Battle of Neighborhoods

1. Introduction

1.1. Background

New York City (NYC) has been described as the cultural, financial, and media capital of the world, significantly influencing commerce,[1] entertainment, research, technology, education, politics, tourism, art, fashion, and sports. Manhattan, often referred to by residents of the New York City area as the City, is the most densely populated of the five boroughs of NYC and has long been the flag bearer for people worldwide coming to the US to live the American Dream [2]. However, with the growing immigration queues and visa issues in the US, a large exodus of technical and financial workers is moving to Canada [3,4], where the immigration rules have been designed to welcome skilled workers [5]. A lot of people who are used to the unique urban Manhattan lifestyle look towards Toronto, an international center of business, finance, arts, and culture, that is recognized as one of the most multicultural and cosmopolitan cities in the world [6]. A couple of my friends have taken job offers and are moving to Toronto from Manhattan and have asked me to help them locate neighborhoods similar to the ones in Manhattan where they currently live. Toronto is a very vibrant city with a lot of neighborhoods, each with unique character. Some neighborhoods have a convenient access to parks and eateries, while others offer a lot of fun and nightlife activities. Choosing a neighborhood to move to in a new city (let alone country) can be complicated, but with the help of location data from Foursquare, the task can be made a little bit easier.

1.2. Business Problem

A couple of my friends are moving to Toronto from Manhattan and have asked me to use Data Science methodology to help them locate neighborhoods similar to the ones in Manhattan where they currently live. The objective of this capstone project is to use data science methodology to analyze and list the neighborhoods in the city of Toronto that are similar to Manhattan to live in.

1.3. Target Audience

People like my friends who are interested in moving to Toronto from Manhattan and looking for a similar neighborhood to what they are used to.

2. Data

2.1. Data Sources

For this project, the following data was used:
• List of neighborhoods in Toronto and Manhattan
• Latitude and longitude coordinates of neighborhoods to get the venue data
• Venues Details
First, BeautifulSoup was used to extract a full list of all Toronto neighborhoods from the Wikipedia webpage "List of postal codes of Canada" [7]. All boroughs with 'Toronto' in their names were parsed to obtain and extract this data. The Manhattan data was extracted from New York City neighborhood data by selecting the rows with 'Manhattan' as the borough [8].
Next, the geopy library was used to obtain the latitude and longitude values for each neighborhood as seen below:

	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
5	Manhattan	Manhattanville	40.816934	-73.957385
6	Manhattan	Central Harlem	40.815976	-73.943211
7	Manhattan	East Harlem	40.792249	-73.944182
8	Manhattan	Upper East Side	40.775639	-73.960508
9	Manhattan	Yorkville	40.775930	-73.947118
10	Manhattan	Lenox Hill	40.768113	-73.958860
11	Manhattan	Roosevelt Island	40.762160	-73.949168
12	Manhattan	Upper West Side	40.787658	-73.977059
13	Manhattan	Lincoln Square	40.773529	-73.985338
14	Manhattan	Clinton	40.759101	-73.996119
15	Manhattan	Midtown	40.754691	-73.981669

Figure 1. Example of New York data with latitudes and longitudes.

Borough	Neighborhood	Latitude	Longitude
Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
Downtown Toronto	St. James Town	43.651494	-79.375418
East Toronto	The Beaches	43.676357	-79.293031
Downtown Toronto	Berczy Park	43.644771	-79.373306
Downtown Toronto	Central Bay Street	43.657952	-79.387383
Downtown Toronto	Christie	43.669542	-79.422564
Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568
West Toronto	Dufferin, Dovercourt Village	43.669005	-79.442259

Figure 2. Example of Toronto data with latitudes and longitudes.

Also, the Foursquare API was used to explore the neighboring venues in Manhattan and Toronto. The explore function was used to get the most common venue categories such as restaurants, gyms etc. in each neighborhood.

2.2. Data cleaning and consolidation

For Toronto data, locations with an unassigned Borough were neglected. The postal code column was also dropped. Manhattan had a total of 40 neighborhoods and Toronto had 39 neighborhoods. Also, the unique venue categories were consolidated into general categories (e.g. different types of restaurants will be combined and listed under 'Food'). The most common general categories for each neighborhood was then used for clustering.

3. Methodology

3.1. Location Data

The geopy library was used to obtain the coordinates of Manhattan and Toronto. These were used to plot maps of both the locations demonstrating their neighborhoods using Folium (**Figures 3** and **4**).

