

Reducing Food Insecurity in the Greater Boston Area: 6.C57 Final Project

Adriana Ladera^{1,2}, Mit Kotak^{1,2}, and Raye Kimmerer^{1,3}

¹Center for Computational Science and Engineering, MIT

²Research Laboratory of Electronics, MIT

³Computer Science and Artificial Intelligence Laboratory, MIT

1 Introduction

Reducing food insecurity is crucial because it directly impacts individual well-being, community stability, and societal progress. Food insecurity leads to severe health consequences, including malnutrition, weakened immune systems, and chronic diseases. For children, it affects cognitive development and academic performance, perpetuating cycles of poverty. Beyond health, food insecurity contributes to social inequities and economic instability, increasing healthcare costs and reducing productivity. This is especially relevant in the Greater Boston Area, where the city of Boston alone has a food insecurity rate of 43% [1] despite being one of the most highly-educated cities in the U.S. This ironic disparity between food poverty and education highlights an issue of essential needs such as food. By addressing food insecurity, societies can promote equity, improve public health outcomes, and build stronger, more resilient communities where everyone has the opportunity to thrive. This fundamental step not only supports humanitarian goals but also fosters long-term economic growth and stability.

In this work, we implement a robust adaptive facility location model formulation with budget uncertainty sets that aims to reduce the food insecurity rate in the Greater Boston Area while optimizing costs. The model determines optimal facility locations in high-need areas, calculates the most efficient distribution routes from warehouses to facilities, and ensures that essential nutrient requirements are met. Through this framework, the project provides a sustainable, cost-effective, and health-conscious strategy for food distribution that can support food banks, non-profits, and government agencies in maximizing their impact within budget constraints. Beyond Boston, this model can be adapted and scaled for broader applications, potentially improving food distribution systems in other cities or regions facing similar challenges.

2 Methodology

To address this problem, we first create a nominal model formulation, which produces an optimal cost without accounting for uncertainties, followed by a budget formulation that implements robustness on the transportation matrix.

2.1 Nominal Model Formulation

We have m warehouse nodes, n food facility nodes, p different types of food items, and q nutrient requirements to satisfy, indexed by i , j , k , and l , respectively.

Decision Variables

- $y_i \in \{0, 1\}$: if food servicing facility is selected or not
- x_{ijk} : quantity of food item k transported from warehouse i to food servicing facility j

Parameters

- t_{ij} : transportation cost per kilogram from warehouse i to food facility j (\mathbf{t} = a 2D matrix $\in \mathbb{R}^{m \times n}$)
- π_k : price of a food item k per kilogram ($\boldsymbol{\pi}$ = a vector $\in \mathbb{R}^p$)
- D_j : total food demand for the target population surrounding the facility j (\mathbf{D} = a vector $\in \mathbb{R}^n$)
- N_l : minimum nutrient requirement that a person needs to meet for a nutrient l (\mathbf{N} = a vector $\in \mathbb{R}^q$)
- $C = \sum_k^p x_{ijk}$: maximum capacity of food items per warehouse for a food item k (a scalar value)
- M : maximum number of facilities that can be selected due to cost constraints (a scalar value)
- V_{kl} : nutritional value of nutrient l for a food item k (\mathbf{V} = a 2D matrix $\in \mathbb{R}^{p \times q}$)
- S : cost to build and operate the each food facility j (a scalar value)

Objective Function

To maximize the number of people fed while minimizing costs (transportation, food, unmet demand), we can structure the objective function as the following:

$$\min \sum_i \sum_j \sum_k \pi_k x_{ijk} + \sum_i \sum_j \sum_k t_{ij} x_{ijk} + S \sum_i y_i$$

