DATA ENGINEERING 101

IMPORTANT SQL FOR DATA CLEANING AND MORE...

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU IDENTIFY DUPLICATE ROWS IN A DATASET USING SQL?

To identify duplicate rows, you can use a GROUP BY clause with a HAVING count greater than one.

For example:

```
SELECT column_name, COUNT(*)
FROM table_name
GROUP BY column_name HAVING COUNT(*) > 1;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW CAN YOU REMOVE DUPLICATE ROWS FROM A SQL TABLE?

To remove duplicates, you can delete rows that have a higher ID (assuming 'id' is a unique identifier).

Example:

```
DELETE FROM table_name
WHERE id NOT IN (  SELECT MIN(id)
FROM table_name  GROUP BY
column_to_check );
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO HANDLE MISSING VALUES IN SQL?

You can use the COALESCE function to handle missing values by providing a default value when there is a NULL.

Example:

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

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU CHANGE INCORRECT DATA IN A COLUMN?

To correct data in a column, use the UPDATE statement.

For example, if correcting a state abbreviation:

```
UPDATE table_name SET column_name = 'TX'
WHERE column_name = 'Tex';
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW CAN YOU TRANSFORM DATA TO A DIFFERENT SCALE IN SQL?

To transform data, for instance scaling a salary column, you can multiply each value by a constant factor:

```
SELECT id, salary, salary * 1.10 AS
adjusted_salary
FROM employees;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO FILTER OUT RECORDS BASED ON A CONDITION IN SQL?

To filter records, use the WHERE clause. For example, to exclude records where the salary is below a certain threshold:

```
SELECT *
FROM employees WHERE salary > 50000;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

WHAT SQL COMMAND IS USED TO SORT DATA?

To sort data, use the ORDER BY clause. For instance, to sort employees by salary in descending order:

```
SELECT *
FROM employees
ORDER BY salary DESC;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO GROUP DATA IN SQL?

To group data by a specific attribute, use the GROUP BY clause.

For example, to group sales by region:

```
SELECT region, SUM(sales)
FROM sales_data GROUP BY region;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU USE SQL TO STANDARDIZE DATA FORMATS?

To standardize formats, such as dates, use conversion functions like TO_DATE:

```
SELECT TO_DATE(date_string, 'YYYY-MM-DD')
AS standardized_date
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO PERFORM CONDITIONAL DATA CLEANING IN SQL?

Use the CASE statement to apply conditional logic. For example, updating status based on age:

```
UPDATE table_name SET status = CASE
WHEN age < 20 THEN 'young'
WHEN age BETWEEN 20 AND 60
THEN 'adult'  ELSE 'senior' END;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU CONVERT STRING DATA TO UPPERCASE IN SQL?

To convert text to uppercase, use the UPPER() function.

Example:

```
SELECT UPPER(column_name) AS uppercase_text
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU SPLIT A COLUMN INTO MULTIPLE COLUMNS IN SQL?

To split a column, you can use the SPLIT_PART() function in PostgreSQL.

Example:

```
SELECT SPLIT_PART(full_name, ' ', 1) AS first_name,  SPLIT_PART(full_name, ' ', 2) AS last_name
FROM employees;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO FIND THE LENGTH OF STRINGS IN A SQL COLUMN?

To find the length of strings, use the LENGTH() function.

Example:

```
SELECT LENGTH(column_name) AS
string_length
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU REPLACE PARTS OF A STRING IN SQL?

To replace parts of a string, use the REPLACE() function.

Example:

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

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO ROUND NUMERIC DATA IN SQL?

To round numeric values, use the ROUND() function.

Example:

```
SELECT ROUND(column_name, decimal_places) AS rounded_value
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU FIND THE MINIMUM AND MAXIMUM VALUES IN SQL?

Use the MIN() and MAX() functions to find minimum and maximum values.

Example:

```
SELECT MIN(column_name) AS min_value,
MAX(column_name) AS max_value
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO CALCULATE THE AVERAGE OF A COLUMN IN SQL?

