

2178 Assignment 2

BoyanZhang

1005266338

1. Introduction:

This analysis seeks to contribute to the understanding of the current landscape of child care services in Toronto by examining data from licensed child care centers.

Our exploration will address three fundamental research questions, serving as guiding principles in discover the potential improvement for children in a safe and nurturing environment:

- **Research Question 1:** Does the auspice type (Non-Profit vs. Commercial) affect the total spaces (TOTSPACE) available in daycare centers in Toronto?
- **Research Question 2:** Is there a significant difference in the total spaces offered (TOTSPACE) among different building types (bldg_type) where daycare centers are located?
- **Research Question 3:** Is there an interaction effect between the auspice type (AUSPICE) and the presence of a subsidy (subsidy) on the total spaces (TOTSPACE) in daycare centers?

A critical element of this assignment involves the application of both one-way and two-way ANOVA tests to determine if there are statistically significant differences between groups within the dataset.

2. Data Cleaning and Consideration

- Observation and Consideration:

We are looking for knowledge of the quantitative analysis in python, The data consists of a pandas DataFrame with **1063 entries** that detail information on licensed child care centres in Toronto.

Brief description for each columns in the dataset:

1. **_id:** A unique identifier for each record in the dataset.
2. **LOC_ID:** The location ID for each child care centre.
3. **LOC_NAME:** The name of the child care centre.
4. **AUSPICE:** The management type of the centre, indicating whether it is a Non-Profit Agency or another entity.
5. **ADDRESS:** The street address of the child care centre.
6. **PCODE:** Postal code associated with the child care centre's location.
7. **ward:** The ward number where the child care centre is situated.
8. **bldg_type:** The type of building the child care centre is located in
9. **BLDGNAME:** The name of the building, if applicable.
10. **IGSPACE:** The number of infant care spaces.
11. **TGSPACE:** The number of toddler care spaces.
12. **PGSPACE:** The number of preschool care spaces.
13. **KGSPACE:** The number of kindergarten care spaces.
14. **SGSPACE:** The number of school-age care spaces.
15. **TOTSPACE:** The total number of care spaces available.
16. **subsidy:** Indicates whether the centre offers subsidized child care.

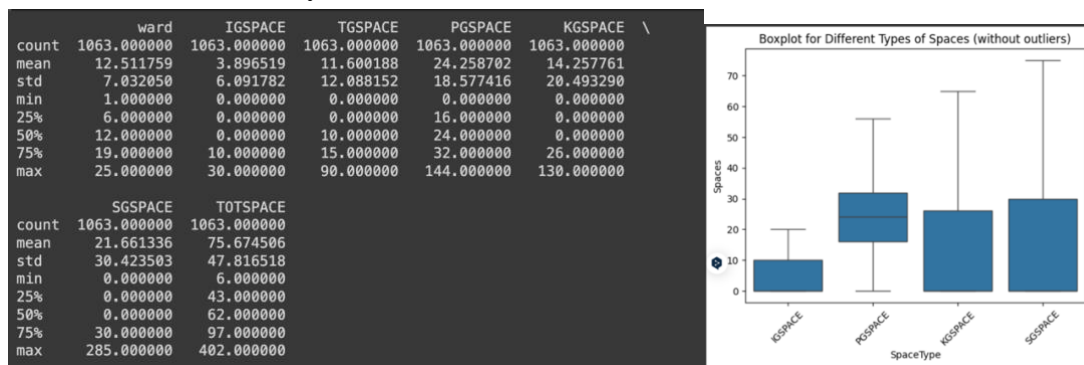
17. **cwelcc_flag**: Indicates whether the centre is part of the City of Toronto's Child Care Fee Subsidy program.

The count indicates there are no missing values for most of the columns, except for **BLDGNAME** which has **348 missing entries**. This suggests that not all child care centres have a building name associated with their record. However, we will use quantitative analysis in this assignment, so we don't need to clean or remove those data.

3. Preliminary Exploratory Data Analysis:

● Summary statistics for all numerical columns

From a statistical summary perspective, the average values for the number of spaces across different age groups (IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE) and the total spaces (TOTSPACE) are provided. And we will use those data to do the later analysis.



The boxplot displays the distribution of spaces across four different types of child care spaces after remove the outliers: Infant (IGSPACE), Toddler (TGSPACE), Kindergarten (KGSPACE), and School Age (SGSPACE). This phase of descriptive analysis provided a patterns across each attribute.

