Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Analysis and clustering of Indian Restaurants in the neighborhood of Boston, MA, USA

Nithin K Mohan
August 17, 2020

1. Introduction: Business Problem

There is a large population of Asian-Indians around Boston. The population is mostly comprised of those employed in various sectors in Massachusetts state (Government, Financial, Healthcare, Hospitals, colleges to name a few), as well as students in prestigious colleges in Boston area. Most of them live in the city of Boston or the suburban towns which are well connected to Boston using Subway/commuter rail/bus system.
(We are using the term Asian-Indians as Indians in USA commonly refers to the American-Indians who are the native tribes of America)

Indian restaurants are of huge demand especially in areas which has larger concentration of Indian population. In this project, we will try to find optimal locations to open a new **Indian** restaurant in the neighborhood of **Boston** in the state of **Massachusetts, USA**.

We will try to detect the locations of already existing Indian restaurants in the locality around Boston. The 'search for venues' end point of **Foursquare API** will be used to get the location details. The search option will be used with Boston along with two more locations (Framingham and Braintree) to get more coverage as the search option can return a maximum of 50 rows only in the result.

This data will be used to find out the locations which has a lesser concentration of Indian restaurants. We will use our data science powers to cluster the Indian restaurants around the area and locate places which has a lesser concentration.

Note: The project can be further expanded by clustering the Asian-Indian population in Massachusetts by city and plotting the cluster of population against cluster of restaurants. This will lead to a much accurate prediction. This cannot be done at this point of time due to lack of availability of population data.

2. Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing Indian restaurants in the neighborhood of Boston.
- location (latitude and longitude) of Indian restaurants.
- concentration of asian Indians in Boston and towns around Boston. (dataset unavailable, hence not used)

The list of Indian restaurants around Boston area along with their location will help us to analyze the locations, cluster them and find out the areas where there is a lesser concentration of restaurants. By knowing the concentration of Asian-Indian population in the towns around Boston, we could predict the best places where the concentration of Asian-Indians are high still the concentration of Indian restaurants are less.

Following data sources will be needed to extract/generate the required information.

- Geo coordinates of the cities Boston, Framingham and Braintree will be obtained using **Geopy Nominatin.**
- Number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

The Get Venue Search endpoint of Foursqauare API will be used to get the list of restaurants.

```
GET https://api.foursquare.com/v2/venues/search
```

Following parameters will be passed to the Foursquare API in addition to the Client ID and Client secret.

query = ‘Indian’
categoryId = ‘4d4b7105d754a06374d81259' # Food (includes restaurants)
ll = Latitude and Longitude of the location
radius = 1000
limit = 100

Note: Since Foursquare explore venue API returns only 50 results per search, we are using 2 more search queries with locations in the periphery of Boston, named Framingham and Braintree to get more coverage.