Finding the Best Coffee Shop Locations in Boston

Phuong Truong

July 28, 2021

Introduction

Boston is one of the oldest and most densely populated cities in the United States. It has a rich history, vibrant culture and is considered a world leader in higher education. It also attracts millions of tourists, young students and professionals every year. Therefore, Boston is the perfect place for entrepreneurs to start a business or open a restaurant.

This project aims to find an optimal location for a new coffee shop in Boston. Specifically, this project is targeted to stakeholders interested in opening a coffee shop or a café in Boston. Since there are already so many coffee shops in Boston, we need to locate areas that have no or few coffee shops and are densely populated. We also prefer locations that have many universities and young students as these are the main customers for a coffee shop.

Foursquare API and k-means clustering algorithm are used to segment the neighborhoods in Boston. Stakeholders can choose the best neighborhoods for their coffee shops based on the above criteria accordingly.

Data

Based on the definition of the problem, factors that influence the decision are: number of existing café or coffee shops in the neighborhood and population of the neighborhood. In order to segment the neighborhoods and explore them, I need a dataset that contains the neighborhoods as well as the zip code, population and latitude and longitude coordinates of each neighborhood. The following data sources were utilized to extract and generate the required information:

- Boston neighborhoods based on zip code charts from The Boston Globe [1] and Zip Code Business Patterns [2]. Some neighborhoods were removed because they are not located in the City of Boston (Cambridge, Brookline and Somerville). Some zip codes that are attached to a single high-volume address were also removed.

- Massachusetts zip codes by population data from US Census [3].

- I used the Geocoder Python package to get the geographical coordinates of each neighborhood. [4]

- I used Foursquare API to explore common venues in Boston's neighborhoods.

Methodology

In this project, I detected areas of Boston that have large population and low number of coffee shops. First, I collected the required data: zip codes, population, latitude and longitude of every neighborhood. I used geopy library to get the coordinates of Boston, and Folium library to create a map of Boston with neighborhoods superimposed on top.

In the next step, I used the Foursquare API to explore the most common venues in each neighborhood. I set the limit of number of venues as 100 and the radius as 500 meters for every neighborhood. Then I used one-hot encoding and grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category to get the top 10 venue categories for each neighborhood.

In the final step, I used k-means clustering algorithm to group the neighborhoods into 5 clusters. I visualized the resulting clusters using Folium library, and examined each cluster to determine if its location meets the criteria to open a coffee shop.

Results

The Foursquare API returned a total of 1402 venues in 225 unique categories for all 29 neighborhoods in Boston. The two figures below show all neighborhoods ranked by population and by total number of venues.

