To start data wrangling for this project, follow these steps:

### **1. Understand the Dataset**

* Go through the dataset to understand its structure and the meaning of each column.
* Identify the key attributes such as ID, Project Code, Country, Vendor, Item Description, Scheduled Delivery Date, Delivered to Client Date, Line Item Quantity, Line Item Value, Freight Cost, etc.
* Recognize which columns are critical for your analysis and which might need cleaning.

### **2. Data Cleaning**

* **Remove or Handle Missing Values:**
  + Check for missing or null values in columns like PO Sent to Vendor Date, Delivered to Client Date, etc.
  + Decide if you need to remove rows with missing data or fill them using strategies like mean, median, or using domain knowledge.
* **Correct Data Types:**
  + Ensure dates (e.g., Scheduled Delivery Date, Delivered to Client Date) are in the correct datetime format.
  + Numeric columns (e.g., Line Item Quantity, Line Item Value, Freight Cost) should be in the appropriate numerical format.
* **Handle Duplicates:**
  + Check if there are duplicate rows in the dataset and remove them, if necessary.
* **Standardize Data:**
  + Standardize text-based data (e.g., Vendor, Product Group) to avoid inconsistencies like spelling variations.

### **3. Feature Engineering**

* **Create New Columns:**
  + **Delivery Time:** Create a new column to calculate the difference between Scheduled Delivery Date and Delivered to Client Date to analyze delivery delays.
  + **Total Shipping Cost:** Add columns that combine Line Item Value, Freight Cost, and Line Item Insurance to calculate the total shipping cost per order.
* **Categorize Data:**
  + You might categorize countries into regions (e.g., Africa, Asia) or group products into broader categories based on the Product Group and Sub Classification.

### **4. Outlier Detection and Handling**

* Look for any outliers in the numerical data (e.g., Freight Cost, Line Item Value) that may affect analysis.
* Decide whether to remove or adjust these outliers based on business logic.

### **5. Data Validation**

* Ensure all data manipulations have preserved the integrity of the original data.
* Validate that dates and numbers make sense within the context (e.g., no negative costs, delivery dates after scheduled dates).

### **6. Exploratory Data Analysis (EDA)**

After wrangling the data, move on to exploring it through visualizations and summaries to uncover patterns and insights.

By following these steps, you can clean and prepare the data for deeper analysis and meaningful insights.

—------------------------------------------------------------------------------------------------------------------------

In the dataset, **Line Item Quantity** being less than **Unit of Measure (Per Pack)** could indicate a few scenarios or data issues. Here’s a detailed breakdown of what could be happening:

### **Key Definitions:**

1. **Unit of Measure (Per Pack)**: This refers to the number of individual units of the product in each pack. For example, it might state that each pack contains 30 units.
2. **Line Item Quantity**: This refers to the **total number of units** of the product being shipped, which might not directly correspond to the pack size. It's the overall total quantity ordered or shipped for that particular line item.

### **Possible Scenarios:**

1. **Line Item Quantity Could Be Less Than Unit of Measure (Per Pack) in Data Mistakes**:
   * If **Line Item Quantity** is less than the **Unit of Measure (Per Pack)**, this might be an error or inconsistency in how the data was recorded or processed. For instance, if the dataset states there are 100 units being shipped, but the **Unit of Measure (Per Pack)** suggests that each pack holds 240 units, the data entry might have a mismatch.
2. **Line Item Quantity Represents Individual Units or Partial Packs**:
   * Sometimes, **Line Item Quantity** might refer to the total number of individual units ordered, not packs, and these individual units may not be a full pack.
   * Example: If each pack contains 240 items (**Unit of Measure (Per Pack)**), and the **Line Item Quantity** is 100, this means the shipment is for **100 individual units** (not whole packs). This could happen when only part of a pack is being shipped, or the dataset is structured to list the individual quantity regardless of how they are packed.
3. **Unit of Measure for Pack vs. Shipment**:
   * It's also possible that **Unit of Measure (Per Pack)** refers to a larger quantity (e.g., 240 units per pack) but the **Line Item Quantity** indicates a smaller quantity that is intended for shipment. If this is a case of partial shipment or different packaging configurations, the two numbers might not align directly.

### **Example from Dataset:**

* For **Line Item Quantity = 100** and **Unit of Measure (Per Pack) = 240**, it could imply:
  + **100** total individual units are being shipped.
  + **240** is how many units each individual pack contains.
  + The shipment consists of fewer than 1 full pack in terms of **packs**, but **100 individual units** of the product are being shipped.

### **Conclusion:**

When **Line Item Quantity** is less than **Unit of Measure (Per Pack)** in your dataset, it could mean:

1. The dataset uses different reference points (individual items vs packs).
2. It could be a data inconsistency or entry error, depending on how the data is structured.
3. Partial pack shipments might be the reason for the mismatch, or it might reflect a different type of measurement system being used.

If this happens frequently in your dataset, it may be worth checking if there’s any documentation or clarification on how these quantities are recorded.

—------------------------------------------------------------------------------------------------------------------------

Here's a detailed breakdown of each term commonly found in a logistics dataset, such as FedEx's, with context for their usage:

1. **ID**
   * **Explanation**: A unique identifier assigned to each record in the logistics database.
   * **Context**: Used to track individual shipments and as a reference for looking up specific transaction data.
2. **Project Code**
   * **Explanation**: A code representing the specific project linked to the shipment.
   * **Context**: Important in project-based logistics to identify shipments associated with a particular initiative or contract.
3. **PQ # (Price Quotation Number)**
   * **Explanation**: Reference number for a price quotation given by a supplier.
   * **Context**: Tracks procurement and negotiation stages for goods being shipped.
4. **PO / SO # (Purchase Order / Sales Order Number)**
   * **Explanation**: The purchase order or sales order number related to the shipment.
   * **Context**: Essential for managing transactions, tracking sales and purchases.
5. **ASN/DN # (Advanced Shipment Notice / Delivery Note Number)**
   * **Explanation**: Reference number for shipment notices (ASN) or delivery notes (DN).
   * **Context**: Allows shipment tracking and ensures transparency between buyer and seller regarding the delivery status.
6. **Country**
   * **Explanation**: The destination or origin country for the shipment.
   * **Context**: Important for customs, legal compliance, and international logistics management.
7. **Managed By**
   * **Explanation**: The department or team overseeing the shipment.
   * **Context**: Assigns responsibility for managing and tracking the logistics process within a company.
8. **Fulfill Via**
   * **Explanation**: Method used for fulfilling the order (e.g., direct shipment or via a fulfillment center).
   * **Context**: Indicates how the goods are processed and dispatched.
9. **Vendor INCO Term**
   * **Explanation**: International Commercial Terms (Incoterms) define the responsibilities of buyers and sellers, such as shipping and insurance.
   * **Context**: Specifies who bears the costs and risks at various stages of the shipping process (e.g., FOB, CIF).
10. **Shipment Mode**
    * **Explanation**: The mode of transport used for the shipment (air, sea, road, etc.).
    * **Context**: Determines the logistics plan, affecting shipping time and costs.
11. **PQ First Sent to Client Date**
    * **Explanation**: The date when the price quotation was initially sent to the client.
    * **Context**: Key for tracking procurement timelines and client interactions.
12. **PO Sent to Vendor Date**
    * **Explanation**: The date the purchase order was sent to the vendor.
    * **Context**: Helps monitor when the purchase process officially began.
13. **Scheduled Delivery Date**
    * **Explanation**: The planned delivery date for the shipment.
    * **Context**: Important for delivery planning, ensuring timely delivery.
14. **Delivered to Client Date**
    * **Explanation**: The date the shipment was actually delivered to the client.
    * **Context**: Used to verify whether shipments were made within the agreed timeframe.
15. **Delivery Recorded Date**
    * **Explanation**: The date the delivery was officially registered in the system.
    * **Context**: Marks the completion of the delivery and finalization in the logistics process.
16. **Product Group**
    * **Explanation**: A broad category that the product falls under (e.g., medical supplies, electronics).
    * **Context**: Used for grouping similar products, helping in inventory management and analysis.
17. **Sub Classification**
    * **Explanation**: A more specific categorization within the product group.
    * **Context**: Differentiates products with slight variations for finer management.
18. **Vendor**
    * **Explanation**: The supplier or manufacturer responsible for the product.
    * **Context**: Tracks which vendor provided the shipped goods.
19. **Item Description**
    * **Explanation**: A detailed description of the item.
    * **Context**: Provides clear information on the contents of the shipment.
20. **Molecule/Test Type**
    * **Explanation**: Specific type of product in pharmaceutical or medical shipments (e.g., HIV, Nevirapine).
    * **Context**: Important for distinguishing between different medical or test products.
21. **Brand**
    * **Explanation**: The brand of the product.
    * **Context**: Helps in identifying product recognition and related regulations.
22. **Dosage**
    * **Explanation**: The strength or concentration of a medical product.
    * **Context**: Essential for ensuring accurate medical or pharmaceutical dosage tracking.
23. **Dosage Form**
    * **Explanation**: The form in which the medical product is provided (e.g., tablet, liquid).
    * **Context**: Determines the method of use or administration of the product.
24. **Unit of Measure (Per Pack)**
    * **Explanation**: Specifies the number of units contained in each pack.
    * **Context**: Important for inventory control and shipment tracking.
25. **Line Item Quantity**
    * **Explanation**: The total quantity of the specific item being shipped.
    * **Context**: Crucial for logistics and inventory calculations.
26. **Line Item Value**
    * **Explanation**: The monetary value of the line item.
    * **Context**: Important for financial and cost analysis.
27. **Pack Price**
    * **Explanation**: The cost of a single pack of the item.
    * **Context**: A key metric for pricing and cost assessment.
28. **Unit Price**
    * **Explanation**: The price of a single unit of the product.
    * **Context**: Useful for understanding and comparing individual unit costs.
29. **Manufacturing Site**
    * **Explanation**: The location where the product was produced.
    * **Context**: Important for compliance with regulations, traceability, and quality control.
30. **First Line Designation**
    * **Explanation**: Indicates if the product is meant for primary treatment (first-line use).
    * **Context**: Common in pharmaceutical and medical logistics, where first-line treatments are prioritized.
31. **Weight (Kilograms)**
    * **Explanation**: The total weight of the shipment.
    * **Context**: Used for calculating shipping costs and ensuring adherence to weight limits.
32. **Freight Cost (USD)**
    * **Explanation**: The cost of shipping the goods, typically in USD.
    * **Context**: Essential for cost analysis and determining the profitability of the shipment.
33. **Line Item Insurance (USD)**
    * **Explanation**: The insurance cost for the individual line item.
    * **Context**: Helps protect against losses during transit.

4o

##### **You said:**

explain PQ # (Price Quotation Number) in details

###### **ChatGPT said:**

ChatGPT

The **PQ # (Price Quotation Number)** is a reference number assigned to a specific price quotation issued by a vendor or supplier in the logistics and procurement process. Here’s a detailed explanation:

### **Explanation**

A Price Quotation (PQ) is a formal document or offer made by a supplier or vendor to a potential buyer, outlining the prices, terms, and conditions for the goods or services being offered. The **PQ #** serves as a unique identifier for this document, ensuring that it can be easily tracked, referenced, and distinguished from other quotations in the system.

The PQ number is typically generated by the vendor or supplier's system and is provided to the buyer as part of the quotation process. It helps both the supplier and the buyer track and manage the details of the quotation, including item prices, delivery terms, and any special conditions tied to the offer.

### **Components of a PQ**

A typical price quotation might include:

* **Item description**: Details about the goods or services being quoted.
* **Quantity**: The number of units or total volume of goods.
* **Unit price**: The cost per unit or measure of the item.
* **Total price**: The overall cost based on the quantity and unit price.
* **Discounts or promotions**: Any special pricing or reductions applied to the quotation.
* **Payment terms**: Conditions under which payment is due (e.g., net 30 days).
* **Delivery terms**: Conditions such as Incoterms (e.g., EXW, CIF) and the expected delivery dates.
* **Validity period**: The time period during which the quoted price is valid.

### **Context and Usage**

* **Tracking and Reference**: The PQ # is primarily used to track specific price quotations in the procurement process. Both the buyer and the supplier can refer to this number to look up all related details of the quotation.
* **Negotiation Tool**: The PQ number allows for clear communication and negotiation. If there are updates to the price or terms, the updated PQ number helps distinguish the revised quotation from earlier versions.
* **Order Management**: If the buyer accepts the quotation, the PQ number may be linked to a Purchase Order (PO) number, which formally authorizes the transaction.
* **Audit and Record Keeping**: The PQ number is critical for accounting, auditing, and compliance purposes, ensuring that the details of every transaction are documented and traceable.

### **Role in Logistics**

In logistics, particularly in international or bulk transactions, the PQ number helps the buyer manage procurement costs and helps the supplier track their offers to various customers. It often plays a role in the supply chain where quotes are needed for budgeting, price comparison, or to ensure that products are available at the agreed prices before moving forward with shipping.

The **PQ #** is also valuable for financial and inventory management, as it links the cost details of goods to the overall financial record for the company.

### **Example Scenario**

Suppose a company is planning to import pharmaceutical products from a vendor. The vendor sends a quotation with detailed pricing for each product, including quantities, unit prices, shipping costs, and terms. The vendor assigns a PQ number, such as **PQ-2024-001**. The buyer references this number when deciding whether to accept the offer or negotiate further, and if they proceed, the PQ number becomes a key reference in the Purchase Order (PO) that authorizes the procurement.

### **Importance**

* **Transparency**: Helps maintain clear communication between suppliers and buyers.
* **Efficiency**: Simplifies tracking of multiple quotes and orders.
* **Compliance**: Ensures all legal and contractual requirements are met by keeping a record of the agreed terms.

In summary, the **PQ #** is a key element in procurement and logistics, providing a unique identifier for a price quotation and ensuring smooth management of the procurement process, from pricing to order fulfillment.

