Right Time, Right Place: Facility Location Optimization

by

Muhammad Bilal Anjum, MIT SCM ’26 Candidate

and

Madeleine Hrycyna, MIT SCM ‘26 Candidate

Submitted on May 8th, 2026, in Partial Fulfillment of the

Requirements for the Degree of Master of Applied Science in Supply Chain Management

at the Massachusetts Institute of Technology

ABSTRACT

This project evaluates a network redesign opportunity for a global retailer operating approximately 1,000 U.S. stores supported by eight Regional Distribution Centers (RDCs). The current push-based allocation model limits flexibility, contributing to excess markdowns, missed sales, and capacity constraints. To improve responsiveness, the analysis proposes introducing Chain Distribution Centers (CDCs) as an intermediate layer to enable delayed allocation and better match inventory to demand. A two-stage approach is used: Geographic Information Systems (GIS)-based screening to identify high-potential locations, followed by optimization modeling to determine the optimal number, placement, and capacity of CDCs across multiple demand planning scenarios. The proposed network redesign is projected to yield markdown savings of approximately 22%, offsetting the investment cost, while network operating efficiencies contribute an estimated 2% in cost savings. These benefits are weighed against a capital and operating investment of approximately $20 million annually in Years 1–5, rising to $60 million annually thereafter. The result is a scalable, data-driven framework for evaluating distribution capacity expansion while improving alignment between inventory allocation and demand growth plans.

Capstone Advisor: Dr. Eva Ponce

Title: Director of the MIT Omnichannel Supply Chain Lab at MIT Center for Transportation & Logistics

Table of Contents

1. Background

Our project sponsor company is a global retailer. Alongside its e-commerce platforms serving customers across the United States and Europe, the company boasts a substantial brick-and-mortar footprint, operating roughly 5,000 stores in nine countries throughout Europe, the Americas, and Australia. Generating annual revenue of over $50 billion USD, the company operates multiple brands targeting different customer segments and categories, consistently offering brand name and designer merchandise at prices 20% to 60% lower than those of traditional full-price retailers. The sponsor company's value proposition is driven by trend-focused merchandise offered at accessible price points and supported by rapid inventory turnover. Frequent assortment updates and short demand cycles encourage repeat visits to their stores, while disciplined inventory management helps limit excess stock and maintain full-price sell-through. To support continued growth under rising demand and uncertainty, the company is focused on reducing costs, improving service responsiveness, and enhancing capacity through operational efficiency and flexibility.

1.1. Motivation

The sponsor company's home fashion division in the United States operates a supply network of eight Regional Distribution Centers (RDCs) serving approximately 1,000 stores across the country, with each RDC supporting between 71 and 185 stores. The company uses a specialized buying model in which products are allocated to specific RDCs at the time of purchase order (PO) placement based on push demand plan forecasts. The lead time between allocation and actual product arrival could be months, depending on supplier lead times and location. When lead times are short (less than three months), allocations generally align with the latest market trends and forecasts. However, longer lead times significantly increase the risk of inefficient allocation due to changes in consumer preferences after initial allocation. These shifts may arise from unpredictable factors such as extreme weather, evolving fashion trends, or other variables.

Due to current system-related limitations, it is not possible to reallocate products to different RDCs if demand plan changes, which is especially problematic given the company's emphasis on high inventory turnover of over 10 turns. This often forces the company to increase markdowns to clear inventory quickly, which leads to a negative financial impact of almost $500 million each year. Conversely, incorrect RDC allocations can also lead to missed sales opportunities and lower levels of customer satisfaction. Although demand plan growth remains strong, the current distribution network, particularly RDC capacity and allocation rigidity, limits the company's ability to scale efficiently.

1.2. Problem Statement

The sponsor company currently is facing challenges in its ability to reallocate goods after they arrive at the RDCs, which negatively affects profitability, inventory, and scalability. To address this issue, the sponsor company seeks a solution that enables just-in-time allocation of merchandise to RDCs and supports overall strategic sales growth.

1.3. Project Goals

The sponsor company hypothesizes that introducing Chain Distribution Centers (CDCs) will increase allocation flexibility and overall capacity, decrease costs, and improve the ability to fulfill demand. By delaying allocation until merchandise reaches the CDC, the company can create a more responsive supply chain that better aligns with dynamic demand planning and enhances customer satisfaction. We believe that the introduction of these CDCs would ensure that the right product is available at the right RDC at the right time and expand network capacity to support sales growth. Our prime objective was to address the following primary questions:

- What is the optimal number of CDCs in this two-tiered distribution system?
- What are the optimal locations of these CDCs in order to balance the cost and service levels?

As a key outcome, we will develop a flexible model that integrates the primary drivers such as demand and transportation to determine the optimal number and location of CDCs, thus minimizing cost, reducing markdowns, and expanding network capacity. To develop the methodology, we needed to address the following secondary questions:

- What approach should be used to initially reduce the number of candidate sites for use in an optimization model?
- Which network configuration would be most effective for minimizing total network cost consisting of transportation, handling and markdowns?
- What is the best size/capacity of the optimal location(s) in terms of small, medium, or large?

2. State of the Practice

Considering we did not have a set of predefined candidate locations for the proposed CDCs, we structured the literature review into two modules. The first focuses on narrowing the universe of potential sites to a manageable set of high-quality candidates. This set then serves as the key input to the second module - an optimization model that minimizes costs and selects the optimal CDC configuration.

2.1. Approaches for Reducing the Number of Candidate Locations

Facility location is a classic subject within supply chain management with a vast body of literature, though much of it focuses on selecting from a pre-identified set of candidate locations. Identifying the right candidates involves both quantitative and qualitative factors, and major approaches include:

- GIS-Based Spatial Screening: Facility location decisions often begin with GIS-based spatial screening, which applies hard, binary geographic and regulatory constraints to eliminate unsuitable areas from consideration and reduce the pool of feasible candidate sites (Bahadori et al., 2022).

- Multi-Criteria Decision Making (MCDM): Facility location decisions carry long-term business implications and require evaluation across multiple, often conflicting, criteria involving uncertainty and subjectivity (Neol et al., 2025). MCDM approaches assign importance weights and rank alternatives, ensuring qualitative factors are not overlooked.

- Mathematical Optimization-Driven Site Reduction: Techniques such as the p-median and Center of Gravity (CoG) models narrow potential locations through cost or distance minimization. While computationally efficient, these methods may not fully capture qualitative or regulatory considerations (Chau & Benson, 2025).

