Comparing Diverse Cities

Comparing locations is valuable to help people relocate

- In business people are always being asked to relocate to new cities. This can be an extremely stressful and overwhelming event for the individual being relocated. To help reduce this stress, it would be nice to help the individuals compare their current location to their new location.
- Before smart phones, social media and the prevalence of machine learning algorithms, a person would have to rely on other people's suggestions or to have to read books about the new city they are going to. They may also choose to take some time to visit the new city. Since cities are big, it could take more than a week to visit all the neighborhoods.
- They may already know that they want to move to a neighborhood that is very similar to their current neighborhood. However, some people like to try new things and they may want to find a neighborhood that is very different. It would be great if they could have the neighborhoods narrowed down to their desired needs. In our case we will assume the individual is being relocated from New York to Toronto.

Data Acquistion

- Venues data
  - Foursquare venue data API (https://api.foursquare.com/v2/venues/explore?)
- Toronto Neighborhood data
  - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- New York neighborhood data
  - https://geo.nyu.edu/catalog/nyu_2451_34572
- Geospatial Data
  - http://cocl.us/Geospatial_data/Geospatial_Coordinates.csv

Data Cleaning

- Removed unassigned data from the New York and Toronto data set
- Removed newline characters from the Toronto data set
- Normalized the columns in both data sets

One way to compare cities is to use a venue-based similarity measure. For example, the person being relocated may require that there are coffee shops, parks and restaurants all within walking distance in their neighborhood.

In order to that, we would want to know what type of venues people are most frequently visited in a specific neighborhood. For example, if people are going to parks a lot, then parks must be a strong attribute of the neighborhood. Conversely, if there is a coffee shop in the neighborhood but it is not frequented as much,then coffee shops would not be an attribute of the neighborhood.

In this project I will explore the use of incidentally generated social network data for the folksonomic characterization of cities by the type of amenities located within them. Folksonomy is also known as collaborative tagging, social classification, social indexing, and social tagging. In our example we will add the venue information in the individuals current New York neighborhoods to the venue information of the Toronto neighborhoods. I will build 8 neighborhood clusters using the 10 top most frequently visited venues per neighborhood.

I will use the mean frequency of visits to determine the to 10 venues. I will run a k-means to cluster the neighborhood into 8 clusters. All Toronto neighborhoods that are in the same cluster as their New York neighborhood would be considered similar. All other cluster would be dissimilar. We added other New York neighborhoods into the analysis since the person may want a neighborhood that is like another New York neighborhood they wish to move into.

Clustering

I will run a k-means to cluster the neighborhood into 8 clusters. All Toronto neighborhoods that are in the same cluster as their New York neighborhood would be considered similar. All other cluster would be dissimilar. We added other New York neighborhoods into the analysis since the person may want a neighborhood that is like another New York neighborhood they wish to move into.

Cluster 1

This cluster contains only New York Neighborhoods that have a large amount of Italian restaurants.

8	NY_Upper East Side	-73.960508	0	Italian Restaurant	Exhibit	Coffee Shop	Art Gallery	Bakery	Juice Bar
9	NY_Yorkville	-73.947118	0	Italian Restaurant	Bar	Gym	Coffee Shop	Pizza Place	Diner
10	NY_Lenox Hill	-73.958860	0	Sushi Restaurant	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Pizza Place	Sporting Goods Shop
12	NY_Upper West Side	-73.977059	0	Italian Restaurant	Bar	Coffee Shop	Indian Restaurant	Bakery	Vegetarian / Vegan Restaurant

Cluster 2

This cluster contains only New York Neighborhoods that have a large amount of Italian restaurants.

	City	Longitude	Cluster Labels	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
70	TO	-79.377529	1	Park	Trail	Playground	Women's Store	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop

Cluster 3

This cluster contains only New York Neighborhoods that have a large amount of Italian restaurants.

	City	Longitude	Cluster Labels	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
63	TO	-79.39442	2	Airport Terminal	Airport Service	Airport Lounge	Boat or Ferry	Airport Gate	Boutique	Harbor / Marina	Sculpture Garden	Airport Food Court
64	TO	-79.39442	2	Airport Terminal	Airport Service	Airport Lounge	Boat or Ferry	Airport Gate	Boutique	Harbor / Marina	Sculpture Garden	Airport Food Court

Cluster 4

This is an all Toronto neighborhood cluster that the primary venue is coffee shops.

	City	Longitude	Cluster Labels	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
26	NY	-73.963896	3	Coffee Shop	Bookstore	Park	American Restaurant	Sandwich Place	Deli / Bodega	Food Truck	Burger Joint
40	TO	-79.360636	3	Coffee Shop	Café	Pub	Park	Bakery	Restaurant	Breakfast Spot	Theater
41	TO	-79.360636	3	Coffee Shop	Café	Pub	Park	Bakery	Restaurant	Breakfast Spot	Theater
44	TO	-79.375418	3	Coffee Shop	Restaurant	Café	Hotel	Bakery	Italian Restaurant	Park	Breakfast Spot

Cluster 5

This cluster contains a unique Toronto neighborhood that has cafés, grocery stores and restaurants

	City	Longitude	Cluster Labels	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
47	TO	-79.422564	4	Café	Grocery Store	Park	Italian Restaurant	Diner	Athletics & Sports	Nightclub	Coffee Shop

Cluster 6

This cluster contains a unique neighborhood that seems to cater to commuters.

	City	Longitude	Cluster Labels	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
37	NY	-73.974052	5	Bar	Park	Boat or Ferry	Baseball Field	Basketball Court	Cocktail Bar	Coffee Shop	Heliport

Cluster 7

This cluster contains Toronto and New York neighborhoods that have a wide variety of venues.

	City	Longitude	Cluster Labels	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
0	NY	-73.910660	6	Discount Store	Coffee Shop	Yoga Studio	Gym	Steakhouse	Big Box Store	Supplement Shop
11	NY	-73.949168	6	Sandwich Place	Coffee Shop	Japanese Restaurant	Dry Cleaner	Deli / Bodega	Outdoors & Recreation	Bus Line

Cluster 8

This cluster contains Toronto and New York neighborhoods that have a a lot of food options.

	City	Longitude	Cluster Labels	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
1	NY	-73.994279	7	Chinese Restaurant	American Restaurant	Cocktail Bar	Dim Sum Restaurant	Bubble Tea Shop	Hotpot Restaurant	Ice Cream Shop	Bakery
2	NY	-73.936900	7	Café	Bakery	Spanish Restaurant	Grocery Store	Mobile Phone Shop	Gym	Sandwich Place	Clothing Store
3	NY	-73.921210	7	Mexican Restaurant	Café	Pizza Place	Lounge	Chinese Restaurant	Restaurant	Bakery	Caribbean Restaurant
4	NY	-73.949688	7	Mexican Restaurant	Café	Coffee Shop	Pizza Place	Deli / Bodega	Sushi Restaurant	School	Sandwich Place

Conclusions

The clustering did show that in general New York (Manhattan) and Toronto are very similiar when looking at the top venues but also that they do have some unique neighborhoods.

To improve the clusters I think you would want to bring in criminal activity data and housing data such as median house price and types of housing. Also you could bring school data especially for families that are being located.