Analyzing Ontario's Child Care Centres Dataset

Zonglin Li
1004910117

Introduction

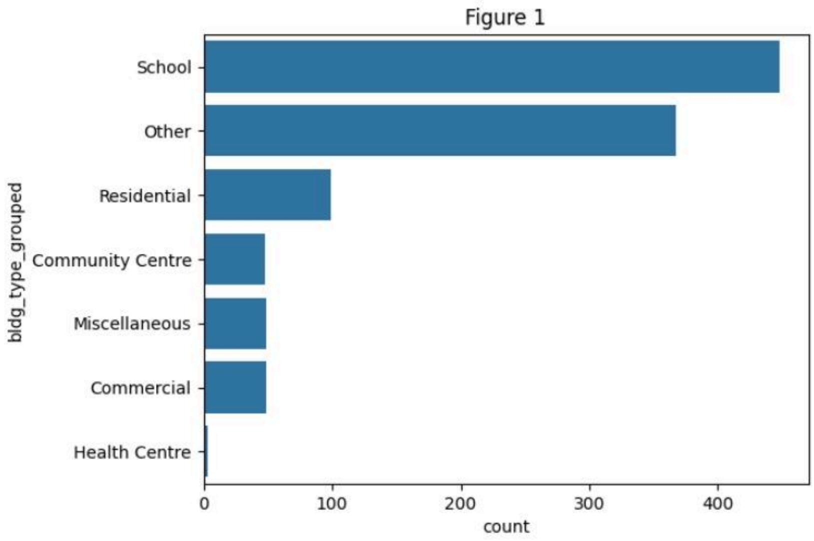
Recognizing the critical need for more accessible and affordable childcare, the Ontario government has made a substantial commitment. Their pledge to create 100,000 new childcare spaces between 2016 and 2019 underscores the urgency and importance of our study. This research aims to illuminate the current situation of childcare centers in Ontario through various data analysis methods. We will employ one-way and two-way ANOVA to determine if there are statistically significant differences between variables in the dataset.

Our data analytics will address two related research questions:

- Do different management types of childcare centers significantly affect the average number of Childcare spaces available for all age groups?
- How do the type of management and the type of building independently and jointly affect the total number of spaces available in childcare centers?

Datta cleaning and data manipulation

The Ontario childcare dataset, with its 17 columns and 1062 entries, each representing one childcare center, is a rich source of information. To ensure the reliability of our findings, we have meticulously cleaned and manipulated the data. We have ignored null values in IGSPACE, PGSPACE, KGSPACE, and 'SGSPACE', and removed irrelevant categorical variables. Our data frame now includes only 'IGSPACE', 'TGSPACE', 'PGSPACE,' 'KGSPACE', 'SGSPACE', 'TOTSPACE', 'ward', 'subsidy', 'AUSPICE', and 'bldg_type_grouped'. The 'bldg_type' represents the type of building childcare centers are located in; there are 30 different types of buildings, and many buildings are only different in name. To enhance data visualization and process efficiency, we have created a function to summarize 30 different building types into seven types: School, Residential, community center, commercial, Health center, other and Miscellaneous.



Exploratory Data Analysis

Our exploratory data analysis has yielded valuable insights. Figure 2, which includes a boxplot and a swarm plot, shows the distribution of the Child care spaces for all age groups across three Operating auspices: Commercial, Non-Profit, and Public agencies. By observing Figure 2, we can discern significant differences. The boxplot of the Non-profit agency has a broader range of 'TOTALSPACE' than other auspices, indicating more considerable variability in the total spaces available within this category. Meanwhile, the Public agency has the narrowest range of 'TOTALSPACE' and interquartile range, suggesting the lowest Childcare spaces for all age groups and the most consistent 'TOTSPACE' values. These findings provide a comprehensive understanding of the current childcare landscape in Ontario.

Figure 3 shows a complex boxplot that illustrates the distribution of childcare spaces for all age groups categorized by 'AUSPICE' (the type of management) and our new variable 'bldg_type_grouped' (Building type). We can have several boxplots for each AUSPICE. Each boxplot has a unique color, representing a different building type. We can tell from the distribution that Non-profit agencies exhibit a wide range of 'TOTSPACE,' with considerable variation in all building types, suggesting diverse operations and capacities, while 'Public Agency' centers, conversely, show the least variation, indicating a more uniform distribution of 'TOTSPACE' irrespective of the building type. Overall, figure 2 and Figure 3 show that both the auspice type and building type contribute to differences in the total capacity of childcare centers, which might be necessary for policy and planning decisions regarding childcare services.

