CLUSTERING THE CITY OF LONDON FOR THE

OPENING OF AN

AFRICAN RESTAURANT

1. INTRODUCTION

1.1 DESCRIPTION OF THE PROBLEM

The City of London is a city and county that contains the historic centre and the primary central business district (CBD) of London. The City is now only a tiny part of the metropolis of London, though it remains a notable part of central London. However, the City of London is not a London borough, a status reserved for the other 32 districts (including London's only other city, the City of Westminster).

The City is a major business and financial centre. The City has a resident population of 9,401 (ONS estimate, mid-2016) but over 300,000 people commute to and work there. About three quarters of the jobs in the City of London are in the financial, professional, and associated business services sectors.

There are many restaurants across the city of London. To name a few: Roganic , CORE by Clare Smyth, The Five Fields and Portland Restaurant. Most of these restaurants offer continental, European or American dishes.

However many of the restaurants do not offer African dishes. Approximately 922,684 Africans live in London. There are decent number of Africans namely Nigerians, South Africans, Kenyans to name a few that reside in the greater London Area. These individuals are engaged in various occupations, business or schooling.

1.2 INTEREST

My project seeks to find the most populous areas in London that would offer prospective investors open restaurants to cater mostly for Africans who will want to have a taste of African dishes.

The project will explore the various neighborhoods in London and segment them and recommend which area is suitable for such business.

1.3 DESCRIPTION OF DATA

To help with this task, I will be getting a list of the various neighborhoods in London. There are various ways to collect the data. They could be extracted from a CSV file and then read into a pandas data frame or extracted from a website that offer such information. This technique is called web / internet scraping.

For the neighborhood data of London, I will be using the link

https://en.wikipedia.org/wiki/List_of_areas_of_London

This web page has a list of the neighborhood data of London.

I'll be getting the data from the website above for the black population and storing it on an FTP Server.

1.4 DESCRIPTION OF THE DATASET, BLACK POPULATION

ENTITY – BLACK POPULATION *HEADERS*

Black African population – the black African population in London.

LONDON BOROUGH – Consists of the cities of London

Black Caribbean population – The black Caribbean population in London.

Other black population - other black population

Total Black Population – the total black population

After the dataset has been read into a data frame, data cleaning will be done to eliminate blank fields or grouping of data if necessary.

To get the geographical coordinates of each neighbourhood in London described above, I will get the data from https://www.freemaptools.com/download-uk-postcode-lat-lng.htm

Different forms (sql, mysql, csv) of the geospatial data is hosted on the above site. I have decided to use the csv file in the link, transform it into an excel file, stored on an FTP site and read into a pandas data frame.

The excel file will be imported into my notebook and read into a pandas data frame.

The data frame for the black population and geospatial data, having the coordinates will be joined so as to enable me wrangle with the data.

1.5 DESCRIPTION OF THE GEOSPATIAL DATA SET

HEADERS

ID – Unique description of each post code in the dataset

POSTCODE – post code of various areas in London / United Kingdom

LATITUDE – The latitude of each postcode / area

LONGITUDE – the longitude of each postcode / area

2.0 DATA CLEANING

After the data has been imported and read into a pandas data frame, some data cleaning will be done. I noticed after the data has been read into the data frame, some rows had some missing values and in some instances missing values were represent with not a number (NAN) values. These rows had to be dropped especially after the two data sets had been merged. This was quite necessary because when the city has been divided and clusters created, the visualization cannot be done on NAN VALUES. I also had to drop some columns in the black population data set as they were not needed in my analysis.

2.1 FEATURE SELECTION

After cleaning the data, I was interested in getting the neighbourhoods of the black population in London. The table below illustrates this:

London Borough	Black African Population	Black Caribbean Population	Other Black Population	Total Black Population	Post Code
Lambeth	35,187	28,886	14,469	78,542	SW2, SW9, SE5
Southwark	47,413	17,974	12,124	77,511	SE1
Lewisham	32,025	30,854	12,063	74,942	SE6
Croydon	28,981	31,320	12,955	73,256	CR0
Newham	37,811	15,050	7,395	60,256	E6, E16, IG11
Brent	24,391	23,723	10,518	58,632	HA0
Hackney	27,976	19,168	9,714	56,858	E8
Enfield	28,222	17,334	8,131	53,687	N11, N14
Greenwich	35,164	8,051	5,440	48,655	SE3, SE12
Haringey	23,037	18,087	6,706	47,830	N11, N22

