

# Error Findings Report

**Title:** First Stage Data Profiling and Cleaning

**Purpose:** This report identifies and summarises various data issues found during the data profiling and cleaning process, along with recommended actions for resolution.

---

## 1. Missing Values

**Issue:** Partial or Incomplete Values in Customer Names

- **Status:** Found
- **Example:** "Yes, Daniel"
- **Action:** Check with the source provider to obtain complete data.

**Issue:** Missing Spaces

- **Status:** Found
- **Action:** Check with the source provider to get the corrected data.

## Shipping Fees Missing

- **Issue:** Order ID 1377 has shipping fees missing.
  - **Status:** Found
  - **Action:** Verify and correct the shipping fees with the source provider.
- 

## 2. Data Outside Range

**Issue:** Quantity Values Less Than 18 and Greater Than 100

- **Status:** Found
- **Example:** Quantity 19,000
- **Action:** Check with the source provider.

**Issue:** Dates on Transactions Earlier Than the Business Start Date

- **Status:** Found
- **Example:** Year 2013
- **Action:** Check with the source provider.

**Issue:** Codes (e.g., Postal Codes or Zip Codes) Outside the Range of Defined Values

- **Status:** Issues Found
  - **Action:** Check with the source provider.
-

### 3. Malformed Data

**Issue:** Truncated Names

- **Status:** Found
- **Example:** "daniel"
- **Action:** Check with the source provider.

**Issue:** Abnormal Length of Words

- **Fields Affected:** Order ID, Quantity, Shipping Price
  - **Example:** Quantity shows 19,000, Product (Sauces) shows -400
  - **Action:** Check with the source provider.
- 

### 4. Business Rules Violations

**Issue:** Order Total Greater Than Customer's Credit/Purchasing Limit

- **Example:** Product (Sauces) shows -400
- **Action:** Check with the source provider.

**Issue:** Values Outside Cutoff Values/Outliers

- **Example:** Quantity shows 19,000, Product (Sauces) shows -400
- **Action:** Check with the source provider.

**Issue:** Negative or Positive Values

- **Example:** "SOPHIA MOORE -84" in Quantity
- **Action:** Check with the source provider.

**Issue:** Incorrect Date Range

- **Example:** Year 2013
- **Action:** Check with the source provider.

**Issue:** Shipped Date Earlier Than Order Date

- **Example:** A date from 1381 in Order ID column, 2013 year date
  - **Action:** Check with the source provider.
- 

### 5. Inconsistent Values

**Issue:** Order Date Contains Order ID

- **Status:** Found
- **Action:** If necessary, check with the source provider.

**Issue:** Order Items Not Associated with an Order

- **Status:** Need to check
  - **Action:** Check with the source provider.
- 

## 6. Duplicate Order IDs

**Issue:** Order ID 1381 appears twice with the same details.

- **Status:** Found
  - **Action:** Check with the source provider to eliminate duplicates.
- 

## 7. Order Date and Ship Date Conflicts

**Issue:** Some orders have shipping dates before the order dates, which is logically incorrect.

- **Example:** Order ID 1370 and 1371 have the same order date and shipping date.
  - **Action:** Check with the source provider to correct the dates.
- 

## 8. Postcode Missing for Ship State

**Issue:** The "Ship State Postcode" is missing in the summary table.

- **Status:** Found
  - **Action:** Check with the source provider to obtain the missing postcode.
- 

## 9. Incorrect City Name Format

**Issue:** Order ID 1376 has "NY-Newyork" which should be "New York, NY".

- **Status:** Found
  - **Action:** Correct the format to "New York, NY" and check with the source provider.
- 

## 10. Category Specifics Missing

**Issue:** Order ID 1384 lists "Jam only" but lacks further details on the category type.

- **Status:** Found
  - **Action:** Check with the source provider to provide complete category details.
-

## 11. Order Quantity Discrepancy

**Issue:** Order ID 1368 shows a quantity of 19000 which seems unusually high.

- **Status:** Found
  - **Action:** Verify with source provider.
-