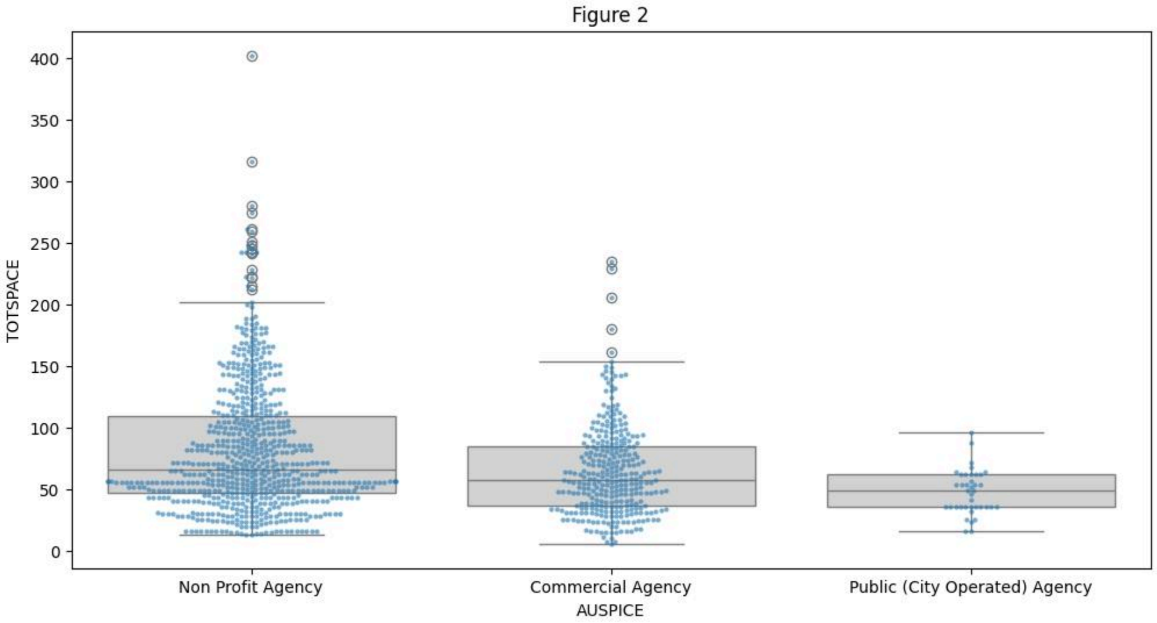
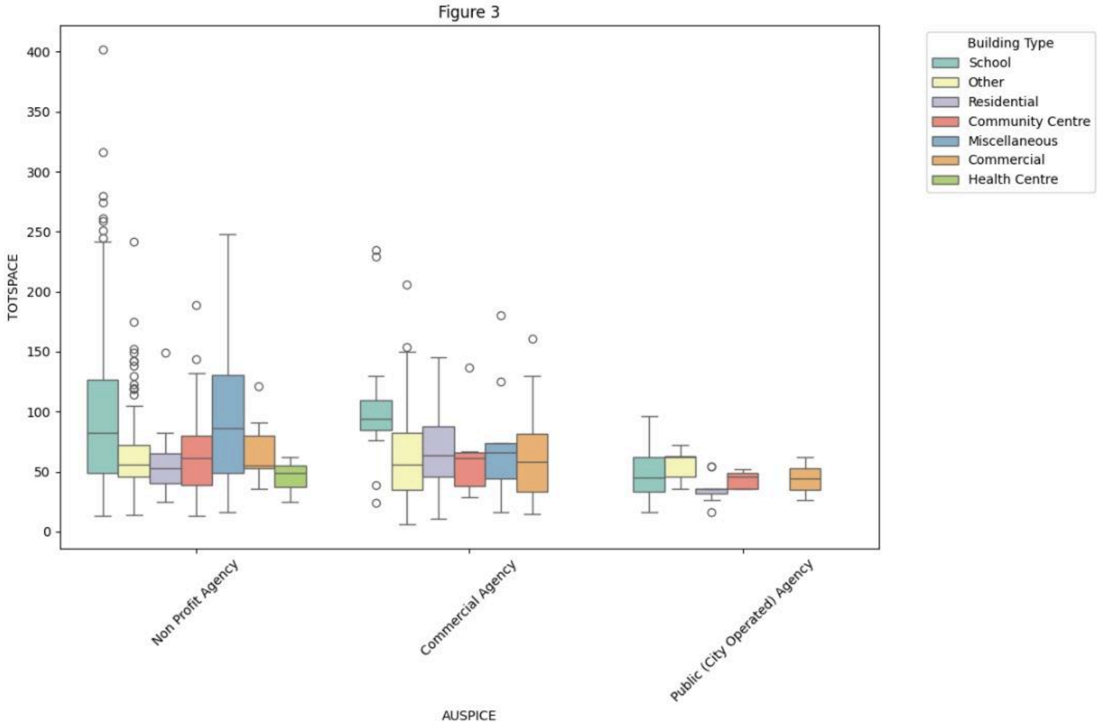


