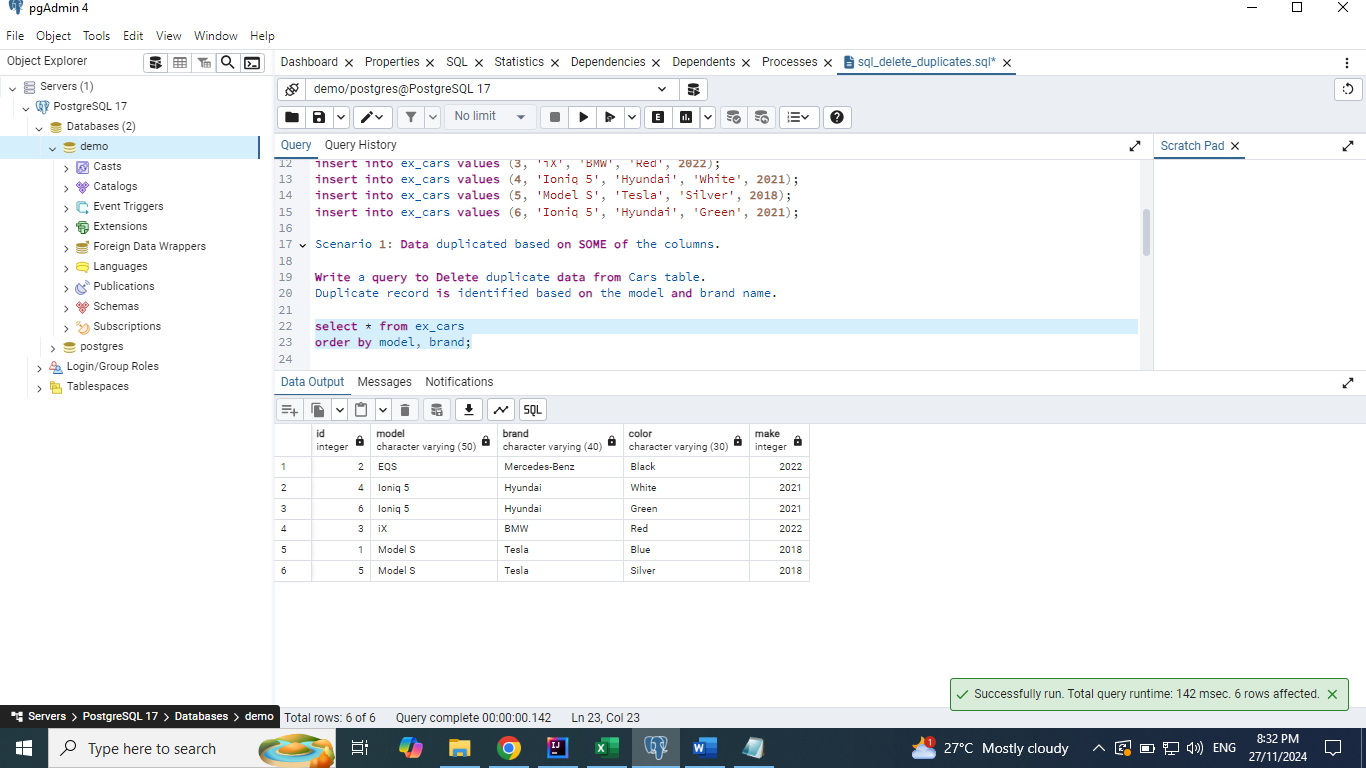
**Delete Duplicate Records**

# Scenario 1: Data duplicated based on SOME of the columns.

# Write a query to Delete duplicate data from Cars table.

# Duplicate record is identified based on the model and brand name.

**select \* from ex\_cars order by model, brand;**



## Solution 1: Delete using Unique Identifier

**select model, brand, count(\*)**

**from ex\_cars**

**group by