Analysing the optimal neighbourhood to open a new restaurant in the Manhattan Borough of New York using Foursquare and New York Location Data

1.0 Introduction

For a large-scale restaurant chain or business, the location of a new restaurant can be very important, where a restaurant is located may determine how many people can access and visit it and therefore, the revenue a restaurant can generate. The number of competing restaurants in an area may also be a determining factor of how successful a restaurant will be. My project intends to provide important information to stakeholders who may be looking to open a new restaurant in the Manhattan Borough of New York City.

Using Foursquare location data and New York City postcode data I will be investigating the ideal neighbourhood to open a new restaurant in Manhattan. To assess this problem, I will be using segmentation and clustering techniques to Identify help which neighbourhood would be a good location for a new restaurant. I will be assessing the frequency of restaurants in each neighbourhood in a given radius and I aim to find a neighbourhood which has a lower frequency of restaurant venues and therefore, could be an excellent location for a new restaurant.

1.1 Data

As outlined in the introduction section of my project, the data that will be used for my project will be New York location data, including information for boroughs, neighbourhoods and Latitude/Longitude. Initially the New York data must be pre-processed to be useable in my project. In addition to this, I will require the use of Foursquare API to link my New York data set to obtain venue information. The New York data set is freely accessible and has been converted to a JSON file. The Foursquare data is obtained by the creation of a Foursquare developer account which will allow me to make a several calls to the API to obtain venue information for the Manhattan Borough of New York.

1.2 Methodology

To conduct an investigation into which neighbourhoods in Manhattan would be favourable for a restaurant to open I had to process and analyse the New York neighbourhood and borough data that I had. Initially, I had to import all relevant packages to allow me to analyse the data, and following this, data had to be transformed to a data frame that could be used and analysed. I then had to extract the neighbourhoods that were part of the Manhattan borough as these would be used for my analysis (Figure 1.0).

