COURSEA IBM DATA SCIENCE CAPSTONE

Where to Set Up Mobile Towers for 5G?

By

B M AZHAR HUSSAIN

PROBLEM DESCRIPTION AND BACKGROUND

Edotco Group is a Mobile Tower Company operating in 8 countries in Asia. Aspiring to be one of the top TowerCos in the world, Like all other tower companies, edotco chooses the best locations for building its mobile towers and rents them Mobile and internet service providers. The bigger the density of an area, the greater the need for data consumption, resulting in higher rental revenue.

edotco plans to have massive expansions of it's mobile tower portfolio and provide greater connectivity over the next few years. Having its head office in Malaysia, edotco is working really close with the local government in implementing the 5G campaign. Malaysia plans to be 5G equipped by 2022 and edotco will be playing a vital role by providing infrastructure support.5G services enable users with really high speed internet and the network setup requires data Transmission equipment to be in very close approximation. The mobile operators are targeting fully enabled 5G services in major public areas to start off with. Places where people would gather the most and require fast data services. Edotco is planning to carefully choose locations to set up it's 5G equipment so that high rental income can be charged.

The objective of this project is to figure out the best locations for Edotco Tower Company Ltd to deploy its 5G towers in Kuala Lampur. Using Data science techniques like clustering, this project can guide Edotco Management committee to decide on which locations to choose for 5G investment. Particularly targeting the shopping centres around kuala lampur,

DATA USED AND SUMMARY METHODOLOGY

The following data will be used to solve this problem:

1. List of neighborhoods in kuala lampur,
2. Location coordinates of these neighbourhoods
3. Venue data, particularly shopping centres.

Source of data will be wikipedia .
https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur lists the neighbourhood in Kuala Lampur. Following this we will log in to our Foursquare API to gain information regarding shopping centres and restaurants to help us solve our queries.
First We will use scraping methods to gain data from this page. Then find out the latitude and longitude coordinates of these neighbourhoods using Python Geocoder package. After this, we shall be using various data science methods such as data cleaning, wrangling, k-means clustering, map visualization etc to come to a conclusion. Detailed methodology will be explained in week 2 section.

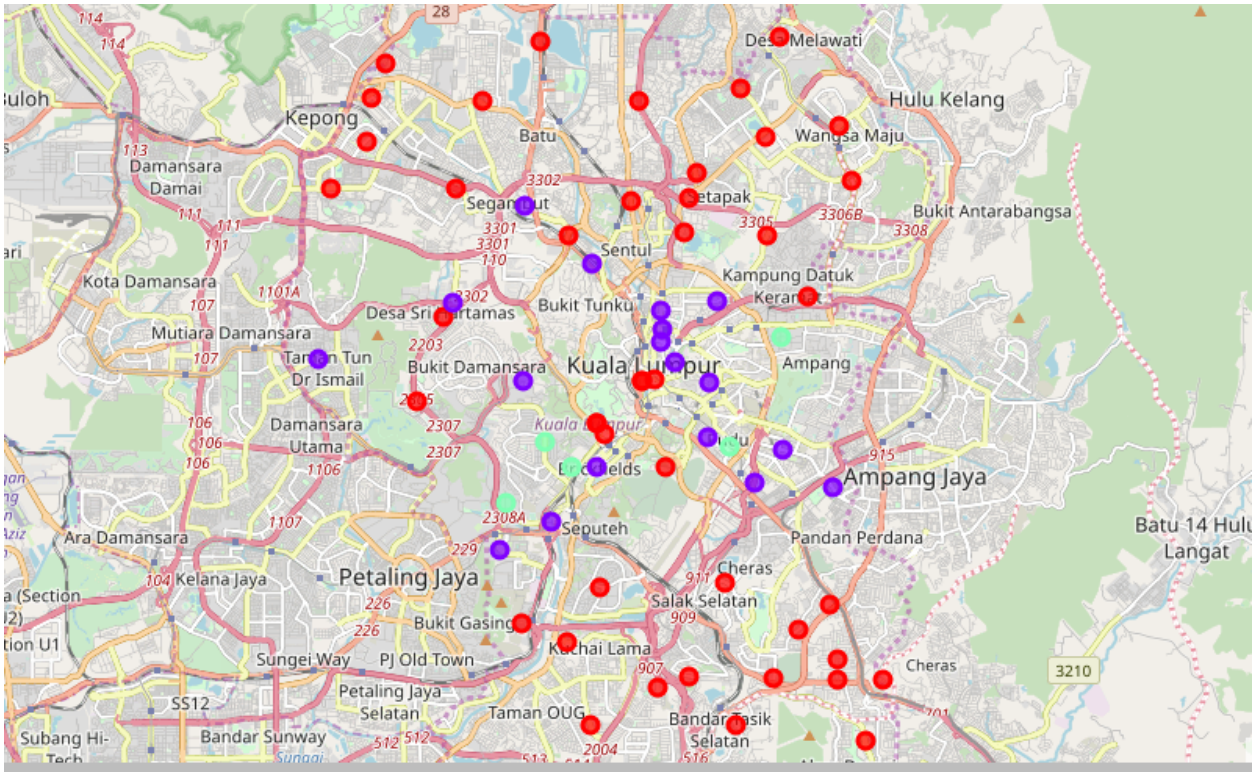
FULL METHODOLOGY

1. Find out list of neighbourhood from reliable source:
   We can find this list on wikipedia:
   https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur
2. To get the names from this website, we use beautifulsoup packages on python.
3. Use geocoder package that will allow this list of names to be converted into geographical coordinates.
4. Visualize the neighbourhood in a map using Folium Package.
5. Register a Foursquare Developer Account
6. Make API calls to Foursquare passing the coordinates of the neighbourhoods.
7. From the results our data will be prepared for clustering , we will filter “Shopping malls” as venue category.
8. Next, we perform clustering by using k-means method and cluster the neighbourhoods into three categories based on which areas have higher concentration of shopping malls. THIS will help us answer which areas are most suitable for deploying mobile towers.

RESULTS

By implementing the K means clustering, we divide the data into 3 categories.

Cluster 0: Neighbourhoods with moderate numbers of Shopping malls (RED)
Cluster 1: Neighbourhoods with low numbers of Shopping malls (PURPLE)
Cluster 2: Neighbourhoods with high concentration of Shopping malls (GREEN)



We can see from the map above that from the shopping centres in Kuala Lampur, highest concentration is in cluster 2, cluster 0 is moderate and cluster 1 has the lowest concentration of shopping centres. This map also suggests the central kuala lampur has ample number of shopping centres for the management to focus on. Nearby suburbs having high concentration of shopping centres is also a very feasible option as this represents local shopping centres with local foot traffic potential.

If the management decides to invest their capital in deploying towers at cluster 0 and 2, they are likely to cover most of the urban areas surrounding KL and might leave the cluster 1 for later expansion.

Point to be noted that this analysis does not take into consideration any other factors such as which type of demographics may include in high foot traffic areas (shopping malls), their age, income level, need for data consumption, high foot traffic times etc. This is only dependent on the hypothesis that higher concentration of shopping malls means higher foot traffic and higher demand for data from the public.

In conclusion, this project gives a justification for edotco management to take decision on picking cluster area 0 and 2 and the suburbs included in them for potential 5G areas.