

Capstone Project-The Battle of Neighborhoods (Week 2)

1. Introduction/Business Problem

The problem I try to solve is in Manhattan, New York, if someone is looking to open a restaurant, where would we recommend that they open it?

This problem is related to all stakeholders who want to open a new restaurant in Manhattan, but not sure where is the best place. For example, if someone wants to open an Italian restaurant, is Little Italy the best place to open a new one or some other places that may also be OK? For Little Italy, it indeed can attract many people who want to eat Italian food. However, there are already too many Italian restaurants in this neighborhood and a new Italian restaurant may not be competitive and attract many customers to it. While for another place where there are a few Italian restaurants around, a new Italian restaurant will definitely attract some people nearby, but where is such a good place in Manhattan and what are the important location factors for choosing such a good place?

We will figure out the above question using the available information in the Foursquare location data and help the stakeholders to find a good place for a new restaurant according to its cuisine. We use Chinese restaurant as an illustration and the procedure can also be used to a new restaurant of other cuisines.

2. Data

By checking all available data from Foursquare location data, I consider the criteria that can be used to determine whether a restaurant is in a good place or not is related to the check-in number / rating number / tip number etc. And the final rating is related to the quality of the restaurant. For example, restaurant A and restaurant B both having 2,000 ratings (or 1,000 check-in numbers, or 500 tips) implies they may be in the similar good/bad places, while restaurant A having a rating of 9.5 and restaurant B having a rating of 8.0 are related to their quality. We checked the dataset and found that the most available criterion is rating number in Foursquare.

As for factors affecting above criteria, we can consider other venues in a range of the existing restaurants as the factors. For opening a Chinese restaurant, we may expect Chinatown can be a good place, but not the unique good place. The important factors can be seen by check the venue percentages in different clusters of the Chinese restaurants with highest rating numbers if we use clustering.

Finally, we will provide some important factors of the good places for stakeholders so that they can find the potential places according to these factors.

2.1 Data Cleaning

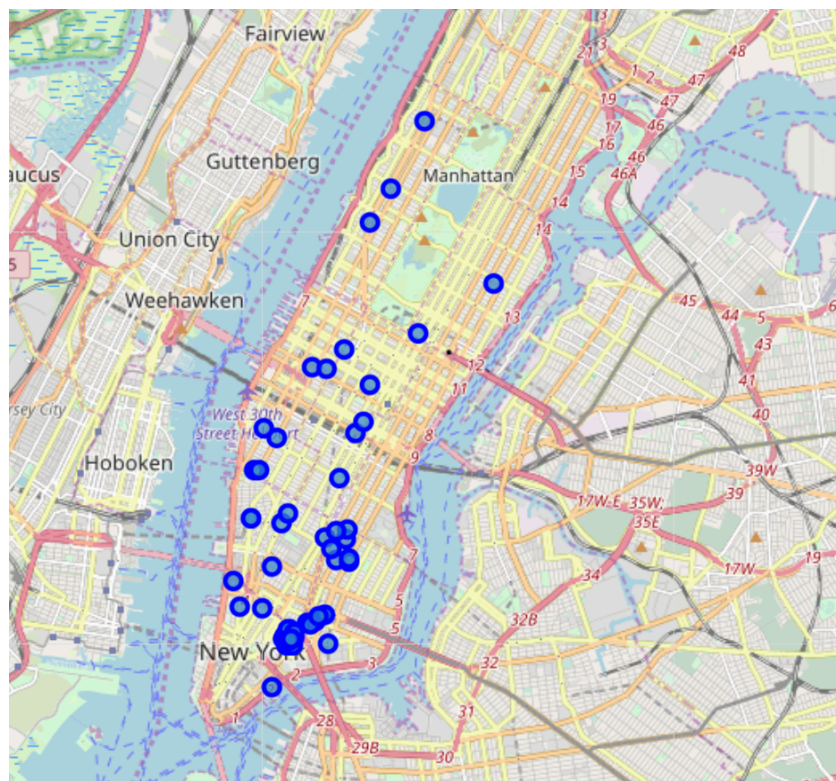
We searched all Chinese restaurant in each neighborhood in Manhattan and first removed some wrongly selected according to their venue category and some duplicates. We had 693 Chinese restaurants available.

