Leveraging Simulation-Based Optimization to Generate Optimal Transportation Plans in the Real World

by

Adriele Pradi
Master of Science, Business Development and Internalization, Umea University

and

Andrew Duncan Mohn
Bachelor of Arts, Mathematics and Computer Science, Boston University

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

June 2023

Signature of Author: ____
Adriele Pradi
Department of Supply Chain Management
May 12, 2023

Signature of Author: ____
Andrew Duncan Mohn
Department of Supply Chain Management
May 12, 2023

Certified by: ____
Dr. Ilya Jackson
Postdoctoral Associate, MIT Center for Transportation & Logistics
Capstone or Thesis Advisor

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

Leveraging Simulation-Based Optimization to Generate Optimal Transportation Plans in the Real World

by

Adriele Pradi

and

Andrew Duncan Mohn

Submitted to the Program in Supply Chain Management
on May 12, 2023 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

This capstone project evaluates the application of stochastic optimization techniques in middle-mile transportation planning, incorporating historical variance in transportation time and yard dwell time. Wayfair's current middle-mile planning process uses advanced forecasting and optimization techniques, but it struggles to account for the randomness and variation of the real world. To address this, the capstone project evaluates whether incorporating sources of variance into the optimization process can outperform traditional models in generating resilient transportation plans. After analyzing 70,000 trips from January 2022 to January 2023, three significant lanes were selected to evaluate changes in the distributions based on day of the week, season, and carrier. For each lane, 20 simulated transportation plans were created using stochastic data. Results confirm that accounting for variance improves outcomes, showing the possibility to incorporate more realistic inputs to the transportation planning process. Managerial insights highlight the model's possibility to representing real-world scenarios, enabling informed decisions on resource allocation, route selection, and scheduling. The scenario-based approach balances speed and efficiency, empowering organizations to manage uncertainty.

Capstone Advisor: Ilya Jackson
Title: Postdoctoral Associate, MIT Center for Transportation & Logistics

ACKNOWLEDGMENTS

We would like to thank our capstone advisor Ilya Jackson for the guidance, support and encouragement. We also would like to acknowledge the availability and guidance of our company partners. During the development of this capstone, it was a pleasure to discuss important matters related to the topic with several MIT colleagues and faculty, in particular Pamela Siska, who provided us with detailed feedback on the writing and presentation. Finally, many thanks to the support of our family and friends.

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

As one of the world's largest home goods retailers, Wayfair has built a large supply chain and logistics network. The e-commerce giant sells over 14 million items from more than 20,000 global suppliers and has two corporate headquarters to support its worldwide expansion, one in Boston and one in Berlin, and large satellite offices in Seattle, Austin and San Francisco. To fulfill global demand and drive customer satisfaction, Wayfair operates 18 fulfillment and 43 delivery centers across North America and Europe (Wayfair, 2023).

Building a resilient operations network is an important part of upholding customer promises. The middle-mile transportation plays a key role in this execution, not only for Wayfair but for any business distributing products to customers. The middle-mile can be defined as the part of the logistics focused on freight transportation with the goal of distributing goods from stocking locations to certain destination facilities, from where the goods will be rerouted to final customers for the last mile operations (Greening et al., 2022). Wayfair's North America middle-mile network contains a complex hierarchy of aggregation points, cross-dock facilities, fulfillment centers, and delivery agents.

As a consequence of the COVID-19 pandemic, a multitude of external forces tested the resilience of supply chains and logistics operations in 2020 and 2021: closures of main ports, labor shortages, imbalance between container demand and availability, facility closures, among others. In this context, companies must prepare to include these relevant scenarios in their planning process. Wayfair faced this challenge and, as the primary motivation for this project, we focused on opportunities to strengthen the existing middle-mile transportation planning process in the face of sources of uncertainties. Furthermore, we believe that the work developed can be transferable to other companies and modalities of transportation that are dealing with uncertainties impacting their operations.

1.1 Company Background

Wayfair developed an e-commerce platform with more than 22 million active customers. According to the 2023 Investors Presentation (Wayfair, 2023), the company manages more than 14 thousand employees and operates a proprietary end-to-end logistics network. The logistics operations are one of the key factors for the company to maintain and grow their supplier network, which is linked to the success of the company's products with customers. The company oversees a logistics portfolio comprising:

1) Forwarding services: Ocean transport services to bring product from manufacturers in Asia closer to customers in North America and Europe,

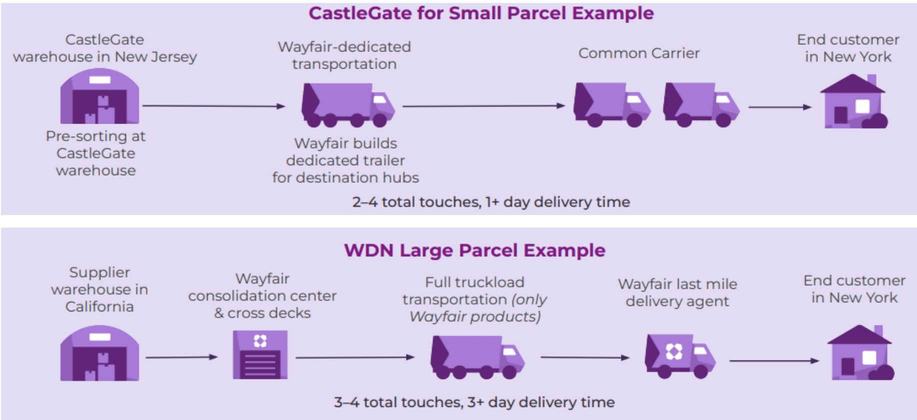
2) Fulfillment operations: Positioning fulfillment center network in 16 locations in North America and 2 locations in Europe.

3) Delivery network: Proprietary middle-mile and last-mile delivery service for large parcel products with 35 locations in North America and 8 in Europe.

Figure 1 illustrates the process of delivering small and large parcels using the Wayfair network targeting consistent and reliable delivery for customers. "CastleGate" is the name of the network developed by Wayfair for small parcels. The acronym WND, which means warehouse and distribution, is adopted to represent the process for distribution for large parcels.

Figure 1

Wayfair Distribution Network



Note: Figure 1. Delivery process adopted by Wayfair for small and large parcels. Retrieved from https://s24.q4cdn.com/589059658/files/doc_financials/2022/q4/2023-_-Investor-Presentation.pdf

1.2 Problem Statement and Research Question

Because of the extensive size of Wayfair's network and customer base, the company adopts advanced practices for transportation planning. Their current system for planning middle-mile logistics involves a complex process of forecasting and optimization techniques such as mixed integer linear programming, dynamic programing, and local search heuristics, leveraging their large amount of well-structured data stored in SQL and no-SQL databases. While these modern tools have allowed companies to build highly efficient operations plans, many companies find it difficult to adequately account for the true randomness of the real world in these processes (Trebilcock, 2022, 09:10). Wayfair is no different - the company's models are run on point forecasts, which are single point estimates of its expected value, and often do not consider key aspects of variation that impact Wayfair's operations. Therefore, the plans created by the model will perform optimally only if forecasts are accurate and there is no variation in the timing of operations. This means there is significant unaccounted-for risk that existing variation could disrupt the plan and cause significant problems. Worse, there is little to no visibility into the most disruptive forms of variance.

For Wayfair's middle-mile planning problem, there are two primary sources of variation: facility operations (yard times, and loading and unloading time) and driving time between facilities. To mitigate potential risk, Wayfair takes a conservative approach to planning, adopting measures to leave space or extra time in case there are issues. While this is an effective solution, it decreases the optimality of the plan, leading to lower truck utilization and higher dwell times, and drives up costs in an already low-margin industry.

The sources of variation described by Wayfair can be incorporated into the transportation planning process using probability distributions. Significant research has been conducted on techniques to account for the probability distributions of inputs while solving optimization problems more precisely. These techniques include stochastic and robust optimization, simulation and scenario testing, and reinforcement learning. While these methods have been thoroughly studied in academic contexts, many companies have found them difficult to implement as it requires extensive expertise in the subject and a high computational budget.

In this capstone project, we answer the following research question:

Does taking into account the two sources of variance in inputs during optimization outperform a traditional optimization model in resilient transportation plan generation?

To answer this question, we will explore to what degree Wayfair can be more realistic in planning truck timing and utilization while still meeting an acceptable level of reliability. We will also explore how the results of the simulation compare to the current model used by Wayfair, which will support identifying possible blind spots or inaccurate assumptions that are used during the current planning process.

1.3 Project Goals and Outcomes

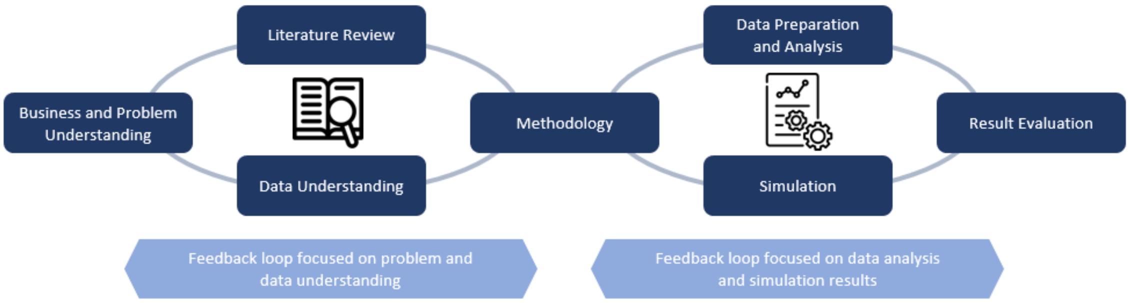
The project's goal is to develop and use a dynamic stochastic optimization model to create more robust transportation plans for Wayfair. To analyze the performance of this model, we will compare our model to the current transportation planning approach used by Wayfair and evaluate to what degree the performance of this plan is improved. This evaluation will give Wayfair insights into how to create stronger transportation plans, deliver on customer promises, and incorporate the sources of variance for alignment with real-world uncertainties. To construct the model, distributions were created using historical data for all inputs of the linear integer programing model. The model is focused on Wayfair's primary market, the United States.

The primary hypothesis tested is whether by using the dynamic simulation-optimization model Wayfair can realize a positive impact on the metric "customer promise". Customer promise is a key metric for Wayfair and companies in the e-commerce space, though cost and speed are good proxies for this metric. This metric is based on how often Wayfair achieves the timeline that was communicated to the customer and can gain the ability to be more aggressive on the expected delivery time proposed to customers.

During the initial stages of the capstone, we mapped phases based on the CRISP-DM methodology. CRISP-DM (the Cross Industry Standard Process for Data Mining) is a framework for data mining projects to make it less costly, more reliable and repeatable (Wirth and Hipp, 2020). Figure 2 displays the steps and actions that were followed to execute the research process. The steps mapped under the same loop were executed in parallel or had dependencies between the phases.

Figure 2

Capstone Execution Phases



2. State of the Art

The core problem of our capstone is how Wayfair can take into consideration sources of variance to outperform traditional optimization models in the middle-mile transportation planning process. To support the understanding of the problem and decisions on the methodology, we focused on four areas of the literature: (1) Middle-mile transportation, (2) Resilience in transportation, (3) Stochastic optimization and programming, and (4) Scenario-based frameworks.

2.1 Middle-mile Transportation

Trucks and rail are major modalities of transport across the world. Bureau of Transportation Statistics figures indicate that in 2021 trucking was the most common domestic mode of transport, responsible for 65% of the freight weight transported domestically in the USA (BTS, 2021). The rail mode followed in second place, carrying around 10% of the freight weight. Among the reasons why trucking is leading compared to other transportation options, Crainic et al. (2021) highlights the fast transit time and guaranteed reliability on the expected delivery time, which directly impacts customer service levels.

Among the applications of domestic freight transport in the USA, the middle-mile has gained importance with the emergence of e-commerce channels. According to Petroianu (2020) middle-mile transportation is focused on cargo moving from one facility to another observing a set of constraints, such as pickup and delivery times. As part of logistics operations, the middle-mile includes bringing consolidated shipments from receiving points to a break-up or cross-dock location, from where it will be rerouted for the last mile of delivery (Seaton, 2018). One of the goals of the middle-mile network is to distribute and allocate inventory to achieve balance of goods distributed across the network, positioning products closer to customers, and decreasing shipping time (Petroianu, 2020).

