Efficient Supply Chain Management for HIV Health Commodities

Group Number: 7

- Rithvik Rangaraj
- Mariya Mansuri
- Harshada Daund

Project Overview

- **Objective**:

Leverage historical shipment data to optimize supply chain operations through predictive analytics.

- **Scope**:

Focus on improving delivery timelines, reducing costs, and identifying reliable vendors.

- **Dataset**:

Shipment records with ~10,000 rows and 33 columns, including details on shipment modes, vendors, and delivery dates.

Goals

- **Prediction**:

Predict shipment delays and classify delay severity.

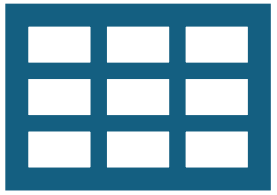
- **Inference**:

Identify factors contributing to delays and high freight costs.

- **Other Goals**:

Analyze vendor performance to optimize supplier decisions.

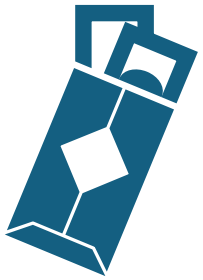
Dataset Overview



Dataset Summary:

~10,000 rows and 33 columns.

Includes shipment details like costs, delivery times, and modes of transport.



Sample Features:

Shipment Mode, Country, Vendor, Weight (Kilograms), Freight Cost (USD), and more.

Methodology

Data Preprocessing
Handled missing values and outliers while encoding categorical variables to prepare the dataset for analysis.

01

02

Exploratory Data Analysis
Conducted visualizations to uncover trends and patterns within shipment data by country and vendor.

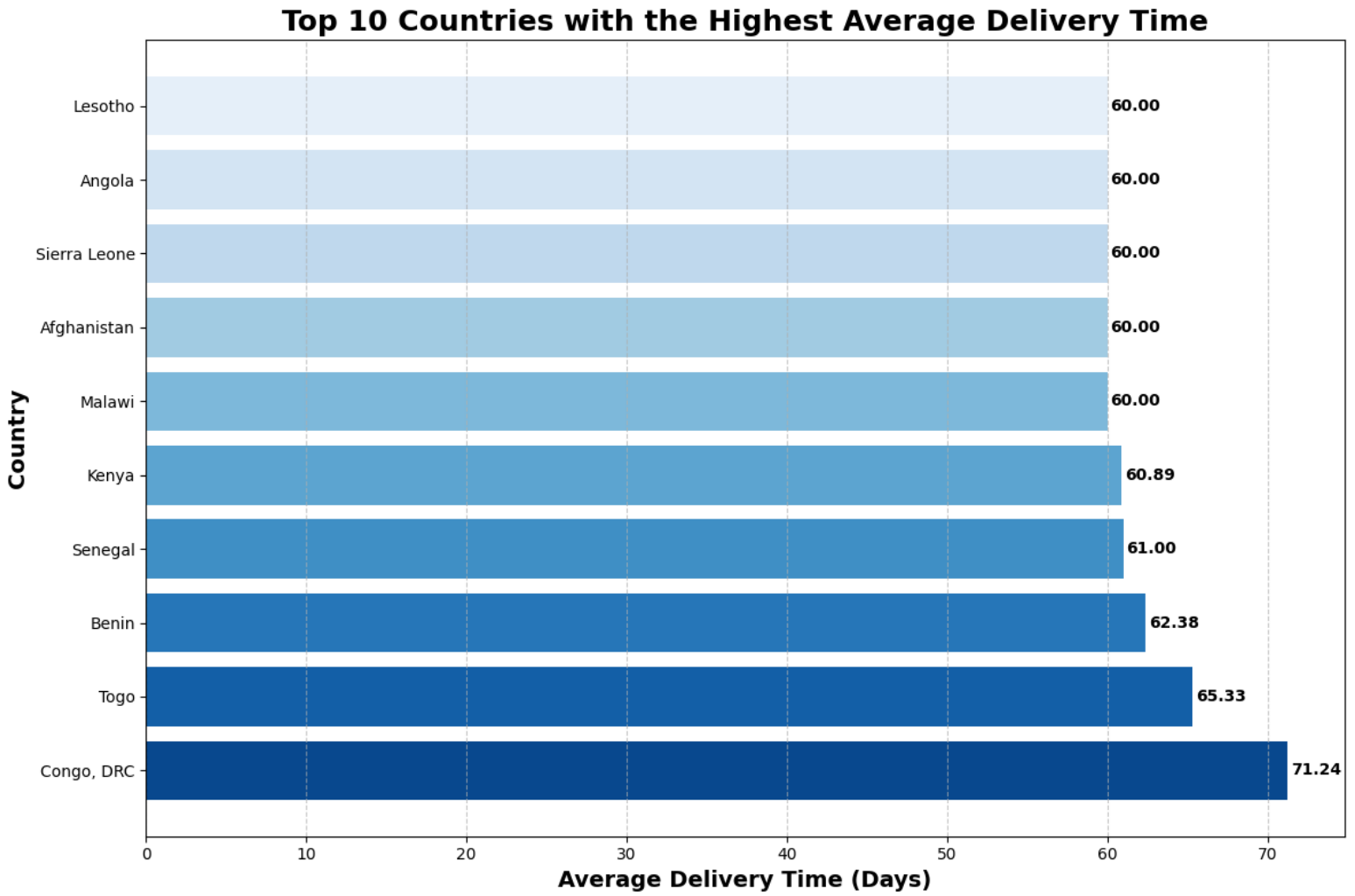
Feature Engineering and Modeling
Recognize and assess which factors have the biggest effects on shipment delays.
Applied Random Forest and Linear Regression to classify shipment delays and predict delivery delay status.

