Capstone Project - The Battle of Neighborhoods (Week 2)

1. Introduction

1.1. Background

London, the capital of Great Britain, and is one of the capital of finance, fashion, arts and entertainment. as one of the most populated city nowadays, it has been grown into a very diverse city. With the people from all over Europe and the world, besides people, you can also get food supplies from every part of the world.

Although, in this busy city already have a wide range of choose for food, with the help of data we can also find an ideal place for this business expansion.

1.2. Problem

One of our client, a successful chain restaurant original in US is looking forward for a business expansion in Europe, starting in London. And they would like to promote a special cultural cuisine of Africa for the busiest Londoner.

As the demographic of London is so big, our client needs to have a deeper insight from the data available in order to make their decision on where to start their first restaurant in London. With the help of the available data, we would able to have a research and provides our client with insights from different data analysis.

1.3. Target Audience

As the population of London is make up of people from with different cultural background, also, thousands of tourists coming from all over the world, the target audience is very broad of range.

2. Data Acquisition and Cleaning

2.1. Data Sources

The data of this project mainly come from two sources, the Wikipedia and Foursquare.

And the data source not just include London, but also the Greater London Area, and analysis all the data within this area, with the London Area Postcode, which the project will focus on the neighborhood within the London Postcode area.

The London Area consists of 32 Boroughs and the City of London, and the data link is https://en.wikipedia.org/wiki/List_of_areas_of_London

2.2. Data Cleaning

Couples of things will be done during the data cleaning.

In the data there are some borough names with [], these are references from wiki page, and we have to remove these.

	Location	Borough	Post-town	Postcode	Dial-code	OSGridRef
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

Here's the sample data after removed.

	Location	Borough	Post-town	Postcode	Dial-code	OSGridRef
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

Due the limitation for the number of calls using Foursquare API, and as London is a big place to handle, after a few research on London, we decided to limit the project to South East London.

2.3. Datasets

2.3.1. Dataset 1

2.3.1.1. Assumption 1

A value will be assigned to another column while the postcode is more than one, and spread to multi-rows.

	Location	Borough	Post-town	Dial-code	OSGridRef	Postcode
0	Abbey Wood	Bexley, Greenwich	LONDON	020	TQ465785	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W3
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W4
10	Angel	Islington	LONDON	020	TQ345665	EC1
10	Angel	Islington	LONDON	020	TQ345665	N1

2.3.1.2. Assumption 2

Create a new dataframe by Location, Borough, Postcode and Post-town for this project.

	Location	Borough	Postcode	Post-town
0	Abbey Wood	Bexley, Greenwich	SE2	LONDON
1	Acton	Ealing, Hammersmith and Fulham	W3	LONDON
2	Acton	Ealing, Hammersmith and Fulham	W4	LONDON
3	Angel	Islington	EC1	LONDON
4	Angel	Islington	N1	LONDON

2.3.1.3. Assumption 3

Only the Boroughs with London Post-town will be included in the search of Location. And the non-post town will be removed.

	Location	Borough	Postcode	Post-town
0	Abbey Wood	Bexley, Greenwich	SE2	LONDON
1	Acton	Ealing, Hammersmith and Fulham	W3	LONDON
2	Acton	Ealing, Hammersmith and Fulham	W4	LONDON
3	Angel	Islington	EC1	LONDON
4	Angel	Islington	N1	LONDON

2.3.1.4. Assumption 4

As a diverse city like London, the project will mainly focus on African markets and accessible facilities, and after studying the information about London, the South-Eastern areas will be considered and analysis. And the area's postcode is starting with SE.

	Location	Borough	Postcode
0	Abbey Wood	Bexley, Greenwich	SE2
1	Crofton Park	Lewisham	SE4
2	Crossness	Bexley	SE2
3	Crystal Palace	Bromley	SE19
4	Crystal Palace	Bromley	SE20
5	Crystal Palace	Bromley	SE26
6	Denmark Hill	Southwark	SE5
7	Deptford	Lewisham	SE8
8	Dulwich	Southwark	SE21
9	East Dulwich	Southwark	SE22

2.3.1.5. Assumption 5

In this section, will be more focus on the demography of London where there will be more multicultural groups. As in Assumption 4, the project will mainly focus on Africa market, and according to the races of London borough, the top 5 Africans group will be shown as below.

