

Student Name: Tony Yang
Student Number: 1003289229

Investigating Childcare Space Availability Among Providers in Toronto

Introduction

Over the past years, the issue of finding suitable childcare services has become challenging in Ontario, particularly in Toronto. As space continues to rise, the availability of licensed and unlicensed childcare spaces remains limited, and high fees further make the situation worse. Toronto Children's Services found that 75% of families can't afford childcare. To help the situation, the government promised to create 100,000 new childcare spaces by 2026. This report presents a thorough exploratory data analysis of Toronto Children's Services, aiming to reveal the underlying trends in childcare across Toronto. Our investigation will focus on three research questions, acting as guiding principles in understanding the trends and patterns of childcare in Toronto to provide insights into childcare data in Toronto and offer a deeper understanding to inform more effective interventions.

1. Research Questions 1: Is there a significant difference in the total space of childcare centers based on auspice type?

2. Research Questions 2: Is there a significant interaction effect between auspice type (AUSPICE) and licensing flag status (cwelcc_flag) on TOTSPACE?

Data Preparation

Upon initial examination, the raw dataset contains 17 columns and 1063 rows. Following my review, I determined that no much data cleaning is necessary for the scope of my analysis. Since my analysis is quantitative, I have reduced our working dataset to the following columns from the raw dataset. Below is a short description of each column:

- AUSPICE: Operating auspice
- IGSPACE: Child care spaces for infants 0-18 months
- TGSPACE: Child care spaces for toddlers 18-30 months
- PGSPACE: Child care spaces for preschoolers 30 months up until they enter grade one
- KGSPACE: Child care spaces for children in full-day kindergarten
- SGSPACE: Child care spaces for children grade one and up
- TOTSPACE: Child care spaces for all age groups
- cwelcc_flag: Indicator of whether the center holds a license (Yes or No)

Missing Value Handling

The only column with null values is "BLDGNAME," which I am not focusing on for our analysis.

Exploratory Data Analysis (EDA)

After data preparation, I delved into a comprehensive Exploratory Data Analysis (EDA) to uncover insights from the dataset. I began by describing our numerical data, as shown in Table 1, which presents a summary of these variables. Additionally, I used box plots and histogram plot to visually represent how these features are distributed in Figure 1 and Figure 2. This detailed analysis provided a clear understanding of the general trends within each feature.

	IGSPACE	TGSPACE	PGSPACE	KGSPACE	SGSPACE	TOTSPACE
count	1063.000000	1063.000000	1063.000000	1063.000000	1063.000000	1063.000000
mean	3.896519	11.600188	24.258702	14.257761	21.661336	75.674506
std	6.091782	12.088152	18.577416	20.493290	30.423503	47.81651
min	0.000000	0.000000	0.000000	0.000000	0.000000	6.000000
25%	0.000000	0.000000	16.000000	0.000000	0.000000	43.000000
50%	0.000000	10.000000	24.000000	0.000000	0.000000	62.000000

75%	10.000000	15.000000	32.000000	26.000000	30.000000	97.000000
max	30.000000	90.000000	144.000000	130.000000	285.000000	402.000000

