

Narrative Report on Data Analysis

Introduction

The analysis below aimed at looking into the rates of occupation together with bed capacities in different types of facilities by use of the dataset provided within TNF2178_A1_data. The objective here was to reveal potential discrepancies among different levels of capacity and bed setups under separate service groups.

Data Preparation and Cleaning

I began by loading the necessary Python libraries including NumPy, Pandas, Scipy and Matplotlib for processing data. The appropriate steps for data preparation entailed generating the rates of occupancy, both bed-based and room capacity, in addition to eliminating any null.

The t-test assumes that the populations from which the samples are normally distributed. However, if the sample size is large, it's not necessary to verify normality due to the Central Limit Theorem, which states that as the sample size increases, the distribution of the sample means approaches a normal distribution. Therefore, the data in this case meet the conditions for conducting a t-test.

Research Questions:

- What is the distribution of occupancy across facility capacity types and how does it vary from one to another?
- Do the averages of occupied rooms between these classes differ in a statistically significant way?
- What patterns of distribution in both occupancy rates and bed capacities provoke further thoughts, especially regarding specific types of services such as emergency or transitional services?

Occupancy Rate Analysis

After cleaning the data, I divided it into two categories: 'Bed Based' and 'Room Based'. To gain further insight from the data, I also performed calculations of measures such as mean, median, minimum value and maximum value for each kind of occupancy rate.

```
Bed Based
Occupancy Rates Statistics:
Min: 0.02
Max: 1.0
Median: 1.0
Mean: 0.93
```

```
Room Based
Occupancy Rates Statistics:
Min: 0.01
Max: 1.01
Median: 1.0
Mean: 0.93
```

Capacity Type Comparison

Furthermore, I compared the capacities between different types of services namely Hotel/Motel Shelter vs. Shelter, and Emergency vs. Transitional). Having calculated the summary statistics for each group.

```
Motel/Hotel Shelter Occupancy Rates Statistics:
Min: 42.0
Max: 131.0
Median: 98.0
Mean: 97.85
```

```
Shelter Occupancy Rates Statistics:
Min: 1.0
Max: 234.0
Median: 24.0
Mean: 29.66
```

```
Emergency Occupancy Rates Statistics:
Min: 1.0
Max: 234.0
Median: 25.0
Mean: 33.83
```

```
Transitional Occupancy Rates Statistics:
Min: 2.0
Max: 83.0
Median: 24.0
Mean: 25.59
```

Welch's T-test and P-value Analysis

I conducted corresponding Welch's t-tests by setting the null hypothesis and alternative hypotheses first:

1. For Bed Based vs. Room Based Occupancy Rates:

H0: There is no significant difference in the occupancy rates between bed-based and room-based accommodations.

Ha: There is a significant difference in the occupancy rates between bed-based and room-based accommodations.

```
Bed Based  
Occupancy Rates Statistics:  
Min: 0.02  
Max: 1.0  
Median: 1.0  
Mean: 0.93
```

```
Room Based  
Occupancy Rates Statistics:  
Min: 0.01  
Max: 1.01  
Median: 1.0  
Mean: 0.93
```

```
t-statistic = -4.498751771925636  
p-value = 6.860477551487939e-06  
Degrees of freedom = 18544
```

2. For Actual Bed Capacity between Motel/Hotel Shelter and Shelter:

H0: There is no significant difference in the actual bed capacity between Motel/Hotel Shelter and Shelter accommodations.

Ha: There is a significant difference in the actual bed capacity between Motel/Hotel Shelter and Shelter accommodations.

```
t-statistic = 37.19940662556716  
p-value = 1.029190825441981e-85  
Degrees of freedom = 175  
There is a significant difference in actual bed capacity between Motel/Hotel Shelter and Shelter.
```

3. For Actual Bed Capacity between Emergency and Transitional:

H0: There is no significant difference in the actual bed capacity between Emergency and Transitional accommodations.

Ha: There is a significant difference in the actual bed capacity between Emergency and Transitional accommodations.

```
t-statistic = 30.04557995319668  
p-value = 8.12328951219356e-195  
Degrees of freedom = 8670  
There is a significant difference in actual bed capacity between Emergency and Transitional.
```

For each Welch's t-test, I calculated the p-value to determine the statistical significance of the differences between groups. As the results are shown above, P-values are calculated as 6.860477551487939e-06, 1.029190825441981e-85 and 8.12328951219356e-195, all three results are less than 0.05, so we can reject the null hypothesis and conclude that there is a significant difference in the mean values between each two independent variables.

