

**FORTUNE ADEKOGBE**

**COURSERA CAPSTONE PROJECT**

**Battle of the Neighborhoods:**

**Location selection for a  
Restaurant start-up**

# INTRODUCTION

Most start-ups find choosing a location for their businesses difficult. Even after streamlining to a populated City or State, narrowing down to a specific Local Government Area is usually an issue. I aim to solve this issue by Engineering a system that analyses the different areas in a State and gives the Entrepreneur the best possible locations to start with.

For this project, **the target audience are restaurants just starting up or planning to open a branch in a new city.** Having had interactions with quite a number of Entrepreneurs facing this dilemma and I am sure the results of this will be useful to them. This analysis assumes that these restaurants target employees of organizations as customers and thus aims to provide feedback that will help in pinpointing areas densely populated with Offices and Organizations. For restaurants start-ups going into food delivery but with limited resources to go all round the city, I will be suggesting groups of Neighborhoods for them to start with.

This can of course be easily tweaked if need be. One of the aims of this project is to create a system that can be easily modified to suit a new situation.

The selected state for this project is Lagos state the financial capital of Nigeria and the state with the highest population density. It is also where I reside and I aim to provide a local solution to this problem.

# DATA

Getting data for this project was a nigh herculean task but I was successful. I was able to obtain [this json file](#) from the [HUMANITARIAN DATA EXCHANGE JSON REPOSITORY](#). It contains the geographical details of Local Government Areas in Nigeria and I extracted those of Lagos from it. A section of the .json file is displayed below:

```
{ "type": "Feature", "geometry": { "type": "Polygon", "coordinates": [[[ 3.332283973693848, 6.648791790008602], [ 3.328931093216056, 6.643056869506893], [ 3.326539993286247, 6.623510837554989], [ 3.31771993637085, 6.620203971862793], [ 3.301352977752799, 6.620390892028865], [ 3.290610074996948, 6.627117156982422], [ 3.286653041839657, 6.653451919555721], [ 3.28961110115057, 6.672671794891414], [ 3.30701494216919, 6.657254219055176], [ 3.317653894424552, 6.644631862640495], [ 3.332283973693848, 6.648791790008602]]]], "properties": { "ID_0": 163, "ISO": "NGA", "NAME_0": "Nigeria", "ID_1": 25, "NAME_1": "Lagos", "ID_2": 507, "NAME_2": "Agege", "HASC_2": null, "CCN_2": 0, "CCA_2": null, "TYPE_2": "Local Authority", "ENGTYPE_2": "Local Authority", "NL_NAME_2": null, "VARNAME_2": null}}, { "type": "Feature", "geometry": { "type": "Polygon", "coordinates": [[[ 3.307024002075252, 6.477795124054069], [ 3.3164160251618, 6.482667922973633], [ 3.326337099075317, 6.482730865478516], [ 3.338913917541618, 6.475834846496696], [ 3.342958927154598, 6.455029010772762], [ 3.331623077392578, 6.452916145324821], [ 3.322557926178035, 6.446716785430908], [ 3.312277078628597, 6.454857826232967], [ 3.306322097778434, 6.471779823303223], [ 3.307024002075252, 6.477795124054069]]]], "properties": { "ID_0": 163, "ISO": "NGA", "NAME_0": "Nigeria", "ID_1": 25, "NAME_1": "Lagos", "ID_2": 508, "NAME_2": "Ajeromi/Ifelodun", "HASC_2": null, "CCN_2": 0, "CCA_2": null, "TYPE_2": "Local Authority", "ENGTYPE_2": "Local Authority", "NL_NAME_2": null, "VARNAME_2": null}}
```

From this data, the different Local Government Areas in Lagos state were obtained. The geopy library was then used to get the latitude and longitude for each one of them. The Polygon coordinates in this json file were used to display a Choropleth to visualize the result of the analysis.

I also used the [Foursquare location data](#) to get information about the offices in the different areas and their corresponding categories. Below is a sample Data Frame of the foursquare search result for Kosofe Local Government Area in Lagos using an office query.

As seen, the names of the offices can be obtained. The category they belong to can also be obtained. With this information and some further analyses, a clustering algorithm can be run to determine Local Government Areas that are similar to each other. The latitude and longitude of each office can also be gotten which is useful for visualization with the folium library.

