Data Validation Summary Report – Week 3

Your Name Here

September 18, 2025

Executive Summary

This report presents the results of data validation checks performed on the shipping dataset. The goal was to ensure accuracy, completeness, and reliability of the records before analysis. Key issues identified include duplicate records, invalid values, missing data, and business rule violations. Addressing these issues will improve the overall quality of the dataset and lead to more trustworthy insights.

1 Objective

The validation process aimed to:

- Verify data consistency and completeness.
- Detect duplicates and anomalies.
- Check compliance with business rules.
- Provide recommendations for data quality improvement.

2 Validation Results

Here is a summary of the validation checks and the issues found.

Table 1: Summary of Data Validation Issues

Validation Check	Issue Detected	Count
Duplicate shipping_id	Same ID appearing more than once	1
Invalid status values	Status = 5 (should only be 0,1,2)	1
Missing customer IDs	NULL values in customer column	1
Missing status	NULL values in status column	1
Negative customer IDs	Invalid negative values detected	1

3 Key Insights

- Duplicate IDs can lead to overcounting of shipments.
- Invalid status values (like 5) indicate data entry or process errors.
- Missing customer/status fields reduce the completeness of analysis.
- Negative customer IDs highlight the need for better data validation rules.

4   Recommendations

Remove or Merge Duplicates Ensure each `shipping_id` is unique in the dataset.

Standardize Status Field Restrict status to 0 = Pending, 1 = Shipped, 2 = Delivered. Update or delete invalid entries.

Fix Missing Data Investigate NULL customer and status fields. Impute missing values where possible or mark as invalid.

Strengthen Data Entry Rules Enforce constraints to prevent negative IDs. Add input validation at the database/application level.

5   Conclusion

The validation process successfully identified multiple issues affecting data accuracy and reliability. By implementing the recommendations, the dataset will become clean, consistent, and business-ready, ensuring that future insights are based on accurate information.