## DESCRIPTIVE STATISTICS

The data shows daily occupancy and capacity for Toronto shelters for 2021. The data contains 50944 rows. The PROGRAM_ID column gives us a unique ID to perform analysis. In 2021, there were 169 unique shelter programs in Toronto. These programs were administered by 35 unique organizations. Most of the programs were geared towards "mixed adults," these are typically co-ed or all gender programs. There are a total of 53 "mixed adult" programs. The shelter programs geared towards Families and Youth were the most limited at only 21 and 25 programs, respectively. See the table and graph below for details.

#	SECTOR	UNIQUE PROGRAM ID COUNT
1	Families	21
2	Men	39
3	Mixed Adult	53
4	Women	31
5	Youth	25
	TOTAL	169
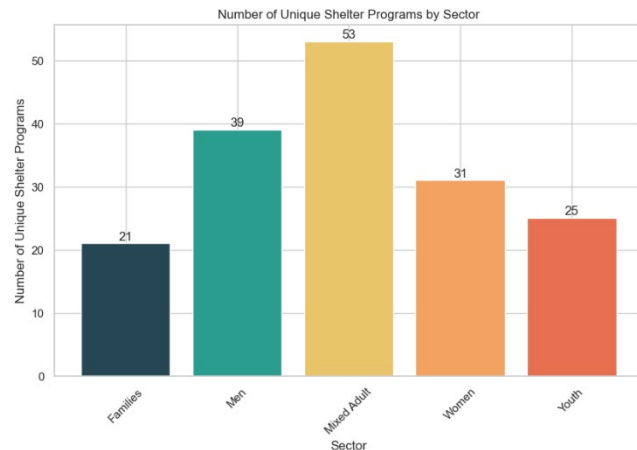
As illustrated by the graph below the mixed adult sector had the higher number of service user counts at only 35,000 shy of 1 million people for a total of 964,801 users. The City of Toronto and other non-profit organizations serviced a total of 2,329,525 users in the year 2021. We do not know if these are unique users as we do not have unique ids for service user counts.
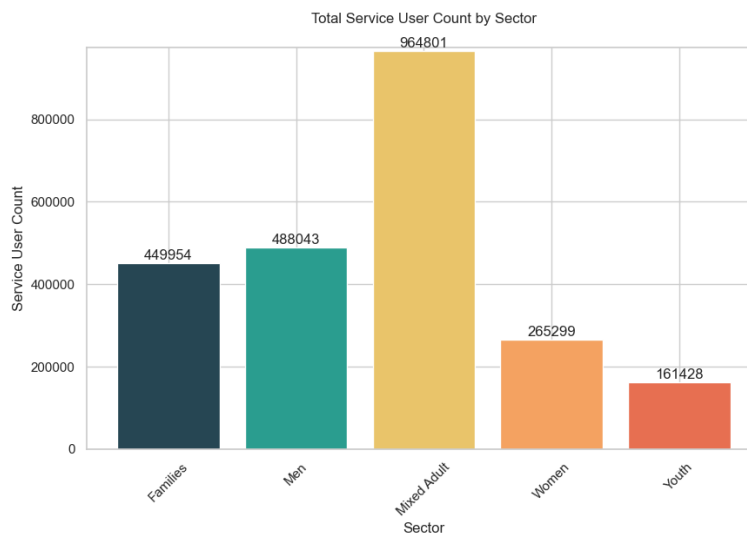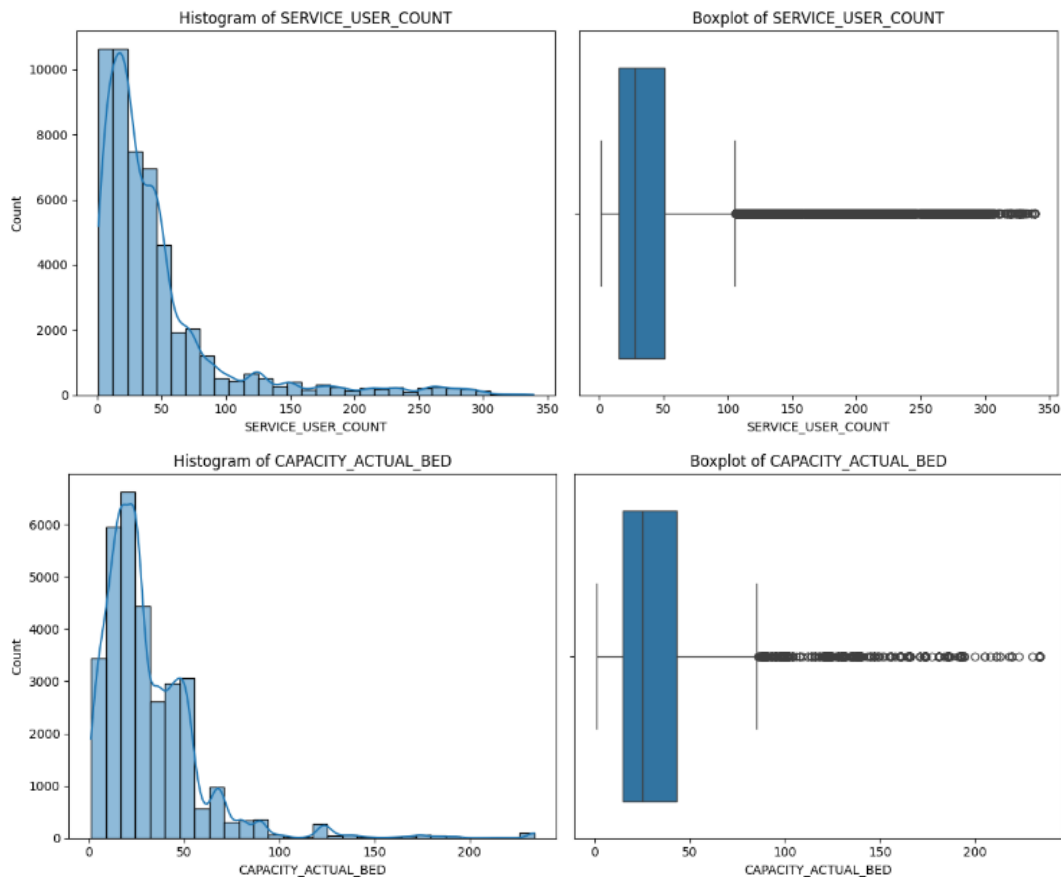

