Capstone Project: The Battle of Neighbourhoods

by Fais Chalo

1. Introduction

1.1. Problem Description

Visiting a new city which seems to be to big to explore all in once is a problem when it comes to metropolies like London. Through a wide range of different cultures London developent economically, socially and politcially.

One great aspect of the diversity is the wide range of different eating possibilities.

But since it can be quite hard to distinguish between a good and bad restaurant and which one to choose there should be parameters in consideration to easen up the search. In my todays analysis i want to take the case of african food restaurants in London, which i recently heard of are quite popular.

1.2 Background and Motivation

Since my coworker is from Ghana and we recently got a new colleague from Senegal, my interest for the african food quitly increased since there are bringing all kind of delicious food to work.

My Ghanese coworker mentioned therefore that he is thinking about opening a Ghanese restaurant in London, but isn't quite sure about the current information provided.

Therefore i will lend him a hand in providing a deep analysis about the possibilities.

1.3. Possible Audience

Like i mentioned there is a high range of diversity in London and so we can conclude that besides the tourist also locals could be highly interested in finding the most visitable restaurant in whole london. Therefore the target audience are people who care about a high sophisticated african restaurant with high quality standarts.

2. Data

2.1 Description of Data ¶

In my Analysis i will take data from the sites Wikipedia and Foursquare.

Dataset 1

In this dataset i will examine data from the Areas of London which was provided by wikipedia.

The London Area has 32 Boroughs and the "City of London". You can find the data here - Greater London Area <https://en.wikipedia.org/wiki/List_of_areas_of_London >

Below we can see the demographic order for our analysis. Since we Lewisham has the highest range of black people we are going to use Lewisham for the analysis.

```
7]: demo_london_sorted.head(5)
```

7]:

	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

Lets check how categories are veloped in the area of lewisham which is our area of analysis.

```
nearby_venues_lewisham_unique.head(5)
```

	Count
Pub	13
Café	9
Gastropub	7
Park	6
Coffee Shop	5

This results clearly show that even though there restaurants in lewisham area, they are not in the top 5 count.

Lets check how categories are developed in multille city points instead like above for one single area.

```
se_venue_unique_count.head(5)
```

	Count
Pub	425
Coffee Shop	302
Café	296
Park	221
Grocery Store	161

3.1. Now we start to create Clusters.

Using the geopy library, the latitude and longitude values of London is obtained.

```
address = 'London, United Kingdom'

geolocator = Nominatim(user_agent="ln_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of London are {}, {}.'.format(latitude, longitude))
```

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

To obtain coordinates, use the folium library.

3.2.5 Optimal Number of Clusters

1. Elbow Method

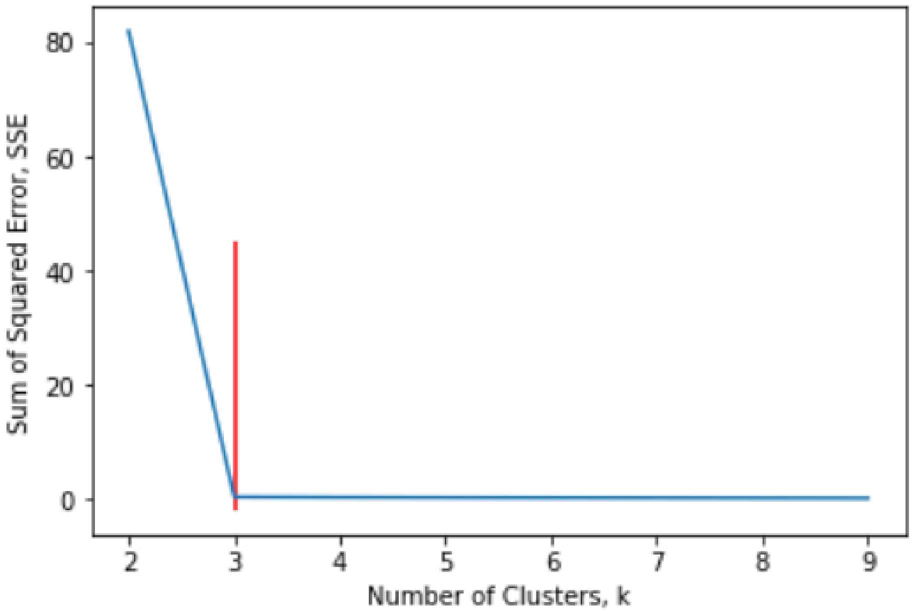
To check for the optimal number of clusters we are going to use the elbow method.

```
%matplotlib inline
import matplotlib
import numpy as np
```

