The Battle of Neighborhoods

The Prospect of a Lunch Restaurant Close to Office Areas in New York City

Fatimah Alhassan

September 8, 2019

Table of Contents

1. Introduction

1.1 Problem Description and Background

New York City is one of the World's 20 largest metropolitan areas and the financial capital of USA. It is a global hub of all kinds of businesses such as banking, retailing, world trade, transportation, tourism, real estate, legal services, accountancy, insurance, theater, fashion, and arts. It is definitely one of the best places to start a new business.

Office areas provide great opportunities for restaurants during daytime. Shops are usually full especially at lunch time (11am - 2pm). Given this scenario I will analyze the prospect of opening a lunch restaurant, close to office areas in the city of New York. The analysis will provide insights about which neighborhoods are suitable to open a lunch restaurant. New York City is made up of 306 neighborhoods, but I will mostly concentrate on the Central Business Districts CBD to target the largest pool of workers. According to ny-offices.com , New York City CBDs:

1. Midtown Manhattan
2. Lower Manhattan, financial district of NYC
3. Downtown Brooklyn
4. Downtown Flushing
5. Long Island City

1.2 Target Audience

- Anyone who is interested in starting a restaurant business in New York City. This analysis can be used as a guide to start or expand restaurants targeting larger pool of office workers during lunch hours.
- Office workers, to find potential lunch places close to office.
- Students who wants to develop their data science skills. This project is a good practical case to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it, and, finally to tell a story out of it.

2. Data

Below is the data to be used in this project:

2.1 New York City Dataset

According to the business problem described in the previous section, the data required for analysis is a dataset of New York City neighborhoods and boroughs. New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, I will need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. The required dataset can be found here. I will use this dataset to compare different neighborhoods within each borough in order to figure out the most suitable neighborhoods to open a lunch restaurant. The figure below shows 20 sample of the dataset.

|	Borough	Neighborhood	Latitude	Longitude
245	Staten Island	Bloomfield	40.605779	-74.187256
20	Bronx	Melrose	40.819754	-73.909422
0	Bronx	Wakefield	40.894705	-73.847201
50	Brooklyn	Gravesend	40.595260	-73.973471
300	Brooklyn	Erasmus	40.646926	-73.948177
232	Staten Island	Midland Beach	40.573527	-74.093483
209	Staten Island	New Springville	40.594252	-74.164960
4	Bronx	Riverdale	40.890834	-73.912585
265	Queens	Pomonok	40.734936	-73.804861
142	Queens	Maspeth	40.725427	-73.896217
263	Queens	Jamaica Hills	40.711460	-73.796465
237	Staten Island	Pleasant Plains	40.524699	-74.219831
126	Manhattan	Gramercy	40.737210	-73.981376
238	Staten Island	Butler Manor	40.506082	-74.229504
284	Staten Island	Manor Heights	40.601810	-74.120594
37	Bronx	Pelham Bay	40.850641	-73.832074
87	Brooklyn	Boerum Hill	40.685683	-73.983748
98	Brooklyn	Ocean Parkway	40.613060	-73.968367
230	Staten Island	Old Town	40.596329	-74.087511
79	Brooklyn	Bath Beach	40.599519	-73.998752

2.2 Foursquare API

Foursquare API is a well-known free API to get details related to specific places based on geographical coordinates and radius. In this project, I will utilize the Foursquare API to obtain the top 100 venues within 1 km of the center of each neighborhood (using *explore* endpoint).

