Looking for the right neighborhoods in two major cities in North America: Toronto vs NYC

Introduction/Business Problem

Moving and relocating in a new city and environment can be a dauting experience. We usually look for certain features to enjoy living in a given neighborhood. How to do this when one is a complete foreigner to a city?

In this scenario, a candidate received a job offer from a company who gives her the chance to work in either their Toronto or New York offices. She is interested to know if these cities offer what she looks for in a potential neighborhood to settle in: parks to enjoy walks, and Chinese restaurants, her favorite kind of food.

We will try to identify neighborhoods with these features so it can help the new employee choose a potential neighborhood which fits her preferences.

Data

For the city of Toronto, list of neighborhoods will be retrieved from a Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), an online encyclopedia. The list of venues and places of interest will come from FourSquare, a geolocation platform.

For the city of New York, list of neighborhoods will be retrieved from an existing online dataset (https://cocl.us/new_york_dataset). The list of venues and places of interest will also come from FourSquare.

Most of the data will be usable as is. The Wikipedia will be pulled out using the BeautifulSoup library.

For both cities, we will list the existing neighborhoods, list the various venues established in them, identify what kind of venues are mostly present in each neighborhood. We will then search for neighborhoods with both parks and Chinese restaurants and see if it yields any results.

Methodology

The following steps were undertaken to arrive to an answer.

Toronto

Data was scraped from Wikipedia to establish a table with a list of neighborhoods in Toronto:

	Postal code	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

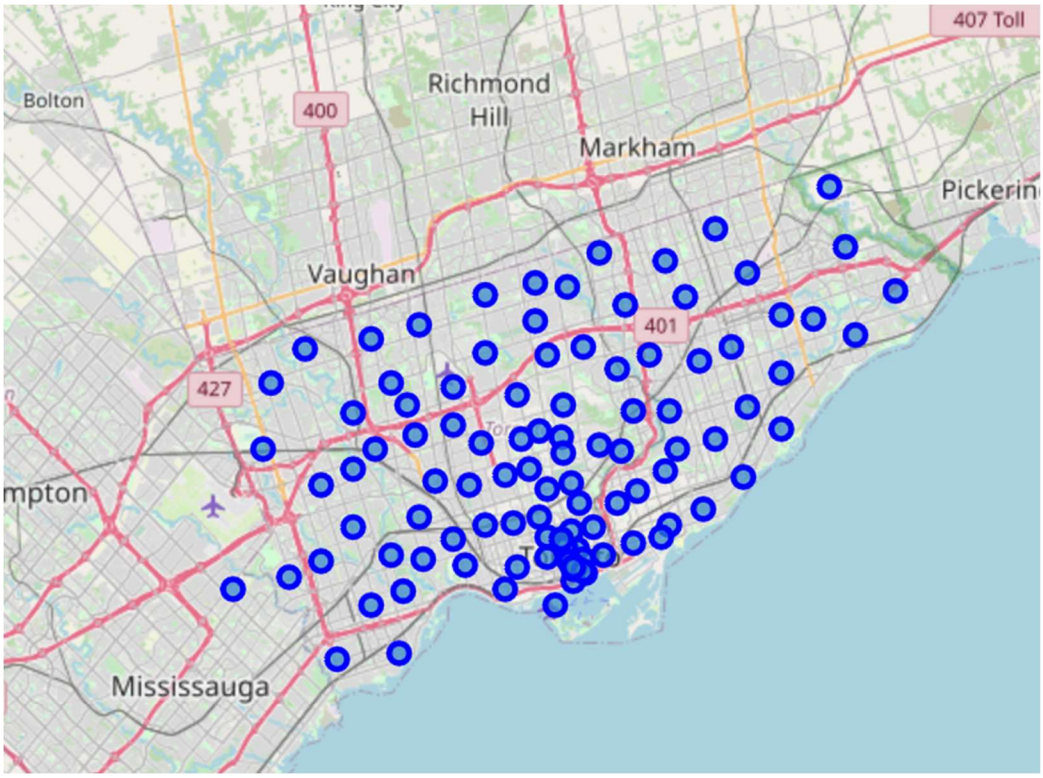
This list was cleaned to ignore “not assigned” values:

	Postal code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Neighborhoods were then grouped by postal codes:

	Postal code	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...
11	M1R	Scarborough	Wexford, Maryvale

A map of all neighborhoods was generated:



Given the large number of neighborhoods, we limited the search to core Toronto neighborhoods:

	Postalcode	Borough	Neighborhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

For which we could generate a close-up map:

Based on information retrieved from FourSquare geolocation platform, we could identify the number of venues, or sites of interests, for each neighborhood:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	58	58	58	58	58	58
Brockton, Parkdale Village, Exhibition Place	23	23	23	23	23	23
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	18	18	18	18	18	18

The type of venue was retrieved for each of them:

	Neighborhood	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop	Bakery
0	The Beaches	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	The Beaches	0	0	0	0	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	0	0	0	0
3	The Beaches	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

This information would be the basis for generating the top 10 most common venues in each neighborhood:

	Neighborhood	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	Berczy Park	Coffee Shop	Cocktail Bar	Pharmacy	Bakery	Seafood Restaurant	Farmers Market	Restaurant	Beer Bar	Cheese Shop	Café
1	Brockton, Parkdale Village, Exhibition Place	Café	Coffee Shop	Breakfast Spot	Yoga Studio	Stadium	Burrito Place	Restaurant	Climbing Gym	Performing Arts Venue	Bakery
2	Business reply mail Processing Centre, South C...	Light Rail Station	Comic Shop	Garden	Brewery	Burrito Place	Spa	Farmers Market	Fast Food Restaurant	Butcher	Restaurant
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport	Boat or Ferry	Plane	Coffee Shop	Rental Car Location	Sculpture Garden	Boutique	Harbor / Marina	Airport Food Court
4	Central Bay Street	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Japanese Restaurant	Bubble Tea Shop	Bar	Thai Restaurant	Salad Place	Burger Joint
5	Christie	Grocery	Café	Park	Restaurant	Candy Store	Italian	Diner	Baby Store	Coffee Shop	Nightclub

From this data frame we can find which neighborhood contained the features we were looking for (parks and Chinese restaurants).

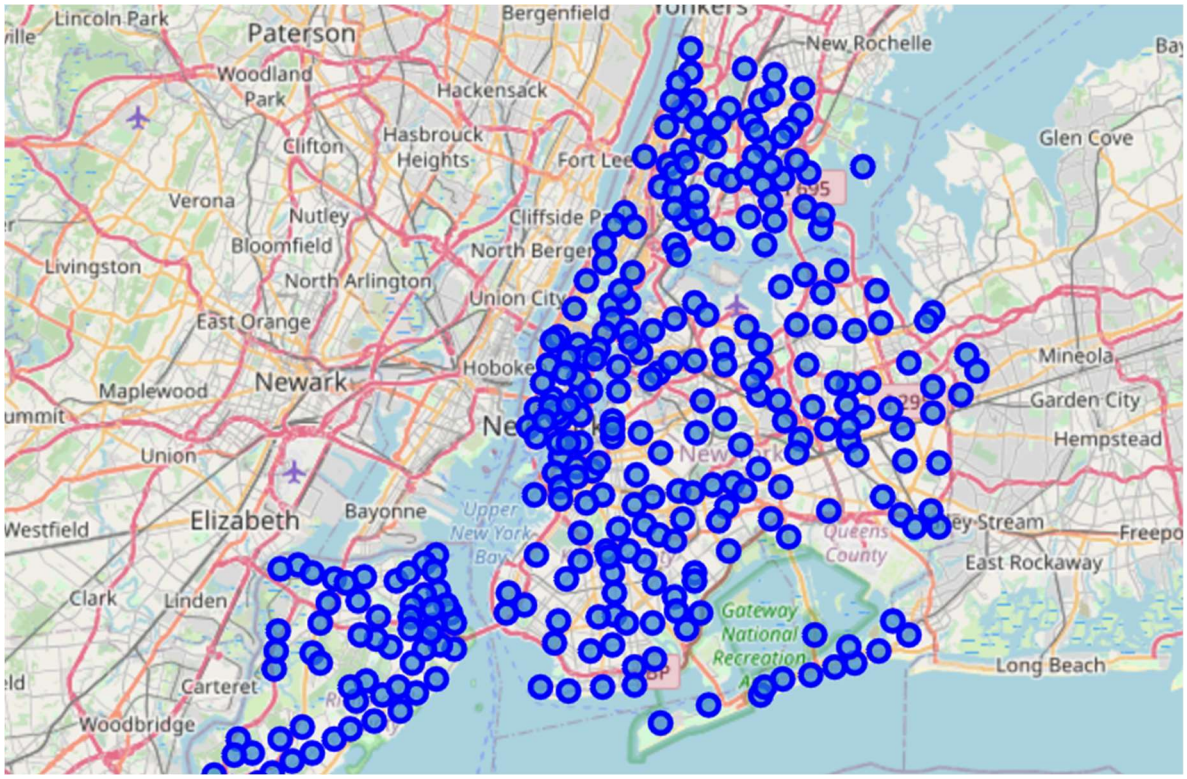
We then followed similar steps to get this information in New York City

