

Final Report - KMeans Clustering to compare Manhattan and Toronto

by Lynden McIntosh

Introduction

Background

Toronto is the most popular and most populated city in Canada, located in the Canadian province of Ontario. It has a population of 2,731,571, which is 1,026,877, approximately 60% higher than Montreal (the Canadian city with the second-highest population). On the other hand, with a population of approximately 8,175,133 people, New York City is the most populated city in the USA. There are five boroughs of New York: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Of the five boroughs, Manhattan is the most popular of the 5 boroughs and serves as the financial capital of the USA. Each year, hundreds of thousands of people migrate to Canada yearly, and the United States of America (USA) is one of the top 5 origins of these immigrants¹. Most of these immigrants settle in the province of Ontario.² The Ontarian city in which most immigrants settle in Toronto. Immigrants migrating from the USA and settling in Toronto may be doing so for a variety of reasons such as work, schooling, and family reunification.

Problem

Many immigrants in Canada are looking for neighborhoods similar to their old Neighbourhood when looking for a home. For example, a family with kids migrating from one neighbourhood to another may want to ensure there are similarly good parks and playgrounds in the new Neighbourhood. Also, a young person migrating for a job opportunity may want to enjoy bars and clubs in the new city as they did in the old city. It would be helpful if we can identify neighborhoods in Toronto that are like an immigrant's Neighbourhood of origin. This information would assist real estate agents in targeting immigrants moving from a given region. Ultimately, real estate agents would be better able to recommend them homes for rent/sale in neighborhoods that suit them based on their previous Neighbourhood. In this report, we consider the Neighbourhoods of Manhattan to Toronto.

To make this determination, we may use locational data related to venues found in each neighbourhood. The datasets should include the Neighbourhoods of each city along with their respective latitudinal and longitudinal coordinates. The goal is to assess and determine neighborhoods located in Toronto that are like those found in Manhattan based on the available venues within the areas by modeling clusters of neighbourhoods to identify similarities.

¹ <https://www.canada.ca/en/immigration-refugees-citizenship/corporate/publications-manuals/annual-report-parliament-immigration-2018/permanent-residents-admitted.html>

² <https://www.canada.ca/en/immigration-refugees-citizenship/corporate/publications-manuals/annual-report-parliament-immigration-2018/report.html>

Interest

This information would be of interest to real estate agents looking to market the sale of homes to individuals settling into Toronto after migrating from Manhattan. Knowing neighbourhoods in Toronto that are similar to specific neighbourhoods in Manhattan can help real estate agents better serve their clients (in this case, the immigrant).

Also note that, although we analyze Manhattan and Toronto data in this report, the analysis can also be replicated in other cities.

Data Sourcing and Processing

Data sources

We source the data of this report from multiple locations. Locational data of Toronto boroughs and cities were scraped from Wikipedia³; however, the coordinate data needed for the analysis is not included. Coursera provided this coordinate data. Locational data of New York boroughs and cities are available on the New York University website.⁴ This data is then used to retrieve venue data from Foursquare API. The data received that is received from Foursquare is a list of all recorded venues in each respective Neighbourhood of Toronto and Manhattan⁵.

Data cleaning

As mentioned above, the data used was downloaded from a combination of websites. This data was then combined into two dataframes for Toronto neighbourhoods and New York neighbourhoods. In both dataframes, the fields are Neighbourhood, Borough, Latitude, and Longitude. The following changes were made in the Toronto dataframe:

- Their zip code grouped neighbourhoods. If two records have the same zip code, those records were combined.
- "Not Assigned" Neighbourhoods were dropped from the dataset.
- Records with "Not Assigned" Boroughs were given the name of their Neighbourhood (s) within it.

The New York and Toronto dataframe was then combined. This new data represents all boroughs of Toronto and Manhattan; however, it also contained data on other boroughs of the city of Toronto and New York. We are concerned with the proper boroughs of Toronto and not its neighbouring municipalities like Mississauga and Scarborough. As it pertains to the New York City data, we are interested in Manhattan for this report. The following boroughs were dropped leaving only the boroughs of Manhattan, East Toronto, Central Toronto, Downtown Toronto, and West Toronto. Table 1. shows those boroughs which were dropped and those that were kept.

³ https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

⁴ https://geo.nyu.edu/catalog/nyu_2451_34572

⁵ <https://developer.foursquare.com/>

Table 1.

| City | Boroughs | Number of Neighbourhood |
|----------|---|-------------------------|
| Kept | | |
| New York | Manhattan | 40 |
| Toronto | East Toronto, Central Toronto, Downtown Toronto, West Toronto | 38 |
| Dropped | | |
| New York | 'Bronx', 'Brooklyn', 'Queens', 'Staten Island' | 266 |
| Toronto | 'Scarborough', 'North York', 'East York', 'York', 'Mississauga', 'Etobicoke | 41 |

Foursquare Data

Using the dataframe of Manhattan and Toronto neighbourhoods along with their respective coordinates, we can request data from Foursquare that is specific to those coordinates. The data that we're interested in relating to the types of venues located in and around a give neighbourhood's coordinate. From Foursquare, we retrieved the top one hundred venues that are in each neighbor within a radius of 500 meters of its respective coordinates.

Further Processing

The resulting dataframe holds 5036 rows of venues (Manhattan: 3326; Toronto: 1710). Each row provides the Neighbourhood, Neighbourhood Latitude, Neighbourhood Longitude, Venue, Venue Latitude, Venue Longitude, and the Venue Category. Taking the data in the Neighbourhood and Venue Category fields, we perform one-hot encoding. This will transform the categorical data in the Venue Category field into quantitative values between 0 and 1 for each Neighbourhood. The value is a function of the frequency at which a venue in each category occurs in the Neighbourhoods. Now there is a dataframe listing the neighbourhoods of Manhattan and Toronto along with values representing the kinds of venues that can be found in the areas. This dataframe is used to form the clusters.

