Introduction

In Ontario, securing childcare is a significant challenge for families due to the high costs and insufficient availability of spaces. With 75% of families unable to afford childcare, the government's initiative to create 100,000 new spaces is a crucial step towards addressing this issue. Our analysis aims to delve into two specific aspects of childcare availability using ANOVA.

Research Questions:

1. Is there a significant difference in the total number of childcare spaces (TOTSPACE) available across different operating auspices (AUSPICE: Commercial, Non Profit, Public)?

2. Is there an interaction effect between the operating auspice (AUSPICE) and the availability of a fee subsidy contract (subsidy) on the total number of childcare spaces (TOTSPACE)?

When conducting ANOVA tests, several assumptions are made regarding the data:
Independence: Each data point is independent and not influenced by others.
Normality: Data should follow a normal distribution within each group.
Equal Variances: All groups should have similar variability.
Scale: The data should be interval or ratio scale.
No Outliers: Extreme values can affect the results, so they're assumed to be minimal or absent.

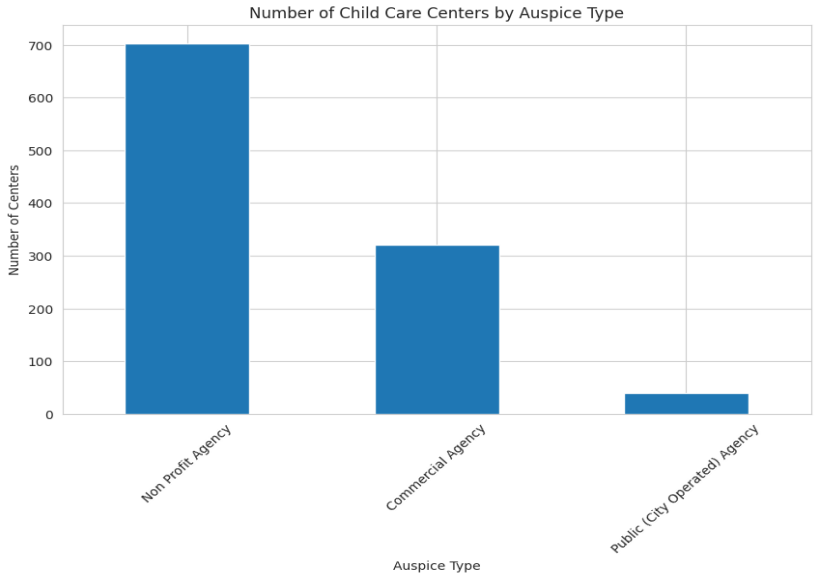
Data Cleaning

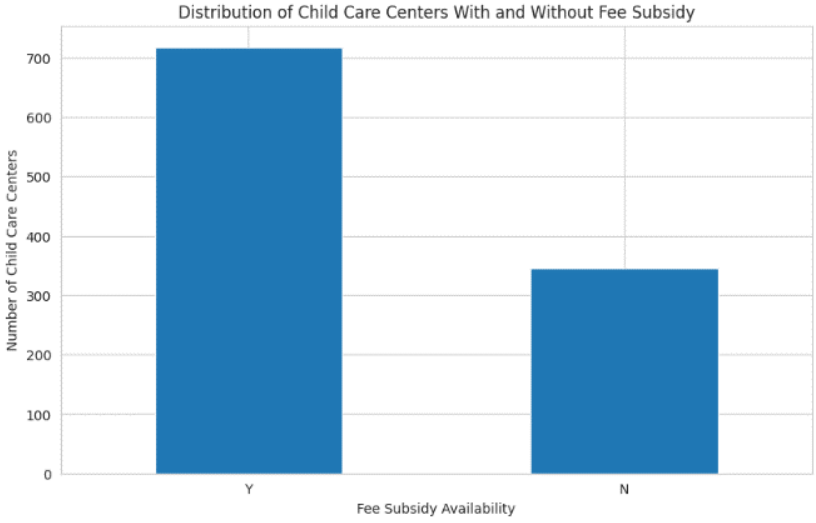
The raw dataset has a total of **17** columns with **1063** rows. Below we outlined our data and its' dictionary of what each column means.

