Tartan Trade Warehouse Construction

Arin Srisajjalertwaja (S2670853), Puttiwat Wanna (S2745852),
Theeramet Sinthumongkolchai (S2652504)

February 14, 2025

THE UNIVERSITY
of EDINBURGH

Contents

1 Aggregation

Due to the large number of candidates and customers, aggregation is necessary. K-means clustering is used as the main tool for this purpose. The advantage of using K-means clustering is its ability to handle large-scale data while providing a practical output—the centroid location, which is useful for logistics problems. However, only customers are chosen for clustering for the following reasons:

1. Aside from reducing computation time, this method better reflects real-life scenarios. Delivery companies often consolidate packages at a single pickup location, which naturally corresponds to the centroid in the model. This approach makes more sense than establishing a new warehouse at the centroid.
2. Oversimplification can be a major issue in clustering, and focusing solely on customers helps prevent this. One clear example is the impact on delivery cost calculations. If we cluster candidates instead of customers, delivery costs must be computed twice—once for transportation from suppliers to warehouses and again from warehouses to customers. In contrast, clustering customers allows for a single calculation, making the process more efficient.
3. Clustering increases the number of variables in the model. If both candidates and customers were clustered, the added complexity might not effectively reduce costs.

In addition, in the stochastic model section, the scenarios are also aggregated. The aggregation is done with the thought that this would be a better way to represent the overall data, instead of using the first k scenarios and cutting the rest out.

2 Question 1: The MECWLP with multiple time periods

2.1 Question 1a: Model Formulation

The problem can be formulated as a Mixed Integer Linear Programming Model. Since it involves in a large number of customers, warehouses, and suppliers, the complexity increases, making it difficult for the model to find a solution. Therefore, an aggregation method is used to reduce the size of customers by grouping them into clusters.

Parts of the model will be described in order, starting with Set Notations, Input Parameters, Output Variables, Constraints, and the Objective Function.

In addition, due to a limited time of the project, the maximum time limit of the model is set to 1 hour and the MIP Gap percentage that is deemed good enough is set to 0.1%.

2.1.1 Set Notations

- W: The set of candidate warehouse locations; W = $\{0, 1, \ldots, 439\}$
- C: The set of customer clusters; C = $\{0,\ 1,\ \ldots,\ 50\}$. By default, customers are grouped into 50 clusters. This number was selected due to a time limit of the project.
- P: The set of product groups; P = $\{0, 1, 2, 3\}$
- S: The set of suppliers; S = $\{0, 1, \ldots, 52\}$
- S_p: The set of suppliers for product group; $\mathrm{S}_p \subset \mathrm{S},\ p \in P$. This set is for constraints related to a supplier and the demand of customers. Since reducing the dimensions of decision variables for efficiency is the main goal, this set is used instead of adding a p dimension to variable $z_{w,s,t}$, which will be defined below.
- T: The set of time periods; $T = \{0, 1, \ldots, 9\}$

2.1.2 Input Variables

- $D_{c,p,t}$: Demand of product p for customer c during period t; $c \in$ C, $p \in$ P, $t \in$ T.
- U_w^W : Maximum capacity of warehouse w; $w \in$ W.
- U_s^S : Maximum capacity of supplier s; $s \in$ S.
- C_w^S : Setup cost of warehouse w; $w \in$ W .
- $C_{w,t}^O$: Operating cost of warehouse w per time period t; $t \in$ T.
- $C_{w,s}^A$: Cost per kg for allocating product from supplier s to warehouse w; $w \in$ W, $s \in$ S.
- $C_{w,c}^D$: Cost per kg for delivering product from warehouse w to customer c; $c \in$ C.

