Quantitative Analysis of the Availability of Licensed Child Care Centres in Toronto

1. Introduction

Finding child care in Ontario, whether licensed or not, challenges many due to steep costs and a scarcity of spots for children. In response, the provincial government has committed to creating an additional 100,000 childcare spaces over a decade, starting in 2016 and concluding in 2026. This report delves into the dataset of '*INF2178_A2_data*', which encompasses a list of child care centers in Toronto with detailed information about the child care center's name, location, auspice, and availability of spaces.

This analysis will discuss two research questions, serving as a guide to find out patterns of availability of child care centers:

1. **Research Question 1:** How does auspice affect the total availability of childcare centers, and is there a significant difference among the availability of non-profit, commercial, and public childcare centers?

2. **Research Question 2:** Does subsidy or not combined with auspice affect the total availability of childcare centers, and is there a significant difference among the availability of non-profit, commercial, and public childcare centers considering the subsidy status?

By looking into these questions, this analysis report aims to gain insights and serves as a guide to aid in the expansion of the childcare space plan of the provincial government.

2. Data Cleaning and Data Wrangling

The raw dataset contains a total of 17 columns and 1,063 entities(rows). After an initial examination of the dataset, it is deemed that the dataset does not require data cleaning. However, with the research questions, only the following columns will be addressed in this analysis:

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

Based on the examination, none of the columns but BLDGMAME has 348 NAN values. However, due to the complementary nature of this column, the NAN values do not affect the analysis of this report.

3. One Way ANOVA

Data Visualization

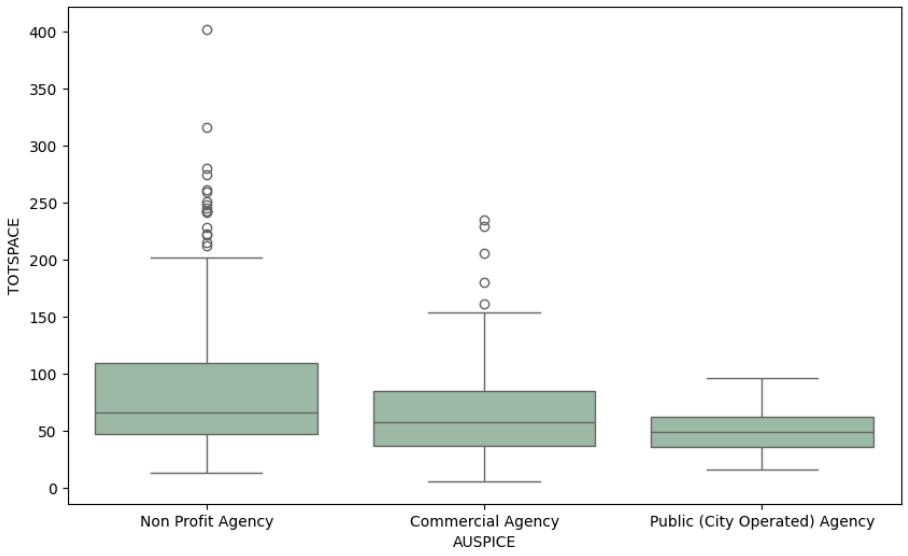

Figure 1: Boxplot for Aupice

In order to understand how auspice affects the total availability of childcare centers, we first need to visualize the space availability distribution across Non-Profit, Commercial, and Public Agencies.

As shown in *Figure 1*. The Non-Profit Agency boxplot exhibits the widest range of Total Space Available, with several outliers indicating some non-profits have significantly more spaces than others and a slightly right-skewed distribution around a median of around 100 spaces. In contrast, the Commercial Agency boxplot shows a more compact interquartile range with less variability, a slightly lower median, fewer outliers, and a more symmetrical data distribution. The Public (City Operated) Agency displays the smallest interquartile range, indicating very little variability in spaces among public agencies, with the lowest median value among the three auspices and no outliers, suggesting public agencies tend to have a more consistent but typically lower number of childcare spaces compared to non-profit and commercial agencies.

