

## **Predicting Shelter Availability in Toronto Using Logistic Regression**

Group 8

Ruisi Bi

Xiaoen Hu

ChiChing Juan

Yu Liang

Sarah Rameez

Cordelia Shan

Professor Tao Wang

INF1344H: Introduction to Statistics for Data Science

## Table of Contents

|   |           |
|---|-----------|
| <i>Introduction.....</i>                      | <i>4</i>  |
| <i>Research Question .....</i>                | <i>5</i>  |
| <i>Review of Similar Research.....</i>        | <i>5</i>  |
| <i>Data Profile .....</i>                     | <i>8</i>  |
| <i>Limitations &amp; Assumptions .....</i>    | <i>10</i> |
| <i>Methodology.....</i>                       | <i>12</i> |
| <b>Descriptive Statistics .....</b>           | <b>12</b> |
| <b>Regression Analysis.....</b>               | <b>12</b> |
| <b>Exploratory Analysis .....</b>             | <b>13</b> |
| <i>Logistic Regression .....</i>              | <i>17</i> |
| <b>Data Transformation .....</b>              | <b>17</b> |
| <b>Model Interpretation .....</b>             | <b>18</b> |
| Neighborhoods .....                           | 18        |
| Sectors .....                                 | 19        |
| Other Variables .....                         | 20        |
| <b>Model Assessment .....</b>                 | <b>21</b> |
| Overall Model Fit: Likelihood Ratio Test..... | 21        |
| Confusion Matrix .....                        | 22        |
| Classification Metrics .....                  | 23        |
| Variance Inflation Factor (VIF) .....         | 23        |
| Pseudo R-Squared .....                        | 24        |
| <i>Discussion and Recommendations.....</i>    | <i>25</i> |

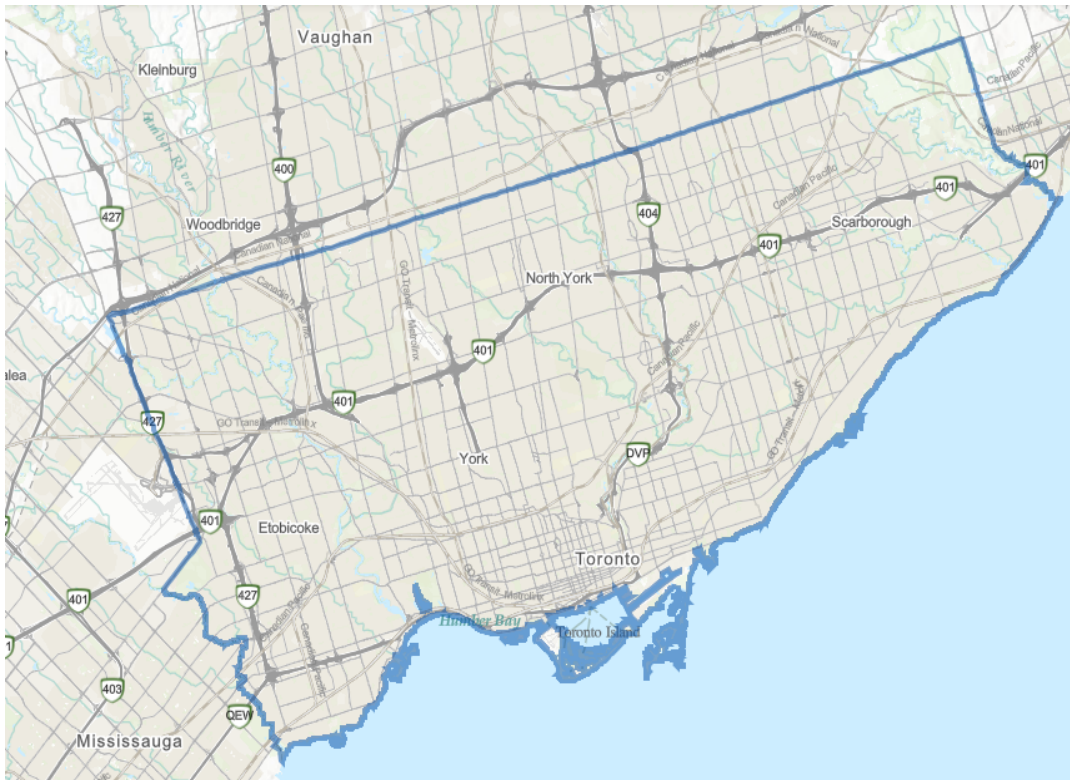
*Conclusion..... 26*

*References ..... 28*

*Appendix..... 30*

## Introduction

The consequences of homelessness, already dire on their own, are only exacerbated in a large metropolitan city with high levels of crime and below freezing minimum temperatures. Toronto's population growth, which has doubled in Ontario since 1971, has led to an increased demand for public services, particularly homeless shelters (Ontario Ministry of Finance, 2024). Currently, Toronto hosts 69 shelters across its 158 neighborhoods (City of Toronto, n.d.).



(City of Toronto, n.d.)

Despite efforts to expand capacity, the shelter system remains overwhelmed by the housing crisis, a surge in refugee claimants, and the opioid epidemic (City of Toronto, 2023). These challenges highlight the critical need for a data-driven approach to understand and predict shelter demand, enabling more efficient allocation of resources.

As per the latest available street needs assessment in 2021, only 90% of those experiencing homelessness reside in sheltered service, with almost 60% identifying as part of a marginalized group and more than three-quarters exhibiting some form of health challenge (City of Toronto, 2021). With limited capacity, it's paramount to determine which

shelters in the city have availability at any given night, based on the socioeconomic and environmental factors that impact shelter demand.

### **Research Question**

The aim of this paper is to determine how socioeconomic, environmental, and demographic factors influence the availability of homeless shelters in Toronto, and whether these factors can be used to develop a predictive model for real-time shelter availability?

This paper attempts to analyze the impact of various surrounding factors on the homeless shelter occupancy rate in Toronto. In addition, we will consider crime rates in specific neighborhoods that could affect access to shelters. Our ultimate goal is to predict the likelihood of an individual finding shelter in Toronto based on their location and personal situation.

We believe our findings will underline the importance of incorporating economic, social, and environmental variables into predictive models to inform policy-making.

### **Review of Similar Research**

Toronto's population has been growing significantly, between 1971 and 2023, Ontario's population has doubled (Ontario Ministry of Finance, 2024). With this growth rate the needs for public healthcare and services, including the availability of emergency and longer-term homeless shelters has rapidly increased (Ontario Ministry of Finance, 2024). Despite the city's best efforts to increase capacity, a gap in shelter availability continues to grow City of Toronto's report (2023).

