INF2178 Experimental Design For Data Science

Technical Assignment #1

Student Name: Liguangxuan He | Student ID: 1006141809

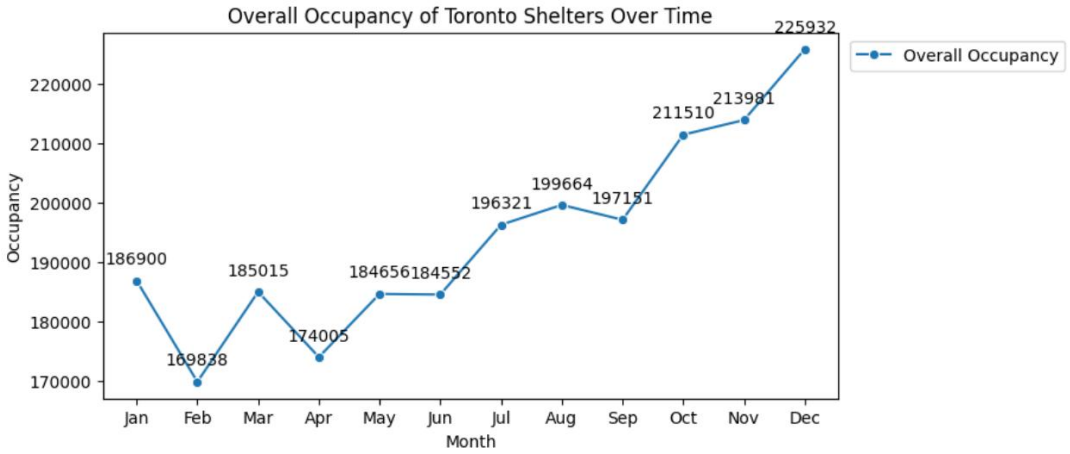
Before commencing any analysis, I typically review basic information about the data-set such as column names, data types, and non-null counts. During this initial assessment, I address missing values in specific columns like program_name, program_model, overnight_service_type, and ad program_area by replacing them with "Not specified." Notably, I observed that the sum of actual bed and room capacities, as well as occupied beds and rooms, equals the capacity type (50944), indicating no need for adjustments regarding null values in these columns because it will either be beds or rooms not simultaneously. As for program_id, considering its lack of utility for analysis purposes, I intend to convert it to a string type to prevent potential confusion with numerical values in the future.

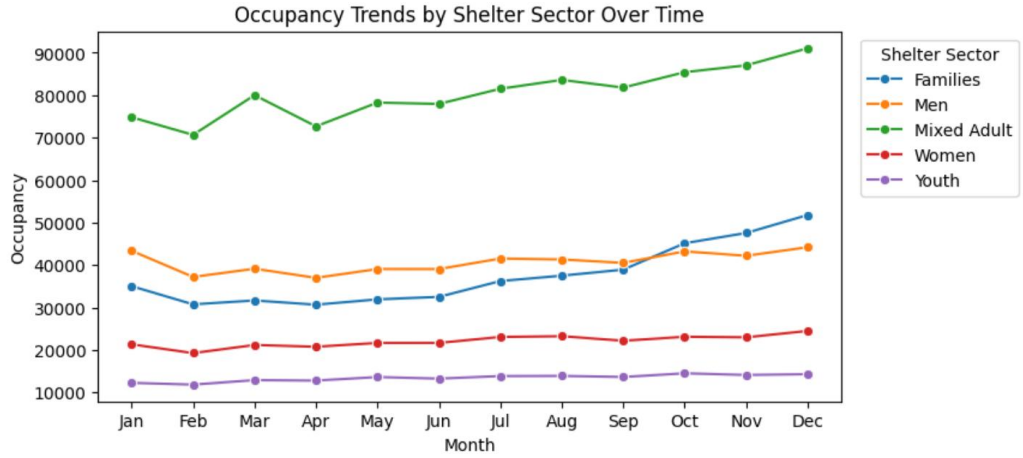
Abstract

This analysis of Toronto's homeless population and shelter system for 2021 reveals critical insights into shelter occupancy trends, service types, and capacity utilization. Despite efforts by the shelter support system to provide for the un-housed, challenges persist with individuals often turned away due to insufficient capacity. The study employs descriptive statistics, trend analysis, and hypothesis testing to understand these dynamics better. Initial findings indicate significant variations in occupancy rates across different times and shelter sectors, a predominance of certain service types, and varying capacity utilization across shelter models. These findings underscore the need for further research to optimize shelter distribution and capacity planning.

