Finding the best Neighbourhood To live between Two Neighbourhoods Using Data Analysis & Data Visualisation

Dheeraj P

April 6,2020

Introduction

Background

In this **capstone project** I will be creating a program to compare the neighbourhoods of the two cities namely **Brooklyn** from New York and **Downtown Toronto** from Toronto and determine how similar or dissimilar they are.

Brooklyn and Downtown Toronto are major cities with plenty of activities and amazing night life. These places have a lot of **restaurants, cafes, pubs** and so on. So these neighbourhood is best for my analysis and visualization.

Problem

I will be comparing all the venues , hotspots to visit, number of residential friendly places, etc. As we know both Brooklyn and downtown Toronto are we very huge and diverse containing people from all around the world. So by comparing these two neighbourhoods we can find a lot of similarities and see how these places vary.

I will use **foursquare api** to collect all the details and explore these regions and will be using **k means** to cluster different cateogaries of data and display it on the exact location using **folium maps**.

Interest

I will be using the two maps to differentiate between the neighbourhoods and see how one varies from the other. This analysis is mainly done when people are moving from one place to another for various purposes. *They would want to have the same facilities in the new neighbourhood as they have in their current neighbourhood.*

Thus by comparing various major cities based on its amenities and venues we make a proper analysis and visualize the data obtained using ***Data Science.***

Data section

Data Source

As I said in the introduction I will be creating a program to analyse Brooklyn and Downtown Toronto.

I will be needing the data on Canadas neighbourhoods

Link had been obtained from previous labs from coursera:

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

I will also be needing the **New York dataset** so that I can extract all the information from **Brooklyn**. I obtained This data set from the lab as well:

https://cocl.us/new_york_dataset

I will also require a **geospatial coordinate** dataset which contains all the latitude and longitude information of the cities. This link was also available in the labs.

http://cocl.us/Geospatial_data

For Reference I checked the following Kaggle website by Muhammad:

https://www.kaggle.com/hark99/clustering-the-battle-of-neighborhoods/notebook

I will be using **jupyter notebook** to create the program. **Foursquare api** will be used to explore the data.

All the modules which I will be using in the program are listed below:

1. **Numpy**
2. **Pandas**
3. **Matplotlib**
4. **Seaborn**
5. **Json**
6. **Geopy**
7. **Requests**
8. **Sklearn**
9. **Folium**

Data cleaning

I will be web scraping to obtain the table of **Canada's neighbourhoods** from Wikipedia and converted it into a Dataset using pandas and cleaned it so that it fits for my analysis.

Later I will make new data frame which comprises of **Downtown Toronto** data only.

Form the New York The json file I cleaned the data and only a dataframe containing only Brooklyn data will be formed.

Feature selection

The canadas neighbourhood dataset was cleaned and the non assigned values where removed. The columns where renamed to postcode , borough and neighbourhoods. The index was set to postcode remaing columns where droped. Later only the Downtown Toronto borough was selected and a new dataframe was created. Using the geospatial dataset we found out the latitude and longitude of each location and added it to the dataframe.

	Borough	Neighborhood	Latitude	Longitude
0	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
1	Downtown Toronto	Queen's Park	43.662301	-79.389494
2	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	Downtown Toronto	St. James Town	43.651494	-79.375418
4	Downtown Toronto	Berczy Park	43.644771	-79.373306

Similarly the json file for new York data set was converted to a data frame. The columns where renamed to postcode , borough and neighbourhoods. The index was set to postcode remaing columns where droped. Later only the Brooklyn borough was selected and a new dataframe was created. Using the geospatial dataset we found out the latitude and longitude of each location and added it to the dataframe.

