

FINDING THE BEST CITY FOR DIFFERENT VENUE CATEGORIES BASED ON POPULARITY

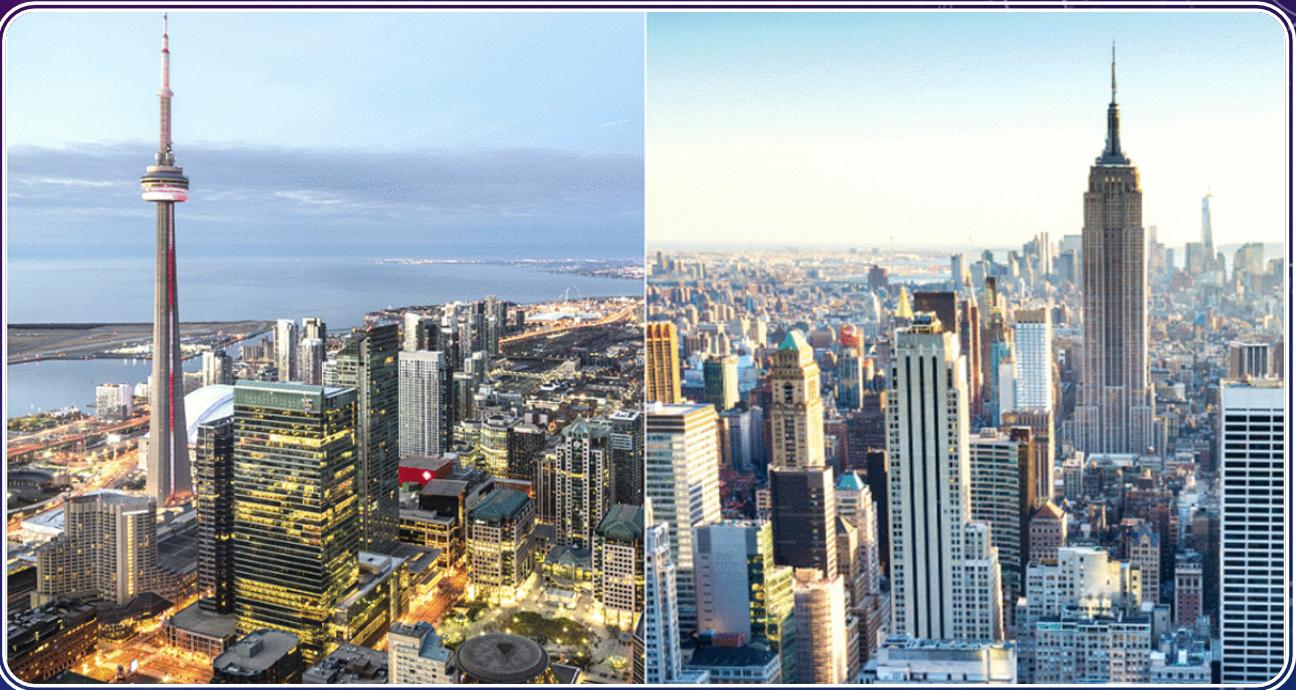
DATA SCIENCE CAPSTONE PROJECT:



BACKGROUND:

- New York City, NY and Toronto, ON are often thought of as “sister” cities due to their common traits of being the biggest cities in their respective countries.

Both cities are celebrated for their fast paced environments, extensive food scene, cultural diversity and being large financial capitals. Thus, it makes sense that Toronto is often thought of as the “New York of Canada” and vice versa.



1.2 Business Problem:

- Both cities are vastly known for their large number of venues and venue types. Both cities are rich in their many venues; from pubs, restaurants, parks, museums and more. In fact, both cities draw hundreds of thousands of tourists every year to explore these venues.
- Therefore, it makes sense for a potential tourist who wants to visit either city (or both cities) to know which city's are more well known for which *types* of venues, so they can plan their visits accordingly.

