Home Sweet Home*

Analyzing the State of Homeless Shelters in Toronto

Ritvik Puri

27 April 2022

Abstract

This paper presents a report on the conditions of various homeless shelters across the city of Toronto. Various factors such as location, secto, program type are taken into account when determing the occupancy rate for the beds and rooms provided by the City of Toronto. Post-pandemic we have seen an influx of people coming into the shelter programs.

Contents

*Code and data are available at: https://github.com/ritvikpuri/toronto-shelters

1 Introduction

As you sit comfortably in your home and look at your screen reading this, it is often easy to miss that homelessness is a big problem in almost every country around the world. Despite being one of the most developed countries in the world, Canada has around quarter of a million people living without a roof on their head. As you may expect, these rates are even higher for a busy metropolitan city such as Toronto. For every 10,000 people, 30 are homeless. While this may not sound like a lot at first glance, there are over 2 million people living in the city. This brings up multiple questions. Who are these people? How do they live? And perhaps most importantly, what is the government doing for them?

According to an article published by the CBC, the shelter system of Toronto is unable to cope up with the sudden influx of homeless people caused by the global COVID-19 pandemic. There have been more than 100 deaths caused due to people lacking proper housing to live in, and the pandemic only made their situation more dire. The people were previously unhappy with the efforts to help the people living on the streets and the rise in COVID cases resulted in the country and the city to focus their efforts on other matters that needed urgent attention. Now that the pandemic has died down after the last wave, the shelter system has once again caught the attention of the city and is in the process of being made more efficient.

We shall be examining the current dataset using R(R Core Team (2020)) and various packages such as ggplot(Wickham (2016)), tidyverse(Wickham et al. (2019)), opendatatoronto(Gelfand (2020)), patchwork(Pedersen (2020)), dplyr(Wickham et al. (2021)) and DiagrammeR(Iannone (2022)). R-Markdown (Allaire et al. (2022)) is used for compilation and presentation of the final paper.

2 Data

2.1 Dataset and Variables

The dataset about Daily Shelter & Overnight Service Occupancy & Capacity is available for use on Open Data Toronto portal.(SSHA (2022)) This dataset is generated using the Shelter Management System, also known as SMIS, which is an advanced, computerized system developed for the city of Toronto. This system is sued to collect Shelter System Flow data that is used to operate shelters and other services throughout Toronto that are funded by the city itself. (Alliance (2021)) Starting from 2021, the data is updated daily and provides information about shelters and overnight service programs.

Various columns are part of the dataset, ranging from the occupancy date, organization name, location, program name, etc. However we will be focusing more on the capacities of these programs. The programs have either room based or bed based capacities. Bed capacity applies to common sleeping areas which are shared by multiple people, each occupying their own bed spot, whereas room capacity applies to individual bedrooms given to people.

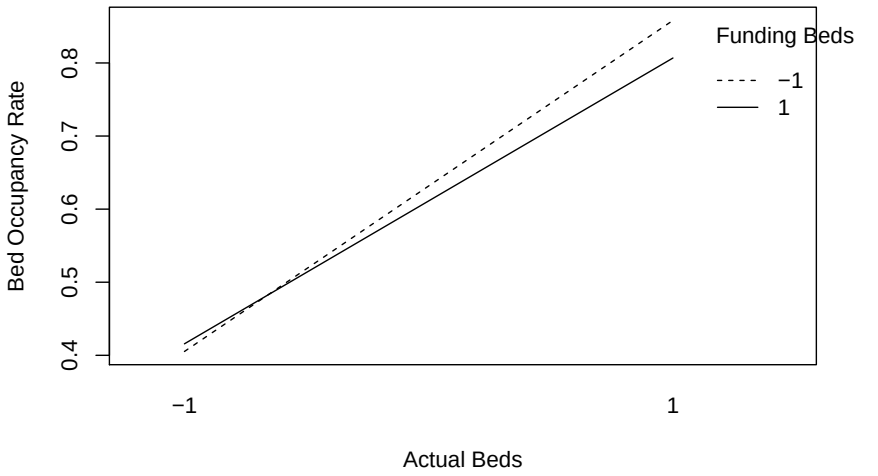
These capacities have two applicable methods of being measured. The actual capacity is the number of rooms or beds in service and are avail be for occupancy. Funding capacity is the number of rooms or beds that the program intends to provide. While both these measures are important, actual capacity has proven to be more accurate and effective method of keeping track of room capacities. These capacities can also be interpreted via the Occupied Beds and Occupied Rooms columns.

