1. A description of the problem and a discussion of the background.

Toronto is the financial capital of Canada and one of the most vibrant and active cities in the world. People always compare it with Manhattan, NY and considers it the equivalent for Canada. Both cities multicultural and are known for their diversity, their cultural background and both cities have been considered prime targets for visitors and tourists.

I will use the data collected for both cities as well as the techniques acquired in this course to make a comparison of both cities based in data rather than perceptions.

The goal is to make a data-based comparison of the kind of venues a visitor can expect to find when visiting both cities and how approachable are these venues (distance-wise) for tourist that usually prefer to go around cities walking.

2. A description of the data and how it will be used to solve the problem.

The data needed to explore, segment, and cluster the neighborhoods in the city of Toronto is not available in the internet in a form ready to be used in data science.
I used the data scraped from the Wikipedia page
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for this analysis.

The data from the Wikipedia site will be completed with geolocation information contained in the link http://cocl.us/Geospatial_data/

For New York. The information for the city of New York and its 5 boroughs and 306 neighborhoods is readily available via the link https://geo.nyu.edu/catalog/nyu_2451_34572 .

This link contains a dataset with all the boroughs and neighborhoods of New York as well as their latitude and longitude.

The two sets of data with neighborhoods of Toronto and New York (Manhattan), will be completed with data collected from Foursquare.

Using Foursquares' API gives access to locations and venues near a particular address or city, recommendations on these venues and much more.