic food, department stores are available in multiple locations across the city. Depending on the location of the user, different options are available.

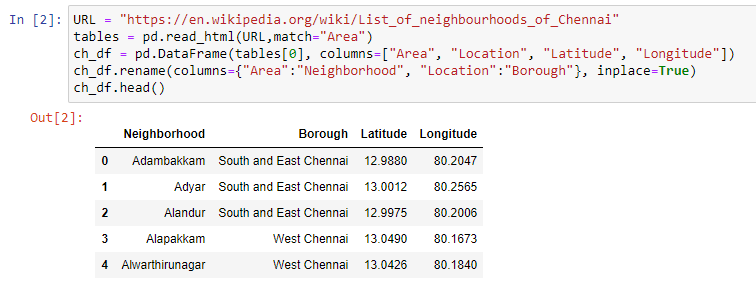
## Business problem

With such a variance of foodies, getting into a good department store is a nice business for people interested. City centers being densely populated and lot of stores that are already available are costly for common man. Our problem now is to find a best suitable place in Chennai (preferably South and East part), which is home to many households of middle class. People go to different parts of the city to work from these areas.

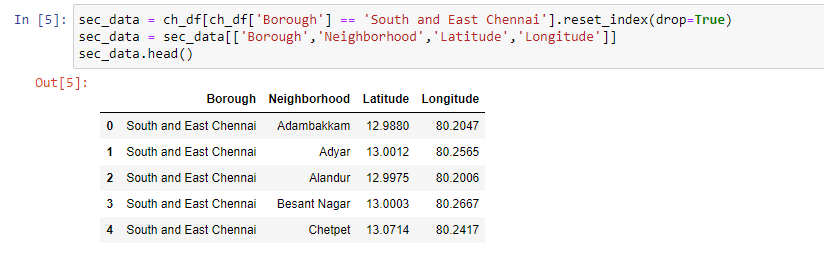
Creating a department store business in such a location will be of great use for the people and also for the stakeholders interested in getting good profit, as people coming from office and local households would be thronging to get the items of interest with low cost. Our task now is to identify the suitable location based on K-means clustering data from information gathered from data sources, cleaning it and analysing it, to get appropriate location for our business requirement.

## Data inputs

From the Chennai city neighborhoods (5), we get data from the required table with below code.



First 5 data in the dataset of the relevant data (South and East Chennai borough) gives below data.



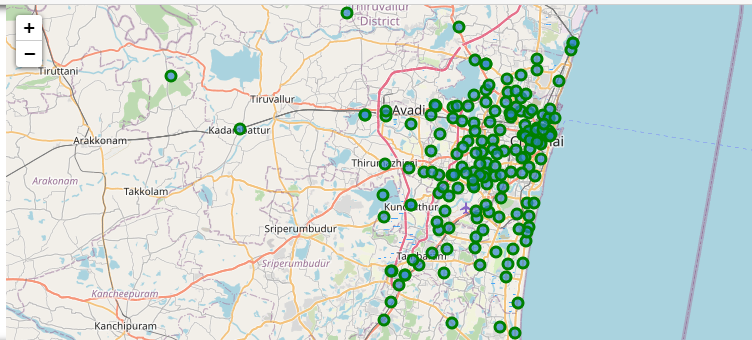
We also use FourSquare API’s (6) to get the map of the neighbourhood, from data of different venue category.

## Methodology

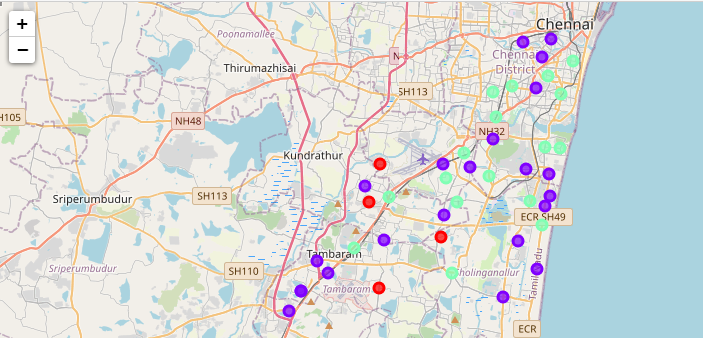
The typical methodology we use here would be like this:

