Capstone Project Report - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Chenmin Ni

5/31/2020

Introduction: Business Problem

As a major American city, **Boston** is home to a wide variety of business establishments. From Central Boston's financial hub to the large hospitals of the Longwood Medical Area to the major research universities in Fenway, **Boston's unique neighborhoods** help the city as a whole serve a global market. Each of Boston's neighborhoods has something unique to offer businesses, and this factors into where industries choose to settle.

In this report, we will try to find an optimal location for a **Korean Restaurant** in Boston. Since there are lots of Korean restaurants in Boston, we will try to find **the neighborhood that are not already crowded with restaurants and with no top Korean restaurants in vicinity**. Here, we will show to the stakeholders about a few most promising neighborhoods based on this criterion for them to make a final decision.

Data

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

- neighborhoods 's information, such as **zip code**, **coordinates** in Boston area
- **number of restaurants** in the neighborhoods
- **number of Korean restaurants** in the neighborhoods, if any

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

- number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**
- coordinate of Berlin center will be obtained using *US Zip Code Latitude and Longitude* from the website: [https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/]

Methodology

- Use geopy library to get the latitude and longitude values of Boston

- create map of Boston using latitude and longitude values

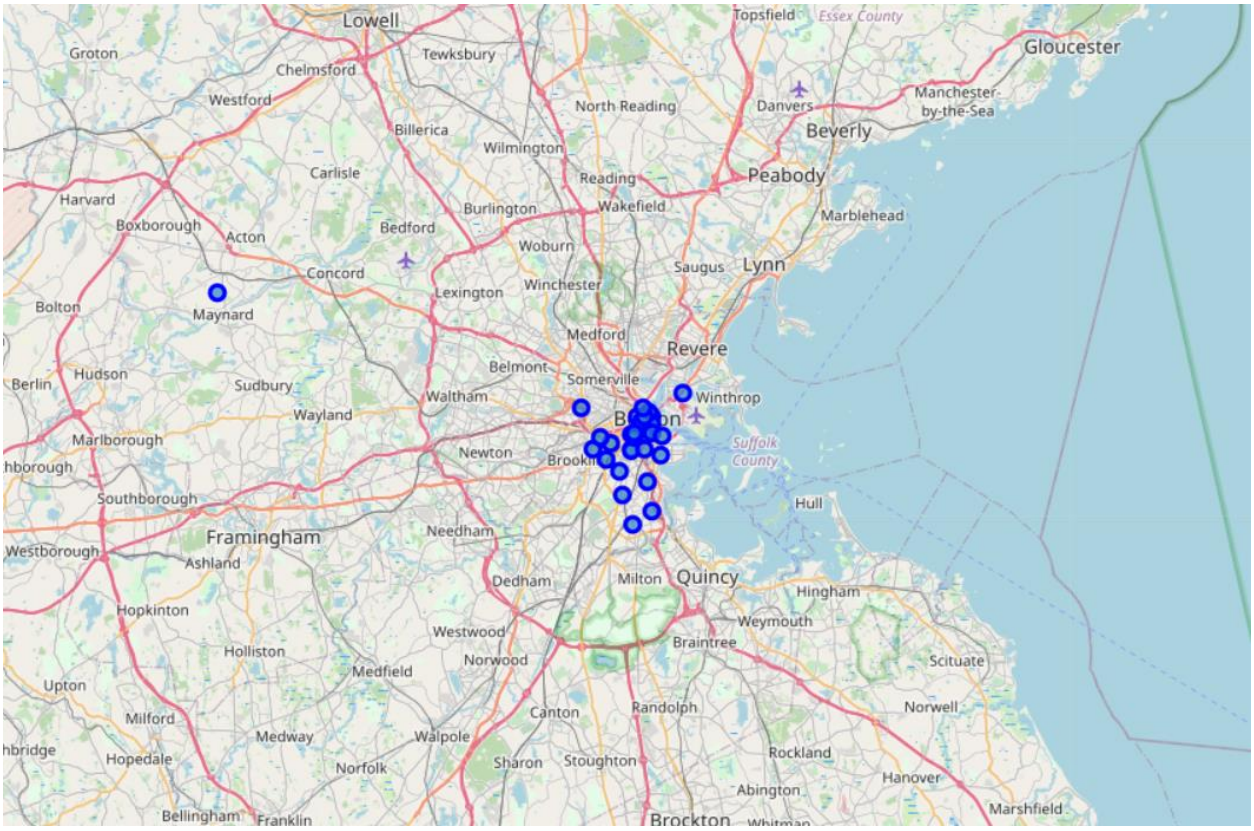

Figure.1 map of Boston using latitude and longitude values

- Import Foursquare Credentials and Version
- Analyze Each Neighborhood
- Get only restaurant venues in all of the Neighborhoods
- put that into a pandas dataframe
- create the new dataframe and display the top 10 venues for each neighborhood.
- Use K-Mean to cluster the Neighborhoods according to their common venue. K-Means algorithm is one of the most common cluster methods of unsupervised learning.
- Create a new dataframe that includes the clusters as well as the top 10 venues for each neighborhood
- Finally, visualize the resulting clusters

