

# INF2178 Assignment 2

## Introduction

In response to the challenges many parents face in securing affordable and available childcare in Ontario, particularly in Toronto, where 75% of families report being unable to afford child care<sup>1</sup>, this report delves into an in-depth analysis of the factors influencing the capacity of licensed childcare centres in the city. Leveraging a dataset provided by the City of Toronto, which encompasses detailed information on the capacities, demographics, and classifications of all licensed childcare centres as of February 12, 2024, our investigation seeks to uncover the underlying dynamics affecting childcare availability. The provincial government's commitment to creating 100,000 new childcare spaces by 2026 highlights the urgency and relevance of this study. Driven by the objective *to understand the multifaceted factors affecting childcare centre capacities in Toronto*, our research is guided by several questions:

- **RQ1:** What impact do the auspices have on the capacity of childcare centres?
- **RQ2:** How do the availability of fee subsidies and the wards (location) influence the capacity of childcare centres?

Through this analysis, we aim to contribute valuable insights to policymakers, stakeholders, and families navigating the complexities of childcare in Toronto, ultimately facilitating more informed decisions and strategies to expand and optimize childcare services across the city.

## Data Cleaning and Wrangling

The dataset comprises 17 columns and 1,064 observations. An initial review indicates that further data cleaning is not required. Notably, the "Building Name" column contains 348 missing values. However, upon detailed inspection, it's clear that these missing entries are deliberate, representing buildings without names. Moreover, these missing values fall outside the scope of our analysis. For this assignment, we will focus exclusively on selected columns from the raw dataset. The table below provides a concise description of each feature (column) utilized:

Column Name	Description
AUSPICE	Operating auspice (Commercial, Non Profit or Public)
TOTSPACE	Child care spaces for all age groups
subsidy	Centre has a fee subsidy contract (Yes/No)
ward	City ward number

---

<sup>1</sup> [TORONTO'S LICENSED CHILD CARE GROWTH STRATEGY For children under 4 2017-2026](#)

## Exploratory Data Analysis

An exploratory data analysis (EDA) was conducted on the curated dataset to uncover any notable patterns or insights. We generated a statistical summary for the continuous variables, focusing on the capacities across various age groups within each childcare center (Table 1). Moreover, boxplots were crafted to depict the distribution of capacities for each age group (Figure 2). A significant observation from this analysis is the uneven distribution across age groups, with the capacity for children aged 0 to 18 months being notably lower than for other age groups.

	0-18	18-30	30+	Full-day	Grade 1+	Total
count	1063.0	1063.0	1063.0	1063.0	1063.0	1063.0
mean	3.9	11.6	24.3	14.3	21.7	75.7
std	6.1	12.1	18.6	20.5	30.4	47.8
min	0.0	0.0	0.0	0.0	0.0	6.0
25%	0.0	0.0	16.0	0.0	0.0	43.0
50%	0.0	10.0	24.0	0.0	0.0	62.0
75%	10.0	15.0	32.0	26.0	30.0	97.0
max	30.0	90.0	144.0	130.0	285.0	402.0

