

The Battle of Neighborhoods: Opening Asian Restaurant in Boston

The Capstone Project

By : Yange Bian

Introduction: The Business Problem and Target Audience

- The city of Boston remains home to the largest number of Asian Americans in the state (60,985) and neighborhoods in Greater Boston in 2016 had at least 4,000 Asian Americans.
- Target Audience:
 - business personnel or local Asian immigrant family interested in opening a restaurant
 - freelancer willing to invest restaurant business
 - Asian American scholars or students looking for dining option
- Why do restaurants matter?
 - The types, quantities, locations and popularity are the true representation of the local neighborhoods' economic activity and ethnic diversity

Key Research Questions

- Where should I open an Asian restaurants to have the best chances of getting profits? Or what is the most popular neighborhood location for Asian cuisine or eatery?
- What Asian restaurants have the highest rating and where are they located? Do they also locate in those popular neighborhoods?
- To what extent does the distribution of Asian restaurants represent the local economic activities? Explore the other top venue/businesses categories in those popular neighborhoods where these Asian restaurants are located.

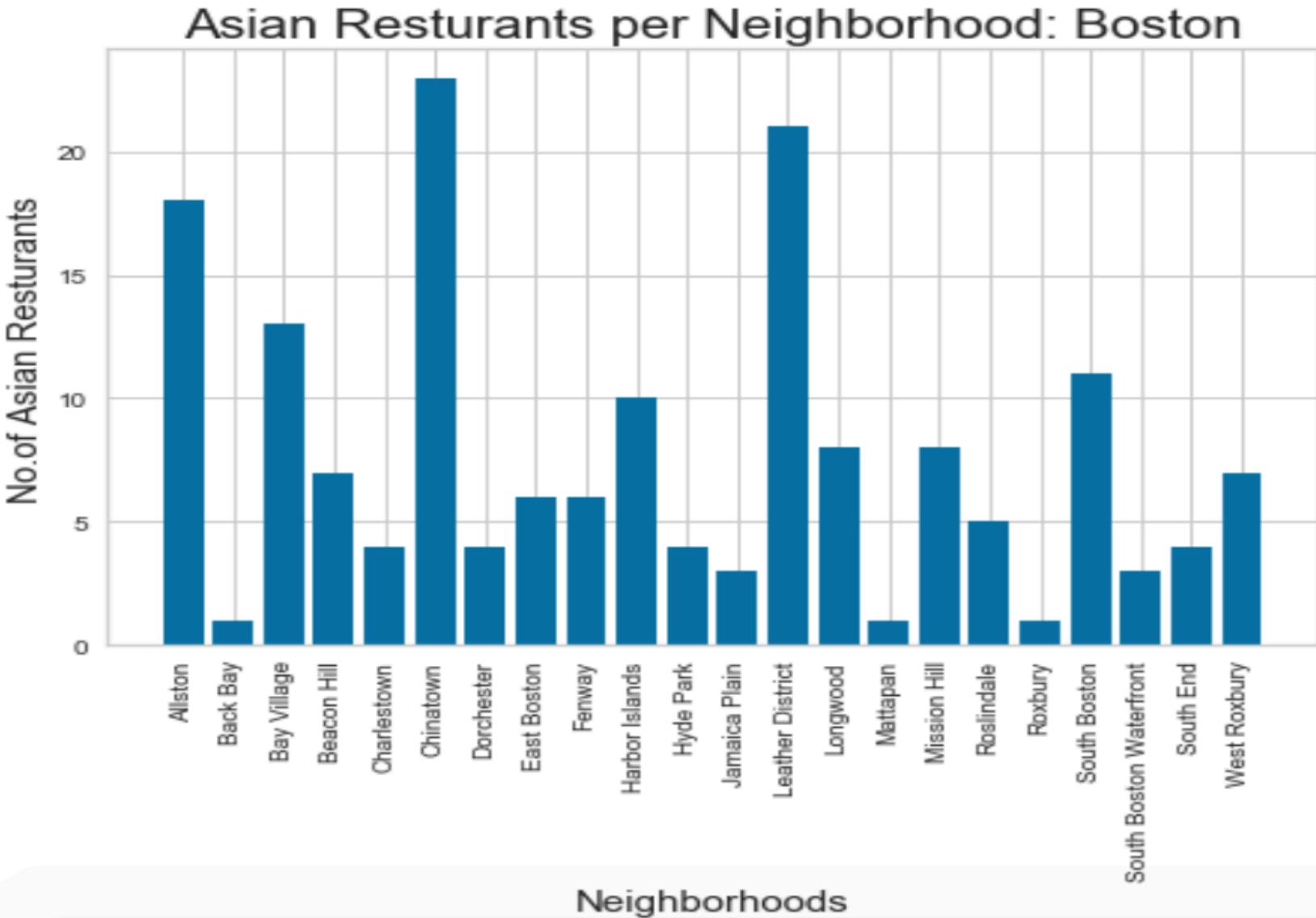
Data Sources

- Boston neighborhoods, boundaries, latitudes and longitudes will be obtained from the data source (**geojson file**): <https://data.boston.gov/dataset/boston-neighborhoods/resource/13ee2b65-6547-4168-b112-83995f138602>
- **Foursquare API** will be used to locate all venues and filtered by Asian restaurants(**i.e. Chinese, Japanese, Korean, Indian, Thai, Vietnamese, Asian Fusion, etc**) via using search, explore or trending queries.
 - **Locations/neighborhoods, ratings, tips and likes** for each restaurant will be obtained and counted, and merged and added to a data frame.

Data Acquisition and Cleaning

- The cleaned data frame extracted from geojson file has 25 neighborhoods including the features(latitude and longitude)
- 22 out of 25 neighborhoods have available data for Asian restaurants
- There are **168 Asian restaurants** returned from Foursquare API
- The top 5 neighborhoods with **the most number of Asian restaurants** are: **Chinatown, Leather Strict, Allston, Bay Village and South Boston.** **They are the highly populated regions for Asian restaurants.**
- The top neighborhoods with **high average ratings** are (>8.0): **South End, South Boston, Bay Village, Charlestown, and Allston.**
- After sorting the data frame, **Bay Village and Fenway** have individual restaurants with **highest rating, tips and likes.**
- **Using the average rating 7.5 as a separator, there are 16 neighborhoods and 226 unique venue categories in each neighborhood**

