Toronto Shelter Data: An Exploratory Data Analysis of Occupancy Rates in 2021

Maya Ghai

1.0 Introduction

The dynamics of shelter occupancy represent a critical aspect of understanding and addressing homelessness within a community. The ebb and flow of individuals seeking shelter can be influenced by a myriad of factors, from weather conditions and natural disasters to economic fluctuations. In this exploratory data analysis, I specifically explore descriptive statistics of shelter occupancy rates, across program models (emergency or transitional) and sectors

The purpose of this summary is to delve into month-to-month changes in shelter occupancy, aiming to identify trends in the demand for shelter services throughout the year. The broader outcome of this data analysis may inform strategic planning for shelter management and resource allocation across a twelve-month span.

1.1 Research Question

The central question guiding this analysis asks if shelter occupancy rates fluctuate across months of the year. We seek to describe the central tendency and spread of occupancy rates per month and analyze if these trends change according to sector and/or program type.

2.0 Initial Dataset

The data frame tracks the daily occupancy of Toronto shelters in the 2021 calender year, and includes 50944 rows from 169 unique programs, across a 365 day span. The following analysis relies on the existing columns from the dataset:

- OCCUPANCY_DATE: The date of shelter occupancy.
- CAPACITY_TYPE: Indicates whether the capacity is based on beds or rooms.
- OCCUPIED_BEDS: Number of beds occupied in the shelter.
- OCCUPIED_ROOMS: Number of rooms occupied in the shelter.
- CAPACITY_ACTUAL_BED: Actual capacity of beds in the shelter.
- CAPACITY_ACTUAL_ROOM: Actual capacity of rooms in the shelter.
- SECTOR: The sector classification of the shelter (e.g., 'Women').
- PROGRAM_MODEL: The program model classification of the shelter (e.g., 'Emergency' or 'Transitional').

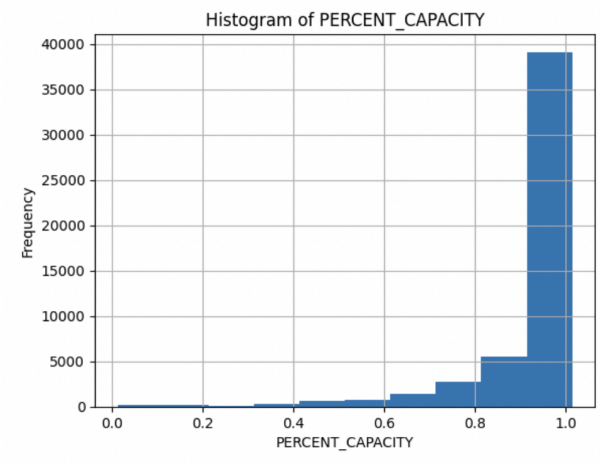
2.1 Data Transformations

The initial dataset only contains a count for the number of occupied beds or rooms per night. In order to make comparisons in rates across the different capacity types, a 'PERCENT_CAPACITY' column is created as the ratio of occupied beds/rooms to the actual capacity of beds/rooms. The statistics for central tendency and variation of the newly created percent capacity variable are described. As a compliment, an AVAILABLE_SPACE column is created, calculated as the remaining percent capacity available. A boolean value for SPACE_AVAILBLE is stored if percent capacity is lower than 1.00 on the given day. Finally, month is extracted as an integer from 1-12 from the date column, and mapped to the name of the month (‘Jan’, ‘Feb’, etc.). Finally, there are two instances with no program model specified. As these represent a minor fraction of the total dataset, we drop these two rows for consistency.

- PERCENT_CAPACITY: Calculated percentage of capacity based on either beds or rooms.
- MONTH: Transformed from 'OCCUPANCY_DATE' to represent the month of shelter occupancy.

