

How to Select Good Location When Open a Restaurant in Manhattan

By Guangqi Chen

June 8, 2019

1. Introduction

1.1 Background

Location data becomes more and more important in any business.

For the City of Manhattan, New York, if someone is looking to open a restaurant, where would you recommend that they open it?

The traditional ways are either to use Yellow Page to search for all restaurant addresses or use Yelp to get general idea how many restaurants in a small area (neighbourhood) or use Google Map to visualize where are those restaurants.

Obviously, the above approaches are time consuming and limited to small areas without seeing the whole picture.

Therefore, it is advantageous for business owners to use location data to view and know how many restaurants, where are they. How do they locate?

1.2 Problem

Because of time and resource limit, Data that might contribute to determining potential include all neighbourhood names, locations, populations; all Manhattan Subway locations; all restaurant locations.

1.3. Interest

Obviously, business owners related to food industry would be very interested in how location data to help them for competitive advantage and business values. Others who are interested in location data are real estate realtors and banks who will give business loan to those business.

2.1 Data Sources

Neighbourhood data for City of Manhattan:

Data Source: https://cocl.us/new_york_dataset

Need to clean data to get Manhattan's all of Neighbourhood's name, Latitude and Longitude information.

Here I will use Course supplied Python Codes to get those information. So, all credits are given to IBM course instructors.

Restaurant data for City of Manhattan:

Use Neighbourhood data above with Foursquare API to get all those venues in each Neighbourhood.

Need to clean data to get all of venues related to restaurant.

Neighbourhood Population data for City of Manhattan:

Data Source: <https://www.worldatlas.com/articles/manhattan-neighborhoods-by-population.html>

Data scraping is used

Manhattan subway station location:

Data Source: <http://web.mta.info/developers/data/nyct/subway/Stations.csv>

Data scraping is used.

This data set will be used to visualize if some Restaurants are close to Subway stations!

2.2 Data Cleaning

How to get Restaurant data?

There are total of 294 unique categories returned from Foursquare Venue Category.

Key words search is used to decide if Venue is related to Restaurant or not. Those key words are: 'Pizza', 'Restaurant', 'Diner', 'Steakhouse', 'Sandwich', 'Noodle House', 'Bistro', 'Veterinarian', 'Taco Place', 'Burrito Place', 'Salad Place'.

Should "Pizza" be considered a restaurant or not?

I Googled and found there is no absolute answer and the answer depends the purpose. So I decided to put Pizza into the lists.

What are not considered restaurant?

Coffee, Donut, Food & Drink Shop, Cafeteria, Frozen Yogurt Shop, Dessert Shop, Bagel Shop, any category related to Bar, e.g. Sake Bar, Cocktail Bar, Snack Place, Juice Bar, Beer Bar.

To get more Venues from Foursquare API search, I set search radius to 2,000 meters. So, there are total of 5 venue locations are out of City of Manhattan.

How to get Neighbourhood Population data for City of Manhattan

Data is scraped from: www.worldatlas.com/articles/manhattan-neighborhoods-by-population.html

After the data is processed, it is found out that some neighbourhood names in this population data set are not existing in the New York's data set (https://cocl.us/new_york_dataset), for example, those neighbourhoods, Alphabet City, Koreatown and Kips Bay, are not existing in New York's data set and also there are some neighbourhoods in New York's data set are not existing in population data set, for examples, Carnegie Hill, Civic Center and Clinton.

One of the reasons for above not matched is some small neighbourhoods in the population data set are inside some neighbourhoods in New York's data set, e.g. Alphabet City is inside East Village, Kips Bay is inside Murray Hill and Koreatown is inside Midtown South.

Some manual updates are performed on population data set and then the data is aggregated again.

How to get subway locations for City of Manhattan?

Data is scraped from: <http://web.mta.info/developers/data/nyct/subway/Stations.csv>.

After the data is processed and visualized on the map, it was found that, the location data is for the whole New York Metro area. Data cleaning is also performed to have only Manhattan's data.

This data set will be used to visualize if some Restaurants are close to Subway stations!

2.3 Feature Selection

After data is cleaned, for restaurant data set, there are total of 1,291 samples with 76 features in the data.

For the neighbourhood population data set, there are total of 37 samples. As mentioned above, neighbourhood names in this population data set are not existing in the restaurant data set and vice versa. Some manual updates are performed on population data. Finally, there are several neighbourhoods in the restaurant data set are without population data.

For Subway location data set, there are total of 152 samples, which is nearly identical to reported in Wikipedia.

3.Data Analysis

3.1 Cluster Neighborhoods based on restaurant information

k-means Clustering is used for the restaurant information, so we have quick view which neighbourhoods are similar to each regarding the restaurant and also to see if there is any location pattern existed. Considering there are total of 40 neighbourhoods in Manhattan, 5 clusters are used in this study.

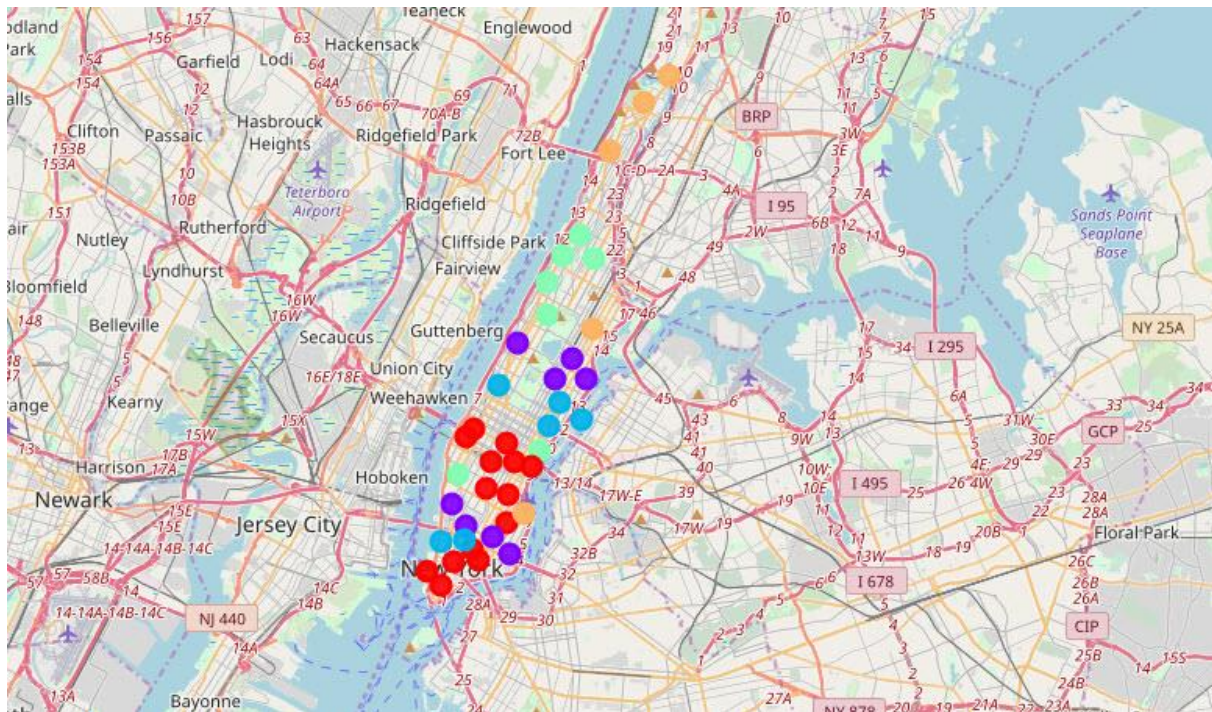


Figure 1. Manhattan Cluster Neighborhoods (restaurant information only) map

Very interesting is the second cluster, Italian restaurant is the number 1 venue in all those neighbourhoods.

