TECHNICAL

ASSIGNMENT 2

A Quantitative Look at

Toronto's Child Care Centers

INF2178

SUBMITTED TO: Dr. Shion Guha

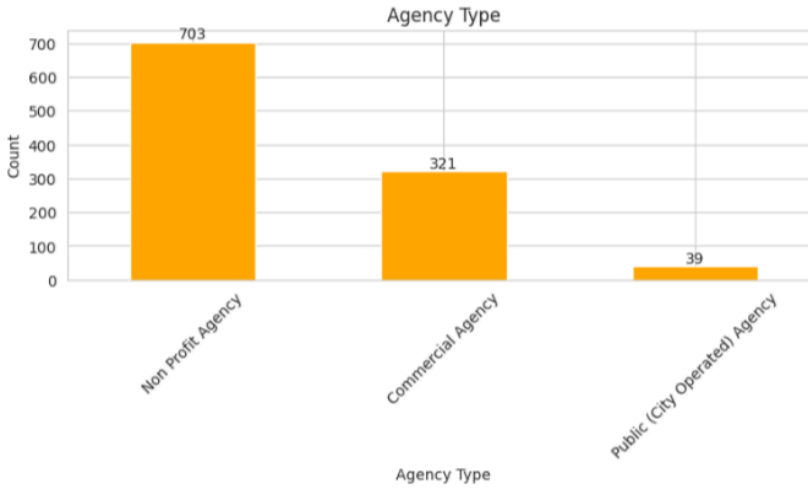
SUBMITTED BY: Adil Mahmood Ashraf

Student Number: 1010727017

1. Introduction

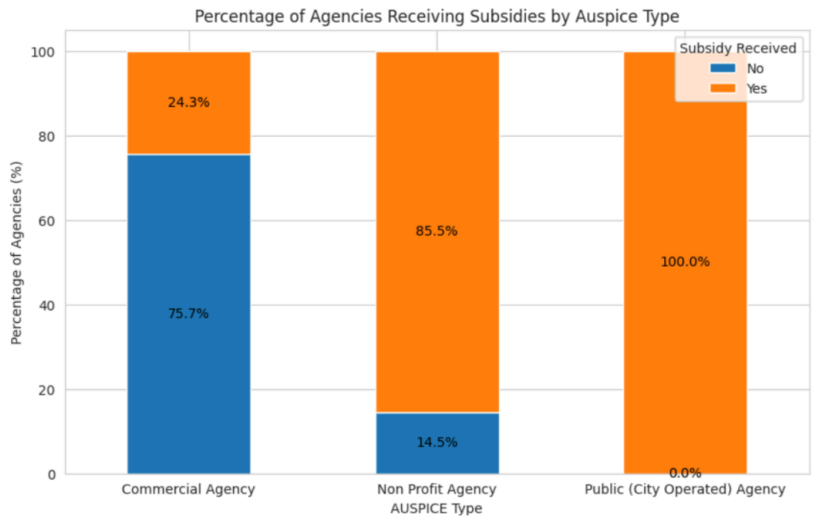
As a father of twin infants, who recently moved Toronto to attend the Master of Information program, the hearsay circulating among families pertaining to the lack of adequate child care services has got me a little concerned. As I came across a data set for technical assignment 2, the opportunity to be able to decipher the data and make an informed decision on where to look for affordable, and good quality child care was somewhat exciting.

<u>Figure 1:</u> Count of Childcare Services by Agency Type



Does the variety of child care facilities—classified as either commercial, non-profit, or public—play a critical role in defining accessibility and quality of child care?
To answer this question, parents need to comprehend how various factors, including agency type (defined as Auspice in this dataset) and subsidy provision, impact the capacity of child care centers.
Understanding the interplay of these variables is crucial to build a robust child care system that caters comprehensively to family needs. As a father, having clarity on the relationship between these variables is also going to help me put in child care applications in the centers where I am more likely to get a spot for my children.

<u>Figure 2:</u> Breakdown of Agencies Receiving Subsidies



Toronto has a total of 1063 child care agencies which are made up of 703 non-profit, 321 commercial and 39 public agencies. Out of these agencies 100% of public agencies, 85.5% of non-profit agencies and 24.3% of commercial agencies have been recipients of some form of a subsidy.

2. Research Questions

My analysis is structured around the following research questions that would be most beneficial to student-parents who are looking for child care centers in Toronto:

- **Research Question 1:** Is there a discernible variation in the total space of child care centers based on their Auspice?
- **Research Question 2:** What is the role of subsidy in conjunction with agency types (Auspice) in determining the capacity of child care centers?