Table 1, Stats Summary Table

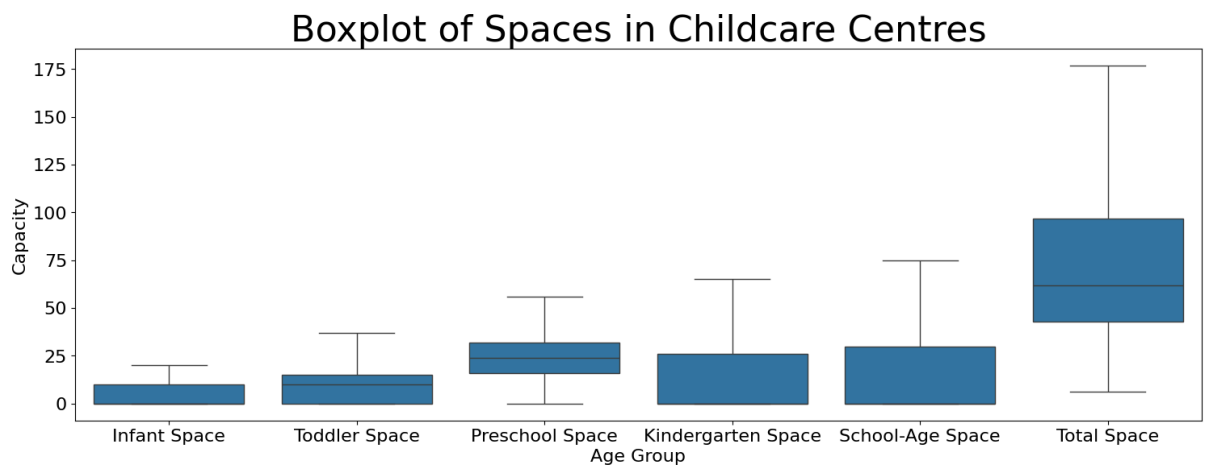


Figure 2, Boxplot for total capacities of childcare centres

Our investigation primarily explores the relationships between each childcare center's total capacity and factors such as auspice, location, and fee subsidy status. To understand the distribution patterns, we plotted a histogram of the total capacities. Although the data exhibits a rightward skew, the large sample size enables us to rely on the Central Limit Theorem (Figure 3). This theoretical framework assures that the normality assumption required for ANOVA tests is satisfied, allowing us to proceed with our analysis confidently.

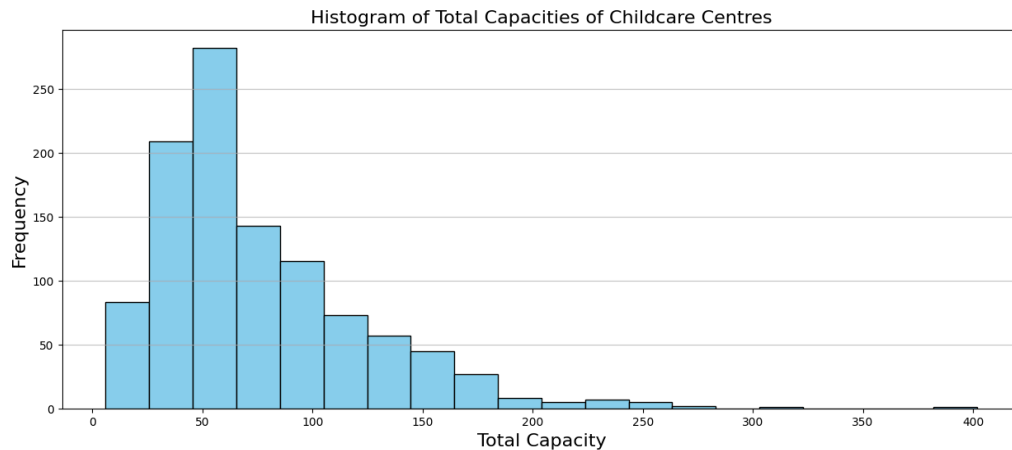


Figure 3, Histogram of the total capacities of the centres

## Total Capacity and Auspice

Childcare centres can be divided into 3 categories based on their auspices, which are non-profit, commercial, and public. We would like to investigate the effects of auspices on the total capacity of childcare centres. To begin with, we started by looking at the count of childcare centred by auspice and then plotted the total capacities of the centres by auspice in Figure 4. Public agencies are the least in numbers and they have a smaller total capacity relative to the other two agencies, which could explain why parents find it difficult to find available and affordable childcare centres.