	Local authority	White	Mixed	Asian	Black	Other
22	Lewisham	53.5	7.4	9.3	27.2	2.6
27	Southwark	54.3	6.2	9.4	26.9	3.3
21	Lambeth	57.1	7.6	6.9	25.9	2.4
11	Hackney	54.7	6.4	10.5	23.1	5.3
7	Croydon	55.1	6.6	16.4	20.2	1.8

2.3.1.6. Assumption 6

Base on the top 5 areas with significantly high Black, Mixed and other races. It remains the following areas, Lewisham, Southwark, Lambeth, Hackney and Croydon.

	Location	Borough	Postcode
0	Crofton Park	Lewisham	SE4
1	Denmark Hill	Southwark	SE5
2	Deptford	Lewisham	SE8
3	Dulwich	Southwark	SE21
4	East Dulwich	Southwark	SE22

2.3.2. Dataset 2

While obtaining the location data, the Geocoder package is used with the *arcgis_geocoder* to obtain the latitude and longitude of the locations needed.

	Location	Borough	Postcode	Latitude	Longitude
0	Crofton Park	Lewisham	SE4	51.46268	-0.03558
1	Denmark Hill	Southwark	SE5	51.47480	-0.09313
2	Deptford	Lewisham	SE8	51.48114	-0.02467
3	Dulwich	Southwark	SE21	51.44100	-0.08897
4	East Dulwich	Southwark	SE22	51.45256	-0.07076

2.3.3. Dataset 3

With the help of the Foursquare API to obtain the South East London Area location for geographical location data. Which will be used to explore the neighborhoods of London.

The location within the neighborhoods of South East London, such as the restaurants and the proximity to amenities will be correlated. Also, the accessibility and the ease of supplies will also be considered.

3. Methodology

3.1. Data Exploration

3.1.1. Single Neighborhood

The initial exploration of the single neighborhood within London area is done to check whether the Foursquare API is working or not.

Partial data is extract here.

	name	categories	lat	lng
0	Street Feast Model Market	Street Food Gathering	51.460209	-0.012199
1	Maggie's Kitchen	Café	51.465380	-0.011213
2	Levante restaurant	Restaurant	51.462072	-0.009491
3	Levante Pide Restaurant	Turkish Restaurant	51.459848	-0.011476
4	Corte	Coffee Shop	51.459776	-0.011554
5	Manor House Gardens	Park	51.456686	0.004684
6	Dirty South	Pub	51.458846	-0.002666
7	Côte Brasserie	French Restaurant	51.467378	0.007176
8	Blackheath Farmers' Market	Farmers Market	51.465913	0.007945
9	Gennaro Delicatessan	Deli / Bodega	51.461765	-0.009726
10	The Sausage Man	Food Truck	51.462507	-0.010248

According to the summary extract, although there are plenty of restaurants in the Lewisham area, but it is not in the top 5.

	Count
Pub	13
Café	8
Gastropub	7
Park	5
Garden	4

3.1.2. Multiple Neighborhoods

Let’s explore multiple neighborhood in the South East London area next.

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bankside	100	100	100	100	100	100
Bellingham	68	68	68	68	68	68
Bermondsey	100	100	100	100	100	100
Blackheath	91	91	91	91	91	91
Brixton	100	100	100	100	100	100
Brockley	100	100	100	100	100	100
Camberwell	100	100	100	100	100	100
Catford	68	68	68	68	68	68
Chinbrook	50	50	50	50	50	50
Crofton Park	100	100	100	100	100	100
Denmark Hill	100	100	100	100	100	100
Deptford	100	100	100	100	100	100
Dulwich	100	100	100	100	100	100
East Dulwich	81	81	81	81	81	81
Elephant and Castle	300	300	300	300	300	300
Forest Hill	100	100	100	100	100	100
Gipsy Hill	200	200	200	200	200	200
Grove Park	50	50	50	50	50	50
Herne Hill	100	100	100	100	100	100
Hither Green	100	100	100	100	100	100
Honor Oak	100	100	100	100	100	100
Ladywell	200	200	200	200	200	200
Lambeth	100	100	100	100	100	100
Lee	50	50	50	50	50	50
Lewisham	100	100	100	100	100	100
New Cross	100	100	100	100	100	100
Newington	200	200	200	200	200	200
Nunhead	100	100	100	100	100	100
Oval	100	100	100	100	100	100
Peckham	100	100	100	100	100	100
Rotherhithe	100	100	100	100	100	100
Selhurst	52	52	52	52	52	52
South Norwood	52	52	52	52	52	52
Southend	68	68	68	68	68	68
St Johns	100	100	100	100	100	100
Surrey Quays	100	100	100	100	100	100
Tulse Hill	200	200	200	200	200	200
Upper Norwood	100	100	100	100	100	100
Walworth	100	100	100	100	100	100
West Norwood	100	100	100	100	100	100

Then summarize into unique categories.

