Cost to Serve

by

Douglas Tjokrosetio
MSc. Computer Science, University of Pennsylvania

and

Matías Opazo
B.S. Industrial Engineering, Pontificia Universidad Católica de Chile

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Signature of Author: ____
Department of Supply Chain Management
May 10, 2024

Signature of Author: ____
Department of Supply Chain Management
May 10, 2024

Certified by: ____
Milena Janjevic
Director, MIT Supply Chain Design Lab
Capstone Advisor

Accepted by: ____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Cost to Serve

by

Douglas Tjokrosetio
MSc. Computer Science, University of Pennsylvania

and

Matías Opazo

B.S. Industrial Engineering, Pontificia Universidad Católica de Chile

Submitted to the Program in Supply Chain Management
on May 10, 2024 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

The capstone project focuses on introducing a novel cost allocation model developed for the sponsor company. Faced with the complexities of serving thousands of clients, the company seeks to refine its weight-based allocation method to accurately identify profitable clients and improve business decision-making. Leveraging the Shapley method from cooperative game theory, the project proposes a model that incorporates geographic distances, weights, and other logistical factors into the cost allocation process. Shapley values are used to assign costs based on each client's marginal contribution to overall transportation costs, representing a significant advancement over the existing weight-based proportional method.

Initial results demonstrate the model's effectiveness in providing fair cost distribution, supported by various methods of analysis that quantify the improvement over the existing allocation policy. A divergence analysis was conducted, where the results of the divergence were captured as monetary values for the company. Furthermore, the model's implementation was augmented through machine learning, enabling predictive insights for cost allocation. This project provides improved operations and strategic planning capabilities and not only addresses the immediate needs of the sponsor company but also sets a precedent for "cost-to-serve" applications in logistics-intensive industries.

Capstone Advisor: Milena Janjevic
Title: Director, MIT Supply Chain Design Lab

ACKNOWLEDGEMENTS

I am grateful for the support and love I have received throughout this capstone project. My family has provided me with encouragement and warmth during challenging times, and my close friends have offered unwavering support and motivation. I also want to thank my girlfriend, Gracella, for being my rock. Lastly, I am profoundly thankful to Milena, whose guidance and insightful direction have been invaluable in navigating the complexities of this project.

DT 2024

I would like to express my deep gratitude to my beloved ones for their love and unwavering support throughout my journey at MIT. They are a significant reason for my presence here. Additionally, I am immensely thankful for the 'Masters of Metrics' scattered wide across MIT, who undoubtedly embody the essence of this institution and inspire with their passion. I also extend my heartfelt thanks to Milena Janjevic for spearheading us in this endeavour. Finally, I express my sincere gratitude to our Sponsor Company for providing us with the opportunity to apply our Supply Chain learnings in the real world.

MO 2024

Table of Contents

1. Introduction

The sponsor company of the capstone project is a company focused on the manufacturing and commercialization of chemical solutions tailored for the construction sector. The prime objectives of these solutions are to reduce the costs of construction with concrete, enhance material properties, and rectify construction defects during the building process (Portland Cement Association, 2023). With the construction sector boasting sales that surpass a trillion dollars, the industry is currently navigating through rough waters of supply chain disruptions, inflation, and elevated interest rates. However, there is anticipated growth in the commercial and infrastructure sectors driven by legislative initiatives such as the *2021 Infrastructure Investment and Jobs Act* and the *2022 Inflation Reduction Act* (Pigott, 2023).

In this context of economic distress and government fomented growth, the sponsor company is looking to assess its transportation cost allocation policy to better understand the cost to serve each of its clients. At present, the company informed that it employs a weight-based cost allocation approach, a methodology stemming from its contractual agreements with its dedicated fleet of freight contractors. Currently, and as informed by the firm, it serves thousands of clients in more than 6,500 locations from multiple manufacturing sites and distribution centers located in the US and Canada with a logistics budget of USD 44 million. By properly allocating the transportation costs to each client, the sponsor will be able to distinguish profitable clients from those that are not, assess the business performance on a client base, and develop data driven business decisions.

Upon implementing a new cost allocation model and freight-cost pricing policy, the sponsor will experience a cascade of tangible benefits, primarily centered around enhanced cost visibility. By using a more robust method and identifying the real costs associated with serving each customer, the sponsor can make better informed decisions about resource allocation. A clear understanding of cost structures is pivotal, as it provides the lens through which profitability can be assessed and the total freight cost can be reduced (Dahlberg et al., 2019). This project delivers valuable insight to the sponsor company on how cost allocations methods affect the business, allowing for more precise analyses and better-informed strategic decision-making.

1.1 Problem Statement

Every delivery route is a puzzle. The company must balance the need for timely service with the backdrop of varying costs and customer profiles. Some customers, though distant from the manufacturing plants, are strategically located near other clients, potentially offering economies of scope. Conversely, some clients might be geographically closer, but may be isolated from other clients, leading to extended and costlier shipments. The sponsor seeks to thoroughly identify the cost implications of serving each customer and further unravel pricing information to develop a robust cost allocation policy for its customer base.

In an effort to achieve better customer allocation, the company is poised to revamp its current methodology. Presently, the approach is anchored in weight-based calculations, concentrating exclusively on the proportion of goods delivered to clients relative to the entire batch sent via a specific route. This simple rule of allocation, however, fails to consider essential cost-influencing elements. The cost of serving customers for the sponsor is a well-defined function of several interconnected factors, primarily the spatial distance from manufacturing plants to customers and the relative proximity of customers to one another. Moreover, factors such as delivery times, weight, and frequency of stops, also play a role in the overarching cost paradigm, underscoring the necessity for their inclusion in cost allocation.

1.1 Scope Definition

The current cost distribution among clients is inequitable; as a result, the sponsor is likely overpaying for certain routes and customers. This imbalance detrimentally impacts their logistic cost margins, emphasizing the necessity for a more transparent approach to freight cost distribution. In response, this project introduces a new allocation model that employs the Shapley method from cooperative game theory (Shapley, 1953). This method calculates the marginal contribution of each client within a delivery route, incorporating variables such as geographical locations, proximity to other clients, order patterns, and distinct shipment data. This approach aims to accurately determine the true cost of serving each client, thereby facilitating a fair and equitable allocation of costs.

Furthermore, a machine learning model has been developed in tandem with the Shapley model to provide the sponsor with predictive capabilities regarding cost allocation for future scenarios and new clients. The capstone project does not seek to optimize delivery networks; rather, it prioritizes providing visibility to the customer landscape, comprehending the cost to serve clients, and enabling equitable cost allocation.

2. State of the Literature

To tackle the cost disparities and challenges of the sponsor, a new model centered around the Shapley method was used. To elucidate why this method was chosen, a study and comparison of various allocation methods is presented in this chapter. The objective is to find the most suitable method to replace the existing weight-based allocation system. The literature on cooperative game theory provides several well-established allocation solutions, among which are the Shapley Values (Shapley, 1953), Nucleolus (Schmeidler, 1969), and Proportional Methods. While each allocation framework seeks to ensure fair allocations, they are grounded in different fairness criteria.

2.1 Traditional Allocation Methods

The Shapley Value, formulated by Shapley, is a concept in cooperative game theory used to fairly distribute the total cost of a coalition among its participants. This method attributes a value to each participant *i* based on their marginal contributions. This is calculated by considering all possible ways a coalition could form and takes the average of the marginal contributions of each player across these coalitions, as shown in *Formula A*.

Formula A : $$\phi_i = \sum_{S \subseteq N | i \in S} \frac{(|S|-1)!(|N|-|S|)!}{|N|!}(c(S) - c(S \backslash \{i\}))$$

In *Formula A*, *N* is the set of all players, *S* is a subset of *N,* and *c* is a characteristic function representing the value of any coalition *S*. This method determines the individual impact or contribution of each player towards the subset *S* of *N*. Crucially, this method hinges on evaluating every possible permutation in which players can join the coalition, ensuring that each player's contribution is assessed in every conceivable way. This approach allows for equitable allocation of the coalition's total value among all participants.

The Nucleolus, introduced by Schmeidler (1969), is a cost allocation concept that aims to minimize the dissatisfaction among players in a cooperative game by focusing on the excess that coalitions experience. The excess in a coalition is interpreted as a measure of dissatisfaction. Hence, the larger the excess, the more dissatisfied the coalition is. The Nucleolus is a distribution of cost allocations that minimizes the largest excess vector, so that no player has an incentive to deviate from the coalition, as shown in *Formula B*.

Formula B : $$\mathcal{N}(Y) = \{x \in Y | \theta(x) \precsim \theta(y) \text{ for all } y \in Y\}, \qquad \mathcal{N}(v) = \mathcal{N}(X)$$

In *Formula B*, *x* is a payoff vector or allocation of values, *Y* is the set of possible payoff vectors, and *v* is a characteristic function that describes how values are assigned to different coalitions of players.

Proportional Methods are various allocation methods where shares of the total cost are assigned to each player *i* based on certain weights or criteria, as shown in *Formula C*.

Formula C : $$x_j = \alpha_j \cdot C(N) \quad \forall j \in N$$

In *Formula C*, *a_j* is a share of the total cost *C(N)*, and the sum of shares of all players is equal to 1. There are different criteria used to determine *a_j*, but the most straightforward is the *Egalitarian Method*. With this method, as defined by Dror (1990), all players are treated equally regardless of their individual contributions, hence *C(N)* is divided equally among the players. In addition to the Egalitarian Method, previous studies have divided costs among players proportionally to demand, distance, volume, capacity, consumption, or willingness to pay (check *Appendix A* for further breakdown of the different proportional allocation schemes).

It is evident that these traditional allocation methods each have distinct procedures as well as fairness criteria. To further reinforce the comprehension of allocation theories and judiciously select the most suitable method to solve the sponsor’s challenges, a systematic study of the existing literature on cost allocation theory was undertaken.

2.2 Literature Study

Searching through existing literature allowed for a deeper understanding of how allocation methods have evolved over time. By weaving together seminal works from renowned researchers such as Dror, Shapley, and Engevall, with more recent scholarly contributions, the body of literature maintains a balance of traditional methods and novel approaches.

The literature was examined through several areas. First, the primary cost allocation method of each article was identified, which was either Shapley, Nucleolus, or Proportional Method. This provides a comprehensive view of the allocation methods most frequently used in prior research and studies. Next, the specific strategy for allocation was identified, such as the *Core, Shapley Values, Demand Nucleolus, Equal Profit Method, Egalitarian Method,* and others. This enables an understanding of the specific techniques and strategies used for cost allocation and allows for an assessment of whether traditional methods remain effective over newer variations or derivatives.

Further, the main problem statement of each article was identified and classified into one of five broad categories: *Traveling Salesman, Vehicle Routing, Horizontal Cooperation, Inventory Systems,* and *Transportation Planning*. This categorization helps determine which problem types are previously tackled through different allocation methods. From the comprehensive analysis, a resulting corpus of 61 academic papers was compiled (see *Appendix A*).

2.3 Results of the Literature Study

The literature study on cost allocation methods within supply chains and logistics showcases a diverse spectrum of studies spanning over four decades. Figure 1 highlights that horizontal cooperation emerges as the dominant theme, indicating a rising interest in the complexities and significance of collaboration in supply chain networks. In contrast, inventory systems have comparatively fewer studies in the context of cost allocation. A closer examination on the dates of publications reveals that the most recent studies concentrate on horizontal cooperation and transportation planning. Conversely, the Traveling Salesman Problem, which has foundational works from as early as 1983, suggests its long-standing relevance to cost allocation methods and theories.

Problem Type	**Literature Count**	**Papers Discussed**
Traveling Salesman	7	Fishburn and Pollak (1983), Dror (1990), Engevall et al. (1998), Engevall et al. (2004), Yengin (2012), Kimms and Kozeletskyi (2016), Levinger et al. (2020)
Vehicle Routing	8	Göthe-Lundgren et al. (1996), Engevall et al. (2004), Krajewska et al. (2008), Crujjssen et al. (2010), Liu et al. (2010), Özener (2014), Zakharov and Shchegryaev (2015), Leenders et al. (2017)

Horizontal Cooperation	12	Özener and Ergun (2008), Agarwal and Ergun (2010), Granot et al. (2011), Lozano et al. (2013), Nguyen et al. (2014), Vanovermeire et al. (2015), Hezarkhani et al. (2015), Otero-Palencia et al. (2018), Algaba et al. (2018), Schulte et al. (2019), Zheng et al. (2019), Wang (2023)
Inventory Systems	5	Wong et al. (2007), Fiestras-Janeior et al. (2012), Özener et al. (2013), Fang and Cho (2014), Vos and Raa (2018)
Transportation Planning	9	Sakawa et al. (2001), Frisk et al. (2010), Massol and Tchung-Ming (2010), Wang et al. (2015), Flisberg et al. (2015), Wu et al. (2017), Gnecco et al. (2020), Bogachev et al. (2021), Estañ et al. (2021)

Figure 1 - *Overview of Literature by Problem Type*

Figure 2 illustrates the frequency of the different allocation strategies used in various problem types. These strategies fall under three main categories: *Shapley*, *Nucleolus*, and *Proportional Methods*. These categories include both traditional methods and its variants that have been developed over the years. For example, the *Shapley* category includes the *Core* concept, *Shapley Values*, *Shapley Monotonic Path (SMP), Line Rule*, and other *Shapley* strategy variants.

Problem Type	**Count of Allocation Methods**		
	Shapley	**Nucleolus**	**Proportional Methods**
Traveling Salesman	6	2	**3**
Vehicle Routing	7	4	**3**
Horizontal Cooperation	10	4	**4**
Inventory Systems	4	1	**2**
Transportation Planning	5	3	**5**
Total	32	14	**17**

Figure 2 - *Allocation Methods by Problem Type*

Figure 3 offers a detailed analysis of the data presented in Figure 2, delineating the specific allocation strategies applied to various problem types.

