In the quest to understand the dynamics of shelter program utilization, an exploratory data analysis was embarked upon, utilizing a rich dataset reflective of the complex landscape that shelter services navigate. The data, a tableau of variables like service user count, bed and room capacity, and occupancy, as well as program models.

The initial phase of the analysis was marked by meticulous data cleaning. Columns peripheral to the analysis were filtered out, missing values were addressed with judicious care, and outliers—those statistical deviants—were reined in. This set the stage for a robust exploration of the data. Occupancy rates for beds and rooms were computed, transforming raw counts into insightful ratios that could speak to the efficiency and demand within shelter services.

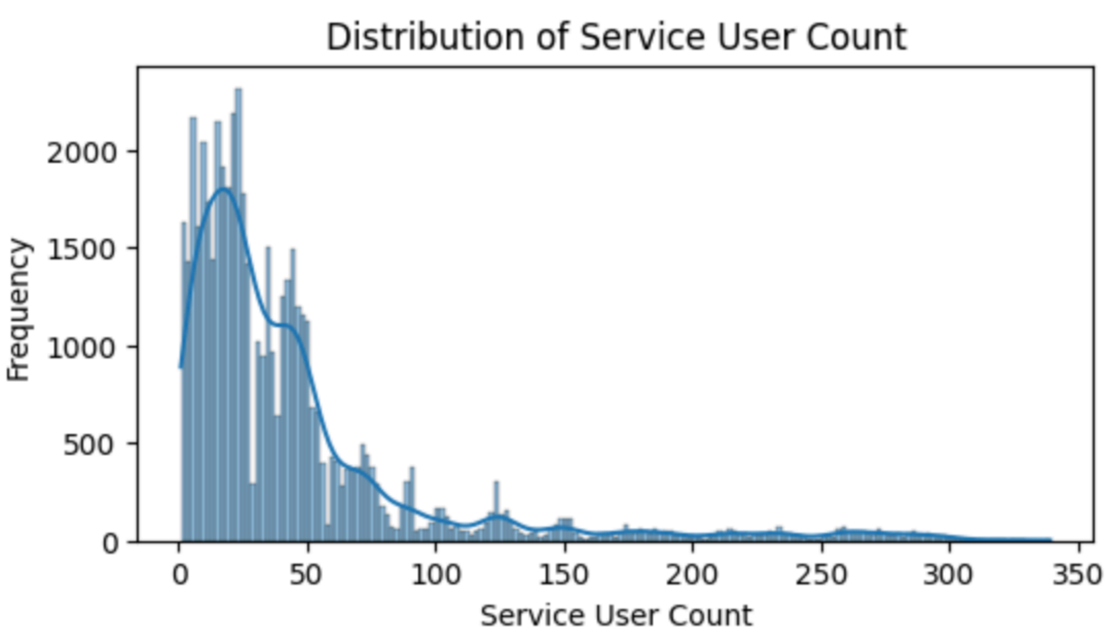
For the second step, I perform quantitative analysis using t-tests. Take the first t-test in the code as an example. The t-test conducted aims to statistically evaluate the difference in occupancy rates between two types of shelter program models: Emergency and Transitional. The occupancy rate, defined as the ratio of occupied rooms to the actual capacity of rooms, serves as a continuous variable that reflects the utilization efficiency of shelter programs. By comparing the mean occupancy rates of these two categories through a t-test, I can determine if the observed differences in occupancy rates are statistically significant or if they could have occurred by chance.

