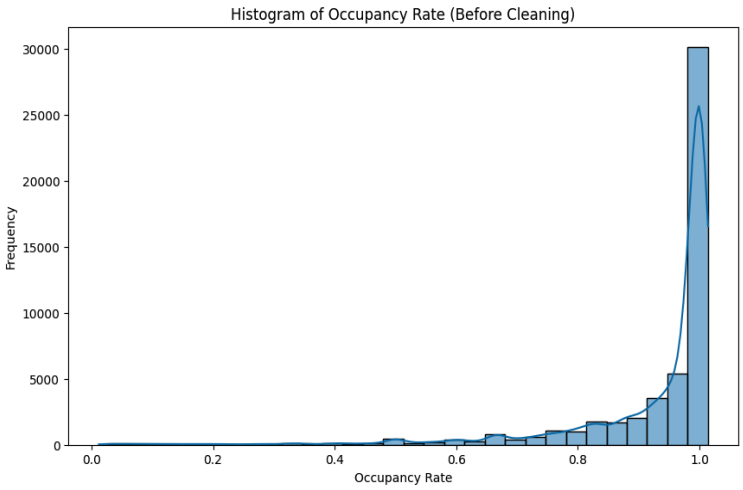
Insights into Shelter Occupancy Trends in Toronto

Introduction

In the heart of Toronto, a city known for its vibrant culture and diversity, there exists a parallel narrative that speaks to the challenges of homelessness and the critical role of shelter programs. The dataset at our disposal (Shelter, Support & Housing Administration, 2024), which tracks the daily occupancy and capacity of Toronto shelters throughout 2021, serves as a window into this world. Through careful exploration and analysis, we've raised the ***research question***: How do the capacity rates of Emergency and Transitional shelter programs compare to those of Shelter and Motel/Hotel service types, and what implications do these rates have for policy development aimed at optimizing shelter capacity and improving the housing stability of individuals experiencing homelessness in Toronto?

Exploratory Data Analysis (EDA)

Our first step was to dive into the dataset, examining variables such as `CAPACITY_TYPE`, `PROGRAM_MODEL`, `SERVICE_USER_COUNT`, and `OCCUPANCY_RATE`, among others. The pandas library (McKinney, 2012) was used to import the data, and the functions .shape, .head, .tail, and .description were then used to analyze it. This made it possible to get a general idea of the dataset's size and the kinds of data that were included in each column (continuous or categorical). Additionally, it made it possible to visually identify probable NaN or null values in the characteristics that required adjustment. Then, we cleaned in accordance with the Tidy data principles (Wang et al., 2020), where each concept was briefly discussed as to whether or not it was met prior to the cleaning process. Since no melting was necessary, no columns needed to be divided, and no observational units needed to be separated, all principles except 5 were met. Certain types that include NaN, however, might not make sense and should be excluded or filled with 0s.Through histograms, box plots, and summary statistics, we uncovered several key insights:

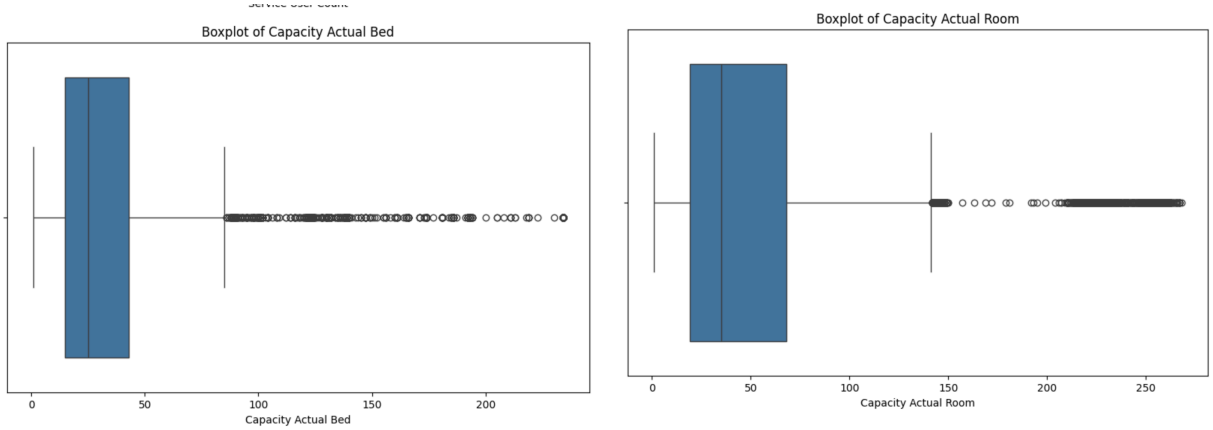


- ***High Occupancy Rates***: The shelters are operating near or at full capacity, with many programs consistently reaching 100% occupancy. This underscores the high demand for shelter services in the city and

points to a potential shortfall in available accommodations for those in need.

