Seafood Restaurants in Boston Neighborhoods

Sujit Kandar, 6-24-2021

Capstone Project for IBM Data Science Certification Curriculum in Coursera

Contents:

1. Introduction/Business Problem

2. Description of the Data

3. Methodology

4. Results

5. Conclusions

6. Acknowledgement

1. **Introduction/Business Problem**

1. Exploratory data analysis on Boston neighborhood venues.

2. Build a simple recommendation system for visitors to choose Boston neighborhoods with the highest concentration of seafood restaurants.

3. Create a map with color marker for each neighborhood.

1. Description of the Data

1. scrap the Boston neighborhoods data from Wikipedia (<u>Neighborhoods in Boston - Wikipedia</u>) using the Beautiful Soup library.

2. restaurant information in Boston neighborhoods from <u>Foursquare</u> using an API.

3.   Merge these two data sets and perform a cluster analysis by the k nearest neighbors regression method.

4.   Create a geographical map using folium library.

3. Methodology

1. The get request and beautiful soup object was created to fetch the neighborhood data from wikipedia.

2. The neighborhood data was stored in the "boston_neighborhoods" list.

3. A pandas data frame was created from the "boston_neighborhoods" list.

4. A function was created and called upon the neighborhoods to get their latitude and longitude coordinates.

5. The coordinates were converted into a pandas data frame from the list.

6. Coordinates of Boston were retrieved using geocoder library and a map of Boston along with the neighborhoods will be created using folium library.
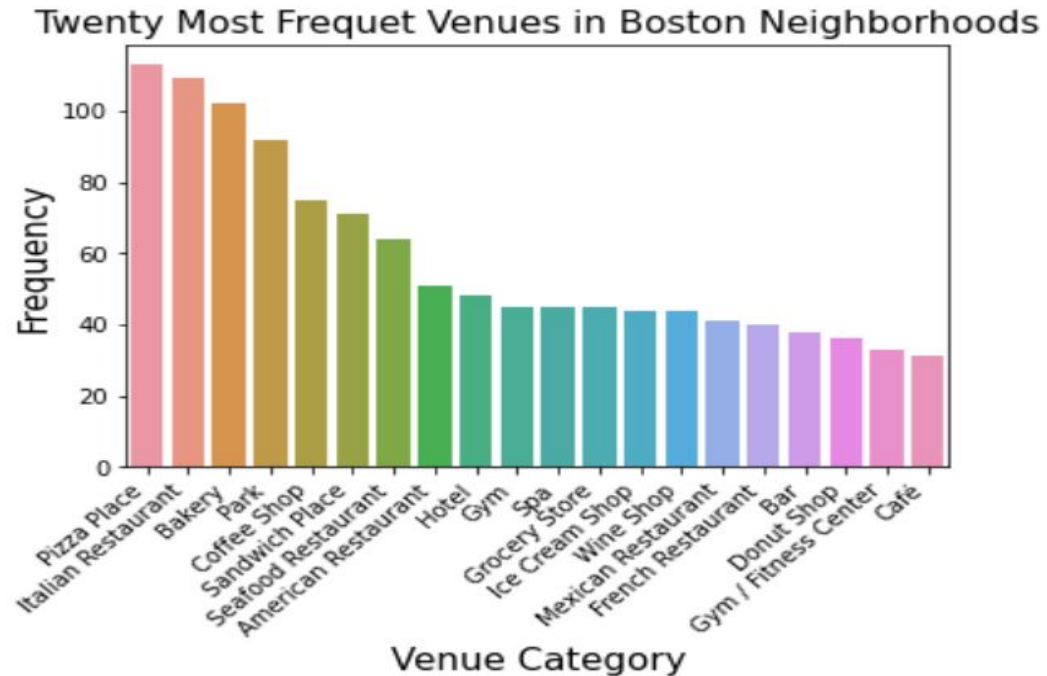
3. Methodology (continued)
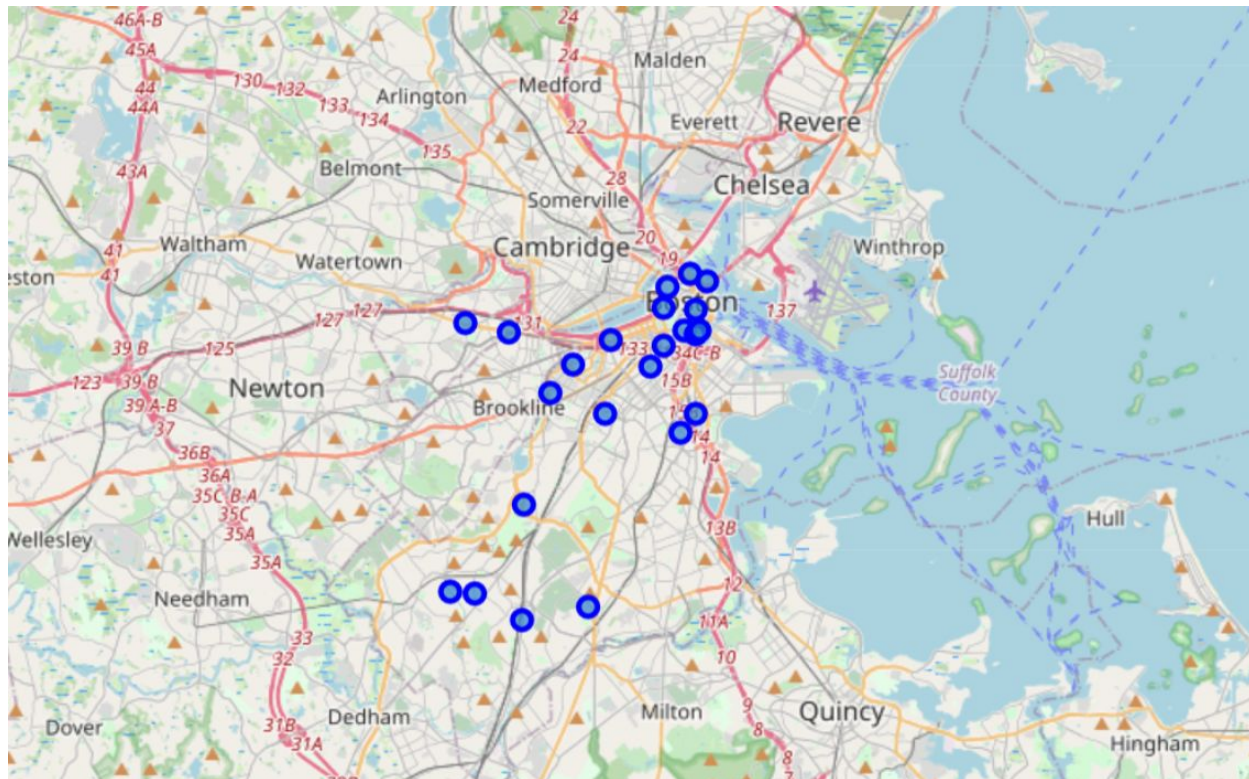
3. Methodology (continued)

