

INF2178 Assignment 1

Overview

This dataset provides a comprehensive view of shelter occupancy across various programs in Toronto, encompassing 50,944 unique entries and 15 variables over the year 2021. It offers a detailed glimpse into the daily occupancy status throughout the seasons, incorporating information on organizations, programs, capacity types, service user counts, and occupancy rates. This analysis aims to delve into the nuances of shelter capacity and occupancy rates, examining factors such as the user-to-bed/room ratios and the impact of gender on shelter utilization.

Preparing the Dataset

The analysis was conducted in Google Colab using Python, with Pandas and Matplotlib serving as the core libraries for data manipulation and visualization. The initial step involved uploading the dataset from local storage, followed by a preliminary review to understand its structure and completeness. As illustrated in Figure 1, the dataset is largely complete, with minor missing values primarily in the columns pertaining to capacity. These missing values were preserved to maintain the integrity of the dataset and ensure a comprehensive analysis.

	Value
Total Rows	50944
Total Columns	15
Column Names	[OCCUPANCY_DATE, ORGANIZATION_NAME, PROGRAM_ID...
Data Types	{'OCCUPANCY_DATE': datetime64[ns], 'ORGANIZATI...
Missing Values	{'OCCUPANCY_DATE': 0, 'ORGANIZATION_NAME': 0, ...
Unique Values	{'OCCUPANCY_DATE': 365, 'ORGANIZATION_NAME': 3...
First Date	2021-01-01 00:00:00
Last Date	2021-12-31 00:00:00

Figure 1, a summary of the dataset

A critical aspect of shelter capacity is the distinction between bed-based and room-based shelters. Accordingly, the dataset was segmented into two subsets based on this criterion. For bed-based shelters, columns irrelevant to this context, such as 'CAPACITY_ACTUAL_ROOM' and 'OCCUPIED_ROOMS', were dropped. A similar approach was applied to the room-based shelters, where columns like 'CAPACITY_ACTUAL_BED' and 'OCCUPIED_BEDS' were removed to streamline the analysis focused on relevant data.

Feature Engineering

In order to extract more meaningful insights from the dataset, two critical features were engineered that were not present in the original data. The first newly created feature is the 'BED(ROOM)_OCCUPANCY_RATE', which is derived by dividing the number of actual

occupied beds or rooms by the total capacity available in the shelters. This occupancy rate is a pivotal metric, offering a clear picture of the shelters' utilization over the course of the year.

The second feature developed focuses on the efficiency of space utilization within the shelters, represented by the user-to-bed ratio for bed-based shelters and the user-to-room ratio for room-based shelters. These ratios are calculated by dividing the 'SERVICE_USER_COUNT' by the respective capacities, encapsulating how effectively each bed or room is being used. The computed ratios are then added to the dataset in a new column titled 'USER_TO_BED(ROOM)_RATIO', enabling a deeper analysis of space utilization across different shelter types.

Shelter Capacity

The graphical representation of Toronto's shelter capacities offers an insightful visualization of how the shelter system accommodates the city's needs throughout the year. As depicted in the line graph (Figure 2), we observe the daily total capacity trends for both bed-based and room-based shelters over the span of 2021. It is evident that the overall capacity for both shelter types has experienced growth within the year. Notably, bed-based shelters consistently exhibit a higher capacity compared to their room-based counterparts. However, a pronounced dip in the capacity of bed-based shelters between March and May signals a potential area of interest that warrants a deeper examination.

