### **Dataset Overview: Before and After Cleaning**

#### **Before Cleaning:**

* **Total Rows**: 541,909
* **Key Issues**:
  1. **Missing Descriptions**: 1,454 rows had missing product descriptions.
  2. **Missing Customer IDs**: 135,080 rows had missing customer IDs.
  3. **Negative Quantities**: There were rows with negative quantities, potentially indicating returns or cancellations.
  4. **Zero or Negative Unit Prices**: Some rows contained zero or negative unit prices, suggesting data anomalies.
  5. **Duplicate Rows**: 5,268 duplicate rows existed in the dataset.

#### **After Cleaning:**

* **Total Rows**: 392,692
* **Actions Taken**:
  1. **Removed Rows with Missing Descriptions**: All rows with missing product descriptions were removed.
  2. **Removed Rows with Missing Customer IDs**: Rows with missing customer IDs were eliminated to ensure the dataset has complete records.
  3. **Removed Rows with Negative Quantities**: Rows with negative quantities (assumed to be incorrect or cancellations) were removed.
  4. **Removed Rows with Zero or Negative Unit Prices**: Rows with anomalies in pricing were eliminated to avoid inconsistencies in financial data.
  5. **Removed Duplicates**: All 5,268 duplicate rows were dropped, leaving only unique records.

#### **Summary:**

The dataset after cleaning has 149,217 fewer rows, with all key anomalies such as missing values, negative quantities, and zero/negative prices resolved. The cleaned dataset is now ready for accurate analysis without risks of distortion from data quality issues.