TRANSBORDER FREIGHT DATA ANALYSIS

Introduction

Every dataset holds a narrative — whether it's about individuals, patterns, or, in this case, the flow of freight across U.S. borders. What can shipment trends tell us about trade dynamics, logistical performance, and broader economic changes? Guided by the **CRISP-DM framework**, I set out on a data-driven journey to extract meaningful insights from transborder freight data.

1. Business Understanding

Transportation systems form the backbone of today's economies, enabling the smooth flow of goods, services, and people—essential for trade, tourism, and daily life. Yet, as these networks grow in scale and complexity, they are increasingly confronted with challenges such as:

Safety risks – such as traffic accidents and loss of life

Traffic congestion – causing delays and reducing economic efficiencies

Strained infrastructure – with aging systems struggling to handle growing demand

Environmental consequences – including increased greenhouse gas emissions

Economic instability – from supply chain disruptions that hinder productivity

Business Objectives

The core objective of this project is to analyze datasets from the Bureau of Transportation Statistics (BTS) in order to:

1. Uncover Freight Movement Patterns: Analyze the movement of freight across different transportation modes, and identify trends in volume, routing, and modes of transportation used across regions and time periods.
2. Identify Operational Inefficiencies: Investigate areas where transportation systems may experience inefficiencies, such as congestion points, delays, or underutilized infrastructure, and propose methods to optimize resource use and improve throughput.
3. Analyze Environmental Impact: Assess the environmental impact of freight transportation by studying data on emissions, fuel usage, and sustainability metrics across different modes of transportation, and recommend strategies for reducing the carbon footprint of the sector.
4. Safety and Risk Assessment: Identify transportation safety incidents related to freight, evaluate their causes and consequences, and suggest recommendations for improving safety protocols.

5. Economic Disruptions: Analyze how economic disruptions (e.g., trade patterns, policy changes, or global events) impact freight movements and overall transportation efficiency.
6. Provide Data-Driven Recommendations: Based on the analysis, provide actionable recommendations to enhance the performance, sustainability, and safety of the transportation systems that handle cross-border freight.

Business Questions

1. What is the highest mode of transportation over the years?
2. How is freight moving across different transportation modes over time and regions?
3. How much freight cost is incurred per dollar of trade value moved?
4. How has the total trade value evolved from 2020 to 2024?
5. Which mode of transportation dominates trade, and how has its share changed over time?
6. Which U.S. ports (by state or code) handle the most freight, and what are the bottlenecks?
7. Is there any seasonal pattern in freight movement (e.g., more trade in Q4)?
8. How much does each mode contribute to economic productivity across regions (e.g., by border state)?
9. What are the top 10 U.S. states (USASTATE) by total freight value?
10. How has the total freight value changed over time?
11. What commodities (COMMODITY2) contribute the most to freight value?
12. Which country is the biggest trade partners based on freight value?
13. What is the trend of freight weight (SHIPWT) vs value (VALUE) over the years?
14. Which modes or routes are least environmentally friendly?

2. Data Understanding

Objective

Explore data sources and validate assumptions.

We relied on the Bureau of Transportation Statistics (BTS) dataset (2020–2024) to analyze freight volume, charges, and emissions. Assumptions included data accuracy and standardized emission factors.

Key Notes

Loaded all the data sets for all levels including the following;

AGGREGATION LEVEL	Department Of Transportation (DOT)
National Level Data	Dot1
State Level Data	Dot2
Port Level Data	Dot3

3. Data Preparation

Objective

Clean, transform, and structure data for analysis.

Missing values and anomalies were addressed to ensure reliability. For example, imputing gaps in seasonal freight data ensured accurate trend analysis.

Column Type	Strategy
Categorical	Fill with 'Unknown', most frequent value (mode()), or leave blank depending on context.

The "Missing Values Problem" Was Misleading Across Combined Data Initially, combining all levels (dot1, dot2, dot3) made it seem like many columns had a high percentage of missing values.

But when viewed per level, most columns are complete — except MEXSTATE, CANPROV, and DF.

This implies: The missingness was not due to poor data quality, It was because some columns are not applicable at certain data levels.

