Assessing the Impact of Tight Delivery Time Windows on Last-Mile Route Efficiency

by

Peng Du
Bachelor of Law in Sociology, Jilin University, 2009

and

Miguel Fazzolari
Bachelor of Business Administration, Universidad Metropolitana, 2011

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 2025

Signature of Author: Peng Du
Department of Supply Chain Management
May 9, 2025

Signature of Author: ____________________
Department of Supply Chain Management
May 9, 2025

Certified by: ____________________
Dr. Selene Silvestri
Research Scientist
MIT Center for Transportation & Logistics
Capstone Advisor

Certified by: ____________________
Dr. Sarah Katharina Schauman
Postdoctoral Associate
MIT Center for Transportation & Logistics
Capstone Co-Advisor

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

Assessing the Impact of Tight Delivery Time Windows on Last-Mile Route Efficiency

by

Peng Du

and

Miguel Fazzolari

Submitted to the Program in Supply Chain Management
on May 9, 2025 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

ABSTRACT

With time windows becoming increasingly prevalent in E-commerce, last-mile delivery faces growing challenges in balancing customer satisfaction and operational efficiency across different geographical contexts. To address this, we leverage the Amazon Last-Mile Routing dataset along with additional real-world rural route data, utilize a solver built on OR-Tools to solve time window instances, and analyze the impact of time windows on route efficiency in both urban and rural areas. Our analysis identifies the different impact time windows have over urban and rural areas, particularly in vehicle use, where rural routes are more sensitive to shorter time window durations and higher time window frequencies. Our findings lead to the major conclusion that optimizing delivery time windows—favoring a smooth and balanced distribution and limiting narrow windows to under 20% of stops—enhances operational efficiency while balancing customer needs, particularly in rural context.

Capstone Advisor: Dr. Selene Silvestri

Title: Research Scientist, MIT Center for Transportation & Logistics

Capstone Co-Advisor: Dr. Sarah Katharina Schaumann

Title: Postdoctoral Associate, MIT Center for Transportation & Logistics

Table of Contents

1. Introduction

1.1 Background

With global e-commerce sales projected to grow from \$5.13 trillion in 2022 to \$8.09 trillion by 2028 (Lin, 2024), last-mile delivery — the final step in the e-commerce logistics process — has become critical for business success. Two key factors contribute to its importance: first, last-mile delivery plays a direct role in shaping the customer experience; and second, it accounts for a substantial portion of the overall logistics costs in e-commerce.

In recent years, e-commerce customers have been increasingly expecting fast, timely, and flexible options for last-mile delivery. Meeting these expectations is crucial for maintaining a competitive advantage, as customer satisfaction directly impacts brand loyalty and repeat business. Research shows that a 5% increase in customer retention rates leads to a profit increase of more than 25% (Gallo, 2014), highlighting the financial benefits of fulfilling customer expectations.

To meet customer demands and maintain a competitive edge, leading e-commerce companies like Amazon and Walmart have introduced delivery time windows, allowing customers to choose the time slot that best fits their schedules, thereby improving flexibility and boosting customer satisfaction. These time windows, which refer to specific time slots during which a delivery is scheduled to occur, can range from narrow options, such as a two-hour slot, to broader options, such as half-day or full-day periods.

Additionally, from an operational perspective, last-mile delivery is the most costly and time-consuming part of the logistics chain. According to Capgemini Research Institute (2019), last-mile delivery accounts for 41% of total logistics costs. A delivery route—an optimized itinerary for distributing parcels to various delivery points—is a critical element of the last-mile delivery process. Therefore, to remain profitable, it is crucial for companies to optimize their routes to achieve a high level of efficiency.

While striking a balance between customer satisfaction and operational efficiency is vitally important for business success, delivery time windows add significant complexity to last-mile operations, making it more challenging to achieve that balance. Companies must be flexible in offering delivery windows to enhance customer satisfaction while simultaneously maximizing route efficiency to ensure profitability and satisfy their shareholders. Failing to achieve this balance can lead to operational bottlenecks, customer dissatisfaction, reduced profitability, and, ultimately, impede business sustainability.

1.2 Problem Statement and Key Questions

Achieving a balance between offering flexible delivery windows and optimizing route efficiency can be challenging. Stricter time slots can complicate route planning and scheduling. Delivery operators may need to prioritize certain deliveries, which could lead to longer overall delivery times and increased costs due to route inefficiencies, higher fuel consumption, and the necessity for more vehicles or personnel to meet demand.

Additionally, the impact of time windows on route efficiency is influenced by the distinct operational dynamics of urban and rural areas. In urban environments, high population density (resulting in shorter travel distances between delivery points), traffic congestion, and infrastructure constraints significantly affect delivery routes. In contrast, rural areas face such challenges as longer distances between delivery points and limited road networks. When combined with time window constraints, these factors can affect route efficiency differently depending on the geographical context.

To find the right balance between delivery windows and route efficiency, it is essential for delivery operators to understand the impact of various time windows on the time and number of vehicles required for delivery across different geographical contexts. Such an understanding can provide insights into the delivery costs associated with different time window settings, and thus enable delivery operators to offer customers appropriate delivery time windows at affordable costs.

In this context, the questions to be answered in this project are:

- What is the impact of enforcing varying tight time-windows on the route efficiency (in terms of operational time to complete a route and number of vehicles used) of last-mile delivery?
- How do time windows affect delivery routes across different geographical regions, and what factors contribute to these variations?
- Based on the answer to the first two questions, how should route operators and management adjust their strategies for setting delivery time windows?

1.3 Project Goals and Expected Outcomes

As e-commerce continues to expand and customers' expectations for faster and more flexible deliveries grow, understanding the impact of varying time windows on delivery efficiency becomes increasingly complex. To effectively manage this challenge, a comprehensive analysis based on real-world data is crucial.

This project aims to analyze the impact of varying delivery time windows on last-mile delivery efficiency by utilizing the Amazon Last-Mile Routing Challenge Dataset (Amazon 2021), which includes urban routes in five metropolitan areas in the United States, along with supplementary rural route data. The analysis will involve generating instances of the Vehicle Routing Problem (VRP) and its variants, such as VRP with Time Windows (VRP-TW) and VRP with Multiple Time Windows (VRP-MTW), and solving them using a state-of-the-art solver.

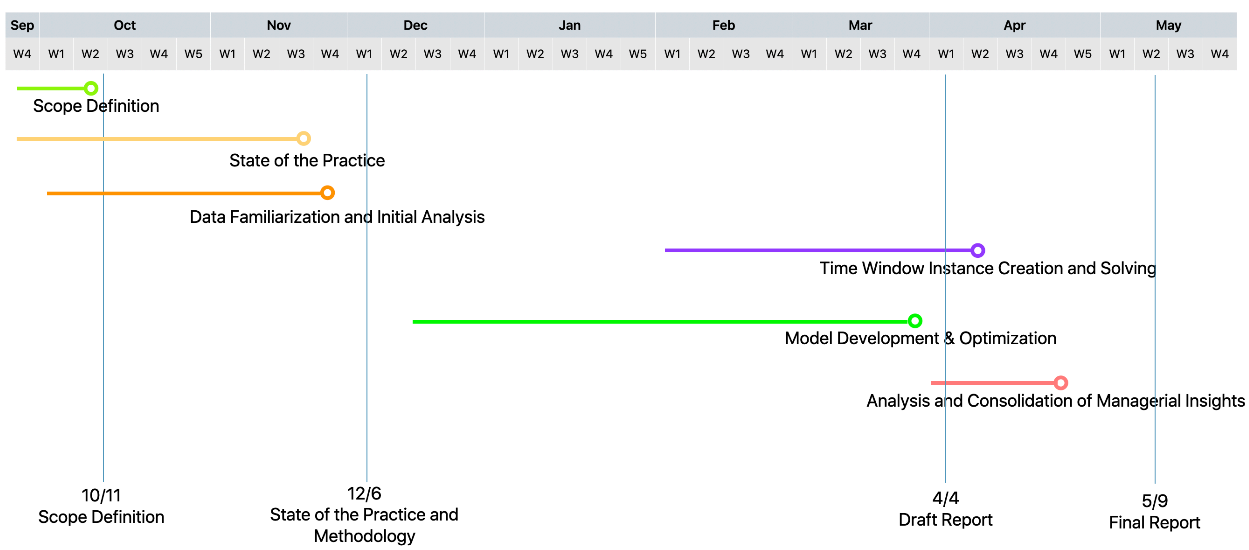
The key deliverables for this project are:

- VRP-TW solutions generated for various instances using a state-of-the-art solver, with detailed analysis of route efficiency.
- A summary of the findings that offer managerial insights to better support operational decision-making regarding delivery time window settings across different geographic contexts.

1.4 Plan of Work

Figure 1 presents the project timeline, outlining all key steps, followed by a detailed work plan for each phase.

Figure 1: Project timeline



Detailed Plan:

1. Scope Definition
   - Background
   - Problem statement and key questions
   - Project goals and expected outcomes

2. State of the Practice
    - Literature review
    - Collect and analyze real-world examples of how similar projects or problems have been addressed
    - Based on findings, suggest ways this capstone can build on or improve previous projects
3. Data Familiarization and Initial Analysis
    - Data collection: acquire datasets required for analysis
    - Data cleaning and extraction: remove inconsistencies and standardize formats
    - Initial study on the datasets
4. Model (Solver) Development and Optimization
    - Build a VRP-TW solving model and optimize the settings to accommodate to the specific needs of this project
5. Time Window Instances Creation & Solving
    - Create VRP-TW instances by varying key parameters, such as time window durations and the percentage of stops requiring time windows
    - Solve VRP-TW instances using the model and collect results focusing on route efficiency
6. Analysis and Consolidation of Managerial Insights
    - Analyze results to uncover trends regarding how time window configurations affect route efficiency
    - Compare rural and urban route performance to identify geographical differences and trends
    - Make conclusion and provide suggestions

2. State of the Practice

The focus of this work is on the vehicle routing problem with multiple time windows (VRP-MTW), a well-known logistics problem involving the determination of the most efficient routes for a fleet of vehicles to service a set of customer locations while adhering to specific time constraints at each stop. The main goal is to analyze and compare the effects of time windows on delivery efficiency in different geographic contexts that include urban and rural areas. Driven by the need to improve the efficiency and cost-effectiveness of logistics operations, meet customer demands, and address geographic-specific challenges, we review the literature on the Vehicle Routing Problem, Time Windows, and their interaction.

2.1 Vehicle Routing Problem

Dantzig and Ramser introduce the Vehicle Routing Problem (VRP) in their paper titled "The Truck Dispatching Problem" in 1959. VRP is a combinatorial optimization problem that aims to determine the most efficient routes for a fleet of vehicles to deliver goods to customers while minimizing total routing costs.

In 1964, Clarke and Wright improve on Dantzig and Ramser's approach with the Savings algorithm, a greedy algorithm that remains widely used today. The core of the Clarke and Wright Savings algorithm is calculating savings measured by the reduction in mileage and time achieved by linking existing nodes, with routes created based on the largest savings value.

The development of exact algorithms for the Vehicle Routing Problem (VRP) starts in 1981 with key papers by Christofides, Mingozzi, and Toth, introducing dynamic programming and mathematical formulations using q-paths and k-shortest spanning trees. A few years later, in 1985, Laporte, Desrochers, and Nobert propose the first cutting plane approach using linear relaxation. These early methods influence later algorithms, with various mathematical programming formulations utilizing vehicle flow, commodity flow, and branch-and-cut techniques. Notable implementations by Fukasawa et al. (2006) and Baldacci et al. (2008) are based on set partitioning models.

The 1990s saw the rise of modern heuristics, particularly metaheuristics, for solving the VRP. Early research focused on tabu search approaches, with some algorithms being overly complex. However, recent developments have led to more rationalized methods. The best metaheuristics combine broad and deep solution searches, often using multiple operators like adaptive large neighborhood search (ALNS) or hybrid genetic algorithms, such as the one proposed by Vidal et al. (2012). These approaches are effective across various VRP variants.

These advancements provide the foundation for this project's use of a state-of-the-art solver to optimize VRP-TW instances in different contexts.

2.2 Effect of Time Windows on Delivery Routes

Although narrow delivery time slots are convenient for the customer, they significantly reduce routing efficiency. Experiments conducted by Agatz et al. in their paper Time Slot Management in Attended Home Delivery (2011) indicate an increase in delivery cost of up to 25% from using two-hour time slots instead of time slots that span an entire morning or afternoon. This illustrates the core trade-off between service and delivery costs.

