

Capstone Project: Comparing Neighborhoods in New York and Boston

Introduction:

Imagine you are a realtor who is working with a new company, ABC123, LLC., that just moved from New York to Boston. The company wants to relocate its workforce from New York to Boston. You and your company have been contracted by ABC123, LLC. to find homes/apartments in neighborhoods that make each employee feel at home. How can this be accomplished? This is a relevant issue for many realtors because the post-buy satisfaction of the homeowners is a key metric in measuring the success of a realtor. Moreover, people are becoming more focused on the available amenities in the area around them rather than the house itself when buying housing. People would be more willing to buy a house that isn't perfect for them if there are all the necessary and desired amenities nearby.

Data:

The following data will be used:

Zip codes (neighborhoods) of Boston and New York City and corresponding latitude and longitude data. Foursquare data of venues based on the latitudes and longitudes of those neighborhoods. The above will allow us to gather data on types of neighborhoods in New York and Boston and segment and cluster them into groups based on their similarities. This will give us the ability to make recommendations on comparable neighborhoods in Boston.

Importing and Cleaning the Raw Data to make final Location DataFrame

Import Zipcodes CSV

```
US_loc_data = pd.read_csv('zip_code_database.csv', converters={'zip': lambda x: str(x)})
```

Cleaning the data and making a table with just New York and Boston Zipcodes and Latitude/Longitude data

```
MAlocdata = US_loc_data[US_loc_data['state']=='MA']
NYlocdata = US_loc_data[US_loc_data['state']=='NY']

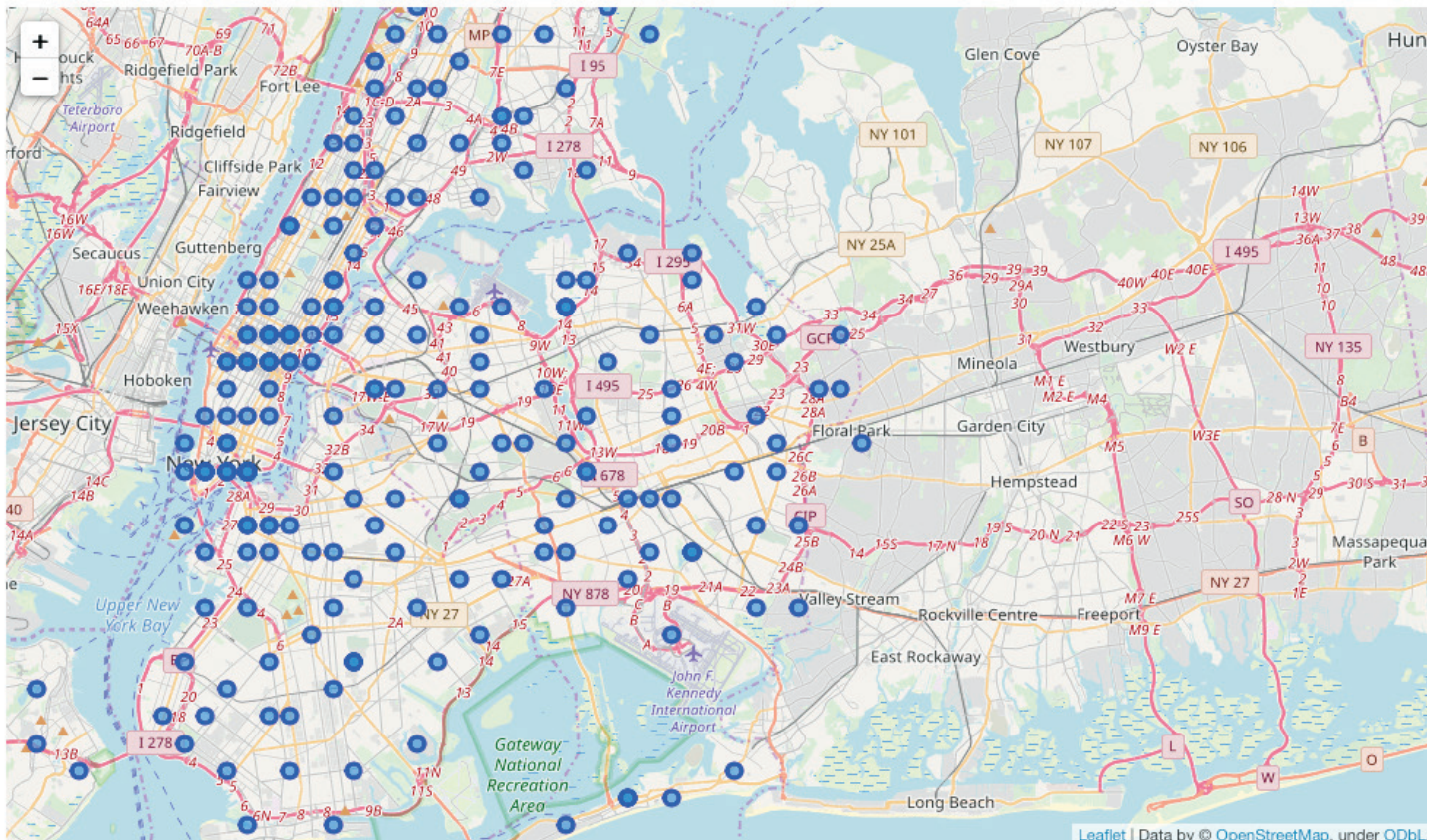
MAlocdata = MAlocdata[(MAlocdata['county']=='Suffolk County')]
MAlocdata.head()
```