2.2 Initial Data Examination

Figure 1: *Distrbibution of Percent Capacity*



Note. Kurtosis: 11.94, Skewness: -3.14

Mean	Median	Min	Max	IQR
0.93	1.0	0.01	1.01	0.08

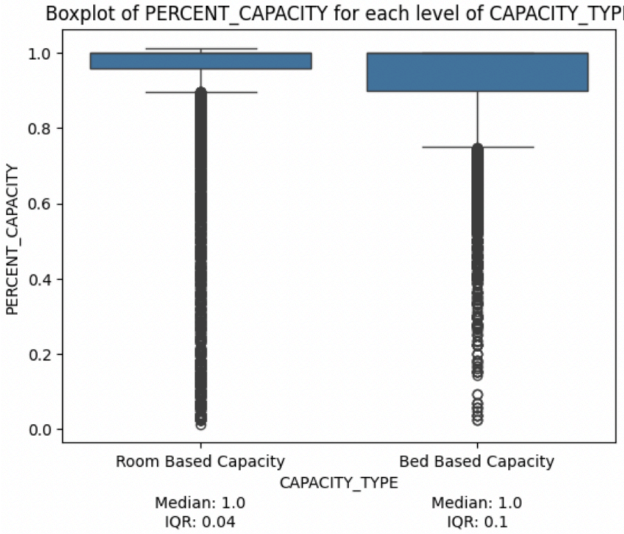
We note four instances where percent capacity is greater than 1.00 (overcapacity). For consistency, these values are updated as 1.00.

2.4.1 Descriptive Statistics: Exploratory Data Analysis

With our processed data, we can begin some initial exploration of occupancy rates. Figure 1, as well as the skew/kurtosis value, illustrates a heavy negative skew in occupancy rates, suggesting a concentration of values within a small range of high occupancy. The characteristic of the data is expected, and aligns with the substantial demand for shelters in Ontario, where facilities are consistently near or at full capacity.

The substantial skewness in the distribution led to a thoughtful consideration of measures for central tendency and spread. Given the skewed nature of the data, median and interquartile range (IQR) have been selected as metrics for the boxplots, while the mean is still reported. We can visualize and describe the occupancy rates for the 2021 calender year across categories of the data.

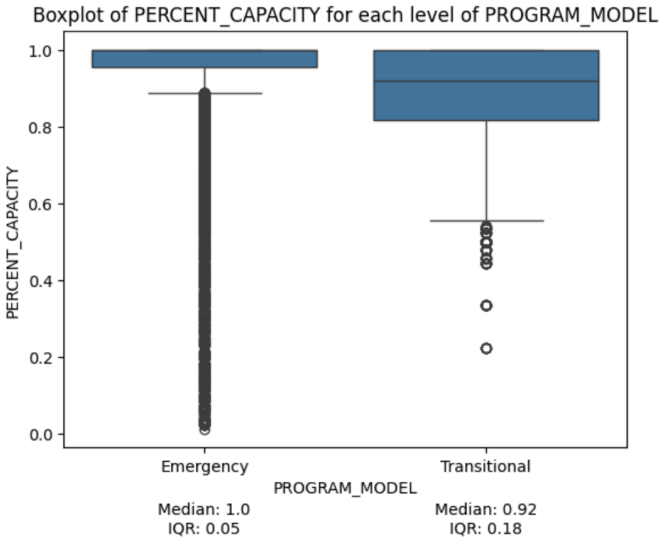
Figure 2: *Summary Statistics of Percent Capacity by Capacity Type*



Capacity Type	Mean	Median	Min	Max	IQR
Bed Based	0.93	1.0	0.02	1.00	0.10
Room Based	0.93	1.0	0.01	1.00	0.04

Upon first analysis, a greater spread, or interqurtile range, in the bed-based capacity indicates higher variance in the occupancy rates for beds in the entire 2021 calender year as compared to rooms.

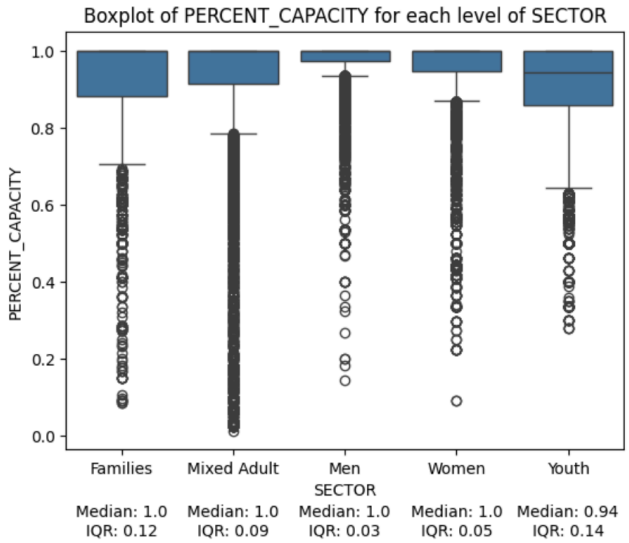
Figure 3: *Summary Statistics of Percent Capacity by Program Model*



Program Model	Mean	Median	Min	Max	IQR
Emergency	0.94	1.0	0.01	1.00	0.05
Transitional	0.88	0.92	0.22	1.00	0.18

Similarily, upon initial analysis we observe greater variance in the use of transitional shelters than emergency shelters.

