Introduction

The problem statement: Help find neighborhoods in Toronto, ON that are similar to a given neighborhood in New York.

Who will be interested: This can be used by relocation agencies to help people who are looking to move to Toronto from New York.

Background: Toronto is a fast growing city in North America. Having said so, there are loads of new opportunities that are sprawling across Toronto that is of interest to people. This has led to a recent growth in immigration to the city. As per this site; of all immigrants in Ontario, 7 out of 10 lived in Toronto. This has led to many immigration services that cater to the city of Toronto,ON. This has also led to a rich culturally diverse Toronto. Huffington Post also came up with this widely read article highlighting the migration from New York to Toronto. Upon reading the article, one thing that strikes is that people do not just move because of new opportunity, but also factor in the lifestyle. Rising cultural diversity of Toronto has been a magnet for people of culturally-diverse New York. This is something relocation agencies cash upon. They have come up with lots of services that help people explore their future neighborhood. We intend to come up with a smart solution that will provide an impetus to this effort. Our algorithm will factor in various lifestyle pointers of a neighborhood and use machine learning to find localities that have matching cultural and lifestyle offerings. This will help find neighborhoods in Toronto, ON that are similar to a given neighborhood in New York.

Data

The data for 2 cities: Toronto and New York will be used to compare neighborhoods, and then we rank the neighborhoods from the selected boroughs for the cities. We will use the previous assignments to retrieve the neighborhood and geo coordinates for New York and Toronto. We will use Machine Learning (unsupervised learning) to cluster the neighborhoods from the two cities. Data sources are *newyork_data.json* & the Toronto Wiki page.

Data sources:

- New York city data: https://gist.github.com/0cd2d7265973edb82dbf9ef2486c9ca1
- Toronto city data: Web scraping https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- FourSquare API to get neighbourhood details for each city

Methodology

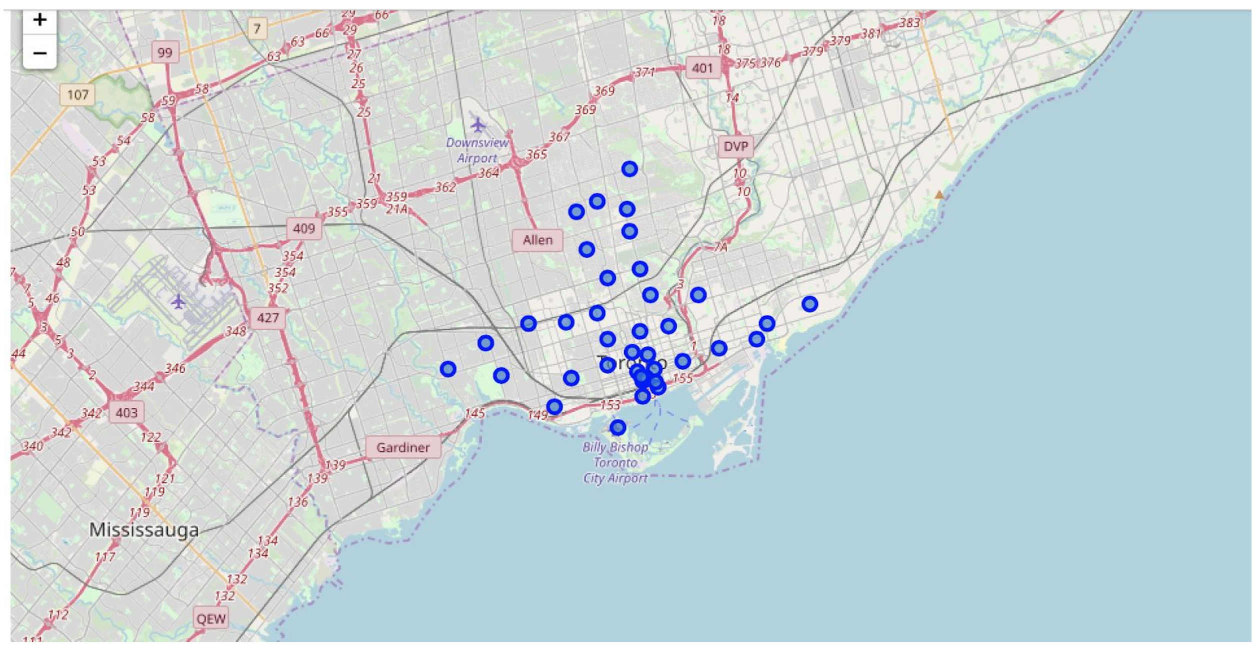
We will follow the following steps:‘

1. Create a dataset that holds the Geo-coordinates for Toronto's neighborhood

   Use BeautifulSoup package to perform web scraping on Toronto Wiki page to build the following dataset.

