IBM DATA SCIENCE CAPSTON

Which neighborhood in Downtown Toronto is best suited for someone coming from Manhattan

NOVEMBER 16, 2020

IBM - COURSERA
Wenyu Xin

Introduction

Problem

For those who live in the United States, we may have encountered many news headlines that say, "Moving to Canada search spiked in Google Trends". In fact, if someone look into the Google Trends for search term "move to Canada" and its searching index for the past 12 months, one will find this search scored 100 out of 100 and spiked on the Google Trends search index during November 1st and 11th [1].

With the current political and social unrests in the United States, it is understandable for Americans living in the United States to think about moving to Canada. On the day of writing this report, American reported over 130 thousand new Covid-19 cases, total of over 11 million cases, over 247 thousand death, and infection rates are on the rise [2]. Moreover, the question of who the next president is still in the air after almost 2 weeks after the election day.

Interest

The closest country that are friendly to and welcome Americans is Canada. It is the nation where the daily new Covid-19 cases number is only in the 4 thousand and total reported cases is just a little bit over 300 thousand [2]. With many similarities and less political and social unrests than the United States, Canada is the best choice for someone to leave the United States.

Data

Data Sources

Neighborhood data for Toronto is from Wikipedia, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.

Geospatial data for both Toronto, https://cocl.us/Geospatial_data, and Manhattan and neighborhood data for Manhattan, https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json, are provided by the Coursera.

Venue data are pulled from Foursquare API, https://developer.foursquare.com/.

Data Cleaning

Neighborhood data for Toronto from Wikipedia was scraped from the website and transformed into a table. Geospatial data is read into a table as well. The first table did not have geospatial data like the latitude and longitude and the second table did not have the borough and neighborhood information. Thus, two tables were merged by the postal index. Furthermore, this merged dataset contained other boroughs other than Downtown Toronto. Thus, further filtering was done to reach the desired dataset.

Neighborhood information for Manhattan was in JSON format. The dataset was read directly into a table. The only cleaning needed was parsing the data and filtering the boroughs to Manhattan only.

The venue data was pulled directly using Foursquare API. It was also in a JSON format that just required parsing.

Methodology

First, the borough, neighborhood, and geospatial data were either scraped or downloaded directly from the data sources mentioned in the Data section. The Toronto dataset was filtered to only include Downtown Toronto information. The New York City dataset was filtered to only include Manhattan information.

	Borough	Neighborhood	Latitude	Longitude
0	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	Downtown Toronto	Queen's Park, Ontario Provincial Government	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

Figure 1. Parsed and cleaned Downtown Toronto Neighborhood dataset

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Figure 2. Parsed and cleaned Manhattan Neighborhood dataset

Second, the maps for each location were plotted using Folium library to get a visual understanding of the neighborhood. There were no particular data analysis done using mapping.

