Explore Childcare Space Availability in Toronto

Kaiwen Suo (1005764816)

1. Introduction

Children's issues have always been the most hotly discussed topic in society. The number and quality of childcare centers is a significant challenge that must urgently be addressed in Ontario. For the past several years, high costs and low numbers have been the status quo for childcare in Ontario. More than 75% of families cannot afford the high cost of childcare. The causes behind the childcare problem have been the subject of significant research.

This report examines information on childcare centers in Ontario using data from 'INF2178_A2_data.xlsx' (updated to February 2024). The dataset is cleaned, organized, and analyzed to examine the current state of childcare in Ontario and explore future developments.

This report focuses on two issues and uses them as research directions for organizing and analyzing the data:

Research Question 1: Do different children care auspices significantly impact the total number of licensed childcare centres spaces available in Toronto?

Research Question 2: How does the combination of childcare facility auspices (Commercial, Non-Profit, Public) and government subsidy status affect the overall availability of childcare spaces in Toronto?

By examining the above two questions, this report seeks to explore the impact of different variables on the overall number of childcare centers in Toronto.

2. Data cleaning and preparation

The "INF2178_A2_data.xlsx" dataset was thoroughly examined to ensure that it was adequately prepared for our analysis. The dataset encompasses a comprehensive list of childcare facilities, capturing various aspects crucial for assessing space availability.

Observations and Considerations:

Data Reduction: To streamline our quantitative analysis, we refined the dataset to include only the most pertinent columns. Descriptions for each are as follows:

1. **_id:** Serves as the unique row identifier within the Open Data database.
2. **LOC_ID:** Acts as a distinct identifier for each childcare centre.
3. **LOC_NAME:** The name of the childcare centre.
4. **AUSPICE:** The operating auspice of the childcare centre, categorized as Commercial, Non-Profit, or Public.
5. **IGSPACE:** The number of childcare spaces allocated for infants aged 0-18 months.
6. **TGSPACE:** The number of childcare spaces allocated for toddlers aged 18-30 months.
7. **PGSPACE:** The number of childcare spaces allocated for preschoolers aged 30 months until they start grade one.

8. **KGSPACE:** The number of childcare spaces allocated for children enrolled in full-day kindergarten.
9. **SGSPACE:** The number of childcare spaces allocated for children in grade one and up.
10. **TOTSPACE:** The total number of childcare spaces across all age groups.
11. **subsidy:** Indicates whether the centre has a fee subsidy contract, with possible values of Yes or No.

Missing Value examine:

After Examine the dataset, only BLDGNAME has 348 missing values. Since we don't need this variable for the further analyze, I neglect this missing and clean this variable.

3. Exploratory Data Analysis (EDA)

First, we conduct the descriptive analysis of variable TOTSPACE:

- **Count:** 1,063 observations in the dataset.
- **Mean:** Each childcare center has average 75.67 spaces available.
- **Standard Deviation:** 47.82 indicates a wide variation in different centers.
- **Minimum:** The smallest number of total spaces at any center is 6.
- **25th Percentile:** 25% of the centers have 43 or less total spaces.
- **Median (50th Percentile):** The median number of total spaces is 62. It means half of the centers have up to 62 spaces, and the other half have more than 62 spaces.
- **75th Percentile:** 75% of the centers have 97 or less total spaces.
- **Maximum:** The largest number of total spaces is 402.

After we examine the descriptive analysis of **TOTSPACE**, boxplots of each age group **'IGSPACE', 'TGSPACE', 'PGSPACE', 'KGSPACE', 'SGSPACE', 'TOTSPACE'** was made to reveal the distribution of spaces for each age group.

