

Student ID: 1010737874
Student name: Han Zheng
Instructor: Dr. Shion Guha

Exploring Child Care Center Capacities in Toronto

1. Introduction

In the bustling urban landscape of Toronto, the demand for quality childcare services continues to grow as families seek supportive environments for their children's early development. Childcare centers play a pivotal role in meeting this need, offering spaces that nurture young minds during their most formative years. However, the capacity of these centers to accommodate children varies significantly across different age groups and is influenced by various factors, including the type of building in which they are housed and the auspice under which they operate.

This report delves into the complexities of licensed childcare centers in Toronto, focusing on their operational and capacity nuances for multiple age groups. Through a comprehensive analysis of data collected from these centers, we aim to uncover patterns and disparities in capacity, offering insights into the current landscape of childcare services in the city. By exploring how building types and auspice categories impact the availability of care for infants, toddlers, preschoolers, kindergartners, and school-age children, this narrative seeks to inform policymakers, stakeholders, and families about the state of childcare in Toronto, fostering a deeper understanding of its challenges and opportunities.

Our exploration will address three fundamental research questions, serving as guiding principles unraveling the troubling nature Toronto's childcare patterns:

1. **Research Question 1:** How does the capacity of childcare vary across different age groups in licensed childcare centers?
2. **Research Question 2:** Does the type of auspice have an effect on the total capacity of childcare centers?
3. **Research Question 3:** Is there an interaction effect between building type and age group on the capacity of childcare spaces available?

By addressing these questions, we aim to contribute insights into the dynamics of childcare in Toronto and provide a deeper understanding that can inform more effective interventions.

2. Data Cleaning and Data Wrangling

The raw dataset has a total of **17 columns** with **1063** entities (**rows**). After initial review of the dataset, we were confident that not much data cleaning was deemed

necessary for the analysis since only BLDGNAME has some missing values. However, we noted any observed discrepancies and defined new features necessary for future analysis. Below we outlined our observation and all new features added to our dataset:

Observations and Considerations:

1. Since our analysis only focused on a few aspects, we've reduced our working dataset to the following columns source from the raw dataset. Below we provided a short description of each column:
 - AUSPICE: The management type of the childcare center, such as Non-Profit Agency, Commercial Agency, and Public Agency.
 - bldg_type: The type of building in which the childcare center is located, such as Public Elementary School and Commercial Building.
 - IGSPACE: The number of infant (0-18 months) spaces available.
 - TGSPACE: The number of toddler (18-30 months) spaces available.
 - PGSPACE: The number of preschool (30 months – grade one) spaces available.
 - KGSPACE: The number of kindergarten child's spaces available.
 - SGSPACE: The number of school-age child's spaces available.
 - TOTSPACE: The total number of spaces available across all age groups.
2. The following column seemed to have many missing values (NaN):
 - BLDGNAME: Name of the building the childcare centre is located inHowever, we are not interested in this qualitative variable, so it has no influence on our analysis.

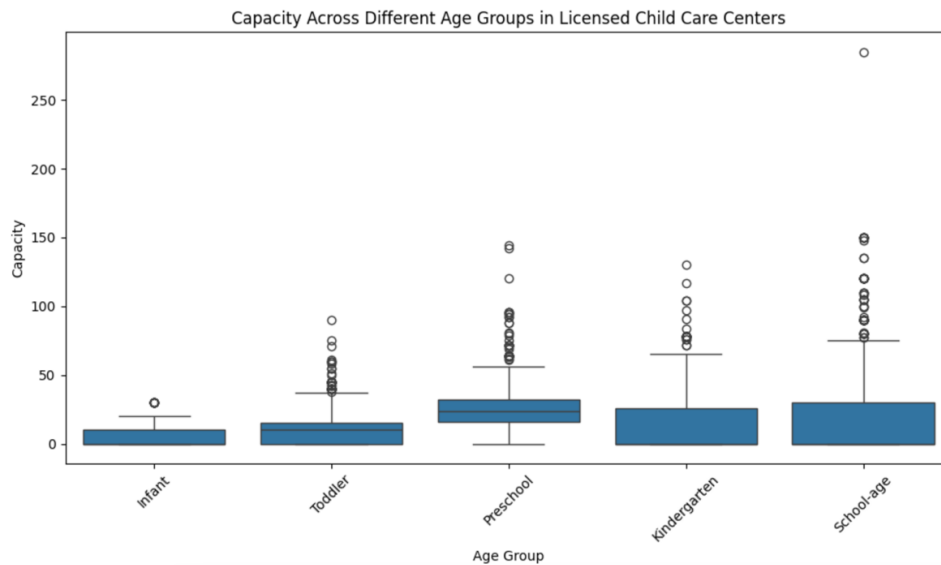
3. Exploratory Data Analysis (EDA)

Our analysis utilized data on licensed child care centers in Toronto, examining variables such as capacity by age group (infant, toddler, preschool, kindergarten, school-age), building type, and auspice (non-profit, commercial, public). We conducted one-way and two-way ANOVAs to explore the effects of these variables on child care capacity, complemented by post-hoc tests for detailed comparisons.

