Write-up: Analysis of Occupancies in the Toronto Shelters

For this project, the first research question was what was the shelters' overall occupancy rate, how is the data about occupancy distributed, and whether there is a significant difference in the amount of each demographic that occupies the shelter? The second research question was what is the overall occupancy rate over time for each demographic that occupies the shelter? To answer these questions, this report will show the outputs and graphs from data analysis work for this project.

Figure 1.1

	OCCUPANCY_DATE	ORGANIZATION_NAME	PROGRAM_ID	PROGRAM_NAME	SECTOR	PROGRAM_MODEL	OVERNIGHT_SERVICE_TYPE	PROGRAM_AREA
0	2021-01-01	COSTI Immigrant Services	15371	COSTI North York West Hotel - Family Program	Families	Emergency	Motel/Hotel Shelter	COVID-19 Response
1	2021-01-01	COSTI Immigrant Services	16211	COSTI North York West Hotel - Seniors Program	Mixed Adult	Emergency	Motel/Hotel Shelter	COVID-19 Response
2	2021-01-01	COSTI Immigrant Services	16192	COSTI North York West Hotel Program - Men	Men	Emergency	Motel/Hotel Shelter	COVID-19 Response
3	2021-01-01	COSTI Immigrant Services	16191	COSTI North York West Hotel Program - Mixed Adult	Mixed Adult	Emergency	Motel/Hotel Shelter	COVID-19 Response
4	2021-01-01	COSTI Immigrant Services	16193	COSTI North York West Hotel Program - Women	Women	Emergency	Motel/Hotel Shelter	COVID-19 Response

Figure 1.2

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50944 entries, 0 to 50943
Data columns (total 14 columns):
 #   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
dtypes: datetime64[ns](1), float64(4), int64(2), object(7)
memory usage: 5.4+ MB
```

Figure 1.3

	PROGRAM_ID	SERVICE_USER_COUNT	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS	\
count	50944.000000	50944.000000	32399.000000	32399.000000	
mean	13986.125844	45.727171	31.627149	29.780271	
std	1705.288632	53.326049	27.127682	26.379416	
min	11791.000000	1.000000	1.000000	1.000000	
25%	12233.000000	15.000000	15.000000	14.000000	
50%	14251.000000	28.000000	25.000000	23.000000	
75%	15651.000000	51.000000	43.000000	41.000000	
max	16631.000000	339.000000	234.000000	234.000000	

The first step was to perform basic data cleaning and data processing to observe what the dataset looks like and gather general information. In the ***Figure 1.1-1.2***, the dataset shows the daily occupancy and capacity of Toronto shelters in 2021. It includes the 50,944 rows and 14 columns. It includes one date variable (OCCUPANCY_DATE), three ID variables (ORGANIZATION_NAME, PROGRAM_ID, PROGRAM_NAME), five numerical variables and five categorical variables. ***Figure 1.3*** shows the descriptive statistic description for each column in a data table.

Figure 1.4

