

Suha Lee
1005155626
09-03-2024

INF2178 Technical Assignment 2

Examining Licensed Child Care Centres in Toronto

1. Introduction

Toronto's childcare sector is in a critical state, grappling with the twin challenges of unaffordability and scarcity of available spaces, issues that have profound implications for families across the city. Due to the growing demand for quality childcare, parents find themselves struggling against rising costs that place a significant strain on household finances. Concurrently, the shortage of spaces exacerbates the problem, forcing many into lengthy waiting lists with no guarantee of placement.

This report delves into the current state of childcare centres in Toronto, examining the availability of services to identify potential pathways towards more accessible solutions. We are primarily interested in investigating two research questions:

Research Question 1 (RQ1): How does the total capacity of child care centres differ among various auspice types in Toronto?

Research Question 2 (RQ2): How do the capacities for different age groups in child care centres vary across auspice types in Toronto, and is there an interaction effect between age group and auspice type on capacity?

By addressing these research questions, we seek to uncover trends in the data that could highlight disparities in service provision, pinpoint potential gaps in available services, and ultimately inform policy decisions and initiatives aimed at improving the accessibility and equity of child care services in Toronto.

2. Data Preprocessing and Exploratory Data Analysis

This report examines a dataset covering licensed child care centres in Toronto detailing location, capacities across age groups, type of operation, as well as additional information on involvement in government programs (last updated February 2024).

The dataset consists of 1063 unique rows spanning 17 columns of features. For our purposes, we are interested in the feature *Auspice* which details the type of organisation for each service's operating agency as well as the various daily capacity values for the different age groups.

Auspice is a categorical variable consisting of three levels - non-profit, commercial, and public. Of the licensed child care centres in the dataset, a majority of them are listed as non-profit agencies (*Figure 1*). Relative to non-profit and commercial agencies, the government does not directly operate many child care centres. The highly skewed nature of the dataset may also be in part responsible for the differences in variances (*Figure 2*), despite the mean values being relatively equal across the groups.

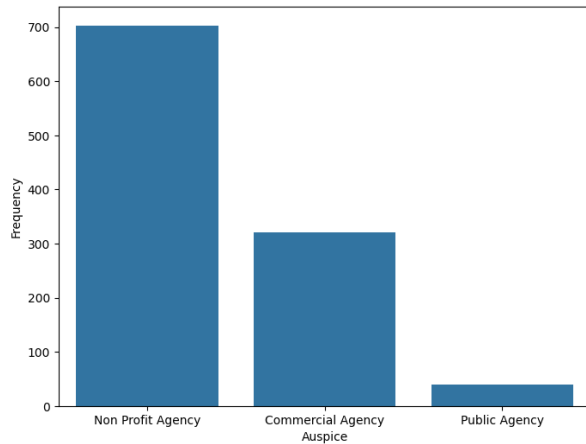


Figure 1: Bar graph of Auspice Frequencies

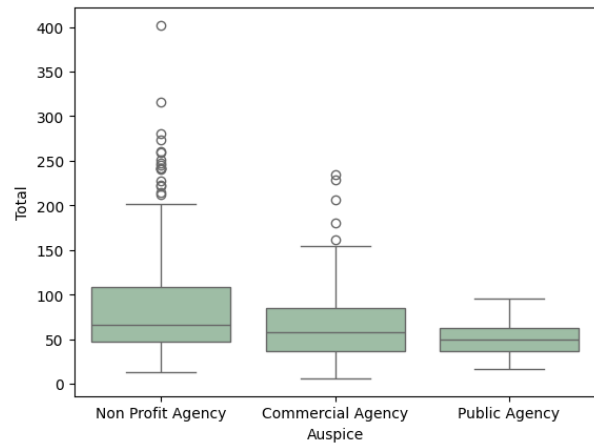


Figure 2: Boxplot of Total Capacity by Auspice

Examining the daily capacities for the different centres, the dataset categorises child care spaces by five age groups - infant, toddler, preschool, kindergarten, and school.

