Danika Mariam (1004014880)
INF2178 – A1 Writeup

Examining the Dataset

The dataset shows information about shelter services in the city of Toronto. After examining the dataset, the variables I chose to look at were the following: 'CAPACITY_TYPE', 'PROGRAM_MODEL', 'SERVICE_USER_COUNT', 'CAPACITY_ACTUAL_BED', 'OCCUPIED_BEDS', 'CAPACITY_ACTUAL_ROOM', and 'OCCUPIED_ROOMS'.

Quantitative Analysis

First, I performed an independent t-test to compare the occupancy rates between emergency and transitional shelter programs within different sectors. The null hypothesis is that there is no significant difference between emergency and transitional shelter programs within different sectors. Since the p-values for each sector's t-test were small enough to be statistically significant, we can reject the null hypothesis. This indicates there is a significant difference in the occupancy rates between emergency and transitional shelter programs in the given sector. This shows us that there is a difference in how these two types of shelter programs are utilized in each sector.

Next, I performed independent t-tests to compare the occupancy rates of different shelter types against the overall mean occupancy rate. For the shelter type, 'Shelter' the p-value was not significant, thus we accept the null hypothesis that there is no significant difference in the occupancy rates between the specific shelter type and the overall mean occupancy rate. The p-values were significant for the following shelter types: 24-Hour Respite Site, Warming Centre, 24-Hour Women's Drop-in, Motel/Hotel Shelter. This indicates that the shelter type has a unique pattern of occupancy compared to the average, and this difference is not likely due to random chance.

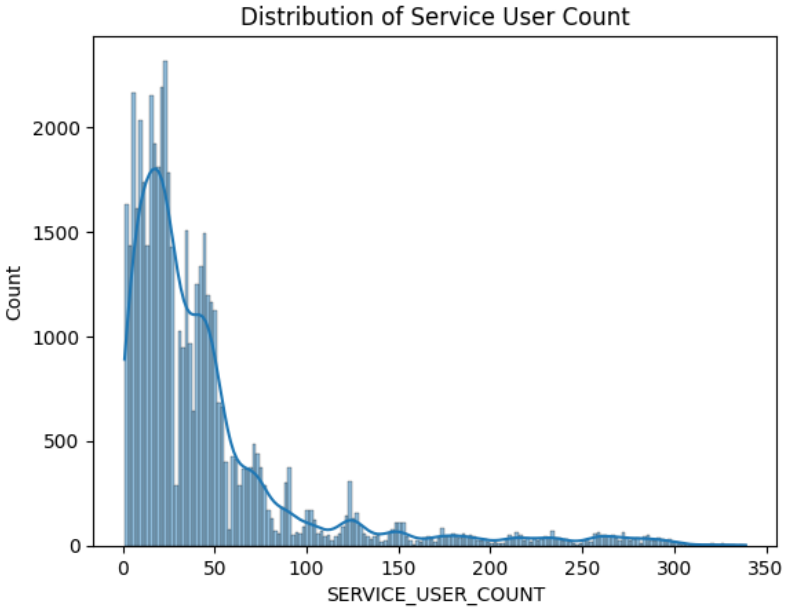
Exploratory Data Analysis

For the exploratory data analysis, I started by looking at the descriptive statistics for some chosen columns in the dataset. This is classified at univariate non-graphical exploratory data analysis. I found the count, mean, standard deviation, minimum and maximum values, as well as the quartiles (25%, 50%, 75%). These values can be found in the table below. There are an average of 46 unique users per program. The average capacity is about 32 beds per program, with around 29 being used, suggesting most beds are occupied with some consistently available. On average, there are about 55 rooms available per program, with around 53 typically in use, indicating a high room occupancy rate. This suggests that there is a high demand of shelter services which matches or exceeds those fulfilled by the current shelter system.

	PROGRAM_ID	SERVICE_USER_COUNT	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS
count	50944.000000	50944.000000	32399.000000	32399.000000
mean	13986.125844	45.727171	31.627149	29.780271
std	1705.288632	53.326049	27.127682	26.379416
min	11791.000000	1.000000	1.000000	1.000000
25%	12233.000000	15.000000	15.000000	14.000000
50%	14251.000000	28.000000	25.000000	23.000000
75%	15651.000000	51.000000	43.000000	41.000000
max	16631.000000	339.000000	234.000000	234.000000

