Capstone Project - The Battle of Neighborhoods (Week 1)

Clustering localities in Bangalore, India based on Restaurants (K-means Clustering)

Introduction:

People visiting new cities would be highly interested in the localities with the best restaurants in the city. People might want to know how good a given restaurant is based on the ratings the restaurant has received and would like to know the price range the Restaurant falls under so that they can make informed budget decisions. Also, they would like to know the best localities where they could find these restaurants. The information of ratings and price range of various restaurants in the city and their localities in form of graphs, charts and maps would help people decide which restaurant to choose among the many restaurants in the city. And also which locality to visit. Also combining the location of the restaurants in the city with their price and rating information would help visitors make easy decisions about the locations they should visit. A map of the restaurants and another map of the localities with specific color attributes will be plotted to highlight their position. Further, we will classify the various locations into different clusters using a Machine Learning Algorithm, the K-means clustering Algorithm. This enables any visitor to take a quick glance and decide what place to visit.

Data Collection:

To get location and other information about various venues in Bangalore, two APIs were used. The Foursquare API and the Zomato API. The Foursquare's explore API was used to fetch venues up to a range of 35 kilometers from the center of Bangalore. The names, categories and locations (latitude and longitude) of these venues were collected. Using the name, latitude and longitude values obtained from the Foursquare API, we used the Zomato search API to fetch data from its database. The Zomato API allows to find only restaurants based on a search criteria using the name, latitude, longitude, etc. The data from the two APIs do not match completely because Foursquare API retrieves all venues in Bangalore and the Zomato API retrieves only restaurants in Bangalore. So, we combine the two datasets to get only Restaurants from the Foursquare API and the corresponding ratings and price information from the Zomato API.

We use various techniques of Data cleaning to get the final dataset.

From Foursquare API, the following for each venue was retrieved:
Name: The name of the venue.
Category: The category type as defined by the API.
Latitude: The latitude value of the venue.
Longitude: The longitude value of the venue.

From Zomato API, the following for each restaurant was retrieved:

Name: The name of the restaurant.
Locality: The locality of the restaurant.
Rating: The average rating of the restaurant given be users.
Price range: The price ranges the restaurant belongs to as defined by Zomato.
Price for two: The average cost for two people dining at the restaurant.
Latitude: The latitude value of the restaurant.
Longitude: The longitude value of the restaurant.
Votes: Number of person provided ratings.