Assignment 1 Report

Student Name and Number : Jiajun Yang 1010128862

Research Question

This report presents a detailed exploratory data analysis (EDA) and statistical testing on a dataset of service facilities, aiming to evaluate the impact of different capacity types (bed-based and room-based) and program models (emergency and transitional) on the facility occupancy rates.

Dataset

INF2178_A1_data.xlsx

Data Loading and Preprocessing

The dataset was loaded from the INF2178_A1_data.xlsx file. To analyze the data contained within this file, I first examined the contents of the file and discovered that the attributes include OCCUPANCY_DATE, ORGANIZATION_NAME, PROGRAM_ID, PROGRAM_NAME, SECTOR, PROGRAM_MODEL, OVERNIGHT_SERVICE_TYPE, PROGRAM_AREA, SERVICE_USER_COUNT, CAPACITY_TYPE, CAPACITY_ACTUAL_BED, OCCUPIED_BEDS, CAPACITY_ACTUAL_ROOM, and OCCUPIED_ROOMS.

	OCCUPANCY_DATE	ORGANIZATION_NAME	PROGRAM_ID	PROGRAM_NAME	SECTOR	PROGRAM_MODEL	OVERNIGHT_SERVICE_TYPE	PROGRAM_AREA	SERVICE_USER_COUNT	CAPACITY_TYPE	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS	CAPACITY_ACTUAL_ROOM	OCCUPIED_ROOMS
0	2021-01-01	COSTI Immigrant Services	15371	COSTI North York West Hotel - Family Program	Families	Emergency	Motel/Hotel Shelter	COVID-19 Response	74	Room Based Capacity	NaN	NaN	29.0	26.0
1	2021-01-01	COSTI Immigrant Services	16211	COSTI North York West Hotel - Seniors Program	Mixed Adult	Emergency	Motel/Hotel Shelter	COVID-19 Response	3	Room Based Capacity	NaN	NaN	3.0	3.0
2	2021-01-01	COSTI Immigrant Services	16192	COSTI North York West Hotel Program - Men	Men	Emergency	Motel/Hotel Shelter	COVID-19 Response	24	Room Based Capacity	NaN	NaN	28.0	23.0
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	25	Room Based Capacity	NaN	NaN	17.0	17.0
4	2021-01-01	COSTI Immigrant Services	16193	COSTI North York West Hotel Program - Women	Women	Emergency	Motel/Hotel Shelter	COVID-19 Response	13	Room Based Capacity	NaN	NaN	14.0	13.0
5	2021-01-01	COSTI Immigrant Services	12251	COSTI Reception Centre CITY Program	Mixed Adult	Emergency	Shelter	Base Shelter and Overnight Services System	6	Bed Based Capacity	8.0	6.0	NaN	NaN
6	2021-01-01	COSTI Immigrant Services	15372	COSTI Uptown Hotel COVID-19 - Family Program	Families	Emergency	Motel/Hotel Shelter	COVID-19 Response	120	Room Based Capacity	NaN	NaN	46.0	44.0
7	2021-01-01	COSTI Immigrant Services	13751	COSTI Uptown Hotel Family Program	Families	Emergency	Motel/Hotel Shelter	Temporary Refugee Response	130	Room Based Capacity	NaN	NaN	46.0	44.0
8	2021-01-01	Christie Ossington Neighbourhood Centre	16111	CONC Etobicoke Hotel Program - Mixed Adult	Mixed Adult	Emergency	Motel/Hotel Shelter	COVID-19 Response	153	Room Based Capacity	NaN	NaN	145.0	145.0
9	2021-01-01	Christie Ossington Neighbourhood Centre	15711	CONC Men's Hotel Program	Men	Emergency	Motel/Hotel Shelter	COVID-19 Response	40	Room Based Capacity	NaN	NaN	41.0	40.0

To conduct the analysis, the following columns were selected for focused analysis: CAPACITY_TYPE, PROGRAM_MODEL, SERVICE_USER_COUNT, CAPACITY_ACTUAL_BED, OCCUPIED_BEDS, CAPACITY_ACTUAL_ROOM, and OCCUPIED_ROOMS.

