The Battle of the Neighborhoods

(Finding Similar Neighborhoods between Different Cities)

1. Business Problem

This project was to try to find similar neighborhoods in two different cities based on the type of venues and popular spots contained in the neighborhoods. Specifically, we looked at neighborhoods in two major cities; (Manhattan) New York, New York USA and Toronto, Ontario Canada.

This project would be of interest to people moving from Toronto proper to Manhattan and are looking to settle in similar areas. This project would also be of interest to companies looking to move headquarters or open up branch offices in Manhattan in areas that are similar to the ones they currently operate in in Toronto.

2. Data

In order to complete our project, we needed the following data:

- Coordinate Data for the city centers of Manhattan and Toronto, available from **GeoPy Nominatim**
- Coordinate Data for the neighborhoods in Manhattan, available at this URL: **https://cocl.us/new_york_dataset**
- Neighborhoods in Toronto by postal code, available on **Wikipedia: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M**
- Coordinate Data for postal codes in Toronto, available at this URL: **http://cocl.us/Geospatial_data**
- Coordinate Data and Venue/spot type from **Foursquare API**

Using the Foursquare API data, the various neighborhoods of Manhattan were put into a matrix based on the most common types of venues in that neighborhood. Then, the same was done with the various neighborhoods of Toronto proper. Finally, the neighborhoods were compared to determine which are 'closest' to one another by examining a matrix of the relative frequencies of each type of venue.

2.1 Manhattan Data

First, the Manhattan neighborhood data was downloaded and saved into a Pandas dataframe. Then **Nominatim** and **Folium** were used to generate a map of Manhattan with the neighborhoods' locations marked. Finally, calls were made to the **Foursquare API** to generate a table containing venue data.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Fig 2.1: Sample Foursquare Data for Manhattan

2.2 Toronto Data

Then, the Toronto neighborhood data was downloaded from **Wikipedia** and saved into a Pandas dataframe along with the coordinate data from this URL http://cocl.us/Geospatial_data. Then **Nominatim** and **Folium** were used to generate a map of Toronto with the neighborhoods' locations marked. Finally, calls were made to the **Foursquare API** to generate a table containing venue data.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

Fig 2.2: Sample Foursquare Data for Toronto

Finally, venues which had venue categories which were not shared by both Manhattan and Toronto were dropped.

3. Methodology

After obtaining dataframes containing venue data from each city, one hot encoding was used to count the number of venues of each venue category in each neighborhood. Then, the mean of each venue category was calculated to the total number of venues in each neighborhood. This told us the relative frequency of that venue category to the total venues in that neighborhood. A value of 1.0 would represent that all venues in that neighborhood are of that category. A value of 0 would indicate that no venues of that category exist in that neighborhood. How similar two neighborhoods are was defined by how closely their column values match each other.

As we were concerned with how 'close' two neighborhoods are to each other the k-Nearest Neighbors Algorithm was used. Since we were only concerned about the 'closest' neighborhood we only needed to set k = 1 in the k-Nearest Neighbors algorithm. Also, since the dimension of each element of our matrix was 187, the calculation was simplified by only calculating the absolute difference of each column value (i.e. the 'Manhattan Distance') to find the most similar Manhattan neighborhood to our Toronto neighborhoods.

$$min\left(\sum_{i}^{n}|x_i - y_i|\right)$$

Fig 3.1: Equation used to determine how similar neighborhoods are where x and y are corresponding venue types

Looking at the Manhattan_grouped and Toronto_grouped dataframes, we could see that the values range from 0 to 1. These values represent the percentage of the total venues in each neighborhood that was of the given venue type. Thus, we could see that two neighborhoods would be completely dissimilar if by subtracting their column values and adding the absolute values of these we get a value of 2. Alternatively, if doing the same process yielded a value of 0, the two neighborhoods were as similar as possible within our definition of similarity. Thus, we could see that by subtracting two neighborhoods column values and adding the absolute values obtained a value that ranged from 0 to 2. This value was subtracted from 2 and then the result was divided by 2 to obtain a percentage of how similar the neighborhoods are

