Tajamul Khan

SQL Data Cleaning

@Tajamulkhann

Handle Missing Values

COALESCE(col, default)
Replace NULL with default.

IFNULL(col, default)
Same as COALESCE in some DBs.

AVG/COUNT with **COALESCE**
Prevents wrong aggregations.

```sql
SELECT COALESCE(email, 'unknown') AS email FROM users;
```

@Tajamulkhann

Remove Duplicates

DISTINCT
Quick way to remove duplicates.

ROW_NUMBER() + PARTITION BY
Keep only the latest/valid record.

```sql
DELETE FROM users
WHERE id NOT IN (
   SELECT MIN(id)
   FROM users
   GROUP BY user_id
);
```

@Tajamulkhann

Standardize Text

UPPER(col), LOWER(col)
Consistent casing.

TRIM(col)
Remove extra spaces.

REPLACE / REGEXP
Clean special characters.

```sql
SELECT LOWER(TRIM(name)) AS clean_name FROM customers;
```

@Tajamulkhann

Fix Date Formats

TO_DATE(string, format)
Standardize formats.

EXTRACT(YEAR | MONTH | DAY FROM date_col)
Break down dates.

```sql
SELECT TO_DATE(order_date, 'YYYY-MM-DD') FROM orders;
```

@Tajamulkhann

Handle Outliers

Identify:
Use AVG() + STDDEV().

Remove:
Delete rows beyond threshold.

Cap:
Limit values instead of deleting.

```sql

SELECT *
FROM sales
WHERE amount > (AVG(amount) + 3*STDDEV(amount));
```

@Tajamulkhann

Detect & Fix Data Entry Errors

REGEXP
Match patterns (phone, email).

Example:
Find invalid phone numbers.

```sql
SELECT phone FROM customers
WHERE phone NOT REGEXP '^[0-9]{10}$';
```

@Tajamulkhann

Standardize Categories

Use **CASE** or **mapping tables** to unify categorical values (e.g., "NY" vs "New York").

```sql

SELECT CASE
         WHEN state IN ('NY','New York') THEN 'New York'
         ELSE state
       END AS clean_state
FROM addresses;
```

@Tajamulkhann

Remove Special Characters

Use **REGEXP_REPLACE()** or **REPLACE()** to strip unwanted chars.

```sql
SELECT REGEXP_REPLACE(address, '[^a-zA-Z0-9 ]', '') FROM users;
```

@Tajamulkhann

Practical Tips

Always clean before joining datasets.

Profile data first (NULL %, distinct values, max/min, etc).

Document assumptions (e.g., how missing values handled).

@Tajamulkhann

Found Helpful ?

Repost

Follow for more!