Lan Li
1005814326

Exploring Toronto Children's Services

1. Introduction

Access to affordable childcare in Ontario is a pressing issue, with high costs and limited spaces leaving 75% of families unable to afford it. This report will provide an exploratory data analysis of the total capacity of Toronto licensed childcare centers based on the dataset called 'INF2178_A2_data.xlsx'.

The paper will be around two research questions:

1. Is there a statistically significant difference in the total capacity among non-profit agencies, commercial agencies, and public agencies in child care services? If differences exist, which types of agencies are significantly distinct in capacity?
2. What is the impact of agency type and subsidy status on the total available child care spaces, and is there an interaction effect between agency type and subsidy status on this capacity?

The questions can help us understand the factors which is affecting the total capacity of child care centers.

2. Data Wrangling

The dataset totally includes **16 columns** and **1063 rows**, but this report only investigates part of the columns. The investigated columns all have 1063 non-null, thus there is no data cleaning needed. The detailed information of target columns is listed below:
- AUSPICE: Operating auspice (Commercial, Non Profit or Public);
- 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;
- subsidy: Centre has a fee subsidy contract (Yes/No).

3. Exploratory Data Analysis (EDA)

By observing the quantitative variables, we found there is a wide range in capacities, especially in the total spaces, indicates that some centres may have sufficient space, but others may have insufficient space to meet the needs of the community. In addition, some centers cannot provide spaces for all age range. This analysis can help adjust policy decisions and resource allocation to better balance the distribution of child care spaces.

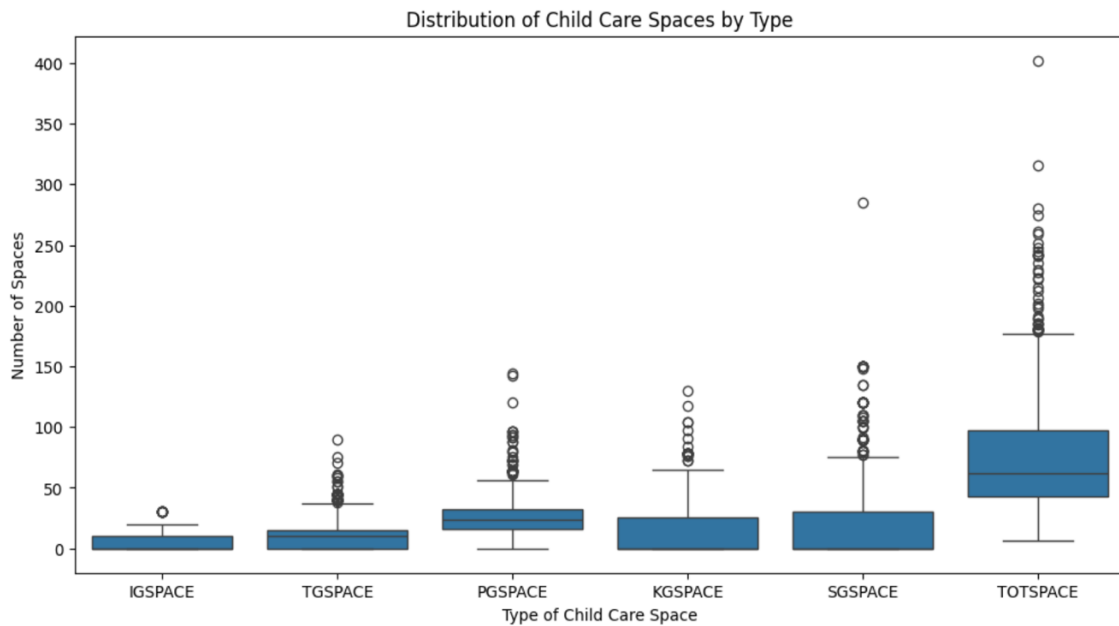| index	IGSPACE	TGSPACE	PGSPACE	KGSPACE	SGSPACE	TOTSPACE
count	1063	1063	1063	1063	1063	1063
mean	3.897	11.600	24.259	14.258	21.661	75.675
std	6.092	12.088	18.577	20.493	30.424	47.817
min	0	0	0	0	0	6
25%	0	0	16	0	0	43
50%	0	10	24	0	0	62
75%	10	15	32	26	30	97
max	30	90	144	130	285	402