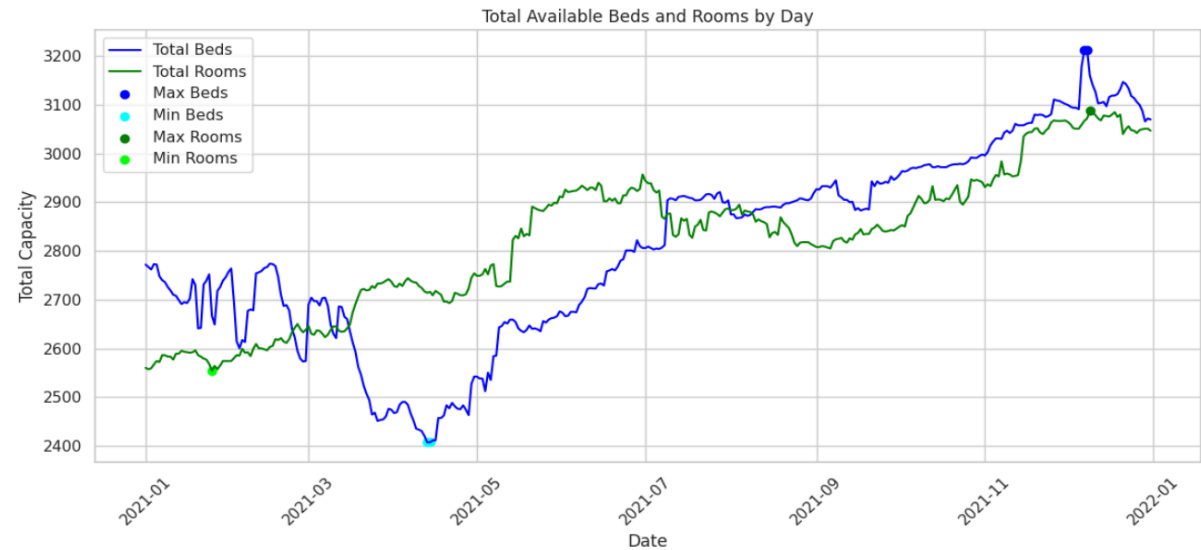


Figure 2, total available beds and rooms by day

Delving further into the dynamics behind the observed fluctuations in shelter capacities, the dataset was meticulously segmented into four distinct subsets, reflecting the combinations of capacity types and program models. The subsequent analysis, visualized through four separate line graphs (Figure 3), reveals a stark contrast in capacity between transitional and emergency programs. Contrary to the general upward trend, the capacity within transitional programs has seen a decline over the course of the year. The earlier noted sharp reduction in the capacity of bed-based shelters can be predominantly attributed to a decrease in the emergency bed-based shelter capacity, highlighting a specific area of concern within the shelter system's capacity management.

