Assessing Daily Shelter Availability in Toronto: An Occupancy and Capacity Analysis*

More shelters for women are needed

Zixuan Yang

April 18, 2024

This research examines the shortage of Toronto's shelter system in meeting the needs of its homeless population. By analyzing shelter occupancy data from Toronto's open data using logistic regression, this paper indicates that resources for shelters in North York, women's shelters, and emergency shelters are insufficient. According to the results, the government should increase resources for women's shelters and add shelter resources in high-demand areas like North York. These findings assist the Toronto government in making appropriate adjustments based on the uneven distribution of resources, maximizing the utilization of shelter resources.

Table of contents

*Code and data are available at: https://github.com/ZixuanYangFrank/Shelter.git

1 Introduction

Like many urban centers around the world, Toronto faces complex issues such as homelessness and providing shelter services for vulnerable populations. With the growing population, Toronto's shelter system plays a crucial role in supporting individuals and families facing housing insecurity. The Shelter and Support Services division in Toronto is responsible for overseeing a network of overnight shelters and allied services aimed at meeting the urgent needs of the homeless population. At the core of this effort is the Shelter Management Information System (SMIS) database, which provides a daily list of active shelters and overnight service programs operated within the city.

Despite the existence of these services, Toronto's shelter system continues to face ongoing challenges related to occupancy rates and capacity. Demand often exceeds available resources, resulting in overcrowding and a lack of sleeping space. The issue of shelter occupancy in Toronto is multifaceted. Insufficient funding or resources may lead to an inability to meet the growing demand for shelter beds or rooms. Additionally, factors such as maintenance, repairs, renovations, outbreaks, or pest control may temporarily lead beds or rooms unavailable, further exacerbating capacity issues.

There is a small portion of users long-term users within Toronto's shelter system,but occupying 40% of shelter capacity. Long-term users pose a challenge to the shelter operation because shelters are intended to provide temporary assistance to those in need until they find permanent housing, rather than offering long-term accommodation (Jadidzadeh and Kneebone 2018). Given these challenges, it is necessary to conduct a detailed analysis of Toronto's housing system to assess whether housing capacity meets requirements and to identify geographical disparities. This analysis can then inform the Toronto government on recommendations for strengthening shelter system services.

1.1 Estimands

Whether a shelter is available is often an important concern for people, so this article will estimate the usage of shelters and focus on whether shelters are at full occupancy. The location of the shelter will affect its occupancy situation. Shelters located in downtown Toronto, where the population is concentrated, face high demand and may have difficulty finding shelters that are not fully occupied. Conversely, shelters in other areas with sparse populations are more likely to have availability. Additionally, the shelter sector caters to different demographics based on gender, age, and household size of the service user group(s) served at the shelter location. There are currently five shelter sectors in Toronto: adult men, adult women, mixed adult (co-ed or all gender), youth, and family. Different sectors serve different populations and needs. Shelter programs include emergency and transitional shelters. Emergency shelters can be accessed by any individual or family experiencing homelessness with or without a referral, while transitional shelters require a referral for individuals and families. Therefore, emergency shelters are more likely to be fully occupied. The capacity of shelters is measured in rooms and beds. Shelters measured in beds often have programs with common sleep areas, while room-based capacity is not shared by people from different households and is more suitable for families. The differences in these shelter characteristics will affect shelter usage. Therefore, the estimand of this study is the impact of these five factors mentioned above on the probability of shelter full occupancy.

This paper's structure is displayed as follows. The data analysis, which will be performed by Section 2, will comprise an overview of the data that contains the variables' definitions, number of variables, and source. The logistic model and its accompanying assumptions will be introduced via Section 3. The study of the logistic regression's findings is the main emphasis of Section 4. Lastly, Section 5 is where a discussion of these findings is suggested.

2 Data

2.1 Data source

The dataset is published on the Toronto open data (Gelfand 2022). This dataset provides a daily list of active overnight shelters and coordinated services from the Toronto Shelter and Support Services division's Shelter Management Information System (SMIS) database. It includes information such as the operator, location, classification, occupancy rate, and capacity of each program. The occupancy rate is calculated by counting the number of people occupying space at 4:00 AM the following day. By comparing the number of occupants to the capacity, we can determine whether the shelter is fully occupied on a given day. For confidentiality reasons, the dataset does not include information about violence against women shelters. These data are directly obtained from the management database and only reflect the recorded status in the database. They have not been audited and therefore may carry a risk of not accurately reflecting the actual situation.

