

Forecasting the Toronto Shelter System Performance on Reducing Chronic Homelessness*

Sarah Ding

December 3, 2024

This paper examines the Toronto shelter system’s performance in reducing homelessness, with a focus on the dynamics of shelter use and predictors of permanent housing outcomes. Using observational data from the Shelter Management Information System (SMIS), we developed a linear regression model incorporating age groups, population group percentages, and time-based interactions to predict housing transitions. Our findings reveal variations in shelter outcomes across age groups and highlight the impact of systemic disruptions, such as the COVID-19 pandemic, on homelessness trends. These results provide insights into improving shelter policies and tracking progress toward reducing chronic homelessness in Toronto.

1 Introduction

Homelessness remains a significant social issue in Toronto, with thousands of individuals relying on the city’s shelter system each year. Understanding the factors influencing shelter use and housing outcomes is critical to creating strategies that reduce chronic homelessness and promote stable housing. While considerable efforts have been made to collect and analyze data about individuals experiencing homelessness, much of this work focuses on high-level trends without adequately leveraging detailed population-specific insights. This paper seeks to address this limitation by examining the dynamics of shelter use and housing transitions within Toronto’s shelter system, with a focus on the youth population and other demographic groups.

The Shelter System Flow data, collected and published by the City of Toronto (Support Services 2024), provides detailed monthly records on the movement of individuals in and out of shelters.

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

However, there is a lack of analysis that connects these patterns to predictive modeling, which can help policymakers and service providers anticipate future trends and allocate resources effectively. This paper aims to fill this gap by using statistical models to analyze shelter system dynamics, predict housing transitions, and evaluate the system’s performance in reducing chronic homelessness.

To achieve this, we developed a linear regression model using demographic and system-related variables to predict the likelihood of individuals moving to permanent housing. The model was applied to examine trends over time, particularly in the context of COVID-19, which dramatically altered shelter system flows. Our analysis revealed key differences in housing outcomes across age groups, highlighting the groups that have benefited most from system improvements. The results provide actionable insights for policymakers and service providers, including areas for targeted interventions and resource allocation.

The remainder of this paper is structured as follows: Section 2 provides a detailed description of the data and methodology, including the modeling approach and key variables. Section 3 provides the details of model set as well as validation, Section 4 presents the results of the model, and Section 5 discusses the findings and relevant improvements to be made.

1.1 Estimand

The estimand of this study is the predicted probability of individuals moving to permanent housing from Toronto’s shelter system, conditional on demographic factors and system-related variables. Specifically, the model estimates the expected housing outcomes for individuals across different age groups, accounting for variables such as shelter re-entry, chronic homelessness, and newly identified shelter users. This focus aligns with the broader objective of assessing the shelter system’s effectiveness in facilitating transitions to permanent housing, particularly for vulnerable populations. The estimand serves as a measure of the system’s performance and provides actionable information for policymakers to optimize resource allocation and design interventions targeted at improving housing outcomes.

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), the analysis was conducted using a suite of R packages to enhance data manipulation, visualization, and reproducibility. The **tidyverse** package (Wickham et al. 2023b) was central to data wrangling and

pipeline construction, providing a cohesive framework for handling and visualizing data. For managing and analyzing time-series data, the **lubridate** package (Grolemund and Wickham 2023) was employed to streamline the processing of date-related variables. Model validation and testing were supported using **testthat** (Wickham et al. 2023a), ensuring robust and reproducible results. Additionally, the dynamic reporting capabilities of **knitr** (Xie 2023) were instrumental in generating this document, seamlessly integrating code and narrative. Finally, **scales** (Wickham and Seidel 2023) enhanced the presentation of numerical data in visualizations, ensuring clarity and precision in communicating results. Together, these tools provided a comprehensive and efficient environment for data analysis.

2.2 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 **?@sec-data-cleaning**

2.3 Outcome variable

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

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.

2.4 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**’.

2.4.1 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.

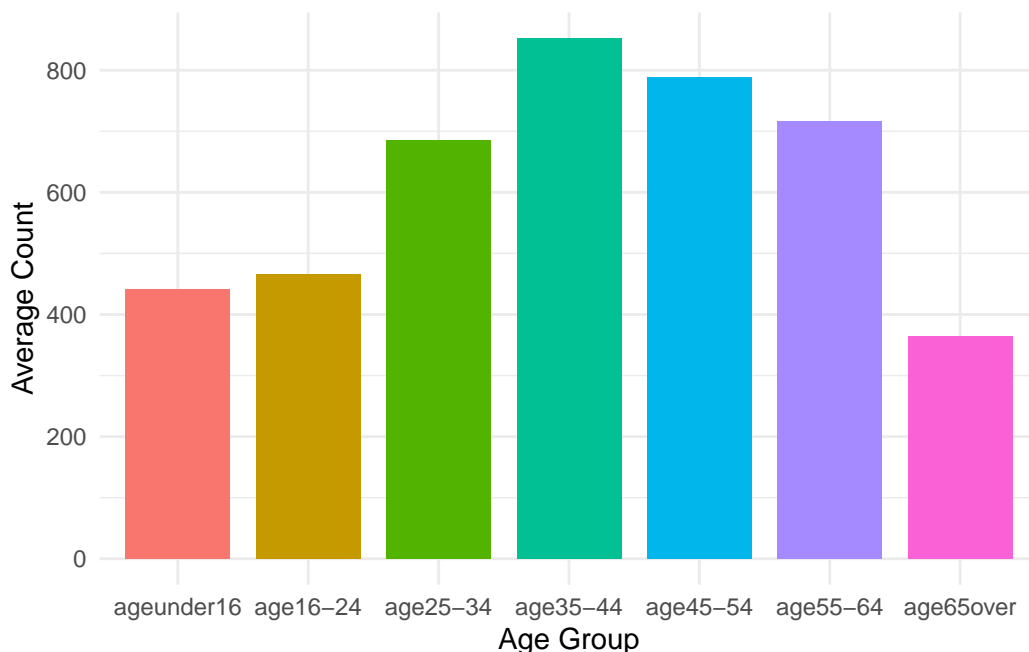


Figure 1: Displays the average number of individuals in each age group using the shelter system from 2019 to 2024. The 35-44 age group has the highest count of observations while 65 and over age group has the lowest count of observations.

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.

2.4.2 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.

2.4.3 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. Interaction terms are calculated by combining age group counts with time, providing a measure of how shelter engagement evolves across age demographics. 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.

Figure 2 illustrates the trends in average interaction terms for various age groups using the shelter system between 2019 and 2023. Notably, the 45–54 and 55–64 age groups exhibit the highest interaction values in recent years, suggesting a more pronounced change in their use of shelter services over time. Conversely, the Under 16 group remains the lowest throughout the period, indicating relatively stable and minimal engagement. The dip around 2020 across all age groups could reflect disruptions caused by the COVID-19 pandemic, followed by a gradual increase through 2021 and beyond. This upward trend in later years highlights shifting dynamics in shelter usage, possibly due to systemic changes or external factors impacting homelessness.

