

STATISTICAL ANALYSIS OF TORONTO CHILDCARE CENTRES

INTRODUCTION

Data from Statistics Canada shows that “while childcare is getting more affordable for parents ... finding it is getting more challenging.”¹ This report analyzes a dataset consisting of 1063 rows to draw conclusions on the availability of childcare spaces.

RESEARCH QUESTIONS

1. Does the operating auspice impact the total number of childcare spaces available in Toronto? In more statistical terms, does the operating auspice have a statistically significant effect on the total number of childcare spaces in Toronto?
2. Does the operating auspice and building class of childcare facilities independently impact the total number of childcare spaces available in Toronto? Do these two factors impact the number of childcare spaces in Toronto interactively?

DATA CLEANING AND DATA WRANGLING

The dataset is generally clean and requires few modifications. The dataset consists of 17 **columns** and 1063 **rows**. The data dictionary reveals that **LOC_ID** is a unique identifier. We can use this to see if the data has any duplicate records. Doing this reveals that there are no duplicates in the data. We made some minor changes by renaming the columns for the sake of stylistic consistency. We also converted all columns with continuous variables to the number format. We also checked the dataset for missing values and only the **BUILDING_NAME** column contains **348** missing values. Since, this column is not useful for our analysis we are not concerned about the missing values.

There was no feature engineering involved as the overall centre capacity numbers were already provided to us in the column labelled **TOTSPACE** we subsequently renamed this column in our data frame to **TOTAL_SPACE**.

TABLE 1: DATA COLUMNS AND DESCRIPTIONS

COLUMN NAME IN DATA FRAME	DESCRIPTION
ID	Unique row identifier for Open Data database
LOC_ID	Unique identifier
LOC_NAME	Name of the childcare centre
AUSPICE	Operating auspice (Commercial, Non-Profit or Public)
ADDRESS	Address street number, street name, street type, street direction, and unit
POSTAL_CODE	Address postal code
WARD	City ward number
BUILDING_TYPE	Type of building
BUILDING_CLASS	Groups different building types into one of five categories (School, Other, Apartment, Place of Worship, Community Centre)
BUILDING_NAME	Name of the building the childcare centre is located in
IG_SPACE	Childcare spaces for infants 0-18 months
TG_SPACE	Childcare spaces for toddlers 18-30 months
PG_SPACE	Childcare spaces for preschoolers 30 months up until they enter grade one
KG_SPACE	Childcare spaces for children in full-day kindergarten
SG_SPACE	Childcare spaces for children grade one and up
TOTAL_SPACE	Childcare spaces for all age groups
SUBSIDY	Centre has a fee subsidy contract (Yes/No)
CWELCC_FLAG	“Y” indicates space participates in CWELCC, “N” indicates it does not

¹ <https://www.ctvnews.ca/canada/parents-finding-daycare-more-affordable-now-but-that-doesn-t-mean-they-can-find-it-data-1.6674068>

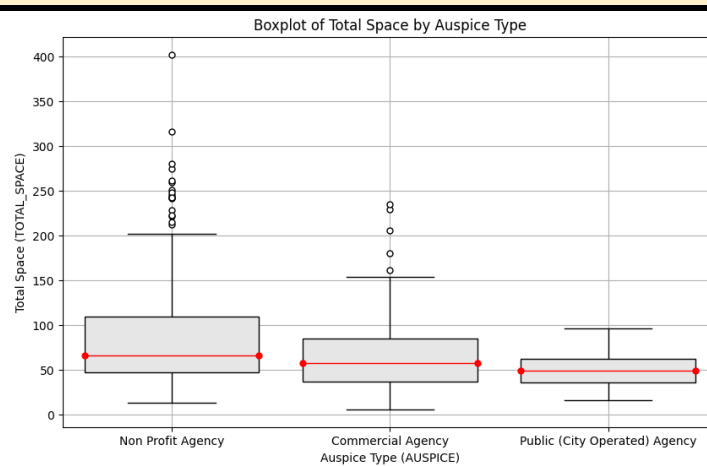
DESCRIPTIVE STATISTICS AND EXPLORATORY DATA ANALYSIS

TABLE 2: DESCRIPTIVE STATISTICS OF CONTINUOUS VARIABLES

COLUMN NAME	COUNT	MEAN	STANDARD DEVIATION	MIN	25%	50%	75%	MAX
IG_SPACE	1063	3.896519285	6.091782085	0	0	0	10	30
TG_SPACE	1063	11.60018815	12.08815236	0	0	10	15	90
PG_SPACE	1063	24.25870179	18.5774162	0	16	24	32	144
KG_SPACE	1063	14.25776105	20.49328983	0	0	0	26	130
SG_SPACE	1063	21.66133584	30.42350292	0	0	0	30	285
TOTAL_SPACE	1063	75.67450611	47.81651783	6	43	62	97	402

Let us begin our analysis with a boxplot of **TOTAL_SPACE** by **AUSPICE** (Auspice Type), this will give us a good idea of variation in median values, and the spread of the data.