	Count
Pub	435
Coffee Shop	292
Café	266
Park	201
Grocery Store	145

3.2. Clustering

In this section, the clustering will be done based on the processed data in previous section for the neighborhood in the South East London.

3.2.1. Libraries

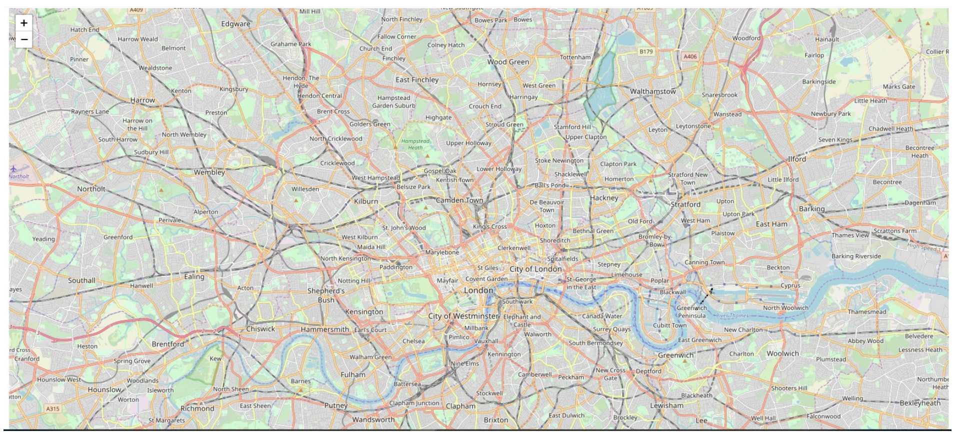
All the necessary libraries have already been import in the previous section.

3.2.2. Map Visualization

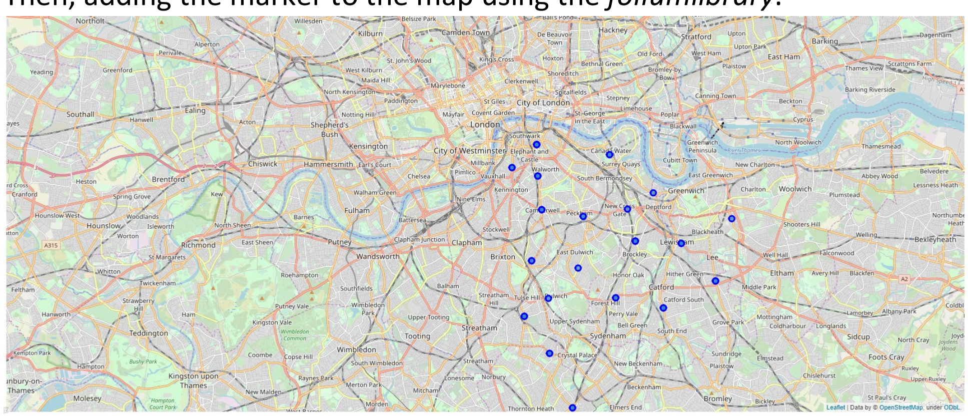
With the *geopy* library, the latitude and longitude values for London is obtained.

Using the *folium* library to obtain the coordinates of London.

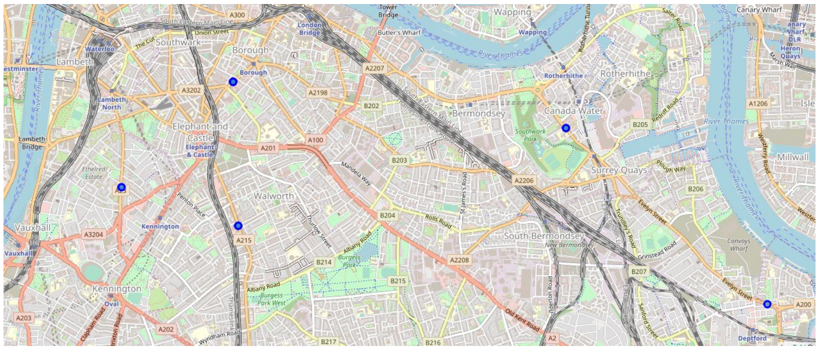
```
The geograpical coordinate of London are 51.5073219, -0.1276474.
```



Then, adding the marker to the map using the *foliumlibrary*.



