Student Name: Jiajun Yang
Student Number: 1010128862

Analysis of Childcare Space Availability by Auspice and Childcare Affordability Initiative in Toronto

1. Introduction

The availability and affordability of child care in Ontario have emerged as significant concerns for many families, exacerbated by high fees and a scarcity of available spaces. This situation has placed a considerable strain on families, with Toronto Children's Services reporting that 75% of families find child care unaffordable. In response to this growing challenge, the provincial government has committed to creating 100,000 new child care spaces from 2016 to 2026, highlighting the urgent need for strategic planning and analysis to ensure that these spaces are distributed in a way that maximizes their impact on the community.

This report sets out to explore the current landscape of childcare provision in Toronto, focusing on the capacity and management models of existing childcare services. Our analysis begins with a one-way ANOVA to examine the relationship between the auspices under which childcare services are provided and their total available spaces. This initial step aims to identify potential disparities in capacity across different management models, providing insight into which models might best support the expansion of child care spaces. Building on these findings, a more nuanced examination through a two-way ANOVA that considers both the auspices of child care services and their participation in the City of Toronto's childcare affordability initiative.

By conducting these analytical approach, the report aims to address the following research questions:

1. How does the total available space vary among different child care auspices?
2. How do the auspices of childcare services and their participation in the City of Toronto's child care affordability initiative interact to affect the total available spaces?

The insights gained from this analysis will contribute to a deeper understanding of the dynamics shaping child care provision in Toronto. Our goal is to inform policy decisions related to the allocation of new child care spaces, ensuring that resources are directed towards the models and programs that offer the greatest potential for expanding access and affordability for Ontario families.

2. Data Cleaning and Data Wrangling

Our dataset comprises a detailed collection of 1,063 records, spanning across 17 columns, capturing a wide array of information on child care facilities within Toronto. This includes unique identifiers, facility names, auspice types, addresses, postal codes, ward numbers, building types, individual age group spaces (infant, toddler, preschool, kindergarten, school-age), total space availability, and indicators of subsidy availability and participation in the City of Toronto's Child Care Affordability and Accessibility Initiative ('cwelcc_flag').

Observations and Consideration:
Given the quantitative nature of our investigation, we cleaned the dataset somewhat to include specific columns extracted from the raw data. These selections helped us to conduct an analysis focusing on the impact of management models and affordability measures on child care center availability. Below we provided a short description of each column:

- AUSPICE: Operating auspice (Commercial, Non Profit or Public)
- TOTSPACE: Child care spaces for all age groups
- cwelcc_flag: 'Y' indicates space participates in CWELCC, blank indicates it does not'

After keeping the columns that needed to be used, we performed a lookup for missing values. We found that there were no missing values that affected our further research.