Constraints

Demand is met for a population near facility j for all food items k sent from all of the warehouses i to facility j :

$$\sum_i x_{ijk} \geq D_j, \quad \forall j, k$$

Minimum nutritional requirements l must be met for each food item k :

$$\sum_k V_{kl} \geq N_l, \quad \forall l$$

Each facility j must not exceed the capacity C :

$$\sum_k x_{ijk} \leq C y_i, \quad \forall i, j$$

Only M facilities can be chosen, concentrating on areas with higher food insecurity:

$$\sum_i y_i \leq M$$

And finally, to impose non-negativity and binary constraints:

$$\mathbf{x} \geq 0, \mathbf{y} \in \{0, 1\}$$

At first glance, this model formulation reflects that of Peters *et. al* [2] which optimizes food distribution under the World Food Programme. However, there are some key differences. Firstly, our formulation is not a network flow, therefore it does not preserve integrality (which we later prove by the value of the optimal cost) and does not introduce intermediate nodes. Additionally, we incorporate features reminiscent to that of adaptive food facility locations, in which there can be up to M facilities built, and there is a startup and operating cost S per facility y_j that is built. There are also some additional constraints that limit the amount of food items x_{ijk} that can be at a food facility y_j (or in other words, capacity constraints). Additionally, we consider a robust model formulation via the introduction of uncertainty sets on the transportation cost matrix.

2.2 Robust Modeling of Transportation Costs via Uncertainty Sets

Sometimes, transportation costs may not be the expected nominal value, as factors such as traffic, delivery problems, and other transport issues can affect the final transport cost. To account for this, we can implement an uncertainty set for the transportation matrix \mathbf{t} using global perturbations based on the budget formulation so that \mathbf{t} lives in the uncertainty set \mathcal{U} :

$$\mathcal{U}_i = \left\{ \mathbf{t}_i : \mathbf{t}_i = \bar{\mathbf{t}}_i + \Delta_i^T \mathbf{u}_i, \|\mathbf{u}_i\|_1 \leq \Gamma \right\}$$

where \bar{t}_i is the nominal value for t_i obtained from the above formulation. Taking the dual of this yields the l_∞ norm and the following derivation:

$$\begin{aligned} & \bar{\mathbf{t}}_i \mathbf{x} + \Gamma \|\Delta_i^T \mathbf{x}\|_\infty \\ & \bar{\mathbf{t}}_i \mathbf{x} + \Gamma \max_j |\Delta_{ij}^T \mathbf{x}_j| \\ & \bar{\mathbf{t}}_i \mathbf{x} + \Gamma |\Delta_{ij}^T \mathbf{x}_j|, \forall j = 1 \dots n \\ & \bar{\mathbf{t}}_i \mathbf{x} + \Gamma z_{ij}, \forall j = 1 \dots n \\ & z_{ij} \geq \Delta_{ij}^T x_j, \forall j = 1 \dots n \\ & z_{ij} \geq -\Delta_{ij}^T x_j, \forall j = 1 \dots n \end{aligned}$$

where Δ is defined as the deviation matrix for \mathbf{t} , such that $\Delta_i = \mathbf{t}_i^U - \mathbf{t}_i^L$, the difference between the upper bound and lower bound of \mathbf{t}_i . We expect the range of transportation costs to vary by roughly 30% less than nominal transportation cost or up to 50% more than the nominal transportation cost: $\mathbf{t} \in [0.7\bar{t}, 1.5\bar{t}]$. Therefore, $\Delta_i = 1.5\bar{t}_i - 0.7\bar{t}_i$. The choice of uncertainty set varies based on the model formulation. Here we choose the budget uncertainty set, which can limit the total deviation of costs across all entries to a budget via the Γ parameter (i.e., only a subset of the entries can reach their worst-case deviations simultaneously), and is useful assuming that it is unlikely all unknowns will take the worst values at the same time.

2.3 Data

Given the above model formulation, datasets must be constructed for the parameters: demand (the food insecure population) D_j , price for a food item π_k , nutritional value per food item V_{kl} , the transportation matrix entries, t_{ij} , and the nutritional requirements for a person N_l .