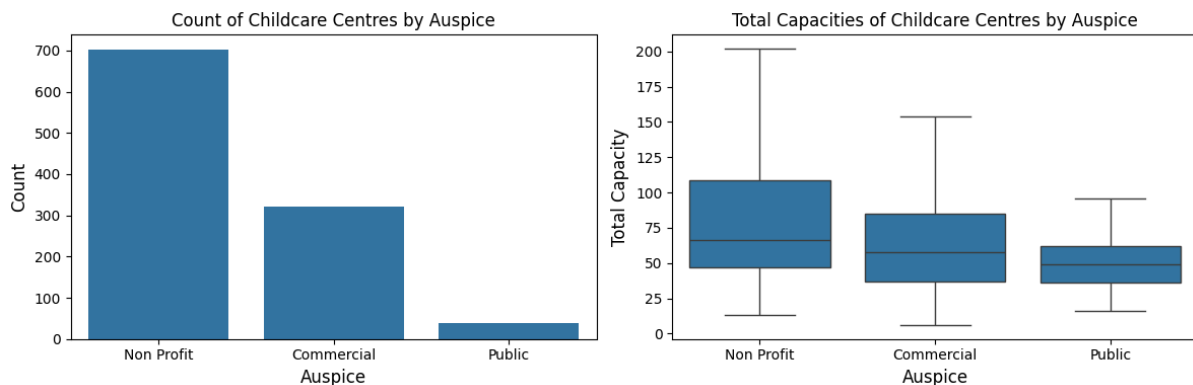


Figure 4

Next, a one-way ANOVA test is conducted, and the hypotheses are as follows:

- H0: There is no significant difference in the mean total spaces among different auspices.
- H1: There is at least one significant difference in the mean total spaces among the different auspices.

The ANOVA yielded a p-value of 5.057716e-10, significantly below the alpha threshold ( $\alpha = 0.05$ ), indicating a statistical difference in total capacities across childcare center auspices (Table 5). Further, Tukey's HSD test highlighted a significant disparity in the total capacities of non-profit agencies compared to commercial and public ones (Table 6).

	df	sum_sq	mean_sq	F	P-value
<b>C(Auspice)</b>	2	9.61e+04	48056.06	21.84	5.06e-10

<b>Residual</b>	1060	2.33e+06	2200.06		
-----------------	------	----------	---------	--	--

Table 5, One Way ANOVA Table

Group 1	Group 2	Diff	Lower	Upper	q-value	p-value
Non-profit	Commercial	16.807	3.994	29.619	4.357	0.006
Non-profit	Public	36.178	8.674	63.682	4.369	0.006
Commercial	Public	19.371	-10.142	48.885	2.180	0.273

Figure 6, Tukey HSD Table

However, the ANOVA's residuals failed to meet normality ( $p=0.0015$ ) and homogeneity of variances ( $p=0.0001$ ) assumptions, as depicted by the Q-Q plot and histogram of residuals (Figure 7). This limitation prompted the use of Welch's ANOVA for a more robust analysis, which corroborated the initial findings by rejecting the null hypothesis ( $p=0.0396$ ) (Table 8). Thus, we conclude that childcare center auspice significantly impacts total capacity, addressing our first research question (RQ1).

