

## THE BATTLE OF THE NEIGHBORHOODS

### GEBRU M W CAPSTONE PROJECT WEEK-2-PART-2

#### I. INTRODUCTION

The City of New York is the most populous city in the United States. With, in 2018, population of around 8.4 million people distributed over a land area of about 302.6 square miles (784 km<sup>2</sup>)

It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States.

This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.



Figure1: New York and its five Boroughs

#### Problem Description:

A restaurant is a business which prepares and serves food and drink to customers in return for money. The City of New York is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history. So, it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied to decide on the Location.

#### Target location:

An Ethiopian company has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which borough/neighborhood of New York city will be best choice to open new Ethiopian restaurant which will serve Ethiopian food for Ethiopians and other customers who love vegan food. The Management also expects to understand the rationale of the recommendations made.

**Success criteria:**

The success criteria of the project will be a good recommendation of borough/Neighborhood choice to Ethiopian Company based on the following criteria:

1. The location should be close enough to other vegan restaurants as most Ethiopian cousins are vegan, opening the restaurant in these areas will most likely have more vegan customers and add more options for these customers.
2. The location should consider the distribution other Ethiopian restaurants in different Boroughs so that areas with less competition from other Ethiopian restaurants can be selected.
3. It should also be close enough to other amenities and subways as well as popular venues.

## II. DATA

This project will analyze one city **New York City** and the following data is analyzed to select the best location for opening the restaurant.

**Data 1:** New York has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. Link to the dataset is:

[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

**Data 2:** For the New York Population, New York City Demographics, description of Boroughs analysis we will get data from Wikipedia as given below:

[https://en.wikipedia.org/wiki/Boroughs\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Boroughs_of_New_York_City)

[https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)

[https://en.wikipedia.org/wiki/Economy\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Economy_of_New_York_City)

[https://en.wikipedia.org/wiki/Portal:New\\_York\\_City](https://en.wikipedia.org/wiki/Portal:New_York_City)

[https://en.wikipedia.org/wiki/List\\_of\\_Michelin\\_starred\\_restaurants\\_in\\_New\\_York\\_City](https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City)

**Data 3:** New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. I will construct a URL to send a request to the API to search for a specific type of venues, to explore a geographical location, and distribution of other Ethiopian restaurants in different Boroughs. Also, I will use the visualization library, Folium, to visualize the results.

<https://foursquare.com/developers/apps>

**Data 4:** The recommended Restaurant location needs to be with less completion from Ethiopian restaurants and have many shopping venues nearby. Convenient public transport is also required. These data can also be found by using Foursquare API.

<https://foursquare.com/developers/apps>

So, based on the above data I will develop a model to cluster the Boroughs/Neighborhoods and recommend a specific location for opening the new Ethiopian restaurant.

**Data cleaning**

Data downloaded from multiple sources were combined into different tables. There were a lot of missing values from these data, I have tried to clean the missing values, remove duplicates and group some of the data.

### III. Methodology and Result

All the required python libraries and packages were downloaded and imported properly. In order to use Foursquare application, I have defined my developer credentials and its Version.

Then started converting the Central New York address to its latitude and longitude coordinates. I did this Initially to get the general distribution of Ethiopian Restaurants in New York and then further identify the restaurant distribution in the five Boroughs.

I then defined a query for **Foursquare API** to search for Ethiopian Restaurant that is within 50000 meters from the Central New York City. Defined the corresponding URL and Sent the GET Request and examined the results. After getting the result I assigned relevant part of JSON to venues and transformed venues into a data frame. Next was defining information of interest and filtering the data frame. The following tasks were made on filtering the data frame:

- keeping only columns that include venue name, and anything that is associated with location
- filtering the category for each row
- cleaning column names by keeping only last term

Then I was able to visualize the Ethiopian restaurants that are nearby.

