Coursera Capstone Project – The Battle of Neighborhoods

By Shuhao Zhang

Oct. 2018

Introduction / Business Problem

Opening a new restaurant can be a really exciting thing. But how do we decide where should we open our restaurant? There could be lots of factors that could potentially affect how the business would go, so choosing the right neighborhood is critical!

Luckily, with the data science tools we have nowadays, we can analyze the neighborhoods and decide which one would be the best for opening a new restaurant. Here are a few factors that we should take into consideration.

Factors that should be taken into consideration

Cuisine Type

Cuisine type is one of the most important things to consider when opening a restaurant. People from different neighborhoods may have very different tastes in food. But at the meantime, the customers who like a specific type of cuisine should share some common interests and characteristics. And these characteristics can be analyzed using the data science tools we have learned in this series of courses.

Demographical Insights of the Neighborhoods

Like the one mentioned above, demographical data can also affect how well the restaurant runs. It is very important to figure out the population of the neighborhoods and what are the preferences that people from different neighborhood have in order to choose the perfect place for a restaurant. By looking at the customers data, we should be able to find out which neighborhood enjoys what kind of food the most and how often do they go to a specific restaurant and so on.

Competing Restaurants

Another factor we have to think is other restaurants. Fortunately, Foursquare can provide us with the data of each neighborhood with the information on their top restaurants and even photos and comments. So we can find out what kind of restaurant will have a potential to be a hit in each neighborhood.

Lease and rent prices in each neighborhood

Money is always important when it comes to business! We have to always keep in mind the expenses and profits. So, analyzing or predicting the profits for potential locations are also very crucial! If we can get the average price for renting spaces for our restaurant, it would really help us to make the predictions and maximize the profits.

Objectives

In this project, due to time and effort limits, I will be mainly focusing on analyzing the most popular restaurants in different neighborhoods in order to provide insights on the most suitable locations for restaurants that serve different types of cuisines.

Data

The majority of the data used in this project will be taken from Foursquare API, which is a crowd sourced, comprehensive geographical data source. With Foursquare API, we will be able to get insight on the most popular venues in each neighborhood, their photos, ratings and customer comments on those venues.

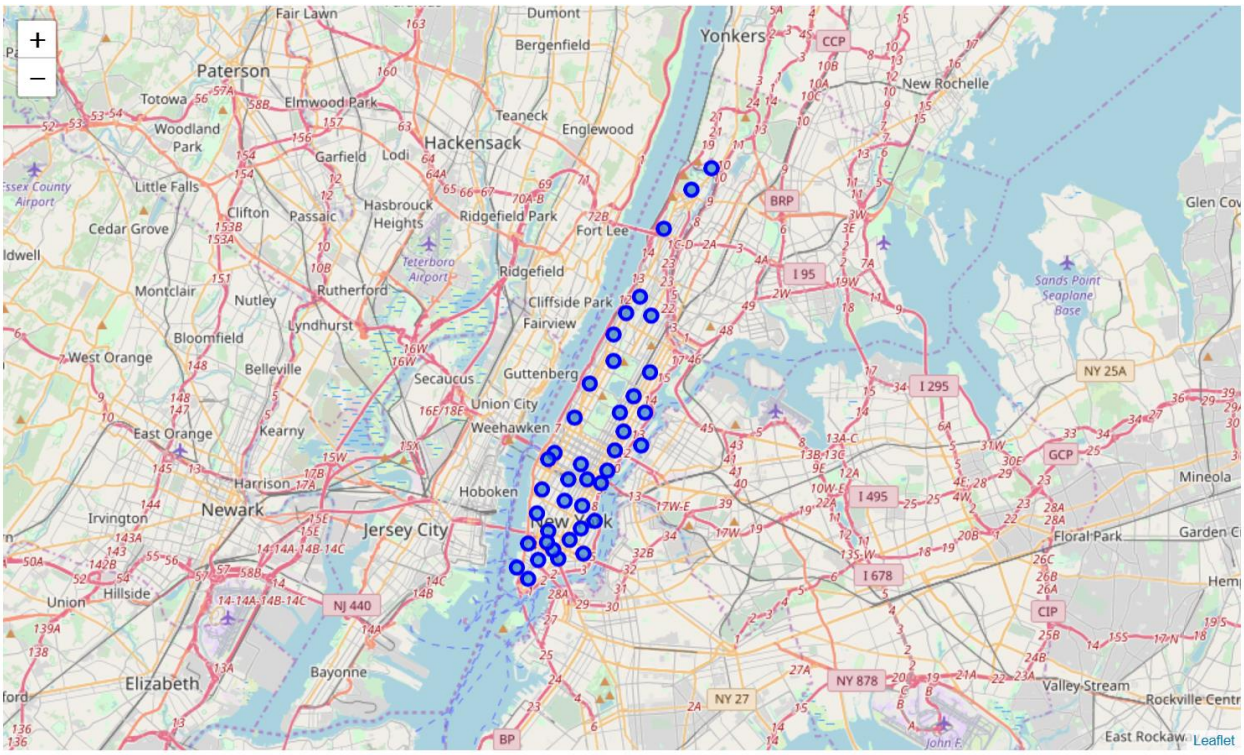
For this project, I will only be analyzing neighborhood in Manhattan because the large amount of restaurants and large diversity in population. I will use the information of these neighborhoods from Foursquare, especially the ratings and rankings for restaurants in different neighborhoods to get knowledge of the preferences of customers in each neighborhood. Their preferences will then help us decide the location of a new restaurant.

