

Introduction

As Toronto's population increases significantly over the year, many parents find themselves having a hard time securing childcare for their young child. The situation becomes even worse due to the low availability of childcare spaces and the high costs associated with them. A report by Toronto Children's Services suggests that childcare remains unaffordable for almost 75% of families, in which the situation is even worse for those with infants to kindergarteners. There is a need for the government to intervene in this situation to help the parents and to alleviate the pressure.

This study uses the data from the City of Toronto's official website to review the patterns behind the city's childcare situations. The data includes all Licensed Childcare Centres in the City of Toronto with the capacities of each childcare centre by different age groups (from infant to children of grade 1 and above), location, auspice, building types, and other information. The study aims to investigate the childcare issue in the City through the following research questions:

Research Question 1: Is there a significant difference in the total availability of childcare spaces among different auspices?

Research Question 2: How do the type of auspices and the childcare categories affect the availability of spaces in childcare centres, especially for children from infants to kindergarteners?

Data

The dataset includes a total of 17 columns and 1063 entries. Based on the research questions, here are a summary of some important data features that will be utilized in further studies:

Table 1.

Column Name	Descriptions
AUSPICE	Operating auspice. This is a categorical variable with 3 fields: Commercial, Non-Profit or Public
IGSPACE	Childcare spaces for infants 0-18 months. Numerical variable.
TGSPACE	Childcare spaces for toddlers 18-30 months. Numerical variable.
PGSPACE	Childcare spaces for preschoolers 30 months up until they enter grade one. Numerical variable.
KGSPACE	Childcare spaces for children in full-day kindergarten. Numerical variable.
SGSPACE	Childcare spaces for children grade one and up. Numerical variable.
TOTSPACE	Childcare spaces for all age groups. Numerical variable.
subsidy	Centre has a fee subsidy contract. This is a categorical variable with 2 fields: Yes/No

After further investigation, all these fields contain 1063 values, which indicates there are no missing values. The data shows 5 types of childcare space counts for children at different ages: IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE. To double-check, an addition was performed to reveal that the column "TOTSPACE" represents the total space at a given childcare centre, in other words, the sum of IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE at the childcare centre. These space types will later be used in the study to create a field called "space_type" for investigating the effect of childcare category on space count. To better visualize the data, Table 2 shows some basic statistical information of the numerical fields, and Figure 1 is a boxplot to visualize the data. From Table 2 and Figure 1, the spread of the boxes and whiskers shows that there is a wide range of spaces available in each category, with notable outliers for

TOTSPACE. Space for infants (IGSPACE) is the lowest, and some childcare centers have a very high total capacity compare to others.

Table 2. Summary of data

STATS	IGSPACE	TGSPACE	PGSPACE	KGSPACE	SGSPACE	TOTSPACE
count	1063	1063	1063	1063	1063	1063
mean	3.90	11.60	24.26	14.26	21.66	75.67
std	6.09	12.09	18.58	20.49	30.42	47.82
min	0	0	0	0	0	6
25%	0	0	16	0	0	43
50%	0	10	24	0	0	62
75%	10	15	32	26	30	97
max	30	90	144	130	285	402

