Kaiwen Suo 1005764816

When I first came to Toronto, I had a lot of doubts about the homeless on the streets. Why are there so many homeless on the streets when there are so many shelters on offer? This time, by analyzing the use of recent shelters, it gave me a new understanding of this phenomenon. In this report I will describe how I processed the data and analyze some graphs.

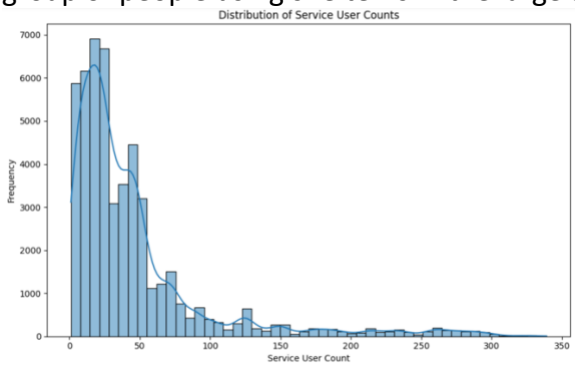
First, I analyzed the structure of this dataset. CAPACITY_TYPE, PROGRAM_MODEL, SECTOR are categorical data and SERVICE_USER_COUNT, CAPACITY_ACTUAL_BED, OCCUPIED_BEDS, CAPACITY_ACTUAL_ROOM, OCCUPIED_ROOMS are continuous data. Some columns include null data which need to be clean for preparation.

I calculate the occupied rate for both beds and rooms to generate continuous data for further T-test analyzing.

	OCCUPANCY_RATE_BED	OCCUPANCY_RATE_ROOM
count	32399.000000	18545.000000
mean	0.927885	0.934087
std	0.122562	0.163241
min	0.022727	0.012048
25%	0.900000	0.958333
50%	1.000000	1.000000
75%	1.000000	1.000000
max	1.000000	1.014085

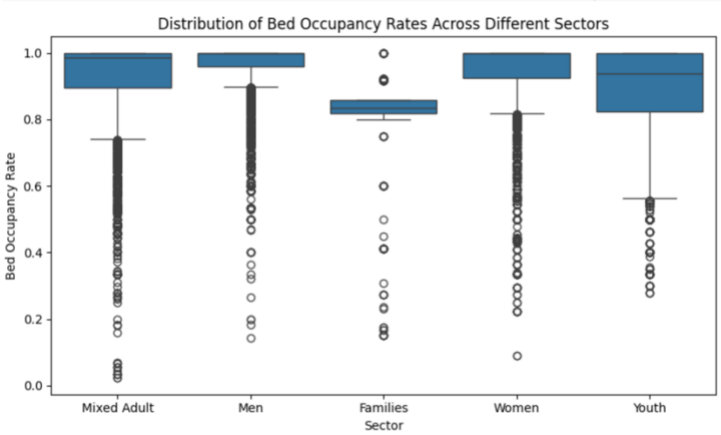
By displaying several simple EDA results, I found more occupancy rate on room using of shelters. Room using have the minimum using rate among whole year.

After conducting the simple EDA, I focused on Service User counts. I create a histogram to analyze. The distribution of service user counts is right-skewed, meaning there are more lower values in this data. This suggests that most of shelters doesn't have large group of people occupied. The peak of the distribution is in the left of the histogram, which indicate the small group of people using shelter is in the large amount.

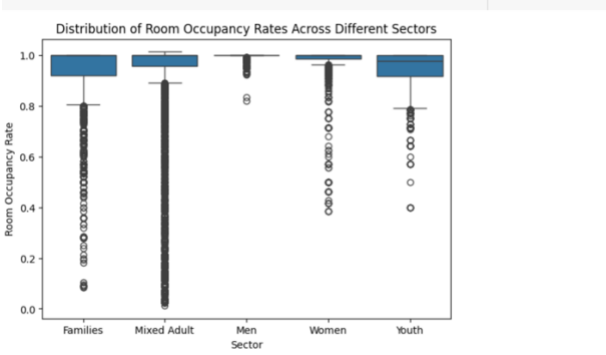


Kaiwen Suo 1005764816

I will list part of boxplots I made of bed and room occupancy rate across different categories.

