

Capstone Project – The Battle of Neighborhoods

Finding a better place to open a Chinese restaurant in New York City

Business Problem

- New York City has attracted a lot of tourists and immigrants every year because of its diversity and activity in business and culture. The international environment has also simultaneously influenced its food culture and cuisines.
- Chinese cuisine is becoming more and more popular in the United States that Chinese restaurants can be found almost everywhere, especially for the large cities like NYC. Opening a Chinese restaurant could bring a good business opportunity for success, however, cautions are needed to be taken when choosing the restaurant location in order to sustain a long-term success from the high competition.

Data Selection

- **New York City Neighborhood**

A dataset that includes all the 5 boroughs and neighborhoods as well as geographical location of each neighborhood is downloaded from the IBM server or from the official website

https://geo.nyu.edu/catalog/nyu_2451_34572

- **Foursquare API Data**

Information about all restaurants around every neighborhood within the radius of 1000 meters is collected from Foursquare, including the specified venue category, the longitude and latitude of the venue.

Methodology

- The data cleansing has been carried out on the dataset of NYC neighborhood to extract the necessary information as follow

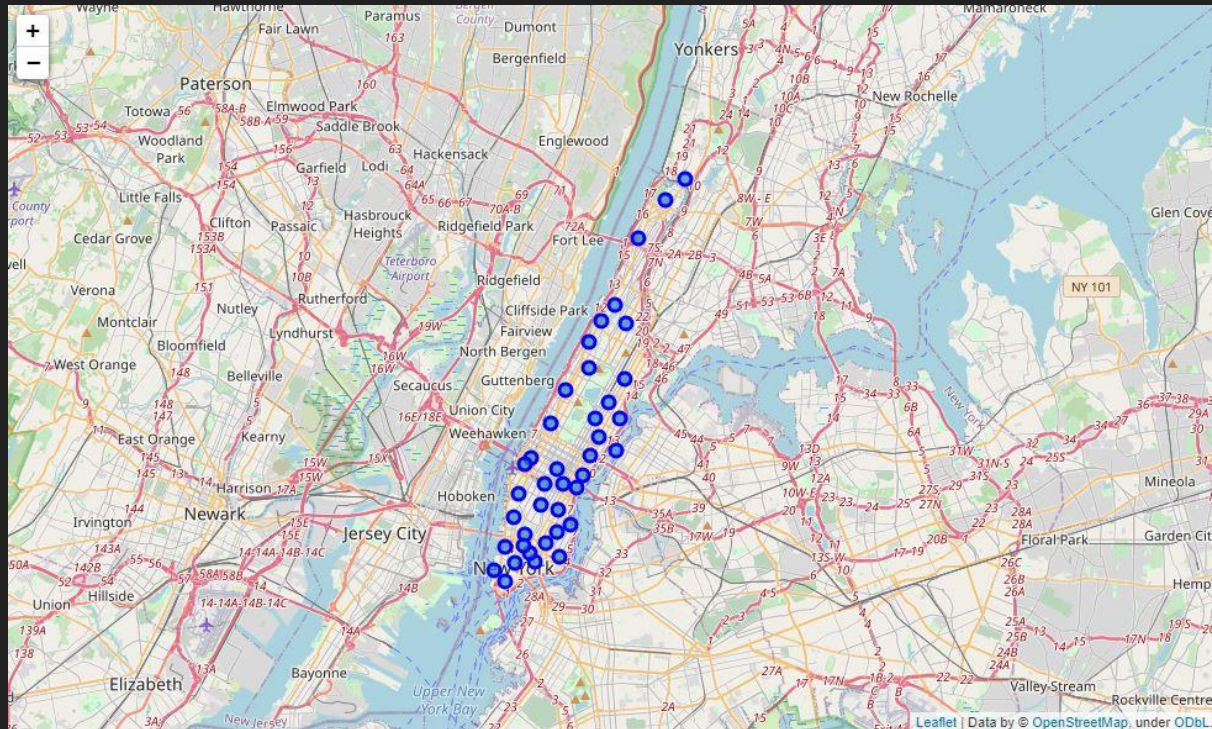
```
In [7]: neighborhoods.head()
```

```
Out[7]:
```

	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

Methodology

- The Foursquare API is applied to explore the neighborhoods in Manhattan, NYC.
- Folium and Geopy will be applied to create a map of NYC with marks for each neighborhood superimposed on top as shown below.