Middle-mile operations are embedded in a complex network that is managed to drive efficiency, avoiding loading delays, and decreasing wasted drive time (Seaton, 2018). Operating these networks incurs significant costs. Walmart, for example, is testing driverless trucks to cut costs on the middle-mile, which, according to the company, is the most expensive part of the whole supply chain and a huge pain point due to driver shortages (Naughton & Boyle, 2019). The pressure to reduce costs is mentioned by Seaton (2018), who studied the current situation of e-commerce distribution with the emergence of "free shipments and free returns". Decisions on how to physically set up the network and evaluate risks, such as working with independent carriers or owning a fleet, are key to maintaining competitiveness.

Research was also done focused on the middle-mile planning decisions. Greening et al. (2022, p. 3) explains that "planning models specify a path for each shipment from its origin to destination, either directly or through intermediate sorting facilities". The model developed by the authors considered waiting times at facilities for solving consolidation network design problems and achieving lower costs. Companies, such as Amazon, are investing heavily in the planning process for the middle-mile. In 2021, the company was awarded an INFORMS prize for operational research to tackle the challenges related to the middle-mile (Amazon, 2021). One component of Amazon's research is working with optimization and machine learning models to arrive at the most optimal decisions for network planning and to find solutions for challenges that are inherently stochastic or unpredictable in nature.

The middle-mile is a vibrant area of research that has gained increasing attention from both academia and companies with the current competitive scenario in e-commerce distribution. In this context, it becomes even more important to understand the forces that can bring disruption and uncertainty to the middle-mile operations.

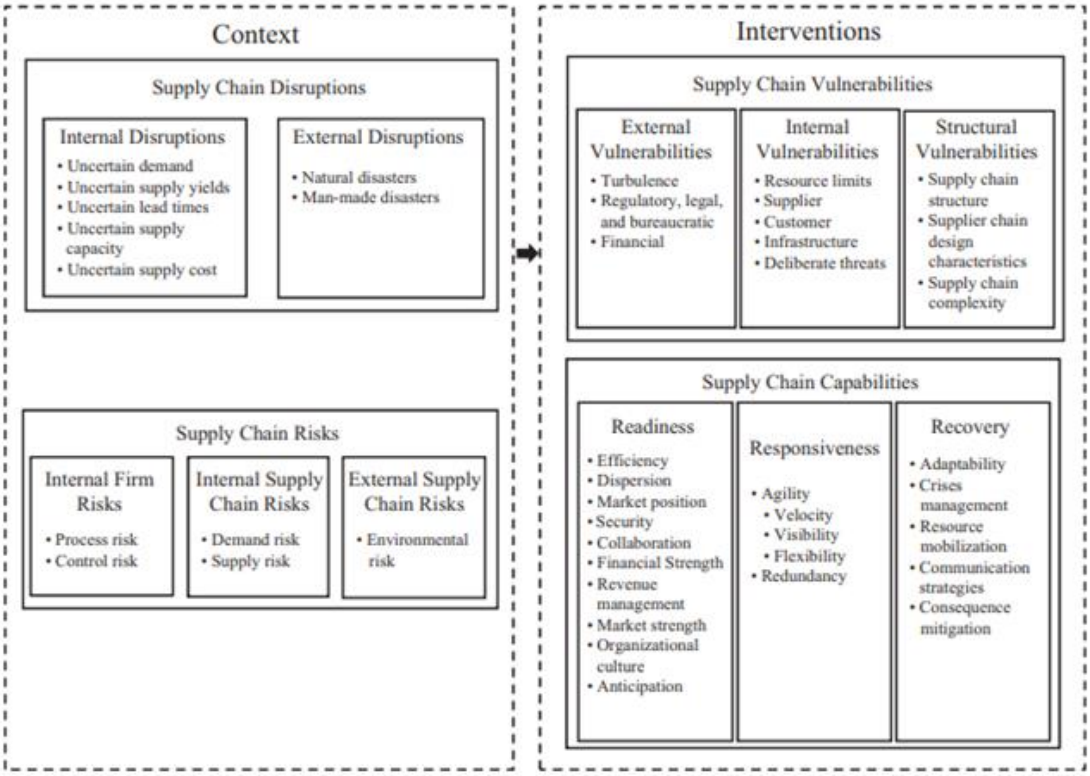
2.2 Resilience in Transportation

E-commerce logistics and transportation operations are key to fulfilling delivery time promises and maintaining competitiveness. Transportation systems should incorporate elements of resilience to function optimally. In a literature review on the topic of supply chain resilience, Kochan and Novick (2018) mapped a framework of aspects of the definition of resilience available in the academic literature. As indicated on Figure 3, some aspects under the internal disruptions are present in the middle-mile network operations, such as the uncertain demand lead times and cost. Petroianu (2020,

p. 5) highlights that "one of the main difficulties of this type of routing is that trucks have to be allocated without the certainty of demand. This situation can lead to last minute cancellations or ad hoc demands that jeopardize planning."

Figure 3

Supply Chain Resilience Typology



Note. This figure was adapted from Kochan and Nowick (2018) to map and display the existing literature in supply chain resilience. From *"Supply chain resilience: a systematic literature review and typological framework"* (p. 845), by Kochan, G and Nowicki, D, 2018, International Journal of Physical Distribution & Logistics Management. Copyright 2018 by Emerald Publishing Limited.

The concept of resilience thinking in transportation was discussed by Wang (2015). The author classifies the disruptions that can impact a transportation system in two dimensions: frequency of occurrence and level of damage. Under those categories, each type of event can be mapped on to a classification of (1) disaster, which includes natural events such as earthquakes; (2) day-to-day variations, such as changes in the demand or capacity; or (3) ongoing long-term changes, such as global warming. The author highlights that the reliability of a network is linked to the ability of managing day-to-day variations.

Similarly, Andersson et al. (2017) states that transport systems are subject to constant disturbances. These disturbances can have several sources, such as traffic incidents or congestion that can be classified as System Killers, Catastrophic Events, Expected Risks, and Contingencies. The author proposes a calculation model focused on estimating the value of transport time variability.

Uncertainty is a constant aspect under the supply chain resilience topic. In a review of existing literature on intermodal transportation, Delbart (2021) identified that stochastic demand and stochastic transit time are the most studied types of uncertainty for planning. The author suggests that there is a gap in the literature for studies that combine several types of uncertainty and address uncertainty on a strategic level.

The concept of resilience in the context of supply chain management is well covered in the academic literature and comprises a range of possible events. While force majeure events are often associated with the discussion about resiliency, as experienced during the COVID-19 pandemic, studying day-to-day uncertainties is also central to support companies maintaining a competitive position on costs and delivery time. To incorporate these variations, we present the concept of stochastic optimization, stochastic programming and scenario-based frameworks.

2.3 Stochastic Optimization

"Stochastic optimization refers to a collection of methods for minimizing or maximizing an objective function when randomness is present" (Hannah, 2015, p. 1). The goal of stochastic optimization is to optimize some decision problem while sufficiently accounting for uncertainty in the problem's specification. Just as there is no "one size fits all approach" to solving deterministic optimization problems, a huge number of techniques are in use for solving these stochastic optimization problems depending on the structure of the problem being solved. Research into this problem includes "stochastic control, dynamic programming (Markov decision processes), stochastic programming or robust optimization, with a variety of related fields using names such as reinforcement learning, approximate dynamic programming, and stochastic search" (Powell, 2016, p. 1459).

This set of techniques has already been applied to the transportation space. So far, most research into this space covers the problem from the perspective of minimizing the risk due to the disruption caused by environmental disasters such as earthquakes, floods, and hurricanes (Sabouhi, Bozorgi-Amiri & Vaez, 2021). However, these models are closer to robust optimization, where the goal is to minimize the impact of disruption in even the worst case, rather than Stochastic Optimization, where low probability events are only given small weights, generally leading to less conservative plans (Maggioni, Potra & Bertocchi, 2017). The only paper we could find that directly dealt with intermodal transportation planning with uncertainty was Baykasoğlu and Subulan (2019), which introduces a fuzzy-stochastic mixed-integer program for a complex, multi-objective transportation planning problem. We evaluate this second strategy for considering uncertainty in optimization models in the next section.

2.4 Stochastic Programming

The approach used in Baykasoğlu and Subulan (2019), and a generally common academic approach for solving these stochastic optimization problems for discrete operations problems, is a two-stage stochastic integer programming model. As the name implies, this technique involves breaking the problem into two stages, which are solved independently. The first stage is solved purely deterministically, without information on the uncertainties of the inputs. The uncertainties are then realized, and the second stage applies some corrective actions to the solutions found in the first stage (Birge & Louveaux 2011). While this technique has been successful in academia, there are significant difficulties in applying this technique to the large-scale problems seen in industry. Specifically, the evaluation of costs in the second stage introduces challenges such as the requirement to integrate the value function of the integer program, solving many NP-hard integer programs, and creation of non-convex or discontinuous second stage programs (Ahmed, 2010). While this technique offers an interesting framework for solving our problem, there is considerable risk that it will not scale to the dozens of facilities and thousands of lanes that our problem involves. As a more scalable concept to address uncertainty, we will explore the scenario-based frameworks option in the next section.

2.5 Scenario-based Frameworks

A scenario-based framework is a methodology for sampling probability distributions and evaluating the performance of optimization models across those scenarios: "This heuristic methodology also allows us to overcome the typical problem of computational intractability of ARC [stochastic optimization] and can be applied to any optimization problem affected by uncertainty." (Maggioni, Potra & Bertocchi 2017, p 7). By considering a finite set of realized scenarios of our random events, we can approximate the results of a two-stage stochastic problem without the challenges described above.

This strategy yields many benefits. First, it directly addresses the potential computational infeasibility of modeling the problem through other stochastic optimization techniques. Second, it allows for a custom balance between stochastic programming and robust optimization by allowing the user to choose the weights of scenarios to either hedge more against risk or save costs by optimizing for more likely events. Finally, the technique can be used regardless of the formulation of the underlying optimization model, requiring only an evaluation function and leaving the rest of the model a black box (Maggioni, Potra & Bertocchi, 2017). This means the methodology has maximum applicability to other types of operations problems and the solution explored in this Capstone can be transferable to other companies facing a similar problem.

Overall, there is a wide range of techniques and methodologies to incorporate uncertainty into the middle-mile planning processes. It is clear that this facet of operations is critical for a company to achieve business goals related to customer satisfaction and costs and highlights the importance of creating a resilient plan. In the methodology section, we explain the rationale behind our decision to create a model that incorporates uncertainty aspects for Wayfair's transportation planning processes.

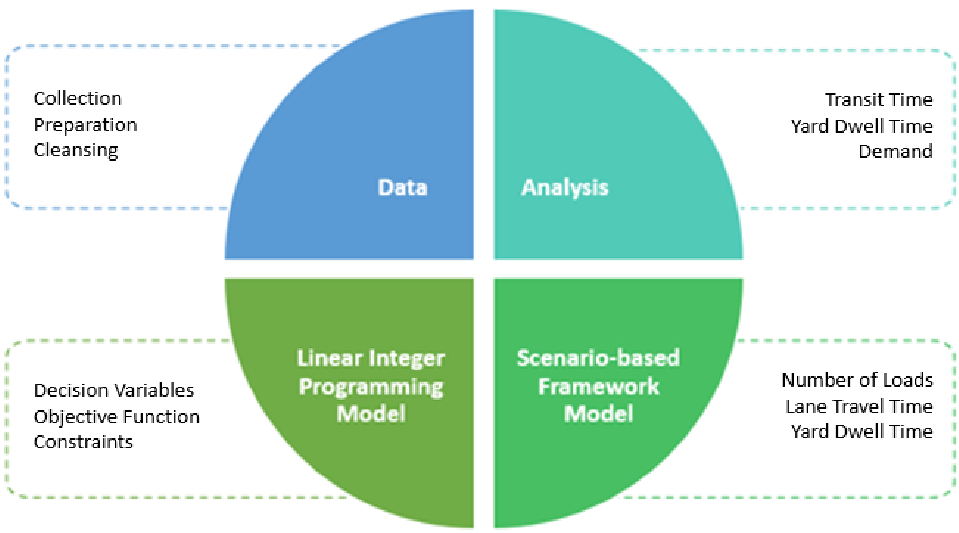
3. Methodology

After reviewing the literature and recognizing the practical challenges of implementing traditional stochastic optimization techniques at scale, we determined that the best approach for solving this problem for Wayfair was not to redesign the existing optimization models for solving transportation problems, but to use a scenario-based framework to leverage the work from previous research. Implementing this framework involved sampling the distributions of the uncertain inputs and repeatedly solving the optimization problem with their realized values.

