A Recommender System for expats looking for a family friendly neighbourhood in Downtown Toronto

IBM Final Capstone Presentation

Anushka Patil

Background knowledge

According to The Economist, Toronto is placed in the world's top ten "most liveable cities". This is one of the reasons why Toronto is full of non-Canadians(expats). Toronto is known to have build a knowledge-based economy by hiring skilled professionals from overseas. This has also led to an extremely diverse population combined with the multiculturalism that stems from Toronto's location. According to the census of 2016, around 46.1% of residents in Toronto were born outside Canada.

Since Toronto is full of expats, it is implied that many families shift to Toronto every year and seek for a family- friendly neighbourhood to live in.

Family-friendly neighbourhoods are characterized by:

01 good entertainment venues nearby (theatres, spa, malls, bookstores etc.)

02 close to daily amenities(grocery store, gym, parlour, bank etc.)

03 amenities for children (school, park, food courts, university etc.)

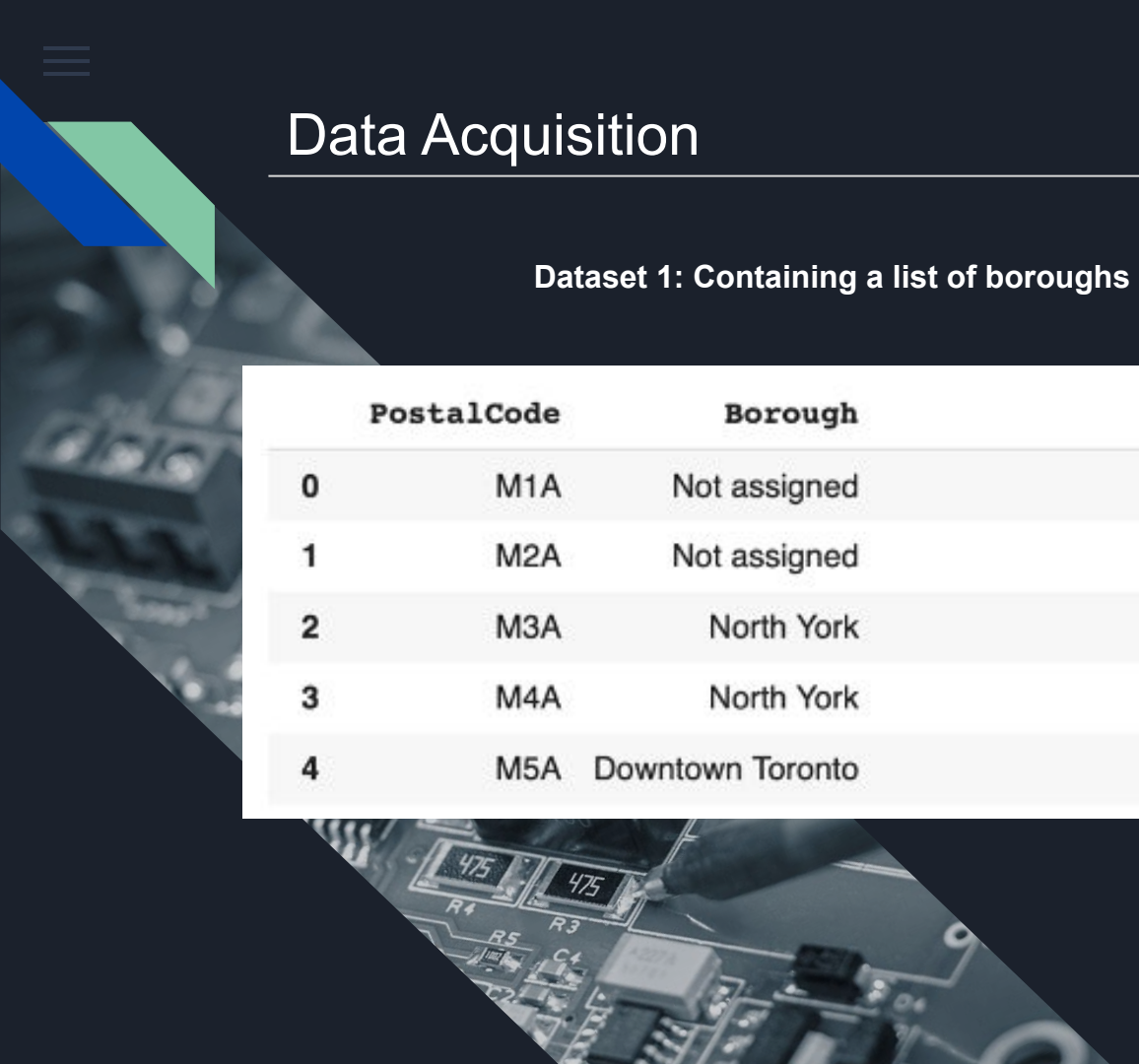
04 having a good transport network (bus station, metro, airport etc.)

