

Capstone Project

Using location data to identify places to open a
new West African restaurant

Ansu Edward Turay
IBM
5 January 2021

1. Introduction

1.1. Background

Based on data retrieved from the London datastore - the population of migrants since 1821 has steadily increased. Cultural integration via intermarriages, friendships built within the workforce or academia etc has led to positive adaptations of social norms. This is evident in food and music. People are more forthcoming in accepting new food delicacies - which otherwise would have only been found abroad. London has had its taste buds ravaging for the Indian, Chinese and Caribbean cuisines and now recent reports, suggest there is a new interest - West African dishes. As a consequence, several West African restaurants have been opening up across the London. This surge maybe due to the dense population of West Africans living these areas.

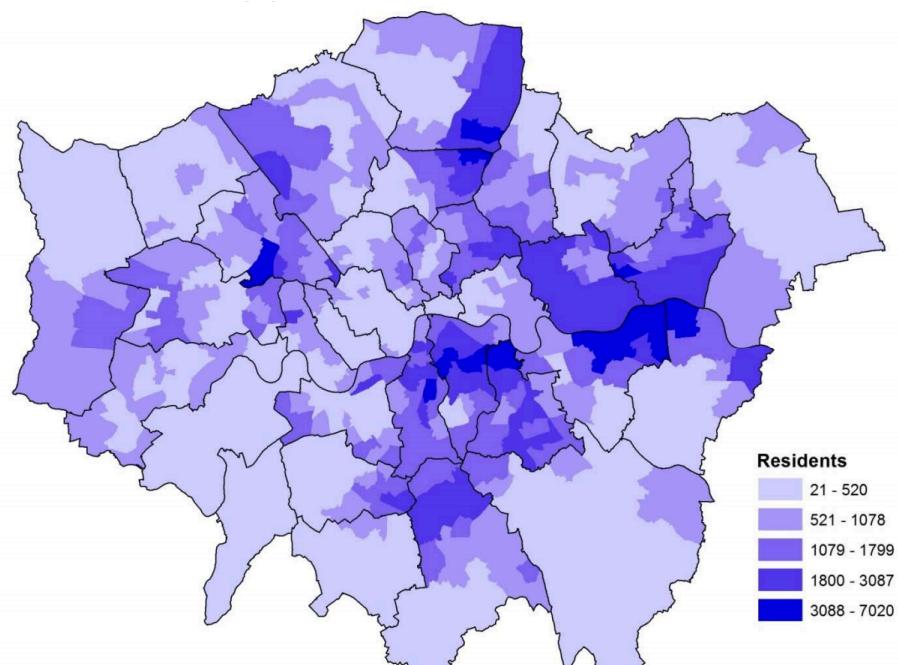


Fig.1 - Distribution of the African population across London 2011 by GLA
intelligence unit

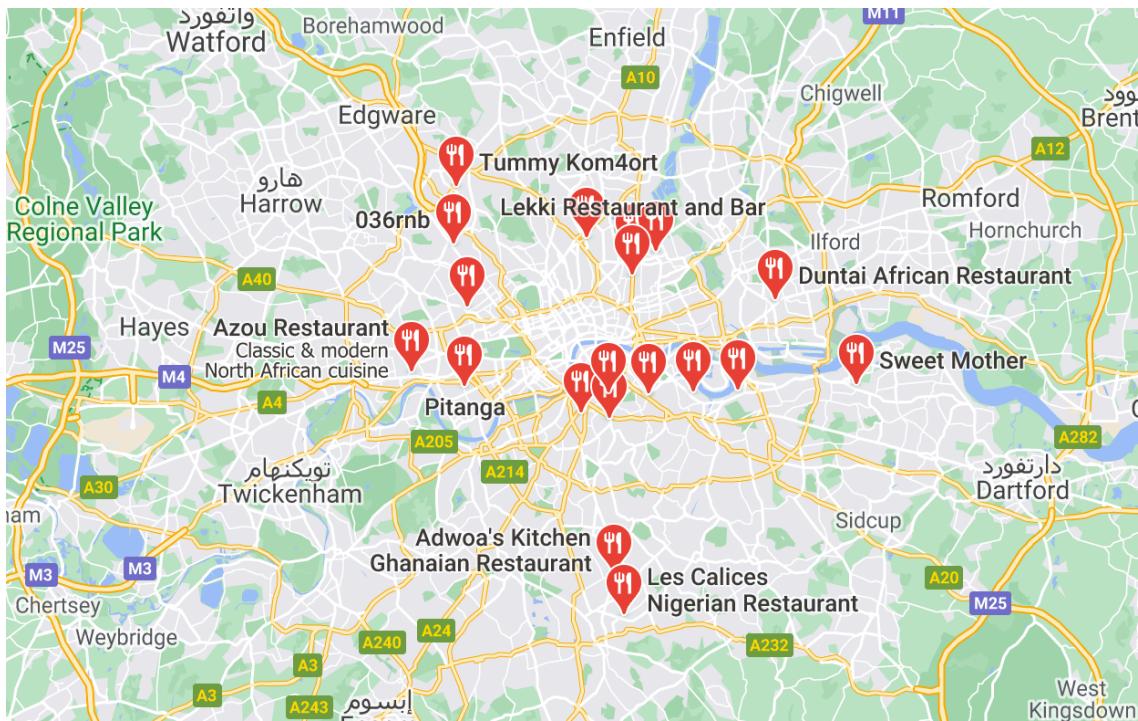


Fig.2 - A few known African restaurants in London 2021.

