SUMMARIZING AND CLEANING DATA IN SQL

Query | Query History

```sql
-- Checking for duplicates in film table

SELECT title, description, release_year, length, replacement_cost, special_features, fulltext,
COUNT (*)
FROM film
GROUP BY title, description, release_year, length, replacement_cost, special_features, fulltext
HAVING COUNT (*)>1;
```

Data Output | Messages | Notifications

title character varying (255)	description text	release_year integer	length smallint	replacement_cost numeric (5,2)	special_features text[]	fulltext tsvector	count bigint

Query | Query History

```sql
-- Checking for duplicates in customer table

SELECT store_id, first_name, last_name, email, address_id,
COUNT (*)
FROM customer
GROUP BY store_id, first_name, last_name, email, address_id
HAVING COUNT (*)>1;
```

Data Output | Messages | Notifications

store_id smallint	first_name character varying (45)	last_name character varying (45)	email character varying (50)	address_id smallint	count bigint

Looking for non-uniform data :

SELECT DISTINCT column_ name

FROM table_name

To amend :

UPDATE table_name

SET column_name = ‘desired value’

WHERE column_name IN (‘variant1’ , ‘variant2’ , ‘variant3’)

SUMMARIZING AND CLEANING DATA IN SQL

Looking for missing data :

```
SELECT *

FROM table_name

WHERE column_name IS NULL
```

Query | Query History

```sql
-- Summarizing film table data

SELECT MIN(rental_rate) AS minimum_rental_rate,
       MAX(rental_rate) AS maximum_rental_rate,
       AVG(rental_rate) AS average_rental_rate,
       MIN(replacement_cost) AS minimum_replacement_cost,
       MAX(replacement_cost) AS maximum_replacement_cost,
       AVG(replacement_cost) AS average_replacement_cost,
       MIN(rental_duration) AS minimum_rental_duration,
       MAX(rental_duration) AS maximum_rental_duration,
       AVG(rental_duration) AS average_rental_duration,
       MIN(length) AS minimum_length,
       MAX(length) AS maximum_length,
       AVG(length) AS average_length,
       MODE() WITHIN GROUP (ORDER BY release_year) AS mode_release_year,
       MODE() WITHIN GROUP (ORDER BY rating) AS mode_rating,
       MODE() WITHIN GROUP (ORDER BY language_id) AS mode_language_id,
       MODE() WITHIN GROUP (ORDER BY special_features) AS mode_special_features,
       MODE() WITHIN GROUP (ORDER BY last_update) AS mode_last_update
FROM film;
```