IBM DATA SCIENCE CAPSTONE PROJECT

Introduction : Business Problem

Manhattan, often referred to by residents of the New York City area as the city, is the most densely populated of the five boroughs of New York City, and coextensive with the County of New York, one of the original counties of the U.S. state of New York. Manhattan serves as the city's economic and administrative center, cultural identifier,[] and historical birthplace. Manhattan Island is divided into three informally bounded components, each aligned with the borough's long axis: Lower, Midtown, and Upper Manhattan.

The project will give an insight on the similarities and dissimilarities by categorizing the clusters of the subdistricts of Upper Manhattan. From the analysis, we can draw conclusions to opt for the optimal location for a new venue of a particular category.

This project will be helpful to any general entrepreneur but may be most useful for entrepreneurs on the food and beverage sector given that location can be the deciding factor for a success.

Data

To analyze the subdistricts of Upper Manhattan, the list of subdistrict is obtained from

List of Manhattan neighborhoods website.

Various venue queries have been made by the Foursquare API based on the locations of the subdistricts. The data is cleaned and with the processed data, we cluster the neighborhoods. The commonality clusters can then provide insight on which type of venue will thrive better on which cluster. K-means clustering algorithm will be used to find pattern between the subdistricts.

To summarize the data needed:

1.Subdistricts of Upper Manhattan

2.Locations of subdistricts

3.Trending venues based on the subdistrict

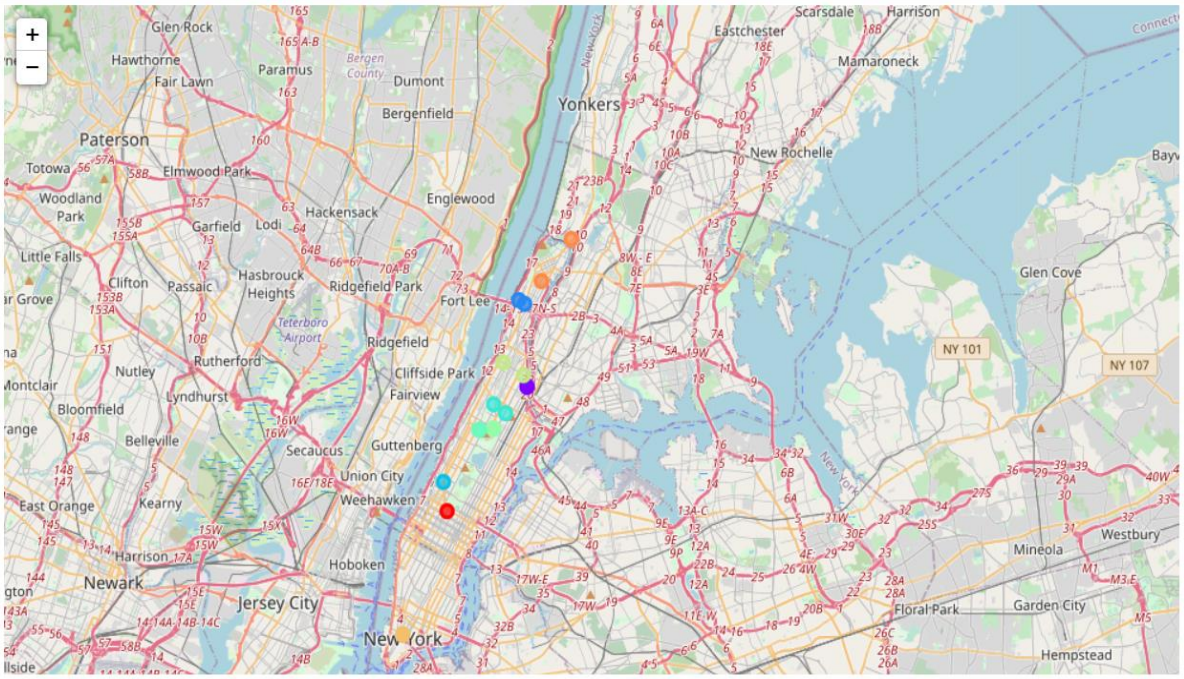
4.Venue Categories

Methodology:

Given our map of Upper Manhattan, we discovered the subdistricts of the neighborhoods and we use the K-Means clustering algorithm to cluster the neighborhoods. We use the one-hot encoder to encode the categorical variables of the venue categories retrieved by the Foursquare API and use the top venues for clustering.

Analysis:

Clusters



Cluster 1

Cluster 1

In [258]:
```
manhattan_merged[manhattan_merged['Cluster Labels']==1].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[258]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
10	Central Harlem	Park	Fountain	Coffee Shop	Playground	Wine Shop	Mexican Restaurant	Bakery	Pizza Place	Plaza	Café
13	Astor Row (Central Harlem)	Park	Fountain	Coffee Shop	Playground	Wine Shop	Mexican Restaurant	Bakery	Pizza Place	Plaza	Café
14	Sugar Hill (Central Harlem)	Park	Fountain	Coffee Shop	Playground	Wine Shop	Mexican Restaurant	Bakery	Pizza Place	Plaza	Café

Cluster 2 and Cluster 3

Harlem)

```
In [259]: manhattan_merged[manhattan_merged['Cluster Labels']==2].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1
```

Out[259]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
7	Hamilton Heights (part of Harlem)	Coffee Shop	Park	Bar	Chinese Restaurant	Tennis Court	Latin American Restaurant	Mexican Restaurant	Deli / Bodega	Sandwich Place	Café
19	Lenox Hill	Bar	Coffee Shop	Park	Café	Baseball Stadium	Tennis Court	Southern / Soul Food Restaurant	Ethiopian Restaurant	Baseball Field	Beer Bar