	zip	type	decommissioned	primary_city	acceptable_cities	unacceptable_cities	state	county	timezone	area
612	02108	STANDARD	0	Boston	NaN	NaN	MA	Suffolk County	America/New_York	617,857,339
613	02109	STANDARD	0	Boston	NaN	NaN	MA	Suffolk County	America/New_York	617
614	02110	STANDARD	0	Boston	NaN	NaN	MA	Suffolk County	America/New_York	508,774,617,781
615	02111	STANDARD	0	Boston	NaN	NaN	MA	Suffolk County	America/New_York	617,781,978,339
616	02112	PO BOX	0	Boston	NaN	NaN	MA	Suffolk County	America/New_York	617

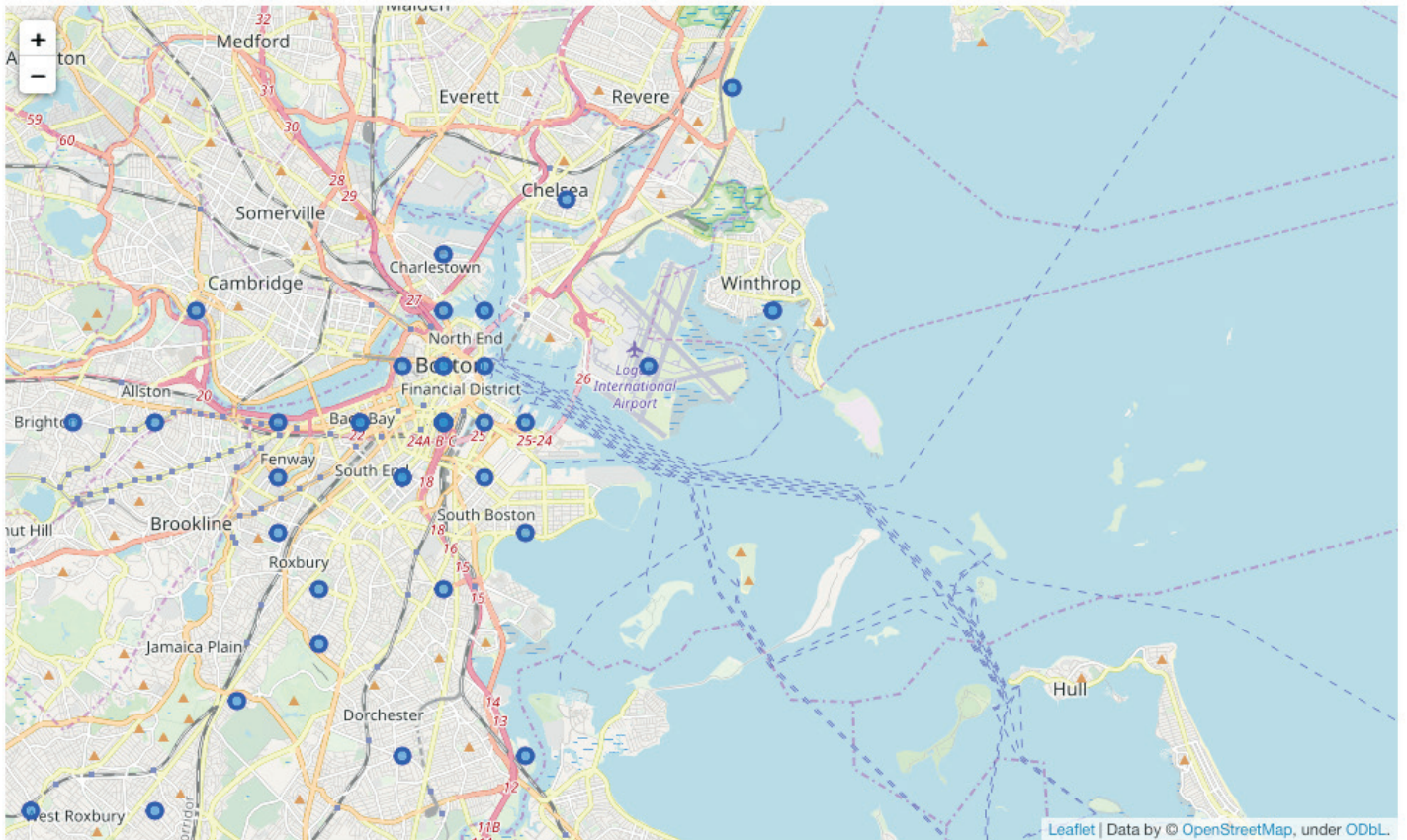
Zip Code dataframe with longitude and latitude data

	zip	city	county	latitude	longitude
0	02108	Boston	Suffolk County	42.36	-71.06
1	02109	Boston	Suffolk County	42.37	-71.05
2	02110	Boston	Suffolk County	42.36	-71.05
3	02111	Boston	Suffolk County	42.35	-71.06
4	02112	Boston	Suffolk County	42.35	-71.06

Zip Codes of New York City and Boston mapped



New York City zip codes



Pulling Foursquare venue data and merging it into a new dataframe

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	10001	40.75	-74.0	Bluestone Lane	40.752254	-73.998824	Coffee Shop
1	10001	40.75	-74.0	Porteño	40.750443	-74.002407	Argentinian Restaurant
2	10001	40.75	-74.0	Jun-Men Ramen Bar	40.747956	-74.000301	Ramen Restaurant
3	10001	40.75	-74.0	26th Street Viewing Spur	40.749825	-74.003352	Scenic Lookout
4	10001	40.75	-74.0	Hudson Market	40.750678	-74.002087	Grocery Store

