

# Clustering Airbnb and Food places in Brooklyn, NY

## 1. Introduction

### a. Background

Brooklyn has its story of diverse populations and immigrants from around the world. Immigrant communities across Brooklyn including Caribbean, Latin America, Middle East, China, and Korea (Scherer, 2015). The New York City borough is home of authentic, quality cuisine from all over the world. According to TripSavvy, Brooklyn is a heaven for foodies (Lowenstein, 2019). As a traveler who is also a foodie, planning where to stay, which has good food places nearby in Brooklyn may not be easy. Therefore, it is helpful to find a place to stay where food places by interest are nearby in advance.

### b. Problem

Using location data to find an ideal place to stay in Brooklyn. An ideal place must near food places of interest. This project aims to cluster places to stay in Brooklyn given food places of interest nearby.

### c. Interest

Amateur travelers, who are into multicultural cuisine would be interested in this project. Others who are interested in the most common cuisine in Brooklyn may be involved.

## 2. Data acquisition

### a. Data sources

There are two datasets in this project. Dataset of New York (1) can be retrieved from the previous sample project in this course. It includes neighborhoods and their location data in New York. The other dataset named 'listings.csv' (2) can be found [here](#). It includes listings of all renting places in New York on Airbnb.

### b. Data cleaning and selection

Dataset of New York (1) is used to filter neighborhoods and their locations in Brooklyn only. As a result, there are 70 neighborhoods in total.

Dataset of Airbnb listings (2) is filtered by Brooklyn borough only and its neighborhoods. The final dataset of listings includes neighborhoods, location

of each property, room type and price. Room type and price of each property are chosen following conditions:

- room\_type = Entire home/apt
- price  $\leq 70 \left( \frac{\$}{\text{night}} \right)$

room\_type and price are features in the listing dataset.

Note: room\_type is chosen based on interest while price is set by the average price per night in Brooklyn that I obtained from my last project [here](#).

### 3. Methodology

Firstly, applying Foursquare API to find latitude and longitude of Brooklyn, NY. Latitude and longitude of Brooklyn is a key factor of this project. Obtaining the location of Brooklyn helps carry out map of the borough along with its neighborhoods. Interactive maps in the project are rendered using Folium library.

Next, calling requests obtain venues data within the neighborhoods. All venues data is from category *food* which requested from Foursquare.

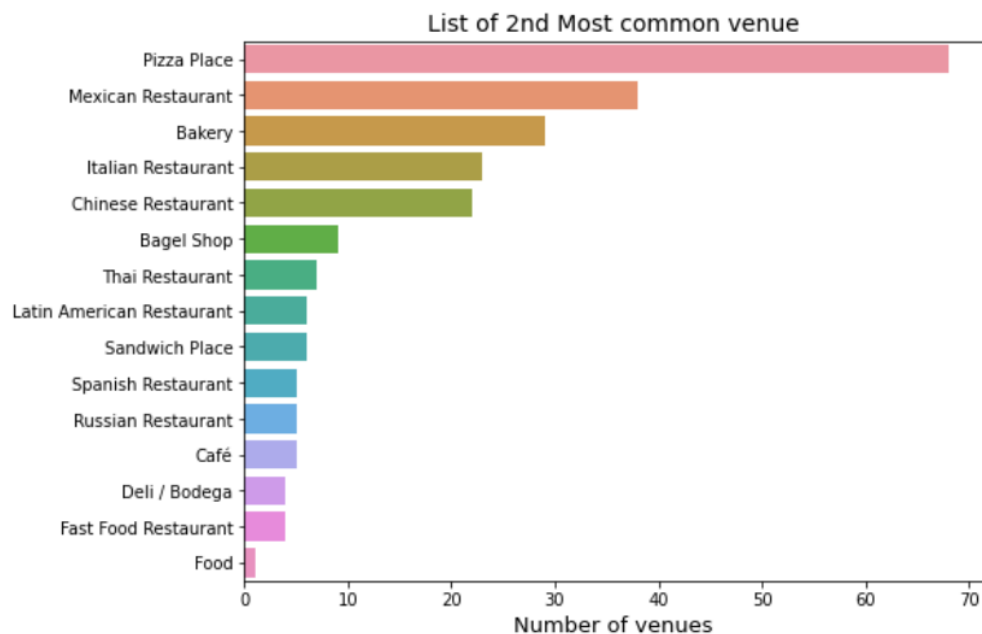
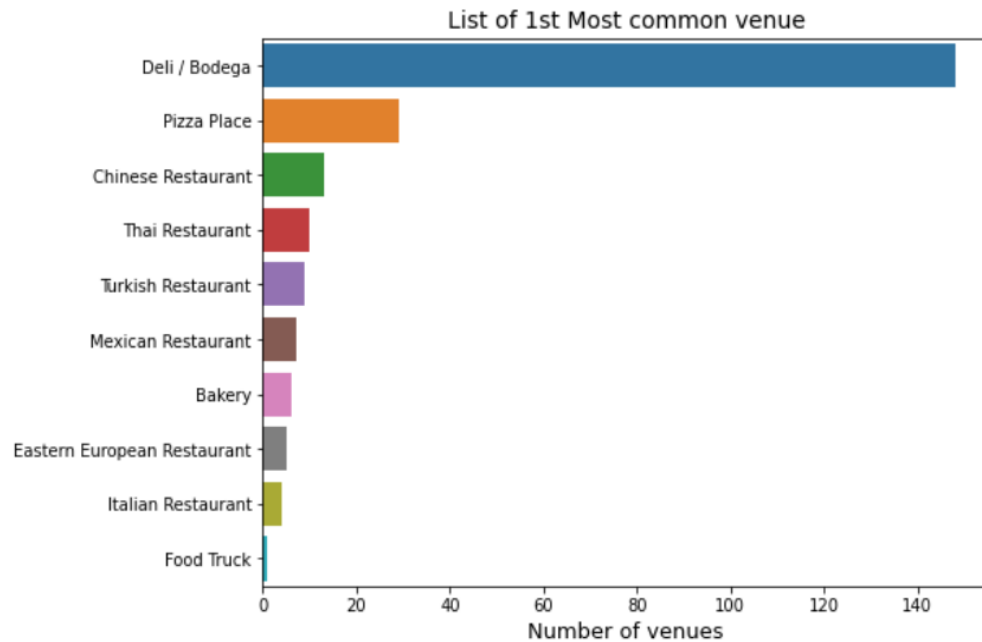
Merging the neighborhoods data, listings data and venues data results in a complete data including neighborhoods and their locations; listings and their locations; venues, venue categories and their locations.