	Borough	Neighbourhood	Latitude	Longitude
37	East Toronto	The Beaches	43.676357	-79.293031
41	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188
42	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572
43	East Toronto	Studio District	43.659526	-79.340923
44	Central Toronto	Lawrence Park	43.728020	-79.388790

   Then, we will further explore the city of Toronto. The following unclustered map view is a representation of it.



2. Use Foursquare APIs to get venues for each of the Toronto's neighborhoods

Next, we use the Foursquare's venues API to get nearby venues. We will capture the following data points: 'Neighbourhood', 'Neighbourhood Latitude', 'Neighbourhood Longitude', 'Venue', 'Venue Latitude', 'Venue Longitude', and 'Venue Category'. A snapshot of the dataset looks like the following. This dataset will form the basis for k-means clustering which we will illustrate later.

	Neighbourhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	...	Theme Restaurant	Thrift / Vintage Store	Toy / Game Store	Trail	Tra Stati
0	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	1	
1	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
2	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
3	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	
4	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	

5 rows × 194 columns

3. Sort through the data to identify top 10 common venue categories for each of the Toronto's neighborhoods

   Dataset snapshot is given below for your reference.

	Neighbourhood	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
0	Adelaide,King,Richmond	Steakhouse	Asian Restaurant	Café	Pizza Place	Hotel	Neighborhood	Lounge	Burger Joint	Seafood Restaurant	Smoke Shop
1	Berczy Park	Seafood Restaurant	Coffee Shop	Cocktail Bar	Beer Bar	Café	Farmers Market	Greek Restaurant	Jazz Club	Basketball Stadium	Fish Market
2	Brockton,Exhibition Place,Parkdale Village	Coffee Shop	Breakfast Spot	Café	Climbing Gym	Stadium	Burrito Place	Restaurant	Caribbean Restaurant	Pet Store	Bakery
3	Business Reply Mail Processing Centre 969 Eastern	Yoga Studio	Fast Food Restaurant	Park	Comic Shop	Pizza Place	Butcher	Burrito Place	Recording Studio	Restaurant	Brewery
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Lounge	Airport Service	Airport Terminal	Harbor / Marina	Sculpture Garden	Airport Food Court	Airport Gate	Bar	Boat or Ferry	Boutique

4. Perform the above for New York's neighborhoods

   Same steps are performed for the city of New York. For conciseness, the final datasets will look something the the following

	Neighbourhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Arcade	Arepa Restaurant	...	Warehouse Store	Waste Facility	Wate
0	Wakefield	0	0	0	0	0	0	0	0	0	...	0	0	
1	Wakefield	0	0	0	0	0	0	0	0	0	...	0	0	
2	Wakefield	0	0	0	0	0	0	0	0	0	...	0	0	
3	Wakefield	0	0	0	0	0	0	0	0	0	...	0	0	
4	Wakefield	0	0	0	0	0	0	0	0	0	...	0	0	

5 rows × 380 columns

	Neighbourhood	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
0	Allerton	Pizza Place	Pharmacy	Spa	Deli / Bodega	Supermarket	Department Store	Fried Chicken Joint	Breakfast Spot	Bus Station	Gas Station
1	Annadale	Pizza Place	Park	Sports Bar	Restaurant	Food	Diner	Train Station	Pharmacy	Field	Event Space
2	Arden Heights	Pharmacy	Coffee Shop	Pizza Place	Bus Stop	Yoga Studio	Financial or Legal Service	Factory	Falafel Restaurant	Farm	Farmers Market
3	Arlington	Bus Stop	Deli / Bodega	American Restaurant	Boat or Ferry	Food	Grocery Store	Fish Market	Farm	Farmers Market	Fast Food Restaurant
4	Arrochar	Deli / Bodega	Pizza Place	Italian Restaurant	Bus Stop	Athletics & Sports	Middle Eastern Restaurant	Bagel Shop	Liquor Store	Supermarket	Hotel

5. This step will be the input. Select a neighborhood in NY for which we are looking for lookalikes in Toronto. Here, we chose Chelsea, NY.

	Neighbourhood	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	...	Veterinarian	Video Store	Warehouse Store	Waste Facility
38	Chelsea	0.0	0.0	0.0	0.0	0.0	0.0	0.029412	0.0	0.0	...	0.0	0.0	0.0	0.0

6. Use K-means on dataset that has all Toronto's neighborhoods plus this neighborhood. Then find the cluster which has the NY neighborhood in it and list all Toronto neighborhoods there. Following dataset shows each neighborhood in Toronto with its assigned cluster label.

