Toronto Child Care Service Analysis

Chenyang Pan
1005131554

1. Introduction

In many of the families in Ontario, the parents may be incapable of providing their children with good education and care due to insufficient fundings. In order to help the parents with this issue, Toronto has offered more and more licensed child care service centers in recent years. The central goal of this analysis is to provide suggestions to the future design of child care centers in Toronto as to let more and more children who are from families that have difficulties offering them enough care to be able to enter the child care center.

In order to encourage more and more children to enter the child care center, the central idea is to increase the total capacity in the future child care service center. In this analysis, we will be focusing on what factor may lead to an increase in the total capacity in the service centers?

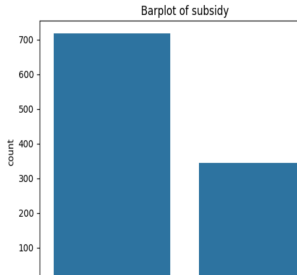
To explore this question, we will perform an in depth analysis based onToronto Licensed Child Care data, updated February 2024. To gain evidence for our suggestion.

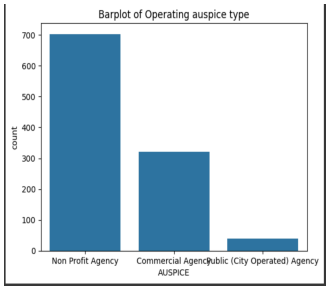
2. Data cleaning and Preliminary Analysis

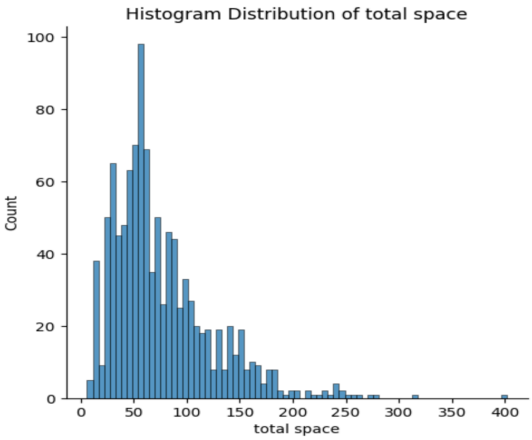
Taking the first glance of the data, there are in total, 17 different features with 9 being numerical and 8 being categorical. Since we only want to investigate interesting and relevant features. 14 features are dropped from the data set. Three features that are kept are “TOTSPACE” the total space capacity, and “subsidy” whether the institution has a subsidy. The AUSPICE, and subsidy are two potential independent variables (categorical) and the TOTSPACE is the potential response variable (numerical) in our analysis. Then the NA terms in the data set are dropped to do further analysis. The reason that the total space of the service center is kept as the only numerical variables is because, as mentioned, increasing the total space of the service center is the priority consideration to be able to let more children enter the shelter. Auspice and subsidy are two most direct potential relevant factors that may be affecting the space in the child service center.

First, two bar plots are made to visually observe the difference of data distribution between the number of service centers with subsidy and without subsidy (Figure 1) and between the number of service centers by each auspice type (Figure 2). By the barplot of Subsidy, we can detect that the child care centers in the society with subsidy are around two times compared to those without subsidy. From the auspice type bar plot, we realize that Non profit agency is the major type of auspice in society and there are only very few public agency service centers.

Figure 1 Figure 2







Then, a histogram of total space is also made to observe the overall data distribution of the total space in each child care service center, shown in Figure 3. By looking at the histogram distribution of the total space in the child care institutions, the histogram is overall right skewed and unimodal. There are also outliers that can be found at 400 capacity spaces.

Figure 3

Next, three side by side boxplots of the total space in the childcare institutions based on the auspice type (Figure 4), subsidy (Figure 5) and subsidy by auspice (Figure 6).

From Figure 4, we can realize that the total space in non profit agency service centers has the most outliers than commercial agency types and also has the highest median among the three types. On the other hand, public agencies have no extreme values and have the lowest median among three types. This provides us some insight that three different auspice types may have different total space. From Figure 5, we realize that the total space in the service center with and without both has extreme values on the right end of the boxplots. Moreover, the median of space in service centers with subsidy is greater than those that without subsidy.