The literature shows that these approaches are typically used in combination to narrow down candidate locations. The most relevant example is Billal et al. (2025), who present a sequential GIS-MCDM methodology that reduces 5,800 candidate locations across Western Canada to a manageable set to be utilized in a subsequent Mixed-Integer Nonlinear Programming (MINLP) optimization. Their multi-stage process uses GIS to eliminate 70% of candidates; applies a Fuzzy Analytic Hierarchy Process (FAHP) to weight 10 preference factors; combines these into a Land Suitability Map (LSM) scored on a 0–10 suitability index; and finalizes 100 candidate sites via ArcGIS Location-Allocation analysis. Rikalovic et al. (2014) applied a principally similar approach for industrial site selection. Bahadori et al. (2022) similarly combined GIS with Analytic Hierarchy Process (AHP) to identify 45 alternative locations for bike-sharing expansion, then ranked them using Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) which selects alternatives that are closest to an ideal solution while farthest away from the least desirable reaffirming the value of integrated GIS-MCDM approaches.

A comprehensive literature review by Battal (2020) found that MCDM techniques, particularly AHP, are the most common approach for facility location problems, while cost, labor, infrastructure, and market access are among the most frequently used decision criteria. By contrast, purely mathematical approaches like the CoG model lack mechanisms to account for qualitative factors and can be overly theoretical. Given these considerations, the integrated GIS-MCDM approach of Billal et al. (2025) was considered most applicable to our project.

2.2. Comparison of Most Appropriate Optimization Framework for a Capacitated Facility Location Problem (CFLP)

Daskin et al. (2005) explained a classical optimization model - Fixed Capacity Facility Location (FCFL). The objective of this model is to determine facility locations and shipment patterns between facilities and customers that minimize the combined facility opening and transportation costs while meeting demand. In the classical FCFL model, capacities are normally introduced as constraints. However, capacity could also be introduced as a decision criterion. FCFLs have historically been solved using multiple approaches depending upon the objective, available data, computational power and time frame. The most common approaches include Deterministic Mathematical Programming, Stochastic Programming, and Robust Optimization. Deterministic Mathematical Programming approaches like Mixed-Integer Linear Programming (MILP) are used to find optimal solutions under specific constraints with decision variables that can be both continuous and discrete. Deterministic methods assume that all parameters and coefficients are fixed and known (Li et al., 2024). Stochastic Programming models explicitly handle uncertainty when the probability distribution of uncertain parameters (like demand or supply) are known (Snoussi et al., 2021) and are highly complex. The goal is typically to maximize the expected value of the objective function across multiple scenarios (Düzgün & Thiele, 2012). However, optimization efforts through stochastic programming suffer from two limitations. First, the determination of the exact distribution of the data (and thus, the enumeration of scenarios that capture this distribution) is rarely satisfied in practice. Second, the size of the optimization model increases considerably with the rise in the number of scenarios also known as the “curse of dimensionality” (Snoussi et al., 2021).

When the volume of scenarios is overwhelming, or when the probability distribution is simply unknown, switching methodologies can bypass the issue entirely. Robust Optimization (RO) eliminates the need to enumerate scenarios or estimate probability distributions. Instead, it defines a deterministic uncertainty set (usually intervals) and finds a solution that optimizes the worst-case scenario within that set (Snoussi et al., 2021). However, a robust solution is not necessarily optimal for the nominal objective function. The robust model sacrifices the potential gains from scenarios better than the worst-case in favor of feasibility and robustness (Snoussi et al., 2021). Considering the advantages and limitations of Deterministic Mathematical Modeling, Stochastic Programming, and Robust Optimization, in the context of our project, Deterministic Mathematical Modeling followed by Robust Optimization across multiple scenarios is a feasible optimization framework since we were provided with multiple demand scenarios ramping from 5 to 15 years in future.

3. Methodology

This section outlines the framework employed to evaluate the strategic integration of CDCs into the existing network. The methodology proceeded in three stages. First, a discovery stage established a comprehensive understanding of the current network state and anticipated future demand conditions. Second, a GIS-based site selection process identified and ranked candidate CDC locations through sequential exclusion, multi-criteria preference scoring, and suitability analysis. Third, a deterministic MILP model was constructed and stress-tested across multiple demand scenarios to identify the network configuration that minimized cost and maintained competitiveness. To effectively identify the optimal number, location, and size of the CDCs we will divide the project into stages, as outlined in Figure 1.

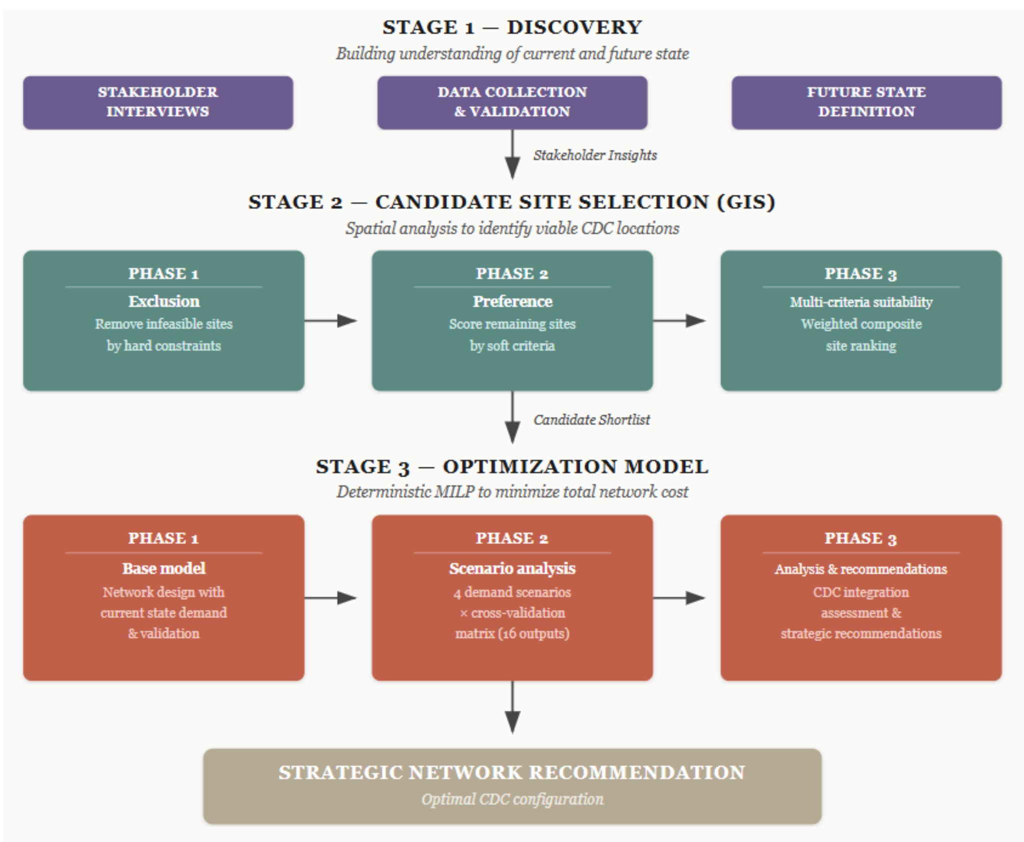

Figure 3: Capstone Methodology

3.1. Stage 1 - Discovery

To build a solid understanding of the business, we conducted multiple virtual interviews to clarify ambiguities regarding the company's business model and rules. As a result, Figure 1 details the current state of the network nodes and flow. Figure 2 drafts the proposed future-state network that includes CDC(s).