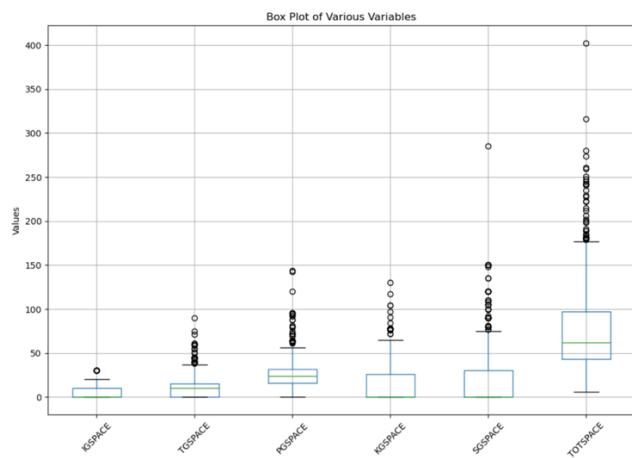


Figure 1

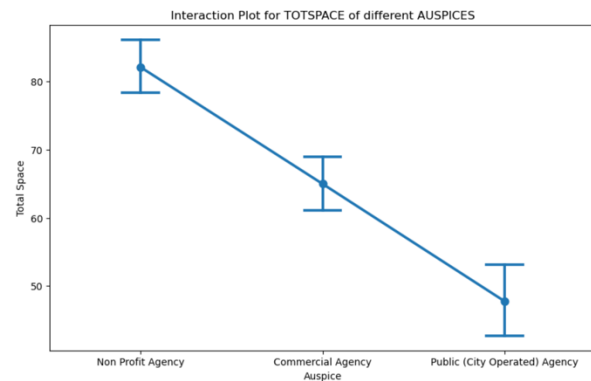


Figure 2

One-way ANOVA – TOTSPACE and AUSPICE interaction

This section will focus on the first research question: the total number of childcare spaces among different operational auspices. In Figure 2, the interaction plot for TOTSPACE of different AUSPICES shows the mean total space available across three types of childcare auspices: Non-Profit, Commercial, and Public (City Operated). The downward trend indicates there might be a difference in the mean total spaces available in childcare centres operated by different auspices. To further investigate whether there is a significant difference in space available among the three auspices, a one-way ANOVA is conducted with the hypotheses below:

Null Hypothesis: There is no significant difference in the total number of childcare spaces available among the different auspices.

Alternative Hypothesis: There is a significant difference in the total number of childcare spaces available among the different auspices.

Table 3. One-way ANOVA Test Result

	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2	96112.11	48056.05715	21.843051	5.05772E-10
Residual	1060.00	2332065.00	2200.06	NaN	NaN

From Table 3, the p-value is extremely small at <0.001 , which is less than the significance level of 0.05. There is strong evidence to reject the null hypothesis, which means there is a statistically significant difference in the total number of childcare spaces available among the different operational auspices. This outcome suggests that the type of auspice has an influence on the availability of childcare spaces. In other words, at least one group differs from the other groups. To better support policy development, a pairwise comparison technique: the HSD test is conducted to analyze the pattern of difference between means.

Table 4. HSD Test result

group1	group2	meandiff	p-adj	lower	upper
Commercial Agency	Non Profit Agency	17.119	0.000	9.704	24.535
Commercial Agency	Public (City Operated) Agency	-17.215	0.078	-35.883	1.453
Non Profit Agency	Public (City Operated) Agency	-34.335	0.000	-52.445	-16.224

For the HSD test, the null hypothesis is the means of the groups in the test are equal. From Table 4, the p-value for Commercial Agency vs. Non-Profit Agency, and Non-Profit Agency vs. Public (City Operated) Agency are <0.001 . This gives us very strong evidence to reject the null hypothesis (significance level = 0.05), which indicates there are significant differences between the total spaces of the groups. The p-value for Commercial Agency vs. Public (City Operated) Agency is 0.078, which is greater than the significance level of 0.05, so we do not have sufficient evidence to reject the null hypothesis that the means are equal. Based on the HSD test, there is a significant difference in the total number of childcare spaces between Commercial Agencies and Non-Profit Agencies, as well as between Non-Profit Agencies and Public (City Operated) Agencies. There is no significant difference between Commercial Agencies and Public (City Operated) Agencies. The city government can consider investigating the reasoning behind the pattern and publish targeted policy to deal with the situation.

