Exploring Space Usage Situation of Childcare Centers in Toronto

Jiachen Liu-INF2178-A2

1.Introduction

In recent years, the city of Toronto has faced significant challenges in providing adequate child care services to meet the demands of its residents. With a large percentage of families unable to afford child care, the issue of space availability and its accessibility has become increasingly prominent. The provincial government's commitment to creating 100,000 new child care spaces has highlighted the need for an in-depth analysis of the current distribution and utilization of these services.

The purpose of this report is to analyze the operation and capacity of licensed child care centers in Toronto, focusing on various demographic aspects, classifications, and locations as of February 2024. Through this analysis, we seek to answer two key research questions:

a. Is there a significant difference in the childcare center space usage among different space types that target different age groups of children?
b. How does the management structure of the childcare centers affect different space types' space usage?

2.Data Cleaning

The initial phase of the analysis involved careful data cleaning to ensure the reliability of the results. The dataset, titled "INF2178_A2_data.xlsx," required several cleaning procedures:

a. The "BLDGNAME" column was found to have missing values, which were addressed appropriately to prevent any bias in the analysis.
b. For ease of reference during coding, the "_id" column was renamed.
c. Data types were standardized by converting the "_id," "LOC_ID," and "ward" columns to string format to avoid any potential issues during the manipulation and analysis stages.

This thorough cleaning process was crucial for the accuracy of the exploratory data analysis and the subsequent ANOVA tests.

3.EDA

In the exploratory data analysis (EDA) phase of this report, I delve into the dataset to uncover patterns and distributions that may influence policy and management decisions. The following tasks were undertaken:

a. Examined the distribution of child care centers by auspice, visualized through a bar chart, to understand the proportional representation of different operating agencies.
b. Explored the distribution of child care centers' capacities by auspice using a box plot, which provided insight into the variability and range of capacities across agency types.
c. Analyzed the distribution of child care centers' capacities across different auspices and subsidy statuses, also presented via a box plot, to investigate any disparities in capacity related to financial support mechanisms.

3.1.Analysis of Figure 3.1: Distribution of Childcare Centers by Auspice

The bar chart presents a clear disparity in the number of child care centers across different auspices. Non-Profit Agencies dominate the landscape with the highest number of centers, followed by Commercial Agencies. Public (City Operated) Agencies represent a significantly smaller portion. This distribution suggests that non-profit organizations are the primary providers of child care services in Toronto. The commercial sector contributes to a lesser extent, and city-operated centers are relatively scarce.
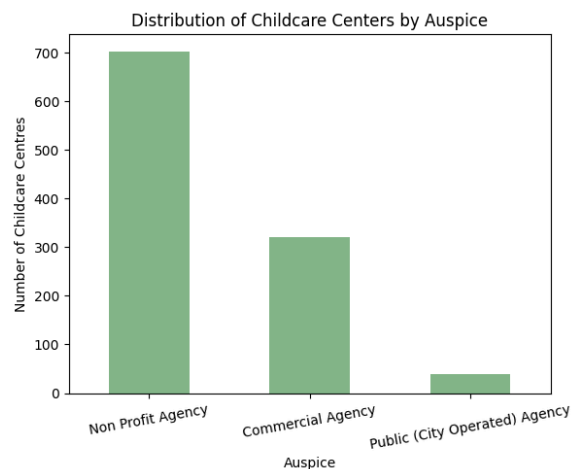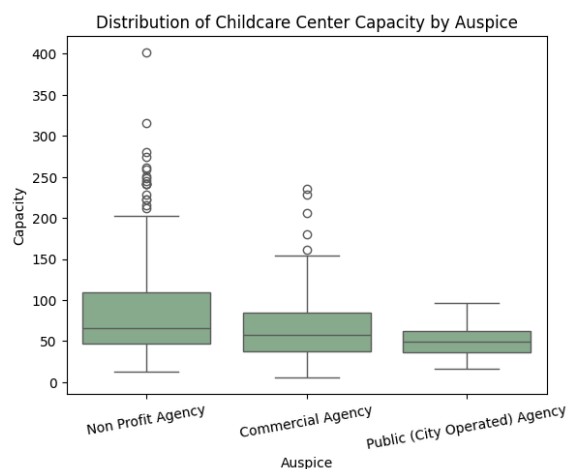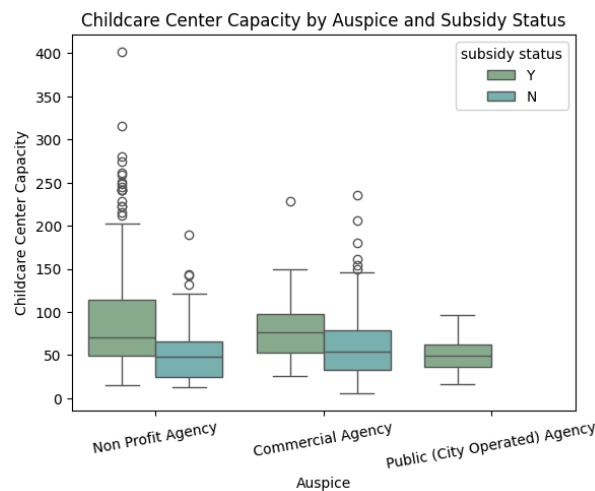

