**Hands on**

**Name : Podutur Lahari - DE126**

**Date : 06-11-2024**

**Data Cleaning Techniques**

**1. Removing Duplicates**

WITH DuplicateEmployees AS (

SELECT \*,

ROW\_NUMBER() OVER (PARTITION BY FirstName, LastName, Department ORDER BY EmployeeID) AS RowNum

FROM Employees

)

DELETE FROM Employees

WHERE EmployeeID IN (

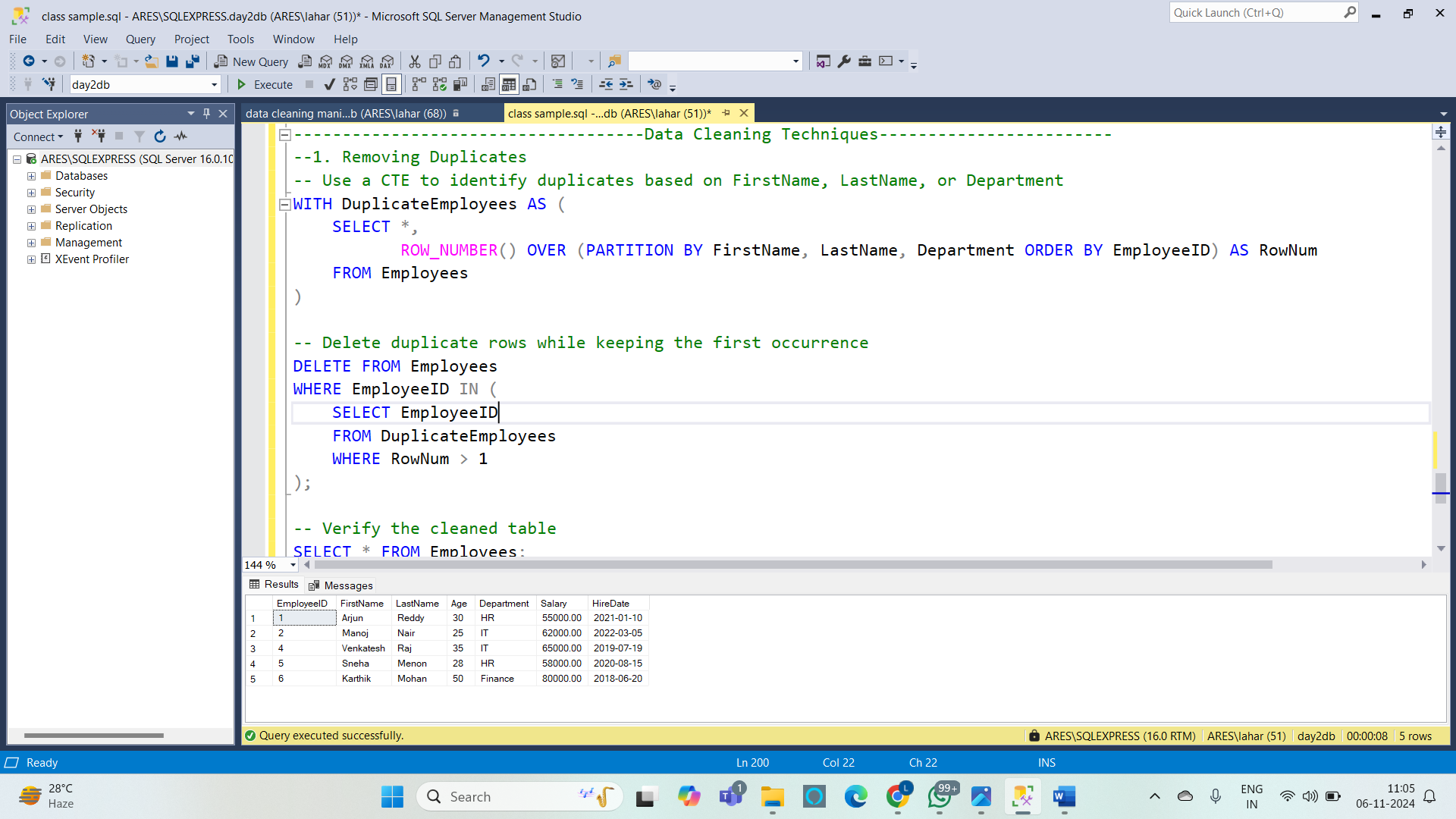
SELECT EmployeeID

FROM DuplicateEmployees

WHERE RowNum > 1

);

