**From Chaos to Control: The Power of Data in Freight Logistics Optimization**

**1.0 Introduction**

A diagram of a pyramid

AI-generated content may be incorrect.

**Figure 1. Framework for Freight Logistics Optimization: Highlighting key focus areas, including freight movement analysis, operational efficiency, environmental impact assessment, safety and risk evaluation, and economic resilience analysis, to drive sustainable and resilient supply chain systems.**

**1.1 Background Information**

Freight transportation is a cornerstone of global trade that facilitates the movement of goods across supply chains and ensuring economic stability. However, the sector faces numerous challenges. These challenges include operational inefficiencies, rising costs, safety concerns, and environmental impacts. Addressing these challenges requires a data-driven approach to optimize logistics, reduce risks, and promote sustainability.

This project utilizes data from the Bureau of Transportation Statistics (BTS) and follows the CRISP-DM framework to systematically explore freight logistics data. It aims to generate actionable insights that enhance efficiency and resilience in freight operations based on analysed patterns in freight movement, assessed environmental impacts, and evaluated economic disruptions.

Recent global events, such as the COVID-19 pandemic and geopolitical tensions, have demonstrated the vulnerability of freight systems. Studies indicate that the pandemic significantly disrupted freight transportation, reinforcing the need for adaptive and robust logistics networks (Schofer et al., 2022). Additionally, geopolitical factors have caused disruptions in global supply chains, underscoring the necessity of proactive freight management strategies (Lawrence et al., 2024; Zheng, 2024).

The primary objective of this project is to leverage advanced analytics to enhance freight transportation systems by identifying inefficiencies, assessing risk factors, and recommending sustainability strategies. These insights will be particularly valuable for logistics managers, policymakers, and business analysts seeking to navigate an increasingly complex economic and regulatory landscape.

**1.2 Key Factors in Freight Transportation Analysis**

The analysis focuses on key categories influencing freight movement efficiency, cost, and sustainability:

* **Freight Movement Patterns:** Analyzing shipment trends across transportation modes and trade types.
* **Operational Efficiency:** Identifying logistical bottlenecks and cost inefficiencies.
* **Environmental Impact:** Evaluating carbon emissions and sustainability measures.
* **Safety and Risk Factors:** Assessing hazards and developing risk mitigation strategies.
* **Economic Disruptions:** Measuring the impact of global events on freight transport.

To effectively analyse these key categories, the following features are identified as crucial to the study:

**1. Shipment Identification & Trade Information**

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **YEAR** | **Year** of the shipment (four-digit AD format). |
| **MONTH** | **Month** of the shipment (1 - 12). |
| **TRDTYPE** | Trade type, indicating whether the shipment is an **export** or **import**. |
| **COMMODITY2** | **2-digit commodity code** categorizing the type of goods being transported. |

**2. Shipment Origin & Destination Details**

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **USASTATE** | U.S. state code where the freight originates or arrives. |
| **MEXSTATE** | Mexican state code, applicable when the shipment involves Mexico. |
| **CANPROV** | Canadian province code, applicable when the shipment involves Canada. |
| **COUNTRY** | Country code indicating the international origin or destination of the shipment (Canada: **1220**, Mexico: **2010**). |
| **DEPE** | Port or district code representing the shipment's processing location. |

**3. Transportation & Logistics Information**

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **DISAGMOT** | Mode of transportation code specifying how the freight is transported (**Vessel, Air, Mail, Truck, Pipeline, Other, Foreign Trade Zones (FTZs)**). |
| **CONTCODE** | Indicates whether the shipment is **containerized (X)** or **non-containerized (0)**. |

**4. Economic & Cost Factors**

|  |  |
| --- | --- |
| **Feature** | **Description** |
| **VALUE** | Total **value of goods** being shipped, measured in U.S. dollars (USD). |
| **SHIPWT** | Total **shipping weight** of the goods in kilograms (Kg). |
| **FREIGHT\_CHARGES** | **Freight cost** associated with transporting the shipment (in USD). |
| **DF** | Indicates whether the **merchandise was produced in the U.S. (1 = Domestic) or outside the U.S. (2 = Foreign)**. |

