INF 2178 – Winter 2024
Technical Assignment 1 – Narrative of findings
Samantha Yin (1004194258)
Feb 4, 2024

Introduction

The homeless population in Toronto has seen a steady rise over the years. Fortunately, the City of Toronto's shelter support system can provide aid to unhoused individuals, acting as a protection in difficult times. However, with the homeless population in Toronto on the rise, this poses significant challenges for the city's resources and policy planning. The dataset we used tracks the daily occupancy and capacity at the Toronto's shelters in the year of 2021, along with supplementary data on various shelter characteristics. This analysis aims to analyze the city's shelter occupancy trends and uncover patterns that may inform more effective responses to the issue of homelessness. Key research questions include: How do occupancy rates vary across different shelter program models and throughout the year? Which types of shelters or programs are more prone to overcapacity, indicating a need for government intervention? Utilizing this data from 2021, the goal is to provide a thorough analysis of the city's shelter occupancy and identify patterns associated with shelters and homelessness.

Data

The dataset provided contains 14 fields and totals 50,944 entries. To facilitate a better understanding through data analysis tools, the data have been categorized into three main types: (1) one date-time field (OCCUPANCY_DATE), (2) 6 numerical fields, and (3) 7 categorical fields. They are displayed in the Table 1 below:

Table 1

Numerical Fields	Categorical Fields
PROGRAM_ID	ORGANIZATION_NAME
SERVICE_USER_COUNT	PROGRAM_NAME
CAPACITY_ACTUAL_BED	SECTOR
OCCUPIED_BEDS	PROGRAM_MODEL
CAPACITY_ACTUAL_ROOM	OVERNIGHT_SERVICE_TYPE
OCCUPIED_ROOMS	PROGRAM_AREA
	CAPACITY_TYPE

Results & Findings

Addition of a Fields

Upon further inspection of the CAPACITY_TYPE field, there are two categories of capacity type within the dataset: bed-based and room-based. Out of the 50,944 records, 32,399 are labeled as bed-based capacity, while 18,545 are designated as room-based capacity. For better analysis, occupancy rates for both beds and rooms have been calculated separately and then merged to give a full picture of shelter utilization. The occupancy rate for the bed-based category is determined by dividing the number of occupied beds by the actual beds available. Similarly, the occupancy rate for the room-based category is calculated by dividing the number of occupied rooms by the actual rooms available. The field is called: OCCUPANCY_RATE.

Numerical Fields

OCCUPANCY_DATE is the date of record, aligning with its data type as DateTime. The recorded dates span from January 1, 2021, to December 31, 2021. Both the mean and median dates fall towards the end of June, indicating an even distribution of data throughout the year. This can serve as a foundation for further investigation on the seasonality of homelessness.

The **Program ID** field serves as a unique identifier for each program, thus does not require further statistical analysis. **SERVICE_USER_COUNT** is the number of users at a shelter location at a specific time, with values ranging from 1 to 339. A standard deviation of 53 indicates variability in user counts among different programs. The Service User Count histogram (Figure 1, (1)) presents a right-skewed distribution, showing that the majority of records feature service user counts at lower count. This is consistent with the calculated mean and median service user counts of 46 and 28, respectively. This information provides insight into the shelter sizes: the largest shelter accommodates 339 users at one time, but most shelters serve around 40 users. Further investigation on the size of shelter will be needed if the government want to invest in more shelter, and deal with homelessness more effectively.

