

# ANALYSIS REPORT

## SALES DATASET

---

Generated: 2026-02-03 15:47

### SUMMARY

Total records analyzed:	500
Records after cleaning:	500
Records with valid data:	500
Instances created:	500

## 1. Exploratory Data Analysis

### 1.1 General Information of the Dataset

The original dataset contains **500** records. After the cleaning process, **500** records were retained, representing a retention rate of **100%**.

The data includes information on sales transactions, including products, quantities, prices, and statuses. The total recorded revenue is **\$144,303.49**.

### 1.2 Missing Values

No missing values found in the dataset.

## 2. Sales Status Analysis

### 2.1 Status Distribution

- Pending: **178**
- Completed: **170**
- Cancelled: **152**

### 2.2 Old Pending Sales

- Sales in Pending status for more than 5 years: **170**

*Note: Sales pending for a long time may indicate issues in the sales closing process.*

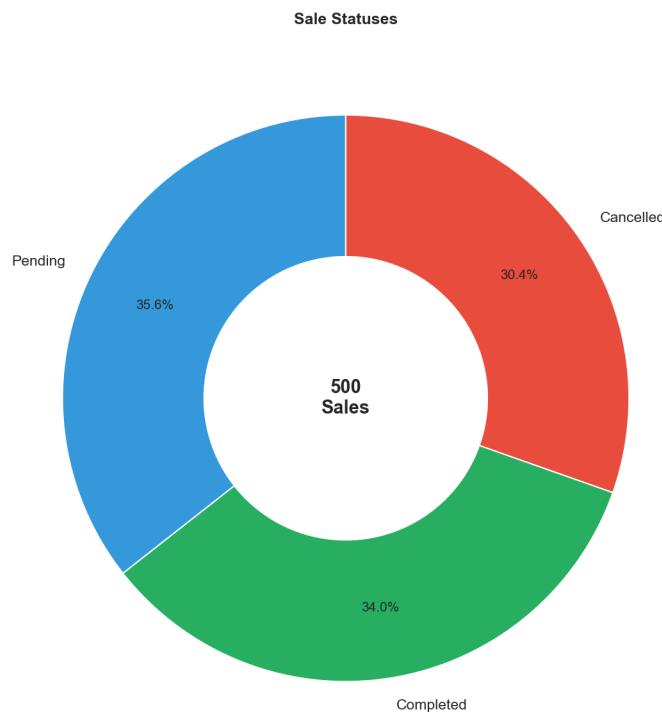


Figure 1: Sales Status

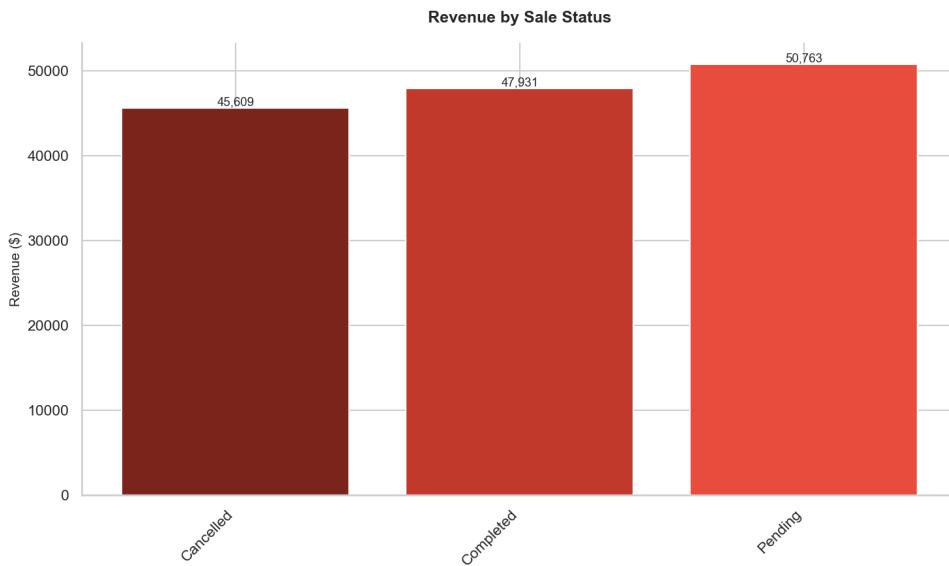


Figure 2: Revenue by Status

### 3. Date and Consistency Validation

---

#### 3.1 Future Dates

- Sales with future dates: **13** (2.6%)

