Optimizing Child Care Accessibility through a Linear Programming Approach: A Strategic Solution to Eliminating Child Care Deserts in New York State

November 6th, 2024

Abstract

Our project addresses the issue of childcare deserts across New York State by developing an optimized strategy to allocate resources for new facility construction and the expansion of existing childcare centers. Using a linear programming model, we incorporated factors such as regional childcare demand, facility expansion costs, and state-specific policies prioritizing access for children under age 5. Our findings indicate that a targeted combination of facility expansion and new constructions is effective in eliminating childcare deserts while remaining within budget constraints. Additionally, our model enforces a fairness constraint, ensuring a balanced distribution of childcare access across all regions and prioritizing areas with higher demand. Under these optimized conditions, we achieved a high social coverage index, significantly improving childcare availability across the state.

Contents

1 Introduction

Child care has extremely important significance for families and even society. For young children, high-quality early childhood education helps shape a favorable environment for their early growth, which in turn helps them achieve long-term success in life(Southern Regional Education Bauer Schanzenbach,2016). In addition, good early childhood education can alleviate parents' stress, help them focus on their work, and enhance their professional competitiveness. For society as a whole, sufficient child care resources help employers retain employees and effectively reduce future crime rates among children(Elango et al., 2015). It can be seen that providing sufficient child care has multiple positive implications, but due to various factors such as income, child care resources are relatively insufficient. According to relevant data, including many parts of New York, 51 of the United States is facing the dilemma of child care deserts. The National Governors Association (NGA) believes that the factors that hinder teenagers from accessing mental health services are high nursing costs, imbalanced resources in mental health services among states.In the survey on the causes of the children's fox desert, the most commonly cited reasons are insufficient operating funds from state and federal governments, as well as the high financial costs required to operate child care centers(Meant,2019). In addition, previous studies have shown that a lack of economic support for providers is one of the main reasons for childcare in deserts. Therefore, using optimization methods to assist the New York City government in developing plans that can provide sufficient child care resources with minimal budget constraints and ensure fairness has positive practical significance.

2 Background

Based on the available information, this paper transforms the design of an optimal solution to address childcare deserts under a limited budget into a linear programming problem. Due to real-world conditions and policy constraints, multiple assumptions are applied in solving the problem.

First, based on household income and parental employment status, different areas in New York City are classified into high-demand and normal-demand childcare regions: regions where at least 60% of parents are employed or the average income is $60,000 or less per year are considered high-demand areas. For high-demand areas, if available childcare slots meet at least half of the number of children aged 2 to 12, the area is not considered a childcare desert. For normal-demand areas, it only needs to reach one-third of the number of children in the same age group. In addition, New York City policies require that available slots cover at least two-thirds of the number of children aged 0 to 5. When expanding childcare slots, the government faces two options: selecting locations to build new facilities (which are of three types, each with different capacities and construction costs) or expanding existing facilities, with a maximum expansion of 500 slots or up to 0.2 times the original capacity, where costs increase with additional slots. This part of the problem is discussed in the first section of the main text.

Second, to address childcare challenges, New York City needs to construct new childcare facilities. To avoid excessive concentration of resources, there are distance constraints between facilities, which is discussed in the second section of the main text.

Finally, considering social equity, the New York City government aims to ensure relatively equitable access to childcare resources across regions within the budget while maximizing the social childcare coverage index. This part of the problem is discussed in the third section of the main text.

In summary, this paper simplifies the solution to the childcare desert issue in New York City into three linear optimization problems, and solves them using Gurobi and Python, discussing the practical implications of the results.

3 Methodology

3.1 The Problem of Budgeting

3.1.1 Model Assumptions

We start with the fundamental assumption that the child care scenario in NYC can be modeled as a linear program with a finite, feasible, and convex solution space. This is essential, as it provides the foundation for our linear programming model. Additionally, we assume that any existing facility with a current capacity of 0 will not be expanded. This assumption aligns with the constraint that a facility's maximum capacity is the lesser of 500 or 1.2 times its current capacity. Since a multiplier of 1.2 on a capacity of 0 remains 0, these facilities cannot be expanded. This assumption is particularly important because a few existing facilities have a capacity of zero.

