Exploring Toronto Licensed Child Care Centers
Guanxiao Chen (#1005717841)

1. Introduction
Access to affordable and available child care is a significant concern for families in Ontario, particularly in Toronto. In this analysis, we delve into a dataset (INF2178_A2_data.xlsx) that provides information on the operation, capacity, and demographic characteristics of licensed child care centers in Toronto, as of February 2024.

Our primary objective is to understand the landscape of licensed child care centers in Toronto and explore potential factors affecting their capacity. To achieve this, we will utilize quantitative analysis techniques, specifically one-way and two-way ANOVAs. By examining the relationships between categorical variables (such as program type or age group served) and continuous variables (such as center capacity), we aim to uncover insights that can inform policy decisions and resource allocation in the child care sector.

To structure our analysis, we have formulated the following research questions:
Research Question 1: How does the operating auspice (Commercial, Non Profit, or Public) of child care centers impact their capacity?
Research Question 2: Is there a significant difference in the capacity of child care centers based on the type of building they are located in?
Research Question 3: How does the distribution of child care spaces vary across different age groups, and is there a significant difference in capacity based on the age group served by the child care center?

2. Data Cleaning
The raw dataset has a total of 17 columns with 1063 entries. After reviewing the dataset columns in relation to the research questions provided, it appears that most columns serve a purpose in addressing the research questions, only a few are not necessary to be included. Below are columns that may not be directly relevant to the research questions or analysis; so could be deleted during data cleaning:
_id_: While it can be useful for tracking individual records, it may not contribute directly to the analysis of child care center capacity or other research questions.
_LOC_ID_: Similarly, not necessary for the analysis of capacity and other factors.
ADDRESS: While location information can be valuable, it may not be directly relevant to the analysis of capacity or operating factors.
PCODE: Similar to the address column, postal codes may not be directly relevant to the analysis.
_ward_: Unless specific analysis is planned based on geographical regions or wards, this column may not be necessary for the current analysis.

The following column seemed to have missing values:
BLDGNAME has 715 non-null and 348 missing values; filling the missing values by applying forward fill method.

3. Exploratory Data Analysis
- To better understand the dataset and analyze the research question, first we explored the distribution of each age group on child care spaces.
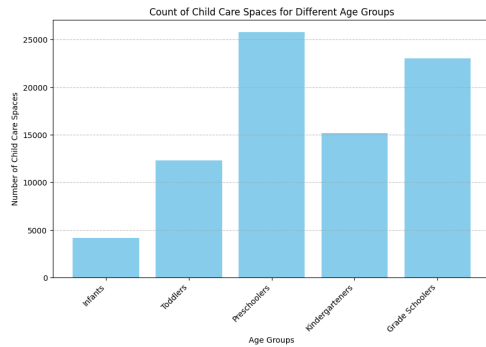
(Figure 1)

The bar plot visually represents the count of child care spaces for each age group, allowing for easy comparison. We observe that the highest number of child care spaces is available for preschoolers (over 25,000), followed by grade schoolers and kindergarteners. Toddlers and Infants have relatively fewer child care spaces available compared to the younger age groups.

- Then, we discovered the average capacity of child care centers based on their operating auspices.
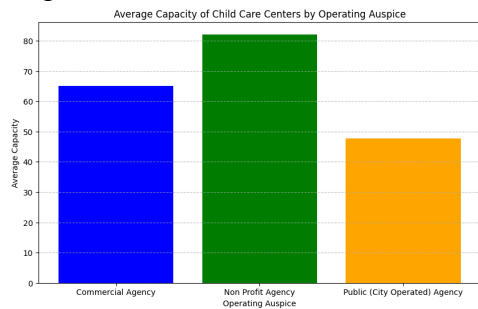
(Figure 2)

The bar graph illustrates the distribution of average capacities across different operating auspices. We observe that child care centers categorized under the 'Non Profit Agency' auspice have the highest average capacity, followed by those under the 'Commercial Agency'' auspice, and finally, centers under the 'Public Agency' auspice have the lowest average capacity.

4. Impact of Operating Auspice on Child Care Center Capacity

Research Question 1: How does the operating auspice (Commercial, Non Profit, or Public) of child care centers impact their capacity?

