Predicting Truck Parking Availability to Improve Driver Safety and Delivery Reliability

by

Bhavya Budhia

and

Samantha Clarke

Submitted on May 8th, 2026 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management
at the Massachusetts Institute of Technology

ABSTRACT

Truck drivers are essential for keeping U.S. supply chains moving, and the availability of safe and adequate parking is essential for keeping drivers moving. Unfortunately, finding parking remains a persistent challenge; recent estimates find there is only one truck parking space for every 11 drivers. A driver's inability to locate adequate parking creates real risks; they are forced to park in unsafe areas or pay steep fines for violating their Hours-of-Service (HOS), which are federally mandated and electronically tracked limits on how long they can drive before they have to stop. Current solutions like building new truck parking stops or deploying Truck Parking Information Management Systems (TPIMS) are prohibitively expensive and time consuming to implement. This capstone leverages current truck stop infrastructure to develop a data-driven predictive model that helps drivers determine when and where to find preferable parking. Using historical parking availability and traffic volume data, a driver's starting and ending locations, preference for amenities, and remaining Hours-of-Service, we calculate a utility score for each stop and rank them from most to least preferable. This gives drivers a clear view of where they should stop based on their personal needs and restrictions. We use these rankings to generate feasibility maps, providing industry stakeholders with powerful visuals that allow them to identify over and under-utilized stops in the current network and determine candidate regions for new truck stop infrastructure. This work demonstrates that a low-cost, simulation-based approach can meaningfully improve parking decision-making and add tangible value to the drivers who keep our supply chains flowing.

Capstone Advisor: Dr. Angela Acocella

Title: Research Scientist, MIT Center for Transportation and Logistics (CTL)

Capstone Advisor: Tim Russell

Title: Research Engineer, MIT Center for Transportation and Logistics (CTL)

TABLE OF CONTENTS

1. Introduction

Our sponsors are the Center for Transportation and Logistics (CTL) FreightLab and the Humanitarian Supply Chain Lab (HSCL) at MIT. FreightLab's overarching mission is to "drive innovation into the freight transportation industry," while HSCL strives "to understand and improve supply chain systems ... to better meet human needs." This capstone addresses a problem at the intersection of these two missions. It builds on a previous capstone project that identified service deficiencies in truck stop infrastructure and explored the challenges drivers face, both in day-to-day and emergency situations.

1.1 Motivation

Truck drivers are essential workers; supply chains everywhere rely on them to transport goods, on time and cost effectively. In 2024, 72.7% of U.S. domestic freight by weight was moved via truck (American Trucking Associations, 2025). Moving from origin to destination, however, can be challenging; drivers face traffic delays on major highways and are subject to strict Hours-of-Service (HOS) regulations. HOS regulations are federally mandated, electronically tracked and dictate how many hours drivers are permitted to be on duty before they must pull over and rest.

It is imperative that drivers be able to easily find safe and affordable parking, while making the most of their available HOS. However, while quality rest time is critical for driver safety on the road, locating parking presents drivers with yet another challenge. A 2025 MIT capstone, "Strengthening Truck Stop Infrastructure for Crisis Resilience," by Cafaro and Hsiang, highlighted the imbalance of required versus available parking. The research found that "90% of drivers struggle to find parking at night, and over 75% report frequent parking challenges" related to safety (Cafaro & Hsiang, 2025). Inability to locate safe parking – especially at night – can be detrimental to a driver's quality of rest. As a result, this can negatively impact their reaction time and ability to handle unforeseen events on the road, creating potentially dangerous outcomes.

1.2 Problem Statement and Key Questions

Our project's main objective is to leverage current truck stop infrastructure within the continental U.S. and develop a predictive model that assists drivers in finding better parking stops. The model will also inform drivers of the optimal time to begin searching for parking.

We believe that with recent developments in forecasting, it is possible to build a model that predicts both the availability and quality of truck parking around a driver's location. Based on these factors, the model advises drivers while ensuring compliance with HOS regulations. Our goal is to enable drivers to consult the model's output during their trips, helping them make more informed decisions about when and where to stop.

In this context, we explore the following topics:

- How can we leverage existing truck stops and predictive modeling techniques to improve the experience of drivers on the road?
- How do region-specific circumstances (e.g., local traffic patterns, infrastructure, geography, etc.) affect a truck stop's availability and selection, and how do they impact our model?

1.3 Project Goals and Expected Outcomes

The project's goal is to provide our sponsoring MIT labs with a robust predictive model that informs drivers' decision-making about where and when to look for parking, given their remaining Hours-of-Service, geographical location, and other relevant factors. This model also provides insights into the larger truck stop network (e.g., visibility to coverage gaps along major routes, data on truck stops frequently at or near capacity, etc.) that may benefit truck stop owners, state Departments of Transportation (DOTs), and other key stakeholders.

Our deliverables include:

- A predictive model that aids drivers when deciding where and when to look for parking
- A visualization tool for further truck stop infrastructure development, which aids relevant stakeholders in driving greenfield expansion

We share our model and results with industry stakeholders. We deliver feasibility frontier maps that drivers can use to reliably identify regions with parking that meets their needs and make best use of their remaining Hours-of-Service. We provide drivers with something they do not have access to today: confidence in where they plan to stop, so they can park safely and rest. Our results provide truck stop operators and DOTs with network visibility, critical for targeted amenity improvements and future evidence-based infrastructure investment support. By reducing the time drivers spend searching for parking, our model improves supply chain efficiency more broadly, enabling timely movement of goods and fewer missed delivery windows, which currently ripple delays throughout the value chain.

2. State of the Practice

The core objective of this capstone is to develop a predictive model that drivers can use to locate optimal parking, given their preferences, while adhering to their remaining Hours-of-Service. To meet this objective, we review literature on several topics. First, we analyze current truck stop industry challenges, as well as technological initiatives targeted at addressing those challenges that are currently being implemented in the field. Second, we review a common challenge researchers face when performing geospatial analysis and the solution we, and many others, are using to address the limitations of more standard approaches. Finally, we review previous research on modeling strategies for predicting truck parking availability.

2.1 Trucking Industry Challenges

Within the trucking industry, the lack of truck parking infrastructure is a well-known issue. Each year, the American Transportation Research Institute (ATRI) conducts a survey to identify critical issues in the trucking industry; in 2024, for the second year in a row, lack of available parking ranked second across all respondents (The American Transportation Research Institute, 2024). The United States DOT estimates that there is “one truck parking space nationally for every 11 truck drivers” (DAT Freight & Analytics, 2022), further supporting the findings in the ATRI survey.

Constructing new truck parking facilities seems like the obvious solution but requires significant financial and time investment. An ATRI study found that “most states (84%) do not have dedicated funding for truck parking”; they therefore rely heavily on federal funding for these initiatives (Fain & Murray, 2025). Furthermore, the same study found that the median cost to construct a single public truck parking space is about $93,500. Due to the significant investment required, states turn to other avenues to attempt to maximize the use of existing infrastructure.