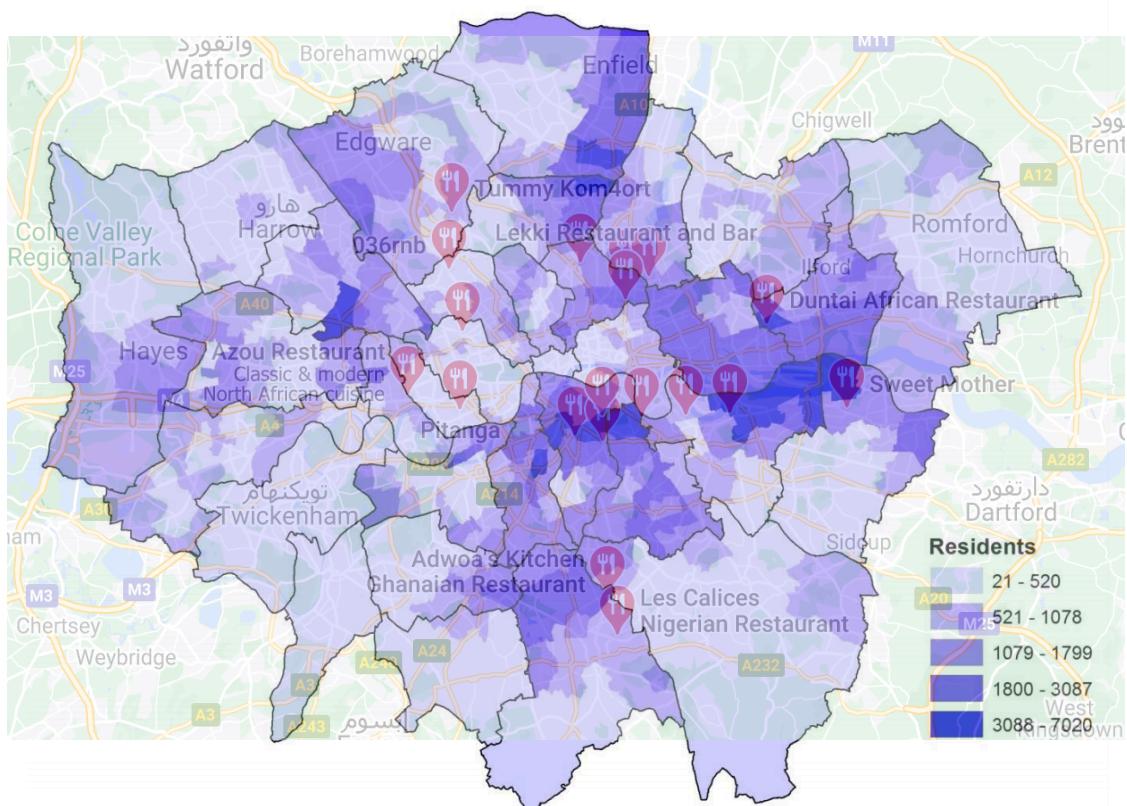


Fig.3 - Fig 1 and 2 superimposed

The aim of this report is to provide investors/business owners suitable areas where they could potentially open a new West African restaurant:

1. In a location that is not heavily populated with other competing restaurants.
2. Within or outside African populated areas.

1.2. Problem

Data will be gathered mainly through *Foursquare* however, many West African restaurants are not captured within it. Therefore, our analysis would be incomplete - if we're to rely on this data alone. Below is a list of potential resources we could investigate alongside *Foursquare*:

1. Instagram
2. JustEat
3. Zomato
4. Tripadvisor
5. Deliveroo
6. UberEats
7. Facebook
8. Yelp

We will aim to extract restaurant information - mainly their name and address - from *Foursquare* and the list of external sources. This will help us plot their locations on a Folium map and determine the best possible locations for opening a new restaurant.

2. Data acquisition and cleaning

2.1. Location

We decided to use regularly spaced grid of locations, centered around London city center, to define our neighborhoods. The following data sources will be needed to extract/generate the required information:

- Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding
- The number of restaurants and their type and location in every neighborhood will be obtained using **Foursquare API** and **Webgrinder**
- Coordinates of central London will be obtained using Google Maps API geocoding

	Latitude	Longitude	UTM X	UTM Y
Central London	51.507	-0.128	-547014	5815565

Table.1 - Latitude and Longitude (Lat/Lon) was generated using the Google Maps API. UTM Cartesian coordinates (X/Y coordinates in meters) were converted from Lat/Lon. This will be used to calculate the distances from potential areas and nearby restaurants.

A grid of area candidates, equally spaced, centered around city center and within ~6km from London was created (*Fig.4*). Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart.

