

# Data Science Capstone Project- Coursera

## The Battle of Neighborhoods



Identifying an optimal location for an African up market restaurant in London.

Richard Chinyama Njolomba

## Contents

Project Description .....	2
Introduction .....	2
The case of an African Restaurant in London .....	2
Problem Identification .....	3
Facts about the London .....	3
Data Description.....	4
Data Source 1 – Neighborhood Data .....	4
Data Source 2 – Sub-Region.....	4
Data Source 3 – Geographical Coordinates .....	4
Data Source 4 – Venue categories.....	4
Data Acquisition .....	5
• Obtained a list of Boroughs and Coordinates.....	5
• Dataset on Race Distribution in London .....	6
• Employment Rates in London.....	7
• The next step was to obtain the coordinates of the 12 districts in Southwark Borough.....	8
• Exploring the Neighborhood using Foursquare API.....	9
Conclusion and Recommendation .....	11

## Project Description

### Introduction

The Project is a partial fulfillment of the Coursera IBM Data Science certification course. The project requirements are to leverage the “Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.” My initial intention was to carry out a project on my city, Lusaka, Zambia. But I soon realized the data challenges I would face with a choice of Lusaka. Even the Foursquare API may have a limited applicability because the app may not be widely used in my city.

The approach taken in this project is as follows:

- Retrieve a list of neighborhoods in London,
- Find the longitude and latitudes of the neighborhoods,
- Use the coordinates identified above as input in the Foursquare API,
- Use the Foursquare API to venue categories in each neighborhood
- Identify any other data such as crime reports, demographic reports and social economic report in the city of London

An analysis of the data will be done, identifying suitable areas where an African restaurant may be established. ZLT will find the information in this report helpful in identifying best location for a restaurant. This information provided will of course be used in conjunction with other information such as financial and economic information.

### The case of an African Restaurant in London

The management of ZLT Limited, believe there is room for a high-end African restaurant in London. The target market is not necessarily just African immigrants, although this is their main target market, but they believe there is a good number of non-Africans who may find African cuisine exotic and an experience to have. A number of factors have convinced them that an African restaurant would be a viable business proposition.

- The population of London is changing rapidly,

- London has been a preferred destination for many foreigners, mostly seeking employment.
- With an increasingly foreign-born population residing in the capital, London's children are more likely than ever to speak English as a second language at home.
- Population projections are that London's population will continue to grow quickly well into the middle of the century.
- There is a redistribution of population in London – shifting from one area to another.

### Problem Identification

We have been commissioned by ZLT Limited, a private consortium of companies, to recommend the best locations to establish an African Restaurant in the city of London. The management of ZLT has highlighted the following as the key success factors for establishing a restaurant:

- The Location of a restaurant must be in a popular place,
- The identified location must not be close to other African restaurants,
- The location must be near or close to African communities,
- The location must be in the most secure areas,
- The Location must be family friendly,
- Easy access to the location by public and private transport.

### Facts about the London

This project is on the Greater London as opposed to City of London. To give us some understanding of the Greater London.

**“Greater London** is a ceremonial county of England that is located within the **London region**. This region forms the administrative boundaries of London and is organized into 33 local government districts—the 32 London boroughs and the City of London, which is located within the region but is separate from the county. The Greater London Authority, based in Southwark, is responsible for strategic local government across the region and consists of the Mayor of London and the London Assembly. The City of London Corporation is the principal local authority for the City of London, with a similar role to that of the 32 London borough councils.

Administratively, Greater London was first established as a *sui generis council area under the* Greater London Council between 1963 and 1986. The county of Greater London was created on 1 April 1965 through the London Government Act 1963. The area was re-established as a region in 1994. The Greater London Authority was formed in 2000.

The region covers 1,572 km<sup>2</sup> (607 sq mi) and had a population of 8,174,000 at the 2011 census.<sup>[6][7][8][9]</sup> The Greater London Built-up Area is used in some national statistics and is a measure of the continuous urban area and includes areas outside the administrative region.” [https://en.wikipedia.org/wiki/Greater\\_London](https://en.wikipedia.org/wiki/Greater_London).