4o mini

##### **You said:**

PO / SO # (Purchase Order / Sales Order Number)

###### **ChatGPT said:**

ChatGPT

The **PO / SO # (Purchase Order / Sales Order Number)** is a crucial identifier in the logistics and supply chain process, representing either a **Purchase Order** (PO) or **Sales Order** (SO) issued by a company. These numbers play a significant role in managing and tracking orders, from procurement to final delivery.

### **1. Purchase Order (PO) Number**

* **Explanation**: A **Purchase Order (PO)** number is a unique identifier assigned to a document issued by a buyer (or purchasing department) to a supplier or vendor, indicating the intent to purchase goods or services at agreed-upon prices and terms.
* **Context**: This document authorizes the supplier to deliver the specified items and serves as the official record of the buyer's request for procurement.
* **Key Components of a PO**:
  + **Buyer Information**: Details about the purchasing company or department.
  + **Supplier Information**: Vendor or supplier's details.
  + **Item Description**: What goods or services are being purchased.
  + **Quantity**: How many units or how much of each product.
  + **Price**: The agreed cost per unit or service.
  + **Delivery Terms**: Expected shipping method, delivery dates, and any relevant terms (e.g., Incoterms).
  + **Payment Terms**: When and how payment will be made (e.g., net 30 days).
* **Usage in Logistics**:
  + **Order Tracking**: The PO # is used to track and manage the procurement process from order creation to goods receipt.
  + **Inventory Management**: It helps maintain accurate stock levels as the order moves through the supply chain.
  + **Budgeting & Accounting**: It ties the procurement process to financial tracking and ensures that costs are recorded and managed correctly.
* **Example**: If a company places an order for 500 units of a product with a supplier, the PO number (e.g., **PO-2024-5001**) is generated. This number is used to track the entire transaction, including delivery and payment.

### **2. Sales Order (SO) Number**

* **Explanation**: A **Sales Order (SO)** number is a unique identifier for a document generated by a seller in response to a customer order. It confirms that the seller has accepted the buyer's order and will deliver the specified goods or services under the agreed terms.
* **Context**: It is typically generated after a customer places an order with a company and is used to track the fulfillment of the sale.
* **Key Components of an SO**:
  + **Customer Information**: Details about the buyer or customer.
  + **Product/Service Information**: What products or services the customer is ordering.
  + **Quantity**: How many units of each item are requested.
  + **Price**: The cost agreed upon for the items or services.
  + **Delivery Details**: Expected delivery timelines and address.
  + **Payment Terms**: Terms for customer payment (e.g., advance payment, installment plans).
* **Usage in Logistics**:
  + **Order Fulfillment**: The SO number tracks the fulfillment process, ensuring that the correct products are picked, packed, and shipped to the customer.
  + **Shipping & Delivery**: It is used in coordination with shipping and tracking systems to ensure that the order is delivered accurately and on time.
  + **Customer Relationship Management**: The SO number helps in managing customer queries, disputes, or changes to the order.
* **Example**: A customer places an order for 300 units of a product with a company. The company generates a Sales Order, such as **SO-2024-3001**, and uses this number to track the shipment of goods to the customer.

### **Differences between PO and SO Numbers:**

* **PO Number** is generated by the buyer to place an order with the vendor or supplier.
* **SO Number** is created by the seller in response to the buyer's order, confirming the sale and beginning the fulfillment process.

### **Context and Importance in Logistics**

* **Tracking Orders**: Both PO and SO numbers provide a way to track orders throughout the supply chain, from order creation to payment and delivery.
* **Order Management**: PO numbers help vendors manage the goods or services they need to supply, while SO numbers help sellers manage customer orders and ensure correct fulfillment.
* **Financial and Inventory Control**: These numbers are linked to financial systems for budgeting, invoicing, and ensuring that inventory levels are managed appropriately.

### **Example Scenario:**

* A company places an order with a vendor for 100 units of product X. The vendor assigns a PO number (**PO-2024-1234**) to the order. Upon receiving the order, the vendor processes it, creates a sales order for the customer, and assigns an SO number (**SO-2024-5678**). These two numbers will be used to track the order as it moves through the supply chain, from vendor procurement to customer delivery.

### **Role in Logistics and Supply Chain**

* **Streamlines Communication**: Both PO and SO numbers act as references for communication between buyers, suppliers, and logistics teams, ensuring that everyone involved in the transaction has a clear point of reference.
* **Audits and Compliance**: These numbers help ensure compliance with procurement and sales policies by maintaining traceable records of every transaction.
* **Invoice and Payment Processing**: PO and SO numbers are critical when processing invoices and payments, linking them to specific orders.

In summary, the **PO / SO #** serves as a key identifier in the procurement and sales process, providing a method for tracking orders, managing inventory, and maintaining financial transparency in logistics operations.

4o mini

##### **You said:**

ASN/DN # (Advanced Shipment Notice / Delivery Note Number)

###### **ChatGPT said:**

ChatGPT

The **ASN/DN # (Advanced Shipment Notice / Delivery Note Number)** is a unique identifier used in logistics and supply chain management to track the movement of goods. It helps ensure that all parties involved (buyer, seller, and logistics provider) are aware of the shipment details, including contents, shipping method, and expected delivery times.

### **1. Advanced Shipment Notice (ASN) Number**

* **Explanation**: An **Advanced Shipment Notice (ASN)** is a document that provides detailed information about an upcoming shipment before the goods physically arrive at their destination. The ASN number is a unique identifier for this notice.
* **Context**: This notice is typically sent by the supplier or shipper to the recipient (buyer, warehouse, or distribution center) ahead of the actual delivery, allowing the recipient to prepare for receiving the goods.
* **Key Components of an ASN**:
  + **Shipment Details**: Information about the products, including SKU numbers, quantities, and descriptions.
  + **Estimated Arrival Date and Time**: Helps the recipient plan for the arrival of goods.
  + **Packaging Information**: Information about how the goods are packed (e.g., pallets, boxes).
  + **Carrier and Tracking Information**: Details about the transport method and carrier used for the shipment.
  + **Delivery Instructions**: Any special handling or delivery instructions.
* **Usage in Logistics**:
  + **Inventory Management**: The ASN allows recipients to update inventory records even before the goods arrive, ensuring that the warehouse is prepared.
  + **Pre-receiving Verification**: It allows the recipient to verify the contents of the shipment ahead of time and identify any discrepancies or issues before the goods are received.
  + **Streamlined Receiving Process**: Knowing the shipment’s arrival details helps streamline the unloading, checking, and storage process upon delivery.
* **Example**: A company ships 100 units of a product to a warehouse, and they send an ASN with the ASN number **ASN-2024-1001** to notify the warehouse of the incoming shipment. The warehouse uses this information to prepare for receiving the goods and ensure accuracy in the delivery process.

