Capstone Project

Analysing and Clustering Neighbourhoods and Venues For New York City & Toronto

Ebrahiem Kippie

April 2021

Project Overview:

In this project we will analyse and compare the cities of New York and Toronto located in the USA and Canada respectively.

The aim of the project is to establish what the major similarities and differences are between the two cities, whether there are businesses that are less or more common in both cities and kind of business is likely to strive in either city.

This would be important information which businesses can leverage to identify new prospective locations for opportunity.

Both New York City and Toronto are considered economic hubs of their respective countries with a rich diverse culture.

New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centre.

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's northwestern shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen's Park to 400-acre High Park and its trails, sports facilities and zoo. The economy of Toronto is the largest contributor to the Canadian economy, at 20% of national GDP, and an important economic hub of the world.

Data Preparation:

The following data sets will be used in this project:

Neighbourhood Data Sets:

New York City = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json"

Toronto = *"https://en.wikipedia.org/w/index.php?*

title=List_of_postal_codes_of_Canada:_M&oldid=1011037969"

Toronto Coordinate Data Set = "*https://cocl.us/Geospatial_data"*

All data sets were supplied as part of *"Coursera's IBM Data Science Certificate"* and contains the listings of the various neighbourhoods for both cities along with its respective latitude and longitude coordinates.

The New York City data set is essentially a JSON file transcribed into Panda's Data Frame. Once the Data Frame has been created, we are able to call the folium package to plot the locations of the various neighbourhood onto a map.