Figure 3.1Distribution of Childcare Centers by Auspice   Figure3.2 Distribution of Childcare Center Capacity by Auspice

3.2Analysis of Figure 3.2: Distribution of Childcare Center Capacity by Auspice

The box plot illustrates the distribution of capacities within child care centers categorized by auspice. This plot suggests that non-profit child care centers not only are more numerous but also tend to offer more spaces, potentially providing a broader service coverage. Commercial centers, while less variable, offer a moderate number of spaces. Public agencies, on the other hand, represent the smallest capacity segment in the child care ecosystem, raising questions about the accessibility and availability of public child care options.

3.3 Analysis of Figure 3.3: Childcare Center Capacity by Auspice and Subsidy Status

This analysis sheds light on the impact of subsidy status on childcare center capacities. Subsidized centers, particularly within the non-profit sector, are generally larger and may therefore play a critical role in providing care to a larger segment of the population. The presence of subsidies might indicate a strategic approach to reach more families, especially those who may not afford care otherwise. Public centers, while consistently smaller, also benefit from offering subsidies by slightly increasing their capacities. Overall, this suggests that financial support mechanisms are pivotal in shaping the capacity and reach of childcare services across different auspices in Toronto.

These analyses together paint a comprehensive picture of the distribution and capacity of childcare centers in Toronto, highlighting significant differences in how services are provided across different types of organizations and financial arrangements.

Figure 3.3 Childcare Center Capacity by Auspice and Subsidy Status

4.ONE-WAY ANOVA

In the one-way ANOVA conducted, the aim was to discern if there are statistically significant differences in the means of childcare center space usage among various space types designed for different age groups of children. This comparison is critical in understanding how child care resources are allocated for age-specific needs.

- Null Hypothesis ($H0$): There are no significant differences in the means of childcare spaces usage across the five types of spaces.
- Alternative Hypothesis ($H1$): At least one of the childcare space usages' means is significantly different from the others.