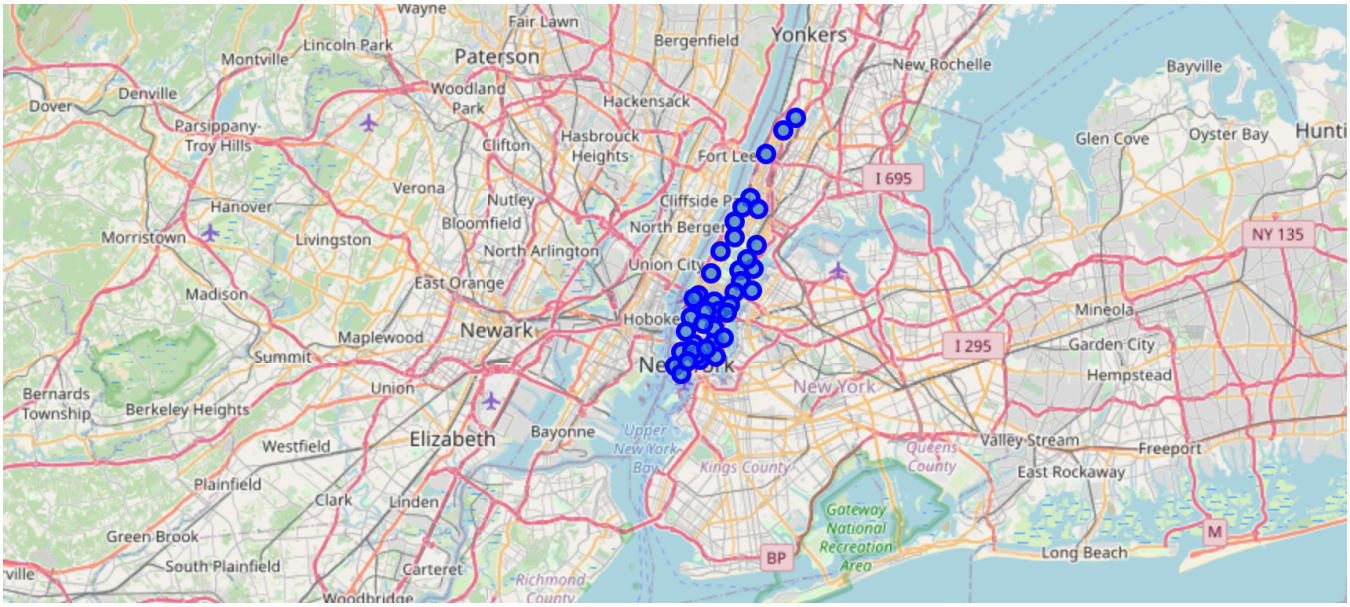

Figure 1.0 – Folium Map showing all neighbourhoods in Manhattan, New York

The next step was to link Foursquare to my New York data set to allow me to obtain information on the venues in New York and more specifically, Manhattan. This was achieved through linking my Foursquare credentials and would enable me to make calls to the API to obtain information about venues in neighbourhoods in Manhattan. Through several steps, I was able to segment the data to obtain the types and frequency of venues with a 500-metre radius of each neighbourhood (Figure 1.1). Once the frequency of each type of venue had been determined this was presented in a data frame which showed the mean frequency of each type of venue in the set radius. This then enabled me to determine the top 10 most common venues for each neighbourhood by the frequency that they had occurred (Figure 1.1). I then looked to determine which neighbourhoods did not contain a restaurant in the top 10 most common venues and could therefore be good locations for a new restaurant to open, a new data frame was created which contained the neighbourhoods that did not have a restaurant present (Figure 1.3).

```
nhattan_grouped
```

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium	Australian Restaurant	Austrian Restaurant
0	Battery Park City	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.013699	0.013699	0.00	0.0
1	Carnegie Hill	0.000000	0.000000	0.00	0.000000	0.011236	0.00	0.000000	0.011236	0.000000	0.022472	0.000000	0.000000	0.000000	0.000000	0.00	0.0
2	Central Harlem	0.000000	0.000000	0.00	0.068182	0.045455	0.00	0.000000	0.000000	0.045455	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0
3	Chelsea	0.000000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.050000	0.000000	0.000000	0.010000	0.000000	0.000000	0.00	0.0
4	Chinatown	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000	0.00	0.0
5	Civic Center	0.000000	0.000000	0.00	0.000000	0.020000	0.01	0.000000	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.01	0.0
6	Clinton	0.000000	0.000000	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0
7	East Harlem	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0
8	East Village	0.000000	0.000000	0.00	0.000000	0.010000	0.00	0.000000	0.010000	0.010000	0.000000	0.010000	0.000000	0.000000	0.000000	0.00	0.0
9	Financial District	0.000000	0.000000	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0
10	Flatiron	0.000000	0.000000	0.00	0.000000	0.050000	0.00	0.000000	0.000000	0.010000	0.000000	0.010000	0.000000	0.000000	0.000000	0.00	0.0

