

APPLIED DATA SCIENCE CAPSTONE PROJECT

FLATTENING THE CURVE AS ECONOMIC ACTIVITIES RESUME: LOCATION DATA ANALYSIS OF LAGOS STATE IN REFERENCE TO THE COVID 19 PANDEMIC



Figure1: Health Workers with a COVID-19 Patient. Source: Nairametrics.com

1.0 DESCRIPTION OF THE PROBLEM: INTRODUCTION

1.1 BACKGROUND

Coronavirus Disease 2019 (COVID-19) is an infectious disease caused by the novel coronavirus.¹ This disease which is currently a pandemic broke out in Wuhan, China in December 2019. As at 27 April 2020, there were 2,878,196 confirmed cases and 198,668 deaths recorded from the disease globally.² Due to its highly infectious nature, countries have had to take up strict social distancing measures such as imposing a lockdown. One of the adverse effects of these measures is the impact on the economy.

In Nigeria, the index case of COVID-19 was confirmed on 27 February 2020, when an Italian tested positive in Lagos two days after arrival from Milan, Italy.³ Based on the Nigerian Center for Disease Control (NCDC)'s COVID-19 data, as at 27 April 2020, the country confirmed 1,337 cases, 255 cases discharged, and 40 deaths recorded.⁴ Total active cases as at 27 April 2020, were 1,042.⁵

Lagos State, being the business centre of the country has been the most hit with 764 confirmed cases, 143 discharged and 19 deaths recorded. The active cases as at 27 April 2020 were 602, representing 57.77% of the country's total active cases.⁵

On the 29th of March 2020, the president of the Federal Republic of Nigeria announced a two-week lockdown of Lagos state, Ogun state and the Federal Capital Territory (FCT), Abuja.³ The lockdown was extended for another two weeks on 13 April 2020, and for another one week on 27 April 2020; however, the president announced that the lockdown will be relaxed after the one-week extension (04 May 2020) but with an indefinite curfew from 8 pm to 6 am.³



Figure2: The City of Lagos. Source: nomadaafricamag.com

1.2 THE PROBLEM

Lagos is a state in South-Western Nigeria. Although it is the smallest in area per km² among all 36 states in Nigeria, it currently has the highest population density.⁶ According to Ekundayo (2013), if Lagos state was a country, it would be the fifth-largest economy in Africa.⁷ Lagos, Nigeria is not the Federal Capital Territory, but it is popularly tagged as the Business/Economic Capital of Nigeria as it is home to various enterprises, investors and headquarters of multinationals/conglomerates. The search for employment and source of living is one of the reasons for its high population density, as there is a constant inflow of citizens from other states. It is obvious that a continuous or an extended lockdown of the state will largely impact the economy negatively. On the other hand, Lagos state is at high risk of a surge in the number of COVID-19 cases once the lockdown is eased due to its population and business activities. A surge in the number of cases could result in repeated lockdowns and an unpalatable cycle that can negatively affect public health and the economy.

The phrase “the new normal” has become popular in the social media space; companies and business owners have been advised to embrace it to adapt and survive the COVID-19 crisis.

The question is: How can Lagos State flatten the curve while economic activities take place?

This project will involve the use of location data of the twenty (20) LGAs in Lagos state and their COVID-19 distribution to identify venues that require strict focus in terms of public health measures to curb the spread. The idea is to curb the spread of COVID-19 in the most common venues in the LGAs as economic activities resume.

AIM: This project aims to develop strategies based on location data analysis, to reduce the spread of the COVID-19 virus in Lagos state as economic activities resume.

OBJECTIVES: The objectives of this project are as follows:

1. Identification of Lagos state Local Government Areas (LGAs) most affected by the COVID-19 virus.
2. Exploration of location data of LGAs with the highest COVID-19 cases using Foursquare API.
3. Segmenting and clustering of LGAs to identify defining features concerning COVID -19 spread by leveraging on Foursquare location data.
4. Development of strategies to reduce the spread of the virus based on the result of the analysis.