05 far from nightlife places (bars, pubs etc.)

Data Acquisition

Dataset 1: Containing a list of boroughs of Canada

	PostalCode	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



Dataset 2: Containing latitudes and longitudes corresponding to different postal codes

	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

Now, this is merged with the first dataset to form a complete dataset, from which we can retrieve the neighbourhoods of the particular borough the expat wants to live in. In this project, we have used Downtown Toronto as the desired borough.

Dataset 3: Obtained using Foursquare

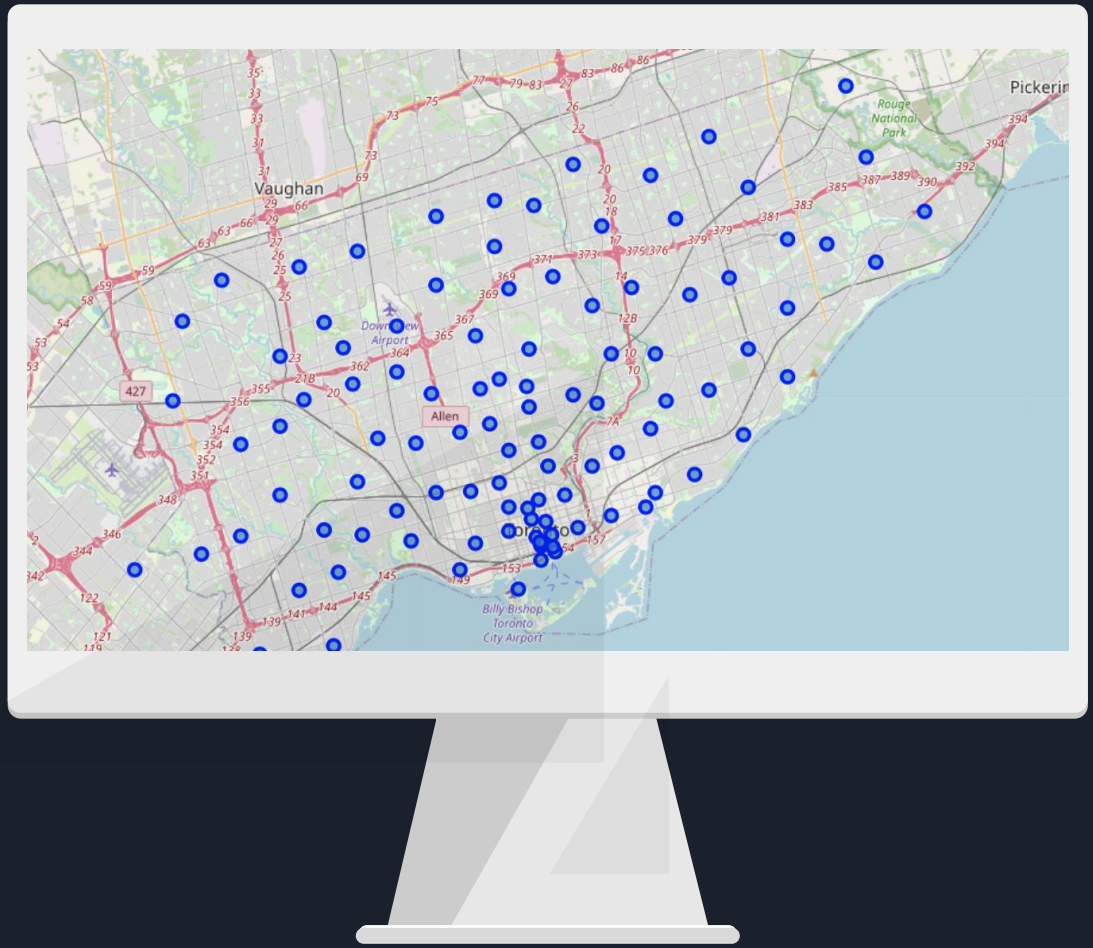
	Postal Code	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Summary	Venue Category	Distance
0	M4W	Rosedale	43.679563	-79.377529	Summerhill Market	This spot is popular	Grocery Store	764
1	M4W	Rosedale	43.679563	-79.377529	Black Camel	This spot is popular	BBQ Joint	994
2	M4W	Rosedale	43.679563	-79.377529	Craigleigh Gardens	This spot is popular	Park	505
3	M4W	Rosedale	43.679563	-79.377529	Toronto Lawn Tennis Club	This spot is popular	Athletics & Sports	896
4	M4W	Rosedale	43.679563	-79.377529	Pie Squared	This spot is popular	Pie Shop	826

We need two types of information from foursquare:

1. Basic: latitude, longitude of the venue to correctly identify the neighbourhood
2. Advance: Category in which the venue falls (for ex: Grocery Store, Park, Theatre etc.)

