

Venue Comparison Analysis between Manhattan & Toronto

Yaohua Xu

May 25, 2019

1. Introduction

1.1 Background

New York City and Toronto are two prosperous cities and they are both financial capitals of their respective cities. From a geographic point, United States adjoins Canada and these two cities are close with even a driving distance. 8.6 million people live in New York City [1] and 2.9 million people live in Toronto [2]. To be more specific, if we just look at Manhattan, which is the center of New York City, there is still 1.7 million people [3]. So, what attracts our interests is that how do these two cities share the similarity and how could they be different from each other. By asking this question, we will have a general overview of these two cities and people who want to choose one of them to live in could be benefited.

1.2 Problem

To understand the characters of each city we should explore the venues and analysis its make-up and categories. Data would contribute to the city exploring should include the detailed neighborhood and venue information. To plot the map of both cities, the data should also contain the coordinates (latitude and longitude) of each neighborhood.

1.3 Interest

This exploring will benefit those people who want to have a better understanding of both cities especially for those people who want to choose one of them to live in. Since the exploring will also focus the detail venues in the city, it could also provide the traveler and native resident a general map of the special properties of the city.

2. Data description and acquisition

2.1 Get each necessary dataset

- Our dataset for New York City already existed in the web https://geo.nyu.edu/catalog/nyu_2451_34572. From this web link we can get the .json file and therefore get the coordinates of each borough. But we can actually be even faster since we already saved it in the server: https://cocl.us/new_york_dataset.
- Our dataset for Toronto will be obtained from the Wikipedia link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. We scrape the table from the webpage and then clean the data.
- We use **Foursquare API** to get the detailed venues of each Borough of both cities.

2.2 Data cleaning

- For New York City (Manhattan) data, we clean the data by choosing “borough”, “Neighborhood”, “latitude” and “longitude” columns and create the data frame that only contains these columns. The problem is there will be too much data for just comparing with Toronto. So we will then only select the data for Manhattan to reduce the dataset scale for better comparing with Toronto. We do this by filtering borough to “Manhattan”.
- After webpage scraping for Toronto, we need to get rid of the “not assign” value and create a data frame with postal code, neighborhood and borough. Luckily we already got the coordinates source (http://cocl.us/Geospatial_data) and all we have to do is to merge this table with our scraped data frame to get a new data frame with each latitude and longitude. Some detail needed to be aware is that both before and after merging data frame we need to standardize some column names like change “Neighbourhood” to be “Neighborhood” to let further exploring be benefited.

2.3 Data usage and methodology for solving problems

From the datasets above we get the corresponding borough, neighborhood and coordinates information. From Foursquare API we get the detailed venues information. The coordinates will help us plot the geographic map of both cities and see the distribution of the neighborhoods. The venues will help us have a general overview of the city facilities. The specific type of venue can also be put into a further discussion and exploration.

We will first explore what is the make-up of those venues and try to figure out the top 5 venues in each area. We will then use those venue data to explore some key venues like restaurant, museum and gym. After that we will focus on one most common venue and explore its categories and corresponding make-ups, like a kind of restaurant. Finally we will be even more specific and try to locate this venue category in the city maps and observe its distributions.

To save all the information we use GitHub repository in this study. We use python BeautifulSoup library to scrape the web data and we use python Folium library to visualize geographic details. We also use python matplotlib to plot some analytical graphs to illustrate the similarity of differences between two cities in venues.

3. Exploratory Data Analysis

3.1 Create the data frame and virtualize the geographic map

By loading the data from the web link above, we create the data frame like below Figure 1 for New York City:

	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

Figure 1. Data Frame of New York City Coordinates

Use Folium library we create the New York City map as Figure 2. As mentioned above, there will be too many data points (neighborhoods) in the whole New York City Area, so we will then slice the data frame to only include Manhattan Borough. The map to show the neighborhood in Manhattan will be like the Figure 3 as below (blue markers):