	CAPACITY_TYPE	PROGRAM_MODEL	SERVICE_USER_COUNT	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS	CAPACITY_ACTUAL_ROOM	OCCUPIED_ROOMS
0	Room Based Capacity	Emergency	74	NaN	NaN	29.0	26.0
1	Room Based Capacity	Emergency	3	NaN	NaN	3.0	3.0
2	Room Based Capacity	Emergency	24	NaN	NaN	28.0	23.0
3	Room Based Capacity	Emergency	25	NaN	NaN	17.0	17.0
4	Room Based Capacity	Emergency	13	NaN	NaN	14.0	13.0
...	...	...	...	...	...	...	...
50939	Bed Based Capacity	Emergency	6	20.0	6.0	NaN	NaN
50940	Bed Based Capacity	Emergency	23	23.0	23.0	NaN	NaN
50941	Bed Based Capacity	Transitional	13	14.0	13.0	NaN	NaN
50942	Bed Based Capacity	Emergency	10	10.0	10.0	NaN	NaN
50943	Bed Based Capacity	Transitional	29	29.0	29.0	NaN	NaN

50944 rows × 7 columns

We observed that some columns contain NaN values, which is attributed to the nature of homeless shelter services and whether the capacity for this program is measured in rooms or beds. When the provided service is based on rooms, the columns relevant to bed-based capacity will have NaN values. These two types of capacity cannot coexist.

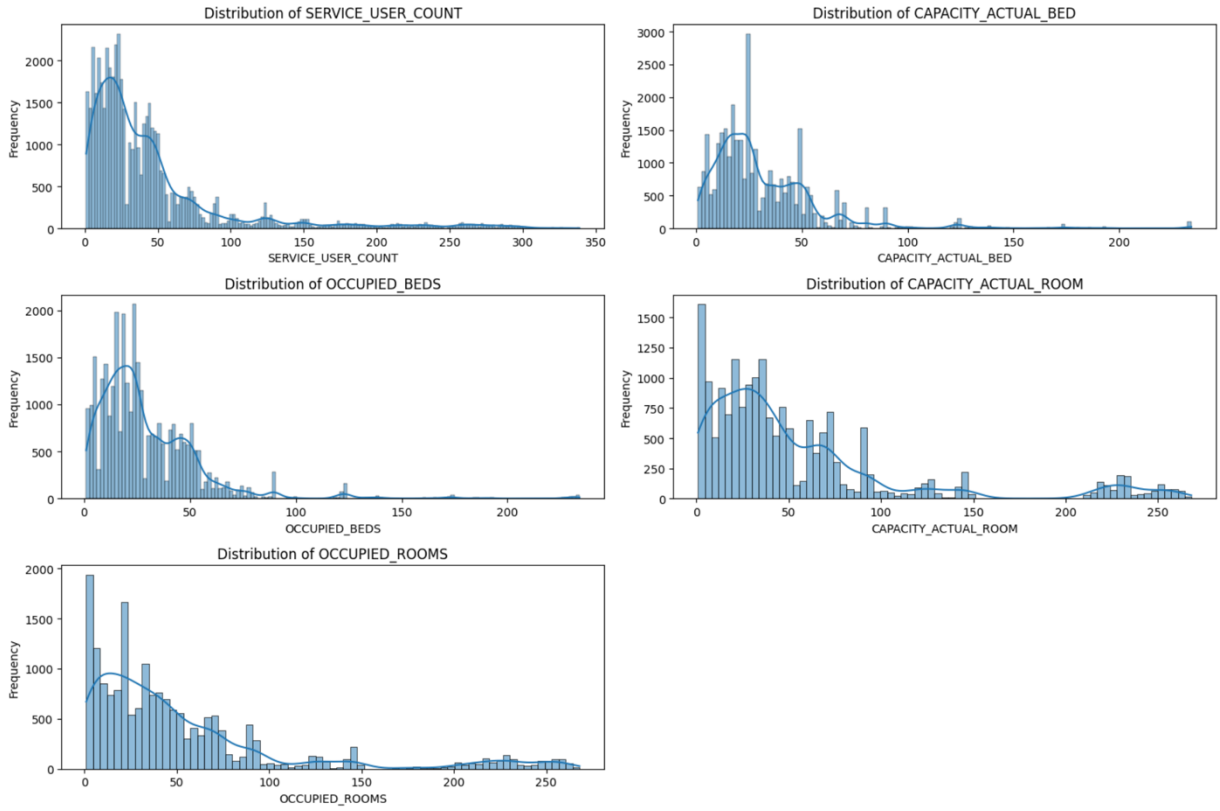
Subsequently, I performed basic descriptive statistics for numerical columns, and the results are presented in the following table.

