Identifying Neighbourhoods in Toronto based on Childcare Support

J

December 2020

Project: Applied Data Science Capstone

# Introduction

Young families who have newborn children or are expecting a child often decide to move into a larger living space in a family friendly neighbourhood to accommodate their growing family. This often involves buying or moving into a new home in a neighbourhood that has services and facilities to support. One service that is especially important for young families to consider when moving to a new neighbourhood is access to childcare services.

Childcare services provide parents an option to return to work and provide infants, toddlers, and young children an avenue to learn, socialize, and play. Having access to these services is critical for families to be able to pursue their careers and provide for their family, while ensuring their young children can develop appropriately.

Understanding the availability and capacity of childcare services can be an important factor in determining what neighbourhood to live in. For Toronto, childcare service capacity is limited and varies considerably by neighbourhood. Reserving space in childcare centres of choice often requires parents to enroll their child when they know they have conceived. This means parents with young children or who are in the process of family planning should take childcare service accessibility and quality into consideration before moving to a new neighbourhood, even if a child is still in the future.

The goal of this project is to identify Toronto neighbourhoods based on the capacity of childcare services available. This will support in decision making for young families planning to move to accommodate their growing family. Additionally, it can be used to identify areas where childcare services are underrepresented.

# Data

## 2.1 Sources

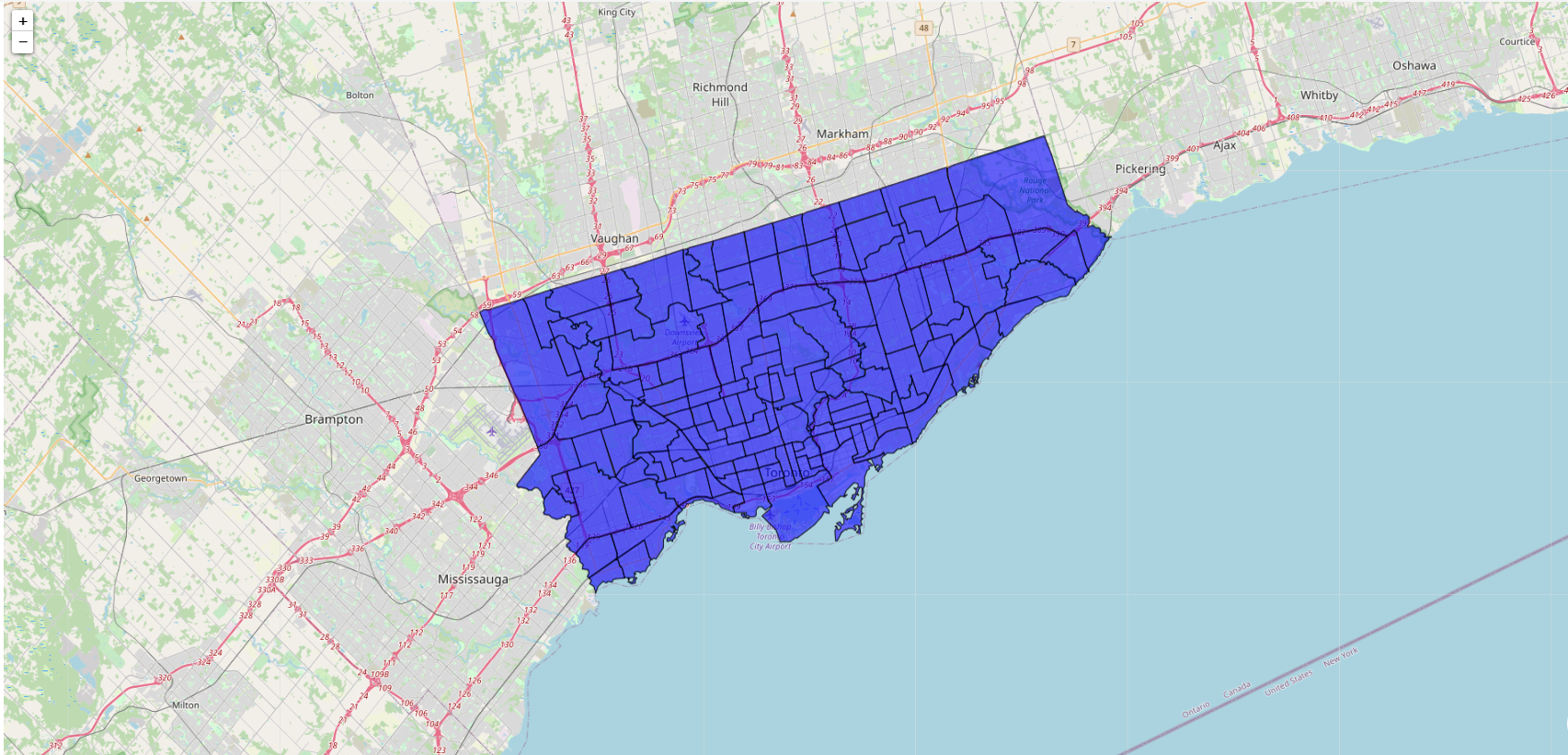
Neighbourhood data on the City of Toronto will be accessed using the [City of Toronto’s Open Data Portal](https://open.toronto.ca) using the [Neighbourhoods](https://open.toronto.ca/dataset/neighbourhoods/) data set. This data set includes the boundaries of each neighbourhood and will be used to identify the neighbourhood each childcare facility belongs to. The data will be imported both as a GEOJSON file to be used to visualize the analysis on a choropleth map once finalized. Additionally, the neighbourhood data will be extracted as a CSV to be used in data cleaning for other parts of the analysis.

