TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FMRI-Y5R3- 21	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle AN INTERACTIVE PLATFORM FOR LARGE SCALE TRUCK ACTIVITY DETECTION AND ANALYSIS USING CONNECTED VEHICLE DATA – PHASE 2		5. Report Date: March, 2024 6. Performing Organization Code:
7. Author(s) Evangelos I. Kaisar, Praveen Edara, Yaw Adu-Gyamfi		8. Performing Organization Report No.
9. Performing Organization Name and Address Florida Atlantic University Freight Mobility Research Institute		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 69A3551747120
12. Sponsoring Agency Name and Address Freight Mobility Research Institute (FMRI) Florida Atlantic University 777 Glades Rd., Bldg. 36, Boca Raton, FL 33431		13. Type of Report and Period: Final 14. Sponsoring Agency USDOT
15. Supplementary Notes		

16. Abstract

Freight transportation plays a vital role in regional economic development by ensuring the timely movement of goods and the efficient delivery of services. However, public and private stakeholders often struggle to leverage freight data effectively due to fragmentation, inconsistent data formats, and limited analytical infrastructure. Existing data management systems within public agencies, such as State Departments of Transportation (DOTs), are not optimized to handle the scale, diversity, and velocity of modern freight data streams. This limitation hinders informed decision-making, infrastructure planning, and sustainable freight management.

This report presents the design, development, and deployment of a GPU-accelerated, web-based freight analytics and visualization platform that addresses these challenges. The platform is built using scalable, open-source technologies and is designed to support high-speed integration, real-time analytics, and multidimensional visual exploration of freight data. Key innovations include a spatiotemporal conflation framework for merging disparate freight datasets—specifically, weigh-in-motion (WIM), freight facility, and connected vehicle (CV) GPS data—as well as a user-centered interface that allows planners to interactively explore trends, activity patterns, and operational bottlenecks.

17. Key Words Freight transportation, Economic development, Data fragmentation		18. Distribution Statement No restrictions. This document is available to the public through Fmri.fau.edu	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 41	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

FREIGHT MOBILITY RESEARCH INSTITUTE

College of Engineering & Computer Science

Florida Atlantic University

Project ID: Y5R3-21

AN INTERACTIVE PLATFORM FOR LARGE SCALE TRUCK ACTIVITY DETECTION AND ANALYSIS USING CONNECTED VEHICLE DATA – PHASE 2

FINAL REPORT

by

Principal Investigator
Evangelos Kaisar, PhD.
Professor
Department of Civil, Environmental and Geomatics Engineering
Florida Atlantic University

Co-Principal Investigator
Praveen Edara, Ph.D., P.E.
Professor
Department of Civil and Environmental Engineering
University of Missouri-Columbia

Co-Principal Investigator
Yaw Adu-Gyamfi, Ph.D.
Assistant Professor
Department of Civil and Environmental Engineering
University of Missouri-Columbia

March, 2024

ACKNOWLEDGEMENTS

This project was funded by the Freight Mobility Research Institute (FMRI), one of the twenty TIER University Transportation Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT).

DISCLAIMER

EXECUTIVE SUMMARY

Freight transportation plays a vital role in regional economic development by ensuring the timely movement of goods and the efficient delivery of services. However, public and private stakeholders often struggle to leverage freight data effectively due to fragmentation, inconsistent data formats, and limited analytical infrastructure. Existing data management systems within public agencies, such as State Departments of Transportation (DOTs), are not optimized to handle the scale, diversity, and velocity of modern freight data streams. This limitation hinders informed decision-making, infrastructure planning, and sustainable freight management.

This report presents the design, development, and deployment of a GPU-accelerated, web-based freight analytics and visualization platform that addresses these challenges. The platform is built using scalable, open-source technologies and is designed to support high-speed integration, real-time analytics, and multidimensional visual exploration of freight data. Key innovations include a spatiotemporal conflation framework for merging disparate freight datasets—specifically, weigh-in-motion (WIM), freight facility, and connected vehicle (CV) GPS data—as well as a user-centered interface that allows planners to interactively explore trends, activity patterns, and operational bottlenecks.

The system's backend leverages cuDF and Dask to perform parallel, in-memory data processing across GPU cores, enabling rapid aggregation, filtering, and transformation of datasets containing tens to hundreds of millions of records. These capabilities are integrated into a browser-based dashboard built with Bokeh, which allows users to dynamically explore maps, charts, and time-series plots without requiring any software installation. Advanced features include cross-filtering, real-time hover insights, heatmaps of freight intensity, and seamless rendering of historical and live data.

Beyond visualization, the report presents a preliminary freight activity detection framework that extracts operational patterns from GPS trajectory data. By combining feature engineering with unsupervised clustering (K-means), the platform identifies four distinct freight activity profiles—from short-haul urban deliveries to long-haul regional transport. These classifications provide valuable insights into freight behavior and infrastructure usage, even in the absence of shipment manifests or detailed logistical metadata.

Benchmarking results confirm that the platform supports real-time interaction and rendering for datasets exceeding 100 million records, with query latencies consistently under 100 milliseconds. This performance, coupled with its modular and low-cost architecture, makes it an ideal solution for transportation agencies seeking to modernize their freight data capabilities without the expense of traditional enterprise systems. Overall, the project demonstrates how modern big data frameworks, GPU computing, and intuitive design can converge to provide a powerful tool for freight planning, policy formulation, and infrastructure investment. The platform not only empowers agencies to unlock actionable insights from existing data assets but also sets a strong foundation for integrating future data sources, such as autonomous vehicle telemetry, 5G-enabled sensors, and real-time incident feeds.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND

The efficient movement of goods and timely provision of services is critical to the economic and sustainable development of a region. Public decision makers require a comprehensive picture of freight movements to understand how freight transportation supports economic development, how land use affects freight transportation, and how transportation infrastructure supply impacts private sector freight and commercial activity. Freight data is available from many public and private sources. However, the data may vary significantly in terms of collection method, data quality, existence of gaps, availability or timeframe (daily, monthly, and quarterly), format (shape files, documents, tables, etc.) and suitability. The lack of coordination among freight data vendors not only prevents the seamless integration of data sources but also hinders data-driven decision making.

Although agencies such as State Departments of Transportation (DOTs) have transportation data management systems for storing and processing data streams, they are not uniquely designed to handle such large, heterogeneous, and multi-resolution data streams. They have limited analytical capabilities that will enable them to integrate, mine, visualize and predict on large, multivariate datasets at reasonable speeds (Mostafa et al. 2018, Richardson et al. 2014). Traditional data warehouses are stretched to the limit due to the enormous size and speed, and the significant variety of datasets across different vendors in terms of collection method, data quality, availability (daily, monthly or quarterly), and format (shapefiles, documents, table, videos, etc.). The need for frameworks that can help integrate and visualize information from existing freight databases is therefore crucial. In the current report, we leverage recent advances in big data and user-centered visualization to develop a scalable framework that allows for multidimensional visualization and analytics to be carried seamlessly on large freight datasets.

Recent advances in Big Data Analytics is enabling organizations to digest large amounts of data and transform them into actionable insights (Kluger et al, 2013, Adu-Gyamfi et al, 2016). This innovation is being fueled by massive open data platforms, driven by machine learning and empowered by low cost cloud computing. This new wave of invention could be leveraged to enable transportation agencies to identify the usefulness of their diverse datasets and to explore previously untapped applications.

1.2 PROJECT OBJECTIVES

Under the above context, the primary goal of this project is to deliver a prototype design and deployment of an interactive, web-based platform that will assist decision makers to seamlessly integrate and analyze its freight datasets. The prototype platform is designed to be significantly faster and cheaper (by using open-sourced software solutions for development) compared to conventional data warehouses, which are heavily reliant on relational databases housed in big, costly enterprise machines. The four key subobjectives that arise from the primary goal are:

1) Develop a spatio-temporal conflation framework that enables seamless integration of three key freight data sources including: weigh-in-motion (WIM), freight facility, and traffic flow data.