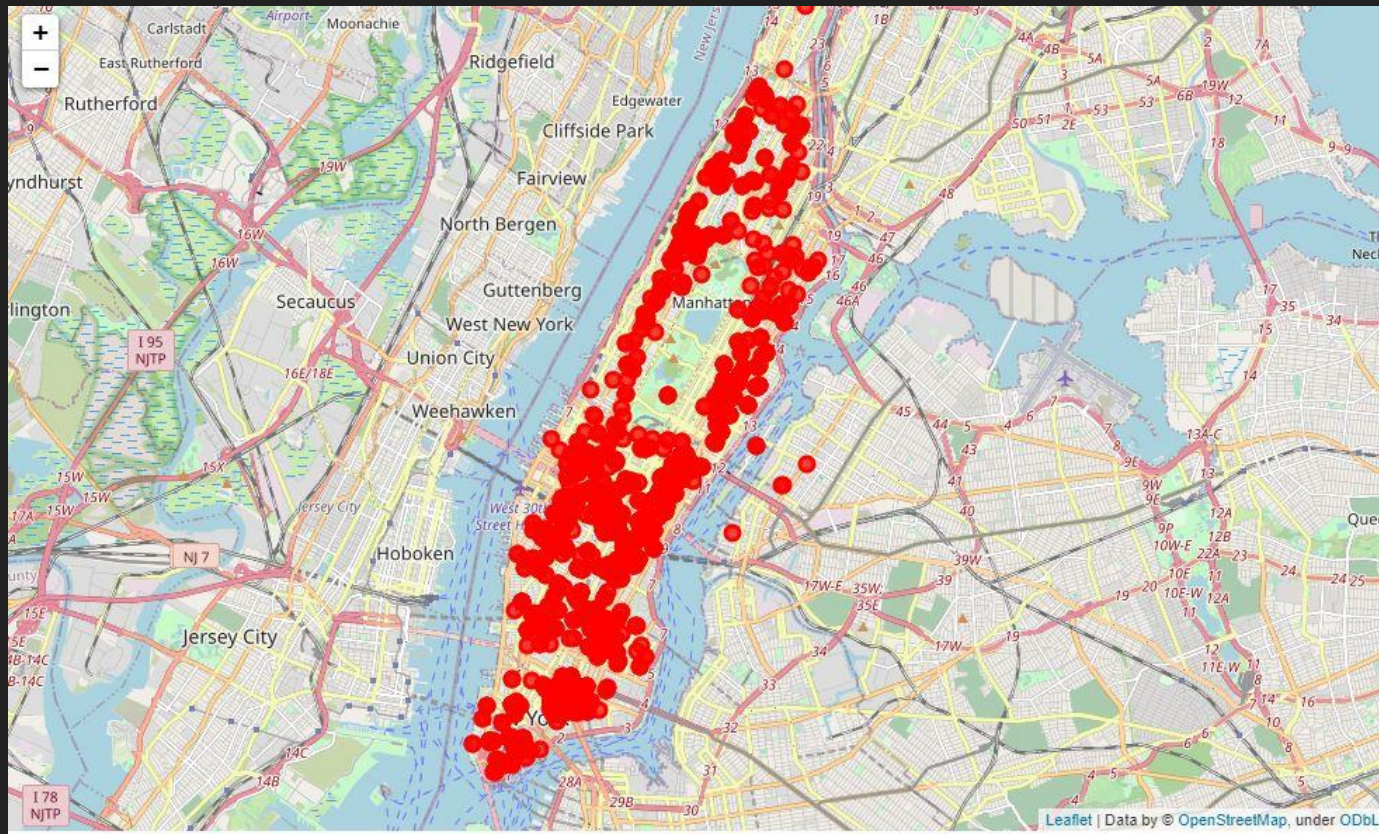
Methodology

- Data about Chinese cuisine in Manhattan is generated by Foursquare API by providing different Chinese restaurants that are near to a neighborhood within the radius of 1000 meters

