

Problem Statement

The retail store provided raw transactional sales data which was not prepared for analysis.

The dataset contained multiple data quality issues such as:

- Missing or blank values
- Inconsistent data formatting
- Duplicate entries
- Unstructured category and item labels
- Outliers in spending and quantity values
- No clear discount logic column present

These problems would negatively affect analysis, so the data needed to be cleaned, standardized, and validated before deriving insights.

Business Requirement:

The task was to:

- Clean and prepare the retail transactional dataset.
- Identify the logic behind discount applicability.
- Create a calculated column for discount based on spending criteria.
- Analyze customer behavior and spending patterns after cleaning.

Discount Rule Applied:

A discount was applied based on total purchase value:

Condition	Discount
Total Spent > 70	30% discount applied
Total Spent ≤ 70	No discount provided

This rule helped determine customer eligibility for discounts and evaluate promotion effectiveness.