Finding a Japanese restaurants in New York

Kaori Ishikawa

03 May 2020

1. Introduction

In this project, we will explore Japanese restaurants in New York. I am originally from Japan and it's been 5 years since I left my country. And, I enjoy trying out a Japanese restaurant when I travel. The reason could be because I am not completely happy with the quality of Japanese restaurants in my city or simply I want to try a new restaurant as the number of the Japanese restaurants in my city is limited.

This report is targeted for people like me, who want to have Japanese food during the trip especially to a big city, New York.

When people look for a restaurant, people mostly check the price, rating and location. Therefore, we will explore Japanese restaurants in New York based on these criteria.

2. Data Acquisition

2.1 Data sources

Geographic coordinate of the boroughs and the neighbourhoods in New York was obtained from here. Details of restaurants including price tier, rating, and location in New York was obtained from Foursquare API. Number of hotels in each Neighborhood Tabulation Areas (NTA) was obtained from here. And, Gis data of Neighborhood Tabulation Areas (NTA) in New York was obtained from here.

2.2 Data cleaning

Data downloaded from multiple sources were combined into a few tables. During the cleaning of the data of Japanese restaurants, there are 3 restaurants out of 83 did not have a rating or price information yet. I removed those restaurants from the data for the further analysis.

3. Exploratory Data Analysis

3.1 Price Tier

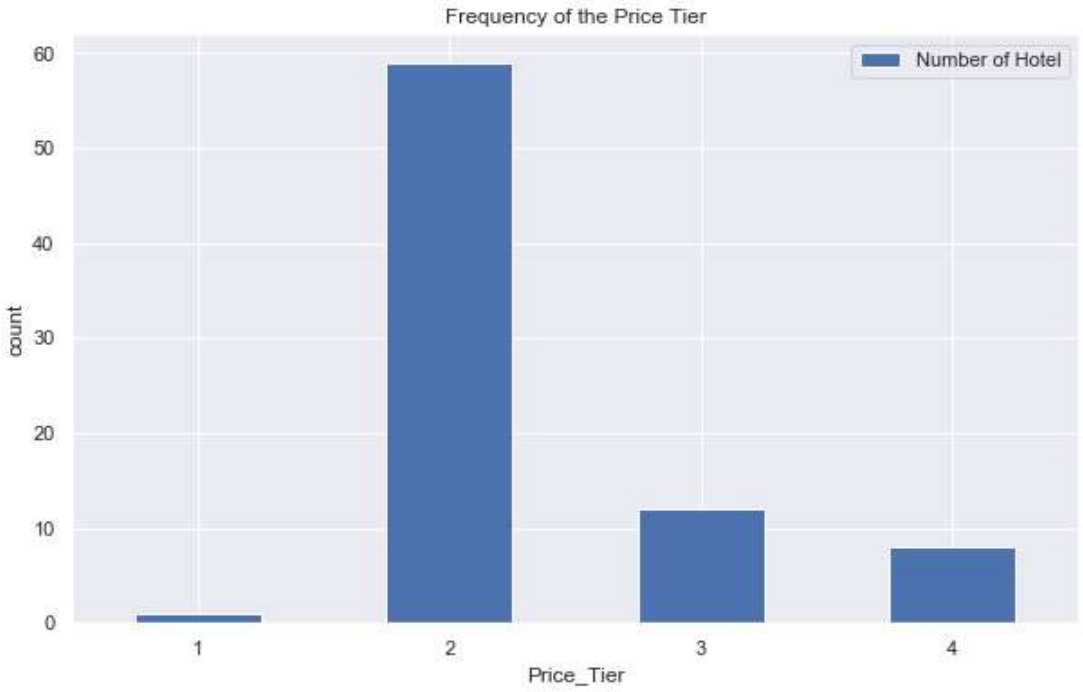
Price tier (Price points) used here is based on the foursquare API. Foursquare define it as following: Currently valid range of price points are [1,2,3,4], 1 being the least expensive, 4 being the most expensive. For food venues, in the United States, 1 is < USD 10 an entree, 2 is USD 10-20 an entree, 3 is USD 20-30 an entree, 4 is > USD 30 an entree.

