

Coursera Data Science Capstone

Opening a sushi restaurant in Manhattan, NYC

Introduction

Business problem: when opening a new restaurant, especially in a large city like New York, or part of it like Manhattan, a business person has a problem of determining the optimal location. Many factors influence the future success (or failure) of a restaurant: vicinity of competitors, crime rate in that neighborhood, surrounding venues, number of residents in that neighborhood, their average income etc. Taking all of these variables into account would surely provide a more precise analysis, but for this project we will only be using the surrounding competitors – e.g. the number and location of other sushi restaurants in Manhattan.

Target audience: entrepreneurs looking to open a sushi restaurant in Manhattan, or any other city or neighborhood.

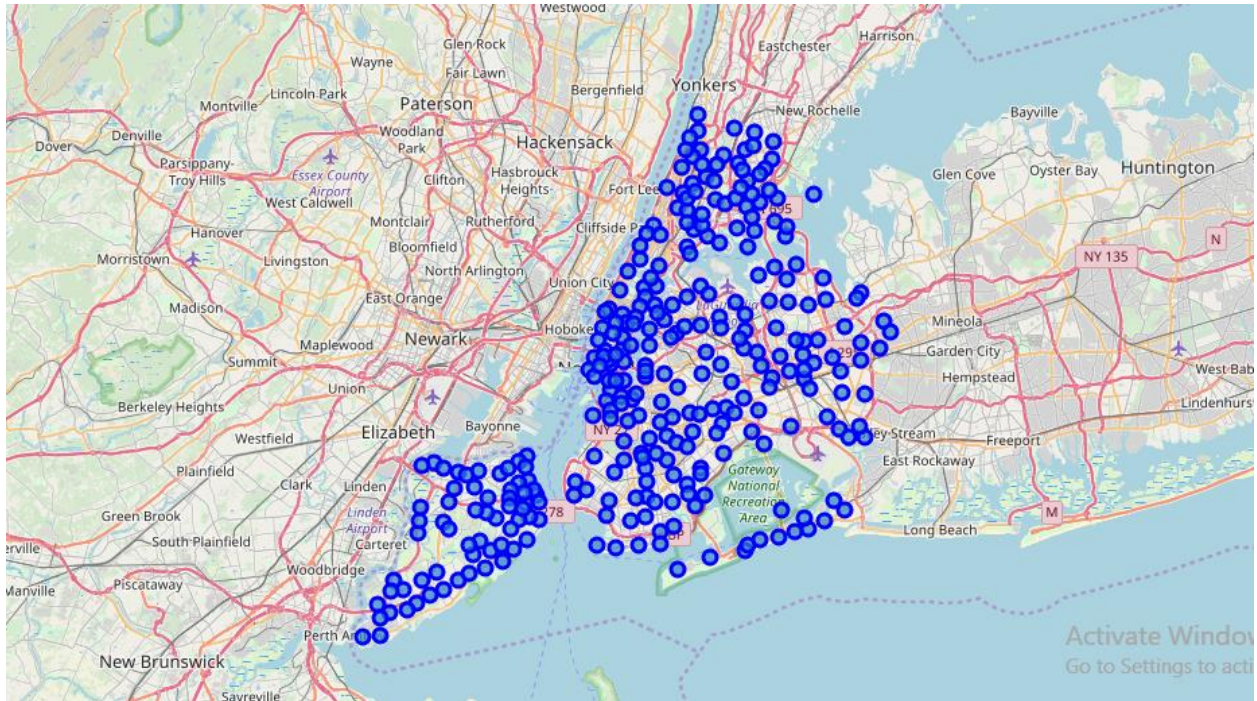
Data which we will be using is the following:

- New York City neighborhoods list (already provided externally earlier in the course),
- Geographical coordinates of NYC neighborhoods,
- Foursquare API location data on most popular NYC (and Manhattan) venues and restaurants.