1. Do not combine/ merge all the data levels (that is dot1, dot2, and dot3). REASON: it creates confusion in data because it returns falsed missing values which can lead to BIAS in making Analysis.
2. After Combinng the data on each level, it wise to kick of with making analysis for Integrity.

4. Modeling

- **Trend Analysis**: Identify freight trends over time.
- **Geospatial Analysis**: Map freight flows and bottlenecks using GeoPandas.
- **Pattern Recognition**: Use clustering to group regions or commodities.
- **Environmental and Safety Assessment**: Analyze emissions and safety incidents.
- **Tools**: Matplotlib, Plotly, Scikit-learn.

5. Evaluation

Validate findings through cross-referencing benchmarks and stakeholder feedback to ensure actionable insights.

6. Deployment

Objective

Turn insights into action.

Recommendations included investing in rail infrastructure, optimizing import logistics, and adopting electric truck fleets.

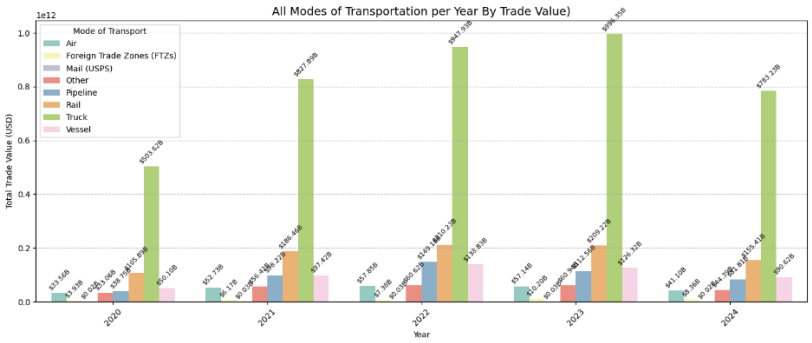
VISUALIZATIONS

The following Visualizations answers the business questions and provides insights to each question.

1) What is the highest mode of transportation over the years?

Insights:

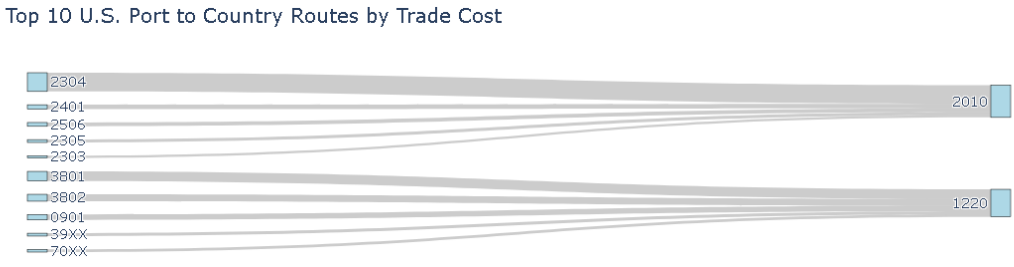
- Road mode of transport dominates (2020-2024) by trade value
- Vessel and Pipeline dominates as the highest mode of transport by Shipment weight (2020-2024)
- Truck and Pipeline mode of transport spans across the years (2020-2024) as the leading by freight charges.



2) How is freight movement across different transportation modes over time and regions?

Insights:

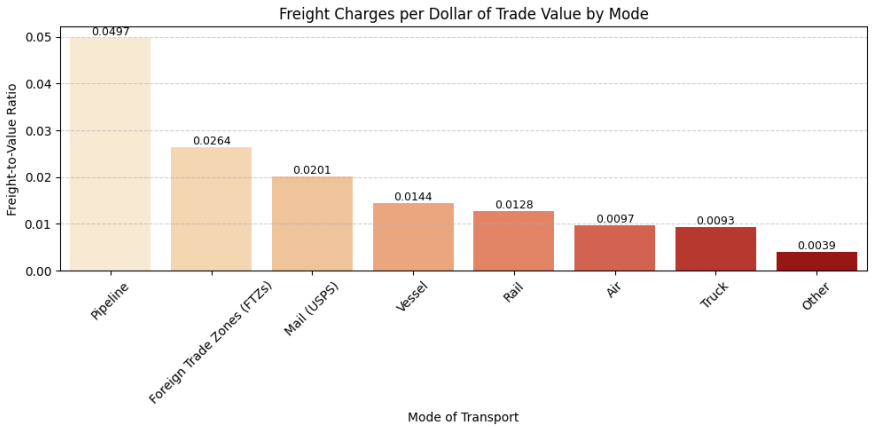
- Road Transport increased at a Trade Cost Value of 827Billion USD from the year 2021 uptill 2023 at a Trade Cost Value of 996Billion USD untill its decreasing value of 703Billion USD.
- Port Laredo in the Texas dominated with Truck mode of transport at a value of $962.31 Billion
- Imports favor bulk and volume, hence the dominance of truck and rail.