Lack of available parking is not the only issue facing drivers today. As part of their work, Cafaro and Hsiang (2025) conducted 10 semi-structured interviews with a variety of stakeholders from the trucking industry, including drivers, owner-operators, and policy leaders. Their goal was to supplement existing research with relevant personal insights and identify where improvements to current infrastructure would be the most beneficial, based on current deficiencies. Three major topics consistently appeared in these interviews: parking, food, and restrooms. Across the country, drivers and industry stakeholders highlighted the importance of food availability and “high quality bathing facilities” like showers with reliable hot water and adequate water pressure (Cafaro & Hsiang, 2025).

2.2 Current Solutions

In the past several years, many states have started implementing Truck Parking Information Management Systems (TPIMS), which “convey real-time information to truck drivers about available parking, thereby maximizing utilization of existing truck parking capacity” (National Coalition on Truck Parking, 2018). This technological approach uses sensors (e.g., in-ground, infrared, video camera) to capture vehicles parked at a location, as well as vehicles leaving and entering. That data is passed to a processing hub, where it is converted to real-time parking-availability data, which can then be passed back to drivers directly to their truck, via roadside signage, or through an app or website.

Starting in 2015, the Mid-America Association of State Transportation Officials (MAASTO) successfully implemented a TPIMS that coordinated truck parking management across Iowa, Indiana, Kansas, Kentucky, Michigan, Minnesota, Ohio and Wisconsin (National Coalition on Truck Parking, 2018). MAASTO implemented various technologies across nine major highways in the region, spanning 130+ public and private parking sites. The TPIMS was fully operational by 2019. While certainly a step in the right direction, a regional TPIMS has several limitations. First, this technology was deployed on a subset of interstates and truck stops; it did not consider the entire available network. Second, implementation was time-intensive; even when considering a portion of the network, it took around three years to fully implement. Third, as drivers move outside of this region, they will face data fragmentation and will have to rely on different websites and applications for their parking information (assuming the next state has deployed a TPIMS). Because our work takes a different approach and uses historical and publicly available data, we move beyond these limitations to have meaningful impact.

2.3 Data Challenges in Geospatial Analysis

Geospatial analysis is a collection of methodologies and tools for viewing, manipulating, and analyzing location-based data to uncover patterns and understand relationships (de Smith et al., 2025). We integrate geospatial analysis into our capstone to better understand our in-scope network of truck stops, major highways, and traffic volumes. A key challenge in geospatial analysis of transportation networks is determining how to aggregate point-level data—such as traffic sensor readings or truck stop locations—into meaningful geographic units for modeling. Standard approaches, such as aggregating by ZIP code, introduce several limitations: these zones vary significantly in shape and size, are subject to change over time, and often create irregular areas poorly suited to capturing physical vehicle movement (Brodsky, 2018).

To address this, our capstone leverages H3, Uber's open-source Hexagonal Hierarchical Spatial Index, as the basis for aggregating traffic volume data. H3 partitions the globe into a uniform grid of hexagonal cells at multiple levels of resolution, enabling consistent and comparable geographic units across different cities and regions (Brodsky, 2018). Hexagons are particularly well-suited to transportation analysis; every hexagon shares an edge with each of its six neighbors, resulting in a single, uniform distance between any cell's center point and those of its immediate neighbors. This geometric property reduces errors when modeling the movement of vehicles through a network.

2.4 Predictive Modeling

Finally, we review prior research on modeling strategies for predicting parking availability. Tamaru et al. (2025) looked to address how insufficient truck parking availability could result in unsafe or unauthorized parking. Unlike the earlier studies analyzed in their research, which focused on predicting availability for single truck stops, the researchers focused on forecasting across multiple sites simultaneously, to capture interrelationships between truck stops within a state. The model they developed was able to consistently outperform baselines, indicating that the introduction of their "Regional Decomposition" factor led to more realistic and actionable predictions. Our capstone will extend their work by considering several additional factors: the parking availability dependencies at truck stops that may exist across state lines, traffic conditions, and truck stop amenities.

2.5 Relevance to Capstone Problem

When considered together, the literature reviewed in Section 2 establishes both the urgency of the problem this capstone seeks to address and the methodological foundation on which our model is built. We explore relevant and pervasive trucking industry challenges (parking scarcity, the high cost of new construction, and amenity deficiencies that truck drivers face in existing facilities), confirming that for a solution to be effective, it must go beyond locating available space. It must also account for quality, which reflects drivers' personal preferences. We then explore solutions currently available in the industry. The deployment of technological innovations like TPIMS has made some meaningful progress but has also proven to be expensive, slow to implement, and limited to a regional scope. These are all limitations that our data-driven, simulation-based approach is designed to overcome. The geographic aggregation framework, reviewed in Section 2.3, informs our data processing methodology. By adopting a hexagonal structure for traffic volume data aggregation, we smooth our traffic data into more consistent and geographically meaningful data, despite the uneven distribution of physical sensors. Finally, the predictive modeling research we review supports our choice of classification engine in our model (discussed further in Section 3.1) and shapes our decision to extend previous work by incorporating factors that existing models are not addressing: cross-state dependencies, traffic conditions, and amenity preferences. These seemingly disparate topics come together to support our overarching goal to develop a predictive model that helps truck drivers determine when and where to look for parking.

3. Methodology

In this section, we explain what kind of data we collect and prepare as well as our choice of methodology. We also discuss how we utilize this methodology to create a model that addresses our core objective, which is locating optimal parking given a driver's remaining HOS and preferences.

3.1 Methodology Selection

After considering our problem and looking at current state-of-the-practice methodologies, we decide to use a combination of geospatial analysis and machine learning techniques. The nature of our prediction task, forecasting whether parking will be available at a given truck stop at a future point in time, is a supervised classification problem with categorical output (parking available, parking unavailable). This structure rules out the use of purely descriptive or rule-based approaches and calls for a method capable of learning complex, non-linear relationships from historical data across many variables simultaneously.

We specifically select machine learning over traditional statistical approaches, such as linear or logistic regression, for several reasons. First, our prediction scope involves many interacting features (e.g., time of day, day of week, geographic location, etc.), many of which exhibit non-linear relationships with parking availability. Because they assume linearity and independence among predictors, traditional regression models are not best suited for handling the current datasets. Second, our historical parking availability source data is noisy and incomplete. Machine learning methods, particularly ensemble approaches like gradient boosting, handle data irregularities well; they are robust to outliers and perform well even when training data is imbalanced or sparsely labeled.

We use these techniques to predict future parking availability at each individual truck stop. We then use these predictions, along with factors such as traffic around truck stop locations, available amenities, and a driver's remaining Hours-of-Service, to calculate a unique utility score for each truck stop (see Section 3.3.2 for detailed calculations). Hundreds of drivers are simulated, and utility scores are aggregated to produce combined rankings.