The shelter system faces significant pressure due to the housing crisis, an influx of refugee claimants, and the opioid crisis, with 8,500 refugee claimants supported through emergency services identified as the major contributors towards the homeless population in the city (Toronto Shelter and Support Services, 2023).. Argintaru et al. (2013) identified significant gaps in healthcare access among homeless populations, findings that remain highly relevant as the population continues to grow at an unmanageable pace. Recent reports from Toronto Shelter and Support Services (2023) indicate persistent health disparities, with ongoing challenges such as opioid overdoses and limited access to mental health services, underscoring the enduring impact of these systemic issues.

The high prevalence of chronic health conditions found by Argintaru et al. (2013) underscores the importance of including unemployment rates in determining shelter availability, as individuals with severe mental and physical health issues are often less likely to maintain long-term employment, which exacerbates their risk of homelessness and reliance on shelter services (Argintaru et al., 2013; Falconer et al., 2024; Jadidzadeh & Kneebone, 2018). These findings are particularly relevant to understanding the broader societal impacts of homelessness in Toronto, where similar challenges likely contribute to shelter occupancy trends and associated social issues. The authors emphasize the need for integrated healthcare and housing policies, which aligns with our research focus on evaluating the effectiveness of existing shelter services and their relationship to community well-being. This comparative perspective helps contextualize local data within national trends, providing a robust foundation for policy recommendations (Argintaru et al., 2013).

Research has shown that socioeconomic status, income inequality, and housing affordability are strongly linked to homelessness rates. For example, a lack of affordable housing often leads to increased shelter use, with marginalized communities (e.g., low-income families, racial minorities, or individuals with mental health or substance use issues) disproportionately affected (Jadidzadeh & Kneebone, 2018). This highlights the importance of unemployment and CPI (Consumer Price Index) as key factors that contribute to homelessness—since the rising cost of living disproportionately affects individuals with lower income and socioeconomic disadvantages, exacerbating their vulnerability to homelessness.

Additionally, youth, LGBTQ+ individuals, and Indigenous populations face higher risks of homelessness due to specific social and cultural factors (Noble et al., 2022). Environmental factors such as weather conditions and seasonal fluctuations in shelter demand play a critical role in shelter occupancy. Research by Jadidzadeh and Kneebone (2018) highlights how Toronto experiences increased shelter demand during colder months, which impacts shelter capacity. This indicates that minimum temperature and precipitation could also play a major role in predicting shelter availability, given that these weather factors directly impact shelter occupancy, especially during extreme cold and

winter weather events. These socioeconomic and demographic factors will be key in understanding the barriers to shelter access and predicting occupancy rates.

Falconer, Kneebone, and Wilkins (Noble et al., 2024) offer significant insights into homeless shelter use patterns in Toronto, which directly support our research question regarding the impact of socioeconomic, demographic, and environmental factors on shelter occupancy. The research reveals that the number of nightly shelter stays in Toronto has more than doubled since 2016, largely due to economic pressures such as rising living costs, insufficient income supports, and increased numbers of refugee claimants and asylum seekers that require shelter services (Falconer et al., 2024). These findings suggest that incorporating variables like housing affordability, unemployment rates, and immigration status into predictive models could improve accuracy in forecasting shelter demand.

The City of Toronto's reports on shelter occupancy and resident deaths provide critical insights for predictive modeling of homelessness trends. These reports reveal significant demographic patterns, such as the majority of shelter resident deaths involving cisgender males and an average age of death notably lower than that of the general population. This information is crucial for identifying high-risk groups and forecasting shelter demand based on age, gender, and health vulnerabilities (City of Toronto, 2023; 2024).

Additionally, health-related trends, including the prevalence of opioid overdoses as a leading cause of death, underscore the intersection between homelessness and broader public health challenges. Integrating variables related to substance use and healthcare access into predictive models can enhance accuracy by reflecting the complex needs of the shelter population (City of Toronto, 2023). Moreover, these insights highlight the critical need for robust mental health and harm reduction services within the shelter system to address these complex challenges (Argintaru et al., 2013; City of Toronto, 2023). These findings highlight the importance of robust mental health and harm reduction services within the shelter system.

Gaps in shelter availability and access to services can be exacerbated by systemic issues such as lack of affordable housing, mental health resources, and discrimination in shelters (Dionne et al., 2023). Seasonal patterns in mortality and shelter usage further

indicate that demand increases during colder months, pointing to the need for adaptive policies and resource allocation strategies. Data on shelter system flow—tracking the inflow of new residents and outflow to permanent housing—provides a dynamic view of homelessness. This data allows for the development of models that anticipate occupancy spikes and identify potential service gaps, enabling more responsive and effective interventions (City of Toronto, 2024).

The literature review serves to identify the key socioeconomic, health, and environmental factors that influence shelter occupancy in Toronto. By examining the existing body of research, we understand how issues like housing affordability, unemployment, chronic health conditions, and environmental factors such as extreme weather contribute to the demand for shelter services. These insights allow us to select relevant variables for our predictive models, helping us better forecast shelter demand and improve service delivery.

In addition, we note that a majority of the studies already conducted opt for a long-term holistic view of homeless shelter availability and do not consider day-to-day variation in their analysis. Our research question accounts for daily variation in the availability of shelter with the goal of creating a model that is able to provide a real-time approximation of finding shelter given one's demographic, socio-economic and environmental circumstances.

### **Data Profile**

Our analysis integrates data from multiple sources, selected based on insights gained from a comprehensive literature review and domain expertise. In the context of a major metropolitan city situated in a cold climate, we identified key factors likely to influence the availability of homeless shelter spaces. These factors were selected to reflect the complex interplay of social, environmental, and economic conditions shaping shelter demand and access.

Given that many factors critical to our analysis can vary significantly across cities, we have chosen to focus specifically on Toronto, Ontario. This decision minimizes the risk of confounding effects arising from divergent trends in different cities, which could obscure meaningful patterns or relationships in the data. By concentrating on a single metropolitan



area, we aim to capture the unique dynamics influencing homeless shelter availability in a cold-climate urban setting, ensuring the results are both focused and relevant to the local context.

*Note: The entire data cleaning process is defined in the appendix.*

- **Daily Shelter Overnight Service and Capacity (City of Toronto, 2024):** This dataset provides comprehensive information on the occupancy and capacity of Toronto's shelter system from January 2021 to September 2024. This dataset includes daily snapshots of the number of individuals accommodated, the types of shelter programs utilized, and the occupancy rates across various sectors, such as family and singles shelters.
  - **Selected variables:** Our dependent variable is provided in this data, a binary variable where 1 indicates availability and 0 indicates no availability in a shelter at any given night. We have also sourced the following independent variables from this data: capacity (number of beds hosted by the shelter), sector (a factor variable categorizing the user groups that the shelter caters to), and the postal code that the shelter is located in.
- **Major Crime Indicators (Toronto Police Service, 2024):** This dataset provides detailed information on crime occurrences, categorized by date and neighborhood, between January 2021 and September 2024. This dataset includes offenses such as assault, break and enter, auto theft, robbery, and theft over \$5,000, recorded at the offense and/or victim level. Each entry corresponds to a reported incident, allowing for analysis of crime patterns across different areas and time periods.
  - **Selected variables:** The independent variable selected for this analysis was the count of crimes occurring at a given night, in the same neighborhood as the location of a shelter.
- **Toronto Weather Records (National Oceanic and Atmospheric Administration, 2024):** The data was retrieved between January 2021 and September 2024 and accessed through a website specializing in historical weather records.

