**TRANSBORDER FREIGHT DATA ANALYSIS**

Ebenezer Omari– [omarieben7@gmail.com](http://omarieben7@gmail.com)

### Introduction

Transportation systems are the backbone of modern economies, enabling the movement of goods, services and people across regions. They play a crucial role in commerce, tourism and daily life, directly impacting economic productivity and societal well-being. However, as transportation networks expand, they face challenges such as safety concerns, congestion, infrastructure stress, environmental impact and economic disruptions.

The Bureau of Transportation Statistics (BTS), a branch of the U.S. Department of Transportation, provides comprehensive and reliable data on various aspects of transportation, including freight movement, passenger travel, safety incidents, infrastructure capacity and environmental impact. This data empowers policymakers, transportation agencies and businesses to make informed decisions that improve efficiency, sustainability and public welfare.

As a data analyst at getINNOtized, with BTS as the company’s largest client, this analysis aims to extract valuable insights from BTS data to identify inefficiencies, uncover patterns and propose actionable recommendations for enhancing transportation systems.

This document outlines the analysis conducted on the Transborder Freight data provided by the Bureau of Transportation Statistics (BTS).

The primary objectives were to:

* uncover freight movement patterns.
* identify inefficiencies.
* analyze environmental impacts.
* safety and risk assessment.
* economic disruptions.
* propose actionable recommendations to improve cross-border freight transportation.

### Key Analytical Questions

1. **What are the top modes of transportation by freight volume and value?**
2. **How does freight movement vary across states and countries?**
3. **What is the trend of freight charges over time?**
4. **Which state contributes the most to freight weight?**
5. **Which country contributes the most to freight value**
6. How does the trade volume differ between Canadian provinces and US States over the past 4 years?
7. **What is the correlation between Freight Value, Shipment Weight and Freight Charges for each year from 2020 to 2024?**
8. **How has the total freight value changed over the years from 2020 to 2024?**

### CRISP-DM Framework

The analysis followed the CRISP-DM methodology, which includes the following stages:

### 1. Business Understanding

The objectives were defined and the above analytical questions were formulated to guide the analysis process.

### 2. Data Understanding

The dataset consisted of 5,242,875 records and 14 variables. Key variables included:

* TRDTYPE: Trade type.
* USASTATE, MEXSTATE, CANPROV: Regions involved in freight movement.
* VALUE: Value of goods in USD.
* SHIPWT: Shipping weight.
* FREIGHT\_CHARGES: Associated transportation costs.
* DISAGMOT: Mode of Transportation
* COUNTRY
* MONTH, YEAR: Temporal attributes.

The dataset covered multiple years and regions, with data on road, rail, air sand water transport modes.

Summary statistics across the dataset (2020 – 2024) revealed:

* **Total Value:** The cumulative value of all freight transactions across the data set is **15.15 trillion USD**. This metric highlights the enormous scale of transborder freight in terms of monetary value.
* **Total Shipment Weight:** The total shipping weight of goods transported is **6.24 trillion units**. This gave an idea of the physical scale of goods being moved.
* **Total Freight Charges:** The total freight charges across all transactions amount to **200.4 billion USD**, showing the financial cost of transportation.

### 3. Data Preparation

To ensure the dataset was ready for analysis, a meticulous data preparation process was carried out as follows:

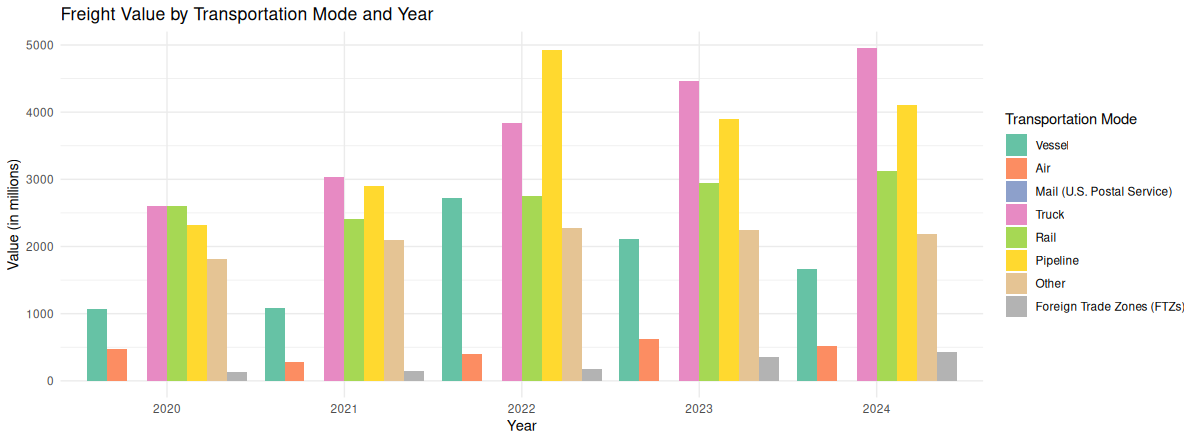
* Monthly data files were organised into corresponding monthly folders.
* These monthly files were merged into a single file for each year (2020 to 2024) using Excel Query. This step streamlined the data and allowed for efficient year-wise aggregation.
* Each year's consolidated dataset was imported into RStudio for further cleaning and transformation.
* The source column was removed as it contained redundant metadata resulting from the aggregation process in Excel.
* Column names across all datasets were carefully checked to ensure uniformity, enabling seamless binding of data across years.
* Empty cells in key columns were identified and replaced with NA values to indicate missing data accurately.
* The cleaned and consolidated datasets from each year were bound together into a single dataset, forming the basis for further analysis.

