Akhil's Data Lab

# 8 SQL TECHNIQUES FOR DATA CLEANING


@akhildubey1004

## Replace NULL values with a default value or remove them based on the context.

```sql
SELECT COALESCE(column_name, 'DefaultValue') AS column_name
FROM table_name;
```

## Identify and delete duplicate rows based on key columns.

```sql
WITH CTE AS (
    SELECT id, column1, column2,
    ROW_NUMBER() OVER (PARTITION BY column1, column2 ORDER BY id) AS row_num
    FROM table_name
)
DELETE FROM table_name
WHERE id IN (
    SELECT id
    FROM CTE
    WHERE row_num > 1
);
```

Akhil's Data Lab

## Convert text to lower or upper case to ensure consistency.

```sql
SELECT LOWER(column_name) AS cleaned_column
FROM table_name;
```

## Remove extra spaces from text fields.

```sql
SELECT TRIM(column_name) AS cleaned_column
FROM table_name;
```

# Convert date strings into a consistent date format.

```sql
SELECT STR_TO_DATE(column_name, '%m/%d/%Y') AS formatted_date
FROM table_name;
```

# Identify and manage outliers in numerical data.

```sql
SELECT *
FROM table_name
WHERE column_name BETWEEN lower_limit AND upper_limit;
```

Akhil's Data Lab

# Replace or remove special characters in text fields.

```sql
SELECT REGEXP_REPLACE(column_name, '[^a-zA-Z0-9]', '') AS cleaned_column
FROM table_name;
```

# Standardize values in categorical columns

```sql
UPDATE table_name
SET column_name = 'Male'
WHERE column_name IN ('M', 'male');
```

# Want more content like this?

Akhil's Data Lab

Tap that follow button and stay in the loop!

Like Comment Share Save

@akhildubey1004