**Analyzing New York City Neighborhoods for the purpose of starting an African Restaurant**

**IBM Applied Data Science Capstone Project**

**By: Emmanuel Ekanem**

**Date: November 28, 2020**

# 1. INTRODUCTION

## 1.1 Background

New York City (NYC) is the capital of the state of New York and one of the most densely populated cities in the world. New York City is the most populous city in the United States, with an estimated 8,336,817 people living in the city, according to U.S. Census estimates dating July 1, 2019 (up from 8,175,133 in 2010; 8.0 million in 2000; and 7.3 million in 1990). This amounts to about 40% of New York State's population and a similar percentage of the metropolitan regional population. New York's two key demographic features are its population density and cultural diversity. According to a 2001 study by Claritas, four of the city's five boroughs ranked among the twenty most diverse counties in USA. Queens ranked 1st, Brooklyn 3rd, Manhattan 7th, and The Bronx 17th. In addition, Hudson County and Essex County, New Jersey, both of which are part of the New York Metropolitan Area, ranked 6th and 15th, respectively. NYC lies on the west coast of USA and attracts heavy tourism from all over the globe every year. Personally, I have friends and family members in New York City whom I visit from time to time, and a major part of my stay is relishing the diverse cuisines. As a major socioeconomic centre in the world, NYC is extremely diverse with people from various culture and cultural affiliations residing here. The multi-cultural nature of New York City is one reason why it has numerous cuisines from all over the world. Immigrant Africans, Caribbeans, and African Americans make up 25.1% of New York City's population. According to the U.S. Census Bureau, there were 2,086,566 black people residing in New York City.

## 1.2 Statement of Problem

Africans love food a lot, and it is often said that the way to a person's heart is through the stomach. Compared to other parts of USA, more Africans reside in NYC. So, having an African Restaurant will give not only Africans but also African Americans and other lovers of African foods; a feeling of home away from home. However, from my stay in NYC and other research, we do not have much African Restaurants there. This problem is a major need in the life of Africans, African Americans, Caribbeans and all those who love African foods in NYC as their longings for traditional African meals are not being satisfied.

### 1.3 Project Purpose

The purpose of this project is to resolve this problem through the study of NYC neighbourhoods to determine possible locations for starting an African Restaurant.

### 1.4 Project Stakeholders

The project stakeholders are all those interested in this project. They include Africans, African Americans, Caribbeans and all those who love African foods in NYC. This project will be useful for entrepreneurs and angel investors who are looking to invest and open an African Restaurant in NYC. Our project will consist of empirical data analysis to provide rewarding recommendations for these stakeholders.

### 2. DATA SECTION

Our data section consists of our data sources, data collection/cleaning and feature selection. New York City has a total of 5 boroughs and 306 neighborhoods. In order to segement the neigh borhoods and explore them, we needed a dataset that contains the 5 boroughs and the neighborho ods that exist in each borough as well as the the latitude and logitude coordinates of each neighb orhood. Our data sources include the following; https://en.wikipedia.org/wiki/New_York_City , https://en.wikipedia.org/wiki/Demographics_of_New_York_City , https://cocl.us/new_york_dataset , Python geopy library and Foursquare API data.

### 3. METHODOLOGY

We began our project by downloading all the dependencies that we need. This included installing and importing the Python libraries we need. The code is displayed below:

**1. Import Python Libraries**

```python
import pandas as pd
import numpy as np

pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import requests
from bs4 import BeautifulSoup
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
#!conda install -c conda-forge folium

import os
import folium # map rendering library
# Matplotlib and associated plotting modules

import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
import matplotlib as mp
import re
import csv
%matplotlib inline


print('Python Libraries imported.')
```

```
Python Libraries imported.
```

