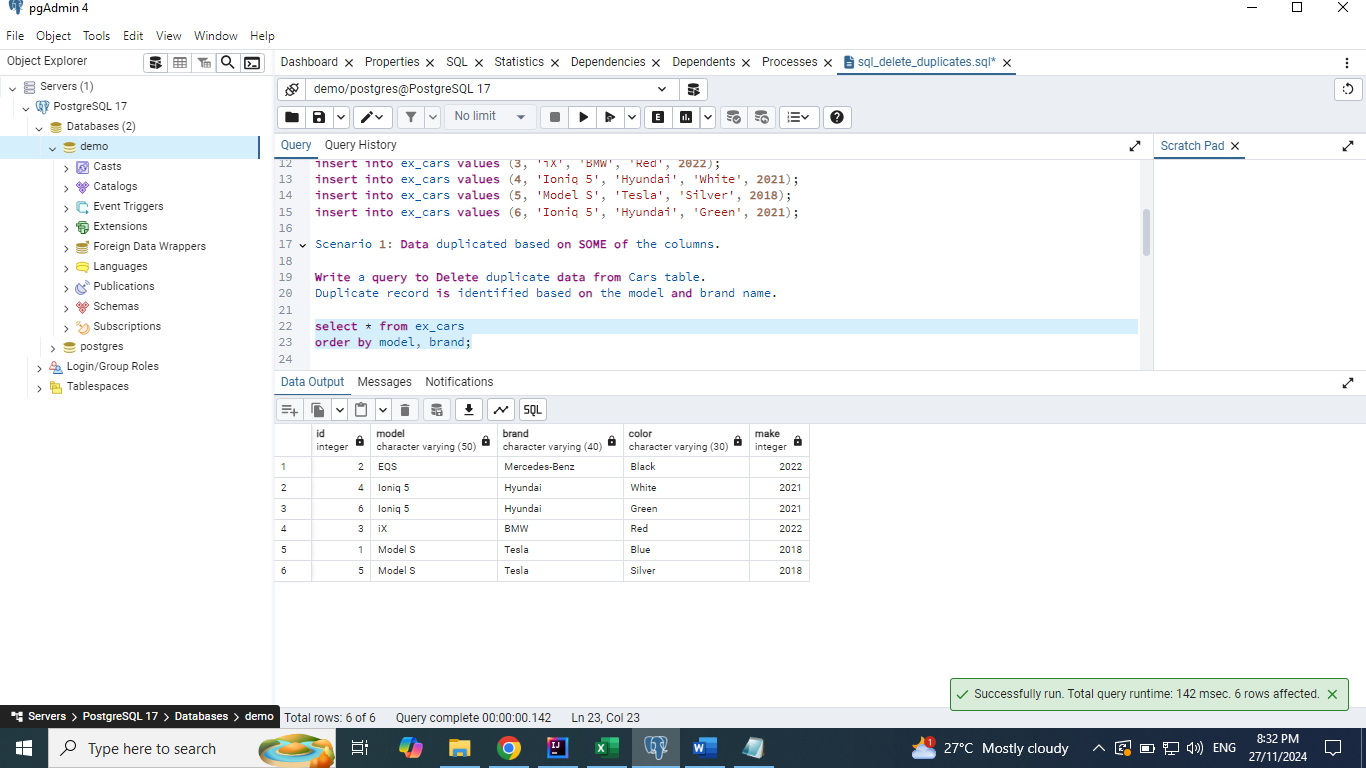
**Delete Duplicate Records**

# 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**

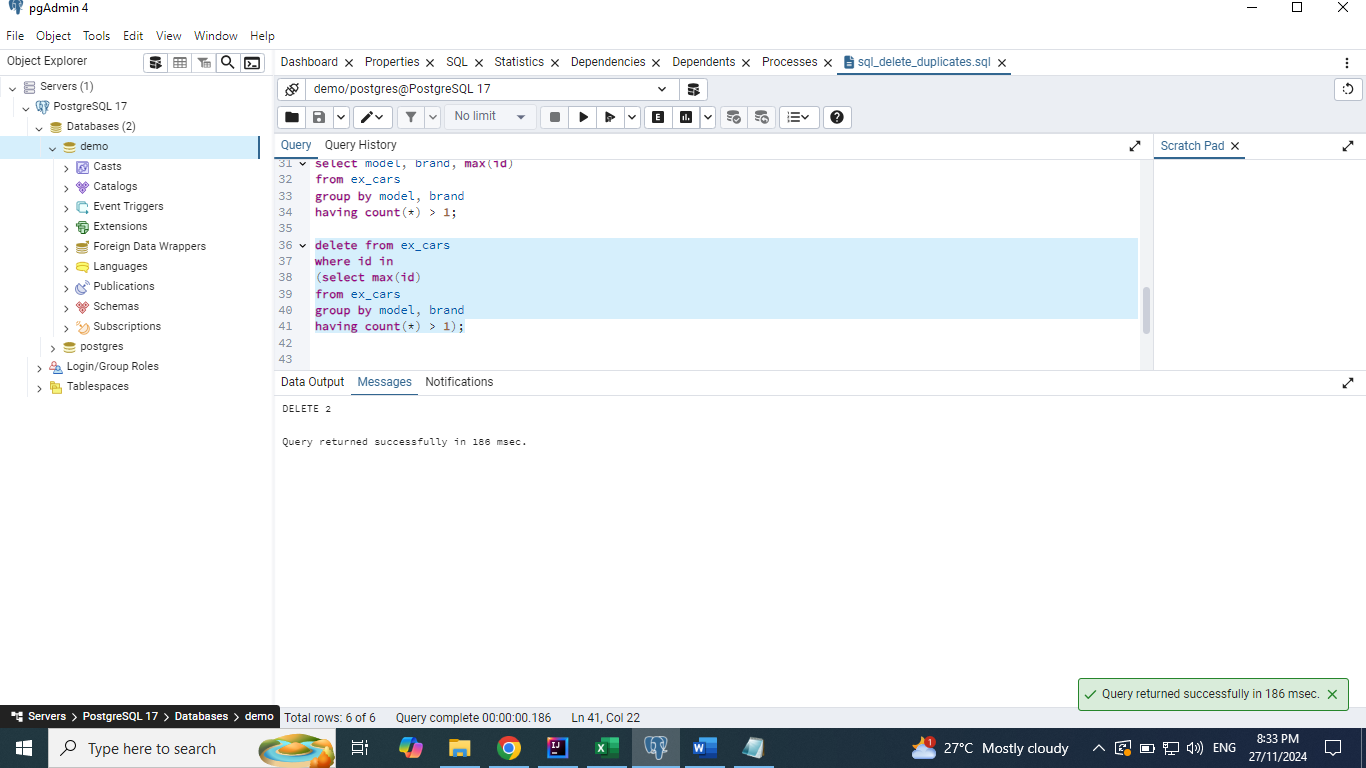
**where id in**

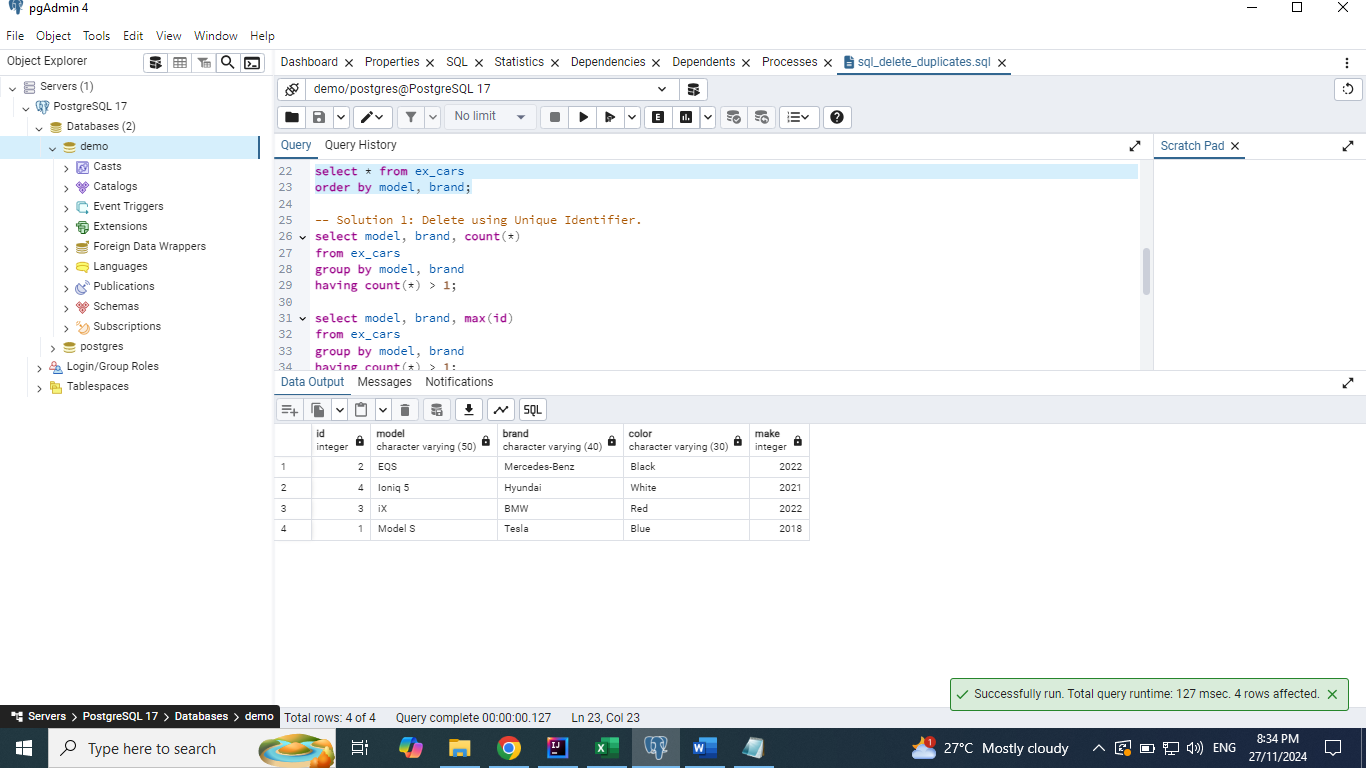
**(select max(id)**

**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**

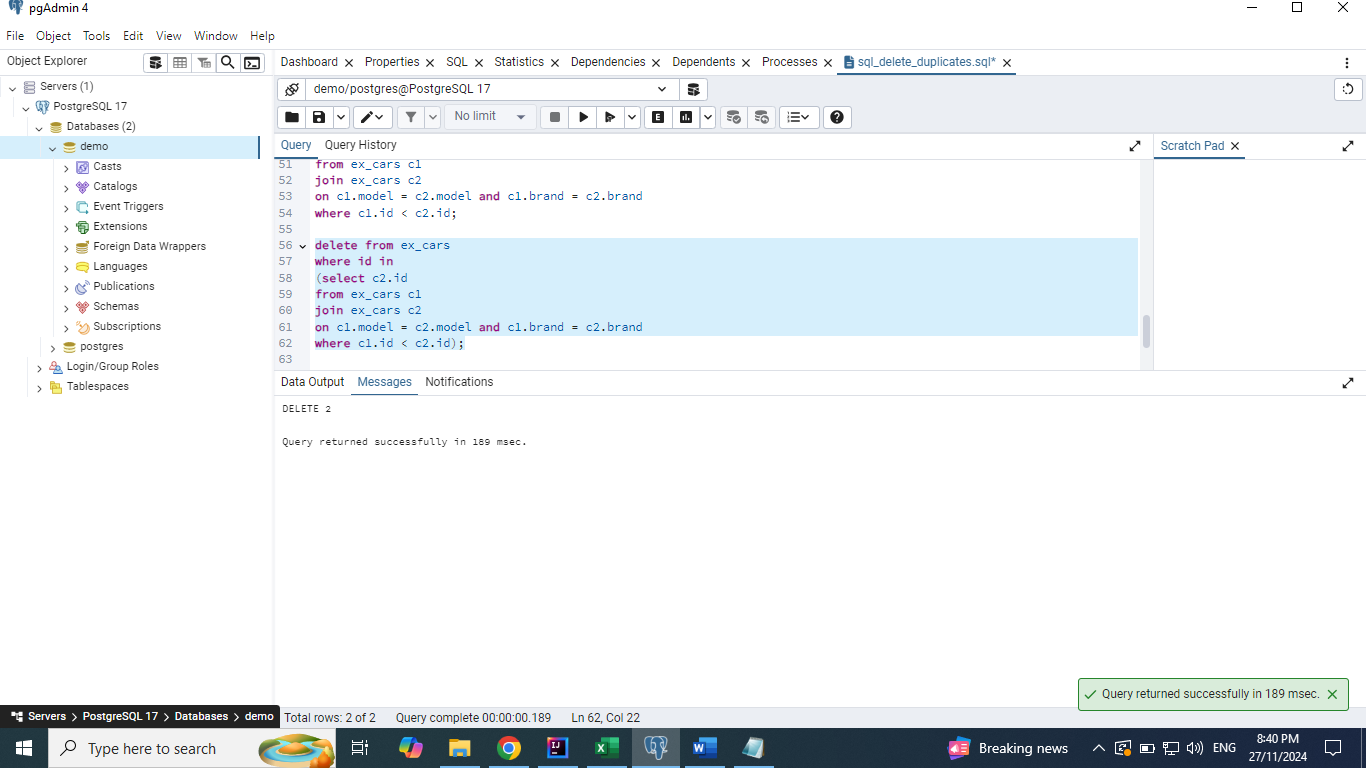
**(select c2.id**

**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**

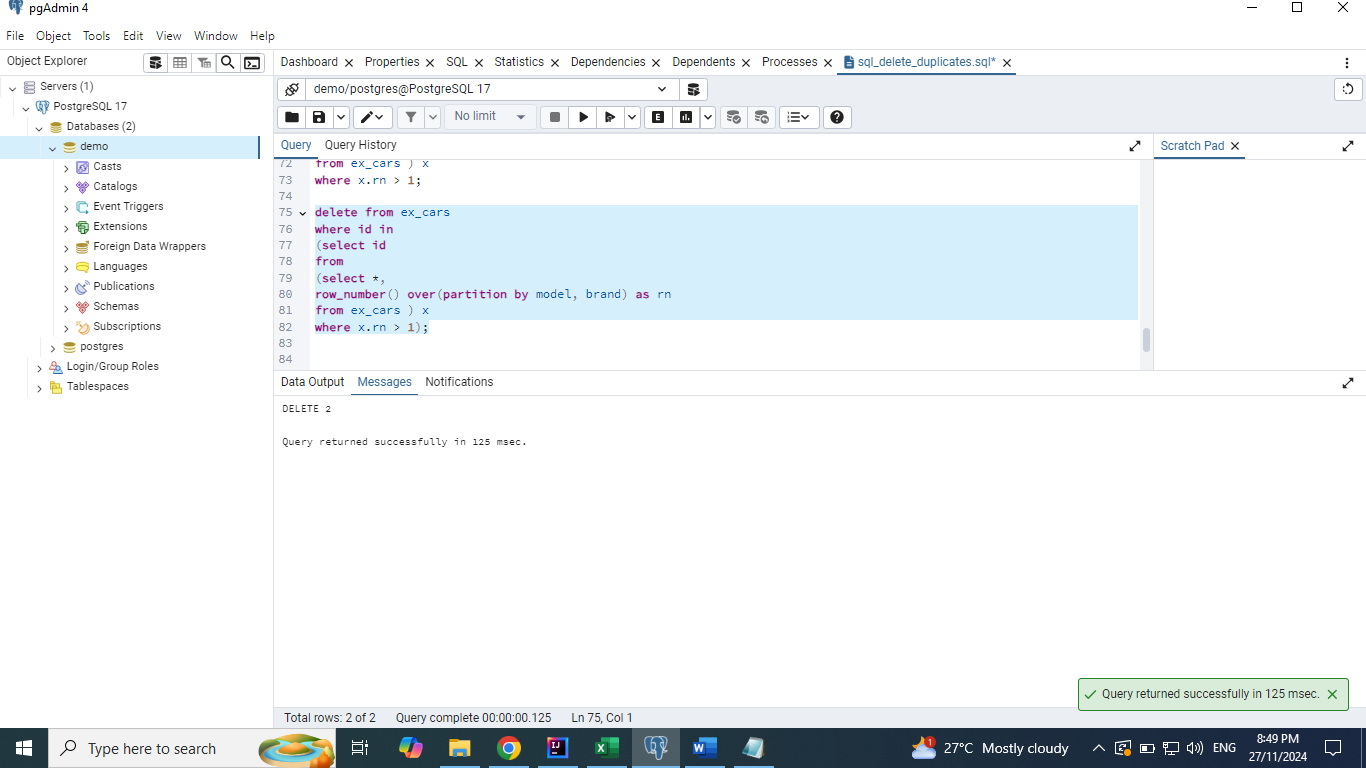
**from**

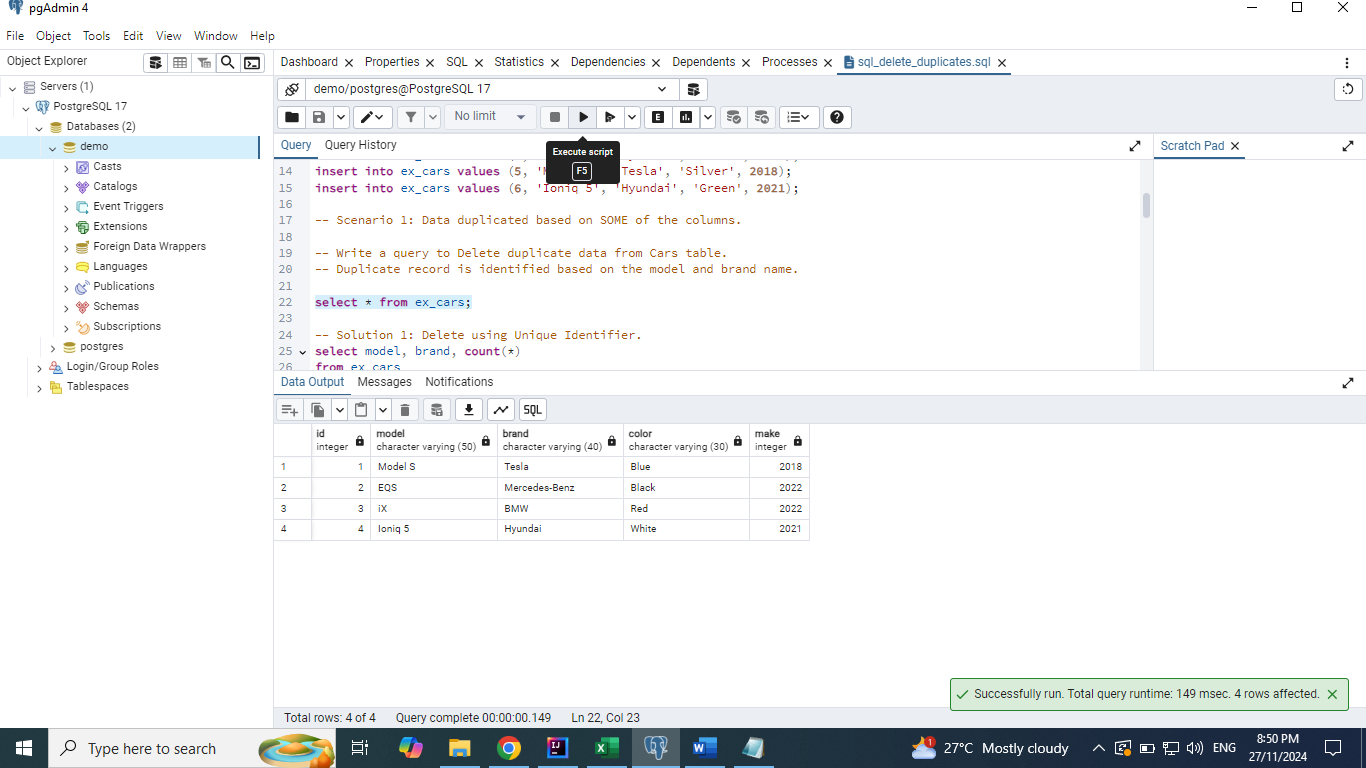
