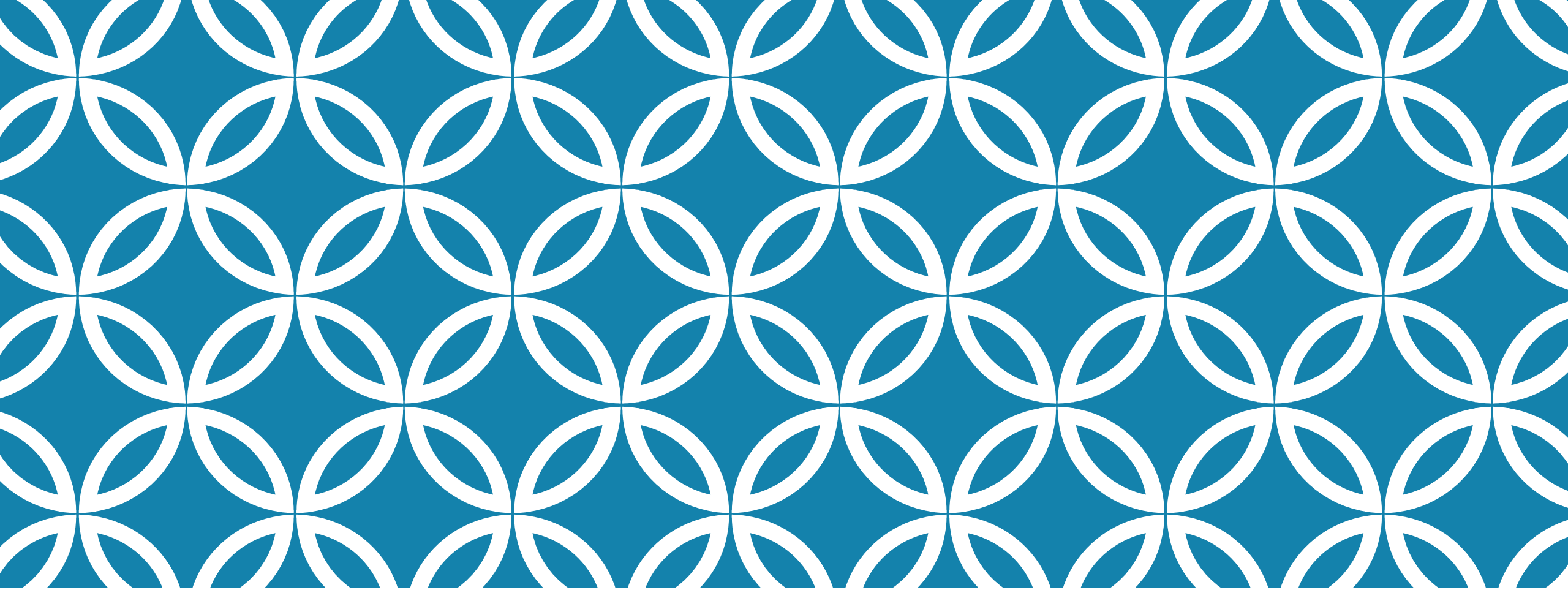

FINDING THE BEST NEIGHBORHOOD
(THE BATTLE OF NEIGHBORHOODS)

Farrokh Manzouri

IBM Capstone Project

INTRODUCTION

In today's dynamic world, it is common that people find a new job and have to move to a new city or neighborhood.

Probably this person would prefer to move to a location that is similar to the place he lives currently in. Because, he can continue to follow his hobbies and habits and can integrate easier and faster.

To this end, here we want to provide a solution and find out what are the similar neighborhoods in the new city that are similar to the current neighborhood.

PROBLEM DEFINITION

As an example we assume one person is living in Midtown, Manhattan in the New York City.

Now he wants to move to Toronto city in Downtown to start a new job and he wants to find the most similar neighbourhood to his current location in Toronto city.

METHODOLOGY

To solve this problem, we need the borough and neighborhood data of the current neighborhood and the destination borough.