SELECT \* FROM Employees;

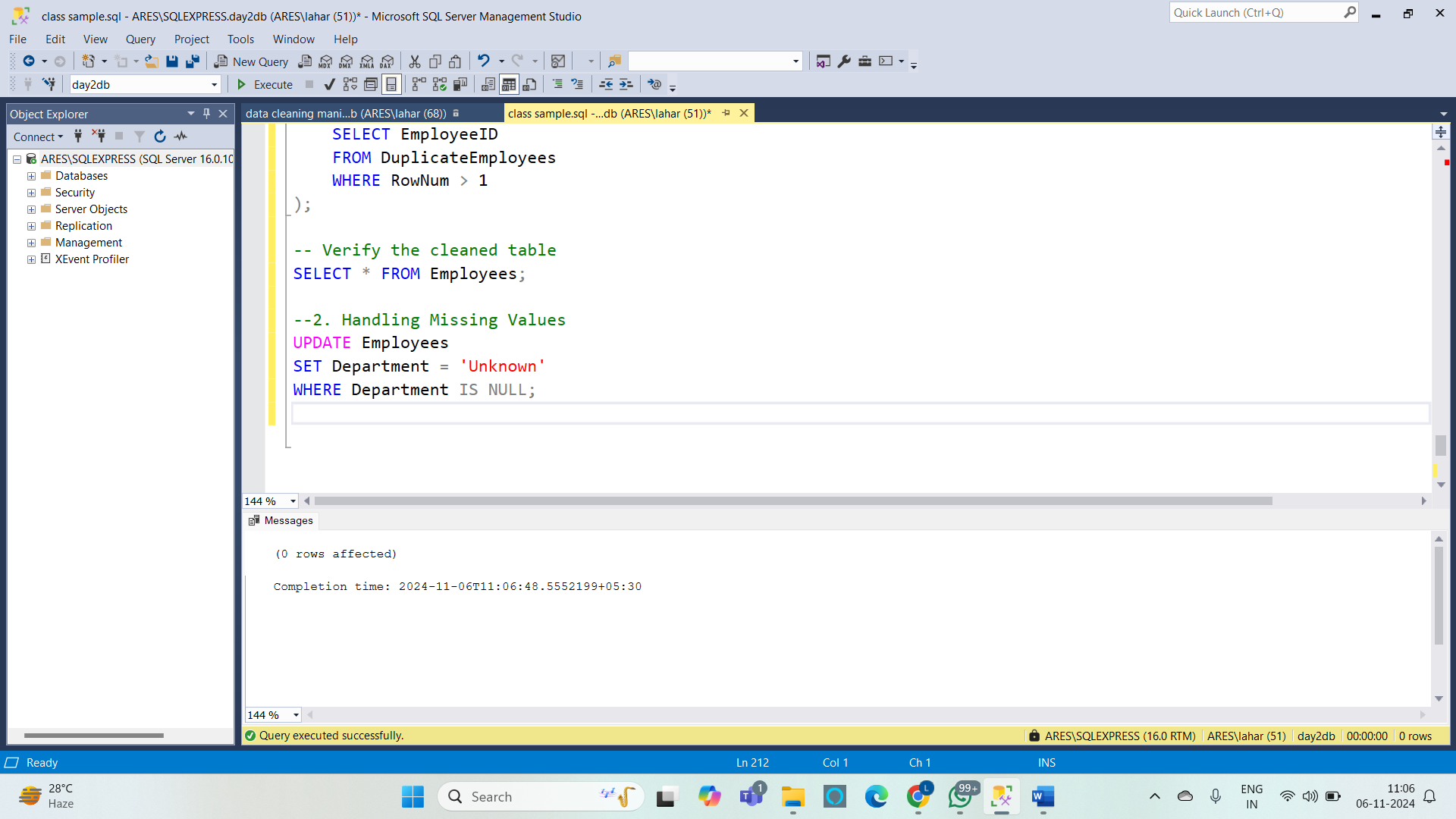


**2. Handling Missing Values :**

UPDATE Employees

SET Department = 'Unknown'