Along with capacities, other areas of interest are Sector, Program Area and Overnight Service Type. The Sectors constitute of genders based sorting, such as programs for women or families only. The Program Areas and Service Types have categories based specifically on the type of services that will suit the City of Toronto the best.

2.2 Visualization

We are going to first take a loot at a very crucial interaction between our variables. We are focused on the Actual Beds, Funding Beds, Bed Occupancy Rate, Actual Rooms, Funding Rooms and Room Occupancy Rate variables. We will consider the four capacity variables as our predictors and the 2 occupancy rates for rooms and beds as our outcomes.

Calculating the mean for each of these variable enables us to classify each column as a new variables as either +1 or -1, if they are greater than or less than their respective column averages. This gives us 6 new variables to plot and look at their interactions and tell if they effect each other by looking at if the lines are parallel or not. From Figure 1, we can see that the lines are in fact not parallel and this relationship can be expressed as a Directed Acyclic Graph. This relationship also shows us that the bigger the shelter programs are, the more they get utilized. This is easy to infer as larger shelters would be more popular and more people, homeless or otherwise, would get to know about it.



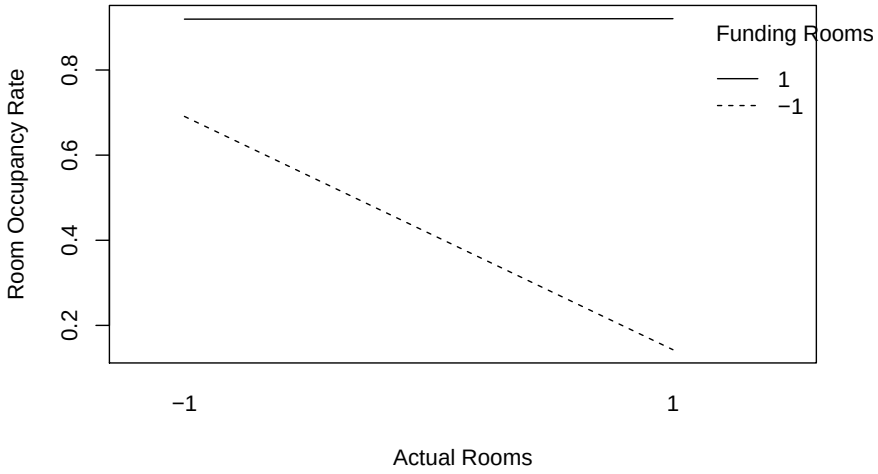


Figure 1: Interaction plots for each combination of predictors on outcomes

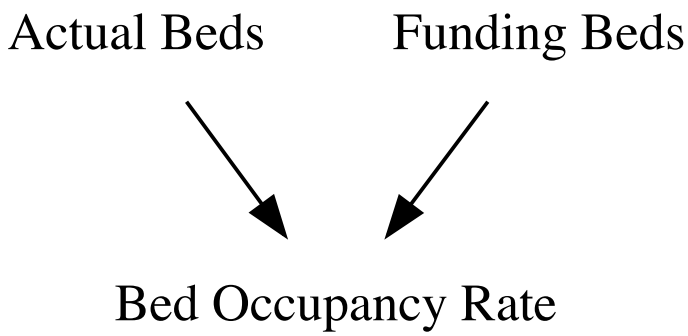


Figure 2: A directed acyclic graph showing the interaction between the 3 variables