On the other hand, the economic loss resulting from waiting for service at home is estimated at $38 billion in the United States in 2011 (Ellis, 2011). This dual perspective—higher operational costs for narrow slots and economic losses for customers—highlights the need for balanced time window strategies.

Recent studies, such as those by Vareias et al., 2020; and Hungerländer et al., 2018; explore different approaches to managing time windows. Vareias et al. propose a flexible strategy that restricts available time slots as new requests come in, while Hungerländer et al. aim to maximize customer choice to encourage orders. These strategies showcase innovative approaches to balancing customer satisfaction with operational efficiency.

Additionally, some studies investigate how pricing policies can manipulate customer preferences to reduce the concentration of requests during peak times. The literature on this topic is extensive and growing, with more than 12 recent contributions focusing on time slot management since 2011, e.g., Vinsensius, Wang et al. (2020); and Archetti and Bertazzi (2020). A key challenge identified is the development of pricing strategies that not only influence customer demand and balance delivery slot availability, but also align with the routing costs involved in distribution.

Encouraging customers to select more than one time window has been shown to significantly reduce costs. For instance, Schaap et al., 2022 suggest that a "2×2h" strategy, which offers the customer a choice between two different time slots, each lasting 2 hours, achieves a balanced trade-off between cost efficiency and customer satisfaction, while broader time window offerings could further lower operational expenses. Strategies that provide limited choices for delivery time windows effectively minimize cost increases while maintaining high levels of customer satisfaction. To strike the right balance between profitability and customer satisfaction, it is crucial to offer customers a small, well-defined set of delivery time window options (Hendrik Schaap, Maximilian Schiffer, Michael Schneider, Grit Walther, 2022).

2.3 Summary of findings

The Vehicle Routing Problem (VRP) has been a significant area of research since its introduction by George Dantzig and John Ramser in 1959. Over the decades, various advancements have been made, including the Clarke and Wright Savings algorithm (1964) and the development of modern heuristics like metaheuristics in the 1990s. More recently, researchers such as Niels Agatz, Ann Campbell, Moritz Fleischmann, and Martin Savelsbergh (2010) have focused on VRP with time window constraints, while Claudia Archetti and Lucca Bertazzi (2020) explored strategies to improve delivery efficiency through effective time window management.

Agatz et al., 2010 and Ellis, 2011 give us an example of why it is so important for delivery service providers and customers to strike a healthy balance in time window offerings that works for both sides. Then in 2022, Schaap et al., came up with the “2x2h” strategy which, in our experience in this field, has become the standard where customers are satisfied and costs are under control for delivery service providers.

However, to the best of our knowledge, no studies have specifically highlighted the different challenges and opportunities associated with VRP-MTW in rural versus urban settings based on analysis of a large real-world dataset. This gap presents an opportunity to analyze how time window constraints impact route efficiency, customer satisfaction, and operational strategies across contrasting geographical contexts.

3. Methodology

In this section, we outline the methodology used to analyze the impact of time window constraints on routing efficiency across urban and rural contexts. Leveraging this methodology, we uncover trends, differences, and operational challenges that serve as a foundation for actionable insights.

The methodology of this project follows six key steps. First, we define urban and rural contexts to establish clear criteria for categorizing delivery routes based on their geographical and operational characteristics. Next, we analyze and extract relevant data from the Amazon Last-Mile Routing Challenge Dataset, along with additional real-world rural route data. To support the rural analysis, we utilize an open-source routing tool to identify travel times (time required to travel between two locations) between stops within a rural route. We then generate a variety of time window instances to represent different delivery constraints and scenarios. In parallel, a solver is developed and applied to optimize routing solutions based on time window configurations. Finally, we evaluate the performance of these solutions and conduct a detailed analysis of the results, quantifying the impact of time window variability on operational efficiency.

The following sections provide a detailed explanation of each step.

3.1 Definition of Urban and Rural Contexts

According to the U.S. Census Bureau, a rural area is located outside of a U.S. Census-designated urban area with a population of 50,000 or more (U.S. Census Bureau, 2010). Based on this definition, a rural route involves deliveries within a community with fewer than 50,000 residents, while an urban route involves deliveries within a community with 50,000 residents or more. In practice, one of the authors, with experience managing delivery operations, notes that rural routes are typically defined by service providers with these characteristics:

- The presence of mostly dirt roads
- Long driving distances between pick-up and delivery stops, with 3+ minutes between stops on average
- Mostly residential pick-up and delivery stops
- Areas with fewer than 50,000 residents

While urban routes are characterized by:

- The presence of mostly paved roads
- Short driving distances between pick-up and delivery stops, with less than 3 minutes between stops on average
- The presence of both commercial and residential pick-up and delivery stops
- Areas with 50,000 or more residents

In this project, we will adopt the above definition of rural and urban routes, as it provides a practical framework incorporating road types, stop distances, and delivery point characteristics. This detailed approach enables a more accurate analysis of the VRP-TW problem across different geographical contexts, ensuring alignment with real-world logistics practices.

3.2 Dataset Analysis and Extraction

Half of the data for this project is sourced from the Amazon Last-Mile Routing Challenge Dataset, which comprises 6,112 training routes and 3,072 evaluation routes across five major metropolitan areas in the United States. The dataset includes route-level and package-level details, along with travel times between stops for each route.

Table 1 provides an overview of the route and package information included in the Amazon training dataset (the source of urban route data for project), showing an average of 148 stops per route. As the Amazon dataset contains extensive information for each route, we extract only the relevant data, including route IDs, stops in each route, stop type (station or drop-off), and location of each stop (longitude / latitude). These data, along with the travel times between stops for each route, make up the complete dataset needed for our analysis. The original time windows in the dataset are ignored, as we generate new instances tailored to our specific purpose.

Table 1. Amazon Last-Mile Routing Challenge Dataset - Training Data Overview

City	Route	Stop	Package ID	With Time Window	Stops per Route	Time Window per Route
Seattle	1,079	155,781	249,456	18,955	144	18
Austin	214	31,274	50,918	4,570	146	21
Boston	929	140,622	213,888	17,555	151	19
Chicago	1,002	162,410	252,541	18,314	162	18
Los Angeles	2,888	414,440	690,372	54,599	144	19
SUM	**6,112**	**904,527**	**1,457,175**	**113,993**	**148**	**19**

Since all routes in the Amazon dataset originate from five major metropolitan areas, we identify 10 additional rural routes managed by one of the authors of this capstone project to study the impact of time windows on rural route efficiency. Geographically, these routes are all located in the vicinity of Austin, Texas. Similar to our work with the Amazon dataset, we extracted the Route ID, stops within a route, the stop type (station or drop-off), and the location of each stop (longitude and latitude). However, the travel times between stops are not available in the original rural dataset; this necessitates that we find a way to obtain them.

3.3 Estimating Travel Times for Rural Route Stops

To identify the travel times between stops within a rural route, we deploy The Open Source Routing Machine (Project OSRM, n.d.), a fast and flexible routing engine that uses OpenStreetMap (OSM) data. OSRM efficiently computes optimal routes and travel times between geographic coordinates. Here's an elaboration of how the process works:

- **Preprocessing:** When OSRM is deployed, it processes raw OSM data through extraction, partitioning, and customization steps to convert it into a routing graph. This graph contains nodes (representing intersections or points along roads) and edges (road segments) with associated attributes such as distance and travel speed. The truck profile (truck.lua) defines default speeds, turn penalties, and other parameters that affect how travel times are computed along each edge.
- **Graph Construction:** OSRM builds an edge-based graph from the node-based graph. In this process, the engine calculates segment travel times by dividing the segment's distance by the vehicle's speed and adding any penalties, such as delays for turns or intersections. Here, "penalties" may include delays for turns, U-turns, or waiting at intersections, as defined in the truck profile.

- **Routing Calculation:** OSRM uses its routing algorithm on the preprocessed graph to compute the shortest (time-optimal) path between every pair of input coordinates by summing the travel times of each segment along the computed route.
- **Assembling the Duration Matrix:** Once the time-optimal routes are calculated, OSRM produces a 2D array (duration matrix) where each element represents the total travel time from one stop to another.

Table 2 displays the parameters we adopt to get the close-to-real-world travel time estimates. These settings enable us to approximate the travel time between any two stops within a route with a high degree of realism.

Table 2. OSRM Parameter Settings

Parameter	Value	Description
Default speed	11 m/s	Uses 11 m/s (~25 mph) as the truck speed when no specific road speed limit is available.
Max speed for map matching	33.3 m/s	Caps truck speed at 33.3 m/s (74.56 mph) during map matching to avoid unrealistic routes.
Routing priority	Routability	Prioritizes roads and turns based on truck suitability over just distance or time.
U-turn penalty	30 seconds	Adds 30 seconds for U-turns to account for the extra time required.
Traffic light penalty	5 seconds	Adds 5 seconds at traffic lights to reflect truck stopping and starting delays.
Side road multiplier	0.70	Reduces speed by 30% on narrow side roads to discourage their use by trucks.
Turn penalty	12 seconds	Adds 12 seconds for each turn to account for the time it takes.
Speed reduction	0.80	Lowers truck speed by 20% to reflect real-world limits like acceleration.
Turn bias	1.08	Slightly favors straighter routes by penalizing sharp turns more heavily.

3.4 Time Window Instances Generation

To conduct experiments for a variety of time window scenarios across urban and rural contexts, we select a total of 20 routes—10 from urban areas and 10 from rural areas. Based on these routes, we generate 2,700 instances—1,350 for each context—by varying three key parameters, as detailed in Table 3.

Table 3. Time Window Instances Configuration

Parameter	Settings	Description
Time Window Frequency	5%, 20%, 40%	Proportion of stops with time constraints, simulating light to heavy loads from 5% to 40%.
Duration of Each Time Window	1 hour, 2 hours, 4 hours	Length of time windows, where shorter durations (1 hour) add constraints and longer ones (4 hours) offer flexibility, reflecting common practices.
Distribution of Time Windows	**Random**: 8:00 AM–6:00 PM **All in the Morning:** 8:00 AM–12:00 PM **40-20-40:** 40% at 8:00 AM–12:00 PM, 20% at 12:00 PM–2:00 PM, 40% at 2:00 PM–6:00 PM	Patterns for assigning time windows, with random spread across the day, all morning within a 4-hour block, or split as 40-20-40 across three periods.

- **Time Window Frequency—Percentage of Stops Requiring Time Windows**: This determines the proportion of stops within a route that have time constraints. We adopt three levels—5%, 20%, and 40%—to simulate a range of real-world time window loads, from light to heavy.
- **Duration of Each Time Window**: The length of the time windows impacts operational flexibility and routing complexity, with shorter durations adding more constraints and longer ones providing greater flexibility. We use three durations—1 hour, 2 hours, and 4 hours—to replicate common real-world practices.
- **Distribution of Time Windows**: This parameter defines how time windows are distributed throughout the day. We include three patterns:
  - Random Distribution: Time windows are randomly assigned between 8:00 AM and 6:00 PM.
  - All-in-the-Morning: All time windows fall within the 8:00 AM–12:00 PM period.
  - 40-20-40: 40% of time windows occur between 8:00 AM–12:00 PM, 20% between 12:00 PM–2:00 PM, and 40% between 2:00 PM–6:00 PM.

By combining these parameters, we construct 540 unique scenarios, corresponding to 20 routes across three frequency levels, three time window durations, and three distribution patterns. To reduce randomness and enhance the robustness of experiments, we generate five distinct randomized instances per scenario, resulting in a total of 2,700 instances.

3.5 Model Development and Optimization

To solve the 2,700 instances of vehicle routing problem with time windows, we build a solving model (solver) with Python. This solver is based on Google OR-Tools, a free software package designed for optimization, which can be customized to address vehicle routing problems effectively (Google, n.d.).

Several key settings are tailored to our specific VRP-TW setup, as detailed in the following paragraphs.

The objective of the solver is to cut down the total operational time—that is, the sum of travel time, waiting time (time a vehicle spends idle at a stop before a time window opens), and service time (time spent at each stop, like 1 minute for drop-offs). The service time at each stop is fixed at 1 minute for every drop-off and thus treated as a constant across all instances, with no impact on the optimization objective.

