

# Comparing two cities New York and Toronto:

## 1. Introduction to business problem:

Nowadays Travel & Tourism growing as a major industry. People around the world are travelling to different places for various reasons. Like any other industry, travel industry also using artificial intelligence and machine learning techniques to help their clients.

New York and Toronto are two important cities because of various reasons. These are cities which are the financial capitals of the respective countries. Those cities are famous for their world class infrastructure, entertainment and leisure, and home to people from various regions across the world. Thousands of people are travelling to these countries every year.

My effort in this project is to help them by finding the most common venues in each neighbourhood of these cities. By doing this, they can find easily which location is better suited for them to visit, stay and can explore some great business ideas too by knowing each neighbourhood better.

## 2. Data Acquisition and Cleaning:

### Data Sources:

Every data science problem is data driven. Here my data sources are :  
[https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset), [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M), [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)  
from where I fetched the data required to do the analysis.

Data Cleaning and Feature selection: We are fetching the necessary features required for our analysis which are Neighbourhood name, Borough, Latitude and Longitude. We create python dataframe from the raw data and filter the unnecessary data and fetch the features required for our analysis.

Dataframe of New York City:

We cleaned the data to get the required data frame which consists the features [**Borough, Neighborhood, Latitude, Longitude**]

```
[12]:
```

	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

Dataframe of Toronto:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.763573	-79.188711
1	M4A	North York	Victoria Village	43.770992	-79.216917
2	M5A	Downtown Toronto	Harbourfront	43.773136	-79.239476
3	M6A	North York	Lawrence Heights	43.744734	-79.239476
4	M6A	North York	Lawrence Manor	43.727929	-79.262029

### 3. Methodologies:

In this project we will direct our efforts on detecting and comparing the neighbourhood of New York and Toronto. We will limit our analysis to top 10 venues in each neighbourhood within an area ~1km around it. In first step we have collected the required data. Second step in our analysis will be finding venues and categorize them. In third and final step we will focus on create clusters of Neighbourhood using kmeans. Create a new dataframe that includes the cluster as well as the top 10 venues for each neighbourhood of Toronto and New York.

Get the top 100 venues that are in New York within a radius of 500 meters. We are using Foursquare API credentials to get latitude and longitude of cities.

4. Exploratory Data Analysis: We are analysing the data to get the unique venues. We are grouping the venues of each neighbourhood and using one hot method to find the average occurrences of each venue in each neighbourhood for both the cities.

For New York,

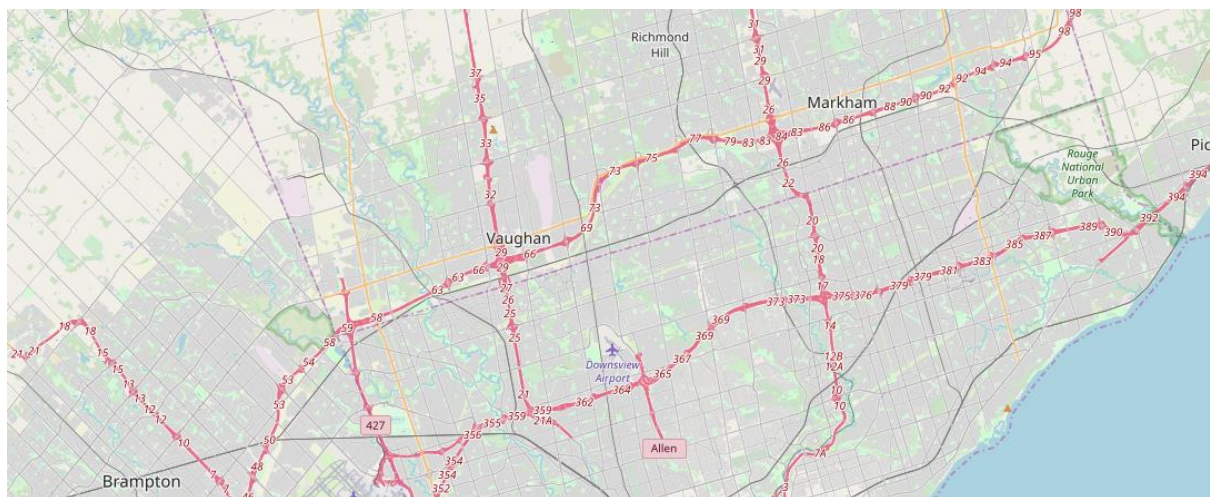
	Neighborhood	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	Allerton	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.016129
1	Annadale	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.111111
2	Arden Heights	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000
3	Arlington	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000
4	Arrochar	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000
5	Arverne	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000
6	Astoria	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.010000
7	Astoria Heights	0.000000	0.000000	0.00	0.000000	0.000000	0.012195	0.02439	0.012195	0.000000
8	Auburndale	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.010000
9	Bath Beach	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.00000	0.000000	0.010000