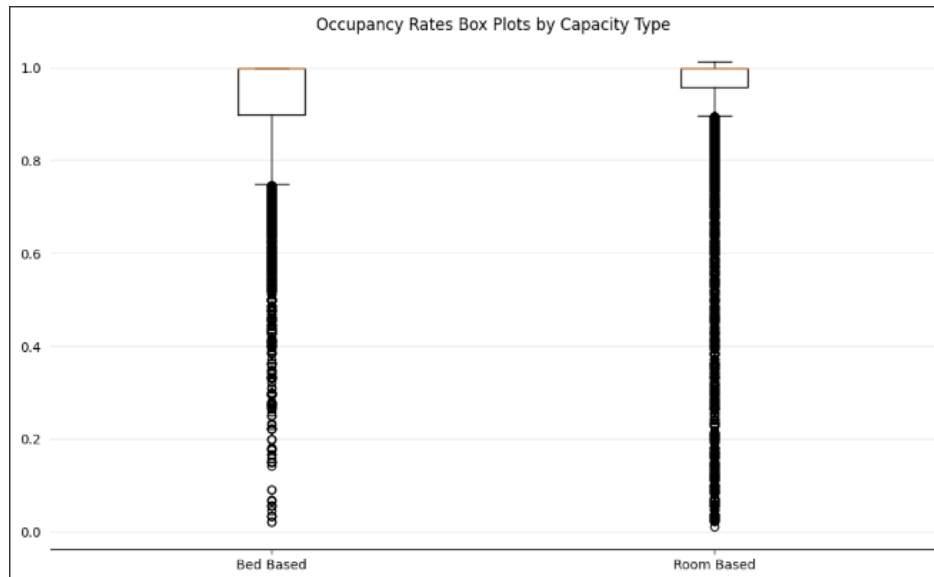
Visualization of Results

The boxplots are good visual representations of the occupancy rates and the actual bed capacities. They offer a precise and easily accessible way to compare medians, ranges, and outlier presence between different groups.

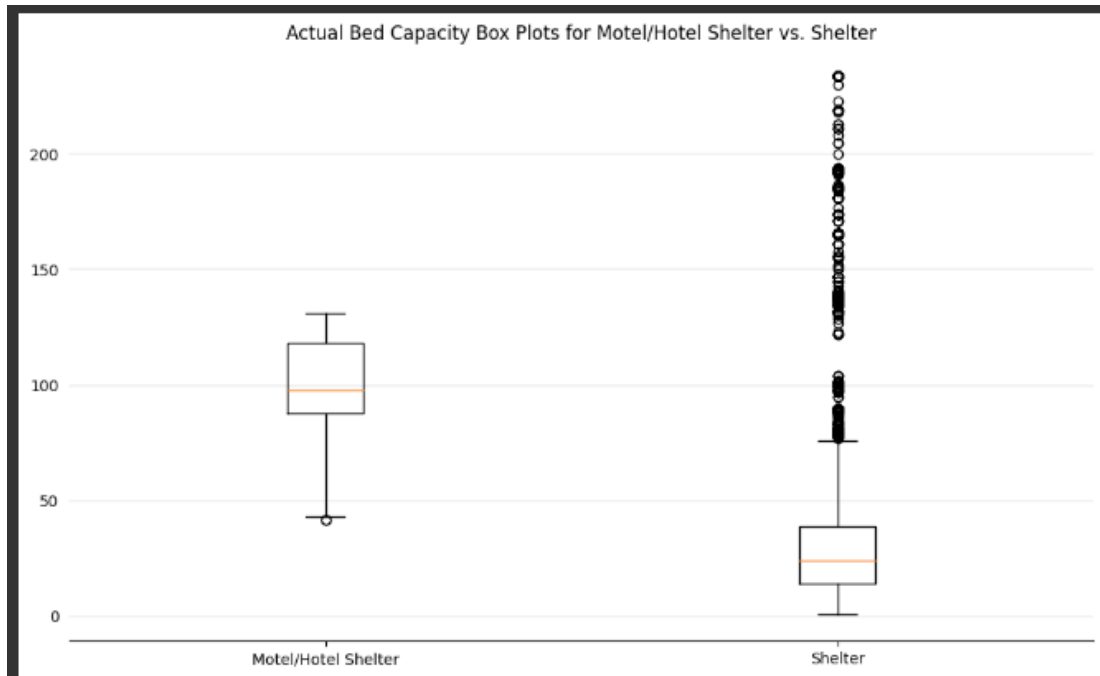
In these boxplots, the central line denotes the median, the box indicates IQR, and the whiskers encompass other parts of the distribution, excluding outliers determined by a highly specific

method based on the IQR. Outliers, however, are very important to the presence of data, providing information on extremes that might be due to special cases or data entry errors, and they are shown as individual points leaving the whiskers.

