

Topic: Toronto Shelter Data set Analysis  
Name: Ruiying Wang  
Student number: 1003059762  
Course: INF2178

The dataset provides a comprehensive overview of Toronto's shelter system over a one-year period (January 2021 to January 2022), detailing capacity, occupancy, and service user statistics across various programs and organizations. Each organization offers multiple programs targeted at different sectors, with a variety of program models (emergency or transitional) and capacity types (room-based or bed-based). This analysis aims to address two key research questions:

- 1) How has the shelter system's performance changed over time?
- 2) Which programs and capacities are in high demand and require enhancements?

To answer these questions, the study employs descriptive statistics and visualizations, including line charts, box plots, and histograms, to examine trends in service user counts and occupancy rates over time, by capacity type, and by program model. This approach helps identify changing occupancy trends and pinpoint primary challenges through time series and box plot visualizations. Additionally, the report includes two statistical tests (t-tests) to compare the mean occupancy rates by program model and capacity type, aiming to ascertain if significant differences exist between them.

### **Exploratory data analysis**

The initial phase of exploratory data analysis includes three steps: data cleansing, preparation, and in-depth analysis. This process involves the strategic exclusion of unused columns such as organization name, program ID, and specific program details, which are not directly related to the core analysis objectives. The identification and removal of records with missing program model data are critical to ensuring the accuracy of the statistical analysis. Furthermore, the computation of occupancy rates from available capacity and occupancy data are calculated for conducting analysis segmented by program model and capacity type, focusing on a range of descriptive statistics including means, standard deviations, and various percentiles.

From the Summary statistics table (Table 1), the average occupancy rate is at 0.93, indicating a 93% usage rate, with most data points falling between 90% and 100%. The average bed and emergency program occupancy rates are similarly high, around 0.93, showing a strong and consistent demand. Transitional program occupancy has a relatively lower mean of 0.88, suggesting slightly less utilization in usage.

### **Diagram 1: Summary statistics table**

#### Overall Summary statistics

	SERVICE_USER_COUNT	Capacity	Occupied	Occupancy_rate
count	50944.000000	50944.000000	50944.000000	50944.000000
mean	45.727171	40.335447	38.159567	0.930142
std	53.326049	43.439493	42.702181	0.138788
min	1.000000	1.000000	1.000000	0.012048
25%	15.000000	16.000000	14.000000	0.923077
50%	28.000000	27.000000	25.000000	1.000000
75%	51.000000	49.000000	47.000000	1.000000
max	339.000000	268.000000	268.000000	1.014085

#### Summary statistics for Room Occupancy Rate

Min: 0.02  
Mean: 0.93  
Max: 1.0  
Stdev: 0.12  
25th percentile: 0.9  
Median: 1.0  
75th percentile: 1.0  
IQR: 0.1

#### Summary statistics for Bed Occupancy Rate

Min: 0.01  
Mean: 0.93  
Max: 1.01  
Stdev: 0.16  
25th percentile: 0.96  
Median: 1.0  
75th percentile: 1.0  
IQR: 0.04

#### Summary statistics for Emergency Program Occupancy Rate

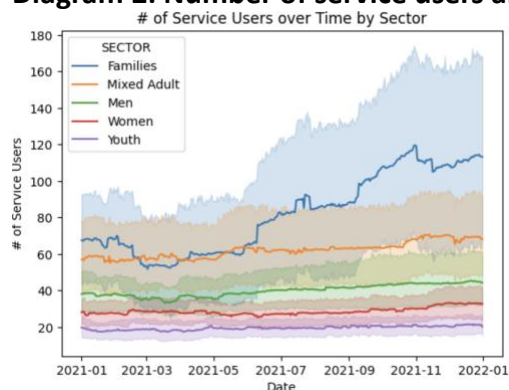
Min: 0.01  
Mean: 0.94  
Max: 1.01  
Stdev: 0.14  
25th percentile: 0.95  
Median: 1.0  
75th percentile: 1.0  
IQR: 0.05

