

# Boston Housing Price Analysis

## Introduction

*Boston is one of the biggest and the 21st most populous city in the US. The city is also placed among the top 30 most economically powerful cities in the world. Consider all these factors, Boston always attracts real estate investor's attention when it comes down to real estate investment. However, Boston consists of several neighborhoods, and each neighborhood has its own characteristic. Therefore, it's important to analyze the relationship between the neighborhood's venue data and its housing price, and see what's the most common venues in high housing price neighborhood.*

## Data Description

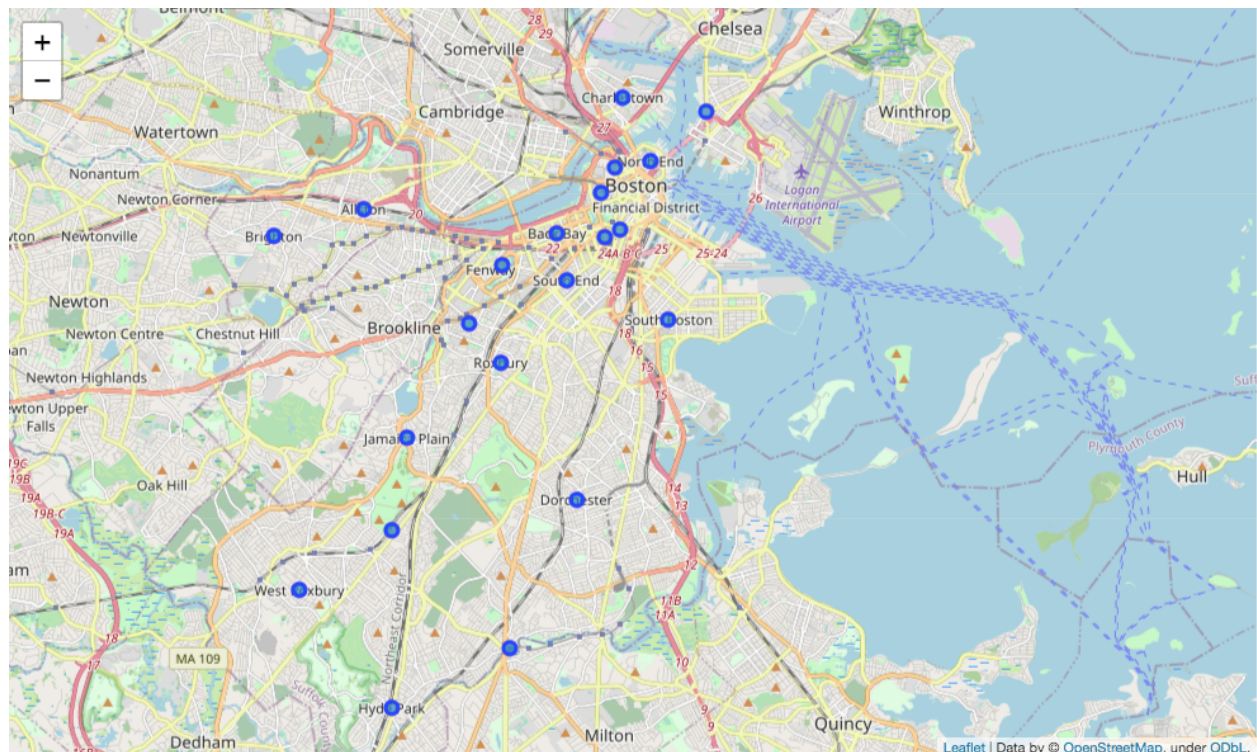
*The Data I am going to use in this project are listed as below:1. The list of the neighborhood in Boston from Wikipedia 2. Foursquare API to get the most common venues of given Neighborhood in Boston3. The average housing price of each neighborhood from redfin. Since it is hard to find the aggregated data online, I manually record the data from the redfin website and put them into a CSV file.*

## Methodology

*As a database, I used GitHub repository in my study. My master data include neighborhood, Housing price per Sq. Ft, each neighborhood's Latitude, and Longitude.*

	City	Neighborhood	Latitude	Longitude	Housing Price per Sq.Ft
0	Boston	Allston	42.3554	-71.1321	950.0
1	Boston	Back Bay	42.3507	-71.0797	1100.0
2	Boston	Bay Village	42.35	-71.0669	1130.0
3	Boston	Beacon Hill	42.3587	-71.0678	1120.0
4	Boston	Brighton	42.3501	-71.1564	552.0
5	Boston	Charlestown	42.3779	-71.062	789.0
6	Boston	Chinatown	42.3513	-71.0626	1120.0
7	Boston	Dorchester	42.2973	-71.0745	471.0
8	Boston	East Boston	42.3751	-71.0392	571.0
9	Boston	Fenway Kenmore	42.3442	-71.0944	1080.0
10	Boston	Hyde Park	42.2557	-71.1245	326.0
11	Boston	Jamaica Plain	42.3098	-71.1203	586.0
12	Boston	Mattapan	42.2676	-71.0924	273.0

I used folium to visualize geographic details of Boston and its neighborhood, and I create a map of Boston with neighborhoods superimposed on top. I used latitude and longitude values to get the visual as below.