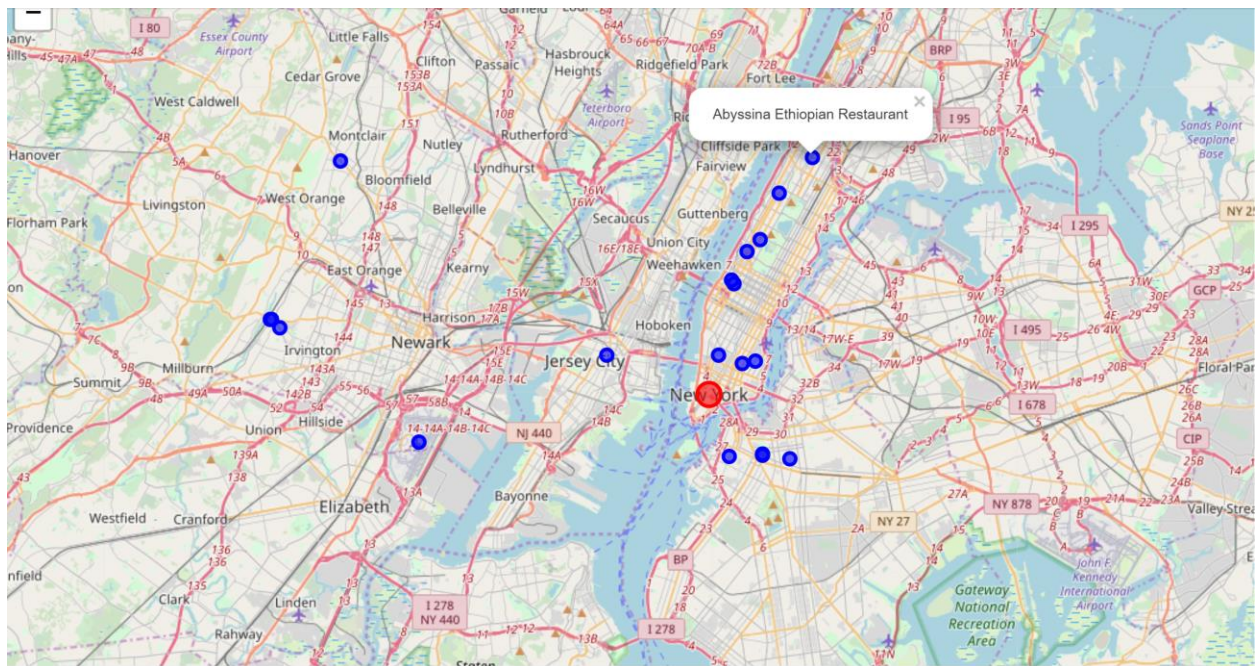


Figure2: Ethiopian restaurants around New York

After cleaning the restaurant data to drop those Ethiopian restaurants which are out of the five New York Boroughs and those duplicate Restaurants, the restaurant data was reduced from 22 to 10 restaurants.

	name	location.city
0	Meskerem Ethiopian Restaurant	New York
1	Awash Ethiopian Restaurant	Brooklyn
2	Haile Ethiopian Cuisine	New York
3	Bati Ethiopian Restaurant	Brooklyn
4	Ethiopian Meskerem Restaurant	New York
5	Bari Ethiopian Kitchen	New York
6	Nile Ethiopian Restaurant	New York
7	Abyssina Ethiopian Restaurant	New York
8	Meske	New York
9	Awash	New York

Figure 3: Ethiopian restaurants in New York

After grouping the restaurants based on their Borough, Brooklyn seem to have the least number of Ethiopian Restaurants compared to Manhattan (New York). The rest of Boroughs do not have any Ethiopian restaurants, most likely because there is no big Ethiopian community in these Boroughs hence no enough customers. So, I avoided other Boroughs to minimize associated risk and focus on Brooklyn.

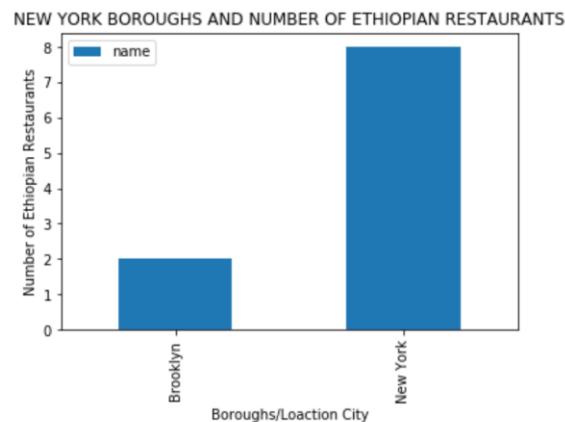


Figure 4: New York Boroughs and number of Ethiopian restaurants

### Downloading and Exploring Dataset for New York

New York has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, I need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. For My convenience, I have downloaded and used the files from Watson and simply run a wget command and access the data.

The data was loaded and transformed into a pandas data frame and grouped by Borough. this will help to get the number of Ethiopian restaurants in each Borough.



