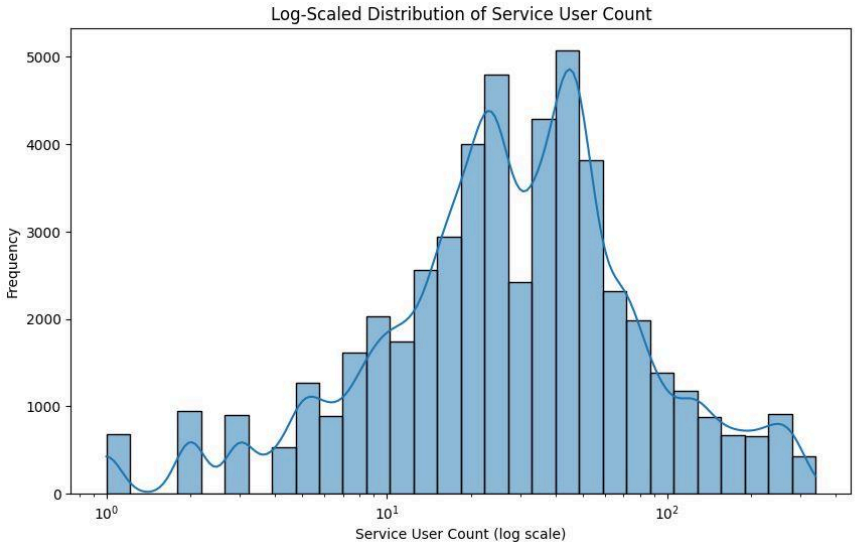
Peiwei Li
INF2178
Shion Guha
04/02/24

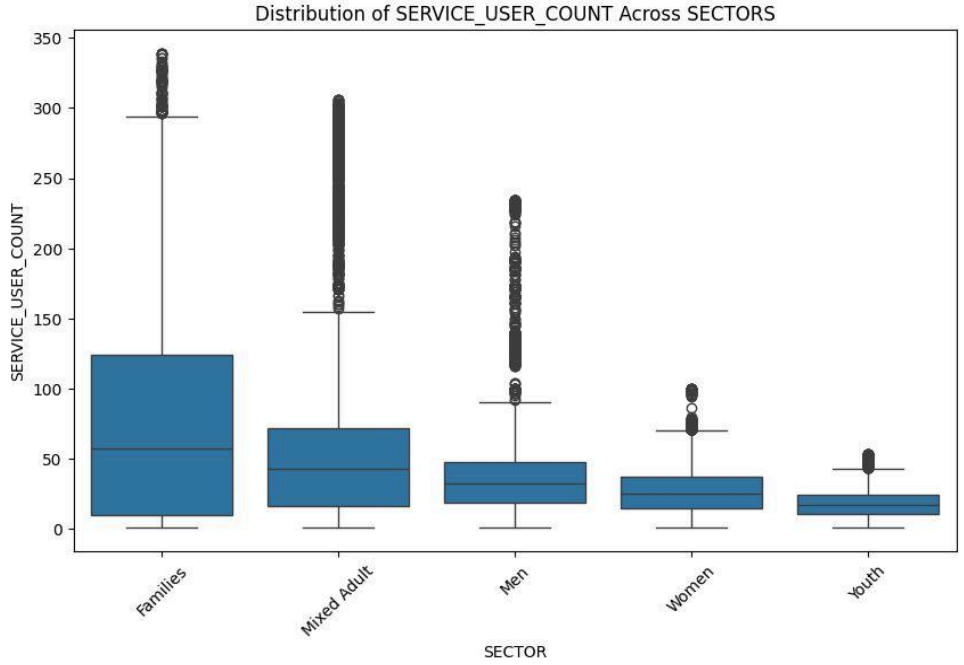
Narrative of Findings

The dataset is an in-depth record of the day-to-day operations of Toronto's system of homeless shelters. It lists a number of important metrics, such as occupancy information and the overall service uptake in great detail. Strategic planning and the optimization of social program support that meets both short-term and long-term housing demands,can gain great help from the insights gleaned from this data.Analyzing the patterns of the dataset is crucial for understanding shelter demand nuances, shaping housing and social welfare strategies, and pinpointing diverse population needs. This scrutiny enables stakeholders to grasp resource distribution, seasonal variances, and socio-economic developments' impact on vulnerable populations and underserved groups. Since determining which industries have a greater number of service users can aid in the effective distribution of resources. I set the research question to be "Is there a statistically significant difference in the average service user count between the 'Families' sector and the 'Mixed Adult' sector?"

