Introduction

The database from the file "INF2178_A1_data.xlsx" contains information related to various programs operated by an organization named COSTI Immigrant Services. The dataset contains 14 columns which are Occupancy Date, Organization Name, Program ID, Program Name, Sector, Program Model, Over Night Type, Program Area, Service User Count, Capacity Type, Capacity Actual Bed, Occupied Beds, Capacity Actual Room, and Occupied Rooms.

The dataset focused on providing detailed information about the occupancy and capacity of various emergency shelter programs for all gender sectors, particularly in response to COVID-19, managed by COSTI Immigrant Services. It includes both room and bed capacity, which are shown separately in the dataset. The data could be used for analysis of shelter usage, capacity planning, or understanding the usage across different sectors. Specifically in this study, I want to explore the patterns of how different sectors habited within the shelters and if there is a significant difference of habitation between gender.

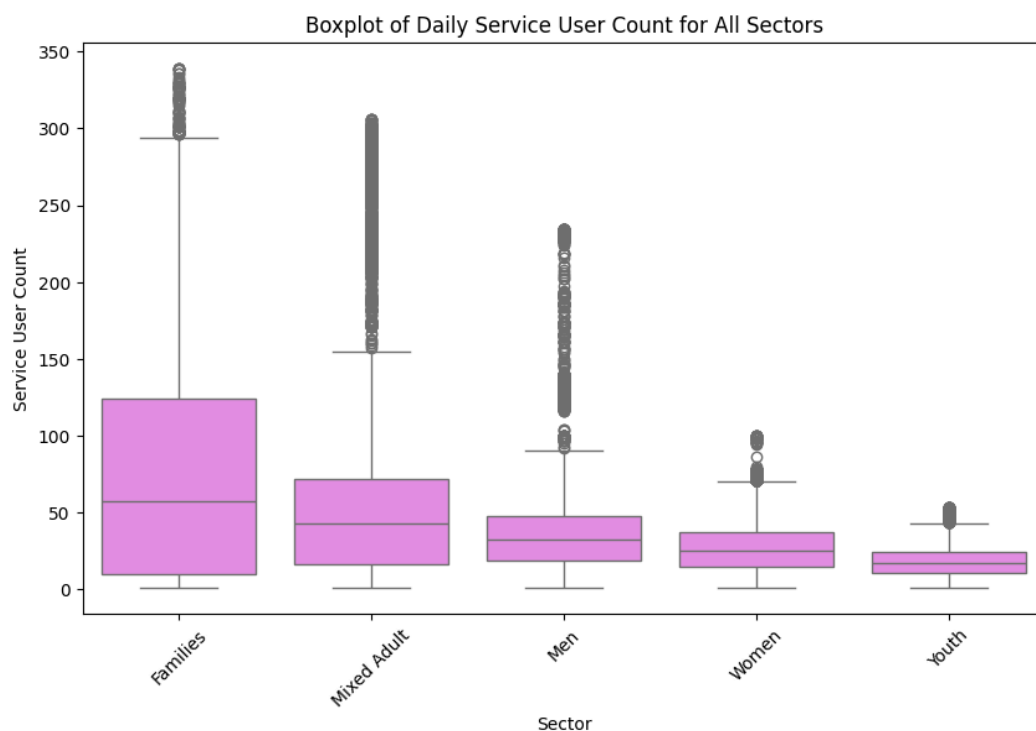**Daily Service User Count for All Sectors**

Figure 1

I would like first to investigate the usage of service per sector. The visualization of Figure 1 showcases the distribution of daily service user counts across different sectors.

The Families Sector appears to have the highest median service user count among all sectors, indicating that it typically has more service users daily. The interquartile range (IQR) is also the largest, suggesting greater variability in the daily counts compared to other sectors. There are many outliers indicating days with unusually high service user counts. This could be the result of the various sizes within a family unit.

For the Mixed Adult Sector, the median service user count is lower than that of the Families sector but higher than the others. The IQR is narrower, which suggests less variability in daily service user counts. There are outliers present, both on the higher and lower ends.

For the Men sector, the median service user count is lower than that of the Mixed Adult sector. This sector shows a small IQR, indicating relatively consistent service user counts from day to day. Numerous outliers are present, suggesting occasional days with significantly different service user counts.