Assumption Checking – One way ANOVA

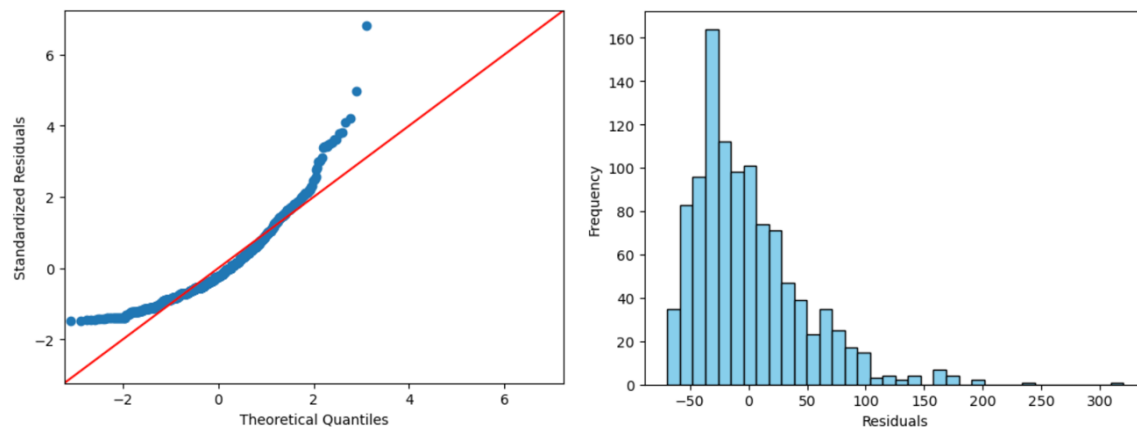


Figure 3

- 1. Normality:** In Figure 3, the QQ plot shows that the data deviates from the 45-degree line and the histogram shows the data is skewed. A Shapiro-Wilk test was conducted with a p-value of <0.001 . This extremely low p-value gives strong evidence to reject the null hypothesis. This suggests that the data might not follow a normal distribution, which violates the assumption of normality.
- 2. Equal Variances:** In Figure 4, the Boxplot of total space by auspice categories shows some outliers. The Levene test was used since the data is not normally distributed. The p-value is <0.001 , which indicates the variances across the groups are not equal and the homogeneity of variance assumption is violated.

3. *Independence*: By observation, each subject has one response. The observations in each group are independent of each other and obtained from a random sample. The sample is independent.

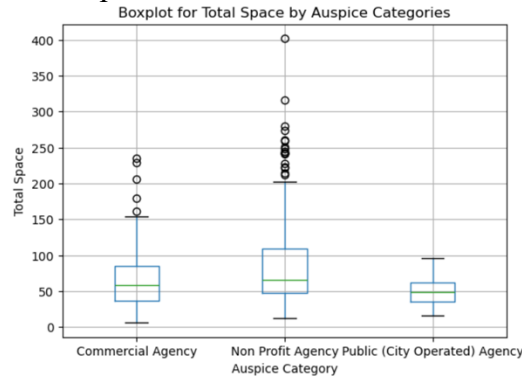


Figure 4 - Boxplot

Based on the above, the assumptions for one-way ANOVA might not hold. The conclusions drawn from the one-way ANOVA should be approached with cautions. Additional tests or other tests can be conducted in further analysis to assess the overall analysis and to ensure that the results are reliable and valid for interpretation.

Two-way ANOVA – Auspice, Space Type and Space Availability

This section will focus on the second research question: the interaction effects between the type of auspice and the childcare category on space availability at childcare centres. By examining the data, a new variable named "space_type" was created to categorize the types of childcare spaces into 'IGSPACE' (infant), 'TGSPACE' (toddler), 'PGSPACE' (preschool), and 'KGSPACE' (kindergarten). The category 'SGSPACE', which represents spaces for children over grade 1, was excluded from this analysis to maintain a focus on the crucial age groups that represent early childhood care. The childcare space availability of these categories was put into a variable called "space". To better investigate the City of Toronto childcare space availability and the factors affecting space availability, a two-way ANOVA was applied to examine the influence of both "AUSPICE" and "space_type" on the variable "space".

The interaction plot in Figure 5 shows the mean of space by AUSPICE and space_type. In the plot, we see 'KGSPACE' (kindergarten) having some intersection with 'IGSPACE' (infant), 'TGSPACE' (toddler). All other lines are not exactly parallel. So, there seems to be significant differences in space availability across the various auspices for different age groups. There might be one factor dependent on the level of the other factor. Hence, we will conduct a two-way ANOVA test with the 3 hypotheses below:

Null Hypothesis 1: There is no difference in space availability across auspices.

Alternative Hypothesis 1: There is a difference in space availability across auspices.