	Neighborhood	Latitude	Longitude
Borough	Bronx	52	52
	Brooklyn	70	70
	Manhattan	40	40
	Queens	81	81
	Staten Island	63	63

Figure:5 New York data Grouped in Boroughs and their Number of Neighborhoods

In order to further simplify, segment and cluster only the neighborhoods in Brooklyn, I needed to slice the original data frame and create a new data frame of the Brooklyn data. I have chosen Brooklyn because Brooklyn seem to have the least number of Ethiopian Restaurants compared to Manhattan (New York) based on the previous data from Foursquare. The rest of Boroughs do not have any Ethiopian Restaurants.

After getting the Geographical Coordinates of Brooklyn I was able to visualize its Neighborhoods.

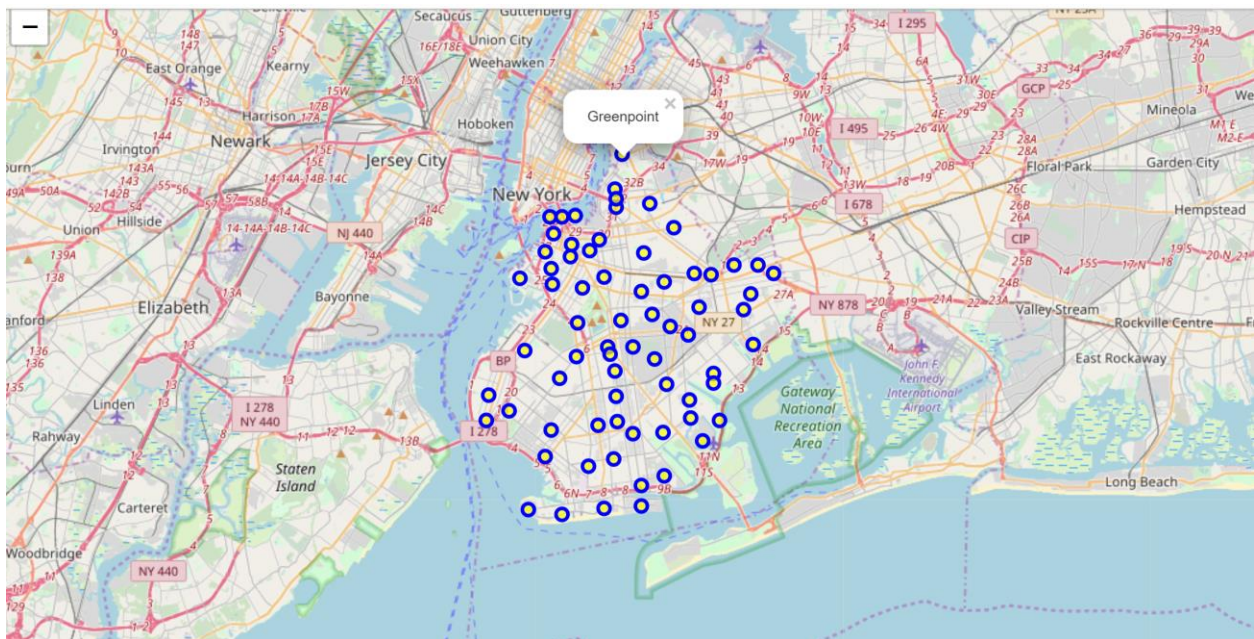


Figure6: Neighborhoods of Brooklyn

The coordinates of each Neighborhood were retrieved and the I created a function to get the venues in all the neighborhoods in Brooklyn.

```
#check the size of the resulting frame
print(Brooklyn_venues.shape)
Brooklyn_venues.head()
```

(2851, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Cocoa Grinder	40.623967	-74.030863	Juice Bar
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	Ho' Brah Taco Joint	40.622960	-74.031371	Taco Place

Figure7: Venues in all the neighborhoods in Brooklyn (only head part).

Then created the new data frame to display the top 5 venues for each neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bath Beach	Pharmacy	Pizza Place	Fast Food Restaurant	Sushi Restaurant	Chinese Restaurant
1	Bay Ridge	Italian Restaurant	Spa	Pizza Place	American Restaurant	Thai Restaurant
2	Bedford Stuyvesant	Pizza Place	Coffee Shop	Café	Deli / Bodega	Bar
3	Bensonhurst	Chinese Restaurant	Donut Shop	Sushi Restaurant	Ice Cream Shop	Italian Restaurant
4	Bergen Beach	Harbor / Marina	Athletics & Sports	Baseball Field	Playground	Donut Shop

Figure 8: Top 5 venues for each neighborhoods of Brooklyn (only head part).