### **2. Delivery Note (DN) Number**

* **Explanation**: A **Delivery Note (DN)** is a document issued by the seller or shipper once the goods have been delivered. It serves as proof of delivery and includes a detailed list of the items delivered.
* **Context**: The DN is typically signed by the recipient to acknowledge receipt of the goods, confirming that they have been delivered as per the agreed terms. The DN number is a unique identifier for this document.
* **Key Components of a Delivery Note**:
  + **Recipient Information**: The name and address of the receiving party.
  + **Product List**: A detailed list of the items delivered, including quantities and descriptions.
  + **Delivery Date**: The date on which the goods were delivered.
  + **Delivery Instructions**: Any additional instructions regarding the delivery, such as handling or storage.
  + **Signature**: The recipient's signature, acknowledging receipt of the goods.
* **Usage in Logistics**:
  + **Proof of Delivery**: The DN serves as a formal record that the goods were delivered to the correct destination and received in good condition.
  + **Order Closure**: The DN is often used to close the order in the system, marking the goods as delivered.
  + **Discrepancy Resolution**: If there is a problem with the delivery (e.g., missing or damaged items), the DN serves as a reference for resolving discrepancies.
* **Example**: Once the 100 units of product are delivered to the warehouse, the delivery driver presents a Delivery Note with the DN number **DN-2024-5001**. The warehouse manager signs it, confirming that the shipment was received.

### **Context and Importance in Logistics**

* **Communication**: Both ASN and DN numbers improve communication between the parties involved in the transaction. The ASN informs the recipient of an impending shipment, and the DN confirms that the delivery has been completed.
* **Tracking and Verification**: These numbers provide a means to track shipments at different stages—before they arrive (ASN) and after they are delivered (DN). This helps resolve any issues quickly and efficiently.
* **Supply Chain Visibility**: The ASN provides visibility into the shipment's status, allowing warehouses and distribution centers to prepare for goods arrival. The DN offers visibility into the completion of the transaction and the receipt of goods.

### **Differences Between ASN and DN**

* **ASN**: Sent in advance of the shipment to inform the recipient of the goods that are on their way. It’s used to prepare for the shipment and ensure that it’s received accurately.
* **DN**: Issued after the goods have been delivered, serving as proof that the items have been delivered to the correct destination and received in good condition.

### **Example Scenario**

* A supplier is shipping 200 units of a product to a retail distribution center. They send an ASN (**ASN-2024-001**) to the distribution center, detailing the shipment contents, expected arrival time, and carrier information. When the shipment arrives, the distribution center receives it and signs a Delivery Note (**DN-2024-001**) to confirm the goods were delivered in full and in good condition.

### **Role in Logistics and Supply Chain**

* **Efficiency**: The ASN helps streamline the receiving process, while the DN ensures that there is official documentation for the completion of the transaction.
* **Dispute Resolution**: If a discrepancy arises (e.g., missing or damaged goods), the ASN and DN numbers help trace the issue back to its source, whether it's during shipment preparation or during delivery.
* **Financial Control**: These documents assist in managing invoices and payments, as the ASN provides details of the shipment, and the DN confirms that the goods were successfully delivered.

In summary, the **ASN/DN #** is a critical element in the logistics process. The ASN provides early notification of an incoming shipment, while the DN confirms the successful delivery of goods, both serving as key reference numbers for tracking, managing, and verifying shipments in the supply chain.

4o mini

##### **You said:**

Fulfill Via

###### **ChatGPT said:**

ChatGPT

The **"Fulfill Via"** field in a logistics or order management system refers to the method or channel used to fulfill a particular order or shipment. It describes the specific process or route through which goods are delivered to the customer or end recipient. This term is particularly important in supply chain management as it helps determine how orders are processed, shipped, and delivered, impacting efficiency, cost, and timing.

### **Explanation and Context of "Fulfill Via"**

The "Fulfill Via" field indicates whether the shipment is being processed directly from the vendor (e.g., direct shipping), from a warehouse (e.g., inventory fulfillment), or through another method such as a drop-shipment or fulfillment center. It provides clarity on the logistical flow of how products are delivered to the customer.

### **Key Fulfillment Methods (Common Values for "Fulfill Via")**

1. **Direct Drop Shipment**:
   * **Explanation**: The product is shipped directly from the vendor or manufacturer to the customer without being handled or stored by a third party (e.g., warehouse or retailer).
   * **Context**: Common in e-commerce or business-to-business transactions where the supplier handles the entire shipment directly to the end customer.
   * **Example**: A customer orders a laptop from an online retailer, and the retailer's supplier ships the laptop directly from their factory to the customer's address.
2. **Fulfillment Center**:
   * **Explanation**: Goods are shipped from a centralized location (a fulfillment center or warehouse) where they are stored until an order is placed.
   * **Context**: Common in retail, e-commerce, and distribution businesses where companies stock goods in warehouses and ship them to customers upon receiving orders.
   * **Example**: An online retailer stores various products in a fulfillment center, and once a customer orders a pair of shoes, the order is processed and shipped from the warehouse.
3. **Third-Party Logistics (3PL)**:
   * **Explanation**: A third-party logistics provider manages the storage, picking, packing, and shipping of goods for a company.
   * **Context**: Used when a company outsources its logistics and warehousing functions to a third-party service provider who handles fulfillment on their behalf.
   * **Example**: A company uses a 3PL service to store its products and ship them to customers when orders are placed, without maintaining its own warehouses.
4. **Drop-Ship via Supplier**:
   * **Explanation**: Similar to direct drop shipment, but explicitly involves a third-party supplier managing the fulfillment process, ensuring that products are shipped directly to the customer.
   * **Context**: Used by retailers or distributors who do not carry inventory but rely on suppliers to fulfill orders.
   * **Example**: An online retailer sells furniture but doesn’t stock the items. Instead, the supplier directly ships the product to the customer after an order is placed.
5. **In-House Fulfillment**:
   * **Explanation**: The company itself manages inventory and fulfillment operations from its own warehouse or store locations.
   * **Context**: This method is used by companies with sufficient resources or operations to handle the logistics and shipping internally.
   * **Example**: A fashion retailer who holds inventory in their store or warehouse fulfills orders directly to customers from that location.
6. **Ship to Store / Pickup In-Store (BOPIS)**:
   * **Explanation**: The order is fulfilled by shipping goods to a physical store, where customers can pick them up at their convenience.
   * **Context**: Often used by retailers to offer convenience to customers who prefer to pick up their orders instead of having them delivered to their homes.
   * **Example**: A customer orders clothes online from a retail chain, and instead of home delivery, the order is sent to a local store for the customer to pick up.
7. **Cross-Docking**:
   * **Explanation**: Cross-docking involves unloading goods from incoming shipments directly onto outbound vehicles with little or no storage in between.
   * **Context**: This method is often used to expedite the shipping process, reducing the need for warehousing.
   * **Example**: Products arrive at a distribution center, are quickly sorted, and then directly loaded onto delivery trucks for rapid delivery to customers.