- Selected variables: For our analysis, we selected the minimum daily temperature (in degrees Celsius) and daily precipitation (in centimeters) for Toronto. These variables were chosen as they directly relate to environmental conditions that can influence the demand for homeless shelters, particularly in a cold-climate urban setting.
- Consumer Price Index Data (Government of Canada, 2007): The Consumer Price Index (CPI) is a key economic indicator that measures the average change over time in the prices paid by consumers for a basket of goods and services. This data was sourced from Statistics Canada for the periods encompassing January 2021 and September 2024, the national statistical agency responsible for producing such economic indicators.
  - Selected variables: For our analysis, we utilized monthly, unadjusted CPI data specific to Toronto, with a base year of 2002. By analyzing this variable, we can assess the inflation trend over time and its potential impact on factors such as shelter availability in Toronto.
- Unemployment Data (Government of Canada, 2024): The unemployment data for this analysis is sourced from Statistics Canada's monthly Labor Force Survey between January 2021 and September 2024. It provides detailed insights into labor market conditions across Canada. By incorporating unemployment data, we aim to explore the connection between labor market trends and the demand for homeless shelter services in Toronto.
  - Selected variables: The independent variable selected for our analysis is Toronto's unemployment rate. This data was manually extracted from each monthly report and compiled into a data frame. It is important to note that the figures provided are seasonally adjusted three-month moving averages of the unemployment rate.

### **Limitations & Assumptions**

We have opted to focus on more recent shelter activity, recognizing that the shelter system has struggled to meet demand in recent years. This approach reflects the evolving dynamics of the current landscape, where pre-pandemic factors may no longer hold the same relevance.

There are several limitations in the project that may affect the comprehensiveness of the findings. While our model integrates variables such as unemployment rates, crime data, and weather conditions, these gaps limit our ability to fully capture the multifaceted factors that drive shelter occupancy. Similarly, while we use the Consumer Price Index (CPI) and unemployment rate as an economic indicator, this variable alone may not fully capture the nuanced economic pressures faced by homeless individuals, such as rent increases or income assistance gaps.

Our focus on Toronto provides valuable localized insights, but limits model generalization to other regions. Cities with different socioeconomic conditions, housing policies, and climate dynamics may exhibit very different shelter occupancy trends.

Our temporal scope emphasizes recent shelter activity to reflect current post-pandemic dynamics. While ensuring relevance, this approach also excludes historical factors that may influence the current shelter system, thus potentially overlooking long-term trends. Data aggregation, such as crime rates at the neighborhood level, may also mask localized variations near shelters, potentially impacting the precision of predictions.

It is important to recognize that certain datasets were not available to us due to our data requirements. Immigration intake, for example, is only available quarterly and only provided at the provincial level rather than the city level. Similarly, we could not incorporate neighborhood populations as they were last updated in 2021 and fail to account for a growing population.

Our analysis also relies on several key assumptions. First, we assume that the datasets used are accurate and representative of Toronto's conditions, though inconsistencies in reporting could introduce bias. The logistic regression model assumes independence of residuals, which may not fully account for spatial dependencies, such as the proximity of shelters influencing occupancy rates. Furthermore, demographic variables such as sector and neighborhood are treated uniformly, which may overlook the unique vulnerabilities of specific groups disproportionately affected by homelessness, including

LGBTQ+ individuals and Indigenous populations. Despite these limitations, the study provides a focused and systematic analysis, forming a solid foundation for understanding and predicting shelter availability.

## **Methodology**

### **Descriptive Statistics**

We utilized descriptive statistics with RStudio to systematically organize and summarize the variability present in our dataset, providing a comprehensive overview of its key quantitative characteristics.

We examine the data from a temporal lens to identify seasonal patterns, as well as a spatial analysis across Toronto's neighborhoods. We also compare the different sectors that shelters cater to in order to examine their effect on shelter availability.

We also perform a correlation analysis of the numerical variables, specifically using Pearson's correlation coefficient ( $r$ ), to examine the relationships between shelter availability (`Availability_Total`) and key continuous predictors, including `Capacity`, `Total_Crime`, `Minimum_Temperature`, and `Precipitation`, `Unemployment_Rate`, and `CPI_all`.

### **Regression Analysis**

Logistic regression was employed in this project to model the relationship between predictors and shelter availability, a binary outcome variable (e.g., 0 and 1 represent two states of an event). Linear regression assumes that the target variable is continuous and that predicted values can span the entire real number range. However, for binary classification problems, linear regression can lead to unreasonable predictions (e.g., negative values or values greater than 1), making it unsuitable for describing the probability of an event.

In contrast, logistic regression uses the logistic (logit) function to map predicted values to the  $[0, 1]$  range, allowing the output to be interpreted as the probability of an event occurring. This probability reflects the likelihood that the event belongs to a particular category (commonly the positive class) given the input variables.

Additionally, the results of logistic regression can be further interpreted using odds ratios. Specifically, the exponentiated regression coefficients represent the effect of a

predictor variable on the likelihood of the target event. For instance, an odds ratio greater than 1 indicates that an increase in the predictor variable raises the probability of the event occurring, while an odds ratio less than 1 suggests a decrease in probability. This interpretability is particularly useful for practical decision-making and understanding variable impacts.

Logistic regression not only provides meaningful and bounded predictions but also helps researchers better understand the influence of predictors, making it an ideal choice for binary classification problems.

### Exploratory Analysis

Summary statistics revealed that the mean and median of 'Availability\_Total' are 0.83 and 0, respectively, indicating the typical availability level is extremely low, with most shelters having no available spaces and only a few contributing to a slightly higher average. The 'Average\_Capacity' of shelters was 107.64, with a standard deviation of 172.96, reflecting variability in shelter sizes.

Key continuous predictors, including 'Total\_Crime', 'Minimum\_Temperature', and 'Precipitation', exhibit substantial variability, reflecting differences in neighborhood safety, seasonal temperature changes, and sporadic precipitation levels.