	SERVICE_USER_COUNT	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS	CAPACITY_ACTUAL_ROOM	OCCUPIED_ROOMS
count	50944.000000	32399.000000	32399.000000	18545.000000	18545.000000
mean	45.727171	31.627149	29.780271	55.549259	52.798598
std	53.326049	27.127682	26.379416	59.448805	58.792954
min	1.000000	1.000000	1.000000	1.000000	1.000000
25%	15.000000	15.000000	14.000000	19.000000	16.000000
50%	28.000000	25.000000	23.000000	35.000000	34.000000
75%	51.000000	43.000000	41.000000	68.000000	66.000000
max	339.000000	234.000000	234.000000	268.000000	268.000000

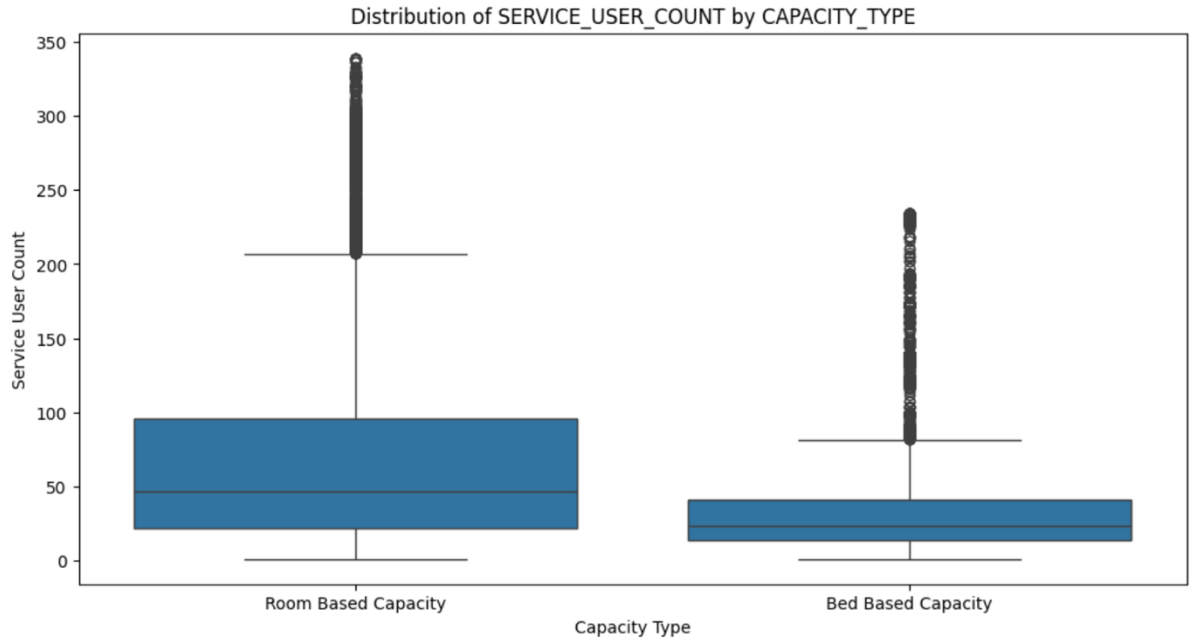
From the descriptive statistics, we can deduce that on average, there are approximately 45 service users, with actual bed capacity and occupied beds at about 31 and 29 respectively, while the actual room capacity and occupied rooms stand at approximately 55 and 52 respectively. The large standard deviations of these metrics reflect significant differences between services in terms of user numbers and resource utilization. The range of minimum and maximum values indicates a wide variation from very low to very high, revealing the challenges different services may face in meeting user needs and improving resource efficiency.

