

Data Cleaning Report

Overview

This report outlines the data cleaning and preprocessing steps applied to the dataset `sales_data_sample.csv`. The aim was to prepare the data for analysis by ensuring completeness, consistency, and correct formatting.

Cleaning Summary

1. Missing Values

- Columns with mostly missing data such as ADDRESSLINE2 were removed.
- All remaining fields now have complete data.

2. Duplicates

- Duplicate rows were checked and removed to ensure each record is unique.

3. Standardized Text

- Fields like STATUS, COUNTRY, and DEALSIZE were cleaned:
 - Converted to title case (e.g., "shipped" → "Shipped")
 - Leading/trailing spaces removed

4. Date Formatting

- The ORDERDATE column was converted to a consistent format: dd-mm-yyyy.

5. Renamed Column Headers

- Column headers were standardized:
 - Converted to uppercase
 - Removed extra spaces
 - Example: "Customer Name" → "CUSTOMERNAME"
 - Contact first name and last name combined into one column

6. Data Types

- Ensured numerical columns (SALES, PRICEEACH, etc.) are of correct types.
- ORDERDATE is recognized as a proper date field.