Figure 1: Quantitative Variables Description

Figure 2: Boxplot of Child Care Spaces by Type

4. Comparison across Agencies (One-Way ANOVA)

A one-way ANOVA is applied to figure out whether there is a significant difference in total spaces of different agencies. The boxplot visualization complements these statistics by showing a broad spread of capacities within Non Profit and Commercial Agencies, according to the length of their boxes and the number of outliers. This suggests a high diversity in the size of these agencies.

AUSPICE	count	mean	std	min	25%	50%	75%	max
Non Profit Agency	703	82.104	51.909	13	47	66	109	402
Commercial Agency	321	64.984	36.588	6	37	58	85	235
Public Agency	39	47.769	17.860	16	36	49	62	96

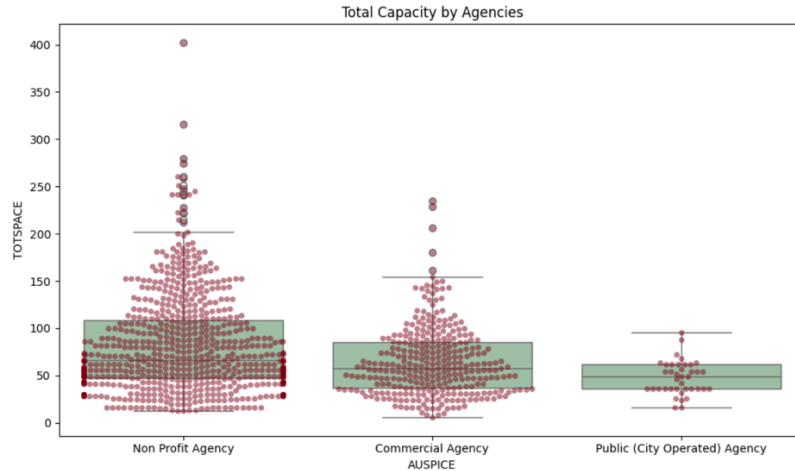Figure 3: Total Spaces Description by Agency

Figure 4: Boxplot of Total Capacity by Agency

Research Question 1: Is there a statistically significant difference in the total capacity among non-profit agencies, commercial agencies, and public agencies in child care services? If differences exist, which types of agencies are significantly distinct in capacity?

One-way ANOVA Test Result
From Figure 5, the p-value is extremely small (<0.001), providing very strong evidence to reject the null hypothesis (0.05 significance level) of equal means across different types of agencies. As shown in Figure 6, there are significant differences in total capacities between Non Profit Agencies and both Commercial and Public Agencies. However, the difference between Commercial and Public Agencies is not statistically significant.

index	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2	96112.11	48056.06	21.843	5.06E-10
Residual	1060	2332065	2200.062	NaN	NaN

Figure 5: ANOVA Test of Total Capacity by Agency

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	16.807	3.994	29.619	4.357	0.006
1	Non Profit Agency	Public Agency	36.178	8.674	63.682	4.369	0.006
2	Commercial Agency	Public Agency	19.371	-10.142	48.885	2.180	0.273