The output, p-value, helps to assess the significance of this difference. A p-value below the significant level (0.05) suggests that the difference in mean occupancy rates between

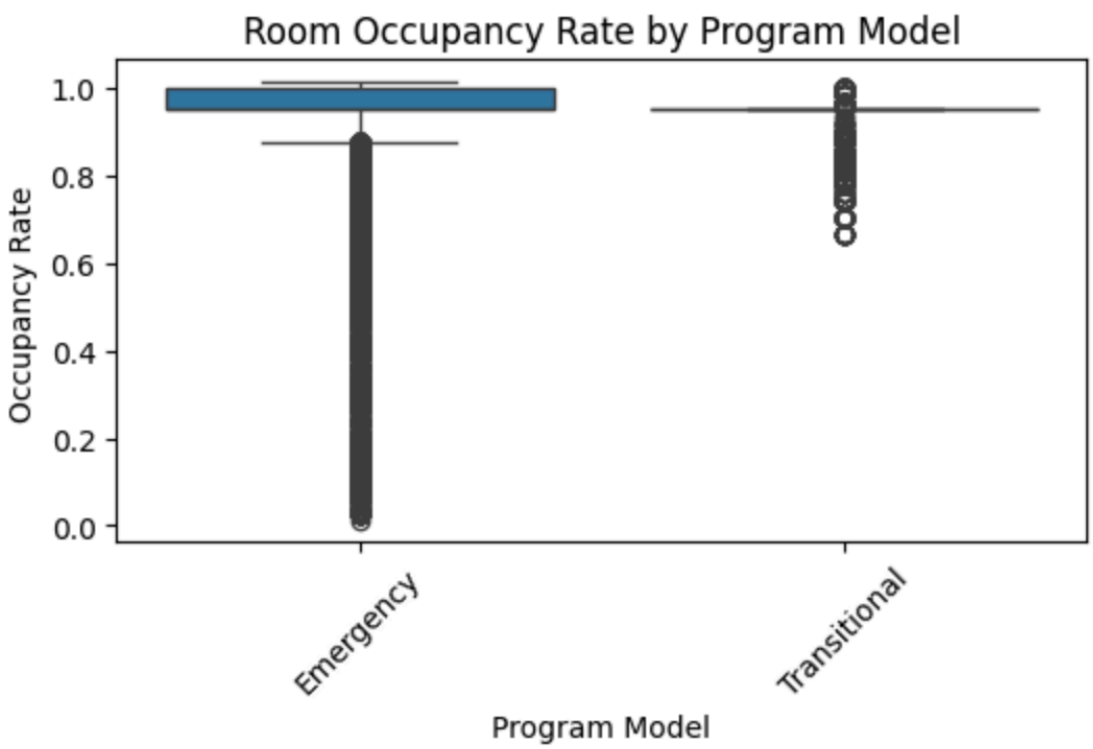
Emergency and Transitional shelter programs is statistically significant, meaning that I can reject the null hypothesis and then conclude that there is difference in the mean occupancy rates between these two types of shelter programs. There are more t-tests in the code performing different comparison and all of those p-values are smaller than 0.05.

For the next step, I use packages, such as seaborn and matplotlib, to perform exploratory data analysis. I firstly made two plots related to Service User Count and Room Occupancy Rate.

The first plot, a histogram overlaid with a kernel density estimate, illustrates the distribution of `Service User Count` across the shelters. The distribution is heavily right-skewed, meaning that most of the shelters serve a relatively small number of users, with the frequency dramatically decreasing as the service user count increases. This pattern suggests that while a few shelters handle large populations, a majority cater to fewer individuals. The concentration of data towards the lower end of the service user count could also indicate that a limited number of shelters are tasked with supporting a large portion of the service user population.

