DATA PROFILING REPORT

SUPERSTORE DATASET ANALYSIS

Report Generated: December 10, 2025
Dataset: Sample - Superstore.csv
Analysis Type: Comprehensive Data Profiling

EXECUTIVE SUMMARY

This report provides a comprehensive profiling analysis of the Superstore dataset, covering data structure, quality assessment, outlier detection, and relationship analysis between key business metrics.

Key Findings:
• Dataset contains 9,994 transactions with complete data quality (no missing values or duplicates)
• 18.72% of transactions result in negative profits, totaling $156,131.29 in losses
• Strong negative correlation (-0.22) between discount rates and profitability
• Sales outliers represent 11.68% of transactions, with profit outliers at 18.82%

SECTION 1: DATASET OVERVIEW

1.1 BASIC STATISTICS

Total Rows: 9,994
Total Columns: 21
Memory Usage: 10.35 MB
Data Quality: 100% complete (no missing values)
Duplicate Records: 0 (0.00%)

1.2 COLUMN STRUCTURE

The dataset contains 21 columns organized into the following categories:

IDENTIFIER COLUMNS (3):
• Row ID int64 Unique transaction identifier
• Order ID object Order reference number
• Product ID object Product SKU identifier

TEMPORAL COLUMNS (2):
• Order Date object Date of order placement
• Ship Date object Date of shipment

CUSTOMER INFORMATION (5):
• Customer ID object Unique customer identifier
• Customer Name object Customer full name
• Segment object Customer segment classification
• Country object Customer country
• City object Customer city
• State object Customer state
• Postal Code int64 Postal/ZIP code
• Region object Geographic region

PRODUCT INFORMATION (4):
• Category object Primary product category
• Sub-Category object Product sub-classification
• Product Name object Full product description

SHIPPING INFORMATION (1):
• Ship Mode object Shipping method used

TRANSACTION METRICS (4):
• Sales float64 Total sales amount ($)
• Quantity int64 Number of units sold
• Discount float64 Discount rate applied (0-1)
• Profit float64 Profit/loss amount ($)

SECTION 2: DATA QUALITY ASSESSMENT

2.1 MISSING VALUES

Result: ✓ EXCELLENT
All 21 columns contain complete data with no missing values across all 9,994 records. This indicates high data collection quality and integrity.

Missing Value Summary:
• Total Missing Values: 0
• Columns with Missing Data: 0
• Overall Completeness: 100.00%

2.2 DUPLICATE RECORDS

Result: ✓ EXCELLENT
No duplicate transactions detected in the dataset.

Duplicate Analysis:
• Total Duplicate Rows: 0 (0.00%)
• Unique Transactions: 9,994

Interpretation: Each record represents a unique transaction, ensuring data integrity for analytical purposes.

SECTION 3: OUTLIER ANALYSIS

3.1 SALES COLUMN OUTLIER DETECTION

Method: Interquartile Range (IQR) Analysis

Descriptive Statistics:
• Mean: $229.86
• Median: $54.49
• Standard Deviation: $623.25
• Range: $0.44 - $22,638.48

Distribution Characteristics:
• Q1 (25th percentile): $17.28
• Q3 (75th percentile): $209.94
• IQR: $192.66

Outlier Thresholds:
• Lower Bound: $-271.71
• Upper Bound: $498.93

Outlier Detection Results:
• Outliers Detected: 1,167 transactions (11.68%)
• Outlier Type: Primarily high-value sales (above $498.93)

Interpretation:
The significant difference between mean ($229.86) and median ($54.49) indicates a right-skewed distribution. The 11.68% of outliers represent high-value transactions that are legitimate business events (bulk orders, enterprise sales) rather than data quality issues.

