IBM Data science

Opening a New Shopping Mall in Kuala Lumpur

Business Problem

- 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.

- Objective: To analyse and select the best locations in the city of Kuala Lumpur, Malaysia to open a new shopping mall.

- This project is timely as the city is currently suffering from oversupply of shopping malls

Data

- Data required
  - ➢List of neighbourhoods in Kuala Lumpur
  - ➢Latitude and longitude coordinates of the neighbourhoods
  - ➢Venue data, particularly data related to shopping malls

- Sources of data
  - ➢Wikipedia page for neighbourhoods https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur
  - ➢Geocoder package for latitude and longitude coordinates
  - ➢Foursquare API for venue data

Methodology

- Web scraping Wikipedia page for neighbourhoods list

- Get latitude and longitude coordinates using Geocoder

- Use Foursquare API to get venue data

- Group data by neighbourhood and taking the mean of the frequency of occurrence of each venue category

- Filter venue category by Shopping Mall

- Perform clustering on the data by using k-means clustering

- Visualize the clusters in a map using Foliu

Results

- Categorized the neighbourhoods into 3 clusters :

  ➢Cluster 0: Neighbourhoods with moderate number of shopping malls

  ➢Cluster 1: Neighbourhoods with low number to no existence of shopping malls

  ➢Cluster 2: Neighbourhoods with high concentration of shopping malls
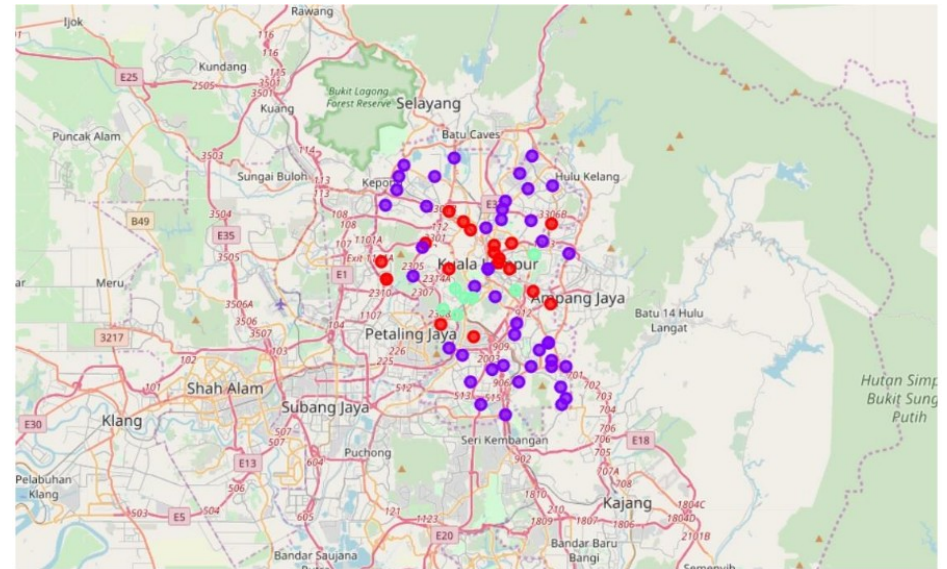
Conclusion

- Answer to business question: The neighbourhoods in cluster 1 are the most preferred locations to open a new shopping mall

- Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall

Thank you