```
In [260]: manhattan_merged[manhattan_merged['Cluster Labels']==3].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1
```

Out[260]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
12	St. Nicholas Historic District, aka Strivers' ...	Coffee Shop	Park	Wine Shop	Hotel	Spa	Café	Memorial Site	Gym / Fitness Center	Cycle Studio	Men's Store
16	Le Petit Senegal (Little Senegal)	Coffee Shop	Park	Wine Shop	Hotel	Spa	Café	Memorial Site	Gym / Fitness Center	Cycle Studio	Men's Store
21	Yorkville	Coffee Shop	Park	Wine Shop	Hotel	Spa	Café	Memorial Site	Gym / Fitness Center	Cycle Studio	Men's Store

Cluster 4 and Cluster 6

In [261]:
```
manhattan_merged[manhattan_merged['Cluster Labels']==4].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[261]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
1	Marble Hill	Pizza Place	Mexican Restaurant	Park	Diner	Café	Coffee Shop	Bar	Bakery	Latin American Restaurant	Deli / Bodega
3	Fort George (part of Washington Heights)	Mexican Restaurant	Latin American Restaurant	Park	Pizza Place	Café	Deli / Bodega	Wine Bar	Restaurant	Spanish Restaurant	Tapas Restaurant

In [262]:
```
manhattan_merged[manhattan_merged['Cluster Labels']==6].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[262]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
20	Carnegie Hill	Theater	Hotel	Concert Hall	Plaza	Performing Arts Venue	Gym	Jazz Club	Zoo	Bakery	Spa

Cluster 7 and Cluster 8

In [263]:
```
manhattan_merged[manhattan_merged['Cluster Labels']==7].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[263]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
6	West Harlem	Park	Café	Southern / Soul Food Restaurant	Coffee Shop	Tennis Court	Mexican Restaurant	Theater	Arts & Crafts Store	Italian Restaurant	Bar
11	Harlem	Park	Café	Southern / Soul Food Restaurant	Coffee Shop	Tennis Court	Mexican Restaurant	Theater	Arts & Crafts Store	Italian Restaurant	Bar
17	East Harlem (Spanish Harlem)	Park	Café	Southern / Soul Food Restaurant	Coffee Shop	Tennis Court	Mexican Restaurant	Theater	Arts & Crafts Store	Italian Restaurant	Bar

In [264]:
```
manhattan_merged[manhattan_merged['Cluster Labels']==8].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[264]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
23	Manhattan Valley, Bloomingdale District	Coffee Shop	Grocery Store	Hotel	Bar	Music Venue	Thai Restaurant	Sandwich Place	Pizza Place	Korean Restaurant	Italian Restaurant

Cluster 9 and Cluster 10

```
In [265]: manhattan_merged[manhattan_merged['Cluster Labels']==9].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[265]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
4	Washington Heights	Pizza Place	Latin American Restaurant	Park	Café	Bakery	Spanish Restaurant	Wine Shop	Coffee Shop	Mobile Phone Shop	Tapas Restaurant
5	Hudson Heights (part of Washington Heights)	Pizza Place	Latin American Restaurant	Park	Café	Spanish Restaurant	Bakery	Tapas Restaurant	Mexican Restaurant	Bar	Wine Shop

```
In [266]: manhattan_merged[manhattan_merged['Cluster Labels']==10].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[266]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
24	Lincoln Square (once San Juan Hill)	Park	Gym / Fitness Center	Bakery	Gym	Theater	Concert Hall	American Restaurant	Indie Movie Theater	Ice Cream Shop	Performing Arts Venue

Cluster 11 and Cluster 12

```
In [267]: manhattan_merged[manhattan_merged['Cluster Labels']==11].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[267]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
9	Morningside Heights	Coffee Shop	Park	Seafood Restaurant	Italian Restaurant	Jazz Club	Southern / Soul Food Restaurant	Wine Shop	Bakery	Playground	French Restau
15	Marcus Garvey Park, Mount Morris Historical Di...	Park	Southern / Soul Food Restaurant	French Restaurant	Jazz Club	Coffee Shop	Cocktail Bar	Café	Grocery Store	African Restaurant	Mexica Restau

```
In [268]: manhattan_merged[manhattan_merged['Cluster Labels']==12].loc[:][manhattan_merged.columns[[0]+list(range(4,manhattan_merged.shape[1]))]]
```

Out[268]:

	Subdistrict	1th Most Common Venue	2th Most Common Venue	3th 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	10th Most Common Venue
18	Upper East Side	Park	Coffee Shop	Grocery Store	Playground	Dog Run	Scenic Lookout	Fountain	Bakery	Bagel Shop	African Restaurant
22	Upper West Side	Park	Coffee Shop	Grocery Store	Playground	Dog Run	Scenic Lookout	Fountain	Bakery	Bagel Shop	African Restaurant

Discussion:

Based on the above clusters we can observe how similar the clusters are in terms of venue categories. We can infer that Manhattan is a very structured city and every cluster has very distinctive features and neighborhoods of a same clusters are very correlated. For example in Cluster 1 the rank of Park, Coffee shops, Pizza Place are same. Thus depending on the business or store, an entrepreneur can decide which locality/neighborhood to select to strive and have a thriving economy. However, the limit of 100 venues from the Foursquare API might affect the results.

Conclusion:

Thus the above clusters tables can give immense information about the most feasible locations to open a particular business.