Coursera Capstone

IBM Applied Data Science

Expanding a High-End Irish Pub to New York City

By: Will Pratt

May 2020

1. Introduction/Business Problem

The city of Boston is well known for many of its citizens having Irish roots. With those roots comes many Irish pubs. The basis for this study is to help a group of investors in Boston expand their successful Pub chain to the New York City area. New York City offers a vast and diverse market that ranks among the most populous cities in the world. Due to the complexity and diversity in the city, the group of investors has asked for assistance in selecting the correct area to expand their restaurant to.

They believe that an area must meet certain criteria for them to consider it:

- It must be considerably wealthy, because of the nature of the high-end restaurant.
- It must have a dense population of people that live near it
- It must not have too many similar restaurants already around their new location.

In the past these criteria have been great indicators on whether the has restaurant fared well or not. The investors believe that leveraging both population data as well as location data, will provide a strong base for a profitable expansion.

2. Data

To solve the problem the following data sources will be used:

- New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

  Source: *https://cocl.us/new_york_dataset*

- GeoSpace Data

  Source*: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm*

  This data will help to get the boundaries for the Boroughs and neighborhoods for visualization

- Venue data

  Source: Foursquare API

  This api will help us to get all the venues from each neighborhood

- Census data

  Source: https://www.kaggle.com/muonneutrino/new-york-city-census-data

This dataset contains demographic and economic data for NYC

- NYC population data

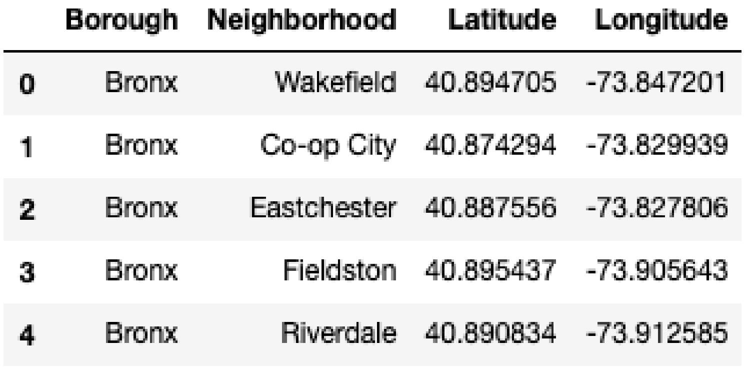
  Source: https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoodpop.htm

  This data contains regions of New York and their subsequent populations

By leveraging these 5 datasets, we will be able to adequately answer all of the investor's criteria marks for each neighborhood and cluster each based on their desirability for expanding the restaurant.

3. Methodology

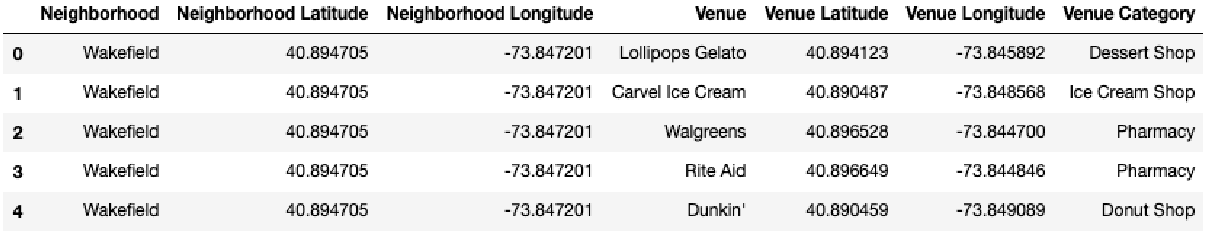
The first part of the project consisted of uploading the initial New York City dataset that contained each Neighborhood and subsequent Borough paired with its longitude and latitude. After initial NYC dataset was uploaded and converted into a pandas data frame, it appeared as figure 1 below. Pandas was used for exploratory data analysis to choose how to segment the neighborhoods, based off demographics and venues in the vicinity of each.



	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 1, NYC Neighborhood Locations

After this was accomplished it was paired with Foursquare API to get venues from each neighborhood. The top 100 venues from each neighborhood were selected and combined into the initial data frame as shown below in figure 2.

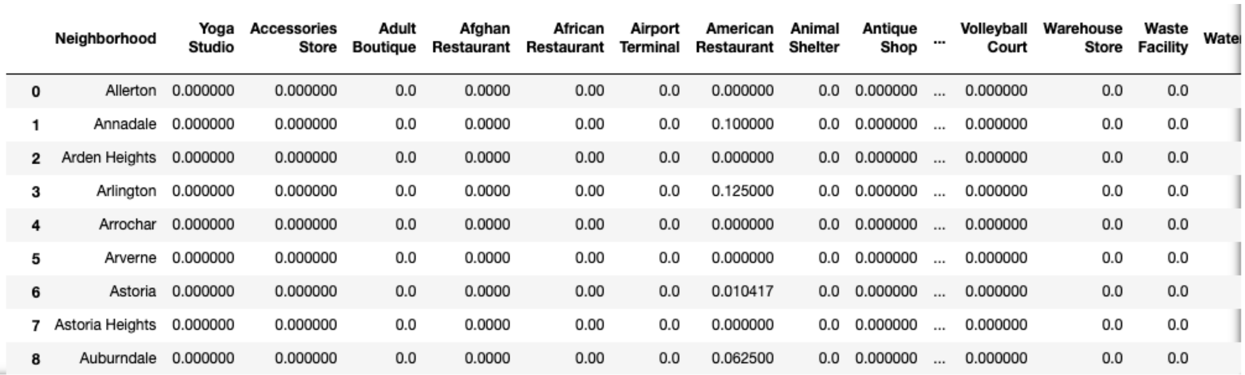


	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Figure 2, Neighborhood Venue Data frame

4. Exploratory Data Analysis and Machine Learning

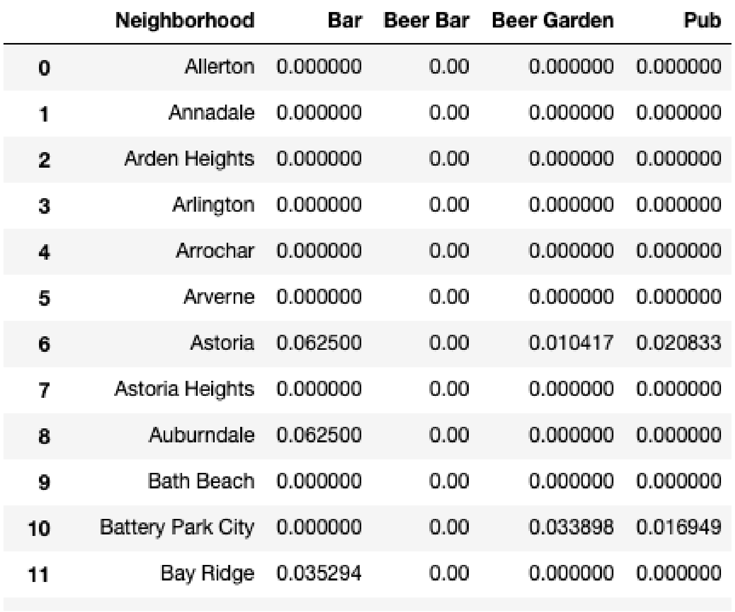
Once the data was properly collected, it was time for the exploratory data analysis to take place. The first step was one-hot encoding to produce a table based on "Venue Category". Everything was grouped based off "Neighborhood" following this. Subsequently, a table was produced like the one shown below.