## Data Description

In this section, I describe the data that will be used in the project. The discussed data sources are not conclusive as during the course of the analysis some more data may be required. The list below, therefore only serves to give a rough estimate of the sources of data needed for this project.

### Data Source 1 – Neighborhood Data

The Greater London, which is the subject of this study rather than just the City of London, has 33 boroughs. We first need to obtain a list of all the locations or neighborhoods in London with their respective boroughs and post codes. This information is available on the following web address: [https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London).

### Data Source 2 – Sub-Region

Data will be required on the rate of employment by ethnic group and also the population distribution in terms of race. This information would be required in order for us to know the areas have the greatest number of people of African descent.

### Data Source 3 – Geographical Coordinates

Geographical coordinates for each neighborhood will be obtained with the aid of GEOPY Library. Each postal code will be assigned a latitude and longitude coordinate.

### Data Source 4 – Venue categories

We will use the Foursquare API to retrieve venues, using the coordinates obtained in Data Source 3 above. We shall further obtain a list using Foursquare API for African Cuisine restaurants in London.

## Data Acquisition

The first step is to acquire the data that would be used in the project. The necessary dependencies were imported as shown below:

### importing dependencies needed for the analysis

```
import requests
import numpy as np # Library to handle data in a vectorized manner

import pandas as pd # Library for data analysis

import json # Library to handle JSON files

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

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

#!conda install -c conda-forge folium=0.5.0 --yes
import folium # map rendering library

import matplotlib.pyplot as plt

from bs4 import BeautifulSoup
from urllib.request import urlopen
import re

print('Libraries imported.')

Libraries imported.
```

It was then possible to scrap data from wikipedia using the beautifulsoup library.

Without going into details, I carried out the following:

- Obtained a list of Boroughs and Coordinates

I first obtained a list of boroughs in London and then scrapped the web for latitude and longitude of all boroughs There are 32 boroughs in London.

```
In [6]: df_boroughs.head()
```

```
Out[6]:
```

	<b>Borough</b>	<b>Latitude</b>	<b>Longitude</b>
<b>0</b>	Barking and Dagenham	51.5607	0.1557
<b>1</b>	Barnet	51.6252	-0.1517
<b>2</b>	Bexley	51.4549	0.1505
<b>3</b>	Brent	51.5588	-0.2817
<b>4</b>	Bromley	51.4039	0.0198

I then found datasets on the population distribution and employment rates in London Boroughs. The datasets were in Excel and needed some massaging. The Datasets were cleaned in Excel ( am an accountant so this came much more naturally).

- [Dataset on Race Distribution in London](#)

```
In [13]: df_race.head()
```

```
Out[13]:
```

	<b>Borough</b>	<b>Black</b>
<b>1</b>	Barking and Dagenham	41000
<b>2</b>	Barnet	14000
<b>3</b>	Bexley	27000
<b>4</b>	Brent	62000
<b>5</b>	Bromley	18000

- Employment Rates in London

```
In [17]: df_emplRate.head()
```

```
Out[17]:
```

	<b>Borough</b>	<b>Rate</b>
<b>1</b>	Barking and Dagenham	0.664587
<b>2</b>	Barnet	0.685637
<b>3</b>	Bexley	0.765823
<b>4</b>	Brent	0.635234
<b>5</b>	Bromley	0.746544

The two DataFrames were merged to come up with the table below. Based on the table below it was concluded that **Southwark Borough** should be the focus of our research. This was based on the fact that Southwark had the highest population of residents of African descent.

```
In [22]: df_data.head(10)
```

```
Out[22]:
```

	<b>Borough</b>	<b>Black</b>	<b>Rate</b>
<b>0</b>	Southwark	71000	0.706081
<b>1</b>	Croydon	70000	0.699552
<b>2</b>	Lambeth	67000	0.762029
<b>3</b>	Lewisham	63000	0.762238
<b>4</b>	Brent	62000	0.635234
<b>5</b>	Greenwich	59000	0.674453
<b>6</b>	Newham	56000	0.600637
<b>7</b>	Hackney	48000	0.616279
<b>8</b>	Enfield	47000	0.676269
<b>9</b>	Waltham Forest	45000	0.640327