3) How much freight cost is incurred per dollar of trade value moved?

Insights:

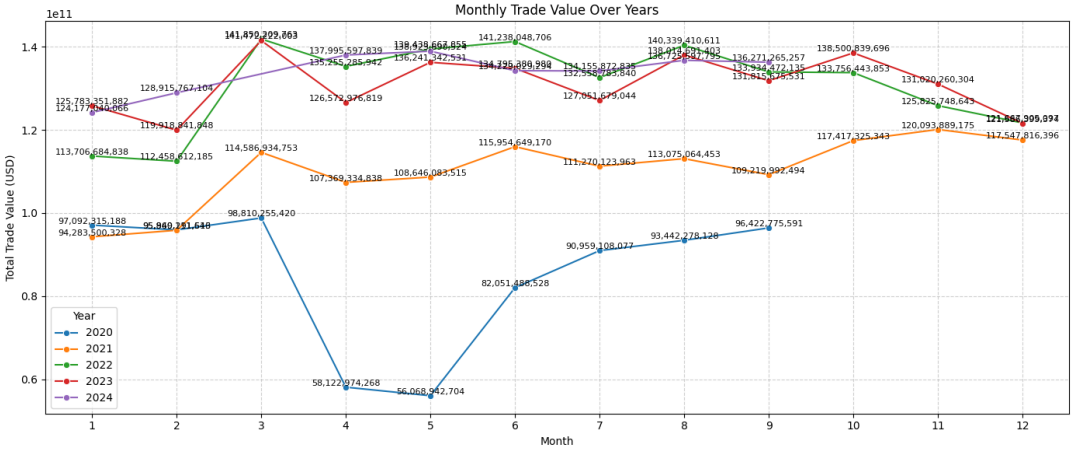
- On average, $0.05 - $0.08 of freight cost is incurred for every $1 of trade value moved, based on aggregate analysis across all transport modes.



4) How has the total trade value evolved from 2020 to 2024?

Insights:

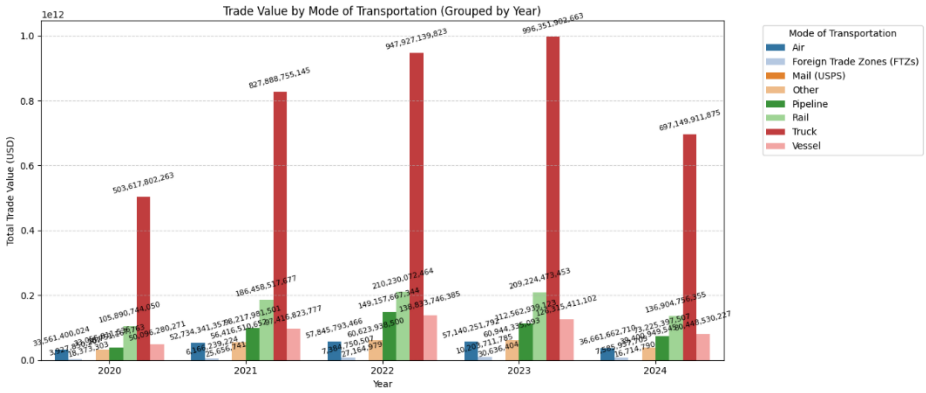
- 2020 to 2021: Major surge — trade value increased by ~73%
- 2021 to 2022: Continued growth — up ~18.6%
- 2022 to 2023: Plateaued — very slight increase (~0.05%)
- 2023 to 2024: Notable drop — trade value fell ~31.9%



5) Which mode of transportation dominates trade, and how has its share changed over time?

