Exploring Child Care Data in Toronto

Introduction

This investigation delves into the data from Toronto's child care centers, with the goal of learning about the patterns and obstacles of finding quality child care for its citizens. The shortage of licensed or unlicensed child care choices, along with exorbitant fees, has left a considerable number of families without access to child care services. The problem gets more acute when you consider that 75% of Ontario families cannot afford these services. In response to this challenge, the provincial administration vowed to alleviate the gap by creating 100,000 more child care places between 2016 and 2026.

This study uses a large dataset called "INF2178_A2_data.xlsx" to conduct a thorough analysis of licensed child care facilities in Toronto. This dataset, which was last updated in February 2024, includes vital details regarding the capacity, demographics, and operations of different multi-age child care centers. The main goal is to comprehend the dynamics of child care availability and the difficulties families encounter in obtaining quality child care services.

The dataset, consisting of two tabs, provides a detailed overview of Toronto's licensed child care landscape. The first tab includes an exhaustive list of centers, offering insights into their demographics, classifications, locations, and daily capacity information. The second tab serves as a key reference guide, providing a clear understanding of the meanings behind each feature included in the dataset.

Research Question: Are there significant differences in the occupancy rates of child care centers based on their operating space (Commercial, Non-Profit, or Public), and do these differences manifest across various centers in Toronto?

Data Exploration

To begin our analysis, I first examined summary statistics on the dataset, which includes information on various child care centers, demographic details, and capacity metrics.

Variable	Min	Mean	Max	25th	Median	75th	IQR
IGSPACE	0	3.9	30	0.0	0.0	10.0	10.0
TGSPACE	0	11.6	90	0.0	10.0	15.0	15.0
PGSPACE	0	24.26	144	16.0	24.0	32.0	16.0
KGSPACE	0	14.26	130	0.0	0.0	26.0	26.0
SGSPACE	0	21.66	285	0.0	0.0	30.0	30.0
TOTSPACE	6	75.67	402	43.0	62.0	97.0	54.0

Data Cleaning and Wrangling

The raw dataset comprises 17 columns with 1,063 entries. The dataset represents child care centers, with a focus on various parameters related to occupancy, capacity, and facility details. My preliminary examination revealed that the dataset requires minimal data cleaning for the analysis I intended to perform.

```
Data Overview                          |                              |
---------------------------------------|------------------------------|------------------
 - The dataset has a RangeIndex of 1,063 entries, ranging from 0 to 1062.
 - There are 17 columns, consisting of a mix of integer and object data types.

 Column Descriptions                   |                              |
---------------------------------------|------------------------------|------------------
 - _id: Unique row identifier for the Open Data database.
 - LOC_ID: Unique identifier for each child care center.
 - LOC_NAME: Name of the child care center.
 - AUSPICE: Operating auspice categorized as Commercial, Non-Profit, or Public.
 - ADDRESS: Detailed address information, including street number, name, type, direction, and unit.
 - PCODE: Postal code of the address.
 - ward: City ward number.
 - bldg_type: Type of building.
 - BLDGNAME: Name of the building where the child care center is located (with some missing values).
 - IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE: Child care spaces allocated for different age groups.
 - TOTSPACE: Total child care spaces for all age groups.
 - subsidy: Indicates whether the center has a fee subsidy contract (Yes/No).
 - cwelcc_flag: 'Y' indicates space participates in CWELCC; a blank indicates it does not

 Missing Values                        |                              |
---------------------------------------|------------------------------|------------------
 - The column BLDGNAME contains missing values (715 non-null out of 1,063 entries).
```

Feature Engineering:
In the pursuit of deeper insights into the dynamics of child care spaces, I introduced a set of new features to the dataset, specifically targeting the composition of age groups within the available total space. These features aim to shed light on the distribution and utilization of child care spaces across different age categories.