### **Importance of "Fulfill Via" in Logistics**

* **Efficiency and Cost Management**: The method chosen for fulfillment directly impacts the cost, timing, and complexity of logistics operations. For example, drop shipping might be cheaper but can take longer, while fulfillment from a warehouse might be faster but involves additional handling costs.
* **Customer Experience**: The fulfillment method influences the delivery time and service level. Faster fulfillment methods (e.g., same-day shipping from a warehouse) enhance customer satisfaction, whereas slower methods (e.g., dropshipping) may increase customer frustration.
* **Inventory Management**: Different fulfillment methods may require varying levels of inventory control. A company using fulfillment centers or third-party services must manage stock levels across multiple locations, while a direct drop shipment model requires minimal inventory management on the seller’s end.
* **Flexibility**: The "Fulfill Via" field also allows companies to choose the most appropriate fulfillment method based on customer needs, available infrastructure, and supply chain flexibility. It helps adapt to different business models, such as e-commerce, retail, and wholesale.

### **Example Scenario**

* **Fulfill Via: Fulfillment Center**: A company selling beauty products through an online store uses a fulfillment center located in a central hub. When a customer orders a skincare kit, the order is processed and shipped from the fulfillment center to the customer's address.
* **Fulfill Via: Direct Drop Shipment**: An online electronics retailer partners with a supplier to ship laptops directly to customers. Once a customer places an order on the website, the retailer sends the order details to the supplier, who ships the laptop directly to the customer’s address without the retailer holding any inventory.

### **Conclusion**

The **"Fulfill Via"** field provides insight into the method used to fulfill an order, which directly affects shipping times, costs, and overall customer satisfaction. It is a vital piece of information in logistics and supply chain management, allowing businesses to optimize their fulfillment processes based on the best method suited for each situation.

4o mini

##### **You said:**

Vendor INCO Term

###### **ChatGPT said:**

ChatGPT

**Vendor INCO Term** refers to the specific **Incoterms (International Commercial Terms)** agreed upon between the buyer and seller (vendor) that govern the responsibilities of each party in an international transaction, particularly regarding shipping, risk, and costs. The term indicates who is responsible for various aspects of the shipment, such as delivery, transportation costs, and insurance, as well as when the responsibility and risk for the goods are transferred from the seller to the buyer.

### **Explanation of "Vendor INCO Term"**

* **Vendor**: The company or supplier that provides goods or services to the buyer.
* **INCO Term**: A standard set of international trade terms used to specify the responsibilities of each party involved in the shipping process (e.g., seller’s responsibilities, buyer’s responsibilities, and points of transfer for risk and costs).

### **Context of INCO Terms in Logistics**

INCO Terms are used to reduce confusion and legal disputes in international trade by establishing clear rules regarding the responsibilities and costs for the seller and buyer. These terms clarify when and where goods should be delivered, and who covers the costs and risks associated with transportation, insurance, customs clearance, and more.

### **Key Incoterms (Common INCO Terms)**

Here are some of the most commonly used **Incoterms** that would be listed as a **Vendor INCO Term**:

1. **EXW (Ex Works)**:
   * **Explanation**: The seller makes the goods available at their premises or another agreed location. The buyer is responsible for all costs and risks associated with transportation, including export duties and freight charges.
   * **Responsibility**: The buyer assumes responsibility for the entire shipment process from the seller's location.
   * **Example**: A seller in Germany makes a product available at their factory. The buyer arranges and pays for all transportation from that point to their destination.
2. **FOB (Free On Board)**:
   * **Explanation**: The seller is responsible for the goods until they are loaded onto the shipping vessel at the port of shipment. The buyer assumes responsibility once the goods are onboard and covers all subsequent costs, including freight, insurance, and delivery.
   * **Responsibility**: Seller’s responsibility ends when goods are loaded onto the ship at the port. Buyer is responsible for shipping and import duties.
   * **Example**: A vendor in India sells steel pipes FOB at the port in Mumbai. The buyer takes ownership when the goods are loaded on the ship.
3. **CIF (Cost, Insurance, and Freight)**:
   * **Explanation**: The seller is responsible for the cost, insurance, and freight to transport goods to the destination port. However, the risk is transferred to the buyer once the goods are loaded onto the vessel.
   * **Responsibility**: Seller pays for freight and insurance, but the risk transfers to the buyer as soon as the goods are on board.
   * **Example**: A supplier in the US sells medical equipment to a hospital in Europe, covering shipping and insurance costs, but the risk transfers to the hospital as soon as the goods are loaded on the ship.
4. **CFR (Cost and Freight)**:
   * **Explanation**: Similar to CIF, but the seller is not required to purchase insurance for the goods. The seller covers the cost and freight to the destination port, while the buyer assumes risk once the goods are on board.
   * **Responsibility**: Seller is responsible for costs and freight to the destination port, but the buyer handles insurance and risk once the goods are shipped.
   * **Example**: A vendor in China ships electronics to a buyer in Africa under CFR terms. The vendor pays for transportation, but the buyer takes responsibility for insurance once the goods are on board the vessel.
5. **DAP (Delivered At Place)**:
   * **Explanation**: The seller is responsible for all costs and risks associated with delivering the goods to a specified location in the buyer’s country. The buyer is responsible for import duties and taxes.
   * **Responsibility**: The seller is responsible for the goods all the way to the agreed destination. The buyer is responsible for customs clearance and import duties.
   * **Example**: A vendor in Japan delivers auto parts to a buyer in the UK. The vendor covers the cost of shipping, but the buyer handles any customs duties and import taxes upon arrival.
6. **DDP (Delivered Duty Paid)**:
   * **Explanation**: The seller takes on all responsibilities and costs associated with delivering the goods to the buyer’s location, including shipping, insurance, and import duties.
   * **Responsibility**: The seller assumes full responsibility for the goods until they reach the buyer’s location, including all customs formalities and duties.
   * **Example**: A vendor in the USA delivers machinery to a buyer in South America under DDP terms. The seller handles everything, including shipping, customs clearance, and import duties.
7. **FCA (Free Carrier)**:
   * **Explanation**: The seller is responsible for delivering goods to a carrier or another nominated person (such as a freight forwarder) at a specified place. After that, the risk is transferred to the buyer.
   * **Responsibility**: The seller arranges transport to the carrier, and the buyer assumes responsibility after the goods are handed over to the carrier.
   * **Example**: A vendor in Canada ships medical supplies to a buyer in Mexico and delivers the goods to a freight forwarder (the nominated carrier) at the airport.
8. **FAS (Free Alongside Ship)**:
   * **Explanation**: The seller’s responsibility ends once the goods are placed alongside the ship at the port of shipment. The buyer assumes responsibility for loading the goods onto the vessel and covering the costs of transportation and insurance.
   * **Responsibility**: Seller is responsible for getting the goods to the port and alongside the vessel; the buyer handles the rest.
   * **Example**: A supplier in Brazil delivers soybeans to the port of Santos, where the buyer arranges for the shipment to be loaded onto the vessel.