Features

At this point, there are 78 samples (number of neighbourhoods) and 379 features (number of venue categories) in the data. These features are the terms in which neighbourhoods are described and are compared. Table 2. is an example of an element of the sample. This Neighbourhood is found in Manhattan called "Clinton." Each number in the column represents the proportion of venues retrieved from Foursquare found in Clinton in that venue category. For example, by reviewing the 'Theater' field, it is understood that 10% of the venues in Clinton retrieved from Foursquare are theaters. Conversely, it is also understood that 0% are yoga studios.

The goal here is to compare features of all the neighbourhoods – grouping together those neighbourhoods with a similar distribution of venues.

Table 2.

| | Theater | Gym / Fitness Center | American Restaurant | Spa | Italian Restaurant | Wine Shop | Hotel | Steakhouse | | New American Restaurant | French Restaurant | ... | Fish Market | Fish & Chips Shop | Filipino Restaurant | Field | Festival | Fast Food Restaurant | Farmers Market | Falafel Restaurant | Exhibit | Yoga Studio |
|---------------|---------|----------------------------|------------------------|------|-----------------------|--------------|-------|------------|--|-------------------------------|----------------------|-----|----------------|-------------------------|------------------------|-------|----------|-------------------------|-------------------|-----------------------|---------|----------------|
| Neighbourhood | Clinton | 0.1 | 0.05 | 0.04 | 0.04 | 0.03 | 0.03 | 0.02 | | 0.02 | 0.02 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

Methodology

Data Exploration

Landscape

As stated above, the data that we are sampling relates to neighbourhoods in the Manhattan and Toronto boroughs. Review Figures 1 and 2 below to see the distribution of neighbourhoods across the two landscapes.

Figure 1.

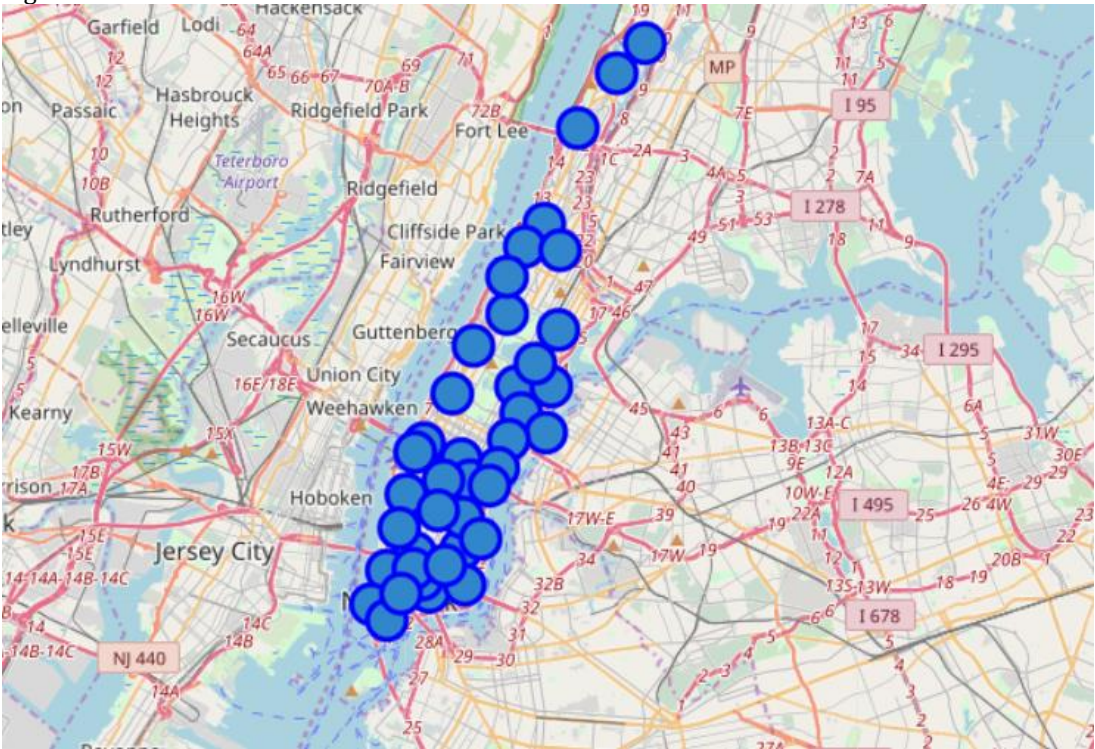
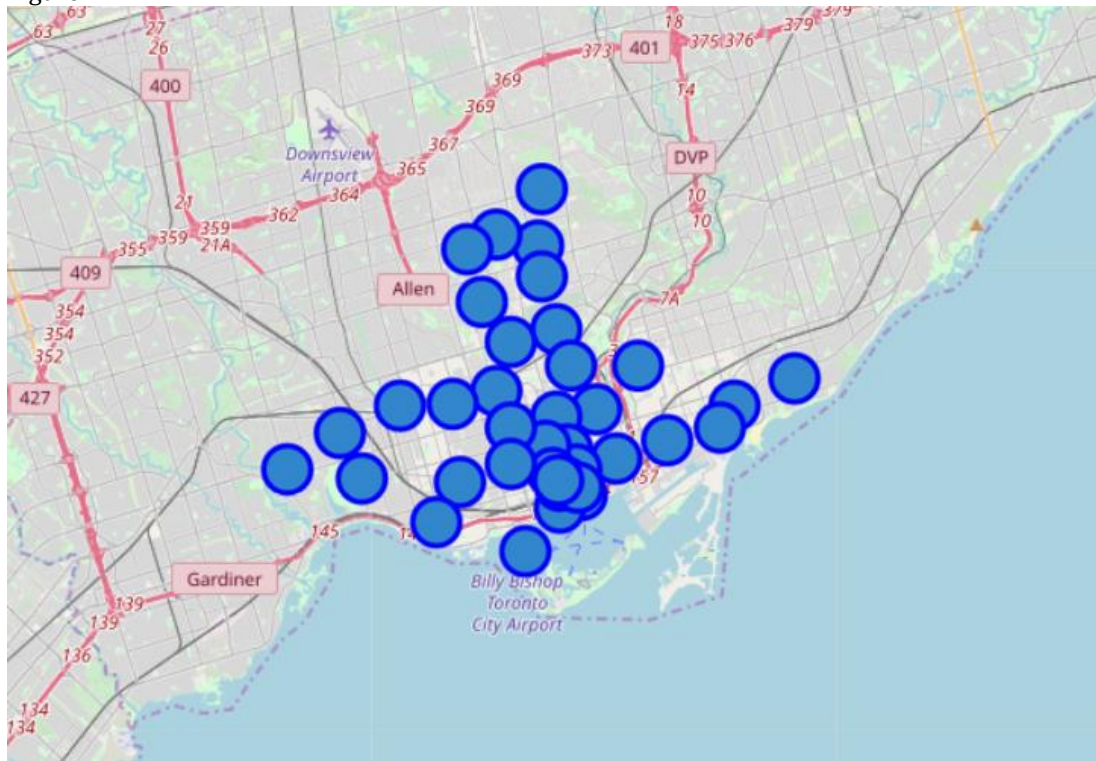


