

Battle of Neighborhoods

A COMPARISON BETWEEN NYC AND TORONTO

YUHAO CHEN

Table of Contents

1.Introduction.....	2
1.1 Background	2
1.2 Description of Problems	2
2.Data	2
2.1 Datasets.....	2
2.2 Foursquare API.....	3
3. Methodology	3
3.1 Importing Data and Data Preprocessing	3
3.2 Explanatory Data Analysis	3
3.3 Modelling.....	5
4. Result	6
5. Discussion	7
6. Conclusion	7

1.Introduction

1.1 Background

Toronto is similar to New York city in many ways. They are both the largest city in their own countries. Besides, both cities are centers for immigrations. According to Quora (John Smith, 2017):

“Both the largest cities in their respective countries.

Both the financial and tourist capitals of their countries.

Both have a high cost of living.

Both have a network of subways, trains and buses serving their greater areas.

Both are known for their skyscrapers.

Both are extremely multi-ethnic.

Both are theatre capitals of the English-speaking world.”

1.2 Description of Problems

In week 3, we explored New York City (NYC) and the city of Toronto by segmenting and clustering their neighborhoods. We noticed that both cities have financial centers and a lot of diversified neighborhoods. Giving that the two cities are very similar, it is interesting for us to examine how similar the two cities are regarding neighborhoods and venues. Through data analysis, we can have better understanding on whether New York city or Toronto is more multicultural.

2.Data

2.1 Datasets

In order to make comparison between NYC and Toronto, we make use of datasets from both cities.

First, we downloaded the data set from "lab: Segmenting and Clustering Neighborhoods in New York City". The data is presented on https://geo.nyu.edu/catalog/nyu_2451_34572. After cleaning the data, we export it into the nyc.csv file for further research.

Besides, we also make use of the Toronto data from "peer assignment: Explore and cluster the neighborhoods in Toronto". The data is produced by merging two datasets from http://cocl.us/Geospatial_data and https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

Both data contains neighborhoods geographical information. The columns in the data sets are:

'Borough','Neighborhood','Latitude','Longitude'

We can use the datasets above to get Venues information through Foursquare API.

2.2 Foursquare API

Foursquare is a location technology platform dedicated to improving how people move through the real world. In this project, we will use the Foursquare developer API to gather information of venues through the geographic information we gather from our two datasets (New York city and Toronto).

This API provides every specific Venue's information respect to given Latitude and Longitude. Foursquare API return the following information of each venue:

1. Neighborhoods
2. Neighborhoods Latitude
3. Neighborhoods Longitude
4. Venue
5. Name of Venue
6. Venue Latitude
7. Venue Longitude
8. Venue Category

The information can help us make comparison between venues in NYC and venues in Toronto. Also, the categories of venues can reflect the degree of multicultural, which we are interested in explore.

3. Methodology

Our data analysis contains the following three steps: 1. Importing Data and Data Preprocessing; 2. Explanatory Data Analysis; 3. Modelling

3.1 Importing Data and Data Preprocessing

We first import the libraries we need for data analysis. We use NumPy and pandas to do data manipulations. We also import json and request to handle JSON file. Besides, we import geopy, folium map and matplotlib to create a geographic visualizations. In the end, we import sklearn for k-mean clustering machine learning algorism.

After importin those libraries, we load geographic information of neighborhoods from nyc.csv and Toronto.csv, storing them in two different data frames (df, df2).

3.2 Explanatory Data Analysis

First, we use create folium map for both cities to visualize their neighborhoods. Although Toronto and New York are big cities, we noticed that NYC is actually much bigger in size through the map. Thus, there are more neighborhoods in New York City than the ones in Toronto.

