BATTLE OF NEIGHBORHOODS

MANHATTAN AND TORONTO

COURSERA CAPSTONE PROJECT

RUCHIT SHAH

INTRODUCTION

- Manhattan, the most densely populated of the five boroughs of NYC, has long been the flag bearer for people worldwide coming to the US to live the "American Dream"

- Growing immigration queues and visa issues in the US

- Large number of technical and financial workers moving to Canada

- Choosing a neighborhood to move to in a new city (let alone country) can be complicated, but with the help of data science, the task can be made a little bit easier.

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

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.

DATA

- BeautifulSoup was used to extract a full list of all Toronto neighborhoods from the Wikipedia webpage "List of postal codes of Canada"

- Manhattan data was extracted from NYU Spatial Data Repository

- geopy library was used to obtain the latitude and longitude values for each neighborhood

- Foursquare API was used to explore the neighboring venues

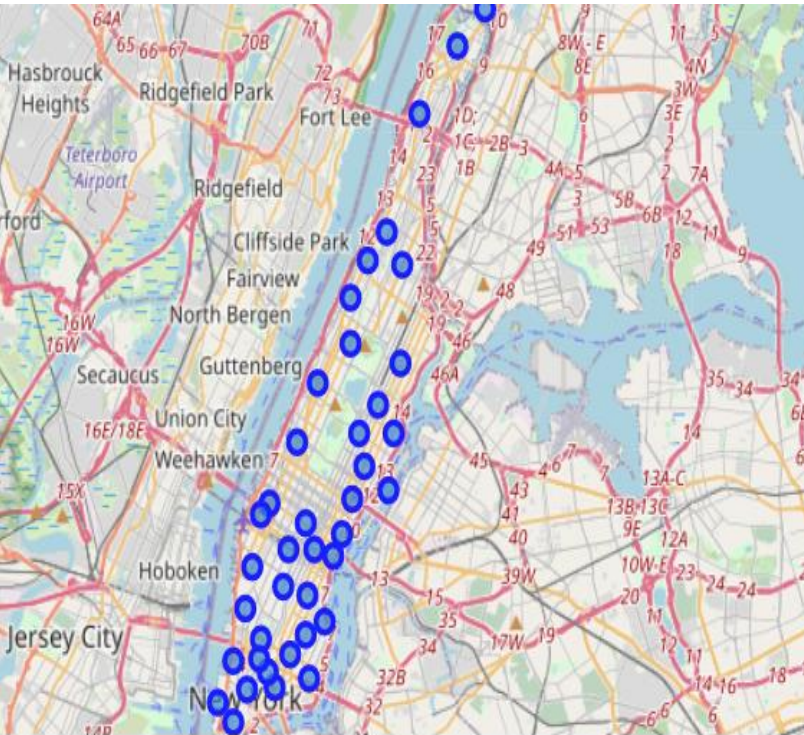- Manhattan had 40 neighborhoods and Toronto had 39 neighborhoods

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

| Borough	Neighborhood	Latitude	Longitude
Manhattan	Marble Hill	40.876551	-73.910660
Manhattan	Chinatown	40.715618	-73.994279
Manhattan	Washington Heights	40.851903	-73.936900
Manhattan	Inwood	40.867684	-73.921210
Manhattan	Hamilton Heights	40.823604	-73.949688
Manhattan	Manhattanville	40.816934	-73.957385