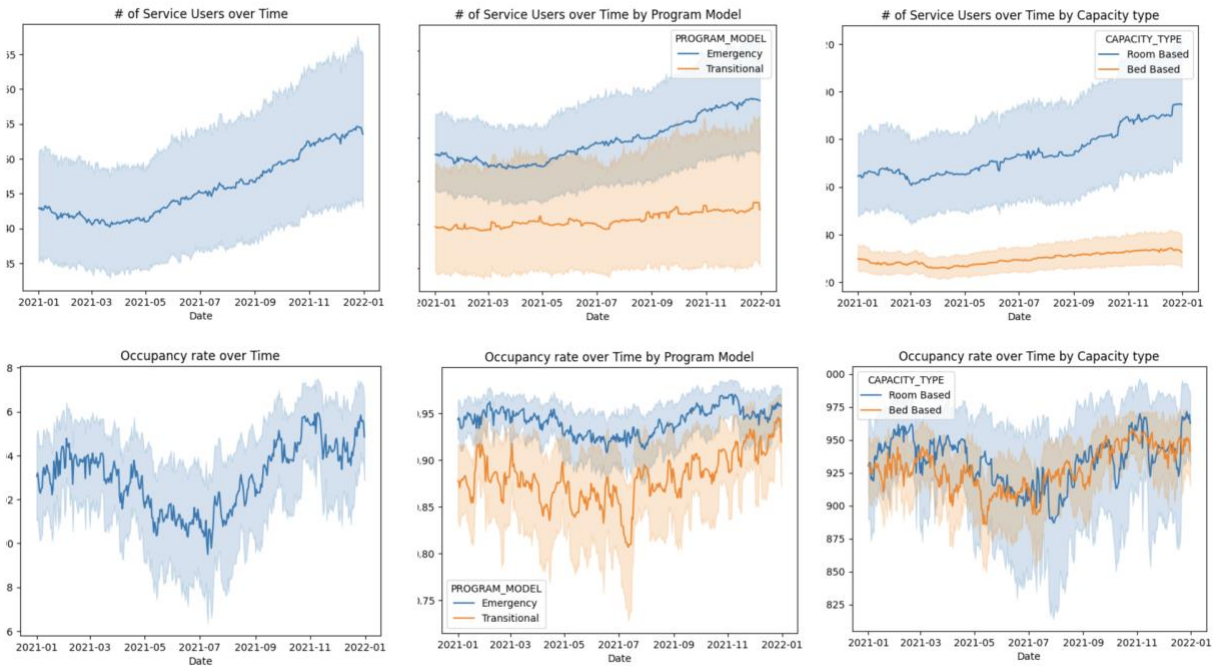
#### Summary statistics for Transitional Program Occupancy Rate

Min: 0.22  
Mean: 0.88  
Max: 1.0  
Stdev: 0.13  
25th percentile: 0.82  
Median: 0.92  
75th percentile: 1.0  
IQR: 0.18

The time series analysis, as shown in Diagram 2, showcases a consistent upward increase in the number of service users throughout the observed period, with a particularly marked increase in the room-based capacity types, in contrast to the relatively stable demand for bed-based capacities. The analysis further reveals a significant upsurge in the utilization of emergency services, notably outpacing the more gradual growth observed in transitional services. Despite periodic fluctuations, occupancy rates have largely remained elevated, with both room-based and emergency services experiencing a minor downtrend as the period draws to a close. This pattern underscores a growing dependency on these critical services, highlighting the emergency and room-based programs are in intensive use throughout the year. Moreover, a noteworthy observation is the narrowing gap in occupancy rates between emergency and transitional programs over time. Despite the steady count of transitional service users, this trend points to a concern of diminishing capacity, which in turn lead to an increase in occupancy rates, signaling an urgent need for strategic capacity expansion in these key areas.