### 4. Modelling and Evaluation

#### Key Findings from the analytical questions:

#### **What are the top modes of transportation by freight volume and value?**

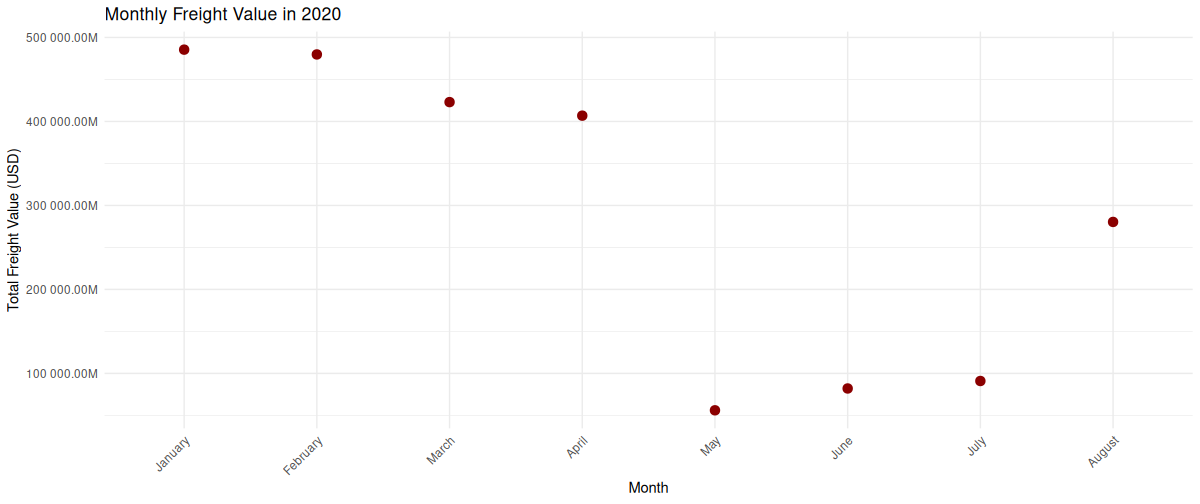
The bar chart below shows the distribution of freight value (in millions of USD) by transportation mode for each year. Each mode is represented by a distinct colour:



**Key Observations:**

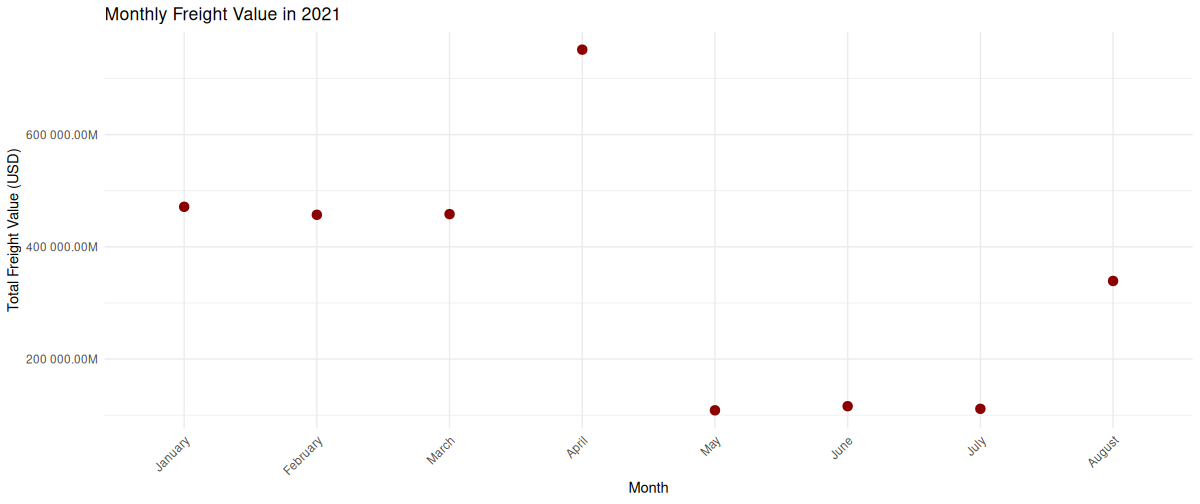
1. Trucks dominate the freight value across most years, indicating their importance in domestic and regional transportation.
2. Air transportation, while less significant in terms of value, consistently contributes a notable portion of the freight value, highlighting its role in high-value shipments.
3. Pipelines and vessels also contribute significantly, particularly for specific years where bulk commodities like oil or goods with high shipping requirements were involved.

#### **What is the trend of freight charges over time?**



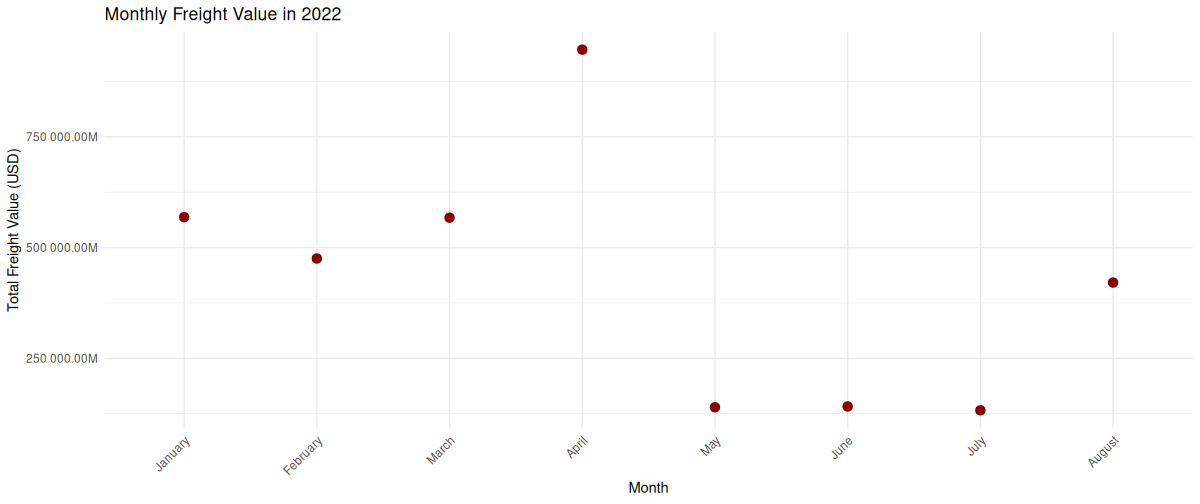
**Analysis for 2020:**

* **High Point:** Freight value peaks in **January**, likely due to increased shipment demand before global COVID-19 lock downs.
* **Low Point: Freight value dips in May, corresponding to the start of significant international shipping restrictions.**



