Anthony Huang
1005681497
INF2178 Assignment - 2
March 6th, 2024

Exploring Toronto Child Care Capacity

Introduction:

Nowadays, finding affordable child care service has become a challenging issue for many families. With limited resources and space in the child care service system, 75% of the families fail to find available services that can take care of their children. While recently the provincial government pledged a large number of child care spaces, how to better allocate the resources to fulfill the public's needs remains to be solved. In order to assist relevant organizations to make smarter decisions, understanding the existing usage and capacity pattern of child care service is crucial.

This report will utilize the data from licensed Toronto child care centers to unravel the hidden pattern using methods such as data visualization and statistical analysis. The focus of this report will be on analyzing the impact of operating auspice on the total service capacity. This report aims to answer the following research questions:

1. Does child care service with different operating auspice have different total space for children?
2. Does operating auspice and subsidy together affect total space?

These research questions will be examined using one-way and two-way ANOVA, followed by a post-hoc test. The hypotheses are that all of these factors have a significant impact on the total capacity of children care service.

Data Preprocessing:

The raw dataset contains 1063 rows and 17 columns. There are 5 columns containing irrelevant or useless information for answering the above research questions, including building address, postal code, name of the center, name of the building and the program id. These columns are removed in order to improve the clarity and process speed for further analysis. The following are the definition for remaining columns:

LOC_ID: unique identifier for each center
AUSPICE: operating auspice type (Commerical, Non-profit, Public)
ward: city ward number
bldg_type: building type
IGSPACE: spaces for 0-18 months children
TGSPACE: spaces for 18 - 30 months children
PGSPACE: spaces for 30s up to grade one children
KGSPACE: spaces for children in kindergarten
SGSPACE: spaces for children in grade one and up
TOTSPACE: spaces for all aged children
subsidy: whether center has a subsidy contract
cwelcc_flag: whether center participate in CWELCC program

None of these columns contains missing value. Since total space (TOTSPACE) is related to the research questions, a boxplot of it is made in order to see if there is any outlier. We noticed there are several outliers above the 3rd quartile (Figure 1), therefore we decided to remove all the outliers before processing to the next stage.

Figure 1. Boxplot of Total Space (TOTSPACE)
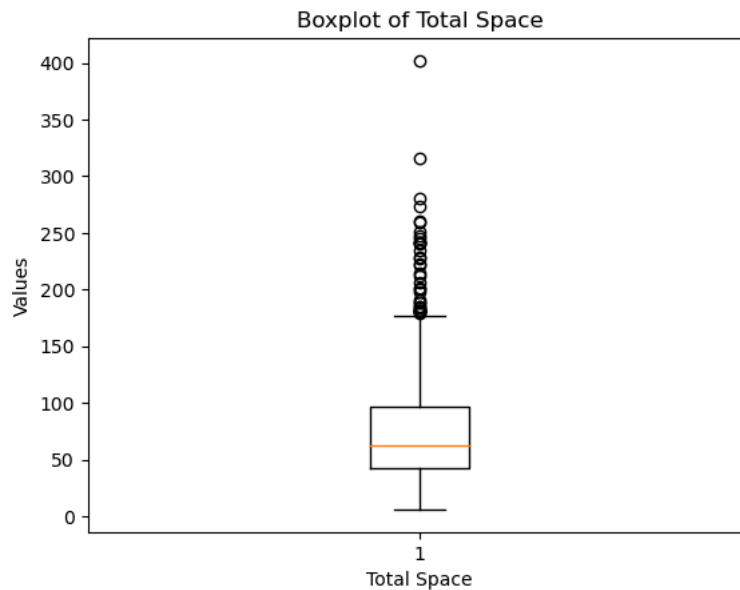
Exploratory Data Analysis:

The statistics overview of relevant quantitative columns is shown in table 1.

Table 1. Descriptive Statistics

	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
mean	8087.888053	12.511759	3.896519	11.600188	24.258702	14.257761	21.661336	75.674506
std	5151.252693	7.032050	6.091782	12.088152	18.577416	20.493290	30.423503	47.816518
min	1013.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	6.000000
25%	1862.000000	6.000000	0.000000	0.000000	16.000000	0.000000	0.000000	43.000000
50%	8826.000000	12.000000	0.000000	10.000000	24.000000	0.000000	0.000000	62.000000
75%	13245.000000	19.000000	10.000000	15.000000	32.000000	26.000000	30.000000	97.000000
max	14504.000000	25.000000	30.000000	90.000000	144.000000	130.000000	285.000000	402.000000