The initial visual exploration of the data was conducted using a boxplot (refer to , which illustrated the distribution of values for each space type. The presence of outliers in the box plot for "PGSPACE" and "SGSPACE" indicated variability in the capacity numbers across the centers, suggesting potential areas of interest for deeper analysis.
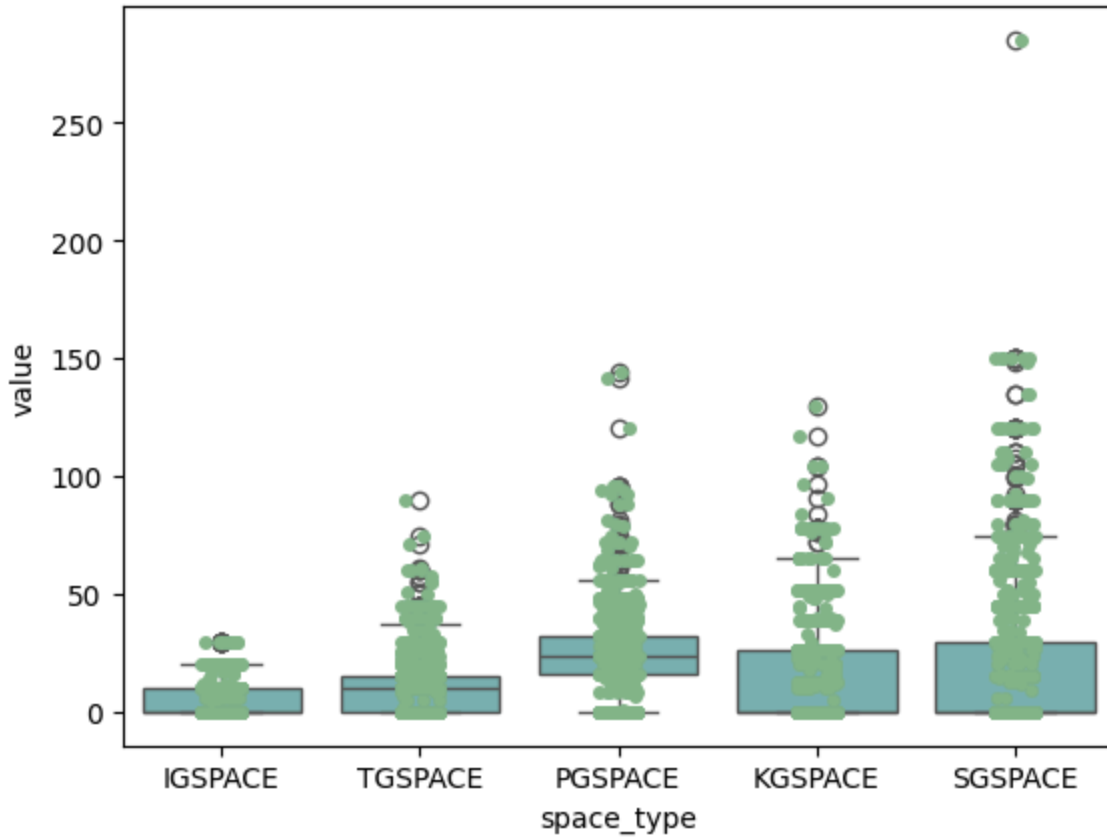
Figure 4.1

An Ordinary Least Squares (OLS) model was fitted to the data, and the resulting ANOVA table (table 4.1) indicated highly significant differences among the group means (p-value < 0.001), leading to the rejection of the null hypothesis. This was further supported by Tukey's Honestly Significant Difference (HSD) post-hoc test (table 4.2), which revealed specific pairs of space types where differences were significant. For instance, "PGSPACE" differed significantly from "IGSPACE", indicating a higher mean capacity for preschoolers compared to infants.

	df	sum_sq	mean_sq	F	PR(>F)
C(space_type)	4.0	2.82e+05	70530.82	188.19	4.52e-151
Residual	5310.0	1.99e+06	374.78	NaN	NaN

Table 4.1 One-way ANOVA summary table

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	IGSPACE	TGSPACE	7.70	5.41	10.00	12.97	0.00
1	IGSPACE	PGSPACE	20.36	18.07	22.65	34.29	0.00
2	IGSPACE	KGSPACE	10.36	8.07	12.65	17.45	0.00
3	IGSPACE	SGSPACE	17.76	15.47	20.06	29.92	0.00
4	TGSPACE	PGSPACE	12.66	10.37	14.95	21.32	0.00
5	TGSPACE	KGSPACE	2.66	0.37	4.95	4.48	0.01
6	TGSPACE	SGSPACE	10.06	7.77	12.35	16.94	0.00
7	PGSPACE	KGSPACE	10.00	7.71	12.29	16.84	0.00
8	PGSPACE	SGSPACE	2.60	0.31	4.89	4.37	0.02
9	KGSPACE	SGSPACE	7.40	5.11	9.69	12.47	0.00

Table 4.2 Turkey's HSD test result based on one-way ANOVA

To meet the assumptions necessary for one-way ANOVA, two tests were conducted:

    a. The Shapiro-Wilk test (table 4.3) for normality of residuals, which indicated a departure from normality (p-value < 0.001). This was also visually supported by the Q-Q plot (figure 4.2) and the histogram of residuals (figure 4.3), which showed deviation from the theoretical quantiles and non-normal distribution, respectively.
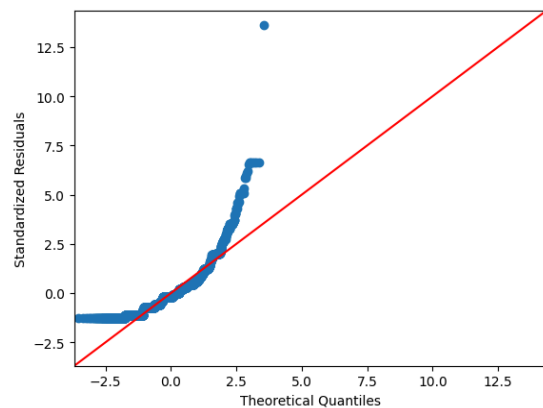

