Settling Down in Toronto

Ramona Rajagopalan

Table of Contents

Introduction

While I'm not at all planning on starting a family anytime soon, I have always been curious about how my parents scouted a good location for them to settle down and raise children. I think I'm more interested because, at the time, tech and techniques like those we've been taught so far in this program were not as accessible to my parents like they are now.

It's easier now for us to do our research remotely—and now that we're in a pandemic, it's also safer. So, the main objective of this project is to explore neighbourhoods that might be of interest to anyone wanting to settle down and start a family.

Problem

Considering that Toronto has 100+ neighbourhoods, how can we narrow them down to the most viable options for places to settle down and start a family?

We'll be exploring Toronto, in particular, and for 2 reasons:

1. Canada is a country that I am genuinely interested in visiting as I've read and heard that quality of life is *relatively* better there than in a lot of places. So, it sounds like a good place to start exploring if you're planning to start a family soon and you have the means to go abroad (or you're already in Canada).
2. I've already got a dataset for it, which is very convenient and saves some time (admittedly, although I've got tons and tons of ideas for this project, I'm on a time crunch).

Who might be interested in this?

Great question! Thanks for asking. You might be interested if:

- you're interested in moving to Toronto. Depending on what you're looking for, this could help with your research.
- you're looking to settle down and start a family soon. This provides a framework that you could apply to your own research—or even give you some ideas about what you might want to consider in your research.
- you're also just starting out in your data science journey and want to work on a fun project similar to this. The techniques used in this notebook are among fundamentals that we've just been taught in this program, so you might find this useful.
- you're my parents, because you want to know that this program was worth it (huge thank you, if you're reading this, and also just want to clarify that no, I'm not planning to run off to Canada and start a family just yet—I might just get a dog but that's about it until I'm even remotely financially stable).

Data

For this project, we used the following data:

- This Wikipedia page with a list of neighbourhoods in Canada and their postal codes.
- This csv file of coordinates of different neighbourhoods in Canada. The file was provided by our instructors in the IBM Data Science Professional Certificate program.
- FourSquare Places API data

The original plan was to use the geocoder library to find the coordinates, but when I used it previously in the Segmenting and Clustering Neighborhoods in Toronto activity, it took a long time. Of course, that could be because of the script I used, but for now, we'll use the csv file for efficiency.

As for the Foursquare API data, we used this to get more insight into the venues (like shops, parks, etc.) that are within these neighbourhoods.

How we utilized this data:

We used the Wikipedia page to get a list of the neighbourhoods assigned to each postal code. Once we had this, we matched the neighbourhoods to their coordinates using the given csv file. After we matched them, we then used the Foursquare Places API to get a breakdown of the venues (specific, e.g. Tim Horton) and the venue categories (general, e.g. Cafe).

Once we had this breakdown, we proceeded to explore the data further in terms of certain minimum requirements we set (i.e. neighbourhood must have a hospital) in the next section.

Methodology

First, we applied **one-hot encoding** and grouped the data by neighbourhood. It was really cool to see the venue categories, because this added a level of uniformity in the keywords that we used to see which neighbourhoods met our minimum.

As a refresher (with additions), the following venue categories in no particular order are those we wanted to have in the neighbourhood at minimum:

- Hospital or Clinic
- School
- Drugstore
- Park
- Farmers Market or Grocery Store

Note: For the purpose of this project, we assumed that housing and source of income were either already taken care of or simply weren't variables we had to factor in.

After we shortlisted neighbourhoods, we clustered (**k-means**) them based on the most common venues (i.e. which venue categories have the highest count). This way, we also got a feel of the sort of people who might become neighbours (if there were many of these venues, then we considered that the target market would reside or often visit the neighbourhood) *and* find out what venues of interest other than our minimums might be accessible.

*For the k-means clustering, we applied k=5 because, although we shortlisted only 9 neighbourhoods, it was a good opportunity for us to better compare them.

Results and Discussion

In this section, we'll discuss the neighbourhoods we've shortlisted, as well as their clusters and how this research can be improved/developed further.

	Neighbourhood	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	R
0	Berczy Park	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	Business reply mail Processing Centre, South C...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Davisville	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Dufferin, Dovercourt Village	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Kensington Market, Chinatown, Grange Park	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
5	St. James Town	0	0	0	0	0	0	0	0	3	0	0	1	0	0	
6	St. James Town, Cabbagetown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	Stn A PO Boxes	0	0	0	0	0	0	0	0	1	1	0	2	0	0	
8	Thorncliffe Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Fig 1. DataFrame of Shortlisted Neighbourhoods (toronto_shortlisted)

One thing we can see straight away in Fig. 1 is that none of these shortlisted neighbourhoods have an airport. So, assuming you're not planning on traveling frequently, this would be great for settling down if we consider how much noise you'd hear from an airport.

