

# INF2178 Technical Assignment 1

## Experimental Design for Data Science

### 1. Understand the dataset

#### Background

The "INF2178\_A1\_data.xlsx" dataset details the escalating homelessness issue in Toronto, marked by an unhoused population that exceeds shelter capacity. This dataset offers daily occupancy and capacity records for 2021 across different shelter programs and organizations. It highlights each program's details and utilization, presenting an in-depth view of the challenges and dynamics in providing shelter to the city's unhoused individuals.

#### Columns chosen

- SECTOR: The group the program serves (e.g., families, women).
- CAPACITY\_TYPE: Measurement type for accommodation (bed or room based).
- PROGRAM\_MODEL: Type of shelter program (e.g., emergency, transitional).
- SERVICE\_USER\_COUNT: Number of people using the program.
- CAPACITY\_ACTUAL\_BED: Available beds in the shelter.
- OCCUPIED\_BEDS: Beds currently in use.
- CAPACITY\_ACTUAL\_ROOM: Available rooms in the shelter.
- OCCUPIED\_ROOMS: Rooms currently in use.

### 2. Data cleaning and calculate occupancy rates

The occupancy rates for beds and rooms have been calculated and added as new columns to the dataset. Bed/Room Occupancy Rate gives the percentage of beds/rooms that are occupied out of the total available beds/rooms.

Then, dropped the rows where both occupancy rates are NaN because they can't be used in t-tests.

	BED_OCCUPANCY_RATE	ROOM_OCCUPANCY_RATE
count	32399.000000	18545.000000
mean	0.927885	0.934087
std	0.122562	0.163241
min	0.022727	0.012048
25%	0.900000	0.958333
50%	1.000000	1.000000
75%	1.000000	1.000000
max	1.000000	1.014085

The statistical summary shows that the shelter system is heavily utilized, with both beds and rooms frequently operating at full capacity. On average, more than 92% of beds and 93% of rooms are occupied, with half of the time shelters being completely full. On average, room type has a higher occupancy rate than bed type, which indicates a greater utilization of available space of rooms. This high level of occupancy underscores the critical demand for shelter services and suggests that the current capacity may be insufficient to meet the needs of the unhoused population in Toronto. The fact that room occupancy rates exceed 100% at times points to instances of overcrowding or potential data discrepancies. This indicates a pressing need for increased shelter capacity or alternative solutions to address homelessness in the city.

### 3. Quantitative analysis using t-tests

Two-sample t-tests have been performed comparing bed occupancy rates and room occupancy rates between 'Emergency' and 'Transitional' program models.

#### **3.1 For the t-test comparing BED OCCUPANCY RATE:**

##### **Hypotheses:**

H0 (Null Hypothesis): There is no significant difference in the bed occupancy rates between Emergency and Transitional program models.

H1 (Alternative Hypothesis): There is a significant difference in the bed occupancy rates between Emergency and Transitional program models.

##### **Analysis:**

t-statistic: 38.78

p-value: 0.0

Given the t-statistic and a p-value of 0.0, we reject the null hypothesis for the bed occupancy rate, which suggests that there is a statistically significant difference in bed occupancy rates between Emergency and Transitional shelters.

#### **3.2 For the t-test comparing ROOM OCCUPANCY RATE:**

##### **Hypotheses:**

H0 (Null Hypothesis): There is no significant difference in the room occupancy rates between Emergency and Transitional program models.

H1 (Alternative Hypothesis): There is a significant difference in the room occupancy rates between Emergency and Transitional program models.

##### **Analysis:**

t-statistic: 18.90

p-value:  $5.92 \times 10^{-79}$  (very close to zero)

Similarly, the p-value is less than 0.05, which indicates that we reject the null hypothesis for the room occupancy rate, which means there is also a statistically significant difference in room occupancy rates between Emergency and Transitional shelters.