**(select \*,**

**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;**

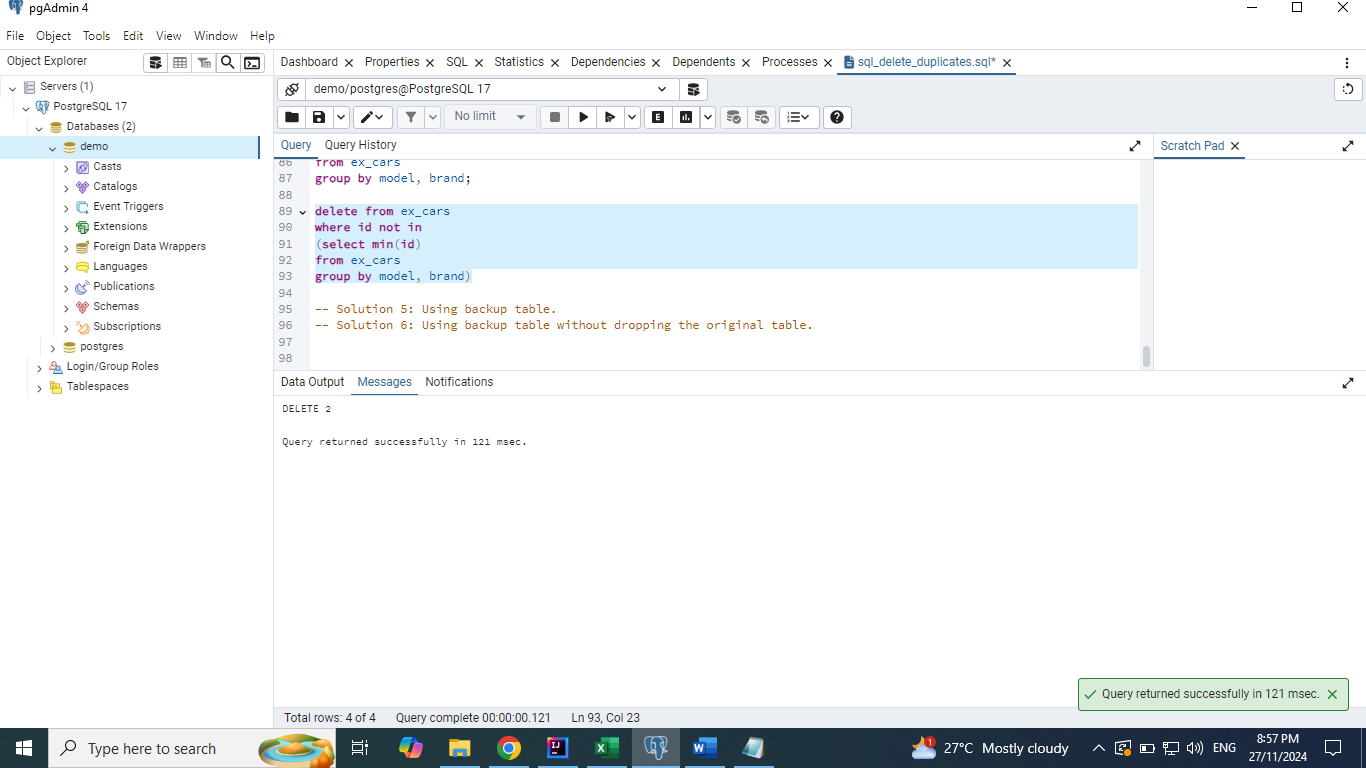
**delete from ex\_cars**

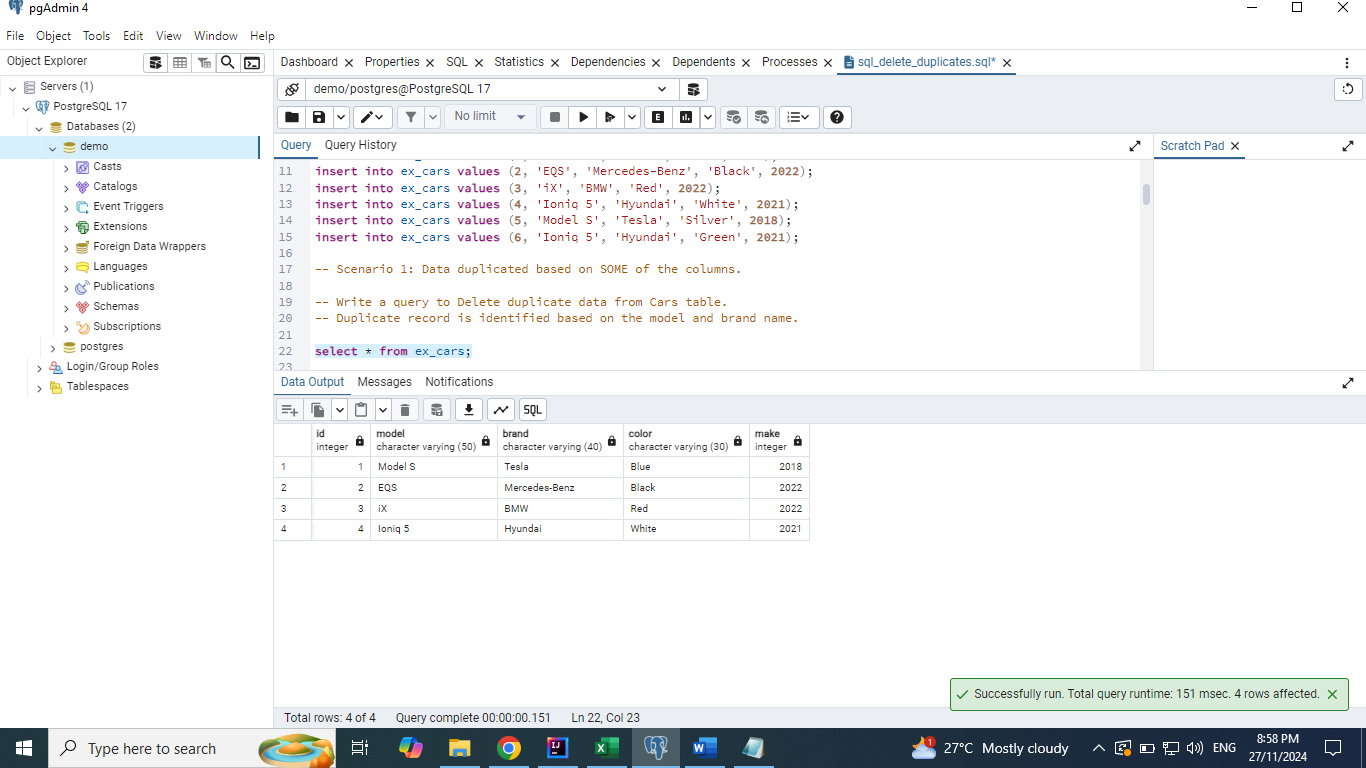
**where id not in**

**(select min(id)**

**from ex\_cars**

**group by model, brand);**

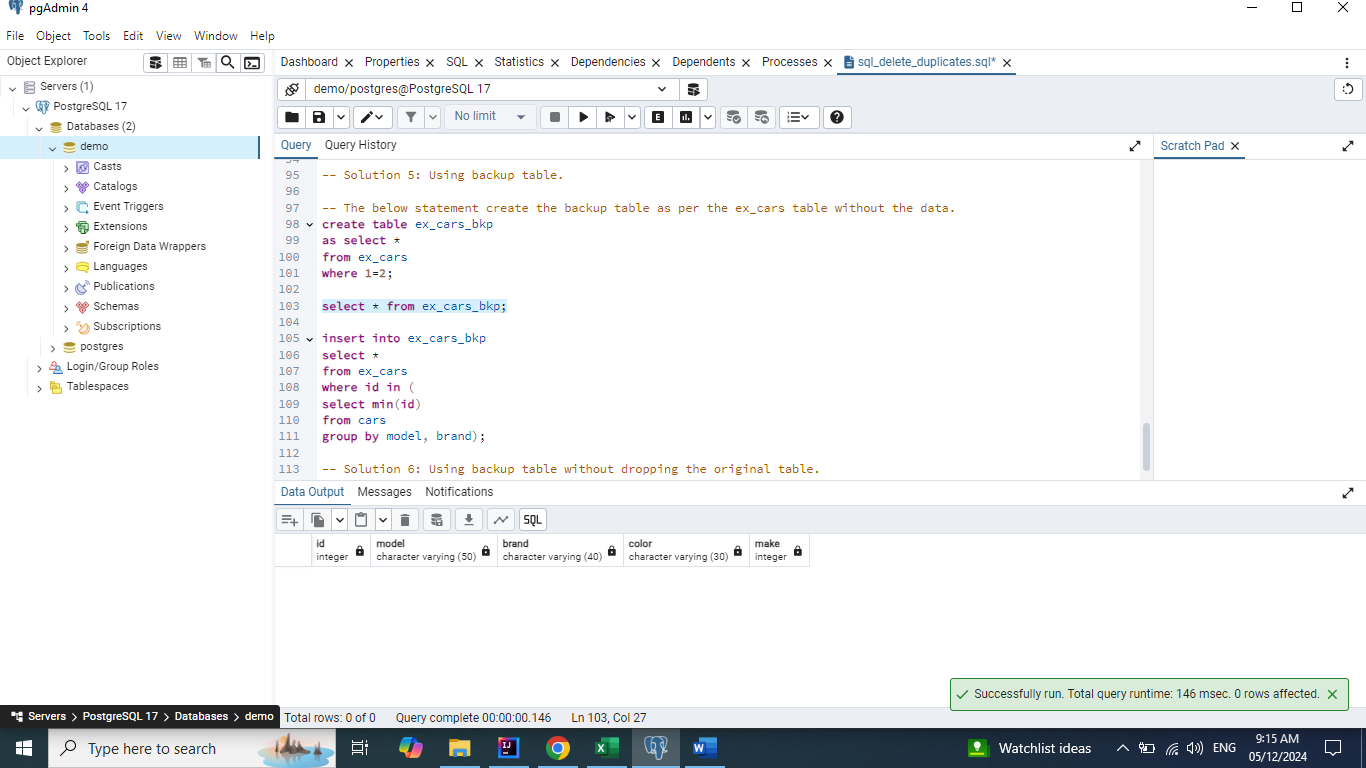




## Solution 5: Using backup table.

**The below statement creates the backup table as per the ex\_cars table without the data.**

**select \* from ex\_cars\_bkp;**



**insert into ex\_cars\_bkp**

**select \***

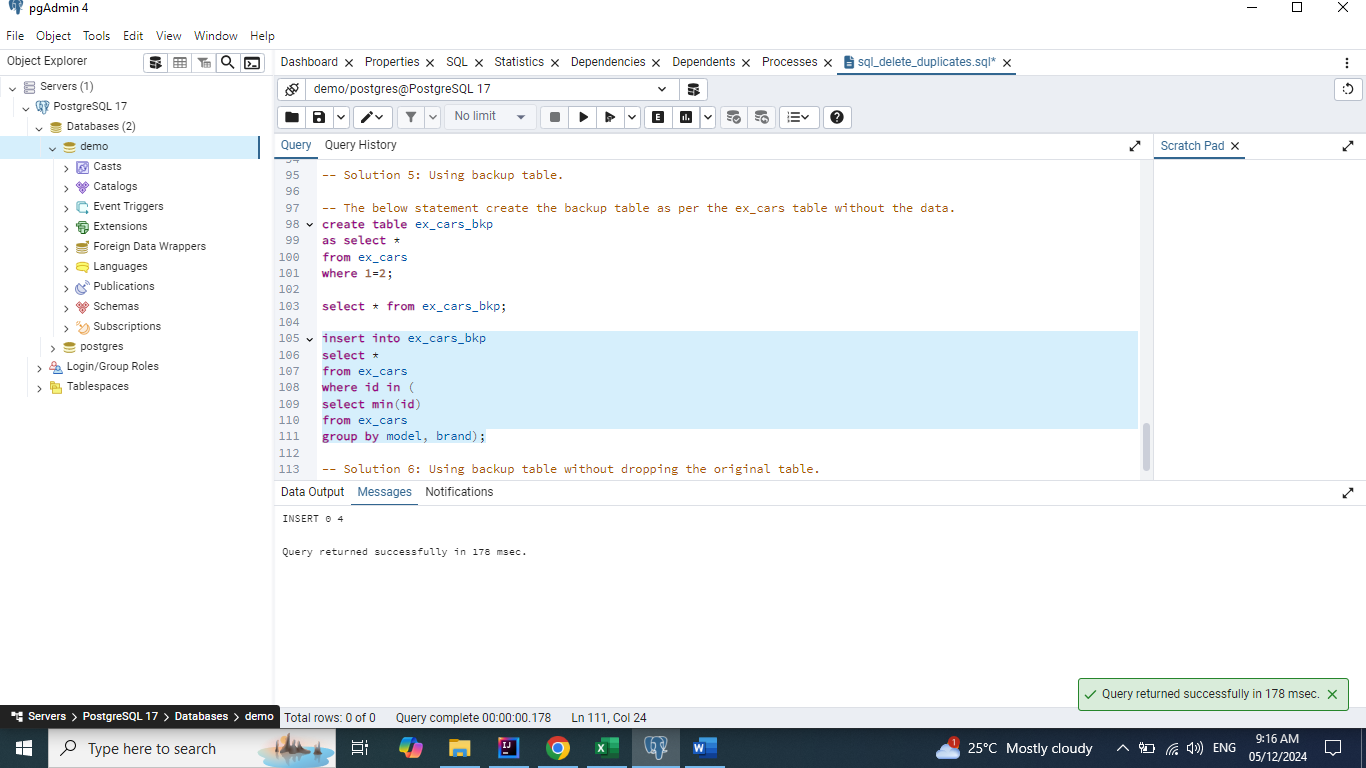
