IBM DataScience Capstone Final Project

Aleksandr V.

Feb 2020

1. Introduction

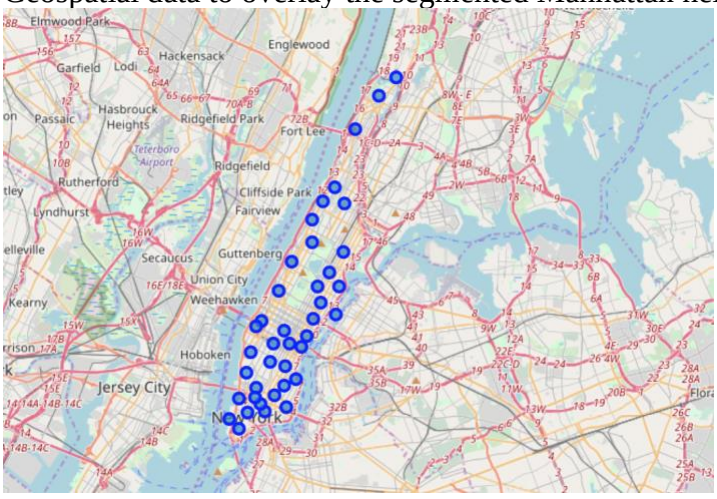
New York is very diverse and is the financial capitals of the United States. It is known as the "City that Never Sleeps" and has many different aspects that it is known for around the world. One of which being one of the most diverse culinary landscapes in the world. When it comes to opening a restaurant having a great product does not equal success, location is extremely important. In this project we will try to find an optimal location for a restaurant. In this project I developed an application that would allow a restauranteur to explore potential locations to open a new restaurant in New York City; specifically Manhattan.

2. Data

The data I will be utilizing will be neighborhood and borough information for New York was taken from the file "newyork_data.json", provided in the coursera class; Geospacial data was pulled from http://cocl.us/Geospacial_data to pair the latitude and longitude of each neighborhood; as well as the foursquare API to access venue data for New York

3. Methodology

In the first part of the project, I utilized the New York Json data create a dataset and separate Manhattan neighborhoods into a separate dataset. Following which I used and Geospatial data to overlay the segmented Manhattan neighborhoods onto a folium map.



In then utilized foursquare API in order to locate the top 100 venues in each neighborhood creating a dataframe to show each restaurant category.

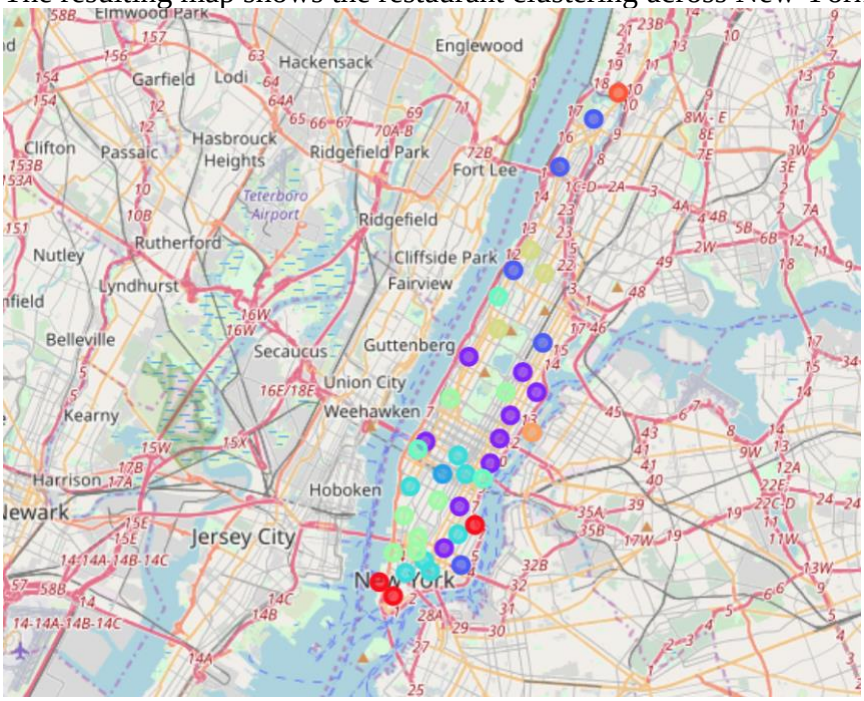
Out[69]:

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery	Belgian Restaurant	Bistro	Brazilian Restaurant	Breakfast Spot	Burger Joint	Burrito Place	Cafeteria	Café	Cajun / Creole Restaurant
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

After gathering this information; I used K-means clustering to partition the 47 neighborhoods in Manhattan into 10 clusters on the frequency of occurrence of each restaurant category. The results were used to provide insights of what restaurants are prominent in certain areas in order to be a starting point for a search for a new restaurant location.

4. Results

The resulting map shows the restaurant clustering across New York neighborhoods.



The clusters are as follows:

Cluster 1

```
In [78]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[78]:

	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	10th Most Common Venue
28	Battery Park City	Pizza Place	Donut Shop	Sandwich Place	Italian Restaurant	Chinese Restaurant	Food Truck	Steakhouse	Seafood Restaurant	Restaurant	Mexican Restaurant
29	Financial District	Pizza Place	American Restaurant	Sandwich Place	Italian Restaurant	Mexican Restaurant	Food Truck	Café	Salad Place	Deli / Bodega	Steakhouse
37	Stuyvesant Town	Pizza Place	Deli / Bodega	Bakery	Diner	Sandwich Place	Ethiopian Restaurant	Mexican Restaurant	South Indian Restaurant	Bistro	Taco Place

Cluster 2

```
In [79]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[79]:

	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	10th Most Common Venue
9	Yorkville	Italian Restaurant	Pizza Place	Thai Restaurant	Deli / Bodega	Indian Restaurant	Japanese Restaurant	Mexican Restaurant	Diner	Sandwich Place	Sushi Restaurant
10	Lenox Hill	Italian Restaurant	Sushi Restaurant	Café	Burger Joint	Pizza Place	Deli / Bodega	Mexican Restaurant	Bakery	Restaurant	Thai Restaurant
12	Upper West Side	Italian Restaurant	Pizza Place	American Restaurant	Bakery	Indian Restaurant	Mediterranean Restaurant	Deli / Bodega	Bagel Shop	Café	Breakfast Spot
14	Clinton	Italian Restaurant	American Restaurant	Thai Restaurant	Pizza Place	Sandwich Place	Mexican Restaurant	Bakery	Restaurant	Chinese Restaurant	Steakhouse
27	Gramercy	Italian Restaurant	Deli / Bodega	Indian Restaurant	Thai Restaurant	Bagel Shop	Mexican Restaurant	Pizza Place	Sandwich Place	Restaurant	Diner
30	Carnegie Hill	Pizza Place	Café	Bakery	Italian Restaurant	Sushi Restaurant	Mexican Restaurant	Japanese Restaurant	French Restaurant	Sandwich Place	Chinese Restaurant
31	Noho	Italian Restaurant	Pizza Place	Japanese Restaurant	Sushi Restaurant	Mexican Restaurant	Bakery	Café	Sandwich Place	French Restaurant	Thai Restaurant
34	Sutton Place	Italian Restaurant	American Restaurant	Indian Restaurant	Pizza Place	Chinese Restaurant	Sushi Restaurant	Japanese Restaurant	Salad Place	French Restaurant	Thai Restaurant
35	Turtle Bay	Italian Restaurant	Japanese Restaurant	Indian Restaurant	Steakhouse	Deli / Bodega	Sushi Restaurant	Greek Restaurant	French Restaurant	Ramen Restaurant	Seafood Restaurant