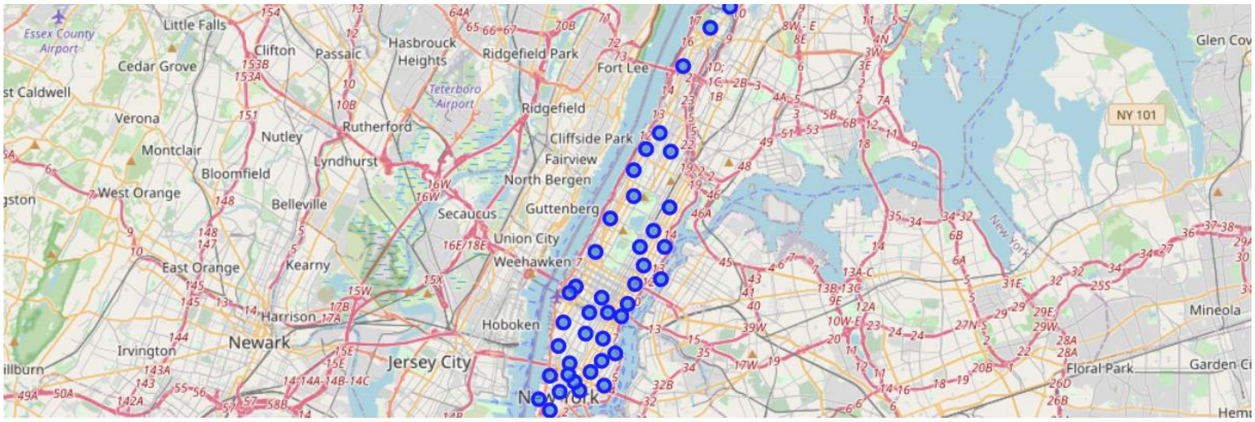

Figure 3. Map of Manhattan Neighborhoods

Figure 4. Map of Toronto Neighborhoods

Once, the coordinates of each neighborhood for both Manhattan and Toronto were obtained, they were saved in *pandas* dataframes. The dataframes were then concatenated to obtain a single dataframe of data that contained all the neighborhoods from both the primary locations. The idea behind this was that when we later perform clustering, the neighborhoods from both Manhattan and Toronto with similar features will be paired together.

3.2. Explore nearby venues for both Manhattan and Toronto using Foursquare API

Foursquare API was used to explore up to 100 nearby venues within a 500m radius of each neighborhood. The category and geographical coordinates for each venue was obtained. This data was then loaded into a *pandas* dataframe. A total of 4827 venues were found and then classified into 9 general categories:

1. Shop & Service
2. Outdoors & Recreation
3. Travel & Transport
4. Food
5. Nightlife Spot
6. Arts & Entertainment
7. Residence
8. College and University
9. Professional and Other places

Given the relatively small number of neighborhoods, this generalization of venue categories was performed to emphasize on the type of venues in a given neighborhood. This was thought to be a more reasonable and focused approach rather than spreading the data over 368 different categories. It could also help to avoid the curse of dimensionality [9].

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	General Venue Category
Marble Hill	40.876551	-73.910660	Arturo's	40.874412	-73.910271	Pizza Place	Food
Marble Hill	40.876551	-73.910660	Bikram Yoga	40.876844	-73.906204	Yoga Studio	Outdoors & Recreation
Marble Hill	40.876551	-73.910660	Tibbett Diner	40.880404	-73.908937	Diner	Food
Marble Hill	40.876551	-73.910660	Dunkin'	40.877136	-73.906666	Donut Shop	Food
Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop	Food
Marble Hill	40.876551	-73.910660	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym	Outdoors & Recreation
Marble Hill	40.876551	-73.910660	Starbucks	40.873755	-73.908613	Coffee Shop	Food
Marble Hill	40.876551	-73.910660	Rite Aid	40.875467	-73.908906	Pharmacy	Shop & Service
Marble Hill	40.876551	-73.910660	Blink Fitness	40.877271	-73.905595	Gym	Outdoors & Recreation
Marble Hill	40.876551	-73.910660	T.J. Maxx	40.877232	-73.905042	Department Store	Shop & Service
Marble Hill	40.876551	-73.910660	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant	Food

