

Logistics Shipments Dataset – Business Rules, Metrics, and Analytical Definitions

Dataset Scope & Purpose

This dataset captures end-to-end shipment activity across warehouses, carriers, and destinations. Each record represents a single shipment event and supports operational and analytical use cases in logistics and supply chain management.

Primary objectives include analyzing shipment volume, cost efficiency, carrier performance, delivery speed, and warehouse throughput.

Table Definition

Table Name: logistics_shipments

Grain: One row per shipment

Column Definitions

Shipment_ID: Unique identifier for each shipment.

Origin_Warehouse: Warehouse where the shipment originated.

Destination: Final delivery location.

Distance_miles: Distance between origin and destination in miles.

Carrier: Logistics provider responsible for shipment.

Shipment_Date: Date shipment was dispatched.

Delivery_Date: Date shipment was delivered.

Transit_Days: Number of days between shipment and delivery.

Weight_kg: Weight of shipment in kilograms.

Cost: Total shipping cost.

Status: Shipment state (Delivered, In Transit, Delayed).

Core Business Metrics

Total Shipments: Total number of unique shipments.

Total Shipping Cost: Aggregate logistics spend.

Average Cost per Shipment: Mean shipment cost.

Average Transit Time: Mean delivery duration.

Cost per Mile: Cost efficiency across routes.

Carrier Performance Metrics

Carrier-level shipment count, total cost, and average transit time are used to evaluate performance and reliability.

Analytical Rules & Constraints

Supported analyses include carrier ranking, warehouse throughput, distance-cost analysis, and delivery trend tracking.

Unsupported analyses include customer churn, profitability, SLA penalties, or emissions modeling due to missing attributes.

Data Quality Rules

Shipment_ID must be unique.

Dates must follow logical order.

Numeric values must be non-negative.

No assumptions or imputations are permitted.

RAG Interpretation Rules

All insights must be strictly grounded in the dataset.

No external assumptions or inferred behaviors are allowed.

Example Questions

- Which carrier has the lowest average transit time?
- What is total shipping cost by warehouse?
- Which routes are most expensive per mile?
- How does shipment volume trend over time?