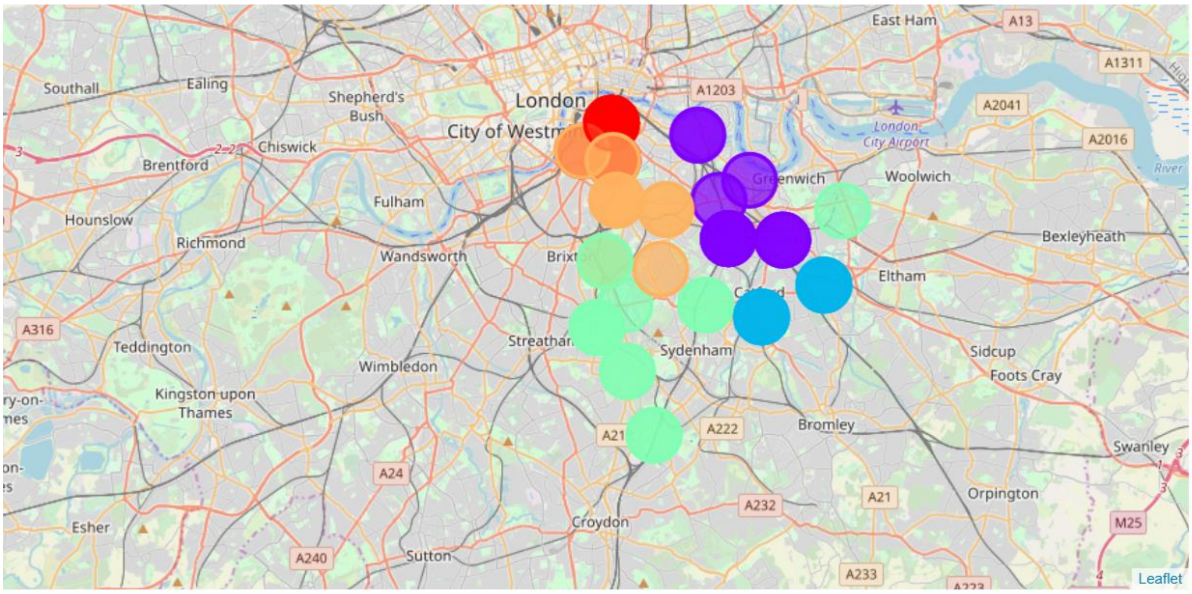
```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# SSE is initialize with empty values
# n_clusters is the "k"
sse = {}
for n_cluster1 in range(2, 10):
    kmeans1 = KMeans(n_clusters = n_cluster1, max_iter = 200).fit(se_grouped_clustering)
    se_grouped_clustering["clusters"] = kmeans1.labels_

    # The inertia is the sum of distances of samples to their closest cluster centre
    sse[n_cluster1] = kmeans1.inertia_
```



We used 200 iterations and concluded that 3 clusters would be quite optimal since the SSE is falling till Cluster number 3 in high range.