For the time dimension setup, we define a "Time" dimension to track the total time along each route. With a callback function that adds up the travel time between stops, we allow waiting at stops by setting a slack of 9999 minutes, meaning a vehicle can wait if it arrives early. The solver enforces time windows by setting constraints on cumulative time variables, ensuring vehicles arrive within each stop's allowed time window (such as 8 AM to 9 AM). For the depot (station), we lock the time window to the operational hours of 6 AM to 10 PM (360 to 1, 320 minutes), and vehicles can start anytime in that window.

The vehicle setup and penalties are configured for up to five vehicles, all starting and ending at the depot. To avoid using more vehicles than needed, we add a penalty for using extra vehicles. The first vehicle has a fixed cost of 0, but every additional vehicle gets a penalty equivalent to 500 minutes to align with the solver's time-based objective. This pushes the solver to use as few vehicles as possible, ideally just one, unless it really needs more to cover all stops within the time windows.

To minimize total operational time across all vehicles, the solver is set to reduce the total operational time, which includes travel time between stops, waiting time if a vehicle arrives early, and service time at each stop. To achieve this, we set the arc costs between stops to zero, ensuring the solver does not focus on travel time alone. Instead, it works to lower the combined time—travel, waiting, and service—for all routes (1 to 5 depending on how many vehicles are used for an instance), aiming to keep the overall time spent by all vehicles as low as possible.

In terms of search strategy, the solver uses a two-step search approach to find good solutions efficiently. First, it builds an initial solution using the Cheapest Insertion strategy, a widely adopted heuristic known for its ability to quickly assign stops to routes across all vehicles at minimal cost (Solomon, 1987). Then, it improves the solution using the Guided Local Search method, which dynamically adjusts penalties to escape local minima and avoid suboptimal solutions during the search (Kilby et al., 1999).

With regard to vehicle start times, we allow flexible scheduling. Vehicles do not have to start at a fixed time; instead, their start times can be anywhere between 6 AM and 10 PM, which the solver decides based on what minimizes the total time. This flexibility helps fit routes into tight time windows more easily.

Lastly, another important setting for the solver is the time limit for solving. Ideally, the longer the solving time, the better the quality of the results we can get. However, considering both the practicality of running a large number of experiments and the time constraints of solving VRP-TW in real-world business practices, we conduct a time limit test to determine the most cost-effective time limit for solving the 2,700 instances. To run this test we use the solver to tackle 144 instances—72 each for urban and rural routes—under eight different time limits: 60 seconds, 150 seconds, 300 seconds, 600 seconds, 900 seconds, 1,200 seconds, 1,500 seconds, and 1,800 seconds. Table 4 shows the improvement in total operational time (sum of travel, waiting, and service time) for each longer time limit compared to a 60-second baseline in 144 instances. It indicates an average improvement of 1.52% at 600 seconds (10 minutes), while 1,800 seconds (30 minutes) yields an average improvement of 1.97%. Table 5 presents the improvement in vehicle use for each longer time limit compared to a 60-second baseline in 144 instances. It demonstrates an average improvement of 2.78% at 600 seconds (10 minutes), while 1,800 seconds (30 minutes) yields an average improvement of 3.47%. Figure 2 converges the improvement in operational time and vehicle use, highlighting that 600 seconds lies at a balanced point of the curve. Based on these results, we set 600 seconds (10 minutes) as the time limit for solving the 2,700 instances to balance solver efficiency and solution quality.

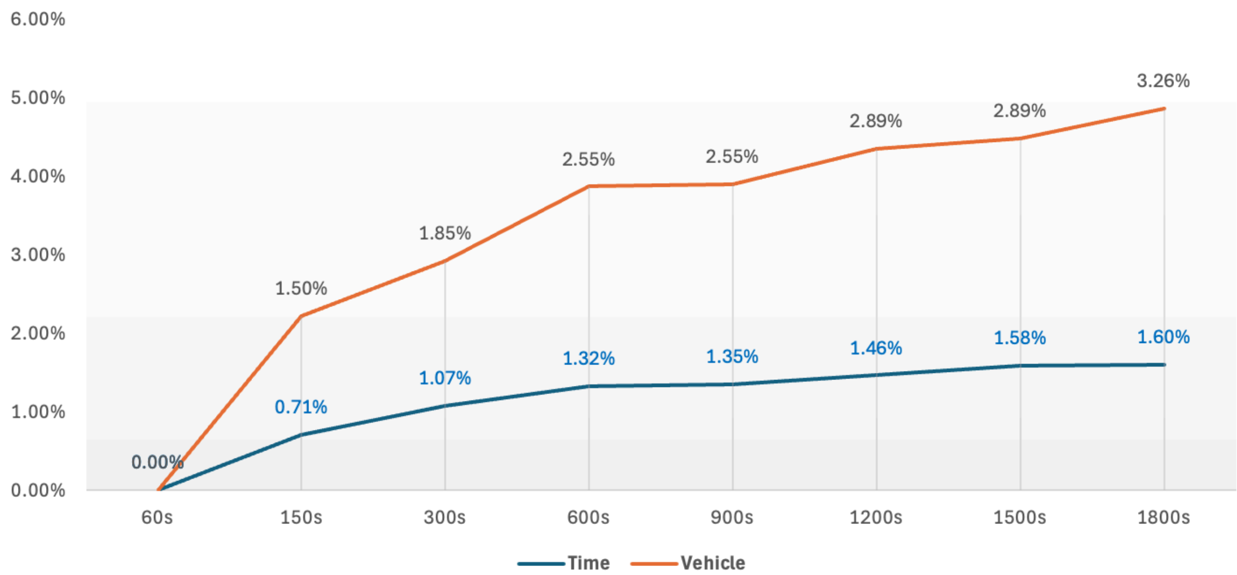
Table 4. Percentage Improvement over 60-Second Baseline – Operational Time

TW Duration - Frequency	60s	150s	300s	600s	900s	1200s	1500s	1800s
1 Hour	**0.00%**	**1.03%**	**1.29%**	**1.50%**	**1.53%**	**1.56%**	**1.69%**	**1.70%**
TW - 5% (Instance 1)	0.00%	0.50%	0.71%	0.66%	0.61%	0.68%	1.06%	1.07%
TW - 20% (Instance 5)	0.00%	1.55%	1.75%	1.91%	1.91%	1.92%	1.92%	1.89%
TW - 40% (Instance 4)	0.00%	1.03%	1.40%	1.94%	2.07%	2.09%	2.09%	2.13%
2 Hours	**0.00%**	**0.58%**	**1.09%**	**1.40%**	**1.45%**	**1.65%**	**1.69%**	**1.67%**
TW - 5% (Instance 5)	0.00%	0.30%	0.53%	0.84%	0.93%	0.99%	0.99%	1.00%
TW - 20% (Instance 3)	0.00%	0.37%	1.12%	1.23%	1.23%	1.31%	1.32%	1.33%
TW - 40% (Instance 1)	0.00%	1.07%	1.64%	2.12%	2.19%	2.66%	2.76%	2.68%
4 Hours	**0.00%**	**0.52%**	**0.83%**	**1.06%**	**1.07%**	**1.16%**	**1.37%**	**1.43%**
TW - 5% (Instance 3)	0.00%	0.56%	0.86%	0.92%	0.92%	1.02%	1.09%	1.14%
TW - 20% (Instance 1)	0.00%	0.38%	0.51%	0.81%	0.75%	0.79%	1.20%	1.23%
TW - 40% (Instance 5)	0.00%	0.61%	1.14%	1.44%	1.55%	1.68%	1.82%	1.91%
Aggregated Average	**0.00%**	**0.71%**	**1.07%**	**1.32%**	**1.35%**	**1.46%**	**1.58%**	**1.60%**

Table 5. Percentage Improvement over 60-Second Baseline – Vehicles Use

TW Duration - Frequency	60s	150s	300s	600s	900s	1200s	1500s	1800s
1 Hour	**0.00%**	**3.82%**	**3.82%**	**4.86%**	**4.86%**	**4.86%**	**4.86%**	**4.86%**
TW - 5% (Instance 1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TW - 20% (Instance 5)	0.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
TW - 40% (Instance 4)	0.00%	3.13%	3.13%	6.25%	6.25%	6.25%	6.25%	6.25%
2 Hours	**0.00%**	**0.69%**	**0.69%**	**1.74%**	**1.74%**	**2.78%**	**2.78%**	**3.89%**
TW - 5% (Instance 5)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TW - 20% (Instance 3)	0.00%	2.08%	2.08%	2.08%	2.08%	2.08%	2.08%	2.08%
TW - 40% (Instance 1)	0.00%	0.00%	0.00%	3.13%	3.13%	6.25%	6.25%	9.58%
4 Hours	**0.00%**	**0.00%**	**1.04%**	**1.04%**	**1.04%**	**1.04%**	**1.04%**	**1.04%**
TW - 5% (Instance 3)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TW - 20% (Instance 1)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
TW - 40% (Instance 5)	0.00%	0.00%	3.13%	3.13%	3.13%	3.13%	3.13%	3.13%
Aggregated Average	**0.00%**	**1.50%**	**1.85%**	**2.55%**	**2.55%**	**2.89%**	**2.89%**	**3.26%**

Figure 2. Improvement to 60 Seconds in Terms of Operational Time and Vehicle Use



These settings make the solver a good fit for the VRP-TW across 2,700 instances. By focusing on total operational time, penalizing extra vehicles, and using a smart search strategy, it finds practical routes that stick to time windows while keeping things efficient.

The solver takes as input the route data (including route ID, stop ID, and stop type), the defined time window instances, and the travel times between every pair of stops within a route. It outputs optimized routing solutions detailing travel time, waiting time, and number of vehicles used. These solutions allow us to assess the impact of varying time window settings on operational efficiency across different geographical contexts

3.6 Analysis of Model Output

In this step, we collect the results of solutions generated by the solver and conduct analyses. Our focus is on the following key performance indicators:

- Travel + waiting time
- Travel time
- Waiting time
- Number of vehicles used

To conduct our analyses, we quantify the impact of time window frequency, duration, and distribution patterns—individually and in combination—on the key performance indicators.

Additionally, we compare the performance of rural routes to urban routes under similar time window settings to identify differences across geographical contexts.

Lastly, we consolidate these findings to provide managerial insights into how logistics service providers can optimize their operations, improve resource allocation, and ultimately deliver higher levels of service to customers across different geographical areas.

4. Results and Discussion

In this section, we present the results generated by the solver and highlight key findings based on those results. We discuss how different time window settings impact operational efficiency. Finally, we offer recommendations for setting proper time windows to balance operational efficiency and customer needs.

4.1 Results of Urban Routes

4.1.1 Results of Urban Routes under Random Distribution

Table 6 shows the results of 450 urban time window instances under Random Distribution, split by window durations and frequencies. For each duration-frequency combination, the result is the average of 50 instances across 10 different routes. A baseline model (with no time windows) is included to compare the results and measure the impact of different time window scenarios.

Table 6. Results for Urban Routes with Random Distribution – 450 Instances

TW Duration - Frequency	Average Time (min) per Stop (Travel + Waiting)	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route
Baseline (no TW)	**1.66**	**1.66**	**0.00**	**1.00**
0%	1.66	1.66	0.00	1.00
1-Hour	**2.84 (70.84%)**	**2.46 (47.81%)**	**0.38**	**1.09 (8.70%)**
5%	2.22 (33.49%)	1.95 (17.20%)	0.27	1.00 (0.00%)
20%	2.97 (78.59%)	2.49 (49.73%)	0.48	1.04 (4.00%)
40%	3.33 (100.48%)	2.93 (76.43%)	0.40	1.22 (22.00%)
2-Hour	**2.23 (33.91%)**	**2.14 (28.74%)**	**0.09**	**1.07 (7.30%)**
5%	1.87 (12.39%)	1.80 (8.36%)	0.07	1.00 (0.00%)
20%	2.27 (36.56%)	2.15 (29.34%)	0.12	1.00 (0.00%)
40%	2.54 (52.86%)	2.47 (48.47%)	0.07	1.22 (22.00%)
4-Hour	**1.76 (5.65%)**	**1.76 (5.65%)**	**0.00**	**1.02 (2.00%)**
5%	1.68 (1.14%)	1.68 (1.14%)	0.00	1.00 (0.00%)
20%	1.76 (5.59%)	1.76 (5.59%)	0.00	1.00 (0.00%)
40%	1.83 (10.22%)	1.83 (10.22%)	0.00	1.06 (6.00%)
Aggregated Average (Baseline Excluded)	**2.28 (36.80%)**	**2.12 (27.40%)**	**0.16**	**1.06 (6.00%)**

On top of the average time and vehicle use, Table 6 also shows the percentage increase from the baseline. For travel times, time windows with a 4-hour duration result in the smallest impact, showing an average increase of just 5.65% across the three frequencies (5%, 20%, and 40%). In contrast, time windows with 1-hour and 2-hour durations lead to a much bigger increase in travel time—averaging 47.81% for 1-hour and 28.74% for 2-hour durations across the frequencies. Even when only 5% of stops require time windows under 1-hour and 2-hour durations, the travel time still jumps significantly, by 17.20% and 8.36%, respectively. For total travel + wait time, the pattern is similar: 4-hour duration has the smallest increase at 5.65%, while 1-hour and 2-hour durations see much larger increases of 70.84% and 33.91%, respectively. This larger increase in total time is because of waiting time, which makes up 13.52% (0.38 out of 2.84) of the total time for 1-hour duration and 3.91% (0.09 out of 2.23) for 2-hour duration, but drops to 0% for 4-hour duration since the longer windows allow the solver to avoid waiting.

