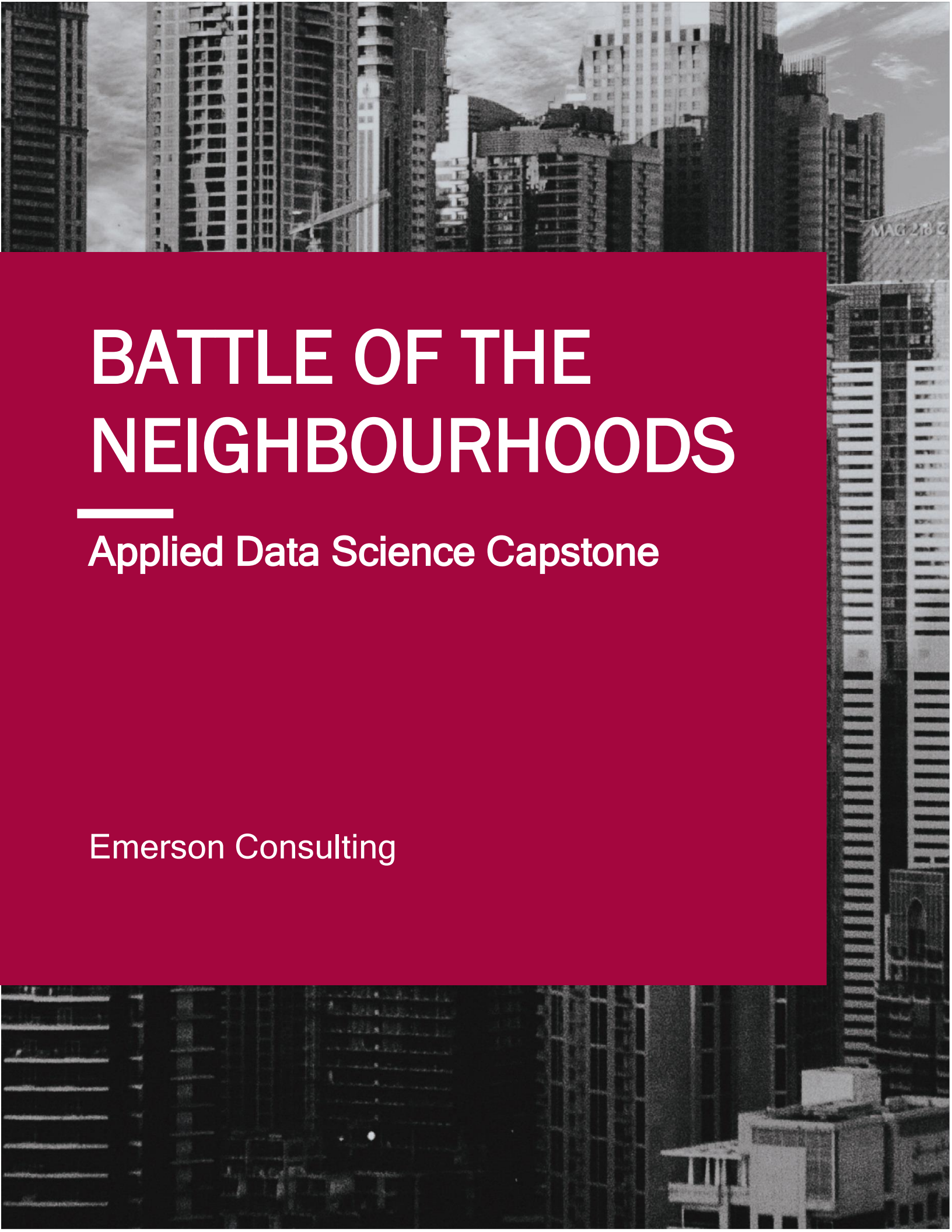


BATTLE OF THE NEIGHBOURHOODS

Applied Data Science Capstone

Emerson Consulting

Table of Contents

Introduction

The purpose of this project is to determine the best area in Toronto, Canada for the client - a residential developer - to build a rental building featuring 3-bedroom units geared towards young families.

