

Assignment 1 Analysis

Name: Jingwen Ji

ID: 1006128101

In 2021, the city of Toronto faced significant challenges in managing its shelter system amidst ongoing social issues and the impacts of the COVID-19 pandemic. The shelter system, which is designed to provide temporary accommodation to individuals and families experiencing homelessness, became a focal point of public health and safety efforts. This official dataset captures key operational metrics of Toronto's shelters throughout the year, makes it possible for us to take a detailed view of the capacity and utilization of these critical resources.

The dataset includes information on the type of capacity (such as beds and rooms), the specific program model under which each shelter operates (such as Emergency or Transitional), the count of service users, and both the actual and occupied capacities. These variables are pivotal in understanding how well the shelter system is responding to the needs of its service users. The distinction between Emergency and Transitional shelters is particularly noteworthy.

Given the dataset's breadth, this analysis focuses on comparing the occupancy rates of beds and rooms between Emergency and Transitional shelters using Welch's t-tests. These statistical tests are chosen for their robustness in handling data with unequal variances and sample sizes, providing a reliable method for assessing differences in mean occupancy rates between the two shelter types. Specifically, the analysis aims to determine whether significant differences exist in how Emergency and Transitional shelters are utilized, as reflected by their bed and room occupancy rates.

First of all, I checked for missing values of specific columns in the dataset.

CAPACITY_ACTUAL_BED	18545
OCCUPIED_BEDS	18545
CAPACITY_ACTUAL_ROOM	32399
OCCUPIED_ROOMS	32399

This indicates that for a substantial number of records, the data on both the actual capacity and the occupancy of beds and rooms is unavailable. This absence of data can pose challenges to conducting a comprehensive analysis of the shelter system's utilization and capacity. The missing data can limit the scope of analysis, particularly in evaluating the utilization rates of shelters and identifying potential shortfalls or surpluses in shelter capacity. It might be necessary to address the missing data through techniques such as imputation or exclusion.

I use the summary statistics of the specific columns related to shelter capacities and occupancies in the Toronto shelter dataset for 2021.

index	SERVICE_USER_COUNT	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS	CAPACITY_ACTUAL_ROOM	OCCUPIED_ROOMS
count	50944.0	32399.0	32399.0	18545.0	18545.0
mean	45.72717101130653	31.627148986079817	29.780270996018395	55.54925856025883	52.79859800485306
std	53.326049257312924	27.127681515041054	26.3794161339531	59.4488052264729	58.79295410058777
min	1.0	1.0	1.0	1.0	1.0
25%	15.0	15.0	14.0	19.0	16.0
50%	28.0	25.0	23.0	35.0	34.0
75%	51.0	43.0	41.0	68.0	66.0
max	339.0	234.0	234.0	268.0	268.0

The dataset contains records for 50,944 service users, indicating the total number of individuals who utilized the shelter system throughout the year.

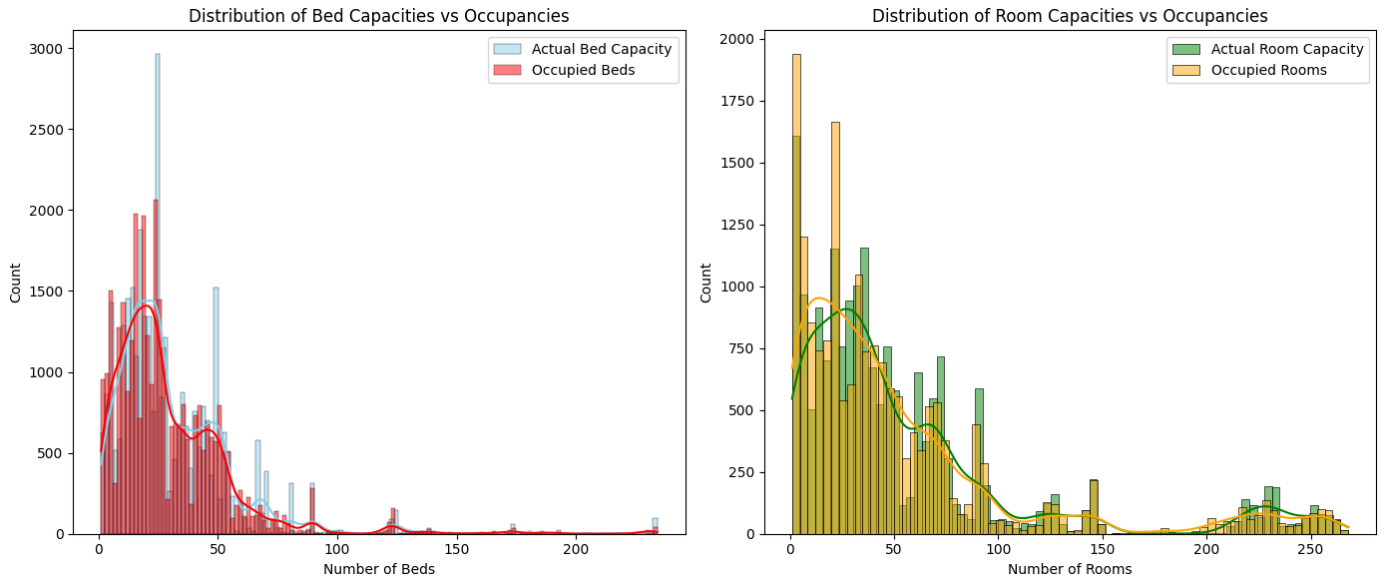
On average, shelters have a bed capacity of approximately 31.63 beds, with an average occupancy of 29.78 beds. This suggests that shelters are operating near their full bed capacity. The standard deviation values for both bed capacities (27.13) and occupancies (26.38) indicate a significant variation among shelters, with some facilities having much larger or smaller capacities and occupancies than others. The minimum value is 1 for both capacity and occupancy, indicating that the smallest shelters in the dataset have a capacity for only one bed. The maximum bed capacity and occupancy are both 234, showing that the largest shelters can accommodate up to 234 individuals.