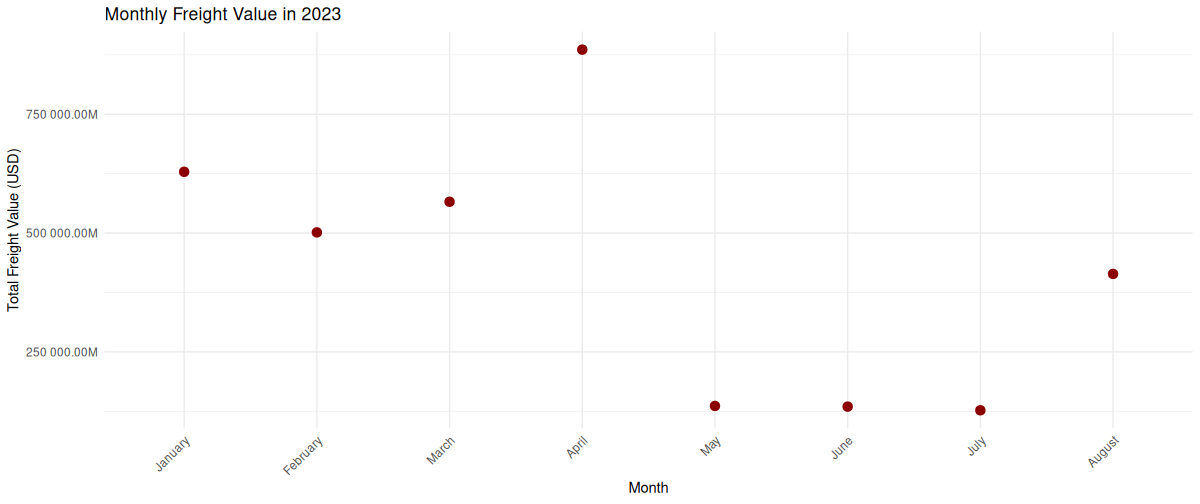
**Analysis for 2021:**

* **High Point: April** has an unusually high freight value.
* **Low Point: May, June and July recorded low freight values.**



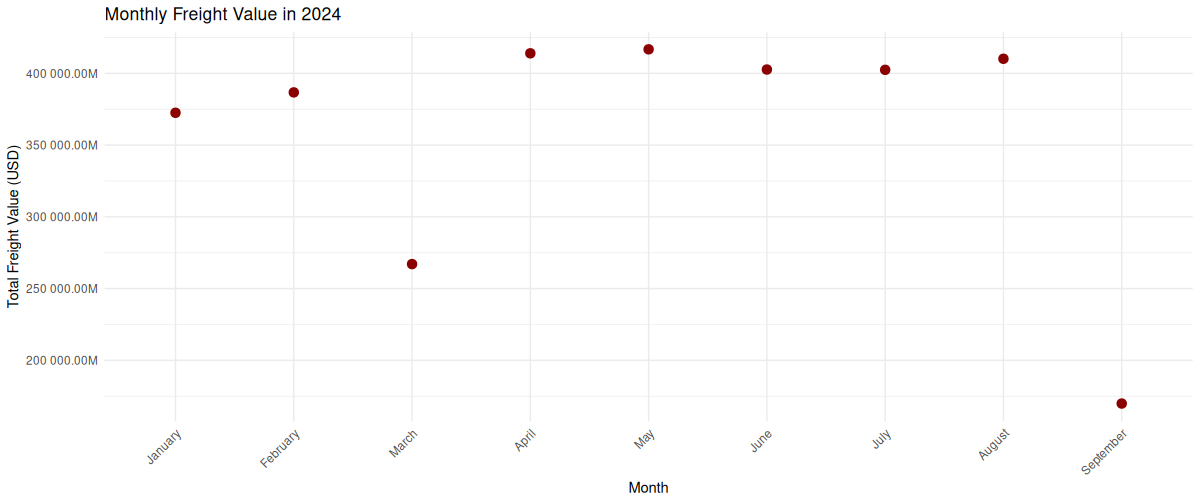
#### ****Analysis for 2022:****

* **Observation:** **April** exhibits a significant rise in freight value, possibly due to easing pandemic restrictions and a rebound in global trade.
* **Low Point:** **May, June, July recorded low freight values,** which might reflect the continuation of supply chain disruptions from 2021.



#### ****Analysis for 2023:****

* **Observation:** **April** continues to have a high freight value, possibly indicating strong summer trade activity.
* **Low Point:** **May, June, July recorded low freight values.**



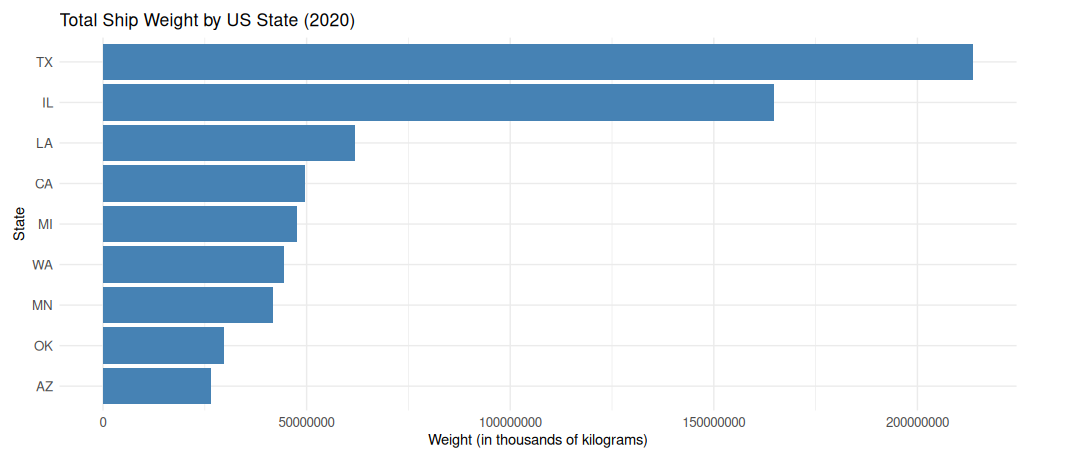
#### ****Analysis for 2024:****

* **Observation:** **May, June, July, August** sees a notable increase.
* **Low Point:** Freight values are relatively low in **September.**

**The data reveals that freight values are influenced by seasonal patterns, economic events and unexpected global disruptions (e.g., COVID-19).**

* **2020 and 2021** exhibit clear impacts from the pandemic, while **2022 onward** shows recovery trends with occasional fluctuations tied to trade cycles.

