

# Marriage Data analysis\*

STA304 WEEK 3

Maggie Zhang

September 27, 2024

This paper investigates the occupancy of shelters in Toronto from 2017 to 2020, focusing on capacity and usage across different shelter sectors. By analyzing the data, we explore how shelter demand {exceeded available resources in certain sectors. This research highlights significant challenges faced by the shelter system, providing insights into the strain on resources and potential gaps in support for the homeless population}. These findings are crucial for informing policies and improving resource allocation to better address homelessness in Toronto.

## 1 Introduction

Homelessness is a persistent and complex issue in many areas in Canada (Gaetz (2010)), and Toronto is no exception. As one of Canada's largest cities, Toronto faces significant challenges in helping homeless people, who are unable to spend on housing largely due to poverty, to meet the necessities of living (Jadidzadeh and Kneebone (2018)). The city operates a range of shelters that serve various groups, including men, women, youth, and families. The balance between availability and the demand for these services is crucial for both the well-being of individuals experiencing homelessness and the government's efforts to reduce homelessness population. Understanding how shelter occupancy changes over time is essential for identifying gaps in resources and shaping effective policy responses.

This paper focuses on analyzing shelter occupancy data from 2017 to 2020 in City of Toronto. The data which offers a detailed look at the city's shelter during this time of period is provided on Gelfand (2022) portal. By examining trends in occupancy through the data, we aim to assess how well Toronto's shelters met the demand for space and where shortages may have occurred. By focusing on specific sectors of the shelter system, pinpoint which groups may have been disproportionately affected by resource limitations.

---

\*Code and data are available at: [https://github.com/MaggieZ111119/CityofToronto\\_Daily\\_Shelter\\_Occupancy.Rproj.git](https://github.com/MaggieZ111119/CityofToronto_Daily_Shelter_Occupancy.Rproj.git)

**FINDING!** This study addresses a clear gap in the current literature by providing an up-to-date analysis of Toronto’s shelter capacity and usage during a time of unprecedented public health and social challenges. \_\_\_\_ **NOTICE!!**

The remainder of this paper are in presented in sections: Section 2 : Data, Section 3 :Results, and Section 4 :Discussion. The data section introduce the dataset used for analysis, expalinin its source and context that shaped its collection.This section thoroughly summarizes the data using visualizations created with `ggplot2` (Wickham et al. (2024)) and tables produced with `knitr` (Xie (2024)). Key variables chosen for the analysis, along with the reasoning behind their selection, are explained. The results section presents the findings drawn from the data, highlighting patterns in shelter occupancy and compare to it’s availability. Following by the discussion section, which explores the implications of these findings, particularly in relation to resource allocation and policy development, offering insights into how Toronto’s shelter system responded to increasing demand.

## 2 Data

All data reviewed and analysis in this paper is the Daily Shelter Occupancy Toronto Shelter & Support Services (2022). The data is essential for evaluating how shelters serve different populations (e.g., men, women, youth, families) and for analyzing the current shelter servise system over time, such as capacity shortfalls or surpluses. It provides inforamtion on all the active shelters exist in the City of Toronto area, collected in forms of four separate datasets corresponding to the years 2017, 2018, 2019, and 2020.

### 2.1 Overview

Each datasets include various characteristics of the shelters. The data was sourced from an Open Data Toronto Portal Gelfand (2022), with each row indicate a unique entry for a specific sector of specific shelter, on a specific date, and has Unique row identifier for Open Data database, “\_id”.

The collection process records information about corresponding name of the non-profit entity that is responsible for the shelter, as well as the name of the facility (e.g. hotel, residence building), along with and their capacity to accommodate homeless clients. The Capacity is being measured as number of bed or a mat/cot available. Something will happen is that for the family sector, it is possible to exceed available capacity depending on the bed number, because family could be accommodated in a room with number beds smaller than number of people in their family. The data was collected every day, 4:00 AM, to record number of homeless clients occupying the shelters. This way to gather information about occupancy provides a consistent snapshot of shelter use across all organizations.

Other similar datasets, especially the ther dataset publish by Toronto Shelter & Support Services, were considered. However, they did not offer the cross-sector view of both occupancy and capacity needed for the focus of this analysis. The selected dataset allows for a broad comparison across time and sectors in different city, providing a fuller picture of trends within the shelter system.

