



# **Applied Data Science Capstone Project : Final Assignment**

---

KIN CHUEN LEW

# Backgroud

---

New York City (NYC) and Toronto are located in North America and are major financial hubs in the world. They are made up of different skyscrapers and business centers. Both are very cosmopolitan and have dynamic life style. Apart from the commercial perspective, they also build with many high-rise residential building. Many Global organization around the world have office located in these 2 countries. Many people often relocate from other countries to these 2 cities and working in the central business district (CBD) areas. They may not be aware of the similarities or differences in these 2 cities. One of the examples is related to the ethnic makeups in NYC and Toronto. NYC has a much larger Black and Latino population, whereas Toronto has proportionally more Asians and Indians. Hence the likelihood of NYC having more America or south America Restaurant than Toronto is higher.

The target audience for this project is the expatriate who will move to either cities and will work on the CBD areas. Hence the scopes will focus on the Manhattan New York and East, downtown, central and West Toronto areas

# Problem and Interests

---

Given the diversity of the culture, this project will compare the following neighbourhoods of these two cities and determine how similar or dissimilar they are. In total,

- Manhattan consists of 40 neighbourhoods
- East, downtown, central and West Toronto consists of 39 neighbourhoods

It will focus on 3 topics

- Difference of the venue category between these 2 cities.
- Difference between the food culture based on the type of restaurant.
- Both cities will be independently split into clusters by neighbourhood. And then comparison between clusters will be done and identify similarity based on the venue category

It meant to provide the information for expatriates who plan to live in the neighbourhoods around the CBD areas so that they can choose the neighbourhoods best suit to their life style and needs.

# Data (a)

---

## Source of Data

Two data sets, one for Manhattan, one for Toronto, created from the previous labs or projects of the training course will be used as the source of data. These datasets have already populated with the information of the boroughs and neighbourhoods of NYC and Toronto as well as the respective latitudes and longitudes.

Before the data analysis, the neighbourhood candidates (NC) need to be filtered from the source of datasets. The outcome will have 2 datasets.

# Data (b)

---

- Neighbourhood Candidates  
Set A - represent the 40 neighbourhoods of Manhattan. The sample data is as follows

	number	Borough	Neighbourhood	Latitude	Longitude
	6	6	Manhattan	Marble Hill	40.876551 -73.910660
	100	100	Manhattan	Chinatown	40.715618 -73.994279
	101	101	Manhattan	Washington Heights	40.851903 -73.936900
	102	102	Manhattan	Inwood	40.867684 -73.921210
	103	103	Manhattan	Hamilton Heights	40.823604 -73.949688

- Neighbourhood Candidates  
Set B - represent the 39 neighbourhoods of East, downtown, central and West Toronto.