	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

Figure1: Pandas Dataframe New York City

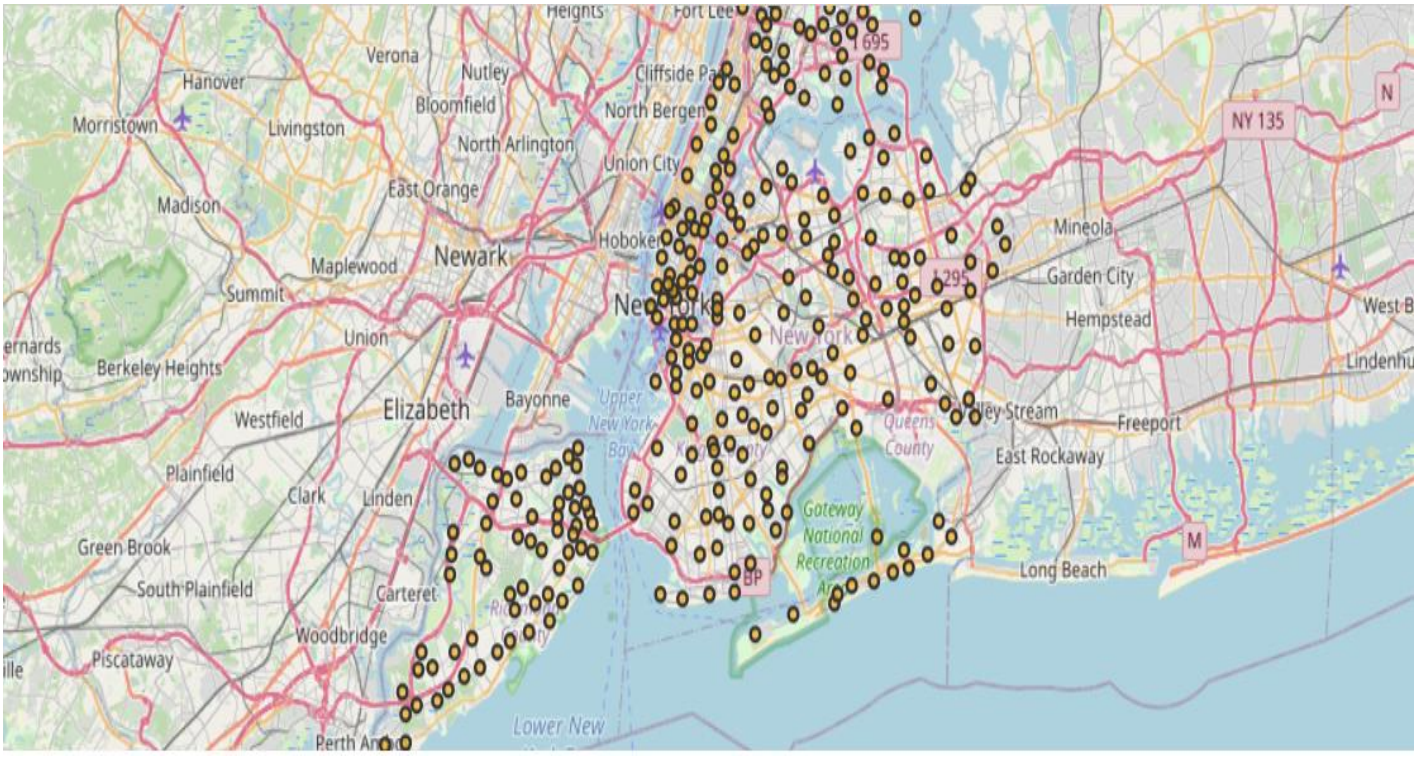


Figure: Map of New York City with Neighbourhoods using the Folium package.

Unlike the New York data set, the source data for the Toronto data set has not been extracted from a JSON file. The Toronto data set is essentially a Wikipedia page with a list of postal codes for Toronto with its associate borough and neighbourhood names. The Pandas Dataframe for Toronto was created by extracting HTML table data from a web page using the *Pandas read_html () function.*

All records with the "Not Assigned" values for Borough and Neighbourhood has been subsequently deleted from the data frame.

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

Figure: Pandas Data frame Toronto

Figure: Map of Toronto with Neighbourhoods using the Folium package.

Venue Data Set:

The data sets contain the listings of the top venues (e.g. the names of restaurants) for both cities' categories into specific neighbourhoods.

Both the neighbourhood and venue data sets will be analysed and retrieved using the Foursquare API.

New York City venue dataframe:

The venue dataframe for New York City was created using a function that takes as input the names, latitude and longitude coordinates for each neighbourhood and returns a dataframe.

By running a venues.shape on the dataframe, we noticed a total of 23483 venues extracted from the dataframe with over 500 venue categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Shell	40.891771	-73.853213	Gas Station
1	Wakefield	40.894705	-73.847201	Pitman Deli	40.896744	-73.844398	Food
2	Wakefield	40.894705	-73.847201	Julio C Barber Shop 2	40.892648	-73.855725	Salon / Barbershop
3	Wakefield	40.894705	-73.847201	Citibank	40.894161	-73.845825	Bank
4	Wakefield	40.894705	-73.847201	Edenwald Liquors	40.890942	-73.850455	Liquor Store

Figure: Venue dataframe for New York City

Toronto venue dataframe:

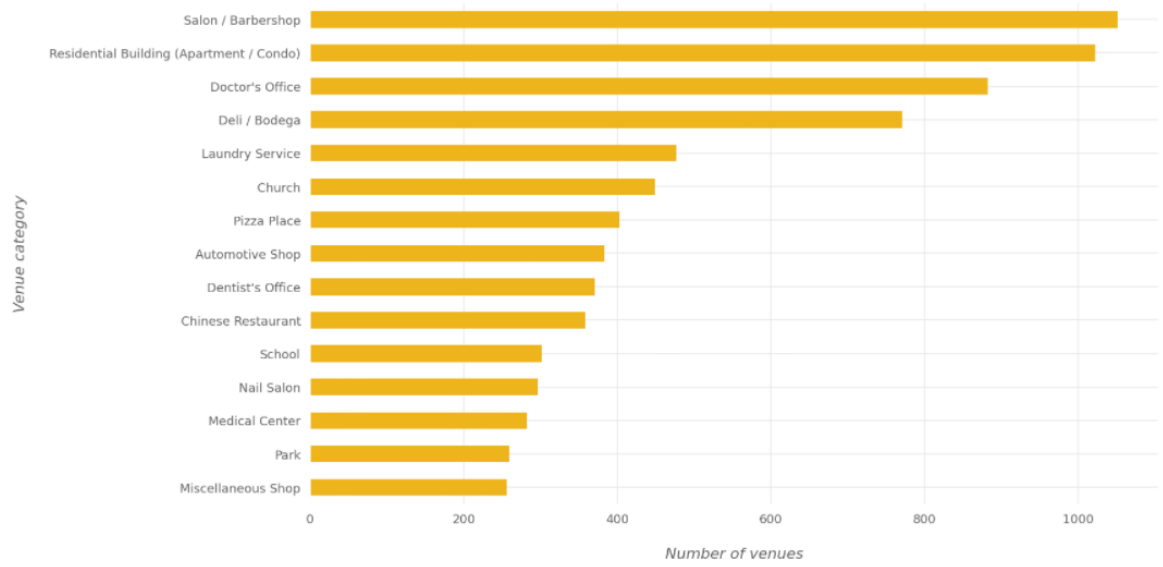
A similar process was followed to create the Toronto venues data frame. The dataframe below retrieved venue data for more than 7000 venues and over 500 unique categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Malvern, Rouge	43.806686	-79.194353	Rouge Park - Woodland Trail	43.801782	-79.200427	Trail
3	Malvern, Rouge	43.806686	-79.194353	Shell	43.803227	-79.192414	Gas Station
4	Malvern, Rouge	43.806686	-79.194353	Tim Hortons / Esso	43.801863	-79.199296	Coffee Shop
5	Malvern, Rouge	43.806686	-79.194353	Pleasant Corner	43.801164	-79.200254	Shopping Mall
6	Malvern, Rouge	43.806686	-79.194353	The Home Depot	43.801594	-79.196069	Hardware Store