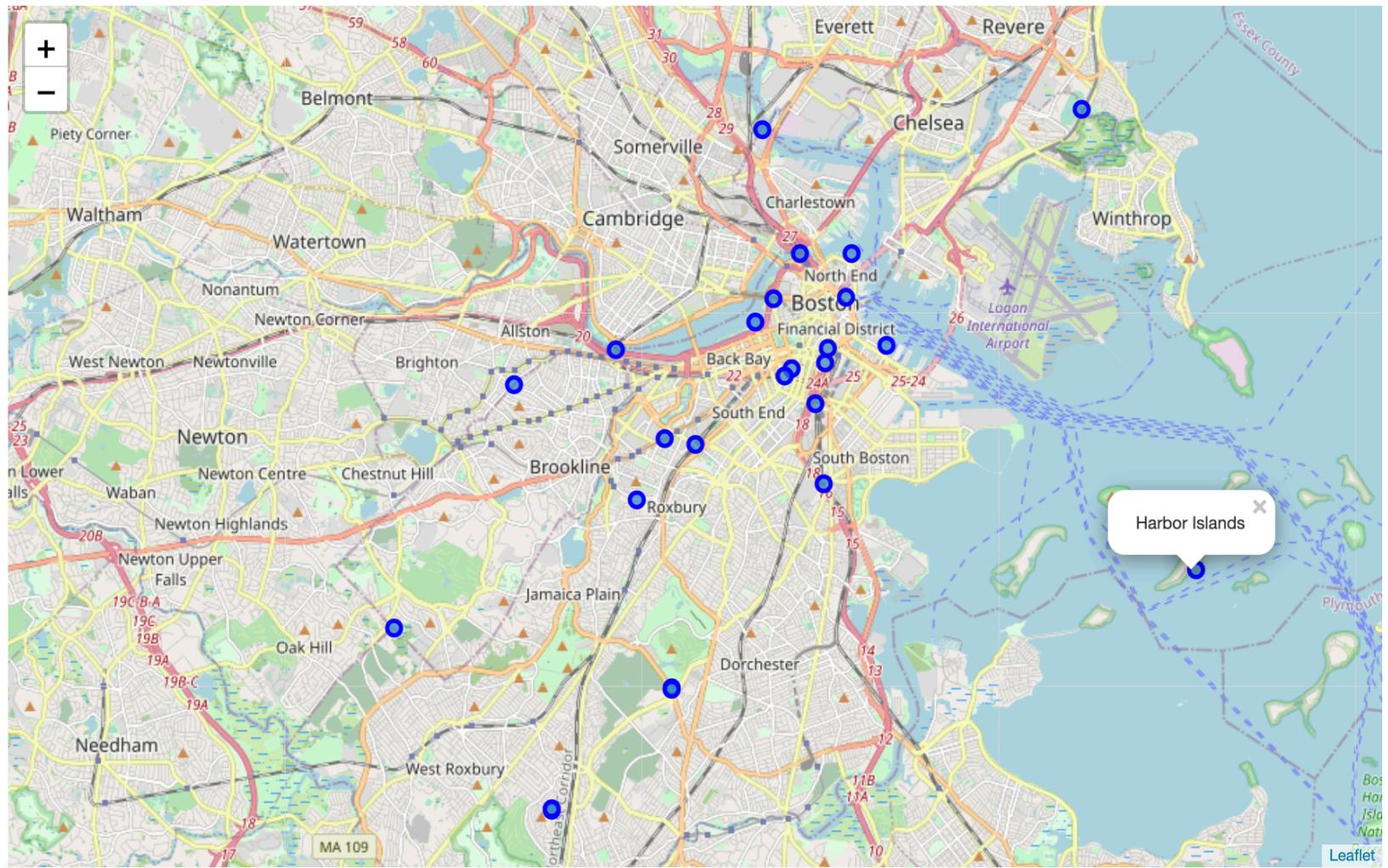
Exploratory Data Analysis Part I



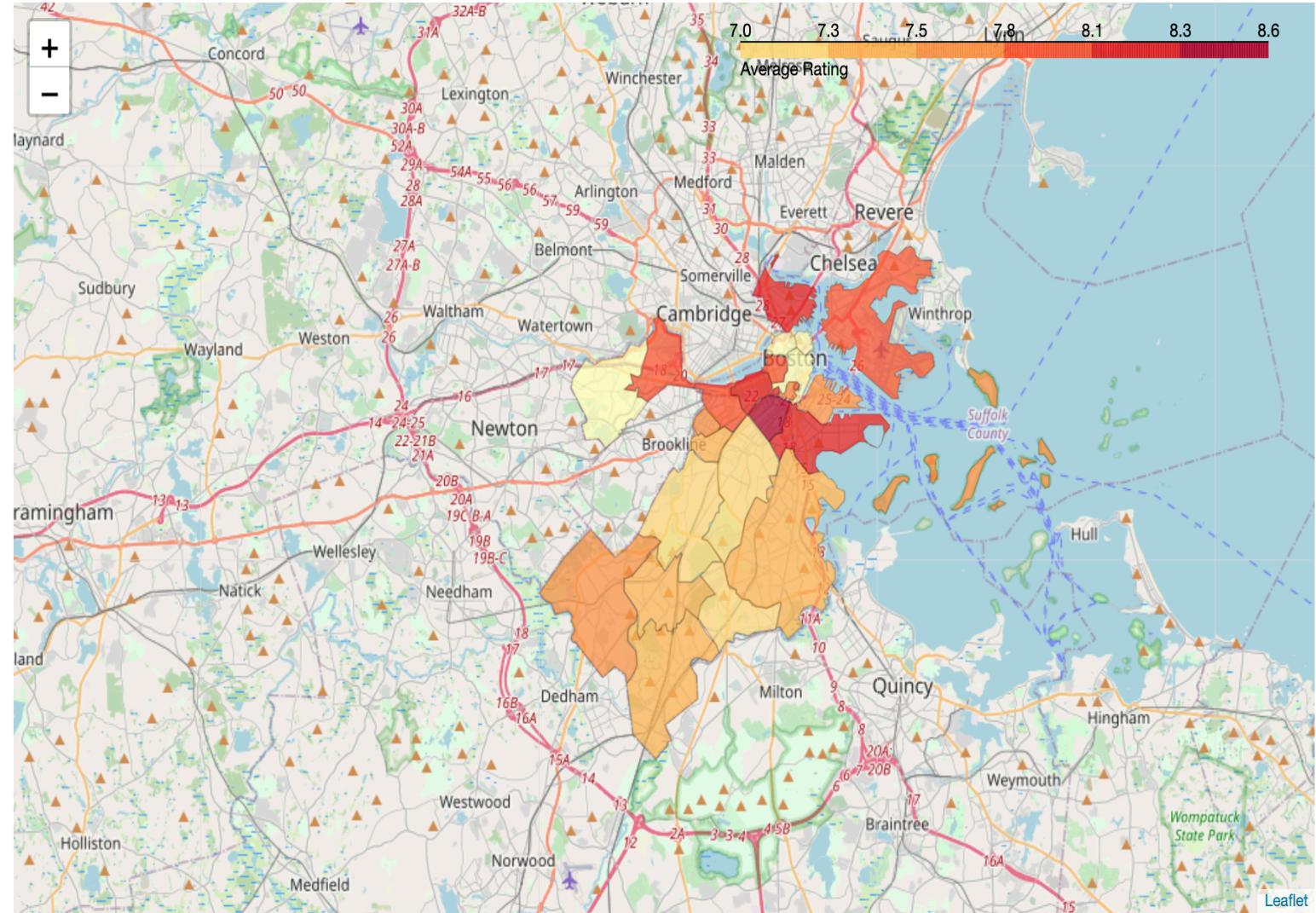
Top Populated Neighborhoods & The Neighborhoods with High Average Ratings



