# Optimizing Strategy for Child Care Facility Expansion Using Mixed-Integer Linear Programming: A Comprehensive Solution to Child Care Deserts

Group 49: Minghao Li, Som Naik, Yemin Tang, Zhexin Wang
IEOR 4004 - Optimization model and methods

November 2, 2024

**Abstract**

The need for child care is growing, especially among employed parents. Recently, more and more research focused on this topic and identified the complexity of child care arrangements. This study focuses on the issue of child care deserts in New York State. The goal is to eliminate the child care desert by considering budget, distance, and fairness respectively. The main optimization method used in this paper is mixed-integer linear programming. Utilizing this powerful technique and data from related information, 3 optimization models have been built to find the best arrangement for child care and the optimal expected cost and social coverage index which can be used to measure fairness. The main finding is the cost will increase when the distance is taking into account and the cost is changing into piecewise. Besides, some specific locations can lead to unfair and cannot find a feasible solution for the fairness case. The recommendation for upper management would be to take these areas more seriously and isolate them for treatment and analysis.
**Keywords**: Optimization, Mixed-Integer Linear Programming, Child Care Deserts

## 1 Introduction

Nowadays, more than half of American children still live in areas where there is a much greater need for licensed child care than there are available child care slots (CAP, nd). Such areas are named child care deserts. The precise definition of Child Care Deserts varies in areas of different needs. In high-demand areas where at least 60% of parents are working or the average income is at most $60000 each year, child care deserts can be defined in such areas based on the number of available slots at most the half population of children aged 2 weeks to 12 years. Conversely, if the region do not satisfy the high-demand standard, it can be defined as a normal-demand area. In such areas, a child care desert can be identified if the available slots are at most one-third of the children's population. Thus, this motivates this study to concentrate on the problem of child care deserts in New York State.

This study will mainly focus on answering the following research questions: How to eliminate child care deserts while minimizing costs? How can the distance factor be taken into account in all of the above questions at the same time while changing to scale-based cost? and how to maximize a social coverage index while trying to solve the child care desert issue? Thus, this study's objective is to use Mix-integer Programming as an optimization technique and find the best solution. In response to answer these research questions and to achieve the goal, this study establishes 3 mixed-integer linear programming models to design the best arrangement of child care.

Therefore, this study aims to go further to address the gaps in this research area and use an operations research technique to deal with this problem, by taking New York State as an example. In addition, this study examined the feasibility of using mixed-integer linear programming modeling to solve child care deserts problems. This study also provides the basic optimization models for solving this problem, it can be applied to other areas. In practice, this research can help upper managers to insight into how to arrange child care facilities.

The rest of this report is organized as follows: Section 2 is the background, which includes the relevant background information and studies. After that, Section 3 will illustrate the mathematical modeling using the mixed-integer linear programming algorithm, and also provide information about the assumptions and data sources used. Section 4 follows, by illustrating the data cleaning process and visualizations to point out the key information of our data and model. The results and findings will be implemented and discussed in Section 5, which also presents some practice meanings and limitations of this study. Finally, Section 6 will reiterate the conclusion briefly and then provide some further studies.

# 2 Background

Child care, also called day care, is the monitoring and care of children, usually between the ages of 2 weeks and 12(Bradley and Vandell, 2007). The need for child care is significant and cannot be ignored. Of infants, 44% are enrolled in a regular nursery program; by the time they are 3 years old, this number rises to 61%(NSECE, 2016). According to a 2002 US Bureau of the Census's estimation, 11596000 children under the age of five regularly spend a considerable amount of time in nonparental care (Johnson, 2005). This service is particularly important for working parents as it provides a professional environment to care for their children when they are working (Adams and Henly, 2020). The subject of child care arrangements has attracted the attention of researchers, they have found that child care arrangements vary widely in terms of quality, quantity, and type (NICHD, 2000). Child care availability in the United States can vary significantly by zip code. As the workforce grows, it is important that families have access to child care facilities. Zip codes which do not have sufficient child care facilities for their population and demographic can be classified as "child care deserts". The government has a longstanding mission of eliminating child care deserts across all zip codes. However, this paper attempts to eliminate child care deserts in the state of New York.

# 3 Methods

## 3.1 Model Assumptions

To make the real-world complex problem manageable and help to construct the model, this study introduces several assumptions:

- Assumption 1: The model makes the assumption that building and expansion expenses are fixed and unaffected by shifts in the market, inflation, or labor costs. This assumption makes sure the model can use the fixed cost values from the data source.
- Assumption 2: Throughout the modeling period, it is assumed that demographic statistics (such as employment rates and age distribution) and child care demand indicators stay unchanged.
- Assumption 3: It is assumed that there is a linear relationship between the cost increase and the newly added slots. Thus, we can employ mixed-integer linear programming to construct the models.

## 3.2 Data Source

This paper uses 5 datasets to analyze the child care deserts issue and build optimization models based on the existing data. The first dataset of this study used contains information of existing child care facilities in New York State for each zip code. The total number of zip codes is 1023. This dataset totally contains 14756 child care facilities with their ID, program type, name, city, distinct, capacity, and location. The second dataset involves the population figures among different age groups. The third one consists of the average income statistics among each zip code area in New York State. The fourth one has information on the employment rate for every zip code in New York State. The third and fourth datasets can help to identify the area as highly-demand or normal-demand. The last one has the choices of potential locations available for newly built facilities. The location information is provided as latitude and longitude.

## 3.3 Problem 1

The New York State (NYS) government aims to eliminate childcare deserts by increasing childcare slots nationwide, ensuring all regions have adequate access. To achieve this, NYS plans to allocate funds for either building new facilities or expanding existing ones, with construction limited to three facility sizes, each with specific capacities and costs. Expansions are capped at 1.2 times the current capacity, up to a maximum of 500 slots, with costs scaling based on facility size. Additionally, slots for children under age five require extra funding for specialized equipment. As consultants, the task is to find the minimum funding needed per area to meet these targets. The full optimization model can be found in Appenix.