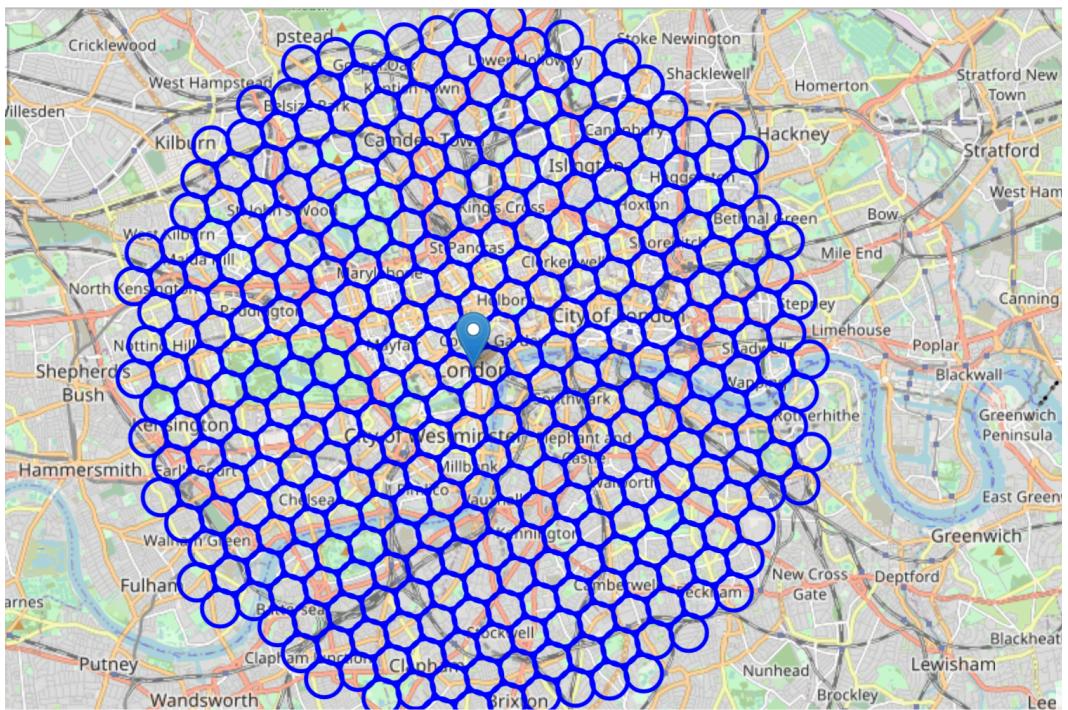


Fig.4 - 364 candidate neighborhood centers generated that are equally spaced and within 6km from the center of London

Address	Latitude	Longitude	X	Y	Distance from center
Warley House, 91 Elms Cres,	51.45	-0.136	-548814	5.81E+10	5992
4 Prague Pl, Brixton,	51.46	-0.128	-548214	5.81E+10	5840
56 Brixton Hill,	51.46	-0.119	-547614	5.81E+10	5747

Table.2 - Some addresses from some of the candidate areas

2.2. Data sources

After some investigation we were able to web scrape the meta data, presented in Table.3, from websites recorded in *Table.3*. With regards to the remaining web sources, mentioned in 1.2, we was unable to extract any unique restaurant data.

Restaurant Name	Address
-----------------	---------

Table.3 - metadata fields

Websites	African Restaurants Scraped
Zomato	235
FourSquare	29
Yelp	227
Tripadvisor	158

Table.4 - Websites used and no. potential African restaurants scraped within London.

2.3. Data cleaning

The *Webgrinder* tool was used to extract the two metadata fields (Table.2) for the known African restaurants from Tomato, Yelp and Tripadvisor. These files were stored as Comma-Separated Values (CSV) and uploaded to the notebook stored on Google Collab - see *Fig.4 below:*

```

yelp.head()

No. potential African Restaurants: 228
No. Columns: 11

   column_2      column_3-link column_1-alt column_3      column_1-imgsw column_9      column_8 column_10
0  ["1."]  ["https://www.yelp.com/biz/african-kitchen-gal..."]  ["African Kitchen Gallery"]  ["African Kitchen Gallery"]  ["https://s3-media0.fl.yelpcdn.com/bphoto/oEaN..."]  ["102 Drummond Street"]  ["020 7383 0918","Euston"]
1  ["2."]  ["https://www.yelp.com/biz/ikoyi-london?osq=Af..."]  ["[ikoyi]"]  ["[ikoyi]"]  ["https://s3-media0.fl.yelpcdn.com/bphoto/TtOk..."]  ["1 St James's Market"]  ["020 3583 4660","St James's"]

["", refined, and has an equal footing amongst...

```

Fig.4 - Restaurant data extracted from Yelp

All three data sets (Yelp, Zomato and Tripadvisor) were combined and data columns such as reviews, star rating, website links were removed (*Table.5*).

No.	Restaurant	Address
223	Tortuga Caribbean Restaurant & Cocktail Bar	197-199 Replingham Road
224	Casablanca	170 Sandringham Road
225	Rudie's	50 Stoke Newington Road
226	Nando's	Cabots Place East
227	Ambassador African Restaurant	1a Spring Street

Table.5 - The last 5 restaurants of the merged dataset (Yelp, Tomato & Tripadvisor)

199 restaurants had missing addresses so we used the Google api to find these addresses and then we searched for Longitude and Latitude for the merged dataset.

```
[ 36 ] addresses[180:190]
```

```
['16 Stukeley Street, Holborn, London WC2B 5LQ, UK',
 '510A High Rd, Tottenham, London N17 9JF, UK',
 '400 Wharfside Way, Jacksonville, FL 32207, USA',
 '2027 Lexington Ave, New York, NY 10035, USA',
 '316 Old St, Shoreditch, London EC1V 9DR, UK',
 '34 Thomas St, Woolwich, London SE18 6HT, UK',
 'Macey House, 40 Surrey Ln, Battersea, London SW11 3PS, UK',
 '38 Choumert Rd, Peckham, London SE15 4SE, UK',
 '117 Bensham Ln, Thornton Heath CR7 7EU, UK',
 'UK Ltd, 91 Brick Ln, Spitalfields, London E1 6QL, UK']
```

Fig.5 - Example of 10 addresses recovered using the Google API

As you can see in Fig.5, some addresses were referencing locations in America. As a consequence of this, these addresses were removed as our scope is only restaurants in London. Once all restaurants had their address populated we used this to obtain the Longitude and Latitude - which will be used to plot on the Folium map later.