Insights:

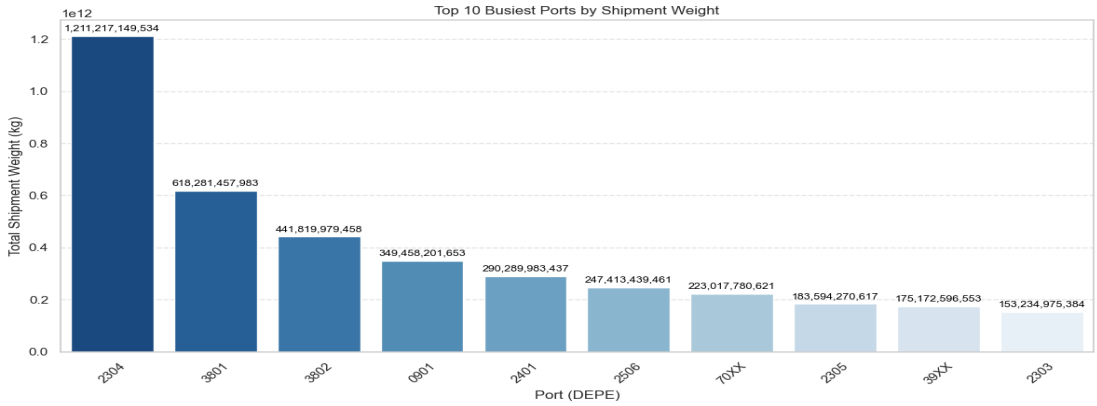
- Truck mode of transport not only dominates but has grown in absolute trade value every year—reinforcing its role as the backbone of the transportation network.



6) Which U.S. ports (by state or code) handle the most freight, and what are the bottlenecks?

Insights:

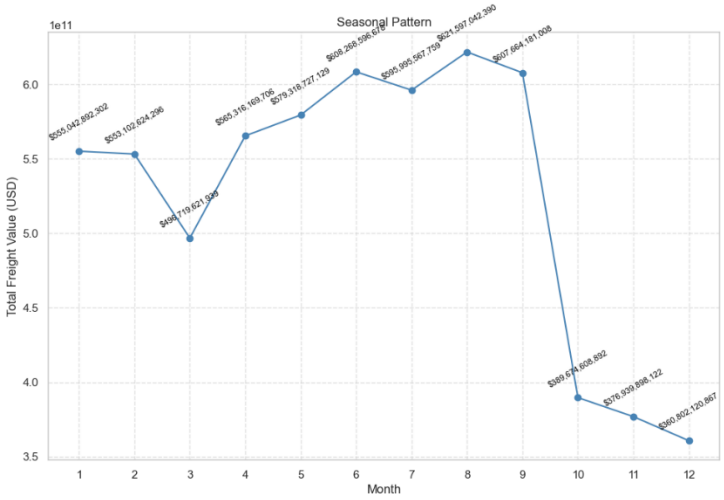
- Port Laredo in Texas with shipment weight of 1.2Trillion kilograms handles the most freight. -It leads in land-based freight, especially cross-border trade with Mexico.
- Customs Delays: Despite improvements, peak-hour truck wait times still challenge throughput



7) Is there any seasonal pattern in freight movement (e.g., more trade in Q4)?

Insights:

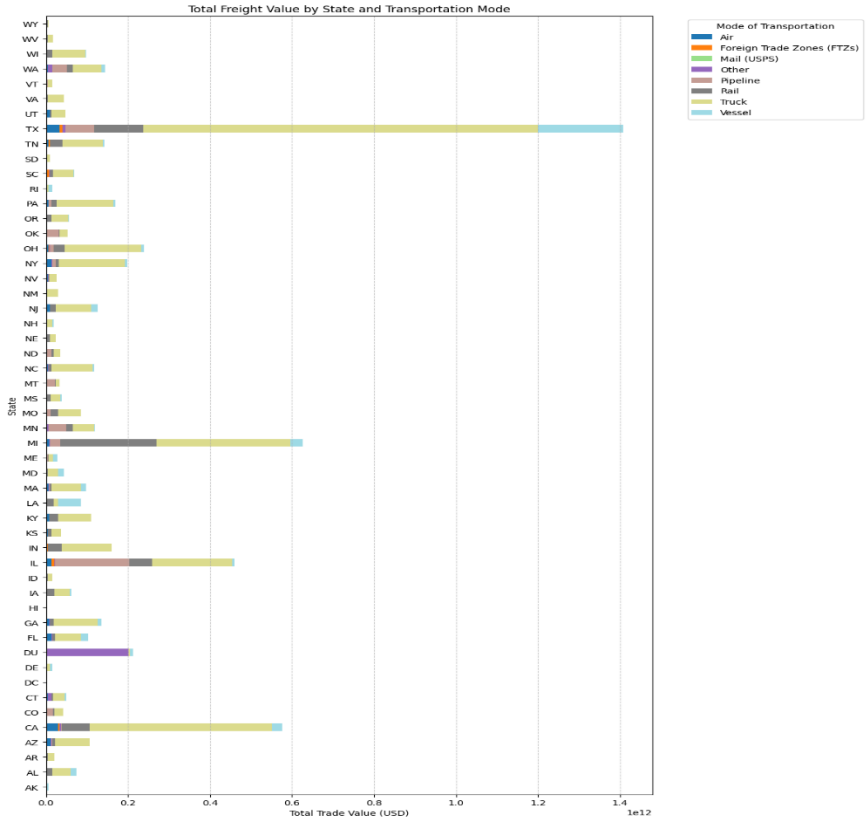
- Higher movement in Q2–Q4 (April–October): Likely due to increased production, trade cycles, and holiday stocking.
- Lower movement in January & February: Common slowdown due to post-holiday demand drop, cold weather, and planning periods. -Freight movement shows a clear seasonal pattern, peaking from May to September, which aligns with economic and trade cycles.



8) How much does each mode contribute to economic productivity across regions (e.g., by border state)?

Insights:

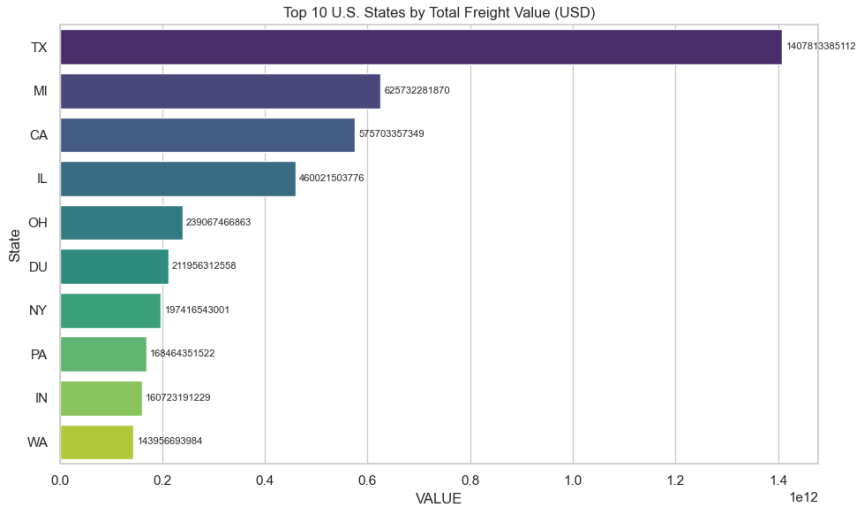
- Texas (TX) dominates by far in total trade value—highlighting its key role as a central freight hub, likely due to extensive border trade with Mexico.
- Michigan (MI), California (CA), and Illinois (IL) also stand out with significant multimodal contributions.



9) What are the top 10 U.S. states (USASTATE) by total freight value?

Insights:

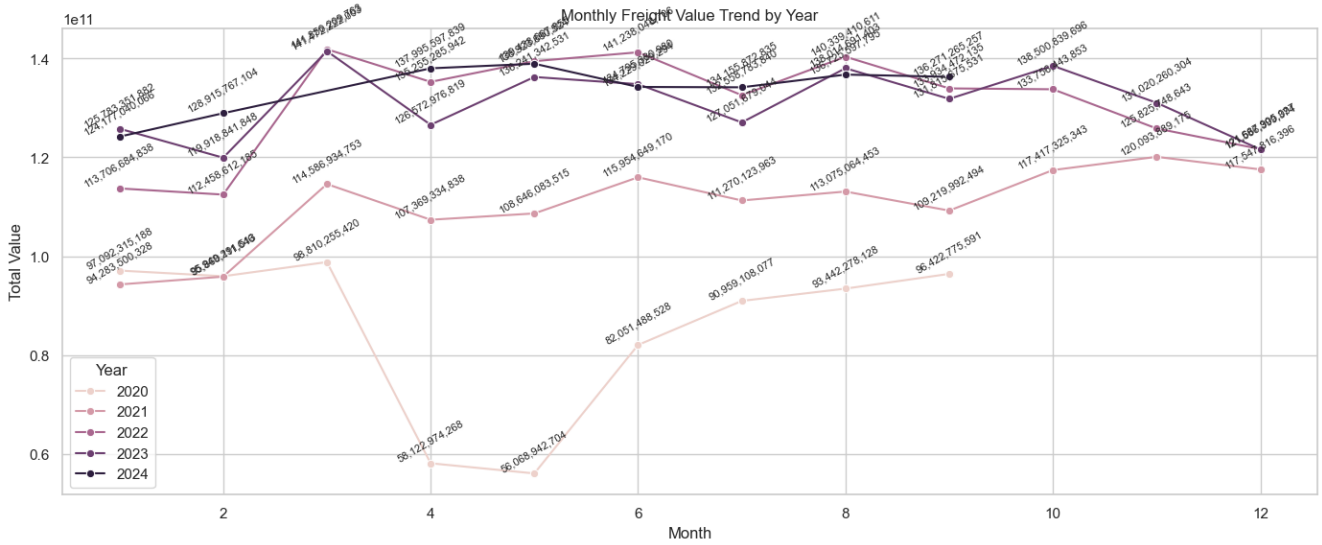
- Texas among the top 10 states has a freight value of 1,407,813,385,112 USD



10) How has the total freight value changed over time?

Insights:

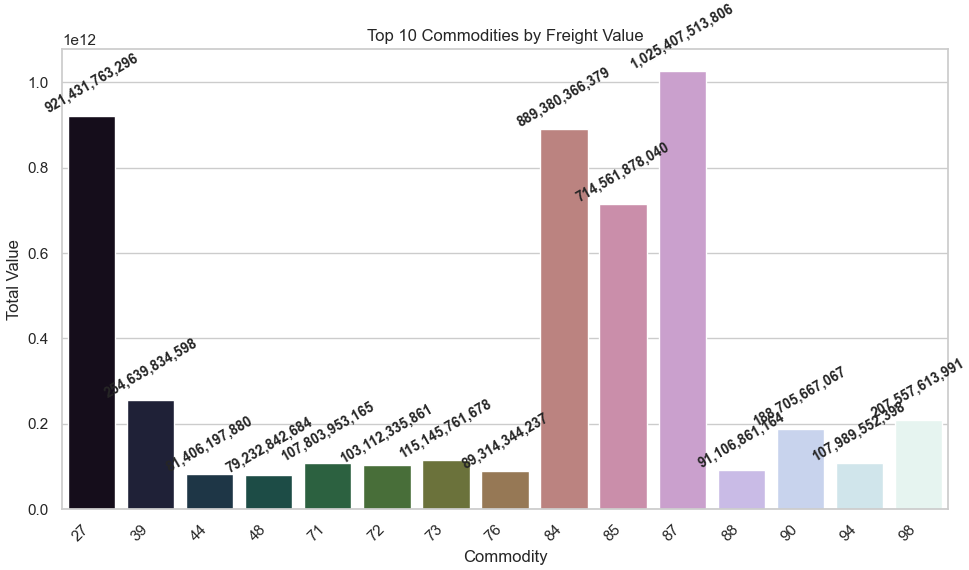
- Across all years, freight value tends to rise mid-year (May–August) and taper off slightly toward December, suggesting seasonal demand cycles.
- Freight value has grown steadily year-over-year, with 2023 and 2024 showing the strongest performance. This reflects expanding trade activity, possibly driven by economic recovery, infrastructure improvements, or increased cross-border flows.



11) What commodities (COMMODITY2) contribute the most to freight value?