Toronto is the largest city in Canada, and is considered to be one of the most multicultural cities in the world [1]. In fact, in 2019 it was in the top 10 'Most Livable Cities' as reported by The Economist [2]. With such high praises also comes a high cost of living – in 2018, the average price among all home types in Toronto was $835,422 [3].

The developer is aware that more and more families are moving out of the city when they have children, due to the rising cost of home ownership, and the shortage of family-sized rental options. However, Toronto has a wealth of schools, libraries, and childcare that are optimal for child-rearing. The developer aims to empower young families - who cannot afford to purchase property - to live in one of these optimal locations.

Thus, the developer wishes to find a shortlist of areas that contain the best mix of schools, libraries, and childcare, so that they may begin to survey the potential sites.

Data

The data used to aid my recommendations to the developer will be obtained using Foursquare API and the City of Toronto's Open Data Catalogue.

The Open Data Catalogue will be used to determine the boundaries of Toronto's neighbourhoods.

The Foursquare API will be used to explore the following:

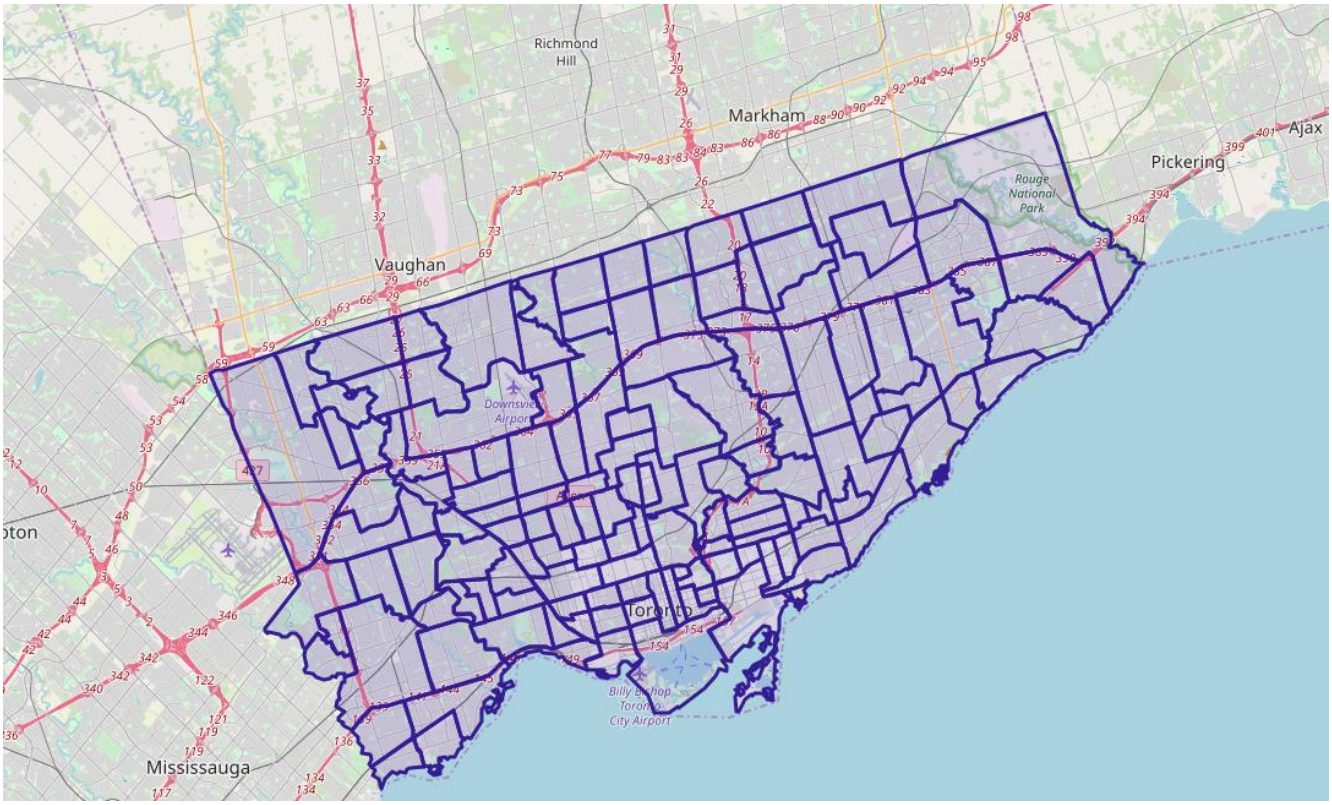
- The number of schools in a neighbourhood
- The number of libraries in a neighbourhood
- The number of playgrounds in a neighbourhood

