Quantitative Analysis

Introduction

In recent years, the homeless population in Toronto has significantly increased. For society's wellness, it is important to ensure that the homeless are guaranteed a basic livelihood. Since the resources are limited, the rational allocation of resources is very important. Understanding the relationships with the occupancy rate would help us to manage the limited resources. Specifically, in this quantitative analysis, I focus on the room occupancy rate between the two different sectors including men and women. The goal of the quantitative analysis is to determine whether there is a statistically significant difference in occupancy rates between men and women. Getting this information would help us to develop an improved program and policy.

Research Question: Is there a significant difference in the occupancy room rate between men and women?

Hypothesis:

$H_0: \mu_{men} = \mu_{women}$

Null Hypothesis: There is no significant difference in the mean room occupancy rates between 'Men' and 'Women'.

$H_a: \mu_{men} \neq \mu_{women}$

Alternative Hypothesis: There is a significant difference in the mean room occupancy rates between 'Men' and 'Women'.

Methods

I first load the necessary library panda for the analysis. Then, I converted the Excel file into a CSV file for analysis and loaded it into the Python file. I calculate the continuous variable room occupancy rate (OCCUPIED_ROOMS/CAPCACITY_ACTUAL_ROOM) and add this column to the dataset. Before conducting the t-test, I removed the missing values in the 'OCCUPANCY_RATE' and 'SECTOR' and then selected the occupancy rate where the sector is men and selected the occupancy rate where the sector is women. In the end, I performed the independent two-sample t-test since the groups of men and women were unrelated. In this t-test, the continuous variable room occupancy rate is the dependent variable, and the categorical variable sector (men and women) is the independent variable.

Data Collection and Preparation

Room occupancy rate is first calculated through the equation OCCUPIED_ROOMS/CAPCACITY_ACTUAL_ROOM. I collected the room occupancy rate where they are in either the men's or women's sector from the shelter service program dataset. Before conducting the t-test, I removed the missing values to makes it ready for analysis.

Results

Through using the python software, I got the t test (difference in the occupancy room rate between men and women) result that the t-statistics is 13.73 and p-value is 2.332×10^{-41}.

Discussion

If the p-value is less than 0.05, we reject the null hypothesis (H0) and conclude that there is a statistically significant difference between the average room occupancy rates in the 'Men' and 'Women' sectors. This result indicate the need for sector-specific strategies to optimise room occupancy.

If the p-value is greater than or equal to 0.05, we cannot reject the null hypothesis (H0), which suggests that there is no significant difference between the occupancy rates of the two sectors. This finding implies that the sectoral effect on occupancy rates is not significant, suggesting that there may be other factors at play or that the current strategy is equally effective across sectors.

Conclusion

Since 2.332×10^{-41} is much smaller than the 0.05. I reject the null hypothesis. There is a statistically significant difference between means of the occupancy rate of the groups (Men and Women) at 5% significant level.

Exploratory quantitative Analysis

Introduction

In recent years, the homeless population in Toronto has significantly increased. For society's wellness, it is important to ensure that the homeless are guaranteed a basic livelihood. In order to get the comprehensive understanding of the shelter's program, I decided to do exploratory quantitative analysis on the dataset. The exploratory quantitative analysis will be divided into three partings including the continuous variable analysis (room occupancy rate, bed occupancy rate), categorical variable analysis(Sector and Capacity type) and explore the relationships between categorical variable and continuous variable (relationship between capacity rate and sector type).

Method

For the continuous variable
In order to explore the central tendencies and distribution of the bed occupancy rate and room occupancy rate, I plan to do the summary statistics of the bed occupancy rate and room occupancy rate and then use the histogram and boxplot as tools to visualize the results.

Data preparation

I first load the necessary libraries including panda, matplotlib.pyplot and seaborn. Then, I converted the Excel file into a CSV file for analysis and loaded it into the Python file. I calculate the continuous variable room occupancy rate (OCCUPIED_ROOMS/CAPCACITY_ACTUAL_ROOM) and bed occupancy rate (OCCUPIED_BEDS/CAPCACITY_ACTUAL_BED. I think the occupancy rate is better for comparison. I added these two columns of room occupancy rate and bed occupancy rate. Before conducting analysis, I removed the missing values in the 'Bed_OCCUPANCY_RATE' and 'Room_OCCUPANCY_RATE'.

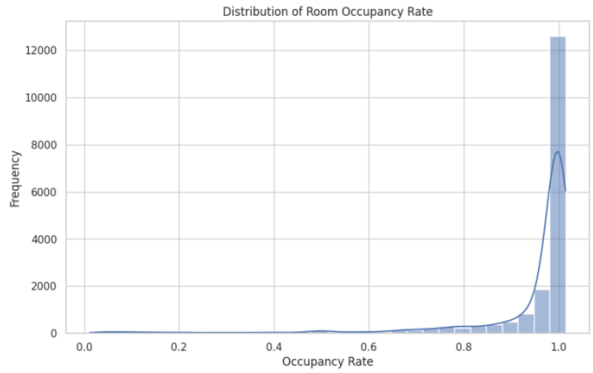
Analysis 1: Continuous Variable (Room occupancy rate and Bed occupancy rate)