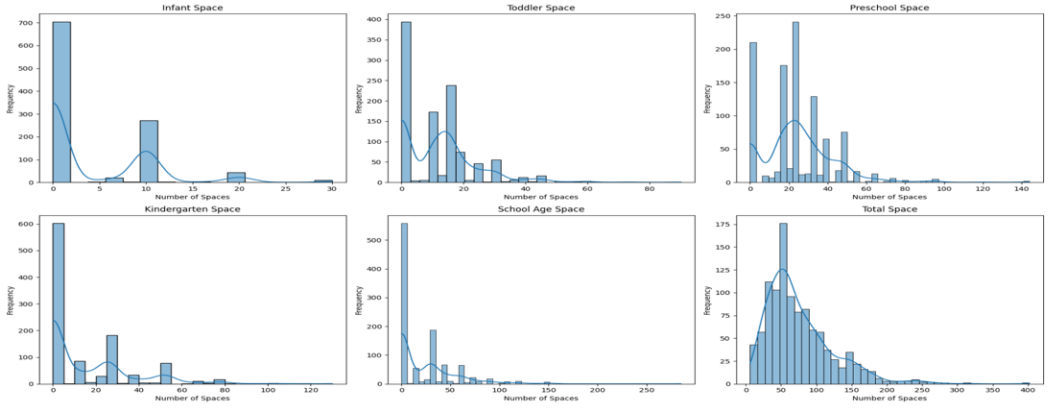


Figure 1: Histogram of different age group space

Examination of space distribution in Ontario's childcare centers indicates a need for more places for infants and toddlers. Many centers have restricted spots, reflecting the increased demand or strict criteria for these age groups. There is a greater abundance of preschool places, which suggests better provision for this group.

Kindergarten spaces exhibit a bimodal distribution, with various offers at different centers. Similarly, school-age spaces also show significant variation across different centers.

The **TOTSPACE** data demonstrates a right-skewed distribution, indicating that although many centers have a large capacity, the majority offer fewer spaces. This highlights the diversified nature of the childcare ecosystem. The distribution of this data indicates possible deficiencies in healthcare for the youngest age groups, highlighting the necessity for focused policy intervention to enhance capacity in areas where it is insufficient.

4. **One-Way ANOVA Analysis:**

Boxplot

First, we conduct the boxplots of **AUSPICES** to see the details of each type of childcare's space.

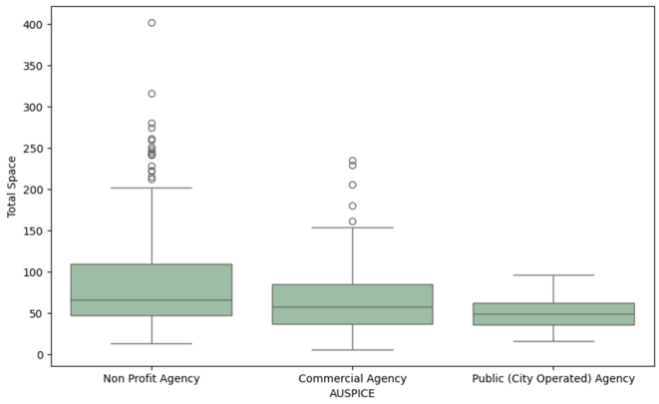


Figure 2: Boxplot of different auspice type

The allocation of overall spaces in childcare centers exhibits substantial variation, which is influenced by the distinction between Non-Profit and Commercial agencies and the availability of subsidies.

Non-profit agencies demonstrate a wide array of capabilities, ranging from modest to large, showcasing their different activities supported by numerous sources of finance. This leads to a significant range in size, encompassing both small and large extremes.

On the other hand, **commercial agencies** typically exhibit a more consistent range of sizes, with a few exceptions that have notably enormous capacity. These organizations have been able to scale up in certain instances successfully.

OLS and One-Way ANOVA table

	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	9.611211e+04	48056.057145	21.843051	5.057716e-10
Residual	1060.0	2.332065e+06	2200.061571	NaN	NaN

Table 1: OLS and One-Way ANOVA Table

After conducting the OLS and One-Way ANOVA test between **AUSPICE** and **TOTSPACE**, we can see the sum of squares of **AUSPICE** is about 9.6 million, which revealed significant variability due to the different categories. By looking at mean square, the value of AUSPICE is much higher than residual which indicate **AUSPICE** has significant impact on total space of childcare center (TOTSPACE). The **F-statistics** is 21.843051 which highly significant. The **P-value** is around 5.06e-10 which is lower than standard thread 0.05.

Based on the statistical result, we can conclude there is a significant difference of Total space across the **AUSPICE** groups.