Frequency of the two categorical variables that related to the research questions are shown through the histograms. (Figure 2 and 3)
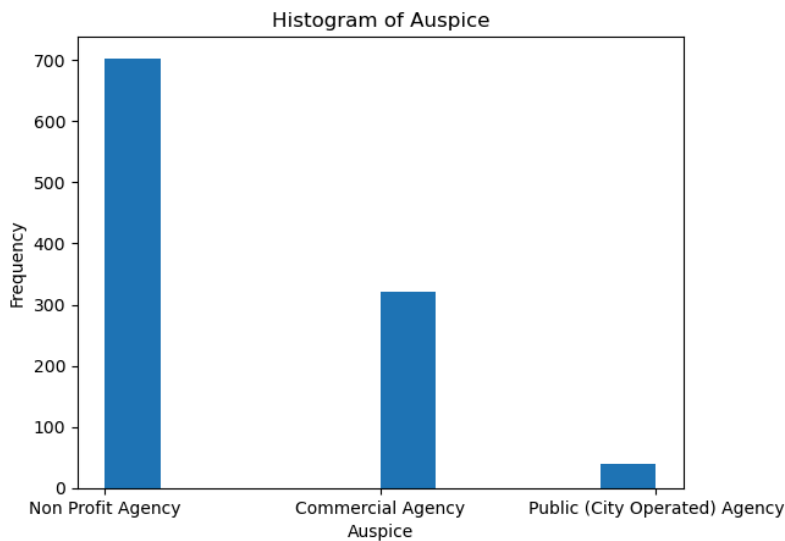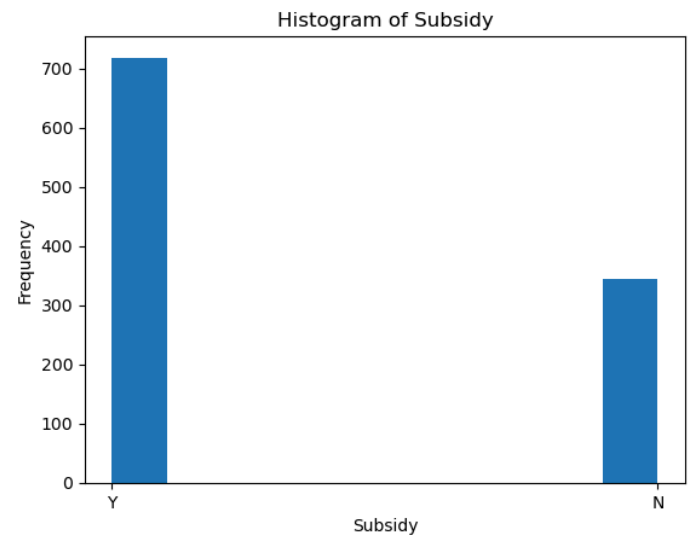
Figure 2. Count of auspice type