By categorizing these expected features, the analysis remains focused on the key aspects of freight transportation:

* **Time-Based Features** (e.g., YEAR, MONTH) allow analysts to study shipment trends over time and assess seasonal fluctuations.
* **Origin & Destination Features** (e.g., USASTATE, COUNTRY) enable tracking of freight movement and identifying regional trade patterns.
* **Logistics Features** (e.g., DISAGMOT, CONTCODE) provide insights into transportation methods, efficiency, and infrastructure utilization.
* **Economic & Cost Features** (e.g., VALUE, SHIPWT, FREIGHT\_CHARGES) help in assessing financial impacts, cost optimization, and economic disruptions.

These structured categories ensure the analysis remains comprehensive, data-driven, and aligned with the core objectives of optimizing freight transportation efficiency, reducing costs, and mitigating risks.

**1.3 Hypothesis**

To establish a structured analytical approach, the following hypotheses are formulated:

* **Null Hypothesis (H₀)**: Transportation mode selection, risk management strategies, and sustainability measures have no statistically significant effect on freight transportation efficiency, environmental impact, or resilience.
* **Alternative Hypothesis (H₁)**: Optimized transportation mode selection, risk management strategies, and sustainability measures significantly improve freight transportation efficiency, reduce environmental impact, and enhance resilience.

By testing these hypotheses, the study aims to quantify the impact of strategic interventions in freight logistics.

**1.4 Key Business Questions**

The analysis is structured around critical business questions that guide insights and decision-making:

1. **What are the key patterns and trends in freight movement across different transportation modes and trade types?**

**Purpose**: To identify inefficiencies and areas for optimization in logistics and supply chain management.

1. **How do different transportation modes impact freight costs and operational efficiency?**

**Purpose**: To determine the most cost-effective and efficient transportation methods for different types of goods.

1. **What are the primary contributors to carbon emissions in freight transportation, and how can they be reduced?**

**Purpose**: To assess environmental impact and explore sustainable alternatives.

1. **What are the most common risks and safety concerns in freight transport, and how can they be mitigated?**

**Purpose**: To enhance freight safety by implementing proactive risk management strategies.

1. **How have economic disruptions (e.g., COVID-19, geopolitical tensions) affected freight movement, and what strategies can improve resilience?**

**Purpose**: To develop adaptive strategies that mitigate the impact of economic fluctuations on freight operations.

1. **How do domestic and foreign freight patterns compare, and what implications do they have for trade policies?**

**Purpose**: To analyse the effects of domestic vs. international freight trends on economic resilience and strategic decision-making.

By structuring the analysis around these questions, the project builds a logical progression toward validating or rejecting the null hypothesis.

**2.0 Methodology**

A structured approach is adopted to ensure that each step aligns with the project’s goal of improving freight transportation through data-driven insights. The methodology involves data exploration, feature analysis, and the development of interactive dashboards using Power BI for effective data visualization and stakeholder engagement.

**2.1 Exploratory Data Analysis (EDA) and Data Cleaning**

**2.1.1 Data Quality Assessment & Exploration**

* **Assess Data Structure**: Used methods like .info(), .head(), and .describe() to explore the dataset characteristics.
* **Check Duplicates**: Removed duplicate records to ensure the dataset integrity.
* **Validate Data Consistency**: Ensured values align with expected ranges (e.g., shipment months, trade types).
* **Identify Missing Values**: Verified and addressed missing data points.

**2.1.2 Features Analysis**

* **Univariate Analysis:** Visualized individual feature distributions using histograms, bar charts, and box plots.
* **Bivariate & Multivariate Analysis:** Examined relationships between features, such as trade type vs. freight charges.
* **Feature Engineering:** Converted encoded variables to readable text, normalized numerical values, and created new features like shipment seasonality indicators and economic impact scores.

**2.2 Interactive Dashboard Development using Power BI**

* **Data Integration:** Imported and structured data within Power BI for efficient visualization.
* **Dashboard Design:** Created interactive charts, maps, and reports to highlight trends and key insights.
* **Stakeholder Engagement:** Ensured visualizations are tailored to logistics managers, policymakers, and analysts for informed decision-making.