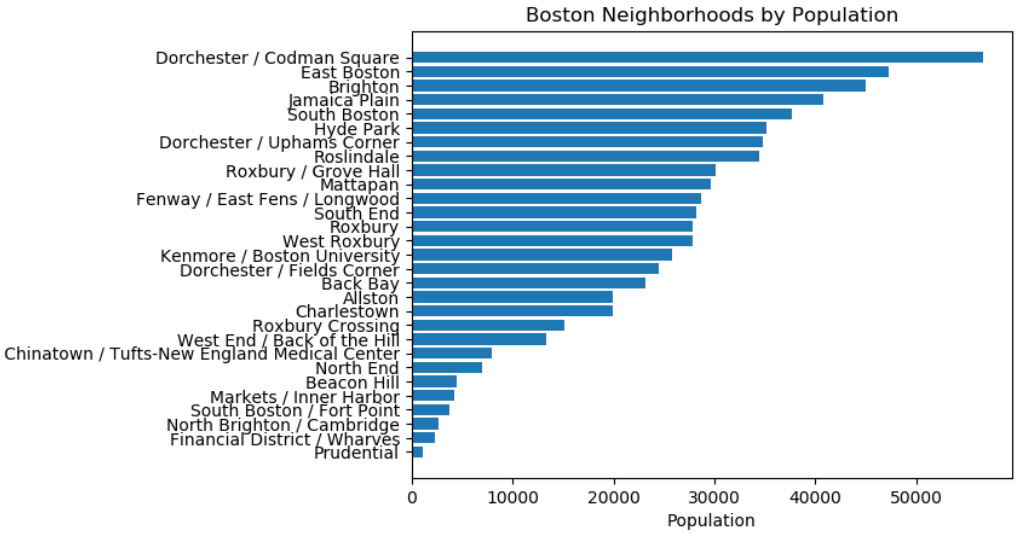

Figure 1: Boston Neighborhoods by Population

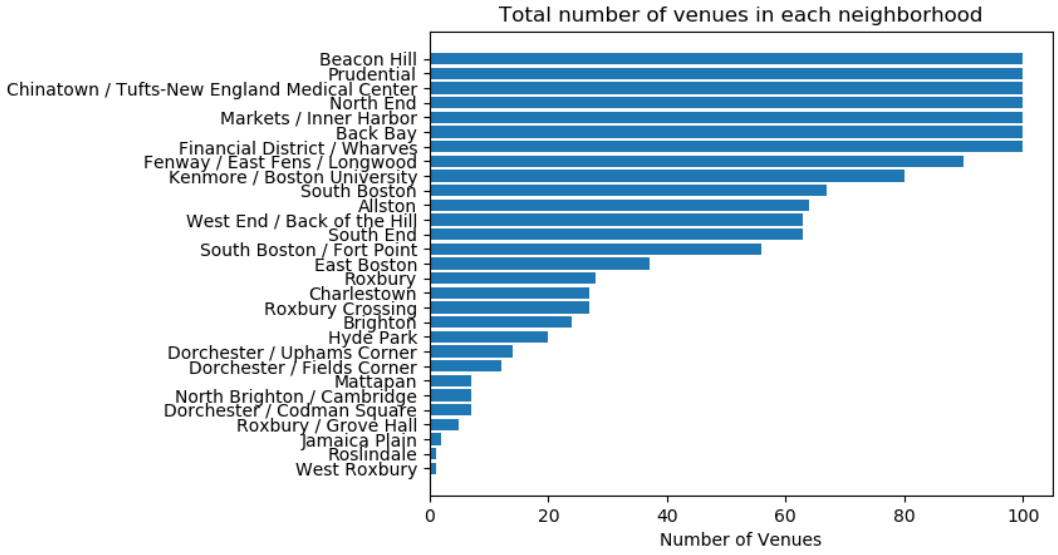

Figure 2: Total number of venues in each neighborhood

From Figure 1 and Figure 2, we can observe that most neighborhoods in the downtown area such as Beacon Hill, Back Bay and Financial District have high number of venues but have small populations, while most residential areas in the suburbs such as Jamaica Plain, Roslindale and Roxbury have low number of venues and large populations.

The map below shows the resulting 5 clusters after running k-means algorithm (Cluster 0: Red, Cluster 1: Purple, Cluster 2: Blue, Cluster 3: Green, Cluster 4: Orange).

