**The Battle of the Neighbourhoods**

1. **Introduction**

Mobile devices have been widely adopted over the last two decades. They are powerful tools for understanding the aggregated behaviour of consumers. Understanding device location opens doors to a wide range of use cases that are unique in many different ways. The location data provided is used to create an accurate representation of your interests, and this can be used to bring more targeted and relevant ads to potential consumers or being able to find a location of interest.

Foursquare is a technology company like Google places and Yelp that has built a massive dataset of accurate location data. It powers location data from Apple’s Maps, Uber, Snapchat, Twitter and others. Their API and location dataset are currently being used by thousands of developers. The Foursquare website also provides a platform for users to comment and rank restaurants, cafés, mall, parks and others.

In this study, we will apply Machine Learning techniques on location data obtained from Foursquare to solve the following problem:

Suppose a user has recently gotten a job in New York City, USA and you are required to move to the Big Apple as you currently live in Toronto, Canada. You are looking to find an area to live that bares some similarity to where you currently live.

We can therefore suggest similar neighbourhoods in New York that suit your preferences with the use of cluster analysis and Foursquare location data by comparing the two cities and determining how similar or dissimilar they are.

1. **Data**

For this study, the Foursquare API will be used to gather information about neighbourhoods located in both New York City and Toronto. Moreover, their respective latitude and longitude coordinates are merged to a dataset of the listed neighbourhoods. We focus particularly on neighbourhoods in the Downtown Toronto borough and compare that to neighbourhoods solely in Manhattan, New York. A Foursquare API GET request is sent in order to acquire the surrounds venues that are within a radius of 500m. The data is formatted using one hot encoding with the categories of each venue. Then, the venues are grouped by neighbourhoods computing the mean of each feature.

We will determine the similarities based on the frequency of the categories found in the neighbourhoods. The similarities found will give the user an idea of which areas would suit their preference and further help the user decide whether to move in a particular neighbourhood in Manhattan or not.