Moreover, we assume that an unlimited number of new facilities can be established in each area, as specific geographic constraints are not part of the problem. Finally, we assume that all children within a given area can access any facility constructed in that area.

3.1.2 Preliminaries

Throughout this section, we will use the following notations.

Z	: Set of zip codes.
F_z	: Set of existing facilities in zip code z.
S	: Set of facility types available for new construction (Small, Medium, Large).
$Demand_z$	: Demand classification in zip code z (High or Normal-Demand).
$P_{z,0-5}$	: Population of children aged 0-5 in zip code z.
$P_{z,5-12}$	: Population of children aged 5-12 in zip code z.
$P_{z,0-12}$	: Population of children aged 0-12 in zip code z.
$C_{f,z,0-5}$	: Current capacity (slots) for children aged 0-5 at facility f in zip code z.
$C_{f,z,5-12}$	: Current capacity (slots) for children aged 5-12 at facility f in zip code z.
$C_{f,z}$	: Total current capacity at facility f in zip code z.
$TotalSlots_s$	: Total capacity (slots) of facility size s.
$Under5Slots_s$	: Number of slots for children aged 0-5 in facility size s.
$Cost_s$	: Construction cost of facility size s.

3.1.3 Decision Variables

$I_{z,f}$	: Number of additional slots to be added through expansion at facility f in zip code z.
$U_{z,f}$	: Number of new under-5 slots added at facility f in zip code z.
$N_{z,f}$	: Number of new facilities of size s to be built f in zip code z.

3.1.4 Model Constraints

First, we deal with total slots constraints for all children under age 12. Assume that we have identified the category of demand in zip code z, we can write the total required number of slots $RequiredTotalSlots_z$ as:

$$RequiredTotalSlots_z = \begin{cases} \frac{1}{2}P_{z,0-12} & \text{if } demand_z \text{ is High} \\ \\ \frac{1}{3}P_{z,0-12} & \text{if } demand_z \text{ is Normal} \end{cases}$$

The first set of constraints dictates that the total slots after construction and expansion should be more than the total required slots in each area z.

$$\sum_{f \in F_z} (C_{z,f,0-12} + I_{z,f}) + \sum_{s \in S} N_{z,s} \times TotalSlots_s \geq RequiredTotalSlots_z \quad \forall z \in Z$$

Similarly, the constraints for total slots under age 5 are:

$$\sum_{f\in F_z}(C_{z,f,0-5}+I_{z,f})+\sum_{s\in S}N_{z,s}\times Under5Slots_s \geq RequiredUnder5Slots_z \quad \forall z\in Z$$

Next, we consider the maximum expansion for each existing facility, which cannot exceed 20% of its capacity.

$$I_{z,f}\leq 0.2C_{z,f,0-12} \quad \forall z\in Z \quad \forall f\in F_z$$

Also, note that the capacity of each facility after expansion cannot exceed 500, assuming $C_{z,f}$ is less than 500.

$$I_{z,f}+C_{z,f}\leq 500 \quad \forall z\in Z \quad \forall f\in F_z$$

Note that if $C_{z,f}$ already exceeds 500, the constraint should be $I_{z,f}=0$ instead. Also, notice that the number of 0-5 slots cannot exceed the number of all slots in each facility. Therefore:

$$U_{z,f}\leq I_{z,f} \quad \forall z\in Z \quad \forall f\in F_z$$

Finally, each variable should be a positive integer.

$$I_{z,f},U_{z,f},N_{z,s}\in\mathcal{Z}^+ \quad \forall z\in Z \quad \forall f\in F \quad \forall s\in S$$

3.1.5 Objective

The objective, which is the total budget, can be divided into two parts. First is the total cost of constructing a new facility, and second part is the total cost of expanding existing ones. Mathematically, it can be formulated as follows.

$$\mathbf{min}\sum_{z\in Z}\left\{\sum_{f\in F_z}(20000+200C_{z,f,0-12})\frac{I_{z,f}}{C_{z,f,0-12}}+100U_{z,f}+\sum_{s\in S}N_{z,s}Cost_s\right\}$$

3.2 The Problem of Realistic Capacity Expansion and Distance

3.2.1 Preliminaries

This problem can be viewed as an extension of previous section. We will continue to use all notations defined before and introduce some new notations for this problem.

