Harlem to Toronto – Where to go?

Prieshecumar Premegi

6 May 2020

Contents

Introduction: Business Problem

Manhattan is the smallest and most densely populated borough of New York City. With 72,033 people per square mile in 2015, this borough has density higher than any individual American city (1).

I currently live in New York and work for a big IT Consultancy company. I've been living in Manhattan for the last 5 year and have moved a few times until I found Harlem. I love everything about this neighbourhood, from restaurants to supermarkets and gym facilities, this is home for me.

My consultancy company won a big project in Old Town Toronto. I have been lucky to be assigned to this project, which is going to run for at least 24 months. I was asked to choose a place to rent in Old Town Toronto. Rent price is not a problem for the company and therefore I want to move to a neighbourhood similar to Harlem.

Both Manhattan and Old Toronto are boroughs of main cities in US and Canada. It's very difficult to compare these two neighbours due to its difference in demographics but also due to the way these boroughs have split their neighbourhoods.

The Old Town Toronto has only 64 neighbourhoods, less than half of Manhattan (2)(3). It's population is 3,169 people per square mile (4), just 5% of Manhattan!

Data acquisition and cleaning

Comparing Neighbourhoods is not an easy task, so comparing Neighbourhoods of two different cities is even more complicate. **How can I compare Harlem neighbourhood to Old Town Toronto neighbourhoods and find the most similar place to live in?**

This problem affects many people the globe who need to move their locations to a new city due to new job opportunities. In this case the origin is Harlem, but it can be changes to any other place.

Data sources and analysis approach:

One way of comparing and segmenting neighbourhoods is to use Foursquare data to rank they types of venues in each neighbourhood. Then, I can segment these venues to identify the neighbourhoods in Toronto that are in the same segment as Harlem.

1) List of neighbourhoods in Old Town Toronto

- I need the list of all Old Town Toronto neighbourhoods, along with its latitude and longitude so that I can use Foursquare to obtain information about the surrounding venue categories and frequency.
- * My data source for this exercise will be the list of Toronto boroughs and neighbourhoods https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M along with the latitude and Longitude of each postcode in Toronto https://cocl.us/Geospatial_data. I will have to scrap this webpage and load only the Old Town Toronto neighbourhoods in a dataframe.

2) Harlem geolocation

- I need the latitude and longitude of Harlem so that I can use Foursquare to obtain information about the surrounding venue categories and frequency.
- My data source is https://www.gps-latitude-longitude.com/gps-coordinates-of-harlem and the latitude is for this is 40.8115504 and longitude is -73.9464769.

3) Get venues in each Neighbourhood using Foursquare

- I will run API calls to obtain the list of all venues in a radius of 750 meter of each neighbourhood geolocation.
- I will clean the data and create a dataframe with the frequency of the top 10 venue categories per neighbourhood.

4) Segment the neighbourhoods to find Harlem like neighbourhoods

- I will use K-means segmentation to find the best number of clusters and group the neighbourhoods.
- Once segmented, I will identify which cluster has Harlem. The neighboughoods in this cluster will be my short-list of candidates for Toronto.

Data cleaning and exploration

The Toronto neighbourhood data was extracted from a webpage and it required some data cleaning. For example, each cell had a new line character in the data, which was bringing issues to data analysis. I also had to filter out all the boroughs that were not in Old Town Toronto.

	Borough	Neighbourhood	Latitude	Longitude
0	Central Toronto	Davisville	43.704324	-79.388790
1	Central Toronto	Davisville North	43.712751	-79.390197
2	Central Toronto	Forest Hill North & West	43.696948	-79.411307
3	Central Toronto	Lawrence Park	43.728020	-79.388790
4	Central Toronto	Moore Park / Summerhill East	43.689574	-79.383160

Figure 1 - Subset of Toronto boroughs

The geolocation data was in a separate dataset, so I had to ensure that I merge the neighbourhood and geolocation data, using the same key – postal code.

Foursquare data provided me the list of 100 venues in each borough, along with its category name and its geolocation.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Davisville	43.7043	-79.3888	Jules Cafe Patisserie	43.704138	-79.388413	Dessert Shop
1	Davisville	43.7043	-79.3888	Thobors Boulangerie Patisserie Café	43.704514	-79.388616	Café
2	Davisville	43.7043	-79.3888	Marigold Indian Bistro	43.702881	-79.388008	Indian Restaurant
3	Davisville	43.7043	-79.3888	XO Gelato	43.705177	-79.388793	Dessert Shop
4	Davisville	43.7043	-79.3888	Viva Napoli	43.705752	-79.389125	Pizza Place

Figure 2 - Foursquare result (fisrt 5 rows)

Sing one hot analysis and I created a dataframe which represented the number of venue categories and frequency (mean) of each venue category per neighbourhood.