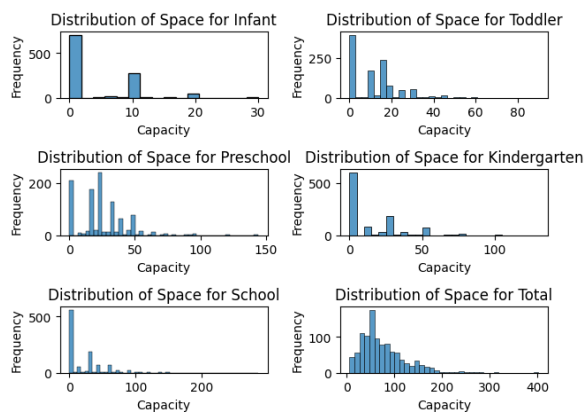


Figure 3: Histograms of Distributions of Capacities

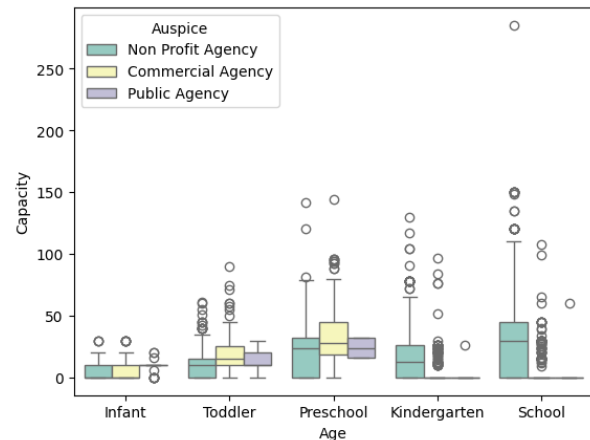


Figure 4: Boxplot of Capacity by Auspice and Age

In each of the distributions shown above (Figure 3), we observe that zero capacity is reported at a very high rate. This suggests that most services specialise in a subset of the age groups and do not accept children outside of their focus. In particular, we note that the distributions for infant, kindergarten, and school are sparser, indicating less availability at both ends of the age range.

Examining capacities by *Auspice* and *Age* (Figure 4), non-profit agencies show the greatest variation in capacity, with a notable number of outliers, suggesting a diverse range of service provisions. Commercial agencies tend to focus on toddlers and preschoolers, suggesting that these groups have the highest demand. Public agencies present the least data, but similarly focus on toddlers and preschoolers, perhaps as an attempt to address pressing needs.

3. One-Way ANOVA

RQ1: How does the total capacity of child care centres differ among various auspice types in Toronto?

To answer our first research question, we examine the *Total* reported capacity for each child care centre, grouped by *Auspice* type. To test for differences in capacity across the three types of *Auspices*, we perform a one-way ANOVA.

Source	df	SS	MS	F-value	p-value
<i>Auspice</i>	2	9.61e+04	48100	21.8	< 0.001
Residual	1060	2.33e+06	2200		

Table 1: Summary of One-way ANOVA for *Auspice* on Capacity

For a significance level of 0.95, the obtained p-value (Table 1) is lower than the alpha and so we can reject the null hypothesis that there is no statistically significant difference among the group means. To identify the specific differences, we perform post-hoc tests using Tukey HSD (Table 2).

<i>Auspice</i> A	<i>Auspice</i> B	Difference	Lower	Upper	q-value	p-value
Non-profit	Commercial	17.1	9.70	24.5	7.66	0.001
Non-profit	Public	34.3	16.2	52.4	6.29	0.001
Commercial	Public	17.2	-1.45	35.9	3.06	0.078

Table 2: Summary of Tukey HSD for *Auspice* on Capacity

For non-profit/commercial and non-profit/public our obtained p-values are lower than the threshold, indicating that we can reject the null hypotheses stating that there is no statistically significant difference between their capacities. Commercial/public resulted in a p-value of 0.078 and so we fail to reject the null hypothesis in this case. Non-profit organizations display significantly larger capacities than both commercial and government services.

To test the appropriateness of our approach we check the two assumptions for ANOVA:

Assumption 1: Residuals are normally distributed

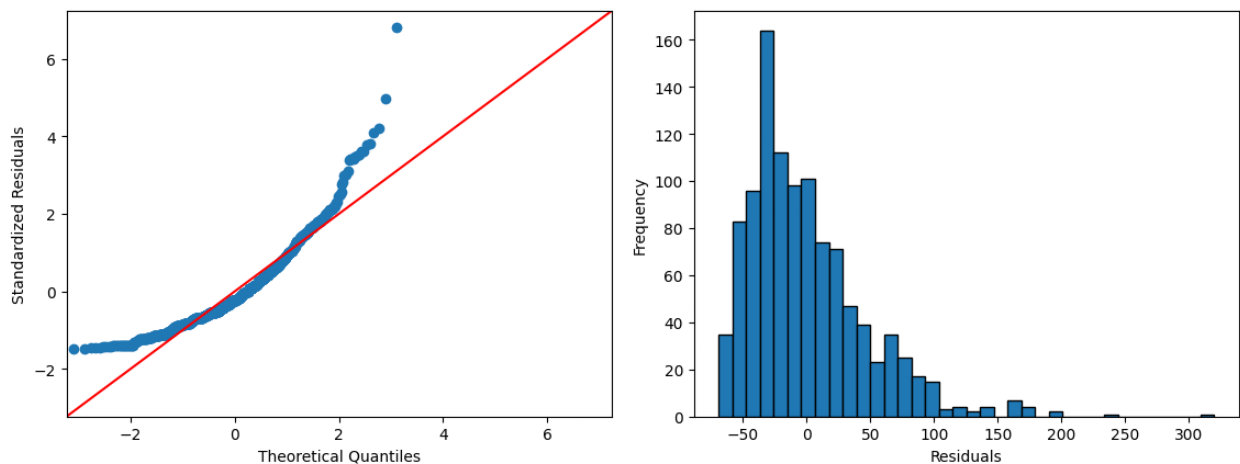


Figure 5: Plots of ANOVA Residuals for *Auspice* on Capacity

To check that the residuals from our one-way ANOVA (*Figure 5*) are normally distributed, we perform a Shapiro-Wilk test of normality. We obtain a test statistic (W) of 0.902 and a p-value < 0.001, indicating that we reject the null hypothesis of normality, failing the first assumption.

Assumption 2: Variances are homogenous

To test for homogeneity of variances we perform a Levene test. The Levene test was chosen over the Bartlett test as it is more robust to deviations from the normal distribution, which we observed in our exploratory data analysis. We obtain a test statistic (W) of 17.9 and a p-value < 0.001, indicating that we reject the null hypothesis of homogeneity, failing the second assumption.

4. Two-Way ANOVA

RQ2: How do the capacities for different age groups in child care centres vary across auspice types in Toronto, and how do the effects of age group and auspice type on capacity interact?

Having observed the effect of the type of *Auspice* on the capacity of a child care centre, we are now interested in considering *Age* as another factor. We perform a two-way ANOVA using *Auspice* and *Age* as categorical variables and *Capacity* as the continuous variable of interest.

Source	df	SS	MS	F-value	p-value
<i>Age</i>	4	2.82e+05	70500	213	< 0.001
<i>Auspice</i>	2	1.92e+04	9610	29.1	< 0.001
<i>Age:Auspice</i>	8	2.21e+05	27600	83.5	< 0.001
Residual	5300	1.75e+06	330		

Table 3: Summary of Two-way ANOVA for *Auspice* and *Age* on *Capacity*

The results of the two-way ANOVA (*Table 3*) suggest that there are significant differences in the capacities of child care centres based on both the type of *Auspice* and the *Age* group of the children. The test also suggests that there is a significant difference based on the levels of the two categorical variables, as reflected in the small p-value (< 0.001) for the interaction between *Age* and *Auspice*.

We can visualize these differences using an interaction plot of the effects of *Auspice* and *Age* on *Capacity* (*Figure 6*). We observe that while commercial and public agencies have generally similar lines, non-profit agencies display a very different interaction with *Age*. Commercial and public agencies peak drastically at preschool while maintaining relatively, flat, low levels for infant, kindergarten, and school. Both of these *Auspice* types also moderately rise at toddler. In contrast, non-profit agencies show high levels at preschool, kindergarten, and school, while dropping at infant and toddler, indicating a focus towards older age groups. The drastic differences at kindergarten and school especially highlight the considerable variance of the effect of *Auspice* on *Capacity* based on its interaction with *Age*.

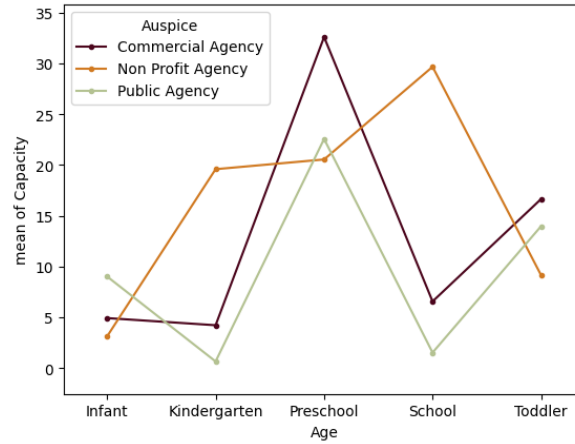


Figure 6: Interaction Plot of Auspice and Age on Capacity

Using Tukey HSD to conduct post-hoc tests based on the results of our two-way ANOVA, we see that there are significant differences in capacity across every pair of Age group (*Table 4*) as well as *Auspice* when considering the interaction of their effects.