WHERE Department IS NULL;

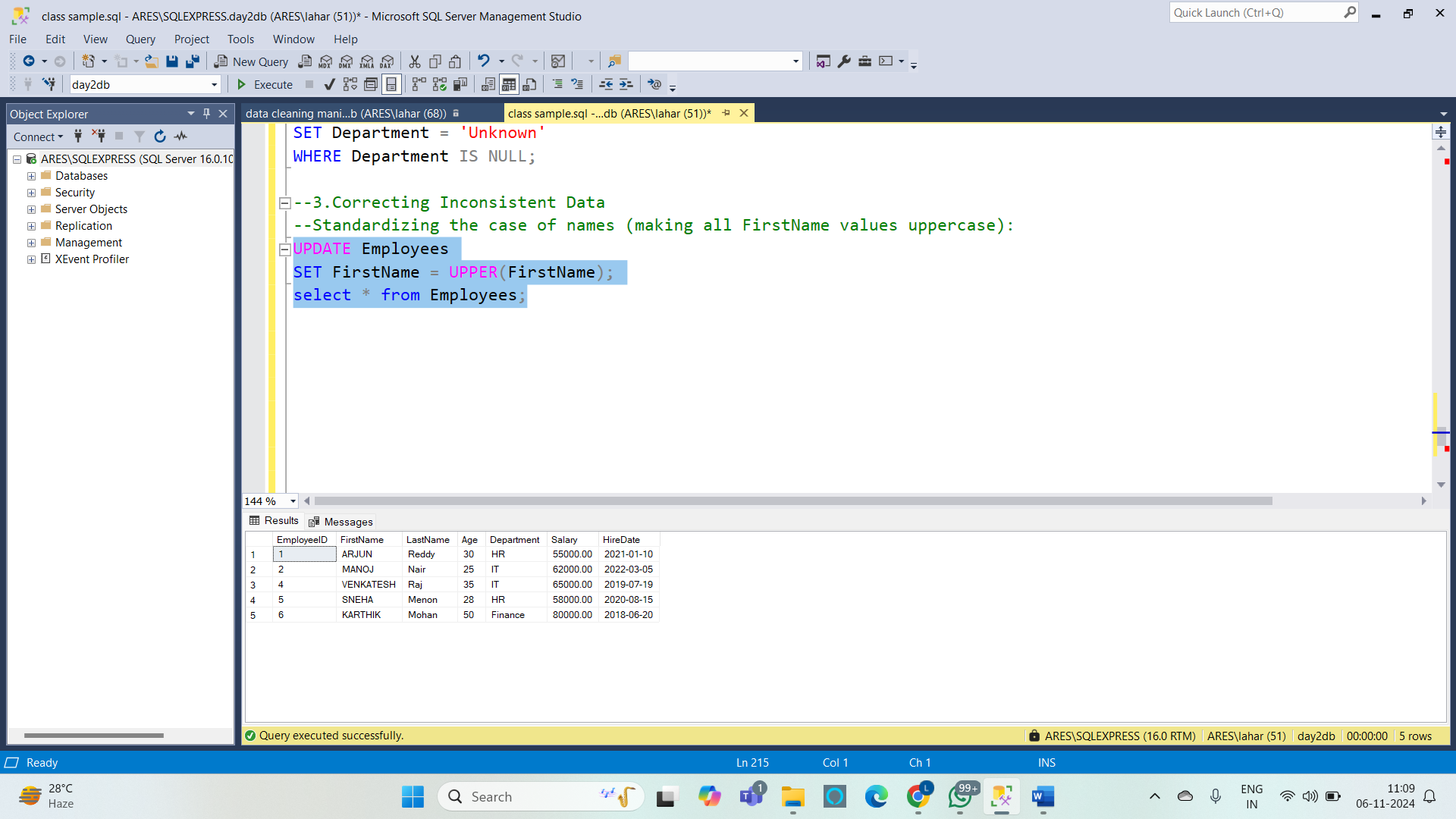


**3. Correcting Inconsistent Data**

**--Standardizing the case of names (making all FirstName values uppercase):**

UPDATE Employees

SET FirstName = UPPER(FirstName);

