Capstone Project
The Battle of Neighborhoods
(Week 1)

2. Data

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

a) Data used for the project are as following:

- New York City dataset:
  https://geo.nyu.edu/catalog/nyu_2451_34572

- New York City latitude and longitude:
  Geolibrary

- Toronto Dataset (Toronto's zip codes, boroughs and neighborhood names):
  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

- Toronto latitude and longitude information:
  http://cocl.us/Geospatial_data

- Explore neighborhoods in Toronto and New York City:
  Foursquare API¶

b) The following tools were used:

- Pandas: Data Analysis
- NumPy: Data handling (vectors)
- JSON: Handling of JSON files
- Geopy: Retrieving location data
- Matplotlib: Data Visualization/Plotting
- Sklearn: Machine Learning; k-Mean calculation
- Folium: Map Rendering

c) Methodology used:

1. Data download
2. Data transformation
3. Adding latitude/longitude
4. Segmentation and Explore neighborhoods
5. Data analyzing
6. Clustering using k-means