**from ex\_cars**

**where id in (**

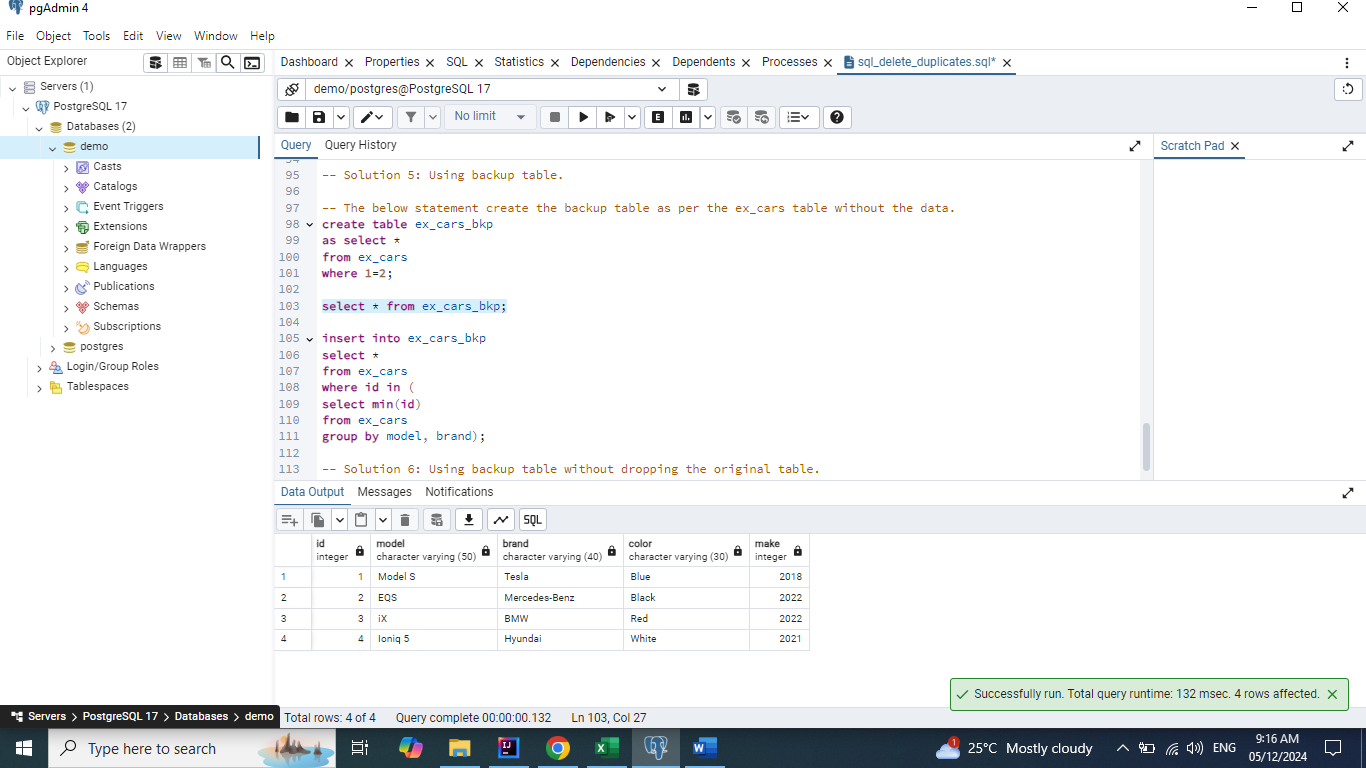
**select min(id)**

**from ex\_cars**

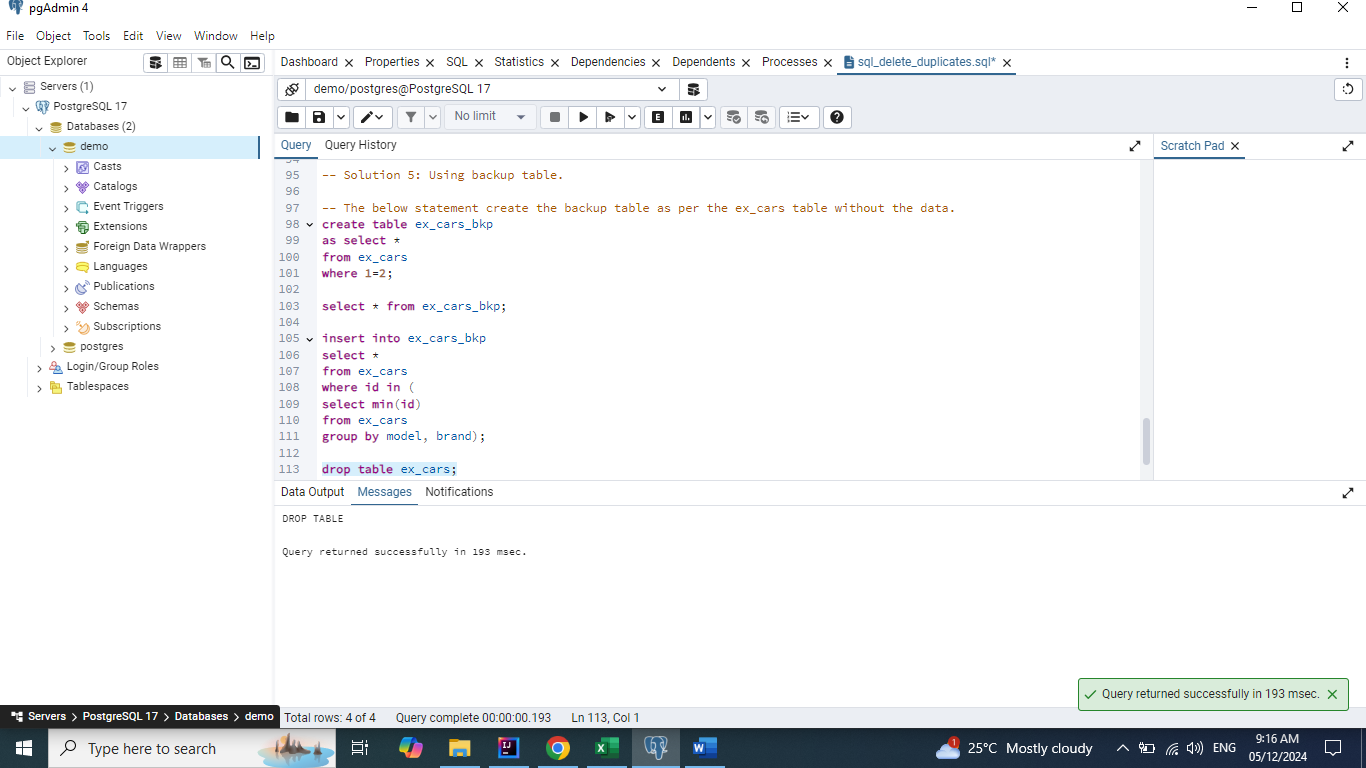
**group by model, brand);**



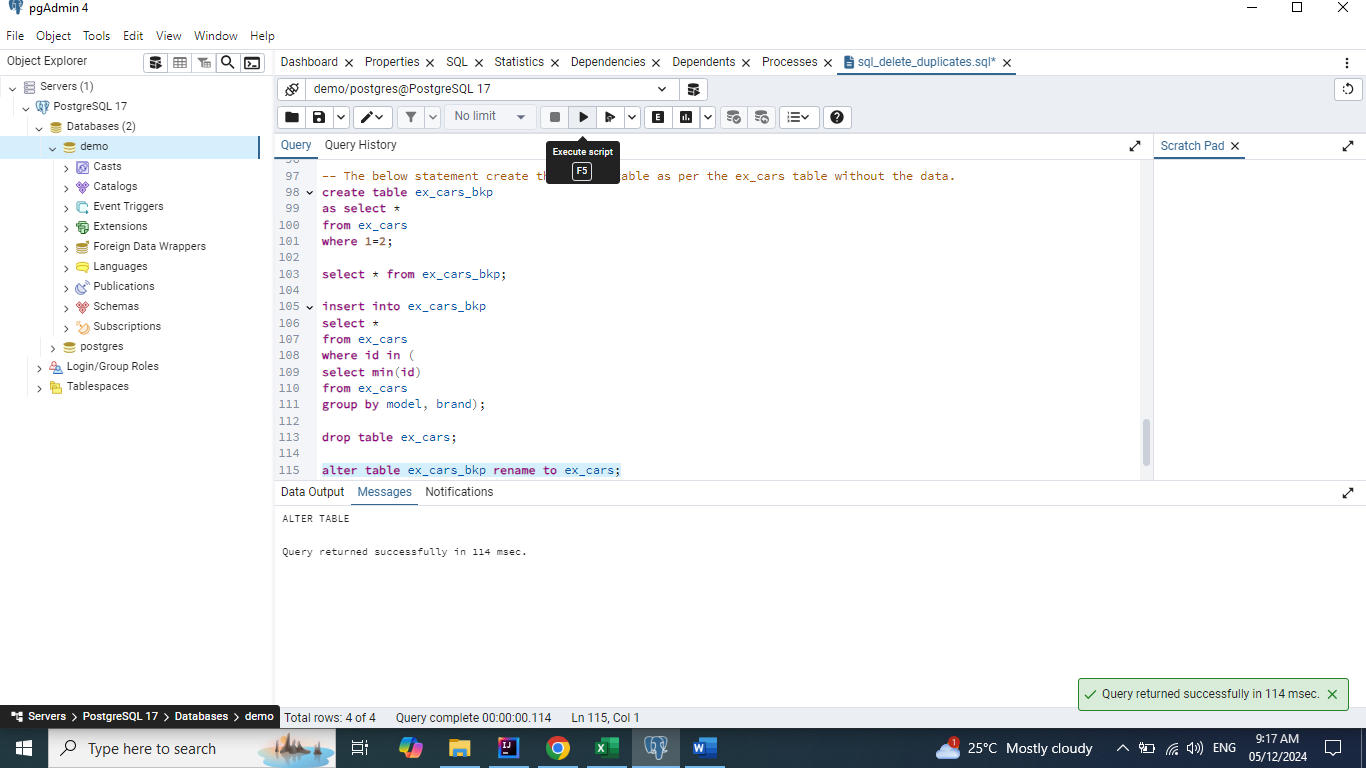
**select \* from ex\_cars\_bkp;**



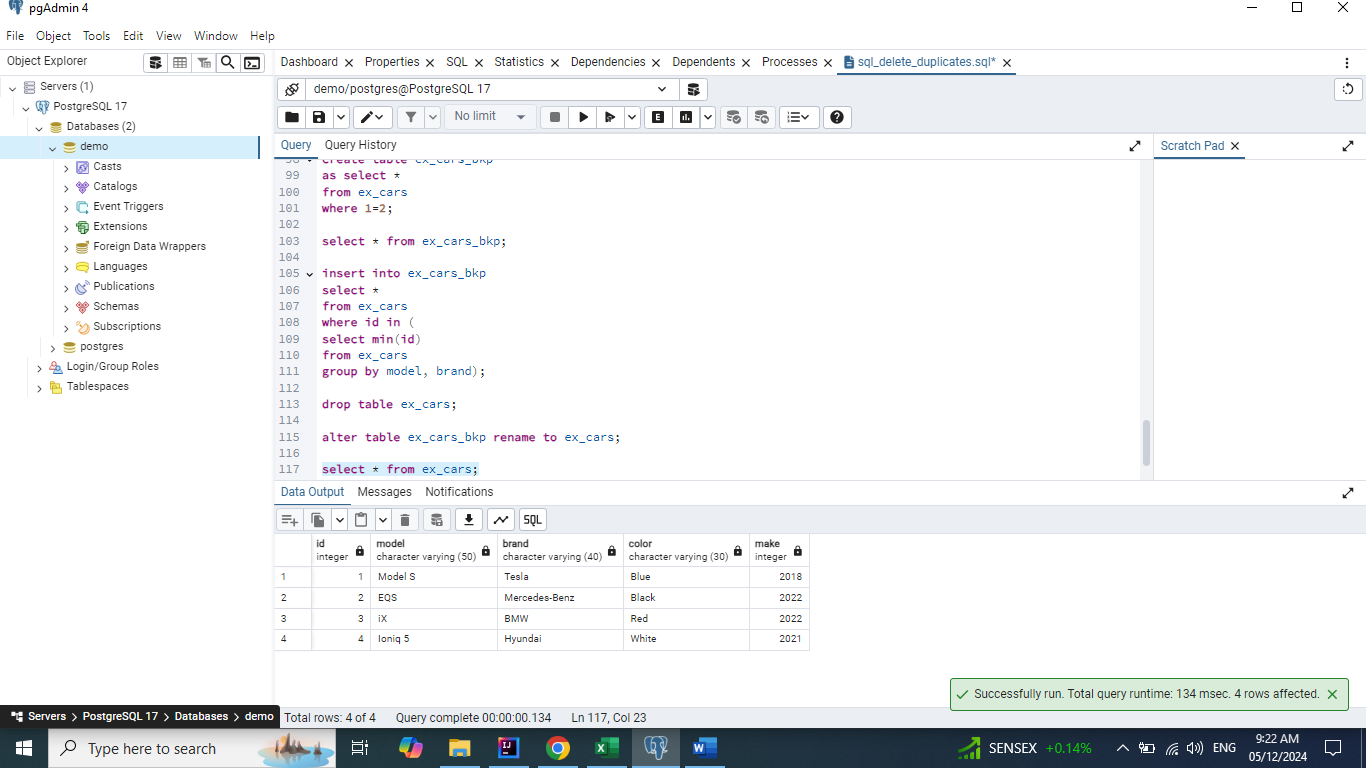
**drop table ex\_cars;**



**alter table ex\_cars\_bkp rename to ex\_cars;**



**select \* from ex\_cars;**



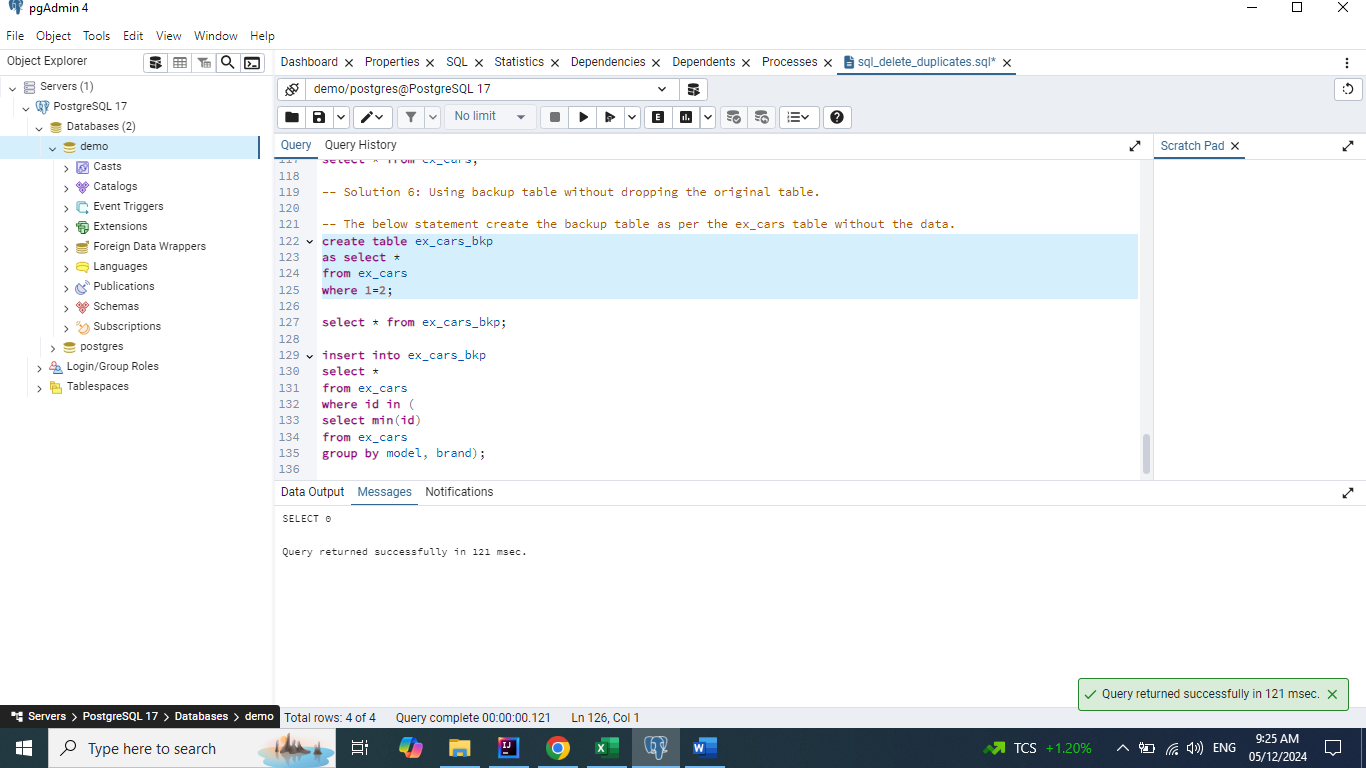
## Solution 6: Using backup table without dropping the original table.

