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 an optimum location 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 acquisition and cleaning

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.

2.1 Data sources

Following data sources were 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 were 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.

2.2 Data Selection

The Get Venue Search endpoint of Foursqauare API was used to retrieve the list of restaurants along with the location and other details. The Get/venue/search call was performed using three locations named Boston (which is in the biggest city in Massachusetts), Framingham and Braintree. By doing a Foursquare search call with Boston location coordinates, we found that the results from Framingham and Braintree were the last in the list (means farthest results from

Boston). Framingham located in the west of Boston and Braintree located in South of Boston have a larger population of Asian Indians as these places are easily accessible from Boston using public transit systems. Hence we wanted to expand our search by doing a separate venue/search calls using Framingham as well as Braintree as location. This helped to expand our input dataset.

The initial intention is to also use the dateset which will have the population of Asian Indian in various towns of Massachusetts. This will help to identify the clusters of locations where the Asian-Indian population in Massachusetts. But such an analysis could not be performed as the 'population by race' was unavailable. Since I am living in Boston for years, I decided top use by experience to address the Asian-Indian population clustering.

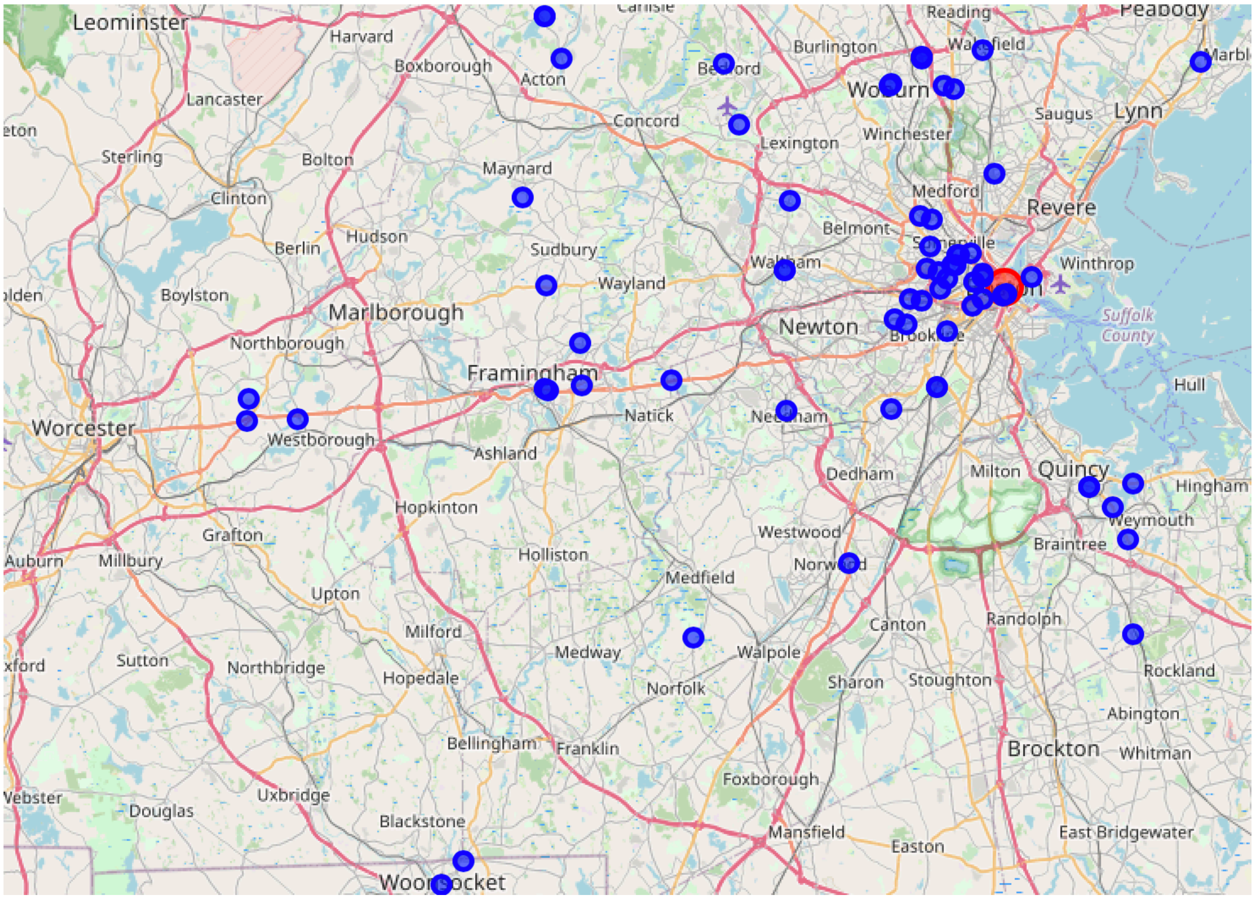
2.3 Data Cleaning

The results of Foursquare API calls using Boston, Framingham and Boston as location are merged into one dataframe for analysis. (The json_normalize option from pandas is used to read the json format API result into dataframe). Only the required details like restaurant name, location (latitude and longitude), address are kept in the data frame and the remaining fields are dropped.

