Final Report of the Battle of Neighborhoods

Applied Data Science Capstone Project

Asian Food in Boston

Aymen Ragguem

April 30, 2021

1. Introduction

1.1 Background

The dispersion of the Asian American population is evident in the high concentration of Asian Americans in various cities and towns in Greater Boston. Figure 2.1 in Link indicates that 16 very different cities and towns had concentrations of Asian Americans of at least 12 percent. While large cities in Greater Boston generally have larger Asian American populations, places with the most rapid Asian American increases have actually been in smaller, suburban locales.

Boston has had **Chinese restaurants** since the late 1800s, most located in Chinatown and catering almost exclusively to the Chinese population; non-Chinese customers went to places that served chop suey and chow mein.Link.

In the late 1950s and early ’60s, fueled by Hawaii’s statehood in 1959, the Lower 48 fell in love with all things Polynesian or from the South Seas. Chinese restaurateurs were quick to **notice the trend**, and so added to their menu tropical drinks and dishes laced with sweet-and-sour sauce and plenty of pineapple. Restaurants were **remodeled** to add a festive, Polynesian touch. Although far removed from Chinese culture, the tiki craze helped make Chinese food or more accurately, **hybrid Chinese-American food** accessible to the wider public.

1.2 Problem Description

The Asian cuisine contains a wide range of cooking practices and traditions, and it varies greatly in taste and flavor to notice the trend.

In this project we will analyze various cuisines in a given location to derive the opportunities of starting up an Asian cuisine to understand the targeted cuisine in Boston, MA such Thai, Chinese and Japanese restaurants and study the level of competition in this field (saturation, growth demand …)

Suffolk, MA was pick up the study for the following reasons :

- One last mention should be made about the Chinese food scene in and around Boston. With a continuous **flow** of Chinese students attending High Schools for nearly 50 years,
- Growth of Asian-American population
- There's tremendous diversity within Boston's Asian American community.

1.3 Interest

For someone interested in starting up a **restaurant business**, this project is a simple starting point on types of data that can be used and the variety of methods to gather them, and how to perform some data analysis to derive the statistics of the various categories of **venues** and the ratio to the **Asian** venues for the cities that falls within the **Boston** greater area.

2. Data description

2.1 Data sources

Each zip code will have latitude and longitude value assigned to a city and each city can have multiple zip code then, the result will grouped by city which will be neighborhood, Suffolk will refer to city.

DataLink : https://simplemaps.com/data/us-zips

The second data set will be the venue data. Details on the venues will be derived from Foursquare.com website via an API to the application.

Foursquare provides a rough guide on the types of cuisine according to a predefined set of categories as documented on its website https://developer.foursquare.com/docs/resources. While it also returns the venues' frequency by neighborhoods which is defined by their zip codes and their respective latitude and longitude.

2.2 Data Cleaning and processing

The first set of data to be processed will be the zip code of USA which contains all states. Suffolk country has **18** neighborhoods, and **33** zip codes. The lat/long of all 33 zip codes will be passed to Foursquares 's API and the venues returned will be grouped by neighborhoods. The map below shows us the distribution of all neighborhoods identified by zip code in the city.