3.2 Data Collection and Preparation

For the purposes of this research, we gather various data from both public and private sources which we utilize to represent the current network and inform our predictive model. Our analysis focuses on four states: Alabama, Florida, Georgia, and South Carolina. Georgia and South Carolina house two critical ports for East Coast importing and shipping operations in the U.S.: Savannah and Charleston, respectively (United States Department of Transportation, 2026). We replicate hundreds of realistic truck routes throughout this region within our model.

3.2.1 Truck Stop Data Collection

We work with an industry partner that provides technological solutions for drivers and truck stop owners to help them gather data on both truck stop geographic information and extensive crowdsourced historical parking availability. They also provide us with information about location-specific truck stop amenities (e.g., payments accepted, food offerings, showers/laundry facilities, etc.).

3.2.2 Traffic Volume Data Collection

We use two different traffic volume data sets from the U.S. Federal Highway Administration (FHWA). The first is the average annual daily traffic volume (AADT) across all U.S. roads for the year 2024 from U.S. DOT's Highway Performance Monitoring System (HPMS) (Federal Highway Administration, 2024a). AADT is the average number of vehicles that pass a specific point on a road each day, averaged over an entire year. The second is hourly traffic volume data collected in the same year across the U.S. (Federal Highway Administration, 2024b).

The HPMS dataset allows us to map traffic volume to the physical road network in the continental U.S. Because of the data size and complexity, and assuming that long-haul trucks will realistically only be travelling on major highways, we only consider the roads and road types which account for the top 80% of truck traffic in the country. Figure 1, for example, shows the in-scope road network and traffic snapshot from the HPMS data for the state of Georgia.

Figure 1: Georgia Traffic Network (Federal Highway Administration, 2024a)

The hourly traffic data captures the continuous traffic flow through the sensors spread across the country. Figure 2 shows the sensor locations in the continental U.S.

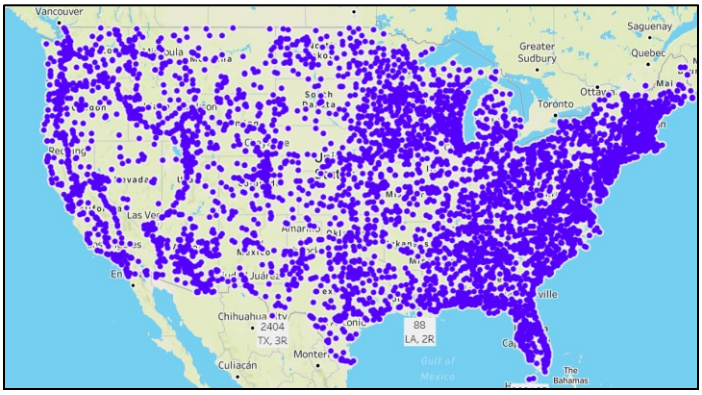

Figure 2: Traffic Sensors in Continental U.S. (Federal Highway Administration, 2024b)

We use this dataset to observe traffic trends and patterns and to create a custom forecast that helps us predict traffic volume at any given truck stop in the future.

3.2.3 Truck Stop Data Preparation

Because most of the truck stop parking data provided is crowdsourced from truck drivers themselves, there are significant data quality issues, including duplicate records and missing information.

For the identification of duplicate records, we reference a study done in Hungary that considered all truck stops within 200 meters of each other as a single truck stop location (Csendes et al., 2021). For our analysis, we use a similar scale of 0.1 miles to identify duplicate records. Additional analysis reveals that minor naming errors and misidentification of stops caused duplicates in our dataset.

We then perform quality checks on this year's data as compared to the truck stop demographic data that was used in a previous capstone (Cafaro & Hsiang, 2025). Here, we focus on validating total parking spots available at a truck stop. Using the same grouping method of 0.1 miles, we identify similar records in our current data versus the 2025 study; where differences exist, we use the truck parking spot count from the 2025 study. This decision is supported by spot checks using Google Maps.

The truck stop amenity data also requires analysis and standardization prior to use in our model. Most fields are a binary response of yes or no (e.g., On-site laundry? Yes. Wi-Fi available? No.). However, there are two columns in this dataset- number of ATMs and number of showers on-site- that reflect the number of amenities available. We aggregate these columns across all truck stops within the same state and standardize to a value between 0 and 1. Once we standardize these disparate data types, we average all values for each truck stop to come up with a composite amenity score (known as $score_{amenities}$, see Section 3.3.2). Currently, we weigh each amenity equally within the score, but further analysis could explore varying weights based on individual drivers' preferences.

Finally, we review and prepare the crowdsourced historical parking availability data. Driver responses indicate the level of parking spots available (many, moderate, or few) at a given truck stop at a given time. Because the data is voluntarily input by drivers on the road, the level of detail per truck stop varies widely. Some truck stops have multiple availability inputs per hour; others only have a single input once every couple of weeks or months. Because of this variability, we aggregate the availability data by hour, day of week, and month and choose the most frequent descriptor. Drivers also have the option to select "paid" when reporting the parking availability at a truck stop; we remove those records, which represent approximately 3.5% of the total data, because they fall outside the scope of our analysis of parking availability.

3.2.4 Traffic Volume Data Preparation

We find instances of duplicate records in the HPMS data that appear to be accidental. For the purposes of this study, we remove the duplicate records and do not believe this negatively impacts our model.

The hourly traffic volume data also requires some processing. Because the traffic-capturing sensors are spread across the country, and their distribution is not uniform throughout the U.S. highway network, we use H3 to aggregate the volume data from the FHWA sensor network. Rather than grouping sensors by three-digit ZIP code (ZIP3)—which produces irregular zones of varying geographic extent—we assign each sensor to an H3 hexagon. We then average traffic volume readings from multiple sensors within the same hexagon, smoothing out anomalies from individual sensors and filling gaps in areas with sparse coverage (known as $score_{traffic}$, see Section 3.3.2). We assume truck stops located within a given hexagon experience broadly similar traffic conditions, providing a consistent and geographically meaningful basis for our prediction models.

3.3 Application of Methodology

In this section, we explain how the remaining components of our utility score are calculated, including a review of how future parking availability and traffic volumes are modeled.

3.3.1 Predicting Traffic and Truck Stop Availability

Using the hourly traffic volume data and historical truck stop availability, we build a model that predicts both traffic and truck parking availability for a given location and time. Currently, we employ a simple but effective method for predicting traffic volume. Our traffic volume data is aggregated into hexagons; we identify the hexagon the truck stop in question is in and retrieve the historical average for traffic within that hexagon at the given time.

