Geolocation Data Analysis of Lagos, Nigeria.

Finding the best location to cite a Tech Start-up office in Lagos…

Lawrence Alaso Krukrubo
December 2019

1. Introduction

1.1 Background

I've worked in Lagos off and on, the longest stretch was three years from 2006 to 2009. No doubts, Lagos is an amazing place. I call it The real School-of-Hard-Knocks… A place where competition is fiercest and survival, for the fittest.
This article explores Lagos state, which is the 4th wealthiest city in Africa, fourth place behind Johannesburg, Cairo and Cape Town (*link*).
Lagos is home to over $(6,800 millionaires, 370 multi-millionaires and 4 billionaires).
It is also the commercial capital of Nigeria (The Most populous black nation), and the commercial hub of West Africa.
I have only recently returned to Lagos city, got enrolled and graduated from The *Founder Institute* accelerator program and currently building an InsureTech Start-up, with a Co-founder.
Therefore I want to use this opportunity to help future Start-ups find the most ideal location in Lagos city, for an office.

1.2 Problem

The problem we're going to solve using Geolocation data analysis and Machine Learning is helping a new Tech Start-Up find the ideal location for its office in The City of Lagos, Nigeria.
Amongst other attributes, the ideal office location for a Start-Up should have:

1. Proximity to Tech-Hubs and talents.
2. High feet traffic for easy interaction with potential customers.
3. Proximity to educational institutions for research and development.
4. Nearness to cafes and restaurants for business meetings or lunch-meets.
5. Nearness to bus-stops, seaports, and airports.
6. Security and safety.
7. A cluster of economic activities and complementing businesses.

1.3 Interest

Definitely, Start-ups and companies across the world that want a foot in Africa, would be interested in this analysis, since Nigeria is the commercial heartbeat of West Africa and Lagos is the commercial nerve centre of Nigeria. Besides research organisations and prospective business ventures would find this analysis really useful.

2. Data Description

The Dataset is the *Wikipedia page* of Lagos state, see the *link*.
We shall explore Lagos city through its respective Local Government Areas (LGA) or Boroughs.
The above link is a web page that shows the respective LGAs in Lagos State and each population figure.
This data will be analyzed through the following steps:-

1. We shall scrape the web page using the beautiful soup library.
2. We shall use The Foursquare API calls to retrieve geolocation data.
3. We shall fetch the text data using the requests library.
4. We shall convert it from JSON to Pandas DataFrame using the json_normalize module.
5. We shall use The folium library to render the maps and plot these via The Matplotlib library.
6. We shall cluster the venue categories per LGA using the Kmeans algorithm.
7. Then we'd explore respective LGAs to find the top LGA for a Startup to cite an office.
8. Finally, to add some fun, we shall use the word-cloud library to display the names of the top categories of venues in Lagos.

3. Methodology

First, let's import the required libraries.

```
from bs4 import BeautifulSoup
import requests  # library to handle requests.
import pandas as pd
import json  # library to handle JSON files.
from pandas.io.json import json_normalize  # transform json files to
pandas dataframes.
from geopy.geocoders import Nominatim  # convert an address into
latitude and longitude values.


# Matplotlib and associated plotting modules.
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors


# import k-means for clustering stage.
from sklearn.cluster import KMeans
#!conda install -c conda-forge folium=0.5.0 --yes
import folium # map rendering library
import numpy as np
import csv
print('All modules imported')
```

B.1 Scraping The Web Page Data:

First, we'd save the weblink for The Lagos data, then get the source code HTML data from the Website and use The Beautifulsoup module to parse it. The next step is to get the table that contains the data we want to scrape from. Then finally, we'd iterate through each link of the table and append the text parts.