### **Importance of Vendor INCO Term in Logistics**

1. **Cost Allocation**: The **Vendor INCO Term** helps define which party will bear the cost of shipping, handling, insurance, and any other logistics-related expenses.
2. **Risk Management**: These terms determine when the risk is transferred from the seller to the buyer, ensuring both parties understand when they are liable for potential damage or loss.
3. **Customs Compliance**: It establishes who is responsible for customs clearance, import duties, and taxes, which is crucial for international shipments.
4. **Clear Responsibilities**: Using Incoterms reduces confusion by clearly defining the duties and responsibilities of both parties involved in the transaction, ensuring smooth communication and reducing disputes.

### **Example Scenario**

* **Vendor INCO Term: CIF**: A company in the United States sells textiles to a buyer in Italy under CIF terms. The seller arranges and pays for transportation, including insurance, up to the Italian port of Naples. However, the risk is transferred to the buyer as soon as the goods are loaded onto the ship in the U.S. The buyer then handles customs clearance and any further transportation costs from the port to their warehouse.

In conclusion, the **Vendor INCO Term** defines the terms of the sale and delivery in international trade, helping to establish the logistics and financial responsibilities between the seller and buyer. This ensures clarity and prevents misunderstandings related to shipping costs, risk, insurance, and customs.

4o mini

##### **You said:**

is Line Item Quantity the amount of packs or the total no. of items in all the packs

###### **ChatGPT said:**

ChatGPT

**Line Item Quantity** typically refers to the **total number of items** in all the packs included in the shipment, rather than just the number of packs.

### **Explanation:**

* **Line Item Quantity** is the total count of individual units or items being shipped, regardless of how many packs or containers they are packed into.

### **Example:**

* If you are shipping 100 units of a product, and those 100 units are packed into 10 boxes (packs), the **Line Item Quantity** would be **100** (the total number of individual units), not **10** (the number of boxes).

Thus, **Line Item Quantity** is useful for understanding the full volume of items in the shipment, while **pack quantity** refers to the packaging or containerization of those items.

4o mini

##### **You said:**

how in this dataset is line itme quantity less than unit of measure (per pack) Dosage Form Unit of Measure (Per Pack) Line Item Quantity Line Item Value Pack Price Unit Price Manufacturing Site First Line Designation Weight (Kilograms) Freight Cost (USD) Line Item Insurance (USD) 0 1 100-CI-T01 Pre-PQ Process SCMS-4 ASN-8 Côte d'Ivoire PMO - US Direct Drop EXW Air Pre-PQ Process Date Not Captured 2-Jun-06 2-Jun-06 2-Jun-06 HRDT HIV test RANBAXY Fine Chemicals LTD. HIV, Reveal G3 Rapid HIV-1 Antibody Test, 30 T... HIV, Reveal G3 Rapid HIV-1 Antibody Test Reveal NaN Test kit 30 19 551.00 29.00 0.97 Ranbaxy Fine Chemicals LTD Yes 13 780.34 NaN 1 3 108-VN-T01 Pre-PQ Process SCMS-13 ASN-85 Vietnam PMO - US Direct Drop EXW Air Pre-PQ Process Date Not Captured 14-Nov-06 14-Nov-06 14-Nov-06 ARV Pediatric Aurobindo Pharma Limited Nevirapine 10mg/ml, oral suspension, Bottle, 2... Nevirapine Generic 10mg/ml Oral suspension 240 1000 6200.00 6.20 0.03 Aurobindo Unit III, India Yes 358 4521.5 NaN 2 4 100-CI-T01 Pre-PQ Process SCMS-20 ASN-14 Côte d'Ivoire PMO - US Direct Drop FCA Air Pre-PQ Process Date Not Captured 27-Aug-06 27-Aug-06 27-Aug-06 HRDT HIV test Abbott GmbH & Co. KG HIV 1/2, Determine Complete HIV Kit, 100 Tests HIV 1/2, Determine Complete HIV Kit Determine NaN Test kit 100 500 40000.00 80.00 0.80 ABBVIE GmbH & Co.KG Wiesbaden Yes 171 1653.78 NaN 3 15 108-VN-T01 Pre-PQ Process SCMS-78 ASN-50 Vietnam PMO - US Direct Drop EXW Air Pre-PQ Process Date Not Captured 1-Sep-06 1-Sep-06 1-Sep-06 ARV Adult SUN PHARMACEUTICAL INDUSTRIES LTD (RANBAXY LAB... Lamivudine 150mg, tablets, 60 Tabs Lamivudine Generic 150mg Tablet 60 31920 127360.80 3.99 0.07 Ranbaxy, Paonta Shahib, India

###### **ChatGPT said:**

ChatGPT

In the dataset, **Line Item Quantity** being less than **Unit of Measure (Per Pack)** could indicate a few scenarios or data issues. Here’s a detailed breakdown of what could be happening:

### **Key Definitions:**

1. **Unit of Measure (Per Pack)**: This refers to the number of individual units of the product in each pack. For example, it might state that each pack contains 30 units.
2. **Line Item Quantity**: This refers to the **total number of units** of the product being shipped, which might not directly correspond to the pack size. It's the overall total quantity ordered or shipped for that particular line item.

### **Possible Scenarios:**

1. **Line Item Quantity Could Be Less Than Unit of Measure (Per Pack) in Data Mistakes**:
   * If **Line Item Quantity** is less than the **Unit of Measure (Per Pack)**, this might be an error or inconsistency in how the data was recorded or processed. For instance, if the dataset states there are 100 units being shipped, but the **Unit of Measure (Per Pack)** suggests that each pack holds 240 units, the data entry might have a mismatch.
2. **Line Item Quantity Represents Individual Units or Partial Packs**:
   * Sometimes, **Line Item Quantity** might refer to the total number of individual units ordered, not packs, and these individual units may not be a full pack.
   * Example: If each pack contains 240 items (**Unit of Measure (Per Pack)**), and the **Line Item Quantity** is 100, this means the shipment is for **100 individual units** (not whole packs). This could happen when only part of a pack is being shipped, or the dataset is structured to list the individual quantity regardless of how they are packed.
3. **Unit of Measure for Pack vs. Shipment**:
   * It's also possible that **Unit of Measure (Per Pack)** refers to a larger quantity (e.g., 240 units per pack) but the **Line Item Quantity** indicates a smaller quantity that is intended for shipment. If this is a case of partial shipment or different packaging configurations, the two numbers might not align directly.

### **Example from Dataset:**

* For **Line Item Quantity = 100** and **Unit of Measure (Per Pack) = 240**, it could imply:
  + **100** total individual units are being shipped.
  + **240** is how many units each individual pack contains.
  + The shipment consists of fewer than 1 full pack in terms of **packs**, but **100 individual units** of the product are being shipped.

### **Conclusion:**

When **Line Item Quantity** is less than **Unit of Measure (Per Pack)** in your dataset, it could mean:

1. The dataset uses different reference points (individual items vs packs).
2. It could be a data inconsistency or entry error, depending on how the data is structured.
3. Partial pack shipments might be the reason for the mismatch, or it might reflect a different type of measurement system being used.

If this happens frequently in your dataset, it may be worth checking if there’s any documentation or clarification on how these quantities are recorded.