The methodology employed revolves around using t-tests for statistical analysis to identify group differences within the dataset. Conducting a t-test to compare the mean SERVICE_USER_COUNT between the 'Families' and 'Mixed Adult' sectors is important for understanding specific demand within the shelter system. Apart from revealing statistical variations in the utilization of shelters, this analysis also sheds light on possible discrepancies in accessibility or necessity that might not be readily noticeable. Gathering these insights is critical for developing targeted strategies, optimizing resource distribution, and ensuring the shelter system meets the unique challenges of different demographic groups. This approach enhances the ability to design specific services, addressing the specific needs towards families and mixed adult groups, and improves the overall effectiveness of social support mechanisms. Additionally, to accommodate the broad range of data values, histograms with a log scale were utilized, enhancing interpretability and reducing skewness from significant numerical values. This way made it possible to conduct a thorough analysis of Toronto's shelter use trends, revealing unique trends and variations between various industries.



The exploratory data analysis with histogram revealed significant variations in the use of shelters, indicating seasonal variations in demand. Delving deeper into the data uncovered associations between shelter occupancy and variables such as economic indicators, seasonal variations, and policy influences. Additionally, the examination of outliers indicated instances of unforeseen crises, demanding immediate responses and adaptations in shelter management. These initial findings emphasize the intricate interplay of factors that shape shelter demand and the vital importance of maintaining a shelter system that can swiftly adapt to evolving circumstances. The line plot that is overlaid on the histogram represents either a fit to the distribution or a smoothed version of the histogram , which is a kernel density estimate. The histogram and line plot exhibit two prominent peaks, indicative of a bimodal distribution. This pattern implies that the services tend to aggregate around two separate prevalent user count values. Consequently, it could be inferred that the dataset encompasses two distinct categories or types of services, each with its own characteristic number of groups.



The distribution of service user counts among five different sectors, that are Families, Mixed Adult, Men, Women, and Youth, depicted in the boxplot. The Families sector exhibits a greater median number of service users accompanied by a wide interquartile range (IQR), suggesting a varied range of values and a few outliers that may indicate exceptionally high counts in certain instances. Conversely, the Mixed Adult sector has a moderately lower median and a narrower IQR but also presents several outliers, suggesting occasional high service user counts. The Men's sector stands out for having the lowest median user count and a very tight IQR, showing minimal variation in the data and no outliers. With a moderate IQR and a single outlier, the Women's sector has a median user count that is higher than Men's but lower than Families and Mixed Adult, indicating a generally consistent user count with a few exceptional values. Finally, there are no outliers in the Youth sector's median user count, which is in close alignment with Men's but has a slightly wider IQR, indicating slightly greater variability. In summary, the boxplot demonstrates that the Families and Mixed Adult sectors generally experience greater service user counts, accompanied by a broader range of values. In contrast, the sectors of Men and Youth are characterized by lower user counts, exhibiting more uniform data with minimal variation. The Women's sector occupies an intermediary position with user counts that are neither particularly high nor low but does display sporadic departures from the typical range.

In this case, the findings suggest a substantial distinction between the average values of the Families and Mixed Adult sectors, as evidenced by a t-statistic of approximately 15.12. The p-value, a critical component in hypothesis testing, plays a pivotal role. It reflects the probability of encountering a t-statistic as extreme as, or even more extreme than, the one computed from the sample data, under the assumption that the means of the two groups are identical. Notably, the p-value is exceptionally minuscule at 4.69 x 10^51, further emphasizing the strong evidence against the null hypothesis of equal means. In this instance, the remarkably small p-value of 4.69 x 10^51 is significantly less than any conventional threshold for statistical significance. According to this, it is extremely unlikely that the observed difference in mean SERVICE_USER_COUNT between the Families and Mixed Adult sectors happened by accident. Therefore, it is safe to say that the mean service user count differs between these two sectors in a statistically meaningful way.

The next steps involve evaluating the practical significance of the observed difference in mean service user counts between the Families and Mixed Adult sectors after statistical significance has been established. This entails determining the underlying causes, assessing the implications for policy, and interacting with stakeholders to develop focused interventions.More deeper analysis, interviews, and ongoing observation will improve the understanding of the dynamics of the shelter system for better solutions.