In both cases, the results show significant differences, meaning the type of program

model (Emergency vs. Transitional) is associated with different occupancy rates, which could have implications for how these types of shelters are managed and funded. For example, emergency shelters often have high occupancy due to urgent housing needs, while Transitional shelters may see varied occupancy related to longer service durations. This discrepancy suggests a need for tailored resource allocation, with possibly more funding for emergency housing or streamlined transitions to long-term solutions. These occupancy trends can guide shelter organizations in strategic planning, operational improvements, and service customization. For example, low Transitional shelter occupancy could prompt reviews of access barriers or the efficiency of moving clients from Emergency shelters.

Also, a two-sample t-test was conducted to compare the "SERVICE\_USER\_COUNT" between "Emergency" and "Transitional" program models.

### **3.3 For the t-test comparing SERVICE\_USER\_COUNT:**

#### **Hypotheses:**

H0 (Null Hypothesis): There is no significant difference in the service user count between Emergency and Transitional program models.

H1 (Alternative Hypothesis): There is a significant difference in the service user count between Emergency and Transitional program models.

#### **Analysis:**

**t-statistic: 29.94**

**p-value:  $3.17 \times 10^{-195}$  (practically zero)**

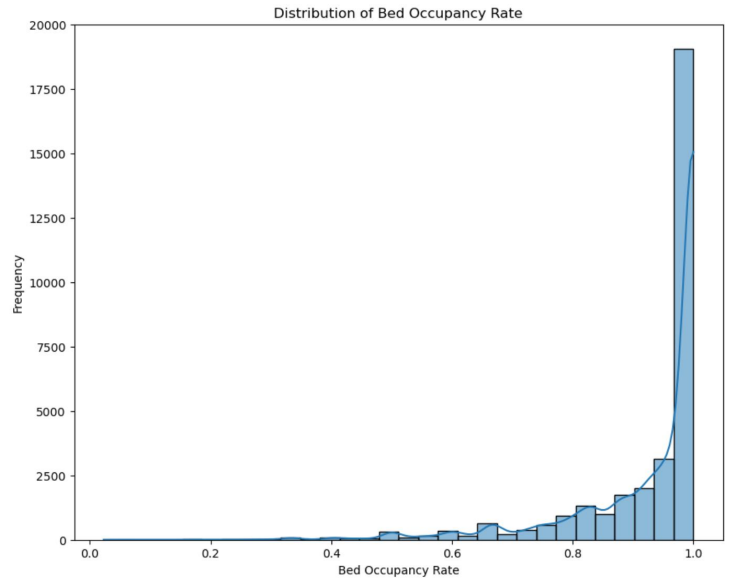
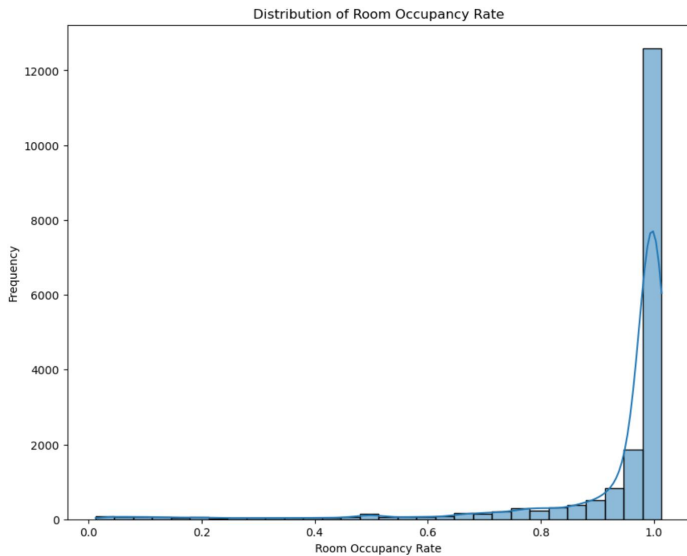
The t-statistic is significantly high, and the p-value is far below 0.05. This result leads us to reject H0 and conclude that there is a statistically significant difference in the service user count between Emergency and Transitional shelters.