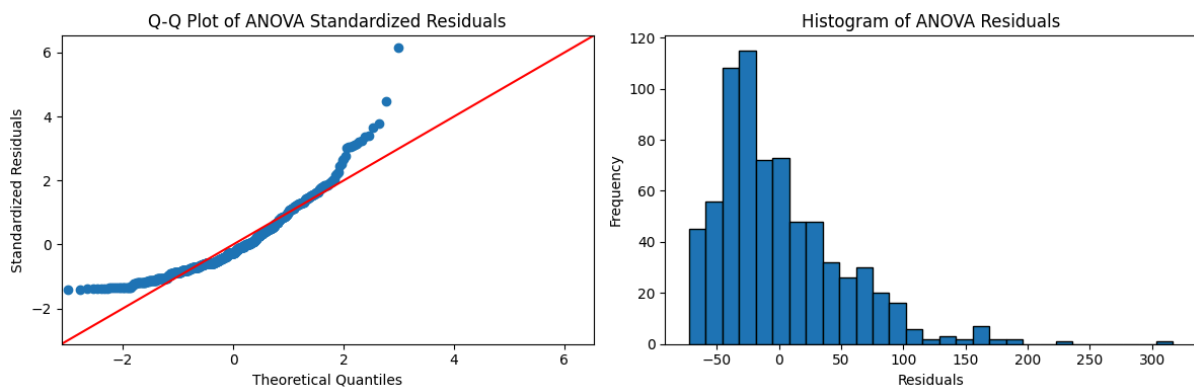


Figure 7

Source	ddof1	ddof2	F	p-unc	np2
Auspice	2	142.886	51.632	1.33e-17	0.0396

Table 8, Welch's ANOVA Table

## Fee Subsidy and Ward

Acknowledging the considerable variance in average income and population among Toronto's wards, and the established influence of family income and neighbourhood environment on childcare centers, we delve into our second research question: How do fee subsidy and location impact the total capacities of childcare centers? The city's dataset, enriched with geographical information, serves as our foundational tool particularly focusing on wards to understand location-based impacts on childcare capacities. This exploration not only aims to address immediate inquiries but also to guide further, more granular location analyses as deemed necessary.

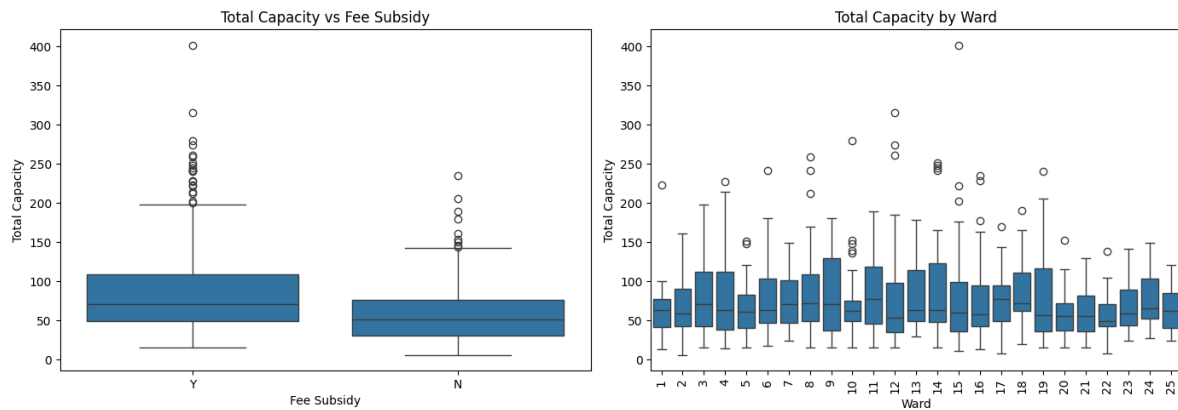


Figure 9, Q-Q plot and histogram of ANOVA residuals

Initial visualizations via box plots contrasting total capacities against fee subsidy enrollment and ward highlight a predominant participation in fee subsidy programs and notable capacity variances across wards (Figure 9). However, these visual aids alone do not suffice to discern potential interaction effects between the categorical factors on total capacities. Thus, we proceed with a two-way ANOVA, articulated through the following hypotheses:

- $H0_1$ : The mean total capacity of childcare centers does not differ significantly between centers that are enrolled in fee subsidy programs and those that are not.
- $H1_1$ : There is a significant difference in the mean total capacity of childcare centers between those that are enrolled in fee subsidy programs and those that are not.
- $H0_2$ : The mean total capacity of childcare centers does not differ significantly across wards.
- $H1_2$ : At least one ward has a mean total capacity that is significantly different from the others.
- $H0_3$ : There is no interaction effect between ward and fee subsidy enrollment on the total capacity of childcare centers.
- $H1_3$ : There is an interaction effect between ward and fee subsidy enrollment on the total capacity of childcare centers.