2) Leverage state-of-the-art big data frameworks to develop a massively parallel database to store the integrated data on a cluster of servers enabled with Graphical Processing Units (GPUs). We leverage the immense computational power of the GPUs to carry out analytics and visual rendering, on-the-fly, via a Structured Query Language (SQL) which interacts with the underlying database.
3) Offer a low-cost, but effective data integration and analytics platform by leveraging open-source software for designing, developing and deploying the platform.
4) Provide user-centered, web-based data visualization to allow for easy interaction with the platform. Additionally, provide near-instant rendering of queries on simple charts and maps to enable decision makers to drill down insights quickly.

1.3 Report Organization

The remainder of this report is organized as follows: Chapter 2 covers information on connected vehicle data and its implication within freight. In Chapter 3, we describe the interactive visualization platform and its key features and opportunities. Chapter 4 highlights the spatial temporal analysis performed on the freight GPS data. Within Chapter 5 a preliminary track activity prediction is performed on the freight GPS data.

CHAPTER 2 CONNECTED VEHICLE DATA

Connected vehicle data refers to the real-time and historical information collected from sensors, telematics systems, and onboard computers in vehicles, transmitted via cellular, satellite, or other wireless networks. In the trucking industry, this data encompasses a wide range of metrics, including vehicle performance, driver behavior, location, fuel consumption, and cargo conditions. The advent of connected vehicle technologies has transformed the trucking sector by enabling real-time monitoring, predictive maintenance, and enhanced operational efficiency. The sections below explore the applications, challenges, and future implications of connected vehicle data within the freight industry.

2.1 Application Of Connected Vehicle Data in Trucking

Connected vehicle data is leveraged in various ways to optimize trucking operations. Key applications include:

1) Fleet Management and Optimization - Telematics systems collect data on vehicle location, speed, and route efficiency, enabling fleet managers to optimize operations. Real-time GPS tracking allows for dynamic route planning, reducing fuel consumption and delivery times (Salazar-Cabrera et al., 2019).
2) Predictive Maintenance - Sensors embedded in trucks monitor engine performance, tire pressure, and other critical components, generating data that can predict maintenance needs. By analyzing this data, fleet operators can schedule maintenance proactively, reducing downtime and preventing costly breakdowns (U.S. Department of Transportation, 2019)
3) Supply Chain Visibility - Shippers and logistics providers use truck data to track freight in real time, improving delivery accuracy and reducing delays.
4) Enhanced Safety & Compliance - Electronic logging devices (ELDs) track hours of service (HOS) to ensure compliance with FMCSA regulations (Federal Motor Carrier Safety Administration, 2023) which ensures adherence to hours-of-service rules. Moreover, Advanced driver-assistance systems (ADAS) use vehicle-to-vehicle (V2V) data to prevent collisions.

2.2 Challenges in Leveraging Connected Vehicle Data

Despite its advantages, the use of connected vehicle data in trucking faces several challenges:

1) Data Privacy and Security - Cybersecurity threats, including hacking and data breaches, pose risks to fleet operations. A report by the National Highway Traffic Safety Administration (NHTSA) highlighted the need for robust encryption and secure data protocols in connected vehicles (National Highway Traffic Safety Administration, 2022).
2) Integration and Interoperability - Trucking companies often use multiple data systems from different vendors, leading to compatibility issues. Standardizing data formats and ensuring interoperability across platforms remain significant hurdles (Datta et al., 2017).
3) Regulatory Hurdles: Different regions have varying data-sharing and privacy laws, complicating cross-border operations.

2.3 FUTURE IMPLICATIONS

The future of connected vehicle data in trucking is promising, driven by advancements in artificial intelligence (AI), 5G connectivity, and autonomous technologies. AI-powered analytics will enable more accurate predictive maintenance and demand forecasting, while 5G will facilitate faster and more reliable data transmission for real-time decision-making (Abdelkader et al., 2021). Additionally, the integration of vehicle-to-everything (V2X) communication will enhance coordination between trucks, infrastructure, and other vehicles, paving the way for platooning and autonomous trucking.

3.1 System Overview

The visualization platform was developed as a modular system combining advanced data processing techniques with modern web-based visual interfaces. Its architecture reflects the project's core principles: scalability, speed, usability, and cost-effectiveness. The general process is shown in Figure 1 below.

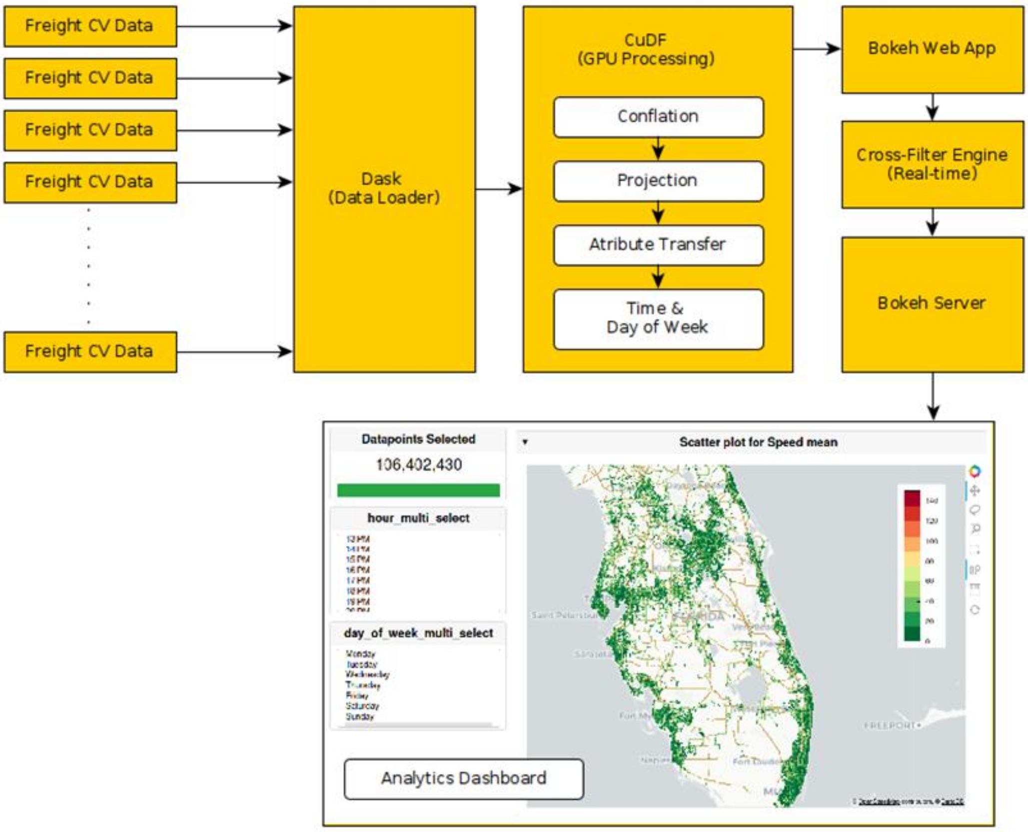

Figure 1. Data Processing Pipeline

The number of records corresponding to each aggregation level is presented in Table 2, providing a view of how temporal granularity affects data volume and, by extension, model complexity and training time.

i. Data Processing Pipeline

The platform ingests large-scale freight datasets and processes them using GPU-accelerated frameworks:

- ❖ cuDF (NVIDIA’s GPU DataFrame library) enables rapid, in-memory processing of tabular freight data.
- ❖ Dask provides parallel and distributed computing capabilities, allowing for the processing of millions of records in real time.