Exploratory data analysis (EDA)

1. Trend Analysis



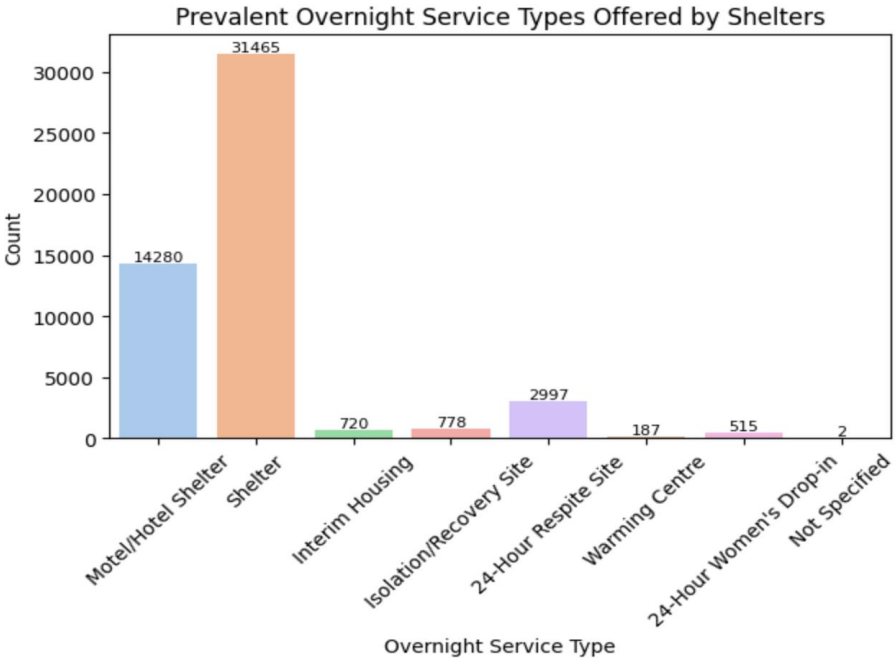


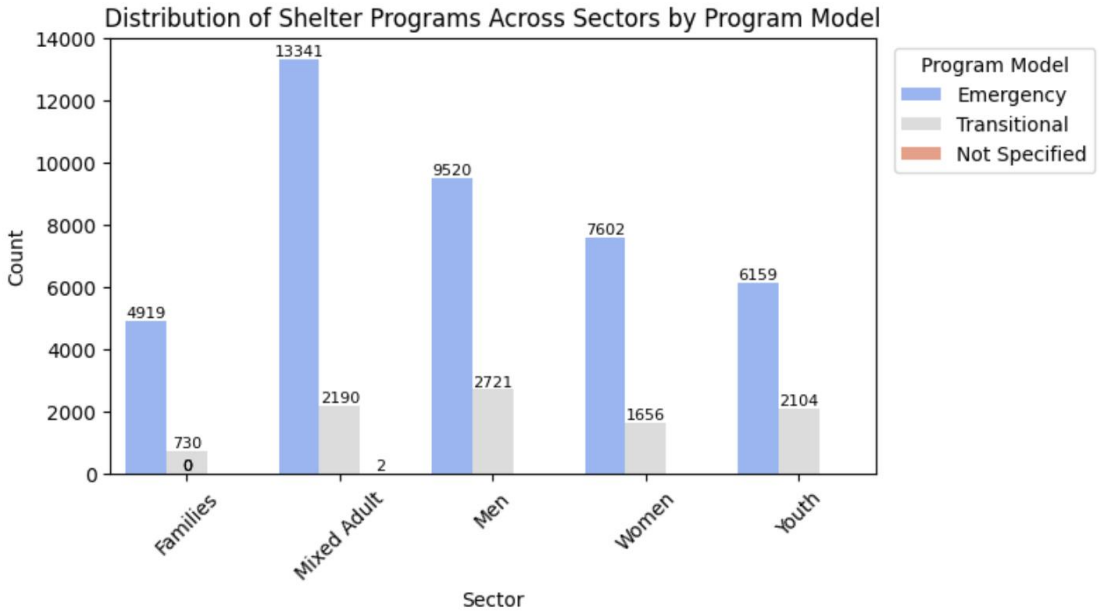
The exploratory data analysis (EDA) of Toronto shelter occupancy over 2021 reveals several key insights. Firstly, examining the overall occupancy trend indicates a notable increase throughout the year, reflecting the impact of the ongoing pandemic and economic instability. Notably, the trend displays a steady rise, possibly indicating persistent challenges faced by vulnerable populations. However, when dissecting occupancy by shelter sector, intriguing patterns emerge. The mixed adult sector stands out with significantly higher occupancy levels, suggesting unique needs or systemic factors driving this trend, which warrants deeper investigation. Conversely, youth and women's sectors exhibit relatively lower and stable occupancy levels, possibly indicating targeted interventions or community support systems. The crossing of the men's and families' occupancy lines in September highlights potential shifts in demand and resource allocation within shelter sectors, prompting the need for dynamic response strategies. Moreover, the parallel nature of women and youth sectors' occupancy suggests a shared vulnerability or common challenges among these demographics, necessitating tailored support programs. Overall, while the general upward trajectory of shelter occupancy underscores persistent societal issues

