The Battle of the Neighborhoods - Report

Introduction

This analysis can help those who are planning to open a new restaurant in Manhattan and need to find the most suitable location.

Background

Selecting the right location for a restaurant is one of the most important problem for running a successful restaurant. Starting a new business in a crowded area such Manhattan, NYC can be challenging. Manhattan has 40 neighborhoods with various venues that attract tourists and people to visit. Neighborhoods of Manhattan are different in terms of different factors that can directly affect the success chance of business. It is important to evaluate different neighborhoods based on the factors that are important for running a successful business such as dining price level, restaurants in the neighborhoods and so on.

Business Problem

In this project, I want to help my client, which are mostly entrepreneurs who want to open restaurant in Manhattan to find the most suitable neighborhood for them to do so. The challenge is to find a "suitable neighborhood". What is the definition of a "suitable neighborhood"? For example, one of the clients wants to open a high end restaurant in Manhattan, it would be wrong to offer this client a location that is surrounded by cheap fast food joints. Therefore, this project aims to solve this problem by analyzing restaurants in different neighborhoods based on their dining price level, popularity and other important features and offer clients a thorough understanding of the dining business layout in Manhattan. Therefore, with the help of this report, clients will understand which neighborhood will fit the type of restaurant the clients want to open.
Audience: Entrepreneurs who want to open restaurant in Manhattan and looking for location advise.

Data

Data 1: For Manhattan neighborhoods data:

Source: https://cocl.us/new_york_dataset

Description: to retrieve neighborhoods' name, coordinates of Manhattan

	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

Data 2: For restaurants related data:

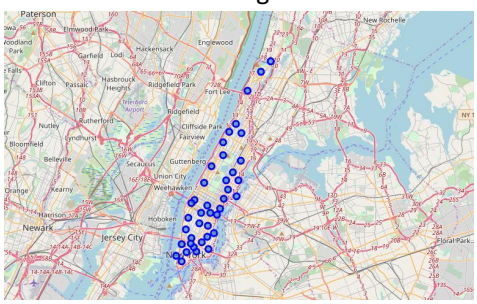
Source: Fousquare API
Two type of Foursquare API calls will be used:
Find popular restaurants:
https://api.foursquare.com/v2/venues/explore

Find detail(price, like counts) for a certain restaurant:
https://api.foursquare.com/v2/venues/venues_id

Description: By using these apis, detail info such as price, popularity of all venues in each neighborhood can be obtained.

3.Methodology :

Business Understanding :
Our main goal is to get optimum location for new restaurant business in New York City for XYZ Company
Analytic Approach :
Manhattan has a total of 40 neighborhoods. This project will use K-means algorithm to cluster each neighborhoods based on mean restaurants price tier, mean restaurants popularity level and mean restaurants category number.
Data 1-Manhattan Neighborhoods Data.



Data 2: popular restaurants at each neighborhood:
The foursquare explore method will be used to retrieve venue tpye id of '4d4b7105d754a06374d81259'(food) for restaurants at each neighborhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue 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'	40.877136	-73.906666	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	Subway Sandwiches	40.874667	-73.909586	Sandwich Place

Data 3: detail of each restaurant, price and likes counts:
for price, 1 being the least expensive, 4 being the most expensive. For food venues, in the United States, 1 is < $10 an entree, 2 is $10-$20 an entree, 3 is $20-$30 an entree, 4 is > $30 an entree.
For likes_counts: the number of thumb up the restaurant received from internet
Both columns are the mean data for all the restaurants data in each neighborhood

Neighborhood	price_tier	likes_count
Central Harlem	2.000000	37.850000
Chelsea	2.277778	513.611111
Chinatown	1.529412	265.823529
Clinton	2.400000	204.133333
East Harlem	1.600000	41.450000

