Capstone Project- The Battle of Neighbourhoods

Introduction

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario.

Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

The aim of this report is to study and analyse the neighbourhoods of Toronto city and group them into similar clusters and, to analyse those clusters to gather meaningful information. That information can be used to find out neighbourhoods that are same as your current neighbourhood or at least similar.

Data

Data used in the analysis are listed below:

· Neighbourhoods in Toronto -- Wikipedia. I cleaned the data and reduced it to boroughs of Toronto so that I can use it to find geological locations for further venue analysis.

· Using **Geopy** to get geological location by address name

· Using **Foursquare API** to get the most common venues of given Borough of New York City.

· Using **Foursquare API** to get the venues' record of given venues of New York City.

Methodology

To start with our analysis, we used the BeautifulSoup package to transform the data in the table on the Wikipedia page into the below pandas data frame.

	Postalcode	Borough	Neighborhood
0	M1A	Not assigned	
1	M2A	Not assigned	
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

We also fetched the coordinate data for all the neighbourhoods in Toronto using the csv file and put it into a data frame.

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

	Postalcode	Borough	Neighborhood
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
8	M9A	Etobicoke	Islington Avenue
9	M1B	Scarborough	Malvern, Rouge
11	M3B	North York	Don Mills
12	M4B	East York	Parkview Hill, Woodbine Gardens
13	M5B	Downtown Toronto	Garden District, Ryerson

Next, we combine both the data frames i.e. adding the coordinate data to the original data frame.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

We use One Hot Encoding, use the neighbourhood to group data, and find out the top ten venues present in each neighbourhood.

	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
0	Berczy Park	Coffee Shop	Cocktail Bar	Beer Bar	Seafood Restaurant	Bakery	Restaurant	Cheese Shop	Café	Jazz Club	Shopping Mall
1	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Coffee Shop	Yoga Studio	Grocery Store	Pet Store	Performing Arts Venue	Nightclub	Italian Restaurant	Intersection
2	Business reply mail Processing Centre	Light Rail Station	Auto Workshop	Skate Park	Brewery	Smoke Shop	Spa	Farmers Market	Fast Food Restaurant	Burrito Place	Restaurant
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Sculpture Garden	Harbor / Marina	Bar	Rental Car Location	Plane	Coffee Shop	Boat or Ferry	Boutique
4	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Thai Restaurant	Bubble Tea Shop	Bar	Burger Joint	Department Store	Salad Place

The below table depicts the clustered data along with the top 10 most common venues in that cluster.

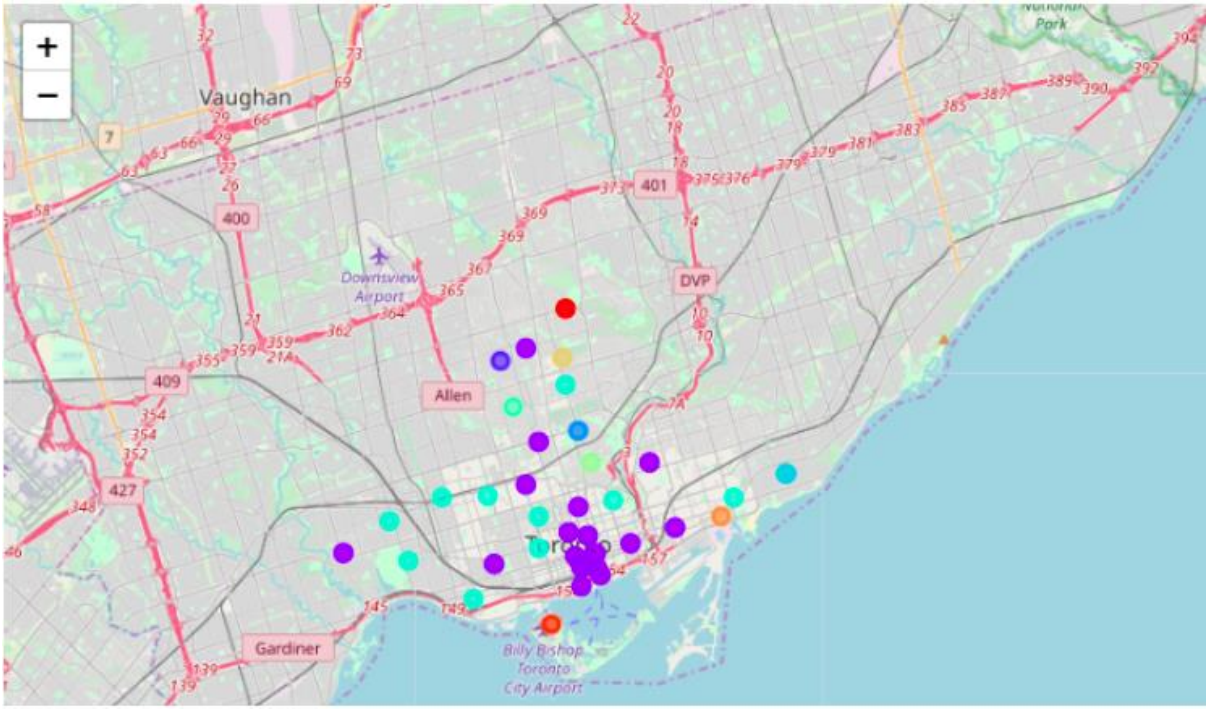
	Postalcode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6t Co
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	Pub	Yoga Studio	Department Store	
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	0	Greek Restaurant	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bookstore	Fu
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572	0	Park	Fast Food Restaurant	Gym	Pub	Brewery	Sa
43	M4M	East Toronto	Studio District	43.659526	-79.340923	0	Café	Coffee Shop	Gastropub	Bakery	Brewery	An Res
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	4	Park	Swim School	Bus Line	Yoga Studio	Dessert Shop	