Figure 4

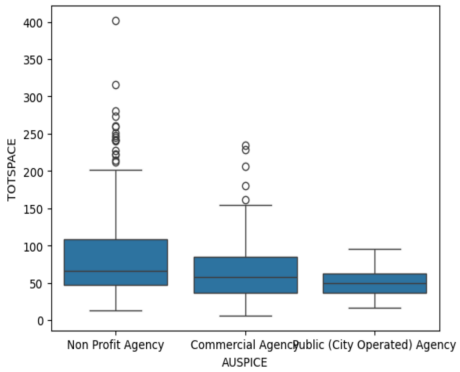
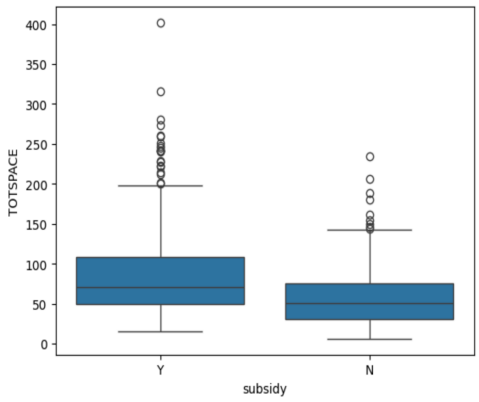


Figure 5



Lastly, from side by side boxplot of the total space based on subsidy by different auspice shown right on Figure 6. What we can conclude from this plot is that. We found that surprisingly the commercial agency service center with subsidy has the highest median in space capacity. This may lead us to think about whether there is an interacting effect between subsidy and auspice type. Moreover, we also see that the public agency service centers recorded in the data set are all supported with subsidy.

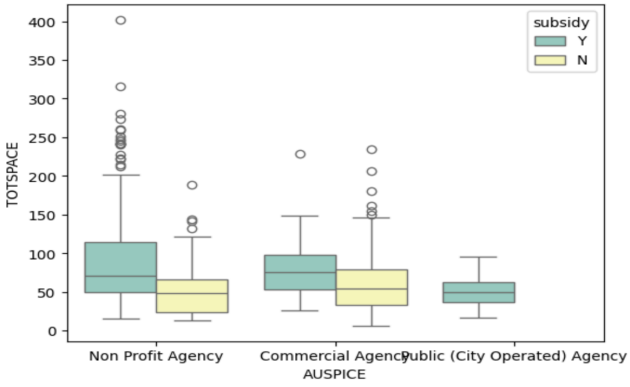


Figure 6

3. One-way ANOVA

Null hypothesis: The mean total space in all auspice type service center is the same
Alternative hypothesis: At least one type of auspice service center has different mean total space capacity

3.1 ANOVA Result

The following ANOVA table (Table 1) shows the result of the one way ANOVA test.

Table 1

	df	Sum of Square	Mean of Square	F statistic	P value
AUSPICE	2.0	9.611e+04	48056.057	21.843	5.057716e-10
Residual	1060	2.332e+04	2200.062		

From the result of the one way ANOVA test, we can detect a P value of 5.057716e-10 which is significantly smaller than the level of significance of 0.05. This means that we reject the null hypothesis which states that the mean total space in all auspice type service centers is the same. This can tell us that when making new designs of the child care service center. We should take consideration of the factor of auspice type to make the right choice.

3.4 Post Hoc Test

Post Hoc Test is performed using Tukey's HSD to further examine the pairwise difference comparison between average space capacity of each type of auspice service

center. The table of pairwise difference comparison p value is shown in table 2 below. Surprisingly, there is no significant difference between commercial agency and public agency because of a high p value of 0.077. On the other hand, Non profit agency shows significant difference when compared to other two indicating that the average total space in Non profit agency is significantly different from the other two.

Table 2

Group 1	Group 2	P value
Non profit agency	Commercial agency	0.001
Non profit agency	Public agency	0.001
Commercial agency	Public agency	0.077

3.4 Model Assumption Check

To ensure the reliability of the model, We performed model diagnostics check to see if the model violated any of the assumptions; Normality assumption, Homogeneity of Variance assumption.

By looking at the QQ plot (Figure 7) and the histogram of residual distribution (Figure 8). We can see that although the dots on two ends of the straight line are off, most of the dots stick to the straight line. However, the histogram if the residual is a bit right skewed. Also, we conducted a Shapiro-Wilk test and found that the P value is extremely small, 1.486e-25. The above clues may imply that there may be a violation of the normality assumption.

Figure 7

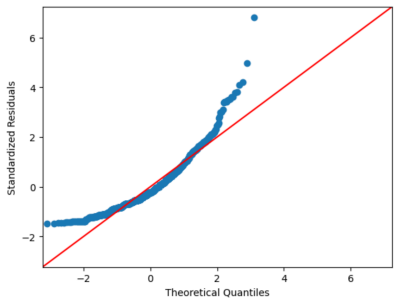
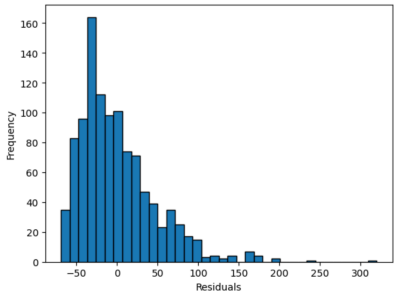


Figure 8



To check the Homogeneity of the variance, we conducted Levene's test to see if the variance is homogenous. The result of the levene's test is shown below in Table 2. We found that the P value is extremely small indicating the violation of the homogeneity of variance assumption.