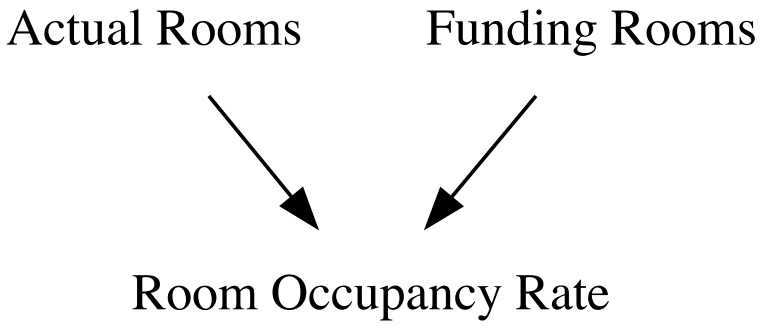


Figure 3: A directed acyclic graph showing the interaction between the 3 variables

3 Model

To find out more about the Occupancy Rates for Beds and Rooms in the programs, we are going to see how they are distributed in each sector. The sectors are as follows: Families, Men, Mixed Adult, Women, Youth and signify what are genders the specific shelter of a program caters towards. These graphs take density of the occupancy rates into account rather than the count as it gives a better measure of the values. As we can infer from Figure 4, all four models show that most of the shelters are functioning at near maximum capacity and in some cases are overflowing.