### ****Which state contributes the most to freight weight?****

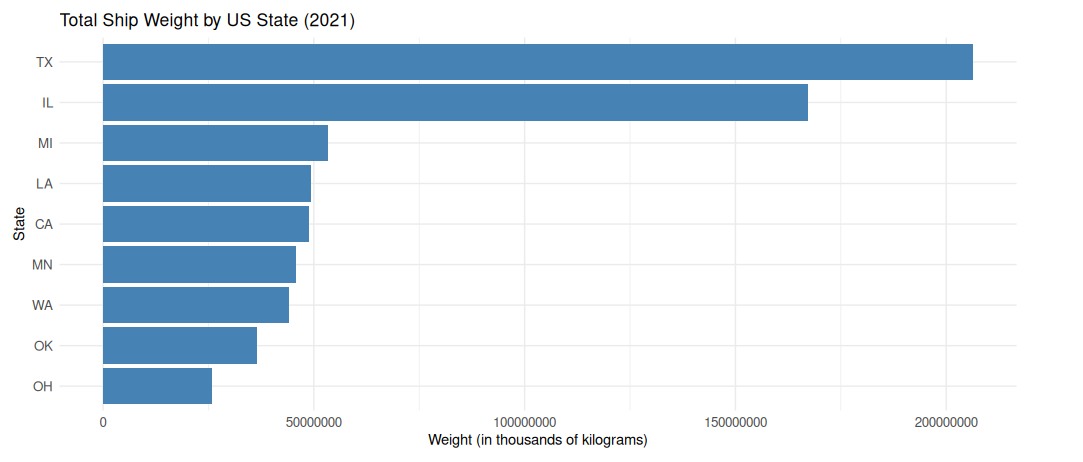


#### ****Analysis for 2020****

The following bar chart illustrates the top 10 states contributing to freight weight in 2020:

**Key Observations:**

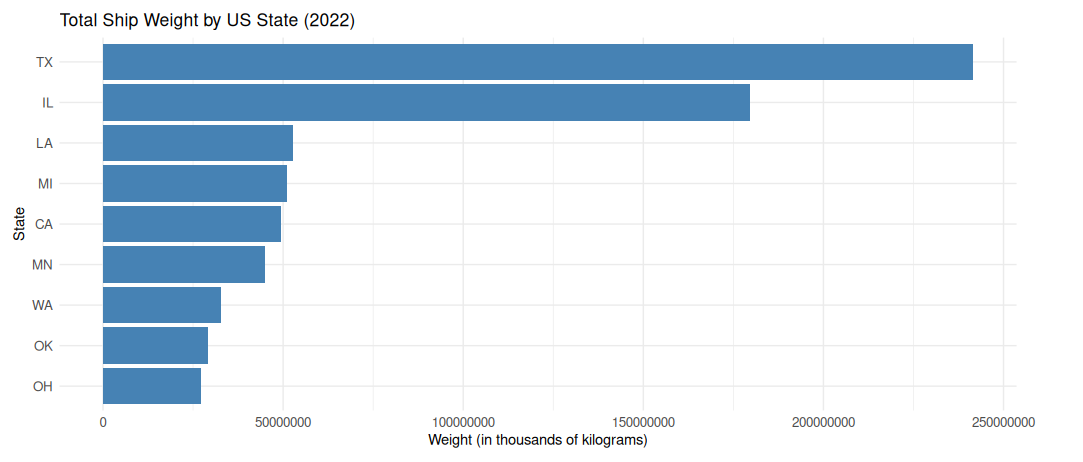
* The states with the highest freight weight include Texas, Illinois and Los Angeles.
* Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.
* States like Los Angeles and California also feature prominently, likely due to their large populations and economic activity.



**Analysis for 2021**

**Key Observations:**

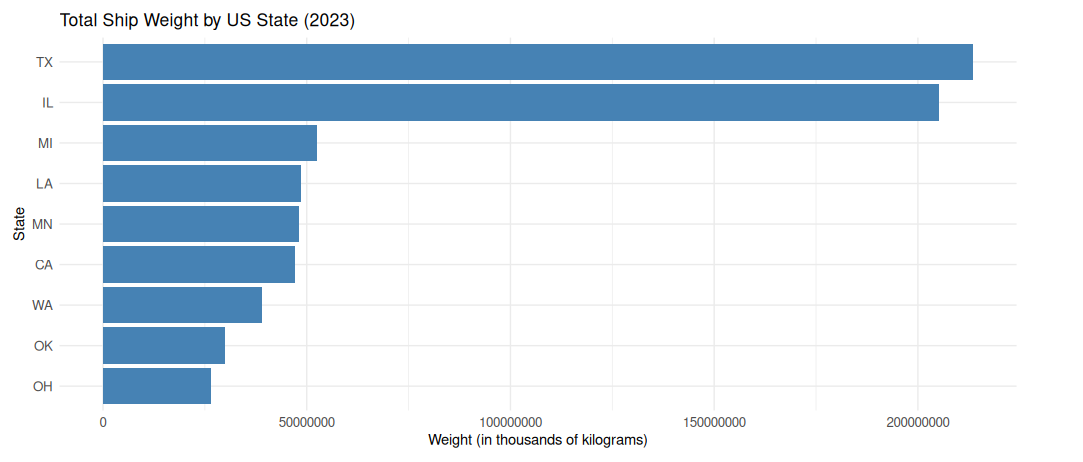
* The states with the highest freight weight include Texas, Illinois and Miami.
* Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.
* **States like Los Angeles and California also feature prominently, likely due to their large populations and economic activity.**



**Analysis for 2022**

**Key Observations:**

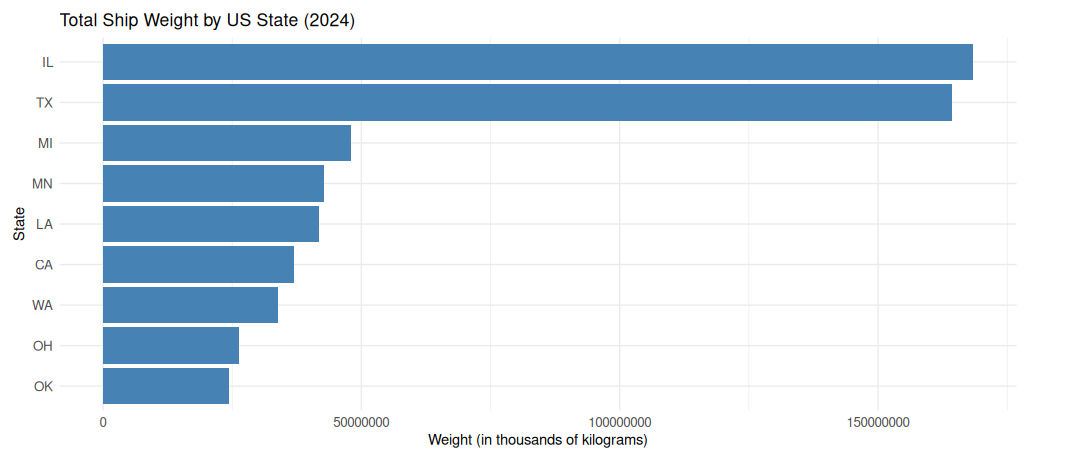
* The states with the highest freight weight include Texas, Illinois and Los Angeles.
* Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.