Since we've already shortlisted these neighbourhoods based on our requirements, let's see the top 10 most common venues in each so that we can get a feel of their profiles.

	Cluster Labels	Neighbourhood	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	0	Berczy Park	Coffee Shop	Restaurant	Cheese Shop	Bakery	Cocktail Bar	Beer Bar	Farmers Market	Seafood Restaurant	Sandwich Place	Breakfast Spot
1	3	Business reply mail Processing Centre, South C...	Gym / Fitness Center	Garden Center	Farmers Market	Fast Food Restaurant	Light Rail Station	Burrito Place	Butcher	Restaurant	Recording Studio	Brewery
2	4	Davisville	Sandwich Place	Dessert Shop	Pizza Place	Gym	Coffee Shop	Café	Italian Restaurant	Sushi Restaurant	Brewery	Greek Restaurant
3	2	Dufferin, Dovercourt Village	Pharmacy	Bakery	Music Venue	Middle Eastern Restaurant	Café	Brewery	Bar	Supermarket	Bank	Park
4	0	Kensington Market, Chinatown, Grange Park	Café	Mexican Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Bar	Vietnamese Restaurant	Dessert Shop	Dumpling Restaurant	Burger Joint	Bakery
5	0	St. James Town	Coffee Shop	Café	Restaurant	Cocktail Bar	Beer Bar	Gastropub	American Restaurant	Seafood Restaurant	Farmers Market	Hotel
6	0	St. James Town, Cabbagetown	Coffee Shop	Restaurant	Café	Pizza Place	Chinese Restaurant	Italian Restaurant	Market	Pub	Bakery	Park
7	0	Stn A PO Boxes	Coffee Shop	Italian Restaurant	Restaurant	Pub	Seafood Restaurant	Café	Beer Bar	Hotel	Japanese Restaurant	Breakfast Spot
8	1	Thorncliffe Park	Sandwich Place	Indian Restaurant	Yoga Studio	Grocery Store	Burger Joint	Bus Line	Restaurant	Coffee Shop	Pizza Place	Pharmacy

Fig 2. DataFrame of Top 10 Most Common Venues in Shortlisted Neighbourhoods (neighborhoods_venues_sorted)

Just from looking at Fig. 2, we can see that the majority of the most common venues are places where people go to eat. But now we want to explore the clusters.

Cluster 1

	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	9th Most Common Venue	10th Most Common Venue
0	Downtown Toronto	0.0	Coffee Shop	Café	Restaurant	Cocktail Bar	Beer Bar	Gastropub	American Restaurant	Seafood Restaurant	Farmers Market	Hotel
1	Downtown Toronto	0.0	Coffee Shop	Restaurant	Cheese Shop	Bakery	Cocktail Bar	Beer Bar	Farmers Market	Seafood Restaurant	Sandwich Place	Breakfast Spot
2	Downtown Toronto	0.0	Café	Mexican Restaurant	Vegetarian / Vegan Restaurant	Coffee Shop	Bar	Vietnamese Restaurant	Dessert Shop	Dumpling Restaurant	Burger Joint	Bakery
3	Downtown Toronto	0.0	Coffee Shop	Italian Restaurant	Restaurant	Pub	Seafood Restaurant	Café	Beer Bar	Hotel	Japanese Restaurant	Breakfast Spot
4	Downtown Toronto	0.0	Coffee Shop	Restaurant	Café	Pizza Place	Chinese Restaurant	Italian Restaurant	Market	Pub	Bakery	Park

Fig 3. Cluster 1, Downtown Toronto

We see here that all 5 are in Downtown Toronto and that, for all of them, their 1st most common venue is some sort of coffee shop.

Considering that the majority of the most common venues here are eateries and there are several bars and pubs, as well, it's possible that people who reside here are relatively young and quite busy. We're also making this assumption based on the fact that nowhere in these top 10 is a pharmacy, drugstore, or hospital.

Cluster 2

	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	9th Most Common Venue	10th Most Common Venue
0	East York	1.0	Sandwich Place	Indian Restaurant	Yoga Studio	Grocery Store	Burger Joint	Bus Line	Restaurant	Coffee Shop	Pizza Place	Pharmacy

Fig 4. Cluster 2, East York

Since there were only 9 neighbourhoods on our shortlist, it makes sense that there would be some clusters with only 1 neighbourhood.

