Halal Restaurants in New York City

IBM Data Science – Capstone Project
The Battle of Neighborhood

Nugraha Syarif Hidayat
nugrahasyarif@outlook.com

June, 2020

I. Introduction

New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long story of international immigration. New York City was home to nearly 8.4 million people in 2018 [1].

New York City has also been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. With its diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, and French, etc. For Muslims specifically, that have many restrictions on their diets, it is important for Muslim to eat halal foods.

According to Institute for Social Policy and Understanding in 2018, there are around 768,767 Muslims live in New York City as of 2016, making up about 8.96% of the city's total population of 8,582,000 [2].

II. Problem

In this project, the author will solve the following problems:

- List and visualize all of the Halal restaurants available in New York City, assigned to their respective neighborhoods.
- Do cluster analysis for all of the Halal restaurants in New York City.
- Find the best three locations to visit or stay in New York City with the greatest number of Halal restaurants available within a 3km radius.

III. Data Section

In this project, we will require the following data for our analysis:

A. *New York City Geographical Data*

The data is obtained from "BetaNYC" website [3] containing boundaries of each New York City neighborhood. This New York City data has a total of 5 boroughs and 310 neighborhoods. We will acquire two sets of data from this source: first is the JSON data that defines the polygon boundaries of each neighborhood, and the second is the datasets containing a list of every neighborhood within their respective boroughs, along with coordinates (Latitude and Longitude) of their central location.

B. *Halal Restaurants data for each neighborhood in New York City*

This data is acquired from Foursquare [4] using its API service. New York City geographical coordinates for each neighborhood will be utilized as inputs for the Foursquare API and be leveraged to provision venues information for each neighborhood. The information we extract for each venue will include: venue ID, venue name, latitude and longitude for each venue, and venue address.

IV. Methodology

The first part of this project is the Exploratory Data Analysis (EDA). EDA was done to New York City geographical data and Halal Restaurant data that was acquired from Foursquare API. In New York City geographical data, there are 5 boroughs and 310 neighborhoods. With EDA we will get a list of every neighborhood within its respective borough along with the central coordinates (latitude and longitude) of each neighborhood. For Halal Restaurant data, the information acquired will include ID, name, location (latitude and longitude), and address for each venue listed in our data. The second part of this project is Clustering Analysis to find the three best locations with the largest number of Halal Restaurants within a 3km radius. The method for cluster analysis is K-means clustering. K-means clustering [5] is a method commonly used to automatically partition data set into *k* groups in which each observation belongs to the cluster with the nearest mean (cluster center or cluster centroid). The last part of this project is to identify the information about the locations from our previous analysis.

A. *Exploratory Data Analysis (EDA)*

1. New York City Geographical Coordinates Data

- Load and explore the data from NYC Neighborhoods geojson file that we obtained from BetaNYC.
- Transform the data of nested python dictionaries into pandas dataframe.

- Obtain the coordinates for every neighborhood by using Polygon library. We use 'centroid' method from Polygon library to calculate the central coordinates within neighborhood's polygon boundaries.
- Sort our dataframe by its borough code and its coordinates.
- Use folium library to create a map of New York City with markers as the center of every neighborhood and border separating each neighborhood.

2. Halal Restaurant Data
   - By using coordinates for each neighborhood, we will do venue search for every Halal restaurant category with foursquare API. The radius of the search will be 2km for each coordinate with a limit of 100 venues for each search.
   - Our initial search will result in 974 lines of unprocessed data with information such as ID, name, latitude & longitude, and address for each venue.
   - Since our search radius is 2km, there are bound to be multiple venues with the same ID being placed in multiple neighborhoods. To get the exact neighborhood for every venue, we will use geopy library and search each venue by its coordinates. We then assign neighborhood locations for every venue to the unprocessed dataframe we obtained previously. We then call this raw dataframe.
   - We begin processing the raw dataframe by first splitting it into two categories: one is where the neighborhood in every venue is listed in our main NYC Neighborhood data and the other where the neighborhood is not listed in our main NYC Neighborhood data.
   - For the first part of our split dataframe, we will drop venues with duplicate IDs and with neighborhood not placed in the correct one.
   - For the second part of our split dataframe, we first change each unique neighborhood name to its corresponding one from NYC Neighborhood data. We can check the correct neighborhood name by visualizing every venue on a map and check it with the NYC neighborhood map. After we change each neighborhood name to the correct ones, we repeat a similar process to the one we did before: dropping multiple venues with the same IDs but placed in an incorrect neighborhood.
   - We merge back our two dataframe to create our Halal Restaurant data by selecting only necessary columns. This results in 146 lines of data meaning that there are 146 Halal Restaurants in New York City.
   - Use folium library to create a map of every Halal Restaurant in New York City.
   - After we finish creating a list of venue for every Halal Restaurant in New York City, we will check the stats of each venue by using venue details request from Foursquare API. The stats include ratings, the number of likes, and the number of tips for each venue.
   - Since venue details request from Foursquare API is premium calls, which are limited to 50 calls for daily usage, we will split our Halal Restaurant data into three parts and we will use three different foursquare developer accounts.
   - After we obtained the number of likes, ratings, and tips for each venue, we then merge that data into the main Halal Restaurants data.

