Toronto vs New York City

Praveen Kumar Y
3-12-2019

1.Introduction

1.0 Background

From an Ancient show that it is characterized by population density, symbolic function, and urban planning, have existed for thousands of years. In Modern day(Today) Cities generally have extensive systems for housing, transportation, sanitation, utilities, land use, and communication. Their density facilitates interaction between people, government organizations and businesses, sometimes benefiting different parties in the process.

A city is distinguished from other human settlements by its relatively great size, but also by its functions and its special symbolic status, which may be conferred by a central authority. The term can also refer either to the physical streets and buildings of the city or to the collection of people who dwell there, and can be used in a general sense to mean urban rather than rural territory.

And mostly Cities have become a main target areas for Business Employment and many other sources of livelihood for individuals migrating from different parts of the world.

1.1 Business Problem

As we see many people migrate to different cities in the world for better livelihood, life, employment, entertainment etc.

Our take in this project is to be able to compare two different cities based on population , ethnicity and similarity of neighborhoods based on the FourSquare Venues Data.

Upon successfully analyzing the data, one should be able to see the similarity or dissimilarity between New York City and the City of Toronto from the perspective of Population diversity, Ethnicity and neighborhood venue data.

1.2 Target Audience:

Target Audience would be any person or business who would be interested in comparing the Cities New York and Toronto based on their Population Diversity, Ethnicity and Neighborhood Data using FourSquare API.

2. Data Gathering and Cleansing

Data collected from different entities such as wikipedia and official data websites presented by respective Countries/Cities.

2.1 Sources of Data:

Population Demographic Data for the Cities were taken from below Sources

City of Toronto Population Demographics and Ethnicity data:
https://en.wikipedia.org/wiki/Demographics_of_Toronto

New York City Population Demographic Data :
https://en.wikipedia.org/wiki/Demographics_of_New_York_City

Population by Gender data : https://newyork.areaconnect.com/statistics.htm

Location Data:

FourSquareAPI :

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
CLIENT_ID,

CLIENT_SECRET,
VERSION,
LATITUDE,
LONGITUDE,
radius,
LIMIT)

How we are going to use the Data retrieved:

Business Problem : To compare two different cities based on location data from FourSquare API and Population Demographics gathered from various Wiki pages.

Using extracted data from above sources :

1. We will present the Total population data for each City(including Gender wise population demographics)
2. We would present the Historical population data on how the population has trended over the years
3. Indicate Top Ethnic Groups in each City for finding similar/dissimilar traditions
4. Latitude & Longitude Location data extracted from FourSquareAPI and
5. Map the venue data by clustering the neighborhoods.

2.2 Data Cleansing

Data downloaded or scraped from multiple sources were combined multiple tables to access and portray as per the need. There were a lot of missing values from the tables as it weren't updated properly over the years.

There are huge set of details in the dataset we have gathered. First, there were a lot of unwanted Columns which may not be useful for the analysis that we are, though we can do extensive analysis , but for the analysis we were conducting it would amount to a lesser number of features even if the data is cleansed, so we have dropped the columns which were not useful.

	NaN	NaN	New York Population:	Sex and Age	Male	Female	Under 5 years	5 to 9 years	10 to 14 years	15 to 19 years	...	Vacant housing units	For seasonal, recreational, or occasional use	Homeowner vacancy rate (percent)	F va (pe
0	NaN	Number	8008278	NaN	3794204	4214074	540878	561115	530816	520641	...	179324	28157	1.7	
1	NaN	Percent	100.00%	NaN	47.38%	52.62%	6.75%	7.01%	6.63%	6.5%	...	5.6%	0.88%	NaN	

2 rows × 99 columns

2.2.0(a) Raw dataset gathered for Gender wise population demographics

We have extracted Ethnicity by country of origin using the belo Ra data Frame just to obtain the Ethnicity and their percentage in the overall population in that particular city.

	Top 20 Ethnic origins in the City of Toronto	Population (2016)[33]	Percentage	Riding with Highest Concentration	Percentage.1
0	Chinese	332830	12.5	Scarborough-Agincourt	47.0
1	English	331890	12.3	Beaches-East York	24.2
2	Canadian	323175	12.0	Beaches-East York	19.7

2.2.0(b) Raw dataset used for finding Ethnicity by Country of origin

We have gathered the data for both the cities from different sources as there is no collective dataset of required data we seek for the analysis . In the gathered dataset, we have chosen the variables/columns like Male/Female/Total Population(above 2.2.0(a)) and similar data cleansing done for gathering the historical population data(below 2.2.1(a), 2.2.1(b)) over the years in both the cities and we have put together only 5 yearly data to be presented in our analysis over the growth of population over the years . We considered data starting 1990 till 2015 from the census data found in each cities/wikipedia pages and collected data put together for the years 1990, 1995, 2000, 2005, 2015.

