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.