

# OPENING A NEW SHOPPING MALL IN KUALA LUMPUR

Coursera Capstone  
IBM Applied Data Science

## BUSINESS PROBLEM

- Location of the Shopping Mall is one of the most important decisions to determine the success or failure of a new Mall opening
- Objective: Analyse and choose the best location in the city of Kuala Lumpur, Malaysia to build a new Shopping Mall
- Underlying question:  
In the city of Kuala Lumpur, if a property developer is looking into open a new Shopping Mall, where would you recommend to open it?

DATA

#### Required Data:

- List of Neighborhoods in Kuala Lumpur
- Latitude and Longitude information of the neighbourhoods
- Venue data, particular related to Shopping Malls

#### Data Sources:

- Wikipedia page for neighborhoods ([https://en.Wikipedia.org/wiki/Category:Suburbs\\_in\\_Kuala\\_Lumpur](https://en.Wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur))
- Geocoder package for latitude and longitude
- Foursquare API for venue data

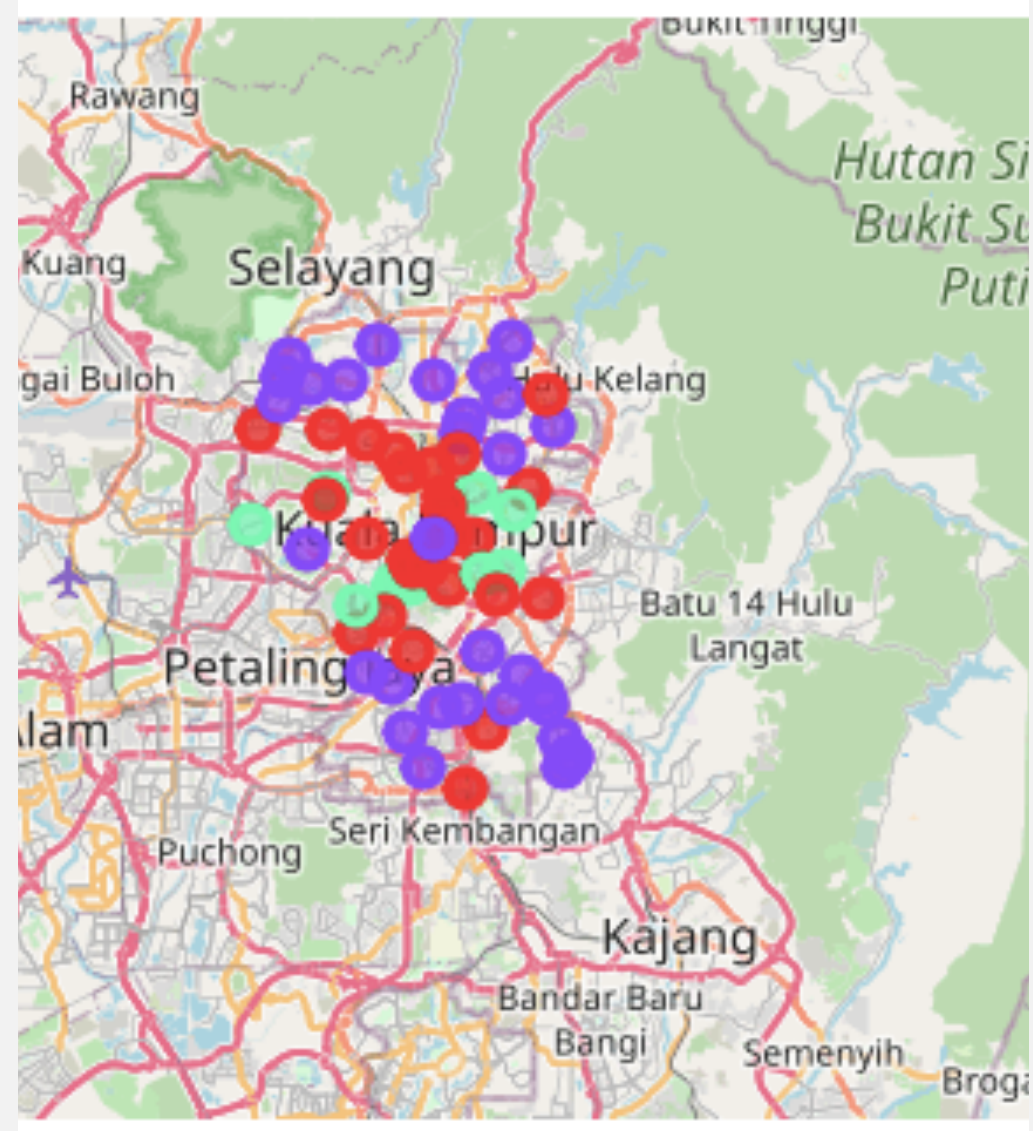
## METHODOLOGY

- Web scraping Wikipedia page for neighbourhood data
- Get lat- and longitude data from geocoder
- Use Foursquare API to get venue data
- Group data and calculate mean of occurrence by venue category
- Filter data to Shopping Mall only
- Run k-mean algorithm to find cluster
- Visualize cluster using a Folium map

# RESULTS

Identification of 3 clusters:

- Cluster 0: Neighborhoods with moderate number of Shopping Malls (red)
- Cluster 1: Neighborhoods with low number to no existence of Shopping Malls (purple)
- Cluster 2: Neighborhoods with high concentration of shopping malls (green)



## DISCUSSION

- Most of the Shopping Malls are concentrated in the central area of the city.
- Largest number in cluster 2 and moderate number in cluster 0.
- Cluster 1 has very low number to no Shopping Malls in their neighborhoods
- Oversupply of Shopping Malls mostly occurring in the central area of the city, with the suburban area with very few Shopping Malls.

## RECOMMENDATIONS

- Further investigation of neighborhoods in cluster 1 in terms of adding more data points i.e. demographics and calculation of business case as they're the most promising areas due to the current data base.
- Gather more data about existing Malls in Cluster 2 and 0 whether they're underperforming and examine measures to exploit those underperforming Malls.

## CONCLUSION

- Answer to underlying question: It is recommended to further investigate the areas in cluster I whether they're having the potential for a successful Mall opening, but more data and business case is required.