Zoom in to see the superimposed areas.



3.2.3. Analyzing Each Neighborhood

This section is used to check and explore the locations in each neighborhood.

3.2.3.1. One Hot Encoding

The dataframe will be re-arrange, and putting the new Neighborhood in the first column.

	Neighbourhood	African Restaurant	American Restaurant	Antique Shop	Aquarium	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	BBQ Joint	Bakery	Bar	Beach	Beer Bar	Beer Garden	Beer Store	Bike Shop	Bistro	Bookstore	Brazilian Restaurant	Breakfast Spot	Brewery	Burger Joint	Bus Station	Bus Stop	Café
0	Crofton Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Crofton Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Crofton Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Crofton Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4	Crofton Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

From the below dataframe, Lewisham and its demography has no African Restaurants within the top spots.

	Neighbourhood	African Restaurant	American Restaurant	Antique Shop	Aquarium	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	BBQ Joint	Bakery	Bar	Beach	Beer Bar	Beer Garden	Beer Store	Bike Shop	Bistro	Bookstore	Brazilian Restaurant	Breakfast Spot	Brewery	Burger Joint	Bus Station	Bus Stop
1881	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1882	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1883	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1884	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1885	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1886	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1887	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1888	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1889	Lewisham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3.2.3.2. Regrouping and Category Statistics

Here's part of the dataframe.

	Neighbourhood	African Restaurant	American Restaurant	Antique Shop	Aquarium	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Australian Restaurant	BBQ Joint	Bakery	Bar	Beach	Beer Bar	Beer Garden	Beer Store	Bike Shop	Bistro	Bookstore	Brazilian Restaurant	Breakfast Spot	Brewery	Burger Joint	Bus Station
0	Bankside	0.00	0.000000	0.0	0.0	0.000000	0.01	0.03	0.0	0.02	0.0	0.0	0.0	0.020000	0.020000	0.01	0.00	0.0	0.01	0.01	0.01	0.01	0.0	0.000000	0.01	0.02	0.0
1	Bellingham	0.00	0.000000	0.0	0.0	0.000000	0.00	0.00	0.0	0.00	0.0	0.0	0.0	0.014706	0.014706	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.0	0.000000	0.00	0.00	0.0
2	Bermondsey	0.00	0.000000	0.0	0.0	0.000000	0.01	0.03	0.0	0.02	0.0	0.0	0.0	0.020000	0.020000	0.01	0.00	0.0	0.01	0.01	0.01	0.01	0.0	0.000000	0.01	0.02	0.0
3	Blackheath	0.00	0.010989	0.0	0.0	0.010989	0.00	0.00	0.0	0.00	0.0	0.0	0.0	0.032967	0.010989	0.00	0.00	0.0	0.00	0.00	0.00	0.00	0.0	0.010989	0.00	0.00	0.0
4	Brixton	0.01	0.000000	0.0	0.0	0.000000	0.01	0.00	0.0	0.00	0.0	0.0	0.0	0.020000	0.030000	0.00	0.03	0.0	0.01	0.01	0.00	0.00	0.0	0.000000	0.03	0.02	0.0

3.2.3.3. Creating New Dataframe

A new dataframe will be create toe put the common location into the *pandas dataframe*.

The new dataframe will put together the 10 most common locations.

	Neighbourhood	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
0	Bankside	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Art Museum	Cocktail Bar	Scenic Lookout	Seafood Restaurant	Garden
1	Bellingham	Grocery Store	Supermarket	Park	Café	Fast Food Restaurant	Pub	Italian Restaurant	Coffee Shop	Train Station	Chinese Restaurant
2	Bermondsey	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Art Museum	Cocktail Bar	Scenic Lookout	Seafood Restaurant	Garden
3	Blackheath	Pub	Grocery Store	Coffee Shop	Park	Clothing Store	Italian Restaurant	Supermarket	Garden	Bakery	Café
4	Brixton	Café	Coffee Shop	Park	Pub	Middle Eastern Restaurant	Indian Restaurant	Cocktail Bar	Italian Restaurant	Brewery	Pizza Place