	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Animal Shelter	Antique Shop	...	Volleyball Court	Warehouse Store	Waste Facility	Wate
0	Allerton	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	
1	Annadale	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.100000	0.0	0.000000	...	0.000000	0.0	0.0	
2	Arden Heights	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	
3	Arlington	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.125000	0.0	0.000000	...	0.000000	0.0	0.0	
4	Arrochar	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	
5	Arverne	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	
6	Astoria	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.010417	0.0	0.000000	...	0.000000	0.0	0.0	
7	Astoria Heights	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.000000	0.0	0.000000	...	0.000000	0.0	0.0	
8	Auburndale	0.000000	0.000000	0.0	0.0000	0.00	0.0	0.062500	0.0	0.000000	...	0.000000	0.0	0.0	

Figure 3, New York One Hot Grouped Data frame

Next, similar venues to Irish Pubs were identified. The 4 places that most closely resembled Irish Pubs were, "Bar", "Beer Bar", "Beer Garden" and "Pub". Based on this, a new data frame was created that contained the frequency of each restaurant in each neighborhood.



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub
0	Allerton	0.000000	0.00	0.000000	0.000000
1	Annadale	0.000000	0.00	0.000000	0.000000
2	Arden Heights	0.000000	0.00	0.000000	0.000000
3	Arlington	0.000000	0.00	0.000000	0.000000
4	Arrochar	0.000000	0.00	0.000000	0.000000
5	Arverne	0.000000	0.00	0.000000	0.000000
6	Astoria	0.062500	0.00	0.010417	0.020833
7	Astoria Heights	0.000000	0.00	0.000000	0.000000
8	Auburndale	0.062500	0.00	0.000000	0.000000
9	Bath Beach	0.000000	0.00	0.000000	0.000000
10	Battery Park City	0.000000	0.00	0.033898	0.016949
11	Bay Ridge	0.035294	0.00	0.000000	0.000000

Figure 4, Frequency Distribution in Neighborhoods

A machine learning algorithm could then be applied to cluster the neighborhoods based off this table. A K-means algorithm was run to sort the neighborhoods into 10 clusters. A map of the clusters is shown below.

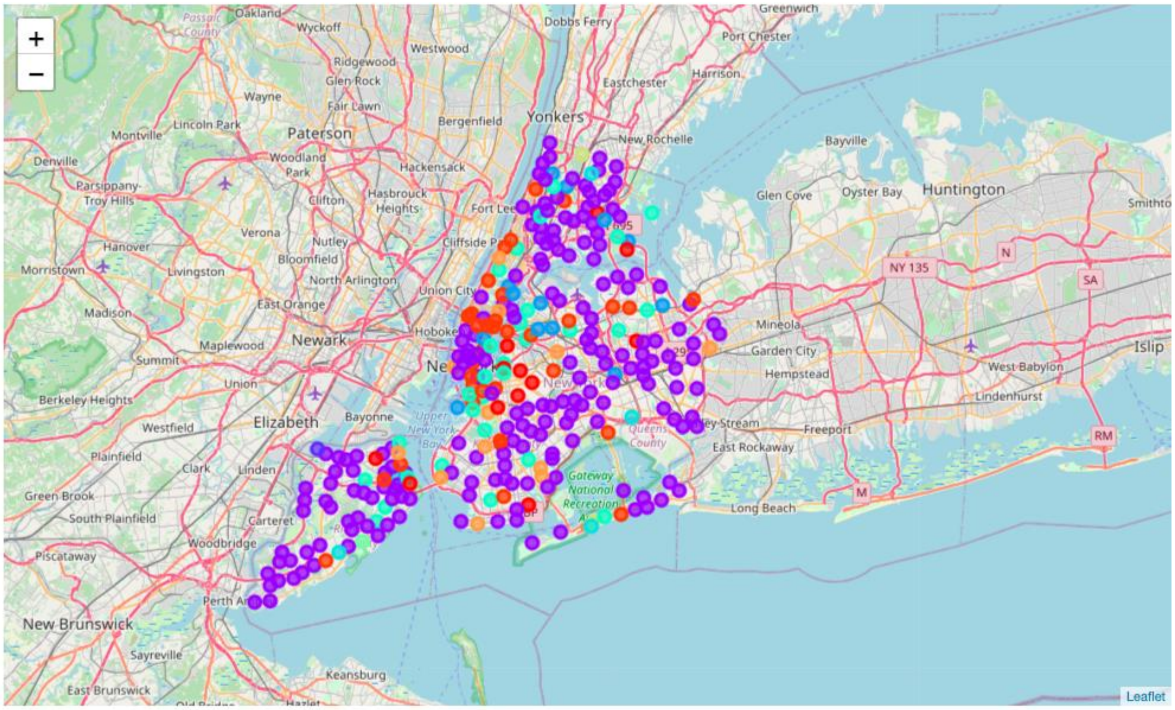
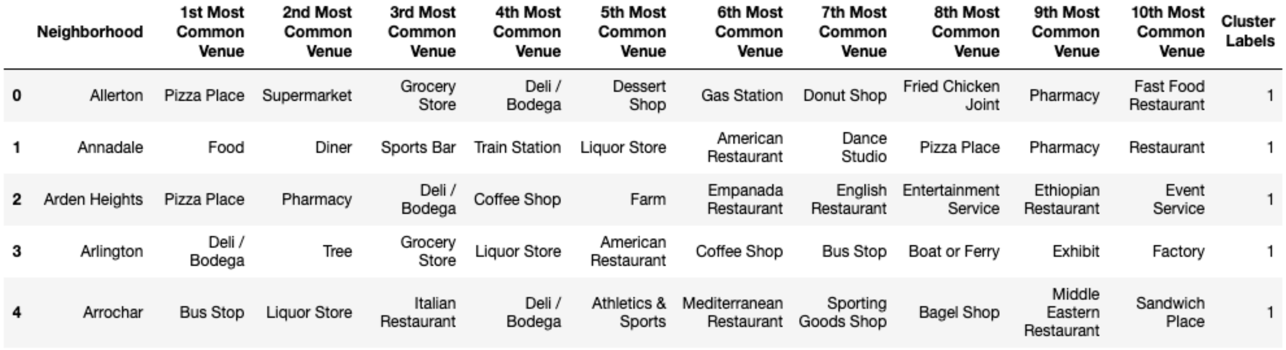


