

DATA CLEANING PROCESS USING SQL LITE

RAW DATA

```
SELECT * FROM SALES_RAW;
```

COUNT BLANKS

```
SELECT
```

```
    SUM(CASE WHEN QUANTITY_ORD = "" THEN 1 END) AS missing_quantity,  
    SUM(CASE WHEN PRICE_EACH = "" THEN 1 END) AS missing_price,  
    SUM(CASE WHEN SALES = "" THEN 1 END) AS missing_sales,  
    SUM(CASE WHEN PRODUCT_LINE = "" THEN 1 END) AS missing_product_line,  
    SUM(CASE WHEN CUSTOMER_NAME = "" THEN 1 END) AS missing_customer,  
    SUM(CASE WHEN CITY = "" THEN 1 END) AS missing_city,  
    SUM(CASE WHEN STATE = "" THEN 1 END) AS missing_state,  
    SUM(CASE WHEN ADDRESS_LINE2 = "" THEN 1 END) AS missing_address
```

```
FROM SALES_RAW;
```

SENTENCE CASE

```
UPDATE SALES_RAW
```

```
SET
```

```
    CUSTOMER_NAME = TRIM(UPPER(CUSTOMER_NAME)),  
    CITY = TRIM(UPPER(CITY)),  
    STATE = TRIM(UPPER(STATE)),  
    COUNTRY = TRIM(UPPER(COUNTRY)),  
    PRODUCT_LINE = TRIM(UPPER(PRODUCT_LINE)),  
    DEAL_SIZE = TRIM(UPPER(DEAL_SIZE));
```

MISSING NUMERIC VALUES TO 0

```
UPDATE SALES_RAW  
SET QUANTITY_ORD = 0  
WHERE QUANTITY_ORD = '';
```

FILLING BLANKS WITH "UNKNOWN"

```
UPDATE SALES_RAW  
SET STATE = 'UNKNOWN'  
WHERE STATE IS NULL OR STATE = '';
```

```
UPDATE SALES_RAW  
SET ADDRESS_LINE2 = 'UNKNOWN'  
WHERE ADDRESS_LINE2 IS NULL OR ADDRESS_LINE2 = '';
```

```
UPDATE SALES_RAW  
SET POSTALCODE = 'UNKNOWN'  
WHERE POSTALCODE IS NULL OR POSTALCODE = '';
```

DATE CLEANNG

```
UPDATE SALES_RAW  
SET ORD_DATE = SUBSTR(ORD_DATE, 1, 10);
```

REMOVING WHITESPACE

```
UPDATE sales_raw  
SET ORD_DATE = TRIM(ORD_DATE);
```

NEW CLEANED VERSION

```
CREATE TABLE CLEANED_TABLE AS  
SELECT *, (QUANTITY_ORD * PRICE_EACH) AS COMPUTED_REVENUE  
FROM SALES_RAW;  
THEN CONFIRM SELECT * FROM CLEANED_TABLE;
```

APPROXIMATING DECIMALS

```
UPDATE CLEANED_TABLE  
SET COMPUTED_REVENUE = ROUND(COMPUTED_REVENUE, 2);
```

REMOVING DUPLICATE

```
CREATE_TABLE Dupli_removal AS SELECT DISTINCT * FROM CLEANED_TABLE
```

CONFIRMATION

```
SELECT QUANTITY_ORD, ADDRESS_LINE1, COMPUTED_REVENUE FROM CLEANED_TABLE
```