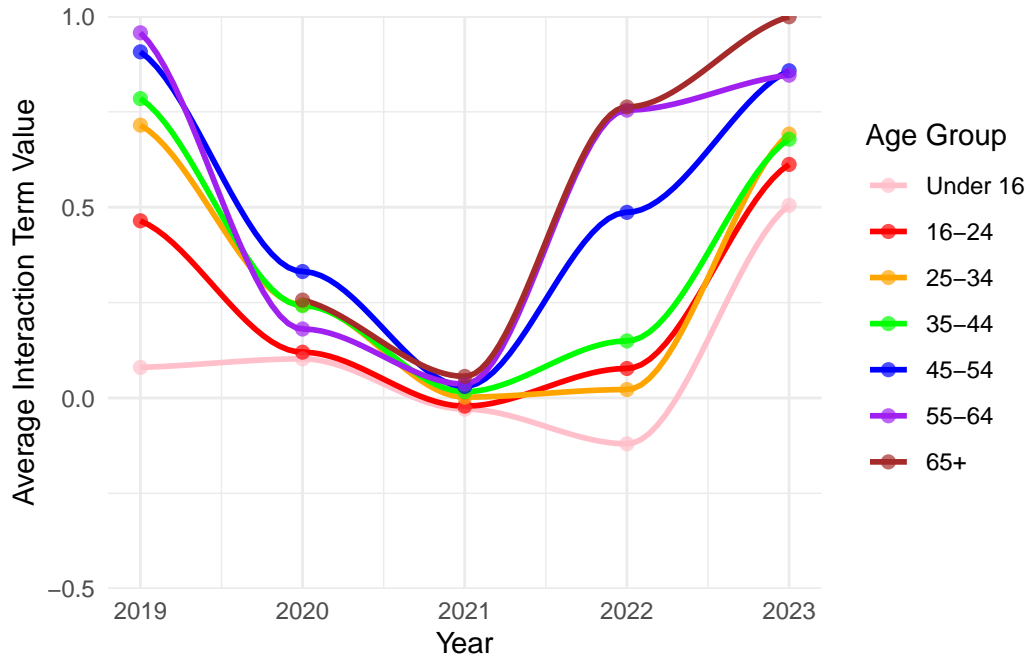


Figure 2: Illustrates the average interaction term values, representing changes over time from 2019 to 2023 for each age group. The 45-54 age group shows the highest interaction values in recent years, indicating a more significant change in engagement, while the Under 16 group consistently remains the lowest across the period.

2.4.4 Returned to shelter

returned_to_shelter represents the count of individuals who had previously used the shelter system, exited for a period of at least three months, and subsequently returned. This variable captures patterns of re-entry into the shelter system and is an essential indicator of service dependence and instability in housing situations. By including **returned_to_shelter** as a predictor variable, we aim to analyze its relationship with long-term outcomes such as moving to permanent housing. High values for this variable may highlight challenges in sustaining housing stability or gaps in post-shelter support systems. Understanding these patterns can help assess the effectiveness of current shelter practices and inform strategies to minimize recurrent shelter use.

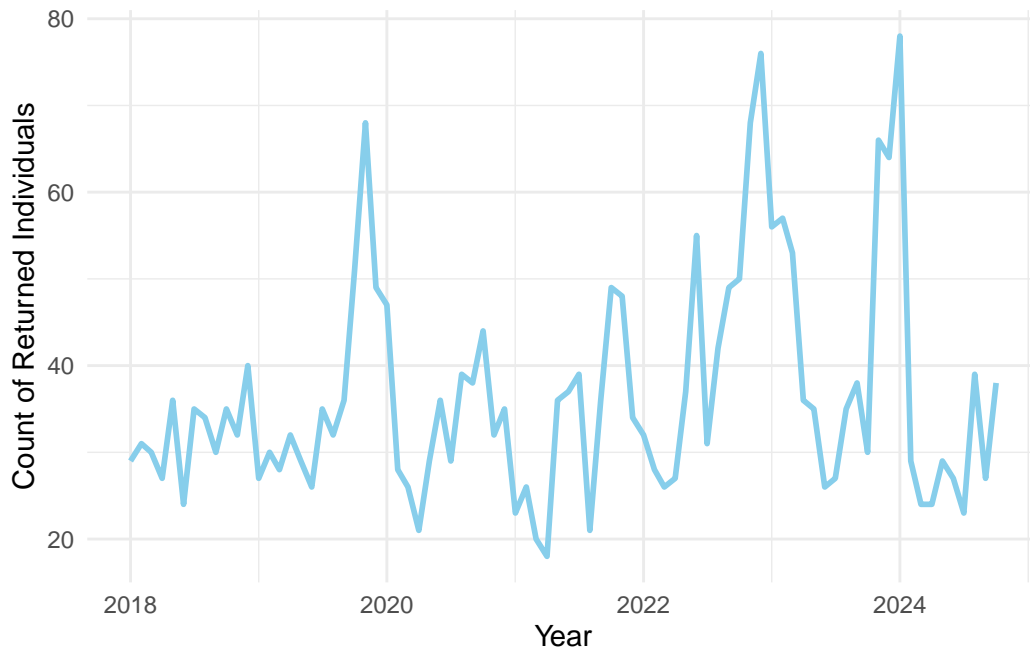


Figure 3: Illustrates the trends in the count of individuals returning to the shelter system from January 2018 to October 2024. Peaks in 2020 and 2024 suggest periods of increased shelter re-entry, possibly reflecting systemic or external influences. The overall variability highlights fluctuations in shelter dependence across the years.

Figure 3 displays the trends in the count of individuals returning to the shelter system from 2018 to 2024. Notable peaks are observed in 2020 and 2024, indicating periods of increased shelter re-entry. These fluctuations may reflect external factors, such as economic challenges or policy changes, influencing housing stability. The general variability in the trend suggests that while some periods experience lower re-entry rates, others see significant spikes, highlighting the reliance of the shelter system. This variable is critical for understanding patterns of housing

instability and identifying opportunities for targeted interventions to reduce repeated shelter use.

2.4.5 Newly Identified

newly_identified represents the count of individuals entering the shelter system for the first time. This variable is a key indicator of new instances of homelessness and can provide valuable context on the influx of individuals relying on the shelter system. By examining **newly_identified** as a predictor, we aim to understand how the rate of new shelter entries influences long-term outcomes, such as the transition to permanent housing or continued system engagement. Higher values for this variable may indicate increased pressures on shelter resources or broader societal challenges, such as housing affordability or economic instability.