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

| | Neighborhood | Common Venue | Common Venue | Common Venue | Common Venue | Common Venue | Common Venue | Common Venue | Common Venue |
|----|-------------------|--------------------|---------------------|-------------------------------|---------------------------------|-----------------------|-------------------------------|-------------------------------|---------------------|
| 8 | Upper East Side | Italian Restaurant | Thai Restaurant | American Restaurant | Mexican Restaurant | Spanish Restaurant | Seafood Restaurant | Vegetarian / Vegan Restaurant | Japanese Restaurant |
| 9 | Yorkville | Italian Restaurant | Pizza Place | Japanese Restaurant | Thai Restaurant | Vietnamese Restaurant | Sandwich Place | Seafood Restaurant | Sushi Restaurant |
| 12 | Upper West Side | Italian Restaurant | American Restaurant | Vegetarian / Vegan Restaurant | Seafood Restaurant | Indian Restaurant | Middle Eastern Restaurant | Chinese Restaurant | Ramen Restaurant |
| 18 | Greenwich Village | Italian Restaurant | Pizza Place | Sushi Restaurant | Seafood Restaurant | American Restaurant | Mediterranean Restaurant | Salad Place | Sandwich Place |
| 20 | Lower East Side | Italian Restaurant | Japanese Restaurant | Mexican Restaurant | Southern / Soul Food Restaurant | French Restaurant | Asian Restaurant | Mediterranean Restaurant | Greek Restaurant |
| 24 | West Village | Italian Restaurant | Pizza Place | American Restaurant | New American Restaurant | Chinese Restaurant | Seafood Restaurant | Salad Place | Steakhouse |
| 30 | Carnegie Hill | Italian Restaurant | Pizza Place | American Restaurant | French Restaurant | Japanese Restaurant | Vietnamese Restaurant | Seafood Restaurant | Asian Restaurant |
| 31 | Noho | Italian Restaurant | Japanese Restaurant | Pizza Place | Vietnamese Restaurant | Seafood Restaurant | Vegetarian / Vegan Restaurant | Thai Restaurant | Mexican Restaurant |

Figure 2: Top restaurants in each neighbourhood at second cluster

3.2 Relationship between neighbourhood populations and restaurant numbers

It is normal for people to think that neighbourhood populations and restaurant numbers.

But looking at the charts for neighbourhood populations and restaurant numbers, the relationship is not very strong.

```
df_pop.tail()
```

| | Rank | Neighborhood | Population |
|----|------|----------------------|------------|
| 32 | 32 | Flatiron District | 8,547 |
| 33 | 33 | NoLita | 5,713 |
| 34 | 34 | Kips Bay | 5,330 |
| 35 | 35 | Meatpacking District | 1,428 |
| 36 | 36 | Little Italy | 1,211 |

Figure 2: Bottom 5 neighbourhoods with population.

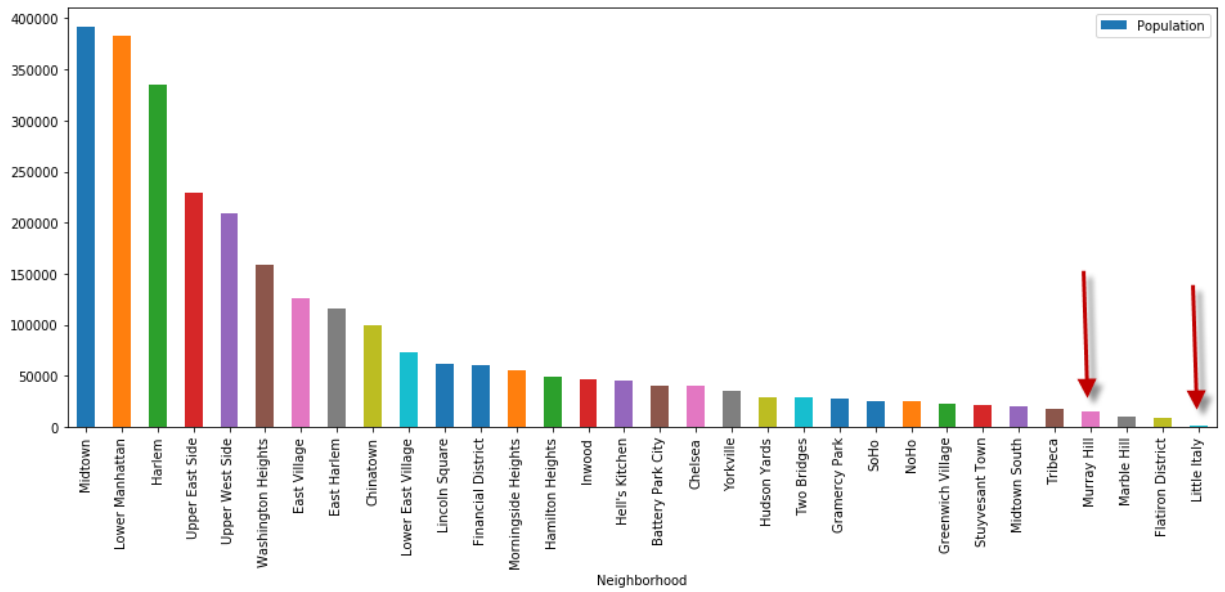


Figure 3: Neighbourhood population in each neighbourhood in Manhattan

Look at above, Little Italy and Murray Hill both have very few populations compared with Midtown's.

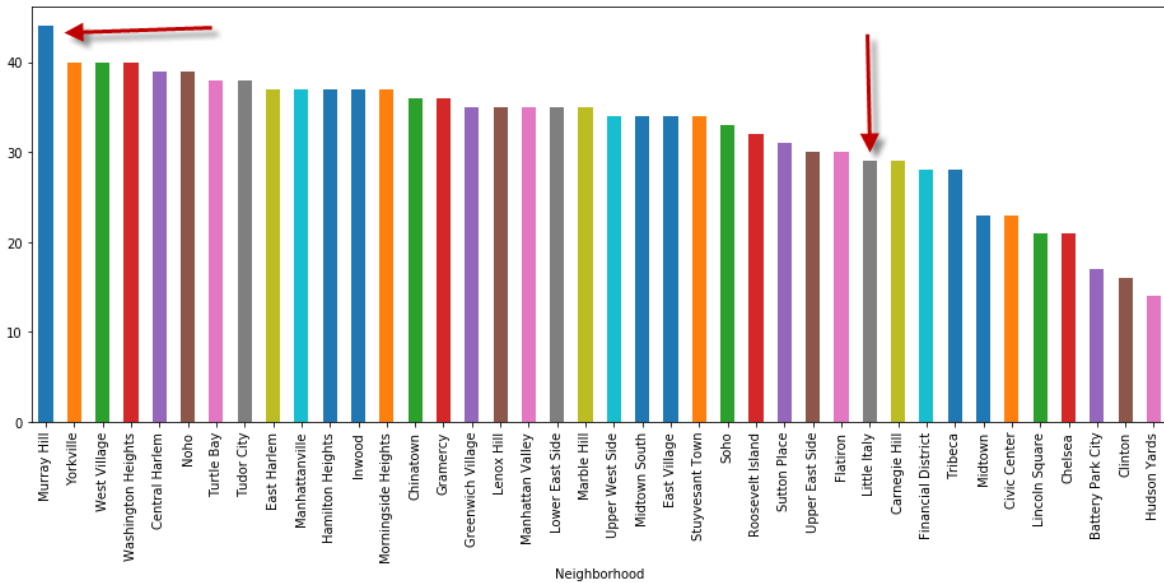


Figure 4: Restaurant numbers by each neighbourhood in Manhattan

4. Data Visualization and Discussion

4.1. What are most popular restaurants in Manhattan?

| Venue Category | |
|---------------------|-----|
| Italian Restaurant | 128 |
| Pizza Place | 100 |
| American Restaurant | 86 |
| Mexican Restaurant | 64 |
| Seafood Restaurant | 55 |

Figure 5: Top 5 restaurants in Manhattan

The top most popular restaurants in Manhattan are Italian Restaurant, Pizza, American Restaurant, Mexican Restaurant and Seafood Restaurant.