B. Clustering Analysis

In this part, we will do clustering analysis by using K-means clustering method.

- We start by importing KMeans module from scikit-learn library to our workspace.
- Get the column of ID, Latitude, and Longitude from Halal Restaurant data.
- Run K-means clustering to our data by setting the number of clusters to seven (7).
- Visualize the results in matplotlib and folium.

C. Finding Locations

From the map we created previously, we can conclude that there are three clusters with the most number of Halal restaurants within 3km radium from the centroid of the cluster.

- We select the top three clusters and obtain their centroid coordinates.
- From the coordinates, run geopy reverse geocoding to find the address for each cluster which includes: the name of the road, neighborhood, postal code, and borough.
- Create a dataframe for those three locations and we put the reference website for finding a place to live in each location.

V. Results and Analysis

We managed to extract 5 boroughs and 310 neighborhoods from our corresponding New York City datasets. The number of neighborhoods for each borough are shown in "Fig. 2". We also managed to get 146 different Halal restaurants in New York City, assigned to each respective neighborhood with our venue search request through Foursquare API. The number of Halal restaurants for each borough are shown in "Fig. 3".

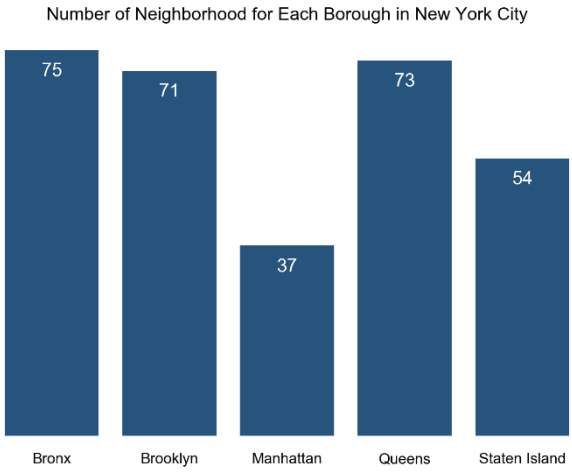

Fig 1. New York City Neighborhoods

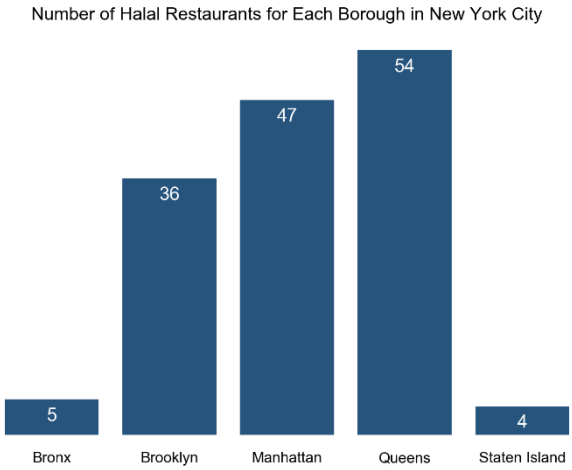

Fig 2. Halal restaurants in each Borough

The top 6 neighborhoods with the most number of Halal restaurants throughout New York City can be seen from "Fig. 4".

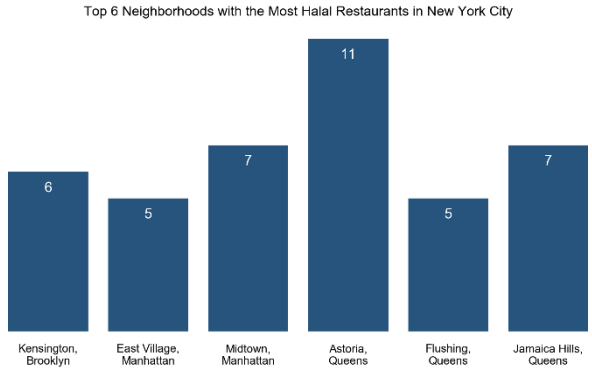

Fig 3. Top 6 Neighborhoods with the Most Halal restaurants

We can see from "Fig. 4" that the top three neighborhoods with the most Halal restaurants are Astoria, Midtown, and Jamaica Hills.

The Choropleth map visualization of the number of Halal restaurant for every neighborhood can be seen from "Fig. 5".

