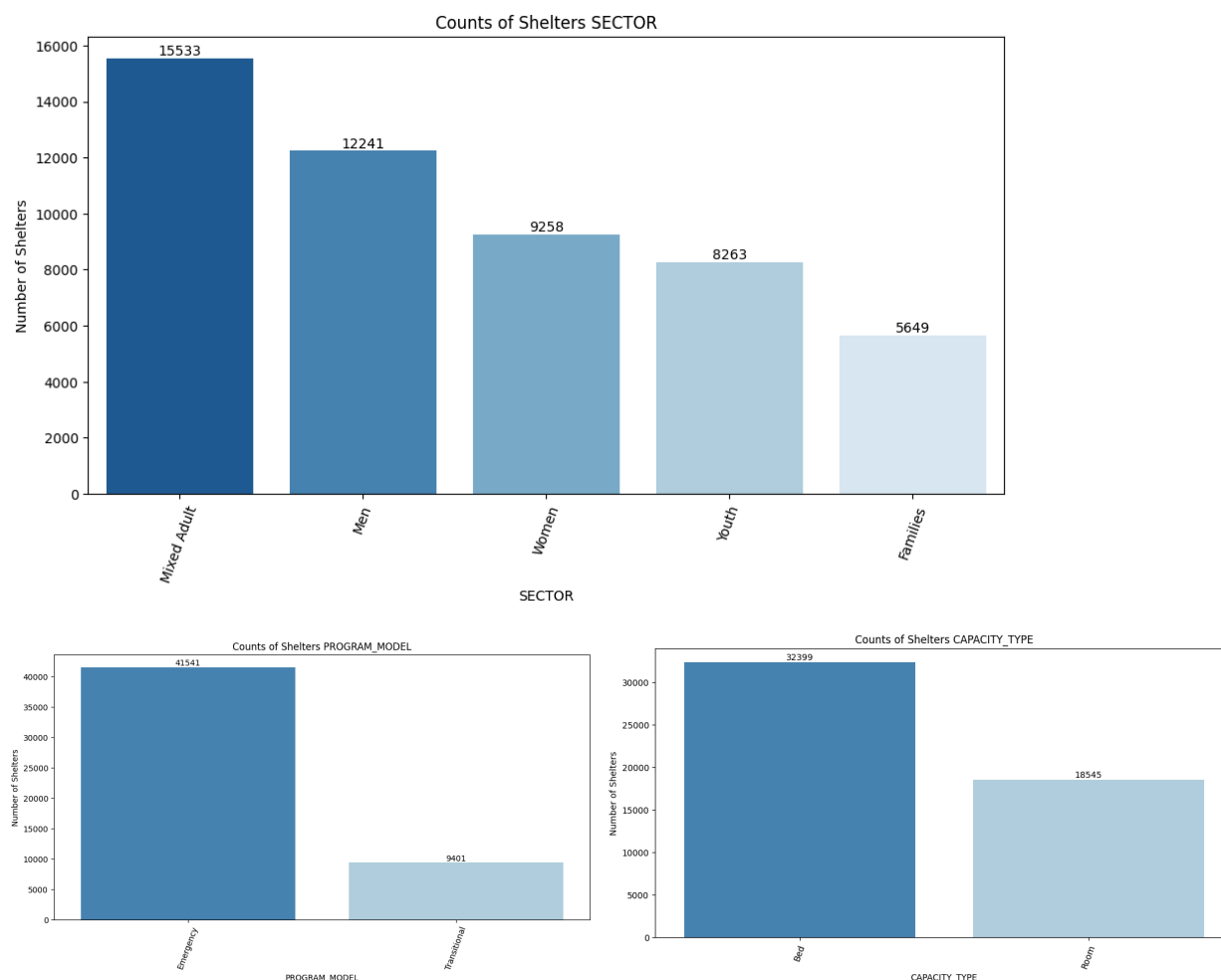


Research question: How can Toronto shelters plan and allocate resources more efficiently?