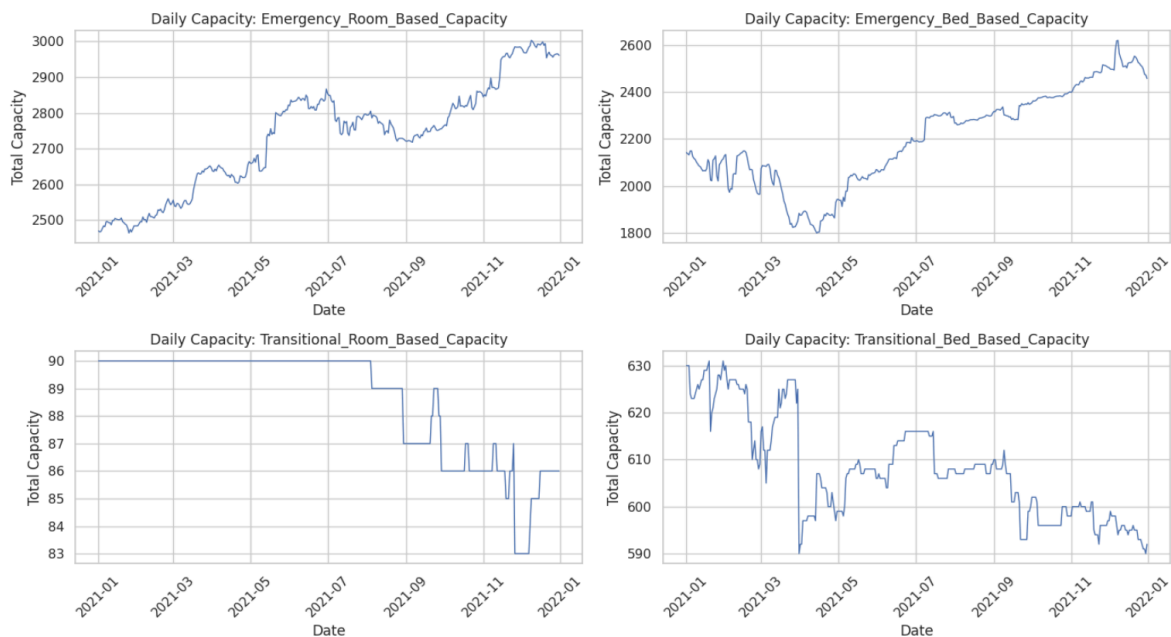


Figure 3

Occupancy Rate

The Occupancy Rate (OCR), which varies from 0 (empty) to 1.0 (full), is a key metric in assessing shelter utilization. Year-long distributions of OCRs for both bed-based and room-based shelters, depicted in histograms, predominantly skew towards full capacity (Figure 4). Bed-based shelters report an average OCR of 0.9279, while room-based shelters are slightly higher at 0.9341, which indicates that room-based shelters generally operate nearer to full capacity.

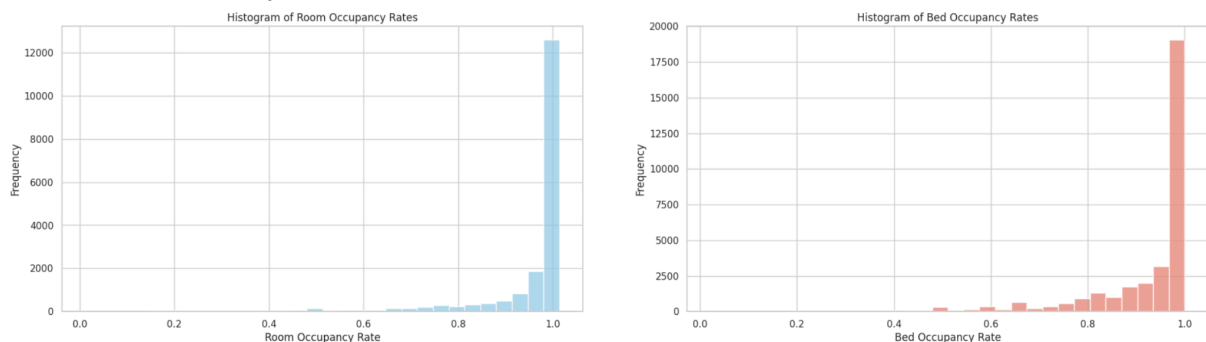


Figure 4, histogram of occupancy rates

Seasonal trends in shelter demand are evident from the analysis of daily average OCRs, with higher occupancy rates observed during the colder months and lower rates in warmer periods (Figure 5). This seasonality is most pronounced in emergency programs, whereas transitional programs show a milder seasonal correlation. Notably, transitional room-based shelters have seen occupancy spikes, occasionally reaching full capacity, potentially impacting accessibility for those in need, especially as their capacities have declined over the year (Figure 6).