3. Exploratory Data Analysis(EDA)

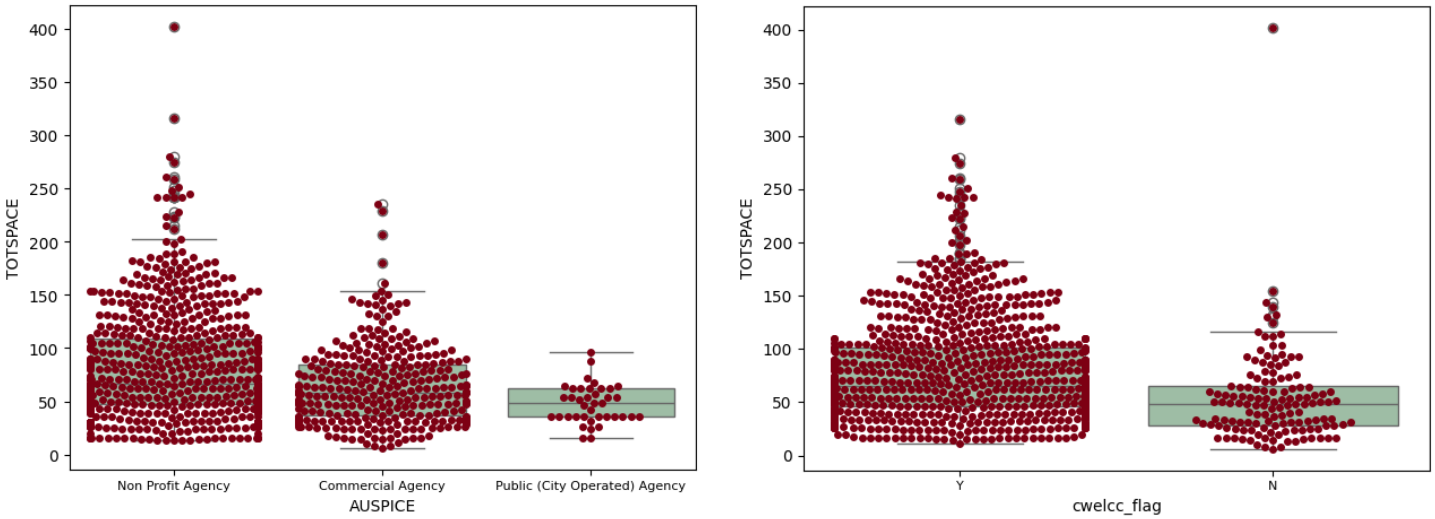

Figure1： Distribution of Total Space Across Childcare Auspices and CWELCC_flag

The two boxplot visualizations provide a clear comparison of the total available childcare spaces ('TOTSPACE') across different management models (auspices) and based on participation in the City of Toronto's childcare affordability initiative (indicated by 'cwelcc_flag').We can find that Non-profit agencies demonstrate a broader range of capacity with higher median 'TOTSPACE', indicating diverse childcare offerings and potential for accommodating more children. In contrast, commercial agencies show a slightly lower median capacity, with public (city-operated) agencies exhibiting the most uniformity and smallest range in available spaces. When examining the impact of the affordability initiative, facilities that participate in the program ('Y') display a greater variance in 'TOTSPACE', with some having significantly high capacities. This contrasts with non-participating facilities ('N'), which have a lower median and less variability in 'TOTSPACE'. The findings indicate a trend where involvement in the Affordability Initiative correlates with an increased ability to offer childcare spaces. This underscores the initiative's significance in broadening the availability of childcare services.

4. One-Way ANOVA

Research questions #1: How does the total available space vary among different childcare auspices?
H0: There is no significant difference in the mean total available space among different childcare auspices.
H1: There is a significant difference in the mean total available space among different childcare auspices.

To address this research question, we employed a One-Way ANOVA to determine if there are significant differences in the total available space across various childcare auspices.

	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	9.61e+04	48056.05	21.84	5.05e-10
Residual	1060.0	2.33e+06	2200.06	NaN	NaN

Table 1: One-Way ANOVA for AUPICE

From the Table 1, the analysis indicates a statistically significant effect of childcare auspices on total available space, as evidenced by a very small p-value ($p < 0.05$). Consequently, we reject the null hypothesis, suggesting that there are differences in total available space among different childcare auspices.

	Group1	Gourp2	Diff	Lower	Upper	q-value	p-value	reject
0	Non Profit Agency	Commercial Agency	17.11	9.70	24.53	7.66	0.001	True
1	Non Profit Agency	Public Agency	34.33	16.22	52.44	6.29	0.001	True
2	Commercial Agency	Public Agency	17.21	-1.45	35.88	3.06	0.077	False

Table 2: Tukey's HSD for AUPICE

The Tukey's HSD analysis demonstrates significant differences in 'TOTSPACE' between Non-Profit Agencies and both Commercial and Public (City Operated) Agencies, with p-values of 0.001, indicating a clear disparity in childcare space capacities among these groups. Specifically, Non-Profit Agencies have a significantly higher mean total space compared to both Commercial and Public Agencies. However, the comparison between Commercial and Public Agencies shows a p-value of 0.077, above the significant level 0.05, which means there is no significant difference in their capacity to provide childcare spaces. This analysis underscores the unique position of Non-Profit Agencies in offering greater childcare space capacities compared to their Commercial and Public counterparts.

To ensure the one-way ANOVA's validity, it's crucial to verify its assumptions.