METHODOLOGY

103 neighborhoods of Canada.

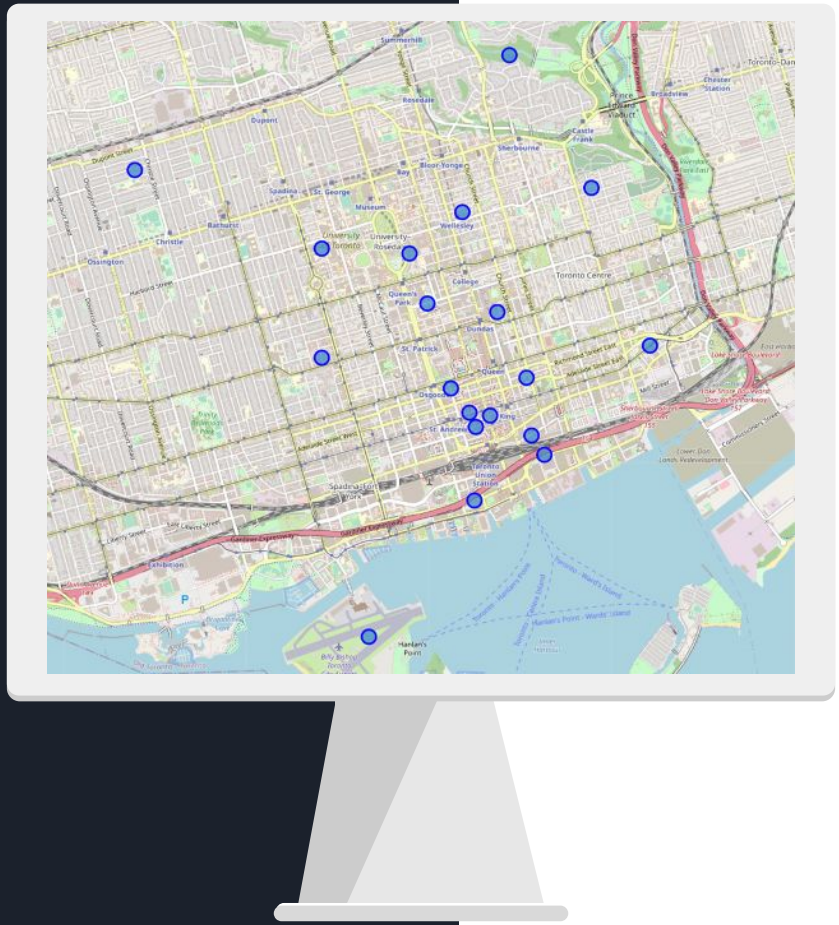


Neighborhoods

of

Downtown Toronto

are extracted.



List of Neighbourhoods in downtown Toronto:

‘Rosedale',
'St. James Town, Cabbagetown',
'Church and Wellesley',
'Regent Park, Harbourfront',
'Garden District, Ryerson',
'St. James Town',
'Berczy Park',
'Central Bay Street',
'Richmond, Adelaide, King',
'Harbourfront East, Union Station, Toronto Islands',
'Toronto Dominion Centre, Design Exchange',
'Commerce Court, Victoria Hotel',
'University of Toronto, Harbord',
'Kensington Market, Chinatown, Grange Park',
'CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport',
'Stn A PO Boxes',
'First Canadian Place, Underground city',
'Christie',
‘Queen's Park, Ontario Provincial Government’

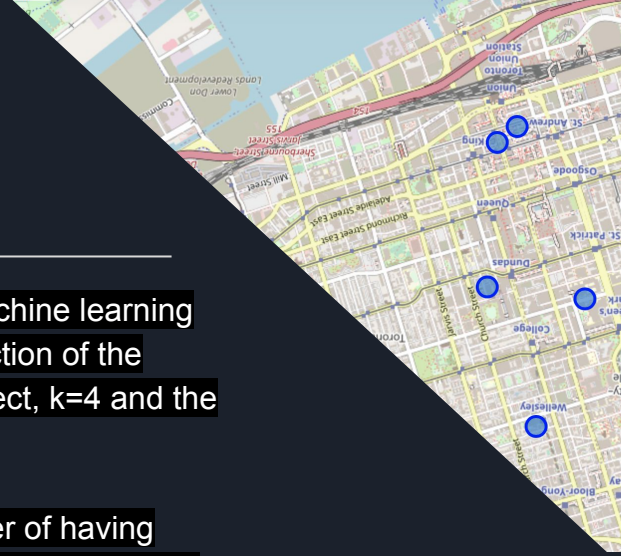
Extraction of unique Categories and One Hot Encoder

Using the third dataset obtained using FourSquare API, 200 unique categories were identified.
This category column is then one hot encoded. What one hot encoding does is, it takes a column which has categorical data, which has been label encoded and then splits the column into multiple columns each corresponding to a particular type of category. The numbers are replaced by 1s and 0s, depending on which column has what value. In our case, we'll get 200 new columns, one for each unique category.