### 3.3.1 Decision Variables

- $x_{ji}$: Number of new facilities of size $j$ built at zip code $i$, where $j = 1, 2, 3$ (representing small, medium, and large facilities, respectively).
- $x_{1pi}$: Number of expanded slots for children aged 0-5 at facility $p$ in zip code $i$.
- $x_{2pi}$: Number of expanded slots for children aged 5-12 at facility $p$ in zip code $i$.

### 3.3.2 Objective Function

$$\text{Minimize } C = \sum_{j=1}^{3} c_j x_{ji} + \sum_{p} c_p^e \left( \frac{x_{1pi} + x_{2pi}}{q_{pi}^{(12)}} \right) + 100 \sum_{p} x_{1pi} \tag{1}$$

where:

- $c_j$: Cost of building a new facility of size $j$.
- $c_p^e$: Expansion cost of facility $p$.
- $q_{pi}^{(12)}$: Capacity of facility $p$ at zip code $i$.
- The term $100 \sum_p x_{1pi}$ represents the additional cost for expanding slots for children under 5.

### 3.3.3 Model Constraints

**Demand Constraint**

$$\sum_{p} q_{pi}^{(12)} + \sum_{j=1}^{3} s_{j1} x_{ji} + \sum_{p} (x_{1pi} + x_{2pi}) \geq k_i a_i, \quad \forall i \tag{2}$$

where:

- $s_{j1}$: Number of slots for children aged 0-12 provided by facility size $j$.
- $k_i$: Child care desert threshold at zip code $i$.
- $a_i$: Number of children aged 0-12 at zip code $i$.
- $q_{pi}^{(12)}$: It is the existing facilities for ages 0-12

**Under 5 Coverage Constraint**

$$\sum_p q_{pi}^5 + \sum_{j=1}^{3} s_{j2} x_{ji} + \sum_p x_{1pi} \geq \frac{2}{3} b_i, \quad \forall i \tag{3}$$

where:

- $s_{j2}$: Number of slots for children aged 0-5 provided by facility size $j$.
- $b_i$: Number of children aged 0 to 5 years at zip code $i$.
- $q_{pi}^{(5)}$: It is the existing facilities for ages 0-5

**Expansion Limits**

$$x_{1pi} + x_{2pi} \leq 0.2 q_{pi}, \quad \forall p, i \tag{4}$$

**Maximum Capacity Constraint**

$$x_{1pi} + x_{2pi} \leq (500 - q_{pi}) o_{pi}, \quad \forall p, I \tag{5}$$

where:

- $o_{pi}$: Binary variable indicating whether facility $p$ in zip code $i$ is expanded (1 if yes, 0 otherwise).

**Non-negativity and Integrality**

$$x_{ji}, x_{1pi}, x_{2pi} \geq 0, \quad \forall j, p, i \tag{6}$$

$$x_{ji} \in \mathbb{Z}, \quad \forall j, i \tag{7}$$

$$o_{pi} \in \{0, 1\}, \quad \forall p, i \tag{8}$$

## 3.4 Problem 2

To better reflect the complexities of child care facility expansion, New York State (NYS) officials recommend a variable cost model for expanding existing facilities and a distance limitation between facilities to avoid over-concentration. For expansions, marginal costs increase with the scale: small expansions (up to 10%) incur a baseline cost ($20,000) plus $200 per existing slot, expansions between 10-15% cost $400 per slot, and expansions between 15-20% cost $1000 per slot due to logistical challenges. Expansion beyond 20% is deemed prohibitively costly. Additionally, to prevent facility clustering, a minimum distance of 0.06 miles between any two facilities is advised. The task is to calculate the minimum funding needed to meet these goals, considering both the staged cost model and the location restrictions. The full optimization model can be found in Appenix.

### 3.4.1 Decision Variables

- $f_{jl}$: Binary variable indicating whether a facility of size $j$ is built at potential location $l$ (1 if yes, 0 otherwise).
- $x_{ji}$: Number of new facilities of size $j$ at zip code $i$, where $j = 1, 2, 3$ (representing small, medium, and large facilities).
- $x_{1pi}$: Number of expanded slots for children aged 0-5 at facility $p$ in zip code $i$.
- $x_{2pi}$: Number of expanded slots for children aged 5-12 at facility $p$ in zip code $i$.
- $\delta_{jpi}$: Number of additional slots created in the $j$th piecewise expansion of facility $p$ in zip code $i$.

#### 3.4.2 Objective Function

$$\text{Minimize } C = \sum_{j=1}^{3}\sum_{l} c_j f_{jl} + \sum_{p}\sum_{j=1}^{3} c^{\delta}_{jp}\delta_{jpi} + 100\sum_{p} x_{1pi} \tag{9}$$

where:

- $c_j$: Cost of building a new facility of size $j$.
- $c^{\delta}_{jp}$: Piecewise expansion cost for the $j$th segment of facility $p$.