Then, we downloaded Wikipedia data for exploration. This was scraped using BeautifulSoup package to get population demographics and cultural diversity of NYC. The outcome of this process is displayed below:

```
print(boroughs)
```

```
[{'borough': 'The Bronx', 'county': 'Bronx', 'population': 1418207.0, 'gdp_billions': 42.695, 'gdp_per_capita': 30100.0, 'land_
sqm': 42.1, 'land_sqkm': 109.04, 'persons_sqm': 33867.0, 'persons_sqkm': 13006.0}, {'borough': 'Brooklyn', 'county': 'Kings',
'population': 2559903.0, 'gdp_billions': 91.559, 'gdp_per_capita': 35800.0, 'land_sqm': 70.82, 'land_sqkm': 183.42, 'persons_sq
m': 36147.0, 'persons_sqkm': 13957.0}, {'borough': 'Manhattan', 'county': 'New York', 'population': 1628706.0, 'gdp_billions':
600.244, 'gdp_per_capita': 368500.0, 'land_sqm': 22.83, 'land_sqkm': 59.13, 'persons_sqm': 71341.0, 'persons_sqkm': 27544.0},
{'borough': 'Queens', 'county': 'Queens', 'population': 2253858.0, 'gdp_billions': 93.31, 'gdp_per_capita': 41400.0, 'land_sq
m': 108.53, 'land_sqkm': 281.09, 'persons_sqm': 20767.0, 'persons_sqkm': 8018.0}, {'borough': 'Staten Island', 'county': 'Richm
ond', 'population': 476143.0, 'gdp_billions': 14.514, 'gdp_per_capita': 30500.0, 'land_sqm': 58.37, 'land_sqkm': 151.18, 'perso
ns_sqm': 8157.0, 'persons_sqkm': 3150.0}]
```

In [11]:
```
df = pd.DataFrame(boroughs, columns=["borough","county", "population", "gdp_per_capita", "persons_sqkm"])
df.head()
```

Out[11]:

| | borough | county | population | gdp_per_capita | persons_sqkm |
|---|---|---|---|---|---|
| 0 | The Bronx | Bronx | 1418207.0 | 30100.0 | 13006.0 |
| 1 | Brooklyn | Kings | 2559903.0 | 35800.0 | 13957.0 |
| 2 | Manhattan | New York | 1628706.0 | 368500.0 | 27544.0 |
| 3 | Queens | Queens | 2253858.0 | 41400.0 | 8018.0 |
| 4 | Staten Island | Richmond | 476143.0 | 30500.0 | 3150.0 |

**We observe the population demographics of NYC boroughs.**

```
print(jurisdictions)
```

```
[{'jurisdiction': 'Queens', 'population_census': '2,229,379', '%_white': 44.1, '%_black_or_african_amercian': 20.0, '%_Asian':
17.6, '%_other': 12.3, '%_mixed_race': 6.1, '%_hispanic_latino_of_other_race': 25.0, '%_catholic': 37.0, '%_jewish': 5.0}, {'ju
risdiction': 'Manhattan', 'population_census': '1,537,195', '%_white': 54.4, '%_black_or_african_amercian': 17.4, '%_Asian': 9.
4, '%_other': 14.7, '%_mixed_race': 4.1, '%_hispanic_latino_of_other_race': 27.2, '%_catholic': 11.0, '%_jewish': 9.0}, {'juris
diction': 'Bronx', 'population_census': '1,332,650', '%_white': 29.9, '%_black_or_african_amercian': 35.6, '%_Asian': 3.0, '%_o
ther': 25.7, '%_mixed_race': 5.8, '%_hispanic_latino_of_other_race': 48.4, '%_catholic': 14.0, '%_jewish': 5.0}, {'jurisdictio
n': 'Staten Island', 'population_census': '443,728', '%_white': 77.6, '%_black_or_african_amercian': 9.7, '%_Asian': 5.7, '%_ot
her': 4.3, '%_mixed_race': 2.7, '%_hispanic_latino_of_other_race': 12.1, '%_catholic': 11.0, '%_jewish': 5.0}, {'jurisdiction':
'NYC Total', 'population_census': '8,008,278', '%_white': 44.7, '%_black_or_african_amercian': 26.6, '%_Asian': 9.8, '%_other':
14.0, '%_mixed_race': 4.9, '%_hispanic_latino_of_other_race': 27.0, '%_catholic': 17.0, '%_jewish': 6.0}]
```

In [8]:
```
df = pd.DataFrame(jurisdictions, columns=["jurisdiction","%_white", "%_black_or_african_amercian", "%_Asian", "%_other", "%_mixed
df.head()
```

Out[8]:

| | jurisdiction | %_white | %_black_or_african_amercian | %_Asian | %_other | %_mixed_race | %_hispanic_latino_of_other_race |
|---|---|---|---|---|---|---|---|
| 0 | Queens | 44.1 | 20.0 | 17.6 | 12.3 | 6.1 | 25.0 |
| 1 | Manhattan | 54.4 | 17.4 | 9.4 | 14.7 | 4.1 | 27.2 |
| 2 | Bronx | 29.9 | 35.6 | 3.0 | 25.7 | 5.8 | 48.4 |
| 3 | Staten Island | 77.6 | 9.7 | 5.7 | 4.3 | 2.7 | 12.1 |
| 4 | NYC Total | 44.7 | 26.6 | 9.8 | 14.0 | 4.9 | 27.0 |