Age A	Age B	Difference	Lower	Upper	q-value	p-value
Infant	Toddler	7.70	5.55	9.85	13.8	0.001
Infant	Preschool	20.4	18.2	22.5	36.5	0.001
Infant	Kindergarten	10.4	8.21	12.5	18.6	0.001
Infant	School	17.8	15.6	19.9	31.9	0.001
Toddler	Preschool	12.7	10.5	14.8	22.7	0.001
Toddler	Kindergarten	2.66	0.507	4.81	4.77	0.007
Toddler	School	10.1	7.91	12.2	18.1	0.001
Preschool	Kindergarten	10.0	7.85	12.2	17.9	0.001
Preschool	School	2.60	0.447	4.75	4.66	0.009
Kindergarten	School	7.40	5.25	9.55	13.3	0.001

Table 4: Summary of Tukey HSD for Age on Capacity

Auspice A	Auspice B	Difference	Lower	Upper	q-value	p-value
Non-profit	Commercial	3.42	2.14	4.71	8.84	0.001
Non-profit	Public	6.87	3.73	10.0	7.26	0.001
Commercial	Public	3.44	0.212	6.67	3.53	0.033

Table 5: Summary of Tukey HSD for Auspice on Capacity

We repeat the procedure we followed for our one-way ANOVA to test the assumptions for our two-way ANOVA.

Assumption 1: Residuals are normally distributed

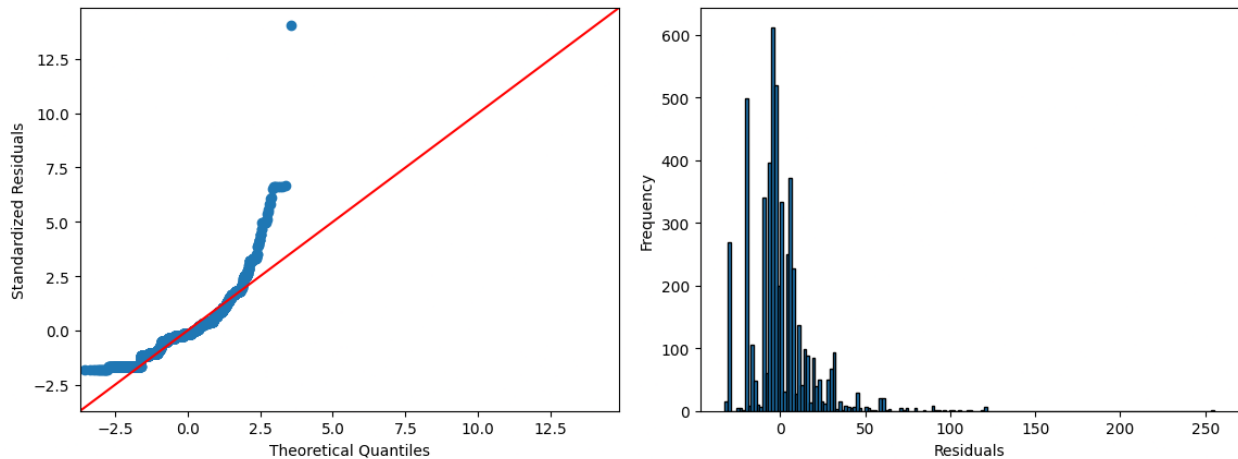


Figure 7: Plots of ANOVA Residuals for Auspice and Age on Capacity

We perform a Shapiro-Wilk test of normality for our two-way ANOVA residuals (Figure 7). We obtain a test statistic (W) of 0.860 and a p-value < 0.001 , indicating that we reject the null hypothesis of normality, failing the first assumption.

Assumption 2: Variances are homogenous

We again elect for the Levene test, obtaining a test statistic (W) of 114 and a p-value < 0.001 , indicating that we reject the null hypothesis of homogeneity, failing the second assumption.

5. Conclusion

Through our investigation of licensed child care centres in Toronto, we sought to investigate the level of service provided by different auspices and how these services vary based on the age group it caters to. We saw that non-profit organizations operate the majority of licensed child care centres in Toronto, mostly catered towards older children already attending some form of school. These demographics are largely ignored by commercial and government agencies. Given the focus on toddlers and preschoolers by for-profit organizations, there appears to be a large demand for services in this area. This theory is supported by the fact that the government also focuses on these groups, perhaps as a means to alleviate service gaps. Non-profits also operate at greater capacities, most likely leveraging existing resources and locations such as schools (which would be convenient given their target demographics).

As a limitation of our study, we note that our choice of ANOVAs was inappropriate given the violations of the assumptions and were employed simply for the purpose of this assignment. When considering accessibility, we should investigate the relation between supply and demand. The chosen dataset omits information on excess demand, such as waiting lists. Furthermore, our dataset only concerns licensed care centres. Given that our major motivation was based on the lack of accessibility to these centres, we should consider that people in need may turn to alternative services.