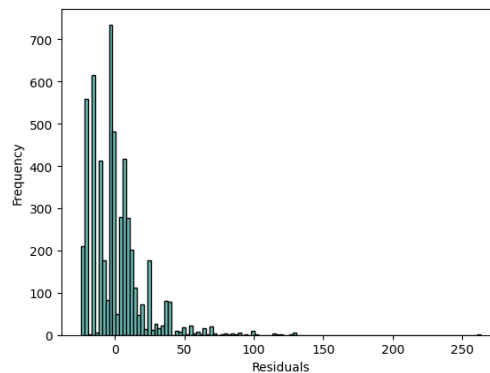Figure 4.2 Q-Q plot of residuals from linear regression

Figure 4.3 Histogram of residuals from linear regression

Test Statistic	P-value
0.843	<0.001

Table 4.3 Shapiro-Wilk test result

b. Tests for homogeneity of variance, including Bartlett's test (table 4.4) (for normally distributed data) and Levene's test (table 4.5) (for non-normally distributed data), both indicated that the variances are not homogeneous (p-value < 0.001).

Parameter	Test Statistics (W)	df	P-value
Value	2461.74	4.0	<0.001

Table 4.4 Bartlett's test result

Parameter	Test Statistics (W)	df	P-value
Value	142.62	4.0	<0.001

Table 4.4 Levene's test result

The violation of the ANOVA assumptions suggests that the significant results should be interpreted with caution. The lack of normality and homogeneity of variances could affect the robustness of the ANOVA results. This may prompt the use of non-parametric methods or transformation of the data as alternative approaches to validate the findings.

Overall, the analysis through one-way ANOVA provided valuable insights into the differences in childcare center capacities across different age groups, highlighting areas where resource allocation could be optimized. The implications of these findings can guide future research and policy-making in the realm of childcare services.

5.TWO-WAY ANOVA

The two-way ANOVA aimed to examine how the type of auspice (Non-Profit Agency, Commercial Agency, Public Agency) interacts with the type of space (IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE) to affect the utilization of child care center spaces.

● Null Hypothesis ($H0$): There would be no significant interaction effect on space utilization between the auspice type and space type.
● Alternative Hypothesis ($H1$): At least one mean space usage is statistically significantly different due to the interaction between these two factors.

