**Comparing New York and Toronto Neighborhoods for XYZ Corp relocation feasibility analysis.**

Anton Anonymous

March 2020

1**. Introduction.**

1.1 **Background.**

The XYZ Corp is looking forward to relocate part of its business operations from NY city

to Toronto.

Among many business considerations is one voiced by HR department.

Bussiness nature of XYZ corp will require moving number of employees from NY city to

Toronto.

HR is concerned about cultural and quality of life changes for these

employees that will move to Toronto.

1.2 **Problem.**

Information on cultural and lifestyle similarities between Toronto

and New York is not readily available and is not easily measurable.

1.3 **Interest.**

HR department of XYZ Corp is deeply concerned that relocating some of employees to Toronto might be a failure due to cultural differences.

2.0 **Data.**

2.1 **Data Sources.**

Foursquare API endpoints provide venue database for both Toronto and New York.This database has over sixty descriptive attributes that very well describe each venue. Borough and neighborhood data for both cities is readily available from many sources such as <https://en.wikipedia.org/wiki/List_of_postal_codes_in_Canada> and <https://geo.nyu.edu/catalog/nyu_2451_34572> .These sources will provide venue data and venue selection within city centre radius.

Python geopy.geocoders package provides tools for geolocation of cities.

2.1 **Data Cleaning.**

There was no need to clean the data because lists od postal codes , geolocation data and FourSquare

API endpoints provide already cleaned data.

2.1 Feature selection

For lookup of neighborhoods of Toronto and New York : Borough, Neighborhood,

Latitude and Longitude were used.

Foursquare API venues endpoint provided venues: name , category,lattitude and

longitude.

Venue category was used to create feature set. This feature set is dependent on set of possible

values for venue category attribute , this can change in time and may be different for different

neighborhoods. Venue category attribute has hundreds of possible values.

3.0 Methodology

I used k-means clustering analysis on cities neighborhoods to determine similarities

between Toronto and New York neigborhoods in their venues.

For each neighborhood I derived new feature - mean of venue category for that

neighborhood. Geolocator package will be used to locate Toronto and New York boroughs

and neighborhoods. This information will be used in turn to select venues for each neighborhood.

The aggregated information on all venues in the neighborhood (for both cities) will be used in

k-means clustering analysis.

I expect that if clusters derived from this analysis contain proportional number of neighborhoods

from both cities it will be safe to claim that neighborhoods in both cities are quite similar with

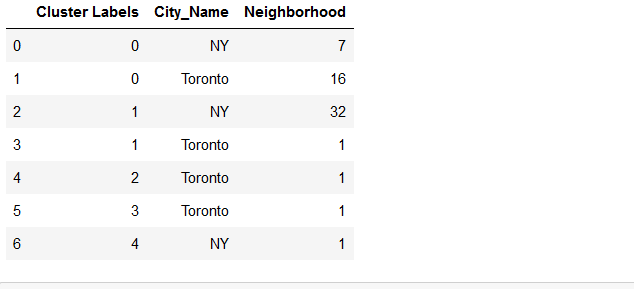
respect to their venues.

4.0 Results.

Neighborhoods were grouped to 5 clusters.

2 clusters were unique for Toronto and 1 for New York.

These contained together only 3 neighborhoods of 59 in both cities.



5.0 **Discussion.**

HR is concerned about relocating its employees from New York to Toronto.

One of the aspects of possible relocation to consider is if living and lifestyle of both

these cities is similar. By analyzing neighborhood venues for both cities I found out

that in this aspect these cities are quite similar and HR should not be overly concerned

about differences. However this analysis tells us nothing on other aspects of life such

as cost of living , number of schools , healthcare etc.

6.0 **Conclusion.**

Both New York and Toronto seem to be quite similar to live in. There are three neigborhoods

that stand out, further analysis of these may be performed if necessary.

To perform deeper analysis other sources of data on cost of living, education , healthcare

will have to be used.