With the bus line being the 6th most common venue, it's possible people here tend to travel quite a bit. This group of venues I find interesting, because it's a mix of places like the Pizza Place and those like the Yoga Studio.

I'm not quite sure what to make of it with such limited data, but considering the Yoga Studio comes in as 3rd most common venue, I think it's rational to posit that people living here are relatively health-conscious.

Cluster 3

	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	9th Most Common Venue	10th Most Common Venue
0	West Toronto	2.0	Pharmacy	Bakery	Music Venue	Middle Eastern Restaurant	Café	Brewery	Bar	Supermarket	Bank	Park

Fig 5. Cluster 3, West Toronto

Interestingly, the most common venue here is the pharmacy. It's possible that there are a lot of elderly people in this neighbourhood, but it's also possible that there are many pharmacies for other reasons (e.g. clinic or hospital in a nearby neighbourhood that didn't make it to our shortlist).

Cluster 4

	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	9th Most Common Venue	10th Most Common Venue
0	East Toronto	3.0	Gym / Fitness Center	Garden Center	Farmers Market	Fast Food Restaurant	Light Rail Station	Burrito Place	Butcher	Restaurant	Recording Studio	Brewery

Fig 6. Cluster 4, East Toronto

So, this has a gym or fitness center as *the* most common venue and the Farmers Market as the 3rd most common venue, so perhaps the people living here are quite health-conscious too, just like in the Sandwich Cluster. I say "perhaps" because the 4th most common venue is Fast Food Restaurant.

Another thing to note here is that the 5th most common venue is the Light Rail Station. It might be a bit noisy here because of that.

Cluster 5

	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	9th Most Common Venue	10th Most Common Venue
0	Central Toronto	4.0	Sandwich Place	Dessert Shop	Pizza Place	Gym	Coffee Shop	Café	Italian Restaurant	Sushi Restaurant	Brewery	Greek Restaurant

Fig 7. Cluster 5, Central Toronto

This cluster, just like Cluster 1, has a lot of eateries for common venues. I'm wondering if the cost of going out to eat might be a bit less here than at the other clusters, simply because of the competition. But, perhaps not. For all we know, the restaurants are a chain and there isn't much competition because they're all different cuisines.

Either way, if you can afford it, you've got a *lot* of options for dining.

Recommendations and Notes

It has become extremely apparent that we've been working with a very limited set of data. What I mean is that the assumptions we've been making have been based purely on how many different venue categories are present in each of our clusters.

The problem with this is that we didn't really get a 360-view of these neighbourhoods. Here is a list of some of the information that anyone wanting to further this might want to consider:

- data on elementary schools, high schools, and daycare centers
- data on housing options in these neighbourhoods
- cost of living in these clusters or even in the boroughs
- job options
- population density
- crime rate

Such information would definitely help in getting better insight into these neighbourhoods, especially in terms of shortlisting which would be most conducive for settling down and raising and family.

Conclusion

The main objective of this project was to narrow down the neighbourhoods of Toronto and figure out which ones might be viable options for those looking to raise a family.

We did some web scraping to get the list of neighbourhoods and matched them to their coordinates from a csv file provided by the course instructors. We then utilized the FourSquare Places API in order to get a summary of sorts of the venues within these places.

Next, we shortlisted some neighbourhoods based on minimum requirements, which are the following in no particular order:

- Hospital or Clinic
- School
- Drugstore
- Park
- Farmers Market or Grocery Store

Although we had set these to be our minimum requirements, we later found that it was possible the venue list we had was incomplete (i.e. no elementary or highschools) and just in general, we were working with pretty limited data so we shortlisted based on which neighbourhoods met at least 3 of our requirements.

Finally, we clustered our shortlisted neighbourhoods in order to find out which were from the same borough and in order for us to see any commonalities in neighbourhoods within each cluster.

Based on our results and discussion, it seems like it would be worthwhile to check out Cluster 1 and Cluster 3 for the following reasons:

- Compared to the other three clusters, these two might be quieter in the sense that neither have a railway station in their top 10 most common venues.

- Cluster 1 neighbourhoods have a *lot* of food options and a *lot* of coffee, which parents of young children (and actually anyone, come to think of it) might find quite convenient.

- Cluster 3 has a bit more variety in its top 10 most common venues. It seems to have quite a few pharmacies and supermarkets, which is handy and several parks, which would be great to relax at.

Again, I refer to the Results and Discussion section for some notes and recommendations on how to improve this research. It would be a lot more insightful to also factor in the sort of data listed there.

THANK YOU!