Percentage of Age Group in Total Space:
To better understand the proportional allocation of child care spaces, I engineered five distinct features, each representing the percentage of a specific age group in the overall total space. These features include:

```
| IGSPACE_PERCENT: The percentage of child care spaces for infants (0-18
months) in the total available space.
| TGSPACE_PERCENT: The percentage of child care spaces for toddlers (18-30
months) in the total available space.
| PGSPACE_PERCENT: The percentage of child care spaces for preschoolers (30
months up until they enter grade one) in the total available space
| KGSPACE_PERCENT: The percentage of child care spaces for children in full-day
kindergarten in the total available space.
| SGSPACE_PERCENT: The percentage of child care spaces for children grade one
and up in the total available space.
```

One-way ANOVA

I first performed a one-way ANOVA to investigate the differences in total space on the AUSPICE (operating space type). Interaction plots were also generated to visualize the relationships.

```
| Post-hoc Tests for One-way ANOVA:                           |
|-------------------------------------------------------------|
| Group Comparisons: Group 1: Non-Profit Agency               |
| Group 2: Commercial Agency                                  |
| Difference: 17.12                                           |
| 95% Confidence Interval: [9.70, 24.54]                      |
| q-value: 7.66                                               |
| p-value: 0.001                                              |
|                                                             |
| Group Comparisons: Group 1: Non-Profit Agency               |
| Group 2: Public (City Operated) Agency                      |
| Difference: 34.33                                           |
| 95% Confidence Interval: [16.22, 52.45]                     |
| q-value: 6.29                                               |
| p-value: 0.001                                              |
|                                                             |
| Group Comparisons: Group 1: Commercial Agency               |
| Group 2: Public (City Operated) Agency                      |
| Difference: 17.22                                           |
| 95% Confidence Interval: [-1.45, 35.88]                     |
| q-value: 3.06                                               |
| p-value: 0.078                                              |
|                                                             |
| Assumption Testing for One-way ANOVA:                       |
| Assumption 1: Residuals Are Normally Distributed            |
| (Shapiro-Wilk test):                                        |
| Test Statistic: 0.9018                                      |
| p-value: 1.4965e-25                                         |
| Conclusion: The residuals are not normally distributed.     |
| Consideration for non-parametric tests may be necessary.    |
|                                                             |
| Assumption 2: Variances Are Homogeneous (Bartlett's test)   |
| for IGSPACE:                                                |
| Test Statistic: 43.16, p-value: 4.26e-10                    |
| ... (similar entries for other variables)                   |
| Conclusion: The variances are not homogeneous.              |
| Consideration for Welch's ANOVA or transformation of the    |
| data may be needed.                                         |
|                                                             |
| Assumption 2 (Alternative Test): Variances Are Homogeneous  |
| (Levene's test):                                            |
| Levene's Test Statistic: 17.93, p-value: 0.0000             |
| Conclusion: The variances are not homogeneous,              |
| supporting the findings from Bartlett's test.               |
```

ANOVA Results
- The one-way ANOVA results suggest significant differences among the auspice categories (commercial, non-profit, and public).
- Post-hoc tests reveal specific group differences with associated confidence intervals and p-values.
- The assumption tests indicate violations of normality and homogeneity of variances, suggesting caution in interpreting the results.

