Data Science Capstone DATS 6501 Sec 13

BREAKING AFRICA'S TRANSPORT BOTTLENECK

WB Group 4 Team Members
Kumar Saurav Jha
Abishek Chiffon Muthuraja
Swathi Kannetti Ramana Reddy

World Bank Team Leads
Charles Kunaka
Kisanet Haile Molla

Special Thanks to Prof. Abdi Awl (GWU)

TABLE OF CONTENTS

1 INTRODUCTION

1.A Introduction

In many African countries, transport costs represent one of the largest barriers to trade competitiveness. The high cost of moving goods—especially food, minerals, and manufactured products—affects both export performance and domestic market accessibility. Understanding and modeling these logistics costs is critical for evaluating transport infrastructure investments and designing effective trade policy interventions.

This project builds on the **already imputed UNCTAD Trade and Transport (T&T) dataset**, which has been integrated into **FlowMax**, an international freight transport model. The dataset provides mode-specific, country-to-country transport costs disaggregated by commodity. We use this enriched dataset to analyze African export logistics, explore income-based disparities in cost and mode usage, and identify trade anomalies that reveal deeper insights into regional trade behavior and infrastructure limitations.

1.B Problem Statement

Project Requirements

Goal of the Project

- Develop a mode choice model for Sub-Saharan Africa by leveraging the UNCTAD dataset.
- The model will help fill missing transport mode data and enable mode choice analysis.
- The final outcome is a calibrated mode choice model that reflects transport patterns in Africa.

Data Collection and Preparation

1. Primary Data Sources

- UNCTAD Trade & Transport Dataset (T&T):
  - Covers 2016–2021, includes 222 exporters and 169 importers.
  - Provides logistics costs (CIF–FOB), commodity, mode of transport, and trade volumes.

- - Limitations: 72% missing data, no hinterland transport costs.

- FlowMax:
    - Offers distance and time matrices for global multimodal routes.
    - Supplies additional trade attributes such as port-hinterland distances, travel time, and trade imbalance.
    - Enables disaggregation by origin, destination, mode, and commodity (O-D-M-C matrix).

2. Data Harmonization and Cleaning

- Mapped all countries to income groups using ISO codes.
- Cleaned missing or invalid rows (e.g., where logistics cost or distance = 0).
- Converted HS-6 commodity codes to 19 standard FlowMax commodity categories to reduce dimensionality and support modeling.
- Filled missing hinterland information by estimating road-based port-hinterland segments.

3. Imputation of Missing Costs

- Trained a Decision Tree Regression model to predict unit logistics costs per distance ($/ton-km).
- Features used:
    - Mode, commodity, bilateral trade volume
    - Port-hinterland travel time and distance
    - Trade imbalance indicator
- Used FlowMax-estimated inputs to impute missing values where UNCTAD lacked coverage.

4. Output

- Generated a complete matrix of unit logistics cost per tonne for all country pairs.
- Includes adjustments for the hinterland segment for maritime routes.
- Ensures that trade scenarios assessed in FlowMax reflect realistic, granular cost structures.

1.C Elaboration

The project aims to support transport policy and trade facilitation efforts by analyzing logistics costs and export flows across African countries. Leveraging an enriched dataset built from the UNCTAD Trade and Transport (T&T) database and integrated into the FlowMax model, the study examines multimodal transport behavior and the underlying factors driving high logistics costs in the region.

Our analysis explores how export volumes, transport modes, commodity types, and destination patterns vary across African economies. Special attention is given to how geography,

infrastructure availability, and trade characteristics affect logistics performance. By identifying inefficiencies, cost anomalies, and strategic trade routes, this project helps uncover opportunities for improving supply chain efficiency, particularly for landlocked and infrastructure-constrained countries.

Ultimately, the project contributes to evidence-based recommendations for improving Africa's trade competitiveness and resilience, especially in the context of growing demand for more integrated and cost-effective transport networks.

1.D Motivation

- Over 90% of traffic in Africa moves by road, which is expensive.
- Food and mineral supply chains are inefficient, and shifting to rail can lower costs.
- Current mode choice models use European data, not African data, so they do not reflect real conditions in Africa.

1.E Project Scope

This project focuses on:

- Analyzing African exports by income group to uncover the drivers behind logistics cost variations.
- Evaluating anomalies and infrastructure limitations, particularly for landlocked and island nations.
- Visualizing key logistics indicators such as mode of transport usage, commodity distribution, and top destination regions.

The outcome will help FlowMax produce more granular, mode-specific cost simulations and better assess the impact of transport policy reforms and infrastructure investments globally.

2 Literature Review

2.A Relevant Research

UNCTAD Dataset
 Offers global insights into transport costs and trade flows, serving as a benchmark for analyzing Africa's logistics efficiency.
https://unctadstat.unctad.org/datacentre/dataviewer/US.TransportCosts

International trade in Africa by type and preferred mode of transportation 2022
 Highlights the preferred transport modes for different commodities in Africa and their impact on trade efficiency.
**https://www.statista.com/statistics/1327892/international-trade-in-africa-by-type-and-preferred-mode-of-transport/**

Technical Documentation GTCDIT June 2023
 Provides an in-depth explanation of the GTCDIT dataset's structure and methodology, detailing variables like costs, flows, and distances. Private sharepoint.

WB methodology Imputation of Missing Values TC Full Dataset
 Describes the data imputation methods (e.g., mirroring) used to address missing values in the transport cost dataset. Private sharepoint.

3. Methodology

Data Analysis Workflow
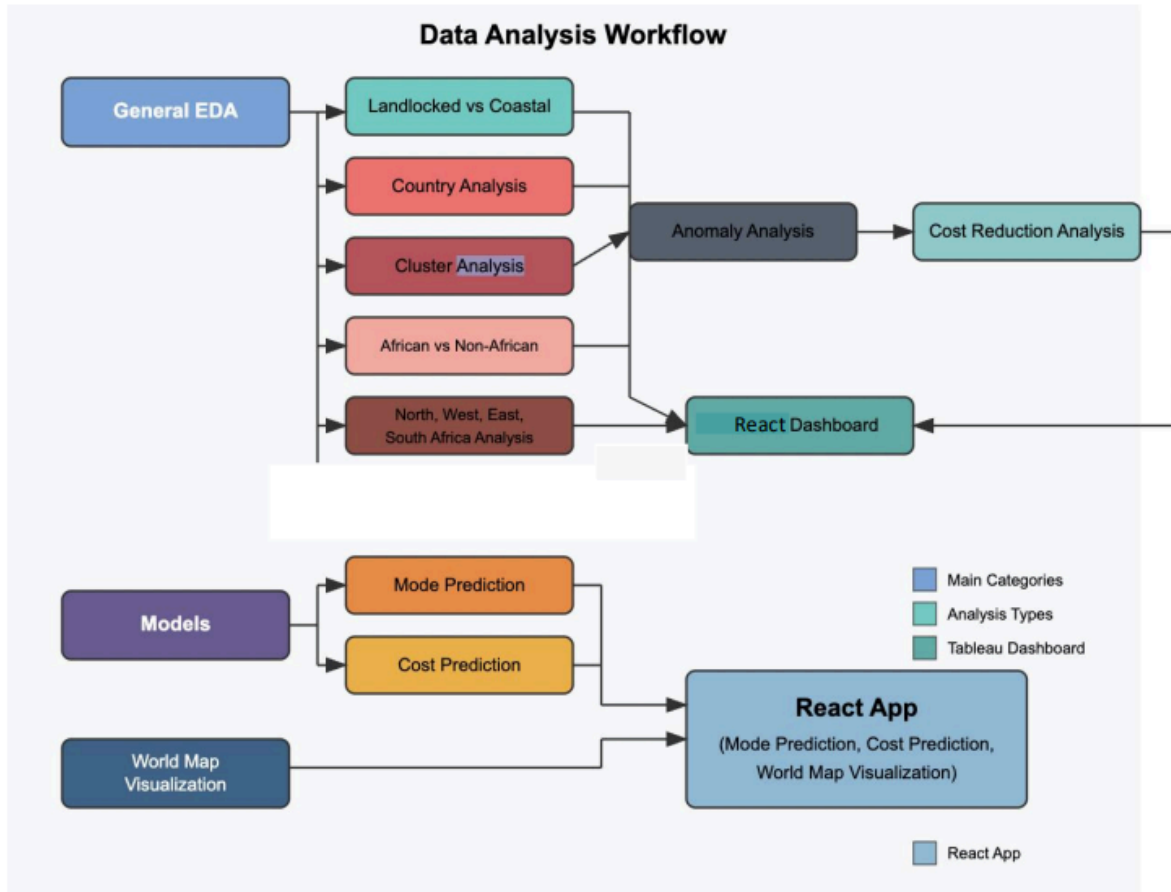
We conducted multiple analyses and exploratory data analyses including comparisons of landlocked versus coastal versus intercontinental trade, country-cost analysis, anomaly detection, African versus non-African analysis, inter-intra regional analysis, import-export analysis, and cost-reduction analysis. We also developed predictive models for both cost and transport mode. Additionally, we created a world map visualization to easily visualize trade flows across countries.

This concise summary effectively outlines the comprehensive scope of your research and the various analytical approaches you've employed.

3.A. Dataset Description

Variable Name	Description	Why It's Used
origin_ISO	ISO3 code of the **exporting** country.	Identifies the origin of goods for bilateral trade flows. Used to group by region, income level, and geography.
destination_ISO	ISO3 code of the **importing** country.	Same as above for the receiving end of the trade flow.
flow(tonne)	Volume of goods transported (in **metric tonnes**).	Used to weight export flows, compute top commodities, mode distribution, and total logistics cost.
distance(km)	Total transport distance from origin to destination (in kilometers).	Used to calculate **unit logistics cost per km** and to assess cost variations by geography.
mode	Numeric encoding of the mode of transport: • 0 = Air • 1 = Rail • 2 = Road • 3 = Sea	Enables easier modeling (e.g., ML) by transforming the categorical transport mode into integers. Helps in feature engineering and classification.
Mode_name	Dominant mode of transport (e.g., Road, Rail, Sea, Air).	Used to disaggregate costs and flows by transport mode. Critical for policy insights on mode choice.
ship_type	Categorical ship classification based on commodity (e.g., container, bulk, tanker).	Relevant for maritime shipments. Helps in analyzing port infrastructure needs and shipment patterns.

`container_type`	Classification of containers used (coded 0–4).	Used to infer the packaging/handling characteristics of goods. Links to transport equipment/infrastructure required.
`IFM_HS`	FlowMax-level commodity group (19 categories, e.g., food, textiles, metals).	Mapped from HS-6 codes to reduce dimensionality and ensure consistent comparison across countries.
`Unit logistics costs ($/ton)`	Final computed cost per tonne, inclusive of transport + time + handling.	Key variable for comparing country competitiveness and logistics efficiency.
Origin type / Destination Type	Classifies the origin country as either continental or overseas	Used to distinguish island nations or remote exporters (e.g., Seychelles) from land-connected ones. Affects available modes and cost assumptions.
Origin region / Destination region	Regional sub-classification countries: • East • West • Central • Southern • North	Allows intra-country trade patterns to be segmented and compared geographically.Useful for heatmaps and region-specific policy insights.
`origin_continent, destination_continent`	Continent of the origin and destination country.	Useful for aggregating analysis geographically (e.g., Africa vs. Non-Africa).
`income_group`	Income classification (Low, Lower-middle, Upper-middle, High) based on World Bank definitions.	Used for stratified analysis to examine cost behavior across different development levels.

3.B. Data Collection

The entire dataset was collected by UNCTAD. For further understanding into the imputational methods involved, refer to the next section.

3.C. Data Preprocessing and Feature Engineering

World Bank Data Imputation Method for Transport Costs

To address the large number of missing values—about 72%—in the UNCTAD Trade and Transport (T&T) dataset, the World Bank team developed a two-step machine learning approach to estimate realistic transport costs across countries, commodities, and transport modes.

Step 1: Model Training on Known Data
 The first step involved training a decision tree-based model using the CIF-FOB margins (the difference between cost-insurance-freight and free-on-board values) reported at the HS-6 product level from 2020. The model used features such as:

- trade volume between countries,

- great-circle and FlowMax route distances,

- port-hinterland distances and travel time,

- trade imbalance indicators,

- landlocked country (LLDC) status, and

- categorical indicators for commodity type and transport mode.

The model's goal was to predict **unit logistics cost per tonne-kilometer**. Several model variants (including Random Forest and Gradient Boosting) were tested, and hyperparameters were optimized using a multi-objective algorithm (NSGA-II) to minimize error and improve fit. The model achieved a good predictive performance, especially for bulk commodity flows, with R^2 values between 0.42 and 0.62 across different commodity groups.

Step 2: Predicting Missing Transport Costs
 In the second step, the trained model was used to predict missing values for over 2.6 million

origin-destination-commodity-mode combinations. The predicted unit costs were then multiplied by route distances and adjusted for port-hinterland legs using road transport rates and time-value estimates. This resulted in a comprehensive matrix of **door-to-door unit transport costs** for the year 2021.

This imputed dataset now enables more complete policy analysis and simulation in FlowMax by providing consistent and comparable cost estimates even for landlocked and underreported regions.

3.D Data Modelling and Visualizations

We built 2 models for cost and mode after selecting the variables using variable importance.

3.D.A: Feature Importance

To understand the key drivers behind transport mode selection, we trained a tuned XGBoost classifier using a randomized hyperparameter search across 10 parameter combinations. The final model achieved strong performance with an overall F1-score of 0.85, indicating reliable predictive ability across classes.

```
              precision    recall  f1-score   support

           0       0.84      0.87      0.85    100497
           1       0.84      0.72      0.78     50121
           2       0.84      0.87      0.85     56366
           3       0.88      0.89      0.88    101668

    accuracy                           0.85    308652
   macro avg       0.85      0.84      0.84    308652
weighted avg       0.85      0.85      0.85    308652
```

Using the trained model, we extracted feature importances to identify the most influential variables. The top predictors were:

- **Flow volume (tonnes)** – the most important factor, accounting for ~30% of model weight,

- **Commodity type (HS-encoded) and Ship type**, jointly shaping routing decisions,

- **Origin-destination pair encoding**, reflecting corridor-specific logistics patterns,

- **Distance**, confirming its core role in shaping transport costs and decisions.

The insights from these features guided our visual analysis and further modeling steps.

SHAP Summary for Mode: air
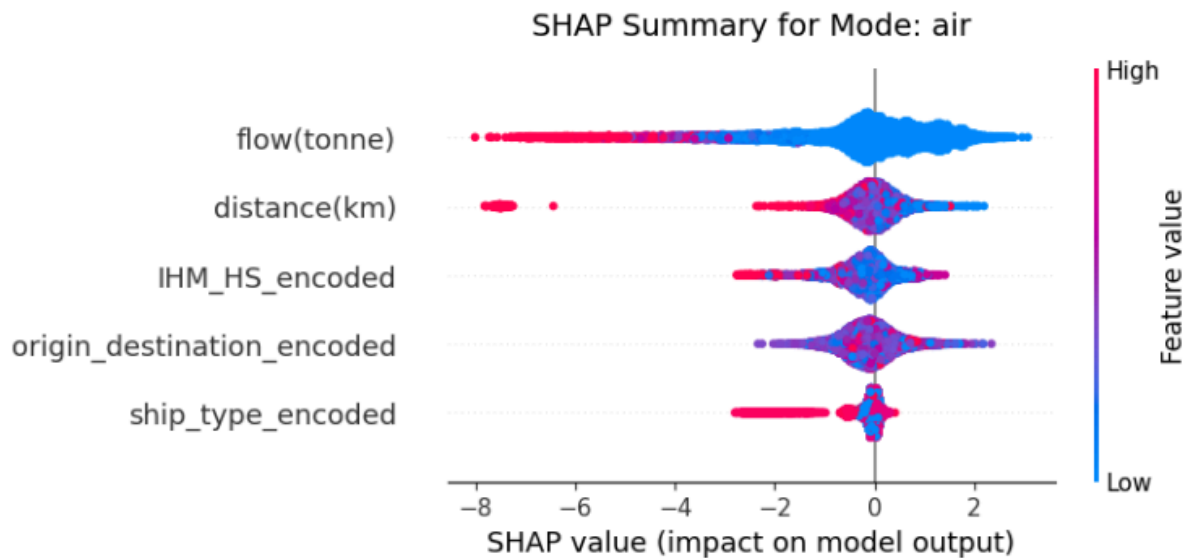
SHAP Summary for Mode: Air
Low shipment volumes (blue) are strongly associated with higher SHAP values, meaning they increase the likelihood of air mode selection—consistent with air being preferred for high-value, low-weight goods.

Distance and route encoding play a nuanced role in air transport selection, where longer distances and certain corridors push the model toward or away from predicting air.

SHAP Summary for Mode: rail
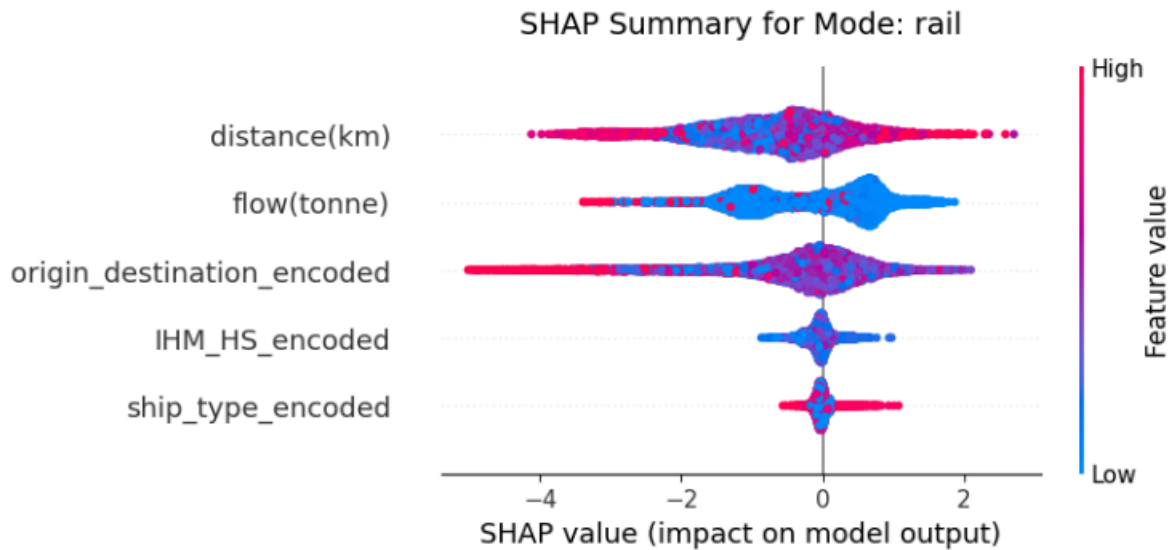
SHAP Summary for Mode: Rail
Longer distances and larger flows increase the likelihood of rail being selected, aligning with rail's comparative advantage in bulk, long-haul transport.

Certain origin–destination pairs show strong positive or negative influence, suggesting rail infrastructure and availability vary significantly by corridor.