The event-based truck parking availability data from our partner company is not consistently comprehensive enough to create a reliable prediction model directly. To overcome this challenge, we use a quasi-Monte Carlo approach to generate synthetic data. We sample random starting locations from our existing data and simulate random starting times (known as t_0) and estimated times to arrival (ETAs), based on distance to the destination, average driving speed, and our calculated traffic volume factor. For each scenario, we then refer to our existing historical availability data to generate the last known availability status for t_0 and the observed parking status nearest to the driver's ETA (within a 60-minute window). We generate the prediction result for each simulated data point from the historical truth found in our source data.

After preparing our data, we then use an XGBoost classification technique to transform our "many, moderate, few" availability data to a prediction of whether a given truck stop will have available parking or not; our confidence in this prediction is known as $score_{parking}$, see Section 3.3.2. Our model achieves an accuracy score of 0.67, meaning we correctly predict the parking status at the time of a driver's ETA 67% of the time. Minimizing both false positives (predicting parking when none was available) and false negatives (predicting no parking when parking was available) is critical for ensuring drivers do not have to go to multiple truck stops to find adequate parking. Our model has a false positive rate of ~29.6% and a false negative rate of ~41.7%. We prefer both values to be lower than they are currently, but we prioritize parameterizing our XGBoost model to minimize the false positive rate because this is the more detrimental error for drivers.

3.3.2 Utility Score Model

For the full model and scenario testing, we simulate different truck drivers in our network. We use the approximate latitude/longitude for each origin and destination city, varying randomly by +/- 0.2 degrees. Remaining HOS is a randomly assigned value ranging from one to six hours. The baseline for personal weighted preference for amenities is 0.2 and is randomly simulated +/- 0.3 (with the lowest possible value being 0.05) to reflect driver behavior (known as $w_{amenities}$). For example, some drivers simply look for a place to park with a restroom and one restaurant or convenience store for food, while others look for more expansive amenities like an on-site gym and laundry facilities. Still others prioritize safety features like on-site security and well-lit parking lots. $W_{amenities}$ captures these variations and steers drivers to truck stops that meet their needs.

To calculate the effect of traffic congestion on travelling speed, we use a foundational volume delay function, developed by the U.S. Bureau of Public Roads (BPR) in 1964, to estimate the relationship between traffic volume and transportation delays (Bureau of Public Roads, 1964).

The BPR function is calculated as follows:

$$T = T_0 \left[1 + \alpha \left(\frac{V}{C}\right)^{\beta}\right]$$

where T is current travel time, T_0 is free-flow speed (we assume 55 miles per hour), V is volume (vehicles/hour), C is capacity (vehicles/hour), α represents how severe congestion gets overall, and β controls how sharply congestion affects the result. Industry accepted values for α and β are 0.15 and 4 respectively and are used here (Bureau of Public Roads, 1964). We use T to determine if a driver will be able to reach a given truck stop within their remaining HOS. For example, if truck stop A is 200 miles away, T is 45 miles per hour, and the driver's HOS is four hours, the truck stop is infeasible because the driver can only travel 180 miles within his remaining time.

Using the simulated driver profiles, processed traffic and truck stop data, and outputs of our traffic-volume and parking-availability prediction models, we calculate a utility score for each individual truck stop under consideration.

The formula used to calculate truck stop utility is:

$$\begin{aligned} Utility\_Score = {} & (w_{parking} \times score_{parking}) + (w_{amenities} \times score_{amenities}) \\ & + (w_{capacity} \times score_{capacity}) + (w_{detour} \times score_{detour}) \\ & + \left(w_{traffic} \times score_{traffic}\right) + (w_{dest_{norm}} \times score_{dest_{norm}}) \\ & + (w_{hos\_utilization} \times score_{hos\_utilization}) \end{aligned}$$

Table 1 defines each parameter in further detail.

Parameter	Definition
$w_{parking}$	Researcher-defined weight applied to $score_{parking}$, assumed 0.20
$score_{parking}$	Predicted probability that parking will be available at a truck stop upon driver arrival
$w_{amenities}$	Researcher-defined weight applied to $score_{amenities}$, ranges from 0.05-0.50
$score_{amenities}$	Composite amenity score, average of standardized amenities
$w_{capacity}$	Researcher-defined weight applied to $score_{capacity}$, assumed 0.15
$score_{capacity}$	Standardized score representing relative parking capacity (number of parking spots) at a truck stop
w_{detour}	Researcher-defined weight applied to $score_{detour}$, assumed 0.20
$score_{detour}$	Distance-based inverse score, penalizes a truck stop that takes a driver beyond their destination[1]
$w_{traffic}$	Researcher-defined weight applied to $score_{traffic}$, assumed 0.10
$score_{traffic}$	Standardized inverse congestion score, derived from historical traffic intensity data
$w_{dest\text{-}norm}$	Researcher-defined weight applied to $score_{dest\text{-}norm}$, assumed 0.9
$score_{dest\text{-}norm}$	Standardized inverse score representing proximity of a truck stop to the destination, prioritizes stops closer to destination
$w_{hos\text{-}utilization}$	Researcher-defined weight applied to $score_{hos\text{-}utilization}$, assumed 0.10

[1] We use the real-world distances to calculate these parameters, using SCGraph (Makowski et al., 2025)

$score_{hos\text{-}utilization}$	Ratio of adjusted travel time to remaining HOS, measuring how efficiently available driving hours are utilized

Table 1: Utility Model Parameters

Truck stops are ranked using these utility scores, with the highest score representing the preferred truck stop for a driver scenario on a lane. We simulate 540 driver scenarios per lane, and we calculate utility scores for all truck stops in each scenario. Next, we calculate a combined utility score for each truck stop.

The formula used to calculate combined utility is:

$$CU_i = (0.50 \times AvgUtility) + (0.20 \times FeasibleRate) + (0.15 \times Top10Rate) + (0.15 \times WorstCaseUtility)$$

Table 2 defines each parameter in further detail.

Parameter	Definition
$\boldsymbol{AvgUtility}$	Normalized mean of simulated scenario utility scores
$\boldsymbol{FeasibleRate}$	Ratio of feasible scenarios divided by total scenarios
$\boldsymbol{Top10Rate}$	Number of times a truck stop was ranked 10th or better divided by total scenarios
$\boldsymbol{WorstCaseUtility}$	Normalized value: across scenarios, bottom 10th percentile utility score

Table 2: Combined Utility Model Parameters

The combined utility score balances average truck stop utility with the fraction of scenarios that were feasible (penalizing stops that are often infeasible), the fraction of scenarios where the stop fell in the top 10 ranking (rewarding consistently high performance), and the bottom 10th percentile utility score across all scenarios (capturing performance in worst-case scenarios).

Finally, we create feasibility frontier maps that highlight preferred regions and truck stops, segregated by a driver's remaining Hours-of-Service (see Section 4.1.3, Figure 3 for an example).