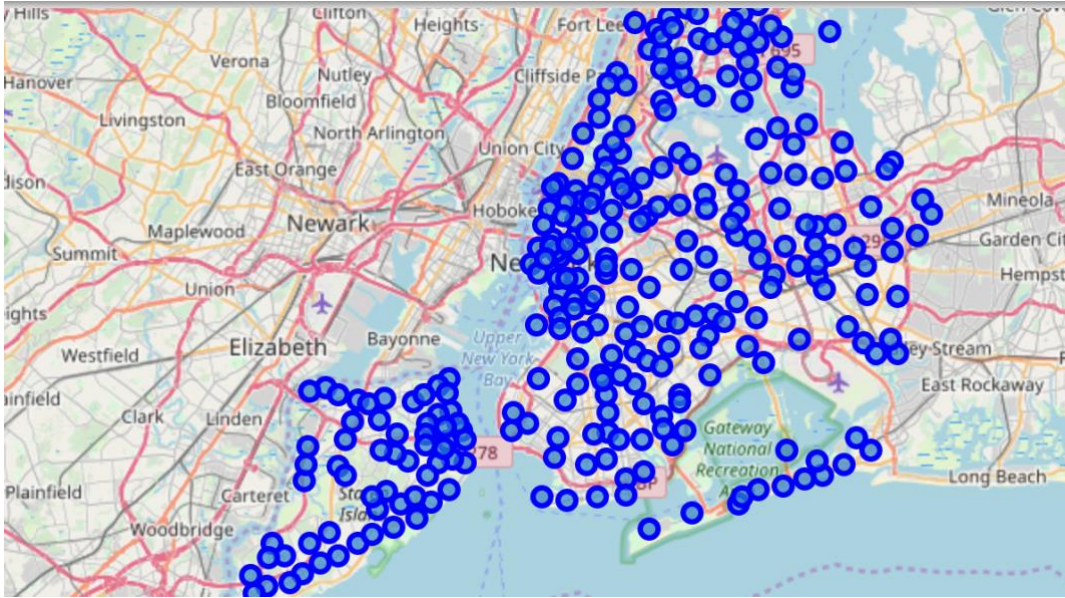


Figure 1: New York City Neighborhoods

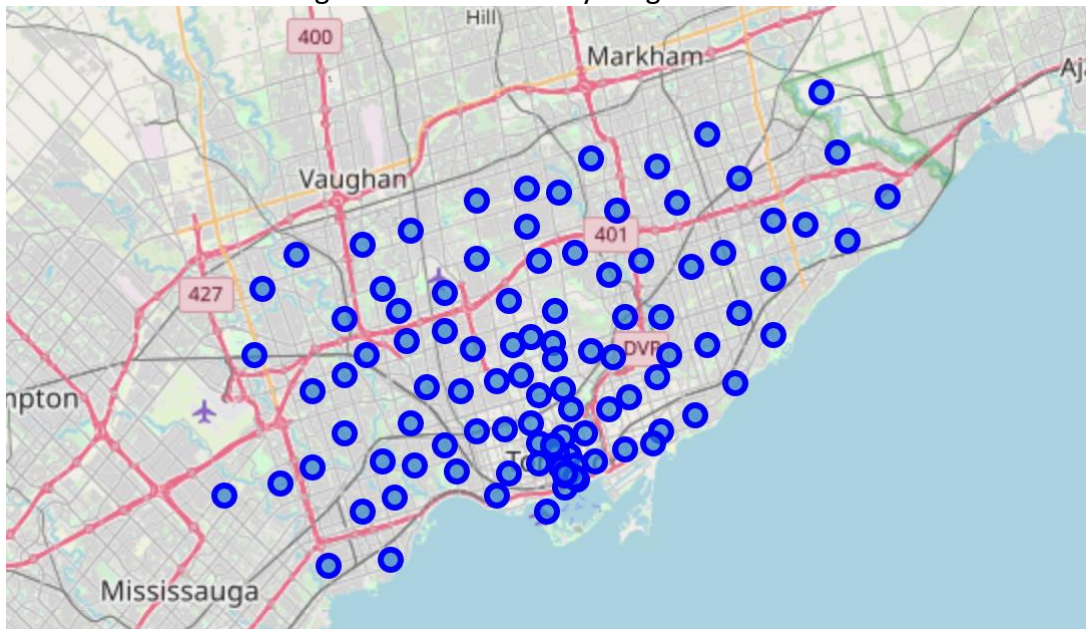


Figure 2: Toronto Neighborhoods

After reviewing our datasets, we start to import data from Foursquare API. Through Foursquare API, we gain information for venues in New York City and Toronto. We store the result of Toronto Venues in 'result' and the one of New York City in 'result2'. We noticed that New York has more Venue data than Toronto.