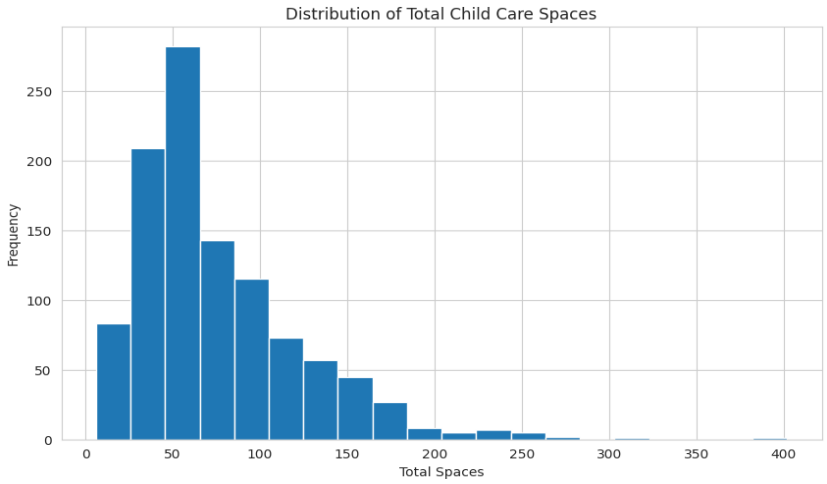
_id	Unique row identifier for Open Data database
LOC_ID	Unique identifier
LOC_NAME	Name of the child care centre
AUSPICE	Operating auspice (Commercial, Non Profit or Public)
ADDRESS	Address street number, street name, street type, street direction, and unit
PCODE	Address postal code
ward	City ward number
bldg_type	Type of building
BLDGNAME	Name of the building the child care centre is located in
IGSPACE	Child care spaces for infants 0-18 months

TGSPACE	Child care spaces for toddlers 18-30 months
PGSPACE	Child care spaces for preschoolers 30 months up until they enter grade one
KGSPACE	Child care spaces for children in full-day kindergarten
SGSPACE	Child care spaces for children grade one and up
TOTSPACE	Child care spaces for all age groups
subsidy	Centre has a fee subsidy contract (Yes/No)
cwelcc_flag	'Y' indicates space participates in CWELCC, blank indicates it does not

For our ANOVA analysis, we only need 3 features: 'AUSPICE', 'subsidy', 'TOTSPACE'. To help us understand the features better, we visualize the categories variables ('AUSPICE' and 'subsidy') using bar plots, and the 'TOTSPACE' using histograms:

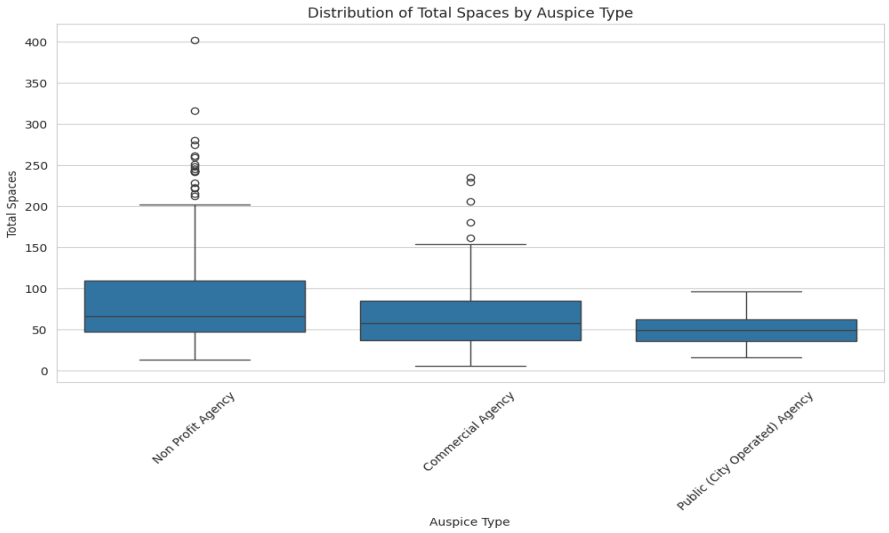






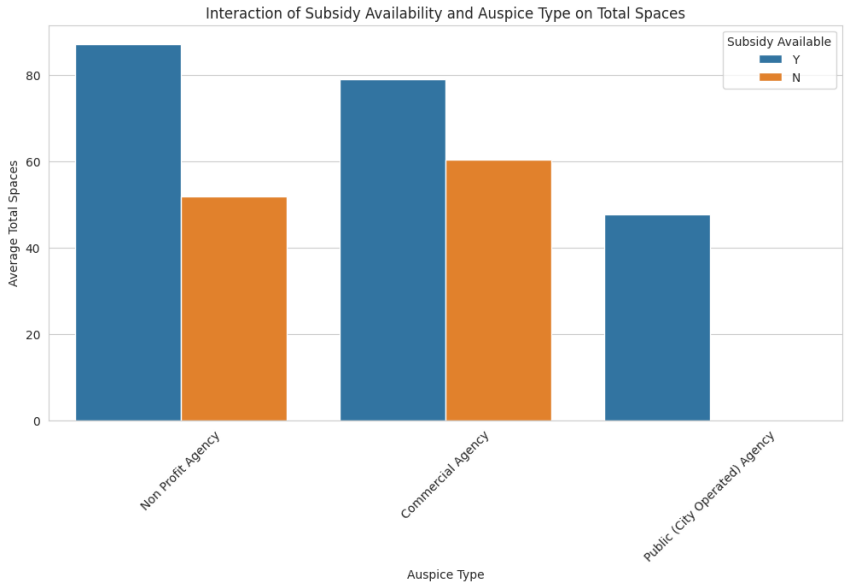
Exploratory Data Analysis

In order to make sure that outliers would not affect our analysis, we need to use boxplot to visualize the distribution of total spaces by auspice.



It seems there are some outliers for the Total space by Auspice type. We need to remove the outliers to make sure they don't affect our analysis.

Also, for the two-way ANOVA, we want to have a sense of the interaction between subsidy availability and auspice type on total spaces, so we would do a bar plot different Auspice with subsidy availability side-by-side.



One-way ANOVA

Let's restate our research questions here: Is there a significant difference in the total number of childcare spaces (TOTSPACE) available across different operating auspices (AUSPICE: Commercial, Non Profit, Public)?

We group the TOTSPACE by AUSPICE and apply the one-way ANOVA from the stats library.

	sum_sq	df	F	Pr(>F)
C(AUSPICE)	9.61e+04	2	21.843	5.06e-10
Residual	2.33e+06	1060		

From the results, we can see a statistically significant difference in the total number of childcare spaces (TOTSPACE) available across different operating auspices (AUSPICE: Commercial, Non Profit, Public). This conclusion is supported by the very low p-value (PR(>F) < 0.001), which is far below the commonly used significance level of 0.05. The F-statistic of approximately 21.84 indicates the between-groups variance is roughly 21 times larger than the within-group variance.

To understand which specific pairs of Auspices differ from each other, we perform a Tukey's HSD, and the following is the result:

Multiple Comparison of Means – Tukey HSD, FWER = 0.05						
Group 1	Group 2	MeanDiff	p-ajd	lower	upper	reject
Commercial Agency	Non Profit Agency	17.120	0.0	9.704	24.535	True
Commercial Agency	Public (City Operated) Agency	-17.215	0.00779	-35.883	1.453	False
Non Profit Agency	Public (City Operated) Agency	-34.335	0.0	-52.445	-16.224	True

For Commercial Agency vs. Non Profit Agency, the mean difference of about 17.12 indicates that Commercial Agencies have, on average, 17.12 more spaces available than Non Profit Agencies, and this difference is statistically significant (p-adj = 0.0).
For Commercial Agency vs. Public (City Operated) Agency, the mean difference of approximately -17.21 suggests that Commercial Agencies have fewer spaces than Public Agencies, but this difference is not statistically significant (p-adj = 0.0779), hence the 'False' in the 'reject' column.
For Non Profit Agency vs. Public (City Operated) Agency, the negative mean difference of around -34.33 indicates that Non Profit Agencies have significantly fewer spaces compared to Public Agencies (p-adj = 0.0), which is consistent with the true in the 'reject' column.

In summary, the post-hoc analysis clarifies that there is a significant difference in the total number of childcare spaces between Commercial and Non Profit Agencies as well as between Non Profit and Public Agencies, but not between Commercial and Public Agencies. This nuanced insight is critical for understanding how the type of operating auspice affects the availability of childcare spaces and can inform policy and management decisions in the sector.

Two-way ANOVA

Let's restate our research questions here: Is there an interaction effect between the operating auspice (AUSPICE) and the availability of a fee subsidy contract (subsidy) on the total number of childcare spaces (TOTSPACE)?