7. Boston neighborhoods venues were explored using foursquare API and the results were stored into a dataframe.

8. A separated data frame was created only for the seafood restaurants along with their geographical coordinates.
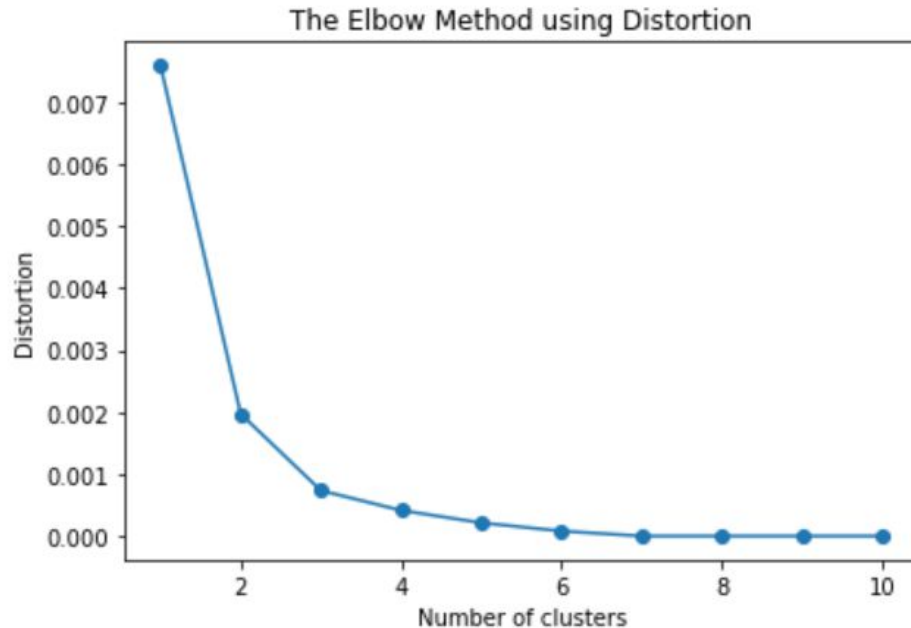
3. Methodology (continued)

9. A list of top 20 venue categories in Boston Neighborhoods was created and plotted as a histogram.

Twenty Most Frequet Venues in Boston Neighborhoods

3. Methodology (K-Nearest Neighbors)

10. Elbow method to determine the optimum number of clusters.


The Elbow Method using Distortion

3. Methodology (continued)

Optimum for k nearest neighbor modelling on this data set = 3

boston_seafood_grouped was merged with boston_data to add latitude/longitude for each neighborhood followed by sorting the data frame.

