Forecasting the Toronto Shelter System Performance on Reducing Chronic Homelessness*

Sarah Ding

December 3, 2024

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

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters 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 Data Cleaning

-Why I am only doing chronic population group. -Constructing Variables

2.4 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

2.5 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.

2.6 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'**.

2.6.1 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.

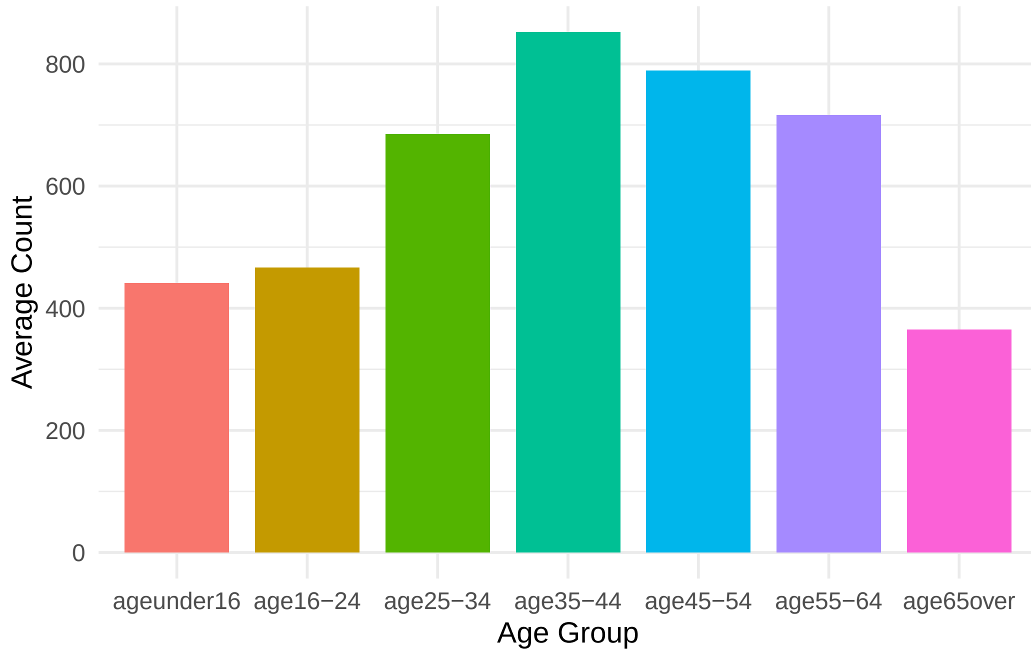


Figure 1: Displays the average number of individuals in each age group using the shelter system from 2019 to 2024. The 35-44 age group has the highest count of observations while 65 and over age group has the lowest count of observations.

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.

2.6.2 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.

2.6.3 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.