Figure 2.



Top Venues: Coffee Shop and Cafés

Caffeine is a very popular drug most often consumed using coffee beans. In Table 3. we see that 'Coffee Shop' is the most frequently observed venue category overall with 'Café' as the second most common. However, by parsing the data, we can see that the most common venue type in Manhattan is 'Italian Restaurant' with 'Café' as a close second. Coffee beverages are also sold at cafés, which further supports the notion that coffee is popular in both Manhattan and Toronto.

Although both coffee shops and cafés are known for selling coffee, for this report, they are represented as separate venue categories. This separation is intentional because although the terms are sometimes informally used interchangeably, the menu of a coffee shop is usually limited to coffee beverages, pastries, and small meals, cafés typical sell fuller meals for lunch and dinner. For this reason, it was decided to leave the type categories separate.

Table 3.

| City | Most Common Venue Category | % | 2 nd Most Common Venue Category | % |
|-----------|----------------------------|-----|--|-----|
| Manhattan | Italian Restaurant | 3.8 | Café | 3.7 |
| Toronto | Coffee Shop | 8.6 | Café | 5.5 |
| Overall | Coffee Shop | 5.4 | Café | 3.5 |

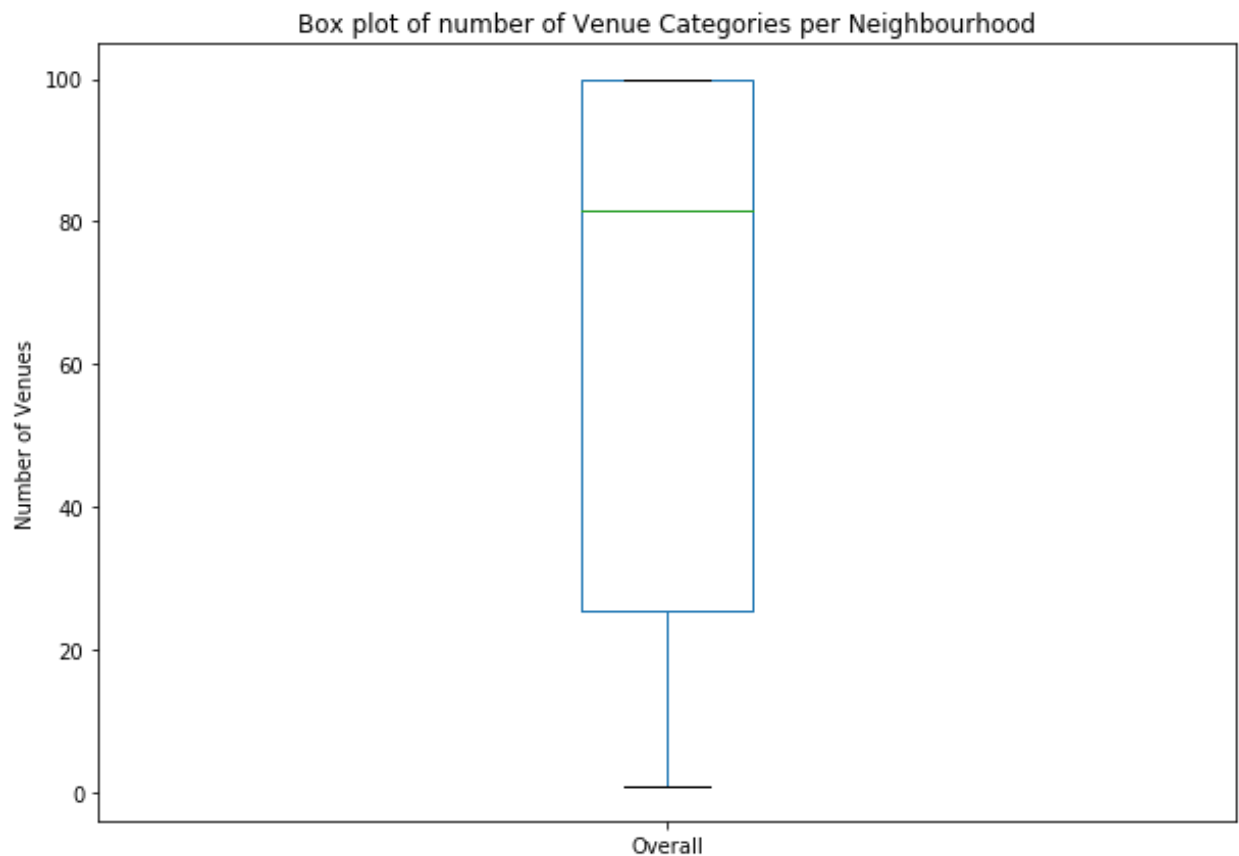
Distribution of Venues per Neighbourhood

