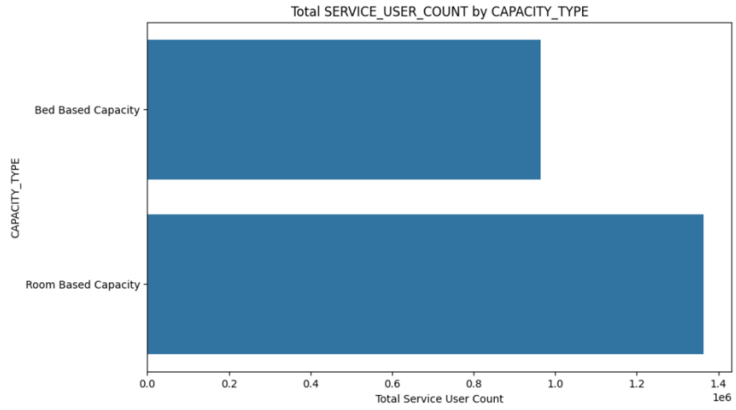
INF 2178 - Assignment 1
Lan Li (1005814326)

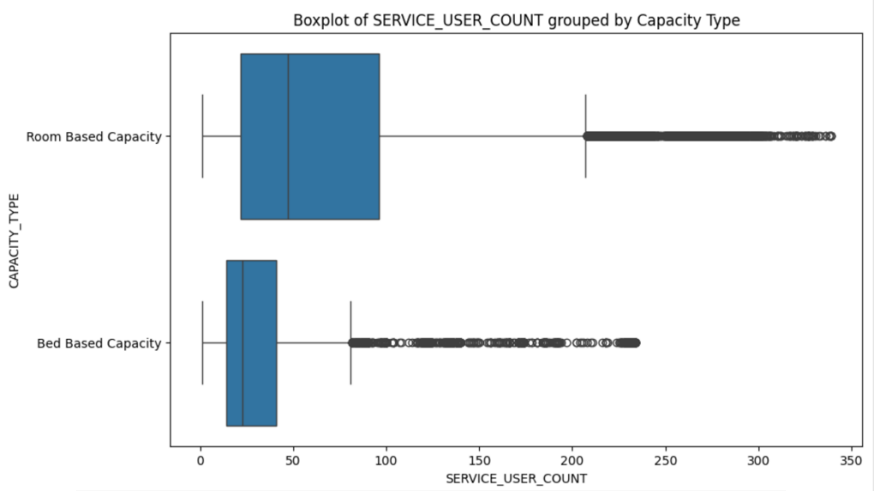
This project is trying to answer what are the differences in service user count and occupancy rates between bed-based and room-based capacities in shelter service of Toronto, and what does this indicate about the demand of the services.

First, the count of service users is investigated by the two different types of capacity. The bar plot illustrates the total number of the service users categorized by bed-based capacity and room-based one. There are 964,851 users using bed-based capacity, and 1,364,674 users using room-based capacity. More users are received by room-based capacity.

```
CAPACITY_TYPE
Bed Based Capacity      964851
Room Based Capacity    1364674
Name: SERVICE_USER_COUNT, dtype: int64
```



The box plot shows the right-skewed distributions of service user count in the two groups. For each program, the room-based capacity has more users than the bed-based one, because the medians of them are so different (47 and 23 respectively).



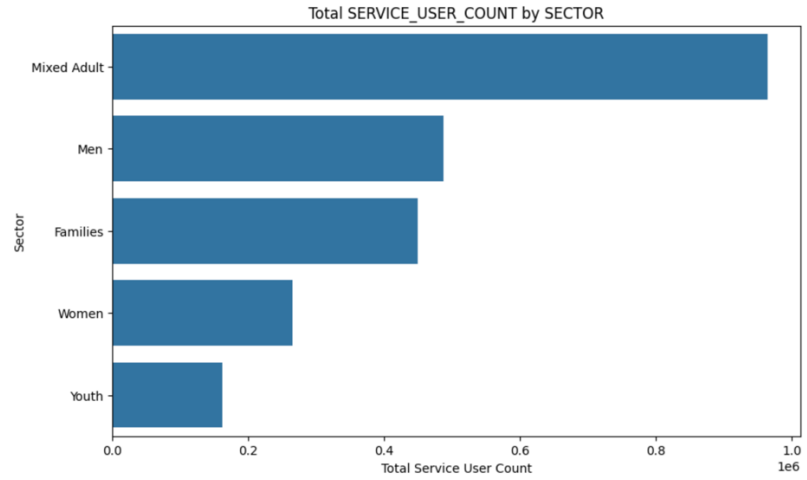
CAPACITY_TYPE	count	mean	std	min	25%	50%	75%	max
Bed Based Capacity	32399.0	29.780271	26.379416	1.0	14.0	23.0	41.0	234.0
Room Based Capacity	18545.0	73.587166	73.319030	1.0	22.0	47.0	96.0	339.0