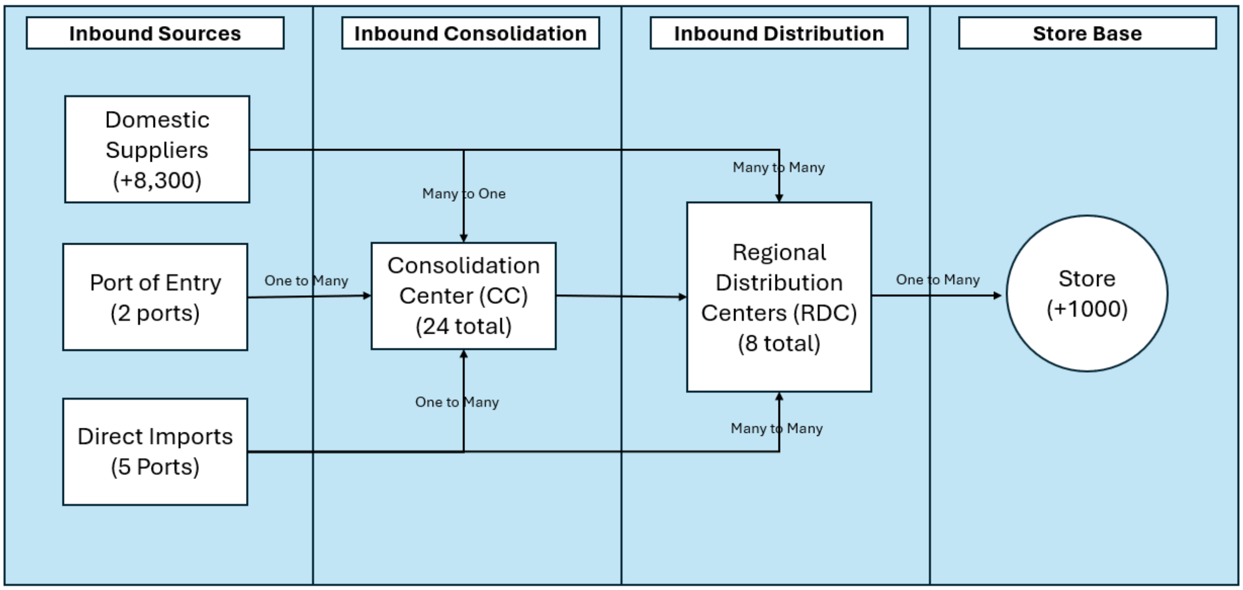

Figure 2: Current Network Flow

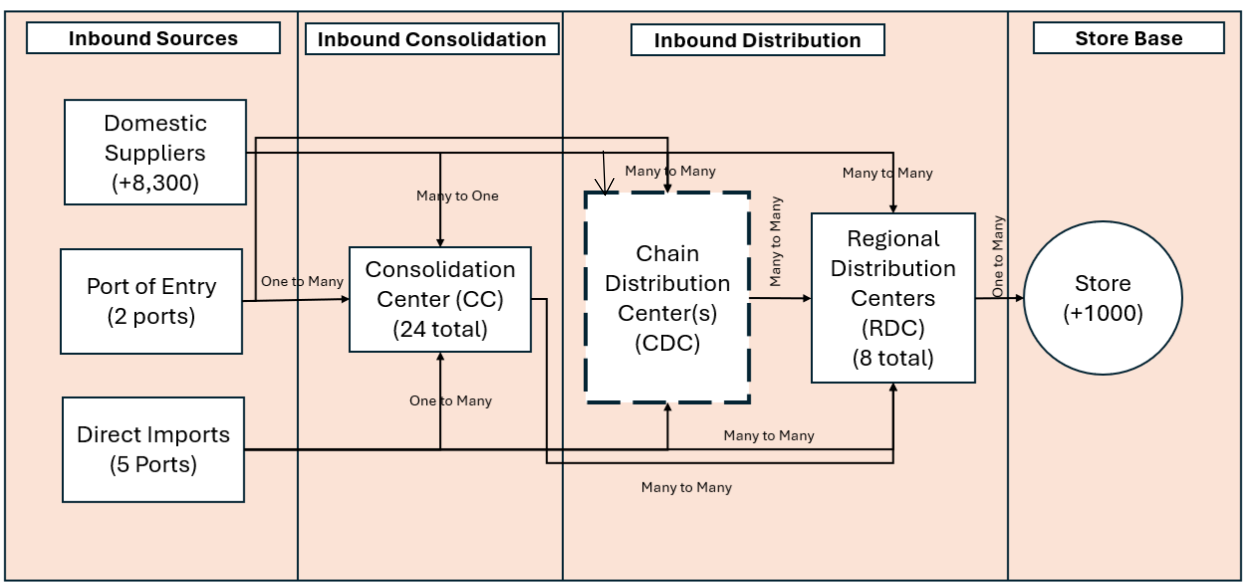

Figure 3: Future State of Network Flow

3.2. Stage 2 and 3: Optimization

The methodological framework of this paper incorporates GIS and MILP tools to model, analyze, and optimize spatially dependent operational decisions for CDC placement.

3.2.1. Methodology Stage 2: Reducing Candidate Locations

The first critical preliminary step involved reducing the solution space to a handful of high-quality candidate locations to serve as inputs for the optimization framework. We followed the process outlined by Billal et al. (2025) as explained in section 2.1. We performed an exclusion, preference, and suitability analysis using ArcGIS Pro, a professional desktop GIS application. Based on information that was available at the time of our project, we were not concerned with identifying the exact coordinates where the potential CDCs could be located. Instead, our core objective was to identify the potential regions that would be the most optimal contenders for a CDC location. Therefore, we chose a resolution level of ~1.86 x 1.86 miles as a cell size in our GIS analysis. Each cell would effectively cover an area of ~3.5 square miles.

3.2.1.1. Stage 2, Step 1: Exclusion Analysis

The preliminary stage of this spatial analysis involved establishing four hard constraints to eliminate regions that are unsuitable for CDC development. To ensure operational relevance we used criteria that were identified and validated with the sponsor company subject matter experts (SMEs) in site selection. See Table 1.