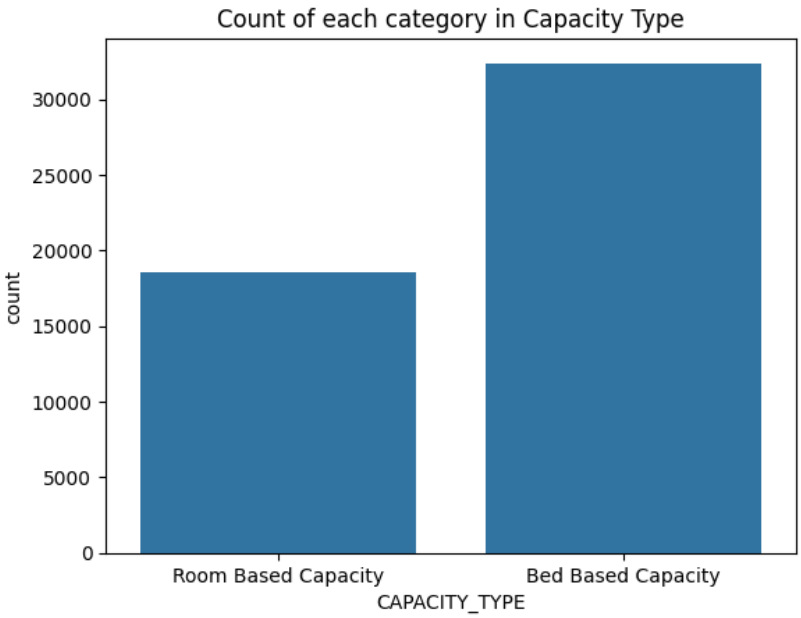
	CAPACITY_ACTUAL_ROOM	OCCUPIED_ROOMS
count	18545.000000	18545.000000
mean	55.549259	52.798598
std	59.448805	58.792954
min	1.000000	1.000000
25%	19.000000	16.000000
50%	35.000000	34.000000
75%	68.000000	66.000000
max	268.000000	268.000000

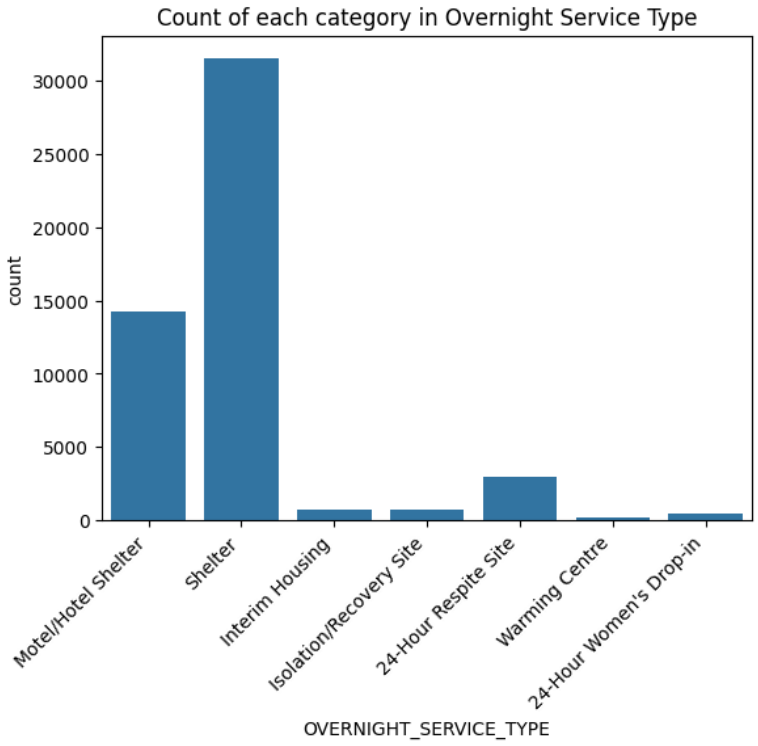
After finding these values, I continued the exploratory data analysis by using visualization tools in Python. I plotted a histogram of the 'SERVICE_USER_COUNT' which shows the distribution of the number of service users across various shelter programs. Most programs have a low user count, with the highest frequency of programs serving between 1 to 25 users. The number of programs decreases as the user count increases, which is evident from the tail of the graph extending to the right. This indicates that while a majority of programs serve a small number of users, there are a few programs that serve a much larger population, up to 350 users. However, these are exceptions rather than the norm. This distribution suggests that the shelter system is made up of many small programs and a few very large ones.