Data 4: most common restaurants 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	Central Harlem	African Restaurant	Chinese Restaurant	Seafood Restaurant	American Restaurant	Fried Chicken Joint	French Restaurant	Pizza Place	Ethiopian Restaurant	Caribbean Restaurant	Café
1	Chelsea	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	French Restaurant	Bakery	New American Restaurant	Chinese Restaurant	Sandwich Place	Middle Eastern Restaurant	Indian Restaurant
2	Chinatown	Chinese Restaurant	Noodle House	Sandwich Place	Pizza Place	Spanish Restaurant	Italian Restaurant	English Restaurant	New American Restaurant	Bakery	Greek Restaurant
3	Clinton	American Restaurant	New American Restaurant	Italian Restaurant	Peruvian Restaurant	Mediterranean Restaurant	Restaurant	Sandwich Place	Seafood Restaurant	Pizza Place	Steakhouse
4	East Harlem	Mexican Restaurant	Latin American Restaurant	Thai Restaurant	Bakery	Pizza Place	Sandwich Place	French Restaurant	Café	Taco Place	Steakhouse

Data 5: processed data that ready to be used for K-means: standard normalization was used to process price_tier and likes counts. One hot encoding was used to process restaurants type.

	Neighborhood	price_tier	likes_count	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Breakfast Spot	Burger Joint	Burrito Place	Cafeteria	Café	Caribbean Restaurant	Chinese Restaurant	Re
0	Central Harlem	0.064272	-0.983757	0.0	0.15	0.100000	0.0	0.000000	0.0	0.05	0.000000	0.0	0.0	0.0	0.05	0.05	0.05	0.100000	
1	Chelsea	0.828416	1.332002	0.0	0.00	0.055556	0.0	0.055556	0.0	0.00	0.055556	0.0	0.0	0.0	0.00	0.00	0.00	0.055556	
2	Chinatown	-1.230278	0.125900	0.0	0.00	0.058824	0.0	0.058824	0.0	0.00	0.058824	0.0	0.0	0.0	0.00	0.00	0.00	0.176471	
3	Clinton	1.164640	-0.174376	0.0	0.00	0.266667	0.0	0.000000	0.0	0.00	0.000000	0.0	0.0	0.0	0.00	0.00	0.00	0.000000	
4	East Harlem	-1.036096	-0.966235	0.0	0.00	0.000000	0.0	0.000000	0.0	0.00	0.100000	0.0	0.0	0.0	0.00	0.05	0.00	0.000000	

