Restaurant Location Selection

----The Battle of Neighborhoods

Gary Liang

October 30, 2019

**A. Introduction**

**A.1. Background & Problem Description**

New York City, the [most populous city](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population) in the United States, one of the greatest metropolises over the world, is a dream place for gourmet to seek delicious cuisine. Here, you may find all types of restaurant from each corner of the world. Its food culture includes an array of international cuisines influenced by the city's immigrant history. [Central](https://en.wikipedia.org/wiki/Central_Europe) and [Eastern European](https://en.wikipedia.org/wiki/Eastern_European) immigrants, especially [Jewish](https://en.wikipedia.org/wiki/Jewish_Americans) immigrants from those regions, brought [bagels](https://en.wikipedia.org/wiki/Bagel), [cheesecake](https://en.wikipedia.org/wiki/Cheesecake#North_America), [hot dogs](https://en.wikipedia.org/wiki/Hot_dog), [knishes](https://en.wikipedia.org/wiki/Knish), and [delicatessens](https://en.wikipedia.org/wiki/Delicatessens) (or [delis](https://en.wikipedia.org/wiki/Delis)) to the city. [Italian](https://en.wikipedia.org/wiki/Italian_diaspora) immigrants brought [New York-style pizza](https://en.wikipedia.org/wiki/New_York-style_pizza) and [Italian cuisine](https://en.wikipedia.org/wiki/Italian_cuisine) into the city, while Jewish immigrants and [Irish](https://en.wikipedia.org/wiki/Irish_diaspora) immigrants brought [pastrami](https://en.wikipedia.org/wiki/Pastrami) and [corned beef](https://en.wikipedia.org/wiki/Corned_beef), respectively. [Chinese](https://en.wikipedia.org/wiki/Chinese_restaurant) and other Asian restaurants, [sandwich](https://en.wikipedia.org/wiki/Sandwich) joints, [trattorias](https://en.wikipedia.org/wiki/Trattoria), [diners](https://en.wikipedia.org/wiki/Diner), and [coffeehouses](https://en.wikipedia.org/wiki/Coffeehouse) are ubiquitous throughout the city. Some 4,000 mobile food vendors licensed by the city, many immigrant-owned, have made Middle Eastern foods such as [falafel](https://en.wikipedia.org/wiki/Falafel) and [kebabs](https://en.wikipedia.org/wiki/Kebab) examples of modern New York [street food](https://en.wikipedia.org/wiki/Street_food). The city is home to "nearly one thousand of the finest and most diverse [haute cuisine](https://en.wikipedia.org/wiki/Haute_cuisine) restaurants in the world", according to [Michelin](https://en.wikipedia.org/wiki/Michelin). As of 2019, there were 27,043 restaurants in the city, up from 24,865 in 2017[1].

As the figures tells, New York City attracts many to start their business in food industry. Before they take action, they need to find out where they would open it? What would they consider when selecting a location? I hope to explore regional characteristics of these restaurants and figure out is the neighborhood of restaurant an important for factor for success of a restaurant with sound analysis.

As I mentioned above, there are hundreds types of restaurants, it is impractical for me to run analysis for each type of restaurant. Based on the maximum total numbers among these restaurants, I choose Pizza Place for the following analysis. The analysis of other types of restaurants can be conducted with the same method. The conclusion on relationship between success and neighborhood of a restaurant can be a useful finding for those who plan to operate a restaurant.

### A.2. Data Preparation

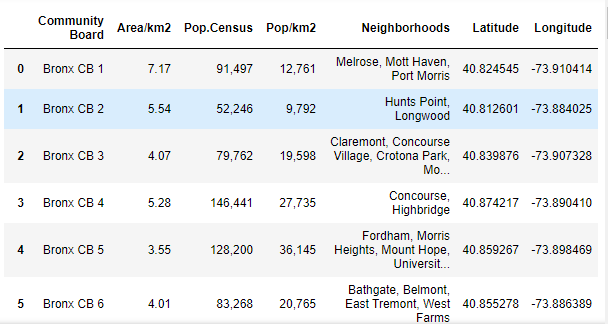
Data used in analysis are listed as below:

* Neighborhoods in New York City from Wikipedia. <https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City> I cleaned the data and reduced it to boroughs of NYC so that I can use it to find geological location for further venues analysis.
* Using **Geopy** to get geological location by address name
* Using **Forsquare API** to get the most common venues of given Borough of New York city.
* Using **Forsquare API** to get the venues record of given venues of New York city.