Cluster 3

```
In [80]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[80]:

	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	10th Most Common Venue
2	Washington Heights	Bakery	Mexican Restaurant	Deli / Bodega	Chinese Restaurant	Latin American Restaurant	Tapas Restaurant	Pizza Place	Café	Donut Shop	Fast Food Restaurant
3	Inwood	Spanish Restaurant	Mexican Restaurant	Pizza Place	Restaurant	Café	Bakery	Latin American Restaurant	Deli / Bodega	Seafood Restaurant	Thai Restaurant
5	Manhattanville	Chinese Restaurant	Mexican Restaurant	Deli / Bodega	Sandwich Place	Seafood Restaurant	Fried Chicken Joint	Bakery	Italian Restaurant	Sushi Restaurant	Donut Shop
7	East Harlem	Deli / Bodega	Mexican Restaurant	Pizza Place	Bakery	Latin American Restaurant	Restaurant	Chinese Restaurant	Thai Restaurant	Burger Joint	Fast Food Restaurant
20	Lower East Side	Deli / Bodega	Chinese Restaurant	Pizza Place	Mexican Restaurant	Café	Japanese Restaurant	Bakery	Sandwich Place	Italian Restaurant	Diner

Cluster 4

```
In [81]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[81]:

	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	10th Most Common Venue
33	Midtown South	Korean Restaurant	Japanese Restaurant	Bakery	American Restaurant	Salad Place	Burger Joint	Italian Restaurant	Sandwich Place	New American Restaurant	Restaurant

Cluster 5

```
In [82]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[82]:

	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	10th Most Common Venue
1	Chinatown	Chinese Restaurant	Bakery	Vietnamese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Hotpot Restaurant	Noodle House	Sandwich Place	Mexican Restaurant	American Restaurant
15	Midtown	American Restaurant	Sandwich Place	Bakery	Steakhouse	Japanese Restaurant	French Restaurant	Burger Joint	Café	Italian Restaurant	Pizza Place
16	Murray Hill	Sandwich Place	Japanese Restaurant	Sushi Restaurant	American Restaurant	Burger Joint	Pizza Place	Italian Restaurant	Bakery	Restaurant	Seafood Restaurant
17	Chelsea	Bakery	American Restaurant	Italian Restaurant	French Restaurant	Pizza Place	Sushi Restaurant	Seafood Restaurant	New American Restaurant	Tapas Restaurant	Mexican Restaurant
19	East Village	Pizza Place	Vietnamese Restaurant	Chinese Restaurant	Mexican Restaurant	Vegetarian / Vegan Restaurant	French Restaurant	Italian Restaurant	Korean Restaurant	Japanese Restaurant	Ramen Restaurant
22	Little Italy	Chinese Restaurant	Italian Restaurant	Café	Bakery	Pizza Place	Mediterranean Restaurant	Thai Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	Sandwich Place
32	Civic Center	Italian Restaurant	French Restaurant	American Restaurant	Chinese Restaurant	Sandwich Place	Café	Bakery	Burger Joint	Vietnamese Restaurant	Pizza Place

Cluster 6

```
In [83]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 5, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[83]:

	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	10th Most Common Venue
26	Morningside Heights	Deli / Bodega	Chinese Restaurant	Italian Restaurant	Pizza Place	Café	Food Truck	Burger Joint	American Restaurant	Mexican Restaurant	Sandwich Place
36	Tudor City	Café	Deli / Bodega	Pizza Place	Food Truck	Mexican Restaurant	Sushi Restaurant	Burger Joint	Japanese Restaurant	Bagel Shop	Sandwich Place
39	Hudson Yards	Italian Restaurant	American Restaurant	Café	Deli / Bodega	Restaurant	Pizza Place	Salad Place	Sandwich Place	Burger Joint	Spanish Restaurant

