# Applied Data Science for Opening a New Chinses Restaurant in Boston

Coursera Capstone Course

Charlie Sun

September 4, 2020

# Introduction

## Business Problem

The objective of this project is to find a proper location for a Chinese restaurant in Boston. Opening a restaurant requires a lot of upfront investment and continuous cost, which requires the restaurant to have a stable source of customers in the course of operation. The main source of guests should be the surrounding communities. Therefore, analyzing location data makes a lot of sense. Location data from Foursquare also reveals the number of existing restaurants and similar food and beverage services in different locations, minimizing the problem of excessive competition in choosing locations for new restaurants.

It is worth noting that this project is focusing on a Chinese restaurant, it's also important to consider different people’s preference for food. People from the same region may have similar tastes. In other words, potential customers of Chinese restaurants are also likely to be regular customers of Thai, Japanese and other Asian restaurants. So, when analyzing location data, it makes sense to consider these Asian restaurant data.

In a word, the question for this project is: in Boston, United States, if someone wants to open a new Chinese Restaurant, which location would you recommend?

## Target Audience

The target audience for this project is the one who is intended to open a new Chinese restaurant in Boston. Location will significantly affect the operation of a restaurant. Whether there are too few restaurants in a location to create a business cluster,

or too many restaurants to increase competition, it can affect the operation of a new restaurant. Also, the taste of the residents in the area where the restaurant is located is another factor. Therefore, a clear perception of all investor locations in the Boston neighborhoods is necessary for the intended audience of this project.

## Data

To solve this problem, the following data is required:

- Neighborhoods in Boston (https://en.wikipedia.org/wiki/Neighborhoods_in_Boston).
- Coordinates of the neighborhoods.
- The race and ethnicity distribution in different neighborhoods. It will help to minimize the range of the data we need to explore (https://statisticalatlas.com/place/Massachusetts/Boston/Race-and-Ethnicity#figure/neighborhood).
- Venue data, especially those are related to restaurants.

## Methodology

### Tools

- Pandas: to create and manipulate data frames.
- Geocoder: to get the coordinates of the location.
- BeautifulSoup and Requests: to do web scraping.
- JSON: to deal with JSON files.

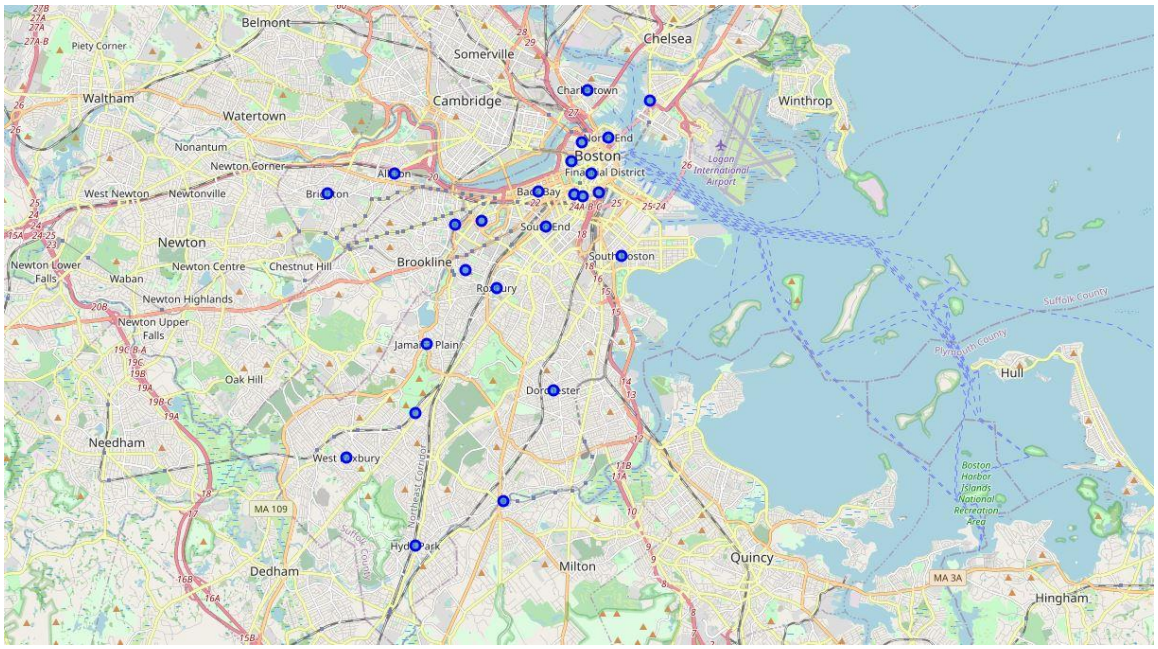
- Folium: to visualize the neighborhoods in the map.
- Scikit Learn: to import k-means clustering.