- ***Diverse Shelter Capacities***: The capacity of shelters varies widely, from small programs with just a handful of beds to large facilities that can accommodate hundreds of individuals each night. This diversity reflects the multifaceted approach Toronto has taken to address homelessness, tailoring shelter solutions to different population

```
Unique values in 'CAPACITY_TYPE':
['Room Based Capacity' 'Bed Based Capacity']
Counts of unique values in 'CAPACITY_TYPE':
CAPACITY_TYPE
Bed Based Capacity     32399
Room Based Capacity    18545
Name: count, dtype: int64
```



segments.Each characteristic underwent EDA, with the Seaborn library producing boxplots and histograms (Waskom, 2017). In order to facilitate distribution comparison and easy visualization, each attribute's data was divided by capacity type and shown on the same set of axes. This is especially important for the next stage, which is to use the self-written new_column function to calculate the numerical data into an overall percentage so that each attribute has its own quantitative column.

- ***Program Model Impact***: A comparison of Emergency and Transitional programs revealed significant differences in occupancy rates. Emergency shelters, designed for immediate, short-term needs, tend to have higher occupancy rates compared to Transitional programs, which offer longer-term support with a focus on reintegration into permanent housing.

```
Unique values in 'PROGRAM_MODEL':
['Emergency' 'Transitional' nan]
Counts of unique values in 'PROGRAM_MODEL':
PROGRAM_MODEL
Emergency       41541
Transitional     9401
Name: count, dtype: int64
```

