Restaurant Genre and Location Selection in New York City

Analysis Report

Analyzed and written by
Jiaping Du

Table of Contents

1. Background:

Every year, 55 million tourists pour into New York City, attracted by its vast options for entertainment, shopping and dining. With 25,000 eating and drinking establishments spread throughout the five boroughs, restaurants are a vital part of the economy, not only providing jobs but luring professionals from other industries to move to the city for its vibrant social life.

But in a city that's more competitive than ever, the restaurant industry has been hurting as it attempts to keep up with rising costs and diners' expectations of top quality at reasonable prices. "The cost to operate a restaurant has been skyrocketing and, in general, the business environment continues to become tougher and tougher," said Andrew Rigie, executive director of the NYC Hospitality Alliance. "Beloved restaurants have closed over the years because they just can't afford to keep their doors open."
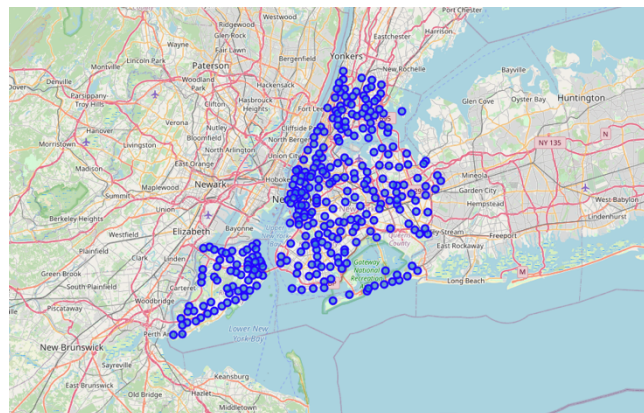
So The selection of the location to the restaurant is much more important in this world that full of competetive.

2. Analysis Data Source

The dataset used in this project is from FourSquare API, and the dataset is about the venue categories in each neighborhood of the New York City. After using these dataset, we will clister the NYC into several clusters and then choose which clusters we should choose to open the restaurant based on the information and the attribution of each cluster (area).

3. The Neighborhoods in New York City:

From our analysis, we can find that there are 5 boroughs and 306 neighborhoods in the New York city, and the distribution of the neighborhoods is shown below.


From the Mapping Plot we can see that there are 306 neighborhoods in the New York City area. The most density area is the Manhattan area. So, we will focus on the Manhattan area as our focus area to select the location of the restaurant.

In the Manhattan area, there are 37 neighborhoods and the density of other areas is much less than the density of Manhattan area.

4. The Venue Situation in each Neighborhoods

From the mapping plot below we can see that there are total 37 Neighborhoods in Manhattan area scattered in the everywhere in the Manhattan area. And the most density area is the southern area of the Manhattan area.
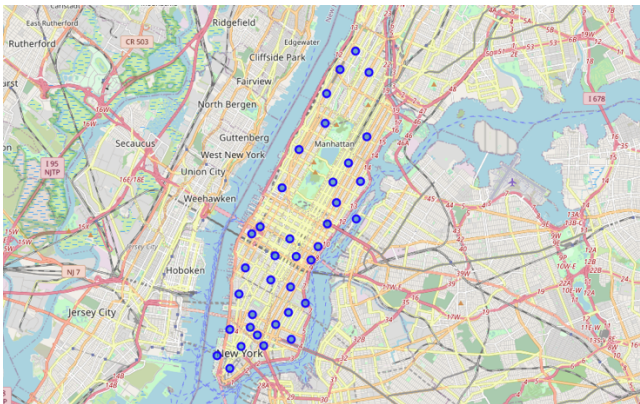
There are many types of restaurants in the Manhattan area including Chinese food, Italian Food and Mexican food, which is the most popular thee type. So, in the following analysis, we will take Chinese food, Italian food and Mexican food restaurant as our type of restaurant to analysis which kind of restaurant can select which area in the Manhattan as their location.

5. Chinese Restaurant in Manhattan Area

From our analysis, we can say that the Chinese Restaurant is mainly locate in only Neighborhood. On other hands, we can have the conclusion that the distribution of the Chinese Restaurant is very concentrated.

Neighborhood	1st Most Common Venue	2nd Most Common Venue
Chinatown	Chinese Restaurant	Bakery
Lower East Side	Chinese Restaurant	Cocktail Bar

Chinatown and the Lower East Side are the two neighborhood that the restaurant in the most popular one of all the Venue.