**Steps**

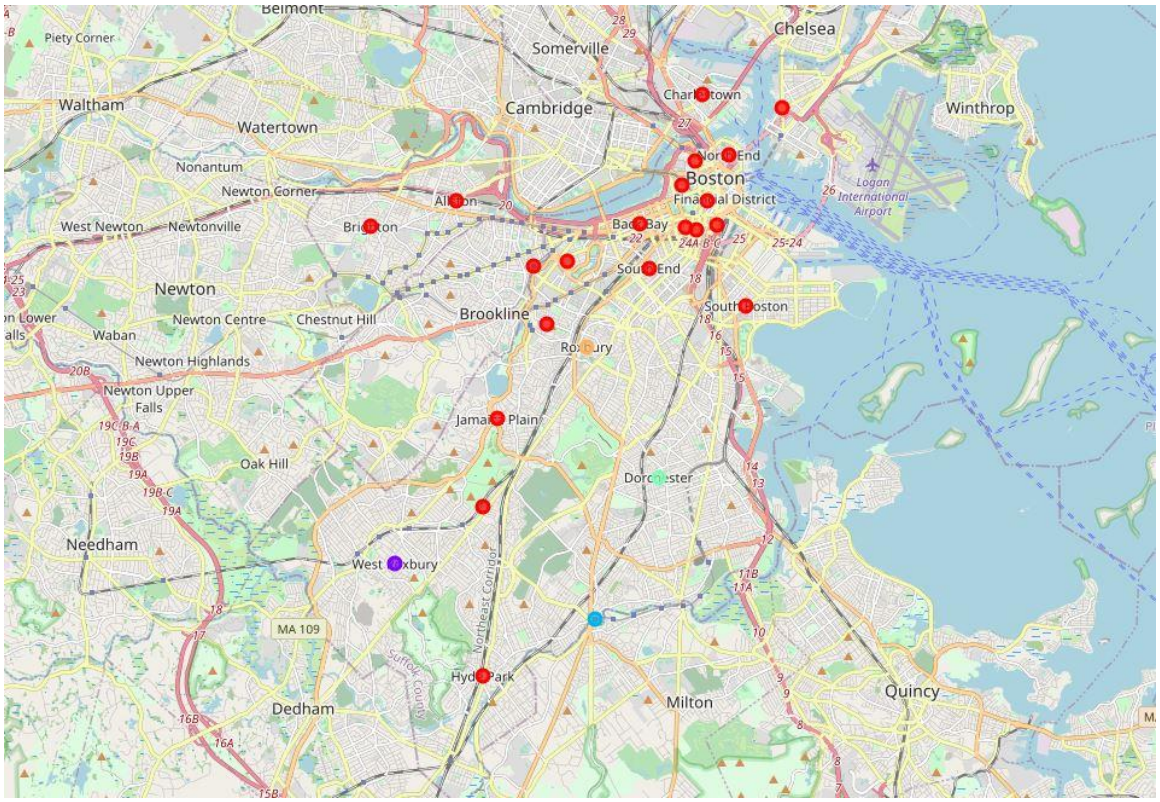
- Use Python requests and BeautifulSoup package to acquire the neighborhoods information of Boston on the Wikipedia page.
- Use Python Geocoder package to get the coordinates of the different neighborhoods in Boston.
- Search for the race and ethnicity data of different neighborhoods. The analysis will focus on the 10 neighborhoods with the largest Asian populations.
- Visualize the different neighborhoods in the map with Folium library.
- Use FourSquare API to get venue data for the focused neighborhoods, especially the location data of the Asian restaurants like Thai, Korean, and Japanese restaurants.
- Do the segmenting and clustering to find the differences and similarities among the clusters.
- Compare and conclude the final observations.

