Exploring the Child Care Centre Spaces in Toronto

Introduction

In the vibrant landscape of Toronto, licensed child care centres play a pivotal role in supporting families and fostering early childhood development. This report embarks on a statistical journey to unravel the influence of two critical attributes - building type and operating auspice - on the total space available in these centres. By delving into this analysis, we aim to shed light on how these factors collectively shape the capacity and, by extension, the accessibility of child care services. Understanding these dynamics is crucial for stakeholders, including policymakers, administrators, and families, as they navigate decisions that impact the availability and quality of child care.

Research Question 1: Does the operating auspice (AUSPICE) affect the total space (TOTSPACE) available in child care centres?

- **Hypothesis 0 (H0)**: There is no difference in total space available across different operating auspices.
- **Hypothesis 1 (H1)**: There is a significant difference in total space available across different operating auspices.

Utility of One-Way ANOVA: One-way ANOVA is useful here to test the differences in a continuous outcome (total space) across multiple categories (operating auspices). It allows us to determine if the type of operating auspice influences the space available for child care, which is crucial for understanding resource allocation.

Research Question 2: Does the interaction between operating auspice (AUSPICE) and building type (bldg_type) have a significant effect on the total space available (TOTSPACE) in child care centres?

- **Hypothesis 0 (H0)**: There is no interaction effect between operating auspice and building type on the total space available.
- **Hypothesis 1 (H1)**: There is a significant interaction effect between operating auspice and building type on the total space available.

Utility of Two-Way ANOVA: Two-way ANOVA helps us explore not only the main effects of each factor (operating auspice and building type) but also how these factors interact with each other to influence the total space. This analysis is crucial for identifying whether the impact of one factor depends on the level of another, providing deeper insights into how child care centres' capacities are determined.

Data Cleansing and Preparation

In order to explain the factors of capacity within Toronto's licensed child care institutions, the dataset needed to be thoroughly analyzed and prepared. Initial examination revealed a compilation of over a thousand items, indicating a significant representation of child care institutions throughout Toronto. The dataset's great diversity offers a fruitful ground for statistical study aiming at revealing deep insights into the spatial dynamics regulated by institutional and architectural factors.

   A. Observations:
1. The dataset originally included a comprehensive list of columns:
   - _id, LOC_ID: Unique identifiers for each row and child care centre.
   - LOC_NAME, ADDRESS, PCODE: Information detailing the name and location of each centre.
   - AUSPICE: The operating auspice of the centre (e.g., Non-Profit, Commercial).
   - ward, bldg_type: Administrative and structural classification of the centre's location.
   - Capacity information for different age groups: IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE.
   - TOTSPACE: The total space available across all age groups.
   - subsidy, cwelcc_flag: Indicators of subsidy availability and CWELCC program participation.
2. For our focused analysis, we honed in on specific columns that directly inform our research questions:
   - AUSPICE; bldg_type; TOTSPACE
   B. Feature Engineering
   To ensure the data was primed for analysis, we embarked on a data cleansing journey, aimed at refining the dataset for enhanced accuracy and reliability.
1. **Identifying and Removing Outliers**: Particularly within the TOTSPACE variable, outliers were identified and removed based on the Interquartile Range (IQR) method. This step was crucial to exclude extreme values that could skew our analysis.
2. **Addressing Missing Values**: We carefully examined the dataset for missing values, especially in the columns critical to our analysis (bldg_type, AUSPICE, TOTSPACE). Decisions were made to either impute missing values where feasible or exclude records when necessary.
3. **Ensuring Data Consistency**: The categories within AUSPICE and bldg_type were standardized to ensure consistency across the dataset, eliminating potential duplications or misclassifications.

Exploratory Data Analysis

At the inception of our investigation, the exploratory data analysis was executed on the original dataset to establish a baseline understanding of the operational landscape of Toronto's child care centres. The initial dataset, rich with numerous data points, provided a panoramic view of the capacities across different auspices, which was depicted through a box plot illustrating the total space (TOTSPACE) variable stratified by operating auspice (AUSPICE).