[41]:

	name	categories	lat	lng
0	Ketu Kosofe Post Office	Post Office	6.592355	3.394546
1	Office Everything	Paper / Office Supplies Store	6.613891	3.357690
2	Ketu Kosofe Post Office	Post Office	6.605957	3.384407
3	Akinwunmi Ambode Campaign Office	Non-Profit	6.550345	3.389895
4	FRSC Office Bariga	Government Building	6.549866	3.393374
5	Dstv office at yaba	Miscellaneous Shop	6.551776	3.378519
6	LASAA Office Aloha Plaza	Building	6.616348	3.386052

## METHODOLOGY

To obtain data about the Local Government Areas (which will be called LGAs subsequently) in Lagos, I installed the geopandas library. With this, I was able to read the Nigerian LGAs .json file into a geopandas data frame. I then filtered the datapoints pertaining to Lagos LGAs into a new data frame and wrote this into a .geojson file for later use.

From this same dataframe, I extracted the feature showing the LGAs in Lagos into a new data frame which is shown below.

```
In [78]: lagos[['ID_2', 'NAME_2']]
```

```
Out[78]:
```

	ID_2	NAME_2
506	507	Agege
507	508	Ajeromi/Ifelodun
508	509	Alimosho
509	510	Amuwo Odofin
510	511	Apapa
511	512	Badagary

I then used the geopy module to obtain the latitude and longitude of these LGAs. Whilst doing this, an `AttributeError` was raised. I thus included a try block to view the LGAs causing the errors.

With this I identified that 2 LGAs were wrongly named.(Badagary instead of Badagry and LagosIsland instead of Lagos Island).

```
In [8]: coordinates = {'LGA':[], 'latitude':[], 'longitude':[]}
for i in lagos_df['NAME_2']:
    address = f'{i}, Lagos'
    geolocator = Nominatim(user_agent="lag_explorer")
    location = geolocator.geocode(address)
    try:
        coordinates['latitude'].append(location.latitude)
        coordinates['longitude'].append(location.longitude)
        coordinates['LGA'].append(i)
    except AttributeError:
        print(f'{i} is a problem')
```

I then used the `pd.DataFrame.replace()` function to correct this and created a data frame with the LGAs as well as their geographical coordinates. The index was set to the LGA column as shown below.

Out[9]:

	latitude	longitude
LGA		
Agege	6.625256	3.311209
Ajeromi/Ifelodun	6.455122	3.335946
Alimosho	6.584343	3.257631
Amuwo Odofin	6.447023	3.266280
Apapa	6.445187	3.368373
Badagry	6.438730	2.907486
Epe	6.524750	4.057246
Eti-Osa	6.460231	3.532181
Ibeju/Lekki	6.471125	3.814750
Ifako/Ijaye	6.660436	3.321539

I moved on to display the world map centered around Lagos with the folium library.

I created 2 functions to help with extracting foursquare location data. The first was named `get_category_type`. This takes in row data from a dataframe and returns the category of that location obtained from foursquare. The second called `get_foursquare` does the actual information retrieval. It takes in a LGA as string and a query string then returns a data frame containing the names, categories, latitudes, longitudes and LGAs of the different retrieved Locations. The function `get_category_type` gets called in the function `get_foursquare`.

With these helper functions, I created an empty data frame to store all retrieved locations. I then looped through the LGAs then a list of queries. Each query in the list represents a common phrase found in office names. For each LGA and query, I called the `get_foursquare` function and cumulated the resulting data frames. A section of this data is shown below.

Out[19]:

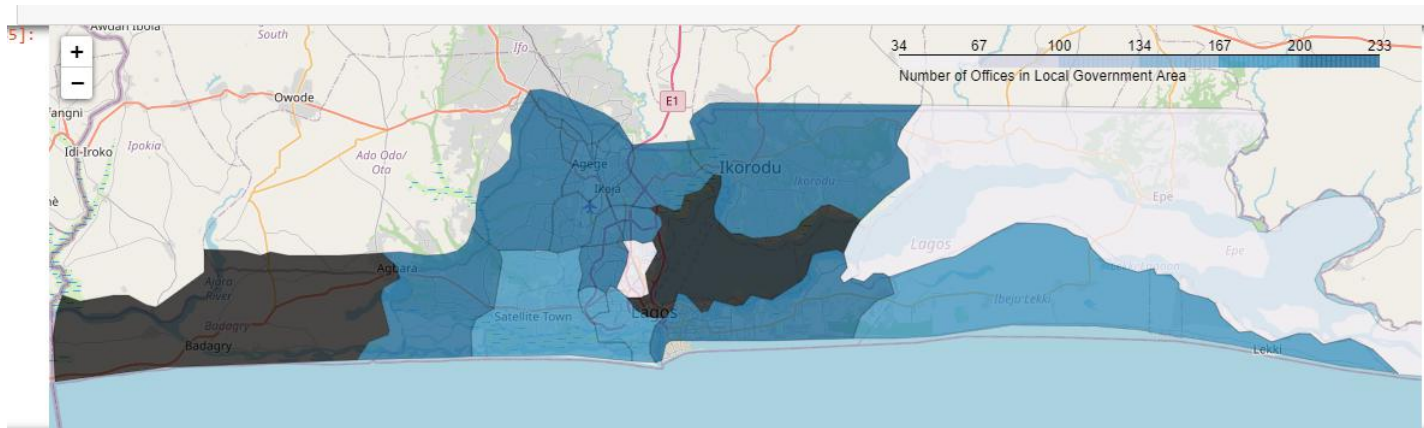
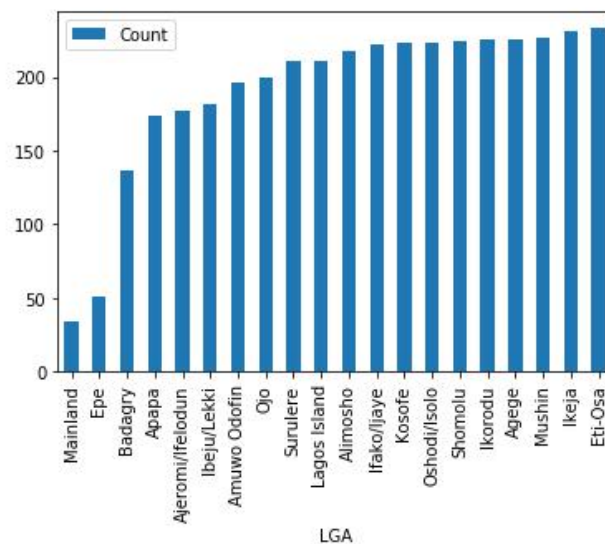
	name	categories	lat	lng	LGA
0	Standards Organisation of Nigeria	Government Building	6.438771	3.456322	Agege
1	SoluTech Business Solutions	None	6.633049	3.306640	Agege
2	Business Class Lounge	Travel Lounge	6.584789	3.332613	Agege
3	Business With Measure Digital Marketing Agency	Advertising Agency	6.586949	3.356983	Agege
4	NECA House Plot A2, Hakeem Balogun Street, Cen...	Conference Room	6.623344	3.358132	Agege

In [20]: `df.shape`

Out[20]: (4604, 5)

I then went through the unique categories and in doing so found some restaurants, classrooms and None values records. Rows with these were not added in a new data frame called the clean\_df. This data frame was grouped by the categories and the frequency of each category was obtained.

A bar plot and a choropleth were made to visualize this. With this information, the top 5 LGAs with the most venues can be recommended to restaurant owners.

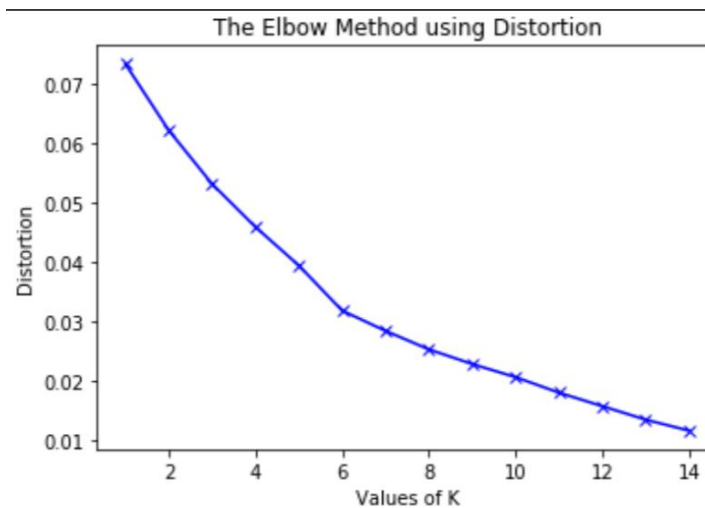




I then went on to obtain clusters of the different local government areas. The aim of the clustering of LGAs is to suggest groups of neighborhoods that a restaurant interested in delivery can focus on as a start-up.

To do this, I one hot encoded the categories feature in the clean\_df and grouped by the local government area then calculated the mean of the one hot encoded features.

I then proceeded to use the Elbow method to find the best number of clusters. The resulting plot can be seen below.



