New York and Toronto comparison for companies that are opening offices in both cities

Breno Santos

December 22, 2020

1. Introduction

1.1 Background

Many companies have locations in New York (United States) and in Toronto (Canada) given that they are the financial centers of their countries. Some examples are Spotify and Fareportal. Companies in the technology sector are a special case in this scenario. A report from a Real Estate company in the US shows that Toronto is "North America's fastest-growing technology market" and many US companies are opening offices there.

1.2 Problem

Data that might contribute potential buyers to understand how a neighborhood in New York relates to a neighborhoods in Toronto, in order to be easier for US companies to choose their locations in Toronto.

1.3 Interest

The audience of this study would be all those US companies that plan to open offices in Toronto.

2. Data acquisition and cleaning

2.1 Data Sources

(a) The list of postal codes of Toronto available in Wikipedia;

(b) Latitudes and Longitudes of each Toronto neighborhood made available in one of the Labs of the Applied Data Science Capstone in Coursera;

(c) Foursquare database of Toronto venues made available by a Foursquare API.

(d) A json file with neighborhoods in New York and their latitudes and Longitudes. This json file was made available in one of the Labs of the Applied Data Science Capstone in Coursera;

(e) Foursquare database of New York venues made available by a Foursquare API.

2.2 Data Cleaning

The Postal Codes that have no neighborhoods assigned to it were removed from the Dataframe. The columns headers were standardized.

2.3 Feature selection

From the Foursquare API, venues locations and categories were relevant to the project.

3. Methodology

3.1 Exploratory Data Analysis

First, we collect Toronto postal codes (a) and merge the data with neighborhoods coordinates (b). Then we pull Fousquare data (c) in those areas.

In part two, we collect New York neighborhoods data (d) and relates it with Fousquare data (e) in those neighborhoods.

Finally, we compare the data on those 2 cities.

3.2 Machine Learning model

To cluster the neighborhoods, it was used an unsupervised learning model widely used for clustering, K-means.

4. Results and recommendations

Once the neighborhoods from Toronto and New York were combined, it was noticed that 'Cluster 0' were made by American and Canadian items. But Cluster 1, 2, 3 and 4 were made only with Canadian elements.

Inside Cluster 0, there were 83 elements from Toronto

5. Conclusion

In this project, I analyzed how similar are the Toronto neighborhoods and the New York neighborhoods. I built a classification model to cluster the neighborhoods from both cities combined. At the end, I was able to list 83 neighborhoods in Toronto that are similar to the neighborhoods in New York.