Forecasting the Toronto Shelter System Performance on Reducing Chronic Homelessness*

Sarah Ding

November 27, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

The dataset used for the current analysis is obtained from Open Data Toronto (Support Services 2024), it is last refreshed on November 15, 2024. Open Data Toronto is an open source accessible to the public,

We use the statistical programming language R (R Core Team 2023)…. Our data (**shelter?**)…. Following Alexander (2023), we consider…

Overview text

*Code and data are available at: [https://github.com/sarahdingg/Toronto_Shelter_System_Forecast).

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

2.3 Variables

The collected data from Open Data Toronto (Support Services 2024) contains several key variables relevant to the analysis:

- **'date'**: The time period (month/year) for which the data is being published (format: mmm-yy)
- **'population_group'**: Represents different groups: all population, chronic, refugees, families, youth, single adult and non-refugees
- **'returned_to_shelter'**: People who were previously using the shelter system, then did not use the system for 3 months or longer, and have now returned. Some other communities may call this indicator "Returned from Inactive"
- **'newly_identified'**: People who entered the shelter system for the first time
- **'moved_to_housing'**: People who were using the shelter system and have moved to permanent housing.
- **'actively_homeless'**: People who have used the shelter system at least one time in the past three months and did not move to permanent housing.
- **'age'**: People's age recorded in the Shelter Management Information System, with ranges of under 16, 16-24, 25-34, 35-44, 45-54, 55-64, and over 65.
- **'population_group_percentage'**: Represents the percentage of each population group (chronic, refugees, families, youth, single adult and non-refugees). The proportion is calculated from all the population.

For the purpose of the current analysis, only the chronic population group is included out of all population groups - for more details on the data cleaning process, refer to Appendix - B.1

2.4 Outcome variables

As the focus is to analyze and predict the performance of Toronto's shelter system in reducing homelessness, the primary outcome variable of interest is the number of individuals who have successfully transitioned to permanent housing, denoted as **'moved_to_housing'**. The **'moved_to_housing'** variable captures the count of people who exited the shelter system and moved into stable, permanent housing arrangements. This variable serves as a direct measure of the shelter system's success in achieving its goal of reducing homelessness to rare, brief, and non-recurring instances. By examining and modeling the **'moved_to_housing'** variable, we can explore effectiveness of the shelter system and identify trends and predictors that influence successful housing transitions.

Table 1: Yearly Averages of Individuals Moved to Permanent Housing

Table 1: Average Moved to Housing by Year

year	average_moved_to_housing
2018	208
2019	277
2020	281
2021	162
2022	220
2023	330
2024	255

Table 1 summarizes the yearly averages of individuals successfully transitioned to permanent housing within Toronto's shelter system from 2018 to 2024. The data reflects consistent efforts in housing transitions, with yearly averages ranging from 162 to 330. The highest average number of moves occurred in 2023, while the lowest was recorded in 2021. These figures provide an overview of the shelter system's progress over time in reducing homelessness. It is important to consider that fluctuations may reflect external factors such as policy changes, economic conditions, or the impact of the COVID-19 pandemic. By examining these trends, the table offers an overview of the shelter system's performance and identifying areas for targeted improvements.

2.5 Predictor variables

The key predictor variables listed below are used in this current analysis to predict the outcome variable: **'age'**, **'date'**, **'age_time_interaction_term'**(constructed), **'returned_to_shelter'**, **'newly_identified'**, **'actively_homeless'**, and **'population_group_percentage'**.

2.5.1 Age

'age' serves as a predictor variable to capture differences in housing outcomes across life stages. The dataset includes age groups such as `ageunder16`, `age16-24`, `age25-34`, `age35-44`, `age45-54`, `age55-64`, and `age65over`, allowing for an understanding of how individuals in different age ranges interact with the shelter system. Each age group provides a window into potential barriers or facilitators that influence transitions to permanent housing. For instance, younger individuals in the `age16-24` range might face hurdles related to education or early career instability, while older adults in the `age65over` group may experience challenges stemming from health concerns or reduced mobility. By incorporating age into the model, the analysis can account for these differences and examine how housing transitions may be shaped by the specific needs and characteristics of each group.

