

## Assignment One

**Introduction:**

The recent increase of the homeless population has become a new challenge for the shelter support service in Toronto. While the capacity for shelter support is also increasing in response to this challenge, many individuals are still unable to have access to these support services due to the fast growing homeless population and limited resources in government. To reduce the pressure in shelter support service, one of the methods is to allocate limited resources to the place where we need it most. And in order to be able to do so, using collected data to understand the usage pattern of shelter support service is necessary.

For this assignment, I want to focus on analyzing the relationship between shelter usage and time of year in order to better understand the temporal pattern of shelter usage. This assignment consist of two research questions:

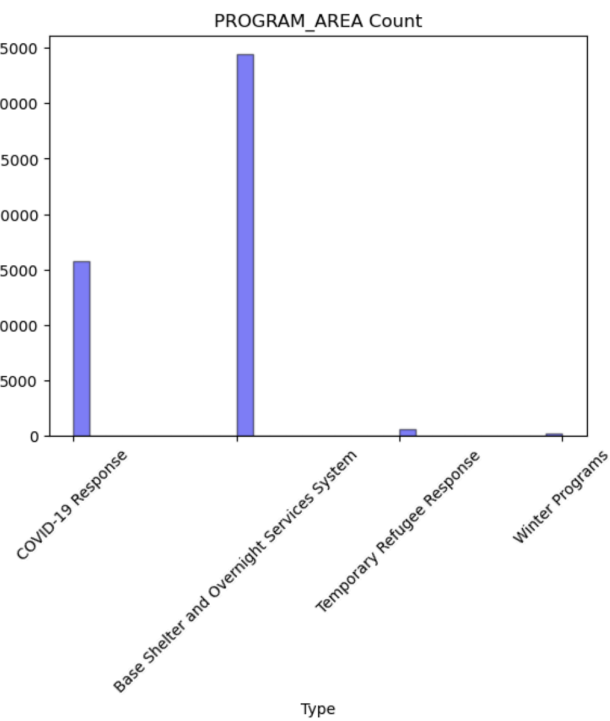
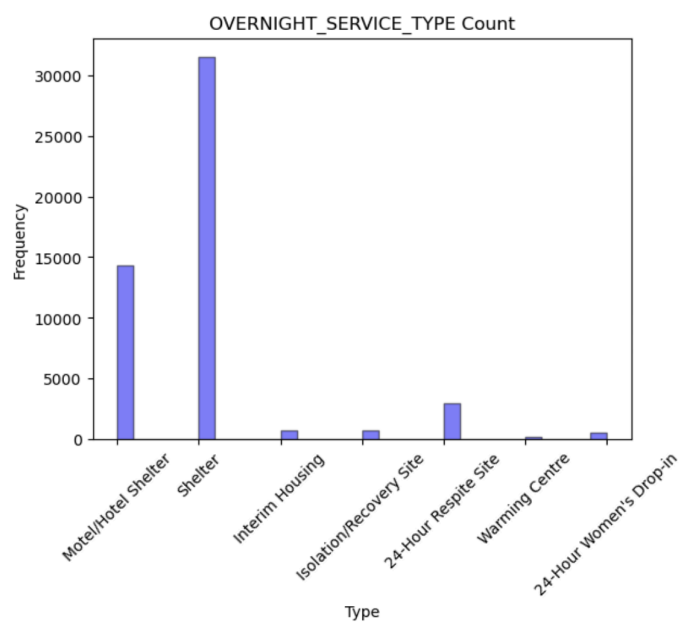
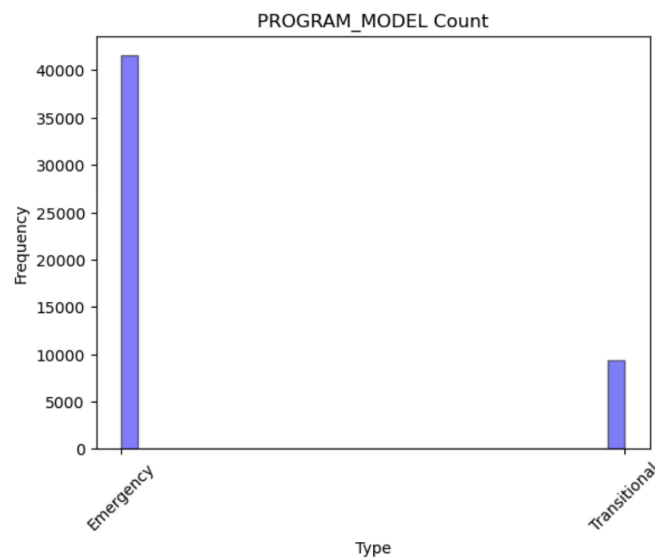
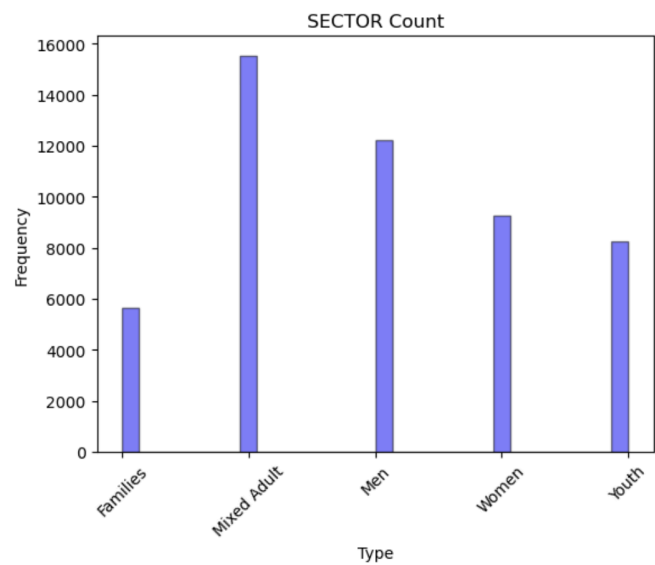
- whether the usage of shelter changes over time in 2021
- whether the capacity of provided shelter program changes enough to adjust to the change of demand