Result

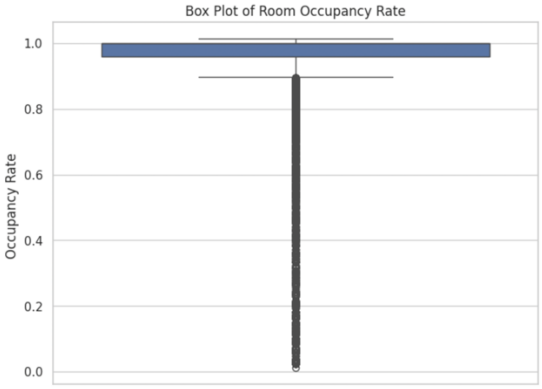
Summary Statistics (Room occupancy rate and Bed occupancy rate)

	Room_OCCUPANCY_RATE	Bed_OCCUPANCY_RATE
count	18545.000000	32399.000000
mean	0.934087	0.927885
std	0.163241	0.122562
min	0.012048	0.022727
25%	0.958333	0.900000
50%	1.000000	1.000000
75%	1.000000	1.000000
max	1.014085	1.000000

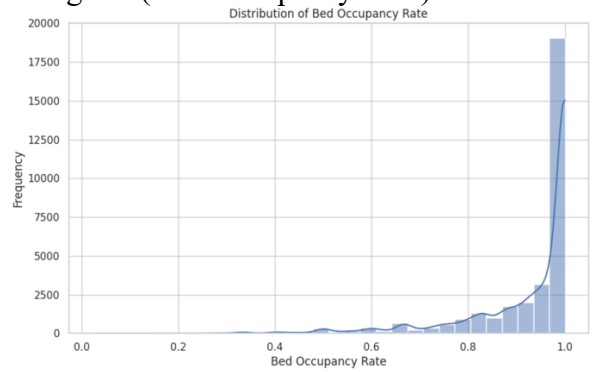
Histogram (Room Occupancy Rate)



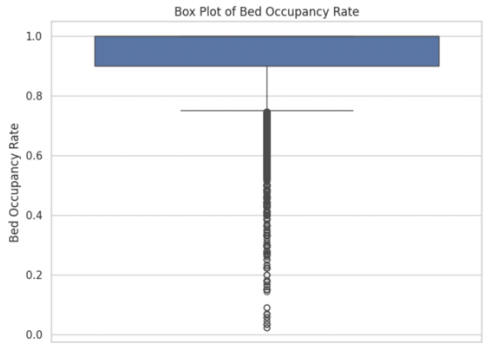
Boxplot (Room Occupancy Rate)



Histogram (Bed Occupancy Rate)



Boxplot (Bed Occupancy Rate)



Discussion

Room Occupancy rate

Summary statistics: There are 18545 observations. On average, rooms are 93.4% occupied. standard deviation of 0.163 shows that there are some variabilities. The minimum room occupancy rate is 1.2%. The maximum occupancy rate is 101.4%. There are some rooms are underutilized. Over 100% occupancy rate is possibly because of the overbooking or errors. The median is around 100%.

Histogram: the distribution of room occupancy rate is left-skewed. Most of the data is around 1.0. High occupancy rate is very common but there are still few low occupancies rate. Mode > Median > Mean shows that lower values pull the mean down more than they affect the median or mode.

Boxplot: The room occupancy rate is concentrated on the high end. This means that the high occupancy rate is very common. The above 1.0 is the outlier. Some outliers below the lower end bound.

Bed Occupancy rate

Summary Statistics: There are 32399 observations. On average, beds are 92.8% occupied. standard deviation of 0.123 shows that there are some variabilities. The minimum room occupancy rate is 2.3%. The maximum occupancy rate is 100%. There are some rooms are underutilized. There is no case that the used bed is higher than the available beds. The median is around 100%.

Histogram: the distribution of room occupancy rate is left-skewed. Most of the data is around 1.0. High occupancy rate is very common but there are still few low occupancies rate.

Boxplot: The room occupancy rate is concentrated on the high end. This means that the high occupancy rate is very common. The outlier is on the lower end of the scale.

Analysis 2: Categorical Variables (Sector and Capacity type)