FIGURE 1: BOXPLOT OF TOTAL SPACE BY AUSPICE TYPE



The boxplot in **FIGURE 1** shows an apparent difference in the median for total space with respect to auspice type. There is also significant spread in terms of total capacity with respect to each auspice type. Furthermore, there are some significant outliers for Non-Profit Agency auspice type. Taken together this suggests that there might be statically significant differences in total capacity with respect to auspice type.

TABLE 3: HYPOTHESES (USING A SIGNIFICANCE VALUE OF 0.05)

RESEARCH QUESTION 1

- H_{null} : There is no difference in the mean total space available between childcare centres operated under the three different auspices, namely (Non-Profit Agency, Commercial Agency, and Public (City Operated) Agency)
- $H_{alternative}$: There is a statistically significant difference in the mean total space available between childcare centres operated under the three different auspices, namely [Non-Profit Agency, Commercial Agency, and Public (City Operated) Agency]

CHECKING ASSUMPTIONS FOR A ONE-WAY ANOVA

TABLE 4: SHAPIRO-WILK TEST FOR NORMALITY

AUSPICE	SHAPIRO STATISTIC	P-VALUE	EXACT P-VALUE
Non-Profit Agency	0.8917580947016522	$p < 0.001$	$8.3401340322369425 \times 10^{-22}$
Commercial Agency	0.9238312522211777	$p < 0.001$	$1.0137738663238549 \times 10^{-11}$
Public (City Operated) Agency	0.951491410440594	$p > 0.05$ (0.09)	0.09225924352564932

Our results reveal that both total spaces for childcare centres under the auspice of non-profit agencies and commercial agencies are not normally distributed. This is evidenced by the fact that **p-values** for both auspice types are **< 0.001**. On the other hand, we can see that the p-value for the public (city operated) agencies is **> 0.05**, meaning that the total spaces for this auspice type might follow a normal distribution. The results are tabulated in the **TABLE 4** above. Even

though the ANOVA test is robust in the face of data that is not normally distributed, it is good practice to conduct such tests.

TABLE 5: LEVENE TEST FOR EQUALITY OF VARIANCES

	LEVENE STATISTIC	P-VALUE	EXACT P-VALUE
LEVENE TEST	17.927068524236084	$p < 0.001$	$2.2036889276494363 \times 10^{-8}$

We also ran a Levene test to test for equality of variances. This is another pre-requisite for running an ANOVA test. According to our results tabulated in [TABLE 5](#) the p-value for our Levene test is **< 0.001**. This means that there is a statistically significant difference in the variances across the three groups.

Even though our data fails both the assumptions for the ANOVA test we can still run the **ANOVA**, but we have to be cautious of how we interpret the results.

TABLE 6: ONE-WAY ANOVA TABLE

	SUM_SQ	DEGREES OF FREEDOM	F-STATISTIC	P-VALUE	PR(>F)
AUSPICE	96112.11429024908	2.0	21.843051025265588	$p < 0.001$	$5.057716322706566 \times 10^{-10}$
RESIDUAL	2332065.264825459	1060.0	NaN		NaN

The results of our ANOVA, tabulated in [TABLE 6](#), show that there are significant differences in dependent variable (**TOTAL_SPACE**) across the **AUSPICE** factor. This is indicated by a **p-value < 0.001**. We can also see from the table that the **F-Statistic** is quite large further strengthening the case against the **null hypothesis** we outlined in [TABLE 3](#). Let us run some post-hoc tests to see how well our model fares.

