INF 2178 Technical Assignment 2

University of Toronto
Yingliang Ding
1009324492

Assumption for One-Way ANOVA #1:

- H0(Null Hypothesis): There is no significant difference in the average total spaces available (TOTSPACE) among child care centers managed by different auspices (Non-Profit vs. Commercial).
- H 1 (Alternative Hypothesis): There is a significant difference in the average total spaces available (TOTSPACE) among child care centers managed by different auspices (Non-Profit vs. Commercial).

Boxplot (see appendix 1)
The boxplot (see appendix 1) visualizes the distribution of total available space for the different types of child care providers (nonprofit, commercial, and public (city-run)). Significant differences in total available space by sponsorship type can be seen in the boxplots, with a greater range of differences and more outliers for nonprofit organizations in particular. This suggests that a one-way ANOVA may reveal significant differences. Non-profit organizations appear to have the highest median total space, while public (city-run) organizations have the lowest median total space.

Checking homogeneity of variances
Please note that the P-value of Performing Levene's test for homogeneity of variances is 4.233362814820158e-06, which far below any conventional significance level (0.05). This may lead to inaccurate ANOVA results.

Based on the result of One-way ANOVA(see appendix 2):
Sum of Squares (sum_sq): This represents the variation due to the groups (between groups for C(AUSPICE)) and the variation within groups (for Residual). For C(AUSPICE), the sum of squares is approximately 96112.11, indicating the variation between the different types of auspices.
Degrees of Freedom (df): There are 2 degrees of freedom for the C(AUSPICE) factor, which suggests that there are three categories being compared (since degrees of freedom are calculated as the number of groups minus one). The degrees of freedom for the residuals are 1060, corresponding to the number of observations minus the number of groups.
F-Statistic (F): The F-statistic is 21.843051, which is used to determine whether the observed variances between means are significant or not. A higher F-statistic typically indicates a greater probability that there are significant differences between the group means.
P-Value (PR(>F)): The p-value is extremely low (5.057716e-10, which is 0.0000000005057716), far below any conventional significance level (0.05). This means that the test found very strong evidence to reject the null hypothesis of no difference in total spaces among the different auspice types.

The ANOVA test strongly suggests that there is a significant difference in the average total spaces available (TOTSPACE) among the different types of auspices managing child care centers. Given the extremely low p-value, we can confidently reject the null hypothesis and conclude that the type of auspice (whether Non-Profit, Commercial, or Public) has a significant impact on the number of total spaces a child care center can offer.

From this preliminary analysis, we can hypothesize that Toronto's child care environment is largely determined by the type of management, which has implications for how the city's child care needs are met. Non-profit and public agencies appear to differ in prioritizing space availability compared to commercial agencies, which may be related to their respective mandates and funding structures. However, the differences within each type of organization

suggest that the factors affecting child care capacity are not limited to the type of organization.

Post-hoc test
The Tukey HSD post hoc analysis nonprofit agencies significantly outperform commercial agencies in terms of space availability. However, the differences between commercial and public (city-operated) agencies are not significant, suggesting comparable space availability. Non-profit agencies also have significantly less space than public (city-operated) agencies. These findings underscore the importance of agency type in determining child care space availability and point to unique operational characteristics of nonprofit agencies within the child care sector.

Assumption for One-Way ANOVA #2:

- H0(Null Hypothesis): The average number of preschool spaces (PGSPACE) is the same across all wards in Toronto.
- H1(Alternative Hypothesis): The average number of preschool spaces (PGSPACE) varies significantly across different wards in Toronto.

Boxplot (see appendix 3)
The boxplot for one-way ANOVA 2 illustrates the distribution of pre-school providers across wards in the region. This visualization facilitates the one-way ANOVA test by clearly depicting the distribution of the number of pre-school establishments across the constituencies and shows that there are differences in the number of pre-school places in different wards, with some wards having a wider range of values and others having a more concentrated distribution. Whilst we cannot draw statistically significant conclusions from the box plot alone, it does suggest that there may be significant differences between wards and therefore further statistical analysis through one-way ANOVA is necessary to test for significance. If the ANOVA confirms that the differences are significant, it may suggest that targeted policy interventions are needed to ensure equitable distribution of preschool space.

