Examining Limited Capacities in Toronto's Childcare Facilities

1. Introduction

In the recent few years, the Government of Ontario faced the difficulties of childcare for both licensed and unlicensed. The challenges like exorbitant fees and limited spaces for children. A stark reality underscores this challenge, emphasized by Toronto's Children's Services that about 75% of families have the financial constraints to afford the childcare costs. The government has pledged to create an additional 100,000 spaces for childcare accommodation. This pressing issue led us to understand the financial difficulties faced by families and the limited capacity within Toronto's Children's Services.

In this report, we will provide a comprehensive one-way and two-way ANOVAs analysis of Toronto licensed childcare centers updated in February 2024 with a goal to discover whether the different type of buildings or agencies has the significant differences of the Total Spaces. To analyze these patterns, we delve into the dataset named "INF2178_A2_data.xlsx" (accessible in the GitHub repository, handling the list of different centers, demographic information, locations, operation, and capacity of these centers for multiple age groups.

Our exploration will focus on three fundamental research questions, serving as the guidance for examining the differences in the total spaces of Toronto Childcare. Also, we will discuss the relevant suggestions about the government policymaking and thoughts.

- **Research Question 1**: Are there significant differences in the Total Spaces of different types of Auspices? If so, what kind of advice you can give to government?
- **Research Question 2**: Are there significant differences of the Total Spaces in the 5 different categories of buildings? Any recommendations to the policymakers?
- **Research Question 3**: Whether different type of Auspices and building categories had a significant difference on the Total Spaces in childcare centers, any interaction between these two factors?

By addressing these research questions, we need to explore deeper into the dynamics of childcare spaces in Toronto and provide with the insightful understanding of our dataset.

2. Data Cleaning and Wrangling

The raw dataset has the total **17 columns** with **1063** entities (**rows**). After loading and reviewing of the dataset, we thought the overall data quality is good and the data cleaning process is also necessary for the further analysis. We observed that there is 2 ID (_id and LOC_ID), 6 Numeric Variables and 9 Categorical variables. We also

outlined the basic observations and all new features added to our dataset:

A. Observation and Considerations:

a). This is a short description of each column of **6 Numerical variables** below:

- **IGSPACE**: childcare spaces for infants 0-18 months.
- **TGSPACE**: childcare spaces for toddlers 18-30 months.
- **PGSPACE**: childcare spaces for preschoolers 30 months up until they enter grade one.
- **KGSPACE**: childcare spaces for children in full-day kindergarten.
- **SGSPACE**: childcare spaces for children grade one and up.
- **TOTSPACE**: childcare spaces for all age groups.

b). Below is the short description of **9 Categorical Variables**:

- **Name and Address**: LOC_NAME, ADDRESS, PCODE, BLDGNAME (many different address information/details)
- **AUSPICE**: Operating auspice (Commercial, Non-Profit, or Public)
- **Ward**: City ward number (1-25)
- **Bldg_type**: Type of building (30 types)
- **Subsidy**: Center has a fee subsidy contract or not (Yes/No)
- **Cwelcc_flag**: Space participates in CWELCC or not (Yes/No)

c). The column '**BLDGNAME**' have **348 missing values** in total. And filled with the words "Not Available" into null values.

B. Feature Engineering:

In this part, we created one feature into our dataset to help in later analysis. The feature is shown below:

Bldg_category: this is a new added column which classifies the total 30 type of buildings into 5 different categories describe below:

- **Schools**: For the words like "School", "University", and "College" in the column of bldg_type
- **Residential Buildings**: For the words like "Apartment", and "House" in the column of bldg_type
- **Community and Recreation Centers**: For the words like "Community", "Recreation" in the column of bldg_type
- **Places of Worship**: For the words like "Worship", "Church", and "Synagogue" in the column of bldg_type
- **Commercial and Other Buildings**: the remain type of buildings in the column of bldg_type

3. Total Spaces across Different Types of Auspices

Research Question #1: Are there significant differences in the Total Spaces of different types of Auspices? If so, what kind of advice you can give to government?