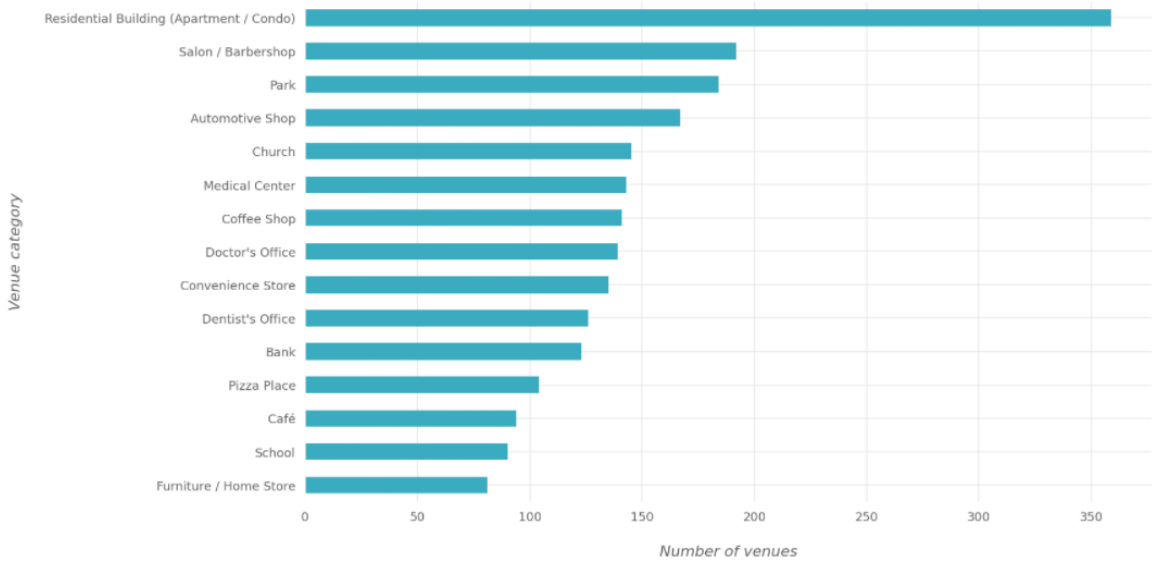
Figure: Venue dataframe for Toronto

Analysing the data:

What are the most common venue categories for New York City?

