# Data Cleaning Report – Sales Dataset (FY 2020–21)

This report summarizes the data cleaning and preprocessing steps performed on the raw sales dataset to ensure consistency, accuracy, and readiness for further analysis.

## Tools Used

• Python (Pandas library)

## Steps Performed

1. Handled missing values using Pandas .fillna() method (filled with 'Unknown').

2. Removed duplicate rows using .drop\_duplicates().

3. Standardized column names to lowercase and replaced spaces with underscores.

4. Checked and converted date columns to datetime format (dd-mm-yyyy).

5. Verified and corrected data types for consistency.

## Summary of Changes

• Missing values before cleaning: 0  
• Missing values after cleaning: 0  
• Duplicate rows before cleaning: 0  
• Duplicate rows after cleaning: 0  
• Final dataset shape: 286392 rows × 36 columns