We begin our analysis by examining the categories of venues in both cities. We examine the count on categories of venues in city centers. Then, we create a dataframe containing the most common venues in every neighborhood for each city. After that, we count the most common venues across different neighborhoods to do comparison between NYC and Toronto.

	name	categories	lat	lng		name	categories	lat	lng
0	The Bar Room at Temple Court	Hotel Bar	40.711448	-74.006802	0	Downtown Toronto	Neighborhood	43.653232	-79.385296
1	The Beekman - A Thompson Hotel	Hotel	40.711173	-74.006702	1	Cafe Plenty	Café	43.654571	-79.389450
2	Alba Dry Cleaner & Tailor	Laundry Service	40.711434	-74.006272	2	Japango	Sushi Restaurant	43.655268	-79.385165
3	City Hall Park	Park	40.712415	-74.006724	3	Rolltation	Japanese Restaurant	43.654918	-79.387424
4	Gibney Dance Center Downtown	Dance Studio	40.713923	-74.005661	4	Sansotei Ramen 三草亭	Ramen Restaurant	43.655157	-79.386501

Table 1 & 2: First 5 results of venues categories in NYC (left) and Toronto (right)

3.3 Modelling

We make k nearest neighborhood (k-mean) clustering method to group the venues into five clusters. The method is imported from the library sklearn. Then, we create folium map to visualize the resulting clusters.

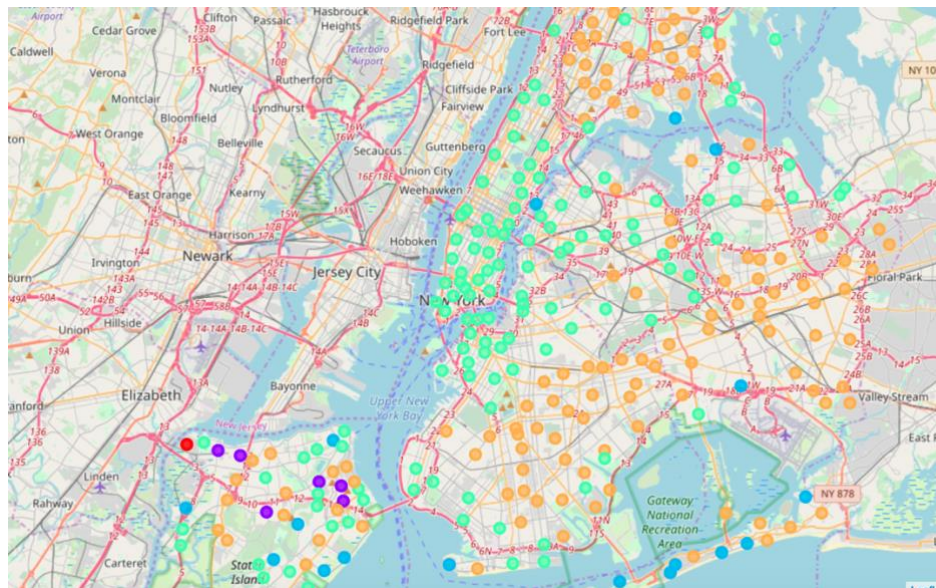


Figure 3: Clustering result of NYC venues

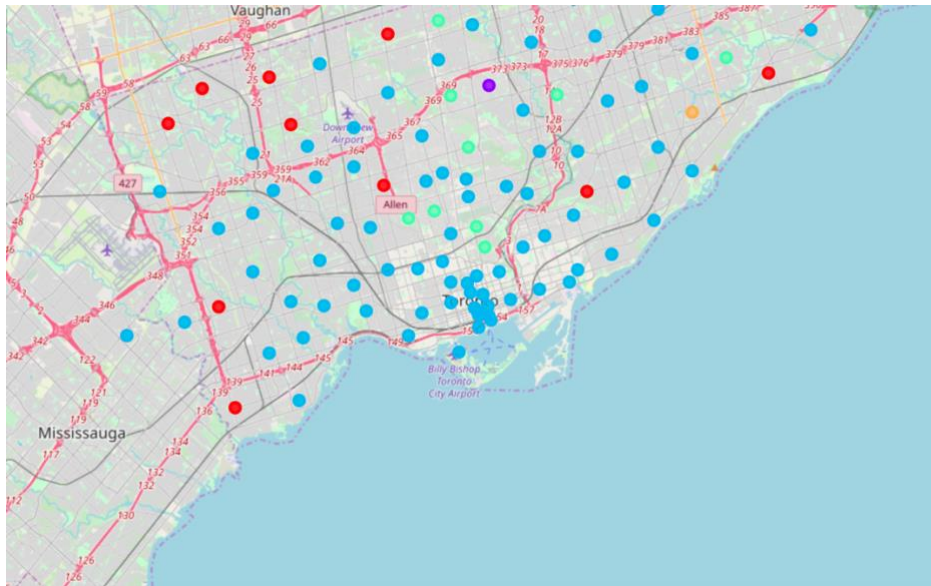


Figure 4: Clustering result of Toronto venues

4. Result

We believe that New York is more multicultural compared to Toronto according to our analysis.