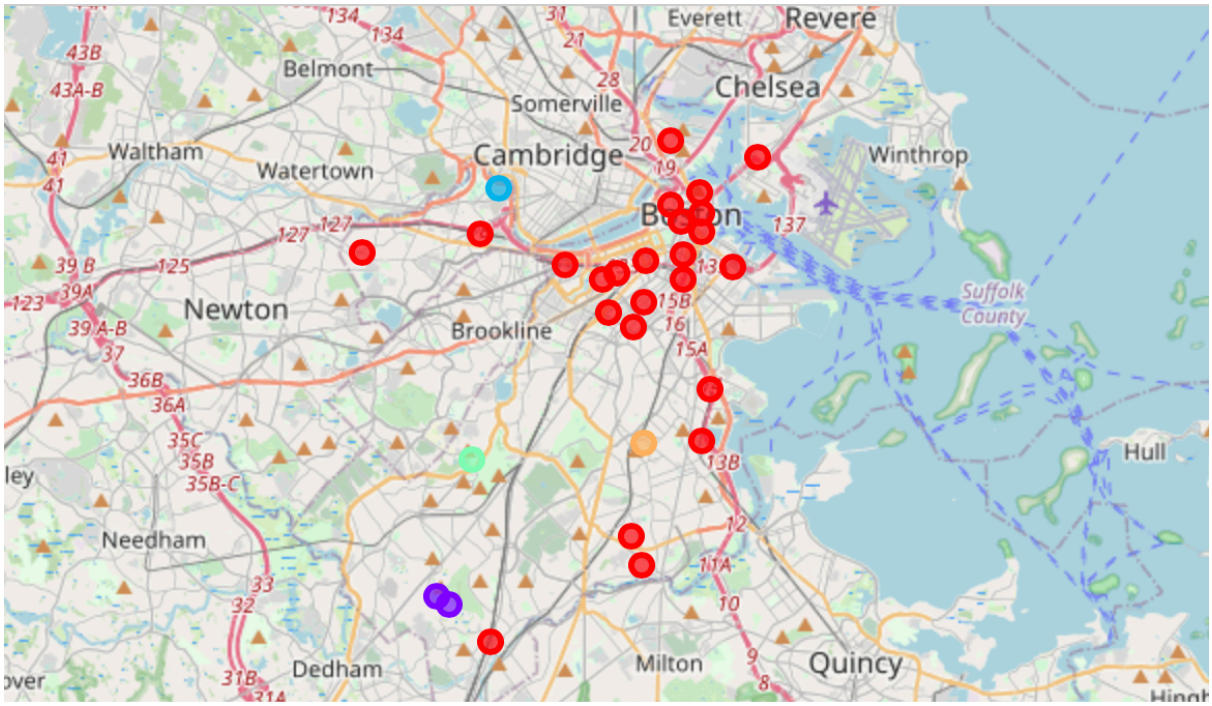

Figure 3: Boston neighborhoods and their clusters

There are 24 neighbors in cluster 0, 2 neighbors in cluster 1, and 1 neighbor in each cluster 2, 3, 4. The tables below show detailed results of each cluster, including neighborhoods, populations, clusters labels, and the 10 most common venue categories.

Neighborhood	Population	Cluster Labels	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
Dorchester / Codman Square	56649	0	Donut Shop	Gym / Fitness Center	American Restaurant	Indian Restaurant	Hardware Store	BBQ Joint	Liquor Store	Farmers Market	Ethiopian Restaurant	Falafel Restaurant
East Boston	47263	0	Latin American Restaurant	Pharmacy	Park	Sandwich Place	Convenience Store	Fast Food Restaurant	Department Store	Liquor Store	Fried Chicken Joint	Deli / Bodega
Brighton	44951	0	Chinese Restaurant	Pizza Place	Park	Sushi Restaurant	Trail	Latin American Restaurant	Noodle House	Bar	Wings Joint	Thai Restaurant
South Boston	37634	0	Japanese Restaurant	Coffee Shop	Italian Restaurant	Grocery Store	Hotel	Art Gallery	Asian Restaurant	Bookstore	New American Restaurant	Pharmacy
Hyde Park	35213	0	American Restaurant	Pizza Place	Grocery Store	Convenience Store	Train Station	Plaza	Caribbean Restaurant	Donut Shop	Sandwich Place	Bar
Dorchester / Uphams Corner	34847	0	Café	Coffee Shop	Pub	Deli / Bodega	Hotel	Bar	Vietnamese Restaurant	Pizza Place	Bank	Liquor Store
Mattapan	29620	0	Convenience Store	Sandwich Place	Baseball Field	Grocery Store	Chinese Restaurant	Pharmacy	Liquor Store	Fast Food Restaurant	Farmers Market	Dumpling Restaurant
Fenway / East Fens / Longwood	28652	0	Clothing Store	Coffee Shop	Ice Cream Shop	Concert Hall	Shoe Store	Hotel Bar	Bookstore	Plaza	Theater	Sandwich Place
South End	28262	0	Pizza Place	Donut Shop	Latin American Restaurant	Café	Sandwich Place	Coffee Shop	Park	Tapas Restaurant	Deli / Bodega	Mediterranean Restaurant

