

# Location Choices of a Restaurant in New York City

Freddy Li

Aug 1, 2019

## Introduction

The project is mainly to solve the question that if someone plan to open a restaurant in New York City, how he should choose the location. To solve the problem, the first step is to find the best borough according to crime and income data because a safe location surrounded by rich people might worth to invest a restaurant. Next, the neighborhoods in selected borough is explored and clustered based on the data from Foursquare to assess the clusters of neighborhoods by the popularity of the venues, which could be the nearby neighbors to absorb the clients to step in the restaurant after visiting that venues, and categories of venues, which implies what kind of restaurants could be better choices. Finally, the list of recommended venues as neighbors from different neighborhoods in selected borough is generated. The restaurant is recommended to open near these venues within the popular neighborhoods of worthy borough for different clusters. The investors, restaurants and food providers would feel interested in this project to help them make decision of location.

## Data

### Data Source:

To process this project, New York Neighborhoods data, New York Crime data, New York Income data and detailed geographic data with venues information are used.

- New York Neighborhoods data: The data is from [“2014 New York City Neighborhood Names”](#) provided by Department of City Planning, NY to give out neighborhoods' names and locations.
- New York crime data: The data is from [“NYPD Complaint Data for Current Year”](#) provided by NYPD to display the detailed crime data by case level.
- New York income data: The data is from [NY QuickFacts](#) provided by Census Bureau to give out the statistics until 2017 by boroughs in New York City.
- Geographic Data: The data is from [Foursquare API](#). The explore endpoint is used to retrieval the recommended nearby venues around neighborhoods.

### Data Cleaning:

To prepare the analysis, all the data scrapped from different sources should be cleaned and formatted to get the crucial information for analysis.

For neighborhoods name data, there are bunches of information not necessary in an unstructured format. Therefore, only borough name, neighborhood name, latitude and longitude were selected from the dataset and transferred into data table.

For crime data, only 3 variables of 35 variables were important to analysis including the crime case ID, happening borough and level of crime. Among 3 levels of crime including felony, misdemeanor and violation, felony was assumed to have the most important for restaurants and used to filtered out the target data.

For statistic data for New York QuickFacts, the data had multiple problem to process. First, there were many columns with null value and should be removed. Second, there were lots of unrelated information to regional income level. Therefore, Median Household Income was selected to stand for regional income level. Third, the names of variables and elements were too long and should be adjusted. Besides, the values of income included dollar sign and comma, which should be removed and the data type of income was adjusted to numeric to calculate.

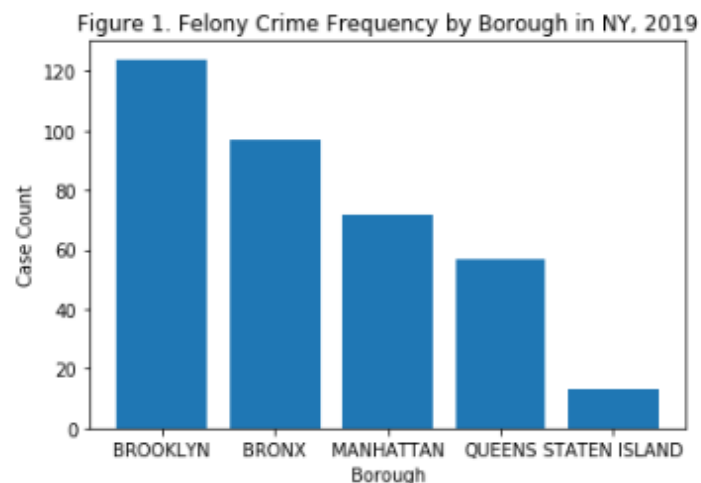
For data from Foursquare API, only the related information was scrapped out including venue name, venue location and venue categories. To query data from the API, neighborhoods name data should be prepared firstly as the inputs. The exploring range was limited to 500 meters from the neighborhoods' coordinates and the number of explored venues was limited to 100. Another problem was that when querying data from Foursquare API, some neighborhoods would have no venues to find. These neighborhoods would be deleted.

