

PROBLEM STATEMENT

A group of investors wants to open a new highend Irish Pub in New York City and has the following criteria for their optimal location:

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



DATA ACQUISITION AND CLEANING PART I

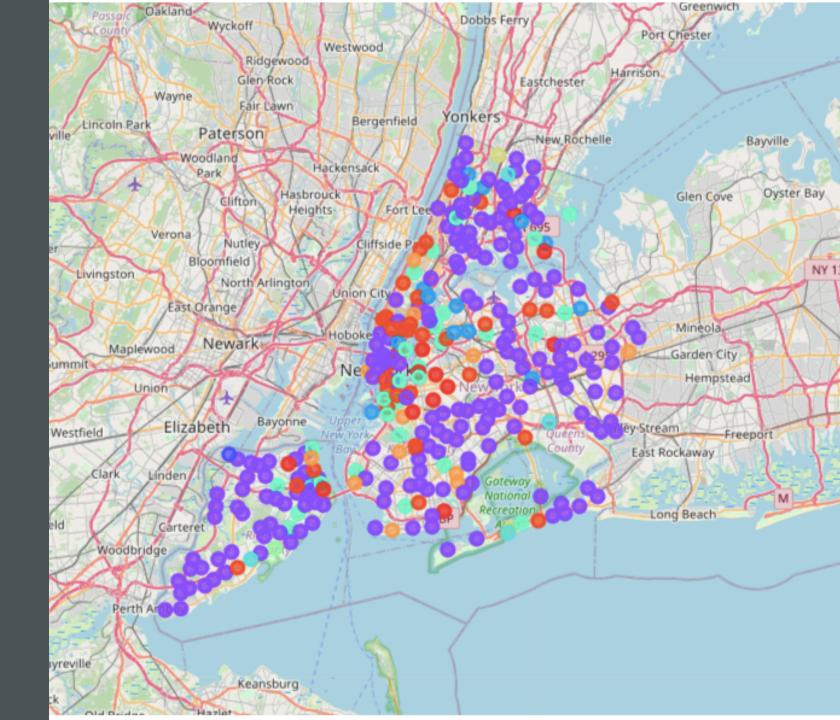
- New York City Borough, Neighborhoods and Longitude/Latitude dataset
- >GeoSpace dataset for boundaries of Boroughs and Neighborhoods
- Foursquare API for acquiring venues
- >NYC Census data from Kaggle
- ➤ NYC Population data from NY.GOV

DATA ACQUISITION AND CLEANING PART 2

- ► In total there are 5 boroughs and 306 neighborhoods in NYC
- Similar venues to pubs were reviewed and "Bar", "Beer Bar", "Beer Garden", and "Pub" were selected
- > A frequency table was created based off each neighborhoods relative number of each kind of venue
- Based of this a k-means clustering algorithm was used to sort each neighborhood

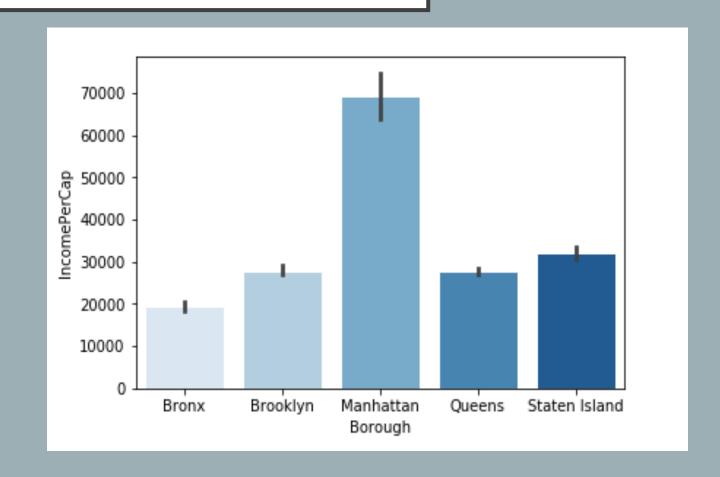
CLUSTERING MAP

- ➤ The clusters appeared to be spread randomly around NYC
- Clusters 1, 2, 5 and 9 contained many more neighborhoods than the others



INCOME DATA ANALYSIS

- Over 2100 rows of data containing demographic and income information from each borough was converted to a data frame and averaged
- The average of each borough's income per capita was plotted, with Manhattan outdoing the others by a large margin.



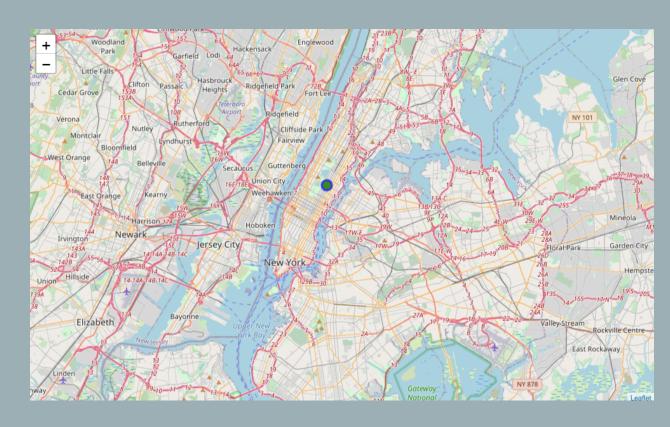
POPULATION DATA ANALYSIS

- The population data for the remaining regions in Manhattan was uploaded and put in descending order
- ➤ Based on the requirement for the largest population, it can be observed that Upper East Side fulfilled that

	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

RESULTS

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**.



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.