**SQL for Data Cleaning: A Practical Guide**

While SQL is primarily used for data retrieval and manipulation, it can also be a powerful tool for data cleaning. Here are some common data cleaning tasks that can be accomplished using SQL:

**1. Handling Missing Values:**

* **Identifying Missing Values:**

SELECT \*

FROM your\_table

WHERE column\_name IS NULL;

* **Filling Missing Values:**
  + **With a Constant Value:**

UPDATE your\_table

SET column\_name = 'N/A'

WHERE column\_name IS NULL;

* **With a Calculated Value (e.g., mean, median):**

UPDATE your\_table

SET column\_name = (SELECT AVG(column\_name) FROM your\_table)

WHERE column\_name IS NULL;

**2. Removing Duplicates:**

DELETE FROM your\_table

WHERE rowid NOT IN (

SELECT MIN(rowid)

FROM your\_table

GROUP BY column1, column2, ...

);

**3. Correcting Inconsistent Data:**

* **Standardizing Text Formats:**

UPDATE your\_table

SET column\_name = UPPER(column\_name);

* **Correcting Typos:**

UPDATE your\_table

SET column\_name = REPLACE(column\_name, 'incorrect\_value', 'correct\_value');

**4. Outlier Detection and Handling:**

* **Using Statistical Functions:**

SELECT \*

FROM your\_table

WHERE column\_name > (SELECT AVG(column\_name) + 3 \* STDDEV(column\_name) FROM your\_table);

* **Using Domain Knowledge:**

UPDATE your\_table

SET column\_name = (SELECT AVG(column\_name) FROM your\_table)

WHERE column\_name > 100; -- Assuming 100 is an unrealistic value