Coursera Capstone

Neighbourhood analysis of Chennai

Introduction:

Chennai is one of the most Important Metropolitan City in India. Being the capital of Tamilnadu, a lot of people continuously Hail to Chennai. Chennai is one among the best City to welcome people to live in, but there are several Neighbourhoods to choose. For people coming from outside Chennai might not know more about its Neighbourhoods.

So, Our business problem is Clustering of Neighbourhoods in Chennai and Analysing the clusters to find the best Neighbourhood on the basis of distribution of several kinds of venues (For Example – some Neighbourhoods won't have all types of venues but some types recurring several times)

Our Stakeholders or people interested in our Project will be people who would want to shift to Chennai from other cities or People who will be interested to understand the Neighbourhoods of Chennai.

Data :

The data that will be used to Solve this problem is Geospatial Data of Chennai.

The data of several Neighbourhoods of Chennai are Scrapped from this Wikipedia page - https://en.wikipedia.org/wiki/Areas_of_Chennai .

With Names of each Neighbourhood, the Latitude and longitudes of each Neighbourhood are scrapped from geopy python library.

The Geospatial data containing the information of venues around each Neighbourhood is collected through Foursquare api.

The dataframe containing the Type of Neighbourhood is scrapped as a csv file from internet.

Methodology :

After Collecting and merging the data there isn't a much of Exploratory Data Analysis done as the Data is Geospatial data. After getting the top venues of each Neighbourhood, the Neighbourhoods are clustered using a machine learning model.

Statistics :

Top venues in Chennai :

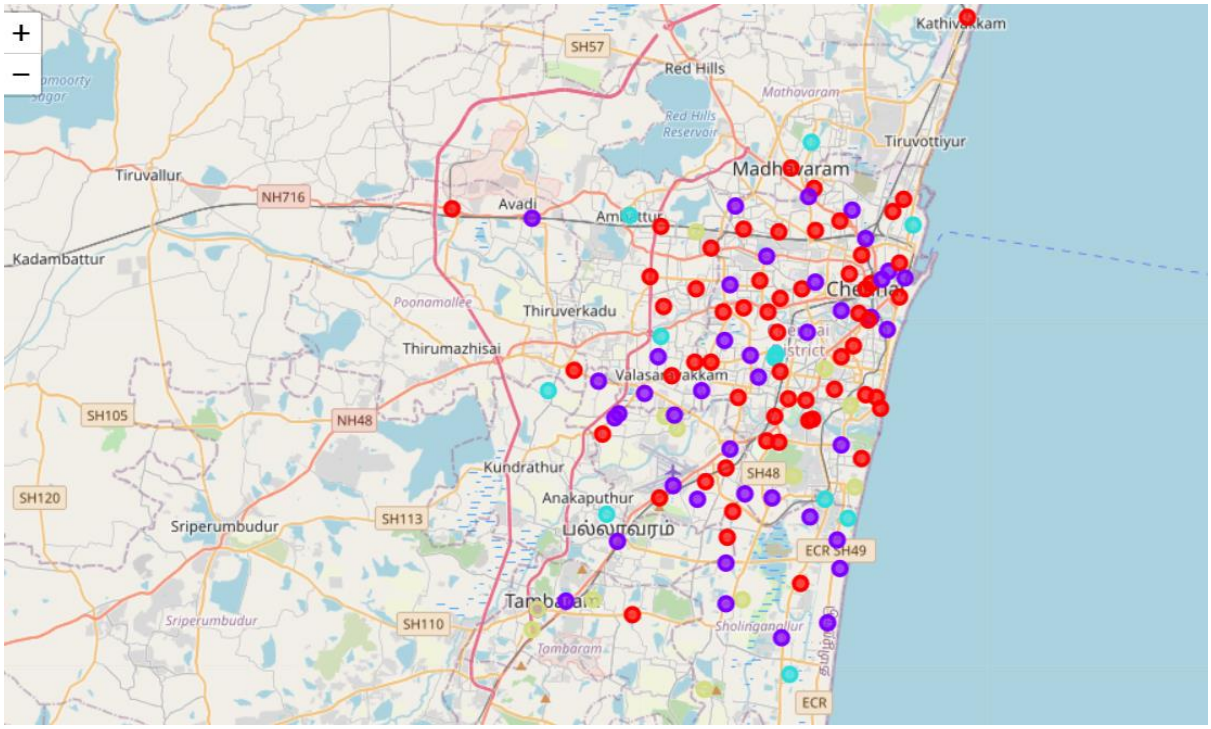
```
Indian Restaurant       119
Café                     36
Fast Food Restaurant     33
Pizza Place              29
Hotel                    27
ATM                      22
Ice Cream Shop           21
Coffee Shop              21
Bakery                   20
Juice Bar                19
Name: Venue Category, dtype: int64
```

Machine Learning Algorithm used :

