An Analysis of daily shelter and overnight service usage in Toronto focusing on shelter occupancy and capacity*

Angel Xu

September 25, 2024

In this report, the overnight service occupancy and capacity of Toronto is analyzed.

Table of contents

1 Introduction

Briefly introduce the analyzing reason and background information. From what perspective and how to analyze Key results Structure introduce

*Code and data are available at: https://github.com/Anjojoo/Toronto-Overnight-Servive-Occupancy-Capacity.

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Alexander (2023).

The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Data Source and Overview

This paper analyzes data from the "Daily Shelter & Overnight Service Occupancy & Capacity" dataset available on the OpenDataToronto portal. The dataset contains a daily record of overnight shelters and service programs collected through Toronto's Shelter Management Information System (SMIS). Data spans from 2021 to 2024, offering current insights, and is updated daily, except for weekends and holidays, with uploads resuming on the next business day.

A similar dataset, "Daily Shelter Occupancy," is also available on OpenDataToronto but only provides data up to March 10, 2022, and includes fewer variables. In contrast, "Daily Shelter & Overnight Service Occupancy & Capacity" features the most recent data, refreshed on 25 September, 2024. It builds upon and enhances the earlier dataset by incorporating additional key variables, including "overnight service type" and "actual capacity." The "overnight service type" variable categorizes the types of overnight services offered in Toronto's shelters, helping analyze variations in user counts by service type. The "actual capacity" variable represents the true available capacity of shelters, as opposed to the funding-based capacity in the previous dataset, offering a more accurate measure of occupancy rates for further analysis.

2.2 Data Tools

The data was simulated, extracted, and cleaned by R(R Core Team 2023) from the opendatatoronto package(Gelfand 2022). Tidyverse(Wickham et al. 2019), here(Müller 2020), ggplot2(Wickham 2016), dplyr(Wickham et al. 2023), lubridate(**lubridate?**), janitor(Firke 2023) and knitr(**knitr?**) are all applied in the analysis.

2.3 Clean Data

Prior to analysis, data cleaning is essential to remove unnecessary columns and create new ones for further examination. The occupancy date variable is converted to a date format, allowing the extraction of the year for subsequent analysis. After this conversion, a new 'year' column is mutated to show trends in service user counts and occupancy rates from 2021 to 2024. Once these steps are complete, relevant variables are selected from the raw dataset to simplify and clarify the analysis. The chosen variables include year, occupancy date, overnight service type,

service user count, capacity type, actual capacity and occupancy rates. NAs exist in actual capacity and occupancy rates, but there are no needs to eliminate them since due to different capacity type of bed and room, NA appears in the opposite actual capacity and occupancy rates.

2.4 Summary Statistics of Variables

To gain a better overview of the dataset, a table including key variables is provided. Table 1(Table 1) presents year, overnight service type, service user count and occupancy rates. Year represents the year where the observation is collected. Overnight service type represents what service has been provided. Service user count records number of users taking service during that night. Occupancy rates record the proportion of actual capacity that is occupied.

Table 1: Table 1: Sample of the key variables in Occupancy & Capacity Data

year	overnight service type	service user count	occupancy rate beds	occupancy rate rooms
2024	Motel/Hotel Shelter	650	NA	100
2024	Motel/Hotel Shelter	113	NA	100
2024	Shelter	8	100	NA
2024	Motel/Hotel Shelter	203	NA	100
2024	Motel/Hotel Shelter	508	NA	100
2024	Motel/Hotel Shelter	152	NA	100

Beside the sample table of ley variables, a summary statistics table of key variables can be provided to clarify some information of the data. Table 2(Table 2) presents the mean and standard deviation of service user counts and two occupancy rates. Service user counts has a mean at 73.73 and a standard deviation at about 91.68% which is pretty high and shows the great difference in user picks between different services. Occupancy rate of beds has a high mean at 98.24% and a standard deviation at 6.08%. This illustrates that the occupency rate of beds are approximately near 100 and are full occupied for most of the situations. Occupancy rate of rooms also has a high mean at 99.53% which is higher than occupancy rate of beds and a even lower standard deviation of 1.77%, suggesting that capacity of shelter rooms reaches nearly full and fuller than shelter rooms. However, the actual capacity of beds and rooms aren't included in this table; thus, the usage of shelter beds and rooms need to be considered more thoroughly by observing their actual capacity numbers.

Table 2: Table 2: Mean and Standard Deviation of Key Variables

Variable	Mean	Standard Deviation
Service User Count	73.74303	91.676591
Occupancy Rate of Beds	98.24426	6.071307
Occupancy Rate of Rooms	99.52812	1.773672

3 Result

Some of our data is of penguins (**?@fig-bills**), from (**palmerpenguins?**).