In 2021 there were 140 Emergency shelter programs and 28 Transitional ones. We have 1 program for which we do not have a program model. This program model has an ID of 16631 and I have labelled this program model as "Unknown."
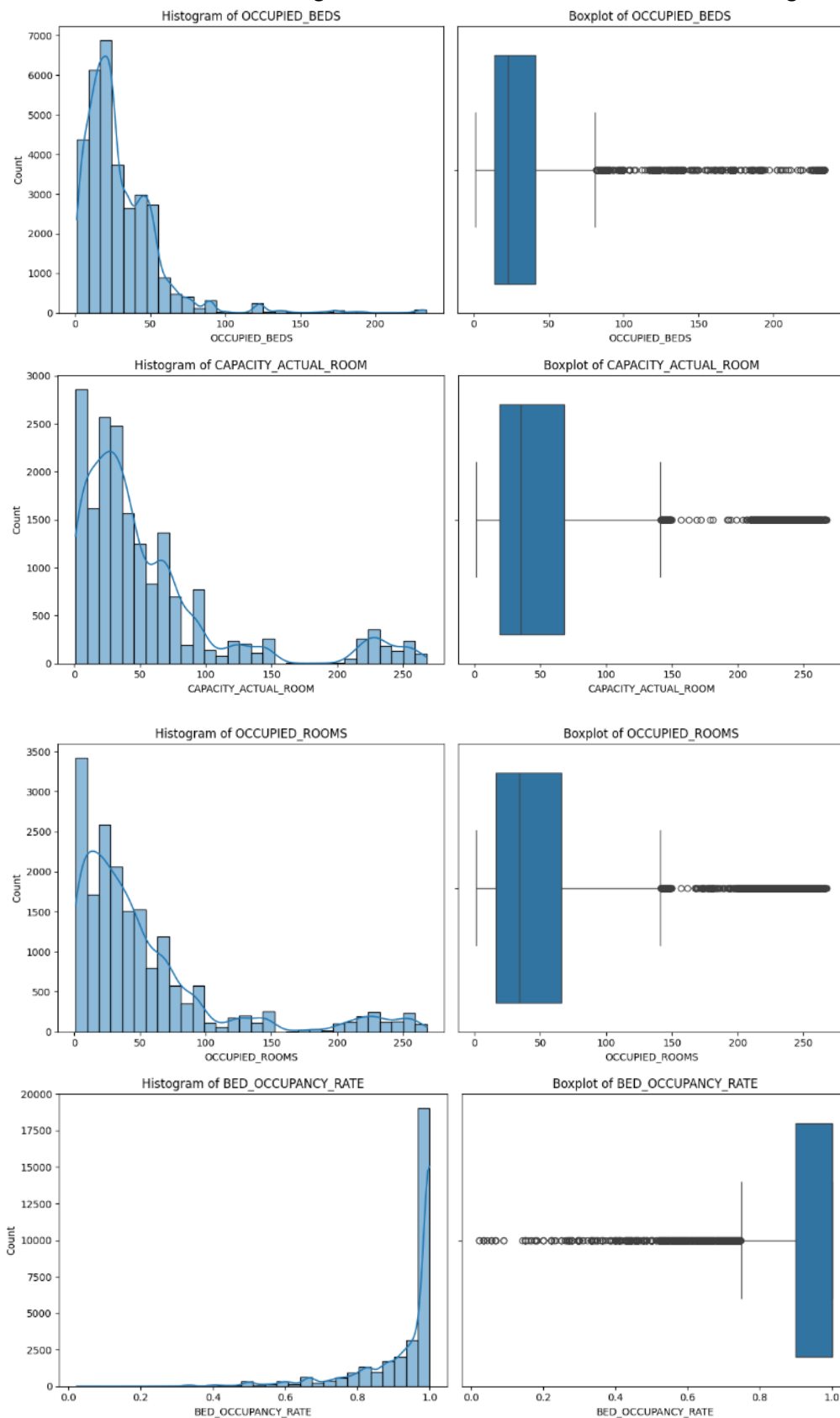
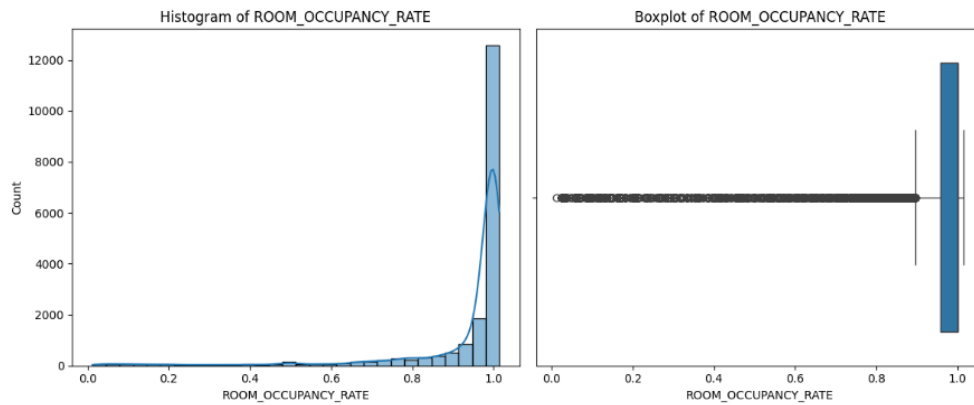| #	PROGRAM MODEL	UNIQUE COUNT
1	Emergency	140
2	Transitional	28
3	Unknown	1
	TOTAL	169