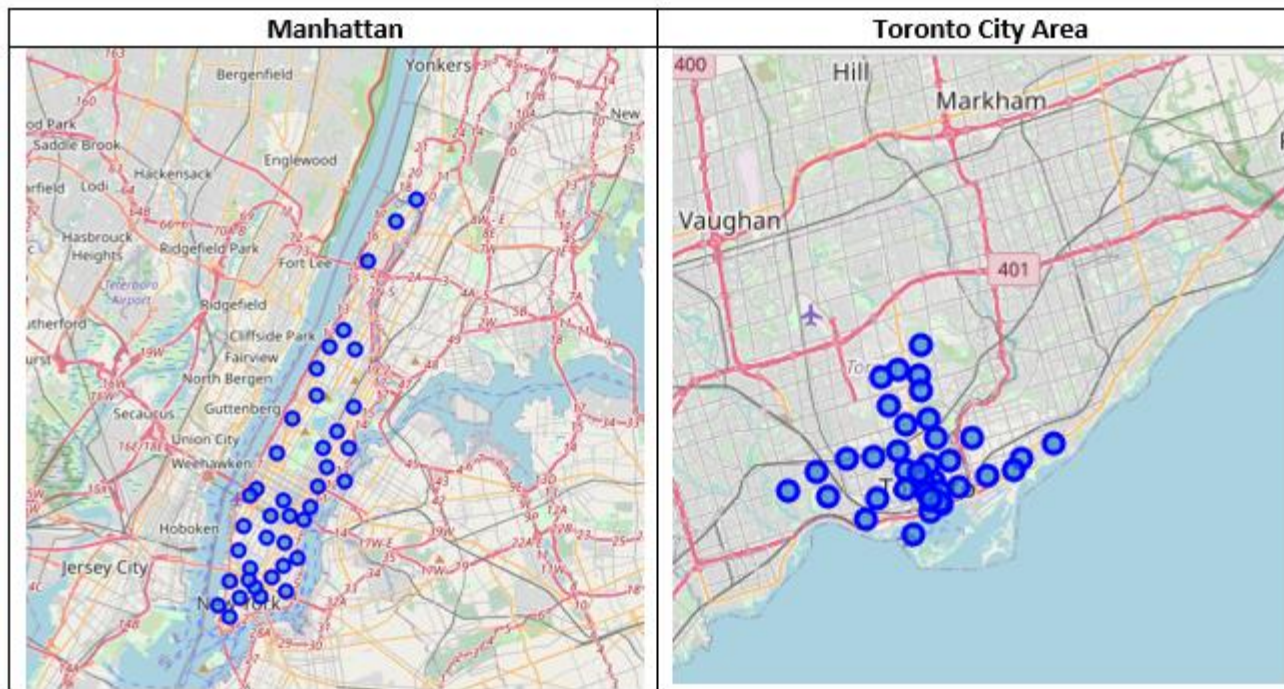
	number	Postal Code	Borough	Neighbourhood	Latitude	Longitude
	38	37.0	M4E	East Toronto	The Beaches	43.67635739999999 -79.2930312
	42	41.0	M4K	East Toronto	The Danforth West, Riverdale	43.6795571 -79.352188
	43	42.0	M4L	East Toronto	India Bazaar, The Beaches West	43.6689985 -79.31557159999998
	44	43.0	M4M	East Toronto	Studio District	43.6595255 -79.340923
	45	44.0	M4N	Central Toronto	Lawrence Park	43.7280205 -79.3887901



# Data (c)

---

Then, the geographical locations of the neighbourhoods will be reviewed to ensure the neighbourhoods are next to each other to ensure they are not scattered too far apart.



## Features selection

The venues and venue categories will be the key features for the analysis. Hence, Foursquare API will be used to extract the revenues and revenue categories of all the neighbourhoods for these 2 cities. These data will combine with the datasets Set A and Set B to create new datasets that have the neighbourhoods and the revenue categories.

# Methodology

---

After the data source have been loaded into the dataframe with data cleansing and filtering, Foursquare API will be used to collect the venues, latitudes, longitudes and venue categories for the neighbourhoods of Manhattan and Toronto City area.

To address the 1st audience interest, multiple datasets will be created to store venue categories followed by using "SET" operations to identify

- The common venue categories for both cities.
- The venue categories existed in Manhattan but not in Toronto City Area.
- The venue categories existed in Toronto City but not in Manhattan.

Difference between the food culture based on the type of restaurant will be the 2nd part of interest in this project. The "Restaurant" will be the key word to extract the records from the previous datasets and conduct an analysis or comparison.

Finally, the similarity of neighbourhood based on the venue category will be assessed. To do that, one hot encoding will be use to split the column which contains numerical categorical data to many columns depending on the number of categories present in that column. Both cities will be independently split into clusters by neighbourhood using cluster algorithm "kmeans"; and the comparing the clusters and surface out the similarity based on the venue category

# Analysis (a)

---

Explore the venues of Toronto City Center and Manhattan

Use Foursquare API to extract near by venues, latitudes, longitudes and the venue categories.

## Difference of the venue category between these 2 cities

- The number of same venues categories for both Manhattan and Toronto is 192.
- Manhattan has 141 venues categories different from Toronto City.
- Toronto City has 44 venues categories different from Manhattan.