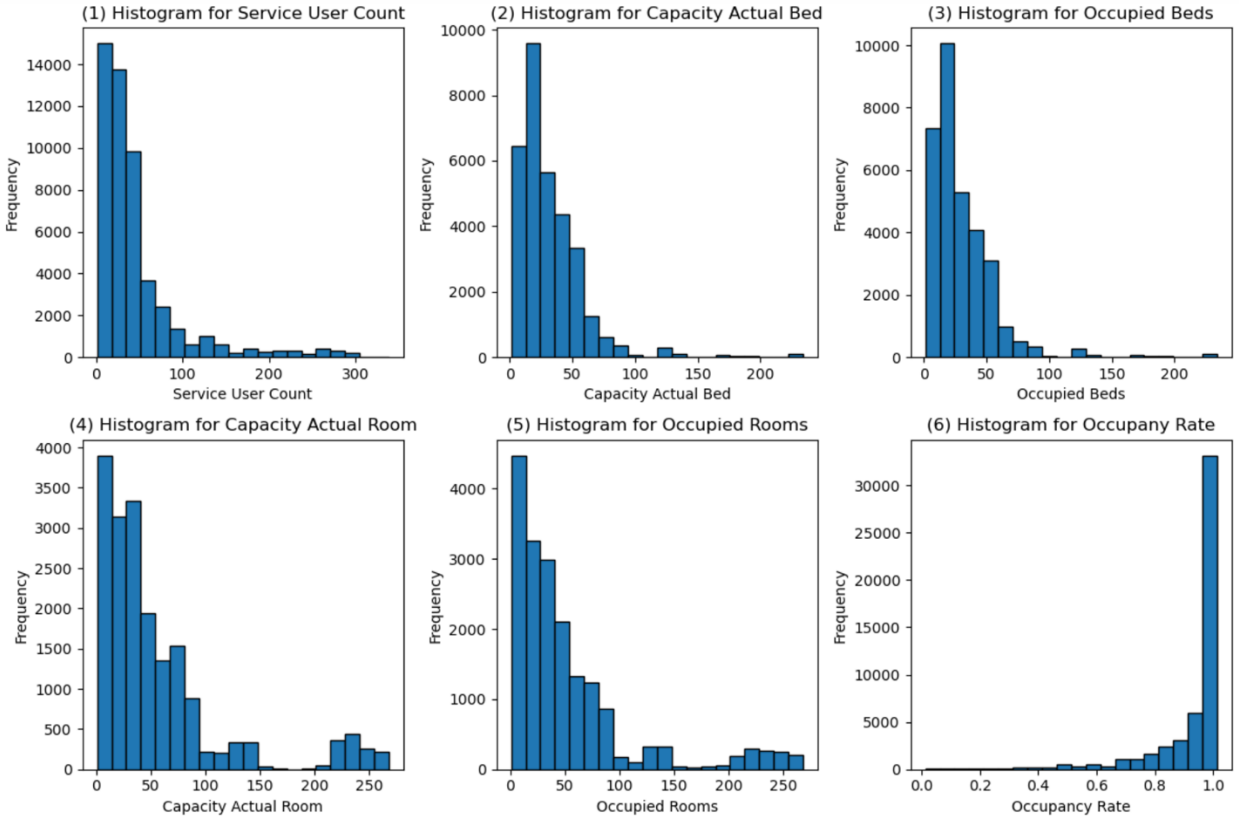

Figure 1

The **bed-based capacity type** reveals the number of available beds (CAPACITY_ACTUAL_BED) at a shelter with an average of approximately 32 beds, while the average for OCCUPIED_BEDS is close to 30 beds. The histograms for both CAPACITY_ACTUAL_BED and OCCUPIED_BEDS (referenced in Figure 1, (2) & (3)), show a right-skewed distribution. The majority of data points for both actual bed capacity and occupied beds cluster towards the lower end of the scale, suggesting that most programs offer a smaller bed capacity. In terms of the **room-based records**,

it shows an average capacity (CAPACITY_ACTUAL_ROOM) of about 56 rooms at a shelter, with an average occupancy (OCCUPIED_ROOMS) of nearly 53 rooms. The histograms for CAPACITY_ACTUAL_ROOM and OCCUPIED_ROOMS (Figure 1, (4) & (5)) display a similar trend to the bed-based type. These figures indicate a high utilization of shelters, operating at a capacity with minimal vacancy. The skewed distribution of rooms and beds means that the majority of programs function on a smaller scale regarding physical space.