This could imply that either Emergency or Transitional programs are more heavily utilized, which in turn could reflect the nature of demand for shelter services. Emergency shelters may have higher turnover and cater to immediate needs, while Transitional programs might support individuals for longer durations with different capacity constraints. This difference is essential for resource allocation, policy-making, and strategic planning for shelter services. Stakeholders may need to consider these differences when designing, funding, and operating shelter programs to effectively meet the needs of the unhoused population. The substantial variance between the two models might also suggest areas where additional support or program expansion could be necessary.

## **4. Exploratory data analysis**

### **4.1 Room/Bed Occupancy Rate Histogram Analysis**

Histograms were created to show the distribution of room/bed occupancy rate.



### Spread and Variability:

The histogram for bed occupancy rates shows a wider spread of values before the peak at 100%. This indicates that there is a greater variability in bed occupancy rates, with more shelters experiencing bed occupancy rates that are not at full capacity.

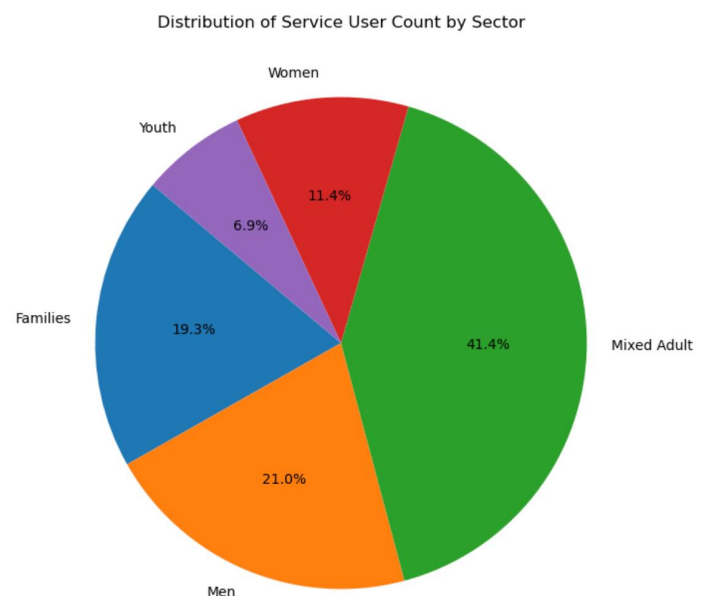
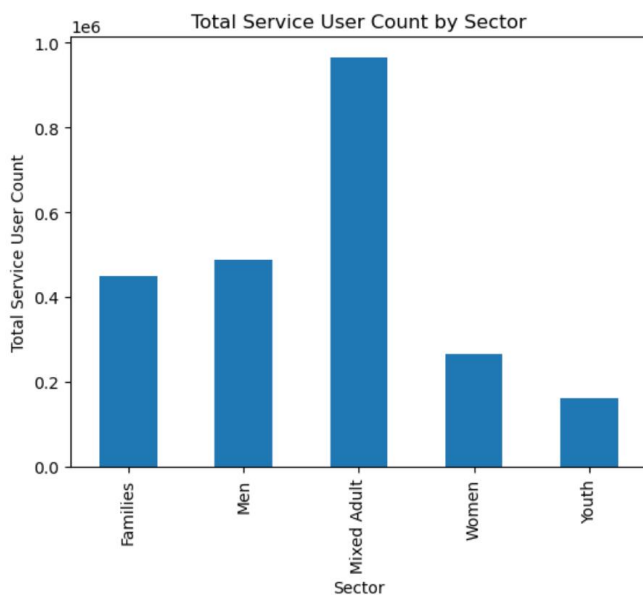
The room occupancy rate histogram shows a narrower spread with a steeper rise to the 100% occupancy rate, suggests that room occupancy rates are more consistently near full capacity, with fewer instances of rooms not being fully occupied.