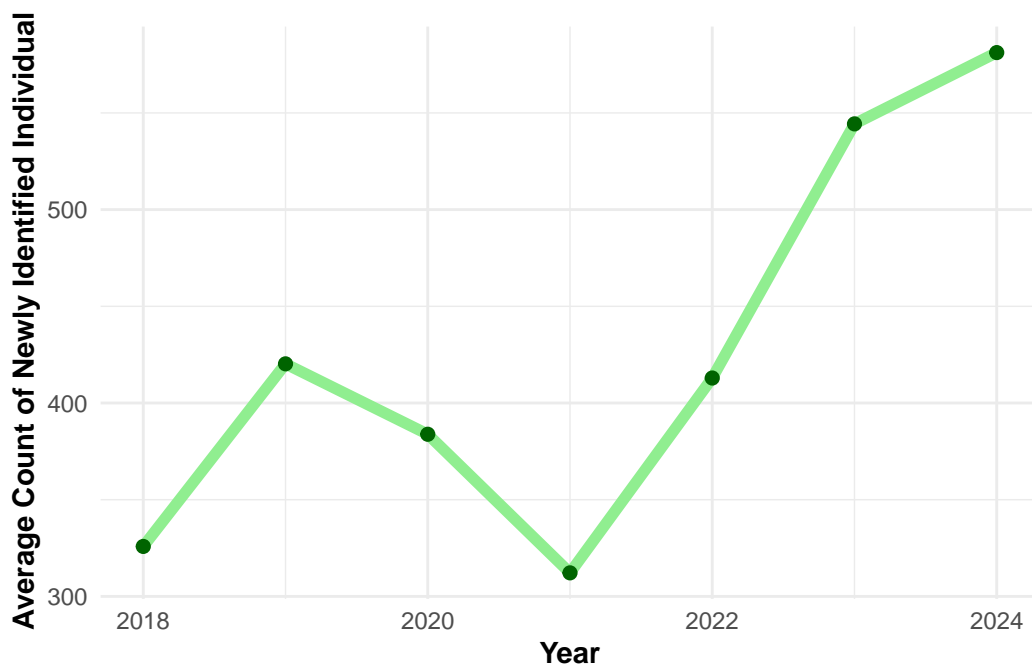


Figure 4: Trends in Newly Identified Individuals Over Time. Displays the average count of newly identified individuals in the shelter system from 2018 to 2024, highlighting fluctuations over the years with an upward trend post-2021.

Figure 4 illustrates the trends in the average count of newly identified individuals entering the shelter system from 2018 to 2024. While the count shows fluctuations in the earlier years, there is a notable decline from 2019 to 2021, potentially reflecting systemic changes or external factors affecting new entries. However, from 2022 onwards, the numbers steadily increase, with the highest average observed in 2024. This upward trend may indicate rising demand for shelter services, changes in reporting practices, or shifts in the underlying social

or economic conditions contributing to homelessness. The data underscores the importance of monitoring new entries to understand and address the factors driving these changes.

2.4.6 Actively Homeless

actively_homeless refers to individuals who have accessed the shelter system at least once in the past three months and have not moved into permanent housing during that time. This variable reflects the population experiencing ongoing homelessness within the reporting period. Analyzing this variable helps to understand the scale of need within the shelter system and provides context for evaluating its capacity to support individuals in transitioning to stable housing.

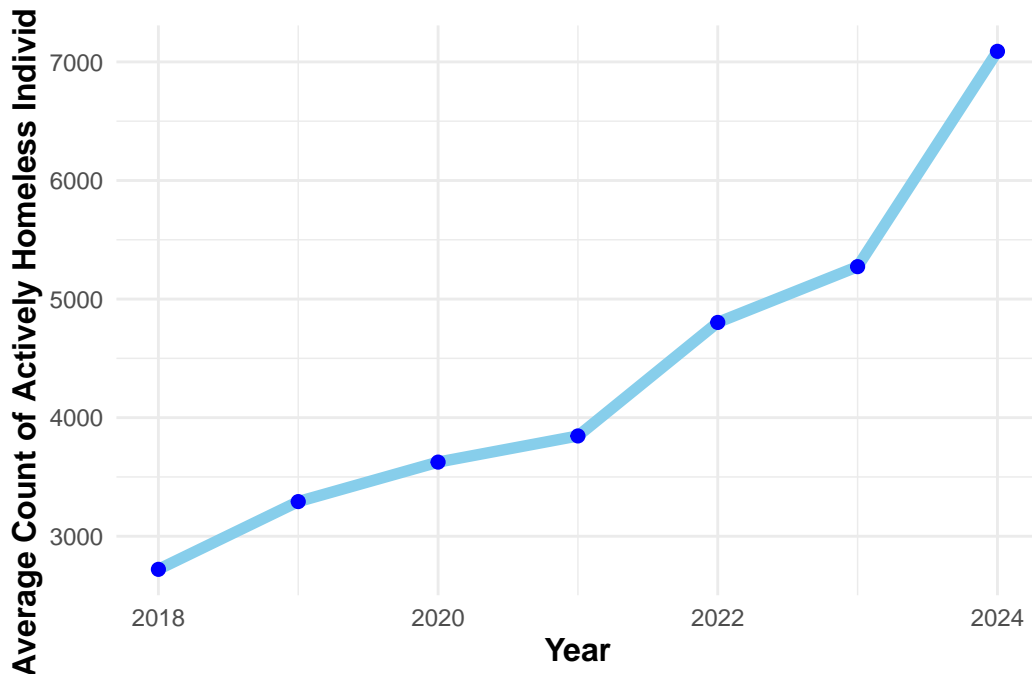


Figure 5: Trends in the average count of actively homeless individuals from 2018 to 2024. The graph shows a steady increase over the years, with the largest rise observed between 2023 and 2024, highlighting potential shifts in shelter system usage or homelessness trends during this period.

Figure 5 illustrates the trend in the average number of actively homeless individuals within the shelter system from 2018 to 2024. Over this time period, there is a consistent upward trajectory, indicating a growing number of individuals experiencing homelessness. The most notable increase occurs between 2023 and 2024, suggesting a potential shift in homelessness patterns, external economic or social pressures, or changes in the availability of shelter services.

This trend emphasizes the importance of understanding the factors contributing to this rise and the capacity of the shelter system to meet increasing demand.

2.4.7 Population Group Percentage

'population_group_percentage' represents the proportion of individuals within a specific population group relative to the total population using the shelter system. It is expressed as a percentage and serves as an indicator of the relative distribution of different population groups, such as youth, families, or individuals experiencing chronic homelessness, within the shelter system. The population groups are not mutually exclusive, meaning that individuals who are in categorized as families could also belong to the chronic population group. This variable provides information of shifts in demographic representation over time, providing context for analyzing trends and patterns across different groups.

