**Gym Data Analytics For Business in Toronto, York Area.**

Elena Grishin

November 5, 2020

1. **Introduction**
   1. **Background**

According to Statistics Canada, the population figure for the York Region in 2019 is 1,191,400. It was constantly growing for the last several years averaging 1.9 percent per year since 2011. The idea behind this project is as the population is growing there are not enough gyms in that region. So, this project is designed to help a group of stakeholders to select a good location to open a gym.

* 1. **Problem**

The York region neighbourhood's data that might be used to find a suitable location to open a gym. Using data science methods such as clustering, this project aims to answers a business question of where in Toronto, York Area stakeholders should open a gym.

1. **Data**

To accomplish the task we will need two data sets merged into one.

First, is a list of postal codes that can be obtained from the Wikipedia page.

* Wikipedia page data source

https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M

Second, is the data set of latitude and longitude of each neighbourhood.

* Latitude and Longitude data source

<http://cocl.us/Geospatial_data>

Data needs to be cleaned to exclude neighbourhoods that have no assigned values.

Also, data needs to be simplified to include only data for Toronto, York region.