

# **The Battle of The Neighborhoods Kubra Yetkin**



## Introduction and Discussion of the Business Objective

---

### Problem Description

- MEG Company produces cosmetic products, they have some stores in Europe but they also want to built a store in Manhattan.
- Before they release the product to the market; MEG Company want to build a system that can help to reccomend new places for their new Cosmetic Shop that will open in Manhattan.

### Criteria

- This analyses based on the idea that is best place to built a new Cosmetic Shop that are near Italian Restaurants, Cafés, Coffee Shops and Hotels. The people live in Manhattan are very social people that frequent these place often.



## Data Requirements

---

### Data Collection

- The number of main districts in Manhattan is 40 . The data regarding the districts in Manhattan needs to be provided for this analyses. I am able to rich the list of Boroughs in New York from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset). The table gives me the list as shown below.

	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



## Data Requirements

---

- New York has 5 boroughs and 306 neighborhoods. After that we need the data includes just Manhattan's neighborhoods, latitudes and longitudes. The data frame as shown below gives us these values. The cleansed data will be used with Foursquare data.

```
manhattan_data = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)
manhattan_data.head()
```

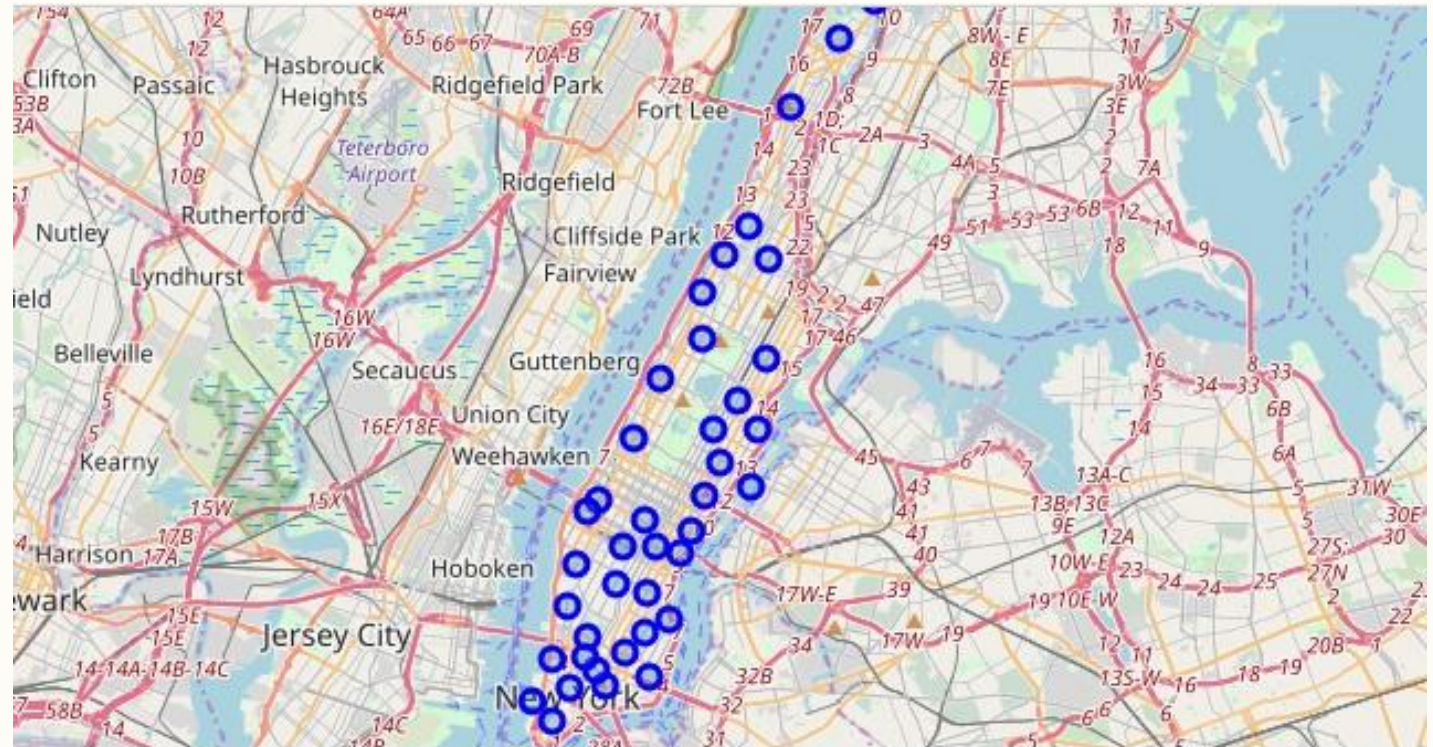
```
]:
```

	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 Visualization

- Create a map of Manhattan with neighborhoods superimposed on top
- The distribution of the neighborhoods in Manhattan as shown below





## Methodology

---

### Visualization

- Analysis and plotting for visualizations.
- Data visualization using various mapping libraries.