This combination allows the system to handle temporal aggregations, spatial transformations, and data cleaning at scale.

ii. Data Loading and Pre-Processing

Efficient data ingestion and preprocessing are foundational to ensuring high-performance analytics and visualization in the freight platform. Given the large volume and heterogeneity of the connected vehicle (CV), weigh-in-motion (WIM), and facility datasets, the data loading pipeline was designed to prioritize parallelism, memory efficiency, and trajectory-aware processing. This section describes the complete workflow for preparing raw data into a clean, structured, and analyzable format using GPU-accelerated tools, notably NVIDIA's cuDF and Dask libraries.

Data Ingestion and Format Standardization

The raw datasets arrived in various formats, including CSV, Parquet, and shapefiles. Each dataset—CV GPS pings, WIM measurements, and freight facility metadata—was loaded using either cuDF's GPU-accelerated I/O operations or Dask's distributed reading functions for large partitions that exceeded GPU memory capacity.

To ensure consistency across datasets:

- Timestamps were parsed and standardized into UTC format.
- Coordinates (latitude, longitude) were validated and cast into a common spatial reference system (typically EPSG:4326).
- Truck identifiers, when available, were cleaned and normalized to a common format (e.g., hashed string IDs).
- Columns with missing or null values were dropped or imputed based on domain-specific rules.

GPU-Accelerated Preprocessing with cuDF and Dask

The preprocessing tasks—ranging from filtering and sorting to type casting and feature generation—were distributed across multiple GPU cores using cuDF (for individual GPU performance) and Dask-cuDF (for multi-GPU parallelism). Key preprocessing steps included:

- Schema Alignment: Column names and data types were aligned across batches to enable downstream joins and fusions.
- Memory Optimization: Data types were cast to optimal formats (e.g., float32 vs. float64, category types for discrete variables) to reduce memory usage and improve speed.
- Filtering Noise: Erroneous or outlier GPS points were removed using rule-based filters (e.g., extreme speeds > 130 mph, null locations, timestamps out of range).

Trajectory Reconstruction and Trip Segmentation

For CV data, which contains timestamped GPS pings per vehicle, a trajectory sorting and segmentation process was performed:

- Sorting by Truck ID and Timestamp: Each truck's GPS records were grouped by ID and sorted chronologically to reconstruct vehicle trajectories.

- Trajectory ID Assignment: Using heuristics such as time gaps and distance thresholds (e.g., >10 minutes between consecutive points or >2 miles), trips were segmented, and a unique Trip_ID was assigned.
- Stop Identification Flags: Points where vehicle speed < 5 mph and displacement < 0.2 miles were flagged as candidate stop locations for later analysis.

This process enabled the generation of trip-level features and the mapping of freight behavior across time and space.

Temporal Resampling and Binning

To harmonize data collected at different frequencies:

- CV data was resampled into fixed temporal bins (e.g., 5-minute, 15-minute intervals) for temporal aggregation of speeds, volumes, and densities.
- WIM and facility data were similarly aggregated into daily or hourly summaries for integration with trajectory-level analyses.

Integration Readiness

After preprocessing, all datasets were stored in GPU-friendly in-memory dataframes using cuDF and persisted in Parquet format for efficient I/O in later stages. Metadata was appended to track the data source, resolution, and preprocessing history to ensure reproducibility.

iii. Geospatial Confirmation Method

A critical component of integrating connected vehicle (CV) data with freight infrastructure is accurately associating vehicle positions with the underlying road network. Due to GPS inaccuracies, signal drift, and spatial resolution differences between datasets, raw GPS coordinates often do not align precisely with mapped roadway geometries. To address this challenge, we implemented a geospatial conflation framework that uses geohash-based spatial matching to link CV trajectory points to a high-resolution Florida reference road network.

Overview of Approach

The conflation process involves encoding both the GPS trajectory data and the Florida road centerlines into spatially indexed geohashes. These geohashes act as compact spatial bins that enable rapid proximity matching between vehicle locations and road segments. This geohash-based approach offers several advantages:

- It reduces the need for expensive point-to-line spatial joins.
- It supports GPU-accelerated indexing and filtering.
- It provides consistent spatial granularity for downstream analyses.

Reference Road Network

The Florida Department of Transportation's (FDOT) road network shapefile was preprocessed to extract road centerlines, segmented by unique road segment identifiers. For each road segment:

- A representative set of geohashes was generated using its bounding geometry.

- Attributes such as road class, directionality, and functional classification were preserved for later enrichment.

GPS Point Encoding and Indexing

Each GPS ping from the CV dataset was:

- Encoded into a geohash at precision level 7–9, balancing spatial resolution (approx. 150m–5m) with computational efficiency.
- Annotated with truck ID, timestamp, speed, and direction of travel.

These geohashes formed the basis for a fast, one-to-many lookup against the road network's geohash index.

Matching Algorithm

The matching process involved the following steps:

- Initial Geohash Join: For each vehicle ping geohash, all overlapping or neighboring geohash-encoded road segments were retrieved. This step dramatically narrowed the candidate set of road segments.
- Distance Filtering (Optional): A secondary filtering step computed haversine distances between the GPS ping and candidate road segments to select the nearest segment if multiple matches occurred. This step was performed selectively, depending on the density of road segments in the area.
- Temporal Continuity Validation: To maintain realistic vehicle trajectories, the algorithm checked for logical continuity in road segment assignments across consecutive pings. Abrupt changes were flagged as potential errors and corrected using trajectory smoothing techniques.
- Final Assignment: Each GPS ping was annotated with a road_segment_id, enabling downstream aggregation by route, corridor, or functional class.

Output and Usage

The output of the conflation process is a GPS-enhanced trajectory dataset, where each record is enriched with road segment metadata, including:

- Segment ID
- Road type (interstate, arterial, local)
- Direction of travel
- Corridor designation (where applicable)

This conflated dataset forms the backbone for all subsequent map-based visualizations, corridor performance analysis, and route-level clustering. It also allows freight planners to identify high-traffic freight corridors, bottlenecks, and underutilized infrastructure segments with spatial precision.

iv. Data Projection and Transformation

To enable accurate spatial analysis and seamless integration of multiple freight-related datasets, all processed data underwent a rigorous projection and transformation workflow. This step is essential for harmonizing coordinate systems, aligning geometry across data layers, and supporting real-time rendering on the web-based visualization platform.

Purpose of Projection

Freight data originates from multiple sources—GPS pings from connected vehicles, shapefiles for reference road networks, WIM sensor locations, and freight facility data—each potentially using a different spatial reference system (SRS). Inconsistent coordinate systems can lead to spatial misalignment, rendering errors, and incorrect analytics. Projection and transformation ensure:

- Spatial coherence among all datasets for conflation and spatial joins.
- Accurate overlay of dynamic vehicle trajectories on static infrastructure.
- Compatibility with web-based rendering engines that require specific geographic projections.

Reference Coordinate System

The projection process was designed to convert all spatial datasets into two main coordinate systems depending on the task:

- EPSG:4326 (WGS84 – Latitude/Longitude): Used for initial alignment of GPS pings and as the baseline for all datasets received in geographic coordinates. This standard ensures compatibility with external geographic data and supports global referencing.
- EPSG:3857 (Web Mercator Projection): Required for rendering within the browser-based visualization platform. Web mapping libraries (such as Leaflet, Mapbox, and BokehJS) utilize this projection for fast tile-based visualization and accurate zooming/panning behavior.

Transformation Workflow

The transformation was applied through the following steps:

- Coordinate Validation: All GPS and shapefile-derived data were validated for missing or malformed coordinates and filtered to ensure they fell within the bounding box of Florida.
- Initial Normalization to EPSG:4326: Regardless of original format, all datasets (including local State Plane shapefiles or facility coordinates in NAD83) were first projected into WGS84 to ensure a common baseline.