For this analysis, we'll conduct a quantitative analysis using both one-way and two-way ANOVA to examine how the operating auspice (Commercial, Non Profit, or Public) of child care centers impacts their capacity. We'll also calculate an overall center capacity variable to provide additional insights.

|	sum_sq	df	F	PR(>F)
C(AUSPICE)	9.611211e+04	2.0	21.843051	5.057716e-10

(Table 1)

The one-way ANOVA results revealed a statistically significant effect of operating auspice on child care center capacity ($F(2, 1060) = 21.843$, $p < 0.001$), indicating that the type of auspice is indeed an influential factor in determining the capacity. Unfortunately, the Levene test for homogeneity of variances returned a 'nan' value, suggesting there was an issue with the input

data that needs to be rectified before the homogeneity of variance assumption can be validated. Moreover, the Shapiro-Wilk test indicated that the residuals from our model deviate significantly from normality ($p < 0.001$), which questions the reliability of the ANOVA results. This suggests that the data may not meet the assumption of normality, and transformations or non-parametric tests could be considered for a more robust analysis.

	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

(Table 2)

In the subsequent two-way ANOVA, where both the auspice type and the presence of a subsidy contract were considered, there was no significant interaction between the auspice type and subsidy status on the centers' capacities ($p = 0.128$). However, the subsidy status alone was a significant factor ($F(1, 1058) = 40.118, p < 0.001$), as was the interaction between operating auspice and subsidy status ($F(2, 1058) = 13.457, p < 0.001$). This suggests that subsidy status has a strong influence on capacity, and its impact varies significantly depending on the operating auspice.
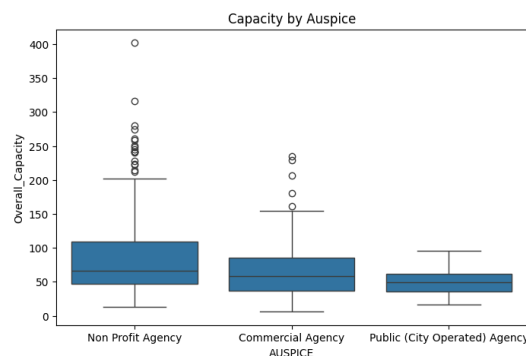
(Figure 3)
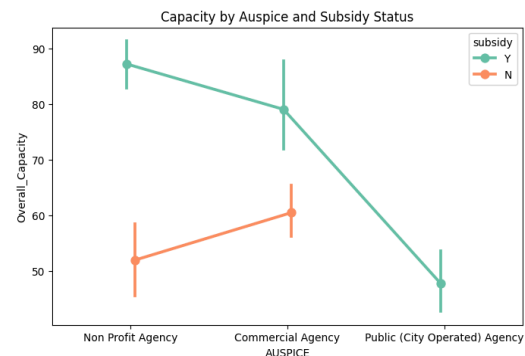
(Figure 4)

The boxplot showing the distribution of overall capacities among child care centers categorized by operating auspice: Non-Profit Agency, Commercial Agency, and Public (City Operated) Agency. The boxplot for Non-Profit Agencies shows a significantly wider interquartile range and higher median capacity compared to the other two categories, with several outliers indicating centers with exceptionally high capacity. The Commercial Agencies have a narrower interquartile range, a lower median, and fewer outliers. The Public (City Operated) Agencies display the narrowest interquartile range and the lowest median capacity, with very few centers having capacities above the upper quartile.

The second graph, a point plot, presents overall capacity by operating auspice and subsidy status. It shows that, on average, Non-Profit Agencies without subsidies have higher capacities than those with subsidies, possibly suggesting that Non-Profit Agencies without subsidies are larger in size or more numerous. For Commercial Agencies, those with subsidies appear to have a

slightly higher average capacity than those without. Interestingly, the Public (City Operated) Agencies show a dramatic decrease in average capacity for those without subsidies, indicating that subsidy status has a considerable impact on capacity within these agencies.