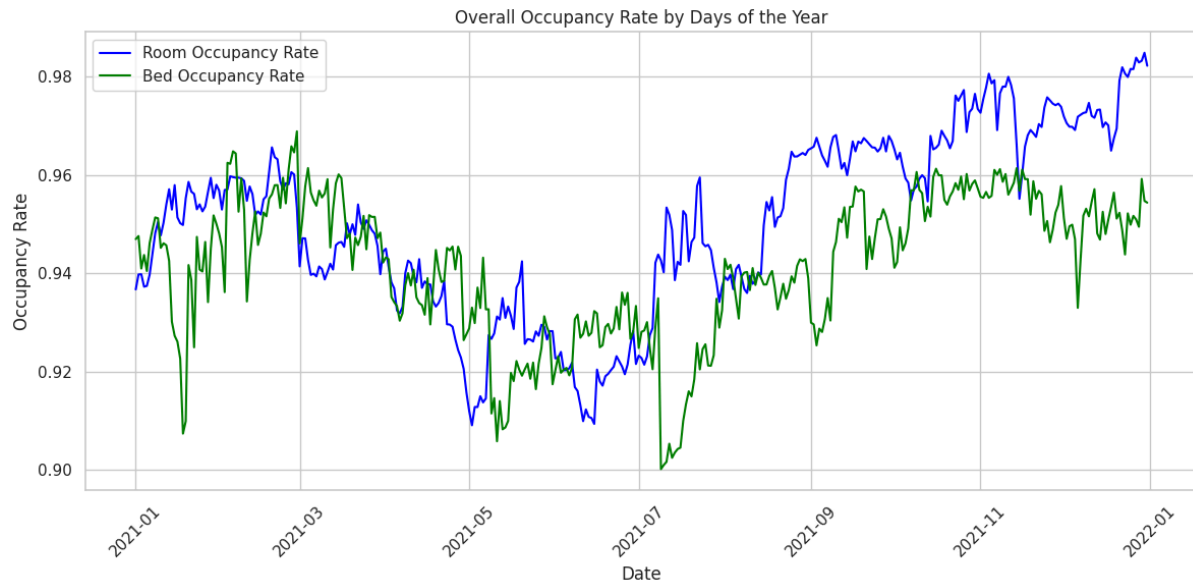


Figure 5

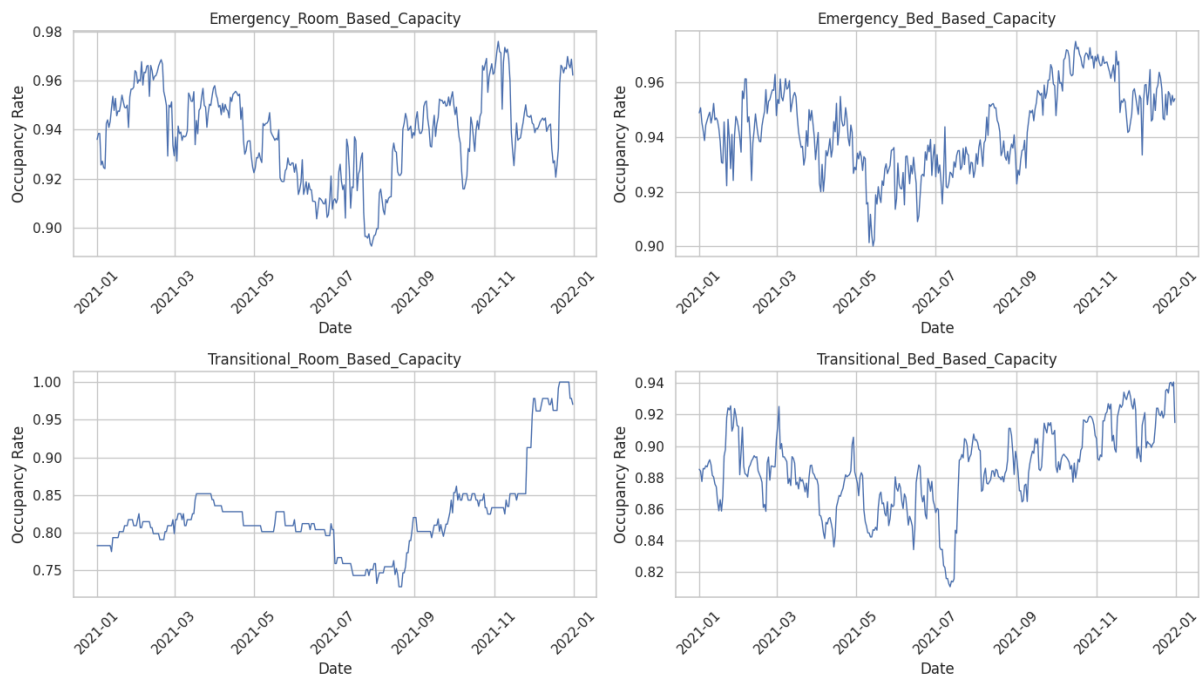


Figure 6

	Standard Deviation	Variance	Max	Min
Emergency_Room_Based_Capacity	0.163904	0.026864	1.014085	0.012048
Emergency_Bed_Based_Capacity	0.115901	0.013433	1.000000	0.022727
Transitional_Room_Based_Capacity	0.092578	0.008571	1.000000	0.666667
Transitional_Bed_Based_Capacity	0.129907	0.016876	1.000000	0.222222

