Battle of the Neighborhoods

Ethan Kurz

June 28, 2019

Introduction

1.1 Background

A restauranteur is interested in opening up a world class restaurant in New York. Why New York? Because of the cultural diversity and the access it has to global clients. This restauranteur knows how to run and operate a world class restaurant, the final issue to discuss is location.

1.2 Problem

Where can the restauranteur open a world class restaurant with an impact felt immediately? In other words, where will the restauranteur have competition but still have access to the bustle and diversity of New York? I will answer these questions by evaluating neighborhoods all across New York, not diving into one borough, and establishing which ones are most popular for restaurants and what types of restaurants are most successful.

1.3 Interest

These questions are important questions because a restaurant wants to be visible to the world and surrounded by competition they can compete with. Putting a new restaurant in an area known for restaurants will make them immediately popular and boost their reputation quickly. I will further recommend what type of restaurant should be opened in that area.

2. Data

2.1 Data Sources

There are 2 sources that I used to create the datasets for this course. The first was simply Coursera, they supplied us with a JSON file giving us information about the boroughs, neighborhoods, and latitude/longitude points.

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756
12	Bronx	City Island	40.847247	-73.786488
13	Bronx	Bedford Park	40.870185	-73.885512
14	Bronx	University Heights	40.855727	-73.910416
15	Bronx	Morris Heights	40.847898	-73.919672
16	Bronx	Fordham	40.860997	-73.896427
17	Bronx	East Tremont	40.842696	-73.887356

The above data was important for me because it helped me connect with the Foursquare API to create the next dataset.

The second source of my data was Foursquare. I connected to their API and requested information about food venues in each of the neighborhoods in all the boroughs.

```
print(restaurants.shape)
restaurants.head(20)
```

```
(16704, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898276	-73.850381	Caribbean Restaurant
1	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
2	Wakefield	40.894705	-73.847201	SUBWAY	40.890656	-73.849192	Sandwich Place
3	Wakefield	40.894705	-73.847201	Pitman Deli	40.894149	-73.845748	Food
4	Wakefield	40.894705	-73.847201	Baychester Avenue Food Truck	40.892293	-73.843230	Food Truck
5	Wakefield	40.894705	-73.847201	Louis Pizza	40.898457	-73.848770	Pizza Place
6	Co-op City	40.874294	-73.829939	Capri II Pizza	40.876374	-73.829940	Pizza Place
7	Co-op City	40.874294	-73.829939	Arby's	40.870518	-73.828657	Fast Food Restaurant
8	Co-op City	40.874294	-73.829939	Townhouse Restaurant	40.876086	-73.828868	Restaurant
9	Co-op City	40.874294	-73.829939	Quality Market	40.875801	-73.828175	Food
10	Co-op City	40.874294	-73.829939	Guang Hui Chinese Restaurant	40.876603	-73.829710	Chinese Restaurant
11	Co-op City	40.874294	-73.829939	Kennedy Fried Chicken & Pizza	40.876655	-73.829767	Fried Chicken Joint
12	Co-op City	40.874294	-73.829939	Kennedy's	40.876807	-73.829627	Fast Food Restaurant
13	Co-op City	40.874294	-73.829939	Capri Deli	40.870285	-73.828602	Deli / Bodega
14	Co-op City	40.874294	-73.829939	Seven Seas	40.870164	-73.828906	Restaurant
15	Co-op City	40.874294	-73.829939	Bartow Pizza	40.870029	-73.828923	Pizza Place
16	Eastchester	40.887556	-73.827806	Fish & Ting	40.885539	-73.829151	Caribbean Restaurant
17	Eastchester	40.887556	-73.827806	Cozy Cottage Restaurant	40.886332	-73.827616	Diner

This data is important for further clustering analysis and counting types of venues per neighborhood.

2.2 Data Pre-Processing

I used the one-hot encoding tool to pre-process the data I generated in the second dataset. The pre-processed data looks like this:

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery	Belgian Restaurant	Bistro	Brazilian Restaurant	Breakfast Spot	E
0	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Wakefield	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

I then removed everything that was not a restaurant, as we are only interested in restaurants. I then created a count of restaurants in each neighborhood, just so we could see which neighborhoods had the most restaurants. These neighborhoods are of importance to us because these are popular restaurant destinations.

	Neighborhood	Count
0	Murray Hill	93
1	Chinatown	78
2	West Village	74
3	Greenwich Village	73
4	East Village	73
5	Flatiron	70
6	Soho	68
7	Noho	68
8	Little Italy	67
9	South Side	65

I then calculated the frequency of occurrence of each category of restaurants per neighborhood.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cambodian Restaurant	Cantonese Restaurant
0	Allerton	0.000000	0.000000	0.040000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
1	Annadale	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
2	Arden Heights	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
3	Arlington	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
4	Arrochar	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
5	Arverne	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
6	Astoria	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.011905	0.011905	0.0000	0.000000
7	Astoria Heights	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
8	Auburndale	0.000000	0.000000	0.083333	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000
9	Bath Beach	0.000000	0.000000	0.000000	0.000000	0.000000	0.047619	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.047619
10	Battery Park City	0.000000	0.000000	0.027778	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.0000	0.000000

I printed each neighborhood along with the top 5 most common venues from this data above.