In 2021, the City of Toronto itself administered 45 unique programs; 124 programs are administered by various non-profits and charities. Three of these organizations combined offered almost as many programs as the City of Toronto. The three organizations are COSTI Immigrant Services at 13 programs, Homes First Society at 19 programs, and The Salvation Army of Canada at 11 programs. This amounts to a total of 43 unique programs administered by three non-profits. The city has 101 shelters with bed-based capacity and 68 shelters with room-based capacity.

#	CAPACITY TYPE	UNIQUE COUNT
1	Bed Based	101
2	Room Based	68
	TOTAL	**169**

Analysis of all the quantitative columns reveals that none of them contain data that is normally distributed. The data is mostly skewed, with some exceptions. This makes sense intuitively as we would expect only a subset of individuals in the population who find themselves in dire circumstances to experience homelessness.

Histogram of OCCUPIED_BEDS

Boxplot of OCCUPIED_BEDS

Histogram of CAPACITY_ACTUAL_ROOM

Boxplot of CAPACITY_ACTUAL_ROOM

Histogram of OCCUPIED_ROOMS

Boxplot of OCCUPIED_ROOMS

Histogram of BED_OCCUPANCY_RATE

Boxplot of BED_OCCUPANCY_RATE

We can see from the histograms above that none of these columns follow a gaussian distribution. Furthermore, the boxplots above indicate that the data is skewed and there are a significant number of outliers.

Both the means of daily service user count and the monthly service user count generally increase as the days and year progresses. In other words, the service user count mean for the month of December is higher than that for the month of November. This is a general trend with some small dips in the data.
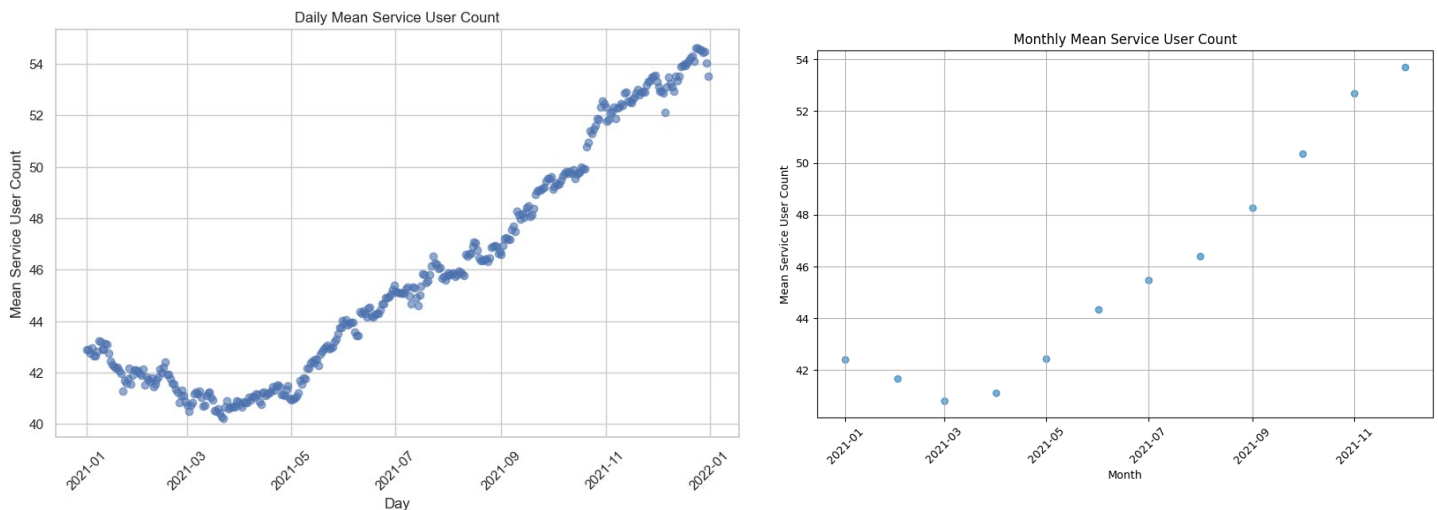

There are only 3 winter programs for the whole city which is somewhat shocking as the weather in Canada can be extremely cold.

INFERENTIAL STATISTICS

Let's examine the boxplots to see what kind of statistical analysis we can perform on the data. The plot below shows Service User Count by Sector. It would be interesting to examine whether the null hypothesis holds in this case. In other words, is it true that there is no statistical difference in the mean service user count between the "Families" sector and the "Mixed Adult" sector. Running a t-test here would tell us whether the difference in means is statistically significant.

We are going conduct t-tests using Welch's t-test. The reason for this is because we are working with real-world data. We have already demonstrated earlier using histograms and boxplots that the equal variances assumption does not hold for our data. We have unequal sample sizes, and the data is heteroscedastic.