1.3 JUSTIFICATION

The COVID-19 pandemic has been established as a threat to both the economy of Lagos state and the health of its populace. A lot of people are scared of death from the virus or its negative impact on their income. As a Biostatistician and Public Health Data Scientist, I am worried about the possibility of a surge in the number of COVID-19 cases in the state when the lockdown is relaxed on 04 May 2020, as stipulated by the Federal Government. However, I also understand that the state cannot afford an economic shutdown as the economy already had challenges pre-COVID 19. An economic shutdown would not only affect the Lagos state economy but also Nigeria's economy at large. The need to devise strategies to flatten the curve without gross negative impact on the economy cannot be overemphasized. I believe without a doubt that this is the solution Lagos state needs at the moment.

1.4 TARGET AUDIENCE

The target audience for this report includes the following:

1. **The Lagos State Government:** This report would be beneficial for decision-makers in the areas of health and commerce in Lagos state. These decision-makers include the Governor of the state, Lagos State Ministry of Health, Lagos State Ministry of Commerce and Industry and Lagos state Ministry of Finance. The recommendations of this report will be useful to them as they strategize together to flatten the curve and keep the economy thriving.
2. **NCDC and Federal Government:** The NCDC and the Federal Government must work with the Lagos state government in strategizing and implementing solutions that would help flatten the curve and protect the economy.
3. **Chief Executives and Business Owners:** Decision makers of companies, conglomerates, multinationals, banks, entrepreneurs need to understand how they can protect their staff, customers and businesses. The analysis using location data would help to identify venues especially business locations that require strict social distancing measures. This report might help some businesses decide to adapt their business to work without a physical location. Ignoring the public health measures for the sake of profit is counterproductive as increased outbreak would only result in

repeated lockdowns. This category needs all the information they can get to balance it all, part of which they will get from this report.

4. **Lagos State Residents:** A COVID-19 LGA analysis based on location data will be beneficial to residents of the states as it can guide them on venues to avoid or venues where they have to take extra caution.
5. **Data Scientist and Public Health Scientists:** This report will be informative for upcoming data scientists, data scientists and public health professionals interested in learning how to solve problems with Data Science.

2.0 DATA ACQUISITION AND CLEANING

2.1 DATA

Based on the definition of the problem, the data needed for this project are:

1. **Lagos State Local Government Area (LGA) Data:** Lagos state is divided into twenty Local Government Areas (LGAs) which is synonymous to boroughs. Location analysis of Lagos state can only be carried out by sourcing data on the LGAs. This data contains the list of LGAs and their population.
2. **Distribution of COVID-19 cases across LGAs:** This contains the number of COVID-19 cases per LGA as at 18 April 2020. This is the available data on COVID-19 distribution across LGAs as at the time of this analysis. This data is needed to compare the number of COVID-19 cases across the LGAs. Location data of LGAs with the highest number of COVID-19 cases would also be explored.
3. **Geographical coordinates of the LGAs:** This data is needed to extract location data of all the LGAs needed for analysis.
4. **Location data of various venues in the 20 LGAs in Lagos state:** This data is needed to identify the top venues in the LGAs of Lagos state. This will help to pinpoint venues to strictly monitor as Lagos state economic activities resume.

2.2 DATA SOURCES

The data required was sourced from the following:

- "Local government areas" from Wikipedia: https://en.wikipedia.org/wiki/Lagos_State#Local_government_areas. This data provides the list of all the 20 LGAs and their population.
- Distribution of COVID-19 cases in the Lagos Metropolitan Area.jpg. https://commons.wikimedia.org/wiki/File:COVID-19_cases_in_the_Lagos_Metropolitan_Area.jpg. The number of cases per LGAs were manually extracted, as the file is in JPG format; this is also the only data available with traceable source on the number of cases per LGA as at the time of the analysis.
- Geopy: A Python 2 and 3 client for several popular geocoding web services. This will be used to get the geographical coordinates of the LGAs.

- Foursquare API: to get location data about the various venues in Lagos state by LGAs.

2.3 DATA ACQUISITION AND CLEANING

The goal of the data acquisition and cleaning is to have a data frame containing the LGAs, COVID-19 cases, Population, Latitude and Longitude values and location data.

