

# Exploring Children Care Centres in Toronto

Xiaojing Liu (#1004263888)

## 1. Introduction

Over the past few years, Toronto has faced a shortage of affordable and accessible childcare. Specifically, there is a long waitlist for childcare spaces, prohibitively high fees that exclude many families, and a scarcity of care options that meet diverse needs. Toronto has one of the highest populations in Canada, so the growing population exacerbated the crisis. It influences not just the immediate well-being of children but also the economic and social fabric of the city. It is because the lack of affordable childcare limits parent's ability to work and contributes to economic strains and perpetuating cycles of poverty. This issue draws the attention of the policy maker. For society's well-being, they have taken some actions. To ensure the issue is properly addressed, understanding and identifying the key factors in Toronto's childcare system is very important.

This report will provide a quantitative analysis of one-way ANOVA and two-way ANOVA of Toronto's childcare system throughout 2024. To explore factors that influence the availability and accessibility of childcare services, the auspice of childcare facilities and the impact of government subsidies play important roles. This report will conduct a quantitative analysis of one-way and two-way ANOVA to examine how these factors individually and collectively influence the availability of childcare space in Toronto. I will delve into the dataset named 'INF2178\_A2\_data.xlsx' (accessible in this GitHub repository) to explore how the type of operating auspice and subsidy influence the availability of the childcare space.

My exploration will address the following two fundamental research questions, serving as guiding principles in exploring how factors individually or collectively influence the availability of the childcare space in Toronto.

- **Research Question 1:** Does the auspice of childcare facilities (i.e., Commercial, Non-Profit, Public) significantly affect the total number of childcare spaces available in Toronto?
- **Research Question 2:** How do the auspice of childcare facilities and government subsidies collectively influence the total number of childcare spaces available in Toronto?

By addressing these questions, we aim to contribute insights into the dynamics of childcare spaces in Toronto and provide a deeper understanding and ideas for policymakers to take action to address and tackle this problem.

## 2. Data Preparation

The raw dataset has a total 1063 entries (rows) and 18 columns. Before doing the one-way ANOVA, I plan to firstly do the data cleaning to ensure that there is no missing values on the targeted columns and rows. Also, Since ANOVA relies on several key assumptions, before conducting the analysis, I will check these assumptions.

### A. observation and Consideration

Since my analysis is quantitative, I am going to focused on the two columns (TOTSPACE and AUSPICE) for the one-way ANOVA analysis, and focused on the three columns (TOTSPACE, subsidy and AUSPICE) for the two-way ANOVA analysis. Below I provided a short description for each targeted column.

- AUSPICE: operating auspice (commercial, non-profit or public)
- TOTSPACE: childcare spaces for all age groups
- Subsidy: centre has fee subsidy contract (Yes/No)

**B. Handling the missing values** in the targeted columns (TOTSPACE, subsidy and AUSPICE)  
Through the research, I found there are no missing values in columns (TOTSPACE, subsidy and AUSPICE).

### 3. One-Way ANOVA Analysis: Boxplot and Swarm plot

By plotting the boxplot and swarm plot before conducting one-way ANOVA, I can quickly detect the distribution of the data including the outliers. Outliers can significantly affect the result of an ANOVA. Also, getting the distribution of data can check if there is a deviation from the ANOVA analysis assumption. Lastly, visually comparing groups can provide an immediate sense of whether there are obvious differences between the groups. Below is a combined box plot and swarm plot for three different agencies (non-profit, commercial, and public).

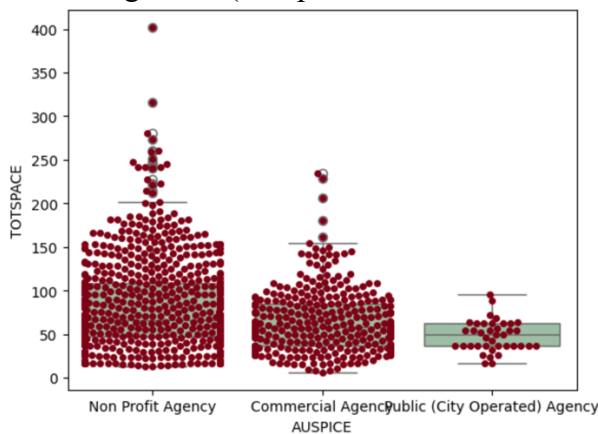


Figure 1: Boxplot and Swarm plot of AUSPICE (non-profit, commercial and public)

#### **Findings:**

**Spread of Data:** The Non-Profit Agency has a wider interquartile range compared to the other two, indicating more variability in the data.