Methodology

First, I loaded the geographical data of neighborhoods in all NYC, then separated data for Manhattan from the whole dataset. The Manhattan dataset looks like this:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688
5	Manhattan	Manhattanville	40.816934	-73.957385
6	Manhattan	Central Harlem	40.815976	-73.943211
7	Manhattan	East Harlem	40.792249	-73.944182
8	Manhattan	Upper East Side	40.775639	-73.960508
9	Manhattan	Yorkville	40.775930	-73.947118
10	Manhattan	Lenox Hill	40.768113	-73.958860
11	Manhattan	Roosevelt Island	40.762160	-73.949168
12	Manhattan	Upper West Side	40.787658	-73.977059
13	Manhattan	Lincoln Square	40.773529	-73.985338
14	Manhattan	Clinton	40.759101	-73.996119
15	Manhattan	Midtown	40.754691	-73.981669
16	Manhattan	Murray Hill	40.748303	-73.978332
17	Manhattan	Chelsea	40.744035	-74.003116
18	Manhattan	Greenwich Village	40.726933	-73.999914
19	Manhattan	East Village	40.727847	-73.982226
20	Manhattan	Lower East Side	40.717807	-73.980890
21	Manhattan	Tribeca	40.721522	-74.010683
22	Manhattan	Little Italy	40.719324	-73.997305
23	Manhattan	Soho	40.722184	-74.000657
24	Manhattan	West Village	40.734434	-74.006180
25	Manhattan	Manhattan Valley	40.797307	-73.964286

Then I plotted these neighborhoods on the map of Manhattan.



The next step was to retrieve restaurants' data for each of the neighborhood from FourSquare.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Restaurant Name	Restaurant Latitude	Restaurant Longitude	Restaurant Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
2	Marble Hill	40.876551	-73.91066	Dunkin' Donuts	40.876993	-73.906507	Donut Shop
3	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
4	Marble Hill	40.876551	-73.91066	Boston Market	40.877430	-73.905412	Comfort Food Restaurant

I grouped the data based on neighborhood and calculated the percentage of restaurant for each category.