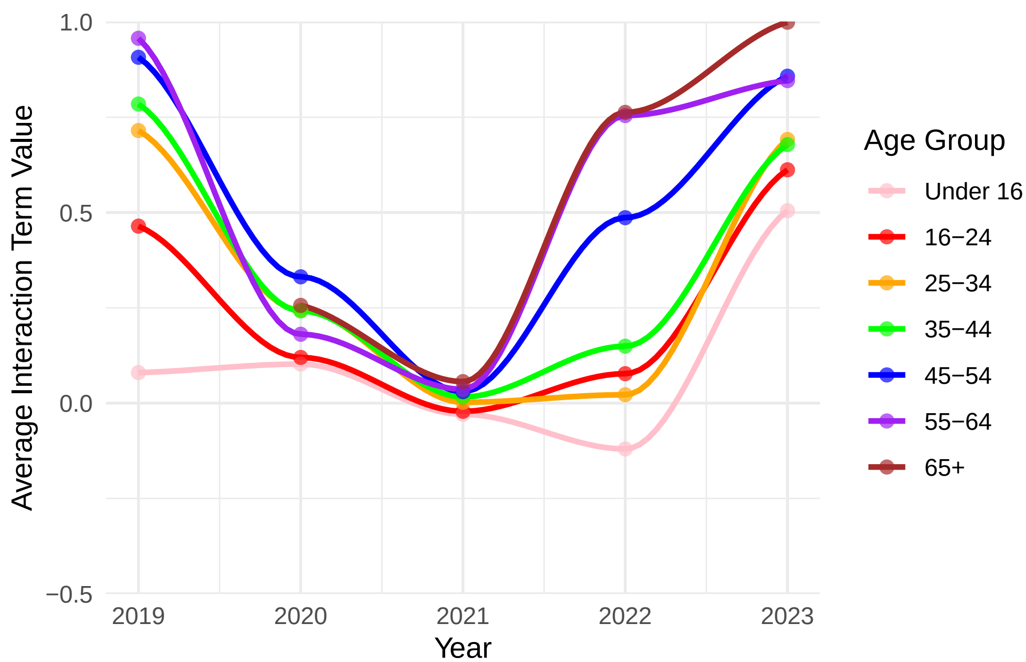


Figure 2: Illustrates the average interaction term values, representing changes over time from 2019 to 2023 for each age group. The 45–54 age group shows the highest interaction values in recent years, indicating a more significant change in engagement, while the Under 16 group consistently remains the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

2.6.4 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.

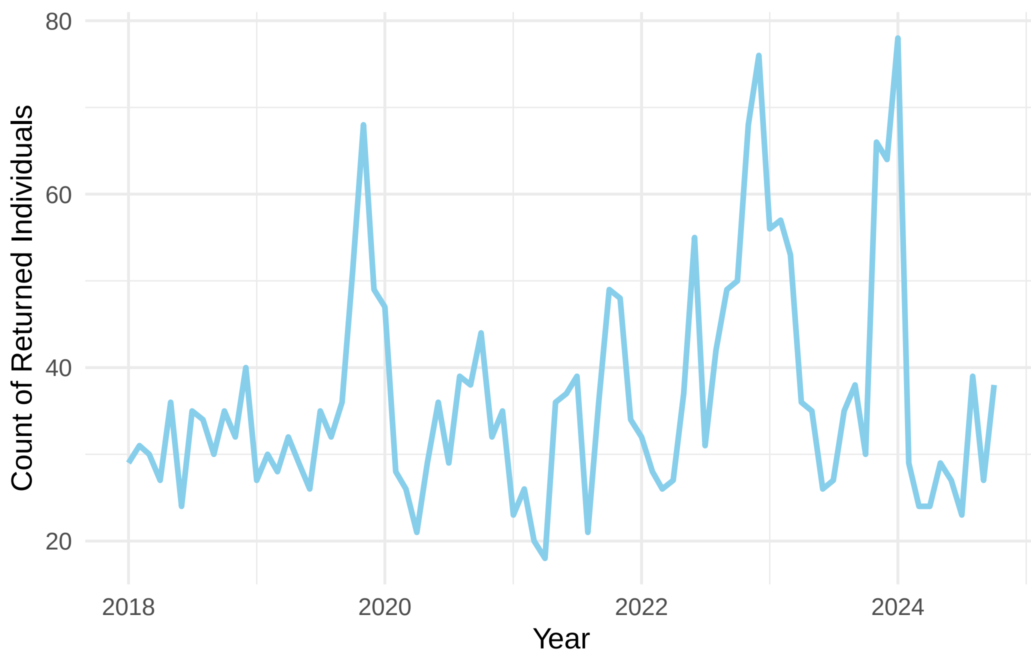


Figure 3: Illustrates the trends in the count of individuals returning to the shelter system from January 2018 to October 2024. Peaks in 2020 and 2024 suggest periods of increased shelter re-entry, possibly reflecting systemic or external influences. The overall variability highlights fluctuations in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

2.6.5 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.

