Course: INF2178
Assignment: Assignment 2
Name: Ruiying Wang
Student number: 1003059762

Exploring Childcare centre Space distribution in Toronto

Introduction

Finding licensed or unlicensed childcare for children in Ontario has been a great challenge for all parents due to high fees and low availability of spaces. According to Toronto Children's Service statistics, over 75% of families cannot afford childcare which make the situation to become even worse.

This report delves into the operations and capacities of licensed childcare centers in Toronto, focusing on various age groups. It aims to uncover disparities in space allocation and explore how age and subsidy types influence this distribution. The findings intend to enhance understanding of the current landscape and assist in governmental resource planning and management.

Research question:

1. Is there a significant difference in the percentage of spaces allocated to different age groups across different centres?
2. How do age group and subsidy type impact the distribution of spaces in different centers?

Data Cleaning and Data Wrangling

To conduct a thorough analysis of space distribution by age group and subsidy type across centers, certain data manipulation steps were undertaken:

1. Percentage of space: This metric facilitates the comparison of space distribution among different age groups. Six new columns were formulated, each representing an age category's percentage out of the total spaces across all centers, denoted as 'age group'_pct.
   - a$ge\ group\ _pct(\%) = \frac{number\ of\ spaces\ by\ age\ gropu}{Total\ spaces\ across\ all\ centres} * 100, round\ by\ 2\ decimal.$
   - $age\ group\ includes: 'Infant', 'Toddler', 'Preschool', 'Kindergarten', 'School\_Age'$

2. DataFrame transformation: The DataFrame was transformed into a long format suitable for ANOVA testing, which requires observations to be structured such that each row corresponds to a single observation with one dependent variable and one or more independent variables.

ANOVA Assumption Check

To answer the research questions that we listed, the one-way and two-way ANOVA will be implemented. These statistical tests require certain assumptions to be met for the results to be considered valid.

Assumption 1: Residuals are normally distributed

To verify the normality of the residuals, we look to Q-Q plots and histograms of residuals (Figure 1). The Q-Q plot (Figure 1 – Left) shows some deviation of residuals from the 45-degree reference line in the tails, suggesting some departure from normality, while the middle parts adhere well to the line. The histogram of residuals (Figure 1 – Right) shows a roughly bell-shaped curve with some irregularities, which indicate a non-perfect tendency towards normality. Both plots suggest that the residuals do not perfectly follow a normal distribution.

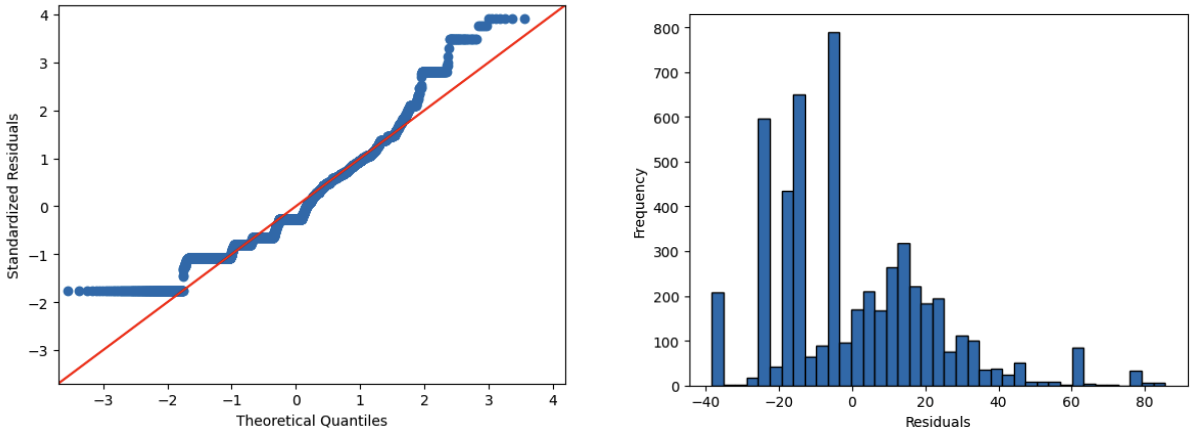


Figure 1: Q-Q Plot (Left); Histogram of Residual (Right)

Figure 2: Shapiro Wilk Test	
Statistical result	**P-value**
0.947	5.517e-40

Shapiro Wilk Test yields a statistical result (W) of approximately 0.947 and a very low p-value, which suggests that the data does not follow a normal distribution.