**Diagram 2: Number of service users and occupancy rates over time**





## Statistical analysis

Before moving on to the t-test analysis, it's important to first look at how the overall occupancy rates and the number of service users are spread out. This is done using histograms for a general view, followed by boxplots that show differences based on the type of program and capacity, as seen in Diagram 3. The histogram for occupancy rates shows that most of the rates are high, with a big peak at 100% occupancy, indicating a significant portion of capacity types are often fully occupied. On the other hand, the histogram for the number of service users shows that most services cater to a smaller number of people, with fewer services supporting a large number of users. This high occupancy rates and the concentration of service users in the lower range underscores a critical capacity shortfall within the shelter system.

**Diagram 3: Histogram Distribution of Occupancy rate and Service User Number**

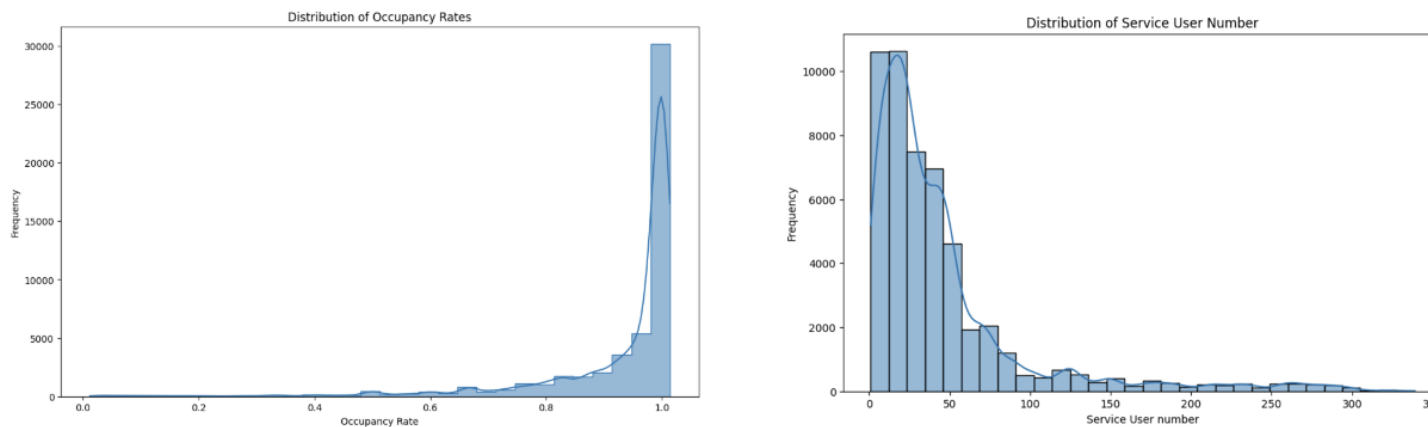
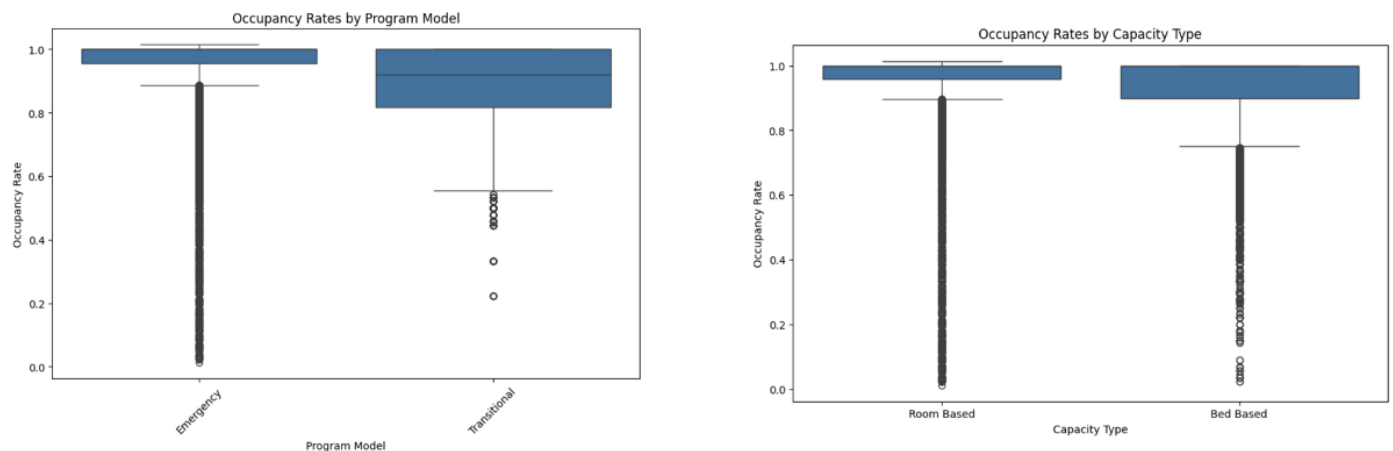