2.2 Data Descriptive Analysis

The original dataset contains 14,415 observations and 32 variables. Initially, variables such as time and location were excluded, while information such as location, sector, capacity type, program model, and capacity were retained. The variable explanations are as follows:

Location (LOCATION_CITY): City where the program is located.

Sector (SECTOR): Categorization of homeless shelters based on the gender, age, and household size of the service user group(s) served at the shelter location. There are five shelter sectors in Toronto: adult men, adult women, mixed adult (co-ed or all gender), youth, and family.

Capacity type (CAPACITY_TYPE): Indicates whether the capacity for this program is measured in rooms or beds. Family programs and hotel programs where rooms are not shared by people from different households are room-based.

Program model (PROGRAM_MODEL): Classification of shelter programs as either Emergency or Transitional. Emergency programs can be accessed by any individual or family experiencing homelessness with or without a referral. Transitional programs provide required, specialized programming and can be accessed by eligible individuals and families experiencing homelessness by referral only.

Capacity: Due to the original data recording capacity separately by capacity type, room-based shelters only have room capacity, while bed-based shelters only have bed capacity. In this study's capacity calculation process, room and bed capacities are merged into a single variable.

Full occupancy (Occupancy_full): This is the response variable of this study, with two levels: 0 and 1. When the occupancy rate of both beds and rooms (OCCUPANCY_RATE_BEDS and OCCUPANCY_RATE_ROOMS) is 100%, it indicates full occupancy in the shelter.

Retaining the aforementioned variables and removing observations with missing values, we obtain 13,759 observations and 6 variables. The summary statistics of the categorical variables are shown in Table 1. Approximately 1/5 of shelters are not fully occupied, indicating that the shelter system still has enough space to meet people's needs. The government needs to focus on how to manage to fully utilize this nearly 20% of idle resources. In terms of shelter location, shelters in the Toronto area are the most prevalent, accounting for over 80% of the total, while Vaughan shelters comprise less than 1%. The significant disparity in shelter distribution between cities and within cities may lead to differences in shelter occupancy rates in different areas. Shelters serving men outnumber those serving women, while mixed adult shelters have the largest proportion, accounting for approximately 29.3% of all shelters. Family shelters have a smaller proportion, comprising only 10%. The majority of shelters have bed-based capacity, approximately 73%, while those for emergency purposes make up over 80%. Shelter capacity is also a factor influencing occupancy rates. Figure 1 illustrates the relationship between shelter capacity and full occupancy, showing that shelter capacity is concentrated below 100, and shelters with capacities exceeding 200 rarely experience not fully

Table 1: Summary statistic

		N	%
Occupancy_full	0	2534	18.4
	1	11225	81.6
LOCATION_CITY	Etobicoke	321	2.3
	North York	1407	10.2
	Scarborough	856	6.2
	Toronto	11068	80.4
	Vaughan	107	0.8
SECTOR	Families	1391	10.1
	Men	3829	27.8
	Mixed Adult	4034	29.3
	Women	2372	17.2
	Youth	2133	15.5
CAPACITY_TYPE	Bed Based Capacity	10081	73.3
	Room Based Capacity	3678	26.7
PROGRAM_MODEL	Emergency	11284	82.0
	Transitional	2475	18.0

occupied situations. The above data indicate that there are regional, gender, capacity type, and program model disparities in the government's provision of shelters. These differences will affect shelter occupancy rates. Below, I will use logistic regression models to study the probability of shelter occupancy being fully occupied.

In this part, I use the readr package (Wickham, Hester, and Bryan 2023), tidyverse package (Wickham et al. 2019) and ggplot 2 package (Wickham 2016) in R (R Core Team 2023) to import the dataset and do the plot of Figure 1. modelsummary package (Arel-Bundock 2022) is used to build the Table 1.

2.3 Measurement

While information regarding shelter location, sector, capacity type, and program model can be directly obtained from shelter registration records, measuring shelter capacity requires special attention. This is because some rooms or beds may be unavailable due to reasons such as maintenance or repair work. Therefore, when calculating shelter capacity, we only consider the rooms or beds that are available on the reporting date.

On the other hand, measuring the occupancy rate involves determining the proportion of actual bed capacity that is occupied on the reporting date. This is calculated by dividing the number of occupied beds or rooms by the actual capacity of beds and rooms available on that day.