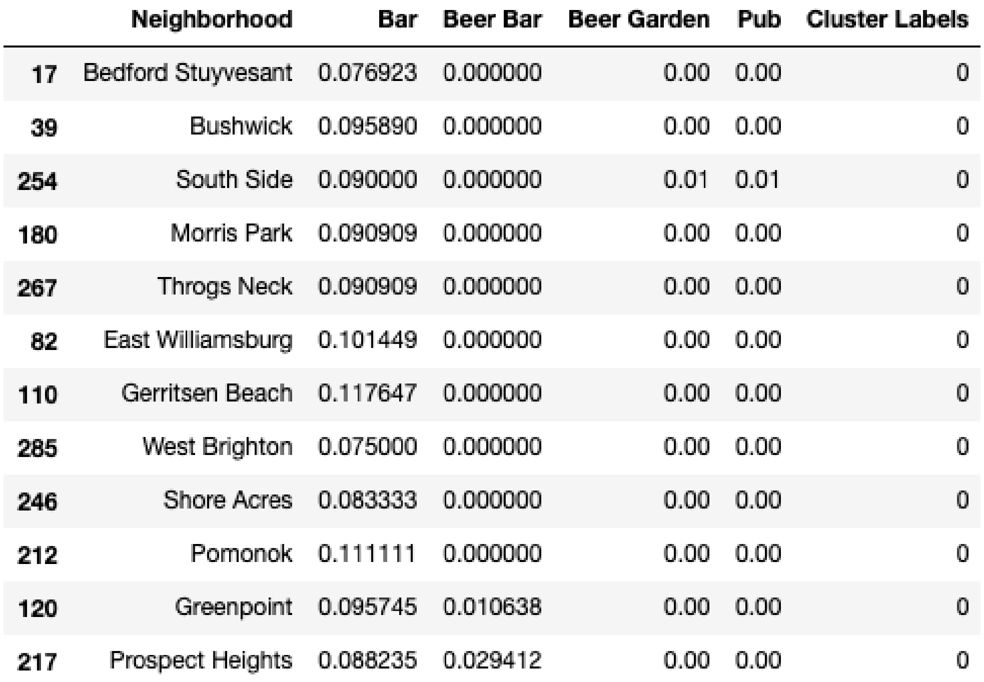
Figure 5, Cluster Mapping in NYC

Exploratory data analysis was then used to decide which clusters to eliminate. A table was combined with most popular venues and their labeled cluster. An additional table was also created that displayed frequencies of each venue as well as the cluster label again. Clusters that had above average frequencies of the 4 similar venues were eliminated.



	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	Cluster Labels
0	Allerton	Pizza Place	Supermarket	Grocery Store	Deli / Bodega	Dessert Shop	Gas Station	Donut Shop	Fried Chicken Joint	Pharmacy	Fast Food Restaurant	1
1	Annadale	Food	Diner	Sports Bar	Train Station	Liquor Store	American Restaurant	Dance Studio	Pizza Place	Pharmacy	Restaurant	1
2	Arden Heights	Pizza Place	Pharmacy	Deli / Bodega	Coffee Shop	Farm	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Service	1
3	Arlington	Deli / Bodega	Tree	Grocery Store	Liquor Store	American Restaurant	Coffee Shop	Bus Stop	Boat or Ferry	Exhibit	Factory	1
4	Arrochar	Bus Stop	Liquor Store	Italian Restaurant	Deli / Bodega	Athletics & Sports	Mediterranean Restaurant	Sporting Goods Shop	Bagel Shop	Middle Eastern Restaurant	Sandwich Place	1

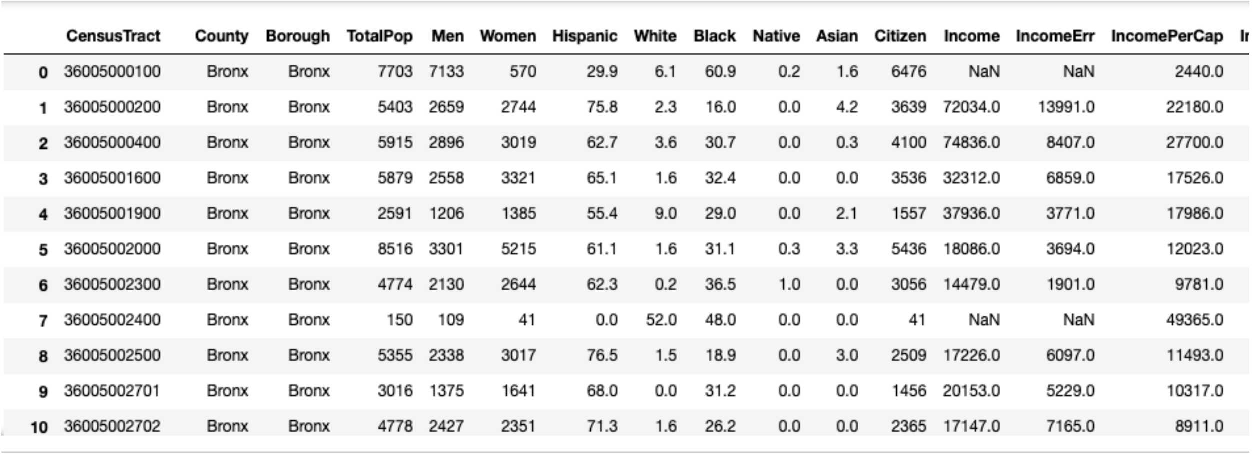
Figure 6, Cluster Label with Common Venue



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
17	Bedford Stuyvesant	0.076923	0.000000	0.00	0.00	0
39	Bushwick	0.095890	0.000000	0.00	0.00	0
254	South Side	0.090000	0.000000	0.01	0.01	0
180	Morris Park	0.090909	0.000000	0.00	0.00	0
267	Throgs Neck	0.090909	0.000000	0.00	0.00	0
82	East Williamsburg	0.101449	0.000000	0.00	0.00	0
110	Gerritsen Beach	0.117647	0.000000	0.00	0.00	0
285	West Brighton	0.075000	0.000000	0.00	0.00	0
246	Shore Acres	0.083333	0.000000	0.00	0.00	0
212	Pomonok	0.111111	0.000000	0.00	0.00	0
120	Greenpoint	0.095745	0.010638	0.00	0.00	0
217	Prospect Heights	0.088235	0.029412	0.00	0.00	0

Figure 7, Cluster Label with Frequency

The following step was uploading the NYC census data to plot out the highest income places in the city. Since the dataset was limited to Boroughs, each Borough data point was averaged for “IncomePerCap” and plotted.



