INF2178 Technical Assignment 1

Name: KA YUEN LEE

Student Number: 1010073974

Professor: Shion Guha

Data Examination

In this data analysis task, I am interested in analyzing the following columns:

OCCUPANCY_DATE, ORGANIZATION_NAME, SECTOR,PROGRAM_MODEL, SERVICE_USER_COUNT,CAPACITY_TYPE,CAPACITY_ACTUAL_BED,OCCUPIED_BED, CAPACITY_ACTUAL_ROO, OCCUPIED_ROOMS.

For this analytical project, my first step will be to filter and examine any missing data across all categories.

```
    OCCUPANCY_DATE              0
    ORGANIZATION_NAME           0
    PROGRAM_ID                  0
    PROGRAM_NAME               35
    SECTOR                      0
    PROGRAM_MODEL               2
    OVERNIGHT_SERVICE_TYPE      2
    PROGRAM_AREA                2
    SERVICE_USER_COUNT          0
    CAPACITY_TYPE               0
    CAPACITY_ACTUAL_BED     18545
    OCCUPIED_BEDS           18545
    CAPACITY_ACTUAL_ROOM    32399
    OCCUPIED_ROOMS          32399
    dtype: int64
```

Summary of Missing Data

In the data examination section, missing values were encountered in several key variables. To address this, a decision was made to impute these missing values with zeros using the fillna(0) method.

Exploratory Data Analysis (EDA)

For the following analysis, I will focus on 3 research questions:

1. What are the factors influencing occupancy rates in shelters, and how do they vary by capacity type, demographic groups, service provider?
2. Does the type of program model (emergency vs. transitional) significantly affect the occupancy rate of housing services?
3. Are room-based capacities more effective than bed-based capacities in serving a higher number of users in housing services?

Research question 1:

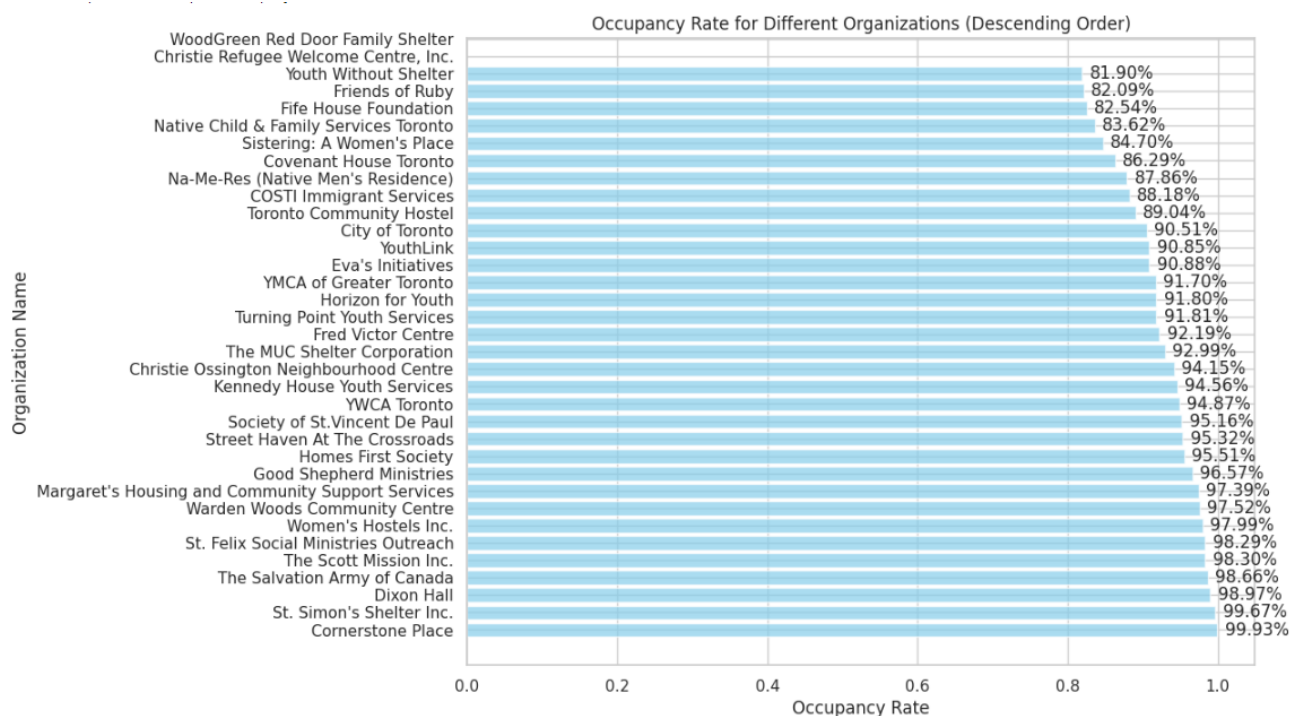| CAPACITY_ACTUAL_BED	OCCUPIED_BEDS	CAPACITY_ACTUAL_ROOM	OCCUPIED_ROOMS	OCCUPANCY_RATE_BEDS	OCCUPANCY_RATE_ROOMS	OCCUPANCY_RATE
NaN	NaN	29.0	26.0	NaN	0.896552	0.896552
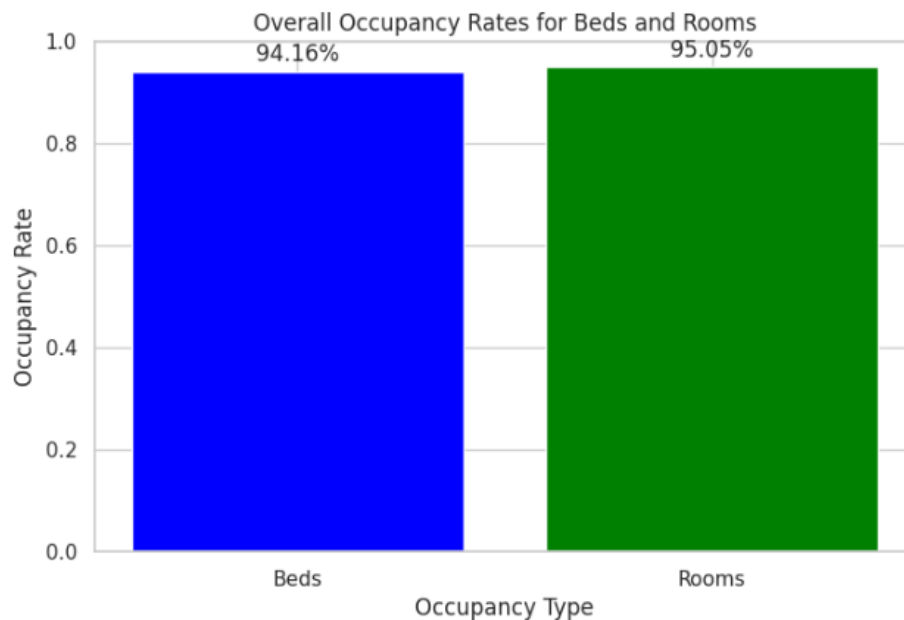
In this analysis, I created new columns to show the Occupancy Rate for Beds and Rooms. The calculated overall occupancy rates for beds and rooms are 94.16% and 95.05%, respectively. It shows that the usage of rooms is higher than beds, and high occupancy rates may indicate that shelters are operating at or near capacity, potentially facing challenges in meeting the demand for shelter services.
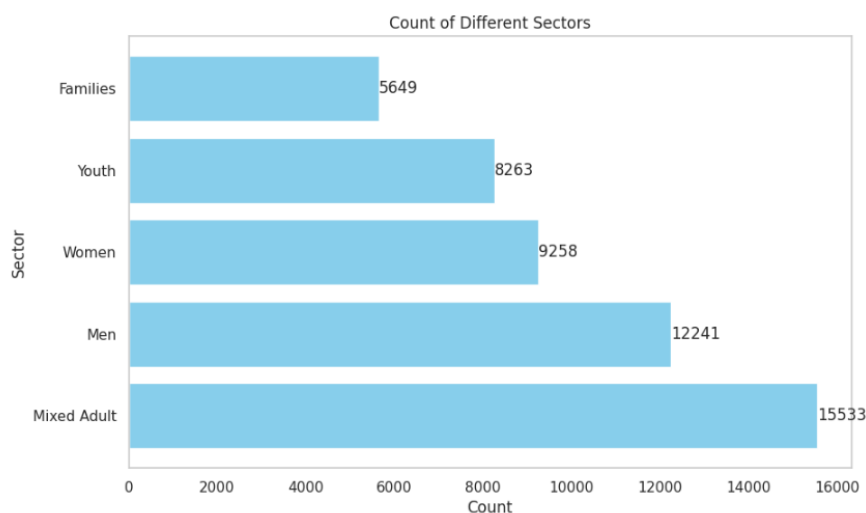

For this figure, mixed adults are the most numerous. There are more males than females, which means that more males are seeking shelter.