**create table ex\_cars\_bkp**

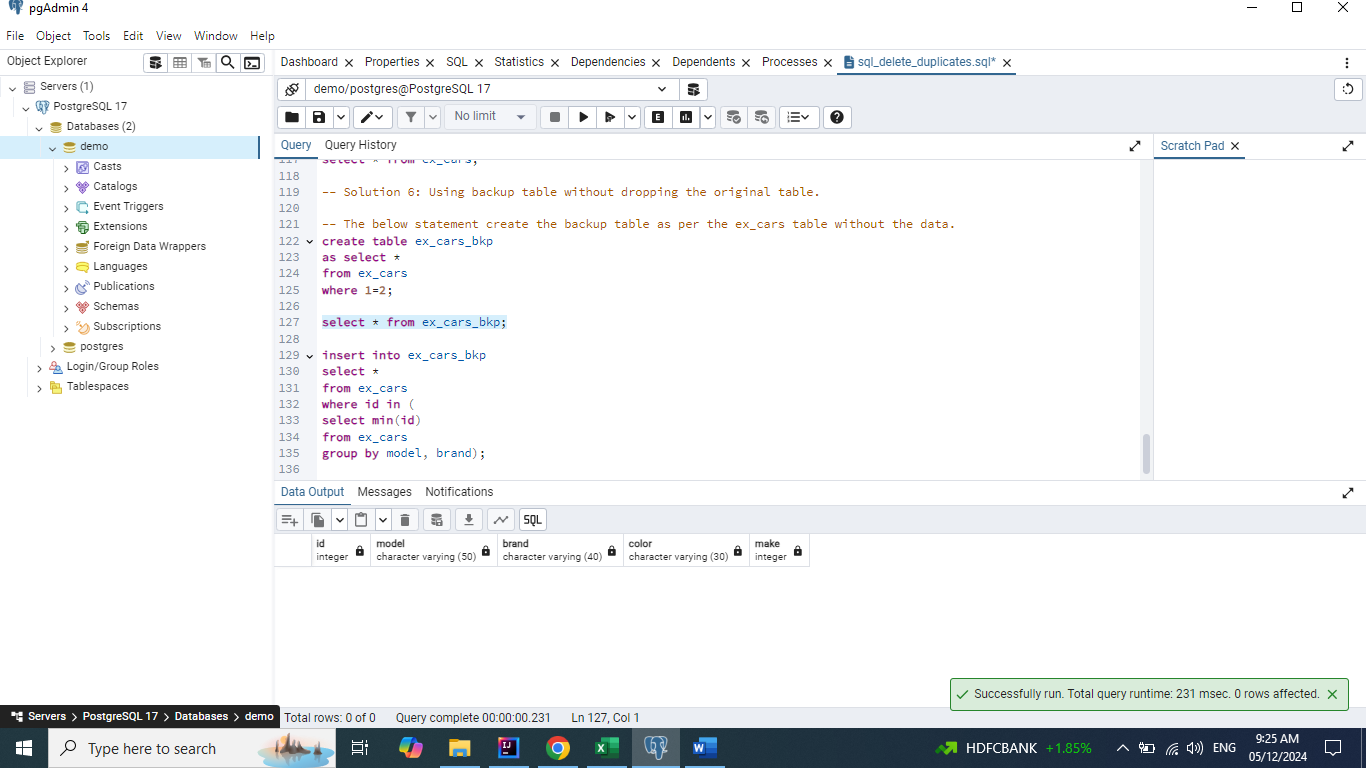
**as select \***

**from ex\_cars**

**where 1=2;**



**select \* from ex\_cars\_bkp;**



**insert into ex\_cars\_bkp**

**select \***

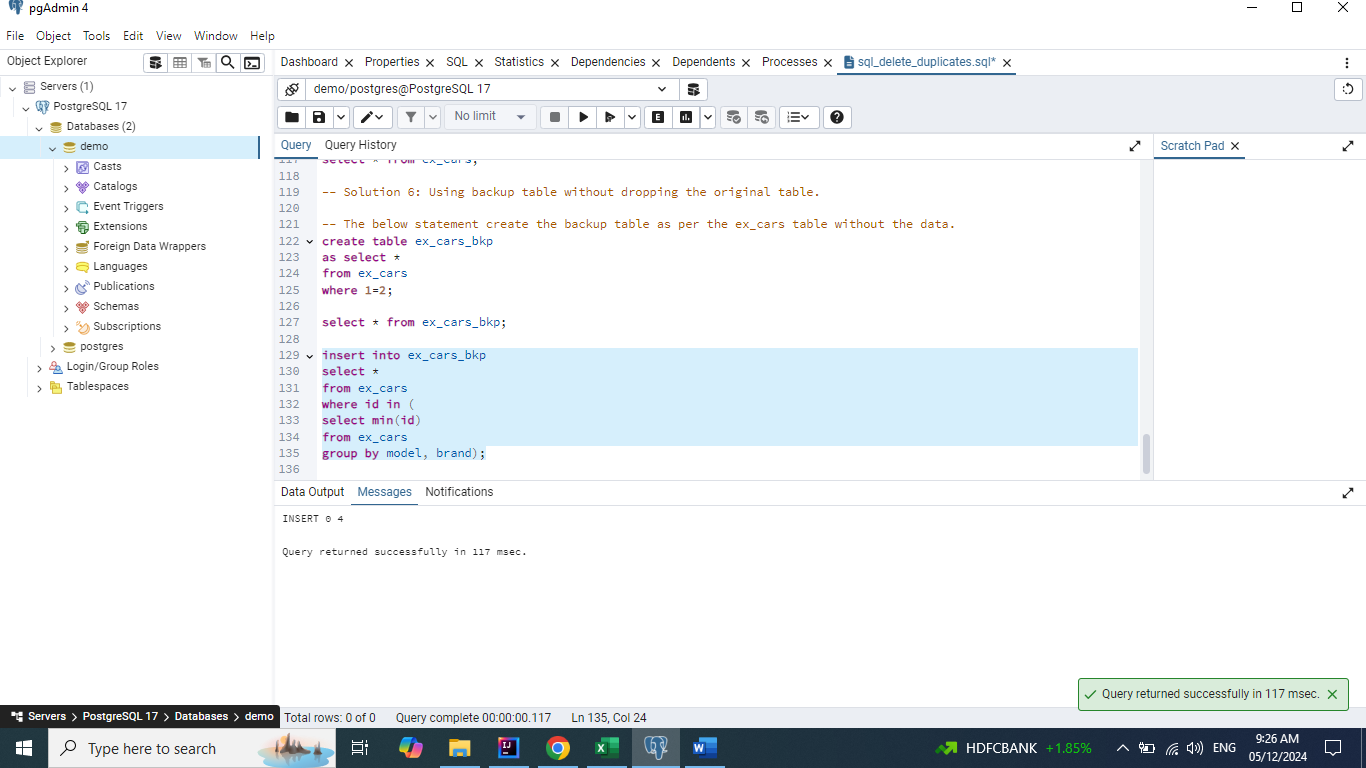
**from ex\_cars**

**where id in (**

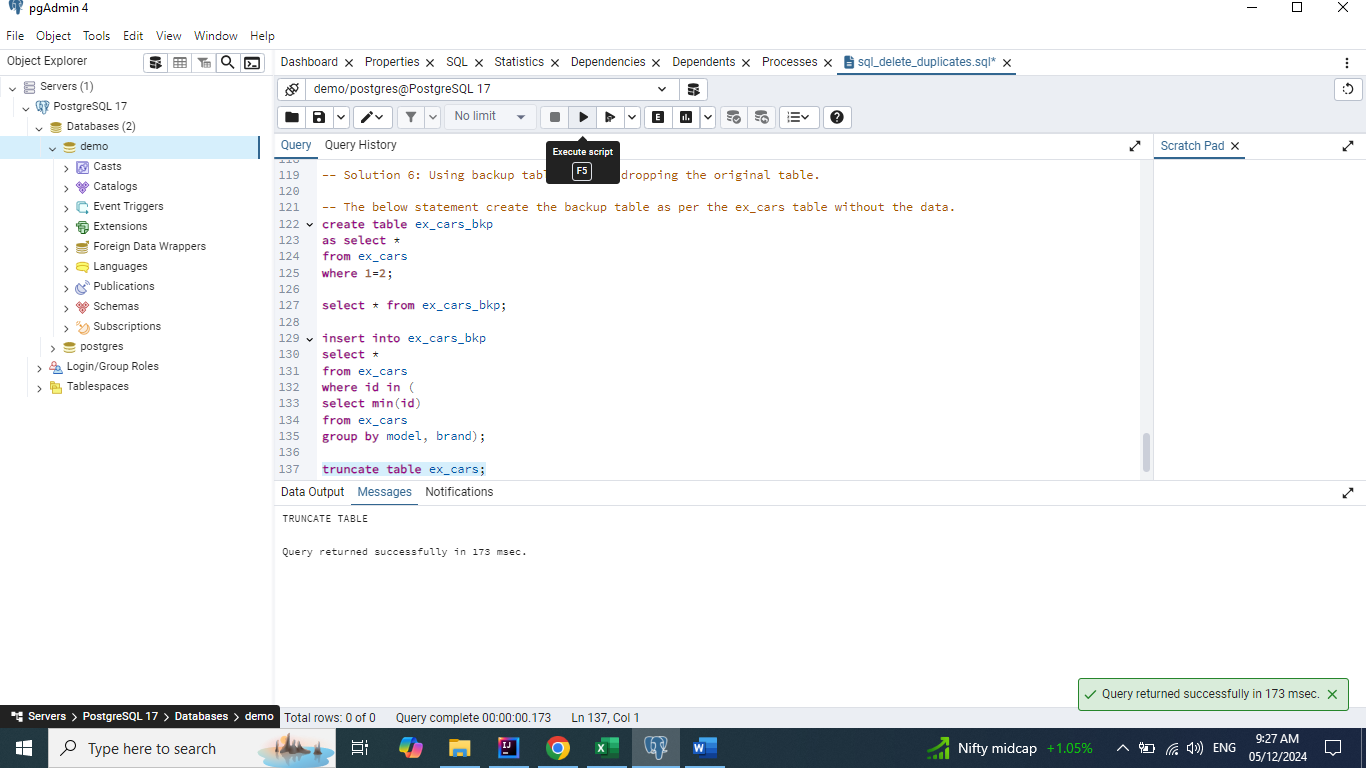
**select min(id)**

**from ex\_cars**

**group by model, brand);**

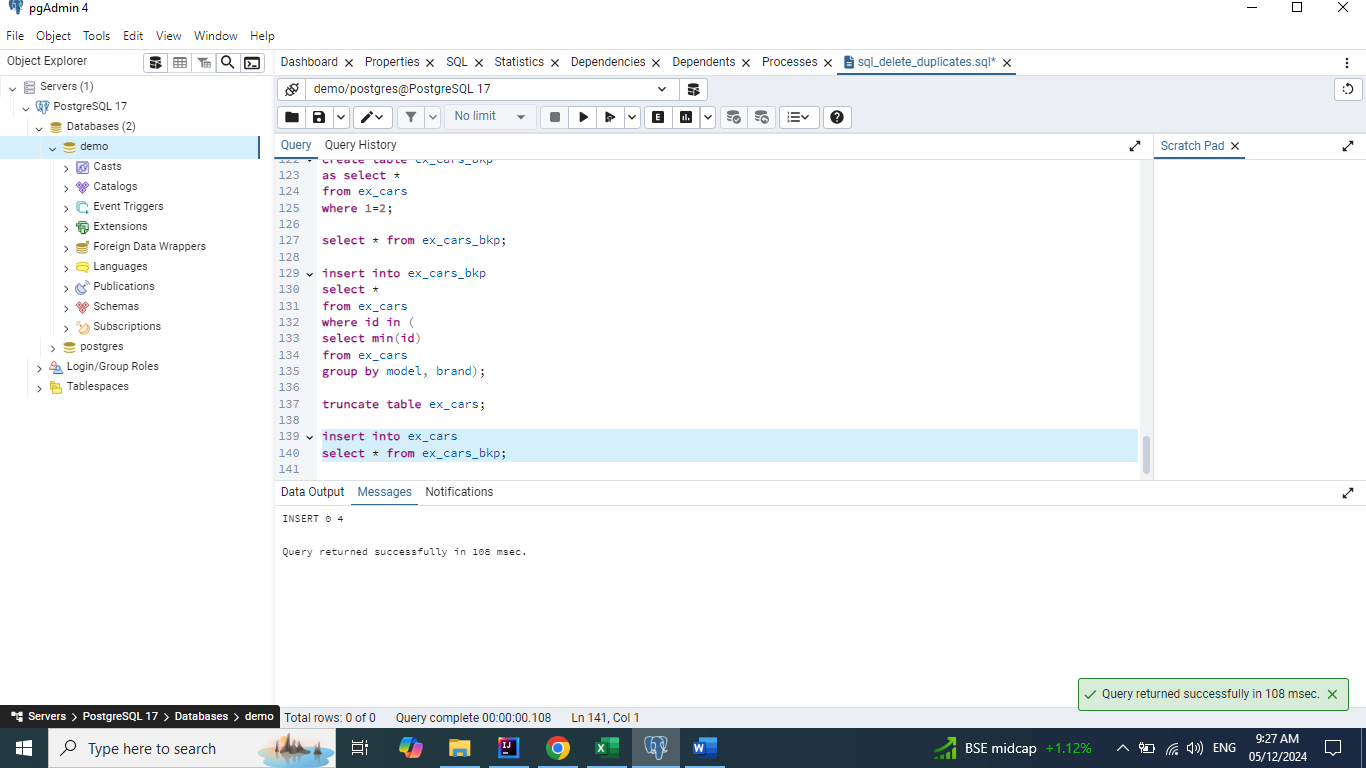


**truncate table ex\_cars;**

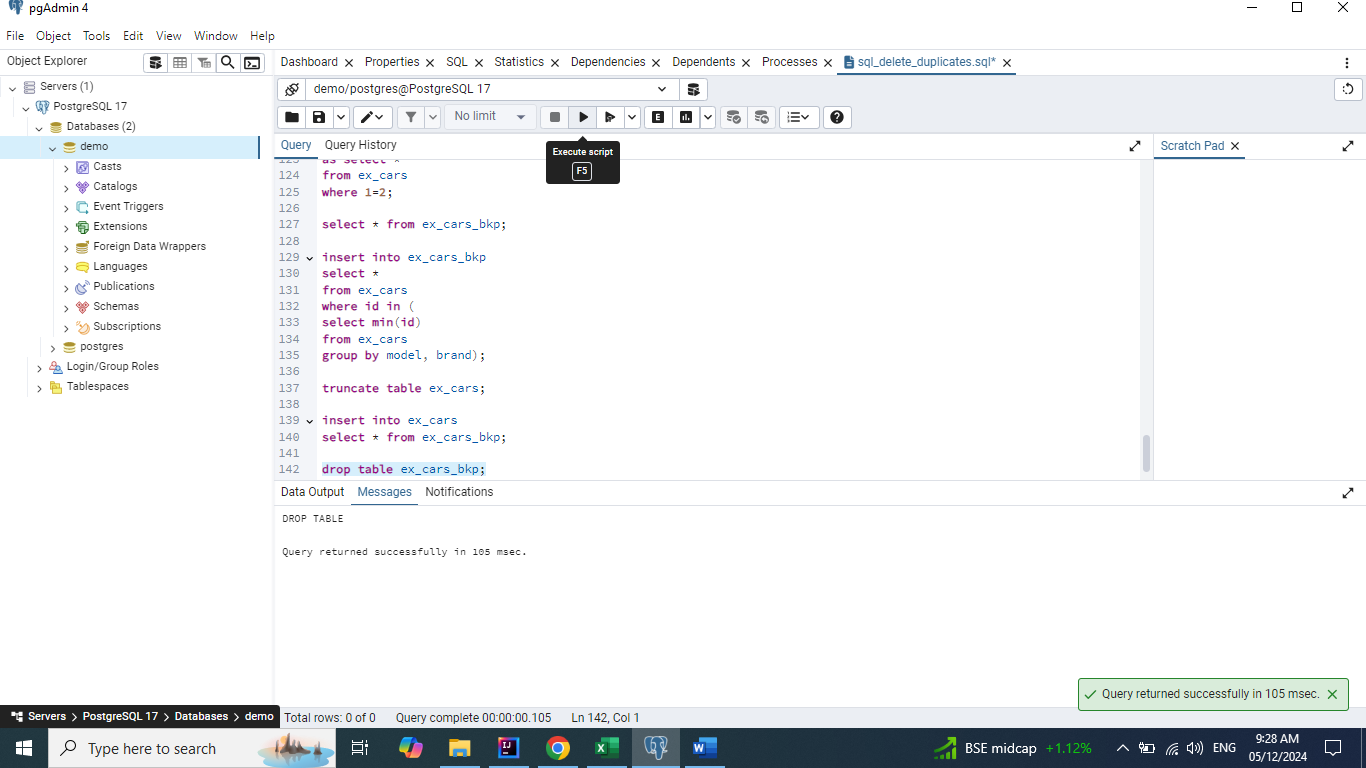


**insert into ex\_cars**

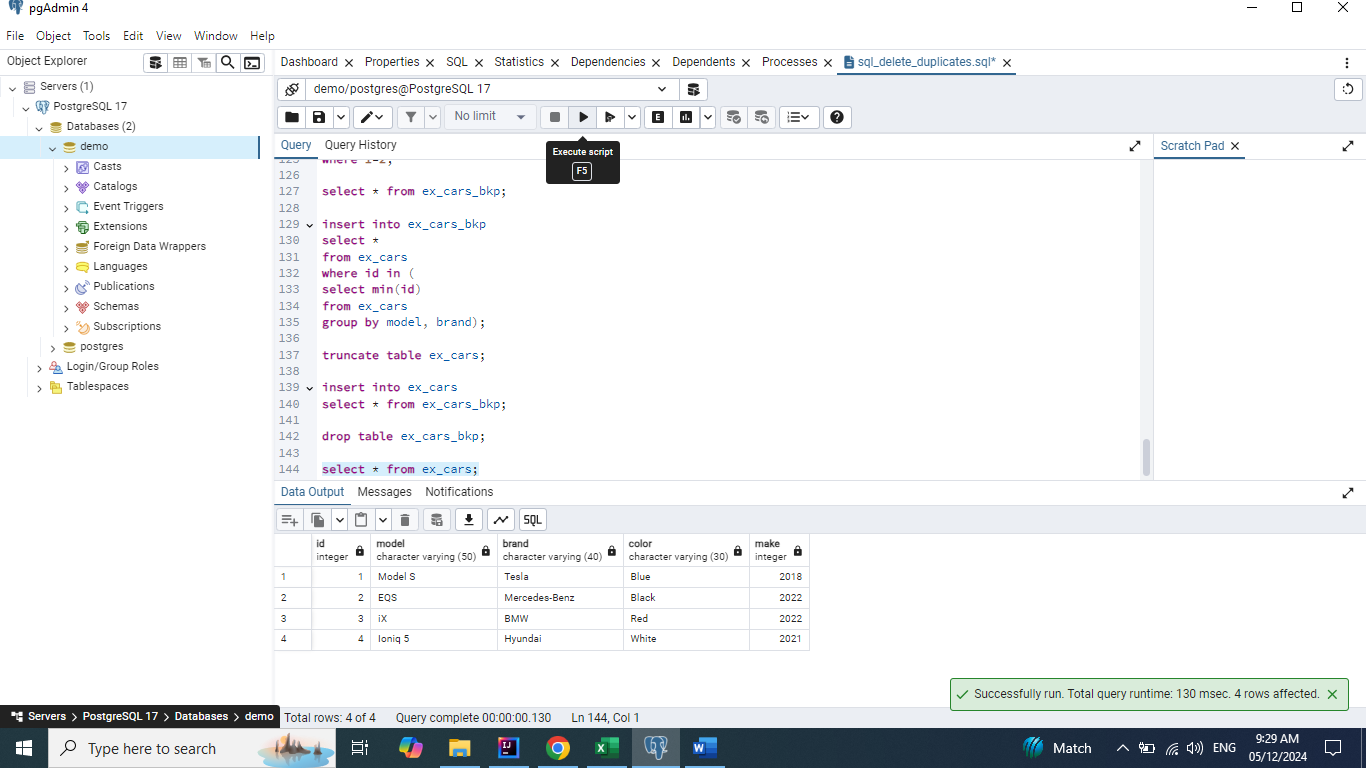
**select \* from ex\_cars\_bkp;**



**drop table ex\_cars\_bkp;**



**select \* from ex\_cars;**

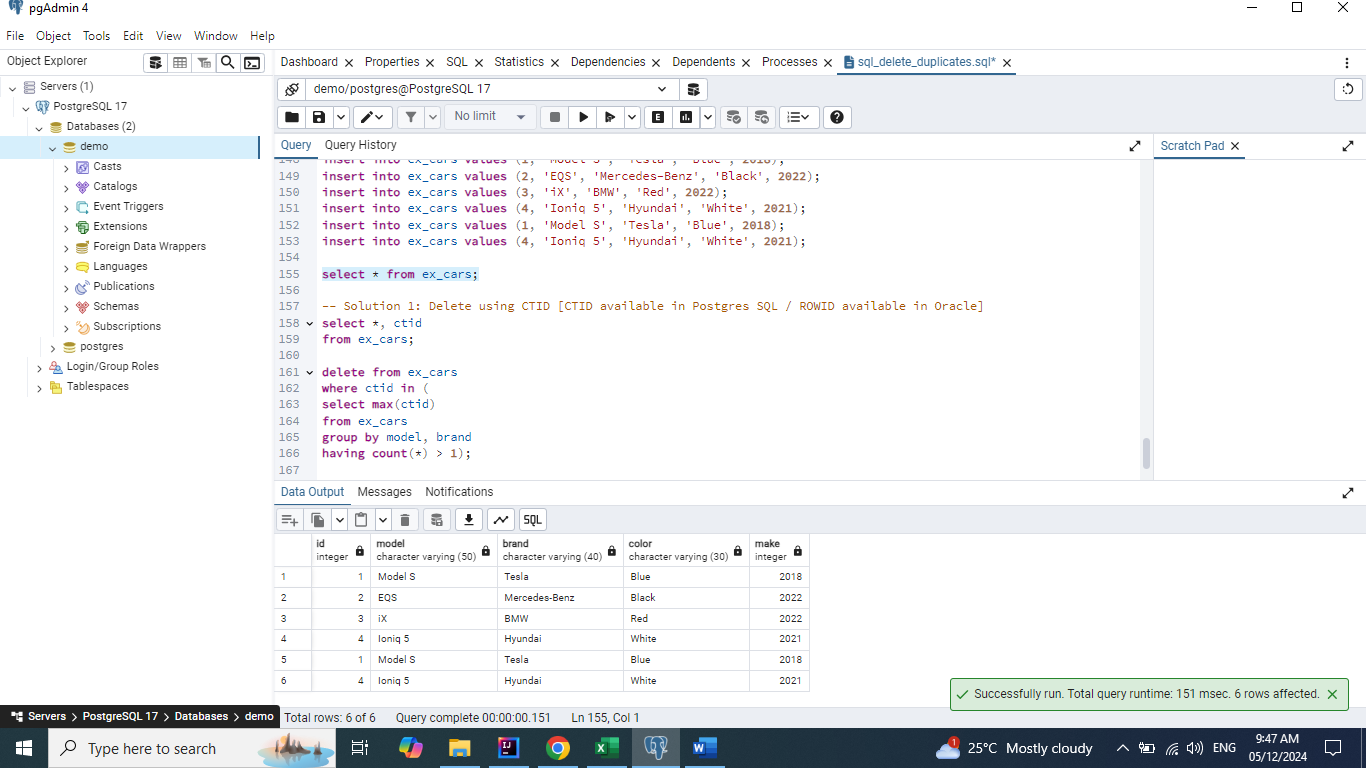


# Delete the records even if the Id is duplicate.

## Solution 1: Delete using CTID [CTID available in Postgres SQL / ROWID available in Oracle]

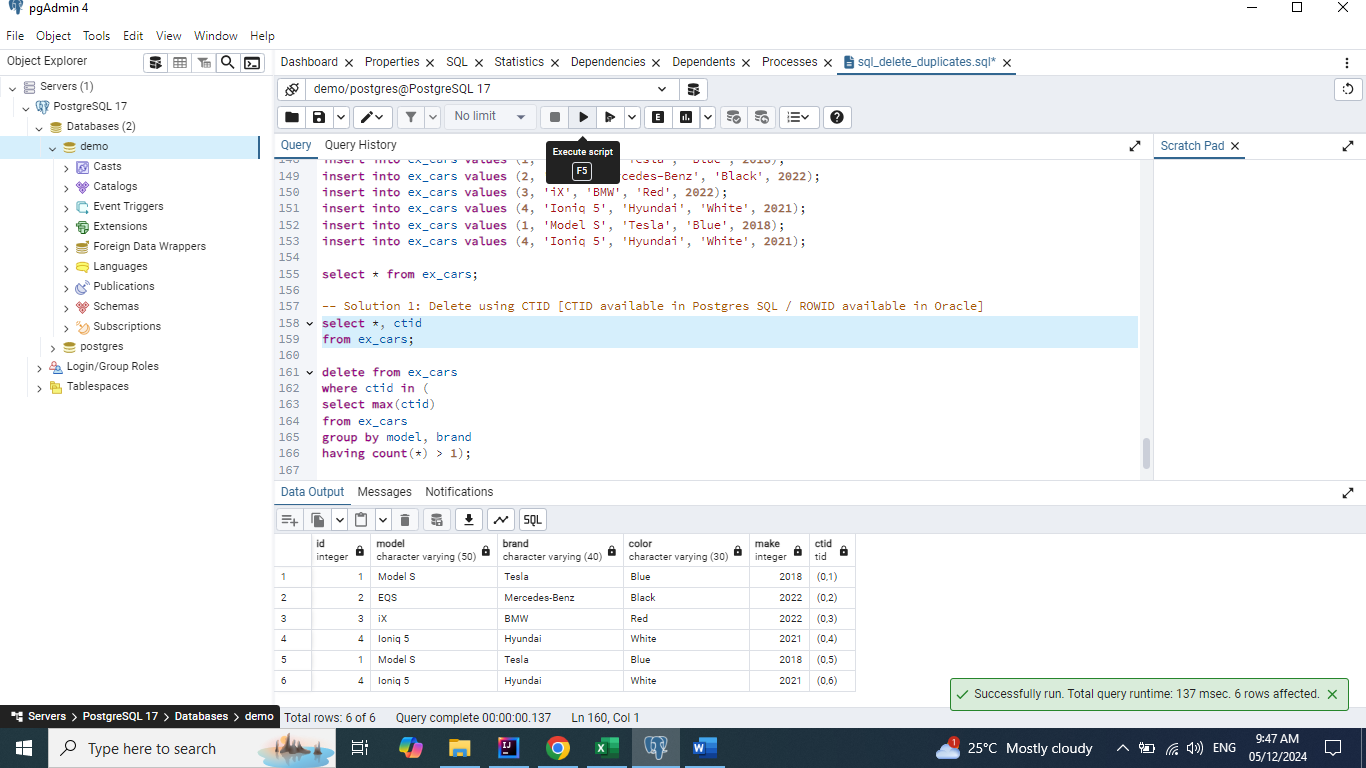
**The above solution will not work in MySQL and Microsoft SQL Server.**