#### 3.4.3 Model Constraints

**New Facility Location Constraint**

$$\sum_{l} m_{li} f_{jl} = x_{ji}, \quad \forall j, i \quad (1) \tag{10}$$

**One Facility per Location**

$$\sum_{j=1}^{3} f_{jl} \le 1, \quad \forall l \quad (2) \tag{11}$$

**Distance Constraint**

$$\sum_{L\neq l} n_{lL}\left(\sum_{j=1}^{3} f_{jL}\right) \le 1, \quad \forall l \quad (5) \tag{12}$$

**Piecewise Expansion Definition**

$$x_{1pi} + x_{2pi} = \delta_{1pi} + \delta_{2pi} + \delta_{3pi}, \quad \forall p, i \quad (6) \tag{13}$$

$$X = \begin{cases} 2000 + 200 \cdot X \\ 2000 + 400 \cdot X \\ 2000 + 1000 \cdot X \end{cases}$$

where:

$$x_{1p_i}, x_{2p_i} \in X$$

**Piecewise Expansion Limits**

$$0 \le \delta_{1pi} \le 0.1 q^{(12)}_{pi}, \quad \forall p, i \quad (7) \tag{14}$$

$$0 \le \delta_{2pi} \le 0.05 q^{(12)}_{pi}, \quad \forall p, i \quad (8) \tag{15}$$

$$0 \le \delta_{3pi} \le 0.05 q^{(12)}_{pi}, \quad \forall p, i \quad (9) \tag{16}$$

**Non-negativity and Integrality**

$$x_{1pi}, x_{2pi} \ge 0, \quad \forall j, p, i \quad (11) \tag{17}$$

$$r_{pi}, \delta_{1pi} \ge 0, \quad \forall j, p, i \quad (12) \tag{18}$$

$$o_{pi}, n_{lL}, f_{jl} \in \{0, 1\}, \quad \forall j, l, L \neq l \quad (13) \tag{19}$$

$$\delta_{1pi} \in \mathbb{Z}, \quad \forall p, i \quad (14) \tag{20}$$

## 3.5 Problem 3

In problem 3, the target is to maximize a social coverage index under a fairness constraint that the difference in the ratio of available child care slots to the total population of children between any two areas does not exceed 0.1. The full optimization model can be found in Appenix.

### 3.5.1 Decision Variables & Objective Function

In this problem, we continue to use the same decision variables and notations as those in former problems. The change lies in the objective function since we are dedicated to maximize the weighted social coverage index. Therefore, we define

$$h_i = \frac{\sum_{j=1}^{3} s_{j1}x_{ji} + \sum_p (x_{1pi} + x_{2pi}) + \sum_p q_{pi}^{(12)}}{a_i} \tag{21}$$

as the child care coverage for all children at a given zip code $i$. Similarly, we can also define

$$g_i = \frac{\sum_{j=1}^{3} s_{j2}x_{ji} + \sum_p x_{1pi} + \sum_p q_{pi}^{(5)}}{b_i} \tag{22}$$

as the child care coverage for children under 5 at a given zip code $i$. Then the corresponding objective function is set to be a weighted sum of the corresponding coverage index at each zip code, i.e.

$$C.I. = \frac{1}{3}\sum_i h_i + \frac{2}{3}\sum_i g_i \tag{23}$$

And the goal is to maximize the overall coverage index.

### 3.5.2 Model Constraints

Since problem 3 stays in the same framework of problem 2, a few modifications in constraints are adopted to fulfill the new requirements. In this section, we only focus on new constraints beyond those in problem 2. Firstly, we add on the constraint that the total cost of expansion as well as newly bulit facilities should not exceed the given budget 1 billion, i.e.

$$\sum_{j=1}^{3}\sum_l c_j f_{jl} + \sum_p \sum_{j=1}^{3} c_{jp}^{\delta}\delta_{jpi} \leq 10^9 \tag{24}$$

Another important constraint is that the difference in the ratio of available child care slots to the total population of children between any two areas does not exceed 0.1. Intuitively, this can be interpreted in the model as

$$h_m - h_n \leq 0.1 \tag{25}$$

for any different zipcodes $n$ and $m$. However, this results in a massive amount of constraints added in the model. Let $N$ denote the number of zip codes in data set. Then in this way, the total number of constraints included in the model is $2\binom{N}{2} = N^2 - N = O(N^2)$, which takes a long time to run the model. Thus, we propose a alternative method to model this constraint. Let $I$ denote the set of all zip codes and define

$$h_{max} := \max_{i\in I} h_i \quad h_{min} := \min_{i\in I} h_i \tag{26}$$

respectively as the maximum and minimum ratio of available child care slots to the total population of children. An equivalent interpretation of the constraint can be written as

$$h_{max} - h_{min} \leq 0.1 \tag{27}$$

To formulate $h_{max}, h_{min}$ in the optimization problem, we shall introduce two sets of binary variables $y_i, z_i$ and let

$$\begin{aligned} h_{max} &\geq h_i \\ h_{max} &\leq h_i + My_i \\ h_{min} &\leq h_i \\ h_{max} &\geq h_i - Mz_i \end{aligned} \tag{28}$$

for any $i \in I$. Here $M$ is large enough positive number. Meanwhile, let

$$\begin{aligned} \sum_{i \in I} y_i &= |I| - 1 \\ \sum_{i \in I} z_i &= |I| - 1 \end{aligned} \tag{29}$$

where $|I|$ denotes the number of elements in set $I$. Hence, if $h_i$ is not the maximum among all zip codes, then $y_i$ must equal to 1. If $h_i$ is indeed the maximum among all zip codes, because of constraint (29), $y_i = 0$ and $h_{max}$ will eventually become $h_i$. Under this circumstance, the number of constraints will dramatically reduce to $4N + 2 = O(N)$. This in turn shortens the time to run the model.

Nonetheless, we will make some adjustments on the optimization model in numerical experiments due to the property of the data. A detailed discussion will be provided in (5.1.3).

# 4 Data Cleaning and Visualization

The raw available datasets needed some cleaning. In order to eliminate a child-care desert in a certain zip code, we must know (1) the population of that zip code, (2) the income of that zip code, (3) the employment rate of that zip code, (4) and all existing child care facilities in that zip code. Hence, we eliminate any zip codes from our datasets which do not have all mentioned information. This leaves us with a final set of 1023 zip codes.