For example: say a tourist, John, wants to see which city is more popular for Korean restaurants vs Indian restaurants. He uses our program to discover that New York is more popular for Korean, and Toronto is more popular for Indian. Therefore, when he's in Toronto, he visits an Indian restaurant and when he's in New York, he visits a Korean restaurant.

1.3 Solution:

- We will observe the types of venues that are present in neighborhoods of each city, and through data analysis, determine which types of venues are more popular in which neighborhoods; and each city as a whole.
- We will then define a few functions that will tell us, for any given venue type, which city is more popular for that venue type, and *if* the user wants, which specific neighborhoods within that city are most popular for that venue type.

2014 New York City Neighborhood Names



Description: This New York City Neighborhood Names point file was created as a guide to New York neighborhoods that appear on the web resource, "New York: A City of Neighborhoods." Label centroids were established at a 1:1,000 scale, but are ideally viewed at a 1:50,000 scale.

Publisher: [New York \(City\). Department of City Planning](#)

Collection: [Bytes of the Big Apple](#)

Place(s): [New York, New York, United States](#)

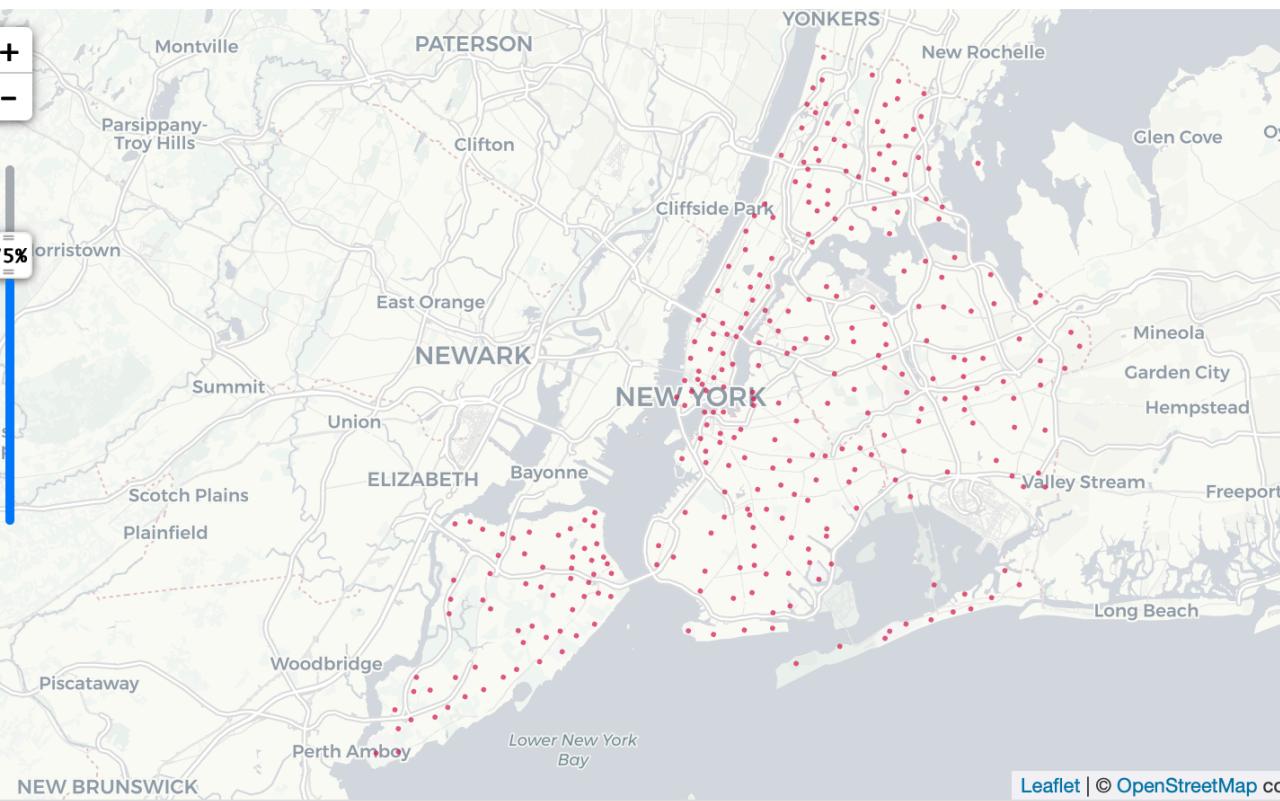
Subject(s): [Neighborhoods](#), [Neighborhood planning](#), and [Communities](#)