We can get a list of New York City Neighborhood Names from the https://geo.nyu.edu/ website.

	Borough	Neighbourhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 1. Sample data from New York dataset

METHODOLOGY

Let's visualize Manhattan and the neighborhoods in it

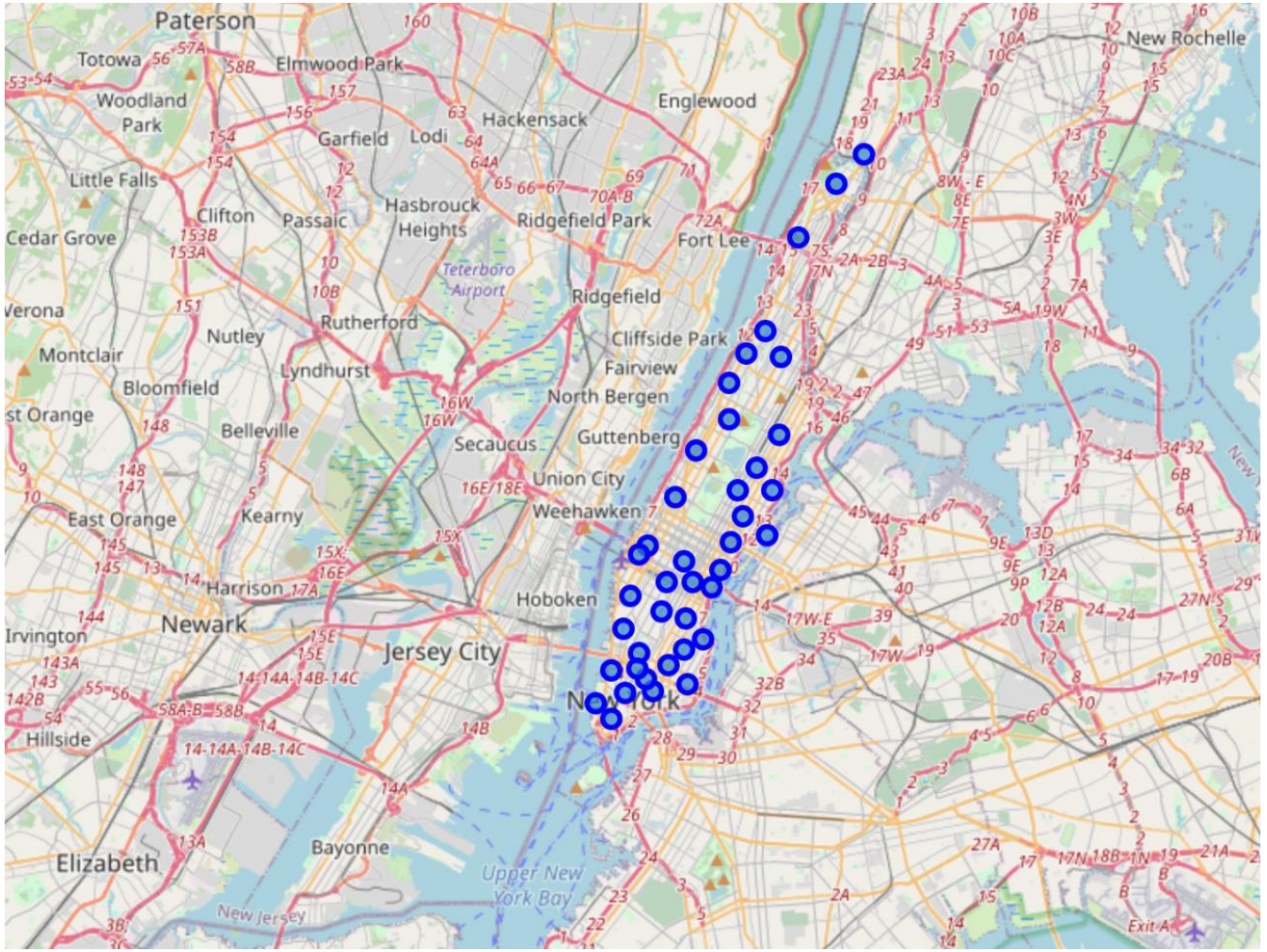

Figure 2. Manhattan, New York. Each neighborhood is pointed with a blue dot

METHODOLOGY

Once we have the list of neighborhoods and corresponding latitudinal and longitudinal information we select our neighborhood of interest.