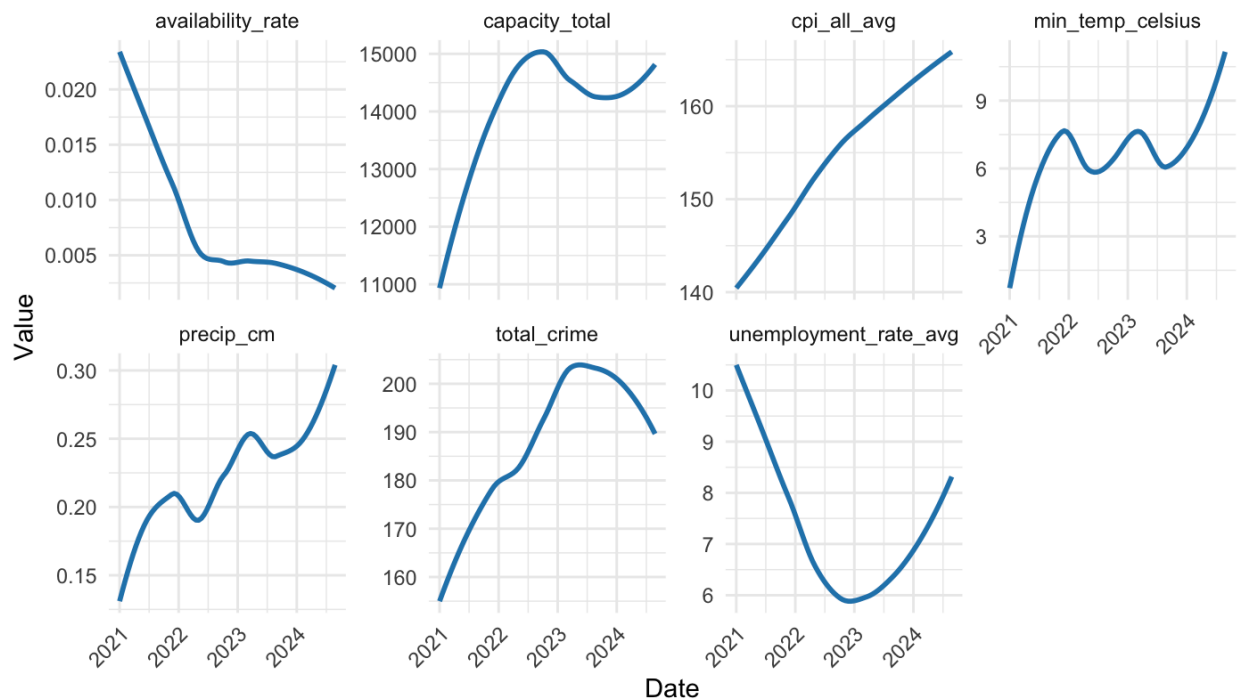
The 'Unemployment Rate' and 'CPI\_All' highlight economic conditions, with variations reflecting shifts in joblessness and cost of living, which could influence shelter demand. Categorical variables such as Sector and Hood\_158 were excluded from numeric analyses but provide important context for spatial distribution and group-level differences in the dataset.

Table 1

| <i>Descriptive Statistics of Key Variables</i> |             |                           |
|--|-------------|---------------------------|
|  | <b>Mean</b> | <b>Standard Deviation</b> |
| <b>Availability Total</b>                      | 0.83        | 2.52                      |
| <b>Availability Rate</b>                       | 0.03        | 0.08                      |
| <b>Capacity Total</b>                          | 107.64      | 172.96                    |
| <b>Total Crime</b>                             | 1.44        | 1.63                      |

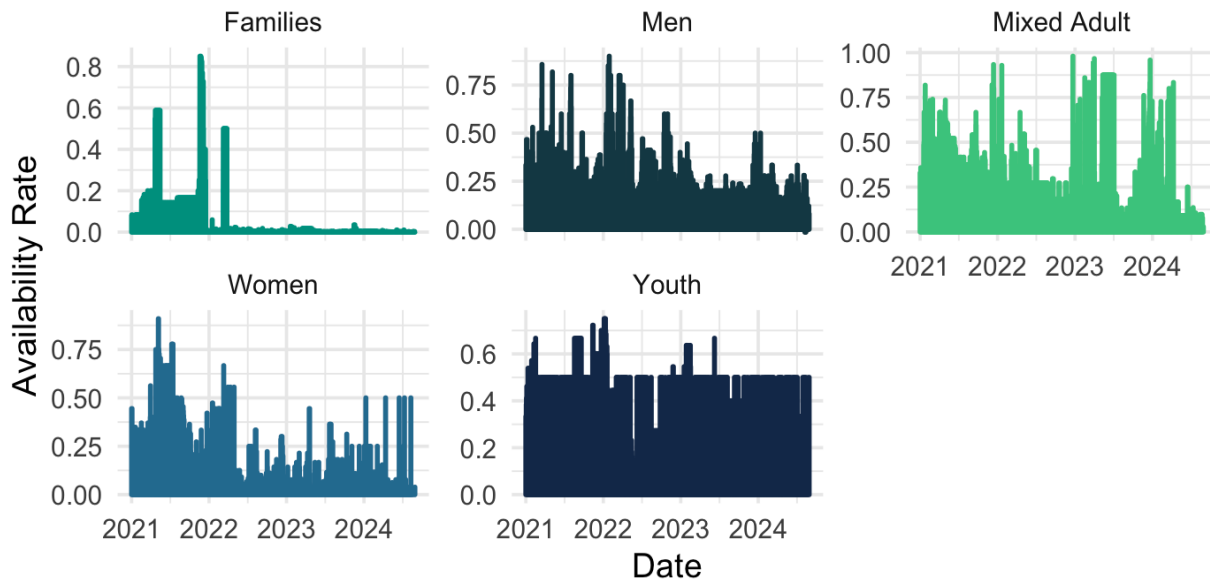
|                                |        |      |
|--------------------------------|--------|------|
| <b>CPI ALL</b>                 | 154.23 | 7.51 |
| <b>Unemployment Rate</b>       | 7.33   | 1.43 |
| <b>Min Temperature Celsius</b> | 6.45   | 9.41 |
| <b>Precipitation cm</b>        | 0.22   | 0.58 |

We conducted a temporal analysis of key variables to locate seasonal patterns. As expected, the availability of shelters has declined over time, despite the notable increase in capacity in an attempt to meet this demand. CPI increases steadily over time, while the unemployment rate demonstrates cyclical trends, potentially impacting shelter availability over time. Low average crime counts per neighborhood but notable variability, indicating safety discrepancies in different areas.