The average room capacity in shelters is 55.55, with an average occupancy of 52.80 rooms. This also suggests high utilization of available room space. The standard deviation for room capacities (59.45) and occupancies (58.79) is even higher than for beds, indicating a greater disparity in room-based shelter sizes and utilization. Both the capacity and occupancy of rooms range from a minimum of 1 to a maximum of 268.

The quartiles for bed capacities and occupancies show a gradual increase, with a noticeable jump in the upper quartile, indicating that a significant number of shelters have higher capacities and occupancies.

The quartiles for room capacities and occupancies follow a similar pattern, with the median values indicating that half of the shelters have capacities and occupancies of 35 and 34 rooms or less, respectively.

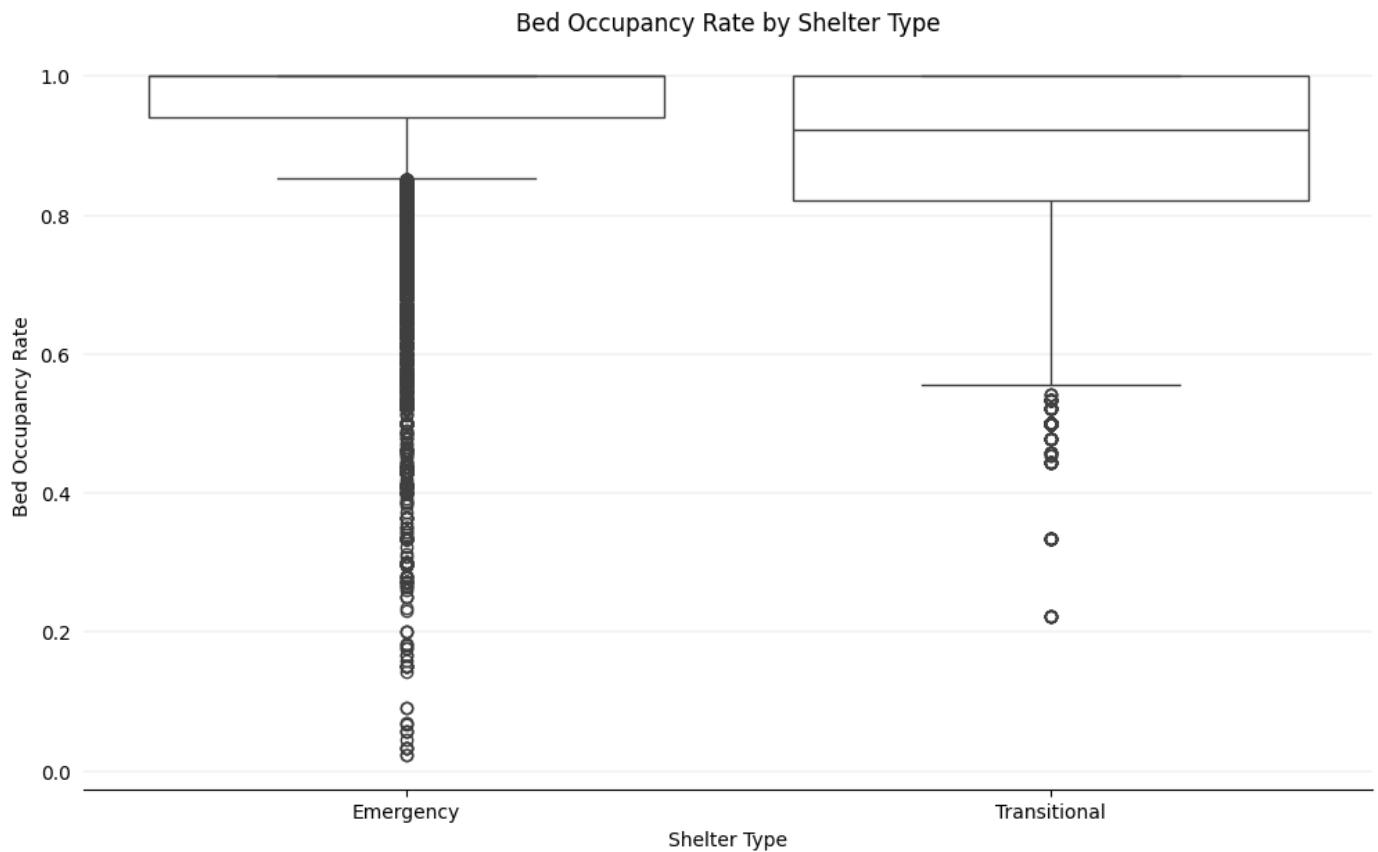
Histograms of distribution of bed and room capacities and occupancies