Figure 5. Example of venues and their categories and general categories

The mean of the frequency of occurrence of each category for each neighborhood was computed and top 5 general categories for each neighborhood was chosen as the feature set for clustering.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Food	Outdoors & Recreation	Shop & Service	Travel & Transport	Professional & Other Places
1	Berczy Park	Food	Shop & Service	Nightlife Spot	Arts & Entertainment	Outdoors & Recreation
2	Brockton, Parkdale Village, Exhibition Place	Food	Shop & Service	Nightlife Spot	Outdoors & Recreation	Arts & Entertainment
3	Business reply mail Processing Centre, South C...	Shop & Service	Outdoors & Recreation	Food	Travel & Transport	Nightlife Spot
4	CN Tower, King and Spadina, Railway Lands, Har...	Travel & Transport	Outdoors & Recreation	Shop & Service	Nightlife Spot	Food

Figure 6. Example of 5 most commonly occurring general categories for each neighborhood

3.3. Clustering

The k-means clustering algorithm was used to compute clusters of similar neighborhoods by using the five most commonly occurring general category for each neighborhood. For simplicity, 4 clusters were chosen. The initial condition was set for 25 trials i.e. the clustering algorithm was run 25 times and the value that occurred most commonly for each neighborhood was assigned to it. The cluster labels were appended to the dataframe consisting of neighborhood, borough, latitude, longitude and the most commonly occurring 5 general categories. These labels were used to plot the neighborhoods on the maps of Manhattan and Toronto for visual inspection.

4. Results

The cluster labels assigned by the clustering algorithm were used to plot and analyze similar Manhattan and Toronto neighborhoods (**Figures 7** and **8**).

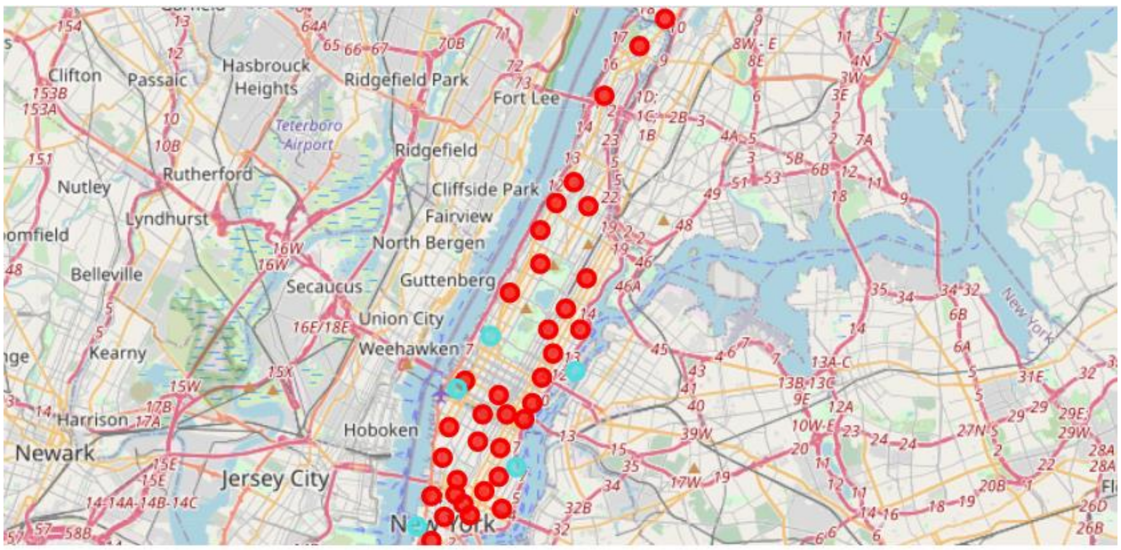

Figure 7. Map of Manhattan showing the neighborhoods assigned to different clusters in different colors. As can be seen, the Manhattan neighborhoods are divided in 2 clusters.