The below table and box plot represent the distribution of the number of Venue Categories per Neighbourhood. A box plot is a useful tool when trying to identify outliers in a distribution of data. Here we can note the following: Based on the 3rd quartile being equal to the maximum value, we can state that 75% of the sampled neighbourhoods have maximum about of venues in the dataset (100). There are some neighbourhoods with only one venue recorded in the data set. Most importantly, there are no outliers in the data set. An outlier here is defined as a value, and that is greater than the third quartile or less than the first quartile by 1.5 times the interquartile range (IQR). The IQR of the data below is 74.5 (the difference between the third and first quartile. As we can see, no values are fitting that description.

Table 4.

| Statistic | Value |
|--------------------------|-------|
| Count | 78.0 |
| Maximum Value | 100.0 |
| 3 rd Quartile | 100.0 |
| Median | 81.5 |
| 1 st Quartile | 25.5 |
| Minimum Value | 1.0 |

Figure 3



Clustering Analysis

There are different kinds of methods used for clustering data. Some of these types are partitioning, hierarchical, and density-based clustering. The concept and objective are the same for these different types; however, the steps taken to achieve those objectives are different. Because of these differences, there are pros and cons to using either approach. For instance, a partitioning method of clustering called k-means requires the user to decide the number of clusters to group the data subjectively. On the other hand, the hierarchical approach does not require the user to choose the number of clusters. However, KMeans is generally considered more efficient than the hierarchical approach. In this report, we use the KMeans clustering method because it's efficient. To avoid the inherent bias in specifying the K number of clusters subjectively, we utilize the 'elbow method' to find the optimal number of 'k' clusters. The YellowBrick extension of the Scikit-Learn API is employed for this purpose.

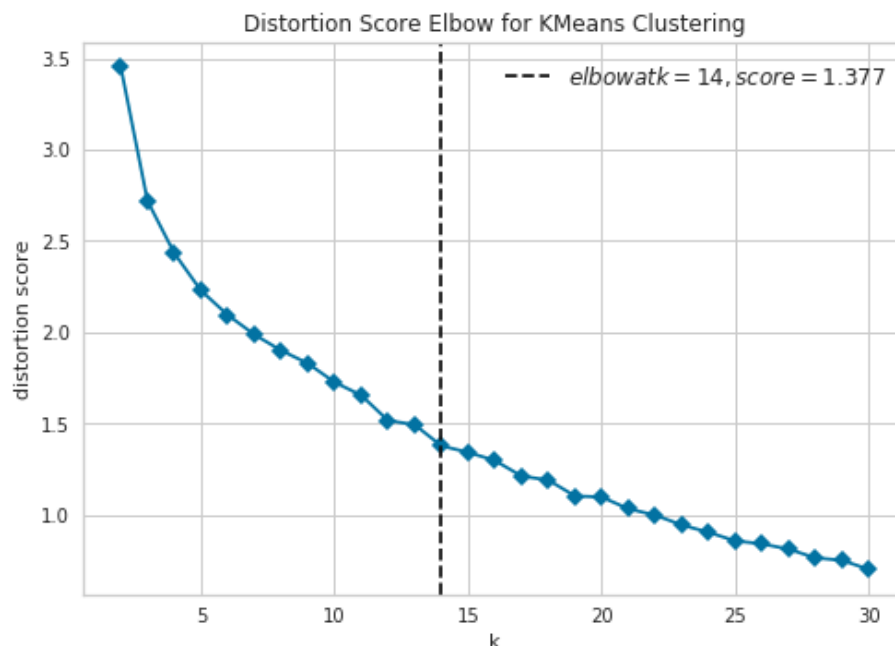
Optimal K Clusters

One method to choose the optimal number of clusters is the elbow method. The objective of the elbow method is to decrease the distance of elements within a given cluster. The steps of the elbow method are to first run the KMeans algorithm multiple times with varying about of clusters/centroids/ k. Then the distortion score is calculated, which is a measurement of the sum-squared-distance from each point to its respective centroid. As k approaches the sample size, the level of distortion decreases. This relationship between k and distortion isn't a bad thing; however, we do not want that many centroids. Thus, the last step is to select the point on the graph at which the marginal improvement begins to decline significantly (the elbow). Thankfully, the KElbowVisualizer of YellowBrick extension finds this point automatically.

Results

In the graph below, we can see that the KMeans algorithm was executed 30 times, and the elbow is located where k=14. Therefore, we proceed with 14 as the optimal number of clusters to model the data.

Figure 4.