### Frequency of Full Occupancy:

Both histograms have their mode at 100%, which indicates that the most common occupancy rate for both beds and rooms is full capacity. However, the peak at 100% occupancy is more pronounced for rooms, suggesting that it is more common for all rooms to be occupied than it is for all beds to be occupied.

## 4.2 Shelter Use by Sector in Toronto

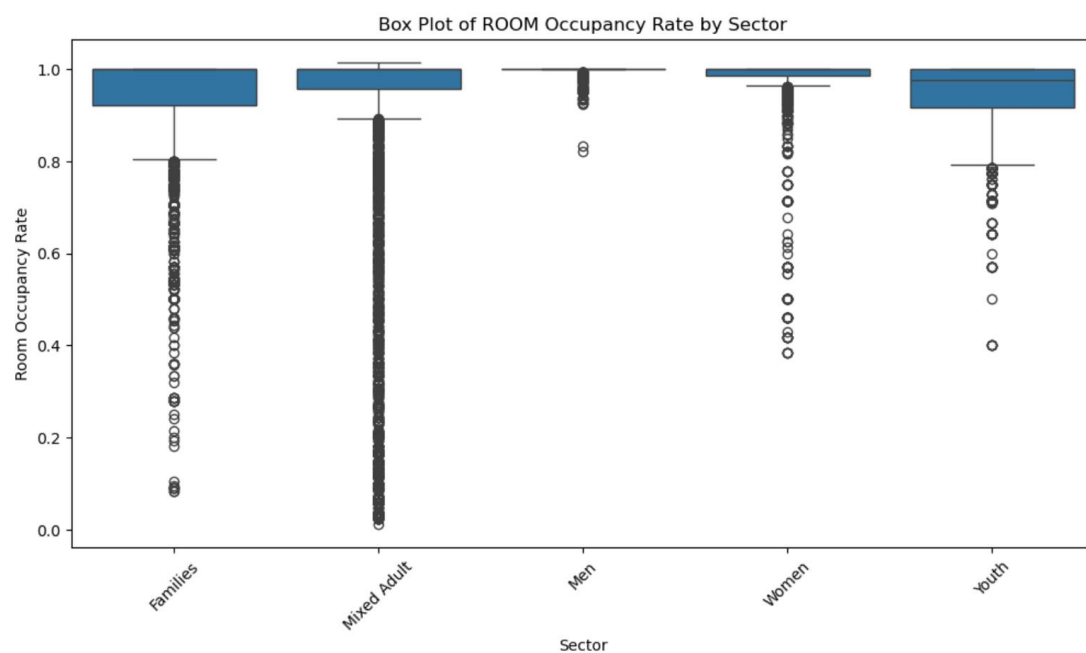
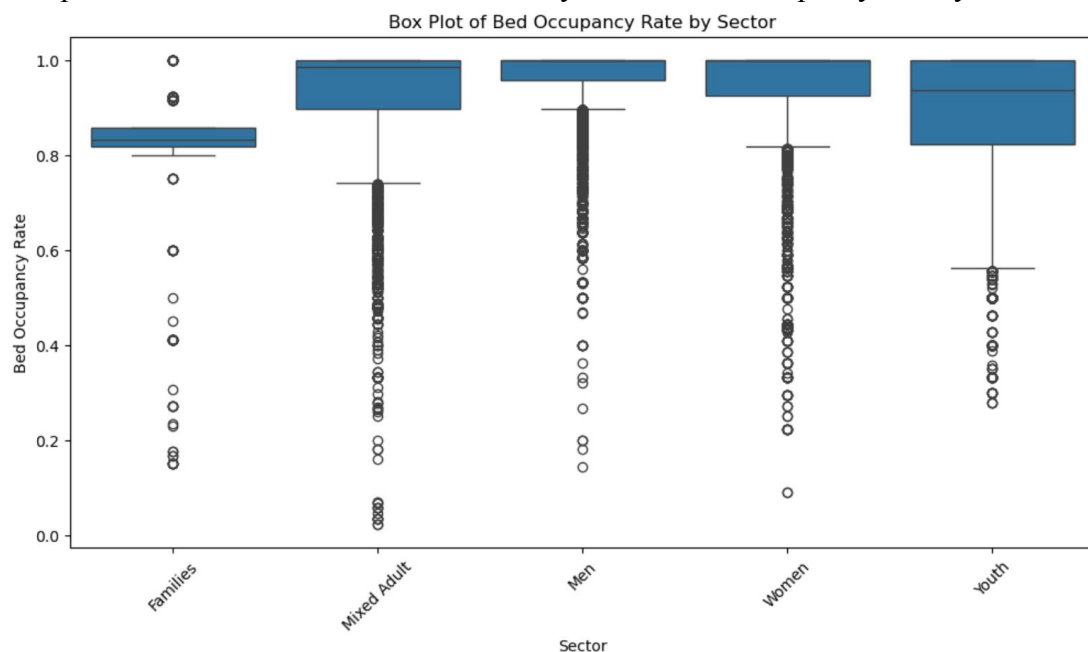
Bar plot and pie chart were created to show the distribution of service user count by sector.