exacerbated by the pandemic, the nuanced patterns across shelter sectors underscore the importance of targeted interventions and adaptive policy frameworks to address diverse needs effectively.

---

2. Program Analysis



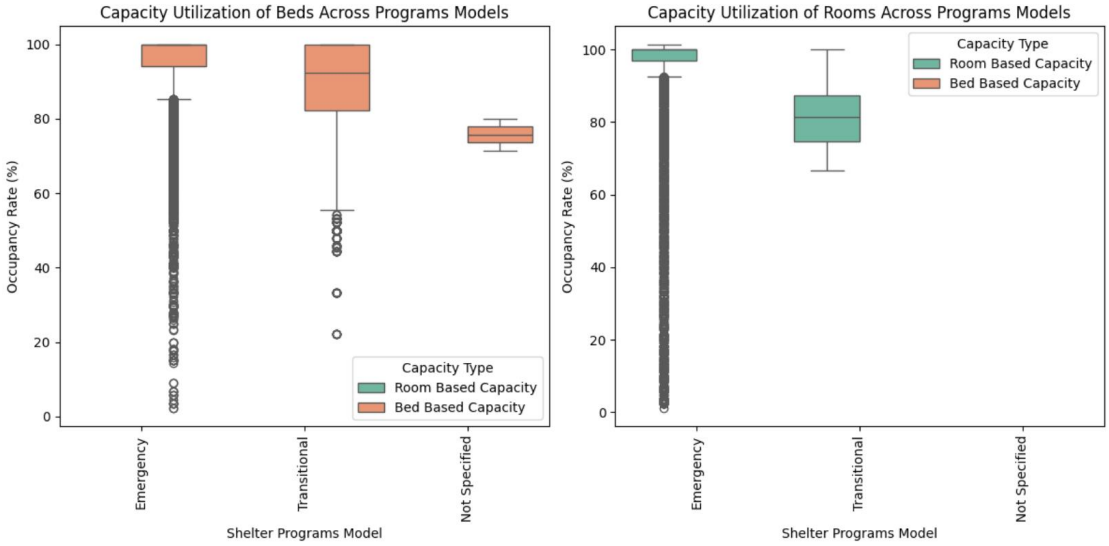


The prevalent overnight service type offered by shelters, as depicted in Figure 1, shows that the most common service is provided by traditional shelters, with 31,465 counts, followed by motel/hotel shelters at 14,280 counts. The distribution of shelter programs across various sectors, as illustrated in Figure 2, reveals that emergency programs significantly outnumber transitional ones across all sectors, indicating a higher demand for immediate and short-term assistance. Notably, the emergency

counts are consistently higher than transitional counts, such as in the family sector (4,919 emergency vs. 720 transitional) and mixed adult sector (13,341 emergency vs. 2,190 transitional). This suggests a pressing need for immediate shelter solutions across demographics. It is crucial to delve deeper into the factors contributing to the higher prevalence of emergency programs, exploring whether it is driven by the nature of the crises faced by individuals or systemic issues in providing transitional solutions. Further investigation into the specific requirements and challenges of each sector, particularly in understanding the varying needs of men, women, and youth, can inform targeted interventions and resource allocations. Additionally, considering the relatively low counts in certain service types, such as warming centers and 24-hour women's drop-ins, warrants an examination of their efficacy and whether there is an unmet demand that needs addressing.

