Home is here too

Applied Data Science Capstone Project by IBM/Coursera

# Introduction

In this project we will find **similar neighborhoods** between two cities. Specifically, this report will be targeted to people who will be reallocated and is interested in finding a **new home in a strange city**. The idea behind finding the most similar neighborhood is to make it easier the adaptation and the sense of belonging in the new city.

For this project we will find the most similar neighborhoods in **Toronto** city to **Brooklyn’s Sunset Park, New York**. We are interested in searching on neighborhoods which are in a **distance limit of 7.5 km from the new workplace**. The new workplace will be located in **129 Spadina Ave**.

We will use our data science powers to generate a few most promising neighborhoods. Advantages of each area will then be clearly expressed so that best possible final location can be chosen.

# Data

Based on the definition of our problem, factors that will influence our decision are:

* Number of each kind of venue per neighborhood: restaurants, grocery stores, parks, etc.
* Distance from the new workplace

Following data sources will be needed to extract/generate the required information:

* Neighborhood's locations of Toronto city from **Wikipedia and external csv file shared on this course**
* Location and type of restaurants, parks and grocery stores in every neighborhood will be obtained using **Foursquare API**
* Coordinate of new workplace will be obtained using **Geocoder** or given since it is unreliable

# Methodology

In our first step we obtained the data of the venues in the neighborhoods of Toronto which are in a range of 7.5 km from the new workplace: 129 Spadina Ave.

1. We obtained the postal code's and neighborhoods of Toronto by scraping Wikipedia
2. We obtained an approximate postal code's latitude and longitude information with an csv file shared on this course
3. We converted the latitude and longitude information to distance from the workplace in km with pyproj library
4. We filtered the postal codes by the ones in a range of 7.5 km from the workplace
5. We included venues information for each postal code with Foursquare

Then, with this information we found the most similar neighborhoods of Toronto in terms of venues to the current neighborhood: Brooklyn's Sunset Park in New York.

1. We calculated a similarity score between neighborhoods by using the dot product
2. We selected the neighborhoods with the higher similarity score

Only for curiosity, we clustered the neighborhoods of Toronto including in that list the current neighborhood in New York to check to which neighborhoods in Toronto the current neighborhood is associated.

# Results

Our analysis shows that the most similar neighborhoods in Toronto to Brooklyn's Sunset Park, New York are:

* Moore Park
* Summerhill East
* The Annex
* North Midtown
* Yorkville
* Summerhill West
* Rathnelly
* South Hill
* Forest Hill SE
* Deer Park
* Rosedale
* St. James Town
* Cabbagetown

However, similarity found was small. The higher similarity found was of only 4.7%.

# Discussion

We proposed 13 neighborhoods in Toronto in a range of 7.5 km from the workplace. Probably other areas further from home could be more similar. Anyways, for next analysis it can be chosen whether the person wants to give more emphasis to some venues which they have in their current neighborhood. Also type of living (house, apartment, etc.), price of properties and size of the properties should be included in further analysis.

When creating the clusters, we found that most neighborhoods of Toronto got classified into the same cluster. This didn't provide us with so much insightful information to verify the results obtained with our similarity score.

# Conclusion

The purpose of this project was to propose neighborhoods in Toronto close to the workplace that were similar to their neighborhood back home: Brooklyn's Sunset Park, New York.

With the proposed neighborhoods, the person who is moving can help themselves to visiting those places first in the process of finding a new home in a strange city.

We proposed a total of 13 neighborhood's in Toronto that where in a distance range of 7.5 km from workplace.

Anyways, the similarity found was small. So, more work needs to be done to improve and explore the proposed neighborhoods, according to what the client values more to have at home.