**2.3 Protocol for Replication**



**Figure 2. Illustration of the data refinement process, showcasing key steps from data quality assessment and duplicate removal to feature analysis and interactive dashboard development for delivering actionable insights.**

**3.0 Results**

**3.1 Univariate Analysis**

**3.1.1 Categorical Features**

**1. Exports Dominate Trade Activity**

* Exports account for 66.48% of total trade, nearly double the volume of imports (33.52%).
* With over 4 million export transactions compared to 2 million imports, this suggests a strong export-driven economy.
* Possible implications include a trade surplus, reliance on exports, or the need for policy adjustments to stimulate import growth.

**2. Geographic Distribution: U.S., Mexico, and Canada**

* **United States (State-Level Trade)**
  + Texas (5.40%), California (4.56%), and Illinois (3.68%) lead in trade volume, benefiting from major ports and logistics hubs.
  + Mid-tier states like Ohio, Florida, and Pennsylvania (2.86% - 2.96%) contribute moderately.
  + Smaller states (e.g., Vermont, South Dakota, West Virginia < 0.67%) and Alaska, Hawaii, D.C. have minimal shares, likely due to geographic and industrial constraints.
  + **Implications:** States with high trade volumes have strategic advantages, while smaller states face logistical and industrial limitations.
* **Mexico (State-Level Trade)**
  + "State Unknown" (3.47%) leads, suggesting unclassified or missing data.
  + **Top contributors:** Estado de México (2.79%), Nuevo León (2.62%), and Chihuahua (1.76%), likely driven by border trade.
  + **Lower contributors:** Zacatecas, Quintana Roo, Oaxaca (<0.21%), possibly due to geographic and economic factors.
* **Canada (Province-Level Trade)**
  + Ontario (14.70%) is the dominant trade hub, followed by Quebec (9.59%), British Columbia (8.22%), and Alberta (6.95%).
  + Smaller provinces and territories (e.g., Northwest Territories, Nunavut) contribute less than 1%.
  + **Implications:** Trade is highly concentrated in major economic regions, highlighting the need for investment in underdeveloped areas.
* **Overall Trade Between the U.S., Canada, and Mexico**
  + Canada leads with 60.39% of total trade, driven by proximity, strong economic ties under the United States-Mexico-Canada Agreement (USMCA), and industries like oil & gas, automotive, and industrial goods (Villarreal, 2020).
  + Mexico accounts for 39.61%, playing a crucial role in manufacturing, automotive supply chains, and agricultural exports, supported by Maquiladora programs and cost-effective labour (Villarreal, 2020).

**3. Trade by Mode of Transport**

* Trucks dominate (65.53%), reflecting a strong reliance on road transport for flexibility and cost-effectiveness.
* Air (17.19%) is used for high-value, time-sensitive goods, while rail (10.14%) handles bulk shipments.
* **Other modes:** Vessels (2.95%), pipelines (0.20%), and mail (0.12%) play smaller roles.
* **Implications:** There is a need for road infrastructure investments, rail and air transport optimization, and sustainability measures to reduce fuel dependency.

**4. Trade by Cargo Type: Containerized vs. Non-Containerized**

* Non-containerized trade leads (55.59%), indicating a reliance on bulk transport for raw materials, heavy equipment, coal, and grain.
* Containerized trade (36.61%) plays a key role in consumer goods, electronics, and manufactured products.
* Unknown category (7.80%) may indicate data inconsistencies or mixed cargo classifications.
* **Implications:** 
  + Bulk transport optimization is essential for large-scale industries.
  + Container trade expansion aligns with global supply chain trends.
  + Improved data classification can enhance logistics planning and efficiency**.**

**5. Trade Distribution by DF (Domestic vs. Foreign Merchandise)**

* **Domestic Merchandise Leads (44.40%):** A strong local production sector suggests self-reliance, industrial growth, and potential government support for local manufacturing.
* **Foreign Merchandise Holds 22.08%:** Imports remain significant, indicating demand for foreign goods, raw materials, or specialized products.
* **Unaccounted 33.52%:** The missing percentage may be due to re-exported goods, re-imported goods, or incomplete data filtering.
* **Implications:**
  + Local industries benefit from reduced import dependency.
  + Foreign trade plays a key role in supplementing consumer demand.
  + Policy adjustments (tariffs, trade agreements) could influence the balance between domestic and foreign goods.

