Shelter Data Analysis: Uncovering Patterns and Insights in terms of Demographic

Ang Zhao(1010171216)

Introduction

In the context of urban planning and social services, understanding the dynamics of shelter usage is crucial for addressing homelessness effectively. My analysis focused on a dataset comprising occupancy rates, service user counts, and demographic information for various shelters. Through exploratory data analysis (EDA), I aimed to uncover patterns, trends, and disparities within the shelter system to inform policy and program development.

Data Overview and Preparation

The dataset included daily records from Toronto's shelter system, capturing data such as occupancy date, organization name, program model (Emergency vs. Transitional), sector (Families, Youth, Men, Women), and service user counts. Initial examination revealed a comprehensive dataset requiring minimal cleaning, primarily involving the handling of missing values and standardizing category names for consistency.
Here's a deeper look into the specific actions taken and their significance:

Handling Missing Values

Missing data can present significant challenges in data analysis, potentially leading to biased estimates or misinterpretations. To mitigate these issues, I employed two distinct strategies based on the nature of the data in each column:
For categorical columns such as **PROGRAM_NAME**, **PROGRAM_MODEL**, **OVERNIGHT_SERVICE_TYPE**, and **PROGRAM_AREA**, missing values were replaced with the label 'Unknown'. For numerical columns related to bed and room capacity (**CAPACITY_ACTUAL_BED**, **OCCUPIED_BEDS**, **CAPACITY_ACTUAL_ROOM**, **OCCUPIED_ROOMS**), missing values were filled with 0.

Exploratory Data Analysis (EDA)

Distribution of Service User Counts and Occupancy Rates

My first step was to explore the distribution of **SERVICE_USER_COUNT** and **Occupancy_Rate**. Histograms revealed that service user counts varied widely across shelters, with a notable skew towards lower counts, indicating that many shelters operate under capacity most of the time. In contrast, occupancy rates were heavily skewed towards 100%, suggesting that when shelters are used, they tend to be near or at full capacity.

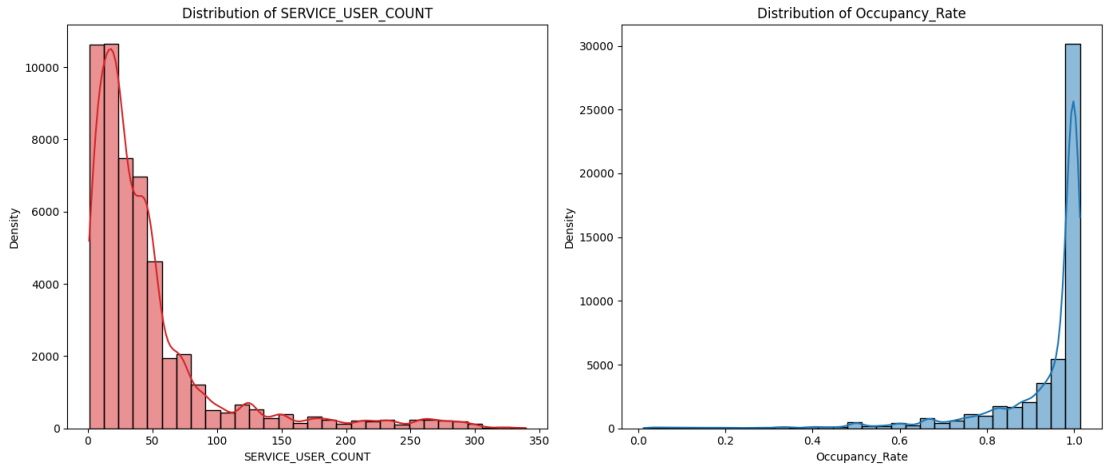

Figure-1 distribution of service user count and occupancy rate

Comparative Analysis by Demographic

Occupancy Rate Differences

The general overview of occupancy rate over sectors is in the figure 2.