2.1.3 Output Variables

- $x_{w,c,t}$: Binary variable indicating whether warehouse w serves customer c in period t ; $w \in$ W, $c \in$ C, $t \in$ T. It has three dimensions since the model is designed to match a customer with a warehouse in a one-to-one relationship. While this might not be optimal, in real world, having multiple warehouses deliver to a single customer is difficult to manage and keep track of. In addition, since each warehouse can store multiple products, the demand for all products of a customer can be handled by a single warehouse.
- y_w: Binary variable indicating whether warehouse w is setup; $w \in$ W. Another dimension of t could be added to show which period of time this warehouse is setup but that would increase a runtime of the optimizer and y_w could be combined with $o_{w,t}$ to find which time period t the warehouse is setup. The main purpose of this y_w is for calculating the setup cost in the objective function.
- $o_{w,t}$: Binary variable indicating whether warehouse w is opened in period t; $w \in$ W, $t \in$ T. This will be used mainly to limit the distribution to only the open warehouse.
- $z_{w,s,t}$: Quantity of product supplied from supplier s to warehouse w in period t; $w \in$ W, $s \in$ S, $t \in$ T. This is a continuous value since each supplier does not send all of their available supply to a single warehouse.

2.1.4 Constraints

- **General Constraints**; Binary variables will be a either 0 or 1 and continuous could be any positive real number. Therefore, we have constraints as follows.

$$z_{w,s,t} \geq 0; \quad \forall w \in \mathrm{W}, \forall s \in \mathrm{S}, \forall t \in \mathrm{T},$$

$$y_w \in \{0,1\} \quad \forall w \in \mathrm{W},$$

$$o_{w,t} \in \{0,1\} \quad \forall w \in \mathrm{W}, \forall t \in \mathrm{T}$$

$$x_{w,c,t} \in \{0,1\} \quad \forall w \in \mathrm{W}, \forall c \in \mathrm{C}, \forall t \in \mathrm{T}.$$

- **Warehouse Opening and Operational Constraints**: The constraint shows if a warehouse is in operation of any time period, that means it got setup. If a warehouse is not in operation at all, then that leaves y_w and option of either 0 or 1. However, due to y_w being used in the objective function in the next section, it will be 0 due to minimization.

$$y_w \geq o_{w,t}, \quad \forall w \in W, t \in T \tag{1}$$

Another constraint means that once a warehouse is running, it needs to continue running in every subsequent time period.

$$o_{w,t} \geq o_{w,t-1}, \quad \forall w \in W,\ t \in T,\ t \geq 1 \tag{2}$$

- **Flow Conservation Constraints**: A supply to a warehouse should be equal to the demand that that warehouse covers.

$$\sum_{s \in S_p} z_{w,s,t} = \sum_{c \in C} (D_{c,p,t} \cdot x_{w,c,t}), \quad \forall w \in W, p \in P, t \in T \tag{3}$$

- **Warehouse Capacity Constraints**: A constraint limits the demand of customers to not exceed the warehouse capacity that it belongs to.

$$\sum_{c\in C}\sum_{p\in P}(D_{c,p,t}\cdot x_{w,c,t}) \leq U_w^W \cdot o_{w,t}, \quad \forall w\in W, t\in T \tag{4}$$

Another alternative way would be to limit the warehouse capacity with the amount of supply to that warehouse, due to the constraint (3) being an equality constraint.

$$\sum_{s\in S} z_{w,s,t} \leq U_w^W \cdot o_{w,t}, \quad \forall w\in W, t\in T \tag{5}$$

After trying with both constraints, it seems that the alternative method (constraint (5)) took longer to solve the model that the former one (constraint (4)). Therefore, constraint (4) is picked.

- **Customer Allocation Constraints**: Each customer is assigned to only one warehouse. This reduces the complexity of assigning customers to multiple warehouses and makes the model more flexible. For example, if we assign a customer to multiple warehouses, if a supplier from, let's say, the second warehouse got cut off, that means the routing needs to be recalculated and would be too complicated to intuitively think of an interim solution. However, if a customer is assigned to only one warehouse, once something bad happens, only a group of customers in that warehouse need to be handled.

$$\sum_{w\in W} x_{w,c,t} = 1, \quad \forall c\in C, t\in T \tag{6}$$

Each customer can only be assigned to an operating warehouse.

$$x_{w,c,t} \leq o_{w,t}, \quad \forall w\in W, c\in C, t\in T \tag{7}$$

An alternative way for constraint (7) would be as follows

$$\sum_{c\in C} x_{w,c,t} \leq M\cdot o_{w,t}, \quad \forall w\in W, c\in C, t\in T \tag{8}$$

where M in this case could be the total number of customers, which is 50 by default. However, this constraint has already been proven with worse runtime, therefore, the former constraint was chosen.