	CensusTract	County	Borough	TotalPop	Men	Women	Hispanic	White	Black	Native	Asian	Citizen	Income	IncomeErr	IncomePerCap	Ir
0	36005000100	Bronx	Bronx	7703	7133	570	29.9	6.1	60.9	0.2	1.6	6476	NaN	NaN	2440.0	
1	36005000200	Bronx	Bronx	5403	2659	2744	75.8	2.3	16.0	0.0	4.2	3639	72034.0	13991.0	22180.0	
2	36005000400	Bronx	Bronx	5915	2896	3019	62.7	3.6	30.7	0.0	0.3	4100	74836.0	8407.0	27700.0	
3	36005001600	Bronx	Bronx	5879	2558	3321	65.1	1.6	32.4	0.0	0.0	3536	32312.0	6859.0	17526.0	
4	36005001900	Bronx	Bronx	2591	1206	1385	55.4	9.0	29.0	0.0	2.1	1557	37936.0	3771.0	17986.0	
5	36005002000	Bronx	Bronx	8516	3301	5215	61.1	1.6	31.1	0.3	3.3	5436	18086.0	3694.0	12023.0	
6	36005002300	Bronx	Bronx	4774	2130	2644	62.3	0.2	36.5	1.0	0.0	3056	14479.0	1901.0	9781.0	
7	36005002400	Bronx	Bronx	150	109	41	0.0	52.0	48.0	0.0	0.0	41	NaN	NaN	49365.0	
8	36005002500	Bronx	Bronx	5355	2338	3017	76.5	1.5	18.9	0.0	3.0	2509	17226.0	6097.0	11493.0	
9	36005002701	Bronx	Bronx	3016	1375	1641	68.0	0.0	31.2	0.0	0.0	1456	20153.0	5229.0	10317.0	
10	36005002702	Bronx	Bronx	4778	2427	2351	71.3	1.6	26.2	0.0	0.0	2365	17147.0	7165.0	8911.0	

Figure 8, NYC Census Data

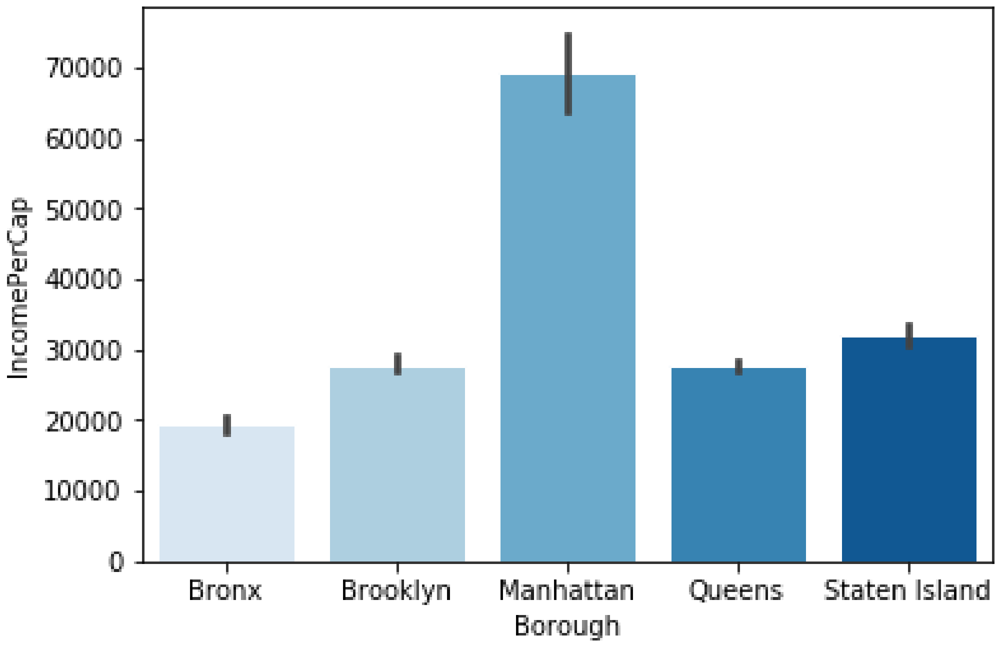
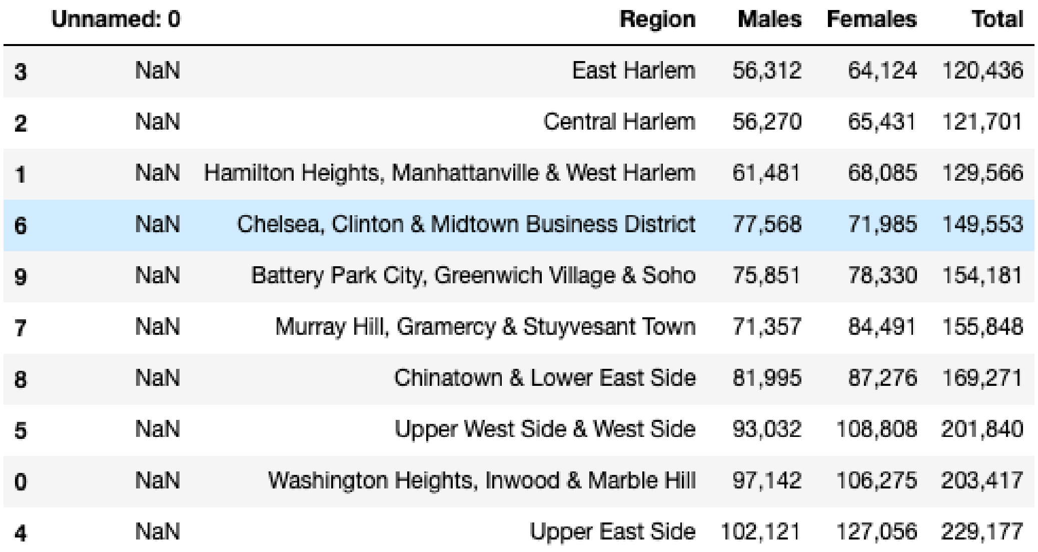


Figure 9, NYC Income by Borough

The wealthiest Borough was selected, and the others were eliminated, too narrow the optimal area further. Finally, the total population of each Manhattan region was loaded and put in descending order to compare.