Common Venue Categories	venue categories in Manhattan but not in Toronto	venue categories in Toronto City but not in Manhattan
'Adult Boutique', 'American Restaurant', 'Antique Shop', 'Art Gallery', 'Art Museum', 'Arts & Crafts Store', 'Asian Restaurant', 'Athletics & Sports', 'BBQ Joint', 'Baby Store', 'Bagel Shop', 'Bakery', 'Bank', 'Bar', 'Bed & Breakfast', 'Beer Bar', 'Beer Store', 'Bike Rental / Bike Share', 'Bistro', 'Boat or Ferry', 'Bookstore',	'Accessories Store', 'Afghan Restaurant', 'African Restaurant', 'Arepa Restaurant', 'Argentinian Restaurant', 'Auditorium', 'Australian Restaurant', 'Austrian Restaurant', 'Badminton Court', 'Baseball Field', 'Basketball Court', 'Beer Garden', 'Big Box Store', 'Bike Shop', 'Bike Trail', 'Board Shop', 'Boxing Gym', 'Bridal Shop', 'Bridge', 'Bus Station', 'Cafeteria',	'Airport', 'Airport Food Court', 'Airport Gate', 'Airport Lounge', 'Airport Service', 'Airport Terminal', 'Aquarium', 'Auto Workshop', 'Baseball Stadium', 'Basketball Stadium', 'Beach', 'Belgian Restaurant', 'Brewery', 'Church', 'College Auditorium', 'College Gym', 'College Rec Center', 'Colombian Restaurant', 'Comfort Food Restaurant', 'Comic Shop', 'Coworking Space',



# Analysis (b)

---

## Difference between the food culture based on the type of restaurant

In order to compare the food culture based on the type of restaurant, the “Sets” created in previous slide will be converted back to data frames and filter all rows with “Restaurant” in the venue categories. Please see the sample outcome in the table.

List of common Restaurants	List of Restaurants in Manhattan but not in Toronto City area	List of Restaurants in Toronto City area but not in Manhattan
Fast Food Restaurant Latin American Restaurant Molecular Gastronomy Restaurant Modern European Restaurant Korean Restaurant Ramen Restaurant Middle Eastern Restaurant Mexican Restaurant Vietnamese Restaurant Chinese Restaurant Mediterranean Restaurant Taiwanese Restaurant Falafel Restaurant Filipino Restaurant Portuguese Restaurant Greek Restaurant Vegetarian / Vegan Restaurant Indian Restaurant	Swiss Restaurant African Restaurant Malay Restaurant Scandinavian Restaurant Jewish Restaurant Cantonese Restaurant Spanish Restaurant Czech Restaurant Japanese Curry Restaurant Kebab Restaurant Dim Sum Restaurant Shanghai Restaurant Lebanese Restaurant Szechuan Restaurant Hotpot Restaurant Arepa Restaurant Afghan Restaurant Empanada Restaurant Paella Restaurant	Comfort Food Restaurant Theme Restaurant Gluten-free Restaurant Belgian Restaurant Doner Restaurant Tibetan Restaurant Colombian Restaurant

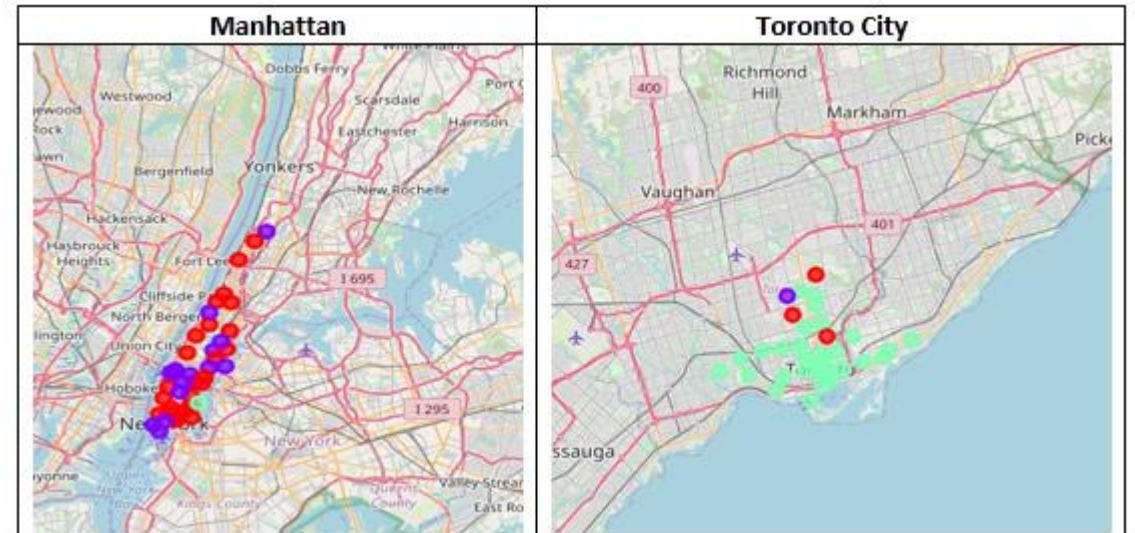
# Analysis (c)