```
----Murray Hill----
                  venue  freq
0    Korean Restaurant  0.19
1  Japanese Restaurant  0.06
2  American Restaurant  0.04
3   Italian Restaurant  0.03
4     Sushi Restaurant  0.03
```

We can see that in the neighborhood with the most restaurants, the most common restaurant is a Korean one. In table form this data looks like:

	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	8
0	Allerton	Chinese Restaurant	American Restaurant	Fast Food Restaurant	Mexican Restaurant	Spanish Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	
1	Annadale	American Restaurant	Restaurant	Sushi Restaurant	Vietnamese Restaurant	Halal Restaurant	Ethiopian Restaurant	Falafel Restaurant	
2	Arden Heights	Vietnamese Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	
3	Arlington	American Restaurant	Caribbean Restaurant	Vietnamese Restaurant	Himalayan Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	
4	Arrochar	Italian Restaurant	Mediterranean Restaurant	Polish Restaurant	Middle Eastern Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Falafel Restaurant	

3. Methodology and Analysis

From the data I have collected and pre-processed, I move forwards with the K-Means clustering algorithm. I chose to use 5 clusters.

```
# set number of clusters
kclusters = 5

restaurants_grouped_clustering = restaurants_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(restaurants_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

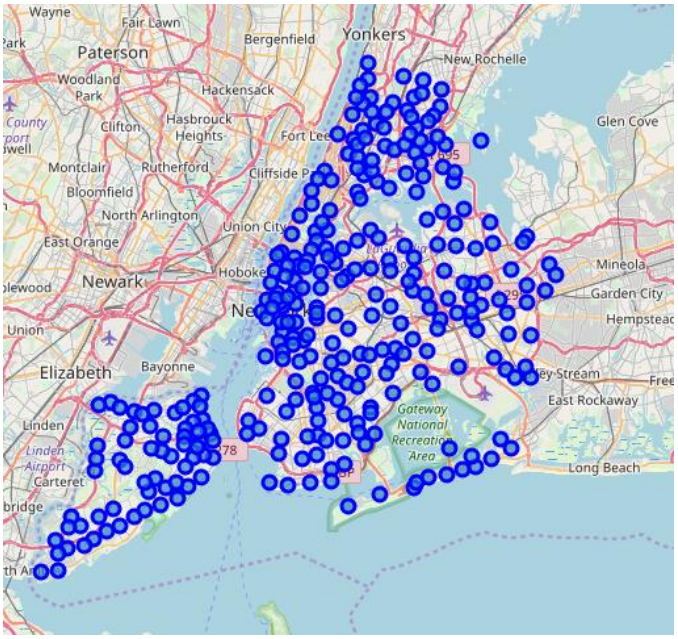
I chose to use K-Means clustering because it chose neighborhoods that were similar to each other in restaurant demographic. This helps me choose which ones are the most popular and what types of restaurants are most popular there.

I spoke briefly before about the exploratory data analysis that I did where I found which neighborhoods had the most restaurants. Not surprisingly, most of the top 10 neighborhoods fall into the same cluster (cluster 0).