Factor	Description
Land Surface Gradient	Certain land masses can be unsuitable for warehouses due to their land inclination or slope. We used criteria of excluding any cell having a slope greater than 7%, as commonly used in such land suitability analysis.
Distance to Airports	To maintain a safe distance from the airports, we established ~2-mile region of buffer around all national and international airports within the U.S.
Flood Risk	Considering the risk of flooding while identifying the suitable regions is a fundamental aspect (Billal et al, 2025). We used historical flood data as well as normal flood channels to identify the region for exclusion.
Federal Lands	Since any potential CDC can only be located on private land, we excluded all federal lands.

Table 1: Exclusion Criteria

Final exclusion analysis results can be found in Figure 4. All areas grayed out (amounting to ~27% of the area under analysis) were excluded and seen as unfit for a CDC location based solely on the defined exclusion criteria, while areas in red with a score of 1 were identified as feasible locations for a CDC.

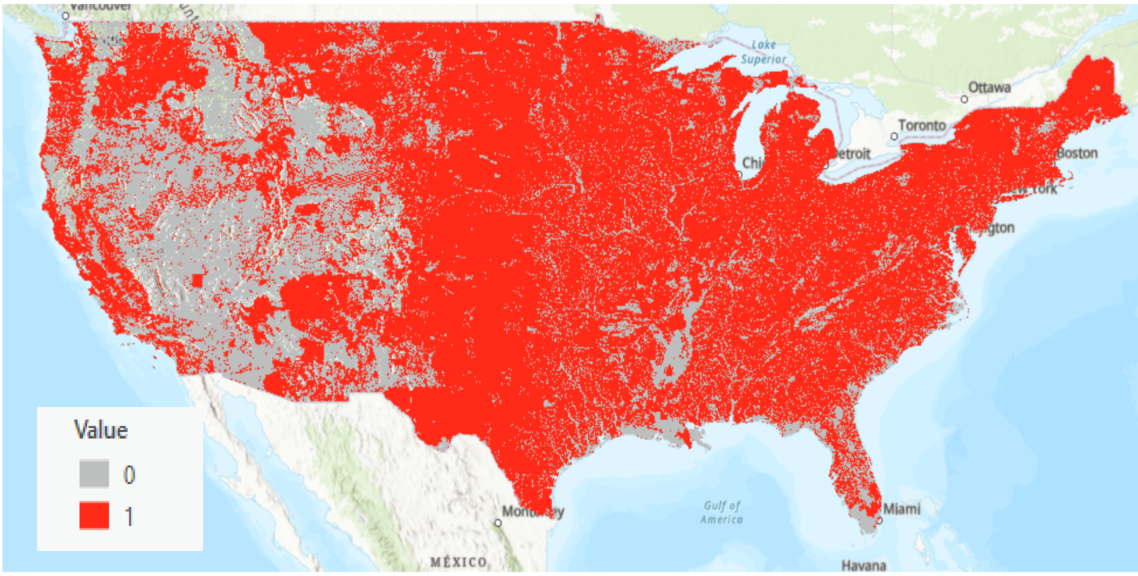

Figure 4: Map of Feasible Land Based on Exclusion Analysis

3.2.1.2. Stage 2, Step 2: Preference Analysis

Following the removal of unsuitable areas, we then considered the qualitative factors necessary for any potential location. These factors were identified through interviews with the sponsor company SMEs and include:

- **Proximity to Highways and Railroads**: The CDCs will be receiving shipments from either ports or domestic suppliers (see Figure 2). It is essential that these centers are well connected to major highways and railroads for optimal travel time. To meet this criterion, we mapped all the major interstate highways in the United States. and assigned a progressively decreasing score for regions farther away from highways or railroads.
- **Labor Density**: The success of a CDC depends largely on meeting the high labor demands of pick-and-pack tasks, which require more manual intervention than typical warehousing. Consequently, our analysis weighed labor availability heavily. We identified major U.S. cities with populations over 50,000 and assigned a suitability score (1–10) based on the density of the material movement occupational class. This ensured our network design accounts for the specific human capital necessary for complex order fulfillment.
- **Ease of Access of Labor**: Presence of labor is not the only relevant human resource consideration; the operational viability of a CDC also depends on the convenience of that labor force commute. To ensure the facility remains an attractive and reachable employer for the local workforce, we prioritized areas within a reasonable travel threshold. Utilizing the network analysis tool in ArcGIS Pro, we delineated service zones based on a 30-minute drive-time radius from major city centers. Locations within this travel-time contour received a maximum score of 10, while scores decreased for areas that had longer commute times, reflecting the diminished likelihood of labor retention and recruitment.
- **Taxation Landscape**: Property and other commercial taxes can become a considerable portion of a distribution center’s operational cost. To minimize this cost, we used industrial property tax rates at the state level to prioritize areas with the lowest taxes.

- **Utility/Power Rates:** U.S. commercial power rates vary significantly from state to state. The operational cost of the CDCs will be greatly impacted by the utility rates of the region in which they are located. Therefore, to have a preferred operational cost, we assigned higher scores to states with lower commercial power rates.

Once these individual layers were created via the scoring rationale described above, we combined these preference factors by assigning weights corresponding to their importance as outlined in Table 2. These weights were determined through interviews with SMEs and executives from the sponsor company.

Factor	**Weights**
Labor Density	30
Labor Access	25
Distance from Main Highways	20
Distance from Railroads	10
Taxation Score	5
Power Score	10

Table 2: Criteria for Preference Analysis

The individual scoring layers were aggregated to produce a cumulative preference heat map (Figure 5). This visualization provides a clear geographic distribution of site suitability across the United States, highlighting hot spots where the combined scores (maximum 1000) for labor availability, infrastructure access and other enlisted factors are highest.

