Exploring Childcare Space Capacity for Licensed Childcare Providers in Toronto

Chenwei Zhu
1006003619

Introduction

Finding affordable childcare for families has been a challenge in Ontario, especially in metropolitan areas like Toronto. The high cost of childcare and the limited number of childcare spaces creates significant social pressure, especially for working families. According to a study by Toronto Children's Services, as many as 75% of families cannot afford childcare, exposing a pressing social issue: the affordability and accessibility of childcare. Responding to this challenge, the Ontario government has committed to creating an additional 100,000 childcare spaces between 2016 and 2026, with the aim of providing families with more choices and reducing their financial burden.

To delve deeper and address this social issue, this study will analyze licensed childcare centers in Toronto using the dataset titled "INF2178 A2 data.xlsx". This dataset, last updated in February 2024, contains operational and capacity information for these centers for multiple age groups. Our analysis will focus on the capacity of childcare centers. Our exploration will address two fundamental research questions that will reveal the impact of different operational factors on the number of children that can be accommodated in childcare centers, with a view to identifying the key factors that influence the supply of childcare services.

1. **Research Question 1**: Whether the different operational nature of childcare centers has an impact on their total childcare spaces.
2. **Research Question 2**: Whether the subsidized status of a childcare center and its membership in CWELCC has an impact on its total childcare spaces，and the presence or absence of mixed effects

The purpose of this study is to provide policy makers with data to support the development of more effective childcare service strategies. Through this analysis, we hope to provide families with more and better childcare options, which will ultimately serve the purpose of easing the burden on families and promoting child welfare.

Data Cleaning and Data Wrangling

In this study of the dataset of licensed childcare centers in Toronto, we began by carefully cleaning and organizing the raw data to ensure the accuracy and validity of the analysis. The purpose of data cleaning was to identify and correct errors or inconsistencies in the data to provide a solid foundation for data analysis. The original dataset contained 1063 rows of entities as well as 17 columns of variables. Our initial review indicated that the original dataset did not require extensive data cleaning within the scope of our analysis. Nonetheless, we noted several observed differences and accordingly defined new features needed for future analyses. Below are the steps has been took to organize the data:

Observations and considerations:

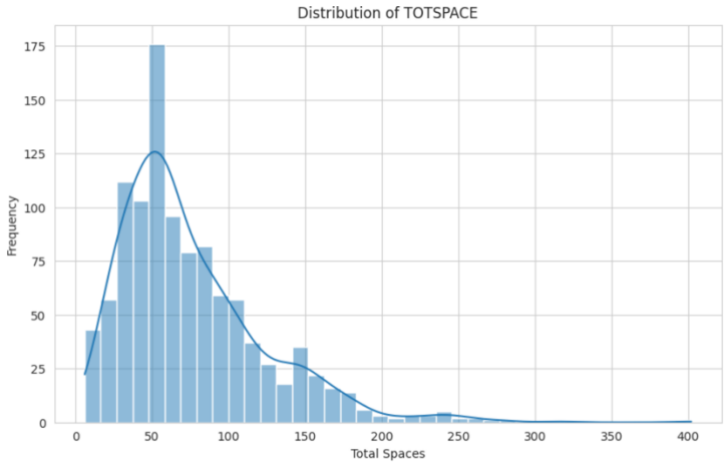
1. Checking for missing values:
   Noticed the presence of 348 missing values (NaN) in the variable 'BLDGNAM'. The variable 'BLDGNAM' stands for 'the Name of the building the childcare center is located in'. This variable will not be used for subsequent research on total childcare space in childcare centers. Therefore after carefully reviewing the data, I chose to set it aside and not process it.
2. Select columns that are closely related to the study objectives, specifically:

AUSPICE	Operating auspice (Commercial, Non-Profit or Public)
TOTSPACE	Childcare spaces for all age groups
Subsidy	Centre has a fee subsidy contract (Yes/No)
cwelcc_flag	'Y' indicates space participates in CWELCC; blank indicates it does not'

Through the above steps, the tidiness and consistency of the data set is ensured, laying a solid foundation for further statistical analysis. These meticulous preparations will help us gain insight into the different factors.