**Method & Procedure:**

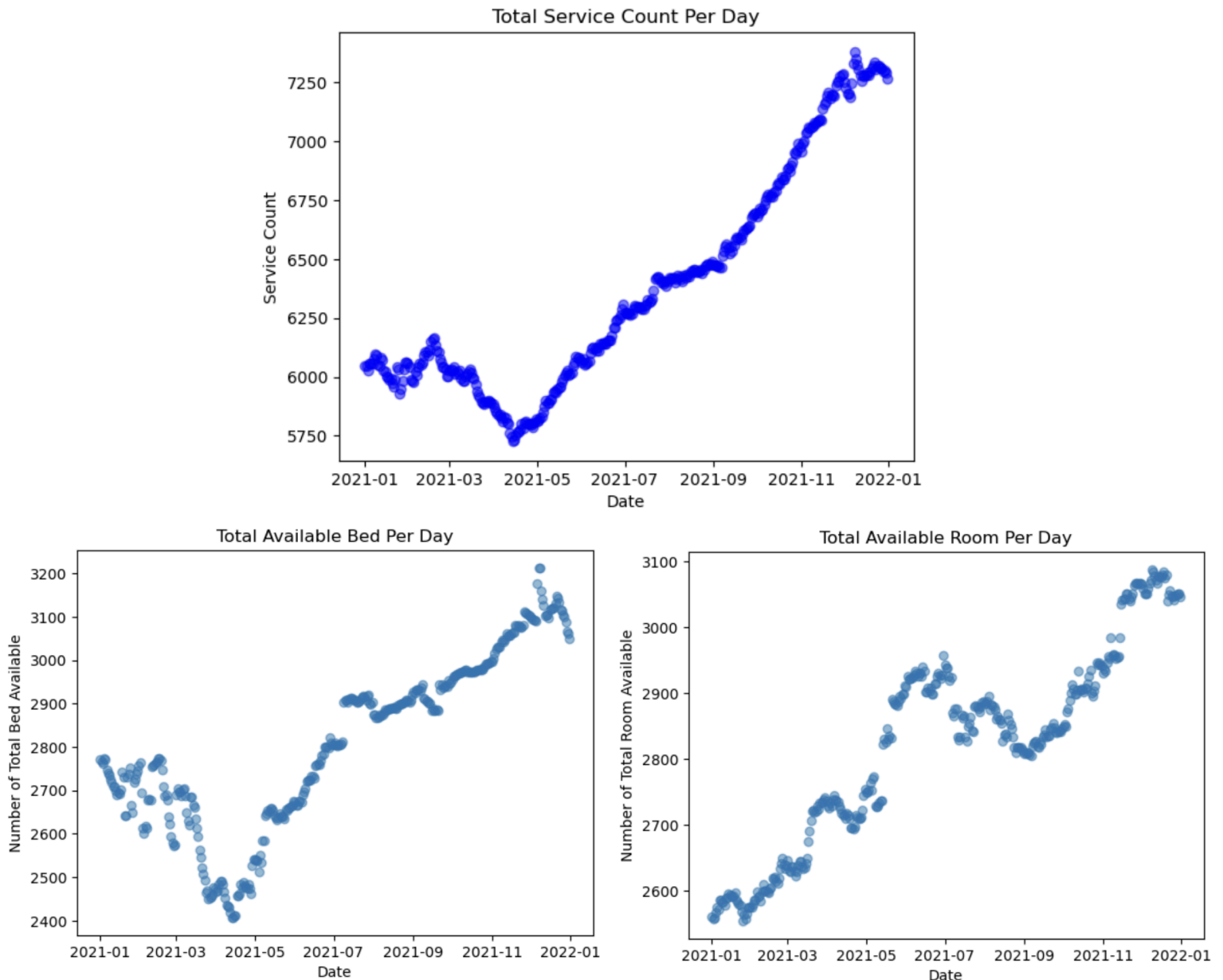
After loading the data, the first step is to check for missing value. Luckily, only a small amount of data is missing in column Program\_Name, Program\_Model, Overnight\_Service\_Type, and Program\_Area. Since this assignment will focus on the temporal pattern of shelter usage, missing program names will not affect much on the result, therefore I decided to keep them. As for the other three columns, since these categorical variables may influence the temporal pattern of shelter usage, I decided to remove them in order to perform further analysis. Only two observations are being removed, which won't affect the result significantly. After cleaning the missing value, I used python to generate the descriptive statistic for this data.