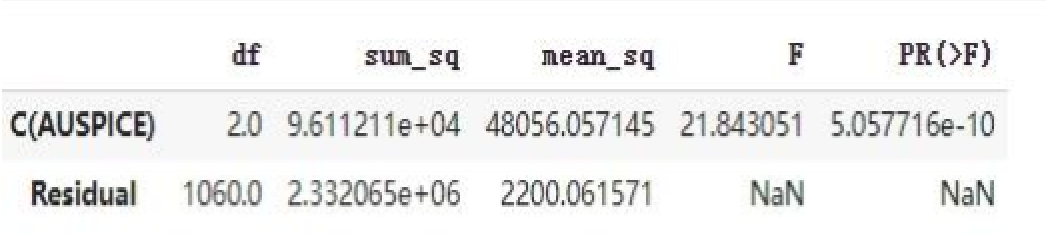
Figure 3



Research 1

Do different management types of childcare centers significantly affect the average number of Childcare spaces available for all age groups?

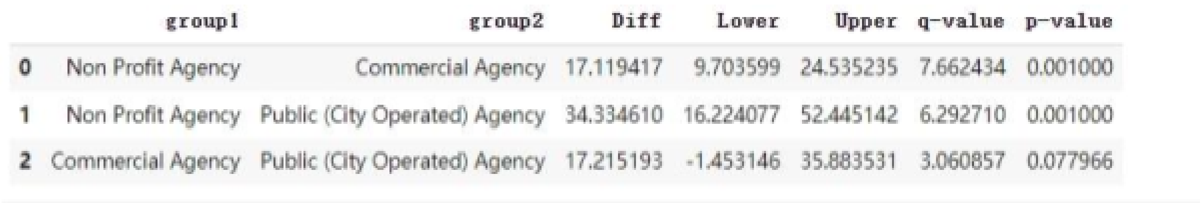
Since 'AUSPICE' is a categorical variable with three different types of management (Nonprofit, Commercial, and Public), One-way ANOVA is an appropriate statistical test when comparing the means of the Childcare spaces for all age groups across different types of management.



	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	9.611211e+04	48056.057145	21.843051	5.057716e-10
Residual	1060.0	2.332065e+06	2200.061571	NaN	NaN

Table 1

After we conducted several data analytic methods in Python, we stored the necessary one-way ANOVA test results in Table 1. We found that the F-statistic is around 21.84, and the p-value is around 5.06e-10, indicating strong evidence that the management type ('AUSPICE') significantly affects the total number of childcare spaces available ('TOTSPACE'). Since management type significantly impacts childcare spaces for all age groups, we conducted Tukey's HSD to determine which management categories differ.



	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	17.119417	9.703599	24.535235	7.662434	0.001000
1	Non Profit Agency	Public (City Operated) Agency	34.334610	16.224077	52.445142	6.292710	0.001000
2	Commercial Agency	Public (City Operated) Agency	17.215193	-1.453146	35.883531	3.060857	0.077966

Table 2

.

In our Tukey's Honestly Significant Difference (HSD) test, we compared all possible pairs of group means to determine which pairs have significant differences.

In Non-Profit Agency vs. Commercial Agency: The mean difference of TOTSPACE is about 17.12 spaces, which means the Non-Profit Agency group has more on average. Since the 95% confidence interval for this difference ranges between 17.11 and 9.7, the p-value is less than 0.05, which is statistically significant.

