

Coursera Capstone
IBM Applied Data Science Capstone

Move to Vancouver or Toronto

By: Chenhan Zhao

March 2020



Vs.



Introduction

For many people, they work their whole life to pursue their dream of working for a big company at a big city. They can have convenient shopping experience, world-class medical care and excellent school district for their kids. However, there are so many big cities which one do you want to go become a problem for many young people. Vancouver and Toronto are two of the biggest cities in Canada which contains 20% population in Canada, and they offer completely different experience for their citizens. One is on the Pacific coast surrounded by mountains and islands, the other is a major Great Lake city. As a result, property developers also taking advantage of their geological environment and built different kinds of venues around the city. The entertainment life contributes a lot to our overall happiness which ultimately affect our productivity. Therefore, choosing the right city that consistent with your lifestyle is crucial to lead a healthy and happy life.

In this report, we aim to use cluster method and venue information extract from Foursquare API to analyze the lifestyle of Vancouver and Toronto. By the analyzed results, we can provide city recommendation for people who are worried which city they should go.

Data

To solve the above problem, the following data are obtained:

- List of postal codes, borough and neighborhoods for both Vancouver and Toronto. These data confines that this project to the city of Vancouver and Toronto.
- Latitude and longitude for each neighborhood. This information is used to obtain the exact location of each neighborhood in order to plot the maps and get venue data.
- Venue data, especially data relate to lifestyles like restaurants, gyms, trails, etc. These data will be used to analyze the lifestyle of each city and each neighborhood.

Data source and extraction method

- List of neighborhoods are extracted from the Wikipedia page for Toronto (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and Vancouver (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V). Web scraping methods are used to extract the tables from web page with Python package *pandas*. In order to only contain the central areas of each city, we only extract neighborhoods whose borough contains the word “Vancouver” or “Toronto”.
- Geological information is obtained from GeoNames website (<http://download.geonames.org/export/zip/>). The package contains all postal codes in Canada and their latitude, longitude is downloaded. This information is imported as *pandas* DataFrame and assigned to corresponding neighborhood.
- Venue information is imported through Foursquare API and after data wrangling, K-means clustering method will be used to analyze each neighborhood and their similarities.