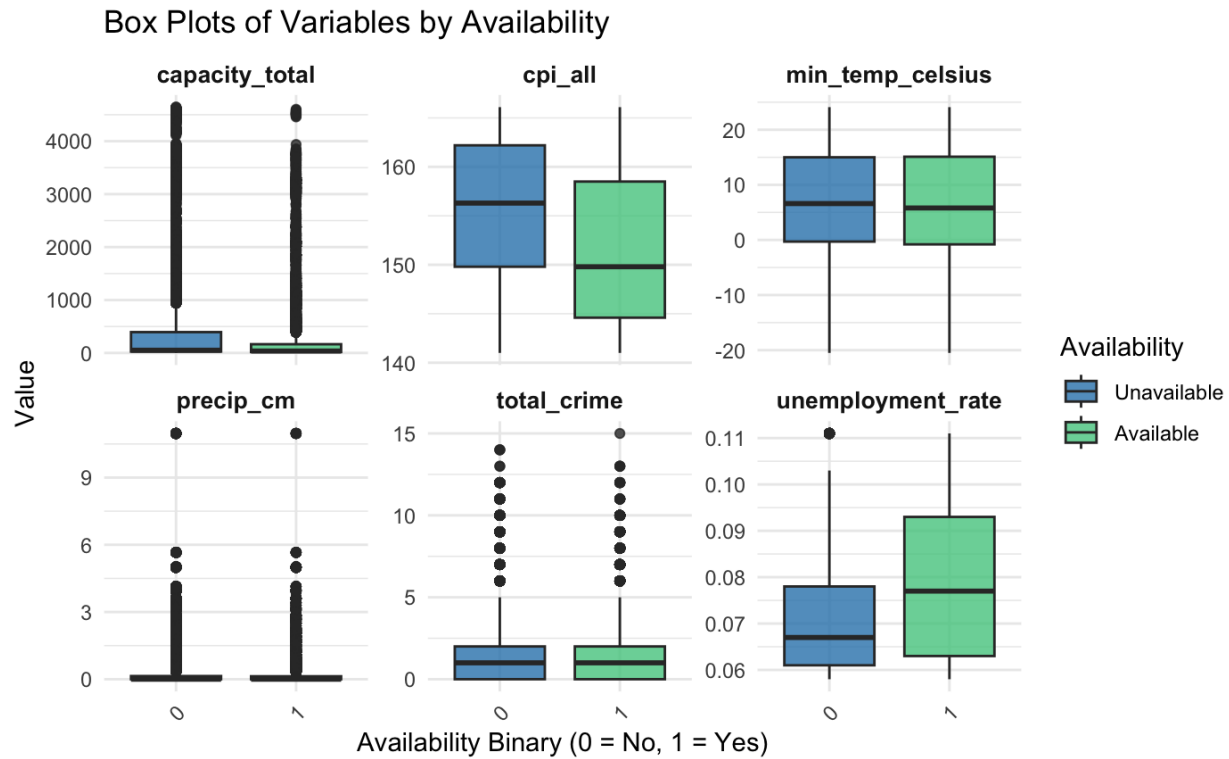
Analyzing the availability rate across sectors that shelters cater to reveals a gradual decline in availability across the family, men and women sectors. Availability across the shelters that cater to youth remains relatively constant across the past three years, while family focused shelters display spikes in availability, indicating the creation of new sectors to cater to the sector, most likely as a result of increased refugee intake following the Ukraine-Russia tensions.

## Availability Over Time Across Sectors



Analysis of spatial patterns reveals disparities in shelter availability across Toronto's neighborhoods. Certain areas experience higher demand due to socioeconomic and crime-related factors. Crime rates, particularly assault and robbery, appear to be correlated with shelters in high-demand neighborhoods. These sector-specific differences highlight opportunities for targeted resource allocation.

We are also interested in viewing the distributions of numerical variables during periods of availability and no availability across Toronto's shelters. While weather and crime appear uniform regardless of availability, the below graph indicates higher CPI when there is no availability in shelters, while unemployment appears to be higher during periods of availability, which calls for a deeper exploration between the two variables.



A correlation analysis reveals weak positive correlation ( $r = 0.18$ ) between unemployment rate and shelter availability suggests potential lag effects of economic downturns on shelter demand. Slight negative correlation ( $r = -0.20$ ) between CPI and availability underscores economic pressures on shelter accessibility. Environmental factors such as Precipitation and Minimum\_Temperature demonstrated negligible correlations ( $r=0.00$  and  $r=0.01$ , respectively), suggesting limited direct influence on shelter availability. Similarly, Total\_Crime had a minimal association with Availability Total ( $r=0.02$ )

**Table 2**

Correlation table for numeric variables

|              | Availabil<br>ity | Capacit<br>y | Crime | CPI | Unemp<br>loymen<br>t | Min<br>Temp | Precipi<br>tation |
|--------------|------------------|--------------|-------|-----|----------------------|-------------|-------------------|
| Availability | 1.00             |              |       |     |                      |             |                   |
| Capacity     | 0.06             | 1.00         |       |     |                      |             |                   |
| Crime        | 0.02             | 0.02         | 1.00  |     |                      |             |                   |



|                      |      |       |       |       |       |      |      |
|----------------------|------|-------|-------|-------|-------|------|------|
| <b>CPI</b>           | -0.2 | 0.05  | 0.09  | 1.00  |       |      |      |
| <b>Unemployment</b>  | 0.18 | -0.05 | -0.07 | -0.69 | 1.00  |      |      |
| <b>Min Temp</b>      | 0.01 | 0.01  | 0.07  | 0.13  | -0.02 | 1.00 |      |
| <b>Precipitation</b> | 0.00 | 0.00  | -0.01 | 0.04  | -0.02 | 0.05 | 1.00 |

Correlation Heatmap with Values



## Logistic Regression

### Data Transformation

**Rescaling:** The unemployment rate was multiplied by 100 to prevent significant skew in the odds ratio.

**Resampling the Dependent Variable:** The dataset was resampled to address the unequal representation of shelter availability versus non-availability. Only 27.4% of the data reflected shelter availability, creating a significant class imbalance. To mitigate this, cases

of availability were resampled to match the number of non-availability cases. While this approach increased the dataset size from 173,642 observations to 242,584, it resulted in noticeable improvements in key model assessment metrics, enhancing the overall performance and interpretability of the logistic regression model.

## Model Interpretation

### Neighborhoods

Top 4 neighborhoods most and least likely to find shelter (using reference neighborhood 172 - Dovercourt village)

|                              | Variable               | Estimate | Std Error | Z Value | P Value | Odds Ratio |
|------------------------------|------------------------|----------|-----------|---------|---------|------------|
| Neighborhoods (Most Likely)  | East Willowdale        | 3.824    | 0.587     | 6.517   | 0       | 45.772     |
|                              | Trinity-Bellwoods      | 3.45     | 0.358     | 9.626   | 0       | 31.503     |
|                              | Downsview              | 2.554    | 0.067     | 38.371  | 0       | 12.859     |
|                              | Morningside            | 2.509    | 0.193     | 12.994  | 0       | 12.297     |
| Neighborhoods (Least Likely) | West Humber-Clairville | -1.992   | 0.068     | -29.506 | 0       | 0.136      |
|                              | Mimico-Queensway       | -3.298   | 0.163     | -20.268 | 0       | 0.037      |
|                              | Oakwood Village        | -3.569   | 0.229     | -15.554 | 0       | 0.028      |
|                              | Wexford/Maryvale       | -3.727   | 0.181     | -20.595 | 0       | 0.024      |

Neighborhood 152 (East Willowdale) has an Estimate of 3.824 with a highly significant P Value ( $<0.001$ ), making its Odds Ratio 45.772. This means residents here are 45 times more likely to find shelter compared to the reference neighborhood, making it the most likely neighborhood to find availability of shelter, holding other factors constant. Similarly,

Neighborhood 081 (Trinity Bellwoods) demonstrates favorable conditions with an Estimate of 3.45 and a very low Standard Error (0.358), ensuring high confidence in its result. The Odds Ratio of 31.503 highlights significantly increased accessibility in this neighborhood. The odds ratio dropped off fairly steeply after this high point, with 36 of the 59 neighborhoods exhibiting odds ratio below 1 for finding availability as compared to the reference neighborhood.