Waltham Forest	18,815	18,841	7,135	44,791	E11
Barking and Dagenham	28,685	5,227	3,228	37,140	IG11
Ealing	17,299	13,192	6,369	36,860	W4
Wandsworth	14,818	12,297	5,641	32,756	SW12
Barnet	19,392	4,468	3,571	27,431	EN5, NW7
Islington	12,622	7,943	5,729	26,294	EC1, N1
Redbridge	12,357	9,064	3,424	24,845	IG2
Hammersmith and Fulham	10,552	7,111	3,842	21,505	SW6
Merton	10,442	8,126	2,243	20,811	SW19
Hillingdon	11,275	4,615	4,192	20,082	UB8

I had to keep the Boroughs, black African population, black Caribbean population, other black population and the total population to have an idea of the African population. Since I won't be making use of the post code column, it was dropped.

Much cleaning was not carried out in the data set, geospatial London which had all the coordinates of the boroughs in London.

The table below describes:

Borough	**Latitude**	**Longitude**
Barking and Dagenham	51.5607	0.1557
Barnet	51.6252	0.1517
Bexley	51.4549	0.1505
Brent	51.5588	0.2817
Bromley	51.4039	0.0198
Camden	51.5290	0.1255
Croydon	51.3714	0.0977
Ealing	51.5130	0.3089
Enfield	51.6538	0.0799
Greenwich	51.4892	0.0648
Hackney	51.5450	0.0553
Hammersmith and Fulham	51.4927	0.2339
Haringey	51.6000	0.1119
Harrow	51.5898	0.3346
Havering	51.5812	0.1837
Hillingdon	51.5441	0.4760
Hounslow	51.4746	0.3680
Islington	51.5416	0.1022
Kensington and Chelsea	51.5020	0.1947
Kingston upon Thames	51.4085	0.3064
Lambeth	51.4607	0.1163
Lewisham	51.4452	0.0209
Merton	51.4014	0.1958

Newham	51.5077	0.0469
Redbridge	51.5590	0.0741
Richmond upon Thames	51.4479	0.3260
Southwark	51.5035	0.0804
Sutton	51.3618	0.1945
Tower Hamlets	51.5099	0.0059
Waltham Forest	51.5908	0.0134
Wandsworth	51.4567	0.1910
Westminster	51.4973	0.1372

As we can see above the table, it has all the coordinates (longitudes and latitudes of the different areas in London).

2.3 EXPLATORY ANALYSIS

The two data set were merged using an inner join on the borough column to create a new data frame named df_combined.

This was the data frame I used to further my analysis.

I first tried to cluster the various neighbourhoods / boroughs. To do this, I utilised the foursquare API to retrieve location data and various restaurant information.

The results of the neighbourhood I got was saved into a json file and read into a pandas data frame.

The results displayed below

BARNET, LONDON
LONDON, BARNET
BEXLEY, SIDCUP
BEXLEYHEATH, LONDON
LONDON, SIDCUP
BECKENHAM, LONDON
SEVENOAKS
KENLEY
THORNTON HEATH
NORTHOLT
HARROW, STANMORE
STANMORE
NORTHWOOD
BRENTFORD
MITCHAM

LONDON, BARKING
CHIGWELL
TEDDINGTON, HAMPTON
TEDDINGTON
WALLINGTON, CROYDON
CARSHALTON
SUTTON/MERTON
WALLINGTON

After the data has been displayed the different neighbourhoods were grouped:

Neighbourhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
BEXLEY, SIDCUP	51.4549	0.1505	Zizzi	51.455834	0.150347	Italian Restaurant
BEXLEY, SIDCUP	51.4549	0.1505	Wilko	51.456257	0.148608	Furniture / Home Store
BEXLEY, SIDCUP	51.4549	0.1505	TK Maxx	51.456481	0.14783	Clothing Store
BEXLEY, SIDCUP	51.4549	0.1505	Prince Albert	51.455171	0.152965	Pub
BEXLEY, SIDCUP	51.4549	0.1505	Bella Italia	51.45633	0.149537	Italian Restaurant

2.4 One – Hot Encoding

I used this technique to apply categorical variable to the neighbourhoods and the different venues in the heart of London. This technique assigns variables to the different venues.