	Name	Address	Latitude	Longitude
0	Afrik'N'Fusion	308-310 North End Rd, Fulham, London SW6 1NQ, UK	51.483017	-0.200791
1	Wolkite restaurant	82 Hornsey Rd, London N7 7NN, UK	51.556129	-0.111073
2	Zeret Kitchen	216-218 Camberwell Rd, Camberwell, London SE5 ...	51.478691	-0.094690

Table.6 - The first 3 restaurants of the merged dataset (Yelp, Tomato & Tripadvisor).

Total restaurants retrieved and deduplication etc. is 183.

Next, the dataset was deduplicated against the restaurant data retrieved with the Foursquare API - targeting the unique IDs and then a combination of "Name" + "Address".

3. Explanatory analysis

3.1. Restaurant mapping

Final dataset consisting of restaurants from Yelp, Zomato, Tripadvisor and Foursquare brought our total of 2688. This was plotted on a map using the Folium plugin (*Fig.6*).

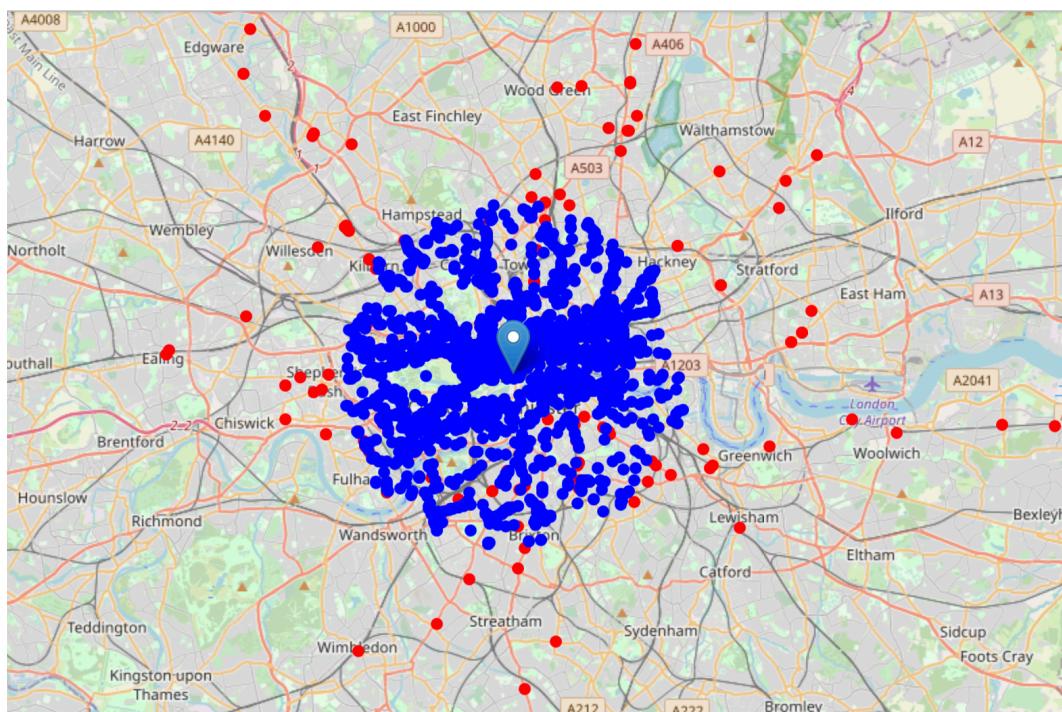


Fig.6 - Plot of all 2688 restaurants. The blue dots represents non African Restaurants and the red represents African restaurants in and outside the 6km radius from the center of London.

Second step in the analysis was calculation and exploration of '**restaurant density**' across different areas of central London - using **heat maps** (*Fig.7*) to identify a few promising areas close to the center of London with a low number of restaurants in general (and no African restaurants in vicinity).

Address	Distance from center	Restaurants in area
Warley House, 91 Elms Cres, Clapham Common,	5992.495307	5
4 Prague Pl, Brixton,	5840.376700	0
56 Brixton Hill, Brixton,	5747.173218	2
57 Rattray Rd, Coldharbour,	5715.767665	0
Bessemere Park, 250 Milkwood Rd,	5747.173218	0
234 Denmark Hill, Brixton,	5840.376700	0

Table.7 - Number of restaurants within 6 candidate areas. The distance from the center is the distance from the nearest restaurant to the center of the area.