- Conflation Preparation: Both vehicle GPS pings and road network geometries were maintained in EPSG:4326 to support geohash-based matching and trajectory reconstruction. This ensured that the geohash encoding used a consistent spatial resolution.
- Web Visualization Projection: After processing and enrichment, the spatial layers (e.g., route segments, truck paths, WIM sites) were transformed into EPSG:3857. This conversion supports high-performance, tile-based rendering in the web dashboard.

Caching Transformed Layers: For performance optimization, transformed layers were cached as vector tiles or JSON-based spatial objects (e.g., GeoJSON) to enable instant rendering and interaction without repeated reprojection.

Benefits of Harmonized Projection

- The projection and transformation framework delivers several advantages:
- Ensures accurate spatial alignment of dynamic and static data layers.
- Reduces latency and rendering errors in the interactive platform.
- Enables precise distance and direction calculations for route analytics.
- Simplifies integration of external basemaps, such as OpenStreetMap or satellite imagery.

v. Web Interface Layer

The web interface serves as the primary access point for exploring, analyzing, and interacting with freight data in real time. The frontend is built using Bokeh, a powerful Python-based interactive visualization library, and is deployed via the Bokeh Server framework. The system is developed in Python 3.10, with direct connectivity to GPU-accelerated backend data pipelines that enable low-latency data exploration.

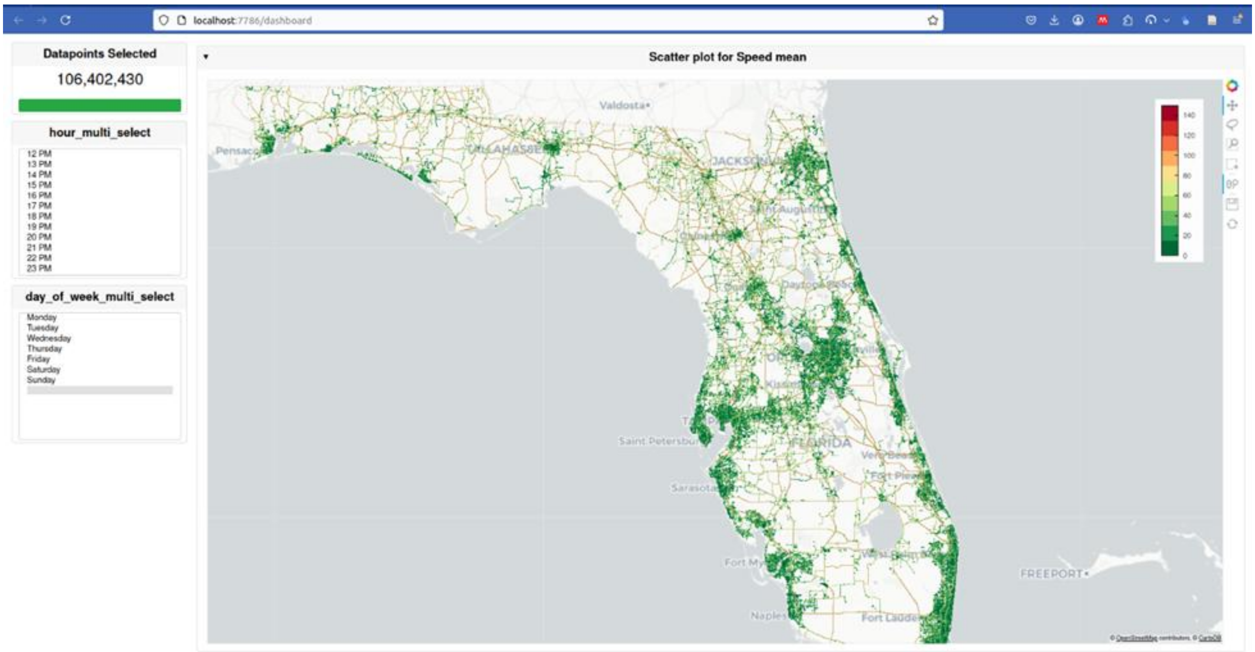

Figure 2. Interactive web dashboard for spatiotemporal analysis of freight data

Key Design Principles

The interface is designed with the following principles in mind:

- Speed – Enabled by GPU-preprocessed data using cuDF for rapid filtering and aggregation.
- Interactivity – Achieved through Bokeh's cross-filtering capabilities across charts and maps.
- Simplicity – Requires no client-side installation; it is browser-based and platform-agnostic.
- Scalability – Can be deployed locally for development or on a remote server for multi-user access.

Technology Stack

- Frontend: Built entirely with Bokeh and JavaScript callbacks for responsive interaction.
- Backend: Served using Bokeh Server, enabling Python callbacks to dynamically update plots based on user input.
- Data Preprocessing: Handled with cuDF, providing in-memory GPU-accelerated transformation of large freight datasets for real-time rendering.

Core Visualization Features

The web interface allows users to interactively explore freight movement patterns across spatial and temporal dimensions. Key components include:

- Map Visualizations: Interactive, tile-based maps show freight routes, GPS trajectories, and activity heatmaps over Florida's road network. Geospatial layers include:
    - Freight truck density by corridor
    - Speed heatmaps by time of day
    - Facility and WIM sensor locations
- Time-Series and Distribution Charts:
    - Hourly/daily speed and volume trends
    - Trip length and stop duration histograms
    - Congestion profiles across peak and off-peak hours
- Cross-Filtering and Interactive Controls, where users can apply filters based on:
    - Time ranges (e.g., day, week, custom span)
    - Geographic zones or map selections
    - Truck class, direction, or road type
    - Stop types and activity profiles

Selections in one component (e.g., a time range slider or facility dropdown) automatically update all related plots and maps, thanks to Bokeh's reactive programming model.

Deployment and Accessibility

The dashboard is served via Bokeh Server, which maintains a live Python session for each user. This architecture enables:

- Real-time updates without page reloads
- Low latency, even with millions of data points
- Multi-user access on secure internal or cloud-hosted environments

The platform is compatible with:

- All major modern browsers (Chrome, Firefox, Safari, Edge)
- Desktop and large-format displays (e.g., control center monitors)
- Touch-enabled tablets for field inspection or planning use cases