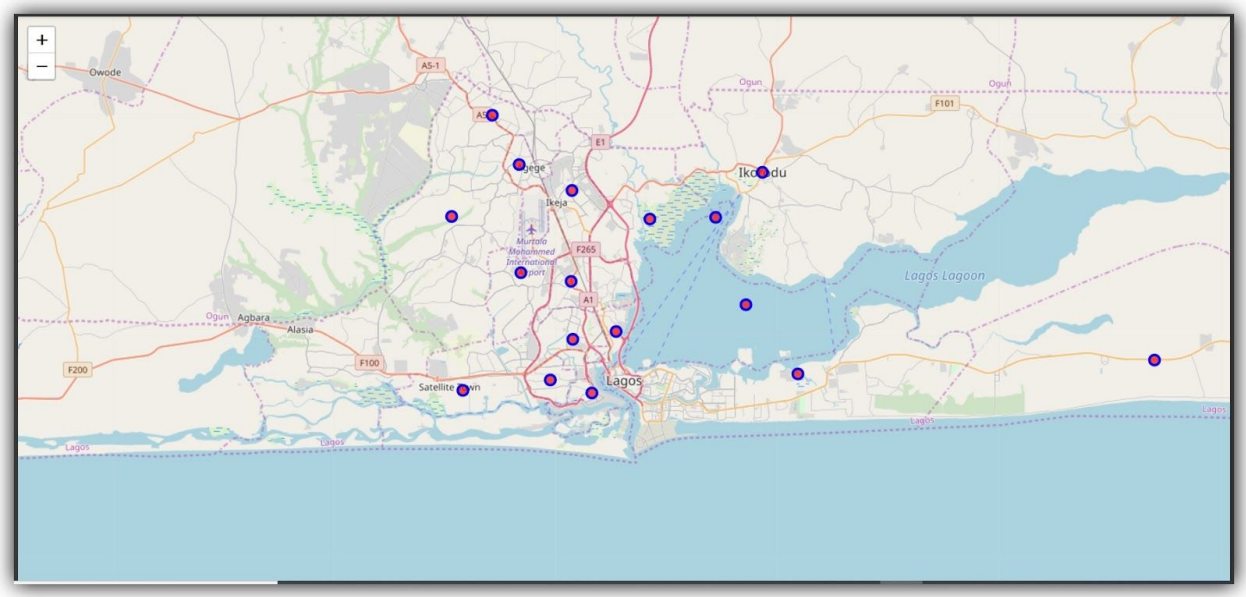
B.2 Creating The Lagos DataFrame:

Let's create a dictionary and append the LGA and corresponding Population data to it. Then we would attach a column of longitude and latitude by using the geopy.nominatim library to get each LGA latitude and longitude data. See image below.

```
[ ]  1 lagos_df.head()
```

	LGA	POP	latitude	longitude
0	Alimosho	5700714	6.58434	3.257631
1	Ajeromi-Ifelodun	1746634	6.45512	3.335946
2	Kosofe	665998	6.58197	3.414836
3	Mushin, Lagos	633543	6.53263	3.352022
4	Oshodi-Isolo	621789	6.54001	3.312415

B.3 Visualizing Lagos State LGAs:



Above is a map of Lagos State with LGAs superimposed, let's render it with the folium library. See codes below:-

```
address = 'Lagos'
geolocator = Nominatim(user_agent="NG_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude


# Render map using the Folium Library.
map_lagos_state = folium.Map(location=[latitude, longitude],
zoom_start=10)

# add markers to map.

for lat, lng, LGA  in zip(lagos_df['latitude'],
lagos_df['longitude'], lagos_df['LGA']):
    label = '{}'.format(LGA)
    label = folium.Popup(label, parse_html=True)
folium.CircleMarker([lat, lng],radius=6,popup=label,color='blue',
fill=True,fill_color='red',fill_opacity=0.7,parse_html=False).add_to(
map_lagos_state)

map_lagos_state
```

B.4 Using The Foursquare API Calls:

We shall use the *Foursquare* API calls to retrieve geolocation info about the venues in each LGA. First, we sign up at *https://developer.foursquare.com/* and create a free account complete with *CLIENT_ID* and *CLIENT_SECRET* parameters for making valid API calls. Next, we'd make a copy of the original lagos_df DataFrame, but this time we will set its index to be the names of The LGAs to make indexing easier.

