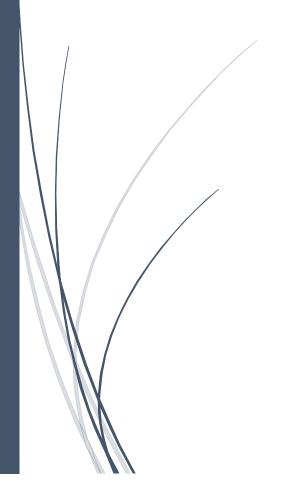5/19/2020

CAPSTONE PROJECT

IBM Professional Certificate

Opening a New Shopping Mall in Colombo, Sri Lanka

Rajitha Senanayake

- # Introduction

For many shopper, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at various fashion outlets, watch movies and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the city of Colombo and many more are being built. Opening shopping malls allows property developers to earn consistent rental income and as with many other business decision, opening a new shopping mall requires serious consideration and is lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

- # Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Colombo, Sri Lanka 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 Colombo, Sri Lanka, if a property developer is looking to open a new shopping mall, where would that they open it?

- # Target Audience of this Project

This project is a particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital of Sri Lanka. This project is timely as the city is currently suffering from oversupply of shopping malls. Data from the National Property Information Centre (NAPIC) released last year showed that the 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. The local newspaper The Daily Mirror also reported in March last year that the true occupancy rates in malls may be as below as 40 per cent in some areas, quoting a Financial Time (FT) article cataloguing the country's continued obsession with building more shopping space despite chronic oversupply.

● Data

To develop a model to cluster neighborhoods, we need the following data:

➢ List of neighborhoods in Colombo. This defines the scope of this project which is confined to the city of Colombo.
➢ Latitude and longitude coordinates of those neighborhoods. This require in order to plot the map and to get venue data.
➢ Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

● Sources of data and method of extraction

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Colombo) contains a list of neighborhoods in Colombo, with a total of 67 neighborhoods. We will use web scrapping techniques to extract the data from the Wikipedia page, with the help of pandas and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhood using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use foursquare API to get the venue data for those neighborhoods. 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, we are particularly Interested in the Shopping Mall category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skill, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that we used.