Exploring Child Care Space in Toronto

1. Introduction

Finding child care in Ontario is challenging for many families due to high costs and limited availability. In response, the Ontario government promised to create 100,000 new child care spaces for children by 2026. Despite the city's efforts to provide adequate child care services, many families still face the challenge of finding affordable and accessible child care. This crucial concern raises a deeper inquiry into understanding the existing child care infrastructure within the city.

This report conducts a thorough quantitative analysis of child care space availability in Toronto, aiming to discern the patterns and underlying factors contributing to the different characteristics. By examining the dataset 'INF2178_A2_data.xlsx', we will track the distribution and utilization of child care spaces, analyzing how they meet the demands of the community.

Our analysis will focus on two primary research questions, which will serve as fundamental principles in exploring the complex nature of Toronto's child care infrastructure patterns.

1. **Research Question 1:** Is there a statistically significant difference in total space among various types of child care auspices?
2. **Research Question 2:** How do the auspice of a child care facility and whether it has a fee subsidy contract independently and interactively affect the total space available?

Through exploring these two questions, our goal is to offer insights into finding some distinct features of child care infrastructure in Toronto, and to provide a more profound understanding that can guide more efficient policies.

2. Data Cleaning and Data Wrangling

Our dataset comprised a total of **17 columns** with **1,063** entities (**rows**), and there are not many missing values in most of the columns. Since we are interested in studying auspices, whether they have fee subsidy contract and child care total spaces, we will focus on related columns and leave out other unrelated columns. Below is the description of each related column:

- **AUSPICE:** Operating auspice (Commercial, Non-Profit or Public);
- **subsidy:** Centre has a fee subsidy contract (Yes/No);
- **TOTSPACE:** Child care spaces for all age groups;

3. Total Spaces among Auspices

Research Question #1: Is there a statistically significant difference in total space among various types of child care auspices?
In order to address this question, we will first look at the boxplot comparing the overall capacities of different auspice types:

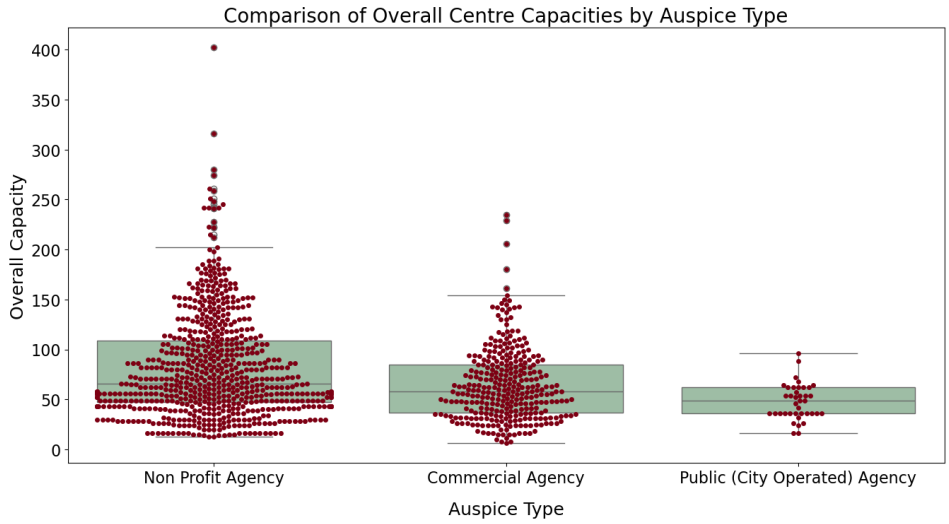

Figure 1: Comparison of Overall Centre Capacities by Auspice Type

From this figure we can see that non-profit agencies show the widest range and highest capacities, commercial agencies have a narrower spread, while public agencies have the least variability in capacity. Therefore, the first conjecture is that there will be significant difference in child care spaces for different auspices, but this requires proof from the **one-way ANOVA test**, and the result is shown below.

