The Battle of Neighborhoods

1 - Introduction/Business Problem

- Working class people are always moving from one place to another either to find better jobs or to change their lifestyles a little bit.
- My cousin got a job offer from a great company with mind-blowing employee benefits and perks but in other country.
- He took the job offer because it was a once on a lifetime opportunity and now he and his family need to change to a whole new location. From New York to Toronto.
- They are urgently in need of someplace exactly or strictly similar to their last location because they have been living there for many years so they love their current neighborhood venue settings and amenities.
- Either way they always get to the same general obstacles for anyone trying to please every family member while scouting for new places to live.
- The essential most important ones are like: the company location or site, schools nearby, health centers, grocery shops, entertainment, recreational parks, restaurants, coffee shops, etc...

How are they going to find another city, borough or neighborhoods with all those conditions for everyone to be happy as they are right now???

We'll find out what makes borough-neighborhoods similar to define a new location. Moving somewhere can be within the city or from one city to another. If moving within the same city then we'll find neighborhoods in that borough which are very similar to their last location. If moving it's being done from one city to another city then my task would be finding the similar neighborhoods.

2 - Data

2.1 - Reading Data

2.1.2 - Gathering New York Data

My first city would be New York City, here's the link to access the data. NY data is in a json format and it's easily transformed into pandas data frame. We'll be using the borough and neighborhood data of two cities. We´ll only keep the important ones like borough, neighborhood, latitude and longitude. For more information of this data you can visit this NYU webpage.

Figure 1: New York Data

	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

2.1.2 - Gathering Toronto Data

Second city would be Toronto city. It can´t be directly downloaded. Postal Code, borough and neighborhood can be scrapped from the Wikipedia page. Used csv file download the csv and read it into pd. There are some rows where “Not assigned” is written. Drop the rows where borough is “Not assigned” (only).

If Neighborhood is 'Not assigned' but Borough is assigned then make corresponding Borough as Neighborhood. Merge the rows if Postal Code and Borough of two or more rows same and merged neighborhood will be separated by comma “,”. Latitude and longitude info taken from here. Finally this data and Toronto data can be merged together. Once both data are ready we’ll use Foursquare API to get the venues near each neighborhood.

Figure 2: Toronto Data

	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

3 - Methodology

After getting above data, we’ll find the nearby venues to each borough and neighborhood pair in both cities. Foursquare API will be used to get the nearby venue. To use the Foursquare API we’ll need the developer account and an app which will give the client id and client secrets. Both city may yield in different number of venue categories, but we’ll take only common venue categories. Now we can find a borough and neighborhood of a city is how much similar to the boroughs of another city. In other words we can find top similar boroughs and neighborhoods in another city. We’ll use cosine similarity to find the similarity between two cities

4 - Getting Venues Using Foursquare API

Foursquare API will be used to get the nearby venues. We’re going to find the venues in 1km range from a borough and neighborhood pair.

4.1 - For New York City

The shape of this data is (306, 474). 302 columns are venue categories columns.

Figure 3: New York Data with Venue Categories

	Borough	Neighbourhood	Latitude	Longitude	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	Staten Island	St. George	40.644982	-74.079353	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.013699
1	Staten Island	New Brighton	40.640615	-74.087017	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.026316
2	Staten Island	Stapleton	40.626928	-74.077902	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
3	Staten Island	Rosebank	40.615305	-74.069805	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
4	Staten Island	West Brighton	40.631879	-74.107182	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.018182

4.2 - For Toronto City

The shape of this data is (103, 339). 334 columns are venue categories columns

Figure 4: Toronto Data with Venue Categories

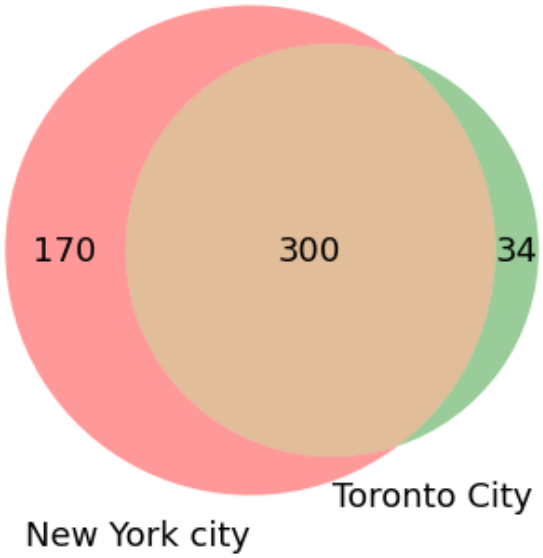
	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Accessories Store	Afghan Restaurant	African Restaurant	Airport	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aq
0	M8V	Etobicoke	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	0.0	0.0	0.0	0.0	0.055556	0.0	0.0	0.0	
1	M8W	Etobicoke	Alderwood, Long Branch	43.602414	-79.543484	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
2	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
3	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	
4	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	

