

# **Capstone Project - The Battle of the Neighborhoods**

## **Analysis and Clustering of Japanese Restaurants in New York, USA**

Cassandra E. Salotti

August 29th, 2020

### **1. Introduction: Business Problem**

As of 2011, about 20,000 Japanese were living in New York City and a total of around 45,000 Japanese were living in the Greater New York City area. Most of this population consists of students, artists and expatriate business workers who are typically posted within the US for three to five years. While the Japanese in New York comprise a smaller demographic than the Chinese or Korean populations respectively, the popularity of Japanese cuisine means that there will never be a lack of demand, both from the largest Japanese community on the East Coast and from the general public, tourists and locals alike.

Japanese restaurants are naturally in demand in areas which feature a larger concentration of Japanese residents. In this project, we will try to find an optimal location to open new Japanese restaurants in New York, NY and the surrounding Greater New York Area.

I will detect the locations of already existing Japanese restaurants within and around New York (extending out to Bergen County, NJ and Long Island). This data will be used to scope out locations which have a lesser concentration of Japanese restaurants.

**Note: This project can be further expanded by clustering the Japanese population in the state of New York by city and plotting the cluster of population against cluster of restaurants. This would lead to a much more accurate prediction, however this will not be explored in this report due to time constraints.**

#### **1.1 Target Audience**

This report and its findings will be useful to anyone wanting to get a detailed overview of Japanese cuisine around New York City and for anyone who wishes to see how data is used to provide prospective small business owners and restaurateurs with the information needed to make informed decisions.

### **2. Data**

Based on definition of our problem, factors that will influence our decision are:

- Number of existing Japanese restaurants in the New York City area.

- Location (latitude and longitude) of Japanese restaurants.
- Concentration of Japanese population in New York and neighborhoods around New York. (again, not expanded upon due to time constraints)

## 2.1 Data Sources

The following data sources were used to extract/generate the required information:

- Geo coordinates of the cities of New York, NY and Scarsdale, NY were obtained using **Geopy Nominatim**
- Number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API**

The Get Venue Search endpoint of Foursquare API was used to get the list of restaurants.

- *GET <https://api.foursquare.com/v2/venues/search>*

The following parameters were passed to the Foursquare API in addition to the Client ID and Client secret:

- *query = 'Japanese'*
- *categoryId = '4d4b7105d754a06374d81259' # Food (includes restaurants)*
- *ll = Latitude and Longitude of the location*
- *radius = 1000*
- *limit = 100*

**Note: Since Foursquare explore\_venue API returns only 50 results per search, I am using one more search query with the city of Scarsdale, NY to get more coverage. Scarsdale has a sizable Japanese community consisting of households with first and second generations, as opposed to communities in the New York Metropolitan area that normally consist of expatriates who are assigned to live there for only three to five years.**

## 2.2 Data Selection

The Get Venue Search endpoint of Foursquare API was used to retrieve the list of restaurants along with the location and other details. The Get/venue/search call was performed using two locations, New York (the largest metropolitan area in the US) and Scarsdale (which has a dense Japanese population outside city bounds). Scarsdale, located in Westchester County to north of the New York Metropolitan area, has a larger Japanese population due to an excellent housing market and good schools. Therefore, I wanted to expand our search by doing a separate venue/search calls using Scarsdale as an additional location. This helped to expand our input dataset.

The initial intention was to also use a dataset which included Japanese populations in various cities in the state of New York. This would help to identify the clusters of locations where the Japanese population in New York is the densest. But such an analysis could not be performed

due to the need to meet the deadline for this project. I decided instead to call on personal experience (I have lived in Bergen County, NJ all my life) to address the Japanese population clustering.