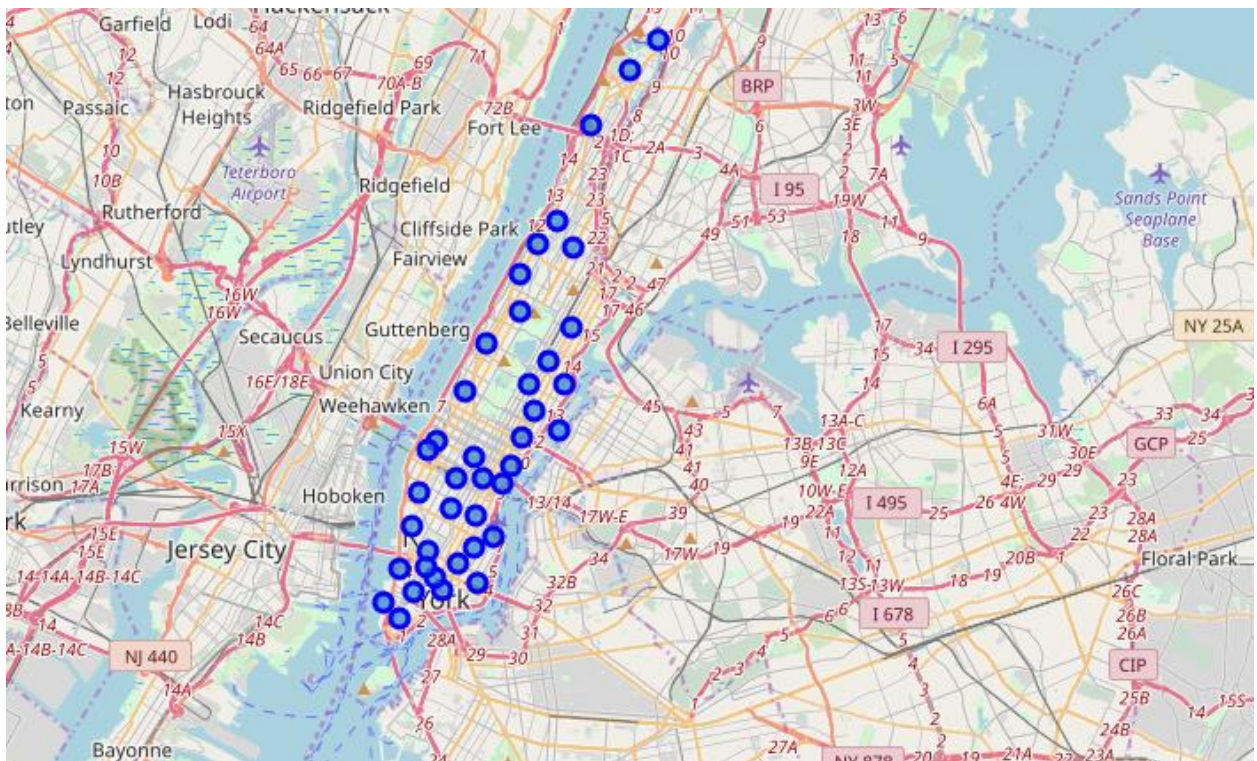
Methodology

The first thing I did was import all of the libraries needed to work with datasets, map visualizations and Foursquare. Then I downloaded the NYC neighborhoods dataset. Since this dataset was in JSON format, I had to make a new dataframe with the columns *Borough*, *Neighborhood*, *Latitude*, *Longitude*. The good thing about this dataset is that it already contains geographical coordinates, so we don't have to deal with that.

To get a better picture of the dataset, I mapped the neighborhoods using Folium:



Next, I concentrated only on Manhattan by keeping only Boroughs that are equal to Manhattan in the dataset. I also mapped the neighborhoods of Manhattan:



Now that I have my Manhattan dataset, I am ready to use Foursquare to fetch location data on Manhattan and merge it with my dataset.

The resulting dataset has 335 unique categories and venues that go beyond restaurants:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

Results

After one-hot encoding the dataset, I further narrowed the dataset only to sushi restaurants, because that's what our analysis is about. I clustered the neighborhoods into 5 clusters, and merged everything into one dataset:

	Cluster Labels	Neighborhood	Sushi Restaurant	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	0	Battery Park City	0.01	40.711932	-74.016869	Battery Park City Esplanade	40.711622	-74.017907	Park
0	0	Battery Park City	0.01	40.711932	-74.016869	Equinox Brookfield Place	40.712704	-74.014995	Gym
0	0	Battery Park City	0.01	40.711932	-74.016869	Hudson Eats	40.712803	-74.016102	Food Court
0	0	Battery Park City	0.01	40.711932	-74.016869	Waterfront Plaza, Brookfield Place	40.713241	-74.016241	Plaza
0	0	Battery Park City	0.01	40.711932	-74.016869	Brookfield Place (BFPL)	40.713240	-74.015193	Shopping Mall

Clusters 0, 1, 2, 3 and 4 are all unique in their own way:

- Cluster 0: **little to no** sushi restaurants, with neighborhoods full of gyms, pizza places, parks and stores,
- Cluster 1: **the most** sushi restaurants, with many other types of restaurants tightly grouped together (Italian, German, Thai, Vietnamese), along with hotels and coffee shops,
- Cluster 2: **medium number** of sushi restaurants, with also many other international restaurants (Italian, Japanese, Indian), dance studios, gyms, Gourmet shops, Jazz clubs – all in all this is a vibrant and upper-class part of Manhattan,
- Cluster 3: the cluster with **the least number** of sushi restaurants, with a lot of stores, coffee shops, cultural places and a few restaurants – some of them being local and some being international,

- Cluster 4: another cluster with **a high number** of sushi restaurants, along with other international restaurants, bars, gyms, spas and coffee shops.

Discussion

The clusters I got in my analysis are all different, and some are not at all good locations for a sushi restaurants, while others have the potential.

After examining all the clusters, I came to the conclusion that Cluster 3 has the smallest number of sushi restaurants, and cluster 1 has the most. If an entrepreneur wants to open a sushi place where there is not much competition, he should choose to open it in one of the neighborhoods belonging to Cluster 3, which could be Soho, for example, where many high-end restaurants exists, but also places that attract tourists such as shops, art galleries and museums.

Conclusion

Of course, the results of this analysis are limited because only factor was taken into account (the number of sushi restaurants). Other factors, such as the crime rate, population, average income etc. could make for a better analysis.