Introduction/Business Problem

A small technology company (“NYCco”) based in New York City has grown too big for its headquarters. The CEO decided to start a satellite office in Toronto, Canada. This will enable it to attract some of the engineering talent graduating from Toronto Universities. It has chosen three key managers (Amy, Brian, and Claire) to move themselves and their families from NYC to Toronto immediately. Given the tight timeline, NYCco has outsourced the relocation and moving logistics to one of NYC’s trusted relocation agencies (A2B inc.). This agency is known to be data-centric and matches the personal interests of each of its clients with a similar neighborhood in their new location.

These three moves are important to the business stakeholders at NYCco. because they need the satellite office to be profitable within 18 months. The way to achieve profitability is to retain their top three managers for the duration of this onramp period. The biggest risk is adding too much strain and stress to the three manager’s lives (and their families). In order for a smooth transition, they need the Manager’s new Toronto neighborhood to feel like home.

In addition to NYCco stakeholders, any individuals that are considering a move from NYC to Toronto will be interested in this project because they can optimize their housing search based on their current interests.

> *The following personal bios of NYCco Managers have been sent to A2B inc.:*
>
> *Amy:*
> *A routine and career-driven executive. She lives by herself and has a strict morning routine. Jog through the park, attends yoga class, and stops by her favorite coffee shop before 8am every day. She doesn’t care much for nightlife as she is often still working or in bed with a book before 9am.*
>
> *Brian:*
> *Brian is a family man. He has four young children that play on competitive sports teams, and a foodie wife that loves to explore hip, new multicultural restaurants with her girlfriends. Their favorite weekend activity is taking all of the kids to the local pizza place after their various team sports followed by ice cream.*
>
> *Claire:*
> *Claire is a young 20-something party-goer. She loves to go out on the town on Friday night, then unwind on Sunday. She loves dancing the night away with her friends on Friday night, and relaxing at her nearby spa on Sunday.*

A successful outcome will be identifying the three Toronto neighborhoods for Amy, Brian, and Claire that are most similar with their current neighborhoods in NYC. If all three managers are happy in their new neighborhoods and remain in Toronto through the duration of the trip, they can help the business stakeholders reach their 18-month profitability target.

Data

The key data needed to solve the business problem are the (1) neighborhood names with associated latitude/longitude coordinates and (2) detailed information on venues throughout each city (NYC and Toronto).

For the New York City data, I will be using the geoJSON file located at https://geo.nyu.edu/catalog/nyu_2451_34572. This file provides the names and latitude/longitude coordinates for all NYC neighborhoods.
For Toronto, the data will be scraped from a Wikipedia page here: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M . Additionally, this table will be joined with postal code location information found here: http://cocl.us/Geospatial_data

Secondly, venue data will be acquired through the social media platform called Foursquare. Foursquare is a popular tool for data science because it contains accurate geospatial coordinates of various venues along with specific venue types, photos, and user-specific information such as written tips. The Foursquare Developer API (via https://foursquare.com/developers/) and it's explore function will help me pull all the corresponding venue type data to determine which Toronto neighborhoods are most similar to the respective New York City neighborhoods.

To solve the problem, I will look at the frequency of certain venue types in each neighborhood, and then use the K-Means clustering algorithm to group similar areas together. This will provide us with clusters that can be summarized in real-world groups such as "Coffee shops", "Nightlife", or "Parks & Rec." We will then match the personal interests of Amy, Brian, and Claire with the appropriate neighborhood that most closely aligns with their described lifestyle.