Exploratory Data Analysis (EDA)

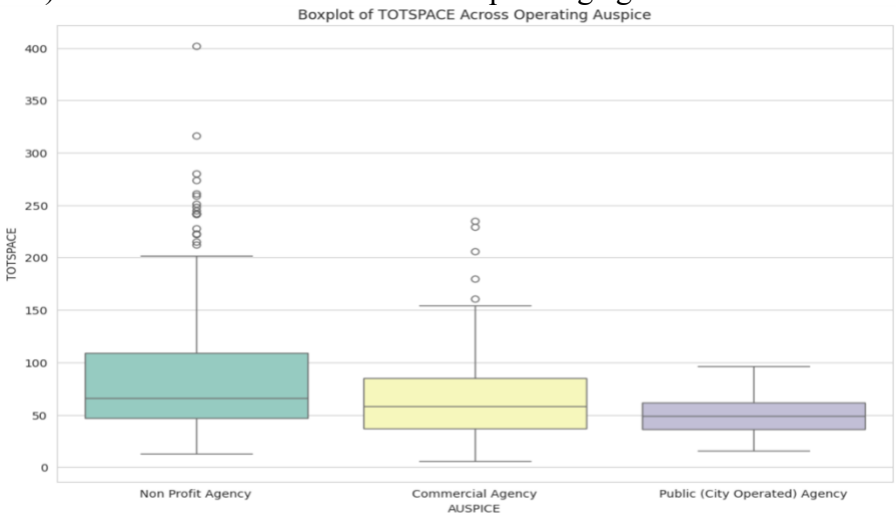
Firstly, use descriptive analysis and histograms were plotted to explore its general distributional trends of "TOTSPACE". The results of the descriptive analysis show that out of 1063 samples, the mean value of "TOTSPACE" is about 75.67 and the standard deviation is about 47.82. The minimum value is 6 and the maximum value is 402. The histogram shows the distribution of "TOTSPACE", which represents the total number of available spaces in the childcare centers. From the histogram [Figure 1], several characteristics of the distribution of the data can be observed: the data seems to be right skewed, which means that there are more childcare centers with less total space and fewer with more space.



[Figure 1]

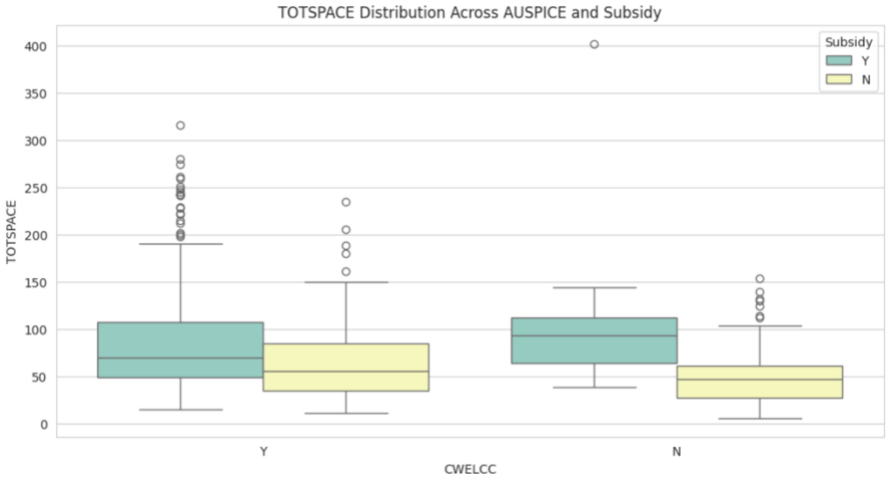
This is evidenced by the fact that the peak of the histogram is located on the left side of the graph, which does not follow a normal distribution. And while there is a wide range of "TOTSPACE" values, most of the data is clustered at the lower end of the range, with fewer centers with very high totals of spaces. This manifests itself as a long tail extending to the right. The distribution appears to be unimodal with only one peak. This suggests that there is a common scale for the total number of spaces in childcare centers, with most centers being close to this scale. And there are some outliers on the right side that need to be noted in the subsequent analysis.

Secondly, create boxplots according to each of the two research questions. 'Boxplot of TOTSPACE Across Operating Auspice' [Figure 2] was changed to show the distribution of TOTSPACE (total number of spaces) in childcare centers of different operating agencies.



[Figure 2]

According to the Figure 2, there may be distributional differences in the total number of spaces in childcare centers operated by different types of agencies. Non-profit agencies provide a more variable number of spaces, while municipally operated childcare centers tend to provide fewer spaces. Whether the differences in the number of total spaces in childcare centers operated by different agencies are statistically significant needs to be analyzed using the One Way ANOVA.



[Figure 3]

Based on the second research question, a Boxplot on TOTSPACE Distribution Across AUSPICE and Subsidy was plotted [Figure 3]. This figure shows the distribution of total space (TOTSPACE) for childcare centers with or without CWLCC enrollment in the case of subsidy (Subsidy Y or N). The figure shows two box-and-line plots for childcare centers that are enrolled in the CWLCC and those that are not enrolled in the CWLCC, one corresponding to the provision of subsidies (Y) and the other corresponding to the non-provision of subsidies (N), respectively. As can be seen from this plot, there may be a difference in the total number of spaces between childcare centers that receive subsidies and those that do not. In particular, subsidized childcare centers appear to have more outliers, suggesting that certain subsidized childcare centers have an unexpectedly high number of total spaces. Also, childcare centers that are not enrolled in the CWLCC have a higher number of total spaces compared to childcare centers that are enrolled in the CWLCC. This may be due to the fact that non-CWLCC-affiliated childcare centers charge higher fees and are therefore better funded. Whether or not there is an effect of joining CWLCC and subsidy status on the total number of spaces and whether there is a mixed effect of them on the total number of spaces, respectively, needs to be analyzed by performing a TWO WAY ANOVA.