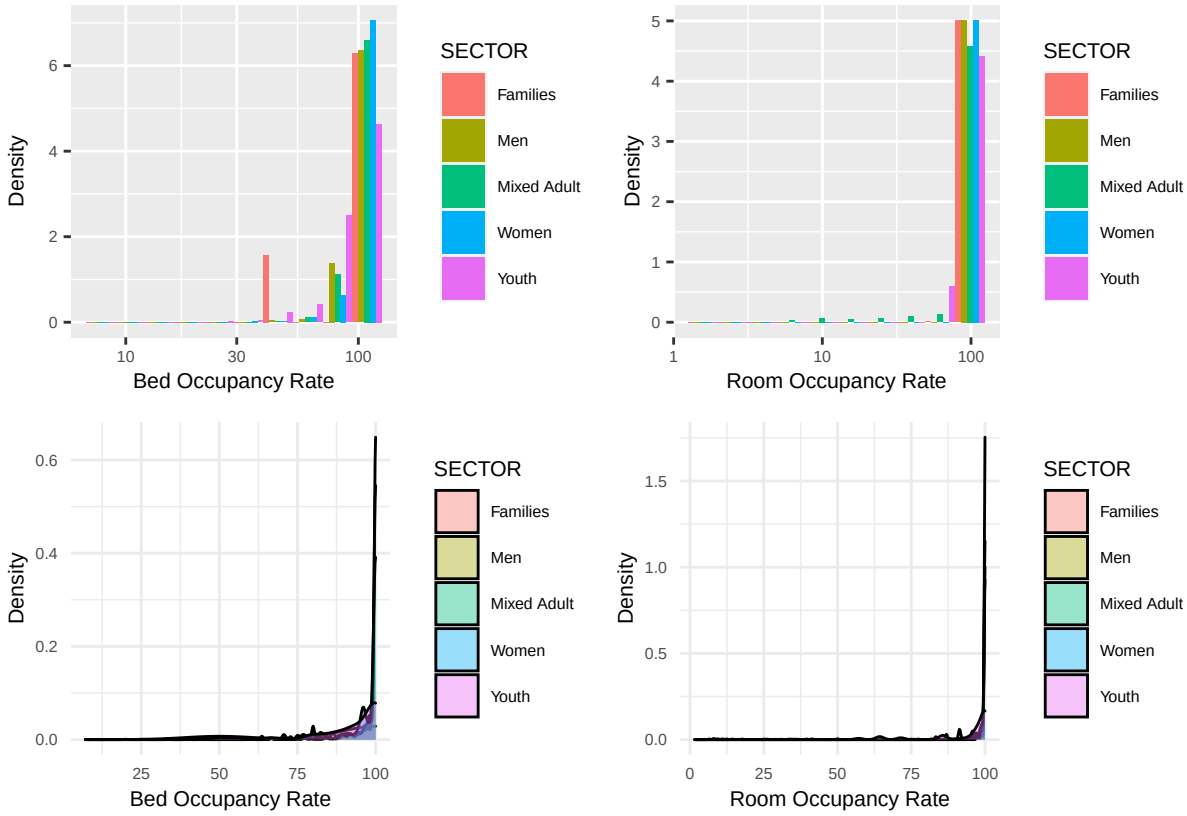


Figure 4: Histogram and Density Models to show the Bed and Room Occupancy Rates

4 Results

Over the last few years, various sources have reported that the Toronto Shelters are overcrowded and overflowing with more people that they can safely accommodate. According to the United Nations Emergency Handbook camp site planning guidelines, there should be around 5 sqm available per person for appropriate living and the minimum per person is 3.5 spm.(UNHCR (n.d.)) Our data suggests that since these shelters are running at near maximum capacity, these universal guidelines might not be met for the people living at these places. Press Progress conducted an investigation in 2019 and found out that four people were living in the area that should only be given to one person.(Pressprogress (2019))

Figure 6 plots the Capacity of Actual Beds Available with the Occupied Beds and plots Capacity of Actual Rooms Available with the Occupied Rooms. While most of the programs have less than 100 rooms and beds available, it should be noted that they are very close to being completely filled and in some cases even being overcrowded and have more people living in them that they can properly support.

