INF 2178 Technical Assignment 1

Jinhang Luo (1005124777)

Faculty of Information
University of Toronto
INF2178: Experimental Design for Data Science
Prof.Shion Guha

Feb 4th, 2024

Introduction

1.1 Background

The dataset, sourced from the Shelter Support and Housing Administration and recorded in 2021, provides a comprehensive analysis of overnight shelter services through the Shelter Management Information System (SMIS) lens. This dataset stands out for its classification of shelters into bed-based and room-based capacities, presenting dual capacity metrics - intended (funding) capacity and the actual in-service capacity. It's instrumental in delving into the complexities and operational patterns of Emergency and Transitional shelters. It offers valuable insights into occupancy patterns and user profiles, crucial for developing effective housing policies and addressing homelessness challenges.

1.2 Dataset description

This dataset, a revised and updated version of the Daily Shelter Occupancy data set, was released, starting with the current year's data for 2021. And it contains 50,944 records in total. The dataset can be found at the following link:

https://open.toronto.ca/dataset/daily-shelter-overnight-service-occupancy-capacity/ (Toronto Open Data, 2021). In the dataset, numerous attributes are present. However, our main focus is exploring the connections between variables such as the program model, number of service users, actual bed capacity, number of occupied beds, actual room capacity, and the number of occupied rooms.

1.3 Research question

In analyzing the dataset, we aim to explore the operational distinctions in how these shelters use their available space. Moreover, we seek to understand the differences in demographic engagement between Emergency and Transitional shelters. Consequently, two research questions have been formed.

1. How do Emergency and Transitional shelter programs differ regarding bed and room occupancy rates?

2. What are the variations in the number of service users between these program models?

Data Cleaning

We tidied up the dataset by removing unnecessary columns and filling missing values with zeros, focusing on the critical areas of bed and room capacities. Then, we calculated bed and room occupancy rates for our main analysis.

Exploratory data analysis (Method & Result)

3.1 Descriptive statistics

Firstly, we performed a descriptive analysis that included the dataset's sample size, mean, and standard deviation for certain attributes.

As shown in **Table 1**, the sample size of the data is 50,944, except for Bed occupancy rates, which are 32,399, and Room occupancy rates, which are 18,545. This variation is likely a result of different data recording methods or criteria used for bed and room information, which could vary based on the specific type of shelter program (e.g., those for families and individuals). Moreover, the variability in means and standard deviations among various variables suggests that the dataset is robust enough to provide meaningful statistical conclusions despite the differences in sample sizes.

Table 1. Descriptive Statistics

Variables	N	Mean	SD
Number of service users	50,944	45.7272	53.3260
Actual bed capacity	50,944	20.1140	26.4498
Number of occupied beds	50,944	18.9394	25.4534
Actual room capacity	50,944	20.2214	44.7312
Number of occupied rooms	50,944	19.2201	43.6309
Bed occupancy rates	32,399	0.9279	0.1226
Room occupancy rates	18,545	0.9341	0.1632

Source: Daily Shelter & Overnight Service Occupancy & Capacity Dataset From Toronto Open Data.

Then, we plotted the boxplots to compare the occupancy rates for beds and rooms across two different shelter program models: Emergency and Transitional (See **Figure 1**).

The left box plot shows emergency shelters with higher median bed occupancy and a narrow interquartile range, indicating consistent usage but with some lower outliers. The right box

plot shows transitional shelters with higher median room occupancy and a broader range, indicating more stable use. Both plots suggest that occupancy is generally high in both types of shelters, with transitional shelters displaying more consistent room occupancy and emergency shelters showing more fluctuation.
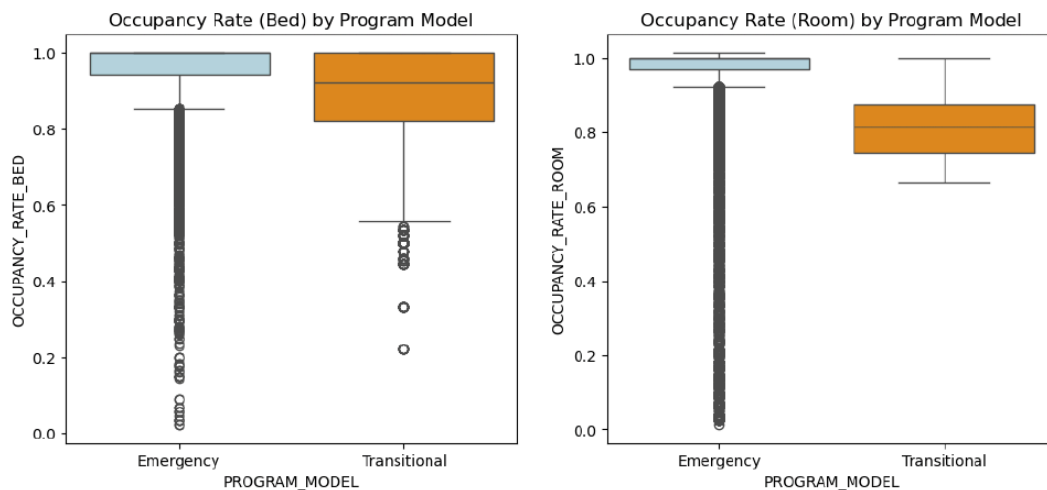