Clustering Data

	Neighborhood	city	county	latitude	longitude	Cluster	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
0	10001	New York	New York County	40.75	-74.00	0	Art Gallery	Coffee Shop	Deli / Bodega	Chinese Restaurant	Pizza Place	Cocktail Bar	Tapas Restaurant	Lounge
1	10002	New York	New York County	40.71	-73.99	0	Café	Malay Restaurant	Art Gallery	French Restaurant	Other Great Outdoors	Motel	Cantonese Restaurant	Mexican Restaurant
2	10003	New York	New York County	40.73	-73.99	0	Japanese Restaurant	Grocery Store	Italian Restaurant	Coffee Shop	Dessert Shop	Ice Cream Shop	Gym	Yoga Studio
3	10004	New York	New York County	40.69	-74.02	8	Food Truck	Seafood Restaurant	Café	Food Stand	Ice Cream Shop	Bike Rental / Bike Share	Beer Garden	Gym / Fitness Center
4	10005	New York	New York County	40.71	-74.01	8	Coffee Shop	Hotel	Memorial Site	Pizza Place	Sandwich Place	Bar	Park	Café

Analyzing Clustering Data

Cluster 0:

```
nybos_cluster0 = nybos_merged[nybos_merged['Cluster']==0]
nybos_cluster0.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
0	10001	New York	New York County	40.75	-74.00	0	Art Gallery	Coffee Shop	Deli / Bodega	Chinese Restaurant	Pizza Place	Cocktail Bar	Tapas Restaurant	Lounge
1	10002	New York	New York County	40.71	-73.99	0	Café	Malay Restaurant	Art Gallery	French Restaurant	Other Great Outdoors	Motel	Cantonese Restaurant	Mexican Restaurant
2	10003	New York	New York County	40.73	-73.99	0	Japanese Restaurant	Grocery Store	Italian Restaurant	Coffee Shop	Dessert Shop	Ice Cream Shop	Gym	Yoga Studio
5	10006	New York	New York County	40.71	-74.01	0	Coffee Shop	Hotel	Memorial Site	Pizza Place	Sandwich Place	Bar	Park	Café
6	10007	New York	New York County	40.71	-74.01	0	Coffee Shop	Hotel	Memorial Site	Pizza Place	Sandwich Place	Bar	Park	Café

Cluster 1:

```
nybos_cluster1 = nybos_merged[nybos_merged['Cluster']==1]
nybos_cluster1.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
379	02211	Boston	Suffolk County	42.35	-71.06	1	Asian Restaurant	Chinese Restaurant	Bakery	Coffee Shop	Sushi Restaurant	Theater	Sandwich Place	Pizza Place

Cluster 2:

```
nybos_cluster2 = nybos_merged[nybos_merged['Cluster']==2]
nybos_cluster2.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
59	10094	New York	New York County	40.71	-74.00	2	Café	Ice Cream Shop	Chinese Restaurant	Italian Restaurant	Dim Sum Restaurant	Bar	Bakery	Grocery Store
65	10102	New York	New York County	40.71	-73.99	2	Café	Malay Restaurant	Art Gallery	French Restaurant	Other Great Outdoors	Motel	Cantonese Restaurant	Mexican Restaurant
98	10150	New York	New York County	40.71	-73.99	2	Café	Malay Restaurant	Art Gallery	French Restaurant	Other Great Outdoors	Motel	Cantonese Restaurant	Mexican Restaurant
99	10151	New York	New York County	40.71	-73.99	2	Café	Malay Restaurant	Art Gallery	French Restaurant	Other Great Outdoors	Motel	Cantonese Restaurant	Mexican Restaurant
101	10153	New York	New York County	40.76	-73.97	2	Spa	American Restaurant	Coffee Shop	Hotel	Boutique	New American Restaurant	Italian Restaurant	Chinese Restaurant

Cluster 3:

```
nybos_cluster3 = nybos_merged[nybos_merged['Cluster']==3]
nybos_cluster3.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
86	10123	New York	New York County	40.71	-73.99	3	Café	Malay Restaurant	Art Gallery	French Restaurant	Other Great Outdoors	Motel	Cantonese Restaurant	Mexican Restaurant
218	11204	Brooklyn	Kings County	40.62	-73.98	3	Pharmacy	Grocery Store	Pizza Place	Middle Eastern Restaurant	Zoo	Farmers Market	English Restaurant	Espresso Shop
227	11213	Brooklyn	Kings County	40.67	-73.94	3	Pizza Place	Grocery Store	Café	Deli / Bodega	Fried Chicken Joint	Candy Store	Clothing Store	Espresso Shop
234	11220	Brooklyn	Kings County	40.64	-74.02	3	Pizza Place	Mexican Restaurant	Grocery Store	Ice Cream Shop	Bakery	American Restaurant	School	Espresso Shop
250	11237	Brooklyn	Kings County	40.70	-73.92	3	Coffee Shop	Mexican Restaurant	Bar	Pizza Place	Deli / Bodega	Latin American Restaurant	Dive Bar	Italian Restaurant

Cluster 4

```
nybos_cluster4 = nybos_merged[nybos_merged['cluster']==4]
nybos_cluster4.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
12	10013	New York	New York County	40.72	-74.00	4	Clothing Store	Women's Store	Café	Yoga Studio	Men's Store	Bakery	Furniture / Home Store	Art
16	10017	New York	New York County	40.75	-73.97	4	Italian Restaurant	Japanese Restaurant	Coffee Shop	Sushi Restaurant	Steakhouse	Park	Asian Restaurant	De Bo
18	10019	New York	New York County	40.77	-73.99	4	Park	Gym / Fitness Center	Theater	Coffee Shop	Sculpture Garden	Hotel Bar	Wine Bar	Ho
24	10025	New York	New York County	40.80	-73.97	4	Pizza Place	Chinese Restaurant	Mexican Restaurant	Coffee Shop	Indian Restaurant	Italian Restaurant	Bagel Shop	Gn Str
29	10030	New York	New York County	40.82	-73.94	4	Southern / Soul Food Restaurant	Bus Station	BBQ Joint	Rental Car Location	Baseball Field	Lounge	Tennis Court	Taj Re

Cluster 5:

```
nybos_cluster5 = nybos_merged[nybos_merged['cluster']==5]
nybos_cluster5.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
222	11208	Brooklyn	Kings County	40.67	-73.87	5	Deli / Bodega	Sandwich Place	Discount Store	Food Truck	Supermarket	Pizza Place	Bus Stop	
224	11210	Brooklyn	Kings County	40.63	-73.95	5	Mobile Phone Shop	Caribbean Restaurant	Pharmacy	Chinese Restaurant	Deli / Bodega	Restaurant	Breakfast Spot	
225	11211	Brooklyn	Kings County	40.71	-73.95	5	Bar	Coffee Shop	Pizza Place	Italian Restaurant	Mexican Restaurant	Bakery	Japanese Restaurant	
226	11212	Brooklyn	Kings County	40.66	-73.91	5	Restaurant	Fried Chicken Joint	Airport Terminal	Park	Discount Store	Pizza Place	Caribbean Restaurant	
228	11214	Brooklyn	Kings County	40.60	-74.00	5	Rental Car Location	Ice Cream Shop	Hookah Bar	Park	Shanghai Restaurant	Donut Shop	Pizza Place	

Cluster 6:

```
nybos_cluster6 = nybos_merged[nybos_merged['cluster']==6]
nybos_cluster6.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
249	11236	Brooklyn	Kings County	40.64	-73.9	6	Caribbean Restaurant	Cosmetics Shop	Taco Place	Bagel Shop	Gym	Pizza Place	Fast Food Restaurant	Mc Ph Sh

Cluster 7:

```
nybos_cluster7 = nybos_merged[nybos_merged['cluster']==7]
nybos_cluster7.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
255	11242	Brooklyn	Kings County	40.64	-73.94	7	Supermarket	Playground	Rental Car Location	Food & Drink Shop	Food	Moving Target	Fast Food Restaurant	