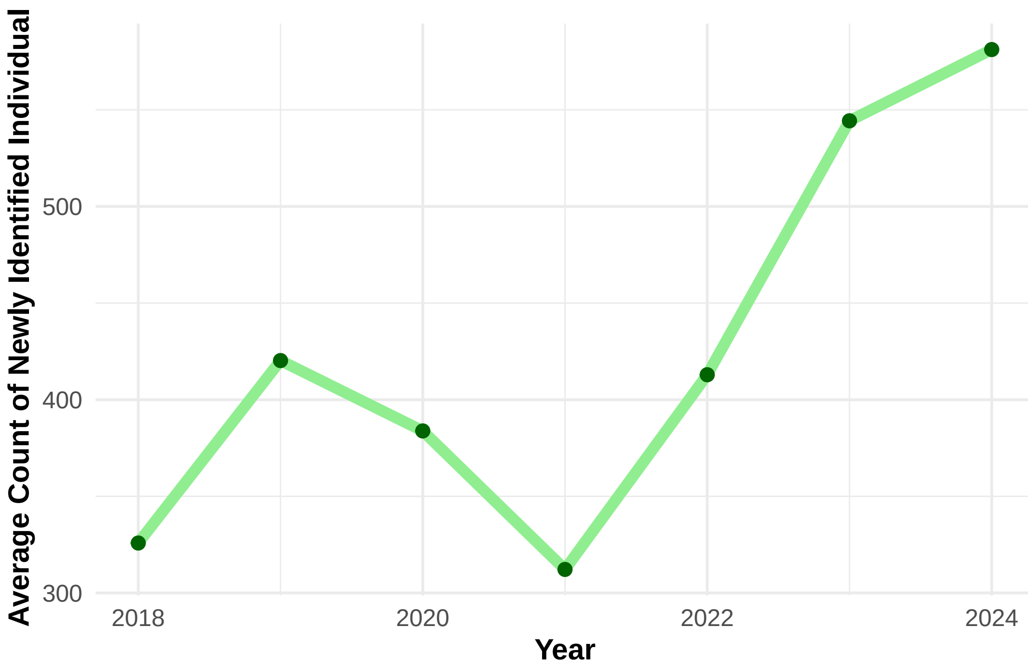


Figure 4: Trends in Newly Identified Individuals Over Time. Displays the average count of newly identified individuals in the shelter system from 2018 to 2024, highlighting fluctuations over the years with an upward trend post-2021.

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.

2.6.6 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.

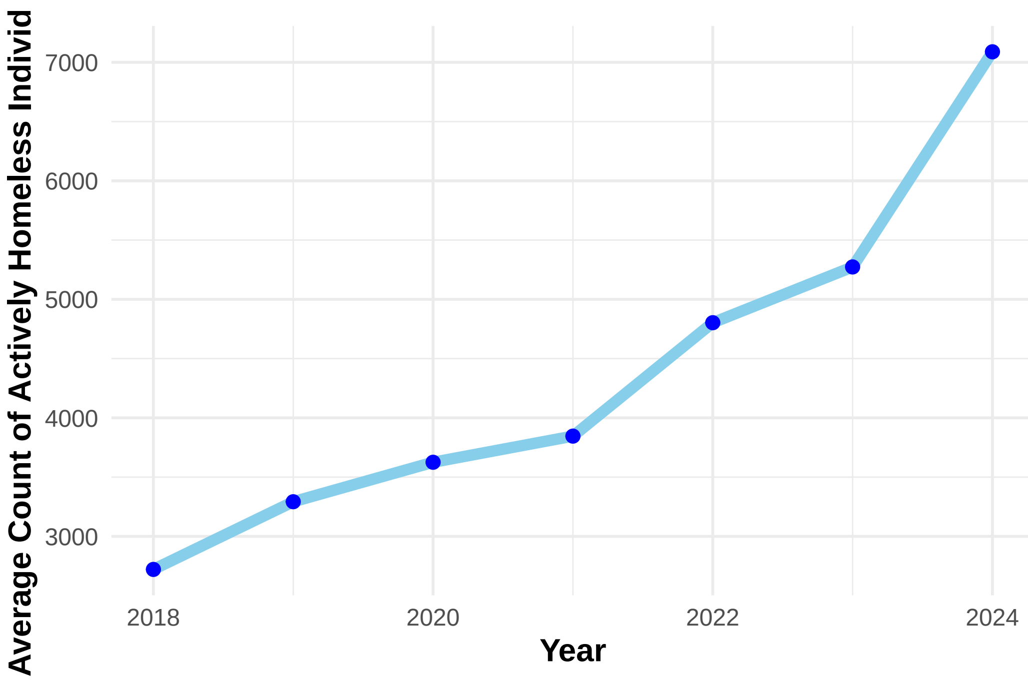


Figure 5: Trends in the average count of actively homeless individuals from 2018 to 2024. The graph shows a steady increase over the years, with the largest rise observed between 2023 and 2024, highlighting potential shifts in shelter system usage or homelessness trends during this period.

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.

2.6.7 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.