We began by analyzing the data set available at Wayfair, including the distribution, average, mean and range. As a next step, we applied the Kolmogorov-Smirnov test to understand whether there was a variance in the distribution based on certain criteria, namely lane, day of the week, and season. Next, we used the Wayfair simulation model to apply the scenario-based framework and understand the outputs. Finally, to provide Wayfair with a solution that can be applied to their business context, we designed a framework to interpret results. Figure 4 displays a summary of the methodology chapter and process adopted.

Figure 4

Methodology Overview



3.1 Data Collection and Preparation

As focus on day-to-day variance was not yet applied to Wayfair's optimization model, the first step was to understand the data to identify relevant aspects that could be part of the simulation model. Wayfair's data presented overall good quality with few outliers. As our work is focused on operations under normal conditions, we chose to eliminate extreme outliers. In a logistics network, the outliers can represent the consequences caused by extreme events such as severe weather. Those conditions are not part of this research; for this reason, we removed all data points below or above the 95% quantile for the selected sources of variance.

After cleaning the data, we created visuals illustrating each possible source of variance. We built charts to help understand the distribution and whether the impact has statistical significance using the Kolmogorov–Smirnov test. According to Grall-Maes (2012), statistical goodness of fit tests identifies discrepancies that are observed between sample values and the values expected in the model. It can be used to test whether an observed sample follows a distribution. The author describes the Kolmogorov–Smirnov test as a widely used test to evaluate the equality of continuous one-dimensional probability distributions.

3.2 Data Description

Wayfair has a large amount of data available that can be explored on different levels. To perform the analysis, we extracted information for each transportation leg executed during the last 13 months. In total, we had approximately 70 thousand rows of data representing the executed movements. For each movement, Wayfair had 110 dimensions with detailed information on several aspects related to the plan and execution. We started the data analysis process by selecting the relevant dimensions, cleaning the data, removing outliers, and conducting exploratory analysis to understand the main sources of variance.

Due to the extensive and detailed number of fields available in the Wayfair dataset, we selected the most relevant and reliable columns. The decision was based on the importance the field has for transportation. After filtering out fields that were less relevant, we selected 52 dimensions to characterize each movement and provide information on the possible sources of variance and day-to-day disruptions.

In addition, the travel time was calculated based on the information from the data set. We added as a new field, number 53, to provide the information on hours spent between pick up time and arrival time.

3.3 Data Cleansing

After understanding the data and assessing the overall data quality, we concluded that the data set required only a few steps to be ready for analysis. As a first step, we removed any rows that had departure dates before 2020 and arrival dates after 2024. As the database only covered the last 13 months, dates before or after those periods of time were likely to be an error. As a second step, we proceeded to address the outliers. As the purpose of the project is to identify and propose transportation plans under normal operational conditions, we opted for removing outliers. Considering that lanes and overall movements can be subject to extreme events that affect the operations, such as travel time, we considered as outliers any movements that presented travel time above or below the 95^{th} percent quantile. After applying the changes, the data was ready for the initial analysis.

3.4 Data Analysis: Transit Time

Based on the lane ID, we selected one lane to be used as a pilot for understanding and exploring the dataset. The lane ID selected represents a lane from the West Coast to the East Coast, and it had more than 700 movements registered during the past 13 months. As this is a long lane, it is subject to several uncertainties during the transport execution. In this section we will present the results of the travel time analysis for one lane. The same test was applied to all lanes available in the Wayfair database.

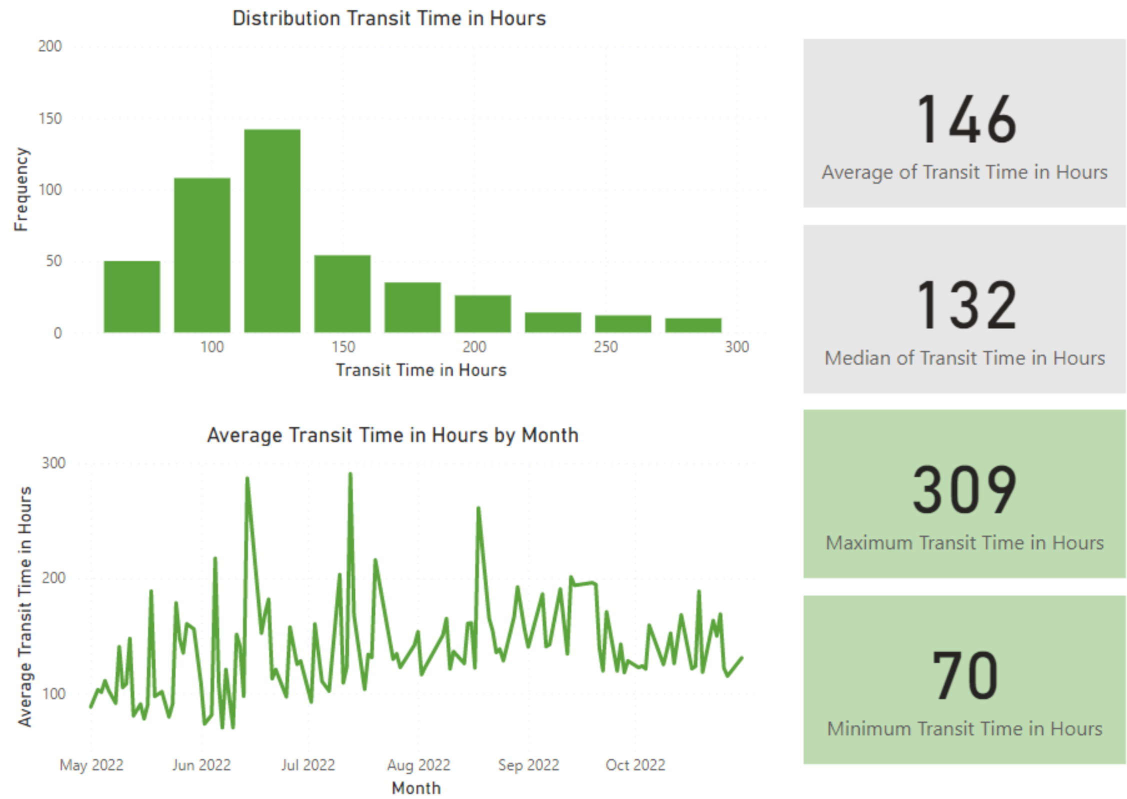
As the travel time is an important aspect for data understanding we explored the distribution and differences based on day of the week, season and carrier.

3.4.1 Transit time distribution

The variance in transit time is visible in the dataset and illustrated in Figure 5. For this specific lane, historical data registered a travel time between 70 and 309 hours. Looking at the distribution by month, it is also possible to identify that every month presents variations.

Figure 5

Transit Time Distribution

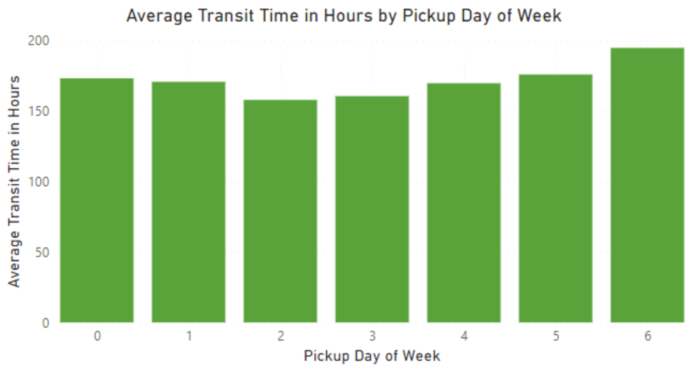


3.4.2 Transit time by day of the week

The distribution of transit time by data of the week, with Monday counted as day 0 and Sunday as day 6, also presents variation. The most impactful one, as illustrated in Figure 6, is the increase in transit time on Saturday and Sunday. We applied the Kolmogorov–Smirnov test to determine whether the distributions are different based on the days of the week and had the following outcome with p-value 0.05.

Figure 6

Transit Time by Day of the Week



Kolmogorov–Smirnov test:

Measure	Test Result	Test Outcome
6 - Sunday	0.000	different
5 - Saturday	0.000	different
4 - Friday	0.012	different
3 - Thursday	0.001	different
2 - Wednesday	0.040	different
1 - Tuesday	0.316	NOT different
0 - Monday	0.565	NOT different

Based on the test, the conclusion is that only Mondays and Tuesdays are not different from the overall distribution. This finding means that, for example, loads departing on a Thursday will have a transit time that is different from the average for the entire dataset.

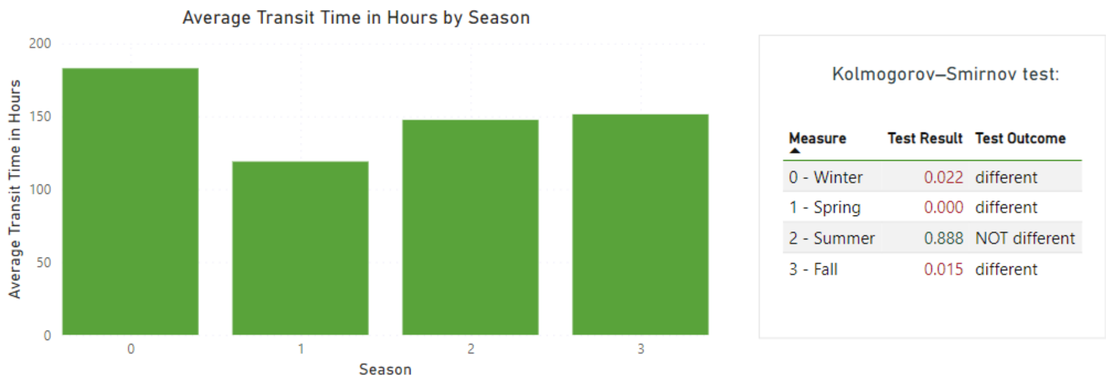
3.4.3 Transit time by season

The transit time by season was also evaluated. The seasons were mapped as Winter (0), Spring (1), Summer (2) and Fall (3). The main variation is in spring, which has the lowest travel time, and Winter, which has the highest travel time. Fall also represents a peak season for Wayfair and similar businesses, which can have an impact on the performance for that lane.

We applied the Kolmogorov–Smirnov test to determine whether the distributions are different based on the seasons and had the following outcome with p-value 0.05, as shown in Figure 7.

Figure 7

Transit Time by Season



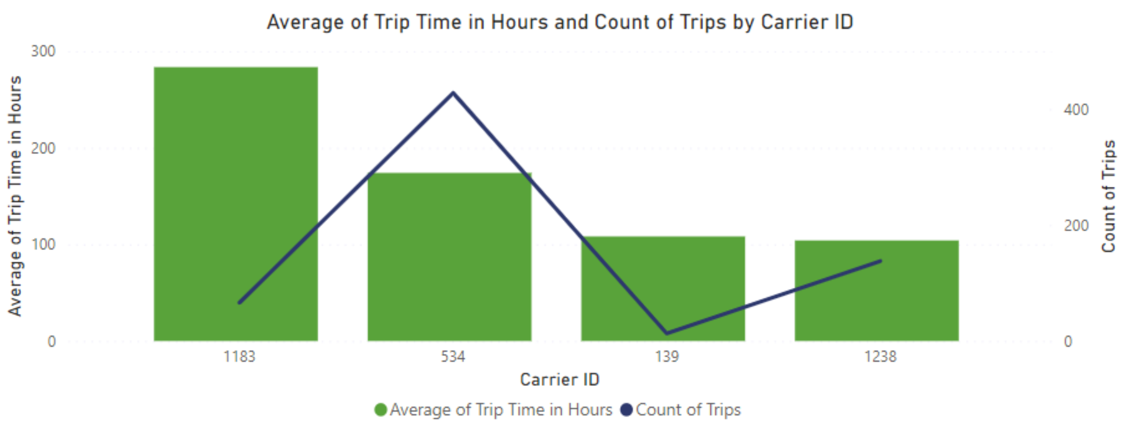
The Kolmogorov-Smirnov test revealed that the distributions of winter, spring, and fall are different compared to the overall distribution. This means that only the distribution of the transit time during summer is not different than the distribution of the overall dataset.