Figure 6: Post Hoc Test One-way ANOVA

Limitation
By observing qq-plot and distribution of residual, they can suggest potential violations of the ANOVA assumptions. And then, the result of Shapiro-Wilk test suggests non-normal distribution of residuals (p-value<0.05), and Levene's test (p-value<0.05) indicates non-homogeneous variances. Due to the violations, the result of one-way ANOVA might be not so accurate. For the future research, Welch's ANOVA or Non-parametric test could be applied to improve the test result.
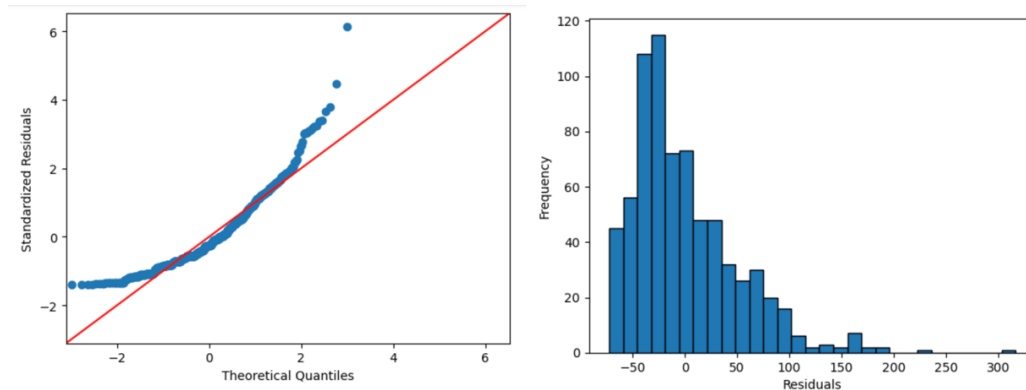
Figure 7: Residual Plots One-way ANOVA

Assumption 1: Shapiro-Wilk test
Test statstics 0.902   p-value 1.496e-25

Assumption 2: Levene's test
Test statstics 9.199   p-value 0.0001

5. Capacity by Agency Type and Subsidy Status (Two-Way ANOVA)

Research Question 2: What is the impact of agency type and subsidy status on the total available child care spaces, and is there an interaction effect between agency type and subsidy status on this capacity?

Two-way ANOVA Test Result
By observing Figure 8, all the public agencies have subsidy, so there is one box plot missing. According to the boxplot, for non-profit and commercial agencies, they may have more child care spaces if the ones have subsidy.

The p-values are telling that there is no statistically significant difference in total child care spaces among the different agency types. However, the presence of a subsidy is associated with a significant difference in total spaces. For the interaction term, the effect of subsidy status on total capacity is significantly different across different agency types. It means the impact of having a subsidy on the total spaces is not the same for all agency types.

Combined the interaction plot with post hoc test result, we can conclude that subsidies make a significant impact on the total spaces for Non-Profit Agencies, but similar comparisons for Public (City Operated) Agencies cannot be made due to the absence of "subsidy N" categories. We can also suggest that subsidies significantly increase capacity in Non-Profit Agencies and that this subsidy effect is not uniform across different types of agencies. Therefore, policymakers could consider design subsidy allocation strategies to maximize the benefit for Non-Profit Agencies, which show a clear positive response in terms of increased capacity.

Limitation
Shapiro-Wilk test suggests non-normal distribution of residuals (p-value<0.05), and Levene's test (p-value<0.05) indicates non-homogeneous variances. Due to the violations of the

assumptions and the absence of "subsidy N" categories, the result of two-way ANOVA might be not so accurate.

Assumption 1: Shapiro-Wilk test
Test statstics 0.902   p-value 1.53e-25

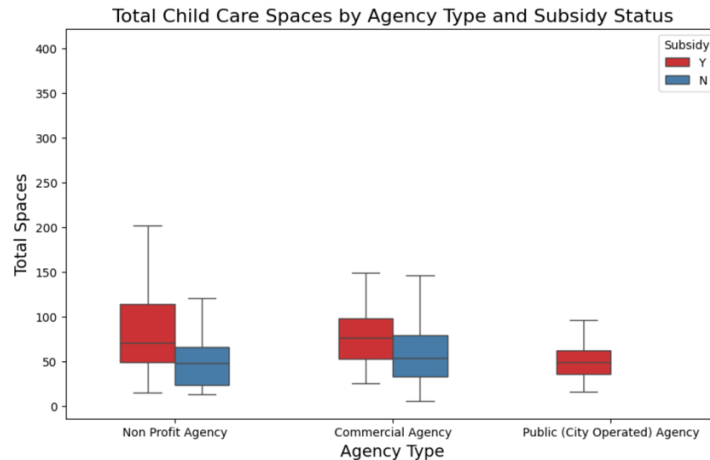Assumption 2: Levene's test
Test statstics 12.985   p-value 2.52e-10