**select \* from ex\_cars;**



**select \*, ctid**

**from ex\_cars;**



**delete from ex\_cars**

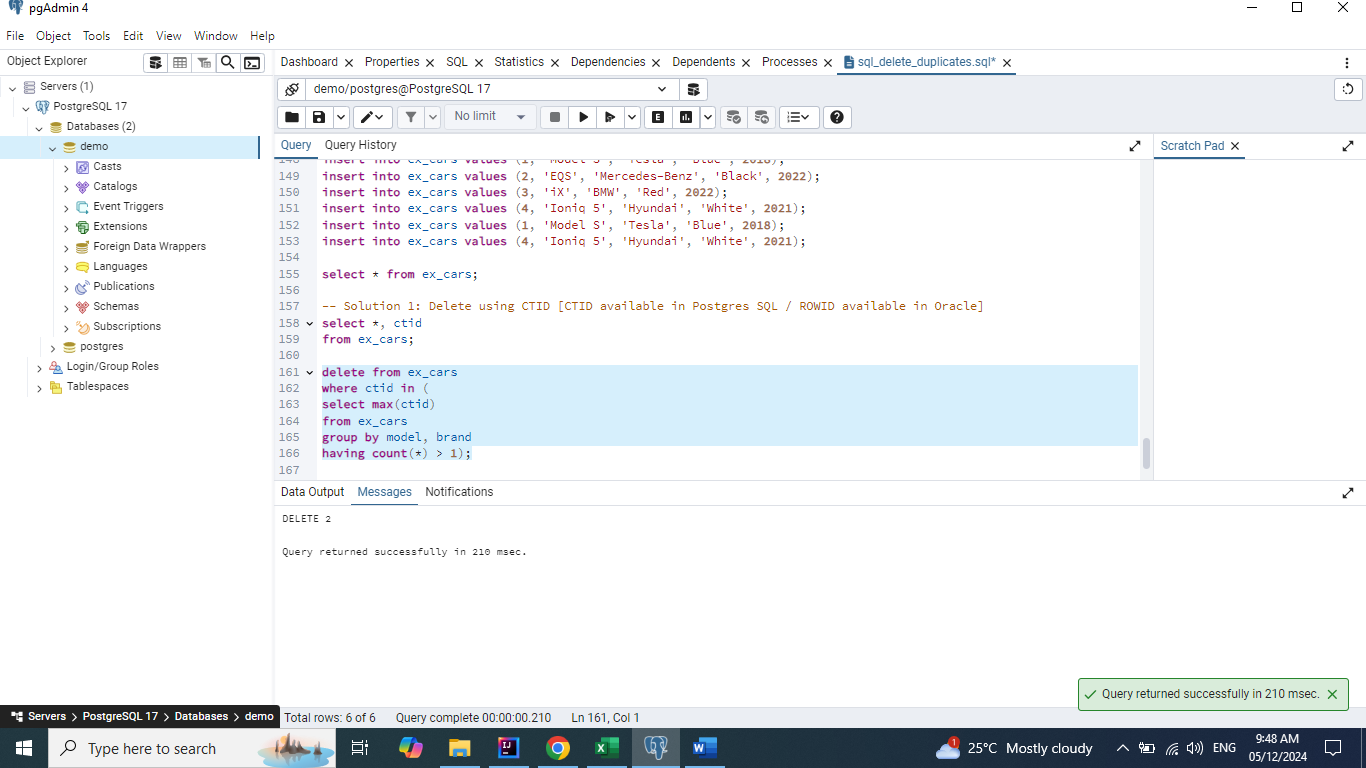
**where ctid in (**

**select max(ctid)**

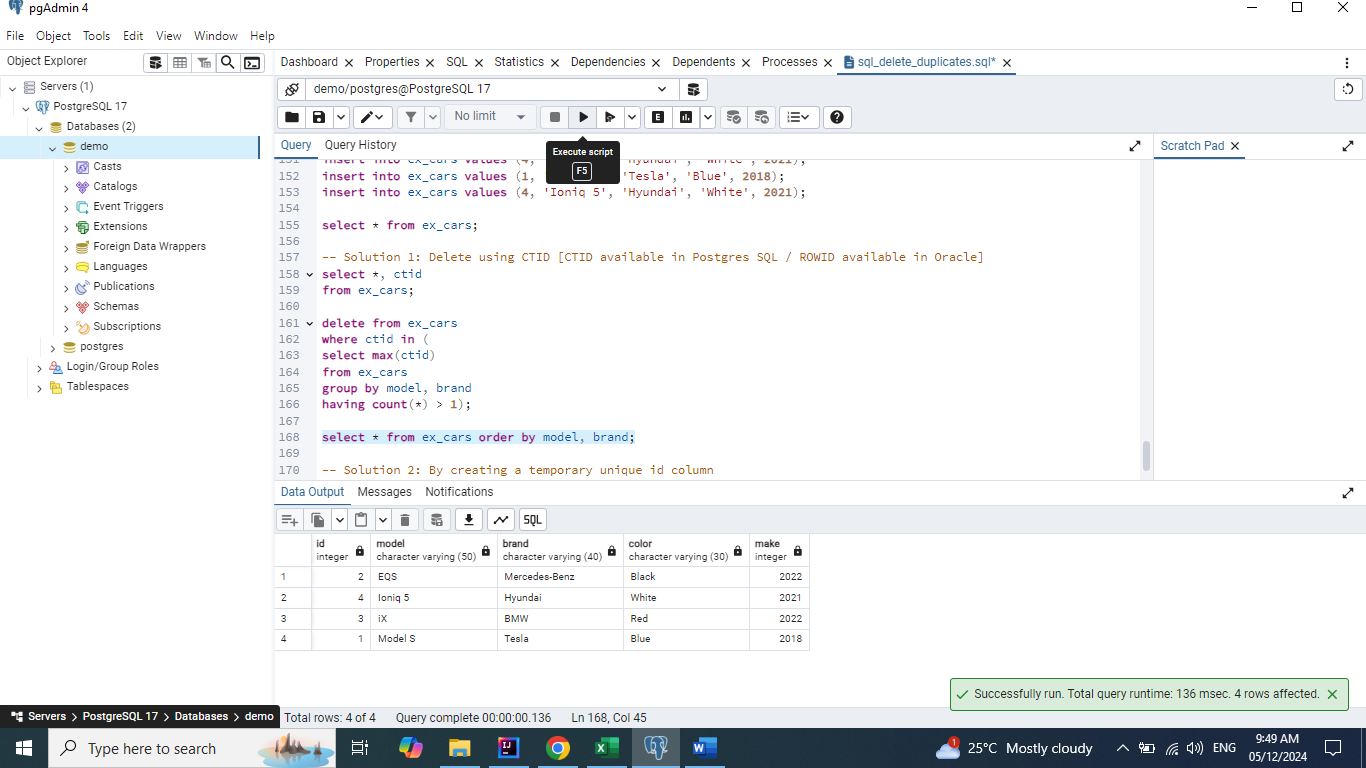
**from ex\_cars**

**group by model, brand**

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

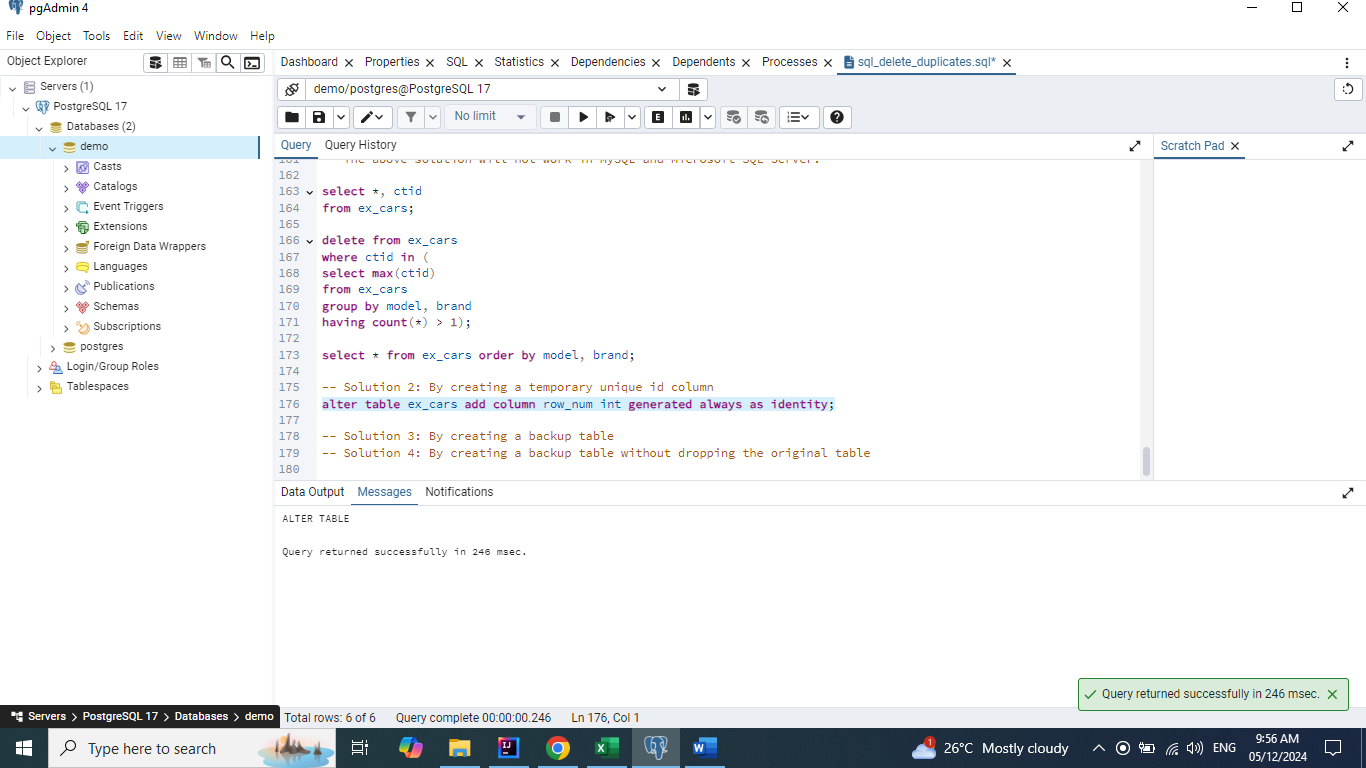


**select \* from ex\_cars order by model, brand;**

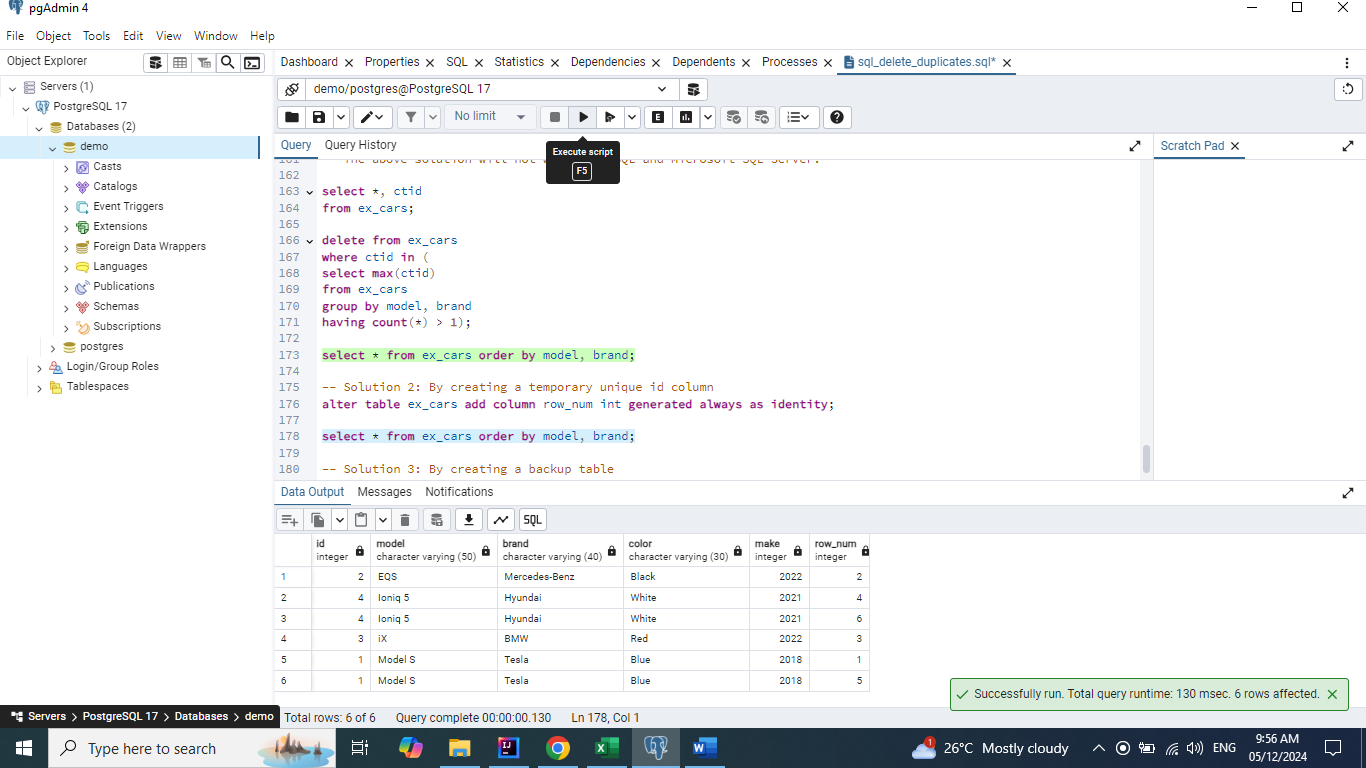


## Solution 2: By creating a temporary unique id column

**alter table ex\_cars add column row\_num int generated always as identity;**



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