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	China Wang	40.874347	-73.910540	Chinese Restaurant
1	Marble Hill	40.876551	-73.91066	Anise	40.885989	-73.910036	Chinese Restaurant
2	Marble Hill	40.876551	-73.91066	Mei Chung Mei Restaurant	40.878272	-73.902981	Chinese Restaurant
3	Marble Hill	40.876551	-73.91066	Pioneer Chinese Restaurant	40.879541	-73.905327	Chinese Restaurant
4	Marble Hill	40.876551	-73.91066	New Yung Hong Chinese Restaurant	40.873272	-73.905128	Chinese Restaurant

Methodology

- The different kinds of Chinese restaurants in Manhattan area is shown as below



Methodology

- One-hot Encoding has been used to find out what are the different kinds of Chinese restaurants present in each neighborhood and then calculate the top 10 Chinese restaurants based on the similarities.

	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	Chinese Restaurant	Food Truck	Xinjiang Restaurant	Street Food Gathering	Dim Sum Restaurant	Mexican Restaurant	Asian Restaurant	Shanghai Restaurant	Seafood Restaurant	Bakery
1	Carnegie Hill	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Indian Chinese Restaurant	Asian Restaurant	Thai Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng
2	Central Harlem	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Noodle House	Dessert Shop	Food Stand	Food Court	Dumpling Restaurant	Dim Sum Restaurant	Cuban Restaurant
3	Chelsea	Chinese Restaurant	Noodle House	Asian Restaurant	Dumpling Restaurant	Dim Sum Restaurant	Cuban Restaurant	Cantonese Restaurant	Bar	Cha Chaan Teng	Hotpot Restaurant
4	Chinatown	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Bakery	Cantonese Restaurant	Shanghai Restaurant	Vegetarian / Vegan Restaurant	Peking Duck Restaurant	Dessert Shop	Taiwanese Restaurant

Results

- The dataset has been clustered into 5 groups using K-Means clustering method, in which each member belongs to the cluster with the nearest mean. Each cluster has been expanded as follow
- Cluster 1

```
In [32]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[32]:

	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	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Asian Restaurant	Cantonese Restaurant	Noodle House	Bar	Bubble Tea Shop	Cha Chaan Teng	Hakka Restaurant
9	Yorkville	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Asian Restaurant	Noodle House	Indian Chinese Restaurant	Cantonese Restaurant	Food Stand	Food Court	Dumpling Restaurant
10	Lenox Hill	Chinese Restaurant	Szechuan Restaurant	Asian Restaurant	Dim Sum Restaurant	Cantonese Restaurant	Noodle House	Bar	Bakery	Bubble Tea Shop	Cha Chaan Teng
11	Roosevelt Island	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Asian Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng	Hotpot Restaurant	Cuban Restaurant
12	Upper West Side	Chinese Restaurant	Asian Restaurant	Sushi Restaurant	Dim Sum Restaurant	Cuban Restaurant	Peruvian Restaurant	Thai Restaurant	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng
14	Clinton	Chinese Restaurant	Szechuan Restaurant	Noodle House	Food Stand	Cuban Restaurant	Dim Sum Restaurant	Asian Restaurant	Cantonese Restaurant	Bakery	Cha Chaan Teng
17	Chelsea	Chinese Restaurant	Noodle House	Asian Restaurant	Dumpling Restaurant	Dim Sum Restaurant	Cuban Restaurant	Cantonese Restaurant	Bar	Cha Chaan Teng	Hotpot Restaurant
24	West Village	Chinese Restaurant	Noodle House	Vegetarian / Vegan Restaurant	Taiwanese Restaurant	Bar	Dim Sum Restaurant	Xinjiang Restaurant	Food Stand	Food Court	Dumpling Restaurant
25	Manhattan Valley	Chinese Restaurant	Asian Restaurant	Szechuan Restaurant	Peruvian Restaurant	Cuban Restaurant	Food Truck	Hotpot Restaurant	Hunan Restaurant	Sushi Restaurant	Seafood Restaurant
26	Morningside Heights	Chinese Restaurant	Szechuan Restaurant	Hunan Restaurant	Peruvian Restaurant	Food Truck	Hotpot Restaurant	Asian Restaurant	Shanghai Restaurant	Seafood Restaurant	Bakery
30	Carnegie Hill	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Indian Chinese Restaurant	Asian Restaurant	Thai Restaurant	Bar	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng
34	Sutton Place	Chinese Restaurant	Asian Restaurant	Dim Sum Restaurant	Hunan Restaurant	Szechuan Restaurant	Peking Duck Restaurant	Noodle House	Food Stand	Bakery	Bar
35	Turtle Bay	Chinese Restaurant	Szechuan Restaurant	Shanghai Restaurant	Asian Restaurant	Food Court	Taiwanese Restaurant	Sushi Restaurant	Noodle House	Dumpling Restaurant	Dim Sum Restaurant
36	Tudor City	Chinese Restaurant	Szechuan Restaurant	Shanghai Restaurant	Asian Restaurant	Food Court	Dim Sum Restaurant	Noodle House	Hakka Restaurant	Sushi Restaurant	Cantonese Restaurant
39	Hudson Yards	Chinese Restaurant	Noodle House	Cantonese Restaurant	Food Stand	Dim Sum Restaurant	Cuban Restaurant	Asian Restaurant	Szechuan Restaurant	Bakery	Cha Chaan Teng

Results

○ Cluster 2