Machine Learning Algorithm: K-means

```
# set number of clusters
kclusters = 5

manhattan_grouped_clustering = manhattan_grouped.drop('Neighborhood', 1)

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

4. RESULTS :

K-Means clustering based on mean occurrence of restaurant category, neighborhood dining price level and popularity :

Cluster 1:

	Neighborhood	price_tier	likes_count	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	1.428571	13.857143	Sandwich Place	American Restaurant	Deli / Bodega	Chinese Restaurant	Seafood Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Diner	Donut Shop	Steakhouse
1	Chinatown	1.529412	265.823529	Chinese Restaurant	Noodle House	Sandwich Place	Pizza Place	Spanish Restaurant	Italian Restaurant	English Restaurant	New American Restaurant	Bakery	Greek Restaurant
2	Washington Heights	1.777778	56.166667	Bakery	New American Restaurant	Deli / Bodega	Café	Pizza Place	Latin American Restaurant	Restaurant	Caribbean Restaurant	Breakfast Spot	Burger Joint
3	Inwood	1.578947	25.947368	Restaurant	Mexican Restaurant	Café	Deli / Bodega	American Restaurant	Bakery	Pizza Place	Spanish Restaurant	Diner	Latin American Restaurant
4	Hamilton Heights	1.450000	39.800000	Café	Pizza Place	Caribbean Restaurant	Bakery	Mexican Restaurant	Food Truck	Gastropub	Indian Restaurant	Italian Restaurant	Sushi Restaurant
7	East Harlem	1.600000	41.450000	Mexican Restaurant	Latin American Restaurant	Thai Restaurant	Bakery	Pizza Place	Sandwich Place	French Restaurant	Café	Taco Place	Steakhouse
11	Roosevelt Island	1.400000	8.500000	Deli / Bodega	Noodle House	Pizza Place	Greek Restaurant	Restaurant	Sandwich Place	Café	Chinese Restaurant	Japanese Restaurant	Falafel Restaurant

Cluster 2:

	Neighborhood	price_tier	likes_count	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
8	Upper East Side	2.550000	167.050000	Italian Restaurant	French Restaurant	Pizza Place	American Restaurant	Mexican Restaurant	Latin American Restaurant	Japanese Restaurant	Burrito Place	Salad Place	Sandwich Place
13	Lincoln Square	2.450000	282.800000	Café	American Restaurant	Italian Restaurant	Mediterranean Restaurant	French Restaurant	Food Truck	Bakery	Mexican Restaurant	Burger Joint	Seafood Restaurant
14	Clinton	2.400000	204.133333	American Restaurant	New American Restaurant	Italian Restaurant	Peruvian Restaurant	Mediterranean Restaurant	Restaurant	Sandwich Place	Seafood Restaurant	Pizza Place	Steakhouse
16	Murray Hill	2.526316	234.315789	Japanese Restaurant	Italian Restaurant	Bagel Shop	Burger Joint	Vietnamese Restaurant	Sandwich Place	Hawaiian Restaurant	Vegetarian / Vegan Restaurant	Jewish Restaurant	Mediterranean Restaurant
21	Tribeca	2.578947	332.315789	Greek Restaurant	American Restaurant	Café	Steakhouse	Bakery	Indian Restaurant	Italian Restaurant	Korean Restaurant	New American Restaurant	Salad Place

Cluster 3:

	Neighborhood	price_tier	likes_count	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
23	Soho	2.25	805.0	Italian Restaurant	Mediterranean Restaurant	Bakery	Spanish Restaurant	Mexican Restaurant	Pizza Place	French Restaurant	Vegetarian / Vegan Restaurant	Food Truck	Filipino Restaurant

Cluster 4:

5]:

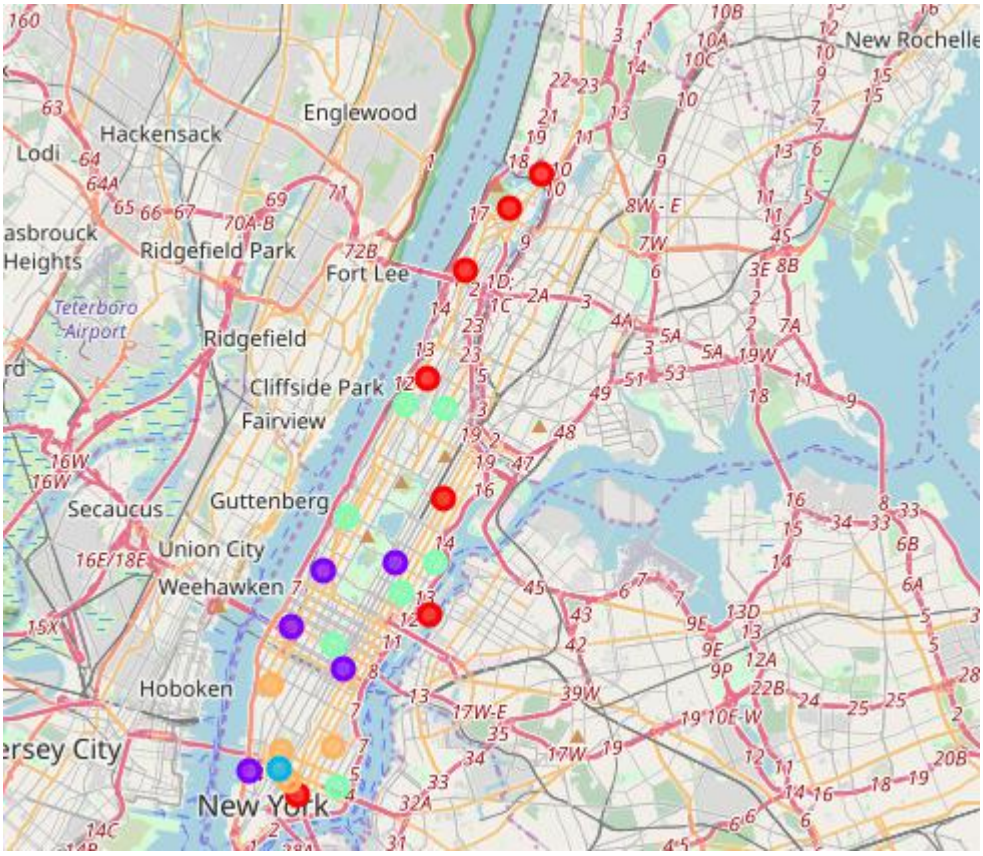
	Neighborhood	price_tier	likes_count	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant	10th Most Common Restaurant
5	Manhattanville	1.842105	152.421053	Italian Restaurant	Mexican Restaurant	Seafood Restaurant	Gastropub	Deli / Bodega	Cuban Restaurant	Dumpling Restaurant	Ramen Restaurant	Café	Falafel Restaurant