Assumption 2: Homoegeneous Variance

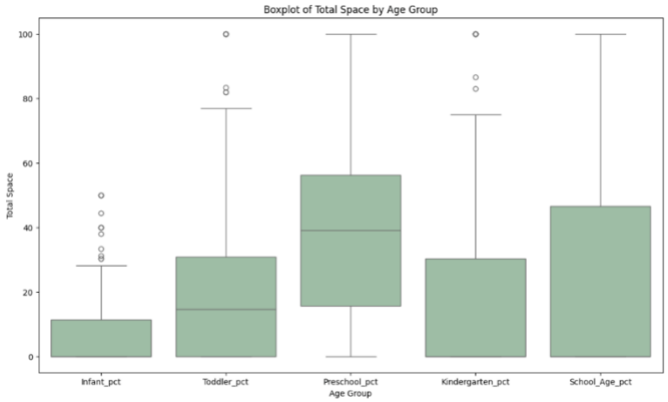
Figure 3: Levenue's test	
Parameter	**Value**
Test statistics (W)	186.606
Degrees of freedom (Df)	4.000
p value	<0.001

Levenue's test (Figure 3) yield a very low p-value, indicates a significant violation of the homogeneity of variances assumption.

The analysis of the two key ANOVA assumptions suggests that the dataset does not fully meet the criteria for ANOVA. Despite some deviations from normality in the residuals and evidence of unequal variances across groups, the large sample size in this study allows for some flexibility. According to the central limit theorem, the distribution of sample means will approximate a normal distribution for large samples, making ANOVA robust to violations of the normality assumption. Additionally, the equal sample sizes among groups can mitigate the impact of unequal variances. Therefore, while proceeding with ANOVA is reasonable, the interpretation of the results should be approached with caution.

Research Question 1: Is there a significant difference in the percentage of spaces allocated to different age groups within each centre?

To aid government efforts in resource distribution, this report examines the variability in space allocation across different age groups within childcare centers, to understand the trends which is vital for aiding centers in strategic planning and optimal space utilization. A One-Way ANOVA test is utilized to investigate significant differences in allocation.



The boxplot shows that Infants have the lowest median space allocation and a tight spread of values, indicating a uniform approach across centers. Toddler spaces show greater variability but a lower median allocation. Preschool and kindergarten spaces have higher medians and more variability. While the school-age group has the highest median allocation and the widest range.

Figure 4: Boxplot of % of spaces by age group across all centres

One-way ANOVA Test

The ANOVA test results (Figure 5) demonstrate significant differences in space allocation among different age groups within childcare centers. With a significantly high F-value of 328.251 and low p-value (lower than the 0.05 threshold), indicates strong statistical evidence that at least one age group's mean space allocation is distinct from the others.

Figure 5: One way ANOVA Test					
	df	**Sum_sq**	**Mean_Sq**	**F**	**FR(>F)**
C (age_group)	4.0	628926.782	157231.695	328.251	8.847e-253
Residual	5310.0	2543479.733	478.998	NaN	NaN

Post-hoc Tukey's HSD Test

The Tukey HSD post-hoc test was applied to identify specific differences between pairs of age groups, and the results (Figure 6) indicates that there are significant differences in the percentage of spaces allocated to different age groups within childcare centers (all p-values are less than 0.05). Looking at the mean difference in the percentage of spaces between the age groups, the comparison between infant and preschool showing the largest mean difference (32.68) and the comparison between toddler and kindergarten showing the smallest (3.20), which implies that preschoolers, on average, are allocated significantly more space than infants by a substantial margin, while the space allocation for toddlers and kindergarteners is more similar compared to other age group comparisons.