- The next step was to obtain the coordinates of the 12 districts in Southwark Borough.

df\_coords

Out[27]:

	Area	Code	Latitude	Longitude
0	Bankside	TQ325795	51.5082	0.1001
1	Bermondsey	TQ335795	51.4986	0.0757
2	Camberwell	TQ325767	51.4736	0.0912
3	Denmark Hill	TQ327760	51.4678	0.0902
4	Dulwich	TQ345725	51.4457	0.0779
5	East Dulwich	TQ345745	51.4620	0.0840
6	Elephant and Castle	TQ319789	51.4950	0.1010
7	Newington	TQ325795	51.4988	0.0901
8	Nunhead	TQ355755	51.4622	0.0508
9	Peckham	TQ345765	51.4714	0.0625
10	Rotherhithe	TQ358796	51.4989	0.0426
11	Surrey Quays	TQ356789	51.4920	0.0445
12	Walworth	TQ325785	51.4898	0.0905

Out[34]:

	Borough	Neighborhood	Code	Latitude	Longitude
0	Southwark	Bankside	TQ325795	51.5082	0.1001
1	Southwark	Bermondsey	TQ335795	51.4986	0.0757
2	Southwark	Camberwell	TQ325767	51.4736	0.0912
3	Southwark	Denmark Hill	TQ327760	51.4678	0.0902
4	Southwark	Dulwich	TQ345725	51.4457	0.0779
5	Southwark	East Dulwich	TQ345745	51.4620	0.0840
6	Southwark	Elephant and Castle	TQ319789	51.4950	0.1010
7	Southwark	Newington	TQ325795	51.4988	0.0901
8	Southwark	Nunhead	TQ355755	51.4622	0.0508
9	Southwark	Peckham	TQ345765	51.4714	0.0625
10	Southwark	Rotherhithe	TQ358796	51.4989	0.0426
11	Southwark	Surrey Quays	TQ356789	51.4920	0.0445
12	Southwark	Walworth	TQ325785	51.4898	0.0905

The above table was used to create a map for Southwark with its 12 districts clearly highlighted.



- Exploring the Neighborhood using Foursquare API

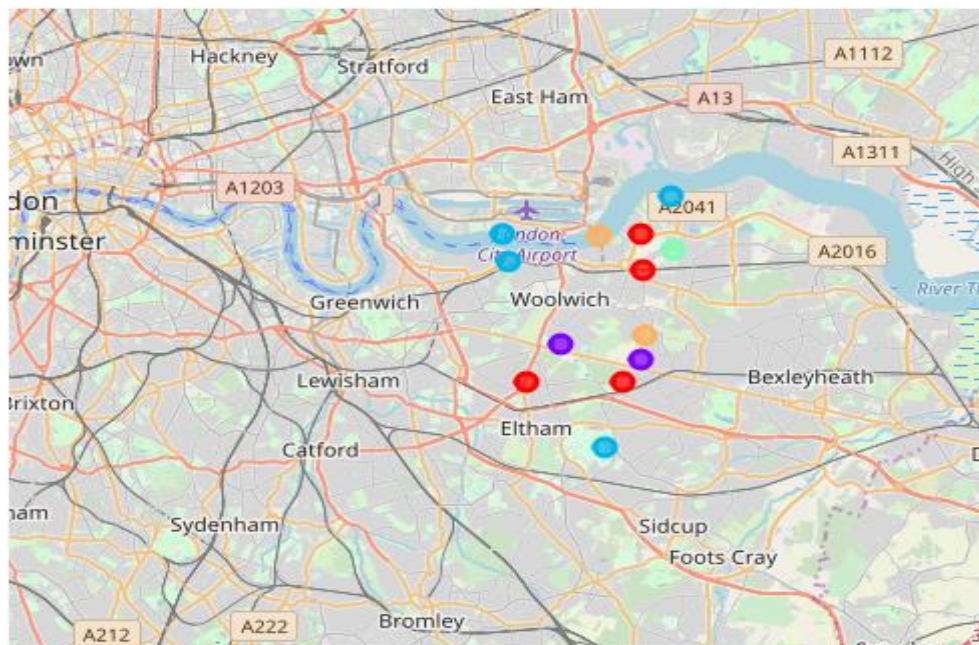
The next stage was to explore the neighborhood using the Foursquare API. I passed the geographical coordinates of each neighborhood to the Foursquare API, which gave us a list of venues in the neighborhood. The results of the exercise is a dataset presented below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bankside	51.5082	0.1001	Morrisons Thamesmead	51.507276	0.105392	Supermarket
1	Bankside	51.5082	0.1001	Wilko	51.505596	0.103845	Furniture / Home Store
2	Bankside	51.5082	0.1001	KFC	51.505567	0.104557	Fast Food Restaurant
3	Bankside	51.5082	0.1001	Next	51.506105	0.103512	Clothing Store
4	Bankside	51.5082	0.1001	McDonald's	51.506112	0.105161	Fast Food Restaurant
5	Bankside	51.5082	0.1001	Argos	51.505427	0.103978	Warehouse Store
6	Bankside	51.5082	0.1001	Pets at Home	51.505520	0.104001	Pet Store
7	Bankside	51.5082	0.1001	Superdrug	51.506883	0.106192	Pharmacy
8	Bankside	51.5082	0.1001	Aldi	51.506065	0.105938	Supermarket
9	Bermondsey	51.4986	0.0757	Royal Victoria Gardens	51.500548	0.072885	Park
10	Bermondsey	51.4986	0.0757	Zainab Express	51.501016	0.072079	Convenience Store
11	Bermondsey	51.4986	0.0757	Grimsby Grove	51.500669	0.070914	Harbor / Marina