Figure 3. Count of Subsidy

For variable auspice, we can see that the total number of each type varies. In order to clearly see whether aupsice has correlation with total space, a histogram comparing the mean total space (Figure 4) and a box plot is made (Figure 5).
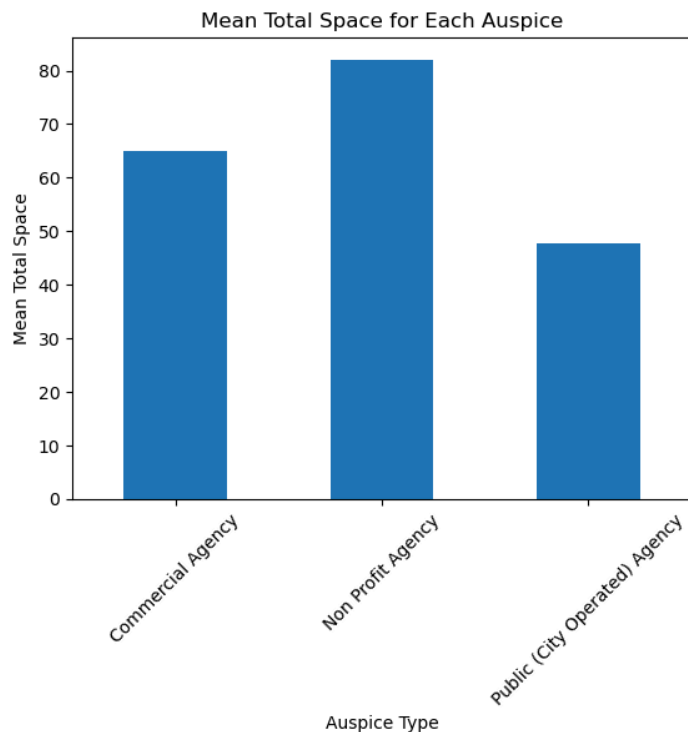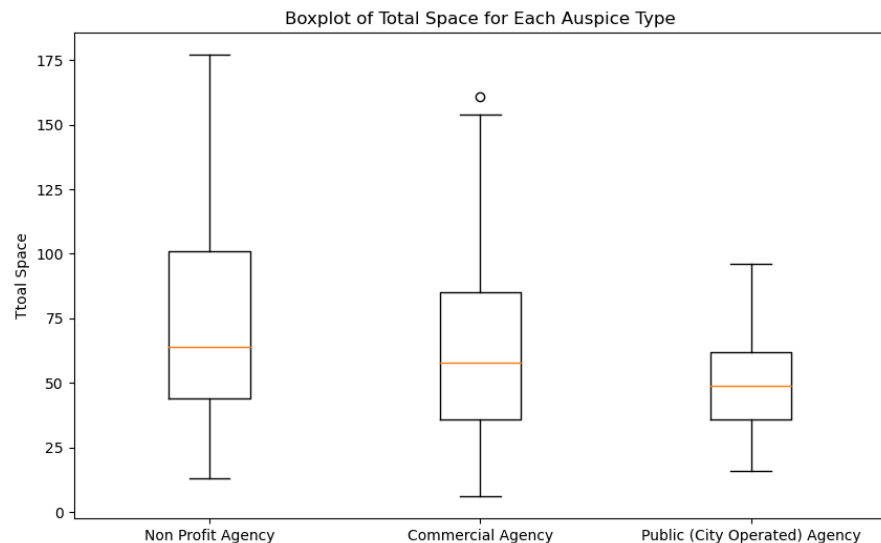
Figure 4. Mean Total Space for Auspice

Figure 5. Distribution of Total Space for Auspice

From the graphs above, we can see that there is a slight difference in the median of total capacity for each auspice type. The variance seems to be inconsistent across auspice types. This will be examined in the following test. Boxplot for subsidy is also made for better answer research question two (Figure 6) We can see the median is about the same and variance also very similar. However, whether there is an interaction between subsidy and auspice is unclear yet. This will also be examined in the test section.

Figure 6. Total Space for Subsidy
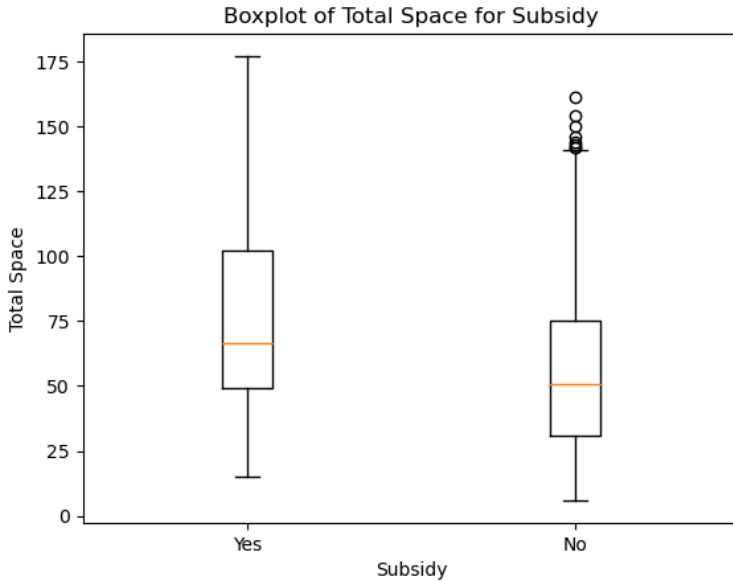
ANOVA TEST:

Research Question One: In order to answer the research questions, ANOVA tests are needed. Before running ANOVA, an assumption test is required. For the one-way ANOVA, shapiro wilk test is performed for checking whether the residual is normally distributed, and levene's test is performed for checking whether variance are homogeneous. Results are shown in the following:

Figure 7. Assumption Test of One-way ANOVA

```
result of shapiro test: (0.9491209983825684, 2.4889163596151464e-18)
reuslt of levene test:
```

	Parameter	Value
0	Test statistics (W)	14.9703
1	Degrees of freedom (Df)	2.0000
2	p value	0.0000

Unlike what we expected, the data failed to pass the assumption tests for performing one-way ANOVA. This is likely due to the fact that real world data sometimes is too complex for using ANOVA. However, ANOVA tests will still be performed due to the limited scope of this report. Due to the fact that assumption tests did not pass, the result of ANOVA may not be very accurate, but it can still give us a comprehensive understanding of the child care capacity pattern. The result of one-way ANOVA is shown below (Figure 8). We can see that the p-value is lower than 0.05, which indicates that not all the group mean are equal.