New York City

From the online dataset (https://cocl.us/new_york_dataset), we generated a data frame with a list of NYC neighborhoods:

	Borough	Neighborhood	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

From this, we could generate a map showing the various NYC neighborhoods:



Given the large number of neighborhoods, we narrowed it down to Manhattan:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Based on information retrieved from FourSquare geolocation platform, we could identify the number of venues, for each neighborhood:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Battery Park City	69	69	69	69	69	69
Carnegie Hill	92	92	92	92	92	92
Central Harlem	45	45	45	45	45	45
Chelsea	100	100	100	100	100	100
Chinatown	100	100	100	100	100	100
Civic Center	100	100	100	100	100	100
Clinton	100	100	100	100	100	100
East Harlem	42	42	42	42	42	42
East Village	100	100	100	100	100	100
Financial District	100	100	100	100	100	100

The type of venue was retrieved for each of them which was the basis to generate the 10 most common venues for each neighborhood:

	Neighborhood	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	Battery Park City	Park	Hotel	Coffee Shop	Gym	Memorial Site	Boat or Ferry	Italian Restaurant	BBQ Joint	Gourmet Shop	Sandwich Place
1	Carnegie Hill	Coffee Shop	Café	Italian Restaurant	Pizza Place	Yoga Studio	Gym	Bar	Bakery	Gym / Fitness Center	Bookstore
2	Central Harlem	African Restaurant	Seafood Restaurant	Chinese Restaurant	American Restaurant	Bar	French Restaurant	Bookstore	Boutique	Market	Southern / Soul Food Restaurant
3	Chelsea	Coffee Shop	Art Gallery	American Restaurant	French Restaurant	Bakery	Ice Cream Shop	Bookstore	Hotel	Bar	Café

Results

Based on the final data frames generated of each city, we then could search which neighborhoods had parks and Chinese restaurants as common venues. It yielded the following results:

	Neighborhood	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
21	North Toronto West, Lawrence Park	Clothing Store	Sporting Goods Shop	Coffee Shop	Yoga Studio	Gym / Fitness Center	Fast Food Restaurant	Diner	Mexican Restaurant	Park	Chinese Restaurant

	Neighborhood	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
15	Inwood	Mexican Restaurant	Lounge	Café	Restaurant	Park	Bakery	Frozen Yogurt Shop	Caribbean Restaurant	Spanish Restaurant	Chinese Restaurant
37	Washington Heights	Café	Bakery	Deli / Bodega	Mobile Phone Shop	Chinese Restaurant	Bank	Sandwich Place	Tapas Restaurant	Supermarket	Park

Based on our search, we found that the following neighborhoods could be suitable in our scenario:

- Toronto: North Toronto West, Lawrence Park
- NYC, Manhattan: Inwood or Washington Heights

Discussion

In our analysis, we limited the number of venues to 100. If we increased that number, it may have yielded in different results, or at least, in more accurate ones. We also limited the search: Toronto was limited in

core neighborhoods around downtown, and New York City was limited to Manhattan. An interesting exercise may have been to search in the outer neighborhoods and suburbs.

We also could have done a more refined search, for example, by restricting the results to neighborhoods where “parks” and “Chinese restaurants” were in the top 5 common venues. But based on the limits we imposed ourselves on the neighborhoods we were looking at, it would have resulted in zero answers.

Conclusion

Searching through large datasets can give interesting answers which can help make a decision. An important amount of work had to be done to clean the data and make it usable. In our scenario of looking a neighborhood suitable to our candidate, we could narrow it down to 1 neighborhood in Toronto, and 2 in Manhattan, New York City where parks and Chinese restaurants can be commonly found. It now can be adapted to answer similar type of questions for other neighborhoods or different features.