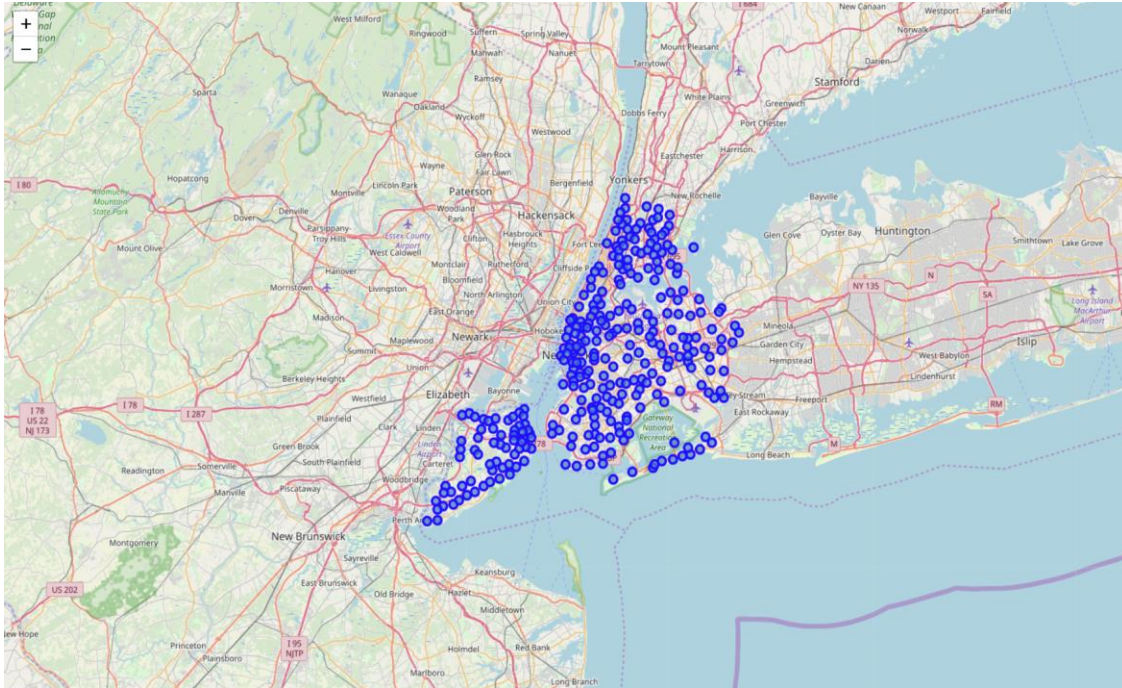


Figure 2. New York City Map with Marks for Neighborhood

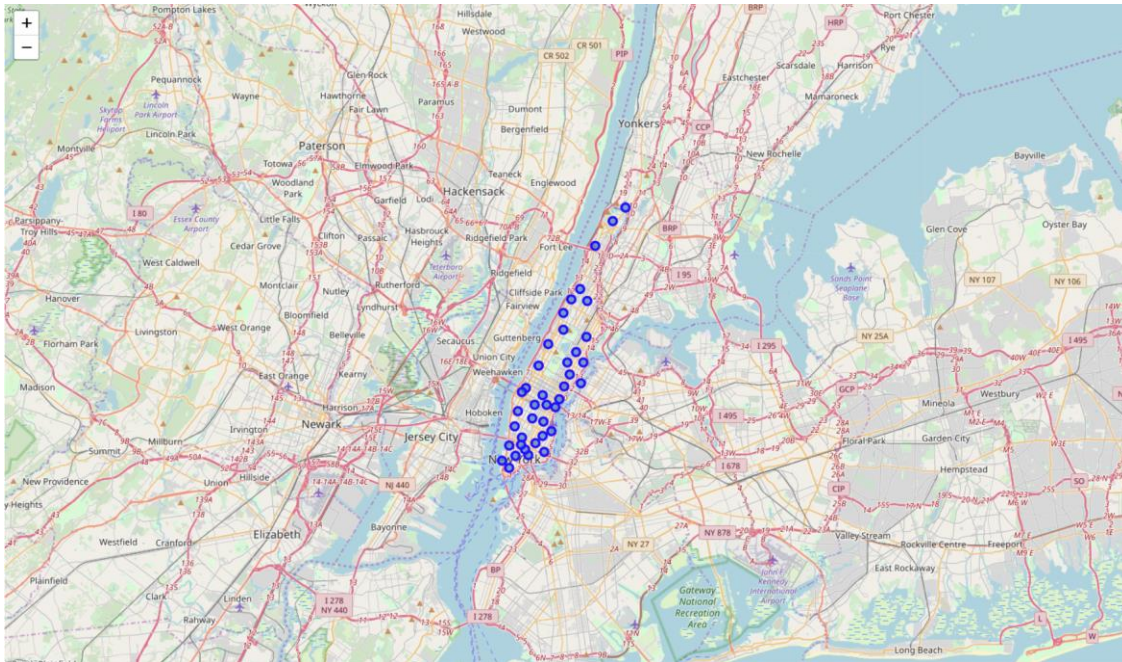


Figure 3. Manhattan Map with Marks for Neighborhood

Follow the same idea, we scrape the Toronto data from the Wikipedia, clean it and merge it with our table with latitude and longitude to create the data frame as below Figure 4:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Figure 4. Data frame of Toronto Coordinates

And then we can virtualize the map for Toronto as below Figure 5 (blue markers for neighborhoods):

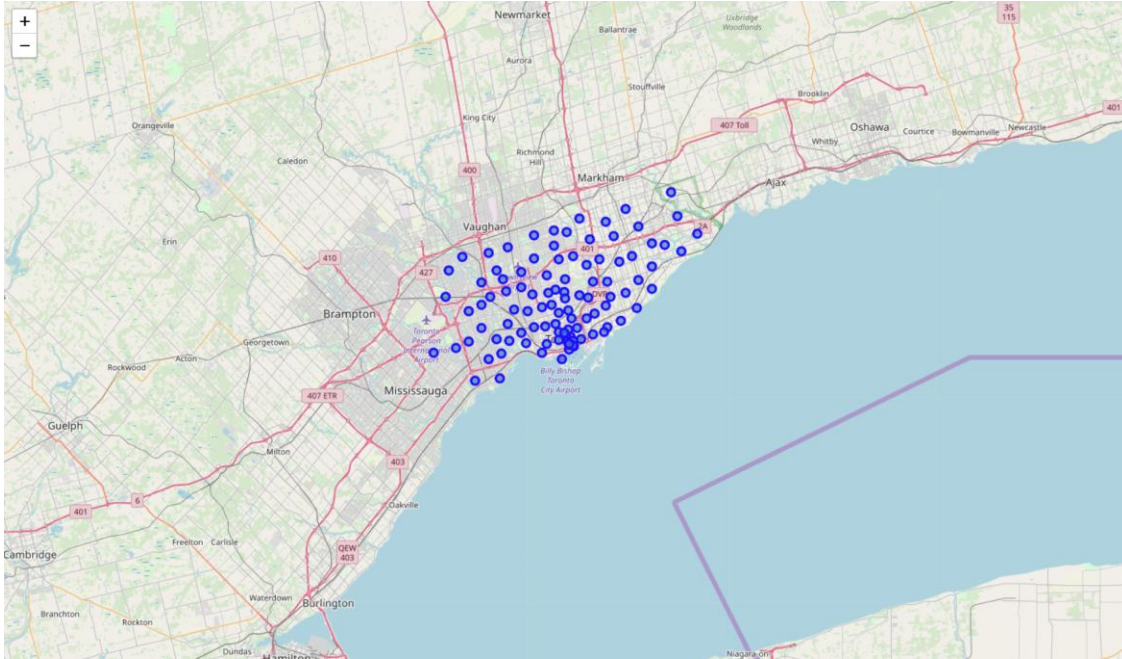


Figure 5. Toronto Map with Marks for Neighborhood

3.2 Get the Venue data from Foursquare API

By using Foursquare API, we can get the data for each kind of venues. Like below Figure 6 for example:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

Figure 6. Manhattan venues Table

We will then count the venues by category and show them in the table like below Figure 7 (as an example for Manhattan venues):

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Animal Shelter	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	ARTS & Crafts Store	Asian Restaurant	Athletics & Sports	Auditorium	Australian Restaurant	Au: Resta
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 7. Manhattan venues count by category

Obviously, we will apply the same process for Toronto and get a similar data frame. And this will be the base for our next discussion.