The neighbourhoods are grouped and the mean frequency of each venue is displayed.

	Neighbourhood	Airport	Airport Lounge	Airport Service	American Restaurant	Asian Restaurant	Bakery	Bar	Books
0	BECKENHAM, LONDON	0.0000	0.0000	0.0000	0.0	0.05	0.000	0.05	
1	BEXLEY, SIDCUP	0.0000	0.0000	0.0000	0.0	0.00	0.050	0.00	
2	BEXLEYHEATH, LONDON	0.0000	0.0000	0.0000	0.0	0.00	0.050	0.00	
3	BRENTFORD	0.0000	0.0000	0.0000	0.0	0.00	0.000	0.00	
4	CARSHALTON	0.0000	0.0000	0.0000	0.0	0.00	0.000	0.00	
5	CHIGWELL	0.0000	0.0000	0.0000	0.0	0.00	0.050	0.00	
6	KENLEY	0.0000	0.0000	0.0000	0.0	0.00	0.125	0.00	
7	LONDON, BARKING	0.0625	0.0625	0.0625	0.0	0.00	0.000	0.00	
8	LONDON, SIDCUP	0.0000	0.0000	0.0000	0.0	0.00	0.050	0.00	
9	MITCHAM	0.0000	0.0000	0.0000	0.0	0.00	0.000	0.00	
10	NORTHOLT	0.0000	0.0000	0.0000	0.0	0.00	0.000	0.00	

The top 5 venues were then generated

```
----BECKENHAM, LONDON----
              venue  freq
0       Pizza Place  0.10
1       Coffee Shop  0.10
2    Clothing Store  0.10
3  Department Store  0.05
4    Sandwich Place  0.05


----BEXLEY, SIDCUP----
                  venue  freq
0           Coffee Shop  0.10
1           Supermarket  0.10
2                   Pub  0.10
3    Italian Restaurant  0.10
4  Portuguese Restaurant  0.05


----BEXLEYHEATH, LONDON----
                  venue  freq
0           Coffee Shop  0.10
```

```
1           Supermarket  0.10
2                   Pub  0.10
3    Italian Restaurant  0.10
4 Portuguese Restaurant  0.05


----BRENTFORD----
                  venue  freq
0  Fast Food Restaurant   1.0
1               Airport   0.0
2                 Hotel   0.0
3    Indian Restaurant   0.0
4             Irish Pub   0.0


----CARSHALTON----
               venue  freq
0      Historic Site   1.0
1              Hotel   0.0
2     Ice Cream Shop   0.0
3  Indian Restaurant   0.0
4          Irish Pub   0.0


----CHIGWELL----
             venue  freq
0   Clothing Store  0.20
1      Coffee Shop  0.10
2  Warehouse Store  0.05
3   Sandwich Place  0.05
4    Movie Theater  0.05


----KENLEY----
                  venue  freq
0           Coffee Shop  0.12
1    Italian Restaurant  0.12
2                Bakery  0.12
3           Supermarket  0.12
4  Fast Food Restaurant  0.12


----LONDON, BARKING----
                venue  freq
0               Hotel  0.31
1  Light Rail Station  0.12
2             Airport  0.06
3  Chinese Restaurant  0.06
4      Airport Lounge  0.06


----LONDON, SIDCUP----
                   venue  freq
0            Coffee Shop  0.10
1            Supermarket  0.10
2                    Pub  0.10
3     Italian Restaurant  0.10
4  Portuguese Restaurant  0.05
```

```
----MITCHAM----
               venue  freq
0                Pub   1.0
1            Airport   0.0
2              Plaza   0.0
3      Ice Cream Shop   0.0
4  Indian Restaurant   0.0


----NORTHOLT----
                   venue  freq
0           Home Service   0.5
1       Business Service   0.5
2  Portuguese Restaurant   0.0
3      Indian Restaurant   0.0
4              Irish Pub   0.0


----NORTHWOOD----
               venue  freq
0            Stables   1.0
1            Airport   0.0
2              Plaza   0.0
3     Ice Cream Shop   0.0
4  Indian Restaurant   0.0


----SEVENOAKS----
              venue  freq
0       Pizza Place  0.10
1       Coffee Shop  0.10
2    Clothing Store  0.10
3  Department Store  0.05
4    Sandwich Place  0.05


----SUTTON/MERTON----
               venue  freq
0      Historic Site   1.0
1              Hotel   0.0
2     Ice Cream Shop   0.0
3  Indian Restaurant   0.0
4          Irish Pub   0.0


----TEDDINGTON----
                 venue  freq
0  American Restaurant   0.2
1        Train Station   0.2
2  Sporting Goods Shop   0.2
3       Soccer Stadium   0.2
4          Coffee Shop   0.2


----TEDDINGTON, HAMPTON----
                 venue  freq
```

```
0  American Restaurant   0.2
1        Train Station   0.2
2  Sporting Goods Shop   0.2
3       Soccer Stadium   0.2
4          Coffee Shop   0.2


----THORNTON HEATH----
                    venue  freq
0             Coffee Shop  0.12
1     Italian Restaurant  0.12
2                  Bakery  0.12
3             Supermarket  0.12
4  Fast Food Restaurant  0.12


----WALLINGTON----
                venue  freq
0       Historic Site   1.0
1               Hotel   0.0
2      Ice Cream Shop   0.0
3  Indian Restaurant   0.0
4           Irish Pub   0.0


----WALLINGTON, CROYDON----
                venue  freq
0       Historic Site   1.0
1               Hotel   0.0
2      Ice Cream Shop   0.0
3  Indian Restaurant   0.0
4           Irish Pub   0.0
```