## 4.1 Data Visualization

Figure 1 shows the distribution of zip codes across income brackets. Most zip codes have an average income of \$60,000 - \$90,000. Outside of this bracket, other zip codes tend to be lower income.

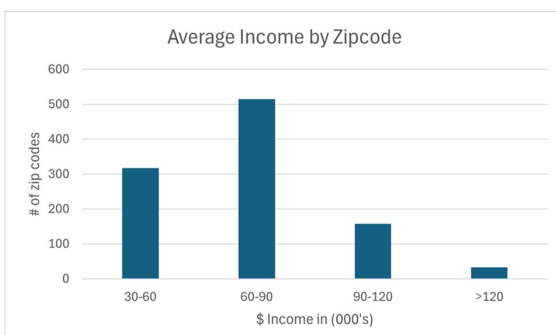


Figure 1: Number of zip codes in income brackets

Figure 2 shows the distribution of zip codes across different employment rates. Most zip codes have an employment rate of 40 to 60 percent. In order to be qualified as a "high demand zip code" the zip code must have an employment rate of greater than or equal to 60 percent, or have an average income of less than 60,000 dollars per year. Using this formula, 35 percent of zip codes are classified as "high demand". We can also roughly observe this from Figures 1 and 2.

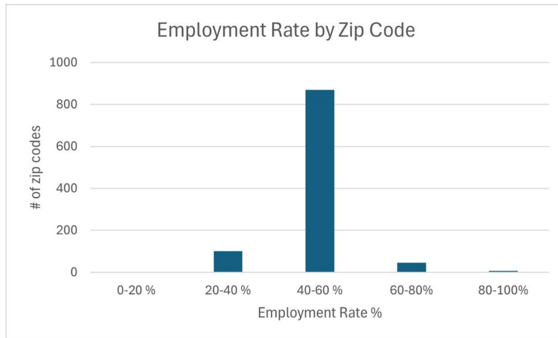


Figure 2: Number of zip codes in income brackets

# 5 Empirical Results and Analysis

## 5.1 Optimization Results

### 5.1.1 Problem 1

Based on the model, we identified that the minimum funding required to reach the target across all areas is approximately \$315932045.9474323 . Additionally, we generated a figure detailing each zip code's minimum cost along with the number of facilities required. The cost distribution among all zip code areas have been shown in Figure 3. The accompanying graph illustrates that, in most areas, the minimum cost needed is under \$874,550. To present the child care arrangement, Figure 4 has been generated. We equally divide all zip codes into 52 groups and show the sum of newly-built facilities in each group for all three facility types. From 4, we can see type 3 is the most numerous in all regions. For Figure 5, focusing on two scenarios: "expand under 5" (representing the expanded facility for children under age 5 ) and "expand all" (representing the expanded facility for all age children). The y-axis shows the total expansion capacity, while the x-axis lists the ZIP Code groups.

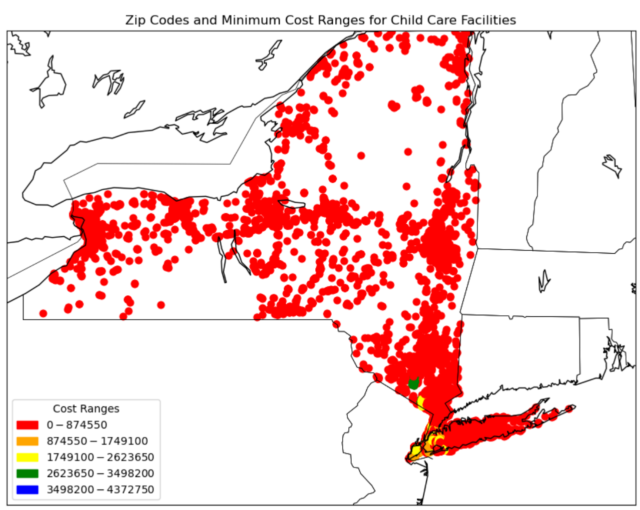


Figure 3: Zip codes and minimum cost range for child care facilities for Problem 1

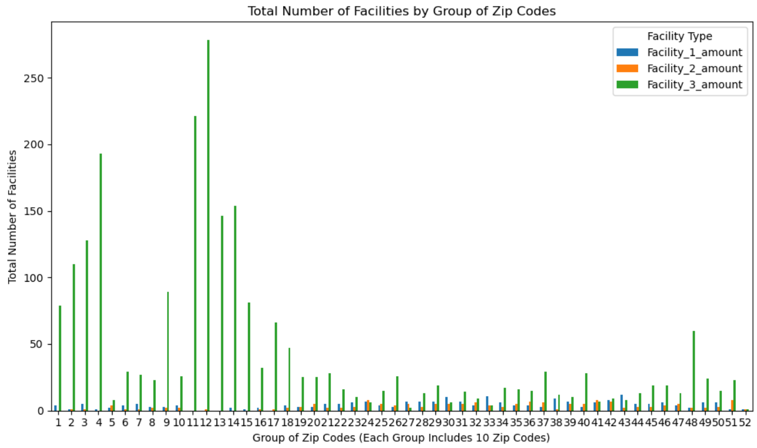


Figure 4: Total number of newly-built facilities by group of zip codes for Problem 1