Table 1 : Quantitative Data Statistics

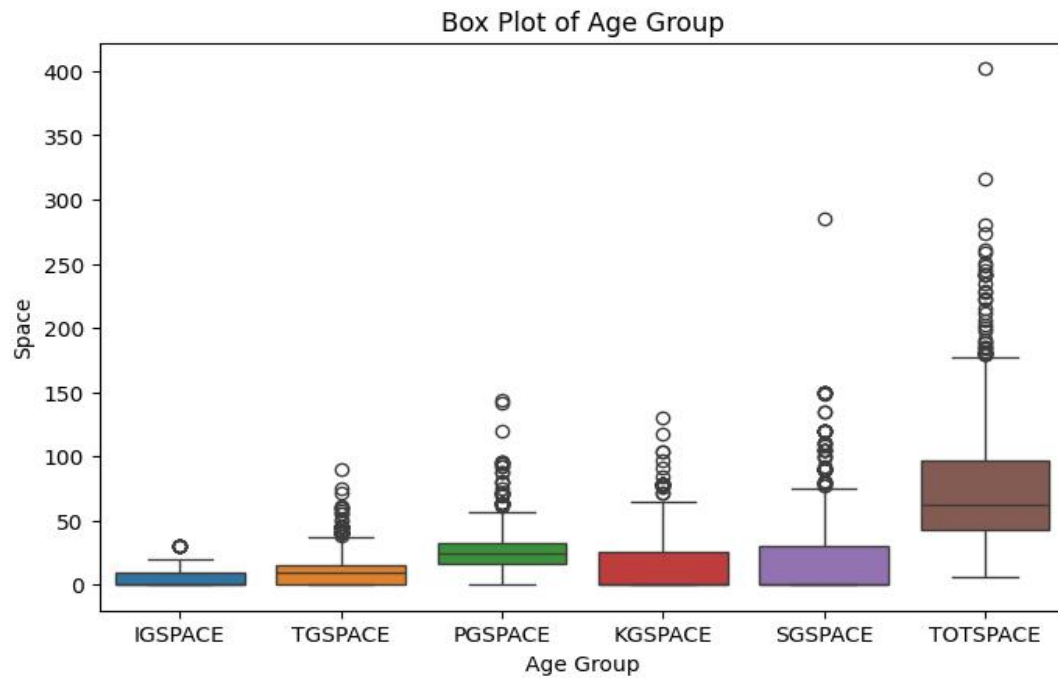


Figure 1: Box Plot Of Different Age Group Space

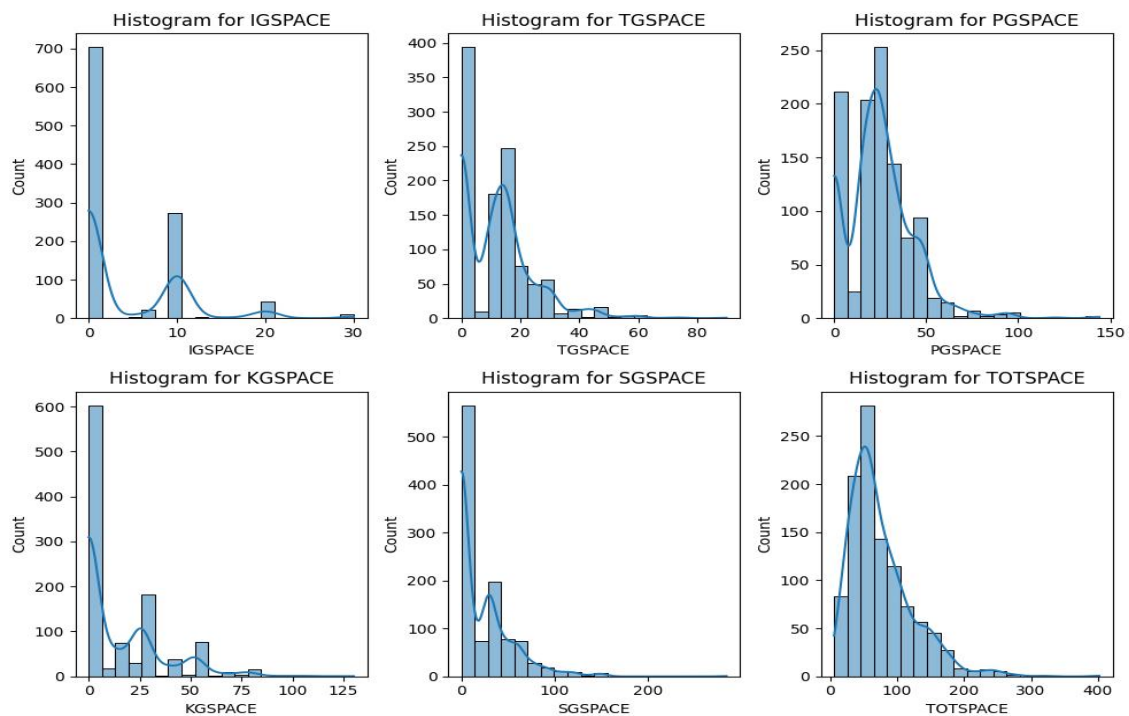


Figure 2: Histogram Of Different Age Group Space

Based on the combination of Table 1, Figure 1 and Figure 2, it indicates clear trends of childcare spaces for different age groups. Infants (IGSPACE) have the fewest spaces, averaging 3.89 spaces, with a maximum of 30 spaces, indicating significant differences among age groups. Toddlers (TGSPACE) have more spaces, averaging 11.6 spaces and reaching as high as 90 spaces. Preschool spaces (PGSPACE) are more common, with an highest average of 24.3 spaces and a maximum of 144, indicating that centers often cater to

this age group. Kindergarten spaces (KGSPACE) are averaging 14.3 spaces with a maximum of 130. School-age spaces (SGSPACE) vary widely, averaging 21.6 spaces and reaching up to 285 spaces. The box plot (Figure 1) provides visual insights of the distribution from Table 1 and it also indicates that our age groups have many outliers. On the other hand, the histogram (Figure 2) illustrates the distribution of all age groups across childcare centers. Each distribution appears to be right-skewed, characterized by a longer tail that extends towards the higher number of spaces.

One-Way ANOVA

