

## **Applied Data Science Capstone Week 4**

### **Introduction**

Toronto is a fast-growing city in Canada, more and more people plan to start their own business there. Hypothetically, there might not so many gyms in Toronto, because working out is a very important way for everyone to keep fit in current society, therefore, an entrepreneur probably wants to take the good opportunity to start his own gym in Toronto. For the entrepreneur, finding a suitable location to open a gym is a big decision. The purpose of this project is to help people who would like to open a Gym in Toronto Area. It will give people some useful suggestions to make an efficient decision on selecting a good location to open a gym. In this project, some data science methods, tools and machine learning algorithms will be used. The project aims to create analysis and provide solutions for the business problem: where is the most suitable location to open a gym for an entrepreneur in the Toronto area?

### **Data Section**

To analyze and solve the business problem, the data we will use is as follows,

1. [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
2. [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)
3. Foursquare API Data

We will use the data from the Wikipedia website in week three and also the data from the Foursquare API. Specifically, we need to have the list of neighborhoods in the Toronto area and the coordinates of these neighborhoods, we should also find the venues which are related to Gym. To acquire the data, web scraping, Geocoder package, Foursquare API, and some other methods will help us to extract these useful data in this project.