Frequency of Different Organization
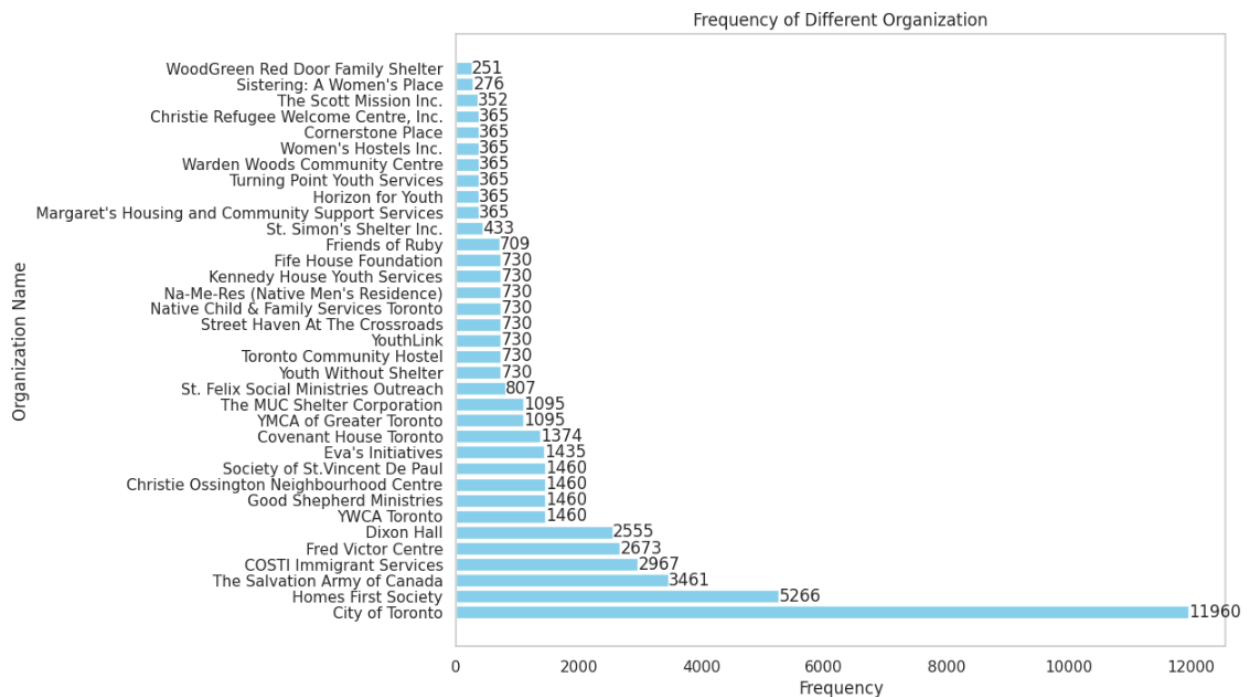
From the graph, we can see that The City of Toronto provides the most services.

Occupancy Rate for Different Organizations (Descending Order)

To know the occupancy rate of beds and rooms in each organization, I created this graph. From this graph, we can see that most of the utilization rates are over 90%, and in two organizations it is over 99% (St. Simon's Shelter Inc. and Cornerstone Place). This means that the occupancy rate is very high and may exceed the capacity of the organization. Therefore, the relevant organizations or the government may consider increasing the resources to prevent some users from not being able to find shelter.