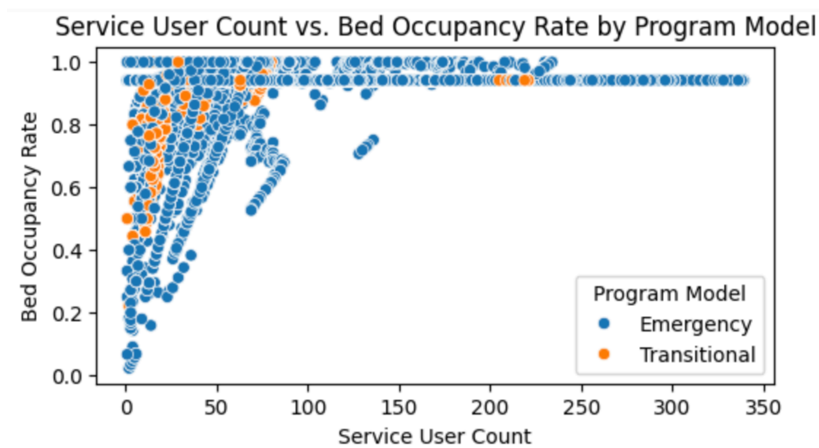


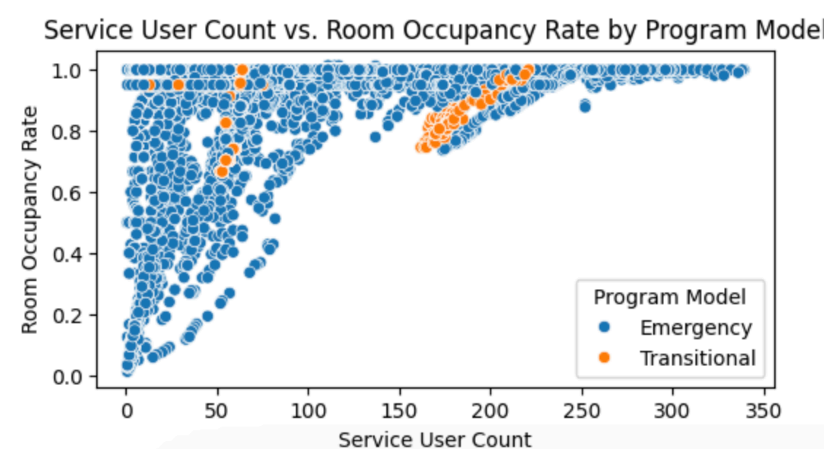
The second plot, a boxplot, provides a comparison of `Room Occupancy Rate` between two different `Program Models`: Emergency and Transitional. The Emergency program model shows a high median occupancy rate with a compact interquartile range, clustered towards full occupancy. This indicates that Emergency shelters often operate at or near full capacity, with less variability in occupancy rates. In contrast, the Transitional program model displays a wider interquartile range, with several outliers indicating instances of both low and high occupancy rates. This variability could reflect a more flexible usage of capacity or a diversity in program types and durations within the Transitional shelters.



Then I move to Find relationship between the SERVICE_USER_COUNT and occupancy rates for both beds and rooms. There are two scatterplots. In the first plot, we observe a dispersion of points that indicates the relationship between the number of service users and the bed occupancy rate. The data points are color-coded to distinguish between Emergency (blue) and Transitional (orange) program models. There seems to be a concentration of data points at lower service user counts for both program models, indicating that a majority of the data has lower service user counts and varied bed occupancy rates. For higher service user counts, the bed occupancy rate tends to cluster closer to 1.0, especially for the Emergency model, which might suggest that when service user counts increase, Emergency shelters tend to reach full bed capacity.

The second plot shows a similar comparison but for room occupancy rates. Again, there's a significant concentration of data points at lower service user counts, with room occupancy rates spread across the range for both program models. As service user counts increase, the room occupancy rate for Emergency shelters shows a trend towards higher occupancy,





almost reaching a rate of 1.0, implying that higher demand in service users correlates with higher room utilization. In contrast, the Transitional program model seems to show a slightly less dense clustering towards full occupancy, suggesting they may not reach full room capacity as consistently as Emergency shelters do at higher service user counts.

These analysis told a story of contrast and capacity, of demand and efficiency. Emergency shelters were the epitome of high utilization, their figures often cresting to full occupancy, a testament to the acute demand they address. Transitional shelters, with their fluctuating occupancy rates, posed questions about their operational models and resource alignment. The insights gleaned here beckoned towards a more granular investigation—a dive into the demography of service users, the temporal rhythms of demand, and the external forces that sway these numbers.