Venue Analysis for a new Italian Restaurant in Brooklyn

A. Introduction

A.1. Background

Italian cuisine is one of the most loved cuisines in the entire world. People of different nations and cultures enjoy it. It combines the health benefits with a wide choice of seasonal ingredients. Who doesn't like dishes like Pizza and Pasta! Having a famous Italian Restaurant at a fine place is a dream of many. Brooklyn is the most populous borough of New York. It offers attractions to every type of traveler. At such a place, finding a suitable place for owning a well-built and decorated restaurant is quite a task. Knowing about the people's likes and dislikes is important before starting such a venture as huge amount of money is invested in from owning a place to interiors to plenty of staffs and so on. Knowing about the competitors is important in any kind of business. Nobody wants to be at loss. Having a platform which could provide all these information would help al those who are new in business.

A.2. Problem

Data might contribute in determining a suitable place for this. Gathering all these information and having a clear and vivid idea about all the challenges in opening a new restaurant is a tedious task but it will help all those who are interested in any kind of such ventures before they decide on going on with their investment. It is always a better idea to have investment of such big budget to be at least risk. Having a restaurant in a neighborhood of high population beats a large fraction of risk. Therefore, we will try to analyze the neighborhoods of Brooklyn which are high in population and try to find out the optimized result. We will be using Foursquare location data to explore the Brooklyn and its neighbors and find an ideal location.

B. Data Description

We need information about the neighborhoods of Brooklyn borough and find the most common venues visited by people in that neighborhood. We can find this by using Foursquare. We will compare the top two most populated neighborhoods of Brooklyn. Once we get the venues, we will be able to decide which neighborhood is more ideal to open a new Italian Restaurant.

1. New York has a total of 5 boroughs and Brooklyn is one of them. In order to segment the neighborhood and explore them, a dataset is needed that contains these boroughs and the neighborhoods that exist in each borough. There is a free dataset available on web which I found on https://geo.nyu.edu/catalog/nyu_2451_34572. This dataset contains all the Boroughs, their Neighborhoods and their Latitudes and Longitudes.

2. I cleaned the data and reduced it to the borough of Brooklyn which I used to find further information about its neighborhoods. Now, it contained only the neighborhoods of Brooklyn and their respective latitudes and longitudes.

3. I found a file at https://www.worldatlas.com/articles/brooklyn-neighborhoods-by-population.html where I found a table which contains the neighborhoods of Brooklyn listed along with their populations. I used this table to make a dataframe and use it according to the problem.

4. Foursquare API is being used to explore neighborhoods of Brooklyn. To find the most common venue categories in the neighborhoods, "explore" function is used. So, the dataframe is now reduced to the venues in the chosen neighborhoods to explore the most common venue categories.

5. I have used Folium for visualizing Brooklyn and its neighborhoods on map.

Data downloaded or scraped were converted into pandas dataframes and put to use.

C. Methodology

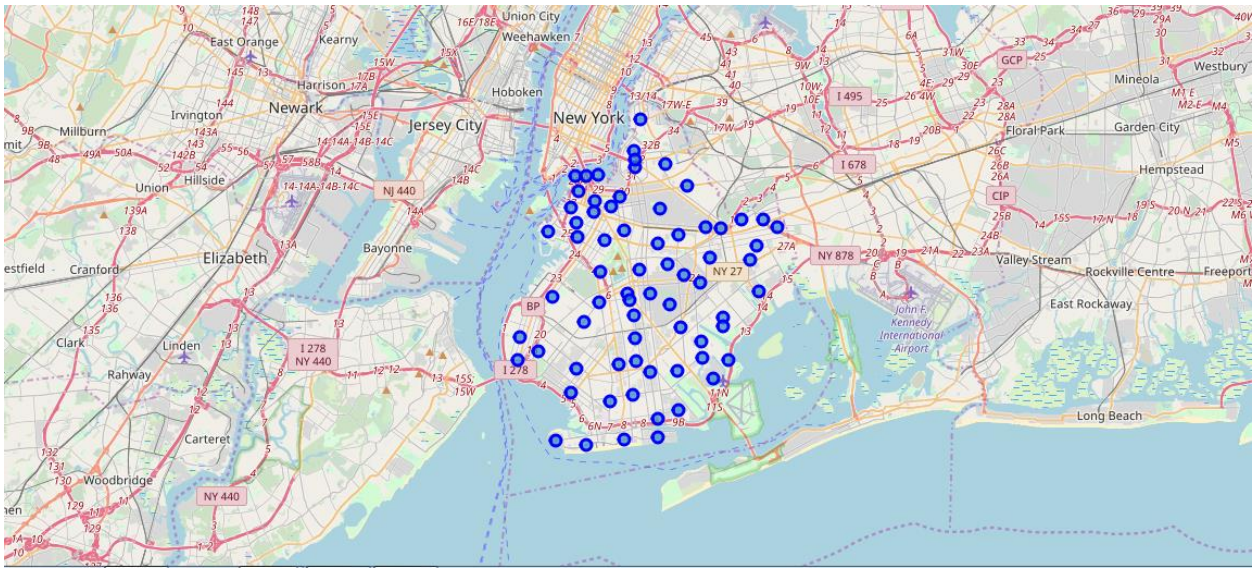
I got the dataset of Boroughs and Neighborhoods of New York from a free dataset available on internet. So, this dataframe contains all the 5 Boroughs and 306 Neighborhoods in New York. After converting it into dataframe, it looked like this:

	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

Now, as we need only the Brooklyn, I made a subset dataframe from the main one which looked like this:

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

I used geopy library to get the latitude and longitude values of Brooklyn. Folium is a great visualization library. I used Folium to create a map of Brooklyn with neighborhoods superimposed on top.