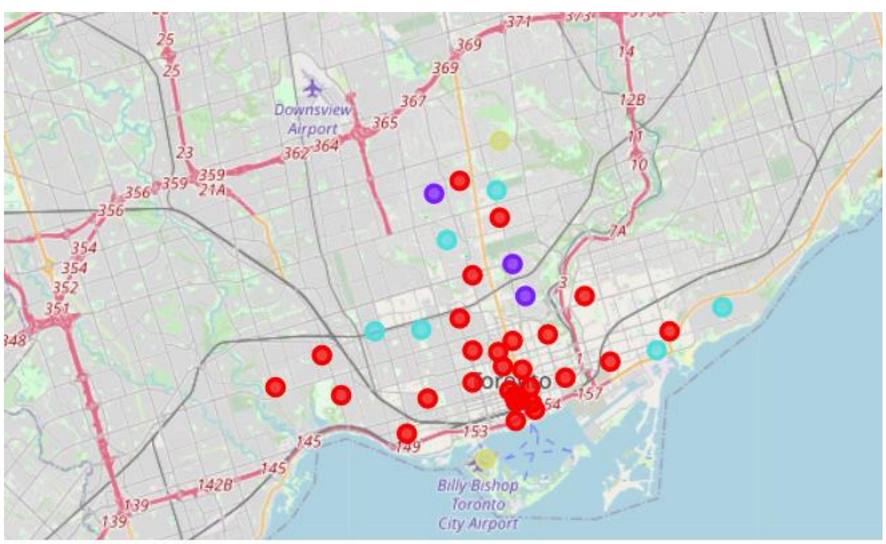

Figure 8. Map of Toronto showing the neighborhoods assigned to different clusters in different colors. As can be seen, the Toronto neighborhoods are divided in 4 clusters.

Each cluster was examined to determine the discriminating general categories and observe the neighborhoods from Manhattan and Toronto that are similar.
Cluster 1 had a total of 64 neighborhoods (36 from Manhattan and 28 from Toronto). The most commonly occurring categories showed Food, Nightlife Spots and, Shop and Service venues to be most prevalent in these neighborhoods (**Figure 9**).

Borough	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Manhattan	Marble Hill	Food	Shop & Service	Outdoors & Recreation	Arts & Entertainment	Travel & Transport
Manhattan	Chinatown	Food	Shop & Service	Nightlife Spot	Arts & Entertainment	Outdoors & Recreation
Manhattan	Washington Heights	Food	Shop & Service	Outdoors & Recreation	Nightlife Spot	Travel & Transport
Manhattan	Inwood	Food	Shop & Service	Nightlife Spot	Outdoors & Recreation	Travel & Transport
Manhattan	Hamilton Heights	Food	Nightlife Spot	Shop & Service	Outdoors & Recreation	Professional & Other Places
Manhattan	Manhattanville	Food	Shop & Service	Outdoors & Recreation	Nightlife Spot	Travel & Transport
Manhattan	Central Harlem	Food	Shop & Service	Outdoors & Recreation	Nightlife Spot	Arts & Entertainment
Manhattan	East Harlem	Food	Shop & Service	Arts & Entertainment	Outdoors & Recreation	Nightlife Spot
Manhattan	Upper East Side	Food	Shop & Service	Outdoors & Recreation	Arts & Entertainment	Travel & Transport
Manhattan	Yorkville	Food	Shop & Service	Outdoors & Recreation	Nightlife Spot	Professional & Other Places
Manhattan	Lenox Hill	Food	Shop & Service	Outdoors & Recreation	Nightlife Spot	Professional & Other Places
Manhattan	Upper West Side	Food	Nightlife Spot	Shop & Service	Outdoors & Recreation	Arts & Entertainment
Manhattan	Clinton	Food	Outdoors & Recreation	Arts & Entertainment	Shop & Service	Nightlife Spot
Manhattan	Midtown	Food	Shop & Service	Outdoors & Recreation	Travel & Transport	Arts & Entertainment
Manhattan	Murray Hill	Food	Shop & Service	Outdoors & Recreation	Travel & Transport	Nightlife Spot
Manhattan	Chelsea	Food	Shop & Service	Outdoors & Recreation	Arts & Entertainment	Nightlife Spot
Manhattan	Greenwich Village	Food	Shop & Service	Outdoors & Recreation	Arts & Entertainment	Nightlife Spot
Manhattan	East Village	Food	Nightlife Spot	Shop & Service	Outdoors & Recreation	Arts & Entertainment
Manhattan	Lower East Side	Food	Shop & Service	Arts & Entertainment	Outdoors & Recreation	Nightlife Spot

Figure 9. Example of neighborhoods assigned to Cluster 1

No Manhattan neighborhoods were assigned to clusters 2 and 4. The second cluster had a total of 3 neighborhoods from Toronto. Outdoor recreation and Travel and Transport appeared to be the most prevalent general categories in this cluster (**Figure 10**).

Borough	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Central Toronto	Roselawn	Outdoors & Recreation	Arts & Entertainment	Travel & Transport	Shop & Service	Residence
Central Toronto	Moore Park, Summerhill East	Outdoors & Recreation	Travel & Transport	Shop & Service	Residence	Professional & Other Places
Downtown Toronto	Rosedale	Outdoors & Recreation	Travel & Transport	Shop & Service	Residence	Professional & Other Places

Figure 10. Neighborhoods assigned to Cluster 2

Five Manhattan and six Toronto neighborhoods were assigned to Cluster 3. Outdoor recreations and Food along with Shop and Service and Residential venues appeared to be the key features for this cluster (**Figure 11**).