Null Hypothesis 2: There is no difference in space availability across space types.

Alternative Hypothesis 2: There is a difference in space availability across space types.

Null Hypothesis 3: There is no interaction between auspice and space type on space availability.

Alternative Hypothesis 3: There is an interaction between auspice and space type on space availability.

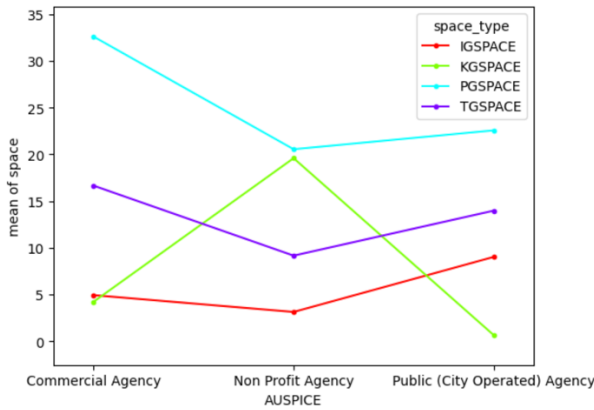


Figure 5 - interaction Plot

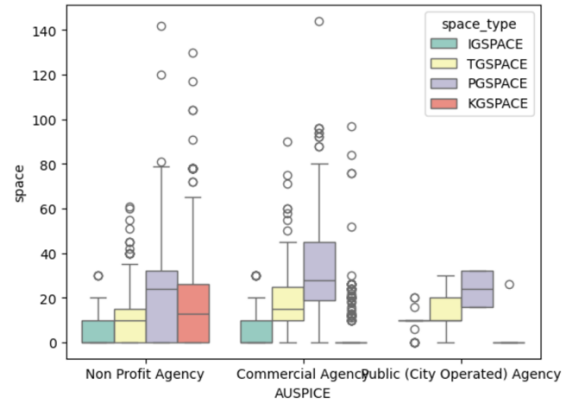


Figure 6 - Boxplot

		df	sum_sq	mean_sq	F	PR(>F)
Hypothesis 1	C(AUSPICE)	2.000	2579.560	1289.780	6.068	0.002
Hypothesis 2	C(space_type)	3.000	225526.026	75175.342	353.695	0.000
Hypothesis 3	C(AUSPICE):C(space_type)	6.000	103362.930	17227.155	81.053	0.000
	Residual	4240.000	901182.437	212.543	NaN	NaN

Table 5. Two-way ANOVA

Table 5 shows the two-way ANOVA test result: the p-value for hypothesis 1 (effect of AUSPICE) is 0.002, hypothesis 2 (effect of space_type) is <0.001, and hypothesis 3 (interaction between AUSPICE and space_type) is <0.001. All p-values are smaller than the significance level of 0.05, which gives us strong evidence to reject the null hypotheses. These results suggest that both the type of auspice and the space type have significant effects on the availability of childcare space for children before grade 1, and there is a significant interaction between auspice and space type.

Based on the two-way ANOVA test, both factors and their interaction significantly influence the space availability in childcare centres. This indicates the need for further investigation into the pattern of difference between means. Similar to one-way ANOVA, an HSD test was conducted with the result as below (Table 6):

Table 6. Screenshots of some HSD test result

group1	group2	meandiff	p-adj	lower	upper	reject
Commercial AgencyIGSPACE	Commercial AgencyKGSPACE	-0.704	1.000	-4.467	3.059	FALSE
Commercial AgencyIGSPACE	Commercial AgencyPGSPACE	27.660	0.000	23.898	31.423	TRUE
Commercial AgencyIGSPACE	Commercial AgencyTGSPACE	11.717	0.000	7.954	15.479	TRUE
Commercial AgencyIGSPACE	Non Profit AgencyIGSPACE	-1.797	0.802	-5.008	1.415	FALSE
Commercial AgencyIGSPACE	Non Profit AgencyKGSPACE	14.656	0.000	11.445	17.867	TRUE
Commercial AgencyIGSPACE	Non Profit AgencyPGSPACE	15.612	0.000	12.400	18.823	TRUE

