

# **Optimal Location for a New African Restaurant in New York**

Eniola Aliu

July 17, 2020

## **1. Introduction**

### **1.1 Background**

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 history of international immigration. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.

Anyone that works in the restaurant, service, and hospitality industry knows that it's not a "get rich quick" type of career. It takes time, hard work, and dedication to be successful. Even still, a study by Ohio State University reveals that nearly 60% of restaurants close or change ownership within the first year and 80% fail within the first 5 years.

### **1.2 Problem**

Since there are lots of restaurants in New York City, we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no African restaurants in vicinity. We would also prefer locations as close to city centre as possible, if first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location.

### **1.3 Interest**

Business personnel who wants to invest or open a restaurant.

Freelancer who loves to have their own restaurant as a side business.

## **2. Data acquisition and cleaning**

### **2.1 Data sources**

For this project, we need the following data:

New York City data that contains Borough, Neighborhoods along with their latitudes and longitudes.

Data Source: <https://cocl.us/newyorkdataset>

All data related to locations and quality of Italian restaurants will be obtained via the Foursquare API utilized via the Request library in Python.

## 2.2 Data cleaning

Data will be collected from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) and cleaned and processed into a pandas dataframe.

Foursquare API will be used to locate all venues and then filtered by African restaurants. Ratings, tips, and likes by users will be counted and added to the dataframe.

Data will be sorted based on rankings.

Finally, the data be will be visually assessed graphically from Python libraries.

## 2.3 Using Foursquare Location Data:

Foursquare data is very comprehensive and its powers location data for Apple, Uber etc. For this business problem I have used, as a part of the assignment, the Foursquare API to retrieve information about the Venue, Venue category with their longitudes and latitudes. The call returns a JSON file and we need to turn that into a data-frame. Here I've chosen 100 popular spots for each neighborhoods a radius of 500 meters. Below is the data-frame obtained from the JSON file that was returned by Foursquare

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

## 3. Methodology

Data will be collected from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) and cleaned and processed into a pandas dataframe.

Use the Geospy and Folium libraries to get the coordinates of all locations and map geospatial data on a New York map.

Foursquare API will be used to collect the top 100 restaurants and their categories for each location within a 500-meter radius.

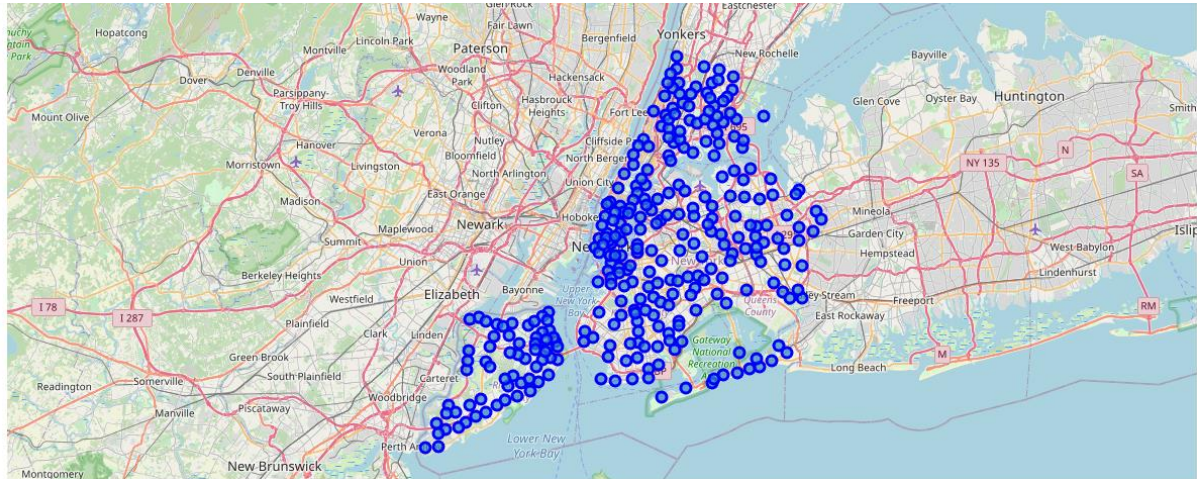
Group collected restaurants by location and by taking the mean of the frequency of the occurrence of each type, preparing them for clustering.