The analysis began with visualizing the distribution of space types across the three auspices using a box plot (figure 5.1), providing an immediate impression of variability and central tendencies within and between groups. The interaction plot (figure 5.2) further illustrated the complex relationship between auspice types and space types, suggesting that the utilization of space does indeed vary according to the type of auspice.
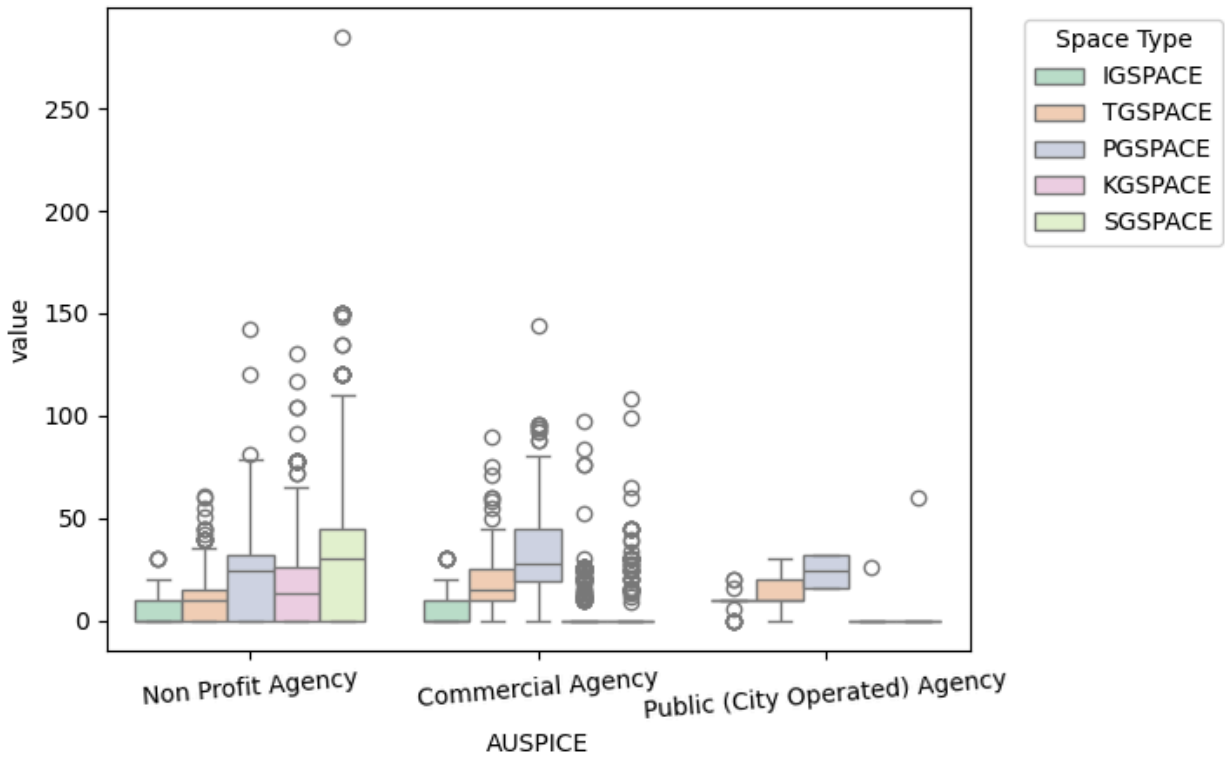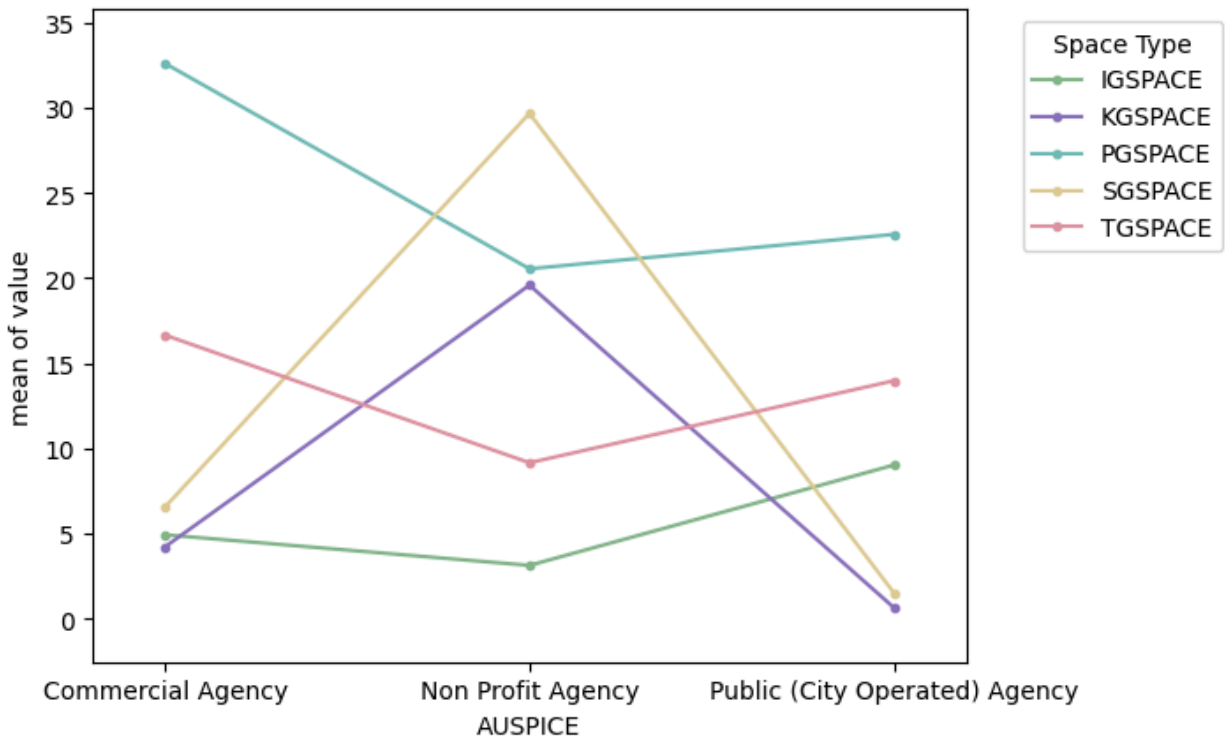
6