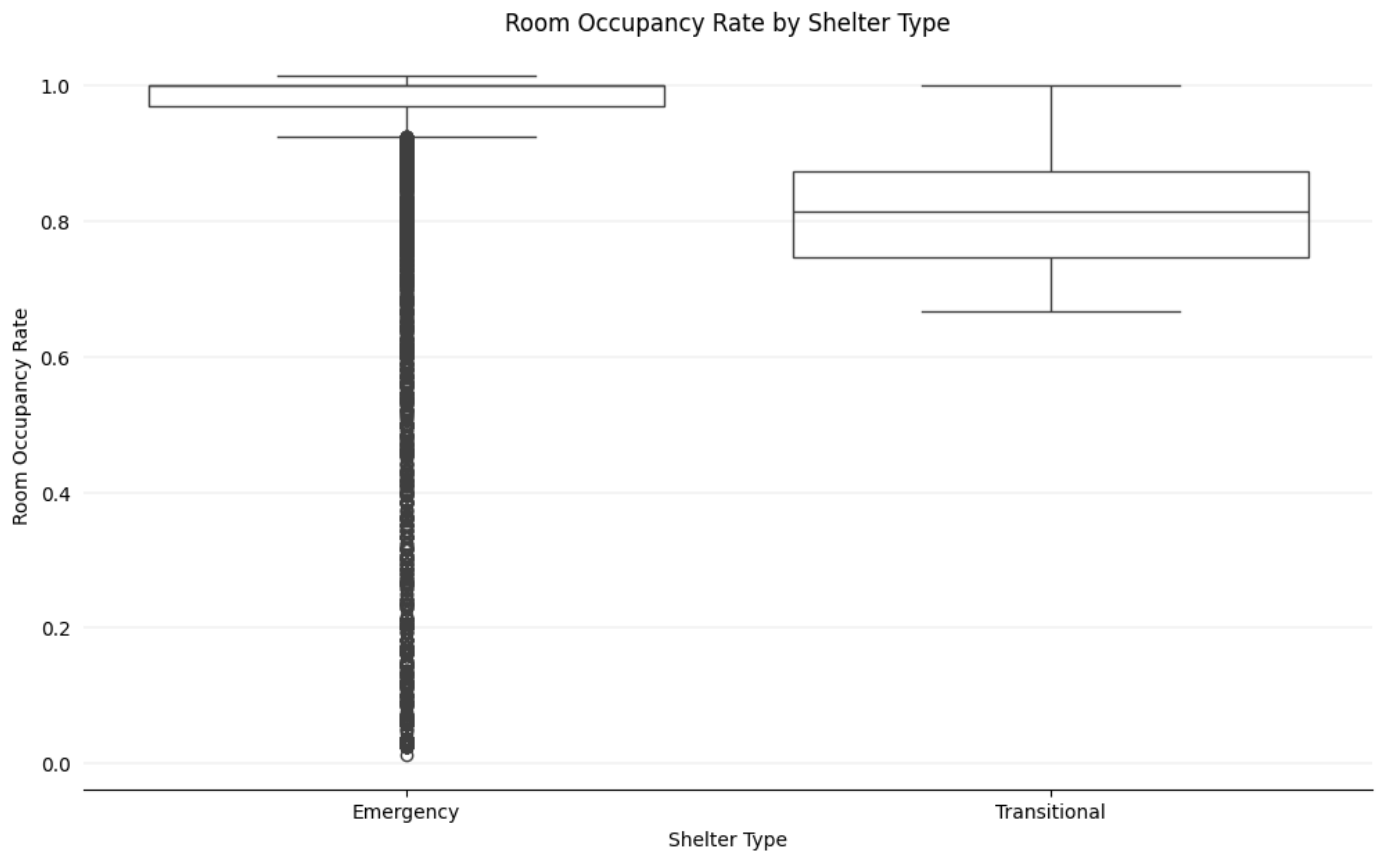
The graphs depict the distribution of bed and room capacities versus their occupancies in the shelter system.