3. Methodology

3.1 Exploring New York City's Central Business Districts

Foursquare data shows that there are 162 unique categories in the five Central Business Districts. However, the focus of this project is on *Restaurant* venue type. By searching for the keyword *Restaurant* in the 500 returned venues, I have found that there are 134 restaurants in the five districts. Next, I used Folium library to plot a leaflet map of these restaurants as illustrated below.
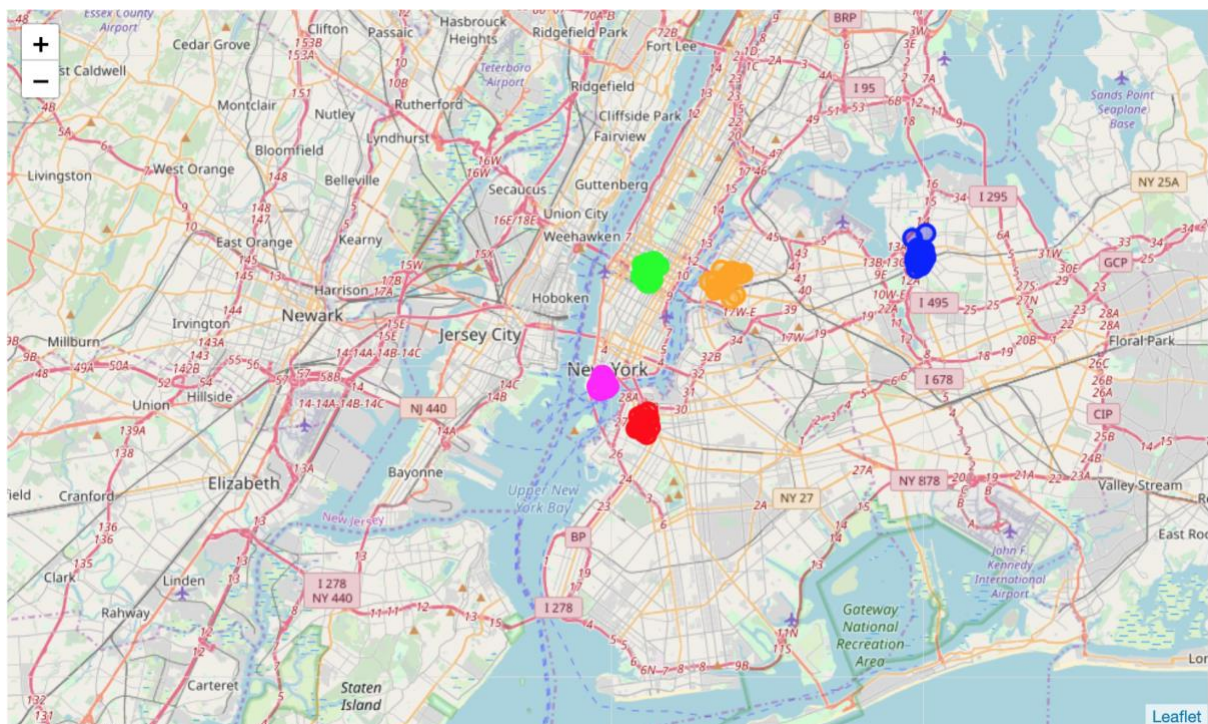
By doing further analysis to find the ten most frequently occurring venues in each of the five districts, I have found that *Coffee Shops* top the charts of the most common venues among all the five districts, followed by *Chinese Restaurant*. A plot of the ten most frequent venues among the five districts is exhibited below.
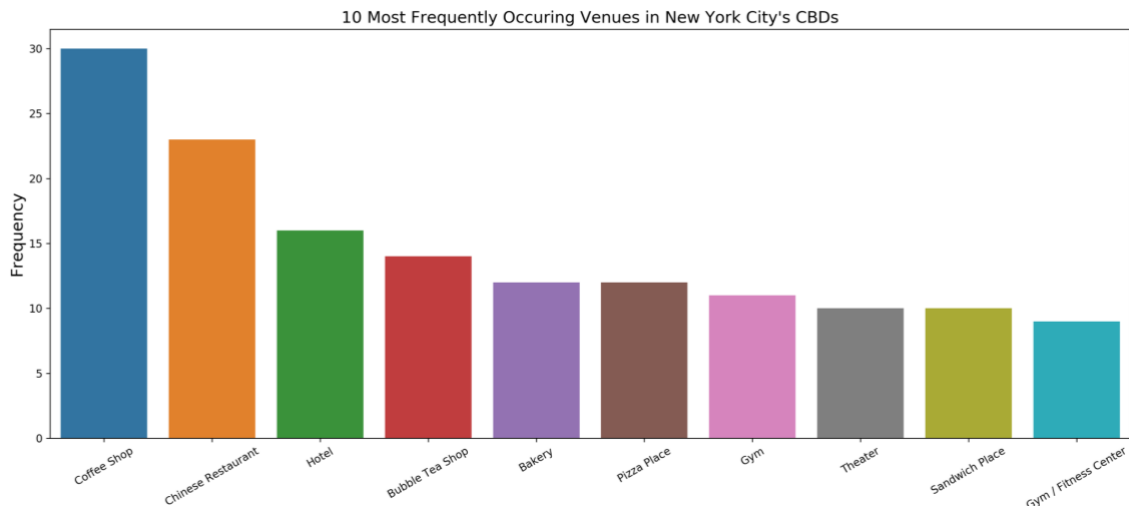
Next step is to analyze each of the five districts by finding the ten most common venues in each district along with the frequency of each venue. The figure below summarizes the results.

```
----Downtown----
                   venue  freq
0             Coffee Shop  0.07
1                  Bakery  0.04
2                     Bar  0.03
3          Sandwich Place  0.03
4     Gym / Fitness Center  0.03
5      Chinese Restaurant  0.02
6   Performing Arts Venue  0.02
7             Pizza Place  0.02
8            Burger Joint  0.02
9    Japanese Restaurant  0.02


----Financial District----
                   venue  freq
0             Coffee Shop  0.08
1     American Restaurant  0.04
2               Wine Shop  0.04
3             Pizza Place  0.04
4          Sandwich Place  0.03
5      Falafel Restaurant  0.03
6     Gym / Fitness Center  0.03
7             Cocktail Bar  0.03
8                    Café  0.03
9             Event Space  0.03


----Flushing----
                   venue  freq
0      Chinese Restaurant  0.18
1         Bubble Tea Shop  0.12
2       Hotpot Restaurant  0.07
3        Korean Restaurant  0.06
4      Dumpling Restaurant  0.05
5                  Bakery  0.05
6         Asian Restaurant  0.04
7    Cantonese Restaurant  0.03
8             Pizza Place  0.02
9              Food Court  0.02
```

```
----Long Island City----
                   venue  freq
0             Coffee Shop  0.10
1                   Hotel  0.09
2      Mexican Restaurant  0.04
3       Indian Restaurant  0.03
4             Pizza Place  0.03
5      Italian Restaurant  0.03
6                    Café  0.03
7     American Restaurant  0.02
8            Dessert Shop  0.02
9              Donut Shop  0.02


----Midtown----
                   venue  freq
0                 Theater  0.08
1             Coffee Shop  0.04
2                     Gym  0.04
3                   Hotel  0.04
4                   Plaza  0.03
5          Sandwich Place  0.03
6         Cuban Restaurant  0.03
7     American Restaurant  0.02
8          Clothing Store  0.02
9      Chinese Restaurant  0.02
```