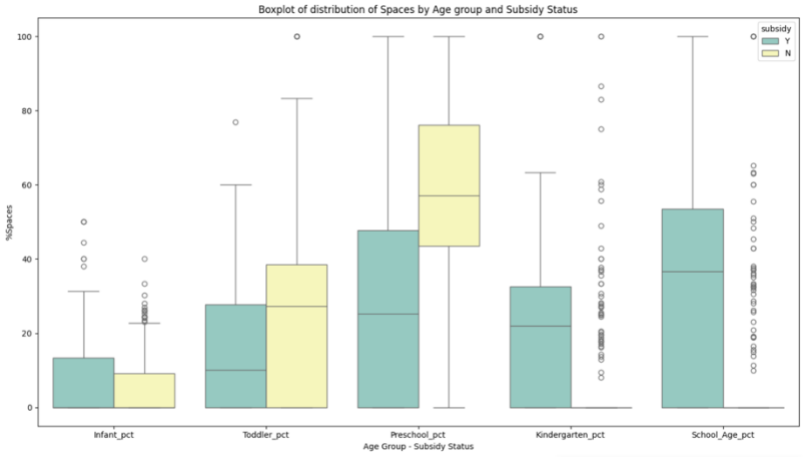
Figure 6 Tukey HSD post-hoc test						
group1	**group2**	**Diff**	**Lower**	**Upper**	**q-value**	**p-value**
Infant	Toddler	11.992	9.401	14.582	17.864	0.001
Infant	Preschool	32.681	30.090	35.271	48.685	0.001
Infant	Kindergarten	8.790	6.200	11.380	13.095	0.001
Infant	School_Age	17.990	15.399	20.580	26.800	0.001

Toddler	Preschool	20.688	18.098	23.279	30.820	0.001
Toddler	Kindergarten	3.201	0.611	5.792	4.769	0.006
Toddler	School_Age	5.998	3.407	8.588	8.935	0.001
Preschool	Kindergarten	23.890	21.300	26.481	35.589	0.001
Preschool	School_Age	14.690	12.100	17.281	21.885	0.001
Kindergarten	School_Age	9.199	6.609	11.790	13.704	0.001

From the One-way ANOVA test, it emphasizes the inequality of the distribution of number of spaces provided by different age groups within each centre, where majority of the childcare services focused on providing services to school age, and lowest to infants who relatively require more cost and manpower compared to the others. To comprehensively address the issue of unequal space distribution, it is crucial to delve into the economic aspects of childcare operations, including the costs and revenues associated with caring for different age groups.

Research question 2: How do age group and subsidy type impact the distribution of spaces in different centers?

Besides the low availability, another key issue that Toronto family is facing is the high fees of the childcare services, that over 75% of families cannot afford childcare and the situation is getting even worse. Offering fee subsidy contract to childcare services may be one solution in providing lower fee services and increasing the number of spaces within the childcare services. To further investigated the relationship between subsidy status and space allocation among various age groups, a two-way ANOVA test is implemented.



The boxplot illustrates a clear connection between subsidy status and the distribution of childcare spaces among age groups, with generally higher median percentages of spaces in subsidized centers for toddlers, preschoolers, and school-age children compared to non-subsidized centers,

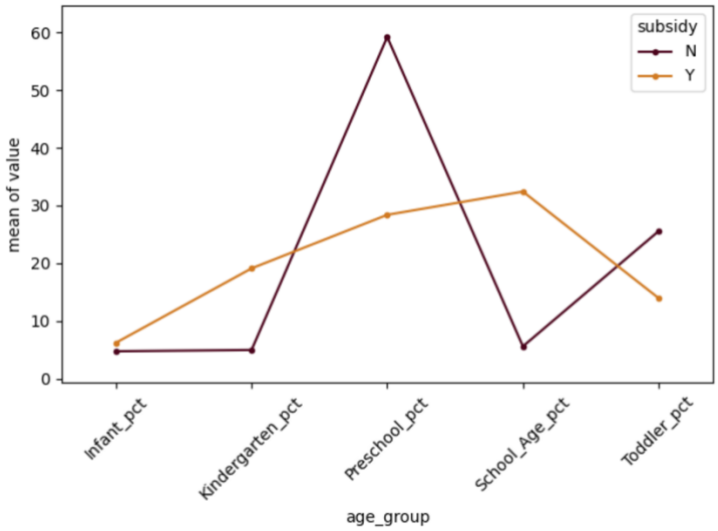
Figure 7: Boxplot of distribution of spaces by age group and subsidy status

Two-way ANOVA Test

From the two-way ANOVA table, it indicates that while subsidy status alone does not significantly affect space allocation (F-value is extremely small, and p-value is 0.999, which is much greater than 0.05), both the age group of the children (F-value is large , and the p-value is near zero) and the interaction between subsidy status and age group (large F-value with a p-value near zero) significantly influence how spaces are distributed in childcare centers. The

interaction effect suggests while subsidies may not alter space allocation across the board, they play a key role when combined with age-specific considerations, and it is crucial to acknowledge these differential impacts and tailor subsidy programs accordingly.