Neighborhood	Manhattan Neighborhood	% Similar
Commerce Court, Victoria Hotel	Murray Hill	0.539731
Toronto Dominion Centre, Design Exchange	Murray Hill	0.511735
King, Adelaide, Richmond	Murray Hill	0.493794
Underground city, First Canadian Place	Financial District	0.492903
Davisville	Yorkville	0.487950
Toronto Islands, Harbourfront East, Union Station	Financial District	0.466117
Kensington Market, Grange Park, Chinatown	East Village	0.464018
Ryerson, Garden District	Lenox Hill	0.460325
Cabbagetown, St. James Town	Hamilton Heights	0.453947
St. James Town	Chelsea	0.448284
Central Bay Street	Murray Hill	0.443861
Studio District	Manhattanville	0.443850
Stn A PO Boxes 25 The Esplanade	Murray Hill	0.436170
Trinity, Little Portugal	East Village	0.430293
Runnymede, Swansea	Yorkville	0.429557
The Danforth West, Riverdale	Sutton Place	0.417077
Church and Wellesley	Upper West Side	0.411111
The Annex, Yorkville, North Midtown	Morningside Heights	0.405435
The Beaches West, India Bazaar	Roosevelt Island	0.388889
North Toronto West	Flatiron	0.384698

Fig 2.3: Final Table containing each assigned Manhattan neighborhood and how similar

4. Results

The Toronto neighborhood (*Commerce Court, Victoria Hotel*) with the most similarity to its assigned Manhattan neighborhood (*Murray Hill*) was checked to see if they share similar types and relative amounts of venues.

Neighborhood	Latitude	Longitude	Manhattan Neighborhood	% Similar	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Commerce Court, Victoria Hotel	43.648198	-79.379817	Murray Hill	0.539731	Coffee Shop	Café	Hotel	Restaurant	American Restaurant	Gastropub	Deli / Bodega	Seafood Restaurant	Steakhouse	Italian Restaurant

Fig 4.1: Most common venues in Toronto neighborhood, Commerce Court, Victoria Hill

+/

Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Manhattan	Murray Hill	40.748303	-73.978332	Coffee Shop	Hotel	Sandwich Place	Japanese Restaurant	Italian Restaurant	Gym	Gym / Fitness Center	French Restaurant	American Restaurant	Bagel Shop

Fig 4.2: Most common venues in Manhattan neighborhood, Murray Hill

As we can see, their most common venues are both **Coffee Shops**, **11%** of the total venues in *Commerce Court, Victoria Hotel* and **5.3%** of the total venues in *Murray Hill*. They also both seem to have a decent number of **Hotels**, **6%** of the total venues in *Commerce Court, Victoria Hotel* and **4%** of the total venues in *Murray Hill*. Within their top most common venues they also have a fair amount of **American Restaurants**, **4%** of the total venues in *Commerce Court, Victoria Hotel* and **3%** of total venues in *Murray Hill*. The final most common venue type is **Italian Restaurants** with **3%** of total venues for both *Commerce Court, Victoria Hotel* and *Murray Hill*, this is how they are most similar.

The overall similarity between Toronto proper and Manhattan was calculated by averaging the % similarity of every Toronto neighborhood with its assigned Manhattan neighborhood. With an overall similarity of about **35%** we can assume that in fact Toronto is not that similar to Manhattan overall. However, if we focused on each borough of Toronto, we found that *Downtown Toronto* is overall **42%** similar. While the other boroughs are overall less similar than the total average with *Central Toronto* at **25%**, *West Toronto* at **33%** and *East Toronto* at **30%**.

5. Discussion

Through our analysis using Foursquare API data and by only looking at venue types that are common between Manhattan and Toronto proper, it can be concluded that Toronto is not very similar to Manhattan. In particular, Manhattan neighborhoods seem to be defined by the type of food available in them while Toronto neighborhoods seem to be far less so. To answer the original question, can we recommend a neighborhood of Manhattan that is similar to each neighborhood of Toronto, the answer appears to be no. However, we can make some fairly good recommendations for neighborhoods for Downtown Toronto in particular. However, there are a few ways we could improve our analysis and make a better recommendation. We could lump categories of food, drinking establishments and the like into one category. We could narrow the categories we look at, for example only retail establishments or only eateries based on the stakeholders needs.

6. Conclusion

The purpose of this project was to find neighborhoods in Manhattan which are similar to neighborhoods in Toronto Proper. In order to give recommendations to people and companies moving from Toronto to Manhattan and looking for a similar neighborhood. We have come to the conclusion that the neighborhoods of these two cities are in fact not very similar overall but we can make the best recommendation for each neighborhood with the data provided by Foursquare. As an extension to this project, it would be a good exercise to look at other large cities and see how well they compare to Manhattan and Toronto. Perhaps London or Paris would make a more suitable match to Toronto.