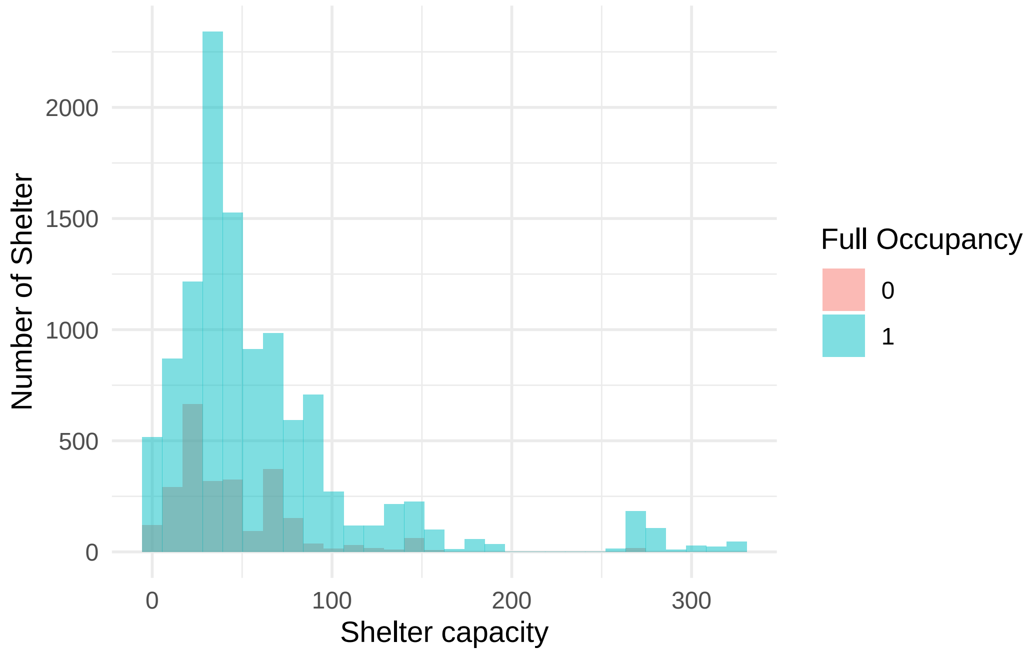

Figure 1: The distribution of Shelter Capacity by full Occupancy

3 Model

Our modeling strategy has two main objectives. Firstly, we aim to establish a model to investigate the impact of different shelter characteristics on the probability of a shelter being fully occupied. Secondly, through model predictions, we aim to help individuals quickly locate shelters with available space, thus avoiding spending significant time searching for shelters.

Since the outcome variable “fully occupied” (Occupancy_full) is binary, a logistic regression model is suitable for modeling the relationship between the fully occupied status and shelter characteristics. I have chosen a Bayesian analysis approach in conjunction with the logistic regression model in the analysis. Bayesian analysis combines prior beliefs with observed data to estimate parameters of interest. It begins with initial assumptions about the parameters, represented by prior distributions. As new data becomes available, these prior beliefs are updated to form posterior distributions using Bayes’ theorem. This approach offers a more flexible and intuitive way to estimate parameters and quantify uncertainty compared to traditional statistical methods. Therefore, we believe that using a Bayesian analysis model with the logistic regression model will be beneficial for our analysis.

3.1 Model set-up

Considering that the output variable (Occupancy_full) y_i is binary, I assume that y_i follows a Bernoulli distribution with probability π_i. To establish the relationship between the predictors and the probability π_i., we employ logistic regression by Wang et al. (2015). The logistic regression model is defined as:

$$y_i|\pi_i \sim Bern(\pi_i)$$
$$logit(\pi_i) = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \beta_3 * X_{3i} + \beta_4 * X_{4i} + \beta_5 * X_{5i}$$
$$logit(\pi_i) = log(\frac{\pi_i}{1-\pi_i})$$

Where Xs are the predictors, representing the characteristic of the shelter.

βs are the estimated parameters, and I assume that the prior distribution of these coefficients as

$$\beta_0 \sim Normal(0, 2.5)$$
$$\beta_1 \sim Normal(0, 2.5)$$
$$\beta_2 \sim Normal(0, 2.5)$$
$$\beta_3 \sim Normal(0, 2.5)$$
$$\beta_4 \sim Normal(0, 2.5)$$
$$\beta_5 \sim Normal(0, 2.5)$$

In order to predict these coefficients, we use stan_glm() from the package rstanarm by Goodrich et al. (2022) in R (R Core Team 2023). The model is written in short-hand, in practice, these categorical variables will be converted into a series of indicator variables with different coefficients.