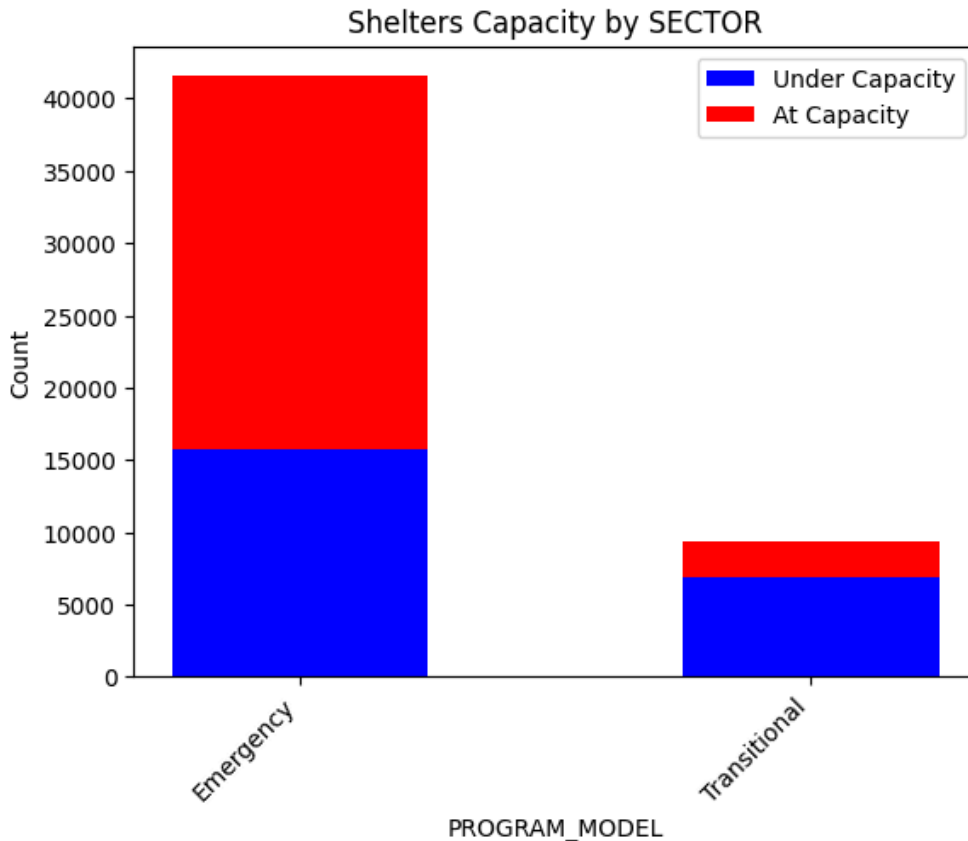
Through analyzing trends and patterns in Toronto shelter usage, I hope to obtain insights on how city shelters can better plan and prepare for changes in occupancy.

Upon examining the dataset, I see the dataset has some missing values. Upon further investigation, I found that the shelters in the dataset are categorized into room-based or bed-based occupancy capacity, where room-based shelters are missing statistics related to bed count and vice versa. To mitigate this, I created a new variable *Occupancy Rate*, calculated as occupied capacity divided by total capacity.

Next, I looked at the categorical variables in the dataset, namely '**SECTOR**', '**PROGRAM_MODEL**', '**OVERNIGHT_SERVICE_TYPE**', '**PROGRAM_AREA**', '**CAPACITY_TYPE**'. I visualized class distribution for each categorical variable, and decided to focus on '**SECTOR**', '**PROGRAM_MODEL**', and '**CAPACITY_TYPE**' for further analysis since they have relatively even distributions among classes and a low number of classes.