SHAP Summary for Mode: road
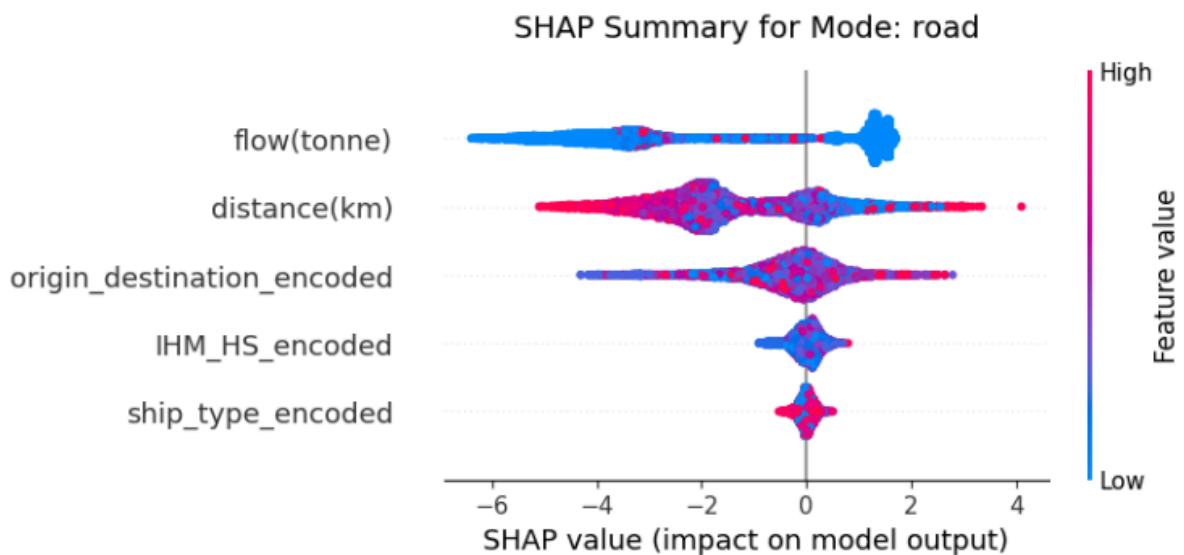
SHAP Summary for Mode: Road
Road is favored for smaller volumes and mid-range distances, as seen by negative SHAP values for high-volume or long-distance shipments.

Route and commodity-specific factors still impact road selection, indicating it remains a flexible mode for diverse flows in short to medium corridors.
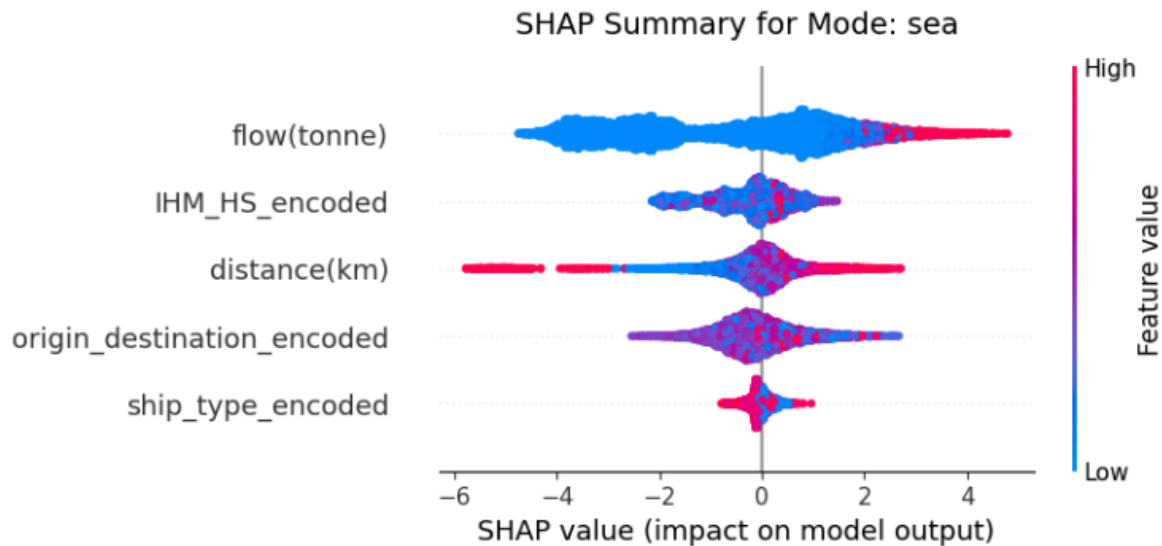

SHAP Summary for Mode: Sea
High-volume shipments (low flow values in blue) strongly drive sea transport predictions, confirming its role in bulk, cost-efficient global trade.

Commodity type and distance also influence sea mode selection, but to a lesser degree, likely due to sea routes being less sensitive to marginal distance changes compared to land modes.

3.D.B: Shipping Mode Classification,
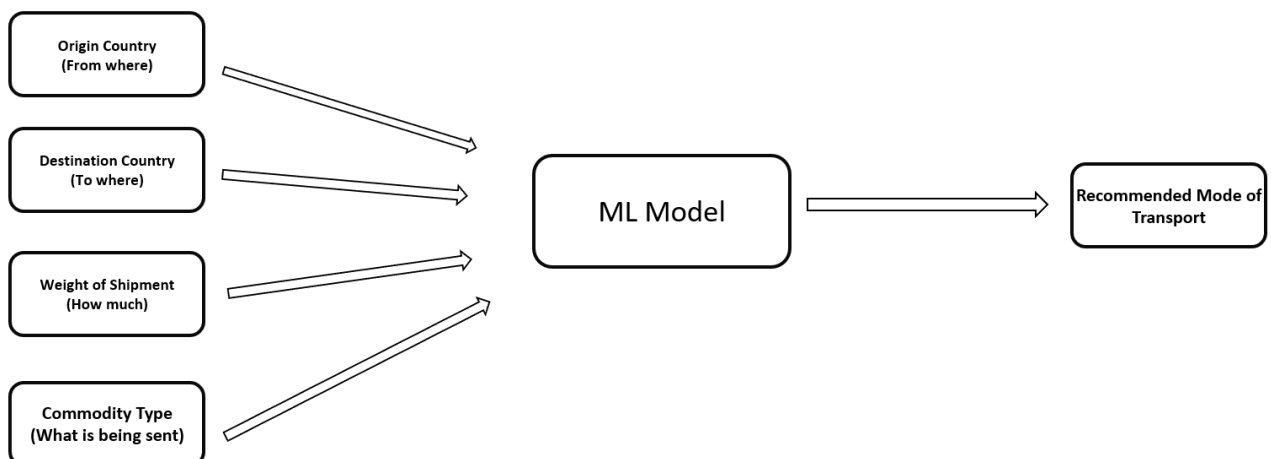
To predict the most likely transport mode (air, sea, road, rail) for a given trade flow, we trained a classification model using the cleaned World Bank logistics dataset.

Data Preparation

We first filtered the dataset to exclude rows with missing or zero values in key cost and flow fields. Outliers in unit logistics cost were removed using the IQR method to ensure model robustness. Next, geospatial distances between origin and destination countries were computed using the Haversine formula with country centroid coordinates. We also engineered a route identifier (origin_destination) and extracted the most common ship type per corridor.

Feature Engineering

Categorical fields like commodity (IFM_HS), ship type, and route ID were encoded using target encoding. The final feature set included:

● Physical attributes: flow volume, distance

● Encoded categorical variables: route, commodity type, and ship type

Model Training

We trained an XGBoost classifier with tuned hyperparameters (n_estimators=200, max_depth=6, learning_rate=0.05). The model was trained on 80% of the data with stratified sampling to preserve class balance.

Results

The model demonstrated strong classification performance across all four modes. Evaluation was done using precision, recall, F1-score, and a confusion matrix. The classifier is now able to reasonably predict the transport mode based on route, shipment characteristics, and commodity type—enabling data-driven insights into trade logistics patterns.

Mode	Precision	Recall	F1-Score	Support
Air	0.68	0.76	0.71	298,668
Rail	0.56	0.37	0.45	183,054
Road	0.45	0.61	0.52	66,547
Sea	0.85	0.85	0.85	628,159
Accuracy			**0.74**	1,176,428

Macro Avg	0.63	0.65	0.63	1,176,428
Weighted Avg	0.74	0.74	0.74	1,176,428

After which, a Gradio app was developed and deployed using HuggingFace. Below is the snapshot of the final app.

Link for the app: https://huggingface.co/spaces/kumarsauravjha/freight_mode_predictor

3.D.C: Predicting Transport Costs for African Exports
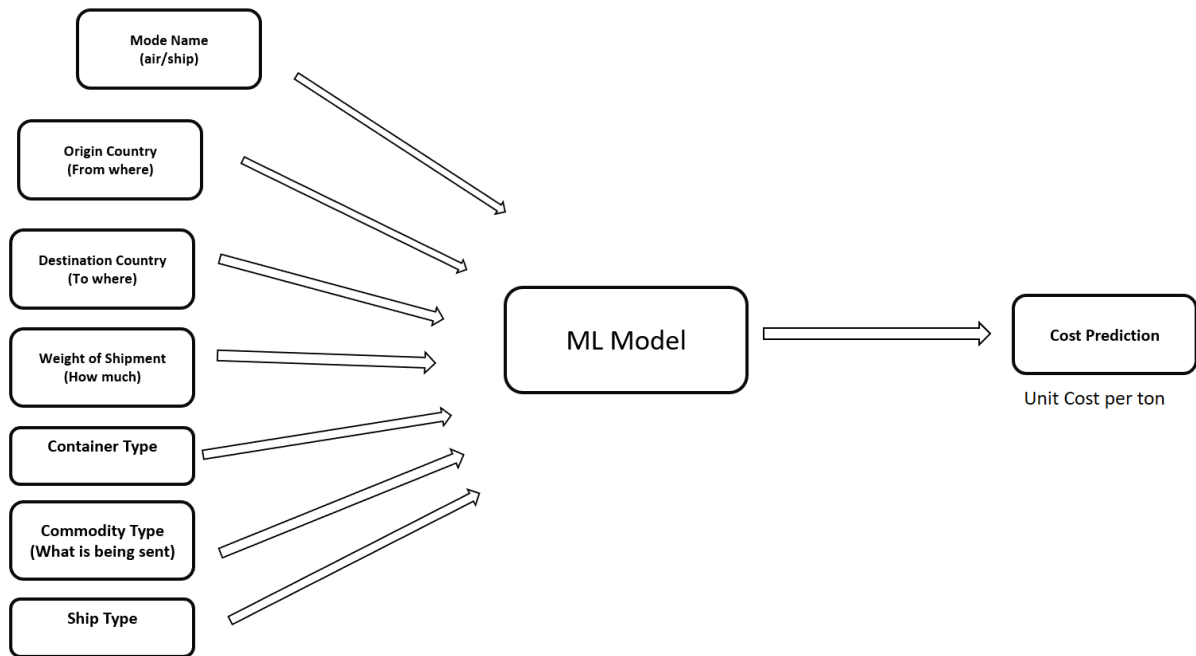
To estimate the unit logistics cost ($/ton) for African exports, we trained a regression model using a cleaned and enriched version of the World Bank transport dataset. The model enables cost benchmarking across countries, regions, and transport modes.

Data Preparation

The dataset was filtered to include only exports originating from 54 African countries in 2020. We excluded records with missing or invalid cost fields. Outliers in unit logistics cost were removed using the 1st–99th percentile range to improve model robustness. This removed 31,564 extreme cases from the full dataset.

We engineered a set of cost-relevant features such as flow volume, distance, transport mode, and commodity type. The final dataset for modeling included:

- **Training samples:** 1,238,051
- **Testing samples:** 309,513

Feature Engineering

The feature set included both numeric and categorical attributes:

Numeric features:

- `flow(tonne)`: Total volume of shipment
- `distance(km)`: GIS-based estimated shipping distance

Categorical features (encoded with target encoding):

- `Mode_name`: Transport mode (Air, Sea, Rail, Road, Other)
- `destination_ISO`: Country of destination
- `IFM_HS`: HS commodity code
- `ship_type`: Type of vessel or vehicle used
- `container_type`: Type of container used

All numeric features were standardized, and all categorical features were target-encoded to maintain performance in XGBoost.

Model Training

We used an **XGBoost regressor** inside a pipeline with preprocessing. The model was trained on 80% of the African exports dataset using default hyperparameters (with random_state=42 for reproducibility).

Results

The model demonstrated strong performance in predicting logistics cost per ton, indicating that flow volume, mode type, and route characteristics explain a substantial proportion of cost variability:

Metric	Value
RMSE	4,554.20
MAE	1,787.23
R² Score	0.8397

The R² score of 0.8397 suggests the model explains ~84% of the variance in export costs, making it a reliable tool for cost estimation and policy simulation.

3.D.D: World Map visualization
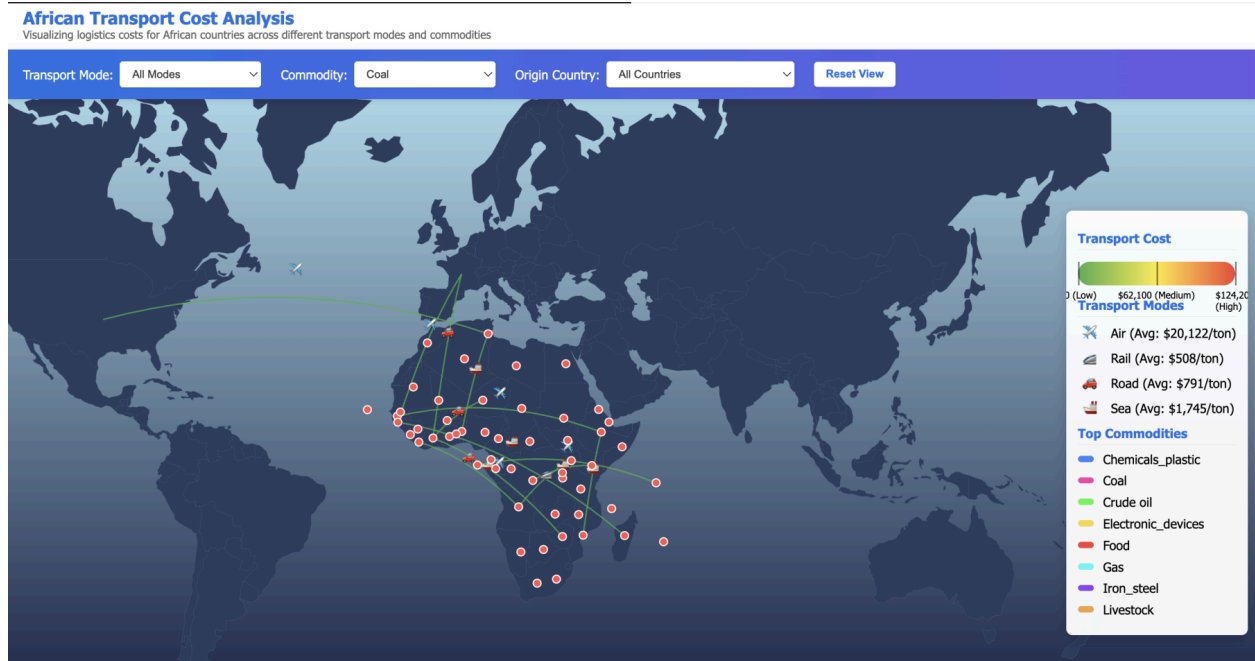
To better understand the spatial dynamics of African trade, we developed an interactive African Logistics Cost Map—a React-based web application that visualizes trade routes, transport modes, and logistics costs between African countries and their major global trading partners.

Key Features

- Color-coded trade routes: Routes are displayed as curved arcs, with color gradients indicating cost intensity—green (low), yellow (medium), and red (high).

- Origin and destination markers: African origin countries are highlighted with larger red dots, while destination countries appear as smaller blue dots.

- Interactive tooltips: Hovering over a route reveals detailed information, including:

  - Origin and destination countries
  - Country codes and names
  - Average transport cost per ton

- Filter controls: Users can filter the visualization by:

- - Transport mode
  - Commodity type
  - Origin or destination country

Technical Implementation

- Built using D3.js and TopoJSON with a Mercator projection centered on Africa.

- Powered by CSV-based logistics cost datasets.

- Includes responsive UI controls for zooming, panning, and filtering to enhance interactivity.

This platform is designed for policymakers, logistics planners, and trade analysts. It helps pinpoint high-cost corridors and inefficiencies, guiding targeted infrastructure investments and trade facilitation strategies across the African continent.

4. Results & Analysis

Our analysis of global transport costs reveals complex patterns that influence international trade efficiency. By examining transport mode distribution, cost structures, geographic factors, and commodity-specific trends, we've identified key insights that explain current inefficiencies and highlight opportunities for optimization.

4.1 Transport Mode Distribution and Cost Structure

4.1.1 Modal Preferences in Global Trade

Air and sea emerge as the dominant transport modes in international trade, followed by road and rail. This distribution reflects both practical limitations and economic considerations in global logistics systems. Air transport, despite having the highest cost per ton, is frequently utilized for high-value, time-sensitive goods. The prevalence of both air and sea transport indicates a bifurcation in global trade - a separation between high-value, time-critical shipments and bulk commodity movements.

Sea transport maintains nearly equal usage to air, demonstrating its crucial role in handling bulk commodities and containerized goods across long distances. The relatively lower utilization of road and rail transport in international trade can be attributed to infrastructure challenges, cross-border complications, and the geographical limitations of land-based transport for intercontinental movements.