Childcare data will be pulled from the City of Toronto’s Open Data Portal using the [Licensed Child Care Centres](https://open.toronto.ca/dataset/licensed-child-care-centres/) data report. In this data lists all licensed childcare service facilities in Toronto and includes information on their location, geo coordinates, capacity by age group, as long with other descriptive data. This will be used as the list of childcare service providers as part of this analysis and used in grouping them by their respective neighbourhood.

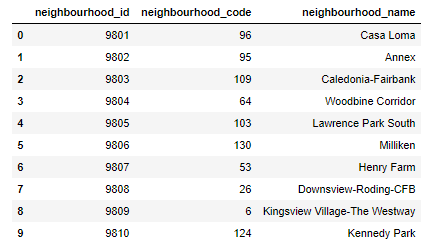
Toronto neighbourhoods have varying geographical areas and shapes, with vastly different populations. This may skew identifying neighbourhoods as family friendly as neighbourhoods with larger populations may naturally have more capacity for childcare services. To normalize, we will pull the most recent population census data by neighbourhood to create a ratio of childcare service capacity by population. This will allow us to compare young family neighbourhoods without natural bias based on population differences between neighbourhoods. The latest population census data is from 2016 and can be accessed in the [Neighbourhood Profiles](https://open.toronto.ca/dataset/neighbourhood-profiles/) data set.

## Cleaning the data

The Toronto Neighbourhood Data primary use was the defining the boundaries of each of the 140 neighbourhoods in Toronto city limits. This did not require cleaning of the data as the GEOJSON file had the necessary coordinate data for the Choropleth map to be used later on.



The secondary use of the Toronto Neighbourhood Data was to define the specific neighbourhoods and their names to be used in the analysis. To accomplish this, the JSON file was converted into a data frame and only the necessary columns were kept: neighbourhood ID, neighbourhood code, and neighbourhood name. This produced 140 unique neighbourhoods for the city of Toronto with a data frame that had a shape of (140, 3). A sample of the data frame can be seen below.



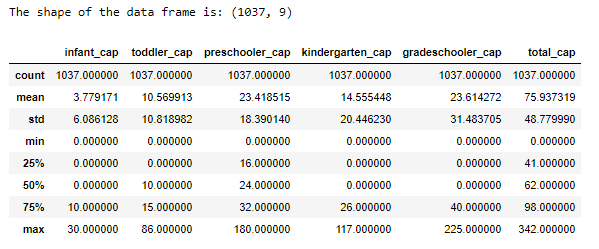
Next, importing and cleaning the Licensed Child Care Centres data was needed to extract the information that will be used in our clustering analysis later on to determine which neighbourhoods provide more childcare services for young families. The Licensed Child Care Centres data has 1037 service providers listed with 22 features (columns) of information for each one. This information included the geo coordinates of each provider. However, it did not include the neighbourhood in which the childcare service provider was located in.