After careful calculation, we get the following result:

	sum_sq	df	F	Pr(>F)
C(AUSPICE)	8.568e+03	2	2.0576	1.283e-01
C(subsidy)	8.353e+04	1	40.118	3.529e-10
C(AUSPICE): C(subsidy)	5.603e+04	2	13.457	1.694e-06
Residual	2.203e+06	1058		

The results reveal some interesting results. The F-statistic is approximately 2.06 with a p-value of 0.128 (greater than the conventional alpha level of 0.05), suggesting that there is no statistically significant difference in the number of total spaces based solely on the operating auspice when not considering the subsidy. The subsidy has a significant effect on the total number of spaces with an F-statistic of approximately 40.12 and a very small p-value (approx. 3.52e-10). This indicates a strong main effect, meaning that the availability of a fee subsidy contract significantly influences the number of childcare spaces available. What's more, the interaction term has an F-statistic of approximately 13.46 and a p-value of about 1.69e-06, which is highly significant. This reveals that the effect of the operating auspice on the total number of childcare spaces depends on whether a fee subsidy contract is available. In other words, the impact of the operating auspice is not consistent across the levels of subsidy; instead, the presence or absence of a subsidy interacts with the auspice to influence the total number of spaces.

To understand how the availability of a subsidy contract (Y for yes, N for no) interacts with the type of operating auspice (Commercial Agency, Non-Profit Agency, Public (City Operated) Agency) to affect the total number of childcare spaces (TOTSPACE), we perform a post-hoc analysis:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Commercial Agency_N	Commercial Agency_Y	18.5991	0.0153	2.3733	34.8249	True
Commercial Agency_N	Non Profit Agency_N	-8.5631	0.5037	-23.273	6.1469	False
Commercial Agency_N	Non Profit Agency_Y	26.7646	0.0	17.2862	36.243	True
Commercial Agency_N	Public (City Operated) Agency_Y	-12.6958	0.4893	-34.2034	8.8119	False
Commercial Agency_Y	Non Profit Agency_N	-27.1621	0.0008	-45.9161	-8.4082	True
Commercial Agency_Y	Non Profit Agency_Y	8.1655	0.5712	-6.8401	23.1711	False
Commercial Agency_Y	Public (City Operated) Agency_Y	-31.2949	0.0045	-55.747	-6.8427	True
Non Profit Agency_N	Non Profit Agency_Y	35.3277	0.0	21.9757	48.6796	True
Non Profit Agency_N	Public (City Operated) Agency_Y	-4.1327	0.9891	-27.6064	19.3409	False
Non Profit Agency_Y	Public (City Operated) Agency_Y	-39.4604	0.0	-60.0631	-18.8577	True

The post-hoc Tukey HSD test indicates that the availability of subsidies significantly influences the number of childcare spaces across different operating agencies. Subsidies lead to an increase in spaces for both Commercial and Non-Profit Agencies. The increase is more pronounced for Non-Profit Agencies when subsidies are available, suggesting that subsidies are particularly effective for them. Commercial Agencies benefit less from subsidies when compared to Public Agencies, which generally offer more spaces.

The significant interaction effect found in the two-way ANOVA (p-value ~ 1.69e-06) suggests that the impact of the operating auspice on total spaces is not consistent but depends on the availability of subsidies. The post-hoc results give us a nuanced understanding of this interaction, indicating that subsidies have a significant impact on the number of spaces available, and this impact varies across different types of agencies.

These findings emphasize the importance of subsidies in influencing the capacity of childcare facilities and suggest that policies aimed at increasing childcare spaces should consider both the type of operating agency and the allocation of subsidies to maximize their effectiveness.

Conclusion

In light of the exploratory and inferential statistical analyses performed on the childcare spaces data in Ontario, the findings are quite telling. The one-way ANOVA and subsequent Tukey HSD post-hoc analysis confirmed significant differences in the availability of childcare spaces among different auspices, specifically between Commercial and Non-Profit, and Non-Profit and Public agencies, with subsidies playing a crucial role. The two-way ANOVA further revealed a significant interaction effect, indicating that the impact of auspice type on childcare space availability is notably influenced by the presence of fee subsidies. These results not only highlight the critical role of subsidy availability in determining childcare space distribution but also suggest that policies aimed at expanding childcare access need to be nuanced and tailored to the type of childcare agency.