Predicting the Location for a New Sushi Restaurant in New York City

Joseph Donnelly

December, 2020

1. Introduction

1.1 Background

New York is the most populated city in the United States with an estimated population of 8.3 million residents. Composed of five boroughs, each of which is a county of the State of New York. The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island—create a most diverse city with a long tradition of different ethnical restaurants. This final project explores the best locations for Sushi restaurants throughout the city of New York. Sushi restaurants have become extremely popular, as they offer a healthy alternative to regular American cuisine. That's why potentially the owner of a new Sushi restaurant can have great success and consistent profit. However, as with any business, opening a new restaurant requires serious considerations and is more complicated than it seems from the first glance. In particular, the location of the restaurant is one of the most important factors that will affect whether it will have success or a failure. So my project will attempt to answer the questions "Where should the investor open a Sushi Restaurant in Manhattan?" and "Where should I go If I want great Sushi?"

1.2 Problem

New York is also one of the most competitive markets in the United States. A new sushi restaurant not only has to compete with other established sushi locations but also all the other cuisines offered throughout the city. The first step in founding a successful new restaurant is picking the optimum location where the business will be able to stand out against the competition and develop a lasting following.

1.3 Interest

New York offers entrepreneurs a great risk and reward scenario for opening a new venture. A successful restaurateurs would be interested in this prediction as a way to find the next location for their portfolio to diversify or for a first time developer to create a foundation for future success. Investors too would be interested in using this information to gauge the potential success or failure a new restaurant can expect in a crowded landscape.

2. Data Acquisition and Cleaning

2.1 Data Sources

In order to answer the above questions, data on New York City neighborhoods, boroughs to include boundaries, latitude, longitude, restaurants, and restaurant ratings and tips are required.

New York City data containing the neighborhoods and boroughs, latitudes, and longitudes will be obtained from the data source: https://cocl.us/new_york_dataset

All data related to locations and quality of Japanese restaurants will be obtained via the FourSquare API utilized via the Request library in Python.

2.2 Data Cleaning

Once we have our data that includes the Borough and Neighbourhoods in New York, as well as the longitude and latitude coordinates of each location. We remove all rows where the borough is unknown. We want to focus on the data that has been assigned to a borough.

Where more than one neighbourhood can exist within the same Borough, we need to combine such rows. We make a single row for each and the neighbourhoods that are within would be put into the same row and separated by commas.

If a particular neighborhood does not have a borough assigned but that row has an assigned borough, then the neighbourhood is considered to be the same as the borough.

3. Methodology

My main target here is to assess which area of the city would have the highest Sushi restaurant density. I used the Four Square API through the venues channel. I used the near query to get venues in the cities. Also, I use the CategoryID to set it to show only Sushi restaurants.

Moreover, I repeated this request and got the top venues. I saved the name and coordinate data only from the result and plotted them on the map for visual inspection.

Next, to get an indicator of the density of Sushi restaurants, I calculated a center coordinate of the venues to get the mean longitude and latitude values. Then I calculated the mean of the Euclidean distance from each venue to the mean coordinates. That was my indicator; mean distance to the mean coordinate.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

I used a geopy library to get the latitude and longitude values of New York City and create a map of New York with neighborhoods superimposed on top.
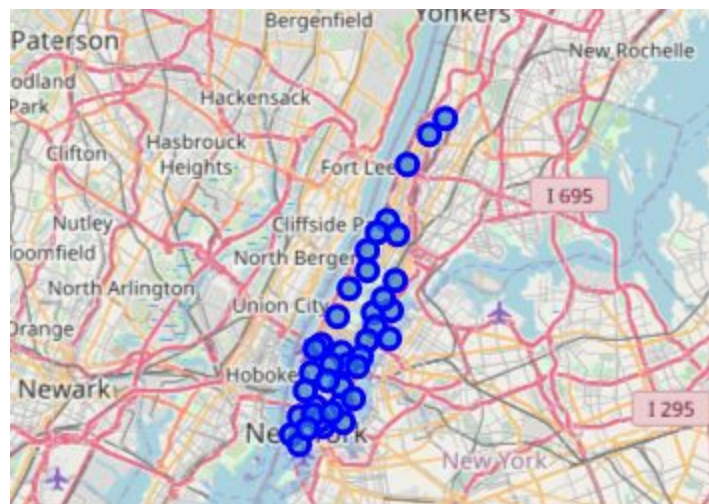
Use Foursquare API to find venues and nearby venues to identify nearby Sushi restaurants.