To add the neighbourhood, a function was created using the **Shapely** python library to pass in the geo coordinates of each provider and identify what neighbourhood they were in utilizing the GEOJSON file of Toronto Neighbourhoods. This information was then added to the Childcare Services data frame, resulting in 23 features.

As most of the features would not be needed for the analysis, all unnecessary features were removed resulting in a data frame with the shape (1037, 9). The features that remained were:

* **provider\_id**: unique provider ID
* **provider\_name**: provider name
* **neighbourhood\_name**: neighbourhood name
* **infant\_cap**: capacity for children 0-18 months
* **toddler\_cap**: capacity for children 18-30 months
* **preschooler\_cap**: capacity for children +30 months to 1st grade
* **kindergarten\_cap**: capacity for children in full-day kindergarten
* **gradeschooler\_cap**: capacity for children in 1st grade plus
* **total\_cap**: capacity for all childcare services

Having a final list of information that will be used in our clustering analysis later on we can run light analysis of the overall data set. Looking at the descriptive statistics of the childcare service, we can see that mean total capacity for all age groups of the typical service provider is 75.9 spaces available, with a standard deviation of 48.8. Additionally, the capacity for infant and toddler care with a mean of 3.8 and 10.6 respectively, is much lower than that of preschool age and above. This means that families looking for infant and toddler care must be cognizant of the limited space available for these services.

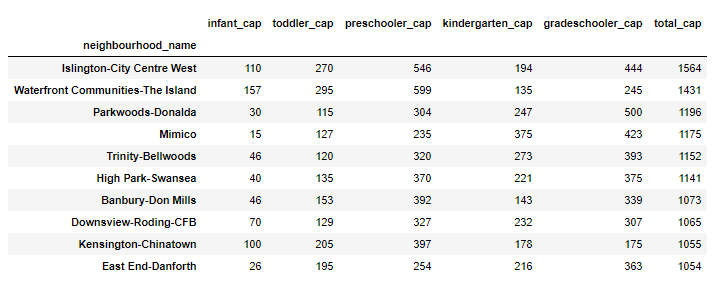


# Methodology

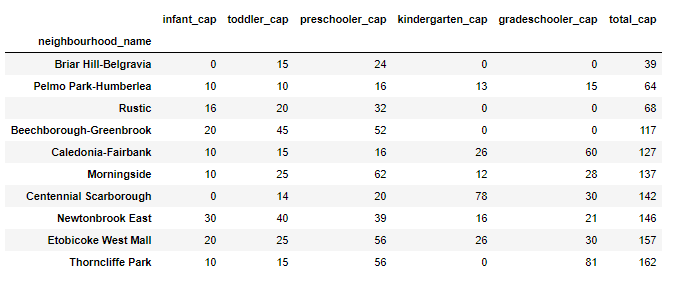
## 3.1 Exploratory Data Analysis

Looking at the Total Capacity for each neighbourhood we can see there is a large deviation in the available childcare services from the top 10 neighbourhoods to the bottom 10 neighbourhoods.

