

Data Analysis of Toronto's Homeless Shelter Usage Trends

Guanxiao Chen (#1005717841)

Project Introduction

This project focuses on examining shelter usage trends by exploring a dataset that tracks the daily occupancy and capacity of Toronto shelters in 2021. The context underscores the challenges faced by the growing homeless population and the inadequacies in the shelter support system. The primary goal is to conduct quantitative analysis using t-test and exploratory data analysis.

Dataset Overview

The dataset, titled "INF2178_A1_data.xlsx," comprises two tabs. The first tab encompasses a detailed list of shelter organizations, programs, daily occupancy, etc. The second tab provides a comprehensive guide to the features present in the dataset.

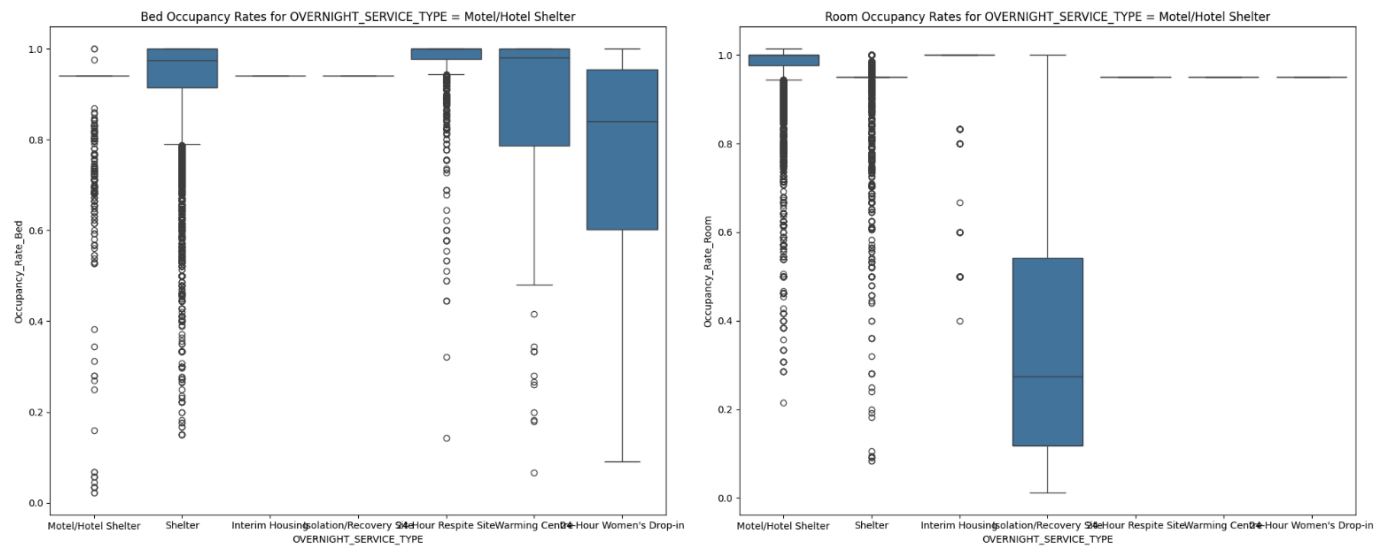
Data Cleaning and Preparation

To ensure the accuracy, consistency, and suitability of the dataset for data analysis, we should check for missing values, remove useless columns and rows, and fill in missing values. Here we drop useless columns: "ORGANIZATION_NAME," "PROGRAM_ID," and "PROGRAM_NAME.". Filling in the missing values of columns: "PROGRAM_MODEL," "OVERNIGHT_SERVICE_TYPE," and "PROGRAM_AREA" with the mode. Filling in the missing values of the numerical columns: "CAPACITY_ACTUAL_BED," "OCCUPIED_BEDS," and "CAPACITY_ACTUAL_ROOM" with the mean. After checking the cleaned dataframe, the dataset is ready to analyze.