To calculate the average, use the AVG() function.

Example:

```
SELECT AVG(column_name) AS average_value
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU CONCATENATE COLUMNS IN SQL?

To concatenate columns, use the CONCAT() function.

Example:

```
SELECT CONCAT(first_name, ' ', last_name) AS full_name FROM employees;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO SELECT UNIQUE RECORDS IN SQL?

To select unique records, use the DISTINCT keyword.

Example:

SELECT DISTINCT column_name FROM table_name;

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU TRIM SPACES FROM DATA IN SQL?

To trim spaces, use the TRIM() function.

Example:

```
SELECT TRIM(column_name) AS trimmed_text
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO HANDLE NULL VALUES USING SQL COALESCE?

Use the COALESCE() function to replace NULLs with a specific value.

Example:

```
SELECT COALESCE(column_name, 'default_value') AS new_value FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU USE SQL TO LIMIT THE NUMBER OF RETURNED RECORDS?

To limit results, use the LIMIT keyword.

Example:

```
SELECT * FROM table_name LIMIT 10;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO JOIN TWO TABLES BASED ON A COMMON COLUMN IN SQL?

To join tables, use the JOIN clause.

Example:

```
SELECT a.*, b.column_name
FROM table_a a JOIN table_b b ON
a.common_column = b.common_column;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU CALCULATE THE SUM OF A COLUMN IN SQL?

To calculate the sum, use the SUM() function.

Example:

```
SELECT SUM(column_name) AS total FROM
table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO UPDATE DATA IN A SQL TABLE?

To update data, use the UPDATE statement.

Example:

```
UPDATE table_name
SET column_name = 'new_value'
WHERE condition_column = 'condition';
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU DELETE RECORDS FROM A SQL TABLE?

To delete records, use the DELETE statement.

Example:

```
DELETE FROM table_name
WHERE condition_column = 'condition';
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO ADD A NEW COLUMN TO AN EXISTING SQL TABLE?

To add a column, use the ALTER TABLE statement.

Example:

```
ALTER TABLE table_name
ADD column_name datatype;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU CREATE AN INDEX IN SQL TO IMPROVE QUERY PERFORMANCE?

To create an index, use the CREATE INDEX statement.

Example:

```
CREATE INDEX idx_column_name
ON table_name (column_name);
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO EXTRACT THE YEAR FROM A DATE COLUMN IN SQL?

To extract the year, use the EXTRACT() function.

Example:

```
SELECT EXTRACT(YEAR FROM date_column)
AS year
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW DO YOU HANDLE DATE RANGES IN SQL QUERIES?

To handle date ranges, use the BETWEEN keyword.

Example:

```
SELECT * FROM table_name
WHERE date_column
BETWEEN 'start_date' AND 'end_date';
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO ENSURE A COLUMN ONLY CONTAINS CERTAIN VALUES IN SQL?

Use a CHECK constraint during table creation.

Example:

```
CREATE TABLE example_table
(   status VARCHAR(10) CHECK (status IN
('active', 'inactive')),   ...  );
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO CHANGE THE DATA TYPE OF A COLUMN IN SQL?

To change the data type, use the ALTER TABLE statement with ALTER COLUMN.

Example:

ALTER TABLE table_name ALTER COLUMN column_name TYPE new_data_type;

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO IMPLEMENT DATA VALIDATION USING SQL?

Implement data validation by using constraints or triggers.

Example for a constraint:

```
ALTER TABLE table_name  ADD CONSTRAINT
check_constraint_name CHECK
(column_name > 0);
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO FIND THE MOST FREQUENTLY OCCURRING VALUE IN A COLUMN?

To find the most common value, use MODE() in a statistical function.

Example:

```
SELECT MODE() WITHIN
GROUP (ORDER BY column_name) AS
most_frequent  FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO REMOVE LEADING ZEROS FROM DATA IN A SQL COLUMN?

Use the LTRIM() function to trim zeros.

