Coursera IBM Capstone Project Report

Introduction

Background

Toronto is the capital city of the Canadian province of Ontario with population 2.7 Million. It is the most populous city in Canada and the fourth most populous city in North America. People have travelled through and inhabited the Toronto area, located on a broad sloping plateau interspersed with rivers, deep ravines, and urban forest, for more than 10,000 years. Now, Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. Some neighbourhoods are Old Toronto, Downtown Core (Central), East End, North End, West End, East York, Etobicoke, North York, Scarborough, York.

NYC is New York City is often called just New York is the most populous city in the United States. Its population is 8.3 Million. Just like Toronto it is filled with people of different nationalities and it is a multicultural city. The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island—were consolidated into a single city in 1898. Some neighbourhoods are Central Bronx, Central Brooklyn, Borough Park, Flatbush, Greenpoint, East Harlem, Upper West Side, Jamaica, Rockaways, West Queens.

Problem Statement

Toronto and NYC are top cities for tourism in the world. The tourism characteristics should be compared in this project. This study or project should be able to use restaurant, museums, galleries for comparison.

Thus, this project should be able to help **business people** who are trying to select a good neighbourhood for open restaurants.

Scope of Study

In case of Toronto, the entire city area is taken for study. For NYC, some neighbourhoods in the study are Central Bronx, Central Brooklyn, Borough Park, Flatbush, Greenpoint, East Harlem, Upper West Side, Jamaica, Rockaways, West Queens.

Data

The number of restaurants, museums, galleries in Toronto and NYC are compared. The data sources that explain the postal codes and corresponding neighbourhoods/ boroughs/ districts of the two locations of interest are extracted from

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for Toronto.

The postal codes for neighbourhoods in NYC are obtained via https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm

The coordinates of the postal codes based on the district/ borough/ neighbourhood of both locations are extracted via 'http://cocl.us/Geospatial_data' (Toronto, Ontario).

The geospatial data, numbers and other details of the restaurants, museums and galleries of both location of interest are extracted via Foursquare API.

Example of data that we get from Foursquare API is the venues, venues latitude and venues longitude when the name of neighbourhood, its latitude and longitude is given

Example: (North York, 43.72588, -79.315572) will give result (The Curry & Roti Restaurant, 43.742554, -79.308792)

Example: (North York, 43.718518, -79.464763) will give the result (Hakka No.1 Restaurant, 43.756800, -79.312850)

Methodology

In this section the data comprises of the number of restaurants, galleries and museums in Toronto and NYC are collected using the search query function available in Foursquare API, with the respective baseline input search queries of 'Restaurant', 'Gallery' and 'Museum' and radius settings of 2000m for Toronto and 2500m for NYC. The radius settings are selected based on the distribution of districts or borough in both location of interests.

In this following section, the data gathered from the previous section will be used to do exploratory data analysis, we will be looking into the data comparisons of both locations via data visualisation as explained below:

Total Number of Restaurants, Museums and Galleries in both locations. Distribution of Restaurants, Museums and Galleries in both locations according to districts / boroughs.

Total Number of Restaurants, Museums and Galleries per district / borough in both locations. Then, the potentiality study of both locations on tourism will be accessed based on the analysed data.

Data wrangling is to identify and handle missing value, to standardize and normalize the data. Data exploratory by analysing neighbourhoods using visualization, descriptive statistical analysis. Model development: k-means will be developed to predict the clusters. A Model will help to understand the exact relationship both cities. The machine learning model should be able to predict the clusters similarity. Here we found how many unique venue categories are there in both locations.

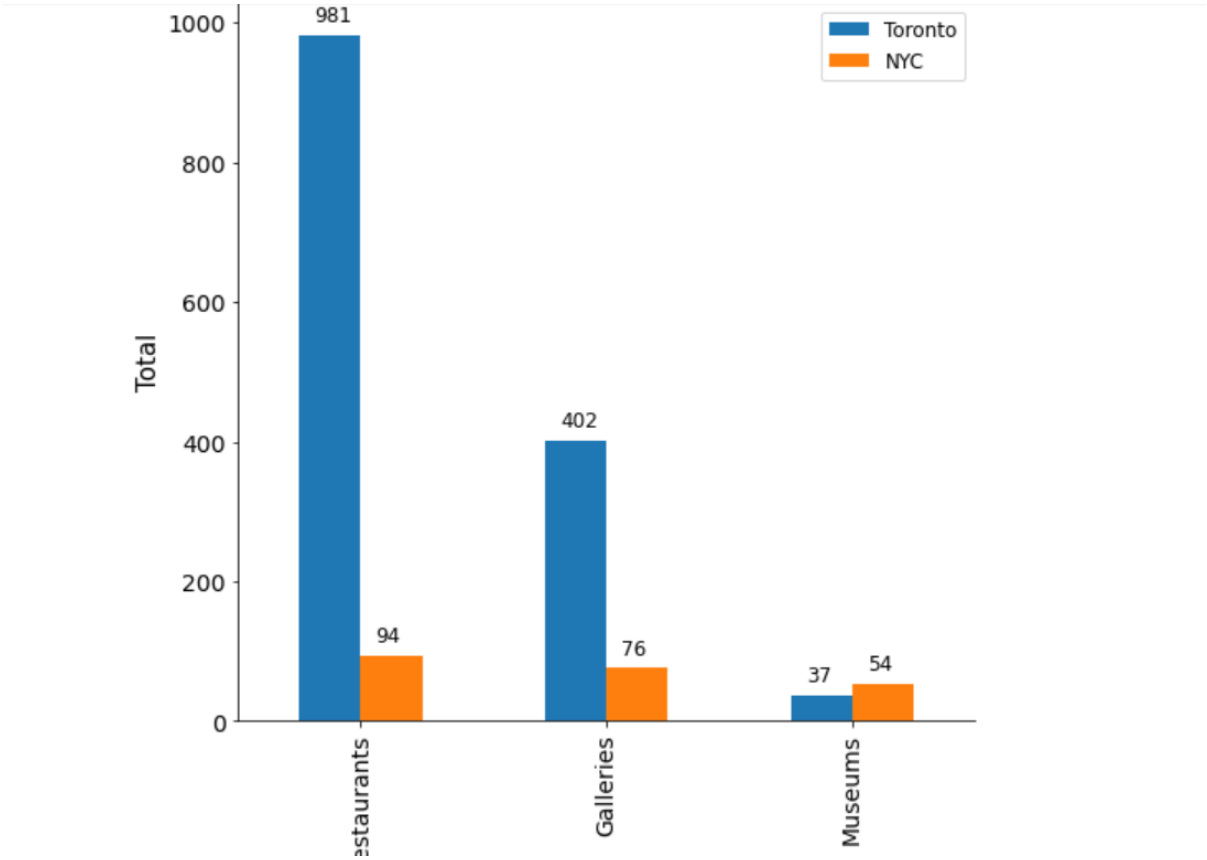
Analysis

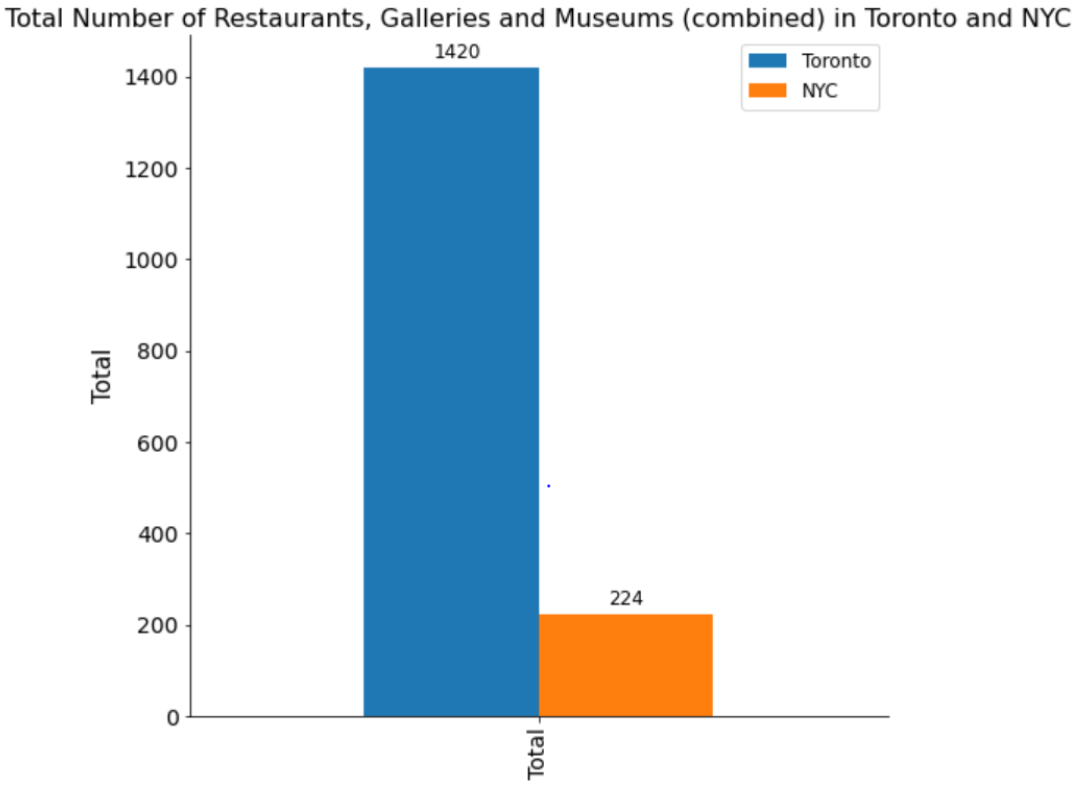
Study on the total number of Restaurants, Galleries and Museums in both Location. Here, the study on the total number of restaurants, galleries and museums are aggregated in the form of dataframes for comprehension.

Study on the Distributions of Restaurants, Galleries and Museums in both Location (District/ Borough) are aggregated in the form of dataframes for comprehension.

Results

As seen from the table below, while comparing between Toronto and 10 neighbourhoods in NYC there are many more restaurants and galleries in Toronto but there are fewer museums in Toronto. We know that NYC is very famous for its museums.

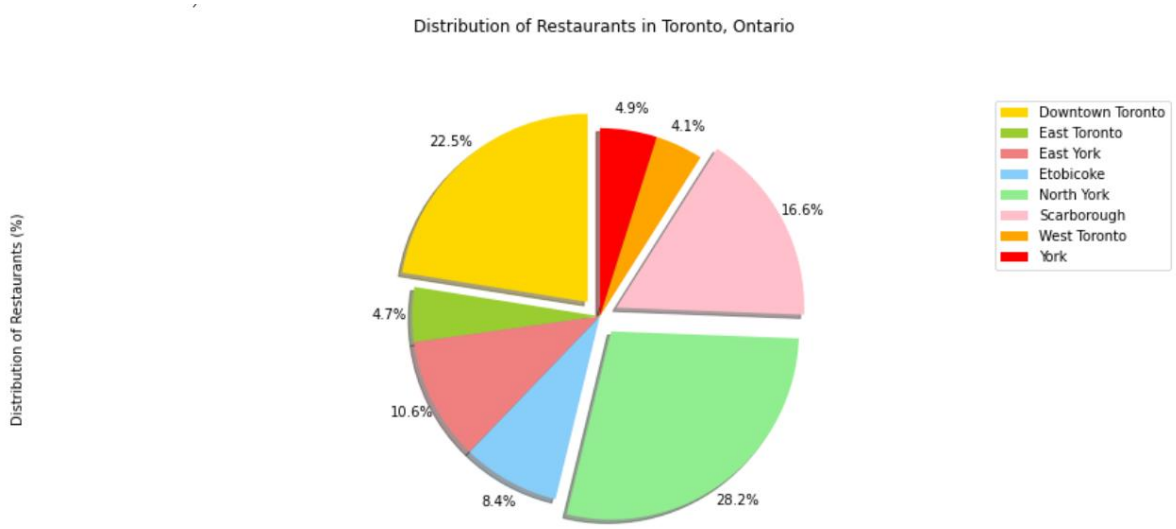


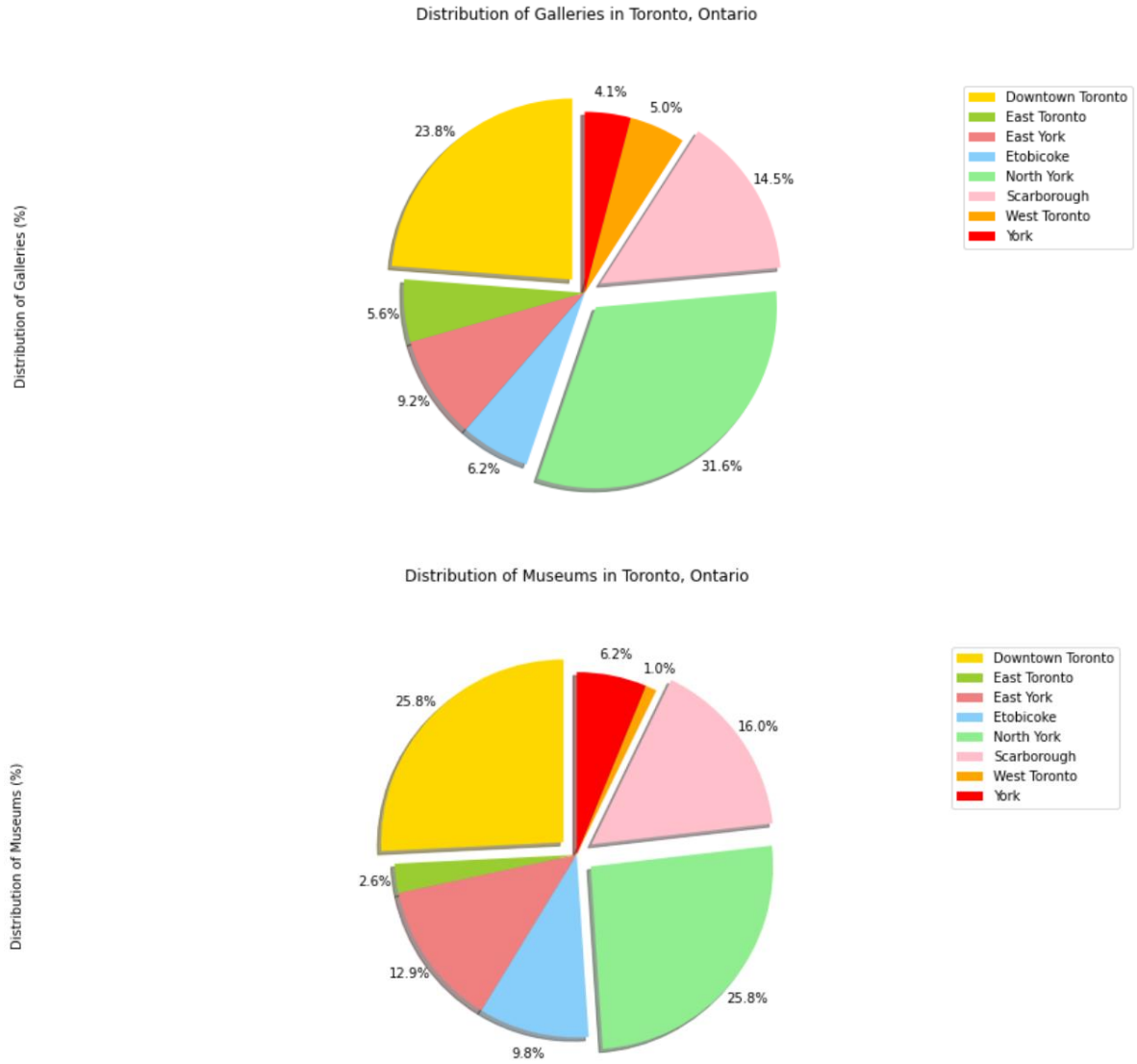


The table below shows the distribution of restaurants, galleries and museums in percent for Toronto.

Neighborhood	Distribution of Restaurants (%)	Distribution of Galleries (%)	Distribution of Museums (%)
Downtown Toronto	22.5	23.8	25.8
East Toronto	4.7	5.6	2.6
East York	10.6	9.2	12.9
Etobicoke	8.4	6.2	9.8
North York	28.2	31.6	25.8
Scarborough	16.6	14.5	16.0
West Toronto	4.1	5.0	1.0
York	4.9	4.1	6.2

Following are the Pie charts for the distribution of restaurants, galleries and museums in percent for Toronto.

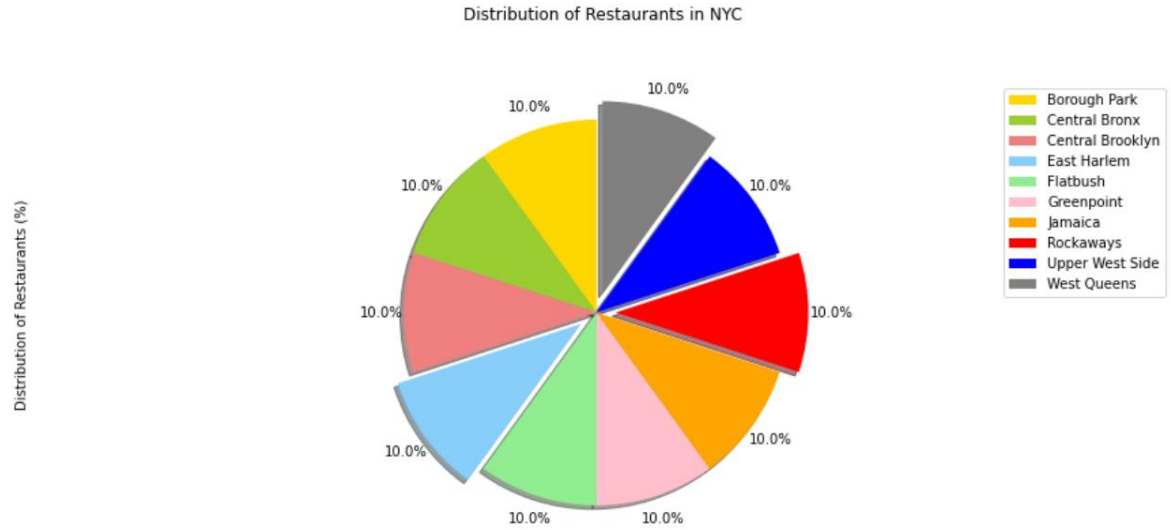


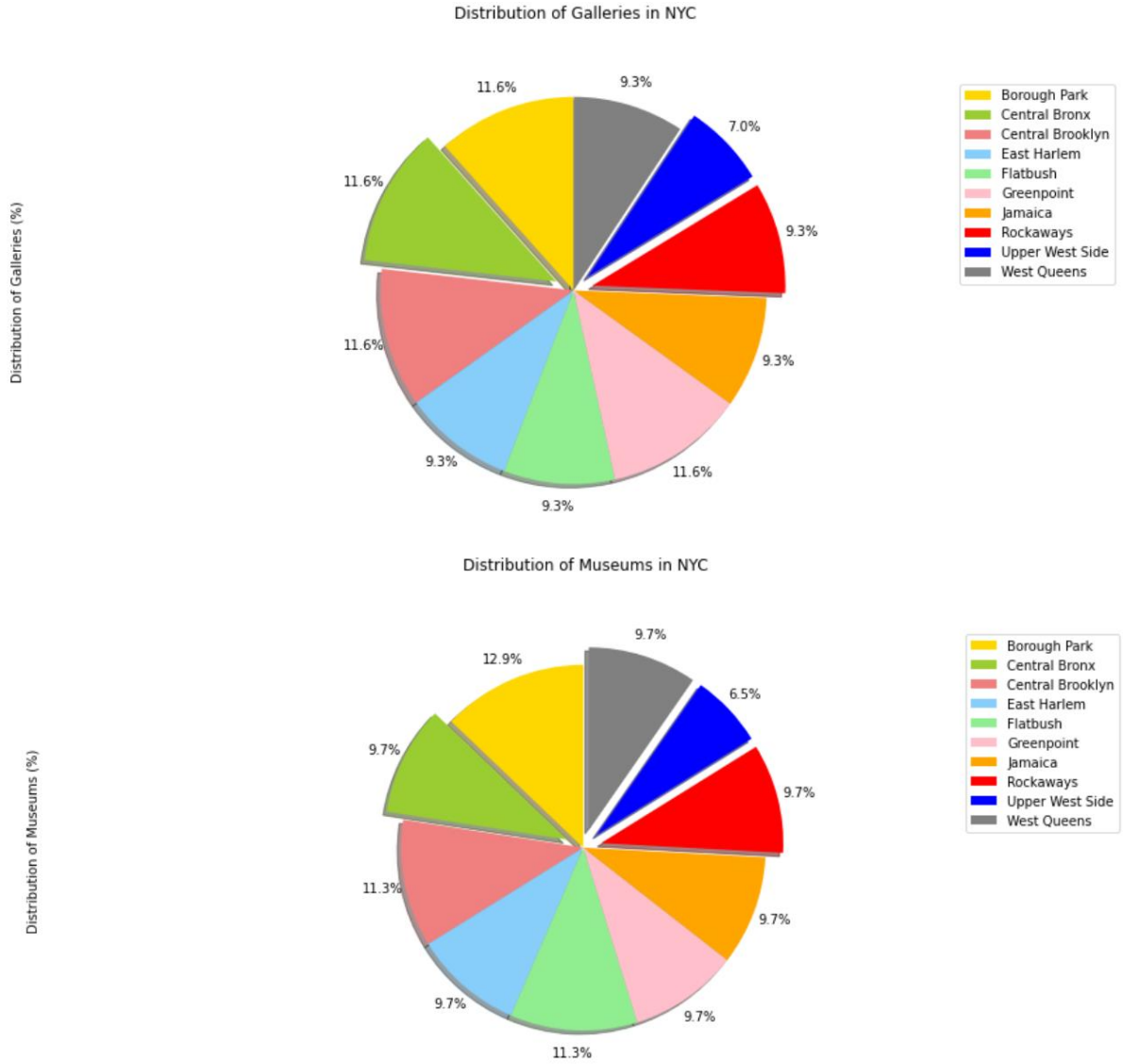


The table below shows the distribution of restaurants, galleries and museums in percent for NYC.

Neighborhood	Distribution of Restaurants (%)	Distribution of Galleries (%)	Distribution of Museums (%)
Borough Park	10.0	11.6	12.9
Central Bronx	10.0	11.6	9.7
Central Brooklyn	10.0	11.6	11.3
East Harlem	10.0	9.3	9.7
Flatbush	10.0	9.3	11.3
Greenpoint	10.0	11.6	9.7
Jamaica	10.0	9.3	9.7
Rockaways	10.0	9.3	9.7
Upper West Side	10.0	7.0	6.5
West Queens	10.0	9.3	9.7

Following are the Pie charts for the distribution of restaurants, galleries and museums in percent for Toronto.





The following dataframes show all the venue categories in Toronto and NYC. We do this to find the unique number of categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North York	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	North York	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	North York	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	North York	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	North York	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central Bronx	40.850656	-73.866524	Bronx Park East	40.849164	-73.868453	Park
1	Central Bronx	40.850656	-73.866524	Park Billiards	40.850970	-73.867792	Pool Hall
2	Central Bronx	40.850656	-73.866524	F & J Pine Tavern	40.848766	-73.862242	Italian Restaurant
3	Central Bronx	40.850656	-73.866524	Cafe Colonial	40.852495	-73.867654	Spanish Restaurant
4	Central Bronx	40.850656	-73.866524	John & Joe's Pizzeria	40.854520	-73.865824	Pizza Place

There are 273 unique venue categories in Toronto and 148 unique venue categories in NYC.

Discussion

We used Folium to plot the venues like restaurants, galleries and museums in NYC and Toronto. In NYC we have randomly taken just 10 neighbourhoods since the data over entire NYC is very large. Even then we have a large number of venues obtained.

We make recommendation that users (tourists) use the maps to find their desired restaurants, galleries, museums.

Also, if a new business person wants to open a venue, we suggest that he can do it in York in Toronto or upper west side because there are fewer venues there.

Conclusion

We observe that since New York is a more populated city than Toronto even just 10 neighbourhoods in NYC have a comparable number of venues with Toronto. Thus, we can use just few numbers of neighbourhoods in NYC to compare with the venues in Toronto. There is a lot of scope for tourist to find a new place to go to in both cities for every day of their stay or even residents to find good venues in their neighbourhood using this project. This project can also help investors interested in opening new venues.