Process Optimization Project Report

Module 3.4: Water Refill Station Planning

Your Name Entry Number: 202XXXXX **Department of Chemical Engineering, IIT Delhi February 17, 2026**

Declaration of Tool Usage

I declare that in completing this assignment:

- I used an LLM-based tool (Gemini) for assistance in:
- Formulating the Capacitated Facility Location Problem (CFLP) and p-median variants.
- Structuring the optimization model for station placement.
- Formatting the report in LaTeX/Markdown.
- I understand the submitted solution fully.
- I can explain and justify every part of my code and reasoning.
- I have verified all results independently.

Contents

1. **Introduction** 2. **Nomenclature** 3. **Assumptions and Justifications** 4. **Mathematical Model Formulation**

- 4.1 Objective Function Construction
- 4.2 Constraints Integration

5. **Optimization Analysis**

- 5.1 Complexity and Convexity
- 5.2 Algorithmic Approach

6. **Preliminary Insights and Discussion** 7. **References**

---

1. Introduction

To promote sustainability and reduce single-use plastic, the festival has adopted a reusable water bottle policy. The success of this initiative depends on the convenient availability of water refill stations. Module 3.4 focuses on optimizing the location and number of these stations. The objective is to minimize the combined cost of installation and user inconvenience (walking distance), subject to station capacity constraints, ensuring that attendees can easily refill their bottles without excessive queuing or travel.

2. Nomenclature

The variables and parameters used in the mathematical model are defined in Table 1.

Table 1: Nomenclature Table

Symbol	**Description**	**Units**	**Type**
i	Index for demand zones (footfall locations), i \in \{1, \dots, m\}	-	Index
j	Index for candidate refill station locations, j \in \{1, \dots, p\}	-	Index
d_i	Footfall / Demand at zone i	persons/hr	Parameter
D_{ij}	Walking distance from zone i to candidate location j	meters	Parameter
f_j	Fixed installation and operation cost of station at j	INR	Parameter
cap_j	Service capacity of station at j	persons/hr	Parameter
C_{walk}	Monetary value of walking distance (Time value)	INR/m	Parameter
y_j	Binary decision: 1 if station is installed at j, 0 otherwise	-	Decision Var
x_{i,j}	Fraction of demand from zone i served by station j	-	Decision Var
Z	Total Cost (Installation + Walking Inconvenience)	INR	Objective Fn

3. Assumptions and Justifications

1. **A1: Capacitated P-Median Variant.**
    - **Justification:** The problem is modeled as a variant of the Capacitated Plant Location Problem (CPLP) or P-Median Problem. We assume the number of stations (*k*) is not strictly fixed but determined by the trade-off between installation cost and walking cost, although the problem statement mentions "Pick k locations", implying *k* might be fixed or bounded. We treat *f_j* as the penalty for adding a station.
2. **A2: Splittable Demand.**
    - **Justification:** We allow demand from a single zone to be split between multiple stations (*x_{i,j} \in [0,1]*). In reality, users choose the station, but at the aggregate level, this models the probability

distribution of users moving to different nearby stations.

3. **A3: Finite Candidate Set.**
   - **Justification:** Stations can only be installed at pre-designated feasible locations (e.g., near plumbing lines), not anywhere in continuous space.

4. Mathematical Model Formulation

The problem is formulated as a mixed-integer linear programming (MILP) model.

4.1 Objective Function Construction

We minimize the Total Generalized Cost (*Z*): 1. **Installation Cost:** Sum of fixed costs for all installed stations.

$$C_{install} = \sum_{j=1}^{p} f_j \cdot y_j \tag{1}$$

2. **User Inconvenience Cost:** Total weighted walking distance.

$$C_{user} = \sum_{i=1}^{m}\sum_{j=1}^{p}(d_i \cdot x_{i,j}) \cdot D_{ij} \cdot C_{walk} \tag{2}$$

Combined Objective:

$$Z = \sum_{j=1}^{p} f_j y_j + \sum_{i=1}^{m}\sum_{j=1}^{p}\alpha_{ij} x_{i,j} \tag{3}$$

Where *\alpha_{ij} = d_i D_{ij} C_{walk}* is the cost coefficient for assigning zone *i* to *j*.

4.2 Constraints Integration

1. Demand Satisfaction: All attendees in every zone must be served.

$$\sum_{j=1}^{p} x_{i,j} = 1, \quad \forall i \tag{4}$$

2. Station Capacity: The total flow assigned to station *j* cannot exceed its capacity.

$$\sum_{i=1}^{m} d_i \cdot x_{i,j} \le cap_j \cdot y_j, \quad \forall j \tag{5}$$

Note: This constraint also acts as the logical link; if y_j=0, capacity is 0, so no flow can be assigned.

3. Integer Constraints: Stations are either installed or not.

$$y_j \in \{0,1\}, \quad \forall j \tag{6}$$

4. Variable Domains:

$$0 \le x_{i,j} \le 1, \quad \forall i,j \tag{7}$$

5. Optimization Analysis

5.1 Complexity and Convexity

This is a classic **NP-hard** combinatorial optimization problem.

- **Convexity:** The LP relaxation (allowing *0 \le y_j \le 1*) is convex, but the integer restrictions create a non-convex feasible region.
- **Scale:** With *m* zones and *p* candidate sites, we have *p* binary variables and *m \times p* continuous variables. For campus-scale problems (*m \approx 50, p \approx 20*), this is easily solvable by Branch-and-Bound algorithms in seconds.

5.2 Algorithmic Approach

1. **Greedy Construction:** Start with 0 stations. Iteratively add the station that yields the max reduction in total cost (considering installation vs. walking savings). 2. **Lagrangian Relaxation:** Relax the demand constraints and penalize violations in the objective function. This decomposes the problem into independent knapsack-like problems for each candidate site, providing tight lower bounds for the optimal solution. 3. **Exact Solution:** Use a solver like Gurobi or CBC. The tight formulation of capacity constraints usually leads

to fast convergence.

6. Preliminary Insights and Discussion

- **Capacity vs. Distance:** If f_j (installation cost) is high compared to walking cost, the solution will favor fewer, large-capacity stations (centralized). If f_j is low, the solution will approach a "station in every zone" distributed topology.
- **Shadow Prices:** The dual variable associated with the capacity constraint of a station j indicates the value of expanding that station's capacity. High shadow prices suggest "bottleneck" stations that should be upgraded.
- **Fairness:** The current objective minimizes *average* walking distance. This might leave some remote zones with very long walks. A "Minimax" objective (minimize the maximum walking distance) could be added as a secondary constraint for equity.

7. References

1. Daskin, M. S. (1995). *Network and Discrete Location: Models, Algorithms, and Applications*. Wiley. 2. Mirchandani, P. B., & Francis, R. L. (1990). *Discrete Location Theory*. Wiley.