Checking homogeneity of variances
Please note that the P-value of Performing Levene's test for homogeneity of variances is 0.09061533070077447, which above significance level (0.05). ANOVA test can be performed normally.

Based on the result of ANOVA(see appendix 4):
Sum of Squares (sum_sq): The sum of squares between groups (C(ward)) is approximately 18,498.39859, and the sum of squares for the residual is about 348,019.458418. This indicates how much variance is explained by the wards and how much is unexplained or due to individual differences within wards, respectively.
Degrees of Freedom (df): The degrees of freedom (df) associated with the wards is 24, suggesting that there are 25 wards being compared (df is one less than the number of groups). The residual degrees of freedom is 1038, which is the total number of observations minus the number of wards.
F-Statistic (F): The F-statistic is 2.298882, which is a measure of the variance between the wards' means over the variance within the wards.
P-Value (PR(>F)): The p-value associated with this F-statistic is extremely small (0.000386), much lower than the common alpha level of 0.05 used for statistical tests.

The results of the ANOVA show that there is a significant difference in the number of preschools across Toronto's wards, which suggests that child care is not evenly distributed across the city.The F statistic shows that ward location has a considerable effect on the number of preschools, and the p-value is highly significant, so we can say with certainty that preschools are not evenly distributed across the city. This means that children's access to pre-schools may be heavily influenced by their constituency.

This variation may stem from a variety of factors, including socioeconomic differences, `urban planning decisions, or differences in the level of investment in public infrastructure. The statistical significance of these results provides a compelling reason for policymakers and urban planners to reassess resource allocation. With the knowledge that certain constituencies may struggle to provide adequate preschool space, targeted interventions can be developed to ensure more equitable provision of ECE services.

Post-hoc test
The Tukey HSD post-hoc analysis for One-Way ANOVA #2 represent different wards in this context. The results indicate that there are no statistically significant differences in the number of preschool spaces available between most of the wards, as the p-adjusted values are greater than the alpha level of 0.05. This is highlighted by the consistent 'False' under the 'reject' column for nearly all pairwise comparisons, signifying that we cannot reject the null hypothesis of equal means for these group comparisons.

Assumption for Two-Way ANOVA

- H0(Null Hypothesis): There is no interaction effect between the type of management (AUSPICE) and the type of building (bldg_type) on the total spaces available in child care centers.
- H1(Alternative Hypothesis): There is an interaction effect between the type of management (AUSPICE) and the type of building (bldg_type) on the total spaces available in child care centers.

Boxplot (see appendix 5)
The boxplot of Bivariate ANOVA 1 illustrates the distribution of total available space by sponsorship type and building type. This visualization tool facilitates a two-way ANOVA, allowing us to detect within-group differences. The distribution of total space within each sponsoring organization and building type category is visible at a glance, showing differences within each subgroup. The next finding, between-group differences, is that by looking at the median values and distributions of the boxes for different groups, we can identify possible differences in the total amount of space available for different types of supporting organizations and building types.
Moreover, the interaction effect can also be reflected in the boxplots, which show the total space available for each combination of mascot and building type. If there were significant interaction effects, we would see non-parallel lines when comparing the medians of these combinations. This may indicate that the effect of one factor depends on the level of another. Finally, there are outliers, the presence of which is easy to recognize. These outliers may affect the results and assumptions of the ANOVA and may require further examination or cleaning of the data.

From this boxplot, we can tentatively conclude that building type appears to have a greater impact on total available space than sponsorship type. However, since the figure shows different distributions of each combination of hosting and building type on total space, this

suggests that the interaction between these two factors may be significant and should be explored with a two-way ANOVA. This could reveal how the effect of management style on space availability depends on the type of building in which the childcare center is located and vice versa.

Checking homogeneity of variances
Please note that the P-value of Performing Levene's test for homogeneity of variances is 9.701217710729298e-18, which far below any conventional significance level (0.05). This may lead to inaccurate ANOVA results.