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

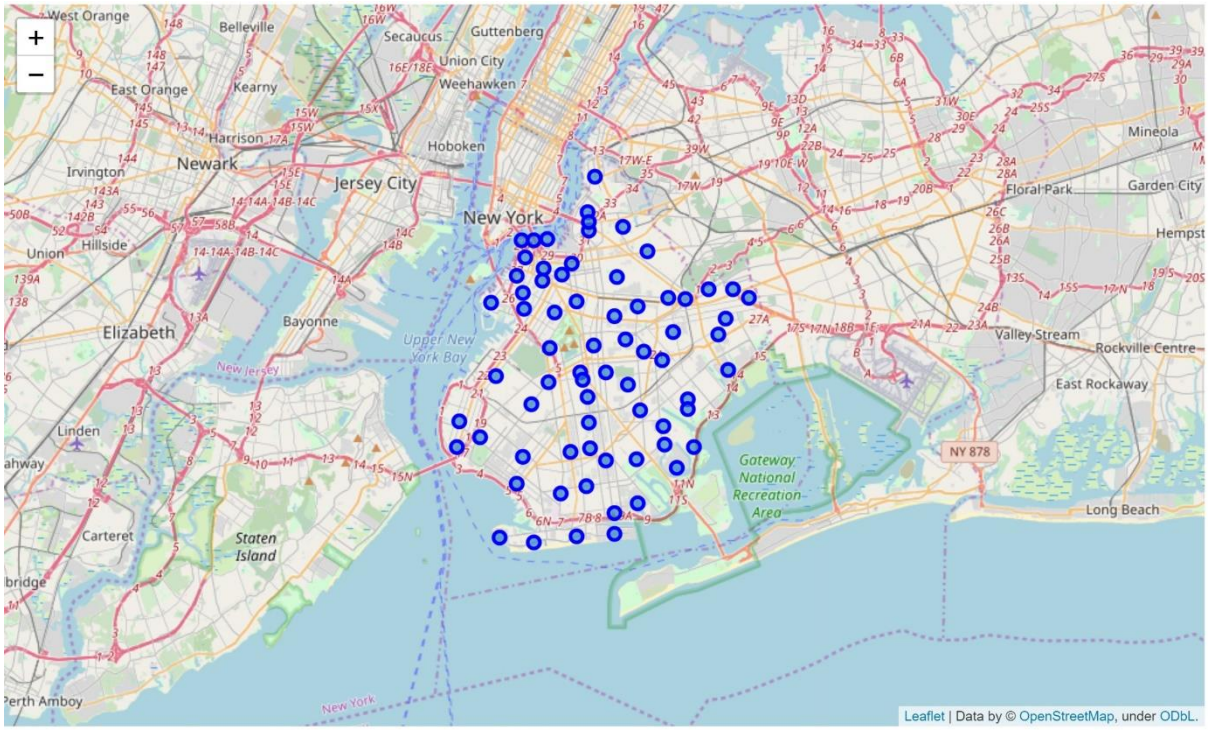
Methodology

Exploratory data analysis

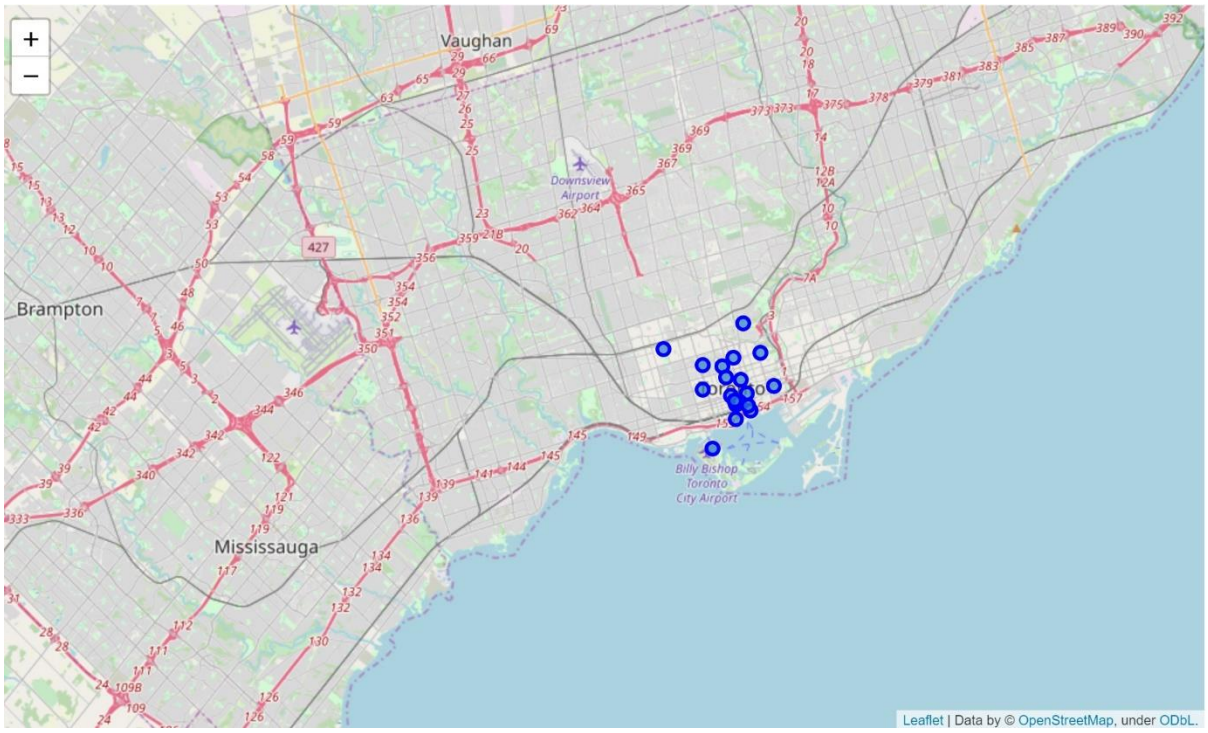
The foursquare api client details and version where initialized to the variables. In order to create a url to get requests.

First I did a exploration through visualization of both the cities of Downtown Toronto and Brooklyn. I requested the venues in that place a marked all of them in their repective maps using folium.

Brooklyn map:



Downtown Toronto map:



Statistical Anaysis

I retrieved all the data life information of rating , location, etc on all the venues located in the above maps using a Jason file for each particular location .

Then performed one hot coding to divide them into particular categories so it will be easy while clustering them later.

Downtown Toronto:

	Yoga Studio	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Baby Store	Bakery	Bar	Basketball Stadium	Beer Bar
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Brooklyn:

	Yoga Studio	Accessories Store	Airport Terminal	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Basel Fi
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Later we created the top 10 venues in a neighbourhood and multiplied it with the one hot coding so that each venue falls in a certain cateogory and is divided into 4 proper clustrers as seen below.

Downtown Toronto:

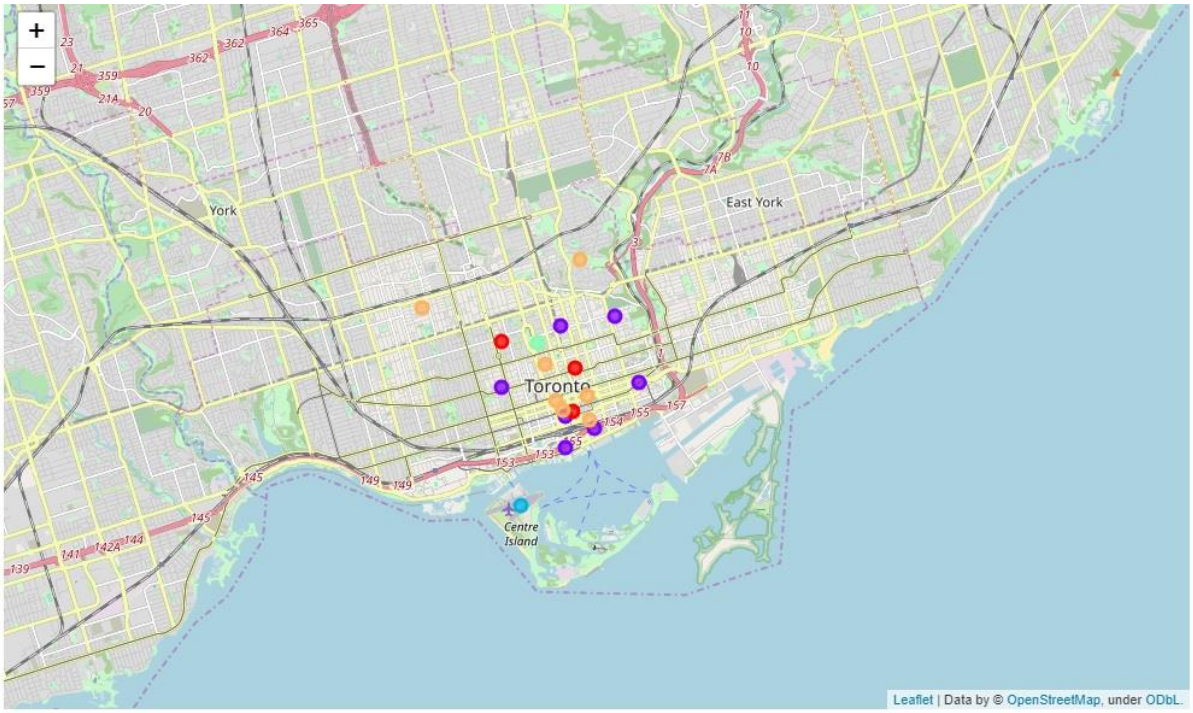
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636	1	Coffee Shop	Bakery	Breakfast Spot	Historic Site	Distribution Center	Pub	Spa	Dessert Shop	Restauran
1	Downtown Toronto	Queen's Park	43.662301	-79.389494	3	Coffee Shop	Yoga Studio	Beer Bar	Distribution Center	Discount Store	Hobby Shop	Diner	Italian Restaurant	Creperie
2	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Café	Coffee Shop	Art Gallery	Movie Theater	Music Venue	Burrito Place	Burger Joint	Clothing Store	Plaza
3	Downtown Toronto	St. James Town	43.651494	-79.375418	4	Gastropub	Restaurant	Coffee Shop	Japanese Restaurant	Food Truck	BBQ Joint	Café	Middle Eastern Restaurant	Cosmetics Shop
4	Downtown Toronto	Berczy Park	43.644771	-79.373306	1	Cocktail Bar	Farmers Market	Beer Bar	French Restaurant	Seafood Restaurant	Park	Coffee Shop	Liquor Store	Bistro

Brooklyn:

