Capstone Project - The Battle of Neighbourhoods
Opening a New Vegetarian Restaurant in Chennai

Introduction:

Vegetarian and vegan dietary practices vary among countries. Differences include food standards, laws, and general cultural attitudes of vegetarian diets. In survey, it states that most of the Indians are Vegetarians. People follow their own diet. In the busy world, most of the people having their launch in restaurants. Hence it is necessary to follow a proper diet to stay healthy and compete the world. The food we take should be hygiene and easily available to the people. Thus, in this Data Analysis Project we going to identify the good place to start our Vegetarian / Vegan Restaurant in and around the neighbourhoods of Chennai.

Business Problem:

The main objective of this Capstone Project is to provide a good Vegetarian Restaurant to the public. The target of this Project is to find a location to start our business in and around location Chennai, India. Even though vegetarians are high in India, it is difficult find a good restaurant within their living area. Thus, through this project we are aiming the neighbourhood with a less or no Vegetarian/ Vegan Restaurant and provide a good environment to have their meals without searching and travelling long distance from their living.

Data:

For our analysis, we need the following data:
- List of Neighbourhoods of Chennai.
- The Latitude and Longitude of the Neighbourhoods.
- Venues Data in the Neighbourhoods.

Source of the Data:

    I.   List of Neighbourhoods of Chennai:

For the List of Neighbourhoods of Chennai, we are collecting the data from Wikipedia (https://en.wikipedia.org/wiki/List_of_neighbourhoods_of_Chennai). This List is not available as for our use directly. Hence, we are applying Beautiful Soup method to extract our requirements and forming a Data Frame using Pandas Library.

    II.   The Latitude and Longitude of the Neighbourhoods:

To get the Latitude and Longitude of the Neighbourhoods we are using "geocoders". We need to find the latitude and longitude for each of neighbourhoods in Neighbourhoods table. Then for our convenience we are merging the latitude and longitude with Neighbourhoods table. Finally, at this stage we have table with Neighbourhood name along with its Latitude and Longitude.

III.　Venues Data in the Neighbourhoods:

To get the Venues in and around the Neighbourhoods, we are using the Foursquare. Thus, we get the response as JSON format. Then we need to extract the only the data we are interested and form a Data Frame to store the data.

Methodology:

1. Data Collection and Cleaning:

The List of Neighbourhoods of Chennai is available in the website (https://en.wikipedia.org/wiki/List_of_neighbourhoods_of_Chennai). We need to process the website to fetch the required data alone. We are using "Beautiful Soup" method to parse the html of the webpage. The data available is now converted as List. Now using the "Pandas" Library the list has been converted to the Data Frame.

	Neighborhood
0	Red Hills
1	Royapuram
2	Korukkupet
3	Vyasarpadi
4	Tondiarpet

Fig. 1

In addition to the List of Neighbourhoods, we also require the Latitude and Longitude of each of the Neighbourhoods. Hence, we are using "geocoder" to fetch the Latitude and Longitude. Finally, we are appending the Latitude and Longitude with the Neighbourhood table. Hence, the table will look like Fig. 2.

	Neighborhood	Latitude	Longitude
0	Red Hills	13.19543	80.18431
1	Royapuram	13.11394	80.29420
2	Korukkupet	13.11680	80.27726
3	Vyasarpadi	13.11778	80.25168
4	Tondiarpet	13.12923	80.28955

Fig. 2

2. Data Analysing:

Foursquare API provides the location-based experience with information about the venues, users, etc. We need to explore the surroundings to get the venues. For this we are using

Foursquare to the Neighbourhood table. We get the response of venues in JSON format. It has been converted to table format for our convenience.

	Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Red Hills	13.19543	80.18431	Hotel Balaji Bavan	13.193716	80.185292	Indian Restaurant
1	Red Hills	13.19543	80.18431	Pvr Cinemas Skls Galaxy Mall	13.191586	80.186100	Movie Theater
2	Red Hills	13.19543	80.18431	Skls Galaxy Mall	13.191500	80.186038	Shopping Mall
3	Red Hills	13.19543	80.18431	Selvaraj Hotel	13.199747	80.178082	Indian Restaurant
4	Red Hills	13.19543	80.18431	Universal Gym	13.190710	80.177946	Gym

Fig. 3

Now from the table, we can analyse the unique venues present in the surroundings of Neighbourhoods. From this step, we are concentrating only on Vegetarian/ Vegan Restaurants available in the surroundings. Thus, extract that information alone and form a Data Frame.

3. K-Means Clustering:

Once the Data has been analysed, we are using K-Means Clustering algorithm to cluster the Neighbourhoods. In our case, we are using k=3. The Neighbourhoods are clustered as follows:
Cluster 0: availability of Vegetarian/ Vegan Restaurants is moderate.
Cluster 1: availability of Vegetarian/ Vegan Restaurants is high.
Cluster 2: availability of Vegetarian/ Vegan Restaurants is less or not available.

