Identifying Ideal Districts in Malaysia for Establishing New Gym Branches Using K-means Clustering

Part 1: A description of the problem and a discussion of the background

Introduction

Malaysia has the highest rate of obesity and overweight among Asian countries with 64% of the male and 65% of the female population being either obese or overweight. [1] The Malaysian government has also taken many initiatives in the past several years to promote healthier lifestyle and living among Malaysians. As a result, there is increasing awareness about fitness among the younger generation of the country. The health and fitness industry in Malaysia are booming as of now. Currently, there is an increasing demand for more gyms and fitness centres in Malaysia.

Realising the market potential of building a successful business in the fitness sector in Malaysia, a hypothetical fitness-based company called "A Fitness" is interested in establishing a gym chain with branches all over the Peninsular of Malaysia. The company is looking to open their gym branches at locations which would attract a huge customer base and guarantee them a regular profit.

Problem statement

The goal of this project is to help the company identify the most ideal locations to open their gym centres. The problem statement is, "Which are the most suitable districts to open new gym centres in the Peninsular of Malaysia?

Discussion of background

It is important to narrow down the scope of the project and to take certain important criteria into consideration. Important aspects to consider are for example:

i) the location area,
ii) the population of the area

iii) the general age of residents in the particular area,
iv) average income of resident in the area,
v) number of gyms or fitness centres already available in the area
vi) number of shopping malls in the area

a) Location

PERLIS
NARATHIWAT
KEDAH
KELANTAN
Sungai Petani
PENANG
TERENGGANU
Taiping
Kuala Dungun
PERAK
Taman Negara
PAHANG
Malaysia
Pekan
Kuala Lumpur
PUTRAJAYA
NEGERI SEMBILAN
MALACCA
Dumai
JOHOR

Image 1 [2]: Map of Peninsular Malaysia

Only the states in the Peninsular of Malaysia are taken into account in this project. The two states in east Malaysia are excluded in this project. There is a total of 11 states and up to 87 districts in Peninsular Malaysia. From the 11 states, the state of Kuala Lumpur (the capital of Malaysia) is excluded also as it already has many gym centres and it would be difficult to make an impact in an already competitive location. It would not make much business sense to open more gyms in an already saturated location. The Klang Valley consists of Kuala Lumpur and areas close to the capital, bordering Selangor.

As of now, the focus of the project would be on the main districts in the states of Perak, Terengganu, Malacca, Johor, Negeri Sembilan, Pahang, Penang, Kedah, Kelantan, and Selangor.

b) Population

The total population of Malaysia in 2020 is estimated to be 32.7 million people [3]. Peninsular Malaysia accounts for the majority (roughly 81.3%) of Malaysia's population [4]. The state of Selangor is the most populous state in Malaysia, while Penang has the highest population density. In general, urbanisation is more prominent in areas with high population and more shopping malls are available at these locations.

The company prefers to open their gym centres at highly populated locations.

c) Age

The target customer base for the gym centres are people within the ages of 20 to 50 years old. It is assumed that Malaysians in their 30's and 40's are generally more health conscious compared to Malaysian of other age groups. Malaysian above 50's generally prefers more lighter forms of exercise such as brisk walking or jogging and may prefer going to parks instead.

Districts with higher population of people in their 30's and 40's are preferred by the company.

d) Income

Standard gym membership fees of gyms in Klang Valley is within MYR 90 to MYR 300 per month [5]. In 2019, the mean income in Malaysia was RM7,901 while Malaysia's median income was RM5,873 [6]. Median monthly household disposable income was RM5,116 in 2019, and mean monthly household disposable income was RM6,764. Mean disposable income comprises of 85.6 per cent of total mean gross income. [6]. People living in urban districts earn more than those who live in rural districts.

Taking into consideration the general cost of living of a regular Malaysian, those who earn lesser then MYR 3000 per household are unlikely to spend on gym membership fees. The target customers are people who earn MYR 4000 and above. It is also assumed that people with higher income would probably be willing to spend more on exclusive gym membership deals.

e) Gyms and Fitness centres

There are already a few big gym chains in Malaysia as of now and most of them are located in Kuala Lumpur or areas in the Klang Valley. It is preferable to open gyms at districts with little to no gym centres currently available. The competition for market share would be very high in areas with many other gym and fitness centres. It would be very difficult to make a profit in a highly competitive area as customers would prefer to frequent gym which charge the lowest possible gym membership fees.

f) Shopping malls

Most of the gyms in the Klang Valley for example, are located in shopping malls or very close to shopping malls. As many people visit shopping malls on a regular basis, it would make sense to open gyms in such locations. The company "A Fitness" prefers areas with a high number of shopping malls.