	df	sum_sq	mean_sq	F	P-value
<b>C(subsidy)</b>	1	2.28e+05	227852.67	111.72	<b>7.67e-25</b>
<b>C(ward)</b>	24	1.46e+05	6070.22	2.98	<b>2.37e-06</b>
<b>C(subsidy):C(ward)</b>	24	5.56e+04	2318.50	1.14	<b>2.95e-01</b>
<b>Residual</b>	1060	2.33e+06	2200.06		

Table 10, two-way ANOVA results

With a significance level set at 0.05, rejection of all three null hypotheses indicates that fee subsidy, ward, and their collective factors significantly influence childcare center capacities. The interaction plot further illustrates specific wards where fee subsidy and ward interactions markedly affect capacities, evidenced by converging or intersecting lines (Figure 11).

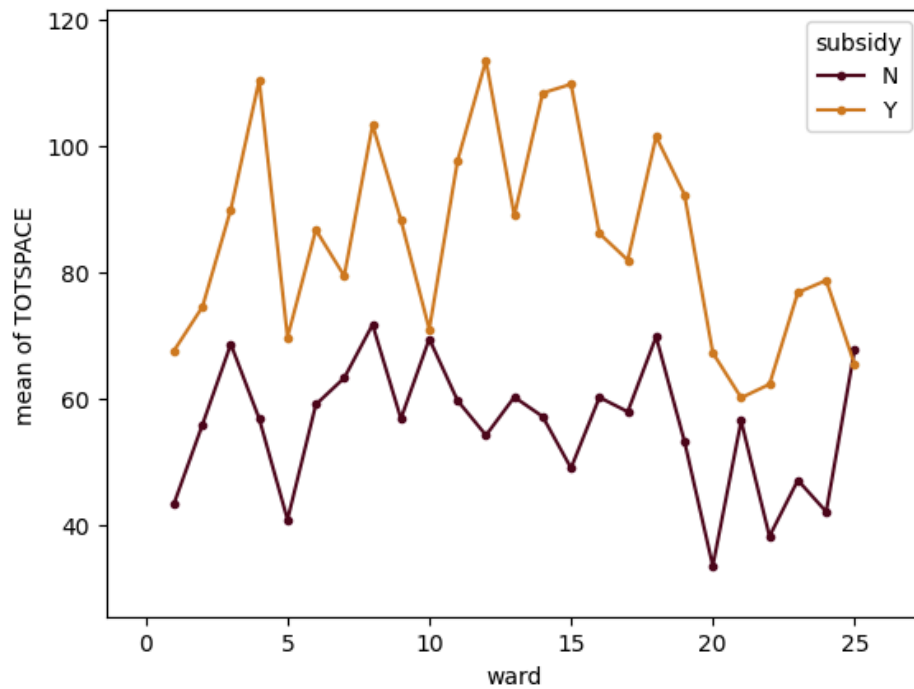


Figure 11, interaction plot

Post hoc analysis via Tukey's HSD test postulates a strong influence of fee subsidy enrollment on capacities, as supported by a notable p-value of 0.001. Conversely, pairwise comparisons between wards and capacities predominantly yield p-values of 0.9, indicating an absence of significant differences. This discrepancy extends to the interaction effects, where null hypotheses concerning interactions between ward and fee subsidy on capacities remain unchallenged due to consistent p-values of 0.9.

The foundational two-way ANOVA's reliability is questioned following failures to meet normality and homogeneity assumptions, as verified by Shapiro-Wilk and Levene's tests (Table 12). While fee subsidy enrollment's impact on capacities is confirmed, the influence of ward and its interaction with fee subsidy necessitates further scrutiny. Before concluding the roles of ward and its synergistic effects with fee subsidy on childcare capacities, additional analytical rigor is essential.

	W-value	P-value
Shapiro-Wilk	0.94	4.29e-21
Levene forward	2.48	0.0001
Levene for <i>subsidy</i>	22.99	1.86e-06

Table 12, Assupetion check results