Quantitative Analysis using t-test

To delve into more detailed comparisons, t-tests are employed. Here we first use an independent t-test method to assess the differences in shelter program occupancy rates based on the variable "OVERNIGHT_SERVICE_TYPE.". Two key continuous variables, "OCCUPIED_BEDS" and "OCCUPIED_ROOMS," are considered to calculate and compare the occupancy rates across different overnight service types. Independent t-tests are good to assess whether there are

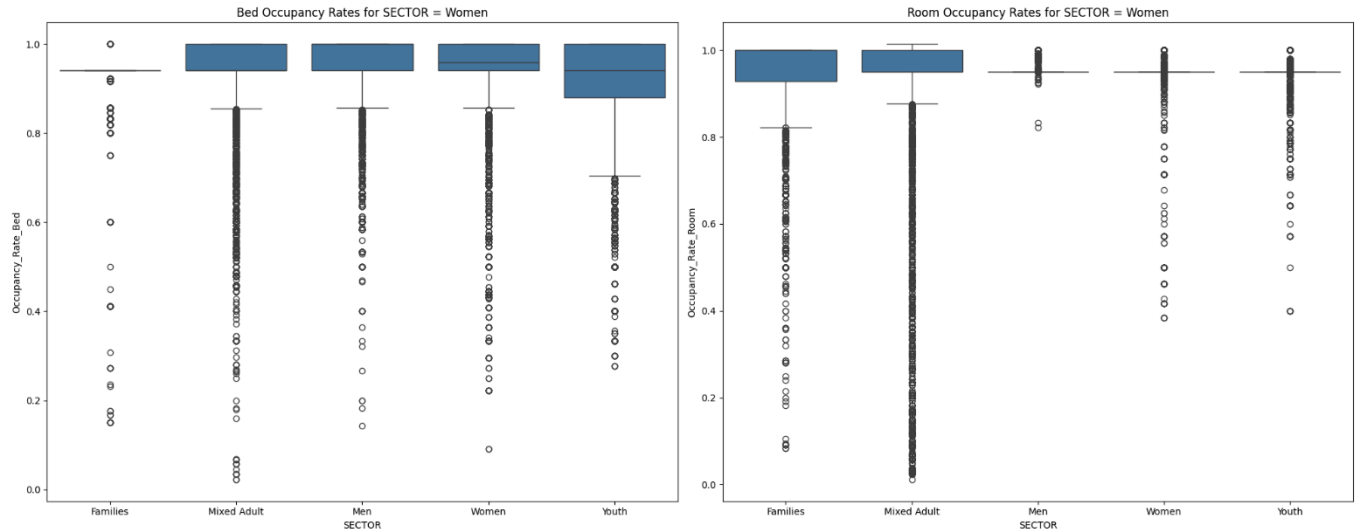
statistically significant differences in occupancy rates for beds and rooms across different overnight service types. To visualize the results, boxplots are created to visually compare the distribution of occupancy rates for beds and rooms across different overnight service types.



The t-test results indicate a significant difference in both bed and room occupancy rates for Motel/Hotel Shelter compared to other service types. The positive t-statistics (Bed: 7.48, Room: 39.23) and extremely low p-values (Bed: 7.68, Room: 0.0) strongly suggest that Motel/Hotel Shelter exhibits a substantially higher mean occupancy rate than the combined rates of other service types.

The t-test results for Shelters reveal noteworthy differences in occupancy rates. The negative t-statistic for bed occupancy (-12.20) and positive t-statistic for room occupancy (5.42) both point towards significant differences. The very low p-values (Bed: 3.48, Room: 5.98) provide strong evidence that the mean bed occupancy rate for Shelters is lower, while the mean room occupancy rate is higher compared to the combined rates of other service types.