Figure 5.1

Figure 5.2

The results of the two-way ANOVA (table 5.1) indicated significant main effects for both the type of auspice and space type (p-value < 0.001 for both), as well as a significant interaction between these two factors (p-value < 0.001), leading to a rejection of the null hypothesis. This implies that the differences in space utilization are not uniform across different auspices, and certain space types are more prevalent or used differently in certain auspice categories.

Source	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	19222.42	9611.21	29.10	2.69e-13
C(space_type)	4.0	282123.30	70530.82	213.58	3.73e-170
C(AUSPICE):C(space_type)	8.0	220645.80	27580.73	83.52	9.97e-131
Residual	5300.0	1750233.00	330.23		

Table 5.1, Two-way ANOVA table

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	IGSPACE	TGSPACE	7.70	5.55	9.85	13.82	0.00
1	IGSPACE	PGSPACE	20.36	18.21	22.51	36.53	0.00
2	IGSPACE	KGSPACE	10.36	8.21	12.51	18.59	0.00
3	IGSPACE	SGSPACE	17.76	15.61	19.92	31.87	0.00
4	TGSPACE	PGSPACE	12.66	10.51	14.81	22.71	0.00
5	TGSPACE	KGSPACE	2.66	0.51	4.81	4.77	0.01
6	TGSPACE	SGSPACE	10.06	7.91	12.21	18.05	0.00
7	PGSPACE	KGSPACE	10.00	7.85	12.15	17.94	0.00
8	PGSPACE	SGSPACE	2.60	0.45	4.75	4.66	0.01
9	KGSPACE	SGSPACE	7.40	5.25	9.55	13.28	0.00

Table 5.2 Tukey HSD post hoc test for differences among five space types

The Tukey HSD post-hoc tests (table 5.2 & table 5.3) were performed to pinpoint where these differences lay. The tests confirmed significant pairwise differences among the space types across the different auspices. Notably, the public (city-operated) agencies showed significant differences compared to non-profit and commercial agencies, indicating a unique pattern of space utilization.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	3.42	2.14	4.71	8.84	0.00
1	Non Profit Agency	Public Agency	6.87	3.73	10.00	7.26	0.00
2	Commercial Agency	Public Agency	3.44	0.21	6.67	3.53	0.03

Table 5.3 Tukey HSD post hoc test for differences among three auspice types

The differences in space utilization have important implications for policy and practice. They highlight the need for tailored strategies that consider the specific conditions and constraints of each type of agency. The significant interaction effect suggests that space allocation policies might need to be customized for different agency types to ensure optimal use of available resources.

However, similar to the one-way ANOVA, the assumptions of normality and homogeneity of variances were tested. The Shapiro-Wilk test, Bartlett's test, and Levene's test would need to be conducted to verify these assumptions for the two-way ANOVA, and potential violations might necessitate adjustments in the analysis or the use of alternative statistical methods.

In summary, the two-way ANOVA provided important insights into the interaction between agency type and space type, which could guide targeted improvements in the distribution and management of child care spaces. Future studies might explore these interactions further, considering additional factors such as geographic location, demographics, and socioeconomic status to paint a more comprehensive picture of the child care landscape in Toronto.

6.Conclusion and Implication

This report has presented a comprehensive analysis of licensed child care centers in Toronto, employing one-way and two-way ANOVA to investigate the distribution and capacity of child care spaces. The findings from the one-way ANOVA suggest significant differences in space utilization across different age groups, indicating that certain age groups may have more or less access to child care spaces. The two-way ANOVA results reveal that these differences are further complicated by the type of agency operating the child care centers, with a significant interaction effect observed between agency type and space type.

For policymakers and stakeholders in the child care sector, the significant disparities in space utilization highlighted by the study point to the need for targeted strategies that address the specific needs of different age groups and consider the operational context of the agencies. The substantial role of non-profit agencies in providing child care spaces, particularly for certain age groups, suggests a potential area of focus for future resource allocation and support.