Research Question 1

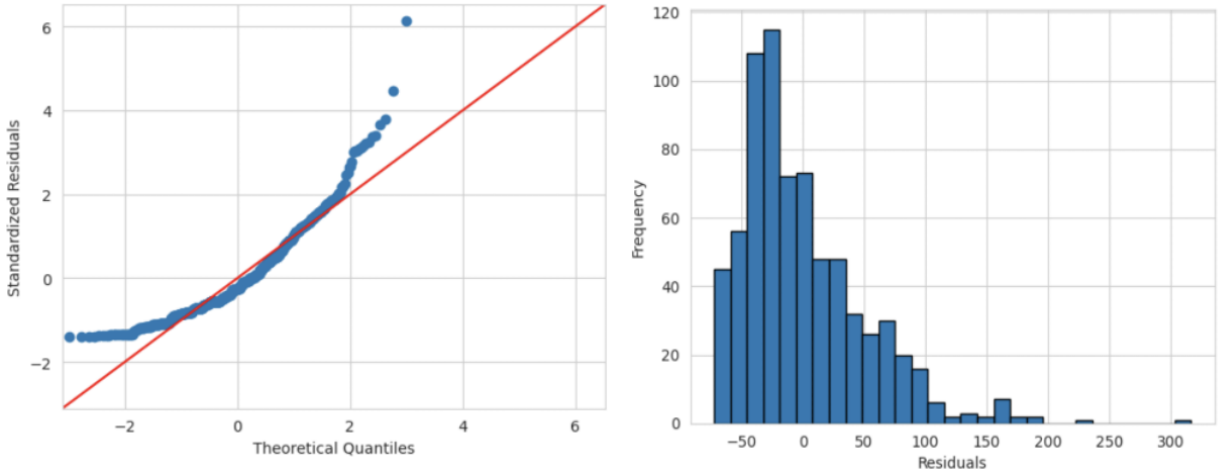
In order to investigate whether the type of childcare center operation has an impact on the total number of spaces in a childcare center, a One-Way ANOVA was used. ANOVA analysis was conducted using AUSPICE (type of operation) as a categorical variable, and TOTSPACE (total number of spaces) as a numerical variable. Based on the base setting of One Way ANOVA we formulate the hypotheses, Null Hypothesis (H0): the mean of all groups is the same, i.e., there is no difference in the number of total spaces in childcare centers with different types of operations. Alternative hypothesis (H1): at least one group has a different mean than the others. After running One-Way ANOVA, the mean, within-group and between-group variance, and the corresponding F-statistic and p-value were calculated for each group. the ANOVA table showed a p-value of 5.057716e-10 ($p\text{-value} < 0.05$). This result indicates that we have enough evidence to reject the null hypothesis. Thus at least one group has a different mean from the others.

In ANOVA we have shown that at least one group has a mean different from the others. Subsequently the post hoc test was used to find out exactly which two groups were directly significantly different. The results of the Post hoc test [table 1] show that there is a difference in the amount of total space between Non-Profit Agency and Commercial Agency ($p\text{-value} < 0.05$), and there is also a significant difference in the amount of total space between Non-Profit Agency and Publicly Operated Agency ($p\text{-value} < 0.05$).

group1	group2	Diff	Lower	Upper	q-value	p-value
Non-Profit Agency	Commercial Agency	16.806538	3.993722	29.619353	4.356853	0.006071
Non-Profit Agency	Public (City Operated) Agency	36.177966	8.673910	63.682022	4.369046	0.005901
Commercial Agency	Public (City Operated) Agency	19.371429	-10.141900	48.884757	2.180132	0.272554

[Table 1]

To ensure that the results of the ANOVA are valid and reliable, we have to test the assumption of the ANOVA. From the Q-Q plot [Figure 4], the sample does not obey the normal distribution, and the result of Shapiro Wilk test (p-value = 1.496e-25 < 0.05) also shows that the sample does not follow normal distribution, so the first assumption of the ANOVA is not valid. Since the sample does not follow a normal distribution, Levene's test was used to check the second Assumption (homogenous variances). the test result shows p-value = 0.0001 (p-value < 0.05), so the second Assumption is also not valid. The failure of Assumption affects the accuracy and reliability of the ANOVA results.