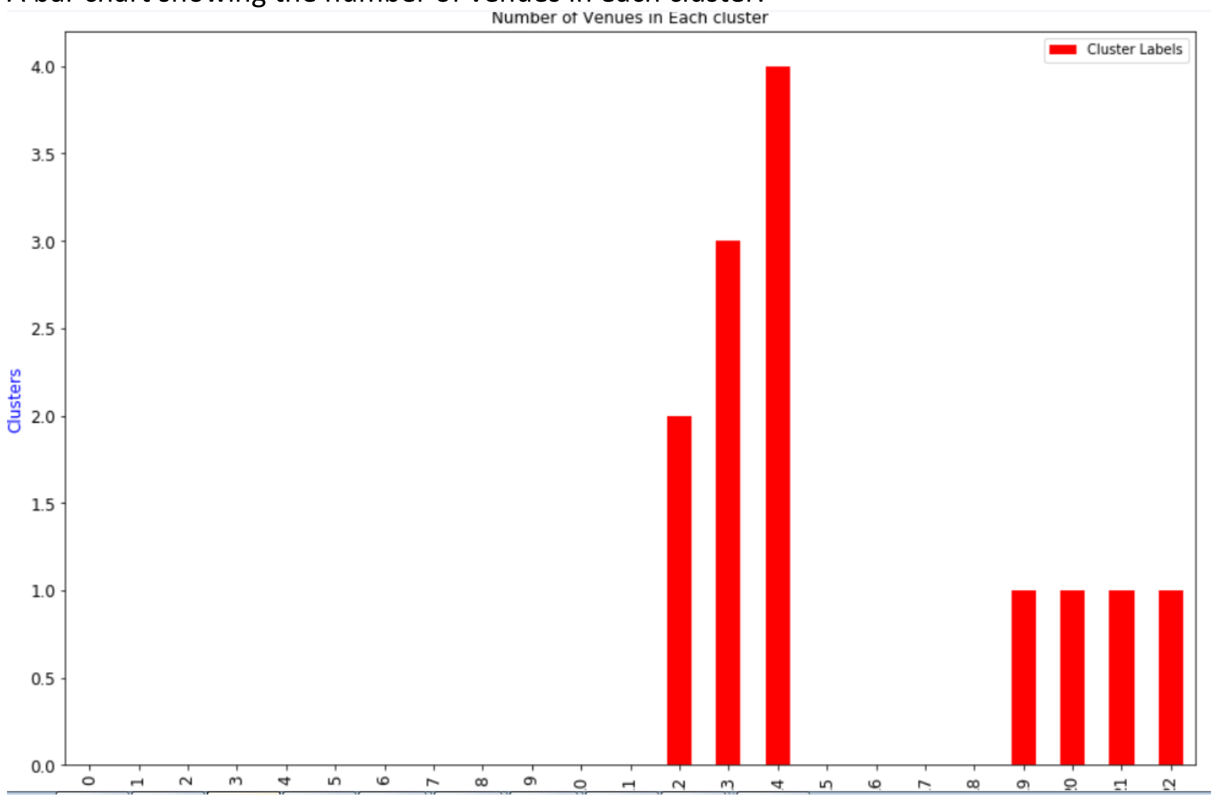
The top 5 neighbourhoods in the most common venue were displayed in the table below