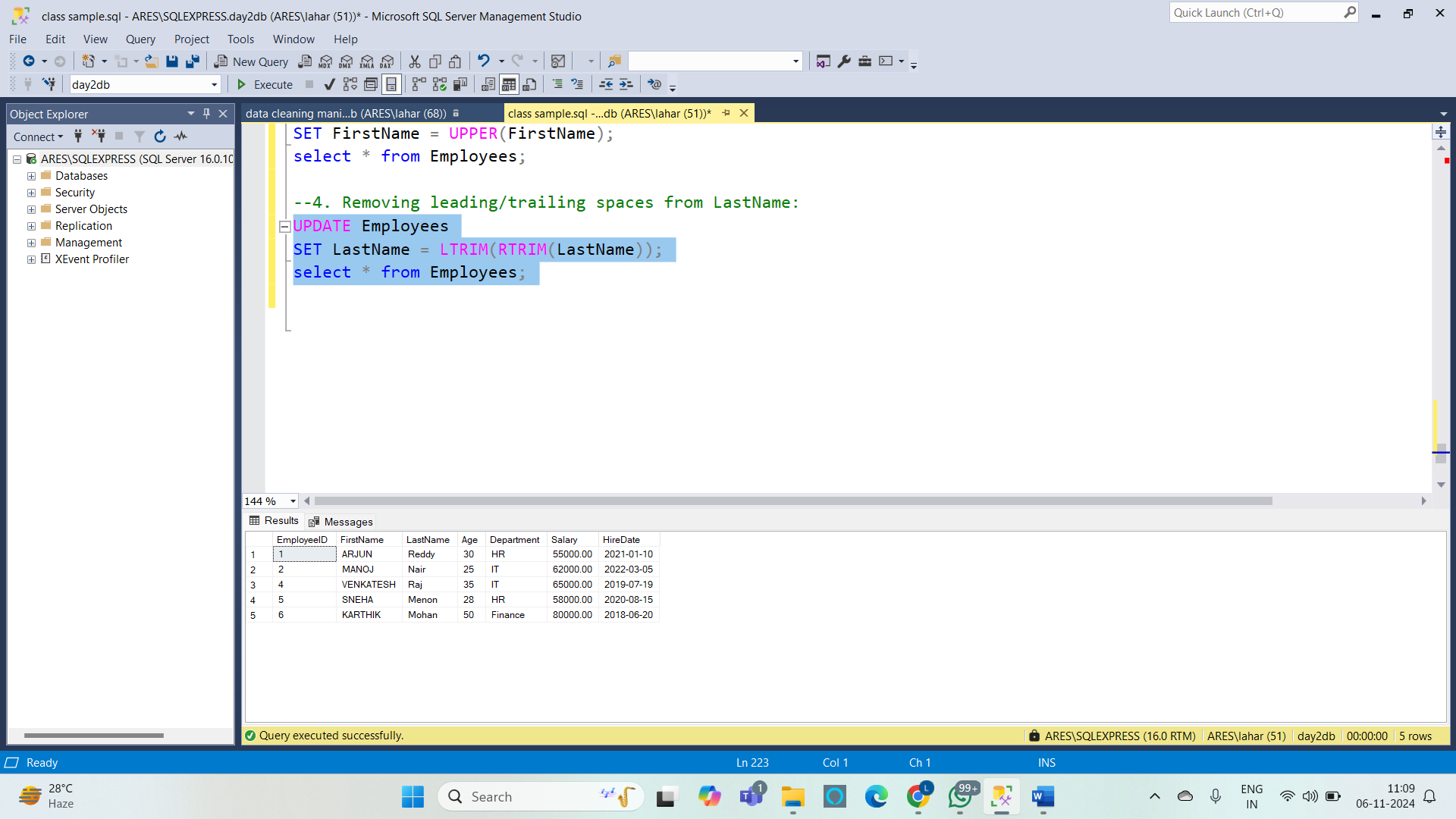
select \* from Employees;

