Navigating Toronto's Homelessness Landscape during Pandemic: A Analysis of the dynamics for age group composition of Toronto Homeless community change with the Spread of Covid between 2020 and 2024*

Yiyi Feng

April 18, 2024

Against the backdrop of the COVID-19 pandemic, our study examines the relationship between the number of COVID-19 cases reported in Toronto from 2020 to 2024 and the count of homeless individuals across different age groups in the city's shelter system. While it's widely recognized that pandemics tend to correlate with an increase in homelessness, their specific impact on various age demographics of the homeless population remains less explored. By generating a series of linear models, our findings reveal a weak negative correlation between the overall homeless population and reported COVID-19 cases, whereas the correlation becomes more positive with increasing age among the homeless. The models predict that as the number of COVID-19 cases decreases, the population of homeless individuals aged 45 and above is projected to decline, while those below 45 are expected to increase. This suggests a potential upward trend in the percentile of younger individuals within the homeless community in the future.

Table of contents

*Code and data are available at: https://github.com/kqlqkqlqF/Trend-of-Toronto-Shelter-System-Flow-with-Covid-in-2020-2024.git.

1 Introduction [1]

Undoubtedly, homeless individuals in Toronto have been facing serious challenges in accessing nighttime accommodation in recent years. As early as 2020, the general manager of the city's shelter, support, and housing administration openly stated that an average of 72 people were being turned away from shelters due to overcrowding every day, a number that only continued to rise (Gibson 2023). By the winter of 2022, this figure had surged to a staggering 168 people (National 2022). For residents living in Toronto, this reality is apparent - despite shelter occupancy rates exceeding 98% (Toronto-Street-Needs-Group 2021), a significant number of individuals still resort to sleeping on the streets, a reality witnessed firsthand by residents and students alike.

The COVID-19 pandemic between 2020 and 2022 undoubtedly dealt a significant blow to many people's economic situations. As a result, the outbreak is widely believed to correlate positively

[1] Please check https://github.com/kqlqkqlqF/Trend-of-Toronto-Shelter-System-Flow-with-Covid-in-2020-2024.git for more information.

with the overall count of homeless individuals; the more COVID-19 cases reported, the more severe the outbreak, and consequently, the higher the number of homeless people. But is this the case? By 2024, the number of COVID-19 infections has been steadily decreasing, making it more likely for us to derive more accurate analyses based on the data compared to when the pandemic first erupted. We aim not only to uncover the relationship between the overall homeless population in Toronto and the pandemic but also to explore the correlation between the number of homeless individuals across different age groups within the homeless community. There are two main reasons for proposing this idea. Firstly, while the relationship between the total homeless population and the pandemic is not a secret, we lack insights into the individual trends within the diverse homeless population. Secondly, COVID-19 has been shown to infect and lead to fatalities more easily in older adults and children with low immunity, thereby demonstrating differential infection rates across age groups. Therefore, we hope to delve into its relationship with the homeless community across different age groups.

We started with deconstructing and analyzing the homeless population in Toronto to understand the trends within this community, including changes in total numbers and demographic composition. Utilizing data obtained from the open databases in Toronto, we examined the situation and demographic composition of the homeless population from 2018 to 2023, focusing on changes such as the number of homeless individuals transitioning between housing and shelter systems, the count of individuals utilizing the shelter system long-term, and demographic characteristics including gender, age, and status.

To elucidate the relationship between the number of homeless individuals in different age groups and the quantity of COVID cases, we sourced another dataset from OpenData Toronto, provided by Toronto Public Health, detailing monthly counts of infectious disease cases. This dataset spans from the initial detection and formal recording of COVID-19 cases in January 2020 to the present, documenting the number of COVID-19 infections each month. In total, we developed six models, each corresponding to the overall homeless population and the populations of five different age groups of homeless individuals, to examine their relationships with the quantity of COVID-19 cases.

Our analysis revealed two key findings. Firstly, our established linear models indicated a weak negative correlation between the overall homeless population in Toronto and the number of COVID-19 cases. However, different age groups of the homeless exhibited distinct correlation patterns, primarily manifested as a negative correlation for those under 45 years old and a positive correlation for those aged 45 and above. Additionally, according to the predictions of these models, as the spread of COVID-19 is effectively controlled in the future and cases decrease, there may be an increase in the proportion of younger individuals within the homeless population.

In summary, in this study, we initially investigated the annual transitions of homeless individuals between housing, shelters, and other statuses within the shelter system in Toronto from 2018 to 2023. Subsequently, we analyzed the composition of the homeless population and trends in age and gender distribution among them. Finally, we fitted linear models between the overall homeless population and five different age groups of the homeless population with

COVID-19 cases. We then analyzed the performance of these six linear regression models and made predictions for the future.

Understanding the trends and dynamics of homelessness in Toronto is of paramount importance, given its implications for public policy and social welfare. By finding out the challenges faced by the Toronto shelter system, we aim to inform policymakers and stakeholders in their efforts to address homelessness in the city.

2 Data

2.1 Data Source

The data for this analysis was collected from the shared Toronto Open Database (Shelter-Support-&-Housing-Administration 2018). This study utilizes and analyzes the dataset titled "About Toronto Shelter System Flow" and "About Monthly Communicable Disease Surveillance Data". The "About Toronto Shelter System Flow" dataset contains information on the transition of homeless individuals in Toronto shelters from January 2018 to March 2024, including gender, age, quantity, and population groups individuals belong to (such as refugees). Additionally, the dataset provides information on the number of people leaving and entering shelter systems. It is released by the Toronto Shelter, Support, and Housing Administration, updated monthly, and has a high level of credibility. The "About Monthly Communicable Disease Surveillance Data" dataset contains several communicable disease cases reported each month from 2016 January to 2024 February in Toronto, while COVID-19 cases were included since 2020 January. Both of the datasets have received full marks for freshness, metadata, accessibility, completeness, and usability on the Opendatatoronto website. Also, the "About Toronto Shelter System Flow" was provided by the Toronto Shelter and Support Agency, while the "About Monthly Communicable Disease Surveillance Data" dataset was published by the Toronto Public Health Agency. Therefore, we consider the content of these datasets to be highly credible and utilize it as the primary data source for this paper. Since the occurrence of COVID-19 cases spans from January 2020 to February 2024, we excluded data on homeless individuals from other periods when constructing the models to maintain consistency in the analysis process, these parts of data remained in the other parts of this study, such as overall analyzation of homeless community composition in Toronto.

2.2 Features

For the dataset of shelter system, upon entering shelters, individuals utilizing shelter services are required to provide their name, age, gender, and group affiliation, which are recorded in the database. This dataset only records homeless individuals using overnight shelter services and does not include those utilizing other welfare policies, such as receiving free food or vaccinations. In the data, homeless individuals are divided into five age groups: under 16, 16 to 24, 25 to 44, 45 to 64, and over 65. Gender is categorized as male, female, and transgender/non-binary/two-spirit. The population group includes chronic, refugees, families, youth, single adults, non-refugee, and indigenous. however, the indigenous group was only included in the statistics starting from January 2022, as the authors of the dataset stated their intention to collect more detailed data, hence adding the subdivision. Here, chronic refers to homeless individuals who have continuously used shelter services for more than 180 days. Regarding the documentation of homeless transitions, the dataset provides six subdivisions: newly identified, return from housing, return to shelter, moved to housing, became inactive, and actively homeless. Newly identified refers to people who entered the shelter system for the first time; returned from permanent housing refers to people who previously used the shelter system, then moved to permanent housing, and have now returned; returned to shelter refers to people who were previously using the shelter system, then did not use the system for 3 months or longer, and have now returned; moved to permanent housing refers to people who were using the shelter system and have moved to permanent housing; became Inactive refers to people who last accessed shelter services three months ago; actively homeless refers to people who have used shelter services at least one time in the past three months and have not moved to permanent housing. The data recording for COVID-19 cases is much simpler. The original dataset contained records of monthly reported numbers for various infectious diseases, including HIV, gonorrhea, and so on. We simply extracted the relevant records for COVID-19 from the original dataset and performed some basic formatting.

2.3 Methodology

The data analysis was conducted using R (R Core Team 2022), a versatile statistical programming language. We utilized a range of packages to enhance our analysis. The tidyverse (Wickham et al. 2019) suite of packages provided a comprehensive toolkit for efficient data manipulation and visualization. Package ggplot2 (Kassambara 2023) allowed us to create compelling visualization. The here (Müller 2020) package simplified file management within our project directory structure. Additionally, kableExtra (Zhu 2021) was employed to generate visually appealing and customizable tables, enhancing the presentation of our findings. For Bayesian analysis, we utilized the rstanarm (Goodrich et al. 2020) package, which provided an elegant interface to Stan, a cutting-edge platform for statistical modeling and computation. This allowed us to estimate relationships within our data using a Bayesian framework, providing valuable insights into our research questions. Report generation was seamlessly managed using knitr (Xie 2023), enabling the integration of R code within our document. Other essential packages included tibble (Müller and Wickham 2022), stringr (Wickham 2020), lubridate (Grolemund and Wickham 2020), janitor (Firke 2023), and testthat (Wickham and RStudio 2020), each contributing to various aspects of our data analysis process, from data manipulation to quality assurance.

As mentioned before, the data cleaning process is simple for the COVID-19 dataset, so for here we will instoduce the data processing procedure for the homeless dataset. Due to the clarity of the data itself, our data cleaning process primarily focused on converting the raw data dates into the yyyy-mm-dd format and selecting the data needed for producing each figure respectively to ensure code was organized and minimize the amount of code in the final QMD file. For the first chart, intended to reveal Trends in the Annual Counts of Homeless Individuals Transitioning within the Toronto Shelter System between 2018 and 2023, we retained the data for the six transition status categories from monthly data and aggregated data within the same year for ease of subsequent chart generation. For the second chart, revealing Trends in the Composition of Homeless Population Groups and Age Distribution within the Toronto Shelter System between 2018 and 2023, we removed all data except for age and population groups. Similarly, for the third chart, Trends in Gender Composition of Homeless Population Across Different Population Groups within the Toronto Shelter System between 2018 and 2023, we retained only gender and group affiliation data. The linear model utilized the cleaned dataset for the second figure, and additional cleaning was performed to remove all data that did not belong to the period 2020 January to 2024 February.

3 Model

3.1 Model Setup

Our objective is to investigate how the number of individuals belonging to different age groups using Toronto's overnight shelter system has changed with the number of COVID-19 cases from January 2020 until 2024 February, and to generate a total of six models to predict trends in this relationship. The six models are as follows: the total number of shelter users in Toronto, the number of shelter users under 16 years old, the number of shelter users aged 16 to 24, the number of shelter users aged 25 to 44, the number of shelter users aged 45 to 64, and the number of shelter users aged 65 and over. These models examine the linear relationship between these shelter user demographics and the number of COVID-19 cases. In these models, the number of homeless individuals using shelters in a given month was represented by y_i, while the number of COVID-19 cases recorded is represented by x_i for each month i:

$$
\begin{aligned}
y_i &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \beta_0 + \beta_1 \times \text{AGE}_i && (2)\\
\beta_0 &\sim \text{Normal}(90, 10) && (3)\\
\beta_1 &\sim \text{Normal}(-0.1, 0.05)\sigma \qquad\qquad \sim \text{Exponential}(1) && (4)
\end{aligned}
$$

The analysis was conducted in R using the rstanarm package, which facilitates the integration of prior beliefs about parameters and the estimation of posterior distributions. We applied normal prior distributions centered around our initial estimates to the intercept (β_0) and slope (β_1) parameters, informed by exploratory data analysis. Our choice of priors aimed to capture reasonable assumptions about parameters before observing the data.

3.2 Model Justification

We assume that the coefficient (β_1) since January 2020 will be positive, as it is commonly understood that the global pandemic severely impacted the economic situation worldwide, leading to increased economic and health issues for many individuals and consequently resulting in a higher number of homeless individuals. Additionally, COVID-19 is more likely to spread within the homeless population due to their lack of fixed abodes, unlike those in shelters where people are densely packed. Therefore, the slope (β_1) is our main focus for verification, as it will determine the expected changes in the number of homeless individuals across different age groups by 2024 as the impact of the pandemic diminishes. Furthermore, we believe that for different age groups, the oldest and youngest homeless populations will maintain a positive correlation with the number of COVID-19 cases. This prediction stems from the higher infection rates of the novel coronavirus among children and the elderly. Moreover, we anticipate that the intercept (β_0) will also remain positive, as many individuals have been utilizing shelter systems since January 2020. Therefore, regardless of whether the outcome shows positive or negative growth, the intercept will be positive.

4 Result

4.1 Trends in Homelessness Transitioning within the Toronto Shelter System

Table 1: Annual Counts of Homeless Individuals Transitioning between Housing, Shelters, and Other Status Within the Toronto Shelter System between 2018 and 2023

Year	Returned from Housing	Returned to Shelter	Newly Identified	Moved to Housing	Became Inactive	Actively Homeless
2018	776	6055	14442	8135	10420	110052
2019	873	6175	13124	8293	11037	117078
2020	891	5308	7617	6094	9302	99777
2021	989	5024	8297	3409	8690	99580
2022	940	4116	9795	4385	8738	117222
2023	688	3197	10183	5927	7507	125274

The data provided in Table 1 and Figure 1 offers insight into the dynamic changes in the homeless population within the Toronto shelter system from 2018 to 2023.

A notable trend observed is the fluctuation in the annual count of newly identified individuals entering the shelter system. Peaking in 2018 and gradually declining over the following years, the trend suggests a persistent need for shelter services, with a gradual increase observed from 2020 onwards.

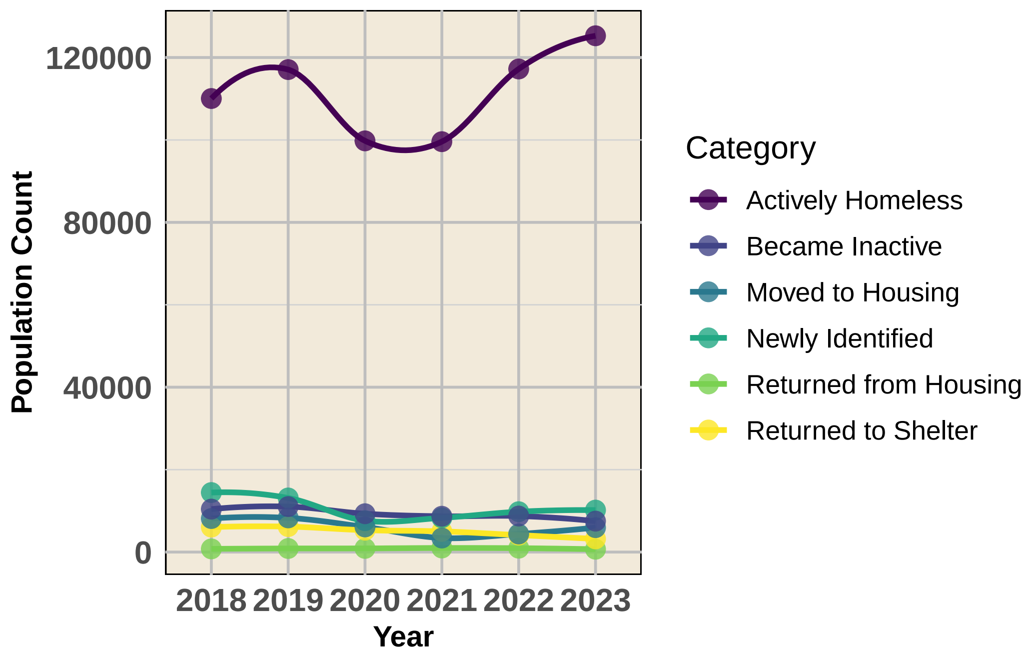

Figure 1: Trends in the Annual Counts of Homeless Individuals Transitioning within the Toronto Shelter System between 2018 and 2023

Another significant aspect of the data is the return of individuals from permanent housing to the shelter system. This phenomenon is concerning as it indicates potential housing instability for the population. Fortunately, the numbers have remained relatively stable from 2018 to 2023, and they are lower than those transitioning to permanent housing each year. However, the declining trend in individuals moving to stable housing from 2018 to 2021 is worrisome. Although there is a slight increase in 2022 and 2023, it remains below the 2018 levels, suggesting a decrease in individuals transitioning to permanent housing.

The data also captures changes in individuals returning to shelters and those becoming inactive within the shelter system. While there is a gradual decline in the number of individuals becoming inactive homeless over the five years, the decrease is minimal, similar to the trend in individuals returning to shelters. Without further explanation, it's challenging to draw reliable conclusions on the decrease in the number of inactive homeless individuals or the reasons behind individuals returning to shelters since their experience after leaving the shelter system is unknown. They could be moving to other cities, finding housing, or dying. Therefore, if there's no further detail for the reason they left the shelter system, it is hard to understand the implications of the data fully, and no conclusion can be drawn.

Lastly, individuals from the actively homeless category highlight the ongoing demand for shelter services and support. Although there was a decline in the number of actively homeless individuals in 2020 and 2021, it surged to new highs in 2022 and 2023. This suggests a significant increase in the number of individuals utilizing overnight shelter services in the past two years, indirectly indicating a potential increase in Toronto's homeless population from 2022 to 2023.

In conclusion, these data paint a complex picture of homelessness within the Toronto shelter system, with a trend of increasing actively homeless individuals.

4.2 Composition of Homeless Population Groups and Correcponding Age Distribution

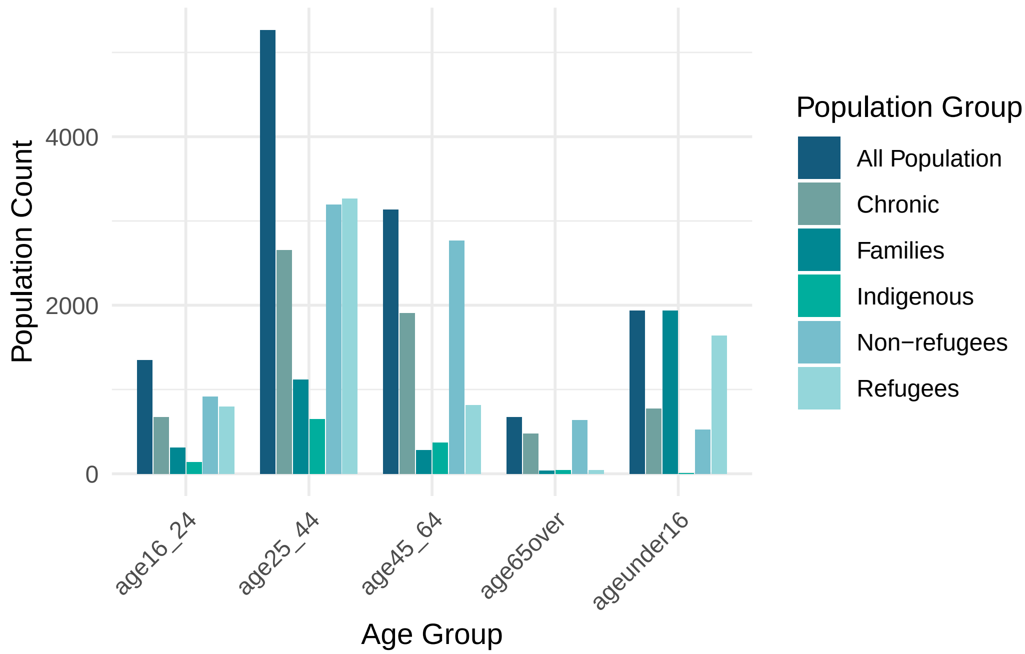

Figure 2: Composition of Homeless Population Groups and Age Distribution within the Toronto Shelter System between 2018 and 2023

Before beginning the analysis of Figure 2, several points need to be explained. Firstly, within the population groups depicted in this graph, apart from refugees and non-refugees, the other options are not mutually exclusive. This means that a person can simultaneously belong to the chronic, families, and refugee groups, but cannot be a refugee and a non-refugee at the same time. This explains why the sum of individuals in these population groups exceeds the number in all population categories. Secondly, as mentioned earlier, the number of individuals in the indigenous group was only recorded starting in 2020, so the actual number of indigenous individuals who have used the shelter system should be higher than what is shown in the graph.

Individuals aged between 25 and 44 constitute the largest portion of the homeless population, while those aged between 45 and 64 represent the second largest group, and those aged 65 and above represent the smallest group. From the perspective of population groups, non-refugees constitute the largest group, followed by refugees and then chronic individuals, with indigenous individuals being the smallest group. From this distribution trend, we can infer the following information:

Firstly, except for the age group under 16, chronic individuals constitute a significant portion of the homeless population in other age groups, indicating that a considerable number of homeless individuals rely on overnight shelter services for an extended period and find it difficult to secure permanent housing, thus perpetuating their homelessness.

Secondly, although refugees constitute a smaller proportion of the homeless population compared to non-refugees overall, in the age group under 16, the number of refugees far exceeds the proportion of non-refugees. This suggests that there are a significant number of parenting teenagers or even younger children among the refugee population who are homeless. Additionally, we observe that the proportion of families in the age group under 16 also significantly increases, slightly surpassing the proportion of refugees and almost equaling the "all population" category. This evidence confirms our previous conclusions and supports the notion that these young refugee homeless individuals are likely to be wandering alongside their refugee family members.

Thirdly, unlike refugees, homeless individuals categorized as non-refugees are mainly distributed in the age group of 25 and above. Particularly in the age group of 65 and above, the number of non-refugees is almost equivalent to the "all population" category.

In summary, chronic individuals are prominent across age groups, highlighting challenges in securing permanent housing. Refugees tend to enter the shelter system as a whole family with their young children, indicating a concerning trend of homelessness among refugee families with children, while the non-refugees dominate the older age groups.

4.3 Gender Composition Across Homeless Population Groups

The data presented in Figure 3 provides insights into the gender dynamics within Toronto's homeless population across different population groups from 2018 to 2023. A consistent trend observed throughout the years is the higher representation of males compared to females and individuals identifying as transgender, non-binary, or two-spirit. This imbalance suggests a prevalent gender disparity among those experiencing homelessness, with males comprising a larger proportion of the homeless population.

However, within specific population groups, variations in gender composition emerge. For instance, in the chronic population group, which likely encompasses individuals experiencing long-term homelessness, both males and females are prominently represented, indicating a diverse demographic within this subgroup. In contrast, the families population group, which

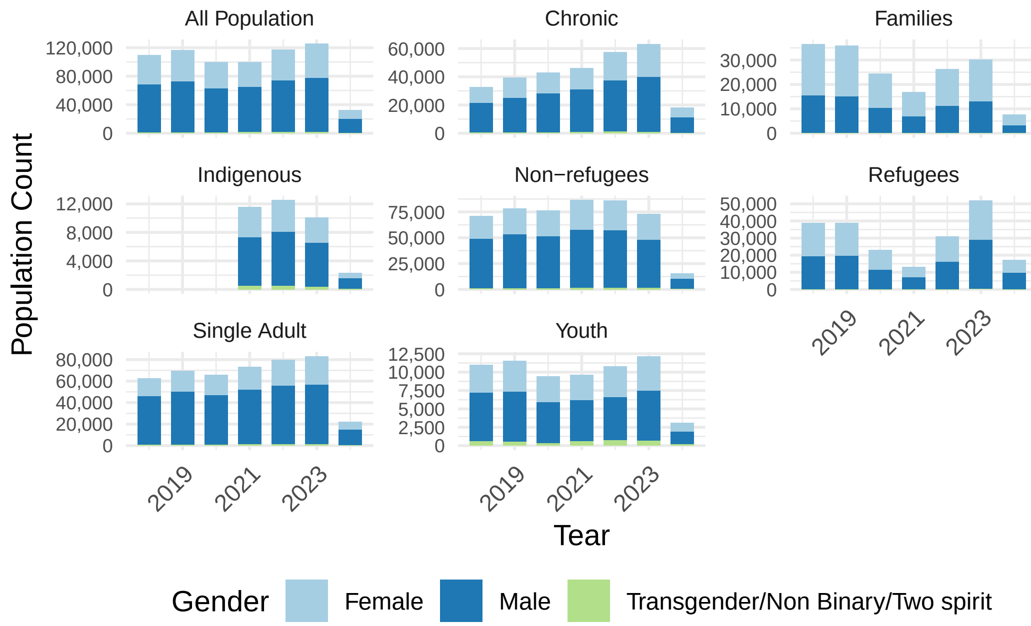

Figure 3: Trends in Gender Composition of Homeless Population Across Different Population Groups within the Toronto Shelter System between 2018 and 2023

likely consists of homeless families with children, demonstrates a higher count of females compared to males and individuals identifying as transgender, non-binary, or two-spirit. This gender distribution within homeless families may be influenced by factors such as caregiving responsibilities and access to support services, indicating the unique dynamics within this subgroup. The "Refugee" population group exhibits a relatively balanced distribution between males and females, with a smaller count of individuals identifying as transgender, non-binary, or two-spirit. This suggests a diverse composition within refugee populations experiencing homelessness, possibly influenced by factors such as migration patterns and resettlement challenges.

Furthermore, the data reveal temporal fluctuations in the chronic population group, which demonstrates a consistent increase in count from 2018 to 2023. This upward trend suggests a growing number of individuals experiencing long-term homelessness.

Trends of the Number of COVID-19 Cases demonstrated in Figure 4 reveal several peaks in the surge of COVID-19 infections, including from September 2020 to January 2021, February to June 2021, and November 2021 to February 2022. Although there were a few minor increases in infection numbers thereafter, overall, the number of infections has been consistently decreasing from February 2022 to February 2024.

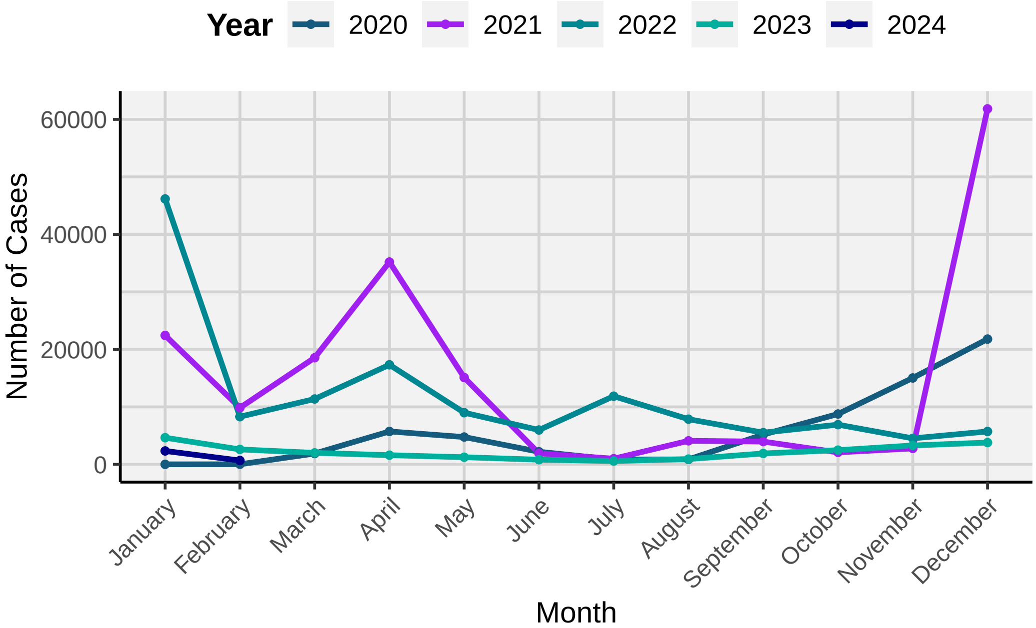

Figure 4: Trends of Number of COVID-19 Cases Reported each Month Since 2020 January to 2024 Febuary

4.4 Result of the Model

4.4.1 Model Coefficients

Table 2: Summary Statistics for the Coefficients for the Linear Models Across different Age Groups

Model	Slope	Intercept
under 16	-23.509	32736.956
16 to 24	-27.076	38385.833
25 to 44	-6.175	31935.634
45 to 64	2.505	1358.041
65 over	36.523	-12580.232
all population	-3.197	38020.076

Intercept (β_0) : The intercept represents the estimated initial count of homeless individuals using shelter overnight service in Toronto at the onset of the observed period, which corresponds to the starting point of the COVID-19 infection data in January 2020. The estimated intercept value varies from 38385.833 to -12580.232 across different age groups, while the intercept for all homeless populations is 38020.076, indicating the large number of homeless in Toronto at the start of the pandemic.

Slope (β_1) : The slope quantifies the average change in the number of homeless individuals using shelter overnight service in Toronto as per unit increase in COVID-19 infections. Across different age groups, the slopes vary from -27.076 to 36.523, indicating that different age groups correlated with the number of cases very differently. For all homeless populations, the estimated slope indicates a decrease of approximately 3.197 homeless individuals for each additional COVID-19 infection, indicating a negative correlation.

Explanation: The analysis suggests an overall negative correlation between the rise in COVID-19 infections and the total count of homeless individuals utilizing overnight shelter services in Toronto. However, this correlation varies significantly across different age groups. While older age groups show a positive correlation with the number of cases, younger age groups exhibit a negative correlation. Interestingly, the correlation becomes more negative as the age group gets younger, while it becomes more positive as the age group gets older. The intercept for the homeless aged 65 and above is -12580.232, which seems unreasonable as the minimum count of people cannot be negative. However, with a slope of 36.523, indicating a substantial increase, we can infer that there was a sharp rise in the homeless population aged 65 and above following the outbreak of the pandemic, leading to a negative intercept in the model.

4.4.2 Model Equation

For all homeless population:

$$y_i = 38020.076 - 3.197x_i + \varepsilon_i \tag{5}$$

For homeless under 16 years old:

$$y_i = 32736.956 - 23.509x_i + \varepsilon_i \tag{6}$$

For homeless between 14 and 24:

$$y_i = 38385.833 - 27.076x_i + \varepsilon_i \tag{7}$$

For homeless between 25 and 44:

$$y_i = 31935.634 - 6.175x_i + \varepsilon_i \tag{8}$$

For homeless between 45 and 64:

$$y_i = 1358.041 + 2.505x_i + \varepsilon_i \tag{9}$$

For homeless 64 years old above:

$$y_i = -12580.232 + 36.523x_i + \varepsilon_i \tag{10}$$

4.4.3 Residual

```
  ageunder16         age16_24          age25_44          age45_64
 Min.   :-16300   Min.   :-13770   Min.   :-12804   Min.   :-8346
 1st Qu.: -2331   1st Qu.: -5183   1st Qu.: -5073   1st Qu.:-6366
 Median : -1154   Median : -1811   Median : -2366   Median :-4215
 Mean   :     0   Mean   :     0   Mean   :     0   Mean   :    0
 3rd Qu.:  1562   3rd Qu.:  1727   3rd Qu.:  2013   3rd Qu.:  982
 Max.   : 50282   Max.   : 51990   Max.   : 52952   Max.   :53140
  age65over         total_population
 Min.   :-8091.58   Min.   :-13035
 1st Qu.:-5914.69   1st Qu.: -5187
 Median :-3879.50   Median : -1990
 Mean   :    0.00   Mean   :     0
 3rd Qu.:  -82.61   3rd Qu.:  1184
 Max.   :51153.07   Max.   : 53404
```

Since there's a total of six models and their performance on residuals is similar, we will only discuss the residuals for the total population model.

1. Minimum Residual: The minimum residual indicates that there is at least one observation for which the model's predicted total population was lower than the actual population by approximately 13035 individuals. This suggests the presence of potential outliers or factors not accounted for by the model.

2. First Quartile Residual: The first quartile (Q1) of the residuals implies that 25% of the observations have total populations that are less than 5187 individuals below the model's prediction.

3. Median Residual: The median residual indicates that half of the observations have residuals below 1990, suggesting a slight positive bias in the model's predictions. This means that, on average, the model tends to slightly underestimate the total population.

4. Third Quartile Residual:The third quartile (Q3) signifies that 75% of the observations have total populations within 1184 individuals of the model's predictions or better. This suggests that the model provides reasonably accurate predictions for the majority of observations.

5. Maximum Residual: The maximum residual suggests that there is at least one observation for which the model's predicted total population was higher than the actual population by approximately 53404 individuals. Similar to the minimum residual, this could indicate the presence of outliers or unaccounted factors.

Interpretation: The distribution of residuals has shown the models' performance and the presence of potential outliers or discrepancies between the model's predictions and the actual observations. Further investigation into the reasons behind significant deviations can help refine the model and improve the accuracy of its predictions. Notably, for homeless aged 65 and above, their 3rd quartile residual is negative, but the maximum value is similar to the other age groups. this further supported our point that the number of individuals aged 65 and above increased significantly after the outbreak of the pandemic.

4.4.4 Summery

Table 3: Summary of Statistical Coefficients for the Linear Models Across different Age Groups

	Std.Error	t.Value	R.Squared	F.Value
under 16	6.796663	-3.4588686	0.2494335	11.9637723
16 to 24	11.482579	-2.3579766	0.1337836	5.5600539
25 to 44	3.009244	-2.0518482	0.1047021	4.2100812
45 to 64	8.145784	0.3075477	0.0026205	0.0945856
65 over	30.079018	1.2142356	0.0393434	1.4743680
all population	1.624356	-1.9679893	0.0971330	3.8729820

The summary statistics for the six linear regression models were shown in Table 3. The model for the age group under 16 demonstrates a relatively low R-squared value of 0.25, suggesting that only 25% of the variability in the number of homeless individuals is explained by the predictor variables, coupled with a high F-value of 11.96, indicating a potential lack of statistical significance. Similarly, the models for age groups 16-24, 25-44, and over 65 exhibit low R-squared values ranging from 0.003 to 0.10, with corresponding F-values suggesting limited explanatory power and statistical significance. Since all R square values were far from 1, this indicates the reliability of this model is limited. For the "under16" model, the standard error is relatively high, suggesting variability in the estimated coefficients and potentially lower precision in predicting the relationship between COVID-19 cases and homeless individuals under 16 years old. The t-value is also negative, indicating that the coefficient may not be statistically significant. Similarly, for the "16to24" and "25to44" models, the standard errors are relatively high, indicating variability in the estimates and potentially lower precision in predicting the relationship between COVID-19 cases and homeless individuals in these age groups. However, the t-values are negative for these models as well, suggesting that the coefficients may not be statistically significant. In contrast, for the "over65" model, the standard error is relatively low, suggesting higher precision in estimating the coefficients, but the t-value is also negative, indicating potential insignificance. Finally, for the "all population" model, the standard error is moderate, suggesting moderate variability in the estimates, and the t-value is negative, indicating potential insignificance. These findings show there's space for further improvement in these models.

4.4.5 Graphing the Model

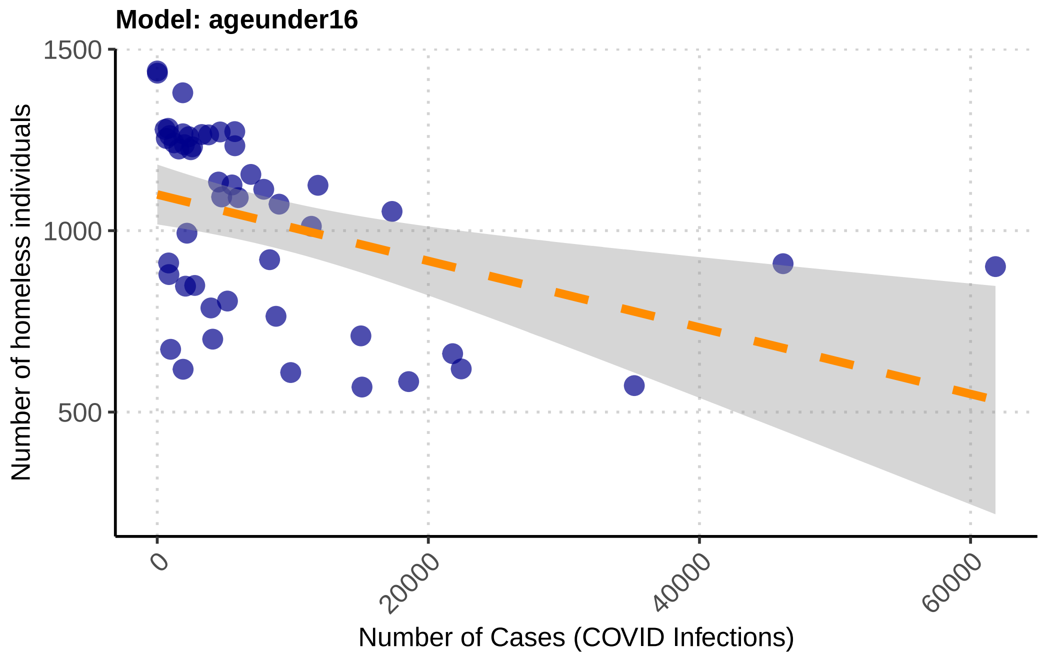

Figure 5: The Linear Relationship between Number of Cases of COVID-19 and the Number of individuals use Shelter Overnight Service in Different Age Group

In this series of figures, as previously mentioned, we observe a positive correlation between the number of homeless individuals aged 45 and above and the quantity of COVID-19 cases, whereas the count of individuals below 45 years old displays a negative correlation. Overall, the total homeless population exhibits a weak negative correlation with COVID-19 cases.

5 Discussion

5.1 Summery of Findings

The analysis of homeless population dynamics in Toronto from 2018 to 2023 revealed fluctuating trends in shelter entry, with a peak in 2018 followed by a gradual decline and a slight increase from 2020 onwards. The data also highlighted potential housing instability through

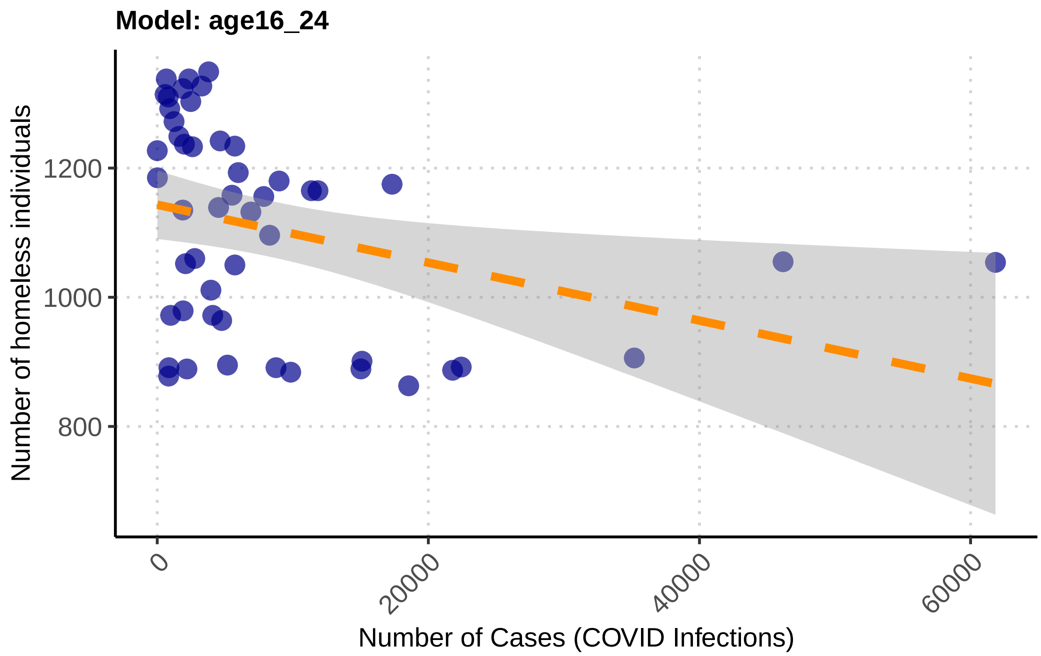

Figure 6: The Linear Relationship between Number of Cases of COVID-19 and the Number of individuals use Shelter Overnight Service in Different Age Group

individuals returning from permanent housing to shelters, although overall stability was observed. Chronic individuals were prevalent across age groups, indicating persistent challenges in securing permanent housing. Refugee families, often with young children, contributed to homelessness trends, while non-refugees dominated older age groups. Gender disparities were evident, with males comprising a larger proportion of the homeless population, especially among chronic individuals, while females were more represented in homeless families. These findings indicate an alarming rise in homelessness, notable shifts in demographic composition, and a gender imbalance, all highlighting the multifaceted nature of homelessness.

From the generated models, we found that the overall trend in the homeless population is weakly negatively correlated with the number of COVID-19 cases, while the age group above 45 exhibits a positive correlation and the age group below 45 shows a negative correlation. Moreover, overall, older age groups of homeless individuals tend to show stronger positive correlations, while younger age groups of homeless individuals tend to exhibit stronger negative correlations. Following this pattern, as the spread of the pandemic is further controlled, the occurrence of COVID-19 cases is expected to decrease, while the total homeless population is anticipated to increase. Additionally, the proportion of older individuals within the homeless population will gradually decrease, while the proportion of younger individuals will gradually increase.

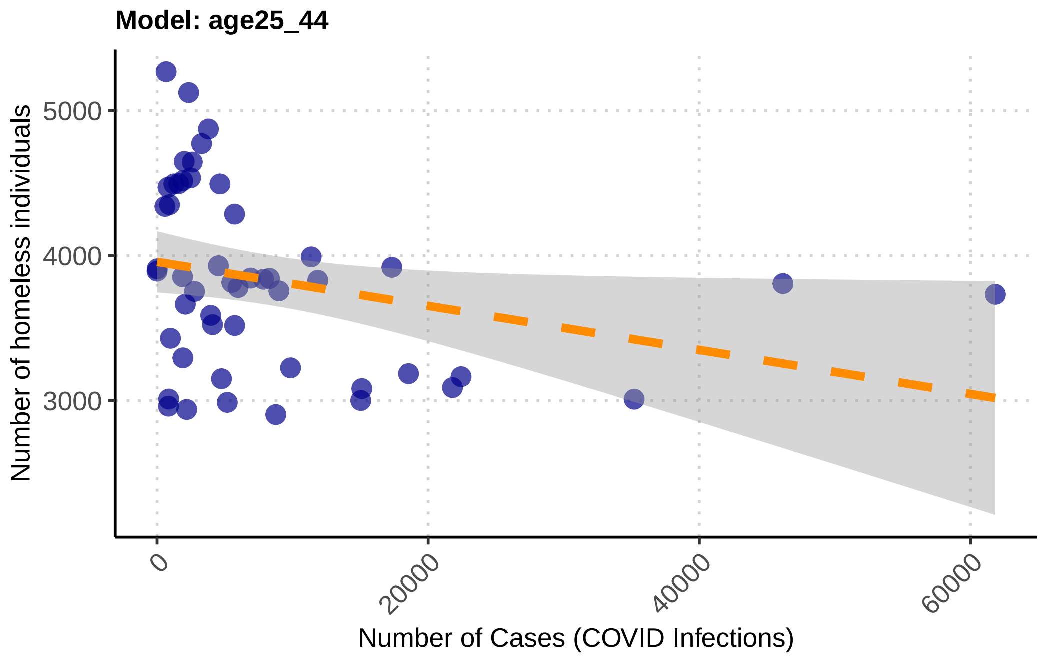

Figure 7: The Linear Relationship between Number of Cases of COVID-19 and the Number of individuals use Shelter Overnight Service in Different Age Group

5.2 Things learned from this study

Firstly, we found that the process of data collection, analysis, and research to reach conclusions is much more complex than we initially anticipated. In the Toronto shelter system Flow dataset we utilized, data is recorded in detail, including information on whether homeless individuals leave shelters to move into permanent housing. However, despite the comprehensive nature of the data, our analysis still falls short. As we raised in the results section, we have no way of knowing where homeless individuals go or what they do after becoming inactive, making it difficult to determine whether the increase or decrease in inactive status is beneficial or detrimental. I believe that in the future, there could be further sharing of data between shelter systems, such as sharing homeless individual information across cities or provinces, which would allow us to track their movements after leaving shelters in another jurisdiction.

Secondly, the proportion of child/youth homeless individuals is much higher than we anticipated, which came as a surprise. From the data, it appears that many of these young homeless individuals are refugees who are traveling with family members, with only a small fraction being residents. Additionally, around one-third of homeless children are classified as chronic, meaning they have been homeless for an extended period. These data indirectly highlight the difficulties refugees face in finding employment and obtaining legal status in a new country, which in turn affects the next generation. We hope that the government can develop relevant

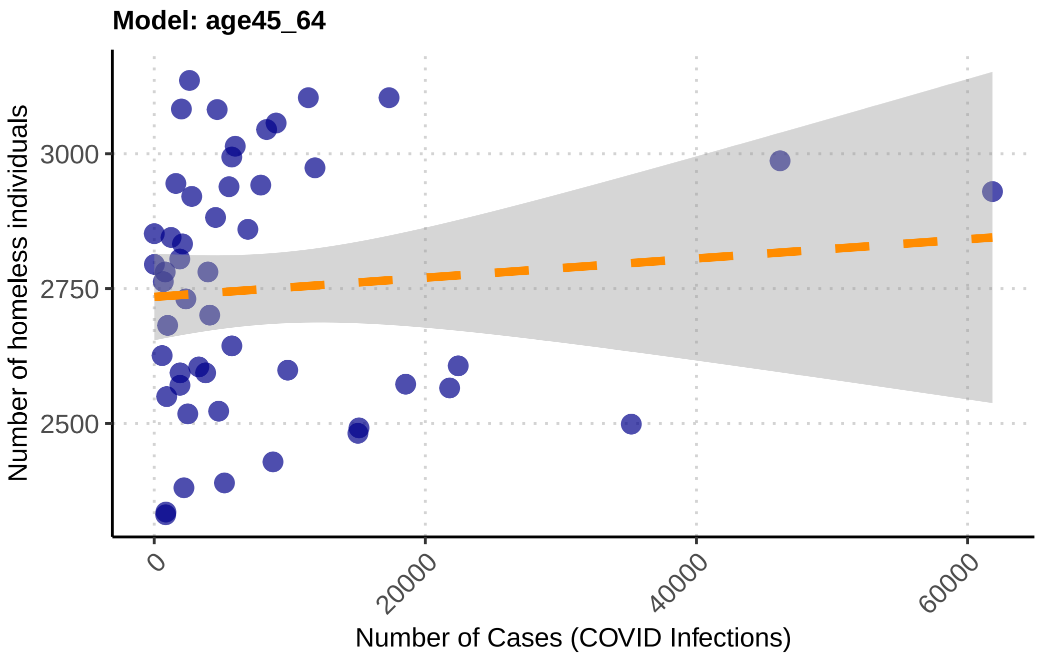

Figure 8: The Linear Relationship between Number of Cases of COVID-19 and the Number of individuals use Shelter Overnight Service in Different Age Group

welfare policies for these refugee children to reduce the number of homeless children.

Thirdly, conducting overall homeless population counts poses significant challenges. Firstly, many homeless individuals do not stay overnight in shelters and spend both day and night on the streets or in parks. This means that we do not have an efficient and accurate way of collecting identity information for these homeless individuals. Secondly, as mentioned earlier regarding refugee issues, many homeless individuals lack legal status or identification documents, making it difficult for staff to determine the reliability of the information they provide. Additionally, the transient nature of homelessness complicates both the accurate collection of homeless individuals' numbers and their identity information. Some individuals may be homeless today, find temporary accommodation the next day, and then suddenly lose their accommodation and resume homelessness after a few days. This instability makes it challenging to accurately collect both the number and identity information of homeless individuals.

5.3 Limitations

Although the analysis provides insights into the dynamics of homelessness in Toronto, there are still some limitations to consider. Firstly, these data primarily reflect individuals utilizing shelter services and may not capture the entire homeless population, including those living on the streets or in unstable housing conditions. Additionally, our data have limitations. For

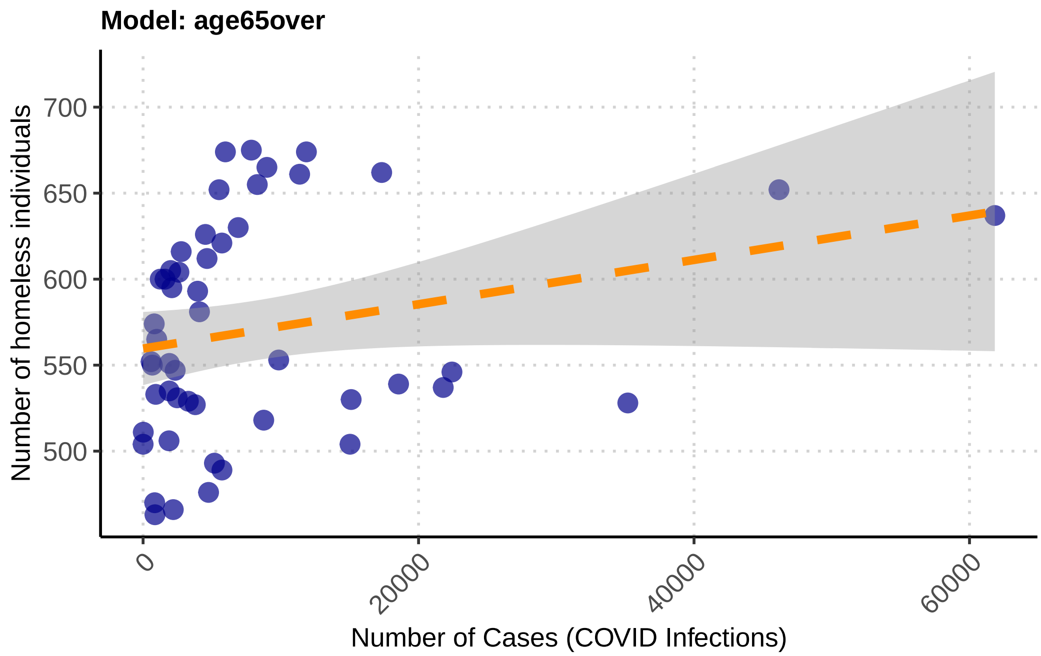

Figure 9: The Linear Relationship between Number of Cases of COVID-19 and the Number of individuals use Shelter Overnight Service in Different Age Group