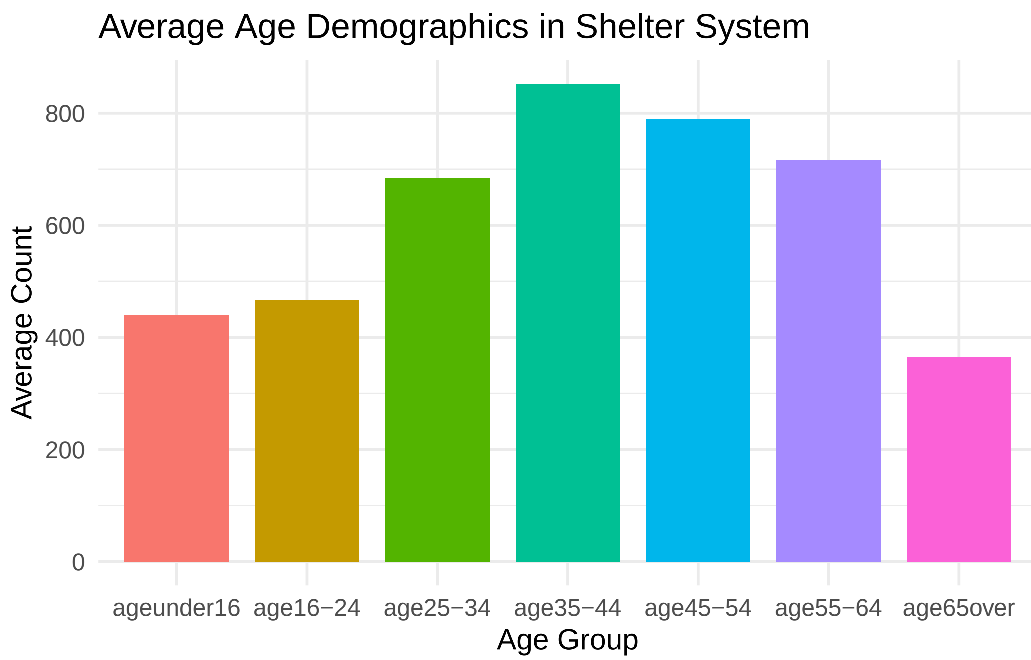

Figure 1: Displays the average number of individuals in each age group using the shelter system from 2019 to 2024. The 35-44 age group has the highest count of observations while 65 and over age group has the lowest count of observations.

Figure 1 illustrates the average distribution of individuals across various age groups within the shelter system. The data shows that middle-aged individuals, particularly those in the `age35-44` and `age45-54` groups, represent the largest average counts, suggesting that these age brackets may face greater challenges in transitioning out of the shelter system. Conversely, individuals in the `ageunder16` and `age65over` categories exhibit lower averages, indicating either fewer individuals from these age groups rely on the shelter system or their stays are less prolonged. The `age16-24` group has moderate representation, possibly reflecting unique barriers related to early adulthood, such as limited resources or support systems. These trends highlight the importance of age-specific interventions and support services to address the distinct needs of each demographic within the shelter system.

2.5.2 Date

'date' is a key predictor variable in this analysis, representing the temporal aspect of transitions within the shelter system. In the dataset obtained from Open Data Toronto (Support Services 2024), variable **'date'** is recorded and refreshed monthly, ranging from January 2019 to October 2024, and is last refreshed on November 15, 2024. By including the date variable, we can capture trends, seasonal effects, and long-term changes that influence the likelihood of individuals moving to permanent housing. For instance, systemic improvements, policy changes, or external factors such as economic conditions and the COVID-19 pandemic may impact shelter system outcomes over time. As a continuous variable, **'date'** can help identify whether the overall performance of the shelter system improves, stagnates, or declines as time progresses. When used in combination with interaction terms, date also allows for the examination of how specific subgroups, such as different age categories, experience changing outcomes over time, offering a dynamic perspective on the shelter system's evolution.