---

Both cities will be independently split into clusters by neighbourhood. And then comparison between clusters will be done and identify similarity based on the venue category

Further analyse the neighbourhood of Manhattan & Toronto City is required. To recreate the clusters of neighbourhoods for both cities, **One hot encoding** will be used to split the column which contains numerical categorical data to many columns depending on the number of categories present in that column. In this case, the categorical data of venue categories will split to multiple columns. The clustering labels will be added.

Once the data frames are ready, K-means clustering machine learning algorithm will be applied to cluster the neighbourhoods for both cities. Folium maps will be prepared to show the clusters for both cities.



Finally, new data frames will be built with the top 10 most common venue categories, which will provide the better understanding about the facilities available in each neighbourhood. This can provide better ways to do the comparison. The clusters are as follows

## Manhattan Clusters

Manhattan Cluster 0									
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
100	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Bubble Tea Shop	Ice Cream Shop	Hotpot Restaurant	Salon / Barbershop	Optical Shop
101	Washington Heights	Cafe	Bakery	Grocery Store	Del / Bodega	Chinese Restaurant	Mobile Phone Shop	New American Restaurant	Latin American Restaurant
102	Inwood	Mexican Restaurant	Lounge	Restaurant	Cafe	Bakery	Spanish Restaurant	Frozen Yogurt Shop	Caribbean Restaurant
103	Hamilton Heights	Pizza Place	Coffee Shop	Cafe	Mexican Restaurant	Del / Bodega	Cocktail Bar	Latin American Restaurant	Sushi Restaurant
104	Manhattanville	Seafood Restaurant	Coffee Shop	Italian Restaurant	Mexican Restaurant	Chinese Restaurant	Del / Bodega	Sushi Restaurant	Climbing Gym
105	Central Harlem	African Restaurant	Cosmetics Shop	French Restaurant	American Restaurant	Bar	Chinese Restaurant	Art Gallery	Seafood Restaurant
106	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Del / Bodega	Spa	Latin American Restaurant	Sandwich Place	Taco Place
108	Yonkers	Italian Restaurant	Gym	Coffee Shop	Del / Bodega	Sushi Restaurant	Bar	Wine Shop	Diner
109	Lenox Hill	Italian Restaurant	Sushi Restaurant	Coffee Shop	Cocktail Bar	Pizza Place	Cafe	Gym / Fitness Center	Gym
111	Upper West Side	Wine Bar	Bakery	Bar	Italian Restaurant	Indian Restaurant	Cafe	Coffee Shop	Pizza Place
112	Lincoln Square	Cafe	Pizza	Italian Restaurant	Gym / Fitness Center	Theater	Concert Hall	Performing Arts Venue	Indie Movie Theater
Manhattan Cluster 1									
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
8	Marble Hill	Gym	Discount Store	Sandwich Place	Coffee Shop	Yoga Studio	Pizza Place	Steakhouse	Shopping Mall
107	Upper East Side	Italian Restaurant	Coffee Shop	Exhibit	Bakery	Gym / Fitness Center	American Restaurant	Spa	French Restaurant
110	Roosevelt Island	Park	Gym	Dry Cleaner	Bubble Tea Shop	Soccer Field	Farmers Market	Supermarket	Metro Station
113	Clinch	Theater	Italian Restaurant	Gym / Fitness Center	Coffee Shop	American Restaurant	Gym	Spa	Wine Shop
114	Midtown	Hotel	Clothing Store	Coffee Shop	Sporting Goods Shop	Theater	Bookstore	Cafe	Steakhouse
115	Murray Hill	Coffee Shop	Sandwich Place	Bar	Japanese Restaurant	American Restaurant	Gym / Fitness Center	Burger Joint	Hotel
125	Morningside Heights	Coffee Shop	Park	American Restaurant	Bookstore	Burger Joint	Cafe	Ice Cream Shop	New American Restaurant
127	Battery Park City	Coffee Shop	Park	Hotel	Clothing Store	Gym	Memorial Site	Shopping Mall	Wine Shop
128	Financial District	Coffee Shop	Pizza Place	Bar	Hotel	Gym	Cocktail Bar	Park	Mexican Restaurant
Manhattan Cluster 2									
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
275	Stuyvesant Town	Park	Bar	Boat or Ferry	Coffee Shop	Helipad	Food Truck	Gas Station	Bistro