## Results

First, we can see all the neighborhoods showing on the map of Boston.




With the help of K-Means clustering, the neighborhoods were divided into 5 clusters, which is displayed in the next picture.



As we can see, the most neighborhoods are shown in red, which means they are in the same cluster. In other words, from the perspective of the location data, most neighborhoods in Boston have some similarities.

The following tables show the most frequented places visited by each community in different clusters.

Cluster 1

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Allston | Korean Restaurant | Pizza Place | Thai Restaurant | Bakery | Vegetarian / Vegan Restaurant | Chinese Restaurant | Thrift / Vintage Store | Fried Chicken Joint | Pharmacy | Bubble Tea Shop |
| 1 | Back Bay | Coffee Shop | Italian Restaurant | American Restaurant | Hotel | Seafood Restaurant | Clothing Store | Ice Cream Shop | Gym / Fitness Center | Juice Bar | Shopping Mall |
| 2 | Bay Village | Hotel | Sandwich Place | Italian Restaurant | Performing Arts Venue | Seafood Restaurant | Bakery | American Restaurant | Theater | Spa | Park |
| 3 | Beacon Hill | Pizza Place | Italian Restaurant | American Restaurant | Museum | French Restaurant | Restaurant | Hotel Bar | Yoga Studio | Hotpot Restaurant | Park |
| 4 | Brighton | Pizza Place | Bank | Bus Station | Bakery | Coffee Shop | Chinese Restaurant | Grocery Store | Pub | Deli / Bodega | Mobile Phone Shop |
| 5 | Charlestown | Pizza Place | Pub | Gastropub | Donut Shop | Coffee Shop | Convenience Store | Yoga Studio | National Park | Plaza | Pharmacy |
| 6 | Chinatown | Chinese Restaurant | Bakery | Asian Restaurant | Seafood Restaurant | Bubble Tea Shop | Performing Arts Venue | Hotel | Restaurant | Coffee Shop | Sandwich Place |
| 7 | Leather District | Chinese Restaurant | Asian Restaurant | Bakery | Coffee Shop | Sandwich Place | Sushi Restaurant | American Restaurant | Food Truck | Dessert Shop | Noodle House |
| 9 | Downtown Boston | Coffee Shop | Chinese Restaurant | Park | Bakery | Historic Site | Sushi Restaurant | Burger Joint | Asian Restaurant | Hotel | New American Restaurant |
| 10 | East Boston | Latin American Restaurant | Convenience Store | Liquor Store | Art Gallery | Pharmacy | Pizza Place | Sandwich Place | Colombian Restaurant | Café | Fast Food Restaurant |
| 11 | Fenway-Kenmore | American Restaurant | Park | Sports Bar | Thai Restaurant | Hotel | Café | Mexican Restaurant | Liquor Store | Greek Restaurant | Coffee Shop |
| 12 | Longwood | Donut Shop | Coffee Shop | Pharmacy | Park | Metro Station | Fast Food Restaurant | Falafel Restaurant | Hotel | Café | Diner |
| 13 | Hyde Park | American Restaurant | Pizza Place | ATM | Italian Restaurant | Donut Shop | Sandwich Place | Bar | Bank | Grocery Store | Theater |
| 14 | Jamaica Plain | Bakery | Art Gallery | Coffee Shop | Donut Shop | Library | Deli / Bodega | Noodle House | Pet Store | Pizza Place | Child Care Service |
| 16 | Mission Hill | American Restaurant | Hotel | Bank | Ice Cream Shop | Indian Restaurant | Greek Restaurant | Liquor Store | Metro Station | Park | Pizza Place |
| 17 | North End | Italian Restaurant | Seafood Restaurant | Park | Pizza Place | Pub | Coffee Shop | Bakery | Café | Historic Site | Sports Bar |
| 18 | Roslindale | Yoga Studio | Cuban Restaurant | American Restaurant | Pizza Place | Pool | Big Box Store | Rental Car Location | Donut Shop | Scenic Lookout | Doctor's Office |
| 20 | South Boston | Liquor Store | Pizza Place | Bar | Italian Restaurant | Convenience Store | Chinese Restaurant | Coffee Shop | Sports Bar | Pet Store | Mexican Restaurant |
| 21 | South End | Italian Restaurant | Coffee Shop | Bakery | Wine Shop | Wine Bar | Gift Shop | Park | Bar | French Restaurant | Yoga Studio |
| 22 | West End | Pizza Place | Hotel | Donut Shop | Bar | American Restaurant | Sandwich Place | Italian Restaurant | Coffee Shop | Sports Bar | Mexican Restaurant |

