IEOR 4004 – Optimization Project 1

Fall 2025, Columbia University

Authors: Tommaso Ravasio, Jiayi Dan, Chenyao Yuan, Zhaojie Yan, Ruolin Bao
Instructor: Yaren Bilge Kaya

---

1 Introduction

To address child care desert elimination in New York State, we develops three sequential mixed-integer linear programming models to determine optimal investment strategies for expanding child care capacity. The first model establishes a baseline budgeting solution, the second incorporates realistic spatial and cost constraints, while the third introduces fairness considerations under budget limitations. Our work provides data-driven insights for public policy planning, demonstrating how operational research techniques can inform resource allocation decisions in social infrastructure development. Code is available at https://github.com/DanJiayi/IEOR-4004-Project-1

2 Background

Child care deserts—areas where demand for licensed care significantly exceeds supply—affect over half of American children. New York State faces particular challenges, with many regions classified as deserts based on specific coverage requirements. The state employs a tiered approach where high-demand areas must meet higher capacity thresholds compared to normal-demand regions. Additionally, specialized provisions ensure adequate coverage for children aged 0-5. Beyond these capacity requirements, spatial considerations through distance constraints between facilities play a crucial role in realistic planning. These policy requirements and operational constraints form the foundation for our optimization models, which aim to identify cost-effective strategies for eliminating child care deserts statewide.

3 Methods

We employ mixed-integer linear programming (MILP) formulated in Python and solved using Gurobi Optimizer. For Problem 1, we adopt a decentralized approach, solving independent MILPs for each ZIP code to minimize total expansion and construction costs. Problem 2 extends this foundation with spatial constraints using Haversine distance calculations. Problem 3 introduces equity constraints and global budget limitations to maximize social welfare. All models incorporate realistic facility expansion limits, construction options, and age-specific capacity requirements.

4 Data Preprocessing

- All ZIP codes are standardized to five digits (left–padded with zeros if needed).
- We restrict the analysis to the intersection of data sources critical: only ZIP codes that appear in both `population.csv` and `potential_locations.csv` are retained. All subsequent requirements and results are computed on this filtered set.
- Missing data (including population, employment rate, and income) are imputed using the mean value.
- For Problem 2 and Problem 3, existing facilities and potential new construction sites are filtered for valid coordinates.

5 Problem 1 – Budgeting

This problem addresses the allocation of public funding to expand child-care capacity across New York State (NYS). Each ZIP code must receive sufficient investment so that it is no longer a *child-care desert* and so that it satisfies the state's 0–5 coverage rule. The objective is to determine, for every ZIP code, the combination of expansions of existing facilities and construction of new ones that **minimizes total cost while meeting the policy requirements**.

5.1 Problem Formulation

Demand rules. For each ZIP i, the total number of children is estimated as

$$P_i^{\mathrm{tot}} = P_i^{(0\text{–}5)} + P_i^{(5\text{–}9)} + \tfrac{3}{5} P_i^{(10\text{–}14)},$$

capping the age range at 12. (The last term is multiplied by $\frac{3}{5}$ to approximate children aged 10–12 within the 10–14 census group, since data are only available in five-year bins.)

A ZIP is labeled *high-demand* if its employment rate is at least 0.6 or its average income does not exceed \$60,000; otherwise it is *normal-demand*. The required numbers of total and 0–5 slots are then

$$R_i^{\mathrm{tot}} = \begin{cases} 0.5\, P_i^{\mathrm{tot}}, & \text{high-demand}, \\ 0.33\, P_i^{\mathrm{tot}}, & \text{normal-demand}, \end{cases} \qquad R_i^{(0\text{–}5)} = 0.67\, P_i^{(0\text{–}5)}.$$

Available actions. Two options can increase capacity:

- *Expansion of existing facilities*, limited to min(20% of current capacity, 500) slots per site.
- *Construction of new facilities* of three standardized sizes:

Type	Total cap.	0–5 cap.	Fixed cost ($)
Small	100	50	65,000
Medium	200	100	95,000
Large	400	200	115,000

Each new 0–5 slot, whether through expansion or new construction, adds $100 in equipment cost.

Optimization model. For each ZIP code i, we formulate a mixed-integer linear program (MILP) that determines the least-cost combination of expansions and new constructions required to satisfy the demand constraints.