FIGURE 2: QQ PLOT OF ANOVA MODEL RESIDUALS

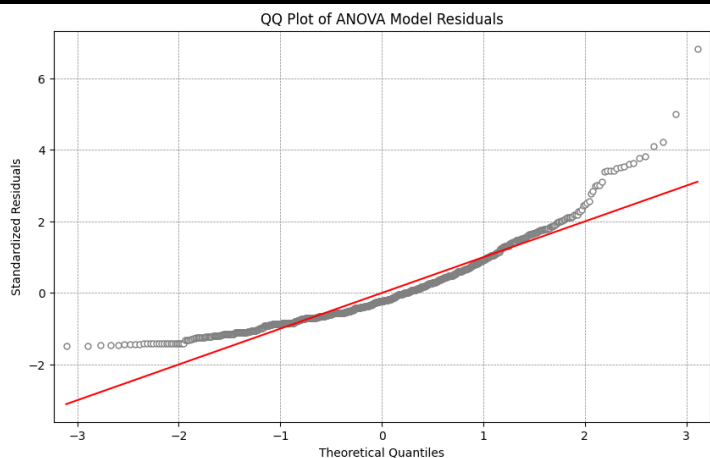
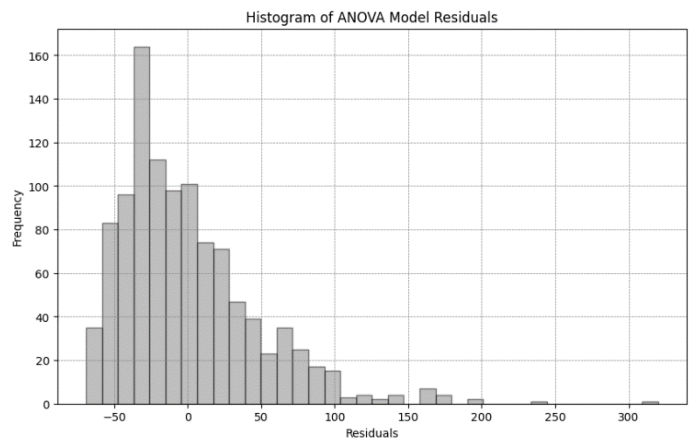


FIGURE 3: HISTOGRAM OF ANOVA MODEL RESIDUALS



We can see from [FIGURE 2](#) that our **ANOVA** model follows the red line closely with some curving on the edges. This suggests that there are some outliers in our data but overall, the residuals are almost normally distributed. The histogram in [FIGURE 3](#) reveals that the residuals are skewed to the right. Both these graphs suggest that results from our **ANOVA** might not be very robust.

TABLE 7: RESULTS OF TUKEY'S HSD

MULTIPLE COMPARISON OF MEANS, FWER=0.05						
GROUP 1	GROUP 2	MEAN DIFFERENCE	ADJUSTED P-VALUES	LOWER	UPPER	REJECT
Commercial	Non-Profit	17.1194	0.0	9.7037	24.5351	True
Commercial	Public (City-Operated)	-17.2152	0.0779	-35.8832	1.4528	False
Non-Profit	Public (City-Operated)	-34.3346	0.0	-52.4448	-16.2244	True

The results of the Tukey HSD in [TABLE 7](#) above reveal that there is a statistically significant difference in the mean **total spaces** between childcare centres under the **Commercial** auspice versus the **Non-Profit** auspice. We can see this because the **p-value = 0.0**. Furthermore, the difference in mean **total spaces** is **NOT** significant between childcare centres under the **Commercial** auspice versus the **Public (City-Operated)** auspice as indicated by **p-value > 0.05**. Finally, there is a

