TECHNICAL

ASSIGNMENT 1

Understanding Toronto's Shelter Dynamics: A Closer Look at Family Needs

INF2178

SUBMITTED TO: Dr. Shion Guha

SUBMITTED BY: Adil Mahmood Ashraf

Student Number: 1010727017

Understanding Toronto's Shelter Dynamics: A Closer Look at Family Needs

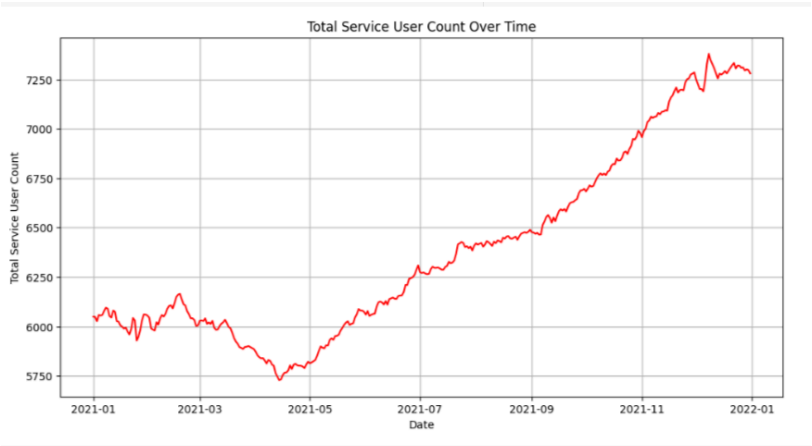
1- Treatment and cleaning the data:

To make the shelter data for Toronto easier to work with, I added a new column called 'CAPACITY' to help with my analysis. I calculated 'CAPACITY' by dividing 'CAPACITY_ACTUAL_ROOM' by 'OCCUPIED_ROOMS', creating a continuous variable for conducting t tests.

Next, I cleaned up the dataset by removing any duplicate entries. Duplicate entries can make the results of the analysis unreliable by increasing the number of observations artificially. By getting rid of these duplicates, I made sure that my analysis was based on unique and accurate records, maintaining the integrity of my findings.

Lastly, I dealt with missing values in the dataset. This step was crucial, especially for columns with only a few missing entries. Instead of tossing out these records or leaving the gaps, which could introduce errors or bias into the analysis, I chose to fill in the missing values with the most common value in their respective columns. This method, called imputation, allowed me to keep as much data as possible while not significantly changing the overall distribution of the dataset.

2- The Rise in Shelter Usage : <u>Figure 1: Total Service User Count Over Time (Jan'21-Jan'22)</u>:

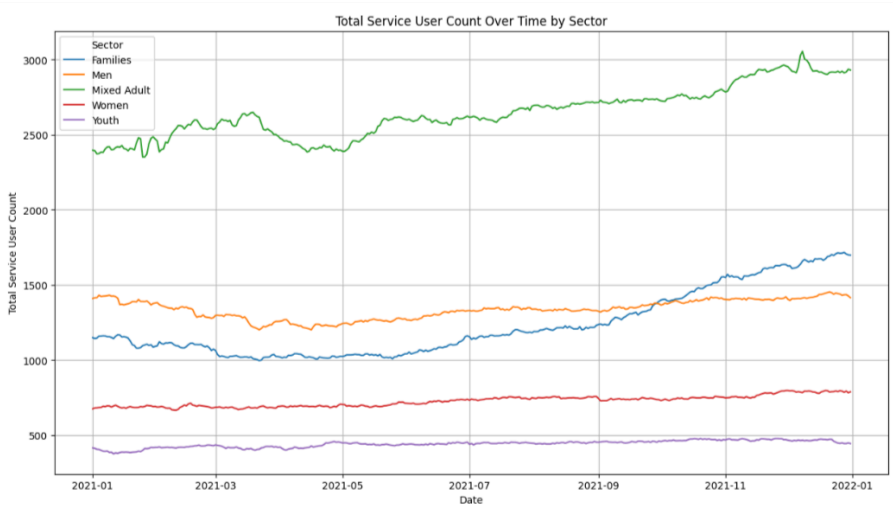


To get my dataset ready for analysis, I made sure all the dates in the

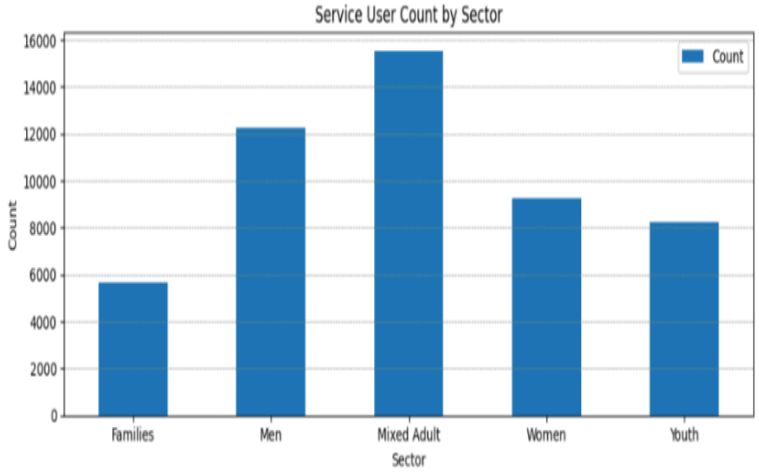
'OCCUPANCY_DATE' column were correctly formatted as datetime objects. This was important because using dates as text or in different formats can cause errors when doing time-based analyses or creating time-series plots. I used the pandas 'to_datetime' function to convert these dates, making it easier to work with time-related data.

It's clear that there's a growing housing problem in Toronto. Over the past year, more and more people have been talking about it, as shown in Figure 1. The rising number of people discussing housing issues highlights the increasing housing crisis in the city.

3- The Rise in Family Shelter Use: <u>Figure 2: Total Service User Count Over Time by Sector</u>



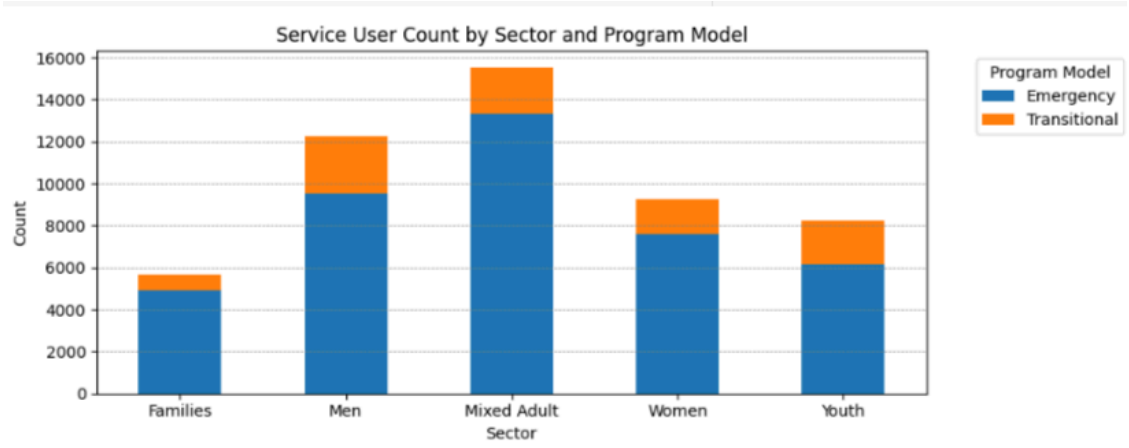
As I dug into the data for analysis, I wanted to figure out which groups benefited the most from housing programs in Toronto. When I looked at the line graphs, one trend really caught my eye: a significant increase in the use of services by families. Their steady increase in shelter use raises questions about the social security systems that are supposed to catch them before they fall.



4- Visualizing Shelter Use by Sector: <u>Figure 3: Service User Count by Sector</u>

To get a deeper understanding of the data, I broke it down into unstacked bar graphs. This allowed me to see how many people were using services in different sectors. At first glance, it looked like families had the smallest number of users (5649). But, there could be a couple of reasons for this. Firstly, it might suggest that many families don't rely on social security for housing. On the other hand, it could mean that families are struggling to access these resources due to financial or social issues related to living in shelters. There's often a stigma attached to not being able to provide for a family, and I wanted to explore this further to better understand what the data was presenting.