model, brand**

**having count(\*) > 1;**

**select model, brand, max(id)**

**from ex\_cars**

**group by model, brand**

**having count(\*) > 1;**

**delete from ex\_cars**

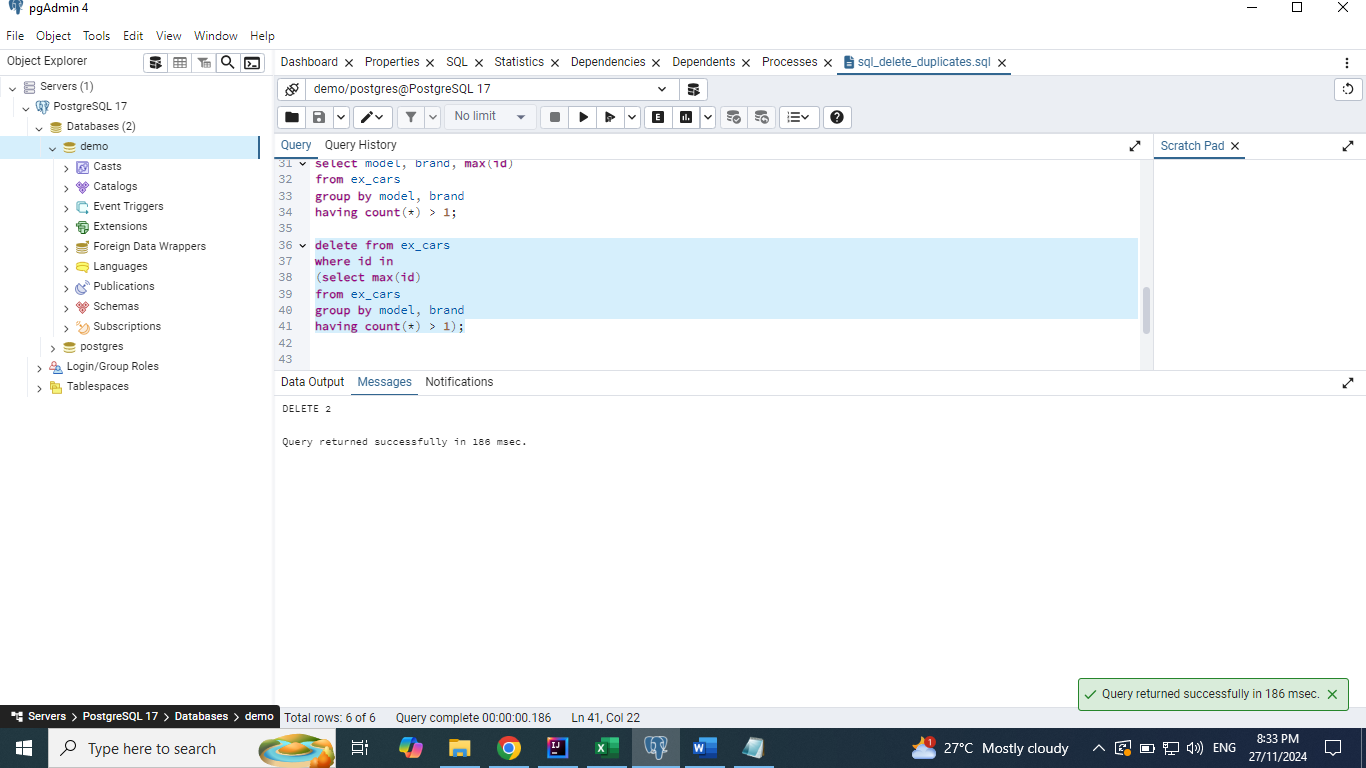
**where id in**

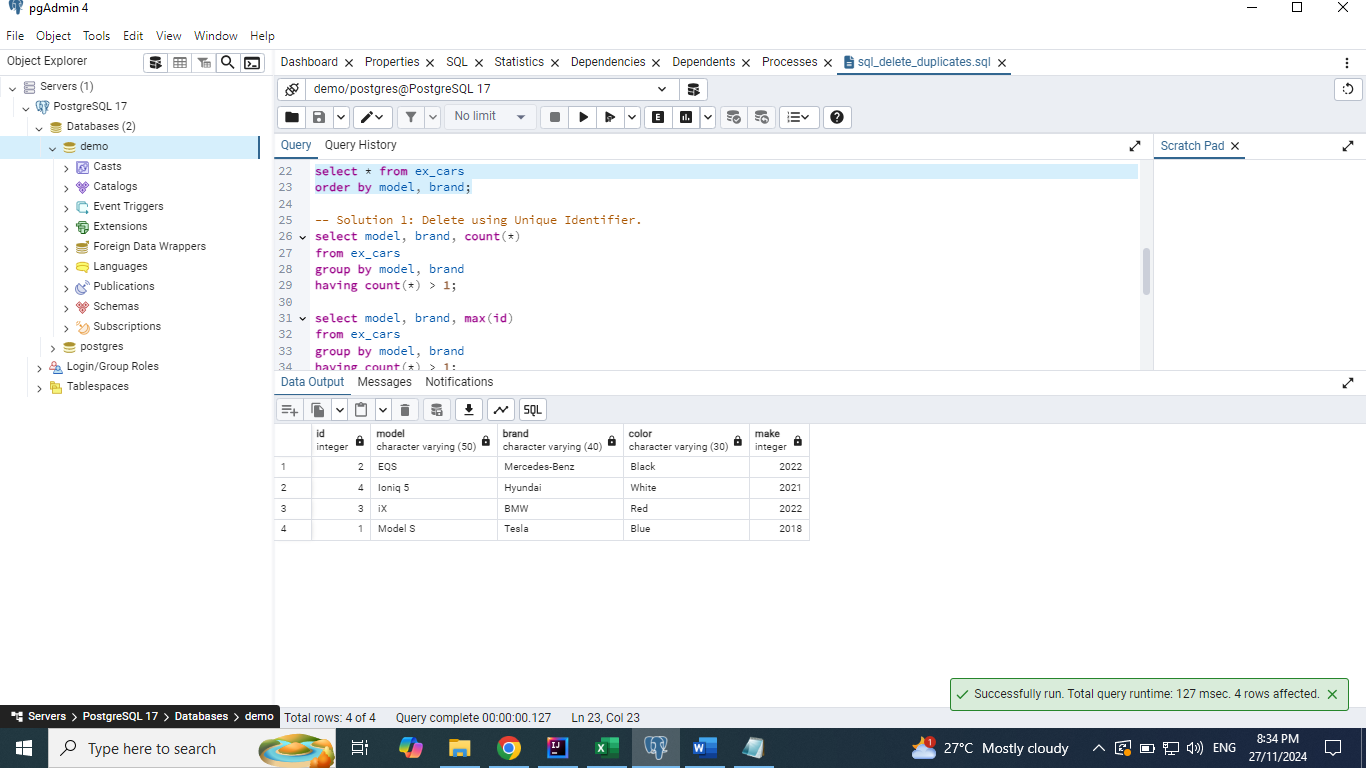
**(select max(id)**

**from ex\_cars**

**group by model, brand**

**having count(\*) > 1);**





## Solution 2: Using Self join

**select \***

**from ex\_cars