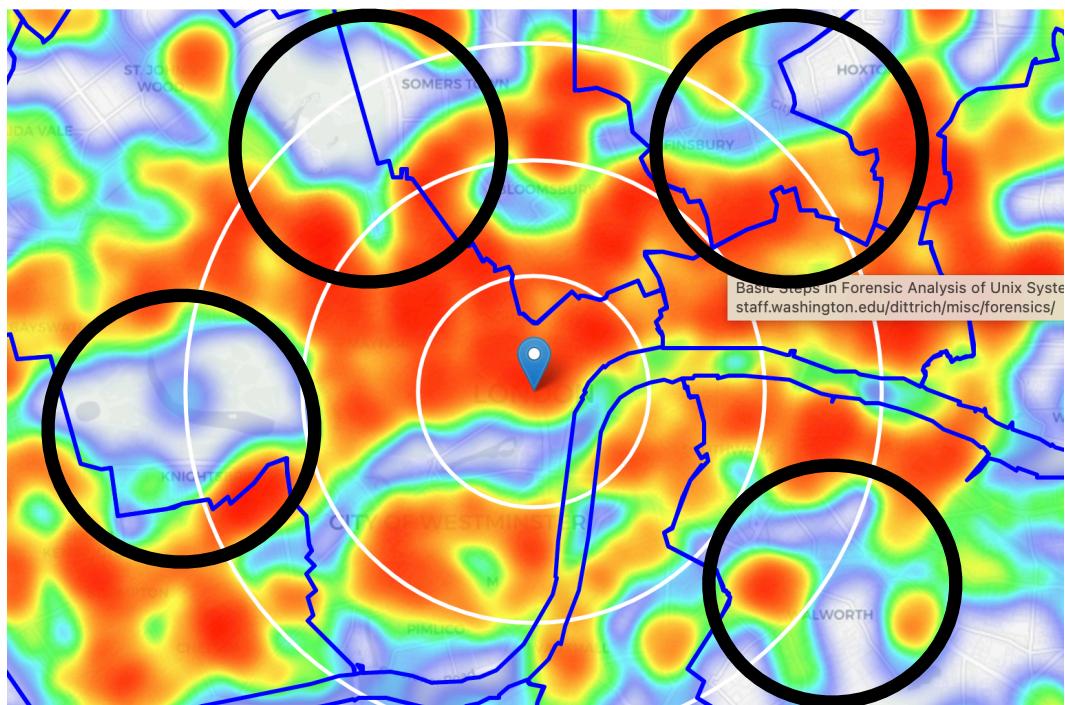


Fig.7 - Heat map of restaurant density in London. The redder the area the more restaurants are located in that area. There is low restaurants in parts of Knightsbridge, Somers Town, Hoxton and Walworth etc.

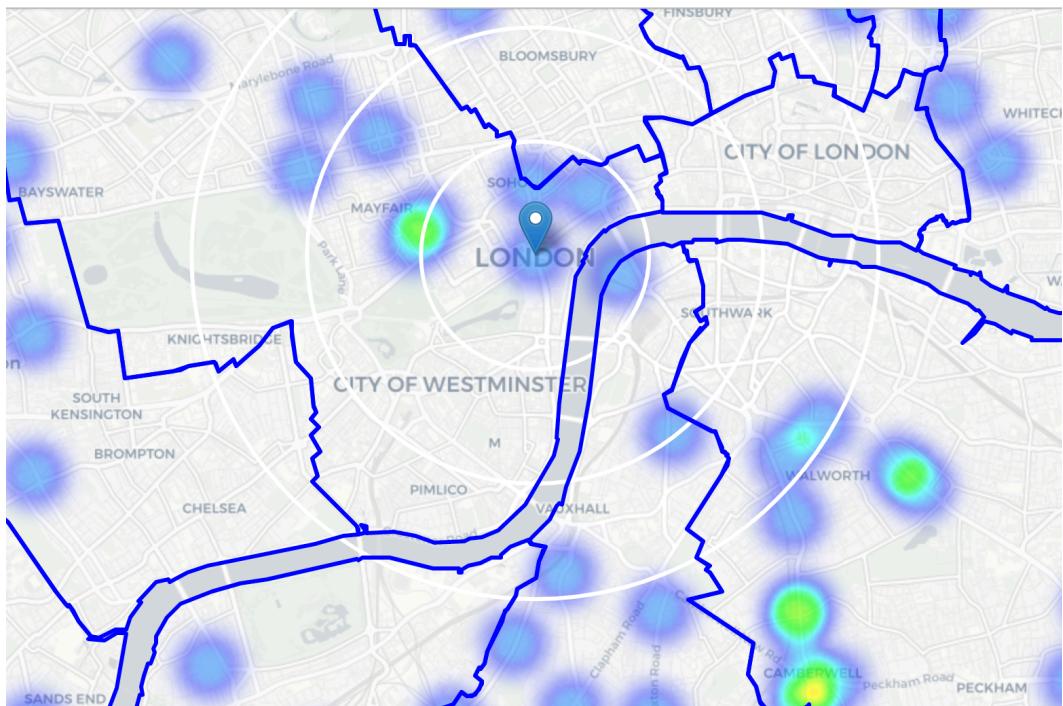


Fig.7 - Heat map of only African restaurant density in London

Based on this, the focus was stirred towards analyzing areas south-east and east of London. Therefore the central area of interest was reduced to have a radius of **2.5km**. This places our location candidates mostly in boroughs **Walworth and City of London** (another potentially interesting borough is **Wandsworth** with large low restaurant density south-west from city center, however this borough is less interesting to stakeholders as it's mostly residential, less popular with tourists and less africans live here).

3.2. New restaurant locations

With a maximum of two restaurants in the vicinity (radius of 250 meters) and no African restaurants within radius of 400 meters was used to determine suitable locations.

Latitude	Longitude	X	Y	Restaurants nearby	Distance to an African restaurant
51.472453	-0.083196	-544764	5.81E+10	4	1050
51.472636	-0.081806	-544664	5.81E+10	4	1113
51.472198	-0.091095	-545314	5.81E+10	18	760
51.472381	-0.089705	-545214	5.81E+10	13	777

Table.8 - location data of restaurants that fall outside the scope. Locations with no more than two restaurants nearby was **1245**. Locations with no African restaurants within 400m was **2060**. And locations with both conditions met was **1104**.

