

Data Profiling Report

Dataset: Sample – Superstore.csv

Total Rows: 9,994 | Total Columns: 21 | Duplicate Rows: 0

1. Dataset Overview

The dataset contains order-level information including customer, product, shipping, and transactional data. All columns contain 0% missing values.

2. Descriptive Statistics Summary

Sales, Profit, Discount, Quantity, and Postal Code exhibit wide variance, reflecting a diverse product mix.

3. Outlier Analysis

Sales Outliers: 1,167 (11.68%)

Profit Outliers: 1,881 (18.82%)

High-value and negative-profit transactions should be validated.

4. Relationship Tests

Discount vs Profit: Correlation -0.219

Sales vs Quantity: Correlation +0.200

5. Recommendations

- Standardize date fields.
- Investigate negative-profit transactions.
- Review discount strategy.
- Ensure consistent category labeling.
- Produce a cleaned dataset for further analysis.