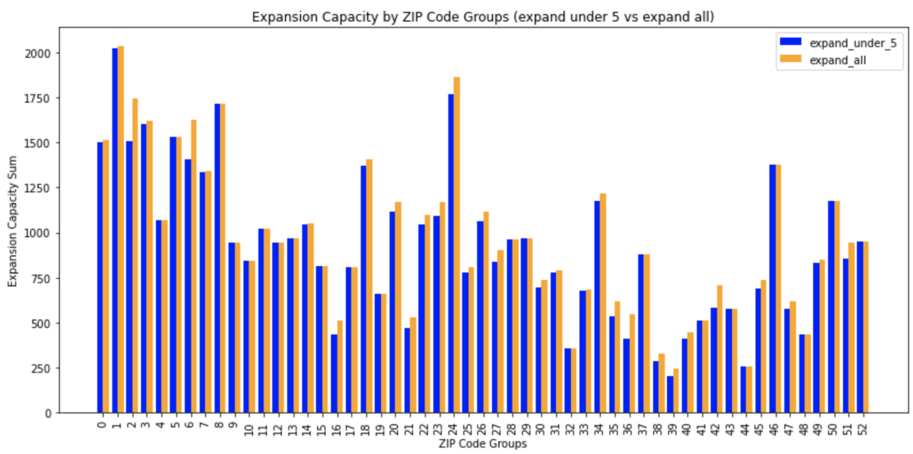


Figure 5: Expansion Capacity by Zipcode for Problem 1

#### 5.1.2 Problem 2

With space limitations in mind, officials noted that the cost of adding slots to existing facilities should adjust based on expansion size—larger expansions result in a higher marginal cost per added slot. To reflect this, we moved from a fixed-cost approach to a more realistic model where costs rise with the scale of expansion. This change, along with a new requirement of at least 0.06 miles between facilities, has increased the total minimum cost of building the necessary facilities to approximately \$321104263.9670968 . Our model focuses on the cost of adding slots, where the cost increases progressively as more slots are added. Figure 6 and Figure 7 have shown the cost allocation and newly-built facilities for problem 2. The observations are similar. The cost remains concentrated on 0-874550 scale, while the type 3 facility has the highest number. Figure 8 provides a comparison of the sum of expanding slots for under 5 and all children. We can see that the most of expanding slots are from the under 5 group. For Figure 9, focusing on two scenarios: "expand under 5" (representing the expanded facility for children under age 5 ) and "expand all" (representing the expanded facility for all age children). The y-axis shows the total expansion capacity, while the x-axis lists the ZIP Code groups.

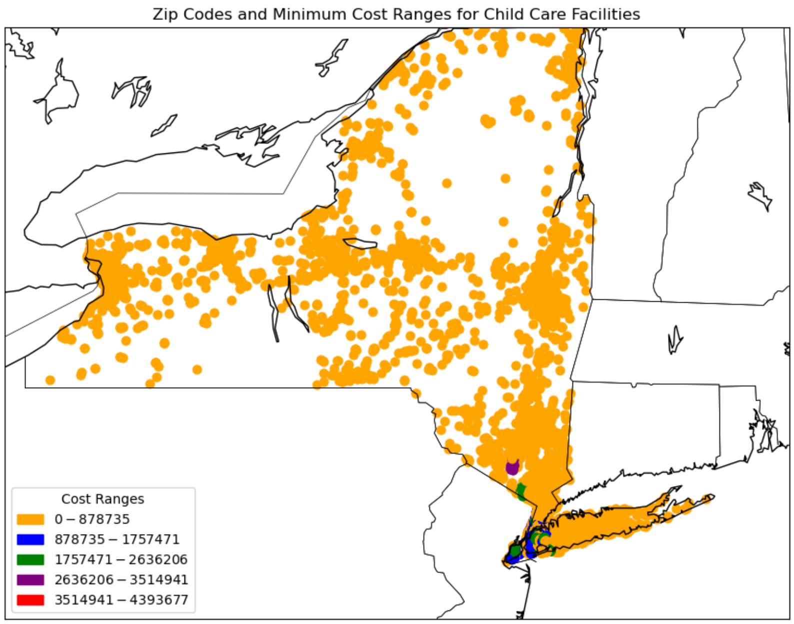


Figure 6: Zip codes and minimum cost range for child care facilities for problem 2

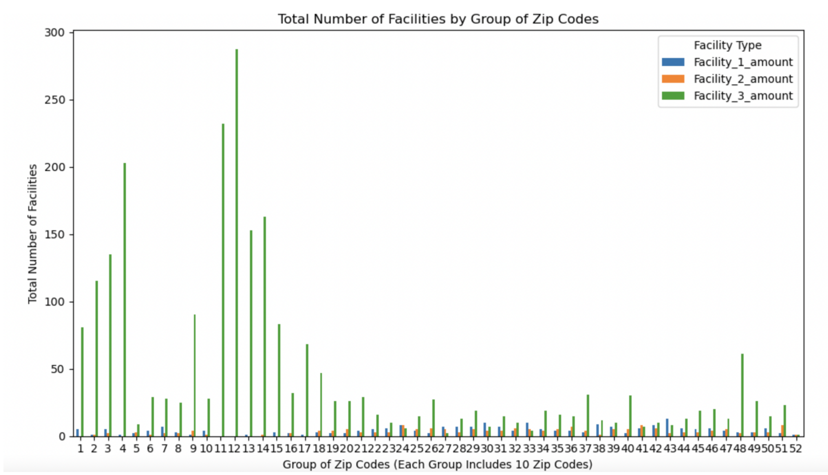


Figure 7: Total number of newly-built facilities by group of zip codes for problem 2