4o mini

##### **You said:**

so the line item quantity can never be less than unit of measure (per pack)

###### **ChatGPT said:**

ChatGPT

Correct, under typical circumstances, **Line Item Quantity** should never be less than **Unit of Measure (Per Pack)** if **Line Item Quantity** refers to the total number of individual units being shipped, and **Unit of Measure (Per Pack)** refers to the number of units per pack.

To clarify:

* **Unit of Measure (Per Pack)**: This is the number of units contained in a single pack. For example, a **pack** might contain **240 units**.
* **Line Item Quantity**: This should represent the **total number of individual units** being shipped. For example, if 100 packs are being shipped and each pack contains 240 units, the **Line Item Quantity** would be **100 × 240 = 24,000 units**.

### **Therefore, the Line Item Quantity should always be greater than or equal to the Unit of Measure (Per Pack) unless:**

1. **The shipment is partial**: For example, if the data is tracking an order where only a partial quantity of the pack is being shipped (though this would still typically be handled by recording a fraction of a pack).
2. **Data errors**: If there is an inconsistency in data entry (e.g., a typo, incorrect units, or a misunderstanding in how quantities are being recorded), you might see **Line Item Quantity** less than **Unit of Measure (Per Pack)**. However, this would usually indicate a mistake rather than a legitimate case.

### **In Conclusion:**

* **Under normal conditions**, **Line Item Quantity** should **never** be less than **Unit of Measure (Per Pack)**, as this would violate the logical structure of how items are packaged and shipped.
* If you encounter such a situation, it would be advisable to check the data for possible inconsistencies or errors.

—------------------------------------------------------------------------------------------------------------------------

how can you do shipment terms analysis and can weight contribute too which shipment mode is used for the product and delays maening more weight means more time to deliver

###### **ChatGPT said:**

ChatGPT

**Shipment Terms Analysis** and **Weight’s Contribution to Shipment Mode and Delays** are both interesting areas that can be explored in your dataset. Here's how you can approach both analyses:

### **1. Shipment Terms Analysis (Vendor INCO Term):**

The **INCO Terms** (International Commercial Terms) define the responsibilities of buyers and sellers for the delivery of goods under a sales contract. These terms can influence shipping costs, delivery times, and other logistical aspects. You can analyze how different **INCO Terms** impact shipping delays and costs by following these steps:

#### **Steps for Shipment Terms Analysis:**

* **Group by INCO Terms**: Group the data by Vendor INCO Term and analyze the distribution of key metrics such as delivery time (difference between Scheduled Delivery Date and Delivered to Client Date), shipping cost (Freight Cost (USD)), and the number of items in the shipment.

**Average Delivery Delay**: Calculate the average shipping delay for each INCO term. Some INCO terms may involve more complex logistics (e.g., Free On Board - FOB vs. Delivered Duty Paid - DDP), which can affect shipping times.  
python  
Copy code  
# Calculate delivery delay

fedx\_df\_copy['Scheduled Delivery Date'] = pd.to\_datetime(fedx\_df\_copy['Scheduled Delivery Date'])

fedx\_df\_copy['Delivered to Client Date'] = pd.to\_datetime(fedx\_df\_copy['Delivered to Client Date'])

fedx\_df\_copy['Delivery Delay'] = (fedx\_df\_copy['Delivered to Client Date'] - fedx\_df\_copy['Scheduled Delivery Date']).dt.days

# Group by INCO Term and analyze delivery delay

inco\_terms\_analysis = fedx\_df\_copy.groupby('Vendor INCO Term')['Delivery Delay'].mean().reset\_index()

print(inco\_terms\_analysis)

**Shipping Cost Comparison**: Compare the average Freight Cost (USD) for different INCO terms to see which terms are associated with higher or lower shipping costs.  
python  
Copy code  
inco\_terms\_shipping\_cost = fedx\_df\_copy.groupby('Vendor INCO Term')['Freight Cost (USD)'].mean().reset\_index()

print(inco\_terms\_shipping\_cost)

**Correlation with Other Factors**: Analyze the relationship between **INCO Terms** and other columns like Shipment Mode, Line Item Value, and Weight (Kilograms) to assess if there are any patterns.  
python  
Copy code  
# Correlation analysis

inco\_terms\_correlation = fedx\_df\_copy.groupby('Vendor INCO Term')[['Weight (Kilograms)', 'Line Item Value', 'Shipment Mode']].mean()

print(inco\_terms\_correlation)

#### **Insights You Can Gain:**

* Some INCO Terms may result in longer delivery times or higher costs due to different responsibilities for transportation.
* Certain terms might be associated with particular shipment modes or destinations that affect delivery speed or cost.

### **2. Weight Contribution to Shipment Mode and Delays:**

The weight of the products can impact the choice of **Shipment Mode** (Air, Sea, Truck) and possibly the delivery time (heavier items might require more time to transport). You can explore how the **Weight (Kilograms)** of products influences shipping mode and delays.

#### **Steps for Analyzing Weight's Impact:**

**Group by Shipment Mode**: Group the dataset by Shipment Mode and analyze the average weight of shipments in each mode. Heavier shipments may tend to use certain shipment modes over others (e.g., Air for lighter, urgent shipments vs. Sea for heavier, non-urgent shipments).  
python  
Copy code  
shipment\_mode\_weight = fedx\_df\_copy.groupby('Shipment Mode')['Weight (Kilograms)'].mean().reset\_index()

print(shipment\_mode\_weight)

**Shipping Delay vs Weight**: Investigate whether heavier shipments experience more significant delivery delays by plotting or calculating the correlation between Weight (Kilograms) and Delivery Delay.  
python  
Copy code  
# Correlation between weight and delivery delay

weight\_delay\_corr = fedx\_df\_copy[['Weight (Kilograms)', 'Delivery Delay']].corr()

print(weight\_delay\_corr)

# Alternatively, plot the relationship

import matplotlib.pyplot as plt

plt.scatter(fedx\_df\_copy['Weight (Kilograms)'], fedx\_df\_copy['Delivery Delay'])

plt.xlabel('Weight (Kilograms)')

plt.ylabel('Delivery Delay (Days)')

plt.title('Shipping Delay vs Weight')

plt.show()

**Weight vs Shipment Mode and Delay**: Check if **heavier shipments** tend to have **longer delays** with certain **shipment modes**.  
python  
Copy code  
shipment\_mode\_delay = fedx\_df\_copy.groupby(['Shipment Mode'])[['Weight (Kilograms)', 'Delivery Delay']].mean().reset\_index()

print(shipment\_mode\_delay)

#### **Insights You Can Gain:**

* **Heavier shipments** might be associated with **slower shipping modes** (e.g., Sea or Truck) and longer delays.
* **Lighter shipments** may tend to use **faster shipping methods** (e.g., Air) with shorter delays.
* If there’s a **positive correlation** between weight and delay, you could infer that heavier shipments might take more time to deliver.

### **Example Flow to Get Insights:**

1. **INCO Term Analysis**:
   * Understand the average delivery delay and cost by INCO Term.
   * Check for patterns in terms of cost or delivery speed across different INCO Terms.
2. **Weight Impact on Shipment Mode and Delay**:
   * Analyze which shipment modes tend to be used for heavier products.
   * Investigate if heavier shipments correlate with longer delivery delays.