Cluster restaurants by k-means algorithm and analyse the top 10 most common restaurants in each cluster.

Finally, the clusters be will be visually assessed graphically on a map thus showing the best locations for opening the potential new restaurant.

### 3.1 Exploring New York City neighborhoods

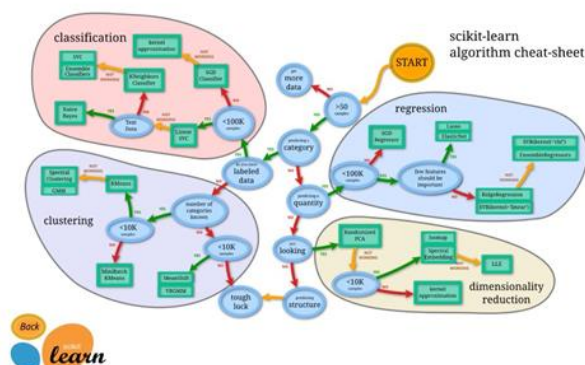
1. Using Wikipedia, we collected all of New York’s neighborhoods.
2. Using geospatial libraries, we added the geographical coordinates.
3. With some cleaning and wrangling, we obtained 288 locations in New York City.



### 3.2 Exploring New York Restaurants

We utilized the Foursquare API to get top 100 restaurants in each neighbourhood. We applied one-hot encoding and grouped them by taking the mean of the frequency of occurrence of each type.

Then we clustered restaurants using the k-means algorithm based on their types similarity. The k-means is an unsupervised machine learning algorithm for clustering unlabelled data.



## 4. Results

### 4.1 Cluster #1

[38]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	157	157	157	157	157	157	157	157	157	157	157
unique	153	53	69	75	76	74	78	76	89	81	91
top	Chelsea	Italian Restaurant	Coffee Shop	Pizza Place	Pizza Place	Bakery	Pizza Place	Bakery	Exhibit	Exhibit	Eye Doctor
freq	2	26	12	12	12	12	8	9	8	8	8

#### Chelsea Neighborhood:

The most common venue: Italian Restaurant.

The least common venue: Eye Doctor.

There are three Pizza Places on the 3rd, 4th and 6th avenues respectively.

Also, a coffee shop is on the 2nd most common venue.

No African Restaurants in this cluster.

### 4.2 Cluster #2

[40]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	2	2	2	2	2	2	2	2	2	2	2
unique	2	1	1	1	1	1	1	1	1	1	1
top	Todt Hill	Park	Women's Store	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Factory	Falafel Restaurant	Farm
freq	1	2	2	2	2	2	2	2	2	2	2

#### Todt Hill Neighborhood:

The most common venue: Park.

There's an Ethiopian Restaurant on the 3<sup>rd</sup> most common venue.

There's also another restaurant on the 9<sup>th</sup> most common venue.

The least common venue: Farm.

We can see that an African Restaurant exists on Todt Hill.

### 4.3 Cluster #3

[42]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	10	10	10	10	10	10	10	10	10	10	10
unique	10	5	8	8	9	8	6	7	8	8	8
top	Breezy Point	Beach	Beach	Deli / Bodega	Event Service	Bus Stop	Exhibit	Eye Doctor	Fabric Shop	Factory	Falafel Restaurant
freq	1	6	2	3	2	2	3	3	3	3	3

#### Breezy Point Neighborhood:

The most common venue is a Beach.

The least common venue: Falafel Restaurant.

There is no African Restaurant in this cluster.

### 4.4 Cluster #4

[44]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	1	1	1	1	1	1	1	1	1	1	1
unique	1	1	1	1	1	1	1	1	1	1	1
top	Port Ivory	Bar	Women's Store	Event Service	Event Space	Exhibit	Eye Doctor	Fabric Shop	Factory	Falafel Restaurant	Farm
freq	1	1	1	1	1	1	1	1	1	1	1

#### Port Ivory:

The most common venue: Bar.