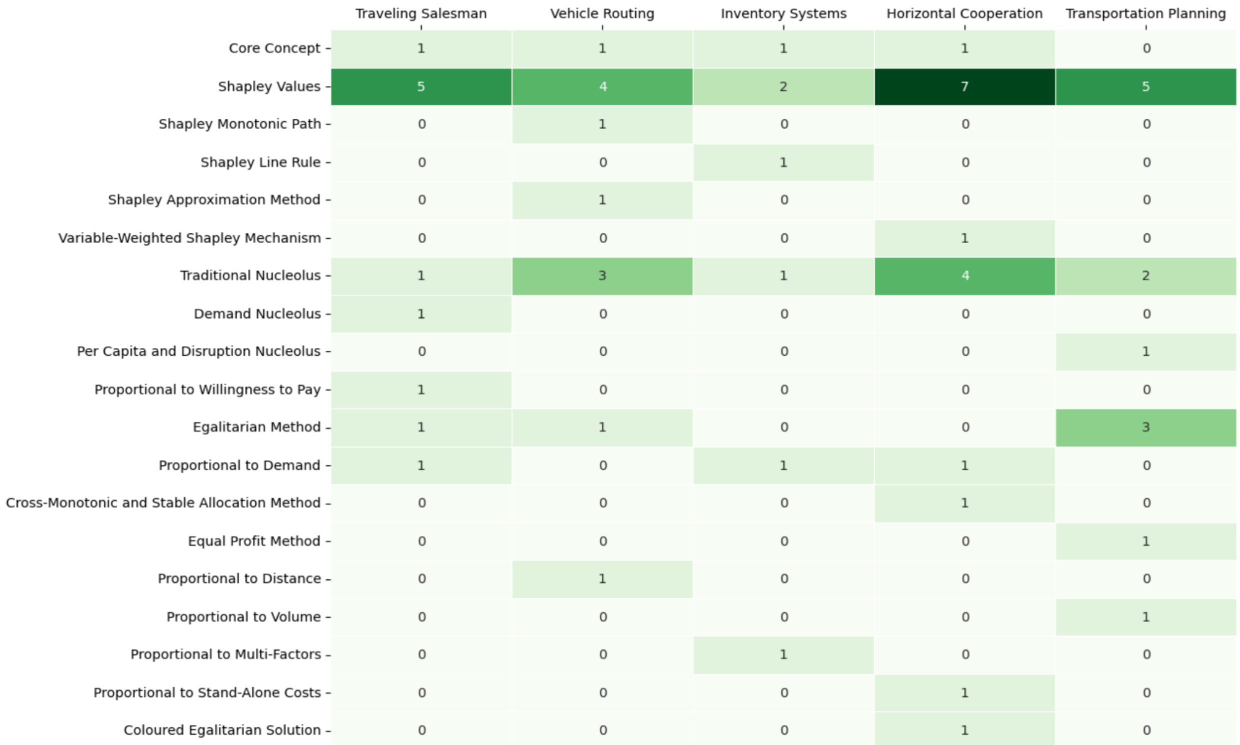


Figure 3 - *Heatmap of Allocation Strategies Used Across Literature*

Figures 2 and 3 indicate that the Shapley method is the predominant strategy in addressing logistics and supply chain cost allocation issues. This method's widespread adoption is largely attributed to its foundation in fairness principles, which considers the marginal contribution of each participant to the total incurred cost of serving all customers in a route. This approach distinctly contrasts with alternative methods, which typically do not connect cost allocation to an individual player's marginal cost.

In particular, the Nucleolus method proves unsuitable for the sponsor company, primarily because the sponsor operates as the singular decision-maker in their transportation network. Consequently, the Nucleolus method's focus on the collective dissatisfaction within a network becomes inapplicable, as the company solely governs the decision-making process. Additionally, proportional methods are inadequate, failing to offer equitable solutions for customers due to their disregard for the myriad factors influencing costs, such as distances, weights, volumes, and time.

Consequently, the Shapley method emerges as the most suitable for cost allocation within the sponsor’s network. It adeptly calculates each client’s unique cost contribution, taking into account crucial factors that influence the sponsor’s logistical cost function. This approach effectively highlights the true impact of each client on the costs associated with every delivery shift. Therefore, the concept of Shapley values has been adopted as the primary strategy to effectively tackle the challenges faced by the sponsor company. However, to ensure that this is the most suitable allocation method chosen for the sponsor company, it is imperative to establish a metric that quantifies the deviation between the Shapley method and the current weight-based allocation method.

2.4 Kullback-Leibler Divergence

The Kullback-Leibler (KL) divergence is a measure from information theory that provides a quantitative measure of how one probability distribution diverges from a second, reference probability distribution, as shown in *Formula D*.

Formula D : $$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x)\ \log\left(\frac{P(x)}{Q(x)}\right)$$

The KL divergence provides a singular numeric indicator that captures the divergence between the allocation distributions as generated by the Shapley method and those produced by the proportional method. This metric is beneficial for comparing how different methods distribute value among elements, such as allocating benefits among participants in a cooperative game (Kullback and Leibler, 1951).

Within the KL divergence theory, a KL value nearing 0 signifies a strong similarity (or minimal divergence) between the distributions, and a measure approaching 1 indicates a pronounced difference. As highlighted in *Formula D*, the logarithmic nature of the equation ensures that if the two allocation distributions *P* and *Q* are the same, the logarithm yields a value of 1, resulting in a KL value of 0, thus denoting that both distributions are identical. Conversely, largely divergent distributions will have an upper limit KL value of 1. This framework makes it feasible to quantitatively determine the extent of agreement or disparity in the allocations made by two allocation methods, which considers fairness and efficiency in resource distribution.

3. Project Methodology

To address the sponsor company's challenges of inequitable cost allocation, the following steps were pursued:

1. **Data Exploration**: This an initial step to comprehend the dataset provided, including its topology and characteristics of the sponsor's delivery network.
2. **Data Cleaning**: The data is then cleaned and refined so that it can be used for the model.
3. **Shapley Model Development**: A model is developed to derive the Shapley value of each client within each route or delivery shift.
4. **Model Execution**: The working model is applied to a single factory test case.

5. **Model Testing and Evaluation**: Upon successful model execution, the model is deployed across the entire network of factories. Analysis is conducted to characterize customers and understand freight dynamics.
6. **Divergence Analysis**: The output of the model is evaluated against the sponsor's existing allocation method. Divergence analysis is performed to assess the impact of the Shapley method versus the proportional method.
7. **Machine Learning Model**: A machine learning model is developed to predict whether the Shapley model or the existing weight-based model is more suitable for certain clients on specific delivery shifts. For customers unsuitable with the proportional method, guidance is given on how to properly charge these customers with respect to the divergence between the Shapley model and the existing weight-based model.
8. **Allocation Policy Development**: The Shapley model, machine learning algorithm, and obtained results is synthesized into a tailored allocation policy.

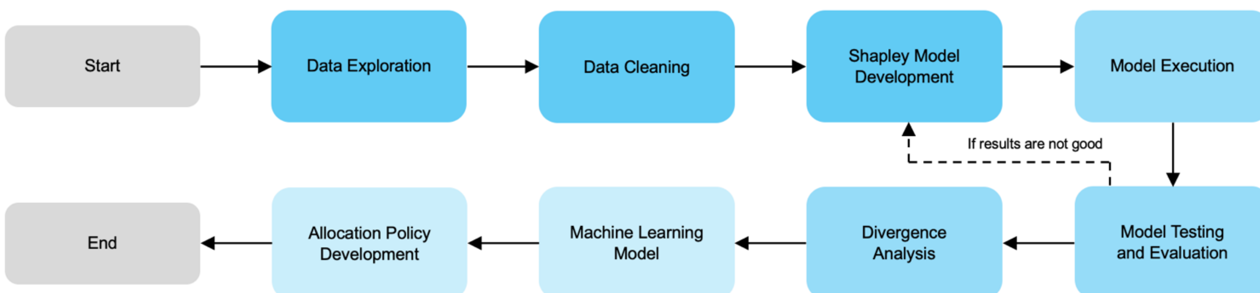


Figure 4 - *Project Methodology Overview*

4. Data Management and Model Development

4.1 Data Exploration and Cleaning

Prior to constructing the Shapley Model to address the challenges faced by the sponsor company, a thorough understanding of the underlying data is essential. An initial exploration revealed that the dataset contains 139,528 entries detailing 25,390 trips and 39,506 stops from June 2021 to July 2023. Each entry provides comprehensive details including information on shifts, deliveries, stops, locations, contractors, costs, trailers, and products.

After exploring the dataset, the data was subsequently cleaned to prepare it for use in the Shapley model. Upon review, the data appears to be correctly formatted according to their intended uses, such as date in date-time format, or costs and coordinates as numerical values. Further, missing and zero values were checked in the dataset. Analysis revealed that 8 rows of data were missing, representing a delivery shift with missing data values. Additionally, 3,529 rows contained zero values for the weight or volume of products shipped, hinting that these products were not delivered. Consequently, rows with zero weight or volume delivered were dropped as well. The missing data and zero values represent 2.53% of the entire dataset, hence the rows were dropped rather than handled with imputation.

After exploring and cleaning the dataset, important portions of the dataset had to be refined and extracted in meaningful ways for the model to run. The model's objective is to calculate the Shapley value for each client on every delivery route. This process involves several key steps: first, the dataset is segmented into smaller subsets that are filtered to each factory; subsequently, a distance matrix for each factory subset is generated; and finally, the results for each factory are used to compute the Shapley values for each client. These steps are depicted visually through a process map in Figure 5 and will be discussed further in the next subsections.

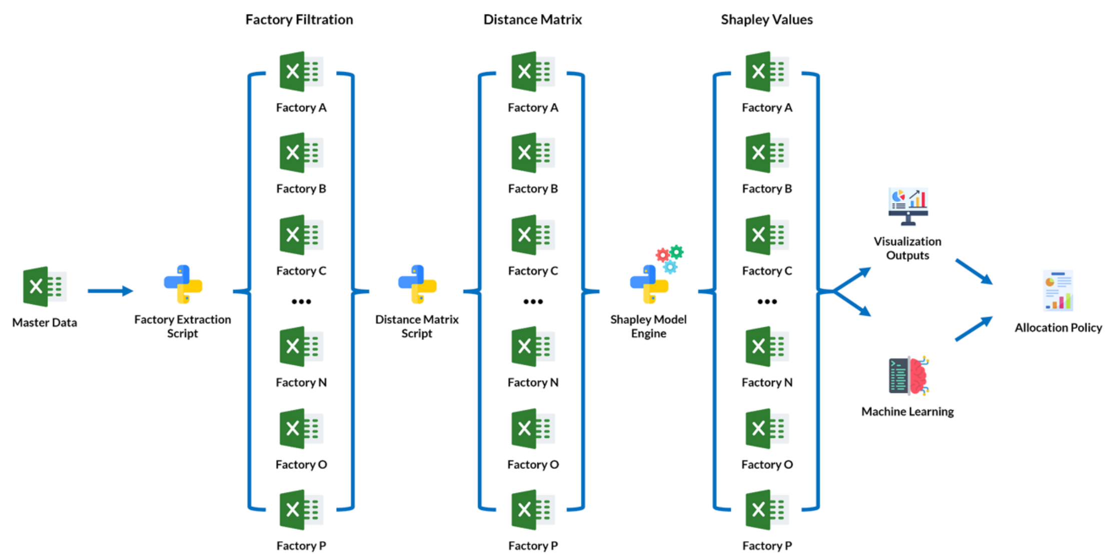


Figure 5 - *Project Process Map*

4.1.1 Factory Filtration

The primary dataset covers routes originating from the distinct factories of the sponsor company. To streamline and enhance the efficiency of calculating Shapley values, the master dataset was partitioned into separate segments, where each segment represents a specific factory of origin. Among these factory segments, the largest network comprises 19,436 entries, while the smallest network has merely 2 entries. On average, each factory network has 8,707 entries. The volume of data corresponding to each factory is illustrated in Figure 6.

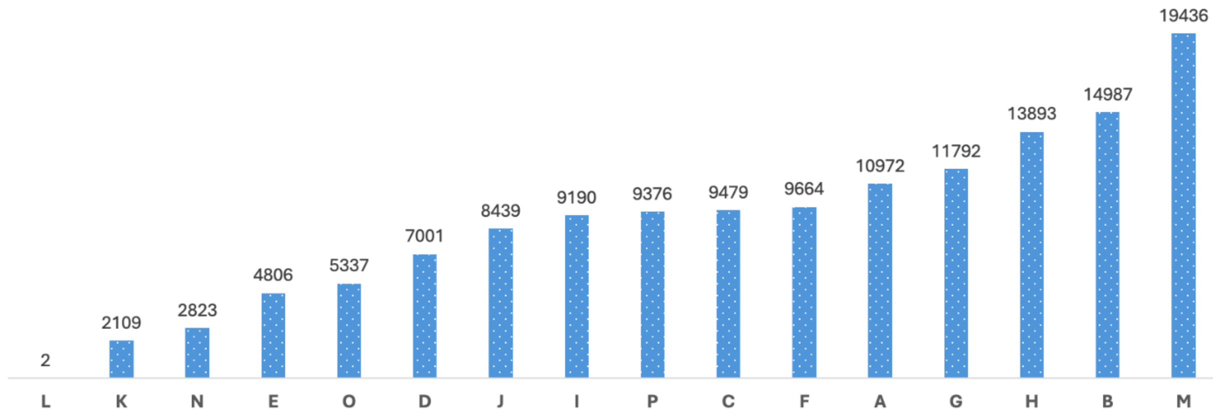


Figure 6 – *Number of Rows for Each Factory Dataset*

4.1.2 Distance Matrix

Upon segmenting the master dataset by factory, a distance matrix for each segment was constructed. This matrix calculates all possible distances between every client serviced by each respective factory using the Euclidean method. This approach computes the geometric distance between two points, as indicated by the Euclidean distance equation in *Formula E*.

Formula E : $$Relative\ Distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The creation of distance matrices between clients is a pivotal element of the analysis, serving as an essential component in the computation of Shapley values. The Shapley value of a client represents the marginal cost of adding a client into a route. Consequently, the distance factor plays a significant role in influencing this cost.

It is important to note that the Euclidean distance serves as a proxy for the traveled distance between routes, representing a lower bound of the actual distance traveled. Consequently, it does not fully capture the true distances covered in practice. This simplification was adopted for ease of use. Ideally, incorporating the actual traveled distances would have provided a more accurate measure, but this information was unavailable during the development of this project.

4.2 Shapley Model

Following the development of distance matrices for each factory dataset, these matrices were integrated into the Shapley model, which was implemented in Python. This model is responsible for calculating the Shapley value for each client on each route originating from each factory. This enables the determination of the actual cost attributed to a client within a route and the allocation of costs across the network accordingly.

The Shapley model incorporates several critical functions. First, it includes a *distance function* that determines the relative distance between clients, utilizing the data from the distance matrices. Second, it features a *logistic cost function* that computes the logistic costs for covering each distance, comprising both a fixed stop cost and a variable distance cost based on the data provided by the sponsor. Additionally, the model encompasses a *game accumulated cost function* that calculates the costs associated with running permutations of the network excluding the client for whom the marginal value is sought. This function is instrumental in determining the marginal cost of a client *i* for a given permutation *p* by calculating the difference between the cost of the route permutation with the inclusion of the client and without it, as illustrated by *Formula F*.

Formula F : $Marginal\ Cost_{(i,p)} = Total\ Route\ Cost_{(i+x)} - Game\ Accumulated\ Cost_{(x)}$

Ultimately, the model employs *Formula F* to find the marginal cost of each client for a given permutation. The Shapley value of a given client is then computed by taking the average of the marginal contributions across all permutations, which is done for every client on every route. This analysis is conducted across all factory datasets and the results are exported as new *Excel* files containing Shapley value information.

In a similar fashion, another model was developed to calculate a client's marginal distance within a route, rather than their marginal cost. Since the distances are known to the model through the distance matrices, an additional feature was implemented to calculate which portion of the distance was contributed by the addition of each client. Since only distance is considered, the computed marginal distances for each client on every route are relevant for further analysis, including the allocation of carbon emissions across routes.