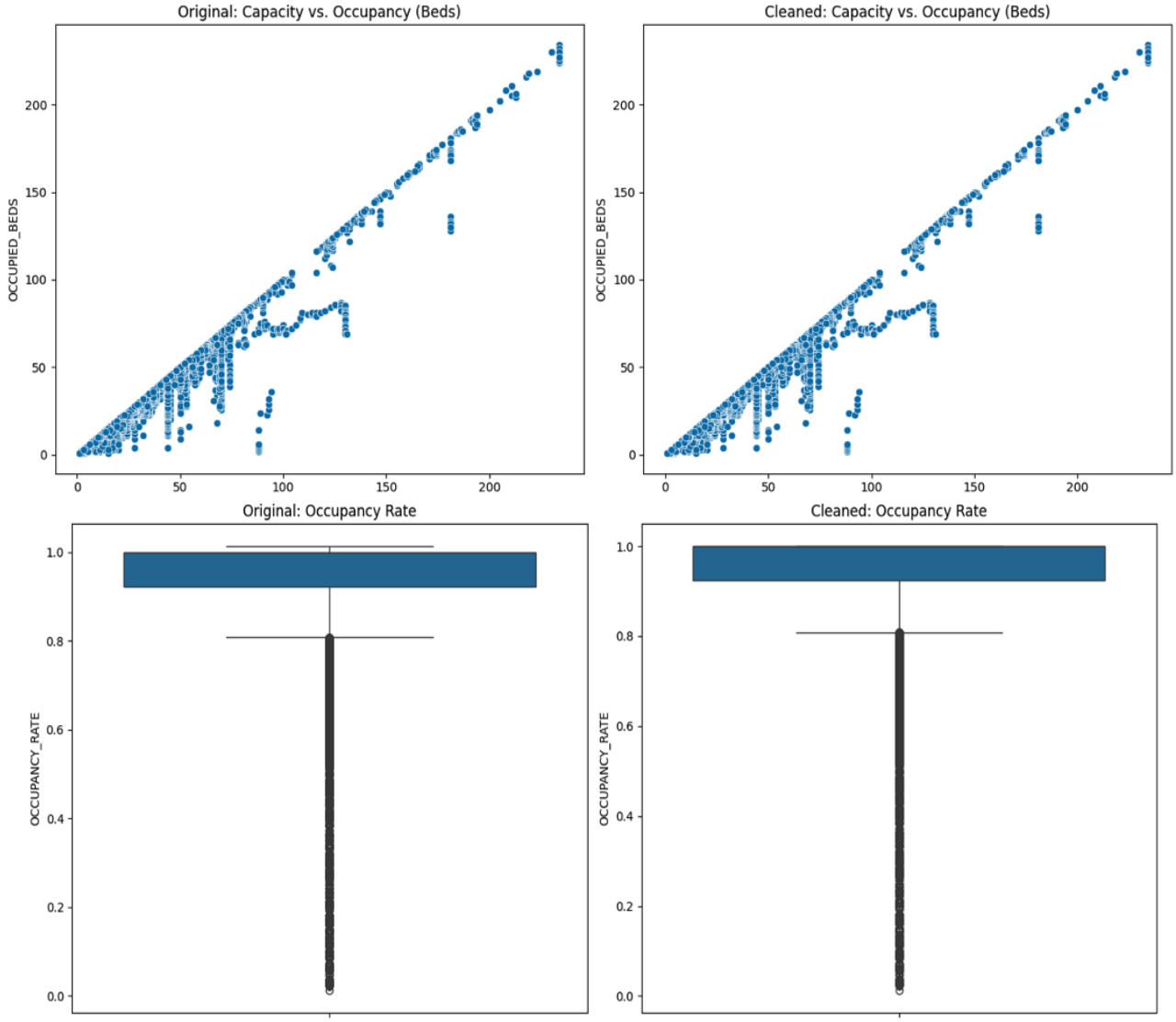
Data Cleaning and T-Test Analysis

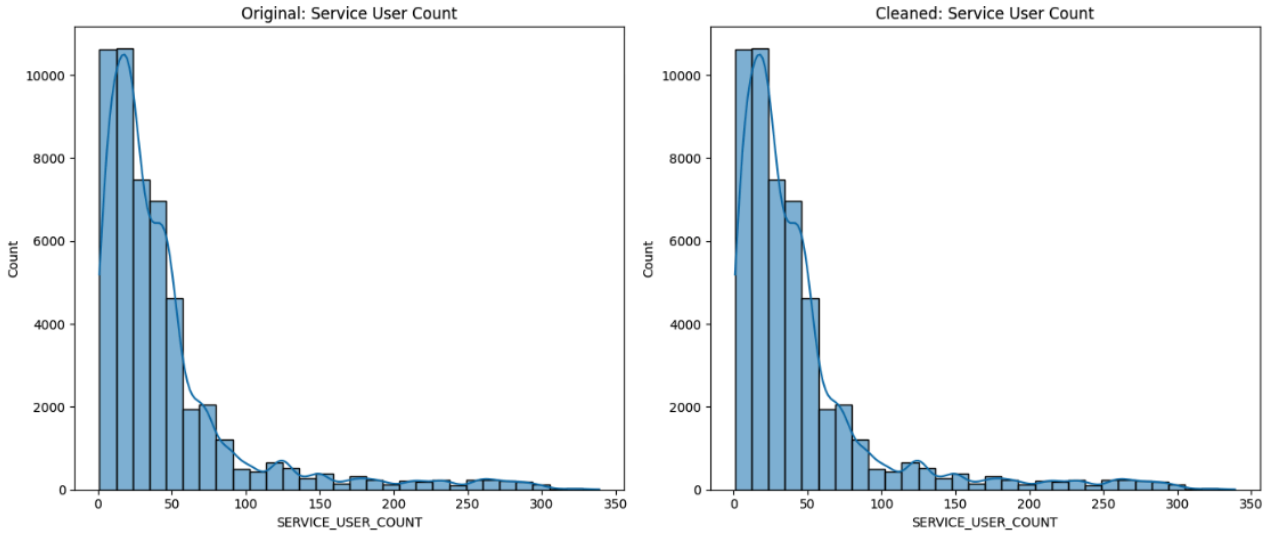
Data Cleaning

Recognizing the importance of data integrity, we meticulously cleaned the dataset, addressing missing values and logical inconsistencies. This process refined our analysis, ensuring reliability in the findings.

As the null value shown on the right illustrates the missing values in 8 columns, for our specific interests, we would like to clean the nulls for CAPACITY_ACTUAL_BED, OCCUPIED_BEDS, CAPACITY_ACTUAL_ROOM, OCCUPIED_ROOMS.

```
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
```





The diagrams above show the comparison between original dataset and cleaned dataset. The count for SERVICE_USER_COUNT and OCCUPANCY_RATE has marginally decreased from 50,944 to 50,940 in the cleaned dataset. This slight reduction indicates that only a few records were removed or adjusted during the cleaning process, likely those identified with issues such as missing or illogical values.

The mean values for SERVICE_USER_COUNT, OCCUPANCY_RATE, CAPACITY_ACTUAL_BED, and OCCUPIED_BEDS have remained nearly identical before and after cleaning. This suggests that the cleaning process did not significantly alter the central tendency of these key metrics, maintaining the dataset's overall integrity.

The minimum values for all variables remain unchanged, suggesting that the cleaning process did not affect the lower bounds of the data distribution.

The maximum value for OCCUPANCY_RATE has changed from 1.014085 to 1.000000, indicating that the cleaning process effectively addressed instances where occupancy rates exceeded logical maximums (e.g., rates over 100%).

Though the difference is hard to trace on the diagram, specific statistics can reveal more in detail in the following table.