Figure 3. Latitudinal and longitudinal data of Midtown, Manhattan

	Borough	Neighborhood	Latitude	Longitude
15	Manhattan	Midtown	40.754691	-73.981669

Next, we use Foursquare API to get the venues in Midtown neighborhood. Here you can see a list of the venues:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Midtown	40.754691	-73.981669	Bryant Park	40.753621	-73.983265	Park
1	Midtown	40.754691	-73.981669	New York Public Library Terrace	40.753017	-73.981480	Plaza
2	Midtown	40.754691	-73.981669	Nat Sherman Townhouse	40.753283	-73.980358	Smoke Shop
3	Midtown	40.754691	-73.981669	Joanna Vargas Skin Care	40.753136	-73.980721	Spa
4	Midtown	40.754691	-73.981669	sweetgreen	40.754640	-73.983102	Salad Place

Figure 4. List of venues in Midtown, Manhattan

METHODOLOGY

For Toronto city as the destination, because the data cannot be directly downloaded, Postal Code, borough, and neighborhood are scrapped from the Wikipedia website.

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
7	M8A	Not assigned	Not assigned
8	M9A	Etobicoke	Islington Avenue, Humber Valley Village
9	M1B	Scarborough	Malvern, Rouge
10	M2B	Not assigned	Not assigned

Figure 5. List of neighborhoods in Toronto

METHODOLOGY

Geographical coordinates of the neighborhoods can be downloaded from this address: https://cocl.us/Geospatial_data.

Next, this information is merged with Toronto data together.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937

Figure 6. List of neighborhoods in Toronto with their coordination

METHODOLOGY

Because the person who is moving from New York to Toronto has found a job in Dow

ntown Toronto, we will select only the neighborhoods in downtown Toronto.

As a result, we slice the original data frame and create a new data frame of the downtown data.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306