Research Questions 1: Is there a significant difference in the total space of childcare centers based on auspice type?"

To address this question, I performed a one-way ANOVA analysis. Before proceeding, I created a bar plot and box plot illustrating the distribution of total space across various auspice types in Figure 3 and Figure 4.

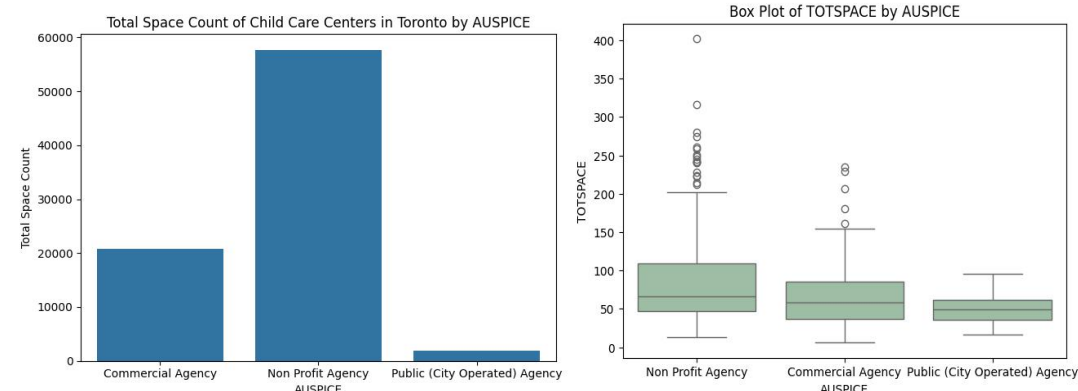


Figure 3: Bar Plot Of Total space vs Auspice Figure 4: Box Plot Of Total Space vs Auspice

Based on figure 3 and 4, it's obvious that the distribution of total space varies significantly across different auspice types. The Non-Profit Agency has the highest space capacity, followed by the Commercial Agency, while the Public Agency has the lowest space capacity. The median of total spaces for Commercial and Non-Profit agencies seems quite similar, but Non-Profit Agencies have a slightly higher median, indicating they might offer more spaces on average. However, Non-Profit Agencies also show more variability in the number of spaces, with some offering a lot more than others. On the other hand, Public Agencies have a much lower median and less variability, meaning they generally offer fewer spaces and Public Agencies don't have any outliers, suggesting they provide a more consistent level of service compared to other categories.