To construct the demand, $D_j \forall j = 1 \dots n$, we refer to the 2024 Greater Boston Food Pantries (GBFB) food map for identifying regions of high food insecurity [1] and the Statistical Atlas for the population density per square mile [3]. In this case, each neighborhood will correspond to a potential facility location j , and we obtain the population density per square mile for each neighborhood (facility) j as δ_j . Taking the number of food insecure households h from the GBFB map [1] and dividing this number by the sum of all of the population densities over j , we obtain $f = 47.69\%$ as the fraction of the population that is food insecure. Therefore, for each j , the resulting population that is food insecure (demand) is equal to f multiplied by the population density per square mile:

$$f = h / \sum_{j=1}^n \delta_j$$

$$D_j = f \delta_j$$

The transportation cost matrix \mathbf{t} is a scaled matrix based on distance matrix \mathbf{d} , where each of the entries are defined the distance between a warehouse i and a food facility j . The x and y coordinates provided for each facility j and warehouse i are provided by Google maps in the form of latitude and longitude. To compute distances accurately, we use the Haversine formula [4] to compute the distances between i and j in miles:

$$d_{ij} = 2R \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_j - \phi_i}{2} \right) + \cos(\phi_i) \cos(\phi_j) \sin^2 \left(\frac{\lambda_j - \lambda_i}{2} \right)} \right)$$

where ϕ is the latitude coordinate and λ is the longitude coordinate, both converted to radians (for example, $\phi_i = i_x \times \frac{\pi}{180}$). $R = 3958.8$, the radius of the Earth in miles. Then, each entry of the transportation matrix t_{ij} is scaled by the \$0.47, the cost per mile of travel. The current price per gallon of gas is \$3.03, and the fuel efficiency of a diesel transport truck is 6.5 miles per gallon:

$$\frac{\$3.03}{\text{gallon}} \times \frac{\text{gallon}}{6.5 \text{ miles}} = \$0.47 \text{ per mile}$$

$$t_{ij} = 0.47 d_{ij}$$

The prices of food items, $\pi_k = [\text{$milk}, \text{$apples}, \text{$beef}\dots]$, are based on the cost of living food prices in Boston for a select amount in kg per food items [5]. Nutritional values of foods V_{kl} are obtained for each food item k and its corresponding amount of the nutrients l from data published by the U.S. Department of Agriculture Agricultural Research Service [6]. The minimum daily nutrient requirements N_l for each nutrient l were obtained from the U.S. Food and Drug Administration [7].

3 Results and Discussion

3.1 Outcomes

Interestingly, the first round of the nominal model formulation described in *Section 3.1* results in a solution that Gurobi determines to be infeasible or unbounded. Testing the model by isolating each constraint reveals that infeasibility/unboundedness comes from the first constraint, for the demand of

the target population around facility D_j , and the second constraint, for nutritional requirements N_l . Still, eliminating these constraints reveals a solution with an optimal cost of 0 and no facilities y_j built (or in other words, the solution vector for \mathbf{y} is a vector of all zeros). This suggested that perhaps the values for D_j and N_l were too stringent given the constructed datasets, and so scaling factors of 0.1 to D_j and 10 to V_{kl} were applied to ensure feasibility and optimality of the model. To target the zero-vector solution of \mathbf{y} and zero-objective value optimal cost, we change the constraint on y_j from a maximum number of M facilities that must be built to a minimum number:

$$\sum_i y_i \geq M$$

With the given modifications, the resulting relationship between M facilities built and corresponding optimal cost are shown in Fig. 1.

