Exploring Segmentation of Toronto Neighborhoods for Prospective Tenants

Coursera Capstone – Aug 2020

Ameya Sohoni *ameyasohoni@outlook.com*

1 Problem Introduction

The city of Toronto is the most populous city in Canada with a recorded population of 2,731,571 in 2016. In addition to being the most populous city and the economic capital of Canada, Toronto is also home to the largest migrant population of any Canadian city. In 2019, Toronto saw an influx of 313,580 immigrants from various corners of the world. For newcomers to a metropolitan area, an essential decision is finding the right place to stay. This decision can be influenced by a plethora of factors including rent rates, safety, convenience, transit to name a few. While individual preferences vary on what weight to place on which parameter, it would be greatly useful to have a scientific segmentation of neighborhoods based on key deciding features, to facilitate the decision.

In this report, I have attempted to segment Toronto neighborhoods based on a few key features which would allow prospective tenants to make an informed decision on which location to select. While this analysis is geared to suit a decision on place of stay, the findings are likely suitable for other scenarios such as places of business and erection of public utilities as well.

2 Dataset and Feature Selection

The data required for this analysis was extracted from multiple sources based on the required feature-set. To make a decision on place of stay the key independent parameters considered are:

1. Average rent (for 2-bedroom apartment) - CA$/month
2. Crime rates – Major Crimes/100k population
3. Convenience – Availability of shops, offices, stores
4. Cultural/ Community – Vicinity to movie halls, parks, outdoor recreation venues
5. Social Life – Vicinity to dining and nightlife venues
6. Education – Availability of schools, college/ university
7. Transit – Vicinity to travel options

The primary list of neighborhoods was extracted from the Toronto police website [1] which segregates the city into 140 neighborhoods. While there are other ways to divide the city into neighborhoods, this particular division is a standard method used by the city of Toronto which provided a very convenient and unbiased dataset. The Toronto police website also provided data for crime rates (see pt. 2 above).

Data for average rent by neighborhood was taken from zumper.com [2] which is a popular portal for rental seekers. However, the neighborhoods used by zumper (let's call them rent neighborhoods) do not match with the standard neighborhood divisions (let's call them base neighborhoods) considered in my primary dataset. As a result, I had to make an approximation using lat/long coordinates. I did this by:

1. Checking which rent neighborhoods fall within the base neighborhood and using the corresponding rent or mean rent (if more than one fall within)
2. For the base neighborhoods within which NO rent neighborhood falls, the nearest rent neighborhood to the base neighborhood center is considered

While this approach is not perfect, in relative terms it provides a good approximation of the average rent for each neighborhood, considering the lack of freely available data.