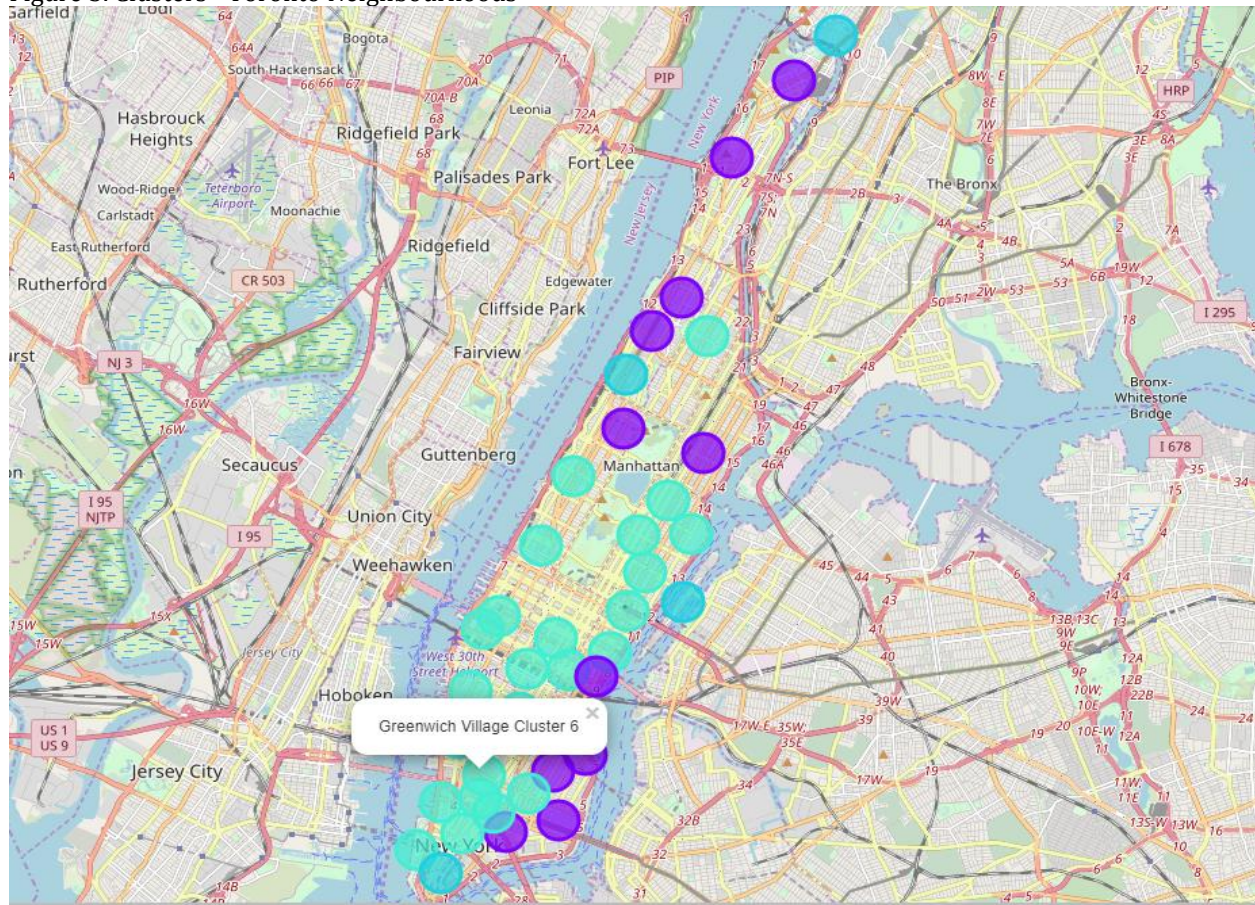


KMeans clustering

Using the KMeans approach with 14 centroids, the clusters depicted on the maps in Figures 5 & 6 were generated. Note the following:

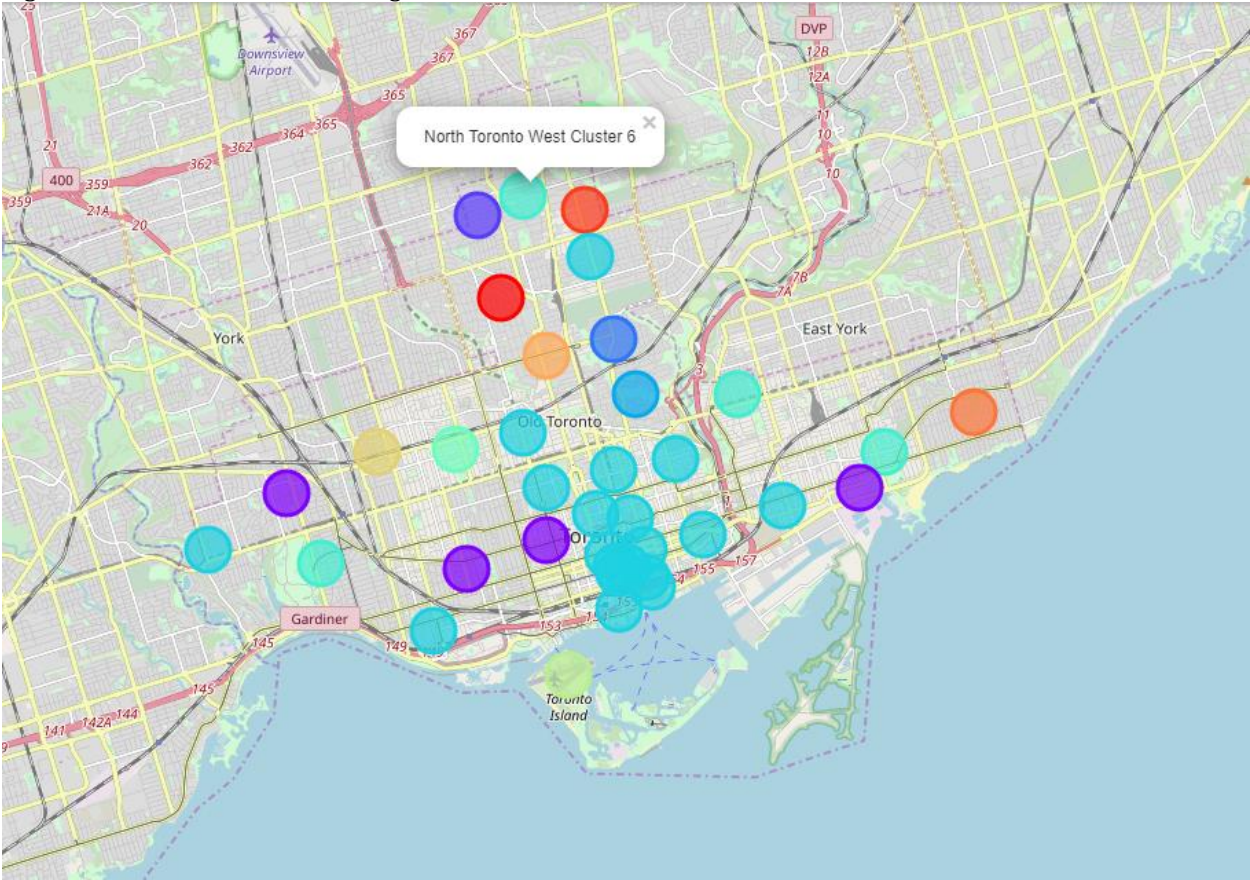
- The colors indicate neighbourhoods in the same cluster. For example, two neighbourhoods colored purple are a part of the same cluster.
- Neighbourhoods on two different maps are still a part of the same cluster if they are the same color. For example, Greenwich Village in Manhattan is of the same cluster as North Toronto West in Toronto.
- Table 5 below shows a sample of 4 neighbourhoods in cluster 1 and the top 5 venues categories of each Neighbourhood.