**Outliers:** There are several outliers in the data for the Non-Profit Agency and some for the Commercial Agency. The Public (City Operated) Agency seems to have fewer outliers.

**Central Tendency:** The median (indicated by the line in the middle of the box) for each group can be compared. The Non-Profit Agency's median appears to be higher than the others, suggesting that this group typically has a higher value for the variable being measured (TOTSPACE).

### 4. One-Way ANOVA Analysis: OLS Model and ANOVA table

OLS is widely used for predicting a quantitative outcome, estimating relationships between variables and testing hypotheses about the relationships. I create the ANOVA table as well to show the relationships. The independent variable is AUSPICE and dependent variable is TOTSPACE.

	df	sum_sq	mean_sq	F	PR(>F)
<b>C(AUSPICE)</b>	2.0	9.611211e+04	48056.057145	21.843051	5.057716e-10
<b>Residual</b>	1060.0	2.332065e+06	2200.061571	NaN	NaN

Figure 2: One-way ANOVA Analysis table

#### **Interpretation:**

**Sum-sq (Sum of Squares):** For C(AUSPICE), the sum of squares is 9611211e+04, indicating the variability due to the group differences. The Residual sum of squares is 2.332065e+06, indicating the variability within the groups.

**F-ratio:** the F-statistic for C(AUSPICE) is 21.843051, which is quite high and indicates that the group means are likely different.

**PR (>F) (p-value):** The p-value for C(AUSPICE) is extremely low (5.057716e-10), which is far below the usual significance level of 0.05. This suggests that the group means are statistically significantly different.

In summary, the ANOVA results indicate that the variable TOTSPACE is statistically significantly different at different groups of AUSPICE. The low p-values mean that these results are unlikely to be due to chance, and therefore there is a meaningful difference in mean TOTSPACE between groups as defined by AUSPICE.

## 5. One-Way ANOVA Analysis: Post-hoc Test

After the ANOVA, we can conclude that at least one group is different, but not which groups are different. Therefore, we should then do the post-hoc test (Tukey's Honest Significant Difference test) to determine exactly which groups are significantly different from each other. Tukey's HSD is a method for making pairwise comparisons while controlling for Type I errors in multiple testing.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	16.806538	3.993722	29.619353	4.356853	0.006071
1	Non Profit Agency	Public (City Operated) Agency	36.177966	8.673910	63.682022	4.369046	0.005901
2	Commercial Agency	Public (City Operated) Agency	19.371429	-10.141900	48.884757	2.180132	0.272554

Figure 3: Tukey HSD post-hoc test result table

Interpretation:

- 1) There is a statistically significant difference in the means of TOTSPACE between the Non-Profit Agency and Commercial Agency. The mean difference is approximately 16.81, and the result is significant at the p-value of 0.006071, which is less than the common alpha level of 0.05.
- 2) There is also a statistically significant difference in the means of TOTSPACE between the Non-Profit Agency and Public Agency. The mean difference here is approximately 36.18, with a significant p-value of 0.005901.
- 3) However, the difference in the means of TOTSPACE between the Commercial Agency and Public Agency is not statistically significant, as indicated by the p-value of 0.272554, which is greater than the alpha level of 0.05.

## 6. Two-Way ANOVA Analysis: boxplot

The reason for making this box plot is the same as the reason we made the box plot for the One-way ANOVA before.