3.4.4 Transit time by carrier

The difference in transit time was also evaluated as displayed in Figure 8, where the bar represents the average transit time, and the line represents the number of trips transported by each carrier. As displayed in the figure, there is no relation between count of trips and average transit time. Which means that using a carrier using a lane very often or carrying a significant amount of cargo, may not lead to them becoming more efficient in their operations times on average.

Figure 8

Transit Time by Carrier



3.5 Data Analysis: Yard Dwell Time

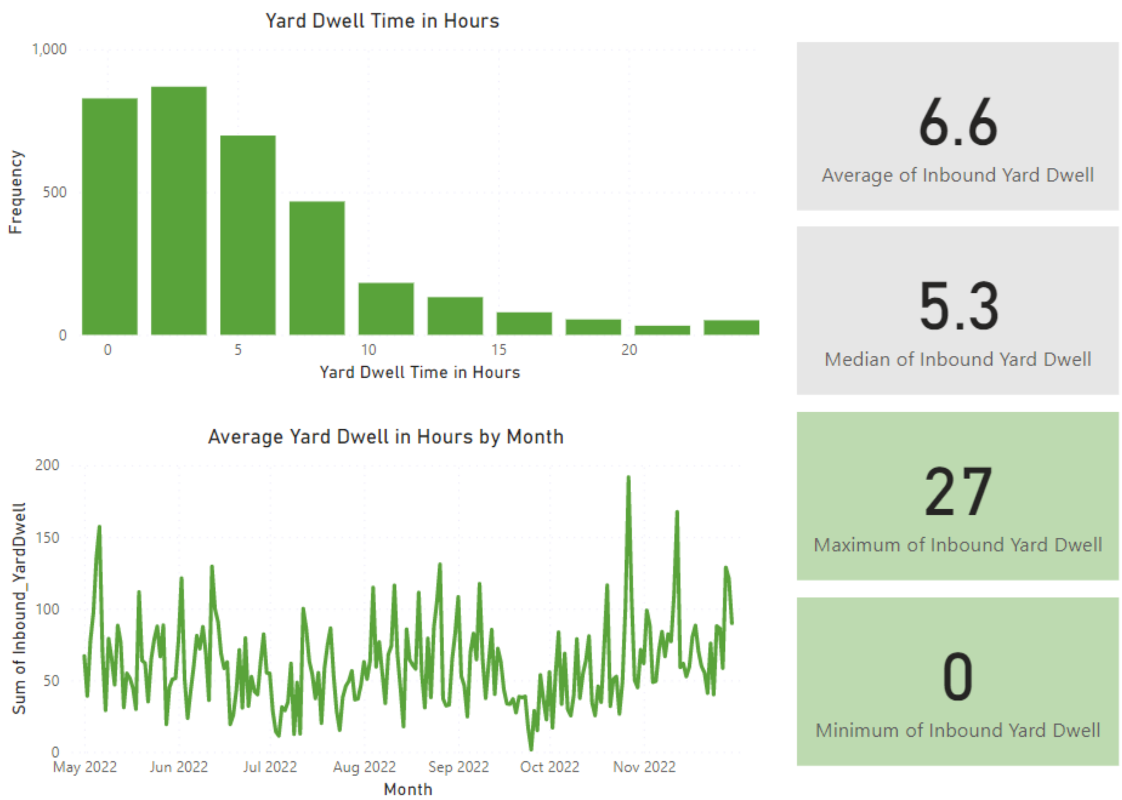
Based on the facility ID, we selected one location to be used as a pilot for understanding and exploring the dataset. The trucking yard dwell time, which is explored in this sub-chapter, refers to the period that a truck spends within a designated yard or terminal before its cargo is unloaded, loaded, or transferred to another mode of transportation. The facility ID selected represents a large cross-dock located on the East Coast and was selected as a top representative of broader Wayfair operations.

3.5.1 Yard dwell time distribution

The variance in the yard time is visible on Figure 9. It represents a variation in the hour's frequency across the entire year as well as patterns on the monthly distribution. As a next step, we evaluated the distribution per day of the week, season, and carrier.

Figure 9

Yard Dwell Time Distribution

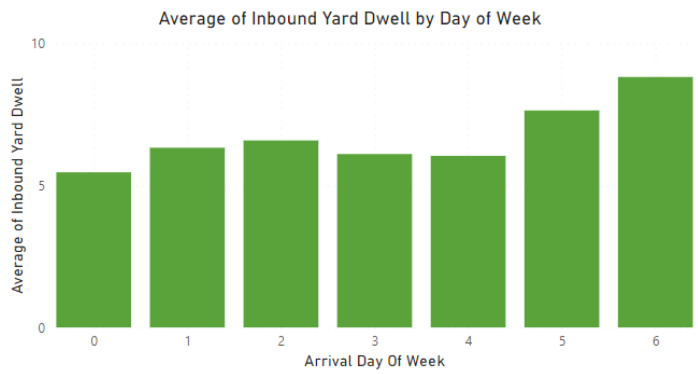


3.5.2 Yard dwell time by day of the week

The distribution of trucking yard time by day of the week, with Monday counted as day 0 and Sunday as day 6, also graphically presents variation. The most impactful variation that can be seen in the plots is the increase in yard time on Saturday and Sunday. We applied the Kolmogorov–Smirnov test to determine whether the distributions are different based on the seasons and had the following outcome with p-value 0.05. The results are shown in Figure 10.

Figure 10

Yard Dwell Time by Day of the Week



Kolmogorov–Smirnov test:

Measure	Test Result	Test Outcome
0 - Monday	0.000	different
1 - Tuesday	0.241	NOT different
2 - Wednesday	0.312	NOT different
3 - Thursday	0.336	NOT different
4 - Friday	0.402	NOT different
5 - Saturday	0.000	different
6 - Sunday	0.000	different

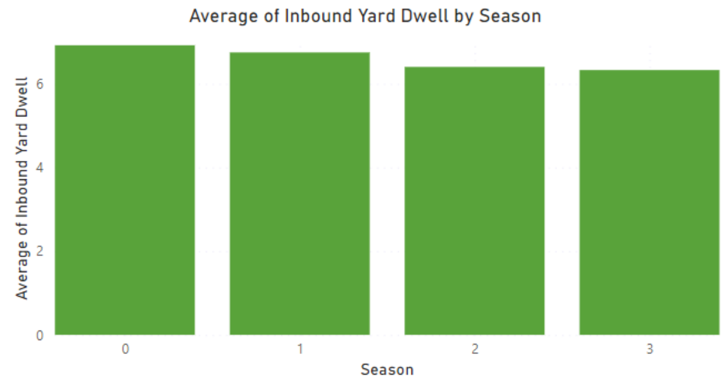
According to the Kolmogorov-Smirnov test, the distribution for Monday, Saturday and Sundays is statistically different compared to the overall distribution. This means that the dwell distribution for a truck arriving to the yard on a Saturday is meaningfully different than arriving on any random day.

3.5.3 Yard dwell time by season

The seasons are mapped as Winter (0), Spring (1), Summer (2) and Fall (3). We applied the Kolmogorov–Smirnov test to determine whether the distributions are different based on the seasons and had the outcome displayed on Figure 11 with p-value 0.05. According to the test, all seasons follow the same distribution regarding the yard dwell time.

Figure 11

Yard Dwell Time by Season



Kolmogorov–Smirnov test:

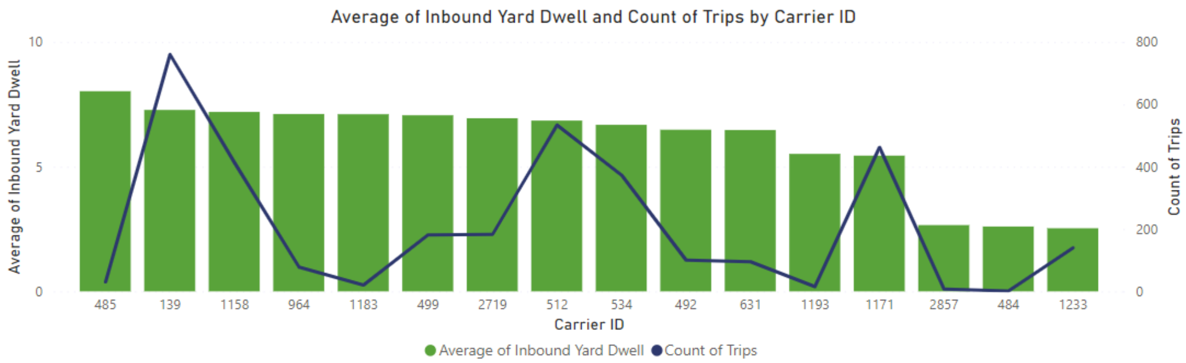
Measure	Test Result	Test Outcome
0 - Winter	0.090	NOT different
1 - Spring	0.878	NOT different
2 - Summer	0.437	NOT different
3 - Fall	0.265	NOT different

3.5.4 Yard dwell time by carrier

As a final analysis, we visualized whether there is a difference in the trucking yard dwell time by the carrier that performed the movement. Figure 12 shows the difference in transit time and volume carried by the several carriers that had cross-dock operations at the location. The analysis also indicates that on average there is no relation between the number of loads executed by one specific carrier and the yard dwell time.

Figure 12

Yard Dwell Time by Carrier



3.6 Linear Integer Programming Model

The middle-mile operations planning is based on a linear integer programming model that considers decision variables to optimize over an objective function, subject to some constraints and penalty functions.

3.6.1 Decision Variables

There are four groups of decision variables. The first group comprises the candidate movement slots, represented by 'Movements', which selects the vehicles that will be used. The decision variables in this category are non-negative integer numbers. The second decision variable group represents the demand key, which represents the candidate movement pairs to assign flow to vehicles, recognized as 'Demand'. The variables in this category are non-negative and continuous as they can represent different dimensions of the demand, for example, volume. The third decision variable, assigned in the model as 'Unfulfilled Demand', relates to tracking any unfulfilled demand and follows the same principle, as a non-negative and continuous variable. The fourth variable, assigned in the model as 'Floor', relates to tracking failure to assign minimums based on soft floor constraints. This constraint is in place to consider the minimum allocation contracts that Wayfair has in place with its freight vendors. The variable is a non-negative integer number.

In summary, the decision variables are:

- X[m]: Non-negative integer decision variable indexed over candidate movement slots. It represents the number of trucks selected for each movement slot m.
- A[k, m]: Non-negative continuous decision variable indexed over demand key (k) and candidate movement (m) pairs. It represents the flow assigned to trucks in each movement slot m for demand key k.
- U[k]: Non-negative continuous decision variable indexed over demand keys (k). It represents the unfulfilled demand for each demand key k.
- F[g]: Non-negative integer decision variable indexed over soft floor constraint groups. It represents the failure to assign the minimum number of trucks in each group g.

3.6.2 Model Constraints

The constraints identified by Wayfair are in place to represent the limitations and requirements that should be followed when applying the model. There are five constraints used in the model. The first constraint is related to the decision variable 'Movements' and is in place to limit the maximum number of vehicles that are available for the movements.

The second constraint relates to the decision variables 'Movements' and 'Demand'. The constraints limit the amount of demand allocated to each vehicle based on the capacity of each possible vehicle available for the movement.

The third constraint is related to the 'Demand' and 'Unfulfilled Demand' decision variables. It sums both variables to make sure that the total volume of demand is considered by the model. The fourth constraint tracks the compliance with the lower bound of the minimum required commitment that Wayfair has in place with its vendors.

The fifth constraint is related to the 'Movements' decision variable and guarantees that the physical flow required by a vehicle will be performed. This is a precedent constraint and takes into consideration that when the vehicle is at point A, it must move physically from point A to point B, before being able to move from point B to point C.

In summary, the constraints apply on the following:

- Group capacity constraint: The sum of X[m] for each group of candidate movement slots should not exceed the upper bound for that truck group
    - Sum(X[m] for m in group) <= UpperBound for each (group, UpperBound)
- Truck capacity constraint: The total flow assigned to trucks in each movement slot m should not exceed the truck capacity times the number of trucks in that slot.
    - Sum(A[k, m] for all k) <= TruckCapacity[m] * X[m] for each candidate movement slot m
- Demand satisfaction constraint: The total flow assigned to each demand key k plus unfulfilled demand should equal the demand volume for that key.
    - Sum(A[k, m] for all m) + U[k] = DemandVolume[k] for all demand keys k
- Soft floor constraint: The total number of trucks assigned to each soft floor group g plus the failure to assign minimums should be greater than or equal to the lower bound for that group.