**6️. Trade Distribution by Month**

* **Peak Trade Months:** August (9.37%), June (9.33%), September (9.30%), March (9.29%), and July (9.28%) experience the highest trade volumes, suggesting stable mid-year activity.
* **Moderate Trade Months:** May (9.25%), April (9.23%), February (9.11%), and January (9.03%) show slightly lower, but still substantial, trade activity.
* **Lowest Trade Months:** October (5.70%), November (5.62%), and December (5.50%) exhibit the lowest trade volumes, likely due to seasonal factors, year-end holidays, and supply chain slowdowns.
* **Implications:**
  + Businesses and policymakers can optimize trade strategies by anticipating peak and slow periods.
  + Supply chain operations may need adjustments to mitigate end-of-year slowdowns.
  + Further analysis can explore industry-specific trends affecting trade seasonality.

**7️. Trade Distribution by Year**

* **Peak Trade Years (2021–2023):** 2023 (22.76%), 2022 (22.50%), and 2021 (22.06%) recorded the highest trade volumes, contributing to about 67% of total trade.
* **Declining Trade in 2024 & 2020:** 2024 (17.01%) shows a decline, while 2020 had the lowest trade volume (15.58%), likely due to COVID-19 disruptions.
* Implications:
  + Stable trade between 2021–2023 suggests a strong economic recovery.
  + 2024’s decline may indicate policy shifts, economic slowdowns, or incomplete data.
  + Further analysis is needed to determine macroeconomic impacts on trade performance.

**3.5 Analytical Questions**

**3.5.1 What are the key patterns and trends in freight movement across different transportation modes and trade types?**

**A graph of a number of different colored bars

AI-generated content may be incorrect.**

**Figure 3. Freight Movement Across Transportation Modes and Trade Types. This bar chart illustrates the volume of exports and imports across different transportation modes. Trucking is the dominant mode, handling the largest freight volume (2.97M exports, 1.30M imports). Air freight exhibits a strong export bias (0.71M exports vs. 0.41M imports), while rail transport remains balanced (0.35M exports, 0.31M imports). Maritime (vessel) and intermodal ("Other") transport play smaller roles, while pipeline and mail contribute minimally. Foreign Trade Zones (FTZs) primarily facilitate imports.**

**1️. Trucking Dominates Freight Transport**

* Trucks handle most trade, with 2.97 million exports and 1.30 million imports.
* This highlights a strong reliance on road transport, likely due to cost efficiency, accessibility, and flexibility for short-to-medium distances.

**2️. Air Freight Favors Exports**

* Exports (0.71 million) exceed imports (0.41 million), with zero recorded imports, possibly due to data limitations or trade patterns.
* Air freight is likely preferred for high-value, time-sensitive goods like electronics and perishables.

**3️. Rail Supports Balanced Trade**

* Exports (0.35 million) and imports (0.31 million) are nearly equal, suggesting rail is viable for both directions.
* Commonly used for bulk goods like raw materials.

**4️. Vessel & Other Modes Play Niche Roles**

* Vessel trade remains low (exports: 0.07M, imports: 0.11M), likely due to trade route constraints.
* Other mode is minimal, with 0.20M exports and 0.02M imports.

**5️. Pipeline & Mail Have Minimal Impact**

* Pipeline trade is almost negligible (0.01M for both exports & imports), indicating low pipeline infrastructure use.
* Mail freight is rarely used, playing a minor role in large-scale trade.

**6️. Foreign Trade Zones (FTZs) Support Imports**

* FTZs handle more imports (0.02M) than exports (0.00M).
* Used primarily for storage and processing before goods officially enter the market.