03

Exploratory Data Analysis

Top 10 Countries with the Longest Average Delivery Times

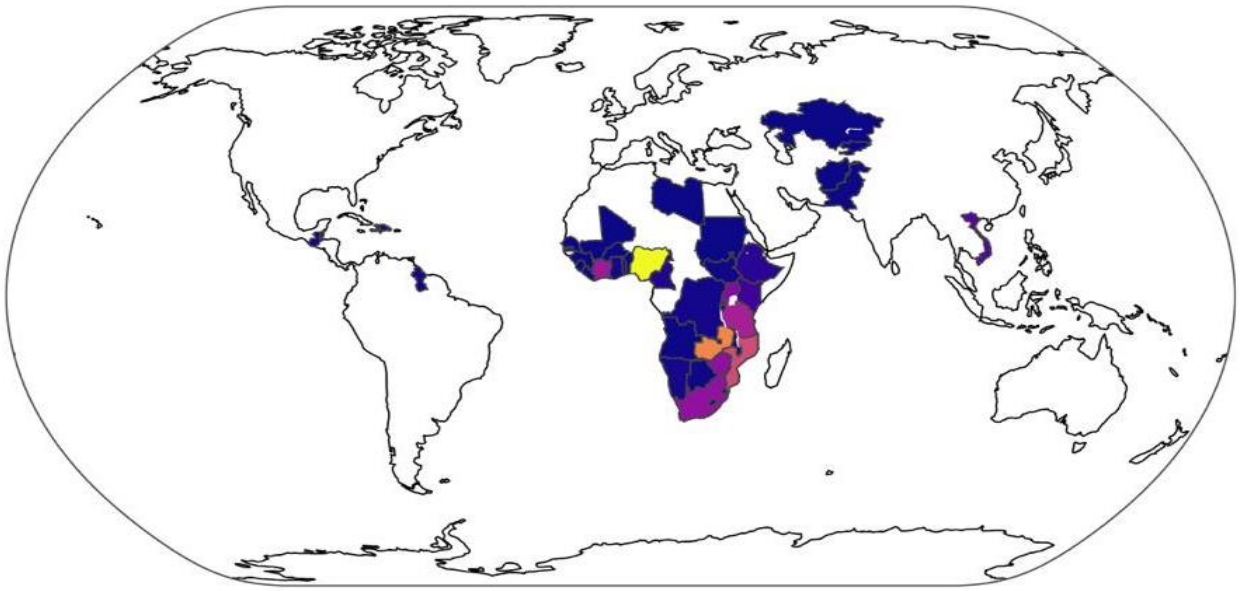
- Congo, DRC has the longest average delivery time (71.24 days), followed by Togo (65.33 days) and Benin (62.38 days).
- Five countries, including Lesotho and Angola, have an average delivery time of 60 days.
- Highlights Delivery Bottlenecks: Identifies countries where delivery processes take the longest.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