## Methodology

### Exploratory Analysis

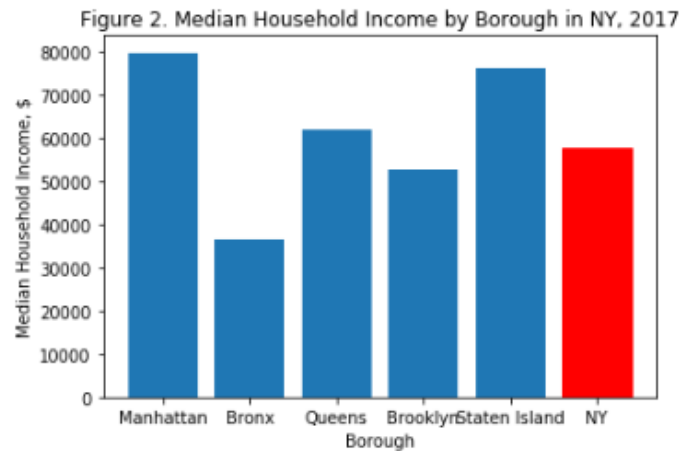
In the dataset of neighborhood names of NY, I found there were 5 boroughs and 306 neighborhoods totally to consider which location could be better to open restaurants.

In the dataset of felony crime data, there were 368 cases this year until July 19<sup>th</sup>, 2019. To compare the crime rate situation, Figure 1 displayed the ordered case counts for each borough. Obviously, Queens and Staten Island had the lowest crime rate this year, especially Staten Island. A safer environment could lower the operation risk and avoid unnecessary damages for restaurants. Therefore, Queens and Staten Island might be good choices.



In the dataset of income level data, the whole New York had median household income of \$57,782 in 2017. Staten Island had the second highest income level of \$ 76,242, following Manhattan, which was higher than the median income level of the whole NY. The median income level of \$62,008 for Queens

was lower than Staten Island and just higher slightly than the average NY. Obviously, the higher median household income, the more likely the restaurants generate high profit.



Combining the analysis of crime rate and median income level among different boroughs in NY, Staten Island was the best borough for restaurant opening. In Staten Island, there were 63 neighborhoods to consider.

### Neighborhoods Exploration in Staten Island:

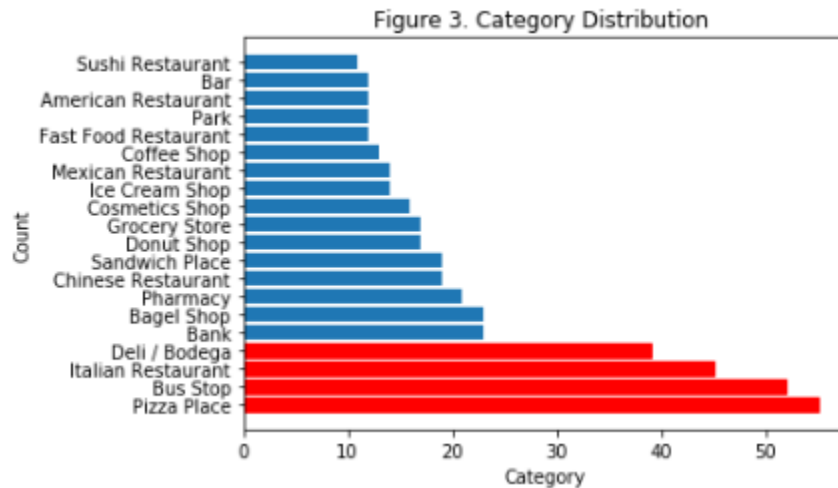
To better understand the neighborhoods, I used Foursquare API to explore the recommended venues around these neighborhoods in Staten Island. There are 825 venues were found out to get the relative information as Table 1 showed. However, there were not any recommended venues in Port Ivory and Howland Hook. For categories, there were 176 unique categories found. The most popular categories were Pizza Place (55 counts), Bus Stop (52), Italian Restaurant (45) and Deli/Bodega(39) as Figure 3 showed. To explore each neighborhood deeper, Bulls Head was the most recommended neighborhood with 42 venues as Table 2 showed.