**We observe the multicultural nature of NYC boroughs.**

The New York dataset and python geopy library was then used to get the geographical coordinates of NYC boroughs and neighbourhoods. This data was also useful for plotting the map of NYC with its neighbourhoods superimposed on top by means of the python folium library at the end of the capstone project. The Foursquare API platform was used to extract and explore venue data. This data which contains venue recommendations for the boroughs and neighbourhoods in NYC was used to analyse African Restaurants in the different neighbourhoods and build the unsupervised learning model to cluster neighbourhoods. There was a limit of 100 venue

recommendations per neighbourhood and a radius of 1000 meters around the geographical coordinates of each neighbourhood. Data visualization techniques were used to explore various data in this project. The outcome is displayed below;

```
# get new york data
new_york_data=get_new_york_data()
```
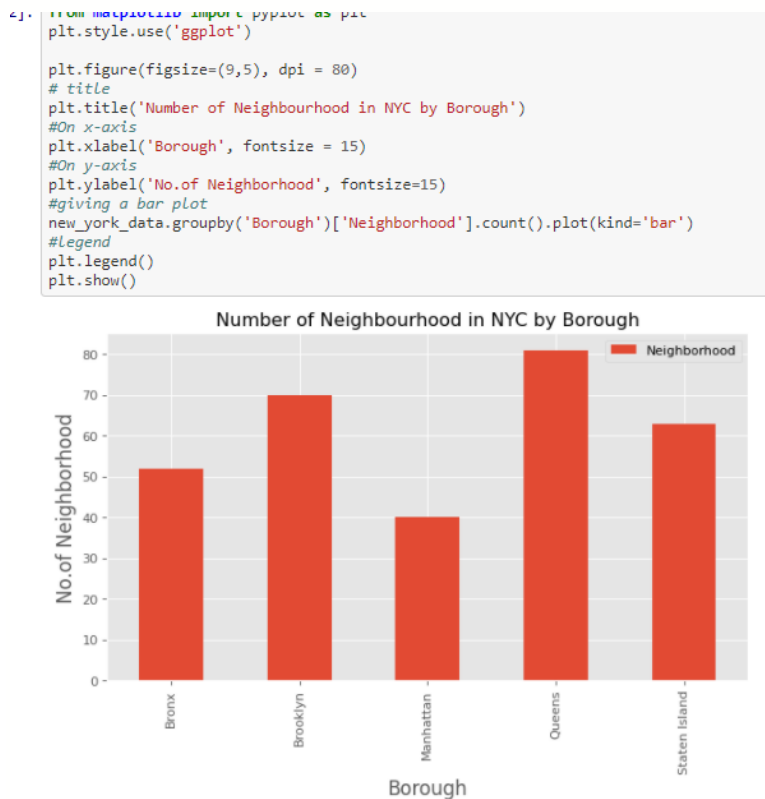
```
new_york_data.head()
```

|   | Borough | Neighborhood | Latitude | Longitude |
|---|---------|--------------|----------|-----------|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

```
new_york_data.shape
```

```
(306, 4)
```

Then, we used matplotlib to visualize the number of NYC neighbourhoods by boroughs;

```
from matplotlib import pyplot as plt
plt.style.use('ggplot')

plt.figure(figsize=(9,5), dpi = 80)
# title
plt.title('Number of Neighbourhood in NYC by Borough')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Neighborhood', fontsize=15)
#giving a bar plot
new_york_data.groupby('Borough')['Neighborhood'].count().plot(kind='bar')
#legend
plt.legend()
plt.show()
```

We then prepared a list that contains all NYC restaurants before segregating and clustering African restaurants. The code is displayed below;

```python
# prepare neighborhood list that contains African restaurant
column_names=['Borough', 'Neighborhood', 'ID','Name']
african_rest_ny=pd.DataFrame(columns=column_names)
count=1
for row in new_york_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    african_resturants=venues[venues['Category']=='African Restaurant']
    print('(',count,'/',len(new_york_data),')','African Restaurants in '+Neighborhood+', '+Borough+':'+str(len(african_resturants)))
    for resturant_detail in african_resturants.values.tolist():
        id, name , category=resturant_detail
        african_rest_ny = african_rest_ny.append({'Borough': Borough,
                                                   'Neighborhood': Neighborhood,
                                                   'ID': id,
                                                   'Name' : name
                                                  }, ignore_index=True)
    count+=1
```