6	Central Harlem	2.000000	37.850000	African Restaurant	Chinese Restaurant	Seafood Restaurant	American Restaurant	Fried Chicken Joint	French Restaurant	Pizza Place	Ethiopian Restaurant	Caribbean Restaurant	Café
9	Yorkville	1.900000	87.300000	Italian Restaurant	Deli / Bodega	Vietnamese Restaurant	Bagel Shop	Mexican Restaurant	German Restaurant	Café	Sandwich Place	Burger Joint	Diner
10	Lenox Hill	2.052632	187.526316	Burger Joint	Thai Restaurant	Bakery	Afghan Restaurant	French Restaurant	Italian Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Café
12	Upper West Side	2.000000	317.850000	Seafood Restaurant	American Restaurant	Italian Restaurant	Bakery	Southern / Soul Food Restaurant	Mediterranean Restaurant	Chinese Restaurant	Ramen Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant
15	Midtown	2.166667	143.222222	Sushi Restaurant	Steakhouse	Vietnamese Restaurant	Salad Place	Hawaiian Restaurant	French Restaurant	Food Truck	Mediterranean Restaurant	Deli / Bodega	Cuban Restaurant
20	Lower East Side	1.833333	274.722222	Café	Japanese Restaurant	Vietnamese Restaurant	French Restaurant	Filipino Restaurant	Italian Restaurant	Latin American Restaurant	Mediterranean Restaurant	Mexican Restaurant	Diner

Cluster 5:

7]:

	Neighborhood	price_tier	likes_count	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
17	Chelsea	2.277778	513.611111	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	French Restaurant	Bakery	New American Restaurant	Chinese Restaurant	Sandwich Place	Middle Eastern Restaurant	Indian Restaurant
18	Greenwich Village	2.157895	477.578947	Italian Restaurant	Sushi Restaurant	French Restaurant	Café	New American Restaurant	Food Truck	Bagel Shop	Snack Place	Sandwich Place	English Restaurant
19	East Village	1.888889	588.500000	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	American Restaurant	Burger Joint	Korean Restaurant	Ramen Restaurant	Caribbean Restaurant	Japanese Restaurant	Scandinavian Restaurant
22	Little Italy	1.800000	501.250000	Café	Pizza Place	Sandwich Place	Snack Place	Italian Restaurant	Dumpling Restaurant	Mediterranean Restaurant	Noodle House	Salad Place	Seafood Restaurant

Cluster Map:



5.DISCUSSION:

From the cluster result, following finding can be useful for our audience:

1. Cluster 2 and Cluster 3 are the best neighborhoods for a high end, popular restaurant. Most neighborhoods in these two cluster have high price tier and popularity level. If you are well funded and want to open a fancy restaurant, study these two cluster first.

2. Cluster 1 is good choice for opening up economic restaurants. Most neighborhoods in this cluster have low price tier and are less popular on the internet.

3. Cluster 4 and cluster 5 are good places for opening up restaurants with wide range of price and menu. Neighborhoods in these two cluster have restaurants that targets all levels income customers.

6.CONCLUSION:

This analysis is performed based on the assumption that our clients want to open restaurant where a similar one already exist. This assumption will help clients mitigate the risk of running a restaurant in a unfitted neighborhood. Therefore, the report used restaurant type and mean price tier and mean popularity level of each neighborhood to cluster 40 neighborhoods into 5 cluster. From the result of K-means clustering, our clients can find useful information about the dining business layout in Manhattan and make more accurate decision based on that.