Housing Sale Prices & Venues Data Analysis of Toronto and New York

Maias Alakramy

23-APRIL-2021

1. Introduction

1.1. Background

Every spot-on earth is unique and there aren't two places that are identical and every place has its geographic information, culture, customs and traditions and as always stockholders want to target the perfect area for their businesses

1.2. Problem

Although every spot-on earth is unique but Stockholders wants to find the similarity between the location because they want to target similar location with similar approaches. Hence in this report we will help the stockholders in Toronto to find new locations to target in New York

1.3. Interest

Obviously, stockholders in Toronto would be very interested in locations in New York that are similar to Toronto in the top venues

2. Data acquisition and cleaning

2.1. Data sources

In this project we will need locational data and venues for both New York and Toronto

New York's data will be collected from a JSON file here since it's a dataset of New York's neighborhoods with the proper geographical coordinate.

For Toronto's data we will web scraping this Wikipedia page of Canadian postal codes and join it with this csv file of a proper geographical coordinate

For each neighborhood we will get all the venues around it in a radius of 500 and this will be obtained using Foursquare API

2.2. Data cleaning

When using Foursquare API, we found that some locations in New York don’t have any venues in the selected radius so we drop it because out project is based on venues

3. Analytical modeling

3.1. Clustering

We will be using K-Means as a clustering method and we choose to cluster the data into 3 clusters in hope to split the data into 3 main parts:

1. Toronto's neighborhoods that aren't similar to any of the ones of the selected New York borough
2. Neighborhoods that are similar
3. New York borough's neighborhoods that aren't similar to any Toronto's once

But the analysis will not be that precise i.e., it will split the data in different way

3.2. Results and Discussion

Our analysis shows that although the three boroughs we studied are very similar to Toronto but we found a borough that is the most similar to Toronto, so we focused our attention to Staten Island.

After directing our attention to Staten Island, we provide more info of its neighborhoods and link the similar neighborhoods in Toronto and Staten Island to each other.

Result of all this is a list of Toronto's neighborhoods and their similar ones in Staten Island. This, of course, does not imply that those neighborhoods are actually optimal locations for a business moved from Toronto to New York! Purpose of this analysis was to only provide info on.

Recommended neighborhoods should be considered only as a starting point for more detailed analysis which could eventually result in location which has also other factors taken into account and all other relevant conditions met.

3.3. Conclusions

Purpose of this project was to find an optimal location for a business moved from Toronto to New York in order to aid stakeholders in narrowing down the search for optimal location in New York boroughs to target.

By using Foursquare to gather data and clustering Toronto's data with each borough's data we have first identified the borough that is the most similar to Toronto (Staten Island), and then we list the similar neighborhoods to be used as starting points for final exploration by stakeholders.

4. Future directions

Final decision on optimal neighborhood to be targeted will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended neighborhood.