Figure 8: Boxplot of Total Spaces by Agency and Subsidy

index	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2	8567.996	4283.998	2.058	0.128
C(subsidy)	1	83527.442	83527.442	40.118	3.53E-10
C(AUSPICE):C(subsidy)	2	56034.454	28017.227	13.457	1.69E-06
Residual	1058	2202809.388	2082.050	NaN	NaN

Figure 9: ANOVA Test of Total Capacity by Agency and by Subsidy

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Y	N	34.121	24.873	43.369	10.244	0.001

Figure 10: Post Hoc Test on Subsidy Status

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency,Y	Non Profit Agency,N	44.986	25.585	64.387	9.370	0.001
1	Non Profit Agency,Y	Commercial Agency,Y	0.145	-28.466	28.756	0.020	0.900
2	Non Profit Agency,Y	Commercial Agency,N	28.371	11.192	45.549	6.674	0.001
3	Non Profit Agency,Y	Public Agency,Y	40.753	8.308	73.198	5.076	0.005
4	Non Profit Agency,Y	Public Agency,N	0.000	-Infinity	Infinity	0.000	0.900
5	Non Profit Agency,N	Commercial Agency,Y	44.841	11.398	78.284	5.418	0.002
6	Non Profit Agency,N	Commercial Agency,N	16.616	-7.776	41.007	2.753	0.376
7	Non Profit Agency,N	Public Agency,Y	4.233	-32.543	41.010	0.465	0.900
8	Non Profit Agency,N	Public Agency,N	0.000	-Infinity	Infinity	0.000	0.900
9	Commercial Agency,Y	Commercial Agency,N	28.225	-3.979	60.430	3.542	0.124
10	Commercial Agency,Y	Public Agency,Y	40.608	-1.756	82.971	3.874	0.069
11	Commercial Agency,Y	Public Agency,N	0.000	-Infinity	Infinity	0.000	0.900
12	Commercial Agency,N	Public Agency,Y	12.382	-23.272	48.036	1.403	0.900
13	Commercial Agency,N	Public Agency,N	0.000	-Infinity	Infinity	0.000	0.900
14	Public Agency,Y	Public Agency,N	0.000	-Infinity	Infinity	0.000	0.900

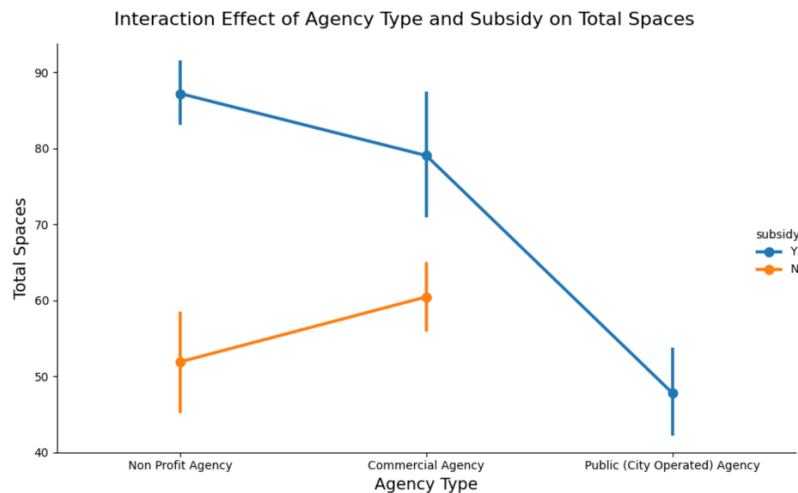Figure 11: Post Hoc Test on Interaction

Figure 12: Interaction Plot

6. Conclusion

The analysis using one-way and two-way ANOVA revealed that subsidies significantly impact the capacity of Non-Profit Agencies' child care spaces. While agency type alone did not lead to an obvious difference, the interaction between agency type and subsidy status was significant, suggesting that subsidies benefit Non-Profit Agencies the most. This indicates that targeted subsidy allocations could effectively increase child care spaces in Non-Profit Agencies.

However, the violation of the assumptions of ANOVA potentially influences the reliability of the results. Policymakers should think over before they accept the findings, and further studies are need for decision making.