Neighborhoods Location Overview



Heat Map (Average Ratings)



K Means Clustering (16 Highly Rated Neighborhoods)



Cluster 1

```
boston_merged.loc[boston_merged['Cluster Labels'] == 0, boston_merged.columns[[0] + list(range(4, boston_merged.shape[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 South End	American Restaurant	Coffee Shop	Theater	Italian Restaurant	Gym	Wine Bar	Performing Arts Venue	Gourmet Shop	Hotel	Park
5 Allston	Bakery	Chinese Restaurant	Gastropub	Mexican Restaurant	Korean Restaurant	Sushi Restaurant	Pizza Place	Bar	Ice Cream Shop	Liquor Store
6 Fenway	Pizza Place	Gym / Fitness Center	Coffee Shop	Grocery Store	Chinese Restaurant	Bakery	Mexican Restaurant	Yoga Studio	Japanese Restaurant	Pub
12 South Boston Waterfront	Seafood Restaurant	Hotel	Park	Italian Restaurant	Steakhouse	Coffee Shop	Gym	Bakery	Bar	Brewery
13 Harbor Islands	Beach	Park	Seafood Restaurant	Chinese Restaurant	Breakfast Spot	Lighthouse	Bakery	Scenic Lookout	Bar	Diner
14 Longwood	Pizza Place	Coffee Shop	Café	Donut Shop	Sandwich Place	American Restaurant	Middle Eastern Restaurant	Art Museum	Art Gallery	Park

Cluster 2

```
boston_merged.loc[boston_merged['Cluster Labels'] == 1, boston_merged.columns[[0] + list(range(4, boston_merged.shape[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
3 Charlestown	Brewery	Bakery	Pizza Place	Gastropub	Gym	Clothing Store	Sporting Goods Shop	Cosmetics Shop	Department Store	Pub
7 Beacon Hill	Science Museum	American Restaurant	Hotel	Hotel Bar	Italian Restaurant	Playground	Gym / Fitness Center	Pizza Place	Bar	Park
8 Leather District	Chinese Restaurant	Asian Restaurant	Bakery	Coffee Shop	Italian Restaurant	American Restaurant	Gym / Fitness Center	Sushi Restaurant	Sandwich Place	Theater
9 East Boston	Italian Restaurant	Pizza Place	Hotel	Sandwich Place	Beach	Ice Cream Shop	Chinese Restaurant	Donut Shop	American Restaurant	Coffee Shop
10 Chinatown	Chinese Restaurant	Coffee Shop	Asian Restaurant	Bakery	Sandwich Place	Sushi Restaurant	Gym / Fitness Center	Mexican Restaurant	Pizza Place	Vegetarian / Vegan Restaurant
11 West Roxbury	Italian Restaurant	American Restaurant	Ice Cream Shop	Pizza Place	Gym / Fitness Center	Park	Burger Joint	Gym	Greek Restaurant	Chinese Restaurant