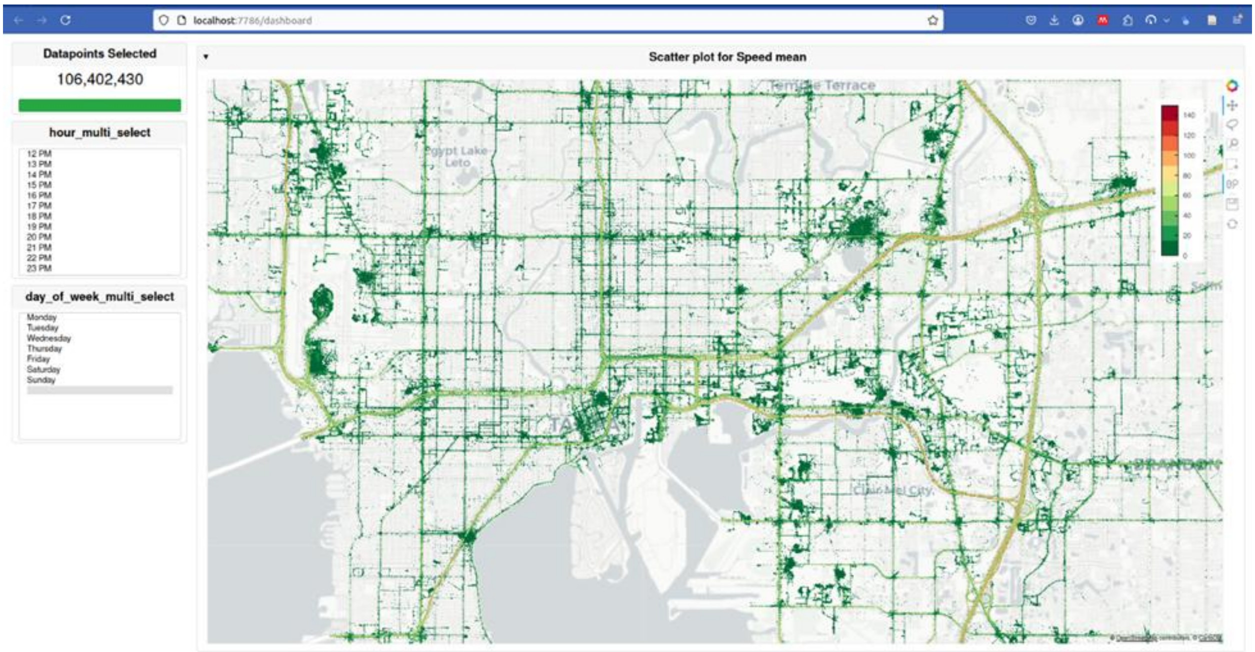

Figure 3. High-resolution freight truck speed heat map in Florida

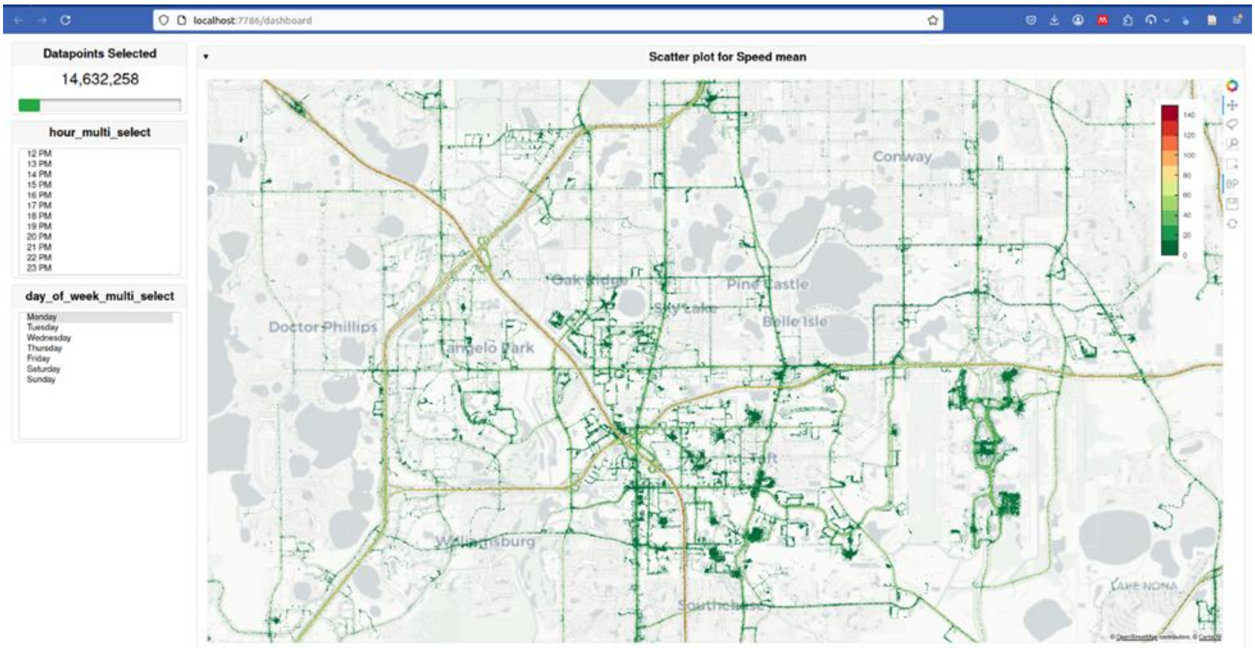

Figure 4. Daily analysis of freight speeds and volume on Florida highways

3.2 Key Features

The platform includes a set of core features that support multidimensional exploration of freight activity data:

i. Map-Based Visualizations

Interactive maps show freight flows and densities across Missouri's road network. These include:

- Dynamic rendering of truck routes based on CV data,
- Density heatmaps of high-activity corridors,
- Visual overlays of freight facilities and WIM locations.

ii. Cross-Filtering Across Views

Filters for time range, vehicle type, location, and facility type are all interconnected. A user's selection in one chart or control updates all other views in real time—providing a seamless analytical experience.

iii. Temporal Trend Analysis

Users can explore time-series charts showing:

- Hourly or daily truck volumes,
- Speeds and congestion trends,
- Freight activity during peak vs. off-peak periods.

These visuals help planners identify behavioral patterns and performance bottlenecks.

iv. Attribute-Based Exploration

The dashboard provides drop-down menus, sliders, and selection tools to filter data by:

- Truck class or weight,
- Geographic zone,
- Time of day or day of week.
- Highway
- City or Location
- Direction

3.3 Performance and Scalability

To achieve near-instantaneous interaction even on large datasets, the platform incorporates several performance enhancements:

i. GPU Acceleration

By leveraging cuDF and Dask, data can be processed in-memory and in parallel across GPU cores, significantly reducing the latency of aggregations, joins, and filtering.

ii. In-Memory Filtering

Cross-filters are powered by in-memory data structures and caching, avoiding the need for repeated database queries. This allows user interactions to reflect instantly across all visual elements.

iii. Scalable Architecture

The platform can scale in multiple directions:

- Vertically, by increasing GPU resources,
- Horizontally, by distributing workload using Dask clusters,
- Functionally, by incorporating new data sources or visualization layers without major architectural changes.

iv. Benchmark Results

Performance testing showed that the system supports real-time visual updates in under 100 milliseconds for datasets exceeding 100 million rows—demonstrating readiness for operational-scale deployments.

3.4 User Interface and Experience

Designed with end-users in mind, the platform emphasizes simplicity, responsiveness, and interactivity.

i. Dashboard Layout

A clean, web-based dashboard presents:

- Central map interface,
- Adjacent time series charts and filtering controls,
- Status indicators for active filters and data subsets.

ii. Interaction Controls

Users can:

- Select time windows with sliders,
- Choose truck types via dropdowns,
- Zoom and pan to inspect specific corridors.

All interactions are intuitive, minimizing training needs and improving accessibility.

iii. Diversity and Browser Compatibility

The platform runs smoothly on:

- Desktop and laptop browsers,
- Tablets and large-format displays,
- Without requiring plugins or proprietary tools.

iv. User-Centered Design Process

Feedback from transportation planners was incorporated during development, ensuring that the system addresses real-world freight planning use cases, and supports both technical and non-technical users.

CHAPTER 4 SPATIAL TEMPORAL ANALYSIS

Effective freight transportation planning requires a deep understanding of not only where freight moves, but also when—and under what conditions—these movements occur. To support this, our platform integrates spatial and temporal dimensions of connected vehicle data through an interactive web-based map dashboard that enables real-time, multidimensional analysis of freight patterns across Florida's highway network.

The spatiotemporal analysis framework combines high-resolution GPS data from freight trucks with geospatial infrastructure layers and temporal metadata to uncover key performance patterns, bottlenecks, and behavioral insights.

4.1 FRAMEWORK AND METHODOLOGY

Data from thousands of connected freight vehicles—processed and conflated with the Florida road network—was aggregated both spatially (by route segment, corridor, or zone) and temporally (by hour, day, or custom time windows). The system supports:

- Dynamic map-based exploration of freight movement
- Real-time speed and volume visualization
- Interactive filtering by geography and time

The frontend map is tightly integrated with the GPU-accelerated backend, enabling users to apply cross-filters (e.g., date range, truck class, location) and receive instantaneous updates across all spatial and temporal visualizations.

