

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 thi

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 should be 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.

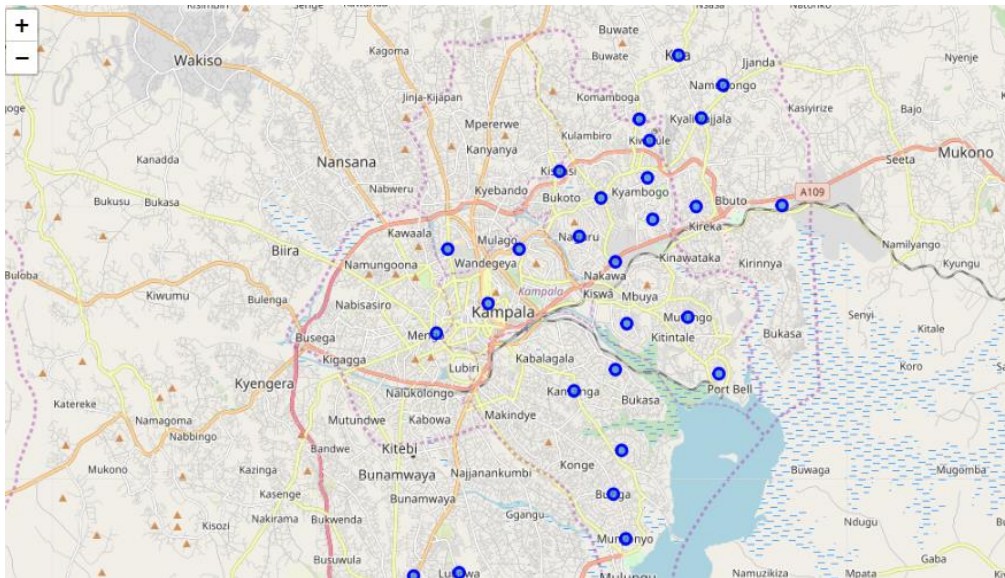
Source of data

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 sources such as [numbeo.com](#). Then I got the geographical coordinates of the suburbs using python Geocoder package which only returned a few of the coordinates. Luckily the data set is small and thus for the rest of the coordinates I had to use different google sites such as [www.latlong.net](#), [fallinggrain](#) which actually returned accurate coordinates as per the map.

I then used 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 returned many categories of venue data, I was particularly interested in the Shopping Mall category in order to help solve the business at hand. This project made use of data science skills, from web scraping, working with API "Foursquare", data cleaning, applying machine learning algorithm and visualization using the folium package.

Methodology

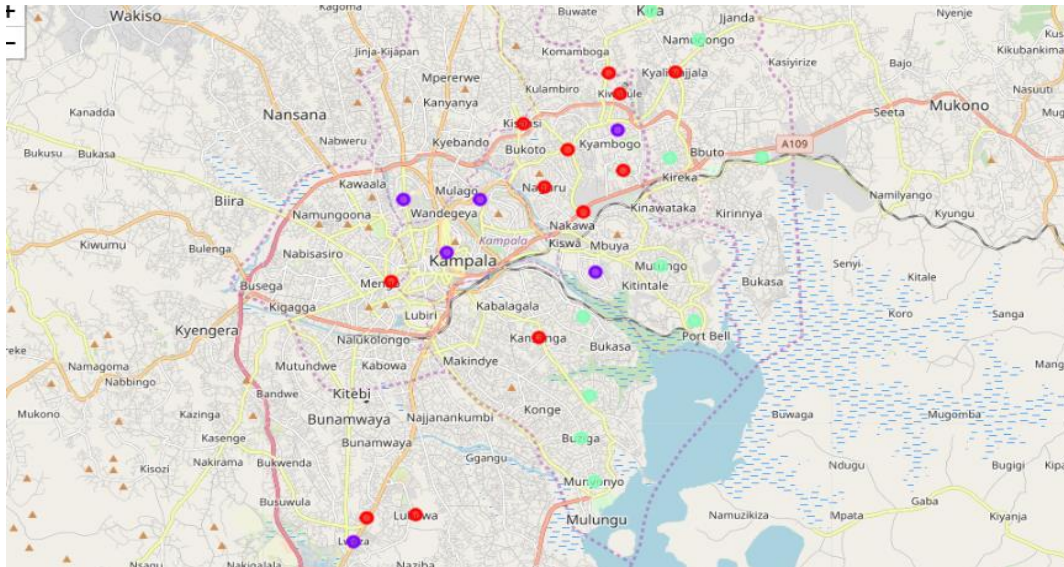
First I needed to list the prominent suburbs around Kampala. Unfortunately, this list is not readily available and I had to scrape for this information from different websites. These included [ecolandprperty](#), [ugbusiness.com](#) and many more. I also needed to get the geographical coordinates in the form of latitudes and longitudes for the different suburb location in order to use the Foursquare API. To do that, I used the Geocoder package that converts addresses into geographical coordinates in the form of latitudes and longitudes though some of the locations were not showing up after running the code. I had to use different google sites such as [www.latlong.net](#), [fallinggrain](#) which actually returned accurate coordinates as per the maps generated by folium. After gathering the data in an excel sheet, I then converted it into a DataFrame in python from which I was able to visualize the suburbs in a map using Folium package. This ensured that the geographical coordinates I was using were correctly plotted in the city of Kampala map.



