Data Cleaning Procedure in Excel

Project: Retail Optimization Analysis (UrbanMart)

Dataset: Sample Superstore

Tool: Microsoft Excel

Objective

To clean and prepare the "Orders" dataset for business analysis using Microsoft Excel by applying structured data handling, formatting, and logic.

Step-by-Step Data Cleaning Workflow

1. Remove Unnecessary Columns

- Keep only relevant fields: Order ID, Order Date, Ship Date, Customer Name, Segment, Region, Product Name, Category, Sub-Category, Sales, Quantity, Discount, Profit

2. Standardize Column Names

- Rename headers to remove spaces and make them readable

Example: `Order Date` → `Order_Date`, `Product Name` → `Product_Name`

3. Check for Missing Data

- Use `Go To Special > Blanks` to highlight null cells
- Fill or remove missing values:
    - If `Customer Name` or `Region` is missing → mark row for review
    - If `Profit` is missing → recalculate or flag

4. Remove Duplicates

- Use `Remove Duplicates` under `Data` tab based on `Order ID` or full row

```
Data > Remove Duplicates
```

5. Convert Dates to Proper Format

- Highlight `Order_Date` and `Ship_Date` columns
- Format as `Short Date`
- Create new columns:
    - `Month`: `=TEXT(Order_Date,"mmmm")`
    - `Year`: `=YEAR(Order_Date)`

6. Calculate Derived Columns (Optional)

- `Profit Margin`: `=Profit/Sales`
- `Days to Ship`: `=Ship_Date - Order_Date`

7. Filter Outliers or Errors

- Use `Conditional Formatting` to:
    - Highlight `Discount` > 0.5
    - Flag negative profits
    - Spot extreme quantities

8. Sort & Filter for Review

- Sort by `Region`, `Sales`, or `Category`
- Filter for `Quantity = 0`, or `Profit < 0`

9. Save Cleaned Version

- Save as: `UrbanMart_Cleaned_Superstore.xlsx`

Notes for Review

- All dates are parsed into proper datetime format
- All column names are standardized (no spaces)
- Dataset is now ready for Pivot Table analysis and visualization in Tableau/Power BI