

## INF218 - Assignment 1

### Exploring Homeless Shelter Occupancy Data in Toronto

#### *Dataset:*

This exploration dives into the homeless shelter occupancy data in Toronto, aiming to gain insights into the patterns and dynamics of shelter usage. In recent years, Toronto has seen a marked increase in the city's homeless population. Although the City of Toronto's shelter support system provides overnight and shelter services for the unhoused, individuals seeking shelter are often being turned away as there are often not enough shelter spaces. The dataset provides information on various aspects, including occupancy rates, program models, and service user counts.

#### *Research Topic:*

Through this exploratory data analysis, we sought to examine the daily occupancy and capacity of homeless shelters in the year 2021. From the analysis, we hope to determine various aspects of Toronto's current shelter program that could be improved, provide insights into homeless shelter services, and support the future improvement of homeless shelter spaces in Canada.

When assessing the data, several key columns were selected including [ capacity—type, program—model, service—user—count, capacity—actual—bed, occupied—beds, capacity—actual—room, occupied—rooms] and two additional variables describing shelter program occupancy rates. (*Exhibit 1*)

#### *Research Questions:*

1. What are the program models' impact on occupancy?
2. How do occupancy trends change throughout the year? (time analysis)
3. How does utilization differ based on capacity type? (*Exhibit 2*)

#### *Research Goals:*

1. I sought to understand if there were significant differences in occupancy rates between the 'Emergency' and 'Transitional' program models. The analysis revealed a substantial dissimilarity, indicating each model's unique utilization patterns and user characteristics.
2. I aimed to uncover monthly or seasonal trends in shelter occupancy and identify any recurring patterns or fluctuations.
3. I explored whether there were variations in occupancy rates between programs with bed-based and room-based capacities. (*Exhibit 2*)

#### *Data Overview (Exhibit 1):*

| #   | Column                 | Non-Null Count | Dtype          |
|-----|------------------------|----------------|----------------|
| --- | -----                  | -----          | -----          |
| 0   | OCCUPANCY_DATE         | 50944 non-null | datetime64[ns] |
| 1   | ORGANIZATION_NAME      | 50944 non-null | object         |
| 2   | PROGRAM_ID             | 50944 non-null | int64          |
| 3   | PROGRAM_NAME           | 50909 non-null | object         |
| 4   | SECTOR                 | 50944 non-null | object         |
| 5   | PROGRAM_MODEL          | 50942 non-null | object         |
| 6   | OVERNIGHT_SERVICE_TYPE | 50942 non-null | object         |

```

7  PROGRAM_AREA                50942 non-null object
8  SERVICE_USER_COUNT          50944 non-null int64
9  CAPACITY_TYPE               50944 non-null object
10 CAPACITY_ACTUAL_BED          32399 non-null float64
11 OCCUPIED_BEDS               32399 non-null float64
12 CAPACITY_ACTUAL_ROOM        18545 non-null float64
13 OCCUPIED_ROOMS              18545 non-null float64
14 OCCUPANCY_RATE_BEDS         32399 non-null float64
15 OCCUPANCY_RATE_ROOMS       18545 non-null float64

```

The following descriptive statistics provide a snapshot of the distribution of key variables, such as service user count, actual bed and room capacities, occupied beds and rooms, and occupancy rates.

#### Key Points

- Wide range of values (1 to 339), indicating variability in the number of service users across facilities.
- Minimal differences between actual and occupied shelter spaces suggest that the facilities are operating near full capacity most of the time.
- The IQR values for Service User Count, Actual Bed Capacity, and Occupied Beds indicate varied utilization patterns among different shelter programs.
- High mean values for occupancy rates imply efficient utilization but could also indicate shelters are unable to meet high demand points

#### Summary Statistics (Exhibit 2):

| Service User Count   | Actual Bed Capacity   | Occupied Beds   | Actual Room Capacity   | Occupied Rooms   | Occupancy Rate for Beds   | Occupancy Rate for Rooms  |
|--|---|---|--|--|---|---|
| Min: 1<br>Mean: 45.73<br>Max: 339<br>25th percentile: 15.0<br>Median: 28.0<br>75th percentile: 51.0<br>Interquartile range (IQR): 36.0 | Min: 1.0<br>Mean: 31.63<br>Max: 234.0<br>25th percentile: 21.0<br>Median: 31.63<br>75th percentile: 31.63<br>Interquartile range (IQR): 10.63 | Min: 1.0<br>Mean: 29.78<br>Max: 234.0<br>25th percentile: 19.0<br>Median: 29.78<br>75th percentile: 29.78<br>Interquartile range (IQR): 10.78 | Min: 1.0<br>Mean: 55.55<br>Max: 268.0<br>25th percentile: 55.55<br>Median: 55.55<br>75th percentile: 55.55<br>Interquartile range (IQR): 0.0 | Min: 1.0<br>Mean: 52.8<br>Max: 268.0<br>25th percentile: 52.8<br>Median: 52.8<br>75th percentile: 52.8<br>Interquartile range (IQR): 0.0 | Min: 0.02<br>Mean: 0.93<br>Max: 1.0<br>25th percentile: 0.93<br>Median: 0.93<br>75th percentile: 1.0<br>Interquartile range (IQR): 0.07 | Min: 0.01<br>Mean: 0.93<br>Max: 1.01<br>25th percentile: 0.93<br>Median: 0.93<br>75th percentile: 0.98<br>Interquartile range (IQR): 0.05 |