We use the weighted frontier distance formula to calculate, on average, how far a driver can travel within each HOS bucket:

$$weighted\ frontier\ distance = \frac{\sum(avg\_truck\_stop_{mi} \times w_{distance})}{\sum w_{distance}}$$

where *avg_truck_stop_mi* is the average distance between origin location and truck stop and $w_{distance}$ is the product of multiplying each stop's combined utility (CU_i) with the *FeasibleRate.*

4. Results and Discussion

In this section, we present the results of our simulation study and discuss the implications for drivers, truck stop operators, and transportation policymakers. We conclude the section with a review of the data and model limitations present that affect the interpretation and application of our results.

4.1 Results

First, we present the parameters and assumptions taken in our simulation runs; we then review in-depth the results.

4.1.1 Simulation Overview

To capture high volume, realistic freight flows throughout the Southeast, we focus our simulated movements to the following major lanes: Jacksonville, FL, to Miami, FL; Charleston, SC, to Atlanta, GA; and Savannah, GA, to Birmingham, AL. On each lane, we simulate 540 freight flows, each flow having a unique driver profile. Across the simulations, Hours-of-Service remaining range from one to six hours, and amenity preference weights range from 0.05 to 0.50. Starting times are distributed across morning, afternoon, and evening hours to capture temporal variation in traffic conditions and parking availability. Overall, we capture a variety of driver behavior across these factors, and the simulations produce a dataset of ranked truck stop recommendations that serve as the foundation of our analysis.

4.1.2 Parking Availability Prediction Performance

To predict parking availability at each truck stop, we train an XGBoost classification model on processed historical availability data and achieve an overall accuracy of 67%. From a practical standpoint, false positives (cases where the model predicts parking availability, but none exists) are the higher risk error type; they result in a driver arriving at a full truck stop with insufficient HOS remaining to reach an alternative. Because of the real-world implications, minimizing this type of error is critical to tuning our model. Our XGBoost framework demonstrates improved predictive power as compared to the baseline, but richer training data would presumably improve accuracy.

4.1.3 Truck Stop Rankings

For each freight lane simulation, we calculate a combined utility score for each truck stop. We then create a feasible frontier map: a visual representation of the preferred and accessible truck stops for drivers, given various HOS constraints. Figure 3 shows the feasible frontier map for a driver travelling from Jacksonville, FL (green flag) to Miami, FL (red flag). Similar visuals for the additional freight lanes we simulated can be found in our project GitHub.

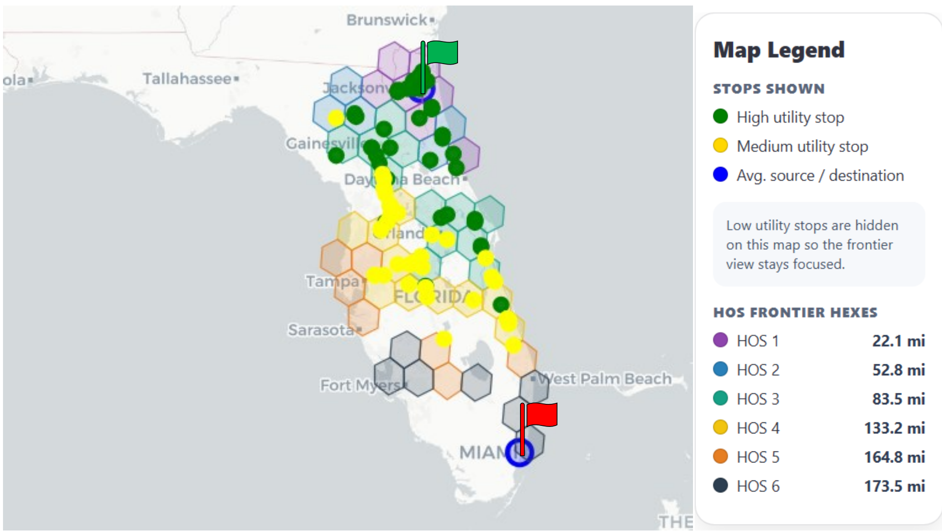

Figure 3: Feasible Frontier Map, Freight Flowing Jacksonville, FL, to Miami, FL

The frontier map in Figure 3 gives truck drivers something they do not have today; this color-coded view acts as a quick reference for lane-specific preferable truck stops, given a driver's remaining HOS. Green markers indicate highly ranked truck stops, while yellow markers indicate moderately ranked stops. For example, if a driver is heading from Jacksonville, FL, to Miami, FL and has four remaining Hours-of-Service, they know they should start looking for parking as they cross from the green to yellow hexagon region. This view enables drivers to balance maximizing their remaining HOS with locating a truck stop with sufficient amenities to meet their preferences.

The feasible frontier maps highlight a key finding from our model. Corridor structure – the physical highway network – strongly affects parking dispersion patterns. Figure 4 shows the feasible frontier map for a driver travelling from Savannah, GA (green flag) to Birmingham, AL (red flag).

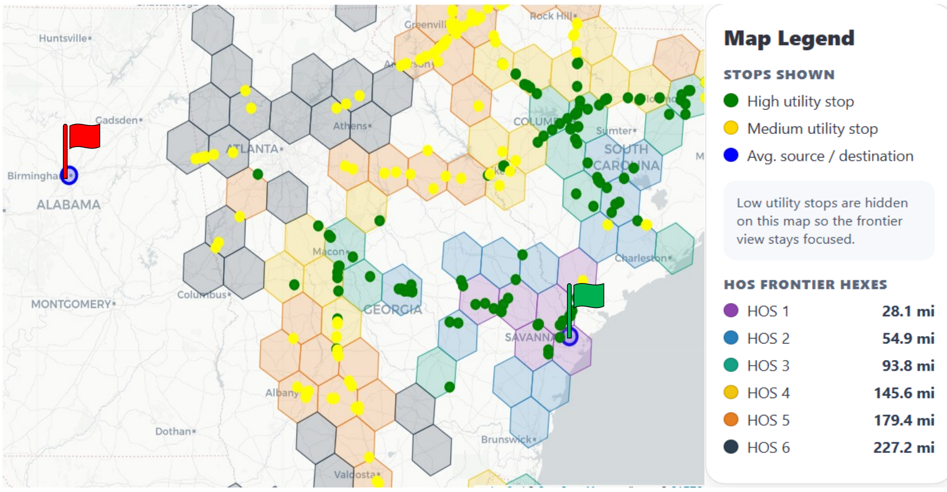

Figure 4: Feasible Frontier Map, Freight Flowing Savannah, GA, to Birmingham, AL

The Savannah to Birmingham corridor looks much different than the corridor going from Jacksonville to Miami. Figure 3, Jacksonville, FL, to Miami, FL, depicts a highly linear corridor with few practical routing alternatives. As a result, preferrable parking options are clustered tightly along the main highway, which has the potential to create higher competition for premium locations and a greater sensitivity to parking shortages. Figure 4, Savannah, GA, to Birmingham, AL, paints a different picture. There are multiple clear inland routing corridors, which allow parking demand to spread across a wider network. This suggests there is greater routing flexibility on this lane, meaning drivers have a lower dependence on single stops.

