IBM Data Science Professional Certificate

Capstone Project

Choosing the Best Location to Open a Gym in Toronto, Canada

Presented by: Zhu Moran

Nov 2020

Project introduction

Nowadays, personal health is becoming a major problem which is ignored by many in the fast-paced society. People are putting a lot of attention in work, study and entertainment while neglecting the importance of having a fit body. Gyms, are becoming more and more popular for people who want to exercise and maintain good personal health. We can see that there are more and more gyms open in CBD areas of big cities. Opening a gym is also considered a good idea for people who want to start their own business. However, the location of gym needs to be carefully examined and it could be very crucial to success of the business.

Business Problem Briefing

The objective of the project is to analyse the neighbourhoods in Toronto and select a best location to open a new gym. The relevant data science methodologies like data cleaning, data sorting and data analysing will be applied. Machine learning techniques like clustering analysis will also be performed. The question to be answered is: if someone or some group wishes to open a new gym in Toronto, where is the best location?

Targeted people

The targeted people or targeted audience will be any individual or groups of people who are interested in opening a new gym in Toronto but having difficulties in selecting the appropriate location.

Data

Firstly, the neighbourhood's data of Toronto could be obtained from Wikipedia page:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

A data frame will be created based on this data set and provide necessary information about neighbourhoods' info of Toronto.

Next, Foursquare API will be used to get the venue data about these neighbourhoods. Among all the available data, we will be interested in data with relevance to gym. Relevant data analysis techniques will be performed to clean and work with the data. Machine learning using K-means clustering will be performed to group the clusters

Lastly, a map with identified clusters will be provided and indicate the suitable locations to open a gym.