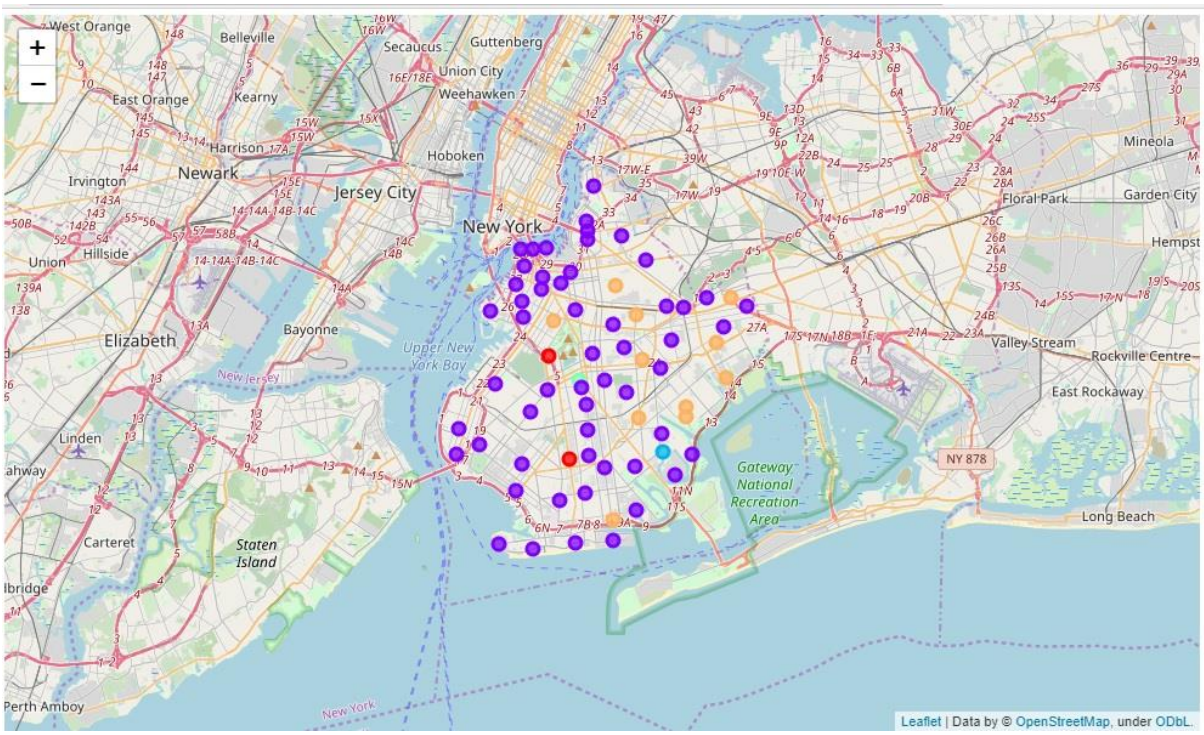
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	1	Spa	Grocery Store	Greek Restaurant	Breakfast Spot	Bagel Shop	Chinese Restaurant	Lounge	Bookstore	Middle Eastern Restaurant
1	Brooklyn	Bensonhurst	40.611009	-73.995180	1	Ice Cream Shop	Italian Restaurant	Sushi Restaurant	Pizza Place	Hotpot Restaurant	Liquor Store	Chinese Restaurant	Noodle House	Flower Shop
2	Brooklyn	Sunset Park	40.645103	-74.010316	1	Pizza Place	Latin American Restaurant	Mexican Restaurant	Bank	Bakery	Grocery Store	Creperie	Café	Mobile Phone Shop
3	Brooklyn	Greenpoint	40.730201	-73.954241	1	Bar	Café	Mexican Restaurant	Yoga Studio	Pizza Place	Polish Restaurant	Record Shop	Cocktail Bar	French Restaurant
4	Brooklyn	Gravesend	40.595260	-73.973471	1	Pizza Place	Lounge	Italian Restaurant	Bakery	Gym	Fish Market	Men's Store	Furniture / Home Store	Martial Arts Dojo

After getting all the information all the clusters where placed on a folium map. This showed us how many different amenities and places are there in the city and exactly where they are located in the neighbourhood.

Downtown Toronto:



Brooklyn:



Based on the maps we can see the various clusters that we split into as: Residential , Comercial, Tourist and cultural.

Result

After analysing the data both the neighbourhoods have a high tourist attraction , lot of restaurants and cafes. They also have a lot of good residential areas like parks etc and proper amenities places like airports and train.

The major difference is that they have different historical places and different origins. This includes monuments ,parliamentary buildings and so on.

Observations and Discussion

Brooklyn is bigger than downtown Toronto it has 3 times more venues than its counterpart as we can see on the map. Not only that the people are diverse as it is evident from the diverse array of restaurants available and the places to visit.

The airport and other transport facilities are nearby and many residential areas are available in Brooklyn than in downtown Toronto.

Conclusion

Based on a the analysis the Brooklyn neighbourhood is more suitable for a highly active and busy person who always wants to enjoy and keep traveling without any financial barriers.

On the other hand if you want a simpler life with balance of nature and entertainment downtown Toronto is the best option and is also a solution financially.