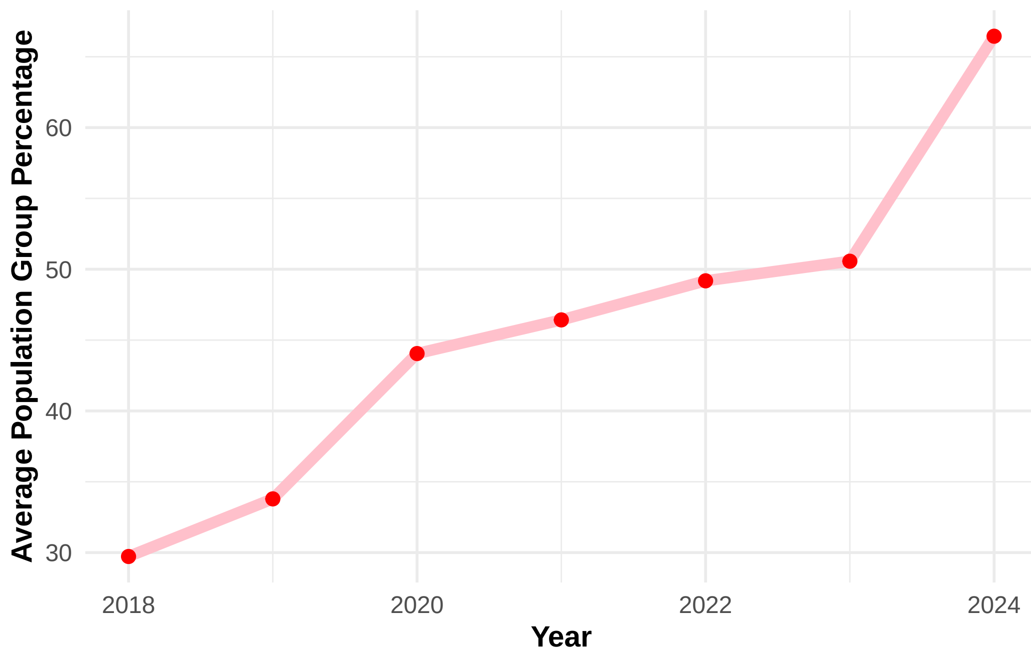


Figure 6: Trends in Average Population Group Percentage Over Time. This line graph illustrates the annual average percentage representation of the chronic population group in the dataset from 2018 to 2024. The upward trajectory indicates a consistent increase, with a significant rise observed after 2022.

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:

$$Outcome = \beta_0 + \beta_1 PopulationGroupPercentage_i + \beta_2 AgeUnder16_i + \beta_3 Age16\text{-}24_i + \beta_4 Age25\text{-}34_i + \beta_5 Age35\text{-}44_i + \beta_6$$

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 presents....

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
RMSE	88.322

Table 3 presents...

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

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

Year	Under 16 (Actual)	Under 16 (Predicted)	16-24 (Actual)	16-24 (Predicted)	25-34 (Actual)	25-34 (Predicted)	35-44 (Actual)	35-44 (Predicted)	45-54 (Actual)	45-54 (Predicted)	55-64 (Actual)	55-64 (Predicted)	65+ (Actual)	65+ (Predicted)
2018	658,306	648,884	927,609	615,958	838,741	826,090	1,149,861	1,630,915	1,349,271	1,030,312	2,260,203	1,244,567	787,086	947,339
2019	1,405,037	1,431,381	1,326,103	1,257,518	2,365,063	1,797,413	3,388,515	1,829,032	3,030,392	2,676,201	2,950,714	2,293,796	966,459	887,475
2020	1,262,981	2,235,211	1,430,293	2,392,141	2,673,786	1,620,431	2,275,312	1,205,872	1,237,528	1,863,102	2,261,306	1,686,351	1,110,311	1,073,903
2021	445,137	470,251	653,886	692,397	1,096,611	1,162,041	1,494,075	1,385,101	1,504,195	1,390,381	1,482,635	1,363,874	348,357	781,950
2022	1,020,712	1,033,831	1,279,603	1,400,890	1,843,233	1,885,982	2,498,205	1,154,476	2,521,285	2,072,130	2,368,326	2,420,031	1,221,742	1,850,302
2023	2,161,478	2,131,122	2,164,029	1,821,503	4,694,531	3,613,647	8,293,848	2,801,331	3,740,631	5,670,936	3,272,752	3,212,815	1,719,136	1,483,810
2024	2,390,021	1,392,971	1,905,699	1,406,853	3,717,163	1,723,453	3,931,138	2,224,342	2,720,487	3,700,717	1,938,499	1,213,015	1,017,210	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.