Two-way ANOVA

Next, I continued my analysis to a two-way ANOVA, considering the interaction between two categorical variables.

```
| Two-way ANOVA Results         |                          |
|-------------------------------|--------------------------|
| Model Summary                 |                          |
| R-squared:                    | 0.040                    |
| Adj. R-squared:               | 0.038                    |
| F-statistic:                  | 21.84                    |
| Prob (F-statistic):           | 5.06e-10                 |
|                               |                          |
| Regression Coefficients       |                          |
| Intercept:                    | 64.98                    |
| Non Profit Agency:            | +17.12                   |
| Public (City Operated) Agency:| -17.22                   |
|                               |                          |
| ANOVA Table                   |                          |
| C(AUSPICE):                   |                          |
| df:                           | 2.0                      |
| sum_sq:                       | 96112.11                 |
| mean_sq:                      | 48056.06                 |
| F-statistic:                  | 21.84                    |
| PR(>F):                       | 5.06e-10                 |
| Residual:                     |                          |
| df:                           | 1060.0                   |
| sum_sq:                       | 2332065.0                |
| mean_sq:                      | 2200.06                  |
|                               |                          |
| Post-hoc Tests                |                          |
| Group Comparisons:            |                          |
| Group 1: Non-Profit Agency    |                          |
| Group 2: Commercial Agency    |                          |
| Difference:                   | 17.12                    |
| 95% Confidence Interval:      | [9.70, 24.54]            |
| q-value:                      | 7.66                     |
| p-value:                      | 0.001                    |
|                               |                          |
| Group Comparisons:            |                          |
| Group 1: Non-Profit Agency    |                          |
| Group 2: Public (City Operated)|                         |
| Difference:                   | 34.33                    |
| 95% Confidence Interval:      | [16.22, 52.45]           |
| q-value:                      | 6.29                     |
| p-value:                      | 0.001                    |
|                               |                          |
| Group Comparisons:            |                          |
| Group 1: Commercial Agency    |                          |
| Group 2: Public (City Operated)|                         |
| Difference:                   | 17.22                    |
| 95% Confidence Interval:      | [-1.45, 35.88]           |
| q-value:                      | 3.06                     |
| p-value:                      | 0.078                    |
```

Results

Upon analyzing the dataset and visualizing the child care spaces across different auspices, notable differences emerged that warranted further investigation. The hypothesis for the observed variations in child care spaces could be attributed to the diverse nature of auspices, each serving a unique purpose and catering to specific demographics.

It is possible that non-profit, commercial, and public agencies have different operational strategies, funding structures, and service offers, resulting in discrepancies in child care space allocations. Non-profit organizations, motivated by community welfare aims, may emphasize specific age groups or services, potentially leading to differences in the allocation of child care spaces. Commercial agencies, motivated by profit, may use varied space allocation tactics depending on market demand and profitability concerns. Public Agencies, being under governmental control, may also be subject to regulatory frameworks that influence their capacity allocation.