## 2.2 Feature Selection and Aggregation

For the purpose of this analysis, several features are chose to be particularly focused on. variable of interest including the the city (named “SHELTER\_CITY” in the dataset) in which shelters are located; the date of data recorded(“OCCUPANCY\_DATE”); the sector(“SECTOR”) of clientele the shelters serve, for example men, women, youth, families, and even co-ed; the shelter’s capacities(“CAPACITY”) and occupancy(“OCCUPANCY”) levels. These variables were chosen for their relevance in which is to investigate shelter occupancy and capacity across different cities and sectors, focusing on the City of Toronto from 2017 to 2020.

There are also seven other detailed feature in the datasets: “ORGANIZATION\_NAME”, “SHELTER\_NAME”, “SHELTER\_ADDRESS”, “SHELTER\_PROVINCE”, “SHELTER\_POSTAL\_CODE”, “FACILITY\_NAME”, and “PROGRAM\_NAME”. Using R programming lanaguage R Core Team (2023), the `janitor` Firke (2023), `tidyverse` Wickham (2023), and `dplyr` Wickham et al. (2023) packages are used in data simulation, downloading, cleaning, and writting test. No new variables were created for this analysis, but the data was aggregated by grouping observations by city, date, and sector. This aggregation facilitated a more comprehensive comparison across cities and sector groups. A random sample of the cleaned data can be seen in Figure 1.

Table 1: Sample of Cleaned Shelter Data

occupancy_date	shelter_city	sector	total_occupancy	total_capacity
2020-09-27	Toronto	Women	370	511
2017-07-11	Toronto	Men	1741	1801
2019-08-30	Toronto	Families	2949	3177
2017-04-30	Toronto	Women	680	697
2018-04-04	Toronto	Families	2535	2661

Figure 1: Sample of Cleaned Shelter Data

## 2.3 Data Breakdown

As shown in Figure 2, The four cities being observed in the dataset are Toronto, Etobicoke, North York, and Scarborough. The sectors recorded include Co-ed (mixture), Families, Men, Women, and Youth. Across all cities, record for Youth age group being presents. In fact, Youth is the only population in Etobicoke and North York. In Toronto, which has the most number of observation recorded, equal number of sectors are being observed, but how these people in distributed inthe total popluation are not clear. This will be further addressed in the following sections.

Table 2: Number of Observations by City and Sector

shelter_city	Co-ed	Families	Men	Women	Youth	Sum
Etobicoke	0	0	0	0	1461	1461
North York	0	0	0	0	1461	1461
Scarborough	1461	0	569	632	207	2869
Toronto	1461	1461	1461	1461	1461	7305

Figure 2: Number of Observations by City and Sector

The following table Figure 3 provides a summary of the average, maximum, and minimum occupancy, as well as capacity metrics for each city. Toronto has most amount of shelter avaiable and also has most people occuupaying in. The Maximum occupany of Toronto in a day, combined all sectors, from 2017 to 2020,is 6803. Wheras Etobicoke only has 53 people occupanying. The fact that Etotobike has only hoemless people in the group of Youth recorded could explain alittle bit. But, the de=ifference still strong enough to see that the homeless population might be significant larger in Toronto compared to aother cities.

Table 3: Statistics for Homeless Shelter Occupancy by City

City	Average Occupancy	Max Oc- cupancy	Min Occu- pancy	Average Capacity	Max Capacity	Min Capacity
Toronto	5443.37577	6803	3001	6042.49281	7241	4460
North York	64.50650	73	20	67.36345	73	32
Scarborough	159.58727	564	60	178.18549	577	67
Etobicoke	48.06913	53	20	50.81383	53	33

Figure 3: Statistics for Homeless Shelter Occupancy by City

The relastion ship of two key feature, occupancy and capacity, are essential for understaidng

current shelter service system. The relationship is reflected in Figure 4. Occupancy and capacity are greatly following a positive correlation, indicating that most there aren't much over-investment to the shelter system, nor heavily over-filled shelter. But there are noticeable deviation from the perfect correlation to above, these are suggesting there are cases where occupancy is higher than capacity, suggesting occupant may living in a overly packed places. More will be discussed in Section 3, including other learning on the shelter service like their trend, and how these shelter information possibly reflect the homeless.