```
In [33]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[33]:

	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	Dim Sum Restaurant	Dumpling Restaurant	Bakery	Cantonese Restaurant	Shanghai Restaurant	Vegetarian / Vegan Restaurant	Peking Duck Restaurant	Dessert Shop	Taiwanese Restaurant
18	Greenwich Village	Chinese Restaurant	Dumpling Restaurant	Bakery	Dim Sum Restaurant	Cantonese Restaurant	Hotpot Restaurant	Vegetarian / Vegan Restaurant	Noodle House	Food Stand	Cha Chaan Teng
20	Lower East Side	Chinese Restaurant	Dumpling Restaurant	Taiwanese Restaurant	Cantonese Restaurant	Bakery	Hotpot Restaurant	Dim Sum Restaurant	Noodle House	Vegetarian / Vegan Restaurant	Hakka Restaurant
22	Little Italy	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Bakery	Cantonese Restaurant	Shanghai Restaurant	Peking Duck Restaurant	Dessert Shop	Taiwanese Restaurant	Hotpot Restaurant
23	Soho	Chinese Restaurant	Dim Sum Restaurant	Bakery	Dumpling Restaurant	Shanghai Restaurant	Vegetarian / Vegan Restaurant	Peking Duck Restaurant	Dessert Shop	Taiwanese Restaurant	Hotpot Restaurant
31	Noho	Chinese Restaurant	Dumpling Restaurant	Dim Sum Restaurant	Cantonese Restaurant	Taiwanese Restaurant	Hotpot Restaurant	Shanghai Restaurant	Vegetarian / Vegan Restaurant	Bakery	Szechuan Restaurant
32	Civic Center	Chinese Restaurant	Dim Sum Restaurant	Bakery	Shanghai Restaurant	Cha Chaan Teng	Dumpling Restaurant	Cantonese Restaurant	Vegetarian / Vegan Restaurant	Peking Duck Restaurant	Dessert Shop