In terms of vehicle use, it is no surprise that the 4-hour time window has the lowest impact across the three durations, with an average increase of 2.00%. Meanwhile, 5%- and 20%-time window frequencies have almost no impact on vehicle use across all durations (0.00% in most cases), but the 40% frequency shows a noticeable impact, even under the 4-hour duration, with increases of 6.00% for 4 hours, 22.00% for 2 hours, and 22.00% for 1 hour.

4.1.2 Results of Urban Routes under "All in the Morning" Distribution

Similar to Table 6, Table 7 presents the results of 450 urban time window instances under the All-in-the-Morning Distribution (Section 3.4). It shows that the All-in-the-Morning Distribution leads to a substantial increase in vehicle use compared to the baseline. Even with a 4-hour time window, the number of vehicles required rises by 35.33%, and this increase becomes more pronounced with tighter constraints—particularly at 1-hour and 2-hour durations, and at higher frequencies like 20% and 40%. In contrast, when time windows are applied to just 5% of stops, there is no change in vehicle use (0.00%), indicating that low-frequency constraints can be absorbed without additional fleet deployment.

Table 7. Results for Urban Routes with All-in-the-Morning Distribution – 450 Instances

TW Duration - Frequency	Average Time (min) per Stop (Travel + Waiting)	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route
Baseline (no TW)	**1.66**	**1.66**	**0.00**	**1.00**
0%	1.66	1.66	0.00	1.00
1-Hour	**2.01 (21.06%)**	**2.01 (21.06%)**	**0.00**	**1.53 (52.67%)**
5%	1.73 (3.81%)	1.73 (3.81%)	0.00	1.00 (0.00%)
20%	2.03 (22.06%)	2.03 (22.06%)	0.00	1.58 (58.00%)
40%	2.28 (37.31%)	2.28 (37.30%)	0.00	2.00 (100.00%)
2-Hour	**1.88 (12.92%)**	**1.88 (12.92%)**	**0.00**	**1.39 (39.33%)**
5%	1.70 (2.03%)	1.70 (2.03%)	0.00	1.00 (0.00%)
20%	1.86 (12.08%)	1.86 (12.08%)	0.00	1.34 (34.00%)
40%	2.07 (24.63%)	2.07 (24.63%)	0.00	1.84 (84.00%)
4-Hour	**1.79 (7.57%)**	**1.79 (7.57%)**	**0.00**	**1.35 (35.33%)**
5%	1.67 (0.38%)	1.67 (0.38%)	0.00	1.00 (0.00%)
20%	1.82 (9.31%)	1.82 (9.31%)	0.00	1.44 (44.00%)
40%	1.88 (13.03%)	1.88 (13.03%)	0.00	1.62 (62.00%)
Aggregated Average (Baseline Excluded)	**1.89 (13.85%)**	**1.89 (13.85%)**	**0.00**	**1.42 (42.44%)**

With regard to time-related impacts, the increase in travel time is more modest but still significant. The 4-hour window results in a 7.57% rise, while the 2-hour and 1-hour durations lead to increases of 12.92% and 21.06%, respectively. The total time per stop (travel plus waiting) mirrors these increases. However, due to the clustering of all time windows within the same morning period (8 a.m. to 12 p.m.), waiting time remains at 0.00% across all scenarios, suggesting that tight scheduling within a common window eliminates idle periods.

4.1.3 Results of Urban Routes under "40-20-40" Distribution

Table 8 summarizes the results as well as the percentage rise from the baseline for 450 urban instances under the 40-20-40 Distribution. As with other distribution patterns, shorter time windows and higher frequencies increase both travel and waiting time. Among the three durations, 1-hour windows cause the sharpest increases — total time per stop jumps by 75.14%, and travel time alone increases by 50.10%. Shorter windows not only extend routes but also lead to more frequent vehicle deployment, with a 40% frequency pushing vehicle use up by 24%. In contrast, 4-hour windows show minimal impact across all metrics, with average increases of just 7.60% in time and 0.00% in vehicles used.

Table 8. Results for Urban Routes with 40-20-40 Distribution – 450 Instances

TW Duration - Frequency	Average Time (min) per Stop (Travel + Waiting)	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route
Baseline (no TW)	**1.66**	**1.66**	**0.00**	**1.00**
0%	1.66	1.66	0.00	1.00
1-Hour	**2.91 (75.14%)**	**2.50 (50.10%)**	**0.42**	**1.09 (8.67%)**
5%	2.43 (46.13%)	2.06 (23.59%)	0.37	1.00 (0.00%)
20%	2.94 (76.48%)	2.49 (49.71%)	0.45	1.02 (2.00%)
40%	3.37 (102.82%)	2.94 (77.01%)	0.43	1.24 (24.00%)
2-Hour	**2.29 (37.45%)**	**2.18 (31.08%)**	**0.11**	**1.04 (4.00%)**
5%	1.97 (18.70%)	1.87 (12.18%)	0.11	1.00 (0.00%)
20%	2.32 (39.58%)	2.21 (32.65%)	0.12	1.00 (0.00%)
40%	2.56 (54.06%)	2.47 (48.42%)	0.09	1.12 (12.00%)
4-Hour	**1.79 (7.60%)**	**1.79 (7.60%)**	**0.00**	**1.00 (0.00%)**
5%	1.70 (2.07%)	1.70 (2.07%)	0.00	1.00 (0.00%)
20%	1.78 (7.23%)	1.78 (7.23%)	0.00	1.00 (0.00%)
40%	1.89 (13.52%)	1.89 (13.52%)	0.00	1.00 (0.00%)
Aggregated Average (Baseline Excluded)	**2.33 (40.06%)**	**2.16 (29.60%)**	**0.17**	**1.04 (4.22%)**

4.1.4 Consolidated Urban Routes Insights

In this section, we consolidate the results from three distinct time window distribution patterns. We propose a scoring method to evaluate the route efficiency of each distribution, duration, and frequency by summing the weighted scores of average operational time (travel + waiting + service) per stop and average vehicles used per route. The weights of these two components are determined based on their relative costs in real-world logistics. In this analysis, we assign a 30% weight to operational time and a 70% weight to vehicle use. These weights are derived from general experience rather than from comprehensive data, as collecting sufficient data to calculate precise weights is challenging and beyond the scope of this project. We include service time in the operational time component because it contributes to labor costs, which are typically measured by the hours a driver spends completing a route and are a critical factor in determining the weights of the two components.

In this scoring framework, we set the baseline (no time windows) score to 100, and the score for each component under other scenarios (e.g., Random Distribution, 5% frequency) is calculated using Equation (1):

$$Score\_c = 100 \times (Baseline\ Metric\_c \div Scenario\ Metric\_c) \times Weight\_c \quad (1)$$

where *Score_c* is the score for component *c* (e.g., time or vehicles), *Baseline Metric_c* and *Scenario Metric_c* are the baseline and scenario values for component *c*, and *Weight_c* is the weight (0.3 for time, 0.7 for vehicles). The total score is the sum of component scores. For example, in urban routes under Random Distribution (Table 9), the baseline operational time is 1.66 minutes per stop and vehicle use is 1.00 per route, while Random Distribution yields 2.28 minutes per stop and 1.06 vehicles per route. Using Equation (1), the time score is 24.39, the vehicle score is 66.70, and the total score is 91.09.

Table 9. Results Comparison for Different TW Distributions – 450 Instances

TW Distribution	Average Travel + Waiting Time (min) per Stop	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route	Average Travel+Waiting+Service Time (min) per Stop	Avarage Vehicles per Stop	Weighted Score
Baseline - No TW	1.66	1.66	0.00	1.00	2.66	0.0070	100.00
Random Distributior	2.28 (37.35%)	2.12 (27.71%)	0.16	1.06 (6.00%)	3.28	0.0074	91.09
40-20-40	2.33 (40.36%)	2.16 (30.12%)	0.17	1.04 (4.00%)	3.33	0.0073	91.60
All in The Morning	1.89 (13.86%)	1.89 (13.86%)	0.00	1.42 (42.00%)	2.89	0.0097	78.38

Table 9 shows the results of the baseline and the aggregated results (consolidating all sub-tier instances) of the three different time window distribution patterns, with each distribution pattern reflecting the average of 150 instances across different time window durations and frequencies. All-in-the-Morning generates the lowest travel time per stop and wait time among the three distribution patterns; however, it yields the highest vehicle use—a 42% increase compared with the baseline, with 40-20-40 at 4% and Random Distribution at 6%. These results indicate that a high density of time windows leads to lower operational time in terms of both travel and waiting. This reduction, however, partly stems from the extra flexibility provided by a second or third vehicle. Additionally, the lower travel time arises because a high density of time windows offers the model more stops to select as the next destination, making the optimal or close-to-optimal selection (in regard to reducing travel time) more attainable. Table 9 shows that All-in-the-Morning achieves an average travel time per stop lower than Random Distribution and 40-20-40, due to the abundance of stops available for selection in a dense morning schedule.

Beyond the flexibility from additional vehicles, the 0-minute wait time under All-in-the-Morning also results from the truck not needing to arrive early and wait at certain stops until the time windows open, since there is an abundance of other stops to choose as the next destination before the time windows open. In contrast, in Random Distribution and 40-20-40 scenarios, the relatively sparse distribution of time windows often leaves the truck with no other appropriate stops to serve, forcing it to arrive early at certain stops and wait there until the time windows open. However, while All-in-the-Morning shows low per-stop travel and waiting time, this perceived efficiency is not rooted in smarter routing, but rather in increased truck deployment to complete deliveries within a compressed time frame—effectively doubling the fleet and driving up both costs and planning complexity.

As previously noted, All-in-the-Morning results in a substantially higher increase in vehicle use compared with the baseline. This is due to the fact that a high density of time windows is much more likely to make serving all stops with only one vehicle infeasible, especially in those instances with high time window frequency. Even in urban contexts, where routes benefit from higher stop density and shorter travel times, tight windows combined with clustering can overwhelm available vehicle capacity, especially during peak delivery hours. As a result, urban logistics planners must weigh the benefits of faster routes against the significant increase in truck deployment. The number of trucks acts as the most consequential lever in managing delivery costs and operational efficiency, and must be carefully monitored under varying constraint scenarios.

Regarding the scores for different time window distributions, the All-in-the-Morning pattern receives the lowest score among the three, which aligns exactly with our expectation. However, it is noteworthy that the 40-20-40 Distribution scores only slightly higher than the Random Distribution. Indeed, both patterns present their own advantages and disadvantages: the 40-20-40 Distribution scores lower with respect to operational time but higher in terms of vehicle use. Consequently, if we assign different weights to these two components, the resulting overall score would be different, highlighting a trade-off between delivery time and truck count that delivery operators must navigate.

Regarding the performance difference between different time window durations, Table 10 indicates that the shorter the duration is, the worse the operational efficiency it generates. As expected, the 1-hour duration produces the worst results in both time consumption and vehicle use, with an average total time per stop of 2.59 minutes and vehicle use of 1.23, compared to 1.78 minutes and 1.12 vehicles for the 4-hour duration. Notably, the average wait time increases significantly, from 0.06 minutes for 2-hour time window durations to 0.27 minutes for 1-hour durations, representing a 350% relative increase. These results demonstrate the critical role of flexibility in route optimization—where narrow windows impose strict timing constraints that increase the likelihood of idle time and force multi-vehicle routing. From a fleet management perspective, 1-hour windows should be limited to premium service cases, or capped to a small subset of high-priority stops to preserve operational sustainability.