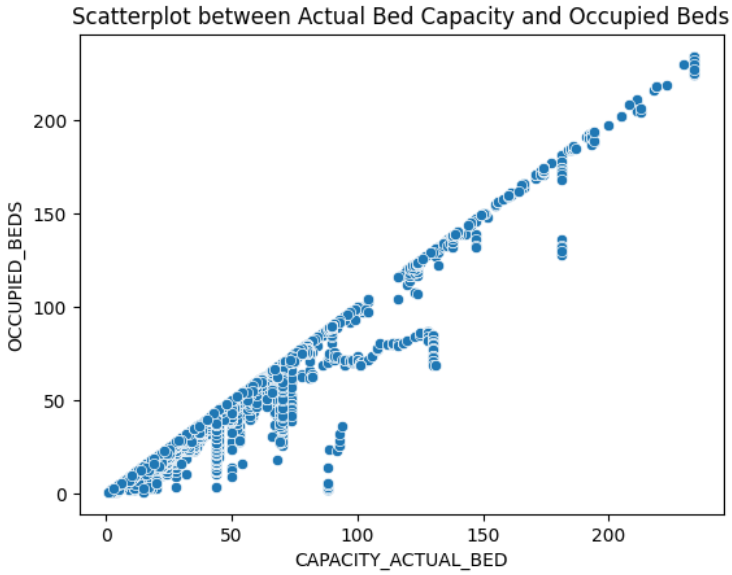
I created bar plots of the categorical variables, 'CAPACITY_TYPE' and 'OVERNIGHT_SERVICE_TYPE'. For CAPACITY_TYPE', we can see that there are many more beds than rooms in the shelters, meaning that many rooms contain more than one bed. For the

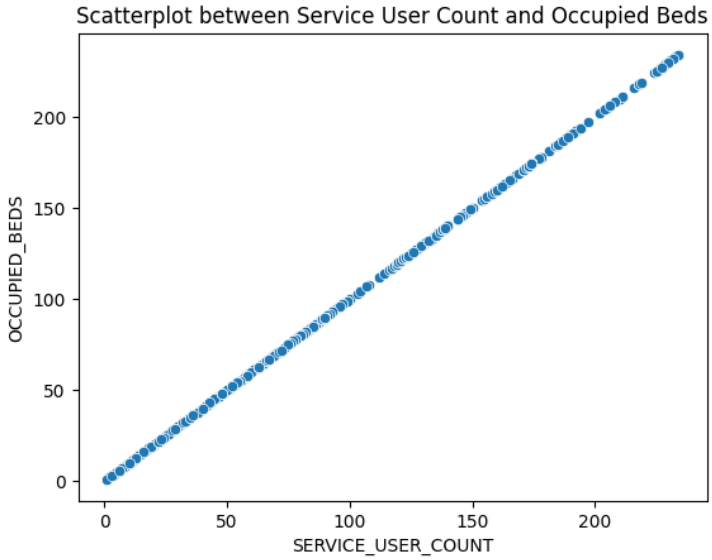
‘OVERNIGHT_SERVICE_TYPE’, we can see that ‘Shelter’ category contains the highest number of programs. The next highest frequency is that of the ‘Motel/Hotel Shelter’ programs. There are much fewer number of programs within the categories of ‘Interim Housing’, ‘Isolation/Recovery Site’, ’24-Hour Respite Site’, ‘Warming Centre’, ’24-Hour Women’s Drop-in’. The category ‘Warming Centre’ has the lowest frequency.





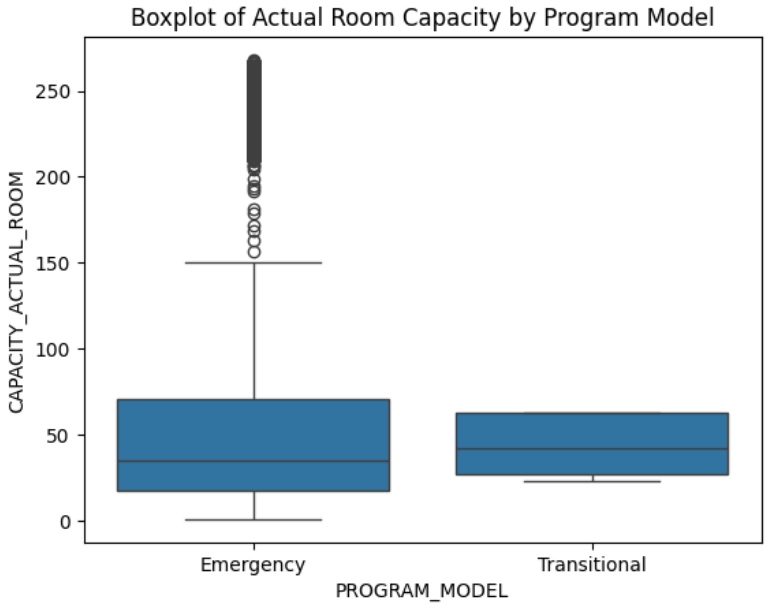
I then created scatterplots between ‘Actual Bed Capacity’ and ‘Occupied Beds’ as well as between ‘Service User Count’ and ‘Occupied Beds’. It looks like both of these plots are linear. For the scatterplot between ‘Actual Bed Capacity’ and ‘Occupied Beds’, the distibution appears to be mostly linear, with some points on the plot falling below the main line. This indicates that most of the programs have are fully occupied, with some shelters still having space for more users. For the scatterplot between ‘Service User Count’ and ‘Occupied Beds’, this line is fully linear because the number of people using shelter services is equivalent to the number or beds that are occupied.