2.5.3 Age & Time Interaction Term

'age_time_interaction_term' is a constructed predictor variable that explores how the relationship between other predictor variables and the outcome variable evolves over time. Specifically, the interaction terms between age groups (e.g., `age16-24`, `age25-34`) and time (`date`) allow us to investigate whether the likelihood of transitioning to permanent housing varies across age demographics as time progresses. These terms are particularly useful for identifying trends that might not be apparent when considering age or time alone. For instance, a positive interaction term for `age16-24` and `date` would indicate that younger individuals are becoming more likely to move to permanent housing over time, potentially reflecting the success of targeted interventions. By examining these interaction terms, the analysis captures how age and time jointly influence housing outcomes, providing a deeper understanding of patterns within the shelter system.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix C.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

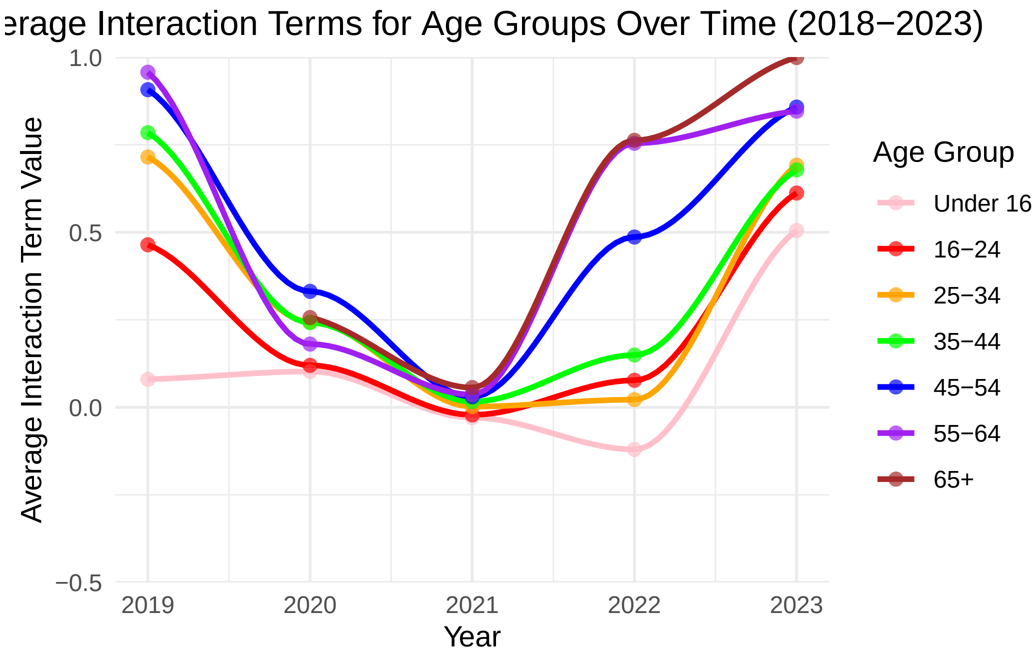

Figure 2: Trends in Average Interaction Terms for Age Groups Over Time (2018–2023)

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in **?@tbl-modelresults**.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 What did we learn about the world? talk about differences before pandemic and after

Which age groups benefit most over time from system improvements. Age groups that require more attention if their outcomes worsen over time. Policy implications, such as whether programs targeted at specific age groups need to be adjusted to meet changing needs.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

A Appendix 1

B Additional data details

how data is cleaned for analysis data and model data, why lagged variables are constructed for time-series analysis and regression models

B.1 Data Cleaning

B.2 Limitations and Improvements

C Model details

C.1 Model Performance

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows…

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows…

C.2 Model Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows… This suggests…

?@fig-stanareyouokay-2 is a Rhat plot. It shows… This suggests…

D Apendix 2

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Support Services, Toronto Shelter &. 2024. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://open.toronto.ca/dataset/toronto-shelter-system-flow/.