The Women Sector has a lower median service user count than the Men sector. The IQR is small, similar to the Men sector, which again suggests consistency in the daily counts. There are a few outliers, indicating some days with higher service user counts.

For the Youth Sector, the median service user count is the lowest among all sectors. The IQR is very small, suggesting that the number of service users does not vary much from day to day. There are no apparent outliers for the Youth sector, which indicates that the daily service user count is very consistent.

Overall, Boxplot 1 shows that Families and Mixed Adult sectors tend to serve more users than the other sectors, with Families showing the most variability. In contrast, the Youth sector serves the fewest users but with the least day-to-day variation. The presence of outliers in all sectors except Youth indicates that there are days with atypical service user counts, which could be due to a variety of factors such as seasonal changes, special events, or changes in service needs. I would like to further investigate the cause of the lower user service rate of those normally considered a minority group if is due to the lack of available service.

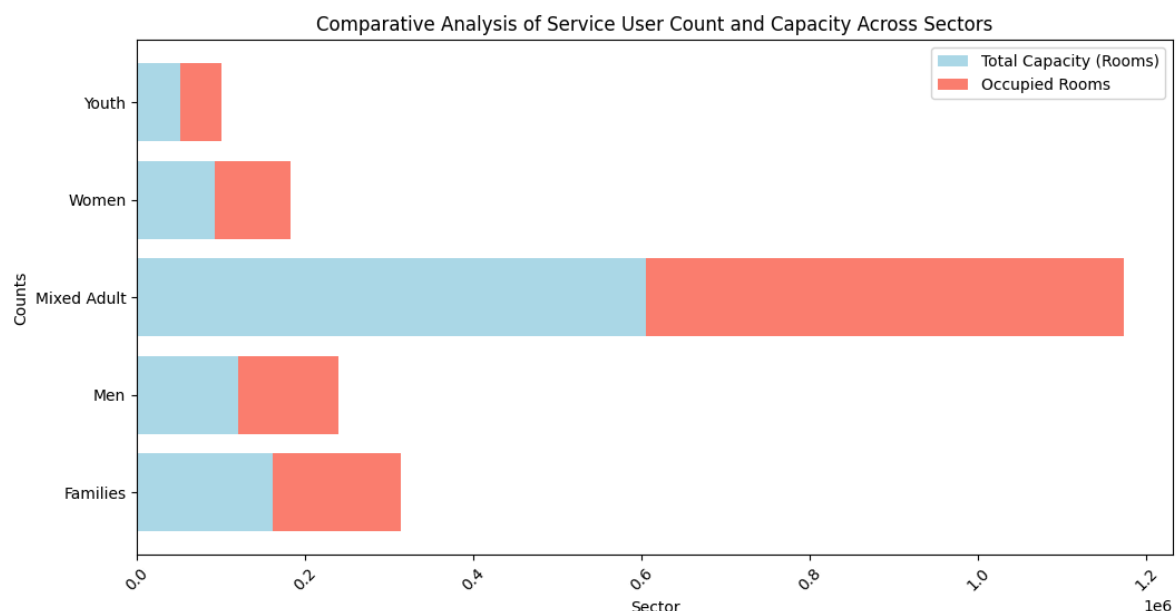## Comparative Analysis of Serice User Count Capcaity Across Sectors

Figure 2

	SECTOR	OCCUPIED_ROOMS	CAPACITY_ACTUAL_ROOM	OCCUPANCY_RATE
0	Families	151957.0	161952.0	0.938284
1	Men	119489.0	119872.0	0.996805
2	Mixed Adult	568075.0	605052.0	0.938886
3	Women	90661.0	92048.0	0.984932
4	Youth	48968.0	51237.0	0.955716

Figure 3

Figure 2 and 3 presents a comparative analysis of service user count and capacity across various sectors: Families, Men, Mixed Adults, Women, and Youth. It compares the total capacity (in terms of rooms available) against the actual number of occupied rooms within each sector. Here's an analysis of the chart:

The Mixed Adults sector has the largest total capacity, yet the lowest occupied rate. The number of Occupied Rooms is 568,075, versus the Actual Room Capacity of 605,052. The occupancy rate is approximately 93.9% (568,075 / 605,052). There is a notable gap between the total capacity and occupied rooms, indicating that there is a significant portion of capacity that is not being used.