These factors will be normalized and summed to create a raw score for each neighbourhood.

Methodology

To answer the question of which neighbourhoods will be most suitable for young families in Toronto, we must first explore the city itself. The City of Toronto is quite large, covering 641 sq.km [3].

Toronto is commonly divided into 140 unique neighbourhoods [4]. Data covering these 140 neighbourhoods can be downloaded from the City of Toronto's Open Data Catalogue as a .geoJSON file. This file contains coordinates of the boundaries of each neighbourhood, as well as the neighbourhood name, and geographical centre. Using this data to create a folium map, the way the city is divided becomes clear.



However, for further analysis, only the geographic centre of each neighbourhood will be utilized.

To explore the venues of interest in each neighbourhood, the Foursquare API will be used. The radius in which each search takes place will be set to 1000m, as that is a reasonable 10 to15 minute walk. The three venues that will specifically be noted are schools, libraries, and child care centres. A dataframe will be created with the count of each instance of each of these types venue in each neighbourhood.

	Neighbourhood	Latitude	Longitude	School	Library	Childcare
0	Wychwood	43.676919	-79.425515	14	3	4
1	Yonge-Eglinton	43.704689	-79.403590	33	5	0
2	Yonge-St.Clair	43.687859	-79.397871	20	3	0
3	York University Heights	43.765736	-79.488883	3	0	0
4	Yorkdale-Glen Park	43.714672	-79.457108	12	0	1
5	Lambton Baby Point	43.657420	-79.496045	10	1	1
6	Lansing-Westgate	43.754271	-79.424748	3	0	0
7	Lawrence Park North	43.730060	-79.403978	21	2	0
8	Lawrence Park South	43.717212	-79.406039	17	3	0
9	Leaside-Bennington	43.703797	-79.366072	13	3	3
10	Little Portugal	43.647536	-79.430323	31	3	3
11	Long Branch	43.592362	-79.533345	3	1	2
12	Malvern	43.803658	-79.222517	12	1	0
13	Maple Leaf	43.715574	-79.480758	3	2	0
14	Markland Wood	43.633542	-79.573432	4	0	0
15	Milliken	43.820691	-79.275009	11	1	0
16	Mimico (includes Humber Bay Shores)	43.615924	-79.500137	10	1	0
17	Morningside	43.782399	-79.207041	5	0	0
18	Moss Park	43.656518	-79.367297	33	8	3

The above chart is a snippet of the generated data. From this data, we can start to get an idea of how appealing each of these neighbourhoods would be to a young family. For instance, Moss Park has 33 schools, 8 libraries, and 3 childcare centres. At a glance, this seems more appealing than Lansing-Westgate, which has 3 schools and no libraries nor childcare.