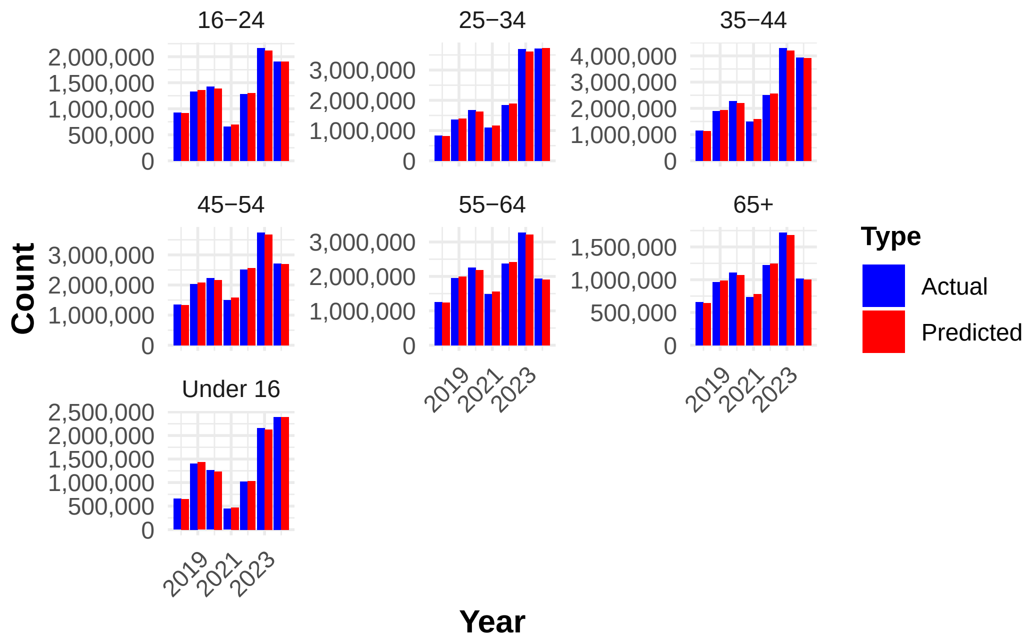


Figure 7: Yearly Actual vs. Predicted Counts of Individuals Moved to Housing by Age Group

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.

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

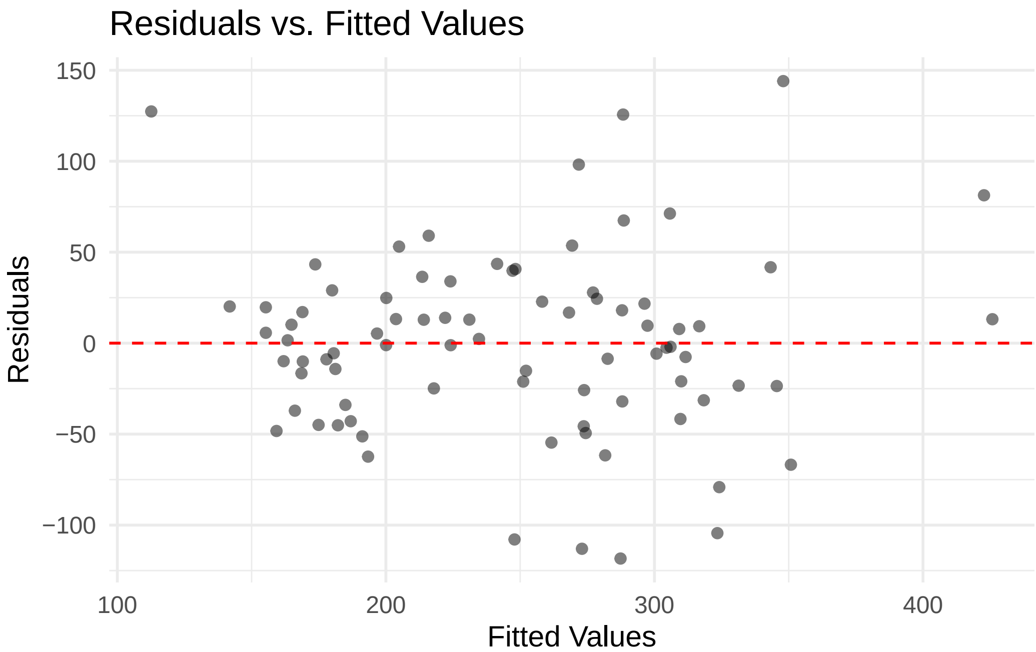
C Model details

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…

```
# Residuals vs. Fitted values
ggplot(data.frame(Fitted = predict(model_toronto), Residuals = residuals(model_toronto)),
       aes(x = Fitted, y = Residuals)) +
  geom_point(alpha = 0.5) +
  geom_hline(yintercept = 0, color = "red", linetype = "dashed") +
  labs(
    title = "Residuals vs. Fitted Values",
    x = "Fitted Values",
    y = "Residuals"
  ) +
  theme_minimal()
```



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/.