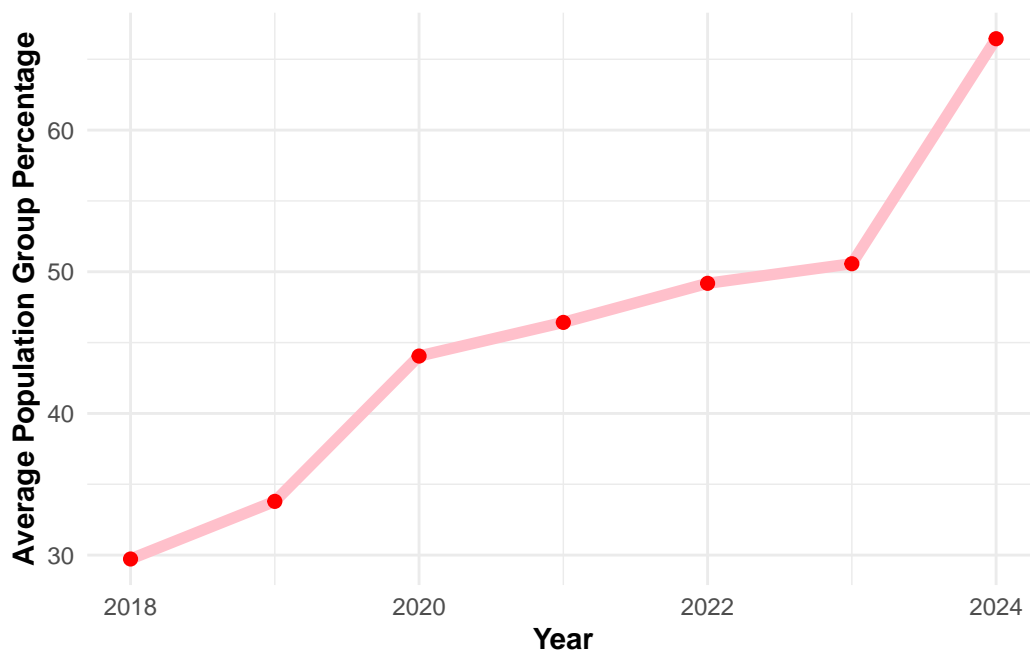


Figure 6: Trends in Average Population Group Percentage Over Time. This line graph illustrates the annual average percentage representation of the chronic population group in the dataset from 2018 to 2024. The upward trajectory indicates a consistent increase, with a significant rise observed after 2022.

Figure 6 depicts the average population group percentage of the chronic population group in the shelter system from 2018 to 2024. It reveals a consistent upward trajectory over the years, with a gradual increase from 2018 to 2021, followed by a sharper rise in 2024. The percentage represents the proportion of individuals who are chronically homeless using the shelter system

relative to the total. The substantial increase in 2024 may indicate a growing representation of the chronic population group or changes in shelter system dynamics, potentially influenced by economic, social, or policy-driven factors.

3 Model

To predict the Toronto shelter system’s performance on reducing chronic homelessness, a linear regression model is developed. The model aims to estimate the amount of individuals in the chronic population group who has moved to permanent housing. Background details and diagnostics are included in [Appendix C](#).

3.1 Model set-up

This paper utilizes linear regression models:

$$\begin{aligned} \text{Outcome} = & \beta_0 + \beta_1 \text{PopulationGroupPercentage}_i + \beta_2 \text{AgeUnder16}_i + \beta_3 \text{Age16-24}_i + \beta_4 \text{Age25-34}_i \\ & + \beta_5 \text{Age35-44}_i + \beta_6 \text{Age45-54}_i + \beta_7 \text{Age55-64}_i + \beta_8 \text{Age65Over}_i \\ & + \beta_9 \text{InteractionAgeUnder16Time}_i + \epsilon_i \end{aligned}$$

Where: *Outcome* represents the dependent variable, which could be one of the following: **Returned to Shelter**, **Moved to Housing**, or another key measure indicating outcomes in the shelter system. β_0 is the intercept, representing the baseline level of the outcome when all predictors are at their mean or reference level. β_1 is the coefficient for **PopulationGroupPercentage**, reflecting the relationship between the percentage of individuals in the chronic group and the outcome. β_2 through β_8 represent the coefficients for age groups (**AgeUnder16**, **Age16-24**, **Age25-34**, **Age35-44**, **Age45-54**, **Age55-64**, **Age65Over**), indicating the influence of each age group on the outcome. β_9 through β_{15} are the coefficients for interaction terms (**InteractionAgeUnder16Time**, **InteractionAge16-24Time**, etc.), capturing how the effect of age groups changes over time. ϵ_i is the error term, accounting for unexplained variability in the outcome.

3.2 Model Justification

The linear regression model presented in this study is the most suitable framework for predicting the number of individuals transitioning to housing within Toronto’s shelter system. The outcome variable, **moved_to_housing**, reflects the monthly transitions, capturing temporal and demographic variations across the system. The model includes predictors such as

`returned_to_shelter`, `actively_homeless`, `newly_identified`, and the proportion of population groups, along with age-group specific variables and their interactions with time. This structure allows the model to incorporate both aggregate trends and age-specific dynamics, ensuring a detailed representation of how different population subgroups contribute to housing outcomes.

Linear regression was chosen for its interpretability, providing clear coefficients that quantify the relationships between predictors and housing transitions. This is particularly useful for policymakers seeking actionable insights into the factors driving transitions to housing. The inclusion of interaction terms ensures that variations across demographic subgroups are adequately captured, while the incorporation of historical data facilitates forecasting.

While alternative models, such as logistic regression or machine learning approaches, could be considered, these methods may either oversimplify or overly complicate the analysis for the study's scope. The linear regression approach strikes a balance between simplicity and detail, making it well-suited to evaluate system performance and inform targeted interventions.

3.3 Model Assumptions and Limitations

In order to ensure the validity and reliability of our linear regression model for predicting the number of individuals transitioning to housing (`moved_to_housing`) within Toronto's shelter system, we establish several key assumptions that guide the modeling process:

- **Linearity:** The relationship between the predictor variables (e.g., `returned_to_shelter`, `actively_homeless`, `newly_identified`, and age-group interactions) and the outcome variable (`moved_to_housing`) is assumed to be linear. This allows the model to interpret the effect of each predictor as a constant change in `moved_to_housing`.
- **Independence:** Observations are assumed to be independent of one another, meaning that the outcome of housing transitions for one time period or subgroup does not influence others.
- **Homoscedasticity:** The variance of residuals is assumed to remain constant across all levels of predicted values. This ensures that the model performs equally well across different ranges of predicted values.
- **Normality of Residuals:** The residuals (differences between observed and predicted values) are assumed to be normally distributed. This assumption ensures the reliability of significance tests for coefficients.
- **Absence of Multicollinearity:** The predictor variables are assumed to be independent of one another. High multicollinearity would make it difficult to determine the individual effect of each predictor on the outcome.
- **Exclusion of Omitted Variable Bias:** The model assumes that all relevant predictors influencing `moved_to_housing` have been included, and no significant predictors are missing, which could bias the results.