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	100	100	100	100	100	100
Brockton / Parkdale Village / Exhibition Place	84	84	84	84	84	84
Business reply mail Processing Centre	51	51	51	51	51	51
Bathurst Quay / South Niagara / Island airport	26	26	26	26	26	26
Central Bay Street	100	100	100	100	100	100
Christie	34	34	34	34	34	34
Church and Wellesley	100	100	100	100	100	100
Commerce Court / Victoria Hotel	100	100	100	100	100	100

Figure 3 - Number of venue categories per neighboughood (snapshot)

There were 290 venue categories in the dataset as you can see in the image below.

	Neighbourhood	ATM	Accessories Store	African Restaurant	Airport	Airport Food Court	Airport Gate
0	Berczy Park	0.0	0.000000	0.0	0.000000	0.000000	0.000000
1	Brockton / Parkdale Village / Exhibition Place	0.0	0.011905	0.0	0.000000	0.000000	0.000000
2	Business reply mail Processing Centre	0.0	0.000000	0.0	0.000000	0.000000	0.000000
3	CN Tower / King and Spadina / Railway Lands / ...	0.0	0.000000	0.0	0.038462	0.038462	0.038462
4	Central Bay Street	0.0	0.000000	0.0	0.000000	0.000000	0.000000

5 rows × 290 columns

Figure 4 - Frequency of venue categories per Neighbourhood (snapshot)

A list of top 10 categories per neighbourhood was created to prepare for k-means analysis.

	Neighbourhood	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	Hotel	Café	Japanese Restaurant	Cocktail Bar	Beer Bar	Restaurant	Bakery	Gastropub	Seafood Restaurant
1	Brockton / Parkdale Village / Exhibition Place	Café	Coffee Shop	Restaurant	Gift Shop	Bar	Bakery	Furniture / Home Store	Thrift / Vintage Store	Supermarket	Music Venue
2	Business reply mail Processing Centre	Fast Food Restaurant	Park	Restaurant	Italian Restaurant	Light Rail Station	Bar	Brewery	Burrito Place	Bakery	Clothing Store
3	CN Tower / King and Spadina / Railway Lands / ...	Rental Car Location	Harbor / Marina	Boat or Ferry	Coffee Shop	Sculpture Garden	Airport Lounge	Airport Service	Boutique	Pier	Park
4	Central Bay Street	Coffee Shop	Café	Clothing Store	Japanese Restaurant	Art Gallery	Italian Restaurant	Ramen Restaurant	Arts & Crafts Store	Diner	Creperie

Figure 5 - Top 10 venue categories per neighboughood (snapshot)

Methodology

In this project, I am focusing in listing and ranking the frequency of venue categories in a radius of 750 meters of each Old Town Toronto neighbourhood. Then, based on the top 10 venue categories, I will cluster and segment them to find the which one is similar to Harlem.

I've decided to use K-means clustering to find the solution for this problem. If I can find the top 10 most common venues for each neighboughood, then I can cluster the data, including Harlem and find out which boroughs are in the same cluster.