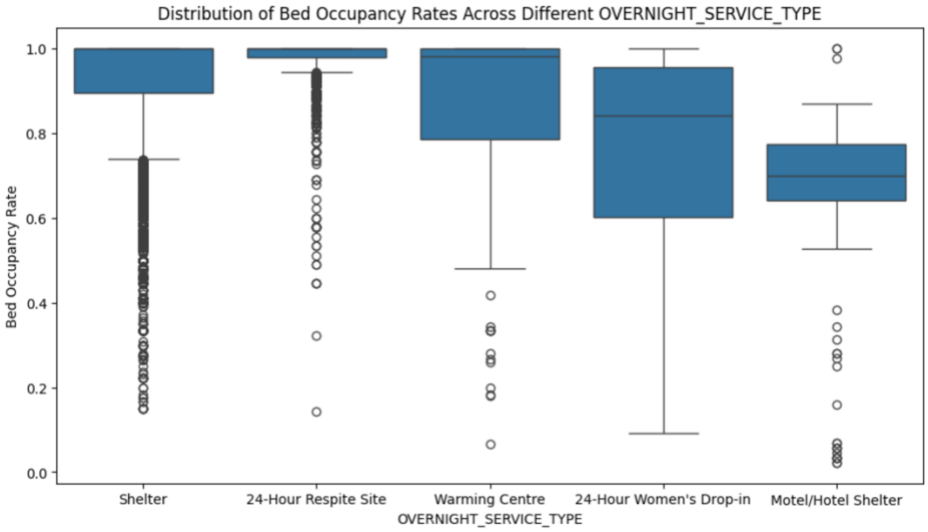


The boxplot for the Mixed Adult sector shows an extensive range in bed occupancy rates, with a significant number of outliers suggesting instances of very low occupancy. The median occupancy rate is approximately 0.8, indicating that the facilities were at or near maximum capacity for a significant portion of the time. If there are outliers below the bottom quartile, it may suggest sporadic underutilization or an uneven distribution of service consumers among facilities.

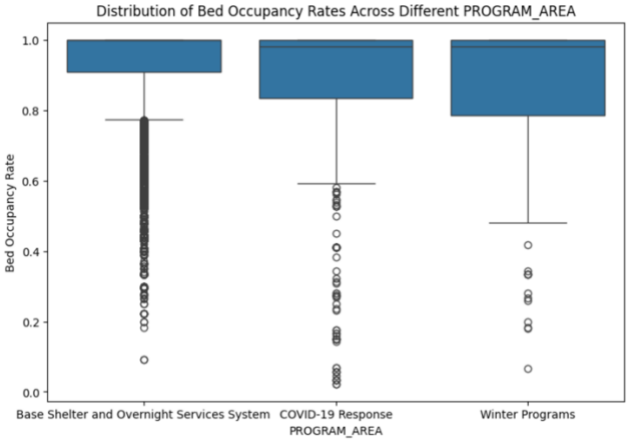


The bed occupancy rates in the men's sector are more constant, as indicated by a narrower interquartile range. The median bed occupancy rate is below 0.6. The presence of several outliers indicates sporadic occurrences of exceptionally high and meager occupancy rates which may indicate shifting demand or capacity-related challenges.

Kaiwen Suo 1005764816

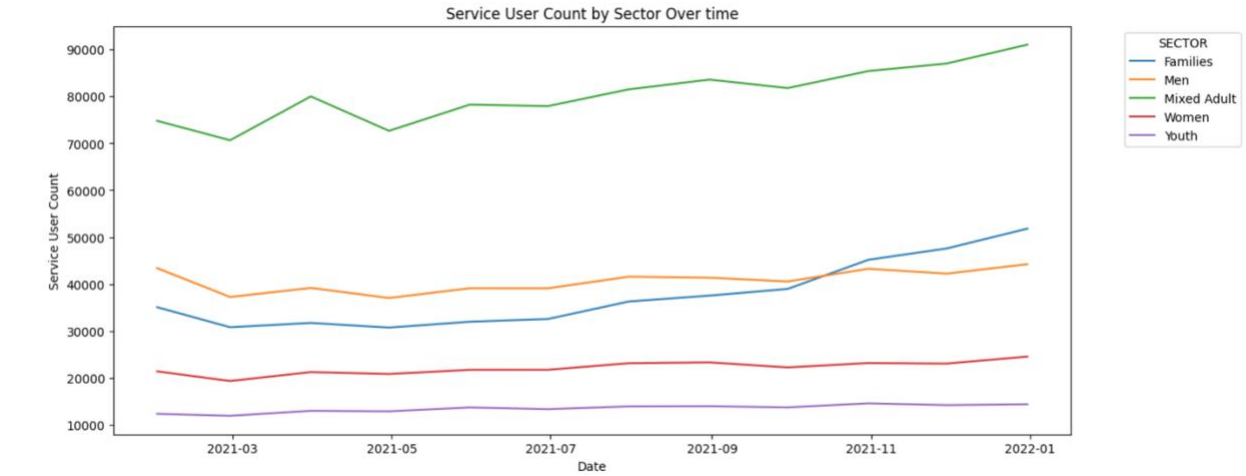


The boxplot displays varied occupancy patterns among different service categories. Shelters and warming centers typically have a consistently high number of people staying in them, with warming centers having a median occupancy rate at maximum capacity and higher fluctuation in numbers. Motel/Hotel shelters have a lower median occupancy rate and a narrower interquartile range, suggesting that their utilization is more constant but generally lower than other services.

