The-battle-of-Neighbourhoods

The battle of neighbourhoods is a project that tries to choose a neighbourhood that is best for citing a restaurant business in either of the Manhattan or the Downtown Toronto of the New York City and the Toronto City respectively

A. Introduction:

A.1. Background

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 center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Brief history about Toronto: People have travelled through and inhabited the Toronto area, located on a broad sloping plateau interspersed with rivers, deep ravines, and urban forest, for more than 10,000 years. After the broadly disputed Toronto Purchase, when the Mississauga surrendered the area to the British Crown, the British established the town of **York** in 1793 and later designated it as the capital of Upper Canada. During the War of 1812, the town was the site of the Battle of York and suffered heavy damage by United States troops. **York** was renamed and incorporated in 1834, as the city of **Toronto**. It was designated as the capital of the province of Ontario in 1867 during Canadian Confederation.

A.2. Problem Description:

Now let me explain the context of this Capstone project through a scenario. Say you live on the west side of the city of Toronto in Canada. You love your neighborhood, mainly because of all the great amenities and other types of venues that exist in the neighborhood, such as gourmet fast food joints, pharmacies, parks, graduate schools and so on. Now say you receive a job offer from a great company on the other side of the city with great career prospects. However, given the far distance from your current place you unfortunately must move if you decide to accept the offer.

Wouldn't it be great if you are able to determine neighborhoods on the other side of the city that are the same as your current neighborhood, and if not perhaps similar neighborhoods that are at least closer to your new job?

A.3. Objective

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

A.4. Target Audience

This information provided by this report would be useful for people who are interested in relocating to a different part of the city and are interested in finding new neighborhoods that are highly similar to their existing neighborhood.

B. Data Description:

To consider the objective stated above, we can list the below data sources used for the analysis.

a) Toronto Neighborhood Data: The following Wikipedia page was scraped to pull out the necessary information: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The information obtained i.e. the table of postal codes was transformed into a pandas data frame for further analysis.

b) Coordinate data for each Neighborhood in Toronto: The following csv file gave us the geographical coordinates of each postal code: http://cocl.us/Geospatial_data

C. Methodology:

C.1. Scrape the Wikipedia page and gathering data into a Pandas dataframe

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 dataframe.

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

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

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Next, we combine both the dataframes i.e. adding the coordinate data to the original dataframe.

	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

C.2. Generating a map of Toronto and plotting the Neighborhood data on it

We first filter the data to find boroughs containing the word “Toronto”,

	PostalCode	Borough	Neighborhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

We then use the python **folium** library to visualize geographic details of Toronto and its boroughs. I created a map of Toronto with boroughs superimposed on top using the latitude and longitude values to get the visual as below:

C.3. Utilizing Foursquare API to explore the neighborhoods

Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them. We set the LIMIT parameter to **100**, which would limit the number of venues returned by the Foursquare API and the radius of 500 meter. Here is a head of the list of Nearby Venues for the first neighborhood i.e. The Beaches.

	name	categories	lat	lng
0	Glen Manor Ravine	Trail	43.676821	-79.293942
1	The Big Carrot Natural Food Market	Health Food Store	43.678879	-79.297734
2	Grover Pub and Grub	Pub	43.679181	-79.297215
3	Upper Beaches	Neighborhood	43.680563	-79.292869

We create a new function that will repeat the process above for all the neighborhoods in Toronto. This function will give us a list of all venues present in Toronto city. Here is a head () value of this dataframe.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

C.4. Analyze each neighborhood

We use One Hot Encoding, use the neighborhood to group data, and find out the top ten venues present in each 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
0	Adelaide, King, Richmond	Coffee Shop	Café	Thai Restaurant	Bar	Steakhouse	Hotel	Restaurant	Sushi Restaurant	Asian Restaurant	American Restaurant
1	Berczy Park	Coffee Shop	Cocktail Bar	Beer Bar	Bakery	Steakhouse	Seafood Restaurant	Café	Cheese Shop	Farmers Market	Clothing Store
2	Brockton, Exhibition Place, Parkdale Village	Café	Yoga Studio	Performing Arts Venue	Breakfast Spot	Coffee Shop	Italian Restaurant	Nightclub	Convenience Store	Intersection	Pet Store
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Yoga Studio	Garden Center	Pizza Place	Recording Studio	Restaurant	Burrito Place	Brewery	Skate Park	Farmers Market
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Service	Airport Terminal	Airport Lounge	Coffee Shop	Boat or Ferry	Boutique	Sculpture Garden	Bar	Plane	Airport Gate

We have some common venue categories in the neighborhoods. We use the unsupervised learning K-means algorithm to cluster the neighborhoods. K-Means algorithm is one of the most common method for clustering in unsupervised learning.

We use a k_**cluster** value of 11 to split the neighborhoods into 11 different clusters based on the similarity they have concerning the venues they contain.

D. Results:

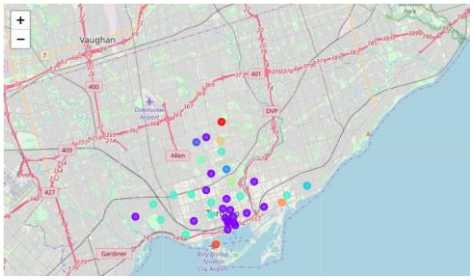
D.1. Adding the Cluster Labels to the Venue Data

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

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	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
The Beaches	43.676357	-79.293031	4	Trail	Health Food Store	Pub	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop
The Danforth West, Riverdale	43.679557	-79.352188	1	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Furniture / Home Store	Pizza Place	Bookstore	Brewery
The Beaches West, India Bazaar	43.668999	-79.315572	5	Park	Gym	Pub	Liquor Store	Board Shop	Fast Food Restaurant	Burger Joint	Fish & Chips Shop
Studio District	43.659526	-79.340923	1	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Yoga Studio	Comfort Food Restaurant	Seafood Restaurant
Lawrence Park	43.728020	-79.388790	0	Park	Swim School	Bus Line	Women's Store	Discount Store	Fast Food Restaurant	Farmers Market	Falafel Restaurant

D.2. Visualizing the resulting Clusters

We use the matplotlib and folium packages to visualize the clusters on a map of Toronto.



E. Discussion:

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

As we analyze the results section, we can analyze the clusters and see similar neighborhoods in different parts of the city. For example, if we compare the different neighborhoods 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 Neighborhood of Studio District in East Toronto, which has venues like cafes, bakeries and restaurants nearby, the person, would like to relocate to a neighborhood 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 neighborhoods in Toronto based on the most common venues in those neighborhoods. Similarly, data can also be used to solve other problems, which most people face in metropolitan cities.