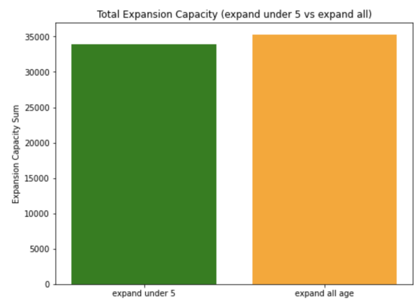


Figure 8: Total Expansion Capacity (expand under 5 vs expand all) for Problem 2

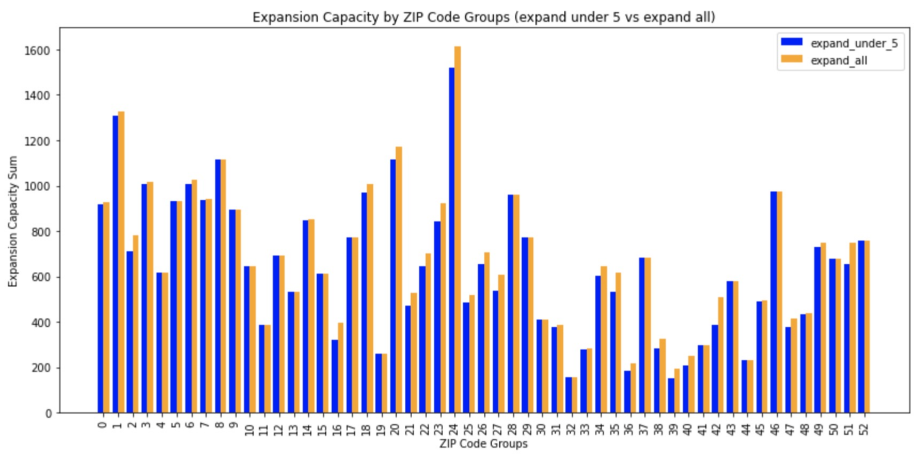


Figure 9: Expansion Capacity by Zipcode for Problem 2

#### 5.1.3 Problem 3

While implementing the model for problem 3, an infeasible solution is reported by Gurobi. A possible explanation is that when existing capacity in a given area is far more than the number of children, the corresponding ratio of available child care slots to the total population of children, denoted as $h$, will attain a rather high value. For example, at zip code 10535, there are existing slots of 100 while the number of children is 10, resulting in $h_{10535} \geq 10$. However, it is impossible for some area to reach such a ratio. Thus, we first try to modify the model, imposing an addition constraint that the maximum ratio at zip code must not exceed 1. It is worth noting that we can not simply add on a constraint that $h_i \leq 1$ for any $i \in I$. A counterexample is as follows

| **Children $< 5$** | **Children $> 5$** | **Slots for $< 5$** | **Slots for $> 5$** |
|---|---|---|---|
| 15 | 1 | 0 | 7 |

Under this circumstance, to satisfy the requirement proposed by NYS, a total number of 10 slots for children under 5 must be built and the ratio $h_i$ will become

$$h_i = \frac{2/3 \times 15 + 7}{15 + 1} = \frac{17}{16} > 1$$

Thus, an alternative method is adopted in the model to avoid this. Let

$$h_i = \min(1, \frac{\sum_{j=1}^{3} s_{j1}x_{ji} + \sum_p (x_{1pi} + x_{2pi}) + \sum_p q_{pi}^{(12)}}{a_i}) \tag{30}$$

and

$$g_i = \min(1, \frac{\sum_{j=1}^{3} s_{j2}x_{ji} + \sum_p x_{1pi} + \sum_p q_{pi}^{(5)}}{b_i}) \tag{31}$$

A similar LP formulation of min function is utilized by introducing two sets of binary variables $y_i^{'}, z_i^{'}$, i.e.

$$\begin{aligned}
&h_i \leq 1 \\
&h_i \leq \frac{\sum_{j=1}^{3} s_{j1}x_{ji} + \sum_p (x_{1pi} + x_{2pi}) + \sum_p q_{pi}^{(12)}}{a_i} \\
&h_i \geq 1 - My_i^{'} \\
&h_i \geq \frac{\sum_{j=1}^{3} s_{j1}x_{ji} + \sum_p (x_{1pi} + x_{2pi}) + \sum_p q_{pi}^{(12)}}{a_i} - M(1 - y_i^{'})
\end{aligned} \tag{32}$$

Meanwhile,

$$\begin{aligned}
&g_i \leq 1 \\
&g_i \leq \frac{\sum_{j=1}^{3} s_{j2}x_{ji} + \sum_p x_{1pi} + \sum_p q_{ip}^{(5)}}{b_i} \\
&g_i \geq 1 - Mz_i^{'} \\
&g_i \geq \frac{\sum_{j=1}^{3} s_{j2}x_{ji} + \sum_p x_{1pi} + \sum_p q_{ip}^{(5)}}{b_i} - M(1 - z_i^{'})
\end{aligned} \tag{33}$$

Moreover, since there does exist areas whose ratio of available child care slots to the total population of children exceeds 1, the minimum ratio across other areas should at least be 0.9 given the fairness constraint. The constraint can be then interpreted as

$$h_i \geq 0.9 \tag{34}$$

for any $i \in I$. Nevertheless, it is still impossible to satisfy the fairness requirement. The key reason is that the ratio of available child care slots to the total population of children in some area is less than 0.9 no matter how much the budget is due to the distance and expansion restriction in previous problems. For instance, at zip code 11219, the maximum ratio $h_{11219} = 0.735 < 0.9$. Therefore, the fairness constraint can never been satisfied.

## 5.2 Discussion

In comparison to previous studies, this research focuses on using the mixed-integer linear programming method to find the optimal solution for child care arrangements. The model construction examines the feasibility of employing mixed-integer programming to solve this problem. The optimal solutions for problems 1 and 2 verify this opinion. Although we found the solution for problem 3 is infeasible, this is not due to a wrong algorithm choice. The study has found some specific areas in which has extremely low social coverage ratios. It is impossible to meet the requirement no matter the budget is. Therefore, senior management can take inspiration from it. For some areas where childcare is sorely lacking, they need focused attention. They can be dealt with and analyzed separately when planning childcare in all regions. This allows for effective budget planning as well as scheduling of expansion.

