

## Data Issue Log

The data issue log below summarizes the key data quality issues identified and the actions taken to resolve them.

COLUMN	ISSUE	DESCRIPTION	DECISION
SHIPMENT_ID	not unique across rows	Shipment_id represents shipment header, not line-level	Created surrogate key "shipment_line_id" using: = <b>shipment_id</b> & "-" & <b>ROW ()</b>
SUPPLIER_NAME	Extra Spaces	Leading/trailing and hidden spaces	Created "supplier_name_clean" using = <b>TRIM(CLEAN(C2))</b>
COUNTRY	Text inconsistency	Different casing & spelling errors	Standardized case using = <b>PROPER(TRIM(CLEAN(E2)))</b> . Then corrected spelling using <b>Mapping Table &amp; XLOOKUP</b>
CITY	Case inconsistency	Upper/lower case mix	Created "city_clean" using = <b>PROPER(TRIM(CLEAN(H2)))</b> .
WAREHOUSE_ID	Case & Format inconsistency	Upper/lower case mix & _ Vs -	Standardized case using = <b>UPPER(TRIM(CLEAN(J2)))</b> . For Format inconsistency create new col "warehouse_id_final" and then apply <b>SUBSTITUTE</b> to replace _ into -
PRODUCT	Case inconsistency	Upper/lower case mix	Created "product_clean" using = <b>PROPER(TRIM(CLEAN(L2)))</b> for Case Standardization.
DELIVERY_STATUS	Case inconsistency	Upper/lower case mix	Created "delivery_status_clean" and then apply = <b>PROPER(TRIM(CLEAN(R2)))</b> for Case Standardization.
ORDER_QUANTITY	Mixed data types	Blank, N/A & Num	Create a new column named "order_quantity_flag" to flag the data issues using this formula: =IF(OR(NOT(ISNUMBER(O2)), O2<=0), "Invalid", "Valid"). Then, create another column named "order_quantity_clean" to transform the invalid values into blanks (nulls) while keeping the valid ones. Use this formula: =IF (P2="Valid", O2, ""). Finally, on the Helper Sheet, I calculated the count of valid versus invalid cases.
UNIT_COST	Mixed data types	Blank, Negative & Num	Same approach as order_quantity
LEAD_TIME_DAYS	Mixed data types	Blank, Negative & Num	Same approach; Zero considered valid

<b>TOTAL_COST</b>	Incorrect calculation	Zero, Negative & Num	Recalculated using clean order_quantity & unit_clean
<b>ORDER_DATE</b>	Date stored as text	Day/Month order mismatch (DMY vs MDY)	Converted using <b>Text to Columns</b> with <b>Date (DMY)</b>