Original Data Summary:

	SERVICE_USER_COUNT	OCCUPANCY_RATE	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS
count	50944.000000	50944.000000	32399.000000	32399.000000
mean	45.727171	0.930142	31.627149	29.780271
std	53.326049	0.138788	27.127682	26.379416
min	1.000000	0.012048	1.000000	1.000000
25%	15.000000	0.923077	15.000000	14.000000
50%	28.000000	1.000000	25.000000	23.000000
75%	51.000000	1.000000	43.000000	41.000000
max	339.000000	1.014085	234.000000	234.000000

Cleaned Data Summary:

	SERVICE_USER_COUNT	OCCUPANCY_RATE	CAPACITY_ACTUAL_BED	OCCUPIED_BEDS
count	50940.000000	50940.000000	32399.000000	32399.000000
mean	45.721339	0.930136	31.627149	29.780271
std	53.324007	0.138791	27.127682	26.379416
min	1.000000	0.012048	1.000000	1.000000
25%	15.000000	0.923077	15.000000	14.000000
50%	28.000000	1.000000	25.000000	23.000000
75%	51.000000	1.000000	43.000000	41.000000
max	339.000000	1.000000	234.000000	234.000000

T-Test Discussion

Emergency vs. Transitional Programs

```
T-statistic: 39.06906374293522, P-value: 0.0
```

The t-test comparing Emergency and Transitional programs demonstrated a statistically significant difference in occupancy rates, with Emergency shelters showing higher utilization. This finding underscores the critical role of Emergency shelters in providing immediate relief but also highlights potential capacity challenges in meeting short-term housing needs effectively.

Statistical Insight: The high occupancy rates in Emergency shelters may signal an urgent need for increased capacity or alternative solutions to alleviate pressure on these essential services.

Shelter vs. Motel/Hotel Services

```
T-statistic: -37.92963658991663, P-value: 5.5121886549851e-310
```

A second t-test comparing traditional shelters with motel/hotel accommodations revealed a higher mean occupancy rate for motel/hotel services. This significant difference suggests a preference or policy inclination towards using temporary accommodations, possibly due to perceived benefits such as privacy, safety, and comfort.

Statistical Insight: The reliance on motel/hotel accommodations, despite potentially higher costs, indicates a strategic use of these resources to meet specific needs within the homeless population. The substantial difference in occupancy rates (-37.93 T-statistic, P-value effectively 0) highlights the importance of understanding and optimizing the use of motel/hotel services within the overall shelter strategy.

Insights

The narrative that emerges from our analysis is one of a city grappling with the complexities of homelessness. The ***high occupancy rates*** across shelter programs highlight the acute need for services and the constant pressure on the shelter system to accommodate an ever-present demand. The variation in ***shelter capacities, service type*** and ***program models*** reflects a strategic, albeit stretched, response to a spectrum of needs within the homeless population.

The difference in occupancy rates between ***Emergency vs. Transitional*** and ***Shelter vs. Motel/Hotel*** programs further illuminates the challenges in transitioning individuals from immediate shelter services to more stable, long-term housing solutions. It underscores the importance of not only providing shelter but also supporting individuals through programs aimed at addressing the root causes of homelessness.

Future Directions

Our preliminary analysis opens several avenues for further research. To fully understand the efficacy of Toronto's shelter system, we need to delve into the factors influencing the transition rates from Emergency to Transitional programs and, ultimately, to permanent housing. Additionally, analyzing the impact of external factors, such as economic conditions and housing market trends on shelter demand and occupancy rates could provide valuable insights into systemic solutions to homelessness.
Despite that the T test performed well in identifying the difference for various types of shelters. More specific analysis could be performed like ANOVA or ANCOVA that could potentially help enhance our confidence in drawing the current conclusion that some hotel/motel shelters are more popular than the traditional shelter.

Conclusion

Through the lens of data, we've glimpsed the challenges and efforts to address homelessness in Toronto. The story told by the 2021 shelter data is one of urgency, complexity, and resilience. It's a call to action for policymakers, community leaders, and citizens to engage in informed dialogue and collaborative efforts to ensure that the city's shelter system not only meets immediate needs but also paves the way for sustainable, long-term solutions to homelessness.

References

McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.".

Shelter, Support & Housing Administration. (2024, February 4). Open data dataset. City of Toronto Open Data Portal. https://open.toronto.ca/dataset/daily-shelter-overnight-service-occupancy-capacity/

Waskom, M., Botvinnik, Kane, D., Hobson, P., Lukauskas, S., Gemperline, D. C., Qalieh, A. (2017). mwaskom/seaborn: v0.8.1 (September 2017). Zenodo. https://doi.org/10.5281/zenodo.883859

Wang, E., Cook, D., & Hyndman, R. J. (2020). A new tidy data structure to support exploration and modeling of temporal data. Journal of Computational and graphical Statistics, 29(3), 466-478.