Relative Cost Comparison by Transport Mode
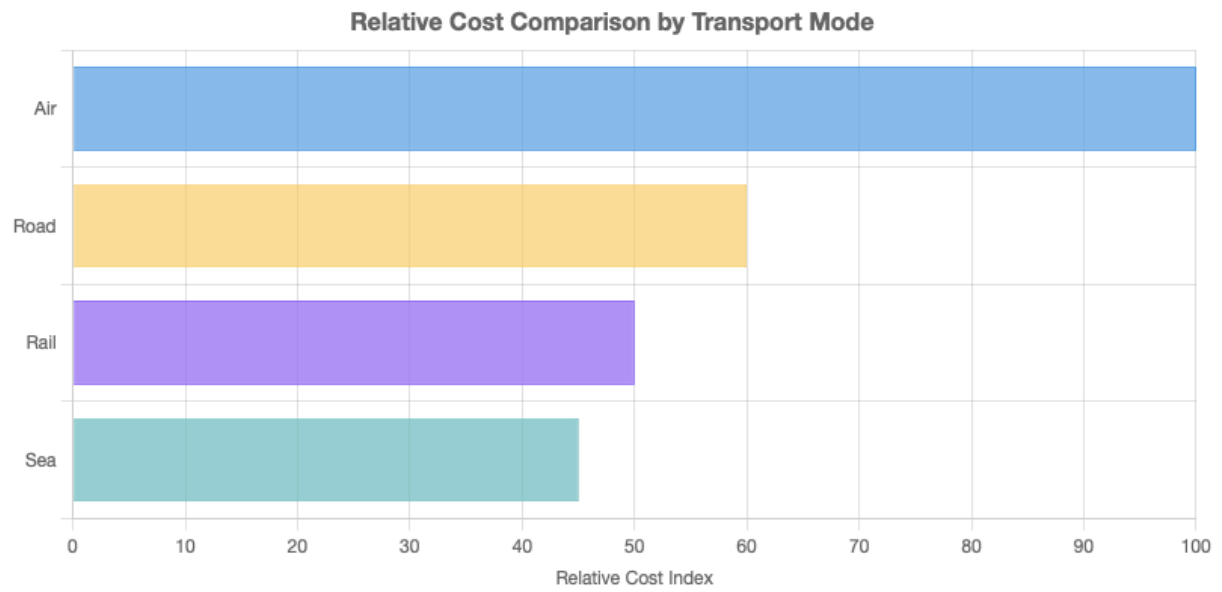
Source: Based on World Bank data analysis on international logistics costs

Distribution of Transport Modes
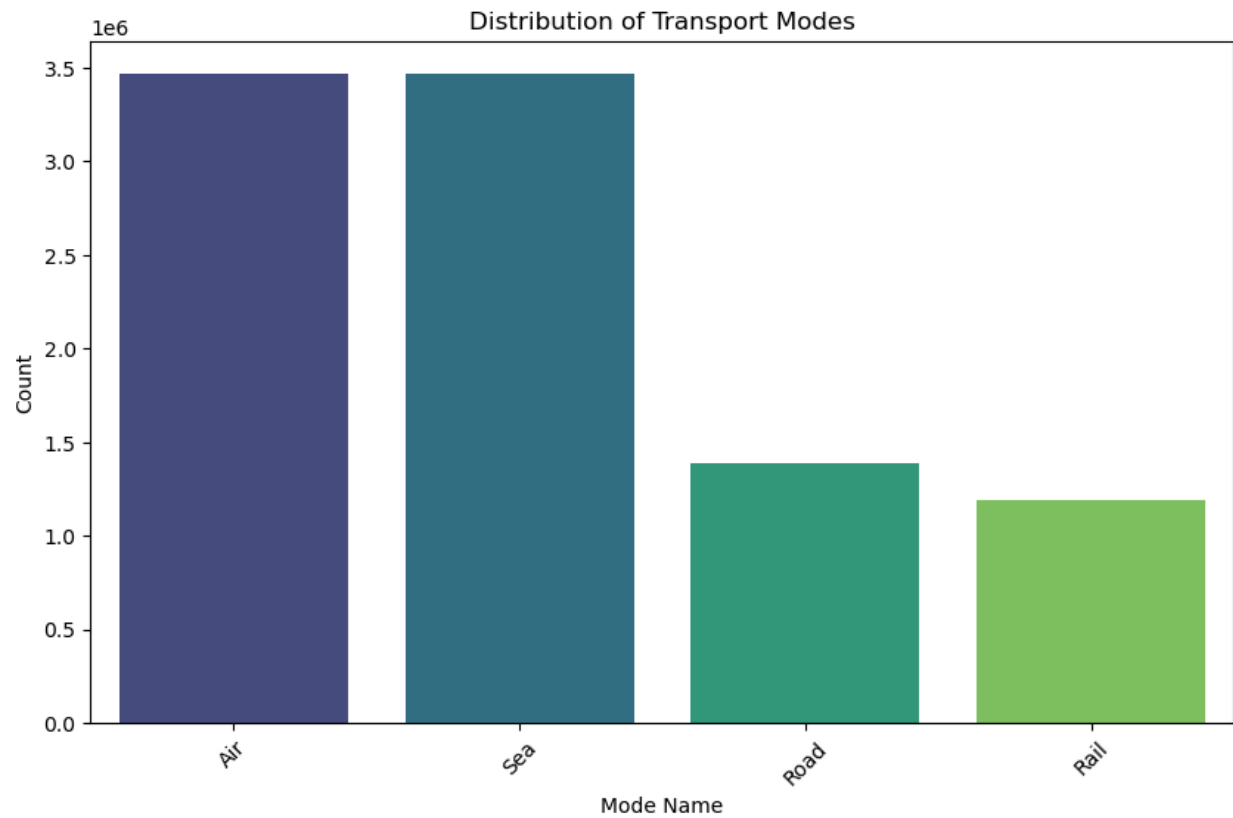
4.1.2 Cost Disparities Between Transport Modes

Our cost analysis reveals dramatic differences in logistics expenses across transport modes:

- **Air transport** demonstrates exceptional cost variability, with outliers exceeding $400,000 per ton and disproportionately high freight rates relative to cargo value. This extreme range reflects the premium placed on speed, security, and specialized handling requirements for high-value goods.
- **Rail transport** consistently emerges as the most cost-effective option on a per-ton basis, despite its limited use. This cost advantage suggests significant untapped potential for cost savings through greater rail utilization where infrastructure permits.
- **Road transport** exhibits moderate costs but with notable inefficiencies, particularly at mid-range distances. Cost per ton-kilometer remains higher than rail for longer distances, likely due to fuel consumption, labor costs, and tolls.
- **Sea transport** proves most cost-efficient for long-distance shipping, with relatively stable per ton-kilometer costs for extended journeys. Its efficiency makes it the backbone of global bulk trade despite slower transit times.

These cost structures create clear transport mode recommendations for different scenarios. Bulk carriers and tankers offer the most cost-effective solutions for intercontinental shipments of industrial materials and liquids, while container ships provide efficient service for general trade goods. For shorter distances, specialized options like RoRo vessels (for vehicles) and rail (for bulk commodities) present optimal choices in their respective niches.

Transport Costs by Mode
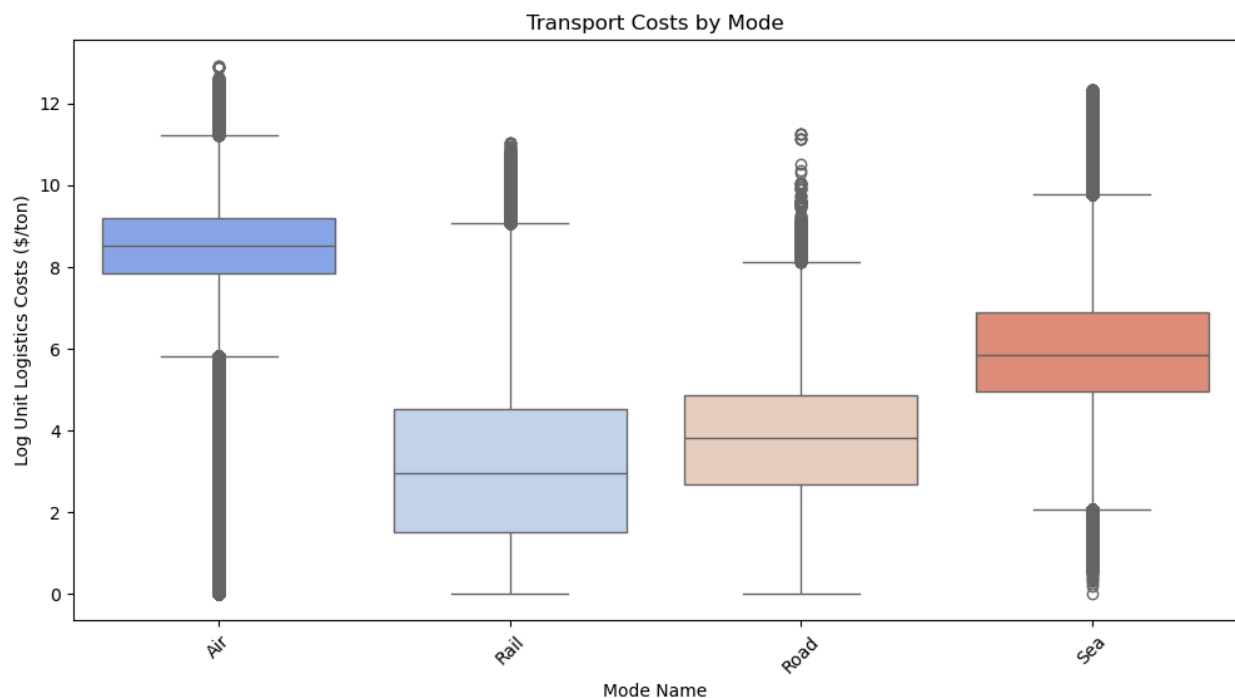
4.2 The Relationship Between Distance, Volume, and Cost

4.2.1 Distance-Based Transport Efficiency
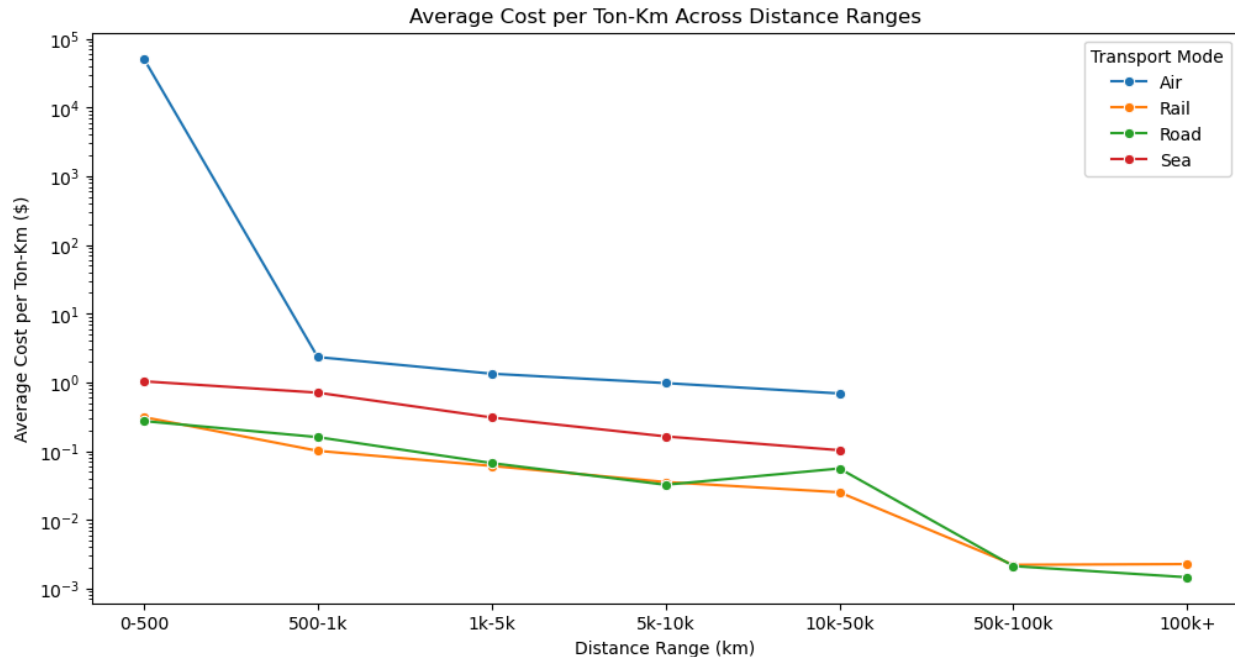
Correlation between average distance and average cost: 0.2645. Our analysis reveals nuanced relationships between distance and transport efficiency:

- **Short-distance transport** (0-500 km) shows surprisingly high cost variability across all modes, with extreme inefficiency for air transport. The most dramatic example appears in ultra-short air shipments between Palestine and Israel, where costs exceed $9.6 million per ton-kilometer despite minimal distance (0.0023 km), likely driven by security concerns, administrative barriers, or emergency delivery requirements.
- **Medium-distance transport** (1,000-5,000 km) begins to demonstrate economies of scale, particularly for sea and rail options, though air and road transport still face significant inefficiencies in this range.
- **Long-distance transport** (10,000+ km) shows cost stabilization, especially for sea and rail modes. This pattern confirms that fixed costs become increasingly distributed over distance, reducing per-unit expenses.

The data supports a clear principle: the longer the distance, the more efficient transport generally becomes on a per-kilometer basis. This efficiency gain is most pronounced for rail and sea transport, while air transport remains costly regardless of distance.

Overseas Trades (flow > 0)
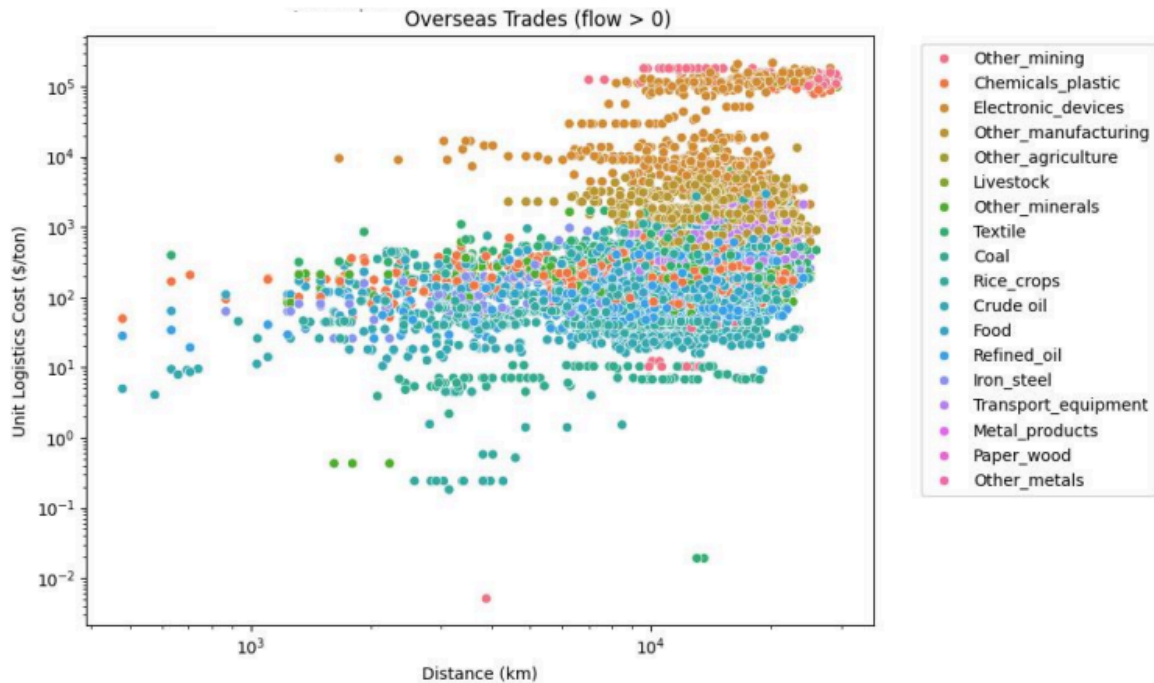
- **Logistics costs increase with distance**, but the spread varies significantly by commodity—**high-value goods like electronics and chemicals** tend to cluster at the upper end of the cost range, especially at longer distances.

- **Commodities like crude oil, food, and refined oil** show tighter, lower-cost clusters, reflecting **greater efficiency and consistency** in their long-haul overseas logistics chains.

4.2.2 Volume-Cost Relationship

Contrary to theoretical expectations, our analysis shows only a weak correlation between shipment volume and transport costs per ton-kilometer. While higher tonnage should theoretically reduce costs through economies of scale and bulk transport efficiencies, other factors appear to exert stronger influence.

The mode of transport emerges as the dominant determinant of cost variability. Air transport remains expensive even for large shipments, while rail and sea transport offer significant cost advantages even for smaller loads. This pattern is further reinforced by commodity characteristics, with high-value, low-weight goods (like electronics) shipped by air showing inflated per ton-kilometer costs, while heavy bulk goods (such as coal or iron ore) demonstrate cost efficiency via sea or rail regardless of individual shipment size.

The surprisingly weak volume-cost correlation suggests that operational bottlenecks, infrastructure limitations, and market structures may outweigh pure scale economies in determining final logistics costs.

Correlation Matrix
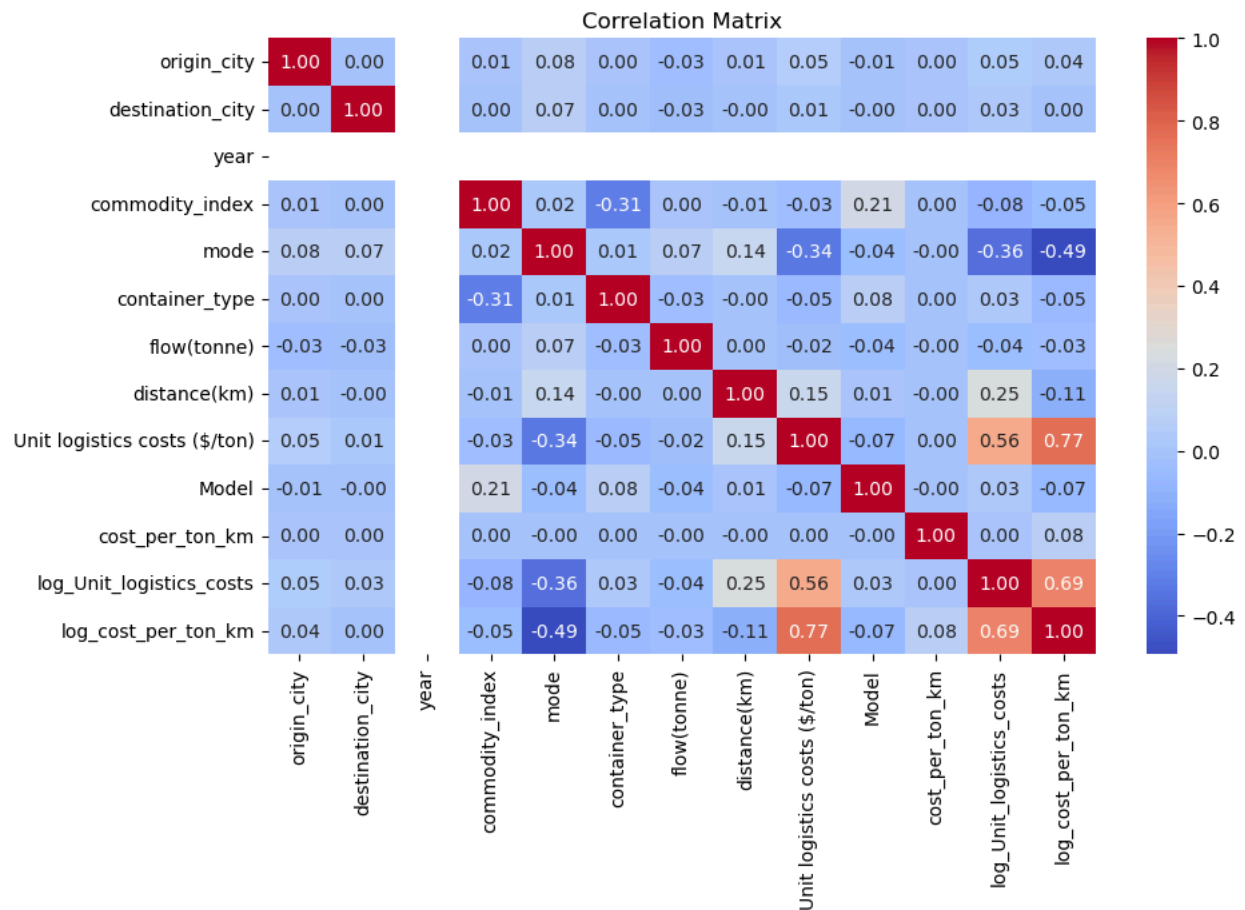
4.3 Geographic and Country-Specific Analysis