Interaction Plot (see appendix 6)
The interaction plot from the Two-way ANOVA#1 depict the relationship between sponsorship type, building type, and total space available. In the interaction plot, non-parallel lines indicate an interaction between the variables. Within each sponsorship category, the lines for different building types are not parallel, suggesting that the relationship between sponsorship type and total available space is influenced by building type. This suggests that there is a clear interaction effect, i.e. the effect of one factor on total available space depends on the level of the other factor.

Secondly the total usable space for certain building types shows greater variation in the comparison between host types as evidenced by the range of values along the line. For example, certain building types, such as "Catholic Elementary School (French)" and "Commercial Building," show greater differences in total space across sponsorship types. The figure also shows that for certain building types, the total available space varies little regardless of the sponsorship type, as evidenced by the proximity of the lines to each other. In contrast, other building types show more pronounced differences between sponsorship types.

In summary, the interaction plots visually confirm the significant interaction of sponsorship and building type on total available space, confirming that the two factors should not be analyzed in isolation. The data suggest that strategies to increase or optimize available space in child care centers may need to be tailored to specific combinations of sponsorship and building type.

Based on the result of Two-Way ANOVA(see appendix 7):
C(AUSPICE): The sum of squares (sum_sq) for `AUSPICE` is a very small negative number, which is unusual and typically indicates a computational error or anomaly. The F-statistic is extremely small and the p-value is 1.0, which means that statistically, there is no significant effect of `AUSPICE` on the dependent variable within the limits of numerical precision used in the calculations.

C(bldg_type): The `bldg_type` factor has a large sum of squares, and the F-statistic is approximately 25.54, with a highly significant p-value ($p < 0.0001$). This indicates that there is a significant effect of `bldg_type` on the dependent variable, suggesting that different building types have different average spaces available.

C(AUSPICE):C(bldg_type) Interaction: The interaction term has a sum of squares that suggests there is variation explained by the interaction between `AUSPICE` and `bldg_type`. The F-statistic is around 2.49, and the p-value is very small ($p < 0.01$), indicating that the interaction effect is statistically significant. This means that the impact of `AUSPICE` on the dependent variable changes depending on the `bldg_type`, and vice versa.

Residual: The residual sum of squares shows the variation not explained by the model, which is relatively large, suggesting that there are other factors not included in the model that also affect the dependent variable.

The results of the two-way ANOVA revealed an intricate and dynamic relationship between management type and building type within Toronto child care centers. Interestingly, management type alone does not have a significant impact on the total square footage available in these centers. In contrast, building type is a key factor that has a significant impact on total square footage. This suggests that physical infrastructure plays a more critical role than management style in determining center capacity.

However, the picture becomes more complex with significant interactions between management style and building type, suggesting that their combined impact cannot be ignored. For example, a nonprofit organization operating in a specific type of building may have more or less usable space than a commercial organization in a comparable building. This subtle interplay suggests that the effectiveness of child care center management depends in part on the physical characteristics of their premises.

This analysis invites stakeholders to look beyond superficial metrics and consider the symbiotic relationship between management practices and physical space. It highlights the need for a differentiated approach to planning and allocating resources for childcare services, recognizing that unique combinations of management and building types can optimize space use.

Post-hoc test

The Tukey HSD post-hoc analysis for the two-way ANOVA indicates significant differences in space availability between certain types of facilities. While most comparisons did not yield significant results-meaning that different building types generally provide similar space capacities-a few cases stood out. For example, "Catholic Elementary School (French)" showed a notable difference when compared to "House of Worship" and "Public High School," suggesting that these particular facility types may affect space allocation. These significant findings highlight areas that may require targeted policy interventions or management strategies. Overall, while facility type often does not affect space availability, the exceptions noted here are critical to understanding specific influences on space allocation in educational settings.

