Efficient Supply Chain Management for HIV Health Commodities

Group:
Group 7

Team Members:
Rithvik Rangaraj
Mariya Mansuri
Harshada Daund

Project Overview

Effective supply chain management is vital for guaranteeing the prompt delivery of HIV health products, including antiretroviral (ARV) medications and HIV testing kits. Shipping delays can hinder healthcare services, raise expenses, and decrease operational effectiveness. This initiative aims to utilize predictive analytics to enhance supply chain logistics. The project seeks to improve overall logistics efficiency by tackling shipment delays, optimizing freight expenses, and boosting vendor performance. Findings from this analysis can inform data-driven decision-making, resource distribution, and enhancements in operations to improve global health results.

Objectives

Prediction Goals

1. Classify shipments into categories: On-Time, Less Delay, Moderate Delay, or Severe Delay.
2. Predict the number of delay days as a continuous metric.

Inference Goals

1. Analyze the relationship between shipment attributes, such as cost, weight, and vendor terms, and delivery delays.
2. Identify key logistical factors contributing to delays in specific regions or shipment modes.

Actionable Goals

1. Provide recommendations for cost reduction and process improvements.
2. Prioritize resource allocation for critical, time-sensitive shipments.
3. Highlight areas for vendor and regional optimization.

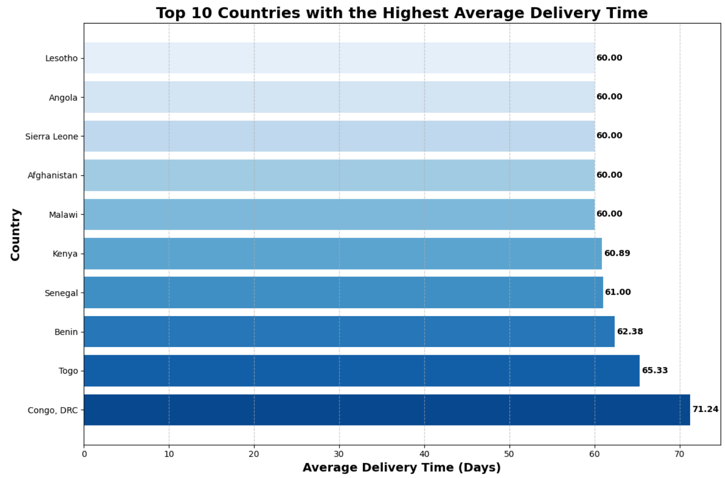
Dataset Overview

The dataset from the **USAID Supply Chain Shipment Pricing Dataset** provides a comprehensive view of global logistics for HIV health commodities.

- **Scope**: Shipment details, including costs, weight, delivery timelines, and vendor information.
- **Size**: ~10,000 rows and 33 columns.
- **Time Frame**: Data spans from **2006 to 2015**, providing a decade-long view of logistics operations.
- **Geographic Coverage**: Global shipments with regional breakdowns.

