The Battle of the Neighbourhoods

1. Introduction/Business Problem

Relocating can be a stressful and scary time in one's life. After studying hard and gaining new life skills being headhunted by several successful companies can be both exciting and rewarding. After careful consideration the choices have been narrowed down to two companies in two different cities. New York or Toronto. In order to help make a final decision it would be helpful to understand the areas around the potential employers. By using Foursquare an informed decision about the differences and similarities between the two areas can be arrived at and will assist in choosing the best company to work for.

2. Data

For the analysis data for New York and Toronto will be used, taking into consideration boroughs and neighbourhoods in each city. The first city looked at will be New York, and the data can be found here. The data is in a json format and will need to be transferred into a pandas dataframe. The data frame will have the below structure:

	Borough	Neighbourhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

The Toronto data will be scrapped from Wikipedia. The data will then be wrangled into a analysis friendly format. Some of the wrangling will include dropping rows were the borough equals "Not assigned", Also, if Neighbourhood is "Not assigned" but Borough is assigned them we will make the corresponding Borough the Neighbourhood. Next, we will merge any duplicate rows (Postal Code and Borough) and then merge the Neighbourhoods to be separated by a comma. Finally, the latitude and longitude data will be downloaded from here and merged with the above data from Toronto.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848

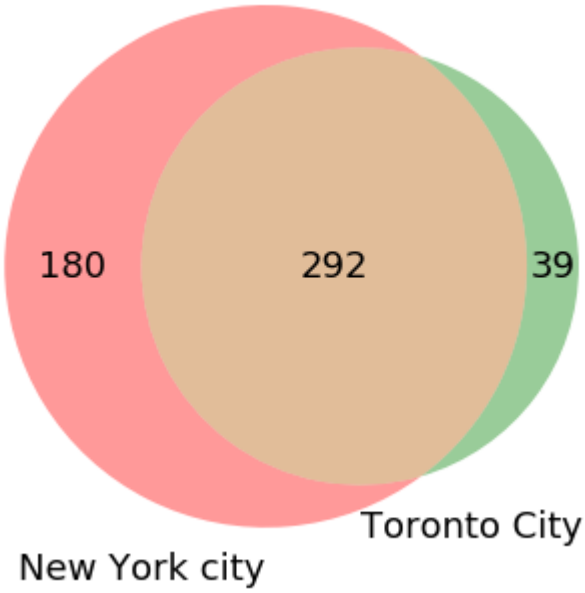
Once the above has been completed we will pull through venue data for the Neighbourhoods using Foursquare API.

3. Methodology and Foursquare API

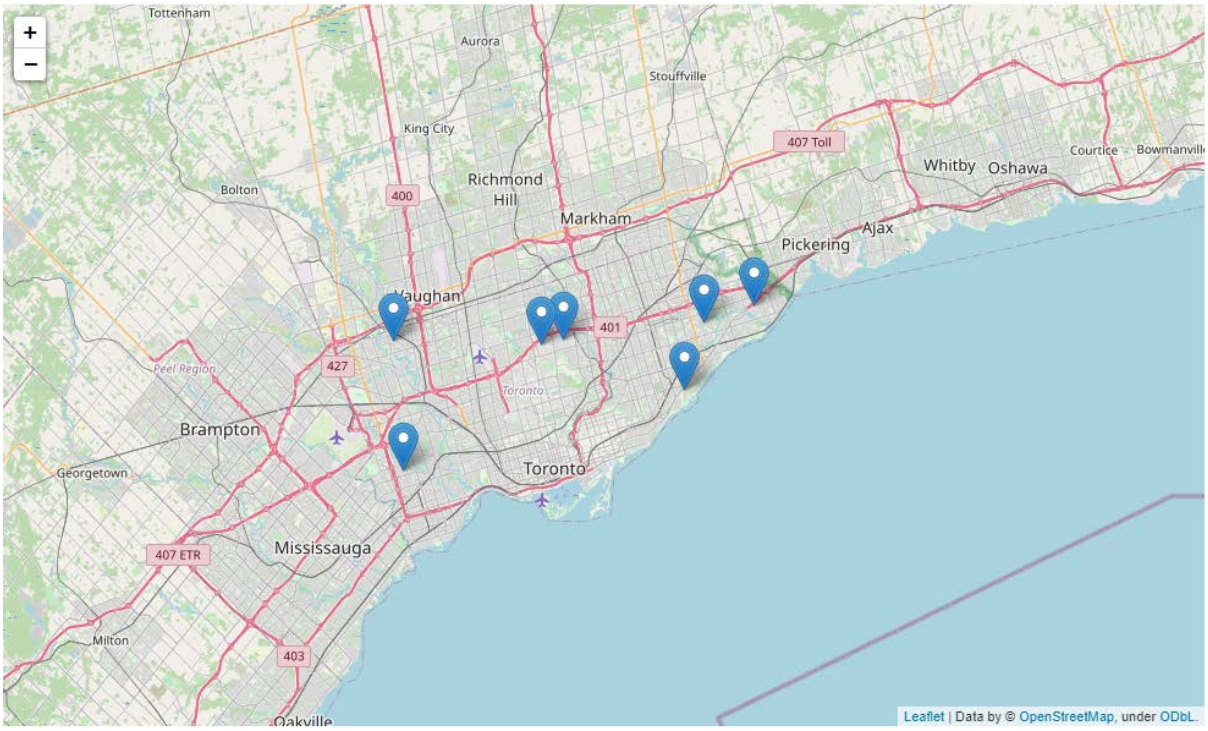
After obtaining the above data and performing the data wrangling, we will look to find nearby venues to each borough and neighbourhood in both cities. Foursquare API will be used to get the venue data, and in order to complete this a developer account is needed along with an app that provides the client id and client secret. As they are different cities, the venues will not be perfectly matched across both, so we will only consider common venue categories. This will allow us to find neighbourhoods in both cities that are similar and help us determine which city will be a better lifestyle match.

4. Analysis

As expected there are similar venues common to both cities, and some that are unique to each city. Below we can see that there are 292 venues that are common across both Toronto and New York. Additionally, New York has 180 unique venues as opposed to 39 unique venues in Toronto.



This could potentially mean that New York would be the better fit as it contains more unique venues, however, Toronto does share most venues with New York. Below are the mapped areas in both cities, showing where these venues are in relation to each other.



In Toronto the venues are spread out somewhat around the city, and this could mean that traveling between the selected venues may not be feasible.

Looking at New York, the venues are also spread out, but occur in groupings that could mean travel between them is easier.

5. Conclusion

Toronto and New York both offer multiple options for selected venues. There are 290 venues that are in both cities, and New York has a larger population of unique venues in relation to Toronto.

Additional analysis would be needed to calculate costs of transport between the venues in each city and is not in the scope of this project.

Based on the data used New York would seem to offer the better option as there are more unique venues that are not present in Toronto. Additionally, several of the venues in New York are located closer together in groupings, whereas in Toronto the venues are more spread out across the city. The largest cluster of venues in New York are in the Bronx and within an easy travel distance to La Guardia airport. As long as the company located in New York is not far from the Bronx or is easily accessible by public transport then New York is the better choice, based on the data used in this analysis.