Format(s): [Shapefile](#)

Year(s): 2014

Held by: [NYU](#)

Preservation record: <http://hdl.handle.net/2451/34572>



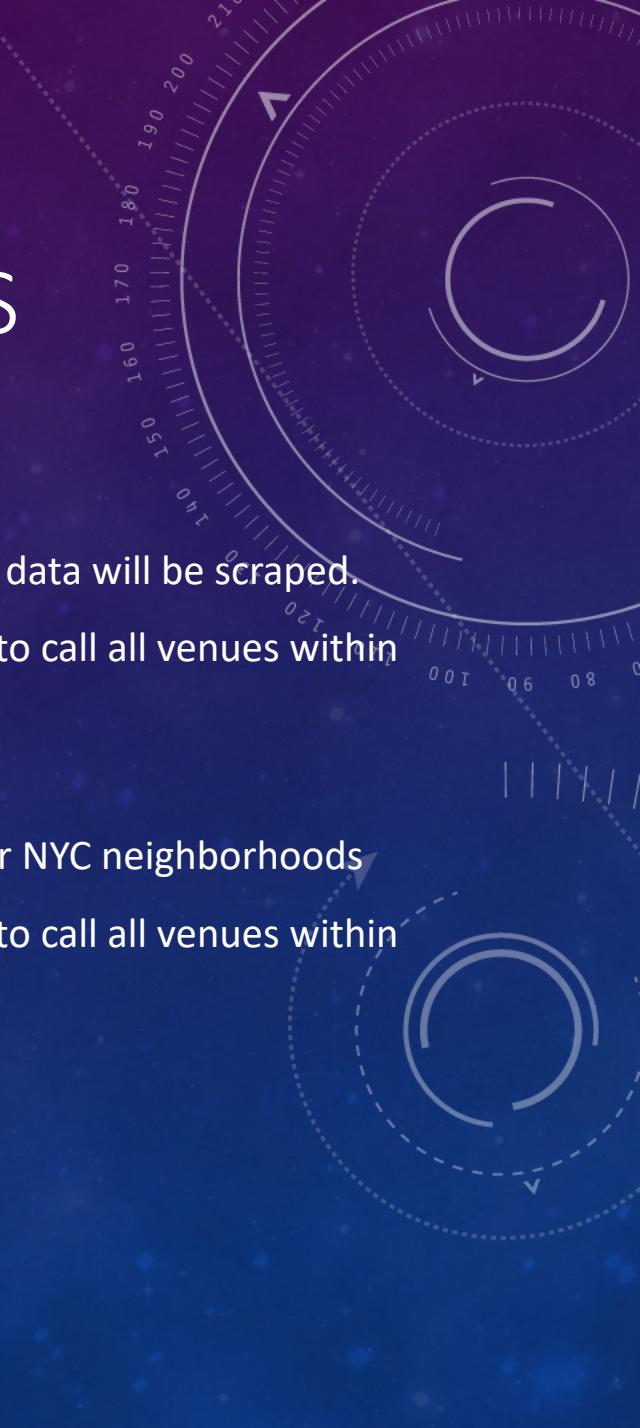
DATA SOURCES

Toronto Data:

- Toronto Wikipedia Page, this data will be scraped.
- FourSquare API will be used to call all venues within each neighborhood.

NYC Data:

- NYU downloaded data set for NYC neighborhoods
- FourSquare API will be used to call all venues within each neighborhood.



Data Acquisition and Cleaning:

Foursquare API was called to determine list of venues for each neighborhood within both cities.