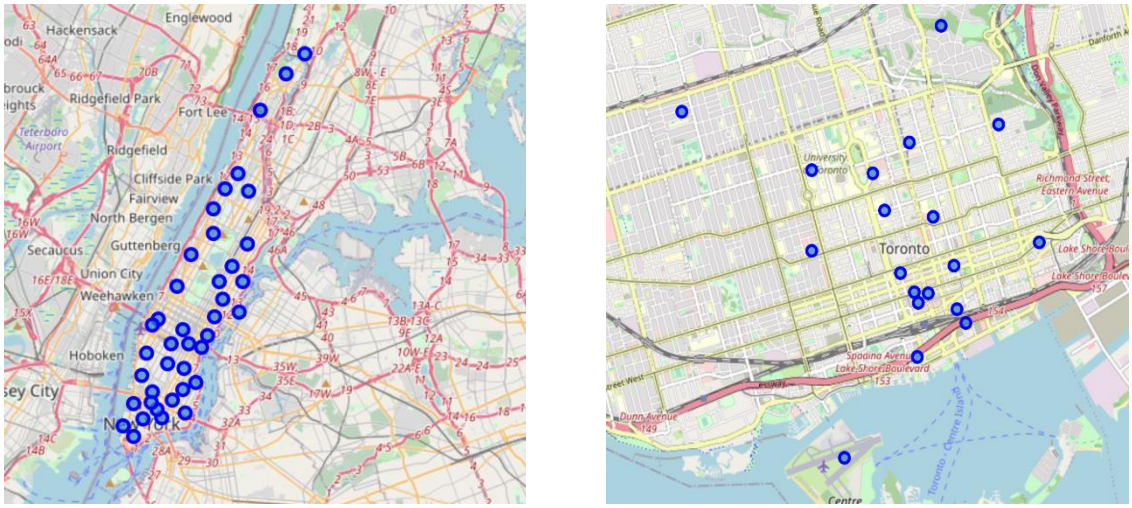

Figure 3. Maps of Manhattan (left) and Downtown Toronto (right) with blue dots representing the cordinates for each neighborhood

After mapping, Foursquare API was used to pull 100 venues from 500 radius of the center of each Toronto and Manhattan neighborhood and those data were stored into two different tables.

:

	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 4. API request venue data for each neighborhood in Manhattan

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

Figure 5. API request venue data for each neighborhood in Toronto

Then the frequency of each venue category for each neighborhood was calculated by dividing the appearance of each venue by total appearances of all venues.

	Neighborhood	Accessories Store	Adult Boutique	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	...	Video Store
0	Battery Park City	0.0	0.0	0.000000	0.000000	0.0	0.0	0.000000	0.000000	0.000000	...	0.0
1	Carnegie Hill	0.0	0.0	0.000000	0.000000	0.0	0.0	0.011628	0.000000	0.011628	...	0.0
2	Central Harlem	0.0	0.0	0.065217	0.043478	0.0	0.0	0.000000	0.021739	0.000000	...	0.0
3	Chelsea	0.0	0.0	0.000000	0.040000	0.0	0.0	0.000000	0.050000	0.000000	...	0.0
4	Chinatown	0.0	0.0	0.000000	0.040000	0.0	0.0	0.000000	0.000000	0.000000	...	0.0

Figure 6. Frequency of venue appearance within each neighborhood in Manhattan

	Neighborhood	Yoga Studio	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	...	Theater	Rest
0	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.(
1	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.000000	0.058824	0.058824	0.058824	0.117647	0.117647	0.058824	0.000000	...	0.000000	0.(
2	Central Bay Street	0.014706	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.(
3	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.(
4	Church and Wellesley	0.026667	0.013333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.013333	...	0.013333	0.(

Figure 7. Frequency of venue appearance within each neighborhood in Toronto

Using the frequency as a selector, top 10 most frequent appeared venues were selected as the top 10 most common venues for each neighborhood were ranked.

	Neighborhood	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	10th Most Common Venue
0	Battery Park City	Park	Hotel	Gym	Coffee Shop	Memorial Site	Shopping Mall	Plaza	Burger Joint	Gourmet Shop	Playground
1	Carnegie Hill	Coffee Shop	Café	Bookstore	Italian Restaurant	Gym / Fitness Center	Gym	French Restaurant	Yoga Studio	Wine Shop	Vietnamese Restaurant
2	Central Harlem	African Restaurant	Bar	French Restaurant	American Restaurant	Chinese Restaurant	Seafood Restaurant	Cosmetics Shop	Cafeteria	Grocery Store	Library
3	Chelsea	Coffee Shop	Art Gallery	Bakery	American Restaurant	Ice Cream Shop	Italian Restaurant	Japanese Restaurant	Park	Cycle Studio	Cupcake Shop
4	Chinatown	Chinese Restaurant	Dessert Shop	Cocktail Bar	Bakery	American Restaurant	Hotpot Restaurant	Spa	Optical Shop	Noodle House	Salon / Barbershop

