1.Introduction

1.1. Business Problem

An investor is looking to open a new gym in Kuala Lumpur. Based on his previous experiences and marketing strategy, he would like to tap-into a mature neighbourhood with high traffic, but low competition. He has listed down the area selection criteria as shown below:

* Area that has hotels or shopping mall or residential (apartments or condo) in vicinity.

* Area that is not already crowded with gyms

To solve this problem, this project is initiated. Data science approach is utilized here to answer the investor's question, which is to locate a potential neighbourhood in Kuala Lumpur.

1.2. Stakeholders

Besides the investor who initiate this project, there are some other stakeholders who might be interested in this project:

- Those that are interested in knowing high traffic attraction areas in Kuala Lumpur
- Those that are interested in knowing the area with sports amenities in Kuala Lumpur
- Those that have interest in using the exploration result in this project

2.Data

Based on the business problem, we would require the following data in this project.

- Neighbourhood list of Kuala Lumpur
- Coordinates of each neighbourhood
- Venues in each neighbourhood

2.1. Neighbourhood list of Kuala Lumpur

First, we gather the list of neighbourhoods in Kuala Lumpur. This information is available at the link below. There are 71 neighbourhoods in Kuala Lumpur. Web scraping technique is applied here to capture these information.

https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur

	0
0	Alam Damai
1	Ampang, Kuala Lumpur
2	Bandar Menjalara
3	Bandar Sri Permaisuri
4	Bandar Tasik Selatan

Figure 1: First 5 rows of KL neighbourhood list

2.2. Coordinates of each neighbourhood

After getting the list of neighbourhoods in Kuala Lumpur, the next step will be to get the coordinates for each of the neighbourhood. This process is realized by geocoding library of Python.

	Neighborhood	Latitude	Longitude
0	Alam Damai	3.057690	101.743880
1	Ampang, Kuala Lumpur	3.148494	101.696729
2	Bandar Menjalara	3.190350	101.625450
3	Bandar Sri Permaisuri	3.103910	101.712260
4	Bandar Tasik Selatan	3.072750	101.714610

Figure 2: First 5 rows of KL neighbourhood list with geo coordinates

2.3. Venues in each neighbourhood

Foursquare location data is utilized to get the list of venues within 3km of radius in each neighbourhood. The information of geo-coordinates collected from previous step will be used to explore venues in each neighbourhood in Foursquare.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Main Category
0	Alam Damai	3.05769	101.74388	Pengedar Shaklee Kuala Lumpur	3.061235	101.740696	Supplement Shop	Supplement Shop
1	Alam Damai	3.05769	101.74388	Jc Deli 皆喜食坊	3.058397	101.748560	Food & Drink Shop	Food & Drink
2	Alam Damai	3.05769	101.74388	Machi Noodle 妈子面	3.057695	101.746635	Noodle House	Noodles
3	Alam Damai	3.05769	101.74388	628火焰蠡茶室	3.058442	101.747947	Chinese Restaurant	Chinese
4	Alam Damai	3.05769	101.74388	Minang Tomyam	3.057185	101.749812	Seafood Restaurant	Seafood

Figure 3: First 5 rows of venues list in each neighbourhood

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Main Category
Alam Damai	76	76	76	76	76	76	76
Ampang, Kuala Lumpur	100	100	100	100	100	100	100
Bandar Menjalara	100	100	100	100	100	100	100
Bandar Sri Permaisuri	100	100	100	100	100	100	100
Bandar Tasik Selatan	100	100	100	100	100	100	100
...	...	...	...	...	...	...	...
Taman Tun Dr Ismail	100	100	100	100	100	100	100
Taman U-Thant	100	100	100	100	100	100	100
Taman Wahyu	47	47	47	47	47	47	47
Titiwangsa	81	81	81	81	81	81	81
Wangsa Maju	100	100	100	100	100	100	100

71 rows × 7 columns

Figure 4: Count of Venues in Each Neighborhood