#### 3.2 Sales vs Hire Date

- Sales made before employee hire date: **0**

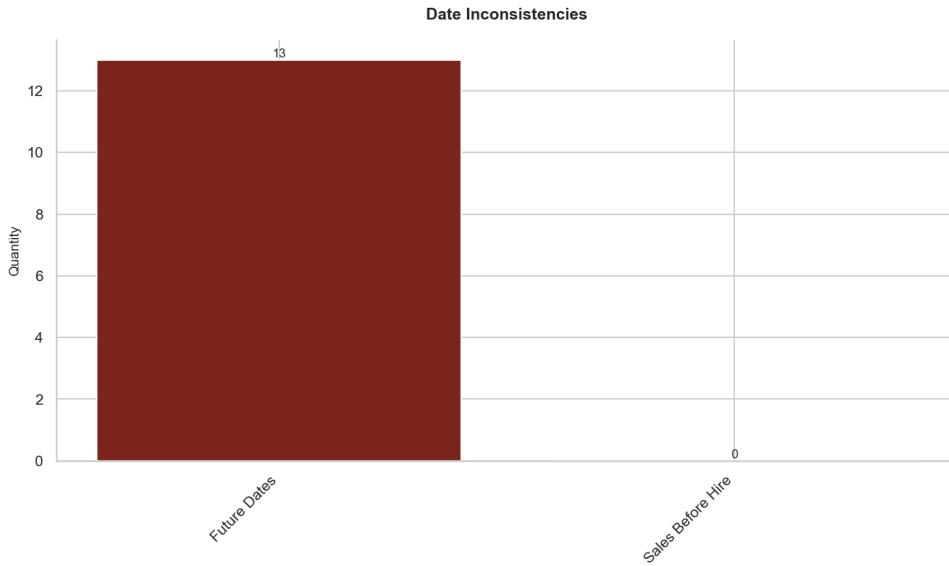


Figure 3: Date Issues

### 4. Applied Cleaning Process

---

#### 4.1 transformations performed

- Date validation (`sale_date > hire_date` of the seller)
- Reassignment of sales to valid employees when applicable
- Conversion of date columns to datetime
- Verification of `seller_employee_id` in employee dataset
- Price consistency validation (`unit_price * quantity`)

#### 4.2 Cleaning Statistics

Table 1: Cleaning Process Summary

Metric	Value
Original records	500
Records after cleaning	500
Records removed	0
Sales reassigned	0
Dates corrected	0

---

### 5. Relations Between Datasets

#### 5.1 Employee ID Validation

- Sales with valid Employee ID: **500**
- Sales with invalid Employee ID: **0**

#### 5.2 Referential Integrity

It was verified that each sale has a valid seller registered in the employee dataset. Sales with invalid IDs may indicate records of deleted employees or data entry errors.

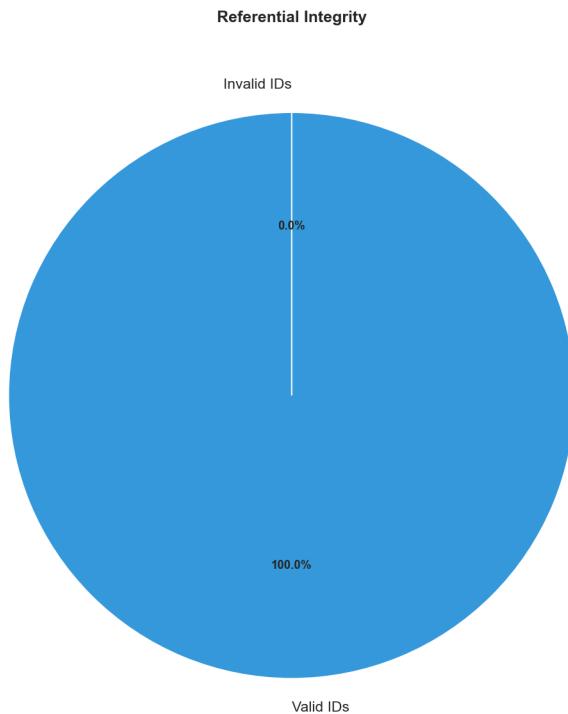


Figure 4: Referential Integrity

## 6. Pipeline Validation Results

### 6.1 Applied Validation Rules

- **is\_pending:** Status = 'Pending'
- **is\_completed:** Status = 'Completed'
- **is\_cancelled:** Status = 'Cancelled'
- **is\_future\_date:** Sale date is not in the future
- **validate\_prices:** Prices are positive
- **validate\_total\_price:** total\_price = quantity \* unit\_price

### 6.2 Validation Results

Table 2: Validation Results by Rule

Validation	VValid	Invalid	N/A	% Success
is_pending	178	322	0	35.6%
is_completed	170	330	0	34.0%
is_cancelled	152	348	0	30.4%
is_future_date	13	487	0	2.6%
validate_prices	500	0	0	100.0%
validate_total_price	500	0	0	100.0%

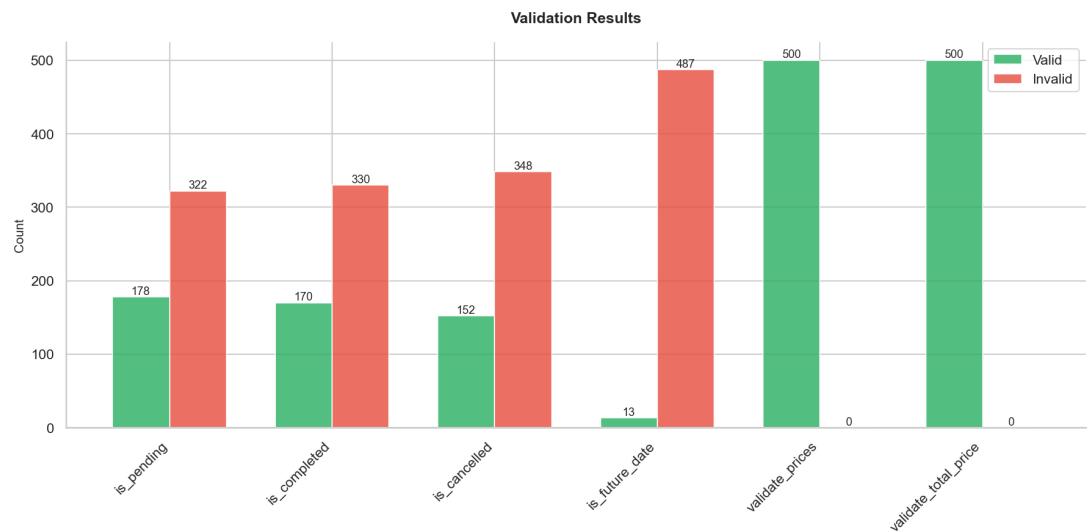


Figure 5: Validation Details