This will be achieved through the following steps:

A. Scraping of Lagos state LGA data from Wikipedia

A table containing Lagos state LGAs and their population data was scraped from the Wikipedia page: https://en.wikipedia.org/wiki/Lagos_State#Local_government_areas. This was achieved using the Wikipedia API and data was read into a Pandas data frame.

```
import urllib.request
!conda install lxml --yes
%pip install Wikipedia
import wikipedia as wp
```

Figure 3: Code snippet showing libraries installed to scrape the data

```
url = "https://en.wikipedia.org/wiki/Lagos_State#Local_government_areas"
page = urllib.request.urlopen(url)
html = wp.page("Lagos_State#Local_government_areas").html().encode("UTF-8")
df = pd.read_html(html, header = 0)[2]
df
```

Figure 4: Code used to scrape data from Wikipedia

The data frame returned, consist of 3 columns (Rank, LGA and Population) and 21 rows containing values for these columns. The last row in the data frame represents the total population of Lagos state.

[2]:	Rank	LGA	Population
0	1.0	Alimosho	5700714
1	2.0	Ajeromi-Ifelodun	1746634
2	3.0	Kosofe	665421
3	4.0	Mushin	633543
4	5.0	Oshodi-Isolo	10621789

Figure 5: Snippet of Data frame scraped from Wikipedia

B. DATA CLEANING

The 'Rank' column was dropped from the data as it is not necessary for the analysis. The last row which contains the total population was also removed as it is also not necessary for the analysis and might negatively affect it if left in the data frame. The types of the data values were checked, and the 'Population' column was found to contain string type data. Population data was changed to type integer using the `astype()` function.

The distribution of COVID-19 cases across the LGAs was added to the data frame by manual extraction from an image file from Wikipedia by creating a column containing a list of the number of cases.

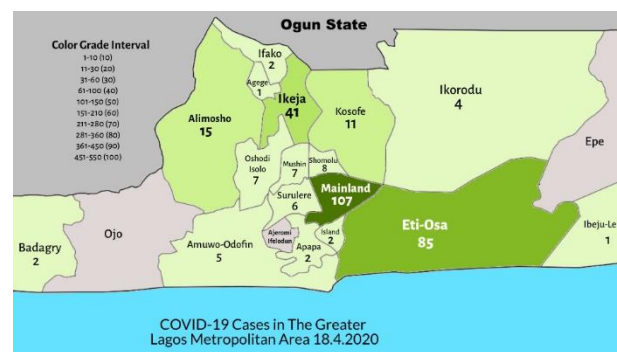


Figure 6: Map image from Wikipedia showing distribution of COVID-19 cases across LGAs

```
[8]: COVID19_Cases = [15, 0, 11, 7, 7, 0, 4, 6, 1, 2, 8, 5, 107, 41, 85, 2, 2, 2, 0, 1]
df['COVID19_Cases'] = COVID19_Cases
df
```

Figure 7: Creation of the COVID19_Cases column

After creating and adding the column for COVID-19 cases, the data frame now had 3 columns namely: 'LGA', 'Population', and 'COVID19_Cases'; and 20 rows containing their values.

```
[8]:
```

	LGA	Population	COVID19_Cases
0	Alimosho	5700714	15
1	Ajeromi-Ifelodun	1746634	0
2	Kosofe	665421	11
3	Mushin	633543	7
4	Oshodi-Isolo	10621789	7

Figure 8: Snippet of data frame showing LGA, population and COVID19 distribution

Extraction of Geographical Coordinates:

The Nominatim function from geopy was used to extract the geographical coordinates of the LGAs. To achieve correct extraction, an 'Address' column was first created specifying that the LGAs are in Lagos. Thereafter, the code to extract the latitude and longitude values of each LGA and merge them with the current data frame was run. The data frame returned 9 columns (LGA, Population, COVID19_Cases, Address, location, point, latitude, longitude and altitude).