● Frequency distribution for categorical columns

Non Profit Agency	703
Commercial Agency	321
Public (City Operated) Agency	39
Name: AUSPICE, dtype: int64	
Public Elementary School	301
Other	227
Catholic Elementary School	113
Place of Worship	82
High Rise Apartment	59
Purpose Built	51
Community/Recreation Centre	35
House	33
Commercial Building	24
Public High School	23
Office Building	18
Church	16
Public Elementary (French)	13
Public (school closed)	8
Catholic Elementary (French)	7
Private Elementary School	7
Low Rise Apartment	7
Community/Rec Centre - City	7
Synagogue	7
Community College/University	5
Catholic High School	4
Industrial Building	4
Multi Human Services Facility	3
Hospital/Health Centre	2
Public Middle School	2
Community/Rec Centre - Private	1
Name: bldg_type, dtype: int64	
Y	718
N	345
Name: subsidy, dtype: int64	

The provided table presents an overview of child care centers in terms of their management type, facility location, and subsidy status.

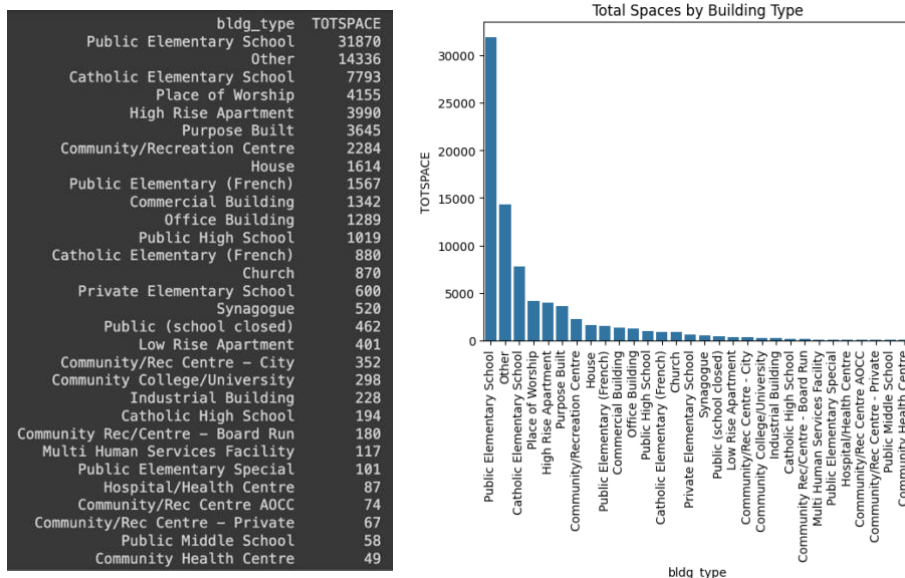
Most centres are run by Non-Profit Agencies (703), followed by Commercial Agencies (321), and a minority by Public (City Operated) Agencies (39).

These centres are predominantly located in public elementary schools (301) and a variety of other venues, including places of worship and high-rise apartments, reflecting a broad integration into community spaces.

Regarding affordability, a majority of the centres (718) offer subsidized care, making child care more accessible to families in need, while 345 do not provide such subsidies.

And those data will use for later ANOVA analysis.

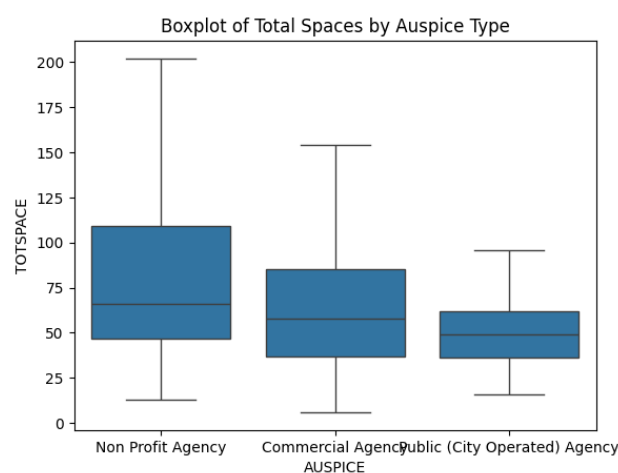
● Total number of child care spaces across various building types



Those two figures show that the total number of child care spaces across various building types in the dataset has been summarized. Public Elementary Schools top the list with the highest number of spaces, accounting for 31,870 spaces. This is followed by buildings categorized as 'Other,' which comprise a total of 14,336 spaces. The variety of buildings used for child care reflects the multifaceted nature of child care provision, with a range of community assets being utilized to meet the needs for child care spaces. Public Elementary Schools, due to their high numbers, seem to be particularly central to child care infrastructure.

4. Data Analysis with ANOVA method:

Question 1: Does the auspice type (Non-Profit vs. Commercial) affect the total spaces (TOTSPACE) available in daycare centers in Toronto?