Only the common venue categories between each city (202 total) were considered for comparison, the rest of the venues were dropped.

Using the common venue categories, the top 10 most frequently occurring of those venues for each neighborhood within each city was determined.

The result was a dataframe for each city which displayed the top 10 most frequent venue types per neighborhood (NOTE: for Toronto, each postal code region is analogous to 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
0	Allerton	Pizza Place	Chinese Restaurant	Deli / Bodega	Spa	Supermarket	Fried Chicken Joint	Donut Shop
1	Annadale	Pizza Place	American Restaurant	Dance Studio	Restaurant	Food	Train Station	Diner
2	Arden Heights	Pharmacy	Deli / Bodega	Coffee Shop	Bus Stop	Pizza Place	Flea Market	Factory
3	Arlington	Deli / Bodega	Coffee Shop	Bus Stop	Home Service	Boat or Ferry	Grocery Store	Yoga Studio
4	Arrochar	Bus Stop	Deli / Bodega	Italian Restaurant	Hotel	Middle Eastern Restaurant	Pharmacy	Bagel Shop

	PostalCode	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	M1E	Electronics Store	Yoga Studio	Discount Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space F
1	M1G	Korean BBQ Restaurant	Yoga Studio	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space F
2	M1H	Caribbean Restaurant	Thai Restaurant	Hakka Restaurant	Athletics & Sports	Bank	Farmers Market	Falafel Restaurant
3	M1J	Playground	Yoga Studio	Diner	Farmers Market	Falafel Restaurant	Event Space	Ethiopian Restaurant
4	M1M	American Restaurant	Motel	Yoga Studio	Discount Store	Farmers Market	Falafel Restaurant	Event Space F

Analysis / Determining the Better City per Venue Type:

- We now have a dataframe for each city, which now consisted of each neighborhood and the most frequently present venues.
- We created a function called `best_city` where we could input any venue category and it would return which city was more popular for that venue type and the neighborhoods within that city where the venue type was most frequent.
- This was determined using a ‘popularity score’ which looked at two things:
 - 1) how popular is that venue type in each city (i.e in how many neighborhoods is it appearing within the most frequently occurring venues?)
 - 2) for each neighborhood that it *is* appearing in the most frequently occurring venues, what is its position? (i.e is it the most popular? Second most? Third most? Etc) – this is called the ‘popularity index’



$$\text{popularity score} = \frac{\text{(sum of all popularity indexes)}}{\text{number of neighborhoods}}$$

PROOF OF CONCEPT:

Determining which city is best for Thai Restaurant, and getting the popularity score within that city +
The neighborhoods where Thai Restaurant is most frequent.

In [184]: `best_city('Thai Restaurant', True)`

Toronto has a higher popularity score for venue type with a score of 6.0
Here are the neighborhoods where Thai Restaurant is most frequent within the city:

Out[184]:

	PostalCode	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
2	M1H	Caribbean Restaurant	Athletics & Sports	Bank	Thai Restaurant	Yoga Studio	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market	Falafel Restaurant
30	M4M	Coffee Shop	Convenience Store	Thai Restaurant	Diner	Comfort Food Restaurant	Clothing Store	Cheese Shop	Seafood Restaurant	Café	Fish Market
34	M4S	Dessert Shop	Coffee Shop	Café	Italian Restaurant	Pizza Place	Toy / Game Store	Sandwich Place	Thai Restaurant	Seafood Restaurant	Sushi Restaurant
44	M5H	Coffee Shop	Steakhouse	Asian Restaurant	Thai Restaurant	Japanese Restaurant	Bar	Pizza Place	Seafood Restaurant	Vegetarian / Vegan Restaurant	Hotel
46	M5K	Coffee Shop	Deli / Bodega	Restaurant	Café	Salad Place	Bakery	Gym	Japanese Restaurant	Thai Restaurant	Bank
48	M5M	Sandwich Place	Italian Restaurant	Coffee Shop	Liquor Store	Thai Restaurant	Restaurant	Sushi Restaurant	Juice Bar	Pub	Comfort Food Restaurant
55	M5W	Cocktail Bar	Restaurant	Breakfast Spot	Moroccan Restaurant	Beer Bar	Coffee Shop	Thai Restaurant	Pub	Sporting Goods Shop	Jazz Club
56	M5X	Coffee Shop	Deli / Bodega	Café	Salad Place	Restaurant	American Restaurant	Seafood Restaurant	Japanese Restaurant	Sushi Restaurant	Thai Restaurant
70	M7A	Coffee Shop	Italian Restaurant	Restaurant	Thai Restaurant	Park	Sandwich Place	Bubble Tea Shop	Sushi Restaurant	Café	Yoga Studio