#### Quantitative Data Analysis (T-Test)

The t-test was conducted to assess whether there are significant differences in the mean occupancy rates and service user counts between the 'Emergency' and 'Transitional' program models. Similar to occupancy rates, the t-test for service user count also aimed to determine whether there were significant differences in the mean counts between program models. The two-sided t-test was chosen because it allows for the

investigation of differences in both directions, considering whether the mean occupancy rates and service user counts are either higher or lower.

The very low p-values (close to zero) in both cases led to the rejection of the null hypothesis, suggesting a significant difference in occupancy rates between program models.

Welch's t-test was also utilized for the overall occupancy rate. The obtained t-statistic and p-value were used to assess whether there was a significant difference in the mean overall occupancy rate between the 'Emergency' and 'Transitional' program models. Again, the extremely low p-value led to rejecting the null hypothesis, indicating a significant difference.

*T-test Statistics:*

*Occupancy rates:*

```
Mean Occupancy Rates (Beds) for Emergency: 0.943500571658339
Mean Occupancy Rates (Beds) for Transitional: 0.8851944654528656
Mean Occupancy Rates (Rooms) for Emergency: 0.9386299840456679
Mean Occupancy Rates (Rooms) for Transitional: 0.8232070500372471
T-statistic for Occupancy Rates (Beds): 36.78483679745313
P-value for Occupancy Rates (Beds): 7.273950955976339e-283
T-statistic for Occupancy Rates (Rooms): 31.71080126309493
P-value for Occupancy Rates (Rooms): 4.4252019739840735e-150
Reject the null hypothesis for Occupancy Rates (Beds): there is a significant
difference between program models.
Reject the null hypothesis for Occupancy Rates (Rooms): there is a significant
difference between program models.
```

*Occupancy rate:*

```
Welch's t-statistic = 36.78483679745313
p-value = 7.273950955976339e-283
```

*Service User Count:*

```
Mean Service User Count (Emergency): 49.06422570472545
Mean Service User Count (Transitional): 30.988724603765558
T-statistic for Service User Count: 38.85174699254652
P-value for Service User Count: 0.0
Reject the null hypothesis: there is a significant difference in the mean of
Service User Count between 'Emergency' and 'Transitional' program models.
```

### **Exploratory Data Analysis (EDA)**

To further investigate the difference between program model types, several visualizations were created to determine the quantitative reasoning behind the difference between the 'Emergency' and 'Transitional' program models. Furthermore, an investigation was conducted into the number of occupied beds throughout the year 2021, occupancy rates by sector, overnight service types, capacity types, and program model counts.

#### *Scatter Plot of Actual Capacity vs Occupied Beds*

A scatter plot was created to visualize the relationship between the actual capacity of beds and the number of occupied beds. The plot indicates shelter spaces with higher bed counts often had higher occupation rates than smaller shelter spaces.

#### *Boxplot for Program Model Analysis*

A boxplot was generated to compare occupancy rates between different program models. From the boxplot, we're able to understand that 'Emergency' shelter spaces often were larger spaces with a higher number of occupied beds.

#### *Histogram for Service User Count Distribution*

A histogram was utilized to examine the distribution of service user counts. From the histogram, we can see most shelters in Canada were built to house >50 residents. This insight could be used alongside efficiency (occupation rate %) and financial data to determine if building larger shelter spaces in the future may improve shelter efficiency.

#### *Temporal Analysis - Line Plot*

A line plot was created for temporal analysis, focusing on the trend of occupied beds throughout 2021. From the line plot, we're able to see a continuous increase in the number of occupied beds from 26 in April 2021, gradually rising month by month to 34 in January 2022.

#### *Boxplot for Sector Analysis*

A boxplot was employed to analyze the distribution of occupied beds across different sectors. The boxplot shows the highest shelter utilization in mixed adult or men's spaces, providing a reason to further research users in that area.