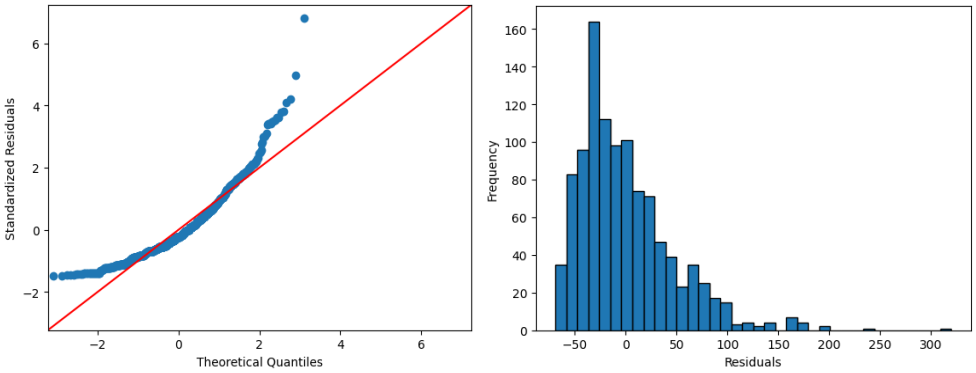

Figure 3: QQ Plot and Distribution Plot of Residuals

Assumption 1: residuals are normally distributed
It appears from the QQ plot from figure3 that the residuals are not normally distributed. If the residuals were normally distributed, they would fall on the red line. However, the points are significantly off the line, especially at the ends, indicating that the residuals are not normally distributed. Additionally, this visual indication of non-normality is corroborated quantitatively by the Shapiro-Wilk test, which yielded a test statistic of 0.90 and a p-value of 1.49e-25. Since the p-value is below 0.05, we reject the null hypothesis that the residuals are normally distributed. Furthermore, the second graph confirms the results of the QQ plot and the Shapiro-Wilk test as the histogram shows that the skew to the right which is not fita normal distribution model. Hence, the first assumption is not valid.

Assumption 2: variances are homogenous:

	Parameter	Value
0	Test statistics (W)	17.92
1	Degrees of freedom (Df	2.00
2	p value	<0.001

Table3: Levene's test for AUSPICE

Since the sample is not normally distributed, we performed Levene's test instead of Bartlett's test to assess the homogeneity of variances assumption for ANOVA. The test yielded a statistic (W) of 17.92 and a p-value < 0.001. With this p-value significantly below the alpha level of 0.05, we reject the null hypothesis that variances across groups defined by 'AUSPICE' are equal. This result indicates that the assumption of homogeneity of variances is not valid.

After testing, we found that the ANOVA does not meet the assumption of normality and the assumption of homogeneity of variances, it indicates that the results of the ANOVA may be unreliable.

5. Two-Way ANOVA
Research questions #2: How do the auspices of childcare services and their participation in the City of Toronto's childcare affordability initiative interact to affect the total available spaces.

H0: There is no interaction between auspice type and affordability initiative participation on childcarespaces.
H1: There is an interaction between auspice type and affordability initiative participation affecting childcare spaces.

To address this assumption, we first generated the interaction plot to assess the potential interaction between 'AUSPICE' and 'cwelcc_flag' on 'TOTSPACE'.

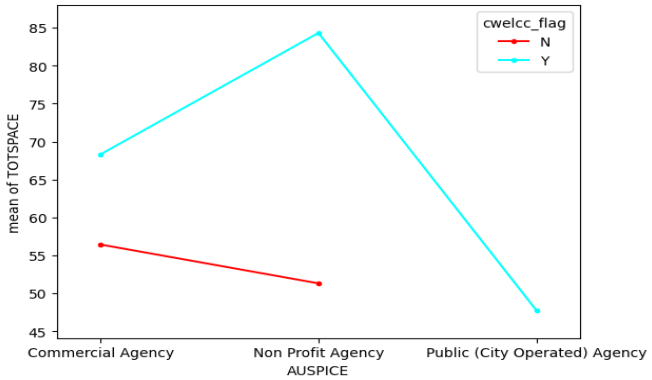

Figure 4: Interaction plot between 'AUSPICE' and 'cwelcc_flag' on 'TOTSPACE'.

From Figure 4, the interaction plot suggests that Non-Profit Agencies see an increase in total spaces with participation in the affordability initiative, whereas Public Agencies show a decrease, indicating an interaction effect between auspice type and program participation on childcare space availability. To verify their interaction effects, we apply the two-way ANOVA and Tukey's HSD test.

	df	sum_sq	mean_sq	F	PR(>F)
C(cwelcc_flag)	1.0	3.76e+04	37688.32	17.52	<0.0001
C(AUSPICE)	2.0	5.41e+04	27083.70	2.59	<0.0001
C(cwelcc_flag):C(AUSPICE)	2.0	2.94e+04	14747.80	6.85	0.001
Residual	1058.0	2.27e+06	2150.46	NaN	NaN