- Sum(X[m] for m in SoftFloorGroup[g]) + F[g] >= LowerBound[g] for g in Soft floor constraints

- Precedence constraint: The number of trucks assigned to a movement slot should satisfy precedence relationships.
  - LeftScalar * X[LeftM] >= RightScalar * X[RightM] for (LeftScalar, LeftM), (RightScalar, RightM) in Precedence constraints

3.6.3 Penalty Function for Orders

The penalty function is a key feature for the integer linear programming model. It defines the mechanism by which a trade-off between cost and speed is defined. Each lane has a single penalty function, which is represented by 'Penalty Function' and is calculated as 'Assigned Arrival Date and Time' minus the 'Arrival Milestone Date and Time'. The property of the penalty function includes three considerations.

The first consideration is that optimal solutions should send orders out as first in first out. If that order is not followed, the operations may not be able to execute the plan and realize the intended value. The goal is that the order arrives no later than the defined milestone time, without considering as a benefit if the order arrives much earlier than expected.

The second consideration is the arrival windows. Each facility present in the model has a certain time each day where they make their delivery routes. Items that arrive after this window will not be sent until the following day. The model penalizes orders that arrive after the delivery window but is indifferent about what time the orders arrive within the time window. This penalty allows the model to consider for planning effects events such as weekend closures or holidays, which will have an impact on the time required for processing the order.

3.6.4 The objective function

The objective function is designed to minimize the sum of 'Movement' costs based on the chosen movement. It considers the assignment penalty for each pair of demand key 'Demand' and candidate 'Movement' slot. It also calculates and attributes a penalty for the 'Unfulfilled Demand' and for assignments below the 'Floor' attributed for the freight vendors.

In summary, the objective function is to minimize the sum of:

- MovementCost[m] * X[m] for each candidate movement slot m
- AssignmentPenalty[k, m] * A[k, m] for each pair of demand key k and candidate movement slot m
- UnfulfillmentPenalty[k] * U[k] for each demand key k
- UnderAssignmentPenalty[g] * F[g] for each soft floor group g

3.7 Scenario-based Framework Model

To execute the scenario-based framework simulation, we used the input of the data analysis and the linear integer programming model described in the methodology chapter. All variables identified as significantly different from the average through data analysis were incorporated into the model. The simulation time was set to run the linear optimization model described earlier a total of 20 times.

The inputs included in the model were deterministic and probabilistic data points. The deterministic inputs included:

- Loads necessary to move through the network
- Facility operating hours
- Facility capacity and staggering constraints
- Lanes available to move loads between facilities
- The carrier guide, service level agreement data (SLA) and the floor commitments.

The service level agreement (SLA) data provides valuable information regarding the agreed-upon transit time in each lane for every carrier. This data plays a crucial role in the current model adopted by Wayfair as it serves as a key indicator of the expected performance in each lane.

The probabilistic inputs included in the model, based on the distribution and statistic tests conducted across the selected lanes, were:

- Number of loads necessary to plan based on demand
- Lane travel times
- Inbound yard dwell time

Once the optimization model was run on the simulated inputs to generate several possible transportation plans, we assessed each plan against a simulated execution for the planned window and evaluated the results.

4. Results

This chapter provides an overview of the input data inserted to the simulation model. The input data is based on the probability distributions presented in the Methodology chapter. In addition, the chapter presents the simulation outcomes for three selected lanes, the corresponding model transportation plan output, and a comparison with the performance under the current model adopted by Wayfair.

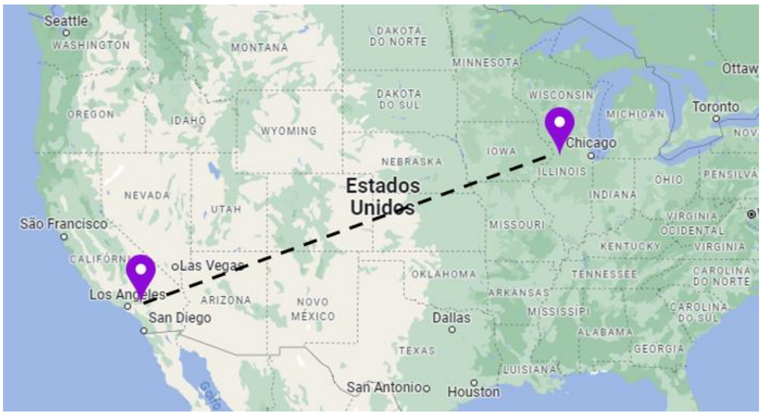
As selecting only one lane for the simulation could lead to bias, we decided to execute the simulation in three separate lanes. In the three selected lanes, we aimed to have examples covering conservative, optimistic and realistic planning approaches based on the service level agreements in place. Two of the lanes represent cross-country lanes, with long transit time, while the third lane represents a relatively short lane. The following subchapters will describe each lane and the outcome of the simulation.

4.1 Lane 1: Pacific Coast to the East North Central Area

For the first model run, we selected a lane that experiences a high volume of traffic and has a service level agreement for travel time provided by the main carrier of 110 hours. This lane stretches from the Pacific Coast to the East North Central Area in the United States covering around 2000 miles. This is a cross-country route, which makes it susceptible to several uncertainties. The carrier operating in this lane is identified as C1. The approximate locations are illustrated by Figure 13.

Figure 13

Lane 1 Representation

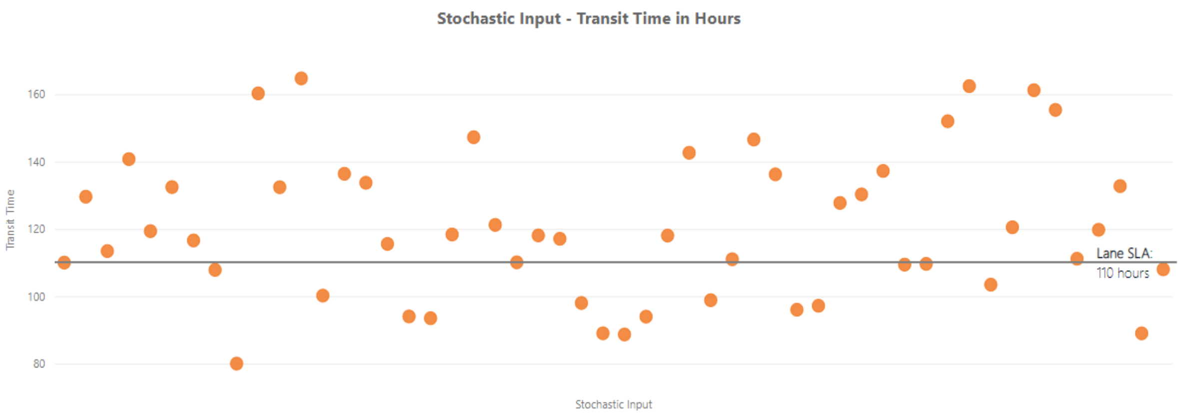


4.1.1 Simulation inputs for lane 1

As one of the stochastic inputs used by the model, we sampled the distribution of the transit times for Lane 1 considering the historical data for 2022. Figure 14 shows the new values incorporated by the model for the stochastic input travel time in hours. The service level agreement for transit time, which is used in the current model adopted by Wayfair, is highlighted in the figure as 110 hours. As captured by the image, most values based on the stochastic distribution registered transit time above the service level agreement.

Figure 14

Stochastic Input for Lane 1 - Transit Time in Hours



4.1.2 Current model for lane 1

As a first step, we used the linear integer programming model in the same way that is in place by Wayfair today. This model uses the service level agreement transit time to build an optimal transportation plan and calculate the expected performance. The optimal plan created at this stage was compared with the plan generated by the simulation. The outcome of the model, as displayed in Figure 15, involves 30 loads for carrier 1 covering the time span of 6 days.

Figure 15

Transportation Plan Based on SLA for Lane 1

Load #	Lane ID	Lane Name	Carrier ID	Local Cut Date Time	Local Arrival Date Time
0	L1	Pacific Coast - East North Central	C1	2023-03-19 - 10:00	2023-03-24 - 02:00
1	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-24 - 20:00
2	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-24 - 20:00
3	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-24 - 20:00
4	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-24 - 20:00
5	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-24 - 20:00
6	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-26 - 10:00
7	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-26 - 10:00
8	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-26 - 10:00
9	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-26 - 10:00
10	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-26 - 10:00
11	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-26 - 10:00
12	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-27 - 00:00
13	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-27 - 00:00
14	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-27 - 00:00
15	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-27 - 00:00
16	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-27 - 00:00
17	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-27 - 00:00
18	L1	Pacific Coast - East North Central	C1	2023-03-24 - 05:00	2023-03-28 - 07:00
19	L1	Pacific Coast - East North Central	C1	2023-03-24 - 05:00	2023-03-28 - 07:00
20	L1	Pacific Coast - East North Central	C1	2023-03-24 - 05:00	2023-03-28 - 07:00
21	L1	Pacific Coast - East North Central	C1	2023-03-24 - 05:00	2023-03-28 - 07:00
22	L1	Pacific Coast - East North Central	C1	2023-03-24 - 05:00	2023-03-28 - 07:00
23	L1	Pacific Coast - East North Central	C1	2023-03-24 - 05:00	2023-03-28 - 07:00
24	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-31 - 15:00
25	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-31 - 15:00
26	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-31 - 15:00
27	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-31 - 15:00
28	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-31 - 15:00
29	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-31 - 15:00

4.1.3 Simulation model for lane 1

The simulation model utilized historical data of interest and incorporated the distribution variables for travel time and yard dwell time. Instead of using predefined service level agreement values, the stochastic input was employed when executing the linear integer programming model described in the methodology section. The model was executed 20 times, resulting in the creation of 20 transportation plans. An example of one of these plans is illustrated in Figure 16.

Figure 16

Transportation Plan Based on Simulation for Lane 1

Load #	Lane ID	Lane Name	Carrier ID	Local Cut Date Time	Local Arrival Date Time
0	L1	Pacific Coast - East North Central	C1	2023-03-20 - 05:00	2023-03-25 - 00:45:00
1	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-25 - 05:46:12
2	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-25 - 05:46:12
3	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-25 - 05:46:12
4	L1	Pacific Coast - East North Central	C1	2023-03-21 - 10:00	2023-03-25 - 05:46:12
5	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-28 - 12:21:00
6	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-28 - 12:21:00
7	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-28 - 12:21:00
8	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-28 - 12:21:00
9	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-28 - 12:21:00
10	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-28 - 12:21:00
11	L1	Pacific Coast - East North Central	C1	2023-03-22 - 10:00	2023-03-28 - 12:21:00
12	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-28 - 06:45:00
13	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-28 - 06:45:00
14	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-28 - 06:45:00
15	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-28 - 06:45:00
16	L1	Pacific Coast - East North Central	C1	2023-03-23 - 05:00	2023-03-28 - 06:45:00
17	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
18	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
19	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
20	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
21	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
22	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
23	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
24	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
25	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
26	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
27	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12
28	L1	Pacific Coast - East North Central	C1	2023-03-25 - 05:00	2023-03-30 - 01:43:12

4.1.4 Comparison of current model and simulation output for lane 1

To compare the outcomes of the models, we focused on the expected dwell time for each of the transportation plans created. Dwell time refers to the duration that a shipment spends in the yard of a facility or distribution center after it arrives but before it is unloaded or processed further. Figure 17 illustrates the dwell time in hours using the current model adopted by Wayfair and the simulation results. As illustrated, the simulation model outcome suggests a higher total dwell time across all solutions if compared to the current model.

The dwell time distribution on Figure 18 represents the number of hours of dwell per unit for the solution following the service level agreement and for one of the transportation plans created as an output of the simulation model. The transportation plan generated by the simulation model provides a more realistic overview of the expected dwell time since it incorporates stochastic inputs. In other words, the plans created using the service level agreement values might be overly optimistic or overly favorable compared to the plans generated with the stochastic input, which takes into account the variability and uncertainty in travel time and yard dwell time.