The Youth sector has the smallest total capacity and occupied rooms among the sectors shown. Occupied Rooms is 48,968 and Actual Room Capacity is 51,237. The occupancy rate for the Youth sector is approximately 95.6% (48,968 / 51,237), which is quite high. This suggests that the majority of the available capacity is being used, with a small percentage of rooms available.

The Women's sector shows a slightly larger than the Youth Sector but a relatively small total capacity among all sectors. The number of Occupied Rooms is 90,661, and the Actual Room Capacity is 92,048. With an occupancy rate of about 98.5% (90,661 / 92,048), the Women's sector also shows the second highest utilization of the available room capacity, with very few rooms left unoccupied.

The total capacity for the Men sector is a lot less than the Mixed Adult sector. The number of Occupied Rooms is 119,489 and the Actual Room Capacity is 119,872. With the highest occupancy rate, nearly all rooms are occupied in this sector, with an occupancy rate of about 99.7% (119,489 / 119,872). This indicates that the capacity closely matches the demand, with very little unoccupied room.

The total capacity of the Families sector is slightly higher than that of the Men sector. The number of Occupied Rooms is 151,957, and Actual Room Capacity is 161,952. This sector has utilized a large portion of its capacity, with about 93.8% (151,957 / 161,952) of the rooms occupied. There remains a small margin of unoccupied rooms, suggesting either a buffer for fluctuations in demand or potential underutilization.

Overall, while the Mixed Adulte sector has the most resources but with lowest absolute usage, the Youth sector, though smallest, has the relative high usage of its available capacity. Men and Women sectors are showing simillar capacity and usage. Therefore I want to further analyze if there is a significant difference of shelter usage between Men and Women sectors.

Analysis on Difference in Servive User Count of Women vs Men

```
T-statistic: -30.50421164151915, P-value: 1.0532742925183108e-199
```

Figure 4

In order to investigate if there is a significant difference in mean service number between Men and Women. Therefore two hypothesis can be proposed to conduct a t-test.

Null Hypothesis (H0): The null hypothesis states that there is no difference in the mean service user count between the Women and Men sectors. It suggests that any difference observed in the sample data is due to random chance.

Alternative Hypothesis (H1): The alternative hypothesis states that there is a difference in the mean service user count between the Women and Men sectors. It suggests that the observed difference in the sample data reflects a real difference in the population.

Based on the T-Test Result shown on Figure 4, A t-statistic of -30.504 suggests that there is a very large difference between the two groups. The negative sign indicates that the mean SERVICE_USER_COUNT for the "Women" sector is lower than the mean for the "Men" sector. A p-value of 1.05×10^{-199} is extremely small, effectively zero for all practical purposes. This p-value is far below any conventional significance level (e.g., 0.05, 0.01, 0.001), indicating that the observed difference in means between the "Women" and "Men" sectors is statistically significant. However, I want to further validate if the usage difference is because the capacity is significantly different.

Analysis on Difference in Room-Based Occupancy Rate of Women vs Men

```
Room Based Capacity:
T-statistic: -13.729207701163253, P-value: 2.3324018592726617e-41
```

Figure 5

In order to further investigate the validity of the above test by conducting another t-test on the room-based capacity rate between Men and Women, and use a line chart to help visualize the result. Two hypothesis has been proposed:

Null Hypothesis (H0): There is no difference in the mean occupancy rate between the Women and Men sectors.

Alternative Hypothesis (H1): There is a difference in the mean occupancy rate between the Women and Men sectors, specifically that the mean occupancy rate for Women is lower than that for Men (indicated by the direction of the negative t-statistic).

Based on the t-test result shown on Figure 5, A T-statistic of -13.729 indicates that the mean occupancy rate for the "Women" sector is lower than that for the "Men" sector. The magnitude of the t-statistic indicates the difference is relatively large.

The P-value of 2.33×10^{-41} suggests that the probability of observing such a large difference in means by random chance is virtually zero if the null hypothesis were true. Therefore the

reason of no significant difference between Men and Women sectors' shelter usage is unlikely because of capacity difference. Next, I want to visualize the difference of occupancy rate between Men and Women sector to further investigate if there is a pattern to explain the lower occupancy rate for Men and Women sector.