Part 2: A description of the data and how it will be used to solve the problem

Description of Data

The main data required for this project are as follows:

i) the location area
ii) the population of the area
iii) the main age of residents in the particular area,
iv) average income of resident in the area,
v) number of gyms or fitness centres already available in the area
vi) number of shopping malls in the area

a) Location Data

The list of districts in each states will be extracted by web scraping the page, https://en.wikipedia.org/wiki/List_of_districts_in_Malaysia [7] using the BeautifulSoup library.

The GeoPy library will be used to obtain the latitudes and longitudes of each district in each state. These values will be stored into a pandas dataframe. Folium will be used to create a map of these location points.

b) Population Data

As there is no one specific website to obtain all the population information of each district, various websites had to be search to obtain the necessary data. The population values are stored into a pandas dataframe.

c) Age of Residents

The main age group of residents of each district will be stored into a dataframe. The age group of residents are divided into several categories such as 20 to 24 years, 25 to 29 years, 30 to 34 years, just to name a few. Since there were no specific data for each district, the age data of each state will be used instead for this category.

d) Income of residents

The average household income of residents of each district will be stored into a dataframe. Since there were no specific data for each district, the median and mean household income data of each state will be used instead for this category. The mean and median household income of each state can be obtained from the Department of Statistics Malaysia Official Portal.

e) Number of gyms in an area

The Foursquare API will be used to collect information on the exact location and number of gyms or fitness centres available in each district. The necessary information will be extracted into a dataframe.

f) Number of shopping malls

The Foursquare API will be used to collect information on the exact location and number shopping malls available in each district. The necessary information will be extracted into a dataframe.

Using data to solve the problem

Preparing and pre-processing dataset

The respective dataframes will be merged together to form one main dataframe with all the information necessary. One-hot-encoding will be used on categorical data such as age group. The overall finalised dataset will then be standardized.

Clustering and Segmenting Districts

K-means clustering algorithm will be used to cluster districts with similar characteristics. The districts will be divided into 4 clusters. The goal of the project is to identify the specific cluster of districts which fulfils the criteria set by the company “A Fitness” such as districts with high population, high income, general age group of 30s to 40s, low number of existing gyms and high number of shopping malls. The districts which fulfils the criteria will be the most ideal locations for the company to establish their gym branches.

Mapping clusters

A map will be created using Folium to visualise the respective clusters.

References

[1] Who.int. 2020. *Malaysia And WHO Call For More Investment In Primary Health Care The 21St Century*. [online]

Available at: <https://www.who.int/malaysia/news/detail/08-04-2019-malaysia-and-who-call-for-more-investment-in-primary-health-care-the-21st-century#:~:text=Malaysia%20has%20the%20highest%20rate,being%20either%20obese%20or%20overweight.> [Accessed 6 December 2020].

[2] Maps, M., 2020. *Malaysia Beaches And Islands (Peninsular Malaysia) – Google My Maps*. [online] Google My Maps.

Available at: <https://www.google.com/maps/d/viewer?mid=1Tp3JDwPo_MZ-JH-_diEytGMOKq4&ie=UTF8&source=embed&oe=UTF8&msa=0&ll=4.289239239796553%2C102.854948140625&z=7> [Accessed 6 December 2020].

[3] Dosm.gov.my. 2020. *Department Of Statistics Malaysia Official Portal*. [online]

Available at: <https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul_id=OVByWjg5YkQ3MWFZRTN5bDJiaEVhZz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09#:~:text=Malaysia's%20population%20in%202020%20is,to%203.0%20million%20(2020).> [Accessed 6 December 2020].

[4] En.wikipedia.org. 2020. *Peninsular Malaysia*. [online]

Available at: <https://en.wikipedia.org/wiki/Peninsular_Malaysia#:~:text=Peninsular%20Malaysia%20accounts%20for%20the,92%25%20of%20total%20population).> [Accessed 6 December 2020].

[5] Ho, F., 2020. *How Much Do Malaysians Need To Spend To Get Fit In 2016?*. [online] iMoney Malaysia. Available at: <https://www.imoney.my/articles/how-much-do-malaysians-need-to-spend-to-get-fit-in-2016> [Accessed 6 December 2020].

[6] Dosm.gov.my. 2020. *Department Of Statistics Malaysia Official Portal*. [online]

Available at: <https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=120&bul_id=TU00TmRhQ1N5TUxHVWN0T2VjbXJYZz09&menu_id=amVoWU54UTl0a21NWmdhMjFMMWcyZz09> [Accessed 6 December 2020].

[7] En.wikipedia.org. 2020. *List Of Districts In Malaysia*. [online]

Available at: <https://en.wikipedia.org/wiki/List_of_districts_in_Malaysia> [Accessed 6 December 2020].