

**IBM DATA SCIENCE
CAPSTONE PROJECT**

Property Development Investment
for a Department Store

Contents

Introduction	3
Background	3
Problem.....	3
Project Interest	3
Data.....	4
Data Requirement.....	4
Source of Data.....	4
Methodology.....	5
Results and Discussion	6
Conclusion.....	6

Introduction

Background

Malaysia is a trending tourism destination in South East Asia increasing arrival growth to over 4.8% in 2019. To accommodate the trending lifestyle and shopping experience for types of individuals, it will be a good opportunity to develop a department store in a feasible area in Kuala Lumpur (KL) to help shoppers spend time in one location to purchase any item they may need.

The current shopping concept in KL is a mall concept. Department Store concept is different from a mall concept and it allows shoppers to purchase quality and branded items at a low cost relatively to products at luxury malls. Property Developers are interested to invest in similar trending concepts due to the increase in tourism activities. But the key concern is finding that premium location which attracts shoppers conveniently and a location that will build demand.

Problem

In order to invest in a multi-million dollar construction project, it is essential to select the geographically best location that is available at a lower cost, and will eventually build demand for shopping, making the investment a success. This capstone will recommend a solution for Property Developers to select an appropriate location to build a Department Store as per their plan.

Project Interest

Property Developers in Malaysia will be the key stake holders of this project part from the communities around the suburb areas in KL. Business investors are another category that will keep an open eye on this project outcome.

Data

Data Requirement

The names of the cities in KL territory and their geometric coordinates

Foursquare venue data in the relevant cities

Source of Data

The cities will be obtained from Wikipedia which will help to extract the geographical coordinates required for this analysis. Beautifulsoup and web scraping techniques will be used along with Geocoder tool to pull up above data.

The Foursquare API will immensely help to explore and obtain venue data of the respective cities of interest.

Once all the data is obtained, they will be going through different types of data manipulation in order to model the required solution for the problem in discussion, namely they are, data clustering, cleaning, segmenting, wrangling and visualization.

Methodology

To begin with, all relevant libraries were imported. Then data was scraped out of **HTML** and XML file using BeautifulSoup from the Wikipedia link which provided the cities in KL. This data was stored in a table, and there are 71 cities in KL territory. Next the geocoder was used to obtain the geographic coordinates of each city in the table. Each city was assigned to a table including its geo coordinates in perspective to longitude and latitude. The geo coordinates for KL was obtained and the map of KL was populated and saved.

The cities were explored by Foursquare API. Before this, a foursquare account has to be created with Foursquare Developer link and obtain the unique Client ID and Client secret ID along with the version.

Next the top 100 venues were searched on Foursquare within the radius of 2000m from KL. The extracted venues are assigned to a table with 7 columns labels. Then it was verified that each city has provided 100 venues each as per the GET request. This venue list has 53 unique venue types which were listed. Then explore and analyze each city to see what types of venues are available. Each city is sorted and grouped according to the available venues and its frequency of occurrence as a mean.

Now the Department Store data occurrence is extracted according to the cities. With K-means clustering, the data is segmented to 3 categories and each cluster category is listed with the respective city. Next, using each cluster analysis is taken place.

Results and Discussion

As per the analysis, the cities are clustered into 3 segment, 0, 1 and 2. Cluster 0 displays few Department Stores that are available in the area, but cluster 1 and 2 does not. The locations in red in the map are the cities where a Department Store is not in demand.

With this result it is highly recommended that Property Developers could invest on a Department Store in cities listed in cluster 1 and 2, as there are no Department Stores in those cities.

Conclusion

In order to simply deliver an construction project of a emmence investment, the Property Developers have to follow an analytical and statistical judgement to make a conscise decision when selecting the loction for a Department Store. With Data Science tolls and techniques, it has made easy for such business investors to make a good pick for projections and investment.

As to the analysis in this scenario, it is recommended to develop a Department Store in the cities listed in cluster 1 and 2 due to the unavailability of department Stores in those areas.