

# The Battle of Neighborhoods

*By Judith Duiniveld*

*Toronto*



*New York*



# Data

---

- The data used for this project will be acquired from the respective cities Wikipedia website pages
- The datasets consists for each neighborhood of:
  - ◆ Postal codes
  - ◆ Neighborhoods names
  - ◆ Latitude and longitude information
- Foursquare API search feature will be used to collect neighborhood venue information
- In addition to Foursquare, various Python packages will be used to create maps and machine learning models to further provide insights

I used the following datasets from these websites:

1. Toronto Neighborhoods:  
[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
2. Toronto Latitude and Longitude:  
[http://cocl.us/geospatial\\_data](http://cocl.us/geospatial_data)
3. New York City neighborhoods:  
[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)
4. New York City Latitude and Longitude:  
Python Geolibrar

# Methodology

1. **HTTP request** → to pull the location information (latitude and longitude) to Foursquare API server using zip codes.
2. **Enabling Foursquare API Search future** → to collect the nearby places of the neighborhoods. Due to http request limitations *the number of places per neighborhood parameter* would reasonably be set to 100 and *the radius parameter* would be set to 700.
3. **Folium** → Python visualization library would be used to visualize the neighborhoods cluster distribution over an interactive leaflet map.
4. **Comparative analysis** → using Python's scientific libraries Pandas, NumPy and Scikit-learn so that two randomly picked neighborhoods would be carried out to derive the desirable insights from the outcomes
5. **Unsupervised machine learning algorithm K-mean clustering** → would be applied to form clusters of different categories of places residing in and around the neighborhoods. These clusters from each of two chosen neighborhoods would be analyzed individually to derive the conclusions.

# Used Python packages

1. Pandas → library for Data Analysis
2. NumPy → Library to handle data in a vectorized manner
3. JSON → Library to handles JSON files
4. Geopy → to retrieve location data
5. Requests → Library to handle http requests
6. Matplotlib → Python Plotting Module
7. Sklearn → Python machine learning library
8. Folium → Map rendering library

Pandas



Requests  
http for humans

matplotlib

Folium