In conclusion, Motel/Hotel Shelter experience significantly higher demand for both beds and rooms, while Shelters demonstrate a lower demand for beds but a higher demand for rooms. These findings emphasize the importance of tailoring resource allocation and support strategies to address the specific needs of different types of shelters in Toronto.



The second quantitative analysis also chooses to apply an independent t-test. Choosing the same continuous variables: “OCCUPIED_BEDS” and “OCCUPIED_ROOMS”, along with the categorical variable “SECTOR” instead. Still using boxplot to visualize the results.

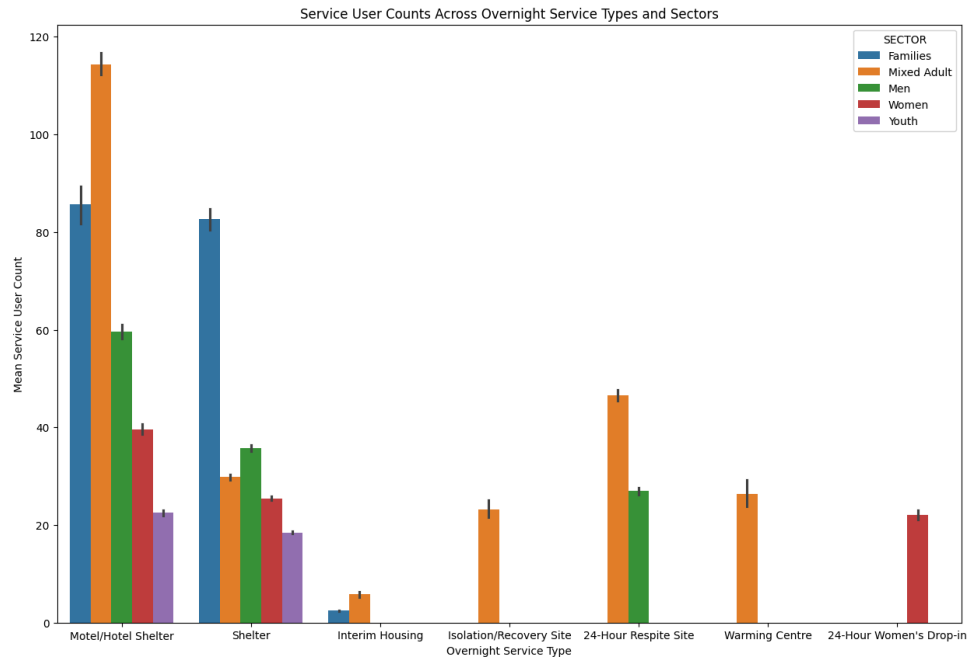
For sector “Mixed Adult”, the negative t-statistics for both bed and room occupancy rates (-5.04 and -23.54) accompanied by extremely low p-values (4.67 and 7.64) indicate a statistically significant lower mean occupancy compared to other sectors. This suggests that the “Mixed Adult” sector experiences distinct shelter utilization patterns, warranting focused attention and targeted interventions.

Similarly, in the case of sector “Men”, the positive t-statistics for bed and room occupancy rates (31.23 and 18.15) with very low p-values (4.27 and 2.31) highlight significantly higher mean occupancy rates. This indicates a substantial difference in shelter utilization patterns for men compared to the combined rates of other sectors.

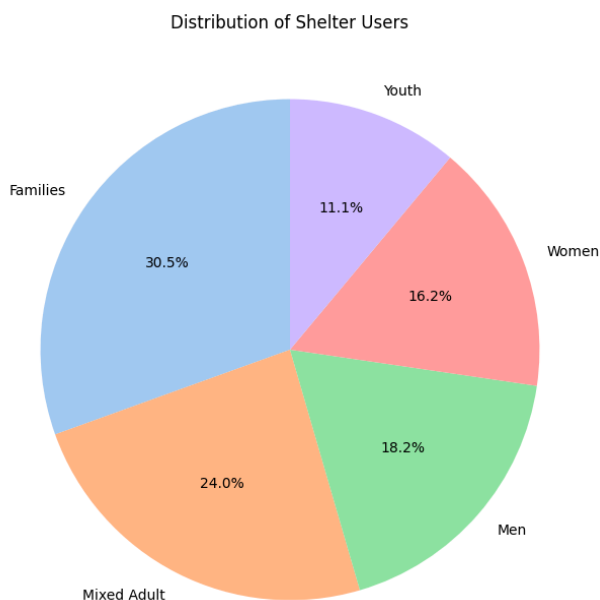
These findings underscore the need for tailored strategies and targeted support services to address the specific needs and challenges faced by these demographic groups in the context of shelter program utilization in Toronto.