	Borough	Neighbourhood	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	10th Most Common Venue
5	Central Toronto	North Toronto West	43.7154	-79.4057	0	Coffee Shop	Sporting Goods Shop	Café	Clothing Store	Grocery Store	Skating Rink	Italian Restaurant	Restaurant	Diner	Dessert Shop
9	Downtown Toronto	Berczy Park	43.6448	-79.3733	0	Coffee Shop	Hotel	Café	Japanese Restaurant	Cocktail Bar	Beer Bar	Restaurant	Bakery	Gastropub	Seafood Restaurant
11	Downtown Toronto	Central Bay Street	43.658	-79.3874	0	Coffee Shop	Café	Clothing Store	Japanese Restaurant	Art Gallery	Italian Restaurant	Ramen Restaurant	Arts & Crafts Store	Diner	Creperie
13	Downtown Toronto	Church and Wellesley	43.6659	-79.3832	0	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Sandwich Place	Restaurant	Men's Store	Diner	Pizza Place	Bookstore	Smoke Shop
14	Downtown Toronto	Commerce Court / Victoria Hotel	43.6482	-79.3798	0	Coffee Shop	Café	Hotel	American Restaurant	Japanese Restaurant	Seafood Restaurant	Restaurant	Concert Hall	Gym	Vegetarian / Vegan Restaurant
15	Downtown Toronto	First Canadian Place / Underground city	43.6484	-79.3823	0	Hotel	Coffee Shop	Café	Japanese Restaurant	Theater	Restaurant	Concert Hall	Seafood Restaurant	Deli / Bodega	Bookstore
16	Downtown Toronto	Garden District / Ryerson	43.6572	-79.3789	0	Coffee Shop	Hotel	Gastropub	Falafel Restaurant	Café	Ramen Restaurant	Japanese Restaurant	Sushi Restaurant	Burger Joint	Park
17	Downtown Toronto	Harbourfront East / Union Station / Toronto Is...	43.6408	-79.3818	0	Coffee Shop	Hotel	Boat or Ferry	Brewery	Park	Japanese Restaurant	Bar	Pizza Place	Café	Plaza
18	Downtown Toronto	Kensington Market / Chinatown / Grange Park	43.6532	-79.4	0	Café	Bar	Coffee Shop	Vegetarian / Vegan Restaurant	Mexican Restaurant	Bakery	Dessert Shop	Art Gallery	Ice Cream Shop	Record Shop
19	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.6623	-79.3895	0	Coffee Shop	Italian Restaurant	Café	Sandwich Place	Sushi Restaurant	Park	Japanese Restaurant	Gastropub	Diner	Ice Cream Shop
20	Downtown Toronto	Regent Park / Harbourfront	43.6543	-79.3606	0	Coffee Shop	Italian Restaurant	Park	Theater	Pub	Bakery	Café	Restaurant	Breakfast Spot	Thai Restaurant
21	Downtown Toronto	Richmond / Adelaide / King	43.6506	-79.3846	0	Café	Hotel	Coffee Shop	Theater	Clothing Store	Restaurant	Tea Room	Plaza	Japanese Restaurant	Cosmetics Shop
23	Downtown Toronto	St. James Town	43.6515	-79.3754	0	Coffee Shop	Café	Seafood Restaurant	Gastropub	American Restaurant	Italian Restaurant	Bakery	Theater	Hotel	Restaurant
25	Downtown Toronto	Stn A PO Boxes	43.6464	-79.3748	0	Coffee Shop	Café	Japanese Restaurant	Hotel	Restaurant	Gym	Cocktail Bar	Beer Bar	Bakery	Italian Restaurant
26	Downtown Toronto	Toronto Dominion Centre / Design Exchange	43.6472	-79.3816	0	Hotel	Coffee Shop	Café	American Restaurant	Concert Hall	Japanese Restaurant	Seafood Restaurant	Gym	Restaurant	Park
33	West Toronto	Brockton / Parkdale Village / Exhibition Place	43.6368	-79.4282	0	Café	Coffee Shop	Restaurant	Gift Shop	Bar	Bakery	Furniture / Home Store	Thrift / Vintage Store	Supermarket	Music Venue
36	West Toronto	Little Portugal / Trinity	43.6479	-79.4197	0	Restaurant	Café	Bar	Vegetarian / Vegan Restaurant	Cocktail Bar	Italian Restaurant	Coffee Shop	Asian Restaurant	Wine Bar	Nightclub
39	Manhattan	Harlem	40.8115504	-73.9464769	0	French Restaurant	Cosmetics Shop	Southern / Soul Food Restaurant	Jazz Club	Cocktail Bar	Burger Joint	Café	American Restaurant	Theater	Pizza Place

Figure 6 - K-means cluster result for Harlem like neighbourhoods

As you can see, we have found a list of potential matching neighboughoods that have venues like Harlem. Let's see them in a map.