5. Analysis of Building Type and Child Care Center Capacity
Research Question 2: Is there a significant difference in the capacity of child care centers based on the type of building they are located in?
We will perform quantitative analysis using one-way and two-way ANOVAs to examine the differences in the continuous variable representing child care center capacity across different building types.

|	sum_sq	df	F	PR(>F)
C(bldg_type)	5.205689e+05	29.0	9.72056	1.715361e-37

(Table 3)

The one-way ANOVA revealed a highly significant effect of building type on the overall capacity of child care centers ($F_{(29, 1033)} = 9.72$, $p < 1.72e-37$), indicating that the building type is a strong determinant of capacity. However, the extremely low p-value in the Levene test ($p < 1.58e-19$) suggests that the assumption of equal variances is violated, which can affect the ANOVA's validity. Similarly, the Shapiro-Wilk test resulted in a very low p-value ($p < 1.02e-20$), indicating that the residuals are not normally distributed, which calls into question the normality assumption.

|	sum_sq	df	F	PR(>F)
C(bldg_type)	9.998990e+05	29.0	1.899015e+01	1.558573e-29
C(AUSPICE)	-1.684039e-06	2.0	-4.637589e-10	1.000000e+00
C(bldg_type):C(AUSPICE)	2.622099e+05	58.0	2.489954e+00	6.381552e-06

(Table 4)

The subsequent two-way ANOVA, which included both building type and operating auspice, indicated significant interaction effects ($p < 6.38e-06$), suggesting that the impact of building type on capacity is different depending on the operating auspice of the center. However, some warnings regarding the rank of the covariance matrix suggest that there may be issues related to multicollinearity or other data anomalies that should be further explored.
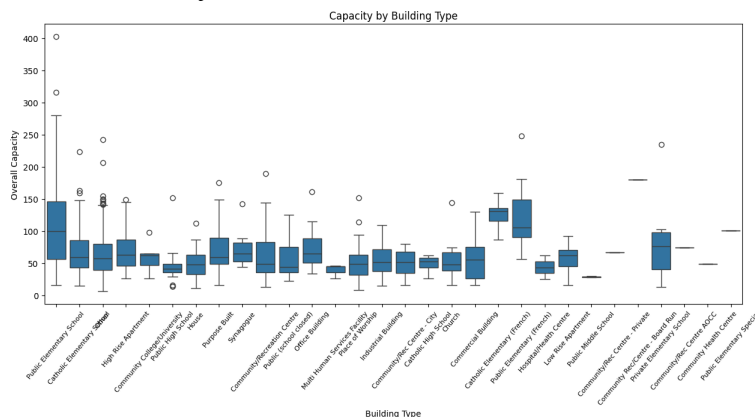
(Figure 5)

The boxplot shows a wide variation in capacity across building types, with some types like 'Public Elementary School' and 'High Rise Apartment' having a larger range of capacities, as evidenced by the length of their boxes and whiskers. Notably, 'Public Elementary School' displays the greatest variation, with some centers having capacities exceeding 250, which could be outliers. This suggests that capacity is not uniform across building types and that certain building types may be more likely to accommodate larger numbers of children.
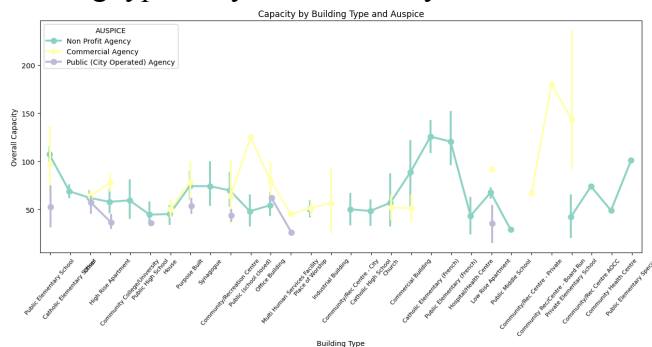
(Figure 6)

The second graph, a point plot, reveals the interaction between building type and operating auspice. It suggests that building type alone does not fully account for the differences in capacity; the operating auspice also plays a role. For example, 'Commercial Agency' shows less variation in capacity across building types compared to 'Non-Profit Agency' and 'Public (City Operated) Agency'. Notably, within certain building types, such as 'Commercial Building', the 'Non-Profit Agency' appears to have a higher capacity than the other types of auspice.

