Applied Data Science Capstone Project

Which boroughs should Sub-Saharan Africans aim to lodge in during their holiday visits to London (UK)?

Olusoji Sogbetun

July 16,2020

Table of Contents

- Introduction
- Business Problem
- Data
- Methodology
- Results
- Discussion and Recommendation
- Conclusion

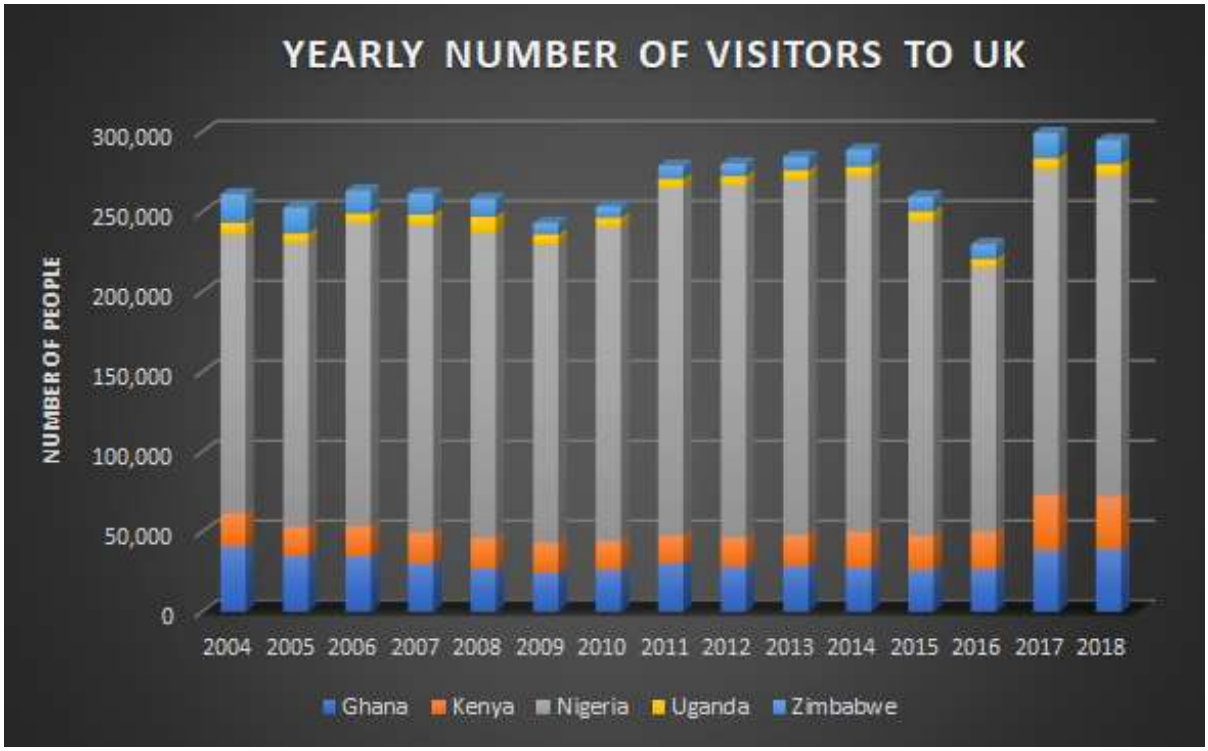
Introduction

London (UK) attracts a great deal of tourists every year and is considered home to a large African diaspora. It has become a popular travel destination for a lot of English speaking sub-Saharan Africans with many coming over to either visit friends and relatives, come for holidays or strictly for business.

According to UK Government records, over a quarter of a million (> 250,000) visitors in total ,i.e. on average, have come annually to the UK from the following sub-Saharan countries i.e. Nigeria, Kenya, Ghana, Uganda and Zimbabwe from 2004 to 2018.

The links between the UK and these African nations are deep, far reaching and rooted both in their shared past and partnership for the future.

The chart below illustrates the yearly numbers of visitors to UK from these five sub-Saharan countries according to reliable UK Government data.



A further dig into the records also reveal a large proportion of the visitors choose to visit London during their visit to the UK.

Business Problem

A Travel Agency in the business of arranging holiday packages in Africa has struggled in the past to satisfy some of its customers on the choice of location for the hotels selected for them during their visit to London UK.

In particular, the dissatisfaction has mostly been prominent amongst clients from five (5) sub-Saharan countries with ties to UK i.e. Nigeria, Ghana, Kenya, Uganda and Zimbabwe.

