

Cousera Capstone

IBM Applied Data Science Capstone

Location for shopping mall in KL



Intro

Shopping malls are

→ **Good investment**

Good investment as they are one-stop destinations for entertainment, shopping and food.

→ **Win-win**

This is a win-win for the retailers as they get a good distribution channel, and also for property developers who get to take advantage of it.

→ **Simple**

Location of a shopping mall is very important which will determine if the same will be a success or failure.

Business Problem

The objective of this project is to analyze and select the best place to open a new shopping mall in Kuala Lumpur. Using data science and machine learning, need to recommend where the shopping mall can be opened.



Data acquisition and cleaning

By the end of this section, your audience should be able to visualize:

→ Data Source

- List of neighbourhoods in Kuala Lumpur
- Latitude and Longitude of the neighbourhoods in Kuala Lumpur.
- Venue data which gives data regarding the shopping malls

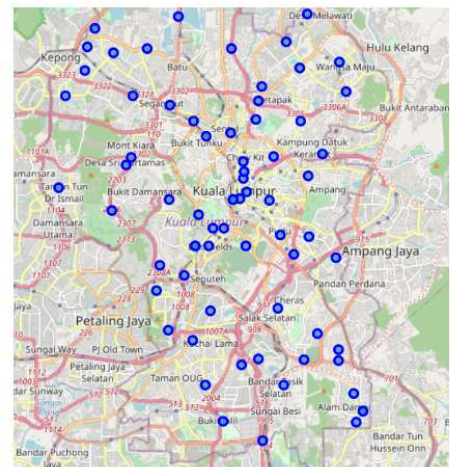
Data acquisition and cleaning

→ Data Cleaning

- Neighbourhoods data can be obtained using the below wiki page and can be extracted using the Beautiful soup packages in python
- Geocoder package can be used to get the latitude and longitude information
- Foursquare website can be used to get the venue details regarding the shopping malls.

Analysis

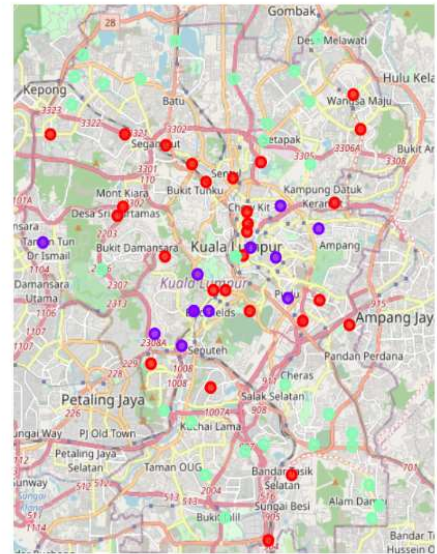
1. Web scraping using python requests and beautifulsoup packages to extract the list of neighbourhood data. Geocoder package can be used to get the latitude and longitude information
2. Using geocoder we will get the coordinates of these neighbourhoods. It can be converted to a dataframe and populated the same in maps using Folium. Foursquare API is used to get the 100 venues within 2000 meters of radius using the foursquare id and key.



Machine learning - K-means

1. We can use K-means clustering for performing clustering on the data. It is the simplest and popular unsupervised ML algorithm and this method suits the problem in hand.

Cluster the neighbourhoods based on their frequency of occurrence of shopping malls



Results

→ **k-means clustering**

- ➔ Cluster 0: Neighbourhoods with moderate number of shopping malls
- ➔ Cluster 1: Neighbourhoods with low number of shopping malls
- ➔ Cluster 2: Neighbourhoods with high number of shopping malls



Conclusion

The neighbourhood in cluster 1 is the most preferred location to open a shopping mall.