Explore Child Care Center Usage Rate In Toronto

1. Introduction:

Securing both licensed and unlicensed child care in Ontario, especially in Toronto, can be very challenging for many families, largely due to high costs and insufficient spaces. In response to this critical issue, the provincial government has promised to create 100,000 new spaces by 2026, addressing the affordability crisis faced by 75% of families.

This report will offer a comprehensive data analysis on Childcare capacity in Toronto. We will use the dataset called “INF2178_A2_data.xlsx”, which highlights the diversity of Toronto's childcare facilities in terms of their management (non-profit agencies and other types), locations, and capacity to serve infants, toddlers, preschoolers, kindergarteners, and school-aged children.

Our analysis will only focus on the childcare capacity condition for toddlers who are 18-30 months old (called toddlers from below), which aims to enhance specialized operation and management in the city's childcare landscape, ultimately contributing to the broader goal of making childcare more accessible and affordable for Toronto's families. Our research questions are:

- Are there any significant differences in capacity rates for toddlers who are 18-30 months old **(TGSPACE_CAPACITY)** of these institutions or centers under different operating types **(AUSPICE)** (e.g., non-profit, commercial, public, etc.)?
- Are there any significant differences in the capacity for toddlers who are 18-30 months old **(TGSPACE_CAPACITY)** provided by different operation **(AUSPICE)** types based on their subsidy status**(subsidy)**?

2. Data Cleaning:

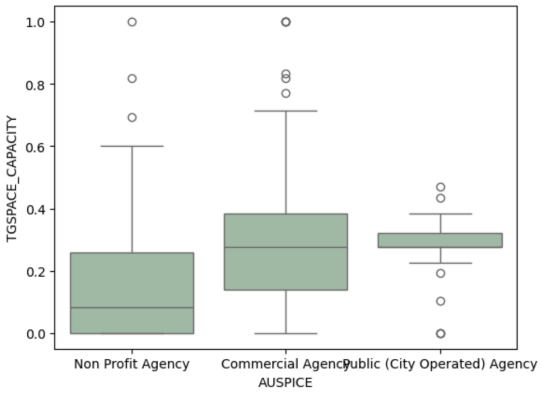
This dataset contains a total of 17 columns and 1063 rows, we will only use 4 of these columns and create a new one to calculate the capacity for our analysis. Below are the descriptions for each column:

- AUSPICE: Operating auspice (Commercial, Non Profit or Public)
- TGSPACE: Child care spaces for toddlers 18-30 months
- TOTSPACE: Child care spaces for all age groups
- Subside: Centre has a fee subsidy contract (Yes/No)
- TGSPACE_CAPACITY: Capacity rates for toddlers who are 18-30 months old.
    - Calculated by: TGSPACE/TOTSPACE

Upon checking, there are no missing values in the dataset so we can start our analysis.

3. One-Way ANOVA Test

We first create a boxplot to visualize a comparative view of the capacity rate (TGSPACE_CAPACITY) in childcare centers across different types of operating agencies (AUSPICE): Non-Profit, Commercial, and Public (city-operated).



The box plot illustrates the distribution of space allocated for toddlers with total space across three types of childcare agencies. Non-profit agencies show a broad range and generally lower median capacity for toddlers, while Commercial and Public (city-operated) Agencies display less variability and higher median capacities. Notable outliers suggest that some agencies deviate significantly from the typical capacity rates within their categories.

After looking at the boxplot, we will look at the one-way ANOVA statistics result:

index	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	4.358882	2.179441	78.267943	1.977e-32
Residual	1060.0	29.51665	0.027846	NaN	NaN

The ANOVA table indicates that the three different operating types significantly affect the childcare toddler capacity, with a very low p-value (1.977e-32) suggesting this is unlikely to be due to random chance. The F-statistic of 78.27 further confirms the importance of the type of childcare center in explaining the variance in toddler capacity.

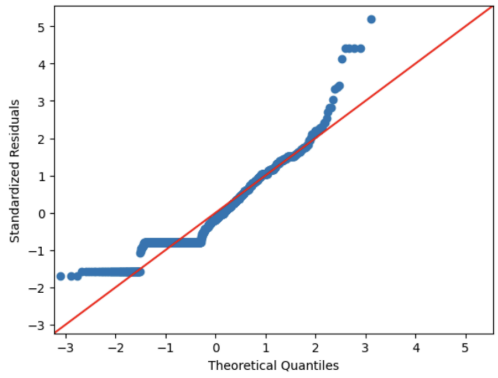
Next, we will use Tukey's Honestly Significant Difference (HSD) test, which is used to compare means between groups after an ANOVA to determine which means are significantly different from each other.