**Analysis for 2023**

**Key Observations:**

* The states with the highest freight weight include Texas, Illinois and Miami.
* **Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.**

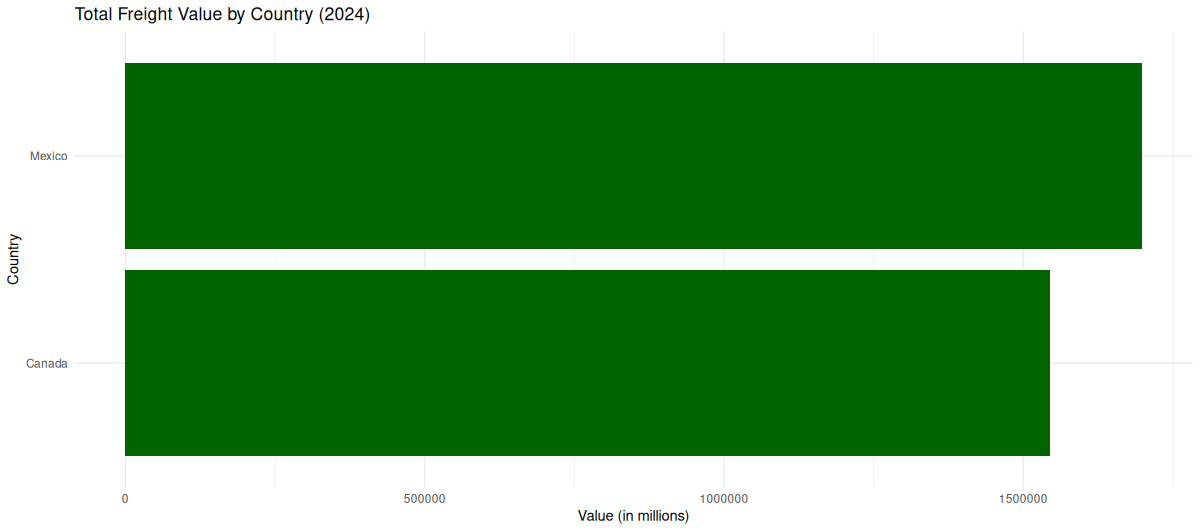
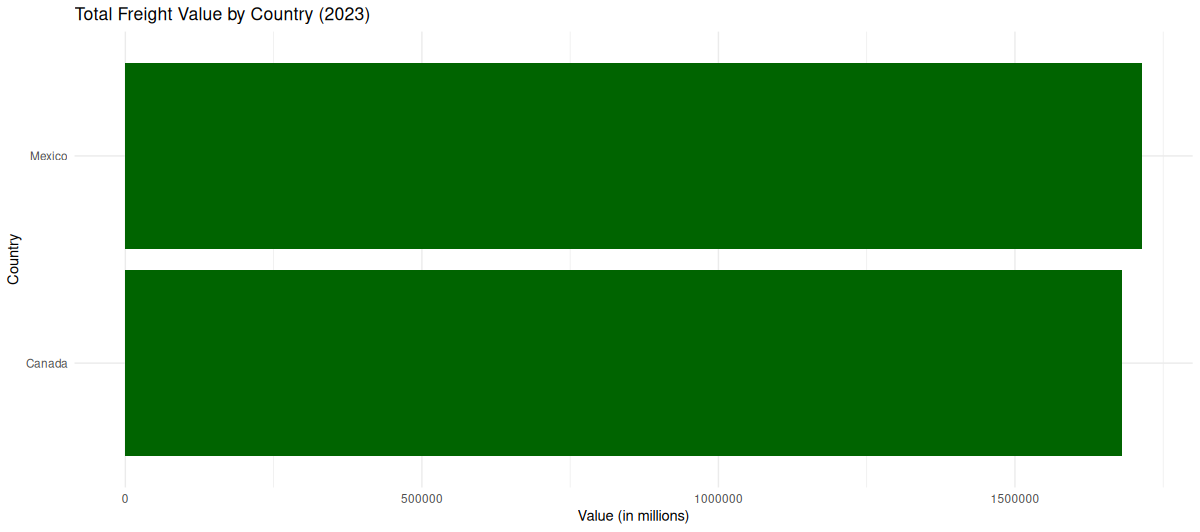
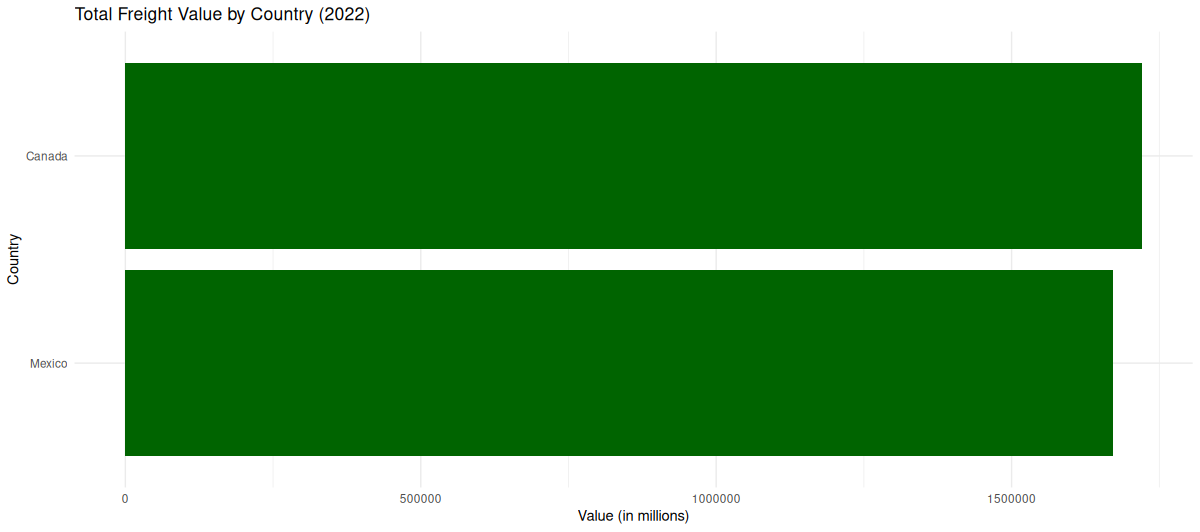
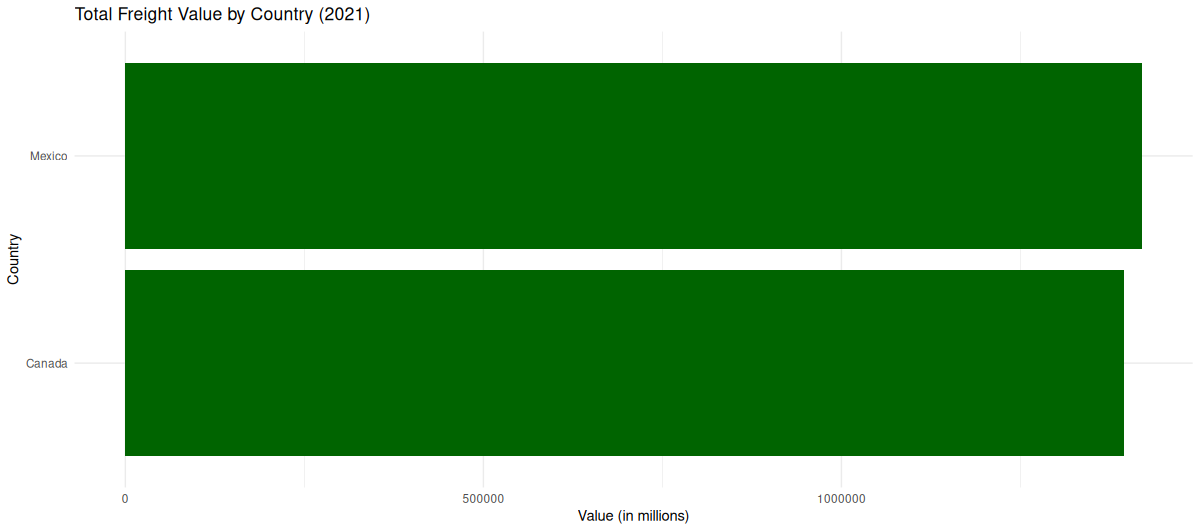
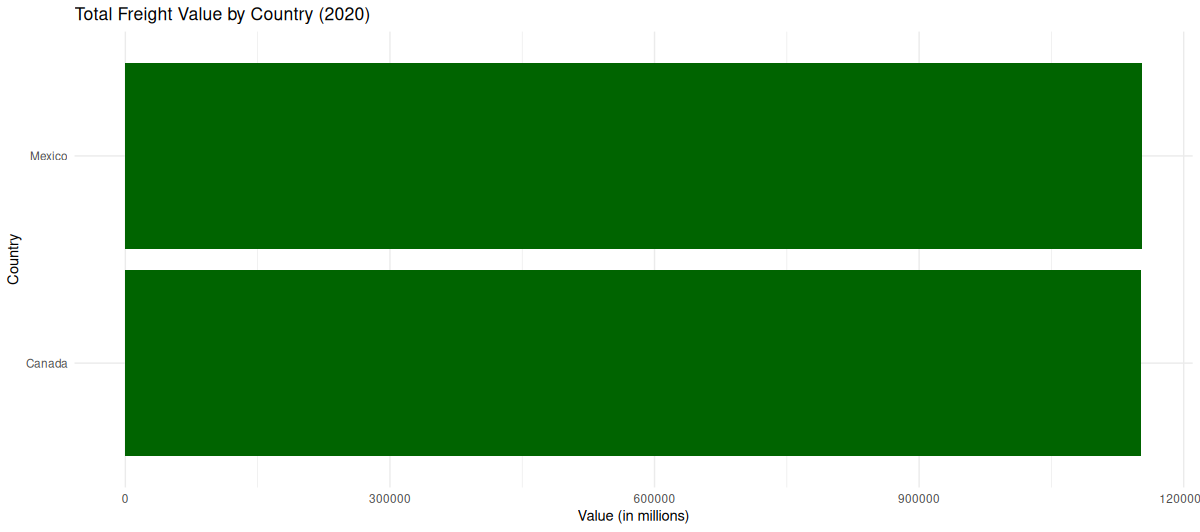