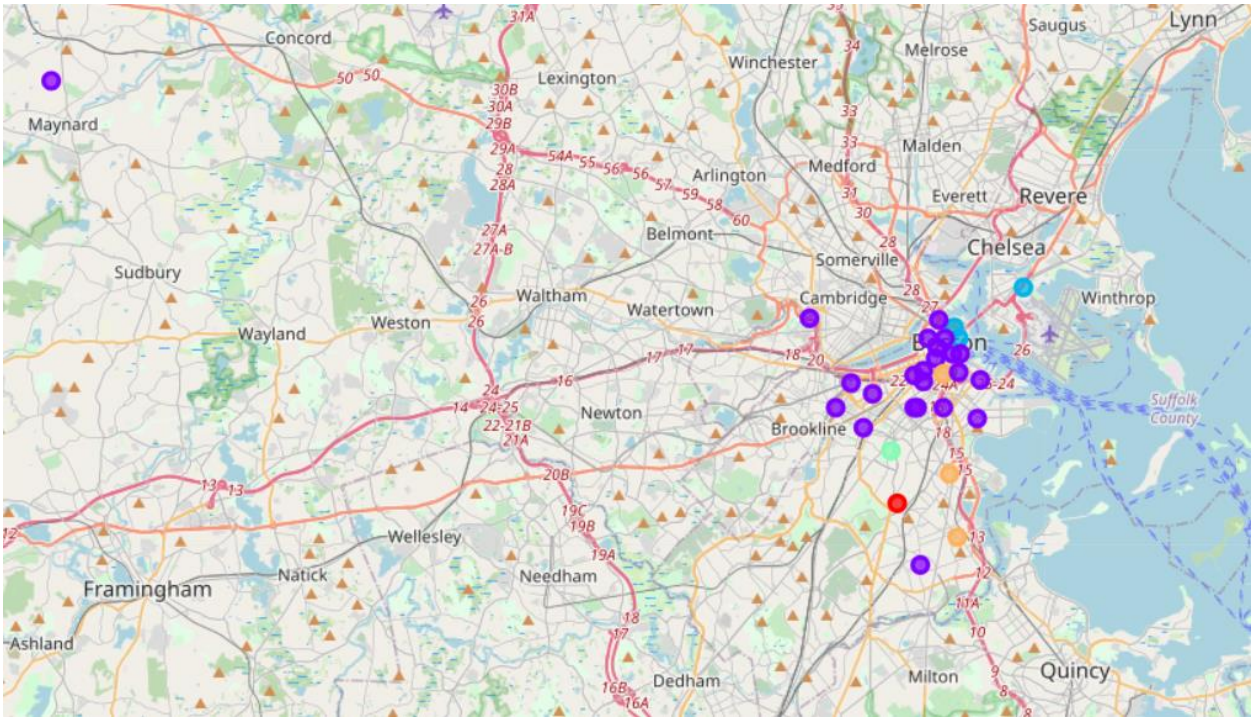

Figure.2 Clusters of Boston restaurant

Results and Discussion

As the clusters are shown above, we now examine each cluster and determine where is the best point for opening a Korean restaurant.

Cluster 1 (red point) is in Roxbury where the top 5 most common restaurant venues are fast food, soul food, Caribbean food, Vietnamese food and Dim Sum. However, this area is a little far away from downtown. Thus, we will not take this cluster into consideration.

Cluster 2 (purple points) has many restaurants, which is the main area for dining and craving. Also, the style of cuisine is divergent and various. So, we will not choose these places as they are crowed with different kinds of restaurants.

Cluster 3 (blue points) is only three different zip code, with a smaller area. Also, this area is close to the downtown Boston. However, as we take a close look at the most common restaurant, Italian, seafood and American restaurants are the top three dining in this area, so we will also not choose to open a Korean restaurant here.

Cluster 4(light green) is a little away from the downtown area, so we will not take it into consideration.

Cluster 5 (orange points) is our choice! Since first, it is near the downtown area but not overcrowded with other restaurants. Second, the most common restaurants in this cluster are both Asian cuisines, meaning that the subculture here is fit for the theme of a Korean restaurants.

Conclusion

From Central Boston's financial hub to the large hospitals of the Longwood Medical Area to the major research universities in Fenway, Boston's unique neighborhoods help the city serve a global market. Each neighborhood does its part to make the city of Boston what it is, while ach maintains its own respective character. Boston's neighborhoods are not static, however, and have changed with the prevailing trends of the city.

This report aims to show the environment of restaurant development in Boston area and help the skateholders to decide where to open a korean restaurant. First, we import information of zip code, neighborhood names, and coordinates (latitudes and longitudes) from the website: [https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/]. And then, we clear the dataframe with removing the NaN values. Second, we import geopy to get a rough map for Boston area regarding to the coordinates. Third, we connect to Foursquare API to get the venue information. Then we select only restaurant information for further analysis.

Here we use K-means to cluster the similar neighborhood base on the restaurant ranking, by doing this, we can clearly know the dining culture in each of the cluster so that we are able to identify where is the best and fit place for our skateholder to open a korean restaurant.