Exploratory Data Analysis

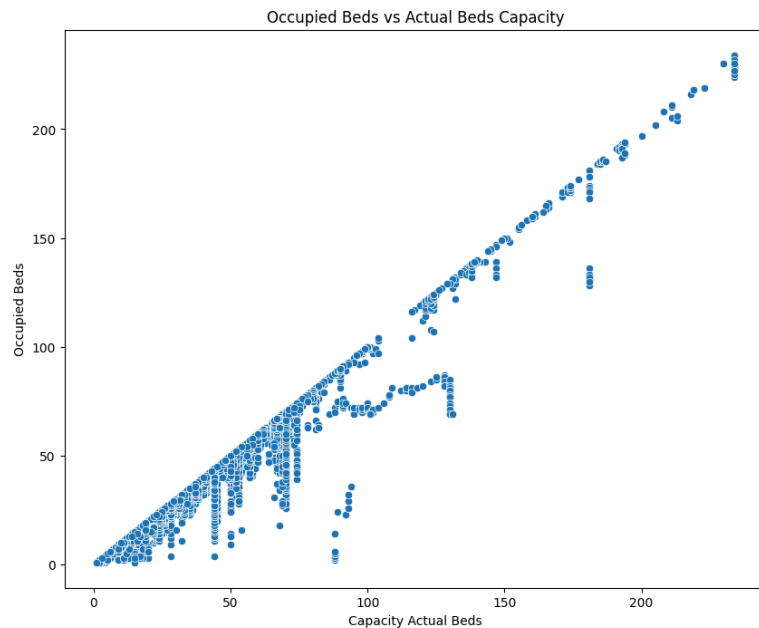
Next, we can generate several exploratory analyses to better visualize and understand the distribution of each variable.



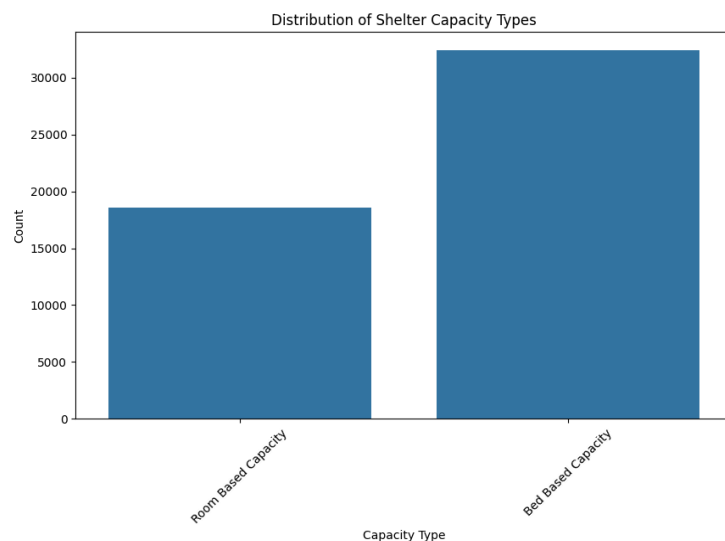
Firstly, to visualize and understand the distribution of service user counts across different overnight service types and sectors in Toronto's shelter system, we create a bar plot that illustrates the mean service user counts for each sector across various overnight service types. From the plot, we can see Motel/Hotel Shelter is the most popular program among all programs; all sectors love to choose this program, but Mixed Adult have the largest number of users. Shelter program is popular too, users also love to choose this program.