**Analysis for 2024**

**Key Observations:**

* The states with the highest freight weight include Illinois, Texas and Miami.
* **Illinois leads Texas not by a significant margin.**

### ****Which country contributes the most to freight value?****



**Analysis:**

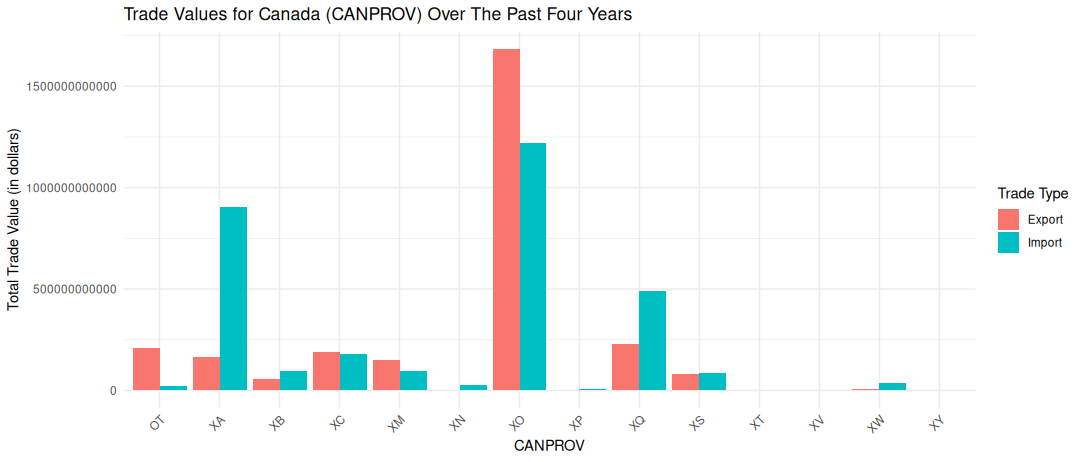
In 2020, Canada and Mexico had nearly the same total freight value. However, in 2021, Mexico surpassed Canada in freight value.

