**Capstone Final Report – Mikel Brabec**

# Description of the problem and discussion of background

In this assignment, I had to explore, segment, and cluster the neighborhoods in the city of Toronto. However, unlike a previous case we did about New York, the neighborhood data is not readily available on the internet. What is interesting about the field of data science is that each project can be challenging in its unique way, so you need to learn to be agile and refine the skill to learn new libraries and tools quickly depending on the project.

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. You will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format like the New York dataset.

Once the data is in a structured format, you can replicate the analysis that we did to the New York City dataset to explore and cluster the neighborhoods in the city of Toronto.

The target audience could be residents that seek new neighborhoods to live in that are similar to their current area. Or restaurants and other shops seeking for new business opportunities in other neighborhoods.

# Description of the data and how it was used to solve the problem

The list that was used was provided by the Canadian Government and can be found on Wikipedia through this link:

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

It is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

Canada Post provides a free postal code look-up tool on its website, via its applications for such smartphones as the iPhone and BlackBerry, and sells hard-copy directories and CD-ROMs. Many vendors also sell validation tools, which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Special features about the data are the three main colums PostalCode, Neighborhood and Borough. This data was then cleaned, restructured and graphed to show the three columns.