Finally, a map of three clusters in the Boston neighborhood using the folium library.

3. Methodology (Boston neighborhood map with three clusters)
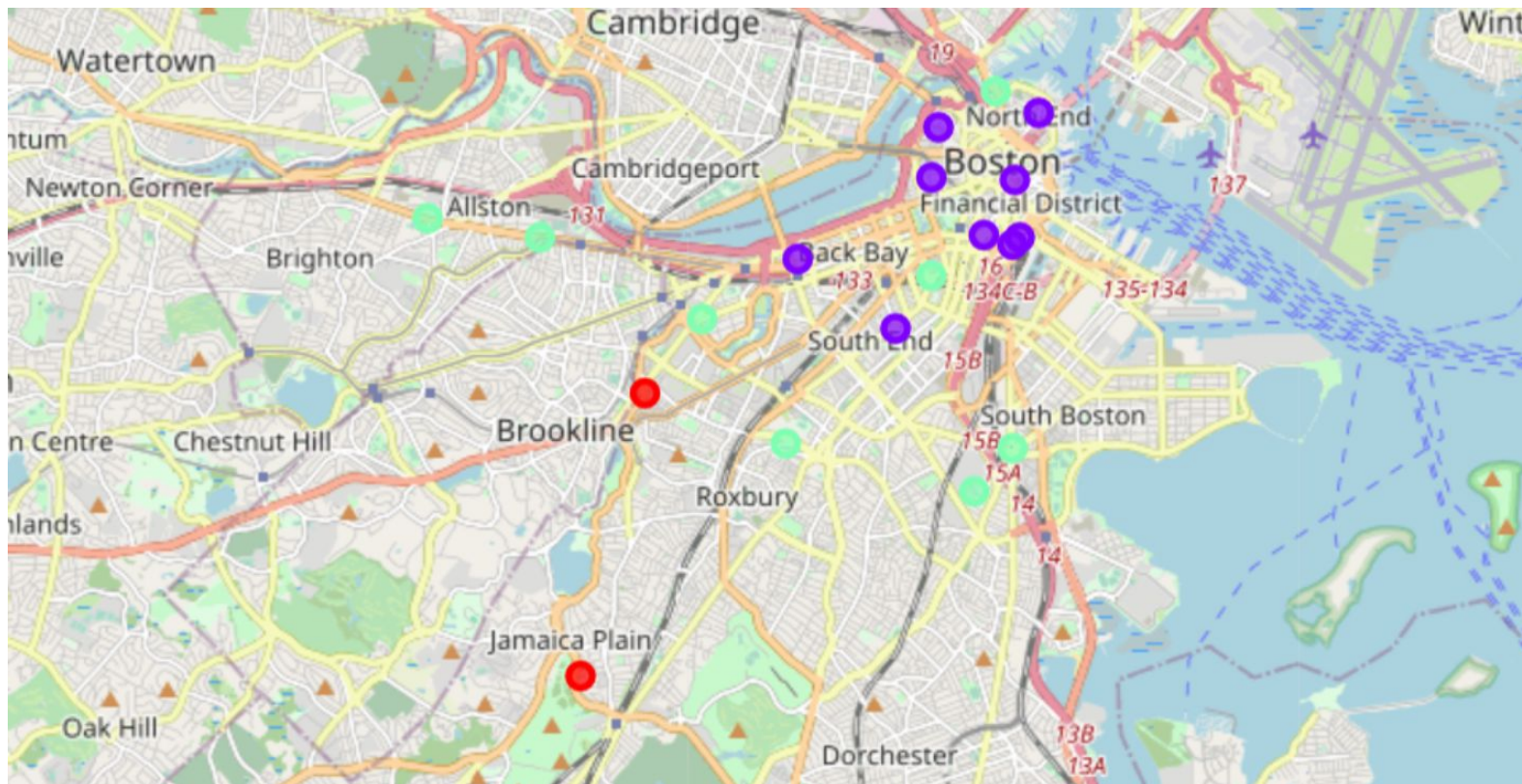
4. Results

1. Cluster 2 (purple circles) has the highest concentration of seafood restaurants in Boston.

2. Cluster 3 (green circles) has the second highest concentration.

3. Cluster 1 (red circles) has the lowest concentration of seafood restaurants in Boston.

5. Conclusions

1. Downtown Boston and the nearby areas have the highest concentration of seafood restaurants.

2. Boston and the surrounding neighborhoods have many indoor and outdoor activities and plenty of food choices.

6. Acknowledgement

1. Many thanks to IBM Data Science Teaching Staff in Coursera.

2. Thanks to geeksforgeeks.