4.3.1 Landlocked vs. Coastal vs Island Countries

Transport Costs for Landlocked vs Coastal Countries in Africa
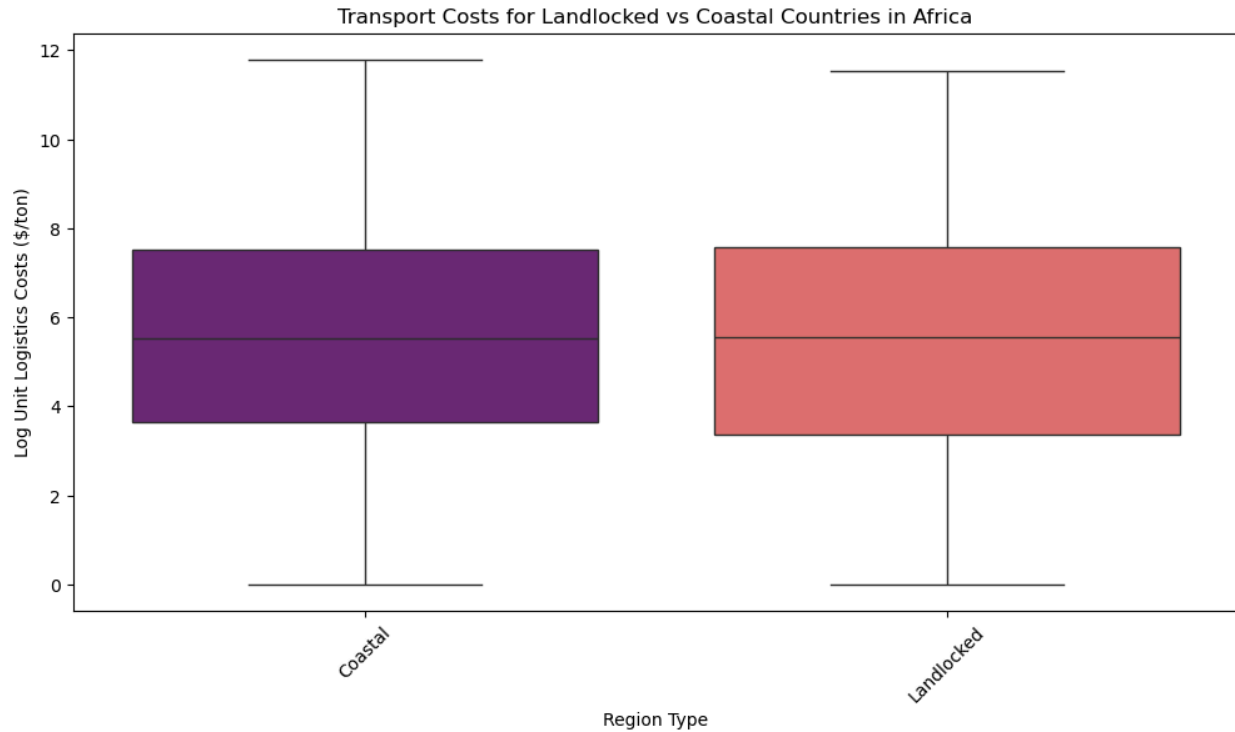
Our analysis confirms that landlocked countries generally face higher transport costs than coastal nations, though with interesting nuances:

- The median transport cost for landlocked countries exceeds that of coastal countries, reflecting the additional complexity and transfers required to reach international markets.
- Transport costs in landlocked regions show greater variability, suggesting inconsistent pricing that may result from monopolistic logistics services or infrastructure limitations.
- Extreme cost outliers exist for both landlocked and coastal regions, indicating that geography alone doesn't determine logistics efficiency.

Distribution of Normal Shipments by Origin Type

Average Logistics Cost by Mode, Origin Type, and Trade Type
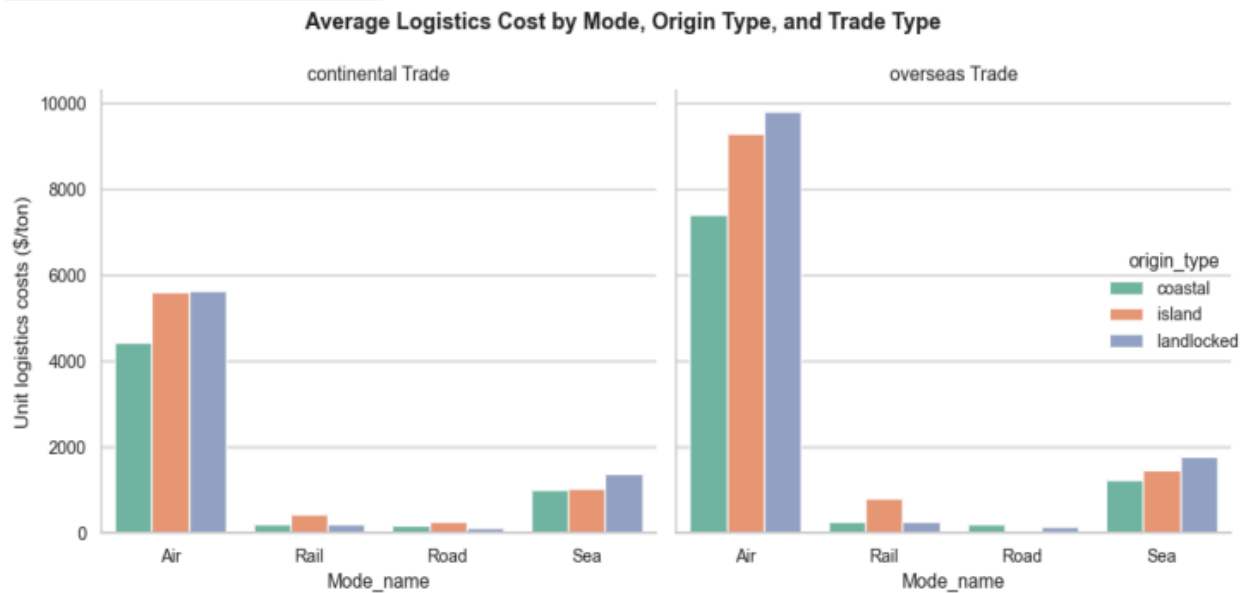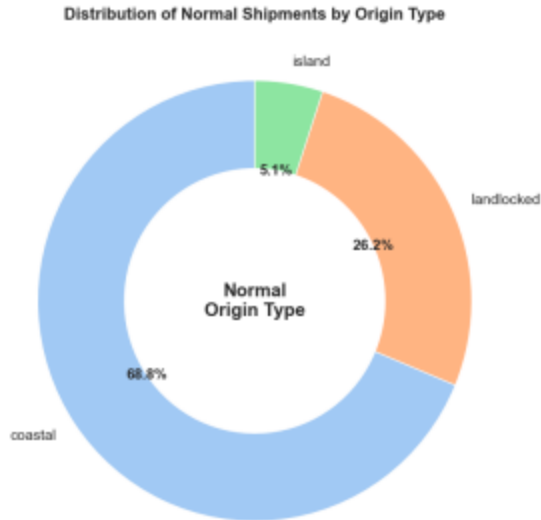
Logistics Cost Comparison by Origin Type and Trade Route

To understand how geography affects logistics performance, we analyzed shipment patterns and costs across three origin types: **coastal**, **landlocked**, and **island** countries.

Distribution of Normal Shipments
As shown in the donut chart:

- **Coastal countries** account for the majority of trade, contributing **68.8%** of normal shipments.

- **Landlocked countries** follow with **26.2%**, while **island states** contribute only **5.1%**. This skew underscores the physical accessibility challenges faced by non-coastal nations.

Cost Comparison by Mode and Trade Type

The bar plots break down **average unit logistics costs** across origin types and two trade contexts:

- **Continental Trade**: Landlocked countries already pay a premium, especially for **sea** and **air** modes.

- **Overseas Trade**: The cost gap widens, with landlocked exporters facing the highest **air transport costs** (~$10,000/ton) due to additional inland and port-handling burdens.

These insights highlight how geography, particularly lack of direct sea access, structurally inflates logistics costs—emphasizing the need for regional connectivity and multimodal infrastructure.

Surprisingly, some landlocked countries (e.g., Botswana, Eswatini) demonstrate lower transport costs than certain coastal nations. This counter-intuitive finding suggests that strong trade agreements, efficient border crossings, and well-managed logistics corridors can overcome geographical disadvantages.

4.3.2 Country-Specific Cost Patterns

Our country-level analysis reveals several distinct patterns:

- **Island nations** like Mauritius, Comoros, and Seychelles face exceptionally high transport costs, reflecting their dependence on long-distance maritime and air connections.
- **Specific African countries** including Sudan, Mozambique, Madagascar, DR Congo, and Zambia show elevated costs across multiple transport modes, suggesting systemic challenges in infrastructure, fuel pricing, or logistics efficiency.
- **Comoros** demonstrates an anomalously high cost for air transport, unusual for a small island nation with relatively short domestic distances, potentially indicating limited air cargo capacity or reliance on expensive charter flights.
- **Infrastructure gaps** appear in countries like Gambia and Lesotho, where absence of certain transport modes (road, rail) points to potential investment opportunities.

The most extreme cost example in our dataset involves routes to American Samoa (ASM), with Saint Helena to ASM representing the single most expensive route at $43,692 per ton. This extraordinary cost reflects multiple compounding challenges: remote origin, remote destination, low shipping volume, limited port infrastructure, and likely multiple transshipments.

4.4 Commodity and Container Type Analysis

4.4.1 Commodity-Specific Transport Patterns

Transport costs and mode preferences vary substantially by commodity type:

- **Chemicals, electronics, and pharmaceuticals** incur significantly higher transport costs, particularly when shipped by air, reflecting their high value, fragility, or hazardous nature.
- **Food, livestock, and minerals** face elevated costs when transported by road, suggesting inefficiencies in these crucial supply chains.
- **Sea transport** demonstrates the most consistent costs across commodity types, highlighting its versatility and reliability for diverse cargo.

These patterns indicate that transport optimization strategies should be tailored to product characteristics rather than applying uniform approaches across all cargo types.

4.4.2 Container and Vessel Type Optimization

Our analysis of container and vessel types reveals clear optimization opportunities:

- **Bulk carriers** offer maximum efficiency for high-volume, low-value industrial materials like coal, iron ore, and grains on intercontinental routes.
- **Container ships** provide the optimal balance for general trade involving retail goods, electronics, and manufactured products due to their standardization and well-connected global infrastructure.
- **RoRo vessels** excel at vehicle transport over short to medium distances, with their built-in ramps and multiple internal decks maximizing cargo storage efficiency.
- **Tankers** maintain the lowest cost per ton-kilometer for liquid transport, making them the obvious choice for oil and gas movements.

The data shows clear modal separation in shipping patterns, with petroleum products never utilizing air transport due to volume and safety considerations. However, we observed the unusual case of RoRo configuration in air transport, likely representing specialized scenarios like Formula 1 car shipments or situations where sea infrastructure is inadequate.

Unit Logistics Costs by Ship Type

Heatmap: Ship Type vs. Transport Mode
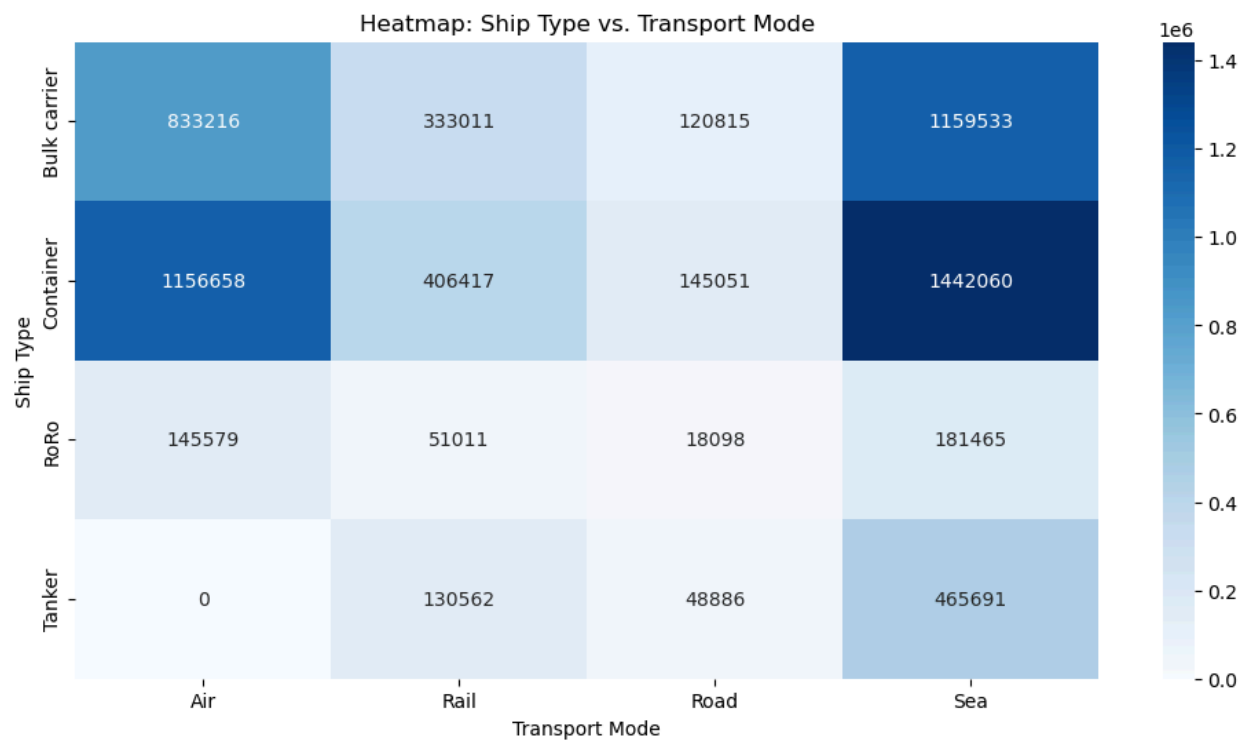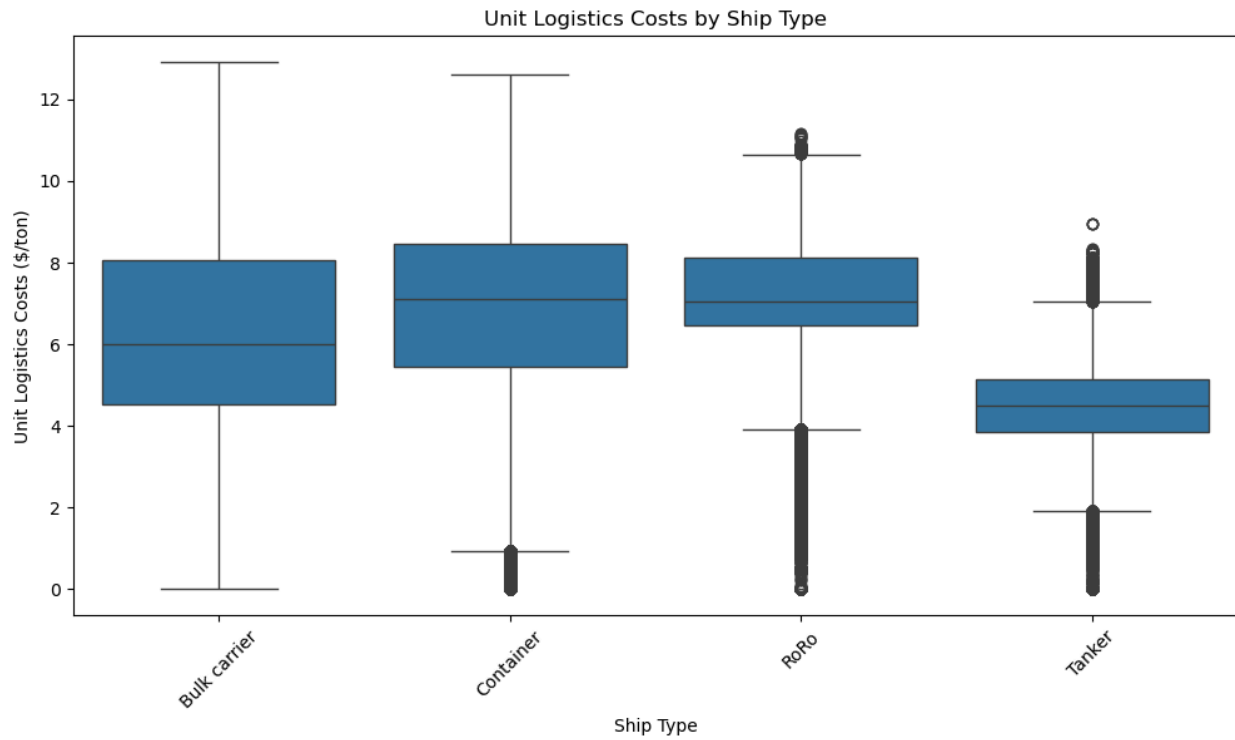
4.5 Statistical Validation and Principal Component Analysis

Our statistical analysis confirms the significance of relationships between transport variables:

- ANOVA tests reveal that transport mode has a significant impact on both flow volume and distance traveled.
- Chi-square tests demonstrate strong dependencies between mode choice and both commodity type and container configuration.

	Statistical Test	p-value	Interpretation
1	ANOVA Test (Mode_name vs. Flow (tonne))	0.0	Significant result (Reject Null Hypothesis)
2	ANOVA Test (Mode_name vs. Distance (km))	0.0	Significant result (Reject Null Hypothesis)
3	Chi-Square Test (Mode_name vs. Commodity Index)	0.0	Significant result (Reject Null Hypothesis)
4	Chi-Square Test (Mode_name vs. Container Type)	0.0	Significant result (Reject Null Hypothesis)

These statistical validations reinforce our observation that transport mode selection follows logical patterns rather than random distribution, driven by a combination of economic, physical, and infrastructural factors.

Principal Component Analysis (PCA) further reveals clear separations between transport modes, confirming that transport mode strongly influences cost and logistics patterns. This multivariate technique demonstrates that the observed patterns are not merely coincidental but reflect fundamental structural differences in how various modes operate within the global logistics system.