Map indicating suburbs of Kampala

Next, I used the Foursquare API to get the top 100 venues within a radius of 2000 meters. To use the Foursquare API we need to register a developer account in order to obtain the Foursquare ID and secret key. I then made an API call passing in the coordinates of the suburbs in a python loop. Foursquare returned the venue data in JSON format from which I extracted Venue name, category, latitude and longitude. This returned 835 venues of which only 51 were shopping malls. I then filter the "Shopping Mall" as venue category for the suburbs and then performed clustering on the data using the "k means clustering". The 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 most popular machine learning algorithm and is particularly well suited for this project. The result allowed me to identify which neighborhoods have the highest concentration of shopping malls and those that have fewer or none at all based on the occurrence of shopping malls in the different suburbs.



Map indicating clustering of the suburbs in Kampala

Further analysis/ Alternative

Our first criteria of the project was to attain Suburbs without shopping malls. I went ahead and removed all the suburbs with shopping malls from the initial list which left me with 14 suburb areas. Further more I filtered out the suburbs to attain the high end high rental neighborhoods. To do this I dropped all the neighborhoods with average monthly rental charges below 1,000,000 ug shillings which is roughly 270\$. This left me with 7 suburbs namely Bukoto, Bunga, Buziga, Luzira, Lweza, Munyonyo and Muyenga

As a native of Kampala city, the above mentioned suburbs are known to reside the wealthy elites of Kampala. A further visual analysis of the map generated a clustering of the areas of Bunga, Buziga, Muyenga and Munyonyo which is indicated within the yellow circle.



Setting up the shopping mall at the center of this circle would be idle since it's in the vicinity of four high rent charge suburbs ie. Bunga, Muyenga, Buziga and Munyonyo. This location can easily serve the residents of these four neighborhoods.