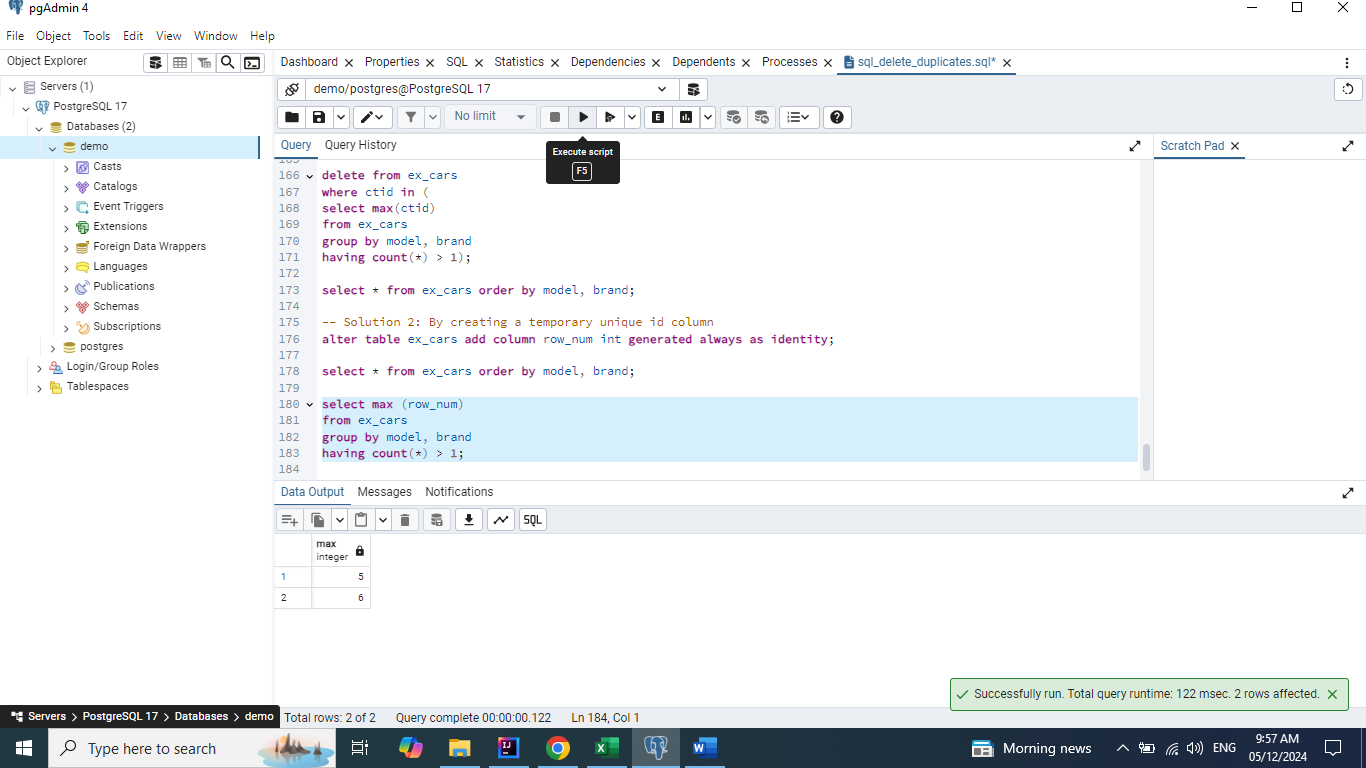


**select max (row\_num)**

**from ex\_cars**

**group by model, brand**

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



**delete from ex\_cars**

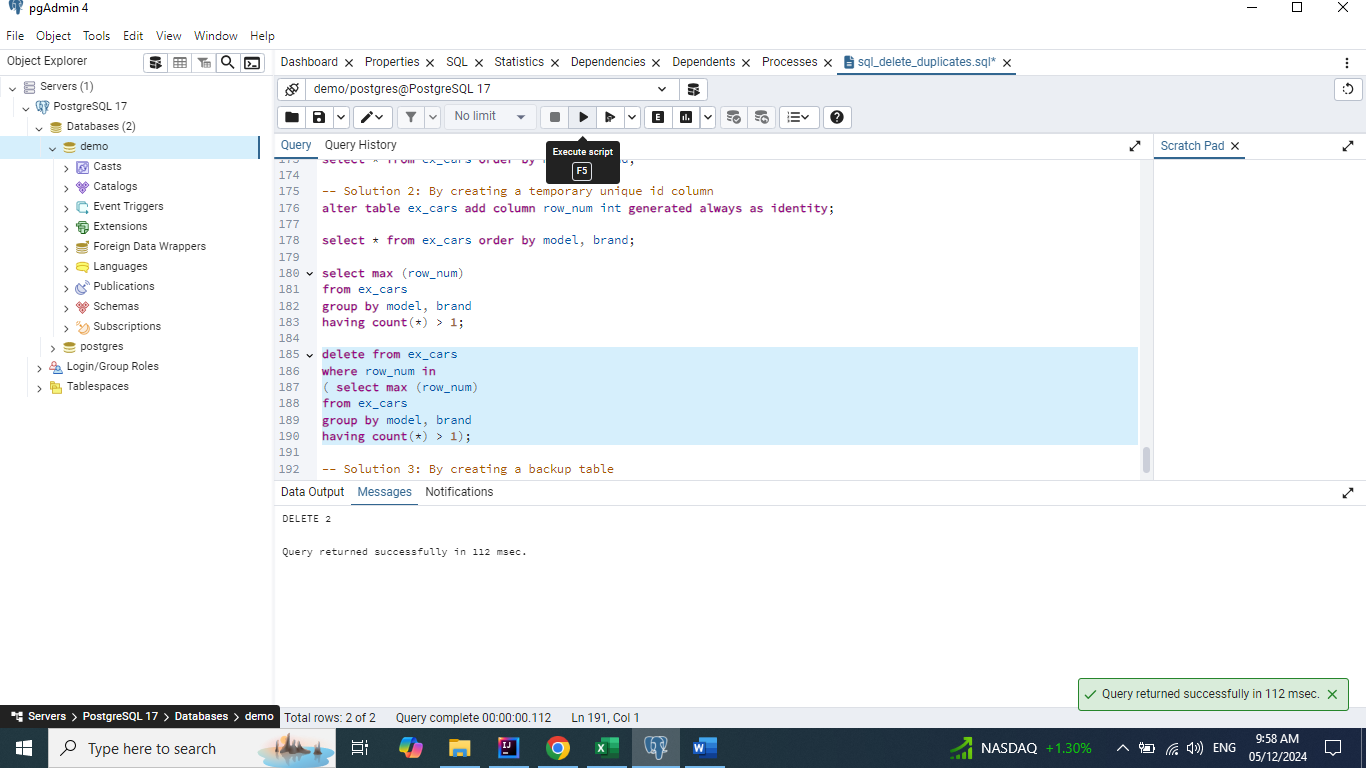
**where row\_num in**

**(select max (row\_num)**

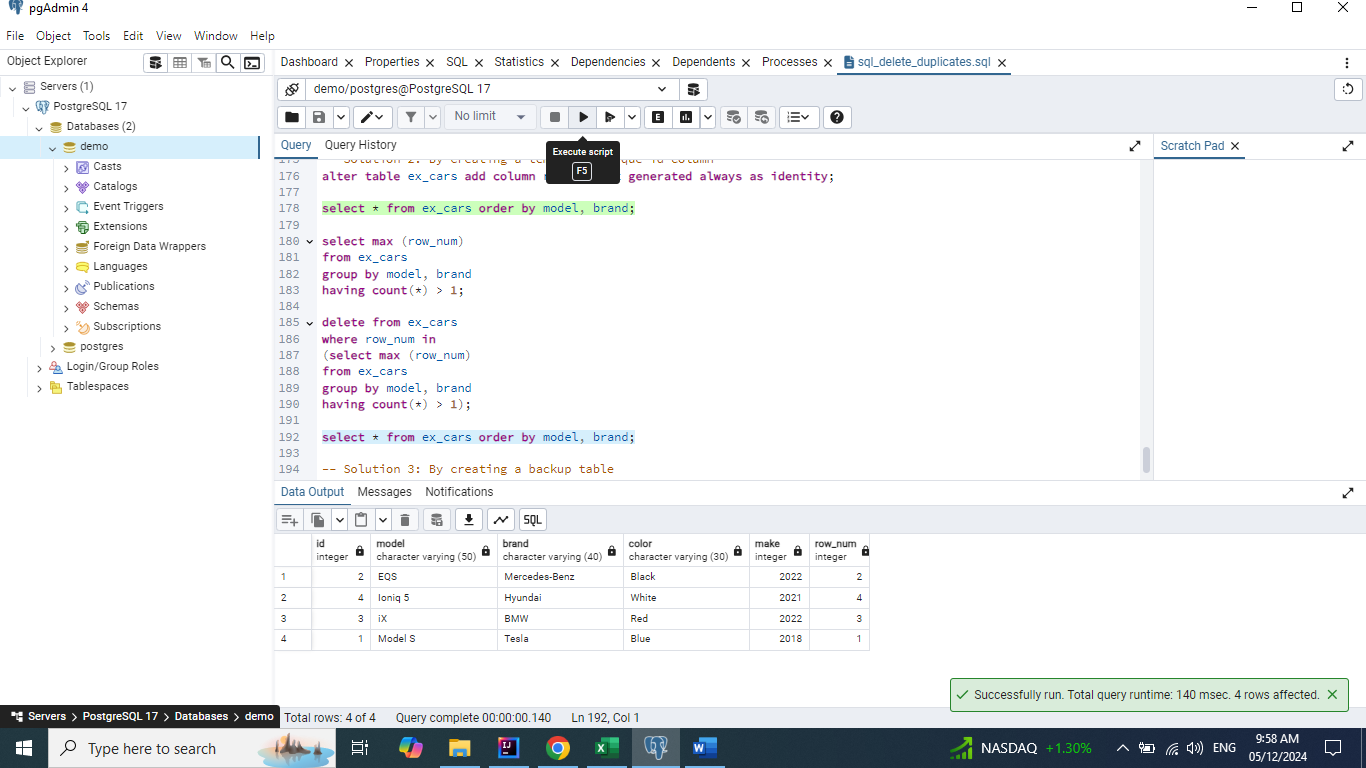
**from ex\_cars**

**group by model, brand**

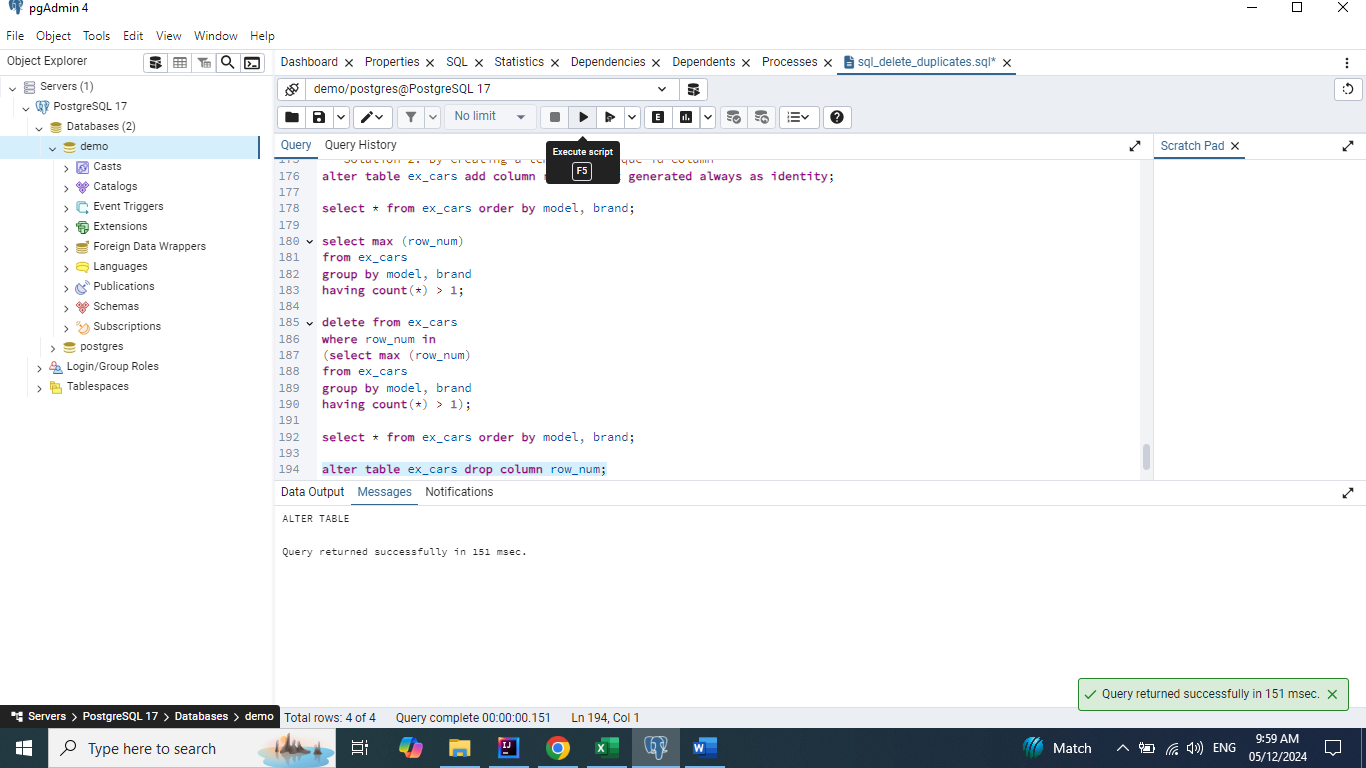
**having count(\*) > 1);**



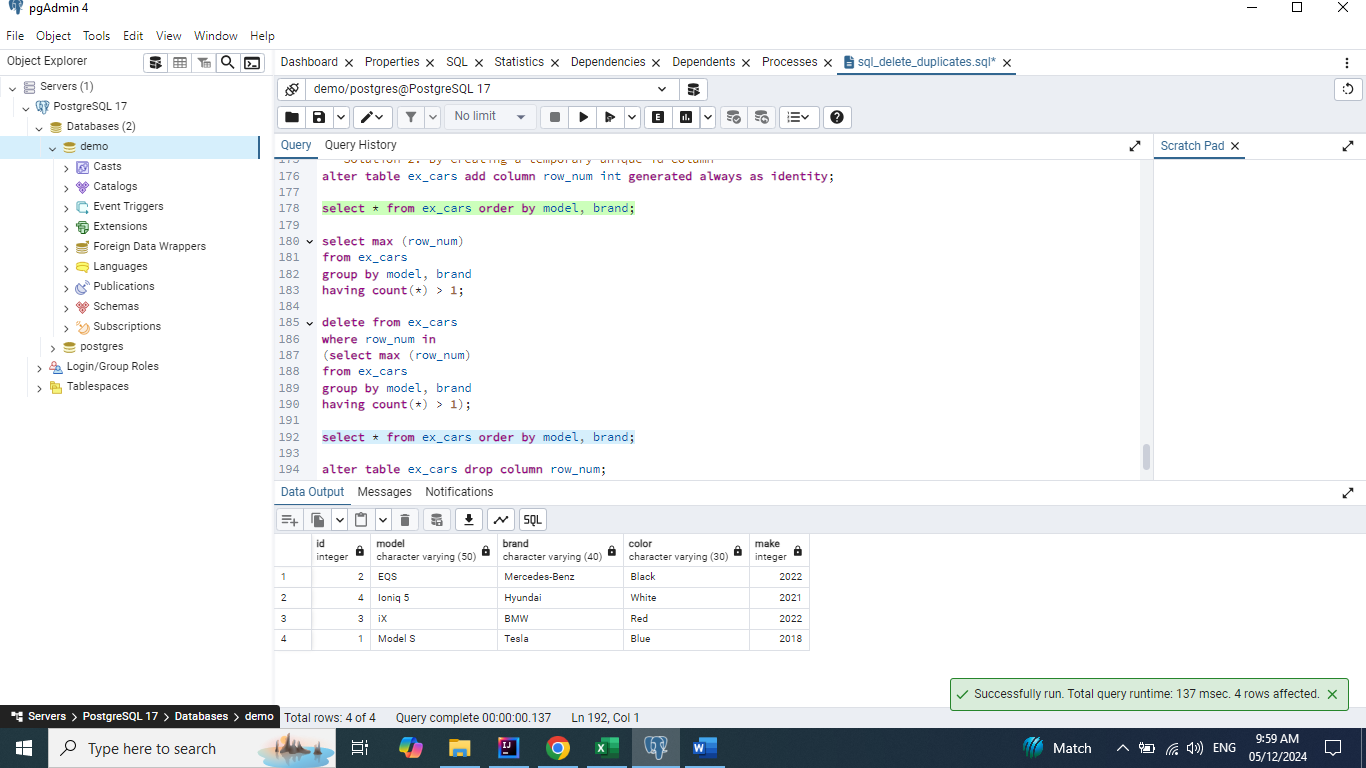
**select \* from ex\_cars order by model, brand;**