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	BECKENHAM, LONDON	Pizza Place	Coffee Shop	Clothing Store	Portuguese Restaurant	Electronics Store
1	BEXLEY, SIDCUP	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store
2	BEXLEYHEATH, LONDON	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store
3	BRENTFORD	Fast Food Restaurant	Warehouse Store	Video Game Store	Grocery Store	Furniture / Home Store
4	CARSHALTON	Historic Site	Video Game Store	Grocery Store	Furniture / Home Store	Fast Food Restaurant

The black neighbourhood dataset was merged with the dataset above yielding the below ta ble:

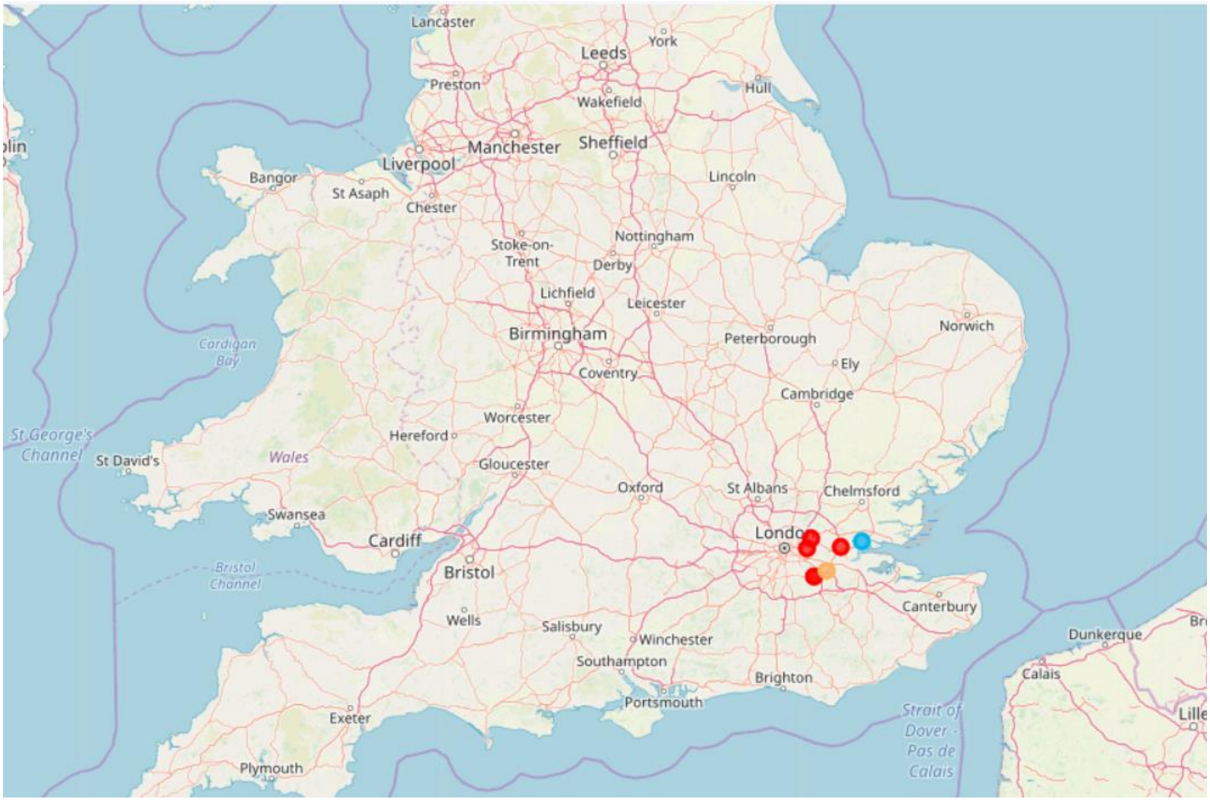
	Borough	Latitude	Longitude	Location	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bexley	51.4549	0.1505	Albany Park	BEXLEY, SIDCUP	0	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store
1	Bexley	51.4549	0.1505	Bexleyheath (also Bexley New Town)	BEXLEYHEATH, LONDON	0	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store
2	Bexley	51.4549	0.1505	Longlands	LONDON, SIDCUP	0	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store
3	Bromley	51.4039	0.0198	Beckenham	BECKENHAM, LONDON	0	Pizza Place	Coffee Shop	Clothing Store	Portuguese Restaurant	Electronics Store
4	Bromley	51.4039	0.0198	Cudham	SEVENOAKS	0	Pizza Place	Coffee Shop	Clothing Store	Portuguese Restaurant	Electronics Store
5	Croydon	51.3714	0.0977	Kenley	KENLEY	0	Pizza Place	Fast Food Restaurant	Supermarket	Bakery	Pub
6	Croydon	51.3714	0.0977	Thornton Heath	THORNTON HEATH	0	Pizza Place	Fast Food Restaurant	Supermarket	Bakery	Pub
7	Ealing	51.5130	0.3089	Northolt	NORTHOLT	0	Home Service	Business Service	Clothing Store	Grocery Store	Furniture / Home Store
8	Hillingdon	51.5441	0.4760	Northwood	NORTHWOOD	2	Stables	Warehouse Store	Chocolate Shop	Furniture / Home Store	Fast Food Restaurant
9	Hounslow	51.4746	0.3680	Brentford	BRENTFORD	3	Fast Food Restaurant	Warehouse Store	Video Game Store	Grocery Store	Furniture / Home Store
10	Merton	51.4014	0.1958	Mitcham	MITCHAM	4	Pub	Warehouse	Chocolate	Furniture /	Fast Food