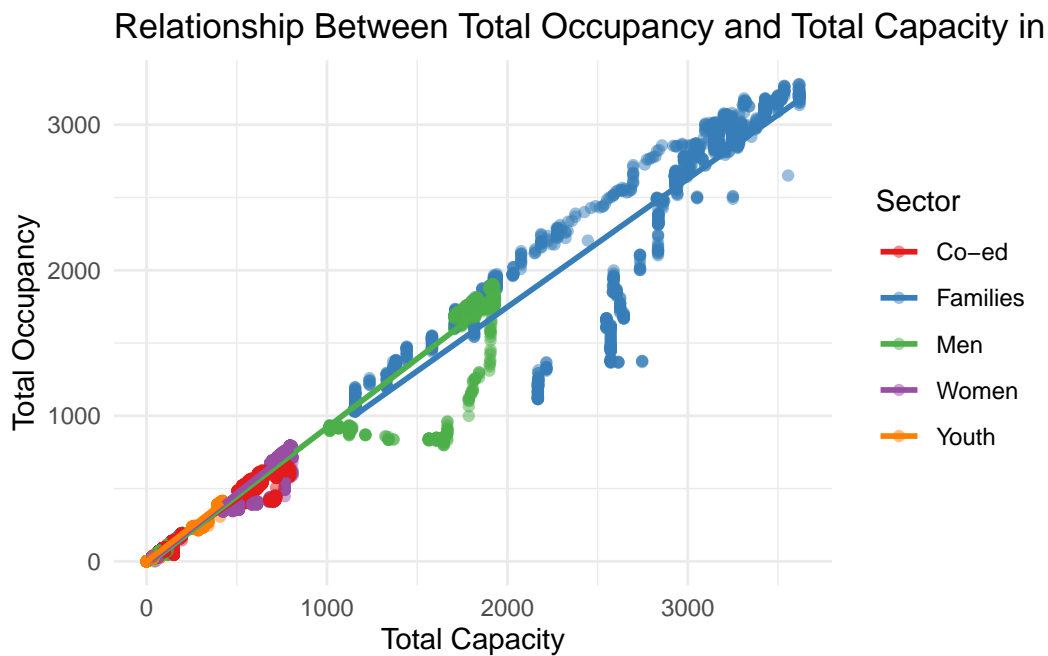


Figure 4: Relationship Between Total Occupancy and Total Capacity in Shelter of the City of Toronto

### 3 Results

Through graphical study of the dataset, some useful information are found and the result will be discussed in this section.

#### 3.1 Shelter Occupancy Trend over Time

The figure Figure 5 as depict the trend of homeless people occupying shelters in each city. A high in occupancy are notice in all four city around the time between late 2019 to early 2020. Especially, there is a obvious rising trend in Scarborough. Combined with the information on

the “About Daily Shelter Occupancy” section in the data website Toronto Shelter & Support Services (2022), it rise may be subjected to the outbreak of COVID\_19. The occupancy number going significantly decrease from mid 2020, across all four cities also suggest this finding.

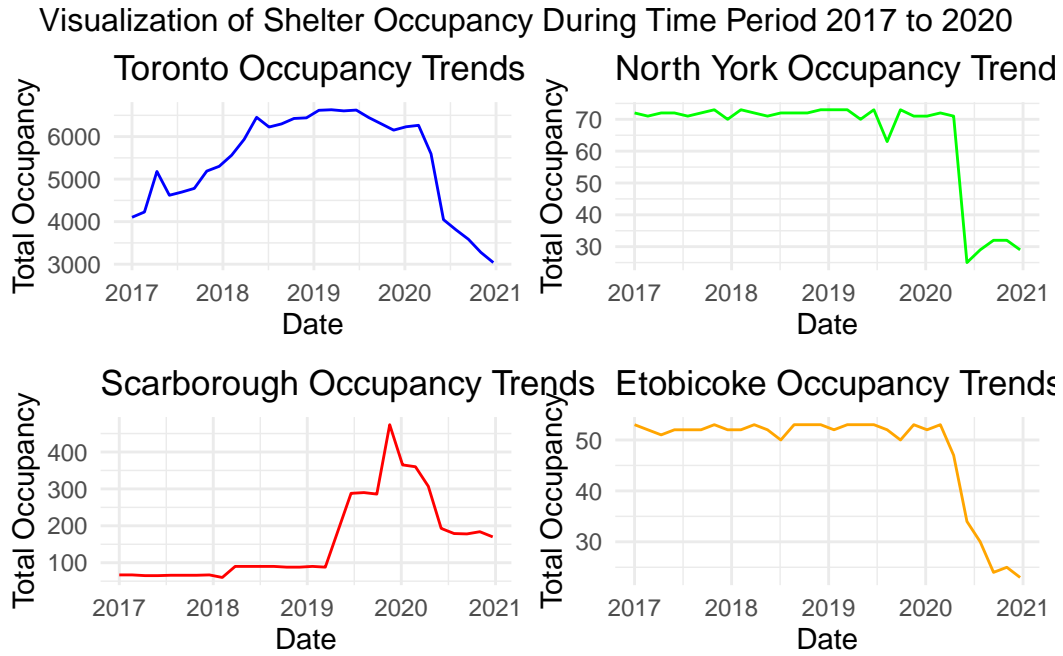


Figure 5: Visualization of Shelter Occupancy During Time Period 2017 to 2020

### 3.2 Shelter Utilization Analysis

The Utilization Rate as being calculated as:

$$\text{Utilization Rate} = \left( \frac{\text{Number of Occupancy}}{\text{Number of Capacity}} \right) \times 100\%$$