{'meta': {'code': 200, 'requestId': '5fc1ed2092d28855b932bf4d'}, 'response': {'suggestedFilters': {'header': 'Tap to show:', 'filters':
[{'name': '$-$$$$', 'key': 'price'}, {'name': 'Open now', 'key': 'openNow'}]}, 'headerLocation': 'Wakefield', 'headerFullLocation': 'Wake
field, Bronx', 'headerLocationGranularity': 'neighborhood', 'totalResults': 49, 'suggestedBounds': {'ne': {'lat': 40.903705185610015, 'ln
g': -73.83531662200086}, 'sw': {'lat': 40.885705167609999, 'lng': -73.85908441909719}}, 'groups': [{'type': 'Recommended Places', 'name':
'recommended', 'items': [{'reasons': {'count': 0, 'items': [{'summary': 'This spot is popular', 'type': 'general', 'reasonName': 'globalI
nteractionReason'}]}, 'venue': {'id': '4c537892fd2ea593cb077a28', 'name': 'Lollipops Gelato', 'location': {'address': '4120 Baychester Av
e', 'crossStreet': 'Edenwald & Bussing Ave', 'lat': 40.894123150205274, 'lng': -73.84589162362325, 'labeledLatLngs': [{'label': 'displa
y', 'lat': 40.894123150205274, 'lng': -73.84589162362325}, {'label': 'entrance', 'lat': 40.89362, 'lng': -73.843737}], 'distance': 127,
'postalCode': '10466', 'cc': 'US', 'city': 'Bronx', 'state': 'NY', 'country': 'United States', 'formattedAddress': ['4120 Baychester Ave
(Edenwald & Bussing Ave)', 'Bronx, NY 10466', 'United States']}, 'categories': [{'id': '4bf58dd8d48988d1d0941735', 'name': 'Dessert Sho
p', 'pluralName': 'Dessert Shops', 'shortName': 'Desserts', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/dessert_', 's
uffix': '.png'}, 'primary': True}], 'photos': {'count': 0, 'groups': []}}, 'referralId': 'e-0-4c537892fd2ea593cb077a28-0'}, {'reasons':
{'count': 0, 'items': [{'summary': 'This spot is popular', 'type': 'general', 'reasonName': 'globalInteractionReason'}]}, 'venue': {'id':
'4d375ce799fe8eec99fd2355', 'name': 'Ripe Kitchen & Bar', 'location': {'address': '151 W Sandford Blvd', 'crossStreet': 'South 11th Avenu
e', 'lat': 40.898151685683544, 'lng': -73.8388749958446, 'labeledLatLngs': [{'label': 'display', 'lat': 40.898151685683544, 'lng': -73.8
3887499584462}, {'label': 'entrance', 'lat': 40.898292, 'lng': -73.838781}], 'distance': 798, 'postalCode': '10550', 'cc': 'US', 'neighbo
rhood': 'South Side', 'city': 'Mount Vernon', 'state': 'NY', 'country': 'United States', 'formattedAddress': ['151 W Sandford Blvd (South
11th Avenue)', 'Mount Vernon, NY 10550', 'United States']}, 'categories': [{'id': '4bf58dd8d48988d144941735', 'name': 'Caribbean Restaura
nt', 'pluralName': 'Caribbean Restaurants', 'shortName': 'Caribbean', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/car

We then converted the segregated data to reveal the boroughs, neighbourhood, ID, name, Likes, Ratings and Tips of African restaurant in NYC.

```python
african_rest_stats_ny.to_csv('african_rest_stats_ny.csv', index=False)
```

```python
african_rest_stats_ny_csv=pd.read_csv('african_rest_stats_ny.csv')
```

```python
african_rest_stats_ny_csv.shape
```