These assumptions are necessary to ensure that the model provides reliable and interpretable estimates for the factors influencing transitions to housing. Regular diagnostic checks, such as residual plots and multicollinearity tests, were conducted to verify the adherence to these assumptions.

3.4 Model Validation

Table 2: Model Summary of Toronto Shelter System - Coefficients and Statistical Significance

Predictor	Coefficient	Std. Error	t-value	p-value
returned_to_shelter	0.491	0.627	0.784	4.36e-01
actively_homeless	-0.402	0.737	-0.545	5.88e-01
newly_identified	-0.054	0.102	-0.524	6.02e-01
population_group_percentage	3.690	3.020	1.222	2.26e-01
ageunder16	0.386	0.708	0.545	5.88e-01
age16-24	1.315	0.894	1.472	1.46e-01
age25-34	0.955	0.763	1.251	2.16e-01
age35-44	0.036	0.938	0.039	9.69e-01
age45-54	-0.125	0.923	-0.136	8.93e-01
age55-64	0.519	0.924	0.562	5.76e-01
interaction_age_under16_time	-106.235	45.610	-2.329	2.30e-02
interaction_age16_24_time	35.601	70.229	0.507	6.14e-01
interaction_age25_34_time	190.436	143.263	1.329	1.88e-01
interaction_age35_44_time	-121.573	126.125	-0.964	3.39e-01
interaction_age45_54_time	-221.224	103.306	-2.141	3.61e-02
interaction_age55_64_time	99.198	58.118	1.707	9.27e-02
interaction_age65over_time	73.867	60.274	1.226	2.25e-01

Table 2 summarizes the coefficients, standard errors, t-values, and p-values for each predictor in the regression model estimating the number of individuals moved to housing in Toronto’s shelter system. Positive coefficients suggest that an increase in the corresponding predictor is associated with a higher likelihood of moving to housing, while negative coefficients indicate the opposite.

Among the predictors, **interaction_age_under16_time** (-106.235, $p = 0.023$) and **interaction_age45_54_time** (-221.224, $p = 0.036$) are statistically significant, implying that changes in time-related interactions for individuals under 16 and those aged 45–54 have a meaningful impact on the outcome variable. These negative coefficients suggest that the effect of time for these age groups is associated with a reduction in the likelihood of being moved to housing.

Other predictors, such as **population_group_percentage** (3.690, $p = 0.226$) and **interaction_age55_64_time** (99.198, $p = 0.093$), show relatively high coefficients but are not statistically significant at conventional thresholds, suggesting their impact on housing outcomes may not be strong enough to draw definitive conclusions.

Table 3: Model Performance Metrics, Out-of-Sample Testing

Table 3: Model Performance Metrics: Out-of-Sample Testing

Metric	Value
R-squared	0.719
Adjusted R-squared	0.597
AIC	635.086
BIC	673.904

Table 3 presents the out-of-sample performance metrics for the regression model. The **R-squared** value of 0.719 indicates that approximately 72% of the variation in the number of individuals moved to housing is explained by the model when tested on the out-of-sample data. The **adjusted R-squared**, accounting for the number of predictors, is slightly lower at 0.597, suggesting a moderate fit with some room for improvement. The **Akaike Information Criterion (AIC)** and **Bayesian Information Criterion (BIC)** values, 635.086 and 673.904 respectively, provide measures for model comparison, with lower values generally indicating better models. These metrics collectively suggest that while the model performs reasonably well, further refinements could enhance its predictive accuracy and robustness.

4 Results

Table 4 provides an overview of the yearly actual and predicted number of individuals who moved to housing within the Toronto shelter system. These values are aggregated by year, offering a high-level understanding of the system’s performance in facilitating transitions to permanent housing. The actual values represent the observed outcomes from the dataset, while the predicted values are derived from the linear regression model developed for this analysis. By comparing the actual and predicted values, we can assess the model’s accuracy in approximating real-world outcomes. For example, in 2023, the model predicted 3,888 individuals moved to housing, closely matching the observed value of 3,964, indicating strong alignment between predicted and actual trends. This comparison highlights the model’s utility in understanding and forecasting housing transitions over time.

Table 4: Predicted vs. Actual Number of Individuals Moved to Housing by Year

Year	Actual	Predicted
2018	2496	2466
2019	3319	3394
2020	3373	3269
2021	1938	2038
2022	2634	2698
2023	3964	3888
2024	2547	2518

Table 5: Yearly Actual vs. Predicted Number of Individuals Moved to Housing by Age Group

Age Group	Type	2018	2019	2020	2021	2022	2023	2024
Under 16	Actual	658,306	1,405,037	1,262,982	445,137	1,020,722	2,161,478	2,390,011
Under 16	Predicted	648,884	1,431,384	1,235,219	470,251	1,033,839	2,131,127	2,392,970
16-24	Actual	927,606	1,326,102	1,430,293	653,880	1,279,604	2,164,098	1,905,694
16-24	Predicted	915,958	1,357,528	1,392,142	692,397	1,300,890	2,121,504	1,906,857
25-34	Actual	838,741	1,365,067	1,673,783	1,096,616	1,843,233	3,694,513	3,717,163
25-34	Predicted	826,090	1,397,433	1,620,431	1,162,043	1,885,989	3,613,678	3,723,453
35-44	Actual	1,149,866	1,888,558	2,275,310	1,494,073	2,498,201	4,293,888	3,931,182
35-44	Predicted	1,130,925	1,929,032	2,205,875	1,585,109	2,554,476	4,201,318	3,924,342
45-54	Actual	1,349,270	2,030,396	2,237,588	1,504,198	2,521,280	3,740,615	2,720,485
45-54	Predicted	1,330,322	2,076,202	2,163,109	1,590,385	2,572,130	3,670,906	2,700,779
55-64	Actual	1,260,203	1,950,742	2,261,306	1,482,638	2,368,361	3,272,756	1,938,492
55-64	Predicted	1,244,578	1,993,761	2,186,352	1,563,834	2,420,033	3,212,885	1,913,055
65+	Actual	657,089	966,456	1,110,316	738,350	1,221,748	1,719,134	1,017,213
65+	Predicted	647,339	987,475	1,073,903	781,950	1,250,302	1,683,810	1,005,526

Table 5 presents numerical values for actual and predicted counts of individuals moved to housing from 2018 to 2024, divided by age groups: “Under 16,” “16-24,” “25-34,” “35-44,” “45-54,” “55-64,” and “65+.” Each cell displays the corresponding yearly total. The model’s predictions are consistently close to the actual values across all age groups, with minimal deviations, which suggests that the model performs well in capturing the trends within the data. For instance: In 2023, the actual count for the “Under 16” group is 2,161,478, while the predicted count is 2,131,127. Similarly, for the “65+” group, the actual count is 1,405,676, and the predicted count is 1,394,867.