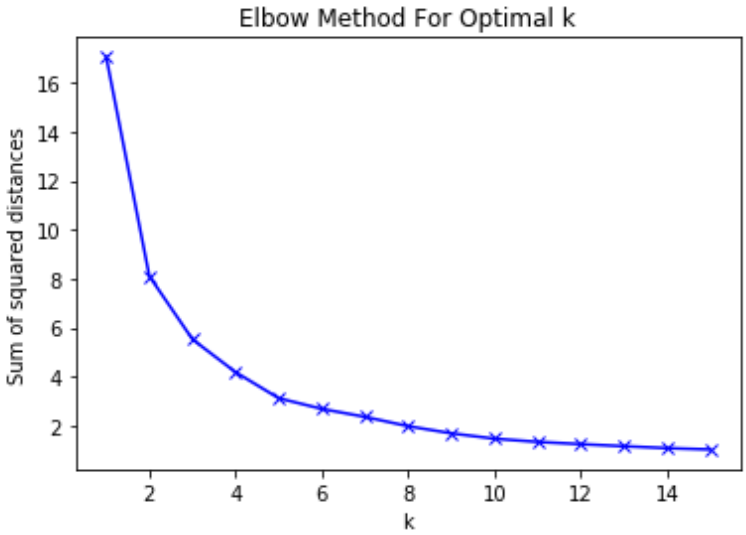
Using this generated data, k-means clustering can be performed to group similar neighbourhoods together. These clusters can potentially be used to classify neighbourhoods as ‘appealing’ or ‘not appealing’.

However, before this k-means clustering is performed, the data first needs to be normalized. This can be achieved by using the sklearn library’s ‘preprocessing’ function to scale the data.

	Neighbourhood	Latitude	Longitude	School	Library	Childcare
0	Wychwood	43.676919	-79.425515	0.285714	0.200000	0.666667
1	Yonge-Eglinton	43.704689	-79.403590	0.673469	0.333333	0.000000
2	Yonge-St.Clair	43.687859	-79.397871	0.408163	0.200000	0.000000
3	York University Heights	43.765736	-79.488883	0.061224	0.000000	0.000000
4	Yorkdale-Glen Park	43.714672	-79.457108	0.244898	0.000000	0.166667

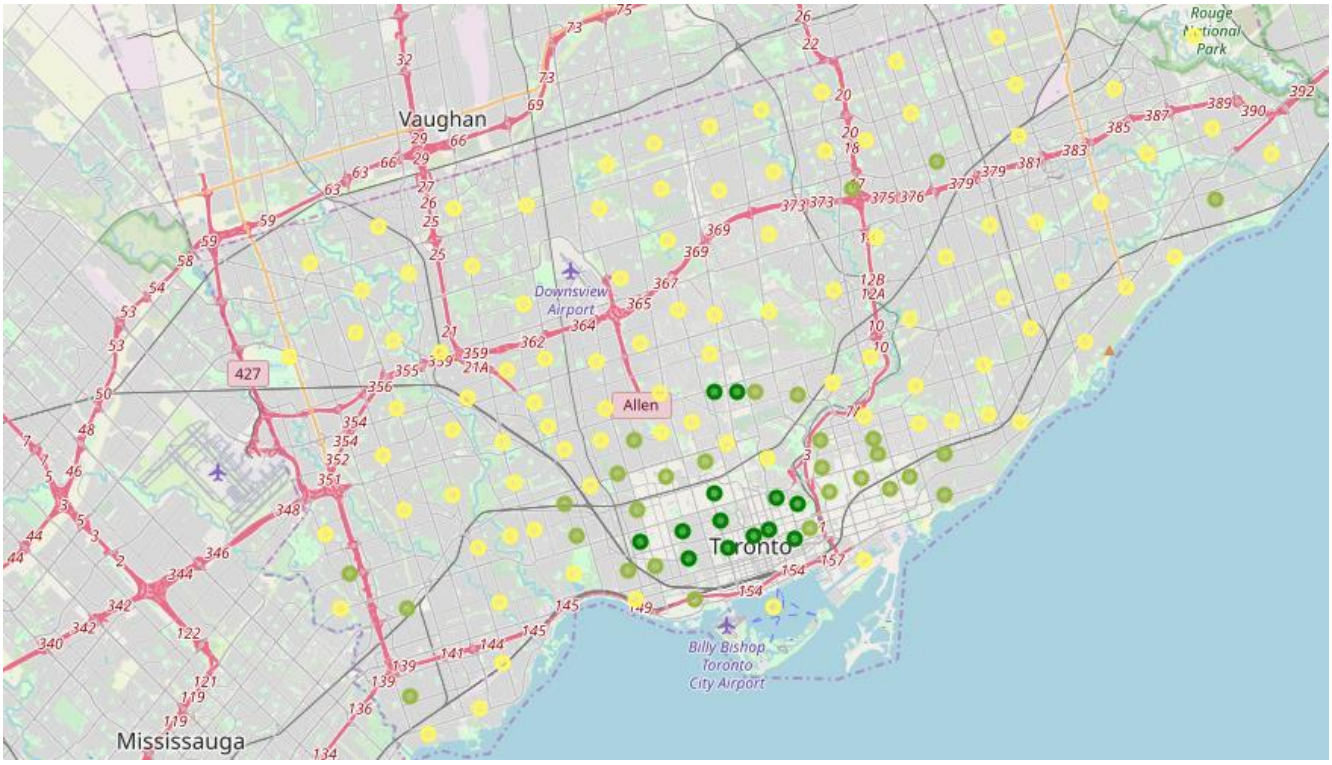
After normalization, the number of each feature is represented as a function of the minimum and maximum existing values for each feature. To start, the k-means clustering will be run using a target of 3 clusters. Once it has been run, the appropriateness of using 3 clusters for the data can be explored using the ‘elbow method’.