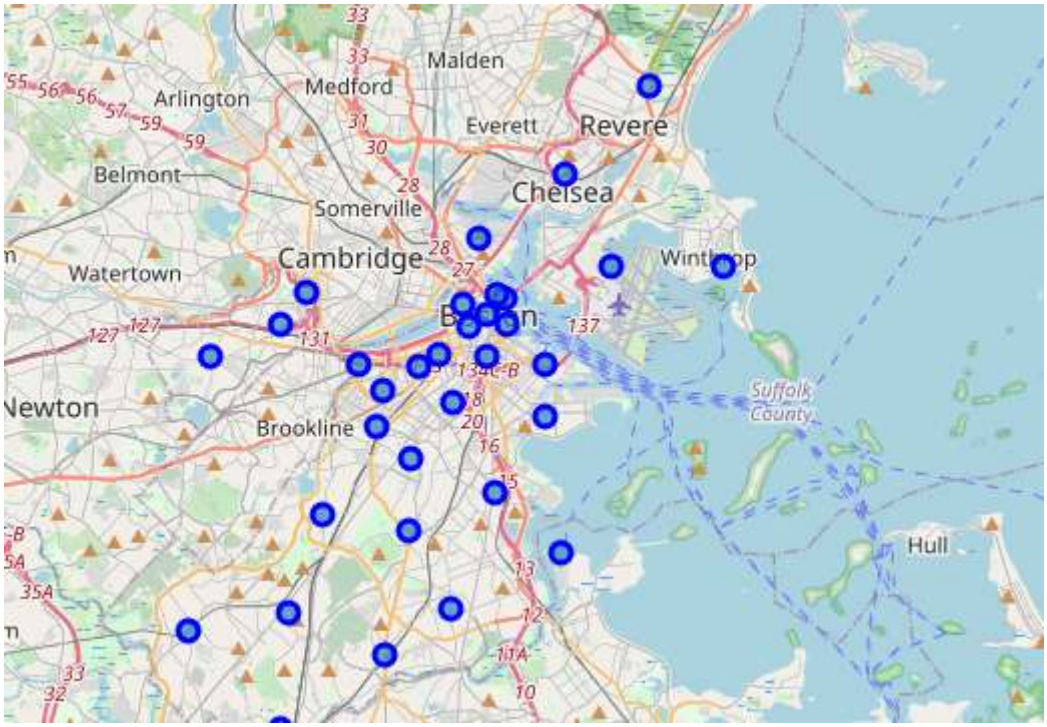

Fig 1. Zip Code distribution across Suffolk country

Foursquare API will provide many categories of the venue data, and we are particularly interested in the Restaurant category in order to help us solve the

business problem. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

Furthermore, the venue data set that Foursquare provides is only a rough guide on the types of cuisine according to a predefined set of categories as Foursquare has documented.

As showing in fig below We search the number of venues for each Neighborhood via Foursquare and we get 81 unique categories

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Allston	100	100	100	100	100	100
Boston	1400	1400	1400	1400	1400	1400
Brighton	100	100	100	100	100	100
Charlestown	100	100	100	100	100	100
Chelsea	100	100	100	100	100	100
Dorchester	277	277	277	277	277	277
Dorchester Center	75	75	75	75	75	75
East Boston	79	79	79	79	79	79
Hyde Park	43	43	43	43	43	43
Jamaica Plain	100	100	100	100	100	100
Mattapan	75	75	75	75	75	75
Revere	97	97	97	97	97	97
Roslindale	100	100	100	100	100	100
Roxbury	100	100	100	100	100	100
Roxbury Crossing	100	100	100	100	100	100
South Boston	100	100	100	100	100	100
West Roxbury	76	76	76	76	76	76
Winthrop	45	45	45	45	45	45

Then we create a new dataframe which contains 10 venues for each neighborhood (18)

2.3 Feature selection

We work on a dataframe uszip.csv which contains the following **18** features :

```
zip                int64
lat              float64
lng              float64
city              object
state_id          object
state_name        object
```

```
zcta                 bool
parent_zcta       float64
population        float64
density           float64
county_fips         int64
county_name        object
county_weights     object
county_names_all   object
county_fips_all    object
imprecise            bool
military             bool
timezone           object
```

We pass to clean up the no necessary row for this project: 'zcta','parent_zcta','density','county_fips','county_weights','imprecise','military', 'timezone'.

We note that population size and density features can help us to identify the Asian people in the country of study

2.4 Analytical Methods

- Understand the correlation between the highest population count and the distribution of Asian venues retrieved by Foursquare API (the ratio of the cuisine and the ratio of the White and Asian demographics)

3. Explore data and analysis

We explore now Asian food by using Foursquare API, for 10 venues and we get the following result :

