**Capstone Project - The Battle of Neighborhoods**

**Ramilya Fazlieva**

1. **Introduction**

This project is a kind of summary of the learned material covered in IBM course of data science. In this work, we used the algorithm – K-means clustering - applied to the neighborhoods in Toronto.

The objective of this Capstone is to learn about segmenting and k-means clustering, to understand how to use the Github platform, to learn the Foursquare API and to try to show the results of studying by publishing the final project on Github. The main goal is to understand the practical side of the application of techniques and algorithms of data science in real life tasks.

1. **Data**

The data has been collected from Wikipedia page, <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> .

This html table was converted to Pandas DataFrame for cleaning and preprocessing. We have dropped the rows where Borough is 'Not assigned' and combined the neighbourhoods with same Postal code.

Then we have imported the csv file conatining the latitudes and longitudes for various neighbourhoods in Canada and merged two tables.

We will analyze all the rows from the data frame which contains Toronto in their Borough.