IBM | Coursera Capstone Project

The Battle of the Neighborhoods

Prosper Popoola

March 26, 2020

1. Introduction

1.1 Background

Lagos is a big city with big potentials and opportunities, but most of these opportunities are not fully tapped, and this may be largely due to lack of data to inform profitable business decisions, and data based decision making on which type of business is good for a certain area in Lagos, but not good for other areas. Some business decisions have been made based on sentiments and/or not so well thought out advises. This project is aimed at using available metadata and Foursquare API data to make informed decisions on possible thriving businesses in certain part of Lagos. Since Lagos is heavily populated, and clustered to certain areas, this project will focus on Alimosho Local Government Area, which is arguably the biggest and most populated LGA in Lagos, Nigeria. At the end of this project, we would have been able to establish which businesses are more common in Alimosho LGA, and which are less common. Then we can arrive at which business is important but less common, which may inform our business decision to break a new ground, or compete with existing thriving businesses.

1.2 Problem

The data will show and rank the existing businesses in our choice area, using independent data from Foursquare. We will rank the top ten

businesses in the choice area and probably for as far as ranking the least ten businesses.

1.3 Interest

Any business conscious person should be interested in the output of this project. Most especially business persons already operating or intending to open a business in our choice area. The output of the project could help in making informed decisions about the need for expansion or change of business type for profitability.

2. Data Acquisition and Cleaning

2.1 Data Sources

The biggest challenge to this project is the unavailability of data to work it. It took a better part of two days to get the required data together from different sources. The list of Lagos LGAs are readily available on the internet, but the list of LGAs with their respective wards or neighborhoods is the real gold mine. You will never find the list of Lagos LGAs and their coordinates anywhere on the internet put together in one place.

The data was without the coordinates, and with inconsistent delimiters and labeling.

2.2 Data Cleaning

Data was downloaded and scraped from wikipedia, which came with more than expected actually. The better part of the project resources were spent on cleaning and wrangling. The table came with wrong columns labels, so that had to be changed.

Out[3]:

	LGA	Wards
0	Agege	Isale/Idimangoro; Iloro/Onipetesi; Oniwaya/Pap...
1	Ajeromi/Ifelodun	Ago Hausa; Awodi-Ora; Wilmer; Olodi; Tolu; Tem...
2	Alimosho	Shasha/Akowonjo; Egbeda/Alimosho; Idimu/Isheri...
3	Amuwo-Odofin	Amuwo-Odofin Housing Estate, Mile 2; Festac 1;...
4	Apapa	Apapa I (Marine Rd. and environs); Apapa II (L...
5	Badagry	Posukoh; Awhanjigoh; Ibereko; Keta-East; Iworo...
6	Epe	Etita/Ebode; Lagbade; Popo-Oba; Oke-Balogun; A...

The Wards column had the required data, thankfully, but with multiple delimiters (; and /), and some neighborhood names written in compound forms. The most important columns to the projects, the Latitude and Longitude columns were not part of the downloaded data, so these were part of the data wrangling.

The downloaded data would in no way return favorable results if passed through GeoCoder in the form they were downloaded, so a little more manipulation was done, by adding Lagos, Nigeria to their names.

The delimiters on the neighborhood (wards) column were removed, and a new dataframe was created for the neighboorhood, leading us to the main project data.

2.3 Feature Selection

The LGA (Borough), Wards(Neighborhoods), Latitudes and Longitudes are the mail features required for this project, though our dataset came with population, we hope to use this feature for the improvement highlighted at the end of this report.

	LGA	Wards	Locale	Latitude	Longitude
0	Agege	Isale/Idimangoro; Iloro/Onipetesi; Oniwaya/Pap...	Agege, Lagos, Nigeria	6.625256	3.311209
1	Ajeromi/Ifelodun	Ago Hausa; Awodi-Ora; Wilmer; Olodi; Tolu; Tem...	Ajeromi/Ifelodun, Lagos, Nigeria	6.455122	3.335946
2	Alimosho	Shasha/Akowonjo; Egbeda/Alimosho; Idimu/Isheri...	Alimosho, Lagos, Nigeria	6.584343	3.257631

3.0 Methodology

3.1 Neighborhoods (Wards) Data Wrangling

Since our target in this project is the neighborhood, and the data from the internet is far from what we want, a lot of data munging was done. Since the LGAs have multiple wards, and are separated by multiple delimiters, we have to find a way to make it usable, and saved to a dataframe,

From this form

	LGA	Wards
0	Agege	Isale/Idimangoro; Iloro/Onipetesi; Oniwaya/Pap...
1	Ajeromi/Ifelodun	Ago Hausa; Awodi-Ora; Wilmer; Olodi; Tolu; Tem...
2	Alimosho	Shasha/Akowonjo; Egbeda/Alimosho; Idimu/Isheri...

to this form

	Neighs	Latitudes	Longitudes
0	Shasha	6.91272	4.34365
1	Akowonjo	6.61048	3.31031
2	Egbeda	5.23472	6.7525
3	Alimosho	6.58434	3.25763
4	Idimu	6.57224	3.25704

for each LGAs, but since our interest and focus is on the biggest LGA in Lagos, Nigeria, we focused only on Alimosho.

4.0 Result

4.1 Business Classification

We were able to classify the businesses in Akowonjo neighborhood, and the result as sampled below were shared into 5 clusters.

Examining the Clusters

In [66]:
```
akowonjo_merged.loc[akowonjo_merged['Avan'] == 0, akowonjo_merged.columns[[1]+list(range(5, akowonjo_merged.shape[1]))]]
```

Out[66]:

	Latitudes	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
4	6.57224	Shopping Mall	Grocery Store	Bus Station	Café	Chinese Restaurant	Clothing Store	Coffee Shop	Convenience Store	Discount Store

In [68]:
```
akowonjo_merged.loc[akowonjo_merged['Avan'] == 1, akowonjo_merged.columns[[1]+list(range(5, akowonjo_merged.shape[1]))]]
```

Out[68]:

	Latitudes	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
1	6.61048	Mobile Phone Shop	Fast Food Restaurant	Shopping Mall	Grocery Store	Bar	Café	Chinese Restaurant	Clothing Store	Coffee Shop

5.0 Conclusions

In this project, we were able to first of create a data that were not readily available on internet, it's a big win as far as I am concerned, and secondly, we were able to get classifications based on top ten common businesses in Alimosho Local Government Area, using the data available through FourSquare.

6.0 Future Improvements

From the result, we can exploit further the least common businesses in Alimosho Local Government, and then using the available population data to further analyse the Local Government, which will depend on our intended goal.