Housing Sales Prices and Venues Data Analysis of Boston Neighborhoods

by Nataliia Romanenko

1. Introduction

Being a recent graduate, I seriously think about relocation from a small town in Eastern Connecticut to a big city with more job opportunities like Boston and New York. There is a lot of information available for choosing a place to live in New York; however, there is significantly less information (at least gathered in one place) about choosing a place to live in Boston. This project will focus on answering the question about choosing a neighborhood in Boston to help people like me who consider a job with relocation to Boston.

Boston is a very diverse city with a population of 694,583 in 2018 on 48 square miles (124 km^2). It is sometimes called a "city of neighborhoods" because of the profusion of diverse subsections; the city government has officially designated 23 neighborhoods [1]. In this project, we will analyze venues data for each neighborhood, segment and cluster the venues, and combine this information with the median price for a condo in each neighborhood, so that people with different preferences will have a full picture for choosing a neighborhood to live in Boston.

2. Data

For the project, we will use the following data and tools:

- Boston Neighborhoods Geospatial Dataset from BostonMaps Open Data managed by the City of Boston's GIS Team [2]. This *geojson* file contains polygon coordinates for each Boston neighborhood that will be used to create choropleth maps.
- Foursquare API to get the most common venues of each Boston neighborhood [3].
- Condo Prices in Boston 2018 Report from Boston magazine showing median condo prices in Boston neighborhoods [4]

3. Methodology

In this project, we will analyze venues data for each Boston neighborhood, segment and cluster the venues, and combine this information with the median price for a condo in each neighborhood.

Geospatial data for Boston neighborhood is taken from Boston Neighborhoods Geospatial Dataset and used to create choropleth map using Python library Folium. The choropleth map shows median condo prices in different neighborhoods in Boston using a color palette, where color corresponds to price value.

Price data is obtained by web-scraping using Python library Beautiful Soup from Condo Prices in Boston 2018 Report from Boston magazine.

Venue data for each neighborhood is obtained using Foursquare API, that returns a list of venues in a given radius from the coordinates of the neighborhood.

Venue Clustering is performed using unsupervised machine learning algorithm k-means, that divides the data into non-overlapping clusters based on feature similarities: intra-cluster distances are minimized, and inter-cluster distances are maximized. We use k-means to group neighborhoods based on the most popular types of venues.

We will start with Data Wrangling: get Boston neighborhoods geo data and median condo price data, combine them and visualize. Then, we will explore Boston neighborhoods using Foursquare API: get the nearby venues for each neighborhood. After getting venues, we will analyze each neighborhood by taking the mean of the frequency of occurrence of each venue category and organize these data into a data frame that has the top 10 venues for each neighborhood. Then, we will cluster the neighborhoods into 5 clusters using k-means machine learning algorithm. Finally, we examine the clusters and visualize all the data we have in a single map.

4. Results

During Data Wrangling step the following information was obtained for Boston neighborhoods:

	Neighborhood	Area	Latitude	Longitude	2017 Median Price
0	Roslindale	2.51	42.291209	-71.124497	450000
1	Jamaica Plain	3.94	42.309820	-71.120330	534000
2	Mission Hill	0.55	42.332926	-71.103214	NaN
3	Longwood	0.29	42.336168	-71.099527	NaN
4	Bay Village	0.04	42.350011	-71.066948	615000
5	Leather District	0.02	42.351049	-71.057969	850000
6	Chinatown	0.12	42.352217	-71.062607	850000
7	North End	0.20	42.365097	-71.054495	570500
8	Roxbury	3.29	42.324843	-71.095016	338000
9	South End	0.74	42.341310	-71.077230	615000
10	Back Bay	0.62	42.350707	-71.079730	1100000
11	East Boston	4.71	42.375097	-71.039217	454500
12	Charlestown	1.36	42.377875	-71.061996	690000
13	West End	0.30	42.363919	-71.063899	570500
14	Beacon Hill	0.31	42.359820	-71.066162	952500
15	Downtown	0.62	42.361339	-71.069152	940000
16	Fenway	0.88	42.343451	-71.097716	571000
17	Brighton	2.88	42.350097	-71.156442	430000
18	West Roxbury	5.49	42.279265	-71.149497	361025
19	Hyde Park	4.57	42.255654	-71.124496	295000
20	Mattapan	2.11	42.280738	-71.090284	244000
21	Dorchester	7.29	42.297320	-71.074495	429950
22	South Boston Waterfront	0.97	42.333431	-71.049495	857500
23	South Boston	2.25	42.333431	-71.049495	678000
24	Allston	1.56	42.355434	-71.132127	480000
25	Harbor Islands	1.29	42.352822	-70.891882	NaN