RESULTS:

We can also use our function to determine: of all the common venues between Toronto and NYC, which city has a higher popularity score for which venues.

	Category	Most Popular City
0	Accessories Store	New York
1	Adult Boutique	New York
2	Airport Terminal	Toronto
3	American Restaurant	Toronto
4	Arepas Restaurant	New York
5	Art Gallery	New York
6	Arts & Crafts Store	New York
7	Asian Restaurant	New York
8	Athletics & Sports	New York
9	BBQ Joint	Toronto
10	Bakery	New York

We can return a dataframe which lists out all of the venue types that are found in both cities, and the more popular city for that venue type in the column right next to it

Results:

- there are 69 / 202 venue types which are more popular in Toronto.
- there are 139 / 202 venue types which are more popular in NYC.
- the full list of these venues can be found in the report/on the notebook file.

Visualizing the cities based on the different venue categories:

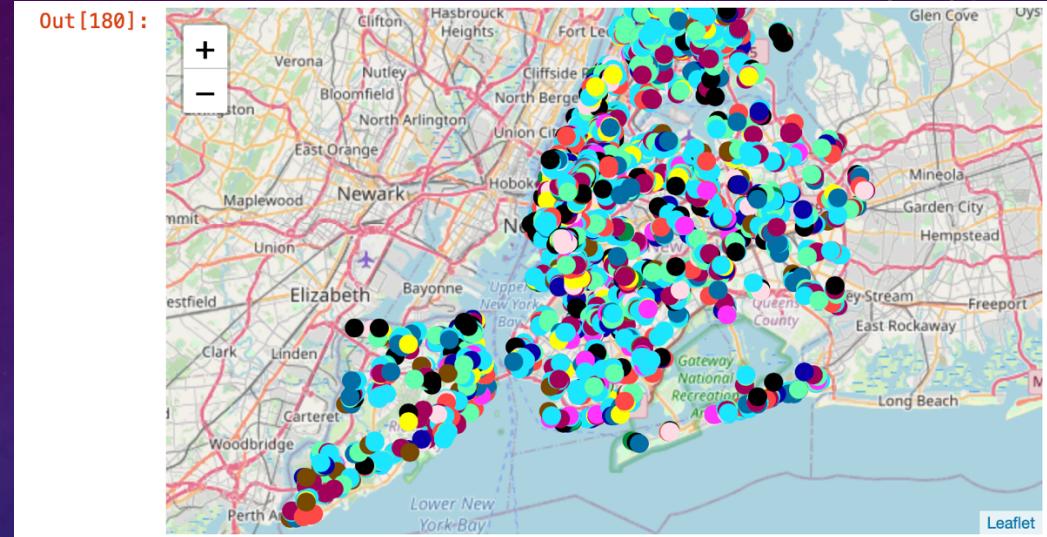
There are 202 common venue categories between NYC and Toronto. We want to display all the venues present in Toronto and NYC, but be able to easily distinguish what venue type/category each venue belongs to. We do this by using distinct colours for each marker that belongs to a certain venue type.

