Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in LAGOS, NIGERIA

By: ABIMBOLA JEMINAT MURITALA

FEBRUARY,2020

Introduction

Shopping malls are a large enclosed shopping area from which traffic is excluded. Shopping malls consists of so many outlets which sells different things ranging from clothes to toys to food to groceries to fashion items. They also consists of fun spots, hangout spots, entertainment, restaurants and so on. Shopping malls are very essential to have in cities for easy access to household needs. For retailers, Shopping malls provide a great avenue for them to showcase or advertise their products and services because of the central locations and the large crowd. Also, it is of great investments to the property developers and investors because of the rates at which outlets are given out as rent. This also has made a lot of them to build Shopping malls around Lagos. There are quite a number of Shopping malls in Lagos but they seem many in some particular locations in Lagos. Opening a shopping mall in Lagos, Nigeria, One needs to be very careful of the location because of competition and as well having a large crowd. The location especially will determine the success or failure of any shopping mall.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Lagos, Nigeria to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Lagos , Nigeria , if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in Lagos Nigeria as there are many shopping malls. Data from the National Property Information Centre (NAPIC) released last year showed that an additional 15 per cent will be added to existing mall space, and the agency predicted that total occupancy may dip below 86 per cent. Tweets from December last year shows that the true occupancy rates in malls may be as low as 40 per cent in some areas, quoting a Financial Times (FT) article cataloguing the country's continued obsession with building more shopping space despite chronic oversupply.

Data

To solve the problem, we will need the following data:

- List of neighbourhoods in Lagos. This defines the scope of this project which is confined to the city of Lagos, Nigeria situated at the west of Africa.

- Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Source of data and methods to extract them

My data was gotten from Wikipedia and it contains a list of neighbourhoods in Lagos, with a total of 14 neighbourhoods. I will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup package. Then I will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give me the latitude and longitude coordinates of the neighbourhoods.

After that, I will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, I am particularly interested in the Shopping Mall category in order to help me solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). The next section will consist of the Methodology section where I will discuss the steps taken in this project, the data analysis that I did and the machine learning technique that was used.

Methodology

The first thing I did was to get the list of neighbourhoods in the city of Lagos. The list is available in the Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighborhoods_of_Lagos). I then did web scraping using **Python requests** and **beautiful soup** packages to extract the list of neighbourhoods data. I then got the geographical coordinates in the form of latitude and longitude in order to be able to use **Foursquare API**. To do so, I used the **Geocoder package** that converted the addresses into geographical coordinates in the form of latitude and longitude. After gathering the data, I populated the data into a pandas data frame and then visualized the neighbourhoods in a map using **Folium** package. This allowed me to perform a check to make sure that the geographical coordinates data returned by Geocoder were correctly plotted in the city of Lagos.

Then, I used **Foursquare API** to get the top 100 venues within a radius of 2000 meters. Then I made API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare then returned the venue data in JSON format and I extracted the venue name, venue category, venue latitude and longitude. With the data, I checked how many venues were returned for each neighbourhood and examined how many unique categories were curated from all the returned venues.

Then, I analysed each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. I also prepared to be used for clustering. Because I was analysing for Shopping Mall data, I filtered Shopping Mall as the venue category for the neighbourhoods.

Lastly, I performed clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. I clustered the neighbourhoods into 3 clusters based on their frequency of occurrence for Shopping Mall. The results then identified which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will helped to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

Results

The results from the k-means clustering show that there are 3 categories of neighbourhoods based on the frequency of occurrence for Shopping Mall:
• Cluster 0: Neighbourhoods with moderate number of shopping malls
• Cluster 1: Neighbourhoods with high concentration of shopping malls
• Cluster 2: Neighbourhoods 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 colour, cluster 1 in purple colour, and cluster 2 in mint green colour.
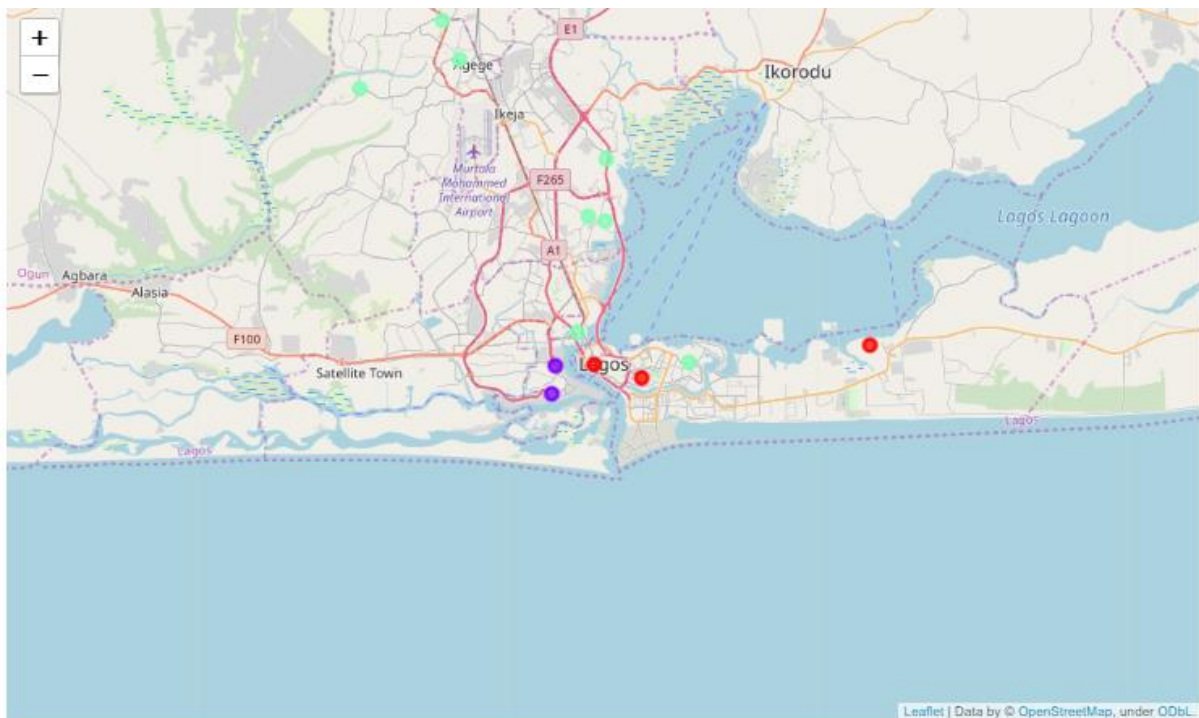
Discussion

As observations noted from the map in the Results section, The shopping malls in cluster 0 have high concentration are in Lagos, Nigeria and they are situated at the central area of the city. As shown in cluster 0 in the map visualization.

On the other hand, cluster 1 have moderate number of shopping mall in the neighbourhoods. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. Meanwhile, shopping malls in cluster 2 do not have shopping malls at all.

From another perspective, the results also show that the oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in cluster 2 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighbourhoods in cluster 1 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 0 which already have high concentration of shopping malls and suffering from intense competition.

Limitations and Suggestions for Future Research

In this project, only one factor was considered i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents that could influence the location decision of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new shopping mall.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 1 are the most preferred locations to open a new shopping mall. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.

REFERENCES

- Category: Neighbourhoods of Lagos Wikipedia. Retrieved from
  https://en.wikipedia.org/wiki/Category:Neighborhoods_of_Lagos

- Foursquare Developers Documentation. Foursquare. Retrieved from
  https://developer.foursquare.com/docs