Table 1 and Figure 1 shows the count of the restaurants. Tier 2 accounts for 73.75% of all.

Table 1

	Price_Tier
2	59
3	12
4	8
1	1

Figure 1



3.2 Rating

Rating is between 0 and 10, and 10 is the best. Mean of the rating is 8, the minimum is 5.9, and the maximum is 9.2.

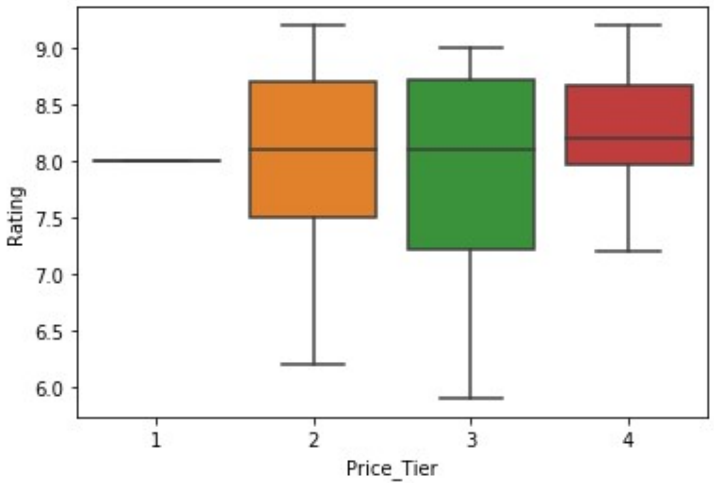
Table 2

count	80.000000
mean	8.013750
std	0.800434
min	5.900000
25%	7.500000
50%	8.100000
75%	8.700000
max	9.200000

3.3 Relationship between Rating and Price Tier

I created the box plot to see if I can find any relationship between rating and price tier. From the figure 2, you can see that there is only a little difference in mean between each price tier, but the low rating of the most expensive tier, Tier 4, is much higher than other price tiers. If you want to have a nice meal, you don't necessary have to go to the expensive restaurants. However, if you go to expensive restaurants, you will less likely dissatisfied with the restaurants based on this data.