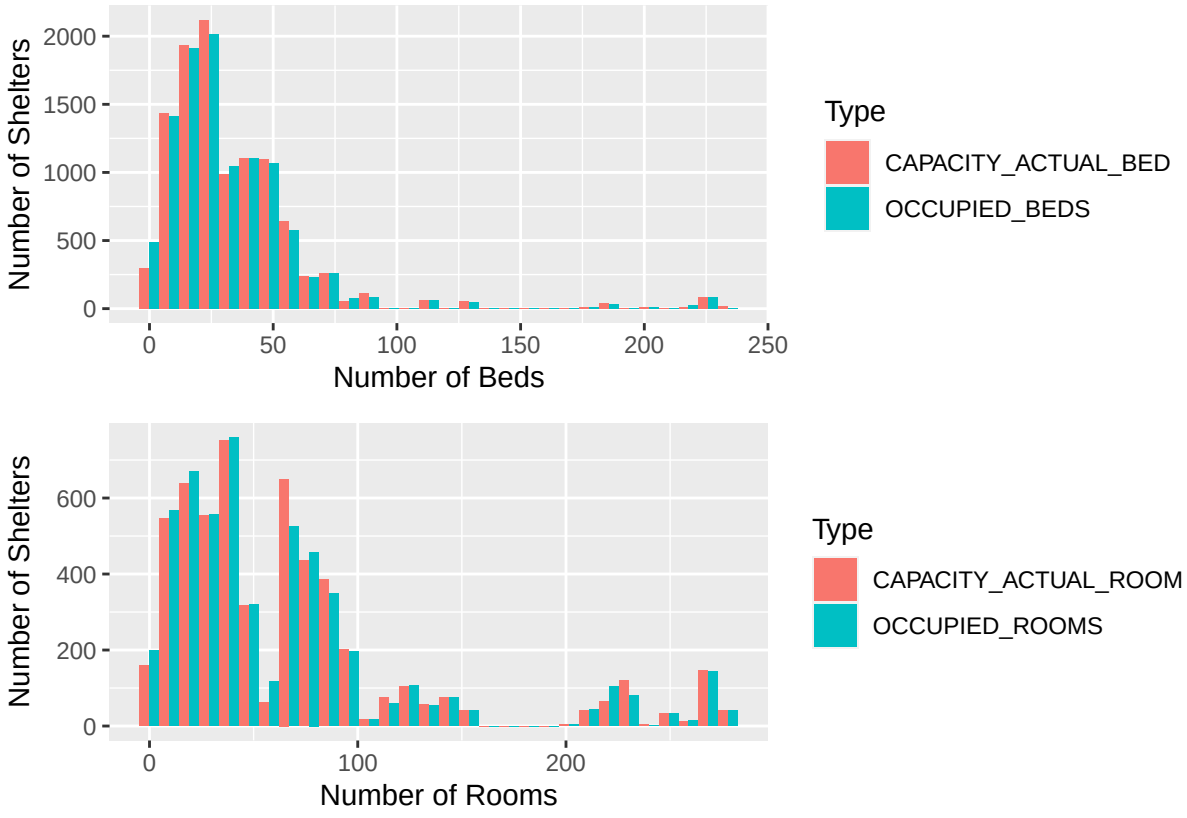


Figure 5: Histogram showing the current amount of Beds and Rooms Occcuiped and Available

5 Discussion

As we see how shelters of various Sectors are overflowing, one of these sectors is more important in my opinion. This would be the homeless families that need housing throughout the city of Toronto. When compared to other sectors, such as Men, as we can see in Figure 6, the number of shelters with beds for families is extremely limited. This may pose a problem for the city as it would be easier to live in a shelter as a single adult man or woman. Shelters have upto a max of 3 beds that can sustain a family, and all such shelters are running at maximum capacity. The shelters which have only 1 bed reserved for families are in fact housing much more families that they can safely keep.

