IBM Applied Data Science Capstone

Final Project

The Gym and Fitness Centres in Metropolis

By Pratyathawi Tripanya
November 2019

Introduction

Nowadays, with Globalization and Urbanization's impacts, many people have to commute between houses and workplaces; almost of them work in the city. So, Traffic congestion becomes one of the most serious problems in many mega cities, without an exception of Kuala Lumpur (the Capital city of Malaysia and one of the mega cities of Asia Pacific Region). However, there is a trend that called "Work Life Balance" spread in many metropolises. The trend which influence people to eat good foods, travel on vacations, and have workouts. Therefore, the trend of "Work Life Balance" especially in workouts and exercises has inspired me to do this final project.

Business Problem

Since the traffic congestion in Kuala Lumpur causes the difficulty of commuting between residences and work offices. It is hard to find the optimal solution for going to gym and fitness centres. In this project, I will try to find an optimal location for a gym and fitness centre in the capital city of Malaysia, Kuala Lumpur. As we can all assume, KL has a lot of fitness centres. I am planning on opening a fitness centre, in the area that suits for metropolitans and without any of them nearby. I will try to find a place where there are fewer of them. Also, that location has to be popular for workers, because I want to solve the pain point of workers who find "Work Life Balance" solution. Hopefully, by applying the knowledge I've gained during this course I will try it by my best.

Data

Based on definition of our problem, I need the following data:

- List of neighbourhoods of the capital city of Malaysia, Kuala Lumpur.
- Latitude and Longitude coordinates of those neighbourhoods.
- Number of existing gym and fitness centres in the neighbourhoods of Kuala Lumpur.

Following data sources will be needed to extract the required information:

- List of neighbourhoods in Kuala Lumpur will be obtained on "https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur".
- Number of gym and fitness centres and locations in every neighbourhoods will be obtained using **Foursquare API.**

Methodology

After extracting data of neighbourhoods in Kuala Lumpur from Wikipedia by web scraping techniques with the help of Python requests and Beautiful Soup packages. Then I obtained the geographical coordinates of the neighbourhoods using Python geocoder package that provided me a set of coordinates, including latitude and longitude, of the area. This was followed by getting venues data for those neighbourhoods by Foursquare API. The application which contains the large database 100 million places provided me many categories of the venue data which I am particularly interested in the gym and fitness centre category. After gathering the data, I populated the data into a Pandas DataFrame and then visualized the neighbourhoods in a map using Folium package and returned me with Geocoder's plotted city map of Kuala Lumpur.

Next, I got the data of top 100 venues that are within a radius of 2000 metres from Foursquare API by making API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in Python loop. Then the API returned the venue data in JSON format and followed by grouping those venues. By doing so, I was also preparing the data for use in clustering. As I was analysing the "Gym and Fitness Centres" data. So, I filtered the "Gym and Fitness Centres" as venue category for the neighbourhoods.

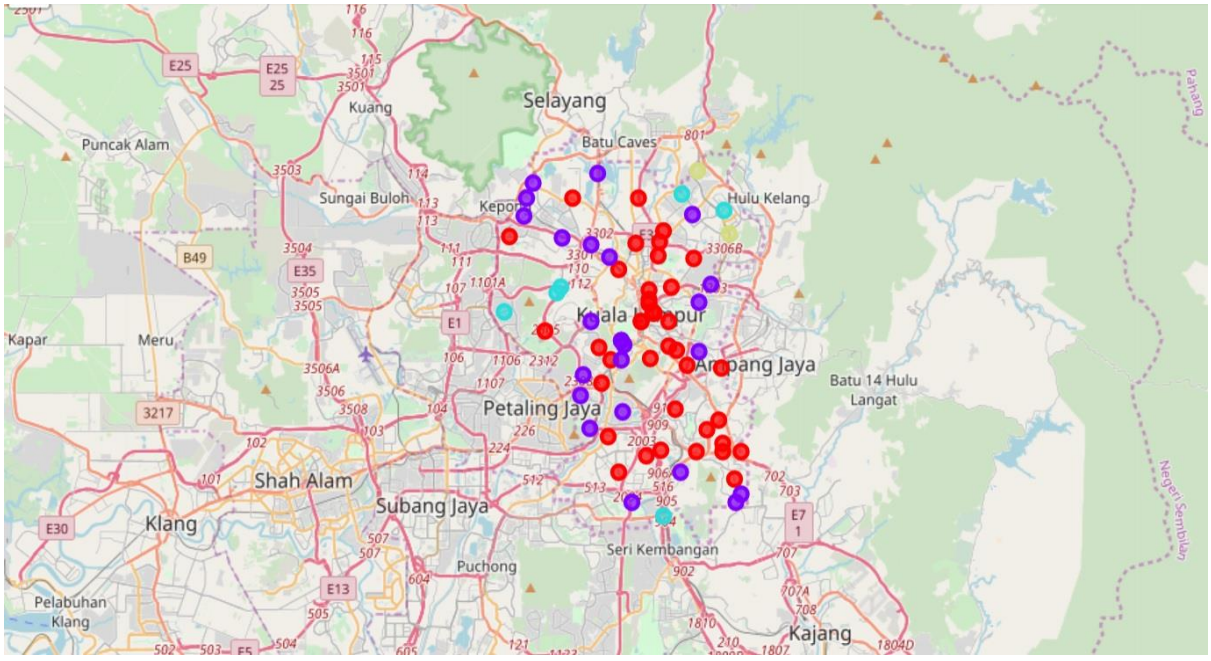
Finally, I did the clustering process by using K-means clustering method. K-means clustering algorithm identified k number of centroids, and then allocates every data point to the nearest cluster i.e. in my project, the number is 4, while keeping the centroids as small as possible. The results will allow us to identify which neighbourhoods have high or low concentration of the gym and fitness centres.

Result and Discussion

The results from K-means clustering method showed 4 clusters which are based on the frequency of occurrence for “The Gym and Fitness Centres”

- Cluster 0: no existence of fitness centres
- Cluster 1: low existence of fitness centres
- Cluster 2: moderate existence of fitness centres
- Cluster 3: high existence of fitness centres

In the visualized map, there are representative dots by four colours naming; red colour for cluster 0, purple colour for cluster 1, blue colour for cluster 2, and yellow colour for cluster 3.



As can be seen, there are a few areas that have high concentration of gym and fitness centres. Those areas are suburb areas for residency of labours in the city. However, there are many areas which have nothing or low concentration of gym and fitness centre. So, these are business opportunities that can solve the business problem or pain point which I mentioned in the introduction.

Conclusion

Main goal of this project was to identify optimal places for opening a new gym and fitness centres in Kuala Lumpur that already has a huge amount of the gym. Throughout this project I was exploring best possible values and finally came to the conclusion and found many places especially in the central of the city that filled all the requirements.

Final decision on optimal fitness centre location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone. Lastly, I think this project will solve the problem of the people who want to make “Work Life Balance”.