Results

The below table depicts the clustered data along with the top 10 most common venues in that cluster.

	Cluster Labels	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 Co V
0	0	Berczy Park	Coffee Shop	Cocktail Bar	Beer Bar	Seafood Restaurant	Bakery	Restaurant	Cheese Shop	Café	Jazz Club	Sho
1	0	Brockton, Parkdale Village, Exhibition Place	Café	Breakfast Spot	Coffee Shop	Yoga Studio	Grocery Store	Pet Store	Performing Arts Venue	Nightclub	Italian Restaurant	Inters
2	0	Business reply mail Processing Centre	Light Rail Station	Auto Workshop	Skate Park	Brewery	Smoke Shop	Spa	Farmers Market	Fast Food Restaurant	Burrito Place	Resta
3	0	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Sculpture Garden	Harbor / Marina	Bar	Rental Car Location	Plane	Coffee Shop	Boat or Ferry	Bo
4	0	Central Bay Street	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Thai Restaurant	Bubble Tea Shop	Bar	Burger Joint	Department Store	Salad

Matplotlib and folium packages are used to visualize the clusters on a map of Toronto as depicted below:



E. Discussion

The intent with which analysis was carried out was to find out similar neighbourhoods for a person relocating within the city.

As we analyse the results section, we can analyse the clusters and see similar neighbourhoods in different parts of the city. For example, if we compare the different neighbourhoods clustered in cluster 2.

	Neighborhood	Borough	Cluster Labels	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
41	The Danforth West, Riverdale	East Toronto	1	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Furniture / Home Store	Pizza Place	Bookstore	Brewery
43	Studio District	East Toronto	1	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Yoga Studio	Comfort Food Restaurant	Seafood Restaurant
46	North Toronto West	Central Toronto	1	Coffee Shop	Sporting Goods Shop	Clothing Store	Burger Joint	Salon / Barbershop	Café	Restaurant	Rental Car Location

As seen in the table above, if a person wished to move from a suburb in East Toronto to Central Toronto. If a person's current location were in the Neighbourhood of Studio District in East Toronto, which has venues like cafes, bakeries and restaurants nearby, the person, would like to relocate to a

neighbourhood like North Toronto West in Central Toronto which also has venues like Coffee Shops and Restaurants. This is just one example of how our data analysis can help people relocate from one part of the city to another which similar to their current localities.

F. Conclusion

In a fast moving world, there are many real life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighbourhoods in Toronto based on the most common venues in those neighbourhoods. Similarly, data can also be used to solve other problems, which most people face in metropolitan cities.