IBM & Coursera

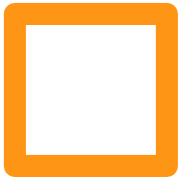
Applied Data Science Capstone

Saksham Kapoor

December 2020

Introduction

There was a time when eating out at a restaurant was considered a luxury only the chosen could afford but thankfully the time has passed and eating out at a restaurant has become a daily habit in the life of New Yorkers. New York has been described as the cultural, financial, media and entertainment capital of the world. New York being the cultural hub is home to many ethical backgrounds, which drives the city's diverse style of restaurant – these include restaurants serving Chinese, Italian, French, Korean Cuisines.... etc. New Restaurants are continuously opening up around the city to meet the high demand and appetite. Although eating out is a popular habit there are certain restaurants that are unfortunately unsuccessful in keep the shop running – perhaps because of poor restaurant location. When opening a new restaurant there are many factors to keep in mind about the location where you set the shop.



Problem

This research project aims to provide a recommendation for the style and location of a new restaurant in New York. There are many factors to consider when creating this recommendation project such as: the number of existing restaurants and their popularity, the popularity of different types of cuisines in New York

Data Utilized

- *For the sake of simplicity, we will only use boroughs with Manhattan in it
- *Table 1: New York Neighborhood data*

	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

Data Utilized

- *Table 2: Manhattan Neighborhood Data*

	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

Methodology

- **Data Cleaning and Exploratory Data Analysis**


```
{'type': 'FeatureCollection',
 'totalFeatures': 306,
 'features': [{'type': 'Feature',
   'id': 'nyu_2451_34572.1',
   'geometry': {'type': 'Point',
    'coordinates': [-73.84720052054902, 40.89470517661]},
   'geometry_name': 'geom',
   'properties': {'name': 'Wakefield',
    'stacked': 1,
    'annoline1': 'Wakefield',
    'annoline2': None,
    'annoline3': None,
    'annoangle': 0.0,
    'borough': 'Bronx',
    'bbox': [-73.84720052054902,
```

Methodology

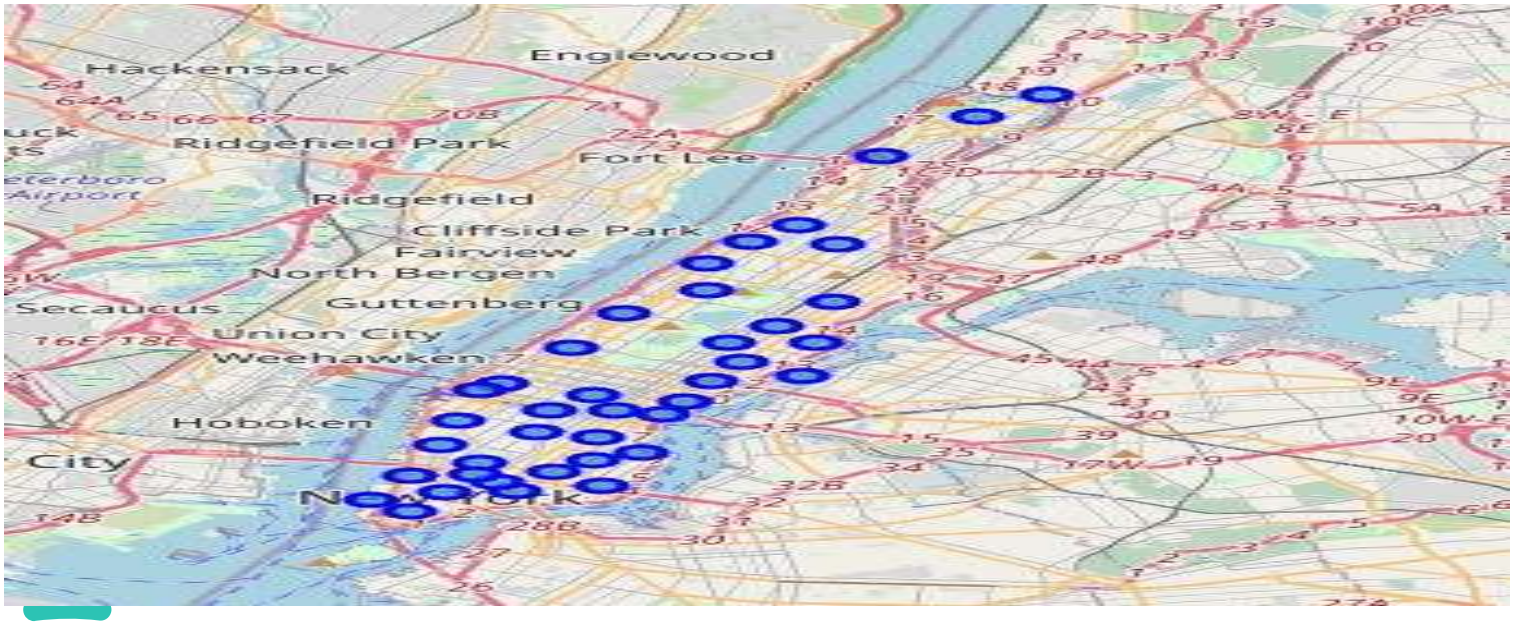
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315