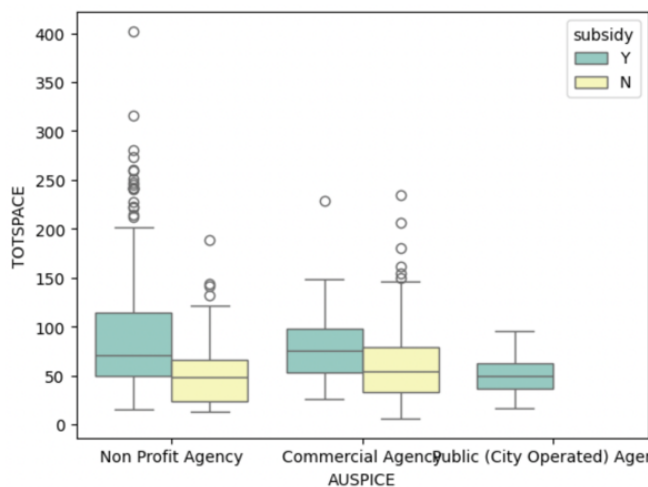


Figure 4: Boxplot TOTSPACE values for the different groups of AUSPICE with subsidy (Y/N)

#### Interpretation:

**Central tendencies:** All of the public agencies are subsidized. Generally, for non-profit agency and commercial agency, the median of total space with subsidy is higher than that of without subsidy.

**Spread of Data:** Specifically, non-profit agencies with subsidy have a wide interquartile range of "TOTSPACE". The interquartile range for commercial range is slightly narrower. The public agencies have the narrowest range of "TOTSPACE".

**Outliers:** non-profit agency has a significant number of outliers, especially for agencies that receive subsidies. It indicates that there are some non-profit agencies with much larger space than the majority. For commercial agency, the outliers are larger for the one with subsidy group than the one without subsidy, but overall fewer than in the non-profit agency. The public agency has the fewest outliers.

## 7. Two-way ANOVA analysis: OLS model and ANOVA table

	df	sum_sq	mean_sq	F	PR(>F)
<b>C(AUSPICE)</b>	2.0	8.567996e+03	4283.998054	2.057586	1.282730e-01
<b>C(subsidy)</b>	1.0	8.352744e+04	83527.441787	40.117876	3.529094e-10
<b>C(AUSPICE):C(subsidy)</b>	2.0	5.603445e+04	28017.227037	13.456555	1.694282e-06
<b>Residual</b>	1058.0	2.202809e+06	2082.050461	NaN	NaN

Figure 5: Two-way ANOVA analysis table

#### **Interpretation:**

- 1) The F-value of 2.057586 and the P-value of 0.128273(0.128730e-01) (It is greater than the commonly used alpha level of 0.05) suggests that there is not a statistically significant effect of AUSPICE on the dependent variable at the 0.05 level.
- 2) The F-value is quite high at 40.117876 while the P-value is extremely low (3.529094e-10 which is almost 0). This indicates that there is very strong statistical evidence that there is a significant difference in the mean value of "TOTSPACE" between the subsidized and unsubsidized groups.
- 3) The F-value of this interaction term is 13.456555 and the p-value is 1.694282e-06. It is less than 0.05. It indicates that there is a significant interaction between institution type and whether it receives subsidies on "TOTSPACE". In other words, the effect of receiving subsidies on "TOTSPACE" varies by AUSPICE type.

In summary, it suggests that there is no significant effect of AUSPICE alone on the dependent variable, but there is a significant effect of subsidy alone, and there is also a significant interaction effect between AUSPICE and subsidy.

## 8. Two-way ANOVA analysis: Interaction plot



#### Interpretation:

For entities with a subsidy, the mean 'TOTSPACE' is higher for 'Non-Profit Agency' than for 'Commercial Agency', and lower for 'Public Agency compared to 'Non-Profit Agency'. For those without a subsidy, the mean 'TOTSPACE' appears to be fairly consistent across all categories, with a slight variation.

Figure 6: Interaction plot for two-way ANOVA analysis

## 9. Two-way ANOVA analysis: Post-hoc Test

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	16.806538	4.406270	29.206805	4.501830	0.004333
1	Non Profit Agency	Public (City Operated) Agency	36.177966	9.559488	62.796444	4.514429	0.004205
2	Commercial Agency	Public (City Operated) Agency	19.371429	-9.191627	47.934484	2.252677	0.249650

Figure 7: Tukey's Honest Significant Difference test-1

**Interpretation:** there are significant differences in 'TOTSPACE' between Non Profit Agencies and both Commercial Agencies and Public agencies (p-value<0.05). However, there is not a significant difference between commercial agencies and public agencies (p-value>0.05).