The HSD test shows the pairwise comparisons between the means of different types of childcare spaces (IGSPACE, PGSPACE, KGSPACE, TGSPACE) across various auspice (Commercial, Non-Profit, Public). Table 6 shows an example HSD test result. The whole table can be found with the codes, only the first several rows are included as reference. Using the 2nd row as an example, the p-value is <0.001, indicating we have strong evidence to reject the null hypothesis that there is no significant difference between the means of the two groups. This means for the auspice Commercial Agencies, there is a statistically difference in the availability of spaces for infants compared to preschool-aged category. In the 4th row, the p-value for Commercial Agency IGSPACE and Non-Profit Agency IGSPACE is 0.802. Hence, we do not have evidence to reject

the null hypothesis, which means there is no significant difference in the mean. These results from the pairwise HSD test provide detailed insights into specific pairs of childcare space types and auspice.

Assumption Checking – Two way ANOVA

1. *Normality*: The QQ plot and histogram in Figure 7 indicate the data is not normally distributed. The Shapiro-Wilk test further confirmed this with a p-value of less than 0.001. The data is not normally distributed, hence, violates the Normality assumption.

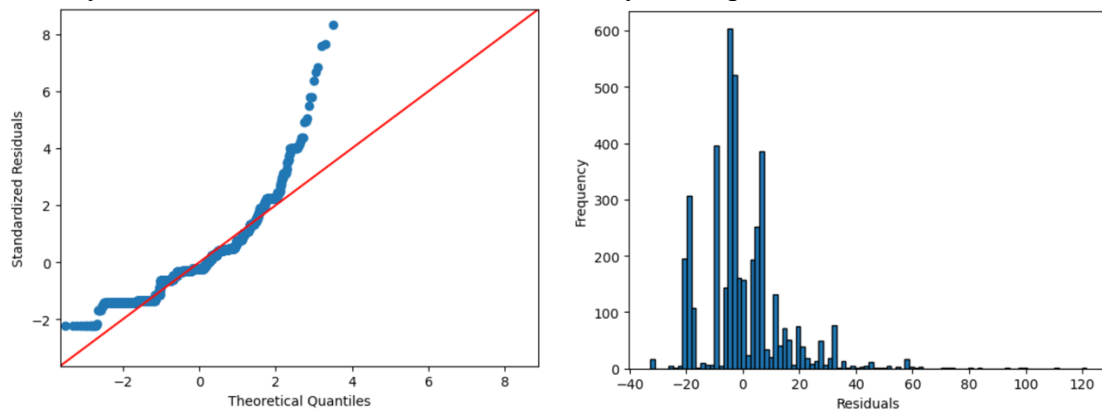


Figure 7

2. *Equal Variances*: In Figure 6, the boxplot suggests variance issues. Again, the Levene test is used due to the non-normal distribution, and the p-value of less than 0.001 confirms the violation of equal variances across groups.

3. *Independence*: The sample is independence.

Once again, the assumption for two-way ANOVA might not hold. Additional caution should be taken when looking at the conclusions from the two-way ANOVA. Further tests can be conducted to analyze the result.

Conclusion

The one-way, two-way ANOVA tests, and the post-hoc tests provide valuable insights into the availability of childcare spaces in the City of Toronto. The one-way ANOVA test revealed significant differences in the total number of childcare spaces among different auspices, with non-profit agencies and public agencies showing different space availabilities compared to commercial agencies. The two-way ANOVA test shows significant interaction effects between auspice and childcare type on the availability of childcare spaces. These findings are crucial for informing policy and operational decisions regarding childcare planning in Toronto. Recognizing which types of agencies are providing more spaces, especially for child at younger age can guide the allocation of resources and support to areas that is most needed. Furthermore, understanding the interaction between auspice types and childcare types can help in planning and expanding childcare services to ensure that all families have equitable access to childcare. Future research can examine the results and address the identified violations of ANOVA assumptions. The study did not have a chance to examine the location and ward information provided in the data due to time constraints. Future studies can investigate the trends of childcare difference by location to better assist policy and management on the availability of childcare spaces.