Feedback from a recent survey carried out by the Travel Agency reveals the top three (3) preferences which tourist from these nations look out for when embarking on a visit to London. These preferences, in addition to lodging in a nice hotel, include their accessibility to;

- Major Shopping Centres
- Popular Attractions / Sites
- Good Restaurants (preferably with menu including spicy options)

In providing a solution to this challenge , it has been proposed to the Travel Agency that machine learning analysis be carried out to help determine the cluster of London Boroughs which fit their customers' preferences.

The underlying principle behind the analysis is that an unsupervised machine learning technique is applied to reliable data in helping to create clusters of boroughs situated in London UK which have similar venue categories within their vicinity.

The results obtained from the analyses provides the Travel Agency with relevant knowledge and information needed in arranging the packaged holidays for their clients.

Data

In carrying out the machine learning clustering analysis required to determine the suitable London Boroughs which fit perfectly with the preferences of the identified sub-Saharan African tourist, relevant data was obtained from the credible sources as follows;

- List of Boroughs located in London UK
    - Credible data was sourced from the UK government website (https://data.london.gov.uk/dataset/london-borough-profiles)

- Geo-coordinates of the Boroughs in London UK
    - Geo-coordinates of the London Boroughs was obtained with the help of the GeoPy library. It was observed during the data wrangling process i.e. before running the extraction of the geo co-ordinates, some of the Boroughs had to be suffixed with "London" to enable the programme retrieve the right and accurate geo-coordinate from the actual Borough located in London UK.

- Top venues in each London Borough
    - Detailed data for the top venues situated in the vicinity of each London Borough was obtained using the Foursquare API.

- Some other informative data mentioned in the report include
    - An indication of UK visitor numbers from sub-Saharan Africa was obtained from a credible UK government website https://www.gov.uk/government/publications/immigration-statistics-year-ending-june-2019

Methodology

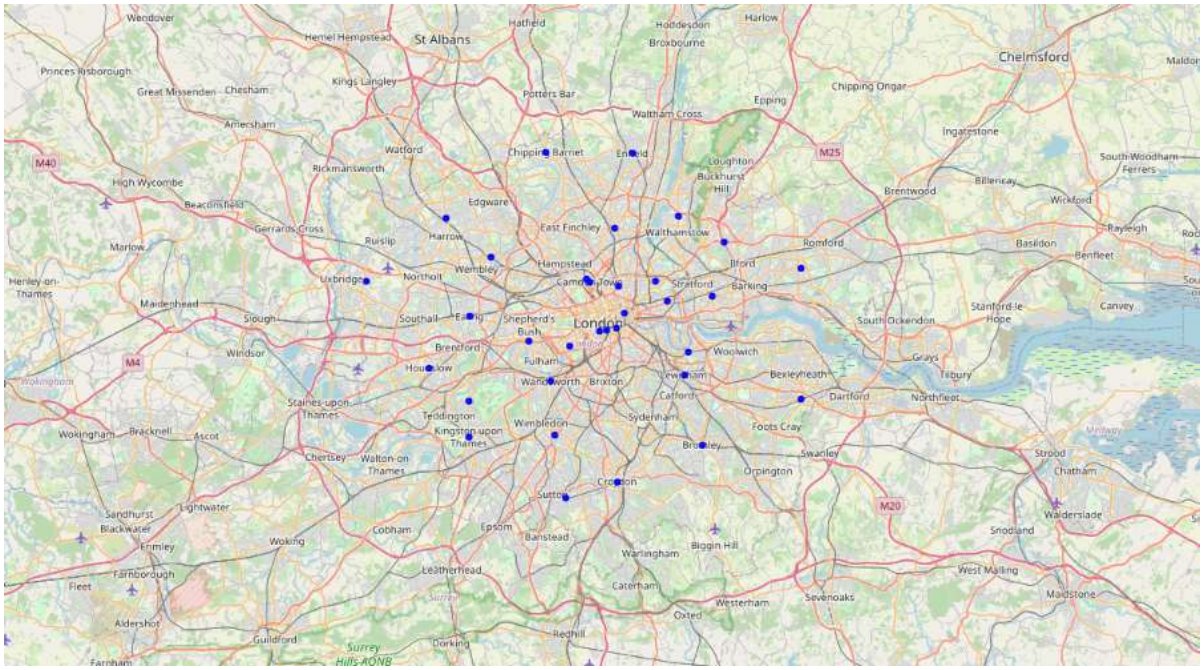
This section discusses the approach adopted in execution of the project using appropriate Python scripts with support from Python Libraries like Pandas, GeoPy, Matplotlib, Scikit-learn and Folium.

