
# Top Data Cleaning Methods in SQL

```sql
SELECT column1, column2
FROM table_name
group by column1, column2;
```

👉 swipe right

# 1.Removing duplicates

Duplicates in a dataset can skew analysis results. You can remove duplicates using the `DISTINCT` keyword or by using the `GROUP BY` clause.

```sql
SELECT column1, column2
FROM table_name
group by column1, column2;
```

# 2.Handling missing values

Missing values can impact the accuracy of analysis. You can identify and handle missing values using SQL functions like `IS NULL` or `COALESCE`, and then decide whether to replace them with a default value or remove them

```sql
SELECT *
FROM table_name
WHERE column_name IS NOT NULL;
```

# 3.Correcting data inconsistencies

Data inconsistencies such as typos or variations in formatting can be corrected using SQL string functions like `UPPER`, `LOWER`, or `REPLACE`.

```sql
UPDATE table_name
SET column_name = UPPER(column_name);
```

# 4.Standardizing data formats

Standardizing data formats ensures consistency across the dataset. SQL functions like `TO_DATE`, `TO_CHAR`, or `CAST` can be used to convert data into a specific format.

```sql
SELECT
        TO_DATE(column, 'YYYY/MM/DD') AS
        formatted_date
FROM table_name;
```

# 5.Removing outliers

Outliers can significantly affect statistical analysis. You can identify and remove outliers using SQL aggregate functions along with statistical techniques like z-score or percentile

```sql
SELECT *
FROM table_name
WHERE column_name BETWEEN
 (SELECT PERCENTILE_CONT(0.05) WITHIN GROUP (ORDER BY column_name)
 FROM table_name)
 AND
 (SELECT PERCENTILE_CONT(0.95) WITHIN GROUP (ORDER BY column_name)
 FROM table_name);
```

Sudhanshu1st

# 6.Data validation

Validating data against predefined rules or constraints helps ensure data integrity. SQL constraints like NOT NULL, UNIQUE, or CHECK can be used during table creation or alteration.

```sql
CREATE TABLE table_name (
 column1 INT NOT NULL UNIQUE,
 column2 VARCHAR(50) CHECK (column2 IN ('value1', 'value2')),
 ..);
```

Sudhanshu1st