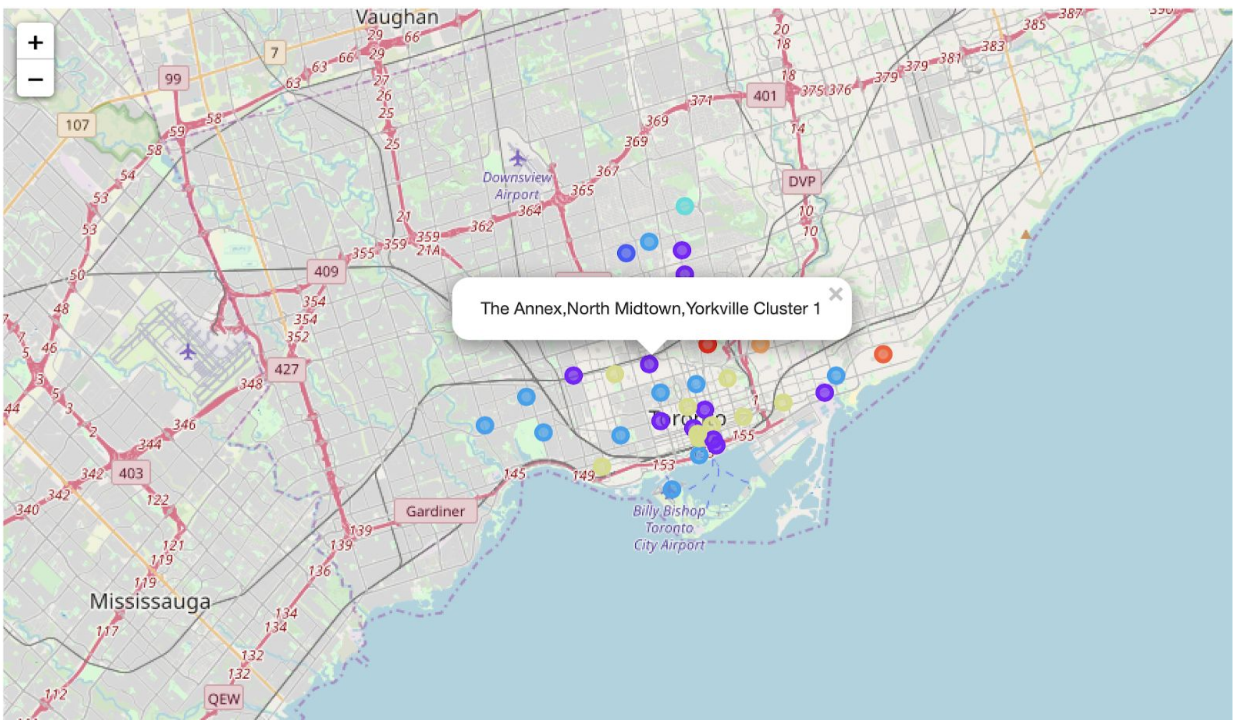
	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8 C
0	East Toronto	The Beaches	43.676357	-79.293031	9	Neighborhood	Other Great Outdoors	Health Food Store	Trail	Pub	Cuban Restaurant	Ethiopian Restaurant	E Re
1	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	8	Greek Restaurant	Ice Cream Shop	Italian Restaurant	Yoga Studio	Bookstore	Restaurant	Spa	
2	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	3	Park	Pet Store	Ice Cream Shop	Liquor Store	Sandwich Place	Burger Joint	Fast Food Restaurant	
3	East Toronto	Studio District	43.659526	-79.340923	7	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Middle Eastern Restaurant	Stationery Store	
4	Central Toronto	Lawrence Park	43.728020	-79.388790	4	Bus Line	Park	Swim School	Dance Studio	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant	D Re

Results

For Chelsea, NY, we found 10 lookalike neighborhoods in Toronto. They are listed below:

```
data_index
0                                     Adelaide,King,Richmond
1                                                Berczy Park
3      Business Reply Mail Processing Centre 969 Eastern
7                 Chinatown,Grange Park,Kensington Market
11                                                Davisville
12                                          Davisville North
15                             Dovercourt Village,Dufferin
30                                 Ryerson,Garden District
32                          Stn A PO Boxes 25 The Esplanade
34                       The Annex,North Midtown,Yorkville
```

Also, if we want to visualize them in maps, we can see them as below (marked in purple)



Discussion

We find that k-means is a very powerful technique to compare neighborhoods. For this algorithm to be successful, we need as much data as possible available for both the cities that speaks about the lifestyle of those areas. Then, we can run ML unsupervised K-means algorithm to find lookalikes.

Please note that this algorithm can be generalized to compare any two cities, given the data is available as mentioned above.

Conclusion

For Chelsea,NY we were able to find lookalike neighborhoods in Toronto, ON. We based it off several lifestyle parameters for the neighborhoods and leveraged Foursquare APIs for the same. This is a powerful algorithm that will help relocation agencies zero in on localities that will fit the bill for their clients. If correctly incorporated in their client onboarding process, this algorithm can have potential positive impacts on the bottom line.