Predicting the Best Location for Arabian Restaurant in New York City

Zia Ur Rehman

April 23, 2020

1. Introduction

There is no doubt that New York City (NYC), which is also called the City of New York or some times simply New York (NY), is the most populous city in the United States. New York is also th e most densely populated major city in the United States with an estimated 2018 population of 8, 398,748 distributed over about 302.6 square miles (784 km^2). The demographics of the city show that it is a large and ethnically diverse metropolis. Starting up a new restaurant in such condense c ity will be a brilliant idea. This project will explore the best location for opening Arabian Restaur ant in New York City.

2. Data

In order to be able to find the best location for Arabian Restaurant in NYC, data on neighborhood s, boroughs, latitude, longitude, restaurants, and the most common and popular ones was needed. New York City Data containing the neighborhoods and boroughs, latitudes, and longitudes could be obtained from the data source link: https://cocl.us/new_york_dataset. On the other hand, New York City Data containing the boundaries of the neighborhoods could be obtained from the data s ource link: https://data.cityofnewyork.us/City-Government/Borough Boundaries/tqmj-j8zm. By u sing the FourSquare API via the Request library in Python, all data related to locations (GeoSpac e Data) and to Restaurants in the different neighborhoods were acquired.

3. Methodology

Data collection process was started by getting them from the following website link: “https://cocl.us/new_york_dataset”. After that, venues were found for each neighborhood by using the FourSquare API and Arabian Restaurants were attempted to be filtered out. In addition to that, some individual neighborhoods were explored to get general idea about the venues in the most popular borough. Next, the rankings of the neighborhoods using Python's Folium library were studied. Some mean values were computed to check the most common type of restaurants in the neighborhoods in general. Finally, k-means clustering algorithm were used to figure out the different prominent location to open a new Arabian Restaurant.

4. Results & Discussion

The initial data was obtained showing that there were 306 different neighborhoods in NYC. Bar chart was a proper selection to visualize and analyze the spread of these neighborhoods in different boroughs as shown in figure 1 below:

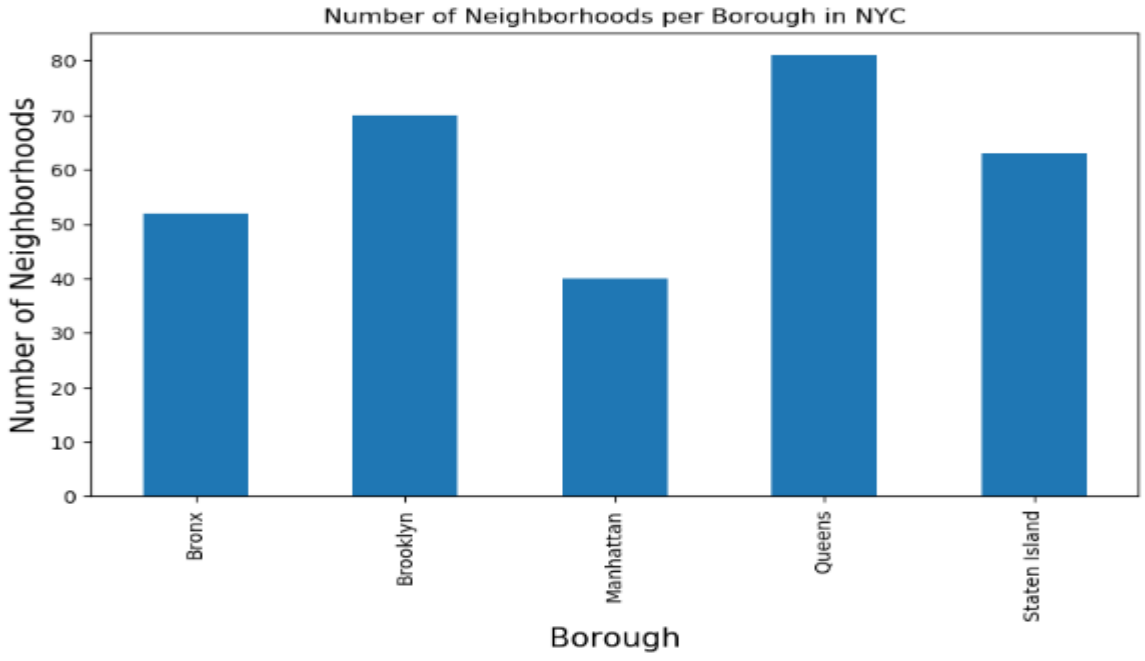

Figure 1: Neighborhoods in different boroughs in NYC

It is obvious from the bar chart shown above that Borough Queens has the highest number of nei ghborhoods (around 80) and Manhattan has the lowest (just below 40). The geographical coordin ates of New York City are 40.7127281, -74.0060152. The map of NYC with different neighborho ods is revealed below in figure 2:

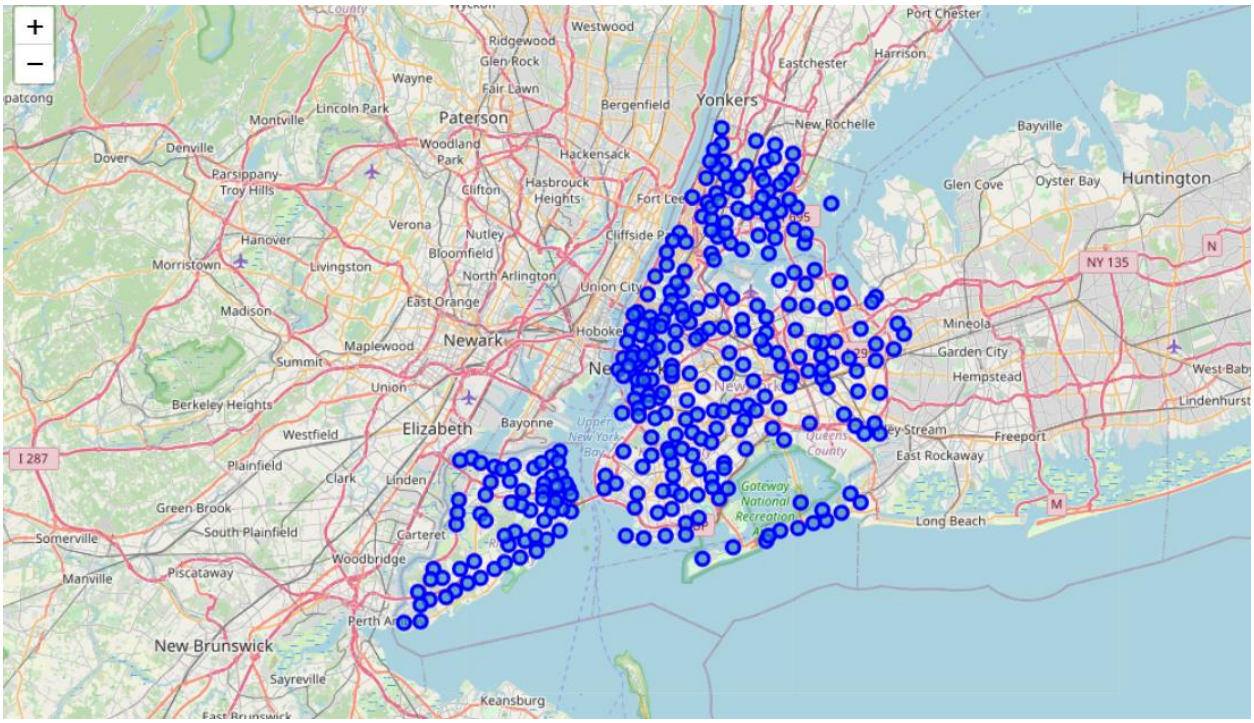

Figure 2: Map of NYC with different neighborhoods

Going further with the analysis by focusing on Queens since it is the Borough with the highest nu mber of neighborhoods. Accordingly, the original dataframe was sliced and a new dataframe was created based on the Queens Data. The different neighborhoods in Queens are shown in figure 3 b elow:

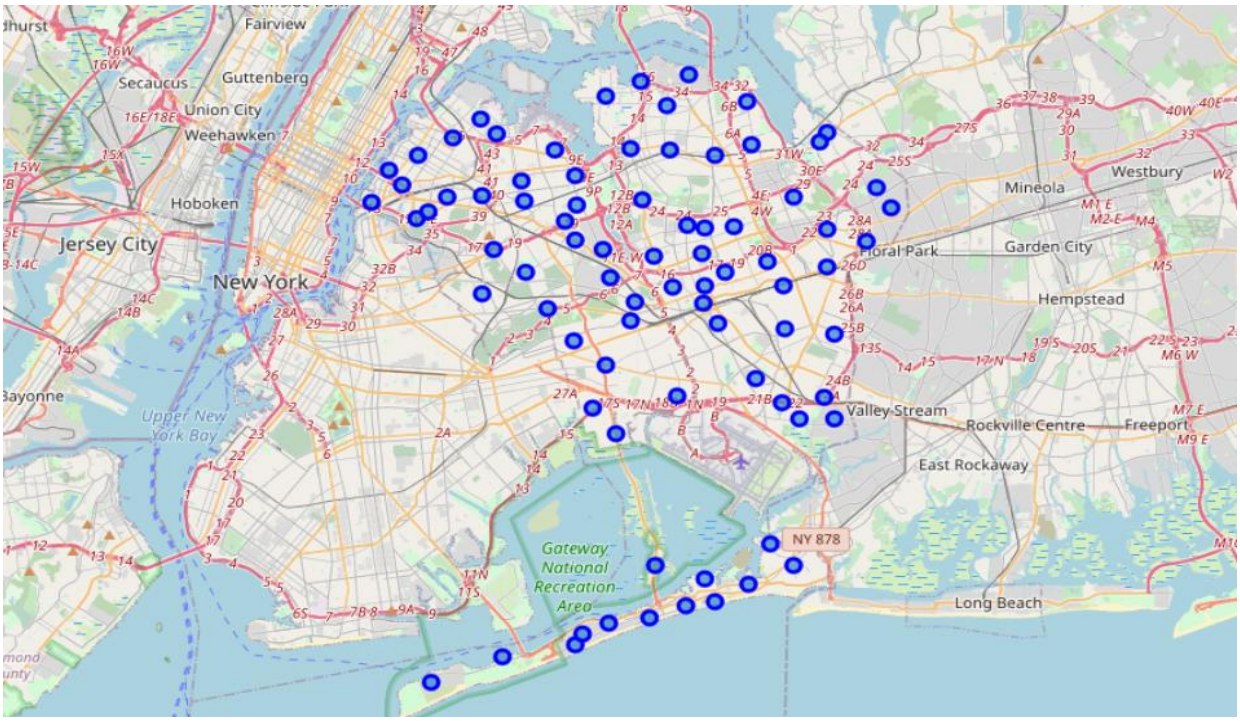

Figure 3: The different neighborhoods in Queens

After exploring and analyzing the different neighborhoods in Queens as in the jupyter notebook related to this capstone project(https://github.com/Zia-AI-Mach/Predicting-the-Best-Location-for-Arabian-Restaurant-in-New-York-City/blob/master/Predicting%20the%20Best%20Location%20for%20Arabian%20Restaurant%20in%20New%20York%20City.ipynb), it was clear that there was no Arabian Restaurant in Queens at all. This could be another attractive reason to start up a restaurant which offers Arabian Foods. However, from the mean values, it was obvious that American Restaurants are the most common and popular in different neighborhoods. It was also important to check other popular venues in order to have an idea about the whole business in different neighborhoods and the crowdedness. Checking for the best five venues in each neighborhood could provide a comprehensive approach.

