

IBM Coursera Data Science Specialization Capstone Project - Exploring neighbourhoods of New York and Toronto (Data section)

Data:

Data for the boroughs of New York will be taken from the Coursera skills lab as New York was one of the cities explored during the lab.

Data for the boroughs of Toronto will be taken from our previous project, the battle of the neighbourhoods, where the data was scraped from a Wikipedia page.

The Foursquare API will be utilised to explore both cities and find venues within a given radius. These venues will be mapped using Folium and then the neighbourhoods will be compared using clustering and segmentation by an unsupervised machine learning technique called kmeans clustering.

After clustering, the clusters will be explored to find similarities and/or to find the top venues per cluster as well as to determine what sort of area the neighbourhoods in the clusters belong to.