**alter table ex\_cars drop column row\_num;**



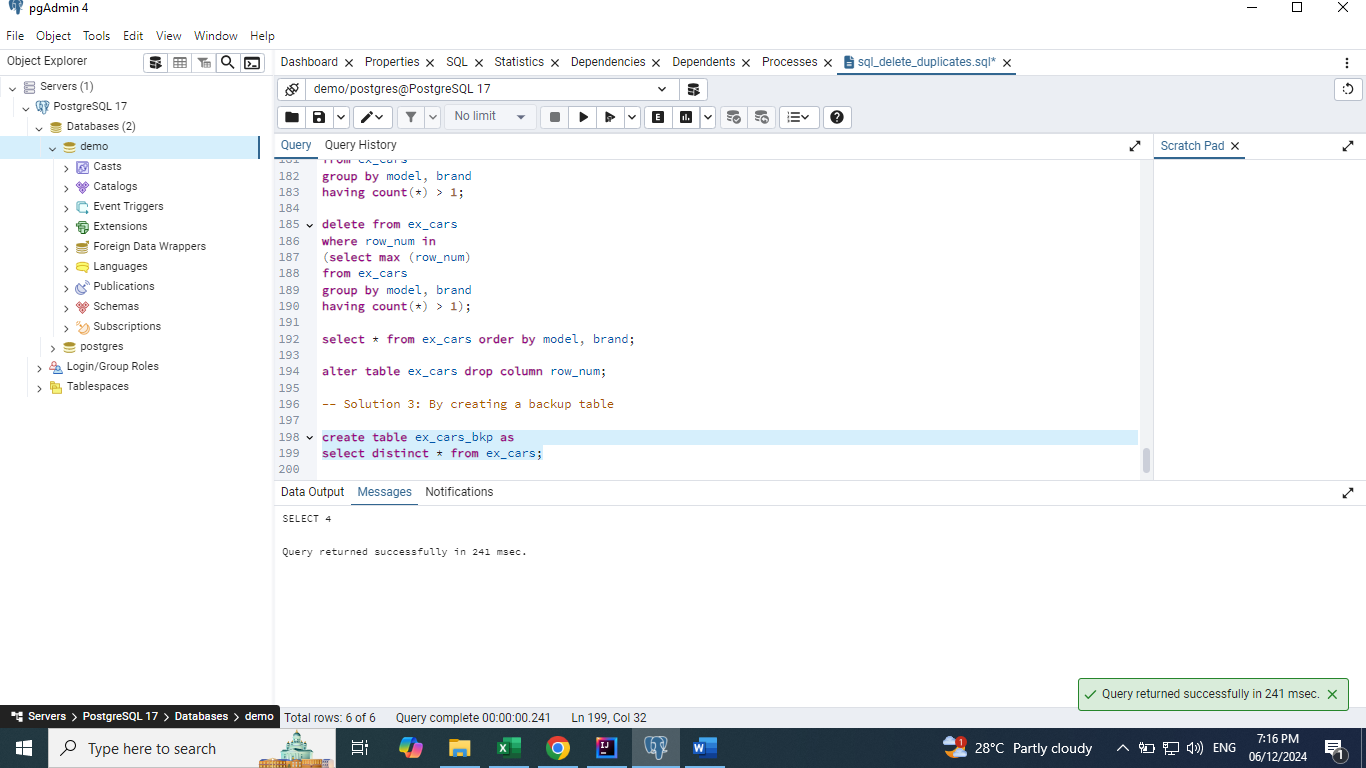
**select \* from ex\_cars order by model, brand;**



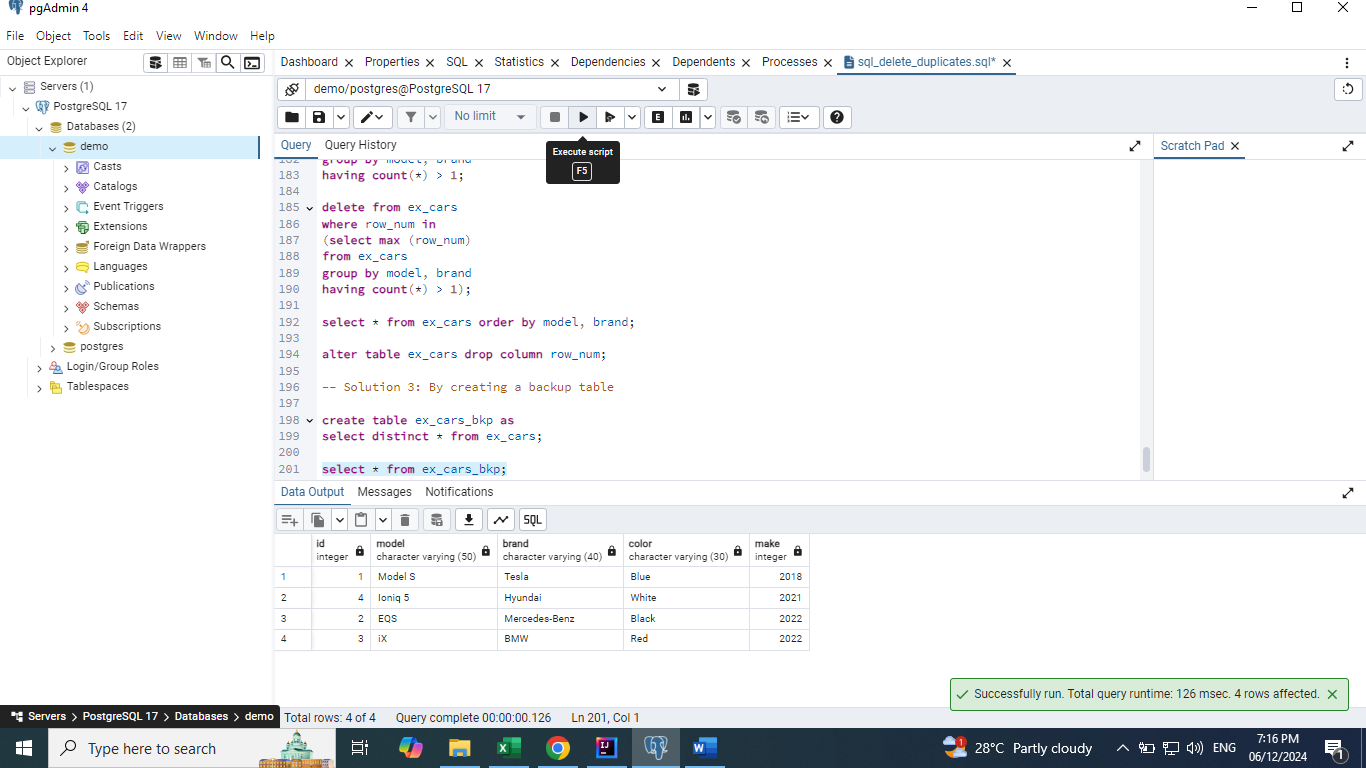
## Solution 3: By creating a backup table