Table 1. Boston Neighborhoods geospatial data with median condo price in 2017

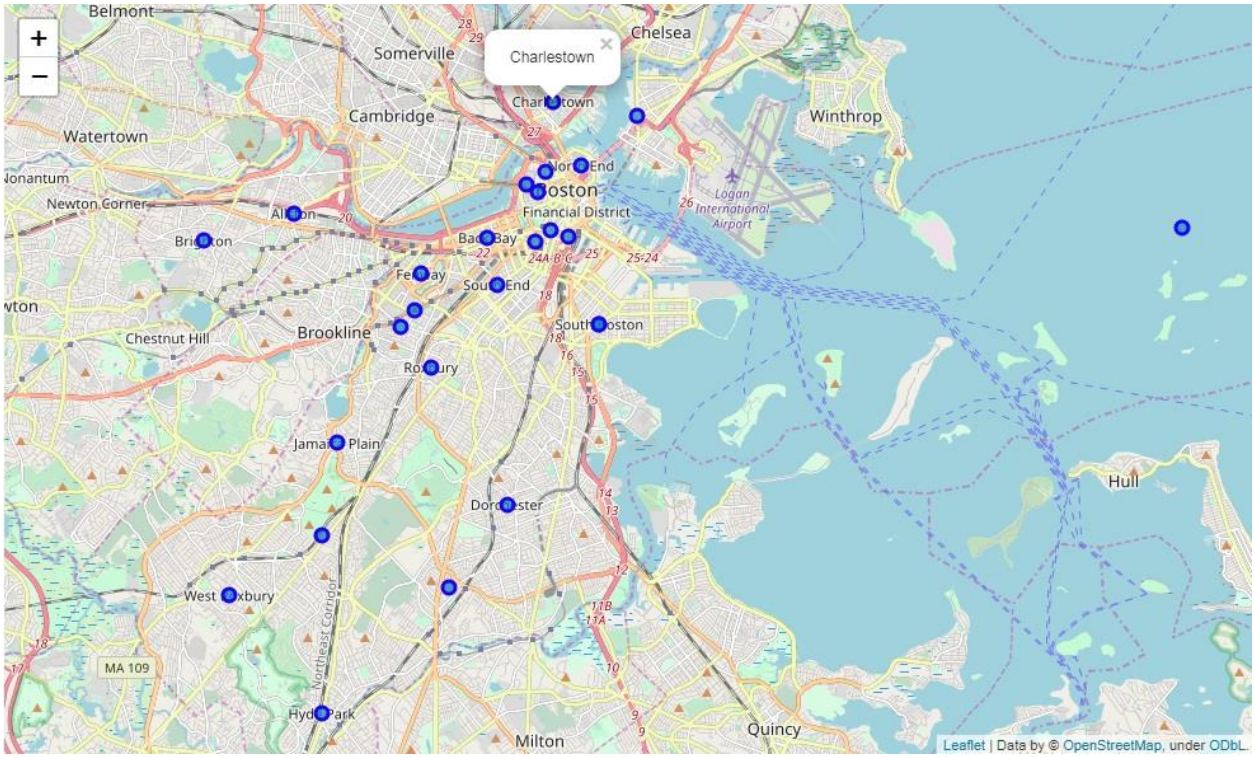

Figure 1. Boston neighborhoods coordinates visualized on a map

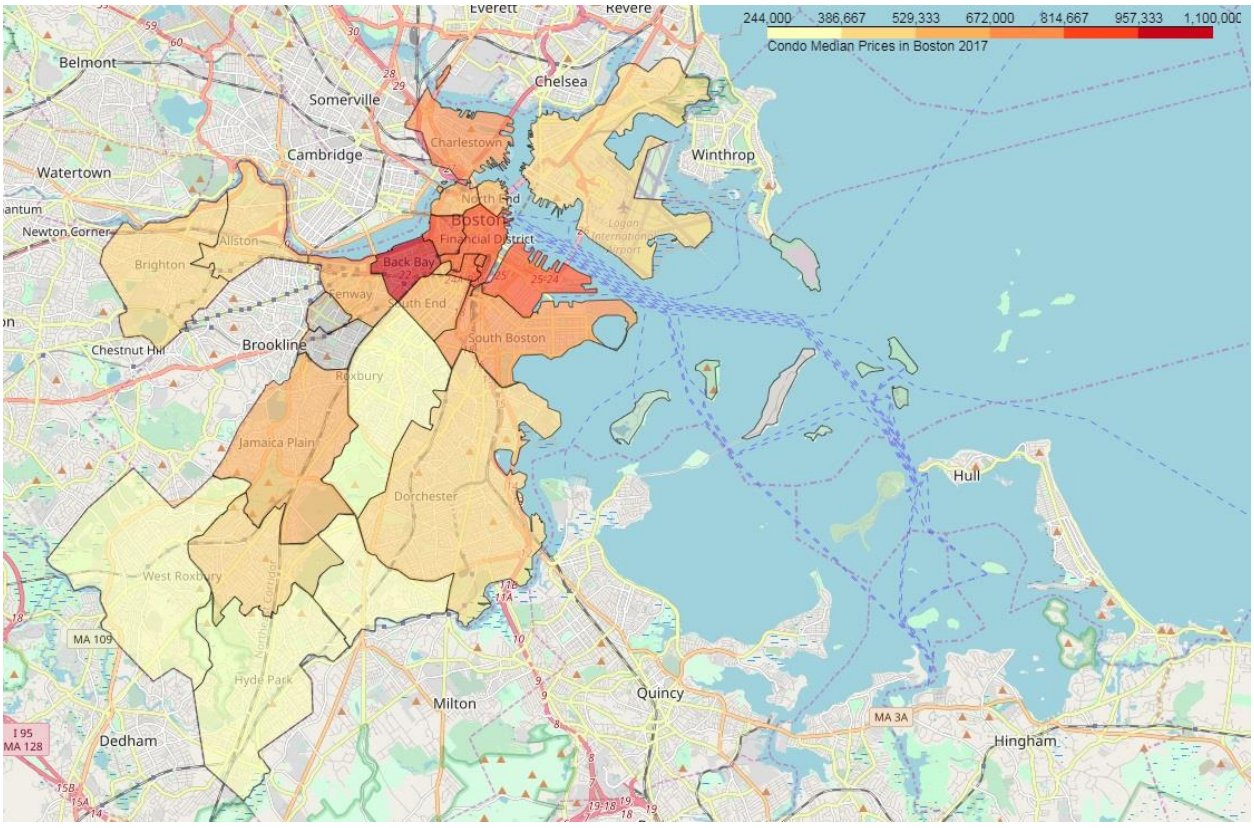

Figure 2 Choropleth map visualizing median condo price in Boston in 2017

While exploring Boston neighborhoods using Foursquare API, the following information was obtained. A total number of venues returned for all Boston neighborhood was 1529. In total, there were 225 unique venue categories. An example of data gathered for each neighborhood is provided below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Roslindale	42.291209	-71.124497	Peters Hill	42.293617	-71.128063	Scenic Lookout
1	Roslindale	42.291209	-71.124497	Roslindale House Of Pizza	42.287989	-71.126549	Pizza Place
2	Roslindale	42.291209	-71.124497	Delfino's	42.287106	-71.129470	Italian Restaurant
3	Roslindale	42.291209	-71.124497	Roslindale Village Farmers Market	42.286534	-71.128509	Farmers Market
4	Roslindale	42.291209	-71.124497	Fornax Bread Company	42.286171	-71.129760	Bakery