c1**

**join ex\_cars c2**

**on c1.model = c2.model and c1.brand = c2.brand**

**where c1.id < c2.id;**

**select c2.\***

**from ex\_cars c1**

**join ex\_cars c2**

**on c1.model = c2.model and c1.brand = c2.brand**

**where c1.id < c2.id;**

**delete from ex\_cars**

**where id in**

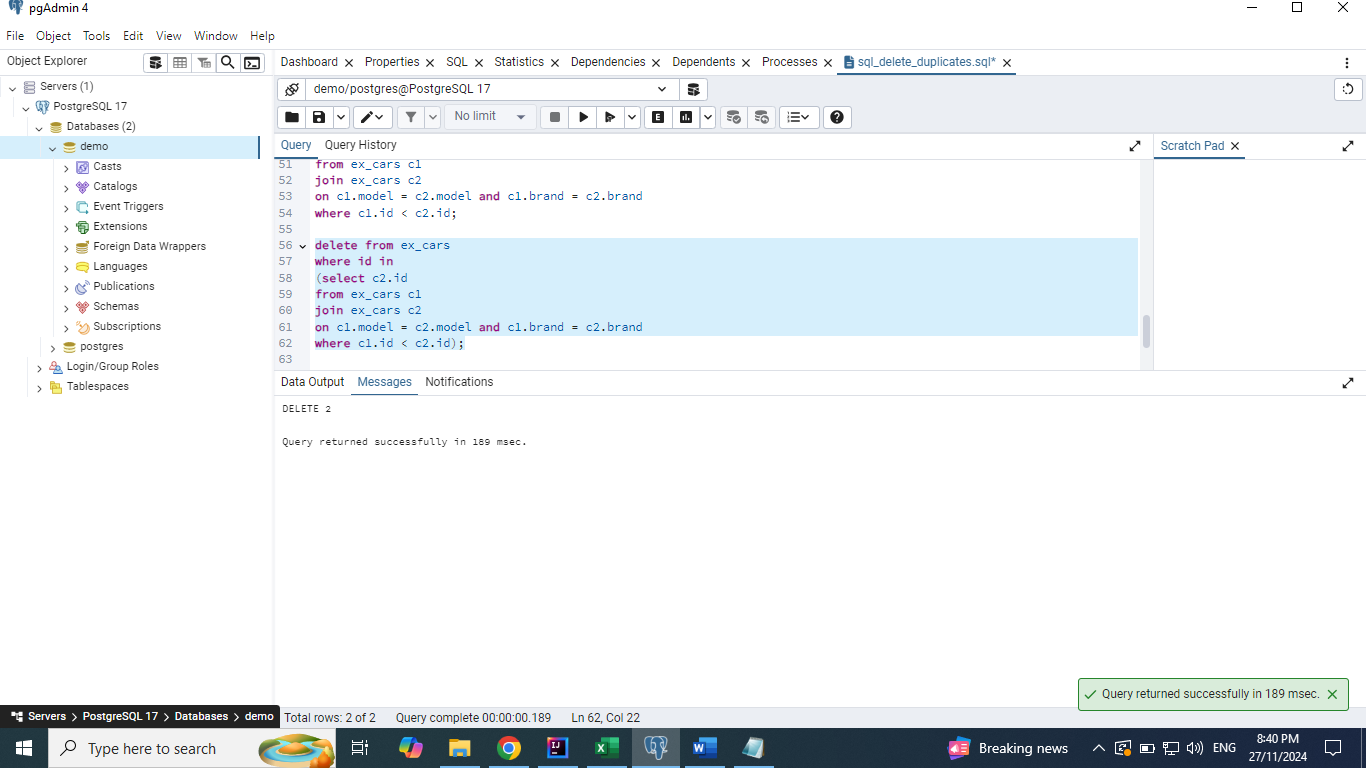
**(select c2.id**

**from ex\_cars c1**

**join ex\_cars c2**

**on c1.model = c2.model and