

New York

City



Toronto

City of Toronto



**Compare the neighborhoods of two cities of New York and Toronto.
August 10th 2019**

Introduction

Background

The City of New York, usually called either New York City (NYC) or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km²), New York is also the most densely populated major city in the United States. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, business services, environmental innovation, food services, and tourism.

Problem

New York City and the city of Toronto are financial capitals of their respective countries of USA and Canada. Both of the city host multi-cultural environments. New York and Toronto are huge cities. We are going to use data to find out how similar or dissimilar they are to each other. This information is going to be useful for anyone who plans on visiting in these cities.

Data acquisition and cleaning

Data Sources

These two cities are huge and the number of factors/features that can be compared between them are too many. We are going to limit the number of data point that we are going to use. We will use sources like rent jungle to scrap of average rental information in these two cities.

But we will be primarily using Foursquare for getting the location information of venues (features).

We will use some or all of these features. The features we are shortlisting are

- Monument / Landmark
- Theaters / art galleries
- Coffee
- Food
- Parks / Trails
- Nightlife

- self defense
- Robbery/Assault

Here are the data and its sources that would be used in this solution.

1. **Neighborhood information:** For base map and location.

a. **Toronto**

Source: For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. We will scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe. The below diagram shows Neighborhood data and coordinates information merged together.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636
3	M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494

(sample data)

b. **New York city**

Source: The 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

	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

(sample data)

2. Foursquare Data:

We make calls to the Foursquare API for this purpose of getting feature information. We will URL to send a request to the API to search for a specific type of venues like, to explore each particular venue (Monument / Landmark, Theaters / art galleries, Coffee, Food, Parks / Trails and Nightlife. Also, we will use the visualization library, Folium, to visualize the results.

For example:

Shown below is the sample dataset for venue of Monument / Landmark for both New York and Toronto respectively.

a. New York

	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country
0	Nathan Hale Monument	Monument / Landmark	City Hall Park	40.712010	-74.005686	[{"label": "display", "lat": 40.71201038109373...	84	10007	US	New York	NY	United States
1	African Burial Ground National Monument	Monument / Landmark	290 Broadway	40.714990	-74.005530	[{"label": "display", "lat": 40.71498975306652...	255	10007	US	New York	NY	United States
	Mirman.		291									

(sample data)

b. Toronto city

	name	categories	lat	lng	labeledLatLngs	distance	cc	country	formattedAddress	address	crossStreet	city	state
1	Sir Adam Beck Monument	Other Great Outdoors	43.651011	-79.387152	[{"label": "display", "lat": 43.65101102690917...	328	CA	Canada	[250 University Ave (Queen St), Toronto ON, Ca...	250 University Ave	Queen St	Toronto	ON
2	Landmark Jewellery	Jewelry Store	43.647640	-79.382905	[{"label": "display", "lat": 43.64764, "lng": ...	784	CA	Canada	[121 King St W, Concourse/underground (King An...	121 King St W, Concourse/underground	King And York	Toronto	ON
3	Egerton Ryerson Monument	Monument / Landmark	43.657864	-79.378758	[{"label": "display", "lat": 43.65786400153166...	807	CA	Canada	[Toronto ON, Canada]	NaN	NaN	Toronto	ON

(sample data)

3. Average rental information:

Source: The data for rents of 1bedroom, 2 bedroom and 3-bedroom averages are available on rentjungle.com. The data for Toronto is limited to few years when compared with New York city. So we will only use data for year that are available for both cities.

<https://www.rentjungle.com/average-rent-in-toronto-rent-trends/>

<https://www.rentjungle.com/average-rent-in-new-york-rent-trends/>

Toronto	New York City
---------	---------------

Month	All Beds	1 Beds	2 Beds		Month	All Beds	1 Beds	2 Beds	
4/2016	1,338	1,295	1,561		4/2016	3,224	2,832	3,618	
5/2016	1,336	1,325	1,541		5/2016	3,248	2,852	3,646	
6/2016	1,339	1,260	1,530		6/2016	3,262	2,848	3,631	
7/2016	1,339	1,268	1,493		7/2016	3,271	2,852	3,653	
8/2016	1,336	1,233	1,489		8/2016	3,253	2,834	3,635	
9/2016	1,325	1,221	1,502		9/2016	3,226	2,837	3,619	
10/2016	1,317	1,248	1,414		10/2016	3,208	2,829	3,618	
11/2016	1,319	1,252	1,403		11/2016	3,111	2,747	3,522	
12/2016	1,362	1,253	1,571		12/2016	3,064	2,700	3,469	
1/2017	1,320	1,233	1,489		1/2017	3,076	2,698	3,461	
2/2017	1,296	1,224	1,489		2/2017	3,073	2,698	3,457	
3/2017	1,311	1,265	1,448		3/2017	3,061	2,711	3,454	
4/2017	1,328	1,276	1,470		4/2017	3,074	2,732	3,510	
5/2017	1,351	1,251	1,495		5/2017	3,109	2,765	3,516	

(sample data)

Data Analysis

Due to limitation of how many calls Four square allows, the problem was converted to selecting top 10 venues in each of the cities.

These are not the same as we had set out in the beginning.

(1690, 7)

18] :	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront,Regent Park	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront,Regent Park	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront,Regent Park	43.65426	-79.360636	Toronto Cooper Koo Family Cherry St YMCA Centre	43.653191	-79.357947	Gym / Fitness Center
3	Harbourfront,Regent Park	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Harbourfront,Regent Park	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot

From the Foursquare lab in the previous module, we know that all the information is in the items key. Before we proceed, let's borrow the `get_category_type` function from the Foursquare lab.

The clusters reveal a very diverse set of venues that do not match the ones we had set out in the beginning.

Examine Clusters

Now, you can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, you can then assign a name to each cluster.

Cluster 1

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] + list(range(4, toronto_merged.shape[1]))]]
```

	Borough	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	10th Most Common Venue
18	Central Toronto	-79.416936	0	Garden	Home Service	Music Venue	Diner	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant

Cluster 2

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[1] + list(range(4, toronto_merged.shape[1]))]]
```

	Borough	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	10th Most Common Venue
17	Central Toronto	-79.38879	1	Photography Studio	Park	Bus Line	Swim School	Dive Bar	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant	Electronics Store

Cluster 3

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[1] + list(range(4, toronto_merged.shape[1]))]]
```

	Borough	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	10th Most Common Venue
0	Downtown Toronto	-79.360636	2	Coffee Shop	Bakery	Park	Mexican Restaurant	Gym / Fitness Center	Breakfast Spot	Restaurant	Café	Pub	Theater
1	Downtown Toronto	-79.378937	2	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Fast Food Restaurant	Restaurant	Italian Restaurant	Ramen Restaurant	Diner	Bookstore
2	Downtown Toronto	-79.375418	2	Coffee Shop	Restaurant	Café	Hotel	Italian Restaurant	Clothing Store	Gastropub	Cosmetics Shop	Beer Bar	Breakfast Spot

Solution to the problem

These two cities are similar in weather and the kind of venues that are available to both the cities. New York being the larger of the two there are more options available. But having said that Toronto is not far lagging behind and has higher percentage of tourism revenue compared to the city that is half the size of New York city