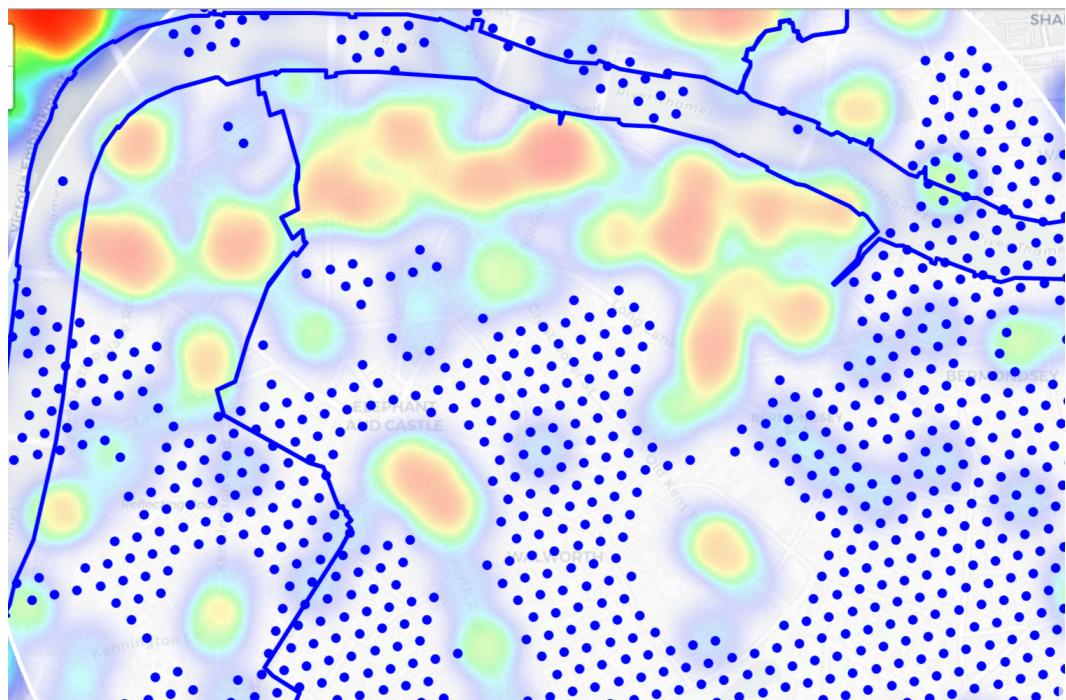


Fig.7 - Heat map of known restaurants along with 1104 locations (blue dots) that fit both conditions mentioned earlier.

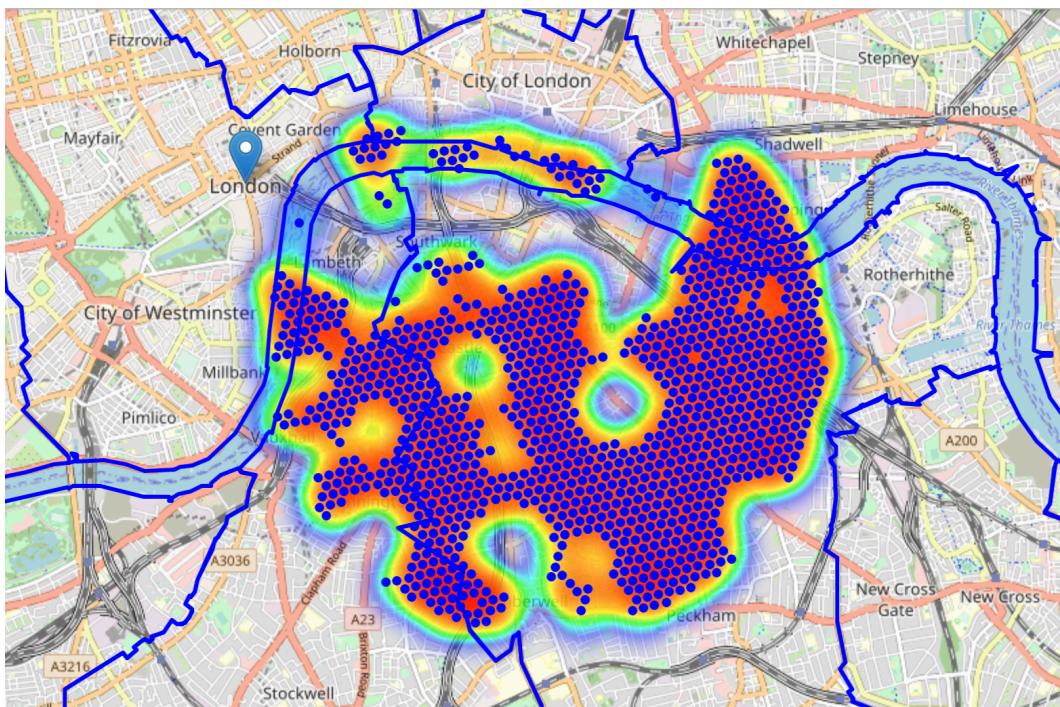


Fig.8 - Heat map representing the 1104 locations only

We then **clustered the above** locations to create **centers of zones containing good locations**. Those zones, their centers and addresses will be the final result of our analysis.

