THE BATTLE OF NEIGHBORHOODS

IBM Applied Data Science Capstone Project

Pramudith Karunarathna

1.Introduction

Background

Newyork and Toronto are largest as well as financial and tourist capitals of the countries United state and Canada respectively.There are roughly about 8.39million residents in Newyork and 2.93 million residents in Toronto. Newyork and Toronto are both huge,diverse and cosmopolitan cities.When comes to similarities,both cities have a high cost of living ,both have a network of subway,trains and buses;both are extreme multi-ethnic and both are theater capitals of English speaking world.Considering Differences main difference is NYC has massive density whereas Toronto is much more spread out.Toronto has somewhat more of an Europhian vibe to it.It has very strong "britishness" to the names of Toronto districts,neighborhoods and streets.Yet another difference;A typical New York City watering hole is a place with a long bar and a baseball game on TV; a typical Toronto watering hole is a British/Irish-type pub with beer served in proper pints.Being nearest countries , there is a great exchange between residents of both cities. This analysis is intended to show which areas of one city resemble those of the other: Newyork and Toronto.

Problem statement

The steep rise in the cost of living,high population density starting number of new york residents to move to Toronto. The influence of Toronto has grown significantly in the past 20 to 30 years .The question for this subset of people is how to get a similar place in Toronto like newyork.
Foursquare is a website where people comment and rank food sites, coffee sites, malls and parks. For instance, let's think that a Foursquare user had to move from New York city, USA to the city of Toronto, Canada. Foursquare location data along with a clustering algorithm can suggest a neighborhood in order to help this user to live in Toronto in a similar place. The neighborhood that will be suggested, will not be a random suggestion, but instead will be a place for his pleasure. Thus, previous data from New York and Toronto will be used to predict a good living neighborhood for him.

2.Data

In order to analyse the cities on a meaningful level, they need to be divided into different areas, e.g. neighborhoods, boroughs.A list of neighborhoods in New York and Toronto is downloaded and their respective location in longitude and latitude coordinates is obtained. The sources are the following:

Newyork
https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json

Toronto
https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&direction=next&oldid=942655364

Foursquare API will be used for this project.Moreover, their specific coordinates are merged. Only Manhattan neighborhoods and boroughs that contain the string "Toronto" are taken into account. A Foursquare API GET request is sent in order to adquire the surrounds venues that are within a radius of 500m. The data is formated using one hot encoding with the categories of each venue. Then, the venues are grouped by neighborhoods computing the mean of each feature.

The similarities will be determined based on the frequency of the categories found in the neighborhoods. These similarities found are a strong indicator for a user and can help him to decide whether to move in a particular neighborhood near the center of Toronto or not.