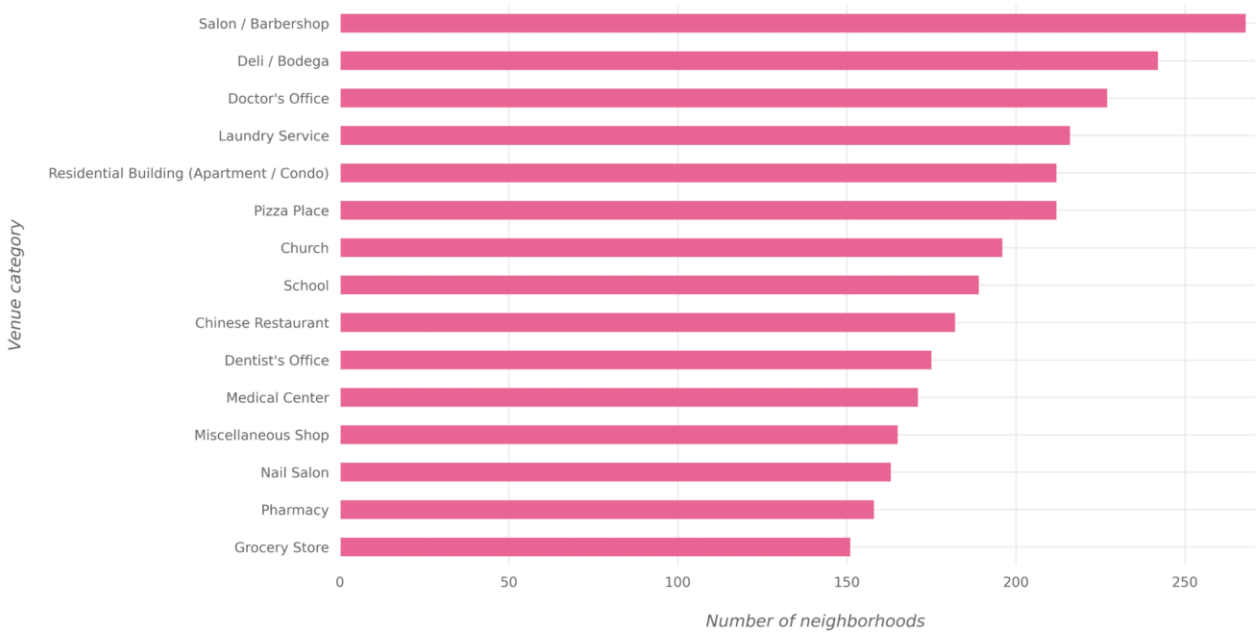


What are the most common venue categories for Toronto?

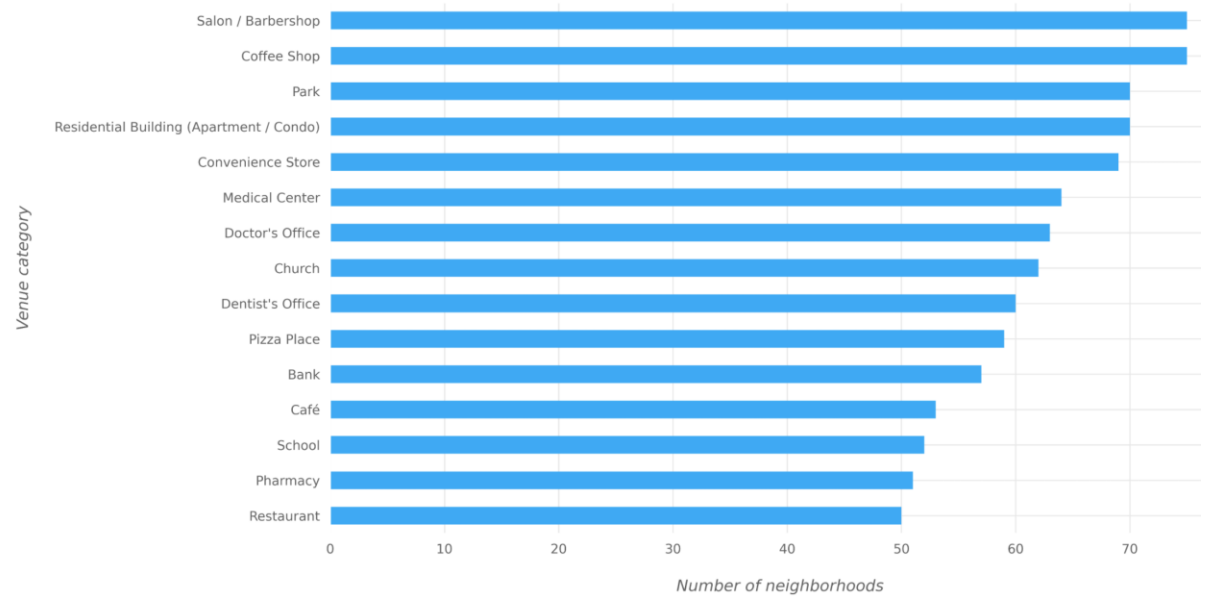


There are many similarities between the most common categories in New York and Toronto such as the frequent occurrence of Salons/Barbershops, Residential Buildings, Pizza places appearing in both cities, although not on the same scale as the numbers in New York City is always going to be much higher compared to Toronto.