**Top 10 Toronto Neighbourhoods by Total Capacity of Childcare Services: All Age Groups**

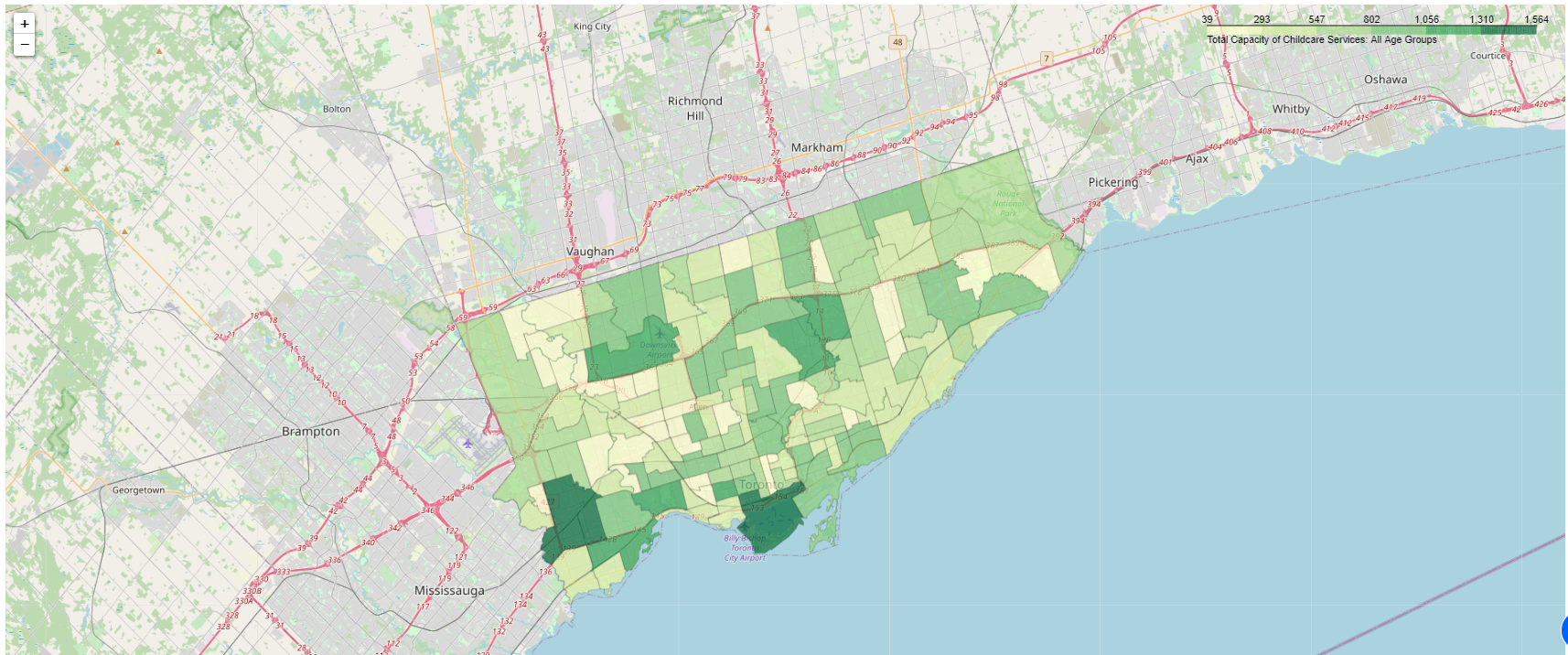


**Bottom 10 Toronto Neighbourhoods by Total Capacity of Childcare Services: All Age Groups**



For easy visualization of the total childcare service capacity neighbourhood, we can add this information to a Choropleth map. This does not yet take into consideration population per neighbourhood, which may skew the initial output to those with larger populations.