Tuckey's Honest Significant Difference Test (HSD)

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

Table 2: Tuckey's HSD

Comparing **Non-profit and Commercial agencies** with a difference of 17.119417 indicates that Non-profit agencies have more childcare space than Commercial agencies. The difference between **non-profit and public agencies** is 34.334610, which is significantly higher, indicating that non-profit agencies have much more space than public agencies. By comparing **Commercial Agencies and Public agencies**, the difference is around 17.215193, and the **P-value** is around 0.077. Results show that non-profit agencies have the most space availability among these auspices.

Assumption Check

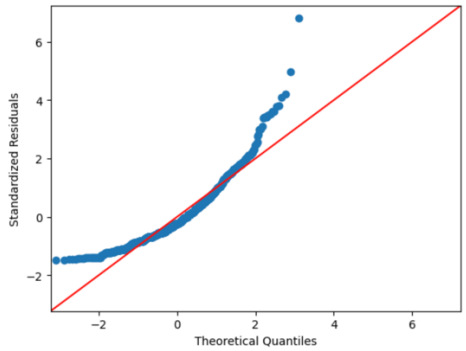


Figure 3: QQ plot of Residual

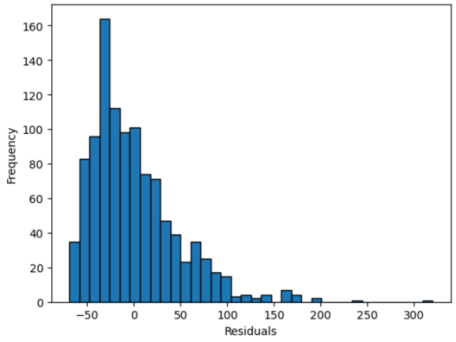


Figure 4: Histogram of Residual

To validate One-Way ANOVA test, we have two assumptions to be check:

Assumption 1: the residuals should be normally distributed.

Assumption 2: variance is homogenous

By using Shapiro Wilk test, we have result for Assumption 1 that **Shapiro Wilk test statistic** is 0.901775598526001 and **P-value** is 1.4964898448030214e-25. Since P-value is smaller than 0.05, assumption 1 not hold.

By using Levene's Test, we have result for Assumption 2 that **Levene's Test statistics (W)** is 17.9271 and **P-value** is smaller than 0.05. Since P-value is smaller than 0.05, assumption 2 not hold.

5. Two-Way ANOVA analysis:

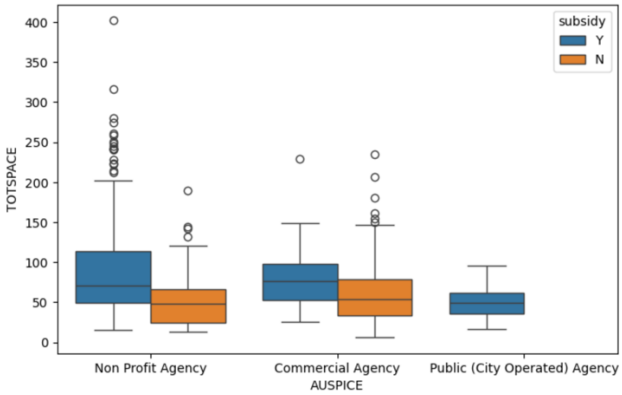


Figure 5: Boxplot of TOTSPACE by AUSPICE and subsidy

In this boxplot, we find that **Subsidized Non-profit agencies** have more extraordinary variability of total space with a high median and larger total space among these groups. There is a big difference in total space between subsidized and non-subsidized in each auspice. There is no subsidized public agency in Toronto.