Figure 17

Average Dwell Time by Simulation Scenario for Lane 1

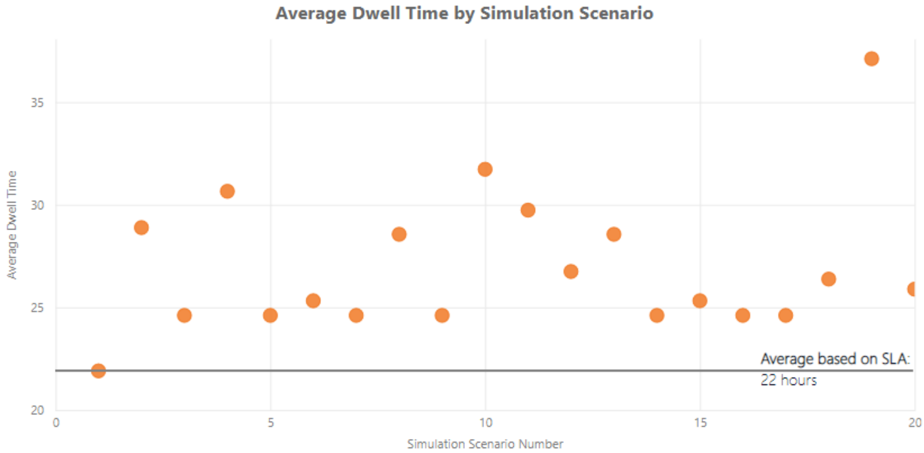
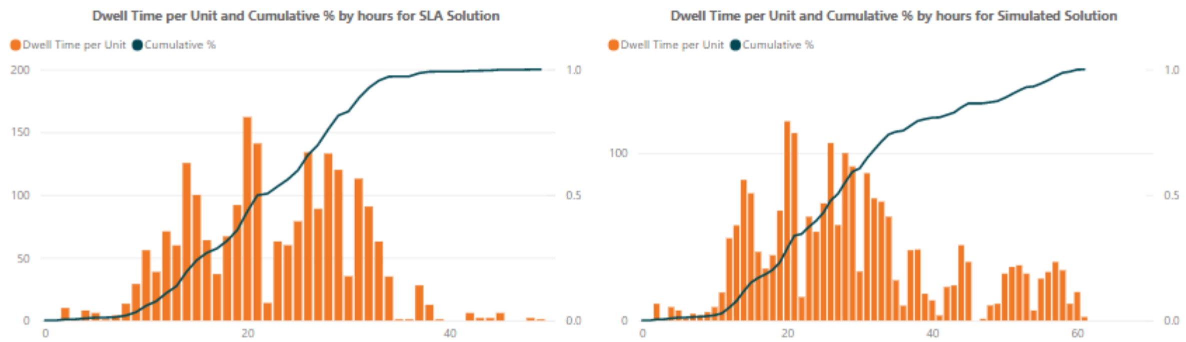


Figure 18

Dwell Distribution Comparison for Lane 1



4.2 Lane 2: Pacific Coast to the Mid-Atlantic Region

Lane 2 is a high-volume lane and has a service level agreement time provided by the main carrier of 153 hours. This lane stretches from the Pacific Coast to the Mid-Atlantic region in the United States and has a travel distance of approximately 2.7 thousand miles. The carrier operating in this lane is identified as C2. The approximate locations are identified by Figure 19.

Figure 19

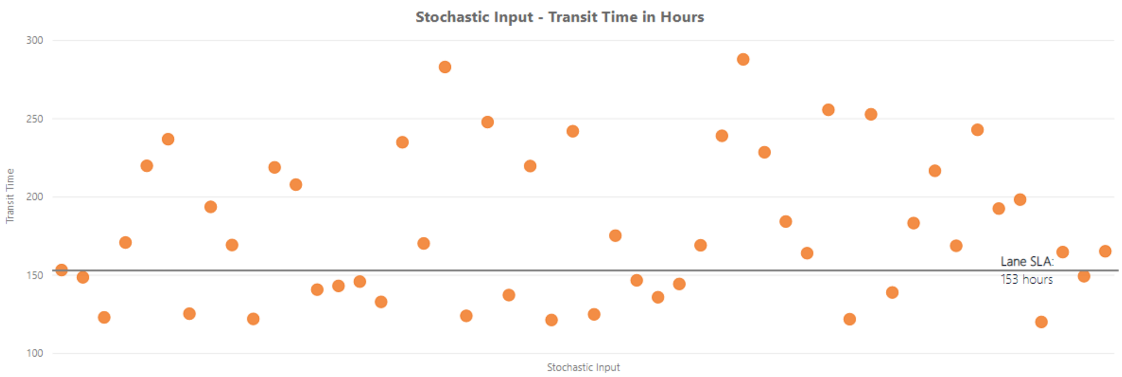
Lane 2 Representation

4.2.1 Simulation inputs for lane 2

For Lane 2, we utilized the distribution of transit times based on historical data from 2022 as a stochastic input for the model. Figure 20 displays the updated values for the travel time in hours, which were incorporated into the model. The service level agreement for transit time, which is currently utilized by Wayfair's model, is indicated in the figure as 153 hours. Similarly, to Lane 1, most of the values generated on the stochastic distribution exceeded the service level agreement.

Figure 20

Stochastic Input for Lane 2 - Transit Time in Hours



4.2.2 Current model for lane 2

Similarly to Lane 1, we developed the transportation plan for Lane 2 taking into account the service level agreements in place between Wayfair and the carrier. Figure 21 illustrates the outcome, which entailed 14 loads for carrier 2 spanning over 6 days.

Figure 21

Transportation Plan Based on SLA for Lane 2

Load #	Lane ID	Lane Name	Carrier ID	Local Cut Date Time	Local Arrival Date Time
0	L2	Pacific Coast - Northeastern Region	C2	2023-03-20 - 05:00	2023-03-27 - 05:00
1	L2	Pacific Coast - Northeastern Region	C2	2023-03-21 - 09:00	2023-03-28 - 13:00
2	L2	Pacific Coast - Northeastern Region	C2	2023-03-21 - 09:00	2023-03-28 - 13:00
3	L2	Pacific Coast - Northeastern Region	C2	2023-03-22 - 09:00	2023-03-29 - 14:00
4	L2	Pacific Coast - Northeastern Region	C2	2023-03-22 - 09:00	2023-03-29 - 14:00
5	L2	Pacific Coast - Northeastern Region	C2	2023-03-22 - 09:00	2023-03-29 - 14:00
6	L2	Pacific Coast - Northeastern Region	C2	2023-03-23 - 05:00	2023-03-30 - 15:00
7	L2	Pacific Coast - Northeastern Region	C2	2023-03-23 - 05:00	2023-03-30 - 15:00
8	L2	Pacific Coast - Northeastern Region	C2	2023-03-24 - 05:00	2023-04-01 - 00:00
9	L2	Pacific Coast - Northeastern Region	C2	2023-03-24 - 05:00	2023-04-01 - 00:00
10	L2	Pacific Coast - Northeastern Region	C2	2023-03-24 - 05:00	2023-04-01 - 00:00
11	L2	Pacific Coast - Northeastern Region	C2	2023-03-24 - 05:00	2023-04-01 - 00:00
12	L2	Pacific Coast - Northeastern Region	C2	2023-03-25 - 05:00	2023-04-01 - 23:00
13	L2	Pacific Coast - Northeastern Region	C2	2023-03-25 - 05:00	2023-04-01 - 23:00
14	L2	Pacific Coast - Northeastern Region	C2	2023-03-25 - 05:00	2023-04-01 - 23:00

4.2.3 Simulation model for lane 2

The simulation model analyzed the relevant historical data and used the stochastic input for travel time in the model for Lane 2. After executing the model 20 times, 20 transportation plans were created. Figure 22 provides an example of one of the transportation plans created.

Figure 22

Transportation Plan Based on Simulation for Lane 2

Load #	Lane ID	Lane Name	Carrier ID	Local Cut Date Time	Local Arrival Date Time
0	L2	Pacific Coast - Northeastern Regio	C2	2023-03-20 - 05:00	2023-03-25 - 13:07:48
1	L2	Pacific Coast - Northeastern Regio	C2	2023-03-21 - 09:00	2023-03-27 - 17:03:00
2	L2	Pacific Coast - Northeastern Regio	C2	2023-03-22 - 09:00	2023-03-27 - 14:01:48
3	L2	Pacific Coast - Northeastern Regio	C2	2023-03-22 - 09:00	2023-03-27 - 14:01:48
4	L2	Pacific Coast - Northeastern Regio	C2	2023-03-22 - 09:00	2023-03-27 - 14:01:48
5	L2	Pacific Coast - Northeastern Regio	C2	2023-03-22 - 09:00	2023-03-27 - 14:01:48
6	L2	Pacific Coast - Northeastern Regio	C2	2023-03-23 - 05:00	2023-03-29 - 02:01:48
7	L2	Pacific Coast - Northeastern Regio	C2	2023-03-23 - 05:00	2023-03-29 - 02:01:48
8	L2	Pacific Coast - Northeastern Regio	C2	2023-03-24 - 05:00	2023-03-29 - 12:43:12
9	L2	Pacific Coast - Northeastern Regio	C2	2023-03-24 - 05:00	2023-03-29 - 12:43:12
10	L2	Pacific Coast - Northeastern Regio	C2	2023-03-24 - 05:00	2023-03-29 - 12:43:12
11	L2	Pacific Coast - Northeastern Regio	C2	2023-03-24 - 05:00	2023-03-29 - 12:43:12
12	L2	Pacific Coast - Northeastern Regio	C2	2023-03-25 - 05:00	2023-03-30 - 21:34:12
13	L2	Pacific Coast - Northeastern Regio	C2	2023-03-25 - 05:00	2023-03-30 - 21:34:12
14	L2	Pacific Coast - Northeastern Regio	C2	2023-03-25 - 05:00	2023-03-30 - 21:34:12

4.2.4 Comparison of current model and simulation output for lane 2

As a next step, after running the simulation model 20 times, we compared the output of the transportation plans. Based on the simulation model, Figure 23 illustrates the yard dwell time in hours using the current model adopted by Wayfair and the simulation results. As seen for lane 1, the transportation plan created for lane 2 based on the service level agreements has also registered a lower expected average dwell time. For Lane 2, some of the simulation results suggest an average dwell time that is more than double the value expected on the service level agreements.

Figure 23

Average Dwell Time by Simulation Scenario for Lane 1

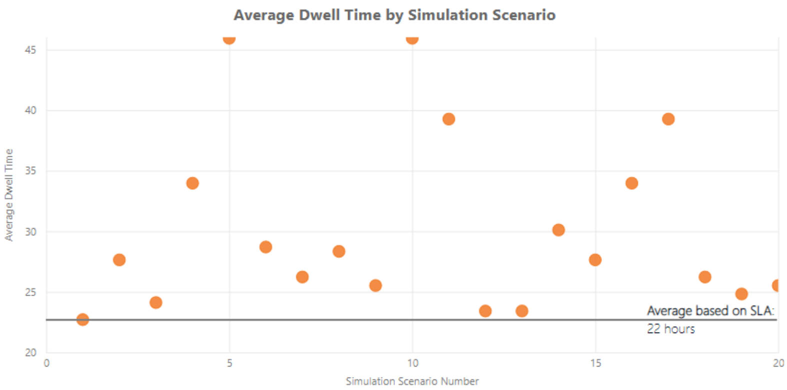
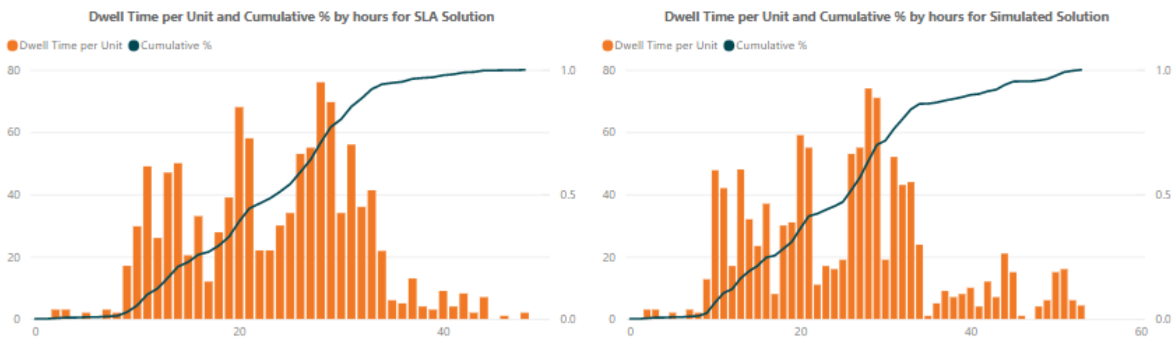


Figure 24

Dwell Distribution Comparison for Lane 2

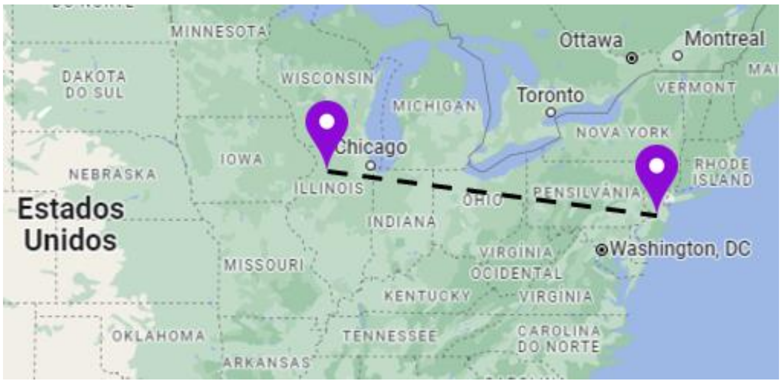


4.3 Lane 3: Mid-Atlantic to the Northeastern Region

Lane 3 is a shorter lane and has a service level agreement for travel time of 33 hours. This lane covers from the Mid-Atlantic to the Northeastern region in the United States and comprises the distance of around 900 miles. The carriers operating in this lane are identified as C3 and C4. The approximate locations are identified by Figure 25.