3.2.3.4. Clustering of Neighborhoods.
Using the *k-means* to create clusters of neighborhood into 5 clusters.

	Location	Borough	Postcode	Latitude	Longitude	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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Crofton Park	Lewisham	SE4	51.46268	-0.03558	1	Pub	Coffee Shop	Café	Park	Gastropub	Bar	Indian Restaurant	Italian Restaurant	Cocktail Bar	Restaurant
1	Denmark Hill	Southwark	SE5	51.47480	-0.09313	4	Café	Coffee Shop	Park	Pub	Middle Eastern Restaurant	Indian Restaurant	Cocktail Bar	Italian Restaurant	Brewery	Pizza Place
2	Deptford	Lewisham	SE8	51.48114	-0.02467	1	Pub	Café	Coffee Shop	Bar	Park	Brewery	Sandwich Place	Italian Restaurant	History Museum	Historic Site
3	Dulwich	Southwark	SE21	51.44100	-0.08897	1	Pub	Café	Bakery	Grocery Store	Park	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Farmers Market	Pizza Place
4	East Dulwich	Southwark	SE22	51.45256	-0.07076	4	Pub	Café	Pizza Place	Gastropub	Park	Italian Restaurant	Mediterranean Restaurant	Cocktail Bar	Restaurant	Coffee Shop

3.2.4. Optimal Number of Clusters for K-mean
As there is a number of ways to have a possible for the evaluation to get the optimal number of clusters be used for the k-mean. The following will be used in this section.

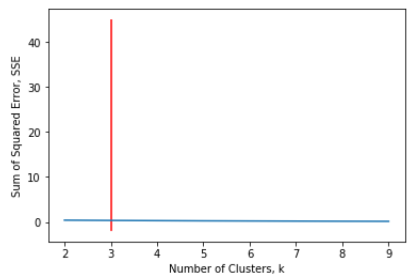
- Elbow Method.
- Sihouette Coefficient

3.2.4.1. Elbow Method

The *elbow method* is used to solve the problem of selecting k. although, it is not perfect but it is able to gives significant insight to choosing the optimal number of clusters by fitting the model with a range of values for K.

The approach is to run the k-means clustering for a range of value k and for each value of k. the Sum of Squared Errors will be calculated, calculate sum of squared errors. After this process, a plot of k and the corresponding sum of squared errors are then made.

According to the below, the number of iteration, the number of k is 3.



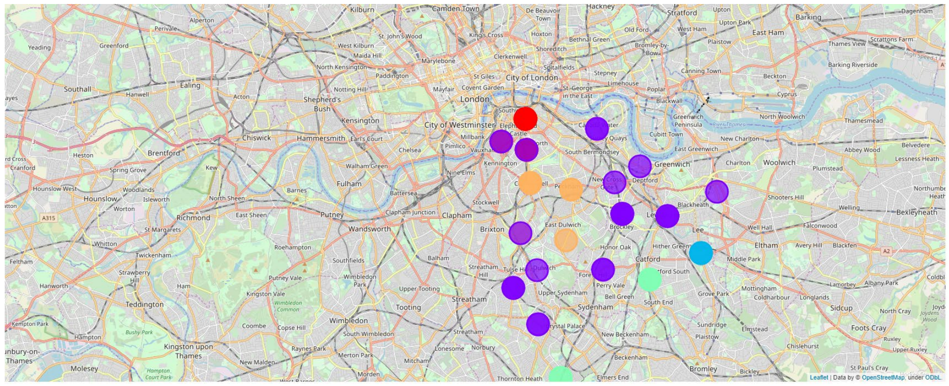
3.2.4.2. Silhouette Coefficient

Finding the optimal value of the number of clusters k, the number of clusters is iterated corresponding *Sihouette Coefficientis* calculated for each of the k-values used. The highest *Sihouette Coefficient* gives the best match of its own cluster.