Figure 7. List of neighborhoods in Downtown Toronto with their coordination

METHODOLOGY

Let's visualize downtown and the neighborhoods in it

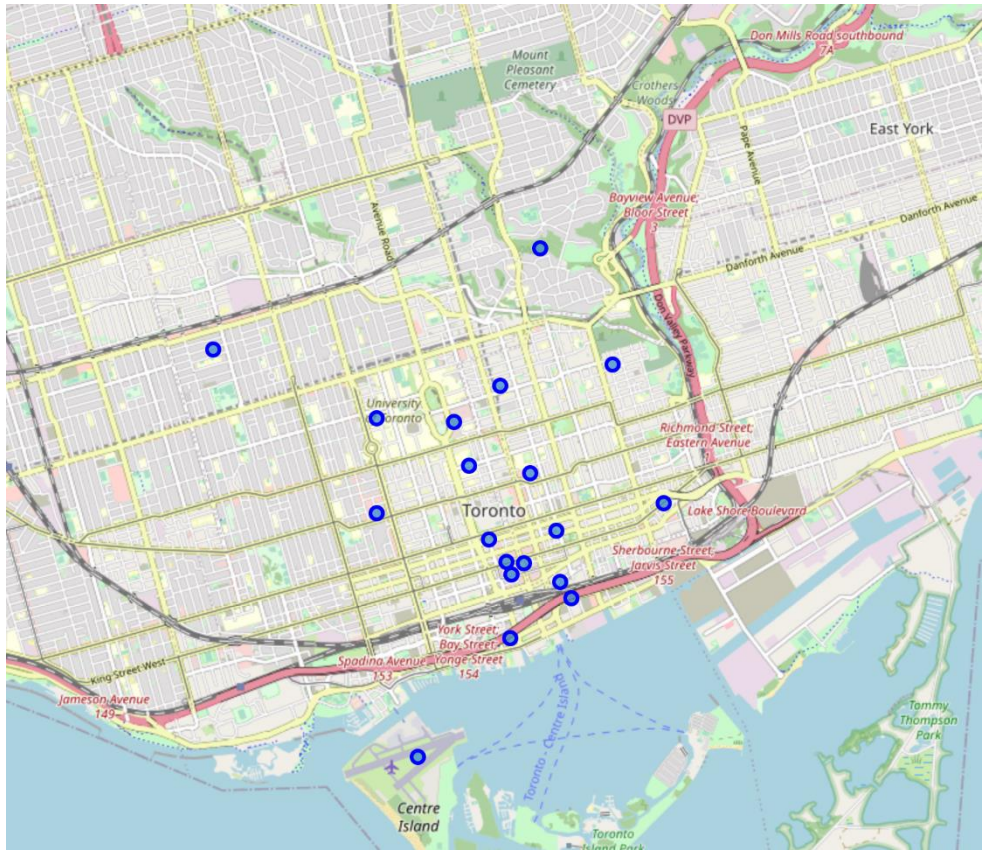

Figure 8. Downtown, Toronto. Each neighborhood is pointed with a blue dot

METHODOLOGY

With the list of neighborhoods and corresponding Geographic information we will use Foursquare API to get the venues near each neighborhood.

To this end, we need to have a Foursquare account to get the required credentials.

Figure 9. Sample data from Toronto dataset

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

METHODOLOGY

Finding nearby Venues

First, we will find the nearby venues to the current neighborhood.

Next, we will find the venues in the neighborhoods of the destination borough.

One-hot Encoding

Subsequently, we will do One-hot encoding for Downtown venues in Toronto and for Midtown venues in New York

METHODOLOGY

Figure 10. One-hot encoding of venues in Downtown Toronto

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	[illegible] & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Baby Store	Bagel Shop	Bakery	Bank	Bar	Baseball Stadium	Basketball Stadium	Beach	Bed & Breakfast	Beer Bar	Beer Store	Belgian Restaurant	Bistro
0	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Regent Park, Harbourfront	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 11. One-hot encoding of venues in Midtown, Manhattan

	Neighborhood	American Restaurant	Art Gallery	Bakery	Bar	Bookstore	Boutique	Boxing Gym	Café	Chinese Restaurant	Clothing Store	Coffee Shop	Concert Hall	Cosmetics Shop	Cuban Restaurant	Cycle Studio	Deli / Bodega	Discount Store	Donut Shop	Fast Food Restaurant	Food Stand	Food Truck	French Restaurant	Golf Course	Gourmet Shop	Grocery Store	Gym	Gym / Fitness Center	Hawaiia Restaurar
0	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
6	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	Midtown	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	Midtown	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	

METHODOLOGY

Common Venue Categories

Considering the fact that some of the venue categories in midtown may not exist in the Downtown Toronto borough, we will select only the common venue categories for analysis.

Calculating Mean Frequency of Occurrence

Next, we group rows by neighborhood and calculate the mean frequency of occurrence for each venue category.