5- Emergency vs. Transitional: <u>Figure 4: Service User Count by Sector and Program Model</u>

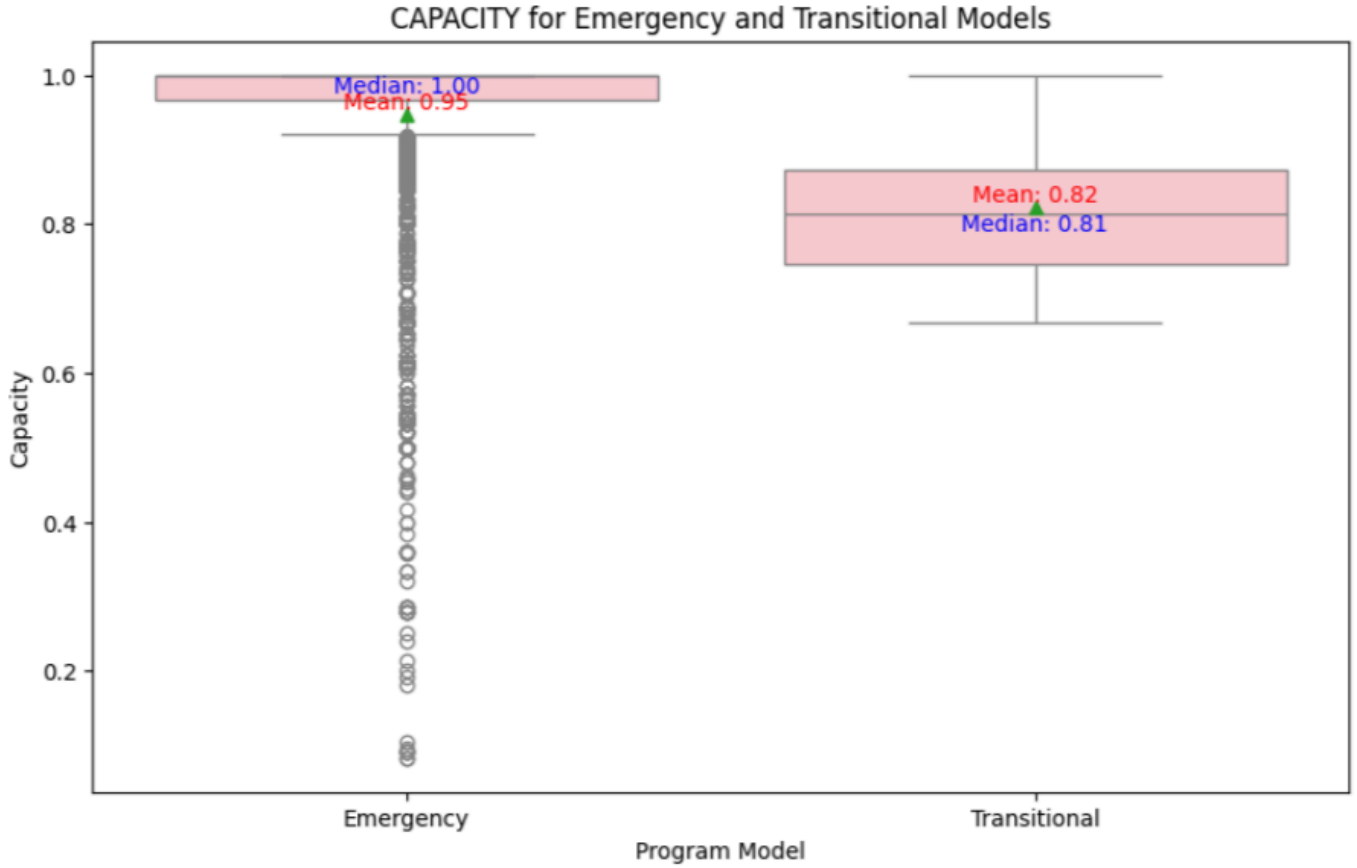


In order to do so, I'm focusing on the shelter programs that families are using. There's a noticeable difference here - families are using emergency shelters more, but they are not as represented in

transitional shelters.

6- T Test, p values and Boxplots: <u>Figure 5: Boxplots for Emergency and Transitional Model</u>

```
Emergency Capacity Normality Test: ShapiroResult(statistic=0.48218250274658203, pvalue=0.0)
Transitional Capacity Normality Test: ShapiroResult(statistic=0.9313963651657104, pvalue=8.823953320312485e-18)
```




```
Welch's t-test Statistic: 32.0056
P-value: 0.0000
```

I used a box plot to compare the 'CAPACITY' of emergency and transitional shelter program models and conducted Shapiro-Wilk normality tests .