OLS and Two-Way ANOVA table

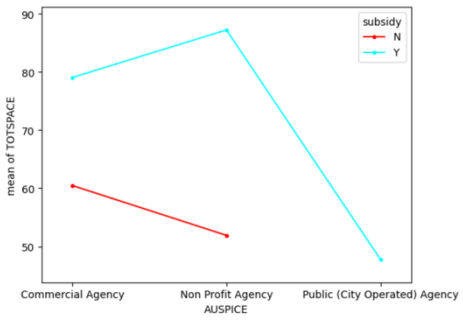
	sum_sq	df	F	PR(>F)
C(AUSPICE)	8.567996e+03	2.0	2.057586	1.282730e-01
C(subsidy)	8.352744e+04	1.0	40.117876	3.529094e-10
C(AUSPICE):C(subsidy)	5.603445e+04	2.0	13.456555	1.694282e-06
Residual	2.202809e+06	1058.0	NaN	NaN

Table 3: OLS and Two-Way ANOVA

AUSPICE alone has 2.057586 as **F-statistics**. With a **P-value** of 0.128273, differences in AUSPICE categories alone do not significantly affect the dependent variable. **Subsidy** alone has higher F-statistics of 40.117876 and a significantly lower P-value with almost 0. It indicates that the subsidy status significantly impacts the total space (TOTSPACE). The F-statistics of the interaction effect of **AUSPICE and subsidy** on total space is 13.456555. The P-value of the interaction effect is lower than 0.05, which indicates the significant impact of the interaction of AUSPICE and subsidy on total space.

Overall, we can conclude there is no significant impact on total space of AUSPICE alone. The effect of subsidy status alone and interaction effect of AUSPICE and subsidy on space is significantly.

Interaction Plot



The subsidy significantly increased the total space for each kind of childcare center. **Non-profit agencies** had the most significant difference between the two subsidy statuses. Childcare centers without subsidies had a lower mean total space. This reveals the importance of subsidies in terms of total childcare space.

Figure 6: Interaction Plot

Post-hoc Test

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	17.119417	9.905213	24.333621	7.876596	0.001000
1	Non Profit Agency	Public (City Operated) Agency	34.334610	16.716448	51.952772	6.468588	0.001000
2	Commercial Agency	Public (City Operated) Agency	17.215193	-0.945610	35.375996	3.146407	0.067567

Table 4: Tukey's Honest Significant Difference test-1

Based on the Post-hoc Test, we can find no big difference in total space between **Nonprofit and Commercial Agencies**. Also, the total space gap between **Commercial and Public Agencies** is small. The total space gap between Nonprofit and Public Agencies is the biggest. The result shows That Nonprofit Agencies have the most total space availability among these three kinds.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Y	N	26.265831	20.400592	32.13107	12.426982	0.001

Table 5: Tukey's Honest Significant Difference test-2

In this table, since the **P-value** is smaller than 0.05, there is a significant difference in the mean total space of childcare centers receiving a subsidy and not. Childcare centers that received subsidies have more space.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(Non Profit Agency, Y)	(Non Profit Agency, N)	35.327657	21.377488	49.277825	10.224542	0.001000
1	(Non Profit Agency, Y)	(Commercial Agency, Y)	8.165515	-7.512442	23.843471	2.102822	0.650278
2	(Non Profit Agency, Y)	(Commercial Agency, N)	26.764597	16.861524	36.667669	10.911871	0.001000
3	(Non Profit Agency, Y)	(Public (City Operated) Agency, Y)	39.460387	17.934574	60.986199	7.401338	0.001000
4	(Non Profit Agency, Y)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900000

Table 5: Tukey's Honest Significant Difference test-3

Based on this table, we can see that non-profit agencies with subsidies have 35 more spaces than non-subsidized agencies. There is no significant difference between a subsidized Non-profit Agency and a subsidized Commercial Agency. Comparing subsidized non-profit agencies and non-subsidized commercial agencies, the difference is more than 26. Subsidized Non-profit agencies have 39 more spaces than subsidized Public Agencies. The P-values of the first, second, and third groups set are all below the standard thread (0.05). There is a significant difference inside each group.

Conclusion

This research shows that non-profit agencies with subsidies have the most total space compared to commercial agencies and public agencies. Conducting the One-Way and Two-Way ANOVA tests is helpful for this research, but the two assumptions still need to be met. Lastly, we have yet to reveal the pattern of the total space distribution of childcare centers. Although assumptions have yet to be met, this result still interests me about the childcare problem in the Toronto area. Further research should be conducted by a more suitable method, such as a Nonparametric Test.