(18, 7)

```python
african_rest_stats_ny_csv.head()
```

| | Borough | Neighborhood | ID | Name | Likes | Rating | Tips |
|---|---|---|---|---|---|---|---|
| 0 | Manhattan | Marble Hill | 0 | 0 | 0 | 0.0 | 0 |
| 1 | Bronx | University Heights | 4c2f8b1b7cc0c9b6a1eaeb9a | Accra Resturant | 9 | 6.9 | 3 |
| 2 | Bronx | University Heights | 4e9308c2b8f7cece23019981 | Ebe Ye Yie | 7 | 7.5 | 2 |
| 3 | Bronx | University Heights | 4cb2beaadedbef3bca7c9d40 | Papaye Restaurant | 11 | 7.5 | 3 |
| 4 | Bronx | Fordham | 4cb2beaadedbef3bca7c9d40 | Papaye Restaurant | 11 | 7.5 | 3 |

```python
african_rest_stats_ny.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18 entries, 0 to 17
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Borough       18 non-null     object
 1   Neighborhood  18 non-null     object
 2   ID            18 non-null     object
 3   Name          18 non-null     object
 4   Likes         18 non-null     object
 5   Rating        18 non-null     object
 6   Tips          18 non-null     object
dtypes: object(7)
memory usage: 1.1+ KB
```

Our initial dataframe revealed errors of object dtype for non-object dtypes. So, we used pandas.Series.astype to convert Likes, Tips and rating object dtypes to float dtypes for ease of data analysis. The outcome is displayed below;

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18 entries, 0 to 17
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Borough       18 non-null     object
 1   Neighborhood  18 non-null     object
 2   ID            18 non-null     object
 3   Name          18 non-null     object
 4   Likes         18 non-null     float64
 5   Rating        18 non-null     float64
 6   Tips          18 non-null     float64
dtypes: float64(3), object(4)
```

Then, we analysed the African restaurants in NYC with the maximum Likes, Tips and Ratings.

**Analyze African Restaurant in NYC with maximum Likes.**

```
# African Restaurant with maximum Likes
african_rest_stats_ny.iloc[african_rest_stats_ny['Likes'].idxmax()]
```

```
Borough                       Brooklyn
Neighborhood              Crown Heights
ID               51d4d2cc454a777ed5a76c57
Name                     Cafe Rue Dix
Likes                            207
Rating                           8.6
Tips                              59
Name: 7, dtype: object
```

**Analyze African Restaurant in NYC with maximum Tips.**

```
# African Restaurant with maximum Tips
african_rest_stats_ny.iloc[african_rest_stats_ny['Tips'].idxmax()]
```

```
Borough                       Brooklyn
Neighborhood              Crown Heights
ID               51d4d2cc454a777ed5a76c57
Name                     Cafe Rue Dix
Likes                            207
Rating                           8.6
Tips                              59
Name: 7, dtype: object
```

**Analyze African Restaurant in NYC with maximum Rating.**

```
# African Restaurant with maximum Rating
african_rest_stats_ny.iloc[african_rest_stats_ny['Rating'].idxmax()]
```