For this analysis, we aimed to first investigate any significant difference of Total Spaces in response to three type of Auspices and whether the different agencies have the differences. In the *Figure 2*, we can observe the Commercial Agency and Public (City

Operated) Agency do not have the significant differences. Secondly, as we can see from the ANOVA table, p-value is very small (<0.001). So, we need to reject the null hypothesis that we have the sufficient evidence that there has the significant difference of Total Spaces with different type of Auspices.

	df	F	P-value
C(AUSPICE)	2	21.843	0.0000001
Residual	1060	NaN	NaN

Figure 1: ANOVA Table of Total Spaces by Auspices

Group 1	Group 2	meandiff	P-value	reject
Commercial Agency	**Non-Profit Agency**	17.119	0.001	TRUE
Commercial Agency	**Public (City Operated) Agency**	-17.215	0.078	FALSE
Non-Profit Agency	**Public (City Operated) Agency**	-34.335	0.001	TRUE

Figure 2: Multiple Comparison of Means (alpha level at 0.05)

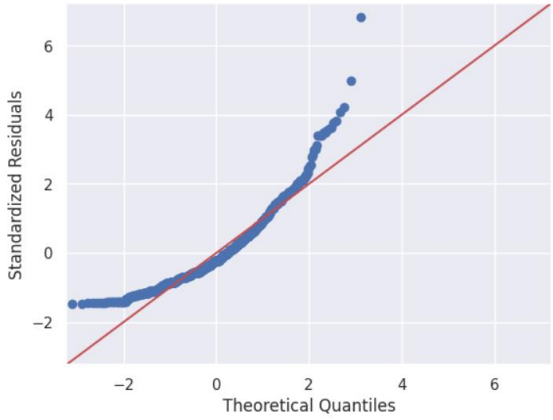

Figure 3: QQ-plot shows the sample data overall distribution (Check normality)

While we can see the QQ-plot above and Shapiro-Wilk test (w-value=0.902) that the sample data **does not adhere to a normal distribution**, as indicated in *Figure 3*. According to Levene's test, we know that two assumptions are not satisfied with our data.

Auspices	Total Spaces
Commercial Agency	64.994
Non-Profit Agency	82.104
Public (City Operated) Agency	47.769

Figure 4: Table of Average (mean) Total Spaces by Different Auspices

To expanding the understanding of the space's variation, we calculated the mean Total Spaces by each type of Auspices (*Figure 4*). Based on the results, it seems that **Non-Profit Agency** have a **higher Total Spaces than other agencies**.

There is a **significant difference** between the Total Spaces offered by Non-Profit Agency and other agencies. The Non-Profit Agency provides more spaces on average. Therefore, government can balance the allocation of limited resources, for example, adjusting the fundings for different types of childcare centers to ensure that all children have access to high-quality services; Given the special role and challenges of Non-Profit Agencies in providing children care services, the government should provide these agencies with additional support to help them expand their services to better meet the needs of more families.

4. Total Spaces across Building Category

Research Question #2: Are there significant differences of the Total Spaces in the 5 different categories of buildings?

Similar to the method used in the first research question, when comparing different types of building category with the Total Spaces, our **ANOVA Table** suggested a statistically significant difference (with alpha = 0.05) between the building category and Total Spaces.

	sum_sq	df	F	P-value
C(bldg_category)	0.00001	4	27.443	0.001
Residual	0.00001	1058	NaN	NaN

Figure 5: ANOVA Table of Total Spaces by bldg_catgory

	w-value	p-value
Shapiro-Wilk Test	0.923	0.001

Figure 6: Shapiro-Wilk test

Parameter	Value
Test statistics (W)	25.732
Degrees of freedom (Df)	4
p value	0.001

Figure 7: Levene's test for check assumption of "Homogeneity of Variances"

Recognizing our data is **violate the normality and homogeneity of variances**, we can see the results from Shapiro-Wilk test (*Figure 6*) and Levene' test (*Figure 7*).

bldg_category	TOTSPACE
Commercial and Other Buildings	*67.199*
Community and Recreation Centers	*65.358*
Places of Worship	*52.819*
Residential Buildings	*60.667*
Schools	*91.947*

Figure 8: Table of Average (mean) Total Spaces by Different Building Category

Based on the results showing on *Figure 8*, Building category – **Schools** has a **higher Total Spaces (91.947) than other categories of childcare building**. The Total Spaces of Schools are the largest part. Policymaker can divert resources or subsidies to Schools. Or consider how other categories of building can be put forward to encourage community centers, places of worship and other buildings to be compatible with childcare services, especially open after school or on the weekends.