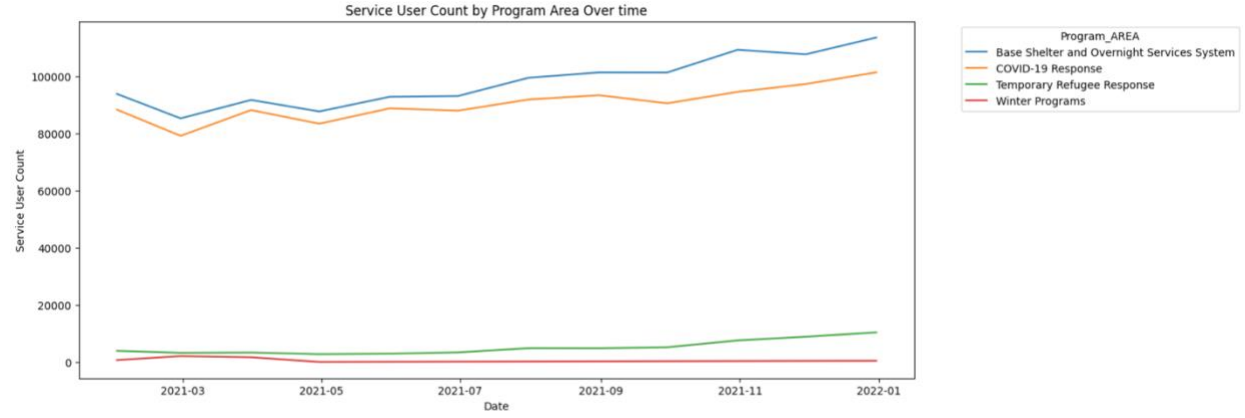


The occupancy rates in different program areas demonstrate a consistent pattern of high utilization, with COVID-19 response facilities frequently operating at maximum capacity, as evidenced by the median value close to or equal to 1. The Base Shelter and Overnight Services System consistently maintains high occupancy rates. However, there may be some fluctuations and infrequent instances of underutilization, as seen by outliers.

Also, I created several graphs about occupancy rate across different categories over time. I will list part of graphs below.



This graph illustrates the utilization of services in five sectors over time. The 'Families' category has the highest number of users, consistently around 80,000. The 'Mixed Adult' category is steadily increasing and is expected to reach over 80,000 by January 2022. The sectors categorized as 'Men' and 'Women' demonstrate a higher level of stability, with approximately 60,000 and 40,000 counts, respectively. The 'Youth' sector has the lowest number, maintaining below 20,000. The data indicates a greater need for services among families, with a consistent increase in the utilization of services by mixed individuals over the years.



This graph depicts the fluctuation in service users across four distinct program areas from January 2021 to December 2021. The 'Base Shelter and Overnight Services System' has the highest number of users, gradually growing over time, topping 80,000 by January 2022. The 'COVID-19 Response' shows a comparable rise, with a smaller user base, reaching its highest point at approximately 60,000. The 'Winter Programs show minimal fluctuation and consistently sustain a user count that is below 20,000. The 'Temporary Refugee Response' first experienced a modest level of activity but had a substantial increase in mid-2021. It stabilizes at a level somewhat below the 'COVID-19 Response'.

Then, I conducted two Two-sample T-tests of bed-occupied rate and room-occupied rate across different program models. The Two-sample T-test for bed-occupied rate result has t-statistic = 38.780694714817365 and p_value = 0. This indicates a statistically significant disparity in the bed occupancy rates between the Emergency and Transitional program models. The significant t-statistic suggests a significant disparity in the means between the two program models.

The Two-sample T-test for room occupied rate result has t-statistic = 18.903262158430557 and p_value = p-value = 5.923255977527666e-79. This also signifies a statistically significant disparity in the room occupancy rates between the Emergency and Transitional program models. The notable outcome suggests that the average disparity is not attributable to random probability.

After this analysis, I have a new understanding of the Canadian Shelter program and a cursory examination of why so many homeless are on the streets despite providing enough shelters. My research determined that existing shelters should be provided or adapted to maximize their use in appropriate seasons and special public health situations.