Given the inherent complexity of the sponsor’s networks, certain factories pose greater challenges in computing Shapley values than others. *Factory H* presents the highest computational challenge, with over 13 shipment points within a single route. Given the necessity of calculating every possible client permutation in a route, this equates to 13 factorial or 6.23 billion permutations to assess to calculate the respective Shapley values. Conversely, *Factory D*, with a more manageable dataset and over 8 factorial permutations comprising 7,001 rows, represents one of the lowest complexity levels for computation outside the trivial case of *Factory L*. Details on the complexity of datasets in relation to model computation are provided in Figure 7.

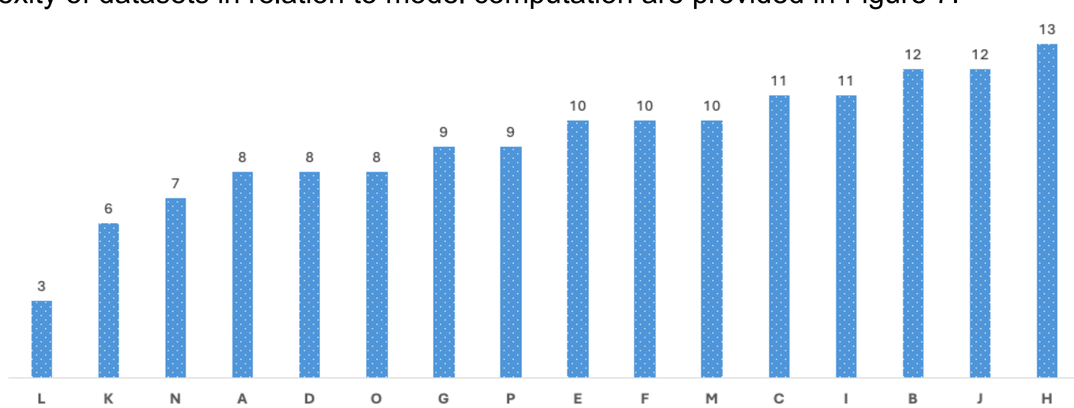


Figure 7 – *Number of Shipment Points for Each Factory Dataset*

With the *Factory D* dataset exhibiting one of the lowest complexities, the model was initially tested with this factory dataset alone. Following the successful error-free computation of Shapley values for each client on each shift from *Factory D*, the Shapley model was then deployed across the entire network of factories. For computational optimization, routes larger than 10 shipment points were filtered out, representing only 0.86% of the data (check *Appendix B* for further details on data filtration). The results obtained from the model are discussed in the following section.

5. Results

5.1 Shapley Values

Following the successful deployment of the Shapley model across the network of factories, distinct Excel files were generated for each factory containing the Shapley values of each client. The data in these files were aggregated on two levels. First, each row represents a delivery point and second, they are further grouped by the number of products. This aggregation process yields smaller factory dataset files, each including the Shapley values of each client. Figure 8 summarizes the overall outputs of the Shapley model.

Average Number of Rows	Total Number of Rows	Minimum Number of Rows	Maximum Number of Rows
6,496	97,442	1,495 (Factory K)	15,409 (Factory M)

Figure 8 – *Summary Statistics of Shapley Model Output*

Collectively, the data points from the model output files total to 97,444 Shapley values. The Shapley values highlight which clients have low marginal contributions on each shift, which ultimately helps in crafting an allocation policy for the sponsor company. However, given the vastness of the entire dataset, it becomes challenging to derive further meaningful insights directly from these values. Therefore, these results will be visualized to obtain a clearer understanding of the Shapley data.

5.1.1 Factory Analysis

It is first essential to gauge the relative contribution of each client within various factory networks. To this end, histograms of Shapley values were generated for each factory, elucidating the overall distribution of the Shapley values of clients served. Overall, the distributions of Shapley values are positively skewed across most factories, indicating that most clients have a relatively low marginal contribution, with a few accounting for significantly higher contributions. Contrastingly, factories like *Factory E* display a more even spread of Shapley values.

In addition, factories such as *Factory J* and *O* exhibit distributions where certain clients possess exceptionally high Shapley values. This may be due to the fact that these clients' routes have very few shipment points with far distances and varying weights. For instance, in *Factory J*, the shift 59492ORS serves only two clients, where client 5562806 has the highest recorded Shapley value in the entire *Factory J* network and also has a weighted distance tenfold greater than the other client in the shift. Additional insights into the distribution patterns of Shapley values across different types of factories are detailed in Figure 9.

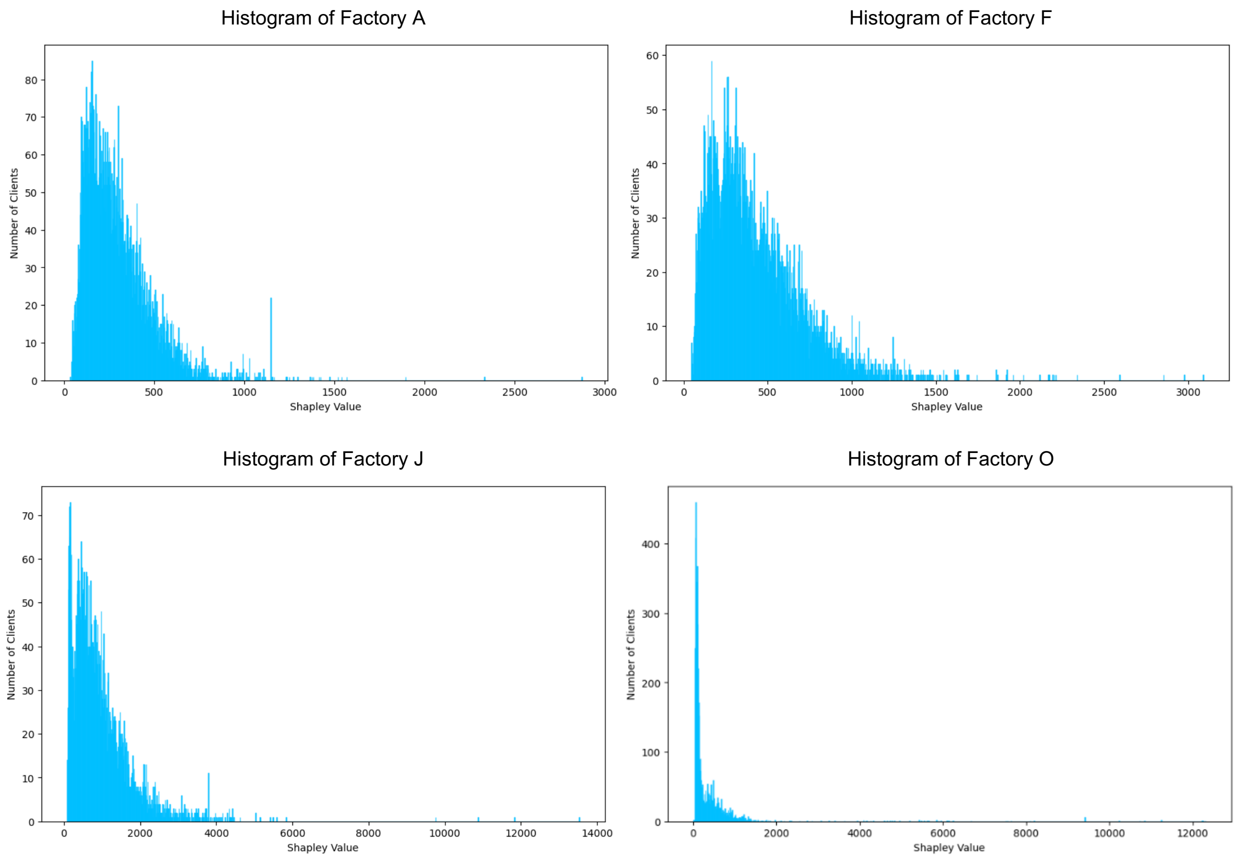


Figure 9 – *Histogram of Shapley Values Across Factories*

Figure 9 highlights different types of Shapley distributions for the sponsor company's factories (check *Appendix C* for the distribution across all factories). A Shapley distribution that *Factory A* follows has lower and more evenly distributed Shapley values, which hints that clients are located closer to the factory. However, a Shapley distribution for *Factory F* shows a more dispersed distribution of Shapley values, where certain clients may purchase more products or delivery trucks are filled with greater volumes. This is even more apparent for *Factories J* and *O* where the range of Shapley values are significantly wider among clients. The overall Shapley distributions reveal that there are certain customers on routes that have disproportionately high Shapley values potentially due to contrasting factors such as volume, distance, or weight.

In addition to understanding the Shapley value distribution of each factory, it is imperative to concurrently examine the geographical distribution and dispersion of clients relative to each factory. The inclusion of clients situated at far distances can significantly influence routes, thereby disproportionately escalating costs for other clients on the same shift. Allocating costs is a zero-sum game, where the gain of one party directly corresponds to the loss of another. Under the sponsor's proportional allocation method, clients closer to factories could essentially be subsidizing the additional expenses incurred by other clients significantly located further away, especially if the volume sold to distant locations is inversely proportional to the total distance traveled.

Figure 10 provides a visual representation of such scenarios, pinpointing clients considered outliers. These outliers are identified based on the empirical rule, specifically those whose locations lie beyond two standard deviations from the mean, as indicated by their labelled positions on the maps. It is evident that there are outliers across the factories' distribution networks that may possibly exert a disproportionate influence on the cost allocation for other clients along the same route (check *Appendix D* for visualization of all factory locations).

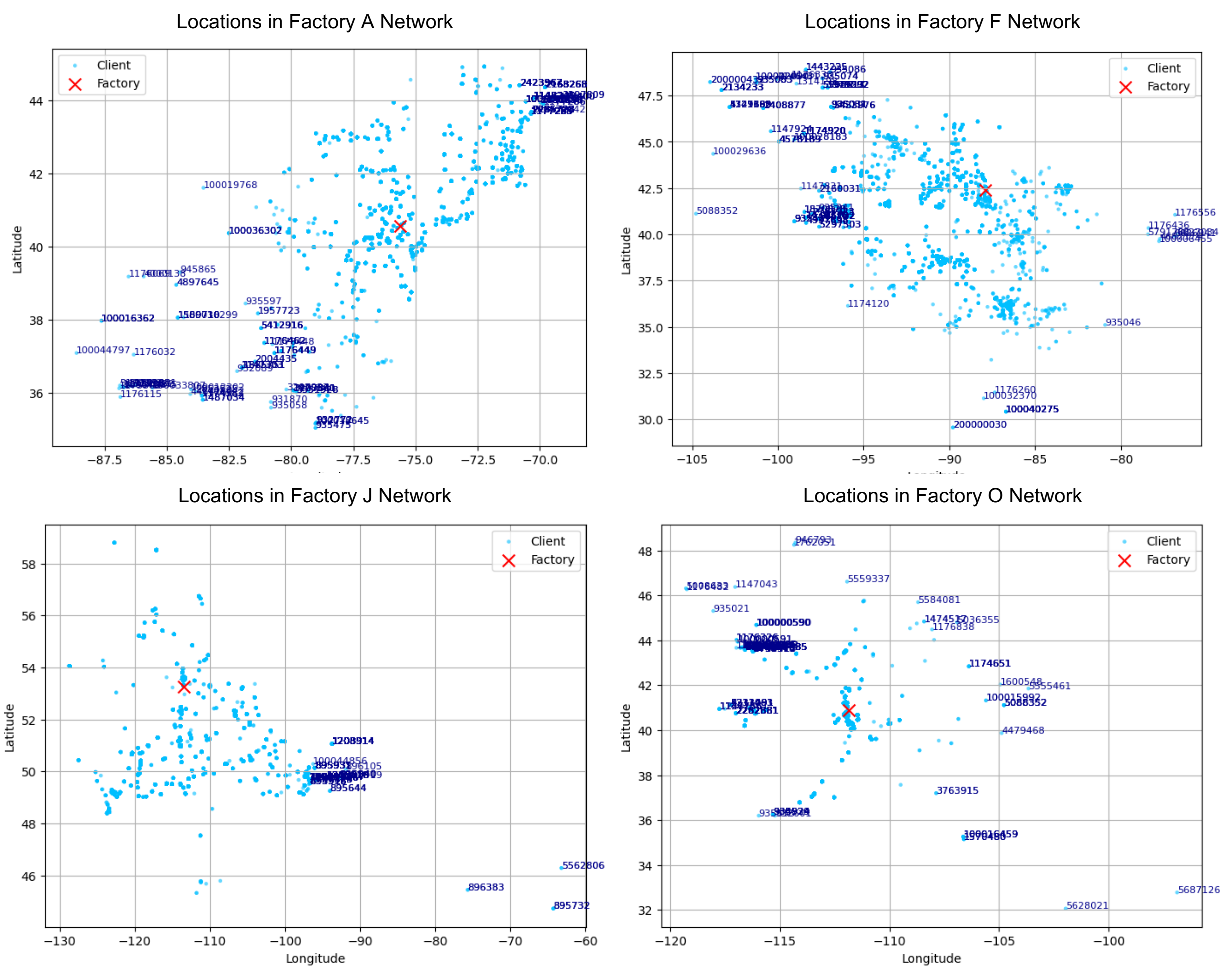


Figure 10 – *Geographical Visualization of Clients Across Factories*

5.1.2 Carbon Emissions

In addition to the Shapley values and geographic distances of locations across the network, the carbon emissions across the networks could also be assessed. This is achieved by computing the ratio of each client's marginal distance on a given route, multiplied by the route's carbon emissions. The marginal distance is calculated using the Shapley concept but focuses on the marginal distance rather than the marginal cost of each client on a route. Figure 11 highlights the Scope 3 emissions (transportation and distribution) across the sponsor company's factories.

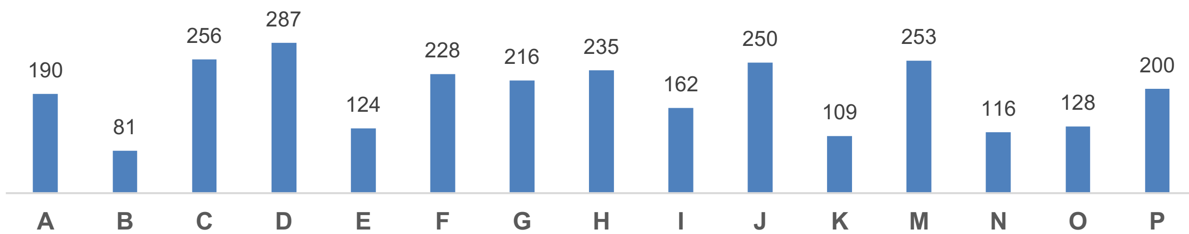


Figure 11 – *Carbon Emissions Across Factories*

Analyzing the Shapley model results provide insights into decision making for cost allocation. Nevertheless, just focusing on Shapley values and geographical distances do not encapsulate the entirety of the cost allocation paradigm. Excluding clients solely based on their low Shapley values or far geographic distances would be an oversimplification. Thus, a more robust approach entails a comparative analysis between the current weight-based proportional method and the Shapley method at the shift level. This approach would enable a detailed understanding of the necessity to employ the Shapley method for specific clients and routes, facilitating a more equitable and precise allocation of costs. To this end, the Kullback-Leibler divergence is used to quantify the divergence between the two allocation methods.