4. Capacity Analysis Across Age Groups

Research Question #1: How does the capacity of childcare vary across different age groups in licensed childcare centers?

For the first research question, we delved into how childcare capacity varies across different age groups, from infants to school-age children. This involved extracting data on the available spaces for each age group and visually representing these differences through interaction plots.



Given the data's non-adherence to normal distribution and variance homogeneity, we opted for a Kruskal-Wallis H test instead of a traditional one-way ANOVA. This non-parametric approach allowed us to identify significant variations in capacity across age groups, which were further explored through post-hoc pairwise comparisons to pinpoint specific differences.

Test result:

Kruskal-Wallis H test result: `KruskalResult(statistic=818.6634260274693, pvalue=6.960346391052689e-176)`

Adjusted Post-Hoc Results:

```
{'Infant vs. Toddler': 1.4545849363185713e-68,
'Infant vs. Preschool': 4.969471920658144e-189,
'Infant vs. Kindergarten': 4.71654843432915e-24,
'Infant vs. School-age': 1.7781549643739318e-37,
'Toddler vs. Preschool': 1.0912886903938264e-83,
'Toddler vs. Kindergarten': 0.37535251261714037,
'Toddler vs. School-age': 0.031704306454949954,
'Preschool vs. Kindergarten': 8.082803911800659e-42,
'Preschool vs. School-age': 3.553441869019119e-15,
'Kindergarten vs. School-age': 1.3597767479119088e-06}
```

To understand which specific age groups differ from each other, we conducted post-hoc analyses using pairwise comparisons adjusted for multiple comparisons. These adjusted post-hoc results highlight significant differences between nearly all pairs of age groups. Here are some example explanations for some of them:

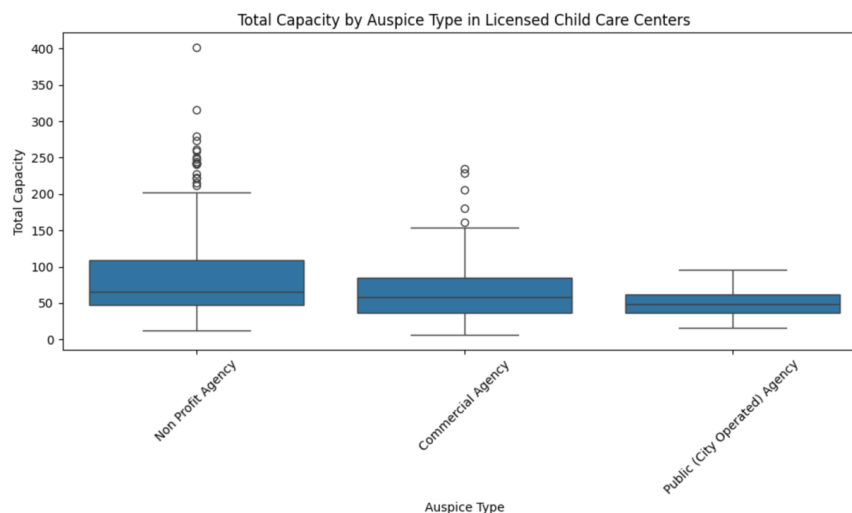
Infant vs. Toddler: The capacity for infants is significantly different from toddlers, with a p-value of $1.45e-68$, indicating lower or higher median capacity for one of these groups compared to the other.

Infant vs. Preschool: A p-value of $4.97e-189$ suggests a substantial difference in capacity between infant and preschool age groups.

5. Impact of Auspice Type on Capacity

Research Question #2: Does the type of auspice have an effect on the total capacity of childcare centers?

For the second research question, we focused on the impact of auspice type on the total capacity of childcare centers, categorizing centers as non-profit, commercial, or public. Through box plots, we visualized the capacity variations by auspice type.



As we can see from the boxplot above, there are a few points we want to focus on. Let us begin with the median line. It revealed that non-profit agencies tend to have a higher median total capacity compared to commercial and public agencies, indicating a potential trend where non-profit childcare centers offer more spaces on average. Then for the spread of the data, as represented by the IQR, illustrated the variability in total capacity within each auspice type, with non-profit agencies showing a wide range, suggesting diversity in the sizes of these centers. Additionally, the presence of outliers in the plots underscored the existence of some childcare centers with capacities significantly different from the norm within their respective auspice category.

