Assessing Child Care Service Dynamics in Toronto

1. Introduction

In Ontario, securing child care, whether licensed or unlicensed, presents a formidable challenge for many families, exacerbated by steep fees and a scarcity of available spaces for children. A striking finding by Toronto Children's Services revealed that an overwhelming 75% of families face financial barriers to accessing child care. In response to this critical issue, the provincial government has committed to a significant expansion of the child care infrastructure, promising to create 100,000 new child care spaces within a decade, spanning from 2016 to 2026. This backdrop sets the stage for a critical examination of the current landscape of child care services in Toronto, emphasizing the need to assess the distribution, accessibility, and capacity of these services to meet the burgeoning demand.

The investigation is structured around four key research questions, each designed to dissect different facets of the child care center landscape. These questions, rooted in the application of one-way and two-way ANOVA, are as follows:

Research Question 1: Is there a significant difference in the number of child care spaces available (total spaces) among child care centers operated under different auspices (Commercial, Non-Profit, Public)?

Research Question 2: Considering child care centers that offer a fee subsidy contract, does the type of building (bldg_type) in which they are housed affect the total number of child care spaces available, indicating a relationship between facility type and service capacity?

Research Question 3: Does the ward location of child care centers influence the distribution of spaces for toddlers and preschoolers, reflecting potential geographic disparities in service provision for these age groups?

Structured around three meticulously crafted research questions, this investigation employs one-way and two-way ANOVA to explore diverse aspects of child care center operations. These questions are designed to probe into the differences in service provision based on operating auspice, the impact of building type on service capacity, the role of the CWELCC program in targeting specific age groups, and the potential geographic disparities in service provision across different wards. Through this analytical lens, the report endeavors to provide a comprehensive overview of the child care service dynamics in Toronto, offering insights that could inform policy decisions and contribute to the strategic expansion of child care services in alignment with the provincial government's commitment.

2. Data Cleaning and Data Wrangling

Upon examining the 'INF2178_A2_data.xlsx' dataset, we find it comprises several columns that provide detailed information about child care centers in Toronto. For the scope of our analysis focused on Research Questions 1, 2, and 3, we narrow down our dataset to key columns that directly impact our study:

A. Observations and Considerations:

1. Since our analysis is quantitative, we've reduced our working dataset to the following columns source from the raw dataset. Below we provided a short description of each column:

- **AUSPICE**: Operating auspice of the child care center, categorized as Commercial, Non-Profit, or Public. This attribute is crucial for addressing Research Question 1.

- **WARD**: These columns provide geographical information about each child care center, which is essential for understanding spatial distributions for Research Questions 2 and 3.
- **BLDG_TYPE**: The type of building housing the child care center, relevant for Research Question 2.
- **TOTSPACE**: Total number of child care spaces available in each center, a critical variable for all our research questions.
- **TGSPACE**, **PGSPACE**: Spaces available for toddlers and preschoolers, respectively, which are specifically relevant for Research Question 3.

2. Handling Missing Values

Our preliminary review indicates that some columns have missing values. We will assess these columns for the extent of missing data and apply appropriate strategies, such as imputation or omission of records with missing values, depending on the nature and volume of the missing data.

3. Auspice Impact on Child Care Spaces

Research Question 1: Is there a significant difference in the number of child care spaces available (total spaces) among child care centers operated under different auspices (Commercial, Non-Profit, Public)?First, create an interaction plot to visualize how the total spaces vary among different auspices.