**4. Removing leading/trailing spaces from LastName:**

UPDATE Employees

SET LastName = LTRIM(RTRIM(LastName));

select \* from Employees;

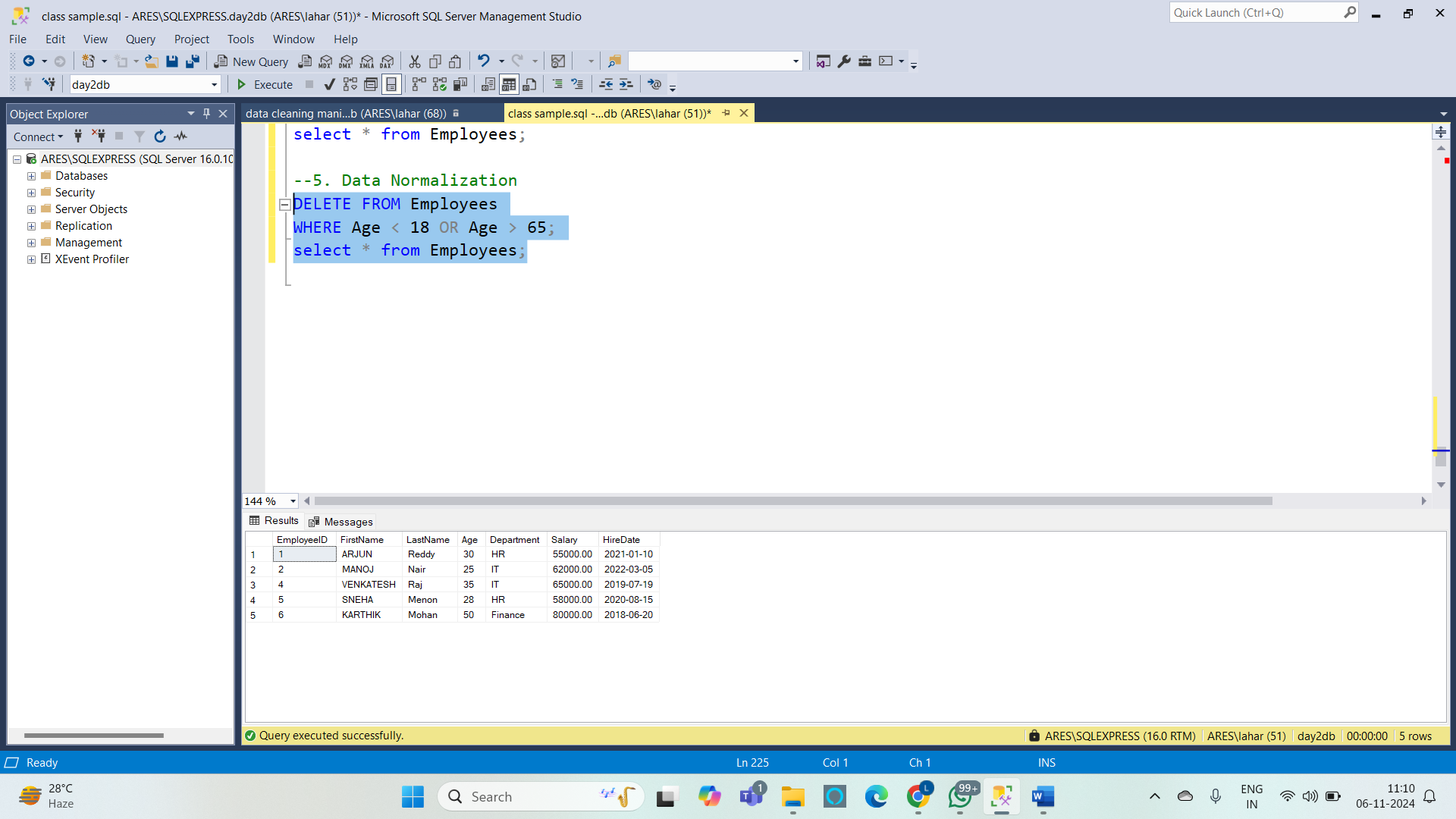


**5.**  **Data Normalization :**

DELETE FROM Employees

WHERE Age < 18 OR Age > 65;

select \* from Employees;