## Toronto Clusters

Toronto Cluster 0									
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
45	Lawrence Park	Park	Swim School	Bus Line	Yoga Studio	Diner	Event Space	Ethiopian Restaurant	Electronics Store
51	Rosedale	Park	Playground	Trail	Yoga Studio	Department Store	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
65	Forest Hill North & West, Forest Hill, Road Park	Park	Sushi Restaurant	Trail	Jewelry Store	Yoga Studio	Dona Shop	Discount Store	Distribution Center
Toronto Cluster 1									
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
64	Roselawn	Home Service	Garden	Yoga Studio	Department Store	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
Toronto Cluster 2									
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
38	The Beaches	Coffee Shop	Health Food Store	Neighborhood	Trail	Pub	Yoga Studio	Dog Run	Diner
42	The Danforth West, Riverdale	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Furniture / Home Store	Liquor Store	Indian Restaurant	Spa
43	Indie Bazaar, The Beaches West	Sandwich Place	Park	Fast Food Restaurant	Coffee Shop	Food & Drink Shop	Light Rail Station	Restaurant	Italian Restaurant
44	Studio District	Coffee Shop	American Restaurant	Bakery	Brasserie	Cafe	Gastropub	Yoga Studio	Fish Market
45	Danforth North	Gym / Fitness Center	Hotel	Breakfast Spot	Food & Drink Shop	Sandwich Place	Department Store	Park	Convenience Store
47	North Toronto West, Lawrence Park	Clothing Store	Coffee Shop	Yoga Studio	Fast Food Restaurant	Italian Restaurant	Cafe	Mexican Restaurant	Salon / Barbershop
48	Danforth	Sandwich Place	Discount Shop	Pizza Place	Coffee Shop	Sushi Restaurant	Cafe	Italian Restaurant	Gym
49	Noon Park, Summerhill East	Louyer	Restaurant	Yoga Studio	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
50	Summerhill West, Ratheny, South Hill, Forest Hill	Coffee Shop	American Restaurant	Liquor Store	Restaurant	Bank	Bagel Shop	Supermarket	Sushi Restaurant
52	St. James Town, Cobbleton	Coffee Shop	Cafe	Pizza Place	Restaurant	Italian Restaurant	Bakery	Pub	Beer Store

# Results & Discussion

---

It is interesting to see the different shapes of the 2 cities; Manhattan is rectangular shaped while Toronto City Area is more squarish shaped.

The number of same venues categories for both Manhattan and Toronto is 192. Manhattan has 141 venues categories different from Toronto City. Toronto City has 44 venues categories different from Manhattan.

The food culture for both cities covered almost all regions: Asia, LATAM, European, Australia, Africa, North and South America. Manhattan has many restaurants, which are related to different provinces of different countries. Toronto City area has some interesting theme restaurants, which not exist in Manhattan.

The neighbourhood of cluster 0 and 1 of Manhattan are similar to cluster 2 of Toronto City, given that these clusters have good mix of venues and facilities such as cafe / coffeeshop, Gym, Spa, restaurants, Part and Hotel.

Cluster 0 of Toronto City has more parks, trails, playgrounds, restaurants, shops that seem to be more suitable for family living.

Both cluster 2 of Manhattan and cluster 0 of Toronto City have only 1 neighbourhood, the comparison is inconclusive, given that the venues are quite different.

# Conclusion

---

Although some differences have been surfaced out from this study, the majority of venue categories are similar. Expatriates stayed in either one of the cities should have no problem to adapt on the other.