	LGA	Population	COVID19_Cases	Address
0	Alimosho	5700714	15	Alimosho,Lagos
1	Ajeromi-Ifelodun	1746634	0	Ajeromi-Ifelodun,Lagos
2	Kosofe	665421	11	Kosofe,Lagos
3	Mushin	633543	7	Mushin,Lagos
4	Oshodi-Isolo	10621789	7	Oshodi-Isolo,Lagos

Figure 9: Snippet of data frame with 'Address' column

```
from geopy.extra.rate_limiter import RateLimiter
geolocator = Nominatim(user_agent="lag_explorer")

# 1 - convenient function to delay between geocoding calls
geocode = RateLimiter(geolocator.geocode, min_delay_seconds=1)
# 2- - create location column
df['location'] = df['Address'].apply(geocode)
# 3 - create longitude, latitude and altitude from location column (returns tuple)
df['point'] = df['location'].apply(lambda loc: tuple(loc.point) if loc else None)
# 4 - split point column into latitude, longitude and altitude columns
df[['latitude', 'longitude', 'altitude']] = pd.DataFrame(df['point'].tolist(), index=df.index)
df
```

Figure 10: Code used to extract geographical coordinates

	LGA	Population	COVID19_Cases	Address	location	point	latitude	longitude	altitude
0	Alimosho	5700714	15	Alimosho,Lagos	(Alimosho, Lagos, Nigeria, (6.58434295, 3.257631097232365, 0.0))	(6.58434295, 3.257631097232365, 0.0)	6.584343	3.257631	0.0
1	Ajeromi-Ifelodun	1746634	0	Ajeromi-Ifelodun,Lagos	(Ajeromi/Ifelodun, Lagos, Nigeria, (6.45512169, 6.455121699999999, 3.3359462518298724, 0.0))	(6.455121699999999, 3.3359462518298724, 0.0)	6.455122	3.335946	0.0
2	Kosofe	665421	11	Kosofe,Lagos	(Kosofe, Lagos, 100234, Nigeria, (6.581974000000001, 6.581974000000001, 3.4148359416178415, 0.0))	(6.581974000000001, 3.4148359416178415, 0.0)	6.581974	3.414836	0.0
3	Mushin	633543	7	Mushin,Lagos	(Mushin, Lagos, Nigeria, (6.5326324499999995, 6.5326324499999995, 3.3520219717987674, 0.0))	(6.5326324499999995, 3.3520219717987674, 0.0)	6.532632	3.352022	0.0
4	Oshodi-Isolo	10621789	7	Oshodi-Isolo,Lagos	(Oshodi/Isolo, Lagos, 100263, Nigeria, (6.5400100000000005, 6.5400100000000005, 3.312414613183944, 0.0))	(6.5400100000000005, 3.312414613183944, 0.0)	6.540010	3.312415	0.0

Figure 11: Snippet of data frame returned by geopy containing geographical coordinates

The columns for address, location and point were dropped as they are repeated variants of the LGA, and latitude-longitude values. The altitude column was also dropped as it is not necessary for the analysis.

```
df.drop(['Address', 'location', 'point', 'altitude'], axis=1, inplace = True)
df
```

	LGA	Population	COVID19_Cases	latitude	longitude
0	Alimosho	5700714	15	6.584343	3.257631
1	Ajeromi-Ifelelodon	1746634	0	6.455122	3.335946
2	Kosofe	665421	11	6.581974	3.414836
3	Mushin	633543	7	6.532632	3.352022
4	Oshodi-Isolo	10621789	7	6.540010	3.312415

Figure 12: New data frame containing geographical coordinates

Foursquare data:

Using the geographical coordinates of the LGAs, location data was extracted using the Foursquare API. The goal was to use the Foursquare API to extract information about the top ten most common venues in the LGAs. The top 3 LGAs based on the number of COVID-19 cases was first explored and data on their top venues and categories were extracted. The next step was to extract all the venues in the LGAs along with their categories and geographical coordinates.

	name	categories	lat	lng
0	Lounge 17	BBQ Joint	6.488902	3.384210
1	Home Zone Bar	Lounge	6.488129	3.384850
2	Ground Zero	Plaza	6.487224	3.384767
3	Debra's Grace	Boutique	6.495049	3.380867

Figure 13: Top venues in Lagos Mainland with the highest number of COVID-19 cases as at 18/04/2020

	name	categories	lat	lng
0	Prince Ebeano	Department Store	6.454818	3.530238
1	Palazzo Dumont	Hotel	6.456057	3.529894
2	Limeridge Hotel Gym	Gym / Fitness Center	6.455751	3.529988
3	Limeridge Hotel	Hotel	6.454438	3.529811