:	OCCUPANCY_DATE	PROGRAM_ID	SERVICE_USER_COUNT	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS	CAPACITY_ACTUAL_ROOM	OCCUPIED_ROOMS
count	50942	50942.000000	50942.000000	32397.000000	32397.000000	18545.000000	18545.000000
mean	2021-06-29 13:21:33.094107136	13986.022005	45.728515	31.628145	29.781400	55.549259	52.798598
min	2021-01-01 00:00:00	11791.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	2021-03-30 00:00:00	12233.000000	15.000000	15.000000	14.000000	19.000000	16.000000
50%	2021-06-28 00:00:00	14251.000000	28.000000	25.000000	23.000000	35.000000	34.000000
75%	2021-09-29 00:00:00	15651.000000	51.000000	43.000000	41.000000	68.000000	66.000000
max	2021-12-31 00:00:00	16611.000000	339.000000	234.000000	234.000000	268.000000	268.000000
std	NaN	1705.241577	53.326660	27.128189	26.379825	59.448805	58.792954

Based on this overview of the quantitative data, we can see that on average a shelter takes 45 people one day, and the average capacity is higher than average usage. To have an overview of the categorical data, I used histograms to visualize the count of different categories.

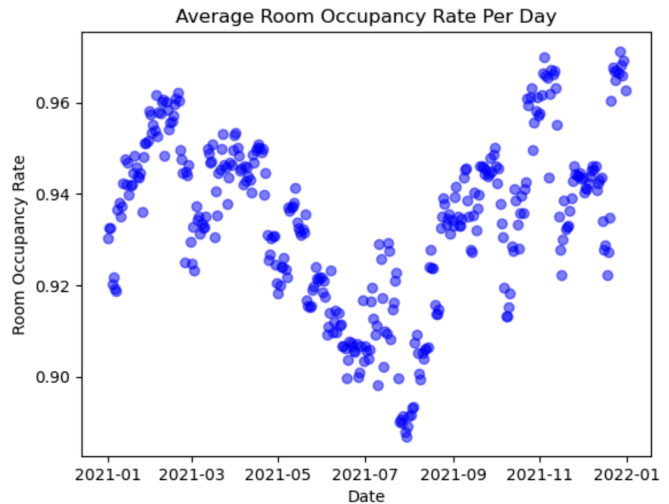
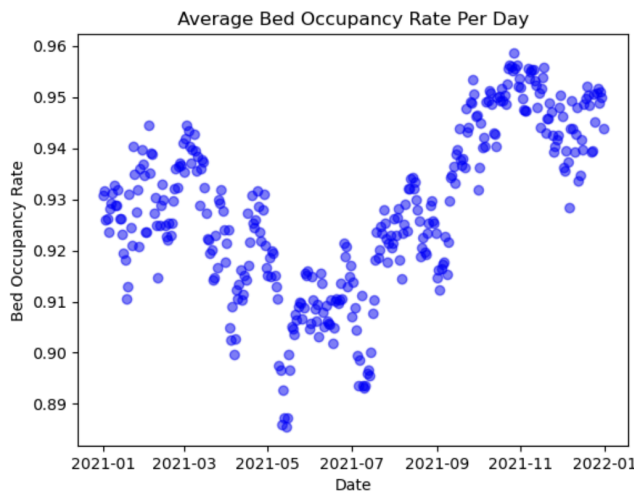