A bar chart showing the number of venues in each cluster:



2.5 DATA VISUALIZED USING FOLIUM MAPS

I used the folium library to visualize the neighbourhoods.



2.6 K- MEANS CLUSTERING

K – Means clustering is grouping similar data points together and discover underlying pattern . The neighbourhoods were grouped into clusters of 5.

Analysis of Clusters

Cluster 1:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Bexley	0	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store	Hotel
1	Bexley	0	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store	Hotel
2	Bexley	0	Italian Restaurant	Pub	Supermarket	Coffee Shop	Clothing Store	Hotel
3	Bromley	0	Pizza Place	Coffee Shop	Clothing Store	Portuguese Restaurant	Electronics Store	Ice Cream Shop
4	Bromley	0	Pizza Place	Coffee Shop	Clothing Store	Portuguese Restaurant	Electronics Store	Ice Cream Shop
5	Croydon	0	Pizza Place	Fast Food Restaurant	Supermarket	Bakery	Pub	Italian Restaurant
6	Croydon	0	Pizza Place	Fast Food Restaurant	Supermarket	Bakery	Pub	Italian Restaurant
7	Ealing	0	Home Service	Business Service	Clothing Store	Grocery Store	Furniture / Home Store	Fast Food Restaurant

Cluster 2:

Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Sutton	1	Historic Site	Video Game Store	Grocery Store	Furniture / Home Store	Fast Food Restaurant
Sutton	1	Historic Site	Video Game Store	Grocery Store	Furniture / Home Store	Fast Food Restaurant
Sutton	1	Historic Site	Video Game Store	Grocery Store	Furniture / Home Store	Fast Food Restaurant
Sutton	1	Historic Site	Video Game Store	Grocery Store	Furniture / Home Store	Fast Food Restaurant

Cluster 3:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Hillingdon	2	Stables	Warehouse Store	Chocolate Shop	Furniture / Home Store	Fast Food Restaurant

Cluster 4:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	Hounslow	3	Fast Food Restaurant	Warehouse Store	Video Game Store	Grocery Store	Furniture / Home Store

Cluster 5:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Merton	4	Pub	Warehouse Store	Chocolate Shop	Furniture / Home Store	Fast Food Restaurant

2.7 RESULTS

We can see from the clusters above, there are no restaurants offering a composition of African food. Most of the restaurants above are Italian Restaurants, fast food chains or bakeries. Most of the black population in London leave in Croydon and Ealing to name a few areas.