### Exploratory and Inferential Analyses

- Clustering algorithms will be used to cluster the data.
- Italian Restaurants, Cafés, Coffee Shops and Hotels: These are the venue types that the client wants to add the analyses because of these venues have a powerful density for the ideal store locations.
- Inferential analysis will be made using the data and this provides us to make recommendations for possible locations to new Cosmetic Shop store.



## Methodology

---

### Exploratory and Inferential Analyses

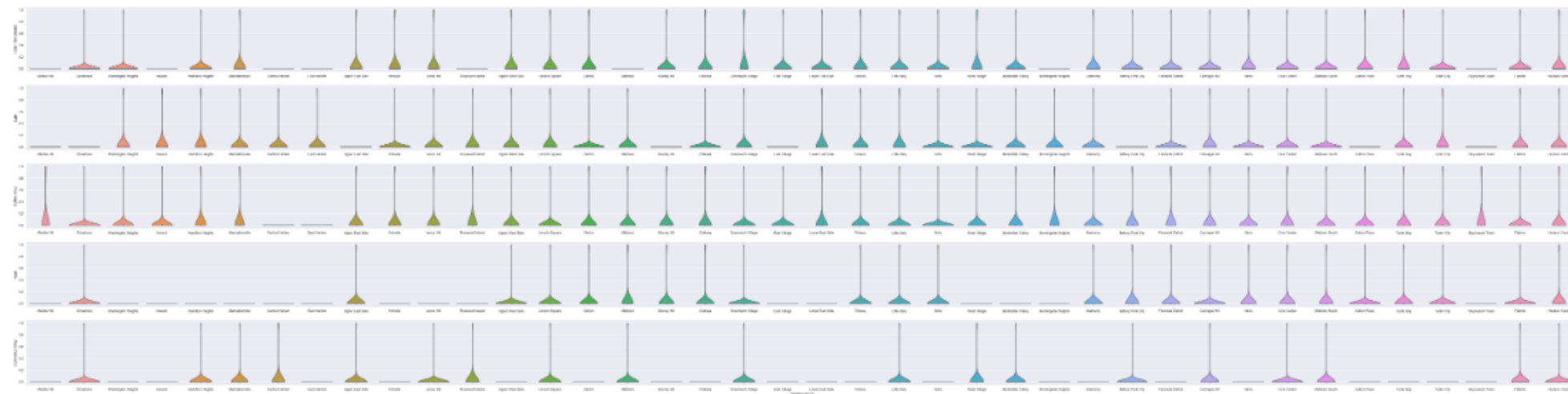
- Create a nearby venues function for all the neighborhoods within a Radius of 500 meters in Manhattan

	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
5	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym
6	Marble Hill	40.876551	-73.91066	Blink Fitness	40.877271	-73.905595	Gym
7	Marble Hill	40.876551	-73.91066	T.J. Maxx	40.877232	-73.905042	Department Store
8	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
9	Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	40.878628	-73.914568	Tennis Stadium

# Methodology

## Exploratory and Inferential Analyses

- We have the 17 neighborhoods that all include the venue category criteria that specified by the client (Italian Restaurants, Cafés, Coffee Shops and Hotels)
- When we included the "Cosmetics Shop" venue category into the analysis, then we are able to make some inferences based on the data to focus the possible districts.
- Frequency distribution for the top 5 venue categories for each neighborhood as shown below.