In 2022, Canada dominated the total freight value by country, while Mexico experienced some growth, suggesting challenges that Mexico might have been facing.

In 2023, Mexico regained dominance in total freight value by country, with Canada closing the gap as it experienced some growth.

In 2024, Mexico solidified its dominance in total freight value by widening the gap over Canada. While Canada continued to experience growth, its rate of increase lagged behind Mexico’s, highlighting potential opportunities or challenges for Canada.

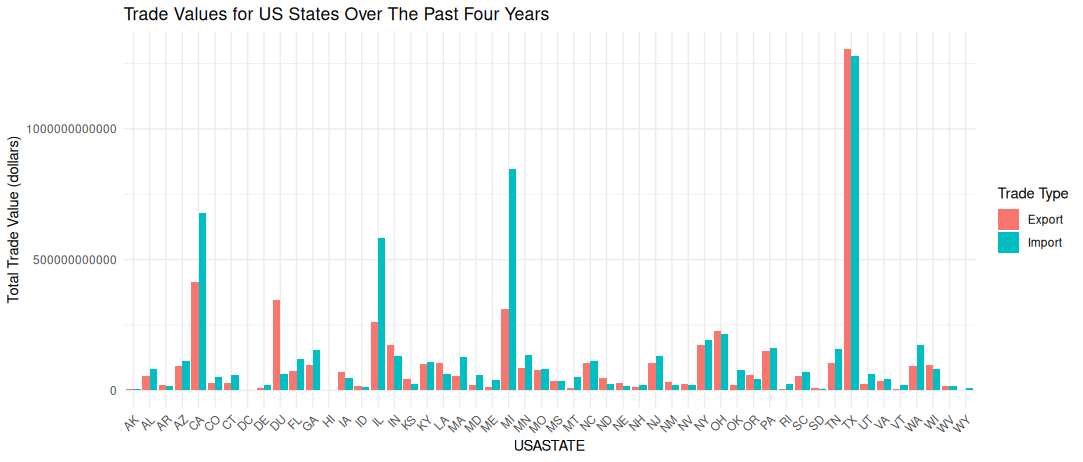
### How does the trade volume differ between US states and Canadian provinces over the past 4 years?



The bar chart illustrates the trade values for Canadian provinces (CANPROV) between 2020 and 2024.

Key insights from the chart include:

* Ontario recorded the highest trade values, with over 1.5 trillion USD in exports and over 1 trillion USD in imports during the period. This highlights Ontario's critical role in cross-border trade.
* Alberta followed as a key trading province, with close to 1 trillion USD in import trade value, demonstrating its importance in the transportation and trade network.
* Trade values for the remaining Canadian provinces were significantly lower, with less than 500 billion USD in total trade value.

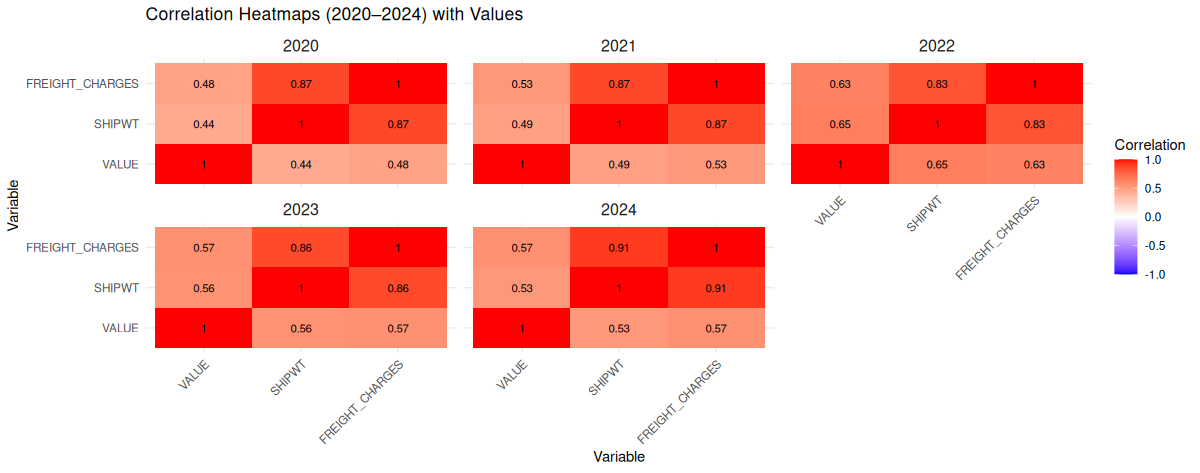


The bar chart illustrates the trade values for US States between 2020 and 2024.

Key Insights from the chart include:

* Texas leads in Trade Value: Texas recorded the highest trade value in both export and import, solidifying its role as a significant player in North American trade.
* Widespread Trade Activity Across States: The majority of U.S. states had notable trade values in both exports and imports, showcasing the connections of cross-border commerce.
* States such as Hawaii recorded no trade values for both exports and imports, likely due to its geographical isolation and limited involvement in freight transportation.

### ****What is the correlation between Freight Value, Shipment Weight and Freight Charges for each year from 2020 to 2024?****



**Analysis**

#### ****2020:****

* **Freight Value vs. Shipment Weight (0.44):** A moderate positive correlation indicates that higher shipment weight is associated with an increase in freight value, though the relationship is not strong.
* **Freight Value vs. Freight Charges (0.48):** A moderate positive correlation suggests that higher freight value is moderately associated with higher freight charges.
* **Shipment Weight vs. Freight Charges (0.87):** A strong positive correlation implies that heavier shipments tend to incur higher freight charges.

#### ****2021:****

* **Freight Value vs. Shipment Weight (0.49):** A moderate positive correlation, slightly stronger than in 2020, shows an increasing association between shipment weight and freight value.
* **Freight Value vs. Freight Charges (0.53):** A moderate positive correlation, also stronger than in 2020, highlights a growing relationship between freight value and freight charges.
* **Shipment Weight vs. Freight Charges (0.87):** The strong positive correlation remains consistent, indicating a continued link between shipment weight and charges.

#### ****2022:****

* **Freight Value vs. Shipment Weight (0.65):** A stronger positive correlation reflects a significant relationship between the value of goods and their shipment weight this year.
* **Freight Value vs. Freight Charges (0.63):** The correlation strengthens, showing a notable association between freight value and charges.
* **Shipment Weight vs. Freight Charges (0.83):** The strong positive correlation decreases slightly but remains substantial.