Figure 5. Clusters - Toronto Neighbourhoods



[73]:

Figure 6. Clusters - Manhattan Neighbourhoods



[74]:

Table 5.

| 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 |
|-----------|--------------------|-----------|------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Manhattan | Chinatown | 40.715618 | -73.994279 | 1 | Chinese Restaurant | Cocktail Bar | Bakery | American Restaurant | Optical Shop |
| Manhattan | Washington Heights | 40.851903 | -73.936900 | 1 | Café | Mobile Phone Shop | Bakery | Mexican Restaurant | Grocery Store |
| Manhattan | Inwood | 40.867684 | -73.921210 | 1 | Café | Mexican Restaurant | Deli / Bodega | Pizza Place | Lounge |
| Manhattan | Hamilton Heights | 40.823604 | -73.949688 | 1 | Café | Mexican Restaurant | Pizza Place | Coffee Shop | Deli / Bodega |
| Manhattan | Marble Hill | 40.876551 | -73.910660 | 5 | Sandwich Place | Coffee Shop | Miscellaneous Shop | Steakhouse | Supplement Shop |

Discussion

Recommendation

We now go back to the initial question, "Which neighbourhoods in Toronto are similar to those neighbourhoods of potential immigrants from Manhattan?" Table 4 lists the 14 clusters and the total number of neighbourhoods in each cluster based on their origin (Manhattan or Toronto). Note that only Cluster 1, 5, and 6 hold neighbourhoods from both Manhattan and Toronto. The remaining clusters are made up of only Toronto neighbourhoods. Those clusters are not useful to us. Moving forward, we shall only focus on elements included in those clusters. With that said, Tables 5, 6, and 7 display the neighbourhoods in clusters 1,5 and 6, respectively. These are Manhattan and Toronto neighbourhoods that are like each other. My recommendation is to market homes in neighbourhoods of Toronto to individuals migrating from neighbourhoods in Manhattan located within the same cluster. The logic is that those neighbourhoods would be more compatible with those migrants based on the idea that the

venues in the neighbourhoods match their former neighbourhoods in Manhattan. For example, if a new immigrant is from Marble Hill, Manhattan, he or she would probably be keen on Toronto's Davisville or King/Adelaide/Richmond neighbourhoods considering that they're both in Cluster 5 and have similar venues. The same can be said about those moving from Chinatown in Manhattan, which is in Cluster 1 along with the Kensington/Grange Town/Chinatown. These clusters may be very useful in narrowing down neighbourhoods matching the perspective buyers.

Table 6.

| Clusters | # Manhattan Neighbourhoods | # Toronto Neighbourhoods |
|----------|----------------------------|--------------------------|
| 0 | 0 | 1 |
| 1 | 12 | 4 |
| 2 | 0 | 1 |
| 3 | 0 | 1 |
| 4 | 0 | 1 |
| 5 | 4 | 19 |
| 6 | 24 | 4 |
| 7 | 0 | 1 |
| 8 | 0 | 1 |
| 9 | 0 | 1 |
| 10 | 0 | 1 |
| 11 | 0 | 1 |
| 12 | 0 | 1 |
| 13 | 0 | 1 |