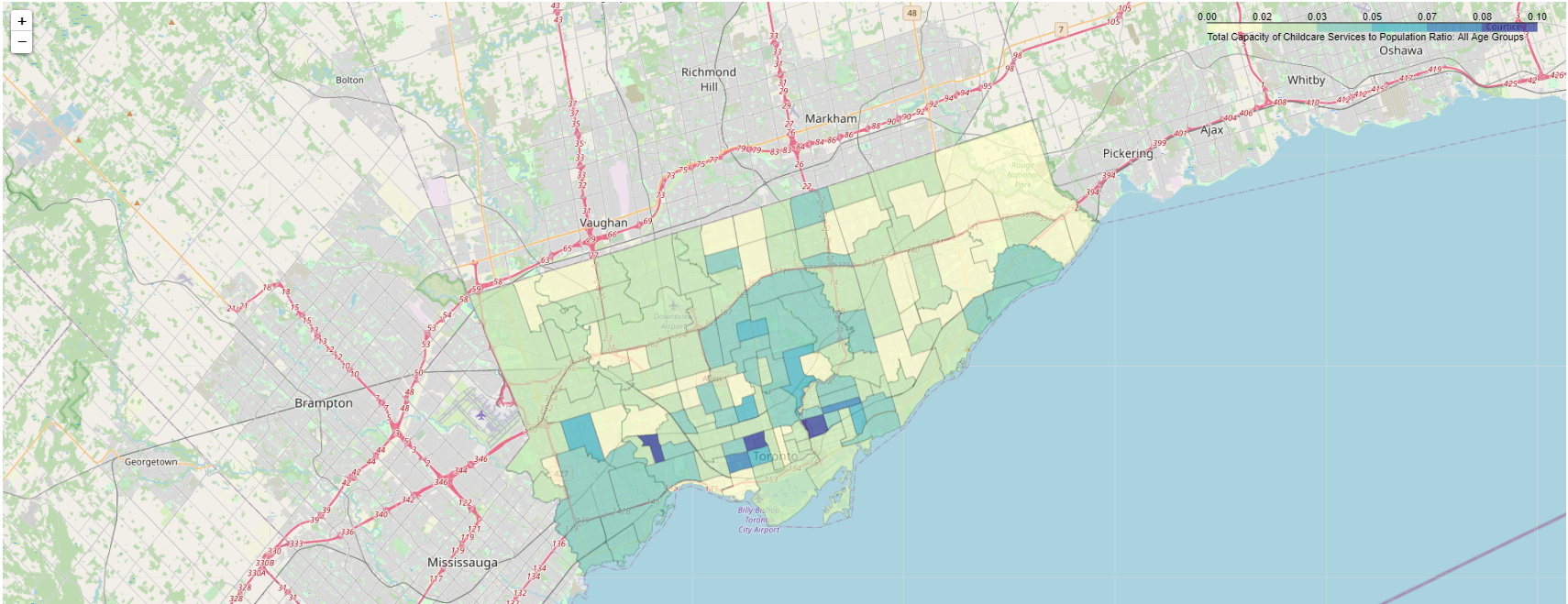
**Toronto Neighbourhoods Total Capacity of Childcare Services: All Age Groups**



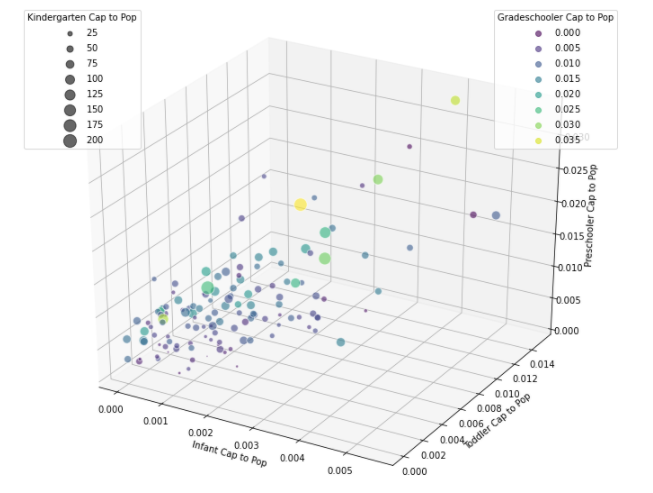
To normalize the capacity by population, we will need to layer in Toronto population census data using the Neighbourhood Profiles data set from Open Data Toronto. This data includes census information by neighbourhood large number of “Characteristics”. As we are only interested in the 2016 Population data, we will only keep this information and remove all the rest. The 2016 population data by neighbourhood will then be appended to the Toronto Childcare Service information we have already cleared.

Looking at the ratio of Total Capacity of Childcare Services to Population, the Choropleth map changes considerably. It shows different groups of neighbourhoods that have higher ratios of Childcare services, which could help identify neighbourhoods for Young Families.

**Toronto Neighbourhoods Total Capacity of Childcare Services to Population Ratio: All Age Groups**



Comparing the neighbourhoods can also be visualized using a scatter plot to see similarities across neighbourhoods by their capacity to population ratios. The 5-dimensional chart below, plots the different age group capacity to population ratios with three youngest age groups represented by the x, y, and z axis and the two older one by size and color. Neighbourhoods with lower childcare capacity to population ratios are represented by small purple dots close to the bottom-left corner. Neighbourhoods with higher capacity to population ratios appear are larger yellow/green dots that are closer to the top-right corner.