Exploratory data analysis(EDA)

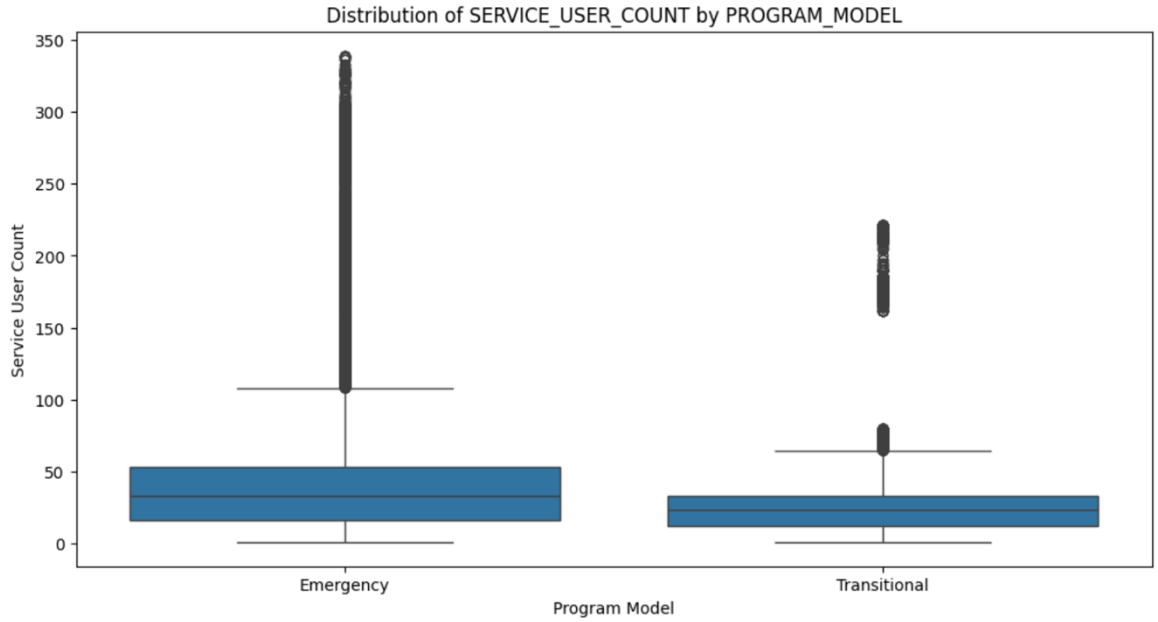
Histograms and box plots were used to explore the distribution of service user count, actual bed and room capacities, and their occupancy. The analysis showed that most indicators are right-skewed, indicating that occupancy is generally low, with some instances of high occupancy.



I plotted the distribution of SERVICE_USER_COUNT for each category in CAPACITY_TYPE. From the chart, we can observe that the median line for room-based capacity is higher than that for bed-based capacity, which suggests that on average, services of the room-based capacity type attract more users. Furthermore, room-based capacity features a larger box, indicating a more significant variance in the number of users utilizing this service.



