**Selecting a successful Restaurant location in New York City**

**Dominic Fiorentino**

**July 6, 2020**

**1. Introduction**

1.1 Background

Opening a restaurant is a very risky and difficult task where a majority of newly opened restaurants will fail within their first year due to lack of planning. This is why proper planning and analysis is vital for an increased probability of success. The old adage “Location, Location, Location” could not be truer when it comes to opening a restaurant especially in the city of New York. The use of data analysis can ultimately result in a more successful business both by selecting a location with less competition and more catering opportunities based off of surrounding venues.

1.2 Problem

My company has been contacted by XYZ Hospitality group who is looking to expand their business and open a new restaurant in New York City. XYZ Hospitality Group specializes in Italian and French Restaurants and they report they have greater success when opening a restaurant when there is less competition from similar cuisines within the surrounding area. The client also reports that they have had greater success with a property that has catering opportunities nearby including event spaces, museums or large offices where lunch contracts can be created.

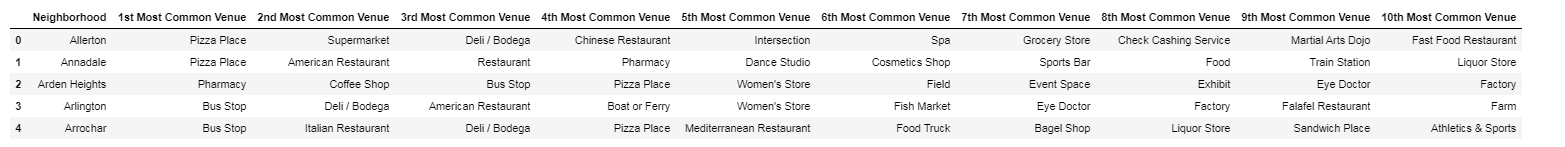
1.3 Interested Parties

The interested party in this particular case will be my client XYZ Hospitality Group and then future potential clients also looking to open restaurants in the New York City area

**2. Data acquisition and cleaning**

2.1 Data Sources

For my analysis I will be using date obtained from NYU that provides centroids of the neighborhoods throughout New York City entitled “New York City Neighborhood Names”. This dataset will be used to identify and segment neighborhoods within New York City to help identify an area of promise to open a potential restaurant in one of the five boroughs of New York City. I will then utilize foursquare data in order to identify different venues in the different neighborhoods of New York City to determine if there is significant competition in the area or if there are catering/partnering opportunities within the area that would be beneficial for the client (XYZ Hospitality group). I will cluster my data into 5 different clusters that will further demonstrate potential competition within the surrounding area that would be detrimental for the client to invest in a certain cluster.

Here is an example of the data set obtained:

The Dataset from NYU containing the New York City data can be located [here](https://geo.nyu.edu/catalog/nyu_2451_34572).

2.2 Data Cleaning and exploratory data analysis

The data was downloaded and inserted into a data frame containing the following columns Borough, Neighborhood, Latitude and Longitude. I then utilized the foursquare API to explore the Neighborhoods within New York City and separate them by venue category. I then analyzed each neighborhood by the one hot encoding method resulting in a data frame size of (10047, 429). I was then able to segment each neighborhood with the top 5 venues and ultimately decided to create a new data frame that displays the top 10 venues for each neighborhood. I then clustered all the neighborhoods into 5 separate clusters and examines each cluster to identify potential locations to invest in new restaurant venues.

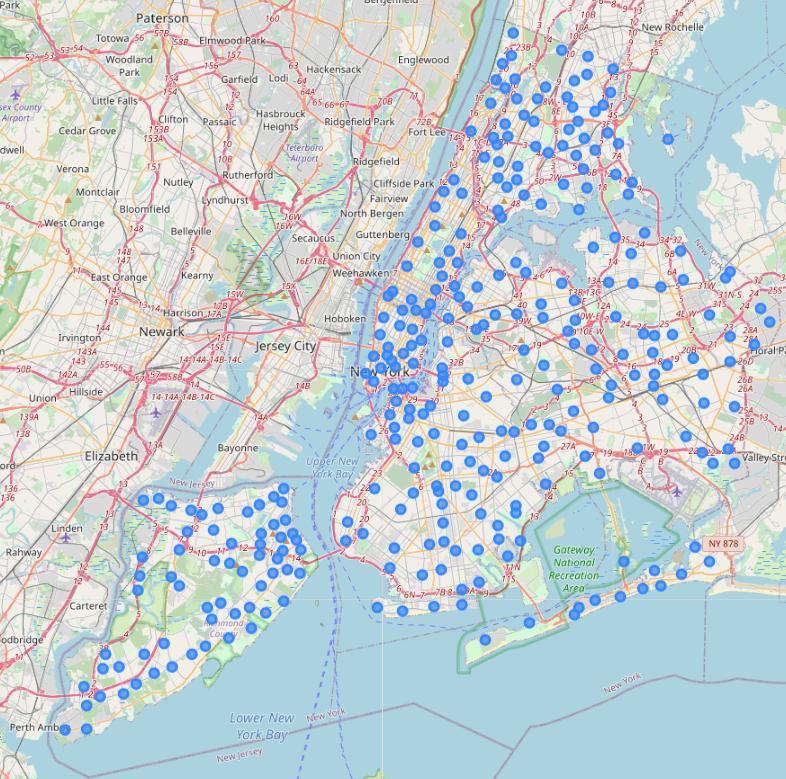


Figure - Map of the Neighborhoods of New York City - Each Dot represents a centroid of the corresponding Neighborhood

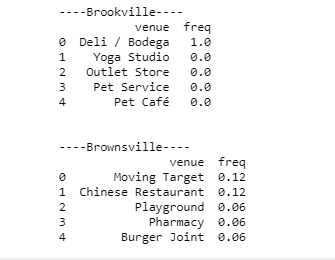


Figure – Example of the top 5 common venues of each Neighborhood

3 Results and Discussion

From the Data, two clusters were identified as having less completion (clusters 3 and 5), however it was decided that because cluster 5 had less food options and the ratio of venues to food options were greater for cluster 5 and specifically for Brookville New York. Thusly it was decided to recommend to the client that Brookville New York would be the best option to locate a restaurant space and market to the surrounding venues.

Cluster 3



Cluster 5



5 Conclusion

The Fifth Cluster identified through my analysis designates Brookville as a potential area to look for space for the potential restaurant venue due to multiple event spaces in the area and limited competition in the area as demonstrated by only having 2 out of ten common venues being food. Of those two common venues they are either falafel restaurants or deli/bodegas. There is also the increased revenue opportunities of partnering with the event and exhibit spaces for catering of events. Finally the proximity to farms which is listed as the 10th most common venue allows for an increased potential savings on food while also being able to advertise as local farm to table restaurant in New York City.