PCA Analysis of Transport Data (Color by Mode_name)
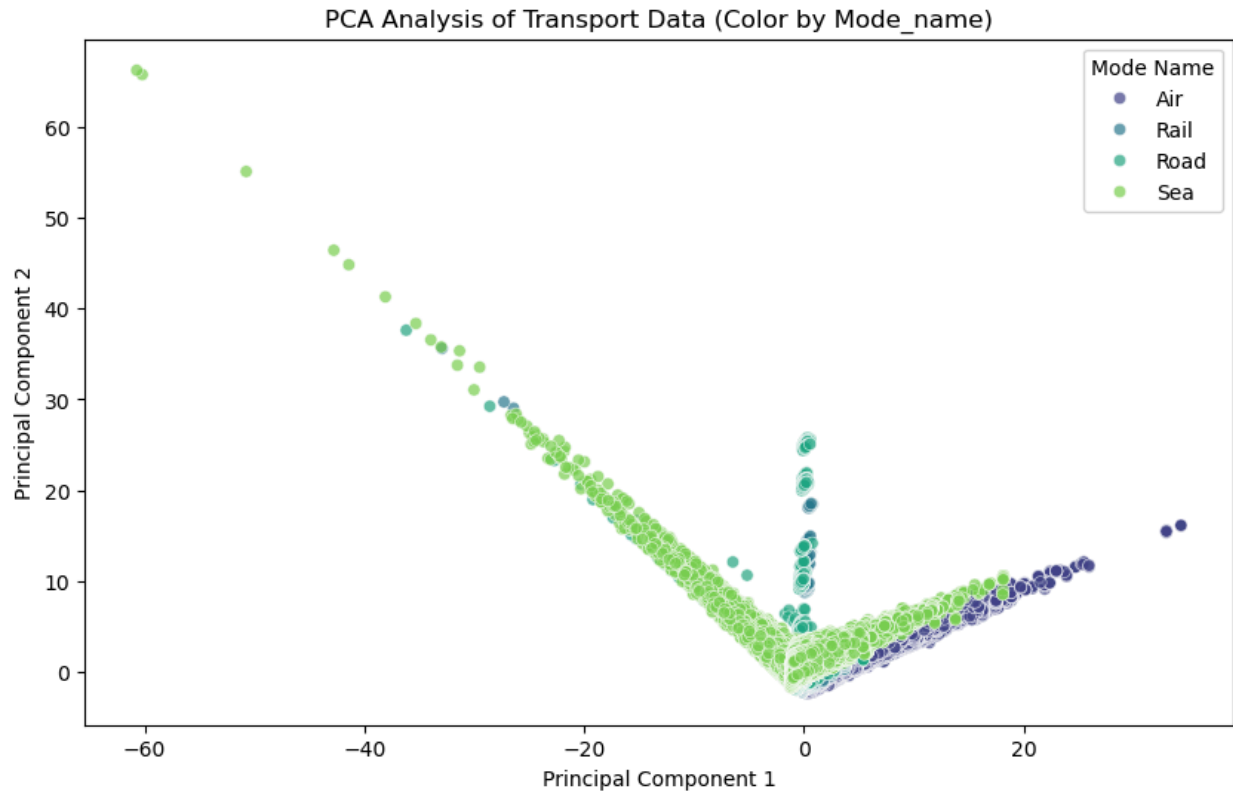
4.8 Country-Level Cost Analysis

Our detailed country-level analysis reveals important systemic patterns in global logistics costs that have significant implications for policy development.

Transport Costs by Mode & Country
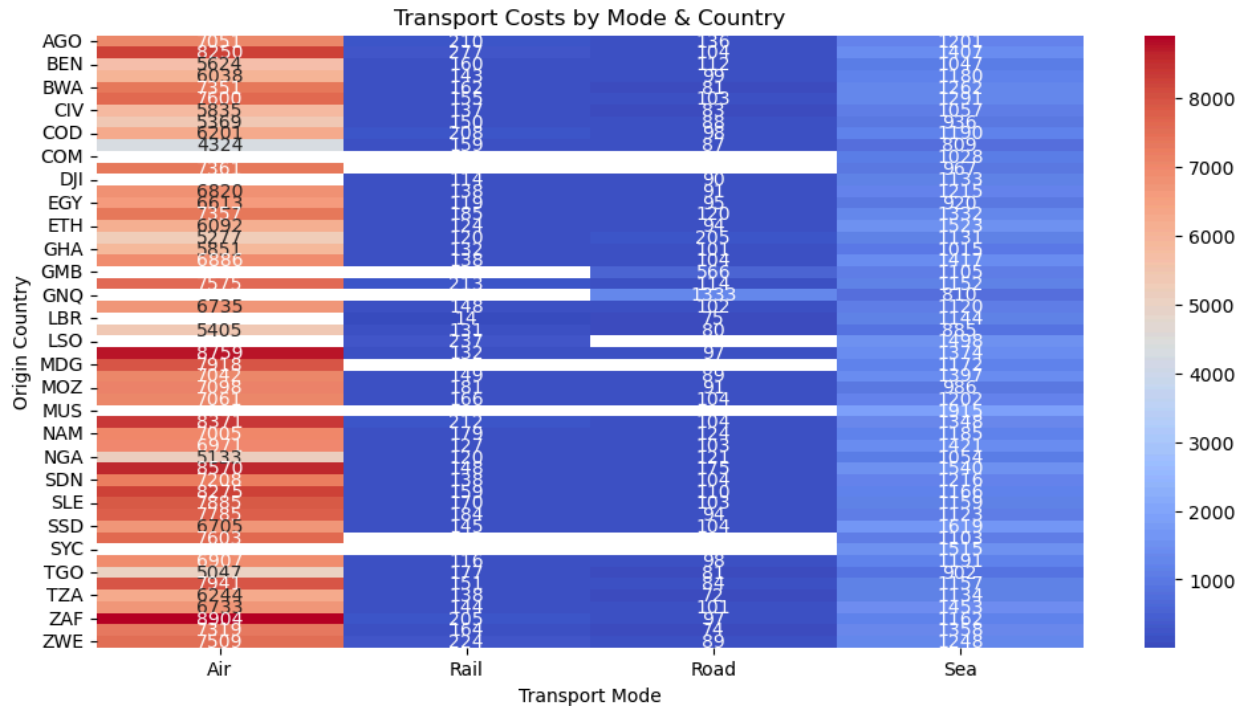
4.8.1 Island Nations and Remote Locations

Small island nations in the Pacific demonstrate extraordinarily high logistics costs, with average expenses ranging from $23,000-27,000 per ton—7-8 times higher than the lowest-cost countries. This extreme cost premium stems from multiple compounding factors:

- Remote geographic locations necessitating longer shipping distances
- Small economies that cannot generate sufficient volume for economies of scale
- Limited transportation infrastructure and connectivity options
- Heavy reliance on air freight and container shipping rather than bulk shipping

The modal analysis for these high-cost countries reveals a striking pattern: exactly 50% air transport and 50% sea transport usage, with zero rail or road utilization. While this aligns with their island geography, the precise 50-50 split suggests a possible dataset characteristic rather than natural trade patterns. This modal constraint represents a structural disadvantage that largely explains their position as the highest-cost exporters in global trade.

4.8.2 Modal Flexibility and Cost Optimization

Countries with the lowest logistics costs (UAE, Belgium, China) demonstrate balanced utilization across all four transport modes, with distributions of approximately 31% air, 18% rail, 19% road, and 31% sea. This multi-modal flexibility provides significant cost advantages, allowing shippers to select the most efficient mode for each specific cargo type and distance.
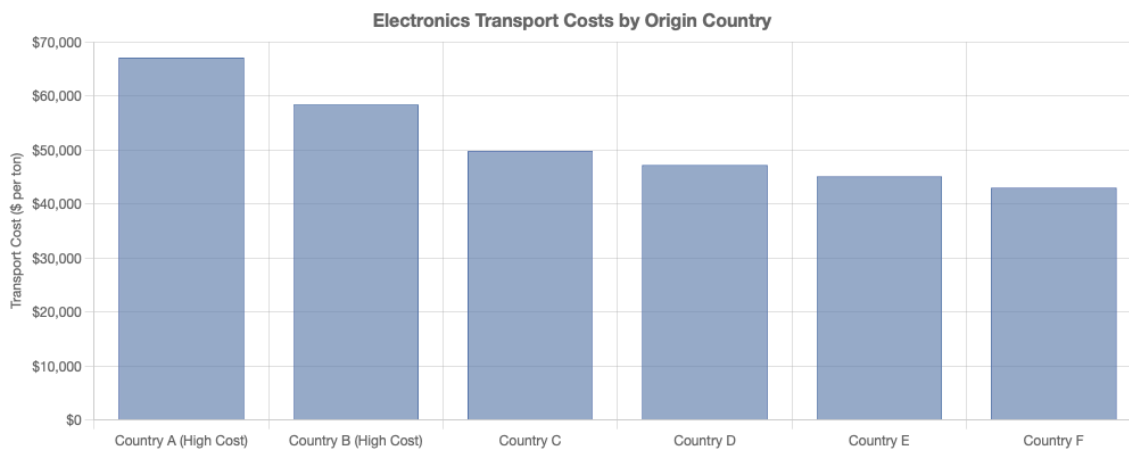
The data provides compelling evidence that modal flexibility significantly reduces overall logistics costs. Countries limited to fewer transport options consistently demonstrate higher average costs, highlighting the importance of developing diverse transport infrastructure rather than concentrating investment in a single mode.

4.8.3 Commodity Impact on Transport Costs

Our commodity analysis reveals that electronic devices generate the highest transport costs, ranging from $49,800 to $67,100 per ton across high-cost countries. This consistent premium occurs because:

- Electronic devices typically have high value-to-weight ratios
- High-value items require air freight for security and speed
- Insurance costs increase with commodity value
- Special handling requirements add to logistics expenses

Interestingly, the cost premium for electronics displays remarkable consistency, with transport costs ranging from $43,000-$50,000 per ton across different origin countries. This suggests standardized air freight pricing models for high-value electronics regardless of origin. The price difference between electronics and the second-most expensive commodity (typically mining products) consistently falls in the 30-50% range.

30-50%	$49.8K-$67.1K	$43K-$50K
Price premium for electronics over second-most expensive commodity	Transport cost range for electronics in high-cost countries (per ton)	Consistent cost range across different origin countries (per ton)

Transport Cost Comparison by Commodity Type

Electronic Devices
Average Transport Cost ($ per ton): $55,000

Electronic Devices · Mining Products · Refined Oil · Chemicals/Plastics · Machinery · Textiles · Agricultural Products

Electronics Transport Costs by Origin Country

Country A (High Cost) · Country B (High Cost) · Country C · Country D · Country E · Country F

4.9 Income-Based Classification of African Export Flows

In this section, we analyze how logistics patterns vary across African countries segmented by World Bank income groups: Low, Lower-middle, Upper-middle, and High income. This classification helps us understand whether wealthier countries are better positioned in terms of logistics cost, trade diversity, and mode of transport.
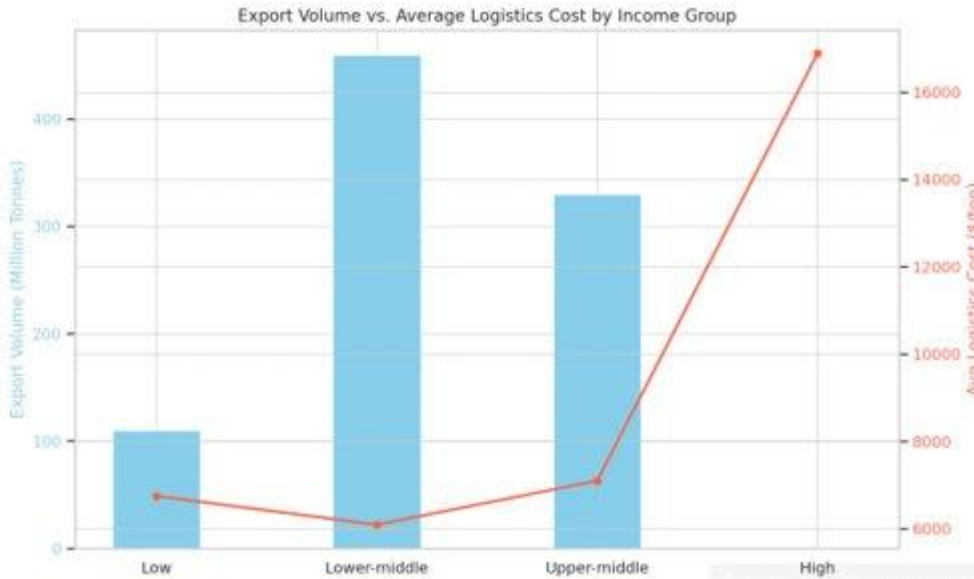
Graph 1: Export Volume vs. Average Unit Logistics Cost by Income Group

- This graph compares total export volume (in million tonnes) with average logistics cost per ton for each income group.
- Key Insight: One might expect that high-income countries would have the lowest average logistics costs, due to better infrastructure and trade facilitation. However, in this case, high-income countries show the *highest* unit logistics cost.
- Interpretation: This counterintuitive pattern arises because Seychelles is the only high-income country in Africa's dataset, and it is a small island nation. As a result, all exports are handled via air or sea—modes typically associated with higher per-ton costs.

- Conclusion: Income level alone does not predict logistics cost; geography and accessibility play a significant role.
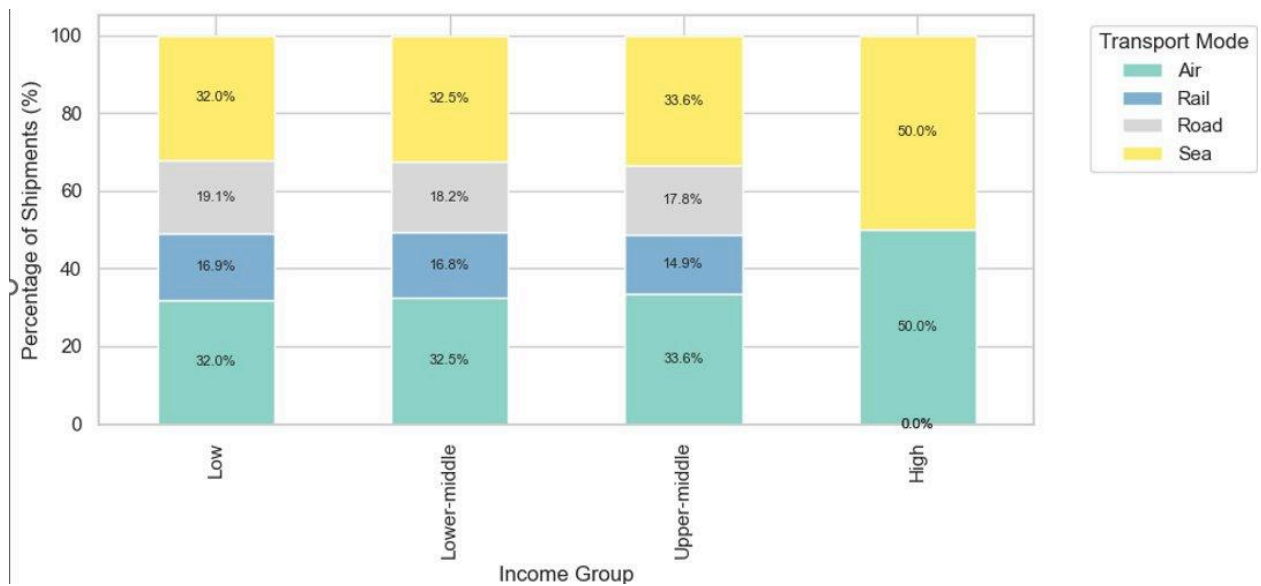
Graph 2: Mode of Transport Distribution by Income Group (Stacked % Bar)

- This graph displays the proportion of total export shipments that used each mode of transport: air, road, rail, or sea, grouped by income level.
- Key Insight: Lower and middle-income countries show greater mode diversity, including road and rail. In contrast, high-income countries (again, just Seychelles) use only air and sea.
- Interpretation: This shows that multimodal connectivity is more limited in island nations, which depend heavily on ports and airports.
- Policy Note: Expanding access to affordable land transport (like regional rail) could help reduce costs for landlocked low-income countries.
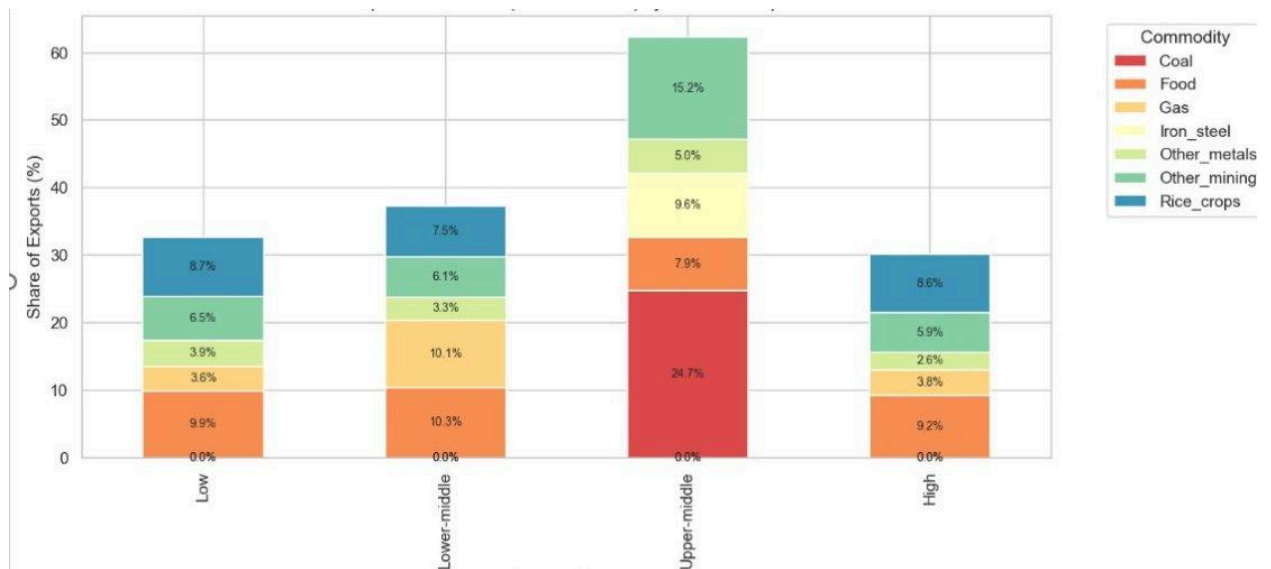
Graph 3: Top 5 Commodities per Income Group (Stacked % Bar)

- This visual shows the top five exported commodities by share of total exports for each income group.
- Key Insight: From low to upper-middle income groups, we see increasing diversity in export commodities. However, this trend reverses in the high-income group, which is again dominated by a narrow set of goods.
- Interpretation: The dominance of a few commodities in high-income group exports reflects Seychelles' specialized economic role, likely in transshipment or re-exports rather than diversified production.
- Conclusion: Export diversity grows with income—but only until a point. Geography and trade role (e.g., hub vs. producer) can distort the pattern.

To summarize, income level influences logistics patterns—but does not tell the full story. A high-income classification might suggest low cost and strong infrastructure, but factors like geographic isolation, mode availability, and trade roles (hub vs. producer) can override income-based assumptions.

> **Seychelles illustrates this perfectly**—as a small island high-income nation with the *highest* logistics cost, narrow commodity spread, and limited transport modes.

4.10 African Export Anomaly Analysis

Our focused analysis of African exports revealed several unusual patterns that provide valuable insights into unique challenges and opportunities across the continent.

4.10.1. Overview of the Anomalous data points

Origin Continent	Anomalous Shipments
Africa (AF)	17,657
Europe (EU)	11,596
Asia (AS)	11,435
Oceania (OC)	10,032
North America (NA)	9,133
South America (SA)	6,413

We used Multivariate anomaly detection using Isolation Forest and below are the findings.