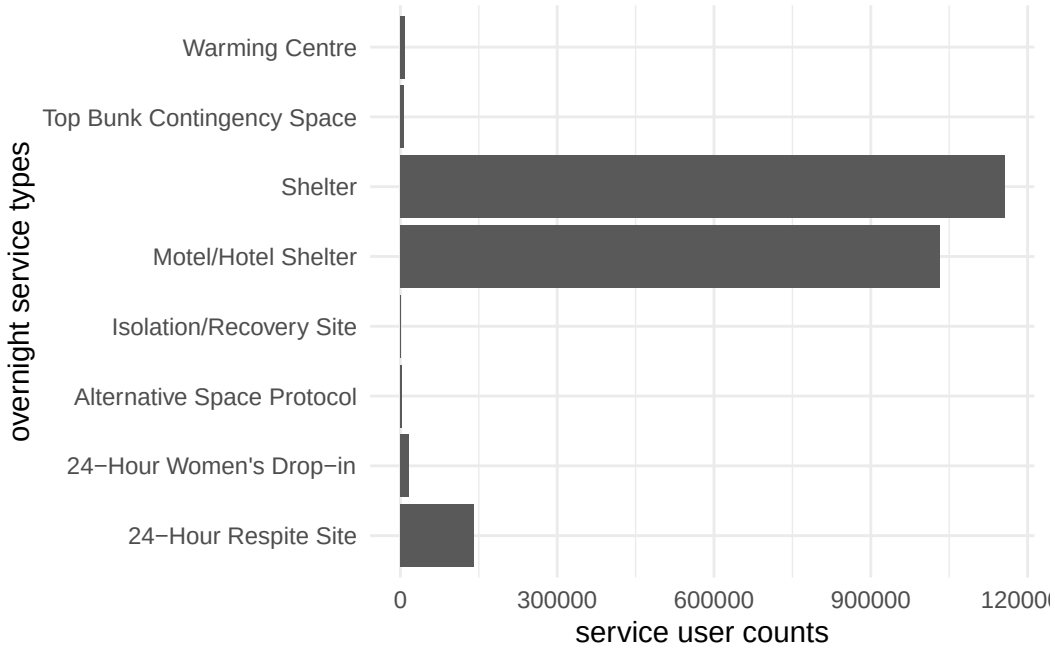

Figure 1: Total service user counts for different overnight service types

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

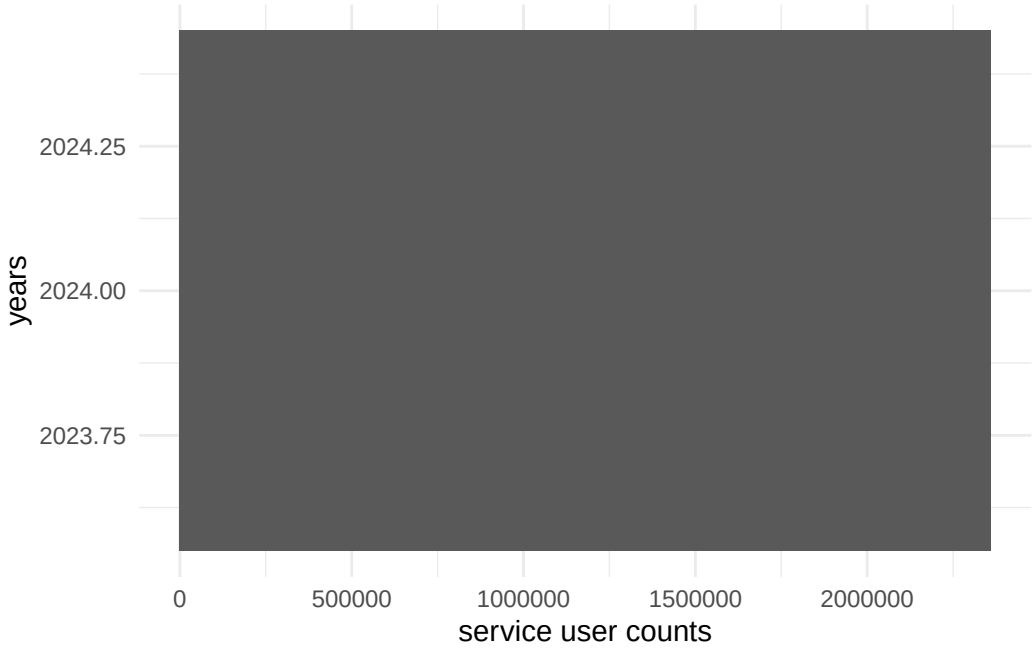

Figure 2: Service user count change from 2021 to 2024

References

Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r*. Chapman; Hall/CRC.

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

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

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

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.