3. Methodology

3.1 Exploratory data analysis

The merged dataframe is analyzed and is plotted on a map using folium function. This helped to get a fair idea of the location if the Indian restaurants around Boston. *(See map below)*

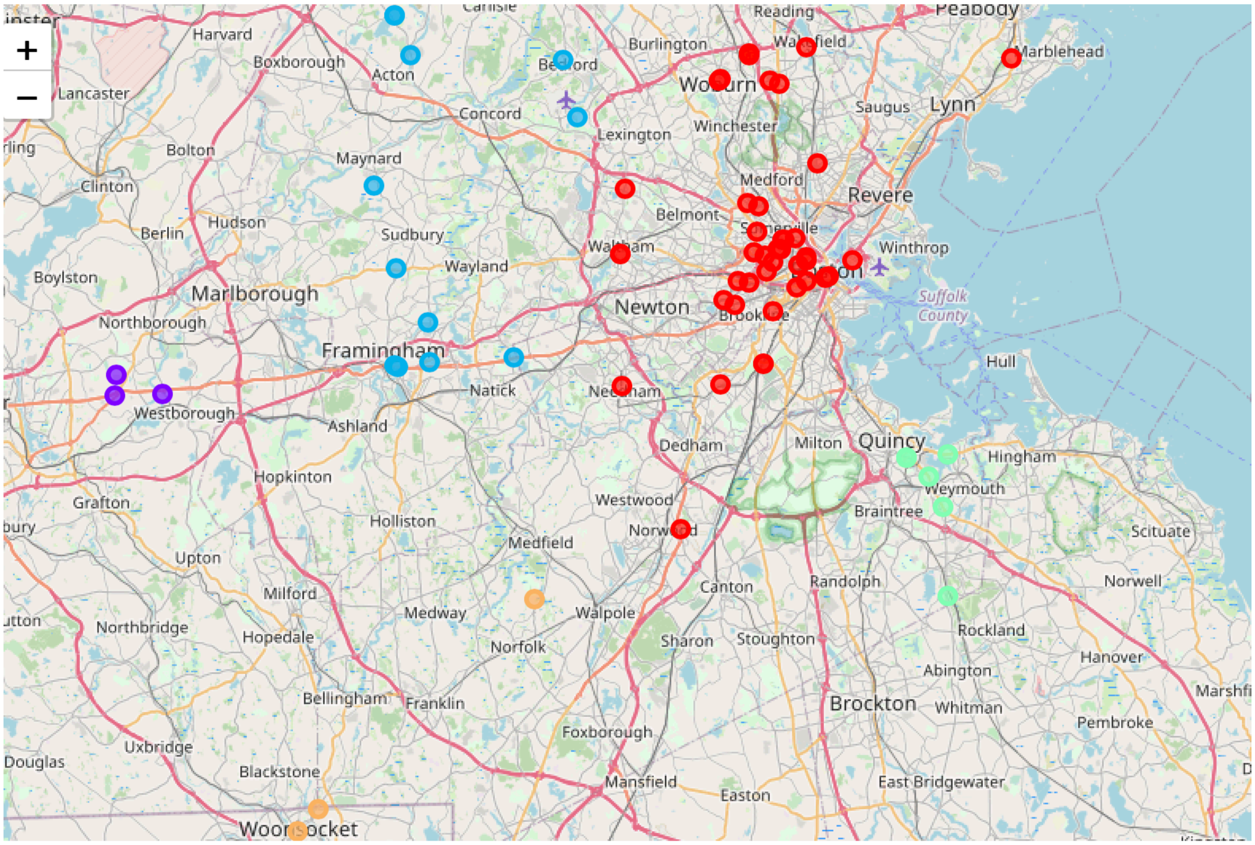


3.2 Machine Learning Algorithm

Now that we have the location details for all Indian restaurants around Boston, we tried try to cluster them using **K-Means Clustering** algorithm. This gave an idea of concentration of Indian restaurants around Boston area. The clustering will help to identify locations with lesser density of Indian restaurants. After trying different values of Ks, an optimum value fo 5 was decided. The location (latitude and longitude) of the restaurant locations were used as the input parameters for the K-Means clustering algorithm.

The K-Means Clusters were plotted on a map using folium for easy visualization.

K-Means Clusters on map



4. Results

Our analysis using the clusters created by K-Means clustering algorithm shows that there is a large concentration of Indian restaurants in Boston, Framingham and Westborough areas. There is a fair concentration in Quincy-Braintree area too. These are definitely areas where there is a larger concentration of Asian-Indian population as well.

The **Norwood-Sharon-Walpole belt** which has a large population of Asian-Indians (mainly due to the good school in these towns) seem to have only very few Indian restaurants.

5. Discussion

Here are a few observations for discussion:

1) The **Norwood-Sharon-Walpole belt** in the south of Boston has a large population of Asian-Indians, but have only very few Indian restaurants.

2) **Framingham and Westborough** area also has a good concentration of Indian population along with famous temples. But both these locations are closely by; so if we take these 2 clusters together, the total count of restaurants is good.

3) The largest concentration of restaurants in **Boston metropolitan area** gives a picture that it is a hot place for an Indian restaurant. But we need to remember that most of the consumers in Boston city area are either the working class or the students. Since most employees and students are following remote working/learning options due to the COVID19 pandemic, there could be a drastic drop in the restaurant business in this area. Hence Boston metropolitan area is definitely not a good choice at least at this point of time.

6. Conclusion

From our analysis, it is evident that the **Norwood-Sharon-Walpole belt** could be an ideal choice to start a new Indian restaurants as this cluster has larger Indian population, but fewer Indian restaurants. Norwood has an advantage being a key shopping place for the people in these areas, but there are already a few restaurants in this area. Hence **Sharon or Walpole** could be a best place to open a brand new Indian restaurant. These two towns are adjacent and easily commutable.