Applied Data Science Capstone project

The Battle of Neighborhoods – Week 1 – Part 1

Introduction & Business problem

Problem Background

The City of New York, is the most populous city in the United States. New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centers. Its iconic sites include skyscrapers such as the Empire State Building and sprawling Central Park.

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's northwestern shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower.

New York & Toronto being the major cities of USA & CANADA, we will have lot of people migrating between these two places.

Problem description

As part of this project, we would like to compare neighborhoods in these two cities and help people in choosing similar neighborhoods when they migrate from New York to Toronto or Vice Versa. We will utilize the Foursquare API to explore the neighborhoods and segment them. We will use k-means to cluster the neighborhoods into 5 clusters.

We would be restricting the comparison to Manhattan Borough which is at the core of the New York City and the Boroughs which have Toronto in its name in Toronto to restrict number of api calls that we make using Foursquare api.

We will be performing the below steps as part of this project.

- Download and Explore Datasets for New York & Toronto cities
- Explore Neighborhoods in New York & Toronto cities
- Cluster Neighborhoods
- Examine Clusters

Note: It's straightforward to remove the condition and compare all the Boroughs in New York and Toronto.

Success Criteria

The success criteria of this project is to help end users identify similar neighborhoods in New York and Toronto cities and make informed choices when they migrate from New York to Toronto or Vice versa.