This study also provides a framework for further child care facility expansion. For example, a similar method can be used in other states and areas. Through the establishment of methods for evaluating demand and optimizing facility placement, the paper provides a practice strategy that may help with larger initiatives to alleviate child care shortages in other locations. Through the process of constructing the optimization model for Problem 1 and 2, this study also examine changing the expansion cost and adding new factor will change the arrangement of child care. The optimal solution of problem 1 and 2 can explain the cost will increase when considering more factors and changing the cost criterion into a piecewise one. The minimum cost difference of Top 20 zip code areas are shown below. Hence, upper management needs to identify the factors to be considered and their ranking of importance before planning begins. Different priorities will lead to different models and different results. This is particularly important when considering budgets.

Top 20 Zip Codes with the Highest Difference in Minimum Cost Between Q1 and Q2:

| | zip_code | Minimum_cost_Q1 | Minimum_cost_Q2 | cost_difference |
|---|---|---|---|---|
| 60 | 10456 | 1.886571e+06 | 1.950436e+06 | 63864.775497 |
| 1 | 10002 | 6.412335e+05 | 7.031657e+05 | 61932.158438 |
| 10 | 10012 | 2.331000e+05 | 2.950000e+05 | 61900.000000 |
| 1013 | 14717 | 7.450000e+03 | 6.500000e+04 | 57550.000000 |
| 56 | 10452 | 1.298733e+06 | 1.354935e+06 | 56201.859577 |
| 218 | 11212 | 1.956393e+06 | 2.010912e+06 | 54519.578320 |
| 213 | 11207 | 2.005502e+06 | 2.058097e+06 | 52595.236331 |
| 849 | 13778 | 1.290277e+05 | 1.800000e+05 | 50972.312373 |
| 61 | 10457 | 1.900291e+06 | 1.947581e+06 | 47290.596605 |
| 57 | 10453 | 1.330304e+06 | 1.376345e+06 | 46041.679102 |
| 2 | 10003 | 4.725523e+05 | 5.157627e+05 | 43210.380758 |
| 241 | 11236 | 1.835365e+06 | 1.878541e+06 | 43176.071833 |
| 128 | 10601 | 2.192183e+04 | 6.500000e+04 | 43078.174894 |
| 71 | 10467 | 1.798758e+06 | 1.841707e+06 | 42948.652396 |
| 67 | 10463 | 1.335095e+06 | 1.377669e+06 | 42573.669739 |
| 267 | 11379 | 5.607385e+05 | 6.031570e+05 | 42418.440849 |
| 262 | 11373 | 1.969290e+06 | 2.011175e+06 | 41884.932045 |
| 338 | 11713 | 1.384286e+05 | 1.800000e+05 | 41571.428571 |
| 205 | 11105 | 7.638894e+05 | 8.050000e+05 | 41110.615446 |
| 868 | 13838 | 2.445766e+04 | 6.500000e+04 | 40542.344498 |

Figure 10: The minimum cost difference of Top 20 zip code areas

The limitations of this study are mainly from the assumptions made before. Assumptions 1 and 2 fixed the data among the whole modeling time period. These assumptions can help construct the model more conveniently, but ignoring some real-world complexity leads to less accurate models. Assumption 3 assumes the linear relationship between the cost and newly added slots but ignores the non-linear factors that may potentially affect the cost. It might underestimate budget requirements by oversimplifying real-world cost escalations, particularly for larger facilities or significant expansions.

# 6 Conclusion

In summary, this study is dedicated to solving child care desert problem. Mixed-integer linear programming has been used as the main technique to optimize the child care expansion while considering the budget, distance, and fairness. While other limitations are considered. For example, the expansion limitation, and demand priority. After data cleaning and analyzing, three optimization models for each case are successfully constructed. The model for problems 1 and 2 finally find the optimal solution. The results showed the allocation of budget and expansion. The problem 2 leads to a higher cost due to the more factors considered. The model for problem 3 does not find a feasible solution due to some special point that extremely affects the child care arrangement. It is recommended that analyze and treat them separately.

The further study could explore several approaches to enhance the accuracy of the optimization model in the usage of child care topics. Firstly, exploring the dynamic model to reflect the changes in data over time. Some rare event such as Covid-19 can significantly affect the child care data. The dynamic model can also consider the demographic shift or change in the economic market. Secondly, the non-linear method can also be introduced for solving the child care desert issue. Other potential factors can also be investigated when building the optimization model. Lastly, researchers can use more data from different data sources and time periods to increase the generalizability and reliability of the model. Various data sources offer a more comprehensive understanding of the variables impacting the demand for child care.

# 7 Appendix

## 7.1 Full optimization model for Problem 1

$$\min C = \sum_{j=1}^{3} c_j x_{ji} + \sum_p c^e_{1pi} \frac{(x_{1pi} + x_{2pi})}{q_{pi}} + 100 \sum_p x_{1pi},$$

subject to:

$$\sum_{j=1}^{3} s_{j1} x_{ji} + \sum_p (x_{1pi} + x_{2pi}) \geq k_i a_i, \quad \forall i$$

$$\sum_{j=1}^{3} s_{j2} x_{ji} + \sum_p x_{1pi} \geq \frac{2}{3} b_i, \quad \forall i$$

$$x_{1pi} + x_{2pi} \leq 0.2 q_{pi}, \quad \forall p, i$$

$$x_{1pi} + x_{2pi} \leq (500 - q_{pi}) o_{pi}, \quad \forall p, i$$

$$x_{ji}, x_{1pi}, x_{2pi} \geq 0, \quad \forall j, p, i$$

$$x_{ji} \in \mathbb{Z}, \quad \forall j, i$$

$$o_{pi} \in \{0, 1\}, \quad \forall p, i$$

## 7.2 Full optimization model for Problem 2

$$\min C = \sum_{j=1}^{3} \sum_l c_j f_{jl} + \sum_p \sum_{j=1}^{3} c^\delta_{jp} \delta_{jpi} + 100 \sum_p x_{1pi},$$