5. Total Spaces with Auspices and Building Category (Two-way ANOVA)

Research Question #3: Whether different type of Auspices and building categories had a significant difference on the Total Spaces in childcare centers, any interaction between these two factors?

	df	F	p-value
C(AUSPICE)	2	4.74	0.030
C(bldg_category)	4	24.13	0.001
C(AUSPICE):C(bldg_category)	8	1.1	0.361
Residual	1049	NaN	NaN

Figure 9: ANOVA Table Summary

In this *Figure 9*, we can observe the p-value of Auspices and building category is smaller than the alpha level at 0.05. So, we need to reject the null hypothesis and we have sufficient evidence that there has **statistically significant of Auspices and building category with the Total Spaces.** However, for the interactions between these two factors. Since the P-value (0.361) is greater than the significance level at 0.05, we fail to reject the null hypothesis. Therefore, we do not have sufficient evidence to claim that there is a statistical significance interaction between "AUSPICE" and "bldg_category". This does not imply that there is no interaction between these two factors; rather, based on our data, we cannot establish enough evidence to assert the presence of a significant interaction.

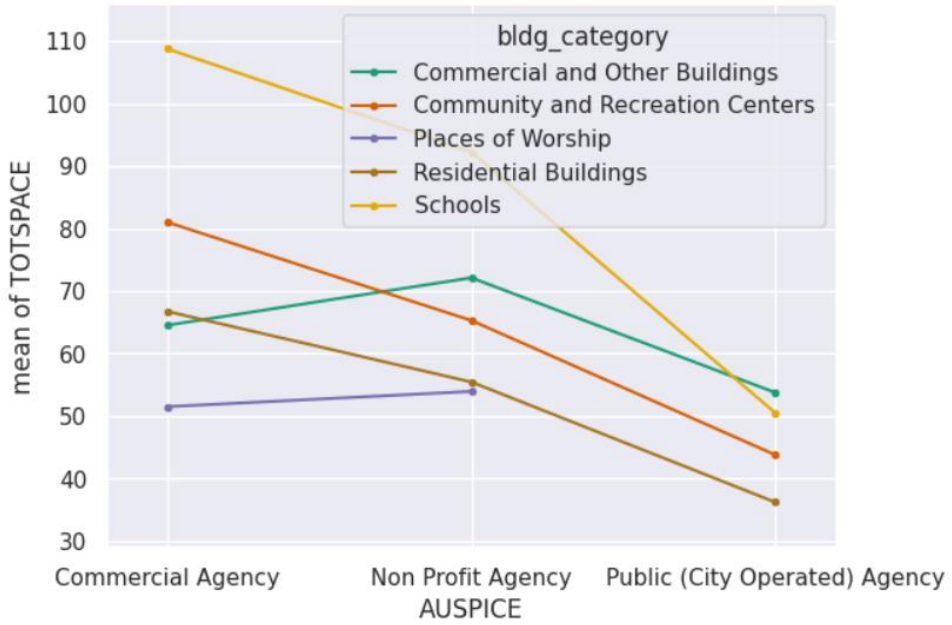

Figure 10: Interaction Plot of Average (mean) Total Spaces by Auspices and Building Category

In response to these findings, governments may need to consider other factors or variables, in understanding and improving the space capabilities and quality of childcare services.

6. Conclusion

Through the quantitative, one-way, and two-way ANOVA analysis, we gained the valuable insights about the spaces distributed within Toronto Childcare Service. Our findings displayed the impact of different types of Auspices, different categories of buildings, and any interaction effect of these two factors. Hopefully, the suggestions can guide policymakers in allocating limited resources more effective of childcare spaces that they can align with the growing needs of families. Although the analysis of one-way ANOVA assumptions is violated, there has a potential avenue for future improvement like data transformation. Given the normality assumption is not satisfied, you may consider transforming the data using logarithmic or square root transformations to make the distribution more normal, which can derive even more awesome insights from the data.