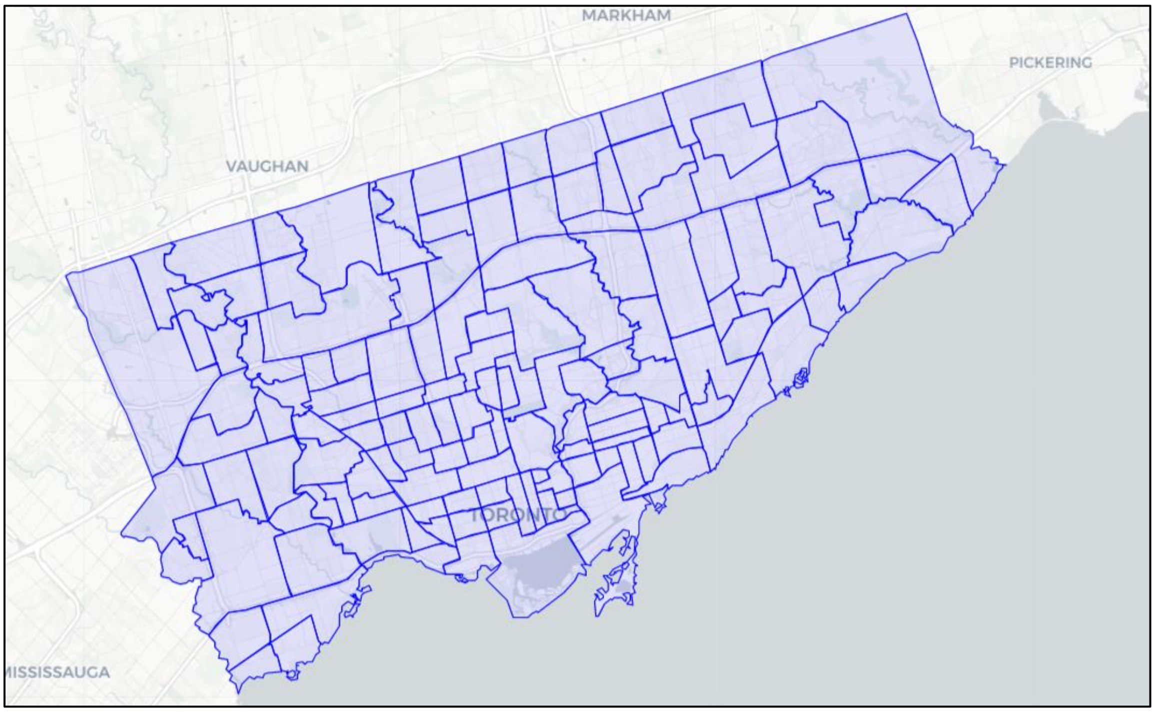

Fig 1: Base Neighborhoods in Toronto

For the remaining parameters *viz. Convenience, Cultural/ Community, Social Life, Education & Transit*, I used the Foursquare.com Places API to extract (upto 100) top recommended venues around the neighborhood center. This will provide a representative sample of the neighborhood's accessibility to similar venues.

The original results of the API GET query to Foursquare however, grouped venues in too many disparate categories. For example, a venue such as a restaurant can be categorized as Afghan Restaurant or Italian Restaurant or Pizza Place. While this is useful in knowing granular details about the venue; it is counter-productive if you are trying to segment the data into decision support parameters such as in our case.

To overcome this problem, I have used Foursquare's own categorization of its venues. This grouping can be queried with a GET request to the API [3]. This grouping reduces the number of venue categories from hundreds to a more manageable ten.

Feature Category	
Arts & Entertainment	66
College & University	38
Event	12
Food	357
Nightlife Spot	26
Outdoors & Recreation	107
Professional & Other Places	110
Residence	5
Shop & Service	178
Travel & Transport	56

Table 1: Truncated list of categories from Foursquare's API call

With this reduced list we can clearly see which categories fit within our pre-defined parameters.

1. **Culture and Community** - Arts & Entertainment, Event, Outdoors & Recreation
2. **Education** - College & University
3. **Socializing** - Food, Nightlife Spot
4. **Convenience** - Professional & Other Places, Residence, Shop & Service
5. **Transit** - Travel & Transport

This gives us our final set of features and the associated venues.

Neighborhood	Venue	Venue Latitude	Venue Longitude	Category
Yonge-St.Clair	The Market By Longo's	43.686711	-79.399536	Convenience
Yonge-St.Clair	The Bagel House	43.687374	-79.393696	Socializing
Yonge-St.Clair	Union Social Eatery	43.687895	-79.394916	Socializing
Yonge-St.Clair	LCBO	43.686991	-79.399238	Convenience
Yonge-St.Clair	Daeco Sushi	43.687838	-79.395652	Socializing
...	...	...	...	...
Briar Hill-Belgravia	Tim Hortons	43.701181	-79.452519	Socializing
Briar Hill-Belgravia	Il vagabondo	43.701480	-79.452443	Socializing
Briar Hill-Belgravia	Country Style	43.702225	-79.452641	Socializing
Briar Hill-Belgravia	Jason's No Frills	43.694789	-79.453038	Convenience
Briar Hill-Belgravia	Sage Wellness Boutique B&B/Hostel	43.695391	-79.450775	Transit

Table 2: Sample table of venues from GET request to Foursquare API call