5.2 Divergence Analysis

A Python script was developed to calculate the Kullback-Leibler (KL) divergence for each shift within each factory dataset, thereby facilitating an objective comparison of allocation methods at the shift level. To accomplish this, two types of ratios from each client are required: the *Weight Ratio* and the *Shapley Ratio*. The *Weight Ratio* of a client is defined as the quotient of the client's weight on the shift to the total weight of the shift, representing the proportion of the total shift weight that is allocated to the client. Similarly, the *Shapley Ratio* is computed by dividing the client's Shapley value by the aggregate of Shapley values across the shift, reflecting the client's proportional contribution to the total value assessed via the Shapley method. The summation of these ratios across a shift must equal 1, indicating a distribution among all delivery points. These distributions serve as inputs to compute the KL divergence, which clearly highlights the difference between the allocation distributions from the Shapley method and the proportional method for each shift.

The results of the Python script revealed that the delivery shifts across factories have a modest yet significant divergence between the Shapley method and weight-based proportional method. This is further reinforced through Figure 11, where the computed average KL divergence value across all factories is 0.175.

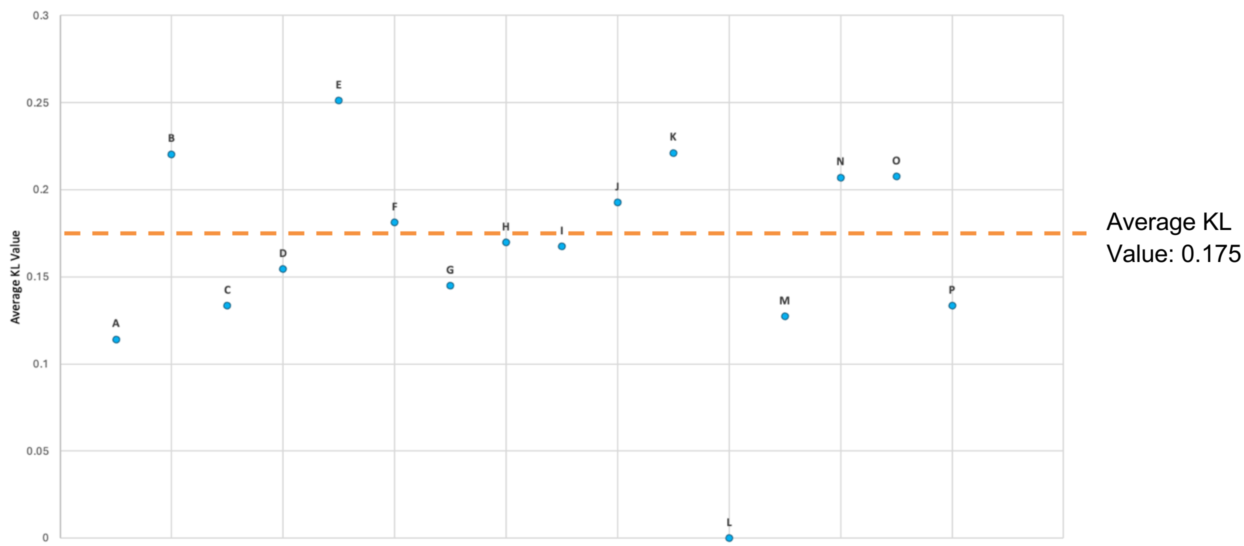


Figure 11 – *Average KL Value for Each Factory Dataset*

The general KL divergence across factory networks exhibits a positive skew. Approximately half of all shifts across the networks demonstrate a divergence of 10% or less when comparing the Shapley and weight-based proportional methods. Figure 12 offers a more detailed visual representation of this distribution. Consequently, a threshold of 10% is established to guide the decision on whether to utilize the Shapley method or not. For cases in which the KL divergence surpasses this threshold, the adoption of the Shapley allocation approach is recommended for the shift. On the other hand, where the KL divergence is below the threshold, the existing weight-based proportional method remains preferable.

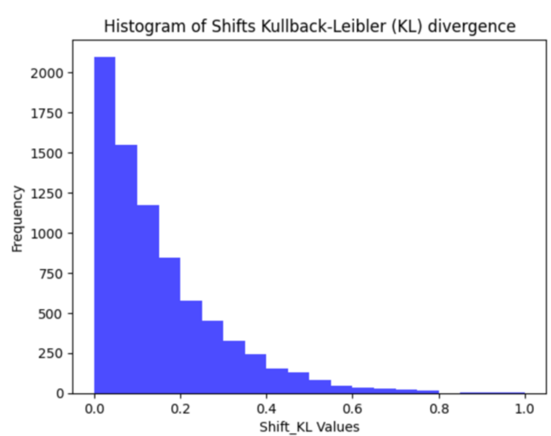


Figure 12 – *Distribution of KL Divergence Across Factory Shifts*

Understanding this divergence is crucial, not only for its criticality in identifying situations where the Shapley method proves beneficial, but also for its practical implications in monetization By applying the insights gained from KL divergence, it is possible to systematically evaluate whether clients are potentially overcharged or undercharged under the current system. This quantitative assessment allows to explore a more equitable cost allocation policy that aligns closer to the actual value provided by the logistics network to each client.

5.3 Monetization

Building on the insights from the KL divergence analysis, the dynamics of monetization within the context of resource allocation are explored. Allocation across distribution networks is inherently representative of a zero-sum game. This asserts that the total amount of resources available for distribution is fixed; thus, any increment in allocation to one client necessarily results in a decrement to another. This finite nature of resources necessitates judicious management to ensure fair and efficient distribution.

Inequitable allocation can inadvertently foster biases in managerial decision-making. Such biases could manifest in favoring certain clients with additional resources to secure their deliveries, potentially at the expense of other clients. For instance, this could lead to scenarios where clients located further away are systematically undercharged, receiving a cost allocation smaller than what is fair, especially for distant customers buying small amount of products. Further, closer clients might find themselves overcharged and essentially subsidizing the gains accorded to the more distant clients.

These biases can distort the overall efficiency and fairness of the allocation system. They undermine the principle of equitable resource distribution based on actual use or contribution, as outlined by the Shapley method, which aims to distribute costs more accurately based on the marginal contribution of each client to the collective cost.

5.3.1 Cost Divergence

To address inefficiencies in the current allocation method, the cost divergence between the Shapley method and the weight-based method is calculated. Understanding the cost divergence is crucial for ascertaining the gains or losses associated with each allocation method, which in turn facilitates the determination of which customers are undercharged or overcharged from an accounting perspective.

To calculate the cost divergence, it is imperative to find the allocation cost of each method. This is initiated by obtaining the Shapley and weight ratios for each client on each route, as done previously when deriving KL divergence. These ratios are indicative of the preference for each allocation method, with a higher Shapley ratio suggesting a greater suitability for Shapley allocation, and vice versa. The allocation costs are then calculated by multiplying the freight cost, obtained from the master data for each client on each route, with their respective allocation ratios.

This procedure results in freight costs for both the allocation methods as shown in *Formulas G* and *H*.

Formula G :

$$\boldsymbol{Cost_{Weight-Based}} = Cost_{Freight} * \frac{Weight_{Client}}{Weight_{Shift}}$$

Formula H :

$$\boldsymbol{Cost_{Shapley}} = Cost_{Freight} * \frac{Shapley\ Value_{Client}}{Shapley\ Value_{Shift}}$$

Subsequently, the cost divergence between the two methods is calculated by subtracting the cost of the Shapley allocation from the cost of the weight-based allocation. The detailed results of the cost divergence for each factory are illustrated in Figure 13. Notably, the highest divergence is observed at the *Factory M*, aligning with its status as the factory with the largest dataset among the factories.

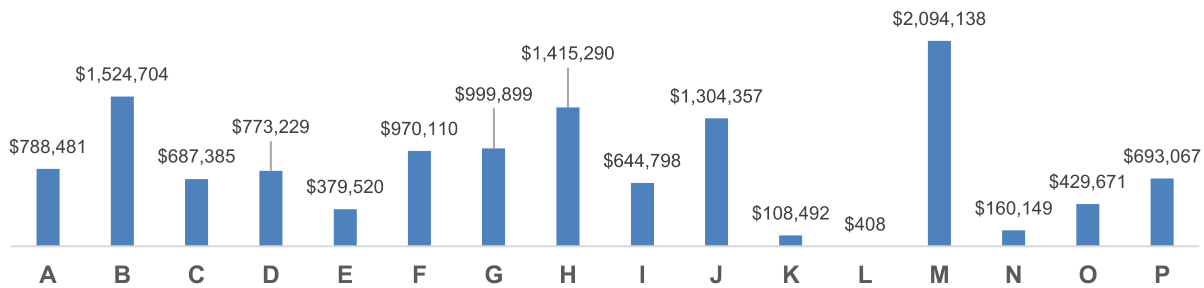


Figure 13 – *Cost Divergence of Allocation Methods Across Factories*

The total cost divergence across all factories amounts to a total monetary value of $12,973,695 over 2 years. This corresponds to $6,486,847.50 per year. This amount represents the potential financial benefit that could be realized if allocation were conducted more equitably, ensuring that customers are charged more accurately.

5.3.2 Client Charges

To further analyze the landscape of how clients are charged, the results from the cost divergence across factories can be utilized. This allows for the assessment of whether clients are charged fairly within each shift. To determine if a client is charged equitably, each client's Shapley ratio is subtracted from their weight ratio. A positive value indicates that the client is overcharged under the weight allocation method, whereas a negative value denotes that the client is undercharged and should be allocated a larger cost under the Shapley method.

To calculate the unfair excess amount paid by overcharged clients, the average cost divergence is divided by the number of overcharged clients within a factory. Similarly, to determine the subsidy enjoyed by undercharged clients, the cost divergence is divided by the number of undercharged clients within a factory, as shown in *Formulas I* and *J*.

Formula I :

$$Excess = \frac{Net\ Cost\ Divergence}{No.\ of\ Overcharged\ Clients}$$

Formula J :

$$Subsidy = \frac{Net\ Cost\ Divergence}{No.\ of\ Undercharged\ Clients}$$

From *Formulas I* and *J*, it is worth noting that while the cost divergence is the same for both the overcharged and undercharged customers, the total amount of overcharged and undercharged customers is not necessarily the same, which leads to varying average excesses and subsidies. Applying this function across the datasets of factories enables the determination of how each customer is charged for every shift. This identifies whether each customer is overcharged or undercharged within a shift, and ultimately what constitutes a fair value to charge them to ensure equitable allocation. Figure 14 displays the excess amounts paid by overcharged clients and the subsidies enjoyed by undercharged clients across factories. Understanding these charges not only facilitates the determination of how to fairly assess the cost to serve each client across shifts but also identifies opportunities to gain additional profits.

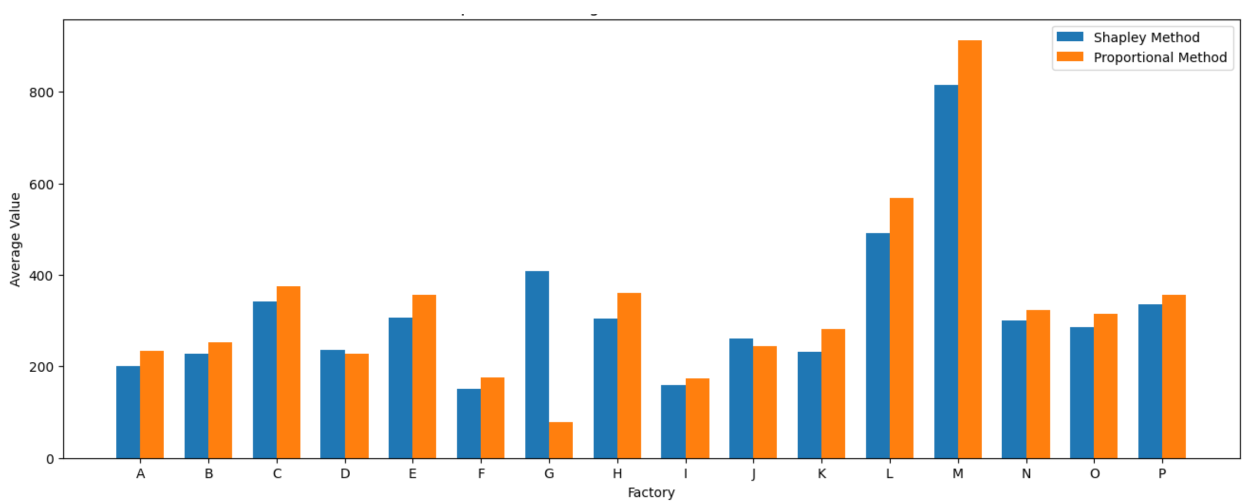


Figure 14 – *Average Client Excess and Undercharge Trends Across Factories Over Time: Proportional vs. Shapley Methods*

By combining the Shapley values, cost divergence, and client charges, a more equitable allocation, fairer charges, and increased profitability can be achieved. In addition, to further enhance this framework and ultimately craft a robust allocation policy, a proposal of a machine learning model has been developed to predict the propensity of whether a client should follow a Shapley or proportional weight-based allocation.

6. Machine Learning

In the previous section, the delivery network and orders were thoroughly characterized. This section focuses on developing a machine learning model to predict whether a new order from a known client should follow either the proportional method or the Shapley method for cost allocation. If an order is predicted to follow the proportional method, the volume or weight quoted can be used as a proxy for the allocation. However, this does not apply if the proportional method varies excessively compared to the Shapley method. Accurate forecasting is crucial due to the significant impact of cost allocation on profitability. By correctly charging for costs, the sponsor company ensures that all clients bear the true logistics expenses incurred in serving them.

6.1 Feature Selection and Engineering

At the time of quotation, the sponsor does not necessarily know the specific shipment that will deliver the order, meaning there is no information on the number of customers or their respective locations. As a result, the features available for a machine learning model are limited. The known features at the time of quoting an operation include the network where the customer is located, the distance from the shipment to the delivery point, the date, and the amount and quantity of the product being quoted.

Besides these available features, a centrality feature was proposed to improve model performance. The rationale behind this feature is that if a given client X has many other clients nearby, these nearby clients are likely to be served on a future route. For the machine learning model, centrality is defined as the average distance from client X to the Z closest historic customers. The centrality of each customer within each network was calculated using different numbers of neighboring customers Z: 25, 50, 75, and 100.

Additionally, factories are recognized as a crucial feature for the model, helping to account for varying parameters among networks. One-hot encoding was applied to represent each factory as a numerical feature.

Conclusively, the features used for the machine learning model are the operational pounds (which represent the total weight of the products being quoted for the order), total products, distance from the shipment point to the client's delivery location, a centrality measure for the Z nearest customers within the same delivery network, and factories encoded as a numerical value.