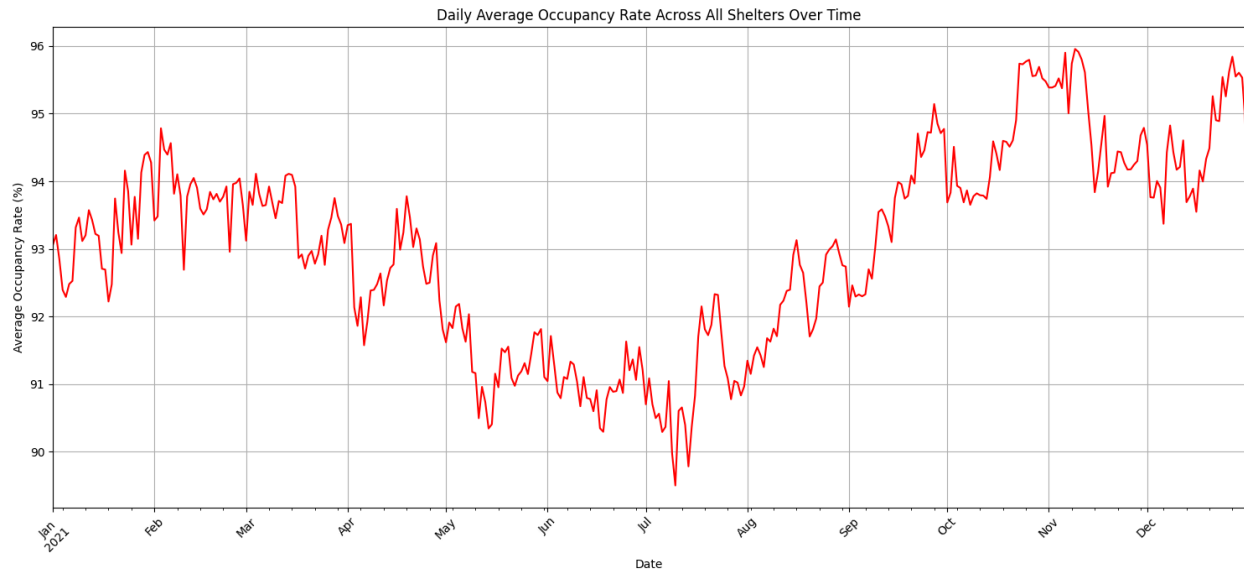


After looking at the distribution of shelter occupancy rate, I found over 50% of shelters in the dataset are at capacity. This confirmed the need to better allocate shelter resources, and inspired me to further investigate the distribution of at-capacity shelters. I found that proportions of shelters at capacity are consistent across capacity types (room-based v.s. bed based) and sectors. However, over 50% of emergency shelters are at capacity, significantly higher than the 25% among transitional shelters at capacity. This difference between at-capacity proportions for emergency v.s. transitional shelters appear to be consistent throughout the year, despite the seasonality in other aspects of shelter occupancy discovered later in the analysis.

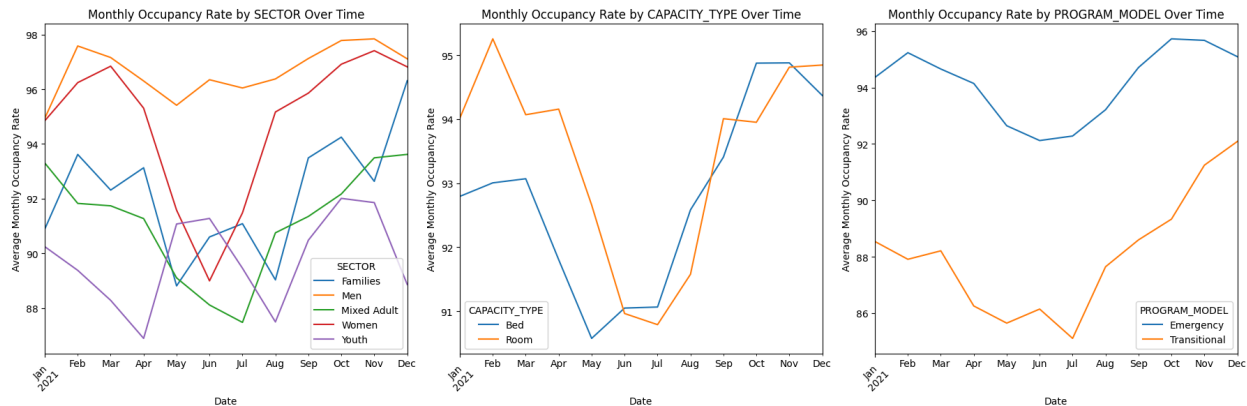


This might indicate a year round need for policy and infrastructure efforts to move occupants out of emergency shelters and into transitional shelters. However, are there underlying reasons that make demand for emergency shelters especially high? First, in an overview of shelter occupancy rate throughout 2021, there appears to be a seasonal trend, with occupancy rates dropping significantly around the mid-year mark before rising again in fall, suggesting a possible correlation with seasonal factors that influence homelessness, like weather conditions. There is an overall increasing trend in occupancy rates towards the end of the year. This could indicate that shelters are experiencing greater demand as the weather becomes colder or due to

holiday-related factors, calling for preparation for higher intake towards the end of the year.



My next question is: is this pattern consistent across types of shelters? To answer this question, I visualized monthly shelter occupancy rate by sector, capacity type, and program model to observe patterns and found relatively consistent seasonal trends with lower occupancy in the summer and high occupancy in the winter.