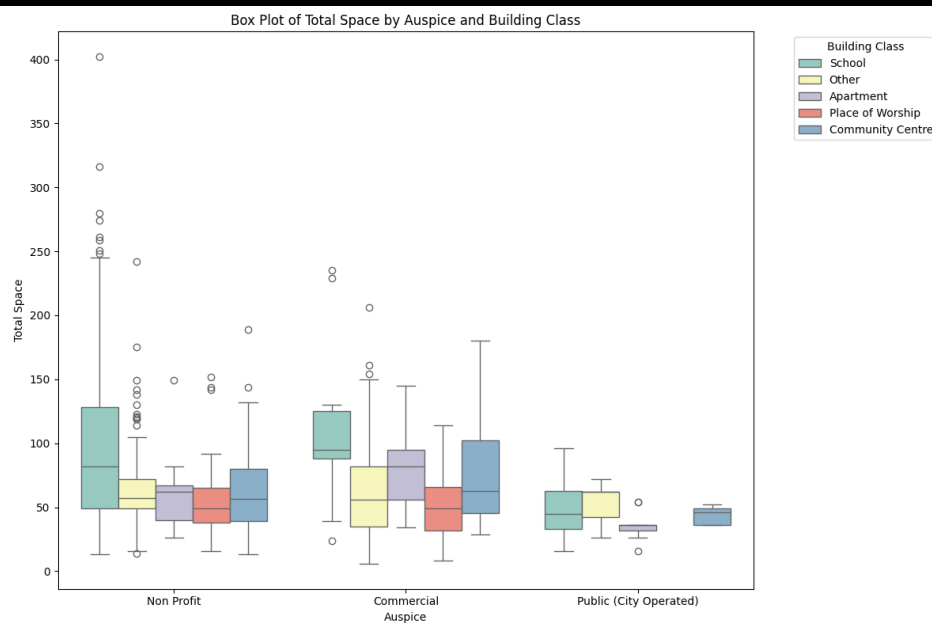
statistically significant difference in the mean **total spaces** between childcare centres under the **Non-Profit** auspice versus the **Public (City-Operated)** auspice as indicated by the **p-value = 0.0**. The results of this one-way **ANOVA** should be taken with caution given that the assumptions for conducting a one-way **ANOVA** have not been met. Having said that we can conclude that in certain cases – specifically **Commercial** versus **Non-Profit** and **Non-Profit** versus **Public (City-Operated)** both in favour of **Non-Profit** – the auspice type impacts the total number of childcare spaces available.

TABLE 8: HYPOTHESES (USING A SIGNIFICANCE VALUE OF 0.05)

RESEARCH QUESTION 2	
MAIN EFFECT OF AUSPICE TYPE	
H_{null1}	: There is no difference in the mean total space available between childcare centres operated under the three different auspices, namely (Non-Profit Agency, Commercial Agency, and Public (City Operated) Agency)
$H_{alternative1}$: There is a statistically significant difference in the mean total space available between childcare centres operated under the three different auspices, namely [Non-Profit Agency, Commercial Agency, and Public (City Operated) Agency]
MAIN EFFECT OF BUILDING CLASS	
H_{null2}	: There is no difference in the mean total space available between childcare centres operated in the five difference building classes, namely (Schools, Places of Worship, Community Centres, Apartments, and Other).
$H_{alternative2}$: There is a statistically significant difference in the mean total space between childcare centres operated in the five difference building classes, namely (Schools, Places of Worship, Community Centres, Apartments, and Other).
INTERACTION EFFECT BETWEEN AUSPICE AND BUILDING CLASS	
H_{null12}	: There is no statistically significant interaction effect between auspice and building class on the mean total childcare spaces available.
$H_{alternative12}$: There is a statistically significant interaction effect between auspice and building class on the mean total childcare spaces available.

To set ourselves up to conduct this two-way **ANOVA** in a reasonable manner we are going to group categories across **BUILDING_TYPE**. There are currently 30 unique building types but many of these can be groups together. We will add an additional column labelled **BUILDING_CLASS** and classify each **BUILDING_TYPE** using a class label. This will allow us to condense the categories and have a more manageable and meaningful way to analyze the data.

FIGURE 4: BOXPLOT OF TOTAL SPACE BY AUSPICE TYPE



The boxplot shown above in **FIGURE 4** reveals that there is considerable variation in the number of total childcare spaces available across **building class**. **Non-Profit** agencies appear to have the most spaces in **Schools** and **Community Centres**. There are no childcare spaces in a place of worship operated under a public auspice type. While this is not something to be concerned about it is an interesting observation. The variation in median values for total space and the spread of the

total spaces across building class reveals that there might be a significant effect of building type on the number of total childcare spaces available.

CHECKING ASSUMPTIONS FOR A TWO-WAY ANOVA

TABLE 9: SHAPIRO-WILK TEST FOR NORMALITY

	VALUE	EXACT VALUE
SHAPIRO STATISTIC	~ 0.93	0.928993915514909
P-VALUE	$p < 0.05$	$5.020620106586926 \times 10^{-22}$

TABLE 10: LEVENE TEST FOR EQUALITY OF VARIANCES

	VALUE	EXACT VALUE
LEVENE STATISTIC	~ 10.6	10.634629477680763
P-VALUE	$p < 0.05$	$1.63118605475703 \times 10^{-21}$

Once again, the tables above reveals that our data does not satisfy the normality assumption for the two-way ANOVA. Even though we have a **Shapiro statistic** close to 1 we also have a **p-value** < 0.05 . Similarly, our **Levene test** has a **p-value** < 0.05 which also confirms that our data violates the homogeneity of variances which is another assumption for conducting an ANOVA test. Despite the fact that our data fails to meet both assumptions we are going to proceed with the two-way ANOVA test as the test is fairly robust.