Figure 14: Top venues in Eti-Osa LGA with the 2nd highest number of COVID-19 cases as at 18/04/2020


```
]:
```

	name	categories	lat	lng
0	University of Suya	BBQ Joint	6.606512	3.349370
1	Best Western Plus Lagos Ikeja Hotel	Hotel	6.597282	3.354239
2	Bakers' World	Bakery	6.600560	3.355194
3	Goat Hunters	African Restaurant	6.601134	3.351368
4	K.F.C.	Fast Food Restaurant	6.600272	3.355484

Figure 15: Top venues in Ikeja LGA with the 3rd highest number of COVID-19 cases as at 18/04/2020

```
i]: print(Lagos_venues.shape)
Lagos_venues.head()

(105, 7)

i]:
```

	LGA	LGA Latitude	LGA Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kosofe	6.581974	3.414836	Cecilia's SteakHouse	6.583084	3.418033	Steakhouse
1	Mushin	6.532632	3.352022	Mushin market	6.527949	3.353511	Flea Market
2	Mushin	6.532632	3.352022	olorunsogo bus stop	6.536227	3.352369	Bus Station
3	Mushin	6.532632	3.352022	KITS Technologies	6.534660	3.346489	IT Services
4	Mushin	6.532632	3.352022	CARWASH	6.539191	3.356523	Rental Car Location

Now that we have data on the venues in the LGAs, let's check the number of venues returned by each of them.

Figure 16: Snippet of venues in Lagos state returned by Foursquare

```
122]:
```

	LGA	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	Agege	Campground	Fast Food Restaurant	Market	Bus Station	Gym	Clothing Store	Fried Chicken Joint	Food	Flea Market	Farm
1	Amuwo-Odofin	Hotel	Pub	Department Store	Steakhouse	Campground	Food	Flea Market	Fast Food Restaurant	Farm	Electronics Store
2	Apapa	Shopping Mall	Fast Food Restaurant	Plaza	Playground	Department Store	Pharmacy	Hotel	Boat or Ferry	Indian Restaurant	Chinese Restaurant
3	Badagry	Hotel	Steakhouse	Chinese Restaurant	Fried Chicken Joint	Food	Flea Market	Fast Food Restaurant	Farm	Electronics Store	Department Store
4	Eti-Osa	Hotel	Gym / Fitness Center	Department Store	Chinese Restaurant	Fried Chicken Joint	Food	Flea Market	Fast Food Restaurant	Farm	Electronics Store

Figure 18: Snippet of Dataframe with Top 10 most common venues in each LGA

The Foursquare data on the most common venues in the LGA would be used to identify the LGAs to strictly monitor as economic activities in Lagos state reopens. This data would also be clustered to identify defining features that can be used to develop strategies based on the location data. The information on the most common venues would be the basis for identifying strategies that can be used to curb the spread of COVID19, these strategies would be highlighted in the 'Recommendations' section of this report. Codes used to generate the Foursquare data can be found in the Methodology section.

REFERENCES

1. World Health Organisation (WHO). 2020. Q & A on coronaviruses (COVID-19). Available at: <https://www.who.int/news-room/q-a-detail/q-a-coronaviruses>.
2. World Health Organisation (WHO). 2020. Coronavirus disease 2019 (COVID-19) Situation Report-98. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200427-sitrep-98-concl-19.pdf?sfvrsn=90323472_4.
3. Wikipedia. 2020. 2020 coronavirus pandemic in Nigeria. Available at: https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Nigeria.
4. Nigeria Centre for Disease and Control (NCDC). 2020. CORONAVIRUS (COVID-19) HIGHLIGHTS. Available at: <https://covid19.ncdc.gov.ng/>.
5. Nigeria Centre for Disease and Control (NCDC). 2020. COVID-19 SITUATION REPORT. Available at: <https://ncdc.gov.ng/themes/common/files/sitreps/d127b988882b6c7fb93335aa6df3bacb.pdf>
6. Wikipedia. 2020. Lagos State. Available at: https://en.wikipedia.org/wiki/Lagos_State
7. Ekundayo, J.M. 2013. *Out of Africa: Fashola: Reinventing Servant Leadership to Engender Nigeria's Transformation*. AuthorHouse. P.135. ISBN9781481790406