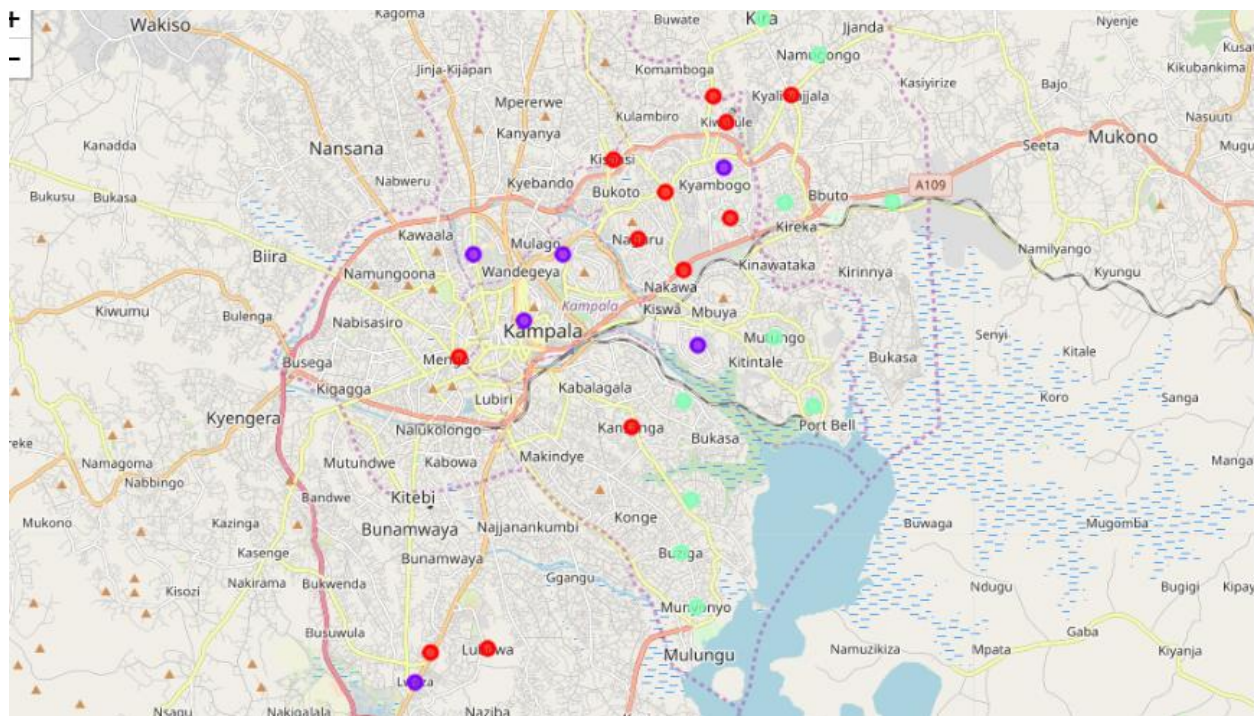
RESULTS

Clustering

The results from the k-means clustering show that we can categorize the neighborhoods into 3 based on frequency of occurrence for “shopping malls”.

- Cluster 0: Suburbs with high concentration of shopping malls.
- Cluster 1: Suburbs with moderate number of shopping malls.
- Cluster 2: Suburbs with no shopping malls.

The results of the clusters are visualized in the map below with cluster 0 in red, cluster 1 in purple and cluster 2 in mint green color.



Map showing clustered suburbs

Further analysis

Farther more, filtering out the suburbs with already existing shopping mall venues and those with rental charges below 1,000,000 Ugandan shillings left me with 7 suburbs namely Bukoto, Bunga, Buziga, Luzira, Lweza, Munyonyo and Muyenga. The best location to set up a high

end shopping mall as per this project would be in the areas of Bunga, Buziga, Munyonyo and Muyenga as indicated in the yellow circle of the map below



Limitations and suggestions for future research

For this project, I only considered two factors i.e existence of a shopping mall in a suburb and average rental charges in that same particular location. There are other factors such as population, availability of social services and even income of residents. However, to the best of this research, such data was not readily available to the suburb level as required by the project. Future research could devise a methodology to estimate such data in clustering algorithm to determine the preferred location. In addition, this project made use of free Sandbox Tier Account of Foursquare API that comes with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

Conclusion

In this project I identified the business problem, specified data requirements, extracted and prepared the data to perform machine learning clustering by clustering the data into 3 clusters based on their similarities. I then provided recommendations to the relevant stakeholders namely the property developer regarding the best location to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project would be the suburbs in cluster 2 as first identified. Further analysis revealed that areas specifically around the suburbs of Bunga, Buziga, Munyonyo and Muyenga would be very ideal. The findings in this project will help the relevant stakeholders capitalize on the opportunities on the high potential locations while avoiding overcrowded areas in their decision to open a new shopping mall in Kampala.

References

Suburbs in Kampala, Uganda

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

knight frank property update

<https://content.knightfrank.com/news/12653/5902-article-1.pdf>

Kampala city

<https://www.ecolandproperty.com/kampala/>

Apendix

Cluster 0

-Kyambogo, Nakawa, Najjera, Naguru, Mengo, Lubowa, Ntinda, Kyaliwajjala, Kiwatule, Kisaasi, Seguku, Kansanga

Cluster 1

-Bugolobi, Kololo, Entebbe, Nalya, Lweza, Makerere, Nakasero

Cluster 2

-Namugongo Bukoto, Bunga, Kitende, Mutungo, Kireka, Buziga, Luzira, Bweyogerere, Kira, Muyenga, Munyonyo