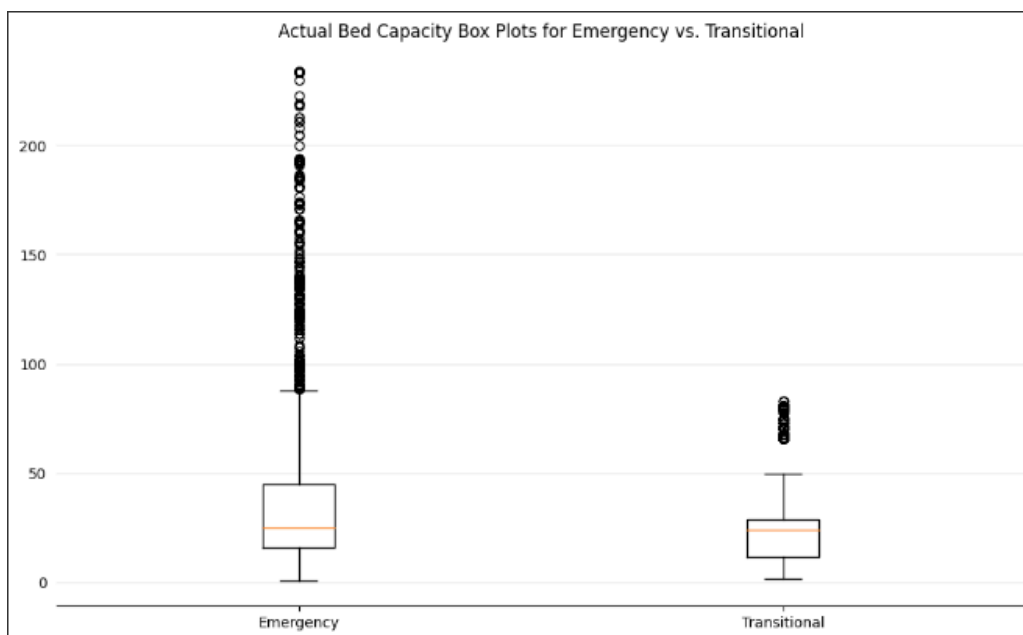
With the help of this visualization method, stakeholders can see statistical summaries and within-category variability, which is instrumental when making decisions related to resource allocation, capacity planning, and understanding how different types of accommodation are behaving.



The first boxplot displays bed-based and room-based occupation rates. All the values of the 'Bed-based' occupancy rates are close to rather low values with no outlier, which means that the rates are steady and constantly low. The 'room-based' occupancy rates show a better median with the 75th percentile reaching 1.0, indicating that large numbers of the room base accommodation are all fully occupied. There are outliers suggesting that some of the room-based accommodations have relatively low occupancy rates.



This box plot shows a distinct difference in the real bed capacity for Motel/Hotel Shelters compared to general Shelters. The median bed capacity of Motel/Hotel Structures is higher, and the interquartile range (IQR) is placed at high values suggesting that they typically provide more beds. There are several outliers for the Motel/Hotel Shelters, some being those with very large bed capacities. In contrast, the median and IQR of Shelters are much lower, with lower outliers, suggesting a generally lower capacity of beds.



The third boxplot compares the real bed capacities of emergency and transitional facilities. Median bed capacity appears to be higher in emergency facilities than in transitional facilities. Nevertheless, the boxplot for Emergency facilities has a large spread which shows a significant

level of variability for the bed capacity. The IQR for transitional facilities is shorter, indicating greater uniformity in their capacities, but also has unspoken outliers on both the high and low ends.

Insights and Conclusions

Through detailed data analysis and statistical testing, I observed:

The occupancy rates of the facilities based on 'Bed Based' and 'Room Based' are significantly different from each other.

'Emergency' and 'Transitional' also differ quite significantly in terms of actual bed capacities.

The p-values obtained from all the comparisons are less than 0.05, which means that the statistical significance of differences observed is not due to chance variations as indicated by these values.

The results discussed in the paper have significant managerial implications for resource use and operational efficiencies of services. For example, the high rate of bed demand in emergency service categories shows that these facilities are well-prepared to handle urgent cases promptly. Furthermore, these analysis findings can be used in the measurement and practice of enhancing bed utilization rates as channels towards productivity service delivery.