The result of the Welch's t-test shows that the p-value is essentially 0, meaning there is a significant difference between the average service user count for room-based and bed-based capacity. We should reject the null hypothesis of equal mean. The t-statistic (-78.51) indicates that the mean of the 'Bed Based Capacity' is lower than the mean of 'Room Based Capacity'.

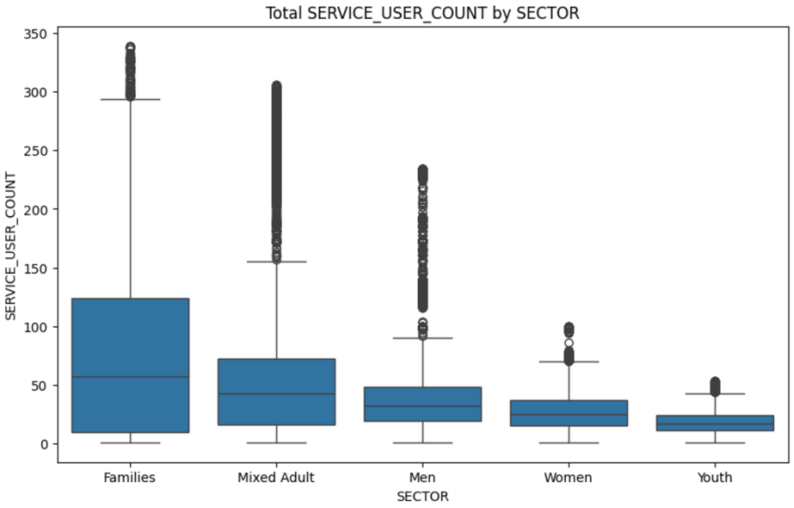
```
T-statistic: -78.50868849938448
P-value: 0.0000000000
```

Second, we count the users in different sectors. The sector which has the largest portion of service users is "mixed adult". The second largest sector is "Men".

```
SECTOR
Mixed Adult    964801
Men            488043
Families       449954
Women          265299
Youth          161428
Name: SERVICE_USER_COUNT, dtype: int64
```

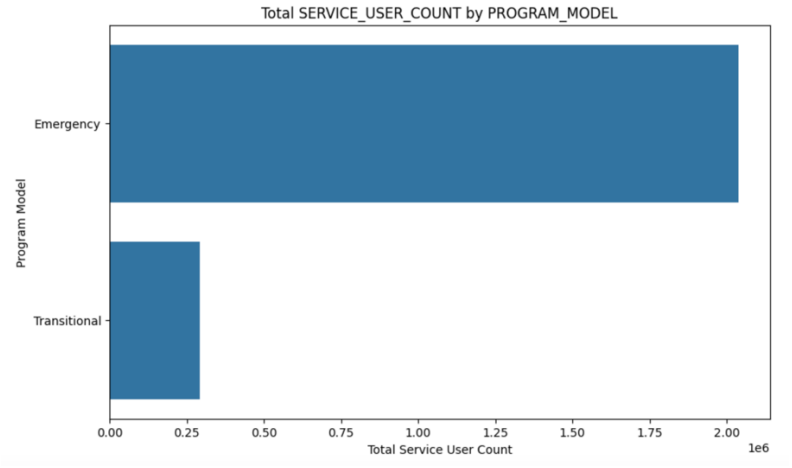


By observing the box plot, for each program, the 'families' sector has the highest number of users, then the 'mixed adult' has the second largest group of users.