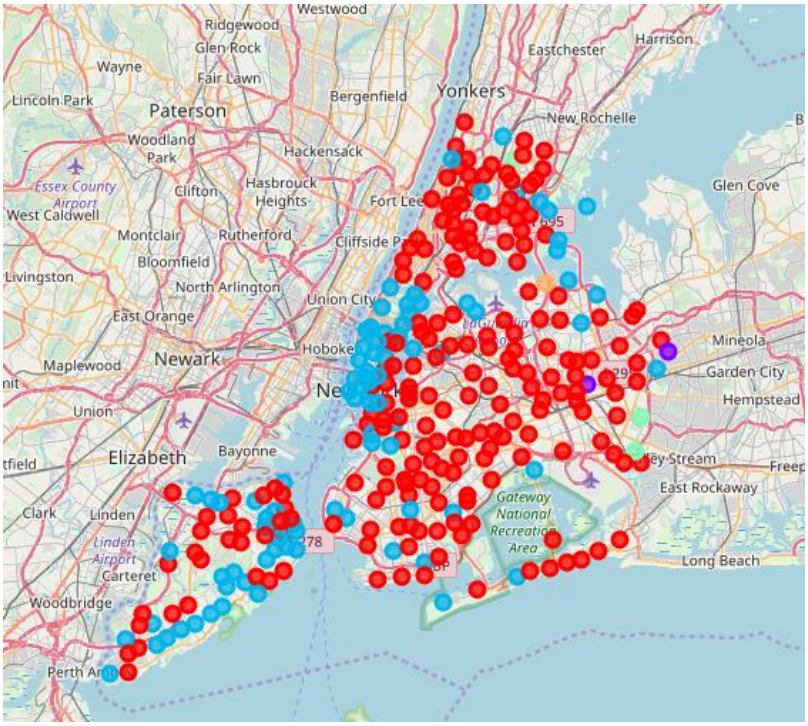
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Wakefield	Caribbean Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	German Restaurant
1	Co-op City	Restaurant	Fast Food Restaurant	Chinese Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Falafel Restaurant	Filipino Restaurant	French Restaurant	German Restaurant
2	Eastchester	Caribbean Restaurant	Seafood Restaurant	Fast Food Restaurant	Chinese Restaurant	Vietnamese Restaurant	English Restaurant	Falafel Restaurant	Filipino Restaurant	French Restaurant
3	Riverdale	Vietnamese Restaurant	Empanada Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	German Restaurant	Gluten-free Restaurant
4	Kingsbridge	Chinese Restaurant	Latin American Restaurant	Mexican Restaurant	Spanish Restaurant	Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Fast Food Restaurant	Caribbean Restaurant
5	Marble Hill	Seafood Restaurant	American Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant
7	Norwood	Chinese Restaurant	American Restaurant	Mexican Restaurant	Fast Food Restaurant	Spanish Restaurant	Restaurant	Caribbean Restaurant	Vietnamese Restaurant	Falafel Restaurant
9	Baychester	American Restaurant	Fast Food Restaurant	Mexican Restaurant	Spanish Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Ethiopian Restaurant	Falafel Restaurant	Filipino Restaurant
12	Bedford Park	Mexican Restaurant	Chinese Restaurant	Spanish Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Halal Restaurant	Ethiopian Restaurant	Falafel Restaurant	Filipino Restaurant
13	University Heights	Chinese Restaurant	Fast Food Restaurant	African Restaurant	Latin American Restaurant	Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Falafel Restaurant	Filipino Restaurant
14	Morris Heights	Spanish Restaurant	Chinese Restaurant	Latin American Restaurant	Vietnamese Restaurant	Halal Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant
15	Fordham	Fast Food Restaurant	Chinese Restaurant	Spanish Restaurant	Mexican Restaurant	Restaurant	American Restaurant	Latin American Restaurant	Seafood Restaurant	Greek Restaurant
16	East Tremont	Chinese Restaurant	Fast Food Restaurant	Latin American Restaurant	Asian Restaurant	Mexican Restaurant	Paella Restaurant	Restaurant	Spanish Restaurant	Vietnamese Restaurant
17	West Farms	Latin American Restaurant	Fast Food Restaurant	Chinese Restaurant	Vietnamese Restaurant	Himalayan Restaurant	Falafel Restaurant	Filipino Restaurant	French Restaurant	German Restaurant
18	High Bridge	Chinese Restaurant	Latin American Restaurant	Seafood Restaurant	Italian Restaurant	Restaurant	Asian Restaurant	Spanish Restaurant	Mexican Restaurant	German Restaurant
19	Melrose	Chinese Restaurant	American Restaurant	Mexican Restaurant	Spanish Restaurant	Vietnamese Restaurant	Hawaiian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant
20	Mott Haven	Spanish Restaurant	Chinese Restaurant	Restaurant	Latin American Restaurant	Fast Food Restaurant	Peruvian Restaurant	Greek Restaurant	Ethiopian Restaurant	Falafel Restaurant
21	Port Morris	Latin American Restaurant	Restaurant	Peruvian Restaurant	Spanish Restaurant	Chinese Restaurant	Greek Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant
22	Longwood	Fast Food Restaurant	Latin American Restaurant	Chinese Restaurant	Vietnamese Restaurant	Himalayan Restaurant	Falafel Restaurant	Filipino Restaurant	French Restaurant	German Restaurant
23	Hunts Point	Spanish Restaurant	Restaurant	Seafood Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant

4. Data Visualization

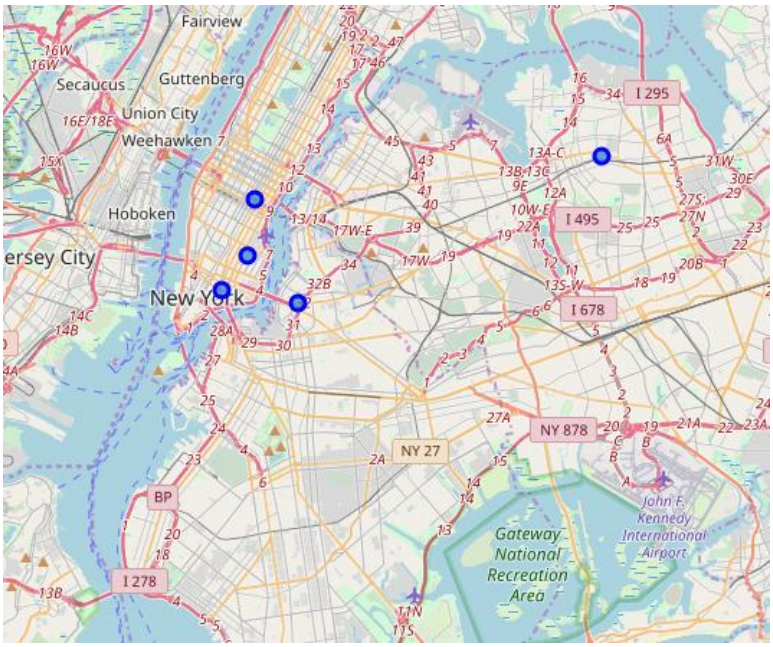
I used folium to visualize data in this project. This first map is just a map of all the neighborhoods in New York:



I then apply the K-Means clustering to it to get this map:



Cross referencing cluster 0 (the red dots) with the top 10 neighborhoods I get this map:



5. Result and Discussion

There are several results from my data analysis. The first is that the neighborhoods with the most restaurants are: Murray Hill, Chinatown, West Village, Greenwich Village, East Village, Flatiron, Soho, Noho, Little Italy, and South Side. Of these 10 locations, 4 are in cluster 0: Murray Hill, Chinatown, East Village, and South Side. These locations already ring a bell to the reader as locations of popular and successful restaurants. The most popular restaurants in these neighborhoods are Chinese Restaurant, Korean Restaurant, Vietnamese Restaurant, Mexican Restaurant, Italian Restaurant, and American Restaurant (in order of most popular to least popular).

6. Conclusion

The conclusion I draw and would explain to a restauranteur is that they should establish a restaurant in Murray Hill. The type of restaurant that would succeed in that neighborhood would be an Asian restaurant, specifically Korean or Japanese. Doing a quick google search about Murray Hill returns that the neighborhood is “a popular home for recent college graduates and young professionals”. Personally, I know this demographic is always looking for a new restaurant and they are not afraid to tell their friends about it. This would be a successful location for any good restauranteur to establish a great restaurant.