Cluster 2

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | West Roxbury | Home Service | Lawyer | Yoga Studio | Doctor's Office | Fast Food Restaurant | Farmers Market | Falafel Restaurant | Event Space | Electronics Store | Dumpling Restaurant |

Cluster 3

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | Mattapan | Yoga Studio | Bakery | Fast Food Restaurant | Mobile Phone Shop | Nail Salon | Pharmacy | Caribbean Restaurant | Southern / Soul Food Restaurant | Convenience Store | Automotive Shop |

Cluster 4

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | Dorchester | Food | Market | Breakfast Spot | Fried Chicken Joint | Plaza | Gym | Café | Pizza Place | Vegetarian / Vegan Restaurant | Southern / Soul Food Restaurant |

Cluster 5

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | Roxbury | Park | Metro Station | Plaza | Art Gallery | Gym | Furniture / Home Store | Doctor's Office | Farmers Market | Falafel Restaurant | Event Space |

## Discussion

Most of the neighborhoods are in cluster 1 where cafe, restaurants, stores are abundant. While the places with high frequency in cluster 2,3,4, and 5 are different. In cluster 2, people visited those places mainly for getting services. In cluster 3, the places that people visit frequently cover a wide range of types, and it can be seen that it is a neighborhood with relatively convenient life. In cluster 4, although lots of places mentioned here are about food, they are mainly related to fast food. In cluster 5, the location people frequently visit are park, metro, plaza, art gallery, gym, and so on. It can be seen that this is a good neighborhood for entertaining. In a word, if I need to find a cluster for opening a new Chinese restaurant, it definitely should be in cluster 1. However, I still need to compare different neighborhoods in cluster 1 to find a more specific location.

## Conclusion

As mentioned above, for the final conclusion, more analysis is still required. In Cluster 1, we can see that in neighborhoods Allston, Chinatown, Leather District, Brighton, and Downtown Boston, Chinese restaurants are frequently visited places. Also, in these neighborhoods, other kinds of Asian restaurants are also frequently visited places. In other words, people who are living in these neighborhoods are of high probability to love Chinese food. Further, in Chinatown, Leather District and Downtown Boston, the frequency of Chinese restaurants is first or second, that is to say, if a new restaurant is to be opened here, the competitive pressure may be very great. Let's compare Allston and Brighton. In both these two neighborhoods, Chinese restaurants are the 6$^{th}$ common venues. However, it obvious that there are more restaurants in Allston than in

Brighton, which means Allston has a more pronounced propensity for business aggregation. In my understanding, business aggregation will help to attractive customers.

Therefore, the final conclusion is that if someone is going to open a new restaurant, I would recommend Allston.