(Notice: The popular neighbor calculation is based on the 1ˢᵗ most common venue and the 2ⁿᵈ most common venue in the neighborhood)

6. Italian Restaurant in Manhattan Area

Neighborhood	1st Most Common Venue	2nd Most Common Venue
Clinton	Italian Restaurant	Theater
Greenwich Village	Italian Restaurant	Sushi Restaurant
Lenox Hill	Coffee Shop	Italian Restaurant
Little Italy	Bakery	Italian Restaurant
Manhattanville	Coffee Shop	Italian Restaurant
Noho	Italian Restaurant	Coffee Shop
Soho	Italian Restaurant	Coffee Shop
Sutton Place	Italian Restaurant	Gym / Fitness Center
Tribeca	Park	Italian Restaurant
Turtle Bay	Sushi Restaurant	Italian Restaurant
Upper East Side	Italian Restaurant	Bakery
Upper West Side	Italian Restaurant	Coffee Shop
West Village	Italian Restaurant	New American Restaurant
Yorkville	Italian Restaurant	Coffee Shop

From our analysis, we can say that the Italian Restaurant has a very widely distribution in the Manhattan Area. The Popular Neighbors scatter on the 14 Neighborhood in the Manhattan, which indicate that the dispersion of Italian restaurants in the Manhattan is very common, and there is no community gathered.

7. Mexican Restaurant in Manhattan Area

From our analysis, we can say that the Mexican Restaurant is mainly locate in 3 Neighborhoods. On other hands, we can have the conclusion that the distribution of the Mexican Restaurant is very concentrated.

Neighborhood	1st Most Common Venue	2nd Most Common Venue
East Harlem	Mexican Restaurant	Thai Restaurant
Inwood	Mexican Restaurant	Restaurant
Tudor City	Park	Mexican Restaurant

East Harlem and the Inwood are the two neighborhoods that the restaurant in the most popular one of all the Venue in Manhattan.
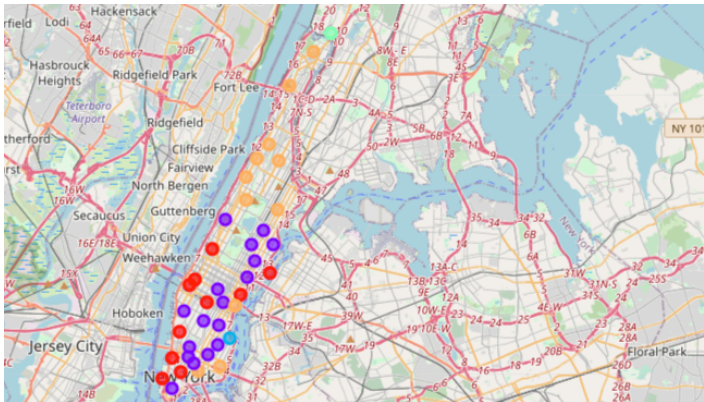
8. Neighborhood Clustering

After we figure out the restaurant distribution in Manhattan area, we decide to make a decision that which location or which area in the Manhattan area is a good place to launch a restaurant and also which type of restaurant can make a better profit in the target area.

So, we decide to use K-mean clustering to cluster these above 30 neighborhoods into 5 cluster in order to better locate the best area to launch the restaurant. To doing so, we first convert the data frame into the following below to better analyze the data in the K-mean model.

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	N
0	Battery Park City	0.000000	0.00	0.00	0.000000	0.014925	0.00	0.000000	0.000000	0.000000	0
1	Carnegie Hill	0.000000	0.00	0.00	0.000000	0.010417	0.00	0.000000	0.010417	0.000000	0
2	Central Harlem	0.000000	0.00	0.00	0.066667	0.044444	0.00	0.000000	0.000000	0.022222	0
3	Chelsea	0.000000	0.00	0.00	0.000000	0.030000	0.00	0.010000	0.000000	0.050000	0
4	Chinatown	0.000000	0.00	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0
5	Civic Center	0.000000	0.00	0.00	0.000000	0.030000	0.01	0.000000	0.000000	0.010000	0
6	Clinton	0.000000	0.00	0.00	0.000000	0.040000	0.00	0.000000	0.000000	0.010000	0
7	East Harlem	0.000000	0.00	0.00	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0
8	East Village	0.000000	0.00	0.00	0.000000	0.020000	0.00	0.010000	0.010000	0.010000	0
9	Financial District	0.000000	0.00	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.000000	0
10	Flatiron	0.000000	0.00	0.00	0.000000	0.030000	0.00	0.000000	0.000000	0.020000	0
11	Gramercy	0.000000	0.00	0.00	0.000000	0.045455	0.00	0.000000	0.000000	0.011364	0

The data above describe the frequency the different venue appear in the given neighborhoods and the frequency has been normalized because we want to make the value in each cell a same scale.