Cluster 8:

```
nybos_cluster8 = nybos_merged[nybos_merged['cluster']==8]
nybos_cluster8.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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
3	10004	New York	New York County	40.69	-74.02	8	Food Truck	Seafood Restaurant	Café	Food Stand	Ice Cream Shop	Bike Rental / Bike Share	Beer Garden	Gyr Fitn Cer
4	10005	New York	New York County	40.71	-74.01	8	Coffee Shop	Hotel	Memorial Site	Pizza Place	Sandwich Place	Bar	Park	Caf
9	10010	New York	New York County	40.74	-73.98	8	Indian Restaurant	American Restaurant	Ice Cream Shop	Italian Restaurant	Bar	Coffee Shop	Grocery Store	Thri Vint Sto
15	10016	New York	New York County	40.75	-73.98	8	Hotel	Coffee Shop	Sandwich Place	Japanese Restaurant	Italian Restaurant	Gym / Fitness Center	Chinese Restaurant	Am Res
23	10024	New York	New York County	40.80	-73.97	8	Pizza Place	Chinese Restaurant	Mexican Restaurant	Coffee Shop	Indian Restaurant	Italian Restaurant	Bagel Shop	Gro Sto

Cluster 9:

```
nybos_cluster9 = nybos_merged[nybos_merged['Cluster']==9]
nybos_cluster9.head()
```

	Neighborhood	city	county	latitude	longitude	Cluster	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 Common Venue
233	11219	Brooklyn	Kings County	40.63	-74.00	9	Chinese Restaurant	Supermarket	Italian Restaurant	Bus Station	Bar	Bank	Even Space
357	02129	Charlestown	Suffolk County	42.38	-71.06	9	Pizza Place	Liquor Store	Boat or Ferry	Monument / Landmark	Gastropub	Historic Site	Historic Museum
378	02210	Boston	Suffolk County	42.35	-71.04	9	Seafood Restaurant	Café	Donut Shop	Steakhouse	American Restaurant	Art Museum	Harbor Marina
381	02215	Boston	Suffolk County	42.35	-71.10	9	Sports Bar	Coffee Shop	American Restaurant	Hotel	Lounge	Hot Dog Joint	Fast Restaurant
384	02222	Boston	Suffolk County	42.35	-71.06	9	Asian Restaurant	Chinese Restaurant	Bakery	Coffee Shop	Sushi Restaurant	Theater	Sandwich Place

Results:

The algorithm gave us clear clusters which could be used to effectively compare neighborhoods in New York and Boston. An interesting point to note is that there were significantly more zip codes in New York compared to Boston. As a result, there were a few clusters with only 1 zip code per city. Listed below is a breakdown of zip codes by cluster.

Cluster 0: 138 NYC, 9 Boston

Cluster 1: 0 NYC, 1 Boston

Cluster 2: 81 NYC, 0 Boston

Cluster 3: 23 NYC, 6 Boston

Cluster 4: 33 NYC, 15 Boston

Cluster 5: 0 NYC, 2 Boston

Cluster 6: 1 NYC, 0 Boston

Cluster 7: 1 NYC, 0 Boston

Cluster 8: 52 NYC, 20 Boston

Cluster 9: 1 NYC, 5 Boston

Discussion:

As we can see, Boston has fewer zip codes than New York. This presents us with a problem as in cluster2 which has 81 NYC zip codes but none for Boston and cluster 0 which is heavily weighted with NYC zip codes. However the clustering data does provide us with comparable areas between the 2 cities in clusters 4 and 8. For example if someone was coming from a New York zipcode in cluster 4, we can clearly recommend any one of the 15 Boston zip codes to those people.

This algorithm has potential for improvement. We could have added more clusters for increased differentiation because, as we can see, cluster 1 are highly overcrowded and it is very hard to be accurate with such a high number of zip codes in one cluster.

Conclusion:

Overall, the above algorithm can be effectively used by the realtor to recommend neighborhoods to customers.