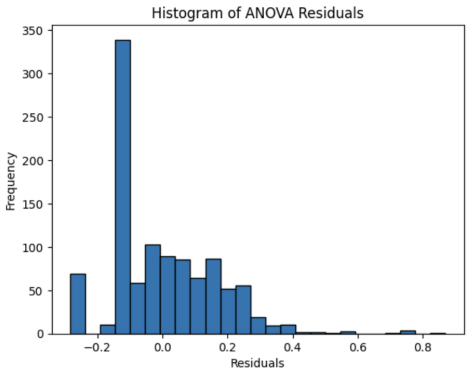
index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	0.13313736803580084	0.10675448670615052	0.15952024936545117	16.749990255496982	0.001
1	Non Profit Agency	Public (City Operated) Agency	0.15149839363041184	0.08706746199948567	0.215929325261338	7.804596097565 4145	0.001
2	Commercial Agency	Public (City Operated) Agency	0.01836102559 4611002	-0.0480543839 5648644	0.08477643514 570844	0.91762428023 2225	0.77254389099 28157

The Tukey HSD test shows that Non-Profit Agencies have significantly lower toddler capacity rates compared to Commercial Agencies (difference of 0.133, p-value = 0.001) and Public Agencies (difference of 0.151, p-value = 0.001). These differences are statistically significant, indicating that Non-Profit Agencies allocate a smaller proportion of space to toddlers. Conversely, the capacity rates between Commercial and Public Agencies are not significantly different (difference of 0.018, p-value = 0.772), suggesting similar space allocation practices between these types of agencies. The significant results imply that agency type influences toddler space allocation.

Now, we need to check the assumptions for one-way ANOVA:

1. Normality Assumption





- The Q-Q plot shows that the standardized residuals do not closely follow the theoretical line, indicating potential deviations from normality, particularly with the presence of outliers.
- The histogram of residuals is skewed, suggesting the data may not be normally distributed, which is another indication that the assumption of normality may be violated in the ANOVA model.
- From the Shapiro-Wilk test for the normality, the w statistics is approximately 0.921, and the p-value is smaller than 0.001, which also supports the normality assumption is violated.

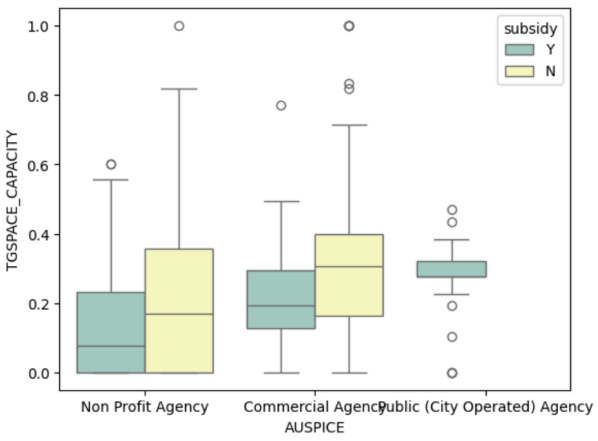
2. Equal variance assumption

We use Levene's test to check the homogeneity when the sample is not normally distributed, and we get a p-value that is smaller than 0.05, which means there is a significant discrepancy in variance in this data group. Thus the assumption of equal variance is violated.

Since the assumptions of one-way ANOVA are not met, the results may be unreliable. This can lead to incorrect conclusions, such as falsely finding a statistically significant effect on the capacity for different agency types.

4. Two-Way ANOVA Test

Same as what we did previously, we will use a boxplot to visualize the distribution of toddler space capacity (TGSPACE_CAPACITY) across different types of childcare agencies (AUSPICE), and it differentiates these distributions based on whether the agency receives a subsidy (Y for yes, N for no).