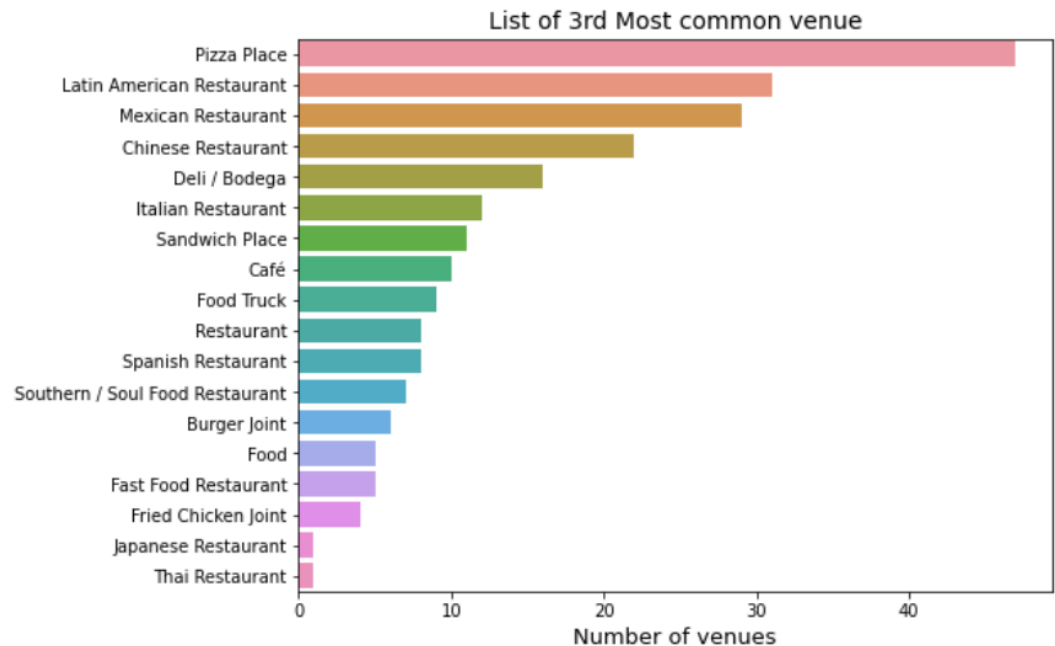
The next step is to apply one-hot encoding to get dummies values associated with each venue category. Counting the occurrence of each venue category in each neighborhood, then applies mean() method to get the frequency of occurrence of each venue category. Sorting the values of frequencies obtained in ascending order to get the most common venue category of each neighborhood.

The final step is to cluster the venue categories by location of listings. In this step, k-means clustering method is applied as the number of clusters set to 5. The result obtained 5 clusters which labeled from 0 to 4. Cluster 1 is picked since it meets the project's interest. Create an interactive map of all listings that are in cluster 1.

## 4. Results

There are five clusters of food categories. Our goal is to find food places of interest, so the cluster labeled 1 is picked. Cluster 1 results following number of venues categories in Brooklyn:





The project reaches its goal that is to find places to stay in Brooklyn which is also near food places of interest by clustering data.

## 5. Discussion

While the project is not a true challenge, however, it is a good practice in terms of working with Foursquare API, Folium library, and clustering method in machine learning. The k-means algorithm is not new to machine learning learners or practitioners. It is useful and mostly known when it comes to clustering problem. Choosing different k-clusters or the number of k-neighborhoods effectively affects the result. The project approaches the k-means method to clustering as the end goal. The project does not intentionally find the best k in terms of k-means.

## 6. Conclusion

In this project, I was able to find locations of places to rent in Brooklyn Airbnb where are also near food places of interest using Foursquare API, Folium, and machine learning approaches. I carried out interactive map for places to rent. I obtained lists of the most common food places in Brooklyn.

## References

- Lowenstein, A. (2019, 06 03). *The 10 Most Iconic Dishes to Eat in Brooklyn*. Retrieved from TripSavvy:  
<https://www.tripsavvy.com/iconic-dishes-of-brooklyn-4122555#:~:text=Brooklyn%20is%20a%20culinary%20destination,is%20a%20haven%20for%20foodies.>
- Scherer, J. (2015, 07 25). *Brooklyn's History of Immigrants*. Retrieved from Brownstoner:  
<https://www.brownstoner.com/brooklyn-life/brooklyns-history-of-immigrants/#:~:text=Today%2C%20Brooklyn%20is%20still%20home,and%20Korea%20across%20the%20borough.>