- **Supplier Capacity Constraints**: Each supply from a supplier cannot exceed the supply capacity of that supplier.

$$\sum_{w\in W} z_{w,s,t} \leq U_s^S, \quad \forall s\in S, t\in T \tag{9}$$

2.1.5 Objective Function

The objective function minimizes the total cost, which includes setup costs, operating costs, and transportation costs (from supplier to warehouse and warehouse to customer):

$$\begin{aligned}\min(&\sum_{w\in W} C_w^S \cdot y_w + \sum_{w\in W}\sum_{t\in T} C_{w,t}^O \cdot o_{w,t} + \\ &\sum_{w\in W}\sum_{s\in S}\sum_{t\in T} C_{w,s}^A \cdot z_{w,s,t} + \sum_{w\in W}\sum_{c\in C}\sum_{p\in P}\sum_{t\in T} C_{w,c}^D \cdot D_{c,p,t}\cdot x_{w,c,t})\end{aligned} \tag{10}$$

2.2 Question 1b: Results

Based on the model above, experiments are conducted by varying the numbers of customer cluster. The results are presented in section 2.2.1, and the analysis of them is provided in section 2.2.2.

2.2.1 Comparisons between Different Number of Clusters

Different numbers of clusters were picked and the results of each were observed as follows.

N-Clusters	Objective Value	Setup Cost	Operation Cost	Transportation Cost (S $\rightarrow$ W)	Transportation Cost (W $\rightarrow$C)	Run Time (seconds)	MIP Gap %
50	39,842,948£	5,846,000£	4,395,800£	4,579,689£	25,021,459£	325	0.092
75	39,835,269£	5,846,000£	4,395,800£	4,582,145£	25,011,324£	260	0.096
100	39,606,368£	5,846,000£	4,395,800£	4,568,233£	24,796,335£	305	0.086
150	39,580,713£	5,846,000£	4,395,800£	4,556,413£	24,782,500£	700	0.098
200	39,654,485£	5,846,000£	4,395,800£	4,556,700£	24,855,985£	1040	0.086

Table 1: Comparing Results between Different Numbers of Clusters Used in Aggregation

The warehouses set up in each time period are the same for each number of clusters used.

Time Period (T)	Warehouses Setup
0	18, 314
1	182
2	-
3	296
4	-
5	47
6	-
7	338
8	-
9	-

Table 2: Warehouse setup decisions over time.

2.2.2 Analysis

1. **Same Warehouse Setup:** Even though different numbers of clusters were used to aggregate the data, the result for the location remains the same. This is possible since the demand of customers remain the same for every number of clusters used. Even if more clusters were used, the center of those demands would still be in the same location (or very close). Therefore, with the warehouse and supplier location being fixed, this makes the distance between a warehouse and a customer cluster in each number of clusters used to be very close to each other. Moreover, having the same warehouses set up makes the setup cost and operation cost the same for each number of clusters and other costs are in the same proportion.

2. **Very Low MIP Gap:** Since the data were aggregated, the model could find the solution more easily. In this case, the solver can always find an optimal solution, making the MIP Gap Percentage very low.

3. **Run Time Difference:** Looking at the runtime, a question might arise about why the lowest amount of clusters, 50, has a higher run time than 75 and 100 clusters. One possible answer could be related to how clusters are decided. Good clustering could help with finding the optimal solution. It could be said that a good range for clustering using K-Mean method is around 50-100 in this case. Looking at 150 and 200 clusters, the run time is a lot higher than the rest. This could mean that there are too many data points for the solver to run, resulting in a long run time. An additional research could be conducted to find another to do the clustering, which might give better results.

4. **Transportation Cost Change:** In 50-cluster to 150-cluster model, the transportation cost decreases for both from a supplier to a warehouse and from a warehouse to a customer. The two costs go in the same direction due to their relation in constraint (3). The decreasing trend could be due to the customer locations were oversimplified, or to say, those numbers of clusters were not enough. However, on the last model, using 200 clusters, both transportation costs slightly go up. This could be the opposite case of under-simplifying the customer groups.