## Clustering Neighborhoods

I used K means clustering and set the number of clusters to 5 and create a new data frame that includes the cluster as well as the top 5 venues for each neighborhood. The following is the table and map created to visualize the clusters in Brooklyn.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	0	Italian Restaurant	Spa	Pizza Place	American Restaurant	Thai Restaurant
1	Brooklyn	Bensonhurst	40.611009	-73.995180	4	Chinese Restaurant	Donut Shop	Sushi Restaurant	Ice Cream Shop	Italian Restaurant
2	Brooklyn	Sunset Park	40.645103	-74.010316	4	Pizza Place	Mexican Restaurant	Bakery	Latin American Restaurant	Bank
3	Brooklyn	Greenpoint	40.730201	-73.954241	0	Bar	Coffee Shop	Cocktail Bar	Pizza Place	Bakery
4	Brooklyn	Gravesend	40.595260	-73.973471	4	Pizza Place	Deli / Bodega	Spa	Bakery	Italian Restaurant

Figure:9 Clusters in Brooklyn

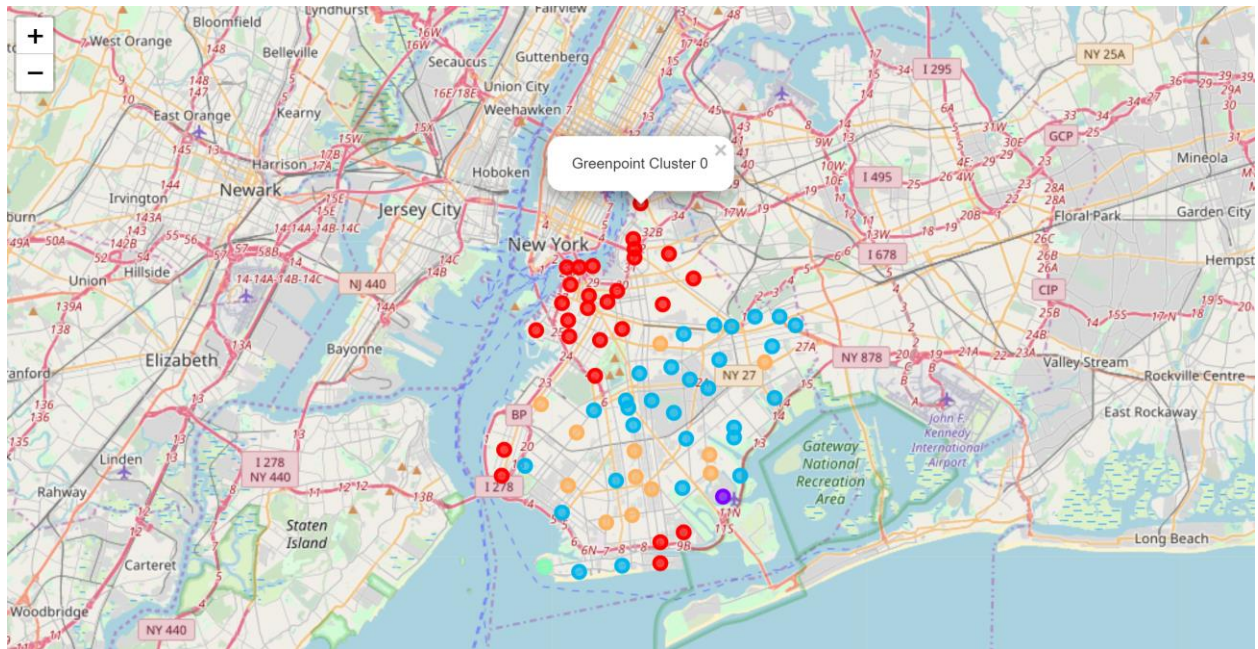


Figure:10 Neighborhoods of Brooklyn

#### IV. Discussion

In this study I used data from different sources to get the distribution of Ethiopian restaurants in New York and its Boroughs in addition to the venues in each Boroughs. Some data retrieved from Foursquare are vague, duplicate and even some business are wrongly labeled as Restaurants. I used K mean Clustering machine learning algorithm for predictive analysis and after clustering the venues in different Boroughs and based on the criteria defined by the stake holder I was able to identify the best location to open the new restaurants. The identified location has great venues and other amenities which are suited for Ethiopian restaurant and the competition from other similar restaurants is not significant. This clustering algorithm can be used to identify specific location for different purposes based on set criteria.

#### V. Conclusion

This project has used different sources of data, exploratory data analysis and algorithm to identify the best location. Cluster 0 or the first cluster seems to have the most venues in Brooklyn and will be the first choice to open the new Ethiopian Restaurant. This location can be further studied and refined using additional data and analysis.