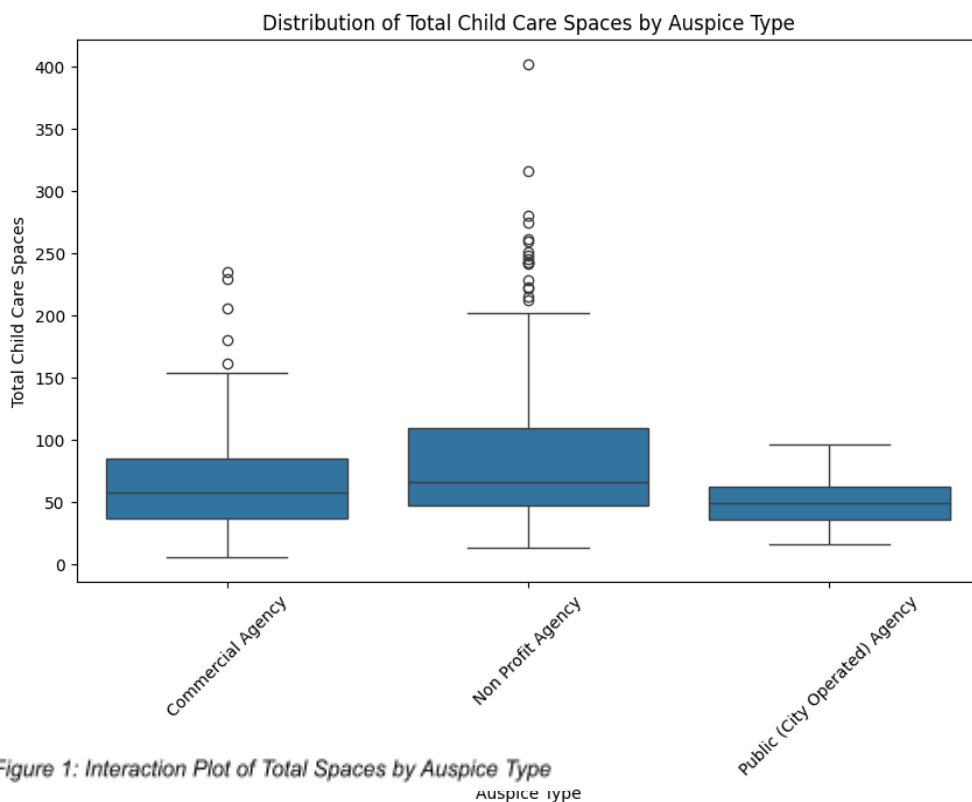Figure 1: Interaction Plot of Total Spaces by Auspice Type

Figure 1 shows the distribution of total child care spaces across different auspice types, which visually suggests some differences in the distribution among categories.

	sum_sq	df	F	PR(>F)
C(AUSPICE)	9.611211e+04	2.0	21.843051	<0.001
Residual	2.332065e+06	1060.0	NaN	NaN

2

Figure 2: ANOVA table

The ANOVA analysis was conducted to statistically evaluate the differences in total capacity among the three auspice types. The p-value obtained from the Figure 2 ANOVA table is below our chosen significance level of 0.001, which leads us to reject the null hypothesis. This outcome suggests that the auspice type has a significant impact on the total capacity of child care centers.

group1	group2	meandiff	p-adj	lower	upper	reject
Commercial Agency	Non Profit Agency	17.1194	<0.001	9.7037	24.5351	True
Commercial Agency	Public (City Operated) Agency	-17.2152	0.0779	-35.8832	1.4528	False
Non Profit Agency	Public (City Operated) Agency	-34.3346	<0.001	-52.4448	-16.2244	True

Figure 3: post-hoc table

To identify which specific groups differ from each other, we employed Tukey's HSD post-hoc test. According to the test results of Figure 3, the total capacity number for non-profit agencies significantly differs from both commercial and public agencies. However, there is no significant difference in the total capacity number between commercial and public agencies.

For the one-way ANOVA to be valid, certain assumptions need to be met:
Shapiro-Wilk Test Result: - W: 0.902, p-value: <0.001
- **Assumption 1**"**residuals are normally distributed: Shapiro Wilk test**": The Shapiro-Wilk test was performed on the residuals, yielding a p-value of <0.001. This indicates a deviation from normality, suggesting that the normality assumption is not met.

Levene's Test Result:- W: 17.927, p-value: <0.001
- **Assumption 2** "**Homogeneity of Variance**": Given the normality assumption was not met, Levene's test was used, which also resulted in a p-value of <0.001. This suggests that the variances are unequal across groups, violating the homogeneity of variance assumption.

