

# Applied Data Science Capstone Project

---

Which Boroughs Should Sub-Saharan Africans aim to Lodge in during visits to London UK?

By

**Olusoji Sogbetun**

**July 16, 2020**

# Table of Contents

---

- **Introduction**
- **Business Problem**
- **Data**
- **Methodology**
- **Results**
- **Discussions and Recommendation**
- **Conclusion**

## INTRODUCTION

### *Total Visitor Numbers from Sub-Saharan Africa to UK (2004 to 2018)\**

- Nigeria – 2,949,000
- Ghana – 459,300
- Kenya – 328,100
- Zimbabwe – 171,460
- Uganda – 101,550

*UK is a popular travel destination for the above English speaking sub-Saharan Africans who come to either visit friends and relatives, come for holidays or strictly for business*

\* Source –  
[www.gov.uk/government/publications](http://www.gov.uk/government/publications)



## BUSINESS PROBLEM

- Travel Agency based in Africa is getting poor reviews from Sub-Saharan African clients regarding the choice of hotel locations for their package holidays to London UK
- Solution proposed is to utilise technology i.e. Data Science / Machine Learning to help determine cluster of London Boroughs which fit customers' preferences

*According to an independent survey carried out by the Travel Agency, in addition to lodging in a nice hotel, the top three (3) preferences tourist from the mentioned sub-Saharan African countries look out for during visits to UK include their accessibility to;*

- Major Shopping Centres
- Popular Attraction / Sites
- Good Restaurant  
(with spicy option on menu)

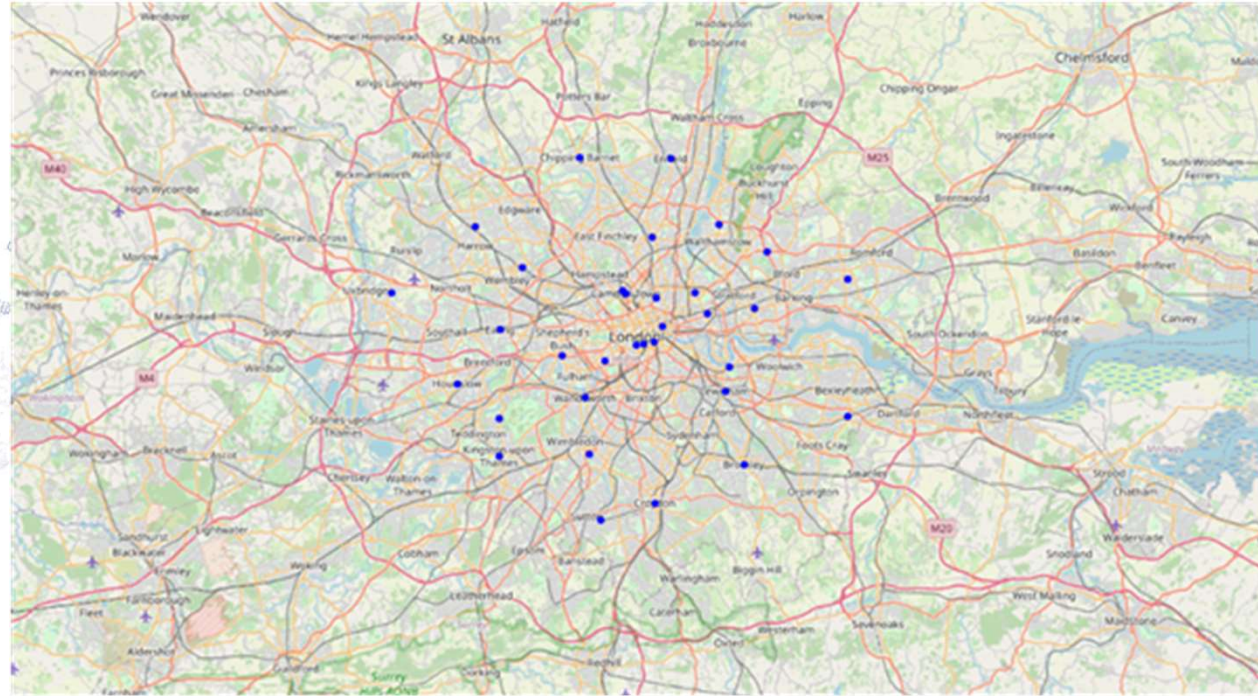


# DATA

## List of London Boroughs\*

- There are 32 Boroughs in London i.e. excluding the City of London
- The Geo-Coordinates for each Borough, including City of London was obtained via the GeoPy Library in the Python script
- Detailed data for the top venues situated in the vicinity of each London Borough was obtained using the Foursquare API

