Harlem to Toronto – Where to go?

Prieshecumar Premegi

6 May 2020

Contents

[Introduction: Business Problem 2](#_Toc39689281)

[Data 2](#_Toc39689282)

[Methodology 2](#_Toc39689283)

[Results and Discussion 2](#_Toc39689284)

[Conclusion 2](#_Toc39689285)

[References 2](#_Toc39689286)

# 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.



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.

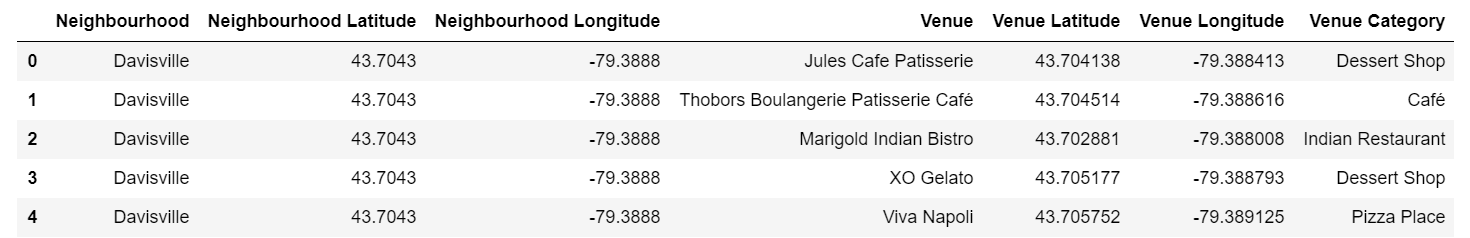


Figure 2 - Foursquare result (fisrt 5 rows)

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.

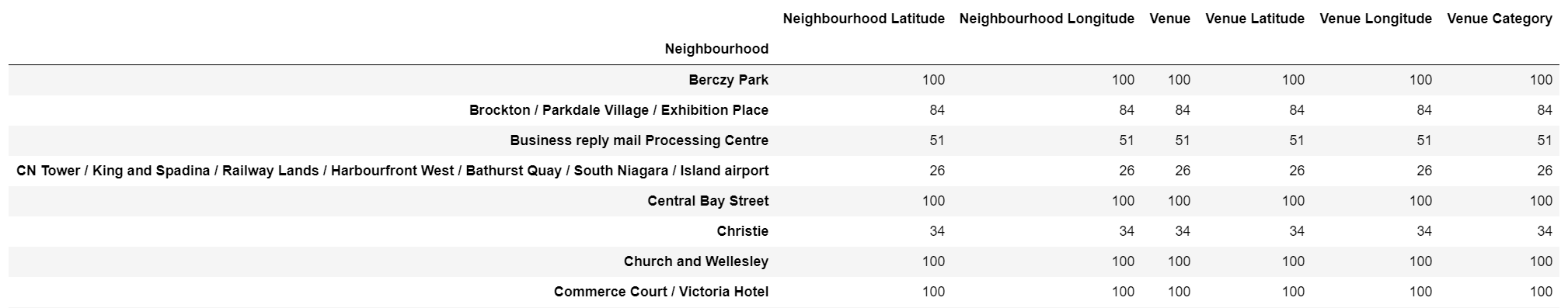


Figure 3 - Number of venue categories per neighboughood (snapshot)

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

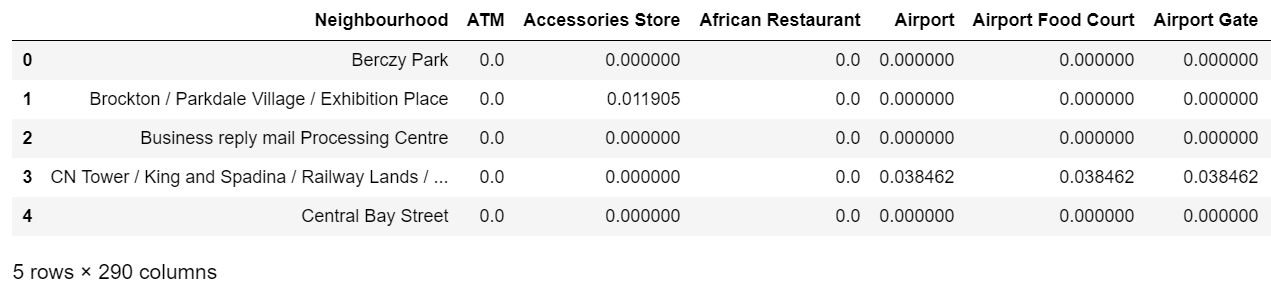


Figure 4 - Frequency of venue categories per Neighbourhood (snapshot)

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



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.



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.

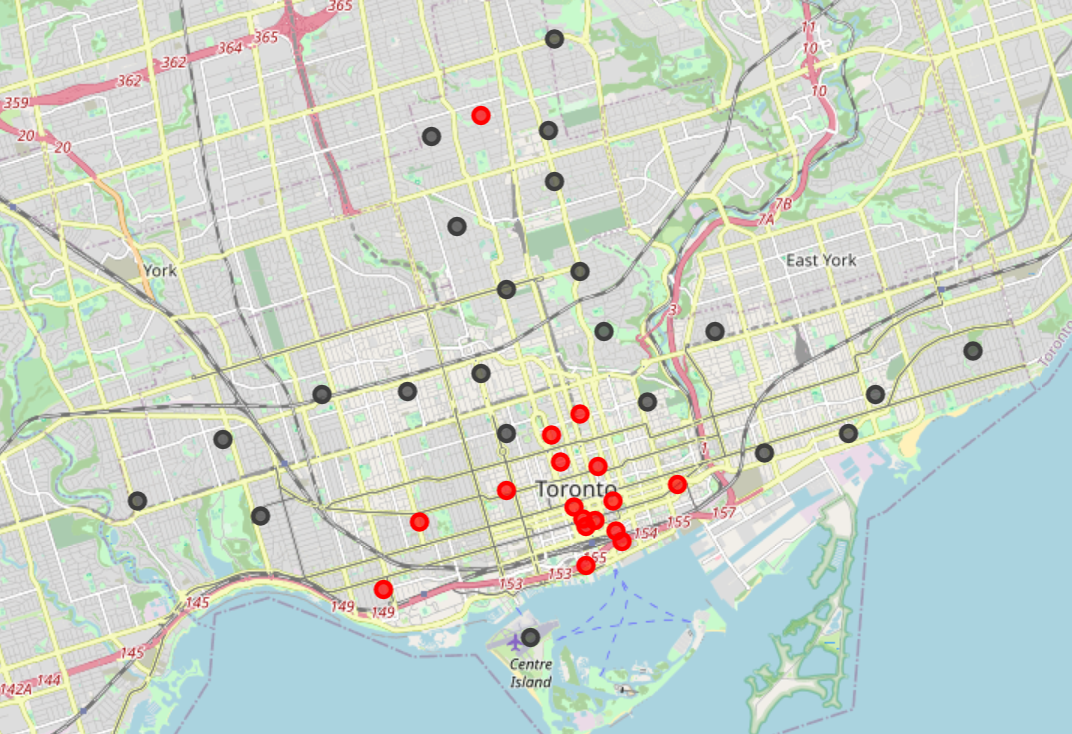


Figure 7 - Harlem like neighbourhoods in Toronto

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>