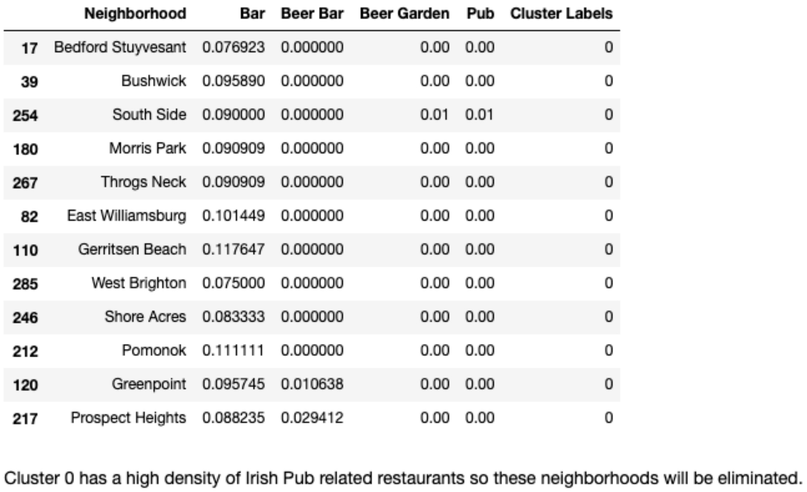


	Unnamed: 0	Region	Males	Females	Total
3	NaN	East Harlem	56,312	64,124	120,436
2	NaN	Central Harlem	56,270	65,431	121,701
1	NaN	Hamilton Heights, Manhattanville & West Harlem	61,481	68,085	129,566
6	NaN	Chelsea, Clinton & Midtown Business District	77,568	71,985	149,553
9	NaN	Battery Park City, Greenwich Village & Soho	75,851	78,330	154,181
7	NaN	Murray Hill, Gramercy & Stuyvesant Town	71,357	84,491	155,848
8	NaN	Chinatown & Lower East Side	81,995	87,276	169,271
5	NaN	Upper West Side & West Side	93,032	108,808	201,840
0	NaN	Washington Heights, Inwood & Marble Hill	97,142	106,275	203,417
4	NaN	Upper East Side	102,121	127,056	229,177

Figure 10, NYC Population by Region in Manhattan

Results

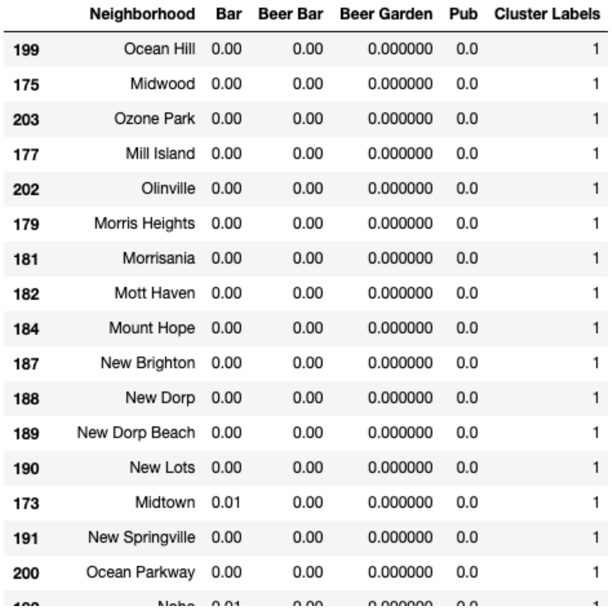
As discussed in the methodology section, through each criterion completed, more neighborhoods were removed. Based off the frequency table that displayed the density of "Bar", "Beer Bar", "Beer Garden" and "Pubs" the 10 clusters were selected through k-means clustering. The resulting frequency tables with reasoning are shown below.



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
17	Bedford Stuyvesant	0.076923	0.000000	0.00	0.00	0
39	Bushwick	0.095890	0.000000	0.00	0.00	0
254	South Side	0.090000	0.000000	0.01	0.01	0
180	Morris Park	0.090909	0.000000	0.00	0.00	0
267	Throgs Neck	0.090909	0.000000	0.00	0.00	0
82	East Williamsburg	0.101449	0.000000	0.00	0.00	0
110	Gerritsen Beach	0.117647	0.000000	0.00	0.00	0
285	West Brighton	0.075000	0.000000	0.00	0.00	0
246	Shore Acres	0.083333	0.000000	0.00	0.00	0
212	Pomonok	0.111111	0.000000	0.00	0.00	0
120	Greenpoint	0.095745	0.010638	0.00	0.00	0
217	Prospect Heights	0.088235	0.029412	0.00	0.00	0

Cluster 0 has a high density of Irish Pub related restaurants so these neighborhoods will be eliminated.

Figure 11, Cluster 0 Results

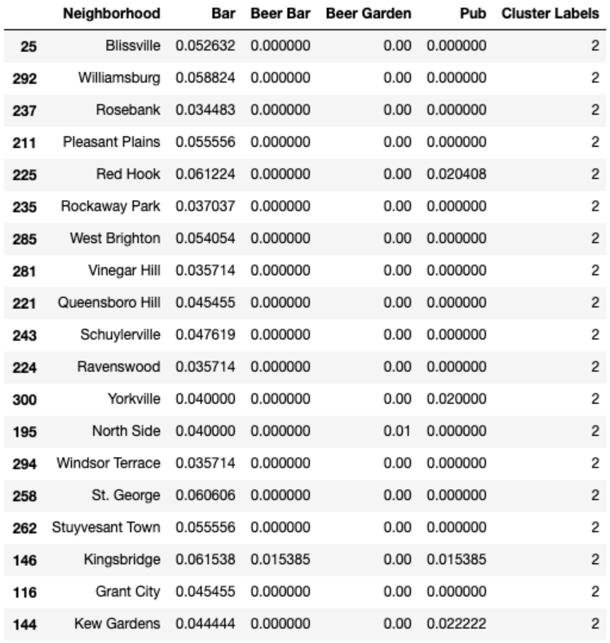


	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
199	Ocean Hill	0.00	0.00	0.000000	0.0	1
175	Midwood	0.00	0.00	0.000000	0.0	1
203	Ozone Park	0.00	0.00	0.000000	0.0	1
177	Mill Island	0.00	0.00	0.000000	0.0	1
202	Olinville	0.00	0.00	0.000000	0.0	1
179	Morris Heights	0.00	0.00	0.000000	0.0	1
181	Morrisania	0.00	0.00	0.000000	0.0	1
182	Mott Haven	0.00	0.00	0.000000	0.0	1
184	Mount Hope	0.00	0.00	0.000000	0.0	1
187	New Brighton	0.00	0.00	0.000000	0.0	1
188	New Dorp	0.00	0.00	0.000000	0.0	1
189	New Dorp Beach	0.00	0.00	0.000000	0.0	1
190	New Lots	0.00	0.00	0.000000	0.0	1
173	Midtown	0.01	0.00	0.000000	0.0	1
191	New Springville	0.00	0.00	0.000000	0.0	1
200	Ocean Parkway	0.00	0.00	0.000000	0.0	1
[illegible]	Noho	0.01	0.00	0.000000	0.0	1

Cluster 1 does not have a high frequency of any of the 4 places, so these neighborhoods will stay