ANOVA

In order to examine the difference in space availability among auspices, an ANOVA test is performed. The result is as follows:

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

Table 1: ANOVA Result

The small p-value of 5.057716e-10 which is smaller than the common threshold of 0.05, suggests that there is a statistically significant difference in the total space available across the different levels of the auspice types. The analysis results indicate that the null hypothesis of all groups having the same population mean can be rejected. This means that operating sponsorship of the childcare centers(whether commercial, non-profit, or public) has a significant impact on the total available space. Therefore, a post-hoc test is needed to determine which specific differences are significant.

Post Hoc Test

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

Table 2: One-Way ANOVA Post Hoc Test Result

The post hoc test result shown in *Table 2* indicates significant differences in the average total spaces when comparing Non-Profit Agencies to both Commercial and Public (City Operated) Agencies. There is no statistically significant difference between Commercial and Public (city-operated) Agencies. **The Non-Profit Agencies tend to have more total childcare spaces compared to the other two types of agencies.**

ANOVA Fitness Testing

In statistical analysis, checking model diagnostics is an important step after performing ANOVA. It is crucial to evaluate whether the assumptions of the ANOVA model are met to ensure the validity and reliability of the analysis results. The following assumptions will be examined:

- **Assumption 1:** The data in each group should be approximately normally distributed(using the Shapiro-Wilk Test).
- **Assumption 2:** The variances of the data across the different groups are homogenous(If not normally distributed, using Levene's Test).

Assumption 1

w	0.901775598526001
p value	1.4964898448030214e-25

Table 3: Shapiro-Wilk Test Result

From *Table 3*, we can learn that the test statistic w is 0.901775598526001 and the p-value is 1.4964898448030214e-25. The small p-value suggests that the null hypothesis of the Shapiro-Wilk test, which assumes that the residuals are normally distributed should be rejected. In other words, the data is not normally distributed. Violation of the normality assumption can affect the validity of the statistical inferences. However, the Shapiro-Wilk test is sensitive to even minor deviations from normality, especially with large sample sizes.

Assumption 2

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

Table 4: Levene's Test Result

From *Table 4*, we can learn that the p-value of 0.0001 is smaller than the typical significance level of 0.05, which suggests that the null hypothesis of equal variances across the different levels of "AUSPICE" should be rejected. In other words, the output indicates that there is statistically significant evidence that the variances of the "TOTSPACE" variable are not equal across the different levels of the "AUSPICE" variable. This violates the assumption of homogeneity of variances required for ANOVA. Violation of the homogeneity of variances assumption can affect the validity of the ANOVA results and lead to an increased risk of Type I errors (false positives) or reduced statistical power.

Assumption Visualization

Figure 2 has verified the data is not normally distributed.