Table 10. Results Comparison for Different TW Durations – 450 Instances

TW Duration	Average Travel + Waiting Time (min) per Stop	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route	Average Travel+Waiting+Service Time (min) per Stop	Avarage Vehicles per Stop	Weighted Score
Baseline (No TW)	1.66	1.66	0.00	1.00	2.66	0.0070	100.00
1-Hour	2.59 (56.02%)	2.32 (39.76%)	0.27	1.23 (23.00%)	3.59	0.0085	80.21
2-Hour	2.13 (28.31%)	2.07 (24.70%)	0.06	1.17 (17.00%)	3.13	0.0081	86.42
4-Hour	1.78 (7.23%)	1.78 (7.23%)	0.00	1.12 (12.00%)	2.78	0.0078	92.03

As for the impact of another key parameter, the time window frequency, Table 11 displays that higher time window frequencies increase both operational time and vehicle use, which aligns with expectations. However, an interesting observation is that the average wait time at 20% frequency (0.13 minutes) is slightly higher than at 40% frequency (0.11 minutes). This suggests that while higher frequencies generally increase operational demands, the higher density of time windows and increased vehicle use provide greater flexibility, allowing the model to reduce wait time. Still, these gains come at the cost of scaling truck deployment to absorb the complexity. Notably, 4-hour time windows stand out as the most optimal option in term of route efficiency—they offer enough flexibility for route planning to remain efficient, while keeping vehicle use near the baseline and preserving cost control.

Table 11. Results Comparison for Different TW Frequencies – 450 Instances

TW Frequency	Average Travel + Waiting Time (min) per Stop	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route	Average Travel+Waiting+Service Time (min) per Stop	Avarage Vehicles per Stop	Weighted Score
Baseline (No TW)	1.66	1.66	0.00	1.00	2.66	0.0070	100.00
5%	1.89 (13.86%)	1.79 (7.83%)	0.09	1.00 (0.00%)	2.89	0.0070	97.69
20%	2.19 (31.93%)	2.07 (24.70%)	0.13	1.16 (16.00%)	3.19	0.0080	86.76
40%	2.42 (45.78%)	2.31 (39.16%)	0.11	1.37 (37.00%)	3.42	0.0094	75.99

In addition to the tables, we present the results of urban routes using heatmaps in Figure 3 and Figure 4. Figure 3 and Figure 4 visualize how different time window strategies impact urban delivery performance. Figure 3 shows that All-in-the-Morning Distributions, especially with tighter windows and higher frequencies, consistently demand more vehicles—visually confirming the sharp fleet increases seen in Table 9. Figure 4 illustrates the benefit of this approach in reducing travel time, though these gains come at the cost of more trucks and compressed scheduling. These plots highlight the core trade-off: tighter constraints improve customer precision but increase vehicle and time costs. By integrating these visuals, we present a more intuitive understanding of how distribution, duration, and frequency interact—helping logistics planners weigh speed, cost, and fleet deployment more effectively.

Figure 3. Urban Routes: Vehicle Use per Route by TW Distribution, Duration, and Frequency

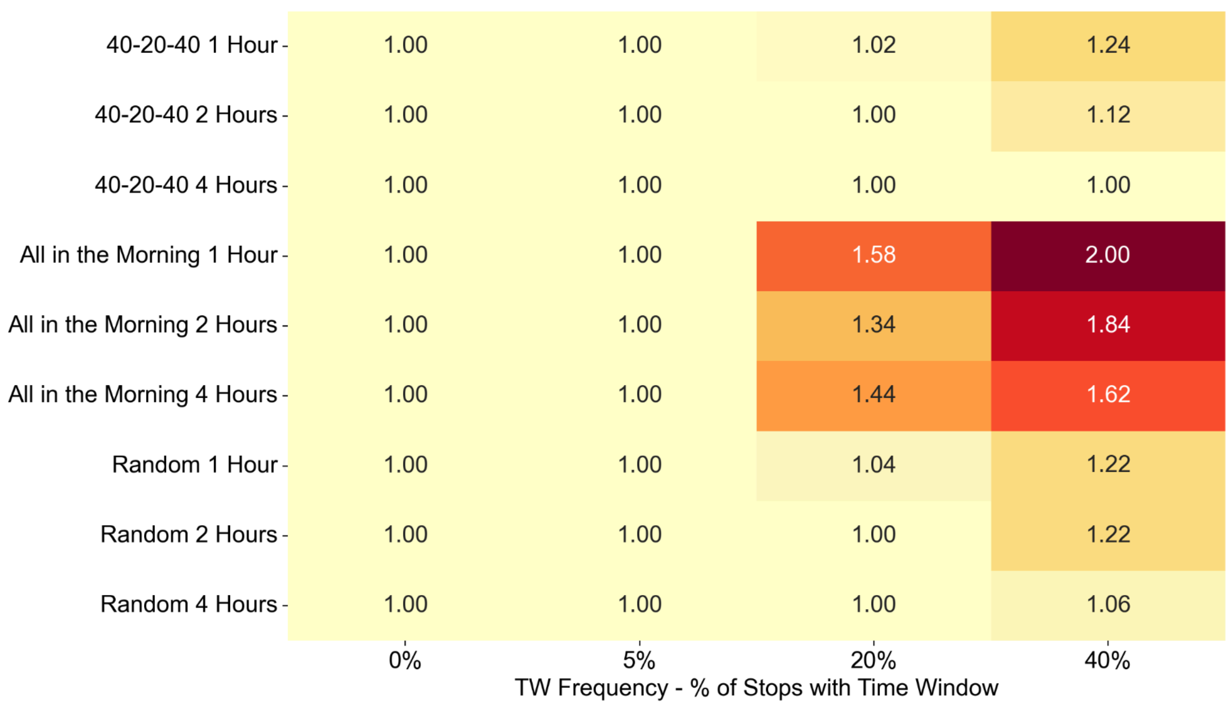
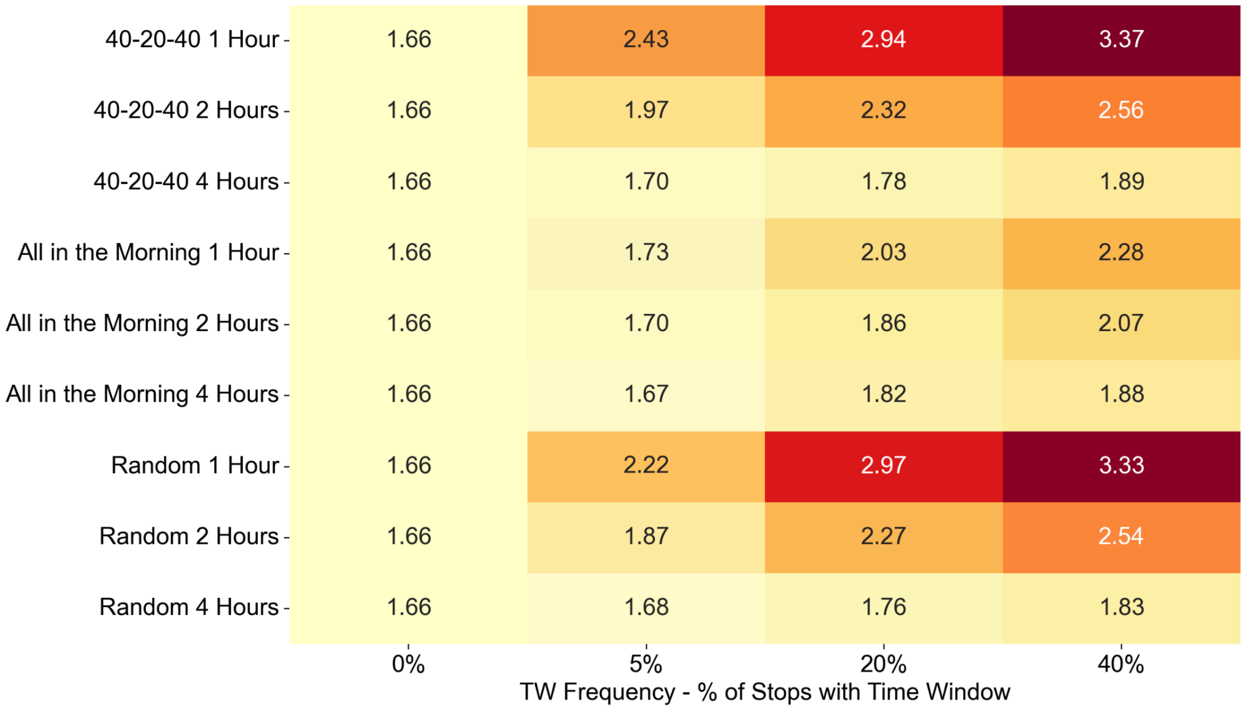


Figure 4. Urban Routes: Travel + Waiting Time per Stop by TW Distribution, Duration, and Frequency



4.2 Results of Rural Routes

4.2.1 Results of Rural Routes under Random Distribution

Table 12 presents the results of 450 rural time window instances under Random Distribution. Each duration-frequency combination reflects an average of 50 instances across 10 routes, with a baseline (no time windows) for reference.

Table 12. Results for Rural Routes with Random Distribution – 450 Instances

TW Duration - Frequency	Average Time (min) per Stop (Travel + Waiting)	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route
Baseline (no TW)	**3.81**	**3.81**	**0.00**	**1.10**
0%	3.81	3.81	0.00	1.10
1-Hour	**5.60 (46.98%)**	**5.55 (45.87%)**	**0.04**	**1.33 (20.61%)**
5%	4.27 (12.09%)	4.25 (11.67%)	0.02	1.10 (0.00%)
20% [1]	5.56 (45.97%)	5.52 (45.08%)	0.03	1.28 (16.36%)
40% [1]	6.96 (82.86%)	6.89 (80.88%)	0.08	1.60 (45.45%)
2-Hour	**4.65 (22.24%)**	**4.65 (22.22%)**	**0.00**	**1.15 (4.85%)**
5%	4.00 (5.07%)	4.00 (5.06%)	0.00	1.10 (0.00%)
20%	4.68 (22.81%)	4.67 (22.80%)	0.00	1.10 (0.00%)
40% [2]	5.29 (38.84%)	5.28 (38.82%)	0.00	1.26 (14.55%)
4-Hour	**4.11 (8.04%)**	**4.11 (8.04%)**	**0.00**	**1.10 (0.00%)**
5%	3.85 (1.13%)	3.85 (1.13%)	0.00	1.10 (0.00%)
20%	4.11 (7.95%)	4.11 (7.95%)	0.00	1.10 (0.00%)
40%	4.38 (15.04%)	4.38 (15.03%)	0.00	1.10 (0.00%)
Aggregated Average (Baseline Excluded)	**4.79 (25.75%)**	**4.77 (25.38%)**	**0.01**	**1.19 (8.48%)**

Note: The superscripts indicate that the model is unable to find a solution for 4 instances (representing less than 1% of the total 450), which are imputed with the average value of other instances with the same route-duration-frequency setting.

Table 12 also details the percentage rise from the baseline for time and vehicle use in rural routes under Random Distribution. Travel times see the least change with 4-hour durations, up by 8.04% on average, while 1-hour and 2-hour durations climb higher, at 45.87% and 22.22%, respectively. Total travel plus wait time mirrors this trend: 4-hour at 8.04%, 1-hour at 46.98%, and 2-hour at 22.24%, with waiting time adding 0.04 (min) for 1-hour and almost none for 2-hour and 4-hour. Vehicle use experiences zero shift with 4-hour durations, whereas 1-hour and 2-hour durations grow by 20.61% and 4.85%, respectively. At 5% frequency, vehicle use barely changes (0%) across all durations, but at 40% frequency, it jumps, reaching 14.55% for 2 hours and 45.45% for 1 hour.

4.2.2 Results of Rural Routes under All-in-the-Morning Distribution

Table 13 presents the results of 450 rural time window instances under the All-in-the-Morning Distribution as well as the percentage rise from the baseline; similar to Table 12, each duration-frequency combination reflects an average of 50 instances across 10 routes, with a baseline (no time windows) for comparison.