---

3. Capacity Analysis




```
Summary statistics of occupancy rate of beds for Emergency programs:
Min: 2.27
Mean: 94.35
Max: 100.0
25th percentile: 94.12
Median: nan
75th percentile: 100.0
Interquartile range (IQR): 5.88


Summary statistics of occupancy rate of rooms for Emergency programs:
Min: 1.2
Mean: 93.86
Max: 101.41
25th percentile: 96.97
Median: nan
75th percentile: 100.0
Interquartile range (IQR): 3.03
```

```
Summary statistics of occupancy rate of beds for Transitional programs:
Min: 22.22
Mean: 88.52
Max: 100.0
25th percentile: 82.14
Median: nan
75th percentile: 100.0
Interquartile range (IQR): 17.86


Summary statistics of occupancy rate of rooms for Transitional programs:
Min: 66.67
Mean: 82.32
Max: 100.0
25th percentile: 74.6
Median: nan
75th percentile: 87.3
Interquartile range (IQR): 12.7


Summary statistics of occupancy rate of beds for Not Specified programs:
Min: 71.43
Mean: 75.71
Max: 80.0
25th percentile: 73.57
Median: 75.71
75th percentile: 77.86
Interquartile range (IQR): 4.29
```

The examination of occupancy rates across various shelter program models unveils distinct usage patterns. Emergency shelters consistently exhibit high utilization, boasting a mean bed occupancy rate of 94.35% and a room occupancy rate of 93.86%. The narrow inter-quartile range (IQR) for both beds (5.88%) and rooms (3.03%) within emergency programs signifies minimal variability in utilization. In contrast, transitional shelters demonstrate slightly lower mean bed (88.52%) and room (82.32%) occupancy rates, characterized by larger IQRs of 17.86% and 12.7%, respectively, suggesting more variability in demand for transitional programs. Programs labeled as "Not Specified" notably display the lowest mean bed occupancy rate (75.71%) and a small IQR of 4.29%, implying consistent but lower overall utilization. These statistics underscore the stability and consistently high demand for emergency shelters, while transitional and unspecified programs experience more variable occupancy rates. To comprehend the usage dynamics further, we calculated the occupancy rates for beds and rooms by dividing the currently occupied beds and rooms by their respective total actual capacities, providing insights into the operational efficiency and demand for each shelter program.

T-test

```
T-test results for Emergency and Transitional programs - Beds:
  - t-statistic = 36.78483679745318
  - p-value = 7.273950955963105e-283
There is a significant difference in beds occupancy rates between Emergency and Transitional programs.

T-test results for Emergency and Transitional programs - Rooms:
  - t-statistic = 31.71080126309496
  - p-value = 4.425201973982312e-150
There is a significant difference in rooms occupancy rates between Emergency and Transitional programs.
```

Independent t-tests are conducted to compare the occupancy rates of beds and rooms between Emergency and Transitional shelter programs. The null hypothesis for each test assumes equal means between the two program types, while the alternative hypothesis suggests a significant difference. The research question revolves around whether there is a statistically significant difference in occupancy rates for beds and rooms between Emergency and Transitional shelter programs. The t-test results indicate extremely low p-values (7.27e-283 for beds and 4.43e-150 for rooms), well below the significance level of 0.05. With such low p-values, the null hypothesis is rejected in both cases. This implies a substantial and statistically significant difference in occupancy rates between Emergency and Transitional programs for both beds and rooms. In practical terms, these results suggest that the utilization of beds and rooms differs significantly between the two types of shelter programs, highlighting the need for tailored resource allocation and program management strategies based on the specific demands of each program model.

Conclusion

In summary, the in-depth examination of Toronto's homeless population and shelter system in 2021 provides crucial insights into occupancy trends, service types, and capacity utilization. Despite ongoing efforts, challenges persist, leading to individuals being turned away due to insufficient capacity. The analysis reveals an upward trajectory in shelter occupancy, reflecting the enduring impact of the pandemic. Varied patterns across shelter sectors emphasize the necessity for targeted interventions and flexible policy frameworks. The prevalence of emergency programs, particularly in high-demand sectors, underscores the immediate need for assistance. The capacity analysis highlights distinct usage patterns, and statistical tests confirm significant differences, emphasizing the call for tailored resource allocation and program strategies. Further research is vital for optimizing shelter distribution and addressing the unique needs of Toronto's homeless population.