Figure 8. Top 10 most common venue in each Manhattan neighborhood

	Neighborhood	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	10th Most Common Venue
0	Berczy Park	Coffee Shop	Bakery	Cocktail Bar	Seafood Restaurant	Cheese Shop	Beer Bar	Farmers Market	Restaurant	Sandwich Place	Breakfast Spot
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Service	Bar	Harbor / Marina	Rental Car Location	Coffee Shop	Boat or Ferry	Historic Site	Sculpture Garden	Airport Terminal
2	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Japanese Restaurant	Thai Restaurant	Department Store	Salad Place	Burger Joint	Bubble Tea Shop
3	Christie	Grocery Store	Café	Park	Athletics & Sports	Baby Store	Coffee Shop	Nightclub	Candy Store	Italian Restaurant	Restaurant
4	Church and Wellesley	Coffee Shop	Gay Bar	Japanese Restaurant	Sushi Restaurant	Restaurant	Yoga Studio	Hotel	Men's Store	Café	Bubble Tea Shop

Figure 9. Top 10 most common venue in each Toronto neighborhood

Based on the ranking of the top 10 most common venues, each venue received a score of 1 to 10. The highest ranked venue got a score of 10, the 2nd highest ranked got 1 point less, or 9. This process repeated until there was no point to assign.

	Neighborhood	Park	Hotel	Gym	Coffee Shop	Memorial Site	Shopping Mall	Plaza	Burger Joint	Gourmet Shop	...	Heliport	Skate Park	Asian Restaurant	Dog Run	F
0	Battery Park City	10.0	9.0	8.0	7.0	6.0	5.0	4.0	3.0	2.0	...	0.0	0.0	0.0	0.0	
1	Carnegie Hill	0.0	0.0	5.0	10.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
2	Central Harlem	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
3	Chelsea	3.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
4	Chinatown	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	

Figure 10. Venue with score of 1 to 10 assigned for Manhattan neighborhood

	Neighborhood	Park	Hotel	Gym	Coffee Shop	Memorial Site	Shopping Mall	Plaza	Burger Joint	Gourmet Shop	...	Heliport	Skate Park	Asian Restaurant	Dog Run	F
0	Berczy Park	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
1	CN Tower, King and Spadina, Railway Lands, Har...	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
2	Central Bay Street	0.0	0.0	0.0	10.0	0.0	0.0	0.0	2.0	0.0	...	0.0	0.0	0.0	0.0	
3	Christie	8.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	
4	Church and Wellesley	0.0	4.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	

Figure 11. Venue with score of 1 to 10 assigned for Toronto neighborhood

Next, a set of unique venue from the Manhattan dataset was parsed before calculating the total scores for each and all pair of neighborhoods. To calculate the total score for each pair neighborhood, each Manhattan neighborhood, or each row, was multiplied column-wise with all the dataset in Toronto dataset. Then each row of the product was summed row-wise to produce a single score for each neighborhood pair.

Result

The result was a matrix of all the neighborhood in Manhattan as rows, all the neighborhoods in Toronto as columns, and their product-sum score as the value. The heatmap showed the higher scores as dark-purple color and lower scores as white color.

