

Capstone project - The Battle of Neighborhoods

Report

By Danijela Zivic

1. Introduction/Business Problem

In this project we will analyze existing venues in neighborhoods between Toronto and New York City. As a European tourist visiting Toronto in Canada or New York City in USA which are both megacities for tourist who lives in a city of 10.000 inhabitants. It would be helpful to decide which of those two big cities is the best city in North America to visit first.

With this project the traveler will explore the similarities and dissimilarities between neighborhoods in those two cities, and determine which neighborhoods are best fit for traveler who likes good food for reasonable price. With implementing recommender system to suggest and locating the best places to visit by finding and filtering different venues, especially restaurants based on the number of likes, prices and rating.

2. Data

To execute traveler idea we will be retrieving data from:

- Toronto data from Wikipedia page - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Toronto Latitude and Longitude - http://cocl.us/Geospatial_data - a csv file that has the geographical coordinates of each postal code
- New York City data, Latitude and Longitude – the link https://geo.nyu.edu/catalog/nyu_2451_34572, but better is from file placed on the server https://cocl.us/new_york_dataset

First we will transform the Toronto data from Wikipedia page into a pandas dataframe and with using the Geocoder package and csv file get latitude and longitude of Toronto. Then New York dataset file, already placed on the server, we will simply run a wget command and access the data, then load and transformed into pandas dataframe. With utilizing the API Foursquare data get the latitude and the longitude coordinates and details about venues, especially restaurants of each neighborhood in Toronto and New York City separately.

Then with Foursquare and various python packages we will explore, cluster, create maps and use machine learning models to get a list of recommended neighborhoods that have the most highest numbers of likes, ratings and for reasonable prices. The bonus that will be gained through the survey is that we can also provide to the traveler a list of nearby attractions.