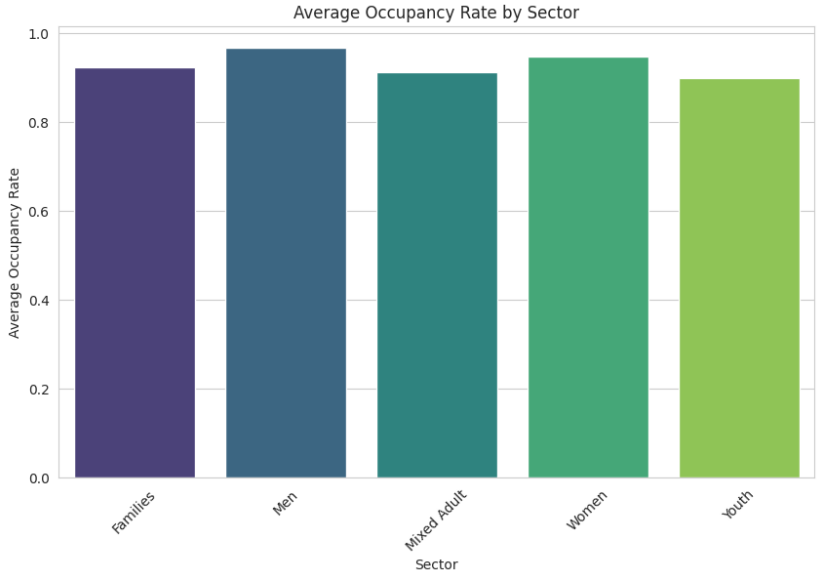

Figure-2 Occupancy rate across all sectors

Boxplots comparing **Occupancy_Rate** across different demographics (Men vs. Women, Families vs. Youth) highlighted significant variances. Men's shelters showed marginally higher occupancy rates than women's shelters, suggesting a higher or more consistent demand. Comparatively, family shelters exhibited significantly higher occupancy rates than youth shelters, indicating a pressing demand for family-oriented accommodations.

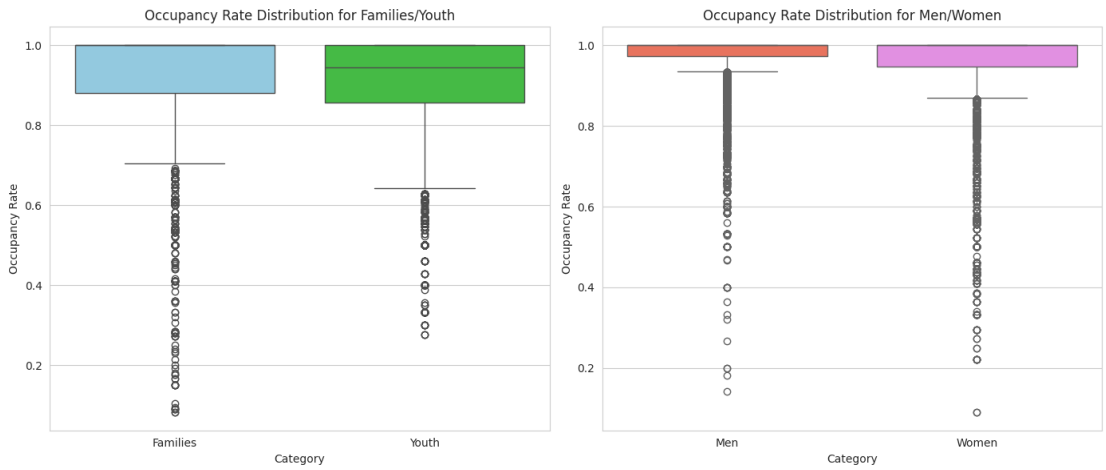

Figure-3 Boxplots of Occupancy Rate Comparisons

Service User Count Variations

The general overview of service user count over sectors in the figure 4.

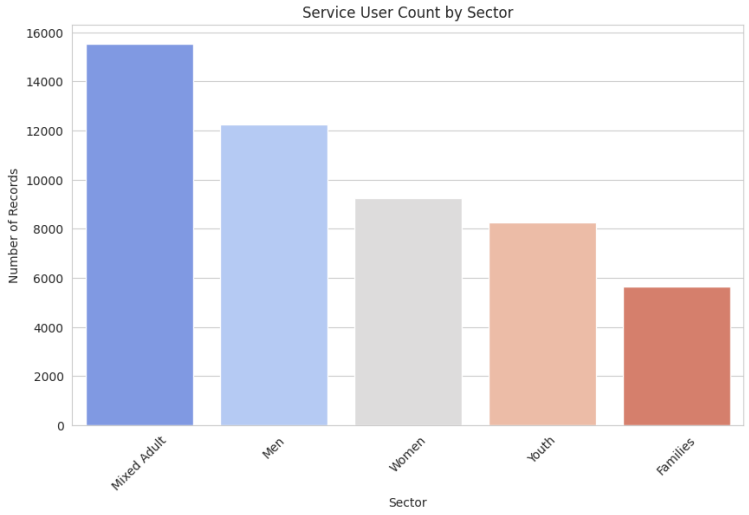

Figure- 4 chart of service user count across all sectors

Similarly, boxplots of **SERVICE_USER_COUNT** demonstrated that men's and family shelters not only had higher occupancy rates but also served more users on average than women's and youth shelters, respectively. This further underscores the unique demands placed on different shelter sectors.