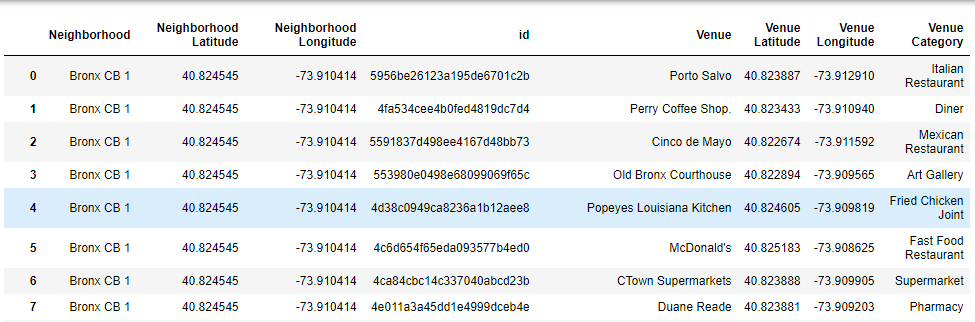
### B. Methodology

I used **BeautifulSoup** to scrape boroughs from Wikipedia, and organize a table containing *Community Board, Area, Pop.Census, Neighborhoods* information of New York City.

And used Geopy to get geological location of each community board. (Because Geopy cannot recognize the address like 'Bronx CB 1', I use the first address in the list of Neighborhood of each community board. If it is still not found, the second address will be used.)



I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 500 meter for each borough from their given latitude and longitude informations. Here is a head of the result, adding venue id, venue name, category, latitude and longitude informations from Forsquare API.

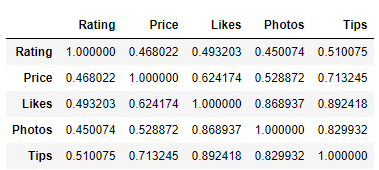


It returns with 2555 records. I summarize venues by category. Among these 2555 records, Pizza Place counts 117 with the maximum total number. Therefore, I choose Pizza Place as an example of restaurants for further analysis.

I utilized the Foursquare API again by pizza places ID to explore detailed record of these pizza places. Select out Rating, Price, Likes, Photos, Tips into a dataframe. And drop those places without a rating.

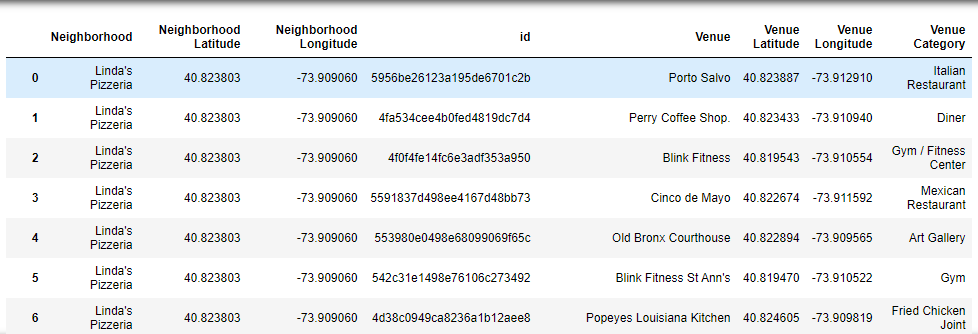


Then I tried to find correlation among these variables:

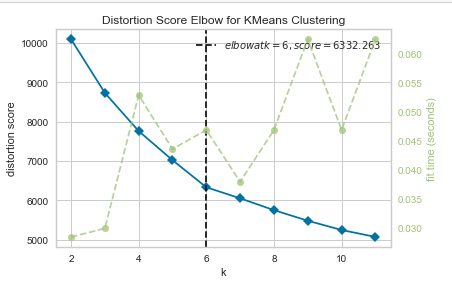


Showing from the correlation matrix, Likes, Photos and Tips are highly correlated to each other. But Likes is not highly related to Rating. Customers who click Likes for some specific reasons but give lower ratings to the general performance might cause this low correlation. Therefore I choose Rating to represent the restaurant. Rating is somewhat correlated to Price, which indicates that the price might not affect impressions of customers on that place significantly.

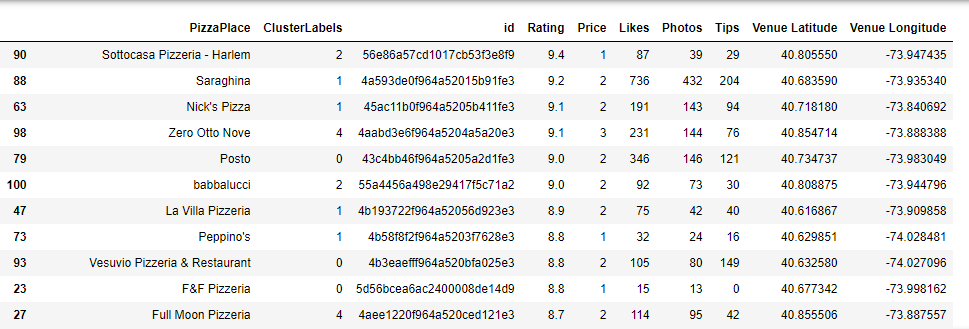
I utilized the Foursquare API centering these pizza places to explore their neighborhoods with a 500-meter radius.