For simplicity, we grouped the 202 common venue categories into 13 unique categories, which we call the 'overall category':

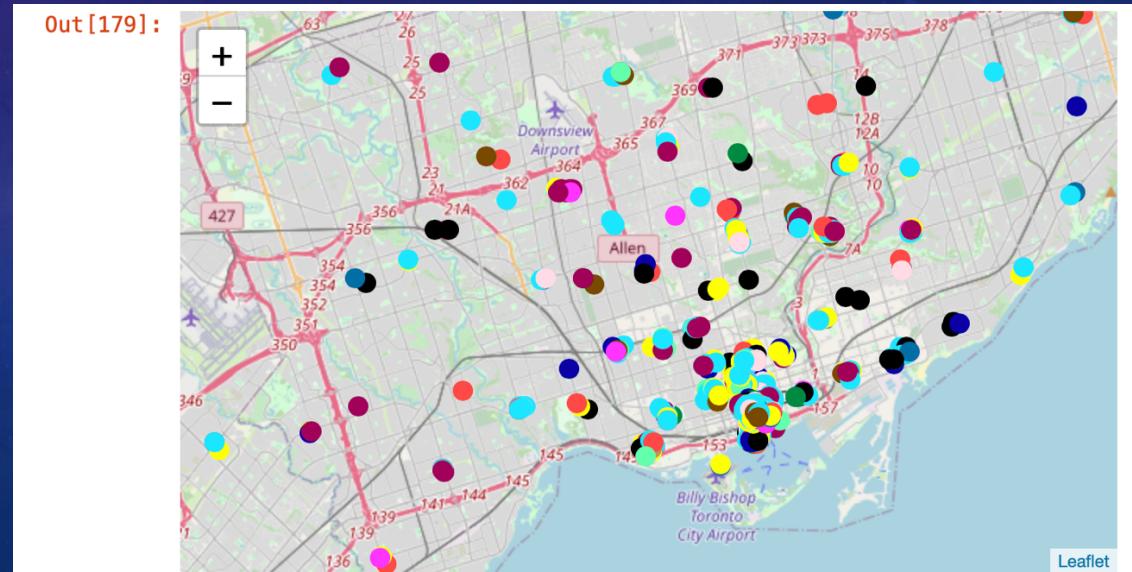
Overall categories = [Outdoor, Café, Restaurant, Health & Beauty, Fitness, Arts & Culture, Travel & Hospitality, Shopping, Nightlife, Services, Entertainment, Food Shopping, Business]

***Each marker colour represents different overall category type for each venue.**

New York:



Toronto:





TORONTO VS. NEW YORK CITY

DISCUSSION + CONCLUSION:

- In conclusion, we can observe that Toronto and New York City, though they are considered ‘sister cities’ due to their vast similarities – they each have their unique subsets of venues which are more popular (i.e have more neighborhoods which these venues are more frequently prominent). Of the 202 venue types that are shared between the two cities, New York City is more popular for 139 of them, and Toronto for 69 of them. The specific venues that each city is more ‘popular’ for can be seen in the results section. The differences between the cities have many possible contributing factors, such as the different demographics of people living there and the overall differences in city populations. NYC itself houses over 8.4 million people, while Toronto’s population is just shy of 3 million.

Areas for Further Improvement/Limitations:

- 1 – Not all neighborhoods are considered ‘equal’ and should have equal weighting. Different neighborhoods within the cities have different populations and sizing, but in this project’s calculations they are all considered equal and have an equal affect on the popularity score of the venue types, when in reality the average calculation should be a weighted average based on metrics like area size and population. For further improvement, data on area size and population should be included and factored in.
- 2 – our basis of popularity is simply on the frequency, however, in reality if someone wants to know which city is more well known or popular for a certain venue category – they might want to not only know frequency, but also how good those venues are. This can be done utilizing the Fousquare API and making premium calls to not only capture the venues, but see additional features like ratings and reviews to see the quality of the venues themselves. This additional data analysis was not performed in this project due to limitations on the number of premium calls for free users of the API, but should heavily be considered for anyone who wants to create a more accurate finding on popularity between the two cities.