So, if you are planning to open restaurant in Manhattan and if you don't have a lot of experience, you may want to avoid those above restaurants since the competition is very high.

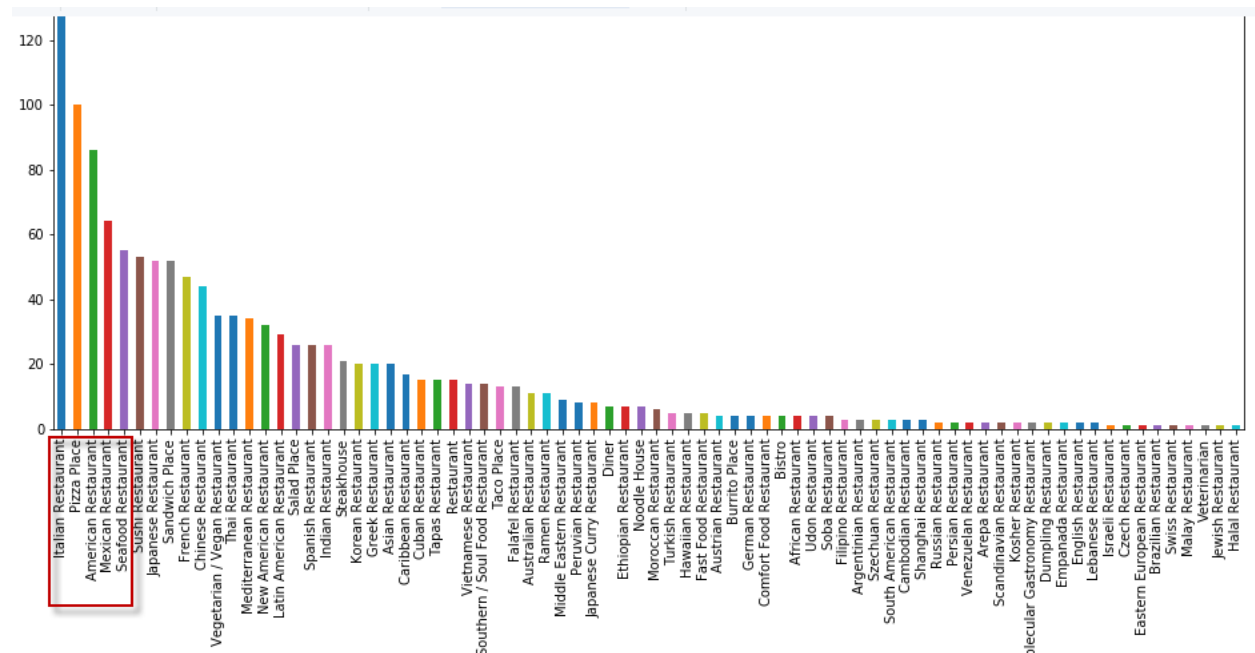


Figure 5: Restaurant number by category in Manhattan

4.2. Where are those restaurants in Manhattan by neighbourhood?

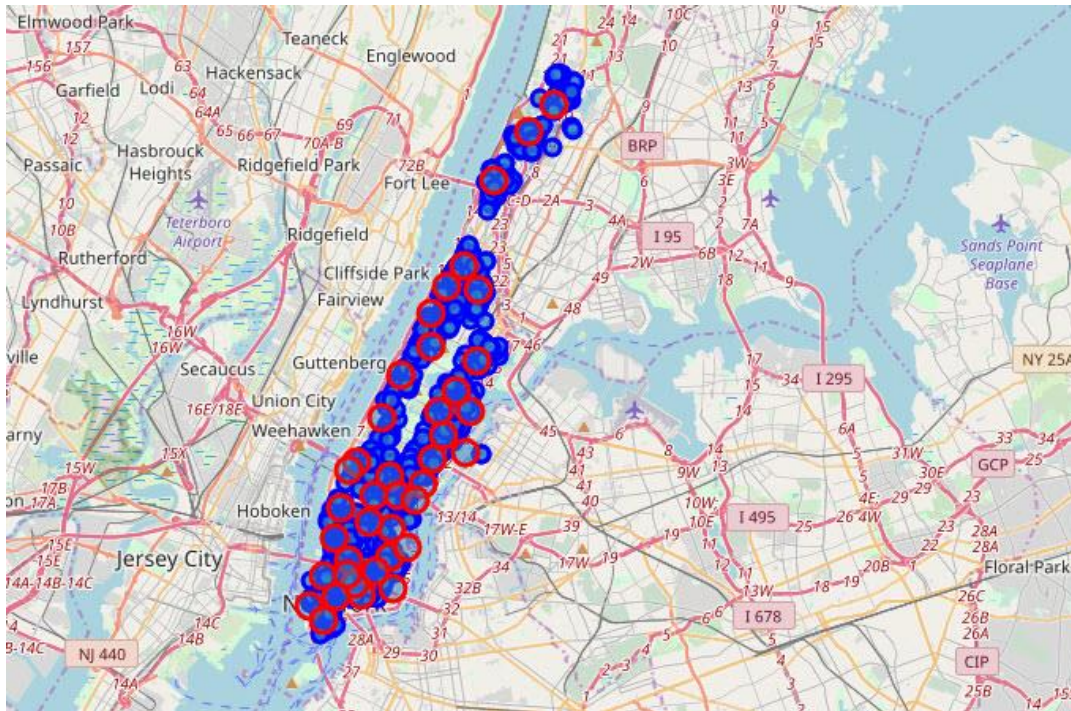


Figure 6: Restaurant locations (blue ones) and neighbourhood locations (red ones) in Manhattan