On the other end, Neighborhood 119 (Wexford/Maryvale) has an Odds Ratio of 0.024, meaning residents here are 96% less likely to find shelter compared to the baseline. Another challenging area is Neighborhood 160 (Mimico-Queensway), with an Odds Ratio of 0.037, showing severe barriers to shelter access.

### Sectors

(In this table, reference sector 'family')

|                                    | Variable    | Estimate | Std Error | Z Value | P Value | Odds Ratio |
|------------------------------------|-------------|----------|-----------|---------|---------|------------|
| Sector (reference sector 'family') | Men         | -0.982   | 0.022     | -44.197 | 0       | 0.375      |
|                                    | Mixed Adult | -0.078   | 0.02      | -3.953  | 0       | 0.925      |
|                                    | Women       | -0.504   | 0.022     | -22.595 | 0       | 0.604      |
|                                    | Youth       | 0.098    | 0.023     | 4.28    | 0       | 1.103      |

Residents in men's shelters are approximately 62.5% less likely ( $1 - 0.375 = 0.625$ ) to find availability compared to family shelters, holding other factors constant. The results are highly significant, as indicated by the P-value and the low standard error (0.029), suggesting high confidence in this finding. Men's shelters thus face the greatest challenges in availability compared to the reference group, according to our model.

Women's shelters show notable but smaller barriers compared to men's shelters, showing a 39.6% lower likelihood ( $1 - 0.604 = 0.396$ ) of finding availability compared to family shelters at a highly significant level. Residents in mixed adult shelters, meanwhile, are far less likely (by 7.5%) to find availability compared to family shelters, though the

effect is small due to having significance only at the 0.05 level. This indicates that mixed adult shelters perform similarly to family shelters, with only a slight reduction in odds. Youth shelters are 10.3% more likely to have availability compared to family shelters. This is a small positive effect but statistically significant, with a moderate standard error of 0.023. This suggests slightly better access to shelters for youth compared to families.

This analysis highlights disparities in shelter availability among different sectors. Men's shelters face significant challenges, while youth shelters show slight improvements in availability compared to the reference group.

#### ***Other Variables***

| <b>Variable</b>            | <b>Estimate</b> | <b>Std Error</b> | <b>Z Value</b> | <b>P Value</b> | <b>Odds Ratio</b> |
|----------------------------|-----------------|------------------|----------------|----------------|-------------------|
| Capacity Total             | 0               | 0                | 8.664          | 0              | 1                 |
| Total Crime                | 0.005           | 0.003            | 1.694          | 0.09           | 1.005             |
| CPI All                    | -0.071          | 0.001            | -77.68         | 0              | 0.931             |
| Unemployment<br>Rate       | 0.085           | 0.005            | 18.346         | 0              | 1.089             |
| Min Temperature<br>Celsius | 0.004           | 0                | 8.163          | 0              | 1.004             |
| Pre-CIP CM                 | 0.014           | 0.008            | 1.77           | 0.077          | 1.014             |

The results indicate that shelter availability is not impacted by total capacity, suggesting that shelters face challenges in providing available beds regardless of their funding or capacity, reflecting the systemic strain on the shelter network.

Contrary to our earlier assumptions, total crime and precipitation have no statistically significant impact on shelter availability, with p-values of 0.09 and 0.077, respectively. While precipitation shows a slightly positive association, this effect is not significant enough to draw reliable conclusions.

Minimum temperature, however, has a small but significant effect (OR = 1.004,  $p < 0.001$ ), suggesting that warmer temperatures marginally improve shelter availability. This

finding likely reflects seasonal demand fluctuations, where milder weather reduces the immediate pressure on shelter systems.

Economic factors like the unemployment rate exhibit a small but significant positive influence, indicating that for every 1-unit increase in the unemployment rate, the odds of finding shelter availability increase by 8.9%, holding all other factors constant. This could be due to a lag effect, where rising unemployment initially reduces demand for shelters as individuals rely on other support systems before seeking shelter. Higher inflation, as measured by CPI, is strongly linked to reduced shelter availability. For every 1-unit increase in CPI, the odds of finding shelter decrease by 6.9%. This finding highlights the economic pressures that inflation exerts on both the shelter system and individuals seeking services, likely reflecting decreased capacity to meet demand during periods of rising costs.

Overall, the findings emphasize that while environmental and economic factors play a role, systemic constraints within the shelter system remain a critical barrier to improving availability.

## Model Assessment

### *Overall Model Fit: Likelihood Ratio Test*

| <b>Model</b>         | <b>Residual Deviance</b> | <b>Degrees of Freedom (Residual)</b> | <b>DF</b> | <b>Deviance Difference</b> | <b>Pr (&gt;Chi)</b> |
|----------------------|--------------------------|--------------------------------------|-----------|----------------------------|---------------------|
| Model 1 (Null Model) | 336,293                  | 242,583                              | -         | -                          | -                   |
| Model 2 (Full Model) | 274,397                  | 242,514                              | 69        | 61,895                     | < 2.2e-16           |

The Likelihood Ratio Test compares two models: model 1 (Null Model) includes only the intercept, assuming shelter availability is random and unrelated to any variables, while model 2 (Full Model): Includes all variables like neighborhood, capacity, and economic factors.

The results provide critical insights into model performance. Model 1, the null model, has a high residual deviance of 336,293, indicating a poor fit to the data. In contrast, Model 2, which includes the selected explanatory variables, demonstrates a significantly improved fit, with a lower residual deviance of 274,397. The improvement in model fit's quantified by a deviance difference of 61,895 between the two models, using 69 degrees of freedom. This improvement is additionally validated by an extremely low p-value ( $< 2.2e-16$ ), confirming that the addition of explanatory variables is statistically significant.

These findings highlight that the full model provides a substantially better explanation of shelter availability compared to the null model. This result underscores the importance of including selected variables such as capacity, neighborhood, and economic factors in the analysis to capture the complexity of shelter availability dynamics effectively.