## Results

---

### Chosen the Neighborhood

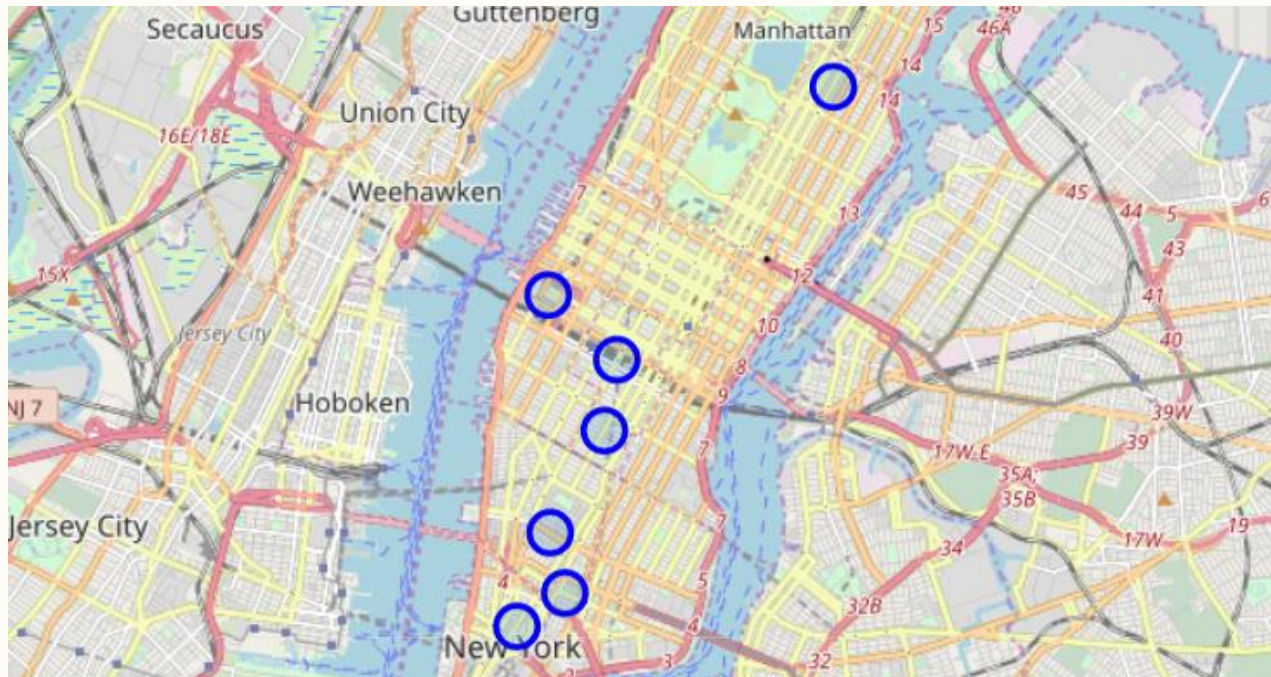
- When the 4 venue types have been added to analyses, we are able to determine the neighborhoods that are positioned with Restaurants, Cafés, Coffee Shops, Hotels and adding Cosmetic Shops to this analysis will gives us the possible neighborhoods.
- So we can increase the types of venues to include Restaurants, Cafés, Coffee Shops, Hotels and Cosmetic Shops, which narrows down our possible neighborhoods and give us the chance to focus the suggested districts for new stores will be located in Manhattan.
- Finally, 7 prospective neighborhoods that include 5 criterias are determined for the new store, these areas listed below :

Hudson Yards  
Flatiron  
Midtown South  
Civic Center  
Carnegie Hill  
Little Italy  
Greenwich Village

	Neighborhood	Latitude	Longitude
0	Hudson Yards	40.756658	-74.000111
1	Flatiron	40.739673	-73.990947
2	Midtown South	40.748510	-73.988713
3	Civic Center	40.715229	-74.005415
4	Carnegie Hill	40.782683	-73.953256
5	Little Italy	40.719324	-73.997305
6	Greenwich Village	40.726933	-73.999914

## Results

- Create a map to show our determined 7 neighborhoods, these districts are possible locations for new store.





## Discussion and Conclusions:

---

- As a result of these inferences, the data is able to provide us for possible location recommendations, but we could not determine the exact point for the new store. There is no accurate answer because this analyses are based on the criteria (types of venues), maybe our options can be narrow down if the different types of venues had been chosed. This results are the best districts to match our criteria.

### Conclusions

- There are many ways to perform this analysis using a different methodology and maybe even different data source. I worked with clustering algorithms to cluster the data but I see that the results have no accurate endpoint because this is based on the criterias that are chosen by us. On the other hand, this data analysis recommendations helped us to narrow down the possible options and it gave to us a startpoint for the possible exact store points.