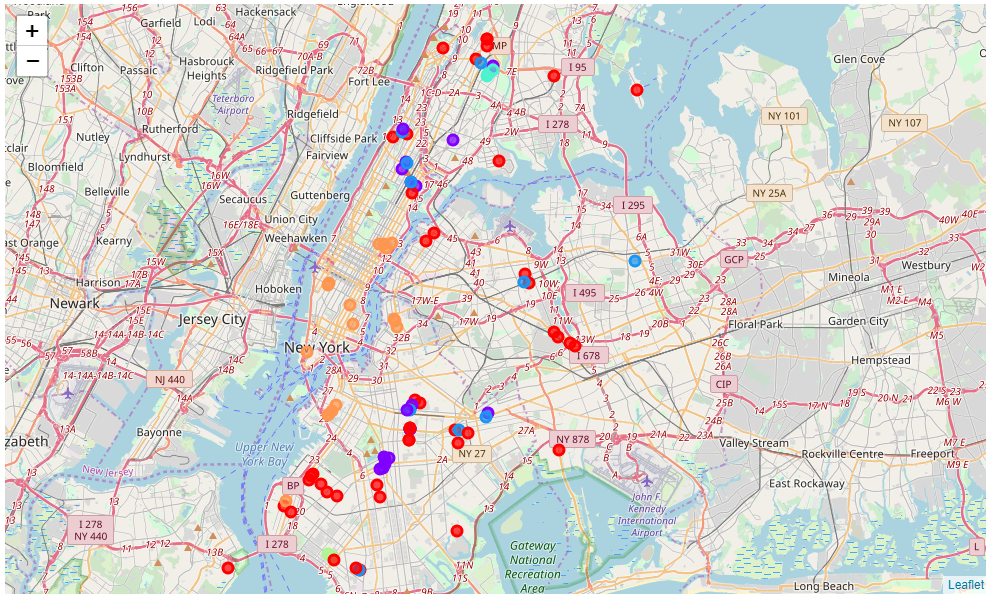
According to venues categories and numbers surrounding each pizza house, I use k-means to cluster pizza house into several groups. I use “elbow” method to help select the optimal number of clusters by fitting the model with a range of values for k. The “elbow” (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. In the visualizer “elbow” k=6 is annotated with a dashed line.



I merged cluster labels of each pizza place with its geological location.

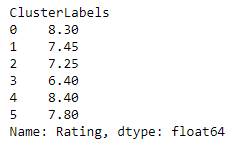


Then I used **folium** to visualize distribution of these pizza places in NYC as below:

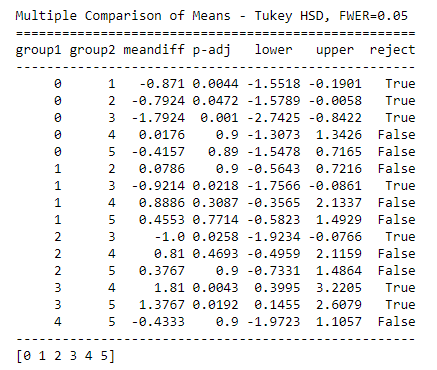


### C. Results

Let’s see if there is different in performance of these pizza places with different clustering labels:



Based on median, it seems there is difference. But is it significant? I ran one-way ANOVA analysis:



Showing in the table, the performances of these 6 types pizza places are significantly different. Places with cluster label 0 and 4 perform best while those with label 1,2 and 3 perform worst. Hence, different neighborhood might affect the impression of the customers on this restaurant. Restaurant owners may search for a similar location to start their business. So what is special in these locations around pizza places?

I select out pizza places with label 0 and 4, and drop categories with empty value, sort venue types by average numbers around pizza places in a descending way.



Pizza place as a classic light meal, normally is opened at a location where different light meals gather, like Cafe, bakery or bar. It explains why places with higher ratings gather around Manhattan. Does the pizza there truly taste better than other places'? Short time spent on tasting could be a reason that render the quality of ratings. Another interesting thing is that Italian restaurant numbers is apparently more than other types of restaurant, especially for pizza places with label 1. These places are in Bronx which is happen to be an Italian area. Pizza as a representative of Italian food seems meet the correct market there.

### D. Conclusion

As a result, for those types of pizza place where providing quick services and moderate flavor with moderate price, they should consider locations in busy area and close to other light meals restaurant.

For those aiming to provide delicious pizza to most picky gourmet, they should open their pizza place at places close to their target customers or customers with certain background.

### E. Discussion

As a recommendation to those who plan to operate restaurant, location selection is only one basic problem to think over. The analysis of this report assumes the type of restaurant is selected, for example, a pizza house. It can not solve the problem about whether a type restaurant is the most popular type and how much customer will visit every day. And as location suggestion, it offer a opportunity analysis but lack risk analysis, like the cost of the location and competition in that area.

Although in this report, it demonstrates the relations between location and ratings, but ratings might not reflect the operation status of the restaurant. Restaurant with a high rating could still be unprofitable, which is unsuccessful from business perspective. So the suggestion is relatively narrow. In order to suggest with more practical and profitable ideas, the relationship between customer reactions and financial performance should be evaluated.

With all these analyses done, the report finally become constructive for a restaurant owner in real business world.