The least common venue: Farm.

We can clearly observe a restaurant on the 9<sup>th</sup> most common venue: Falafel Restaurant.

No African Restaurant.

### 4.5 Cluster #5

[46]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
count	136	136	136	136	136	136	136	136	136	136	136
unique	136	43	52	59	68	66	73	67	75	73	72
top	Crown Heights	Pizza Place	Chinese Restaurant	Pizza Place	Pizza Place	Sandwich Place	Sandwich Place	Pizza Place	Fast Food Restaurant	Bank	Eye Doctor
freq	1	22	11	10	13	11	7	7	6	9	7

#### Crown Heights Neighborhood:

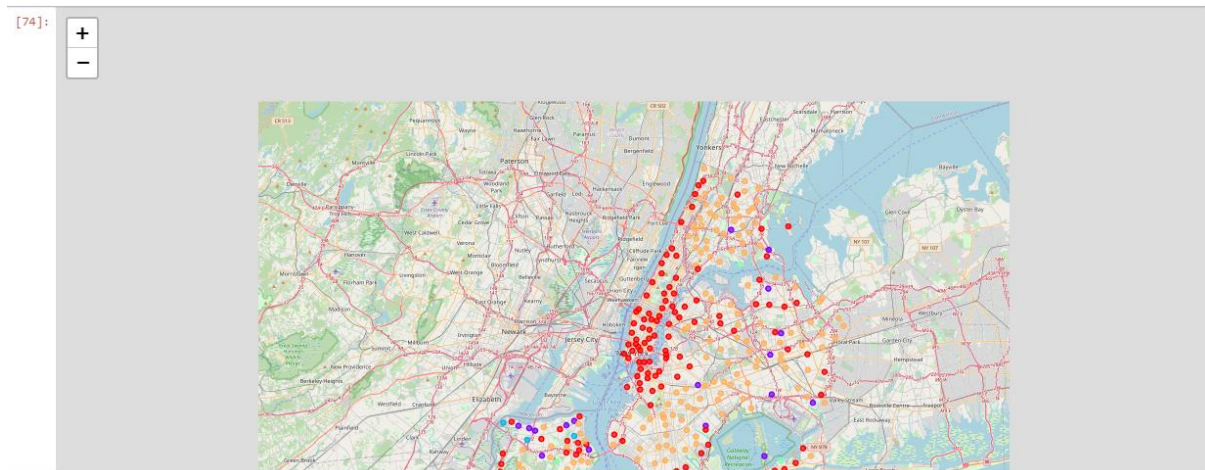
The most common venue is a Pizza Place.

The least common venue is an Eye Doctor.

There are many restaurants and pizza spots in this neighborhood.

No African restaurant in this cluster.

## 4.6 Visualizing Clusters



- Cluster 1 – red
- Cluster 2 – purple
- Cluster 3 – blue
- Cluster 4 – green
- Cluster 5 – orange

## 5. Discussion

Analysing the most popular restaurants in each cluster, the stakeholder should select the least popular types within the top 10 as a safe choice. In our recommendations, we advise selecting from the 10<sup>th</sup> or 9<sup>th</sup> positions. This selection is a reasonable balance between being too popular and having no customers.

**Recommendations**, based on description of each cluster:

Cluster 1 Neighborhood: No most or least common African Restaurant, so it's not advisable to open in Chelsea Neighborhood.

Cluster 2 Neighborhood: There is an Ethiopian Restaurant in Todt Hill which makes a good location to open an African restaurant as African customers visit this venue.

Cluster 3 to 5 neighbourhood have no African restaurant which is also a bad idea to open a new one at any of the venues.

So, cluster 2 is the ideal location to open an African restaurant.

## 6. Conclusion

In this report, we established a methodology to determine what the most promising type of restaurant is, and where it should be opened.

This type of analysis can be applied to any city of your choice that has available geospatial information.

This type of analysis can be applied to any type of venue that is available in the Foursquare database.