4.2 SPATIAL ANALYSIS OF FREIGHT SPEEDS

Spatial speed distribution was calculated using average truck velocities on geohash-matched road segments across Florida. The interactive map visualizes:

- Freight truck speed by segment
- Traffic density and route utilization
- Hotspots of congestion or underperformance

Figure 5 illustrates the spatial distribution of average truck speeds on Florida highways. Segments are color-coded from red (slowest) to green (fastest), allowing users to immediately spot urban congestion zones or rural free-flowing corridors.

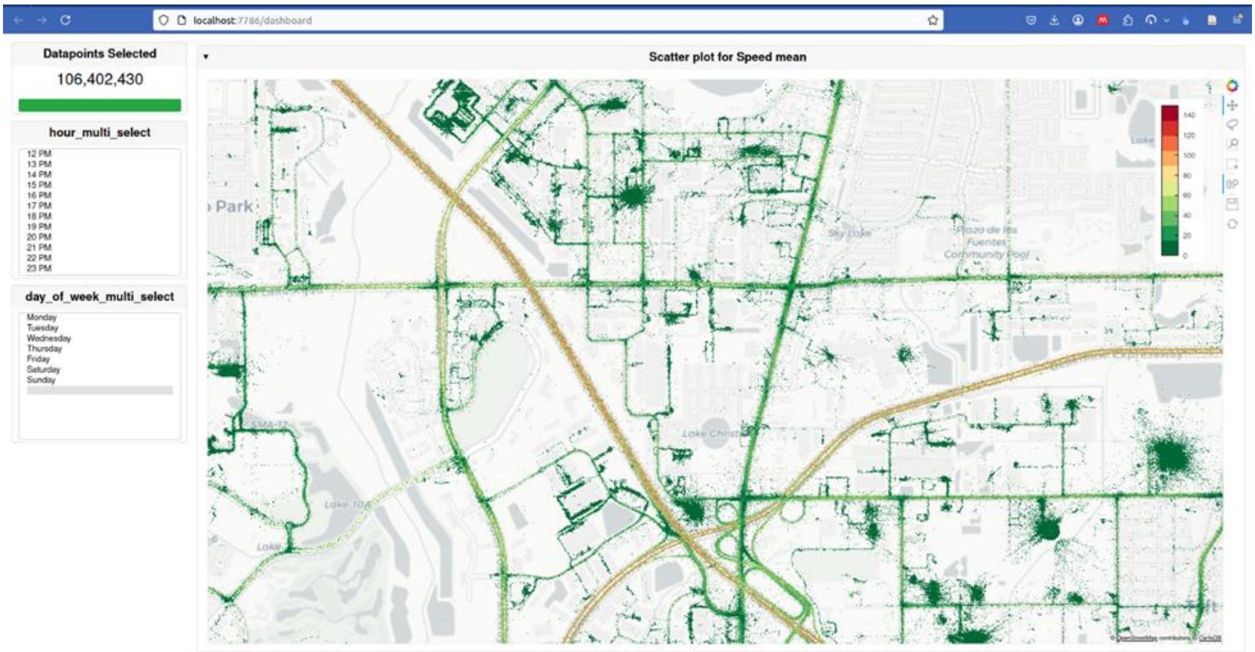

Figure 5. Spatial analysis of freight speeds on Florida highways

4.3 Temporal Trends and Peak Analysis

Temporal variations in speed and volume were analyzed using time-series aggregations. Data was grouped by:

- Hour of day
- Day of week
- User-defined time intervals

Figure 6 displays a high-resolution heatmap showing hourly changes in average freight truck speeds. Morning and evening peak periods show clear slowdowns, particularly on major intercity corridors. Off-peak hours (especially late nights) show improved speed profiles, consistent with line-haul logistics behavior.

Temporal features include:

- Comparison between weekday and weekend performance
- Identification of recurring rush-hour patterns

4.4 Interactive Spatiotemporal Visualization

A core innovation of the system is the interactive spatiotemporal map dashboard, which enables users to simultaneously analyze:

- Where freight is moving (spatial)
- When it's moving and how conditions change (temporal)

Key capabilities include:

- Time slider controls to animate or drill into specific windows
- Map brushing to focus on particular zones or corridors
- Cross-filtering that links map selections with time-series plots and charts

This tightly coupled interface empowers users to pose and answer complex questions—such as:

- "How does freight speed in Miami-Dade change between 6 AM and 9 AM on weekdays?"
- "Which interstate corridors show consistent congestion over multiple weeks?"
- "How do freight volumes change near port facilities at different times of day?"

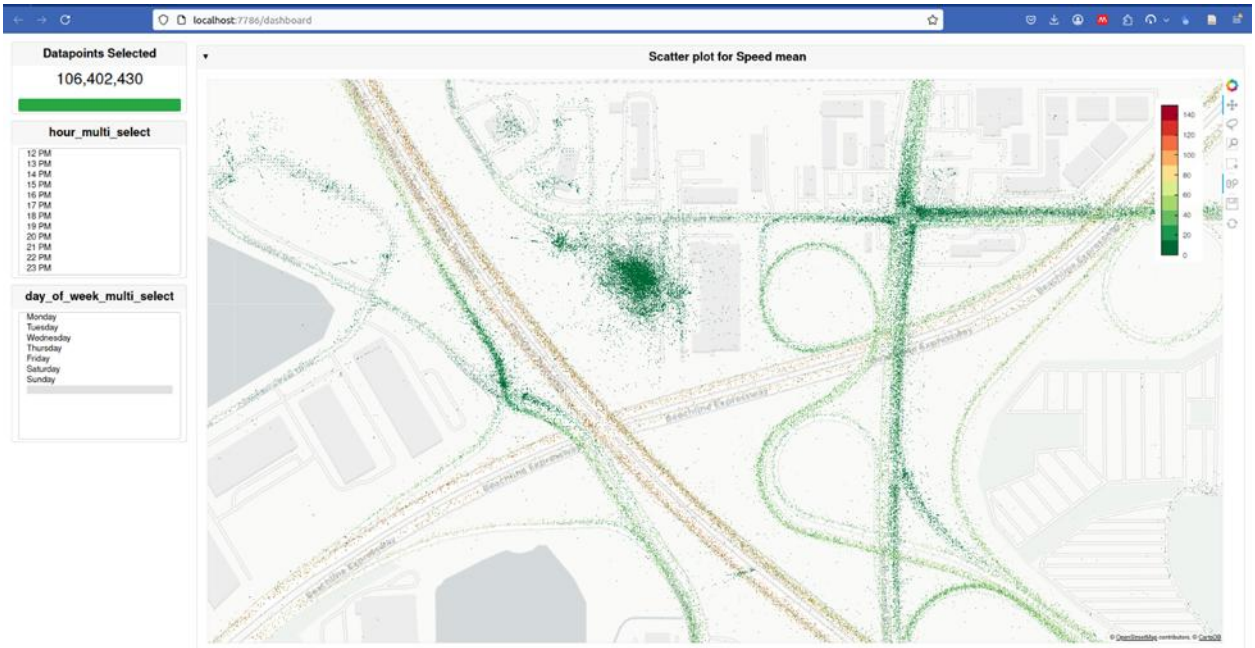

Figure 6. High-resolution spatial analysis of freight speeds on Florida highways

The extraction of freight activity patterns from GPS data presents a complex challenge due to the heterogeneity of vehicle behaviors, variability in operational schedules, and the need to distinguish between different types of freight activities such as pickups, deliveries, loading, and idling. This process is further complicated by noisy or incomplete GPS signals, the lack of contextual data (e.g., shipment logs or facility information), and the difficulty of associating stops with specific logistical functions without additional semantic information. Thus, leveraging GPS data sources to classify freight activity has been explored in the literature, with authors adopting heuristic methods and machine learning algorithms for the task (Yang et al., 2014; Akter et al., 2023). Within this preliminary study in the extraction of truck activity data, Akter et al.'s method of freight activity classification is adopted and modified for the data presented.

