SQL Functions for Data Cleaning

Clean and Optimize Your Data Using SQL



Why Data Cleaning is Important?

 Ensures accuracy and consistency in your data.

 Prepares data for analysis and reporting.

 Eliminates errors, duplicates, and inconsistencies.

Data Issues that Need Cleaning

- Duplicates
- Inconsistent Formats
- Null Values
- Incorrect Data Types
- Outliers
- Extra Whitespaces

Overview of Useful SQL Functions

- TRIM: Removes leading and trailing spaces.
- REPLACE: Substitutes part of a string.
- COALESCE: Replaces null values.
- CAST/CONVERT: Changes data types.
- SUBSTRING: Extracts part of a string.
- UPPER/LOWER: Standardizes text to uppercase or lowercase.

Removing Duplicate Rows

- Function: SELECT DISTINCT
- Example:

SELECT DISTINCT column_name FROM table_name;

Use Case: Ensures each record in a result set is unique.

Replacing Null Values

Function: COALESCE

Example:

SELECT COALESCE(column_name, 'default_value')
FROM table_name;

 Use Case: Replace nulls with default values for clean reports.

Consistent Data Formatting

Functions: UPPER, LOWER, CAST

• Example:

SELECT UPPER(column_name), CAST(column_name AS INT) FROM table_name;

 Use Case: Ensure consistent format across your dataset.

Cleaning Extra Whitespaces

• Function: TRIM

• Example:

SELECT TRIM(column_name)
FROM table_name;

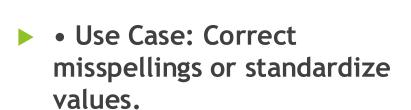
 Use Case: Remove unwanted spaces from text data.

Replacing Incorrect Text

• Function: REPLACE

Example:

SELECT
REPLACE(column_name,
'old_text', 'new_text')
FROM table_name;





Summary of SQL Cleaning Functions

- TRIM: Removes spaces.
- REPLACE: Substitutes strings.



- COALESCE: Handles nulls.
- DISTINCT: Removes duplicates.
- CAST: Changes data types.



► Stay connected for more SQL tips and data cleaning techniques!