To use the elbow method, matplotlib is imported and utilized to create a plot of the sum of squared distances between each cluster at each given k (number of clusters) value.



The value that occurs on the plot at the 'elbow' – that is, the point directly after which the sum of squared distances stops decreasing in a linear fashion – is the best k-value. In this case, the best k-value is indeed 3, so 3 clusters are the most appropriate for this data-set.

These clusters can be visualized on the map using folium, as below.



The dataset can then be sliced into 3 – the most ideal neighbourhoods, moderately ideal neighbourhoods, and unideal neighbourhoods. Saving the dataframe of most ideal neighbourhoods, this subset can be further probed to develop a list of the top 5 neighbourhoods for the developer to consider building in.

To rank these neighbourhoods in the ‘most ideal’ cluster, a sum of the scores of individual categories can be tallied. The data can then be sorted by this total score, and ranked from 1 to 5.

Results

The final ranking of the neighbourhoods from the 'most ideal' cluster found that the top five neighbourhoods were Bay Street Corridor, University, Church-Yonge Corridor, Kensington-Chinatown, and North St. James Town.

	Neighbourhood	Total_Score
0	Bay Street Corridor	2.170068
1	University	2.133333
2	Church-Yonge Corridor	2.090476
3	Kensington-Chinatown	2.059184
4	North St.James Town	1.773469

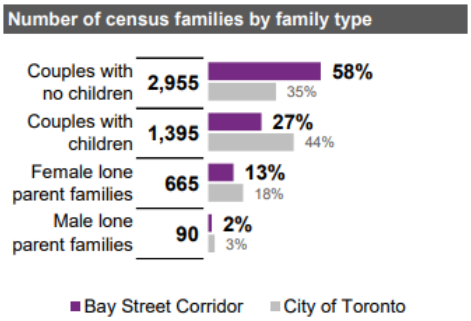
When this Top 5 list is transposed back to view the individual venues around each neighbourhood, it becomes quite clear why they were chosen. The high number of schools, libraries, and childcare options make them likely to be more attractive to a young family.