**create table ex\_cars\_bkp as**

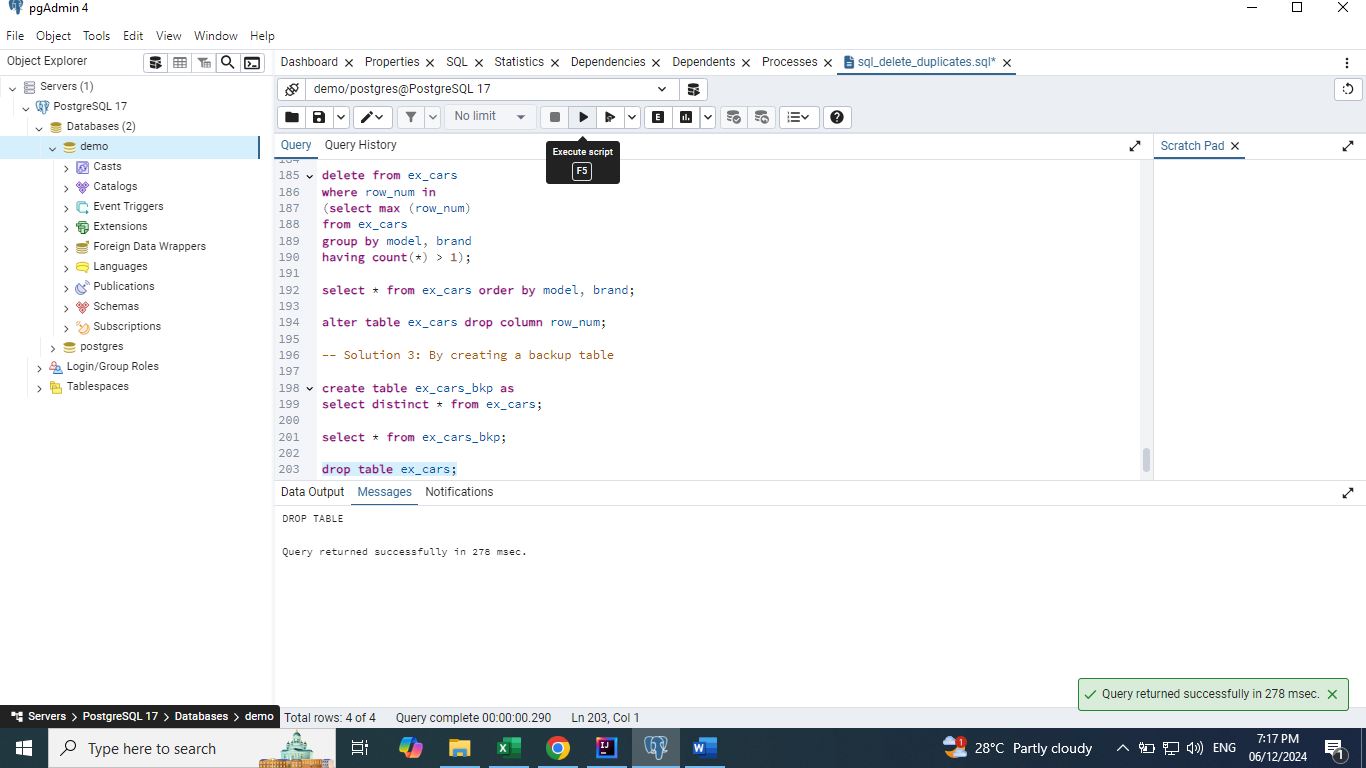
**select distinct \* from ex\_cars;**



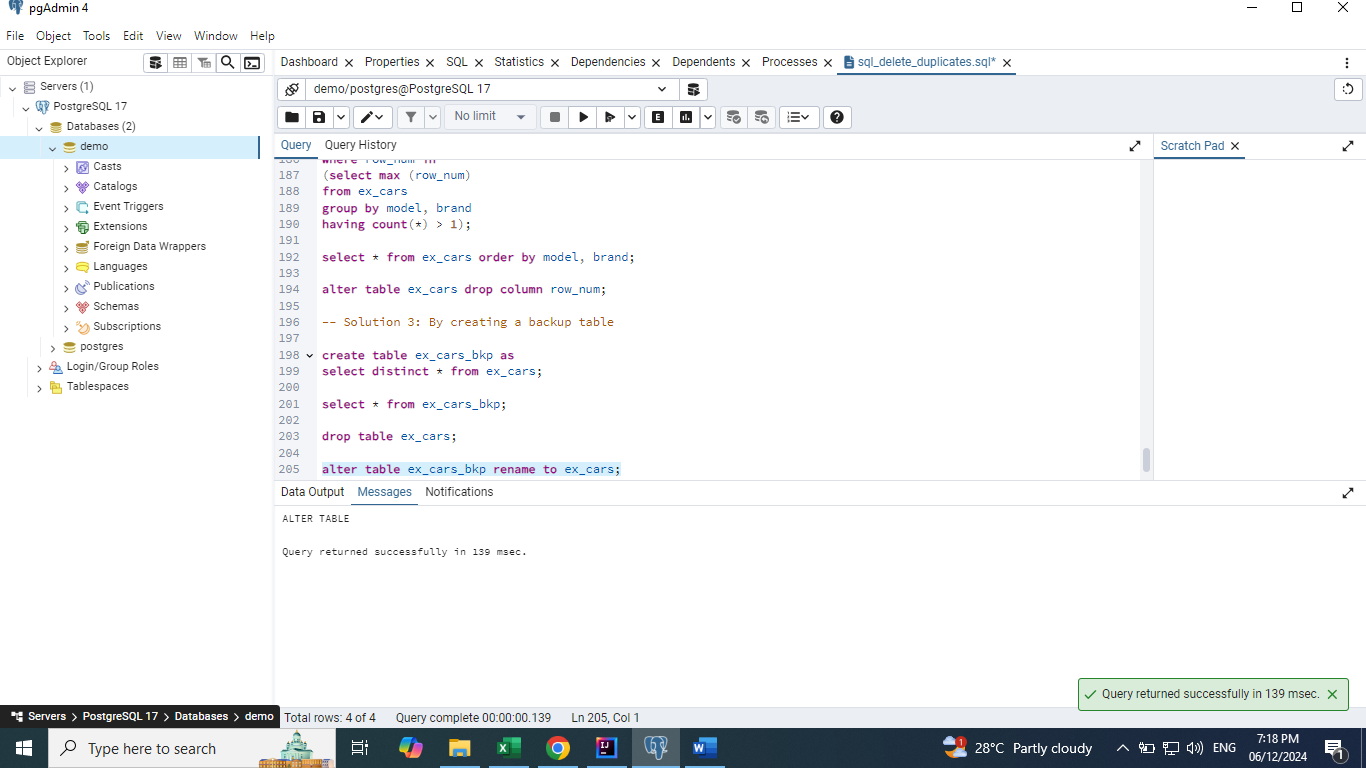
**select \* from ex\_cars\_bkp;**



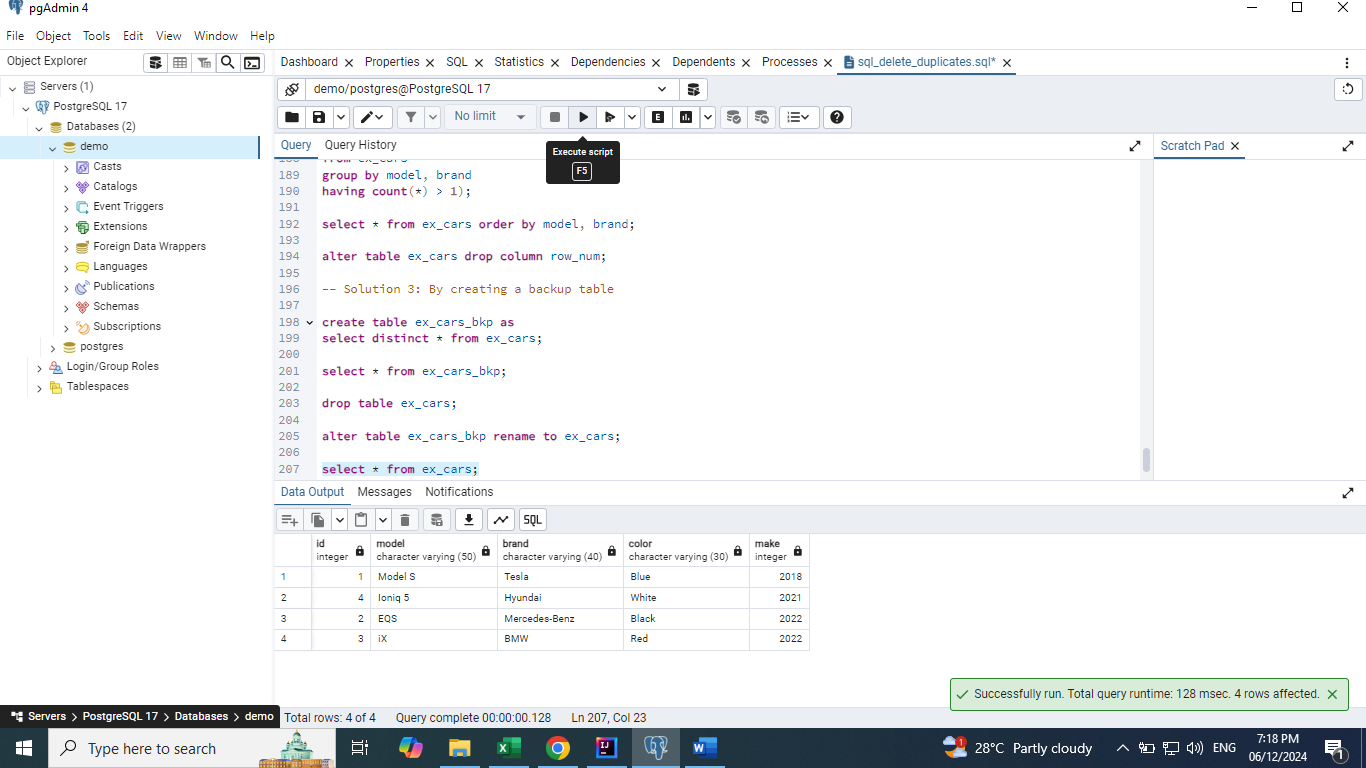
**drop table ex\_cars;**



**alter table ex\_cars\_bkp rename to ex\_cars;**



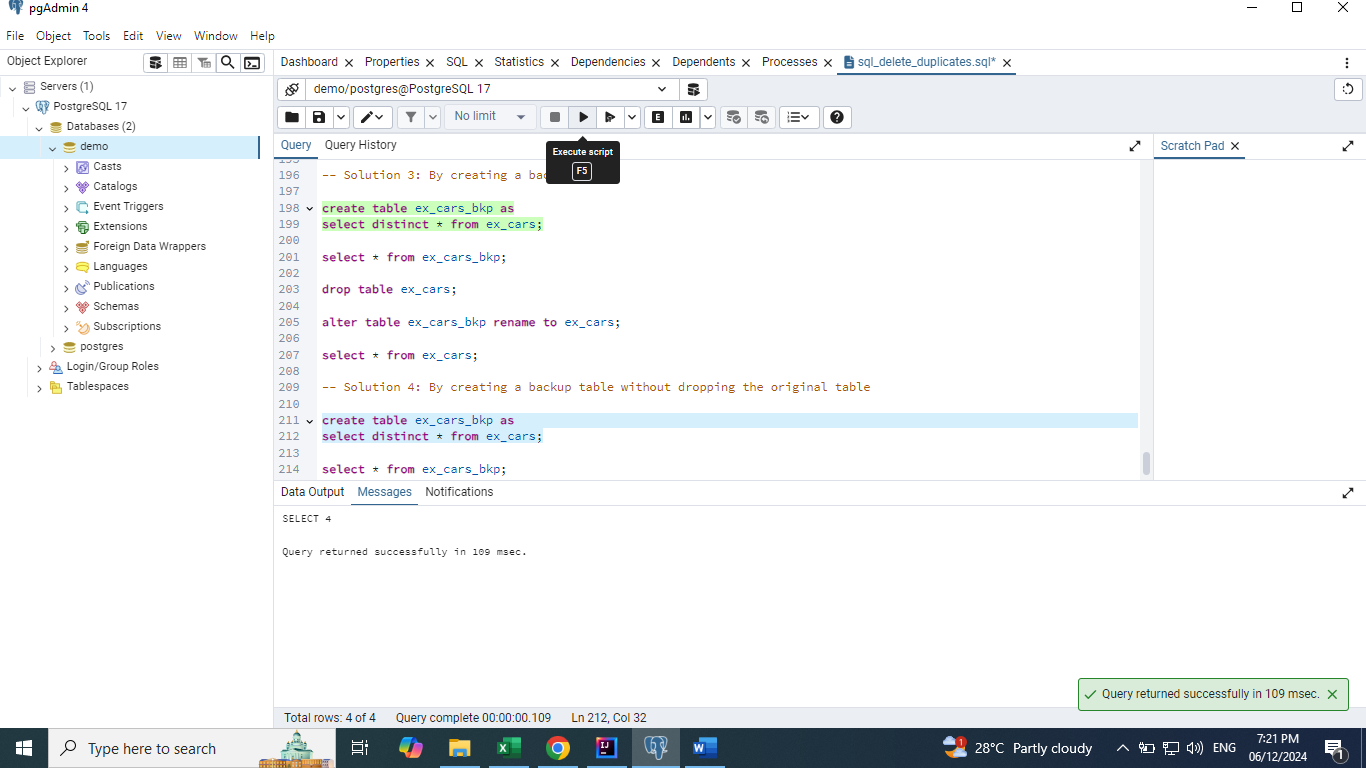
**select \* from ex\_cars;**



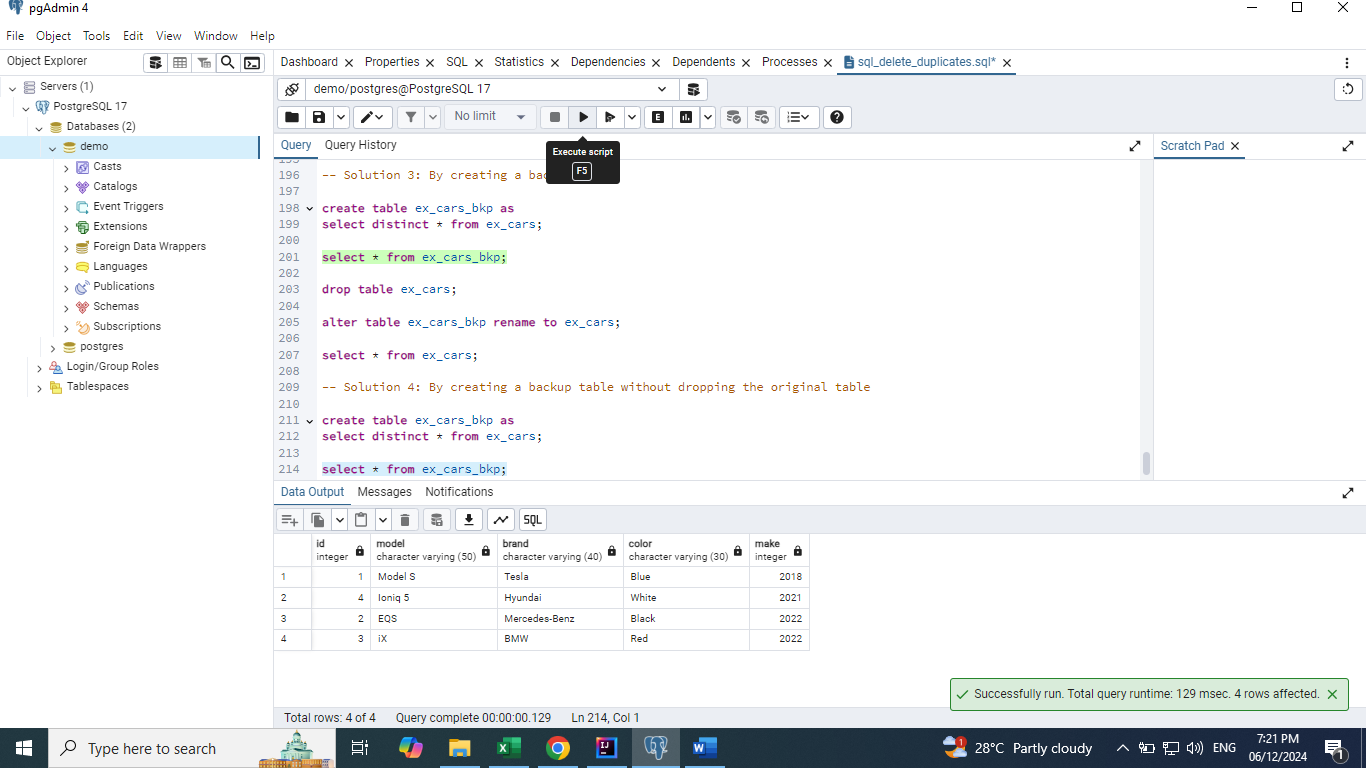
## Solution 4: By creating a backup table without dropping the original table

**create table ex\_cars\_bkp as**

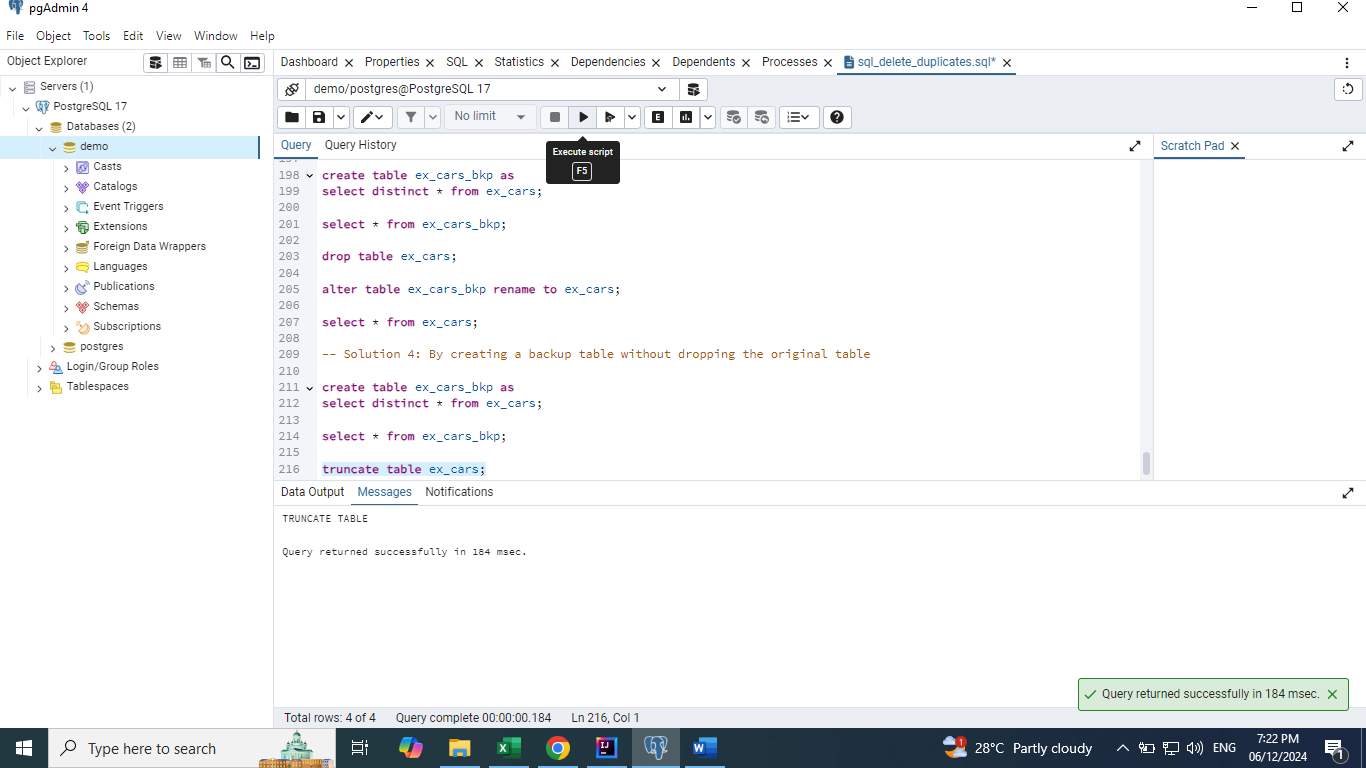
**select distinct \* from ex\_cars;**



**select \* from ex\_cars\_bkp;**

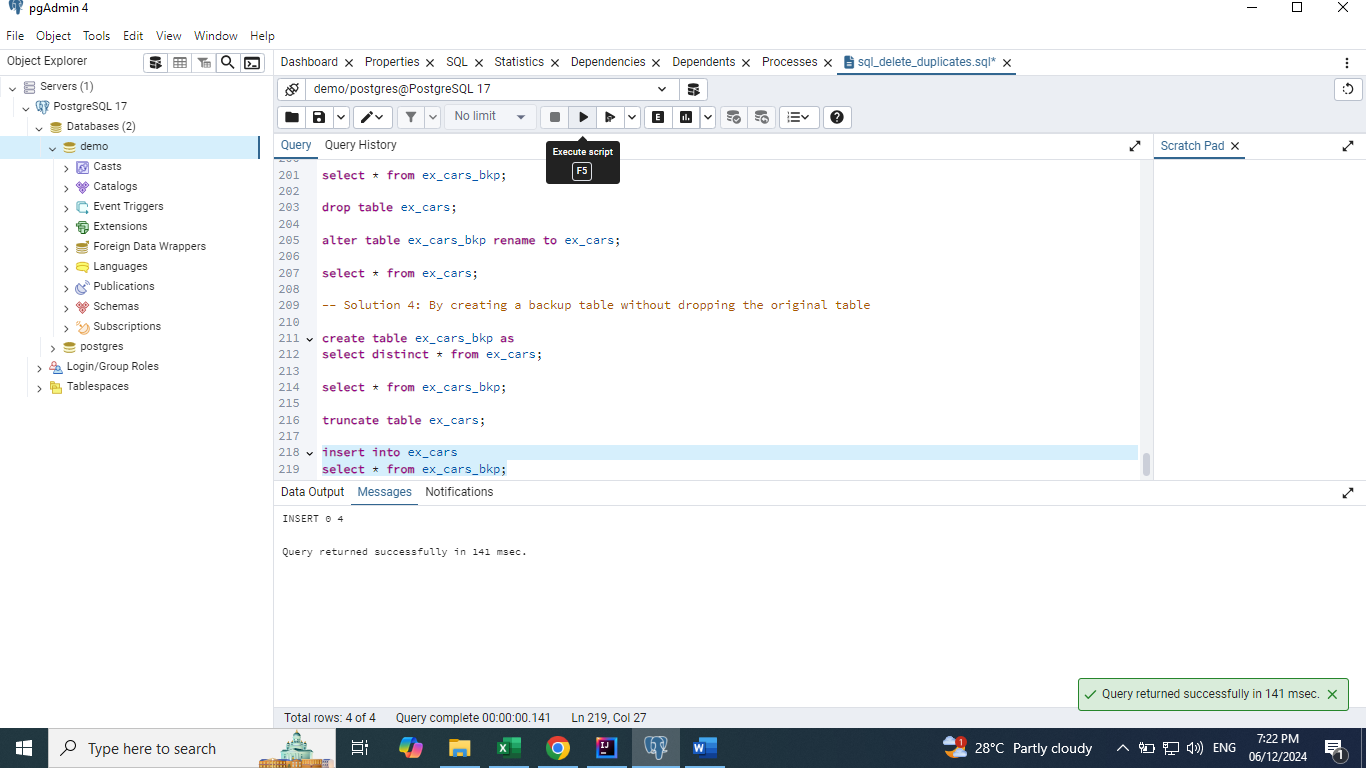


**truncate table ex\_cars;**

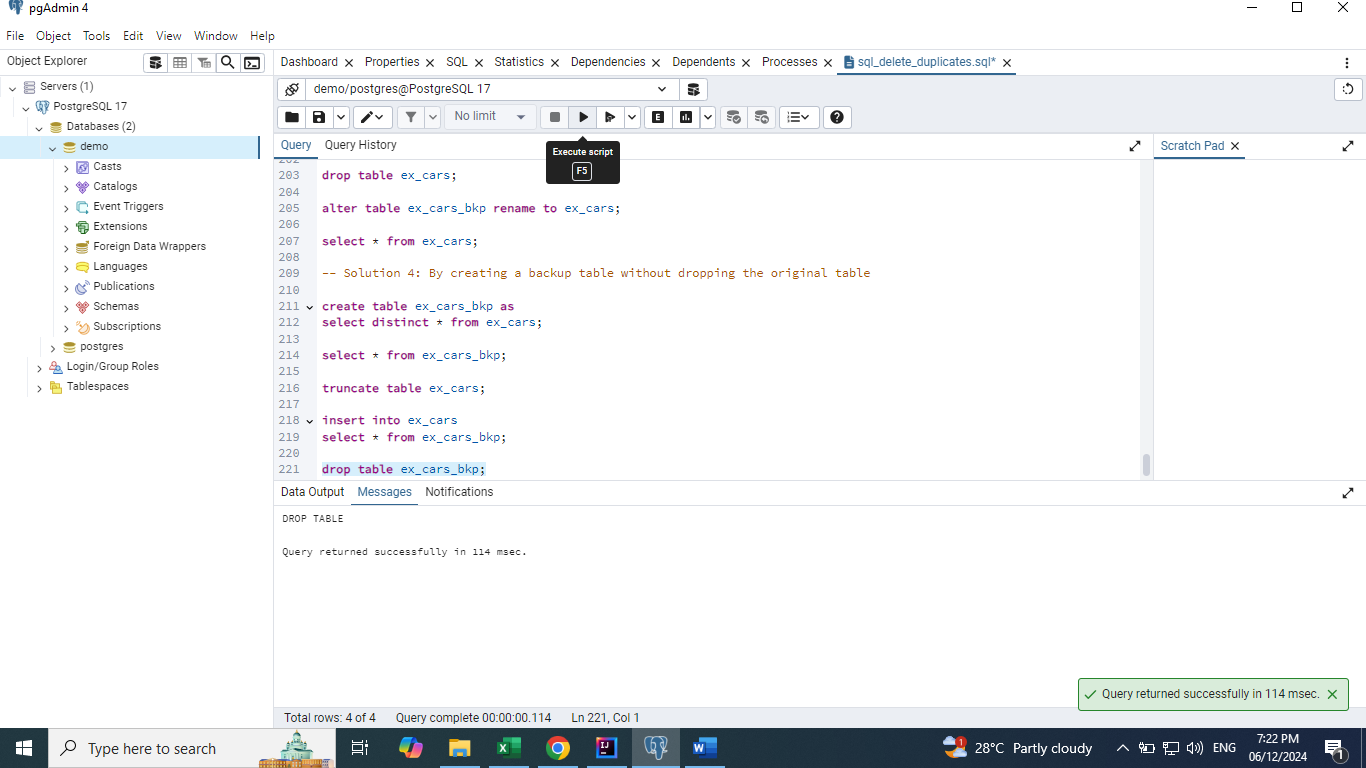


**insert into ex\_cars**

**select \* from ex\_cars\_bkp;**



**drop table ex\_cars\_bkp;**



**select \* from ex\_cars;**