3.2 Model justification

I expect that the demand for shelters in Toronto will be the highest due to its dense population. Shelters with stricter review processes are more likely to be fully occupied. For example, emergency shelters are expected to fill up more quickly than transitional shelters. Since room-based capacity offers private spaces, the demand for such shelters is likely to be greater, making them more scarce compared to bed-based capacity shelters. Individuals and families tend to prioritize shelters with larger capacities, believing that they will have a better chance of securing a spot. However, whether this assumption holds true requires further investigation.

4 Results

Table 2 presents the results of the Bayesian analysis approach in conjunction with the logistic regression model. The positive coefficient for capacity indicates that shelters with larger capacities are more likely to be fully occupied. The standard deviation of the coefficient is shown in parentheses below the coefficient.

Regarding the location coefficients, with Etobicoke as the reference group, the results show that the coefficients for North York and Scarborough are positive, while the coefficients for Toronto and Vaughan are negative. The comparison between Toronto and Etobicoke is not big. However, shelters in Vaughan are more likely to experience not fully occupied situations. The location results indicate differences in shelter occupancy across different cities.

In the sector variable, family is the control group. There are no large differences observed when comparing sectors with men and sectors with youth to the control group. However, the coefficient for sectors with women is 0.965 with a standard deviation of 0.14, indicating a higher probability of full occupancy in women's shelters. This suggests a potential issue of inadequate supply in women's sector shelters. Furthermore, the coefficient for mixed adult shelters is -0.285, and the result indicates a lower probability of full occupancy in mixed adult shelters.

Regarding capacity type, there is no difference observed between room-based capacity and bed-based capacity. However, the coefficient for transitional shelters is -2.116 with a standard deviation of 0.056, indicating that emergency shelters are more likely to be fully occupied.

Figure 2 displays the credible intervals for different coefficients. Credible intervals located to the right of x = 0 indicate that the corresponding factor increases the probability of shelters being fully occupied. When the credible interval includes x = 0, it suggests that the effect of this factor can be disregarded. We observe that Scarborough and North York are both positioned to the right of 0, with the credible interval for North York being further from 0. Additionally, the Women Sector is also located to the right of x = 0. On the other hand, the Transitional program model and the city of Vaughan are situated to the left of x = 0. The results in Figure 2 are consistent with those in Table 1, providing a clearer depiction of the impact of each factor on the probability of shelters being fully occupied.

5 Discussion

5.1 Large capacity shelter is more likely to be fully occupied

The shelter capacity size is one of the most important indicators of shelter services. A larger shelter capacity means people have more opportunities to choose, so homeless individuals prioritize larger shelters when making their choices. However, if everyone believes that larger shelters offer more opportunities, it leads to an increase in demand for larger shelters, thereby

Table 2: Results of logistic regression on the full occupancy

	full occupancy
(Intercept)	2.018
	(0.205)
Capacity	0.002
	(0.001)
LOCATION_CITYNorth York	0.806
	(0.194)
LOCATION_CITYScarborough	0.385
	(0.185)
LOCATION_CITYToronto	−0.122
	(0.153)
LOCATION_CITYVaughan	−3.104
	(0.289)
SECTORMen	0.168
	(0.118)
SECTORMixed Adult	−0.285
	(0.114)
SECTORWomen	0.965
	(0.140)
SECTORYouth	−0.179
	(0.129)
CAPACITY_TYPERoom Based Capacity	−0.091
	(0.096)
PROGRAM_MODELTransitional	−2.116
	(0.056)
Num.Obs.	13 759
R2	0.199
Log.Lik.	−5398.160
ELPD	−5409.5
ELPD s.e.	68.7
LOOIC	10 819.0
LOOIC s.e.	137.4
WAIC	10 819.0
RMSE	0.35