Exploratory Data Analysis

Aggregate Total Line Item Value by Country

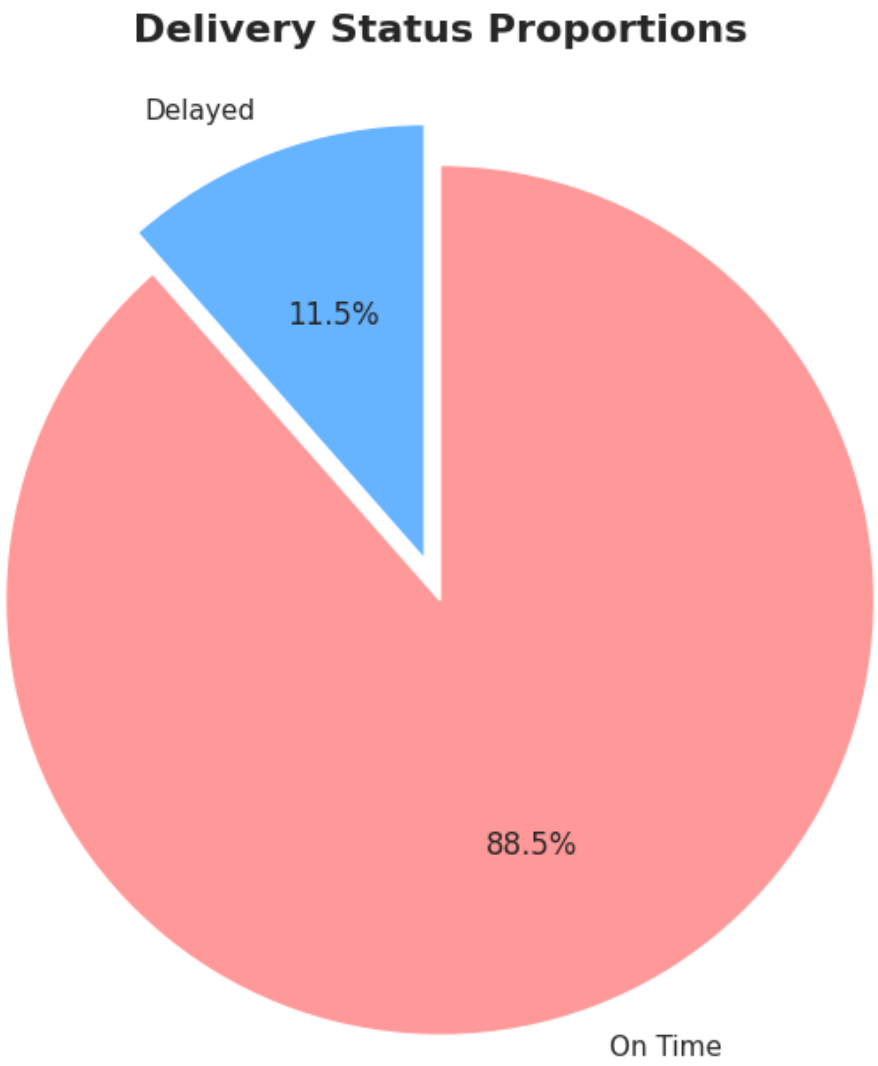
- High-Value Regions: Southern Africa (e.g., South Africa, Botswana) shows the highest total line item values.
- Moderate Values: Central and East Africa exhibit moderate activity.
- Southern Africa dominates due to high demand and shipment volume.
- Sub-Saharan Africa is a key region for shipment activity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