Average Occupancy Rate for Different Months
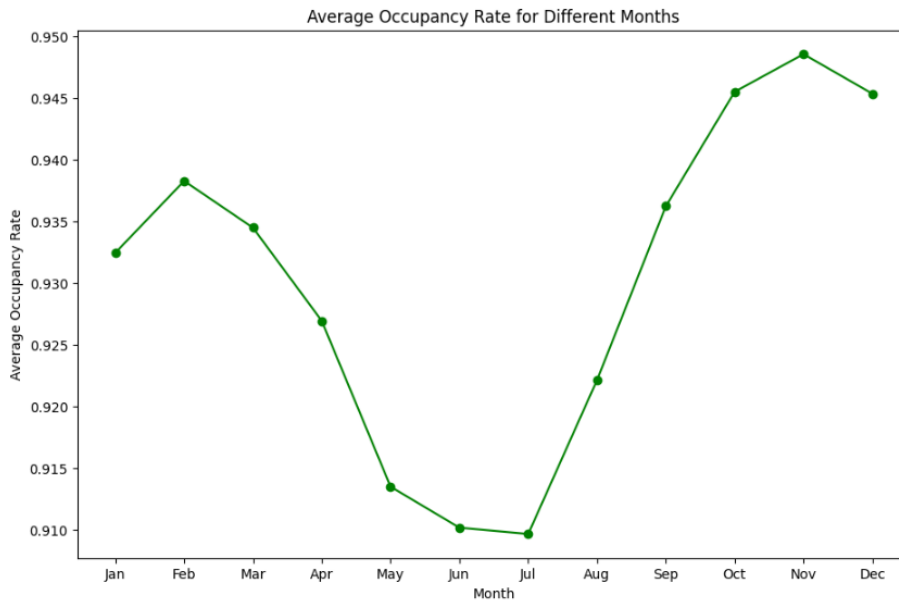
This graph shows the trend in occupancy rates for different months, and it shows how the demand for shelters may change depending on the season. For example, during the winter months (January to March), occupancy stays above 93%. However, during the summer months (May to July), occupancy drops to around 91%. This means that more people may need to seek shelter services due to the cold weather. Therefore, the organization may need to arrange more resources and services for them during the winter.

Research question 2:

Boxplot of Occupancy Rates for Program Models
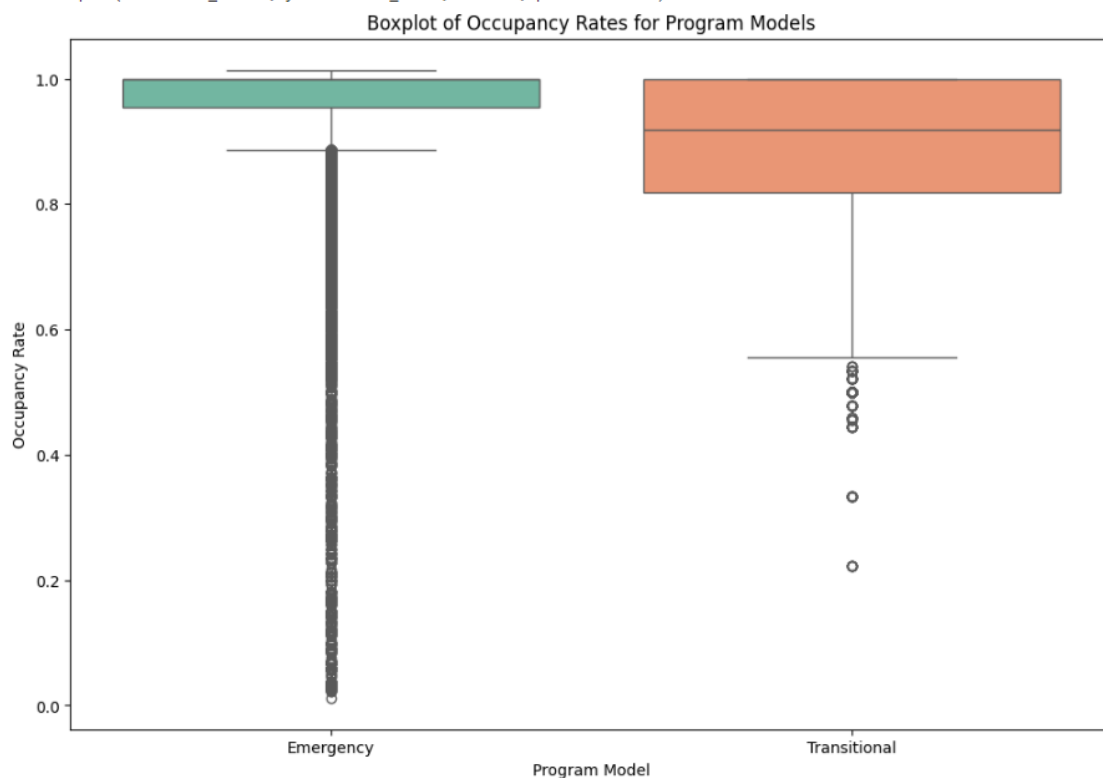
In the boxplot, the Emergency program model shows a very tight interquartile range (IQR) near the top of the scale, indicating that most of the occupancy rates are high and very consistent. There are no outliers, and the median is at or very close to 1.0, which suggests that Emergency programs are running at or near full capacity.