\* Source –  
[www.data.London.gov.uk/dataset/London-borough-profiles](http://www.data.London.gov.uk/dataset/London-borough-profiles)

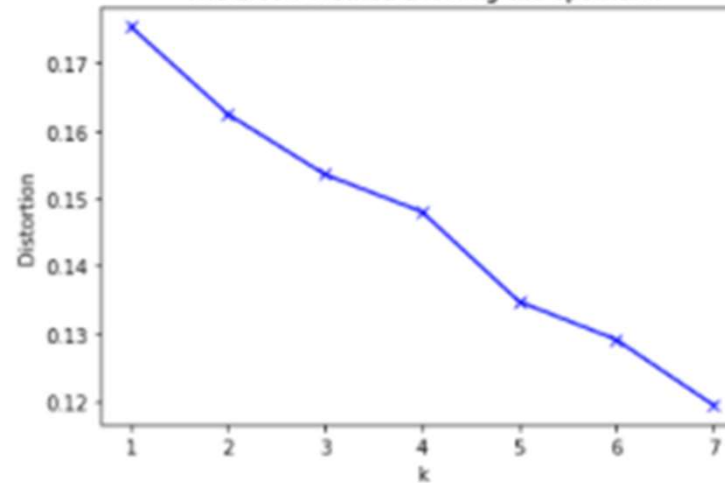




# METHODOLOGY

- The Machine Learning analyses is executed using appropriate Python scripts with support from Python libraries like Pandas, GeoPy, Scikit-Learn and Folium.
- Data is obtained from reliable sources, processed and analysed using the K-means clustering algorithm model
- The optimal number of clusters was determined using inertia (elbow rule ) method

The Elbow Method showing the optimal k



```
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	51.515518	-0.091995	0	Coffee Shop	Gym / Fitness Center	Hotel	Cocktail Bar	Garden	Scenic Lookout	French Restaurant	Steakhouse	Boating Dym	Modern European Restaurant
1	Barking and Dagenham	Outer London	51.554117	0.155554	0	Bus Stop	Gas Station	Convenience Store	Grocery Store	Liquor Store	Market Arts Digs	Gym	Gym / Fitness Center	Park	Supermarket
2	Barnet	Outer London	51.653060	-0.200225	3	Pub	Coffee Shop	Grocery Store	Bookstore	Fast Food Restaurant	Pizza Place	Pharmacy	Soccer Stadium	Park	Gym / Fitness Center
3	Bexley London	Outer London	51.441579	0.150455	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	51.563625	-0.275750	0	Coffee Shop	Clothing Store	Bar	Hotel	Supermarket	Sandwich Place	Indian Restaurant	Sporting Goods Shop	Grocery Store	Warehouse Store
5	Bromley	Outer London	51.402305	0.014814	2	Clothing Store	Pub	Coffee Shop	Cafe	Burger Joint	Indian Restaurant	Park	Pizza Place	Supermarket	Department Store
6	Camden London	Inner London	51.542305	-0.135550	3	Pub	Coffee Shop	Cafe	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	51.512555	-0.305195	2	Coffee Shop	Pub	Park	Italian Restaurant	Hotel	Bakery	Burger Joint	Pizza Place	Thai Restaurant	Cafe
9	Enfield	Outer London	51.652055	-0.081015	0	Pub	Clothing Store	Coffee Shop	Indian Restaurant	Department Store	Pharmacy	Park	Fish & Chips Shop	Supermarket	Pizza Place

# RESULT, DISCUSSION AND RECOMMENDATION

- Going through the results, it is apparent that cluster “0” will be of most interest to the Travel Agency
- The top 10 venue categories associated with this cluster matches the preferences considered to be of priority to the clients from sub-Saharan Africa.



## 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]))]]
```

	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

## CONCLUSIONS

- The presentation provides an overview of business problem, proposed solution to the challenge, sources of reliable data processed and analysed using unsupervised machine learning algorithm model (K-means).



# THANK YOU

---

- The Python Script for this project can be found via the GitHub link presented below

***<https://github.com/sojey71/Applied-Capstone-London/blob/master/Applied%20DataScience%20Capstone%20Project.ipynb>***