

DATA

The data scientist at DrivenByData has started assessing this business problem. He has concluded that the following data points will be required to provide a thorough analysis of both regions and recommendation to the client:

- Geo coordinates of both NYU and UofT
- Number of Art related venues around each of the NYU and UofT
- Distance of these venues from each of the universities
- Natural grouping of venue data on the basis of similarity

Data Description:

- FourSquare explorer API to search for Art related venues around NYU and UofT in their respective cities of New York City and Toronto.
- Publicly available dataset that contains neighbourhoods data of Manhattan (and New York City) at https://cocl.us/new_york_dataset
- Web-scraping of the wikipedia page on Toronto neighbourhoods: [https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'\)](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'))

Approach to solve the problem:

- Convert addresses into their equivalent latitude and longitude values.
- Use the Foursquare API to explore neighbourhoods in New York City and Toronto
- Use the Foursquare API explore function to get the most common venue categories in each city
- Group the neighbourhoods into clusters using k-means clustering algorithm
- Use Folium library to visualize neighbourhoods and their emerging clusters.