Decision variables:

- $n_s \in \mathbb{Z}_{\geq 0}$: number of new facilities of size $s \in \{\text{small}, \text{medium}, \text{large}\}$;
- $x_e \in \mathbb{Z}_{\geq 0}$: number of additional slots created by expanding existing facility e;
- $y_s^{\text{new}} \in \mathbb{Z}_{\geq 0}$: number of 0–5 slots assigned within new facilities of type s;
- $y_e^{\text{exp}} \in \mathbb{Z}_{\geq 0}$: number of 0–5 slots assigned within the expansion of facility e;
- $z_e \in \{0, 1\}$: binary variable equal to 1 if the expansion of facility e reaches at least its original capacity (triggering the baseline cost).

Objective: minimize the total cost of all actions,

$$\min \underbrace{\sum_e \text{Cost}_{\text{exp}}(x_e)}_{\text{expansion cost}} + \underbrace{\sum_e C_{\text{base},e} z_e}_{\text{baseline cost}} + \underbrace{\sum_s C_s n_s}_{\text{construction cost}} + \underbrace{100\left(\sum_e y_e^{\text{exp}} + \sum_s y_s^{\text{new}}\right)}_{\text{equipment cost}}.$$

Constraints:

$$\begin{aligned}
&\text{(i) Total slots requirement:} && \text{Existing}_i + \sum_e x_e + \sum_s \text{cap}_s n_s \;\geq\; R_i^{\text{tot}}, \\
&\text{(ii) 0–5 slots requirement:} && \text{Existing}_i^{(0\text{–}5)} + \sum_e y_e^{\text{exp}} + \sum_s y_s^{\text{new}} \;\geq\; R_i^{(0\text{–}5)}, \\
&\text{(iii) Capacity limits:} && y_e^{\text{exp}} \leq x_e, \quad y_s^{\text{new}} \leq \text{cap}_s^{(0\text{–}5)} n_s, \\
&\text{(iv) Expansion bounds:} && x_e \leq \min\{0.2\,\text{cap}_e,\, 500\}, \\
&\text{(v) Baseline trigger:} && x_e \geq \text{cap}_e z_e, \quad x_e \leq (\text{cap}_e - 1) + M z_e, \quad M \text{ large.}
\end{aligned}$$

Expansion costs $\mathrm{Cost}_{\mathrm{exp}}(x_e)$ are modeled through a piecewise-linear function with decreasing marginal cost, capturing economies of scale. The model is solved independently for each ZIP code to minimize computational complexity while preserving feasibility across the entire state.

Assumptions. ZIP codes are optimized independently, since Problem 1 has no global budget constraint. Missing data are imputed by column means to ensure feasibility. High- versus normal-demand classification relies solely on employment and income data. All cost parameters follow the values given in the project brief.

5.2 Mathematical Formulation

Let $\mathcal{I}$ denote the set of ZIP codes retained after preprocessing:

$$\mathcal{I} \;=\; \{\, i \;:\; i \in \mathrm{ZIP}(\texttt{population.csv}) \cap \mathrm{ZIP}(\texttt{potential\_locations.csv}) \,\}.$$

Let $S = \{\text{small}, \text{medium}, \text{large}\}$ be the set of facility sizes and E_i the set of existing facilities in ZIP i.

We solve the budgeting problem independently for each ZIP code $i \in \mathcal{I}$.

5.2.1 Parameters (data).

- R_i^{tot}, $R_i^{(0\text{–}5)}$: required total and 0–5 slots.
- $\mathrm{Existing}_i$, $\mathrm{Existing}_i^{(0\text{–}5)}$: currently available total and 0–5 slots in ZIP i.
- For each existing facility $e \in E_i$: current capacity cap_e and the per-site expansion bound $U_e := \min\{0.2\,\mathrm{cap}_e,\ 500\}$.
- For each size $s \in S$: total capacity cap_s, 0–5 capacity $\mathrm{cap}_s^{(0\text{–}5)}$, fixed construction cost C_s.
- Baseline expansion cost at facility e: $C_{\mathrm{base},e}$ (triggered if expansion reaches cap_e).
- Equipment cost per 0–5 slot: \$100.