These analyses can provide insights into how shipment terms and product weight affect shipping performance, cost, and delivery delays.

—------------------------------------------------------------------------------------------------------------------------

what about line item insurance

###### **ChatGPT said:**

ChatGPT

**Line Item Insurance** (or **Line Item Insurance (USD)**) is another key column that can play an important role in your logistics and financial analysis, especially in the context of **shipping cost** and **product protection**. Here's how you can integrate **Line Item Insurance** into your analysis:

### **1. Line Item Insurance and Shipping Costs:**

Line Item Insurance is typically used to insure the value of the goods being shipped. It could be useful to analyze the relationship between the insurance cost and the overall shipment cost (including the **Freight Cost** and **Line Item Value**).

#### **Steps for Analysis:**

**Total Shipping Cost**: You can create a new column to represent the total cost of the shipment, including the **Line Item Value**, **Freight Cost (USD)**, and **Line Item Insurance (USD)**. This will allow you to analyze the overall cost of each shipment.  
python  
Copy code  
fedx\_df\_copy['Total Shipping Cost'] = fedx\_df\_copy['Line Item Value'] + fedx\_df\_copy['Freight Cost (USD)'] + fedx\_df\_copy['Line Item Insurance (USD)']

**Average Insurance Cost by Shipment Mode**: You can check how much insurance is paid on average for different **Shipment Modes** (Air, Sea, Truck) to see if certain shipment modes involve higher insurance costs.  
python  
Copy code  
insurance\_by\_mode = fedx\_df\_copy.groupby('Shipment Mode')['Line Item Insurance (USD)'].mean().reset\_index()

print(insurance\_by\_mode)

**Insurance vs Total Shipping Cost**: Investigate how **Line Item Insurance** compares to the **total shipping cost** for each shipment, and whether it represents a significant portion of the total cost.  
python  
Copy code  
insurance\_percentage = (fedx\_df\_copy['Line Item Insurance (USD)'] / fedx\_df\_copy['Total Shipping Cost']) \* 100

fedx\_df\_copy['Insurance Percentage'] = insurance\_percentage

print(fedx\_df\_copy[['ID', 'Line Item Insurance (USD)', 'Total Shipping Cost', 'Insurance Percentage']].head())

#### **Insights You Can Gain:**

* You can see whether insurance significantly contributes to the **total shipping cost**.
* The analysis can reveal if there’s a pattern in **higher insurance costs** for certain **shipment modes** (e.g., Air freight might have higher insurance costs due to the value or fragility of goods).
* This can also indicate if **more expensive products** (higher **Line Item Value**) tend to have higher **insurance costs**.

### **2. Missing Values in Line Item Insurance:**

If you have missing values in **Line Item Insurance** (as in your case), it’s important to determine how to handle them, especially since **insurance** is crucial for risk management in logistics. Here are a few strategies:

**Remove rows with missing insurance**: If the **Line Item Insurance** is essential to your analysis, you might want to remove the rows with missing values. However, ensure that this won’t significantly reduce the dataset or introduce bias in the analysis.  
python  
Copy code  
fedx\_df\_copy\_clean = fedx\_df\_copy.dropna(subset=['Line Item Insurance (USD)'])

* **Impute missing values**: If you don't want to drop rows, you can fill missing **insurance** values using strategies like:

**Mean/Median Imputation**: Fill missing insurance values with the mean or median of the non-missing values.  
python  
Copy code  
fedx\_df\_copy['Line Item Insurance (USD)'] = fedx\_df\_copy['Line Item Insurance (USD)'].fillna(fedx\_df\_copy['Line Item Insurance (USD)'].median())

* + **Domain Knowledge**: If possible, impute missing insurance values based on **Line Item Value** (for instance, insurance may be proportional to the value of the goods being shipped).

### **3. Line Item Insurance and Shipment Delays:**

You can also explore if there’s any correlation between **Line Item Insurance** and **Delivery Delays**. For instance, higher-value goods might be insured more, and you may observe patterns where **higher-value shipments** (and therefore higher insurance costs) are associated with **longer delivery times** due to the complexity of handling or additional security requirements.

#### **Steps for Analysis:**

**Correlation between Insurance and Delivery Delay**: Analyze if there’s any correlation between **Line Item Insurance** and the **Delivery Delay**.  
python  
Copy code  
insurance\_delay\_corr = fedx\_df\_copy[['Line Item Insurance (USD)', 'Delivery Delay']].corr()

print(insurance\_delay\_corr)

**Plot Insurance vs Delay**: A scatter plot or boxplot can be helpful in visually exploring if higher insurance amounts tend to correlate with higher delivery delays.  
python  
Copy code  
import matplotlib.pyplot as plt

plt.scatter(fedx\_df\_copy['Line Item Insurance (USD)'], fedx\_df\_copy['Delivery Delay'])

plt.xlabel('Line Item Insurance (USD)')

plt.ylabel('Delivery Delay (Days)')

plt.title('Line Item Insurance vs Delivery Delay')

plt.show()

#### **Insights You Can Gain:**

* You may discover that **higher insurance** is associated with **more complex shipping** that may result in **longer delays**.
* Conversely, there could be a situation where goods with **higher insurance** are shipped more urgently, and thus have **shorter delivery delays**.

### **4. Line Item Insurance and Vendor or Country Analysis:**

You can also look at how **Line Item Insurance** varies across different **Vendors** or **Countries**. Certain vendors or countries may have higher insurance premiums due to the risk associated with shipping to specific regions or due to the nature of the goods.

#### **Steps for Analysis:**

**Insurance by Vendor**: Check the average insurance paid for different **Vendors** to see if there are any vendors with consistently high insurance costs.  
python  
Copy code  
insurance\_by\_vendor = fedx\_df\_copy.groupby('Vendor')['Line Item Insurance (USD)'].mean().reset\_index()

print(insurance\_by\_vendor)

**Insurance by Country**: Similarly, you can analyze how **Line Item Insurance** varies across different **Countries**.  
python  
Copy code  
insurance\_by\_country = fedx\_df\_copy.groupby('Country')['Line Item Insurance (USD)'].mean().reset\_index()

print(insurance\_by\_country)

#### **Insights You Can Gain:**

* Some **Vendors** might have higher insurance costs due to the value or risk associated with their products.
* Certain **Countries** might be associated with higher shipping risks, which could lead to higher insurance costs.

### **Summary of What You Can Do with Line Item Insurance:**

* **Analyze its impact on overall shipping cost** by calculating its contribution to the total cost of shipping.
* **Identify patterns in insurance costs** across different **shipment modes**, **vendors**, or **countries**.
* **Handle missing values** in insurance carefully (either drop or impute based on business logic).
* **Explore correlations** between **insurance** and **delivery delays**, as well as other numerical columns like **Line Item Value**.

By performing these analyses, you can gain valuable insights into the logistics and cost structure of shipments and how insurance impacts both the financial and logistical aspects of the process.