Table 1. Snapshot of Recommended Nearby Venues Information

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	St. George	40.644982	-74.079353	Beso	40.643306	-74.076508	Tapas Restaurant
1	St. George	40.644982	-74.079353	Staten Island September 11 Memorial	40.646767	-74.076510	Monument / Landmark
2	St. George	40.644982	-74.079353	Richmond County Bank Ballpark	40.645056	-74.076864	Baseball Stadium
3	St. George	40.644982	-74.079353	A&S Pizzeria	40.643940	-74.077626	Pizza Place
4	St. George	40.644982	-74.079353	St. George Theatre	40.642253	-74.077496	Theater

Table 2. Top 10 Popular Neighborhoods

	Neighborhood	Venue Count
0	Bulls Head	42
1	Eltingville	40
2	West Brighton	37
3	Charleston	31
4	St. George	30
5	Stapleton	29
6	New Dorp	27
7	Tompkinsville	27
8	New Springville	27
9	Grant City	26



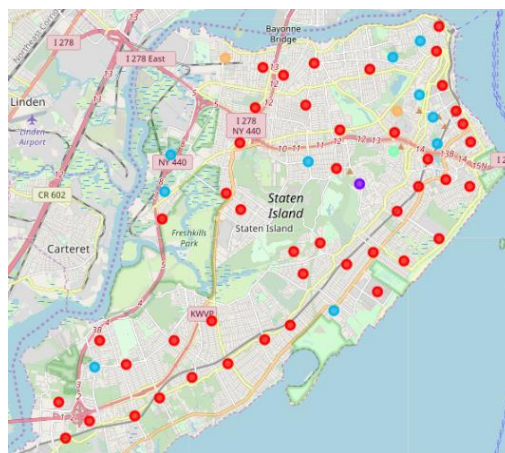
### Neighborhoods Clustering:

To research the features of these 61 neighborhoods further, I used one-hot encoding to transform all categories for venues into dummy variables and average the values of categories for each neighborhood to assess the common level of categories for each neighborhood. Finally, the top 10 common venues were chosen for each neighborhood.

Next, I used K-means clustering methods to classify all neighborhoods into 5 clusters based on their category feature and label the data. For each cluster, I compared the venue count with 10. If venue count of the specific neighborhood was less than 10, the values of extra least common venues were adjusted to null because these venues labeled by categories didn't exist actually.

Finally, cluster 1 included 47 neighborhoods, cluster 2 had 1, cluster 3 had 10, cluster 4 had 1 and cluster 5 had 2. The first cluster owned most neighborhoods with similar features. The clustering was labeled in the following map.

Figure 4. Clustering Map



## Results

For my clustering analysis, I finally got the following clusters shown sequentially from Table 3 to Table 7.

Table 3. Snapshot of Cluster 1

Neighborhood	Venue Count	Cluster Labels	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
Bulls Head	42.0	0	Pizza Place	Bus Stop	Café	Pharmacy	Deli / Bodega	Chinese Restaurant	American Restaurant	Supermarket	Liquor Store	Diner
Eltingville	40.0	0	Pizza Place	Sushi Restaurant	Italian Restaurant	Diner	Gourmet Shop	Bank	Fast Food Restaurant	Pharmacy	Train Station	Chinese Restaurant
West Brighton	37.0	0	Bank	Coffee Shop	Italian Restaurant	Music Store	Bar	Pharmacy	American Restaurant	Café	Chinese Restaurant	Sandwich Place
Charleston	31.0	0	Cosmetics Shop	Big Box Store	American Restaurant	Mobile Phone Shop	Supermarket	Boutique	Furniture / Home Store	Diner	Sporting Goods Shop	Burger Joint
St. George	30.0	0	Clothing Store	Italian Restaurant	Bar	American Restaurant	Steakhouse	Sporting Goods Shop	Snack Place	Bus Stop	Scenic Lookout	Pizza Place
Stapleton	29.0	0	Pizza Place	Discount Store	Mexican Restaurant	Sandwich Place	Bank	Fast Food Restaurant	Spanish Restaurant	Skate Park	Sri Lankan Restaurant	Chinese Restaurant
New Springville	27.0	0	Deli / Bodega	Mobile Phone Shop	Chinese Restaurant	Pizza Place	Bagel Shop	Liquor Store	Martial Arts Dojo	Men's Store	Mexican Restaurant	Grocery Store

Table 4. Cluster 2

Borough	Neighborhood	Venue Count	Cluster Labels	1st Most Common Venue
Staten Island	Todt Hill	1.0	1	Park

