THR BATTLE OF NEIGHBORHOOD
Applied data science capstone

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

COURSERA

Created by: Dheemanth Shenoy D

INTRODUCTION

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

Business question

- In the city of Kuala Lumpur, Malaysia, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

DATA AND SOURCE

Data required

- List of neighborhoods in Kuala Lumpur
- Latitude and longitude coordinates of the neighborhoods
- Venue data, particularly data related to shopping malls

Source

- 1.Wikipedia page for neighborhoods (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)
- 2.Geocoder package for latitude and longitude coordinates
- 3.Foursquare API for venue data

METHODOLOGY

- Web scraping Wikipedia page for neighborhoods list
- Get latitude and longitude coordinates using Geo-coder
- Use Foursquare API to get venue data
- Group data by neighborhood 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 Folium

RESULT
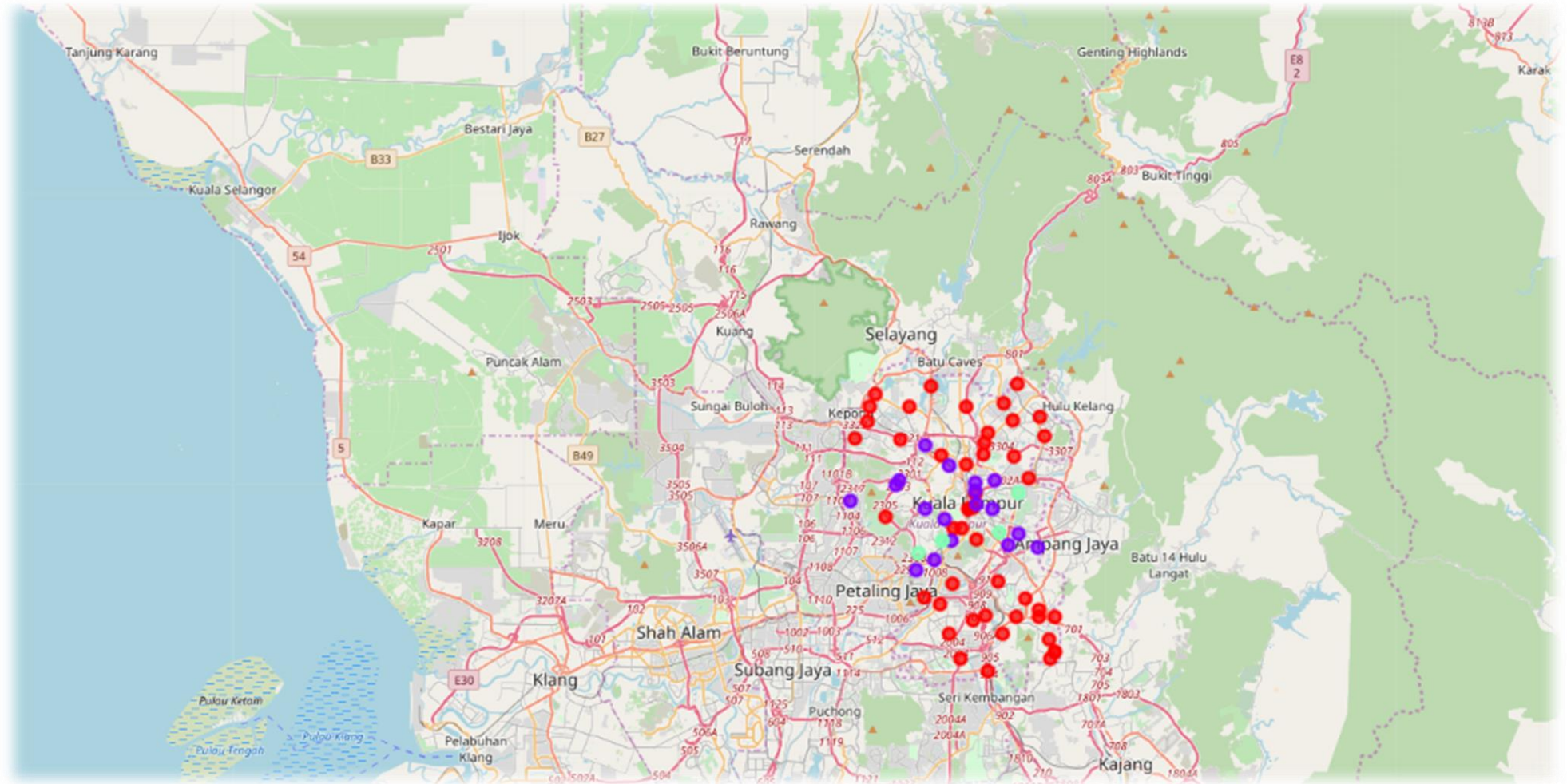
CONCLUSION

- Most of the shopping malls are concentrated in the central area of the city

- Highest number in cluster 2 and moderate number in cluster 1

- Cluster 0 has very low number to no shopping mall in the neighborhoods

- Oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls

- Profitable to start shopping mall in cluster 0