Next thing we need to do is create a method called ***return_venues***, that takes each LGA address, converts it to a latitude and longitude list using the *Nominatim* module and then we use the *Foursquare* API calls to retrieve the top 200 categories of venues available in a 10 KM radius around each LGA.
We repeat for each LGA and save the unique list of venue categories to a variable, using a simple method called ***store_venue_categories,*** defined below.
Next, we pass the copy_lagos_df DataFrame as a parameter to the method above and assign the result to a variable that stores all existing venue categories in Lagos State.

```
all_venue_categories = store_venue_categories(copy_lagos_df)

# Let's see howmany venues we have altogether
print(len(all_venue_categories))
>>
101
# so we have 101 different categories of venues altogether.
```

B.4 Clustering Lagos State Venues Using Kmeans algorithm:

Let's create about 5 Clusters from Lagos State and choose the most viable cluster in terms of available venue categories. Then out of this cluster, we shall select the top LGA based on certain parameters we would define.

First, we create a method called *getNearbyVenues,* that returns a DataFrame consisting of every venue detail in each LGA, all attached together. We pass this method the list of each LGA name and corresponding latitude and longitude data as parameters and save the output Dataframe in a variable called *lagos_state_venues*.

Next, we one-hot encode the venue_category column of the above DataFrame, and then we group by LGA, taking the mean frequency occurrence of each venue category per LGA. With this grouped DataFrame showing the mean occurrence of each venue category per LGA, we would create our clustering algorithm.

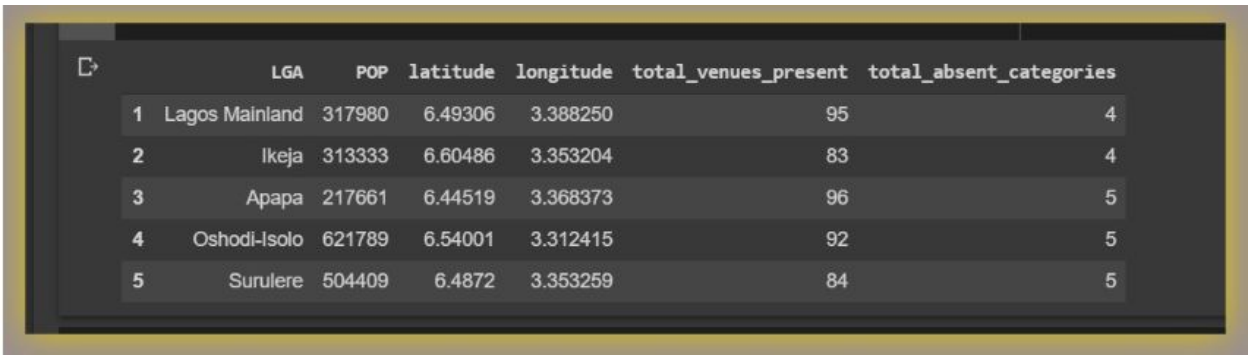
```
# set number of clusters

kclusters = 5

lagos_grouped_clustering = lagos_grouped.drop('LGA', 1)

# run k-means clustering

kmeans = KMeans(n_clusters=kclusters,
random_state=0).fit(lagos_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_
```

The various clusters of Lagos venues. Clearly the cluster of red circles is the dominant or ideal cluster.

4. Results

From the result of our analysis, Lagos Mainland LGA has the least number of absent top categories of venues (4), while having the highest number of available venues(95). Next to it is Ikeja LGA, which has the same number of absent top categories of venues(4).

See table below showing the Top 5 LGAs to cite an office in Lagos:-



	LGA	POP	latitude	longitude	total_venues_present	total_absent_categories
1	Lagos Mainland	317980	6.49306	3.388250	95	4
2	Ikeja	313333	6.60486	3.353204	83	4
3	Apapa	217661	6.44519	3.368373	96	5
4	Oshodi-Isolo	621789	6.54001	3.312415	92	5
5	Surulere	504409	6.4872	3.353259	84	5

4. Discussion:

All models are wrong, but some are useful… George Edward Pelham

Taking a cue from *George Pelham*, this is by no means an exhaustive or unerring analysis of Lagos State. We simply used the data from the *Wikipedia page of Lagos*. We scraped this data and applied Machine Learning and Exploratory Data Analysis(EDA) in line with certain parameters we created, to arrive at a plausible result concerning the most ideal location to cite an office in Lagos, which is The *Lagos Mainland LGA*.