The boxplot shows the distribution of total spaces (TOTSPACE) for the three categories of auspice types. Non-Profit Agencies tend to have a higher median of total spaces compared to Commercial and Public (City Operated) Agencies. The boxplot for Commercial Agencies shows that they have a narrower IQR than Non-Profit

Agencies, indicating less variability, but also generally have fewer spaces. Public (City Operated) Agencies have the smallest IQR and the lowest median, suggesting they operate with the least number of total spaces on average.

One-Way ANOVA Analysis:

	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

We can use One-Way ANOVA test to determine if there are any statistically significant differences between the means of the three groups.

From your ANOVA results, we can see that the F-statistic is 21.843051 and the p-value is very small (5.057716×10^{-10}), which is far below the common alpha level of 0.05.

This very small p-value indicates that there is a **statistically significant difference** in the mean total spaces provided by the different auspice types.

Tukey's Honest Significant Difference (HSD) test

There's a statistically significant difference between Commercial and Non-Profit Agencies, with Non-Profit Agencies having more spaces on average.

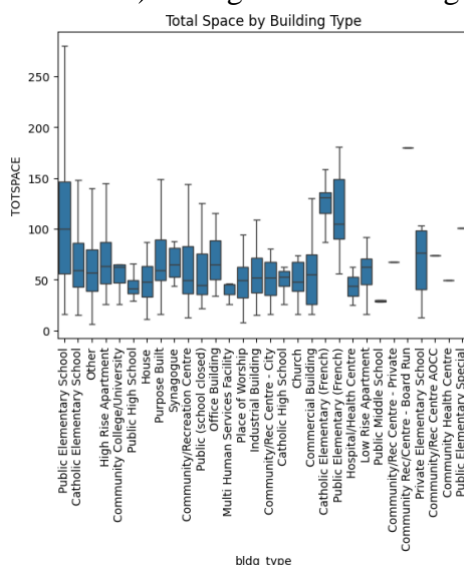
No significant difference in spaces between Commercial Agencies and Public (City Operated) Agencies was observed.

Non-Profit Agencies have significantly more spaces than Public (City Operated) Agencies.

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

Question 2: Is there a significant difference in the total spaces offered (TOTSPACE) among different building types (bldg_type) where daycare centers are located?

We will use the boxplot to explore the differences in total spaces available (TOTSPACE) among various building types where daycare centers are located.



The boxplot shows a comparison of total spaces across different types of buildings. It is evident that some building types, such as Public Elementary Schools, seem to have a wider range of TOTSPACE, which indicates more variability in the number of spaces they offer. Other building types show narrower IQRs, suggesting more consistency in the number of spaces available.

One-Way ANOVA Analysis:

	sum_sq	df	F	PR(>F)
C(bldg_type)	5.205689e+05	29.0	9.72056	1.715361e-37
Residual	1.907608e+06	1033.0	NaN	NaN

A One-Way ANOVA test has been performed to determine if there are statistically significant differences in the mean values of TOTSPACE across the different building types.

The ANOVA results show a significant F-statistic (9.72056) and a very small p-value (1.715361e-37), which is much less than the standard significance level of 0.05.

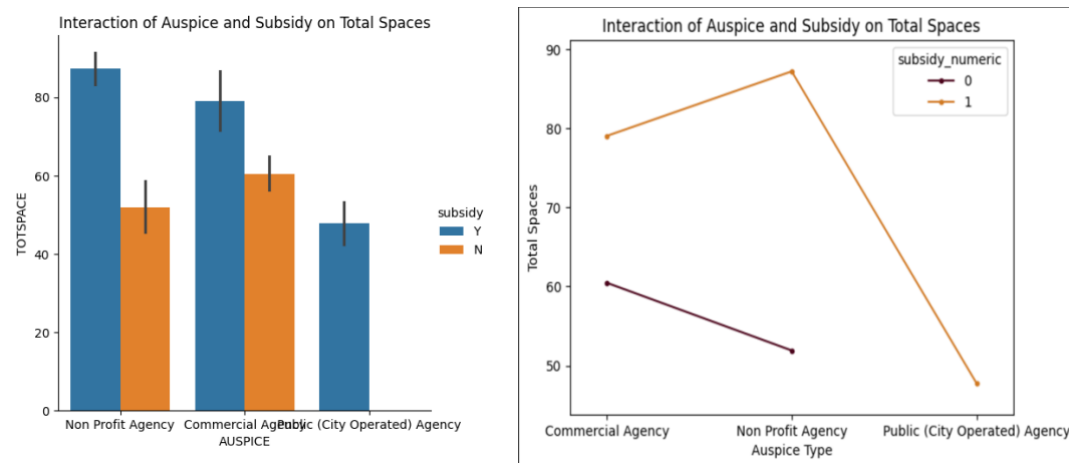
This small p-value indicates that there are **statistically significant differences** in the average number of total spaces provided by different building types.