As the scope of the project is on restaurant venue type, I will explore next which district has the highest number of restaurants as the most common venue type. The plot below shows the answer.
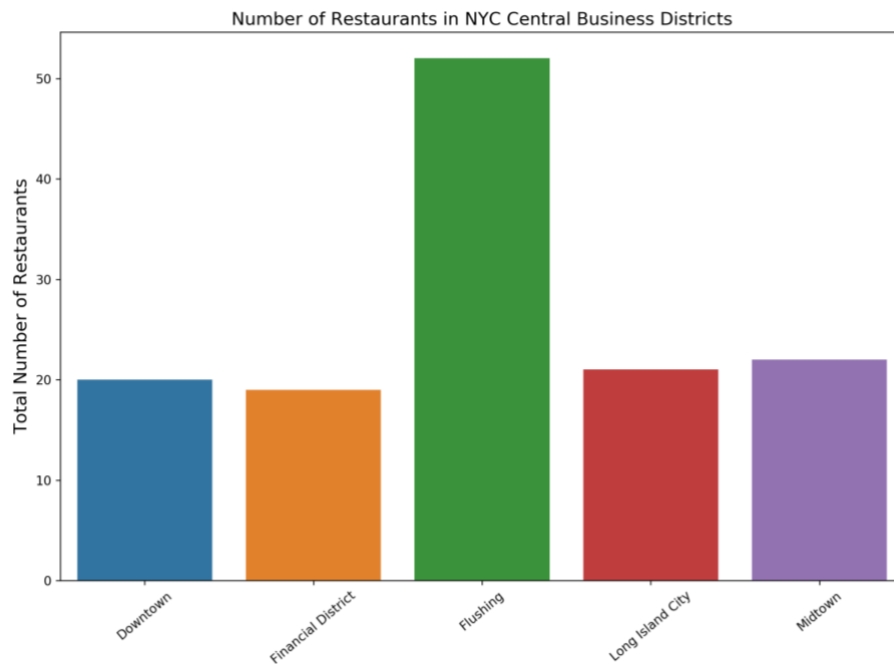
3.2 Clustering