	Year	Population	percentage change
12	1991	2275771	+3.8%
13	1996	2385421	+4.8%
14	2001	2481494	+4.0%
15	2006	2503281	+0.9%
16	2011	2615060	+4.5%
17	2016	2731571	+4.5%

2.2.1(a) Population History Raw Data

	year	Toronto	NYC
0	1990	2275771	7322564
1	2000	2481494	8008288
2	2010	2615060	8175133
3	2015	2731571	8491079

2.2.1(b) After cleansing

Then we have gathered the location coordinate of each city from the identified sources and listed out the neighbourhoods in each city according to the data available at hand. We have extracted data for NYC neighborhoods from here and the Toronto data is gathered from here .

Further we have gathered FourSquare Location data using the FourSquare API available and gathered the venue data for both the cities (New York CIty and Toronto) and map them in a Word cloud to represent the top venues located in each of the cities for comparison over the type of venues that are placed in each of the cities .

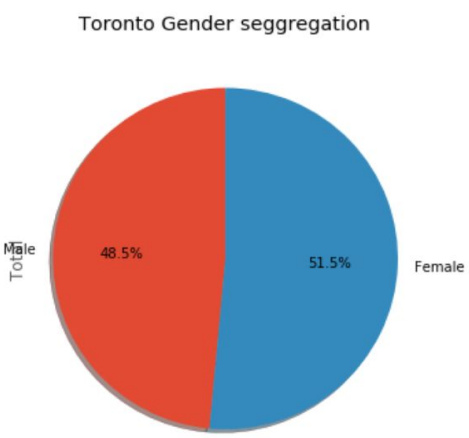
3. Analyzing the Data

Once we have cleaned the Raw data we are left with the data shown below, Firstly, we have looked at the data based on the population and we have combined the data including the Gender segregation in each of these cities out of Total population.

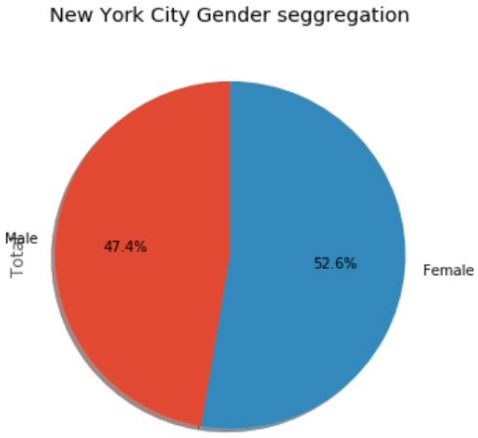
	Total	Male	Female
Toronto	1906495	925070	981420
NYC	8008278	3794204	4214074

3.0 Population data cleaned and merged with Gender segregation

Once got the above data ready we have made a pie chart out of both using a method for creation of the Pie representation of the above data for each city separately as shown in below charts



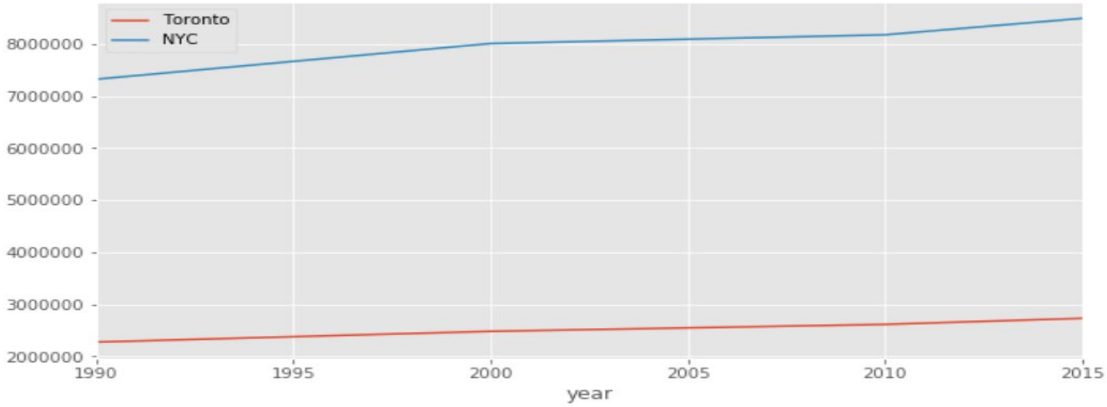
3.0.(a) Toronto



3.0.(b) New York City

When we see the above pie charts for both the Cities we see that both the Cities were almost equally biased based on Gender ratio of the cities, but when we look at the Total Population Data, one can easily state that New York City as densely populated than the City of Toronto i.e., NYC Total pop. is at 8008278 and Toronto Total pop. is at 1906495 , which states that NYC is ~4 times densely populated that City of Toronto.