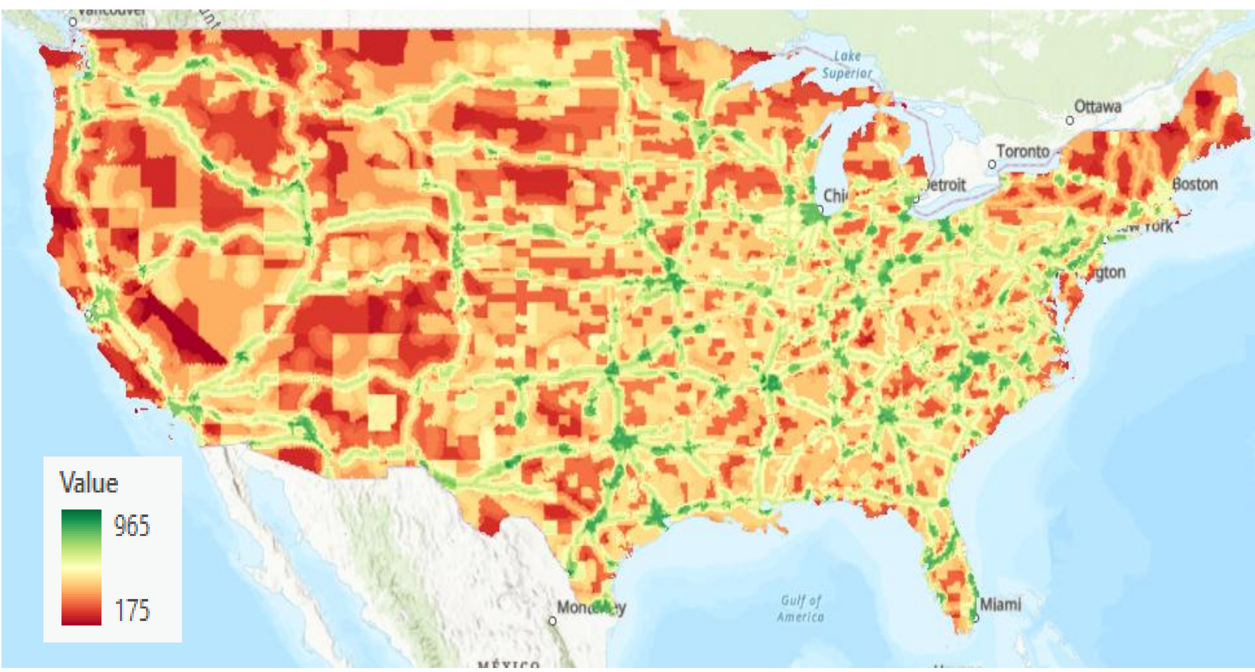

Figure 5: Results of Preference Analysis

3.2.1.3. Stage 2, Step 3: Suitability Analysis

In the final step of our GIS analysis, we combined the results of our exclusion and preference analysis to generate suitability index for identifying the most appropriate regions for a CDC. Utilizing ArcGIS, each feasible cell of ~3.5 miles size received a score. Each score represented a combination of the preference factors outlined in section 3.2.2.

$$Suitability\ Index = Exclusion\ Score\ x\ Preference\ Score$$

A binary exclusion scoring approach was applied, where unsuitable cells were assigned a value of 0 and suitable cells a value of 1. To transition from a granular site-level analysis to a broader strategic framework, we aggregated the suitability scores by state. By calculating the mean suitability index for all grid cells within each state boundary, we identified the most viable states for CDC placement, as illustrated in the map in Figure 8. In this visualization, a color ramp is utilized where darker red shading indicates a higher concentration of favorable conditions. To further refine these findings into actionable territories, we segmented the analysis into five key operational regions (Texas, Midwest, Northeast, Southeast and West). This regional breakdown of scores in Figure 7 should be interpreted along the same lines as in Figure 6. The results provide a definitive geographic guide for our recommended CDC locations and enable regional selection to be used for the optimization model, isolating the specific hubs within each zone that outperformed the average.

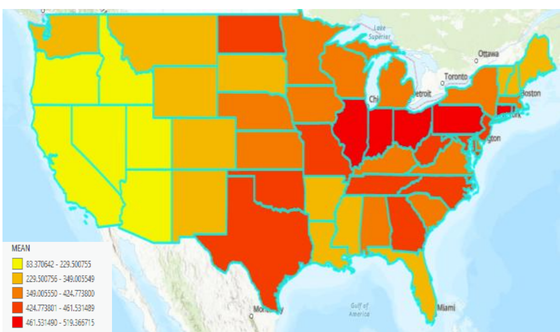

Figure 6: State Preference Analysis Results

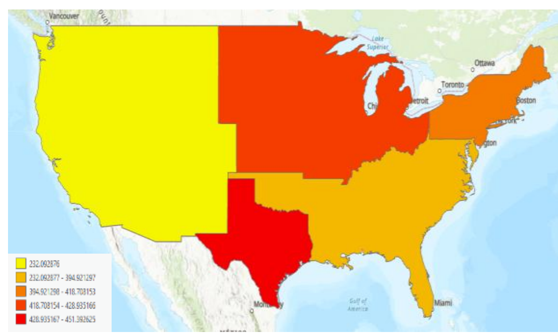

Figure 7: Regional Preference Analysis

In order to identify actual CDC locations for feeding into the optimization model, we identified the top two cities with maximum suitability score across all the five regions and considered them as the potential CDC locations as agreed on by executives at the sponsor company. Details of these cities are covered in Figure 8.

Region	City	State	Site Suitability Score
Midwest	Chicago	IL	675
Midwest	Akron	OH	625
Northeast	Baltimore	MD	578
Northeast	Pitts burg	PA	574
Southeast	Highpoint	NC	672
Southeast	Atlanta	GA	660
Texas	McAllen	TX	665
Texas	Odessa	TX	634
West	Aurora	CO	600
West	Spokane	WA	574

Figure 8: Top Candidate Sites

3.2.2. Methodology Stage 2: Optimization Model

Following the selection of candidate locations, we developed a deterministic optimization model designed for rigorous scenario assessment. This methodology represents an industry-standard approach for evaluating complex facility location problems. The model was constructed in Python using the Gurobi Optimizer to solve a Mixed Integer Linear Programing (MILP) problem. We defined a cost-minimization objective function that accounts for fixed facility cost, variable handling costs, transportation across network arcs, and markdown costs. The model optimizes facility activation and product flow while adhering to the following key constraints at each node: facility capacity limits, total demand satisfaction, and flow conservation. This quantitative model provides a tractable yet comprehensive representation of the distribution network, identifying cost-efficient CDC configurations that remain operationally feasible under varying conditions.

3.2.2.1. Optimization Model Formulation

In this section, we will be covering the details of the network optimization model we created. This is a multi-echelon supply network consisting of ports, suppliers, consolidation centers, CDCs and RDCs as its nodes. Within this network, the RDCs are considered the "customer" nodes having a definitive demand fulfillment plan at a category level per week. In this model, products belonging to different categories are considered unique, having their own distinct weight and volumetric qualities. This demand plan is to be fulfilled by the supply available at numerous ports and supplier nodes.

Sets and Indices:

To provide a better understanding of the model, the sets and indices developed are represented in Tables 3 and 4.

Variable	Definition
S	Set of Suppliers and Ports $S^{port} U\ S^{dom}$
P	Set of Processing Facilities including CCs & CDCs. $P^{con} U\ P^{cdc}$

C	Set of Customer Locations i.e. RDCs
K	Set of Product Categories
A	Set of all Potential Arcs $(i,j) \epsilon A$
R	Set of CDC Configurations

Table 3: Sets Used in Model

Variable	Definition
i	Origin Node $i \in S \cup P$
j	Destination Node $j \in P \cup C$
l	Processing Facility Node $l \in P$
r	Facility Size $r \in \mathrm{R}$
k	Product Category $k \in \mathrm{K}$

Table 4: Indices Used in Model

Decision Variables:

Before diving into the individual cost elements and their corresponding components, the decision variables are detailed in Table 5. These are the decisions the model will make to optimize cost.