Finally, I will use **K-Means** to cluster these five districts. Districts that have similar venue categories will be clustered together. By Initializing the K-Means model with k=3, I will get the three clusters shown below.
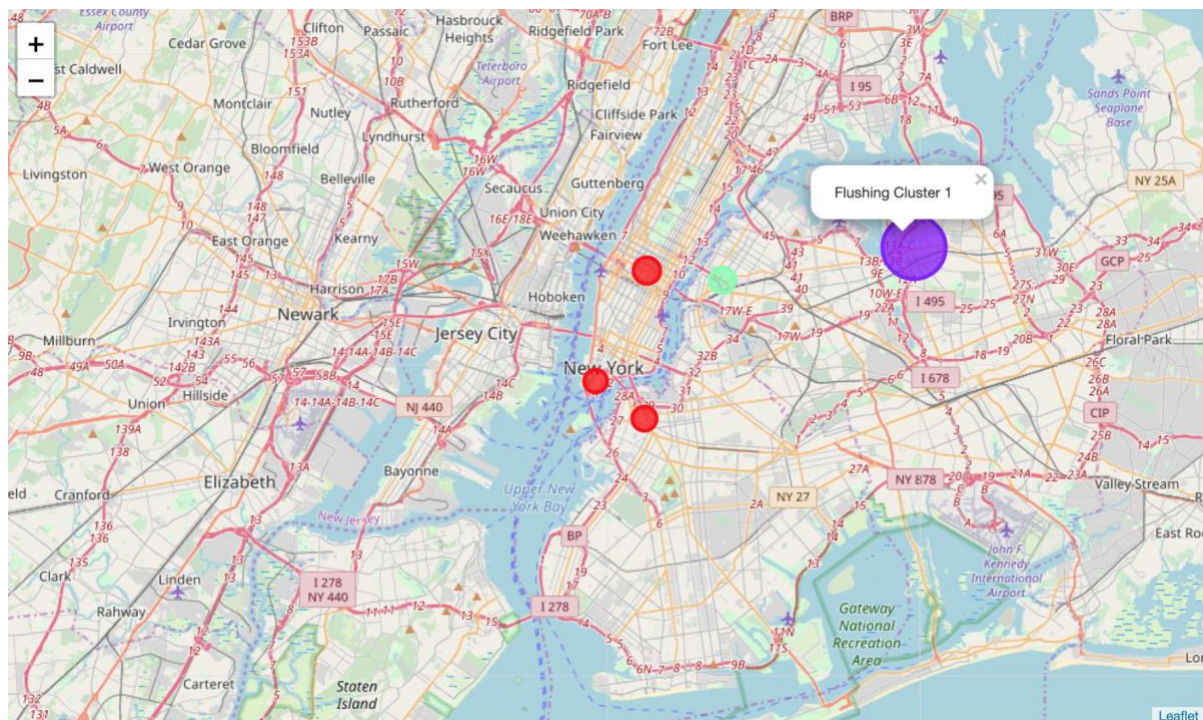
In the leaflet map illustrated above, the radius of each circle is proportional to the number of restaurants in the corresponding district. Flushing district located in the northern part of Queens borough, scored the highest number of restaurants (52), and the Financial District at the lower part on Manhattan has the lowest number of restaurants (19).

For the most common venues, these clusters make sense as Midtown Manhattan, Financial District and Downtown Brooklyn are dominated by coffee shops, gyms/ fattiness centers and sandwich place, fall under the red cluster. Whereas Flushing district dominated by Chinese restaurants stands alone under the purple circle. Lastly, Long Island district dominated by hotels is also in a separated cluster (green circle).

4. Results

The result of the analysis can be summarized in the following points:

- Coffee shops tops the charts in the five districts as the most frequently occurring venue type.
- Flushing district located in the northern part of Queens borough, scored the highest number of restaurants (52). Whereas the Financial District at the lower part on Manhattan has the lowest number of restaurants (19).
- Midtown Manhattan, Financial District in lower Manhattan and Downtown Brooklyn are dominated by coffee shops, gyms/ fattiness centers and sandwich places as the most common venues.
- Flushing district dominated by Chinese restaurants as the most common venue.
- Long Island district in the borough of Queens is dominated by hotels.
- As the clustering was based only on the most common venue category, Midtown Manhattan, Financial District and Downtown Brooklyn fall under the same cluster. Whereas Flushing district stands alone in another cluster. Long Island is also in a separated cluster.

5. Discussion

The pitfall of this analysis is that clustering is solely based on the most common venues obtained from Foursquare data. Land price, distance, number of potential customers, ingredients suppliers and staffing cost are some of the factors that could play a major role on the final decision. Hence, this analysis can't be considered conclusory. However, it

definitely provides some very important preliminary information on the prospect of opening a lunch restaurant around the central business districts of New York City.

According to the analysis, the Flushing district in the borough of Queens has the highest number of restaurants among the five districts, and hence it will be highly competitive to start a restaurant business in this district. The remaining districts, Midtown Manhattan, Financial district, Long Island City and Downtown Brooklyn have quite similar number of restaurants (22, 19, 20 and 21 respectively). **Downtown Brooklyn** will be the most suitable district for an upcoming lunch restaurant as Coffee shop is the most common venue type and the frequency of restaurants is very low compared to the remaining districts.

 A quick google search would also reveal that the cost of spaces in Manhattan are comparatively high, whereas Downtown Brooklyn has lower cost of rents. Therefore, Downtown Brooklyn could be a good candidate to open a decent lunch restaurant, targeting the large pool of customers (office workers and college students).

6. Conclusion

In conclusion, the capstone project helped me to get some glimpse of how real-life Data Science projects look like. In this project I have made use of several methodologies learned during the Data Science Specialization such as data wrangling with pandas, basic data visualization, machine learning techniques and the Foursquare API.

For further analysis, different factors should be considered such as average price per square foot, distance of the venues from the closest station, number of potential customers and ingredients suppliers. Other clustering techniques can also be used to obtain better results.