The first step in carrying out the project involved the importation of these various Python support libraries before the data is then read, explored and processed for machine learning analyses.

The data ,which has been obtained from reliable sources, will be prepared for machine learning analyses principally to segment the London Boroughs into clusters based on the unique venues associated with them.

The data preparation process involves the extraction and addition of each borough's geo-coordinates (Latitude and Longitude)using the GeoPy library into the dataframe.

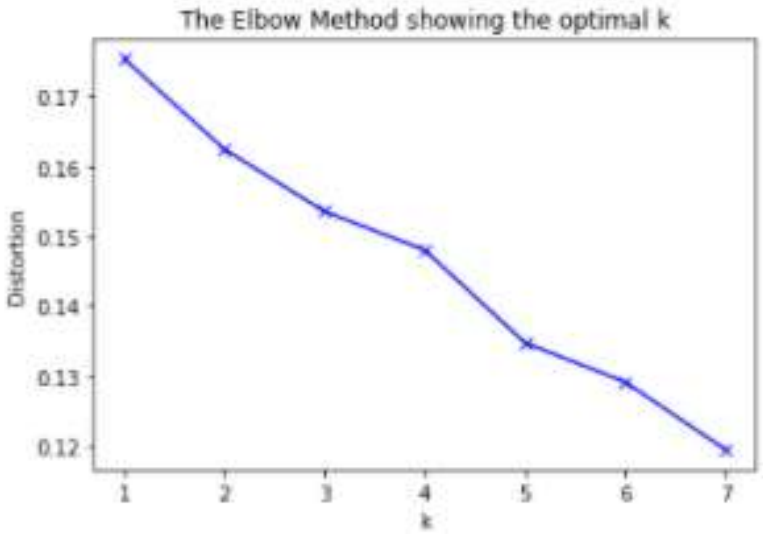
For visualisation purposes, a map of London was created and each identified boroughs superimposed on top of the map. This is presented below



As the data was being further explored, I proceeded to obtain details of the unique venues located within a 1km radius of each borough using the Foursquare API and incorporated the results into the dataframe.

The dataframe, now containing details of the unique venue categories associated with each borough, was further processed to enable the application of the unsupervised Machine Learning technique for segmentation of the boroughs into clusters. The machine learning technique used during the execution of the project is K-means clustering algorithm.

To obtain and identify the optimal number of clusters to apply during the machine learning process, I made use of the inertia (elbow rule) method and selected four (4) optimal number of clusters to be used.



The data was then analysed using the K-means clustering algorithm and the optimal number of clusters obtained above applied.