Room-Based Occupancy Rate Over Time for Men and Women
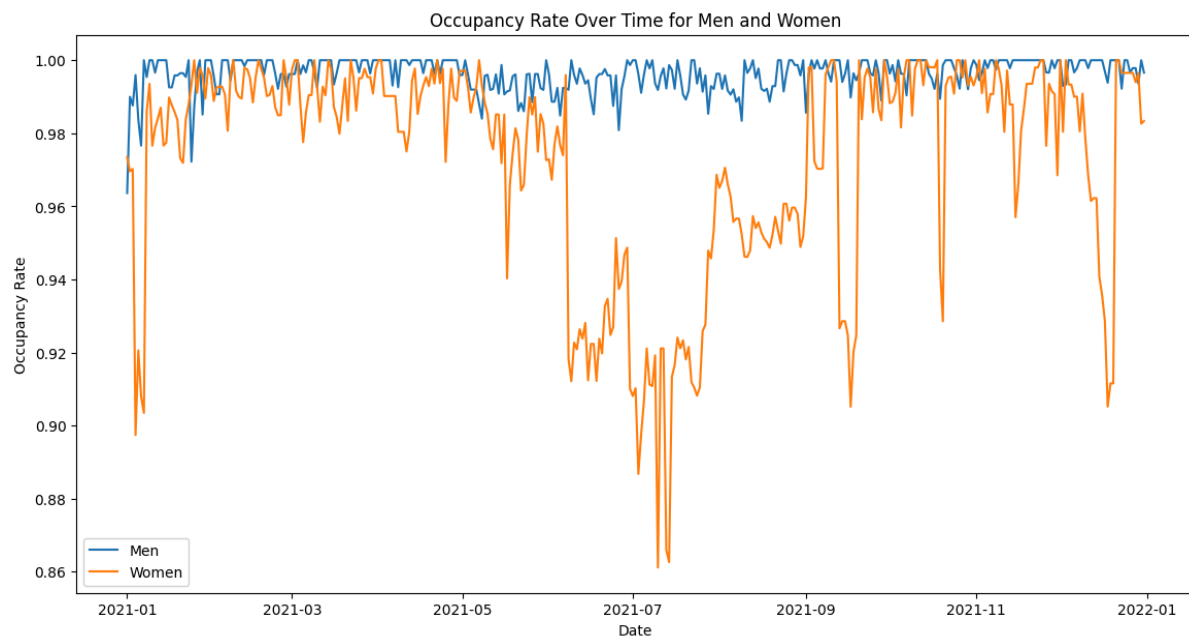

Figure 6

Figure 6 further validate the t-test performed above, demostrated the trends in occupancy rates for the "Men" and "Women" sectors over a period of time that appears to span from January 2021 to approximately January 2022.

Both "Men" and "Women" sectors show fluctuations in occupancy rates over the given time period. The occupancy rates for both sectors largely move in parallel, suggesting that factors affecting occupancy may impact both sectors similarly. The occupancy rates for both sectors are high, staying mostly above 90%. There are several sharp drops (notably around mid-2021 and early 2022) where the occupancy rate for the "Women" sector falls below that of the "Men" sector.

The "Men" sector generally maintains a slightly higher occupancy rate compared to the "Women" sector throughout the period. The "Women" sector has more pronounced dips in occupancy rate, indicating periods where the occupancy rate decreased more significantly.

The "Women" sector shows greater volatility with sharper increases and decreases. The "Men" sector appears to have a more stable occupancy rate with fewer sharp fluctuations. The drops in "Women" sector's occupancy rates could correspond to seasonal patterns, specific events, or changes in policy/operations. The specific reasons for these fluctuations remain unclear from the chart alone and would require additional context or data to interpret accurately.

In summary, while both sectors show high occupancy rates, the "Women" sector appears to experience more significant changes over time. The "Men" sector, while also fluctuating, does not seem to have as many sharp changes as the "Women" sector.

Conclusion

In conclusion, the data indicates a pronounced difference in the service utilization between the sectors, with 'Families' and 'Mixed Adult' typically serving more users, though with notable variability and a degree of underutilization in terms of capacity. The t-tests have shown statistically significant differences in both service user count and occupancy rates between the 'Women' and 'Men' sectors, with 'Women' consistently showing lower figures in comparison. The occupancy rate analysis over time reveals that while the trends for both sectors generally move in parallel, the 'Women' sector experiences more pronounced volatility with significant dips that could suggest sensitivity to external factors or internal policy changes. For next step, I wish to further investigate how does season affect Women sector's shelter usage.