Further Analysis

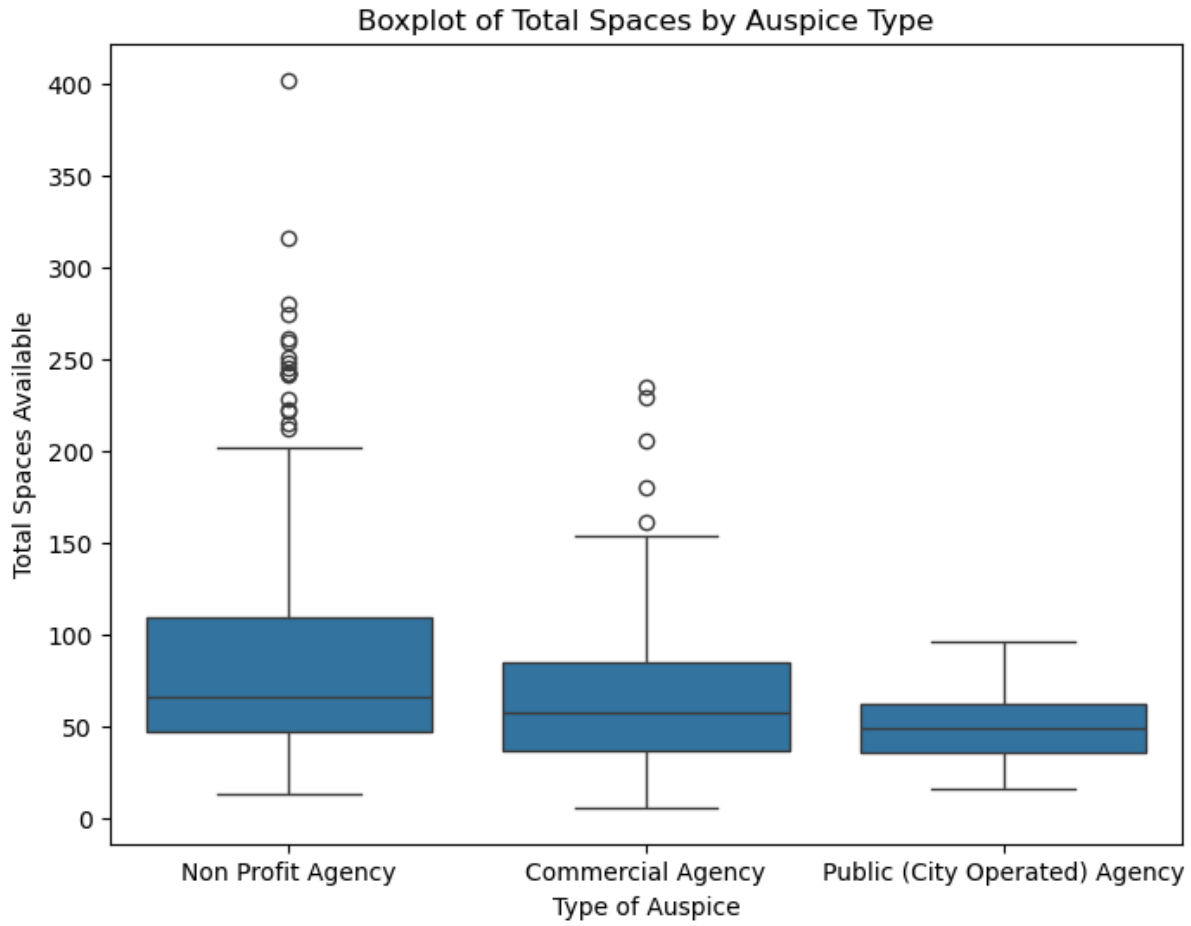
Further analysis should explore the nuances and subtleties that underlie the distribution of child care spaces across Toronto, building on the findings from the one-way and two-way ANOVAs. It would be prudent to explore more deeply how specific combinations of these factors correlate with space availability, given the significant interaction between building type and management style found in the two-way ANOVA.

In order to understand the variation not explained by the current model, subsequent analyses could include a detailed examination of the residuals. This could be an examination of additional variables such as location-specific factors, socio-economic status of the wards, or proximity to community services. Regression diagnostics could also be used to identify potential model violations such as nonlinearity, heteroscedasticity, or influential outliers that could have an impact on the ANOVA results.

In addition, further disaggregation by county could shed light on geographic patterns or disparities in service provision, given the variation in the number of preschool places indicated by the one-way ANOVA. For policymakers seeking to address inequities and optimize resource allocation, these findings would be valuable. Thus, in order to refine the understanding of the child care landscape and support data-driven decision making to improve child care services, the following analytical steps would be taken.

