

# Toronto Overoccupied Shelters Analysis

Jennifer Liu - 1006702310

02/02/2022

## Abstract

Daily Shelter & Overnight Service Occupancy & Capacity is used from the City of Toronto Open Data Portal to analyze Toronto shelters capacity under the pandemic and the sector or program that needs more shelters. It is found that Toronto shelters have a capacity rate that approaches the maximum and is currently in crisis. Although youth shelters are slightly less occupied, women shelters need more attention. The city has not created any winter shelter program for women and has the lowest number of COVID-19 response shelter. However, biases and the lack of historical data might affect the result. With the current finding, the city ought to bring more attention to this shelter crisis to prevent more COVID-19 outbreaks and hospitalization.

1

## Introduction

Global News has reported that Toronto is currently facing a shelter capacity crisis with the cold weather and COVID-19 impact. This has resulted in an increasing number of homelessness with hypothermia and frostbite Delitala (2022). Although the world is fighting this pandemic together, homelessness people need extra attention as they are the ones who lack the most of sanitary resources. In fact, more than 50 Toronto shelters are facing COVID-19 outbreaks Draaisma (2022). Therefore, it is worth examining whether Toronto shelters have enough capacity to meet its demand given the current situation.

The dataset of Daily Shelter & Overnight Service Occupancy & Capacity (*About Daily Shelter & Overnight Service Occupancy & Capacity* 2022), taken from the City of Toronto Open Portal website will be analyzed (*Open Data* 2022). It contains the most updated information about shelters' program, usage and capacity for shelters in Toronto. This 2022 data reflects shelters' current situation. We are going to first analyze whether Toronto has enough shelters for the homelessness and which sector and program need to be added or removed in order to meet the demand. It is also important to consider and analyze whether women and youth have enough shelters as they need more accommodation such as domestic violence support (*Shelters Offer More Than Safety* 2015).

This report first discusses about the data collection and cleaning process for the dataset. Its various limitations, biases and characteristics are also discussed. Then, a brief data summary shows some important findings. It is shown that both room-based and bed-based shelters have a median capacity rate of 100, which proved a shelter crisis the city is currently facing. Youth's shelters are slightly less occupied than the other ones, but women shelters have less programs available: there is no winter program and have the lowest number of COVID-19 response shelters.

This report is analyzed using R R Core Team (2020), using tidyverse Wickham et al. (2019), dplyr Wickham et al. (2020) and opendatatoronto packages Gelfand (2020). All the tables and graphs are created using ggplot2 Wickham (2016).

---

<sup>1</sup>This is my GitHub repository: <https://github.com/jenniferliu888/STA304>

# Data Collection and Cleaning

## Data Source

This report uses data collected from the Shelter Support and Housing Administration (SSHA) division's Shelter Management Information System (SMIS) database (*About Daily Shelter & Overnight Service Occupancy & Capacity* 2022). This division manages housing and homelessness support in Toronto and provides services such as street outreach and emergency shelters (*About City-Operated Services* 2022). The SMIS is a tool that requires shelter users to enter their information to better serve them by avoiding the loss of important identification paper, facilitating shelters' transfer and many more (*About City-Operated Services* 2022). This dataset (*About Daily Shelter & Overnight Service Occupancy & Capacity* 2022) is retrieved from the City of Toronto Open Data Portal (*Open Data* 2022) using the R package *opendatatoronto* Gelfand (2020). The current dataset is last updated on February 2, 2022.

## Methodology, Data Collection and Limitations

The 2022 dataset contains information about active shelters of the SSHA, which captures daily shelters' usage, capacity, location, classification and operator. It has been upgraded from the Data Shelter Occupancy dataset, also found on the Toronto's Open Data Portal, that contains data from 2017 to 2021 (*About Daily Shelter* 2021).