* Get the data from the required table using web scraping
* Cleanup the data and get the required data and put it in pandas dataframe
* Since the data has latitude and longitude of whole city, with foursquare API, we get different venue category for each of the neighborhood of the borough (here we decided to go for South and East Chennai).
* With that neighbourhood based on category, using k-means clustering, cluster the data to see where or which cluster would be the appropriate place for our new department store.

Below diagram of the Four-Square API, shows the neighborhoods of Chennai city.



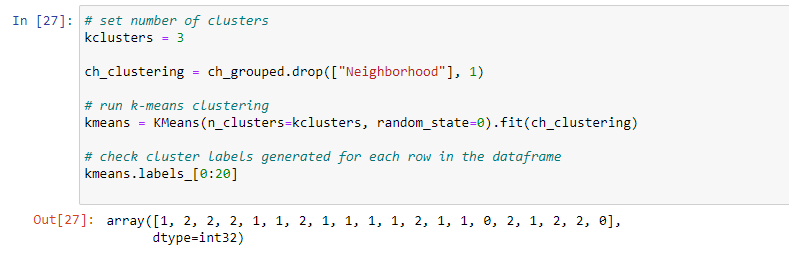
Below Four-Square API segmentation shows the clusters of location neighborhoods, segregated based on the “department store” venue type in South and East Chennai borough, which is the hot spot for many businesses, as the city center is already having lot of department store in other clusters.



## Results

After getting the neighbourhood data from foursquare API, now we have to cluster the data.

Using K-means clustering, we cluster the neighborhood with k value of 3.



Adding the cluster labels to get the map with different colors (for different clusters), we join the venues sorted data set and original data set.



When evaluating the results, cluster 0, which comprises Kovilambakkam, Nagalkeni, Pozhichalur and Selaiyur are having few department stores and would be the right place for stakeholders to open a new one.

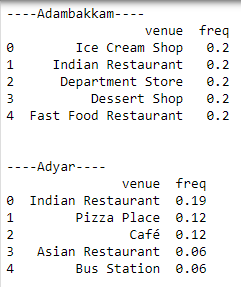


## Discussion

From foursquare API, we find the below unique venue categories around each neighbourhood and from this, department store would be more suitable for South India, as more people travel to far distances for groceries. We can cut down the travel time by creating a new department store, in location where population would also be there and not far from city.



Also, from foursquare API, we get nice data like below, showing the top 5 venue types in each of the neighbourhood, based on the frequency of its use.



And, from clustering exercise, we identified that clusters 1 and 2 are the locations where already department stores are found plenty and are in much densely populous areas. Moreover, those are the locations of office places which are not of our interest due to density of population and pollution in the areas.

## Conclusion

Since starting a department store in South and East Chennai is our goal (we have elaborated the need for selecting this area), by using the webscraping, foursquare API’s and k-means clustering we identified that the following localities would be ideal for our department store (Kovilambakkam, Nagaleni, Pozhichalur or Selaiyur). Being the place of regular households and also nearby city, these locations would be right choice for the stakeholders, as people in these areas would not have to spent time and energy travelling and spending money to get same quality products from other localities. Thus we would suggest or recommend to start department store in these locations.

## Acknowledgements

1. <http://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm>
2. <http://go.euromonitor.com/rs/805-KOK-719/images/2017%20Top%20100%20Cities%20Destinations%20Final%20Report.pdf>
3. <http://go.euromonitor.com/rs/805-KOK-719/images/wpTop100Cities19.pdf>
4. <https://www.thehindu.com/news/national/tamil-nadu/chennai-rated-the-safest-city-in-india/article8277935.ece>
5. <https://en.wikipedia.org/wiki/List_of_neighbourhoods_of_Chennai>
6. <https://developer.foursquare.com/>