From the result, the high *n_clusters* is the better silhouette coefficient. In this project, the cluster value of 5 will be used.

```
Index(['Location', 'Borough', 'Postcode', 'Latitude', 'Longitude',
       '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',
       '7th Most Common Venue', '8th Most Common Venue',
       '9th Most Common Venue', '10th Most Common Venue'],
      dtype='object')
```

3.2.5. Visualizing the Result Clusters



3.2.5.1. 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	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Southwark	0	Pub	Café	Coffee Shop	Hotel	Italian Restaurant	Theater	Park	Pizza Place	Street Food Gathering	Garden
6	Southwark	0	Pub	Café	Coffee Shop	Hotel	Italian Restaurant	Theater	Park	Pizza Place	Street Food Gathering	Garden
7	Southwark	0	Pub	Café	Coffee Shop	Hotel	Italian Restaurant	Theater	Park	Pizza Place	Street Food Gathering	Garden
8	Southwark	0	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Art Museum	Cocktail Bar	Scenic Lookout	Seafood Restaurant	Garden
18	Lambeth	0	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Art Museum	Cocktail Bar	Scenic Lookout	Seafood Restaurant	Garden
22	Southwark	0	Pub	Coffee Shop	Hotel	Italian Restaurant	Café	Theater	Cocktail Bar	Pizza Place	Brewery	Garden
23	Southwark	0	Pub	Coffee Shop	Hotel	Italian Restaurant	Café	Theater	Cocktail Bar	Pizza Place	Brewery	Garden
30	Southwark	0	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Art Museum	Cocktail Bar	Scenic Lookout	Seafood Restaurant	Garden

3.2.5.2. 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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Lewisham	1	Pub	Coffee Shop	Café	Park	Gastropub	Bar	Indian Restaurant	Italian Restaurant	Cocktail Bar	Restaurant
2	Lewisham	1	Pub	Café	Coffee Shop	Bar	Park	Brewery	Sandwich Place	Italian Restaurant	History Museum	Historic Site
3	Southwark	1	Pub	Café	Bakery	Grocery Store	Park	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Farmers Market	Pizza Place
9	Lewisham	1	Pub	Coffee Shop	Café	Grocery Store	Park	Supermarket	Gym / Fitness Center	Japanese Restaurant	Pizza Place	Pharmacy
10	Lambeth	1	Pub	Coffee Shop	Grocery Store	Café	Park	Bakery	Italian Restaurant	Pizza Place	Train Station	Gym / Fitness Center
11	Lambeth	1	Pub	Coffee Shop	Grocery Store	Café	Park	Bakery	Italian Restaurant	Pizza Place	Train Station	Gym / Fitness Center
14	Lewisham	1	Pub	Café	Gastropub	Park	Garden	Food Truck	Fish & Chips Shop	Coffee Shop	Restaurant	Turkish Restaurant
15	Lewisham	1	Pub	Coffee Shop	Café	Grocery Store	Park	Supermarket	Gym / Fitness Center	Japanese Restaurant	Pizza Place	Pharmacy
16	Lewisham	1	Pub	Coffee Shop	Café	Gastropub	Park	Bar	Restaurant	Italian Restaurant	Food Truck	Fish & Chips Shop
17	Lewisham	1	Pub	Coffee Shop	Café	Gastropub	Park	Bar	Restaurant	Italian Restaurant	Food Truck	Fish & Chips Shop
20	Lewisham	1	Pub	Café	Gastropub	Park	Garden	Food Truck	Fish & Chips Shop	Coffee Shop	Restaurant	Turkish Restaurant
21	Lewisham	1	Pub	Coffee Shop	Café	Italian Restaurant	Bar	Pizza Place	Gastropub	Park	Brewery	Indian Restaurant
25	Lambeth	1	Café	Pub	Hotel	Park	Coffee Shop	Theater	Street Food Gathering	Gay Bar	Bar	Cricket Ground
28	Southwark	1	Pub	Brewery	Park	Coffee Shop	Bar	Beer Bar	Café	Gym / Fitness Center	Food Truck	Vietnamese Restaurant
33	Lewisham	1	Pub	Coffee Shop	Café	Park	Gastropub	Bar	Indian Restaurant	Italian Restaurant	Cocktail Bar	Restaurant
34	Southwark	1	Pub	Brewery	Park	Coffee Shop	Bar	Beer Bar	Café	Gym / Fitness Center	Food Truck	Vietnamese Restaurant
35	Lambeth	1	Coffee Shop	Pub	Café	Grocery Store	Bakery	Park	Pizza Place	Brewery	Gym / Fitness Center	Market
36	Lambeth	1	Coffee Shop	Pub	Café	Grocery Store	Bakery	Park	Pizza Place	Brewery	Gym / Fitness Center	Market
37	Croydon	1	Pub	Coffee Shop	Park	Café	Italian Restaurant	Train Station	Bakery	Grocery Store	Gastropub	Breakfast Spot
38	Southwark	1	Pub	Café	Coffee Shop	Italian Restaurant	Park	Pizza Place	Brewery	Art Gallery	Theater	Hotel
39	Lewisham	1	Pub	Grocery Store	Coffee Shop	Park	Clothing Store	Italian Restaurant	Supermarket	Garden	Bakery	Café
40	Lambeth	1	Pub	Coffee Shop	Grocery Store	Café	Bakery	Park	Gym / Fitness Center	Pharmacy	Pizza Place	Brewery
42	Lewisham	1	Pub	Coffee Shop	Café	Park	Gastropub	Bar	Indian Restaurant	Italian Restaurant	Cocktail Bar	Restaurant