	Sum of Squares	df	Mean Square	F	PR(>F)
Auspice	96112.10	2	48056.05	21.843	<0.001
Residuals	2332065.00	1060	2200.06		
Total	2428177.10	1062			

Table 1: ANOVA for Comparing Auspice Type

The null hypothesis is that there is no difference in the mean overall centre capacities between the different auspice types, while the alternative hypothesis suggests there is a difference. From this ANOVA table we can see that F statistic is greater than 1 and p-value is extremely low, thus, we reject the null hypothesis and say that there are significant differences in the mean scores of different groups.

To determine which specific group or groups differ from each other, a **post-hoc test** would be necessary, as ANOVA tells us that at least two groups differ, but not which ones specifically.

Here we use **Tukey HSD test** as a way to test the difference. From the table below we can see that Public (City Operated) Agency has significantly less space on average for child care services compared to other two agencies, and one suggestion for the policy makers is that investments could be directed towards helping Public Agency increase their child care space.

Group 1	Group 2	Mean Difference	PR(>F)	Lower Bound	Upper Bound	Reject Null Hypothesis

Commercial Agency	Non-Profit Agency	17.119	<0.001	9.704	24.535	TRUE
Commercial Agency	Public (City Operated) Agency	-17.215	0.078	-35.883	1.453	FALSE
Non-Profit Agency	Public (City Operated) Agency	-34.335	<0.001	-52.445	-16.224	TRUE

Table 2: Tukey HSD Post-Hoc Comparisons for Auspice Types

Shapiro-Wilk	
Statistic	**PR(>F)**
0.902	<0.001

Table 3: Shapiro-Wilk Test of Normality for Different Auspices

In order to properly complete the ANOVA test, we also need to check if **assumptions** are satisfied. To begin with, the **normality check** was done by the **Shapiro-Wilk test**, and the result is in the table on the left. From the normality check table, we can see that the normal distribution assumption is violated.

Levene's Test	
Statistic	**P-Value**
17.927	<0.001

Table 4: Levene's Test for Homogeneity of Variances for Difference Auspices

Then we conduct **Levene's test** to check the homogeneity of variances across all groups in different auspices for their total child care spaces. From the test result table on the left, we reject the null hypothesis and there is a statistically significant difference in variances between the different auspices. Thus, the assumption of equal variances is also violated for the groups being compared.

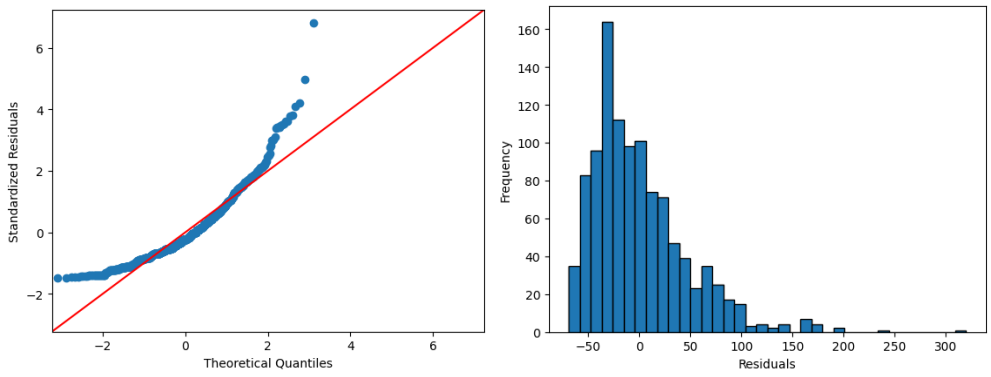

Figure 2: Q-Q Plot and ANOVA Residuals Analysis

The above **Q-Q plot** also indicates that the standardized residuals do not follow a normal distribution, and the histogram of residuals reveals a right-skewed distribution as well. Both plots suggest that the assumption of normally distributed residuals may not hold.

The last assumption is that the observations (and also errors) are independent of each other, here we assume this assumption is satisfied. Given the above violation of assumption findings, however, we still proceed our analysis and conclude that there a statistically significant difference in total space among various types of child care auspices.

4. Auspices, Subsidy Status and Their Effect on Total Spaces

Research Question #2: How do the auspice of a child care facility and whether it has a fee subsidy contract independently and interactively affect the total space available?