As we have the population history data ready lets see the population history comparison of the cities in a trend chart using matplotlib libraries , below trend was formed as part of the population history data e have seen in fig. 2.2.1(b)



3.1.0 Population Trend chart for NYC and Toronto between 1990 - 2015

Now it's time we use the ethnicity data based on country of origin and form a word cloud based on the results we got from the Ethnicity dataframe for each city.

3.1.1 Ethnicity by Country of Origin for the CIty of Toronto

From the technical standpoint we see the Primary ethnic groups present in the City of Toronto are of the Origin Chinese,English,Canadian,Irish,Scottish,East Indian, Italian, Filipino,German,Polish etc.,

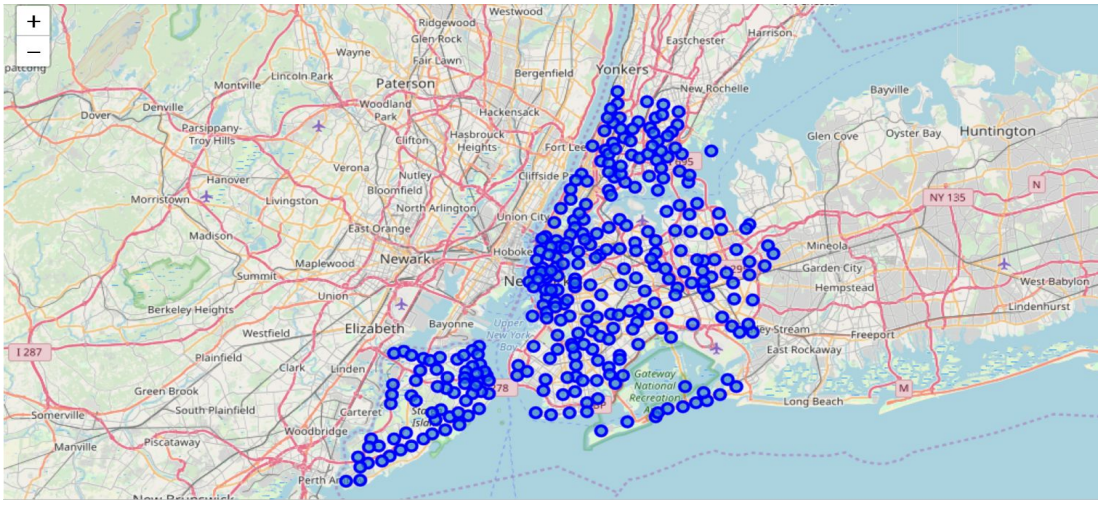
And also we can see the top ethnic groups present in the New York City were of the origin mainly from Countries like Dominican Republic, China, Mexico, Jamaica,

India,Ecuador, Guyana, India,El Salvador, Trinidad & Tobago, Korea, Haiti, Italy etc., as depicted in the figure 3.1.2 below

3.1.2 Ethnicity by Country of Origin for New York City

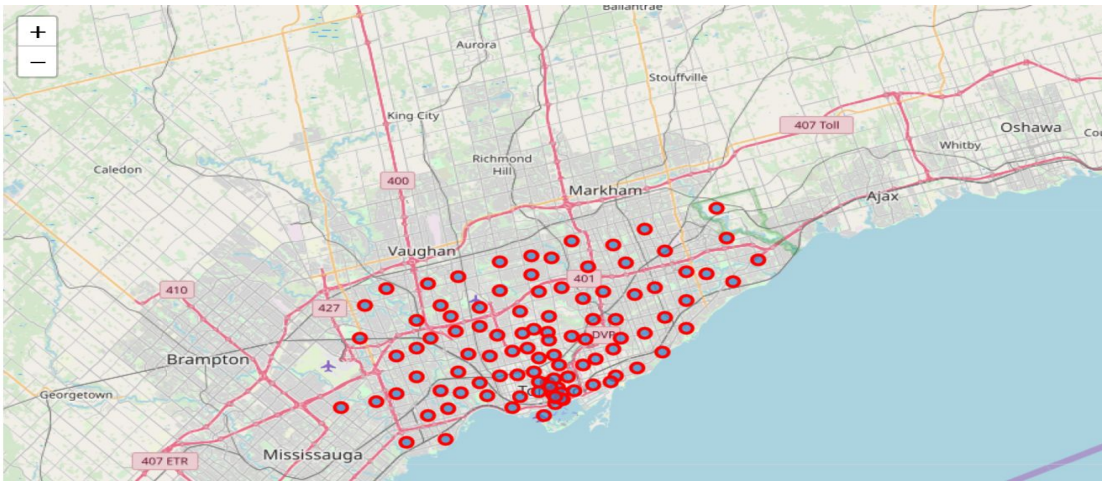
When we compare both 3.1.1 & 3.1.2 we see that a different set of Ethnic groups present in the Cities one comprising more from Europe end(like English, Scottish, Italy, Filipino, French etc.,) , where the other City(New York) mostly comprise individuals from China, Dominican Republic, Mexico, Jamaica, Ecuador, Guayana etc.,

As we have seen the population demographic data for both cities and Ethnicity present in each of the cities , let's look at the Map view of both cities with neighborhoods mark down using Folium Maps Libraries.Below we see the Folium Maps for both NYC and Toronto .