3.2 PROFIT COLUMN OUTLIER DETECTION

Method: Interquartile Range (IQR) Analysis

Descriptive Statistics:
• Mean: $28.66
• Median: $8.67
• Standard Deviation: $234.26
• Range: $-6,599.98 - $8,399.98

Distribution Characteristics:
• Q1 (25th percentile): $1.73
• Q3 (75th percentile): $29.36
• IQR: $27.64

Outlier Thresholds:
• Lower Bound: $-39.72
• Upper Bound: $70.82

Outlier Detection Results:
• Outliers Detected: 1,881 transactions (18.82%)
• Outlier Types: Both high profits and significant losses

3.3 NEGATIVE PROFIT ANALYSIS

■■ CRITICAL FINDING:

Negative Profit Transactions: 1,871 (18.72% of all transactions)
Total Loss Amount: $-156,131.29

Interpretation:
Nearly 1 in 5 transactions result in losses. This is a significant business concern that warrants investigation into:
• Pricing strategies for discounted items
• Cost structure analysis
• Product-specific profitability assessment

SECTION 4: RELATIONSHIP ANALYSIS

4.1 DISCOUNT vs PROFIT RELATIONSHIP

Statistical Correlation:
• Pearson Correlation Coefficient: -0.2195 (weak negative correlation)

Interpretation:
There is a statistically significant negative relationship between discount rates and profitability. As discounts increase, profits tend to decrease.

Profit by Discount Bracket Analysis:

Discount Bracket	Mean Profit	Median Profit	Count
0-10%	$96.06	$54.32	94
10-20%	$24.74	$6.53	3,709
20-30%	$-45.68	$-25.38	227
30-50%	$-156.28	$-81.90	310
50-100%	$-89.44	$-11.66	856

Key Insights:
• Transactions with discounts ≥20% typically result in losses
• Low discount rates (0-10%) yield the highest average profit ($96.06)
• Most transactions (3,709) fall in the 10-20% discount bracket
• High discount brackets (30-100%) show substantial average losses

■■ BUSINESS RECOMMENDATION:
Re-evaluate discounting strategy for rates above 20%. Current discount policies in higher brackets are eroding profitability.

4.2 SALES vs QUANTITY RELATIONSHIP

Statistical Correlation:
• Pearson Correlation Coefficient: 0.2008 (weak positive correlation)

Additional Insights:
• Average Sale per Unit: $60.66
• Average Quantity per Transaction: 3.79 units

Interpretation:
There is a weak positive relationship between quantity and sales, which is expected but not perfectly linear. The correlation coefficient of 0.20 suggests that factors other than quantity (such as product mix, pricing tiers, and discounts) significantly influence transaction values.

The average sale per unit of $60.66 provides a baseline metric for pricing analysis and revenue forecasting.

SECTION 5: KEY FINDINGS & RECOMMENDATIONS

5.1 DATA QUALITY ASSESSMENT

✓ STRENGTH: Excellent data quality with 100% completeness and no duplicates
→ Recommendation: Maintain current data collection standards

5.2 PROFITABILITY CONCERNS

■■ CRITICAL ISSUE: 18.72% of transactions unprofitable
→ Recommendation: Implement profit margin checks before approving high discount rates
→ Action: Review pricing strategy for products frequently sold at steep discounts

5.3 DISCOUNT STRATEGY

■■ ISSUE: Discounts ≥20% consistently result in negative profits
→ Recommendation: Set maximum discount threshold at 20%
→ Action: Require management approval for discounts exceeding 20%
→ Opportunity: Focus promotional efforts on 10-20% discount range which maintains profitability while driving volume

5.4 OUTLIER MANAGEMENT

✓ OBSERVATION: Sales and profit outliers are legitimate business transactions
→ Recommendation: Monitor high-value outlier transactions for pattern analysis but do not filter/remove
→ Action: Create separate reporting for bulk/enterprise transactions

5.5 BUSINESS METRICS

✓ INSIGHT: Average transaction metrics established
→ Recommendation: Use $60.66 per unit and 3.79 units per transaction as benchmarks for forecasting
→ Action: Track deviations from these metrics to identify trends

END OF REPORT

For questions or additional analysis requests, please refer to the source dataset: Sample - Superstore.csv

Report prepared by: Data Profiling Analysis System
Analysis Date: December 10, 2025