Table 5. Cluster 1

| Cluster Labels | Borough | Neighbourhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----------------|------------------|---|-----------------------|-------------------------------|-----------------------|-----------------------|---------------------------|
| 1 | Manhattan | Chinatown | Chinese Restaurant | Cocktail Bar | Bakery | American Restaurant | Optical Shop |
| 1 | Manhattan | Washington Heights | Café | Mobile Phone Shop | Bakery | Mexican Restaurant | Grocery Store |
| 1 | Manhattan | Inwood | Café | Mexican Restaurant | Deli / Bodega | Pizza Place | Lounge |
| 1 | Manhattan | Hamilton Heights | Café | Mexican Restaurant | Pizza Place | Coffee Shop | Deli / Bodega |
| 1 | Manhattan | Manhattanville | Coffee Shop | Mexican Restaurant | Chinese Restaurant | Italian Restaurant | Park |
| 1 | Manhattan | East Harlem | Mexican Restaurant | Bakery | Deli / Bodega | Pizza Place | Latin American Restaurant |
| 1 | Manhattan | East Village | Bar | Wine Bar | Ice Cream Shop | Mexican Restaurant | Pizza Place |
| 1 | Manhattan | Lower East Side | Pizza Place | Café | Coffee Shop | Sandwich Place | Japanese Restaurant |
| 1 | Manhattan | Manhattan Valley | Indian Restaurant | Coffee Shop | Pizza Place | Szechuan Restaurant | Bar |
| 1 | Manhattan | Gramercy | Bar | Mexican Restaurant | Italian Restaurant | American Restaurant | Pizza Place |
| 1 | Manhattan | Tudor City | Park | Café | Mexican Restaurant | Greek Restaurant | Deli / Bodega |
| 1 | Manhattan | Stuyvesant Town | Bar | Park | Tennis Court | Fountain | Coffee Shop |
| 1 | Downtown Toronto | Kensington Market,Grange Park,Chinatown | Café | Vegetarian / Vegan Restaurant | Chinese Restaurant | Bar | Vietnamese Restaurant |
| 1 | West Toronto | Trinity,Little Portugal | Bar | Coffee Shop | Men's Store | Asian Restaurant | French Restaurant |
| 1 | West Toronto | High Park,The Junction South | Mexican Restaurant | Café | Thai Restaurant | Bookstore | Arts & Crafts Store |
| 1 | East Toronto | Business Reply Mail Processing Centre 969 Eastern | Light Rail Station | Comic Shop | Park | Brewery | Spa |

Table 6. Cluster 5

| Cluster Labels | Borough | Neighbourhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----------------|------------------|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 5 | Manhattan | Marble Hill | Sandwich Place | Coffee Shop | Miscellaneous Shop | Steakhouse | Supplement Shop |
| 5 | Manhattan | Roosevelt Island | Coffee Shop | Sandwich Place | Outdoors & Recreation | Greek Restaurant | Restaurant |
| 5 | Manhattan | Morningside Heights | Coffee Shop | Park | Bookstore | American Restaurant | Burger Joint |
| 5 | Manhattan | Financial District | Coffee Shop | Pizza Place | Wine Shop | Hotel | Gym |
| 5 | East Toronto | Studio District | Café | Coffee Shop | Bakery | Italian Restaurant | American Restaurant |
| 5 | Central Toronto | Davisville | Sandwich Place | Pizza Place | Dessert Shop | Italian Restaurant | Coffee Shop |
| 5 | Downtown Toronto | Cabbagetown,St. James Town | Restaurant | Café | Coffee Shop | Park | Italian Restaurant |
| 5 | Downtown Toronto | Church and Wellesley | Coffee Shop | Japanese Restaurant | Sushi Restaurant | Gay Bar | Restaurant |
| 5 | Downtown Toronto | Regent Park,Harbourfront | Coffee Shop | Café | Pub | Park | Bakery |
| 5 | Downtown Toronto | Ryerson,Garden District | Coffee Shop | Clothing Store | Cosmetics Shop | Café | Fast Food Restaurant |
| 5 | Downtown Toronto | St. James Town | Coffee Shop | Café | Restaurant | Italian Restaurant | Hotel |
| 5 | Downtown Toronto | Berczy Park | Coffee Shop | Cocktail Bar | Farmers Market | Bakery | Café |
| 5 | Downtown Toronto | Central Bay Street | Coffee Shop | Café | Italian Restaurant | Sandwich Place | Fried Chicken Joint |
| 5 | Downtown Toronto | King,Adelaide,Richmond | Coffee Shop | Café | Steakhouse | American Restaurant | Thai Restaurant |
| 5 | Downtown Toronto | Toronto Islands,Harbourfront East,Union Station | Coffee Shop | Hotel | Aquarium | Café | Scenic Lookout |
| 5 | Downtown Toronto | Toronto Dominion Centre,Design Exchange | Coffee Shop | Café | Hotel | Restaurant | Bar |
| 5 | Downtown Toronto | Commerce Court,Victoria Hotel | Coffee Shop | Café | Hotel | Restaurant | American Restaurant |
| 5 | Central Toronto | The Annex,Yorkville,North Midtown | Sandwich Place | Café | Pizza Place | Coffee Shop | Indian Restaurant |
| 5 | Downtown Toronto | University of Toronto,Harbord | Café | Bakery | Restaurant | Italian Restaurant | Japanese Restaurant |
| 5 | Downtown Toronto | Stn A PO Boxes 25 The Esplanade | Coffee Shop | Café | Restaurant | Italian Restaurant | Fast Food Restaurant |
| 5 | Downtown Toronto | Underground city,First Canadian Place | Coffee Shop | Café | Hotel | Steakhouse | Restaurant |
| 5 | West Toronto | Exhibition Place,Brockton,Parkdale Village | Breakfast Spot | Café | Coffee Shop | Restaurant | Stadium |
| 5 | West Toronto | Runnymede,Swansea | Café | Coffee Shop | Pizza Place | Restaurant | Sushi Restaurant |