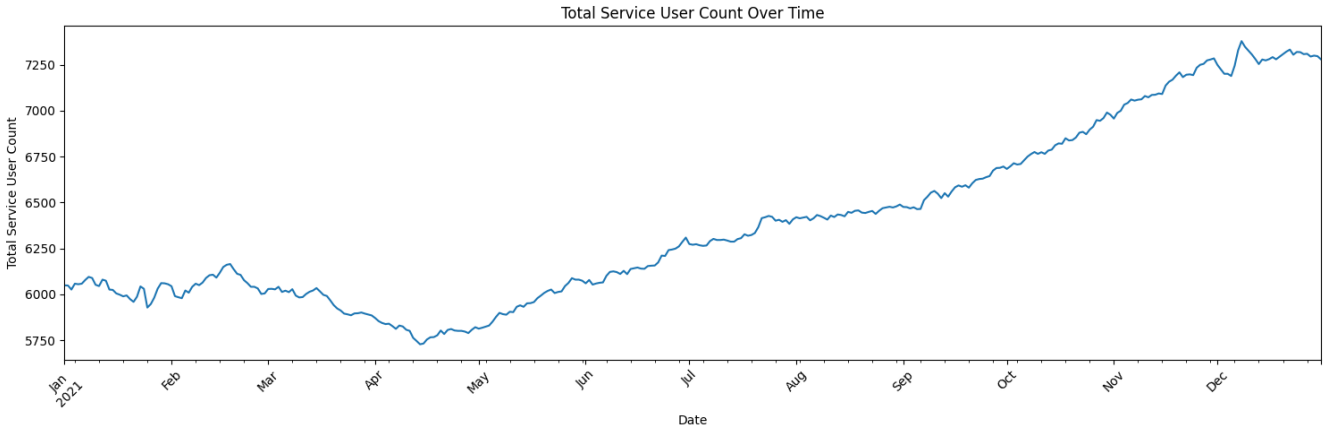


Additionally, I created a boxplot of the room capacity for each program model. This boxplot shows us that the means and interquartile ranges for the room capacities of emergency vs. transitional program models are similar, however there are many outliers for the emergency programs, and few outliers for the transitional programs. The similarity in means and interquartile ranges suggests that, on average, the room capacities for emergency and

transitional program models are not significantly different. The presence of many outliers for emergency programs indicates that there are instances where emergency programs have significantly higher or lower room capacities compared to most programs. The presence of few outliers for transitional programs suggests that room capacities for transitional programs are more consistently distributed, with fewer extreme values. This indicates that there are some unique characteristics or challenges within emergency shelter facilities compared to transitional shelter facilities.



Lastly, I wanted to look at the trends in the number of shelter users over time. We can see that the line reaches its lowest point in mid-April then steadily increases. This shows that there is an increase in unhoused people. Since we know that the Service User Count is linearly correlated with the number of Occupied Beds, and that this is in turn, mostly linearly correlated with the Actual Bed Capacity, we can infer that Toronto Shelters are near capacity, and more beds are probably needed.



Conclusion

In conclusion, the analysis of the dataset on shelter services in Toronto has revealed valuable insights into the utilization patterns and characteristics of various shelter programs. Statistical tests demonstrated significant differences in occupancy rates between emergency and transitional programs across different sectors, indicating variations in their utilization. Exploratory data analysis highlighted a substantial demand for shelter services, with most programs serving a small number of users while a few handle larger populations. The presence of outliers in emergency program room capacities suggested unique challenges or characteristics within these facilities compared to transitional programs. The temporal analysis demonstrated a concerning increase in the number of shelter users over time, emphasizing the urgency of addressing the growing homelessness crisis. Overall, these findings underscore the need for tailored strategies for different shelter types, as well as an immediate focus on expanding capacity to meet the rising demand for shelter services in the city.