```
Borough                      Manhattan
Neighborhood              East Harlem
ID               5b855d2da0215b002c09d9fa
Name                         Teranga
Likes                             19
Rating                           8.8
Tips                               7
Name: 13, dtype: object
```

Also, we explored the average rating of African restaurants in NYC boroughs. The outcome is visualized below;
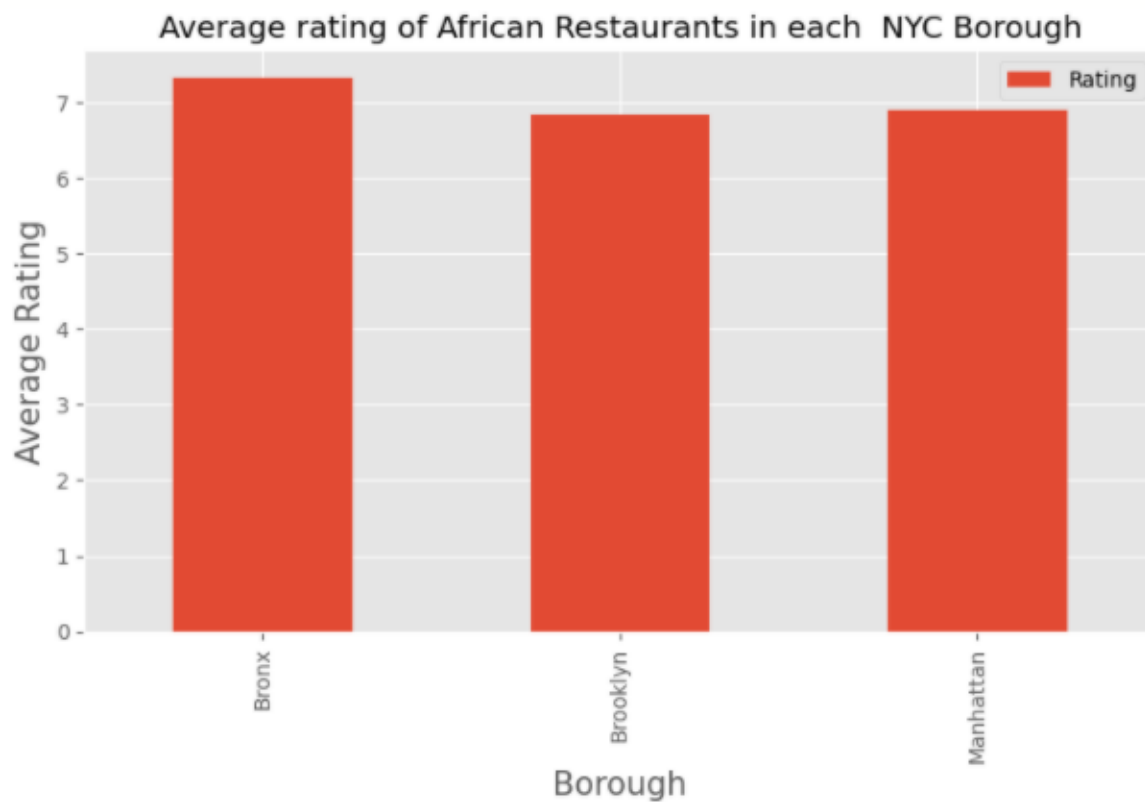
Then we explored African restaurants in NYC boroughs with ratings > 7. The result of our exploration is displayed below.

| | Borough | Neighborhood | Latitude | Longitude | Average Rating |
|---|---|---|---|---|---|
| 3 | Manhattan | East Harlem | 40.792249 | -73.944182 | 8.8 |
| 5 | Manhattan | Hamilton Heights | 40.823604 | -73.949688 | 8.7 |
| 2 | Brooklyn | Crown Heights | 40.670829 | -73.943291 | 8.6 |
| 1 | Manhattan | Central Harlem | 40.815976 | -73.943211 | 7.7 |
| 0 | Bronx | Belmont | 40.857277 | -73.888452 | 7.5 |
| 4 | Bronx | Fordham | 40.860997 | -73.896427 | 7.5 |
| 6 | Bronx | University Heights | 40.855727 | -73.910416 | 7.3 |

Finally, we used folium to visualize the neighbourhoods in New York City with the highest average rating for African Restaurants. The map is displayed below.
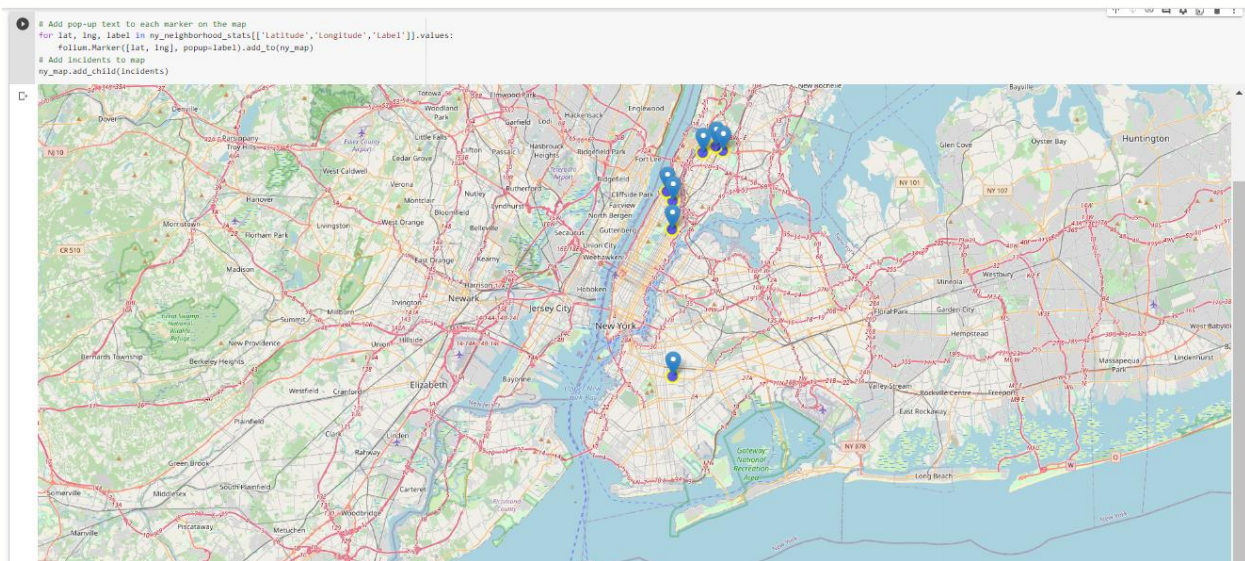


## 4. RESULTS SECTION

Our analysis of the neighborhoods and boroughs of NYC for the purpose of determining the location for a new African restaurant produced several results. Our exploration of the Wikipedia data emphasized the cosmopolitan nature of NYC, and this justifies our initial hypothesis for the need for more African restaurants there. Our exploration of the NYC dataset revealed to us that

Queens borough has the highest number of neighborhoods in NYC (80), closely followed by Brooklyn, Staten Island, Bronx and Manhattan. Our results also show that of the over 300 restaurants in NYC, we only have 18 African restaurants in 3 boroughs; Bronx (7), Manhattan (7) and Brooklyn (4) as Queens and Staten Island had none. This confirmed our earlier problem statement of the need for more African Restaurants in NYC. We also observe that though the Bronx and Manhattan has the highest number of African Restaurants, Central Harlem, Manhattan has 4 African Restaurants; the most for a neighborhood in New York City. Our analysis also revealed that though the African Restaurant with the maximum Likes and Tips is in Crown Heights, Brooklyn, the African Restaurant with the maximum Rating is in East Harlem Manhattan. We also observe that though the Bronx has the highest average rating for African Restaurants in NYC Boroughs, only 7 African restaurants (out of 18) had a Rating greater than 7.