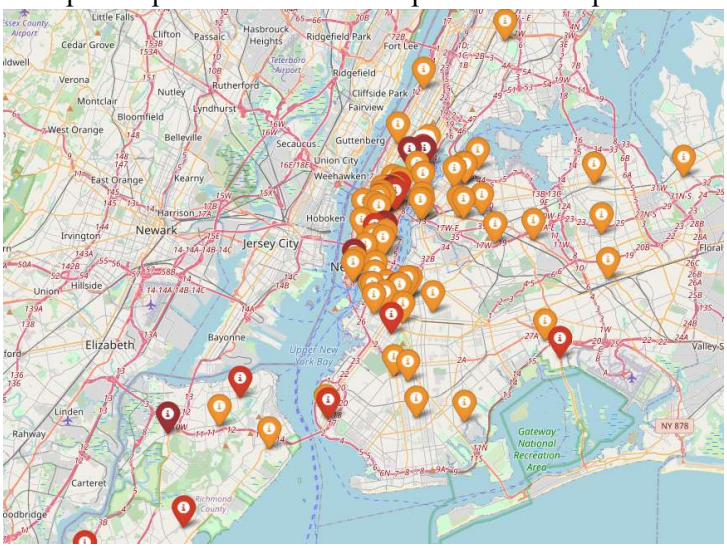
Figure 2: Box Plot



3.4 Relationship between location and price tier

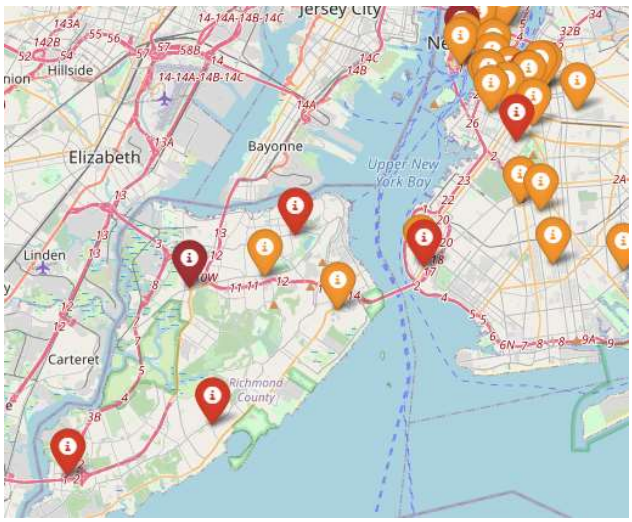
I create a map of the venue with the cooler based on the price tier. From the Map 1, you can see that most of the restaurants are in Manhattan area. Map 2 is zoomed in the Manhattan area. You can see that that in Manhattan, East 39th near Grand Central Terminal and East 83rd Street have high-end restaurants. Also, from Map 3, Staten island has high average price although it has only a few Japanese restaurants.

Map 1 : Japanese restaurant map based on the price tier



Map 2 : Manhattan area

Map 3 : Staten island area



3.5 Relationship between Rating and location

I create a map of the venue with the cooler based on the Rating. From the Map 4, you can see that the restaurants in Manhattan area has higher rating. And, if you go further from Manhattan, the rating tends to go lower.

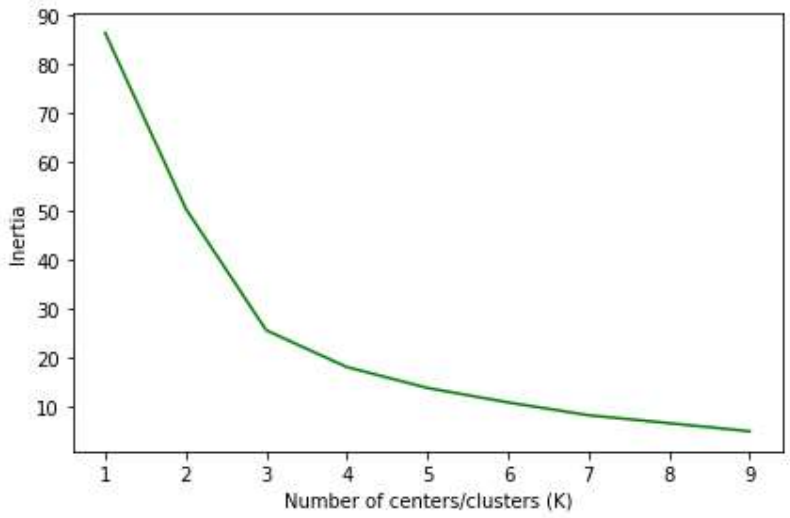
Map 4 : Japanese restaurant map based on the rating

4. Clustering

4.1 Finding the best k

I clustered by k-mean method. To decide the best k by elbow method, I plot the inertia with the different k. (Figure 3) I will use k = 4 in this project.

Figure 4



4.2 Clustering Japanese restaurants

I made 4 clusters based on rating and price tier. Table 3 is the characteristic of each tier. Restaurants in cluster 0 have good rating but price tier is low. On the other hands, restaurants in cluster 2 has good rating and price tier is relatively high. The other 2 cluster seems less balanced between rating and price, and I will use cluster 0 and 2 as recommended restaurants with 2 different budget.