METHODOLOGY – LOCATION DATA

MANHATTAN NEIGHBORHOODS

- 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

- Dataframes for both locations were 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.
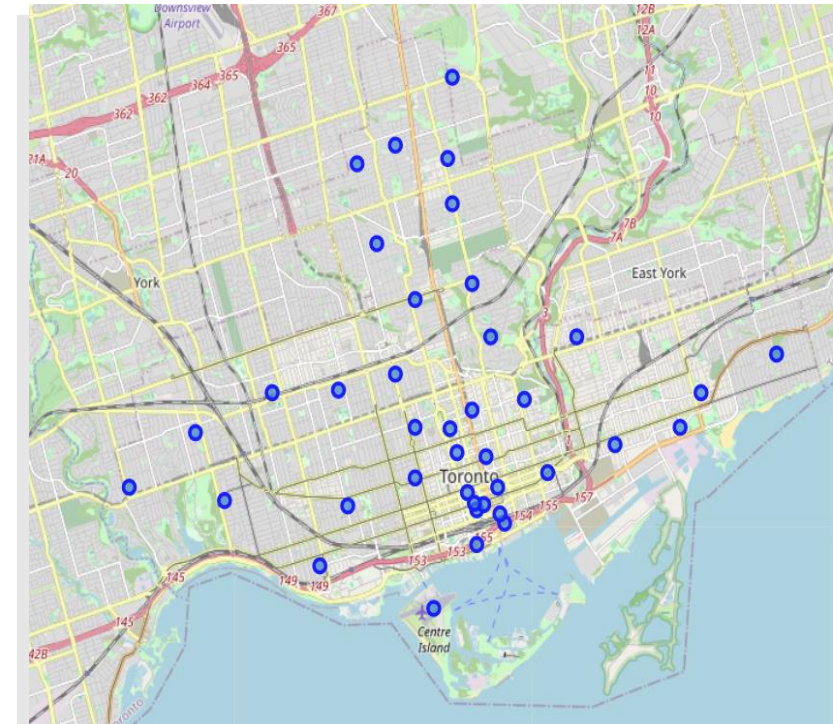
TORONTO NEIGHBORHOODS

METHODOLOGY – EXPLORE NEARBY VENUES

- 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

- 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

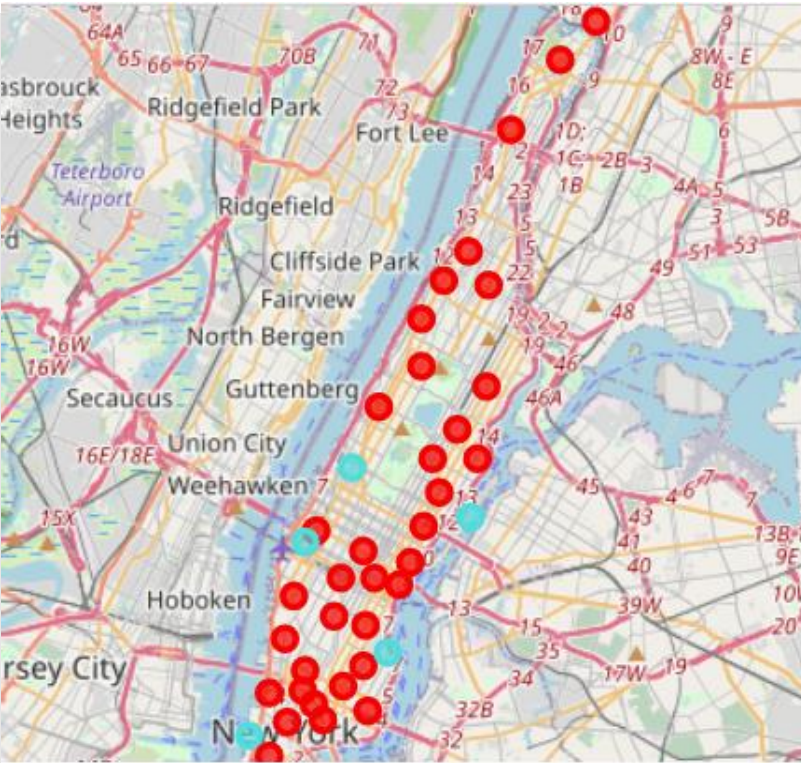|	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Food	Outdoors & Recreation	Shop & Service	Travel & Transport	Professional & Other Places
1	Berczy Park	Food	Shop & Service	Nightlife Spot	Arts & Entertainment	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