Table 4: Two-Way ANOVA for AUPICE & cwelcc_flag

The two-way ANOVA presented in Table 4 indicates significant main effects for both the childcare affordability initiative ('cwelcc_flag') and the type of childcare agency ('AUSPICE') on the total available spaces ('TOTSPACE'), with F-values of 17.52 and 2.59 respectively, and p-values all less than 0.0001. Furthermore, there is a significant interaction between these factors,since p value equal to 0.001, suggesting that the effect of 'AUSPICE' on 'TOTSPACE' differs depending on the facility's participation in childcare affordability initiative. This interaction means that the impact of the affordability initiative on available spaces is not uniform across different types of agencies.

	group1	Group2	Diff	Lower	Upper	q-value	p-value
0	(Y, Non Profit Agency)	(Y, Commercial Agency)	15.99	5.84	26.12	6.37	0.001
1	(Y, Non Profit Agency)	(Y, Public Agency)	36.54	14.71	58.36	6.76	0.001
2	(Y, Non Profit Agency)	(N, Non Profit Agency)	32.99	12.99	52.98	6.66	0.001
3	(Y, Non Profit Agency)	(N, Commercial Agency)	27.87	12.99	42.75	7.56	0.001
4	(Y, Non Profit Agency)	(N, Public Agency)	<0.001	-inf	inf	<0.001	0.900
5	(Y, Commercial Agency)	(N, Public Agency)	20.54	-2.37	43.46	3.61	0.108
6	(Y, Commercial Agency)	(N, Non Profit Agency)	16.99	-4.18	38.18	3.23	0.198
7	(Y, Commercial Agency)	(N, Commercial Agency)	11.88	-4.56	28.33	2.91	0.308
8	(Y, Commercial Agency)	(N, Public Agency)	<0.001	-inf	inf	<0.001	0.90
9	(Y, Public Agency)	(N, Non Profit Agency)	3.54	-25.12	32.22	0.49	0.90
10	(Y, Public Agency)	(N, Commercial Agency)	8.66	-16.71	34.04	1.37	0.90
11	(Y, Public Agency)	(N, Public Agency)	<0.001	-inf	inf	<0.001	0.90

12	(N, Non Profit Agency)	(N, Commercial Agency)	5.11	-18.71	28.94	0.86	0.90
13	(N, Non Profit Agency)	(N, Public Agency)	<0.001	-inf	inf	<0.001	0.90
14	(N, Commercial Agency)	(N, Public Agency)	<0.001	-inf	inf	<0.001	0.90

Table 5: Tukey's HSD for AUPICE & cwelcc_flag

From Table 5, which detailing the results from Tukey's HSD test, provides insight into the relationships between different groups of child care facilities based on their participation in the City of Toronto's affordability initiative ('cwelcc_flag') and their operating type ('AUSPICE'). The analysis reveals significant differences in 'TOTSPACE' between Non-Profit Agencies participating in the initiative and other groups, indicating that these agencies notably benefit from the initiative in terms of available spaces. Specifically, Non-Profit Agencies show significantly higher spaces compared to Commercial Agencies and Public Agencies within the same participation category, with all significant p-values at 0.001. However, no significant differences were found between Commercial and Public Agencies across participation status, where p-values exceeded the 0.05. This suggests a specific advantage for Non-Profit Agencies participating in the initiative, highlighting the program's role in enhancing child care space availability within this sector.

After applying the two-way ANOVA and Tukey's HSD test, verifying the assumptions underlying ANOVA is crucial to ensure the analysis's validity and reliability, similar to the process in one-way ANOVA. We firstly conducted the Shapiro-Wilk test. The test reuslt showed a statistic of 0.89 and a p-value of approximately 4.58e-26, indicating the data does not follow a normal distribution, thus invalidating this assumption. Regarding the assessment of homogeneity of variances, the Bartlett's test was used. However, the results provided are incomplete, with 'NaN' values for both the test statistic and p-value, making a conclusive evaluation impossible. This issue might stem from the absence of the 'N' for 'cwelcc_flag' in the public agency data. Based on the results of the above hypothesis tests, the reliability of the two-way ANOVA results may be affected.

6. Conclusion

In conclusion, this research aimed to explore the dynamics of child care space availability in Toronto, particularly focusing on the impact of different management models and the participation in the City of Toronto's child care affordability initiative. Utilizing one-way and two-way ANOVA analyses, followed by Tukey's HSD tests, the study found that Non-Profit Agencies, particularly those enrolled in the affordability initiative, generally provide more child care spaces than Commercial and Public (City Operated) Agencies. These findings suggest that policies favoring or aiding Non-Profit child care centers might be an effective strategy for increasing access to child care services. However, the analysis also highlighted that the assumptions of normality and homogeneity of variances were not satisfied, which warrants a careful interpretation of the results.