Variable	Definition
x_{ijk}	Continuous variable quantifying the flow of products belonging to category $k \epsilon K$ on arc $(i,j) \epsilon A$ [Units]
y_{lr}	Binary variable, equal to 1 if CDC facility l is utilized with Capacity $r \epsilon R$

Table 5: Decision Variables in the Model

The continuous decision variables represent the product flow along these arc types: ports/suppliers to CC, CDC, and RDC, and inter-facility flows from CC and CDC to RDC.

Parameters:

The parameters of the model in Table 4 explain the fixed data points which have been used in the optimization. These include the supply data at the supply nodes (which are ports and domestic suppliers) demand plan data at the customer locations (i.e. RDCs); the fixed cost of opening a CDC of a specific capacity; the handling cost of product at each processing node (i.e. CCs and CDCs); the transportation cost along each arc; and the markdown savings.

Variable	Definition
f_{lr}	Fixed Cost of Opening a CDC $l \epsilon P^{cdc}$ With Capacity $r \epsilon R$
s_{ik}	Available Supply of Product Category $k \epsilon K$ on Supply Nodes $i \epsilon S$
γ_{jk}	Demand Fulfillment plan of Product Category $k \epsilon K$ at the RDCs $j \epsilon C$

q_{ik}	Unit Handling Cost of Product $k \epsilon K$ at Node $i \epsilon P$
c_{ijk}	Unit Cost of Transporting Product $k \epsilon K$ on Arc $(i,j) \epsilon A$
γ_{lr}	Markdown Savings due to Introduction of a CDC (25% cost savings)

Table 1: Parameters Used in Model

Objective Function:

The MILP model is formulated to minimize the total cost across the distribution network, expressed as the sum of all cost components using the variables defined in Table 4. The formulation consists of an objective function, which defines the cost expression to be minimized, and a set of constraints that enforce the operational and logical requirements of the system. Together, these components ensure that the model identifies a feasible and cost-optimal network configuration, as presented below.

$$\begin{aligned} \min Z = & \sum_{r \in R} \sum_{l \in P^{cdc}} f_{lr} \cdot y_{lr} + \sum_{k \in K} \sum_{(i,j) \in A} q_{ik} \cdot x_{ijk} + \sum_{k \in K} \sum_{(i,j) \in A} c_{ijk} \cdot x_{ijk} \\ & + \sum_{r \in R} \sum_{l \in P^{cdc}} x_{ljkr} \cdot \gamma_{lr} . MarkdownRate * Price \\ & + \sum_{r \in R} \sum_{l \in P^{con}} x_{ljkr} \cdot 1 . MarkdownRate * Price \end{aligned}$$

Components of Objective Function:

As outlined previously, the four key elements of the objective function are a sum of all products through each node, presented below in more detail:

$\sum_{r \in R} \sum_{l \in P^{cdc}} f_{lr} \cdot y_{lr}$ = Facility Opening Cost

$\sum_{k \in K} \sum_{(i,j) \in A} q_{ik} \cdot x_{ijk}$ = Total Handling Cost

$\sum_{k \in K} \sum_{(i,j) \in A} c_{ijk} \cdot x_{ijk}$ = Total Transportation Cost

$\sum_{r \in R} \sum_{l \in P^{cdc}} x_{ljkr} \cdot \gamma_{lr} . MarkdownRate * Price + \sum_{r \in R} \sum_{l \in P^{con}} x_{ljkr} \cdot 1 . MarkdownRate * Price$
= Markdown Cost

Supply Constraint:

The sum of the product flow coming out of any supply node cannot be more than the available supply for a given product at that node.

$$\sum_{j \in P \cup C} x_{ijk} \leq s_{ik} \quad \forall\, i \in S,\ \forall k \in K$$

Demand Fulfillment Constraint:

The sum of the product flow going to each of the customer nodes should be at least fulfilling the demand at the corresponding node.

$$\sum_{i \in S \cup P} x_{ijk} \geq \gamma_{jk} \quad \forall\, j \in C,\ \forall k \in K$$

Flow Conservation at CC and CDC Constraint:

The sum of the product flowing in and out of a CC and CDC should be equal to conserve the flow.

$$\sum_{i \in S} x_{ilk} - \sum_{j \in C} x_{ljk} = 0 \quad \forall\, l \in P^{con}, \forall\, k \in K$$

$$\sum_{\mathrm{i} \in \mathrm{S}} \mathrm{x}_{ilk} - \sum_{\mathrm{j} \in \mathrm{C}} \mathrm{x}_{ljk} = 0 \quad \forall\, \mathrm{l} \in \mathrm{P}^{\mathrm{cdc}}, \forall\, \mathrm{k} \in \mathrm{K}$$

Direct Flow Constraint:

The sum of products flowing directly from ports and suppliers to RDCs, bypassing CCs and CDCs, is equal to or less than 20% in line with historical trends.

$$\sum_{k \in K} x_{ijk} \leq 0.20 * x_{ij} \quad \forall\, i \in S,\ \forall j \in C$$

Capacity Constraints:

The sum of products flowing into CCs and CDCs is less than or equal to their capacity.

$$\sum_{\mathrm{i} \in S} \sum_{k \in K} x_{ijk} \leq CAP_j^{CC} \quad \forall\, j \in P^{cc}$$

$$\sum_{\mathrm{i} \in S} \sum_{k \in K} x_{ijk} \leq \sum_{r \in R} CAP_r^{CDC} * y_{lr} \quad \forall\, j \in P^{cdc}$$

Non-Negativity Constraints:

In addition to the binary restrictions on the CDC opening variables, all continuous flow variables are bound by zero, ensuring that product flow remains physically interpretable.

$$x_{ijk} \geq 0 \qquad \forall (i,j) \in A,\ \forall k \in K$$

$$\mathrm{y}_{\mathrm{lr}} = \{0,1\} \qquad \forall\, \mathrm{l} \in \mathrm{P}^{\mathrm{cdc}}$$

3.2.2.2. *Optimization Scenario Modeling and Analysis Approach*

Following base model validation, the framework is extended to assess the strategic integration of CDCs under a structured scenario analysis. Given the supply chain's complexity and demand plan volatility, four scenarios are stress-tested against demand fluctuations and infrastructure shifts, aligned with company growth targets over 5-, 8-, and 12-year horizons. From baseline current state, 30%, 50%, and 80% growth all with additional RDC. Optimal results per scenario are outlined in table 5.

