

## “Data Cleaning & Preparation”

---

### Data Cleaning & Preparation (Excel + Power BI)

Before building the dashboard, the dataset was cleaned and prepared using **Microsoft Excel** and **Power BI Power Query**.

This ensures accuracy, reliability, and consistency of insights.

---

#### 1. Excel-Based Data Cleaning Steps

The following steps were performed in Excel **before loading the dataset into Power BI**:

##### ✓ 1. Checked for missing values

- Verified no blank or null entries in Sales, Profit, Discount, Quantity, Category, Region, etc.
  - Ensured all required fields had valid data.
- 

##### ✓ 2. Validated data types

- Order Date → Converted to proper Date format
  - Sales, Profit, – Set as Number format (2 decimals)
  - Quantity → Whole number
  - Discount → Percentage / Decimal
- 

##### ✓ 3. Removed duplicates

- Checked for duplicate rows using Order ID
  - Ensured each order is unique
- 

##### ✓ 4. Standardized column names

- “Product Name” → Product\_Name
  - “Order Date” → Order\_Date
  - Avoided spaces for easier DAX development
- 

##### ✓ 5. Outlier & anomaly check

- Verified extreme values in Sales and Profit
- Checked Discount values (should be 0.0–1.0)

## ✓ 6. Trimmed & cleaned text fields

Using Excel tools:

- Removed trailing/leading spaces
  - Ensured consistent text casing for:
    - Category
    - Segment
    - Region
    - Customer Name
- 

## 2. Power BI Data Preparation (Power Query)

Once loaded into Power BI:

### ✓ Step 1 — Data profiling

Turned on:

- Column quality
- Column distribution
- Column profile

To check for:

- Errors
  - Empty values
  - Unique values
  - Distribution issues
- 

### ✓ Step 2 — Ensured data types were correct

Converted:

- Date → Date type
  - Sales & Profit → Decimal Number
  - Quantity → Whole number
  - Region/Category/Customer Name → Text
- 

### ✓ Step 3 — Created a proper Date Table

This is mandatory for time intelligence (YoY, MoM trends).

```
DateTable = CALENDAR(MIN('Sheet1'[Order Date]), MAX('Sheet1'[Order Date]))
```

Marked as **Date Table** and connected relationships.

---

#### ✓ Step 4 — Removed unneeded columns (optional)

This improves performance.

---

#### ✓ Step 5 — Created DAX measures

Used separate measure table (recommended practice):

- Total Sales
- Total Profit
- Profit Margin
- YoY Growth
- Total Sales LY
- Quantity
- Average Discount

---

### 3. Summary of Data Preparation

This entire preparation workflow ensures:

- ✓ Clean dataset
- ✓ Uniform data types
- ✓ Reliable KPIs
- ✓ Accurate time intelligence
- ✓ High-quality visuals
- ✓ Professional BI workflow