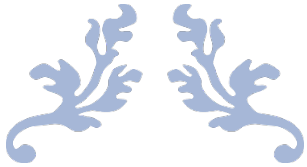

NEW YORK AND TORONTO - AN ANALOGY

COURSERA - CAPSTONE PROJECT REPORT

JUNE 24, 2019

SUDHINA S.

Contents

1. Abstract

This project deploys unsupervised machine learning algorithms to compare the cities of New York and Toronto using venue information of the two cities from FourSquare, a Location Based Social Network site (LSBN). It tries to understand the degree of similarity between the two cities while grouping similar neighbourhoods into clusters. This neighbourhood-based comparison of the two cities would be highly helpful to tourists, migrants, businesses and government bodies in recommending locations, urban planning, in the analysis of cities and urban computing.

2. Introduction

2.1 Background

'Stranger you are, Stranger am I...

Nomads in the jungle of civilizations formed by our nomadic forefathers!'

People flock from city to city, country to country in search of better job, better living conditions, better opportunities, sometimes to quench their traveller spirit and so on...To explore new places, new cultures is often life-changing an experience that walk with us to the grave and beyond! So, we see many around us possessed by the gypsy spirit - those who '*travel for travel's sake*!' (R.L. Stevenson, 1878) When we voyage, we often seek novelty. Imagine spending your time, money and effort to end up in a place which is strikingly similarly to the last place you had been to, which has nothing new to offer!

Nuances of 'migration' extend the limits of just a geographical shift to a cultural shift. Being a-few-days visitor is far different from being some-long-years resident. To uproot one's self from the conversancy of homeland to the boons and perils that the novelty offers is easier said than done. Even as enticing and developed the new city/country/neighbourhood may be or perilous and under-developed motherland might be, one's heart always yearns for home. Like people say, a part of your home will always travel with you. Afterall '*homeland is one of the magical fantasy words like unicorn and soul and infinity that have now passed into language*'(Zadie Smith, 2000). How wonderful it is to be able to relocate to a better city which is reminiscent of our home - to be able to fly for our dreams under the aegis of familiarity that our homeland offers!

2.2 Business Problem

This project seeks to cushion the transition for people migrating between the cities of Toronto and New York. Toronto and New York are urban cynosures coveted by many-an-urbanizing youth. They house major industrial, educational and cultural hubs. These cities are growing every passing second! So are the number of people travelling/relocating to/between the two cities!

The exercise clusters the two cities and compares the similarities and dissimilarities between the two based on their major venues. It addresses the question 'How similar are Toronto and New York?' It takes the goal a step further to group similar neighbourhoods in the two cities, for instance, which neighbourhoods in New York are more similar (read ' more comfortable') for a person from East York, Toronto?

2.3 Business Interests

Comparison across cities helps business, people and government alike!

Owing to the growing globalization (and hence the growing numbers of MnC's) and urbanization, workforces transcend geographic boundaries. Business today demands travel and relocation (however long or short a period it may be). Employees are deployed to distant cities frequently and quickly! So, the migration trends are booming. To add to that, people today (in general) relish travel and change. They yearn for heights and are ready to strive for it, shift for it. Any person travelling to a new city looks it up in advance, tries to prepare oneself for the change. While tourists try to check out what is different in the place, people who are relocating try to select a place as close to home. In short, all of us benefit from knowing what is similar and what is not between cities.

City-to-City comparisons helps a firm to zero in on its best location to start/expand to, a tenant to decide on the best rent available in a new place, a professional to identify prime zones/cities to work etc.

Comparing cities also helps in urban planning - to understand what is going on where, to realize best inter-city plans, to analyse what strategy might do well and what might not, to help maintain balanced, decentralised growth and so on...

3. Data Acquisition and Cleaning

3.1 Data sources

To compare and cluster the cities of New York and Toronto, Canada postal code dataset available in the wiki page (coupled with geospatial data) and New York city neighbourhood dataset (source: New York city Spatial Data Repository site) are used. Further, venue information for each city is obtained from FourSquare, a Location Based Social Network site. The venue information is easily accessible through a public REST API.

3.2 Data Cleaning

The borough, neighbourhood, latitude and longitude data obtained from the first three sites are combined, cleaned and formatted to two separate dataframes, one for each city. Only the rows of data that have a borough assigned are processed for further use. All cells with no borough value or value of 'Not Assigned' are ignored. For rows having borough data but no neighbourhood data, the neighbourhood is taken to be the same as the borough.

Venue information is obtained through Foursquare API using the latitude-longitude data for each neighbourhood. This data is then sorted and grouped to get 10 most common venues per neighbourhood.