3. Data Cleaning, Selection and Data Wrangling

This report presents a quantitative statistical analysis of the capacities of child care centers across different Auspices within Toronto. No new columns or variables were added for data analysis. However, while performing two-way ANOVA, I did manipulate the data in python to drop NaN values; convert Convert 'AUSPICE' and 'subsidy' to categorical data types; and Convert 'TOTSPACE' to a numeric type.

A- Quantitative Variables:

With the analysis firmly quantitative in nature, I am going to be focused on key columns in the dataset titled "INF2178_A2_data.xlsx" that provide measurable data, enabling the precise application of statistical tests to yield meaningful insights.

- **TOTSPACE**: The quantitative variable TOTSPACE represents the total licensed capacity of child care centers, this variable is pivotal for our quantitative assessment and is central to answering Research Questions 1 and 2.

B- Categorical Variables:

Integral to our ANOVA comparisons are categorical variables which add depth to our statistical exploration. I chose to convert AUSPICE and Subsidy into categorical variables using Python code for Two-Way ANOVA.I also dropped all the NaN values before proceeding to perform the Two-Way ANOVA. However, the variables remained untouched for one-way ANOVA.

- **AUSPICE**: This variable delineates the operating auspice of the child care center into Commercial, Non-Profit, or Public categories, a classification crucial for addressing Research Question 1 and instrumental for the two-way ANOVA in Research Question 2.
- **Subsidy**: The subsidy column is classified simply as 'Yes' or 'No'. This variable indicates a center's receipt of government subsidies and is crucial for investigating potential interactions with Auspice in our analysis for Research Question 2.

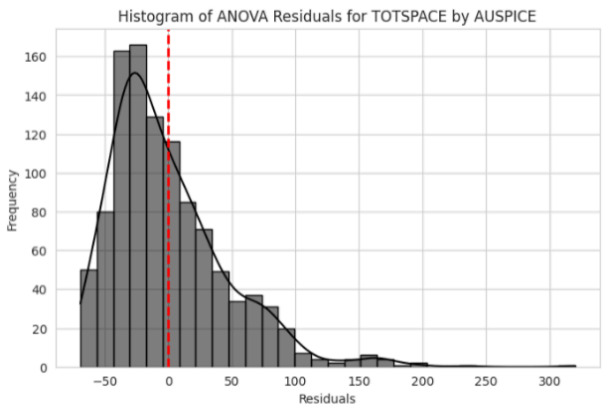
3. Answering Research Question 1 using One-Way ANOVA: Is there a discernible variation in the total space of child care centers based on their Auspice?

Let's begin our quest to answer this question by stating our null hypothesis.

Ho: There is no significant difference in the mean total space of child care centers among different Auspice types.

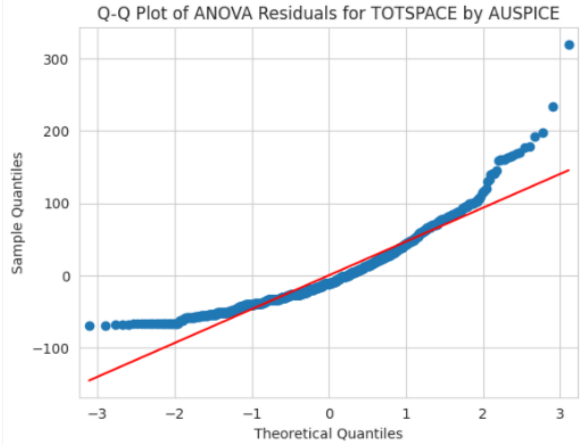
To address our first research question regarding the variation in total child care spaces based on Auspice type, I employed a One-Way ANOVA, a robust statistical method suited for comparing means across multiple groups. Prior to conducting the ANOVA, I tested the underlying assumptions essential for its valid application.

<u>Figure 3:</u> Histogram of Residuals



<u>A- Residuals:</u> The residuals from the ANOVA model, as shown in Figure 3, capture the differences between observed values and those predicted by the model, and were scrutinized for normality. The Q-Q plot in Figure 4 illustrates a deviation from the expected normal line (red), particularly at the tails, indicating potential outliers or a non-normal distribution. The histogram of residuals, with the overlaid normal distribution curve, further reveals a skew in the data, corroborated by the Shapiro-Wilk test suggesting a violation of the normality assumption. The QQ plot for normality also seemed to violate normal distribution.

<u>Figure 4:</u> QQ Plot for Normality



<u>Normality:</u> The Shapiro-Wilk test was used to assess the normality of the distribution of total spaces within each category of Auspice. The results showed that for Commercial Agency the p-value was 1.0e-11, for Non-Profit Agency it was 8.3e-22, and for Public (City Operated) Agency it was 9.0e-10, indicating a potential deviation from normality in our data.