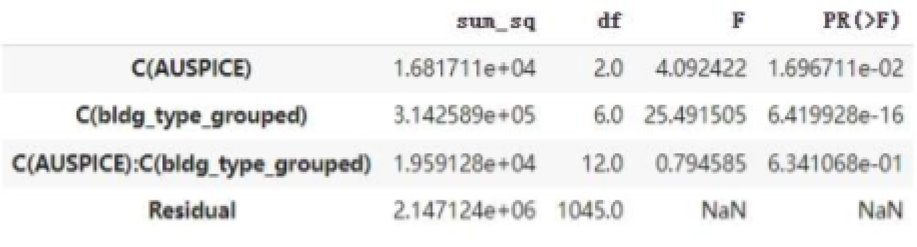
The mean difference between non-profit and public agencies' TOTSPACE is around 34.33, which also shows that Non-Profit Agencies have more spaces on average than Public Agencies.
Meanwhile, with the 95% confidence interval for this difference between 16.22 and 52.33 and the p-value 0.001, we conclude that this difference is also statistically significant.

In commercial Agency vs Public Agency, The difference is around 17.22 spaces, which shows the Commercial Agency has slightly more on average. Since the confidence interval of commercial Agency vs Public Agency crosses zero (about -1.45 to 35.88), and the p-value is 0.077966, the mean differences between commercial Agency and Public Agency here are not statistically significant at the 0.01 level.

Research 2

How do the type of management and the type of building independently and jointly affect the total number of Child care spaces for all age groups?

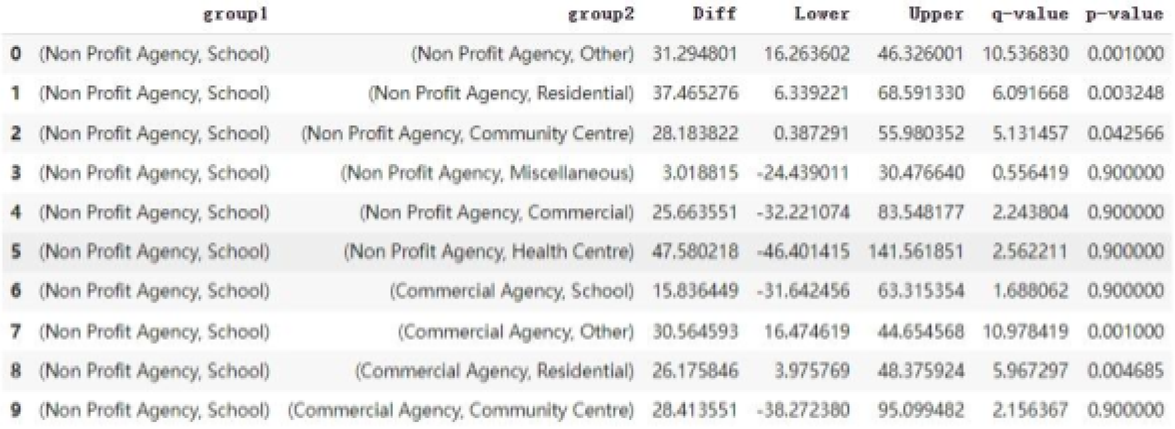
Since we have two categorical variables, 'blog_type_grouped' and 'AUSPICE', and we want to study the independent and joint effects of the Child care spaces for all age groups, we will use two-way ANOVA.



	sum_sq	df	F	PR(>F)
C(AUSPICE)	1.681711e+04	2.0	4.092422	1.696711e-02
C(bldg_type_grouped)	3.142589e+05	6.0	25.491505	6.419928e-16
C(AUSPICE):C(bldg_type_grouped)	1.959128e+04	12.0	0.794585	6.341068e-01
Residual	2.147124e+06	1045.0	NaN	NaN

Table 3

Table 3 is the two-way ANOVA table. For 'AUSPICE' on 'TOTSPACE', we found that the F-statistic is 4.09 and a small p-value of 0.0169, which suggests the type of agency managing a childcare center is systematically associated with the capacity of that center. For 'bldg_type_grouped' on 'TOTSPACE, the F-statistic is approximately 25.49, and we get a p-value < 0.05. We can conclude that Different types of buildings offer different average numbers of childcare spaces. For the Interaction effect 'AUSPICE':'bldg_type_grouped, the interaction term has an F-statistic of approximately 0.75 with a p-value of about 0.6341, which is insignificant at the 0.05 level. This suggests that there is no evidence of a significant interaction effect between the type of management and the type of building on the number of spaces.