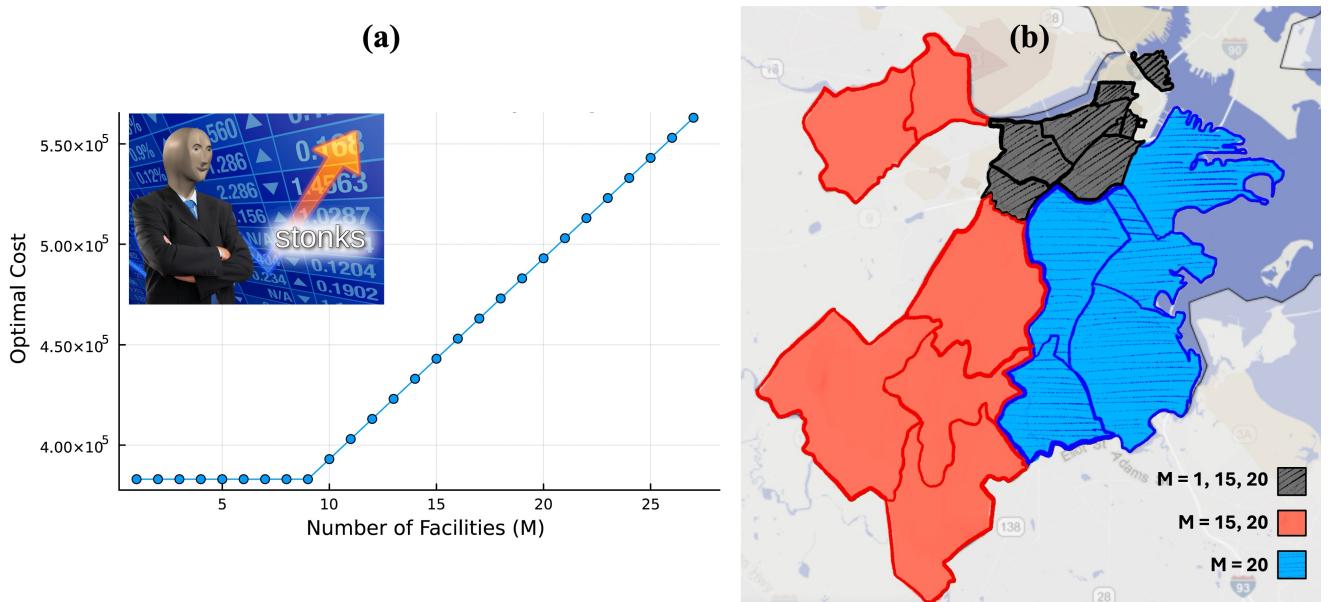


Figure 1: (a) Optimal Cost vs. M , the minimum number of facilities that must be built. (b) Map of regions where a facility y_j is to be built for select values of M . Black diagonally-shaded region shows facilities y_j that are selected when $M = 1, 15, 20$. Red region shows y_j that are only selected when $M = 15, 20$. Blue horizontally-shaded region shows y_j that are only selected when $M = 20$.

Fig. 1(a) reports the optimal cost as a function of the parameter M , the minimum number of facilities that can be built. When $M = 1$, the total number of facilities y_j built is 9, visualized as the diagonally-shaded regions in black in Fig. 1(b). Unsurprisingly, these are the regions that are the most densely-populated (i.e. the North End, Beacon Hill, the Leather District, Chinatown, Back Bay, Bay Village, Fenway-Kenmore, South End, and Mission Hill) and nearest to one another, therefore, the nominal model ensures that the transportation costs between the nearest neighboring warehouse i and these neighboring, highly-populated regions j are minimized. The optimal cost as a function of M exhibits a piecewise linear relationship, with the second linearization occurring at $M > 10$. This suggests that the model prefers not to establish more facilities y_j unless the constraint on the minimum number of M forces it to. This is reflected in the visualization of the regions selected in Fig. 1(b), for $M = 15, 20$, the red regions, and $M = 20$, the horizontally-shaded blue regions, which are less densely-populated and also farther apart from one another in distance. They may also be farther to the nearest neighboring warehouse i than the regions selected when $M = 1$ (black, diagonally shaded), which may explain why

the model opts to select these regions in order to reduce transportation and facility startup and operating costs.

