

To ensure complete, accurate, and reliable data in the anticipated datasets (Customer, Order, and Shipping), I recommend outlining the following requirements for each dataset:

1. Customer Dataset Requirements

The Customer dataset should provide complete, consistent, and accurate information about customers, enabling reliable tracking and analysis of orders and shipments. The following data components are essential:

Customer_ID (Primary Key): Unique identifier for each customer.

Name: Full name of the customer.

Age: Accurate age information for demographic analysis.

Country: Customer's country for regional analysis.

Email Address: For communication and identification.

Phone Number: To support contact verification and shipping queries.

Customer_Status: (e.g., Active, Inactive) for segmentation.

Registration Date: Date when the customer account was created.

Last Order Date: The most recent order placed by the customer (to identify active/inactive customers).

Payment Method: Preferred payment method (to analyze customer preferences).

Data Quality Considerations:

Ensure no Customer_ID duplication.

Validate contact information (email and phone) for completeness and accuracy.

Verify that country names or codes are standardized.

2. Order Dataset Requirements

The Order dataset should provide complete, accurate, and traceable information regarding customer orders. The necessary components include:

Order_ID (Primary Key): Unique identifier for each order.

Customer_ID (Foreign Key): Link to the Customer table.

Order_Date: Date the order was placed.

Product_ID: Identifier for the product ordered (can be linked to a product catalog).

Quantity: Number of units ordered.

Order_Amount: Total monetary value of the order.

Payment Status: (e.g., Paid, Unpaid) to track order completion.

Order_Status: (e.g., Confirmed, Shipped, Delivered, Canceled) to trace the order lifecycle.

Delivery_Method: Method used for delivering the order.

Shipping_ID (Foreign Key): Link to the Shipping dataset to track shipment details.

Data Quality Considerations:

Ensure each Order_ID is unique and linked to valid Customer_ID.

Validate that the order statuses are updated regularly to avoid stale or misleading data.

Ensure accuracy in order amount calculations, especially with respect to discounts or taxes.

3. Shipping Dataset Requirements

The Shipping dataset must accurately capture shipping statuses and link to relevant customer orders to allow end-to-end tracking of shipments. Key components:

Shipping_ID (Primary Key): Unique identifier for each shipment.

Customer_ID (Foreign Key): Link to the Customer table to identify which customer placed the order.

Shipping_Status: Current status of the shipment (e.g., Pending, In Transit, Delivered, Delayed, Canceled).

Shipment_Date: The date when the shipment was dispatched.

Delivery_Date: The actual or estimated date of delivery.

Carrier: Name of the shipping company handling the delivery.

Tracking_Number: Unique number for shipment tracking purposes.

Shipping_Method: (e.g., Standard, Express, Same-Day).

Shipping_Cost: The cost associated with the shipment.

Order_ID (Foreign Key): Link to the Order dataset for order-tracking purposes.

Data Quality Considerations:

Ensure no missing or inconsistent shipping statuses.

Provide timestamps for status updates to ensure accurate lifecycle tracking.

Avoid duplicate or redundant entries (especially when a shipment is pending for a long time).

4. Anticipated Datasets Structure

To integrate all three datasets effectively, the following relationships should be defined:

Customer → Order: One customer can have multiple orders. (Customer_ID in both tables should ensure data consistency.)

Order → Shipping: One order can correspond to multiple shipments, especially for large orders. (Order_ID in both tables should ensure traceability.)

5. General Data Requirements

Timestamp Data: Ensure that all dates and times (e.g., `Order_Date`, `Shipment_Date`) are recorded with proper time zones and precision to avoid confusion, especially in international transactions.

Data Validation & Cleaning: Implement validation rules (e.g., for email, phone, addresses) and regularly clean the data to remove duplicates or outdated information.

Data Consistency & Standardization: Enforce standardization (e.g., country codes, status values) to prevent discrepancies.

Data Updates & Freshness: Ensure real-time or near-real-time updates for key fields (e.g., `Shipping_Status`) to maintain data accuracy.

Error Logging & Handling: Track any issues, such as missing or delayed shipments, in error logs or audit trails for troubleshooting.