Example of 5 most commonly occurring general categories for each neighborhood

- The mean of the frequency of occurrence of each category for each neighborhood was computed

- Top 5 general categories for each neighborhood was chosen as the feature set for clustering

METHODOLOGY – CLUSTERING

- 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 used to plot the neighborhoods on the maps of Manhattan and Toronto for visual inspection.
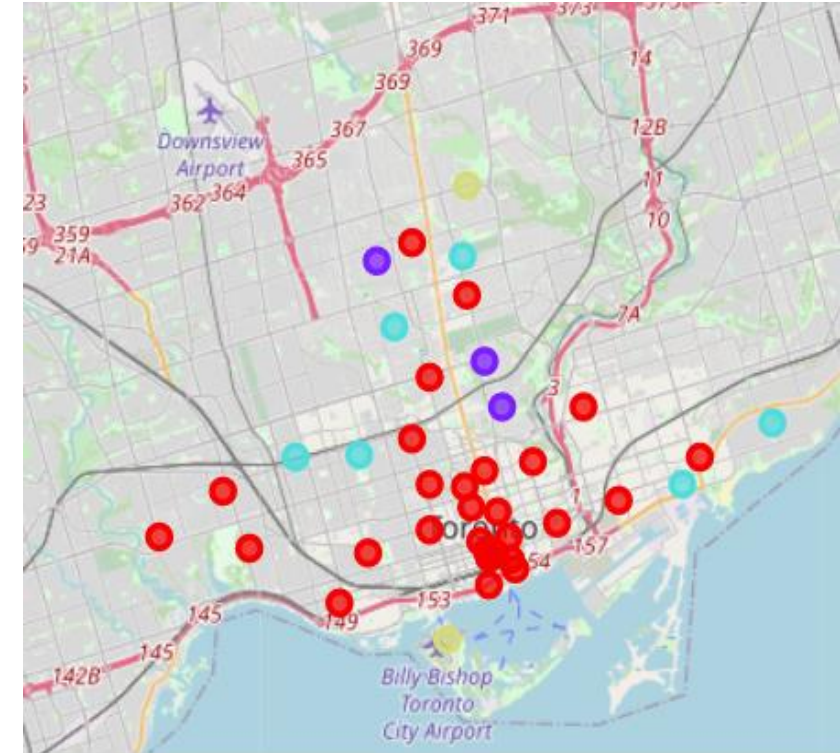
RESULTS

MANHATTAN NEIGHBORHOODS

- Cluster labels assigned by the clustering algorithm were used to plot and analyze similar Manhattan and Toronto neighborhoods

- No Manhattan neighborhoods were assigned to clusters 2 and 4

TORONTO NEIGHBORHOODS

RESULTS- CLUSTER ANALYSIS

CLUSTER 1

- Total of 64 neighborhoods (36 from Manhattan and 28 from Toronto)
- Most commonly occurring categories showed Food, Nightlife Spots and, Shop and Service venues to be most prevalent in these neighborhoods

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

Example of neighborhoods assigned to Cluster 1

RESULTS- CLUSTER ANALYSIS

CLUSTER 2

- 3 Toronto neighborhoods
- Outdoor recreation and Travel and Transport appeared to be the most prevalent general categories in this cluster

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

Example of neighborhoods assigned to Cluster 2

RESULTS- CLUSTER ANALYSIS

CLUSTER 3

- Five Manhattan and six Toronto neighborhoods
- Outdoor recreations and Food along with Shop and Service and Residential venues appeared to be the key features for this cluster

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

Example of neighborhoods assigned to Cluster 3

RESULTS- CLUSTER ANALYSIS

CLUSTER 4

- 2 Toronto neighborhoods

- 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

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

Example of neighborhoods assigned to Cluster 4

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

- Analysis in this project was a crude

- 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.

CONCLUSION

- With the help of Data Science tools 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.

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-trun

[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