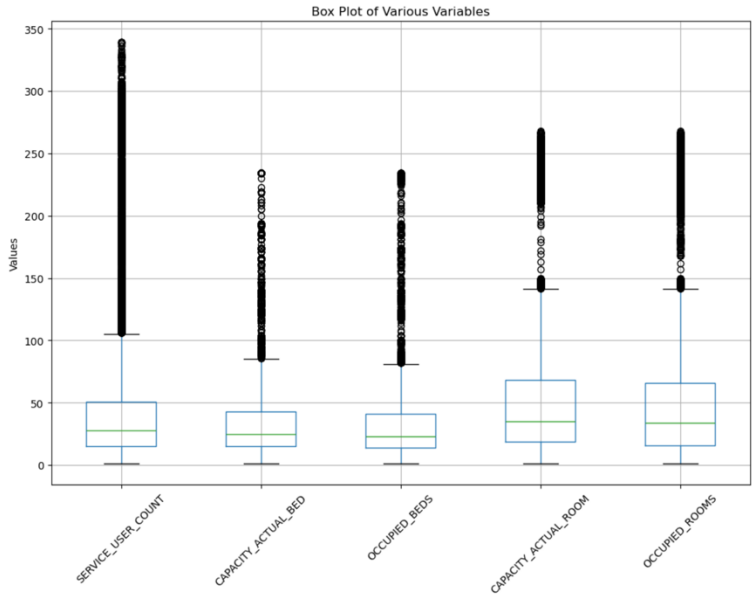

Figure 2

The boxplot (Figure 2) for these variables showcases similar trends, with each distribution displaying a concentration of values in the lower range, and a numerous outlier that extend upwards.

The **Occupancy Rate** is the most important variable in this study. The mean occupancy rate stands at 93%, indicating a high level of occupancy across the shelter programs. With a standard deviation of 13.8%, the data shows a moderate variation around this mean. Occupancy rates range from a minimum of 1.2% to a maximum that slightly surpasses full capacity at 101%. The existence of values beyond the 100% threshold suggests potential inaccuracies within the dataset, as the occupancy rate should fall between 0% and 100%. Although special circumstances may exist that cause the over 100% occupancy rate, we could consider trim the dataset later for better result. After a closer examination of the median and the 75th percentile, it's clear that both are at 100% occupancy. The histogram for occupancy rates (Figure 1 (6)) reveals a left-skewed distribution, with most of the data clustered towards the upper end, indicating that the majority of shelters operate at or near full capacity. This situation highlights the high demand for shelter services and space. As a result, the government could consider either expanding shelter capacities or adding new shelters to meet the high demand levels in the city.

Categorical Fields

Now, looking at the categorical features, the categories for each feature are listed in Table 2:

Table 2

Fields	Categories
SECTOR	Families, Mixed Adult, Men, Women, Youth

PROGRAM_MODEL	Emergency, Transitional, nan
OVERNIGHT_ SERVICE_TYPE	Motel/Hotel Shelter, Shelter, Interim Housing, Isolation/Recovery Site, 24-Hour Respite Site, Warming Centre, 24-Hour Women's Drop-in, nan
PROGRAM_AREA	COVID-19 Response, Base Shelter and Overnight Services System, T emporary Refugee Response, Winter Programs, nan
CAPACITY_TYPE	Room Based Capacity, Bed Based Capacity