Non-profit agencies show a wide range of CAPACITY, with no clear difference between subsidized and non-subsidized facilities. Commercial Agencies exhibit a higher median capacity in the absence of subsidies than the other two types. Public (City-Operated) Agencies display the least variation in capacity, with non-subsidized agencies having higher median capacities than their subsidized counterparts. There are outliers in both Non-Profit and Commercial Agencies, which indicates that there are some facilities with capacities that are unusual. These observations suggest that the presence of subsidies does not consistently correlate with higher capacities across all types of agencies.

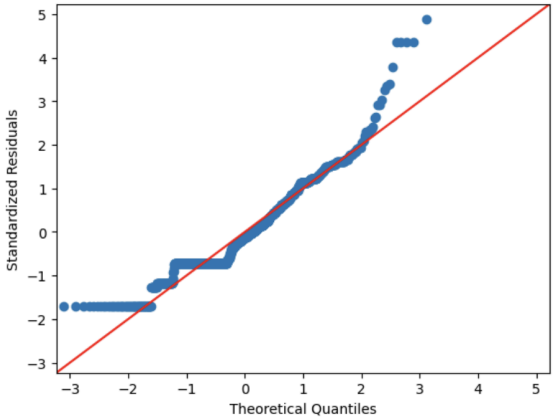
After looking at the boxplot, we will look at the two-way ANOVA statistics result:

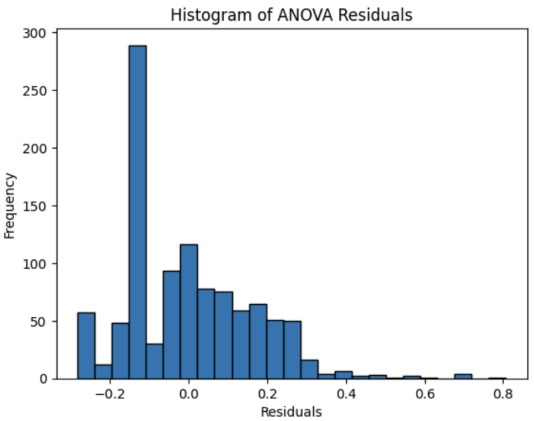
index	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	1.80576235279101	0.902881176395505	33.214	1.023e-14
C(subsidy)	1.0	0.605849973964890 6	0.605849973964890 6	22.287	2.663e-06
C(AUSPICE):C(subsidy)	2.0	0.162435061080 11316	0.081217530540 05658	2.988	0.051
Residual	1058.0	28.760089459189878	0.027183449394319357	NaN	NaN

The two-way ANOVA results highlight significant effects on toddler space capacity from both agency type (AUSPICE) and subsidy status, evidenced by F-values of 33.21 and 22.23, with corresponding p-values well below 0.05. The interaction between these factors has an F-value of 2.98 and a p-value of 0.05, which indicates a possible interaction that should be interpreted with caution. Residual variance remains (sum of squares = 28.76), suggesting there are other unaccounted factors affecting capacity. Overall, these results show that both the type of childcare agency and the presence of subsidies are important determinants of the space allocated to toddlers.

Now, we need to check the assumptions for Two-way ANOVA:

1. Normality Assumption





The assumption of normality is not met as evidenced by the Q-Q plot showing substantial deviation from the line and the Shapiro-Wilk test yielding a highly significant p-value(5.191), indicating the residuals do not follow a normal distribution.