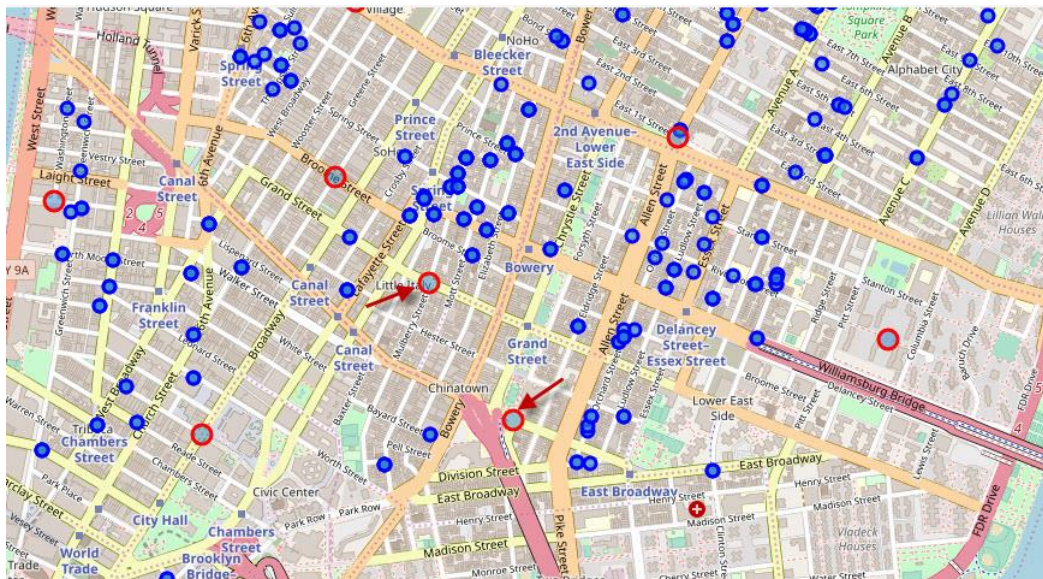


Figure 7: Restaurant location in Little Italy and Chinatown in Manhattan

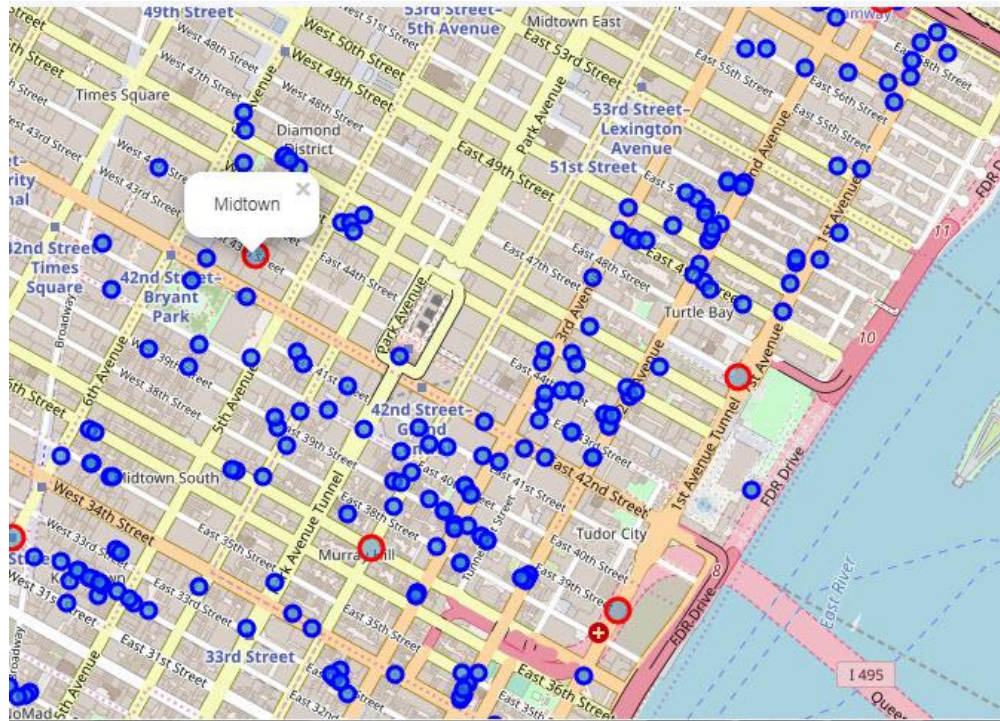


Figure 8: Restaurant location around Midtown and Murray in Manhattan

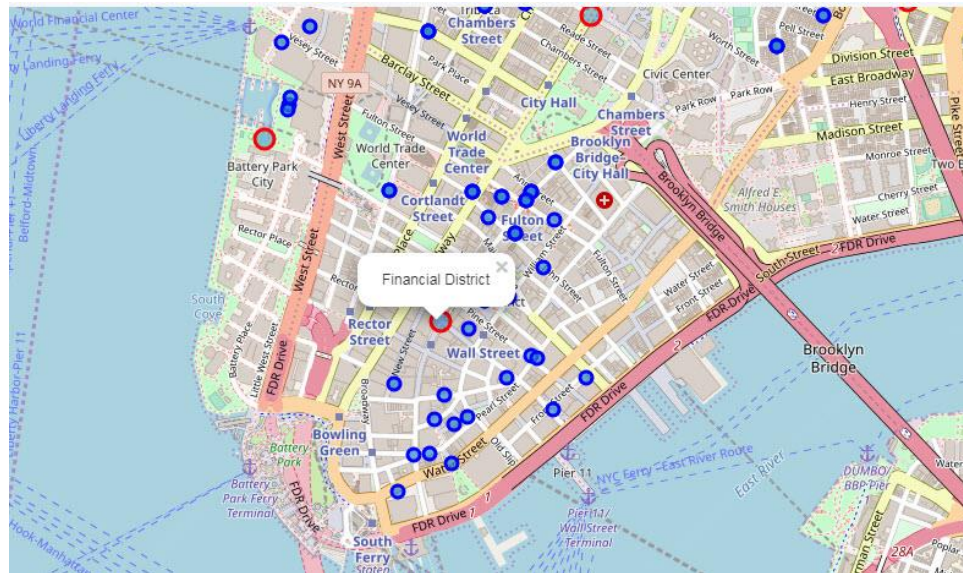


Figure 9: Restaurant location around Financial District in Manhattan

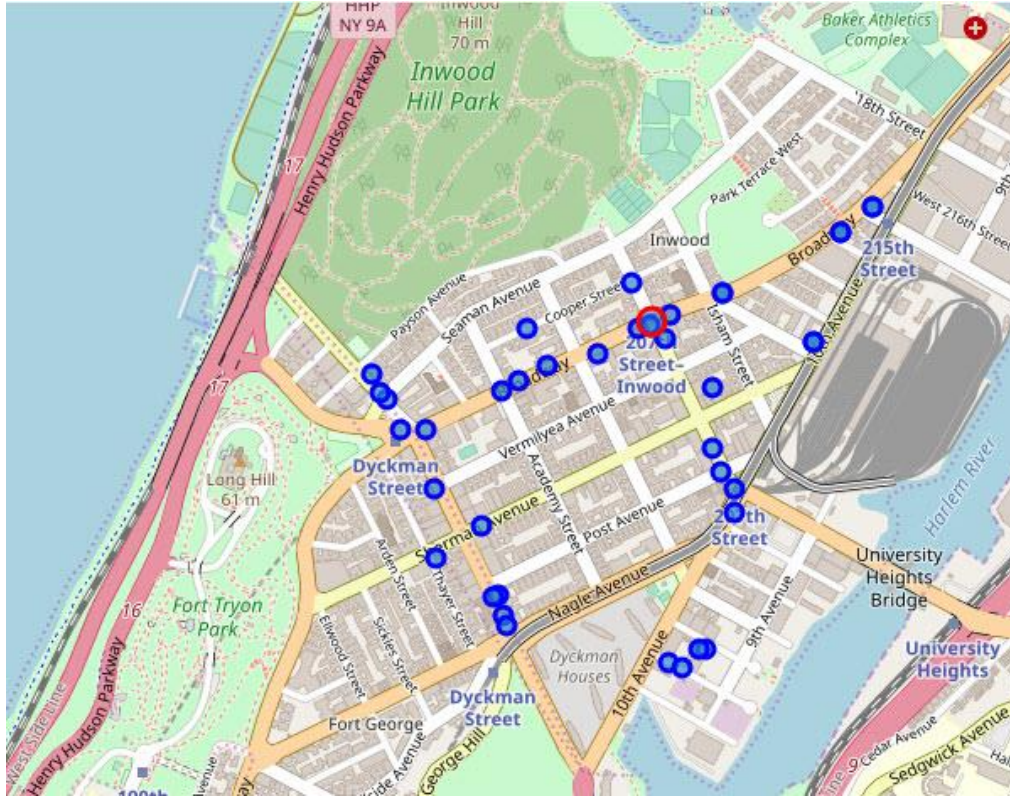


Figure 10: Restaurant location around Inwood in Manhattan

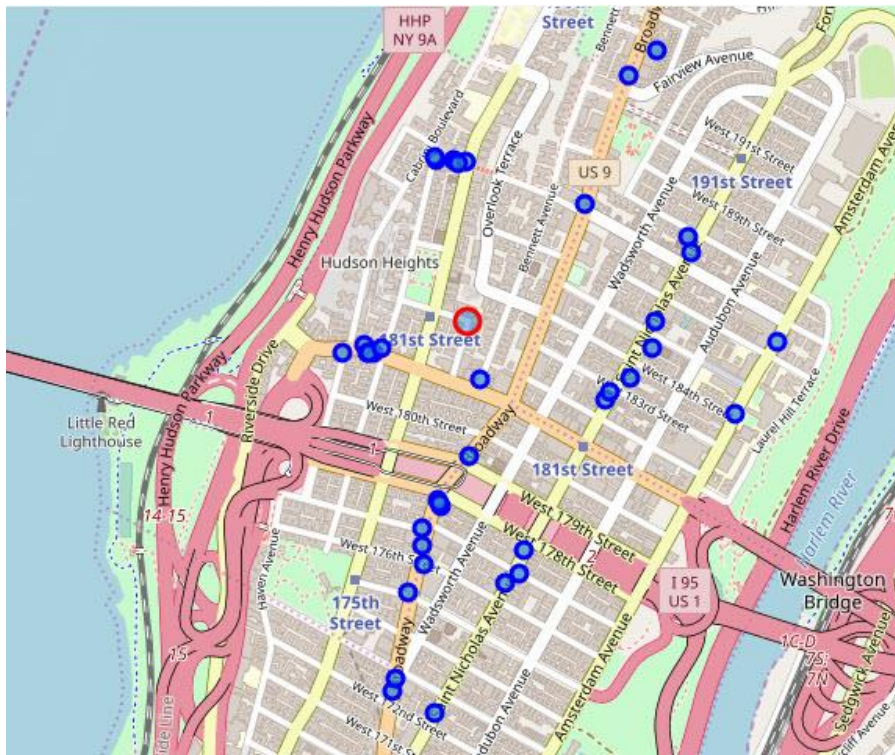


Figure 11: Restaurant location around Washington Heights in Manhattan

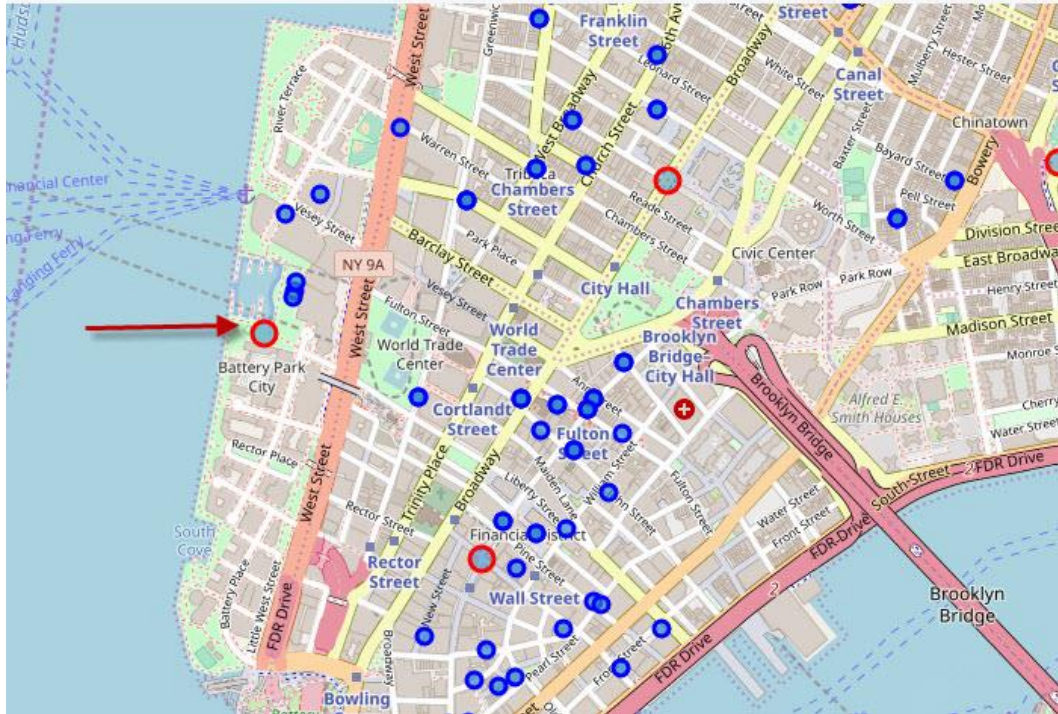


Figure 12: Restaurant location around Battery Park in Manhattan

4.3. Does the neighbourhood population contribute to Restaurant numbers in Manhattan?

This is the very hard question and it is not a YES and NO question.

From Section 3, it showed that there is no strong relationship between population numbers and restaurant numbers (e.g. Little Italy and Murray).

One possible reason to explain is Manhattan is a tourist City, where million of visitors come to Manhattan at all seasons. In this study, with limited resources, Hotel location is not used since FourSquare data has limited information about this.

But in general, population should make positive contribute to Restaurant numbers. For this study, the population will not be used as the factor to decide restaurant location.