Table 7. Cluster 6

| Cluster Labels | Borough | Neighbourhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|----------------|-----------------|-------------------------------|-----------------------|-------------------------|-----------------------|------------------------|-----------------------------|
| 6 | Manhattan | Central Harlem | American Restaurant | French Restaurant | Chinese Restaurant | Art Gallery | Bar |
| 6 | Manhattan | Upper East Side | Exhibit | Italian Restaurant | Bakery | Art Gallery | Juice Bar |
| 6 | Manhattan | Yorkville | Italian Restaurant | Gym | Coffee Shop | Bar | Pizza Place |
| 6 | Manhattan | Lenox Hill | Coffee Shop | Italian Restaurant | Pizza Place | Sushi Restaurant | Cosmetics Shop |
| 6 | Manhattan | Upper West Side | Italian Restaurant | Coffee Shop | Wine Bar | Bar | Indian Restaurant |
| 6 | Manhattan | Lincoln Square | Theater | Italian Restaurant | Gym / Fitness Center | Café | Plaza |
| 6 | Manhattan | Clinton | Theater | Gym / Fitness Center | Italian Restaurant | Spa | American Restaurant |
| 6 | Manhattan | Midtown | Hotel | Steakhouse | Coffee Shop | Theater | Clothing Store |
| 6 | Manhattan | Murray Hill | Coffee Shop | Hotel | Sandwich Place | Japanese Restaurant | Italian Restaurant |
| 6 | Manhattan | Chelsea | Coffee Shop | Ice Cream Shop | Bakery | Italian Restaurant | Nightclub |
| 6 | Manhattan | Greenwich Village | Italian Restaurant | Clothing Store | Sushi Restaurant | Indian Restaurant | Café |
| 6 | Manhattan | Tribeca | American Restaurant | Park | Italian Restaurant | Spa | Café |
| 6 | Manhattan | Little Italy | Bakery | Café | Italian Restaurant | Clothing Store | Mediterranean Restaurant |
| 6 | Manhattan | Soho | Clothing Store | Boutique | Art Gallery | Women's Store | Shoe Store |
| 6 | Manhattan | West Village | Italian Restaurant | New American Restaurant | American Restaurant | Wine Bar | Park |
| 6 | Manhattan | Battery Park City | Park | Coffee Shop | Memorial Site | Gym | Hotel |
| 6 | Manhattan | Carnegie Hill | Pizza Place | Coffee Shop | Bar | Café | Yoga Studio |
| 6 | Manhattan | Noho | Italian Restaurant | Cocktail Bar | French Restaurant | Art Gallery | Mexican Restaurant |
| 6 | Manhattan | Civic Center | Gym / Fitness Center | Sandwich Place | Italian Restaurant | Hotel | French Restaurant |
| 6 | Manhattan | Midtown South | Korean Restaurant | Hotel | Hotel Bar | Japanese Restaurant | Cosmetics Shop |
| 6 | Manhattan | Sutton Place | Gym / Fitness Center | Indian Restaurant | Italian Restaurant | Furniture / Home Store | Gym |
| 6 | Manhattan | Turtle Bay | Italian Restaurant | Coffee Shop | Sushi Restaurant | Steakhouse | Wine Bar |
| 6 | Manhattan | Flatiron | Gym / Fitness Center | Yoga Studio | Gym | Japanese Restaurant | American Restaurant |
| 6 | Manhattan | Hudson Yards | American Restaurant | Gym / Fitness Center | Italian Restaurant | Café | Restaurant |
| 6 | East Toronto | The Danforth West,Riverdale | Greek Restaurant | Coffee Shop | Italian Restaurant | Ice Cream Shop | Furniture / Home Store |
| 6 | East Toronto | The Beaches West,India Bazaar | Movie Theater | Italian Restaurant | Sandwich Place | Park | Sushi Restaurant |
| 6 | Central Toronto | North Toronto West | Clothing Store | Sporting Goods Shop | Coffee Shop | Restaurant | Spa |
| 6 | West Toronto | Roncesvalles,Parkdale | Breakfast Spot | Gift Shop | Bookstore | Restaurant | Eastern European Restaurant |

Limitation / Future Opportunity

There are a few limitations to this report; however, these limitations present an opportunity for further research in the area.

- The assumption here is that people migrating from Manhattan to Toronto would like to live in a neighbourhood like the one they are leaving. It can also be true that they want to live in a new neighbourhood, considering that they are moving to a new city.
 - Real estate agents may resolve this by first asking their clients whether they want to live in a neighbourhood like the ones they are leaving. If the answer is 'yes', then the results from this report would help narrow that search down.
 - Also, further survey research can be done to assess what factors most influence immigrants' preference regarding where to Toronto would like to live.
- Elements of the sample representing neighbourhoods in Toronto grouped neighbourhoods based on zip code. For example, Kensington Market, Grange Park, and Chinatown are all treated as one neighbourhood into the dataset; although, there may be some differences between Kensington Market and Chinatown.
 - Further research may be done to consider whether there's any material difference in the results of this report if neighbourhoods in Toronto are not grouped based on zip code.
- KMeans was used to create the clusters in the data.

- Further research could be done to consider whether there's any material difference in the results of this report if other clustering methods were used, such as hierarchical clustering and DBSCAN.

Conclusion

In this report, we analyzed the similarities between neighbourhoods in Manhattan and Toronto based on the types of venues found in those neighbourhoods. This was done using locational data provided Foursquare, which included the venue's categories and their coordinates. We identified 14 clusters in which neighbourhoods in Manhattan and Toronto can be grouped by using the KMeans approach. Manhattan neighbourhoods fit into three clusters, and within those clusters, there were neighbourhoods in Toronto that were also included. Those neighbourhoods in Toronto match the respective neighbourhoods in Manhattan. This information can be useful in helping real estate agents select the most suitable Neighbourhood for clients immigrating from Manhattan to Toronto.