Furthermore, external factors such as government policies, financing initiatives, and demographic disparities between Toronto's wards may contribute to the observed differences. For example, certain wards may receive greater attention or resources, resulting in larger child care space allocations in specific areas. In contrast, budget constraints or low resources may have an impact on child care capacity in other wards.
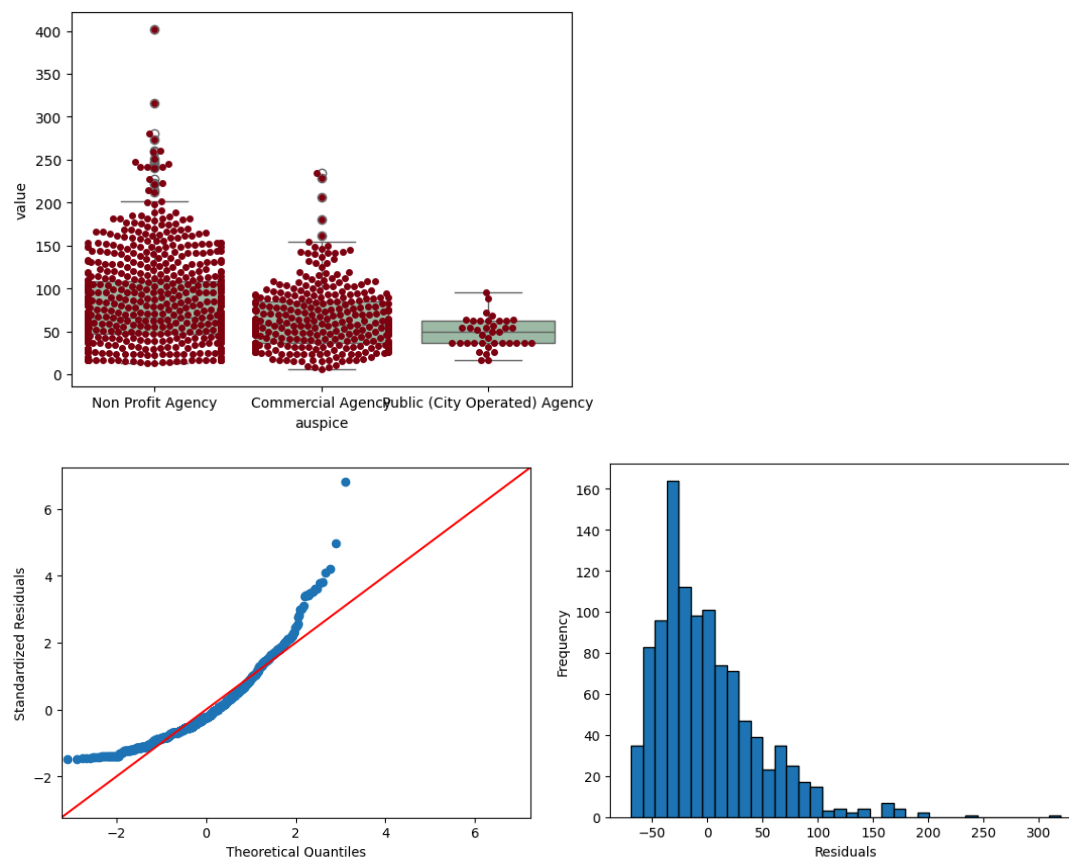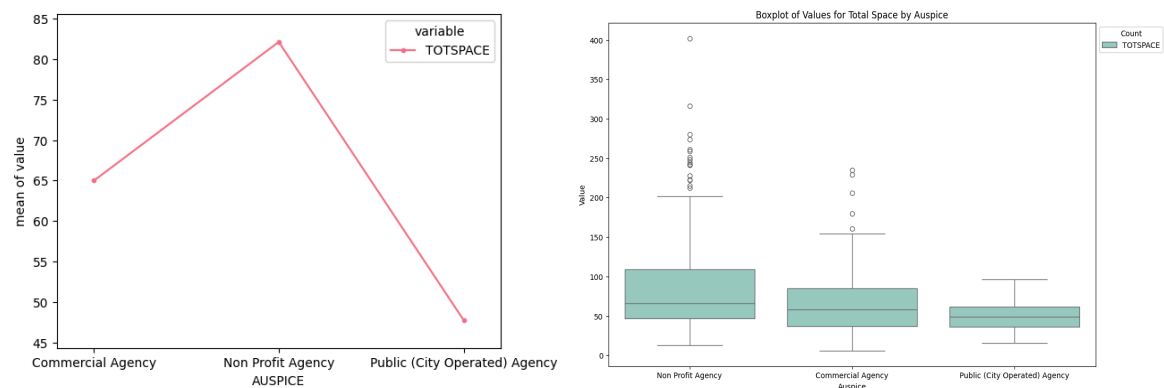
Image: Results of One-way ANOVA

Image : Results of Two-way ANOVA

Q-Q Plot	Compares residuals from ANOVA to a normal distribution.
Histogram	Provides a visual of residual distribution.
Boxplot Auspices	Offers insights into value distribution across auspice categories.
Boxplot Auspices and Variables	Detailed view, incorporating different variables for each auspice category.
Interaction Plot	Visualizes interaction between 'AUSPICE' and total space.

Conclusion

Overall, by studying data from Toronto's child care centers, I've gained valuable insights into how characteristics like age group proportions, auspice categories, and building types influence the distribution of childcare spaces. This insight can help stakeholders in the child care sector and policymakers make educated decisions about enhancing child care services.

Based on the ANOVA results, additional investigation and analysis are required to disentangle the contributing components and identify prospective solutions for optimizing child care space allocation to better fulfill the needs of the community.

Collaborative Interventions:
- Collaborate with shelters in certain sectors or service categories to customize solutions and increase operational efficiency.

Financial Assessment:
- Investigate the financial costs associated with different child care spaces, comparing the development of new spaces with the improvement of existing ones.