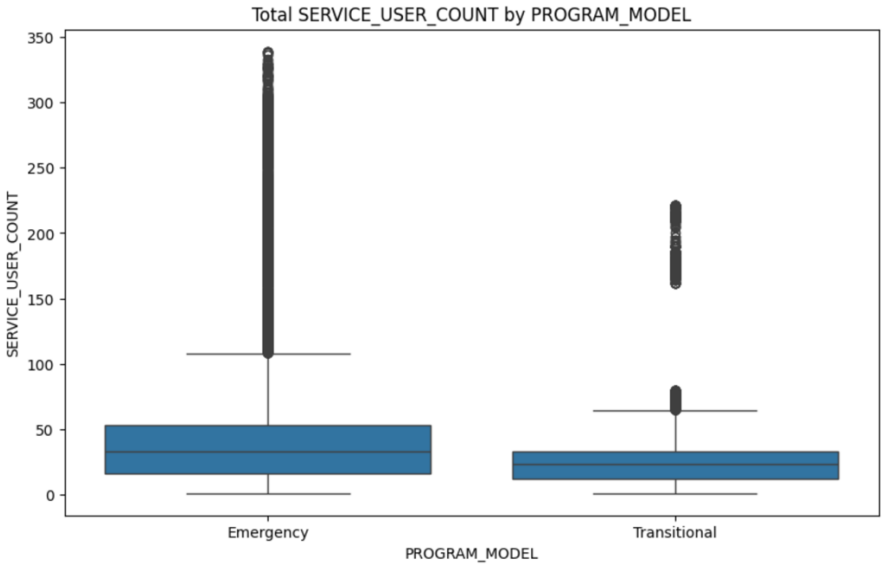


The bar chart of total count of service users grouped by program model exhibits the majority of the users use emergency program, and the minority of them use transitional program.

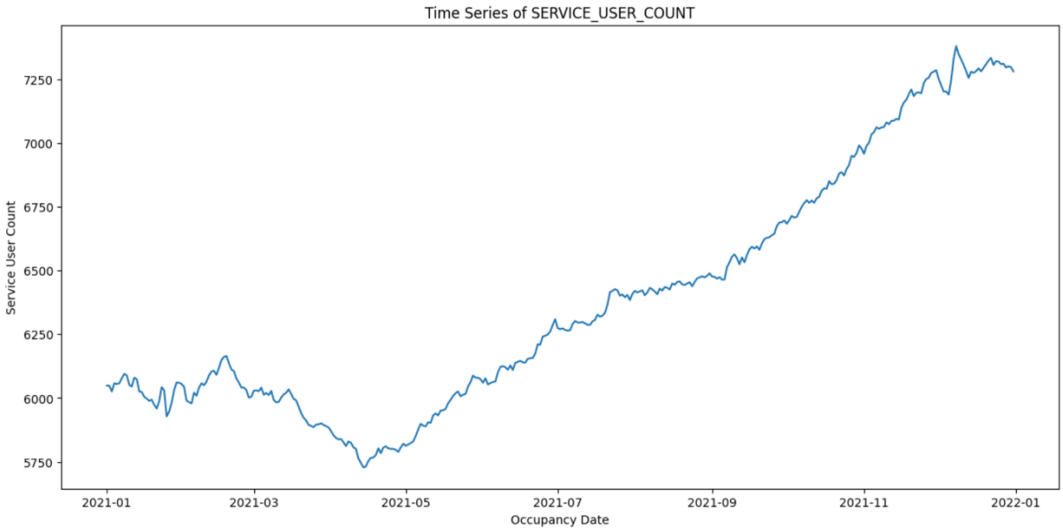
```
PROGRAM_MODEL
Emergency       2038177
Transitional     291325
Name: SERVICE_USER_COUNT, dtype: int64
```



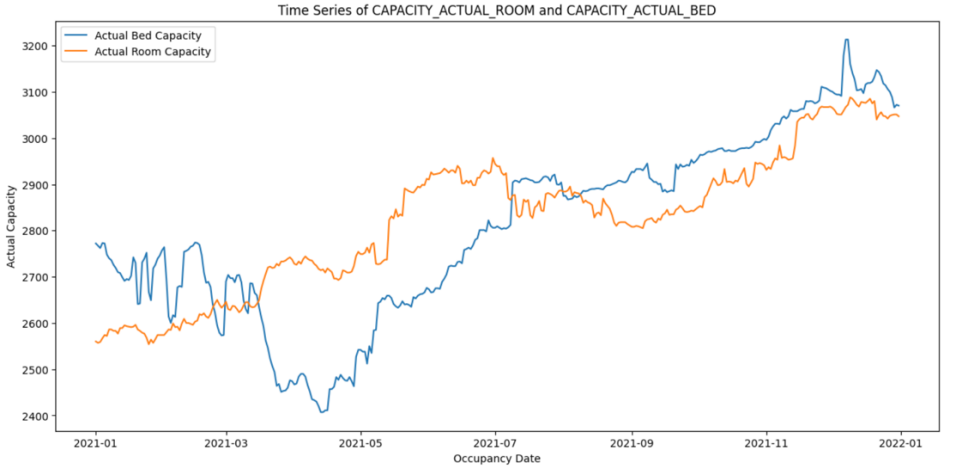
The box plot shows there are slightly more users received by the emergency programs.