We then used premium call to download their rating numbers. We removed all restaurants with with NaN rating numbers and 360 restaurants left. We then investigated the top restaurants with highest rating numbers in these 360 restaurants.

3. Methodology

3.1 Exploratory data analysis

We would like to first check where the most popular restaurant are located. Let's check the top 60 with largest rating numbers on the map.



As shown in the above map, we can obviously see that there are indeed several regions which have more people to visit, and Chinatown is among the most popular place for people to get Chinese food. It is worth for us to investigate the reason why these several small regions attract more people. Thus, we explore the around venues of these 60 Chinese restaurants. Let's print some restaurants along with the top 5 most common venues.

----456 Shanghai Cuisine----			----Buddakan----		
	venue	freq		venue	freq
0	Chinese Restaurant	0.14	0	Bakery	0.07
1	Bakery	0.07	1	Coffee Shop	0.05
2	Italian Restaurant	0.05	2	Hotel	0.05
3	Salon / Barbershop	0.04	3	Seafood Restaurant	0.04
4	Dessert Shop	0.04	4	Nightclub	0.03

----99 Favor Taste 99號餐廳----			----Buddha Bodai 佛菩提素菜----		
	venue	freq		venue	freq
0	Chinese Restaurant	0.04	0	Chinese Restaurant	0.13
1	American Restaurant	0.04	1	Dim Sum Restaurant	0.07
2	Cocktail Bar	0.04	2	Bakery	0.07
3	Bar	0.04	3	Bubble Tea Shop	0.04
4	Bakery	0.03	4	Ice Cream Shop	0.04

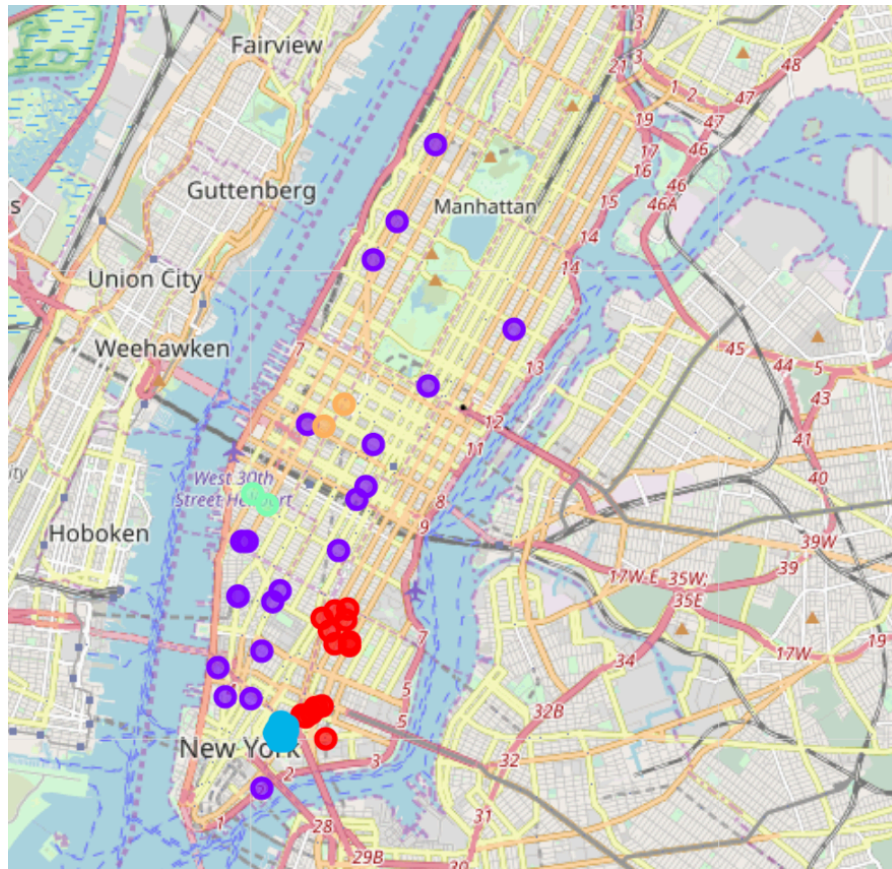
----Baohaus----			----Café China----		
	venue	freq		venue	freq
0	Pizza Place	0.05	0	Korean Restaurant	0.06
1	Japanese Restaurant	0.05	1	Hotel	0.05
2	Dessert Shop	0.04	2	Japanese Restaurant	0.04
3	Ramen Restaurant	0.03	3	Coffee Shop	0.04
4	Sushi Restaurant	0.03	4	Bakery	0.04