#### *Program Model Analysis - Boxplot and Violin plot*

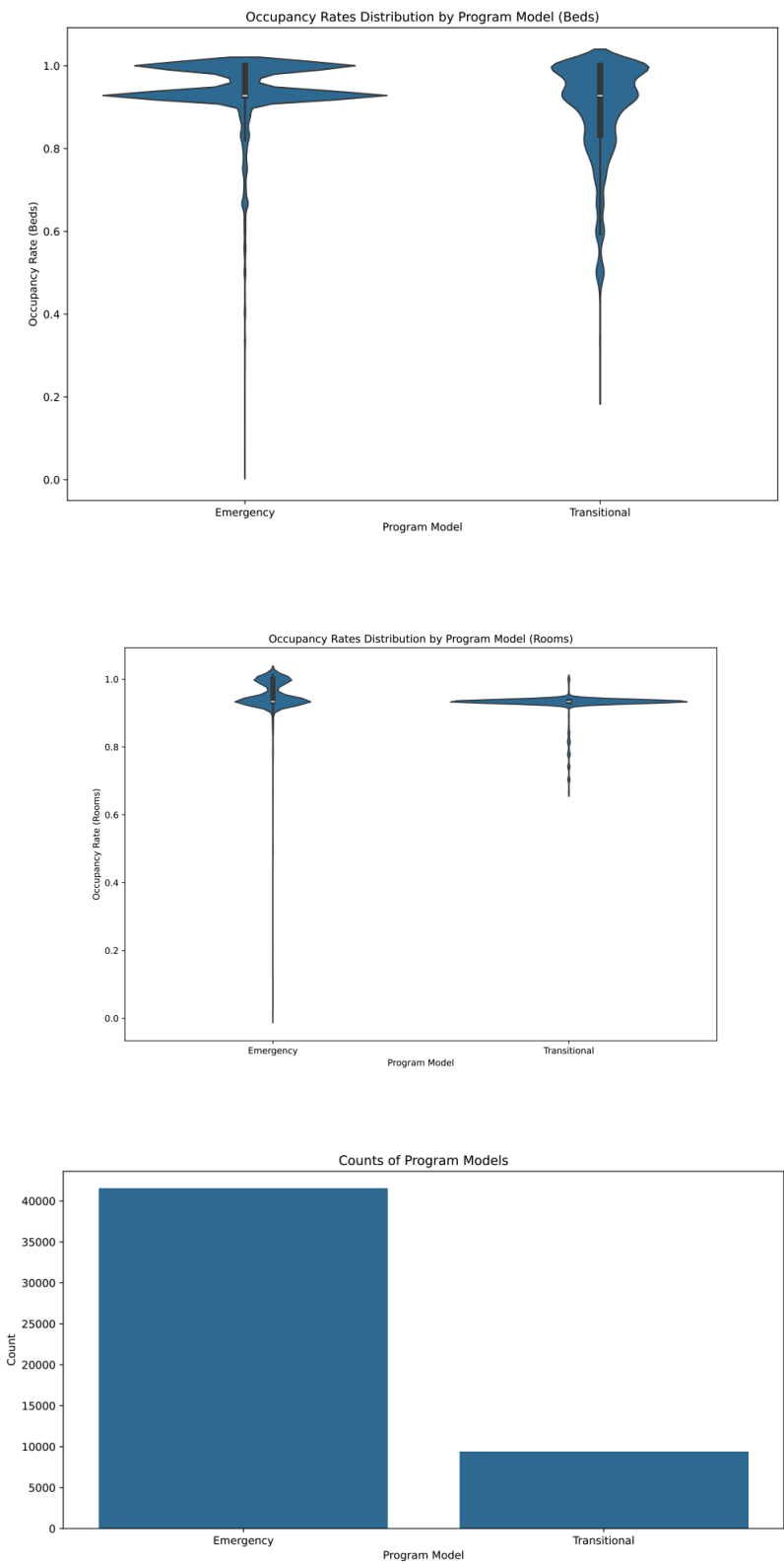
Boxplots and violin plots were employed to analyze the distribution of occupancy rates for beds across different program models. From the plots, we're able to see high occupation rates (near 100%) for both program models but a higher range among transitional models which are often operating near 80% capacity.

#### *Barplot for Program Model Count*

A barplot was generated to display the count of different program models showing a significant difference between the utilization of 'Emergency' and 'Transitional' services.

From the several graphs that were created, I was able to recognize clear differences between 'Emergency' and 'Transitional' services. Although transitional programs offered more complex services to residents, they were often not near full capacity while many emergency services likely had to turn away residents. A discussion to be had could be in the further development of emergency services or the utilization of transitional services by emergency residents during periods to high utilization.

Exploratory Data Analysis (EDA) Figures:



**Conclusion:**

For further analysis, there are several research topics/recommendations I would like to research further to have a stronger understanding of shelter spaces in Canada including:

- Conducting a detailed analysis of temporal patterns to identify the factors influencing fluctuations in occupancy (economics, immigration, weather patterns, etc.).
- Explore demographic data to understand the user characteristics contributing to differences in program models.
- Collaborate with shelters in specific sectors or service types to tailor interventions and improve operational efficiency.
- Financial costs associated with different shelter spaces and the development of new spaces versus the improvement of existing shelter spaces