#### ***Confusion Matrix***

| <b>Confusion Matrix Results</b>  |                               |                              |  |
|----------------------------------|-------------------------------|------------------------------|--|
| <b>Prediction</b>                | <b>Actual: No Shelter (0)</b> | <b>Actual: Shelter (1)</b>   | <b>Explanation</b>   |
| <b>Predicted: No Shelter (0)</b> | True Positives (TP): 85,599   | False Negatives (FN): 36,014 | Correctly and incorrectly predicts no shelter availability (majority class). |
| <b>Predicted: Shelter (1)</b>    | False Positives (FP): 35,693  | True Negatives (TN): 85,278  | Correctly and incorrectly predicts shelter availability.                     |

The matrix indicates that while model performs exceptionally well for the majority class ("No Shelter"), with a high number of true positives, it struggles with the minority class ("Shelter Availability"), as indicated by the large number of false negatives (31,008). This suggests it frequently incorrectly classifies cases where shelter availability exists, which reflects class imbalance in the dataset, where "No Shelter" cases dominate, given that availability in shelters across Toronto are sparse.

The confusion matrix highlights that while the model performs well for the majority class ("No Shelter"), it faces challenges with the minority class ("Shelter Availability"). This is evident from the high number of 36,014 false negatives. Additionally, the model incorrectly predicts 35,693 false positives, further indicating difficulty in distinguishing between classes. Despite resampling the data to counter false predictions, the overall performance continues to reflect the inherent imbalance in the dataset, where "No Shelter" cases dominate. This imbalance likely stems from the sparse availability of shelters across Toronto, emphasizing the need for techniques to improve minority class prediction.

### ***Classification Metrics***

| <b>Classification Metrics</b> |              |   |
|-------------------------------|--------------|---|
| <b>Metric</b>                 | <b>Value</b> | <b>Explanation</b>  |
| Accuracy                      | 70.44%       | Overall proportion of correct predictions.                                    |
| Sensitivity                   | 70.57%       | True positive rate: the model effectively predicts "No Shelter Availability." |
| Specificity                   | 70.31%       | True negative rate: struggles to predict "Shelter Availability" correctly.    |
| F1 Score                      | 70.40%       | Balances precision and recall for "Shelter Availability."                     |

The results indicate moderate accuracy, with the model correctly predicting 70.44% of cases overall. However, this performance must be interpreted cautiously, as the balanced accuracy across sensitivity (70.57%) and specificity (70.31%) suggests the model performs similarly for both "No Shelter Availability" and "Shelter Availability." The F1 Score (70.40%) reflects a balance between precision and recall, the model's performance highlights room for improvement in distinguishing between the two classes.

### ***Variance Inflation Factor (VIF)***

| <b>Variance Inflation Factor (VIF)</b> |            |                                    |
|--|------------|------------------------------------|
| <b>Variable</b>                        | <b>VIF</b> | <b>Interpretation</b>              |
| Capacity Total                         | 1.91       | Low multicollinearity, no concern. |

|                         |       |   |
|-------------------------|-------|---|
| Sector                  | 9.96  | Moderate multicollinearity, acceptable for multiple categories.         |
| Neighborhood            | 19.97 | High multicollinearity, likely due to multiple neighborhood categories. |
| Total Crime             | 1.26  | Very low multicollinearity.   |
| CPI All                 | 2.22  | Minimal multicollinearity, not a concern.                               |
| Unemployment Rate       | 2.16  | Minimal multicollinearity, no issue.                                    |
| Min Temperature Celsius | 1.04  | Negligible multicollinearity.   |
| Pre-CIP CM              | 1     | Negligible multicollinearity.   |

Most variables have low VIF values, indicating minimal collinearity issues. The neighborhood variable shows high VIF (19.97), suggesting potential redundancy among its categories. This may need further analysis or dimensionality reduction.

#### ***Pseudo R-Squared***

| <b>Pseudo R-Squared</b>              |               |  |
|--------------------------------------|---------------|--|
| <b>Metric</b>                        | <b>Value</b>  | <b>Explanation</b>   |
| Log-Likelihood (llh)                 | -1.371987e+05 | Indicates how well the full model fits the data. Higher (less negative) is better.           |
| Log-Likelihood (llhNull)             | -1.681464e+05 | Indicates fit for the null model (intercept only).   |
| G <sup>2</sup> (Deviance Difference) | 6.189540e+04  | Improvement in model fit when predictors are added.  |
| McFadden R <sup>2</sup>              | 0.1841        | Indicates moderate fit (values between 0.2-0.4 are good).                                    |
| r2ML                                 | 0.2252        | Explains 18.34% of the variance in the dependent variable.                                   |
| r2CU                                 | 0.3003        | Explains 26.76% of the variance in the dependent variable (adjusted pseudo-R <sup>2</sup> ). |



Pseudo R-squared evaluates the goodness-of-fit for logistic regression models. We have already confirmed from our analysis of deviance table that the full model significantly outperforms the null model, evidenced by higher log-likelihood and deviance improvement. The large value for the  $G^2$  indicates that the predictors meaningfully contribute to explaining shelter availability.

The McFadden  $R^2$  is often used in logistic regression to assess the proportion of variance explained, and indicates a moderate model fit, while the  $r^2_{ML}$  and  $r^2_{CU}$  indicate higher levels of variation being explained by the model. The  $r^2_{CU}$  tends to provide a more optimistic view by accounting for variance that could otherwise be underestimated.

While the model demonstrates substantial improvement over the null model, the moderate pseudo- $R^2$  values suggest that other factors may also contribute to shelter availability, highlighting the complexity of the problem at hand.

### **Discussion and Recommendations**

The results provide critical insights into how socioeconomic, environmental, and demographic factors influence shelter availability in Toronto, offering important insights that could inform public policy. Geographic disparities are striking; neighborhoods such as East Willowdale and Trinity-Bellwoods demonstrate greater odds of availability compared to areas like Wexford/Maryvale and Mimico-Queensway, which may reflect unequal resource distribution and systemic inequities. Similarly, sector analysis reveals that men's shelters face the greatest challenges, highlighting the need for tailored interventions across sectors.

Economic factors also play an important role in determining shelter availability. Rising unemployment correlates with marginal improvements in availability, possibly due to delayed reliance on shelters, while higher inflation significantly reduces availability, indicating economic pressures on both shelters and individuals seeking services. Environmental conditions, such as warmer temperatures, show minor but significant impacts, likely reflecting reduced seasonal demand, though precipitation and crime appear to have no meaningful influence.

Despite incorporating multiple, varied predictors, the model highlights underlying systemic challenges, with capacity lacking in any meaningful relationship with availability. This suggests that the shelter system is overburdened as a whole, and increasing capacity alone may not resolve accessibility issues without addressing the operational inefficiencies inherent in the system, most likely due to a lack of resources for non-profit and public services like shelter organizations.