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Y	N	34.121178	24.873006	43.36935	10.244062	0.001

Figure 8: Tukey's Honest Significant Difference test-2

**Interpretation:** there is a statistically significant difference in the means of 'TOTSPACE' between the groups that received a subsidy and those that did not (p-value <0.05). Specifically, the group that received a subsidy ('Y') has a higher mean 'TOTSPACE' compared to the group that did not receive a subsidy ('N').

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(Non Profit Agency, Y)	(Non Profit Agency, N)	44.986164	25.584897	64.387430	9.370204	0.001000
1	(Non Profit Agency, Y)	(Commercial Agency, Y)	0.145138	-28.465689	28.755965	0.020500	0.900000
2	(Non Profit Agency, Y)	(Commercial Agency, N)	28.370552	11.192465	45.548638	6.674107	0.001000
3	(Non Profit Agency, Y)	(Public (City Operated) Agency, Y)	40.752830	8.307960	73.197700	5.075887	0.004757
4	(Non Profit Agency, Y)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900000

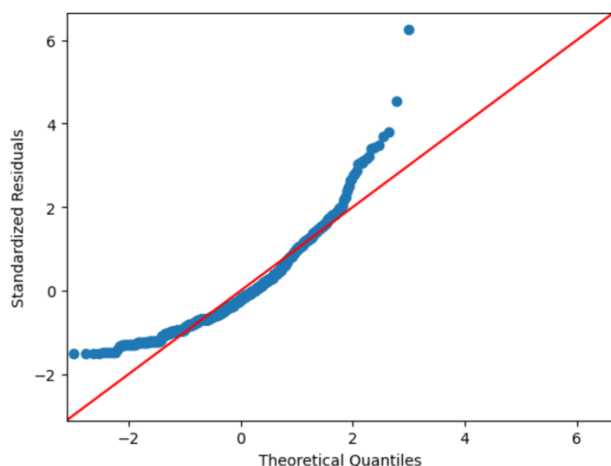
Figure 9: Tukey's Honest Significant Difference test-3

**Interpretation:** Comparing 'Non-Profit Agency, Y' with 'Non-Profit Agency, N', there is a very significant mean difference in 'TOTSPACE' (p-value = 0.001). Comparing 'Non-Profit Agency, Y' with 'Commercial Agency, Y', there is no significant mean difference (p-value = 0.9). Comparing 'Non Profit Agency, Y' with 'Commercial Agency, N', there is a significant mean difference (p-value = 0.001). Comparing 'Non-Profit Agency, Y' with 'Public (City Operated) Agency, Y', there is a significant mean difference (p-value = 0.004757). The comparison between 'Non-Profit Agency, Y' and

'Public (City Operated) Agency, N' has a difference of 0, which implies that there is no difference between these two groups, and this is reflected in the p-value (p-value = 0.9).

## 10. ANOVA assumption check

### Assumption 1: the residuals should be normally distributed



#### **Interpretation:**

The Q-Q plot is used to check Assumption 1 of ANOVA, which is that the residuals should be normally distributed. If the points on the Q-q plot closely follow the reference line, the assumption is considered to be met. From the graph, we can find that there is a deviation from the red line in the tails. It may violate the assumption of normality for the residuals. The outliers influence the model's estimate.

### Assumption 2: variance is homogenous: Levene's test

It is an important assumption check for ANOVAS assumption (variance are homogenous).

	Parameter	Value
0	Test statistics (W)	9.1994
1	Degrees of freedom (Df)	2.0000
2	p value	0.0001

**Interpretation:** Given the low p-value, you would reject the null hypothesis of Levene's test, which is that the group variances are equal. This indicates that the variances of 'TOTSPACE' for different 'AUSPICE' groups are not equal, which is a violation of one of the ANOVA assumptions.

## 11. Conclusion

There is a significant difference in the total number of childcare spaces available based on the type of operating auspice (commercial, non-profit, public). Subsidies have a significant impact on the availability of childcare spaces, and their effect varies by the type of agency. The interaction between the type of operating auspice and the presence of government subsidies significantly influences the availability of childcare spaces.

Despite some deviations in the assumptions of normality and homogeneity of variances, which may influence the reliability of the ANOVA results, the findings provide insightful implications for policymakers. In conclusion, both the type of operating agency and government subsidies play important roles in the availability of childcare spaces in Toronto, with their interaction revealing nuanced impacts that could inform future policy and action.