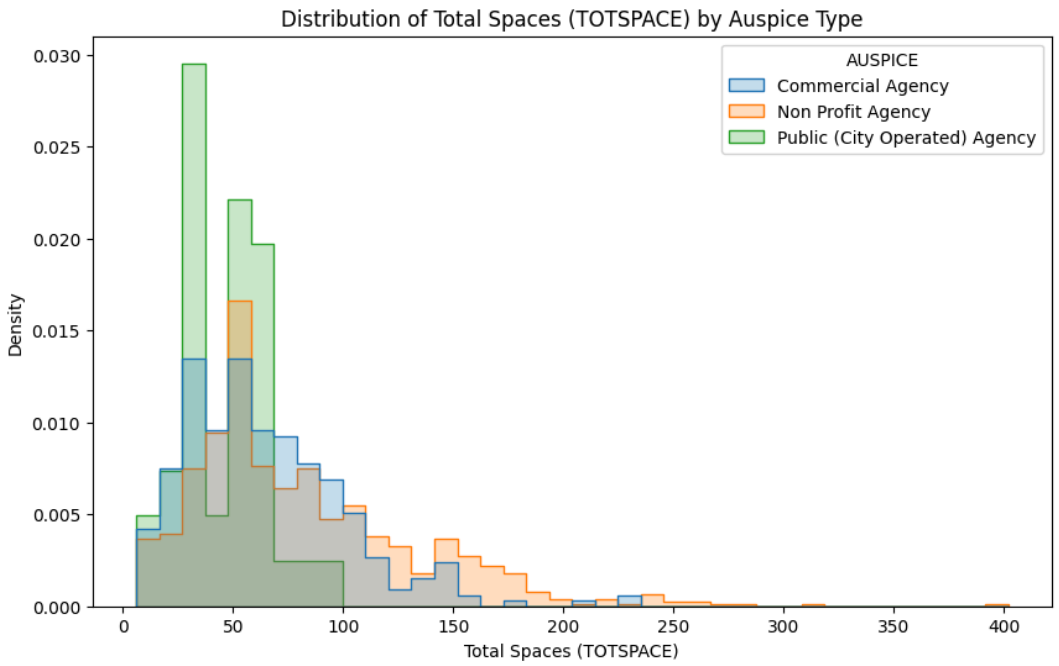

Figure 2: Histogram for Distribution of Total Space by Auspice Type

4. Two Way ANOVA

Data Visualization

The analysis of the second research question starts with visualizing the distribution of the interaction of Auspice and Subsidy on total spaces.

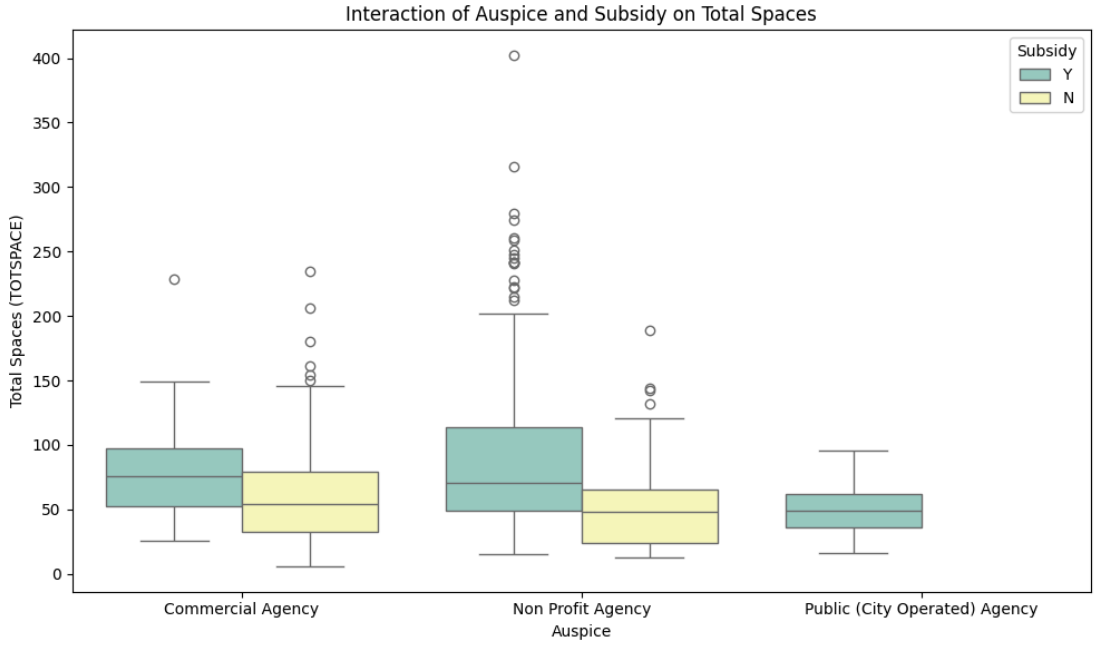

Figure 3: Boxplot for Interaction of Auspice and Subsidy on Total Spaces

As shown in *Figure 3*. Commercial Agency centers tend to have the lowest total spaces overall. Non-Profit Agency centers generally have the highest total spaces. Public (city-operated) Agency centers fall in between the other two auspice types.

For both Commercial Agency and Non-Profit Agency centers, the subsidized centers tend to have higher total spaces compared to non-subsidized centers. For Public (city-operated) Agency centers, there doesn't seem to be a substantial difference in total spaces between subsidized and non-subsidized centers.

Overall, this boxplot suggests that the auspice type (Commercial, Non-Profit, or Public) and whether the center receives a subsidy or not both influence the total available spaces in childcare centers, with Non-Profit Agency centers generally having larger spaces on average. However, there is also considerable variability within each group, and the presence of potential outliers should be further investigated.