	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	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	Korean Restaurant	Vietnamese Restaurant	Noodle House	Café	Food Truck
1	Boston	Asian Restaurant	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	Vietnamese Restaurant	Korean Restaurant	Bakery	Dim Sum Restaurant	Food Truck
2	Brighton	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Sushi Restaurant	Korean Restaurant	Japanese Restaurant	Noodle House	Vietnamese Restaurant	Café	Hunan Restaurant
3	Charlestown	Asian Restaurant	Chinese Restaurant	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Vietnamese Restaurant	Korean Restaurant	Noodle House	Restaurant	Breakfast Spot
4	Chelsea	Chinese Restaurant	Vietnamese Restaurant	Asian Restaurant	Sushi Restaurant	Japanese Restaurant	Dim Sum Restaurant	Food Truck	Food Court	Dumpling Restaurant	Diner
5	Dorchester	Chinese Restaurant	Vietnamese Restaurant	Asian Restaurant	Sushi Restaurant	Seafood Restaurant	Japanese Restaurant	Ramen Restaurant	Food Truck	Noodle House	Taco Place
6	Dorchester Center	Vietnamese Restaurant	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Seafood Restaurant	BBQ Joint	Bakery	Breakfast Spot	Café	Grocery Store
7	East Boston	Chinese Restaurant	Asian Restaurant	Vietnamese Restaurant	Sushi Restaurant	Japanese Restaurant	BBQ Joint	Dim Sum Restaurant	Food Truck	Food Court	Dumpling Restaurant
8	Hyde Park	Chinese Restaurant	Thai Restaurant	Vietnamese Restaurant	Convenience Store	Fried Chicken Joint	Food Truck	Food Court	Dumpling Restaurant	Diner	Dim Sum Restaurant
9	Jamaica Plain	Chinese Restaurant	Asian Restaurant	Sushi Restaurant	Ramen Restaurant	Thai Restaurant	Japanese Restaurant	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Food Truck	Noodle House
10	Mattapan	Chinese Restaurant	Thai Restaurant	Vietnamese Restaurant	Convenience Store	Fried Chicken Joint	Food Truck	Food Court	Dumpling Restaurant	Diner	Dim Sum Restaurant
11	Revere	Chinese Restaurant	Asian Restaurant	Thai Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Fried Chicken Joint	Food Truck	Food Court	Dumpling Restaurant	Diner
12	Roslindale	Chinese Restaurant	Thai Restaurant	Asian Restaurant	Sushi Restaurant	Japanese Restaurant	Noodle House	Convenience Store	Food Court	Dumpling Restaurant	Diner
13	Roxbury	Asian Restaurant	Sushi Restaurant	Chinese Restaurant	Japanese Restaurant	Thai Restaurant	Vietnamese Restaurant	Restaurant	Coffee Shop	Convenience Store	Food Court
14	Roxbury Crossing	Asian Restaurant	Chinese Restaurant	Thai Restaurant	Sushi Restaurant	Japanese Restaurant	Vietnamese Restaurant	Noodle House	Korean Restaurant	Food Truck	Indian Restaurant
15	South Boston	Asian Restaurant	Chinese Restaurant	Sushi Restaurant	Japanese Restaurant	Thai Restaurant	Vietnamese Restaurant	Bakery	Dim Sum Restaurant	Food Truck	Noodle House

The Asian venues frequency is built using the "Asian Restaurant" as venue category. Notice that all the neighborhoods that have Chinese and Asian restaurants below top 5, all have high percentage of Asian populations. Noticed that Foursquare returns :

Asian restaurant as a **category**

Chinese restaurant which is a **subcategory**

Ramen Restaurant which is a **subsubcategory**.

4. Interpretation of the results

- **Classification** of Asian cuisine is defined by the source of data, as we see Foursquare returns categories like Asian restaurants (global

results) and then in next iteration it returns Chinese or Japanese restaurants which are subcategory and after this it returns some results like Ramen and dim sum … restaurants which are sub-subcategory; and here there is very difference between all Asian cuisine returned to the this hierarchy.

- **Demographic density**

  There is a correlation between population (**students**) and the count of venues of Asian cuisine, if an area has a very low population count, it will have a low count of venues and naturally a lower count of Asian venues, too. These areas can be considered low opportunity a high risk

- **Conclusion**

  Neighborhood with higher percentage of Asians population such as Boston are worth exploring.