Reviewing these two lanes, there is also a notable difference in the average number of miles a driver can travel and still reach a preferred truck stop (truck stop with high or medium utility). On the Jacksonville, FL, to Miami, FL lane, the average distance a driver can travel is approximately 174 miles, assuming they have six Hours-of-Service remaining. We contrast that with the Savannah, GA, to Birmingham, AL lane, where drivers can travel approximately 227 miles if their remaining HOS is six hours. This difference of approximately 50 miles is caused by the interaction between corridor structure, traffic volume, and available amenities. Because there is only one direct corridor to travel between Jacksonville and Miami, the impact of traffic congestion is felt more strongly, particularly as drivers approach and move through major cities like Orlando, FL and West Palm Beach, FL. On the other hand, there are multiple direct paths from Savannah, GA, to Birmingham, AL; traffic volume is spread out between them (meaning the impact of congestion is less significant) and drivers have more choices in terms of truck stops and available amenities, allowing drivers on this lane to travel farther within the same period of time.

Next, we analyze the impact that time of day has on utility rankings; we segment our simulations into three categories: morning (4AM through 12PM), afternoon (12PM through 8 PM), and evening (8PM through 4AM); we choose these ranges based on generally accepted time-of-day groupings and density of available data. This allows us to visualize temporal effects on truck stop utility, again taking into consideration the driver's remaining Hours-of-Service. Figure 5 represents the high, medium, and low-ranking truck stops for a driver with 6 Hours-of-Service remaining, travelling from Jacksonville, FL, to Miami, FL across the three-time categories.

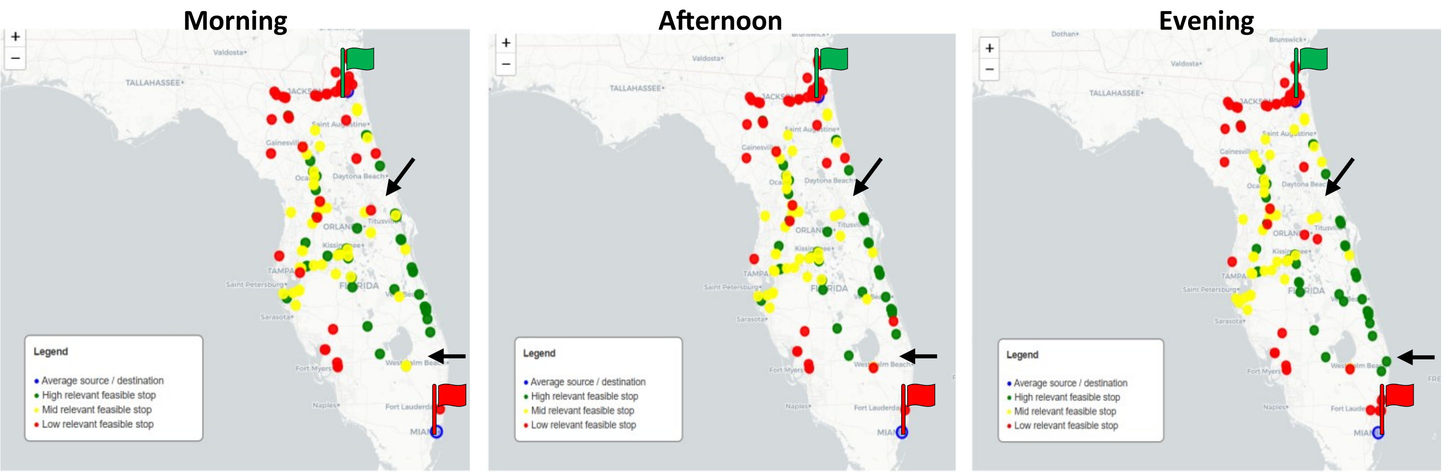

Figure 5: Preferred Truck Stop Network, Freight Flowing Jacksonville, FL, to Miami, FL

While there are some differences across the time horizons (indicated with black arrows), the three visuals are extremely similar, indicating that time-of-day changes matter less than expected, particularly at high-ranking truck stops. This result was consistent across the three major lanes we simulated. High-ranking truck stops remain relatively stable throughout the day; within the top three ranked stops, there was limited shifting in position. This finding supports the idea that the most preferable truck stops in the network have structural advantages that render them resilient, not just temporary advantages due to time of day.

After simulating hundreds of unique drivers across three major Southeastern freight lanes, corridor structure emerges as one of the dominant drivers of truck stop accessibility and overall preference ranking; the physical highway network exerts a high impact on how far drivers can feasibly travel and how demand is distributed across the network. Time of day, by contrast, has a lower impact than expected. High-ranking truck stops prove structurally resilient, maintaining low variability in utility rankings regardless of when drivers are on the road.

4.2 Implications

First and foremost, the most direct beneficiaries of this model are the truck drivers themselves. Our simulation results highlight how constrained the parking search window truly is: on the Jacksonville, FL, to Miami, FL corridor, drivers operating at maximum HOS have a feasible range of only approximately 174 miles. Because that lane has almost no practical routing alternatives, available parking clusters tightly along a single highway, intensifying competition for the best stops. Currently, navigating the search for available truck parking is a source of stress for drivers; they have limited real-time information and a high degree of uncertainty whether parking will be available when they reach a truck stop. Additionally, Hours-of-Service constraints put significant pressure on drivers; if they arrive at a truck stop that is full towards the end of their HOS clock, they may be forced to park illegally or in an unsafe area to avoid an HOS violation. The feasible frontier maps our model produces address this directly: rather than wasting time searching, drivers receive a color-coded, lane-specific view of preferred stops matched to their remaining HOS. This has the potential to reduce both HOS violations and safety incidents by facilitating earlier, better-informed parking decisions. Our model creates consistent visibility to reliable and quality parking options; it has the potential to improve overall driver rest quality, which in turn would contribute to safer outcomes on the road.

Second, this model has impactful implications for stakeholders who play a critical role in addressing the truck parking shortage. For truck stop operators, our model creates visibility to their competitive position within the current network. Notably, our results show that high-ranking truck stops remain stable across morning, afternoon, and evening hours, suggesting that the most preferable stops hold structural advantages, not just temporary benefits tied to time of day. This means that investments in additional capacity or amenity development would have durable impact. If owners can identify where their operational or facility offerings are lacking, they could more confidently focus efforts on closing those gaps. When considered from the perspective of state DOTs or transportation policymakers, this model is a low-cost, data-driven solution which can be used to inform future infrastructure investment priorities without requiring the time-consuming deployment of technological solutions like TPIMS. For example, our feasibility maps identify the Jacksonville, FL, to Miami, FL corridor as a single-lane route where parking demand clusters tightly along one highway; this is a key freight lane where additional investment could have significant impact.