- L_z : All potential locations for new facilities in zip code z.
- E_z : All locations of existing facilities in zip code z.
- T : Expansion tiers (1, 2, 3).
- $D_{i,j}$: Distance between location i and location j.

As the problem introduces distance constraints, the number of new facilities in each area is now restricted by a certain number. We can calculate each region's maximum number of new facilities by a separate linear program.

3.2.2 Sub-Problem: Maximum Facility Per Area

We will use the following integer program to compute the maximum number of facilities that can be built in a certain area z. For all potential locations $l \in L_z$, we define a binary variable y_l.

$$y_l = \begin{cases} 1 & \text{if build in location } l \\ 0 & \text{Otherwise} \end{cases}$$

The constraints are that for each pair of locations if the distance between them is smaller than 0.06 miles, only one of them can build a new facility.

$$\text{If} \quad D_{i,j} \leq 0.06, \quad y_i + y_j \leq 1 \quad \forall i \in L_z \quad \forall j \in L_z \quad i \neq j$$

Also, no new facility can be built near an existing facility.

$$\text{If} \quad D_{i,j} \leq 0.06, \quad y_i = 0 \quad \forall i \in L_z \quad \forall j \in E_z$$

The objective is:

$$\mathbf{max} \sum_{l \in L_z} y_l$$

The optimal value of this linear problem will be the maximum number of new facilities to be built in area z. We denote this as m_z.

3.2.3 Decision Variables

Similarly, we will only state new variables that are introduced in this problem. All other decision variables in the former problem will be contained in this problem as well.

$\delta_{z,f,t}$: Binary variable indicating whether tier t is fully utilized at facility f. $\forall z \forall f \forall t$

$$\delta_{z,f,1} = \begin{cases} 1 & \text{if expansion rate} \leq 10\% \\ 0 & \text{Otherwise} \end{cases}$$

$$\delta_{z,f,1} = \begin{cases} 1 & \text{if } 10\% \leq \text{expansion rate} \leq 15\% \\ 0 & \text{Otherwise} \end{cases}$$

$$\delta_{z,f,1} = \begin{cases} 1 & \text{if } 15\% \leq \text{expansion rate} \leq 20\% \\ 0 & \text{Otherwise} \end{cases}$$

3.2.4 Model Constraint

In addition to the constraints in the previous problem, we should also consider the following set of constraints for this model. First, we consider the piecewise cost function constraint, where only one of $\delta_{z,f,1}, \delta_{z,f,2}, \delta_{z,f,3}$ can be 1, which indicates that the expansion rate is in the corresponding region.

$$I_{z,f} \leq 10\% \cdot C_{z,f,0-12}\delta_{z,f,1}$$

$$10\% \cdot C_{z,f,0-12} \cdot \delta_{z,f,2} \leq I_{z,f} \leq 15\% \cdot C_{z,f,0-12} \cdot \delta_{z,f,2}$$

$$15\% \cdot C_{z,f,0-12} \cdot \delta_{z,f,3} \leq I_{z,f} \leq 20\% \cdot C_{z,f,0-12} \cdot \delta_{z,f,3}$$

$$\sum_{i=1}^{3} \delta_{z,f,i} = 1 \quad \forall z \in Z \quad \forall f \in F_z$$

Note that $C_{z,f,0-12}$ is a constant and therefore the constraints are linear. Moreover, these constraints model the changing rate of expansion cost. Next, we need to specify a constraint for the total number of facilities that can be built in each area.

$$\sum_{s \in S} N_{z,s} \leq m_z \quad \forall z \in Z$$

Finally, the non-negative constraints.