A two-pronged approach is adopted in the classification of truck activity. Firstly, the data is cleaned and trip and stop features extracted from the data. Using extracted features, clustering is performed to determine the unique activity patterns.

5.1 Data Cleaning and Feature Extraction

Data collected is filtered based on individual trucks and within each truck, all trip information is extracted. For individual trips, filtering is further done to eliminate incomplete trips which are categorized based on a minimum travel distance of 1.2 miles and travel time of less than 20 minutes. Following this initial filtering, information on the number of stops, their durations and the time of day in which they occurred are obtained.

Within this preliminary study, stops are defined as GPS cluster pings where the speed is less than 5mph and distance traveled is < 0.2miles. Furthermore, the time difference between the commencement of the stop and its finality are defined to be greater than or equal to 5minutes. This threshold in stop duration time aims to eliminate traffic signal related stops. Five stop features for each trip ID are obtained i.e.,

- Number of stops < 30 minutes
- Number of stops between 30 minutes to 8hrs
- Number of stops > 8hrs
- Duration of stops within the daytime (6AM to 6PM)
- Duration of stops within the nighttime (12AM – 6AM & 6PM – 12AM)

Stop-related features reflect behavioral characteristics that are used to distinguish typical activity patterns. For example, stops lasting less than 30 minutes generally indicate short breaks, such as food stops, restroom use, refueling, or brief deliveries. In contrast, stops lasting between 30 minutes and 8 hours are typically associated with longer pickup or delivery activities, but do not represent extended rest periods (Jing, 2018).

Beyond recording of stop related features, trip related features are also recorded. Features recorded include:

- Number of trips < 1hr
- Number of trips between 1-4hrs
- Number of trips > 4hrs
- Number of trips < 30miles
- Number of trips between 30-100 miles and
- Number of trips > 100miles.

This information is critical as trips with lengths less than 30 miles and/or durations under 1 hour are considered indicative of short-haul movements, whereas trips exceeding 100 miles in length and/or lasting more than 4 hours are classified as long-haul truck movements

5.2 DATA CLUSTERING

Following the cleaning of data and extraction of features, clustering is performed. Clustering aims to classify truck activities without the need for contextual information. Different clustering algorithms have been developed, however, for the task at hand K-means clustering is adopted. A total of 247 individual trucks with 11,321 total trips are utilized to demonstrate the performance.

Prior to clustering, determining the optimal number of centroids is critical to the model's performance. The elbow method is used as it helps identify the point at which adding more clusters doesn't significantly improve the model — like the "elbow" of a bent arm. Therefore, given the 11 features derived from 247 number of trucks, the elbow method resulted in 4 number of cluster centers as shown in the Figure 7 below.

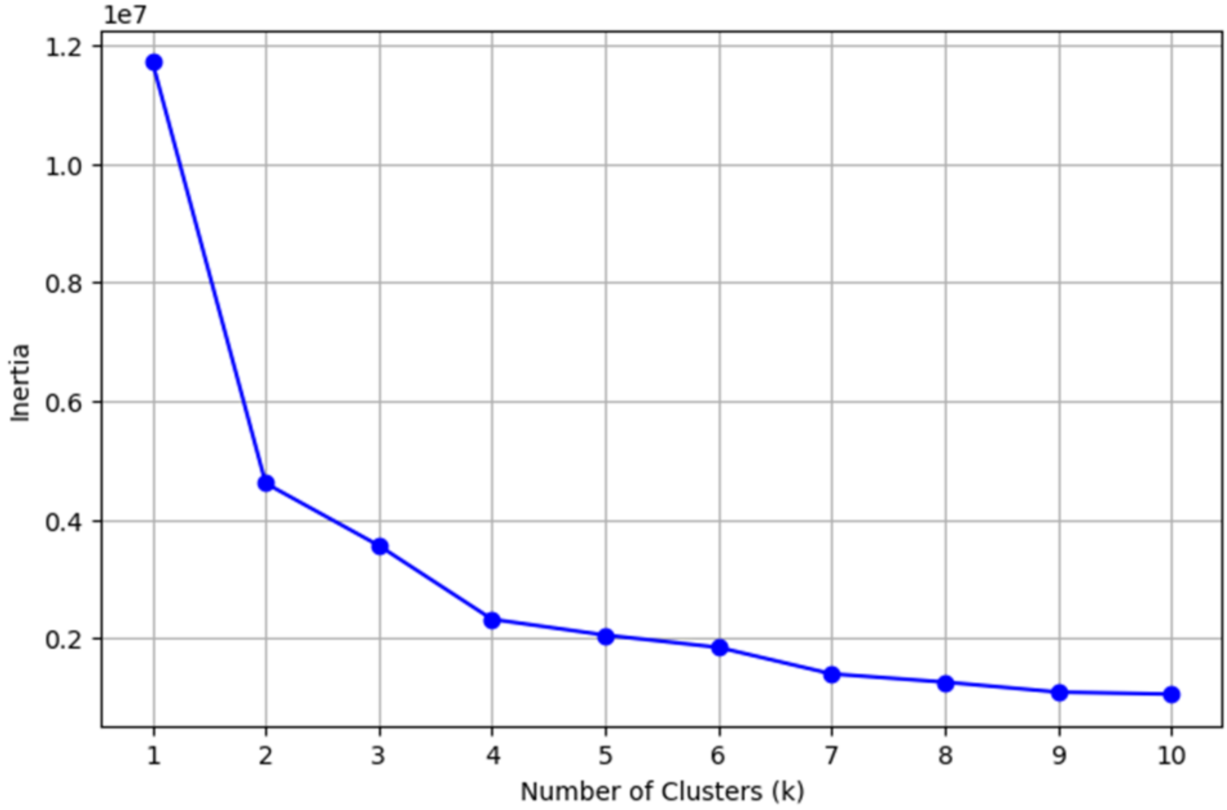

Figure 7. Elbow Method for Optimal Clusters

Utilizing a total of 4 clusters, representing 4 freight activity patterns, the clusters below were obtained (Figure 8).