## K-means Clustering

The purpose of this analysis is to help identify neighbourhoods based on the childcare service capacity they offer for young families. K-means Clustering identifies data points based on the distance between the data point and its nearest neighbouring data points. This helps identifies data points with commonalities amongst them and works well with our analysis as we are only utilizing one-dimensional numeric data.

The result of the clustering applied to the childcare service data will be neighbourhood clusters based on childcare service capacity for the varying age groups. For the purposes of this analysis we will use 5 clusters to label the neighbourhoods. Once we have cluster labels for each of the neighbourhoods, we will be able to identify which ones offer higher level of childcare service capacity to compare to others. This can ultimately be used to help identify which neighbourhoods have more capacity to support young families.

Utilizing the k-means model in the scikit-learn python library with the neighbourhood data, 5 distinct clusters were identified. Examining the descriptive statistics of each cluster, we can apply a “Cluster Name” to each of the clusters to describe the general capacity of childcare services for neighbourhoods in the cluster. This was done by primarily looking at the **mean** **capacity to population** ratios across each of the age groups in the cluster. From this each cluster was given a name ordered from the lowest total capacity to population mean to highest.

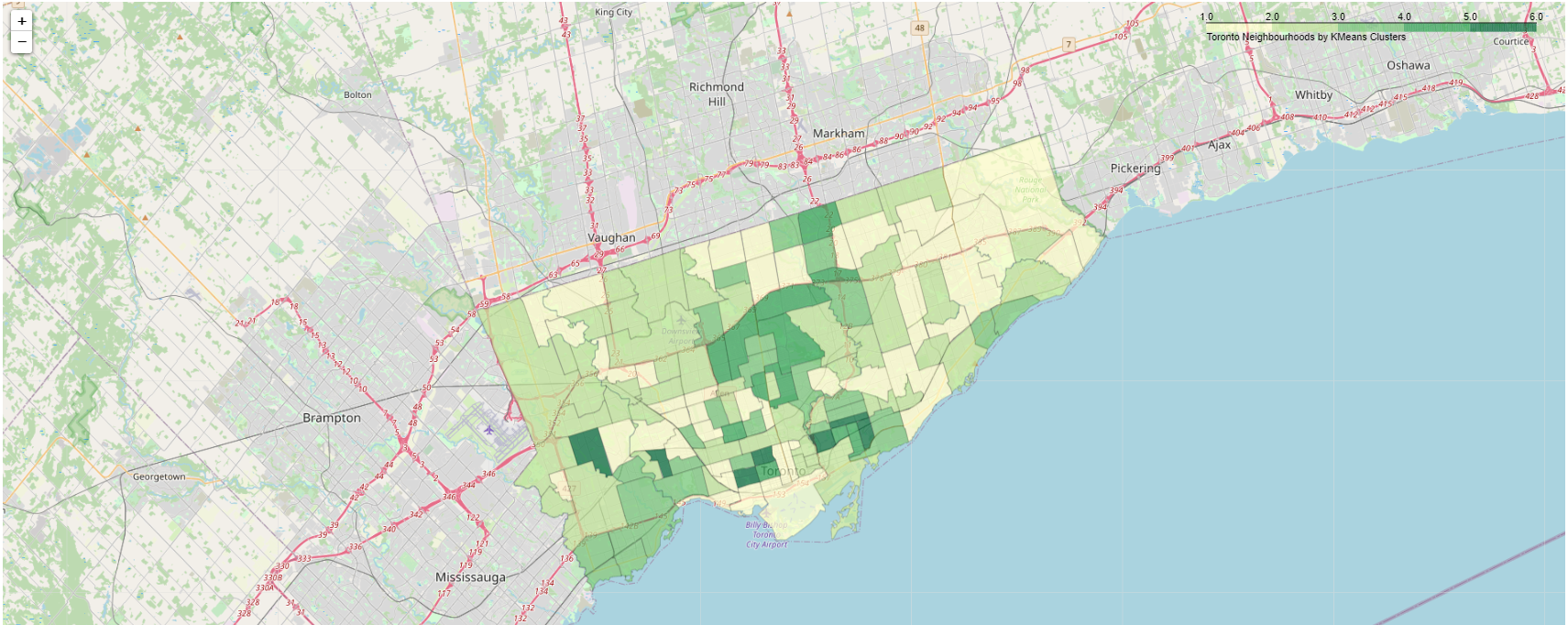
|  |  |  |
| --- | --- | --- |
| **Cluster Assigned** | **Cluster Name** | **Total Cap to Pop Ratio - Mean** |
| 1 | Low Childcare Support | 0.0157 |
| 2 | Low to Moderate Childcare Support | 0.0287 |
| 3 | Moderate and Balanced Childcare Support | 0.0418 |
| 4 | High Newborn Childcare Support | 0.0452 |
| 5 | High Childcare Support | 0.0765 |