6.2 Machine Learning Model

To classify and predict whether an order should follow the Shapley method or the proportional method, a logistic regression was developed. Logistic regression, commonly used for binary classification tasks, estimates the probability that a given observation belongs to a particular class. In this case, the logistic regression model aims to determine the most suitable allocation method for each order.

However, to utilize the logistic regression model effectively, the data must first be labelled. This labelling enables the model to train and learn, and eventually make predictions on which allocation method an order should follow. Two different approaches for labelling the data were considered. The first is based on the KL method, which measures the similarity between the proportional and Shapley methods. The second approach is based on the difference in monetary value between both allocations methods. These two labelling approaches were explored to identify the optimal labelling strategy that would yield a balanced dataset and high-accuracy model.

6.2.1 Labelling Based on KL Values

Under the KL approach, the logistic regression model leverages the KL divergence value of each shift for labelling data. If the KL value exceeds a certain threshold, the shift is classified to follow the Shapley allocation, whereas if it falls below that limit, it is considered to follow the proportional method. A lower KL divergence means that the two methods are similar, and therefore the proportional method should suffice for prediction. When KL is larger, the methods differ more significantly. However, defining a threshold is not straightforward, and multiple values for this KL value threshold were tested to calibrate the model.

As highlighted in Figure 15, various KL divergence thresholds were systematically evaluated to determine the most appropriate classification boundary. By varying these thresholds, the ratio of Shapley allocations to proportional method classifications fluctuates across shifts, which helps enable the identification of the most suitable model for cost allocation.

KL Divergence Threshold	% of Orders that follow Proportional Method	% of Orders that follow Shapley Method
5%	0.19	0.81
7.5%	0.28	0.72
10%	0.37	0.63
12.5%	0.44	0.56
15%	0.51	0.49

Figure 15 – *Percentage of Allocation Distribution for Various KL Thresholds*

It is evident from Figure 15 that a low threshold has significantly high data imbalance between both allocation ratios. Unbalanced data can negatively affect the performance of logistic regression because the model may become biased towards the majority class (Japkowicz, 2000). This denotes that the model might predict the Shapley method for most instances, regardless of the actual distribution of orders that follow the proportional method. Consequently, this results in poor classification performance, particularly for the proportional method class, and favors classifying orders as following the Shapley method.

6.2.2 Labelling Based on Monetary Difference

An alternative method for labeling the data is to consider the monetary value difference between the Shapley and proportional methods. This approach subtracts the cost of the Shapley method from that of the proportional method for each order. If the number is negative, the order is classified to follow the Shapley allocation; if the monetary difference is positive, the order is classified to follow the proportional method.

Using this method allows the identification of customers who are under-allocated based on the proportional method so that appropriate measures can be taken. Conversely, if the Shapley method indicates that a customer should pay more, those operations will be labeled to follow the Shapley allocation, and proportional method if not. By using this approach, the dataset becomes more balanced.

6.3 Model Outcomes and Testing

To evaluate the model's effectiveness, various scenarios were defined using different feature sets. Model performance was assessed using the Area Under the Receiver Operating Characteristic Curve (AUC) and the distribution of orders between the Proportional and Shapley methods. The AUC is a performance metric used to evaluate the predictive ability of binary classification models, including logistic regression. An AUC score of 0.5 indicates that the model is no better than random guessing, while a score of 1.0 represents a perfect classifier.

The logistic regression model was run for both the KL labelling approach and the monetary difference labelling approach, while using the features selected in Section 6.1. The results from both approaches are compared to assess model performance.

KL Divergence Threshold	Number of Z customers for Centrality	AUC	% of Orders that follow Proportional Method	% of Orders that follow Shapley Method
5%	Granular (25, 50, 75, 100)	0.56	0.2	0.8
5%	100	0.52	0.2	0.8
7.5%	Granular (25, 50, 75, 100)	0.56	0.29	0.71
7.5%	100	0.49	0.29	0.71
10%	Granular (25, 50, 75, 100)	0.55	0.38	0.62
10%	100	0.58	0.38	0.62
12.5%	Granular (25, 50, 75, 100)	0.56	0.45	0.55
12.5%	100	0.56	0.45	0.55
15%	Granular (25, 50, 75, 100)	0.57	0.52	0.48
15%	100	0.57	0.52	0.48

Figure 16 – *Results from KL Labelling Approach*

Number of Z customers for Centrality	AUC	% of Orders that follow Proportional Method	% of Orders that follow Shapley Method
100	0.79	0.47	0.53

Figure 17 – *Results from Monetary Difference Labelling Approach*

Figures 16 and 17 clearly illustrate that the Monetary Difference method results in a more balanced classification distribution between the proportional and Shapley methods compared to the KL method. Maintaining a balanced dataset is crucial for logistic regression models, as imbalanced data can significantly distort predictions and undermine overall accuracy. Furthermore, employing the Monetary Difference method substantially enhanced the AUC, or predictive capability, of the model to 0.79. In addition to improved predictive performance, the Monetary Difference method offers the advantage of simplicity by providing a clear understanding of the monetary value, as opposed to the KL value. Given these considerations, it is evident that the Monetary Difference method is better suited for policy formulation.

7. Allocation Policy

By using the machine learning model, an allocation policy could be developed. First, a logistic regression model determines whether an order should follow the proportional or Shapley allocation method. If the order of a known client is classified to follow the proportional weight-based allocation method, the existing allocation practices of the sponsor company should be maintained. However, if the order is classified to follow the Shapley allocation method, the weight-based allocation method must still be used, yet a surcharge to cover the true estimated cost of serving a given client needs to be charged over this allocation. Figure 18 outlines the overall allocation policy process.

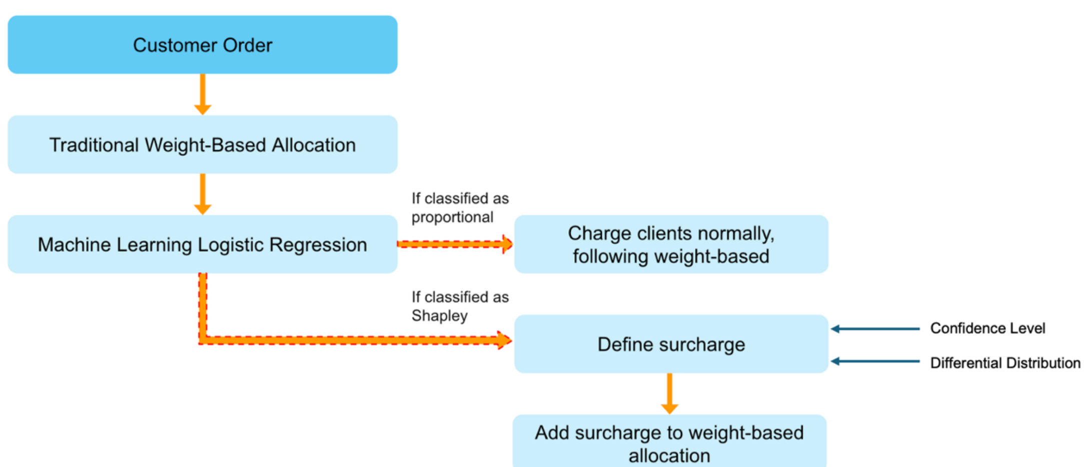


Figure 18 – *Flow for Allocation Policy*

To calculate the surcharge, a straightforward approach is proposed. Given that the surcharge is calculated for orders predicted to follow the Shapley method, the policy focuses exclusively on customers for whom this method is applicable. By examining historical deliveries that exhibit a divergence between the two methods, where the Shapley method should have yielded a higher value than the proportional method, it is possible to construct a probability distribution for the divergence value for each factory. This distribution is then added to the proportional method.

Figure 19 presents a representative distribution of the difference between methods for cases where the Shapley method should have been employed at Factory F. It is important to note that the divergence is represented as a negative value, as it denotes the difference between the two methods, considering only cases where the Shapley method resulted in a higher value. From this analysis, it can be inferred that, in most cases, the overcharge for the proportional method should have been less than $200 to align with the Shapley method.

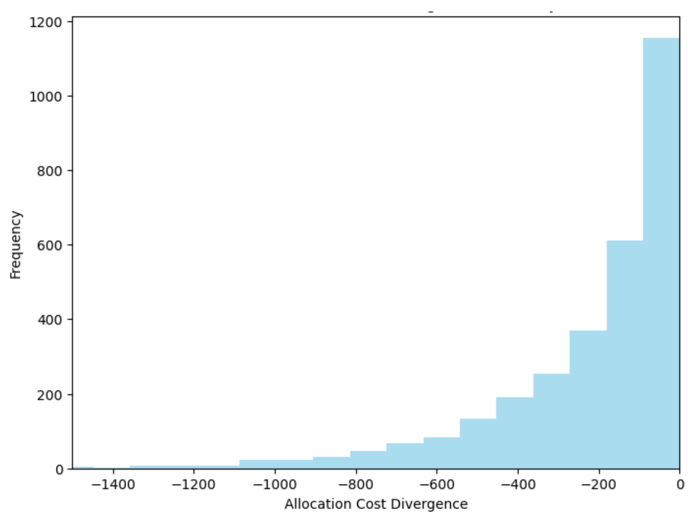


Figure 19 – *Distribution of Monetary Differences for Cases Favoring Shapley Method (Fact. F)*

The company should define a confidence level and use it with the previously defined cumulative distribution function (CDF) for the differential between both allocation policies for each network. Defining a confidence level ensures that the actual costs of serving a customer are covered. Furthermore, the cumulative distribution function can be adjusted conditionally based on distance, the number of products being quoted, or a combination of these factors. Ultimately, the policy can include these variables as well.

8. Conclusion

A comprehensive literature review of allocation methods was conducted. As a result, the Shapley method is used to allocate costs fairly among clients, as it accounts for the marginal contribution of each client to the total cost on each route. Each order on a given route is assessed and labeled based on the monetary difference between the proportional method and the Shapley method. Significant divergences between the two methods have been observed, both in relative percentages and monetary value.

The analysis estimated a difference of approximately $13 million between the two methods across the network over time, affecting half of the customers with a divergence of 10% or greater. Although transforming this difference directly into profit is challenging, given that allocation is a zero-sum game, it is possible if customers undercharged by the proportional method can be identified before finalizing business deals and are charged accordingly.

A logistic regression model was employed to predict whether a future client would be best allocated using the proportional method or the Shapley method. Once the allocation policy is established, implementing the proportional method is straightforward, whereas the Shapley method requires further analysis. For orders of a client that are predicted to follow the Shapley allocation, the proposed approach involves considering the historical distribution of differences between the methods where the proportional method fell short and setting a confidence level to ensure that a defined percentage of cases is covered. This value is then surcharged to the current weight-based policy that the company follows.

By only modifying the policy for the customers that are predicted to follow the Shapley allocation, the company effectively builds a safety budget, ensuring that this policy cannot underperform compared to the traditional method. Furthermore, it partially captures the potential profit mentioned previously if the business context allows for establishing the surcharges.

Appendices

Appendix A - Literature Review Database

Allocation Method	Author	Year	Title	Problem Statement	Allocation Strategy Focus
Shapley	Dror	1990	Cost allocation: the traveling salesman, bin packing, and the knapsack.	Traveling Salesman	Shapley values for the *knapsack* problem.
Shapley	Engevall et al.	1998	The traveling salesman game: an application of cost allocation in a gas and oil company.	Traveling Salesman	The core concept and Shapley values.
Shapley	Engevall et al.	2004	The heterogeneous vehicle-routing game.	Traveling Salesman, Vehicle Routing	The core concept with constraint generation.
Shapley	Wong et al.	2007	Cost allocation in spare parts inventory pooling.	Inventory Systems	The core concept.
Shapley	Krajewska et al.	2008	Horizontal cooperation among freight carriers:request allocation and profit sharing.	Vehicle Routing	Shapley values.
Shapley	Agarwal and Ergun	2010	Network Design and Allocation Mechanisms for Carrier Alliances in Liner Shipping	Horizontal Cooperatio n	Shapley values.
Shapley	Crujjssen et a l.	2010	Supplier-initiated outsourcing: a methodology to exploit synergy in transportation.	Vehicle Routing	Introduced the Shapley Monotonic Path (SMP) with Shapley values.
Shapley	Frisk et al.	2010	Cost allocation in collaborative forest transportation.	Transporta tion Planning	Shapley values compared to the *Equal Profit Method.*
Shapley	Liu et al.	2010	Allocating collaborative profit in less-than-truckload carrier alliance.	Vehicle Routing	Shapley values compared to the *Weighted Relative Savings Model.*
Shapley	Massol and Tchung-	2010	Cooperation among liquefied natural gas	Transporta tion	Shapley values.

	Ming		suppliers: is rationalization the sole objective.	Planning	
Shapley	Granot et al.	2011	On Chinese postman games where residents of each road pay the cost of their road	Horizontal Cooperation	The core concept and considers the traditional nucleolus procedure.
Shapley	Fiestras-Janeior et al.	2012	Cost allocation in inventory transportation systems.	Inventory Systems	Introduced the *Line Rule* (Shapley values with less computational effort).
Shapley	Yengin	2012	Characterizing the Shapley value in fixed-route traveling salesman problems with appointments.	Traveling Salesman	Shapley values.
Shapley	Lozano et al.	2013	Cooperative game theory approach to allocating benefits of horizontal cooperation.	Horizontal Cooperation	Shapley values.
Shapley	Özener et al.	2014	Developing a collaborative planning framework for sustainable transportation.	Vehicle Routing	Modified Shapley values, by creating approximation method.
Shapley	Fang and Cho	2014	Stability and endogenous formation of inventory transshipment networks.	Inventory Systems	Shapley values (in the form of the *MJW* value).
Shapley	Vanovermeire et al.	2014	Integration of the cost allocation in the optimization of collaborative bundling.	Horizontal Cooperation	Shapley values.
Shapley	Hezarkhani et al.	2015	A competitive solution for cooperative truckload delivery.	Horizontal Cooperation	Shapley values.
Shapley	Wang et al.	2015	A methodology to exploit profit allocation in logistics joint distribution network optimization.	Transportation Planning	Shapley values.
Shapley	Zakharov and Shchegryaev	2015	Stable cooperation in dynamic vehicle routing problems.	Vehicle Routing	Shapley values.
Shapley	Kimms and	2016	Shapley value-based cost	Traveling	Shapley values to develop a

	Kozeletskyi		allocation in the cooperative traveling salesman problem under rolling horizon planning	Salesman	model with lower NP-completeness.
Shapley	Wu et al.	2017	Using Shapley value for city bus route scheduling	Transportation Planning	Shapley values.
Shapley	Leenders et al.	2017	Emissions allocation in transportation routes	Vehicle Routing	Shapley values.
Shapley	Vos and Raa	2018	Stability Analysis of Cost Allocation Methods for Inventory Routing	Inventory Systems	Shapley values.
Shapley	Otero-Palencia et al.	2018	A stochastic joint replenishment problem considering transportation and warehouse constraints with gainsharing by Shapley Value allocation	Horizontal Cooperation	Shapley value function.
Shapley	Schulte et al.	2019	Scalable Core and Shapley Value Allocation Methods for Collaborative Transportation	Horizontal Cooperation	Shapley values to develop a model with lower NP-completeness.
Shapley	Zheng et al.	2019	Coordinating a closed-loop supply chain with fairness concerns through variable-weighted Shapley values	Horizontal Cooperation	Introduced a *variable-weighted* Shapley value mechanism.
Shapley	Levinger et al.	2020	Computing the Shapley Value for Ride-Sharing and Routing Games	Traveling Salesman	Shapley values.
Shapley	Gnecco et al.	2020	Public transport transfers assessment via transferable utility games and Shapley value approximation	Transportation Planning	Shapley values.
Shapley	Wang	2023	A collaborative approach based on Shapley value for carriers in the supply chain distribution	Horizontal Cooperation	Shapley values.
Nucleolus	Göthe-Lundgren et al.	1996	On the nucleolus of the basic vehicle routing game.	Vehicle Routing	Traditional nucleolus procedure with constraint generation.