Finally, our model results have broader implications for supply chains in general. The main goal of this model is to help drivers consistently find reliable and quality parking within the existing network. The more time a driver spends looking for parking, the less time they have left in their HOS clock for transporting goods. Often, these extended search periods result in delays that compound across supply chains and can result in missed delivery windows, increased costs, and reduced asset utilization for carriers. Our model allows drivers to more effectively manage their time and reduce costly delays.

4.3 Limitations

While the model demonstrates meaningful predictive capability and produces impactful insights for key stakeholders, there are several limitations in the underlying data, model design, and geographic scope that must be considered when reviewing the results and assessing real-world implications.

We begin by assessing the limitations within the various data sources used. First, the historical parking availability data is crowdsourced from truck drivers; as a result, it is unevenly distributed across time of day, day of week, month, and truck stop. Some truck stops have consistent and abundant reporting; others have one input per month. This introduces noise into both our training and validation phases, decreasing the generalizability of our end model. To combat this, we implement the quasi-Monte Carlo approach detailed in the methodology section (with a reasonable accuracy result), but a more consistent, higher volume of availability data would have yielded a stronger model.

Second, in the current model, amenity data is binary for all but two attributes, and composite amenity scores treat all amenities as equally weighted. This model choice most likely does not reflect real drivers' preferences as they exist in the current market. It is reasonable to assume, for example, that longer distance truck drivers would place greater preference on amenities like gyms or on-site laundry than shorter haul drivers would. Richer research and analysis on drivers' preferences by amenity type would be required to set more accurate amenity specific weights.

We will next review limitations of the model itself. First, we are able to achieve a prediction accuracy of 67% with our XGBoost classification; while this represents an improvement over the logistic regression baseline predictions (accuracy of 65%) and uses a methodology better suited to our data types, it also means that approximately one in three parking availability predictions generated is incorrect, a rate that may be too high for drivers in real-world time-critical situations. To increase confidence in predictive power, significant additional availability data and further model refinement would be required.

Second, the utility function's penalty weights are currently researcher-defined, rather than empirically calibrated against actual driver routing decisions. Without validation against actual data from real drivers, there is a risk that the relative weighting of these parameters does not accurately reflect the tradeoffs that truck drivers make in practice, potentially causing the model to consistently over- or under-prioritize certain truck stop characteristics.

Third, the quasi-Monte Carlo simulation approach does have some limitations. While the simulation approach is effective for generating a large and diverse population of synthetic scenarios, it produces routes that approximate the distribution of real freight lanes rather than replicating them exactly. Due to factors like carrier mandated preferred stops, contractual fuel agreements, or familiarity bias toward well-known truck stop chains, it is reasonable to assume that actual driver behavior on Southeast corridors would deviate systematically from these simulated distributions. This in turn would indicate that the model's rankings may not fully reflect the choices drivers would make in practice.

Finally, the current model evaluates truck stops independently at a single point in time rather than optimizing across the full sequence of stops a driver might make on a multi-day trip. This single-stop framing simplifies the recommendation engine but misses the interdependencies between successive parking decisions that characterize real-world long-haul driving, where the choice of one stop shapes the geographic and temporal constraints facing the driver at every subsequent decision point. In order to model more realistic multi-day freight flows, the model could be extended to consider multiple truck stop recommendations in sequence.

The geographic scope of this analysis is an important boundary condition on the generalizability of its findings. From the outset, we develop the model and validate exclusively across four Southeast states, a region characterized by relatively flat terrain and dense interstate network. Whether the model's predictive performance and utility score rankings would hold in regions with meaningfully different infrastructure density and topography, like the Northeast or Midwest regions of the U.S., remains an open question that future work should address. Within the focus region itself, the analysis is further bound by its focus on interstate and major highway corridors. Even though these roads account for the majority of long-haul truck traffic, the in-scope network excludes secondary and off-highway truck stops that a meaningful set of drivers use and whose omission means the model may not capture the full range of parking options available to drivers on these key freight lanes.

5. Recommendations

In this section, we will translate our model findings into specific recommendations, organized into two parts: management recommendations directed at industry practitioners and policymakers and a roadmap for future work aimed at researchers and technology partners who can address the limitations previously highlighted. In both sections, the recommendations share a common thread; the truck parking shortage is not simply an infrastructure problem to be solved by construction alone. It is a data and decision-making problem that stronger analytical tools, targeted investment and improved information sharing between drivers, operators and policymakers can meaningfully reduce.

5.1 Management Recommendations

For truck stop operators, the most actionable takeaway from this analysis is the ability to benchmark individual stop performance against the broader network using utility score rankings as a diagnostic tool. Operators of stops that consistently rank poorly across simulated scenarios, particularly those where low scores are driven by amenity deficiencies as compared to traffic volume outside of their control, should conduct a targeted review of their facility offerings. This is especially relevant on high-competition corridors like Jacksonville, FL, to Miami, FL, where drivers are highly reliant on a single highway, as compared to the Savannah, GA to Birmingham, AL corridor, where drivers have multiple inbound highway options. Additional amenity market research would be beneficial, but a focused review of gaps will provide insights into amenity development and investment that will most increase the stop's overall utility for drivers. In this way, our simulation framework provides a repeatable, evidence-based method for determining whether operational changes translate into improved network utility over time.

For state DOTs and transportation policymakers, the most immediate recommendation is to partner with academic institutions to incorporate simulation-based demand modeling into their arsenal for truck parking infrastructure planning, alongside or in place of more expensive sensor-based TPIMS deployments. Our feasibility frontier maps provide a visual, evidence-based starting point for prioritizing regions where new truck stop capacity would generate the greatest utility for drivers. Due to the significant capital investment required to construct just one truck parking spot, directing scarce capital toward locations with the highest level of unserved driver demand is both analytically sound and financially responsible. At the federal level, the cumulative output of this modeling framework, especially as it is extended into additional regions, could support the development of a national truck parking demand map that gives agencies and state partners a consistent, comparable basis for allocating infrastructure funding across the contiguous U.S.

For industry partners and technology providers operating in the truck parking industry, the central recommendation is to explore integrating the utility score ranking framework into existing driver-facing applications as a real-time or near-real-time recommendation engine. The model's architecture is designed to complement, rather than replace, existing data infrastructure; the model's predictive ability would be bolstered with increased data availability and quality. To fully realize the potential of this model, technology providers should prioritize expanding crowdsourced data collection at under-reported stops, through avenues such as driver incentive programs or automated data partnerships. Carriers and fleet operators represent an additional and underutilized partner in this effort. Integrating the model's recommendations into electronic logging device platforms or transportation management systems would allow parking decisions to be made proactively rather than reactively.

5.2 Future Works

In this section, we suggest future work that could be pursued, building off the predictive model we create in this capstone.