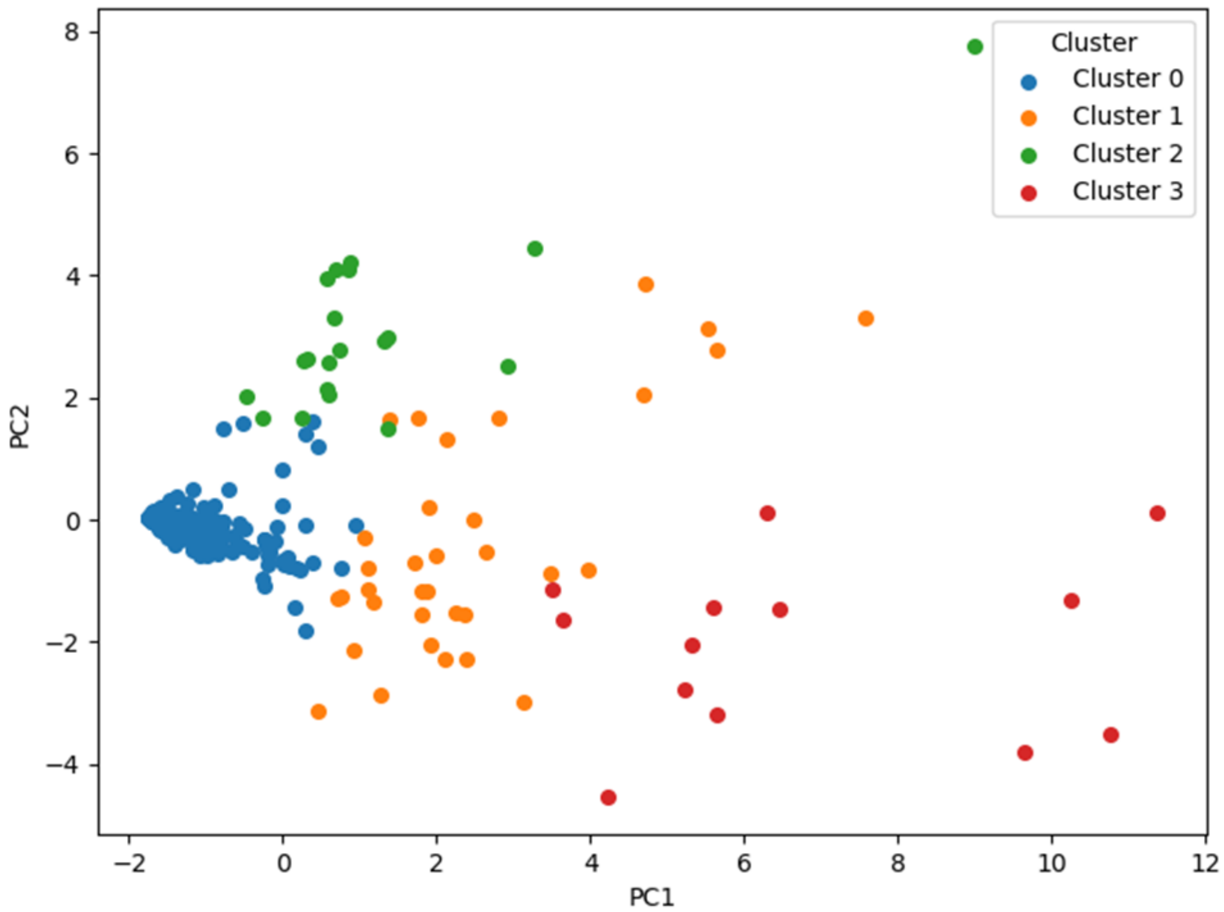

Figure 8. Trip Clusters

Upon analysis of the centroids of individual clusters the information below was obtained. From the table, the four activity patterns can better be summarized as:

Cluster 0: Low-Activity, Short-to-Medium Range, Daytime Local Delivery

- Low trip frequency, short-to-medium range (<100 miles), short stops, daytime-focused.
- Example: Local delivery services, small-scale logistics.

Cluster 1: Moderate-Activity, Mixed-Range, Daytime Regional Distribution

- Moderate trip frequency, mix of short and medium trips with some long hauls, frequent short stops, primarily daytime with some nighttime.
- Example: Regional distribution centers, logistics hubs.

Cluster 2: High-Activity, Short-Range, Daytime Urban Delivery

- High trip frequency, short-range (<30 miles), very short trips and stops, daytime focused.
- Example: Last-mile delivery, urban freight services.

Cluster 3: High-Activity, Long-Range, Mixed-Time Long-Haul Transport

- High trip frequency, mix of all trip distances with emphasis on medium and long hauls, frequent short stops, significant daytime and nighttime activity.
- Example: Long-haul freight, inter-regional transport.

Table 1. Cluster Centroid Features

	Trips						Stops				
Cluster	<30 mi	30-100 mi	>100 mi	< 1hr	1-4hrs	>4hrs	<30 min	30 min-8hr	>8hr	Day	Night
0	Low (3.18)	Low (2.44)	Low (2.87)	Low (2.93)	Low (4.59)	Low (0.96)	Low (2.98)	Low (0.43)	Low (0.00)	Low (44.25)	Low (14.44)
1	Medium (25.35)	Medium (10.85)	Medium (9.44)	Medium (19.00)	Medium (21.65)	Medium (5.00)	Medium (21.24)	High (3.65)	High (0.06)	Medium (359.83)	Medium (160.09)
2	High (79.70)	Medium (11.40)	Low (0.20)	High (62.55)	Medium (28.65)	Low (0.10)	Medium (9.00)	Medium (0.55)	Low (0.00)	Low (79.62)	Low (32.97)
3	Medium (31.85)	High (40.77)	High (27.08)	Medium (23.54)	High (64.31)	High (11.85)	High (49.54)	High (4.15)	Low (0.00)	High (458.31)	High (305.27)

CHAPTER 6 SUMMARY AND CONCLUSION

Connected vehicle data is revolutionizing the trucking industry by enabling data-driven decision-making, improving efficiency, and enhancing safety. While challenges such as data privacy, integration, and regulatory hurdles are present, the benefits of cost savings, customer satisfaction, and sustainability make it a critical tool for modern fleets. This project therefore presents a first step towards increased utility of connected vehicle data through the introduction of an interactive web visualization platform and preliminary analysis of truck GPS data.

By combining GPU-accelerated processing with dynamic, browser-based analytics, the interactive web visualization platform delivers real-time exploration of integrated freight data—including connected vehicle records, WIM, facility, and traffic flow datasets. Its capabilities support planners and decision-makers in:

- Understanding freight trends,
- Identifying hotspots and inefficiencies,
- Making data-driven infrastructure and policy decisions.

As part of the TITAN system, this platform demonstrates that open-source, scalable technologies can power the next generation of transportation analytics.

Similarly, the preliminary study presented effectively demonstrates how GPS data can be leveraged to infer truck freight activity patterns using unsupervised learning. By extracting trip- and stop-level features and clustering them, the approach identifies four distinct operational behaviors among freight vehicles. The framework provides a scalable methodology for freight pattern recognition in the absence of semantic data, laying the foundation for deeper insights into urban logistics, regional distribution, and long-haul transport operations.

REFERENCES

Abdelkader, G., Elgazzar, K., & Khamis, A. (2021). Connected vehicles: Technology review, state of the art, challenges and opportunities. *Sensors*, 21(22), 7712. https://doi.org/10.3390/s21227712

Akter, T., & Hernandez, S. (2023). Representative truck activity patterns from anonymous mobile sensor data. International Journal of Transportation Science and Technology, 12(2), 492-504.

S. K. Datta, J. Haerri, C. Bonnet and R. Ferreira Da Costa, "Vehicles as Connected Resources: Opportunities and Challenges for the Future," in IEEE Vehicular Technology Magazine, vol. 12, no. 2, pp. 26-35, June 2017, doi: 10.1109/MVT.2017.2670859

Federal Motor Carrier Safety Administration. (2023). Electronic Logging Devices. U.S. Department of Transportation. https://www.fmcsa.dot.gov/hours-service/elds/electronic-logging-devices

National Highway Traffic Safety Administration. (2022). *Cybersecurity best practices for the safety of modern vehicles*. U.S. Department of Transportation. https://www.nhtsa.gov/sites/nhtsa.gov/files/2022-09/cybersecurity-best-practices-safety-modern-vehicles-2022-pre-final-tag_0_0.pdf

R. Salazar-Cabrera, Á. P. De La Cruz and J. M. Madrid Molina, "Fleet Management and Control System from Intelligent Transportation Systems perspective," 2019 2nd Latin American Conference on Intelligent Transportation Systems (ITS LATAM), Bogota, Colombia, 2019, pp. 1-7, doi: 10.1109/ITSLATAM.2019.8721347.

U.S. Department of Transportation. (2019). Telematics technology provides many benefits to trucking fleets. ITS Knowledge Resources. https://www.itskrs.its.dot.gov/2019-b01394

Yang, X., Sun, Z., Ban, X. J., & Holguín-Veras, J. (2014). Urban Freight Delivery Stop Identification with GPS Data. Transportation Research Record, 2411(1), 55-61. https://doi.org/10.3141/2411-07 (Original work published 2014)