Cluster 7

```
In [84]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 6, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[84]:

	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	10th Most Common Venue
8	Upper East Side	Italian Restaurant	American Restaurant	Bakery	Pizza Place	Diner	Café	Sushi Restaurant	Salad Place	Burger Joint	Mexican Restaurant
13	Lincoln Square	Italian Restaurant	Food Truck	Café	French Restaurant	American Restaurant	Pizza Place	Chinese Restaurant	Bakery	Mediterranean Restaurant	Mexican Restaurant
18	Greenwich Village	Italian Restaurant	Café	Sushi Restaurant	American Restaurant	Indian Restaurant	Pizza Place	Chinese Restaurant	French Restaurant	Sandwich Place	Seafood Restaurant
21	Tribeca	Italian Restaurant	American Restaurant	Deli / Bodega	French Restaurant	Café	Pizza Place	Mexican Restaurant	Asian Restaurant	Sushi Restaurant	Steakhouse
23	Soho	Italian Restaurant	Café	French Restaurant	Mediterranean Restaurant	American Restaurant	Sandwich Place	Bakery	Vegetarian / Vegan Restaurant	Pizza Place	Asian Restaurant
24	West Village	Italian Restaurant	American Restaurant	New American Restaurant	Japanese Restaurant	Pizza Place	French Restaurant	Seafood Restaurant	Gastropub	Steakhouse	Indian Restaurant
38	Flatiron	Italian Restaurant	American Restaurant	New American Restaurant	Sandwich Place	Mediterranean Restaurant	Japanese Restaurant	Café	Mexican Restaurant	Vegetarian / Vegan Restaurant	Sushi Restaurant

Cluster 8

```
In [85]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 7, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[85]:

	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	10th Most Common Venue
4	Hamilton Heights	Deli / Bodega	Pizza Place	Mexican Restaurant	Chinese Restaurant	Donut Shop	Sandwich Place	Café	Food Truck	American Restaurant	Fast Food Restaurant
6	Central Harlem	Deli / Bodega	Fried Chicken Joint	Southern / Soul Food Restaurant	African Restaurant	Seafood Restaurant	Caribbean Restaurant	Chinese Restaurant	Pizza Place	French Restaurant	Sandwich Place
25	Manhattan Valley	Deli / Bodega	Indian Restaurant	Pizza Place	Mexican Restaurant	Chinese Restaurant	Latin American Restaurant	Thai Restaurant	Gastropub	Café	Bagel Shop

Cluster 9 ¶

```
In [86]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 8, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[86]:

	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	10th Most Common Venue
11	Roosevelt Island	Café	Greek Restaurant	Sandwich Place	Japanese Restaurant	Caribbean Restaurant	Food	Deli / Bodega	Pizza Place	Dumpling Restaurant	Eastern European Restaurant

Cluster 10

```
In [87]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 9, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[87]:

	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	10th Most Common Venue
0	Marble Hill	Sandwich Place	Pizza Place	Deli / Bodega	American Restaurant	Donut Shop	Latin American Restaurant	Café	Mexican Restaurant	Diner	Spanish Restaurant

5. Discussion

Our analysis shows that although there is a great number of restaurants in Manhanntan, there are pockets of low restaurant diversity represented by outlier clusters containing singular neighborhood clusters (Clusters 4,9,10). Highest diversity of restaurants categories was detected in Clusters 2 & 7 which are areas that offer a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. This does not imply that these areas are actually optimal locations for a new restaurant. Purpose of this analysis was to only provide information on areas which currently have a strong restaurant presence. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition, but also other factors taken into account and all other relevant conditions met.

6. Conclusion

Purpose of this project was to identify areas in Manhattan and segment what restaurants were present in order to aid a restauranteur in narrowing down the search for optimal location for a restaurant. By utilizing Foursquare data to identify restaurant distribution we have first identified general areas that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding

existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors such as competition quality, real estate availability/pricing, social and economic dynamics of every neighborhood.