Diving into demographic sectors, box plots underscore that most shelter types routinely approach or hit maximum capacity, with medians nearing 1.0. An exception is observed with

bed-based shelters for families, which notably exhibit lower occupancy rates, likely due to a preference among families for room-based accommodations (Figure 7).

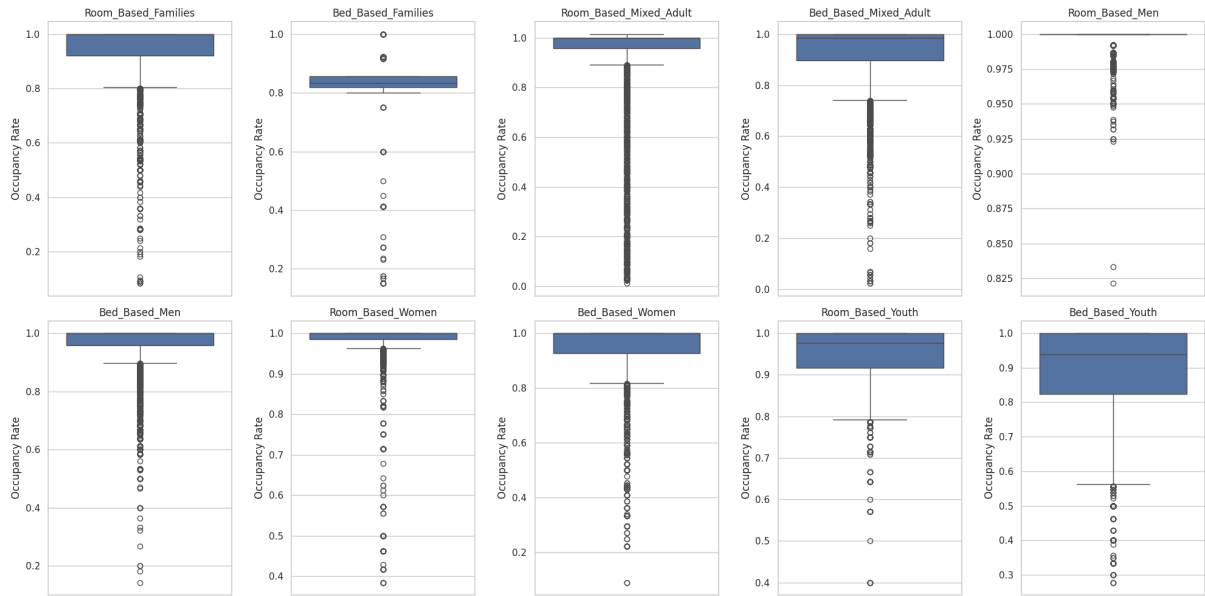


Figure 7, box plots of occupancy rates by demographic groups and capacity types

Statistical comparisons via t-tests between bed-based and room-based shelters across different sectors further illuminate the occupancy dynamics. The results, highlighted by exceptionally low p-values for each sector, confirm statistically significant differences in OCRs between shelter types within each demographic group. This underscores the profound influence of shelter capacity type on occupancy rates, affirming that both bed-based and room-based shelters cater distinctively to the diverse needs of various demographic groups.

	Sector	P-Value
0	Men	0.000000e+00
1	Women	5.961275e-38
2	Youth	1.496019e-133
3	Families	8.170315e-40
4	Mixed Adult	3.937029e-11

Shelter Utilization

This final section of the analysis delves into the relationship between the service user count and the capacity of shelters, with a focus on the user-to-bed/room ratios. These ratios, computed by dividing the service user counts by the corresponding capacity, offer insights into the utilization efficiency of shelters. Scatter plots for each day of the year reveal that bed-based shelters typically exhibit a user-to-bed ratio near 1.0, with occasional dips below 1.0, indicating instances of underutilization. Conversely, room-based shelters display two prominent linear trends in their user-to-room ratios: one cluster around a ratio of 3 and another around 1.2. Further investigation into room-based shelters with user-to-room ratios exceeding 1.5 shows that these facilities predominantly serve families, where rooms are allocated to family units of at least two members, thus elevating the user-to-room ratio.