Exploratory Data Analysis

Delivery Status Proportions

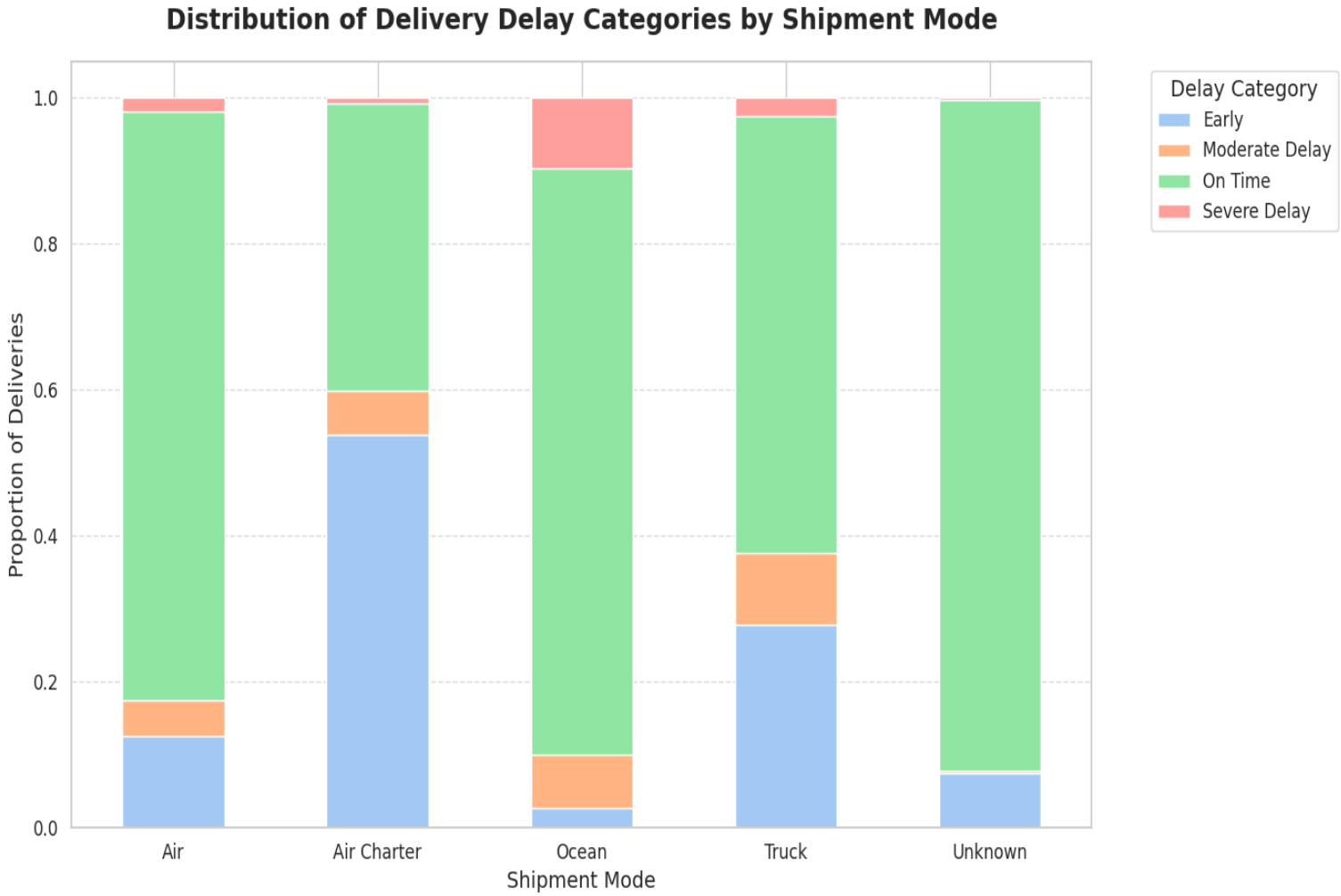
- 88.5% On Time: Majority of deliveries were timely, indicating an efficient logistics system.
- 11.5% Delayed: Highlighted for further analysis to improve performance.