**6. Handling Outliers :**

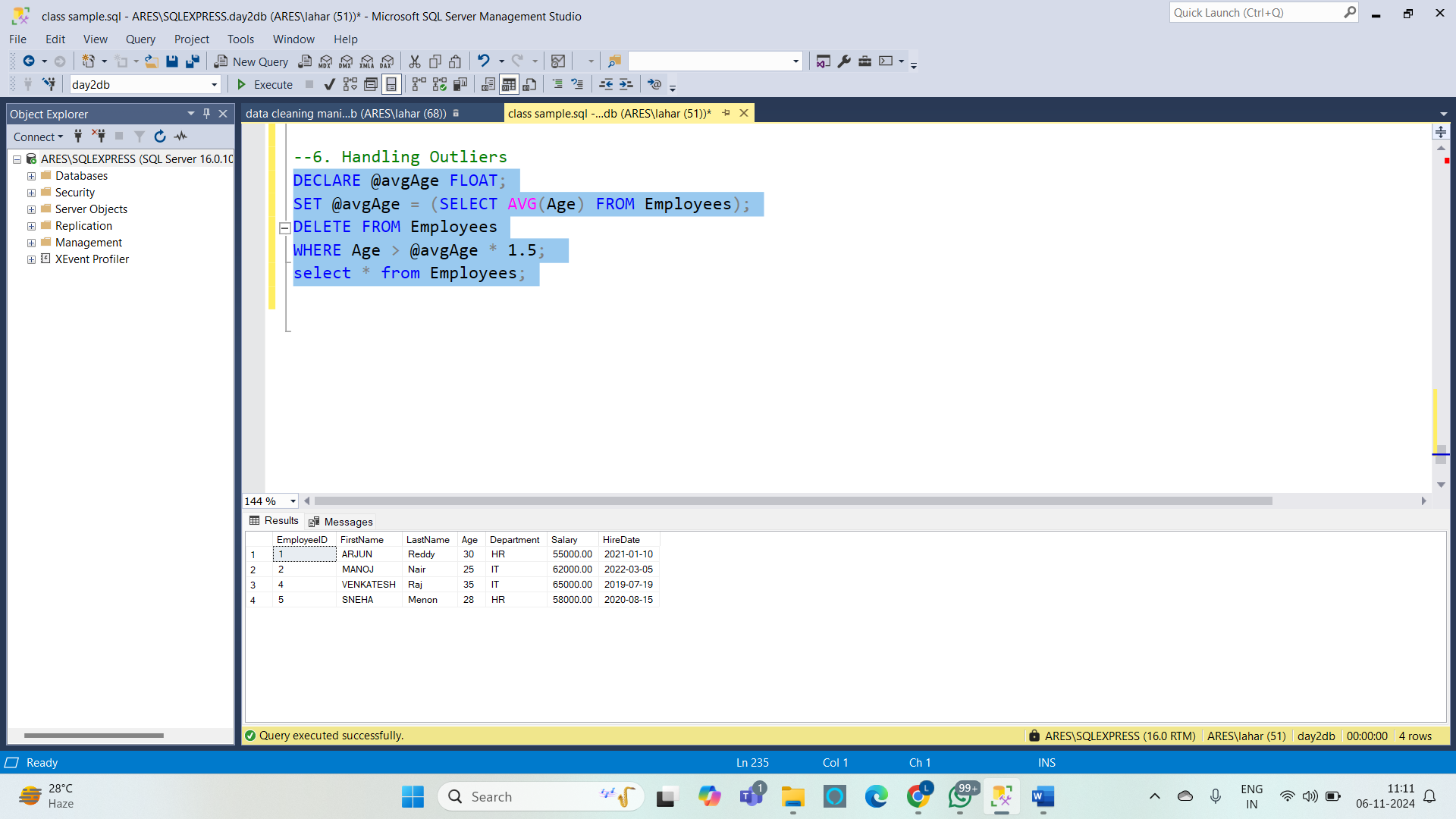
DECLARE @avgAge FLOAT;

SET @avgAge = (SELECT AVG(Age) FROM Employees);

DELETE FROM Employees

WHERE Age > @avgAge \* 1.5;

select \* from Employees;