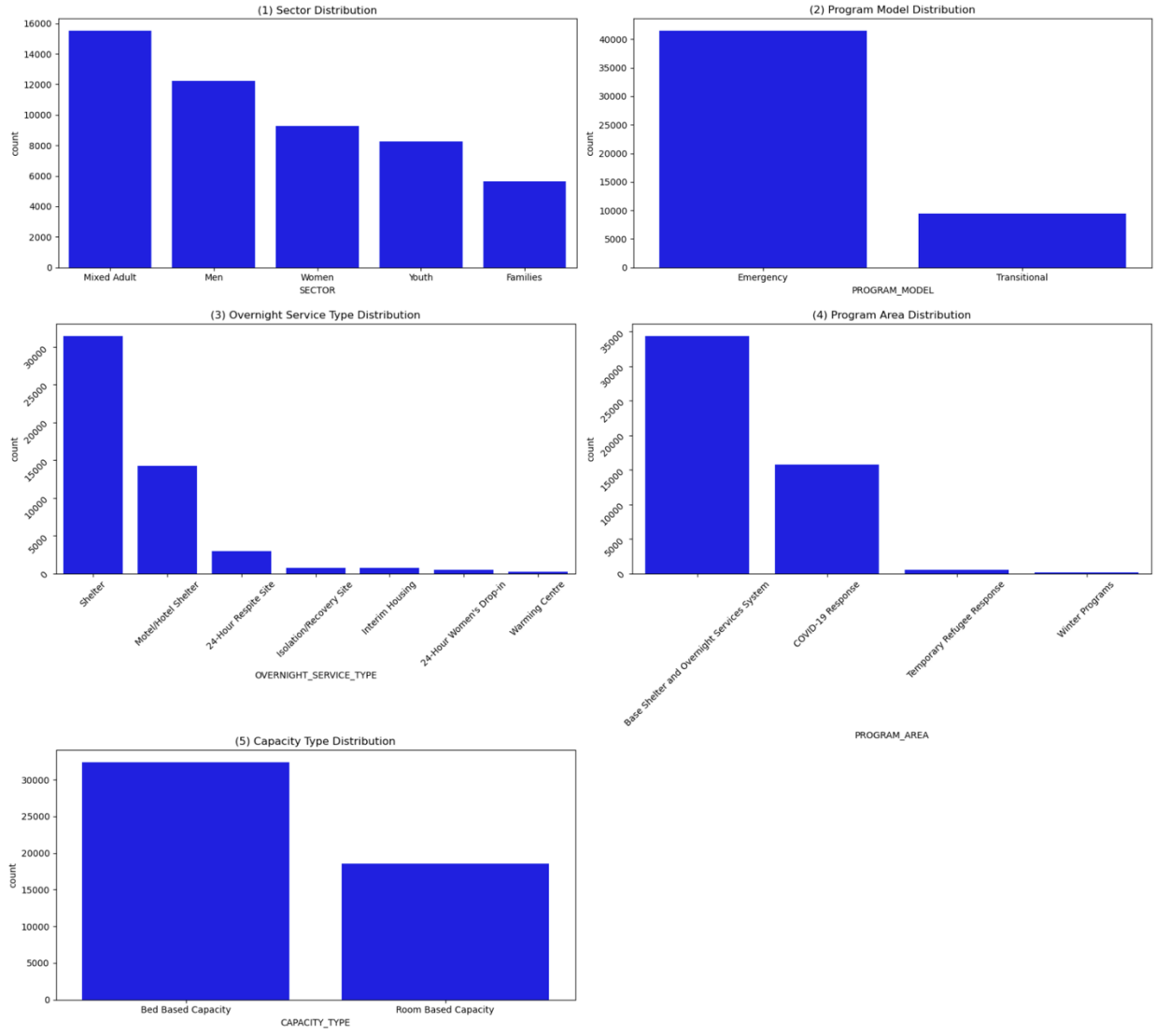

Figure 3

Five bar charts (Figure 3) were created for data visualization. In the Sector distribution graph (Figure 3 (1)), the 'Mixed Adult' sector has the highest count, followed by 'Men' and 'Women', with 'Families' and 'Youth' having a smaller portion of the shelter population. In terms of Program Model (Figure 3 (2)), shelters under the 'Emergency' model significantly outnumber the 'Transitional' category. There seems to be a need for immediate, short-term shelter options. Further analysis is needed to understand which type of shelter program model is more in need. The

Overnight Service Type graph (Figure 3 (3)) shows that the 'Shelter' category is most common, while alternatives such as 'Interim Housing' and '24-Hour Women's Drop-in' are less common. The Program Area graph (Figure 3 (4)) illustrates a significant count for the 'Base Shelter and Overnight Services System', in contrast to other areas like 'COVID-19 Response' and 'Winter Programs'. Lastly, according to the Capacity Type analysis in Figure 3 (5), 'Bed Based Capacity' is more than 'Room Based Capacity.' Given that this category was used in creating and merging the occupancy rate field, it will be the focus of further analysis.

Data Cleaning

In the previous analysis, several invalid data points within our dataset were identified. To better assist the study on homelessness patterns in the City of Toronto, there was a need to perform a series of data cleaning steps. These steps included addressing percentages outside the valid range and dealing with "NaN" values that could skew our analysis. Any occupancy rate data exceeding 100% or falling below 0% were removed, resulting in the elimination of 4 rows from our dataset. Further examination revealed that the "PROGRAM_MODEL" field contained "NaN" values. All records with "NaN" values in this field were removed, thereby excluding 2 records from the study. These adjustments are unlikely to impact the overall findings significantly but will enhance the accuracy of our analysis.