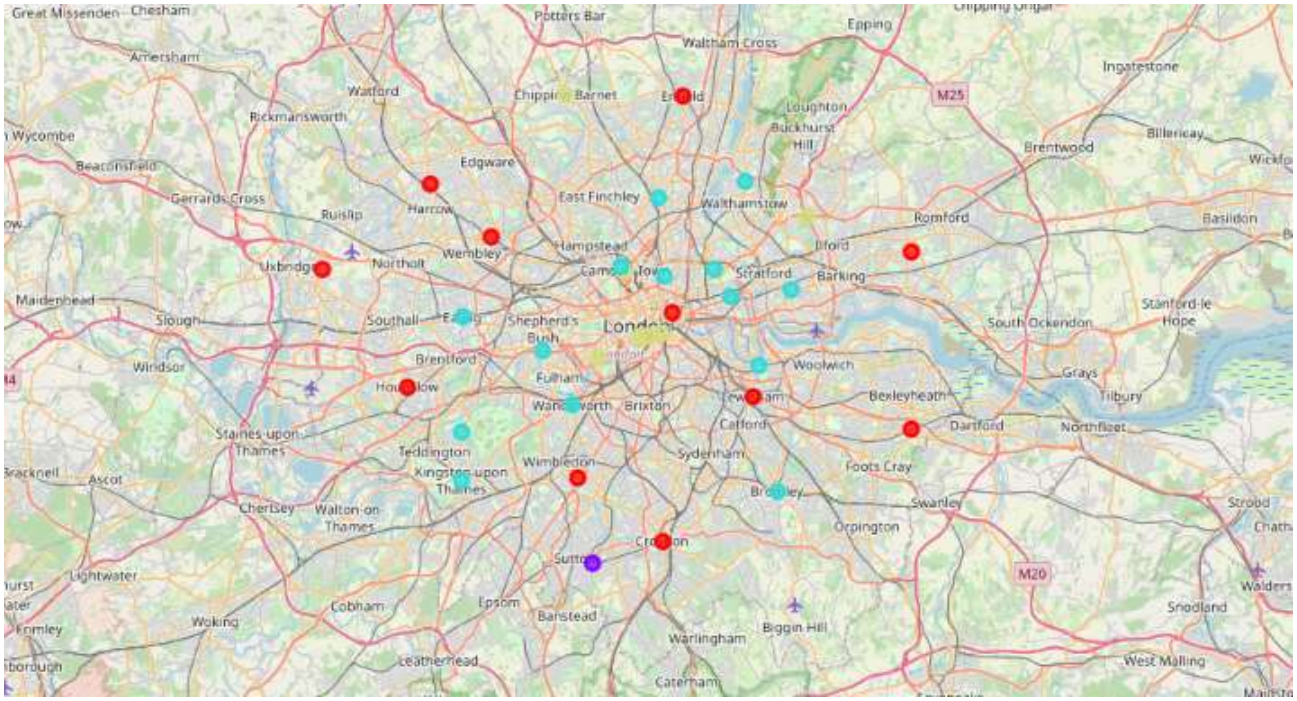
The results generated cluster labels which was then assigned to each Borough features as part of the Pandas dataframe.

```
In [61]: # add clustering labels
         Borough_merged['Cluster Labels']= kmeans.labels_
         Borough_merged.head(10)
```

Out[61]:

	Borough	Inner/Outer London	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	City of London	Inner London	51.515618	-0.091998	0	Coffee Shop	Gym / Fitness Center	Hotel	Cocktail Bar	Garden	Scenic Lookout	French Restaurant	Steakhouse	Boxing Gym	Modern European Restaurant
1	Barking and Dagenham	Outer London	51.554117	0.150504	0	Bus Stop	Gas Station	Convenience Store	Grocery Store	Liquor Store	Martial Arts Dojo	Gym	Gym / Fitness Center	Park	Supermarket
2	Barnet	Outer London	51.653090	-0.200226	3	Pub	Coffee Shop	Grocery Store	Bookstore	Fast Food Restaurant	Pizza Place	Pharmacy	Soccer Stadium	Park	Gym / Fitness Center
3	Bexley London	Outer London	51.441679	0.150488	0	Pub	Toy / Game Store	Fast Food Restaurant	Chinese Restaurant	Museum	Breakfast Spot	Greek Restaurant	Tennis Court	Italian Restaurant	Steakhouse
4	Brent London	Outer London	51.563826	-0.275760	0	Coffee Shop	Clothing Store	Bar	Hotel	Supermarket	Sandwich Place	Indian Restaurant	Sporting Goods Shop	Grocery Store	Warehouse Store
5	Bromley	Outer London	51.402805	0.014814	2	Clothing Store	Pub	Coffee Shop	Café	Burger Joint	Indian Restaurant	Park	Pizza Place	Supermarket	Department Store
6	Camden London	Inner London	51.542305	-0.139560	3	Pub	Coffee Shop	Café	Ice Cream Shop	Music Venue	Market	Burger Joint	Bar	Vegetarian / Vegan Restaurant	Grocery Store
7	Croydon	Outer London	51.371305	-0.101957	0	Coffee Shop	Pub	Clothing Store	Hotel	Platform	Italian Restaurant	Bookstore	Park	Mediterranean Restaurant	Sandwich Place
8	Ealing	Outer London	51.512655	-0.305195	2	Coffee Shop	Pub	Park	Italian Restaurant	Hotel	Bakery	Burger Joint	Pizza Place	Thai Restaurant	Café
9	Enfield	Outer London	51.652085	-0.081018	0	Pub	Clothing Store	Coffee Shop	Indian Restaurant	Department Store	Pharmacy	Park	Fish & Chips Shop	Supermarket	Pizza Place

For visualisation purposes, a map of London was created with the resulting clusters for the London Boroughs superimposed on top of the London map. This is presented below.