Two Way ANOVA

In order to examine how subsidized or not affects the auspice's total space availability, a Two-way ANOVA test is performed. The result is as follows:

	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	8.567996e+03	4283.998054	2.057586	1.282730e-01
C(subsidy)	1.0	8.352744e+04	83527.441787	40.117876	3.529094e-10
C(AUSPICE): C(subsidy)	2.0	5.603445e+04	28017.227037	13.456555	1.694282e-06
Residual	1058.0	2.332065e+06	2082.050461	NaN	NaN

Table 5: Two-Way ANOVA Result

Based on the p-values, the most significant factor affecting the response variable is the subsidized or not, indicating that the subsidized and non-subsidized groups have significantly different means. However, the interpretation of the main effects should be done cautiously due to the significant interaction effect between auspice and subsidy. The significant interaction effect suggests that the impact of subsidization on the response variable depends on the level of auspice types or conversely, the effect of auspice types on the response variable differs between subsidized and non-subsidized groups. The interaction plot shown in Figure 4 provides a visualization of the subsidy status.

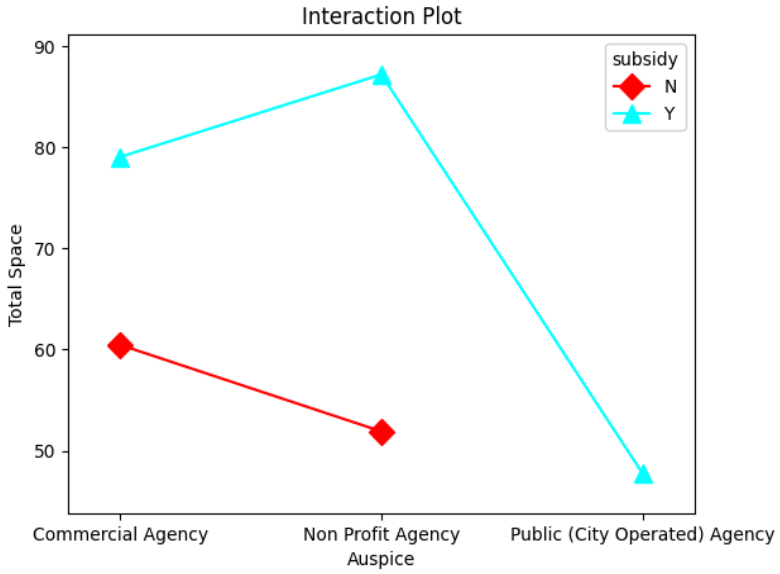

Figure 4: Interaction Plot

To fully understand the nature of the interaction and the specific differences between groups post-hoc tests need to be performed.

Post Hoc

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

Table 6: Two-Way ANOVA Post Hoc Test Result

Table 6 highlights that subsidized Non-Profit Agencies tend to have significantly higher mean Total Space Available compared to most other groups, except for subsidized Commercial Agencies, where the difference is not statistically significant. Additionally, the subsidization status appears to have a significant impact on the mean Total Space Available for Non-Profit Agencies and Commercial Agencies, but not for Public (City Operated) Agencies.

ANOVA Fitness Testing
To ensure the validity and reliability of the analysis results, the following assumptions will be examined:

- **Assumption 1:** The data in each group should be approximately normally distributed(using the Shapiro-Wilk Test).

w	0.9018619656562805
p-value	1.5311055543621852e-25

Table 7: Shapiro-Wilk Test Result

The small p-value(smaller than 0.05) suggests that the null hypothesis cannot be rejected which means the data is not normally distributed. The conclusion drawn from this ANOVA test should be considered with caution.

5. Conclusion

In conclusion, the findings indicate that auspice and subsidy status significantly influence the total number of available childcare spaces. Non-profit agencies, particularly those receiving subsidies, tend to have higher capacity compared to commercial and public agencies. However, the analysis also revealed violations of key assumptions like normality and homogeneity of variances, which could impact the validity of the results. To better support the provincial government's goal of expanding childcare spaces, it is recommended to prioritize funding and incentives for non-profit organizations to establish new centers or expand existing ones, while also exploring strategies to increase capacity within the public and commercial sectors. Moreover, implementing a data-driven approach to identify underserved areas could help target resources more effectively and ensure equitable access to childcare across Toronto.