example, we cannot track the whereabouts of all homeless individuals who leave the shelter system; the data only record numbers without monitoring their activities. The reasons for individuals becoming inactive homeless may vary significantly, such as death, disappearance, or relocation, leading to confusion and ambiguity in predicting future dynamics. Thirdly, our dataset is based on the trends of homelessness from 2018 to 2023, which were affected by the multifaceted impacts of the COVID-19 pandemic, including but not limited to: increased difficulty in data collection due to social distancing measures, the health, economic, and social impacts of COVID-19 on the homeless population, and the potential closure of shelters due to the spread of the disease. These factors may reduce the effectiveness and accuracy of data collection. Moreover, we can observe a significant decrease in shelter utilization in 2020 and 2021 during the pandemic, which contradicts our initial assumptions. Therefore, we have some reservations about the reliability of the data during this period. Furthermore, there exist significant differences in sample sizes among different age groups. As the number of homeless individuals aged 65 and above is inherently low, the credibility of model predictions for this group is naturally lower compared to the 25-44 age group. However, this issue is particularly concerning because the scarcity of homeless individuals aged 65 and above is a fact, and we cannot collect a similar volume of data as we can for the 25-44 age group due to the limited presence of elderly homeless individuals.

In terms of model analysis, as mentioned earlier, the COVID-19 pandemic is not the sole factor

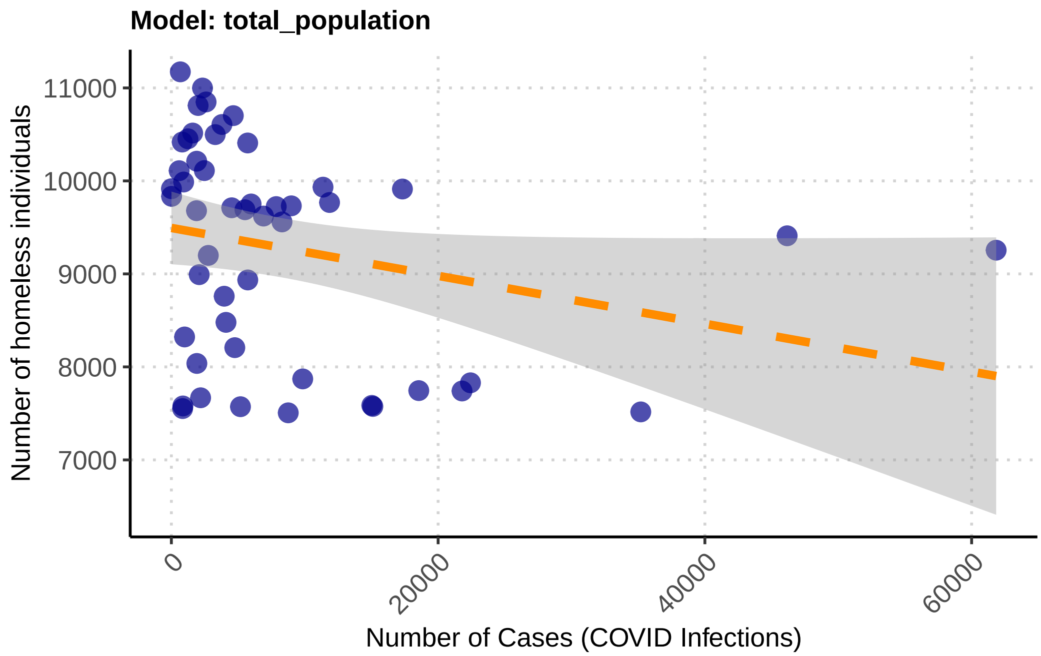

Figure 10: The Linear Relationship between Number of Cases of COVID-19 and the Number of individuals use Shelter Overnight Service in Different Age Group

influencing the homeless population, and their correlation is unlikely to be entirely linear. The number of homeless individuals is closely tied to various factors such as economic conditions, housing prices, welfare policies, and more. These factors impose significant limitations on our models. Additionally, since the outbreak and containment of the pandemic occurred within approximately three years, our dataset and data reliability are relatively limited. This project itself is not conducive to long-term monitoring (e.g., over five to ten years), as determined by the specific form of viral epidemic. Moreover, due to the sharp increase in infection numbers during concentrated outbreaks of the virus, there is significant variability in monthly infection counts. This results in a plethora of low-infection count data points and scarce data during periods of high infection counts, affecting the overall accuracy of the model predictions.

5.4 Future Study

As the homeless population continues to grow steadily and faces increasingly dire challenges in obtaining permanent housing, we urge the government to intervene urgently to expand housing capacity, reduce housing costs, and strengthen support services. Future research should explore the fundamental causes of homelessness, identify effective interventions to prevent homelessness from recurring and assess the long-term impacts of housing policies and support programs. By understanding the root causes of homelessness and implementing targeted interventions, people can be protected from homelessness, thereby reducing its occurrence at a more fundamental level. As time passes, the impact of the COVID-19 pandemic on homelessness dynamics is expected to diminish. If similar studies were to be conducted in the future, I believe it would be beneficial to first incorporate more factors, such as the average rental prices and cost of living in Toronto. Additionally, using non-linear models for prediction may yield better results. However, regardless of the approach, generating models for the homeless population poses significant challenges. Furthermore, it is essential to incorporate a wider range of comprehensive factors into our research when collecting data, which may include reasons for homelessness and dynamic trajectories, to provide deeper insights into which factors have a more profound impact on homelessness. Moreover, the government needs to establish a more comprehensive data collection and measurement system, including tracking the dynamics of inactive homeless individuals and conducting categorized surveys of homeless families. This will enable a more accurate assessment of the challenges faced by the homeless population and facilitate the development of targeted and effective measures to address these issues.

Bibliography

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gibson, Victoria. 2023. "Why the Homelessness Crisis Could Get Even Worse." 2023. https://www.thestar.com/news/gta/why-the-homelessness-crisis-could-get-even-worse/article_9054cdcc-f846-5a33-a9e4-f972860686a1.html.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2020. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Grolemund, Garrett, and Hadley Wickham. 2020. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

Kassambara, Alboukadel. 2023. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2022. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

National, CBC News. 2022. "Toronto Shelters Turning Away Average of 168 People a Day." 2022. https://www.youtube.com/watch?v=vUZhZIS-XRo&t=2s.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Shelter-Support-&-Housing-Administration. 2018. *About Toronto Shelter System Flow*. https://open.toronto.ca/dataset/toronto-shelter-system-flow/.

Toronto-Street-Needs-Group. 2021. "Toronto 2021 Street Needs Assessment Result," 11–18. https://www.toronto.ca/legdocs/mmis/2021/ec/bgrd/backgroundfile-171729.pdf.

Wickham, Hadley. 2020. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and RStudio. 2020. *Testthat: Unit Testing for r*. https://CRAN.R-project.org/package=testthat.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.