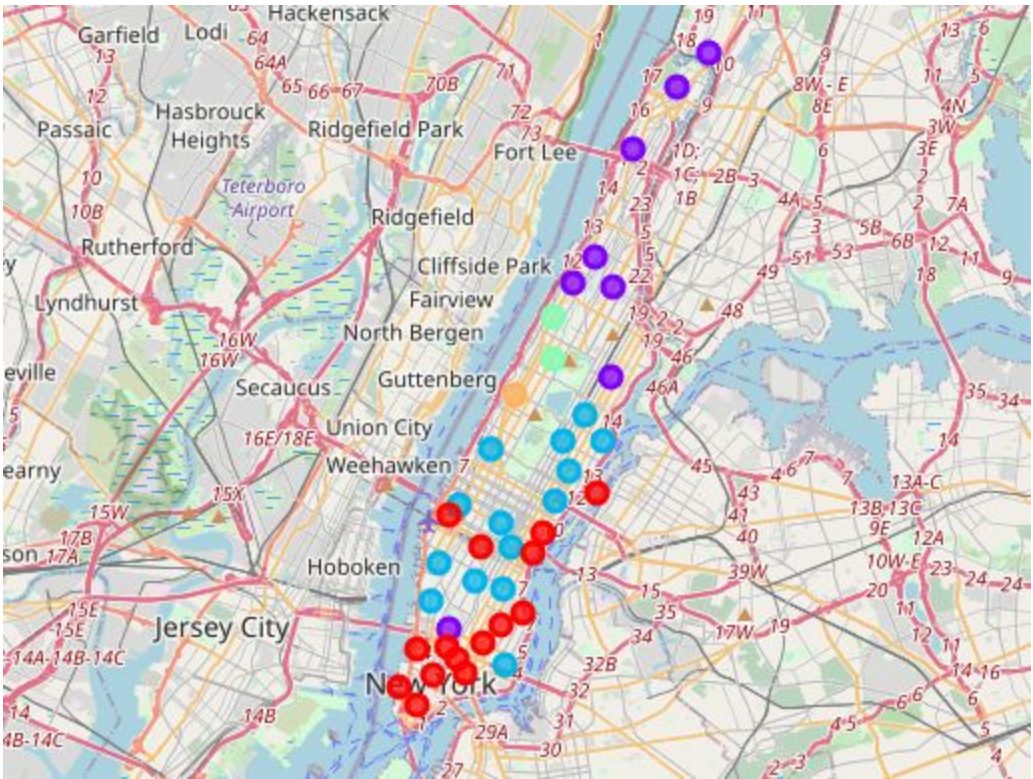|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.910660	Planet Tokyo	40.886233	-73.909479	Sushi Restaurant
1	Chinatown	40.715618	-73.994279	Nakaji	40.715912	-73.996597	Sushi Restaurant
2	Chinatown	40.715618	-73.994279	Shinsen	40.715608	-73.996611	Japanese Restaurant
3	Chinatown	40.715618	-73.994279	Bondi Bar	40.721247	-73.996264	Sushi Restaurant
4	Chinatown	40.715618	-73.994279	Sushumai Asian Fusion	40.721155	-73.987337	Sushi Restaurant

Add to Map.

Analyze each neighborhood and return Most Common Venues.

	Neighborhood	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	Battery Park City	Sushi Restaurant	Japanese Restaurant	Noodle House	Theme Restaurant	Hawaiian Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market
1	Carnegie Hill	Sushi Restaurant	Japanese Restaurant	Indian Chinese Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Smoothie Shop	Seafood Restaurant	Sake Bar	Ramen Restaurant	Noodle House
2	Central Harlem	Sushi Restaurant	Vegetarian / Vegan Restaurant	Indian Chinese Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant
3	Chelsea	Sushi Restaurant	Japanese Restaurant	Asian Restaurant	Fish Market	Vegetarian / Vegan Restaurant	Steakhouse	Smoothie Shop	Seafood Restaurant	Sake Bar	Ramen Restaurant
4	Chinatown	Sushi Restaurant	Japanese Restaurant	Indian Chinese Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant

Cluster Neighborhoods.

	Borough	Neighborhood	Latitude	Longitude	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
0	Manhattan	Marble Hill	40.876551	-73.910660	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Indian Chinese Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	0	Sushi Restaurant	Japanese Restaurant	Indian Chinese Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Indian Chinese Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Indian Chinese Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	1	Sushi Restaurant	Vegetarian / Vegan Restaurant	Indian Chinese Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Fish Market	Grocery Store	Hawaiian Restaurant

Map Clusters.

4. Results

The results of the exploratory data analysis and clustering is summarized below.

-Analyzing the data we can affirm that sushi restaurants are a very common venue in Manhattan.
-After clustering the neighborhoods of the borough we establish that cluster 1 (colour purple in the map) has the highest quantity of sushi restaurants. -Cluster 0 (colour red in the map) is the second cluster with more sushi restaurants.

-Otherwise cluster 2 (colour orange in the map) and cluster 4 (colour green in the map) have the least number of sushi restaurants.

5. Discussion

After analyzing the results we check that from Central Park to Manhattan Bridge (South Side Manhattan) is a touristic and financial area, that is the reason we can find more Sushi restaurants in this area. This area is the most important zone of Manhattan. This area includes clusters purple and red, just the two clusters with the most number of Sushi restaurants. The rest of the area of Manhattan, from Central Park to Inwood Hill Park (The north side Manhattan) is a residential area. There are not as many businesses as in the south touristic zone. Based on the result of the project an investor should open a sushi restaurant in the area of the cluster purple or in case that can not be afforded it would also be interesting to open it in the red cluster. Basically this is all the touristic and financial area where greater consumption is generated.

Conclusion:

Finally, to conclude this project, we have got a small glimpse for how real-life Data science project looks like. I have used some frequently used python libraries to handle JSON file, plotting maps, and other exploratory data analysis. Use Foursquare API to major boroughs of New York City and their neighborhoods. Potential for this kind of analysis in a real-life business problem is discussed in great detail. As a final note, all of the above analyses is depended on the adequacy and accuracy of Four Square data