```
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
```

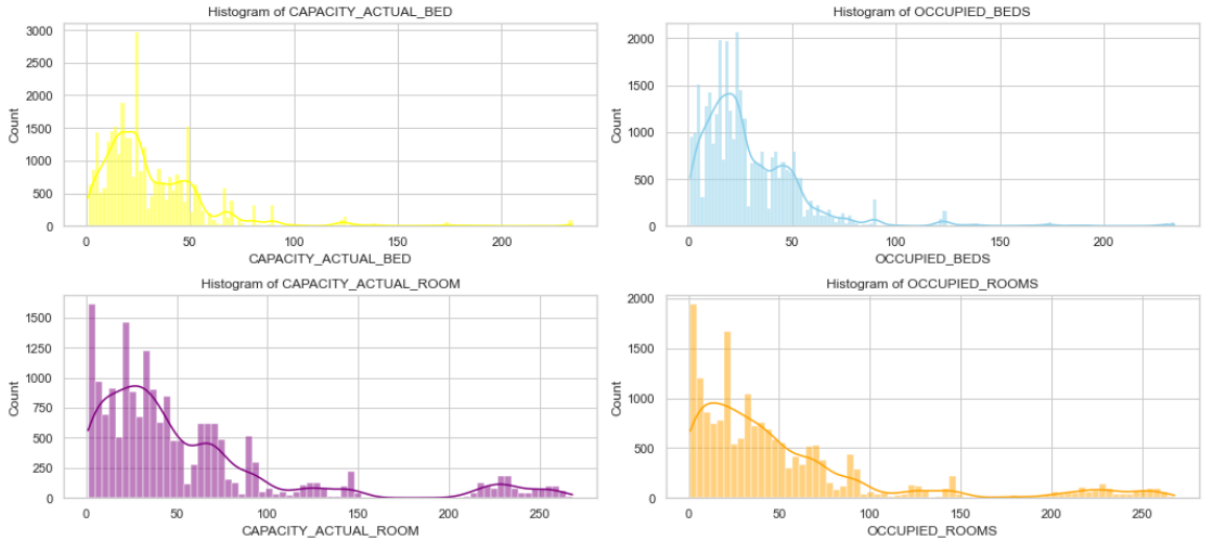
In the ***Figure 1.4***, the missing values are addressed by identifying any potential outliers which will affect the further analysis.

Figure 1.5

	OCCUPIED_RATE_FOR_BEDS	OCCUPIED_RATE_FOR_ROOMS	OCCUPIED_RATE
0	0.0	0.896552	0.896552
1	0.0	1.000000	1.000000
2	0.0	0.821429	0.821429
3	0.0	1.000000	1.000000
4	0.0	0.928571	0.928571

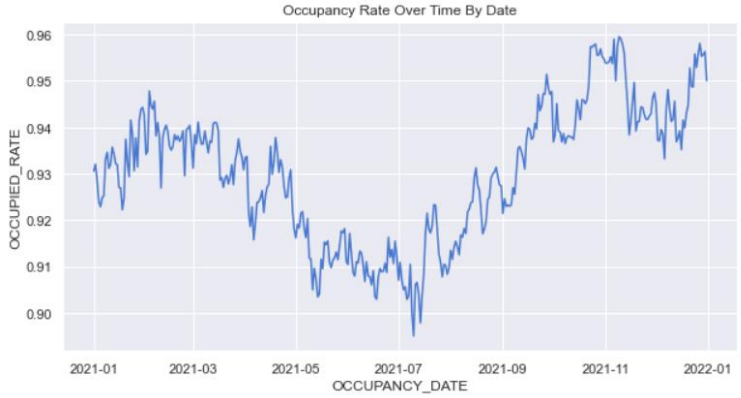
In this data table, it shows the calculate occupancy rate for the shelters. The occupancy rate is set up from 0 to 1, where any number between 0 and 1 represents the number of homeless people living in the shelter.

Figure 1.6



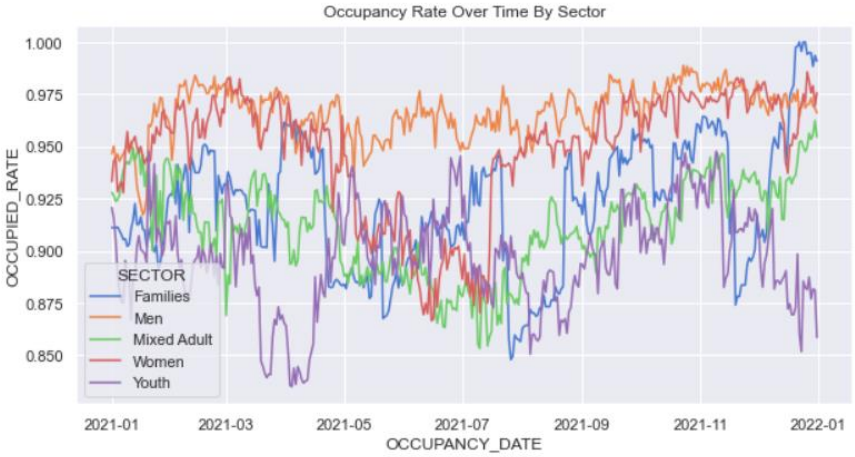
In the graph, it uses data visualization technique through four numerical variables (CAPACITY_ACTUAL_BED, OCCUPIED_BEDS, CAPACITY_ACTUAL_ROOM, OCCUPIED_ROOMS). The overall distribution is right skewed with a longer tail. However, this does not mean that there are many outliers. This can be explained that a shelter with more space will likely have more rooms and beds. Therefore, it can be inferred that most shelters in Toronto are small and medium-sized, larger shelters are limited.

Figure 1.7



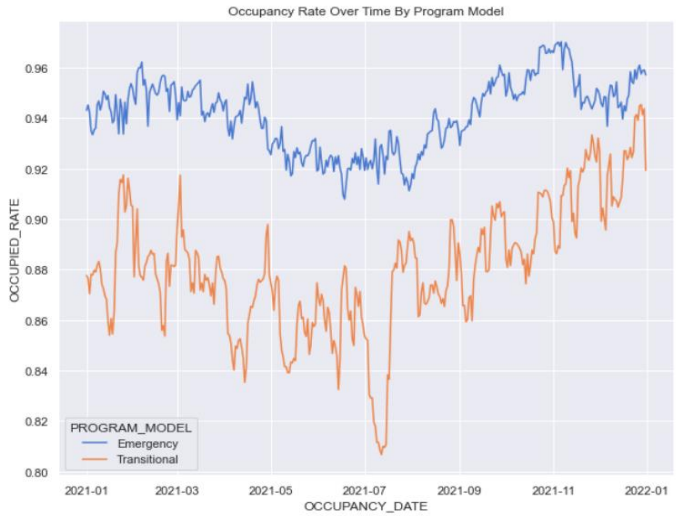
This graph illustrates the overall trend of the occupancy rate for the whole year of 2021. The exploratory variable is "OCCUPANCY_DATE" and the response variable is "OCCUPIED_RATE". The occupancy rate may differ and fluctuate due to seasonal factors. January 2021 and January 2022 had a relatively higher occupancy rate. In contrast, the period between May 2021 and July 2021 has a significant drop at that point. This can be explained by the tendency of most people to seek residency in shelters during the winter season when it is colder, compared to the warmer season where it may not be as necessary.

Figure 1.8



This graph shows how the occupancy rate changes over time for each demographic. The categories of "Men" and "Women" exhibit similar trends over time. However, the "Youth" category consistently has a low occupancy rate, which may show how youth have parental guidance and support from government agencies, making them less reliant on shelters. Thus, the majority of the shelter population are adults and family members.

Figure 1.9

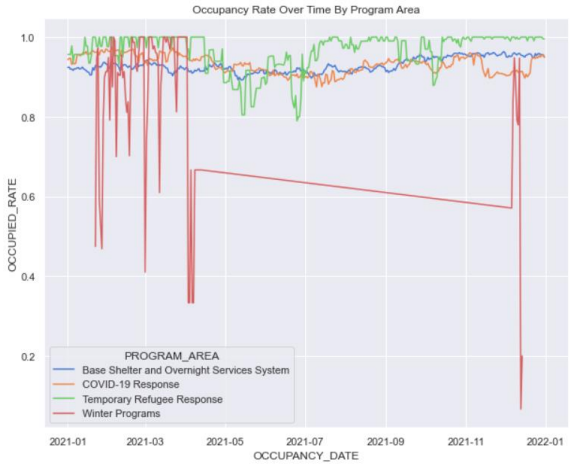


The occupancy rate over time can be separated into two different types of shelter programs (Emergency and Transitional). Based on the graph, the Emergency program consistently maintains a higher and smoother occupancy rate. In contrast, the Transitional program shows more fluctuations over time, with a significant drop around July 2021. The occupancy rate for different programs suggests that people with Emergency programs are more likely to receive accommodations in shelters.

Figure 1.10

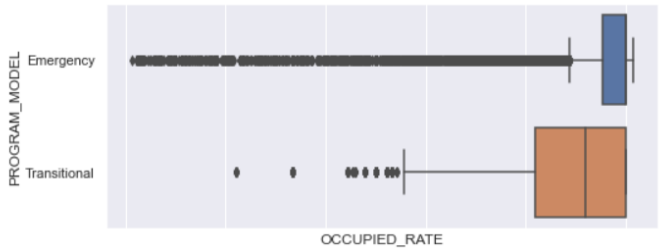
This graph presents the changes in occupancy rates by various Overnight Service Types. According to the line plot, the Isolation/Recovery Site shows a higher occupancy rate from January 2021 to May 2021, likely due to the spread of COVID-19. Additionally, overnight services such as 24-Hour Respite Site, Intern Housing, Shelter, and Motel/Hotel Shelter have nearly the same occupancy rates, indicating that people prefer these four types of shelters. This suggests differences in homeless people's preferences and the accessibility of services.

Figure 2.1



This graph displays the occupancy rate with regard to different Program Areas. The Winter Programs exhibited significant fluctuations during 2021. This is because the Winter Programs seems to get popular during a specific season (winter), while the occupancy rate for this area is low during other seasons. Thus, this report recommends that that government consider discontinuing the Winter Program in non-winter seasons to save costs and allocate more resources during the winter period to provide additional room or bet capacity for homeless people.

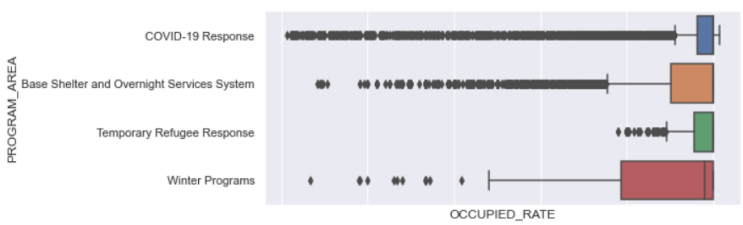
Figure 2.2 (Boxplot with Program Model)