5.2.2 Decision variables.

- $n_s \in \mathbb{Z}_{\geq 0}$: number of new facilities of type $s \in S$.
- $x_e \in \mathbb{Z}_{\geq 0}$: expansion slots added at existing facility $e \in E_i$.
- $y_s^{\mathrm{new}} \in \mathbb{Z}_{\geq 0}$, $y_e^{\mathrm{exp}} \in \mathbb{Z}_{\geq 0}$: 0–5 slots assigned within new facilities of type s and within the expansion at e, respectively.
- $z_e \in \{0, 1\}$: baseline–cost trigger for facility e (1 if $x_e \geq \mathrm{cap}_e$).

5.2.3 Objective function and constraints.

Piecewise linear expansion cost funtion. To model the economies of scale in expansion costs, we employ a piecewise linear function formulation following the standard MILP modeling technique [1]. Let $\{0 = \kappa_0 < \kappa_1 < \cdots < \kappa_K\}$ be expansion breakpoints (e.g., $20, 50, 100, 200, 500$) and let c_k be the unit cost on segment $[\kappa_{k-1}, \kappa_k]$. Introduce nonnegative auxiliary variables $t_{e,k}$ ($\forall e \in E_i$, $k = 1, \ldots, K$) measuring how much of segment k is used at facility e.

Define binary variables $w_{e,k}$ for $k = 1, \ldots, K-1$, where

$$w_{e,k} = \begin{cases} 1 & \text{if segment } k \text{ is at its upper bound} \\ 0 & \text{otherwise} \end{cases}$$

Then these constraints can be represented as:

$$\begin{aligned} x_e &= \sum_{k=1}^{K} t_{e,k}, \\ (\kappa_1 - \kappa_0) w_{e,1} \le t_{e,1} &\le (\kappa_1 - \kappa_0), \\ (\kappa_k - \kappa_{k-1}) w_{e,k} \le t_{e,k} &\le (\kappa_k - \kappa_{k-1}) w_{e,k-1}, \quad \text{for } k = 2, \ldots, K-1, \\ 0 \le t_{e,K} &\le (\kappa_K - \kappa_{K-1}) w_{e,K-1}, \end{aligned}$$

with $w_{e,0} \equiv 1$ for notational convenience.

The (linear) expansion cost is $\sum_{e \in E_i} \sum_{k=1}^{K} c_k \, t_{e,k}$.

MILP for ZIP i.

$$\min \quad \underbrace{\sum_{e \in E_i} \sum_{k=1}^{K} c_k \, t_{e,k}}_{\text{expansion cost}} + \underbrace{\sum_{e \in E_i} C_{\text{base},e} \, z_e}_{\text{baseline}} + \underbrace{\sum_{s \in S} C_s \, n_s}_{\text{construction}} + \underbrace{100 \left(\sum_{e \in E_i} y_e^{\text{exp}} + \sum_{s \in S} y_s^{\text{new}} \right)}_{\text{equipment}} \tag{1}$$

$$\begin{aligned}
\text{s.t.} \quad & \text{Existing}_i + \sum_{e \in E_i} x_e + \sum_{s \in S} \text{cap}_s \, n_s \ \ge \ R_i^{\text{tot}}, && (2) \\
& \text{Existing}_i^{(0\text{–}5)} + \sum_{e \in E_i} y_e^{\text{exp}} + \sum_{s \in S} y_s^{\text{new}} \ \ge \ R_i^{(0\text{–}5)}, && (3) \\
& y_e^{\text{exp}} \le x_e \quad (\forall e \in E_i), \qquad y_s^{\text{new}} \le \text{cap}_s^{(0\text{–}5)} \, n_s \quad (\forall s \in S), && (4) \\
& 0 \le x_e \le U_e \quad (\forall e \in E_i), && (5) \\
& x_e \ge \text{cap}_e \, z_e, \quad x_e \le (\text{cap}_e - 1) + M \, z_e \quad (\forall e \in E_i), && (6) \\
& x_e = \sum_{k=1}^{K} t_{e,k}, \quad 0 \le t_{e,k} \le \kappa_k - \kappa_{k-1} \quad (\forall e \in E_i, \ \forall k = 1, \ldots, K), && (7) \\
& n_s \in \mathbb{Z}_{\ge 0} \ (\forall s \in S), \ x_e, t_{e,k} \in \mathbb{Z}_{\ge 0} \ (\forall e \in E_i, \forall k), \ y_s^{\text{new}}, y_e^{\text{exp}} \in \mathbb{Z}_{\ge 0}, \ z_e \in \{0, 1\}. && (8)
\end{aligned}$$