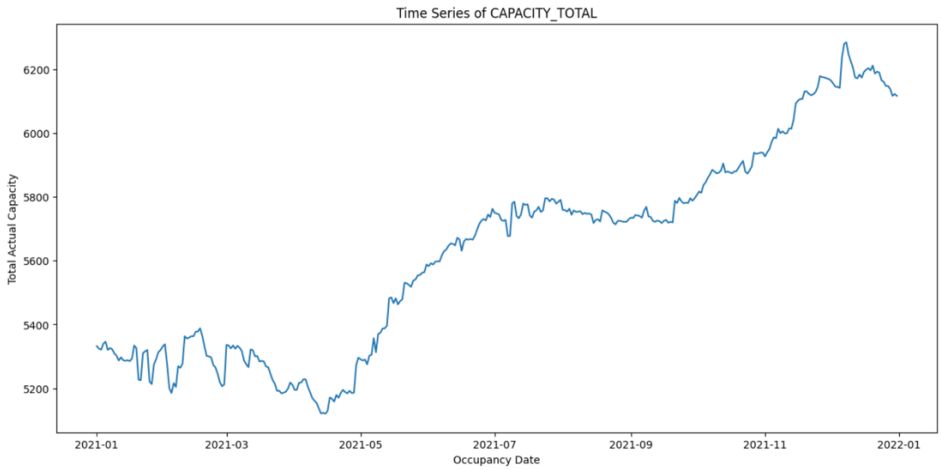


The time series chart of total count of service users shows the period from 2021 to 2022. From January to May, the number of homeless in Toronto using the shelter service decreased. However, from May to the end of year, the number was increasing dramatically, so the increasing demand of the shelter service followed.

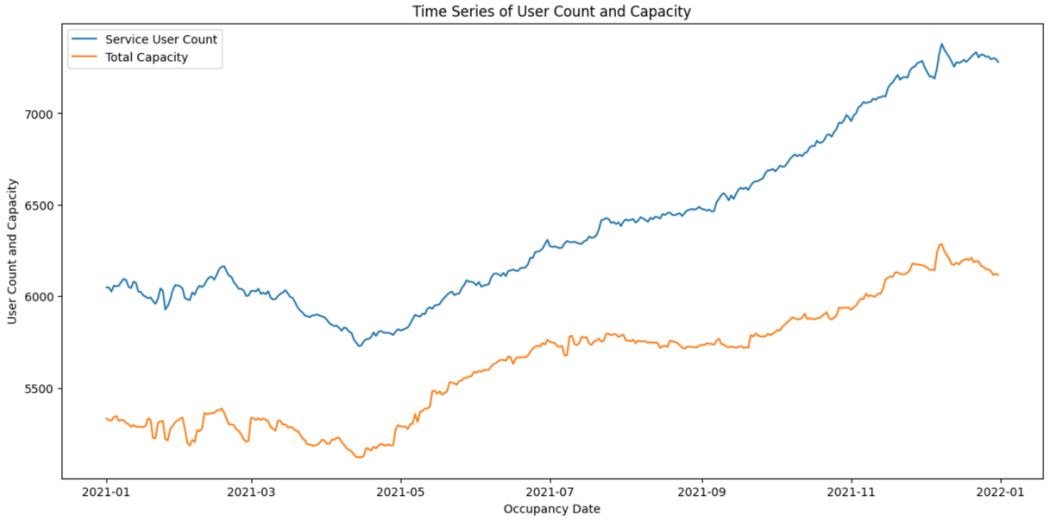


The time series chart below shows how the actual capacity of bed and room change during the whole year. The trend of actual bed capacity slightly goes up. The trend of actual room capacity decreases first but increase later. The second time series chart below shows the total capacity of bed and room.