Cost Component	Baseline - No CDC	Baseline - With CDC	% Savings	Sc1. +30% Growth	Sc2. +50% Growth	Sc3. +80% Growth
Transportation	$4.9	$5.0	-3%	$9.5	$11.4	$13.9
CDC Fixed Cost	$0.0	$0.8		$1.3	$1.3	$1.7
Net Handling	$11.1	$10.8	3%	$14.3	$16.6	$19.9
CC Handling	*$1.5*	*$1.1*		*$1.3*	*$1.6*	*$1.8*
CDC Handling	*$0.0*	*$1.4*		*$2.1*	*$2.0*	*$2.7*
RDC Handling	*$9.6*	*$8.3*		*$10.9*	*$13.0*	*$15.4*
Markdown	$4.8	$3.8	22%	$4.6	$5.6	$6.5
Total Cost	**$20.8**	**$20.4**	**2%**	**$29.7**	**$34.9**	**$42.0**

Table 5: Scenario Optimal Results

To ensure the robustness of our recommendations, we will utilize a cross-validation matrix. After deriving the optimal solution for each individual scenario, that specific configuration is fixed and re-run against the parameters of the other three scenarios. This process generates a 4x4 matrix, totaling 16 distinct outputs. This comparative evaluation allows us to identify which network design maintains the highest operational efficiency and lowest cost penalties across a spectrum of potential future scenarios.

4. *Results and Discussion*

This section outlines the final geographic and economic results derived from the integrated ArcGIS and Gurobi optimization MILP modeling process. We discuss the comparative performance of each solution across four distinct demand plan and infrastructure scenarios. These results serve as the evidentiary basis for the final network design, balancing cost-efficiency with the necessary flexibility to meet future growth.

4.1. *Overall Results*

The integrated modeling process identified an optimal CDC network configuration minimizing total weekly transportation, handling, and markdown cost while achieving 100% demand fulfillment across all four scenarios. Benchmarked against a no-CDC baseline of $20.8 million per week, introducing two small CDCs in Baltimore, MD and Odessa, TX, reduced total weekly cost by $371,000 (1.8%), yielding estimated annual savings of $19.3 million over the first five years. These savings were driven primarily by reductions in RDC handling and markdown costs, which more than offset the $841,000 weekly CDC fixed rental cost. As the demand plan scaled across Scenarios 1 through 3, the model expanded the CDC network from two to four facilities while maintaining 100% fulfillment. At 30% volume growth, a third CDC was added in Atlanta, GA to address increased Southeast demand. At 50%, Aurora, CO replaced Odessa, TX as demand shifted toward the Midwest corridor. At 80% growth, a fourth CDC activated in Chicago, IL, bringing total weekly cost to $42.0 million. Baltimore, MD and Atlanta, GA emerged as the most strategically robust locations, appearing in every CDC scenario. A cross-validation cost matrix, seen in Figure 9 shows the final rankings and confirmed that the solutions with three-CDCs performed best across multiple scenarios, underscoring the financial resilience of the core network locations across varying demand and infrastructure conditions. To interpret the matrix in Figure 9, each CDC configuration was ranked from 1 (Most Optimal) to 4 (Least Efficient) for every demand scenario column. The results demonstrate that:

- Scenario 2 Configuration: Performed optimally (Rank 1 or 2) across nearly all cross-validated scenarios.

- Scenario 1 Configuration: Followed closely as the second-most resilient option across different demand scenarios.

#	CDC locations & sizes	Demand scenario applied: Baseline	Scenario 1 (+30% vol.)	Scenario 2 (+50% vol.)	Scenario 3 (+80% vol.)	Average rank
1	Base CDCs: Baltimore (S) + Odessa (S)	**1**	3	4	4	3.0
2	Sc1 CDCs: Baltimore (S) + Atlanta (S) + Odessa (S)	3	**1**	2	3	2.3
3	Sc2 CDCs: Baltimore (S) + Atlanta (S) + Aurora (S)	2	2	**1**	2	**1.8**
4	Sc3 CDCs: Baltimore (S) + Atlanta (S) + Odessa (S) + Chicago (S)	4	4	3	**1**	3.0

Figure 9: Cross Validation Matrix for Scenario Evaluation

4.2. Implications

The results of this analysis carry several strategic implications for the organization's network design decisions. First, site selection extends beyond geographic optimization. The recommended CDC locations align with regions that offer favorable labor market conditions as discussed in section 3.2.1, which directly affect operational throughput and long-term cost stability. Second, the multi-scenario analysis demonstrates the value of cross examination in site selection to ensure robust analysis over a single scenario-specific optimization. The recommended configuration delivers consistent performance across all demand fulfillment plans and infrastructure conditions. Third, the introduction of CDCs into the existing network creates capacity to fulfill the growth in volume by redistributing goods across the network in ways that reduce RDC pressure and position the organization to absorb demand growth without proportional cost increases. Finally, by identifying the conditions under which this location becomes economically justified, the model equips leadership with a concrete, scenario-tested basis for evaluating the impact of adding CDCs to the network.

4.3. Limitations

While this model provides a robust foundation for network design decision-making, several limitations warrant consideration. First, as a deterministic model, it optimizes against fixed inputs rather than capturing the variability inherent in real-world conditions, particularly meaningful given the high seasonality of the retail industry, where demand plan fluctuations can significantly affect the performance of any single optimized configuration. Second, the analysis excludes certain cost factors, namely tax incentives, zoning considerations, and variability in state and local labor markets, that can meaningfully influence facility operating costs but are difficult to standardize across candidate locations. Additionally, the model assumes a small CDC size configuration (430,000 cartons) across all scenarios; evaluating medium- and large- capacity configurations may yield further cost efficiencies, particularly as demand approaches the 50–80% growth range where all facilities operate at full utilization. Finally, the model does not account for delivery to stores from RDCs, which represent a growing share of total fulfillment expenses. Future iterations should incorporate stochastic demand modeling to better reflect seasonal variability, expand the cost parameter set to include location-specific fiscal and regulatory factors, and integrate last-mile cost estimates to produce a more complete picture of total network cost.

5. Recommendations

Integrating the results of the GIS suitability index and the MILP optimization, we present a definitive network design for the proposed CDCs. These recommendations prioritize locations that demonstrate not only the lowest total network cost but also the highest resilience against volatile demand surges. By aligning facility placement with proven labor and infrastructure details, the following strategy provides a scalable roadmap for the sponsor company’s expansion.