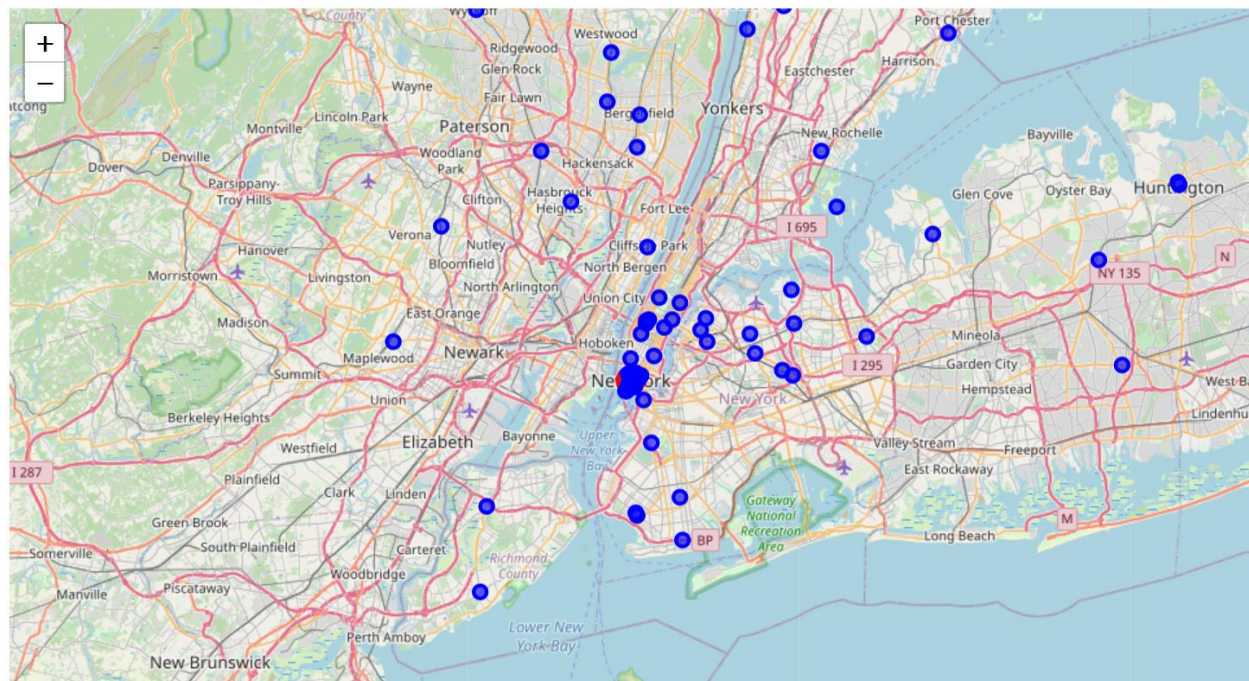
## 2.3 Data Cleaning

The results of Foursquare API calls using New York and Scarsdale as locations were merged into one data frame for analysis. (The `json_normalize` option from pandas was used to read the json format API result into data frame). Only the required details like restaurant name, location (latitude and longitude), address were kept in the data frame and the remaining fields were dropped.

# 3. Methodology

## 3.1 Exploratory Data Analysis

The merged data frame is analyzed and is plotted on a map using the folium function. This helps to get a fair idea of the location of Japanese restaurants around New York. *(See map below)*