Results

A summary of the results obtained after undergoing the unsupervised machine learning K-means clustering is as shown below;

a) Cluster "0"

Cluster "0" Results

```
In [56]: Borough_merged.loc[Borough_merged['Cluster Labels'] == 0, Borough_merged.columns[[0] + [1] + list(range(5, Borough_merged.shape[1]))]]
```

Out[56]:

	Borough	Inner/Outer London	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	City of London	Inner London	Coffee Shop	Gym / Fitness Center	Hotel	Cocktail Bar	Garden	Scenic Lookout	French Restaurant	Steakhouse	Boxing Gym	Modern European Restaurant
1	Barking and Dagenham	Outer London	Bus Stop	Gas Station	Convenience Store	Grocery Store	Liquor Store	Martial Arts Dojo	Gym	Gym / Fitness Center	Park	Supermarket
3	Bexley London	Outer London	Pub	Toy / Game Store	Fast Food Restaurant	Chinese Restaurant	Museum	Breakfast Spot	Greek Restaurant	Tennis Court	Italian Restaurant	Steakhouse
4	Brent London	Outer London	Coffee Shop	Clothing Store	Bar	Hotel	Supermarket	Sandwich Place	Indian Restaurant	Sporting Goods Shop	Grocery Store	Warehouse Store
7	Croydon	Outer London	Coffee Shop	Pub	Clothing Store	Hotel	Platform	Italian Restaurant	Bookstore	Park	Mediterranean Restaurant	Sandwich Place
9	Enfield	Outer London	Pub	Clothing Store	Coffee Shop	Indian Restaurant	Department Store	Pharmacy	Park	Fish & Chips Shop	Supermarket	Pizza Place
14	Harrow	Outer London	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Platform	Grocery Store	Afghan Restaurant	Sandwich Place	Bus Stop	Park	Supermarket
16	Hillingdon	Outer London	Fast Food Restaurant	Chinese Restaurant	Plaza	Pharmacy	Coffee Shop	Construction & Landscaping	Park	Grocery Store	Pub	Food Stand
17	Hounslow	Outer London	Indian Restaurant	Clothing Store	Coffee Shop	Hotel	Fast Food Restaurant	Grocery Store	Pharmacy	Chinese Restaurant	Supermarket	Pub
22	Lewisham	Inner London	Pub	Clothing Store	Coffee Shop	Restaurant	Gym	Supermarket	Food Truck	Café	Bus Stop	Street Food Gathering
23	Merton	Outer London	Coffee Shop	Sushi Restaurant	Clothing Store	Park	Lebanese Restaurant	Thai Restaurant	Grocery Store	Bar	Supermarket	Café

b) Cluster "1"

Cluster "1" Results

```
In [57]: Borough_merged.loc[Borough_merged['Cluster Labels'] == 1, Borough_merged.columns[[0] + [1] + list(range(5, Borough_merged.shape[1]))]]
```

Out[57]:

	Borough	Inner/Outer London	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
28	Sutton London	Outer London	Grocery Store	Italian Restaurant	Park	Pub	Supermarket	Flea Market	Flower Shop	Food & Drink Shop	Food Court	Food Truck

c) Cluster "2"

Cluster "2" Results

```
In [58]: Borough_merged.loc[Borough_merged['Cluster Labels'] == 2, Borough_merged.columns[[0] + [1] + list(range(5, Borough_merged.shape[1]))]]
```

Out[58]:

	Borough	Inner/Outer London	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	Bromley	Outer London	Clothing Store	Pub	Coffee Shop	Café	Burger Joint	Indian Restaurant	Park	Pizza Place	Supermarket	Department Store
8	Ealing	Outer London	Coffee Shop	Pub	Park	Italian Restaurant	Hotel	Bakery	Burger Joint	Pizza Place	Thai Restaurant	Café
10	Greenwich	Outer London	Pub	Grocery Store	Garden	Café	Pizza Place	Park	Gym / Fitness Center	Trail	Bar	Bakery
11	Hackney	Inner London	Pub	Café	Coffee Shop	Bakery	Brewery	Cocktail Bar	Park	Vietnamese Restaurant	Gastropub	Modern European Restaurant
12	Hammersmith and Fulham	Inner London	Pub	Coffee Shop	Café	Italian Restaurant	Gym / Fitness Center	Thai Restaurant	Sandwich Place	Cocktail Bar	Bakery	Park
13	Haringey	Inner London	Turkish Restaurant	Café	Pub	Mediterranean Restaurant	Burger Joint	Furniture / Home Store	Bookstore	Grocery Store	Supermarket	Gift Shop
15	Havering London	Outer London	Pub	Coffee Shop	Café	Ice Cream Shop	Vegetarian / Vegan Restaurant	Grocery Store	Market	Bar	Yoga Studio	Flea Market
18	Islington	Inner London	Pub	Bakery	Coffee Shop	Gastropub	Park	Mediterranean Restaurant	Boutique	Burger Joint	French Restaurant	Cocktail Bar
20	Kingston upon Thames	Outer London	Pub	Coffee Shop	Café	Clothing Store	Thai Restaurant	Italian Restaurant	Burger Joint	Park	Department Store	Bakery
24	Newham	Inner London	Grocery Store	Café	Pub	Bus Stop	Fish & Chips Shop	Gym / Fitness Center	Market	Toy / Game Store	Boutique	Park
26	Richmond upon Thames	Outer London	Pub	Boat or Ferry	Garden Center	Café	Tea Room	Golf Course	German Restaurant	Scenic Lookout	Trail	Park
29	Tower Hamlets London	Inner London	Pub	Park	Café	Burger Joint	Art Gallery	Pizza Place	Convenience Store	Persian Restaurant	Light Rail Station	Breakfast Spot
30	Waltham Forest London	Outer London	Café	Coffee Shop	Pub	Park	Turkish Restaurant	Pool	Mini Golf	Tea Room	Golf Driving Range	Grocery Store
31	Wandsworth	Inner London	Pub	Hotel	Coffee Shop	Clothing Store	Gym / Fitness Center	Bus Stop	Pharmacy	Park	Supermarket	Breakfast Spot

d) Cluster “3”

Cluster "3" Results

```
In [59]: Borough_merged.loc[Borough_merged['Cluster Labels'] == 3, Borough_merged.columns[[0] + [1] + list(range(5, Borough_merged.shape[1]))]]
```

Out[59]:

	Borough	Inner/Outer London	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
2	Barnet	Outer London	Pub	Coffee Shop	Grocery Store	Bookstore	Fast Food Restaurant	Pizza Place	Pharmacy	Soccer Stadium	Park	Gym / Fitness Center
6	Camden London	Inner London	Pub	Coffee Shop	Café	Ice Cream Shop	Music Venue	Market	Burger Joint	Bar	Vegetarian / Vegan Restaurant	Grocery Store
19	Kensington and Chelsea	Inner London	Bakery	Ice Cream Shop	Italian Restaurant	French Restaurant	Bookstore	Plaza	Burger Joint	Coffee Shop	Restaurant	Pub
21	Lambeth	Inner London	Theater	Hotel	Coffee Shop	Bar	Café	Art Gallery	Pub	Plaza	Sandwich Place	Scenic Lookout
25	Redbridge	Outer London	Gym / Fitness Center	Hotel	Metro Station	Golf Course	Pizza Place	Park	Soccer Field	English Restaurant	Eastern European Restaurant	Pub
27	Southwark	Inner London	Theater	Pub	Hotel	Coffee Shop	Park	Seafood Restaurant	Scenic Lookout	Street Food Gathering	Italian Restaurant	Portuguese Restaurant
32	Westminster	Inner London	Hotel	Café	Coffee Shop	Plaza	Monument / Landmark	Park	Outdoor Sculpture	Hotel Bar	Garden	Historic Site

Discussion and Recommendations

Going through the details of the venues associated with each Borough cluster i.e. the top 10 venue categories associated with the London boroughs, it becomes apparent that cluster ‘0’ will be of most interest to the Travel Agency.

This is because the top 10 venue categories associated with this cluster matches the identified preferences considered to be of priority to the clients from sub-Saharan Africa countries i.e. Nigeria, Ghana, Kenya, Uganda and Zimbabwe during their visit to London UK.

Conclusion

This write-up provides the reader with an overview of a business problem which relates to a Travel Agency requiring access to strategic information on London Boroughs so as to enable it make appropriate arrangements for its holiday packages specifically aimed at its sub-Saharan African customers.

The unsupervised machine learning analyses was performed via Python scripts with support from Python Libraries like Pandas, GeoPy, Matplotlib, Scikit-learn and Folium.

Reliable data was collated and processed and the output result of the analyses can be said to provide a fit-for-purpose solution to the business challenges of the end client i.e. Travel Agency.