Coursera Capstone
IBM Applied Data Science Capstone
Opening a New Shopping Mall in Kuala Lumpur, Malaysia

By: SANJAY SIDDHARTH

APRIL 2020

Introduction

- Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and
- The objective of this capstone project is to analyse and select the best locations in the city of Kuala
- Lumpur, Malaysia to open a new shopping mall. Using data science methodology and machine
- learning techniques like clustering, this project aims to provide solutions to answer the 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?
- services.

Target Audience of this project

- This project is particularly useful to property developers and investors looking to open or invest in
- new shopping malls in the capital city of Malaysia i.e. Kuala Lumpur.

DATA

- To solve the problem, we will need the following data:
- • List of neighbourhoods in Kuala Lumpur. This defines the scope of this project which is
- confined to the city of Kuala Lumpur, the capital city of the country of Malaysia in South East
- Asia.
- • Latitude and longitude coordinates of those neighbourhoods. This is required in order to
- plot the map and also to get the venue data.
- • Venue data, particularly data related to shopping malls. We will use this data to perform
- clustering on the neighbourhoods.

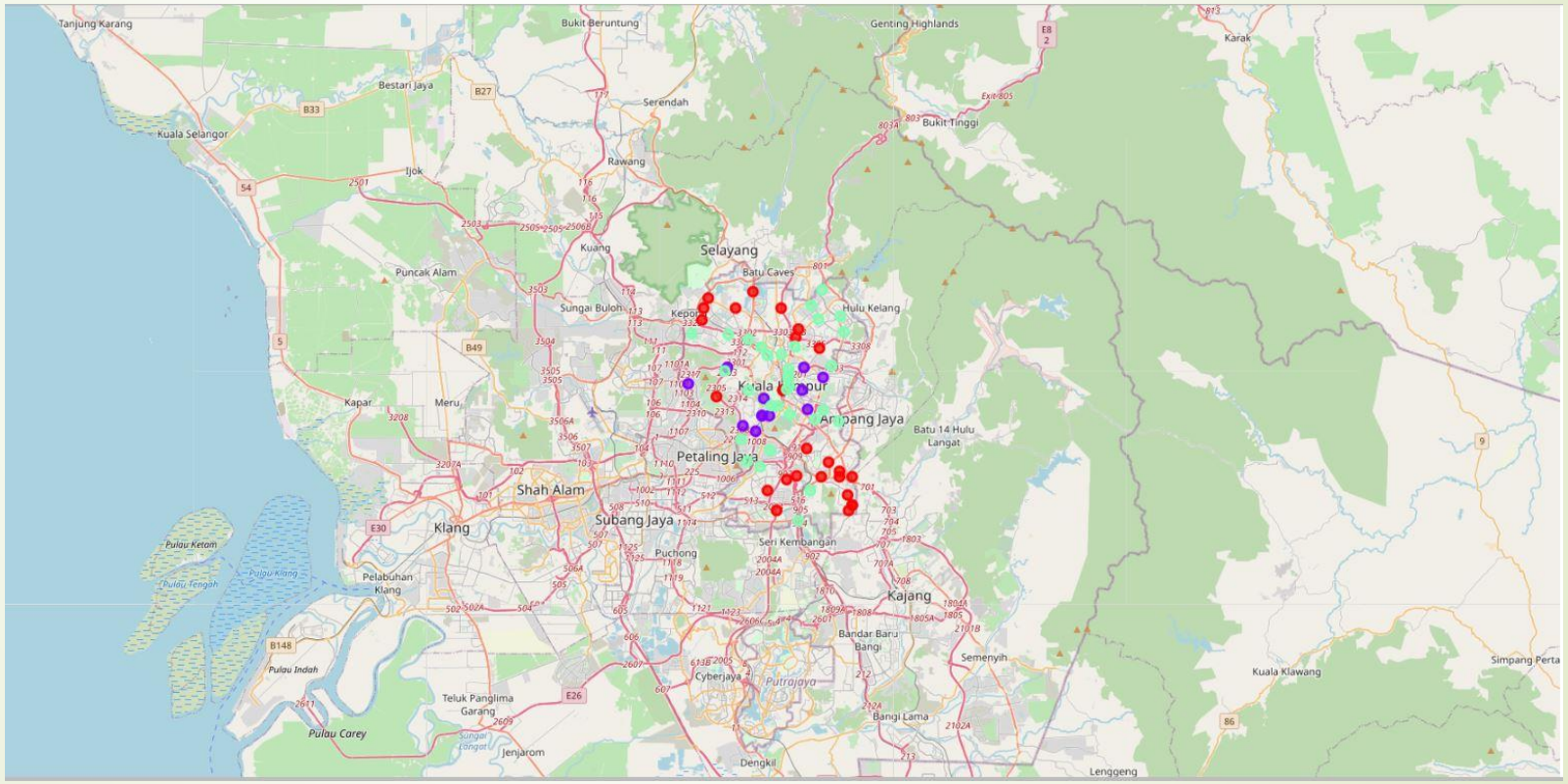
Sources of data and methods to extract them

- This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur) contains a list of neighbourhoods in Kuala Lumpur, with a total of 70 neighbourhoods.
- After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.
- Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward.

RESULTS

- The results from the k-means clustering show that we can categorize the neighbourhoods into 3
- clusters based on the frequency of occurrence for “Shopping Mall”:
- • Cluster 1: Neighbourhoods with moderate number of shopping malls
- • Cluster 2: Neighbourhoods with low number to no existence of shopping malls
- • Cluster 3: Neighbourhoods with high concentration of shopping malls
- The results of the clustering are visualized in the map below with cluster 1 in red colour, cluster 2 in
- purple colour, and cluster 3 in mint green colour.

RESULTS


Tanjung Karang
Bukit Beruntung
Genting Highlands
Karak
Bestari Jaya
Serendah
Bukit Tinggi
Kuala Selangor
Rawang
Ijok
Kuang
Selayang
Batu Caves
Puncak Alam
Sungai Buloh
Hulu Kelang
Kapar
Meru
Ampang Jaya
Batu 14 Hulu Langat
Petaling Jaya
Shah Alam
Subang Jaya
Klang
Seri Kembangan
Puchong
Pulau Ketam
Pulau Tengah
Pulau Klang
Pelabuhan Klang
Kajang
Bandar Baru Bangi
Semenyih
Pulau Indah
Cyberjaya
Putrajaya
Kuala Klawang
Simpang Perta
Teluk Panglima Garang
Pulau Carey
Jenjarom
Bangi Lama
Dengkil
Lenggeng

THANK YOU!!