Based on this, the option of 6 clusters was chosen. The kmeans clustering algorithm was used. A kmeans instance was created and fit on the data. The Cluster labels were obtained and added to the data frame.

```
0]: lagos_grouped.head()
```

```
t[90]:
```

clinging cility	Rental Car Location	Residential Building (Apartment / Condo)	Salon / Barbershop	School	Shipping Store	Spa	Student Center	Tech Startup	Tourist Information Center	Toy / Game Store	Trade School	Train Station	Travel Agency	Travel Lounge	University	Video Game Store	Warehouse	Cluster Labels
0.0	0.0	0.0	0.000000	0.044444	0.000000	0.004444	0.0	0.044444	0.008889	0.0	0.000000	0.0	0.0	0.004444	0.000000	0.0	0.000000	1
0.0	0.0	0.0	0.005650	0.056497	0.000000	0.000000	0.0	0.028249	0.000000	0.0	0.005650	0.0	0.0	0.005650	0.000000	0.0	0.000000	0
0.0	0.0	0.0	0.000000	0.018349	0.004587	0.000000	0.0	0.041284	0.004587	0.0	0.000000	0.0	0.0	0.004587	0.000000	0.0	0.000000	1
0.0	0.0	0.0	0.005102	0.040816	0.000000	0.000000	0.0	0.025510	0.000000	0.0	0.005102	0.0	0.0	0.005102	0.000000	0.0	0.000000	0
0.0	0.0	0.0	0.000000	0.051724	0.000000	0.000000	0.0	0.017241	0.000000	0.0	0.000000	0.0	0.0	0.005747	0.005747	0.0	0.005747	0

## RESULTS

From the analysis of the number of offices in the different LGAs, the following LGAs were the top 5:

- 1.Eti-Osa
- 2.Ikeja
- 3.Mushin
- 4.Agege
- 5.Ikorodu

As a result of clustering the Local Government Areas, the following images show the top 2 clusters when the number of offices are considered.

---

The Local Government Areas in cluster 0 are:

```
Out[55]: 3      Amuwo Odofin
          12      Kosofe
          15      Mushin
          16      Ojo
          17      Oshodi/Isolo
          18      Shomolu
          19      Surulere
          Name: LGA, dtype: object
```

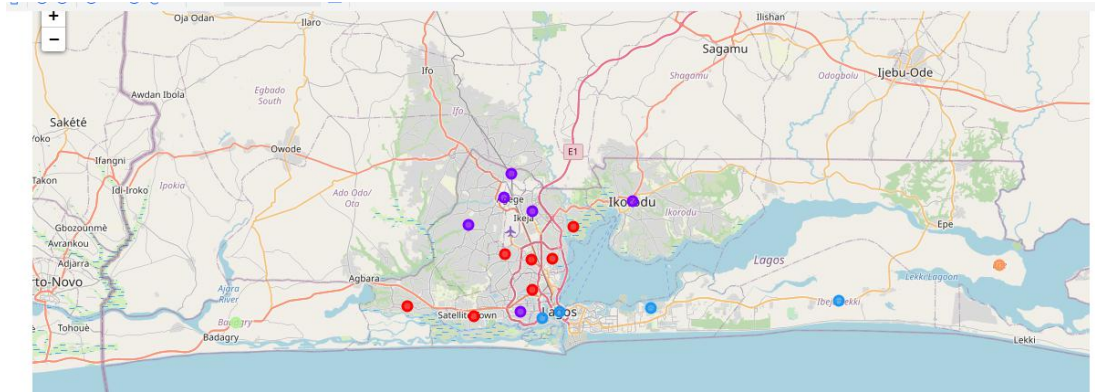
---

---

The Local Government Areas in cluster 1 are:

```
Out[56]: 0      Agege
          1      Ajeromi/Ifelodun
          2      Alimosho
          9      Ifako/Ijaye
          10     Ikeja
          11     Ikorodu
          Name: LGA, dtype: object
```

A folium Map was generated showing the different clusters. It was seen that clusters 0 and 1 with red and purple pointers were predominant.



## DISCUSSION

With Lagos being the financial capital of Nigeria and having the highest population density in the country, it is no surprise that a lot of restaurants are interested in having branches here.

From the results obtained, I recommend the Local Government Areas: **Eti-Osa, Ikeja, Mushin, Agege and Ikorodu** as locations in Lagos Nigeria where a restaurant can situate their business.

For Entrepreneurs interested in food delivery to limited range of locations, I recommend starting with the neighborhoods: **Amuwo Odofin, Apapa, Kosofe, Mushin, Oshodi/Isolo Ojo, Shomolu and Surulere OR Agege, Ajeromi/Ifelodun, Alimosho, Ifako/Ijaye, Ikeja and Ikorodu.**

Keeping in mind that I limited this analysis to 11 search queries, its usefulness can be improved with a wider range of queries. This path can also be taken for Entrepreneurs and Organizations with other interests with some modifications.

This analysis can also be tweaked to take into consideration the level of competition in the different areas.

## **CONCLUSION**

Restaurant owners with the results of this analysis for their more specific situations can situate their businesses in areas where they are most likely to get customers.

A similar application in a different sectors will facilitate the growth of the economy and the country as a whole.