The results were then prepared for analysis using the one hot encoding to pivot the venue categories from rows to column. I calculated the mean of the frequency of occurrence of each category in the table.

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
<b>Neighborhood</b>						
<b>Bankside</b>	9	9	9	9	9	9
<b>Bermondsey</b>	5	5	5	5	5	5
<b>Camberwell</b>	4	4	4	4	4	4
<b>Denmark Hill</b>	4	4	4	4	4	4
<b>Dulwich</b>	4	4	4	4	4	4
<b>East Dulwich</b>	5	5	5	5	5	5
<b>Elephant and Castle</b>	3	3	3	3	3	3
<b>Newington</b>	5	5	5	5	5	5
<b>Nunhead</b>	5	5	5	5	5	5
<b>Peckham</b>	4	4	4	4	4	4
<b>Rotherhithe</b>	5	5	5	5	5	5
<b>Surrey Quays</b>	6	6	6	6	6	6
<b>Walworth</b>	9	9	9	9	9	9

The final stage was to ran the k-means clustering algorithm on the dataframe to derive clusters of neighborhoods using 5 as the number of clusters. The results of the clustering is shown in the map below:



## Conclusion and Recommendation

After a thorough analysis we can conclude that the most appropriate location to setup up an African Restaurant would be in the first cluster (Cluster 0). Firstly, it is evident that there are no restaurants catering for the African community in the neighborhood. As a matter of fact, there are no African Restaurants in the vicinity. As indicated in the research, Southwark has the highest number of residents of African origin and that the employment rate is fairly high for minorities.

The research has partially satisfied the requirements of management as indicated in the preamble. In the preamble we indicated that management want the following conditions meant for establishing a restaurant:

- \* The identified location must not be close to other African restaurants,
- \* The location must be near or close to African communities,
- \* The location must be in the most secure areas,
- \* The Location must be family friendly,
- \* Easy access to the location by public and private transport

The first two conditions are the most crucial considerations and have been meant. The last are easy to establish, although for lack of time we did not investigate how secure the neighborhood are. Access to good transport is implied in the findings as most of the areas have bus stations or train as being some of the most frequented places.

However, there are other factors that may be required in order to assist management to come up with a decision. Our findings should therefore be used in conjunction with other information. Consideration such as cost of renting a premise for a restaurant should be considered.