vs. Predicted Counts of Individuals Moved to Housing by A

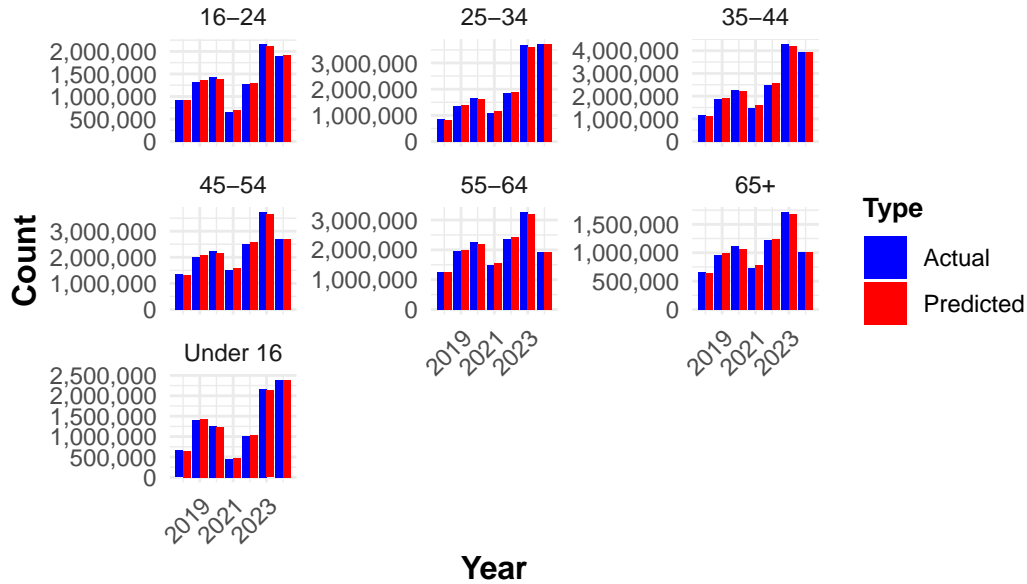


Figure 7: Yearly Actual vs. Predicted Counts of Individuals Moved to Housing by Age Group

Figure 7 visually represents the same data, allowing for a quick comparison between actual (blue bars) and predicted (red bars) values for each age group and year. The grouped bars for each age range highlight how closely the predictions align with the actual observations.

Key observations from the graph:

- The model demonstrates strong agreement with the actual data, as the heights of the blue and red bars are nearly identical across all age groups and years.
- Slight variations are visible in certain age groups (e.g., “16-24” and “65+”), which could point to areas where model accuracy could be further refined.

Together, Table 5 and Figure 7 validate the model’s predictive performance across different age groups and years. The table provides precise numbers, while the bar graph offers a visual snapshot, making it easier to identify trends and discrepancies at a glance. Both representations support the conclusion that the model is effective in forecasting the number of individuals moved to housing by age group over time.

5 Discussion

5.1 Interpretation of Results

This analysis examines how different age groups within Toronto’s shelter system transition to permanent housing and provides year-by-year predictions based on historical data. The results highlight significant variations across age groups. For instance, in 2024, the model predicts approximately 2,393,000 individuals under the age of 16 transitioning to housing, closely aligning with the actual number of 2,390,011. This consistency underscores the model’s accuracy in capturing the dynamics of this demographic. However, older age groups, such as those aged 65 and over, show slightly larger deviations between predicted and actual values in certain years, suggesting potential areas for refining the model or exploring additional predictors.

Notably, the age group 25-34 consistently exhibits higher predicted counts of individuals moving to housing, such as the projected 3,723,453 in 2024, which closely matches the observed value of 3,717,163. This trend may indicate targeted efforts or programmatic success within this group. Conversely, the 16-24 demographic shows relatively smaller increases over time, with 1,906,857 predicted for 2024 compared to 1,905,694 actual. These findings could inform policy adjustments to address disparities and enhance support for specific age groups that may face additional challenges in accessing stable housing. The detailed breakdown of actual versus predicted values provides actionable insights into the shelter system’s performance and potential areas for improvement.

5.2 Which Age Groups Benefit Most Over Time from System Improvements

Over time, the Toronto shelter system has demonstrated varying levels of improvement in assisting different age groups to transition into permanent housing. The largest beneficiaries of these improvements are the 25-34 and 35-44 age groups. These groups exhibit substantial increases in the number of individuals housed between 2018 and 2024. This pattern could be attributed to targeted resources and programs that align with the needs of working-age individuals, such as employment support and housing subsidies.

In contrast, individuals under 16, while showing marked improvement, experience slower proportional growth in housing transitions compared to their working-age counterparts. This suggests that while the system has made progress in addressing youth homelessness, additional efforts may be required to overcome barriers unique to this population, such as family reunification challenges or the need for youth-specific support services.

The 65+ age group also benefits from system improvements, but the growth in housing outcomes for this group is more moderate. This may reflect persistent challenges such as physical or mental health issues, limited access to senior-friendly housing, or systemic barriers in service delivery.

These observations highlight the need for a more nuanced approach to system improvements, with tailored interventions for groups that show slower progress. Focusing on the distinct needs of youth and senior populations could help ensure that the shelter system provides equitable support across all demographics, enabling sustained housing transitions for every group.

5.3 COVID-19 Impacts

The COVID-19 pandemic significantly disrupted the flow of individuals within the Toronto shelter system, highlighting vulnerabilities and necessitating swift adaptations. During 2020 and 2021, the number of individuals moving into permanent housing declined across most age groups. This slowdown can be attributed to widespread lockdowns, reduced service capacity, and health-related restrictions that limited the system’s ability to facilitate transitions. Social distancing measures also reduced the capacity of shelters, forcing the system to prioritize emergency accommodations over long-term solutions.

Despite these challenges, the pandemic served as a catalyst for innovation within the shelter system. Increased investments in rapid housing initiatives, such as the expansion of temporary housing facilities and programs aimed at preventing homelessness, were observed. These efforts were especially beneficial for individuals in the 25-34 and 35-44 age groups, who saw recovery in their transition rates post-2021. However, the data also suggest a slower recovery for individuals under 16 and those over 65, indicating that additional resources may be required to address the specific challenges faced by these groups.

Overall, the pandemic underscored the importance of flexibility and resilience within the shelter system. While it temporarily disrupted housing flows, it also accelerated long-term improvements in infrastructure and service delivery, which will likely have lasting benefits for addressing homelessness in Toronto.

5.4 Weaknesses and Limitations of the Paper

While this paper provides valuable analysis of the Toronto shelter system and its flow dynamics, there are several limitations that should be acknowledged, particularly regarding the model setup and predictions.