Methodology

- This dataset was explored on a map using the folium library in order to get a better understanding of the number of neighborhoods in the New York City, and their distribution in the City.

Methodology

*To simplify our work, we only used the Manhattan borough data and performed exploratory analysis on it similar to the figure above



Methodology

	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

Methodology

- This dataset was explored on a map using the folium library in order to get a better understanding of the number of neighborhoods in the New York City, and their distribution in the City.

Methodology

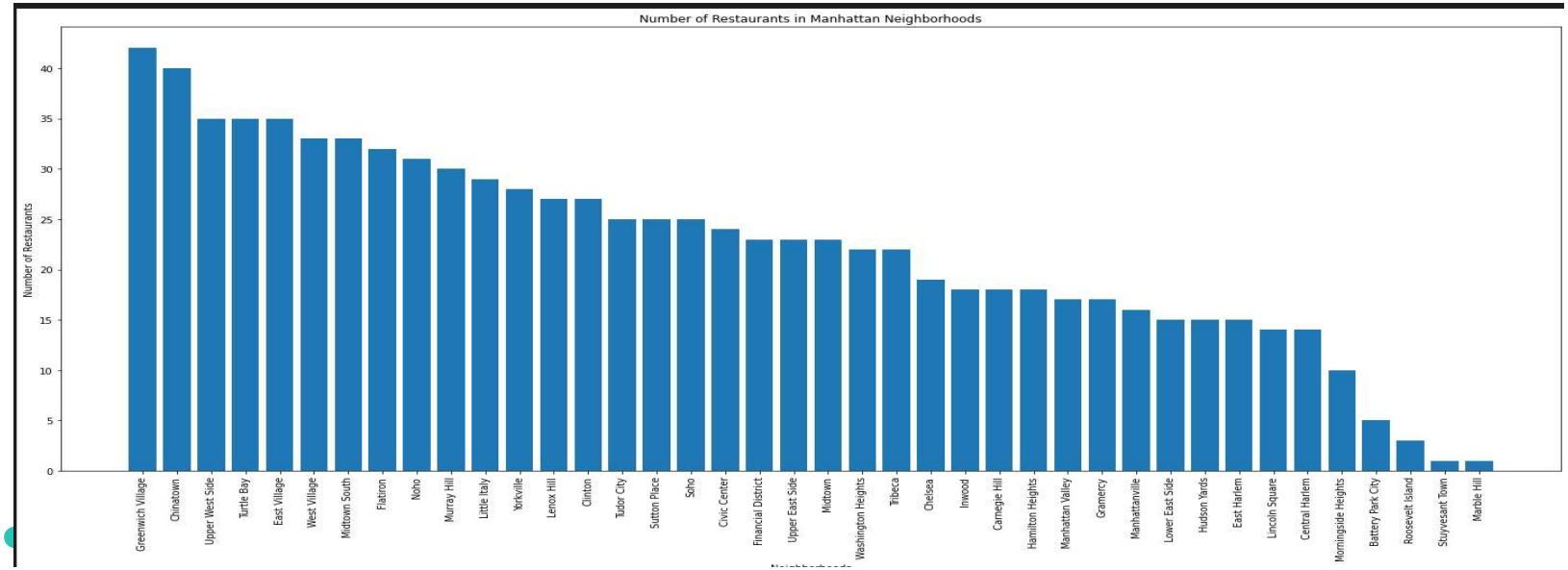
- The FourSquare API was called upon to retrieve information on venues local to each of these neighborhoods