Figure 1: The Box-Plot of Occupancy Rates (Bed & Room) by Program Model

Furthermore, as shown in **Figure 2**, the bar plot indicates the average number of service users in the emergency and transitional programs model. The blue bar for emergency shelters indicates a higher average service user count compared to the orange bar for transitional shelters. This suggests that, on average, emergency shelters accommodate more individuals than transitional ones.
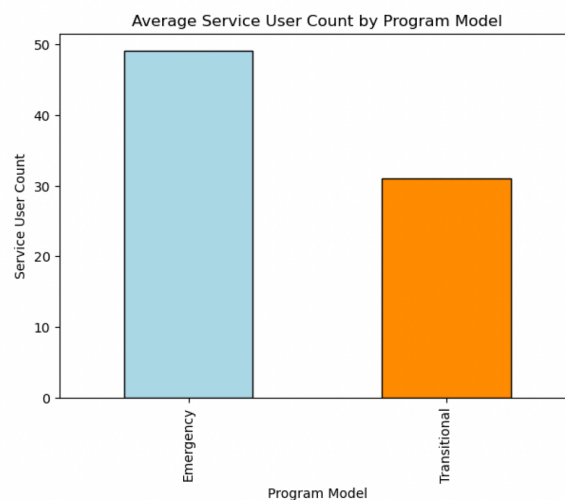

Figure 2: The Bar-Plot of Average Service User Count by Program Model

In the end, as shown in **Figure 3**, the histograms compare service user counts between shelter types. The emergency model shows a steep decline from a high frequency of shelters with fewer users to a few shelters with many users. The transitional model displays a narrower distribution, indicating a smaller, more consistent number of users. This suggests emergency shelters occasionally reach high occupancy, whereas transitional shelters typically maintain a steadier, lower occupancy.
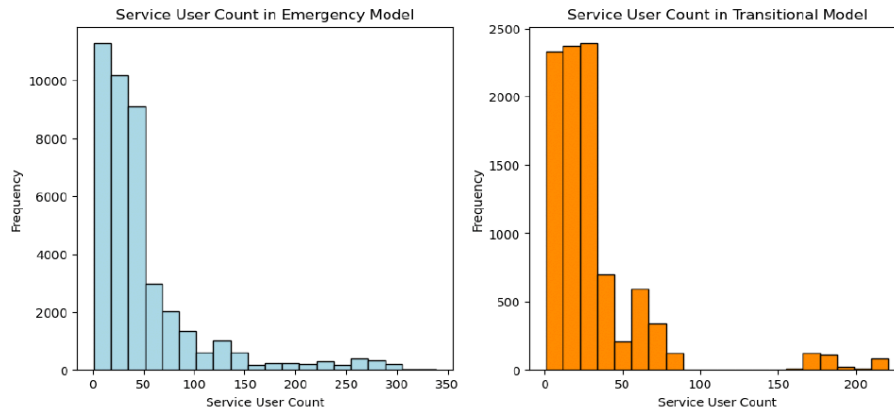

Figure 3: The Histogram of Service User Count by Program Model

3.2 T-Test

The t-test serves as a valuable tool for verifying the presence of a significant distinction between two groups and for exploring the connection within the data, ultimately leading to a conclusion. When the p-value falls below 0.05, it means a statistically significant difference between the two groups.

Program Model and Occupancy Rate (Bed)

We conducted a t-test to analyze whether the occupancy rate (bed) differed by program model. The hypotheses are the following:

H_0(Null Hypothesis) $= \mu_0$ (The means of the two groups are equal).

H_A(Alternative Hypothesis) $\neq \mu_0$ (The means of the two groups are not equal)

The results we obtained show that the p-value is about 0. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can be rejected.

4

Program Model and Occupancy Rate (Room)

We conducted a t-test to analyze whether the occupancy rate (room) differed by program model. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups are equal).

H_A(Alternative Hypothesis) ≠ μ_0 (The means of the two groups are not equal)

The results we obtained show that the p-value is about 5.923e-79. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can be rejected.

Program Model and Service User Count

Like the above test, we conducted a t-test to analyze whether the service user count differed by program model. The hypotheses are the following:

H_0(Null Hypothesis) = μ_0 (The means of the two groups are equal).

H_A(Alternative Hypothesis) ≠ μ_0 (The means of the two groups are not equal)

The results we obtained show that the p-value is about 3.172e-195. And we set the alpha at 0.05. Since the p-value is less than 0.05, there is a statistically significant difference. Thus, the null hypothesis can be rejected.

Discussion

Our analysis of the dataset shows clear trends in how shelters are used. Emergency shelters have bed occupancy rates that change a lot because they need to respond quickly to immediate needs, which might go up and down with the seasons or sudden events. Transitional shelters are different; their room occupancy rates are mostly the same, likely because they're meant for people to stay longer as they move toward more permanent living situations. When we looked at the number of people using these services, we found that emergency shelters sometimes get very full. However, usually, they're helping fewer people

at a time. Transitional shelters have a steady flow of users, which fits their role in giving ongoing help.

Limitation

The findings from this dataset may be limited to other contexts or populations, given that they are based on a specific set of shelters within a certain timeframe.

Conclusion

Our investigation suggests that emergency and transitional shelters play their part in the broader system, catering to the different needs of those they serve. Emergency shelters need to be able to scale up or down quickly because of how demand can swing. In contrast, transitional shelters need a steady flow of support to keep offering consistent help. For those making policies and running these services, our study points out that they have to think differently about managing resources and designing programs for each type of shelter. Further research could investigate the underlying causes of occupancy variability and explore the impact of specific interventions on service user outcomes.

Rerefence List

Shelter, Support & Housing Administration. (2021). Toronto Open Data.
https://open.toronto.ca/dataset/daily-shelter-overnight-service-occupancy-capacity/