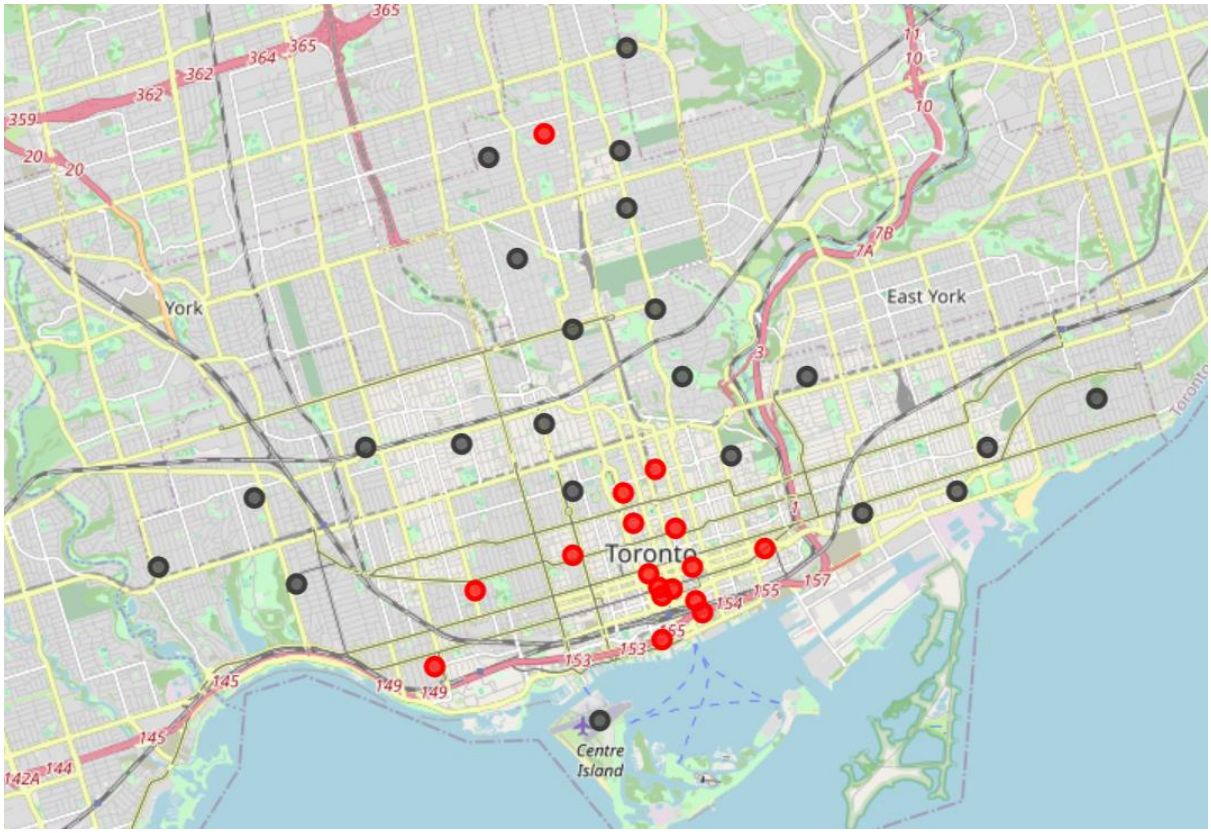

Figure 7 - Harlem like neighbourhoods in Toronto

In the image above, you can see the Harlem like neighbourhoods marked in red. These are mainly concentrated in the downtown area of Toronto with minor exceptions.

Results and Discussion

The analysis shows that the borough of Downtown Toronto is by far the borough with more neighbourhoods classified and clustered as like Harlem. There some neighbourhoods in West Toronto and Central Toronto that share the same profile. You can see the most common top 10 venue categories in each neighbourhood that are part of Harlem's cluster as well as a map with the location of each borough (in red).

I can also observe that Harlem's top 4 common venue categories are not in any of the suggested boroughs. You can only see a match from the 5th venue category, "Cocktail Bar", matching the boroughs of Little Portugal, Trinity and Berczy Park.

In fact, there is no borough in Toronto that has "French restaurant" in the first place of most common venue category. This ultimately suggests that the experience in Toronto is going to be perhaps different from Harlem since there is not a strong correlation of common venues in Toronto.

Conclusion

This project was create to help me identify neighbourhoods in Old Town Toronto that are similar Harlem (Manhattan) due to a recent change in career that is forcing me to leave the place I love. To help me fid the best places in Toronto, that I will enjoy as I do in Harlem, I have used data from

Foursquare to list the most common venues around 750 meters of each neighbourhood in Toronto, to find those which match with Harlem's most common venues.

Downtown Toronto is the place to start for, as it has the most neighbourhoods in the same cluster as Harlem.

References

(1) New York City Neighbourhoods - https://en.wikipedia.org/wiki/New_York_City#Geography

(2) Manhattan demographics - https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods

(3) Toronto Neighbourhoods - https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto

(4) Toronto demographics
- https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods