Step by step Guide on Cleaning the data set

**1. Open the File**

* Open messy\_console\_sales\_data.xlsx in Excel.

**2. Remove Duplicates**

* Select all data (Ctrl + A).
* Go to the **Data** tab → Click **Remove Duplicates**.
* Select all columns or just key ones like Order ID, Customer Name, Console Model.
* Click **OK**.

**3. Standardize Text Formatting**

* Use formulas or Flash Fill to clean up inconsistent text:
  1. Customer name =PROPER(TRIM(A2))
  2. Email =LOWER(TRIM(B2))
  3. Console Model =Use a mapping table + VLOOKUP to standardize names
  4. Region Use =PROPER(TRIM(C2)) or a mapping table
  5. Retailer =Same as above

**4. Fix Date Formats**

* Select Order Date and Delivery Date.
* Use Text to Columns or =DATEVALUE () to convert inconsistent formats.
* Format cells as Short Date.

**5. Handle Missing Values**

* Use **Filters** to find blanks.
* Fill with:
  + "Unknown" for text
  + 0 or average for numbers
  + Use =IF(ISBLANK(A2), "Unknown", A2) to fill blanks

**6. Convert Text to Numbers**

For columns like Units Sold, Unit Price, Discount, Tax, Total Amount:

* Remove non-numeric characters using SUBSTITUTE() or VALUE()
* Example: =VALUE(SUBSTITUTE(A2, "USD ", ""))

**7. Standardize Categories**

Use **Data Validation** or a mapping table to clean:

* Payment Method (e.g., "paypal" → "PayPal")
* Shipping Status (e.g., "delivrd" → "Delivered")
* Return Status (e.g., "YES" → "Returned")

**8. Check for Outliers or Errors**

* Use **Conditional Formatting** to highlight:
  + Negative values
  + Unusual ratings or feedback scores

**9. Create a Clean Summary Table**

Use **PivotTables** to summarize:

* Total revenue by console model
* Average rating by region
* Units sold by retailer
* **Cleaning formulas** for:
  + Customer Name: =PROPER (TRIM (cell value))
  + Email: =LOWER (TRIM (cell value))
  + Order Date: =TEXT (...,"dd-mm-yyyy")
  + Units Sold: =IFERROR(VALUE(...),"Check")
  + Unit Price: =VALUE(SUBSTITUTE(...))
  + Payment Method, Region, Shipping Status: Standardized with PROPER, TRIM, and IF