Realistically, this model is best-fit to serve a much smaller radius, where the population that experiences food insecurity is more highly concentrated. This occurs when $M = 1$, and 9 facilities are built in the black diagonally-shaded regions of the North End, Beacon Hill, the Leather District, Chinatown, Back Bay, Bay Village, Fenway-Kenmore, South End, and Mission Hill.

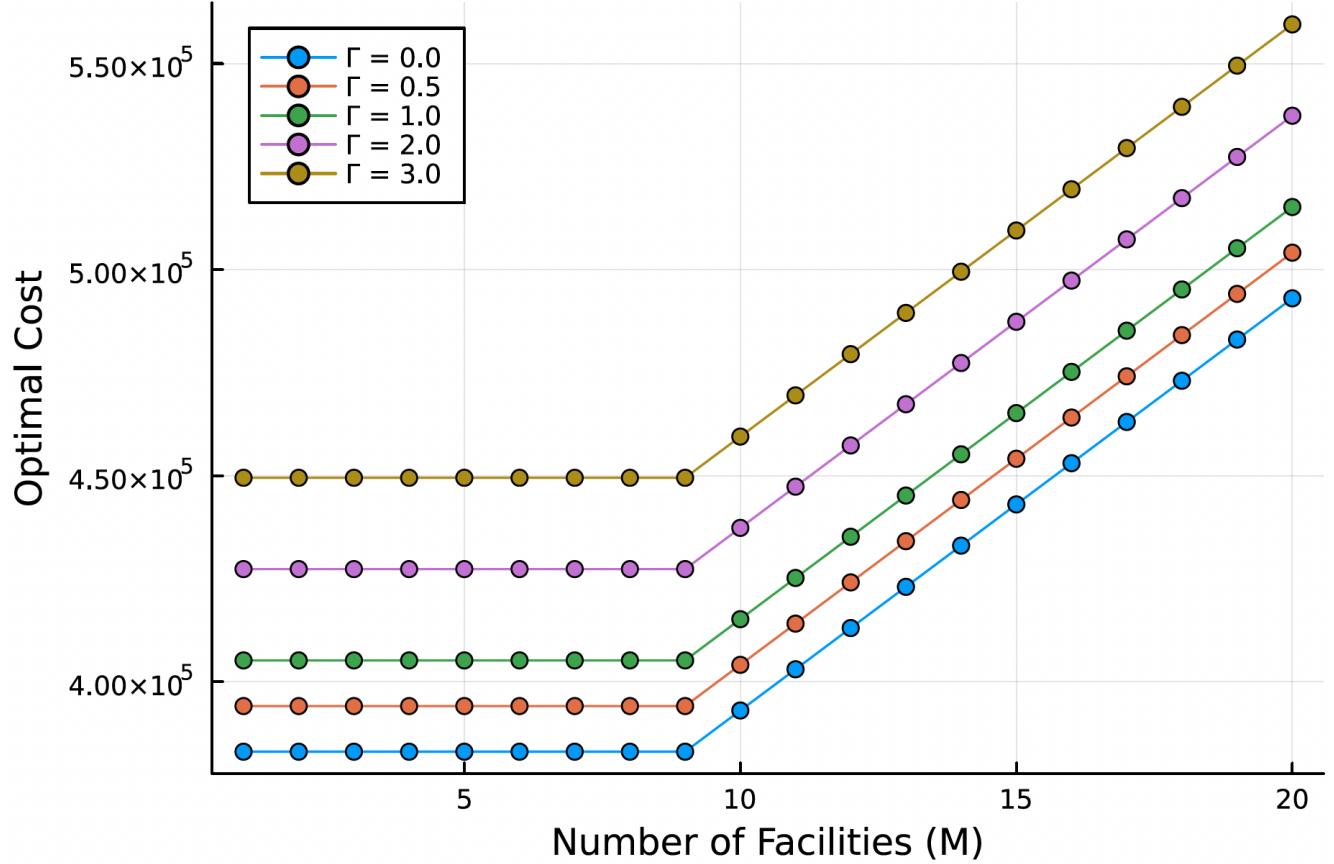


Figure 2: Optimal Cost vs. M , the minimum number of facilities that must be built, for different parameters values of Γ .