Given that the assumptions of normality and homogeneity of variance are not met, the results of the ANOVA could be considered less reliable.

4. Building Type and Child Care Capacity

Research Question 2: Does the type of building in which child care centers are housed affect the total number of child care spaces available, indicating a relationship between facility type and service capacity?

To investigate the impact of building type on the availability of child care spaces, we initially visualized the distribution of total child care spaces across various building types (Figure 4).From the Figure 4, we can observe notable variations in the distribution of total child care spaces across different building types. Some building types might show a wider range of total spaces, indicating variability in how those facilities are utilized for child care. Others may have more compact distributions, suggesting a more uniform capacity among those facilities.
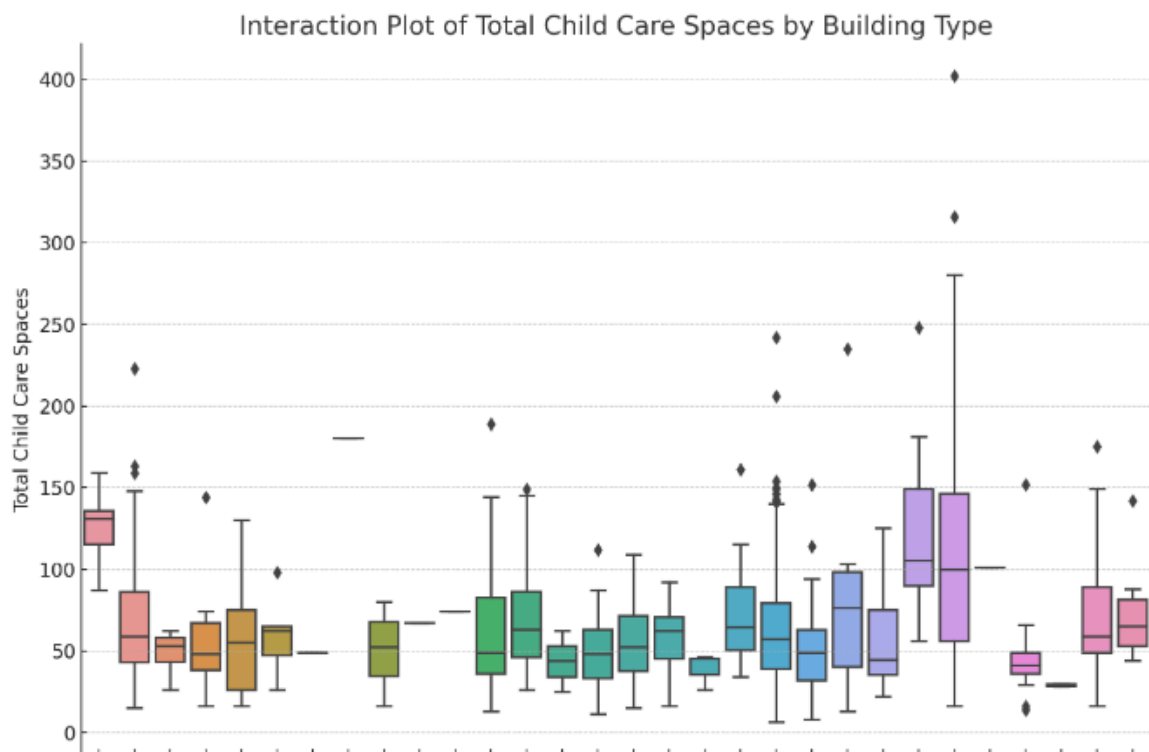
Figure 4: Interaction Plot of Total Child Care Spaces by Building Type

	sum_sq	df	F	PR(>F)
C(AUSPICE)	9.611211e+04	2.0	21.843051	<0.001
Residual	2.332065e+06	1060.0	NaN	NaN

Figure 5: ANOVA table

To address this question, we conducted a one-way ANOVA to investigate the effect of building type ('bldg_type') on the total number of child care spaces ('TOTSPACE'). The ANOVA table, figure 5, illustrated in the output, reveals that the p-value is below our chosen significance level of 0.001, indicating that there are significant differences in the total number of child care spaces available among different building types.