Shelter users are required to complete a series of questions in order to meet SMIS's requirement that helps to facilitate services provided. Although most information is confidential, the system generates a data that picture the shelter's occupancy rate at 4AM (*About Daily Shelter & Overnight Service Occupancy & Capacity* 2022). It is uncertain whether all homelessness completed the questions as the collection method is not specified whether it is mandatory. Moreover, this data is unaudited and may not reflect the actual situation given COVID and shelters' instability.

The SMIS database only records shelters that are funded or managed by SSHA (*Shelter Management Information System*, n.d.). Although most shelters are funded by the city, some shelters might be privately funded and are not included on the list. It is uncertain the exact number as no data has been found about this matter. Thus, the dataset used for the report does not represents all shelters in Toronto. There is a sampling/selection bias as the sample selection process is not random and privately-funded shelters that are not included in the sample could have distinct characteristics.

The dataset also contains missing data because shelters occupancy are measured by room-based or bed-based capacity. Room-based is for family program or motel-shelter that offer sleeping rooms. Therefore, the analysis and the result section ought to analyze room-based and bed-based capacity shelters independently. When analyzing data for room-based capacity, bed-based occupancy is going to be filtered out and vice-versa. By doing so, the missing value is not going to affect our analysis and we could compare the difference in occupancy rate between room-based and bed-based shelters.

This dataset has considered ethical factors as privacy information has been removed. However, it is important to consider whether homelessness were forced to provide the information as this might entail an ethical debate.

Because data is updated daily, this dataset reflects the current state of shelters' occupancy under the pandemic. However, due to the difference in variables and collecting method between 2022 dataset and previous ones, it is not possible to compare the data across different years. Thus, the analysis of the report is only relevant for the current state and would change when there is no pandemic and when weather is warmer.

## Data Characteristics

The Daily Shelter & Overnight Service Occupancy & Capacity dataset contains 4587 observations and 32 variables. In order to keep only the variables that are relevant to the report, 12 variables are kept for the

analysis: sector, program area, capacity of the bed, number of occupied beds, unoccupied bed, unavailable beds, beds occupancy rate, capacity of rooms, unavailable rooms, occupied rooms and rooms occupancy rate. Shelters have 5 sectors (mixed adults, men, women, families and youth) and 3 programs (base shelter, COVID response and winter programs). The capacity measure the actual number of room/bed available. The capacity of the same shelter might vary because there are also a number of unavailable rooms and beds that cannot be occupied. The occupancy rate is calculated by the system and is the actual capacity divided by the occupancy number. The original dataset provides two type of capacity: actual and funding. Funding refers to the capacity the shelter was intended to provide and does not reflect the actual state of the shelter and is therefore removed from our dataset.

All shelters are either bed-based or room-based. Bed-based capacity is for shelters with common sleeping areas and room-based capacity is for family programs and hotel programs. For the purpose of the report, observations are separated into two dataset: room-based with 1591 observations and bed-based with 3136 observations.

## Data Summary

Here is a summary of the mean, median, standard deviation and IQR for the occupancy ratio of the dataset.

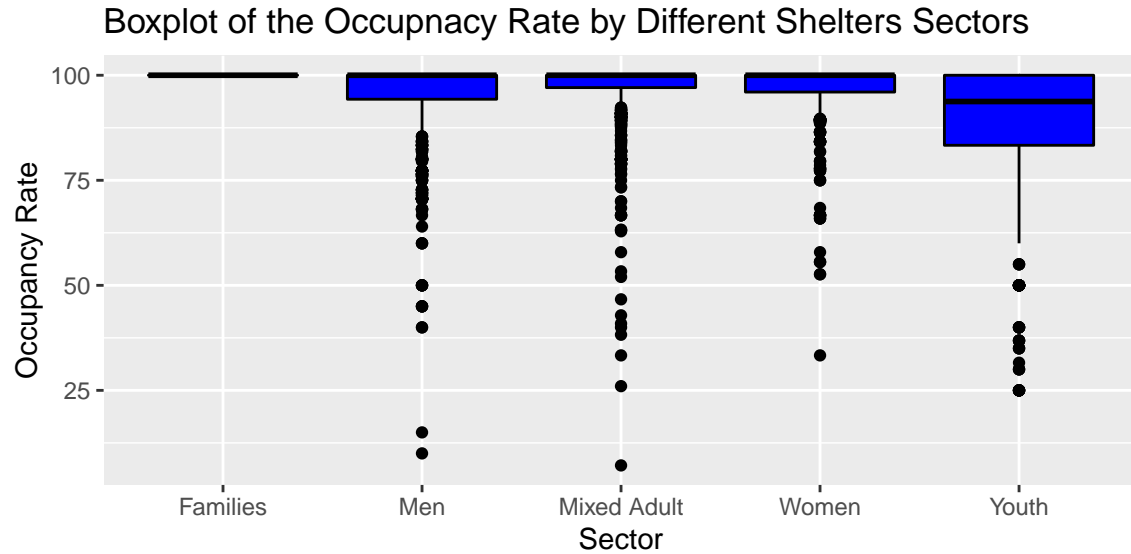
Table 1: Summary of Occupancy Rate of Beds for Shelter