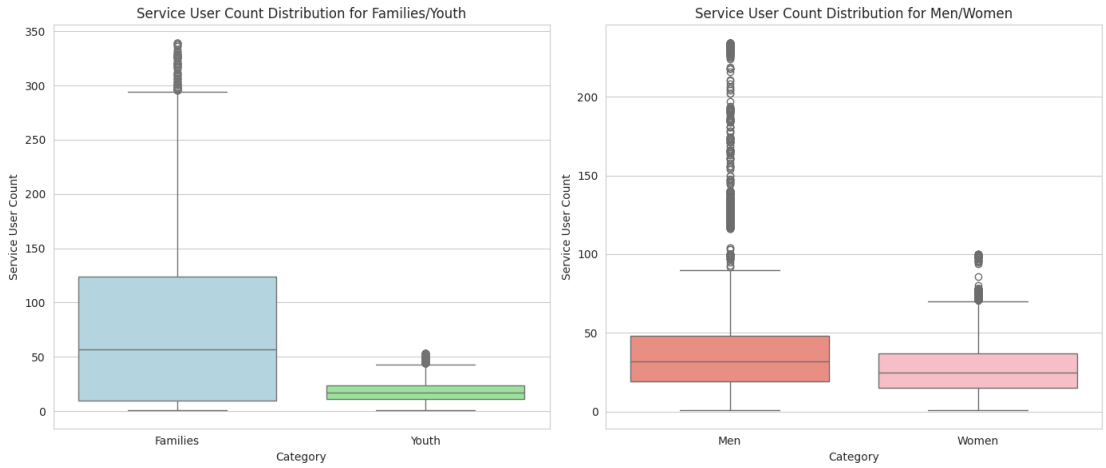

Figure-C Boxplots of Service User Count Comparisons

Statistical Analysis

T-tests were conducted to ascertain if the observed differences were statistically significant. The analysis confirmed significant differences in both occupancy rates and service user counts across all compared groups (Men vs. Women, Families vs. Youth), with p-values far below the 0.05 threshold, indicating that these differences were not due to chance.

Table 1: T-Test Results for Men vs. Women Shelters

Metric	Men's Shelters	Women's Shelters	T-Statistic	P-Value
Sample Size	12,241	9,258	-	-
Mean Occupancy Rate	0.966	0.947	14.518	1.90e-47
Mean Service User Count	39.87	28.66	30.504	1.05e-199

Table 2: T-Test Results for Family vs. Youth Shelters

Metric	Family Shelters	Youth Shelters	T-Statistic	P-Value
Sample Size	5,649	8,263	-	-
Mean Occupancy Rate	0.921	0.898	10.016	1.60e-23
Mean Service User Count	79.65	19.54	58.352	0.0

Insights and Implications

Interpretation:

Men vs. Women Shelters: The significantly higher mean occupancy rate and service user count in men's shelters as compared to women's shelters suggest a higher demand or utilization rate. The extremely low p-values indicate that these differences are statistically significant and unlikely due to chance.

Family vs. Youth Shelters: Family shelters exhibit notably higher mean occupancy rates and service user counts compared to youth shelters. The differences are statistically significant, as evidenced by the negligible p-values, underscoring the particularly acute

demand for family-oriented shelter services.

The statistical analysis underscores marked disparities in shelter utilization across different demographic groups, highlighting the critical need for targeted resource allocation and programmatic interventions.

Conclusions and Recommendations

The EDA provided valuable insights into the complexities of shelter usage and demand. To further refine these findings and develop actionable recommendations, I suggest:

- **Targeted Resource Allocation**: Prioritizing resources to sectors under the most strain, as identified by occupancy and service user counts, such as families.
- **Program Development**: Designing specialized programs to address the unique needs of the most affected demographics, such as men and families.
- **Further Research**: Conducting qualitative studies to understand the underlying causes of the observed trends and to explore the experiences of shelter users.

Reflection

This exploratory analysis has not only shed light on the operational dynamics of Toronto's shelter system but also highlighted the critical role of data in informing social policy and program development. The insights garnered underscore the necessity of a nuanced approach to addressing homelessness, one that considers the varied and specific needs of different demographic groups within the shelter system.