c1.brand = c2.brand**

**where c1.id < c2.id);**



## Solution 3: Using Window function

**select \*,**

**row\_number() over(partition by model, brand) as rn**

**from ex\_cars;**

**select id**

**from**

**(select \*,**

**row\_number() over(partition by model, brand) as rn**

**from ex\_cars ) x**

**where x.rn > 1;**

**delete from ex\_cars**

**where id in**

**(select id**

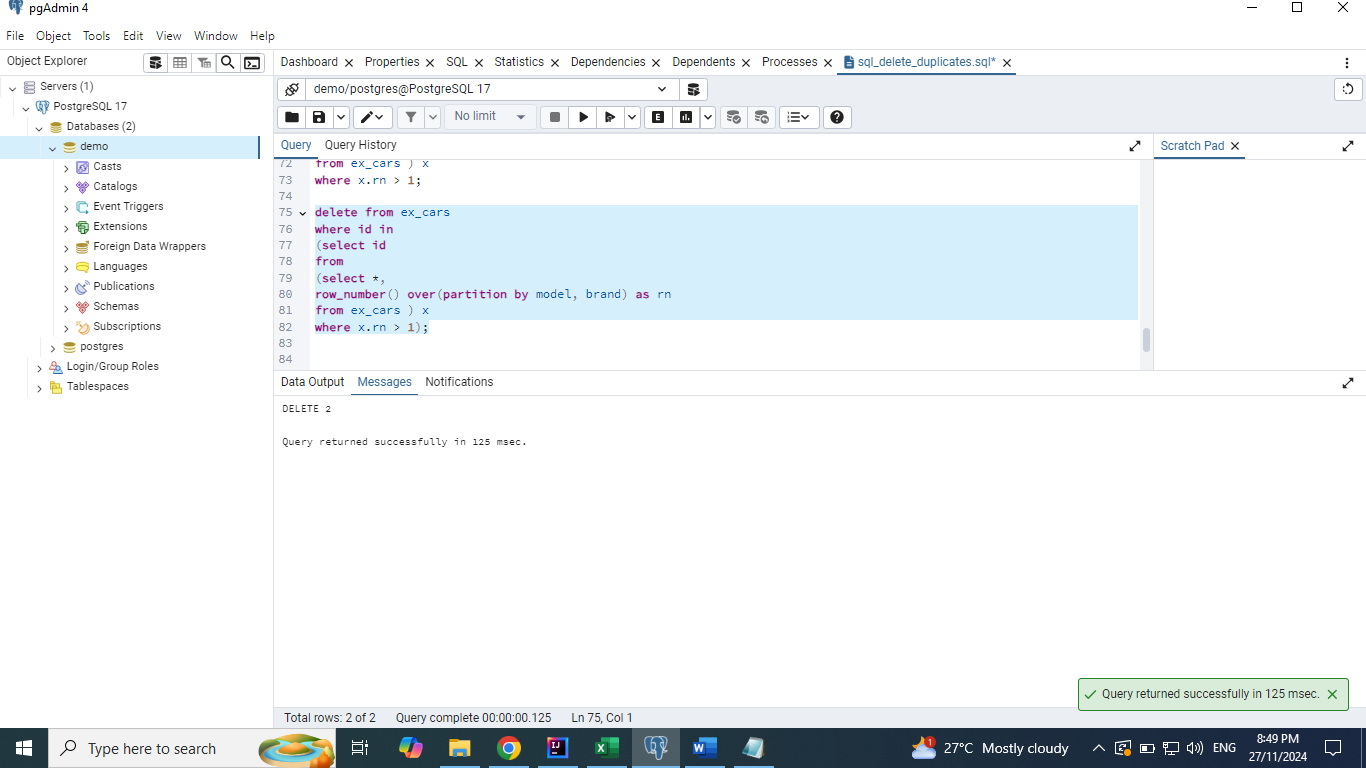