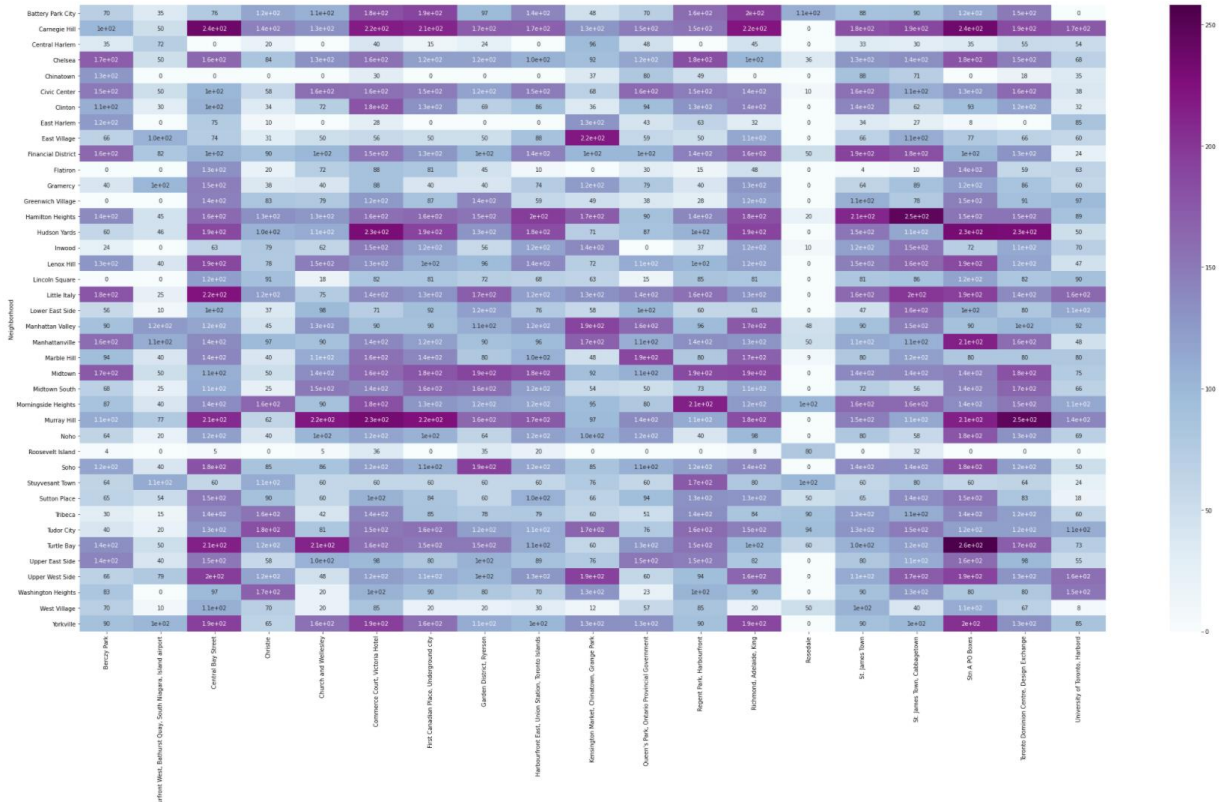

Figure 12. Heatmap with Manhattan neighborhood on the Y-axis and Toronto neighborhood on the X-axis. The color indicates the score levels between each pair of neighborhoods. The darker the color, the higher the score.

Figure 13. A sample of using Soho neighborhood in Manhattan and paired with Garden District, Ryerson neighborhood in Toronto

Discussion

User can select the Manhattan neighborhood which he or she reside on the Y-axis. Then user select the cell with darkest color or the cell with highest cell value. Finally, move down the column to see which Toronto neighborhood best match the venues from Manhattan neighborhood.

Example #1: Bob lives in Chinatown (the 5th row on the Y-axis). The best neighborhood in Toronto to live in is Berczy Park. In fact, based on Google map, Toronto's Chinatown is couple streets off the center of Berzy Park.

Example #2: Jody lives in Soho (the 11th row from the bottom up on the Y-axis), a place with many high-end shopping places. The best neighborhood in Toronto to live in is Garden District, Ryerson, which is a place surrounded by shopping complexes.

Some limitation of this project includes limited to only two city comparison, restricted to only 10 most common venues for each neighborhood, and the questions of arbitrary score assigning to the 10 most common venues.

Conclusion

As more and more unrests happen in the United States, more and more people will want to move to neighboring country, Canada. This project is a good starting point for people to investigate where in Canada is best suitable for Americans.

Reference

[1] https://trends.google.com/trends/explore?geo=US&q=move%20to%20canada

[2] https://coronavirus.jhu.edu/us-map