1. By Mode of Transport

- **Air dominates anomalies (~85%)**, driven by high-value, low-weight shipments where even minor errors or pricing quirks lead to large unit-cost distortions.

- Volatility is especially high for **perishable or compact goods**, making air freight prone to misclassification and cost anomalies.

2. By Continent of Origin

- **Africa leads** with 17,657 anomalies, followed by Europe and Asia.

- Reflects reliance on expensive modes (like air), sparse overland infrastructure, and inconsistent customs reporting in certain regions

Top 10 Anomalous Commodities from Africa
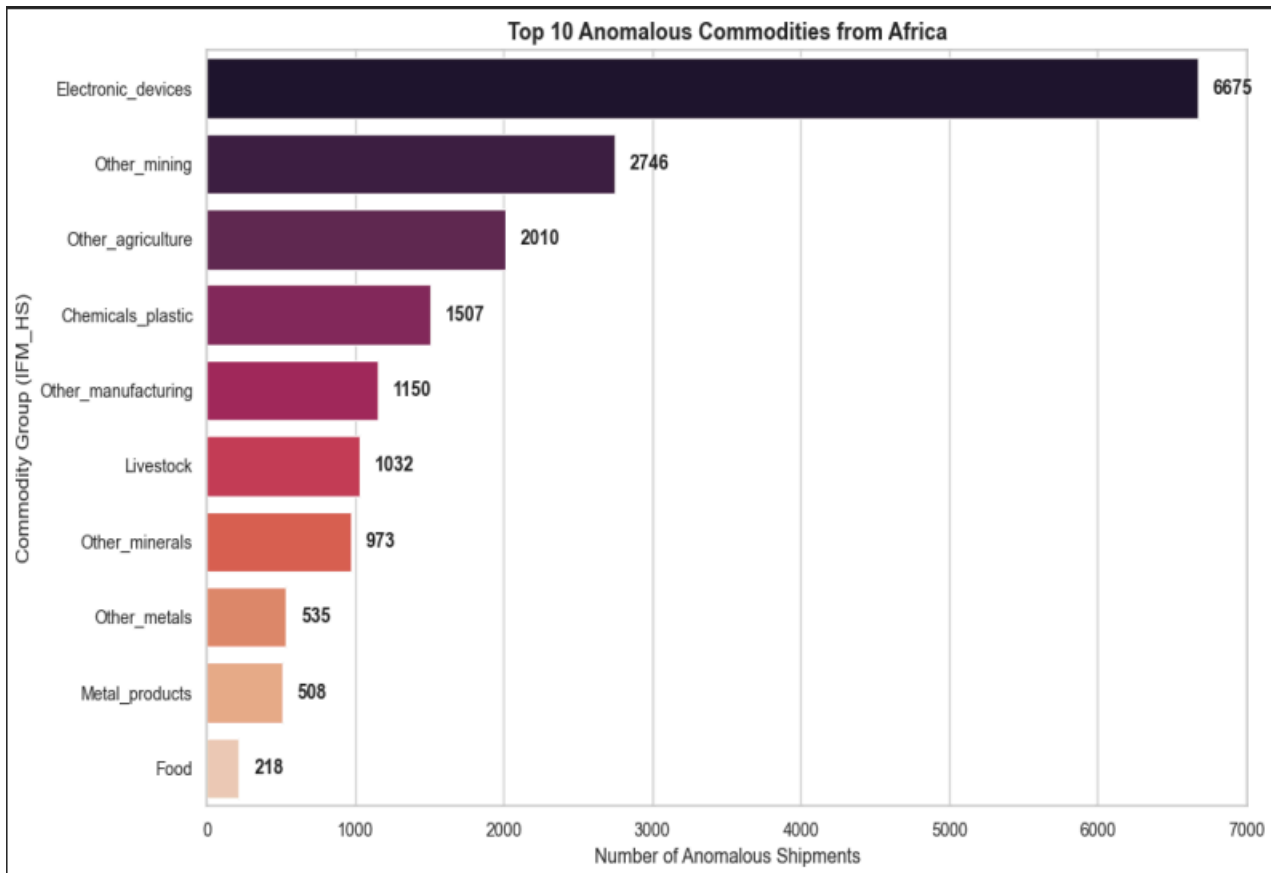
3. Top Anomalous Commodities (Africa)

- **Electronics (6,675 anomalies)**: Compact, high-value—small reporting errors inflate cost metrics.

- **Mining (2,746)** and **Agriculture (2,010)**: May involve misreported weights or non-standard routing.

- **Others (e.g., chemicals, livestock, plastics)**: Suggest mix of data-entry issues and time-sensitive shipments driving up costs.

Top 20 Anomalous Corridors (Count of Anomalies)
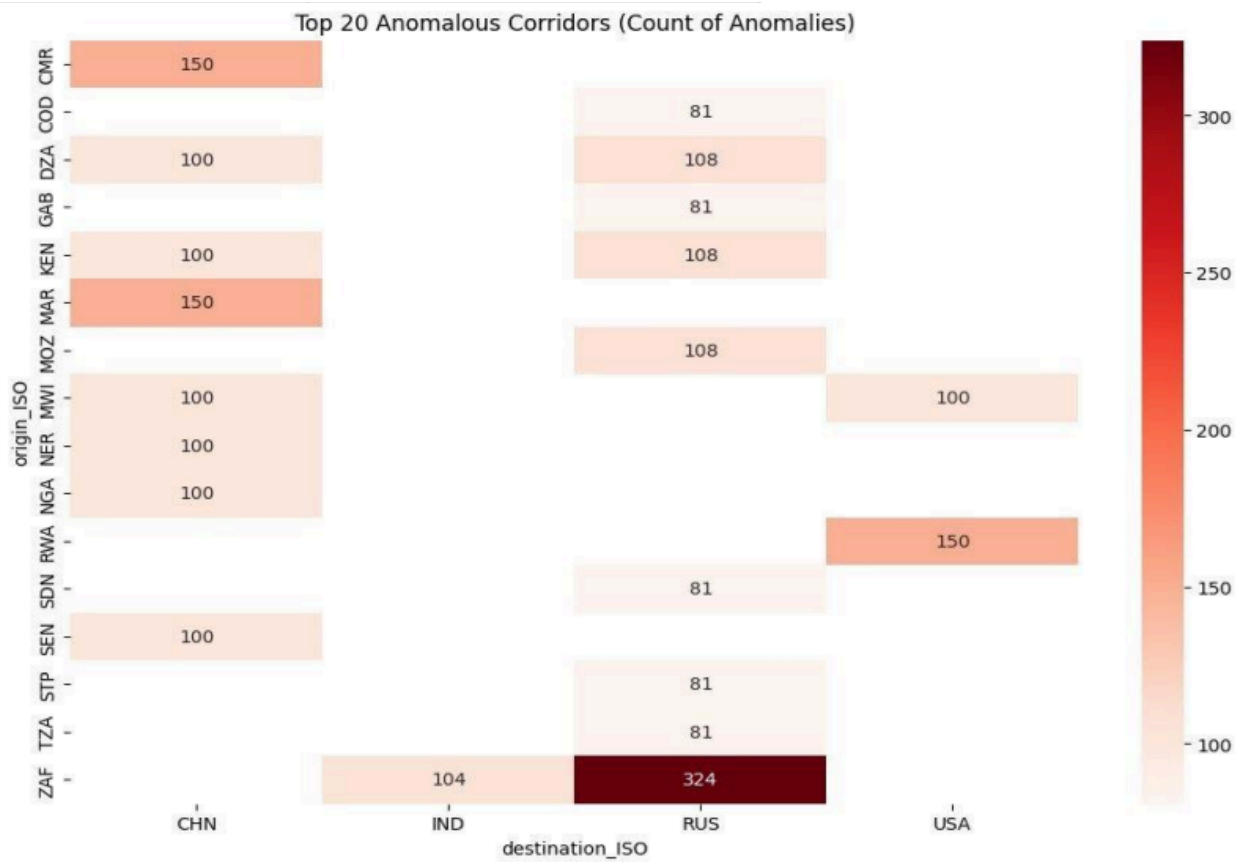
4. Anomalous Corridors

- **South Africa → Russia (ZAF–RUS)**: Highest anomaly count (324), likely due to long routes and air freight surcharges.

- Other flagged corridors (e.g., **Morocco → Kenya**, **Gabon → Algeria**) show high anomalies despite low traffic, indicating that outliers stand out more in sparse flows.

    Consistently anomalous corridors may signal **systemic data-quality issues** or **real-world inefficiencies**—warranting deeper validation and policy attention.

4.10.2 Modal Distribution Patterns and Geographic Constraints

African island nations (Mauritius, Comoros, Seychelles) demonstrate the same 50-50 split between air and sea transport observed in other island regions, reinforcing their structural disadvantage of being limited to the two most expensive transport modes.

However, Equatorial Guinea stands out with substantial road usage (23.83%) despite its geographically isolated position. This unusual pattern suggests a strategic advantage: Equatorial Guinea appears to be leveraging mainland connections through neighboring countries (likely Cameroon), creating a unique hybrid logistics model that continental islands like

Madagascar don't employ. This approach offers a potential model for other coastal nations seeking to optimize their modal mix.

Perhaps most surprising is the similar transport mode distribution across all African regions (approximately 31-34% air, 31-34% sea, 15-18% rail, 16-19% road) despite vastly different infrastructure development levels and geographical characteristics. This unexpected uniformity suggests standardized logistics patterns that may not be optimally tailored to each region's unique characteristics and constraints.

4.10.3 Specialized Logistics Channels and Commodity Patterns

Our analysis identified several specialized logistics patterns unique to African exports:

- 100% of RoRo air shipments from Africa transport "Transport_equipment," suggesting a specialized logistics channel likely dedicated to high-value vehicles such as luxury cars, racing vehicles, or specialized machinery.
- South Africa dominates these shipments (1,708 out of 27,066), aligning with its established automotive manufacturing industry.
- Electronic devices consistently command the highest premiums across African exporters, regardless of origin country or destination.

These specialized patterns demonstrate niche optimization in certain segments despite overall logistics challenges.

4.10.4 Rethinking the Landlocked Disadvantage

Contrary to conventional assumptions, landlocked countries exhibit lower average logistics costs ($5,855/ton) than coastal countries ($6,361/ton), with the difference being statistically significant ($p < 0.0001$). This finding highlights that geography alone does not determine trade efficiency.

Possible explanations include:

- Many landlocked countries, particularly in Europe, benefit from excellent infrastructure and proximity to ports.

- They often trade with neighboring countries, reducing transport distances.

- Their exports tend to be high-volume, low-value goods like agriculture.

- Strong regional integration and cross-border logistics networks help offset physical constraints.

In Africa, the cost gap is smaller—landlocked countries face only a 6% premium—further suggesting that infrastructure and policy can mitigate geographic challenges.

An exception is Lesotho, a landlocked enclave within South Africa, which shows disproportionately high export costs. Despite being surrounded by a highly efficient economy, Lesotho's case illustrates how political borders and customs frictions can outweigh even geography in creating trade inefficiencies.

4.11 Comparative Continental Analysis: The African Context

Our comparative analysis of transport costs across continents reveals distinct patterns that help explain Africa's unique logistics challenges and provide context for targeted interventions.
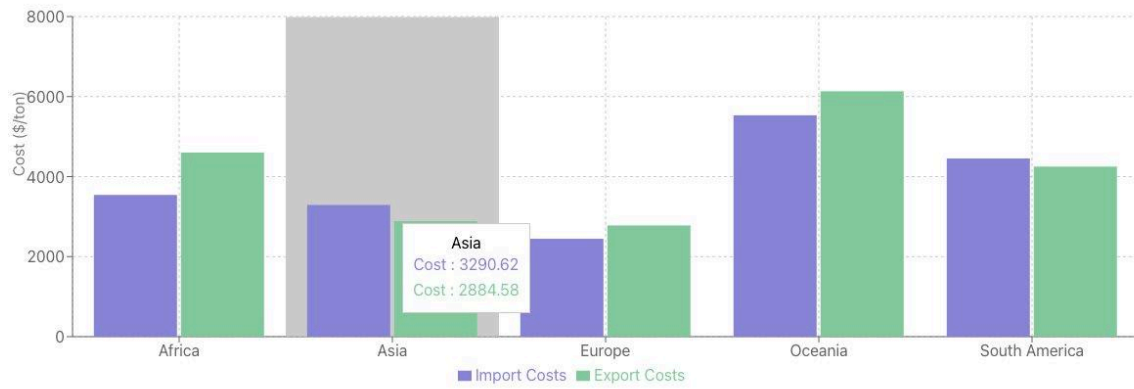
4.11.1 Import-Export Cost Imbalances

The continent-level analysis reveals a striking pattern in how logistics costs are distributed:

- Africa maintains the second-highest export costs globally (after Oceania), with a heavily skewed import-export balance
- While Oceania's high costs are explained by its extreme geographic isolation, Africa's challenges stem from different structural factors
- Europe and Asia maintain relatively low costs and healthier trade balances through dense multimodal networks, top-ranked ports, and high shipping volumes

Our analysis identified a critical feedback loop that perpetuates high costs in Africa: unlike Asia, where massive export volumes make outbound shipping cost-effective (despite expensive imports), Africa faces the opposite situation. Low export volumes mean ships often sail out half-empty, forcing carriers to recover fixed costs by inflating export prices. This reverse imbalance, combined with the oligopolistic structure of global shipping markets, traps Africa in a high-cost, low-volume cycle that undermines trade competitiveness and development potential.

Import vs Export Logistics Costs by Continent

Comparing the costs of importing goods to a continent versus exporting from that continent.

Cost ($/ton)

Asia
Cost : 3290.62
Cost : 2884.58

Africa   Asia   Europe   Oceania   South America

■ Import Costs ■ Export Costs

Export-Import Cost Balance by Continent

Negative values indicate higher export costs than import costs. Bubble size represents shipping volume.

Export-Import Balance ($/ton)

Africa   Asia   Europe   Oceania   South America
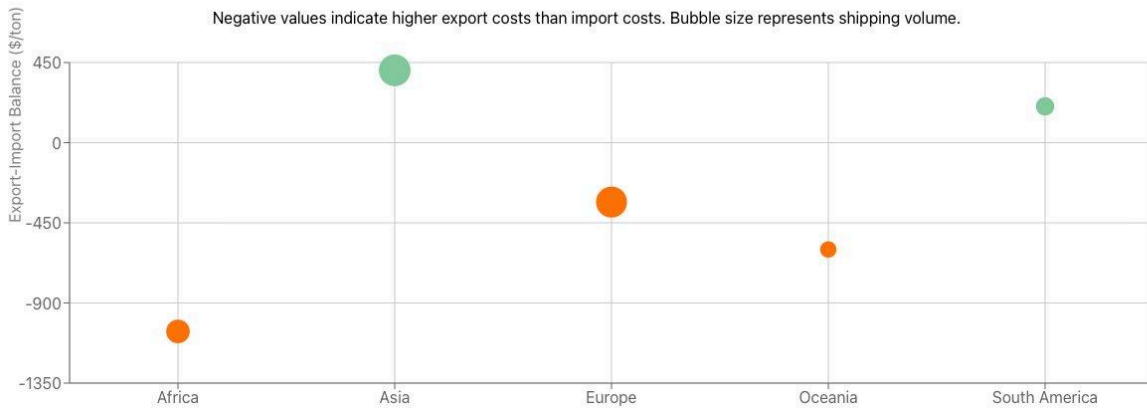
Export-Import Balance by Region

Balance (Export – Import) $ — Export/Import Ratio

Mean vs Median Costs (bubble size represents shipment volume)

● Regions

Oceania
Latin America
North America
South Asia
East Asia

Other
Southeast Asia
Africa
Middle East
Europe

4.11.2 Intra-Regional vs. Inter-Regional Trade Costs

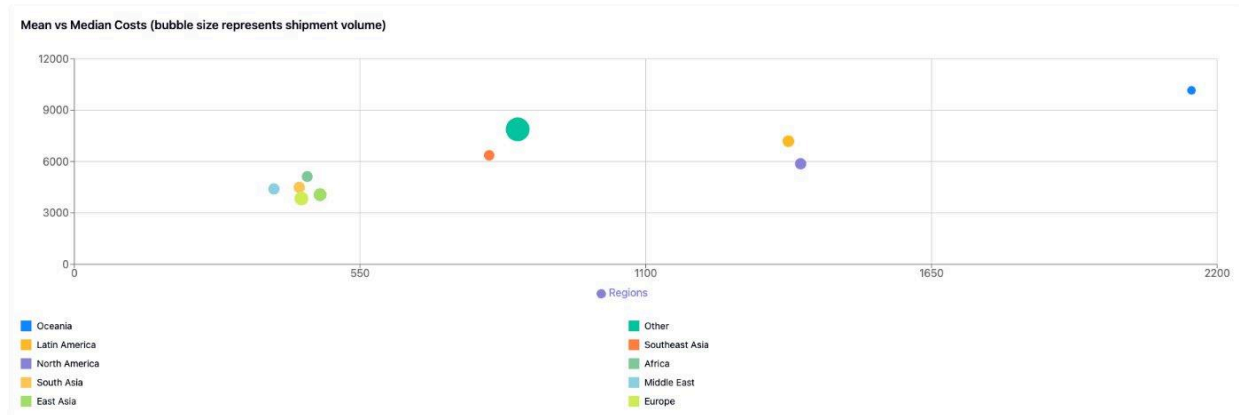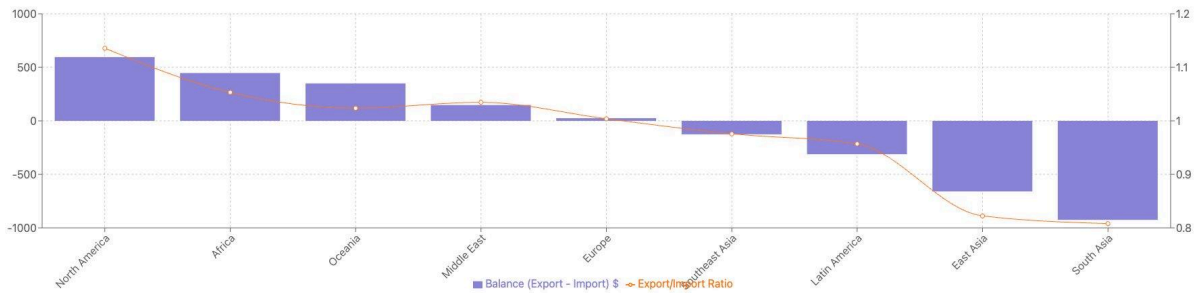Regional analysis reveals substantial cost disparities, with Oceania showing the highest average costs ($10,157/ton)—more than 2.5 times higher than Europe ($3,839/ton). Origin-destination pair analysis demonstrates that intra-regional trade is generally much cheaper than inter-regional movements, with Europe achieving the lowest intra-regional costs ($820/ton).

However, one of our most counter-intuitive findings relates to African intra-regional trade costs:

- Intra-African shipping ($2,174/ton) costs nearly twice as much as intra-European shipping ($1,207/ton)
- Intra-African trade represents only 16-18% of total African exports, compared to 71% for Europe and 55% for Asia-Pacific
- This pattern defies conventional wisdom that regional proximity should reduce transport costs

Intra-Regional vs Inter-Regional Shipping Costs

Comparing shipping costs within regions versus between different regions.