Nucleolus	Engevall et al.	1998	The traveling salesman game: an application of cost allocation in a gas and oil company.	Traveling Salesman	Introduced the *Demand Nucleolus* (where excess is multiplied with total demand of the coalition).
Nucleolus	Sakawa et al.	2001	Fuzzy programming and profit and cost allocation for a production and transportation problem.	Transportation Planning	Traditional nucleolus procedure.
Nucleolus	Engevall et al.	2004	The heterogeneous vehicle-routing game.	Traveling Salesman, Vehicle Routing	Traditional nucleolus procedure with constraint generation.
Nucleolus	Agarwal and Ergun	2010	Network Design and Allocation Mechanisms for Carrier Alliances in Liner Shipping	Horizontal Cooperation	Traditional nucleolus procedure.

Nucleolus	Frisk et al.	2010	Cost allocation in collaborative forest transportation.	Transportation Planning	Tradtional nucleolus procedure compared to the *Equal Profit Method.*
Nucleolus	Liu et al.	2010	Allocating collaborative profit in less-than-truckload carrier alliance.	Vehicle Routing	Traditional nucleolus procedure compared to the *Weighted Relative Savings Model.*
Nucleolus	Massol and Tchung-Ming	2010	Cooperation among liquefied natural gas suppliers: is rationalization the sole objective.	Transportation Planning	Traditional nucleolous procedure, *"Per Capita"* Nucleolus, and the D*isruption Nucleolus.*
Nucleolus	Granot et al.	2011	On Chinese postman games where residents of each road pay the cost of their road	Horizontal Cooperation	Traditional nucleolus procedure and considers the core concept.
Nucleolus	Lozano et al.	2013	Cooperative game theory approach to allocating benefits of horizontal cooperation.	Horizontal Cooperation	Traditional nucleolus procedure.
Nucleolus	Hezarkhani et al.	2015	A competitive solution for cooperative truckload delivery.	Horizontal Cooperation	Traditional nucleolus procedure.
Nucleolus	Leenders et al.	2017	Emissions allocation in	Vehicle	Traditional nucleolus

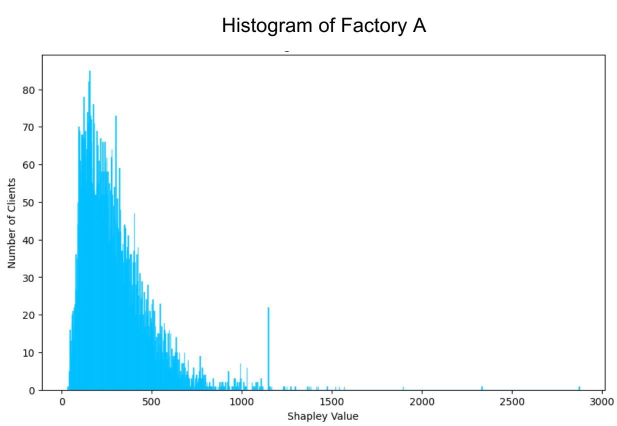
			transportation routes	Routing	procedure.
Nucleolus	Vos and Raa	2018	Stability Analysis of Cost Allocation Methods for Inventory Routing	Inventory Systems	Traditional nucleolus procedure.
Proportional	Fishburn and Pollak	1983	Fixed-route cost allocation	Traveling Salesman	Proportional to willingness to pay (WTP) scheme.
Proportional	Dror	1990	Cost allocation: the traveling salesman, binpacking, and the knapsack.	Traveling Salesman	Egalitarian method.
Proportional	Engevall et al.	2004	The heterogeneous vehicle-routing game.	Traveling Salesman, Vehicle Routing	Proportional to demand scheme.
Proportional	Wong et al.	2007	Cost allocation in spare parts inventory pooling.	Inventory Systems	Proportional to demand scheme.
Proportional	Özener and Ergun	2008	Allocating costs in a collaborative transportation procurement network.	Horizontal Cooperation	Developed a cross-monotonic and stable allocation method.
Proportional	Frisk et al.	2010	Cost allocation in collaborative forest transportation.	Transportation Planning	Developed the *Equal Profit Method* (EPM) based on stable allocation theory.
Proportional	Liu et al.	2010	Allocating collaborative profit in less-than-truckload carrier alliance.	Vehicle Routing	Egalitarian method.
Proportional	Massol and Tchung-Ming	2010	Cooperation among liquefied natural gas suppliers: is rationalization the sole objective?	Transportation Planning	Egalitarian method, proportional to non-cooperative profits, and proportional to shipments schemes.
Proportional	Özener et al.	2013	Allocating cost of service to customers in inventory routing.	Inventory Systems	Proportional to several factors (distance, capacity, consumption) scheme.
Proportional	Nguyen et al.	2014	Consolidation strategies for the delivery of perishable products.	Horizontal Cooperation	Proportional to demand scheme.

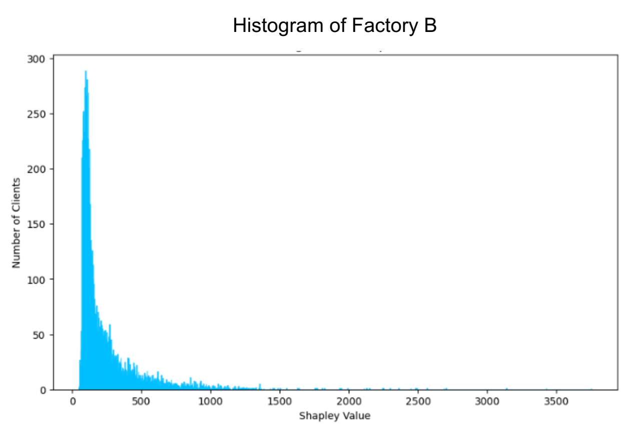
Proportional	Özener	2014	Developing a collaborative planning framework for sustainable transportation.	Vehicle Routing	Proportional to distance scheme.
Proportional	Flisberg et al.	2015	Potential savings and cost allocations for forest fuel transportation in Sweden: a country-wide study.	Transportation Planning	Proportional to volume scheme.
Proportional	Hezarkhani et al.	2015	A competitive solution for cooperative truckload delivery.	Horizontal Cooperation	Proportional to stand-alone costs of minimal essential deliveries.
Proportional	Algaba et al.	2018	Horizontal cooperation in a multimodal public transport system: The profit allocation problem	Horizontal Cooperation	Developed the *Coloured Egalitarian Solution* and *Coloured Cost Proportional Solution.*
Proportional	Bogachev et al.	2021	Comparative Analysis of the Use Egalitarian and Utilitarian Approaches in the Freight Transportation Optimization Problem	Transportation Planning	Egalitarian method.
Proportional	Estañ et al.	2021	On how to allocate the fixed cost of transport systems	Transportation Planning	Egalitarian method.

Appendix B – Data Filtration for Computational Optimization

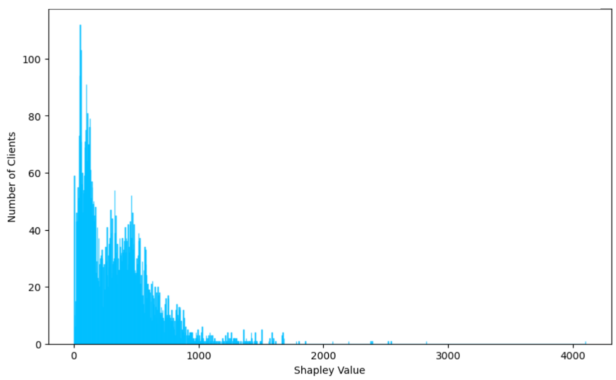
Factory	Number of Rows	% of Total Rows	Rows Filtered Out	% of Dataset Filtered Out
A	10,972	7.88%	0	0.00%
B	14,987	10.76%	187	0.71%
C	9,479	6.80%	12	0.13%
D	7,001	5.03%	0	0.00%
E	4,806	3.45%	0	0.00%
F	9,664	6.94%	0	0.00%
G	11,792	8.46%	0	0.00%
H	13,893	9.97%	624	4.49%
I	9,190	6.60%	49	0.53%
J	8,439	6.06%	331	3.92%
K	2,109	1.51%	0	0.00%
L	2	0.00%	0	0.00%
M	19,436	13.95%	0	0.00%
N	2,823	2.03%	0	0.00%
O	5,337	3.83%	0	0.00%
P	9,376	6.73%	0	0.00%