Table 3

Parameter	Value
Test statistics	17.927
Degree of Freedom	2
P value	0.0000

4. Two Way ANOVA

Null hypothesis: 1. Whether subsidy exist or not, there is no main effect caused by auspices to the total space of the service center. 2. Whether auspice exists or not, there is no main effect caused by subsidy to the total space of the service center. 3. There is not interaction effect between subsidy and auspice toward the total space of the service center
Alternative hypothesis: 1. Whether subsidy exist or not, there is a main effect caused by auspices to the total space of the service center. 2. Whether auspice exists or not, there is a main effect caused by subsidy to the total space of the service center. 3. There is an interaction effect between subsidy and auspice toward the total space of the service center

4.1 ANOVA Result

By looking at the result of the two way anova, we can find that the P value associated with auspice is 1.283e-01 which is larger than the level of significance, so that we fail to reject first null hypothesis that Whether subsidy exist or not, there is no main effect caused by auspices to the total space of the service center. The P value associated with subsidy is 3.529e-10 which is much smaller than the level of significance of 0.05, so that we reject the second null hypothesis that there is no main effect caused by subsidy to the total space of the service center. Lastly, the P value associated with the interaction between auspice and subsidy is also extremely small and is smaller than the significance level. So that we reject the last null hypothesis that There is not interaction effect between subsidy and auspice toward the total space of the service center the ANOVA table is shown on Table 4

Table 4

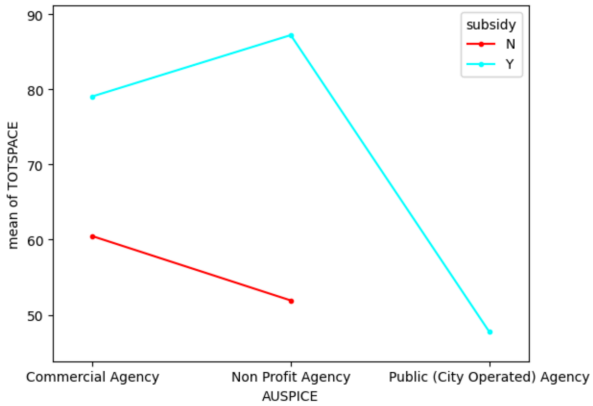
	DF	Sum of Square	Mean of Square	F	P value
AUSPICE	2	8.570e+03	4283.998	2.058	1.283e-01
subsidy	1	8.353e+04	83527.442	40.118	3.529e-10
AUSPICE: subsidy	2	5.603e+04	28017.227	13.457	1.694e-06
Residual	1058	2.203e+06	2082.050		

4.2 Interaction plot

By looking at the interaction plot between AUSPICE, subsidy and TOTSPACE. What can be found is that the two lines are not parallel from each other indicating that

there is a significant difference in relationship between auspice and total space of the child care service center when the subsidy is presented or missing. Shown in Figure 9

Figure 9



4.3 Post Hoc Test

I have conducted a Post Hoc test using Tukey's HSD. There are some important observations that can be found from the result of the test. First, similar to the one way ANOVA, Non profit agency shows difference in mean total space compared to the other two types of auspice, the other two types do not have such difference between each other. There is a significant difference in total space in the service center between centers with subsidy and without subsidy due to a small p value of 0.001. Non profit Agency service centers with subsidies have a significant difference in total space compared to other types of service center due to small p values.

4.4 Model Assumption check

Similar to the One way ANOVA, Normality assumption, Homogeneity of Variance assumption are checked with the same method and found to have similar results as the one way ANOVA test. I have detected the violation of both assumptions from QQ plot, off the straight line, and the histogram of residuals shows to be right skewed and the shapiro wilk test shows a result P value of 1.531e-25 which is significantly small. These indicate the violation of normality assumption. By conducting a levene's test to test the homogeneity of variance. What can be found is also similar as before, that the variances are not homogenous. These both assumptions are violated.

5. Conclusion

Although The analysis suggested that child care service centers that are non profit agencies with subsidies seem to have the highest average total space capacity for children which lead the suggestion to future design of the child care service center to be non profit agencies with subsidy, the assumptions of the two ANOVA test are all violated which may lead to the inaccuracy of the the result. Moreover, there may be more factors which affect the service center total capacity needed to be considered.