Table 3

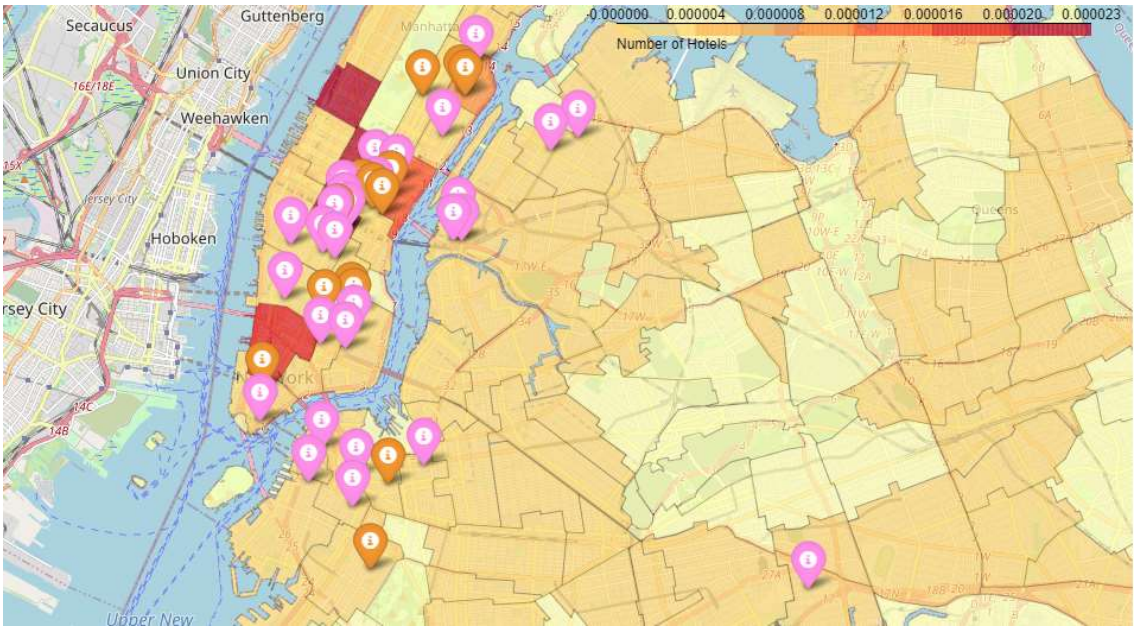
	Cluster	Count	Mean Price Tier	Mean Rating
0	0	28.0	2.000000	8.671429
1	1	11.0	2.272727	6.563636
2	2	16.0	3.500000	8.406250
3	3	25.0	2.000000	7.664000

4.3 Relationship between cluster, location and the hotel density

I map the restaurants in cluster 0 and 2 in a choropleth maps based on the hotel density. People prefer to have dinner not too far from the hotel where they are staying. To give an idea when you select the are to stay, I used hotel density per Neighbourhood Tabulation

Areas. In this analysis, I will not look into the detail of hotels. From map 5, you can see that most of the restaurants are in Manhattan area. Especially, Midtown-Midtown South area and Turtle Bay-East Midtown have high density of hotels, and also well-balanced Japanese restaurants nearby.

Map 5 : Japanese restaurants in cluster 0 and 2



5. Results and Discussion

In New York, most of the Japanese restaurants are affordable. You can have an entree between USD 10-20 in 74% of the restaurants. And, there is very little difference in the average rating between each price tier, and it shows that you can find a nice restaurant within your budget. By k-mean clustering, we selected 2 clusters of restraints with good rating and 2 different budgets. From the map, you can see that Midtown-Midtown South area and Turtle Bay-East Midtown have a lot of hotels, and also well-balanced Japanese restaurants nearby. For future study, I'd like to also look into the detail of hotels such as location, stars, price and so on. And, as the number Japanese restaurants are relatively small, it is also important to compare the data with other restaurants or other cities in order to get the threshold of the rating above which is considered "good restaurant". However, I could not get venues details of all restaurants this time due to the limitation of foursquare account I have. Also, rating system is subjective and the same restaurant can be rated quite differently between by the people from Japan and by others. This also should be kept in mind.

6. Conclusion

In this project, I studied Japanese restaurants in New York. While the number of Japanese restaurants is continuously growing in many international cities, not all the restaurants have a good quality. However, New York has many great Japanese restaurants with different price ranges and you can find a restaurant you can enjoy.