In the clusters above for example the most popular restaurants are Pizza Restaurants and in the case of Ealing, there are not many restaurants.

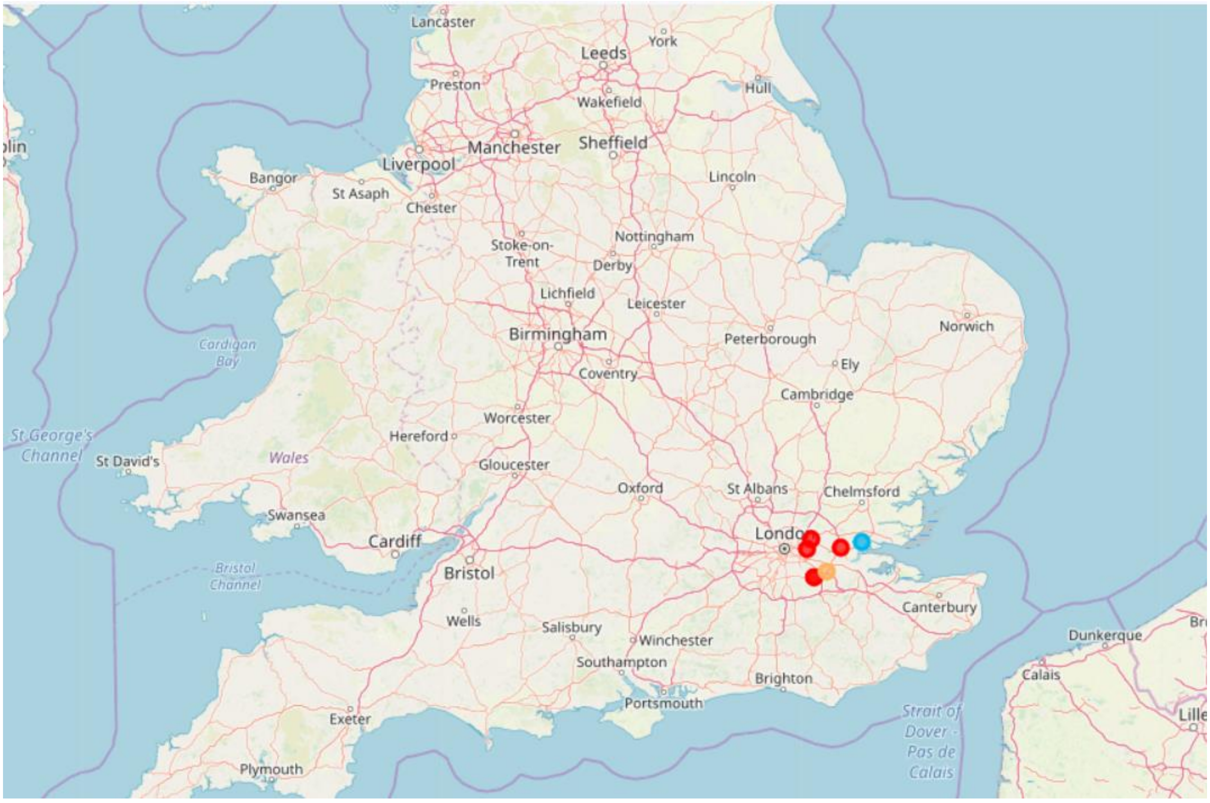
Such areas for example have an estimated number of 110,116 Africans living in those areas.

2.8 DISCUSSION

London is a big and vibrant metropolitan city and a hot spot for businesses. Different classification methods can be used to cluster the neighbourhoods.

I used the k-means algorithm in the study. I set the K-means value to 5 which yielded 18 venues.

I ended the study using visualization to display the data in a map using folium library.



2.9 CONCLUSION

The areas discussed above are a good hunting ground for prospective investors who may want to open an African restaurant for different nationals from countries such as South Africa, Kenya and Sierra Leone to name a few.

3.0 REFERENCES

www.maptools.com

https://en.wikipedia.org/wiki/List_of_London_boroughs

https://en.wikipedia.org/wiki/Ethnic_groups_in_London