After doing K-means clustering, clusters of every Halal restaurants in New York City can be seen from "Fig. 6". The top three locations for the most Halal restaurants are listed in "Fig. 7".

VI. Conclusions

There are not many differences between the results from our observation using EDA and Cluster analysis regarding the best neighborhood to visit or stay if we prefer Halal cuisines.

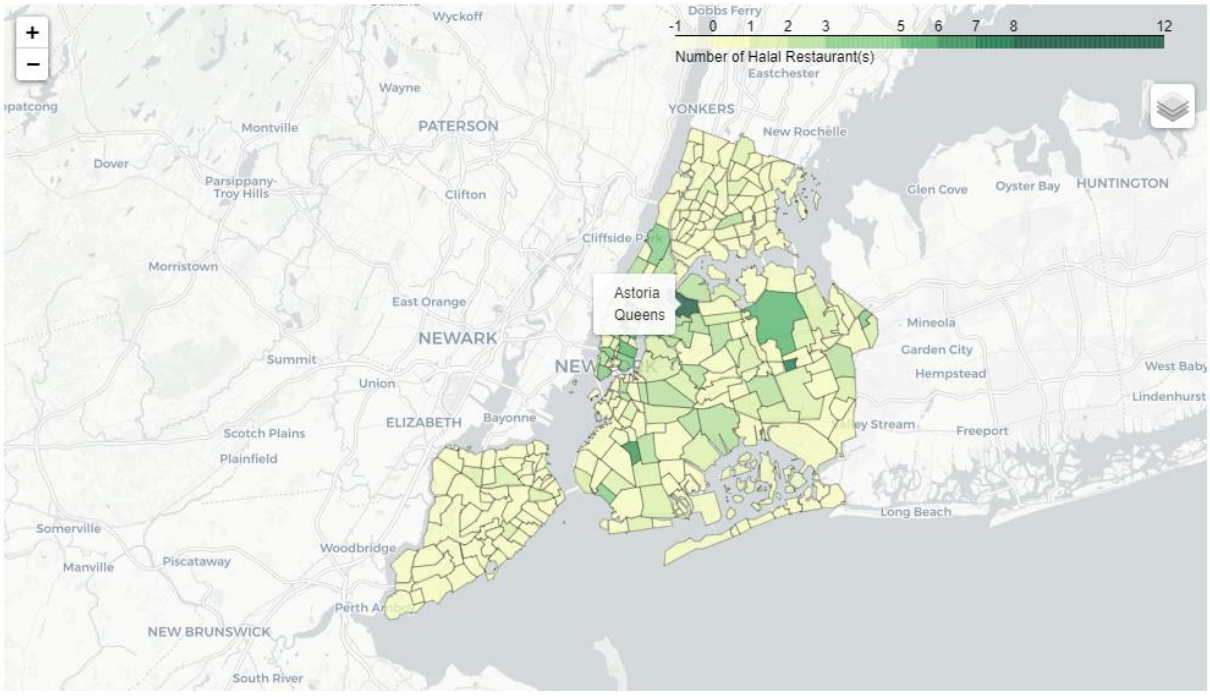

Fig 5. Choropleth Map

Fig 6. Halal restaurants clusters

From our EDA, the top three neighborhoods are Astoria Queens, Midtown Manhattan, and Jamaica Hill Queens, with the number of Halal restaurants in the neighborhood, which are 11, 7, and 7 respectively.

	Road	Neighborhood	Postal Code	Borough	Reference Website
0	1st Avenue	East Village	10009	Manhattan	https://www.redfin.com/zipcode/10009
1	Union Turnpike	Kew Gardens	11367	Queens	https://www.redfin.com/zipcode/11367
2	29th Street	Astoria	11102	Queens	https://www.redfin.com/zipcode/11102

Fig 7. Best Three Locations

But in our cluster analysis, we can pinpoint it more clearly within a 3km radius, the best locations with the most number of Halal Restaurants are Astoria 11102 NY, East Village 10009 NY, and Kew Gardens 11367 NY with 18, 28, 13 number of Halal Restaurants respectively.

VII. References

[1] "Planning-Population-Current and Future Populations - DCP," NYC Department of City Planning, [Online]. Available: https://www1.nyc.gov/site/planning/planning-level/nyc-population/current-future-populations.page.

[2] E. Becker, "Impact Report of Muslim Contributions to NYC," Institute for Social Policy and Understanding, Michigan, 2018.

[3] datHere, "NYC Neighborhoods," 2020. [Online]. Available: https://data.beta.nyc/dataset/pediacities-nyc-neighborhoods.

[4] "Foursquare Developer," Foursquare, 2020. [Online]. Available: https://developer.foursquare.com/.

[5] J. B. MacQueen, "Some method for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1: Statistics, pp. 281-297, 1967.