----Big Wong King 大旺----			----Cha Chan Tang 茶餐廳 (Cha Ch		
	venue	freq		venue	freq
0	Chinese Restaurant	0.14	0	Chinese Restaurant	0.13
1	Bakery	0.07	1	Bakery	0.06
2	Italian Restaurant	0.05	2	Dim Sum Restaurant	0.05
3	Salon / Barbershop	0.04	3	Dessert Shop	0.05
4	Dessert Shop	0.04	4	Bubble Tea Shop	0.05

For some restaurants, Chinese Restaurant is the 1st most common venue. We can guess these restaurants are located in Chinatown. However, for many restaurants Chinese Restaurant is even not top 5 most common venues for them. We should consider some technique to see the groups with different most common venues. We use Clustering to cluster the top 60 Chinese restaurants with highest rating numbers.

3.2 Cluster Restaurants

We run *k*-means to cluster the restaurants into 5 clusters. Let's first visualize the resulting clusters.



4. Results

We can examine each cluster with the most common venues together with above map to determine the discriminating venue categories that distinguish each cluster.

Cluster 1 (red in the map)

Cluster Labels	Restaurant	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	Wah Fung Number 1 Fast Food 華豐快飯店	Chinese Restaurant	Sandwich Place	Ice Cream Shop	Noodle House	Cocktail Bar	Vietnamese Restaurant	Dumpling Restaurant	American Restaurant	Hotpot Restaurant	Bakery
0	Shu Jiao Fu Zhou Cuisine 潭頭王福州小吃 (Shu Jiao Fu ...	Cocktail Bar	Bar	Mexican Restaurant	Asian Restaurant	Pizza Place	Art Gallery	Sandwich Place	Ice Cream Shop	American Restaurant	Australian Restaurant
0	Hot Kitchen	Japanese Restaurant	Ice Cream Shop	Dessert Shop	Italian Restaurant	Gym	Seafood Restaurant	Mexican Restaurant	Grocery Store	Coffee Shop	Yoga Studio
0	Màlà Project	Ice Cream Shop	Vegetarian / Vegan Restaurant	Cocktail Bar	Coffee Shop	Japanese Restaurant	Seafood Restaurant	Ramen Restaurant	Chinese Restaurant	Vietnamese Restaurant	Wine Bar
0	Tim Ho Wan 添好運	Japanese Restaurant	Yoga Studio	Ice Cream Shop	Vegetarian / Vegan Restaurant	Dessert Shop	Coffee Shop	Cosmetics Shop	Grocery Store	Café	Clothing Store
0	Mimi Cheng's	Japanese Restaurant	Dessert Shop	Pizza Place	Ice Cream Shop	Grocery Store	Italian Restaurant	Coffee Shop	Ramen Restaurant	Bar	Chinese Restaurant
0	99 Favor Taste 99號餐廳	American Restaurant	Cocktail Bar	Bar	Chinese Restaurant	Bakery	Pizza Place	Café	Furniture / Home Store	Boutique	Sandwich Place
0	Spicy Village	Chinese Restaurant	American Restaurant	Cocktail Bar	Optical Shop	Sandwich Place	Café	Bakery	Speakeasy	Pizza Place	Salon / Barbershop
0	Congee Village 粥之家	Bar	Pizza Place	Mexican Restaurant	Sandwich Place	Speakeasy	American Restaurant	Boutique	Asian Restaurant	Art Gallery	Italian Restaurant
0	The Bao	Japanese Restaurant	Ice Cream Shop	Coffee Shop	Grocery Store	Gym	Dessert Shop	Chinese Restaurant	Sushi Restaurant	Tea Room	Italian Restaurant