Tukey's Honest Significant Difference (HSD) test

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
Catholic Elementary (French)	Catholic Elementary School	-56.7497	0.1543	-119.6741	6.1747	False
Catholic Elementary (French)	Catholic High School	-77.2143	0.4913	-178.4734	24.0448	False
Catholic Elementary (French)	Church	-71.3393	0.0685	-144.5495	1.8709	False
Catholic Elementary (French)	Commercial Building	-69.7976	0.0465	-139.195	-0.4003	True
Catholic Elementary (French)	Community College/University	-66.1143	0.6864	-160.7104	28.4818	False
Catholic Elementary (French)	Community Health Centre	-76.7143	0.9986	-249.4224	95.9938	False
Catholic Elementary (French)	Community Rec/Centre - Board Run	54.2857	1.0	-118.4224	226.9938	False
Catholic Elementary (French)	Community/Rec Centre - City	-75.4286	0.2043	-161.7826	10.9255	False
Catholic Elementary (French)	Community/Rec Centre - Private	-58.7143	1.0	-231.4224	113.9938	False
Catholic Elementary (French)	Community/Rec Centre AOC	-51.7143	1.0	-224.4224	120.9938	False
Catholic Elementary (French)	Community/Recreation Centre	-60.4571	0.1512	-127.3467	6.4324	False
Catholic Elementary (French)	High Rise Apartment	-58.0872	0.1582	-122.6695	6.4952	False
Catholic Elementary (French)	Hospital/Health Centre	-82.2143	0.8504	-211.7453	47.3168	False
Catholic Elementary (French)	House	-76.8052	0.0067	-144.0317	-9.5787	True
Catholic Elementary (French)	Industrial Building	-68.7143	0.7441	-169.9734	32.5448	False
Catholic Elementary (French)	Low Rise Apartment	-68.4286	0.4021	-154.7826	17.9255	False

The Tukey HSD test results show that, in comparison to French Catholic Elementary buildings, only Commercial Buildings and Houses have significantly fewer daycare spaces. There are no significant differences in the number of spaces for the other types of buildings when compared to French Catholic Elementary buildings.

Question 3: Is there an interaction effect between the auspice type (AUSPICE) and the presence of a subsidy (subsidy) on the total spaces (TOTSPACE) in daycare centers?



The provided bar graph illustrates the interaction effect between the type of auspice (AUSPICE) and the availability of subsidy (subsidy) on the total spaces (TOTSPACE) in daycare centers.

The bar graph shows that Non-Profit Agencies with subsidies (indicated by 'Y') have the highest mean total spaces, followed closely by Commercial Agencies with subsidies. For both Non-Profit and Commercial Agencies, the availability of a subsidy is associated with a higher mean number of total spaces than those without a subsidy (indicated by 'N').

The interaction plot suggests that subsidies have varying impacts on daycare spaces depending on the agency type. Non-Profit Agencies show a marked increase in total spaces with subsidies, while Commercial Agencies exhibit a minor decrease.

Surprisingly, Public Agencies show a significant decrease in spaces when subsidized. This indicates that subsidies might benefit Non-Profit daycare centers the most, while potentially having an adverse effect on Public ones.

	sum_sq	df	F	PR(>F)
C(AUSPICE)	8.567996e+03	2.0	2.057586	1.282730e-01
C(subsidy)	8.352744e+04	1.0	40.117876	3.529094e-10
C(AUSPICE):C(subsidy)	5.603445e+04	2.0	13.456555	1.694282e-06
Residual	2.202809e+06	1058.0	NaN	NaN

Two-Way ANOVA Analysis:

- The p-value for the main effect of AUSPICE is above the typical alpha level of 0.05 ($p = 0.1282730$), indicating that, on its own, AUSPICE does not have a statistically significant effect on TOTSPACE.
- The main effect of subsidy is highly significant ($p = 3.529094e-10$), suggesting that the presence of a subsidy is associated with a significant difference in TOTSPACE, regardless of the AUSPICE type.
- The interaction term (C(AUSPICE):C(subsidy)) has a very small p-value ($p = 1.694282e-06$), indicating that the effect of subsidy on TOTSPACE significantly differs depending on the AUSPICE type.

Tukey's Honest Significant Difference (HSD) test

The Tukey's HSD test results indicate significant differences in total spaces between Commercial and Non-Profit Agencies, with Non-Profits offering more spaces, and between Non-Profit and Public Agencies, where again Non-Profits offer more. No significant difference was found between Commercial and Public Agencies.

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