Example:

```
SELECT LTRIM(column_name, '0') AS
trimmed_column  FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO CREATE A BACKUP OF A SQL TABLE?

To create a backup, use a SELECT INTO statement to copy the entire table.

Example:

```
SELECT * INTO backup_table_name
FROM original_table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO REVERSE A STRING IN SQL?

Use the REVERSE() function to reverse a string.

Example:

```
SELECT REVERSE(column_name) AS reversed_string  FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO FIND RECORDS THAT MATCH PATTERNS IN SQL?

Use the LIKE or SIMILAR TO operators for pattern matching.

Example:

```
SELECT *
FROM table_name
WHERE column_name LIKE 'pattern%';
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO NORMALIZE DATA IN SQL?

Normalize data by adjusting values to a common scale.

Example for zero to one normalization:

```
SELECT (column_name - MIN(column_name))
/ (MAX(column_name) - MIN(column_name))
AS normalized_column
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO MERGE TWO SQL TABLES WITHOUT DUPLICATES?

Use the UNION operator to combine and remove duplicates.

Example:

```
SELECT column_name
FROM table_one
UNION
SELECT column_name
FROM table_two;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO COUNT DISTINCT VALUES IN A SQL COLUMN?

Use COUNT(DISTINCT column_name) to count unique values.

Example:

```
SELECT COUNT(DISTINCT column_name) AS unique_count
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO CHECK IF A COLUMN CONTAINS ONLY NUMERIC DATA IN SQL?

Check for non-numeric data using NOT LIKE and wildcard.

Example:

```
SELECT *
FROM table_name
WHERE column_name NOT LIKE '%[^0-9]%';
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO EXCLUDE ROWS BASED ON MULTIPLE CONDITIONS IN SQL?

Use the AND operator to combine conditions in the WHERE clause.

Example:

```
SELECT *
FROM table_name
WHERE condition_one AND condition_two;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO ROUND NUMBERS TO THE NEAREST THOUSAND IN SQL?

Use the ROUND() function with negative scale for rounding.

Example:

```
SELECT ROUND(column_name, -3) AS
rounded_value  FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

HOW TO SPLIT A COMMA-SEPARATED STRING INTO MULTIPLE ROWS IN SQL?

Use string splitting functions and UNNEST.

Example in PostgreSQL:

```
SELECT
UNNEST(STRING_TO_ARRAY(column_name,
',')) AS split_values
FROM table_name;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO AUTOMATICALLY UPDATE A TIMESTAMP COLUMN IN SQL?

Use a trigger to update a timestamp column upon modifications.

Example:

```
CREATE TRIGGER update_timestamp
BEFORE UPDATE ON table_name
FOR EACH ROW
EXECUTE FUNCTION
update_modified_column();
```

SHWETANK SINGH

GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO CREATE A VIEW FOR SIMPLIFIED QUERYING IN SQL?

Use CREATE VIEW to simplify data access.

Example:

```
CREATE VIEW view_name
AS
SELECT column_one, column_two
FROM table_name  WHERE condition;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO HANDLE CASE SENSITIVITY IN SQL QUERIES?

Use LOWER() or UPPER() to normalize case.

Example:

```
SELECT *  FROM table_name
WHERE LOWER(column_name) =
LOWER('SomeValue');
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO PERFORM DATE SUBTRACTION IN SQL?

Subtract dates to find intervals.

Example:

```
SELECT CURRENT_DATE - INTERVAL '1 day' AS yesterday;
```

SHWETANK SINGH
GRITSETGROW - GSGLEARN.COM

gsglearn.com

HOW TO EXTRACT THE FIRST LETTER OF EACH WORD IN A STRING IN SQL?

Use string functions to extract initials.

Example:

```
SELECT
STRING_AGG(LEFT(SPLIT_PART(full_name, ' ',
generate_series), 1), '') AS initials
FROM table_name,
GENERATE_SERIES(1,
ARRAY_LENGTH(STRING_TO_ARRAY(full_name
, ' '), 1))
```

SHWETANK SINGH

GRITSETGROW - GSGLEARN.COM

THANK YOU