Travel times show the smallest increase with 4-hour durations, up by 4.06% on average, while 1-hour and 2-hour durations rise more significantly at 19.94% and 10.36%, respectively. Total travel plus wait time follows a similar pattern: 4-hour at 4.06%, 1-hour at 19.94%, and 2-hour at 10.36%, with no waiting time across all durations due to the morning-concentrated time windows. Vehicle use shows no change at 5%, but increases notably at higher frequencies, reaching 69.09% for 4-hour at 40% frequency, 96.82% for 1-hour at 40% frequency, and 90.91% for 2-hour at 40% frequency.

Table 13. Results for Rural Routes with All-in-the-Morning Distribution – 450 Instances

TW Duration - Frequency	Average Time (min) per Stop (Travel + Waiting)	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route
Baseline (no TW)	**3.81**	**3.81**	**0.00**	**1.10**
0%	3.81	3.81	0.00	1.10
1-Hour	**4.57 (19.94%)**	**4.57 (19.94%)**	**0.00**	**1.72 (56.52%)**
5%	3.97 (4.22%)	3.97 (4.22%)	0.00	1.10 (0.00%)
20% [2]	4.63 (21.58%)	4.63 (21.58%)	0.00	1.90 (72.73%)
40% [1]	5.10 (34.02%)	5.10 (34.01%)	0.00	2.17 (96.82%)
2-Hour	**4.20 (10.36%)**	**4.20 (10.36%)**	**0.00**	**1.59 (44.24%)**
5%	3.87 (1.57%)	3.87 (1.57%)	0.00	1.10 (0.00%)
20% [1]	4.26 (11.78%)	4.26 (11.78%)	0.00	1.56 (41.82%)
40% [2]	4.48 (17.75%)	4.48 (17.75%)	0.00	2.10 (90.91%)
4-Hour	**3.96 (4.06%)**	**3.96 (4.06%)**	**0.00**	**1.43 (30.30%)**
5%	3.82 (0.33%)	3.82 (0.33%)	0.00	1.10 (0.00%)
20%	3.95 (3.79%)	3.95 (3.79%)	0.00	1.34 (21.82%)
40%	4.11 (8.06%)	4.11 (8.06%)	0.00	1.86 (69.09%)
Aggregated Average (Baseline Excluded)	**4.24 (11.45%)**	**4.24 (11.45%)**	**0.00**	**1.58 (43.69%)**

Note: The superscripts indicate that the model is unable to find a solution for 6 instances (representing 1.33% of the total 450), which are imputed with the average value of other instances with the same route-duration-frequency setting.

4.2.3 Results of Rural Routes under 40-20-40 Distribution – 450 Instances

Table 14 details the performance of 450 rural time window instances under 40-20-40 Distribution as well as the percentage rise from the baseline, grouped by duration and frequency. Travel times grow least with 4-hour durations, while 1-hour and 2-hour durations increase more significantly. Total travel plus wait time follows a similar trend, with waiting time contributing slightly for 1-hour and 2-hour durations but almost none for 4-hour. Vehicle use remains unchanged for 4-hour durations but rises for 1-hour and 2-hour durations, especially at higher frequencies.

Table 14. Results for Rural Routes with 40-20-40 Distribution

TW Duration - Frequency	Average Time (min) per Stop (Travel + Waiting)	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route
Baseline (no TW)	**3.81**	**3.81**	**0.00**	**1.10**
0%	3.81	3.81	0.00	1.10
1-Hour	**5.57 (46.44%)**	**5.49 (44.33%)**	**0.08**	**1.27 (15.76%)**
5%	4.34 (14.03%)	4.28 (12.31%)	0.07	1.10 (0.00%)
20% [1]	5.53 (45.14%)	5.48 (44.07%)	0.04	1.14 (3.64%)
40%	6.86 (80.14%)	6.72 (76.59%)	0.14	1.58 (43.64%)
2-Hour	**4.68 (23.03%)**	**4.68 (23.01%)**	**0.00**	**1.13 (3.03%)**
5%	4.04 (6.24%)	4.04 (6.23%)	0.00	1.10 (0.00%)
20% [1]	4.67 (22.79%)	4.67 (22.76%)	0.00	1.10 (0.00%)
40% [1]	5.33 (40.04%)	5.33 (40.04%)	0.00	1.20 (9.09%)
4-Hour	**4.19 (10.15%)**	**4.19 (10.14%)**	**0.00**	**1.10 (0.00%)**
5%	3.88 (2.04%)	3.88 (2.04%)	0.00	1.10 (0.00%)
20%	4.20 (10.31%)	4.20 (10.31%)	0.00	1.10 (0.00%)
40%	4.50 (18.09%)	4.50 (18.09%)	0.00	1.10 (0.00%)
Aggregated Average (Baseline Excluded)	**4.82 (26.54%)**	**4.79 (25.83%)**	**0.03**	**1.17 (6.26%)**

Note: The superscripts indicate that the model is unable to find a solution for 3 instances (representing 0.67% of the total 450), which are imputed with the average value of other instances with the same route-duration-frequency setting.

4.2.4 Consolidated Rural Routes Results Insights

In this section, we consolidate the results from three distinct time window distribution patterns in rural context to identify the high-level trends. Consistent with Section 4.1.4, we also incorporate the route efficiency score using the same calculation method.

Table 15 presents the aggregated results (consolidating all sub-tier instances) of the baseline and the three different time window distribution patterns, with each distribution pattern reflecting the average of 150 instances across different time window durations and frequencies. All-in-the-Morning generates the lowest travel time per stop (4.24 minutes) and wait time (0 minutes) among the three distributions; however, it results in the highest vehicle use—a 43.69% increase compared with the baseline, whereas 40-20-40 and Random Distribution show much smaller increases (6.26% and 8.48%, respectively). While this may suggest improved timing efficiency under All-in-the-Morning, it's important to recognize that the gains are achieved not through smarter routing, but by significantly increasing fleet deployment to meet the compressed delivery schedule. As highlighted in the broader rural analysis, the time efficiency is genuine—but it comes at a high operational cost. This underscores the need to look beyond vehicle travel time alone and consider the broader implications for resource allocation, labor, and sustainability.

Table 15. Results Comparison for Different TW Distributions – 450 Instances

TW Frequency	Average Travel + Waiting Time (min) per Stop	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route	Average Travel+Waiting+Service Time (min) per Stop	Avarage Vehicles per Stop	Weighted Score
Baseline - No TW	3.81	3.81	0.00	1.10	4.81	0.0111	100.00
Random Distribution	4.79 (25.75%)	4.77 (25.38%)	0.01	1.19 (8.48%)	5.79	0.0121	89.22
40-20-40	4.82 (26.54%)	4.79 (25.83%)	0.03	1.17 (6.26%)	5.82	0.0118	90.51
All in The Morning	4.24 (11.45%)	4.24 (11.45%)	0.00	1.58 (43.69%)	5.24	0.0160	76.06

In terms of vehicle use, the All-in-the-Morning Distribution shows a much higher increase compared with the baseline than the other two distribution patterns, as previously observed. As with other rural scenarios, the higher density of time windows is much more likely to make it infeasible to serve all stops with only one vehicle—particularly in instances with narrow windows or high constraint frequency. This strain is further amplified by the lower stop density and longer distances typical of rural routes, which reduces the system's ability to absorb timing constraints through route optimization alone.

Regarding the scores for different time window distributions, the All-in-the-Morning pattern receives the lowest score among the three, which is expected. Meanwhile, the 40-20-40 Distribution (at 90.51) scores slightly higher than the Random Distribution (89.22). This reflects the additional cost of unpredictability in Random assignments, which introduces inefficiencies in scheduling and sequencing compared to the more structured spread of 40-20-40.

As for the performance difference between different time window durations, Table 16 indicates that the shorter the duration, the worse the operational efficiency, which is consistent with intuition. This is particularly relevant in rural contexts, where tighter delivery windows quickly escalate vehicle needs. For example, 1-hour windows with 40% stop constraints can push vehicle requirements to more than double the baseline, especially in clustered scenarios like All-in-the-Morning. These scenarios compress deliveries into narrow time bands that simply exceed what a single vehicle can realistically achieve.

Table 16. Results Comparison for Different TW Durations – 450 Instances

TW Frequency	Average Travel + Waiting Time (min) per Stop	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route	Average Travel+Waiting+Service Time (min) per Stop	Avarage Vehicles per Stop	Weighted Score
Baseline (No TW)	3.81	3.81	0.00	1.10	4.81	0.0111	100.00
1-Hour	5.25 (37.79%)	5.20 (36.71%)	0.04	1.44 (30.96%)	6.25	0.0146	76.30
2-Hour	4.51 (18.54%)	4.51 (18.53%)	0.00	1.29 (17.37%)	5.51	0.0131	85.59
4-Hour	4.09 (7.41%)	4.09 (7.41%)	0.00	1.21 (10.10%)	5.09	0.0122	91.86

Another key parameter—the time window frequency—influences route efficiency to a large extent. Table 17 shows that higher time window frequencies increase both operational time and vehicle use. This finding reaffirms the sensitivity of rural operations to rigid constraints: even moderate increases in constrained stop percentage (e.g., from 20% to 40%) can cause disproportionate rises in vehicle deployment. However, more flexible delivery windows, such as 2-hour and 4-hour slots, significantly mitigate this effect, allowing most routes to stay within a 1.1 to 1.3 truck range even with higher coverage under Random and 40-20-40 distributions. This reinforces a central takeaway: Delivery flexibility is directly proportional to routing efficiency, especially in rural networks where travel time consumes a greater portion of the route schedule.

Table 17. Results Comparison for Different TW Frequencies – 450 Instances

TW Frequency	Average Travel + Waiting Time (min) per Stop	Average Travel Time (min) per Stop	Average Waiting Time (min) per Stop	Average Vehicles per Route	Average Travel+Waiting+Service Time (min) per Stop	Avarage Vehicles per Stop	Weighted Score
Baseline (No TW)	3.81	3.81	0.00	1.10	4.81	0.0111	100.00
5%	4.00 (5.19%)	4.00 (4.95%)	0.01	1.10 (0.00%)	5.00	0.0111	98.82
20%	4.62 (21.35%)	4.61 (21.12%)	0.01	1.29 (17.37%)	5.62	0.0130	85.22
40%	5.22 (37.21%)	5.20 (36.58%)	0.02	1.55 (41.06%)	6.22	0.0157	72.46

In addition to the tables, we present the results of rural routes using heatmaps in Figures 5 and 6, offering a clearer, more intuitive view of how rural delivery performance is impacted by different time window configurations. Figure 5 highlights the sharp increase in vehicle use, especially under All-in-the-Morning and 1-hour windows at high frequencies—visually reinforcing the 44% vehicle increase reported in Table 15. In low-density rural areas, even modest constraint increases can force multi-vehicle routing. Figure 6 complements this by showing how travel plus waiting time per stop escalates with tighter and more frequent time windows. The clustered heatmap gradients show how rural routes quickly become inefficient under rigid constraints, due to longer travel distances and fewer stop alternatives.