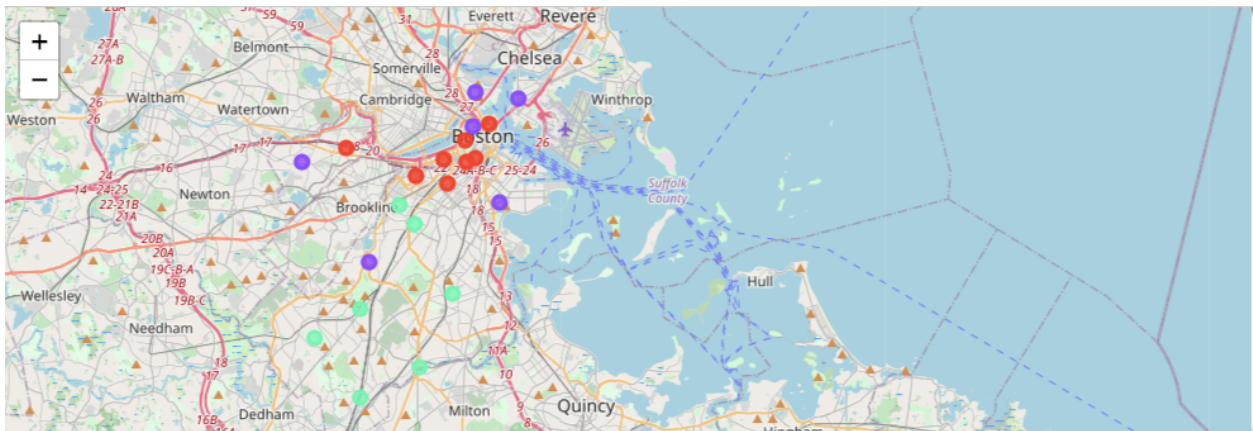


I also utilized the Foursquare API to explore the boroughs and segment them. I designed the limits as 100 venue and the radius 500 meter for each neighborhood from their given latitude and longitude information. Below is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

	name	categories	lat	lng
0	Lulu's Allston	Comfort Food Restaurant	42.355068	-71.134107
1	Azama Grill	Falafel Restaurant	42.354422	-71.132358
2	Allston Diner	Diner	42.354979	-71.134295
3	Fish Market Sushi Bar	Sushi Restaurant	42.353039	-71.132975
4	Kaju Tofu House	Korean Restaurant	42.354329	-71.132374

In summary of this data, 844 venues were returned by Foursquare.

In order to analyze the venue data, I used unsupervised learning K-means algorithm to cluster the neighborhood. First, I use `get_dummies` function to turn the categorical value into numeric value. Then I import `StandardScaler` from `Sklearn` to standardize the data. This process make sure the housing price value won't have much more influence on clustering than any other numeric value. Then I set the number of cluster to 3 and label the each neighborhood. The result shows below.



## Result

By calculating the average housing price of each cluster, the result shows that three cluster's housing price has significant differences.

**Cluster0 Average Housing Price: 1067.875**

**Cluster1 Average Housing Price: 647.0**

**Cluster2 Average Housing Price: 405.7142857142857**

Therefore, I named the cluster 0 as high price area, cluster 2 as middle price area, and cluster 3 as low price area.

In high price area venue data set, we can see that the most common venue for each neighborhood usually Korean restaurant, Italian restaurant, Chinese restaurant, and hotel.