For Toronto,

	Neighborhood	Zoo	Accessories Store	Afghan Restaurant	Airport	Airport Lounge	American Restaurant	Amphitheater	Antique Shop	Argentinian Restaurant
0	Adelaide	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00
1	Bathurst Manor	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00
2	Bayview Village	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00
3	Berczy Park	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00
4	Bloordale Gardens	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.00
5	Brockton	0.00	0.000000	0.000000	0.000000	0.000000	0.062500	0.00	0.00	0.00
6	CFB Toronto	0.00	0.000000	0.000000	0.000000	0.000000	0.058824	0.00	0.00	0.00

ML algorithm used:

We are using K means clustering algorithm to cluster our neighbourhood based on top venues, where we set our cluster size as 5. We visualize our cluster with the help of folium.



## 4. Results & Discussion:

We analysed the neighbourhoods of Toronto and New York and found how similar these cities are. The top venues of cluster1 of Toronto is:

	Borough	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	North York	3.0	Pizza Place	Coffee Shop	Fast Food Restaurant	Sports Bar	Filipino Restaurant	Smoothie Shop	Food & Drink Shop	Supermarket	Electronics Store	Fried Chicken Joint
1	North York	3.0	Park	Coffee Shop	Fast Food Restaurant	Pharmacy	Indian Restaurant	Business Service	Chinese Restaurant	Mobile Phone Shop	Filipino Restaurant	Fireworks Store
2	Downtown Toronto	3.0	Bakery	Coffee Shop	Indian Restaurant	Pharmacy	Yoga Studio	Gas Station	Caribbean Restaurant	Thai Restaurant	Bank	Fried Chicken Joint
3	North York	3.0	Fast Food Restaurant	Restaurant	Train Station	Auto Garage	Grocery Store	Coffee Shop	Sandwich Place	Bowling Alley	Japanese Restaurant	Pizza Place
4	North York	3.0	Coffee Shop	Chinese Restaurant	Discount Store	Fast Food Restaurant	Convenience Store	Pizza Place	Bank	Pharmacy	Light Rail Station	Sandwich Place
5	Queen's Park	3.0	Intersection	Bakery	Bus Line	Coffee Shop	Diner	Fast Food Restaurant	Soccer Field	Bank	Beer Store	Mexican Restaurant
7	Scarborough	3.0	Fast Food Restaurant	Coffee Shop	Electronics Store	Asian Restaurant	Burger Joint	Pharmacy	Bakery	Indian Restaurant	Chinese Restaurant	Furniture / Home Store
9	North York	3.0	Coffee Shop	Convenience Store	Pizza Place	Fast Food Restaurant	Sandwich Place	Park	Intersection	Shopping Mall	Bakery	Bank

The top venues of New York is:

	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
11	Pelham Parkway	Pizza Place	Deli / Bodega	Donut Shop	Bakery	Sandwich Place	Bank	Italian Restaurant	Burger Joint	Ice Cream Shop	Gas Station
16	Fordham	Italian Restaurant	Pizza Place	Mobile Phone Shop	Spanish Restaurant	Deli / Bodega	Coffee Shop	Shoe Store	Donut Shop	Supplement Shop	Liquor Store
28	Throgs Neck	Italian Restaurant	Deli / Bodega	Pizza Place	Donut Shop	Restaurant	Ice Cream Shop	Sports Bar	Tennis Court	Park	Beach
29	Country Club	Italian Restaurant	Sandwich Place	Deli / Bodega	Bank	Pizza Place	Bakery	Donut Shop	Fast Food Restaurant	Playground	Dive Bar
34	Belmont	Italian Restaurant	Pizza Place	Bakery	Deli / Bodega	Zoo	Dessert Shop	Café	Shoe Store	Latin American Restaurant	Plaza
37	Pelham Bay	Italian Restaurant	Fast Food Restaurant	Donut Shop	Sandwich Place	Deli / Bodega	Pizza Place	Convenience Store	Bank	Bakery	Gym / Fitness Center

## Conclusion:

In this project we have clustered the neighbourhoods of two popular cities New York and Toronto and identified the most common venues in each area. Some area are popular for restaurants, others are popular for beaches, Bakery...etc.

In New York City cluster 1's most popular venues are beaches, where as cluster 2 is popular for Restaurants. In New York Italian pizza is a very common venue where as in Toronto there are different combination of restaurants like Chinese, Vietnamese, Malay can be seen.

Toronto's cluster 1 is popular for restaurants where as cluster 2 is popular for dog run, coffeshops.