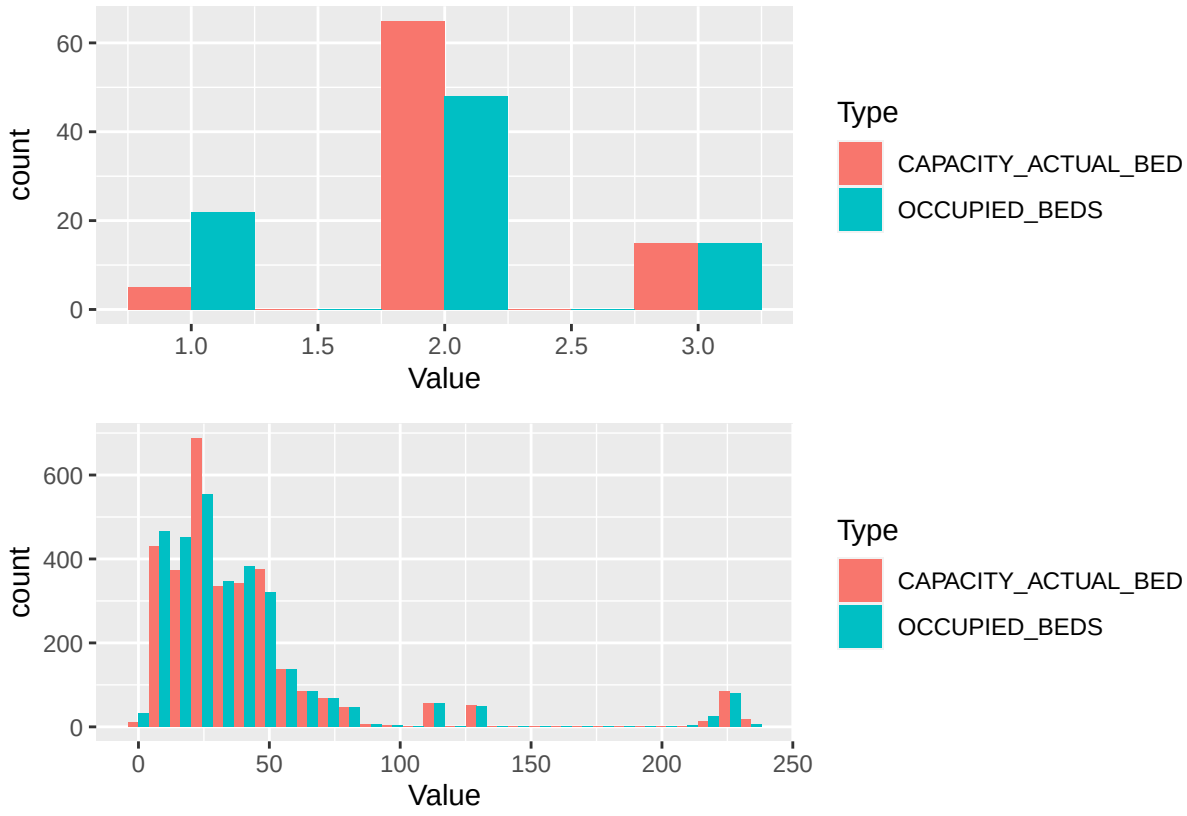


Figure 6: Histograms showing the difference between the Men Sector and Family Sector

Such shelter programs are often categorized if they have been specially set up for specific issue at hand. While most of the shelters programs in the City of Toronto are Basic Shelter and Overnight ones, over the last few years a lot of COVID-19 response have been set up. This is still an adapting program, and aims to deliver support the people affected by unprecedented situations that may have risen due to the global pandemic.(Toronto (2020)) These shelters are currently open for Families, Men and Mixed Adults and from the Figure 7 we can see the spikes due to people having to take shelter as a result of rapidly changing situations. These shelters have been extended to be kept running until April 2023 as of the most recent announcement by the Shelter, Support and Housing Administration (SSHA). There are also a lot of Winter Programs that are only set up during the winter months to provide an alternative to people sleeping on the cold streets of Toronto.

It is evident that these shelters are currently being completely utilized by the numerous homeless people across the entire city. The City Council needs to set up more such camps and programs in order to provide relief for homeless people According to a recent report, after the COVID-19 camps have been dissolved