Figure 5. Rural Routes: Vehicle Use per Route by TW Distribution, Duration, and Frequency

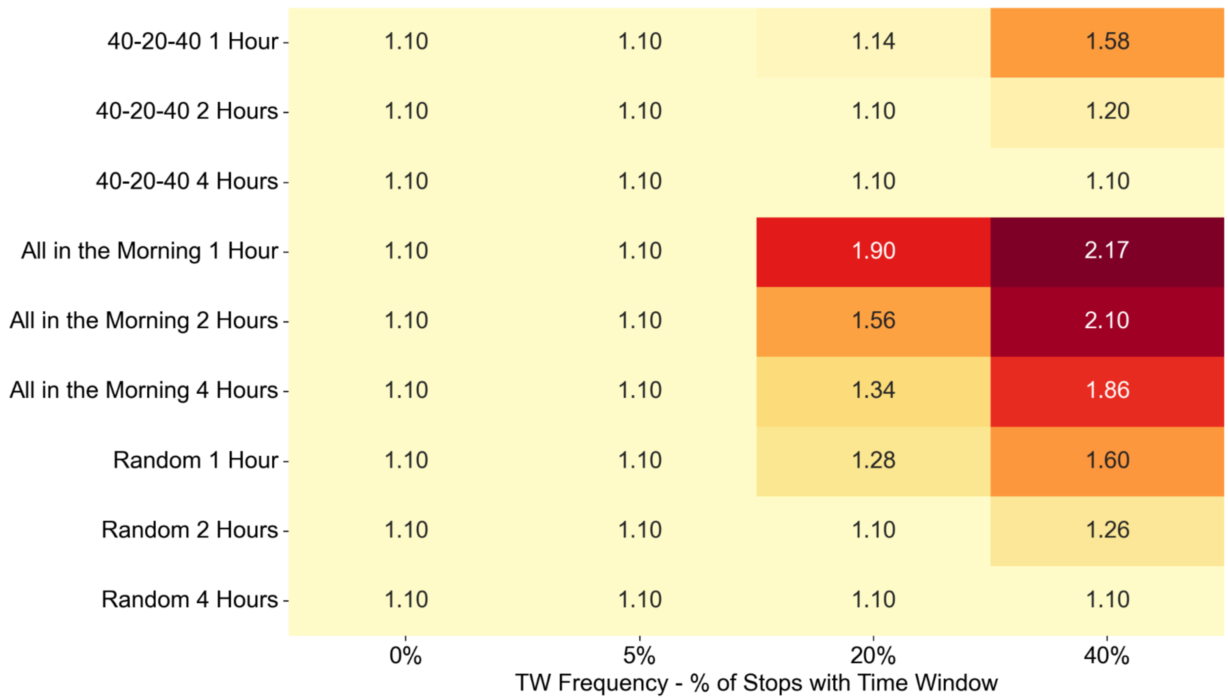
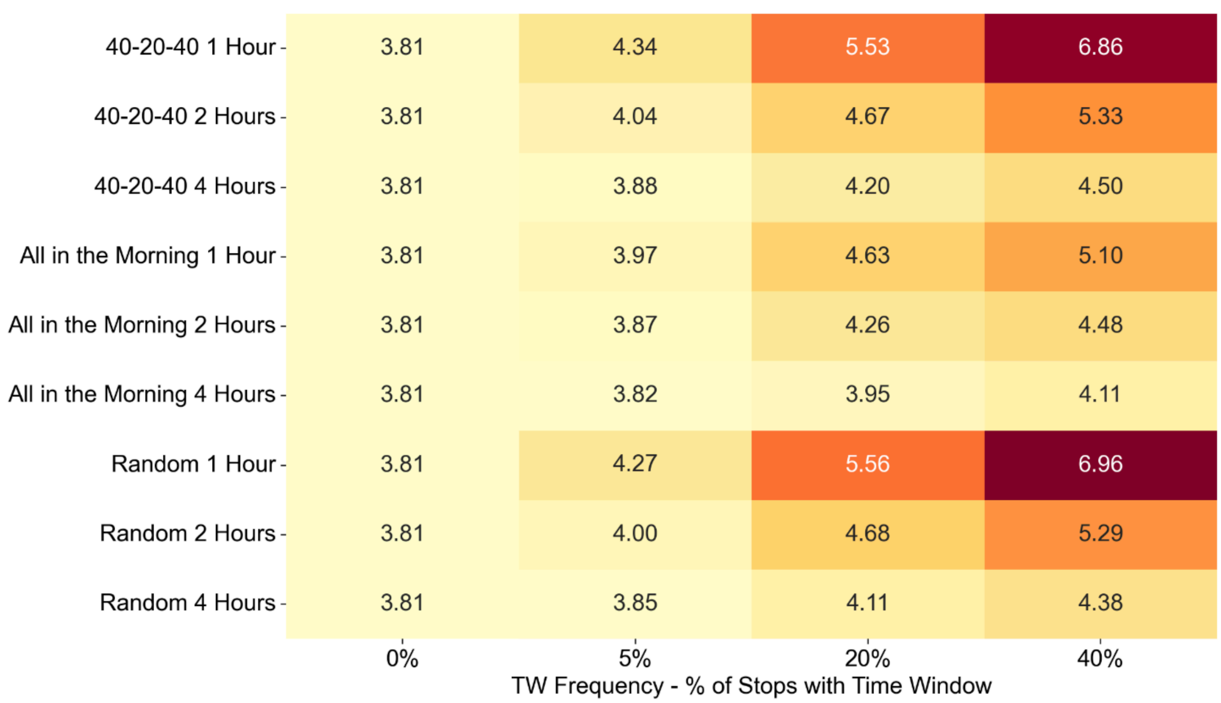


Figure 6. Rural Routes: Travel + Waiting Time per Stop by TW Distribution, Duration, and Frequency



Together, these visuals make the core takeaway clear: flexibility is essential for sustainable rural routing, and tight or dense time windows, while offering better service precision, lead to higher costs and more fragmented routes.

4.3 Urban vs. Rural Routes Analysis

This section compares the results between urban and rural routes. Tables 18, 19, and 20 display the comparisons under different time window distributions, durations, and frequencies, respectively.

In both urban and rural routes, a common trend emerges: the All-in-the-Morning Distribution requires the highest number of vehicles—1.42 for urban and 1.58 for rural routes (Table 18). However, this distribution also achieves the lowest travel and waiting times, with 1.89 minutes per stop and no waiting in urban areas, and 4.24 minutes per stop with no waiting in rural areas.

Similarly, a 1-hour duration yields the lowest route efficiency among the three time window durations, with the highest operational times and vehicle use, such as 2.59 minutes per stop and 1.23 vehicles in urban routes, and 5.25 minutes per stop and 1.44 vehicles in rural routes (Table 19). Likewise, a 40% frequency leads to the highest average travel time and vehicle use among the three time window frequencies—for example, 2.42 minutes per stop and 1.37 vehicles in urban routes, and 5.22 minutes per stop and 1.55 vehicles in rural routes (Table 20). These results affirm that higher densities of time windows, shorter durations, and tighter delivery commitments place similar strains on both rural and urban systems, driving up vehicle use and elongating delivery times. Despite these commonalities, there are meaningful structural and performance differences between urban and rural routes, which are elaborated in the following paragraphs.

As shown in Table 18, the baseline average travel time per stop is significantly higher in rural routes (3.81 minutes) compared to urban routes (1.66 minutes), which aligns with expectations due to the longer distances and lower stop density in rural areas. Interestingly, under both the Random and 40-20-40 Distributions, the percentage increase in travel plus waiting time is greater in urban routes—e.g., 37.35% for urban versus 25.75% for rural under Random Distribution, and 40.36% for urban versus 26.54% for rural under the 40-20-40 Distribution. This disparity stems primarily from higher increases in waiting time in urban settings (e.g., 0.16 minutes for urban versus 0.01 minutes for rural under Random Distribution), which highlights an important consideration: in denser environments where travel time is inherently lower, time window requirements lead to higher idle time. This insight is further supported when examining total travel

+ service time in baseline scenarios. In urban routes, the average total travel + service time in the baseline is approximately 391.29 minutes (6.52 hours) across 147 stops, while in rural routes, it is 479.37 minutes (8.00 hours) across 100 stops. Under Random Distribution, the average total travel + service time for urban routes increases to 458.77 minutes (7.65 hours), while rural routes increase to 576.74 minutes (9.61 hours). Given the operational start time of 6 AM and the latest time windows opening at 5 PM (11 hours after the start time) for 1-hour durations, 4 PM (10 hours after the start time) for 2-hour durations, and 2 PM (8 hours after the start time) for 4-hour durations, vehicles in urban settings—where less driving is required— are more likely to arrive at stops before the designated delivery windows open, leading to longer waiting time. This underlines the importance of aligning time window settings with vehicle speed and stop density (factors that determine travel and service time)—issues that affect urban delivery far more prominently than rural.

When it comes to vehicle use, while both environments start from the same baseline assumption—completing routes with as few vehicles as possible in the absence of time windows—the rate at which each system becomes strained differs significantly. Rural settings experience a more immediate and pronounced impact from delivery constraints. For instance, with 1-hour time windows implemented to 40% of stops, rural routes under the All-in-the-Morning Distribution require 2.17 vehicles on average, compared to 2.00 in urban routes. The rationale behind this difference is straightforward: both the average travel time per stop and the total travel time per route are substantially higher in rural contexts than in urban contexts. Consequently, it is more challenging for the model (or delivery operators in practice) to meet time window requirements with a single vehicle under any distribution pattern in rural settings, as the time capacity of the first vehicle is more heavily utilized in rural routes compared to urban routes; this often necessitates additional vehicles to complete deliveries within the required windows. In real-world practices, the implications of additional vehicles are more severe for rural deliveries, where the addition of a second vehicle comes with more mileage, longer return times, and fewer opportunities for resource consolidation. Rural systems are thus more prone to constraint-induced fragmentation, with tight time windows effectively splintering efficient routing sequences and forcing delivery operators to break deliveries into separate routes to comply

Urban routes, by contrast, exhibit a more buffered and gradual response. Higher stop density and shorter distances allow urban vehicles to remain more adaptable under pressure. Even under 40% constraint with 2-hour or 4-hour windows, urban scenarios often manage with 1.00–1.22 vehicles, except for All-in-the-Morning scenarios, where peaks of 1.62 or 1.84 emerge. This suggests that urban density enables more efficient batching and sequencing, particularly when broader delivery windows are in place. However, All-in-the-Morning clustering remains problematic across both settings, as it forces deliveries into an early congestion window, front-loading resource demand and straining vehicle availability at the start of the day.

Table 18. Urban vs. Rural under Different TW Distributions

TW Distribution	Average Travel + Waiting Time (min) per Stop		Average Travel Time (min) per Stop		Average Waiting Time (min) per Stop		Average Vehicles per Route	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Baseline - No TW	1.66	3.81	1.66	3.81	0.00	0.00	1.00	1.10
Random Distribution	2.28 (37.35%)	4.79 (25.75%)	2.12 (27.71%)	4.77 (25.38%)	0.16	0.01	1.06 (6.00%)	1.19 (8.48%)
40-20-40	2.33 (40.36%)	4.82 (26.54%)	2.16 (30.12%)	4.79 (25.83%)	0.17	0.03	1.04 (4.00%)	1.17 (6.26%)
All in The Morning	1.89 (13.86%)	4.24 (11.45%)	1.89 (13.86%)	4.24 (11.45%)	0.00	0.00	1.42 (42.00%)	1.58 (43.69%)

In terms of relative strain acceleration, rural routes also show a sharper curve. As constraints increase from 20% to 40%, rural routes exhibit a steeper rise in vehicle requirements than urban ones. This reflects a core operational vulnerability: rural networks lack the routing flexibility to absorb added constraints without rapidly escalating fleet demands. In practical terms, this makes rural delivery operations less tolerant of delivery rigidity, and more sensitive to seemingly moderate increases in constraint levels.

Table 19 reinforces this point through vehicle use under different time window durations. While urban and rural routes show similar relative increases under 2-hour and 4-hour durations (e.g., 17.00% for urban and 17.37% for rural under 2-hour duration), the 1-hour window produces a far larger gap: 30.96% for rural compared to 23.00% for urban. This highlights the compounding cost effect of narrow windows in rural areas, where each additional vehicle has a higher marginal cost and more logistical limitations. Delivery operators should be particularly cautious when introducing tight time windows in rural settings, unless the service level requirement or customer value clearly justifies the trade-off.

Table 19. Urban vs. Rural under Different TW Durations

TW Duration	Average Travel + Waiting Time (min) per Stop		Average Travel Time (min) per Stop		Average Waiting Time (min) per Stop		Average Vehicles per Route	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Baseline (No TW)	1.66	3.81	1.66	3.81	0	0.00	1.00	1.10
1-Hour	2.59 (56.02%)	5.25 (37.79%)	2.32 (39.76%)	5.20 (36.71%)	0.27	0.04	1.23 (23.00%)	1.44 (30.96%)
2-Hour	2.13 (28.31%)	4.51 (18.54%)	2.07 (24.70%)	4.51 (18.53%)	0.06	0.00	1.17 (17.00%)	1.29 (17.37%)
4-Hour	1.78 (7.23%)	4.09 (7.41%)	1.78 (7.23%)	4.09 (7.41%)	0.00	0.00	1.12 (12.00%)	1.21 (10.10%)

Finally, Table 20 illustrates how different frequencies of time window constraints affect each setting. At 40% frequency, average travel and waiting time increases by 45.78% in urban routes, compared to 37.21% in rural routes. However, the vehicle use impact is higher in rural routes, with a 41.06% increase versus 37.00% in urban. These results support the idea that urban environments are more susceptible to time inefficiencies (such as waiting), while rural environments are more affected by fleet inefficiencies. This implies a strategic divergence in planning: urban networks require better time alignment and scheduling, whereas rural networks must focus on vehicle optimization and coverage planning.