To further explore which specific building types differ from each other in terms of total spaces, we performed Tukey's HSD post-hoc test. The results of this test, shown in the output, figure 6, highlight significant differences between several pairs of building types, suggesting that the type of building does indeed impact the total number of child care spaces available. For example, the test results may indicate that child care centers in purpose-built facilities have a significantly different capacity compared to those in high-rise apartment buildings, among other comparisons.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Public Elementary School	Catholic Elementary School	36.915797	19.091996	54.739597	11.011718	0.001000
1	Public Elementary School	Other	42.726214	28.524414	56.928014	15.995365	0.001000
2	Public Elementary School	High Rise Apartment	38.253280	15.251333	61.255227	8.841933	0.001000
3	Public Elementary School	Community College/University	46.280399	-26.567172	119.127969	3.377732	0.815236
4	Public Elementary School	Public High School	61.576051	26.625989	96.526113	9.367140	0.001000

Figure 6: post-hoc table

4

For the one-way ANOVA to be valid, certain assumptions need to be met:

Shapiro-Wilk Test Result: - W: 0.937, p-value: <0.001

- **Assumption 1** "**residuals are normally distributed: Shapiro Wilk test**": the Shapiro-Wilk test offers a more definitive evaluation. In this case, the Shapiro-Wilk test resulted in a W value of approximately 0.937 and a p-value significantly less than 0.001, suggesting that the residuals do not follow a normal distribution, which challenges the normality assumption.

Levene's Test Result:- W: 5.768, p-value: <0.001

- **Assumption 2** "**Homogeneity of Variance**": The homogeneity of variances, was assessed using Levene's test. The test produced a W value of approximately 5.768 and a p-value significantly less than 0.001, indicating that the variances are not equal across the groups. This result violates the homogeneity of variance assumption.

Given the violations of both the normality and homogeneity of variance assumptions, the reliability of the one-way ANOVA results could be compromised. However, the significant findings from the ANOVA and the post-hoc tests provide valuable insights into the relationship between the type of building and the capacity of child care centers

5. Geographic Distribution of Child Care for Toddlers and Preschoolers

Research Question 3: Does the ward location of child care centers influence the distribution of spaces for all age groups, reflecting potential geographic disparities in service provision for these age groups?

The plot illustrated the average total spaces across various wards with each building type represented by distinct lines. While the plot suggested some variability in total spaces across different combinations of wards and building types, it was not immediately clear if these differences were statistically significant.