Constraints (2) and (3) enforce the policy requirements on total and 0–5 capacity. (4) ties 0–5 assignments to available capacity (you cannot assign more 0–5 slots than you add or build). (5) imposes the per-facility expansion limit $U_e = \min\{0.2\,\mathrm{cap}_e, 500\}$. (6) activates the baseline cost once expansion reaches the original capacity (big-M with M large enough). (7) and (1) implement the linear PWL expansion cost with decreasing marginal costs, all other costs are fixed or linear. Decision variables include both integer and continuous variables, so the overall model is a MILP.

5.3 Solution Approach

For Problem 1, the optimization problem for each ZIP code is independent of the others; i.e., the budget allocation in one region does not affect the optimal solution in another. Therefore, we optimize each ZIP code independently using Gurobi Optimizer. This approach reduces computational complexity and avoids solver scalability issues.

5.4 Results and Discussion

Main results. The minimal total cost and average cost per ZIP are **$483,486,000** and **$293,734**, respectively. Table 1 shows representative outputs for selected ZIP codes. For each area, the table reports its demand category, minimum total funding, and the main investment actions chosen by the optimizer.

Table 1: Representative optimal solutions for selected ZIP codes. All ZIPs are five-digit and appear in both `population.csv` and `potential_locations.csv`.

ZIP Code	Demand Type	Min. Funding ($)	New Facilities	Expansions
10001	Normal	301,800	1 Large	1 site (+350 slots)
10005	Normal	262,300	2 Large	None

The optimizer prioritizes expansions when existing facilities can be enlarged cost-effectively, as in ZIP 10001, where a single expansion complements a new large center to meet demand at minimal cost. Conversely, when expansion opportunities are limited or absent, as in ZIP 10005, the model relies exclusively on new construction. This behavior reflects the decreasing marginal cost structure of expansions and the fixed setup cost of new facilities, leading to economically balanced and policy-compliant investment plans.

Interpretation. The model tends to favor expansions when existing facilities have available capacity because marginal expansion costs decrease with scale. Where no facilities exist or existing ones are already at their limit, the solution relies on constructing new centers. The baseline cost trigger discourages full doubling of facilities, so expansions often stop just below the threshold cap_e to avoid paying the fixed baseline fee, an effect consistent with the model's intended logic and our expectation. Empirically, these results also align with our broader observations:

- Expansions often stop just below the baseline trigger.

- The 0–5 coverage constraint typically binds before the total-capacity rule, confirming the model's expected behavior.

Conclusion. The implemented MILP efficiently identifies the minimum investment needed for each ZIP code to eliminate child-care deserts under the state coverage rules. By combining fixed-cost construction decisions with scalable expansions, the model provides transparent and data-driven funding estimates for local planning. The approach achieves a clear and interpretable allocation framework that satisfies all requirements and can be readily extended to larger datasets or policy updates.

6 Problem 2 – Realistic Capacity Expansion and Location

In this problem, the decision-making process is designed to determine how to meet child-care capacity requirements at minimum cost, by combining expansion of existing facilities with new construction at candidate sites.

6.1 Problem Fromulation

6.1.1 Decision Variables.

- n_s: Number of new facilities of size s (small, medium, large) to be built in each ZIP code.
- x_e: Number of additional slots added by expanding existing facility e.
- y^{new}: Number of new slots specifically for children aged 0–5 in new facilities.
- y^{exp}: Number of new slots specifically for children aged 0–5 in expanded facilities.
- $y_{p,s}$: Binary variable equal to 1 if a new facility of size s is built at candidate site p; 0 otherwise.

This structure allows the model to flexibly allocate capacity between cost-efficient expansions and new facility construction, ensuring that all coverage and spatial requirements are met.

6.1.2 Constraints.

Building upon the foundation of Problem 1, the constraints additionally incorporate distance constraints between facilities. It guarantee that the final solution is *feasible*, *demand-responsive*, and *spatially valid*.