However, a few key insights to note are as below:

- Men's shelters have a consistently high occupancy rate throughout the year, suggesting a persistent need for men's accommodations.
- Women shelters show the most significant decrease in occupancy during the summer months. Mixed adult shelters have a similar pattern with a less significant change.
- Youth shelter occupancy rates show the most inconsistent seasonal pattern when compared to the rest of the shelters, with decreased occupancy in Jan-April and June-August, and a rise in occupancy between April-May and August-October.
- Room based shelters and bed based shelters show similar trends throughout the year with a significant decrease during summer months and significant increase for winter months.

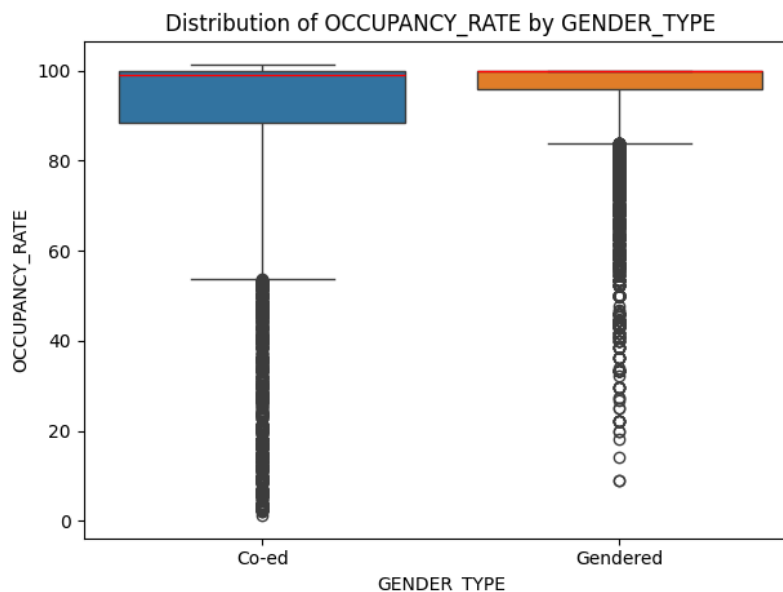
However, room based shelters are consistently more utilized than bed based shelters for the first half of the year.

- Emergency shelters have significantly higher occupancy rates than transitional shelters throughout the year, which confirms increasing needs of emergency shelters.

To investigate if these differences are statistically significant, I decided to perform t-tests to compare occupancy rates between groups.

First, I performed a Welch's t-test on overall mean occupancy rate for room-based shelters and bed-based shelters, which found significant difference in average occupancy rates between room-based shelters. However, since visualizations show that occupancy rate distributions for both groups are not normal, I decided to draw random samples from both groups. The means of these samples form a normal distribution according to Central Limit Theorem, and a t-test of these means shows that there is no significant difference between two group means. Further analysis of monthly average occupancy rate, combined with random sampling method, shows a significant difference between occupancy rates for room-base vs bed-based shelters in the early months of the year. However, more data from following years is needed to further investigate this trend.

Next, I performed the same t-test on gender-based shelters v.s. Co-ed shelters. The test result shows a significant difference in overall occupancy rates between gender-based and co-ed shelters. This combined with the box plot confirms a higher average occupancy rate for co-ed shelters compared to gender-based shelters.



However, t-tests of monthly occupancy rate between the two groups revealed that co-ed shelters have significantly higher monthly occupancy rate than gender-based shelters for only 4 months of 2021 and significantly lower monthly occupancy rate for 5 months of 2021. What potentially led to the overall lower occupancy rate in gender-based shelters was the acute decrease in female

shelter occupancy during certain months. Further visualizations show a particularly steep decrease in room-based and transitional women's shelters in the summer months, and a steady increase towards pre-summer occupancy levels right after. These seasonal fluctuations are not observed in men's shelter occupancy, which points to the possibility of certain factors that are impacting the homeless women population. Further research on potential causes such as weather, job availability, and policy changes. It would also be informative to study the trends of women shelter occupancy rate spanning multiple years.

Given the distribution of occupancy is extremely skewed, a potential next step would be to apply transformation techniques and outlier detection techniques to better understand the structure of the dataset. Time series technique could be applied to the data for more precise trend analysis. Due to the complexity of Toronto's homelessness, introducing data on other socioeconomic factors, such as crime rate and job data, might uncover more dynamic ways Toronto shelters can plan and allocate shelter resources.