TABLE 11: TWO-WAY ANOVA TABLE

	SUM_SQ	DEGREES OF FREEDOM	F-STATISTIC	P-VALUE	PR(>F)
AUSPICE	3.470572×10^2	2.0	0.085732	$p > 0.05$	7.697322×10^{-01}
BUILDING_CLASS	2.527908×10^5	4.0	31.223053	$p < 0.05$	2.463033×10^{-19}
AUSPICE:BUILDING_CLASS	1.914627×10^4	8.0	1.182410	$p < 0.05$	2.463033×10^{-19}
RESIDUAL	2.123251×10^6	1049.0	NaN		NaN

The results of the TWO-WAY ANOVA displayed in TABLE 11 above indicate that BUILDING_CLASS is a significant factor in determining the number of childcare spaces. Our ONE-WAY ANOVA revealed that AUSPICE was a significant factor but the results from the TWO-WAY ANOVA reveal that the effect of AUSPICE is not significant when the effect of BUILDING_CLASS and the interaction between AUSPICE and BUILDING_CLASS is considered. Having said that we will need to conduct a Tukey HSD and construct an interaction plot before we can definitively discount the AUSPICE factor. This is because AUSPICE may have an impact for certain levels of BUILDING_CLASS.

FIGURE 5: QQ PLOT OF TWO-WAY ANOVA MODEL RESIDUALS

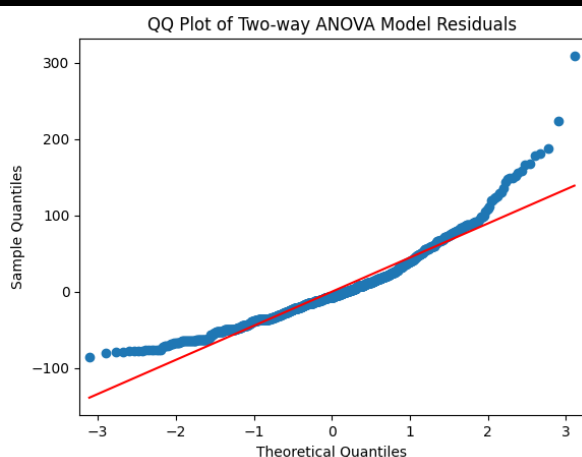
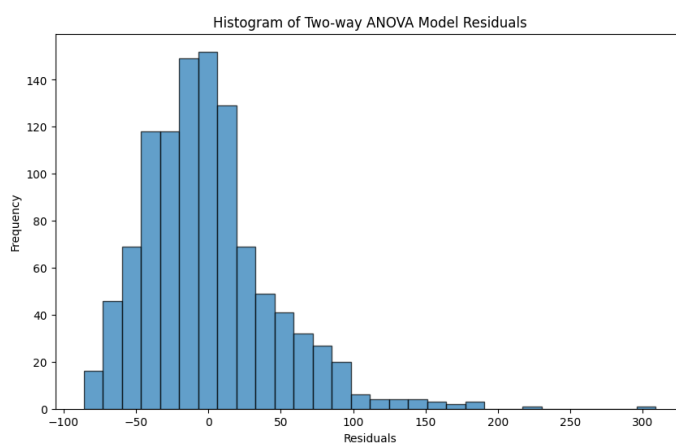


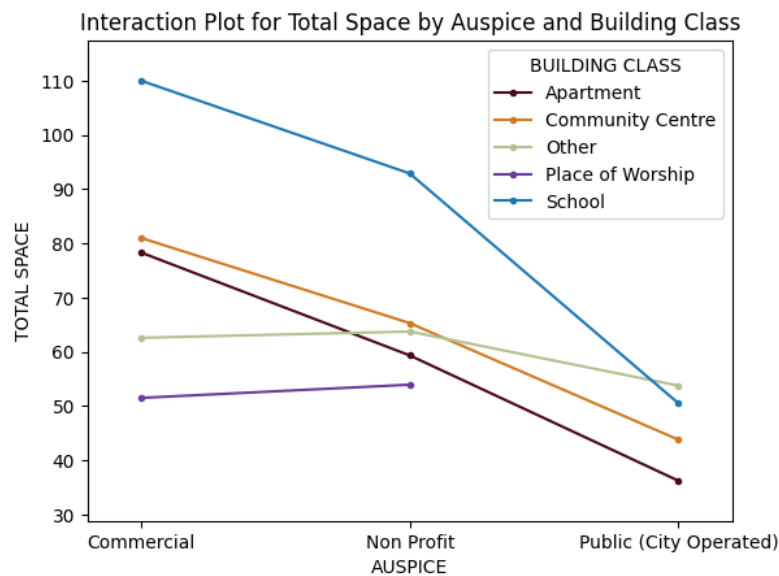
FIGURE 6: HISTOGRAM OF TWO-WAY ANOVA MODEL RESIDUALS