Total capacity requirement.

$$\text{Existing}_i + \sum_e x_e + \sum_s \text{cap}_s n_s \;\geq\; R_i^{\text{tot}}$$

0–5 slots requirement.

$$\text{Existing}_i^{(0\text{–}5)} + \sum_e y_e^{\text{exp}} + \sum_s y_s^{\text{new}} \;\geq\; R_i^{(0\text{–}5)}$$

Expansion limits.

$$x_e \leq \min\{0.2\,\text{cap}_e,\ 500\}$$

and for age-specific expansions:

$$y_e^{\text{exp}} \leq x_e$$

New facility age capacity.

$$y_{i,s}^{\text{new}} \leq cap_s^{(0-5)} \times n_s$$

Site rules. Each candidate site can host at most one facility:

$$\sum_{s \in S} y_{p,s} \leq 1 \quad \forall p \in P_i$$

Distance constraints.

- If two candidate sites are closer than 0.06 miles, only one can be selected:

$$\sum_s y_{p,s} + \sum_s y_{p',s} \leq 1 \quad \text{if } d(p,p') < 0.06$$

- If a candidate site is too close to an existing facility, it cannot be selected:

$$\sum_s y_{p,s} = 0 \quad \text{if } d(p,e) < 0.06$$

where $d(p, p')$ is computed using the **Haversine distance formula** following [2]:

$$d = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

$$R = 3958.7613 \text{ miles}$$

6.1.3 Objective function.

The optimization seeks to **minimize total cost**, composed of:

Expansion cost.

$$f_e(x) = \begin{cases} (20{,}000 + 200 \cdot cap_e) \cdot \dfrac{x}{cap_e} & \text{for } x \le 0.1\, cap_e \\ (20{,}000 + 400 \cdot cap_e) \cdot \dfrac{x}{cap_e} & \text{for } x \le 0.15\, cap_e \\ (20{,}000 + 1000 \cdot cap_e) \cdot \dfrac{x}{cap_e} & \text{for } x \le 0.20\, cap_e \end{cases}$$

New facility cost.

$$\sum_{s \in S} C^s \cdot n_s$$

where C^s is the fixed cost of building a facility of size s.

Age-specific capacity cost.

$$100 \times \left(\sum_{e \in E_i} y_e^{\text{exp}} + \sum_{s \in S} y_s^{\text{new}} \right)$$

Overall objective.

$$\min \left[\sum_{e \in E_i} f_e(x_e) + \sum_{s \in S} C^s \cdot n_s + 100 \times \left(\sum_{e \in E_i} y_e^{\text{exp}} + \sum_{s \in S} y_s^{\text{new}} \right) \right]$$

6.2 Solution Approach

Similar to Problem 1, the problem was solved using ZIP-level decomposition. Before optimization, all candidate sites were pre-processed: unique site_ids were assigned, latitude and longitude were converted to numerical values, and pairwise distances were computed using the Haversine formula to enforce minimum distance requirements.

During optimization, the model consistently maximized the expansion of existing facilities up to their allowed 20% capacity, as expansion has a lower marginal cost compared to new construction. Once expansion capacity was fully utilized, the model strategically selected new facility sites and sizes to cover remaining demand. The distance constraints ensured that the placement of new facilities was spatially feasible and realistic.

6.3 Results and Discussion

Main results. The optimization resulted in a minimum total cost of: **$451,366,600** (which is slightly lower than the total cost in Problem 1 due to differences in the cost function). This value reflects the optimal mix of expansion and construction decisions across all ZIP codes.

Interpretation. The minimal cost solution of $50.9 million emerged from a hybrid capacity strategy. Expansion accounted for approximately 18–20% additional capacity per ZIP code, which is the maximum allowed by the model. This uniform expansion pattern is a direct reflection of the model's cost minimization logic, not an imposed requirement.

However, expansion alone was insufficient to meet demand. To fill the remaining capacity gap, 20 new facilities were constructed across the selected ZIP codes. ZIP 10009 had the highest number of new facilities (5), reflecting its larger residual demand, while ZIPs with lower gaps required fewer new builds. All chosen sites complied with the minimum distance constraints, ensuring that facilities were well distributed.

This combination of fully utilized expansion and selective construction directly minimized total cost: expansion minimized marginal costs first, while construction allowed full coverage where necessary.

