**TRANSBORDER FREIGHT DATA ANALYSIS**

Ebenezer Omari– omarieben7@gmail.com

### **Introduction**

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 states or countries contribute the most to freight weight/value?**
5. **What patterns exist between shipment weight and charges?**
6. **What are the environmental implications of freight transportation?**
7. **How can inefficiencies in the freight system be reduced?**

### **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: Transaction type.
* USASTATE, MEXSTATE, CANPROV: Regions involved in freight movement.
* VALUE: Value of goods in USD.
* SHIPWT: Shipment weight.
* FREIGHT\_CHARGES: Associated transportation costs.
* 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:** $14.57 trillion USD
* **Total Shipment Weight:** 6.2 trillion kilograms
* **Total Freight Charges:** $191.93 billion USD

### 3. Data Preparation

* Empty values in columns such as MEXSTATE, CANPROV, DF were identified and addressed.
* The data was cleaned and aggregated to provide meaningful insights at state, country and mode-of-transport levels.
* Converting categorical and numerical variables to appropriate data types.

### 4. Analysis and Visualization

#### Key Findings from the analytical questions:

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

**2020:**

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

**2021:**

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

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

* **Observation:** **May** 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.

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

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

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

* **Observation:** **April, 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.

### 5. Deployment

Insights and visualizations were shared via:

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