Appendix C – Shapley Distribution Across Factories

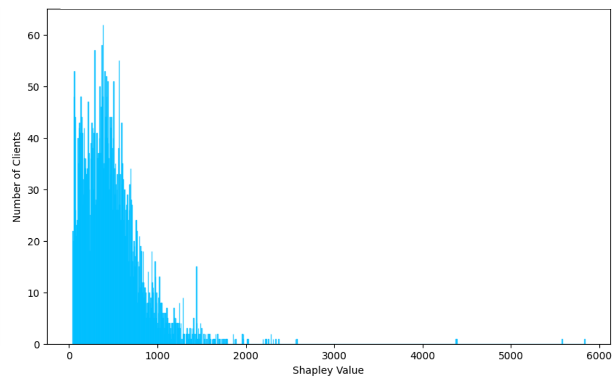




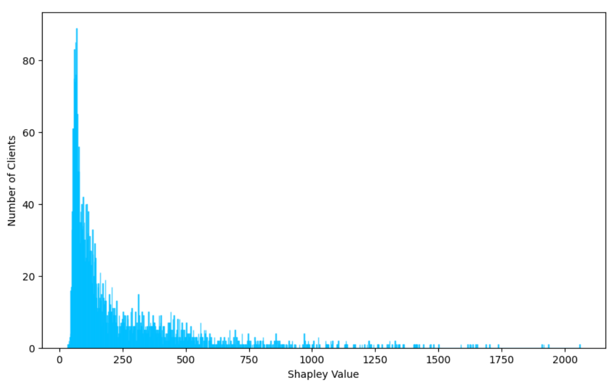
Histogram of Factory C



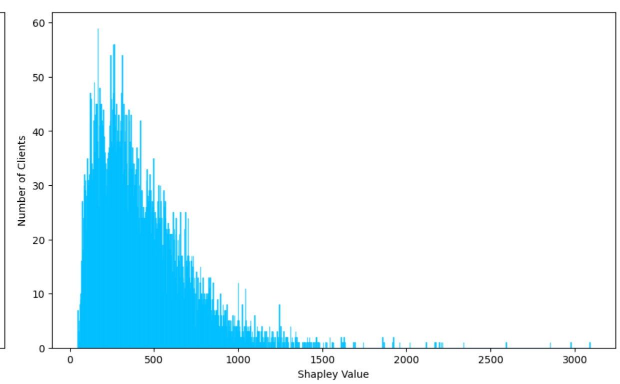
Histogram of Factory D



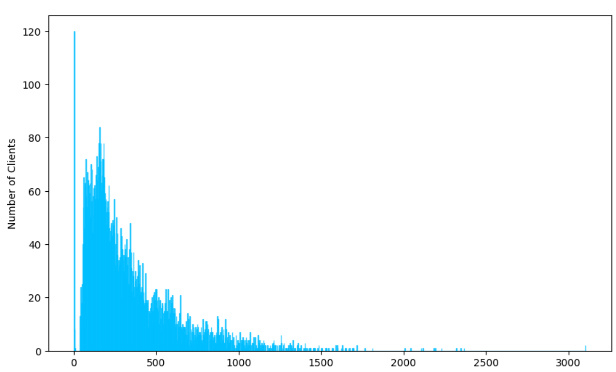
Histogram of Factory E



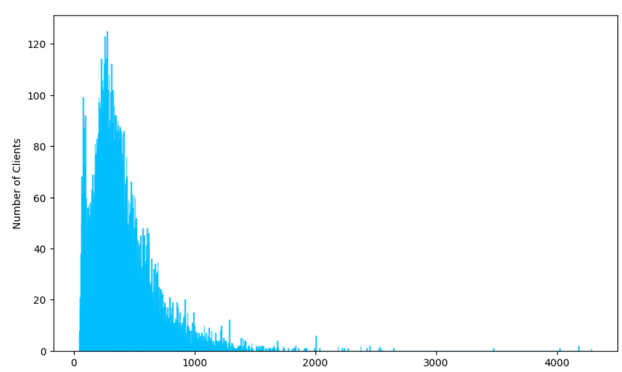
Histogram of Factory F

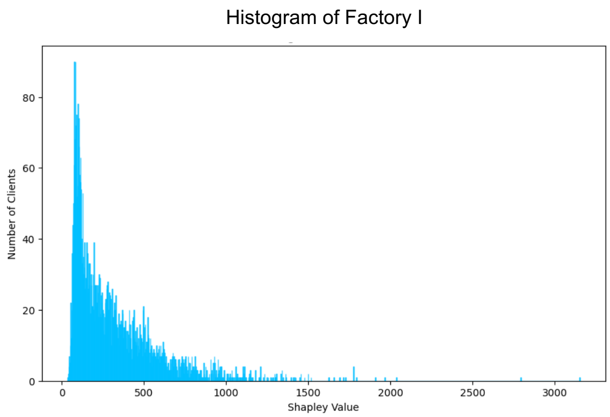


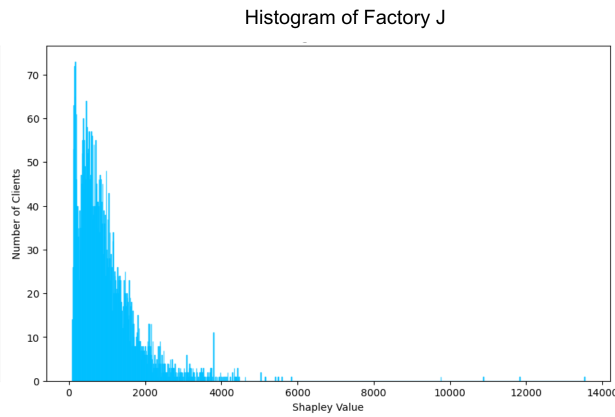
Histogram of Factory G

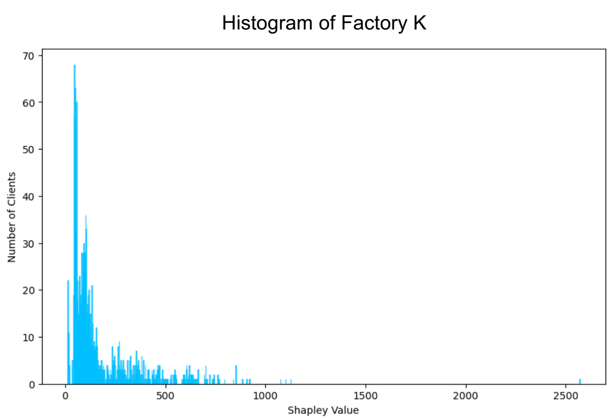


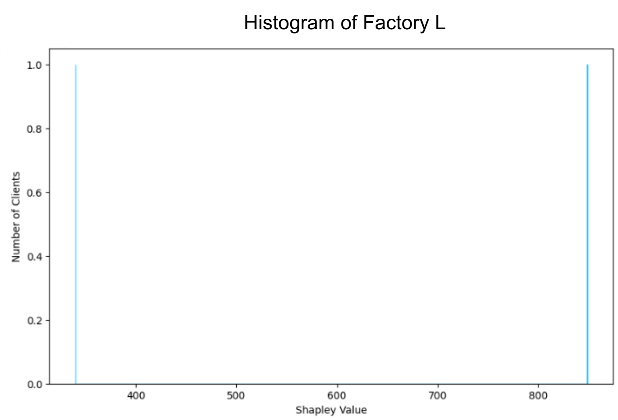
Histogram of Factory H

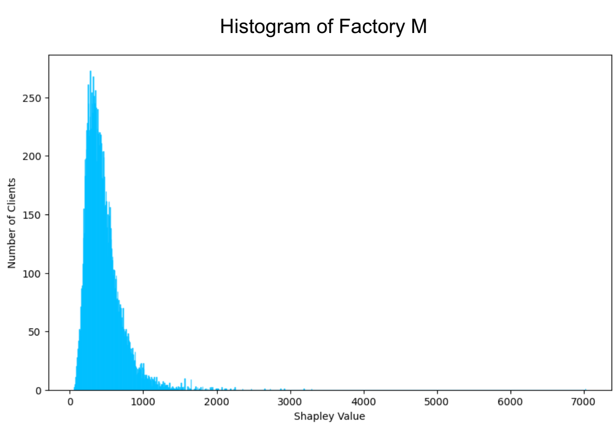


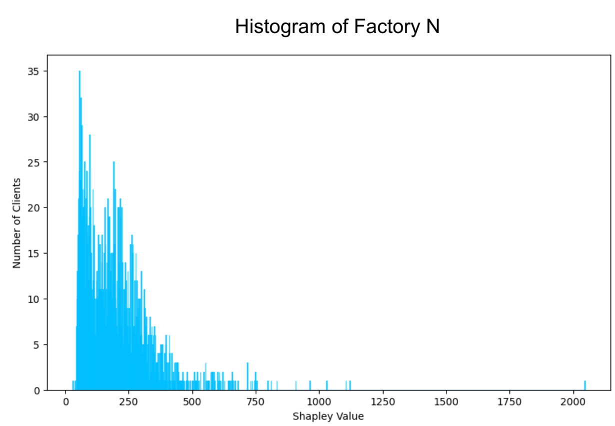

Histogram of Factory I
Number of Clients
Shapley Value

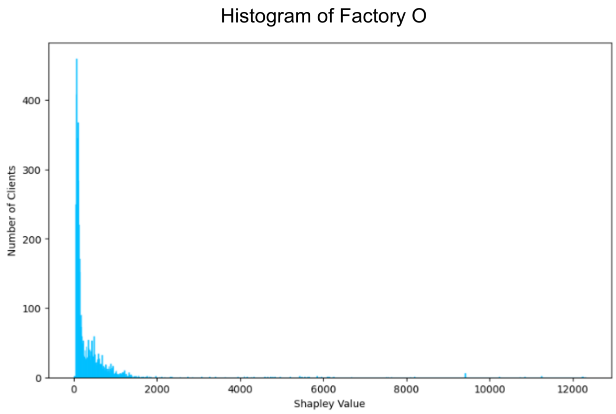

Histogram of Factory J
Number of Clients
Shapley Value

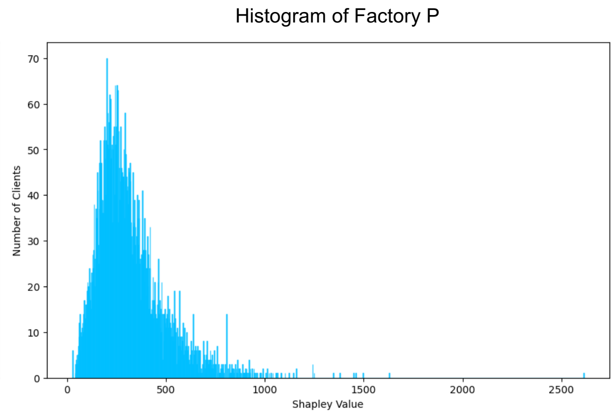

Histogram of Factory K
Number of Clients
Shapley Value


Histogram of Factory L
Number of Clients
Shapley Value

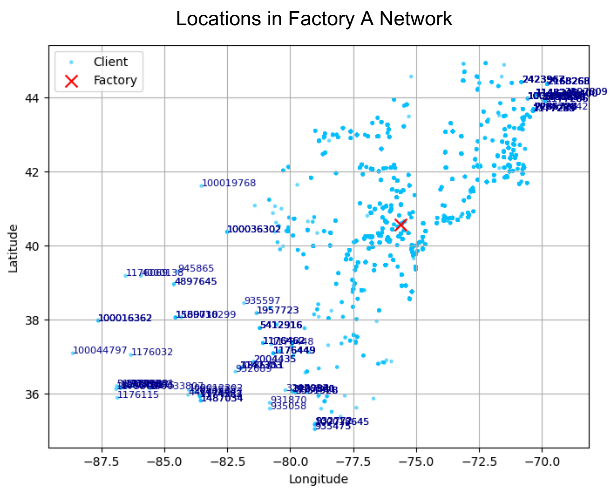

Histogram of Factory M
Number of Clients
Shapley Value

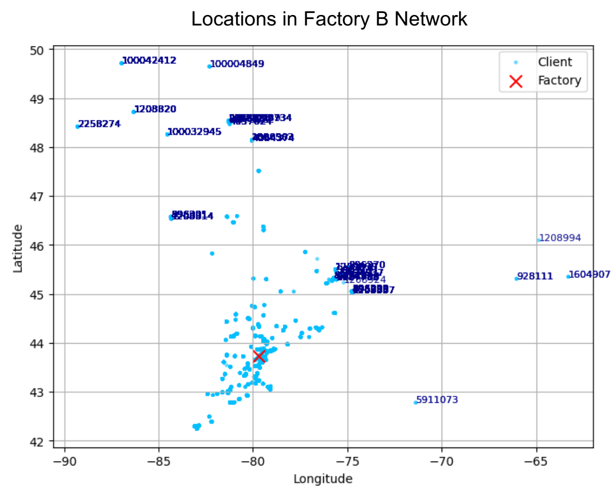

Histogram of Factory N
Number of Clients
Shapley Value

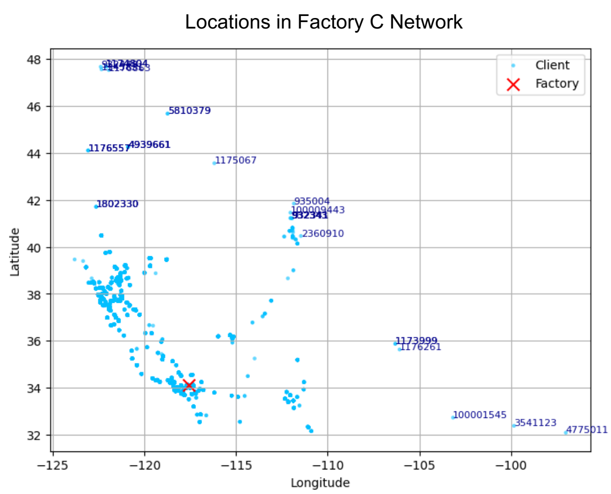


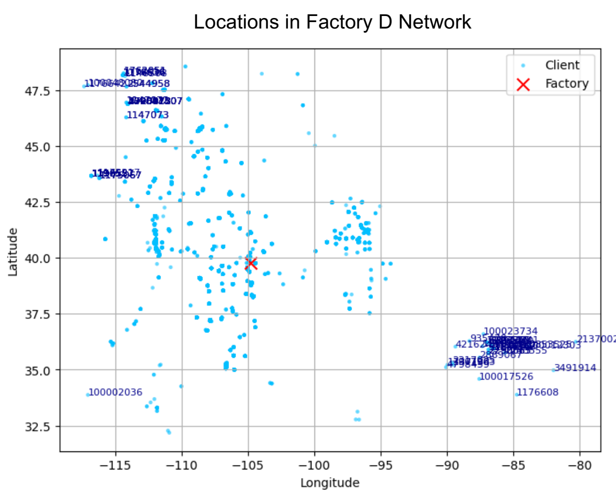


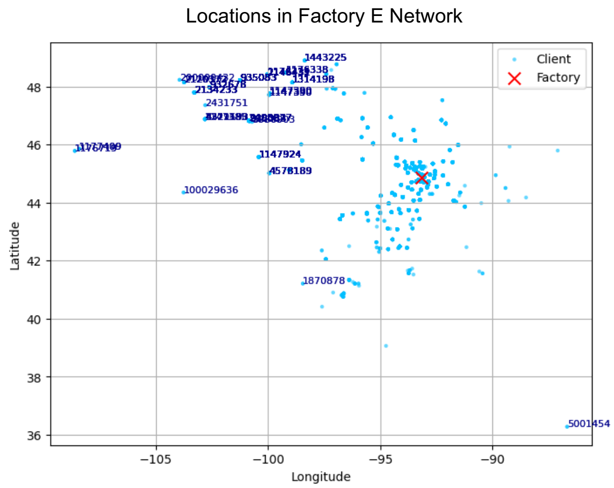
Appendix D – Geographical Visualization of Locations Across Factories