In summary, while both urban and rural routes follow common trends under increased delivery constraints, the operational consequences differ. Urban systems offer greater flexibility and can tolerate moderate levels of constraint with less disruption, especially when time windows are broader and clustering is avoided. Rural systems, on the other hand, exhibit a faster breakdown in efficiency under tight windows or higher constraint percentages. For both, the most sustainable strategies involve limiting narrow windows to under 20% of stops, prioritizing 2–4 hour durations, and avoiding dense early-morning clustering—unless sufficient vehicle capacity and operational budget are available to support the added complexity.

Table 20. Urban vs. Rural under Different TW Frequencies

TW Frequency	Average Travel + Waiting Time (min) per Stop		Average Travel Time (min) per Stop		Average Waiting Time (min) per Stop		Average Vehicles per Route	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Baseline (No TW)	1.66	3.81	1.66	3.81	0	0.00	1.00	1.10
5%	1.89 (13.86%)	4.00 (5.19%)	1.79 (7.83%)	4.00 (4.95%)	0.09	0.01	1.00 (0.00%)	1.10 (0.00%)
20%	2.19 (31.93%)	4.62 (21.35%)	2.07 (24.70%)	4.61 (21.12%)	0.13	0.01	1.16 (16.00%)	1.29 (17.37%)
40%	2.42 (45.78%)	5.22 (37.21%)	2.31 (39.16%)	5.20 (36.58%)	0.11	0.02	1.37 (37.00%)	1.55 (41.06%)

4.4 Opportunities for Model Extension and Experimental Refinement

This study provides a solid foundation, but several opportunities exist to extend the numerical experiments and improve representativeness.

First, expanding time window frequency scenarios would enhance data granularity and allow for clearer trend identification. In this project, we tested 5%, 20%, and 40% frequencies to represent low, medium, and high constraint levels. However, including additional values like 10%, 30%, and 50% would help trace the curve of how operational time and vehicle requirements evolve more precisely.

Second, solving time could be extended. The 600-second time limit was a practical compromise to run 2,700 instances efficiently. Yet, test results suggest that increasing the limit to

1,800 seconds could yield slightly more optimal solutions—improving travel time by 0.27% and vehicle use by 0.69%—and offer more accurate comparisons across time window configurations.

Third, the experiment's geographic diversity could be expanded. All current rural routes are drawn from areas near Austin, Texas. Including more routes (e.g., 30–50 per context) and sourcing rural data from a wider variety of regions would enhance generalizability, especially in areas with very different infrastructure or delivery conditions.

Finally, future work could explore routes that fall between purely urban or rural classifications. Many real-world routes—such as those in suburban areas or on city outskirts—don't fit neatly into one category. Introducing "grey area" scenarios could offer valuable insights into how mixed-density environments respond to delivery constraints and help create more adaptable routing strategies for hybrid geographies.

5. Conclusion

To summarize the capstone, this section focuses on providing high-level insights and recommendations for delivery operators to consider when addressing the challenge of balancing route efficiency with customer needs. Additionally, we include other relevant factors that should be taken into account in real-world business practices.

5.1 Managerial Insights

A key finding from this project is that the 40-20-40 Distribution, which delivers a relatively smooth and balanced spread of time windows throughout the day, consistently achieves the highest route efficiency score across both urban and rural scenarios. This superior performance is largely attributed to its efficient use of vehicles, as it avoids the peak clustering seen in All-in-the-Morning or the unpredictability of Random Distributions. From a managerial standpoint, this suggests that flattening the distribution of delivery windows—even if customer preferences lean toward specific times—can yield significant gains in efficiency and cost savings. Delivery operators should proactively influence customer selection behaviors by offering incentives or structured slot availability that guides demand toward more balanced windows.

Additionally, operators must pay close attention to the total time range spanned by delivery windows across the business day. Overly long time spans can create bottlenecks, particularly toward the end of the delivery schedule, resulting in increased idle time and operational delays. As referenced in Sections 4.1.4 and 4.3, these extended ranges can inflate waiting times and underutilize fleet capacity. To mitigate this, companies should first establish a reliable baseline for total route duration without time windows, and then design delivery time window options that minimize late-day waiting time, enhancing both productivity and customer satisfaction.

Another critical insight is the disproportionate impact of short-duration windows—particularly 1-hour durations—and high-frequency constraints (e.g., 40%) on overall route efficiency. These configurations substantially drive up vehicle requirements, increasing both direct costs (labor, fuel, vehicle wear) and indirect costs (route planning complexity, risk of delivery failure). For delivery operators, this underscores the importance of cost-benefit evaluation before implementing strict time windows across a significant portion of deliveries. Unless there is a compelling business case—such as premium delivery fees or strategic service differentiation—the operational strain often outweighs the customer service benefits.

Lastly, it is important to recognize that rural routes are more sensitive to time window constraints than urban ones, especially when it comes to vehicle use. The limited route density and geographic spread make it more expensive to deploy additional trucks. Therefore, in rural environments, tight time windows or high constraint frequencies should be used sparingly. Delivery operators should reserve such constraints for premium accounts or very profitable deliveries, where the added service cost is justified by strong revenue or long-term strategic value.

To summarize, the findings and recommendations from this project offer practical value for logistics companies facing the challenge of balancing customer expectations with operational constraints. As e-commerce grows and customers demand tighter delivery windows, many companies struggle with rising last-mile costs, inefficient fleet use, and route fragmentation—especially in rural and suburban areas. This project addresses that tension by quantifying the trade-offs between service level and operational efficiency under various time window constraints. By offering a framework to evaluate these trade-offs and adjust decision parameters—such as the weights of operational time and vehicle use—this research provides a scalable approach that planners and operators can adapt to their own contexts. Implementing such models can help businesses optimize resources and make more strategic decisions on how to serve customers profitably and sustainably. As the delivery landscape evolves, these insights support more resilient, data-informed, and customer-focused operations

5.2 Potential Next Steps

One important avenue for future exploration involves refining the weights assigned to the two core components—operational time and vehicle use—in the calculation of route efficiency scores. In this project, a fixed weighting scheme was used (30% for operational time and 70% for vehicle use), based on general industry intuition about their relative cost impact. However, in real-world logistics, these weights can and should be tailored to reflect the specific economic context and strategic priorities of the delivery operation.

To develop a more accurate and adaptable weighting framework, delivery operators must consider a variety of real-world cost drivers. These include fuel prices, which can fluctuate significantly and impact cost-per-mile calculations; labor rates, which vary regionally and can influence the relative cost of time spent per stop; and vehicle-related expenses, such as maintenance, leasing or depreciation, insurance, and fleet management overhead. The combination of these cost elements contributes directly to the total cost-to-serve and, by extension, should influence how operational time and vehicle use are prioritized in efficiency scoring.

Additionally, factors like vehicle availability constraints, service-level agreements (SLAs), and local urban vs. rural delivery dynamics could further inform these weights. For example, in fleet-limited or high-traffic areas, vehicle use may be far more costly or operationally constrained than time, while in labor-scarce markets, minimizing driver hours might become the overriding priority. As such, a one-size-fits-all model may be too simplistic for effective decision-making across diverse operating environments.

To address this, future work could focus on building a flexible, data-driven model that allows operators to dynamically adjust these weights based on business-specific parameters. This could involve creating a cost model that incorporates regional or seasonal inputs, simulating scenarios under different economic assumptions, or integrating historical cost data to back-calculate more accurate weightings. Ultimately, by customizing the efficiency scoring system to reflect true operating conditions, logistics planners would be better equipped to evaluate trade-offs and make more informed, cost-effective routing decisions.

Another promising direction for future work would be to test time window performance under fixed fleet availability constraints. In the current study, vehicle count was allowed to adjust freely (up to 5) in response to operational demands. However, in many real-world situations—especially during peak seasons or in remote areas—companies may not be able to scale their fleet on short notice. Modeling scenarios with a fixed number of vehicles would reveal how time window constraints affect service quality, route feasibility, and waiting time when capacity is limited. This could help planners understand not just how many trucks are needed for a given service level, but also what performance trade-offs arise when those trucks simply aren't available.

REFERENCES

Agatz, N., Campbell, A., Fleischmann, M., & Savelsbergh, M. (2011). Time Slot Management in Attended Home Delivery. *Transportation Science, 45*(3), 285–449.

Amazon. (2021). Last Mile Routing Research Challenge. Retrieved from https://pubsonline.informs.org/doi/10.1287/trsc.2022.1173

Archetti, C., & Bertazzi, L. (2020). Recent Challenges in Routing and Inventory Routing: E-commerce and Last-Mile Delivery. *Networks, 77*(2), 255–268.

Baldacci, R., Christofides, N., & Mingozzi, A. (2008). An Exact Algorithm for the Vehicle Routing Problem Based on the Set Partitioning Formulation with Additional Cuts. *Mathematical Programming, 115*, 351–385.

Capgemini Research Institute. (2019). The Last-Mile Delivery Challenge. Retrieved from https://www.capgemini.com/insights/research-library/the-last-mile-delivery-challenge/

Christofides, N., Mingozzi, A., & Toth, P. (1981a). Exact Algorithms for the Vehicle Routing Problem, Based on Spanning Tree and Shortest Path Relaxations. *Mathematical Programming, 20*, 255–282.

Christofides, N., Mingozzi, A., & Toth, P. (1981b). State-Space Relaxation Procedures for the Computation of Bounds to Routing Problems. *Networks, 11*(2), 145–164.

Clarke, G., & Wright, J. R. (1964). Scheduling of Vehicle Routing Problem from a Central Depot to a Number of Delivery Points. *Operations Research, 12*(4), 568–581.

Cwioro, G., Hungerländer, P., Maier, K., Pöcher, J., & Truden, C. (2018). An Optimization Approach to the Ordering Phase of an Attended Home Delivery Service. *arXiv preprint*. https://doi.org/10.48550/arXiv.1805.00448

Dantzig, G., & Ramser, J. (1959). The Truck Dispatching Problem. *Management Science, 6*(1), 80–91.

Ellis, B. (2011). Waiting for the Cable Guy is Costing Us $38 Billion. *CNN Money.* Retrieved from https://money.cnn.com/2011/11/03/pf/cost_of_waiting/index.htm

Fukasawa, R., Longo, H., Lysgaard, J., de Aragão, M. P., Reis, M., Uchoa, E., & Werneck, R. F. (2006). Robust Branch-and-Cut-and-Price for the Capacitated Vehicle Routing Problem. *Mathematical Programming, 106*, 491–511.

Gallo, A. (2014, October 29). The Value of Keeping the Right Customers. *Harvard Business Review*. https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

Google. (n.d.). OR-Tools. https://developers.google.com/optimization

Kilby, P., Prosser, P., & Shaw, P. (1999). Guided local search for the vehicle routing problem with time windows. In S. Voss, S. Martello, I. Osman, & C. Roucairol (Eds.), *Meta-heuristics: Advances and trends in local search paradigms for optimization* (pp. 473–486).

Lin, Y. (2024, October 20). Global Ecommerce Sales Growth Report. *Shopify*. https://www.shopify.com/blog/global-ecommerce-sales

Project OSRM. (n.d.). Open Source Routing Machine. http://project-osrm.org

Schaap, H., Schiffer, M., Schneider, M., & Walther, G. (2022). A Large Neighborhood Search for the Vehicle Routing Problem with Multiple Time Windows. *Transportation Science, 56*(5), 1369–1392.

Solomon, M. M. (1987). Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations Research*, 35(2), 254–265.

U.S. Census Bureau. (2010). Rural America. Retrieved from https://mtgis-portal.geo.census.gov/arcgis/apps/storymaps/collections/189aa1dbd64c4c81b3b4a2b71124f6c6?item=1

Vareias, A., Repoussis, P., & Tarantilis, C. (2018). Assessing Customer Service Reliability in Route Planning with Self-Imposed Time Windows and Stochastic Travel Times. *Transportation Science, 53*(1), 256–281.

Vidal, T., Crainic, T. G., Gendreau, M., Lahrichi, N., & Rei, W. (2012). A Hybrid Genetic Algorithm for Multidepot and Periodic Vehicle Routing Problems. *Operations Research, 60*(3), 611–624.

Vinsensius, A., Wang, Y., Chew, E. P., & Lee, L. H. (2020). Dynamic Incentive Mechanism for Delivery Slot Management in E-Commerce Attended Home Delivery. *Transportation Science, 54*(3), 567–587.