Intra vs. Inter-Continent Cost Comparison
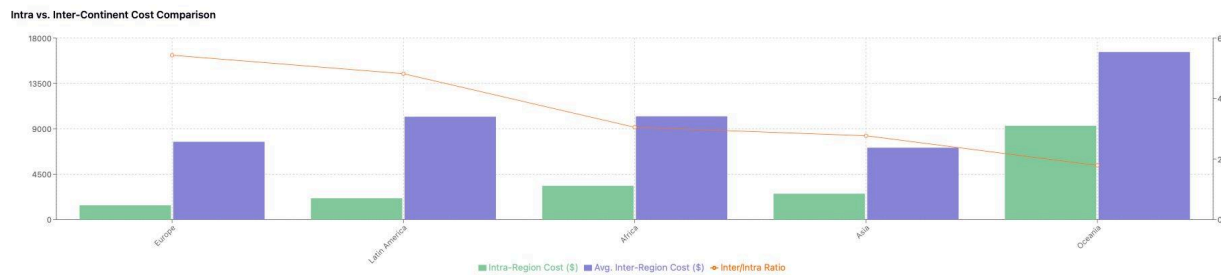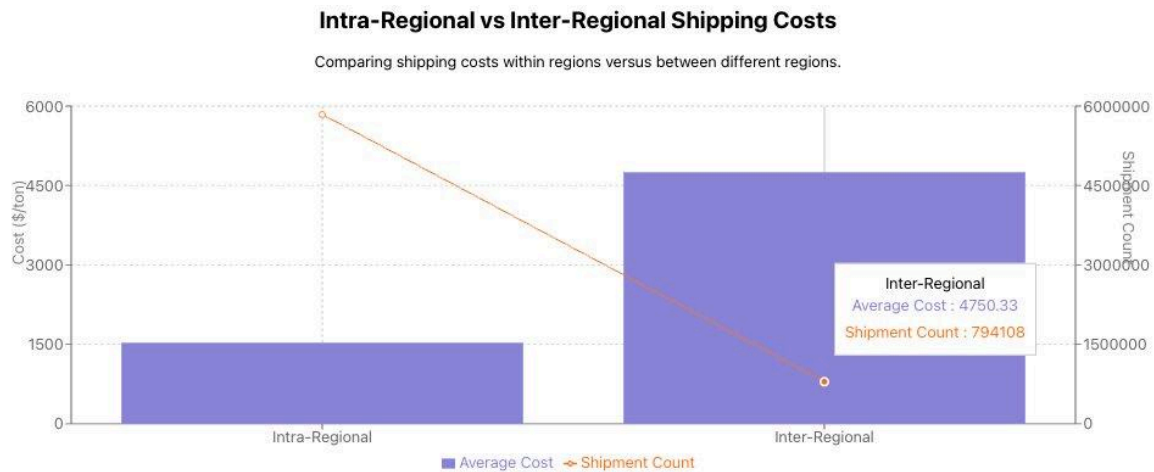
Within Africa itself, we observed significant regional variations. North Africa achieves the lowest logistics costs ($4,505/ton) while exporting the second-highest volume, demonstrating superior logistics efficiency. By contrast, East Africa's costs are 71% higher ($7,71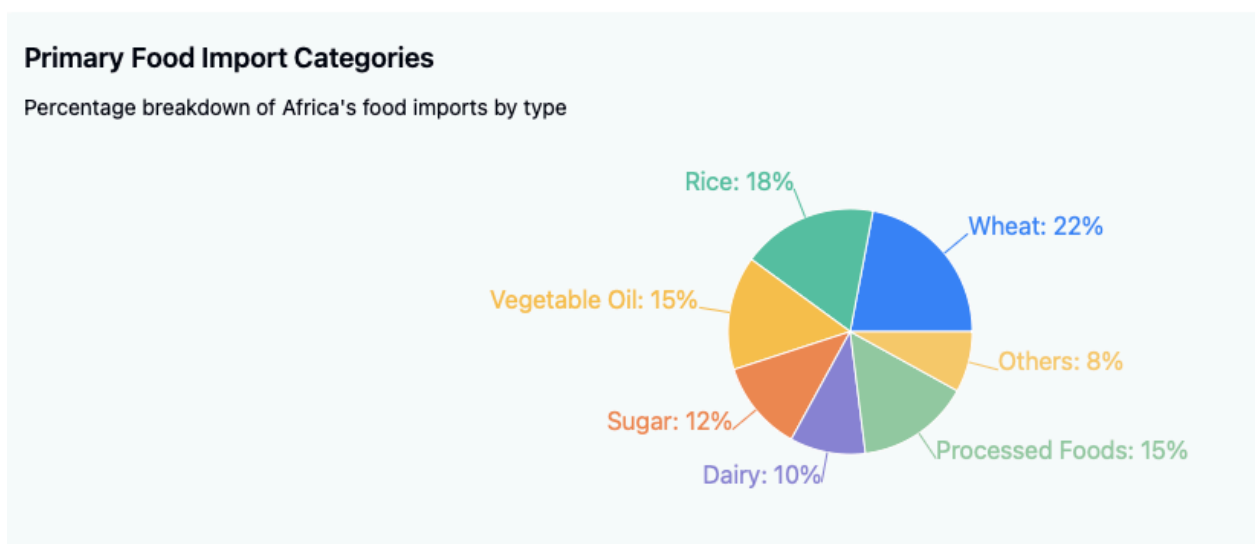2/ton) despite moving comparable volume, suggesting significant infrastructure or operational inefficiencies that represent major improvement opportunities.

African Export Statistics and Transport Mode Distribution

Export Volume and Unit Logistics Costs by Region

Southern and North Africa have the highest export volumes, while East Africa faces the highest logistics costs
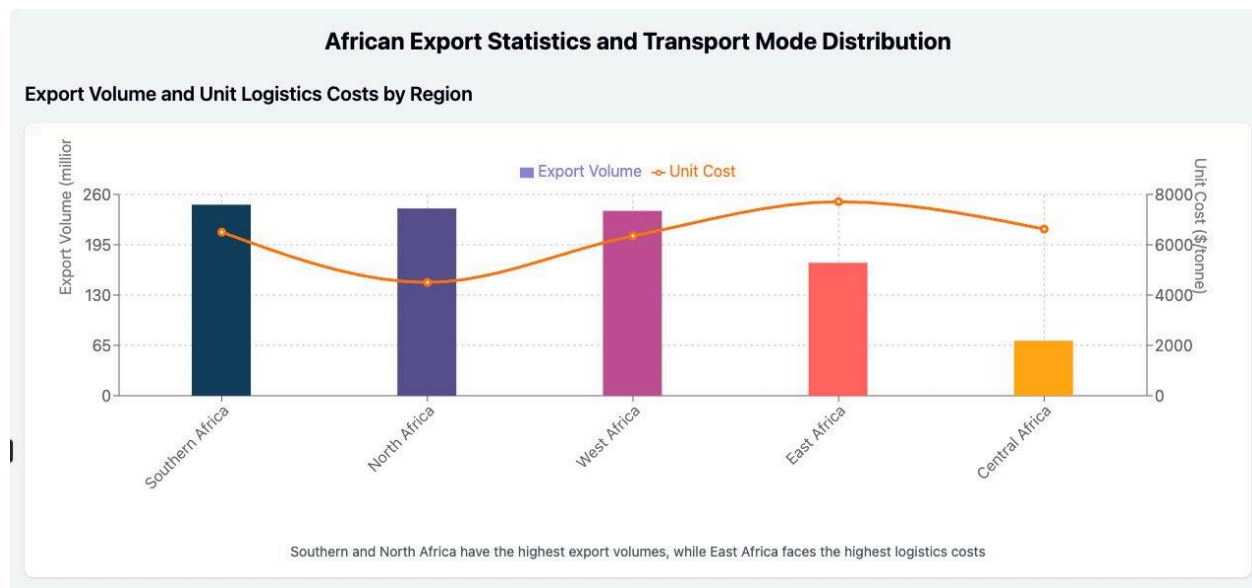
The intra-regional cost anomaly in Africa can be largely attributed to historical trade patterns established during colonial periods, which oriented African economies toward exporting raw materials to external markets while importing finished goods. This extractive model has created a persistent structural trap where:

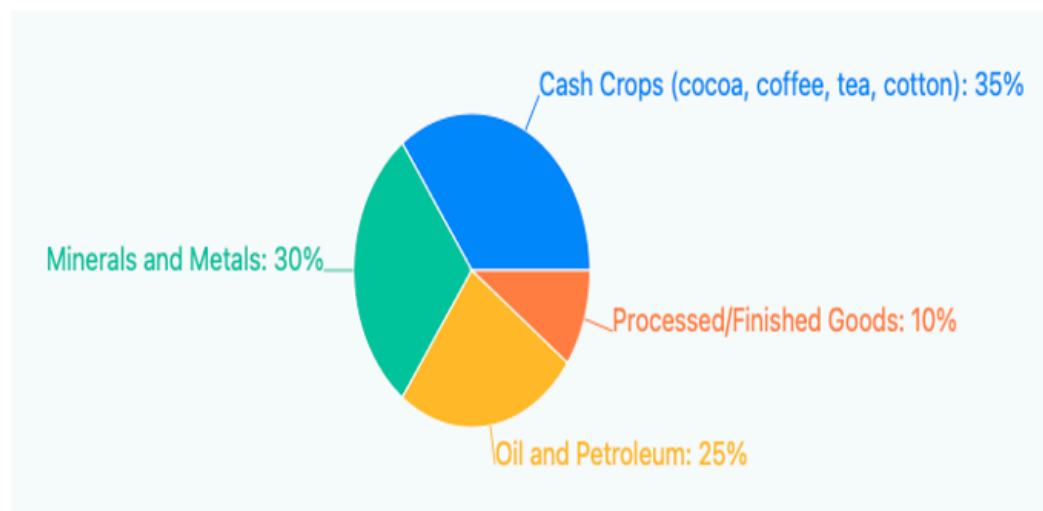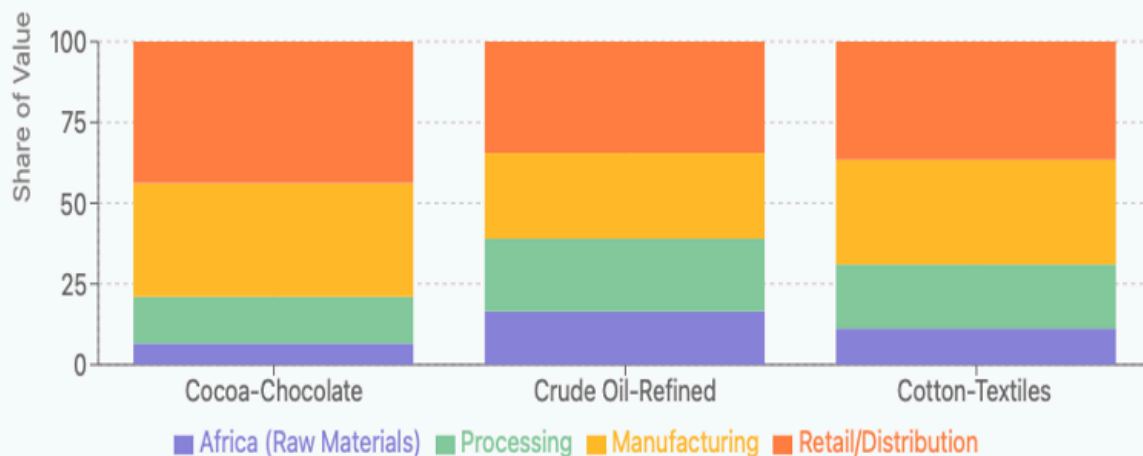1. African countries export raw materials (e.g., cocoa, minerals) to developed markets
2. Value-addition occurs abroad through processing and manufacturing
3. African countries import finished products (e.g., chocolate, processed foods) at marked-up prices
4. Domestic industries remain underdeveloped due to this competitive disadvantage

Primary Food Import Categories

Percentage breakdown of Africa's food imports by type

Value Captured at Each Stage

Percentage of final value captured at each stage

Legend: ■ Africa (Raw Materials) ■ Processing ■ Manufacturing ■ Retail/Distribution

Cash Crops (cocoa, coffee, tea, cotton): 35%
Minerals and Metals: 30%
Processed/Finished Goods: 10%
Oil and Petroleum: 25%

The data shows the direct economic consequences of this pattern: Africa spends $35-50 billion annually importing processed foods despite its agricultural potential. The high cost of intra-regional transport reinforces this dependency, as it's often cheaper to import processed foods from North America than to trade them across African borders. This creates a self-perpetuating cycle where colonial-era trade patterns are reinforced by current logistics cost structures.

The most expensive trade routes globally occur between Oceania and "Other" regions ($13,682/ton), confirming that geographic proximity plays a crucial role in determining logistics costs for most regions—making Africa's intra-regional cost premium all the more significant as a structural anomaly.

Most and Least Expensive Routes

Top 5 Most Expensive Routes

Top 5 Least Expensive Routes
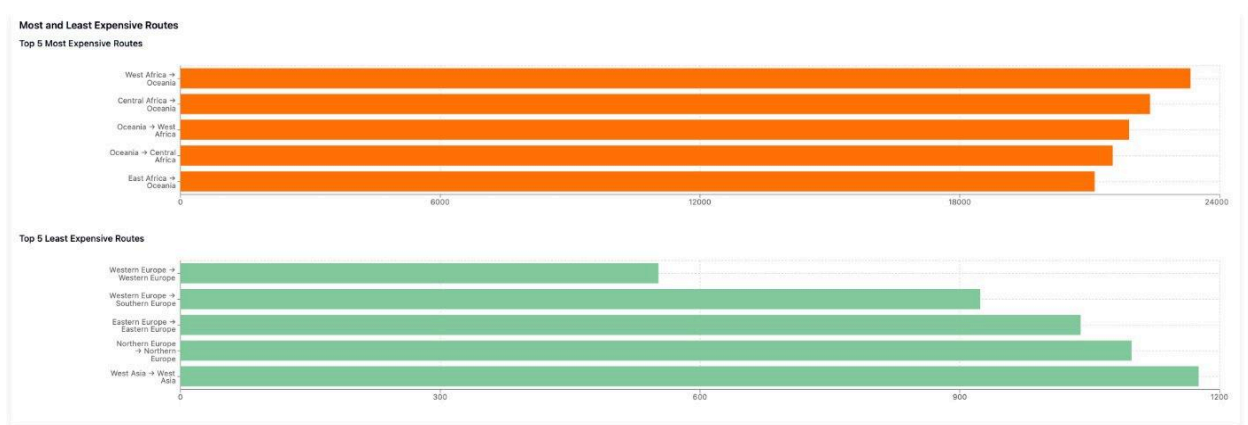
4.11.3 Volume-Cost Relationship and Regional Variations

Our analysis identifies a moderate negative correlation (-0.1968) between trade volume and logistics costs, indicating that countries with higher trade volumes tend to achieve somewhat lower logistics costs, likely through economies of scale. While not as strong as might be expected, this relationship confirms that volume consolidation remains an important strategy for cost reduction.

Regression Curve: Shipment Volume vs. Cost per Ton

Small shipments face drastically higher costs, especially from Africa
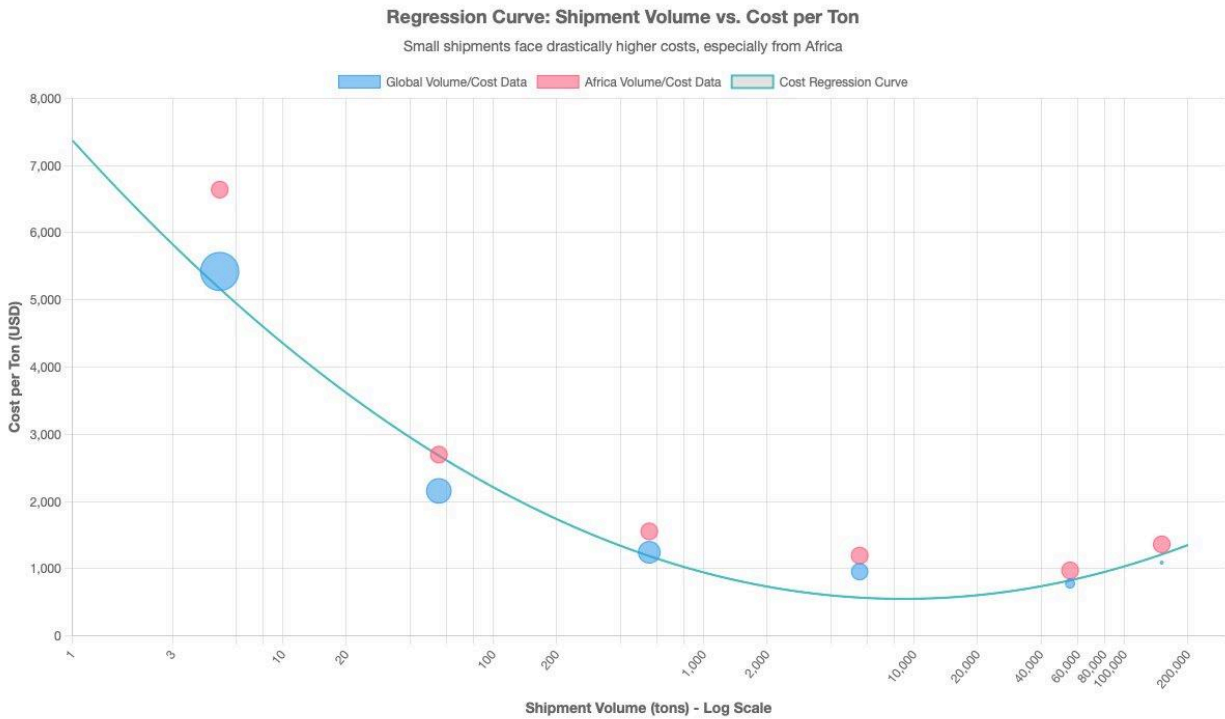
Our regression analysis reveals more nuanced patterns:

- Globally, shipping costs decrease predictably as volume increases (following expected economies of scale)
- African low-volume shipments (0-10 tons) face disproportionately higher costs compared to global averages
- At higher volumes, African shipping costs converge with global averages

Top 10 Most Expensive Region-Volume Combinations

Small shipments (0-10 tons) from developing regions are particularly expensive.
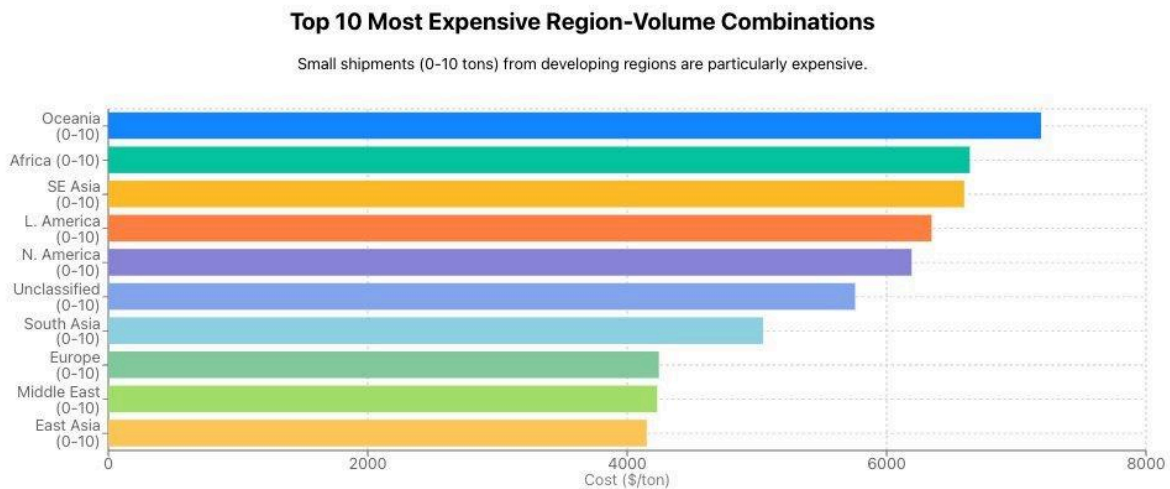
This pattern suggests that fragmented supply chains represent a significant opportunity for immediate cost reduction through volume consolidation. However, our regional comparison within Africa demonstrates that scale alone cannot explain cost variations:

- North Africa moves approximately the same freight volume as Southern Africa but at 32% lower unit cost
- East Africa ships less volume yet pays 71% more than North Africa

Efficiency analysis based on cost per ton-kilometer metrics reveals that infrastructure quality and logistics capabilities often outweigh simple distance in determining costs:

- Least efficient countries: Israel and Palestine stand out with extreme inefficiency (over $3,300 per ton-km)
- Most efficient countries: China demonstrates remarkable efficiency ($0.26 per ton-km), followed by India and Malaysia

These dramatic efficiency differences and stark regional variations prove that governance quality, port efficiency, and logistics management capabilities matter at least as much as pure economies of scale in determining final shipping costs.