4.4. Subway location and Restaurant Location

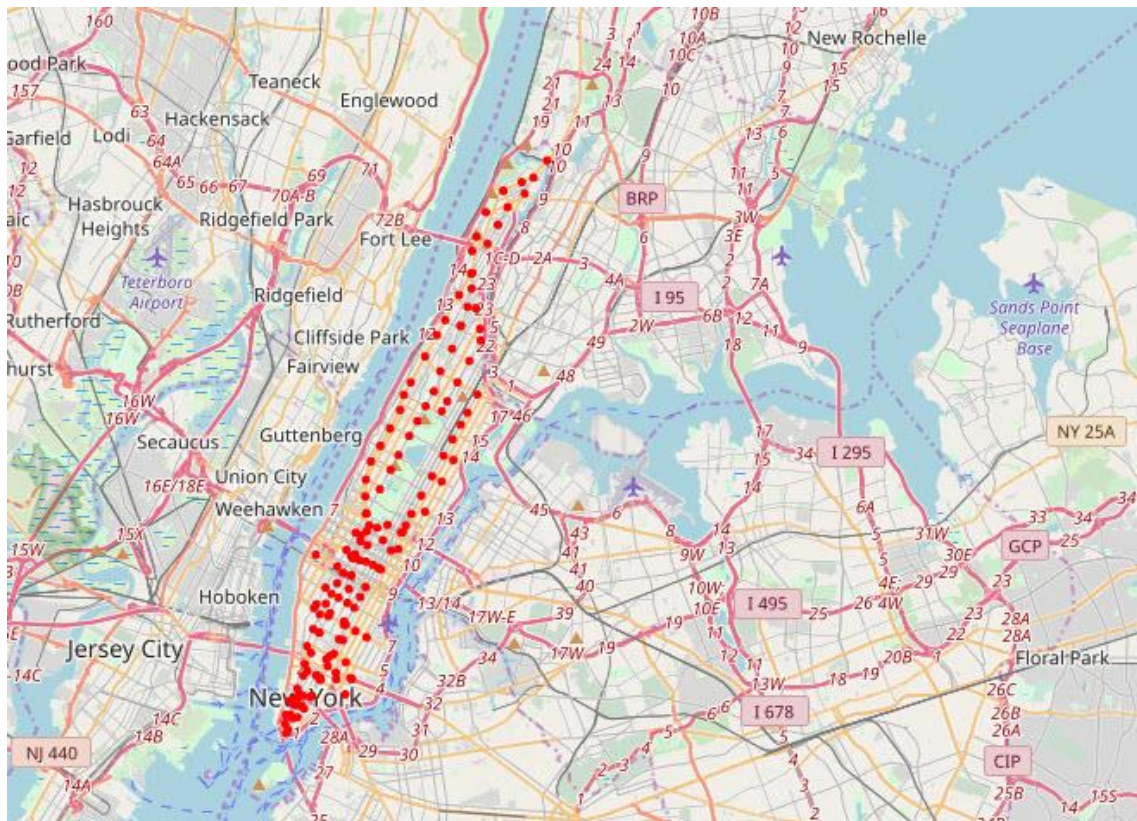


Figure 13: Subway locations (Red Colour) in Manhattan

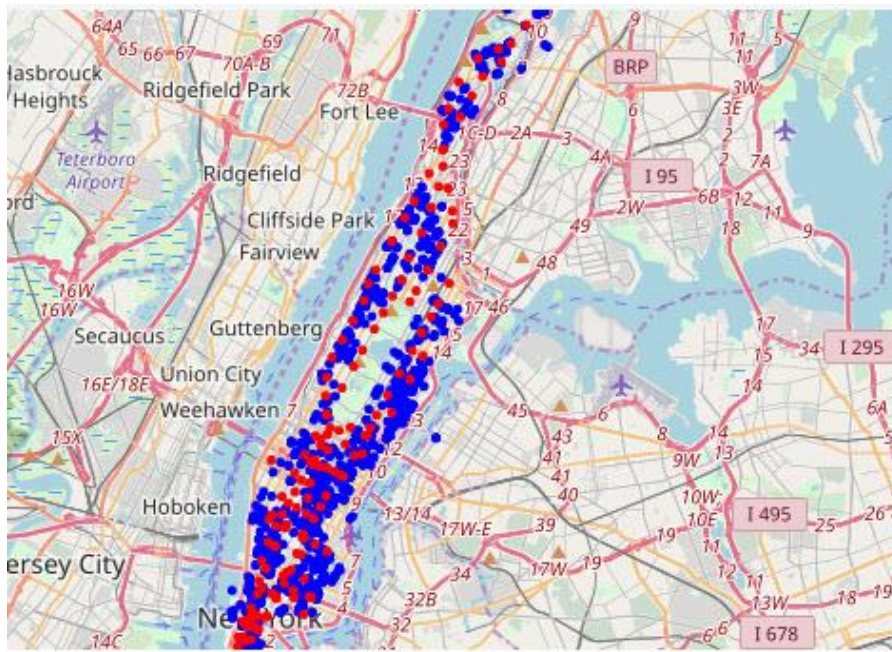


Figure 14: Subway locations (red ones) and Restaurant locations (blue ones) in Manhattan

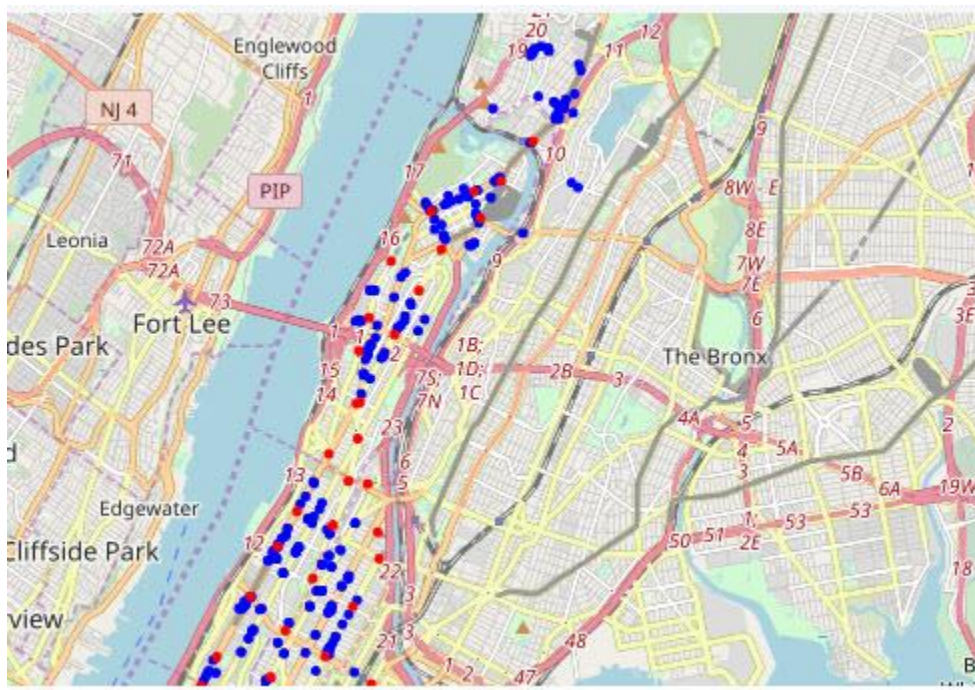


Figure 16: Subway locations and Restaurant location in Manhattan North

Observed from above maps, most of restaurants are clustered around subway stations. Which also makes senses since driving is not an easy job in Manhattan and most visitors will take subway in Manhattan. This is the key factor to consider when open a restaurant in Manhattan, walking distance from Subway.

4.5. Where to open the new restaurant?

From Figure 6 to Figure 11, we can see location patterns:

Restaurants are clustered close to each other

Restaurants are clustered around most of Subway Stations

Restaurants in Manhattan south are denser than in the north (so competition is high)

To find the good location, we need to find some areas where restaurants are not very close to each other in less dense area with Subway Stations close by.

Here are some potential areas after view the location maps:

Potential area 1: area in between Inwood and Washington Heights

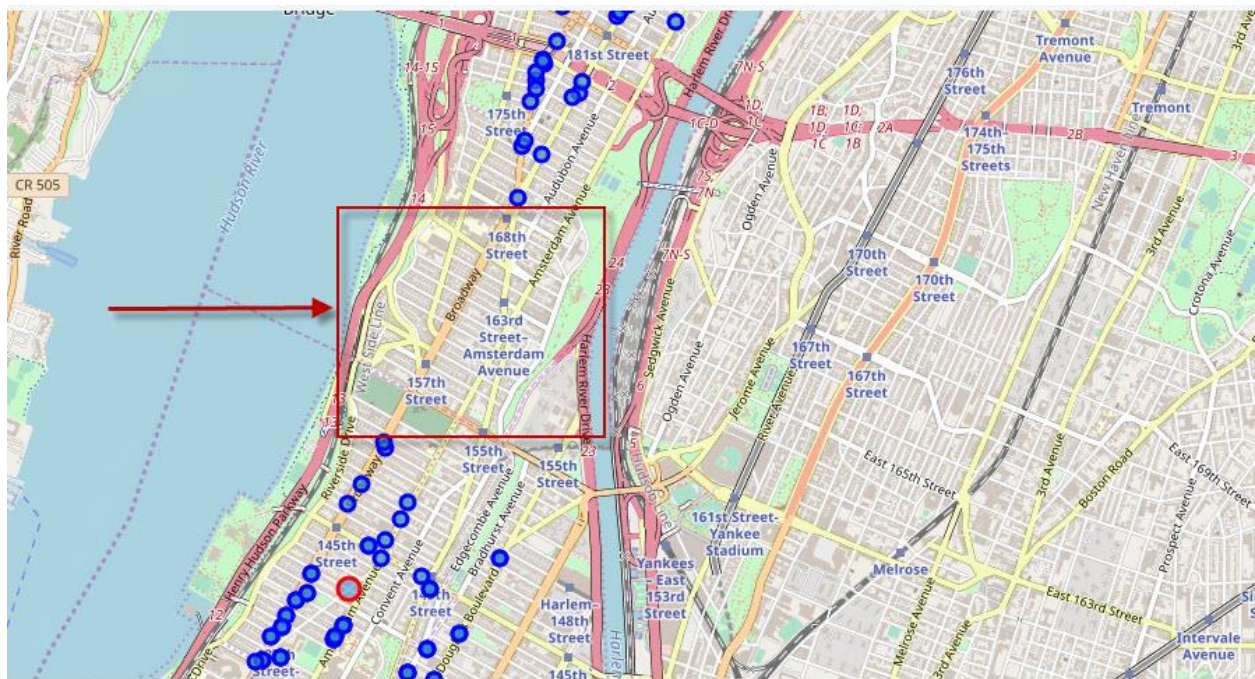


Figure 17: Potential area in between Inwood and Washington Heights

Shown from Figure 17, this area in between 165 street and 157 street, this area should be the best location in Manhattan. The reason is simple, there is no any restaurant in that area. So the business owner should pay attention to this area, ask City about Restaurant permit in that area.