Table 2 An Example of venue data gathered for each Boston neighborhood

A number of venues for each Boston neighborhood is summarized below:

Neighborhood	Venue
Allston	82
Back Bay	100
Bay Village	95
Beacon Hill	64
Brighton	81
Charlestown	66
Chinatown	100
Dorchester	20
Downtown	48
East Boston	42
Fenway	100
Hyde Park	28
Jamaica Plain	57
Leather District	100
Longwood	24
Mattapan	24
Mission Hill	21
North End	100
Roslindale	42
Roxbury	37
South Boston	77
South Boston Waterfront	42
South End	44
West End	85
West Roxbury	50

Table 3 Number of venues for each Boston neighborhood

After analyzing venue data, the following results were obtained:

	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	Pizza Place	Bakery	Korean Restaurant	Chinese Restaurant	Gym / Fitness Center	Bar	Café	Sushi Restaurant	Diner	Coffee Shop
1	Back Bay	American Restaurant	Coffee Shop	Hotel	Italian Restaurant	Clothing Store	Seafood Restaurant	Salon / Barbershop	Spa	Sporting Goods Shop	Cosmetics Shop
2	Bay Village	Sandwich Place	Hotel	Italian Restaurant	Theater	Steakhouse	American Restaurant	Performing Arts Venue	Gym / Fitness Center	Bakery	Vegetarian / Vegan Restaurant
3	Beacon Hill	Hotel Bar	Pizza Place	Italian Restaurant	Gift Shop	American Restaurant	Plaza	Hotel	Sushi Restaurant	Coffee Shop	Gourmet Shop
4	Brighton	Pizza Place	Convenience Store	Café	Coffee Shop	Pub	Bakery	Donut Shop	Chinese Restaurant	Dry Cleaner	Greek Restaurant
5	Charlestown	Café	Park	Gastropub	Pizza Place	Gym	History Museum	Grocery Store	Pub	Donut Shop	Athletics & Sports
6	Chinatown	Chinese Restaurant	Asian Restaurant	Bakery	Sushi Restaurant	Theater	Coffee Shop	Sandwich Place	Pizza Place	Performing Arts Venue	Seafood Restaurant
7	Dorchester	Pharmacy	Pizza Place	Liquor Store	Golf Course	Park	Diner	Discount Store	Sandwich Place	Fast Food Restaurant	Metro Station
8	Downtown	Hotel Bar	Gift Shop	Italian Restaurant	Gourmet Shop	Pizza Place	Hotpot Restaurant	Liquor Store	Ice Cream Shop	Kids Store	Korean Restaurant
9	East Boston	Mexican Restaurant	Park	Italian Restaurant	Latin American Restaurant	Pizza Place	Seafood Restaurant	Café	Bar	Burrito Place	Fast Food Restaurant
10	Fenway	Sports Bar	Pizza Place	Coffee Shop	Lounge	Baseball Field	Restaurant	American Restaurant	Thai Restaurant	Liquor Store	Mexican Restaurant
11	Hyde Park	Pizza Place	Park	Pharmacy	American Restaurant	Grocery Store	Platform	Theater	Sandwich Place	Donut Shop	Gas Station
12	Jamaica Plain	Park	Bakery	Coffee Shop	Pizza Place	Art Gallery	Bookstore	Thrift / Vintage Store	Yoga Studio	American Restaurant	Accessories Store
13	Leather District	Coffee Shop	Asian Restaurant	Chinese Restaurant	Bakery	Sandwich Place	Sushi Restaurant	Vegetarian / Vegan Restaurant	American Restaurant	Food Truck	Hotpot Restaurant
14	Longwood	Sandwich Place	Donut Shop	Italian Restaurant	Platform	Pub	Falafel Restaurant	Café	Bookstore	Liquor Store	Gastropub
15	Mattapan	Caribbean Restaurant	Pizza Place	Liquor Store	Scenic Lookout	Hot Dog Joint	Ice Cream Shop	Indian Restaurant	Hardware Store	Gym / Fitness Center	Donut Shop
16	Mission Hill	Pizza Place	Sandwich Place	Sushi Restaurant	Convenience Store	Coffee Shop	Falafel Restaurant	Café	Caribbean Restaurant	Gastropub	Donut Shop
17	North End	Italian Restaurant	Bakery	Pizza Place	Seafood Restaurant	Park	Café	Sandwich Place	Market	Playground	Coffee Shop
18	Roslindale	Pizza Place	American Restaurant	Grocery Store	Bar	Plaza	Italian Restaurant	Sandwich Place	Liquor Store	Rental Car Location	Discount Store
19	Roxbury	Donut Shop	Pizza Place	Italian Restaurant	Convenience Store	Skating Rink	Recreation Center	Plaza	Bed & Breakfast	Supermarket	Furniture / Home Store
20	South Boston	Pizza Place	Bar	Donut Shop	Sandwich Place	Italian Restaurant	Coffee Shop	Gym	Convenience Store	Sports Bar	Beach
21	South Boston Waterfront	Pizza Place	Italian Restaurant	Bar	Sports Bar	Chinese Restaurant	Liquor Store	Coffee Shop	Donut Shop	Sushi Restaurant	Dog Run
22	South End	Italian Restaurant	Coffee Shop	Wine Shop	Wine Bar	Park	Bar	Gift Shop	Bakery	Yoga Studio	Salon / Barbershop
23	West End	Sandwich Place	Pizza Place	Donut Shop	Coffee Shop	Bar	Hotel	Café	Italian Restaurant	Sports Bar	Gym / Fitness Center
24	West Roxbury	Pizza Place	Pharmacy	Convenience Store	Liquor Store	Bank	Park	Gift Shop	Donut Shop	American Restaurant	Grocery Store