The Machine Learning Algorithm used is **K – means Clustering.**

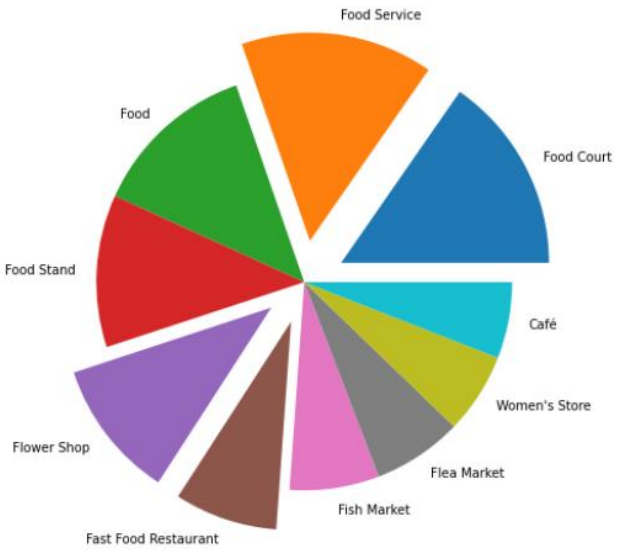
K – means Clustering is the best suited algorithm for Clustering this Neighbourhoods data (unlabeled data). The number of Clusters opted for this data is 4 Clusters.

After clustering Neighbourhoods :



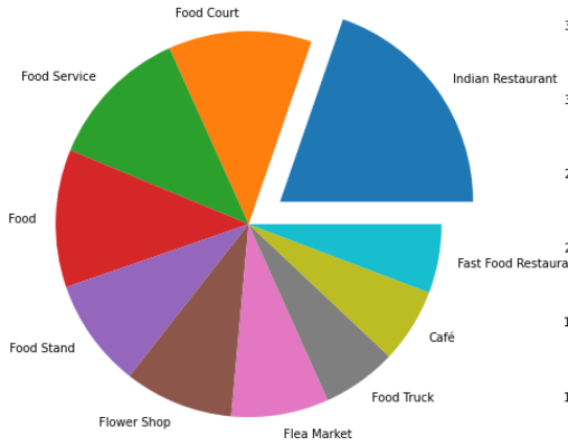
Analysing Each Cluster :

Cluster 1 venue Types –



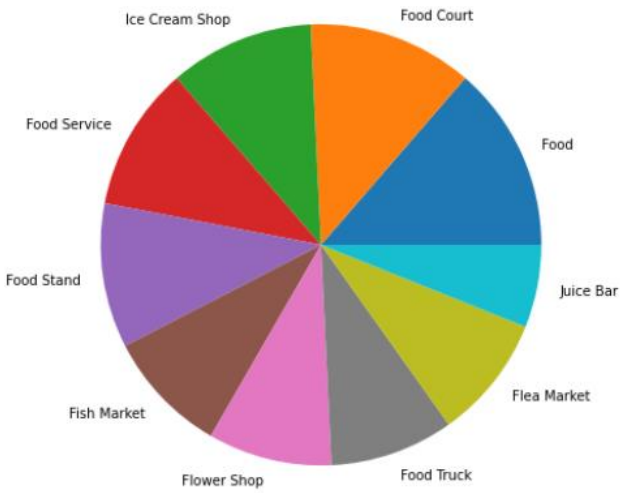
This Cluster majorly consists of Restaurants and food related-venues than any other venues.

Cluster 2 venue types –



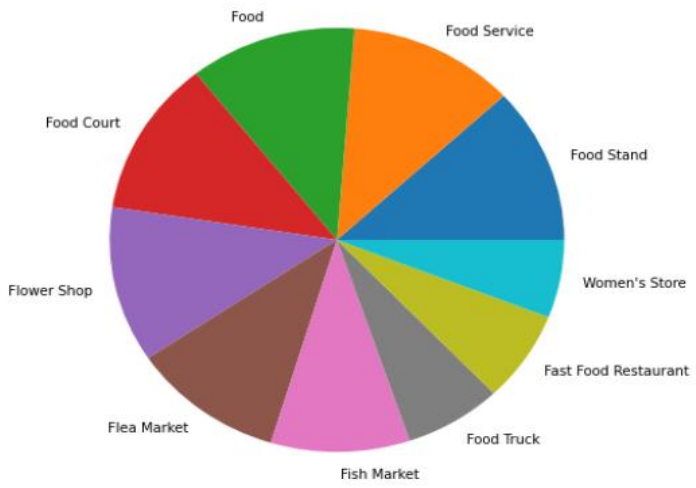
This Cluster also consists of Indian Restaurants and other restaurants as major.

Cluster 3 venue types –



This Cluster seems to give a equal Distribution between several of the venue Categories.

Cluster 4 venue types –



This cluster consists of very less restaurants but more of market and store type venues.

Discussion :

After Clustering the Neighbourhood, the most stable Neighbourhood is chosen as the best Neighbourhood. Everyone would have their own Preferences of Neighbourhoods. Some can choose restaurants filled clusters, some can choose stores, markets filled clusters and so on. So it is up to the customers to choose one among the Clusters as they preferred .

Cluster 3 is Selected as the best Cluster as it has equal Distribution of all Neighbourhoods. This Cluster consists of 7 Neighbourhoods. Among the seven “Maduravoyal” is only the Municipality Area. Municipality Areas have higher value than census Town or Panchayats. So “Maduravoyal” is selected.

Also, this project can help people to find/understand the Neighbourhoods in Chennai and implement this in their Business model.

Conclusion :

The Best Neighbourhood in Chennai for living is “Maduravoyal” followed by “pallikaranai” .