Table 1: First 9 neighborhoods, populations, and 10 most common venues in cluster 0

	Neighborhood	Population	Cluster Labels	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
7	Roslindale	34419	1	Lake	Yoga Studio	Dumpling Restaurant	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market	Fast Food Restaurant
13	West Roxbury	27801	1	Lake	Yoga Studio	Dumpling Restaurant	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market	Fast Food Restaurant

Table 2: Neighborhoods, populations, and 10 most common venues in cluster 1

	Neighborhood	Population	Cluster Labels	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
26	North Brighton / Cambridge	2621	2	Park	College Hockey Rink	Pool	Residential Building (Apartment / Condo)	Gym	College Stadium	Dumpling Restaurant	Electronics Store	Ethiopian Restaurant	Falafel Restaurant

Table 3: Neighborhood, population, and 10 most common venues in cluster 2

	Neighborhood	Population	Cluster Labels	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	Jamaica Plain	40749	3	Monument / Landmark	Home Service	Yoga Studio	Dumpling Restaurant	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish Market

Table 4: Neighborhood, population, and 10 most common venues in cluster 3

	Neighborhood	Population	Cluster Labels	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	Roxbury / Grove Hall	30159	4	Food	Brewery	Garden	Discount Store	Ethiopian Restaurant	Food Service	Food Court	Food & Drink Shop	Flower Shop	Flea Market

Table 5: Neighborhood, population, and 10 most common venues in cluster 4

Discussion

After examining each cluster, we can observe that:

- Cluster 0: This cluster includes neighborhoods that already have a lot of coffee shops, cafes and restaurants. These neighborhoods are located in the downtown area (commercial zone) or residential areas with large populations. It's best to avoid these areas when opening a new coffee shop.

- Cluster 1: These are residential areas in the suburbs of Boston. The most common venues are lakes, flower shops, flea and fish markets. There are not many coffee shops or cafes in these neighborhoods. The populations of these neighborhoods are quite large (ranked 7 and 13). These are good locations to open a coffee shop.

- Cluster 2: This is the university area with many parks, sports venues, and apartment buildings. There are not many coffee shops in this neighborhood. This is a good place to open a coffee shop as there are many young college students in this area.

- Cluster 3: This is the one of the most heavily populated (ranked 3) residential areas in Boston. There are not many coffee shops of cafes in this area. There are some landmarks in this area which may attract tourists. This is a very good location to open a coffee shop.

- Cluster 4: This is the residential area in the suburb of Boston. It has a relatively large population (ranked 8). There are many food venues but not many coffee shops. This is also a good location to open a coffee shop.

The resulting clusters may not imply that these areas are actually optimal locations for a new coffee shop. Our analysis only focused on areas that are densely populated but are not crowded with coffee shops. There could be many reasons why opening a coffee shop in these areas might not be feasible even if there are few competitors, such as high rents or poor public transport accessibility. Besides, choosing a different number of k can yield different clusters and we may detect other suitable areas for a coffee shop. Therefore, additional factors need to be considered in future studies.

Conclusion

Using Foursquare API and k-means clustering algorithm, my project successfully identified the areas in Boston that are heavily populated but are not crowded with coffee shops. Stakeholders can carefully consider these recommended neighborhoods and their advantages as a starting point to make an appropriate decision on choosing a final, optimal location to open a coffee shop in Boston. In future works, different clustering algorithms and additional factors such as residents' average income, transport accessibility and rental price may be implemented in order to improve the correctness of this analysis.

Reference

1. http://archive.boston.com/news/local/articles/2007/04/15/sixfigurezipcodes_city/
2. Boston Redevelopment Authority Research Division, May 2016.
3. https://www.massachusetts-demographics.com/zip_codes_by_population
4. https://geocoder.readthedocs.io/index.html.