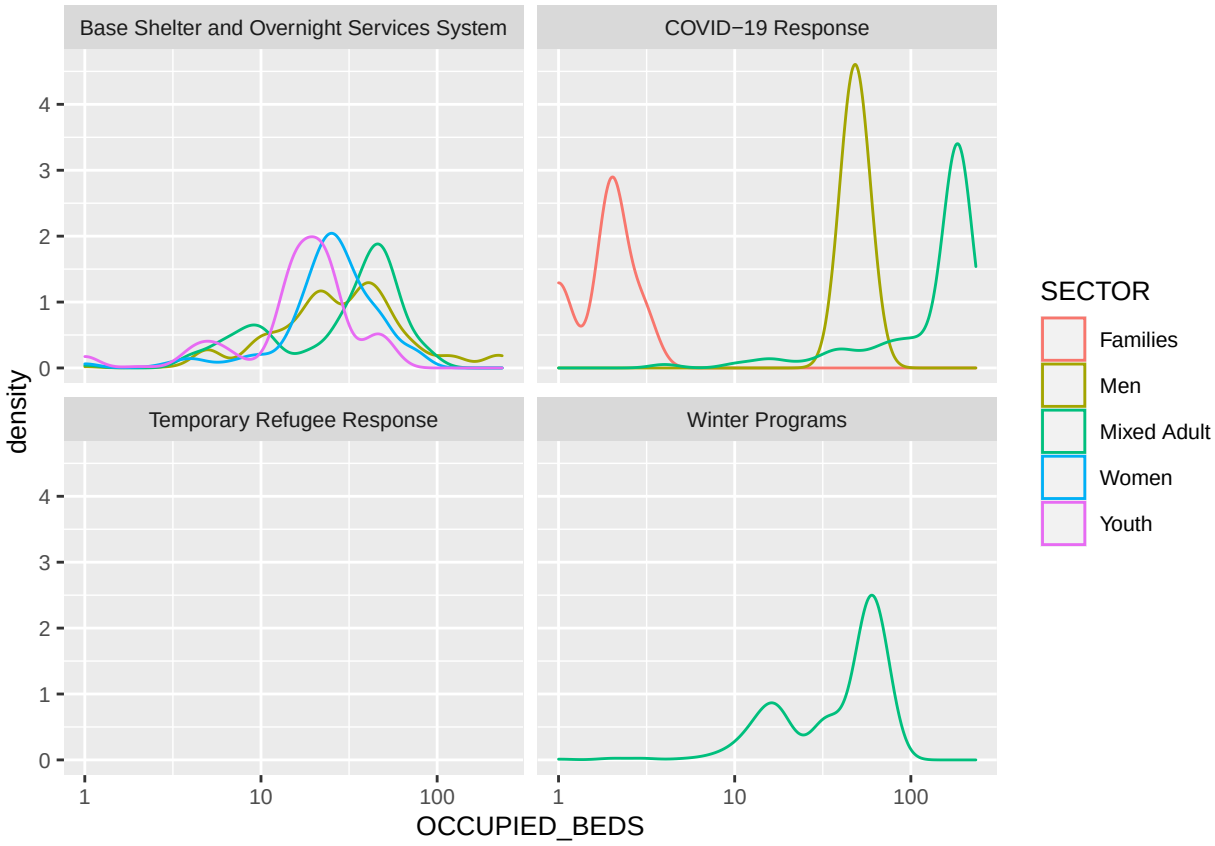


Figure 7: Occupied Beds depicted in a Density chart grouped by their Program Area

as the pandemic subdues, the City is planning to extending the lease of all others. 27 facilities that are completely funded by city will be extended upto the end of 2023 and a select few would even be extended till 2024.(DeClerq (2022))

Along with these extensions, the City Council should also try to increase the affordability of housing in the cities as well as increase the number of jobs that homeless people can be applicable for. Even low paying government jobs should be able to alleviate their current situation and help them secure their own home for the first time in their life.

Appendix

A Datasheet

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
    - The dataset was created to help make the Shelter System in the City of Toronto more effective.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
    - The dataset was created by Shelter, Support and Housing Administration (SSHA)
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
    - The creation was funded by the City of Toronto
4. *Any other comments?*
    - This data set provides a daily list of active overnight shelter and allied services in the Shelter Support and Housing Administration division's Shelter Management Information System (SMIS) database and hosted on Open Data Toronto Website.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
    - The instances represent the various properties of each shelter program such as Location, Sector, Capacity, etc.
2. *How many instances are there in total (of each type, if appropriate)?*
    - There are 29 instances
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
    - The dataset contains all possible instances as it is updated everyday.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
    - There are 27 discrete variables and 2 continous variables.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
    - No such labels.
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
    - There are mulitple instances where an `NA` value exists. Such data is omitted in modelling.

7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
    - There are no relationships between individual instances.
8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
    - There are no recommended data splits.
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
    - There are no errors, sources of noise, or redundancies in the dataset.
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*
    - The dataset is self-contained.
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
    - There is no confidential data, and the dataset is publicly available.
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
    - The dataset talks about Homelessness in the city of Toronto which may be a sensitive topic for some people.
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
    - The dataset containes people from all genders.
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
    - It is not possible to identify individuals in any way.
15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
    - No specific columns can be identified as sensitive.
16. *Any other comments?*
    - None.

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

- The data associated is collected via the SMIS database and is updated daiy.