6. The Influence of Age Groups and Operating Auspice

Research Question 3: How does the distribution of child care spaces vary across different age groups, and is there a significant difference in capacity based on the age group served by the child care center?

|	sum_sq	df	F	PR(>F)
C(AgeGroup)	2.821233e+05	4.0	188.190768	4.517383e-151

(Table 5)

In exploring the distribution of child care spaces across different age groups, our analysis has revealed insightful patterns. Utilizing one-way ANOVA, we discovered a highly significant difference in capacity across age groups ($F(4, 5310) = 188.19$, $p < 0.001$). This substantial variation suggests that the capacity allocated within child care centers is distinctly age-dependent, highlighting the diverse needs and resources allocated for infants through school-age children.

|	sum_sq	df	F	PR(>F)
C(AUSPICE)	32108.667729	2.0	50.888535	7.889039e-22

(Table 6)

Furthering our investigation with a two-way ANOVA focusing on preschooler space as a function of operating auspice, we found a significant effect of auspice on capacity ($F(2, 1060) = 50.89$, $p < 0.001$). This indicates that the auspice under which a center operates is a major determinant in the capacity available for preschool-aged children, potentially reflecting

differences in funding, philosophy, or operational strategy among commercial, non-profit, and public agencies.
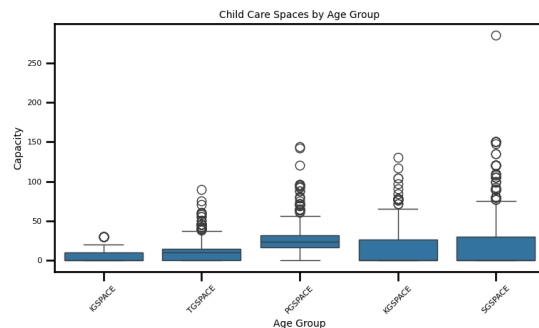
(Figure 7)
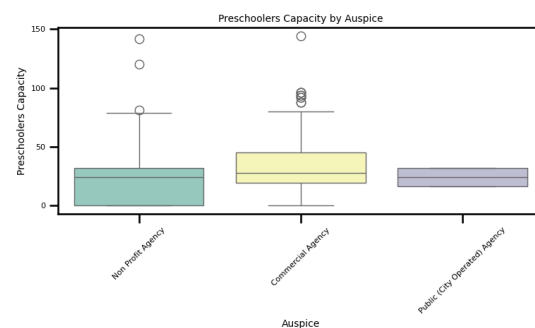
(Figure 8)

This graph displays a comparison of capacity across different age-defined spaces within child care centers. Infant spaces have the lowest median capacity, with relatively few outliers, suggesting a uniformity in infant care provision. Toddler and preschool spaces have higher medians and more variability, as shown by the spread of the data points and outliers. Kindergarten and school-age spaces exhibit the highest median capacities, with school-age spaces showing a particularly large range and many outliers, indicating that these spaces vary widely in capacity.

The second graph presents the capacity distribution for preschoolers across different operating auspices. The Non-Profit Agency category shows the broadest interquartile range, suggesting a high variation in capacity among these centers, while the Commercial Agency has a narrower range, but both exhibit outliers that indicate some centers have much higher capacities. The Public Agency has the smallest range and no outliers above the upper whisker, indicating more consistency in capacity among these centers, though it does present an outlier below the lower whisker.

7. Conclusion

In conclusion, the comprehensive analysis of the factors influencing child care center capacities in Toronto has yielded several key findings that have significant implications for policy and operational strategies within the child care sector. The one-way ANOVA results indicated a substantial difference in capacity across various building types, underscoring the importance of physical infrastructure in the provision of child care services. Furthermore, the interaction effects observed in the two-way ANOVA pointed to a nuanced relationship between operating auspice and building type in determining capacity, highlighting the complex interdependencies within the child care system.

Our findings reveal that Non-Profit Agencies tend to offer a wider range of capacities, possibly reflecting their diverse operational models and funding mechanisms. Meanwhile, Public (City Operated) Agencies show a narrower capacity range, indicating a more uniform approach to service provision. Additionally, the presence of a fee subsidy contract was found to be a significant factor influencing capacity, suggesting that financial support mechanisms play a crucial role in the ability of centers to accommodate children.