Borough	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Manhattan	Roosevelt Island	Outdoors & Recreation	Food	Shop & Service	Residence	Professional & Other Places
Manhattan	Lincoln Square	Food	Outdoors & Recreation	Arts & Entertainment	Shop & Service	College & University
Manhattan	Battery Park City	Food	Outdoors & Recreation	Shop & Service	Travel & Transport	Professional & Other Places
Manhattan	Stuyvesant Town	Outdoors & Recreation	Shop & Service	Travel & Transport	Nightlife Spot	Food
Manhattan	Hudson Yards	Food	Outdoors & Recreation	Shop & Service	Travel & Transport	Nightlife Spot
East Toronto	The Beaches	Outdoors & Recreation	Shop & Service	Nightlife Spot	Travel & Transport	Residence
Downtown Toronto	Christie	Shop & Service	Food	Outdoors & Recreation	Nightlife Spot	Travel & Transport
West Toronto	Dufferin, Dovercourt Village	Shop & Service	Food	Nightlife Spot	Outdoors & Recreation	Arts & Entertainment
Central Toronto	Davisville North	Outdoors & Recreation	Shop & Service	Food	Travel & Transport	Arts & Entertainment
Central Toronto	Forest Hill North & West, Forest Hill Road Park	Outdoors & Recreation	Shop & Service	Food	Travel & Transport	Residence
East Toronto	Business reply mail Processing Centre, South C...	Shop & Service	Outdoors & Recreation	Food	Travel & Transport	Nightlife Spot

Figure 11. Neighborhoods assigned to Cluster 3

Two Toronto neighborhoods were assigned to the fourth cluster. Travel and transport, outdoor recreation and Shop and Service seemed to be the most prevalent common neighborhood venue categories for these neighborhoods. Residential and Professional venues seemed to be a distinctive feature as well (**Figure 12**).

Borough	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Central Toronto	Lawrence Park	Travel & Transport	Professional & Other Places	Outdoors & Recreation	Shop & Service	Residence
Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	Travel & Transport	Outdoors & Recreation	Shop & Service	Nightlife Spot	Food

Figure 12. Neighborhoods assigned to Cluster 4

5. Discussion

The clustering results show that neighborhoods similar to Manhattan in terms of their neighboring venues exist in Toronto. Food and, shop and service seem to be the most prevalent neighboring venues that are most similar between the neighborhoods of the two locations. It also depicted that some neighborhoods in Toronto offer a slightly different living experience in terms of their proximity to famous venues from Travel and Transport, and residential categories.

Future Studies

While the analysis in this project was a crude comparison of the neighborhoods of two world-famous locations, it can be used as a basis for an in-depth analysis to help in various other decisions. For example, if someone wants to open a business of a particular category (say restaurant) then one can just focus on venues with restaurants and their types to do a much-detailed analysis. Other factors to add in for someone looking to move or open a business would be income and population data, crime data and so on.

6. Conclusion

With the help of Data Science and Machine learning, we were able to compare neighborhoods from Manhattan and Toronto and identify similarities between them in terms of the neighboring venues. This analysis can be used by people looking to migrate from Manhattan to Toronto (and vice versa) in terms of identifying similar or new neighborhoods based on their preferences of Food, Shop and Service, Outdoor Recreation and/or Travel and Transport venues.

7. References

[1] Barry, Dan. "A Nation challenged: in New York; New York Carries On, but Test of Its Grit Has Just Begun" Archived March 24, 2020, at the Wayback Machine, The New York Times, October 11, 2001. Accessed November 20, 2016. "A roaring void has been created in the financial center of the world."
[2] https://www.encyclopedia.com/education/educational-magazines/immigrant-american-dream
[3] https://www.vox.com/2019/3/19/18264391/us-tech-jobs-canada-immigration-policies-trump
[4] https://www.ctvnews.ca/world/jobs-will-move-to-canada-if-u-s-immigration-policies-don-t-change-duolingo-ceo-says-1.5015618

[5] https://www.cicnews.com/2020/03/canada-to-welcome-up-to-390000-immigrants-in-2022-0313872.html#gs.mdnbvb
[6] Robert Vipond (April 24, 2017). Making a Global City: How One Toronto School Embraced Diversity. University of Toronto Press. p. 147. ISBN 978-1-4426-2443-6.
[7] https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
[8] https://geo.nyu.edu/catalog/nyu_2451_34572
[9] https://en.wikipedia.org/wiki/Curse_of_dimensionality