3.3 Overview of the venues in both cities

In the area we searched, Manhattan has 3317 venues in total and Toronto has 2242. We can present the top 5 venue categories in each city as below Figure 8 and Figure 9:

Top 5 venues in Manhattan:

	count
venues	
Italian Restaurant	127.0
Coffee Shop	124.0
American Restaurant	80.0
Café	78.0
Bakery	72.0

Figure 8. Top 5 Manhattan venues

Top 5 Venues in Toronto

	count
venues	
Coffee Shop	186.0
Café	105.0
Restaurant	62.0
Pizza Place	56.0
Park	55.0

Figure 9. Top 5 Toronto venues

We can see difference here. In Manhattan the No.1 venue is Italian restaurant and in Toronto the most is coffee shop. But we can also notice the similarity that both cities have Coffee Shop as one of their most common venues.

3.4 Explore some key venues of each city

To better understand the special properties of each city we will explore four different kinds of venues: restaurant, museum, gym and bar. We know that these venues might reflect different lifestyle of each city. When select venues by categories, we use the keywords like “Restaurant” or “Museum” and select all rows that contain the key words. This method might not include all the related category but

will maximize the accuracy of our results. By applying matplotlib library in python, the portions of each venues among the total venue are shown below as Figure 10:

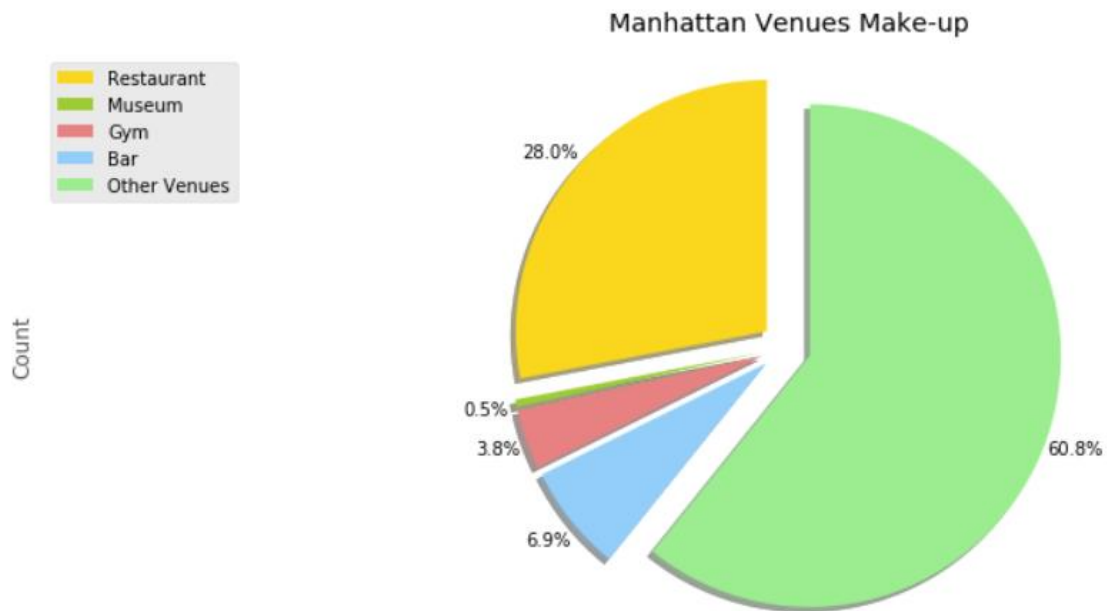


Figure 10. Manhattan Venues Make-up

As we can see in Manhattan, 28% of the venues are restaurant. And 6.9% of the venues are bars. Follow this idea, we can also plot for another city, as shown in the Figure 11 for the venues make-up in Toronto:

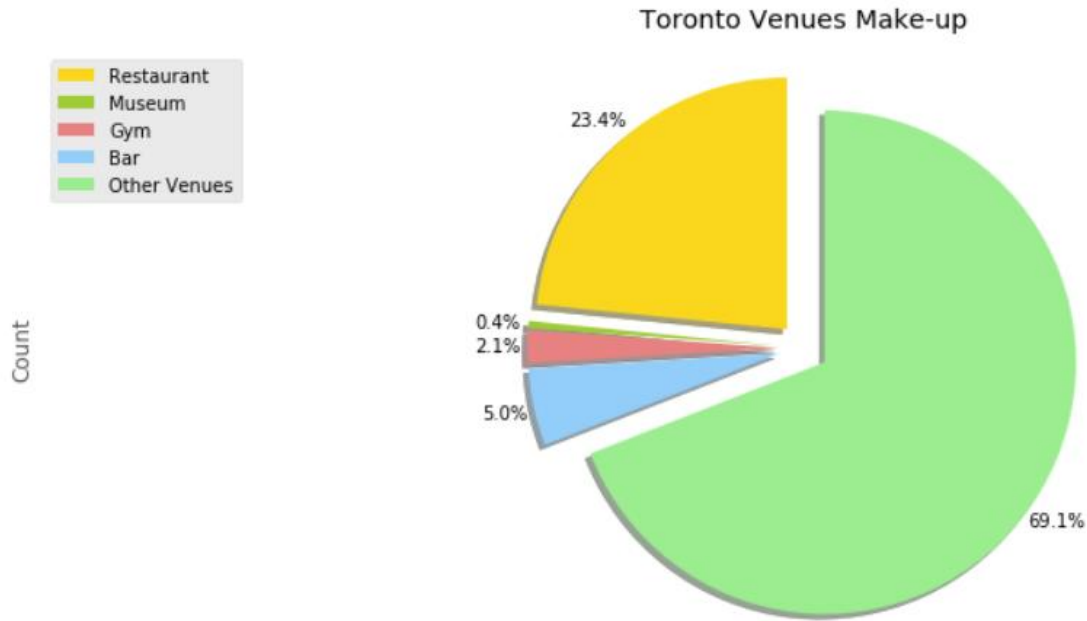


Figure 11. Toronto Venue Make-up

As we can see in Toronto, 23.4% of the venues are restaurant. And 5.0% of the venues are bars. Obviously, Manhattan has more restaurant and bar portions in total venues. From the discussion above we can find out the difference that might show the different lifestyle: Manhattan has more bar and Toronto has more Coffee Shop.

But to be specific, let's focus on one big category now: restaurant.

3.5 Explore restaurants of each city

From the above graphs we can tell that restaurant has a very big portion of the total venues in both cities. So it drives our interests that what is the make-up of these restaurants? What is the portion of each restaurant categories? These questions might help understand the different preference of each city. The restaurant make-up and actual restaurant count by category for Manhattan is shown in the Figure 12 and Figure 13 as below:

From the graphs above we can tell that Italian restaurant are the most common restaurant in Manhattan. And American restaurant is the second.

The restaurant make-up for Toronto is shown in the Figure 14 and Figure 15 as below:

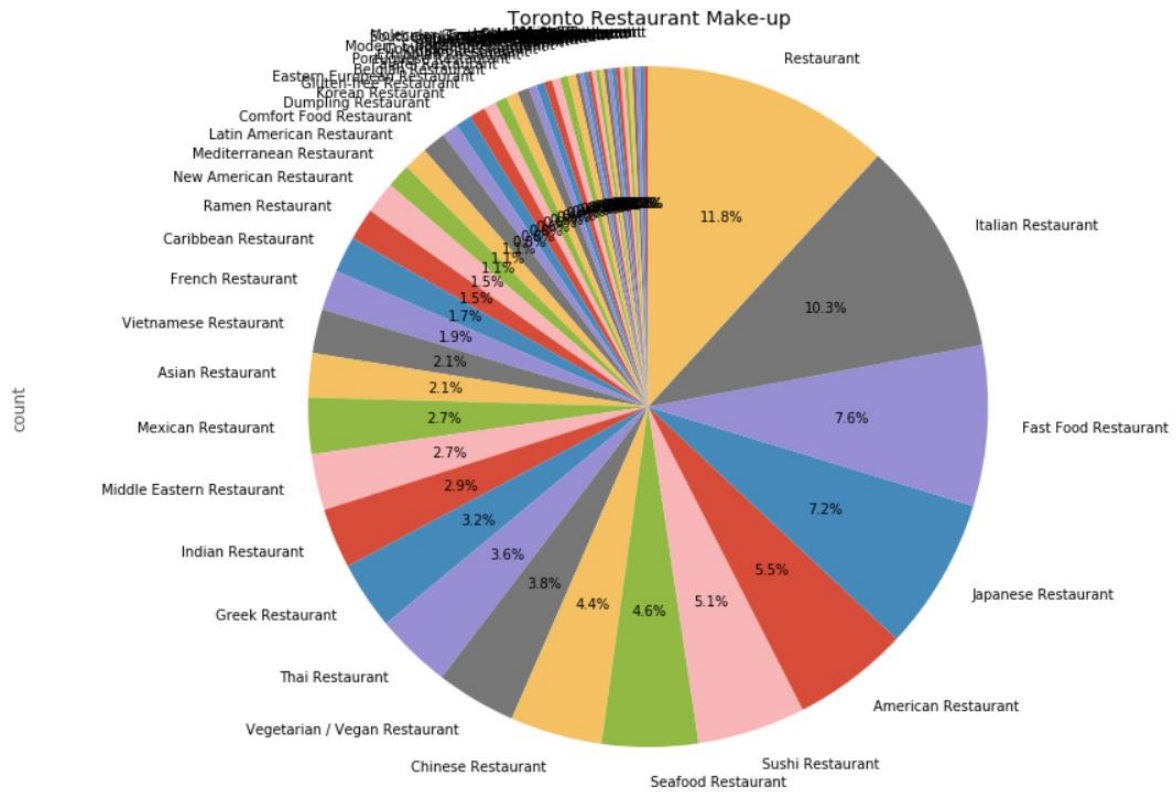


Figure 14. Toronto Restaurant Make-up

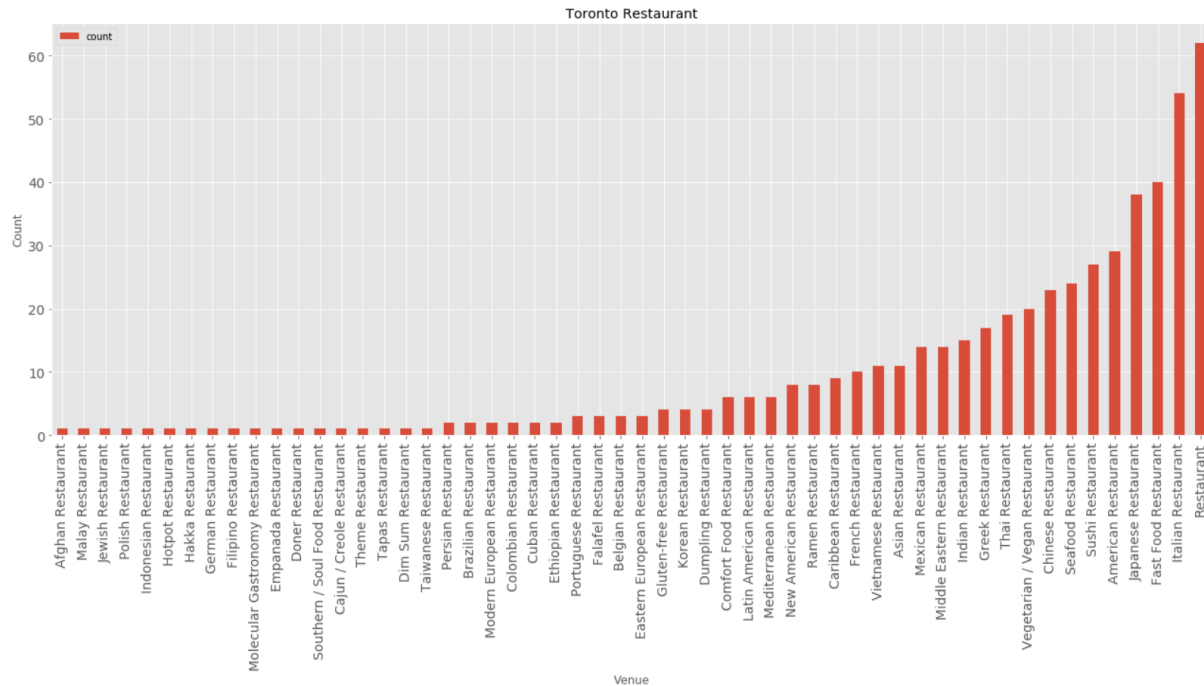


Figure 15. Toronto Restaurant Count by Category

From the graphs above we see that “Restaurant” is in the first place. This is because there are still some restaurants unidentified for categories in the raw data. We can ignore this and when we look at the second one, it’s still Italian restaurant. But we also noticed that fast food followed by Italian restaurant instead of “American Restaurant”, different from Manhattan.