2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
   - The data is collected manually through the numbers of each shelter daily.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
   - No such sampling was conducted.
4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
   - The data collection process is done by people working for the SSHA.
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
   - The data is collected daily.
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
   - Ethical review processes were not conducted.
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
   - I obtained the data from Open Data Toronto website: open.toronto.ca/dataset
8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*
   - The individuals voluntarily interviewed with data collectors. The notice of data collection is not available.
9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
   - The individuals consented to the collection and use of their data. The exact language to which consent was granted is not available.
10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
    - A mechanism to revkoe consent was not provided.
11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*
    - An analysis of the potential impact of the dataset and its use on data subjects was not conducted.
12. *Any other comments?*
    - None.

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*
    - The data was originally obtained in CSV format and no cleaning process was required.
2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*
    - The raw data can directly be downloaded using opendatatoronto package for R.
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
    - R Software is avalaible at https://www.R-project.org/
4. *Any other comments?*
    - None

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
    - The dataset has not been used for other tasks yet.
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
    - https://github.com/ritvikpuri/toronto-shelters
3. *What (other) tasks could the dataset be used for?*
    - The dataset can be used for examining the state of Shelter Systems in the City of Toronto.
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
    - Since the data is already well formatted by the uploading party, no special composition needs to be accounted for.
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
    - The dataset might not be appropriate for tasks that do not revolve around Toronto Shelter Programs.

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
    - No, the dataset is openly available and being used for personal uses only.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*

- The dataset will be distributed using Github.

3. *When will the dataset be distributed?*
   - The dataset will be distributed in April 2022.
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
   - The dataset will be released under the MIT license
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
   - There are no restrictions
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
   - No such controls or restrictions are applicable.
7. *Any other comments?*

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
   - Shelter, Support & Housing Administration
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*
   - Can be contacted via email at: sshadata@toronto.ca
3. *Is there an erratum? If so, please provide a link or other access point.*
   - There is no erratum available currently.
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
   - The dataset is updated daily by the SSHA
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
   - There are no such applicable limits.
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
   - The older versions would not be hosted. Dataset consumers will be able to check whether the dataset has been updated through Open Data Toronto website.
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
   - There is no mechanism for accepting contribtuions from other users as of now.

B Additional details

Code and data are available at: https://github.com/ritvikpuri/toronto-shelters

References

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2022. *Rmarkdown: Dynamic Documents for R*. https://github.com/rstudio/rmarkdown.

Alliance, Toronto. 2021. "Toronto Alliance to End Homelessness." Government. *Toronto Alliance to End Homelessness*. Toronto Alliance. https://taeh.ca/frequently-asked-questions.

DeClerq, Katherine. 2022. "Toronto to Decommission 5 Temporary Covid- Shelters This Year, Extends Leases of All Others." CTV News. https://toronto.ctvnews.ca/toronto-to-decommission-5-temporary-covid-shelters-this-year-extends-leases-of-all-others-1.5851635.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Iannone, Richard. 2022. *DiagrammeR: Graph/Network Visualization*. https://CRAN.R-project.org/package=DiagrammeR.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

Pressprogress. 2019. "Toronto's Shelters Are Now Consistently Above 90." *PressProgress*, January. https://pressprogress.ca/torontos-shelters-are-now-consistently-above-90-capacity-and-that-is-extremely-dangerous/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

SSHA. 2022. "Daily Shelter & Overnight Service Occupancy & Capacity 2022." Government. *Open Data Toronto*. SSHA. https://open.toronto.ca/dataset/daily-shelter-overnight-service-occupancy-capacity/.

Toronto, City of. 2020. "Temporary Covid-19 Shelter Sites." Government. *City of Toronto*. https://www.toronto.ca/community-people/community-partners/emergency-shelter-operators/about-torontos-shelter-system/new-shelter-locations/temporary-covid-19-shelter-sites/.

UNHCR. n.d. "Site Planning for Camps." Organization. *UNHCR Emergency Handbook*. UNHCR. https://emergency.unhcr.org/entry/35943/site-planning-for-camps.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

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, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.