We have decided to cluster the neighborhoods into five. After that we could choose the best location to open a new Arabian Restaurant. We started with examining each cluster and determining the discriminating venue categories that distinguish each cluster. Based on the defining categories, we could then assign a best location for the restaurant to each cluster. See the map of the five clusters in figure 4:

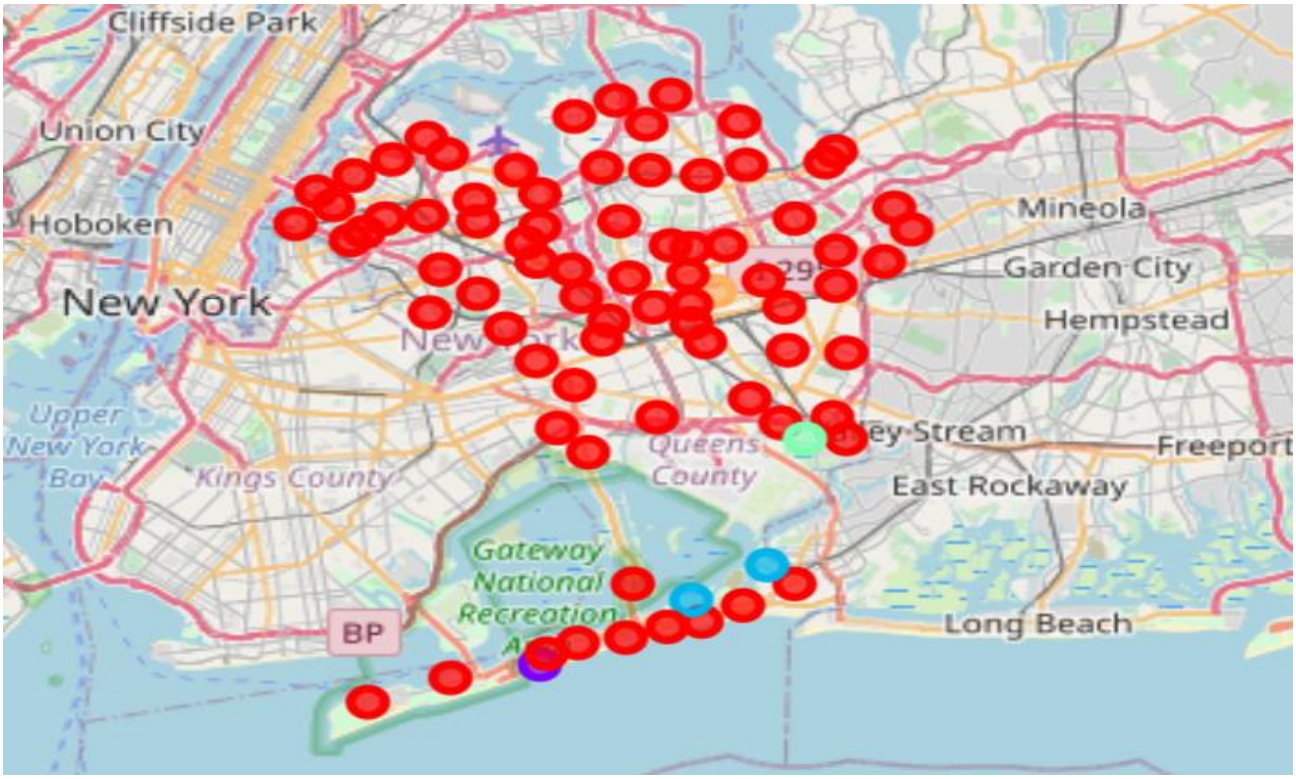

Figure 4: The five clusters of the neighborhoods in Queens

5. Conclusion

Undoubtedly, cluster one with the red color is the most condense and logically opening a new res taurant in the middle of it will be very appropriate. In fact, there should be other factors taken in c onsideration before proceeding with such establishment. It seems the location in the middle of the Queens Borough is the best as it looks very dense with dots and venues where people will be atte nding regularly and order meals.