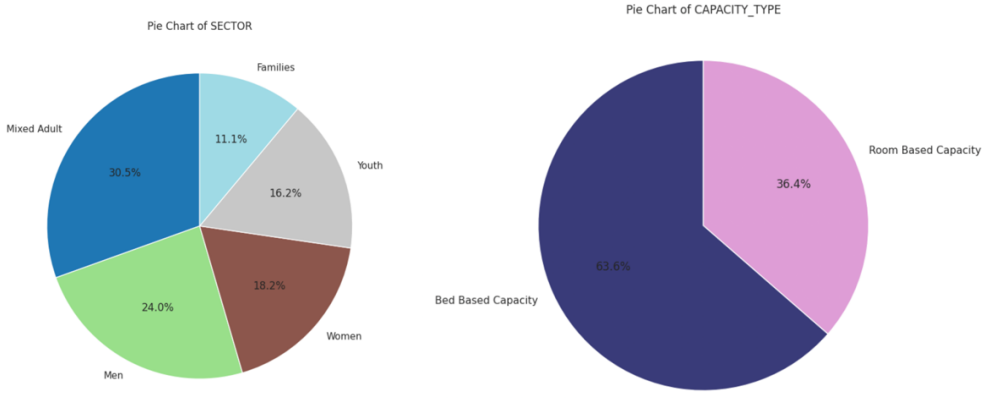
Result

Frequency Counts (Sector and Capacity type)

```
Frequency Counts for SECTOR:
Mixed Adult    15533
Men            12241
Women           9258
Youth           8263
Families        5649
Name: SECTOR, dtype: int64
```

```
Frequency Counts for Capacity type:
Bed Based Capacity     32399
Room Based Capacity    18545
Name: CAPACITY_TYPE, dtype: int64
```

Visualization (pie chart)



Discussion

Sector type

There are 5 sector types of homeless including Mixed adult, men, women, youth, families. The mixed adult is the most common sector type with 30.5%. There is a huge shelter service demand

for the adults. The proportion of men and women are very close (men: 24%, women: 18.2%). The relatively close proportions suggest a balanced approach to addressing the needs of both genders, though programs for men are slightly more prevalent. The Families sector takes 11.1% of the observations, the smallest among the categories analyzed. It shows that while there is support for families, the focus or availability of dedicated family services is less compared to services aimed at individuals or gender-specific groups.

Capacity type

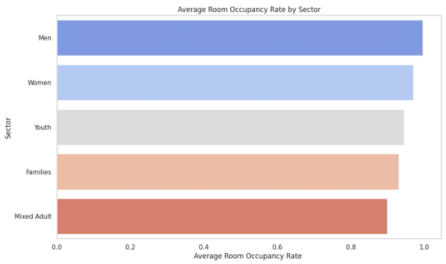
There are two types of capacity including the bed-based capacity and room based capacity. Bed based capacity takes the higher proportion than the room-based capacity. Bed based capacity is more common than the room-based capacity. It means that a majority of the programs are designed to maximize the use of space and accommodate as many individuals as possible in shared sleeping arrangements.

Analysis 3: Explore Relationships (Capacity rate and sector type)

Result

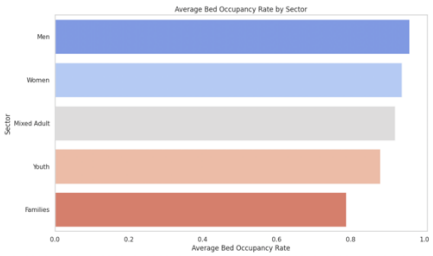
Room occupancy rate by sector

```
        SECTOR  Room_OCCUPANCY_RATE
1          Men             0.996345
3        Women             0.969882
4        Youth             0.945445
0     Families             0.930657
2  Mixed Adult             0.899828
<ipython-input-66-2e41ede89c22>:11: FutureWarning:
```



Bed Occupancy rate by Sector

	SECTOR	Bed_OCCUPANCY_RATE
1	Men	0.960008
3	Women	0.939212
2	Mixed Adult	0.920749
4	Youth	0.880683
0	Families	0.788377



Discussion

Room Occupancy rate by sector

Findings: For the room occupancy rate, the men have the highest room occupancy rate of (99.6%). This suggests that programs designed for men are operating very close to their full room capacity, indicating high demand or efficient management of resources within this demographic. The room occupancy rate in the female sector is 96.9%, which is slightly lower than in the male sector, but still shows a high utilisation rate. This reflects a strong demand for female-specific programmes, but also shows that these services are also close to saturation. The Families sector has the lowest room occupancy rate at 89.98%. This shows that underutilization of bed-based capacities in family-specific programs.

Further Analysis: In order to provide more efficient service for the homeless, we could explore more about the reasons why the family has the lowest room capacity rate.

Bed Occupancy rate by sector

Findings: In terms of bed occupancy, males have the highest bed occupancy rate (96.0%). This indicates that the bed occupancy rate for programmes designed for males is very close to saturation, which suggests that there is a high level of demand from this population or that resources are being managed effectively. The bed occupancy rate for the female sector, at 93.92%, is slightly lower than the male sector, but still shows a high rate of use. This reflects a strong demand for female-specific programmes, but also indicates that these services are also approaching saturation. The lowest bed occupancy rate was recorded in the family sector at 78.83%. It means that underutilisation of beds in family specific projects.

Further Analysis: In order to provide more efficient service for the homeless, we could explore more about the reasons why the family has the lowest bed capacity rate.

Relationship Interesting findings: I found that the overall room capacity rate is higher that the overall bed capacity rate. For the future study, I can explore the reasons behind it. Homeless prefer the room than the bed.

Possible Research Question

1. How well do current program capacities align with the demand across different sectors?
2. What factors contribute to the varying occupancy rates observed across different sectors (e.g., Men, Women, Families)?
3. How do occupancy rates fluctuate throughout the year, and what implications does this have for program operation and planning?