(3230, 7)

	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	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

Methodology

- Linking this dataset back to the business problem of determining a suitable location for a restaurant, I filtered this dataset to only include venues that included the term ‘Restaurant’.
- I utilized the ‘group-by’ data analysis method to understand how many different ‘Restaurant’ venues there are in Manhattan, and how they are distributed among the various neighborhoods. I utilized a bar graph visual to enhance my understanding.



Methodology

- The one-hot encoding methodology was applied in order to sort the dataset. The five most popular restaurant types in each neighborhood were displayed.

```
----Carnegie Hill----
                venue  freq
0    Japanese Restaurant  0.11
1      Indian Restaurant  0.11
2      French Restaurant  0.11
3      Kosher Restaurant  0.06
4   Fast Food Restaurant  0.06
```

Methodology

- The Clustering machine learning technique was utilized to group Manhattan neighborhoods based on the top ten most popular restaurants in each neighborhood. Six cluster categories were used for simplicity

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Manhattan	Marble Hill	40.876551	-73.910660	1	Seafood Restaurant	Vietnamese Restaurant	Gluten-free Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	French Restaurant	German Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	3	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Asian Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Greek Restaurant	Hotpot Restaurant	Malay Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	3	Chinese Restaurant	Italian Restaurant	New American Restaurant	Tapas Restaurant	Latin American Restaurant	Spanish Restaurant	Sushi Restaurant	Caribbean Restaurant	Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Restaurant	Spanish Restaurant	Caribbean Restaurant	Chinese Restaurant	Empanada Restaurant	American Restaurant	Latin American Restaurant	Fast Food Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	3	Mexican Restaurant	Latin American Restaurant	Indian Restaurant	Sushi Restaurant	Caribbean Restaurant	Chinese Restaurant	Italian Restaurant	Fast Food Restaurant	Japanese Restaurant
5	Manhattan	Manhattanville	40.816934	-73.957385	0	Italian Restaurant	Mexican Restaurant	Seafood Restaurant	Dumpling Restaurant	Spanish Restaurant	Indian Restaurant	Falafel Restaurant	Ramen Restaurant	Cuban Restaurant
6	Manhattan	Central Harlem	40.815976	-73.943211	0	Chinese Restaurant	African Restaurant	American Restaurant	French Restaurant	Seafood Restaurant	Tapas Restaurant	Ethiopian Restaurant	Southern / Soul Food Restaurant	Caribbean Restaurant
7	Manhattan	East Harlem	40.792249	-73.944182	3	Mexican Restaurant	Thai Restaurant	Latin American Restaurant	Spanish Restaurant	New American Restaurant	Cuban Restaurant	French Restaurant	Restaurant	Hawaiian Restaurant
8	Manhattan	Upper East Side	40.775639	-73.960508	0	Italian Restaurant	French Restaurant	American Restaurant	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	Latin American Restaurant	Seafood Restaurant
9	Manhattan	Yorkville	40.775930	-73.947118	0	Italian Restaurant	Sushi Restaurant	Japanese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Thai Restaurant	Latin American Restaurant	Peruvian Restaurant	French Restaurant
10	Manhattan	Lenox Hill	40.768113	-73.958860	0	Italian Restaurant	Sushi Restaurant	Thai Restaurant	Afghan Restaurant	Chinese Restaurant	Greek Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Eastern European Restaurant