Figure 25

Lane 3 Representation

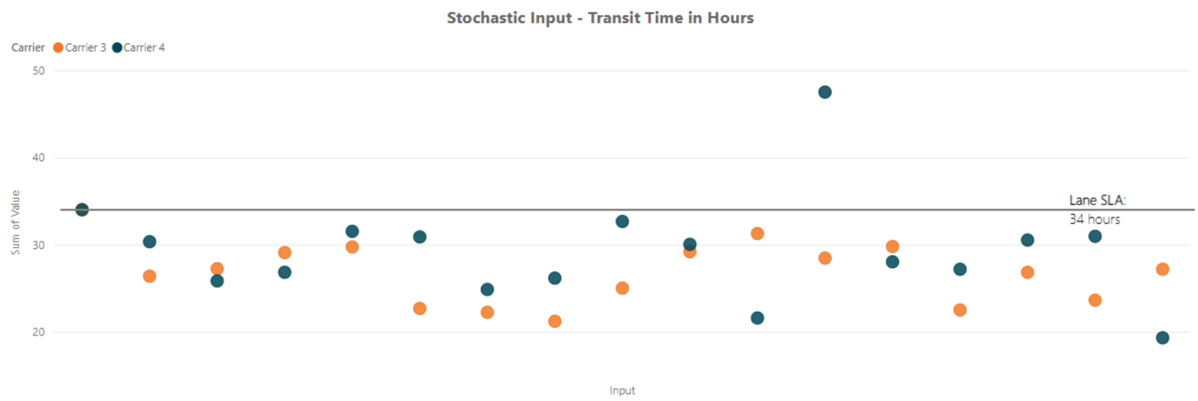


4.3.1 Simulation inputs for lane 3

As one of the stochastic inputs used by the model, we sampled the distribution of the transit times for Lane 3. Figure 26 shows the new values incorporated by the model related to the travel time in hours. The transit time established by the service level agreement, which is used in the current model, is highlighted in the chart as 34 hours. As there are two carriers operating in this lane, we sampled the distribution for each one of them. Unlike the distributions observed for lanes 1 and 2, the data points for lane 3 were mostly situated below the service level agreement.

Figure 26

Stochastic Input for Lane 3 - Transit Time in Hours



4.3.2 Current model for lane 3

The integer linear programming model was executed based on the service level agreements for Lane 3. Figure 27 illustrates the simulation's outcome, which entailed 14 loads for carrier 3 and 4 spanning over 6 days.

Figure 27

Transportation Plan Based on SLA for Lane 3

Load #	Lane ID	Lane Name	Carrier ID	Local Cut Date Time	Local Arrival Date Time
0	L3	Northeastern Region - East North Central	C3	2023-03-20 - 14:00	2023-03-21 - 23:00
1	L3	Northeastern Region - East North Central	C3	2023-03-20 - 18:00	2023-03-22 - 03:00
2	L3	Northeastern Region - East North Central	C3	2023-03-23 - 16:00	2023-03-25 - 01:00
3	L3	Northeastern Region - East North Central	C3	2023-03-23 - 16:00	2023-03-25 - 01:00
4	L3	Northeastern Region - East North Central	C3	2023-03-23 - 16:00	2023-03-25 - 01:00
5	L3	Northeastern Region - East North Central	C3	2023-03-23 - 16:00	2023-03-25 - 01:00
6	L3	Northeastern Region - East North Central	C3	2023-03-23 - 16:00	2023-03-25 - 01:00
7	L3	Northeastern Region - East North Central	C3	2023-03-25 - 19:00	2023-03-27 - 04:00
8	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 08:00
9	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 08:00
10	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 08:00
11	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 08:00
12	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 08:00
13	L3	Northeastern Region - East North Central	C4	2023-03-21 - 18:00	2023-03-23 - 03:00
14	L3	Northeastern Region - East North Central	C4	2023-03-23 - 20:00	2023-03-25 - 05:00

4.3.3 Simulation model for lane 3

The simulation model analyzed the relevant historical data and used the stochastic input for travel time in the model for Lane 3. After executing the model 20 times, 20 transportation plans were created. Figure 28 provides an example of one of the transportation plans created.

Figure 28

Transportation Plan Based on Simulation for Lane 3

Load #	Lane ID	Lane Name	Carrier ID	Local Cut Date Time	Local Arrival Date Time
0	L3	Northeastern Region - East North Central	C3	2023-03-20 - 02:00	2023-03-21 - 03:37
1	L3	Northeastern Region - East North Central	C3	2023-03-22 - 18:00	2023-03-23 - 19:07
2	L3	Northeastern Region - East North Central	C3	2023-03-22 - 18:00	2023-03-23 - 19:07
3	L3	Northeastern Region - East North Central	C3	2023-03-22 - 18:00	2023-03-23 - 19:07
4	L3	Northeastern Region - East North Central	C3	2023-03-22 - 18:00	2023-03-23 - 19:07
5	L3	Northeastern Region - East North Central	C3	2023-03-23 - 00:00	2023-03-24 - 01:07
6	L3	Northeastern Region - East North Central	C3	2023-03-23 - 11:00	2023-03-24 - 12:37
7	L3	Northeastern Region - East North Central	C3	2023-03-23 - 11:00	2023-03-24 - 12:37
8	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 00:37
9	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 00:37
10	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 00:37
11	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 00:37
12	L3	Northeastern Region - East North Central	C3	2023-03-25 - 23:00	2023-03-27 - 00:37
13	L3	Northeastern Region - East North Central	C4	2023-03-22 - 09:00	2023-03-23 - 17:37
14	L3	Northeastern Region - East North Central	C4	2023-03-23 - 15:00	2023-03-25 - 00:07

4.3.4 Comparison of current model and simulation output for lane 3

As a next step, after running the model 20 times, we compared the output of the transportation plans. Based on the simulation model, Figure 29 illustrates the yard dwell time in hours using the current model adopted by Wayfair and the simulation results. Unlike lane 1 and 2, lane 3 average dwell time is lower for most of the simulation scenarios compared to the service level agreement.

The transportation plan generated by the simulation model offers a more realistic perspective on the expected dwell time as it incorporates stochastic inputs. In contrast, the plans created using the service level agreement values depicted in Figure 30 might be overly pessimistic, meaning they could be more conservative or cautious in their estimations compared to the plans generated with the stochastic input. The stochastic input accounts for the variability and uncertainty in travel time and yard dwell time, providing a more comprehensive and reliable assessment.

Figure 29

Average Dwell Time by Simulation Scenario for Lane 3

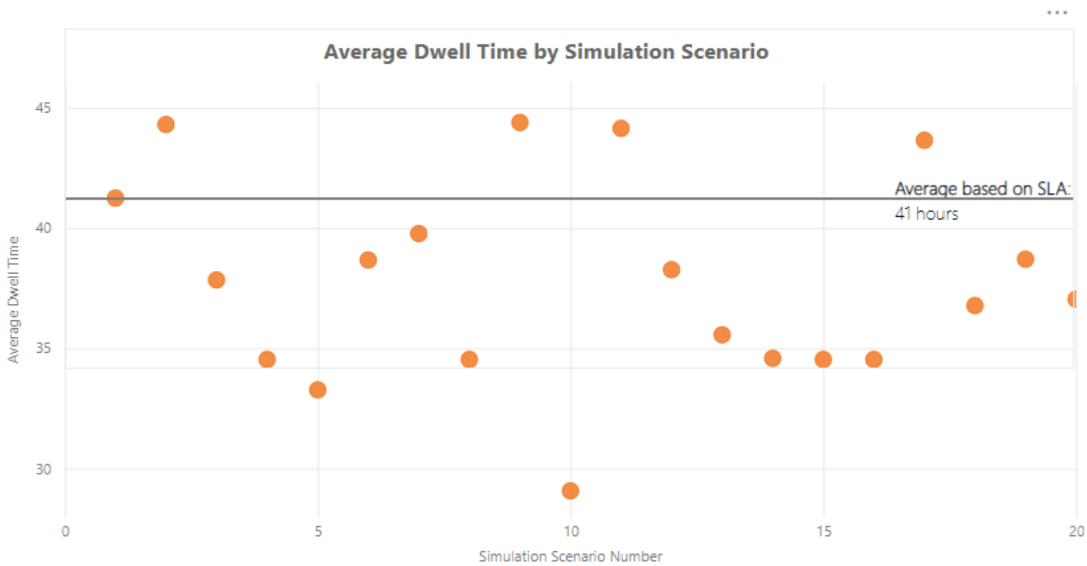
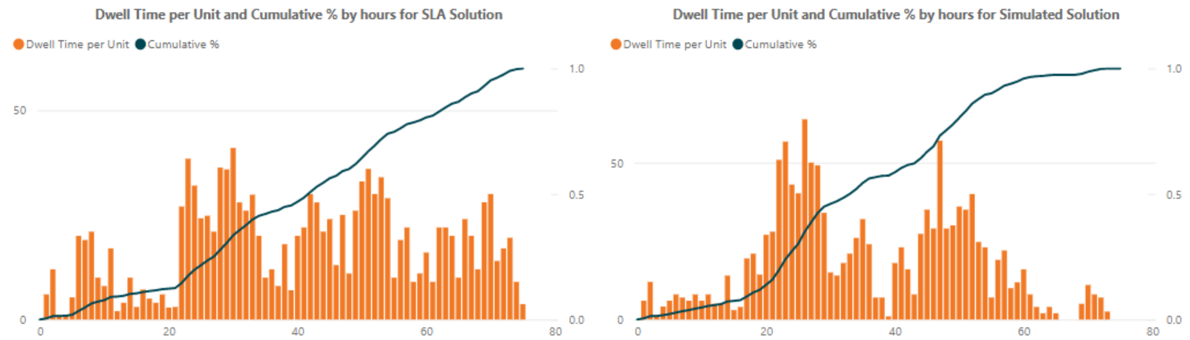


Figure 30

Dwell Distribution Comparison for Lane 3

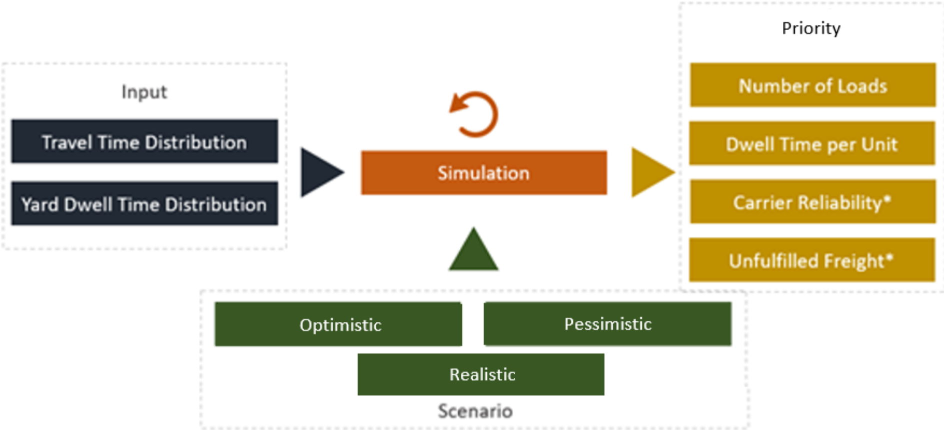


4.4 Framework for Output Evaluation

Upon generating simulation results, it becomes critical to assess and determine the output that exhibits optimal performance and can best serve the company's interests. To facilitate this decision, we have outlined a three-stage process comprising two preparatory stages and one post-simulation stage.