Figure 7: Ward and Building Type on Total Spaces
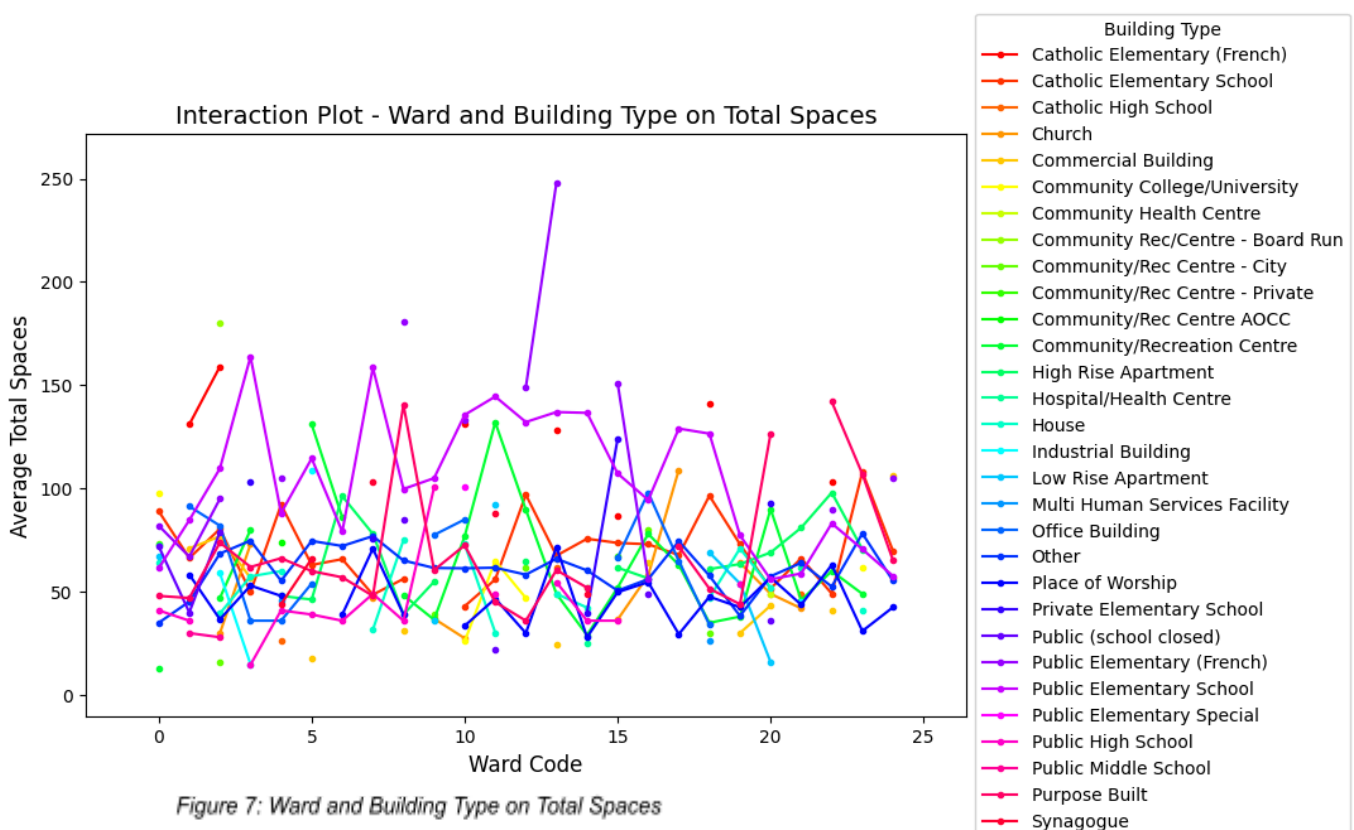
	sum_sq	df	F	PR(>F)
C(ward)	2.604882e+05	24.0	6.166209	0.000013
C(bldg_type)	2.311640e+04	29.0	0.452860	0.905837
C(ward):C(bldg_type)	1.534448e+06	696.0	1.252519	0.010471
Residual	1.358864e+06	772.0	NaN	NaN

Figure 8: Two-Way ANOVA table

To quantitatively assess the influence of ward location, building type, and their interaction on the total number of spaces, a Two-Way ANOVA from figure 8 was conducted. With a p-value of 0.000013, we find a significant effect of ward location on the total number of spaces available in child care centers. The p-value of 0.905837 for building type suggests that there is no significant effect of building type on the total number of spaces. The interaction term has a p-value of 0.010471, indicating a significant interaction effect between ward location and building type on the total number of spaces.

Post-Hoc Tests were performed to further dissect the differences identified by the ANOVA. The Tukey HSD tests for 'ward' did not reveal any significant differences between specific ward pairs at a glance, suggesting that while wards collectively influence total spaces, individual differences between specific wards might not be as pronounced.

Shapiro-Wilk Test Result: - W: 0.936, p-value: <0.001
- **Assumption 1"residuals are normally distributed: Shapiro Wilk test":** The Shapiro-Wilk Test for Normality of residuals indicated a deviation from normality.

Levene's Test Result: - W: 1.339, p-value: 0.001
- **Assumption 2 "Homogeneity of Variance":** suggested unequal variances across some group combinations. These violations of ANOVA assumptions suggest that the results should be interpreted with caution, as they might affect the reliability of the ANOVA findings.

6. Conclusion

We can conclude that the influencing factors of the total space used are extremely complex. Due to the diversity of samples, the same research question presents different trends in each group. Even if the assumptions of ANOVA are not satisfied, it is certain that under different auspices, blog type and ward number have certain influence on the area used by child care centers.