Based on the graph, we can see that the majority of shelter are mix adult, emergency, regular shelter, and belong to the base shelter and overnight services system. Since this assignment focuses on temporal pattern analysis, further detailed examination of the relationship between these groups will not be performed. The total number of users is generated in the scatter plot, the maximum capacity is also presented in the scatter plot.



From the graphs, we can see that capacity and usage of shelter gradually increase from spring to winter, and go down from winter to spring. While these graphs tell us the usage and capacity of every day, it is hard to see whether the increase in capacity matches the increase in users. To reflect this, choosing the correct metrics is important. For this data, I decided to use bed occupancy rate and room occupancy rate, which are calculated based on room/bed occupied and maximum capacity of shelter. After creating

two columns that store these metrics, I used a scatter plot to visualize the usage trend per day over time.



Based on these graphs, we can see that there seems to be a seasonal pattern in shelter usage. The shelter usage tends to be high during winter and low during summer. But is this relationship significant enough? To answer this question within the scope of this assignment, I decided to use t-tests to examine the differences between seasons. Based on month, I categorized them into spring, fall, winter, and summer. The hypotheses are as follow:

Test One:

- $H_0$ : Summer's average bed occupancy is the same as fall and spring
- $H_a$ : Summer's average bed occupancy is lower than fall and spring

Test Two:

- $H_0$ : Winter's average bed occupancy is the same as fall and spring
- $H_a$ : Winter's average bed occupancy is higher than fall and spring

Test Three:

- $H_0$ : Summer's average room occupancy is the same as fall and spring
- $H_a$ : Summer's average room occupancy is lower than fall and spring

Test Four:

- $H_0$ : Winter's average room occupancy is the same as fall and spring
- $H_a$ : Winter's average room occupancy is higher than fall and spring

## Results:

Test One:

t-statistic = -9.069980380945879  
p-value = 6.390947000043207e-20

Test Two:

t-statistic = 1.871045069238  
p-value = 0.030675430703284757

Test Three:

t-statistic = -9.087629178579402  
p-value = 5.733006878615375e-20

Test Four:

t-statistic = 2.843462260364103  
p-value = 0.0022345708167123795

We can see that compared to the other two seasons, summer does have a significant decrease in occupancy rate. Moreover, both tests for winter get a p-value that is lower than 0.05. This suggests that there is significant evidence to support the all alternative hypothesis.

## Discussion:

While the capacity is increasing, shelters are still getting more crowded in winter, proven by the significant increase in occupancy rate. On the other hand, summer has less occupancy rate than other seasons. Ideally, the occupancy rate should be the same across different seasons if enough shelter service resources are allocated correctly. However, the t-tests result shows that the growth of shelter capacity is not consistent with the growth of need for shelter. In other words, the increasing shelter service during winter fails to maintain its quality in other seasons, and summer seems to have redundant resources. This suggests that the government may consider to increase total spending especially in winter, or slightly decrease the spending in summer and allocate more resources to winter, in order to better help homeless people.

One of the limitations of this assignment is the use of t-test. This assignment only analyzes the data in one dimension using two sample t-tests. For future improvement, the interaction between each category and their effect on occupancy rate should be considered. More advanced techniques such as correlation analysis or regression may be used to further understand the pattern in this data.