After this, a list is created manually to select categories considered relevant to a family-friendly neighbourhood. The categories are chosen based on the characteristics mentioned in the business problem section. The selected categories are as follows:

'Grocery Store','Park','Bank','Playground','Sandwich Place','Candy Store','Metro Station','Diner','Restaurant','Bakery','Farm','Pet Store','Gift Shop','Garden','Dance Studio','Pool','Theater','Performing Arts Venue','Bookstore','Salon / Barbershop', 'Arts & Crafts Store','Ice Cream Shop','Historic Site','Supermarket','Yoga Studio','Health & Beauty Service','Furniture / Home Store','Video Store','Comic Shop','Clothing Store','Shopping Mall','Cosmetics Shop','Gym','Dog Run','Museum', 'Farmers Market','Chocolate Shop','Dessert Shop','Spa','Gym / Fitness Center','Shoe Store','Event Space','Food Truck','Gym Pool','Electronics Store','Skating Rink', 'Pharmacy','Music Venue','Department Store','Monument / Landmark','Art Museum','Poutine Place','Concert Hall','Church','Fountain','Tailor Shop','Basketball Stadium','Sporting Goods Shop','Beach','Lake','Train Station','University','Movie Theater','Aquarium','Baseball Stadium','Indie Movie Theater','Organic Grocery','Health Food Store','Music Store','Women's Store','Food Court','Optical Shop','Airport','Harbor / Marina','Sculpture Garden','Cupcake Shop','Rock Climbing Spot'.

K-Means Clustering
– A machine learning technique



K-means clustering is one of the simplest and a popular unsupervised machine learning algorithms. ‘k’ refers to the number of centroids and cluster refers to collection of the datapoints are aggregated together based on their similarities. In this project, k=4 and the datapoints are nothing but the neighborhoods

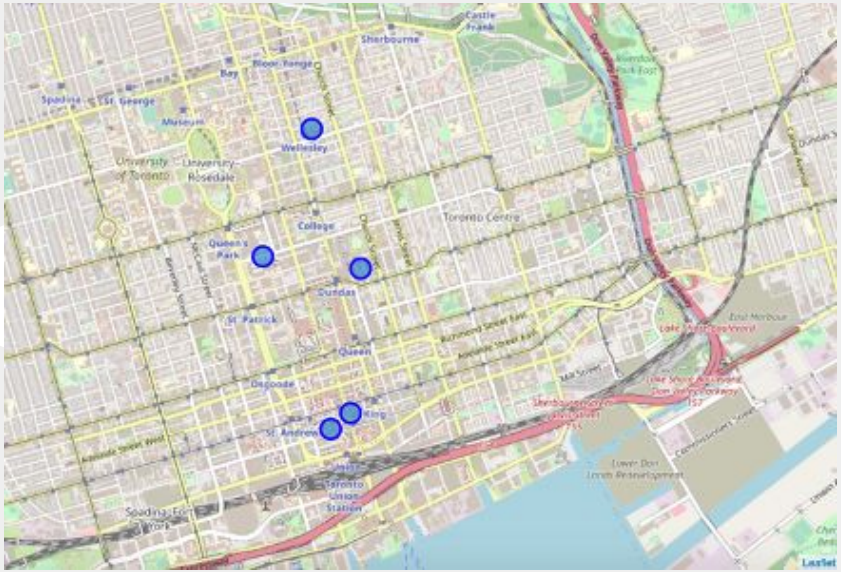
After performing k-means clustering, the groups are printed out in the order of having maximum sum. Here, maximum sum corresponds to the group of neighborhoods having maximum number of family friendly venues nearby.

Sum:	36.2	36	34.5	14.5
Group:	Group 2	Group 3	Group 4	Group 1

CONCLUSION

Group 2 is the set of most family friendly neighborhoods in the Downtown borough of Canada.

	Neighbourhood	Group
2	Church and Wellesley	2
4	Garden District, Ryerson	2
7	Central Bay Street	2
10	Toronto Dominion Centre, Design Exchange	2
11	Commerce Court, Victoria Hotel	2



In the future, the same project can be extended to accommodate a factor of safety using the crime dataset made available on Kaggle.

Thank you!