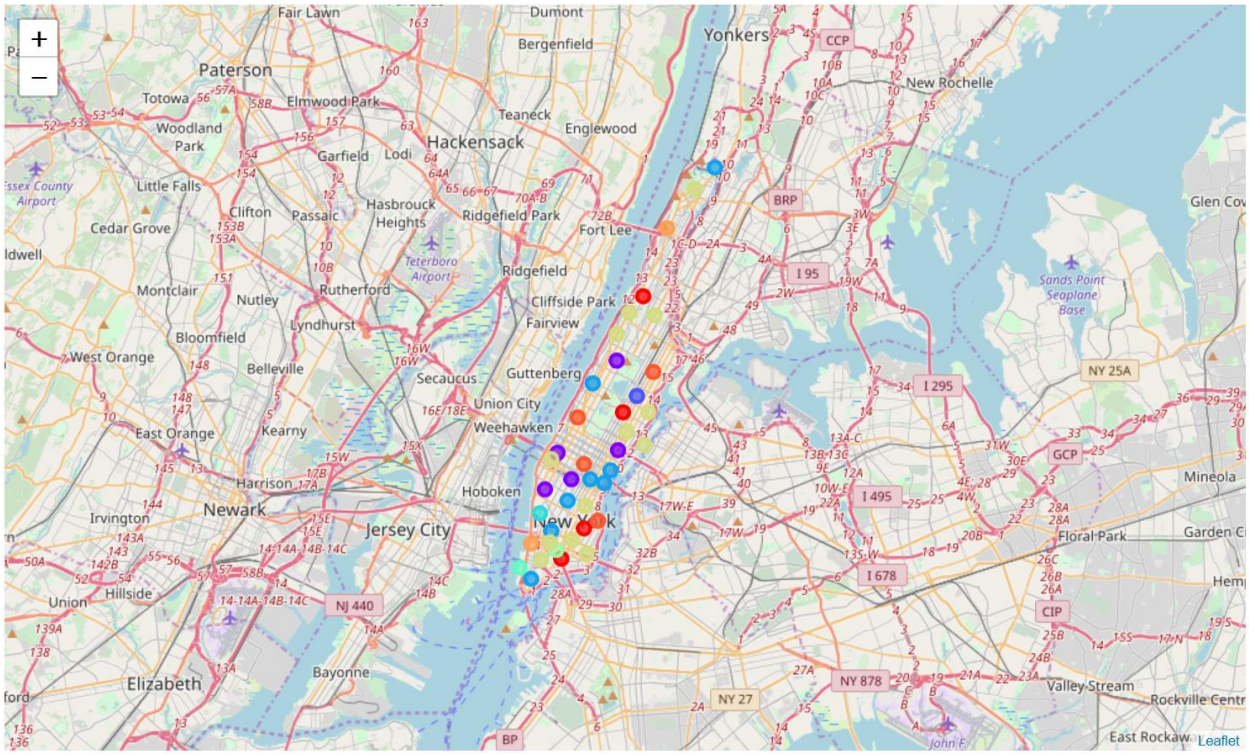
	Neighborhood	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	...	Theme Restaurant	Tibetan Restaurant	Tonka Restau
0	Battery Park City	0.000000	0.060606	0.00	0.000000	0.00	0.00	0.00	0.060606	0.000000	...	0.0	0.0	
1	Carnegie Hill	0.000000	0.014493	0.00	0.014493	0.00	0.00	0.00	0.000000	0.014493	...	0.0	0.0	
2	Central Harlem	0.065217	0.043478	0.00	0.000000	0.00	0.00	0.00	0.043478	0.021739	...	0.0	0.0	
3	Chelsea	0.000000	0.040000	0.01	0.000000	0.03	0.01	0.00	0.010000	0.020000	...	0.0	0.0	
4	Chinatown	0.000000	0.030000	0.00	0.000000	0.03	0.01	0.01	0.000000	0.010000	...	0.0	0.0	

Next, I found out the 10 most common restaurants in each neighborhood based on the frequency they appeared in each neighborhood.

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	Battery Park City	Pizza Place	Italian Restaurant	Burger Joint	American Restaurant	Sandwich Place	BBQ Joint	Bakery	Chinese Restaurant	Burrito Place	Steakhouse
1	Carnegie Hill	Pizza Place	Café	Sushi Restaurant	Bakery	Italian Restaurant	Japanese Restaurant	French Restaurant	Mexican Restaurant	Restaurant	Deli / Bodega
2	Central Harlem	Fried Chicken Joint	Deli / Bodega	Chinese Restaurant	African Restaurant	Pizza Place	Seafood Restaurant	Caribbean Restaurant	Southern / Soul Food Restaurant	Sandwich Place	French Restaurant
3	Chelsea	Bakery	Italian Restaurant	Pizza Place	Tapas Restaurant	Seafood Restaurant	Café	Mexican Restaurant	Japanese Restaurant	American Restaurant	Breakfast Spot
4	Chinatown	Chinese Restaurant	Bakery	Dim Sum Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Mexican Restaurant	Hotpot Restaurant	Café	American Restaurant	Vegetarian / Vegan Restaurant
5	Civic Center	Italian Restaurant	Sandwich Place	French Restaurant	American Restaurant	Mexican Restaurant	Bakery	Sushi Restaurant	Korean Restaurant	Café	Indian Restaurant