Neighbourhood	Latitude	Longitude	School	Library	Childcare
Bay Street Corridor	43.657511	-79.385721	41	15	2
University	43.662506	-79.401180	49	12	2
Church-Yonge Corridor	43.659649	-79.379017	42	11	3
Kensington-Chinatown	43.653554	-79.397240	47	14	1
North St.James Town	43.669623	-79.375247	33	9	3

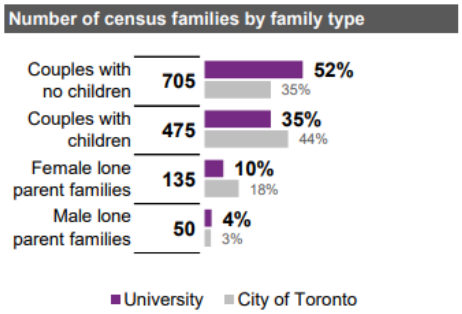
Discussion

When one thinks of a young, urban-dwelling couple deciding to start a family, one often imagines that they must move out of the city in order to access the necessary services. However, these results show that is clearly not the case – all 5 of the recommended neighbourhoods are well within the core of Toronto.

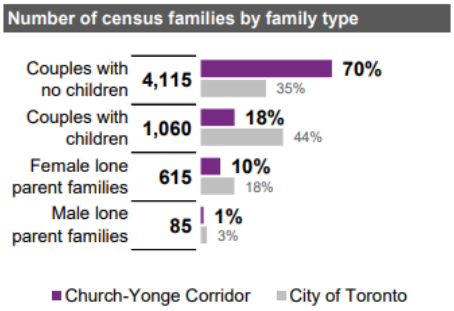
When the demographic information from the 2016 census is examined for these 5 neighbourhoods, most are occupied primarily by couples with no children. Given the number of schools, libraries, and child care options in these neighbourhoods compared to the proportion of families, it is likely that some families currently commute into these neighbourhoods. This may present as an opportunity for the developer to convince these young families to move into the neighbourhood if suitable housing is built.



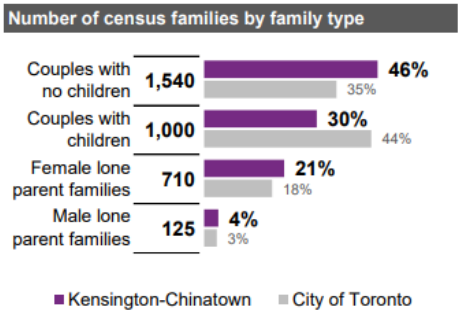
1. Bay Street Corridor [5]



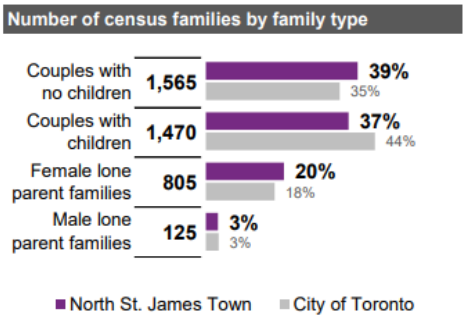
2. University [5]



3. Church-Yonge Corridor [5]



4. Kensington-Chinatown [5]



5. North St. James Town [5]

Conclusion

This study analyzed the 140 neighbourhoods of the City of Toronto to determine which would be the most appealing for a developer to build young family-oriented rental housing. While in this case the top five suggested neighbourhoods were built on a clustered analysis of the presence of schools, libraries, and childcare options, these venue variables could easily be changed or added to if the developer chooses to focus in on a more specific cross-section of the demographic.

Of course, there is significant further analysis that could be done for this developer. The income levels, land cost, current rental vacancy rate, and a plethora of more complicated factors could be taken into consideration to create a more nuanced list of suggested neighbourhoods.

References

[1] https://www.toronto.ca/community-people/moving-to-toronto/about-toronto/
[2] https://www.eiu.com/public/topical_report.aspx?campaignid=liveability2019
[3] https://www.toronto.ca/city-government/data-research-maps/toronto-at-a-glance/
[4] https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/
[5] https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/