Conclusion of the optimization strategy. The minimal cost solution arises from a two-stage economic mechanism that emerges naturally from the model structure.

First, since expansion is cheaper, the solver expands each facility to the maximum allowed threshold of 20%. This creates a uniform expansion ratio across all ZIP codes and provides a low-cost capacity buffer.

Second, new facilities are built strategically in locations that maximize coverage while minimizing the number of sites needed, given cost and distance constraints. This ensures that the more expensive construction option is used only where strictly necessary.

This strategy mirrors real-world infrastructure planning: leverage existing assets first, then invest in new infrastructure only to cover residual demand. The piecewise-linear cost function effectively pushes the solver to avoid overspending on expansion, while distance rules prevent unrealistic clustering of facilities.

7 Problem 3 – Fairness and Budgeted Expansion

This problem extends the previous budgeting optimization by introducing two key policy objectives:

1. **Spatial fairness:** ensure that the difference in coverage ratio (slots per child) across ZIP codes does not exceed 10%;

2. **Global budget constraint:** limit the total cost of all investments to $100 million.

The goal is to maximize the overall *social welfare*, defined as a weighted average coverage of children aged 0–12, while satisfying fairness and budget constraints.

7.1 Problem Formulation

The optimization is formulated as a mixed-integer linear program (MILP) solved using `Gurobi`. It builds on Problem 1 by incorporating both global fairness and budget limits.

7.1.1 Decision Variables.

- $x_i \geq 0$ — number of expansion slots added to existing facility i;
- $y_j \in \{0, 1\}$ — whether to construct a new facility at candidate site j;
- $z_{j,s} \in \{0, 1\}$ — binary indicator if site j is built with size $s \in \{\text{small}, \text{medium}, \text{large}\}$;
- cov_z — coverage ratio (slots per child) in ZIP code z;
- $\text{cov}_{\min}, \text{cov}_{\max}$ — lower and upper coverage bounds enforcing fairness.

7.1.2 Constraints.

Distance Feasibility. No two newly built facilities can be within 0.06 miles, and new sites cannot overlap existing ones:

$$y_j + y_{j'} \leq 1, \quad \text{if } \text{dist}(j, j') < 0.06 \text{ miles.}$$

Demand Satisfaction. Each ZIP must meet capacity requirements for total and under-5 children:

$$\text{Existing}_z + \sum_{i \in E_z} x_i + \sum_{j \in N_z} \sum_s \text{cap}_s z_{j,s} \geq R_z^{\text{tot}},$$

$$\text{Existing}_z^{(0\text{–}5)} + \sum_{i \in E_z} x_i + \sum_{j \in N_z} \sum_s \text{cap}_s z_{j,s} \geq R_z^{(0\text{–}5)}.$$

High-demand ZIPs (employment rate ≥ 0.6 or income $\leq$ \$60,000) must meet at least 50% coverage; others must meet 33%.

Facility Size Selection. Each new site can have at most one size:

$$\sum_s z_{j,s} = y_j \quad \forall j.$$

Expansion Costs (Piecewise). Expansion costs follow a piecewise-linear structure in the expansion ratio x_i/cap_i, modeled via Gurobi's PWL function:

$$0,\ 0.1,\ 0.15,\ 0.2$$

represent breakpoints, with increasing slope to capture economies of scale.

Fairness Constraint. The difference between the highest and lowest ZIP-level coverage ratios must be within 10%:

$$\text{cov}_{\text{max}} - \text{cov}_{\text{min}} \leq 0.1.$$

Budget Constraint. The total expansion and construction cost cannot exceed $100 million:

$$\text{TotalCost} \leq 100{,}000{,}000.$$

7.1.3 Objective Function

The model maximizes a weighted social welfare function prioritizing younger children:

$$\max \sum_{z \in \mathcal{Z}} \left[\frac{2}{3} \frac{P_z^{(0\text{–}5)}}{P_z^{\text{tot}}} + \frac{1}{3} \frac{P_z^{(5\text{–}12)}}{P_z^{\text{tot}}} \right] \times \text{cov}_z.$$

7.2 Results and Discussion

Main resultss. Under the current configuration, the statewide MILP remains *infeasible* within the time limit. This infeasibility appears to be *structural* rather than numerical: no feasible combination of expansions and new constructions can simultaneously satisfy (i) all ZIP-level coverage targets, (ii) the statewide fairness gap constraint $\text{cov}_{\text{max}} - \text{cov}_{\text{min}} \leq 0.1$, (iii) the 0.06-mile spacing rule for facilities, and (iv) the overall budget ceiling. This outcome is consistent with earlier Problem 1 findings, where the minimum feasible statewide cost already approached $400M even without fairness or spacing restrictions.