While the predictive model demonstrates moderate fit and accuracy in its assessment, the challenges and limitations in identifying shelter availability cases reflect the need for additional data and more sophisticated modeling techniques. These findings suggest that resource allocation should prioritize neighborhoods and sectors facing the greatest barriers to access, with targeted investments during inflationary periods to alleviate economic pressures. Operational audits of shelter systems could identify inefficiencies and optimize resource utilization, while integrating immigration trends and broader socioeconomic data could enhance future predictive models. By addressing these systemic issues, public policy can more effectively ensure equitable and accessible shelter services across Toronto.

We also recommend that future iterations of the model consider a longer time period, spanning 10 years instead of 3, which would also reduce the overall influence of covid-specific factors. It would also be beneficial to approximate city level immigration and population changes on a monthly level. We also suggest in depth interviews with those managing Toronto's shelters to curate an improved selection of key variables, for more accurate model performance.

## **Conclusion**

This research provides a nuanced understanding of the factors influencing shelter availability in Toronto, revealing both systemic challenges and opportunities for targeted intervention. Geographic and sectoral disparities emphasize the inequities in resource allocation, while economic pressures such as inflation further strain the shelter system. By performing a multiple logistic regression, we have determined that increasing capacity by itself will not resolve accessibility issues, as availability is impacted by broader operational inefficiencies and economic constraints rather than sheer capacity limits.

The model demonstrates moderate predictive success but highlights the complex interplay of shelter dynamics with our selected variables. While socioeconomic and environmental variables offer valuable insights, they also underscore the need for more comprehensive data and advanced analytical techniques to fully capture the multifaceted nature of shelter availability.

Ultimately, this study points to the urgent need for policy measures that address not only capacity expansion but also systemic inequities and inefficiencies. By leveraging data-driven insights, Toronto can develop more equitable, responsive, and sustainable solutions to meet the growing demand for shelter services, and better address the needs of the homeless population of Toronto.

## References

- Argintaru, N., Chambers, C., Gogosis, E., Farrell, S., Palepu, A., Klodawsky, F., & Hwang, S. W. (2013). A cross-sectional observational study of unmet health needs among homeless and vulnerably housed adults in three Canadian cities. *BMC Public Health*, 13(1), 577.
- City of Toronto. (2021). Social procurement program update: Progress and next steps [Background file].
- City of Toronto. (2023). Annual review of statistics on deaths of shelter residents.
- City of Toronto. (2023, 2024). Annual review of statistics on deaths of shelter residents. City of Toronto. Retrieved from <https://www.toronto.ca>
- City of Toronto. (2024). Daily Shelter & Overnight Service Occupancy & Capacity [Dataset]. <https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/shelter-census/>
- City of Toronto. (2022, April 12). Find Your Neighbourhood (Toronto, Ontario, Canada). City of Toronto; City of Toronto. <https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/find-your-neighbourhood/>
- City of Toronto. (n.d.). Shelter census.
- City of Toronto. (n.d.). Toronto maps.
- Dionne, M.-A., Laporte, C., Loeppky, J., & Miller, A. (2023). A review of Canadian homelessness data. *Income Research Paper Series*.
- Falconer, R., Kneebone, R., & Wilkins, M. (2024). Social policy trends: Homeless shelter use in Toronto. *The School of Public Policy Publications*, 17(1).
- Government of Canada, S. C. (2007, June 19). Consumer Price Index, monthly, not seasonally adjusted. <https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1810000401>
- Government of Canada, S. C. (2024). Ontario's labour market | ontario.ca. <https://www.ontario.ca/page/labour-market#reports>
- Jadidzadeh, A., & Kneebone, R. (2018). Patterns and intensity of use of homeless shelters in Toronto. *Canadian Public Policy*, 44(4), 342–355.

National Oceanic and Atmospheric Administration. (2024). Toronto Weather Records.

<https://www.extremeweatherwatch.com/cities/toronto>

Noble, A., Owens, B., Thulien, N., & Suleiman, A. (2022). The impact of the COVID-19 pandemic on youth experiencing homelessness in Toronto, Canada. PLOS ONE, 17(8), e0273502.

Ontario Ministry of Finance. (2024). Chapter 1: Demographic trends and projections 2024.

In Ontario's long-term report on the economy 2024.

Toronto Police Service. (2024). Major Crime Indicators Open Data [Dataset].

<https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-open-data/about>

## Appendix

### Data Pre-processing Documentation

1. The shelter data was loaded, and the date values were converted to the correct data types - 180,767 records.
2. The major crime indicators data was loaded, irrelevant columns removed, date values converted to date type, and observations outside our time scope (2021-2024) were removed - 159,632 records.
3. The location indicator provided in the daily shelter occupancy data is using the postal codes for the addresses, while the location indicator for the crime data is by neighborhood. In order to converge the two by extracting the address for each shelter ID, and manually locating the neighborhood number at each address using the Toronto Neighborhood Finder (City of Toronto, 2022).
4. The mapped locations were used to left-join the crime data to the shelter data, and removed shelter records outside of Toronto's boundaries - 173,642 records.
5. Loaded the weather data, converting temperature and precipitation variables to numeric data types, and left-join the combined data with the weather data - 173,642 records.
6. Loaded the CPI data, create new variables in both the CPI data and combined data for year and month, and left join the combined data with the CPI data - 173,642 records.
7. Loaded the unemployment data, create new variables for year and month, and left-join the combined data with the unemployment data - 173,642 records.
8. As half the shelters in the shelter data record data in terms of rooms, and half in terms of beds, we needed to standardize the data. As it was easier to generalize one bed per person, we isolated all shelters that record data using rooms, and called them to request the number of beds per room. Out of 32 shelters, 3 provided the information, and we located the number of beds in five other shelters using online resources (shown below) - we applied the average beds per room to standardize the capacity and availability variables - 173,642 records.

| <b>Shelter</b>                        | <b>Rooms</b> | <b>Beds</b> | <b>Beds per room</b> | <b>Source</b>  |
|---------------------------------------|--------------|-------------|----------------------|--|
| Toronto Community Hostel              | 3            | 34          | 11.33                | <a href="#">Source</a><br>(last updated Nov 2023)  |
| Christie Refugee Welcome Centre, Inc. | 29           | 75          | 2.59                 | <a href="#">Source</a><br>(inferred from the daily occupancy rate and the average residents per night) |
| Homes First Society                   | 98           | 337         | 3.44                 | <a href="#">Source</a>   |
| City of Toronto (Robertson House)     | 32           | 90          | 2.81                 | <a href="#">Source</a><br>(updated Dec 2023)   |
| Street Haven At The Crossroads        | 33           | 32          | 0.97                 | <a href="#">Source</a>   |
| Fort York Residence                   | 39           | 116         | 2.97                 | Call   |
| Thistledown Community Centre          | 27           | 83          | 3.07                 | Call   |
| Covenant House Toronto                | 34           | 96          | 2.82                 | Call   |

9. Removed all NA observations from the data frame - 167,079 records.
10. Converted all categorical variables to factors - 167,079 records.
11. Resample the minority class to create an equal balance between availability and no availability - 242,584 records (final data frame).