**7. Optimisation Insights for Logistics Efficiency**

* **Truck congestion and dependency risks:** Since trucking dominates, optimizing routes, fuel efficiency, and cross-border logistics is crucial.
* **Potential for air freight expansion:** Given the current bias towards exports, investing in air cargo infrastructure could help balance import trade.
* **Rail as an alternative for cost savings**: A more balanced rail trade suggests opportunities for intermodal freight expansion.
* **Maritime trade inefficiencies:** Low vessel usage could mean opportunities exist for bulk transport cost reduction.
* **FTZs for supply chain optimization:** Their primary role in imports suggests strategic use in import tax deferral and warehousing.

**4.0 Discussion**

**4.1 Understanding Trade Patterns and Economic Implications**

Trade activity in North America is largely driven by exports, which account for 66.48% of total trade. This figure is nearly double the volume of imports, suggesting an export-driven economy with a strong production base. Over four million export transactions were recorded compared to two million imports, which could indicate a trade surplus. This imbalance raises important considerations about economic reliance on exports and whether policy adjustments are needed to stimulate import growth.

**4.2 Geographic Distribution: Key Trade Players in North America**

Trade across the United States, Mexico, and Canada follows distinct regional patterns. In the U.S., states with major ports and logistics hubs lead in trade activity. Texas, California, and Illinois rank highest, each contributing significantly due to their strategic locations. Conversely, states like Vermont, South Dakota, and West Virginia account for less than 0.67% of total trade, reflecting their geographical and industrial limitations.

In Mexico, data shows a peculiar trend: "State Unknown" contributes 3.47% to total trade, pointing to unclassified or missing data. Nonetheless, the country’s most active trade regions include Estado de México, Nuevo León, and Chihuahua, which suggests their border proximity and role as manufacturing hubs. On the other hand, states like Zacatecas and Oaxaca contribute less than 0.21%, likely due to their distance from major trade corridors.

Canada's trade landscape is more centralized, with Ontario alone responsible for 14.70% of the country’s trade. Quebec, British Columbia, and Alberta follow closely behind. Smaller provinces, such as the Northwest Territories and Nunavut, contribute less than 1%, emphasizing the concentration of economic activity in major industrial hubs.

When comparing total trade volumes, Canada leads with 60.39%, largely due to strong economic ties under the United States-Mexico-Canada Agreement (USMCA). Key industries driving this trade include oil & gas, automotive, and industrial goods. Meanwhile, Mexico accounts for 39.61% of trade, playing a crucial role in manufacturing and agricultural exports, aided by Maquiladora programs and cost-effective labour (Villarreal, 2020).

**4.3 The Dominance of Trucking in Freight Transportation**

One of the most striking observations in this analysis is the overwhelming reliance on trucking, which handles 65.53% of all trade. This preference is driven by trucking’s flexibility and cost-effectiveness, making it ideal for short-to-medium distance shipments. Air freight, while less commonly used (17.19%), plays a critical role in transporting high-value, time-sensitive goods such as electronics and perishables. Rail transport accounts for 10.14%, supporting bulk goods like raw materials. Other transport modes, including vessels (2.95%) and pipelines (0.20%), play smaller roles.

This heavy dependence on trucking underscores the need for improved road infrastructure and optimized logistics to mitigate congestion and reduce costs. Additionally, opportunities exist to optimize air and rail transport for better trade efficiency. Investing in sustainable transport solutions could also help reduce fuel dependency and environmental impact.

**4.4 Containerized vs. Non-Containerized Trade**

The nature of freight movement also reveals interesting insights. Non-containerized trade dominates with 55.59%, reflecting a high reliance on bulk transport for raw materials, heavy equipment, and agricultural products. Meanwhile, containerized trade comprises 36.61%, focusing on consumer goods and electronics. A notable 7.80% of trade falls under an "Unknown" category, highlighting potential inconsistencies in data classification.

To improve efficiency, optimizing bulk transport logistics is essential, particularly for industries dealing with raw materials. Expanding container trade aligns with global supply chain trends and could streamline cross-border commerce. Additionally, addressing data inconsistencies would enhance logistics planning and decision-making.