2. Equal variance assumption

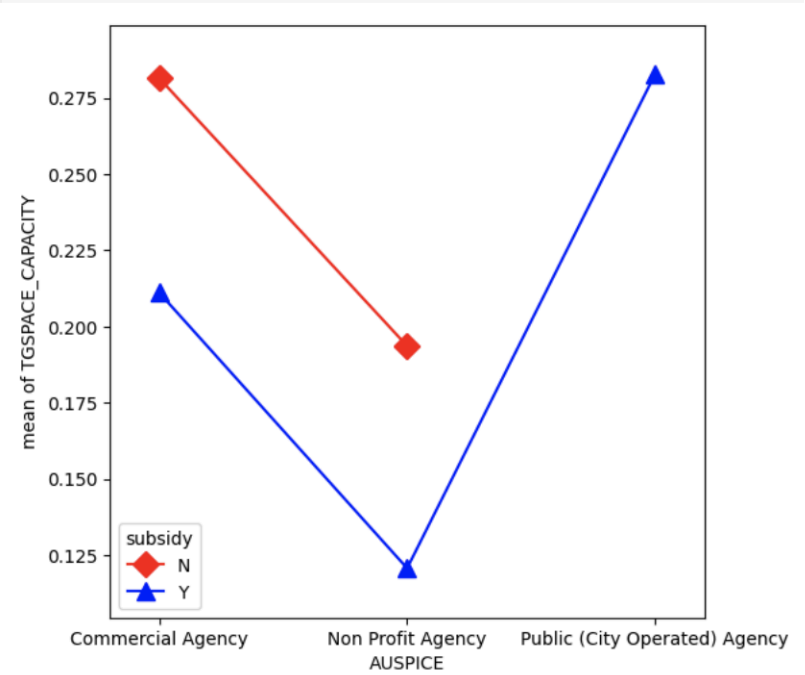
We use Levene's test to check the homogeneity of variance, results with a statistic of 14.23 and a highly significant p-value <0.01 indicate that the assumption of equal variances across groups is violated for the ANOVA model.

We now look at the Tukey's Honestly Significant Difference (HSD) test

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency,Y	Non Profit Agency,N	0.072954155 20233192	0.022547705 47645476	0.123360604 92820908	5.843491377 134526	0.001
1	Non Profit Agency,Y	Commercial Agency,Y	0.090446489 20131326	0.033796984 24431655	0.147095994 15830998	6.446204506 071117	0.001
2	Non Profit Agency,Y	Commercial Agency,N	0.160823396 7525002	0.125040407 3544212	0.196606386 1505792	18.14600789 454597	0.001
3	Non Profit Agency,Y	Public (City Operated) Agency,Y	0.162083491 54028076	0.084303799 95583946	0.239863183 12472205	8.413584317 837179	0.001
4	Non Profit Agency,Y	Public (City Operated) Agency,N	0.0	-Infinity	Infinity	0.0	0.9

The Tukey's Honestly Significant Difference (HSD) test indicates that Non-Profit Agencies receiving subsidies have a statistically significant difference in mean of capacity than Non-Profit Agencies without subsidies, with a modest difference in means (p-value=0.001). Subsidized Non-Profit Agencies also show significantly different capacities compared to both Commercial Agencies without subsidies and Commercial Agencies with subsidies (p-values=0.001 for both comparisons), suggesting that subsidies have a more pronounced effect on the capacity of Non-Profit Agencies. Interestingly, there is no statistically significant difference in capacity between subsidized Non-Profit Agencies and Public (city-operated) Agencies without subsidies (p-value=0.900), pointing to a nuanced interaction between agency type, subsidy status, and capacity.

Lastly, we will look at the interaction plot:



This interaction plot shows the relationship between the type of agency and the effect of subsidies on toddlers's mean capacity in childcare centers. For commercial agencies, the mean of capacity is lower in receiving subsidies compared to those that do not receive subsidies. Similarly, Non-Profit Agencies that receive subsidies also show a lower mean in

capacity compared to those without subsidies. The toddler capacity mean of commercial agencies is generally higher than that of non-profit agencies no matter whether subsides or not. The city-operated agency is probably operated by the government, so it is fully subsided and with the highest mean of capacity.

5. Conclusion

To answer the first research question, the one-way ANOVA result shows there are significant differences in toddler space capacity across different types of operating agencies. Non-profit agencies typically allocate less space to toddlers compared to Commercial and Public Agencies, while the difference in capacity rates between commercial and public is not statistically significant.

To answer the second research question, the two-way ANOVA results demonstrate that both types of operating agencies and subsidy status independently and significantly influence toddler space capacity, as indicated by their respective F-values and p-values. The interaction effect between these two factors, suggests a possible differential impact of subsidies on space capacity across different agency types.

It is important to note, however, that the assumptions for both one-way and two-way ANOVA were not fully met. The normality assumption was violated, as shown by the Q-Q plot and the Shapiro-Wilk test, and the homogeneity of variances was not confirmed by Levene's test. These violations mean that the results of the ANOVA tests should be interpreted with caution. Suggests that alternative analyses or transformations of data might be necessary for more robust conclusions. Despite these limitations, the analyses provide valuable insights into the factors that influence childcare capacity in Toronto, which can inform policy and operational decisions to enhance childcare accessibility and affordability.