**Realistic Business Problem Statement (Fashion Retail Context)**

**Business Problem:**  
The company is experiencing issues with stock management and pricing strategy across its fashion product lines. There is a need to understand which products are most likely to go out of stock quickly, and how pricing (actual vs. compare-at price) influences this. Additionally, the business wants to optimize inventory planning by identifying popular sizes, colors, and materials for each demographic (e.g., Women, Girls). Cleaning and analyzing this dataset can help in developing a data-driven **inventory optimization and dynamic pricing strategy**.

**Objective:**  
To clean and analyse sales and inventory data in order to:

* Identify top-selling products and attributes (size, color, material, brand)
* Detect pricing inconsistencies or discounts
* Understand demand trends by time and customer segment
* Predict stock-out risks to improve restocking and pricing decisions

**Step-by-Step Excel Cleaning Guide**

**1. Standardize Column Names**

* Rename columns like:
  + Variant Price → variant\_price
  + Product Type → product\_type
  + Remove spaces and make all headers lowercase with underscores

**2. Fix Date Formats**

* Select the entire **Date** column
* Use Data > Text to Columns > choose **Date (DMY)** format
* Or use formula:

=TEXT(A2,"yyyy-mm-dd")

* Apply consistent formatting (yyyy-mm-dd)

**3. Clean Text Fields**

For columns like brand, dominant\_material, dominant\_color, product\_type, ideal\_for, and is\_in\_stock:

* Use =PROPER(TRIM(A2)) to clean and title-case values
* Drag down for the entire column
* For is\_in\_stock, replace blanks using **=IF(A2="", "Others", A2)**

**4. Convert Price Columns to Numbers**

* For variant\_price and compare\_at\_price:
  + Remove any non-numeric characters (use Find & Replace, e.g., remove commas)
  + Change column format to **Number**

**5. Standardize Sizes**

* Use =UPPER(TRIM(A2)) in the size column

**6. Filter & Handle Missing Data**

* Use filters to check missing values (blanks)
* Fill or handle them appropriately (e.g., fill is\_in\_stock with "Unknown")

**7.Convert the decimal numbers under the time columns to actual time in AM/PM**

* Select the cell with the decimal value.
* Press Ctrl + 1 (or right-click → Format Cells).
* Choose Custom under the "Time" tab.

**8. Save a Cleaned Copy**

* Save the final cleaned version as Fashion\_Inventory\_Cleaned.xlsx

**Complete Excel Data Cleaning Checklist**

**1. Standardize Column Headers**

* Make all column names **lowercase**
* Replace spaces with underscores (variant\_price, product\_type, etc.)
* Ensure **consistency** across all fields

**2. Fix Inconsistent Date Formats**

* Use TEXT() or Text to Columns to ensure all dates are yyyy-mm-dd
* Confirm that Excel recognizes them as **dates**, not text
* Optional: Extract **Month/Year** for seasonal analysis using:
  + =TEXT([@date], "mmm yyyy") or =MONTH([@date])

**3. Clean Text Columns**

Apply to columns like brand, dominant\_color, product\_type, etc.:

* Remove leading/trailing spaces: =TRIM(A2)
* Standardize casing: =PROPER() for names, =UPPER() for codes
* Use filters to identify inconsistent entries (e.g., cotton vs. Cotton)

**4. Convert Price Columns to Numbers**

* Remove non-numeric symbols (e.g., ₹, $, commas)
* Format variant\_price and compare\_at\_price as **Number**
* Optional: Create a **discount column**:

=[@variant\_compare\_at\_price] - [@variant\_price]

**5. Handle Missing or Incomplete Data**

* For is\_in\_stock, replace blanks:  
  =IF([@is\_in\_stock]="", "Unknown", [@is\_in\_stock])
* Identify and highlight any missing critical fields (like price or date)
* Use **Conditional Formatting** to spot blanks or zeros

**6. Normalize Sizes**

* Use =UPPER(TRIM(A2)) to standardize sizes
* Group sizes for analysis (e.g., XS-S-M vs. age-based like 5-6Y)
* Optional: Add a **"Size Category"** column (e.g., Kids, Adults)

**7. Validate Categorical Fields**

* Create a **drop-down list** for fields like ideal\_for, is\_in\_stock, brand
* Helps prevent typos and inconsistencies

**8. Remove Duplicates**

* Use Data > Remove Duplicates on product\_id or id if needed

**9. Add Calculated Columns (Optional but Helpful)**

* **Discount %**:

=([@variant\_compare\_at\_price] - [@variant\_price]) / [@variant\_compare\_at\_price]

* **Is Discounted** (Yes/No):

=IF([@variant\_price] < [@variant\_compare\_at\_price], "Yes", "No")

**10. Save Cleaned Version**

* Save as: Fashion\_Inventory\_Cleaned.xlsx
* Keep a backup of the original

**Ready for Analysis!**

Once these steps are done, your dataset will be fully ready for:

* Sales trend analysis
* Inventory planning
* Pricing strategy optimization
* Customer preference insights