Problem Description:

Analyze the similarity of neighborhoods in New York and Toronto, based on commercial amenities available surrounding a specific address.

Background

Understanding a new urban location is a complex undertaking. Municipalities are differentiated by their size, the density of housing, the mix and locations of commercial and industrial sites, and by the demographics of the city’s population. The approach used here will be an analysis of the foursquare dataset using venue categories and number of different types of venues to define the nature of a neighborhood.

While the physical size of Toronto and New York is comparable, the population of New York is almost three times that of Toronto, implying higher population density, which might lead to higher numbers of venues in a given neighborhood’s area. Toronto covers an area of 243.3 square miles, or 630 square kilometers and has a population of 2.93 million residents. Manhattan covers an area of 22.7 square miles (59 km2).and has a population of 1.629 million residents. New York City, as a whole, covers 302.6 mi² (784 km2) and has a population of 8.623 million residents.

We can query data from foursquare to get venues surrounding a point. The project will conduct a search or searches for venues in a neighborhood, then use clustering tools to group neighborhoods in each city. Clustering neighborhoods across the two cities would allow us to determine if neighborhoods tend to more commonly cluster with others in their own city, or if there are clusters that have representatives in both municipalities.

These clusters will be compared and contrasted to determine if similar neighborhoods exist in both cities, and to understand the composition or characterization of neighborhoods that have matches in both regions.