The boxplot shown on the left is the original dataset before data cleaning; it was extracted from the original dataset, revealing a broad dispersion in TOTSPACE among Non-Profit, Commercial, and Public (City Operated) agencies. This graph highlighted a considerable spread in the Non-Profit category, with a few notable outliers that extended far beyond the upper quartile,
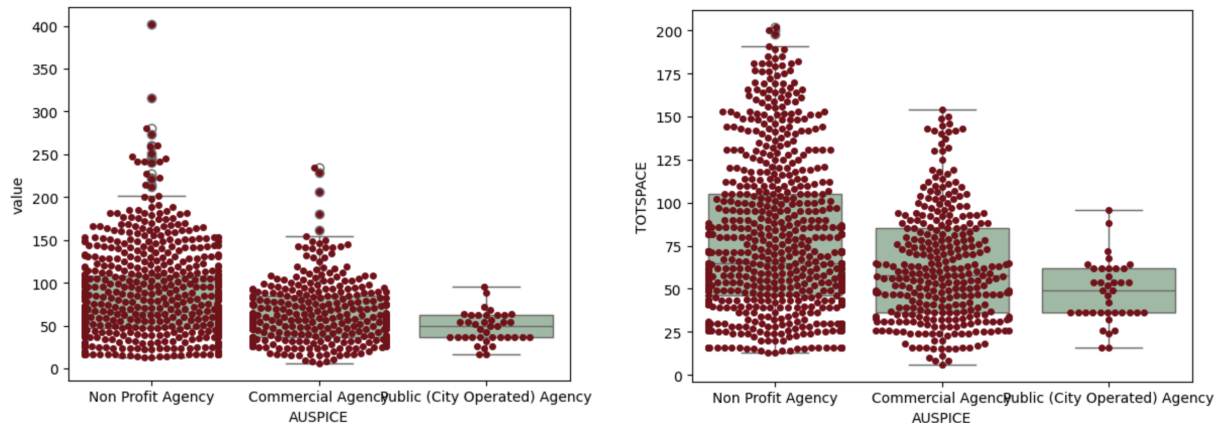
suggesting a diversity of centre sizes under this auspice. The Commercial and Public categories exhibited a more compact interquartile range but still showed variability worthy of further examination.

Post-cleansing, the second box plot painted a more precise picture. With the outliers removed, the total space distributions within each auspice category were more compact, and the extremes were significantly curtailed. This refinement allowed for a clearer comparative analysis between the auspice types, with reduced variance and a more standardized representation of centre capacities.

|	_id	LOC_ID	ward	IGSPACE	TGSPACE	PGSPACE	KGSPACE	SGSPACE	TOTSPACE
count	1063.000000	1063.000000	1063.000000	1063.000000	1063.000000	1063.000000	1063.000000	1063.000000	1063.000000
mean	532.000000	8087.888053	12.511759	3.896519	11.600188	24.258702	14.257761	21.661336	75.674506
std	307.005972	5151.252693	7.032050	6.091782	12.088152	18.577416	20.493290	30.423503	47.816518
min	1.000000	1013.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	6.000000
25%	266.500000	1862.000000	6.000000	0.000000	0.000000	16.000000	0.000000	0.000000	43.000000
50%	532.000000	8826.000000	12.000000	0.000000	10.000000	24.000000	0.000000	0.000000	62.000000
75%	797.500000	13245.000000	19.000000	10.000000	15.000000	32.000000	26.000000	30.000000	97.000000
max	1063.000000	14504.000000	25.000000	30.000000	90.000000	144.000000	130.000000	285.000000	402.000000

The summary table also shows a broad range in TOTSPACE, from a minimum of 6 to a maximum of 402, indicating a significant disparity in the total capacity of child care centres. The median TOTSPACE is 62, which may be reflective of the space requirements for a typical child care centre in Toronto. The standard deviation in the TOTSPACE is quite high at approximately 47.82, hinting at the presence of child care centres with unusually high or low total capacities. The maximum value of TOTSPACE being 402, which is substantially higher than the 75th percentile (97), indicates the existence of outliers which might significantly impact the average and suggest that a few centres have an exceptionally high capacity.

One Way ANOVA

Analyzing the ANOVA results and Tukey's post-hoc tests, we find significant insights into our research questions about the impact of operating auspice on the total space of child care centers. The ANOVA results before and after data cleansing show that the operating auspice (AUSPICE) has a statistically significant effect on the total space (TOTSPACE).

Comparing the two sets of diagnostic plots before(left) and after (right) data cleansing, we can conclude that the data cleaning process had a significant impact on the normality of the residuals. Before cleansing, there is a left skewed distribution with residuals shifts to the right, which is less ideal for ANOVA that assumes normally distributed residuals.The right diagram of after cleaning suggests that the cleaned data was reasonably well-behaved with only minor deviations from normality.
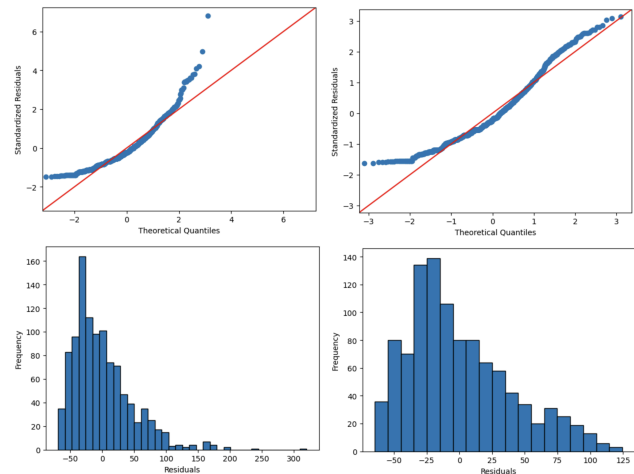
ANOVA Results:

The ANOVA results indicate that the differences in mean TOTSPACE across the levels of AUSPICE are statistically significant, as demonstrated by very low p-values (p < .00001) in both the original (on the left) and the cleansed datasets (on the right). This strong significance holds despite the slight difference in the sum of squares and F-statistic before and after cleansing,

	sum_sq	df	F	PR(>F)		sum_sq	df	F	PR(>F)
C(AUSPICE)	9.611211e+04	2.0	21.843051	5.057716e-10	C(AUSPICE)	7.060043e+04	2.0	22.292641	3.317236e-10
Residual	2.332065e+06	1060.0	NaN	NaN	Residual	1.642081e+06	1037.0	NaN	NaN

which is attributed to the removal of outliers and potential noise in the data.

Tukey's HSD Results:

Tukey's post-hoc test provides further granularity by comparing the means between each pair of auspice categories. Both before (top table) and after (bottom table) data cleansing, we observe consistent patterns:

```
            Multiple Comparison of Means — Tukey HSD, FWER=0.05
===============================================================================
     group1                    group2          meandiff p–adj   lower    upper  reject
-------------------------------------------------------------------------------
Commercial Agency            Non Profit Agency  14.7283    0.0    8.377 21.0796   True
Commercial Agency Public (City Operated) Agency –15.0441 0.0671 –30.8957  0.8076  False
Non Profit Agency Public (City Operated) Agency –29.7724    0.0 –45.1478 –14.397   True
-------------------------------------------------------------------------------
```

```
                  Multiple Comparison of Means - Tukey HSD, FWER=0.05
================================================================================
      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) Agency   -34.3346    0.0 -52.4448 -16.2244   True
--------------------------------------------------------------------------------
```

There is a significant difference in TOTSPACE between Non Profit agencies and Commercial agencies, as well as between Nonprofit and Public (City Operated) agencies, with Non Profit agencies typically having more space. The difference between Commercial agencies and Public (City Operated) agencies is not statistically significant, which is indicated by the 'False' in the 'reject' column of Tukey's output. This suggests that Commercial and Public (City Operated) agencies have comparable total spaces.

Conclusions:

By comparing the ANOVA and Tukey's results before and after data cleansing,, we can reject the null hypothesis (H0) that there is no difference in total space available across different operating auspices. The alternative hypothesis (H1) that there is a significant difference in total space available across different operating auspices is supported.

These results substantiate our research question by confirming that the operating auspice indeed affects the total space of child care centers in Toronto. This conclusion holds true even after cleansing the data, which implies that the detected effect is robust against potential outliers and strengthens the validity of our analysis.