Fig. 2 displays the robust budget model formulation results. We introduce the parameter Γ as a way to "tune" the amount of uncertainty that can affect the model, with $\Gamma = 0$ effectively being analogous to no uncertainty in the model (i.e. the nominal model). We see that as we increase the parameter Γ , the optimal cost of the model increases by a constant amount for all values of M . This makes sense, as allowing more uncertainty into the model means that the model will assume worst-case scenarios, which, in terms of our model formulation, should increase optimal cost. Additionally, for all values of Γ , the plots of optimal cost as a function of M facilities are all very similar, featuring piece-wise linear functions that all change at $M \geq 9$, which may be due to the budget uncertainty set formulation, which can capture the expected correlated or limited deviations in transportation costs. Nevertheless, the robust formulation still agrees with the nominal model results in that $M = 1$ provides the lowest optimal cost, in which 9 food serving facilities y_j are built to service the populations experiencing food insecurity in the selected regions. As determined from the original nominal model results, the robust formulation reinforces the

idea that this model is best-fit to serve neighboring, high-density food-insecure populations, illustrated as the black regions on the Fig. 1(b) map.

3.2 Model Limitations, Impact, and Practical Implications

Applying a scaling factor of 0.1 to all values in D_j in order to make the model produce an optimal solution means that is too small given the amount of food insecure people in the Boston area, which means that the data-accurate demand for each region j was not actually met. Additionally, applying a scaling factor of 10 to the nutritional values matrix entries V_{kl} in order to maintain feasibility of the model formulation does not reflect the true nutritional values of the food items from the dataset sources.

By taking f as the fraction of the population that is food insecure and multiplying the population density per square mile of each region j by f , we also assume that the distribution of the population that experiences food insecurity is even across regions (i.e. all regions j have a population of which f experiences food insecurity). Certainly, this is not true—more densely-populated regions such as Fenway-Kenmore and Beacon Hill may be more likely to experience food insecure regions, whereas the opposite may be true for less densely-populated regions with a larger span of land.

Additionally, the model does not compare various uncertainty sets, which may strengthen and verify the results of the robust budget formulation. Specifically, it is quite odd that the robust budget formulation shows very similar optimal costs as a function of M facilities despite having tuned values of Γ . Having additional uncertainty sets such as the budget-box and box formulations could help identify and debug issues that may be causing this result in the robust formulation.

This model does not take into account that different people will have different needs for the amount of nutrients they would require. Additionally, the variety of food items per person is fixed (i.e. everyone gets a package of rice, meat, cheese, etc.), contingent on the fact that the upfront quantity of food items purchased is a good estimate for how many people will be served, and in hopes that there will be no surplus or lack of food items to build a nutrition package.

The expected results from solving this optimization problem should provide an actionable plan to address food insecurity in the Boston area. Specifically, the outcomes should include finding optimal food servicing facility locations, a distribution plan for food items, a breakdown of total costs (transportation costs, food procurement costs for each food item, unmet demand penalties), maximizing the number of people fed, as well as minimizing the unmet demand. In a practical setting this model would aid in efficient resource allocation for food aid programs, improve budget management for food security initiatives, and improve the health and nutritional impacts of food insecure populations in the Boston area. It could also be used in disaster or crisis planning, where rapid and efficient food distribution is needed. More broadly, extrapolating this model could be scalable to larger applications, such as minimizing food insecurity over all of Massachusetts.

3.3 Future Outlook

While this model formulation primarily succeeds in meeting the food needs of densely-populated neighboring regions concentrated in the downtown Boston area, it sets up a framework that could improve on the food security rate of the Greater Boston area, and, while simplified, sets up a framework for future improvements on the current solution that is more representative of realistic scenarios.