On one side, we have the 'Emergency' model, where the mean and median capacity values hover around the full capacity mark of 1.00 (100%). This suggests that emergency shelters are frequently operating at their maximum, a testament to either their efficiency or a pressing

demand that keeps them consistently full. For the 'Emergency' model, the test returns a statistic of approximately 0.482 with a striking p-value of 0.0, flatly rejecting the notion of a normal distribution.

Transitioning to the 'Transitional' model, the mean capacity dips to 0.82, and the median follows closely at 0.81, suggesting that, on average, operate below full capacity. This finding could imply either a systemic efficiency, allowing room for sudden influxes, or it could signal an underutilization that warrants further investigation. Similarly, the 'Transitional' model's normality is questioned with a statistic of around 0.931 and a p-value approximately 8.23e-18. These p-values, starkly lower than the conventional alpha level of 0.05, firmly dismiss the null hypothesis of normal distribution for both models.

In the context of the 'Emergency' and 'Transitional' shelter capacity data, we observe the following:

- **Non-Normal Distribution:** The Shapiro-Wilk test has indicated that the data for both groups do not follow a normal distribution.
- **Mean and Median Differences:** The boxplot analysis shows that the mean and median are quite different in the 'Transitional' model.
- **P-Values from Shapiro-Wilk Test:** The extremely low p-values suggest that the capacity data for both groups deviate significantly from a normal distribution.

Given these points, the decision to use a Welch's t-test was taken after considering the large sample sizes of Families who use shelter services and due to the unequal variance between emergency and transitional services. The Welch’s t-test indicates a t-statistic of 32.0056 with a p-value of 0.0000, which is highly significant. This means there is a statistically significant difference between the capacity of emergency and transitional models. The result is consistent

with the visible differences in the box plot distributions.

7- The Capacity Conundrum and The Way Forward: <u>Figure 05: Service User Count by Sector and Capacity Type</u>

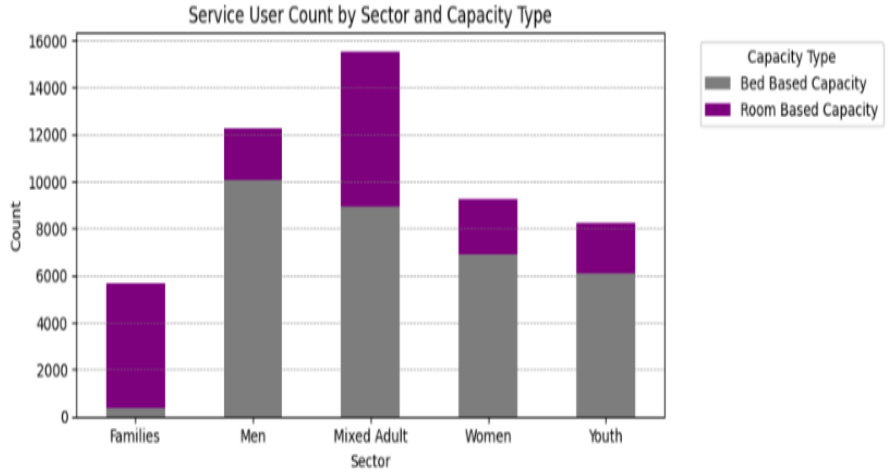


Figure 5 paints a picture of the User Service Count and Capacity Types available in Toronto. When we focus on families, we see that bed-based capacity is only a small part of the total service user count. Looking at the different types of capacities provided, it's clear that families are less likely to use bed-based services. This raises an important question: This raises an important questions that requires qualitative data analysis: What situations do families often find themselves in that lead to a disproportionate need for emergency shelters as opposed to transitional ones? Is the shelter system adequately prepared to provide struggling families with more than just a bed for the night? Why do families not utilize Transitional services, despite availability of capacity?

This data isn't just a bunch of numbers; it's the story of families navigating a complex and possibly inadequate system. This insight pushes me to advocate for a shelter program that's as adaptable and diverse as the lives it intends to support.The story told by these figures is a call to ensure that every family has access to a shelter, but also a chance for social stability.