Results

○ Cluster 3

```
In [34]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[34]:

	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	Chinese Restaurant	Xinjiang Restaurant	Hotpot Restaurant	Bakery	Bar	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng	Cuban Restaurant	Dessert Shop
2	Washington Heights	Chinese Restaurant	Asian Restaurant	Hotpot Restaurant	Bakery	Bar	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng	Cuban Restaurant	Dessert Shop
3	Inwood	Chinese Restaurant	Dumpling Restaurant	Xinjiang Restaurant	Hotpot Restaurant	Bakery	Bar	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng	Cuban Restaurant
4	Hamilton Heights	Chinese Restaurant	Xinjiang Restaurant	Hotpot Restaurant	Bakery	Bar	Bubble Tea Shop	Cantonese Restaurant	Cha Chaan Teng	Cuban Restaurant	Dessert Shop
5	Manhattanville	Chinese Restaurant	Japanese Restaurant	Noodle House	Xinjiang Restaurant	Dessert Shop	Food Stand	Food Court	Dumpling Restaurant	Dim Sum Restaurant	Cuban Restaurant
6	Central Harlem	Chinese Restaurant	Asian Restaurant	Japanese Restaurant	Noodle House	Dessert Shop	Food Stand	Food Court	Dumpling Restaurant	Dim Sum Restaurant	Cuban Restaurant
7	East Harlem	Chinese Restaurant	Thai Restaurant	Cuban Restaurant	Food Truck	Food Stand	Food Court	Dumpling Restaurant	Dim Sum Restaurant	Dessert Shop	Xinjiang Restaurant

Results

○ Cluster 4

In [35]: `manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]`

Out[35]:

	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
13	Lincoln Square	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Shanghai Restaurant	Food Truck	Food Stand	Taiwanese Restaurant	Peking Duck Restaurant	Noodle House	Cuban Restaurant
15	Midtown	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Noodle House	Asian Restaurant	Shanghai Restaurant	Cuban Restaurant	Hunan Restaurant	Sushi Restaurant	Seafood Restaurant
16	Murray Hill	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Asian Restaurant	Noodle House	Dumpling Restaurant	Cuban Restaurant	Hakka Restaurant	Shanghai Restaurant	Sushi Restaurant
19	East Village	Chinese Restaurant	Taiwanese Restaurant	Dim Sum Restaurant	Asian Restaurant	Dumpling Restaurant	Vegetarian / Vegan Restaurant	Hotpot Restaurant	Hunan Restaurant	Szechuan Restaurant	Seafood Restaurant