Figure 4: *Summary Statistics of Percent Capacity by Sector*

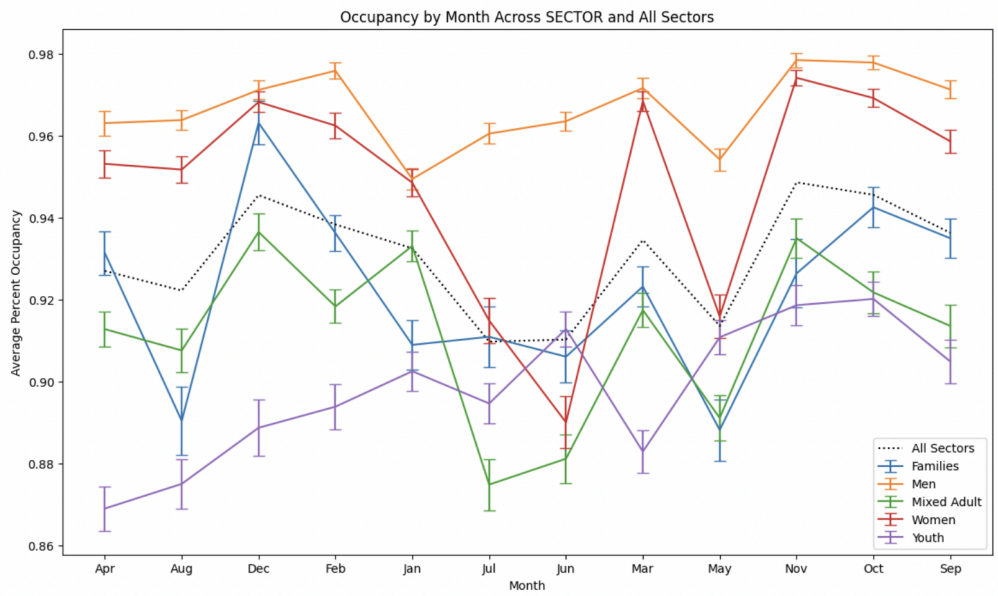


Sector	Mean	Median	Min	Max	IQR
Families	0.92	1.00	0.08	1.00	0.12
Mixed Adult	0.91	1.00	0.01	1.00	0.09
Men	0.97	1.00	0.14	1.00	0.03
Women	0.95	1.00	0.09	1.00	0.05
Youth	0.90	0.94	0.28	1.00	0.14

Finally, analyzing across sector we can derive some initial ideas about the nature of data. Notably, initial analysis that the Men sector exhibits the highest average percent capacity, potentially suggesting a consistent demand for shelter space. Moreover, this sector displays the lowest variance in occupancy rates across the months, indicating a relatively stable demand pattern. On the other hand, the Youth sector demonstrates a larger spread in occupancy rates and a lower median value, implying a more dynamic demand for shelter space with fluctuations throughout the year. Nonetheless, these initial insights are purely descriptive– to understand the statistical significance of these claims we conduct further analysis on the data.

2.4.1 Exploratory Data Analysis Across Month

Figure 5: *Mean Occupancy by Month Across Sector*



Note. Error bars represent the standard error of the mean.

The visualization highlights a compelling trend: a substantial decrease in average percent occupancy during the summer months. This trend aligns with common observations in many regions, where individuals experiencing homelessness may find alternative arrangements during warmer weather, such as sleeping outdoors. As expected from Figure 4, the men's sector displays a less pronounced fluctuation in average occupancy rates across seasons. On the other hand, this seasonal drop in occupancy rates is most pronounced in the sector for women. For the remaining analysis, I will specifically examine the Women's sector to delve deeper into this specific data.

3.1.1 Independent T-Test

To explore whether variations in occupancy rates for Women are significantly different across different months, independent samples t-tests are employed. The comparisons are made with respect to program model. Given the potential unequal variances between groups, Welch's t-test is utilized to ensure robustness in the analysis. To mitigate the potential for an inflated Type I error rate due to multiple comparisons, a Bonferroni correction is applied. The hypotheses tested for each month are structured as follows:

- **Null Hypothesis (H_0):** The mean occupancy rates are equal for the given month across Emergency and Transitional programs.
- **Alternative Hypothesis (H_A):** The mean occupancy rates are not equal for the given month.