Table 5. Cluster 3

Neighborhood	Venue Count	Cluster Labels	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
New Brighton	13.0	2	Bus Stop	Deli / Bodega	Park	Discount Store	Convenience Store	Playground	Bowling Alley	Home Service	Eastern European Restaurant	Flower Shop
Sandy Ground	8.0	2	Bus Stop	Intersection	Art Gallery	Home Service	Liquor Store	Market	Fast Food Restaurant	Event Space	NaN	NaN
Park Hill	6.0	2	Bus Stop	Hotel	Athletics & Sports	Coffee Shop	Gym / Fitness Center	Yoga Studio	NaN	NaN	NaN	NaN
Randall Manor	6.0	2	Bus Stop	Park	Playground	Bagel Shop	Yoga Studio	Falafel Restaurant	NaN	NaN	NaN	NaN
Willowbrook	6.0	2	Bus Stop	Chinese Restaurant	Jewish Restaurant	Pizza Place	Event Space	Food Truck	NaN	NaN	NaN	NaN
Bloomfield	5.0	2	Recreation Center	Theme Park	Discount Store	Bus Stop	Park	NaN	NaN	NaN	NaN	NaN
Chelsea	4.0	2	Steakhouse	Bus Stop	Sandwich Place	Park	NaN	NaN	NaN	NaN	NaN	NaN
Fox Hills	4.0	2	Bus Stop	Sandwich Place	Grocery Store	Yoga Studio	NaN	NaN	NaN	NaN	NaN	NaN
Grymes Hill	3.0	2	Bus Stop	Dog Run	Men's Store	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Oakwood	3.0	2	Bus Stop	Playground	Bar	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Table 6. Cluster 4

Borough	Neighborhood	Venue Count	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Staten Island	Emerson Hill	3.0	3	Moving Target	Food	Gym

Table 7. Cluster 5

Borough	Neighborhood	Venue Count	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
Staten Island	Silver Lake	4.0	4	American Restaurant	Burger Joint	Bus Stop	Golf Course
Staten Island	Arlington	4.0	4	Intersection	Bus Stop	Boat or Ferry	American Restaurant

## Discussion

Based on above analysis, I firstly chose Staten Island as the best borough in NY to open restaurants. Then, I found a list of the most popular neighborhoods by recommended venues. To explore the features of lots of neighborhoods to support the assessment for neighborhoods and find more information, I clustered the neighborhoods into 5 clusters based on the category features.

For cluster 1, there were 47 neighborhoods and most were very popular with many recommended venues. The good thing was there were a crowd of potential clients for restaurants. However, there were diversified restaurants in these locations already, which contributed to the fierce competition from the incumbents. Therefore, it's recommended to open a restaurant with more special features to differentiate from incumbents. More deeper market research should be deployed in this cluster.

For cluster 2, the only neighborhood had only one venue which was a park. A big restaurant is not a good choice to open here but a small food or drink shop could be a better choice.

For cluster 3, Bus stop and park were the main venues. Most restaurants were fast food so fast food restaurant or food truck which can let the clients pick up foods quickly and easily could be absorbing for the clients these areas. Among this cluster, New Brighton was the most popular neighborhood with 13 recommended venues.

For cluster 4, only one neighborhood with only 3 venues without significant characteristics. It's worth to do some field research to assess whether it's suitable to open restaurants here.

For cluster 5, the both neighborhoods were quite same popular although Silver Lake had owned more popular restaurant. Relatively speaking, Arlington was a better choice to open a restaurant to provide fast food for people who came here to take boats and bus, or drove by to avoid the competition from incumbents because it seemed it's a popular transportation place.

## Conclusion

All in all, different recommendations to open restaurants are given for different clusters. Combining the analysis, the recommendations are given:

- Staten Island is the best choice of borough in New York City to open small-sized restaurants because of the lower felony crime rate and higher median household income.
- The neighborhoods in Cluster 1 have lots of potential clients for restaurants but face the highly competition. Therefore, a medium- or big-sized restaurants with differential features could be good choice for these locations. Considering the popularity, Bulls Head, Eltingville and West Brighton are best choices of neighborhoods.
- Todt Hill is a good neighborhood to provide fast food, health food or drinks to people who love park or outdoor activities.
- Cluster 3 is a group of neighborhoods to open fast food restaurants or food trucks to provide people who are waiting buses. Considering popularity, New Brighton is the best choice.
- Cluster 4 & 5 are less recommended locations to open restaurants due to the lower popularity.