Figure 8: Two-way ANOVA Results					
index	**df**	**sum_sq**	**mean_sq**	**F**	**PR(>F)**
C(age_group)	4.0	628926.782	157231.695	401.936	7.007e-303
C(subsidy)	1.0	2.740e-05	2.740e-05	7.004e-08	0.999
C(age_group):C(subsidy)	4.0	468242.006	117060.501	299.245	2.058e-232
Residual	5305.0	2075237.726	391.185	NaN	NaN



The interaction plot indicates that the relationship between age group and space allocation is influenced by subsidy status, with the peak at kindergarten. The differences in trends for preschool and school-age groups suggest that subsidy status has a different impact on space allocation for these age groups.

Figure 9: Interaction plot of mean % values for different age groups by subsidy status

Post-hoc Tukey's HSD Test

The Tukey HSD post-hoc (Figure 10) test following the ANOVA reveals marked differences in the allocation of childcare spaces among various age groups, particularly within subsidized centers. Age has a significant influence, with the greatest disparity occurring between infants and preschoolers, while subsidy status notably doesn't impact infants' space allocation, as indicated by the non-significant p-value. (p-value of 0.900). This suggests that while older children in subsidized centers benefit from more spaces, subsidies don't appear to affect the distribution of spaces for infants (p-value of 0.9). These results also underscore the need to evaluate the effectiveness of subsidies in addressing the needs of the youngest age group and suggest that age-specific considerations should inform the optimization of space allocation strategies in childcare centers.

Figure 10: Post-hoc test – Subsidy						
group1	**group2**	**Diff**	**Lower**	**Upper**	**q-value**	**p-value**
Y	N	0.00015	-1.135	1.136	0.00037	0.9

Figure 10: Post-hoc test – Age group						
group1	**group2**	**Diff**	**Lower**	**Upper**	**q-value**	**p-value**
Infant	Toddler	11.992	9.651	14.333	19.768	0.001
Infant	Preschool	32.6810	30.340	35.022	53.873	0.001
Infant	Kindergarten	8.790	6.449	11.131	14.490	0.001
Infant	School_Age	17.990	15.649	20.331	29.655	0.001
Toddler	Preschool	20.688	18.347	23.029	34.104	0.001
Toddler	Kindergarten	3.201	0.860	5.542	5.277	0.0017
Toddler	School_Age	5.998	3.657	8.339	9.887	0.001
Preschool	Kindergarten	23.890	21.549	26.231	39.382	0.001
Preschool	School_Age	14.690	12.349	17.031	24.217	0.001
Kindergarten	School_Age	9.199	6.858	11.540	15.165	0.001

Figure 10: Post-hoc test – Subsidy status and Age Group						
group1	**group2**	**Diff**	**Lower**	**Upper**	**q-value**	**p-value**
Y,Infant	Y,Toddler	7.741	4.4376	11.045	10.488	0.001
Y,Infant	Y,Preschool	22.191	18.887	25.495	30.064	0.001
Y,Infant	Y,Kindergarten	12.904	9.600	16.208	17.482	0.001
Y,Infant	Y,School_Age	26.222	22.918	29.526	35.526	0.001
Y,Infant	N,Infant	1.4752	-2.625	5.576	1.610	0.9

Conclusion

The above analysis of childcare space distribution in Toronto has highlighted significant disparities across different age groups and the impact of subsidy status on space allocation within centre. The One-Way and Two-Way ANOVA tests established that there are notable differences in space allocation among the age groups, while the provision of subsidies is shown to significantly affect space allocation for toddlers, preschoolers, and school-age children, suggesting that policy measures such as subsidies could indeed incentivize childcare centers to allocate more spaces to these age groups. However, the same is not observed for infants, where subsidy status does not significantly impact space allocation, as demonstrated by the non-significant difference in spaces between subsidized and non-subsidized groups for infants. These findings underscore the necessity for policymakers to refine subsidy programs, ensuring that the benefits of such policies are more evenly distributed across all age groups. By doing so, they can better meet the needs of families across Toronto and support the development of a more inclusive and accessible childcare system.