5 - Analysis

5.1 - Common Venue Categories

There are some common venue categories in both data. Figure-5 is showing that there are total 300 common venue categories in both cities. So we'll take only these common categories of data. Other than common categories of the data we'll be dropped.

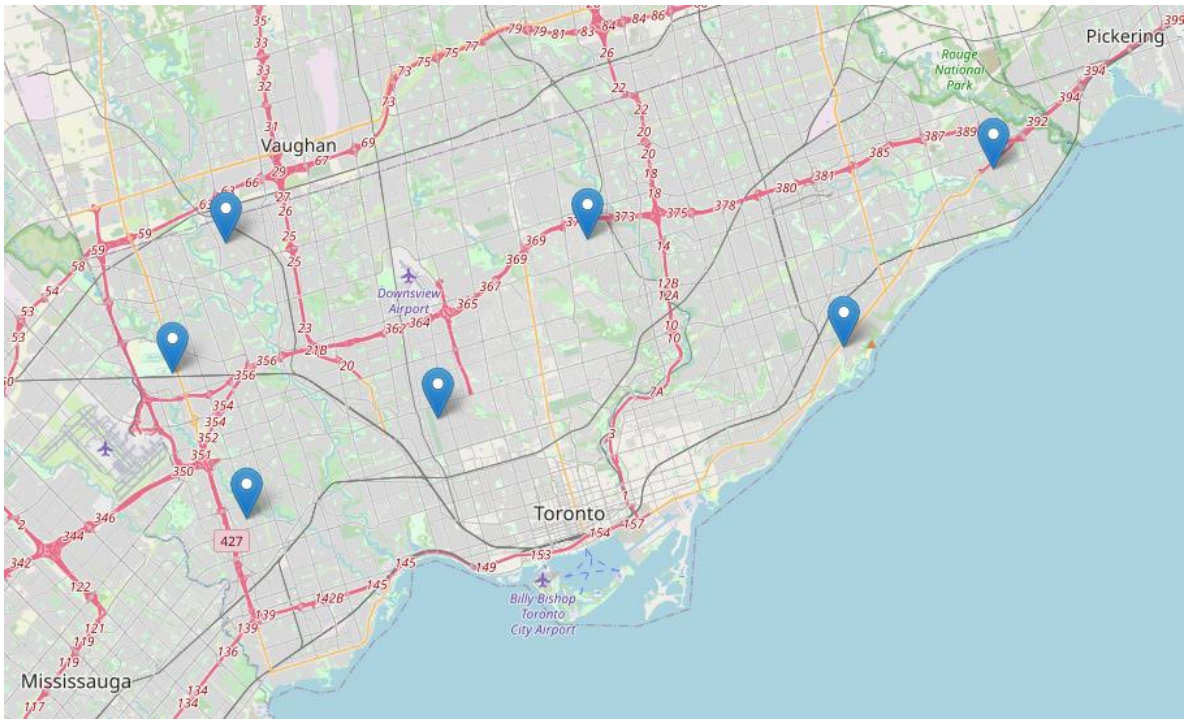
Figure 5: Venn-Diagram of Venue Categories of both cities



5.2 - Top 7 similar boroughs in Toronto City

Let's say current location is Bronx, Riverdale, New York. We'll get the index; latitude and longitude of this location form the New York City data. We can get the all common venue categories data of corresponding index and then we'll multiply with Toronto City data. This will give how much each borough is similar to current location. We'll sort these values and get the top 7 boroughs. Then we'll visualize it using the folium.

Figure 5: Venn-Diagram of Venue Categories of both cities



6 - Conclusion

Cosine similarity is used here to find out the similarities between two boroughs. This model can be implemented within a city also. This will help to find a suitable place for people. This will give most similar borough and neighborhood.