Results

Using the frequencies as variables, I grouped all the neighborhoods in Manhattan into 10 clusters and plotted them on the map again. Each different color represents the cluster the neighborhood belongs to.



Here are some sneak peaks for some of the clusters and which neighborhoods are in the cluster.

Cluster 1:

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
1	Chinatown	Chinese Restaurant	Bakery	Dim Sum Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Mexican Restaurant	Hotpot Restaurant	Café	American Restaurant	Vegetarian / Vegan Restaurant
4	Hamilton Heights	Deli / Bodega	Pizza Place	Mexican Restaurant	Café	Donut Shop	Chinese Restaurant	Sandwich Place	Indian Restaurant	Sushi Restaurant	Bakery
8	Upper East Side	Italian Restaurant	American Restaurant	Diner	Pizza Place	French Restaurant	Sushi Restaurant	Deli / Bodega	Bakery	Bagel Shop	Salad Place
19	East Village	Pizza Place	Mexican Restaurant	Japanese Restaurant	Café	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Italian Restaurant	Deli / Bodega	French Restaurant

Cluster 4:

	Neighborhood	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
0	Marble Hill	Sandwich Place	Deli / Bodega	Steakhouse	Diner	Restaurant	Pizza Place	Seafood Restaurant	Comfort Food Restaurant	Chinese Restaurant	Donut Shop
12	Upper West Side	Italian Restaurant	Vegetarian / Vegan Restaurant	Bakery	Indian Restaurant	French Restaurant	Bagel Shop	Burger Joint	Restaurant	Breakfast Spot	Pizza Place
16	Murray Hill	Sandwich Place	American Restaurant	Japanese Restaurant	Pizza Place	Italian Restaurant	Burger Joint	Café	Sushi Restaurant	French Restaurant	Restaurant
18	Greenwich Village	Italian Restaurant	American Restaurant	Café	Sushi Restaurant	Seafood Restaurant	Chinese Restaurant	Indian Restaurant	French Restaurant	Pizza Place	Sandwich Place
29	Financial District	Sandwich Place	Italian Restaurant	Food Truck	American Restaurant	Mexican Restaurant	Steakhouse	Café	Pizza Place	Deli / Bodega	Salad Place
35	Turtle Bay	Italian Restaurant	Deli / Bodega	Café	Steakhouse	Japanese Restaurant	Food Truck	Sushi Restaurant	French Restaurant	Asian Restaurant	Diner
36	Tudor City	Café	Greek Restaurant	Pizza Place	Deli / Bodega	Diner	Mexican Restaurant	Food Truck	Sandwich Place	American Restaurant	Bagel Shop
38	Flatiron	Italian Restaurant	American Restaurant	New American Restaurant	Japanese Restaurant	Sandwich Place	Bakery	Food Truck	Mediterranean Restaurant	Mexican Restaurant	Korean Restaurant

Discussion

By clustering the neighborhoods in Manhattan, we were able to tell the taste differences between different neighborhoods and also able to find out which neighborhood are similar to other ones. This gave us some insights in regard to choosing the optimal location for our new restaurant.

For example, the first cluster we had above, which had Chinatown, Hamilton Heights, Upper East Side, and East Village, showed high interests in Asian cuisines. The fourth cluster preferred Italian food over other cuisines.

Conclusions

With the help of machine learning techniques, we were able to merge different neighborhoods into clusters based on their preferences in restaurants. This also provided insights for the problem we tried to solve in the beginning - how to choose the best place to open a restaurant. By analyzing the clusters, we were able to identify neighborhoods that are better for opening a specific type of restaurant.