Compare neighbourhoods of New York and Toronto

Introduction

In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighbourhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighbourhoods of the two cities and determine how similar or dissimilar they are.

And this project we will try to find out the similarity of the two city and we can identify similar districts of the two city. If people would like to move from Toronto to New York or vice versa, they can choose a similar district to reside, so that they do not need too much time to integrate to the community.

Data Collection

We use the same data set used in module 3.

Toronto's Data

We collect neighbourhoods data from Wikipedia (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and location data from https://cocl.us/Geospatial_data

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

New York's Data

We collect neighbourhoods data from https://cocl.us/new_york_dataset

]:

	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

Apart from the location data, we also collected venue data from foursquare API.

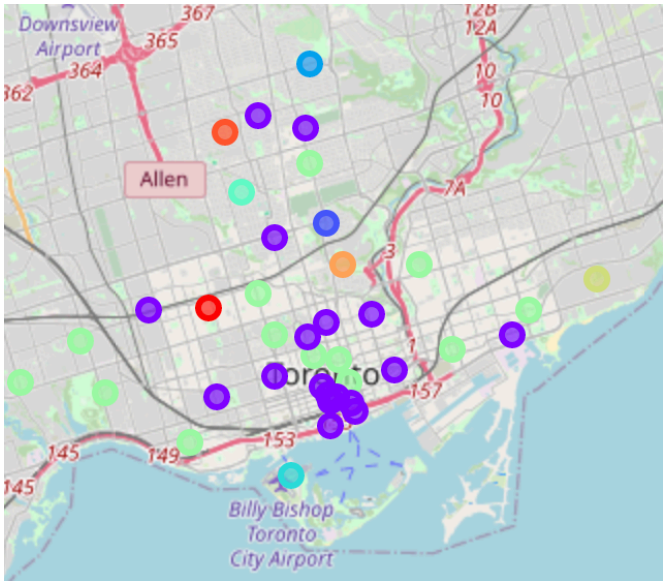
Methods

We first encode the data of different type of venue in different district using one-hot encoding. And then grouping to 10 different group different district using k-mean squared method.

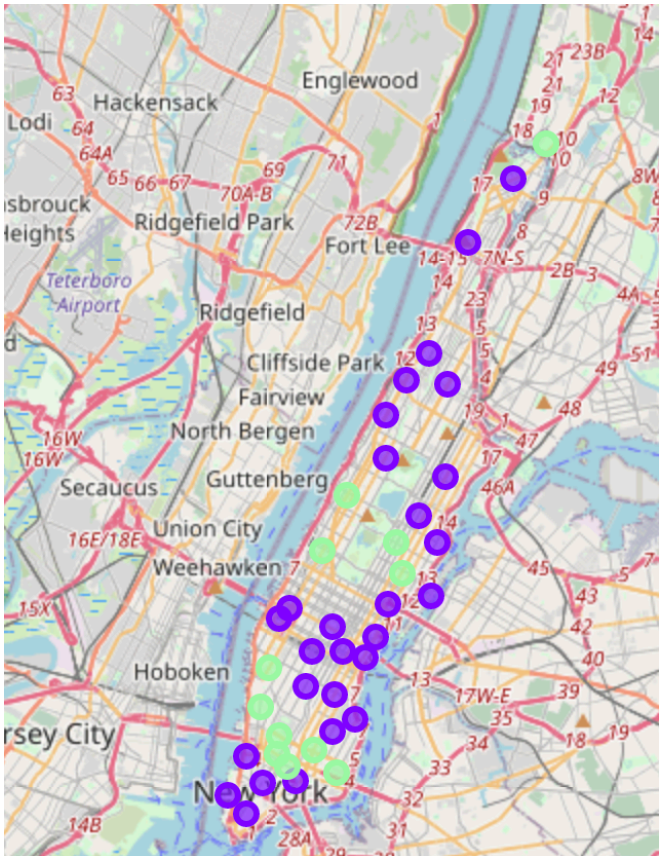
	Neighborhood	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	...	Video Game Store	Video Store	Vietnamese Restaurant	Volleyb Cou
0	Battery Park City	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0
1	Berczy Park	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0
2	Brockton, Parkdale Village, Exhibition Place	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0
3	Business reply mail Processing Centre, South C...	0.062500	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0
4	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.000000	0.0	0.0	0.0	0.055556	0.055556	0.055556	0.111111	...	0.000000	0.00	0.000000	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
74	Upper East Side	0.033708	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0
75	Upper West Side	0.012195	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.012195	0
76	Washington Heights	0.000000	0.011364	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.011364	0.00	0.000000	0
77	West Village	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00	0.000000	0
78	Yorkville	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.01	0.020000	0

After clustering, we have got the following result:

Toronto:



Manhattan:



Most of the neighbourhoods in Toronto and Manhattan fall into Cluster 2 and 7 (as in the jupyter notebook). And every district in Manhattan is similar to most of the district in Toronto. With only a few district in Toronto is unique.
People should easily move from one place to another without any issues.