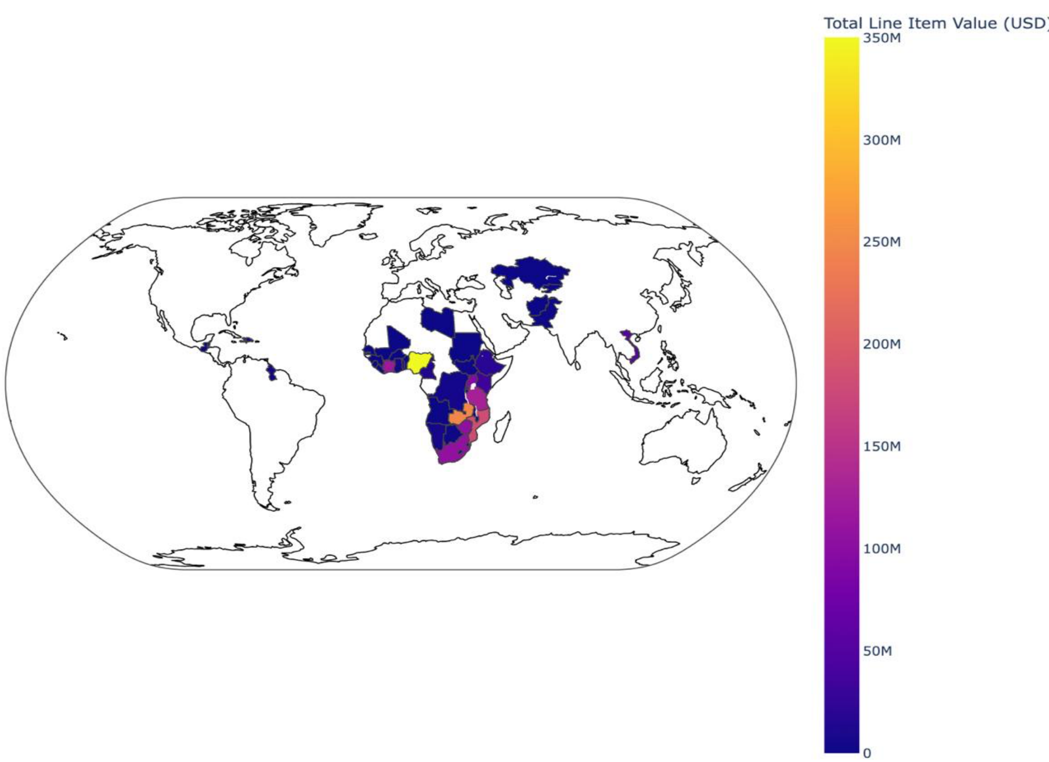
Data Exploration

1. **Top 10 Countries with the Longest Average Delivery Times**



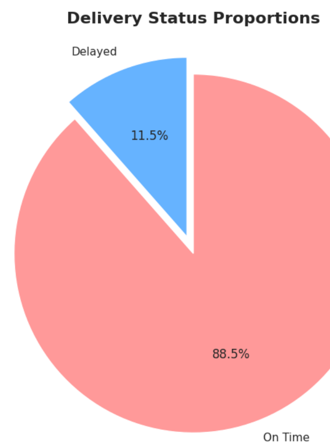
- Congo, DRC has the longest average delivery time (71.24 days), followed by Togo (65.33 days) and Benin (62.38 days).
- Five countries, including Lesotho and Angola, have an average delivery time of 60 days.
- Highlights Delivery Bottlenecks: Identifies countries where delivery processes take the longest.
- Guides Optimization Efforts: Helps stakeholders focus on improving delivery timelines in specific regions.

2. **Aggregate Total Line Item Value by Country**



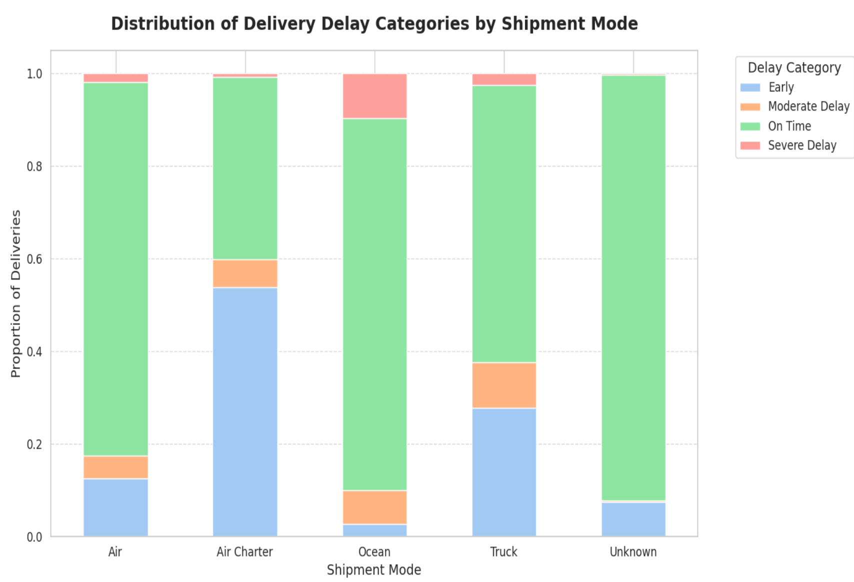
- High-Value Regions: Southern Africa (e.g., South Africa, Botswana) shows the highest total line item values.
- Moderate Values: Central and East Africa exhibit moderate activity.
- Southern Africa dominates due to high demand and shipment volume.
- Sub-Saharan Africa is a key region for shipment activity and financial value.

3. **Delivery Status Proportions**



- 88.5% On Time: Majority of deliveries were timely, indicating an efficient logistics system.
- 11.5% Delayed: Highlighted for further analysis to improve performance.