We can see Café, NightClub, Cocktail Bar, Bar are in the most common venues, if you check the map you can see they are mainly in East Village, etc., where are indeed the nightlife areas.

Cluster 2 (purple in the map)

Cluster Labels	Restaurant	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
1	Xi'an Famous Foods	American Restaurant	Gym	Italian Restaurant	Restaurant	Coffee Shop	Gym / Fitness Center	New American Restaurant	Yoga Studio	Café
1	Ollie's Sichuan Restaurant	Theater	Italian Restaurant	Hotel	Burger Joint	Gym / Fitness Center	Cocktail Bar	American Restaurant	Pizza Place	Lounge
1	Pinch Chinese	Italian Restaurant	Clothing Store	French Restaurant	Sushi Restaurant	Seafood Restaurant	Coffee Shop	Café	Dessert Shop	Cosmetics Shop
1	Fish Market	Italian Restaurant	Café	American Restaurant	Juice Bar	Seafood Restaurant	Salad Place	Restaurant	Vegetarian / Vegan Restaurant	Deli / Bodega
1	Very Fresh Noodles	Coffee Shop	Bakery	Hotel	Seafood Restaurant	Nightclub	Steakhouse	Art Gallery	Boutique	Italian Restaurant
1	Mr. Chow	American Restaurant	Italian Restaurant	Spa	Coffee Shop	Boutique	Park	Men's Store	Cocktail Bar	Playground
1	Philippe	Hotel	Boutique	French Restaurant	Jewelry Store	Italian Restaurant	Furniture / Home Store	Department Store	Spa	Art Gallery
1	Xi'an Famous Foods	Italian Restaurant	Sushi Restaurant	Thai Restaurant	Bar	Dessert Shop	Turkish Restaurant	Ice Cream Shop	Gym	Mexican Restaurant
1	Hao Noodle and Tea by Madam Zhu's Kitchen	Italian Restaurant	Jazz Club	Coffee Shop	Massage Studio	Café	Cocktail Bar	Wine Bar	American Restaurant	Gay Bar
1	Sammy's Noodle Shop	Italian Restaurant	Gym / Fitness Center	American Restaurant	Coffee Shop	Jazz Club	Massage Studio	Café	Arts & Crafts Store	Wine Bar

We may see Clothing store, Cosmetics Shop, Department Store, Men's store in the most common venues, by checking the map, they are mainly located around Fifth Avenue Midtown area and SoHo, which are both shopping areas.

Cluster 3 (blue in the map)