There are 36 Neighborhoods in Brooklyn. As we need one of the populated neighborhoods for our Restaurant, we needed data which could give us that information. So, we got it from a document on web, whose table got us a new dataframe which looked like this:

	Rank	Neighborhood	Population
0	1	Bedford-Stuyvesant	157530
1	2	Bensonhurst	151705
2	3	Bushwick	129239
3	4	Sunset Park	126000
4	5	Borough Park	106357

I observed that following are the two most populated neighborhoods in Brooklyn:

1. Bedford-Stuyvesant
2. Bensonhurst

So, all the next analysis is based on this information.

Now I defined the Foursquare credentials and version to explore these two neighborhoods. From the Foursquare lap, we used the following functions:

1. get_category_type
2. getNearbyVenues
3. return_most_common_venues

get_category_type and getNearbyVenues functions were used to find the following information about the two chosen neighborhoods:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bedford Stuyvesant	40.687232	-73.941785	Sincerely Tommy	40.686066	-73.944294	Boutique
1	Bedford Stuyvesant	40.687232	-73.941785	Bed-Vyne Brew	40.684751	-73.944319	Bar
2	Bedford Stuyvesant	40.687232	-73.941785	Bed-Vyne Wine & Spirits	40.684668	-73.944363	Wine Shop
3	Bedford Stuyvesant	40.687232	-73.941785	Anchor Coffee	40.684145	-73.941015	Coffee Shop
4	Bedford Stuyvesant	40.687232	-73.941785	Eugene & Co	40.683899	-73.944023	New American Restaurant

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bensonhurst	40.611009	-73.99518	il Colosseo	40.612221	-73.997436	Italian Restaurant
1	Bensonhurst	40.611009	-73.99518	Coco Nails	40.614735	-73.994620	Cosmetics Shop
2	Bensonhurst	40.611009	-73.99518	Il Fornaretto Bakery	40.613940	-73.999359	Bakery
3	Bensonhurst	40.611009	-73.99518	Bagel Town	40.610199	-73.989510	Bagel Shop
4	Bensonhurst	40.611009	-73.99518	Frank And Sal's Prime Meats	40.610232	-73.999227	Butcher

D. RESULT

For further analyzing the neighborhoods, I created a dataframe that could represent the probability of every kind of venue in the two neighborhoods:

	Neighborhood	Yoga Studio	Adult Boutique	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	Bar
0	Bedford Stuyvesant	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.035714	0.035714	0.000000	0.0	0.071429
1	Bensonhurst	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.033333	0.0	0.000000	0.033333	0.033333	0.0	0.000000

I calculated the frequency of each category-type for the top 5 Most Common Venues for each of the location so as to have an idea of the likes of the people of respective neighborhoods:

```
----Bedford Stuyvesant----
           venue  freq
0    Coffee Shop  0.11
1  Deli / Bodega  0.11
2    Pizza Place  0.07
3           Café  0.07
4            Bar  0.07


----Bensonhurst----
                venue  freq
0  Chinese Restaurant  0.10
1  Italian Restaurant  0.07
2      Ice Cream Shop  0.07
3          Donut Shop  0.07
4                 Spa  0.07
```

For the two locations, following are the observations:

1. There is already a famous and favourite Italian Restaurant in Bensonhurst and it stands 3rd in the Most Common Venues.

2. There are no special cuisine restaurants in the top most common venues for Bedford-Stuyvesant. So, it is a better option than Bensonhurst for starting a new Italian Restaurant.

3. However, we can also notice that there is a Pizza Place which is often visited by the people in the neighborhood which tells us that people are interested in Italian cuisine.

   As a result, Bedford Stuyvesant is the neighborhood in Brooklyn which the investor should choose to go on with the proposal of starting a new Italian Restaurant.

E. Discussion

New York has 5 boroughs and Brooklyn is one of them. Brooklyn is one of the most populated boroughs. It has 36 Neighborhoods. I used Folium to visualize all the neighborhoods in Brooklyn. Data from web was used to get the two most populated neighborhoods in Brooklyn: Bedford Stuyvesant and Bensonhurst.

I used get_category_type and getNearbyVenues functions of Foursquare lab to get the nearby venues and the venue-category of each of the two most populated Neighborhoods. Using a function, the top 5 Most Common Venues in each neighborhood was found.

It was observed that there is no Italian Restaurant in Bedford Stuyvesant but there was one Italian Restaurant in Bensonhurst as the 3rd most common venue. So, it can be said that it will be a fair option to start the Italian Restaurant in Bedford Stuyvesant rather than Bensonhurst.

It was also observed that though there is no Italian Restaurant in Bedford Stuyvesant, there is a Pizza place which is 3rd most common venue. Therefore, it can also be concluded that the people in the neighborhood of Bedford Stuyvesant do like Italian cuisine and so it would be ideal to start business there.

F. Conclusion

Reminding of the requirement: To find a perfect venue to start an Italian Restaurant in Brooklyn. For least risk, we found two the most populated neighborhood for further analysis. Later, it was found that one of them already has a famous Italian Restaurant, apparently the most common among the people around. So, we checked for the other one and found that it doesn't have Italian Restaurant but people do like Italian food as a Pizza Place is among the top 5 most common venues. So, we conclude that Bedford Stuyvesant is the most ideal location for the start of the business.

We can say that analyzing in the same way for different neighborhoods and the venues around and also trying to figure out the likes of people is a good way to know the optimized location for the investment in a new business at a new location.