4. **Distribution of Delivery Delay Categories by Shipment Mode**



- On-Time Deliveries: Air and Unknown modes dominate with reliable, timely deliveries.

- Early Deliveries: Air Charter shows significant early arrivals, suggesting overestimated timelines.
- Moderate Delays: Ocean and Truck modes have noticeable moderate delays due to slower transit.
- Severe Delays: Minimal but slightly higher in Ocean and Truck shipments.

Data Preprocessing

1. **Handling Missing Values**:
    - Imputed missing categorical values using mode and numerical values using the median.
2. **Date Standardization**:
    - Cleaned and standardized date columns for consistency.
3. **Balancing Classes**:
    - Applied **SMOTE (Synthetic Minority Over-sampling Technique)** to address the imbalance in delay categories.
4. **Feature Engineering**:
    - Derived metrics like **Scheduled-to-Delivered Days**, **PQ to PO Days**, and **Freight Cost per Weight Unit**.

Feature Analysis

Statistical Insights

1. **ANOVA Results**:
    - Significant features included
    - **Line Item Quantity** (P-value = 1.65e-49)
    - **Pack Price** (P-value = 5.87e-57)
    - **Freight Cost** (P-value = 2.05e-26).
2. **Correlation Analysis**:
    - Strong correlations: **Freight Cost ↔ Weight ↔ Line Item Value**.
    - Moderate correlations: **Unit Price ↔ Pack Price**.
3. **Chi-squared Tests**:
    - Categorical features like **Shipment Mode** (P-value = 5.91e-248) and **Vendor Terms** (P-value = 0.0) were confirmed as critical predictors.

Selected Features

1. **Time-Related**:
    - Scheduled-to-Delivered Days.
2. **Cost-Related**:
    - Freight Cost, Weight, Line Item Value.
3. **Logistics Attributes**:
    - Shipment Mode, Vendor Terms, Product Group.

Modeling Techniques

Random Forest Classifier

1. **Purpose**:
    - Categorize shipments into delay categories.
2. **Steps**:
    - Addressed class imbalance with SMOTE.
    - Optimized hyperparameters using GridSearchCV.
3. **Performance**:
    - **Accuracy**: 99.58%.
    - Near-perfect precision, recall, and F1 scores.
4. **Insights**:
    - Captured complex relationships effectively.
    - Best suited for categorical predictions.

Linear Regression

1. **Purpose**:
    - Predict days of delay as a continuous metric.
2. **Performance**:
    - **R^2 Score**: 0.07 (features explained only 7% of variance).
    - High errors: **MAE = 13.84 days**, **MSE = 699.06**.
3. **Insights**:
    - Linear regression struggled due to the dataset's inherent non-linearity.

Model Comparison

1. **Random Forest**:
    - Strengths: High accuracy and robustness to imbalanced data.
    - Limitations: Computationally intensive.
2. **Linear Regression**:
    - Strengths: Simple implementation and interpretation.
    - Limitations: Ineffective for non-linear relationships.

Challenges

1. **Data Quality**:
    - Addressed missing and inconsistent shipment records through imputation and cleaning.
2. **Class Imbalance**:
    - Resolved using SMOTE to balance the dataset for delay classification.
3. **Non-Linearity**:
    - Linear regression failed to capture complex relationships, underscoring the need for advanced models.

Actionable Insights

1. **Optimize Shipment Modes**:
    - Prioritize **Air Mode** for time-sensitive deliveries.
    - Reassess reliance on **Ocean Mode** for critical shipments.
2. **Vendor and Regional Focus**:
    - Mitigate delays in regions like **Congo** and **Togo** by addressing logistical bottlenecks.
    - Strengthen partnerships with high-performing vendors.
3. **Cost Management**:
    - Streamline processes for heavy shipments with high freight costs.
4. **Process Improvements**:
    - Refine planning for modes prone to delays.
    - Incorporate advanced models like XGBoost for improved predictions.

Results Summary

1. **Prediction Goals Achieved**:
    - High accuracy (99.58%) in delay classification using Random Forest.
2. **Insights Gained**:
    - Clear understanding of how cost, weight, and shipment mode influence delays.
3. **Actionable Recommendations**:
    - Provided a roadmap for improving logistics efficiency.

Citations

- **Dataset Source**: https://data.usaid.gov/HIV-AIDS/Supply-Chain-Shipment-Pricing-Dataset/a3rc-nmf6/about_data