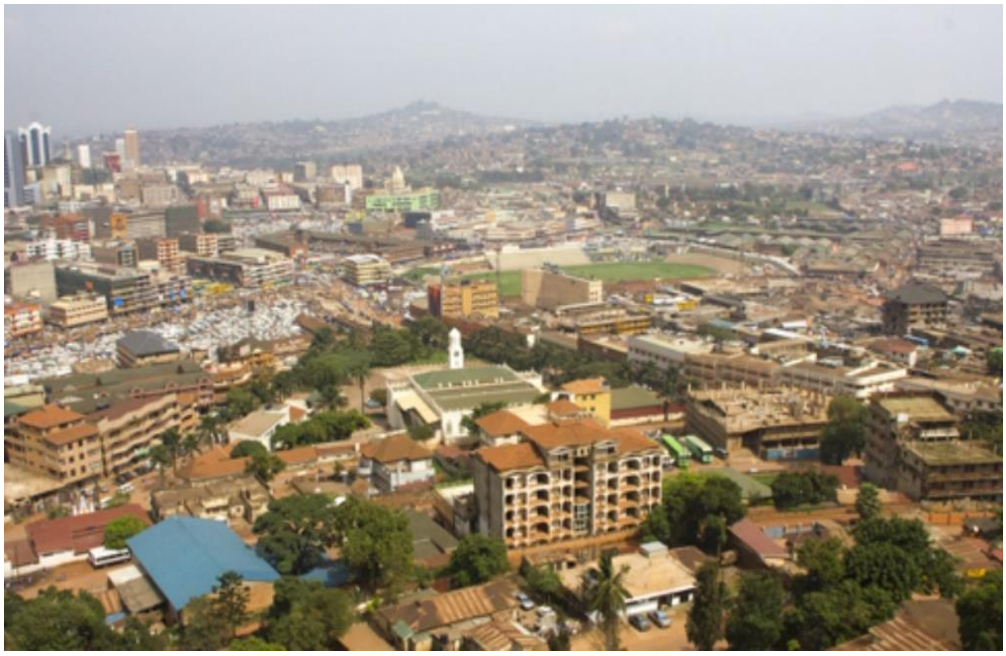
COUSERA CAPSTONE PROJECT

IBM APPLIED DATA SCIENCE CAPSTONE

OPENING A NEW SHOPPING MALL IN KAMPALA, UGANDA

BY: GODFREY OBONG

MARCH 2019



Arial view of Kampala city center and surrounding neighborhoods

INTRODUCTION

Visiting a mall is always a great way to relax and enjoy for many shoppers over the weekend, holidays and even weekdays. Malls are known to offer great experiences such as large grocery stores, butcheries, bakeries, and home stores, multi-cuisine cheerful and expansive food court, family dining restaurant, cinema, gym and health club, parking and security, clothing stores, shoe stores, and many more. Shopping malls are a one stop destination for the public with many different interest in mind. For the business owners the central location and the large crowds provide a great distribution Chanel to market their products and services. Property investors are also keen on this trend due to the potential to earn consistent rental income from

the business owners. In the last decade many malls and shopping centers both multi National and locally owned have sprang up in the central district and the neighboring suburbs of Kampala city. Opening a new shopping mall requires serious consideration. Particularly the location of the shopping mall is one of the most important decisions that determines the success of the venture.

Business problem

The main objective of the capstone project is to analyze and select the best location to set up the next high end shopping mall in one of the suburbs of Kampala city, Uganda. Using data science methodology and machine learning, I will analyze locational venue data from the foursquare site to come up with the best current location to set up a shopping mall. The objective is to have a high end location that is in a suburb that has little to no shopping mall within and it's the surrounding neighborhoods. It would also be plausible that the surrounding neighborhoods do not also have shopping malls within. The location should also be in a high end to middle income suburb. Since the mall is intended to be high end with classy stores and restaurants, the target customers will have to be the close by neighborhood dwellers who should view visiting the mall as routine and necessity. For this I will consider neighborhoods with high value property and rental rates. In Uganda the individuals that can afford the high property rates are the ones that frequent the malls while the low income earners prefer to stick to the local markets and neighborhood small supermarkets as these avenues are viewed to be less expensive.

Target

This particular project is particularly useful to property developers that are looking to open up a new shopping mall in the city of Kampala. Currently Uganda is in the development stage of shifting into a middle class economy. The middle class of Uganda is growing by the day and its this group that can be found in some of the affluent neighborhoods of Kampala. This class residing in the upscale suburbs are the principle target population for this particular property developer.

DATA

The following data was used to solve the problem

- A list of suburbs around Kampala city. The scope of the project is defined by the suburbs around Kampala.
- The latitudes and longitudes coordinates of the mentioned suburbs which will be used in plotting the maps and acquiring venue data.
- Venue data from foursquare API using locational latitude and longitude coordinates.
- Average rental prices in the different neighborhoods to be used as an inference to the general social standing in the different suburb areas.

Source of data

The suburbs of Kampala city can be found listed on sites such as " wiki 1" and "source". Since we are targeting to set up the mall in an affluent suburb. The average rental charges used can be got from the rental index published by Knight Frank a property development company in Uganda and other web. We shall then consider only the high rental suburb areas. Then I will get the geographical coordinates of the suburbs using python Geocoder package which will give us the latitudes and longitude coordinates of the suburb.

I shall then use the Foursquare API to get the venue data for those suburbs. Foursquare has one of the largest database of 105 plus million venues used by over 125,000 developers. Foursquare API will return many categories of venue data, I am particularly interested in the Shopping Mall category in order to help solve the business at hand. This project will make use of data science skills, from web scraping, working with API "Foursquare", data cleaning, applying machine learning algorithm and visualization using the folium package.