D.1 Lagos Mainland LGA:

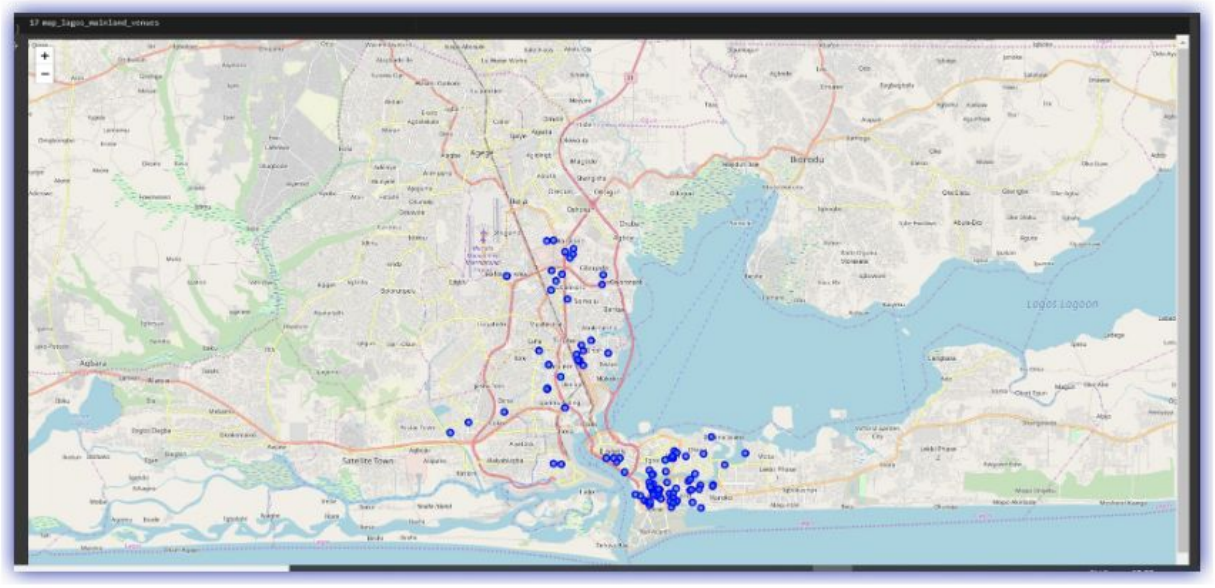
Lagos Mainland was founded by *Chief Olofin* and is peopled by the same group as found in Lagos especially the Egbas and Aworis.

Lagos Mainland developed from settlements such as Ebute-Metta, Ido-Otto, and Ijora. See *link*

The communities of Lagos Mainland include:

Yaba, Ebute-Metta, Iddo-Otto, Iwaya, Akoka, Makoko, Abule-nla.

Yaba is a beehive of Tech activities, with the presence of Tech partners and co-working services such as *Co-creation hub*, *NG Hub from Facebook*, *Hub One*, *Yaba ICT Hub*, *Pancake Hub*, *LeadSpace* , *Mindthegap Incubation Hub*, and a bunch of others.



Lagos Mainland LGA, showing some venues.

5. Conclusion:

In this article, we've explored a bit of Lagos City and we've helped any Start-up looking to relocate or expand to Lagos to find the ideal locations for its offices.

The entire analysis backing this article is available in my *GitHub repo*.

Finally, let's see a word cloud object showing the most common venue categories all across Lagos State.

Ladies and Gentlemen, the most common venues in Lagos City are… You guessed right!…

Hotels and Bars!!

Cheers…