4.0 Results

The table provides the results of independent t-tests comparing occupancy rates in Emergency and Transitional program models for the Women sector across different months. Key statistical measures include the t-statistic, uncorrected p-value, adjusted p-value (using Bonferroni correction), and an indication of significant differences.

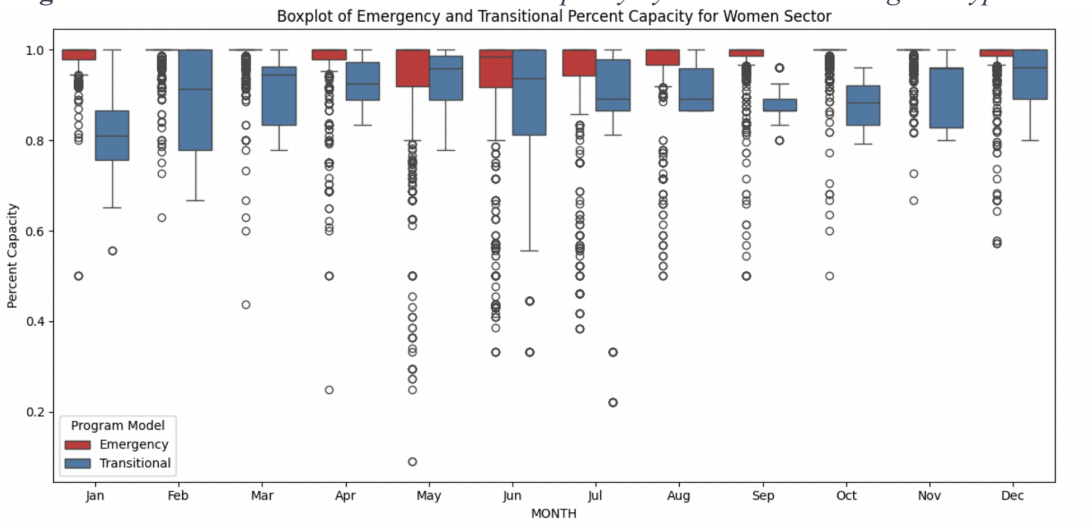
Month	T-statistic	P-value (uncorrected)	Adjusted P-value (Bonferroni)	Significant Difference
Jan	17.913	0.000	0.000	* Yes
Feb	10.583	0.000	0.000	* Yes
Mar	10.909	0.000	0.000	* Yes
Apr	5.859	0.000	0.000	* Yes
May	-1.340	0.181	2.172	No
Jun	2.562	0.011	0.134	No
Jul	3.941	0.000	0.001	* Yes
Aug	9.333	0.000	0.000	* Yes
Sep	20.395	0.000	0.000	* Yes
Oct	18.804	0.000	0.000	* Yes
Nov	11.793	0.000	0.000	* Yes
Dec	6.799	0.000	0.000	* Yes

Notably, the analysis indicates significant differences in occupancy rates for both program types during the majority of the calendar year.

3.3 Discussion

For the Women's sector, the months of May and June, however, display interesting dynamics. May and June shows no significant difference in occupancy rates ($p > 0.05$), suggesting a potential convergence of utilization between the two program types during this period, or less overall underutilization of emergency shelters during summer months. The occupancy rates for emergency and transitional programs for Women are visulized with a boxplot.

Figure 6: *Women's Sector: Mean Shelter Occupancy by Month Across Program Types*



Note. May and June show no significant difference in occupancy rates between emergency and transitional shelters ($p > .05$). There is a statistcially significant difference in occupancy rates between program models for all other months ($p < .05$*).

3.4 Future Work: Expanding Insights

Future research endeavors may focus on exploring the underlying factors contributing to the observed dynamics in May and June within the Women's sector. This could involve qualitative investigations, stakeholder interviews, or a more granular analysis of contextual variables to provide a comprehensive understanding of trends. Moreover, extending similar analyses to other sectors is a logical next step, as seasonal trends may differ across demographic groups. Such insights can inform resource allocation, program enhancements, or policy adjustments to better address the unique needs of individuals seeking shelter during specific months.