What are the most common venues in New York City?



What are the most common venues in Toronto?



Rare Categories in New York City:

Venue Category	Count
EV Charging Station	1
Australian Restaurant	1
Israeli Restaurant	1
Tex-Mex Restaurant	1
Well	1
Tree	1
Music Festival	1
Circus	1
Tibetan Restaurant	1
Costume Shop	1
Buddhist Temple	1
Night Market	1
Stadium	1
Buffet	1
Airport	1

Rare Categories in Toronto:

Venue Category	Count
Perfume Shop	1
Luggage Store	1
Baggage Claim	1
Corporate Coffee Shop	1
Fish Market	1
Rehab Center	1
Mattress Store	1
Ski Chalet	1
Entertainment Service	1
College Hockey Rink	1
Sake Bar	1
College Basketball Court	1
Argentinian Restaurant	1
Ski Lodge	1
Museum	1

Results of Clustering:

Below are the results of the most common categories for the respective neighbourhoods using the K-Means clustering algorithm:

New York City:

	Neighborhood_	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
1	Co-op City	School	Residential Building (Apartment / Condo)	Fast Food Restaurant	Dentist's Office	Park	High School	Liquor Store
3	Fieldston	College Academic Building	College Administrative Building	College Residence Hall	Residential Building (Apartment / Condo)	Tech Startup	Historic Site	College Classroom
5	Kingsbridge	Laundry Service	Ice Cream Shop	Park	Bank	Salon / Barbershop	Discount Store	Bagel Shop
10	Baychester	Gas Station	Bank	Automotive Shop	Chinese Restaurant	Donut Shop	Burger Joint	Residential Building (Apartment / Condo)
12	City Island	Harbor / Marina	Italian Restaurant	Antique Shop	Pharmacy	Deli / Bodega	Park	Residential Building (Apartment / Condo)
22	Port Morris	Factory	Bar	Government Building	Financial or Legal Service	Storage Facility	Automotive Shop	Hardware Store
27	Clason Point	Park	Housing Development	Lounge	Salon / Barbershop	Automotive Shop	Intersection	Event Space
28	Throgs Neck	Other Great Outdoors	Deli / Bodega	Chinese Restaurant	Residential Building (Apartment / Condo)	Italian Restaurant	Tattoo Parlor	Japanese Restaurant

Toronto:

	Borough	Cluster_Labels2	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category
0	Scarborough	0	Automotive Shop	Factory	Coffee Shop	Dentist's Office	Medical Center	Sandwich Place	Convenience Store
1	Scarborough	0	Automotive Shop	Medical Center	Park	General Entertainment	Salon / Barbershop	Bar	Arts & Crafts Store
10	Scarborough	0	Automotive Shop	Furniture / Home Store	Auto Garage	Hardware Store	Storage Facility	Miscellaneous Shop	Factory
12	Scarborough	0	Automotive Shop	Auto Garage	Storage Facility	Doctor's Office	Church	Coffee Shop	Tech Startup
34	North York	0	Automotive Shop	Residential Building (Apartment / Condo)	Auto Dealership	Miscellaneous Shop	School	Bank	Transportation Service
53	Downtown Toronto	0	Automotive Shop	Furniture / Home Store	Italian Restaurant	Light Rail Station	Park	Design Studio	Hardware Store
81	Toronto/York	0	Automotive Shop	Furniture / Home Store	Doctor's Office	Miscellaneous Shop	Gas Station	Event Space	Park
96	North York	0	Automotive Shop	Gas Station	Bank	Furniture / Home Store	Italian Restaurant	Church	Gym
102	Etobicoke	0	Factory	Automotive Shop	Auto Dealership	Hardware Store	Medical Center	Transportation Service	Café

Please have a look at the notebook for a more complete analysis.

Conclusion:

The results of this reports can be used by aspiring businesses to gain an understanding what business are most popular in the respective neighbourhoods of both New York City and Toronto, where there are differences, but also where similarities exist.

The reports can be used to gain a general understanding of what business is likely to strive in a particular neighbourhood and where competition exists before investing in a particular business