5.1. Management Recommendations

Based on the results of the optimization analysis, three primary recommendations emerge from this capstone project. First, the introduction of CDCs into the network is financially justified even at current demand plan levels. The two-CDC configuration in Baltimore, MD and Odessa, TX generates an estimated $19.3 million in annual savings over the first five years, increasing to $63 million per year thereafter, with savings derived from reduced handling and markdown costs that more than offset fixed facility expenses. Leadership should prioritize a decision on CDC adoption in the near term, as the cost benefits will materialize regardless of whether demand growth is realized. Second, Baltimore, MD, and Atlanta, GA, should be treated as the foundational CDC locations for any phased implementation strategy. These sites demonstrated consistent selection across all demand scenarios, reflecting their strategic positioning relative to high-density RDC corridors in the Northeast and Southeast. Third, for the final CDC location, the cross-validation cost matrix supports Aurora, CO as the recommended site, given its stronger performance at elevated demand levels. However, it is worth noting that Odessa, TX, performed competitively in close comparison, particularly at baseline and moderate demand levels, and therefore should not be dismissed as an alternative. Taken together, the recommended near-term configuration is a three-CDC network anchored in Baltimore, MD; Atlanta, GA; and Aurora, CO, which the analysis identifies as the most cost-resilient arrangement across a range of demand and infrastructure conditions. See figure 10 for final network map.

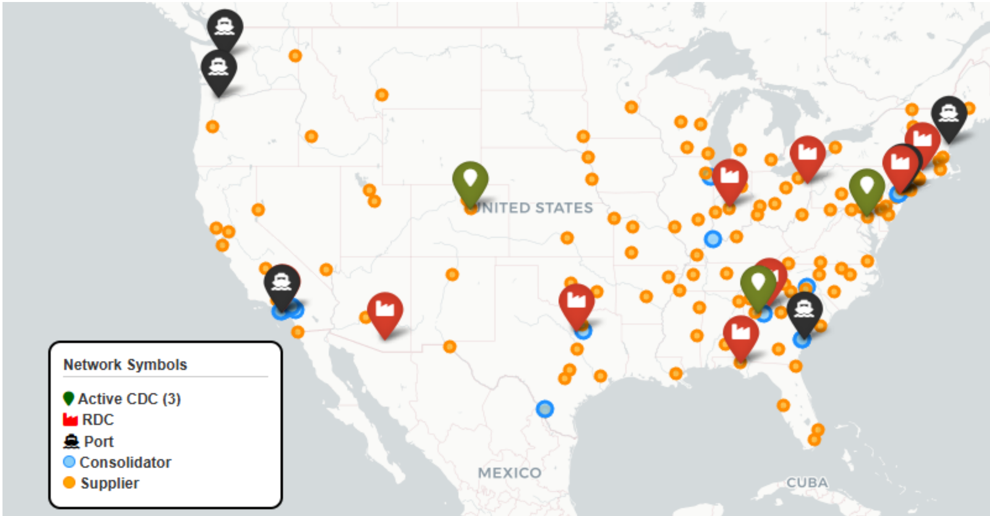

Figure 10: Proposed Future State Network Design

As illustrated in Figure 10, the proposed network strategically repositions consolidation capacity to reduce RDC dependency while improving geographic coverage and serviceability across the distribution system. The recommended three-CDC configuration, anchored in Baltimore, MD; Atlanta, GA; and Aurora, CO, serves as the upstream bridge that will enable greater flexibility in volume routing as the demand plan grows. Leadership is encouraged to prioritize Baltimore, MD and Atlanta, GA as the initial implementation sites, given their consistently optimal transportation positioning and demand corridor coverage across all scenarios tested. Aurora, CO can subsequently be phased in as volume thresholds and capital conditions warrant, with Odessa, TX remaining a viable alternative should Midwest demand materialize more than projected. This phased approach allows the organization to capture near-term cost savings while preserving strategic flexibility to scale the network in a financially disciplined manner.

5.2. Future Work

While this analysis establishes a strong macro-level foundation for CDC network design, two primary areas present opportunities for deeper investigation. First, future research should pursue a more granular, micro-level site selection analysis incorporating local zoning regulations, tax incentives, existing infrastructure, and facility-specific cost structures within the recommended locations. This level of specificity would transform the strategic network design into actionable site acquisition and development decisions. Second, transitioning from a deterministic to a stochastic modeling framework would meaningfully strengthen the robustness of future optimization efforts, allowing the model to explicitly account for demand uncertainty, cost variability, and supply disruptions rather than optimizing against fixed inputs alone. Third, future iterations should explore a constrained version of the model that incorporates lease terms, capital expenditure timelines, and phased network build-out scenarios, translating the optimization output into a financially grounded implementation roadmap. Together, these extensions would elevate the analytical foundation from a strategic planning tool to a fully operational decision support system capable of guiding both near-term implementation and long-term network evolution.

References:

Bahadori, M. S., Gonçalves, A. B., & Moura, F. (2022). A GIS-MCDM method for ranking potential station locations in the expansion of bike-sharing systems. *Axioms*, *11*(6), 263. https://doi.org/10.3390/axioms11060263

Battal, T. (2020). Understanding the logistics centre location decision: what are the main decision criteria and evaluation methods? *International Journal of Logistics Economics and Globalisation, 8(*2), 154–192.

Billal, M. M., Sebastian, R. M., & Kumar, A. (2025). Developing a GIS-based MINLP framework for optimizing the waste-to-energy supply chain with plant scale considerations. *Journal of Cleaner Production, 490,* 144709. https://doi.org/10.1016/j.jclepro.2025.144709

Chau, N. N., & Benson, G. E. (2025). Case article—Locating a truck terminal in Texas. *INFORMS Transactions on Education, 25*(2), 90–98. https://doi.org/10.1287/ited.2022.0042ca

Daskin, M. S., Snyder, L. V., & Berger, R. T. (n.d.). Facility location in supply chain design. In *Logistics systems: Design and optimization* (Chapter 2, pp. 17–42).

Dissanayake, I., & Gunathilake, M. (2024). A comprehensive survey of recent advances in facility location problems: Models, solution methods, and applications. *Journal of Sustainable Development of Transport and Logistics, 9*(2), 78–91. https://doi.org/10.14254/jsdtl.2024.9-2.6

Düzgün, R., & Thiele, A. (2012). Multi-Range Robust Optimization vs Stochastic Programming in Prioritizing Project Selection. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.2188286

Li, W., Finsa, M. M., Laskey, K. B., Houser, P., Douglas-Bate, R., & Verner, K. (n.d.). *Optimizing well placement for sustainable irrigation: A two-stage stochastic mixed integer programming approach* [Preprint]. Preprints.org.

Neol, D., Smith, R., & Idowu, B. (2025). Comparison of hybrid MCDM techniques in sustainable facility location under uncertainty. *Operations Research Perspectives, 12*, 100306. https://doi.org/10.1016/j.orp.2025.100306

Rikalovic, A., Cosic, I., & Lazarevic, D. (2014). GIS based multi-criteria analysis for industrial site selection. *Procedia Engineering, 69*, 1054–1063. https://doi.org/10.1016/j.proeng.2014.03.090

Snoussi, I., Hamani, N., Mrabti, N., & Kermad, L. (2021). A robust MILP model for sustainable collaborative distribution. *Mathematics, 9*(19), 2318. https://doi.org/10.3390/math9192318