The "Mixed Adult" sector leads in service usage, with "Men" and "Families" following, while "Women" and "Youth" sectors show considerably lower usage. This pattern suggests a higher demand or capacity for mixed adult services, and possibly lower demand or availability for women's and youth services. Mixed adults form the largest service user group, highlighting the sector's primary role in service consumption. The significant figures for "Men" and "Families" indicate they are also key user groups. Despite lower numbers, "Women" and "Youth" are critical demographics needing focused resources and services. The data underlines the necessity for tailored resource allocation and policy development to ensure equitable and adequate service provision across all sectors. The lesser engagement of "Women" and "Youth" may indicate a need to assess the adequacy of services and address potential access barriers for these groups.

### 4.3 Occupancy Variability in Toronto Shelters by Sector

Box plots were created to show the summary statistics of occupancy rate by sector.



Both bed and room occupancy rates are generally high, suggesting that the shelters are operating near or at full capacity for most sectors. The 'Mixed Adult' and 'Men' sectors show greater variability in occupancy rates, which may indicate a fluctuating demand or different capacity management strategies within these sectors. The consistency of high occupancy rates in the 'Youth' sector could indicate a stable and continuous demand for youth-specific shelter services. The presence of outliers, particularly in the lower ranges, could warrant further investigation to understand the factors contributing to these lower occupancy rates, such as seasonal variations, accessibility issues, or other operational factors.

## 5. Further analysis

- **Capacity and Demand:** Evaluate whether shelter capacities align with demand and if high occupancy stems from insufficient capacity or effective usage.
- **Seasonal Variability:** Examine seasonal trends affecting occupancy, including the necessity for temporary shelters during peak times.
- **Demographic Studies:** Analyze shelter usage demographics to identify specific needs and assess if certain groups are more prone to being turned away.
- **Outlier Examination:** Investigate outliers, particularly lower occupancy instances, to uncover factors like operational days, events, or program shifts causing these variances.
- **Trend Analysis:** Conduct a study over time to track occupancy rate changes, exploring long-term trends and implications for future shelter requirements.

## 6. Conclusion

There is a critical challenge of shelter occupancy and demand in Toronto, revealing that more than 92% of beds and 93% of rooms are typically occupied, indicating overcrowding. This high utilization rate underscores a pressing need for increased shelter capacity or alternative solutions to manage homelessness effectively. Key findings include statistically significant differences in bed and room occupancy rates between Emergency and Transitional shelters, with t-statistics of 38.78 and 18.90, respectively, and both p-value are less than 0.05, highlighting the need for differentiated resource allocation and policy strategies. The analysis also points to substantial service user count differences between these shelter types, underscoring the diverse needs of Toronto's unhoused population. In conclusion, it should call for tailored approaches in resource distribution, strategic planning, and policy formulation to address the identified gaps and improve shelter services across all sectors, with a particular focus on understanding and mitigating barriers to access for women and youth.