3.2.5.3. 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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
12	Lewisham	2	Grocery Store	Pub	Park	Café	Supermarket	Italian Restaurant	Soccer Field	Coffee Shop	Train Station	Indian Restaurant
19	Lewisham	2	Grocery Store	Pub	Park	Café	Supermarket	Italian Restaurant	Soccer Field	Coffee Shop	Train Station	Indian Restaurant
45	Lewisham	2	Grocery Store	Pub	Park	Café	Supermarket	Italian Restaurant	Soccer Field	Coffee Shop	Train Station	Indian Restaurant

3.2.5.4. 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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
26	Lewisham	3	Grocery Store	Supermarket	Park	Café	Fast Food Restaurant	Pub	Italian Restaurant	Coffee Shop	Train Station	Chinese Restaurant
29	Croydon	3	Pub	Grocery Store	Platform	Café	Park	Supermarket	Coffee Shop	Clothing Store	Caribbean Restaurant	Chinese Restaurant
31	Croydon	3	Pub	Grocery Store	Platform	Café	Park	Supermarket	Coffee Shop	Clothing Store	Caribbean Restaurant	Chinese Restaurant
32	Lewisham	3	Grocery Store	Supermarket	Park	Café	Fast Food Restaurant	Pub	Italian Restaurant	Coffee Shop	Train Station	Chinese Restaurant
44	Lewisham	3	Grocery Store	Supermarket	Park	Café	Fast Food Restaurant	Pub	Italian Restaurant	Coffee Shop	Train Station	Chinese Restaurant

3.2.5.5. 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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Southwark	4	Café	Coffee Shop	Park	Pub	Middle Eastern Restaurant	Indian Restaurant	Cocktail Bar	Italian Restaurant	Brewery	Pizza Place
4	Southwark	4	Pub	Café	Pizza Place	Gastropub	Park	Italian Restaurant	Mediterranean Restaurant	Cocktail Bar	Restaurant	Coffee Shop
13	Lambeth	4	Coffee Shop	Pub	Café	Market	Pizza Place	Brewery	Tapas Restaurant	Bakery	Park	Restaurant
24	Southwark	4	Pub	Pizza Place	Café	Park	Coffee Shop	Bar	Gastropub	Burger Joint	Cocktail Bar	Italian Restaurant
27	Southwark	4	Pub	Pizza Place	Café	Park	Coffee Shop	Bar	Gastropub	Burger Joint	Cocktail Bar	Italian Restaurant
41	Lambeth	4	Café	Coffee Shop	Park	Pub	Middle Eastern Restaurant	Indian Restaurant	Cocktail Bar	Italian Restaurant	Brewery	Pizza Place
43	Southwark	4	Café	Coffee Shop	Park	Pub	Middle Eastern Restaurant	Indian Restaurant	Cocktail Bar	Italian Restaurant	Brewery	Pizza Place

4. Discussions

Here are some highlights about the above 5 clusters.

- Pubs, Café and Coffee Shops are popular in the South East London area.
- For the restaurants, the Italian Restaurants are very popular in the South East London, especially in Southwark and Lambeth areas.
- In the Lewisham area which is most condensed area of Africans in the South East London, and from the top 10 locations, which restaurants can hardly be seen in the top 5 locations.
- Although there are some variations in the clusters, but the pubs really got a dominate category in the cluster result.

5. Conclusion

The cluster 2 and 3 are the most viable cluster for creating a brand of African Restaurant. This two clusters do not have top restaurants that could rival their standards. And the proximity of resources needed is paramount as Lewisham and Lambeth are not far out from Southwark.

If more data are analysis in this project such as crime data, or the restaurants customers' feedback can be including in the analysis will be more helpful in order to provide more insight while select a better location.