5.2.1 Model Enhancements

One of the most impactful near-term enhancements to the model would be to expand the collection of driver-sourced availability data; this would make the predictive component of our underlying model more robust and reliable and enable the generation of more accurate recommendations as conditions change during a driver's trip. Additionally, the utility function's penalty weights should be empirically calibrated using data gathered directly from drivers (either through surveys/focus groups or through performing a deep sentiment analysis on extensive truck stop review data that could be provided through our industry partner) to ensure that tradeoffs built into the model reflect how drivers actually make parking decisions rather than researcher assumptions.

The model's single-stop recommendation capability is a strong candidate for extension into a multi-stop optimization approach, where the model plans an optimal sequence of stops across a multi-day trip. This would more accurately reflect the realities faced by long-haul truck drivers and substantially increase the model's practical use for drivers managing complex multi-day routes.

Finally, expanding the model's geographic coverage beyond the four state Southeast region focused on here would test the generalizability of the current framework and progressively build toward the national truck parking demand map that policymakers and federal agencies could use to inform infrastructure investment decisions.

5.2.2 Resilience and Scenario Testing

Our model, focused on the Southeast region of the U.S., provides a solid foundation for resilience and scenario testing. Further work should formally model the impact of hurricane-season disruptions on corridor-level parking availability. Using historical storm tracks and their documented effects on freight flows, one could simulate how the network responds when a subset of truck stops becomes inaccessible and demand is redistributed across the remaining infrastructure. In addition to natural disasters, our model framework could be extended to other exogenous shocks relevant to supply chain resilience: port closures, major highway incidents, or sudden demand surges driven by emergency situations. These stress tests would not only validate the model's behavior under edge cases but would also generate insights useful to state DOTs and emergency management agencies about which corridors and truck stops are most critical to maintain access to during a regional disruption. It would also provide visibility into where pre-positioning of capacity or resources would have the greatest impact on freight continuity.

5.2.3 Stakeholder Engagement

To realize the full potential of this model, sustained engagement with the stakeholders who are best equipped to validate, refine, and deploy it in practice is paramount. As an immediate next step, the simulation results should be shared with state DOTs in Alabama, Georgia, South Carolina, and Florida to solicit feedback on whether the model's predictive power aligns with their operational experience and to receive initial reactions relative to the model's potential for driving greenfield recommendations. These stakeholders will have better visibility to land use, zoning and right-of-way constraints that will impact greenfield selection but that this model does not currently consider.

Piloting the utility score ranking framework with a willing fleet operator or third-party logistics provider would create an invaluable real-world test of whether the model's recommendations translate into adoption by drivers. It would also facilitate a test of usability and behavioral barriers that cannot be identified through simulation alone.

Engaging directly with drivers through structured interviews or surveys would complement the quantitative outputs of the model with the kind of qualitative insights that shape truly useful tools. Understanding how drivers currently make parking decisions, what information they trust and under what conditions they would allow an algorithmic recommendation to drive their decisions are all components that would enhance this model. Beyond the immediate value to our sponsoring labs at MIT's Center for Transportation and Logistics, this work demonstrates that a low-cost, data-driven, simulation-based approach can meaningfully address an infrastructure-scarcity problem that has historically been treated as solvable only through expensive physical construction or sensor-heavy real-time monitoring. For the trucking industry, our framework offers a complement—or, in many regions, a viable alternative—to TPIMS deployments, accelerating progress and addressing a problem that consistently ranks among drivers' most pressing daily concerns. This framework contributes directly to highway safety, HOS compliance, and the wellbeing of the drivers. For supply chain functions more broadly, our combined predictive and utility-scoring methodology illustrates how publicly available and crowdsourced data can be unified to generate decision-grade insights at both the operational level (a driver's next stop) and the strategic level (where to invest in new capacity). By translating a scarce and uncertain resource into a predictable, navigable network, this capstone advances both of our sponsors' missions and provides truck drivers with something they do not have today: confidence in where they are going, so they can park safely, rest, and get back on the road.

References

American Transportation Research Institute. (2024). *Critical issues in the trucking industry*. https://truckingresearch.org/2024/10/critical-issue-in-the-trucking-industry-2024/

Brodsky, I. (2018, June 27). *H3: Uber's Hexagonal Hierarchical Spatial Index*. Uber Blog. https://www.uber.com/en-EG/blog/h3/

BPR (Bureau of Public Roads). 1964. *Traffic assignment manual*. Washington, DC: Urban Planning Division, US Department of Commerce.

Cafaro, C., & Hsiang, H. L. (2025). *Truck stops as critical infrastructure: Enhancing resilience and driver safety in times of crisis* (Master's Capstone, Massachusetts Institute of Technology). https://landline.media/wp-content/uploads/2025/07/Cafaro_Hsiang_Capstone_2025-05-09-Final.pdf

Csendes, B., Albert, G., Szander, N. & Munkácsy, A. (2021). Where Truck Drivers Stop – Application of Vehicle Tracking Data for the Identification of Rest Locations and Driving Patterns. *Promet - Traffic&Transportation, 33* (6), 821-832. https://doi.org/10.7307/ptt.v33i6.3962

DAT Freight & Analytics. (2022). *Dry van report: There's only one parking space for every 11 truckers nationwide*. Retrieved December 7, 2025 from https://www.dat.com/blog/dry-van-report-theres-only-one-parking-space-for-every-11-truckers-nationwide

de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2025). *Geospatial analysis: A comprehensive guide to principles, techniques and software tools* (7th ed.). The Winchelsea Press. https://www.spatialanalysisonline.com/

Fain, A., & Murray, D. (2025). *Expanding truck parking at public rest areas*. The American Transportation Research Institute.

Federal Highway Administration. (2024a). *Highway Performance Monitoring System (HPMS) - policy*. Retrieved December 7, 2025 from https://www.fhwa.dot.gov/policyinformation/hpms.cfm

Federal Highway Administration. (2024b). *Special Tabulations: U.S. Traffic Volume Data - policy*. Retrieved December 7, 2025 from https://www.fhwa.dot.gov/policyinformation/tables/tmasdata/

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Makowski, C., Saragih, A., Guter, W., Russell, T., Heinold, A., & Lekkakos, S. (2025). *SCGraph: A Dependency-Free Python Package for Road, Rail, and Maritime Shortest Path Routing Generation and Distance Estimation* (SSRN Scholarly Paper No. 5388845). Social Science Research Network. https://doi.org/10.2139/ssrn.5388845

National Coalition on Truck Parking. (2018). *Truck parking availability detection and information dissemination*. Federal Highway Administration.

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, (pp. 1394–1401). IEEE. https://doi.org/10.1109/ICMLA.2018.00227

Tamaru, R., Cheng, Y., Parker, S., Perry, E., Ran, B., & Ahn, S. (2025). *Truck parking usage prediction with decomposed graph neural networks* (arXiv:2401.12920). arXiv. https://doi.org/10.48550/arXiv.2401.12920

United States Department of Transportation. (2026, February 24). *Latest Supply Chain and Freight Indicators*. https://explore.dot.gov/views/FreightIndicators/ContainerShipCapacity