Cluster 1

```
se_clusters.loc[se_clusters['Cluster Labels'] == 0, se_clusters.columns[[1] + list(range(5, se_clusters.shape[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	Café	Coffee Shop	Pub	Hotel	Theater	Italian Restaurant	Park	Bar	Art Gallery	Street Food Gathering
6	Southwark	0	Café	Coffee Shop	Pub	Hotel	Theater	Italian Restaurant	Park	Bar	Art Gallery	Street Food Gathering
7	Southwark	0	Café	Coffee Shop	Pub	Hotel	Theater	Italian Restaurant	Park	Bar	Art Gallery	Street Food Gathering
8	Southwark	0	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Bar	Art Museum	Seafood Restaurant	Street Food Gathering	Scenic Lookout
18	Lambeth	0	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Bar	Art Museum	Seafood Restaurant	Street Food Gathering	Scenic Lookout
22	Southwark	0	Coffee Shop	Pub	Italian Restaurant	Hotel	Café	Theater	Street Food Gathering	Bar	Pizza Place	Art Gallery
23	Southwark	0	Coffee Shop	Pub	Italian Restaurant	Hotel	Café	Theater	Street Food Gathering	Bar	Pizza Place	Art Gallery

Cluster 2

```
se_clusters.loc[se_clusters['Cluster Labels'] == 1, se_clusters.columns[[1] + list(range(5, se_clusters.shape[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
0	Lewisham	1	Pub	Coffee Shop	Café	Park	Bar	Gastropub	Pizza Place	Italian Restaurant	Gym / Fitness Center	Turkish Restaurant
2	Lewisham	1	Pub	Coffee Shop	Café	Bar	Park	Garden	Vietnamese Restaurant	History Museum	Gym / Fitness Center	Brewery
14	Lewisham	1	Pub	Café	Gastropub	Park	Coffee Shop	Garden	Food Truck	Supermarket	Fish & Chips Shop	Restaurant
16	Lewisham	1	Pub	Café	Coffee Shop	Park	Gastropub	Food Truck	Gym / Fitness Center	Fish & Chips Shop	Italian Restaurant	Bar
17	Lewisham	1	Pub	Café	Coffee Shop	Park	Gastropub	Food Truck	Gym / Fitness Center	Fish & Chips Shop	Italian Restaurant	Bar
20	Lewisham	1	Pub	Café	Gastropub	Park	Coffee Shop	Garden	Food Truck	Supermarket	Fish & Chips Shop	Restaurant

Cluster 3

```
se_clusters.loc[se_clusters['Cluster Labels'] == 2, se_clusters.columns[[1] + list(range(5, se_clusters.shape[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
12	Lewisham	2	Pub	Grocery Store	Café	Park	Italian Restaurant	Fish & Chips Shop	Fast Food Restaurant	Coffee Shop	Supermarket	Train Station
19	Lewisham	2	Pub	Grocery Store	Café	Park	Italian Restaurant	Fish & Chips Shop	Fast Food Restaurant	Coffee Shop	Supermarket	Train Station
26	Lewisham	2	Grocery Store	Park	Supermarket	Café	Pub	Coffee Shop	Fast Food Restaurant	Gas Station	Italian Restaurant	Bus Stop
32	Lewisham	2	Grocery Store	Park	Supermarket	Café	Pub	Coffee Shop	Fast Food Restaurant	Gas Station	Italian Restaurant	Bus Stop
44	Lewisham	2	Grocery Store	Park	Supermarket	Café	Pub	Coffee Shop	Fast Food Restaurant	Gas Station	Italian Restaurant	Bus Stop
45	Lewisham	2	Pub	Grocery Store	Café	Park	Italian Restaurant	Fish & Chips Shop	Fast Food Restaurant	Coffee Shop	Supermarket	Train Station

4. My Results

The following are the highlights of the 3 clusters above:

1. Pubs, Coffee shops and Grocery Stores are highly popular in Lewisham and area of Southwark.
2. For restaurants we can say that italian restaurants seem to be quie popular in SE.
3. This results suprisingly shows that even though there is such a high range of different ethnicities in SE, theres no clear direction or tendency for african restaurants.
4. Publs are quite well presented .

5. Conclusion

Since we just had a limited range of data available we could not conclude if the best option would be to start an african business in the area of Lewisham. What we can say is that other kind of social spots are more popular than typical "african restaurants.

Further on we have to accept the limitations of our data in kind of different category features for our polation such as crime rate, interest of food etc.