For the graph of bed capacities vs occupancies, the blue histogram represents the actual bed capacity of shelters, while the red histogram shows the number of occupied beds. Both histograms appear to be right-skewed, indicating that a majority of shelters have a smaller number of beds, with fewer shelters having a larger bed capacity. The peak at the lower end of both distributions suggests that many shelters have a low bed capacity, with a rapid fall in the number of shelters as bed capacity increases. There is also a notable peak for actual bed capacity at around 25 beds, which may suggest a common shelter size.

For the graph of room capacities vs occupancies, the green histogram represents the actual room capacity, while the yellow histogram depicts the number of occupied rooms. These histograms are also right-skewed, indicating that most shelters have fewer rooms, with the count decreasing as the number of rooms increases. Similar to the bed capacities, there is a significant peak at the lower end, indicating many shelters have a limited number of rooms available and occupied.



The boxplot for Emergency shelters indicates that the median bed occupancy rate is high, close to 1.0, which means that many Emergency shelters are operating at or near full bed capacity. The range of occupancy rates in Emergency shelters is tight around the median, but there is a substantial number of outliers, indicating that some Emergency shelters have much lower occupancy rates. Transitional shelters have a wider interquartile range, indicating more variability in occupancy rates. The median is lower than in Emergency shelters, suggesting that, on average, Transitional shelters have a lower occupancy rate. There are also outliers for Transitional shelters, but they are less extreme than those for Emergency shelters.



For Emergency shelters, the room occupancy rate also has a median close to 1.0, implying that rooms in Emergency shelters are typically highly occupied. The distribution of room occupancy rates in Emergency shelters is similar to the bed occupancy rates, with a high median and a cluster of lower outliers. Transitional shelters show a lower median room occupancy rate compared to Emergency shelters, and the interquartile range is broader, suggesting greater variability in room occupancy rates among Transitional shelters. There are outliers in the Transitional shelter data, indicating some shelters with particularly low room occupancy rates.

Overall, emergency shelters, on average, tend to have high occupancy rates for both beds and rooms, often reaching full capacity. This suggests a strong demand for immediate, short-term shelter options. Transitional shelters display more variability in occupancy rates and generally have lower median occupancy rates than Emergency shelters. This could reflect the programmatic nature of Transitional shelters, where the duration of stay may be longer, and the need for specialized services may result in a more diverse occupancy pattern. The presence of outliers, especially in the Emergency shelter data, indicates that there are exceptions to the general trend of high occupancy, which needs further investigation.

Welch's t-test for Bed Occupancy Rate:

Transitional vs. Emergency: t-stat = -36.78483679745313, p-value = 7.273950955976339e-283

Welch's t-test for Room Occupancy Rate:

Transitional vs. Emergency: t-stat = -31.71080126309493, p-value = 4.4252019739840735e-150

The results of the Welch's t-tests for both bed and room occupancy rates between Transitional and Emergency shelters are highly statistically significant, indicated by the extremely small p-values in both cases.

For bed occupancy rate, the t-statistic of approximately -36.78 for bed occupancy rates suggests a significant difference in the occupancy rates between Transitional and Emergency shelters, with Emergency shelters having higher occupancy rates. The negative sign of the t-statistic indicates that the mean occupancy rate for Transitional shelters is lower than for Emergency shelters.

For room occupancy rate, the t-statistic of approximately -31.71 for room occupancy rates also suggests a significant difference between Transitional and Emergency shelters, with Emergency shelters showing higher occupancy rates. The negative t-statistic indicates that the mean occupancy rate for rooms in Transitional shelters is lower than in Emergency shelters.

Both of the p-values are extremely close to zero, indicating a very high level of statistical significance for the difference in bed and room occupancy rates.

These t-test results strongly suggest that Emergency shelters are operating at or near full capacity more consistently than Transitional shelters, reflecting the immediate and urgent need for housing that Emergency shelters fulfill. The much lower occupancy rates in Transitional shelters could be due to a variety of factors, such as the programmatic nature of these shelters, which may have specific entry criteria and services that lead to a more controlled intake process and potentially longer stays. The findings may also point to capacity and demand mismatches in the shelter system, with Emergency shelters possibly being under higher pressure, which might impact their ability to provide services effectively. For policymakers and service providers, these results underscore the need to address the challenges within Emergency shelters and to reevaluate the resources and support provided to Transitional shelters to optimize their utilization.