Hypothesis:

H0: There is no significant difference among the means of the groups

H1: At least one group mean is significantly different from the other

I conducted ANOVA analysis by generating the OLS (Table 2) and ANOVA tables (Table 3). These two tables provide valuable information on assessing the significance of relationships between variables and differences between groups.

Table 2: OLS Table

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

Table 3: ANOVA table

	sum_sq	df	F	PR(>F)
--	--------	----	---	--------

C(AUSPICE)	9.611211e+04	2.0	21.843051	5.057716e-10
Residual	2.332065e+06	1060.0	NaN	NaN

The result from both OLS table and ANOVA table indicate there's a notable difference in the total childcare spaces across different auspice. The F-statistic value is 21.84 with a p-value of $5.057716e-10 < 0.05$ which indicates that there is strong evidence to reject the null hypothesis. To delve deeper into the differences, Tukey's HSD test was performed, as illustrated in Table 4.

Table 4: Tukey's HSD test table

group1	group2	Diff	Lower	Upper	q-value	p-value
Non Profit Agency	Commercial Agency	17.119417	9.703599	24.535235	7.662434	0.001000
Non Profit Agency	Public (City Operated) Agency	34.334610	16.224077	52.445142	6.292710	0.001000
Commercial Agency	Public (City Operated) Agency	17.215193	-1.453146	35.883531	3.060857	0.077966

Table 4 clearly indicates that the total space mean of Non-Profit agencies significantly differs from both Commercial and Public agencies, as evidenced by the p-value of 0.001, which is less than the significance level of 0.005. However, the p-value of 0.078 for the comparison between Commercial and Public agencies is greater than 0.05, suggesting that there is no significant difference in the means of total space between these two types of agencies.

To validate the one-way ANOVA, certain assumptions need to be met.

Assumption 1: residuals are normally distributed: Shapiro Wilk test.

Prior to conducting the test, residual plots were generated for enhanced visualization, both figures indicate the potential violations of normality as shown in Figures 5 and Figure 6.

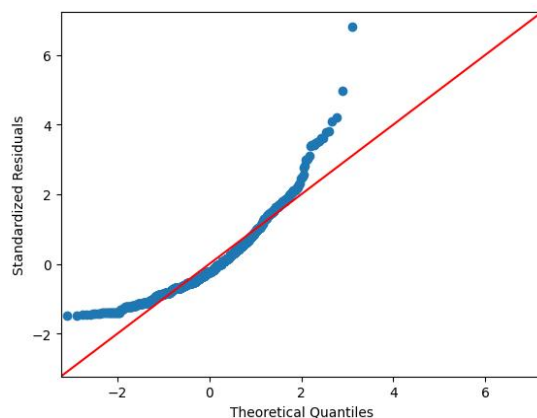


Figure 5: QQ Plot Of Residuals

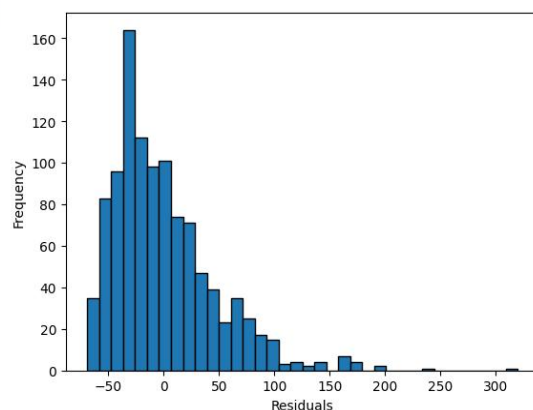


Figure 6: Histogram Of Residuals

Shapiro Wilk test Result - W : 0.901775598526001 , P-VALUE: $1.4964898448030214e-25$

The p-value from Shapiro Wilk test is less than 0.05 which indicates that the Data normality assumption is not met.

Assumption 2: homogeneity of variance - normality is not met : Levene's test

Levene's test result -Test statistics (W) is 17.9271 and p-value is $0.000 < 0.05$ which indicates that the assumption of equal variance is not met.

