



## STEP 1 : Data Cleaning

### Initial Steps

1. Importing Data:
  - Imported the dataset into Excel for analysis and cleaning.
2. Applying Filters:
  - Applied filters to all columns to facilitate the identification of inconsistencies and anomalies.
3. Checking for Duplicate Values:
  - Reviewed the dataset for any duplicate entries and removed them to ensure data integrity.
4. Identifying Null Values:
  - Scanned the dataset for null or missing values and noted their locations for further action.

### Data Cleaning Actions

1. Gender Column:
  - The gender column contained inconsistent values: "M," "Men," "W," and "Women."
  - **Action Taken:** Changed all instances of "M" to "Men" and "W" to "Women" to ensure uniformity.
2. Channel Column:
  - There were spelling mistakes in the channel names.
  - **Action Taken:** Corrected "Amenazon" to "Amazon" and "Meenasho" to "Meesho".
3. Category Column:
  - The category "Womenester dress" was incorrectly labelled.

- **Action Taken:** Changed "Womenester dress" to "Western dress" for accuracy.
- 4. **Size Column:**
  - The size column had inconsistent entries for men's sizes.
  - **Action Taken:** Changed all instances of "Men" to "M" for standardisation.
- 5. **Quantity Column:**
  - The quantity column contained both numerical and alphabetical data.
  - **Action Taken:** Converted entries such as "one" to "1" and "two" to "2" to ensure consistency in numerical representation.
- 6. **Shipping State Column:**
  - The shipping state column contained multiple misspelt state names.
  - **Action Taken:** Reviewed and corrected as many misspelt state names as possible to ensure accuracy and standardisation.

#### **Final Review:**

- Conducted a final review of the dataset to verify that all corrections were accurately implemented and that the data was ready for analysis.

## **STEP 2 : Data Processing**

**Ques** - What is data processing ?

**Ans** - This term encompasses the actions taken to manipulate and analyse data, including the creation of new columns, applying formulas, performing calculations, and transforming data to extract insights.

**Steps Taken :**

**Step(1)**: As part of the data processing tasks, I created a new column titled **"Age Group"**.

In this column, I applied a formula to categorise individuals based on their age. The formula used was:

**=IF(E2>=50, "Senior", IF(E2>=30, "Adult", "Teenager"))**

This formula classifies individuals into three categories:

- "Senior" for those aged 50 and above,
- "Adult" for those aged 30 to 49, and
- "Teenager" for those under 30.

After applying the formula in the first cell, I proceeded to copy and paste it down the entire column to ensure that all relevant data was categorised accordingly.

**Step (2) :** To address the owner's inquiry regarding the month with the highest sales and orders, I extracted the month from the existing Date column in the dataset. I utilised the following formula:

```
=TEXT(G2, "mmm")
```

This formula converts the date in cell G2 into a three-letter abbreviation representing the month. By applying this formula to the entire Date column, I created a new column that displays the corresponding month for each entry, facilitating further analysis of sales and orders by month.

### **STEP 3 : Data Visualisation**

**Step (1) :** To analyse the sales and orders data further, I created a Pivot Table.

In the Values area, I added the Amount (using the Sum function) and the Order ID (using the Count function). I then included the extracted Months in the Rows area.

To simplify the presentation, I disabled the Grand Totals for both rows and columns.

I accessed the Design panel and selected the Grand Total option to turn off the totals.

Next, I utilised the Pivot Table Analyze tab to insert a Pivot Chart. Since only one chart was required, I chose a Combo Chart.

Given that the Order IDs were in millions, they appeared on the x-axis. To ensure clarity in the representation, I opted to use a Secondary Axis for the Order IDs.

To improve the visual aesthetics of the chart, I removed the field buttons by right-clicking on the represented values and selecting Hide All Field Buttons from the Chart. I then added a chart title to enhance understanding.

As the values on the left side of the chart contained more than three zeroes, it would be challenging for viewers to interpret the numbers accurately. To address this, I double-clicked on the left panel, which displayed the Order ID values, and selected

Format Axis from the dropdown menu. Under Format Axis > Number, I modified the Format Code from the default setting to 0.00,"m" (noting the importance of including the double commas) to represent the values in millions.

This step enhanced the readability of the chart, allowing for a clearer presentation of the data.

**Step (2)** : I created a new worksheet titled "Desi Trends Report 2022" to compile the analysis results.

I then copied the chart created in Step 3 and pasted it into the newly created sheet. This allows for a consolidated view of the sales and orders analysis, facilitating easier access and presentation of the findings.

**Step (3)** : In the new worksheet, I added a header titled "Desi Trends Annual Report 2022" in one of the columns. To enhance the presentation, I selected all the columns up to column U and utilised the Merge and Center feature from the Home toolbar. This action combined the selected cells into a single cell, effectively centering the title across the top of the report and providing a clear designation for the visualisation.

**Step (4)** : After adding the title to the sheet, I customised it to my preference by adjusting the colours and adding borders. On the remaining sheets, I navigated to the View toolbar and removed the gridlines by unchecking the Gridlines option in the Show toolbar.

**Step (5)** : Created an additional sheet to analyse and represent the question of who shopped more—men or women. To achieve this, I generated a new pivot table, adding the Gender column to the Rows and the Sum of Amount to the Values. I then created a chart using the same method as before and customised it according to my preferences.

- Following that, I created multiple additional charts to fulfil the project requirements. Each chart was then incorporated into the final sheet, providing a detailed and comprehensive view of the data.

**Step (6)** : After completing and consolidating all the charts into the 'Desi Trends Report 2022' sheet, I selected one of the charts and, from the PivotChart Analyze tab, chose 'Insert Slicer.' A dropdown menu appeared with each column name, and I selected 'Month,' 'Channel,' and 'Category,' resulting in three slicers being added to the sheet.

To ensure these slicers were connected to each chart, I right-clicked on each slicer, selected 'Report Connections,' and checked the boxes for the relevant charts on the sheet. I repeated this process for the remaining two slicers."