Cluster 3

```
boston_merged.loc[boston_merged['Cluster Labels'] == 2, boston_merged.columns[[0] + list(range(4, boston_merged.shape[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
1 South Boston	Coffee Shop	American Restaurant	Bakery	Italian Restaurant	Theater	Wine Bar	Asian Restaurant	Pizza Place	Pet Store	Chinese Restaurant
2 Bay Village	Coffee Shop	American Restaurant	Asian Restaurant	Theater	Spa	Chinese Restaurant	Italian Restaurant	Wine Shop	Hotel	Pizza Place
4 Back Bay	Spa	Italian Restaurant	Ice Cream Shop	Gourmet Shop	Hotel	Plaza	Bakery	Park	Jewelry Store	Coffee Shop

15 Dorchester	Coffee Shop	Bar	Sandwich Place	Italian Restaurant	American Restaurant	Pizza Place	Liquor Store	Gym / Fitness Center	New American Restaurant	Gym
---------------	-------------	-----	----------------	--------------------	---------------------	-------------	--------------	----------------------	-------------------------	-----

Low Rated Neighborhoods for Asian Restaurants

	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	Hyde Park	American Restaurant	Italian Restaurant	Pharmacy	Pizza Place	Ice Cream Shop	Bakery	Grocery Store	Discount Store	Liquor Store	Mexican Restaurant
1	Jamaica Plain	Pizza Place	Park	Gym	Italian Restaurant	Donut Shop	American Restaurant	Sandwich Place	Liquor Store	Coffee Shop	Mexican Restaurant
2	Mattapan	American Restaurant	Ice Cream Shop	Bakery	Coffee Shop	Park	Brewery	Caribbean Restaurant	Pizza Place	Gym / Fitness Center	Italian Restaurant
3	Mission Hill	Café	Sandwich Place	Pizza Place	American Restaurant	Middle Eastern Restaurant	Concert Hall	Italian Restaurant	Art Museum	Thai Restaurant	Coffee Shop
4	Roslindale	American Restaurant	Italian Restaurant	Pharmacy	Pizza Place	Ice Cream Shop	Bakery	Grocery Store	Discount Store	Liquor Store	Mexican Restaurant
5	Roxbury	American Restaurant	Ice Cream Shop	Bakery	Coffee Shop	Park	Brewery	Caribbean Restaurant	Pizza Place	Gym / Fitness Center	Italian Restaurant

Results (summary of K- means Clustering)

- 3 clusters are identified
- **Clusters 1 and 3** are generally recommended regions for opening restaurants
 - the top common venues for those neighborhoods are **complementary and supportive for restaurants dining**.
 - Less competitive than the populated neighborhoods in Cluster 2
 - Also have average ratings >7.5 for Asian restaurants → **the local customers still prefer Asian cuisine and these restaurants can still easily attract customers**.
- **Within the clusters, the recommended neighborhoods are: South End, Allston, Fenway, South Boston Waterfront, Harbor Islands, Longwood, South Boston, Bay Village, Back Bay and Dorchester**
- The following neighborhoods are therefore **not recommended: Hyde Park, Jamaica Plain, Mattapan, Mission Hill, Roslindale and Roxbury**.
 - average rating less than 7.5
 - Top common venues are American/Italian/western restaurants → **the local preferences are western cuisine, and opening Asian restaurants to attract local customers isn't as ideal**.

Discussion

- Business Recommendations: number of Asian restaurants, level of competition, and average ratings
- South End: highly rated, and not as competitive and populated(K-means clustering)
 - Benefits from surrounding populated neighborhoods (possibility of high customer visits)
- Most neighborhoods from Cluster 1 and 3 are also recommended
- Drawbacks: Foursquare API
 - Not guaranteed to be up-to-date
 - Potentially missed out some data-→less comprehensive data coverage
 - Can be resolved if similar data interfaces are introduced-→compare data source coverage

Conclusion

Part I

- Objective: to identify the best neighborhoods for opening Asian restaurants in Boston
- Data Sources: geojson and Foursquare API
- Methodology:
 - Exploratory Data Analysis (descriptive statistics and basic visual)
 - Data Cleaning and Wrangling (Data frames creation and sorting)
 - Data Visualization (bar plots, and geospatial folium and heat map)
 - K means Clustering
- Analysis:
 - Location, frequency distribution, and average rating for Asian restaurants in each neighborhood;
 - top common venue categories are sorted to analyze local economic activities for opening Asian restaurants
 - K-means clustering are used to group similar neighborhoods based on venue categories

Conclusion

Part II

- Improvement:
 - multiple data courses to compare and increase the data information coverage;
 - Different machine learning algorithm can be introduced and compared with K-means clustering performance
- Overall, the analysis procedure is detailed, appropriate and insightful to answer the business objective/problem.
 - Similar business problem such as opening a new gym, coffee shop or school could be analyzed using the same procedure
 - A comprehensive, data-driven approach to solve complex, real-life problem