[Figure 4]

Research Question 2

A Two-Way ANOVA was used to explore whether a childcare provider's subsidy status and its membership in CWELCC had an effect on its total number of spaces and whether there was a mixed effect. The steps of the analysis were broadly like Research Question 1. Null Hypothesis (H0): there is no effect of subsidy status and CWELCC membership on the total number of spaces in childcare centers and there is no interaction effect. Alternative hypothesis (H1): at least one factor influences the total number of spaces, or there is an interaction effect between the two factors. The results of the analysis [Table 2] show that there is a significant effect of both receiving a subsidy and being a member of the CWELCC on the total number of spaces in the childcare center (p-value < 0.05) and a significant effect of the mixed effect between the two on the total number of spaces in the childcare center. (p-value < 0.05)

	df	sum_sq	mean_sq	F	PR(>F)
C(cwelcc_flag)	1.0	6.723429e+03	6723.428955	3.176381	7.499648e-02
C(subsidy)	1.0	9.816181e+04	98161.809243	46.375040	1.633653e-11
C(cwelcc_flag):C(subsidy)	1.0	1.910868e+04	19108.682949	9.027604	2.721895e-03
Residual	1059.0	2.241580e+06	2116.694877	NaN	NaN

[Table 2]

We then determined by post hoc test that there was a significant difference in the number of total spaces between subsidized and non-subsidized childcare centers (p-value = 0. 0001). There was a significant difference in the number of total spaces between CWELCC members and non-CWELCC members only (p-value = 0. 0001). In addition to testing the single effects of the factors on the total number of spaces using the post hoc test [Table 3], their mixed effects were also tested.

group1	group2	Diff	Lower	Upper	q-value	p-value
(Y, Y)	(Y, N)	29.128683	14.090822	44.166544	7.054287	0.001000
(Y, Y)	(N, Y)	30.320515	-9.030303	69.671332	2.806087	0.194903
(Y, Y)	(N, N)	40.489249	22.077829	58.900670	8.008861	0.001000
(Y, N)	(N, Y)	59.449198	18.030838	100.867558	5.227228	0.001344
(Y, N)	(N, N)	11.360566	-11.133302	33.854434	1.839307	0.555438
(N, Y)	(N, N)	70.809764	28.050863	113.568666	6.030939	0.001000

[Table 3]

The results show the average TOTSPACE (Total Number of Spaces) difference between combinations of different subsidy statuses (Y for providing subsidies and N for not providing subsidies) and CWELCC membership statuses (Y for being a member of CWELCC and N for not being a member). For example, the mean TOTSPACE difference between the combination (Y, Y) (i.e., subsidy provided, is a CWELCC member) and (Y, N) (subsidy provided, is not a CWELCC member) is 29.13 and this difference is statistically significant ($p < 0.05$). For reasons of space, the results of the post hoc test will not be repeated, please refer to Table3.

Similarly, to ensure the validity and reliability of the analytical results, the assumption of ANOVA needs to be tested. The result of Shapiro Wilk test (p-value = 3.3379e-26, which is p-value < 0.05) suggests that the sample does not obey the normal distribution. The results of Levene's test (p-value = 0.0001) indicates that homogenous variances do not hold in this case. Failure of Assumption can lead to a decrease in the reliability and validity of the results of the ANOVA analysis.

Conclusion

Through exploratory analysis and ANOVA, we obtained valuable information about the factors that influence the amount of total space in licensed childcare centers in Toronto. Our findings indicate that the nature of the childcare center's operation (AUSPICE) has a statistically significant impact on the number of total spaces in a childcare center. there is a difference in the amount of total space between Non-Profit Agency and Commercial Agency and there is also a significant difference in the amount of total space between Non-Profit Agency and Publicly Operated Agency. whether or not they receive a subsidy and whether they are enrolled in CWELCC are also two significant influences on the total space, and there is a mix of these two factors, with the mix also having a significant effect on the amount of total space in childcare centers. This information can help policymakers better understand what factors affect the total number of spaces in licensed childcare centers in Toronto so that they can target measures to alleviate the shortage of total spaces in childcare centers. There are also limitations to this study; this study used ANOVA for the analysis of the research questions. However, two assumptions of the ANOVA were not satisfied in this study. When the data deviate significantly from the normal distribution, the results of the ANOVA may be affected. Especially with small sample sizes, deviations from normality may increase the risk of making Type I or Type II errors. If the variances of different groups are significantly different (known as variance), the results of the ANOVA may be affected. groups are significantly different (known as variance heterogeneity), this may affect the accuracy of the F-statistic and therefore the reliability of the results. Thus, the accuracy of the results of the current study may be open to question, but it also points to a general direction for subsequent research and helps subsequent researchers and policy makers to better understand the factors that influence the total number of spaces in licensed childcare centers in Toronto.