Figure 31

Framework for Selecting Transportation Plans



The initial stage of the proposed process involves identifying and selecting appropriate inputs for the simulation. It is clear by the dataset distribution that using service level agreement numbers as part of the transportation planning process is likely not reflecting the reality. Thus, the selection of input variables, such as travel time distribution and yard dwell time distribution, assumes a significant role in the decision-making process. By considering factors like exposure to potential disruptions and alterations to historical data, the company can make informed decisions regarding the input selection and range of historical data that should be used.

In addition to input selection, it is essential to flag potential scenarios that may impact transportation planning. Even under normal circumstances, the company may anticipate mild disruption, such as a storm or congestion. In such cases, it is recommended to consider the impact of those during the planning process and select appropriate data points that accurately reflect extreme conditions. One example would be to utilize data points that reflect longer travel times and design transport plans that are suited to such conditions when anticipating mild disruptions.

As seen on the three lanes evaluated in this capstone, it is possible to identify patterns in which the lane has a performance that is optimistic, pessimistic, or realistic compared to historical data. If these scenarios are identified, they can be incorporated in the model as indicators of potential bias or conservative approach for that specific lane.

The third stage involves making decisions based on the simulation results and selecting the goal. The selection of a plan must consider both qualitative and quantitative aspects. The number of loads and average yard dwell time per unit are directly linked to the costs incurred while executing the plan. Higher yard dwell time can lead to an increase in cost as units are not dispatched efficiently and are placed at the staging area. The number of loads also impacts the number of trucks required for transportation, which in turn affects costs. While carrier reliability and unfulfilled freight can affect delivery reliability metrics, this study did not focus on these aspects, and further research is recommended to incorporate them into the planning process.

5. Discussion

In this capstone project, we explored the application of stochastic optimization techniques to middle-mile transportation planning, specifically focusing on the incorporation of variance in transportation time and dwell. As a starting point, we engaged in an interview with a company representative, followed by the analysis of historical data. Based on our findings, we formulated a hypothesis suggesting that variations in time and dwell are significant factors. Subsequently, we conducted tests to validate and confirm our hypothesis. As an outcome, our findings suggest that accounting for these uncertainties in the planning process can lead to improved plans, while also protecting the organization from potential risks.

Our data analysis confirmed that transit time and yard dwell time distribution vary depending on the season and day of the week, which is crucial information for Wayfair's planning process. Currently, the company relies on service level agreements and point forecasts based on historical averages. This could lead to risk or suboptimal outcomes if average expected travel times are used for planning transportation.

Because of this variance, simulation results reveal differences between the current transportation plan outcomes and those generated by the scenario-based model. For example, Lane 1 shows considerably higher yard dwell times in the simulation compared to the service level agreement-based plan, whereas Lane 3 exhibits the opposite trend. This suggests that service level agreements might be overly optimistic or conservative for certain lanes.

Our research demonstrated that stochastic optimization models are capable of capturing the inherent variability in transportation time and dwell, which are often overlooked in deterministic models. By considering these uncertainties, our approach enabled more robust transportation plans that better reflect real-world conditions. This is particularly important in the context of middle-mile transportation, where delays and unforeseen events can have significant impacts on supply chain performance.

Interestingly, the updated model did not significantly impact truck utilization or allocation metrics, as simulation outcomes were similar to those based on service level agreements. Further investigation is needed to determine the reasons for this.

While our study offers valuable insights into the benefits of stochastic optimization for middle-mile transportation planning, it is important to note some limitations. First, the quality of the results is dependent on the accuracy of the input data, such as the estimated distributions of transportation time and dwell. Future research could focus on improving the quality of these estimates, potentially by incorporating real-time data and advanced analytics. Second, the computational complexity of stochastic optimization models can be a barrier for practical implementation, especially for large-scale transportation networks. Further research could explore ways to improve the efficiency of these models, making them more accessible and feasible for widespread adoption.

From a managerial perspective, we recommend that Wayfair incorporate the simulation approach for key lanes with critical variations. Adopting season-based expected transit times could lead to more accurate delivery time calculations for customers. Additionally, considering variations such as operations on different weekdays can be useful for internal operations planning.

6. Conclusion

Within this research project, we evaluated the application of stochastic optimization techniques to middle-mile transportation planning, with a focus on incorporating variance in transportation time and dwell. Our findings reveal that considering these uncertainties in the planning process can lead to significant improvements in plans, while also protecting the organization from potential risks. While extreme situations have been studied for resilience, our research focused on the day-to-day variations which we observed can also have a significant impact on the performance of plans.

6.1 Computational Results

While the adoption of stochastic optimization has proven to be a valuable tool for addressing the inherent variability and uncertainties associated with middle-mile transportation, the state of art for stochastic programming and optimization indicates that models involving a large and complex system, such as middle-mile logistics, can become very complex. This approach can lead to potential computational infeasibility and solutions that are not scalable. For this reason, we decided to adopt a scenario-based framework, using the existing linear integer programming approach developed by Wayfair and sampling distributions that can be used to investigate a variety of scenarios.

After analyzing 70.000 trips from January 2022 to January 2023, we selected three significant lanes and evaluated the distribution of the historical dataset. The key aspects that were evaluated were changes in transit time and yard dwell time depending on the day of the week, season and carrier performing the transport. We used the Kolmogorov-Smirnov test to identify statistically significant variations on the distribution, indicating which stochastic input should be incorporated in the model.

After sampling the distributions, we ran the integer linear programming model combining the stochastic inputs with deterministic inputs provided by Wayfair, which included the number of loads to move through the network, facility operating hours, capacity constraints and carrier guide service level agreements. The outcome of these transportation plans was compared to the existing transportation plans currently in use by Wayfair. The result of the comparison provides confirmation of the research question that taking into account the three sources of variance can potentially lead to better results. We found that the simulation outcome expected higher yard dwell time than the results incorporating only the average. It means that the current transportation plans might be more optimistic compared to reality. As a result, it can lead to a negative impact on the metric “customer promise”.

6.2 Managerial Insights

By offering a more accurate and robust representation of real-world scenarios, the model enables organizations to make better-informed decisions regarding resource allocation, route selection, and scheduling. These capabilities facilitate the development of more adaptive and resilient transportation plans, minimizing disruptions and associated costs while maximizing customer satisfaction.

Moreover, the insights gained from our research underscore the importance of striking a balance between speed, efficiency, and risk management in transportation planning. The scenario-based approach provides a comprehensive framework for understanding these trade-offs, empowering organizations to develop more effective strategies for managing uncertainty and ensuring supply chain success.

In summary, our capstone project has revealed the significant benefits of applying stochastic optimization to middle-mile transportation planning, showcasing its potential for creating more robust, adaptive, and risk-averse plans. By embracing this approach, organizations can better navigate the complexities and uncertainties of today's business environment, ensuring a more resilient and efficient supply chain that ultimately drives competitive advantage and long-term success.

6.3 Limitations and Future Research

While our study has demonstrated the potential of stochastic optimization for middle-mile transportation planning, further research is required to overcome the limitations identified, such as improving input data quality and enhancing computational efficiency. Future work should also explore the integration of real-time data and advanced analytics to further refine these models and facilitate their widespread adoption.

REFERENCES

Ahmed, S. (2010). Two-stage stochastic integer programming: A brief introduction. *Wiley encyclopedia of operations research and management science*, 1-10.

Amazon (2021, April 21st). *How Amazon's Middle-mile team helps packages make the journey to your doorstep.*https://www.amazon.science/latest-news/how-amazons-middle-mile-team-helps-packages-make-the-journey-to-your-doorstep.

Andersson, M., Berglund, M., Flodén, J., Persson, C., & Waidringer, J. (2017). A method for measuring and valuing transport time variability in logistics and cost benefit analysis. *Research in Transportation Economics*, 66, 59-69.

Baykasoğlu, & Subulan, K. (2019). A fuzzy-stochastic optimization model for the intermodal fleet management problem of an international transportation company. *Transportation Planning and Technology*, 42(8), 777–824. https://doi.org/10.1080/03081060.2019.1675316.

Birge, & Louveaux, F. (2011). *Introduction to Stochastic Programming.* Springer New York.

Bureau of Transportation Statistics. (2021). *Freight Shipments by Mode.* https://www.bts.gov/topics/freight-transportation/freight-shipments-mode.

Cao, X. (2007). *Stochastic Learning and Optimization : A Sensitivity-Based Approach (1. Aufl.)*. Springer US. https://doi.org/10.1007/978-0-387-69082-7.

Crainic, T. G., Gendreau, M., & Gendron, B. (Eds.). (2021). *Network Design with Applications to Transportation and Logistics*. Springer.

Delbart, T., Molenbruch, Y., Braekers, K., & Caris, A. (2021). Uncertainty in intermodal and synchromodal transport: Review and future research directions. *Sustainability*, 13(7), 3980.

Fu, C., Fu, C., & Michael, M. (2015). Handbook of simulation optimization. Springer.

Greening, L. M., Dahan, M., & Erera, A. L. (2022) Lead-Time-Constrained Middle-mile Consolidation Network Design with Fixed Origins and Destinations.

Grall-Maës, E. (2012). Use of the Kolmogorov–Smirnov test for gamma process. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 226(6), 624-634.

Hannah, L. A. (2015). Stochastic optimization. *International Encyclopedia of the Social & Behavioral Sciences*, 2, 473-481.

Kochan, C. G., & Nowicki, D. R. (2018). Supply chain resilience: a systematic literature review and typological framework. *International Journal of Physical Distribution & Logistics Management.*

Maggioni, Potra, F. A., & Bertocchi, M. (2017). A scenario-based framework for supply planning under uncertainty: stochastic programming versus robust optimization approaches. *Computational Management Science*, 14(1), 5–44. https://doi.org/10.1007/s10287-016-0272-3.

Naughton, K., & Boyle, M. (2019). Walmart Targets Automated 'Middle-Mile' Delivery to Cut Shipping Costs. *Transport Topics News* (19/06/2019 2019). https://www.ttnews.com/articles/walmart-targets-automated-middlemile-delivery-cut-shipping-costs.

Petroianu, L. P. G. (2020). *Exact and heuristic approaches to middle and last mile logistics*. University of Washington.

Powell, & Meisel, S. (2016). Tutorial on Stochastic Optimization in Energy-Part I: Modeling and Policies. *IEEE Transactions on Power Systems,* 31(2), 1459–1467. https://doi.org/10.1109/TPWRS.2015.2424974.

Powell. (2009). What you should know about approximate dynamic programming. *Naval Research Logistics,* 56(3), 239–249. https://doi.org/10.1002/nav.20347.

Sabouhi, Bozorgi-Amiri, A., & Vaez, P. (2021). Stochastic optimization for transportation planning in disaster relief under disruption and uncertainty. *Kybernetes*, 50(9), 2632–2650. https://doi.org/10.1108/K-10-2020-0632.

Seaton, H. E. (2018). The new e-commerce/home delivery retail distribution paradigm. *Journal of Transportation Management*, 29(1), 3.

Trebilcock, Bob (Host). (2022, April 12th). Planning Under Uncertainty. It's A New World for Supply Chain Planners. [Audio podcast episode]. In *247 Podcast.* SupplyChain247. https://www.supplychain247.com/article/planning_under_uncertainty._its_a_new_world_for_supply_chain_planners#.

Wang, J. Y. (2015). Resilience thinking in transport planning. *Civil Engineering and Environmental Systems*, 32(1-2), 180-191.

Wayfair. (2022, October 1). *About us.* https://www.aboutwayfair.com.

Wayfair. (2022). 2023 Investor Presentation. Retrieved from https://s24.q4cdn.com/589059658/files/doc_financials/2022/q4/2023-_-Investor-Presentation.pdf.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-39).