# IBM Data Science Professional Certificate Capstone Project

# Analyzing And Clustering Toronto Neighborhoods To Find The Best Place Gym/Fitness Center

# By: Mohammad Odeh

# Introduction

Toronto is the largest city in Canada, the capital of Ontario, and the home to more than five million people, this gives it a big advantage for businessmen and investors who want to establish a profitable business.

A business man is interested in opening a gym in the city of Toronto and he is looking for the best place for the business that will lead to more customers, so the objective of this project is do full analysis of the neighborhoods in Toronto by clustering them based on similar businesses (venues), this analysis will help locate areas with less gyms which indicate a great opportunity to open a gym and provide such a service for residents of that area.

# Data

To find the solution and perform detailed analysis on the topic, we need data that contains neighborhoods and locations coordinates of the city of Toronto, and we need data about the available gyms centers ( venues ) in each neighborhood, there for we must:

1. Web scrape the lists of neighborhoods from Wikipedia page :

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

A screenshot of a cell phone

Description automatically generated

1. Get the coordinates of each neighborhood using a csv file contains all coordinates of Toronto neighborhoods ( <http://cocl.us/Geospatial_data> )

A screenshot of a cell phone

Description automatically generated

1. Get the list of available gyms in each neighborhood by using Foursquare API

A screenshot of a computer

Description automatically generated

# Methodology

For Finding the best location for opening a new gym, a clustering algorithm (K-Means clustering) is applied which will cluster the neighborhoods into clusters depending on the number of gyms in that neighborhood.

The data that will be used for that is the cleaned/edited list of neighborhoods with their corresponding coordinates and number of gyms on each one :

A screenshot of a cell phone

Description automatically generated

Next in order to find the optimal K ( number of clusters ), elbow method is used by applying k means for different sets of “ K “, then the sum of square distance ( Inertia ) is calculated for each “ K “

A picture containing screenshot

Description automatically generated

So, K=3 is the optimal number of clusters, there for we apply the algorithm to the data with k=3

# Result

The resulted clusters showed in the map below, show that there is a concentration of gyms on a specific area while there is a lot of neighborhoods that could be a potential place for opening a gym on.

A close up of a map

Description automatically generated

# Conclusion

The results showed that the heart of the gym business is concentrated near the coast and the more we go to the edge of the city the more opportunity there will be to open a gym that will serve the community and the people there.

With the results that shows the neighborhoods and places where it has no gyms, and with further analysis using the population and demographic distribution data among the neighborhoods, we can determine which of these areas are more profitable and will serve more people.