

Comparing Neighborhoods of two Major Cities in North America

Michael Obi

March 8, 2020

1. Introduction

1.1 In this project, I explore, segment, and cluster the neighborhoods in the city of Toronto in Canada and compare it to the City of New York in the United States of America. However, unlike New York, the neighborhood data of Toronto is not readily available on the internet.

1.2 Visitors and tourists across the cities would be very interested in knowing what similarities or differences exist and what to expect as they plan their travel for adaptability and cost control.

2. Data acquisition and cleaning

2.1 New York Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. The dataset will be taken from the web at https://geo.nyu.edu/catalog/nyu_2451_34572

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. I will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a *pandas* dataframe so that it is in a structured format like the New York dataset.

Once the data is in a structured format, I will perform the analysis of the two city (Toronto and New York) datasets to explore and compare the neighborhoods in the cities.

2.2 Addresses will be converted into their equivalent latitude and longitude values. Also, I will use the Foursquare API to explore neighborhoods in both Cities. I will use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. I will use the *k*-means clustering algorithm to complete this task.

Then, I will use the Folium library to visualize the neighborhoods in the Cities and their emerging clusters.