min	Q1	median	Q3	max	mean	standarddeviation
7.14	93.75	100	100	100	94.50866	11.14594

Table 2: Summary of Occupancy Rate of Rooms for Shelter

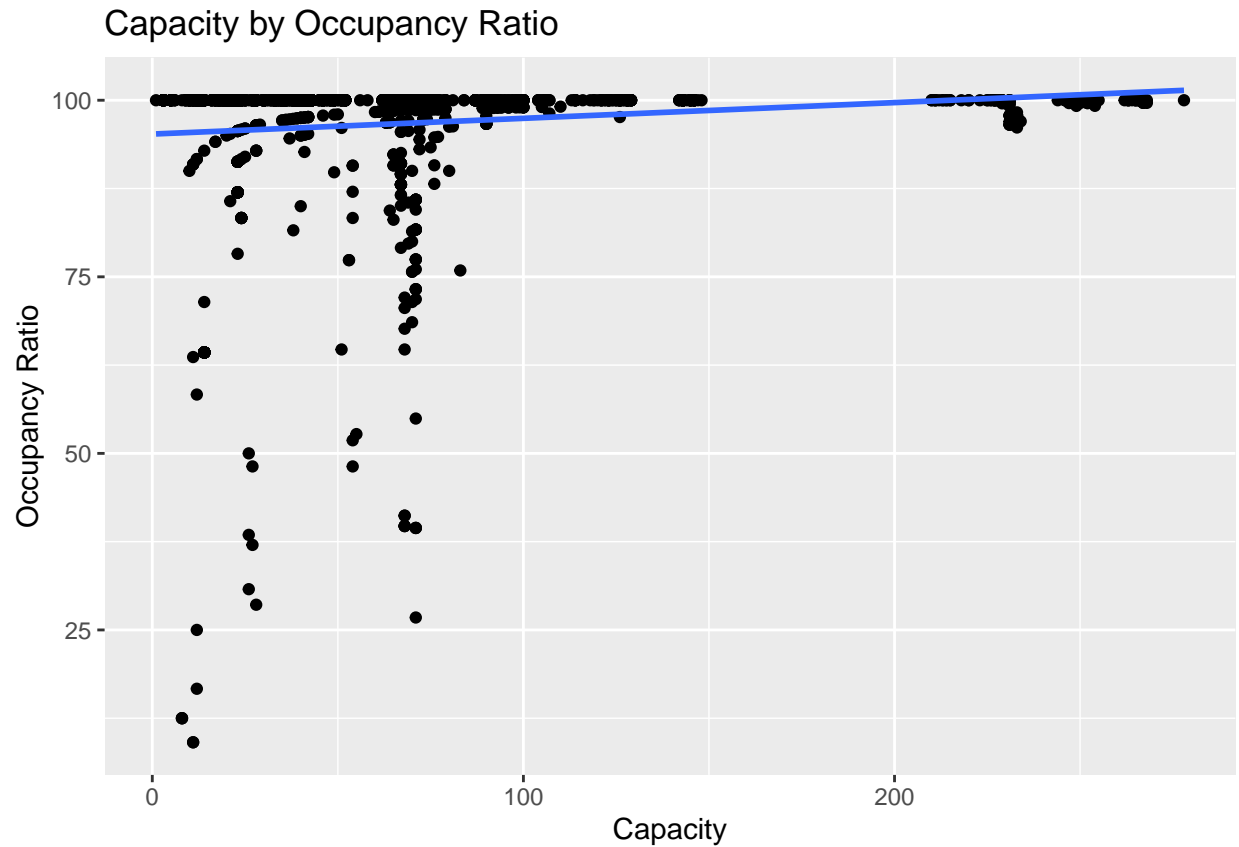
min	Q1	median	Q3	max	mean	standarddeviation
9.09	100	100	100	100	96.71002	10.52368