Then, we employed the Kruskal-Wallis H test to analyze the data, circumventing the limitations posed by the initial assumption violations for ANOVA.

```
{('Non Profit Agency': ShapiroResult(statistic=0.8917579650878906, pvalue=8.339870811481173e-22),
'Commercial Agency': ShapiroResult(statistic=0.9238301515579224, pvalue=1.013540900512977e-11),
'Public (City Operated) Agency': ShapiroResult(statistic=0.9514912962913513, pvalue=0.09225869178771973)},
LeveneResult(statistic=17.927068524236084, pvalue=2.2036889276494363e-08))
```

To explain the test results above, let us go with the p-value. It indicates that the total capacities for both non-profit and commercial agencies significantly deviate from a normal distribution, as the p-values are both less than 0.05. This suggests that the data for these groups are not normally distributed. However, for public agencies, the p-value is greater than 0.05, which means we cannot conclude that their capacity distribution significantly deviates from normality.

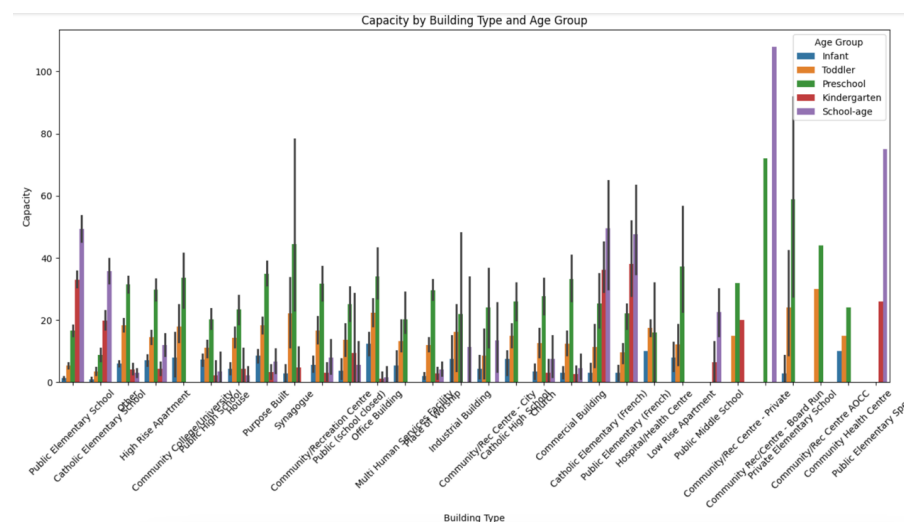
The Levene's test for homogeneity of variances gives us a p-value much lower than 0.05, showing that the variances of total capacity are significantly different across the three types of agencies. This implies that the variability in center size is not consistent among non-profit, commercial, and public childcare centers.

Combining the boxplot and test results above, we can say that non-profit agencies may operate with a different scale and resource allocation compared to their commercial or public counterparts, affecting the number of children they can accommodate.

6. Interaction Between Building Type and Age Group on Capacity

Research Question #3: Is there an interaction effect between building type and age group on the capacity of childcare spaces available?

For the last research question, we investigated the potential interaction effect between building type and age group on childcare capacity. We firstly made an interaction plot to aggregate data across these dimensions to achieve visually assess capacity variations.



The graph above broken down by building type and age group, and each building type has a set of bars representing the number of spaces available for infants, toddlers, preschoolers, kindergarteners, and school-age children. The length of each bar corresponds to the median capacity for that age group within the building type, and the error bars suggest variability in the data, indicating the range of capacities across different centers. To explain the graph in common words, we can say that high rise apartments show a notable number of toddler and preschool spaces, while public elementary schools appear to provide substantial capacity across all age groups, particularly for school-age children.

We also applied two-way ANOVA in this research question as the test results shown below:

Kruskal-Wallis Test Result for Preschool Age Group: KruskalResult(statistic=249.32680577357215, pvalue=1.9633096847924947e-40)				
	sum_sq	df	F	PR(>F)
C(bldg_type)	1.041138e+05	29.0	15.527176	1.579914e-73
C(AgeGroup)	2.821233e+05	4.0	305.043080	6.918114e-236
C(bldg_type):C(AgeGroup)	6.917570e+05	116.0	25.791573	0.000000e+00
Residual	1.194230e+06	5165.0	NaN	NaN

With a test statistic of approximately 249.33 and a small p-value of 1.96e-40, the test emphatically suggests variation in capacity among different building types. Additionally, the test points to significant differences across age groups, indicating that preschool age group capacities vary notably.

7. Conclusion

Overall, this thorough examination sheds light on the pressing need for targeted policy interventions and resource allocation to balance the availability of childcare across Toronto. By understanding the specific capacity constraints and advantages presented by different age groups, building types, and auspice categories, stakeholders can make informed decisions to enhance the quality and accessibility of childcare services for all families in the city.