$$I_{z,f,t} \in \mathcal{Z}^+ \quad \forall z \in Z \quad \forall f \in F \quad \forall t \in T$$

$$\delta_{z,f,i} \in \{0,1\} \quad \forall z \in Z \quad \forall f \in F \quad \forall i \in \{1,2,3\}$$

3.2.5 Objective

The objective has changed since the cost function has changed for expanding existing facilities. However, the cost of building new facilities is unchanged.

$$\mathbf{min} \sum_{z \in Z} \left\{ \sum_{f \in F_z} (20000 + ExpansionPerSlotCost_{z,f} C_{z,f,0-12}) \frac{I_{z,f}}{C_{z,f,0-12}} + 100U_{z,f} + \sum_{s \in S} N_{z,s} Cost_s \right\}$$

where $ExpansionPerSlotCost_{z,f} = 200\delta_{z,f,1} + 400\delta_{z,f,2} + 600\delta z, f, 3$

3.3 Fairness Problem

3.3.1 Preliminaries

We will introduce the following notation for this problem:

$Coverage_z$	:	The coverage rate of all children under age 12 for area z.
$Coverage5_z$	:	The coverage rate of all children under age 5 for area z.

These parameters can be calculated by the following formulas.

$$Coverage_z = \frac{\sum_{f \in F_z} (C_{z,f,0-12} + I_{z,f}) + \sum_{s \in S} N_{z,s} \times TotalSlots_s}{P_{z,0-12}}$$

$$Coverage5_z = \frac{\sum_{f \in F_z} (C_{z,f,0-5} + I_{z,f}) + \sum_{s \in S} N_{z,s} \times Under5Slots_s}{P_{z,0-5}}$$

3.3.2 Decision Variables

No additional decision variables are needed for this problem.

3.3.3 Model Constraint

There are two more constraints we need to add to this problem. First, the coverage between any two areas should not exceed 1. This can be formulated as follows:

$$Coverage_{z_1} - Coverage_{z_2} \leq 0.1 \quad \forall z_1 \in Z \quad \forall z_2 \in Z \quad z_1 \neq z_2$$

Secondly, the total budget is 1 billion dollars. We can take the objective function from the previous problem.

$$\sum_{z \in Z} \left\{ \sum_{f \in F_z} (20000 + ExpansionPerSlotCost_t \cdot C_{z,f,0-12}) \frac{I_{z,f}}{C_{z,f,0-12}} \right. \\ \left. + 100U_{z,f} + \sum_{s \in S} N_{z,s} \cdot \text{Cost}_s \right\} \leq 1000000000 \quad (1)$$

3.3.4 Objective

The objective is to maximize average social index, which can be formulated as follow.

$$\mathbf{max} \sum_{z \in Z} Coverage_z + 2 \sum_{z \in Z} Coverage5_z$$

4 Data Cleaning and Visualization

4.1 Data Cleaning

In this study, we focus on areas, characterized by zip codes, in New York City. The data provided contains 1024 areas. Furthermore, there are three available options for building a new facility, as specified in Table 1.

Facility Size	**# of Slots (Ages 0-5)**	**Cost of New Facility ($)**
100 slots (Small)	50 slots	65,000
200 slots (Medium)	100 slots	95,000
400 slots (Large)	200 slots	115,000

Table 1: Construction cost estimates for different sizes of child care facilities

The data set provided contains population for ages 0 to 5, 5 to 9, and 9 to 14. Since we will only use population from 0 to 12, we estimate this by adding ages 0 to 5, 5 to 9, and three-fifths of ages 9 to 14.

We estimate 0-5 capacity in an existing facility by adding the capacities of infant, toddler, preschool, and five-twelfth of children.

We also drop any existing facilities with a total capacity of zero, as they cannot be expanded anyways.

4.2 Visualization

4.2.1 Children Population Heatmap

To better visualize the need in each area, we created a heatmap that shows the children population in each area. As in Figure 1.