Figure 12, Cluster 1 Results

***Note this is only part of the cluster**

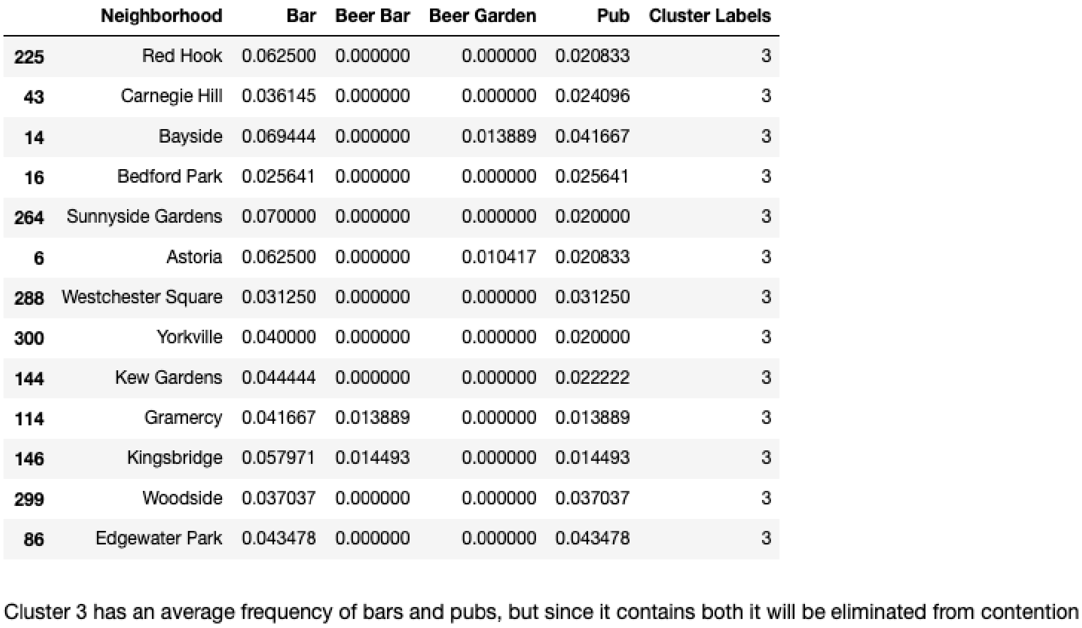


	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
25	Blissville	0.052632	0.000000	0.00	0.000000	2
292	Williamsburg	0.058824	0.000000	0.00	0.000000	2
237	Rosebank	0.034483	0.000000	0.00	0.000000	2
211	Pleasant Plains	0.055556	0.000000	0.00	0.000000	2
225	Red Hook	0.061224	0.000000	0.00	0.020408	2
235	Rockaway Park	0.037037	0.000000	0.00	0.000000	2
285	West Brighton	0.054054	0.000000	0.00	0.000000	2
281	Vinegar Hill	0.035714	0.000000	0.00	0.000000	2
221	Queensboro Hill	0.045455	0.000000	0.00	0.000000	2
243	Schuylerville	0.047619	0.000000	0.00	0.000000	2
224	Ravenswood	0.035714	0.000000	0.00	0.000000	2
300	Yorkville	0.040000	0.000000	0.00	0.020000	2
195	North Side	0.040000	0.000000	0.01	0.000000	2
294	Windsor Terrace	0.035714	0.000000	0.00	0.000000	2
258	St. George	0.060606	0.000000	0.00	0.000000	2
262	Stuyvesant Town	0.055556	0.000000	0.00	0.000000	2
146	Kingsbridge	0.061538	0.015385	0.00	0.015385	2
116	Grant City	0.045455	0.000000	0.00	0.000000	2
144	Kew Gardens	0.044444	0.000000	0.00	0.022222	2

Cluster 2 has a noteworthy number of bars; thus, it will be eliminated.

Figure 13, Cluster 2 Results

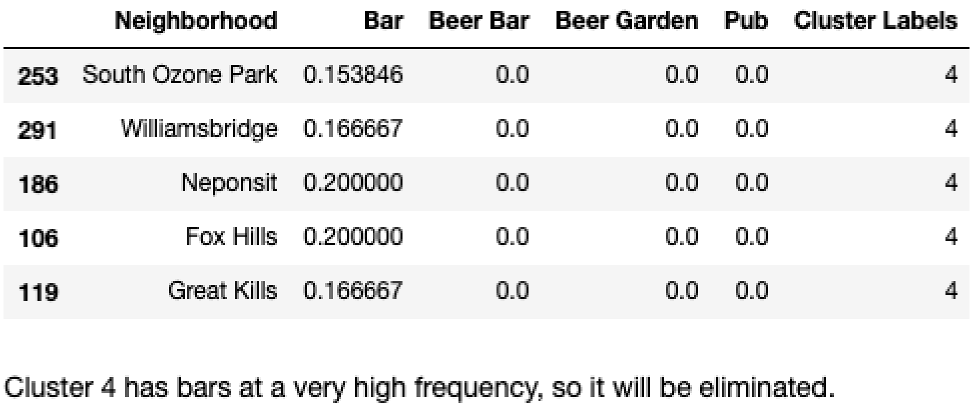
***Note this is only part of the cluster**



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
225	Red Hook	0.062500	0.000000	0.000000	0.020833	3
43	Carnegie Hill	0.036145	0.000000	0.000000	0.024096	3
14	Bayside	0.069444	0.000000	0.013889	0.041667	3
16	Bedford Park	0.025641	0.000000	0.000000	0.025641	3
264	Sunnyside Gardens	0.070000	0.000000	0.000000	0.020000	3
6	Astoria	0.062500	0.000000	0.010417	0.020833	3
288	Westchester Square	0.031250	0.000000	0.000000	0.031250	3
300	Yorkville	0.040000	0.000000	0.000000	0.020000	3
144	Kew Gardens	0.044444	0.000000	0.000000	0.022222	3
114	Gramercy	0.041667	0.013889	0.000000	0.013889	3
146	Kingsbridge	0.057971	0.014493	0.000000	0.014493	3
299	Woodside	0.037037	0.000000	0.000000	0.037037	3
86	Edgewater Park	0.043478	0.000000	0.000000	0.043478	3

Cluster 3 has an average frequency of bars and pubs, but since it contains both it will be eliminated from contention

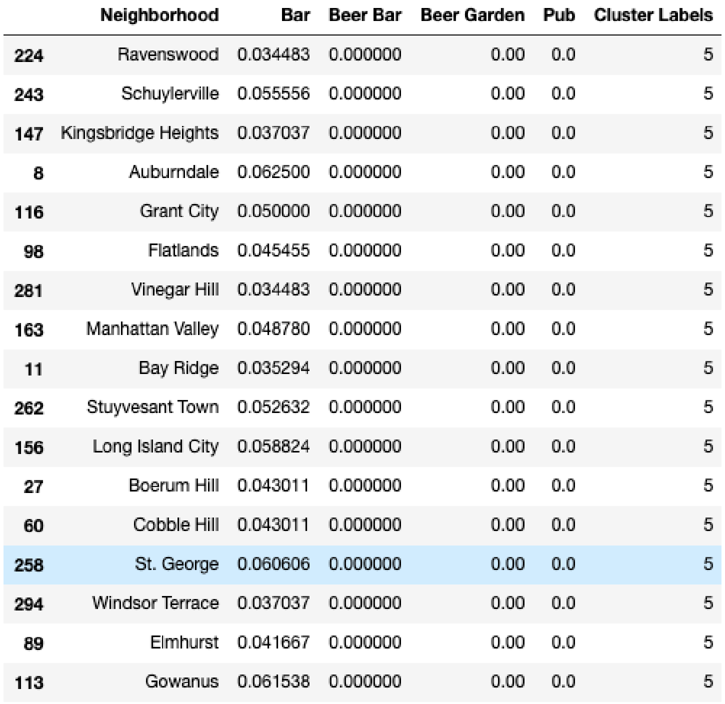
Figure 14, Cluster 3 Results



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
253	South Ozone Park	0.153846	0.0	0.0	0.0	4
291	Williamsbridge	0.166667	0.0	0.0	0.0	4
186	Neponsit	0.200000	0.0	0.0	0.0	4
106	Fox Hills	0.200000	0.0	0.0	0.0	4
119	Great Kills	0.166667	0.0	0.0	0.0	4

Cluster 4 has bars at a very high frequency, so it will be eliminated.

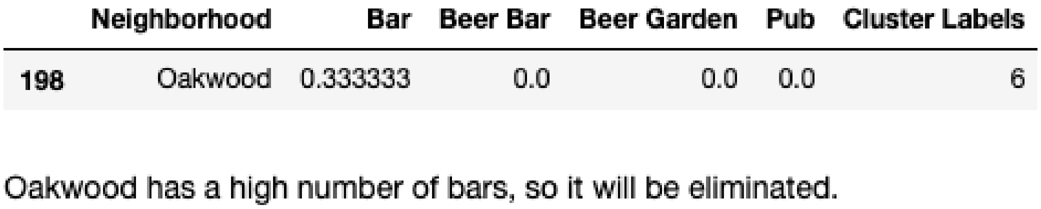
Figure 15, Cluster 4 Results



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
224	Ravenswood	0.034483	0.000000	0.00	0.0	5
243	Schuylerville	0.055556	0.000000	0.00	0.0	5
147	Kingsbridge Heights	0.037037	0.000000	0.00	0.0	5
8	Auburndale	0.062500	0.000000	0.00	0.0	5
116	Grant City	0.050000	0.000000	0.00	0.0	5
98	Flatlands	0.045455	0.000000	0.00	0.0	5
281	Vinegar Hill	0.034483	0.000000	0.00	0.0	5
163	Manhattan Valley	0.048780	0.000000	0.00	0.0	5
11	Bay Ridge	0.035294	0.000000	0.00	0.0	5
262	Stuyvesant Town	0.052632	0.000000	0.00	0.0	5
156	Long Island City	0.058824	0.000000	0.00	0.0	5
27	Boerum Hill	0.043011	0.000000	0.00	0.0	5
60	Cobble Hill	0.043011	0.000000	0.00	0.0	5
258	St. George	0.060606	0.000000	0.00	0.0	5
294	Windsor Terrace	0.037037	0.000000	0.00	0.0	5
89	Elmhurst	0.041667	0.000000	0.00	0.0	5
113	Gowanus	0.061538	0.000000	0.00	0.0	5

Figure 16, Cluster 5 Results

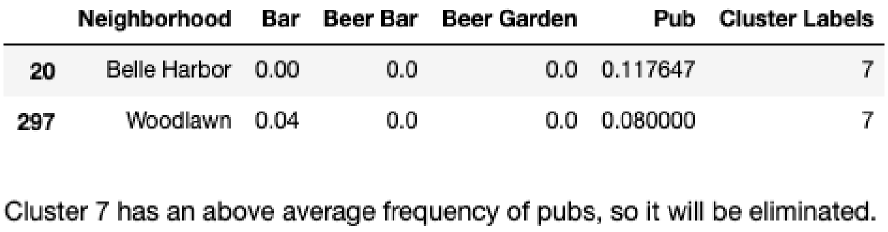
***Note this is only part of the cluster**



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
198	Oakwood	0.333333	0.0	0.0	0.0	6

Oakwood has a high number of bars, so it will be eliminated.

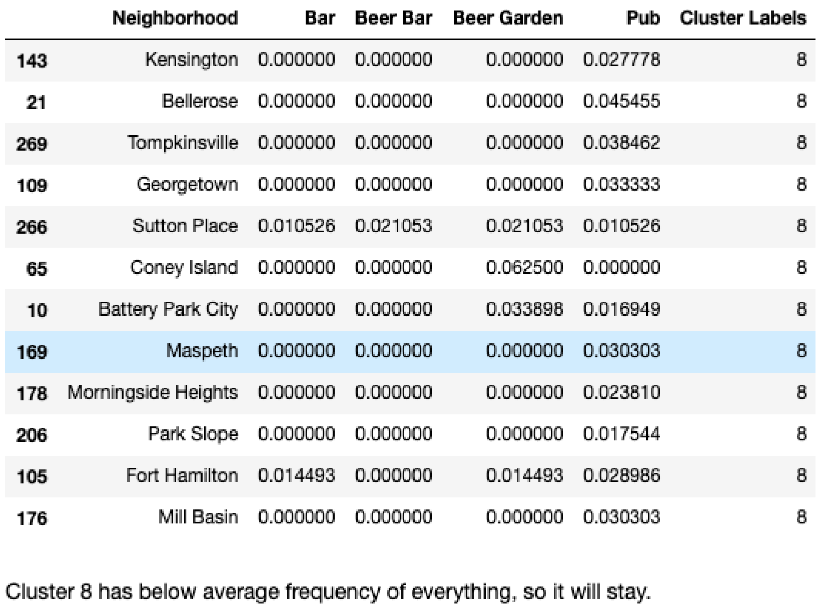
Figure 17, Cluster 6 Results



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
20	Belle Harbor	0.00	0.0	0.0	0.117647	7
297	Woodlawn	0.04	0.0	0.0	0.080000	7

Cluster 7 has an above average frequency of pubs, so it will be eliminated.