```
t-value = 39.04455484711919
p-value = 0.0
```

In this part of the project, a t-test was conducted on two different types of Program Models (Emergency and Transitional). The t-value is large while the p-value is extremely small. It represents that there is statistical significance between the two types of occupancy rates. From observing shelter occupancy rates in Emergency and Transitional situations, it is found that Emergency hotel occupancy rates are significantly higher than Transitional ones. This shows that when homeless people face extreme weather or financial crisis, they are more likely to stay in shelters, leading to higher occupancy rate in Emergency shelters where demand exceeds the supply.

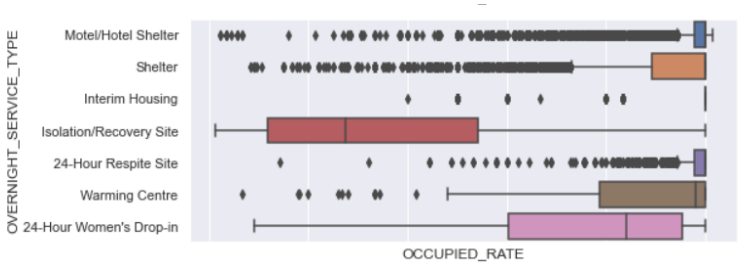
Figure 2.3 (Boxplot with Program Area)