Potential area 2: Marble Hill Area

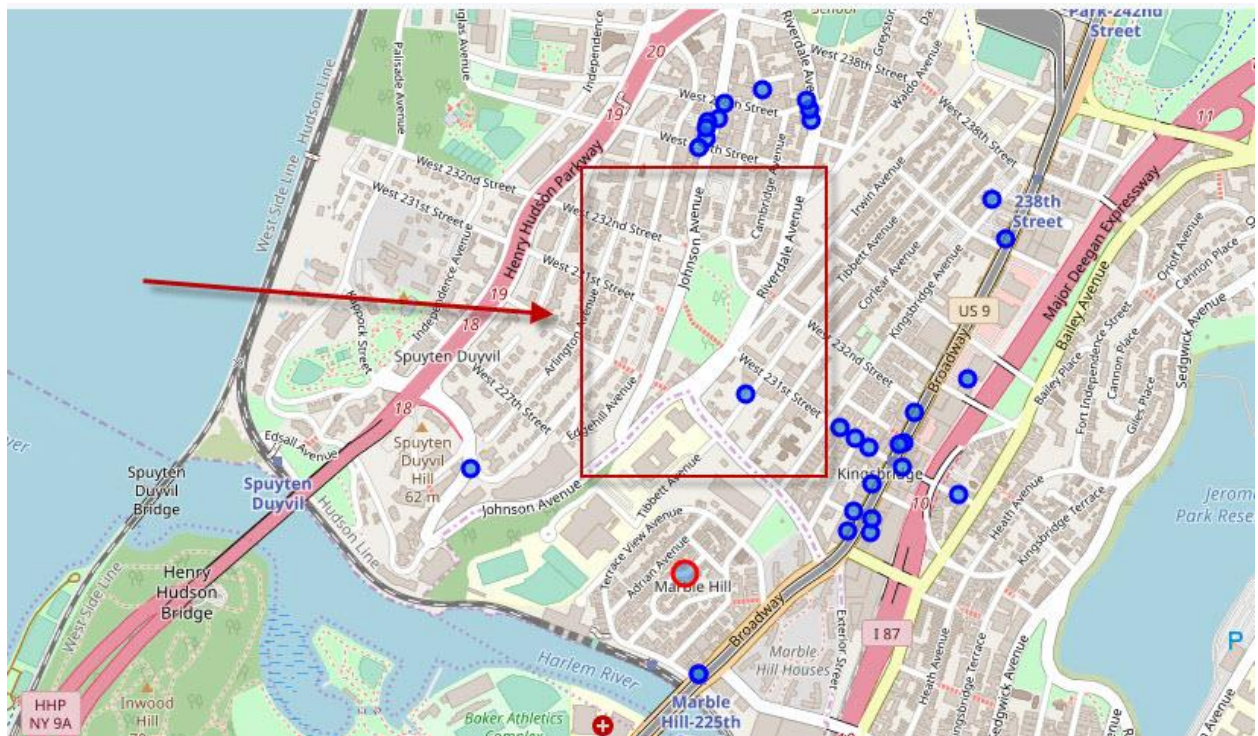


Figure 18: Potential area in Marble Hill Area

Shown from Figure 18, compared with other neighbourhoods, the competition in this Marble Hill area is not high, with limited restaurants. Which will give business owners an opportunity. So, ask City about Restaurant permit in this area.

Potential area 3: Area between Clinton and Lincoln Square



Figure 19: Potential Area between Clinton and Lincoln Square

Compared with first 2 areas, the competition in this area is very high, a lot of restaurants in this area. Still, there could be an opportunity for business owner.

Potential area 4: Area Manhattan South Area

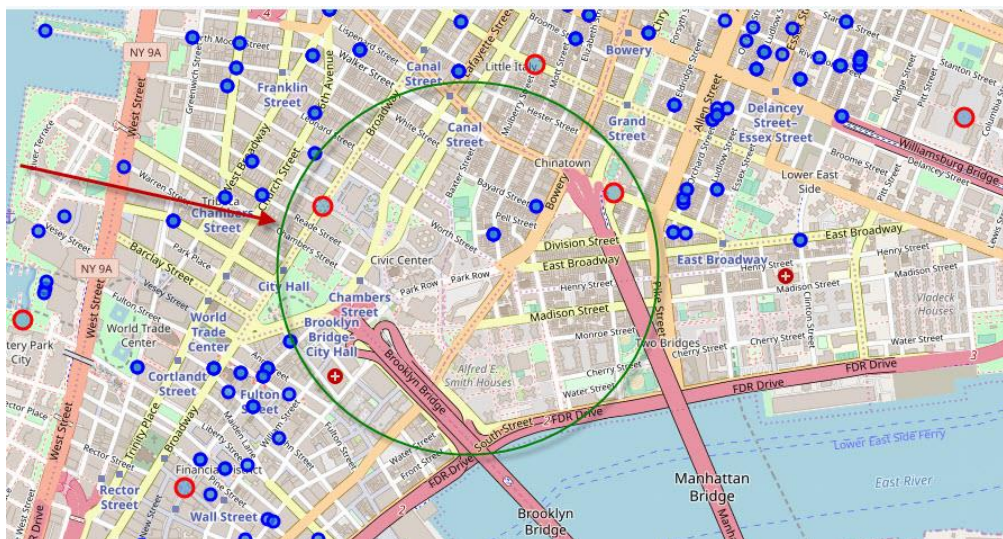


Figure 20: Potential area in Manhattan South Area

The competition in this area is extremely high, a lot of restaurants in this area. But if the business owners want to open in the South of Manhattan, this should be good area to look at since you can find some gaps between restaurants.

5. Conclusion

In this course project study, using location data from FourSquare API and other data is explored to help potential business owners and others to find good locations to open a restaurant in the City of Manhattan. k-means Clustering is also used for the restaurant information, so we have quick view which neighbourhoods are similar to each regarding the restaurant.

For those of business owners, there are two options:

Option 1: Less competition: Go to North

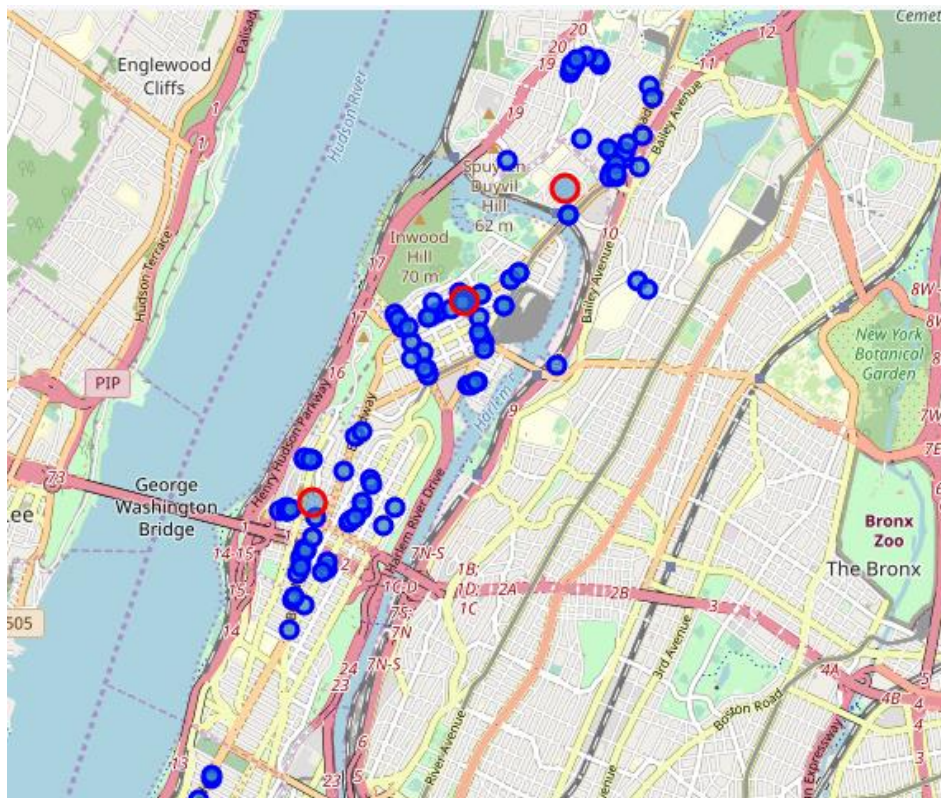


Figure 21: North Neighbourhoods have less competition than south


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

]:

| | 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 | Pizza Place | Mexican Restaurant | Sandwich Place | Latin American Restaurant | Spanish Restaurant | Diner | Japanese Restaurant | Thai Restaurant | Caribbean Restaurant | Restaurant |
| 2 | Washington Heights | Latin American Restaurant | Pizza Place | Mexican Restaurant | Tapas Restaurant | Sandwich Place | New American Restaurant | Spanish Restaurant | Chinese Restaurant | Restaurant | Seafood Restaurant |
| 3 | Inwood | Latin American Restaurant | Mexican Restaurant | Pizza Place | Spanish Restaurant | Sushi Restaurant | Chinese Restaurant | American Restaurant | Seafood Restaurant | Restaurant | Mediterranean Restaurant |
| 7 | East Harlem | Pizza Place | Mexican Restaurant | Thai Restaurant | Italian Restaurant | Latin American Restaurant | Taco Place | Moroccan Restaurant | Spanish Restaurant | American Restaurant | Sandwich Place |
| 37 | Stuyvesant Town | Pizza Place | Vietnamese Restaurant | Vegetarian / Vegan Restaurant | Italian Restaurant | Ramen Restaurant | Greek Restaurant | Mexican Restaurant | Chinese Restaurant | Eastern European Restaurant | Middle Eastern Restaurant |

Figure 22: Top Restaurants in each Neighbourhood in Cluster 5

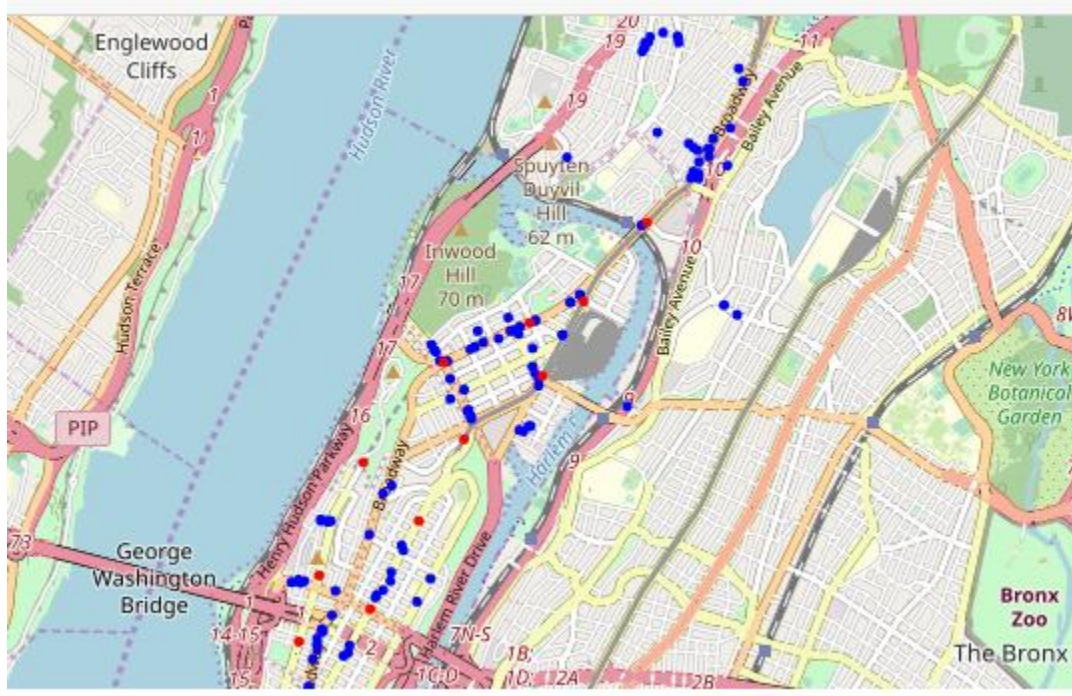
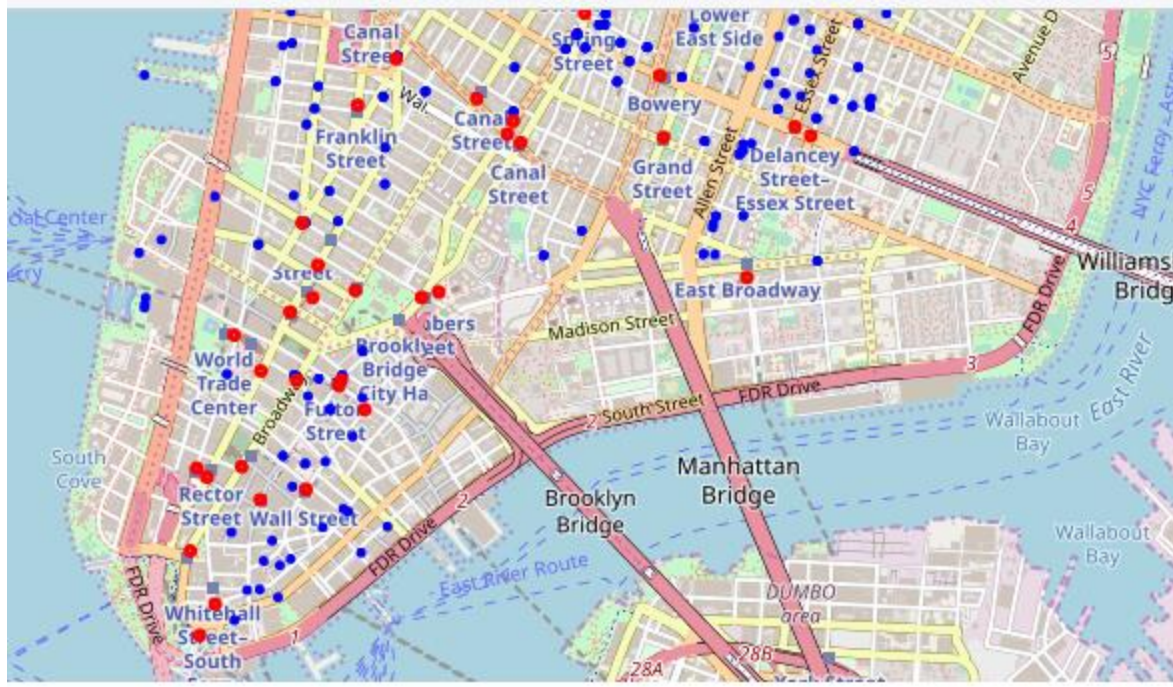


Figure 23: Subway locations and Current Restaurant locations in North Manhattan

Option 2: Strong Competition: Go to South but not too many areas



6. Future Direction:

Permit Information: This data will give us where are commercial zones, specifically, if the restaurant is allowed in any location.

Hotel Information: This data will show us where are most visitors stay

Restaurant Detail information: This data is about restaurant opening hours, ratings and others.