Introduction: Business Problem

In this project we will **compare the Neighbourhoods of the two cities** and determine how **similar** or **dissimilar** they are.

We will retrieve the top 100 venues that are in the Neighbourhoods of a city within a radius of 500 meters. We will explore, analyse and cluster the Neighbourhoods of the two cities based on their latitude and longitude. The top ten venues around the Neighbourhood would be used to compare the cities.

New York and Toronto are very diverse and are the financial capitals of their respective countries. We will use data science to determine how similar or dissimilar the Neighborhoods are of the two cities New York and Toronto and come up with findings.

Data

Based on definition of our problem, factors that will influence our decision are:

- Venues of any type in the Neighbourhood
- Top ten most common Venues

Following data sources will be needed to extract/generate the required information:

- The dataset containing the Borough, Neighbourhood with Latitude and Longitude coordinates for each city
- The dataset containing the Borough, Neighbourhood with Latitude and Longitude for Toronto is not readily available, but can be obtained by scrapping the wiki page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,
- The dataset containing the Borough, Neighbourhood with Latitude and Longitude coordinates for New York city is available for free from the website https://geo.nyu.edu/catalog/nyu_2451_34572
- The venues in every Neighbourhood will be obtained using **Foursquare API**