# **The battle of the neighborhoods Toronto vs New york**

**Problem statement**

A major real estate company is looking at projects in North America and looking to build a 500-unit apartment complex in either New Yok or Toronto. The senior manager of the project management team approaches the data scientist to do analysis to identify which City the apartment should be built based on the surrounding venues and amenities. The areas identified which have the most promise are Queens in NY and Scarborough in Toronto. However, the project manager wants the data scientist to prove that these two are the best boroughs in both respective cities and which one should be chosen for the project. This analysis will focus on the best place to build an apartment complex based on the surrounding amenities. The data scientist will then present the findings at a conference to both senior leadership and the project manager – to inform the decision on which city they would be spending millions of dollars on this project.

**Data**

The data that will be used for this project

* Toronto data - https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M
* New York data - <https://geo.nyu.edu/catalog/nyu_2451_34572>
* Foursquare API data
* Longitude and Latitude - http://cocl.us/Geospatial\_data

We will use the above data sources to extract data for our model of comparison. The raw data of the boroughs in both New York and Toronto will be extracted from Wikipedia and the public NYU catalog. We will also use the Foursquare API to get valuable locations and venue information about the surrounding amenities for each of the boroughs that we identify as suitable for this project.

The algorithm that we will use to compare the cities will be K means clustering to identify categories of places around the neighborhoods in question. Queens and Scarborough will be analyzed side by side and as comparatively to make sure a decision is made on the best area to build this apartment complex.