Navigating Toronto's Homelessness Landscape: A Analysis of Toronto Shelter System Occupied Situation with the Spread of Covid and Average Housing Price in Toronto from 2018 to 2023

Yiyi Feng

April 2, 2024

In the context of economic struggles and the impact of the pandemic between, this analysis deeply examines the number of indivisuals recorded in Toronto's shelter system from 2018 to 2023. It explores changes in the homeless population across population groups, genders, and age groups, studying population dynamics closely, and specifically uncover the relationship between housing price, pandemic and number of homeless. A model was generated to predict the trend for the number of individuals using shelter systems in 2024-2026. The modeling predicts an increase in homelessness as the rising of housing price, emphasizing the need for proactive approaches. These discoveries reveal the necessity for more shelter facilities to address the growing challenges ahead.

Table of contents

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.

Our understanding of the changing trend in the composition of the homeless community in Toronto is little. While the overall number of homeless individuals is documented, the specific demographic shifts within this population, such as variations in age groups and gender distribution, have not been comprehensively explored. Understanding these trends is essential for developing targeted interventions and resource allocation strategies to address the diverse needs of the homeless population in Toronto more effectively.

To do this, we analyzed Toronto's homeless population, to understand the trends in this community, including changes in overall numbers and demographic composition. Leveraging data obtained from the Toronto Open Database, we aim to find the complexity and evolving trends of Toronto's homeless community. We analyze the status and demographic composition

[1] Please check https://github.com/kqlqkqlqF/Toronto-Shelter-System-Flow-between-2018-and-2023-with-Further-Predictions.git for more information.

of Toronto's homeless population from 2018 to 2023, focusing on changes in status, such as the number of homeless individuals transitioning between housing and shelters, and long-term use of the shelter system, as well as demographic characteristics including gender, age, and identity.

A series of structured analyses were conducted. We begin by examining the annual counts of transitions of homeless individuals between housing, shelters, and other statuses within the Toronto shelter system from 2018 to 2023. Following this, we analyzed the composition of homeless population groups and the trend of age distribution within the homeless population. Similarly, the trend of gender composition across different population groups in the homeless community was also analyzed. Finally, we present the predictions from our linear model for future trends in the number of homeless individuals in Toronto.

Our analysis reveals several key findings. Firstly, the linear model we built suggests a steady increase in the number of homeless individuals in Toronto, indicating the need for expansion of the shelter system to accommodate more people. Additionally, our examination of the data through various figures provides insights into the nuanced dynamics of homelessness in the city. These include trends in the annual counts of homeless individuals transitioning within the Toronto shelter system, trends in the composition of homeless population groups and age distribution, and trends in gender composition across different population groups within the shelter system.

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". The dataset contains lots of information on the transition of homeless individuals in Toronto shelters from January 2018 to January 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. Moreover, the dataset has received full marks for freshness, metadata, accessibility, completeness, and usability on the Opendatatoronto website. Therefore, we consider the content of this dataset to be highly credible and utilize it as the primary data source for this paper. However, due to the limited data available for January

2024, we did not include this portion in the data analysis, as the smaller dataset size for 2024 may lead to unexpected analysis results.

2.2 Features

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.

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.

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 final linear model utilized the cleaned dataset for the second figure, and additional cleaning was performed. It is important to note that we excluded data from January 2024 in all chart data to avoid errors.

3 Model

3.1 Model Setup

3.2 Model Justification

4 Result

4.1 Trends in Homelessness Transitioning within the Toronto Shelter System

4.2 Composition of Homeless Population Groups and Correcponding Age Distribution

4.3 Gender Composition Across Homeless Population Groups

4.4 Result of the Model

4.4.1 Model Coefficients

Intercept (β_0) : The intercept represents the ..

Slope (β_1) : The slope quantifies the average change in the number of …

Explanation: The analysis showed a positive correlation between ..

4.4.2 Model Equation

4.4.3 Residuals

1. Minimum Residual: The minimum residual is -1826.5, indicating that there is at least one observation for which the model's predicted total population was lower than the actual population by approximately 1826.5 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 is -655.6, implying that 25% of the observations have total populations that are less than 655.6 individuals below the model's prediction.

3. Median Residual: The median residual is 226.4, indicating that half of the observations have residuals below 226.4, 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) is 712.4, signifying that 75% of the observations have total populations within 712.4 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 is 1185.1, suggesting that there is at least one observation for which the model's predicted total population was higher than the actual population by approximately 1185.1 individuals. Similar to the minimum residual, this could indicate the presence of outliers or unaccounted factors.

Interpretation: The distribution of residuals provides insights into the model's 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. Additionally, assessing the model's assumptions, such as linearity, homoscedasticity, and normality of residuals, is essential to ensure the validity of the statistical inferences drawn from the model.

4.4.4 Graphing the Model

5 Discussion

5.1 Summery of Findings

5.2 Limitations

5.3 Future Study

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.