Coursera Capstone

IBM Data Science Professional Certificate

Planning to open a New Shopping Mall in Kuala Lumpur, Malaysia

Week 4

Prepared By: Divin Divakar
Dated: 23 May 2021

Introduction

Shopping malls are a popular way for many people to unwind and enjoy themselves on weekends and holidays. They can shop for groceries, eat at restaurants, shop at various fashion outlets, watch movies, and engage in a variety of other activities. Shopping malls provide as a one-stop shop for a variety of shoppers.

The shopping malls' central location and vast crowds provide an excellent distribution route for shops to sell their products and services. Property developers are also capitalizing on this trend by constructing more shopping malls to meet demand. As a result, Kuala Lumpur has a plethora of retail malls, with more being erected all the time.

Property developers can earn continuous rental income by developing shopping malls. Of course, opening a new retail mall, like any other business decision, necessitates careful analysis and is far more complicated than it appears. The site of the shopping mall is one of the most critical factors that will determine whether the mall succeeds or fails.

Business Problem / Problem Statement

The goal of this capstone project is to research and pick the finest locations for a new retail mall in Kuala Lumpur, Malaysia. This project seeks to deliver solutions to the business question using data science methodology and machine learning techniques such as clustering.

If a property developer wanted to build a new retail mall in Kuala Lumpur, Malaysia, where would you suggest they build it?

Target Audience of this project

Property developers and investors planning to open or invest in new shopping malls in Malaysia's capital city, Kuala Lumpur, will find this project particularly valuable. This initiative comes at a good moment because the city is now experiencing an overabundance of shopping complexes.

According to data issued last year by the National Property Information Centre (NAPIC), existing mall space will be expanded by 15%, and total occupancy might fall below 86%.

In March of last year, the local newspaper The Malay Mail stated that genuine mall occupancy rates in some locations could be as low as 40%, citing a Financial Times (FT) piece detailing the country's persistent infatuation with creating more shopping space amid chronic oversupplies.

Data

We will need the following information to solve the problem:

- A list of Kuala Lumpur neighborhoods. This outlines the project's scope, which is limited to the city of Kuala Lumpur, Malaysia's capital, and largest metropolis in Southeast Asia.
- Coordinates of those neighborhoods' latitude and longitude. This is essential for both plotting the map and retrieving the venue information.
- Statistics about venues, notably data on commercial malls. This information will be used to cluster the neighborhoods.

Data Sources and Methods of Extraction

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs in Kuala Lumpur) offers a list of 70 different neighborhoods in Kuala Lumpur. With the help of Python requests and beautiful soup packages, we will extract data from the Wikipedia page using web scraping techniques. Then, using the Python Geocoder library, we will get the geographical coordinates of the neighborhoods, which will give us their latitude and longitude coordinates.

Following that, we will use the Foursquare API to gather venue data for those areas. Foursquare is used by over 125,000 developers and has one of the largest databases of 105+ million places.
The Foursquare API will provide a variety of venue data categories, but we are particularly interested in the Shopping Mall category to help us address the business challenge we have presented.

This project will require a wide range of data science abilities, including web scraping (Wikipedia), API work (Foursquare), data cleaning, data wrangling, machine learning (K-means clustering), and map visualization (Folium).

The Methodology part will cover the procedures performed in this project, the data analysis that was done, and the machine learning methodology that was utilized.

---------- End of Document ----------