Predicting the Best Neighborhood for Opening a Shopping Mall in Lagos, Nigeria

Ayodeji Jackson Folusho

May 21, 2020

A Peer Graded Assignment for Applied Data Science (CAPSTONE)

Introduction

1.1 Background

Given the peculiarity of Lagos being the largest commercial city, in Nigeria and West Africa, a lot of investors show keen interest in the entertainment industry especially in shopping malls to address needs of the growing middle class and the rapid urbanization of Lagos.

Location is a critical factor in determining the failure or success of such investment and we will attempt to provide guidance to investors as well as property developers. This will be carried out using data science methodology such as Machine Learning, Clustering etc.

The target audience for this report would be investors, property developers and Government bodies looking to invest.

1.2 Problem

The project aims to solve the problem of identifying the best neighborhoods in Lagos for set up of a shopping mall. Best Neighborhood implies a suburb or cluster of suburbs that would yield the best return and profitability for investors through repeated patronage of services provided by the mall

1.3 Interest

Interests are property developers and investors looking to make investments in Lagos targeted at the entertainment / shopping sector

2. Data acquisition and cleaning

2.1 Data Methodology

1. Obtain data by building a dataframe of neighborhoods in Lagos, Nigeria by web scraping the data from Wikipedia page,

2. Get the geographical coordinates of the neighborhoods and obtain venue data for the neighborhoods from Foursquare API.

3. Explore and Cluster the neighborhoods

4. Select the best cluster to open a new shopping Mall

2.2. Data Sources

A web based source , Wikipedia https://en.wikipedia.org/wiki/Category:Local_Government_Areas_in_Lagos_State which contained a list of Local Government Areas also known as Neighborhoods was used for Web scraping and invariably extracting the required data using Python requests, beautifulsoup packages etc. Foursquare API was used to get the venue data while geographical coordinates were extracted using the Python geocoder package (Latitude and Longitude)

Our Area of interest is the Shopping mall category .

Other data science techniques used include but are not limited to

- Data Cleaning
- Data Wrangling
- K-means clustering (Machine Learning)
- Map Visualization (Folium)

3. Data Analysis

As earlier stated the first step was to get the data from the Wikipedia page by web scraping using Python requests and beautifulsoup packages to extract the list of neighborhoods data.

The list of name wouldn't suffice for the scope of this project and it was therefore important to get the geographical coordinates using the geocoder package. This also enabled us to use the Foursquare API in gathering venue information.

The data was populated into the pandas df (dataframe) which was visualized using the Folium package (Map)

The Foursquare API was a very useful tool as it enabled us to get the top 100 venues that are within a radius of 2000 meters using my Foursquare ID and Secret Key obtained by registering an account on Foursquare Developer. When API calls to Foursquare in a python loop, it returns the venue date as a *json* file format which enabled me to extract the venue name, venue category, venue longitude and latitude

Another step involves grouping the rows taking into account mean frequency of occurrence. This is a first step in preparing the data for use in clustering. Data was also filtered using "Shopping Mall" which is our focal category for this project

Clustering is then carried out using k-means clustering. The Neighborhoods were clustered into 2 based on their frequency of occurrence using the "shopping mall" filter

4. Results

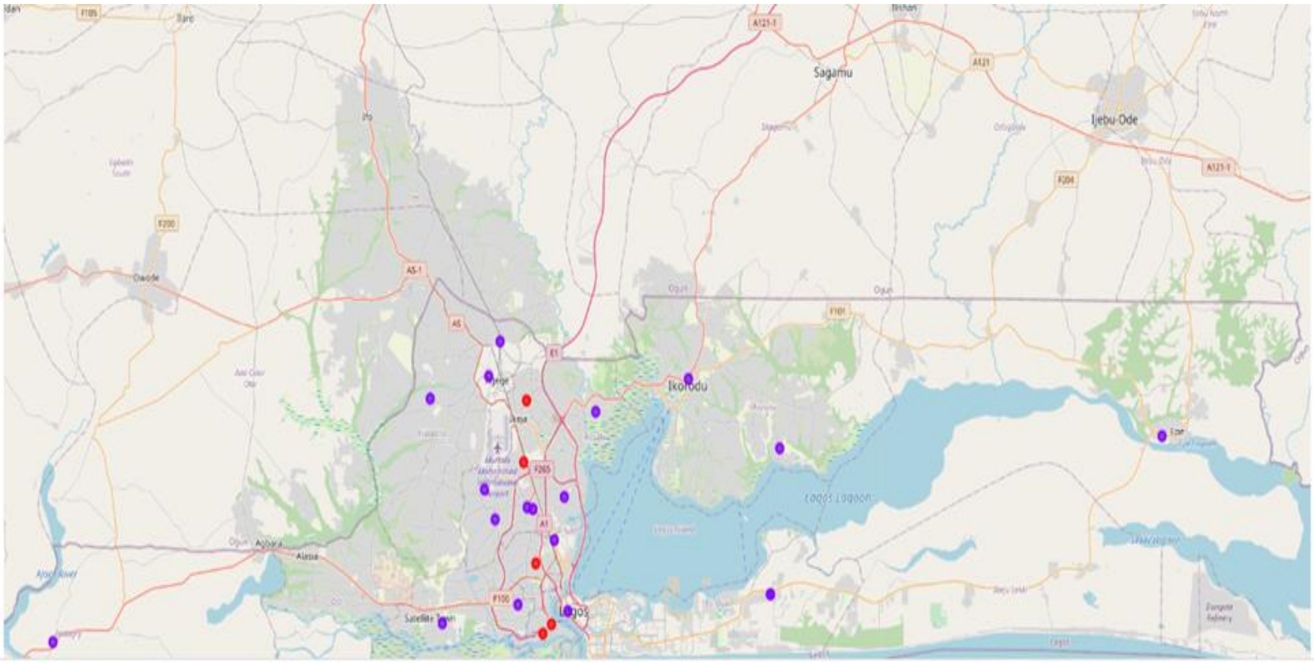
The results from the k-means clustering show that we can categorize the neighborhoods into 2