Figure 1.1 – Data frame showing the mean frequency of venues by neighbourhood

	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	Battery Park City	Coffee Shop	Hotel	Park	Gym	Clothing Store	Memorial Site	Playground	Sandwich Place	Gourmet Shop	Plaza
1	Carnegie Hill	Coffee Shop	Café	Bookstore	French Restaurant	Cosmetics Shop	Yoga Studio	Gym	Wine Shop	Pizza Place	Japanese Restaurant
2	Central Harlem	African Restaurant	French Restaurant	Seafood Restaurant	Bar	Chinese Restaurant	Art Gallery	Cosmetics Shop	American Restaurant	Fried Chicken Joint	Caribbean Restaurant
3	Chelsea	Coffee Shop	Bakery	Art Gallery	American Restaurant	French Restaurant	Seafood Restaurant	Ice Cream Shop	Wine Shop	Cupcake Shop	Cosmetics Shop
4	Chinatown	Bakery	Chinese Restaurant	Dessert Shop	Cocktail Bar	Hotpot Restaurant	Spa	American Restaurant	Bubble Tea Shop	Salon / Barbershop	Ice Cream Shop
5	Civic Center	Coffee Shop	Gym / Fitness Center	Spa	Cocktail Bar	Hotel	Yoga Studio	French Restaurant	Italian Restaurant	Bakery	Park
6	Clinton	Italian Restaurant	Gym / Fitness Center	Theater	American Restaurant	Sandwich Place	Coffee Shop	Wine Shop	Hotel	Cocktail Bar	Spa
7	East Harlem	Mexican Restaurant	Thai Restaurant	Bakery	Sandwich Place	Park	Deli / Bodega	Latin American Restaurant	Steakhouse	Cocktail Bar	Taco Place
8	East Village	Bar	Pizza Place	Mexican Restaurant	Italian Restaurant	Wine Bar	Korean Restaurant	Coffee Shop	Ice Cream Shop	Seafood Restaurant	Salon / Barbershop
9	Financial District	Coffee Shop	Italian Restaurant	Cocktail Bar	Bar	Pizza Place	Gym / Fitness Center	Gym	Salad Place	American Restaurant	Mexican Restaurant
10	Flatiron	Italian Restaurant	American Restaurant	Japanese Restaurant	Gym / Fitness Center	New American Restaurant	Wine Shop	Coffee Shop	Sporting Goods Shop	Spa	Furniture / Home Store
11	Gramercy	Bar	Italian Restaurant	Pizza Place	American Restaurant	Bagel Shop	Cocktail Bar	Playground	Wine Shop	Grocery Store	Mexican Restaurant
12	Greenwich Village	Italian Restaurant	Clothing Store	Sushi Restaurant	Café	Indian Restaurant	Gym	American Restaurant	Boutique	Bakery	Gourmet Shop
13	Hamilton Heights	Pizza Place	Coffee Shop	Deli / Bodega	Café	Mexican Restaurant	Yoga Studio	Sushi Restaurant	Liquor Store	Caribbean Restaurant	Latin American Restaurant

Figure 1.2 – Data frame showing the top 10 venues by neighbourhood

1.3 Results

The results of the analysis of New York data showed that 40 neighbourhoods were present in Manhattan and linking these neighbourhoods (listed in figure 1.2) to venue data using the Foursquare API produced a data frame containing the top 10 most common venues for each of these listed neighbourhoods. 38 Neighbourhoods in total contained a restaurant in the top 10 most common venues within a 500-metre radius, leaving 2 neighbourhoods which did not contain a restaurant within the top 10 most common venues. The neighbourhoods containing no restaurant were: Battery Park City and Stuyvesant Town (Figure 1.3).

	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	Battery Park City	Coffee Shop	Hotel	Park	Gym	Clothing Store	Memorial Site	Playground	Sandwich Place	Gourmet Shop	Plaza
1	Stuyvesant Town	Park	Bar	Coffee Shop	Gym / Fitness Center	Farmers Market	Playground	Fountain	Harbor / Marina	Gas Station	Cocktail Bar

Figure 1.3 – Data frame Identifying neighbourhoods not containing a restaurant venue in the top 10 most common venues.

1.4 Discussion

The results of this project showed two potential locations for a new restaurant in the Manhattan area of New York City, based on a restaurant not being a top 10 most common venue for these neighbourhoods. As outlined in the results section, these neighbourhoods were: Battery Park City and Stuyvesant Town. I believe these to be favourable locations for a future restaurant to be opened based on the analysis conducted. It is likely that a restaurant built in one of these neighbourhoods is likely to face less competition from other restaurant venues and potentially higher popularity due to the lack of this particular service in these neighbourhoods compared with others. However, further analysis could be conducted using a wider radius to provide a more detailed picture of these neighbourhoods. In addition to this, further analysis could also be conducted to determine what type of restaurant is most popular in Manhattan and is therefore likely to thrive in a new neighbourhood.

1.5 Conclusion

In this study I analysed which the neighbourhoods in Manhattan that would be a good location for a new restaurant. Using New York location data, Foursquare API and the use of segmentation and clustering techniques I was able to extract, analyse and group venue data to determine the frequency of different types of venues in 40 neighbourhoods in Manhattan. The study provides useful information that could be used by developers/businesses looking to open a new restaurant in Manhattan.