**Data Cleaning & Revenue Logic (Final)**

**Data Cleaning Steps**

* Removed blank rows and duplicate records across the dataset.
* Standardized inconsistent values (e.g., variations in “Cancelled,” “Pending,” etc.) and unified text capitalization.
* Created a **Flagged** helper column:
  + **OK** → records valid for analysis.
  + **Anomaly** → flagged when Quantity = 0 with any non-zero Amount or when both Quantity = 0 and Amount = 0.
* Cleaned and standardized categorical columns (Status, Category, Style, etc.) to ensure consistency.
* Adjusted data types where needed for correct analysis.

**Revenue Calculation Rules**

* **Excluded**: Cancelled, Returned, Damaged/Lost, and Pending orders.
* **Included**: Delivered, Shipped, and In Transit (considered valid revenue because many are prepaid, even if not delivered yet).
* Orders where **Quantity = 0 and Amount > 0 (or = 0)** were flagged as anomalies and excluded.
* Orders where **Quantity > 0 but Amount = 0**:
  + If status = Cancelled, Returned, Damaged, Lost → kept as valid (can happen with zero refunds).
  + If status = Shipped, In Transit, or Delivered → kept as valid (can reflect promotions/free orders, does not affect revenue).