

Dataset Introduction – E-commerce Sales

The dataset contains transactional records from an e-commerce platform, capturing customer purchases and related details. It includes **TransactionID** as a unique identifier, **CustomerID** and **ProductID** to track buyers and products, and **ProductCategory** to classify items. Transaction details such as **Quantity**, **Price**, **Discount**, and **TransactionDate** are recorded, along with **PaymentMethod**. Customer information includes **CustomerLocation**, **CustomerAge**, **CustomerGender**, **CustomerIncomeGroup**, and **CustomerLoyaltyScore**. This dataset is suitable for analyzing sales patterns, customer behavior, and loyalty trends.

Column-by-Column Excel Cleaning Paths

1. TransactionID

- **Check blanks:**
Home → Editing → Find & Select → Go To Special → Blanks → Delete Row
 - **Remove duplicates:**
Data → Data Tools → Remove Duplicates → select TransactionID → OK
-

2. CustomerID

- **Check blanks / trim spaces:**
Home → Editing → Find & Select → Go To Special → Blanks → Delete Row
Formulas → Text → TRIM(A2) (use helper column)
 - **Ensure consistent text/numeric:**
Home → Number → Format Cells → Text or Number
-

3. ProductID

- **Check blanks / trim spaces:**
Home → Editing → Find & Select → Go To Special → Blanks → Delete Row
Formulas → Text → TRIM(A2)
 - **Standardize casing:**
Formulas → Text → UPPER(A2) or PROPER(A2)
-

4. ProductCategory

- **Trim spaces & proper case:**
Formulas → Text → TRIM(A2) → PROPER(A2)
 - **Standardize categories / fix typos:**
Home → Editing → Find & Select → Replace
 - **Group rare categories:** Manually replace → “Other”
-

5. Quantity

- **Convert to number:**
Home → Number → Number (0 decimals)
Data → Text to Columns → Finish if needed
 - **Filter for invalid values:**
Data → Sort & Filter → Filter → Number Filters → Less Than 1 → Highlight/Correct
-

6. Price

- **Convert to number:**
Data → Text to Columns → Finish Or Home → Number → Currency
 - **Check negatives / zeros:**
Home → Conditional Formatting → Highlight Cells Rules → Less Than → 0
-

7. Discount

- **Convert % or decimal:**
Home → Number → Percentage OR Number
 - **Check range:**
Home → Conditional Formatting → Highlight Cells Rules → Less Than → 0
-

8. TransactionDate

- **Convert text to date:**
Data → Text to Columns → Next → Next → Date: DMY → Finish
- **Format date consistently:**
Home → Number → Short Date / Custom → dd-mm-yyyy

- **Filter invalid/future dates:**
Data → Sort & Filter → Filter → Date Filters → Before/After
-

9. PaymentMethod

- **Trim & proper case:**
Formulas → Text → TRIM(A2) → PROPER(A2)
 - **Standardize values:**
Home → Editing → Find & Select → Replace
 - **Fill blanks:**
Home → Editing → Find & Select → Go To Special → Blanks → Type “Unknown” → Ctrl+Enter
-

10. CustomerLocation

- **Trim & proper case / upper case:**
Formulas → Text → TRIM(A2) → PROPER(A2) or UPPER(A2)
 - **Standardize country/city names:**
Home → Editing → Find & Select → Replace
 - **Fill blanks:**
Same as PaymentMethod
-

11. CustomerAge

- **Convert to number:**
Home → Number → Number (0 decimals)
 - **Filter invalid/out-of-range values:**
Data → Sort & Filter → Filter → Number Filters → Less Than / Greater Than
 - **Fill blanks:** Optional, use median manually
-

12. CustomerGender

- **Trim & proper case:**
Formulas → Text → TRIM(A2) → PROPER(A2)
- **Standardize values:**
Home → Editing → Find & Select → Replace (M → Male, F → Female)
- **Fill blanks:** Replace with “Unknown”

13. CustomerIncomeGroup

- **Trim & proper case:**

Formulas → Text → TRIM(A2) → PROPER(A2)

- **Standardize typos / missing values:**

Home → Editing → Find & Select → Replace

Home → Editing → Find & Select → Go To Special → Blanks → Type “Unknown” → Ctrl+Enter

14. CustomerLoyaltyScore

- **Convert to number:**

Home → Number → Number (0 decimals)

- **Filter invalid values:**

Data → Sort & Filter → Filter → Number Filters → Less Than 0

- **Fill blanks:** Optional: median or leave blank
-

15. Derived Columns

- **TotalValue:**

Insert new column → type formula: =[@Quantity]*[@Price]*(1-[@Discount]) (for table)

- **SuspiciousFlag:**

=IF(OR([@Quantity]<1,[@Price]<=0,[@TotalValue]<=0),"Check","OK")

Additional Excel Tools to Help

- **Conditional Formatting:** Home → Conditional Formatting → Highlight Cells Rules → [Choose Rule]
- **Data Validation (restrict categories):** Data → Data Validation → List → select allowed values
- **Remove duplicates:** Data → Data Tools → Remove Duplicates
- **Text to Columns:** Data → Text to Columns → choose format
- **Find & Replace:** Ctrl+H