Table 4 Most common venues for each neighborhood

Using k-means we obtained five clusters. After examining each cluster, we determined the discriminating venue categories that distinguish each cluster:

- The most common venues for cluster 1:
  Cafes, Restaurants, Parks, and Playgrounds
- The most common venues for cluster 2:
  Fast Food Restaurants, Pubs/Bars, and Parks
- The most common venues for cluster 3:
  Asian Cuisine Restaurants and Art venues
- The most common venues for cluster 4:
  Grocery / Convenience Stores, Restaurants / Cafes, Liquor Stores
- The most common venues for cluster 5:
  Parks, Hotels, and Gift/Gourmet Shops

After overlaying venue clusters on median condo price choropleth, we generated the following map, where area color corresponds to condo price and marker color corresponds to venue cluster:

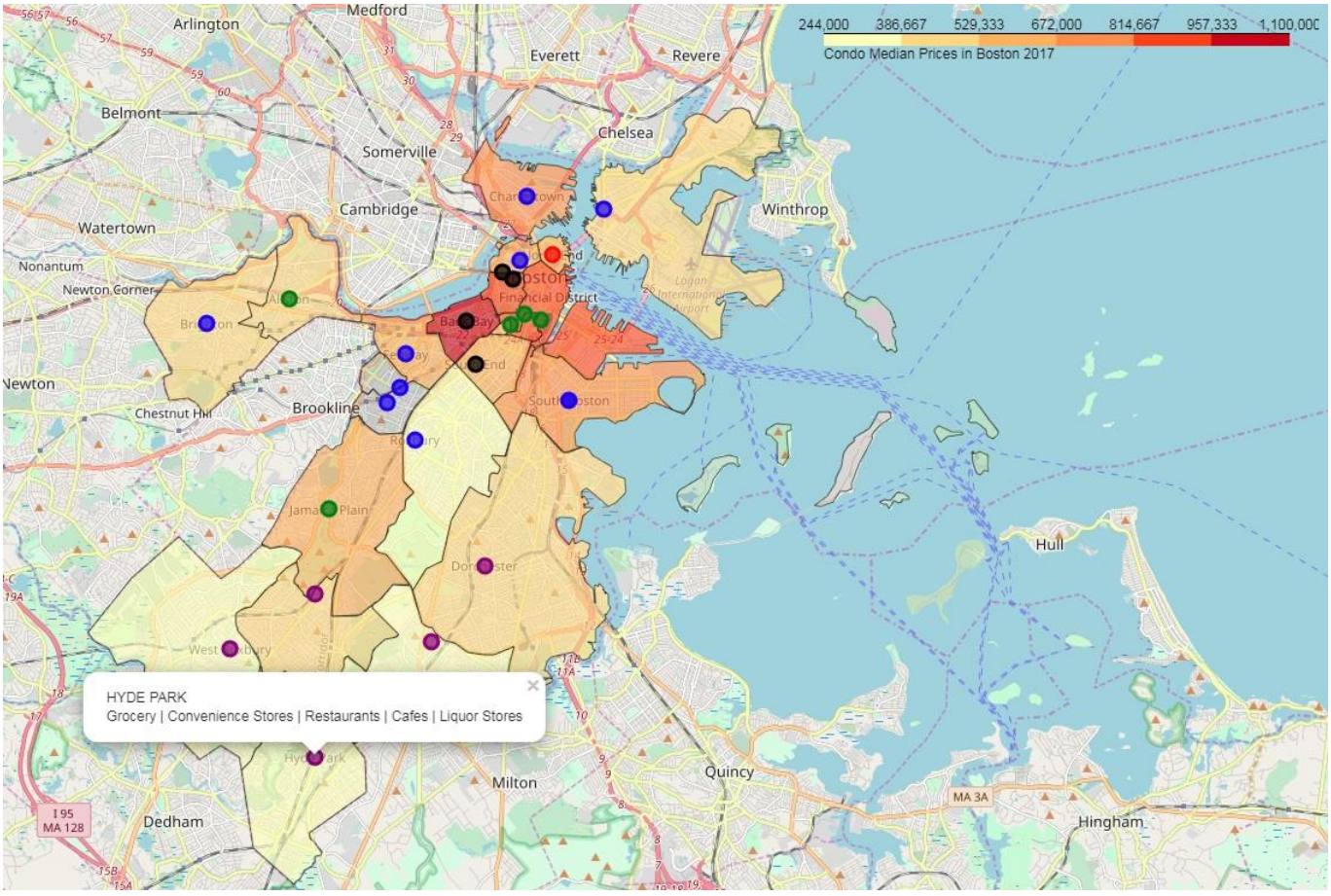

Figure 3 Median Condo Price and Venue Clusters in Boston

5. Discussion

Boston is really a very diverse city, sometimes called a "city of neighborhoods" because of the profusion of diverse subsections. This project focused on answering the question about choosing a neighborhood in Boston to help people like me who consider a job with relocation to Boston.

We started with Data Wrangling step and obtained Boston neighborhoods geo data and median condo price data, combined them and visualized the price data using choropleth map, where color represents price value.

We also performed venue analysis: we explored Boston neighborhoods using Foursquare API to get the nearby venues for each neighborhood. After getting venues, we analyzed each neighborhood by taking the mean of the frequency of occurrence of each venue category and organize these data into a data frame that has the top 10 venues for each neighborhood. Then, we clustered the neighborhoods into 5 clusters using k-means machine learning algorithm. After examining each cluster, we determined the discriminating venue categories that distinguish each cluster. Finally, we visualized all the data we have in a single map by overlaying venue clusters on median condo price choropleth, where area color corresponds to condo price and marker color corresponds to venue cluster. Using this map one can explore different options and see which neighborhood can be a good fit given a popular venue profile and a housing (condo) price.

6. Conclusion

More and more people are moving to big cities because they provide more job opportunities. Sometimes, it is hard to choose a neighborhood to live, and the type of analysis performed here can be a good starting point at choosing a place to live when people can make an informed decision based on housing price and venue profile of the neighborhood.

References:

1. Wikipedia - Boston
2. Boston Neighborhoods Geospatial Dataset
3. Foursquare API
4. Condo Prices in Boston 2018 Report
5. the Elliman report