Service User Counts by Sector
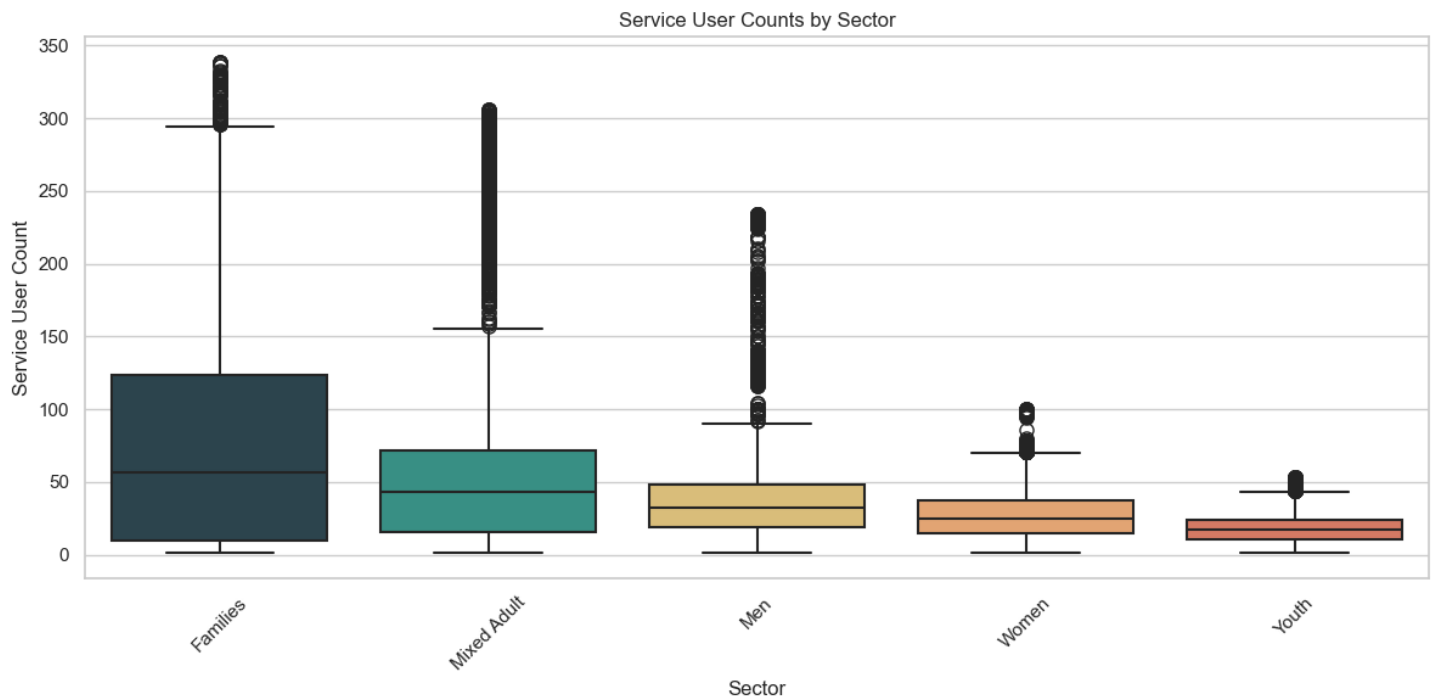
Running the T-test described above gives us a t-statistics of 15.12 and a p-value of 4.69×10^{-51}. The t-statistic indicates that there is significant difference between the means of the service user count for "Families" and "Mixed Adult." The p-value of less the 0.05 indicates to us that this difference is statistically significant. In other words, we have strong evidence to reject the null hypothesis we articulated above. We can conclude form this that the higher service user count for the "Families" sector is statistically significant. This could indicate that the need for shelters for families is greater than people commonly expect. We know that the service user count for the "Mixed Adult" sector is much higher at 964,801 when compared to the service user count for the "Families" sector which is at 449,954. Clearly, our results provide strong evidence to suggest that the City of Toronto should fund more shelters programs targeted towards the "Families" sector.