3.5 Explore Italian restaurants of each city

From the above we know that Italian restaurants are the most famous restaurant in both cities. Then it will be interesting to so what are its distributions in the neighborhoods of each city? To do this we use Folium library to plot the geographic map again and mark the neighborhood with Italian restaurant in it. The blue markers stand for each neighborhood and the red one highlights the place with Italian restaurant. The Figure 15 below shows Manhattan neighborhood and Figure 16 shows Toronto neighborhood:

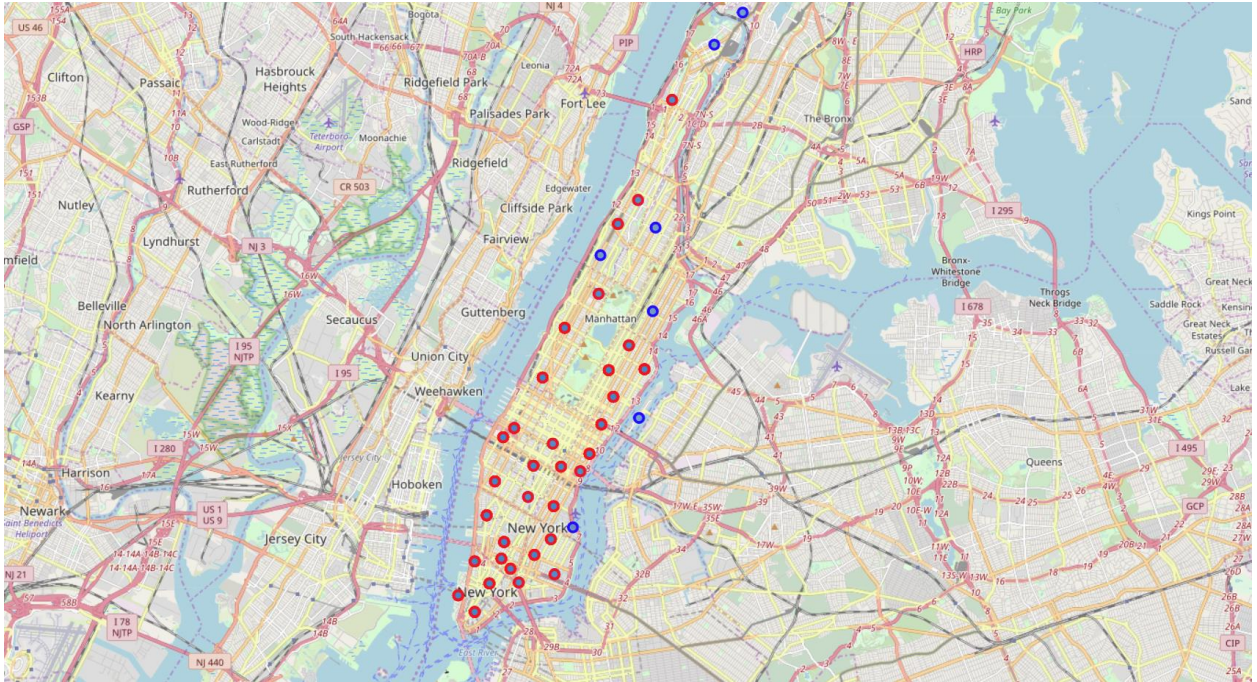


Figure 15. Manhattan Italian Restaurant Distribution

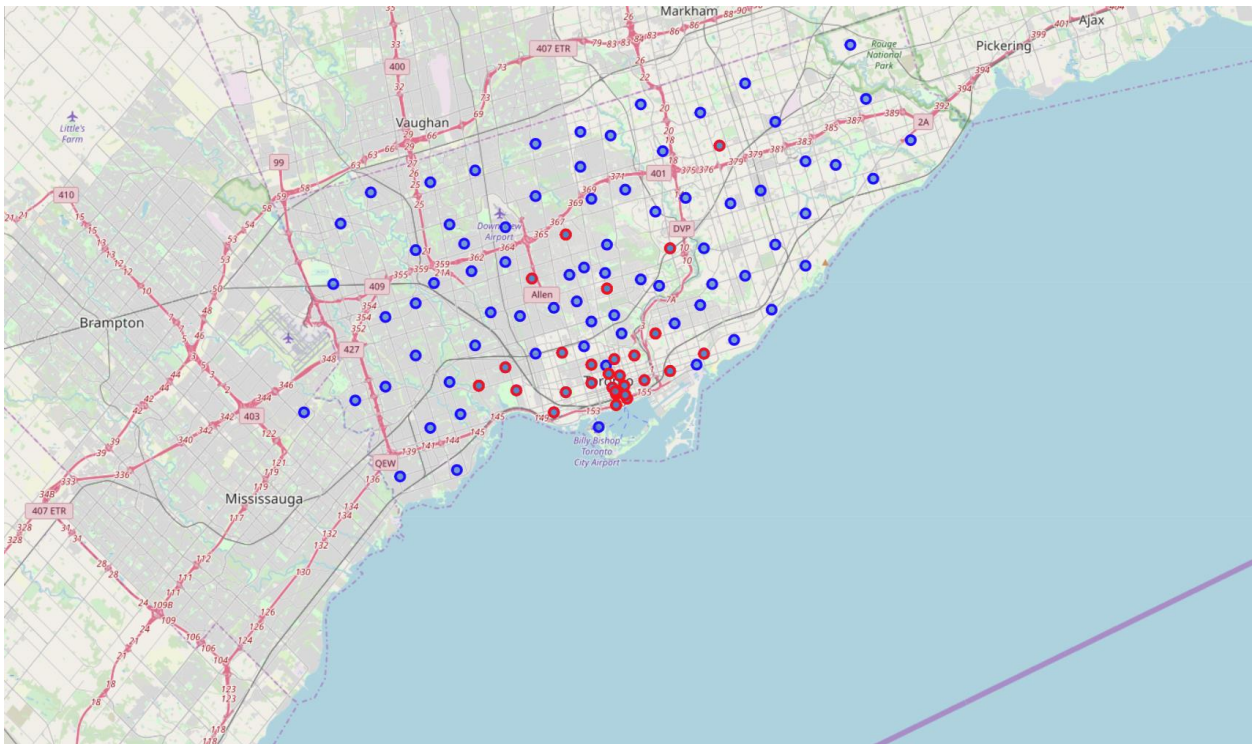


Figure 16. Toronto Italian Restaurant Distribution

From the maps above we see that most of the neighborhoods in Manhattan have Italian restaurants. They are almost everywhere. For Toronto however, the Italian restaurants are more centralized, as most of them are in the downtown or city center.

4. Discussion

As mentioned, these two cities are very close in both geographic locations and economic status of their countries. To explore them we virtualize the map and count the venues. We have the general view first and do research more specific step by step. From both the overview of the venues and the detailed type of restaurants, we did the comparison.

We used the Folium library to virtualize the map. We used the Foursquare API to get the venue information. We also used the matplotlib library in python to plot the analytical graphs. We use GitHub repository to store our research.

5. Conclusion

Manhattan, as the center of New York City, does share some similarities with Toronto. Italian restaurant is the most common restaurant in both cities. And venues like Coffee Shop and Japanese restaurants are both make up a big portion among all the venue categories.

The two cities do have a lot of differences in some detail. For example, Toronto has more Coffee Shop and Manhattan has more bars. Toronto has more “Fast Food Restaurant” while Manhattan more focus more on “American Restaurants”. For Italian restaurant, Manhattan almost has it everywhere. But in Toronto they are gathered in the city center or downtown.

It might not be enough to say if Toronto is a more relaxing place with more Coffee Shop than restaurant comparing to Manhattan, since Manhattan contains more venues that make it to be more diverse. But it’s always a good reference to look at these results to explore what we are interested about the city life.

6. References:

- [1] https://en.wikipedia.org/wiki/New_York_City
- [2] <https://en.wikipedia.org/wiki/Toronto>
- [3] <https://en.wikipedia.org/wiki/Manhattan>