	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	Housing Price per Sq.Ft
0	Allston	Korean Restaurant	Pizza Place	Bakery	Chinese Restaurant	Thai Restaurant	Mexican Restaurant	Fried Chicken Joint	Italian Restaurant	Dive Bar	Asian Restaurant	950.0
1	Back Bay	Hotel	Italian Restaurant	Coffee Shop	Clothing Store	Seafood Restaurant	American Restaurant	Grocery Store	Plaza	Ice Cream Shop	Shopping Mall	1100.0
2	Bay Village	Sandwich Place	Bakery	Hotel	Hotel Bar	Asian Restaurant	Spa	Theater	Performing Arts Venue	Italian Restaurant	Restaurant	1130.0
3	Beacon Hill	Italian Restaurant	Pizza Place	French Restaurant	Gift Shop	Gourmet Shop	Hotel	Hotel Bar	Park	Outdoor Sculpture	Other Repair Shop	1120.0
4	Chinatown	Chinese Restaurant	Asian Restaurant	Bakery	Coffee Shop	Performing Arts Venue	Sushi Restaurant	Bubble Tea Shop	Café	Theater	Sandwich Place	1120.0
5	Fenway Kenmore	Pizza Place	Sports Bar	Lounge	American Restaurant	Thai Restaurant	Hotel	Japanese Restaurant	Restaurant	Café	Coffee Shop	1080.0
6	North End	Italian Restaurant	Seafood Restaurant	Bakery	Pizza Place	Park	Spa	Sandwich Place	Grocery Store	Coffee Shop	Market	983.0
7	South End	Italian Restaurant	Wine Shop	Wine Bar	Park	Bakery	Bar	Coffee Shop	French Restaurant	Gift Shop	Playground	1060.0

*In middle price area venue, the most common venue are usually pizza place and coffee shop, bakery and pharmacy.*

	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	Housing Price per Sq.Ft
0	Brighton	Bakery	Bank	Pizza Place	Grocery Store	Dry Cleaner	Bus Station	Chinese Restaurant	Pub	Coffee Shop	Sandwich Place	552.0
1	Charlestown	Pizza Place	Convenience Store	Pub	Coffee Shop	Gastropub	Donut Shop	Yoga Studio	National Park	Plaza	Pharmacy	789.0
2	East Boston	Pharmacy	Art Gallery	Sandwich Place	Pizza Place	Latin American Restaurant	Park	Community Center	Department Store	Deli / Bodega	Convenience Store	571.0
3	Jamaica Plain	Coffee Shop	Bakery	Art Gallery	Bar	Seafood Restaurant	Bookstore	Library	Liquor Store	Pizza Place	Noodle House	586.0
4	South Boston	Pizza Place	Chinese Restaurant	Sports Bar	Coffee Shop	Italian Restaurant	Liquor Store	Bar	Cosmetics Shop	Park	Sandwich Place	793.0
5	West End	Pizza Place	Sandwich Place	Hotel	Donut Shop	Café	Bar	Sports Bar	Italian Restaurant	Coffee Shop	Mexican Restaurant	591.0

*When we look at the low price area, the most common venue are usually metro station, sandwich place, park, and home service.*

	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	Housing Price per Sq.Ft
0	Dorchester	Pizza Place	Caribbean Restaurant	Breakfast Spot	Vegetarian / Vegan Restaurant	Southern / Soul Food Restaurant	Market	Gym	Plaza	Fried Chicken Joint	Donut Shop	471.0
1	Hyde Park	American Restaurant	Pizza Place	Platform	Grocery Store	Fast Food Restaurant	Donut Shop	Discount Store	Plaza	Bar	Bank	326.0
2	Mattapan	Metro Station	Pizza Place	Bakery	Pharmacy	Convenience Store	Caribbean Restaurant	Mobile Phone Shop	Shoe Store	Southern / Soul Food Restaurant	Bank	273.0
3	Mission Hill	Sandwich Place	Pizza Place	Sushi Restaurant	Bank	Café	Caribbean Restaurant	Donut Shop	Coffee Shop	Pub	Gastropub	501.0
4	Roslindale	Yoga Studio	Big Box Store	Donut Shop	Cuban Restaurant	Pool	Rental Car Location	Scenic Lookout	Pizza Place	American Restaurant	Athletics & Sports	446.0
5	Roxbury	Park	Gym	Metro Station	Plaza	Art Gallery	Rental Car Location	Furniture / Home Store	Doctor's Office	Falafel Restaurant	Electronics Store	404.0
6	West Roxbury	Home Service	Wine Shop	Doctor's Office	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Dry Cleaner	419.0

## Discussion

*By looking at the result, we can see that the neighborhood's housing price are certainly tied to the category of venue. The neighborhood with more Italian restaurant, Korean restaurant, and Chinese restaurant have higher housing price compare with other neighborhood, where doesn't have many these restaurant.*

*The housing price is affected by many other factors other than category of venue. However, by analyzing the category of venue in certain neighborhood, we can definitely get some insights regarding to its housing price.*

## Conclusion

*When a real estate investor want to invest in certain neighborhood, he/she can start analyzing the neighborhood 's venue category to gain some insights regard this neighbor. However, the housing price analysis can be completed by adding more other factors such as education recourse, population density to get more accurate result.*