**5. DISCUSSION**

We embarked on this project based on the need for more African restaurants in NYC given the massive population of those who crave and relish African foods in NYC. To resolve this problem, we did an elaborate study of NYC neighbourhoods to determine possible locations for starting an African Restaurant. Using various tools for data analysis, inferential statistic testing and machine learnings, our initial hypothesis is justified that there is a need for more African restaurants in NYC. We observed that Queens borough, which is the largest borough in NYC is second to Brooklyn only in terms of being the second most populous borough and second highest African American community. However, our results show that despite it being the most culturally diverse borough in NYC, it had no African restaurant. This was a huge surprise to us. So, from our results, we have determined that Queens borough will be the best location for setting up a new African restaurant given its population demographics/cultural diversity coupled with the fact that it currently has no African restaurant. So, we will be recommending that entrepreneurs and angel investors explore this huge investment prospect and set up an African restaurant in Queens, NYC.

## 6. CONCLUSION

Though this project contributes to the body of knowledge on this subject matter, it was bedevilled by some limitations (scope, dataset, etc) which we believe future research can build on. Some recommendations will be that investors and future researchers explore the legal requirements and cost of setting up an African restaurant in NYC, carry out feasibility studies on the taste preference of New Yorkers for African foods and consider setting up an African food store along with a restaurant. This research and entire course have been a great learning experience for me, and I am grateful to God, my family, Coursera, IBM and wonderful peers for helping me through it all. For a better view of the codes and output of this project, please refer to the notebook on my GitHub repository at [Coursera_Capstone/The_Battle_of_Neighborhoods_1.ipynb at main · Emmyeka1/Coursera_Capstone (github.com)](github.com). Thank you.

# REFERENCES

Demographics of New York City retrieved 11/25/2020 from [Demographics of New York City - Wikipedia](#)

New York City retrieved 11/25/2020 from [https://en.wikipedia.org/wiki/New_York_City](https://en.wikipedia.org/wiki/New_York_City)

New York Dataset retrieved 11/20/2020 from [https://cocl.us/new_york_dataset](https://cocl.us/new_york_dataset)

The Battle of Neighborhoods 1 retrieved 11/28/2020 from [Coursera_Capstone/The_Battle_of_Neighborhoods_1.ipynb at main · Emmyeka1/Coursera_Capstone (github.com)](#)