The median of the occupancy rate for shelters with rooms and beds are both 100. This indicates that shelters in Toronto are extremely occupied or fully occupied most of the time. Although the mean for bed-based shelter is slightly lower than room-based with 94.49 compared to 96.75, the difference is small. The standard deviation of 11.17 for bed-based and 10.32 for room-based shelters is also low. The spread of the dataset is close to the median, which again signals that the city would need new shelters as most of the shelters have an occupancy ratio very close to 100. When the occupancy ratio is 100, it means that the occupancy and capacity number are the same.

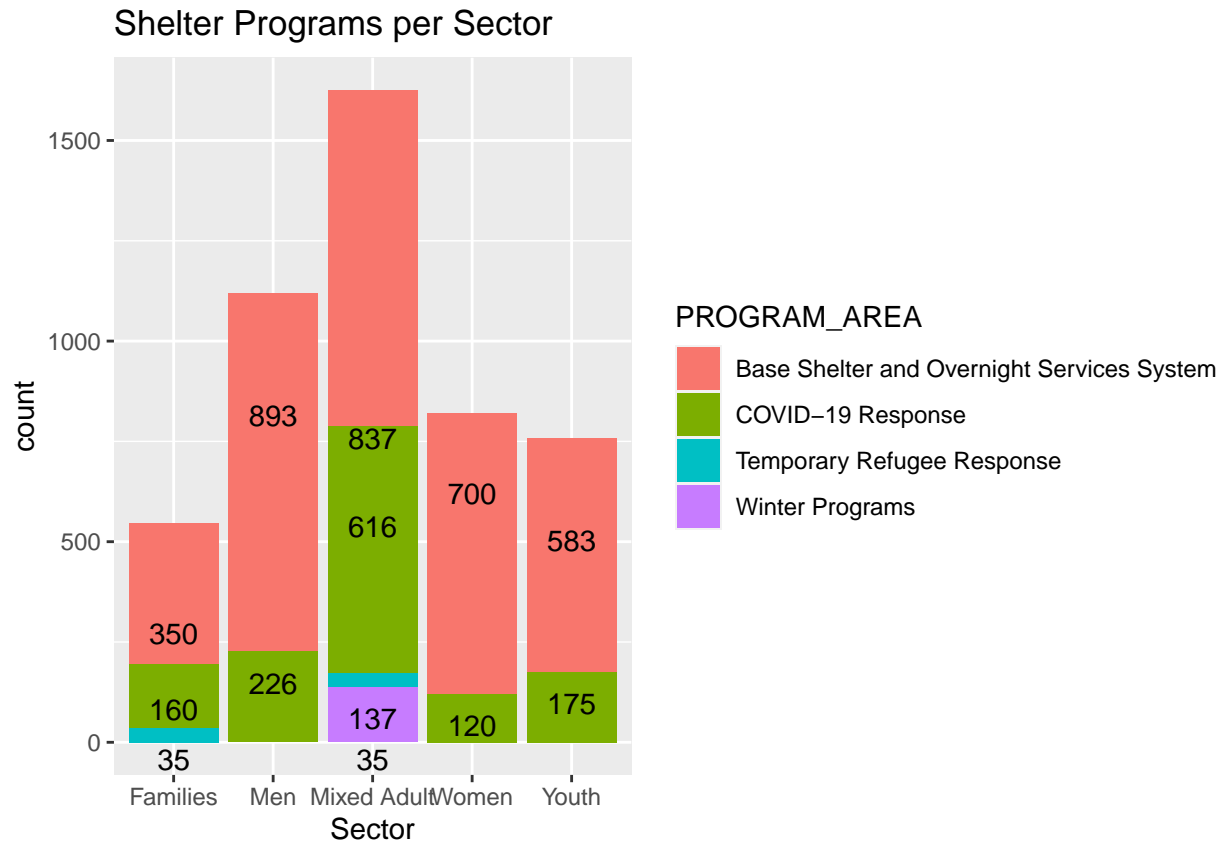


This graph features 5 boxplots of numbers of shelters' unoccupied beds according to their different sector. Men, mixed adults and women sector shelters are skewed to the left, with their median approaching 100, which reflects the very high occupancy ratio obtained from table 1 @. On the other hand, the shelter that serves youth is more symmetric, with a median approaching 88. Thus, it is reasonable to conclude that the shelters for youth have a lower occupancy ratio than the other ones. There are also many outliers in terms of the occupancy ratio for all sectors, which means that it also happens the shelters are under occupied. The boxplot for families is not considered as we are analyzing bed-based shelters whereas families shelters are most room-based.

Thus, using the above data, it is possible to make a hypothesis that the city should increase the number of shelters in Toronto as the boxplot for men, mixed adults and women are extremely left skewed. Although youth shelters have a high occupancy rate, they are still doing better than others shelters and some resources can be allocated if needed.



This is a scatterplot of occupancy ratio and the shelter's capacity in order to determine if there exists a relationship between the two. For example, if bigger shelter would have a higher occupancy ratio or not. The variables occupancy ratio and capacity have a weak positive correlation. This means that the capacity of the shelter might affect the occupancy ratio and when adding new shelters in Toronto, the size of the shelter could be considered when thinking about the occupancy rate.