This will be done for both datasets

METHODOLOGY

	Neighborhood	American Restaurant	Art Gallery	Bakery	Bar	Bookstore	Boutique	Café	Chinese Restaurant	Clothing Store	Coffee Shop	Concert Hall	Cosmetics Shop	Deli / Bodega	Discount Store	Donut Shop	Fast Food Restaurant	Food Truck	French Restaurant	Gourmet Shop	Grocery Store	Gym	Gym / Fitness Center	Health & Beauty Service	Historic Site
0	Berczy Park	0.000000	0.018182	0.036364	0.000000	0.000000	0.0000	0.018182	0.000000	0.000000	0.090909	0.018182	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.018182	0.018182	0.018182	0.000000	0.000000	0.000000	0.000000
1	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.000000	0.000000	0.062500	0.000000	0.0625	0.000000	0.000000	0.000000	0.062500	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Central Bay Street	0.000000	0.000000	0.000000	0.000000	0.014706	0.0000	0.058824	0.000000	0.000000	0.176471	0.000000	0.000000	0.000000	0.014706	0.014706	0.000000	0.000000	0.014706	0.000000	0.000000	0.000000	0.014706	0.000000	0.000000
3	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000	0.187500	0.000000	0.000000	0.062500	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000	0.000000	0.000000	0.000000
4	Church and Wellesley	0.013333	0.000000	0.000000	0.000000	0.013333	0.0000	0.026667	0.013333	0.013333	0.093333	0.000000	0.000000	0.000000	0.000000	0.000000	0.013333	0.000000	0.000000	0.000000	0.000000	0.013333	0.000000	0.013333	0.000000
5	Commerce Court, Victoria Hotel	0.040000	0.010000	0.010000	0.000000	0.010000	0.0000	0.060000	0.000000	0.000000	0.130000	0.010000	0.000000	0.030000	0.000000	0.000000	0.010000	0.010000	0.010000	0.000000	0.000000	0.040000	0.010000	0.000000	0.000000
6	First Canadian Place, Underground city	0.030000	0.010000	0.010000	0.020000	0.010000	0.0000	0.070000	0.000000	0.000000	0.110000	0.020000	0.000000	0.030000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.040000	0.010000	0.000000	0.000000
7	Garden District, Ryerson	0.000000	0.010000	0.010000	0.000000	0.020000	0.0000	0.040000	0.010000	0.090000	0.090000	0.000000	0.030000	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000	0.000000	0.000000	0.010000	0.010000	0.000000	0.000000

Figure 12. Mean of the frequency of occurrence of each venue category in downtown, Toronto

	Neighborhood	American Restaurant	Art Gallery	Bakery	Bar	Bookstore	Boutique	Café	Chinese Restaurant	Clothing Store	Coffee Shop	Concert Hall	Cosmetics Shop	Deli / Bodega	Discount Store	Donut Shop	Fast Food Restaurant	Food Truck	French Restaurant	Gourmet Shop	Grocery Store	Gym	Gym / Fitness Center	Health & Beauty Service	Historic Site	Hotel	Indian Restaurant	Italian Restaurant
0	Midtown	0.01	0.01	0.05	0.01	0.03	0.01	0.02	0.01	0.05	0.05	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.06	0.01	0.01

Figure 13. Mean of the frequency of occurrence of each venue category in midtown, Manhattan

METHODOLOGY

Similarity Measurement

Using a cosine similarity measure, we can compare the similarity between the current neighborhood and the neighborhoods in the destination borough to find a neighborhood that has the highest similarity.

Cosine similarity

is defined as a measure of similarity between two non-zero vectors of an inner product space.

The cosine of two non-zero vectors can be calculated by using the Euclidean dot product formula:

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \, \|\mathbf{B}\| \cos\theta$$ Equation. 1

RESULTS

The results of the similarity measurement shows that in our case the best match is **Toronto Dominion Centre, Design Exchange** with the following mean of the frequency of occurrence:

Here we can see the most common venue categories in the elected neighborhood:

	Neighborhood	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
17	Toronto Dominion Centre, Design Exchange	Coffee Shop	Hotel	Café	Salad Place	Japanese Restaurant	American Restaurant	Steakhouse	Bakery	Bar	Concert Hall

Figure 14. Most common venues in the elected neighborhood in Downtown Toronto

RESULTS

We can compare the results with most common venue categories in the current neighborhood in Midtown, Manhattan:

1- Hotel
2- Coffee Shop
3- Bakery
4- Clothing Store
5- Theater
6- Steakhouse
7- Sporting Goods Shop
8- Sandwich Place
9- Bookstore
10- Pizza Place

CONCLUSION

- Finding two similar neighborhoods using the available data about the venues can solve a number of problems including moving to a new city

- To this end, we calculated the mean frequency of occurrence for each venue category in each neighborhood, and then measured similarity using cosine similarity

- The search area and other parameters can be adjusted for each specific problem. Besides, there are several similarity measures that can be used as an alternative including Euclidean distance or Jaccard similarity.