3 Question 2: The MECWLP with multiple time periods and uncertain demands

3.1 Question 2a: Stochastic Model Formulation

In this section, the deterministic model is extended to two-stage stochastic model by introducing the scenario set, then, adjusting some input, output variables, constraints, and objective function to deal with uncertainty based on each scenario. The clustering for customer locations is fixed to 50 clusters in this model. The scenarios were clustered into 3, 5, 10, and 15 clusters and the results are compared in the later section. In addition, due to a limited time of the project, the maximum time limit of the model is set to 1 hour and the MIP Gap percentage that is deemed good enough is set to 0.1%.

3.1.1 Additional Sets and Notation

- Ξ : Set of clustered scenarios $\Xi = \{0, 1, 2\}$. The default number of clusters for scenarios is 3.

3.1.2 Adjusted and Additional Input Parameter

- $D^{\xi}_{c,p,t}$: Demand of product p for customer c during period t under scenario ξ; $c \in$ C, $p \in$ P, $t \in$ T, $\xi \in \Xi$.
- σ^{ξ} : Probability for scenario ξ to happen; $\xi \in \Xi$.

3.1.3 Adjusted Output Variables

Now that scenarios are introduced, related variables need to be adjusted.

- $x^{\xi}_{w,c,t}$: Binary variable indicating whether warehouse w serves customer c in period t under scenario ξ; $w \in$ W, $c \in$ C, $p \in$ p, $t \in$ T, $\xi \in \Xi$.

3.1.4 Adjusted Constraints

- **Customer Allocation Constraints:** Since a dimension of scenarios is added, the constraints need to be adjusted accordingly.

$$x^{\xi}_{w,c,t} \leq y_{w,t}, \quad \forall w \in W, c \in C, t \in T, \xi \in \Xi \tag{11}$$

$$\sum_{w \in W} x^{\xi}_{w,c,t} = 1, \quad \forall c \in C, t \in T, \xi \in \Xi \tag{12}$$

- **Flow Conservation Constraints**: A supply to a warehouse should be sufficient, instead of equal, to the uncertain demand that warehouse covers.

$$\sum_{s \in S_p} z_{w,s,t} \geq \sum_{c \in C} (D^{\xi}_{c,p,t} \cdot x^{\xi}_{w,c,t}), \quad \forall w \in W, p \in P, t \in T, \xi \in \Xi \tag{13}$$

- **Warehouse Capacity Constraints**: The warehouse maximum capacity needs to be sufficient according to the added scenarios.

$$\sum_{c \in C} \sum_{p \in P} (D^{\xi}_{c,p,t} \cdot x^{\xi}_{w,c,t}) \leq U^W_w \cdot o_{w,t}, \quad \forall w \in W, t \in T, \xi \in \Xi \tag{14}$$

3.1.5 Objective Function

The objective function for this stochastic model is formulated below,

$$\begin{aligned} \min \sum_{w \in W} C^S_w \cdot y_w + \sum_{w \in W} \sum_{t \in T} C^O_{w,t} \cdot o_{w,t} + \sum_{w \in W} \sum_{s \in S} \sum_{t \in T} C^A_{w,s} \cdot z_{w,s,t} + \\ \sum_{\xi \in \Xi} \sigma^{\xi} \cdot (\sum_{w \in W} \sum_{c \in C} \sum_{p \in P} \sum_{t \in T} C^D_{w,c} \cdot D^{\xi}_{c,p,t} \cdot x^{\xi}_{w,c,t}) \end{aligned} \tag{15}$$

The first three term formulation is the same as the deterministic version. They are decided by the first-stage variables which do not depend on uncertainty. However, the last term is determined by the second-stage variables, which are based on demand uncertainty. Therefore, by taking the expectation of this term, the recourse function is obtained.

3.2 Question 2b: Results

Similarly to Section 2.2, another experiment using stochastic model is conducted by adjusting the numbers of scenario clusters.