We can see from the QQ Plot and the Histogram in FIGURES 5 & 6 above that the residuals are not normally distributed, and this violates the normality assumption for an ANOVA. These plots suggest that the results from our ANOVA may not be very robust. The results of the ANOVA in TABLE 11 show that AUSPICE is not a significant factor in the total number of childcare spaces available. However, BUILDING_CLASS is a significant factor. We can also see that the interaction effect between AUSPICE and BUILDING_CLASS is negligible despite the fact that $p < 0.05$. The reason for this is that F-STATISTIC which explains how much of the variation is a result of the interaction is very low. We also ran a Tukey HSD on the significant factor (BUILDING_CLASS) to see what the levels of this factor can tell us about the number of total childcare spaces available in Toronto.

TABLE 12: RESULTS OF TUKEY'S HSD

MULTIPLE COMPARISON OF MEANS - TUKEY HSD, FWER=0.05							
GROUP1	GROUP2	MEANDIFF	P-VALUE	P-ADJ	LOWER	UPPER	REJECT
Apartment	Community Centre	-1.1825	$p > 0.05$	0.9999	-24.9041	22.5392	False
Apartment	Other	-3.9391	$p > 0.05$	0.9664	-20.4695	12.5912	False
Apartment	Place of Worship	-13.7208	$p > 0.05$	0.3007	-33.1215	5.68	False
Apartment	School	26.1185	$p < 0.05$	0.0001	9.9125	42.3244	True
Community Centre	Other	-2.7567	$p > 0.05$	0.9951	-22.089	16.5756	False
Community Centre	Place of Worship	-12.5383	$p > 0.05$	0.5178	-34.3757	9.2991	False
Community Centre	School	27.3009	$p < 0.05$	0.0009	8.2453	46.3566	True
Other	Place of Worship	-9.7816	$p > 0.05$	0.2904	-23.4714	3.9081	False
Other	School	30.0576	$p < 0.05$	0.0	21.4755	38.6397	True
Place of Worship	School	39.8392	$p < 0.05$	0.0	26.543	53.1354	True

FIGURE 7: INTERACTION PLOT FOR TOTAL SPACE BY AUSPICE TYPE AND BUILDING CLASS

We can see from the **TUKEY HSD** ([TABLE 12](#)) and the **INTERACTION PLOT** ([FIGURE 7](#)) above that there is a significant interaction between **AUSPICE** and **BUILDING CLASS** at least for buildings classed as **SCHOOLS** and **COMMUNITY CENTRES**. It is interesting to note that the total number of childcare spaces is consistently higher under the **COMMERCIAL AUSPICE** this is true across **building classes**. There is also an interaction between **AUSPICE** and **BUILDING_CLASS** for the **APARTMENT** and **OTHER building class** as the lines for these two levels intersect. As mentioned in the introduction the cost of childcare in Toronto is stabilizing with childcare getting more affordable. However, there is still a shortage of childcare spaces. Perhaps, one way the government can increase the supply of childcare spaces is by empowering commercial providers to increase the supply. This can be done through subsidies and tax breaks. The majority of childcare spaces are available in **SCHOOLS**. With the most being available under the **COMMERCIAL** and **NON-PROFIT auspices**. An interesting way to solve the capacity problem might be to allow commercial childcare providers to utilize schools to deliver childcare.

LIMITATIONS

The **BUILDING_CLASS** column abstracts away the differences between various building types this makes it easier to run the **TWO-WAY ANOVA**. However, a more detailed analysis that directly uses the **BUILDING_TYPE** column might provide even greater insights into the interaction effects between **AUSPICE** and the **types of buildings** being used to provide childcare facilities.