Table 2. Result of One-way ANOVA

|	sum_sq	df	F	PR(>F)
C(AUSPICE)	5.046271e+04	2.0	17.716959	2.729002e-08
Residual	1.455467e+06	1022.0	NaN	NaN

To further understand which group has different meanings, a post-hoc test is conducted. From table 3 we can see the result of the post-hoc test. The p-values are all smaller than 0.05, suggesting that nonprofit agency's space is significantly higher than the other two types, and commercial agencies have more space than public agencies.

Table 3. Result of Post-hoc test for one-way ANOVA

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	11.806717	5.767004	17.846431	6.488924	0.001000
1	Non Profit Agency	Public (City Operated) Agency	27.160515	12.569289	41.751742	6.178827	0.001000
2	Commercial Agency	Public (City Operated) Agency	15.353798	0.322946	30.384650	3.390721	0.043983

Research Question Two: The boxplot of each interaction group is shown below. We can see there are five groups since all public agencies have subsidies. The medians and variance are similar across different groups.

Figure 8. Boxplot for two-way ANOVA group
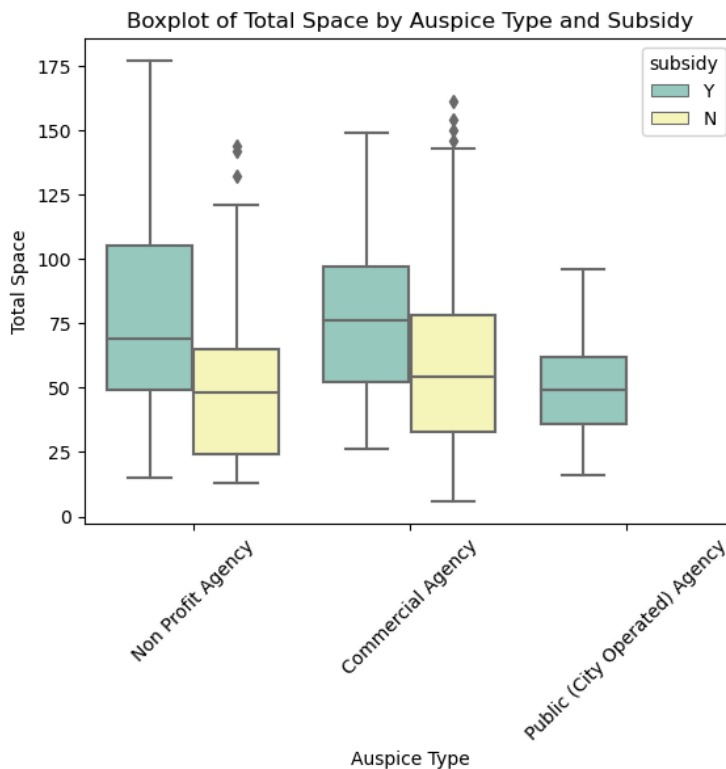
The same assumption tests are performed prior to performing two-way ANOVA. Similar to the previous issue, both tests did not pass, which might decrease the validity of the two-way ANOVA.

Figure 9. Result of Assumption Test for Two-way ANOVA

```
Shapiro-Wilk test statistic: 0.9479749202728271
p-value: 1.5315137372036613e-18
Levene's test statistic: 11.02305686626293
p-value: 7.60311963065639e-09
```

Based on the result from two-way ANOVA and post-hoc test (Table 4 & 5), there are several groups that have significant higher space than other groups:

- For a non-profit agency with subsidy, it has significantly higher total space than a non-profit agency with no subsidy, a commercial agency with no subsidy, or a public agency.
- Non-profit agencies without subsidies have more space than commercial agencies with subsidies.
- Commercial agency with subsidy has more space than public agency and commercial agency without subsidy

In general, subsidy is correlated to the increase in the total number of child care service spaces, and this effect is stronger for non-profit agencies.