Based on the results of the assumption tests conducted above, it suggests that the reliability of the one-way ANOVA result may be compromised.

Two-Way ANOVA

Research Questions 2: Is there a significant interaction effect between auspice type (AUSPICE) and licensing flag status (cwelcc_flag) on TOTSPACE?

To address this question, I performed a Two-way ANOVA analysis. Before proceeding, I created box plot illustrating the distribution of total space across various auspice types with cwelcc_flag in Figure 7.

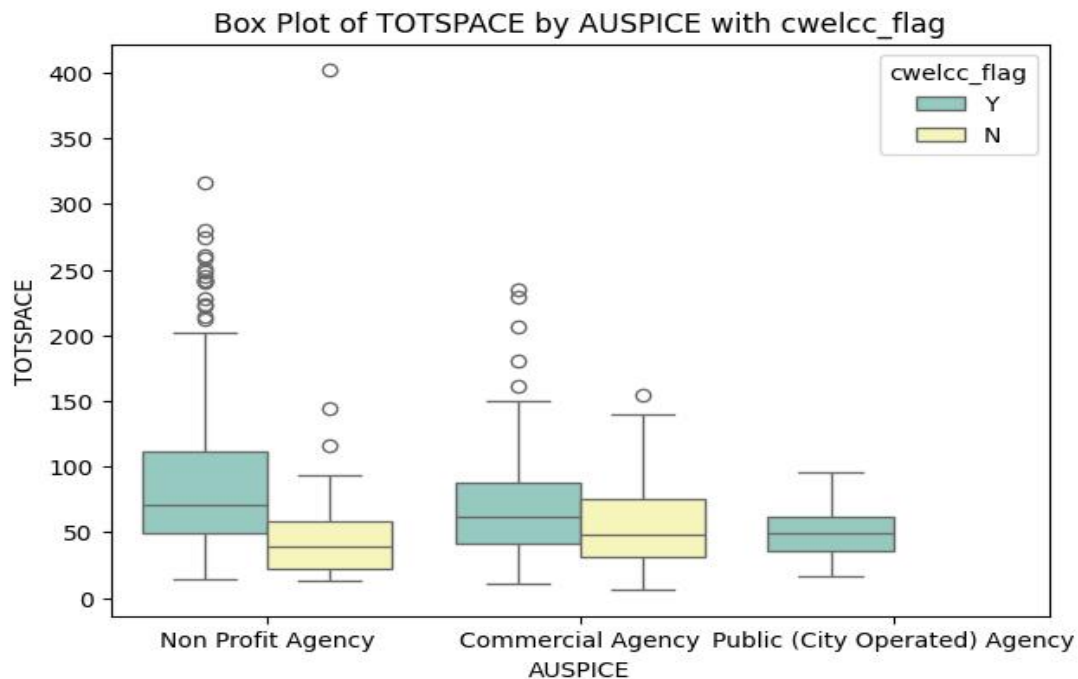


Figure 7. Box Plot of TOTSPACE by AUSPICE with cwelcc_flag

Figure 7 clearly illustrates that the distribution of total space significantly differs across various auspice types with cwelcc_flag. Both Licensed Non-Profit Agencies and commercial agencies exhibit higher space capacity and mean compared to unlicensed agencies. The licensed Non-Profit Agency has the highest space capacity, followed by the licensed Commercial Agency, However, the unlicensed commercial agency displays higher variability and mean than the unlicensed Non profit agency. The Public agency has only one box indicating that all public agencies require a license to operate.

I conducted ANOVA analysis by generating the ANOVA tables (Table 5). The result from ANOVA table indicate There is significant interaction effect between the auspice type (AUSPICE) and licensing flag status (cwelcc_flag) on the total space capacity (TOTSPACE) of childcare centers. The p-values are all less than 0.05 which indicates that there is strong evidence to reject the null hypothesis. The interaction plot also support the result as shown in Figure 8.