And emergency model has more outliers than the transitional model.On the other hand, the Transitional program model's boxplot displays a wider IQR, with the median lower than the Emergency model's, which indicates that the occupancy rates for Transitional programs vary more and, on average, are lower. There are several outliers below the main cluster of the data, which suggests that some Transitional programs have notably lower occupancy rates than others.

In summary, the Emergency program model has a higher occupancy rate than the Transitional program model. This could indicate that more people seek shelter in emergency situations.

```
T-Statistic: 39.07496980654136
P-Value: 0.0
```

Also, a t-test was conducted on Program Models which is a statistical method used to determine if there is a significant difference between the two groups. In this case, a T-statistic of 39.07 is extremely large, which suggests a very large difference between the two Program Models of occupancy rates.

Also, a P-value of 0.0 indicates that the null hypothesis is rejected. This means that there is statistically significant evidence to say that there is a difference between the two Program Models' of occupancy rates.

Research question 3:
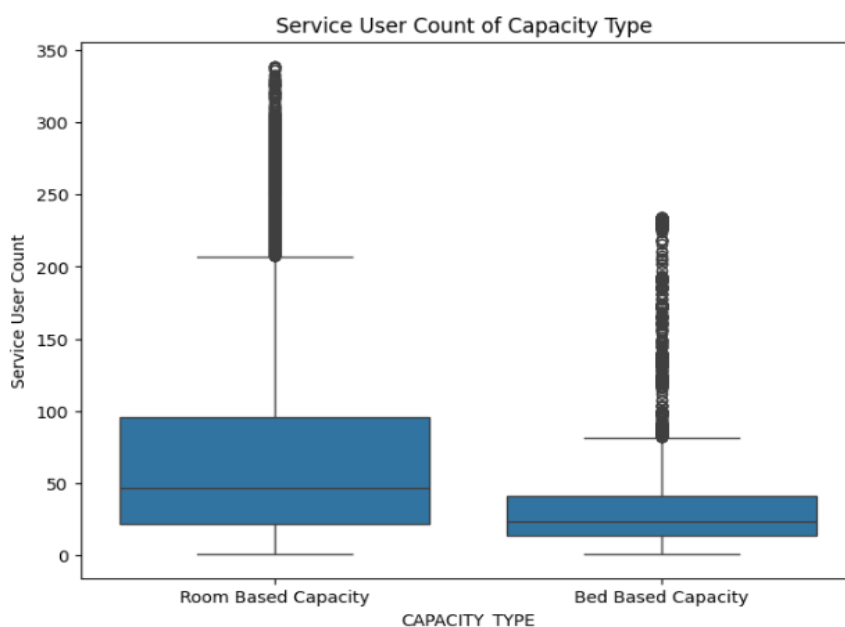
This boxplot compares the 'Service User Count' between two capacity types: 'Room Based Capacity' and 'Bed Based Capacity'. First, the 'Room Based Capacity' shows a much wider range of values, with a higher median service user count compared to the 'Bed Based Capacity'. This suggests that room-based services, on average, serve more users than bed-based ones.

Also, the 'Room Based Capacity' also has a significant number of outliers, indicating there are instances where the service user count is exceptionally high. The presence of outliers suggests that the service user count for room-based services can vary greatly. Otherwise, the 'Bed Based Capacity' has a smaller interquartile range (IQR), indicating less variability in service user count. The median is lower than the room-based services, and there are a few outliers.

This analysis indicates a trend of higher occupancy rates in room-based services compared to bed-based services. However, it is important to note that this observed trend could be influenced by factors not accounted for in the analysis, such as program accessibility, population needs, or other external variables.

```
t-statistic: 97.1228806296218
p-value: 0.0
```

Also, a t-test was conducted on capacity types which is a statistical method used to determine if there is a significant difference between the two groups. In this case, the t-statistic is extremely high (97.12), indicating a very large difference in means between the two groups.

The p-value is 0.0, which suggests that the observed difference is statistically significant. This means that we can reject the null hypothesis that there is no difference in service user count between room-based and bed-based capacities.

This insight can inform decision-making and resource allocation, helping organizations better meet their service user needs based on the chosen capacity type.