Apart from auspice, whether the child care facility has obtained a fee subsidy contract may also significantly influence its operational accessibility and overall service quality. Facilities with subsidy contracts often have the resources to offer more affordable care, thereby expanding access to families of various socioeconomic backgrounds. Therefore, analyzing the impact of fee subsidy contract with yes or no alongside auspice offers a comprehensive view of the factors that contribute to the effectiveness and reach of child care services.

We will first look at the **boxplot** which shows the distribution of these two features. From the diagram below we observe that Non-Profit Agencies with subsidies generally have a wider range of total space with more outliers, suggesting higher variability and some facilities with significantly more space, and Commercial Agencies display a similar pattern. Overall, subsidies appear to be associated with increases in both the median and the variability of total space available in Non-Profit and Commercial Agencies. Another thing to notice is that Public (City Operated) Agencies only show data for those receiving subsidies, and there is no data presented without subsidies, and this might have potentially affect to our analysis.

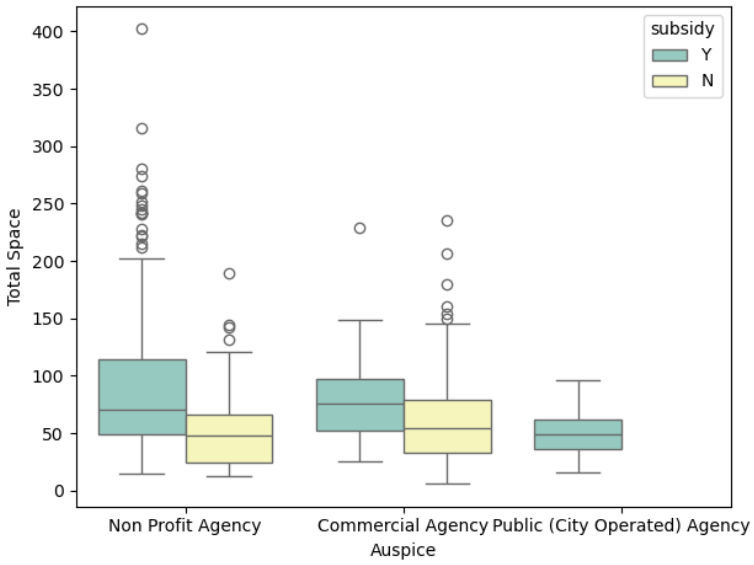

Figure 3: Side-by-side Boxplots for Child care Space Example

To investigate the research question further, we use a **two-way ANOVA** to see if there are interaction effect between these two features. One first glance from the **interaction plot** below suggests that there might exist an interaction effect between the type of agency (Auspice) and whether or not they have fee subsidy contract because of the non-parallel lines. We will then apply the interaction **two-way ANOVA model** to check the relationship between them.

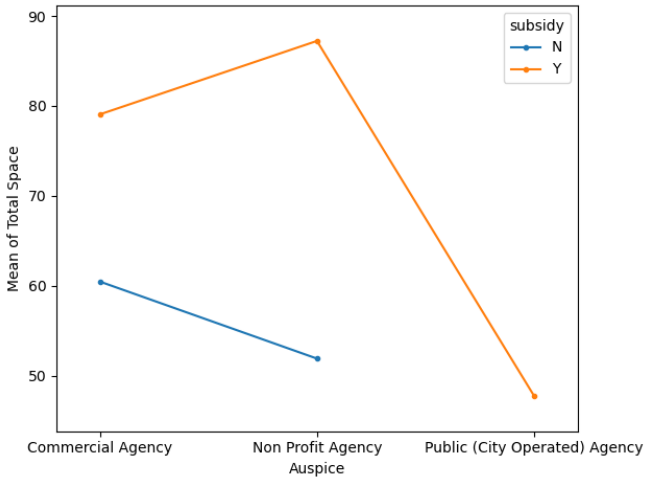

Figure 4: Interaction Plot for Child care Space with Different Subsidy Status

From the **two-way ANOVA table**, we can see that the **F statistic** for the interaction effect is 13.46, much greater than 1, and p-value is less than 0.05. Therefore, the adequacy of the additive model (**null hypothesis**) is **rejected** and both factors (auspice and subsidy) affect the outcome, which means there is an interaction between the explanatory variables. However, as mentioned above, for Public (City Operated) Agency there doesn't exist no subsidy status, and it might affect the result of this statically significance and there could be further research in this area.