Exploratory Data Analysis

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.



Feature Analysis and Statistical Insights

Purpose:

Understand and identify the most impactful factors causing shipment delays.

Approach:

- Analyzed time-related features like pq first sent to client date days, po sent to vendor date days, scheduled delivery date days, delivered to client date days.
- Evaluated cost-related features such as Freight Cost, Line Item Value, and Weight.
- Applied statistical tests including ANOVA, correlation analysis, and Chi-squared tests to uncover patterns.

Key Findings from ANOVA:

- Line Item Quantity (P-value = 1.65e-49): Larger shipments often face logistical challenges, impacting delays.
- Pack Price (P-value = 5.87e-57): Differences in pricing affect cost dynamics and delays.
- Freight Cost (P-value = 2.05e-26): Strongly linked to shipment delays.

Correlation Highlights:

- Strong Links: Freight Cost ↔ Weight ↔ Line Item Value.
- Moderate Links: Unit Price ↔ Pack Price.

Insights:

Prioritized features with high statistical relevance to minimize redundancy and improve model efficiency.

Modeling Approach

Models Used:

Random Forest Classifier: To predict the level of delays (On-Time, Moderate, Severe).

Linear Regression: To predict the number of days delayed (continuous target).

How We Did It:

Preprocessed data: Filled missing values and applied SMOTE for balanced classes.

Encoded features: Used One-Hot Encoding for categorical data.

Evaluated performance: Metrics included accuracy, precision, recall, F1-score, R^2, MAE, and MSE.

Random Forest Model

What We Did:

- Tuned hyperparameters using GridSearchCV for optimal results.
- Balanced class distribution using SMOTE.

Results:

Accuracy: 99.58%.

Important Features: Shipment Mode, Line Item Value, Pack Price.

Performance Metrics: Near-perfect precision, recall, and F1-scores.

Why It Worked:

Successfully captured complex patterns and handled class imbalances effectively.

Linear Regression Model

Objective: Predict the number of days delayed.

Results:

Accuracy: 38.67 %

R^2 Score: 0.07 (features explained only 7% of the variance).

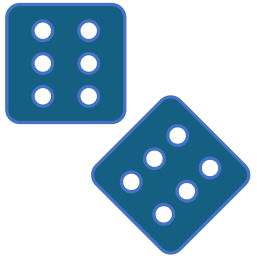
Errors:

- Mean Absolute Error: 13.84 days.
- Mean Squared Error: 699.06.

Key Features: Freight Cost, Weight, Unit Price.

Challenges: The linear model struggled to capture non-linear relationships inherent in the dataset.

Model Comparison



Random Forest:

Strengths: Exceptional accuracy and robust handling of imbalanced data.

Limitations: Computationally intensive and complex.

Linear Regression:

Strengths: Simple and easy to implement.

Limitations: Poor performance due to oversimplified assumptions about relationships.

Key Insights

- Features like Freight Cost, Shipment Mode, and Line Item Value play a major role in predicting delays.
- Statistical tests ensured robust and relevant feature selection.
- Random Forest proved to be highly effective, achieving an accuracy of 99.58%.
- Linear Regression struggled, highlighting the importance of non-linear models for this data.
- Insights provide actionable improvements for logistics and vendor management.

Thank You