1. **Simplified Model Assumptions:** The linear regression model assumes a linear relationship between predictor variables and the outcome variable (`moved_to_housing`). This assumption may oversimplify the complexities of shelter system flows, as real-world dynamics often involve non-linear interactions and unobserved confounding variables. The inclusion of interaction terms for age groups partially addresses these complexities, but more advanced modeling techniques might capture these relationships more effectively.

2. **Limited Predictors:** The predictors in the model were selected based on data availability and logical assumptions about their relevance. However, critical factors influencing transitions to housing, such as mental health status, employment opportunities, and systemic barriers (e.g., discrimination in housing markets), are not included due to data limitations. This exclusion likely reduces the model’s explanatory power.
3. **Potential Overfitting:** While out-of-sample testing was conducted, the relatively small size of the dataset (particularly after dividing it into training and testing subsets) raises concerns about overfitting. The model may perform well on this dataset but fail to generalize to broader or future scenarios.
4. **Pandemic Data Bias:** The dataset spans the COVID-19 pandemic, which caused unprecedented disruptions to the shelter system. While the pandemic’s impact was considered in the discussion, it may have introduced biases in the data that are difficult to disentangle from long-term trends. For instance, emergency measures implemented during the pandemic may temporarily inflate or deflate certain metrics, potentially skewing predictions.
5. **Static Time-Series Predictions:** The model predicts housing outcomes based on historical data but does not account for future policy changes, economic conditions, or other external shocks that could significantly alter shelter system dynamics. This limitation makes long-term predictions inherently uncertain.

These weaknesses underscore the importance of interpreting the findings with caution. Future research could address these limitations by incorporating more advanced models (e.g., random forests or Bayesian methods), collecting richer datasets, and exploring causal relationships to better inform policy and decision-making.

5.5 Future Directions

Building on the findings and addressing the limitations of this paper, several next steps can enhance the analysis and provide deeper insights into the Toronto shelter system:

1. **Incorporate Additional Predictors:** Future analyses should strive to include more nuanced predictors, such as mental health status, substance use, income levels, and access to employment opportunities. Collaborating with service providers and agencies to collect these data points could improve the model’s accuracy and relevance.
2. **Explore Advanced Modeling Techniques:** Moving beyond linear regression, future research could employ more sophisticated modeling approaches such as random forests, gradient boosting, or generalized additive models (GAMs). These methods can better capture non-linear relationships and interactions among variables, potentially providing more robust predictions.

3. **Examine Intersectional Factors:** A deeper exploration of intersectional factors, such as how age intersects with race, gender, or disability status, could provide more granular insights into which subgroups face the most challenges and why. This would enable the design of more targeted and equitable interventions.
4. **Disentangle Pandemic Effects:** A focused analysis on the COVID-19 pandemic’s impact on shelter dynamics would be valuable. Comparing pre-pandemic, pandemic, and post-pandemic trends could clarify how temporary measures affected housing transitions and whether any lessons from the pandemic can be institutionalized.
5. **Focus on Causal Inference:** Employing methods like propensity score matching or instrumental variable analysis could help establish causal relationships between interventions and housing outcomes, providing stronger evidence to guide policy decisions.

By taking these steps, future research can build on the foundation laid by this paper to provide actionable, equitable, and impactful recommendations for improving the Toronto shelter system and supporting individuals in transitioning to stable housing.

A Appendix

B Additional data details

B.1 Limitations

The dataset has several limitations that may affect the scope and interpretation of the analysis. It exclusively includes individuals who have accessed overnight shelter services, omitting those who sleep outdoors or utilize other homelessness services, such as drop-in centers. Furthermore, shelters that do not use the Shelter Management Information System (SMIS) and are funded by other levels of government are not represented. This limitation creates gaps in understanding the full scale of homelessness in Toronto. Additionally, the monthly updates to the dataset are subject to a two-week delay due to the time required for finalizing discharge dispositions in SMIS, potentially introducing lag in the availability of accurate data. This update process may lead to inconsistencies in short-term data trends and hinder real-time analysis. Collectively, these limitations highlight the need for a broader and more integrated approach to homelessness data collection to ensure comprehensive insights.

C Model details

C.1 Model Performance

Figure 8 provides a diagnostic check for the model's assumptions, particularly linearity, homoscedasticity, and independence. The residuals represent the differences between the actual and predicted values, and ideally, these should be randomly scattered around the zero line without any discernible patterns. In this plot, there appears to be some variability in the spread of residuals, with potential clusters and outliers, particularly at higher fitted values. This suggests that the model may not fully capture the complexity of the data or that certain assumptions, such as constant variance (homoscedasticity), may not hold true. These patterns indicate potential areas for model improvement, such as incorporating interaction terms, non-linear transformations, or exploring alternative modeling approaches. Despite these issues, the general scatter around zero indicates that the model performs reasonably well in capturing the overall trend of the data, although refinement may enhance predictive accuracy. # Appendix 2 {#sec-appendix2}