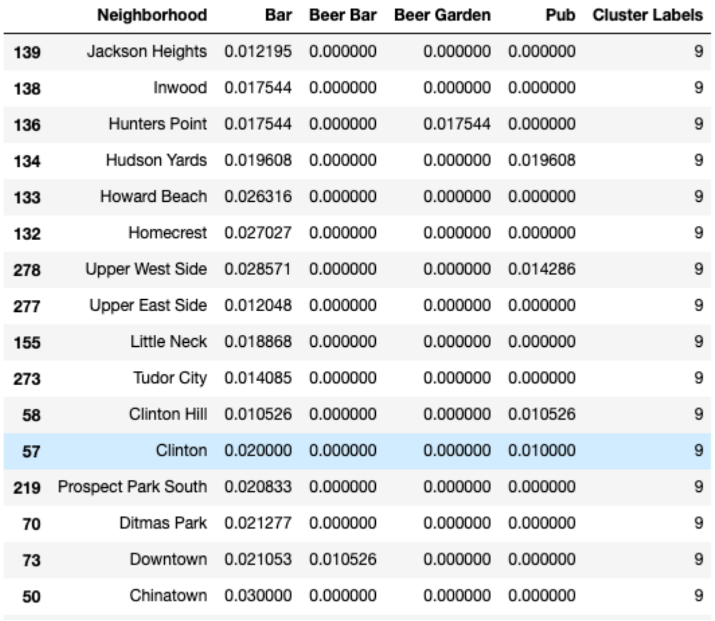
Figure 18, Cluster 7 Results



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
143	Kensington	0.000000	0.000000	0.000000	0.027778	8
21	Bellerose	0.000000	0.000000	0.000000	0.045455	8
269	Tompkinsville	0.000000	0.000000	0.000000	0.038462	8
109	Georgetown	0.000000	0.000000	0.000000	0.033333	8
266	Sutton Place	0.010526	0.021053	0.021053	0.010526	8
65	Coney Island	0.000000	0.000000	0.062500	0.000000	8
10	Battery Park City	0.000000	0.000000	0.033898	0.016949	8
169	Maspeth	0.000000	0.000000	0.000000	0.030303	8
178	Morningside Heights	0.000000	0.000000	0.000000	0.023810	8
206	Park Slope	0.000000	0.000000	0.000000	0.017544	8
105	Fort Hamilton	0.014493	0.000000	0.014493	0.028986	8
176	Mill Basin	0.000000	0.000000	0.000000	0.030303	8

Cluster 8 has below average frequency of everything, so it will stay.

Figure 19, Cluster 8 Results



	Neighborhood	Bar	Beer Bar	Beer Garden	Pub	Cluster Labels
139	Jackson Heights	0.012195	0.000000	0.000000	0.000000	9
138	Inwood	0.017544	0.000000	0.000000	0.000000	9
136	Hunters Point	0.017544	0.000000	0.017544	0.000000	9
134	Hudson Yards	0.019608	0.000000	0.000000	0.019608	9
133	Howard Beach	0.026316	0.000000	0.000000	0.000000	9
132	Homecrest	0.027027	0.000000	0.000000	0.000000	9
278	Upper West Side	0.028571	0.000000	0.000000	0.014286	9
277	Upper East Side	0.012048	0.000000	0.000000	0.000000	9
155	Little Neck	0.018868	0.000000	0.000000	0.000000	9
273	Tudor City	0.014085	0.000000	0.000000	0.000000	9
58	Clinton Hill	0.010526	0.000000	0.000000	0.010526	9
57	Clinton	0.020000	0.000000	0.000000	0.010000	9
219	Prospect Park South	0.020833	0.000000	0.000000	0.000000	9
70	Ditmas Park	0.021277	0.000000	0.000000	0.000000	9
73	Downtown	0.021053	0.010526	0.000000	0.000000	9
50	Chinatown	0.030000	0.000000	0.000000	0.000000	9

Figure 20, Cluster 9 Results

***Note this is only part of the cluster**

Given those results, the dataset was narrowed down to clusters 1, 8 and 9. After this the income per capita was found to be by far the largest in Manhattan based off census data after being explored. All other Boroughs were removed. Lastly the population data was loaded, and the Upper East Side was discovered to have the largest population in the remaining neighborhoods. This was the area that was selected as the optimal place for the investors to expand their high-end Irish Pub to. A map of the marked location is shown below.

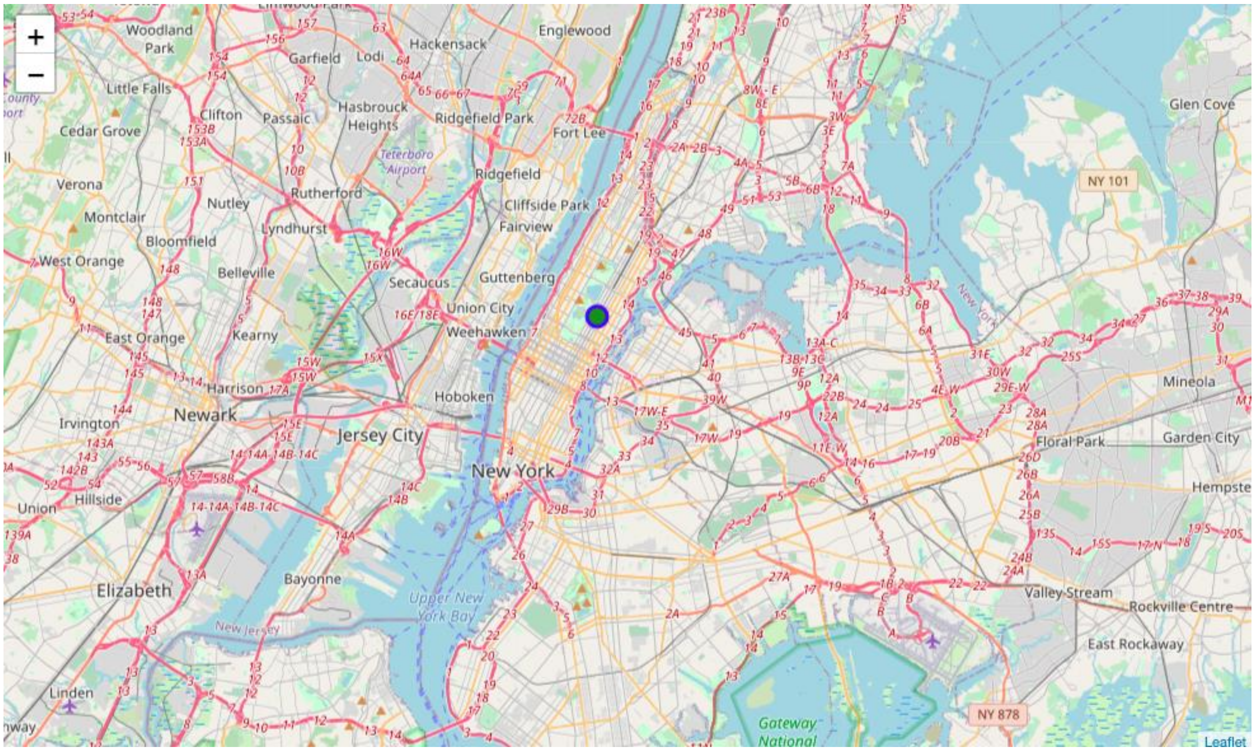


Figure 21, Upper East Side

Discussion

From the results discovered and observed, it can be concluded that many of the neighborhoods in Manhattan would have been sufficient for expanding the Irish Pub. Much of Manhattan had a low density of Pub related restaurants, while simultaneously being wealthy and densely populated. Though, the optimal location taking all three into account is the **Upper East Side**.

It was also clear from clustering that cluster 1 had by far the most neighborhoods, which seemed to be spread pretty randomly throughout New York City.

Conclusion

In this study we attempted to find the optimal spot for a high-end Irish Pub in New York City. The Boston investors selected 3 criteria to explore and use as mechanisms to lower the pool of viable candidates. Although useful for getting a neighborhood that would generally fare well with the expansion, this was surely not an exact science and could have had many things improved. A big factor that could've helped improve was if we had income data by neighborhood instead of borough. In the end, with the data collected, analyzed and clustered, we will stick to the recommendation made for expanding to the Upper East Side.