	Neighborhood	Vegetarian / Vegan Restaurant	Cluster Labels	Latitude	Longitude
0		0.035714	0	13.08362	80.28252
1	Adyar	0.012048	2	13.00305	80.25193
2	Alandur	0.000000	2	13.00013	80.20060
3	Alapakkam	0.090909	0	13.04610	80.16499
4	Alwarthirunagar	0.100000	0	13.05055	80.18397

Fig. 4

4. Visualization:

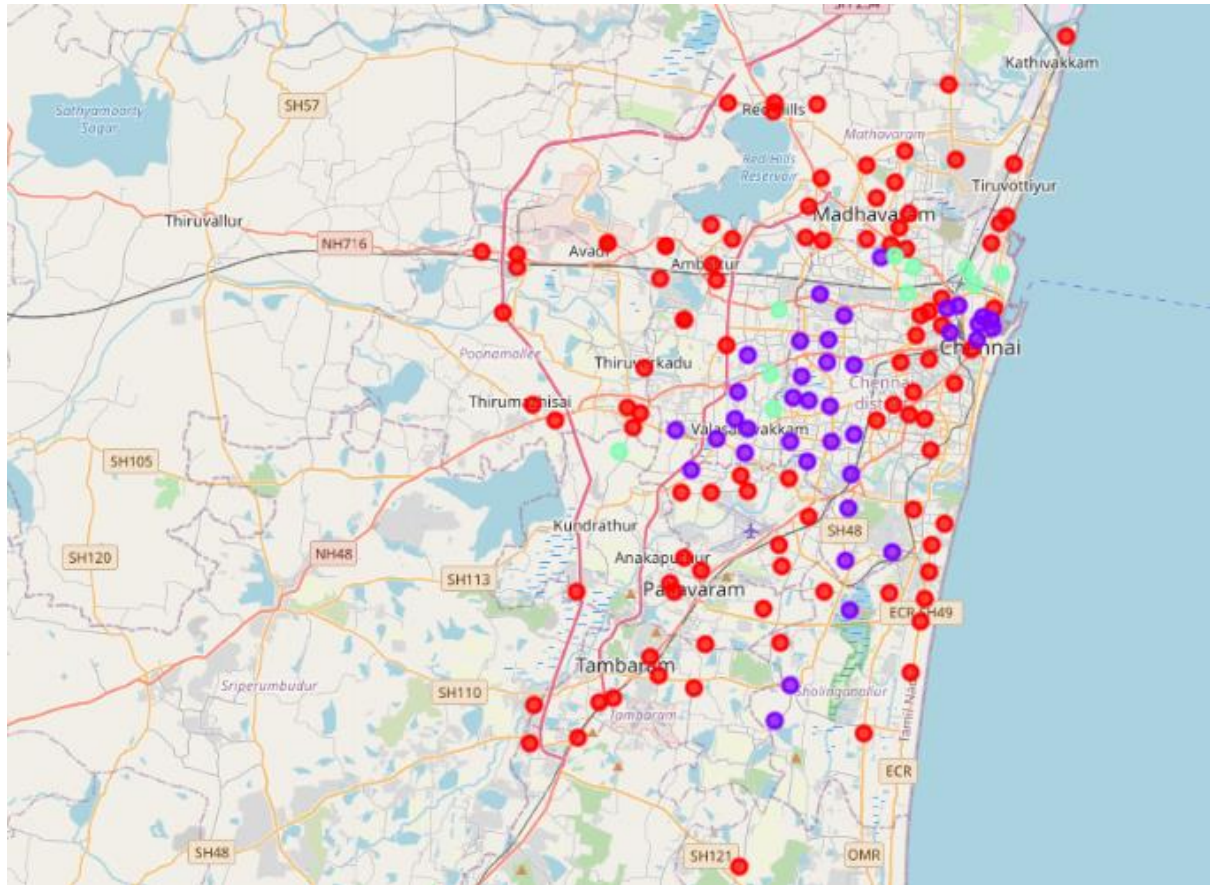The Map is been used to visualize the cluster of this project result for our better understanding.

Fig. 5.

Observation:

From this project, we can clearly understand that we can start our Vegetarian/ Vegan Restaurant in the cluster 2 area since it only has a less or no restaurant. It will a high competition when we open our Vegetarian/ Vegan Restaurant in cluster 1 Neighbourhoods.

Future Directions:

In this Project, we only considered one factor whether the Restaurant is already available or not in the Neighbourhoods. In future, we can specify more accurately how profit we can earn when we select a neighbourhood by considering more factors like number of vegetarians living in that area.