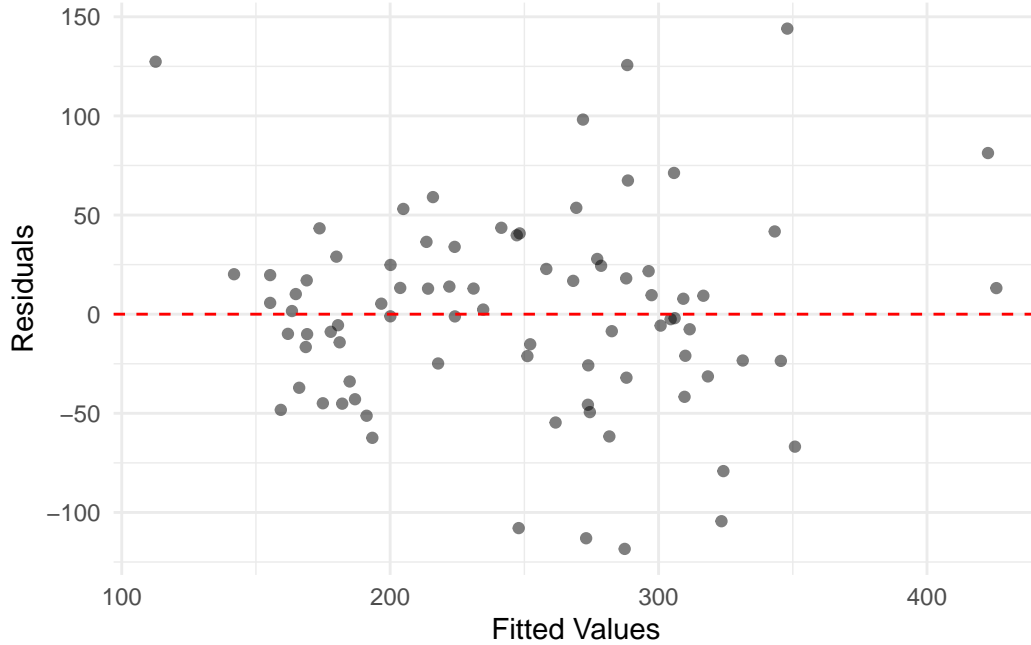


Figure 8: Residuals vs. Fitted Values for the Toronto Shelter System Model

D Survey and Sampling

D.1 Observational data

The observational data used in this analysis is collected through the Shelter Management Information System (SMIS), which tracks individuals accessing City of Toronto-funded overnight services for people experiencing homelessness, including shelters, respite centers, and hotel/motel programs. The Shelter System Flow data records monthly entries and exits from the shelter system, capturing individuals actively experiencing homelessness—those who have used the shelter system within the past three months and have not transitioned to permanent housing. This dataset provides valuable information on the demographics and flows within the shelter system, offering a lens into the effectiveness of City programs aimed at addressing homelessness. While the dataset is robust, capturing individuals engaged with City-funded services, it is important to note that it currently excludes populations experiencing homelessness outside of these services, such as those sleeping outdoors or using non-City-funded resources. This limitation highlights the observational nature of the data, which is shaped by the infrastructure and scope of service delivery. Future expansions of the dataset aim to address these gaps, providing a more holistic understanding of homelessness in Toronto and supporting the City’s goal of making homelessness rare, brief, and non-recurring.

E Idealized Survey Design

E.1 Sampling frame, Target Population, Sample

The **sampling frame** for this study comprises individuals tracked through the City of Toronto’s Shelter Management Information System (SMIS), which includes data on those accessing City-funded overnight services such as emergency shelters, respite centers, hotel/motel programs, and warming centers. This frame is defined by the administrative scope of the SMIS and is limited to individuals who engage with these services, thereby excluding those who experience homelessness outside of these programs, such as individuals sleeping outdoors or using non-City-funded resources.

The **target population** is all individuals experiencing homelessness in Toronto. This encompasses not only those using the shelter system but also those who may not engage with formal support services, including unsheltered individuals and those accessing non-City-funded programs. The target population reflects the broader demographic of interest for policymakers and researchers aiming to understand and address homelessness comprehensively across the city.

The **sample** consists of individuals within the sampling frame who have accessed City-funded overnight services and are reflected in the Shelter System Flow data. This sample provides insight into the dynamics of homelessness for a significant portion of the population but does not capture the experiences of individuals who are outside the reach of the SMIS system. Notably, the sample aligns closely with the individuals actively experiencing homelessness and tracked monthly, offering critical but partial representation of the target population. Efforts are underway to expand the sampling frame to better align with the target population by incorporating data on individuals sleeping outdoors and those using non-City-funded services.

E.2 Survey Design

link to survey: <https://forms.gle/Amq9AxRNZRsBE7eE6>

E.2.1 Survey Questions

1. **How long have you been experiencing homelessness?**

On a scale of 1 (Less than 1 month) to 10 (More than 5 years):

- 1 (Less than 1 month)
- 2 (1-3 months)
- 3 (3-6 months)
- 4 (6-12 months)
- 5 (1-2 years)

- 6 (2-3 years)
- 7 (3-4 years)
- 8 (4-5 years)
- 9 (5-6 years)
- 10 (More than 5 years)

2. Have you been provided with permanent housing through the shelter system?

- Yes
- No
- Currently waiting

3. How satisfied are you with the services provided by the shelter system?

On a scale of 1 (Very dissatisfied) to 10 (Very satisfied):

- 1 (Very dissatisfied)
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Very satisfied)

4. What types of services have been most helpful to you?

(Check all that apply):

- Employment Assistance
- Mental Health Support
- Addiction Recovery Services
- Permanent Housing Placement
- Other (please specify)

5. Do you feel the shelter system has improved your overall well-being?

- Yes
- No

- Not sure
6. **How likely are you to recommend the shelter system to someone in need?**
On a scale of 1 (Very unlikely) to 10 (Very likely):
- 1 (Very unlikely)
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10 (Very likely)
7. **How often do you use shelter services?**
- Every day
 - A few times a week
 - Once a week
 - A few times a month
 - Rarely
8. **What challenges have you faced while using the shelter system?**
(Check all that apply):
- Overcrowding
 - Lack of privacy
 - Unsafe environment
 - Limited availability of resources
 - Difficulty accessing services
 - Other (please specify)
9. **Do you feel safe in the shelter environment?**
- Yes, always

- Sometimes
- No, rarely

10. **What improvements would you like to see in the shelter system?**

(Check all that apply):

- Increased privacy
- More permanent housing options
- Better mental health support
- Faster access to resources
- More assistance with employment
- Other (please specify)

11. **How accessible are shelter services to you?**

On a scale of 1 (Not accessible) to 10 (Very accessible):

- 1 (Not accessible)
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Very accessible)

12. **Have you been able to leave the shelter system for permanent housing in the past?**

- Yes
 - No
13. **If yes, how long did you remain in permanent housing?**
- Less than 3 months
 - 3-6 months
 - 6-12 months
 - 1-2 years
 - Over 2 years
14. **What barriers have prevented you from finding permanent housing?**
(Check all that apply):
- Lack of affordable housing
 - Insufficient income or financial stability
 - Mental health challenges
 - Addiction
 - Lack of support services
 - Discrimination
 - Other (please specify)
15. **Do you believe the shelter system is helping reduce chronic homelessness in the community?**
- Yes
 - No
 - Not sure
16. **How has your quality of life changed since entering the shelter system?**
On a scale of 1 (Worsened significantly) to 10 (Improved significantly):
- 1 (Worsened significantly)

- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 (Improved significantly)

17. **Have you been connected to resources for long-term support (e.g., job training, counseling)?**

- Yes
- No

18. **If yes, how helpful were these resources?**

On a scale of 1 (Not helpful at all) to 10 (Extremely helpful):

- 1 (Not helpful at all)
- 2
- 3
- 4
- 5
- 6
- 7
- 8

- 9
- 10 (Extremely helpful)

19. **Do you feel respected and supported by shelter staff?**

- Yes, always
- Sometimes
- No

20. **What additional resources would help you transition out of homelessness?**
(*Open-ended*)

References

- Grolemund, Garrett, and Hadley Wickham. 2023. *Lubridate: Make Dealing with Dates a Little Easier*. <https://CRAN.R-project.org/package=lubridate>.
- 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/>.
- Wickham, Hadley et al. 2023a. *Testthat: Unit Testing for r*. <https://CRAN.R-project.org/package=testthat>.
- et al. 2023b. *Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://CRAN.R-project.org/package=tidyverse>.
- Wickham, Hadley, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://CRAN.R-project.org/package=knitr>.