Table 5: ANOVA table

	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	1.083348e+05	54167.406210	25.188746	6.095400e-07

C(cwelcc_flag)	1.0	3.768832e+04	37688.323276	17.525698	3.070092e-05
C(AUSPICE):C(cwelcc_flag)	2.0	2.949561e+04	14747.806131	6.857975	1.098398e-03
Residual	1058.0	2.275187e+06	2150.460631	NaN	NaN

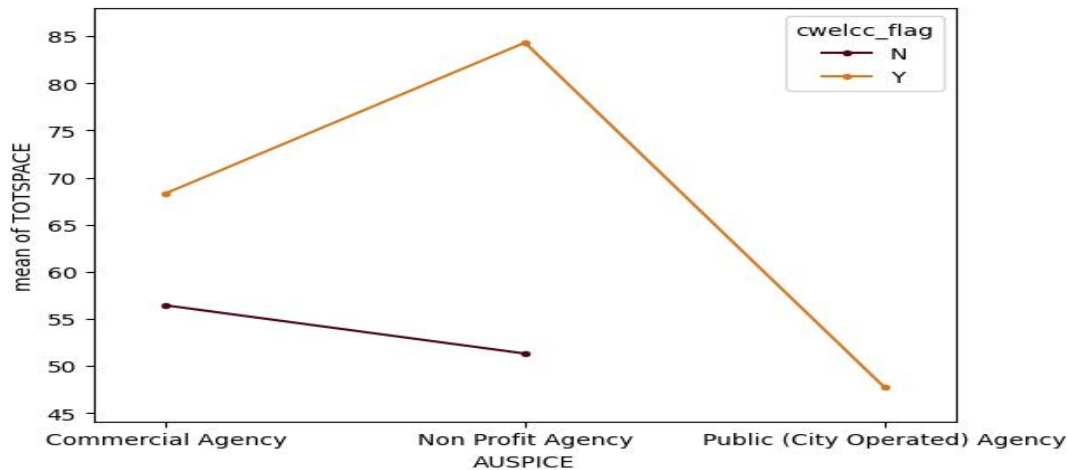


Figure 8 Interaction plot

To delve deeper into the interaction effect, Turkey's HSD test was performed, as illustrated in Table 6. Table 6 clearly indicates there is significant interaction effect between auspice type (AUSPICE) and licensing flag status (cwelcc_flag) on TOTSPACE as evidenced by the all p-value of 0.001, which is less than the significance level of 0.005. With the result of Shapiro Wilk test: $w = 0.8972285389900208$, $P\text{-value} = 4.5834627426744554e-26$ indicates the violation of normality. Moreover, $P\text{-value} = 5.754326496635005e-08$ in Levene's test is less than 0.05. Based on the results of the assumption tests conducted above, it suggests that the reliability of the Two-way ANOVA result may be compromised.

Table 6. Turkey's HSD Table

group1	group2	Diff	Lower	Upper	q-value	p-value
(Non Profit Agency, Y)	(Non Profit Agency, N)	32.990302	12.999232	52.981373	6.662830	0.001
(Non Profit Agency, Y)	(Commercial Agency, Y)	15.993434	5.864514	26.122354	6.375101	0.001
(Non Profit Agency, Y)	(Commercial Agency, N)	27.876118	12.994318	42.757918	7.562847	0.001
(Non Profit Agency, Y)	(Public (City Operated) Agency, Y)	36.540220	14.719584	58.360857	6.761020	0.001
(Non Profit Agency, Y)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900

Conclusion

The One-way and Two-way ANOVA analyses offered valuable insights into the dynamics of childcare space availability in Toronto. While these analyses provided valuable information, we faced challenges as the tests did not meet all the expected assumptions, underscoring the complexity of utilizing ANOVA in this context. Despite the assumptions underlying ANOVA, our study still uncovers important insights into childcare, highlighting the impact of variables such as the type of childcare center, space capacity, and CWELCC participation on space distribution. These findings contribute to our understanding of childcare provision dynamics and inform potential interventions to enhance childcare access in the community.