subject to:

$$\sum_l m_{li} f_{jl} = x_{ji}, \quad \forall j, i$$

$$\sum_{j=1}^{3} f_{jl} \leq 1, \quad \forall l$$

$$\sum_{j=1}^{3} s_{j1} x_{ji} + \sum_p (x_{1pi} + x_{2pi}) \geq k_i a_i, \quad \forall i$$

$$\sum_{j=1}^{3} s_{j2} x_{ji} + \sum_p x_{1pi} \geq \frac{2}{3} b_i, \quad \forall i$$

$$\sum_{L \neq l} n_{lL} \left( \sum_{j=1}^{3} f_{jL} \right) \leq 1, \quad \forall l$$

$$x_{1pi} + x_{2pi} = \delta_{1pi} + \delta_{2pi} + \delta_{3pi}, \quad \forall p, i$$

$$0 \leq \delta_{1pi} \leq 0.1 q^{(12)}_{pi}, \quad \forall p, i$$

$$0 \leq \delta_{2pi} \leq 0.05 q^{(12)}_{pi}, \quad \forall p, i$$

$$0 \leq \delta_{3pi} \leq 0.05 q^{(12)}_{pi}, \quad \forall p, i$$

$$x_{1pi} + x_{2pi} \leq (500 - q_{pi}) o_{pi}, \quad \forall p, i$$

$$x_{1pi}, x_{2pi} \geq 0, \quad \forall j, p, i$$
$$r_{pi}, \delta_{1pi} \geq 0, \quad \forall j, p, i$$
$$o_{pi}, n_{lL}, f_{jl} \in \{0, 1\}, \quad \forall j, l, L \neq l$$
$$\delta_{1pi} \in \mathbb{Z}, \quad \forall p, i$$

### 7.3 Full optimization model for Problem 3

$$
\begin{aligned}
\max \quad & C.I. = \frac{2}{3}\sum_i h_i + \frac{1}{3}\sum_i g_i, \\
\text{s.t.} \quad & \sum_{j=1}^{3}\sum_l c_j f_{jl} + \sum_p \sum_{j=1}^{3} c_{jp}^{\delta}\delta_{jpi} \leq 10^9 \\
& h_i = \frac{\sum_{j=1}^{3} s_{j1}x_{ji} + \sum_p (x_{1pi} + x_{2pi}) + \sum_p q_{pi}^{(12)}}{a_i} \quad \forall i \\
& g_i = \frac{\sum_{j=1}^{3} s_{j2}x_{ji} + \sum_p x_{1pi} + \sum_p q_{pi}^{(5)}}{b_i} \quad \forall i \\
& h_{max} - h_{min} \leq 0.1 \quad \forall n, m \\
& h_{max} \geq h_i \quad \forall i \\
& h_{max} \leq h_i + My_i \quad \forall i \\
& h_{min} \leq h_i \quad \forall i \\
& h_{max} \geq h_i - Mz_i \quad \forall i \\
& \sum_{i \in I} y_i = |I| - 1 \\
& \sum_{i \in I} z_i = |I| - 1 \\
& \sum_l m_{li} f_{jl} = x_{ji} \quad \forall j, i \\
& \sum_{j=1}^{3} f_{jl} \leq 1 \quad \forall l \\
& h_i \geq k_i \quad \forall i \\
& g_i \geq \frac{2}{3} \quad \forall i \\
& \sum_{L \neq l} n_{lL}(\sum_{j=1}^{3} f_{jL}) \leq 1 \quad \forall l \\
& x_{1pi} + x_{2pi} = \delta_{1pi} + \delta_{2pi} + \delta_{3pi}, \quad \forall p, i \\
& 0 \leq \delta_{1pi} \leq 0.1 q_{pi}^{(12)} \quad \forall p, i \\
& 0 \leq \delta_{2pi} \leq 0.05 q_{pi}^{(12)} \quad \forall p, i \\
& 0 \leq \delta_{2pi} \leq 0.05 q_{pi}^{(12)} \quad \forall p, i \\
& x_{1pi} + x_{2pi} \leq (500 - q_{pi}) o_{pi} \quad \forall p, i \\
& x_{1pi}, x_{2pi} \geq 0 \quad \forall j, p, i \\
& r_{pi}, \delta_{1pi} \geq 0 \quad \forall j, p, i \\
& y_i, z_i o_{pi}, n_{lL}, f_{jl} \in \{0, 1\} \quad \forall i, j, l, L, L \neq l \\
& \delta_{1pi} \in \mathbb{Z} \quad \forall p, 1
\end{aligned}
\tag{35}
$$

## Bibliography

Adams, G. and Henly, J. R. (2020). Child care subsidies: Supporting work and child development for healthy families.

Bradley, R. H. and Vandell, D. L. (2007). Child care and the well-being of children. *Archives of pediatrics & adolescent medicine*, 161(7):669–676.

CAP (n.d.). *Child Care Deserts*. `https://www.americanprogress.org/series/child-care-deserts/`.

Johnson, J. O. (2005). *Who's Minding the Kids?: Child Care Arrangements, Winter 2002*. US Department of Commerce, Economics and Statistics Administration, US ....

NICHD (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science*, 4(3):116–135.

NSECE (2016). Early care and education usage and households' out-of-pocket costs: Tabulations from the national survey of early care and education (nsece). *OPRE Report 2016-09. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US Department of Health and Human Services.*