Similarly, the boxplots below show Program Model by Bed Occupancy Rates and Program Model by Room Occupancy Rates. Let's compare the bed occupancy rate between the two programs models. Let us also compare the room occupancy rate between the two program models. This would be beneficial in revealing whether one program model results in better utilization of resources offered through shelter programs.
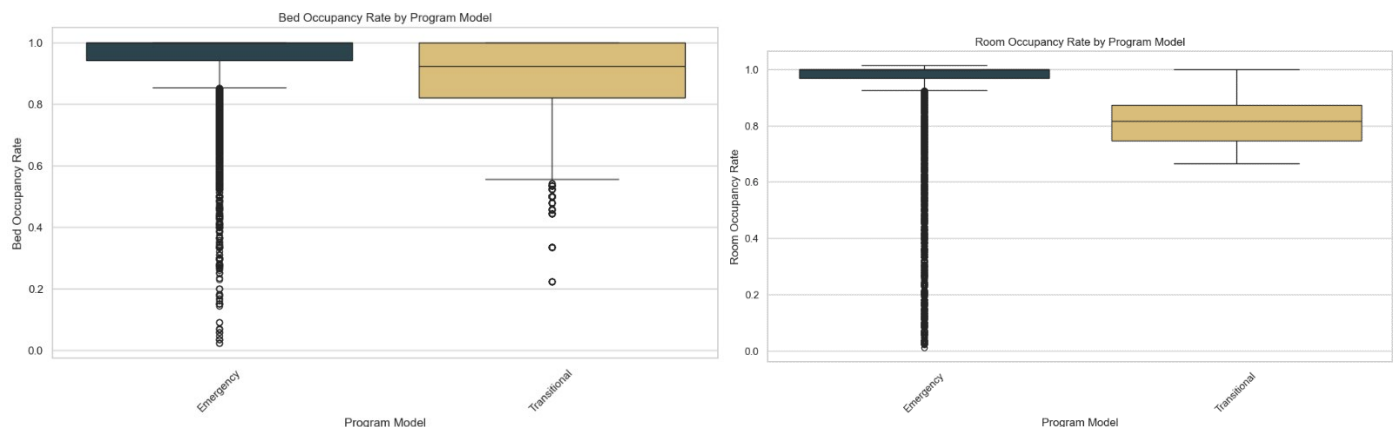
Again we are going to use the Welch's T-test in this scenario as the data is incredibly skewed. Furthermore, the sample size between the two program models is also very skewed. The T-test for bed occupany rate between the two program models give us a t-statistic of 36.78 and a p-value of 7.27×10^{-283}. On the other hand, the T-test for room occupancy rate between the two program models gives us a t-statistic of 31.71 and a p-value of 4.43×10^{-150}. Our null hypothesis in both cases is that different programs models have no affect on bed occupancy rate or room occupancy rate. We can see from our T-test which give us a p-value of less that 0.05 in both scenarios that there is evidence to suggest that programs models have a statically significant impact on bed occupancy and room occupancy rates.

Our first T-test comparing bed occupany rates shows that bed occupancy rates under the emergency program model is higher than under the transitional program model. Similarly, the results of the second T-test reveals that the room occupany rate under the emergency program model is higher than under the transitional program model. Clearly there is evidence to suggest that the resources under the emergency program model are better utlitized as opposed to the resources under the transitional model. According to the feature description homeless shelters under the transitional model can only be "accessed by eligible individuals and families experiencing homelessness by referral only."[1] The resources under the emergency model can be accessed "by any individual or family experiencing homelessness with or without a referral."[2] This explains the discrepancy in the occupany rates between both models.

TODO

I would like to perform a deeper analysis on the data to test for normality. I supposed homeless is not normally distributed and therefore use of resources meant for those who are homeless will also not be normally distributed. This makes sense intutitively but I would like to perform more tests to see whether the data is normally distritbuted or not. The Shaprio-Wilk test is ideal for this purpose. It might help to get even more data to anaylyze to see if the distribution changes. For example, if we have data from several years would anything change. Analyzing the data pre-covid and post-covid would also produce some useful insights.

---

[1] Toronto, City of. 2022. 'Shelters'. City of Toronto. City of Toronto. Toronto, Ontario, Canada. 28 December 2022. https://www.toronto.ca/community-people/housing-shelter/homeless-help/shelters/.
[2] Toronto, Shelters, 2022.