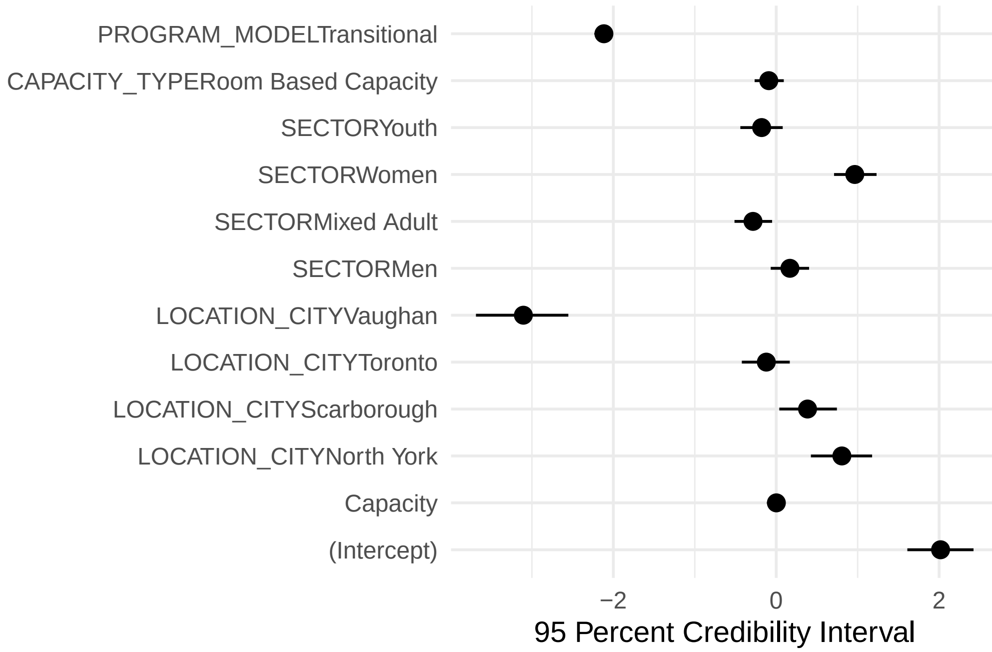

Figure 2: Credible intervals for predictors of shelter fully occupied

increasing the probability of larger shelters being fully occupied. Therefore, when homeless individuals consider choosing shelters, they should not solely prioritize shelter size but should consider other factors such as geographic location and other characteristics of the shelter.

5.2 The uneven distribution of shelters in cities may contribute to the wastage of shelter resources

From Table 1, we can see that shelters in Toronto account for over 80%. However, the results from Table 2 and Figure 2 indicate potential wastage of shelter resources in Toronto compared to Scarborough and North York. Conversely, although shelters in Vaughan represent less than 1% of the total, their estimated coefficients are much lower than 0, suggesting resource wastage. Meanwhile, Scarborough and North York experience resource shortages. Due to geographical differences, homeless individuals tend to reside in densely populated shelter areas. Despite vacancies in Vaughan shelters, their limited capacity easily leads to resource scarcity due to external factors. Conversely, in densely populated shelter areas, even if one shelter is fully occupied, homeless individuals can seek alternatives nearby. Toronto, with its dense concentration of shelters, harbors a large homeless population, posing security risks to the city. To address this issue, the government should consider distributing shelters across different cities to prevent the concentration of homeless individuals and mitigate the scarcity of shelter resources, thereby attracting homeless individuals for placement.

5.3 There is a shortage of shelters for women

There is a clear shortage of shelters for women. Table 1 indicates that the percentage of shelter sector for women is 17.2%, while the percentage for men is 27.8%. This imbalance in shelter services provided to women and men is evident. Single-gender shelters are perceived as safer by homeless individuals, especially for women, who prefer staying in shelters specifically for women rather than mixed adult shelters. Therefore, in Table 2 and Figure 2, the coefficient for women shows a significant positive trend. Since the coefficient for mixed adult shelters is negative, indicating an oversupply of mixed adult shelters, the government can consider converting some mixed adult shelters into shelters specifically for women to more effectively utilize the shelter system.

5.4 Emergency shelters is more important then Transitional shelters

The proportion of Emergency program shelters has reached 82% in@tbl-table1 , but Table 2 shows that Transitional program shelters still have a significant probability of not being fully occupied. Transitional program shelters require permits for application, which many homeless individuals may not meet. Therefore, they can only choose Emergency program shelters, leading to a shortage of resources in Emergency program shelters and an oversupply of resources in Transitional program shelters. To address this issue, the government should reconsider the eligibility criteria for Transitional program shelters and may consider relaxing them to attract more people to use Transitional program shelters.

5.5 Weaknesses and next steps

This study points out that there are uneven distribution of resources in the Toronto shelter system and has proposed some suggestions for the Toronto shelter system. However, the research has some weaknesses. Considering that the shelter system is typically updated daily. Individuals need to reapply for continued stay. There is a risk of dependency because people like to repeat applications for the same shelter. This could result in correlated observations between observations, violating the independence assumption of the logistic regression model. Additionally, the data used in this study is unverified, which may introduce systematic bias during data entry, leading to inaccurate estimations. Therefore, in the next step, it is crucial to conduct data verification to ensure that the collected data reflects the true occupancy of shelters.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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/.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31 (3): 980–91.

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, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.