<u>B- Variances:</u> For examining homogeneity of variances—an assumption that variances are

equal across groups—I implemented Levene's Test. The test yielded a statistic of 17.93 and a p-value of 2e-8, indicating that the variances are not equal across the groups.

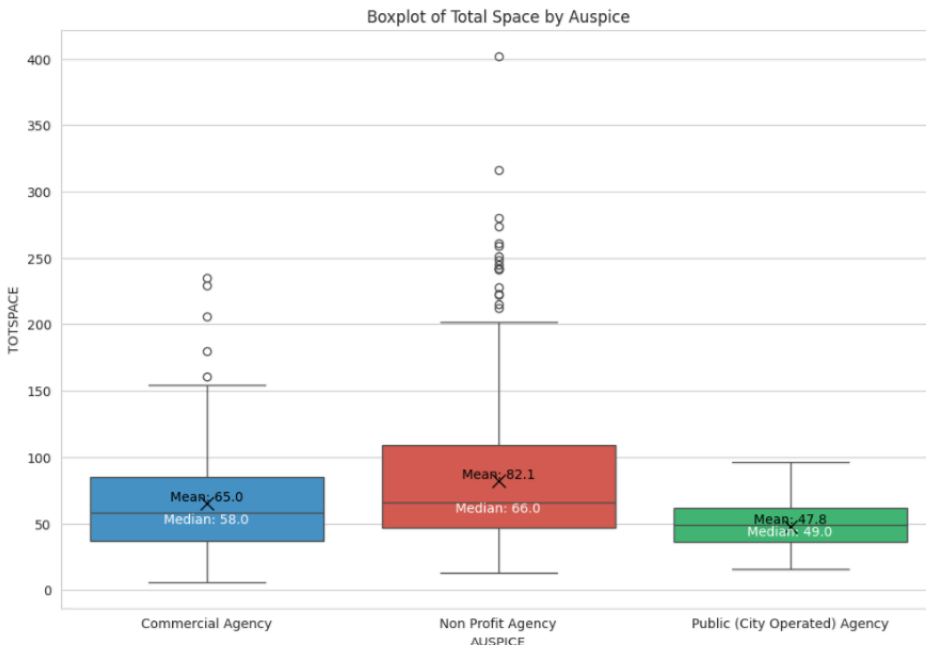
C- Post Hoc Testing: The primary phase of our analysis with a one-way ANOVA revealed that Non-Profit and Commercial agencies differ significantly in their total space, underscoring the profound impact of organizational models on child care capacity.
Upon confirming the presence of significant differences in the ANOVA, with a F-statistic of 21.84 and a p-value of less than 0.001 from **Table 1**, I proceeded with Post Hoc testing to explore pairwise differences between Auspice groups. Given the significant F-statistic of 21.84 and a p-value far below the alpha threshold in our one-way ANOVA results, I moved forward with post hoc analyses to pinpoint the specific Auspice groups among which differences occur. The substantial p-value indicates that not all Auspice groups offer the same total space in child care centers, necessitating a closer look to inform targeted improvements within these services.

D- Conclusion : Based on these insights, we are likely to reject the null hypothesis, indicated by the significant F-statistic of 21.84 and the p-value of less than 0.001 obtained from the one-way ANOVA. Therefore, I would reject the null hypothesis and conclude that there is a discernible variation in the total space of child care centers based on their Auspice.

Table 1: Values for One Way ANOVA

	sum_sq	df	f	PR(>F)
C(AUSPICE)	96112.11e+04	2	21.84	5.057716e-10
Residual	2332065.26e+06	1060.0	NaN	NaN

Figure 5: Boxplot of Total Child Care Spaces by Auspice Type



4. Answering Research Question 2 using Two-Way ANOVA: What is the role of subsidy in conjunction with agency types (Auspice) in determining the capacity of child care centers?

We begin our search for the answer to this question by stating the alternative hypothesis for this question (H1).

H1 : There is a significant interaction between Auspice type and subsidy status on the total space capacity of child care centers.