Insights:

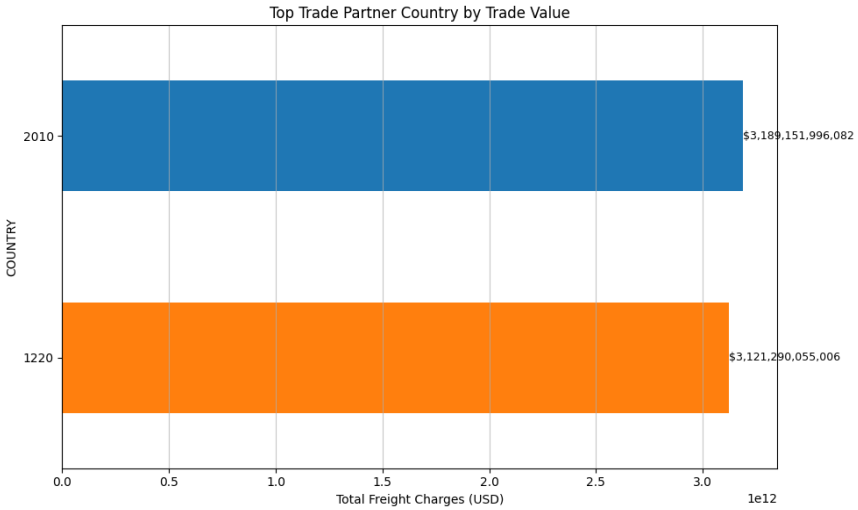
- Vehicles (87) lead the pack, likely driven by strong automotive trade with Mexico and Canada.
- Mineral fuels (27) reflect the high value of energy commodities, especially pipeline and vessel shipments.
- Machinery (84 & 85) show the backbone of industrial and consumer electronics trade.



12) Which country is the biggest trade partner based on freight value?

Insights:

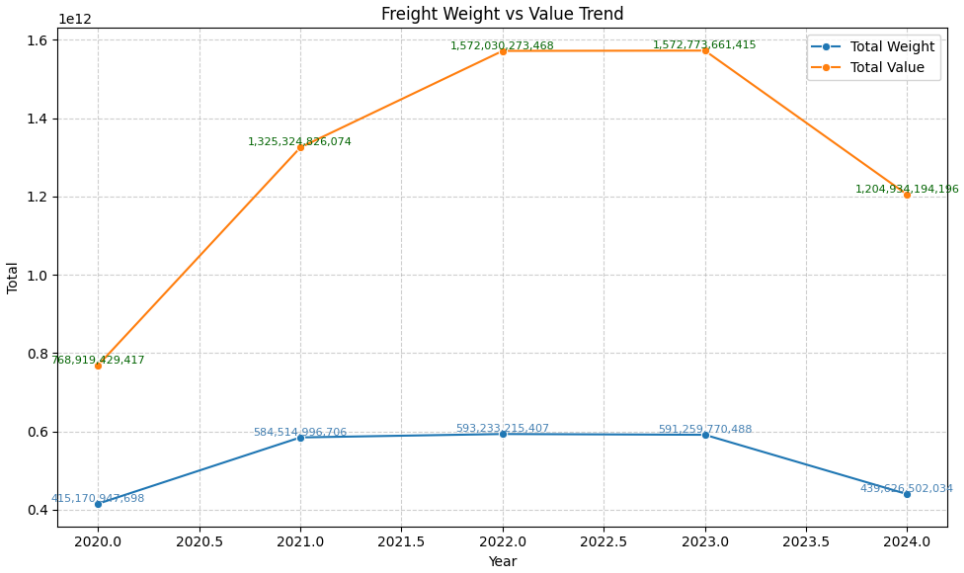
- Mexico is the top trading partner to the United State with a trade value of $3,189,151,996,082.



13) What is the trend of freight weight (SHIPWT) vs value (VALUE) over the years?

Insights:

- Divergence: If VALUE is increasing faster than SHIPWT, it indicates shipping of higher-value goods. (e.g., electronics vs raw materials).
- Parallel Movement: If both increase or decrease together, volume and economic activity are scaling proportionally.
- Yearly Anomalies: Sudden drop in 2020, might indicate events like COVID-19's impact on trade.



14) Which modes or routes are least environmentally friendly?

- Cross-border routes like New Jersey→Mexico and Kentucky→Canada are among the least offenders—likely due to heavy reliance on air/truck and high freight costs per kg.