#### ****2023:****

* **Freight Value vs. Shipment Weight (0.56):** A moderate positive correlation, lower than 2022, indicates the relationship between value and shipment weight weakened slightly.
* **Freight Value vs. Freight Charges (0.57):** A consistent moderate positive correlation reflects a stable association between freight value and charges.
* **Shipment Weight vs. Freight Charges (0.86):** The strong positive correlation indicates the relationship remains solid between shipment weight and charges

#### ****2024:****

* **Freight Value vs. Shipment Weight (0.53):** A moderate positive correlation, slightly lower than in 2023, suggests a declining association.
* **Freight Value vs. Freight Charges (0.57):** The consistent correlation indicates that freight value and charges remain moderately linked.
* **Shipment Weight vs. Freight Charges (0.91):** A very strong positive correlation indicates an almost linear relationship, suggesting that heavier shipments are nearly directly proportional to higher freight charges.

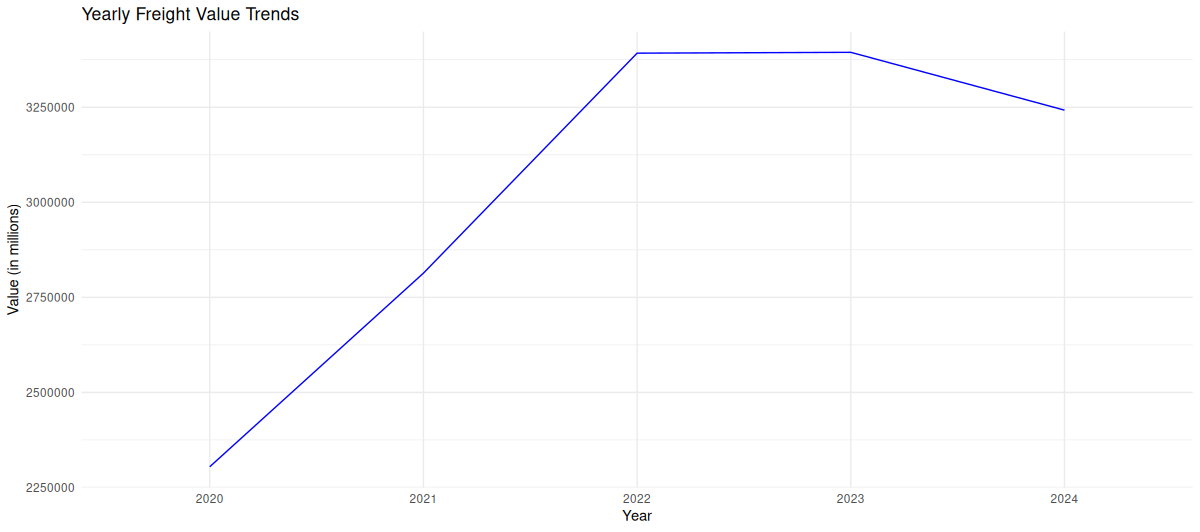
### ****Overall Interpretation:****

1. **Shipment Weight and Freight Charges:** Across all years, there is a consistently **strong positive correlation (above 0.83)**, with the strongest in 2024 (0.91). This indicates that shipment weight is a critical determinant of freight charges.
2. **Freight Value and Shipment Weight:** The correlation is generally moderate, ranging from 0.44 in 2020 to a peak of 0.65 in 2022, and then slightly declining afterward. This suggests that heavier shipments tend to have higher values, but the relationship is not as strong or consistent as with freight charges.
3. **Freight Value and Freight Charges:** The correlation is moderate across all years, ranging from 0.48 to 0.63, indicating that while higher-value shipments often result in higher charges, other factors also contribute.

### ****Insights:****

* Shipment weight is the dominant factor influencing freight charges.
* While freight value contributes to determining charges, its impact is moderate compared to shipment weight.
* The year 2022 shows the strongest correlations across all variables, possibly indicating unusual or specific market conditions.

### ****How has the total freight value changed over the years from 2020 to 2024?****



### ****Analysis of Freight Value Trends (2020-2024):****

### From the data:

### The freight value increased steadily from 2020 to 2022.

### There was a notable jump in 2021 and 2022, with 2022 to 2023 having the highest freight value.

### Slight decline in 2024.

### This suggests consistent growth in freight activities or higher-value shipments over the years.

### ****Insights:****

* The significant year-on-year increase from 2020 to 2024 could indicate:
* Increased global trade.
* An expansion in freight services, especially after 2020 (potentially post-pandemic recovery)

**5. Deployment:**

**Actionable Strategies and Recommendations**

* Increase investment in under-performing transportation modes or routes

Certain transportation modes or routes may be underutilised, leading to inefficiencies and missed economic opportunities. By allocating resources to improve infrastructure, logistics, and operations in these areas, overall freight performance can be enhanced.

An example is investing in better rail infrastructure can reduce costs for bulk freight, such as heavy materials or agricultural products.

The benefit will be reducing congestion in other modes (example road) and provides alternative transport options for businesses.

* Monitor freight charges and operational costs more closely to identify inefficiencies

Large variations in freight charges across modes or regions can indicate inefficiencies, such as unnecessary delays, sub-optimal routing, or excessive charges by service providers. Establishing a robust monitoring framework can identify these issues and enable targeted interventions.

An example is using an automated system for analysing freight charges against industry benchmarks can flag areas for optimisation. This will reduce operational costs for businesses.

* Implement Policies to Reduce Environmental Impact

Freight transport contributes significantly to emissions and environmental degradation. Policies promoting sustainable practice such as using low-emission vehicles, shifting freight to less polluting modes (example rail or water).

This will improve sustainability and alignment with environmental regulations.

* Develop Data-Driven Logistics Solutions

Introducing advanced data analytics can optimise logistics operations by identifying bottlenecks, predicting demand and improving resource allocation. This includes real-time tracking and machine learning-based forecasting models.

An example is the using predictive analytics to anticipate peak freight periods and allocate resources proactively. This will reduce delays, customer satisfaction and minismise operational cost

* Encouraging and introducing the use of technology

Adopting emerging technologies can help streamline freight operations and boost transparency. A typical example is the use of IoT sensors to monitor cargo conditions. This will reduce losses, improve accountability and enhance operational efficiency.

Insights and visualizations were shared via:

* **GitHub Repository: Includes documentation, R script and data files.**