The map above is the cluster result of our analysis, as we can see that we group the neighborhoods into 5 clusters, and each cluster has its own attribution.

For the cluster 1 (red), the venue in this cluster is mainly park, plaza and restaurant. For the cluster 2 (purple), the venue in this cluster is mainly restaurant, bar and coffee shop. For the cluster 3 (blue), the venue in this cluster is mainly Park and Boar or Ferry. For the cluster 4 (green), the venue in this cluster is mainly gym. For the cluster 5 (yellow), the venue in this cluster is mainly Park, Bar, Restaurant (especially Chinese Food).

Neighborhood	1st Most Common Venue	2nd Most Common Venue
Roosevelt Island	Park	Bubble Tea Shop
Lincoln Square	Plaza	Café
Clinton	Italian Restaurant	Theater
Tribeca	Italian Restaurant	Park
West Village	Italian Restaurant	New American Restaurant
Battery Park City	Park	Hotel
Civic Center	Coffee Shop	Cocktail Bar
Midtown South	Korean Restaurant	Hotel
Turtle Bay	Sushi Restaurant	Italian Restaurant
Hudson Yards	Gym / Fitness Center	American Restaurant

Venue in Cluster 1

Neighborhood	1st Most Common Venue	2nd Most Common Venue
Upper East Side	Italian Restaurant	Bakery
Yorkville	Italian Restaurant	Coffee Shop
Lenox Hill	Coffee Shop	Italian Restaurant
Upper West Side	Italian Restaurant	Wine Bar
Midtown	Coffee Shop	Clothing Store
Murray Hill	Sandwich Place	Japanese Restaurant
Chelsea	Coffee Shop	Art Gallery
Greenwich Village	Italian Restaurant	Sushi Restaurant
East Village	Bar	Ice Cream Shop
Little Italy	Bakery	Italian Restaurant
Soho	Italian Restaurant	Coffee Shop
Gramercy	Bar	Pizza Place
Financial District	Coffee Shop	Pizza Place

Venue in Cluster 2

Neighborhood	1st Most Common Venue	2nd Most Common Venue
Stuyvesant Town	Boat or Ferry	Park

Venue in Cluster 3

Neighborhood	1st Most Common Venue	2nd Most Common Venue
Marble Hill	Gym	Coffee Shop

Venue in Cluster 4

Neighborhood	1st Most Common Venue	2nd Most Common Venue
Chinatown	Chinese Restaurant	Bakery
Washington Heights	Café	Bakery
Inwood	Mexican Restaurant	Lounge
Hamilton Heights	Pizza Place	Café
Manhattanville	Coffee Shop	Bar
Central Harlem	African Restaurant	Chinese Restaurant
East Harlem	Mexican Restaurant	Thai Restaurant
Lower East Side	Chinese Restaurant	Cocktail Bar
Manhattan Valley	Bar	Yoga Studio
Morningside Heights	Park	Bookstore
Tudor City	Mexican Restaurant	Café

Venue in Cluster 5

9. Restaurant Type Index (RTI) of each Cluster

Grouping the different neighborhoods into 5 clusters is not our eventually target, we did that is because we want to locate the restaurant into the area there restaurant is the popular venue because it can attract more potential customers into the restaurant, which is also called "Combined Effect". It refers to the economic effects produced by the spatial concentration of various industries and economic activities and the centripetal force that attracts economic activities to a certain area, which are the basic factors leading to the formation and continuous expansion of cities.

The Combined effect is a common economic phenomenon, such as the agglomeration effect of industries. The most typical example is the US Silicon Valley, where dozens of global IT giants and countless small and medium-sized high-tech companies have gathered.

Next, we create an index called Restaurant Type Index (RTI) to get to know the popular degree of different type of restaurant in each cluster. The RTI result is shown below:

	1	2	3	4	5
CH_score	0.000000	0.00000	0.0	0.0	0.047628
IT_score	0.052632	0.09269	0.0	0.0	0.000000
MX_score	0.000000	0.00000	0.0	0.0	0.090909

(RTI in 5 clusters)

From the result, we can see that the cluster 1 and cluster 2 have a very large popularity in Italian food and the cluster 5 have a large popularity in both Chinese food and Mexican food.

10. Location Selection based on the type of restaurant

From all the analysis above, it is clear that if our client wants to start a Italian restaurant, the area in cluster 2 is a good choice for him to enjoy profit from the Combined Effect. And if our client wants to start a Chinese food restaurant or Mexican food restaurant, the area in cluster 5 is a good choice for him to enjoy profit from the Combined Effect.