This histogram gives a clear summary of the different shelter programs by sectors. We can see that there has the highest number of mixed adults program, followed by men, with families shelter the lowest. Winter program is only available for mixed adult shelters whereas temporary refugee response program is only for mixed adult and families shelter. Women have the lowest number of COVID response shelters.

## Reference

- About City-Operated Services.* 2022. City of Toronto. <https://www.toronto.ca/community-people/community-partners/emergency-shelter-operators/about-torontos-shelter-system/see-our-shelters/>.
- About Daily Shelter.* 2021. Open Data Toronto. <https://open.toronto.ca/dataset/daily-shelter-occupancy/>.
- About Daily Shelter & Overnight Service Occupancy & Capacity.* 2022. Open Data Toronto. <https://open.toronto.ca/dataset/daily-shelter-overnight-service-occupancy-capacity/>.
- Delitala, Albert. 2022. *Relentless Cold Has Put Toronto Shelter Capacity in Crisis, Advocates Say.* Global News. <https://globalnews.ca/news/8581880/toronto-shelter-capacity-crisis/>.
- Draaisma, Muriel. 2022. *Resident of Toronto's Largest Homeless Shelter Dies with Covid-19 as 50 Shelters Battle Outbreaks.* CBC. <https://www.cbc.ca/news/canada/toronto/shelter-resident-death-covid-19-outbreaks-toronto-1.6317072>.
- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal.* <https://CRAN.R-project.org/package=opendatatoronto>.
- Open Data.* 2022. City of Toronto. <https://open.toronto.ca>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

*Shelter Management Information System*. n.d. City of Toronto. <https://www.toronto.ca/community-people/community-partners/emergency-shelter-operators/shelter-management-information-system/>.

*Shelters Offer More Than Safety*. 2015. Canadian Women’s Foundation. <https://canadianwomen.org/blog/shelters-offer-more-safety/>.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.