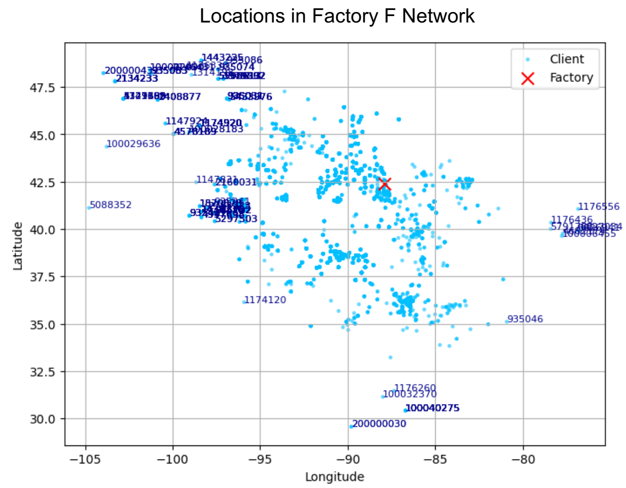


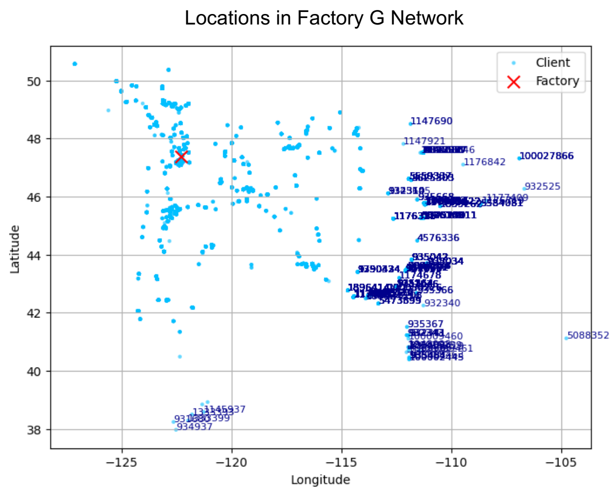


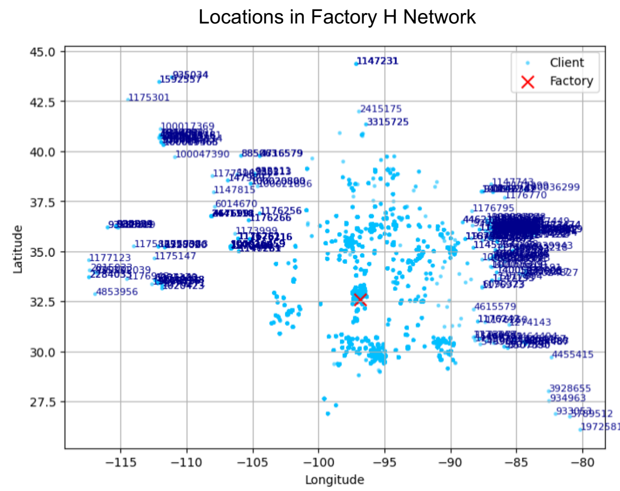




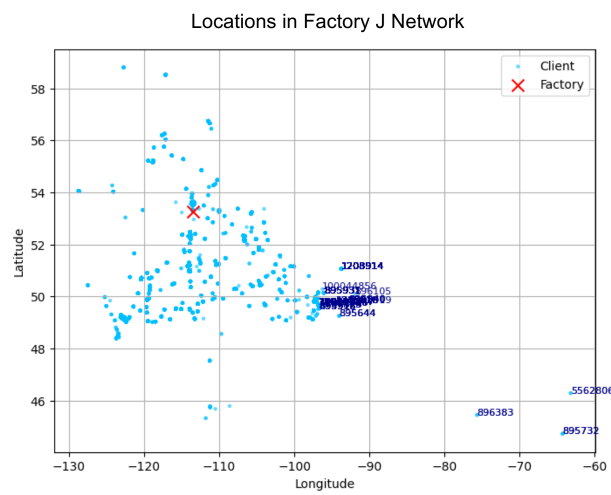

Locations in Factory E Network
Client
Factory
Longitude
Latitude

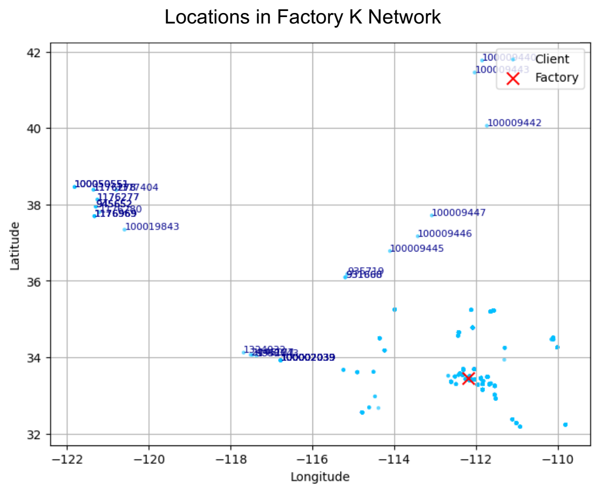

Locations in Factory F Network
Client
Factory
Longitude
Latitude

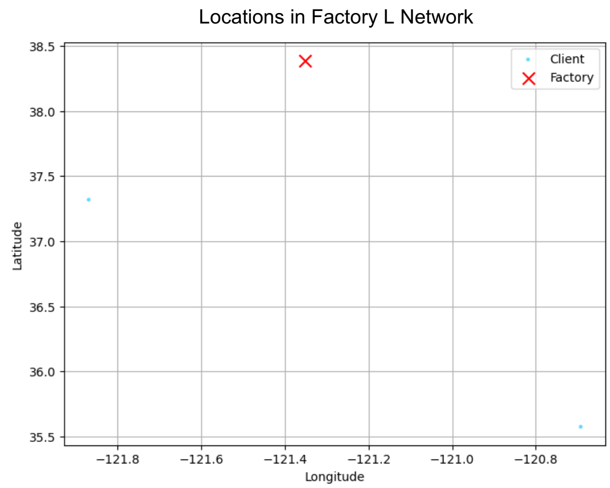

Locations in Factory G Network
Client
Factory
Longitude
Latitude

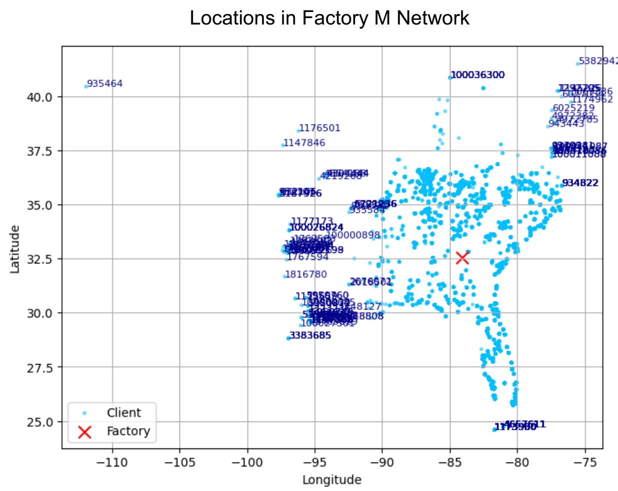

Locations in Factory H Network
Client
Factory
Longitude
Latitude

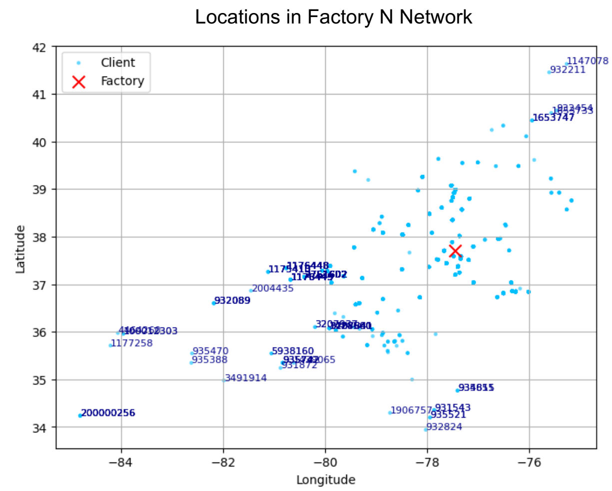

Locations in Factory I Network
Client
Factory
Longitude
Latitude

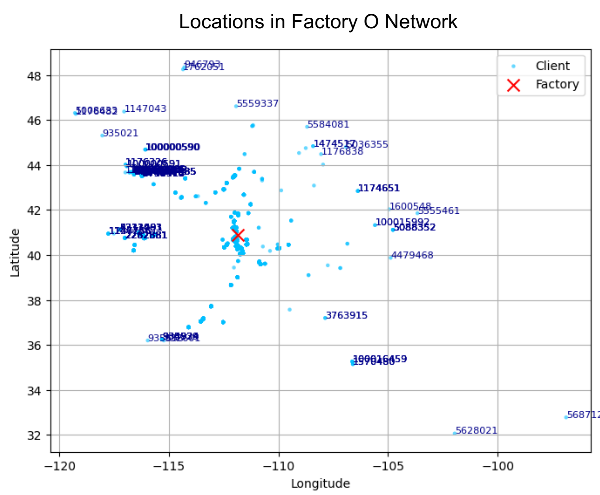

Locations in Factory J Network
Client
Factory
Longitude
Latitude

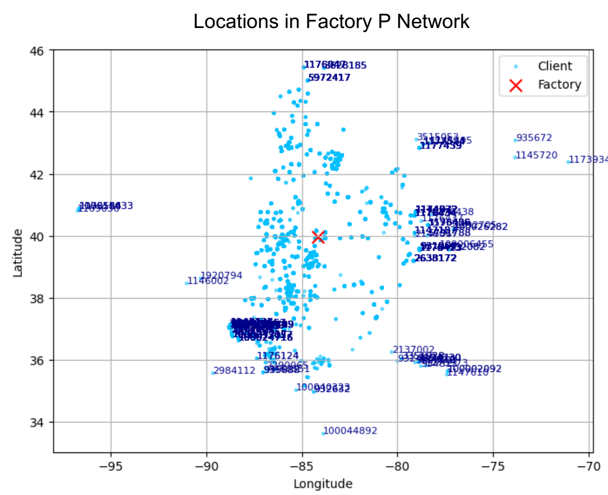

Locations in Factory K Network
Client
Factory
Longitude
Latitude


Locations in Factory L Network
Client
Factory
Longitude
Latitude


Locations in Factory M Network
Client
Factory
Longitude
Latitude


Locations in Factory N Network
Client
Factory
Longitude
Latitude


Locations in Factory O Network
Client
Factory
Longitude
Latitude


Locations in Factory P Network
Client
Factory
Longitude
Latitude

References

Agarwal, R., Ergun, Ö., 2010. Network design and allocation mechanisms for carrier alliances in liner shipping. *Operations Research* 58, 6, 1726–1742.

Algaba, E., Fragnelli, V., Llorca, N., Sánchez-Soriano, J., 2019. Horizontal cooperation in a multimodal public transport system: The profit allocation problem. *European Journal of Operational Research* 275, 2, 659-665.

Bogachev, V., Kravets, A., Bogachev, T., 2021, Comparative Analysis of the Use Egalitarian and Utilitarian Approaches in the Freight Transportation Optimization Problem. *International Conference on Industrial Engineering, Applications and Manufacturing.* DOI 10.1109/ICIEAM51226.2021.9446374.

Cruijssen, F., Borm, P., Fleuren, H., Hamers, H., 2010. Supplier-initiated outsourcing: a methodology to exploit synergy in transportation. *European Journal of Operational Research* 207, 2, 763–774.

Dahlberg, J., Engevall, S., Göthe-Lundgren, M., Jörnsten, K., Rönnqvist, M., 2018. Incitements for transportation collaboration by cost allocation. *Central European Journal of Operations Research* 27, 1009–1032.

Dror, M., 1990. Cost allocation: the traveling salesman, binpacking, and the knapsack. *Applied Mathematics and Computation* 35, 2, 191–207.

Engevall, S., Göthe-Lundgren, M., Värbrand, P., 1998. The traveling salesman game: an application of cost allocation in a gas and oil company. *Annals of Operations Research* 82, 203–218.

Engevall, S., Göthe-Lundgren, M., Värbrand, P., 2004. The heterogeneous vehicle-routing game. *Transportation Science* 38, 1, 71–85.

Estañ, T., Llorca, N., Martínez, R., Sánchez-Soriano, J., 2021. On how to allocate the fixed cost of transport systems. *Annals of Operations Research* 301, 81-105.

Fang, X., Cho, S.-H., 2014. Stability and endogenous formation of inventory transshipment networks. *Operations Research* 62, 6, 1316–1334.

Fiestras-Janeiro, M.G., Garc´ıa-Jurado, I., Meca, A., Mosquera, M.A., 2012. Cost allocation in inventory transportation systems. *Top* 20, 2, 397–410.

Fishburn, P., Pollak, H., 1983. Fixed-route cost allocation. *American Mathematical Monthly* 90, 6, 366–378.

Flisberg, P., Frisk, M., Rönnqvist, M., Guajardo, M., 2015. Potential savings and cost allocations for forest fuel transportation in Sweden: a country-wide study. *Energy* 85, 353–365.

Frisk, M., Göthe-Lundgren, M., Jörnsten, K., Rönnqvist, M., 2010. Cost allocation in collaborative forest transportation. *European Journal of Operational Research* 205, 2, 448–458.

Gnecco, G., Hadas, Y., Sanguineti, M., 2020. Public transport transfers assessment via transferable utility games and Shapley value approximation. *Transportmetrica A: Transport Science* 17, 4, 540-565.

Göthe-Lundgren, M., Jörnsten, K., Vãrbrand, P., 1996. On the nucleolus of the basic vehicle routing game. *Mathematical Programming* 72, 1, 83–100.

Granot, D., Hamers, H., Kuipers, J., Maschler, M., 2011. On Chinese postman games where residents of each road pay the cost of their road. *Games and Economic Behavior* 72, 2, 427–438.

Hezarkhani, B., Slikker, M., van Woensel, T., 2015. A competitive solution for cooperative truckload delivery. *OR Spectrum*, DOI 10.1007/s00291-015-0394-y.

Japkowicz, N. (2000). The Class Imbalance Problem: Significance and Strategies. *Proceedings of the International Conference on Artificial Intelligence*, 56, 111-117.

Kimms, A., Kozeletskyi, I., 2016. Shapley value-based cost allocation in the cooperative traveling salesman problem under rolling horizon planning. *EURO Journal on Transportation and Logistics* 5, 4, 371-392.

Krajewska, M.A., Kopfer, H., Laporte, G., Ropke, S., Zaccour, G., 2008. Horizontal cooperation among freight carriers: request allocation and profit sharing. *Journal of the Operational Research Society* 59, 11, 1483–1491.

Leenders, B. P. J., Velázquez-Martinez, J. C., Fransoo, J. C., 2017. Emissions allocation in transportation routes. *Transportation Research* Part *D* 57, 39-51.

Levinger, C., Hazon, N., Azaria, A., 2020. Computing the Shapley Value for Ride-Sharing and Routing Games. *Adaptive Agents and Multi-Agent Systems*, 1895-1897.

Liu, P., Wu, Y., Xu, N., 2010. Allocating collaborative profit in less-than-truckload carrier alliance. *Journal of Service Science and Management* 3, 1, 143–149.

Lozano, S., Moreno, P., Adenso-Díaz, B., Algaba, E., 2013. Cooperative game theory approach to allocating benefits of horizontal cooperation. *European Journal of Operational Research* 229, 2, 444–452.

Kullback, S., Leibler, R. A., 1951. On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79-86.

Massol, O., Tchung-Ming, S., 2010. Cooperation among liquefied natural gas suppliers: is rationalization the sole objective? *Energy Economics* 32, 4, 933–947.

Nguyen, C., Dessouky, M., Toriello, A., 2014. Consolidation strategies for the delivery of perishable products. *Transportation Research* Part *E* 69, 108–121.

Otero-Palencia, C., Amaya-Mier, R., Yie-Pinedo, R., 2018. A stochastic joint replenishment problem considering transportation and warehouse constraints with gainsharing by Shapley Value allocation. *International Journal of Production Research* 57, 10, 1-25.

Özener, O. Ö., 2014. Developing a collaborative planning framework for sustainable transportation. *Mathematical Problems in Engineering* 2014, art. ID 107102.
Wong, H., Van Oudheusden, D., Cattrysse, D., 2007. Cost allocation in spare parts inventory pooling. *Transportation Research* Part *E* 43, 4, 370–386.

Özener, O. Ö., Ergun, Ö., 2008. Allocating costs in a collaborative transportation procurement network. *Transportation Science* 42, 2, 146–165.

Özener, O. Ö., Ergun, Ö., Savelsbergh, M., 2013. Allocating cost of service to customers in inventory routing. *Operations Research* 61, 1, 112–125.

Pigott, M., 2023. Construction in the US. *IBIS World.* Retrieved the 2nd of October 2023 from: https://my-ibisworld-com.libproxy.mit.edu/us/en/industry/23/at-a-glance
Portland Cement Association, 2023. Chemical Admixtures. Retrieved the 2nd of October 2023 from:
https://www.cement.org/cement-concrete/concrete-materials/chemical-admixtures

Sakawa, M., Nishizaki, I., Uemura, Y., 2001. Fuzzy programming and profit and cost allocation for a production and transportation problem. *European Journal of Operational Research* 131, 1, 1–15.

Schmeidler, D., 1969. The Nucleolus of a Characteristic Function Game. *SIAM Journal of Applied Mathematics*, 17, 1163-1170.

Schulte, F., Lalla-Ruiz, E., Schwarze, S., Ramirez, R. G. G., 2019. Scalable Core and Shapley Value Allocation Methods for Collaborative Transportation. DOI 10.13140/RG.2.2.21550.87369/1.

Shapley, L. (1953) A Value for n-Person Games. *Contributions to the Theory of Games II*, P*rinceton University Press*, 307-317.

Vanovermeire, C., Sörensen, K., 2014. Integration of the cost allocation in the optimization of collaborative bundling. *Transportation Research* Part *E* 72, 125–143.

Vos, B. D., Raa, B., 2018. Stability Analysis of Cost Allocation Methods for Inventory Routing. *IFAC-PapersOnLine* 51, 11, 1682-1688.

Wang, Y., Ma, X., Xu, M., Wang, L., Wang, Y., Liu, Y., 2015. A methodology to exploit profit allocation in logistics joint distribution network optimization. *Mathematical Problems in Engineering* 2015, art. ID 827021.

Wang, Y., 2023. A collaborative approach based on Shapley value for carriers in the supply chain distribution. *Heliyon* 9, 7, 1-16.

Wong, H., Van Oudheusden, D., Cattrysse, D., 2007. Cost allocation in spare parts inventory pooling. *Transportation Research* Part *E* 43, 4, 370–386.

Wu, C. K., Lin, W. P., 2017. Using Shapley value for city bus route scheduling. *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS).* DOI 10.1109/ISPACS.2017.8266512.

Yengin, D., 2012. Characterizing the Shapley value in fixed-route traveling salesman problems with appointments. *International Journal of Game Theory* 41, 2, 271–299.

Zakharov, V., Shchegryaev, A., 2015. Stable cooperation in dynamic vehicle routing problems. *Automation and Remote Control* 76, 5, 935–943.

Zheng, X., X., Li, D., F., Liu, Z., Jia, F., Sheu, J. B., 2019. Coordinating a closed-loop supply chain with fairness concerns through variable-weighted Shapley values. *Transportation Research* Part *E* 126, 227-253.