3.2.1 Comparisons between Different Number of Scenario Clusters

N-Clusters Scenarios	Objective Value	Setup Cost	Operation Cost	Transportation Cost (S $\rightarrow$ W)	Transportation Cost (W $\rightarrow$C)	Run Time (seconds)	MIP Gap %
3	40,677,340£	7,714,000£	6,268,800£	4,484,065£	22,210,475£	1183	0.098
5	41,461,047£	8,098,000£	5,997,600£	4,504,428£	22,861,019£	3600	2.173
10	50,211,4827£	12,504,000£	7,930,400£	5,900,465£	23,876,617£	3601	18.102
15	111,273,3027£	11,676,000£	8,114,700£	14,308,881£	77,173,722£	3600	100.000

Table 3: Comparing Results between Different Numbers of Scenario Clusters Used in Aggregation

The warehouses set up in each time period are as follows.

Time Period (T)	Warehouses 3-Clusters	Warehouses 5-Clusters	Warehouses 10-Clusters	Warehouses 15-Clusters
0	42, 103	42, 314	42, 314	12, 16, 17, 18, 19
1	182, 203	182, 203	18, 182	23, 24
2	-	-	-	-
3	47	383	103	25, 26
4	-	-	47, 203, 400	28
5	314	47	-	-
6	-	338	338, 383	29
7	338	-	231	30
8	-	53	53	-
9	-	-	-	32

Table 4: Warehouse Setup in Stochastic Model

3.2.2 Analysis

1. **MIP Gap Percentage**: Looking at different numbers of scenario clusters, the MIP Gap Percentage keeps increasing with the number of clusters. This is due to an increase in the number of data points when including scenarios. At first, with 3 scenario clusters, it could be compared to the 50 customer clusters in the deterministic model on Table 1. It has longer run time due to an increase in the data points from the added scenarios dimension. However, this increase in the data points is not linear, which is what makes the spike in MIP Gap percentage as the number of clusters grows.

2. **Runtime Increase**: Due to an increase in dimension, the run time also increases. It could be seen that the run time for 5, 10, and 15 clusters were capped at 3600, which is the maximum time limit set in the model. Relating to the MIP Gap Percentage, with the same amount of run time, lower number of clusters has lower MIP Gap Percentage. This supports the theory that increased data points lead to increased run time.

3. **Value of Stochastic Solution**: This stochastic model provides solutions for handling demand uncertainty. The decisions from this model, as shown in Table 4, differs from those in the deterministic model, in Table 2, leading to a higher cost to account uncertainty. Hence, the effectiveness

of incorporating randomness can be evaluated by calculating the Value of the Stochastic Solution (VSS), for example, in the 3-Cluster Scenario cases,

$$\text{VSS} = \text{Obj}^{Sto} - \text{Obj}^{Deter} = 40,677,340 - 39,842,948 = 834,392.$$

This value represents approximately 2 percent of the deterministic objective value, indicating that the stochastic model, with 3 clustered scenarios, is beneficial for managing randomness.

3.3 Question 2c: Relaxed Formulation

To deal with high-demand scenarios that would lead to unserved demand or excess capacity, the stochastic model can be formulated with constraint relaxation. For example, when relaxing the maximum capacity constraint, the original stochastic model is extended to identify which warehouse will possibly violate this capacity constraints.

3.3.1 Additional Input Parameter

- $\lambda_{w,t}$: Penalty Cost for excessive capacity of warehouse w at period t; $w \in \text{W}, t \in \text{T}$.

3.3.2 Additional Output Variables

- $\alpha^{\xi}_{w,t}$: Excess capacity of warehouse w at period t under scenario ξ; $w \in \text{W}, t \in \text{T}, \xi \in \Xi$.

3.3.3 Adjusted Constraints

- **Warehouse Capacity Constraints**: Since the constraint is relaxed using new variables, it should be as follows.

$$\sum_{c \in C} \sum_{p \in P} (D^{\xi}_{c,p,t} \cdot x^{\xi}_{w,c,t}) \leq U^{W}_{w} \cdot o_{w,t} + \alpha^{\xi}_{w,t}, \quad \forall w \in W, t \in T, \xi \in \Xi \tag{16}$$

3.3.4 Additional Objective Function

To prevent this model from violating the constraint excessively, a penalty term for this violation is added in the objective function. Since this violation is based on uncertainty, the penalty term is included in the recourse function, so, the additional objective function is as follows.

$$\sum_{\xi \in \Xi} \sigma^{\xi} \cdot (\sum_{w \in W} \sum_{t \in T} \lambda_{w,t} \cdot \alpha^{w}_{w,t}) \tag{17}$$