\*See Appendix A for complete details

# 4. Results

Adding the labels to each of the neighbourhoods, a Choropleth map of Toronto neighbourhoods can now be created with each of the Clusters Assigned. In the Choropleth map each color represents a cluster with yellow representing cluster 1 and dark green representing cluster 5.

**Toronto Neighbourhoods by Childcare Service K-means Clusters**



Examining the Choropleth map above, it is apparent that there are neighbourhoods with higher support for childcare services than others.

In the immediate downtown neighbourhoods of Church-Yonge Corridor, Baystreet Corridor, Waterfront Communities and Islands, there is low support for childcare services. Once you move away from the downtown core neighbourhoods the support for childcare services increases. Specifically, pockets of increased childcare service capacity by population exist in neighbourhoods:

* Directly east of the city and Don Valley Parkway
* West of the city close to High Park, Mimico, and the Junction, and
* North / North East of the city

As you move Northwest and to the neighbourhoods on the boundaries of city limits, childcare support services show lower support. Many of the neighbourhoods in these areas fall under the clusters “Low Childcare Support” or “Low to Moderate Childcare Support”.

Looking at the 5-dimensional plot again and substituting the "Cluster Labels" for the "Grade-schooler to Cap Ratio" we can see the clusters largely align with the initial exploratory analysis.

Smaller dots appear more often in the bottom left corner of the plot (neighbourhoods with lower cap to pop ratios) while the larger dots appear closer to the top right corner.



Using the k-means clustering, we have successfully able to identify neighbourhoods in Toronto based on their childcare service capacity. This can help in the decision support for Young families as they identify which areas in Toronto to move to.

# Discussion

For Young Families planning on moving and need childcare services, this analysis can be used to identify the neighbourhoods that have the most capacity to support them. Neighbourhoods that appear in clusters 3-5 will likely have the childcare services needed to support a young family at varying age groups. Even neighbourhoods in cluster 3 are generally bordering neighbourhoods that have higher capacity. This information can support decisions on where to live for those that need to consider childcare services carefully in their planning.

Additionally, this information could be used to help identify neighbourhoods that are potentially underserviced for childcare services. Those interested in setting up a new Licensed childcare service to support underserviced neighbourhoods in Toronto could layer this information on top of other demographic, socioeconomic, and other data to help identify the best neighbourhoods for that support.

# Conclusion

Deciding on what neighbourhood to move to or potentially buy a home in is not always a straightforward decision. It requires examining a variety of factors that align with individual needs and lifestyle of those choosing to make move. Young families in Toronto who are often faced with the decision to move to meet their growing family, want to be informed on areas that offer the childcare service support that they need. The analysis here can help them narrow their search by identifying neighbourhoods that have the highest capacity of childcare services and support their decision on where to live in Toronto.

# Disclaimers

This project is limited in scope to only include publicly accessible information made available through the City of Toronto and other digital resources.

It is a one-dimensional analysis and only investigates the capacity of listed childcare service providers to determine Identify Family Neighbourhoods. It does not take into consideration other factors that impact the access to childcare services in a neighbourhood including childcare service providers ratings, income, education, socioeconomics, demographics, other public facilities and many other important social determinants.

# Appendix A: Jupyter Notebook Descriptive Statistics of each Cluster