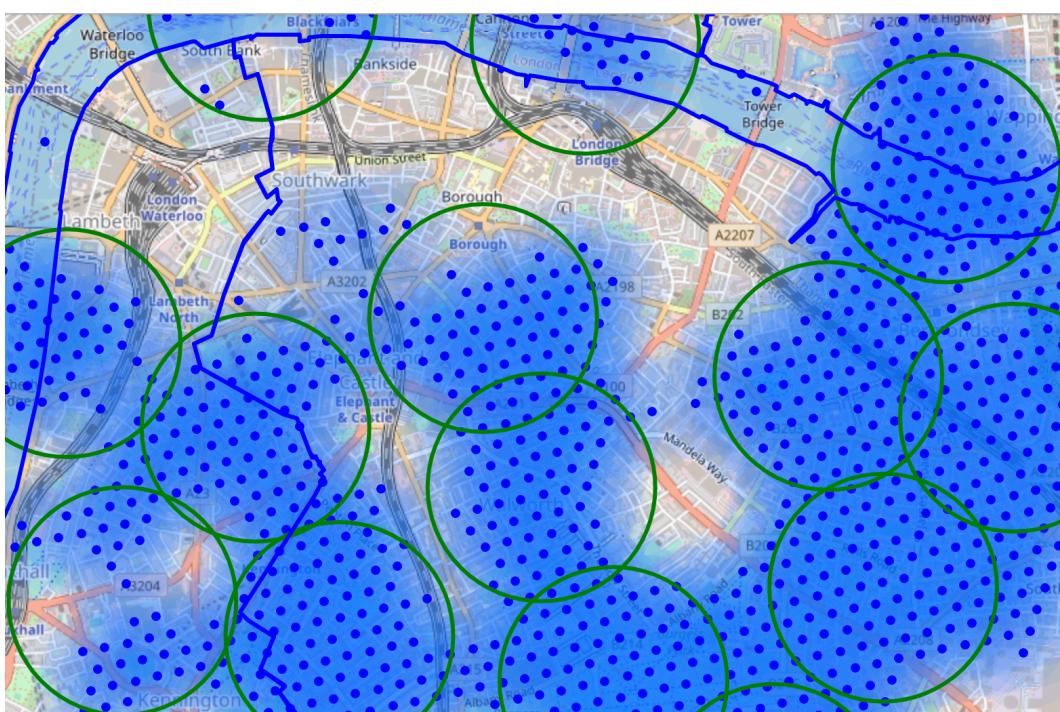


Fig.9 - 15 clustered locations

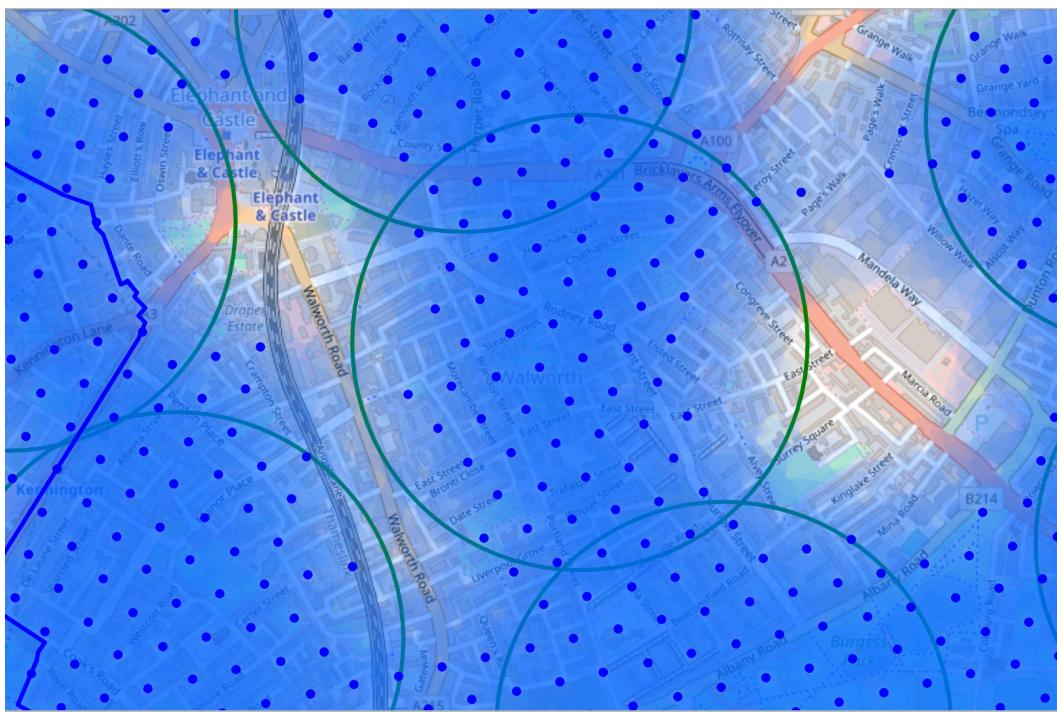


Fig.10 - Zoomed in view on the cluster within Walworth

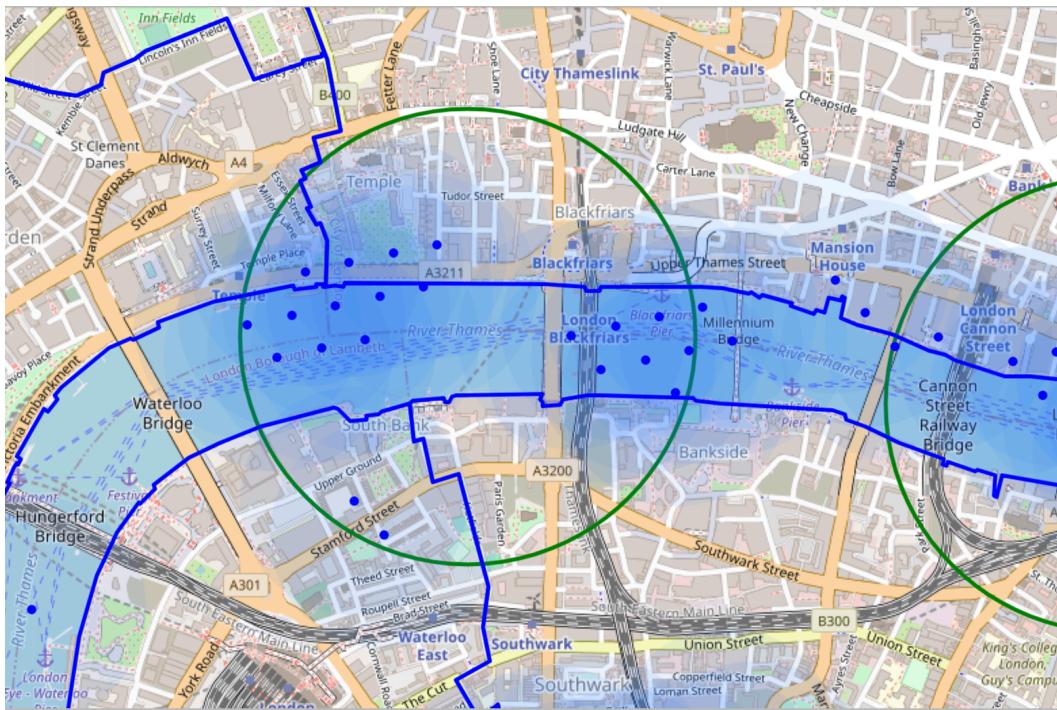


Fig.10 - Zoomed in view on the cluster within Blackfriars

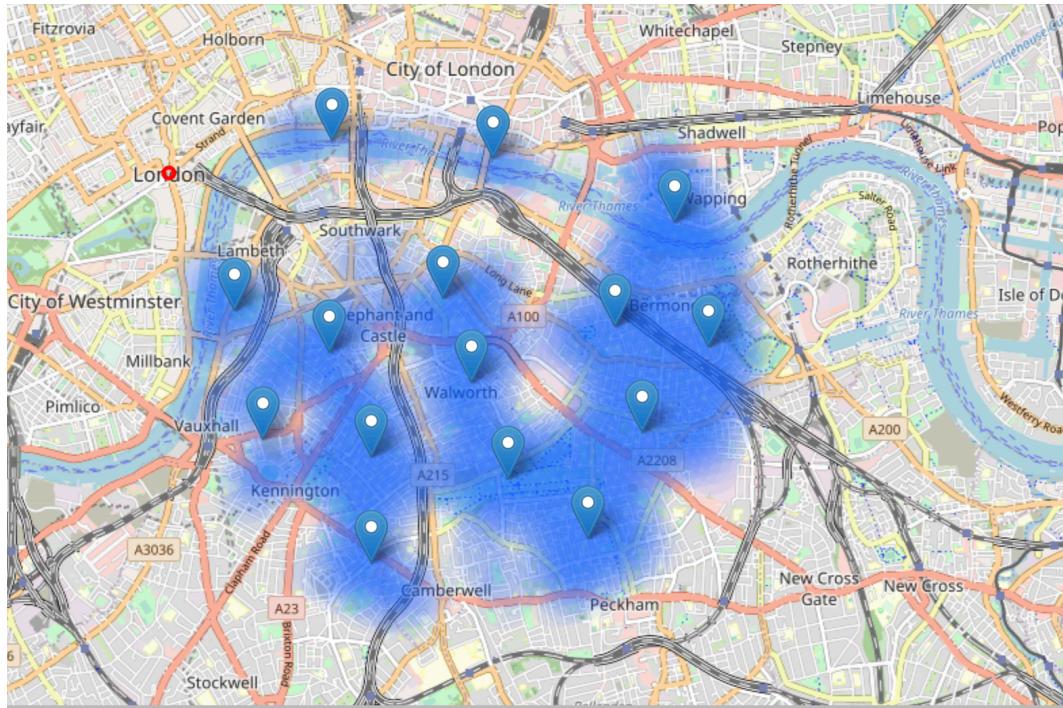


Fig.11 - 15 candidate addresses to open a new African restaurant

Address	DFC	Address	DFC
121 Cator St, London SE15 6PD	4.9	Dawes House, 117 Orb St, London SE17 1ES	3.3
31 Sullivan Rd, Prince's SE11 4UH	2.1	The Foundry, 17 Oval Way, Oval, SE11 5RR	2.5
8 Layard Rd, Bermondsey, SE16 2JE	5.0	12 Chumleigh St, SE5 ORN	4.1
Dickens Fields, 10 Dickens Square, SE1 4JL	2.7	30 Lorrimore Square, SE17 3QU	3.1
Langton Mews, 53 Langton Rd, Camberwell, SW9 6UG	3.9	55 Goodwin Cl, Bermondsey SE16 3TL	4.2
9 Ainsdale Dr, SE1 5JY	4.8	Adelaide House, Bridge, EC4R	2.9
163 Lambeth Palace Rd, South Bank, SE1 7JU	1.4	2 Pier Head, Wapping High St, St Katharine's & Wapping, E1W 1PN	4.5
Rennie Court, 20 Stamford St, South Bank, SE1 9LP	1.5		

Table.9 - 15 potential addresses. The Distance From Central London (DFC) is in km.

4. Discussion and Conclusion

4.1 Discussion

The limitation we found whilst conducting our analysis was that we were only able to take in account restaurants that had registered with one of the known providers. Therefore, any traditional restaurants were not taken into account. Also there are some mobile caterers like that could be direct competitors but we was unable to extract their data efficiently from areas like Instagram etc.

The distance from the nearest African restaurant to the center of one of the candidate areas is based on the data taken from Foursquare (*Table.9*) - thus, it does not include the 183 African restaurants taken from Yelp, Zomato and Tripadvisor. However, *Fig.6* captures all 2688 restaurants in and around London. Therefore, a combination of *Fig.6* and *Fig.11* in order to determine whether the final 15 addresses are worth investigating further.

4.2 Conclusion

Purpose of this project was to identify central London areas close to center with low number of restaurants (particularly African restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new African restaurant. By calculating restaurant density distribution from Foursquare data we have first identified general boroughs that justify further analysis (Walworth and Blackfriars), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest

number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.