Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Kuala Lumpur, Malaysia

By: Shivam Raman Sawhney

Introduction

For many buyers, a shopping center is a great way to relax and have fun on weekends and holidays. They can go shopping, eat in restaurants, shop at various thrift stores, watch movies and do many other events. Supermarkets are one destination for all types of customers. For retailers, many shopping centers are an excellent sales channel for marketing their products and services. Entrepreneurs also use this trend to create more malls to meet demand. As a result, there are several shopping centers in Kuala Lumpur, and many others are under construction. The opening of a shopping center allows property developers to receive a fixed rental income. Of course, as with any business decision, opening a new market requires serious consideration and is more complicated than it seems. In particular, the location of the market is one of the most important decisions that will determine whether the market is successful.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Kuala Lumpur, to open up a new shopping mall. Using data science methodology and the techniques learned along with machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Kuala Lumpur, 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 the capital city of Kuala Lumpur. 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. The local newspaper The Malay Mail also reported in March last year 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 neighborhoods in Kuala Lumpur. This defines the scope of this project which is confined to the city of Kuala Lumpur, the capital city of the country of Malaysia in South East Asia.
- Latitude and longitude coordinates of those neighborhoods. 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 neighborhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur) contains a list of neighborhoods in Kuala Lumpur, with a total of 70 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods 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 databases 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 skills, 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 was used.