Methodology

- The dataset was later separated by cluster, in order to try and understand the data attributes that make up each cluster group:

```
## Cluster 1

manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	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
5	Manhattanville	Italian Restaurant	Mexican Restaurant	Seafood Restaurant	Dumpling Restaurant	Spanish Restaurant	Indian Restaurant	Falafel Restaurant	Ramen Restaurant	Cuban Restaurant
6	Central Harlem	Chinese Restaurant	African Restaurant	American Restaurant	French Restaurant	Seafood Restaurant	Tapas Restaurant	Ethiopian Restaurant	Southern / Soul Food Restaurant	Caribbean Restaurant
8	Upper East Side	Italian Restaurant	French Restaurant	American Restaurant	Sushi Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Tapas Restaurant	Latin American Restaurant	Seafood Restaurant
9	Yorkville	Italian Restaurant	Sushi Restaurant	Japanese Restaurant	Mexican Restaurant	Vietnamese Restaurant	Thai Restaurant	Latin American Restaurant	Peruvian Restaurant	French Restaurant
10	Lenox Hill	Italian Restaurant	Sushi Restaurant	Thai Restaurant	Afghan Restaurant	Chinese Restaurant	Greek Restaurant	Mexican Restaurant	Middle Eastern Restaurant	Eastern European Restaurant
12	Upper West Side	Italian Restaurant	Indian Restaurant	French Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	Thai Restaurant	American Restaurant	Mediterranean Restaurant	Mexican Restaurant
13	Lincoln Square	Italian Restaurant	French Restaurant	American Restaurant	Seafood Restaurant	Mediterranean Restaurant	Mexican Restaurant	Greek Restaurant	Chinese Restaurant	Vietnamese Restaurant
14	Clinton	Italian Restaurant	American Restaurant	New American Restaurant	Mediterranean Restaurant	French Restaurant	Restaurant	Korean Restaurant	Chinese Restaurant	Ethiopian Restaurant
15	Midtown	Mediterranean Restaurant	Japanese Restaurant	Turkish Restaurant	French Restaurant	Indian Restaurant	Cuban Restaurant	Sushi Restaurant	Greek Restaurant	Fast Food Restaurant
16	Murray Hill	Italian Restaurant	Japanese Restaurant	American Restaurant	Jewish Restaurant	Restaurant	Cuban Restaurant	Chinese Restaurant	Hawaiian Restaurant	Greek Restaurant
17	Chelsea	French Restaurant	American Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Paella Restaurant	Asian Restaurant	Israeli Restaurant	Indian Restaurant
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	Indian Restaurant	American Restaurant	French Restaurant	Seafood Restaurant	Chinese Restaurant	Vietnamese Restaurant	Ramen Restaurant
21	Tribeca	Italian Restaurant	American Restaurant	Greek Restaurant	French Restaurant	Sushi Restaurant	New American Restaurant	Seafood Restaurant	Modern European Restaurant	Korean Restaurant
22	Little Italy	Italian Restaurant	Chinese Restaurant	Mediterranean Restaurant	Japanese Restaurant	Hotpot Restaurant	Asian Restaurant	Seafood Restaurant	Mexican Restaurant	Middle Eastern Restaurant
23	Soho	Italian Restaurant	Mediterranean Restaurant	Asian Restaurant	Vegetarian / Vegan Restaurant	French Restaurant	Seafood Restaurant	Japanese Restaurant	Spanish Restaurant	Mexican Restaurant

Methodology

- Again, utilizing the Folium library, these clusters were displayed on a map of Calgary to gain an appreciation for any geographical influences on this dataset. Note: red circles represent Cluster 0; purple circles represent Cluster 1, blue circles represent Cluster 2, light green circles represent Cluster 3, yellow circles represent Cluster 4, Orange circle represents Cluster 5.

Result & Discussion

- From the data presented and analyzed, it is clear that there is demand for Category 0 (Cluster 0) – type restaurants, which are most likely to serve European-Asian Cuisine such as Italian, Chinese, French options.
- There is an obvious lack of restaurants near Cluster 1. From these results, it has been recommended that European – Asian cuisine restaurant should be opened
- Most of the neighborhoods in Manhattan that had FourSquare restaurant data associated with them appear to have been categorized into Cluster 0. Based off of Figure 8, this cluster appears to show a high frequency of European-Asian cuisine options, including Italian, Chinese, Japanese, Korean, French.

Thank you

Saksham Kapoor

sakshamkapoort@gmail.com