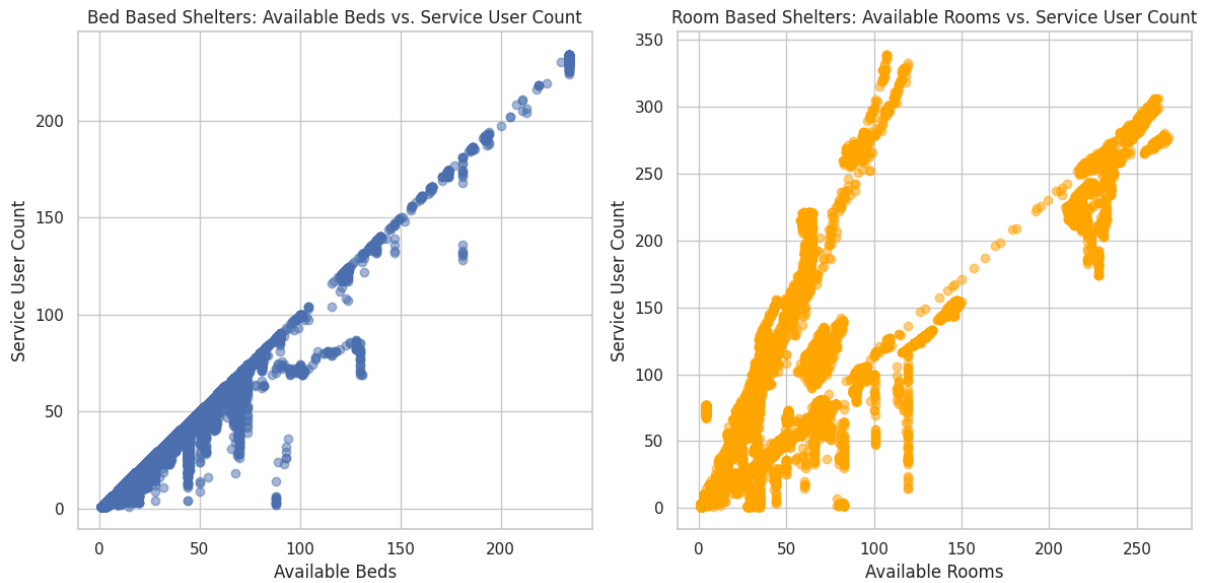


Figure 8, scatter plot of user-to-room ratio

To assess the impact of demographic categories on shelter utilization, t-tests were performed comparing the user-to-room/bed ratios between bed-based and room-based shelters across different sectors. The results, summarized in low p-values, underscore significant disparities in the user-to-room/bed ratios between the two types of shelters within each demographic sector. This underscores the critical role of shelter capacity type in determining the accommodation efficiency per unit, highlighting distinct utilization patterns across demographic groups.

Sector P-Value (User-to-Room/Bed Ratio)		
0	Families	0.000000e+00
1	Mixed Adult	8.045580e-268
2	Men	9.406820e-34
3	Women	5.465853e-111
4	Youth	1.496019e-133

Conclusion

In conclusion, this analysis of Toronto's shelter occupancy and utilization unveils several key insights. First of all, occupancy rates vary significantly with shelter capacity type, demographic sector, and season. Secondly, room-based shelters for families are often at full capacity and bed-based shelters experience underutilization. Statistical tests confirm these differences are significant across demographics, emphasizing the need for tailored shelter management strategies. This study highlights the necessity of adaptive, data-informed approaches to optimize shelter allocation and enhance the effectiveness of Toronto's shelter system in meeting diverse community needs.