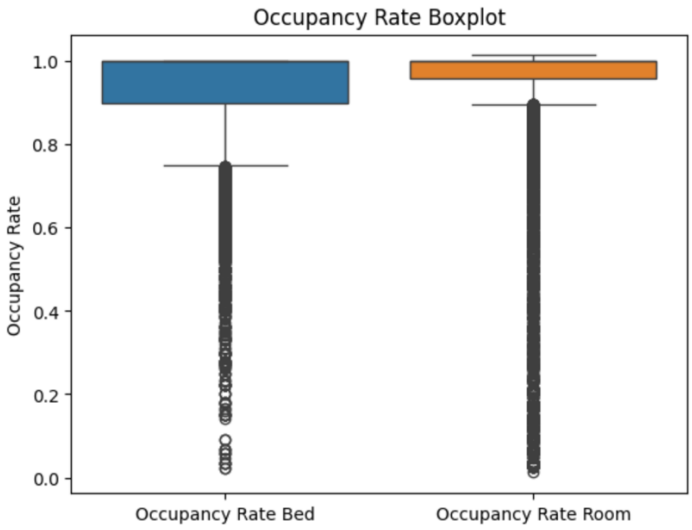
When I put the total count of users and the total capacity in one time series chart, I discover the trend and shape of the two variables are similar. However, the actual capacity is always lower than the service user count, meaning the shelter service is insufficient for the homeless population.



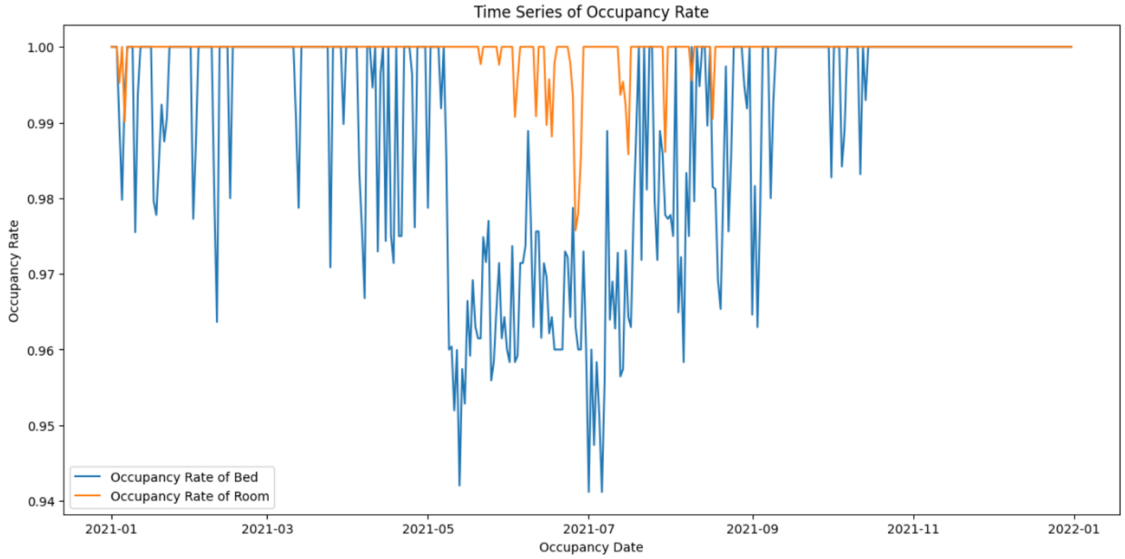
I create two new columns called OCCUPANCY_RATE_BED and OCCUPANCY_RATE_ROOM. They are used to do the t-test. The Welch's t-test results suggest that there is a statistically significant difference between the mean occupancy rates of bed and room. Since the t-statistic is negative, it also means that the mean occupancy rate for bed may be lower than that for room.

```
T-statistic: -4.498751771925636
P-value: 6.860477551487939e-06
```

The box plot also illustrate that the occupancy rate of room is mainly higher than the rate of bed.



When I draw the time series by using the median of the occupancy rate for bed and room, the line representing the occupancy rate of room is more often located on the horizontal line ‘1.00’. Most of the time, room-based capacity has full occupancy rate. Thus, the government should consider the improvement of room-based shelter.



In summary, based on the time series plots, we know there is a big gap between the total count of users and the actual capacity. The capacity is not sufficient for admitting all the homeless users. Toronto government should consider the policy which increase the capacity of the shelters to fill in the gap. Furthermore, there is a significant difference between the mean of occupancy rate for bed and room. The room-based capacity is more often to reach the ceiling. Thus, the government should pay more attention to the room-based shelter improvement.