

DATA CALIBRATION EXPLANATION

Data calibration refers to the process of correcting, standardizing, and validating raw data values to ensure that derived metrics are accurate, consistent, and trustworthy before analytics and reporting.

In retail analytics, calibration is especially critical for financial fields, where even small inconsistencies can lead to incorrect revenue reporting.

Without calibration:

- Revenue metrics may be incorrect
- Dashboards may show misleading KPIs
- Business decisions could be impacted

The raw sales transaction data originates from multiple sources and contains:

- Incorrect or inconsistent total_amount values
- String-typed numeric fields
- Invalid or negative values
- Inconsistent timestamp formats

So we recalculate to identify inconsistencies

- total_amount = quantity × unit_price – discount
- The total_amount field from the source is not trusted
- It is recalculated using validated numeric fields
- Rounded to 2 decimal places to enforce financial precision

This is the code snippet that corresponds to calibrating data

```
silver_prepared = (
    silver_casted
    .withColumn(
        "transaction_timestamp",
```

```
F.to_utc_timestamp("transaction_timestamp", "UTC")  
)  
.withColumn(  
    "total_amount",  
    F.round(  
        F.col("quantity") * F.col("unit_price") - F.col("discount"), 2  
    )  
)  
)
```