Source	Sum of Squares	df	Mean Square	F	PR(>F)
AUSPICE	8568.00	2	4284.00	2.06	0.13
SUBSIDY	83527.44	1	83527.44	40.12	<0.001
AUSPICE*SUBSIDY	56034.45	2	28017.23	13.46	<0.001
RESIDUAL	2202809.00	1058	2082.05		

Table 5: Two-Way ANOVA for The Child Care Space Study

For an interaction two-way ANOVA model, **post-hoc tests** are generally conducted on the main effects as well as the interaction part. **Table 2 above** has already shown the test result for auspice, and **Table 6 below** provides the result for different subsidy status. It can be seen that having subsidy contract can significantly increase total space numbers, with a mean difference of more than 26.

Group 1	Group 2	Mean Difference	PR(>F)	Lower Bound	Upper Bound	Reject Null Hypothesis
N	Y	26.266	<0.001	20.324	32.208	TRUE

Table 6: Tukey HSD Post-Hoc Comparisons for Subsidy

Finally, we will conduct **Tukey HSD test** on the pairs of two variables, and the results are overall consistent with the above analysis. There exist significant differences in commercial agencies with and without subsidy, while with subsidy they have more overall capacity of child care spaces on average, and No-Profit Agency shows similar results.

Group 1	Group 2	Mean Difference	PR(>F)	Lower Bound	Upper Bound	Reject Null Hypothesis
Commercial Agency, N	Commercial Agency, Y	18.599	0.015	2.373	34.825	TRUE
Commercial Agency, N	Non-Profit Agency, N	-8.563	0.504	-23.273	6.147	FALSE
Commercial Agency, N	Non-Profit Agency, Y	26.765	<0.001	17.286	36.243	TRUE
Commercial Agency, N	Public (City Operated) Agency, Y	-12.696	0.489	-34.203	8.812	FALSE
Commercial Agency, Y	Non-Profit Agency, N	-27.162	<0.001	-45.916	-8.409	TRUE
Commercial Agency, Y	Non-Profit Agency, Y	8.166	0.571	-6.840	23.171	FALSE
Commercial Agency, Y	Public (City Operated) Agency, Y	-31.295	<0.001	-55.747	-6.843	TRUE
Non-Profit Agency, N	Non-Profit Agency, Y	35.328	<0.001	21.976	48.680	TRUE
Non-Profit Agency, N	Public (City Operated) Agency, Y	-4.133	0.989	-27.606	19.341	FALSE
Non-Profit Agency, Y	Public (City Operated) Agency, Y	-39.460	<0.001	-60.063	-18.858	TRUE

Table 7: Tukey HSD Post-Hoc Comparisons for Auspice and Subsidy

Just like one-way ANOVA, we should perform **assumption check** to ensure the validity of the test's results, and below are the results tables of the normality and homoscedasticity check. It can be seen from the two tables below that the dataset failed both assumptions. We can still assume errors are independent to each other. As above, we still conclude that auspice and whether it has a fee subsidy contract independently and interactively affect the total space available.

Shapiro-Wilk	
Statistic	PR(>F)
0.902	<0.001

Table 8: Shapiro-Wilk Test of Normality for Different Auspices

Levene's Test	
Statistic	P-Value
12.98	<0.001

Table 9: Levene's Test for Homogeneity of Variances for Difference Auspices

5. Conclusion

To conclude, our statistical investigation into Toronto's child care spaces has revealed key factors affecting facility capacity such as different auspices and subsidy status. The study showed that facilities with subsidies generally offer more space, especially among non-profit agencies. This suggests that subsidies are instrumental in enhancing child care service provision. While the interaction between facility auspice and subsidy status is significant, we might need to re-examine the dataset using another way as it currently violates assumptions, and Public (City Operated) Agency only has data with subsidy. Future studies could help deepen understanding and offer more profound insights for shaping child care policies as well.