T-test 1

PROGRAM_MODEL	count	mean	std	min	25%	50%	75%	max
Emergency	41537.0	0.941405	0.138563	0.012048	0.954545	1.000000	1.0	1.0
Transitional	9401.0	0.880381	0.128472	0.222222	0.818182	0.918919	1.0	1.0

Figure 4

Given the critical state of shelter occupancy, with most facilities operating at or near full capacity, the city might need to consider allocating additional funds for the development of shelter infrastructure. Previous examination on the breakdown of occupancy rates by the two program model types (Figure 4) reveals a higher occupancy rate in the Emergency model. This observation leads to a critical question: Does the program model influence occupancy rates, or is the observed difference due to some random variation? To address this question, an independent samples t-test was conducted with the primary goal to assess whether there is a statistically difference in the occupancy rates between the Emergency and Transitional program models.

We set the hypothesis as below:
Null Hypothesis (H0): There is no statistically difference in the occupancy rates between Emergency and Transitional program models.
Alternative Hypothesis (H1): There is a statistically difference in the occupancy rates between Emergency and Transitional program models.

The t-test results showed a p-value very close to 0. For this analysis, a significance level of 0.05 was adopted, as it is commonly used in similar studies. The p-value very close to 0 is significantly lower than 0.05, there is very strong evidence to reject the null hypothesis. This indicates a statistically significant difference in the occupancy rates between shelters operating under the Emergency program model and shelters under the Transitional model. The results highlight an imbalance in occupancy rates across different shelter program models in the city. The higher

occupancy rates in the Emergency model may signal a greater demand for immediate shelter. These findings can support the City's strategic planning, offering a factual basis for decision-making regarding program planning and resource allocation for shelter services. The city can consider prioritizing investments to expand the capacity of shelters, especially for the Emergency program model. This can help alleviate the current states on shelter capacities and better support people who are in urgent need.

T-test 2

Building on the insights from a prior investigation, a critical discovery was made: occupancy rates appear to vary across different times of the year, indicating a need to better understand these fluctuations. To explore this, the data was divided into 2 groups: the first half of the year (January 1 to June 30) and the second half (July 1 to December 31) for 2021. The objective was to discover any statistical differences that might result from the temporal differences due to seasonality and its influence on shelter occupancy. To achieve this, an independent samples t-test was conducted with the hypotheses set as follows:

Null Hypothesis (H0): There is no statistically difference in the occupancy rates between the first half and the second half of the year 2021.
Alternative Hypothesis (H1): There is statistically difference in the occupancy rates between the first half and the second half of the year 2021.

With a p-value of 1.270644163653919e-11, which is well below the significance threshold of 0.05, there is strong evidence to reject the null hypothesis. This outcome suggests that the time of year has an influence on shelter occupancy rates, with variations likely tied to seasonal dynamics. Clearly, the shelter needs are not constant but fluctuate with the seasons. This finding can be very useful in city planning and resource allocation for shelter services. The city can take the temporal factor into account when allocating resources and planning shelter capacities to ensure the system is responsive to the cyclical nature throughout the year. This might include setting up temporary shelters during peak months to accommodate the increased need. Adjusting shelter resources can enhance the efficiency of the city's resource use and provide more effective support to people who are need shelter services.

Conclusion

This analysis of Toronto's shelter system in 2021 aimed to reveal some critical insights into homelessness and shelter utilization in the city. The city's shelters act as a support system for unhoused individuals, working hard to satisfy the growing demand for shelter spaces. The analysis reveals a statistically difference in occupancy rates between the Emergency and Transitional program models, which indicates the type of shelter program can influence occupancy rates. The second t-test shows that occupancy rates vary at different times of the year, with a seasonal impact on shelter utilization. The study aims to address homelessness more effectively by highlighting areas for targeted intervention and resource allocation in Toronto. By incorporating these insights into city planning and shelter management strategies, Toronto can do better at meeting the fluctuating and increasing demands for shelter. Due to time constraints, the study did not dive into specific demographic characteristics of shelter users and occupancy rate. Future studies can consider exploring whether specific shelter sectors (Mixed Adult, Men, Women, Youth) affect occupancy rates to further refine targeted interventions and resource allocations.