Appendix

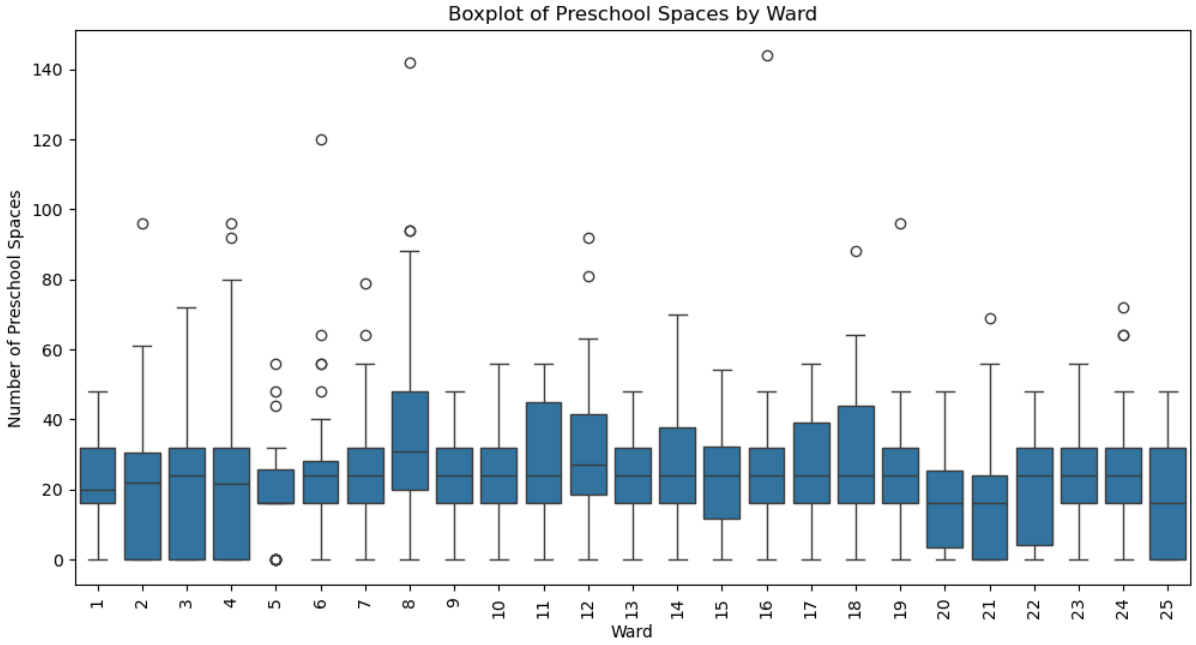
Appendix 1:



Appendix 2:

	sum_sq	df	F	PR(>F)
C(AUSPICE)	9.611211e+04	2.0	21.843051	5.057716e-10
Residual	2.332065e+06	1060.0	NaN	NaN

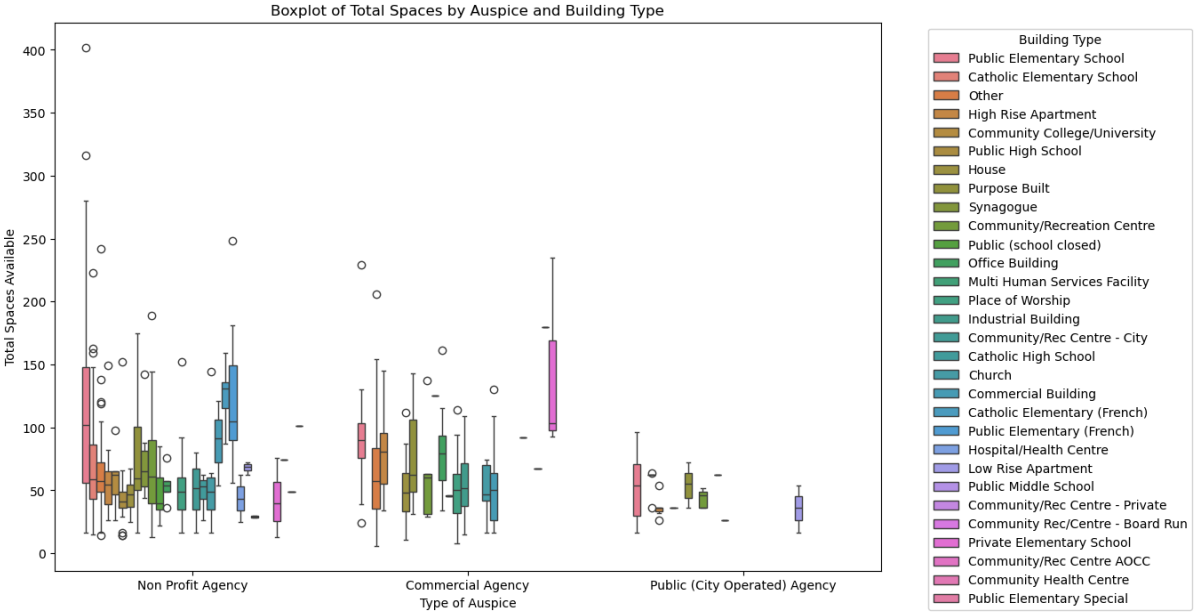
Appendix 3:



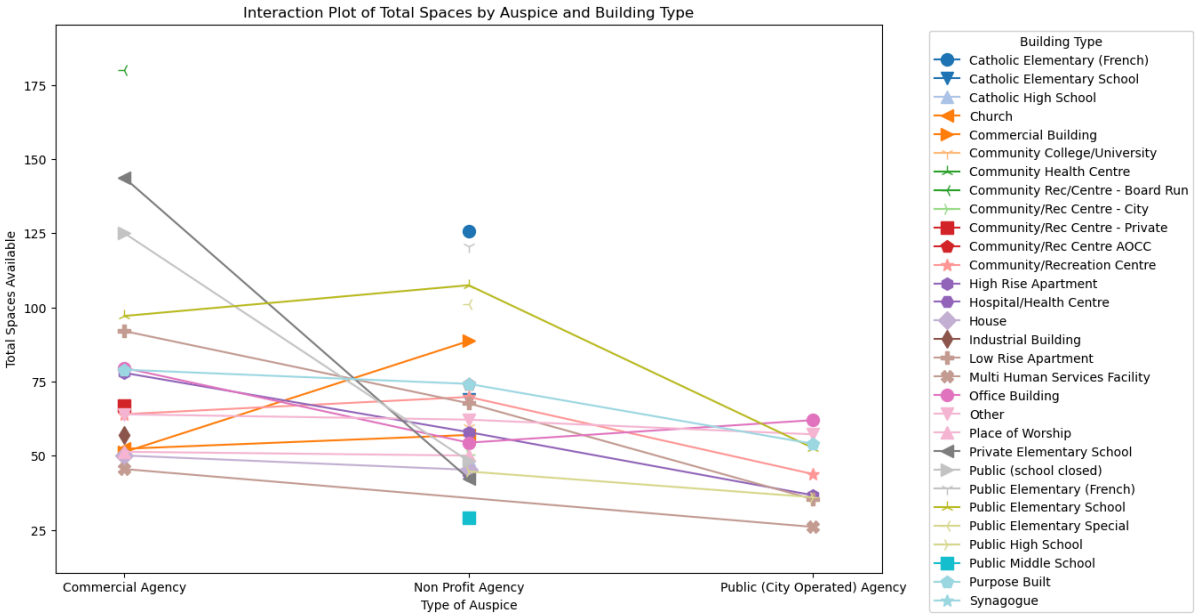
Appendix 4:

	sum_sq	df	F	PR(>F)
C(ward)	18498.398591	24.0	2.298882	0.000386
Residual	348019.458418	1038.0	NaN	NaN

Appendix 5:



Appendix 6:



Appendix 7:

	sum_sq	df	F	PR(>F)
C(AUSPICE)	-6.190936e-07	2.0	-1.704891e-10	1.000000e+00
C(bldg_type)	1.344907e+06	29.0	2.554256e+01	2.527883e-24
C(AUSPICE):C(bldg_type)	2.622099e+05	58.0	2.489954e+00	6.381552e-06
Residual	1.835612e+06	1011.0	NaN	NaN