From Figure 6, it's the utilization rate in Toronto remains the lowest over 4 years almost all the time compared to other cities, but still in a relatively high number. This suggest that shelter service across all four cities are not over-investing, their service are actually helping the homeless population since a lot of people are using them. There is a significant decrease in Utilization Rate during

### 3.3 Over Capacity Cases

Back to the original dataset, identify cases where shelter capacity is exceeded (i.e., occupancy > capacity) and understand the frequency and distribution of these events. see Figure 7, under

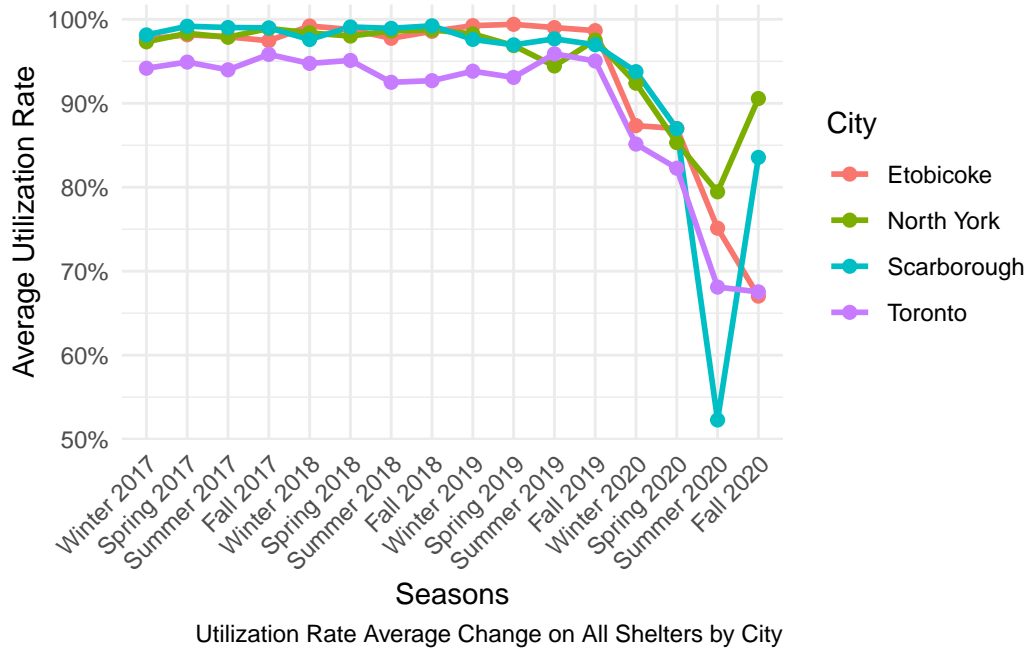


Figure 6: Yearly change in Utilization Rate Average on All Shelters by City (2017-2020)

Table 4: Overcrowded Cases by City and Sector

shelter_city	Families	Youth	Co-ed	Men	Women
Toronto	87	0	0	0	0
Etobicoke	0	0	0	0	0
North York	0	0	0	0	0
Scarborough	0	0	0	0	0

Figure 7: Overcrowded Cases by City and Sector

## **4 Discussion**

### **4.1 First discussion point**

from the study, the Marriage number across Toronto is .....

### **4.2 Second discussion point**

[My second point]

### **4.3 Third discussion point**

[My third point]

### **4.4 Weaknesses and next steps**

The study is not finalized, more detailed and careful data examination is needed.



## Appendix

### References

- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://github.com/sfirke/janitor>.
- Gaetz, Stephen. 2010. “The Struggle to End Homelessness in Canada: How We Created the Crisis, and How We Can End It.” *The Open Health Services and Policy Journal* 3 (1).
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Jadidzadeh, Ali, and Ron Kneebone. 2018. “Patterns and Intensity of Use of Homeless Shelters in Toronto.” *Canadian Public Policy* 44 (4): 342–55.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto Shelter & Support Services. 2022. “Daily Shelter Occupancy.” <https://open.toronto.ca/dataset/daily-shelter-occupancy/>.
- Wickham, Hadley. 2023. *Tidyverse: Easily Install and Load the Tidyverse*. <https://tidyverse.tidyverse.org>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, and Teun van den Brand. 2024. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.