Additionally, I plotted the distribution of SERVICE_USER_COUNT for each category in PROGRAM_MODEL. We can conclude that users tend to prefer the Emergency program model.

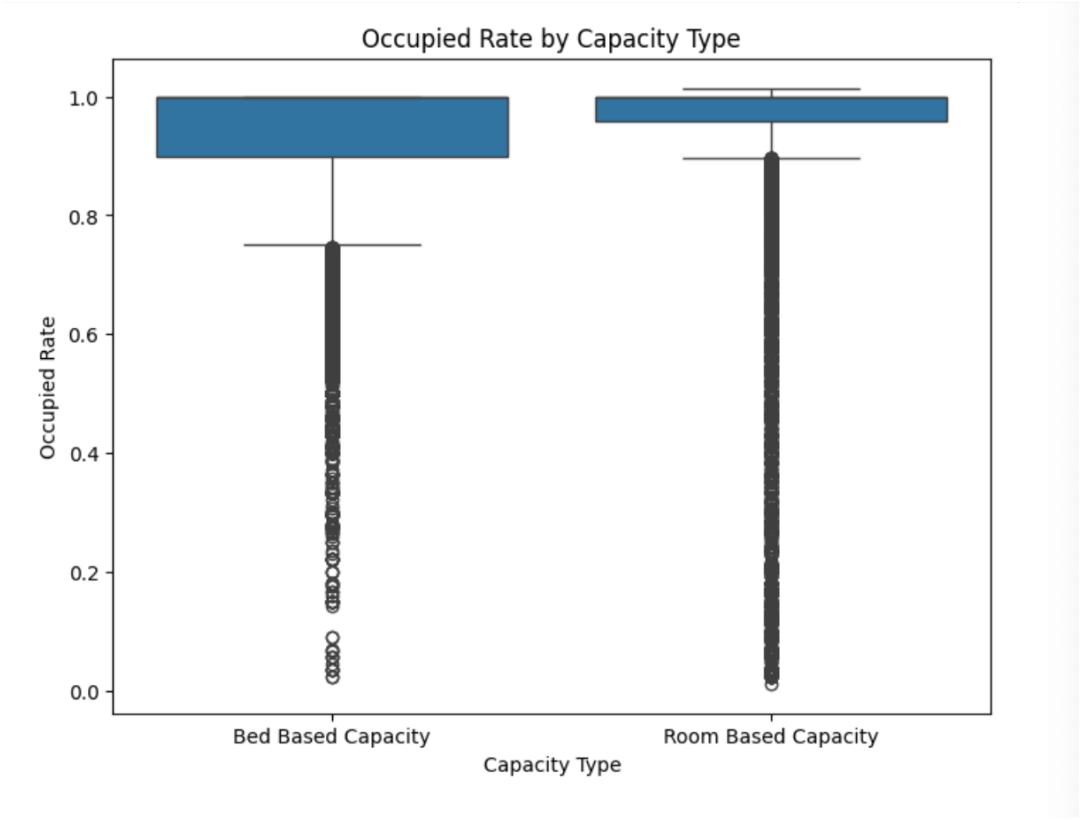


Overall, people tend to prefer room-based capacity and the emergency program model. This preference may be attributed to the fact that most shelters accommodate families rather than individuals, and the emergency model, which can be used unconditionally, is more popular than the transitional model, which requires certain conditions to be met for use.