3.1.3 Map View of New York City (Folium Maps)

In the above fig. 3.1.3 We can see how densely populated the New York City really is and the number of neighborhoods present in it are ~306 , which in itself states that New York City is a very highly populated City with a lot of footprints of people from different countries situated in a number of neighborhoods placed across the City. When we have a look at the City of Toronto , geographically as we see on the map below Toronto seems to be a much smaller City than New York and definitely a far less populated with just ~103 neighborhoods (merely with a population of about 1906495), which is about 1/3rd of the total neighborhoods present in the New York CIty.



3.1.4 Map View of the City of Toronto (Folium Maps)

The above data is completely extracted from the data sources mentioned in the “Data Gathering and Cleansing” Section where we put together the data already available online about the Neighbourhood/Borough details with exact Location coordinates for mapping of the above Folium Maps for each City, individually depicting the neighborhoods present in them.

By Comparing the Folium Maps above for each of the cities listed, we can conclude that the area or the Geographical expansion of the Cities, New York City is more when compared to the City of Toronto.

Now that we have the location and neighborhood data, we can pull in the venue data we have obtained from FourSquare API for further analysis based on the type of venues present in both Cities(neighborhoods of each city). What we see below is a Word cloud form of the location data showcasing the Variety of Venues present in the New York CIty . Top 100 venue data received as part of the FourSquare API data

extraction. We have limited the results to 100 as it has retrieved a lot of Venue data for New York City

3.1.5 Top 100 type of Venues in the New York City within a radius of 500 meters

Below would be representation of the variety of venues present in the heart of the City of Toronto . The data we have received is just around 5 to 7 venues when we keep the radius as 500 meters from the Toronto Coordinates, so we have increased it to 1000 meters to present a better picture of the Venue data for Toronto alone(still the number venues we had as only ~26, which is , when compared to New York City is a fraction the total venues present in the NYC).

3.1.6 Top 100 type of Venues in the City of Toronto within a radius of 1000 meters

4. Results & Discussion

Our analysis shows that there is a huge difference between the cities in terms of Population Demographics , when compared to New York City , Toronto seem to be very less populated than NYC, though the gender ratio seems to be of almost the same ratio(b/w Male vs Female).

When we look at the population history of both the Cities It trended similarly, upwards in the trend chart presented above. Although we see a huge gap in between population density that both the cities possess.

Ethnicity by country of origin of the people in both the Cities share a variety of ethnic groups like the NYC mostly consists of the Dominican Republic, China, India, Jamaica, Ecuador, Korea, Guyana etc., Where as the Toronto's Ethnic groups consists of people from English, Canada, China, Germany, Italy, French etc..

Upon looking at the location data from FourSquare we found that there were only fraction of venues in Toronto (~7) than compared to NYC (>100) , which obviously states that NYC is a very renowned City in terms of population & cultural diversity , ethnicity, Business.

By the above factors we can also state that New York City has far more opportunities than Toronto in terms of Business/Shifting to another City, if you were looking to shift to a new City for Employment opportunities. We see that it's a pretty good option to shift to the NYC as the analysis suggests they got more opportunities than in Toronto, In a manner , the above data can suggest a person/individual to decide the top categories based on population demographics .

Now the decision is left with the Individual who compare the data available for both of these cities and decide whether to shift to a New City or venture in suitable one in these cities.

5. Conclusion

Purpose of this project was to compare two cities and find the differences and mostly seeking to find a variety of differences in terms of population density , ethnicity, Business availability. Final decision has to be made by individuals/stakeholders based on specific characteristics of neighborhoods and locations in the city, taking into consideration additional factors like attractiveness of each location, Educational quotient, real estate availability, prices, social and economic dynamics of every neighborhood etc.

6. Future Scope

Think the analysis presented above shows only a short fraction of the many varieties of data that an individual/Business can gather and put it to a greater realistic purpose, like the idea that we can put together a variety of data say Crime Data, Employment rate in those cities, Status of the people living in the cities, Amenities available, Real estate availability..etc., there will be a million possible things we can put together and put a real case for use of such data in a very constructive way useful for future generations of possible improvements, challenges future predictions on employment, Crime patterns, looking at possible improvements in livelihood of the people living those cities based on the actual/real time data & many more.

I leave it to your imagination as the possibilities are vast…

THE END