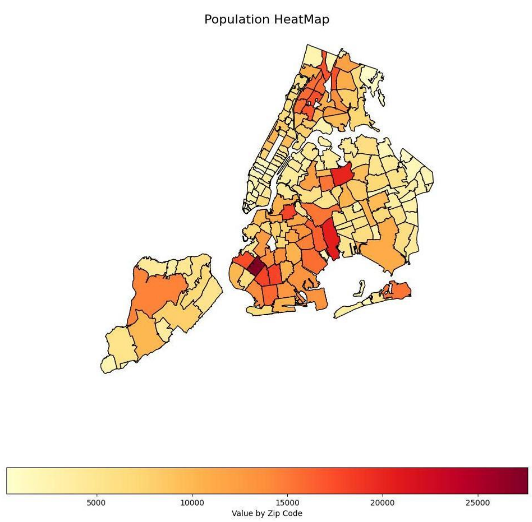

Figure 1: Heatmap of NY Child Population Distribution

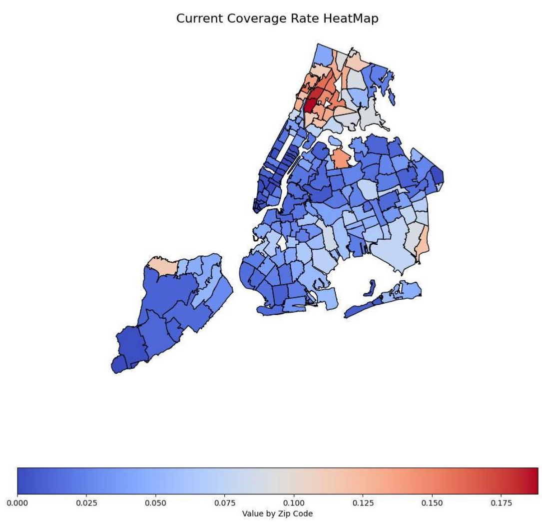

Figure 2: Heatmap of Original Social Coverage Rate

4.2.2 Current Coverage Rate

Figure 2 shows the current child care coverage rate (before optimization), calculated by $\frac{ExistingSlots}{Population}$. From the heatmap, we observe that the coverage rate in all areas is lower than 0.125, therefore all areas are below the requirement.

5 Case Study and Results

5.1 Problem of Budgeting Results

After optimization, we can successfully meet all demands and constraints by using **\$316,248,854**. To analyze our data we first begin by grouping our zip codes into groups of 11 by their proximity. This makes the data less granular and easier to view in charts. Once the data is aggregated into groups of 11, we first look at how often each size of building was built in each group:

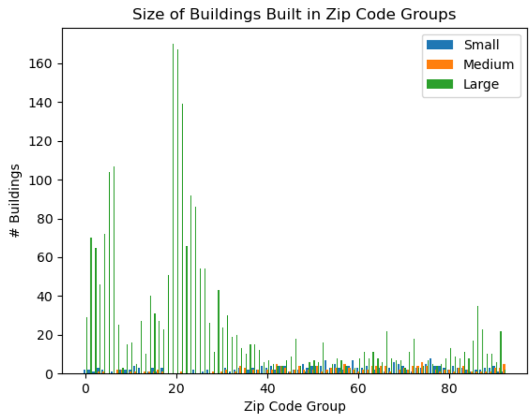

Figure 3: Distribution of New Facility Sizes

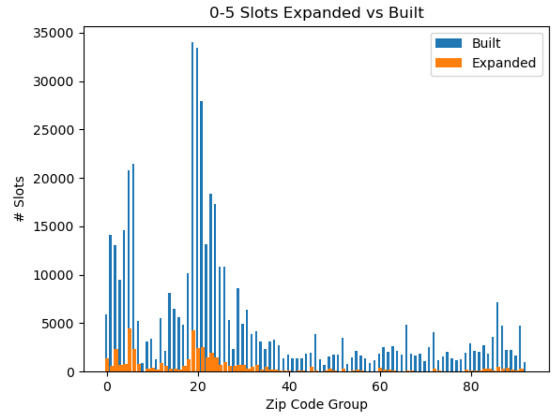

Figure 4: Expanded Slots Breakdown By Size

From Figure 3, we can see that the overwhelming majority of the time, large buildings are built in our zip codes. This makes sense as they provide a very large number of slots for less than double the cost of small facilities that are built. We also see that the amount of small and medium buildings is mostly flat across all the zip code groups, while the large buildings frequency has a discernible shape. One unique point is that the peak of the large buildings being built happens in zip code group 25, which corresponds to zip codes in the Northern Poughkeepsie area. This area may be susceptible to childcare deserts due to its low socioeconomic status in its resident population. Next, we look at expansion vs building new facilities. The results are seen below:

In Figure 4 we can see that most slots are added by building new facilities rather than expanding current facilities. Looking deeper, we can see that both distributions roughly have the same shape. This implies that when slots are added, usually you must build new facilities and expand existing ones, you just obtain more slots by building. It also implies that there is no zip code or geography where it is disproportionally better to expand vs build new slots, the shapes are roughly the same. Finally, we look at expansion in detail, analyzing which type of slots are usually added when facilities are expanded. The results are shown in Figure 5:

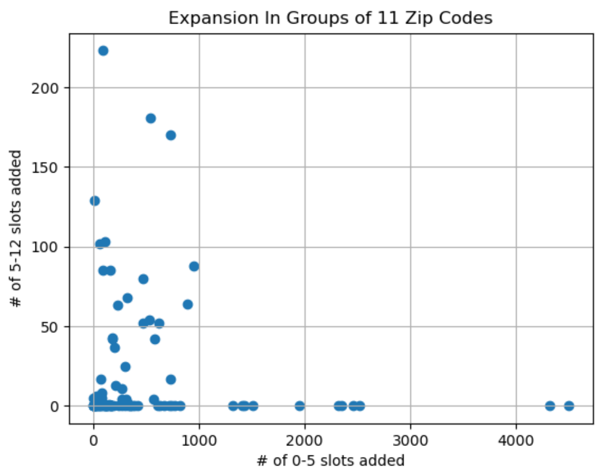

Figure 5: Expanded Slots Breakdown By Size

In Figure 5, we see a density of points on the x-axis, implying that most of the time you will be adding 0-5 slots. This makes sense as most existing facilities had a shortage of 0-5 slots. We also see that the range of the 0-5 slots is a lot higher, reaching 4000 in some zip code groups. Whereas for 5-12 slots, we add at most 250 in a zip code group. The last deduction that can be made from the data is that 5-12 slots are only really built when not a lot of 0-5 slots are built. This indicates that perhaps facilities may have a shortage in one type of slot or the other. As a comparison to Figure 2, the heatmap of social coverage rate can be found in Figure 6.

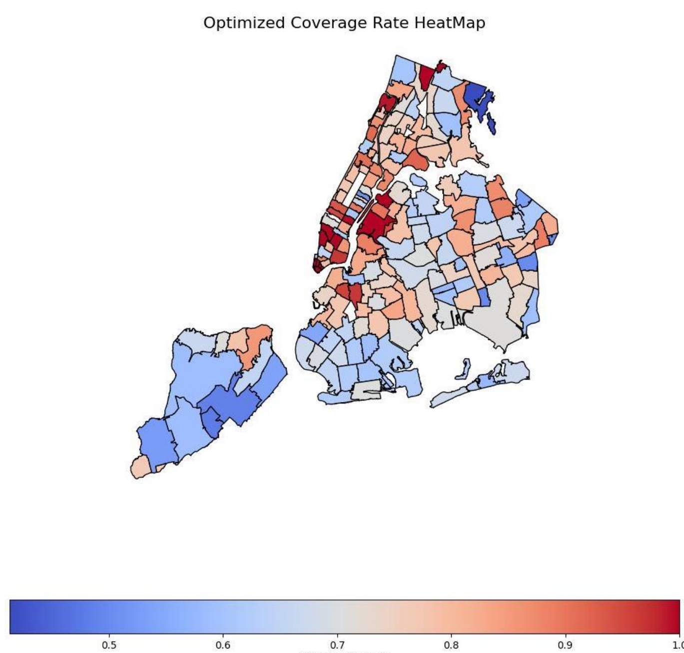

Figure 6: Expanded Slots Breakdown By Size

5.2 Problem of Realistic Capacity Expansion and Distance Results

After optimization, we can successfully meet all childcare demands and facility constraints by using **$320,532,485**. Once again, to analyze our data we first begin by grouping our zip codes into groups of 11 by their proximity. This makes the data less granular and easier to view in charts. Once the data is aggregated into groups of 11, we first look at how often each size of building was built in each group:

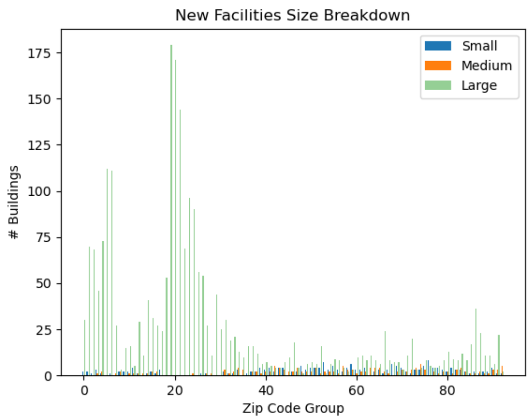

Figure 7: Expanded Slots Breakdown By Size

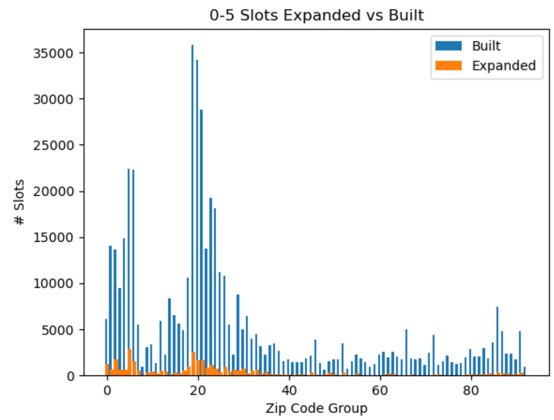

Figure 8: Expanded Slots Breakdown By Size

From the chart, we can see that there is a similar pattern to problem 1 despite there being new constraints with locations where new facilities can be built. The overwhelming majority of new facilities that are built are still large. This makes sense as the fixed cost for building a new facility has not changed from problem 1 and thus a large building still provides a large number of slots for a disproportionate premium to small buildings. However, one new observation in problem 2 is that slightly more large facilities are built in the geographic zone around zip code 20. We have a peak of 175 large facilities in those groups as opposed to 160 large facilities from problem 1. This could be due to the fact that the piecewise cost introduced in problem 2 encourages building in those zip codes rather than expanding their sparse existing facilities. In other zip code groups, slight expansion in the lowest cost bracket could have been enough to close the desert. Next, we look at slots that were added through expansion versus by building new facilities:

In this chart we see a similar trend as problem 1, where most slots come from building new facilities as opposed to expanding. The distribution follows the same bimodal shape as well, with peaks around zip code group 8 and zip code group 23. The key distinction is that the number of 0-5 slots that come from expansion is now fewer than problem 1. It is still the same shape, meaning that expansion still usually happens in the same geographies, but the magnitude is lower. This makes sense as larger expansions now cost proportionally more per slot. This incentivize the model to prefer building over expanding as the cost for building new facilities did not change from problem 1. From looking at the shape of the distributions, we can also deduce that the new location constraints did not hugely affect the location in which slots are added. For example, if a certain zip code group had a large concentration of existing facilities, it would be hard to build any new locations there and thus the shape of our chart in that zip code group would be lower. However, the shape seems to largely be the same. To confirm this, further research must be done into the granular locations of the data. Finally, we can look at the age demographic for which we expanded slots in our facilities:

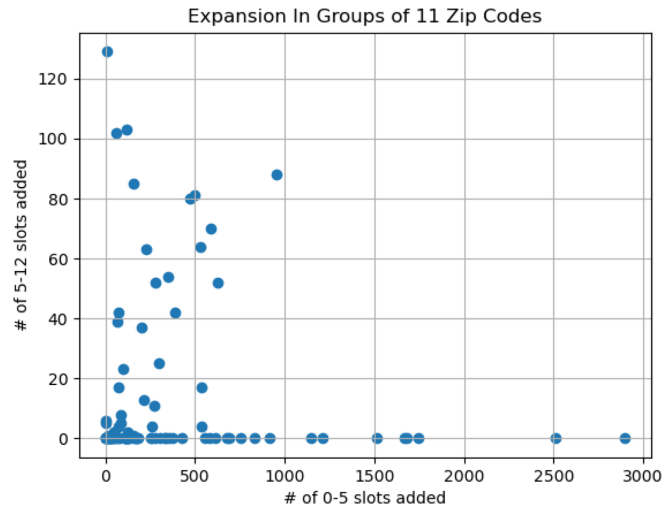

Figure 9: Expanded Slots Breakdown By Size

From this chart, we see a density in the x-axis, implying that when slots are added, most of the are in the 0-5 age capacity. This makes sense as most existing facilities had a shortage in the 0-5 age capacity slots. Next we see an elbow shape which implies that in most scenarios, you are either expanding 0-5 slots or 5-12 slots, but not both. We also see a peak in the 5-12 slots when you are not adding a lot of 0-5 slots. In particular, we only see a large number of 5-12 slots being added when only a few 0-5 slots are added. We also see that the range of the x-axis is far larger than that of the y-axis. This means that in most scenarios we had to add far more 0-5 slots than 5-12 slots. This makes sense as the demand requirement for 0-5 slots is stricter for NYS as mentioned in the instructions. One difference from problem 1 is that the largest points in this scatter plot are lower than those from problem 1. This makes sense as with the new piecewise cost function, it can get more expensive to expand and thus building is preferred. This trend is also seen in the previous bar graph where more slots came from new facilities rather than expanding existing facilities.

5.3 Problem of Fairness Results

The model of the problem of fairness turns out to be infeasible, which means that with the given funding ($1 billion), we cannot address the problem of fairness.

6 Conclusion

First, in the first part of the main text, this paper provides an optimal budget solution to meet childcare needs by expanding existing facilities and constructing new ones, assuming there are no geographical constraints or distance limitations. This is solved using a linear programming model. Under this scenario, the minimum budget achieved is approximately 316 million USD.

Based on the first scenario, the model in the second part of the main text incorporates distance limitations between facilities and the non-linear cost function associated with expanding existing facilities in real-world settings. To address the distance constraint, two strategies are employed: first, by calculating the distance between potential new facilities and existing facilities, where if the distance is less than the specified minimum of 0.6 miles, it is recorded as a constraint, prohibiting construction of a new facility at that location; second, for each potential location within a zip code, the distance between them is calculated, and if it is less than 0.6 miles, a constraint is added to allow only one facility to be constructed between those two locations. Under this scenario, the minimum budget required by the model is approximately 321 million USD.

In the third part of the main text, the model considers the equity of resource distribution to ensure minimizing disparities in childcare services between different areas. Additionally, the New York City government aims to maximize the social childcare coverage index within the budget constraints. To achieve this, new constraints are introduced, and a linear model is developed with the objective of maximizing the social childcare coverage index. However, it is ultimately found that the issue of equity cannot be resolved under the current conditions.

This issue highlights the advantages of data-driven decision-making in helping the government make effective use of fiscal spending, providing solutions, and promoting social equity. In the future, this approach is expected to help communities build an environment conducive to children's growth, enhance the government's ability to provide public resources, and has significant real-world implications.

References

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

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

[3] Bauer, L., & Schanzenbach, D. W. (2016). The long-term impact of the head start program. The Hamilton Project

[4] Means, S. F. (2019). Kentucky Child care deserts and where to find them. Prichard Committee for Academic Excellence Ed Blog.

[5] Elango, S., Garcia, J. L., Heckman, J., & Hojman, A. (2015). NBER working paper series: Early childhood education. National Bureau of Economic Research.