Clusters based on the frequency of occurrence for "Shopping Mall":

- **Cluster 0:** Neighborhoods with moderate number of shopping malls

- **Cluster 1:** Neighborhoods with low number to no existence of shopping malls

The results of the clustering are visualized in the map below with cluster 0 in red color, cluster 1 in purple color,



5. Discussions

Most of the shopping malls are concentrated in Neighborhoods around cluster 0 with little concentration in Cluster 1. Cluster 1 represents a good opportunity for set up of a shopping mall due to little or no competition. property Developers and investors could have a quick return on investment by investing in this area. This in itself would however require further strategic studies to understand why shopping malls are not well situated in the cluster but from a stand point of data science. Cluster 1 presents a very good opportunity

6. References

https://en.wikipedia.org/wiki/Category:Local_Government_Areas_in_Lagos_State

Report: https://cocl.us/coursera_capstone_report

Notebook: https://cocl.us/coursera_capstone_notebook

Presentation: https://cocl.us/coursera_capstone_presentation

7. Appendix

A. Number of Venues

Number of venues returned from each Neighborhood

```
In [120]: venues_df.groupby(["Neighborhood"]).count()
```

Out[120]:

Neighborhood	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
Agege	9	9	9	9	9	9
Ajeromi-Ifelodun	4	4	4	4	4	4
Alimosho	4	4	4	4	4	4
Amuwo-Odofin	5	5	5	5	5	5
Apapa	6	6	6	6	6	6
Badagry	3	3	3	3	3	3
Ejigbo, Lagos	4	4	4	4	4	4
Epe, Lagos	1	1	1	1	1	1
Eti-Osa	8	8	8	8	8	8
Eti-Osa East	8	8	8	8	8	8
Ifako-Ijaye	5	5	5	5	5	5
Ijede	1	1	1	1	1	1
Ikeja	39	39	39	39	39	39
Ikorodu	8	8	8	8	8	8
Kosofe	4	4	4	4	4	4
Lagos Island	22	22	22	22	22	22
Lagos Mainland	24	24	24	24	24	24
List of Lagos State local government areas by population	21	21	21	21	21	21
Mushin, Lagos	11	11	11	11	11	11
Ojo, Lagos	8	8	8	8	8	8
Ojuwoye, Mushin	10	10	10	10	10	10
Oshodi-Isolo	5	5	5	5	5	5
Somolu	11	11	11	11	11	11
Surulere	22	22	22	22	22	22

B. Cluster 0

Cluster Examination

Cluster 0

```
In [140]: lg_merged.loc[lg_merged['Cluster Labels'] == 0]
```

Out[140]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
23	Surulere	0.136364	0	6.48932	3.358000
4	Apapa	0.166667	0	6.43795	3.364360
18	Mushin, Lagos	0.090909	0	6.44498	3.372547
17	List of Lagos State local government areas by ...	0.095238	0	6.56296	3.346040
12	Ikeja	0.076923	0	6.60776	3.348540

C. Cluster 1

Cluster 1

```
: lg_merged.loc[lg_merged['Cluster Labels'] == 1]
```

41]:

	Neighborhood	Shopping Mall	Cluster Labels	Latitude	Longitude
21	Oshodi-Isolo	0.000000	1	6.521350	3.318630
20	Ojuwoye, Mushin	0.000000	1	6.528992	3.354942
19	Ojo, Lagos	0.000000	1	6.530016	3.349559
16	Lagos Mainland	0.041667	1	6.506430	3.375530
15	Lagos Island	0.045455	1	6.454700	3.388760
14	Kosofe	0.000000	1	6.599990	3.415090
13	Ikorodu	0.000000	1	6.623560	3.504830
0	Agege	0.000000	1	6.625610	3.312620
10	Ifako-Ijaiye	0.000000	1	6.651110	3.323290
9	Eti-Osa East	0.000000	1	6.466680	3.583260
8	Eti-Osa	0.000000	1	6.466680	3.583260
7	Epe, Lagos	0.000000	1	6.582122	3.960848
6	Ejigbo, Lagos	0.000000	1	6.542900	3.308255
5	Badagry	0.000000	1	6.432160	2.892650
3	Amuwo-Odofin	0.000000	1	6.445430	3.267540
2	Alimosho	0.000000	1	6.609270	3.255800
1	Ajeromi-Ifelodun	0.000000	1	6.459410	3.340550
22	Somolu	0.000000	1	6.537850	3.385340
11	Ijede	0.000000	1	6.573000	3.592500