Diagnosis. A closer inspection suggests that several interacting constraints drive the infeasibility:

1. *Proximity restrictions.* The requirement that no new facility be located within 0.06 miles of any existing site eliminates most of the 215,000 candidate locations, especially in dense metropolitan ZIPs.
2. *Global fairness constraint.* ZIPs with extremely low baseline capacity but large populations require large proportional increases to meet statewide uniformity, which conflicts with the proximity and budget restrictions.
3. *Budget constraint.* Given that Problem 1 already required roughly $400M to eliminate deserts without fairness constraints, the current cap is insufficient by design.
4. *Expansion limits and cost curvature.* The 20% cap on per-site expansion, combined with the convex piecewise-linear cost function, further restricts flexibility in capacity redistribution.

Policy-oriented relaxations. To reconcile the model's policy objectives with computational feasibility, several theoretically consistent relaxations are conceivable:

1. *Geographic flexibility:* retain the *new–new* separation rule but relax or remove the *existing–new* ban, or substitute a soft penalty in the objective.

2. *Equity formulation:* widen the fairness tolerance (e.g., $\text{cov}_{\max} - \text{cov}_{\min} \leq 0.2$) or embed equity as a penalty term rather than a hard constraint.

3. *Budget phasing:* increase the budget ceiling or adopt a multi-period planning framework (e.g., two-stage expansion) to allow gradual convergence toward coverage equity.

4. *Demand smoothing:* permit partial satisfaction of demand in ZIPs with no feasible expansion or construction sites, to prevent local infeasibility from propagating globally.

Interpretation. The infeasibility does not invalidate the modeling framework; rather, it highlights the tension between policy ambition and physical or financial limits. The results imply that, at the current scale and data granularity, enforcing perfect fairness and tight spatial constraints simultaneously is mathematically inconsistent with realistic budget levels. Relaxing or rebalancing these constraints—particularly by softening fairness or phasing investment—would likely yield tractable formulations without sacrificing interpretability.

Takeaways. The analysis underscores that infeasibility in large-scale equity optimization often carries substantive policy meaning: it reveals where structural limitations, rather than solver performance, prevent equitable resource allocation. Future work should explore multi-objective formulations that treat fairness, accessibility, and cost as trade-offs rather than absolute requirements. Such extensions would transform the current infeasible statewide model into a decision-support tool capable of guiding incremental, data-driven childcare investment planning across New York State.

8 Conclusion

In this project, we developed a series of mixed-integer linear programming (MILP) models to support data-driven decision-making for eliminating child care deserts in New York State. Across three progressively complex problems, our analysis integrated cost minimization, spatial feasibility, and fairness considerations into a unified optimization framework.

- **Problem 1:** We identified cost-efficient investment strategies by optimizing each ZIP code independently. The model revealed that expansions often stop just below the baseline trigger to avoid fixed costs, and that the 0–5 age coverage constraint is generally more restrictive than the total capacity rule.

- **Problem 2:** By adding spatial distance constraints, we found that the optimal statewide strategy follows a two-stage mechanism—fully utilizing low-cost expansions before strategically constructing new facilities where necessary. This ensures spatial realism while minimizing total expenditure.

- **Problem 3:** Incorporating fairness and a global budget constraint exposed structural infeasibility under current policy targets. The findings suggest that simultaneous enforcement of equity, budget, and spatial rules exceeds feasible limits, highlighting the need for phased or relaxed fairness formulations.

Overall, this project demonstrates how optimization modeling can translate policy goals into actionable, cost-effective, and interpretable investment plans. Future extensions could incorporate multi-period budgeting or soft fairness penalties to achieve practical, equitable outcomes across diverse communities.

References

[1] Wayne L Winston. *Operations research: applications and algorithm.* Thomson Learning, Inc., 2004.

[2] Roger W Sinnott. Virtues of the haversine. *Sky and telescope*, 68(2):158, 1984.