**Coursera Capstone Project**

**Utilizing Foursquare API to analyze Chennai’s Neighborhoods**

Introduction:

Chennai is the capital of Tamilnadu and it is one of the metropolitan cities of India. Chennai is the 6th Largest city in India with the population of 9,714,000 as of 2015 with Land area of 971 km2.Diversity of cuisine available is reflective of social and economic diversity of Chennai.

The aim of project is to find neighborhoods with a high frequency of restaurants/food stalls/cafés in Chennai, India. Chennai is one of the Metropolitan City in India. To find popular restaurants or food stalls or cafes, respective coordinates need to be retrieved, so that Foursquare can be used to find nearby venues.

Problem Description:

Chennai has many Ancient temples like Triplicane parthasarathy temple, Mylapore Sivan temple and Hertiage sites like Mahabalipuram, St.Geroge Fort etc. Due to these, many foreign tourists across globe and people from various parts of India are visiting Chennai daily. The main problem they are facing is good lodging and food.

In such situation, they mainly relied on Internet and searching various restaurants nearby their lodging and there are few questions to be addressed:

1. How many types of foods available in the restaurant?
2. Which is the most nearest me having good rating?
3. How many “similar” restaurants are available nearby?
4. Do the “similar” restaurants cost more? If so what is the specialty they will have?

Expectation from proposed recommender system is to address these questions and also to recommend new places based on their rankings.

Target Audience:

Target Audience is not particular to a touristor who are new to Chennai city but also to everyone who is living in Chennai city also since no one aware of popular restaurants in all the places. People who rarely visits restaurants would prefer to have the most rated restaurants nearby them and all this could be handled by our recommender system.

Data

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 metrics. Folium visualization library can be used to visualize the clusters superimposed on the map of Chennai city. These clusters can be analyzed to help small scale business owners select a suitable location for their need such as Hotels, Shopping Malls, Restaurants or even specifically Indian restaurants or Coffee shops.

Chennai has multiple neighborhoods. The chennaiiq.com website has a dataset which has the list of locations in Chennai along with their Latitude and Longitude. In order to obtain the venue details in each neighborhood Foursquare API is used. https://chennaiiq.com/chennai/latitude\_longitude\_areas.asp https://foursquare.com/ There is a total of 105 neighborhoods. But the Latitude and Longitude data obtained are in Degrees Minute Seconds format which needs to be converted to Decimal Degrees Format. The details obtained from Foursquare API are Venue, Venue Latitude, Venue Longitude and Venue Category.

Methodology

Now, we have the neighborhoods data of Chennai (105 neighborhoods). We also have the most popular venues in each neighborhood obtained using Foursquare API. A total of 1130 venues have been obtained in the whole city and 145 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. Here K - Nearest Neighbor clustering technique have been used. 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. This data can be used to suggest business people, suitable locations based on the category.

Result:

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 a good Data Science 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.So one can easily find out the food outlets using this project but still we can see that foursqure is not upto date and in future this can resolved using better data source with latest update.