Two Way ANOVA
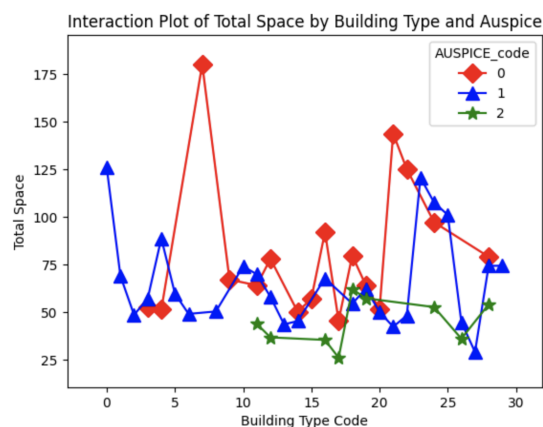
To further analyze and answer the research question: Does the interaction between operating auspice (AUSPICE) and building type (bldg_type) have a significant effect on the total space available (TOTSPACE) in child care centres? The interaction plot illustrates how the effect of building type on total space varies across the three categories of operating auspice. Non-parallel lines indicate that the interaction between these two factors is significant, meaning the influence of building type
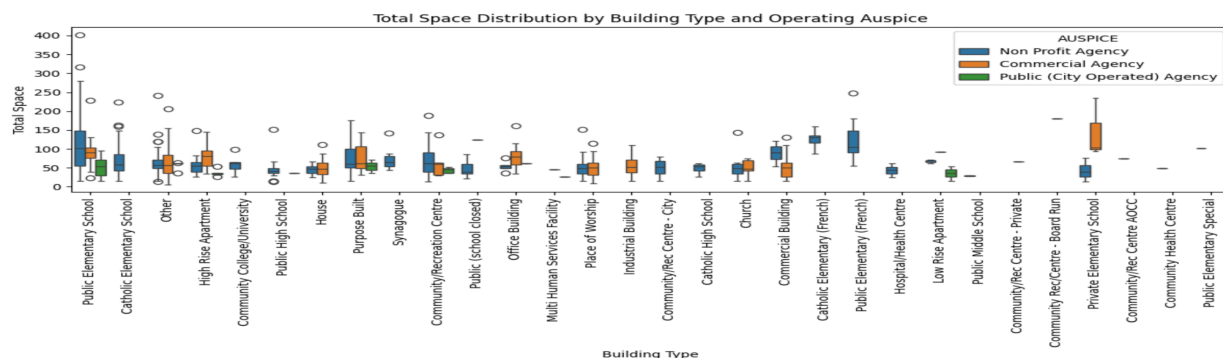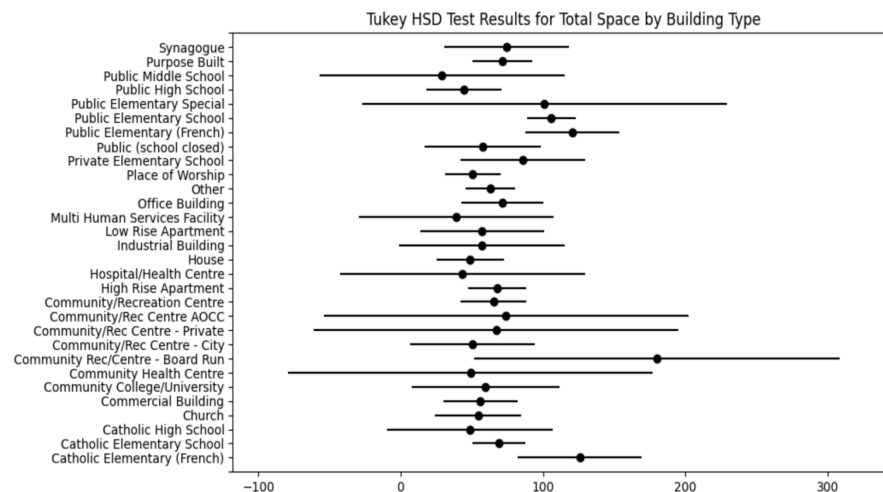
on TOTSPACE changes depending on the operating auspice. Certain building types show a more pronounced variation in TOTSPACE among the different operating auspices, while others exhibit more uniformity across auspices.

The box plot displays the distribution of total space (TOTSPACE) for child care centers across different building types (bldg_type), grouped by operating auspice (AUSPICE). The spread and median of the data points within each building type category vary by operating auspice, suggesting that TOTSPACE is influenced not just by the building type, but also by the management under different auspices. The presence of outliers in several building type categories indicates that there are centers with either exceptionally high or low TOTSPACE, which could impact the average space calculations and potentially skew the analysis.

The ANOVA results below show that the factor "building type" has a highly significant effect on the total space (TOTSPACE) available in child care centers ($p < .00001$). However, the effect of the "operating auspice" (AUSPICE) alone is not significant (p-value ~ 1). Most importantly, the interaction between building type and operating auspice is significant ($p < .00001$), suggesting that the influence of building type on total space varies depending on the

```
                            sum_sq      df             F         PR(>F)
C(bldg_type)             9.582108e+05    29.0  1.819840e+01   2.921140e-28
C(AUSPICE)              -2.755934e-08     2.0 -7.589427e-12   1.000000e+00
C(bldg_type):C(AUSPICE)  2.622099e+05    58.0  2.489954e+00   6.381552e-06
Residual                 1.835612e+06  1011.0           NaN            NaN
```


Tukey HSD Test Results for Total Space by Building Type

operating auspice.

Given the significance of the interaction term in the ANOVA results, we have enough evidence to reject the null hypothesis (H0) and accept the alternative hypothesis (H1). The implication is that the total space available in child care centers is influenced by a combination of building type and operating auspice, and their effect cannot be considered independently.

Conclusion

The one way and two way ANOVA models and the corresponding plots support the conclusion that when planning for or evaluating the **total space** of child care centers, it is not sufficient to consider **building type** or **operating auspice** in isolation. The interaction of these factors significantly influences the space available, which is crucial information for stakeholders in child care center administration, policy-making, and infrastructure development.