**4.5 Domestic vs. Foreign Merchandise: A Balancing Act**

A closer look at trade distribution reveals that domestic merchandise leads at 44.40%, signifying a strong local production sector. Foreign merchandise, on the other hand, accounts for 22.08%, indicating a substantial demand for imported goods. The remaining 33.52% of trade volume remains unaccounted for, possibly due to re-exports, re-imports, or gaps in data tracking.

This distribution suggests that local industries are benefiting from reduced import dependency. However, foreign trade still plays a crucial role in supplying specialized products and raw materials. Adjustments in trade policies, including tariffs and trade agreements, could significantly impact this balance.

**4.6 Trade Fluctuations Across Time**

Trade activity is not constant throughout the year. Peak trade months occur in August, June, September, March, and July, where volumes exceed 9% per month. Conversely, October, November, and December experience the lowest trade volumes, likely due to seasonal slowdowns and holiday-related disruptions.

Understanding these fluctuations is valuable for businesses and policymakers, as it allows for better planning and supply chain optimization. Anticipating peak and slow periods can help mitigate bottlenecks and improve logistics efficiency.

From a yearly perspective, trade volumes peaked between 2021 and 2023, with each year contributing over 22% of total trade. However, 2024 has seen a decline to 17.01%, raising concerns about economic slowdowns, policy shifts, or incomplete data reporting. The lowest trade volume occurred in 2020 (15.58%), largely attributed to disruptions caused by the COVID-19 pandemic.

**4.7 Freight Movement Across Transportation Modes and Trade Types**

Examining trade by transportation mode reveals further insights. Trucks dominate freight transport, with nearly 3 million export transactions and 1.3 million imports. This heavy reliance on trucking suggests an urgent need to optimize routes and improve fuel efficiency to avoid congestion.

Air freight exhibits a strong export bias, with exports (0.71 million) far exceeding imports (0.41 million). This suggests that air transport is primarily used for high-value outbound shipments. Rail trade, in contrast, remains balanced, indicating its viability for both exports and imports, particularly for bulk goods.

While vessel and pipeline transport play minor roles, they could offer cost-saving opportunities if better utilized. Foreign Trade Zones (FTZs) are mostly used for imports, highlighting their strategic role in warehousing and deferred taxation.

**4.8 Optimizing Freight Transportation for the Future**

These findings suggest several key takeaways for improving North America’s freight transportation network:

* Reducing Truck Dependency: Optimizing trucking routes and enhancing cross-border logistics could mitigate congestion and improve efficiency.
* Expanding Air Freight Infrastructure: Given the current export bias, investing in air cargo facilities could help balance import trade.
* Leveraging Rail Transport: A more balanced rail trade presents opportunities for expanding intermodal freight solutions.
* Enhancing Maritime Trade: The relatively low vessel usage signals untapped potential for bulk transport cost reductions.
* Maximizing Foreign Trade Zones: FTZs play a crucial role in import handling and can be strategically expanded for better trade facilitation.

**5.0 Conclusion**

By addressing these logistical challenges and opportunities, North America can enhance its freight transportation efficiency, reduce costs, and bolster its position in global trade. Future research should further explore macroeconomic factors affecting trade performance and assess the long-term impact of emerging trade policies and technological advancements on the industry.

**References**

Lawrence, M., Homer-Dixon, T., Janzwood, S., Rockstöm, J., Renn, O., & Donges, J. F. (2024). Global polycrisis: the causal mechanisms of crisis entanglement. *Global Sustainability*, *7*, e6. https://doi.org/10.1017/SUS.2024.1

Schofer, J. L., Mahmassani, H. S., & Ng, M. T. M. (2022). Resilience of U.S. Rail Intermodal Freight during the Covid-19 Pandemic. *Research in Transportation Business & Management*, *43*, 100791. https://doi.org/10.1016/J.RTBM.2022.100791

Villarreal, M. A. (2020). *U.S.-Mexico Economic Relations: Trends, Issues, and Implications* . https://crsreports.congress.gov/product/pdf/RL/RL32934

Zheng, Z. (2024, December 19). *Enhancing supply chain resilience amid rising global risks*. UN Trade and Development (UNCTAD) Transport and Trade Facilitation Newsletter. https://unctad.org/news/enhancing-supply-chain-resilience-amid-rising-global-risks