**from**

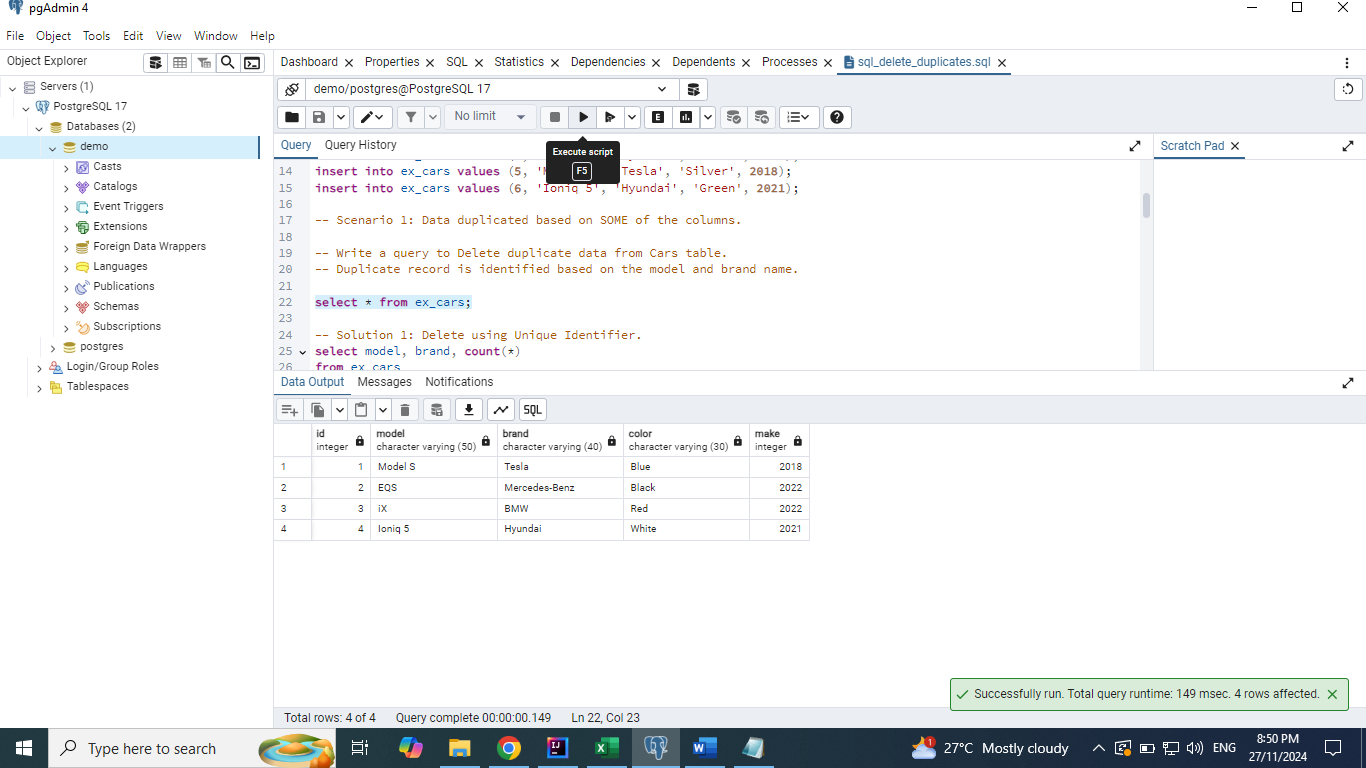
**(select \*,**

**row\_number() over(partition by model, brand) as rn**

**from ex\_cars ) x**

**where x.rn > 1);**





## Solution 4: Using Min function. This deletes even multiple duplicate records

**select min(id)**

**from ex\_cars**

**group by model, brand;**

**delete from ex\_cars**

**where id not in**

**(select min(id)**

**from ex\_cars**

**group by model, brand);**