21	Tribeca	Chinese Restaurant	Bakery	Cantonese Restaurant	Asian Restaurant	Street Food Gathering	Shanghai Restaurant	Cha Chaan Teng	Noodle House	Dessert Shop	Dim Sum Restaurant
27	Gramercy	Chinese Restaurant	Dim Sum Restaurant	Taiwanese Restaurant	Szechuan Restaurant	Dumpling Restaurant	Asian Restaurant	Hotpot Restaurant	Hakka Restaurant	Shanghai Restaurant	Sushi Restaurant
33	Midtown South	Chinese Restaurant	Szechuan Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Asian Restaurant	Shanghai Restaurant	Cuban Restaurant	Hakka Restaurant	Sushi Restaurant	Taiwanese Restaurant
37	Stuyvesant Town	Chinese Restaurant	Taiwanese Restaurant	Szechuan Restaurant	Asian Restaurant	Dim Sum Restaurant	Lounge	Dumpling Restaurant	Bakery	Bar	Bubble Tea Shop
38	Flatiron	Chinese Restaurant	Dim Sum Restaurant	Dumpling Restaurant	Noodle House	Asian Restaurant	Hakka Restaurant	Sushi Restaurant	Szechuan Restaurant	Taiwanese Restaurant	Bar

Results

○ Cluster 5

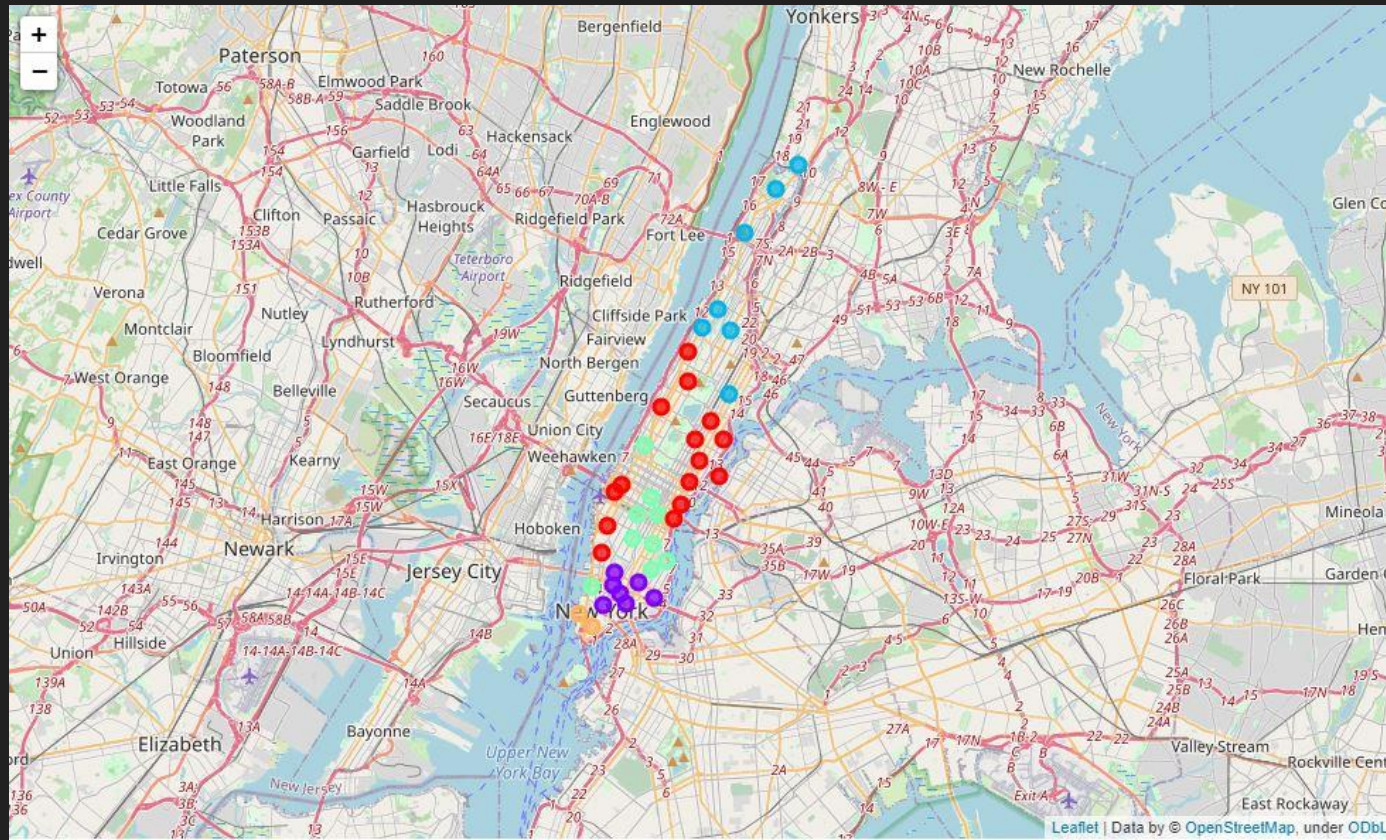
```
In [36]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[36]:

	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	Chinese Restaurant	Food Truck	Xinjiang Restaurant	Street Food Gathering	Dim Sum Restaurant	Mexican Restaurant	Asian Restaurant	Shanghai Restaurant	Seafood Restaurant	Bakery
29	Financial District	Chinese Restaurant	Food Truck	Xinjiang Restaurant	Seafood Restaurant	Dim Sum Restaurant	Mexican Restaurant	Asian Restaurant	Street Food Gathering	Shanghai Restaurant	Bakery

Results

- K-Means clustering algorithm has been applied to group the similar neighborhoods into clusters together based on the distribution of existing Chinese restaurants



Conclusions

- It is more preferable to open a new Chinese restaurant around northeast area in Manhattan, i.e., the intersection between cluster 1 and cluster 3, or central area, i.e., the intersection between cluster 1 and cluster 4, with lowest competition and risk.
- The analysis could be more robust and convincing by considering some other factors such as transportations, housing prices and demographics.
- In addition, other unsupervised learning methods can be used to solve the problem as density-based clustering.