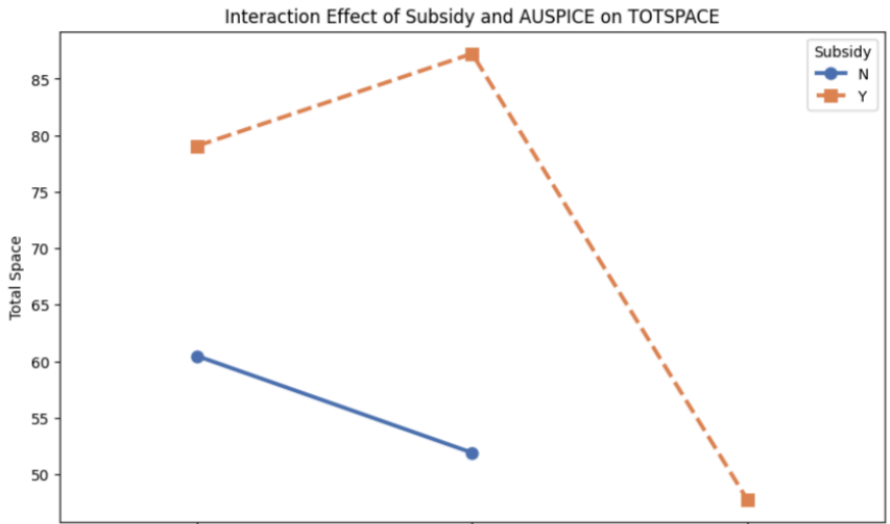
To approach our second research question, which probes into the combined effect of Auspice type and the presence of government subsidies on total child care spaces, I conducted a Two-Way ANOVA. This sophisticated statistical analysis enables us to discern not only the individual impacts of each factor but also whether there is an interaction between them that could influence the total space capacity of child care centers.

A- Normality: As with our previous analysis, I began by validating the assumptions critical to the Two-Way ANOVA. Please note that these values have been derived from the output in the .ipynb file. The Shapiro-Wilk test for normality indicates that:

- Commercial Agencies have a test statistic of 0.92 with a p-value of approximately 1.0e-11, suggesting a potential deviation from normality.
- Non-Profit Agencies have a test statistic of 0.89 and an extremely low p-value of roughly 8.3e-22, reinforcing this deviation.
- Public (City Operated) Agencies show a test statistic of 0.95 with a p-value of about 9.0e-10, which also suggests a likely departure from the normal distribution.

B-Homogeneity: Levene's Test for Homogeneity of Variance yields a statistic of 17.93, with a p-value of approximately 0.000000020, indicating a significant difference in variances across these groups and thus a violation of the assumption of homogeneity of variances.

Figure 6: Visualization of Two Way ANOVA Showcasing the Interaction Effect of Total Child Care Spaces by Auspice Type and Subsidy Status



C-Visualization of Interaction Plot:

Our investigation revealed significant interactions between Auspice type and subsidy

status on child care center capacities. Although the Two-Way ANOVA faced challenges with normality and homogeneity assumptions, the results still provide valuable insights. We identified a substantial main effect of subsidies and a nuanced interaction with Auspice types, indicating that subsidies do not uniformly affect all agency types. The interaction plot illustrates that the influence of subsidies on the available space in daycare facilities varies by agency type. Non-Profit Agencies are noted to experience a significant boost in total spaces when receiving subsidies, while Commercial Agencies show a slight reduction. Notably, Public Agencies display a marked decline in spaces when subsidized, suggesting that while subsidies appear to enhance the capacity of Non-Profit daycares, they might inversely impact Public ones. These findings highlight the complexity within child care resource allocation and suggest the need for nuanced policy development that considers the distinct needs of different Auspice categories.

D- Post Hoc Test and Conclusion:

Table 2: Multiple Comparison of Means

group1	group2	meandiff	p-adj	lower	upper	reject
Commercial Agency	Non Profit Agency	17.1194	0.0	9.7037	24.5351	True
Commercial Agency	Public (City Operated) Agency	-17.2152	0.0779	-35.8832	1.4528	False
Non Profit Agency	Public (City Operated) Agenc	-34.3346	0.0	-52.4448	-16.2244	True

The Tukey's HSD (Honestly Significant Difference) test, a commonly used Post Hoc test following ANOVA, provided further insights. It confirmed significant differences in the mean total space between Commercial Agency and Non-Profit Agency (mean difference 17.12, p-adj < 0.001), and between Non-Profit Agency and Public (City Operated) Agency (mean difference -34.33, p-adj < 0.001), but not between Commercial Agency and Public (City Operated) Agency (mean difference -17.22, p-adj = 0.0779). This comprehensive approach allowed us to ensure the robustness of our findings and to draw well-founded conclusions about the impact of Auspice on the total capacity of child care centers, thus failing to reject the null hypothesis.

5. Conclusion of the Study:

The study revealed that Non-Profit and Commercial agencies significantly differ in their total space, emphasizing the profound impact of organizational models on child care capacity. Additionally, the investigation into the combined effect of Auspice type and subsidy status indicated significant interactions, with subsidies not uniformly affecting all agency types. While subsidies were found to enhance the capacity of Non-Profit agencies, they had varying effects on Public agencies, even potentially reducing their capacity suggesting the need for nuanced policy development that considers the distinct needs of different agency types.