T-test

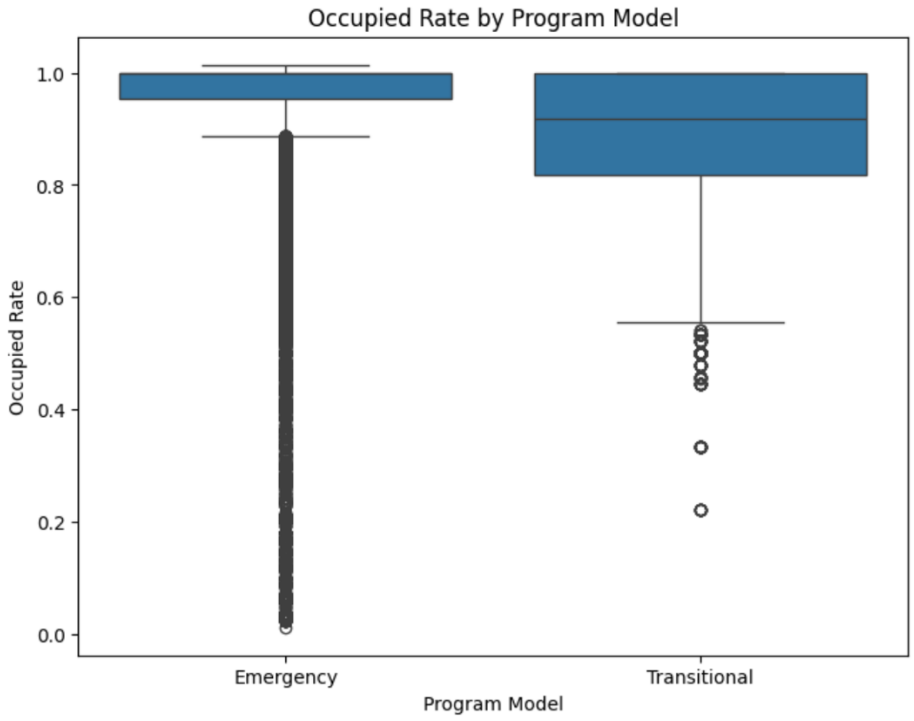
Capacity Type:

To conduct a t-test, I calculated the occupancy rates for bed-based and room-based capacity types and divided the data into two datasets: one containing only bed-based data (bed_based_data) and another containing only room-based data (room_based_data). I then performed a t-test and Welch's t-test on the occupancy rates of bed-based versus room-based capacity types to assess whether there was a statistically significant difference between these two types of occupancy rates. The results of the t-test showed a p-value of 1.2128933183471424e-06, and the results of Welch's t-test showed a p-value of 6.860477551487939e-06. Both test results indicate a significant difference in occupancy rates with p-values far less than 0.05, meaning we reject the null hypothesis that the occupancy rates of the two capacity types are the same. This suggests that the occupancy rates for bed-based and room-based capacities are significantly different statistically. The box plots allow us to see that the occupancy ratio of users utilizing room-based homeless shelter services is higher than that of bed-based services.



Program Model:

I separated the data based on the Program model: 'Emergency' and 'Transitional', to explore whether the differences in program models have a significant impact on occupancy rates. Through a t-test, the obtained p-value was 0.0, which is far below 0.05. Therefore, the differences in program models have a statistically significant impact on occupancy rates. The box plots enable us to see that the occupancy ratio of users utilizing Emergency type homeless shelter services is higher than that of Transitional services.



Conclusion:
The analyses in this report have highlighted the effects of different capacity types and program models on occupancy rates within homeless shelter services. Notably, the distinction between bed-based and room-based capacity types significantly influences occupancy rates. These findings underscore the necessity of optimizing resource allocation and enhancing service efficiency.

Discussion:
Based on the results it was seen that room-based capacity tends to have higher occupancy rates and may provide more security, features that are particularly attractive to families or individuals seeking shelter. This preference underscores the importance of designing shelter services that are tailored to the needs and preferences of the people they serve, which may lead to more effective and efficient use of resources. The comparison of occupancy rates between the emergency and transitional program models also reveals user preferences and needs. Higher occupancy rates in emergency shelters indicate a pressing need for immediate, no-strings-attached accommodations, reflecting the urgency of homelessness for many individuals and families. Therefore, I believe the government could plan to favor shelters that offer room types and emergency models to efficient use of resources.