```
t-value = 5.6741534258206725
p-value = 1.4179919102362063e-08
```

Since the research question explores how the occupancy rate changes by different categories, the graph above provides an overall distribution of four different shelter area types. ***Figure 2.3*** shows that the boxplot of Winter Programs has some outliers below the 1.5 IQR. However, although these points are treated as outliers, they may be affected by other factors such as seasonal changes. For example, homeless people tend to live in shelters during the cold winter and move out when the warmer weather arrives, which can explain why there was a lower occupancy rate with Winter Programs. The t-test conducted compares the groups of COVID-19 Response and Winter Programs. Based on the results, it can be concluded that the occupancy rate for the "COVID-19 Response" area significantly differs from the "Winter Programs" area. This difference is likely due to the response to the spread of COVID-19, which occurred in 2021.

Figure 2.4 (Boxplot with Overnight Service Type)




```
t-value = 16.796204818698236
p-value = 1.0244830363910635e-62
```

This figure presents the overall distribution of occupancy rates for different overnight service types, with a focus on analyzing and comparing the occupancy rates of shelters with different types of overnight services, such as Motel/Hotel Shelter and Warming Centre. The analysis also used a t-test to determine if there were statistically significant differences. From the boxplot, the distribution of occupancy rates between these two services is similar enough to conclude that they are more popular services for homeless people. The t-value is approximately equal to 17 and the p-value is extremely smaller than the significant level at 5% (0.05), which indicates there is a significant difference between these two types of services in terms of occupancy rate. Based on the results of the t-test, these two types of services (Motel/Hotel Shelter and Warming Centre) were more crowded, and Motel/Hotel Shelter had significantly higher occupancy rates than traditional Shelter, possibly due to providing services more tailored to the needs of some homeless people.