The first areas of improvement would be to restructure the choice of food items k such that the nutrients for those food items V_{kl} are able to meet the daily nutrition requirements for a person N_l without using scaling factors to maintain model feasibility, as well as choosing the locations of the warehouses i such that all regions, not just the densely-populated clustered regions are considered. Additionally, we could also enforce soft constraints, wherein fulfilling slightly less of the daily nutritional values N_l for a

person could allow the food needs of many more people to be met. However, this would require more complex ethical conversations on how much less nutrients a person can receive if it means reaching more people who experience food insecurity.

Additionally, we could implement other robust modeling formulations, using box and budget-box formulations to compare the optimal cost results against the original budget formulation and nominal model. This could also help us identify if there was an error in the original budget formulation that was causing different parameters of Γ to produce very similar piece-wise linear plots. In general, we will prioritize robustness over optimality, as a strong robust model formulation will be well-prepared for any worst-case scenarios that uncertainties may bring.

We could also implement two-stage decision modeling using different scenarios that are sampled from a historical collection of our current data. This could be achieved via a deterministic optimization based on expected outcomes, adaptive optimization based on second stage scenarios, or “wait and see” solutions based on realized second-stage outcomes. However, while the “wait and see” model may produce the lowest optimal cost of the three two-stage decision optimizations, this is a more long-term commitment that is necessary to obtain more complete and accurate information on the future dynamics.

Of course, while all of these improvements on the current model formulation would be insightful, some of these in-depth research questions would require a thesis worth of research to fully answer, and sadly, we are but fleeting students in a semester-long optimization class.

4 Conclusions

This work presents a robust adaptive facility location model with budget uncertainty sets, designed to reduce food insecurity in the Greater Boston Area while minimizing costs. The model identifies optimal facility locations in high-need areas, determines efficient distribution routes from warehouses to facilities, and ensures essential nutrient requirements are met. Our findings reveal that modifications such as scaling factors and minor constraint adjustments are necessary to achieve feasibility. Additionally, for both the nominal and robust formulations, the model results suggest that facility location placement is most effective within a smaller service radius, where the population that experiences food insecurity is more concentrated. We conclude by suggesting future directions that could involve two-stage decision modeling or correcting datasets to prevent the need for scaling factors that maintain feasibility. This framework offers a sustainable, cost-effective, and health-focused approach to food distribution, supporting food banks, nonprofits, and government agencies in maximizing their impact within budgetary limits. Furthermore, the model can be adapted and scaled for broader applications, enhancing food distribution systems in other cities or regions facing similar challenges.

Acknowledgment

Shout out to Benjamin Boucher, who graded our project proposal, has consistently given very helpful feedback throughout the project design and model implementation process, and was very patient with all of our questions. Also, special thanks to Prof. Jacquillat for providing thorough and insightful lectures that were helpful for building this project.

References

- [1] E. Schieffelin. Fy 2024 closing the meal gap. *The Greater Boston Food Bank*, February 2024.

- [2] Koen Peters, Sérgio Silva, Rui Gonçalves, Mirjana Kavelj, Hein Fleuren, Dick den Hertog, Ozlem Ergun, and Mallory Freeman. The nutritious supply chain: Optimizing humanitarian food assistance. *INFORMS Journal on Optimization*, 3(2):200–226, 2021.
- [3] Statistical Atlas. Population of boston, massachusetts. *Statistical Atlas*, September 2018.
- [4] Sonia Rode. How to calculate distances between points with the maps javascript api. <https://mapsplatform.google.com/resources/blog/how-calculate-distances-map-maps-javascript-api/>, 2019. Accessed: November 15, 2024.
- [5] Sonia Rode. Food prices in boston, ma, united states. <https://www.numbeo.com/food-prices/in/Boston>. Accessed: November 15, 2024.
- [6] Susan E. Gebhardt and Robin G. Thomas. Nutritive value of foods. *United States Department of Agriculture Agricultural Research Service*, October 2002.
- [7] Food and Drug Administration. Daily value and percent daily value on the nutrition and supplement facts labels.