4.12 Integration of Findings:

4.12.1: The Determinants of Transport Efficiency

Integrating our various analyses, we can identify several key determinants of transport efficiency:

1. **Mode Selection**: The choice of transport mode emerges as the single most powerful determinant of cost structure, with rail offering the best overall efficiency while air transport represents a necessary but expensive option for high-value, time-sensitive goods.
2. **Distance Optimization**: Each transport mode has an optimal distance range where it performs most efficiently. Short distances face disproportionately high costs per kilometer across all modes, while long distances enable better fixed-cost distribution.
3. **Infrastructure Quality**: Countries with well-developed multimodal infrastructure demonstrate consistently lower costs, even when facing geographical disadvantages such as landlocked status.
4. **Commodity Characteristics**: The nature of goods being shipped significantly influences optimal mode selection and resultant costs, with different commodities gravitating toward their most suitable transport options.

5. **Geographic Factors**: While island nations and remote locations face inherent cost disadvantages, effective logistics management and infrastructure investment can partially mitigate these challenges.
6. **Operational Efficiency**: The presence of extreme cost outliers across all categories suggests that operational factors—including customs procedures, port efficiency, and logistics coordination—may sometimes outweigh structural factors in determining final costs.

These determinants interact in complex ways, creating unique logistics profiles for different countries, commodities, and trade routes. Understanding these interactions provides the foundation for developing targeted interventions to improve transport efficiency across the global trade system.

4.12.2: Key Insights and Implications

Our analysis yields several crucial insights with significant implications for policy and investment:

1. **Air transport demonstrates extreme inefficiency for short distances**, suggesting potential for substantial savings through mode shifting where alternative infrastructure exists.
2. **Sea and rail transport offer the best efficiency for long distances**, with costs per ton-kilometer stabilizing as distance increases, reinforcing their role as the backbone of global trade.
3. **Short-distance transport faces disproportionate inefficiencies across all modes**, indicating potential for targeted interventions in border procedures, local infrastructure, and regional logistics coordination.
4. **Land routes (road and rail) provide optimal efficiency at mid-range distances**, highlighting the importance of regional transport corridor development.
5. **Extreme outliers in certain routes** (notably Palestine-Israel and routes to American Samoa) suggest that political, security, or extreme geographic factors can override economic considerations in transport patterns.

These insights point toward strategic opportunities to enhance global transport efficiency through targeted infrastructure development, policy reforms, and logistics optimization initiatives. By addressing specific inefficiencies rather than applying generic solutions, stakeholders can maximize the impact of limited investment resources.

4.13 Cost Reduction Opportunities in African Exports

Our analysis of potential cost reduction strategies revealed several high-impact opportunities unique to African exports.

4.13.1 Data Quality Concerns in Rail Cost Reporting

The analysis identified multiple air-to-rail optimization opportunities with exactly 0.0 optimal costs, indicating perfect 100% savings. This statistical impossibility likely indicates systematic data collection errors for rail costs on certain routes. These "phantom efficiencies" require urgent data verification before investment decisions, particularly for Liberia-originating shipments, to avoid misallocated infrastructure spending.

4.13.2 Regional Infrastructure Opportunities

Several high-potential infrastructure investment opportunities emerged from our analysis:

- **Southern African Rail Corridors**: Zimbabwe-Zambia rail connections show significant cost reduction potential (up to 100% savings versus air). This intra-regional Southern African route could serve as a proof-of-concept for the African Continental Free Trade Area (AfCFTA), demonstrating how strategic rail investments between neighboring countries can eliminate air transport premiums.
- **Egyptian Port Model Replication**: Egypt emerges as Africa's most versatile and cost-effective export gateway, ranking as the most efficient port for five of six global regions. This extraordinary cross-regional efficiency advantage (25-67% cost savings versus alternatives) suggests Egypt has developed transferable best practices in port operations that should be studied and replicated across other African coastal nations.
- **Atlantic Gateway Development**: Cape Verde demonstrates significant potential as Africa's second most cost-effective gateway to North America ($3,356/ton versus Egypt's $2,817/ton). This suggests Atlantic-facing ports have a natural efficiency advantage for American-bound exports that remains underutilized by many West African nations with similar geographic positioning.
- **Eastern Corridor Prioritization**: Four of the five most expensive transport corridors originate from either Ethiopia or Lesotho, indicating that East and Southern African landlocked countries face disproportionately higher logistics challenges than their West African counterparts. The development of efficient eastern corridors should be prioritized in continental infrastructure planning.

4.13.3 Policy Recommendations by Region

Central Message: "Regional hub development can dramatically reduce transport costs across multiple corridors."
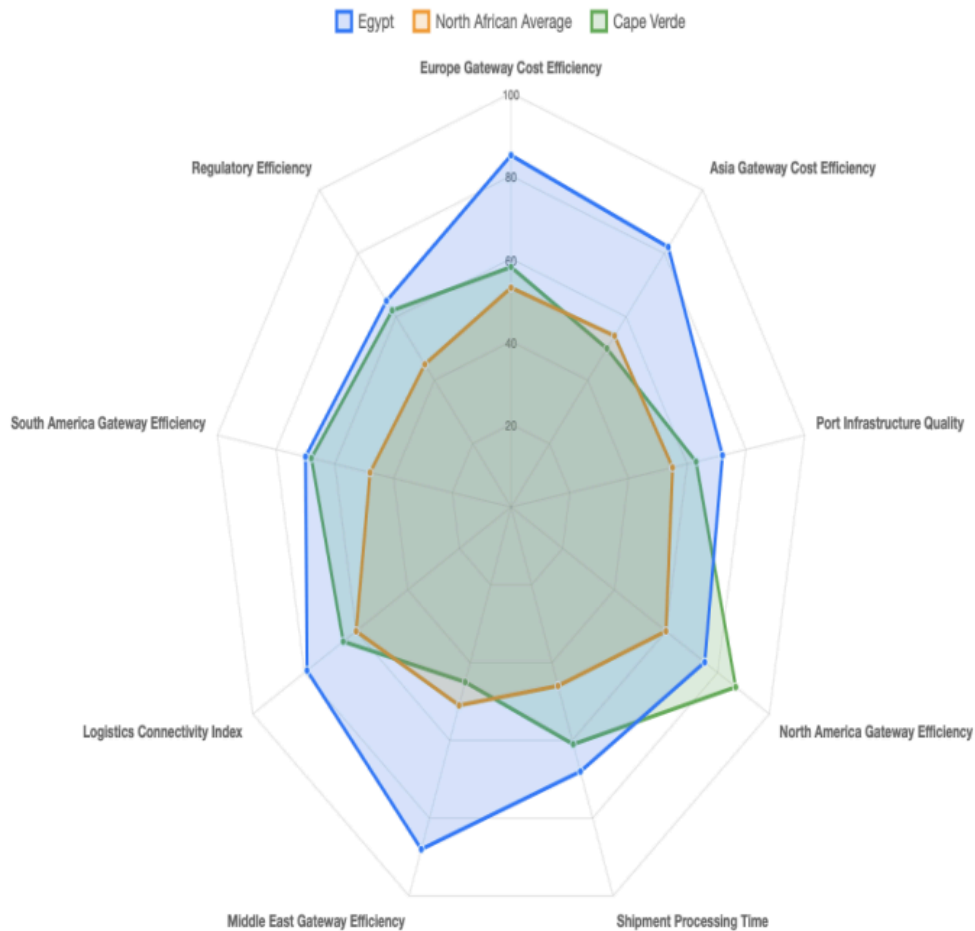
Based on our regional and anomaly analysis, we propose the following targeted interventions to reduce Africa's logistics costs and improve trade efficiency:

Regional Priorities

- **North Africa**: Scale the **Egyptian model** of port efficiency and multimodal integration. Egypt's export dominance offers replicable practices for other coastal nations.

- **West Africa**: Invest in **Atlantic-facing gateways**, especially underutilized hubs like **Cape Verde**, to boost transatlantic trade with the Americas.

- **Southern Africa**: Prioritize **rail corridor investments**—starting with high-impact links like **Zimbabwe–Zambia**—to enable efficient modal shifts.

- **East Africa**: Develop **Special Economic Zones (SEZs)** near ports and rail with **export tax incentives** to offset the region's high logistics costs.

Cross-Cutting Policy Levers

- **Scale Before You Sail**: Encourage consolidation into **10,000+ ton shipments**, which yield the highest cost savings. Tailor transport modes to commodity types:

    - *Cold-chain air* for perishables

    - *Rail* for ores

    - *Bulk sea freight* for grains

- **Corridors Over Coordinates**: Africa's exports cost ~$1,000/ton more than imports—not due to distance, but due to **bottlenecks** and **incomplete rail links**. Focus on fixing broken corridors.

- **Modal Shift = Billion-Dollar Opportunity**: Even a **modest shift from air/road to rail** could save **over $1B/year**, especially for **food** and **critical minerals**.

- **Target the Outliers First**: Ultra-short **air shipments** and **island supply chains** account for a disproportionate share of costs. These are **low-hanging fruit** for policy reform and innovation.

Additional Insights

- Improve **data quality** on anomalous corridors (e.g., ZAF–RUS, GAB–DZA) to detect real inefficiencies vs. reporting errors.

- Expand **inland logistics hubs** and **dry ports** for landlocked countries to strengthen last-mile access.

- Optimize **air freight for high-value, low-weight exports** (e.g., electronics, perishables) via route planning, consolidation, and cold-chain investment.

Global Trade Risks

Any new global tariffs could further raise Africa's already-high export costs, threatening economies that are deeply embedded in global raw material supply chains. Without support, these regions risk being **economically stranded**, with severe consequences for livelihoods and growth.

5. Conclusion

A. Conclusion

This analysis of the UNCTAD provides valuable insights into the patterns, determinants, and inefficiencies in global transport systems, with particular relevance for developing regions like Sub-Saharan Africa. Several key conclusions emerge from this research:

1. **Transport mode selection follows predictable patterns** influenced by commodity type, distance, infrastructure availability, and economic considerations. Air and sea dominate international trade, with air transport reserved for high-value, time-sensitive goods despite its extreme cost premium. Sea transport offers the most cost-efficient option for longer distances, while rail demonstrates the lowest overall costs but remains underutilized in many regions.
2. **Geographic factors create significant cost disparities**, with landlocked countries facing higher and more variable transport costs than coastal nations. However, the relationship between distance and cost is weaker than expected, suggesting that infrastructure quality, border procedures, and market structures may be more important determinants of transport efficiency than simple geography.
3. **Modal flexibility significantly reduces logistics costs**. Countries with access to all four transport modes demonstrate substantially lower costs than those limited to fewer options, highlighting the importance of developing diverse transport infrastructure rather than focusing exclusively on a single mode.
4. **Economies of scale drive cost efficiencies** at multiple levels. Larger shipment volumes, higher national trade volumes, and more established trade routes all correlate with lower per-unit transport costs. This creates advantages for larger economies and challenges for smaller countries with limited cargo volumes.
5. **Regional integration offers substantial cost benefits**. Intra-regional trade routes demonstrate significantly lower costs than inter-regional connections, emphasizing the economic value of strengthening regional trade agreements and infrastructure corridors.
6. **Transport costs vary significantly across commodities** and regions, creating both challenges and opportunities for targeted interventions. The persistence of extremely high costs for certain routes and commodity types suggests potential for substantial efficiency improvements through better mode selection, infrastructure development, and logistics optimization.
7. **North Africa outperforms other African regions** in transport cost efficiency, demonstrating that improvements are achievable within the African context. Egypt's successful port model and superior logistics performance offer valuable lessons for neighboring countries.

These findings have important implications for policy development, infrastructure investment, and logistics strategy in developing regions. By understanding the determinants of transport mode choice and identifying specific inefficiencies, stakeholders can develop targeted interventions to reduce costs, enhance competitiveness, and support economic development.

B. Project Limitation

Despite the comprehensive nature of this analysis, several limitations should be acknowledged:

1. **Data Completeness Challenges**: The dataset contained significant missing values for transport modes, requiring predictive modeling to fill these gaps. While the prediction model demonstrates strong performance, these estimates introduce an element of uncertainty into the analysis.
2. **Limited Time Period**: The dataset covers only the year 2020. This relatively short time period may not capture longer-term trends or the full impact of major infrastructure investments that take many years to complete.
3. **Aggregation Level**: The analysis operates at the country level, potentially masking important sub-national variations in transport costs and infrastructure quality. Regional differences within large countries are not captured in this country-pair approach.
4. **Limited Visibility into Non-Tariff Barriers**: While the dataset captures monetary costs, it provides limited insight into other important factors such as time delays, reliability issues, and administrative barriers that significantly influence mode choices in practice.
5. **Focus on International Trade**: The dataset primarily addresses international trade flows, with limited coverage of domestic transport costs and patterns that form critical components of overall logistics expenses, particularly for large countries.
6. **Anomalous Data Patterns**: Several unusual patterns were identified, such as the exact 50-50 split between air and sea transport for island nations and zero-cost rail anomalies, suggesting possible data collection or processing issues that could affect specific findings.
7. **Limited Infrastructure Quality Metrics**: While the dataset includes information on the presence of infrastructure, it lacks detailed quality metrics that would provide more nuanced understanding of infrastructure effectiveness.

These limitations suggest opportunities for future research to develop more granular, time-extensive, and comprehensive datasets that could further enhance understanding of transport mode choice determinants.

C. Future Research

Based on the findings and limitations of this study, several promising directions for future research emerge:

1. **Develop More Granular Sub-National Models**: Future research should explore transport costs and mode choices at sub-national levels, particularly for large countries where internal logistics can represent a substantial portion of total costs. This would

require gathering data on domestic transport networks and costs to complement the international focus of the current dataset.

2. **Incorporate Time and Reliability Metrics**: Expanding the analysis to include non-cost factors such as transport time, reliability, and predictability would provide a more comprehensive understanding of mode choice determinants, as these factors often outweigh pure cost considerations for certain goods and markets.

3. **Conduct Longitudinal Studies of Infrastructure Investments**: Tracking specific infrastructure corridors or projects over longer time periods could provide valuable insights into the impact of investments on transport costs, mode shifts, and trade volumes. This would help quantify the return on investment for different types of transport infrastructure.

4. **Develop Commodity-Specific Optimization Models**: Creating specialized models for key commodity groups (such as agricultural products, minerals, or manufactured goods) would enable more targeted recommendations for specific supply chains that face unique constraints and opportunities.

5. **Analyze the Impact of Trade Agreements and Policies**: Investigating how changes in trade policies, customs procedures, and regional agreements affect transport costs and mode choices could help identify non-infrastructure interventions that might yield significant efficiency gains.

6. **Explore Emerging Transportation Technologies**: Assessing the potential impact of new technologies such as autonomous vehicles, drone delivery, or blockchain-based logistics management could help anticipate future changes in optimal mode choices and infrastructure requirements.

7. **Develop Multi-Modal Optimization Tools**: Creating practical decision support tools that help shippers identify optimal mode combinations for specific routes and cargo types could translate the research findings into practical applications that deliver immediate cost savings.

8. **Conduct Comparative Studies of Regional Success Stories**: More detailed analysis of regions that have achieved superior transport efficiency, such as North Africa compared to other African regions, could identify transferable best practices and policy approaches.

By pursuing these research directions, future studies can build on the current findings to develop more nuanced, practical, and forward-looking insights into transport mode optimization, ultimately contributing to more efficient global trade and economic development.

6. References

1. UNCTAD. (n.d.). Global Transport Costs Dataset for International Trade. UNCTAD Stat Data Center. Retrieved from

   https://unctadstat.unctad.org/datacentre/reportInfo/US.TransportCosts

2. World Bank. (n.d.). Transport Costs and Economic Development. World Bank Knowledge Repository. Retrieved from

   https://openknowledge.worldbank.org/bitstreams/0c3172ea-7ff1-4d95-8adf-2a6b7c25b73a/download

3. Brookings Institution. (n.d.). Unpacking the Misconceptions About Africa's Food Imports. Retrieved from https://www.brookings.edu/articles/unpacking-the-misconceptions-about-africas-food-imports/

4. United Nations. (2017-2018). Towards Food Secure Africa. Africa Renewal Magazine, December 2017-March 2018. Retrieved from https://www.un.org/africarenewal/magazine/december-2017-march-2018/towards-food-secure-africa

5. **UN and International Organization Data Sources**

   African Development Bank (AfDB) -

   Official statistics on Africa's food imports and agricultural sector Feed Africa Initiative Data on $35 billion annual food import expenditure

   United Nations Conference on Trade and Development (UNCTAD)

   Review of Maritime Transport
   Facts & Figures on Africa
   Comprehensive statistics on commodity trade flows

   Food and Agriculture Organization (FAO)
   Why has Africa become a net food importer?
   Longitudinal data on agricultural production and trade

   International Monetary Fund (IMF)
   Primary Commodity Prices
   Data on commodity price fluctuations affecting African exports

6. **Research Institutions and Government Data**
   Brookings Institution
   Unpacking the misconceptions about Africa's food imports
   Country-level agricultural trade analysis
   World Bank

Open Data Portal
Africa's Food Markets Could Create One Trillion Dollar Opportunity by 2030

BIMCO Container Trade Statistics

•Container shipping imbalance data showing 2.3:1 import-export ratio in Sub-Saharan Africa