1 Model

Let $\mathbf{O}$, $\mathbf{T}$, $\mathbf{W}$ be the set of origin (o-nodes), transshipment (t-nodes) and warehouse (w-nodes) nodes, respectively. Let, π be the unit cost of missed demand, d_{ij} be the distance nodes i and j. Let D_w be the demand continuous random variable at w-node with density and distribution functions $f_{D_w}(x)$ and $F_{D_w}(x)$, with mean m_w and standard deviation σ_w. We define three sets of decision variables: (i) X_{ot}, number of packages delivered between o-node and t-node, (ii) Y_{ow}, number of packages delivered between o-node and w-node, (iii) Z_{tw}, number of packages delivered between t-node and w-node. We have developed a two-phase model to identify the optimal number of packages to be delivered from the origins to the warehouses. The first phase is a non-linear cost minimization problem with linear constraints. The cost function is comprised of three parts: (i) the cost of distribution of the packages, (ii) the cost of missed demands at the warehouse nodes, (iii) the cost of inequity in the number of packages received by the warehouses.

1.1 Cost of distribution

The cost of distribution is defined analytically as follows.

$$P_1 = \left(\sum_{o\in\mathbf{O}} \sum_{t\in\mathbf{T}} \frac{X_{ot}}{1000}(100 + 2 \cdot d_{ot}) + \sum_{o\in\mathbf{O}} \sum_{w\in\mathbf{W}} \frac{Y_{ow}}{1000}(100 + 2 \cdot d_{ow}) + \sum_{t\in\mathbf{T}} \sum_{w\in\mathbf{W}} \frac{Z_{tw}}{1000}(100 + 2 \cdot d_{tw}) \right)$$

Note that, we have combined the cost per package and distance per package criteria from the document in the same objective of cost of distribution. We assume that minimizing the cost of distribution will optimize the cost per package and distance per package criteria.

1.2 Cost of missed demand

The cost of missed demand is defined as a classic Newsvendor Loss function. Let, R_w be the total amount received by w-node, $w \in \mathbf{W}$, i.e., $R_w = \sum_{o\in\mathbf{O}} Y_{ow} + \sum_{t\in\mathbf{T}} Z_{tw}$. For any given w-node $w \in \mathbf{W}$, the Newsvendor Loss function can be defined as follows.

$$\mathcal{L}_w = \int_{R_w}^{\infty} (x - R_w) f_{D_w}(x) dx$$

Thus the total cost of missed demand can be expressed as follows: $P_2 = \pi \left(\sum_{w\in\mathbf{W}} \mathcal{L}_w\right)$

1.3 Cost of inequitable distribution

We define an equitable distribution as follows. A distribution is equitable when all warehouses receive packages in proportion to their respective "average" demand. In other words, the equitable proportion of packages received by any given w-node $w \in \mathbf{W}$ can be expressed as follows.

$$\mathcal{E}_w = \frac{m_w}{\sum_{w\in\mathbf{W}} m_w} \cdot \sum_{w\in\mathbf{W}} R_w$$

Thus the total cost of inequitable distribution can be defined as follows: $P_3 = \sum_{w\in\mathbf{W}} (R_w - \mathcal{E}_w)^2$.

The overall cost function to identify optimal delivery policy can thus be written as follows.

$$\begin{aligned} min \quad & P_1 + P_2 + P_3 && (1) \\ s.t. \quad & \sum_{w\in\mathbf{W}} Z_{tw} \le \sum_{o\in\mathbf{O}} X_{ot}, \qquad \forall t \in \mathbf{T} && (2) \end{aligned}$$

Equation 2 ensures that each t-node cannot ship more than their received number of packages from the o-nodes. After solving the first phase problem, we determine the number of FTL (Full truckload) and LTL (less than truck-load) required across each arc. For the second phase, we only consider the LTL quantities and develop a VRP (Vehicle Routing Problem) model to determine the optimal routes and truck numbers required to ship the LTL quantities.

2 Methods and Results

2.1 Demand allocation

To identify the parameters of the demand distribution at each w-node, we at first fit distribution to all customer zip-code demands. We have seen that, demands at each zip-code follows normal distribution. Then we assign each zip-code to its closest warehouse. To estimate the demand parameters for a given warehouse, we pool the demand at the zip-codes assigned to that warehouse. We consider the correlations within the zip-codes demand while pooling and estimate the mean and standard deviation parameters accordingly.

2.2 Two-phase model

In phase-1, we have developed a nonlinear optimization model in python and solve the model using "trust-region constrained algorithm" from the "Scipy" library in Python. We calculate the number of trucks for the FTL quantities (number of packages) across each arc in the network and separate the LTL quantities to be solved by phase-2 or VRP model. In the second phase, we develop a VRP model in Matlab by using the library "Matlog" to identify the optimal routing solution for the LTL quantities. Note that, the routing do not increase the cost of missed demand or the cost of inequity solved in phase 1.

We acknowledge the fact that finding optimal number of trucks to distribute the optimal number of packages to be shipped to distribution centers requires solving a integer programming problem. Since the simplest of integer programming problems, i.e, the knapsack problem, is NP-hard, we take a continuous approach here. In phase-1 model, the number of packages are considered continuous. After solving phase-1 model, we take the optimal number of packages and round up to make them usable in a real setting. As the number of packages are continuous variable, the number of trucks required to distribute the packages are also continuous. For the FTL quantities across each arc, we can calculate the integer number of trucks by dividing the number of packages by 1000. We then take the LTL quantities and feed them to a VRP model. The VRP model generates routes considering optimal pickup and distribution. To show the effectiveness of the VRP model, we have calculated the cost of distribution when the LTL quantities are distributed in full truck-loads, i.e., the continuous number of trucks from Phase-1 are rounded up to the nearest integer. We have seen that the inclusion of VRP reduces the cost of distribution by 12% (in total, 2200 dollars). The VRP model generates an integer number of trucks for each route.

There is another motivation behind the discussion of rounded-up number of trucks from Phase-1. While VRP reduces the cost of distribution, it is computationally expensive, i.e., NP-Hard problem. If the decision maker does not have enough computational complexity, they can just use the results from Phase-1 model and round up the number of trucks.

2.3 Results

All the codes can be found in this GitHub link. The solution, i.e., the number of packages and the number of trucks, can be found attached or in this GitHub link under "Final Tables" folder. Under "TruckAssignement.xlsx" file, sheet "FTL_assignments" contains the FTL shipment information, and the "Routes" sheet contains the routes generated for the LTL quantities. We have seen that our approach generates a cost of distribution of 15792 USD, cost of missed demand of 313 USD and a cost of inequity of 0.12 USD. On the other hand, if the decision maker does not prefer the VRP model due to computational complexity, the rounded-up solution for Phase 1 generates a cost of distribution of 17943 USD, without any increase in cost of missing demand and cost of inequity. The solution can be found in "WC_Assignements" sheet under "TruckAssignement.xlsx" file. The "CustomerAssignment.xlsx" file contains the assignment of the customer Zip codes to the warehouses.