Diagram 4 presents a clear comparison of occupancy rates across different program models and capacity types. The left boxplot shows the occupancy rates for emergency programs, which are consistently high, with most values clustering close to full capacity. This indicates that emergency shelters are often near or at their maximum capacity, highlighting their critical demand. In comparison, the boxplot for transitional programs shows a broader spread of occupancy rates and a lower median, suggesting that these programs experience more fluctuation in their occupancy and are generally less full than emergency programs. This variability points to a more uneven demand for transitional shelter services.

On the right, the boxplot for room-based capacities reveals a tight grouping of values near the maximum occupancy rate of 1, meaning that room-based options are frequently at or near full occupancy, indicating a strong and consistent demand for room-based shelters. Conversely, bed-based capacities have a wider spread of occupancy rates and a lower median, indicating a greater variability in usage which suggests that the demand for bed-based accommodations fluctuates more than for room-based ones, reflecting varied needs and preferences among service users.

**Diagram 4: Boxplot of Occupancy rate by Program model and Capacity Type**



## Welch's t-test

Given the unequal variances between groups, which contravenes the standard t-test assumption, we opted for Welch's t-test to individually assess differences in mean occupancy rates by program model and capacity type, assuming a 95% confidence level.

### 1. Occupancy rate by program model

*Hypothesis:*

*H0: There is no difference in mean occupancy rate between Emergency program and Transitional Program.*

*H1: There is difference in mean occupancy rate between Emergency program and Transitional Program.*

From Diagram 5, the positive t-statistic indicates that the mean occupancy rate of the emergency program is higher than that of the transitional program. The p-value of 0, which indicates a value extremely close to zero, suggests that there is a statistically significant difference in mean occupancy rates between emergency and transitional program models.

**Diagram 5: T-test: Mean Occupancy rate of Program type: Emergency and Transitional**  
T-test: Mean Occupancy rate of Program type: Emergency and Transitional

t-statistic = 40.981115372199206

p-value = 0.0

## **2. Occupancy rate by capacity type**

*Hypothesis:*

*H0: There is no difference in mean occupancy rate between Room capacity type and Bed capacity type.*

*H1: There is difference in mean occupancy rate between Room capacity type and Bed capacity type.*

From Diagram 6, The negative t-statistic indicates that the mean occupancy rate of the room capacity type is lower than that of the bed capacity type. The p-value is extremely low which reject null hypothesis and suggests that there is a statistically significant difference in mean occupancy rates between room and bed capacity types.

**Diagram 6: T-test: Mean Occupancy rate of Room capacity Type and Bed capacity Type**  
T-test: Mean Occupancy rate of Room capacity Type and Bed capacity Type

t-statistic = -4.498751771925636

p-value = 6.860477551487939e-06

## **Conclusion**

From the analysis above, it is noticed that there is a growing demand for Toronto's shelter services, particularly spotlighting emergency, and room-based programs are in critical need. The observed concentration of service users within lower ranges signals a significant challenge of insufficient program capacity to meet escalating demands. Addressing this gap shows necessitates in strategic expansions in emergency and room-based capacities, especially for family-oriented services, which have evidenced the most significant growth in demand. Future improvement should take broader demographic analyses, stakeholder feedback incorporation, predictive demand modeling, and comparative studies into consideration to fostere a more comprehensive understanding of shelter system requirements and facilitating informed decision-making for strategic enhancements.