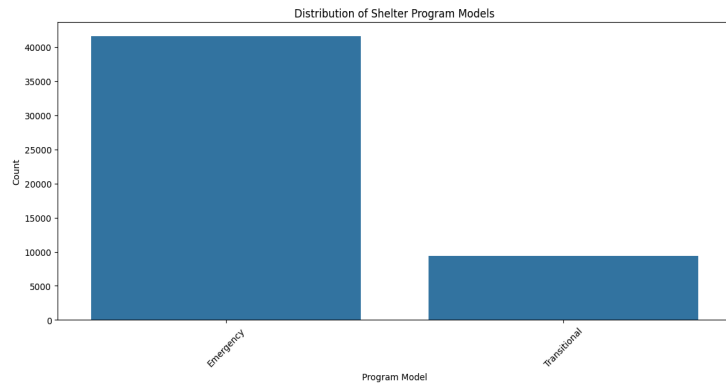


Secondly, to visualize and understand the distribution of shelter users across different sectors in Toronto's shelter system, we create pie charts to illustrate it. The pie chart shows Families account for the largest proportion of users, Youth is the least.

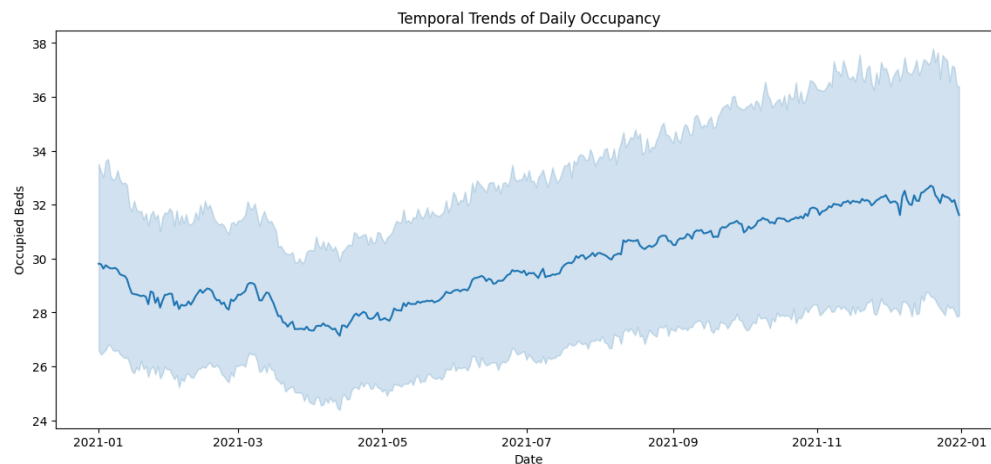


The scatter plot visually illustrates the relationship between the actual bed capacity and the occupied beds, which can show the efficiency with which bed capacities are utilized. It can be seen that the overall trend is proportional and have a positive slope which means the program and capacity is efficient.





Using count plots to see the distribution of shelter capacity types and program models, we can see Bed Based Capacity count greater than Roo Based Capacity and most program model is Emergency instead of Transitional.



This line plot visualizes the trends in daily occupancy over time. It can be seen that there is an overall upward trend in 2021 and a low point around April 2021. Overall, the number of occupied beds is increasing, and we can also see the demand for beds and the capacity of beds are increasing too.

Conclusion

The t-tests were instrumental in assessing differences in shelter program occupancy rates based on categorical variables. The exploratory analysis provides a foundation for understanding shelter usage dynamics, enabling managers to make informed decisions to address the evolving needs of Toronto's homeless population. The insights gained serve as a valuable starting point for more in-depth investigations to enhance the effectiveness of the shelter support system.