	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(Non Profit Agency, School)	(Non Profit Agency, Other)	31.294801	16.263602	46.326001	10.536830	0.001000
1	(Non Profit Agency, School)	(Non Profit Agency, Residential)	37.465276	6.339221	68.591330	6.091668	0.003248
2	(Non Profit Agency, School)	(Non Profit Agency, Community Centre)	28.183822	0.387291	55.980352	5.131457	0.042566
3	(Non Profit Agency, School)	(Non Profit Agency, Miscellaneous)	3.018815	-24.439011	30.476640	0.556419	0.900000
4	(Non Profit Agency, School)	(Non Profit Agency, Commercial)	25.663551	-32.221074	83.548177	2.243804	0.900000
5	(Non Profit Agency, School)	(Non Profit Agency, Health Centre)	47.580218	-46.401415	141.561851	2.562211	0.900000
6	(Non Profit Agency, School)	(Commercial Agency, School)	15.836449	-31.642456	63.315354	1.688062	0.900000
7	(Non Profit Agency, School)	(Commercial Agency, Other)	30.564593	16.474619	44.654568	10.978419	0.001000
8	(Non Profit Agency, School)	(Commercial Agency, Residential)	26.175846	3.975769	48.375924	5.967297	0.004685
9	(Non Profit Agency, School)	(Commercial Agency, Community Centre)	28.413551	-38.272380	95.099482	2.156367	0.900000

Table 4

We applied Tukey’s HSD test to present a detailed comparison of the mean 'TOTSPACE' between pairs of building types. Table 4 displays the first ten rows of the total result. One exciting finding is that schools have a significantly higher mean 'TOTSPACE' than 'Other', 'Residential', 'Community Centre', 'Miscellaneous', and 'Commercial' building types, ranging from 30 to 47 spaces. Also, the significant p-values (less than 0.05) for paired comparisons (Schools vs. Residential, Schools vs. Commercial) confirm that these are not random differences but are statistically significant. The significant p-values imply that the building type is an essential factor in determining the capacity of childcare centers.

Conclusion

Through statistical analysis and data manipulation, We found significant differences in "total space" for the "AUSPICE" category, suggesting that different management types affect the average amount of space available for all ages in childcare centers. Immediately following Tukey's HSD test, we find that Nonprofit Agencies tend to have more childcare spaces on average than Commercial Agencies and Public Agencies. We then used a two-way ANOVA to explore whether the factors of management type ("AUSPICE") and building type ("bldg_type_grouped") significantly affect the total number of childcare spaces for all age groups. The results of the two-way ANOVA show no significant interaction between them, suggesting that each factor affects the number of childcare spaces independently of the other factors. For example, whether a facility is managed by a nonprofit organization, a commercial entity, or a municipality does not change the effect of building type on the number of spaces. Moreover, in a diagnostic analysis of the data, we found that the ANOVA residuals did not follow a normal distribution, which may negatively affect our research result.

Limitation:

In Figure 4, we plot the theoretical quantiles against the standardized residuals from your model. The points would lie approximately along the red reference line if the data were normally distributed. However, the points in Figure 4 did not lie along the reference line, especially at both the lower and upper ends of the distribution, which means that the residuals are not normally distributed.

Figure 5 shows the frequency distribution of the standardized residuals. Our histogram shows a right-skewed distribution, not a bell-shaped curve, which means the residuals are not normally distributed. Both Figure 4 and Figure 5 suggest that the assumption of normality for the residuals does not hold.

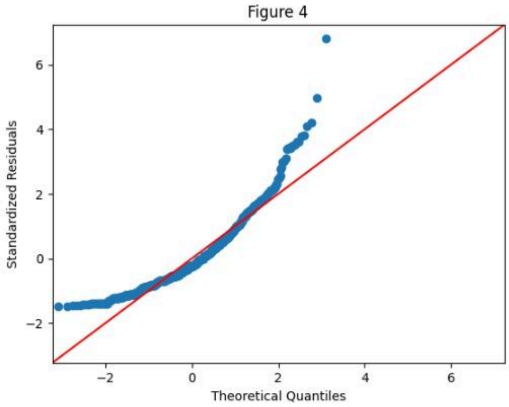


Figure 4

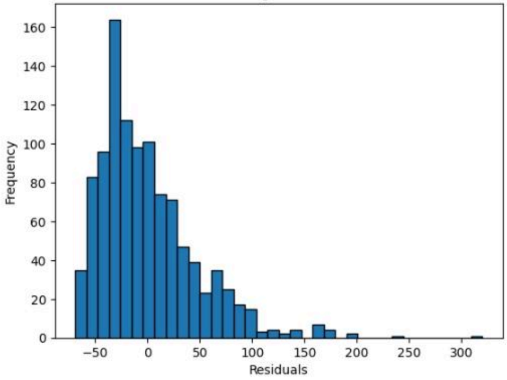


Figure 5