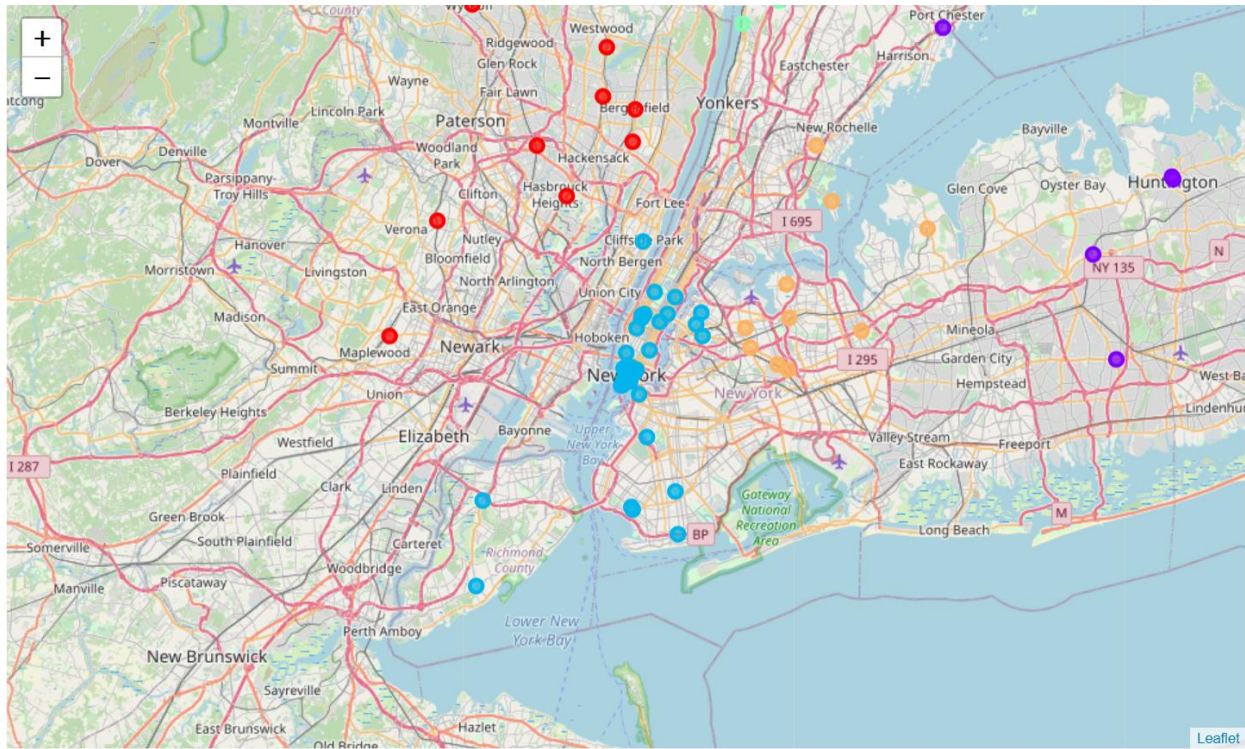


## 3.2 K-Means Clustering

Now that I have the location details for every Japanese restaurant around New York, I now have to cluster them using K-Means Clustering algorithm. This gave an idea of concentration of Japanese restaurants around the New York area. The clustering will help to identify locations with lesser density of Japanese restaurants. After trying different values of Ks, an optimum value for 5 was decided. The location (latitude and longitude) of the restaurant locations were used as the input parameters for the K-Means clustering algorithm.

The K-Means Clusters were then plotted on a map using folium for easy visualization.

*(K-Means Clusters on map)*



## 4. Results

Our analysis using the clusters map shows that there are a greater number of Japanese restaurants in Midtown and Downtown Manhattan, Staten Island, Brooklyn, Westchester County, and Bergen County NJ. This makes sense as there are established Japanese communities in these areas. There are also fair concentrations in Huntington and Port Chester, which was a pleasant surprise to me as I'm almost never out by these areas.

We also see from the map that while there is likely a notable Japanese population in Queens (centered around Flushing, Jackson Heights and Forest Hills), there seem to be very few Japanese restaurants serving this population.

## 5. Discussion

Here are some observations which warrant further discussion:

1. The borough of **Queens** is home to what appears to be a good-sized Japanese population but is lacking in Japanese restaurants.
2. **Cliffside Park NJ, Staten Island, and Brooklyn** all have a correlation of Japanese population with Japanese restaurants. Since all of these areas are easily accessible to each other via a number of bridges, tunnels and ferries, the number of restaurants here can be said to meet the demands of the population.

3. At the risk of sounding obvious, the large concentration of Japanese restaurants in **Manhattan** would normally suggest that this is the ideal locale to open a new Japanese restaurant. The Japanese food scene in Manhattan covers a broad spectrum, from Michelin-star sushi restaurants in Midtown and Tribeca to down-to-earth izakayas and ramen joints in East Village and Bowery, respectively. However, it must be stated that at the time of this report the restaurant world at large is experiencing some extraordinary circumstances. The COVID-19 pandemic is putting a massive dent in a service industry that relies heavily on tourists as well as locals. In addition, most Japanese consumers in the metropolitan area are either expatriates or college students. Since most of these consumers are either returning to Japan or learning/working remotely, the demand for restaurants that cater to these transient populations will likely slow down considerably if they haven't already. Therefore, Manhattan is decidedly not an ideal choice for opening new Japanese restaurants at this time.

## 6. Conclusion

From this in-depth analysis, it is evident that the borough of Queens could be an ideal choice to open new Japanese restaurants as this cluster has a large Japanese population, but fewer Japanese restaurants. Aside from having the highest number of neighborhoods in all five Boroughs, the northeastern and central areas of Queens are home to the largest Asian American population outside of the United States West Coast with over one-in-five residents (22.0%) of Asian descent, most of whom are intimately familiar with Japanese cuisine regardless of whether or not they identify as Japanese. With travel on hold until the pandemic is brought under control, the smartest move that potential small business owners and restaurateurs can make right now is to market to population-dense residential areas that would recognize and appreciate Japanese food well enough to support any new restaurants that crop up.