Table 4. Result of Two-way ANOVA

	sum_sq	df	F	PR(>F)
C(AUSPICE)	2.026500e+03	2.0	0.757258	4.692140e-01
C(subsidy)	6.001180e+04	1.0	44.850144	3.508106e-11
C(AUSPICE):C(subsidy)	3.882915e+04	2.0	14.509586	6.117030e-07
Residual	1.364813e+06	1020.0	NaN	NaN

Table 5. Result of Two-Way Post-hoc Test

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(Non Profit Agency, Y)	(Non Profit Agency, N)	28.721291	17.443140	39.999441	10.282633	0.001000
1	(Non Profit Agency, Y)	(Commercial Agency, Y)	2.148962	-10.533980	14.831904	0.684143	0.900000
2	(Non Profit Agency, Y)	(Commercial Agency, N)	20.632512	12.591965	28.673058	10.361074	0.001000
3	(Non Profit Agency, Y)	(Public (City Operated) Agency, Y)	31.496614	14.208503	48.784725	7.356218	0.001000
4	(Non Profit Agency, Y)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900000
5	(Non Profit Agency, N)	(Commercial Agency, Y)	26.572329	10.772091	42.372566	6.790539	0.001000
6	(Non Profit Agency, N)	(Commercial Agency, N)	8.088779	-4.298347	20.475905	2.636639	0.426903
7	(Non Profit Agency, N)	(Public (City Operated) Agency, Y)	2.775324	-16.913995	22.464643	0.569143	0.900000
8	(Non Profit Agency, N)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900000
9	(Commercial Agency, Y)	(Commercial Agency, N)	18.483550	4.805055	32.162045	5.456138	0.001691
10	(Commercial Agency, Y)	(Public (City Operated) Agency, Y)	29.347652	8.821337	49.873968	5.772988	0.001000
11	(Commercial Agency, Y)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900000
12	(Commercial Agency, N)	(Public (City Operated) Agency, Y)	10.864103	-7.167065	28.895270	2.432810	0.515720
13	(Commercial Agency, N)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900000
14	(Public (City Operated) Agency, Y)	(Public (City Operated) Agency, N)	0.000000	-inf	inf	0.000000	0.900000

The following is the interaction plot that demonstrate that subsidy effect is stronger for non-profit agency.

Figure 10. Interaction Plot for Two-way ANOVA
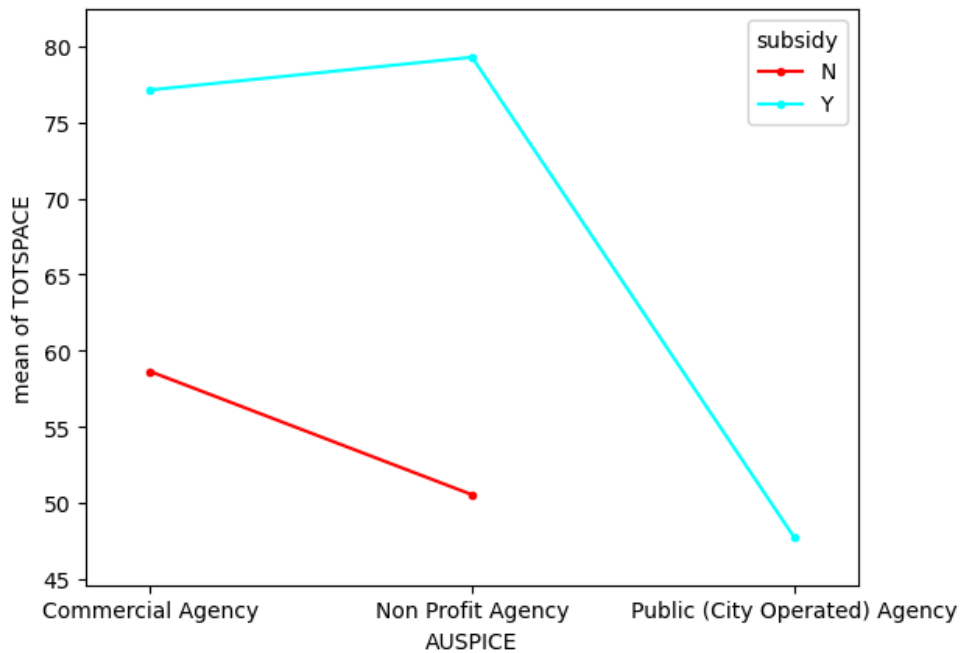
Conclusion:

To summarize, through exploratory data analysis and ANOVA testing, we found that child care services that are non-profit agencies have significantly more total space than other types of agencies. Moreover, having subsidies may lead to increases in total space, and this phenomenon is stronger for non-profit agencies as well. In terms of how to solve the current challenge that people cannot find available child care service, increasing the number of non-profit centers that have a subsidy contract may be the most effective way.