Cluster Labels	Restaurant	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	
2	Oriental Garden 福臨門海鮮酒家	Chinese Restaurant	Bakery	Dim Sum Restaurant	Dessert Shop	Salon / Barbershop	Italian Restaurant	Vietnamese Restaurant	Spa	Japanese Restaurant	Si
2	Tasty Dumpling	Chinese Restaurant	Dim Sum Restaurant	Bakery	Dessert Shop	Bubble Tea Shop	Salon / Barbershop	Optical Shop	Vietnamese Restaurant	Malay Restaurant	
2	Taiwan Pork Chop House 臺灣武昌好味道	Chinese Restaurant	Bakery	Dim Sum Restaurant	Bubble Tea Shop	Dessert Shop	Noodle House	Ice Cream Shop	Spa	Vietnamese Restaurant	C
2	Deluxe Green Bo Restaurant	Chinese Restaurant	Bakery	Dessert Shop	Dim Sum Restaurant	Salon / Barbershop	Noodle House	Vietnamese Restaurant	Malay Restaurant	Spa	
2	456 Shanghai Cuisine	Chinese Restaurant	Bakery	Italian Restaurant	Dessert Shop	Salon / Barbershop	Vietnamese Restaurant	Bubble Tea Shop	Spa	Dim Sum Restaurant	C
2	Ping's Seafood	Chinese Restaurant	Bakery	Dim Sum Restaurant	Bubble Tea Shop	Dessert Shop	Spa	Cocktail Bar	Optical Shop	Ice Cream Shop	I
2	Buddha Bodai 佛菩提素菜	Chinese Restaurant	Bakery	Dim Sum Restaurant	Ice Cream Shop	Bubble Tea Shop	Dessert Shop	Vietnamese Restaurant	Cocktail Bar	Optical Shop	
2	Cha Chan Tang 茶餐廳 (Cha Chan Tang)	Chinese Restaurant	Bakery	Dessert Shop	Dim Sum Restaurant	Bubble Tea Shop	Salon / Barbershop	Optical Shop	Vietnamese Restaurant	Malay Restaurant	
2	Peking Duck House	Chinese Restaurant	Bakery	Dim Sum Restaurant	Dessert Shop	Bubble Tea Shop	Spa	Malay Restaurant	Optical Shop	Noodle House	I
2	Dim Sum Go Go	Chinese Restaurant	Bakery	Dim Sum Restaurant	Bubble Tea Shop	Dessert Shop	Ice Cream Shop	Noodle House	Cocktail Bar	Vietnamese Restaurant	

It is very obvious that Chinese Restaurant is the most common venues for all restaurants in this cluster. By checking the map, they are indeed in Chinatown.

Cluster 4 (green in the map)

Cluster Labels	Restaurant	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
3	Chop Shop	Art Gallery	Park	Coffee Shop	Chinese Restaurant	Hotel	Wine Shop	Bookstore	Italian Restaurant	Gym	Lounge
3	Grand Sichuan International	Art Gallery	Theater	Café	Pizza Place	Coffee Shop	Italian Restaurant	Grocery Store	Bakery	Bagel Shop	French Restaurant

Cluster 5 (orange in the map)

Cluster Labels	Restaurant	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
4	Kung Fu Little Steamed Buns Ramen	Theater	Coffee Shop	Sandwich Place	Bar	Thai Restaurant	Italian Restaurant	Hotel	American Restaurant	Bakery	Performing Arts Venue
4	Hakkasan	Theater	Hotel	Italian Restaurant	Burger Joint	American Restaurant	Sandwich Place	Bakery	Indie Theater	Japanese Restaurant	Pizza Place

For cluster 4 and 5, the most common venues are Art Gallery and Theater, we consider them together as art areas.

5. Discussion

From the results, we can see that the good place for opening a new Chinese restaurant is not only the Chinatown. Other good places such as potential locations near nightlife areas, art areas or shopping area may also be considered.

The results will be helpful for the stakeholders who want to open a new restaurant in Manhattan. If you want to open a new Chinese restaurant and there is a similar restaurant already in Chinatown. You don't have to compete with it by opening the new one in the Chinatown and may consider other places with above discussed factors, such as around Fifth Avenue Midtown area. It can also attract many customers, especially who are near your new restaurant.

This methodology can be improved by using a more powerful criterion if available. And it is the first and the most important step for helping the stakeholders. A further step I consider may be once the stakeholders have some good potential places and are not sure which is the best, a regression model can be fitted with criterion as the outcome and venue factors as the covariates. The Lasso regression model can be used to reduce the factors numbers. Moreover, a more complicated model can also consider such as rent cost, safety, etc.

6. Conclusion

Using Foursquare data and through Clustering technique, we provide a method to help the stakeholders who want to open a new restaurant in Manhattan choose a potential good place. The method can be applied to a new restaurant with any cuisine and give the stakeholders some important factors to find the potential good places.