We find that Toronto's venues are less diversified in central Toronto: most of them are restaurants. Both cities have a lot of coffee shop. However, Toronto has more a lot of Japanese/sushi restaurants, following by other Asian food restaurants. In contrast, New York City has more western food restaurants (French restaurants, Italian restaurants and steak house).

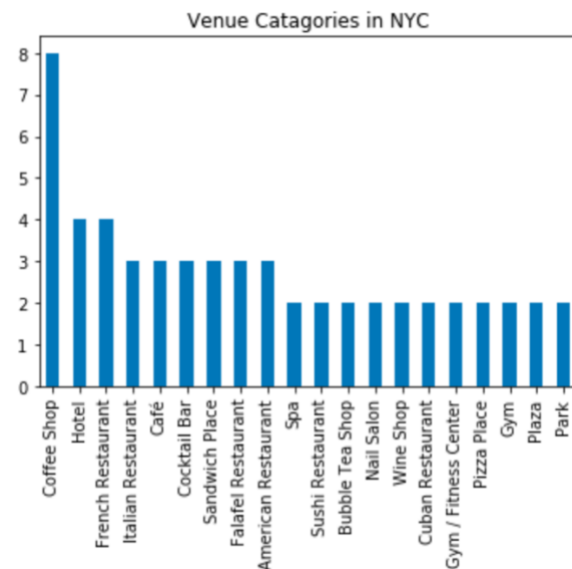
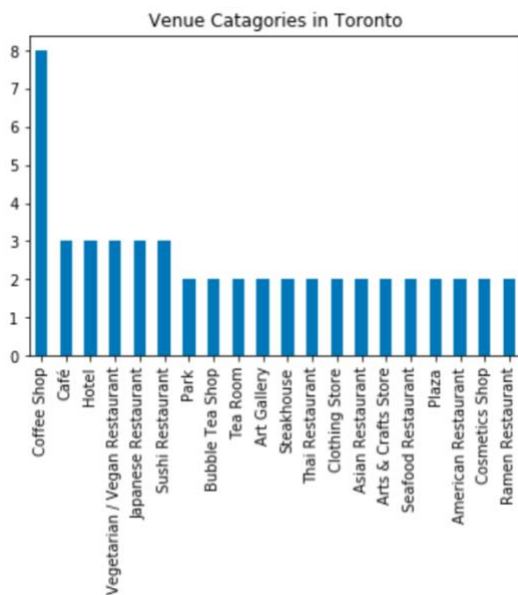


Figure 5 & 6: Categories of Venues in central NYC and Toronto

We also noticed that, about half of the most common venues in NYC neighborhoods are pizza place, following by various venues. While the most common venues in Toronto are coffee shops. We also noticed that some neighborhoods in NYC are more diverse: they have not only restaurants as the most common venues but also park and bar. While the most common venues in Toronto are restaurant, especially Asian food restaurants.

Pizza Place	56	Coffee Shop	22
Italian Restaurant	35	Park	13
Deli / Bodega	14	Café	7
Caribbean Restaurant	14	Pizza Place	6
Park	12	Fast Food Restaurant	4
Bar	10	Indian Restaurant	4
Bus Stop	10	Home Service	3
Chinese Restaurant	9	Hockey Arena	2
Donut Shop	8	Japanese Restaurant	2
Coffee Shop	8	Middle Eastern Restaurant	2

Table 3 & 4: Count on most common venue category in each neighborhood of New York City (left) and Toronto (right)

5. Discussion

In our study, we compare NYC and Toronto regarding the degree of multicultural through neighborhoods data and venues data from Foursquare API.

The main short coming of the study is that the venue data only reflect part of the multicultural elements. There are other essential components which we do not include in our analysis, such as population, schools and cultural centers, etc. Future study can combine more related elements regarding multicultural to establish a more comprehensive research.

6. Conclusion

Since New York City and Toronto are two similar multicultural cities, we want to know more about which city is more multicultural respective to venues. Thus, we make use of NYC and Toronto data, as well as Foursquare API, to analyze neighborhoods venues data in both city. Then, we use k-mean clustering method to cluster all the venues into five groups. We find that, compared to Toronto, New York City is much bigger in size and has more venues. Also, New York City has more various types of venues. The venues in New York City are more multicultural while the ones in Toronto consist of mainly Asian restaurants. Thus, we conclude that New York is more multicultural than Toronto.