

# Coursera Capstone

## IBM Applied Data Science Capstone

### ***Opening a New Shopping Mall in Kuala Lumpur, Malaysia***

By: Rishabh Singh March 2020

# Business Problem

- The location of the shopping mall is one of the most significant decisions that will decide if the mall will be a failure or a success.
- Objective: To break down and choose the best areas in the city of Kuala Lumpur, Malaysia to open another shopping mall
- This task is convenient as the city is right now experiencing oversupply of shopping malls
- Business question
  - In the city of Kuala Lumpur, Malaysia, if a property developer is looking to open another shopping mall, where might you suggest that they open it?

# Data

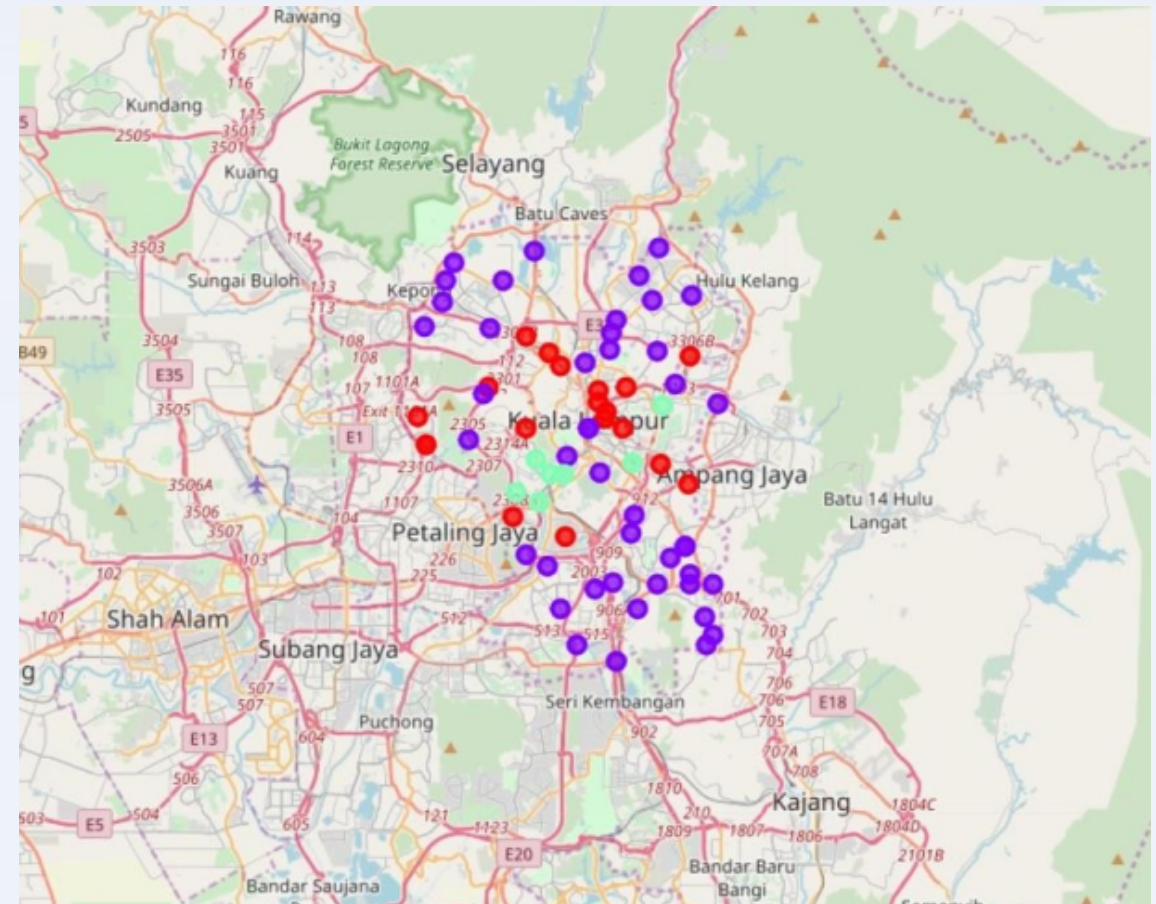
- Data required
  - List of neighbourhoods in Kuala Lumpur
  - Latitude and longitude coordinates of the neighbourhoods
  - Venue data, particularly data related to shopping malls
- Sources of data
  - Wikipedia page for neighbourhoods  
([https://en.wikipedia.org/wiki/Category:Suburbs\\_in\\_Kuala\\_Lumpur](https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_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 Folium

# Results

- The outcomes from the k-means clustering show that we can order the areas into 3 clusters dependent on the recurrence of event for "Shopping Mall":
  - 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





# Discussion

- A large portion of the shopping malls are gathered in the focal region of the city
- Most elevated number in cluster 2 and moderate number in cluster 0
- Cluster 1 has low number to no shopping mall in the neighbourhoods
- Oversupply of shopping malls for the most part occurred in the focal zone of the city, with the suburb region despite everything have not very many shopping malls

# Recommendations

- Open new shopping malls in neighborhoods in cluster 1 with practically zero rivalry
- Can likewise open in neighborhoods in cluster 0 with moderate rivalry if have special offering suggestions to stand apart from the opposition
- Stay away from neighborhoods in cluster 2, effectively high convergence of shopping malls and exceptional rivalry

# Conclusion

- Answer to business question: The areas in cluster 1 are the most favored areas to open another shopping mall
- Discoveries of this task will assist the applicable stakeholders with capitalizing on the open doors on high potential areas while maintaining a strategic distance from stuffed territories in their choices to open another shopping mall