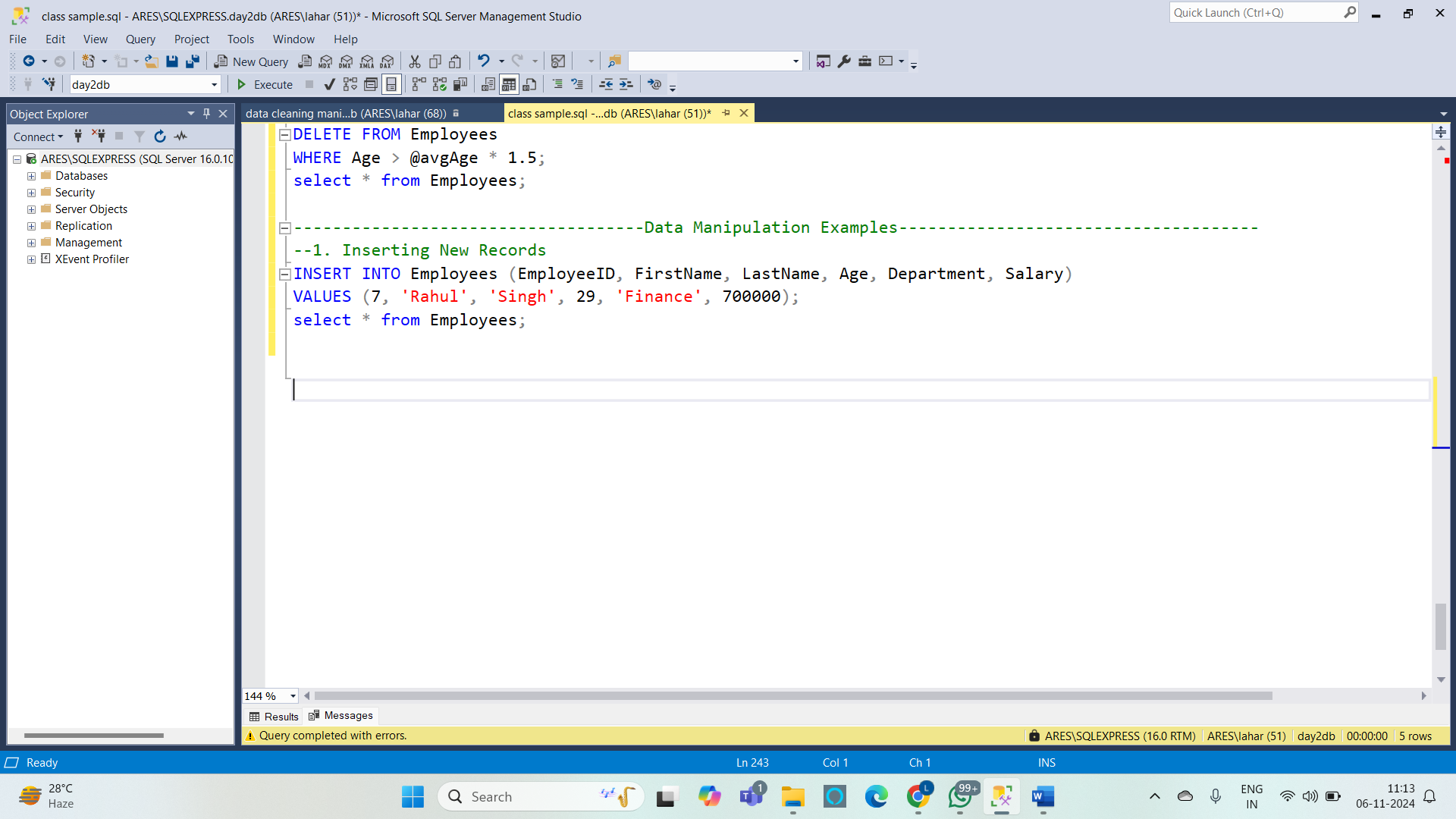
**Data Manipulation Examples**

**1. Inserting New Records :**

INSERT INTO Employees (EmployeeID, FirstName, LastName, Age, Department, Salary)

VALUES (7, 'Rahul', 'Singh', 29, 'Finance', 700000);

select \* from Employees;



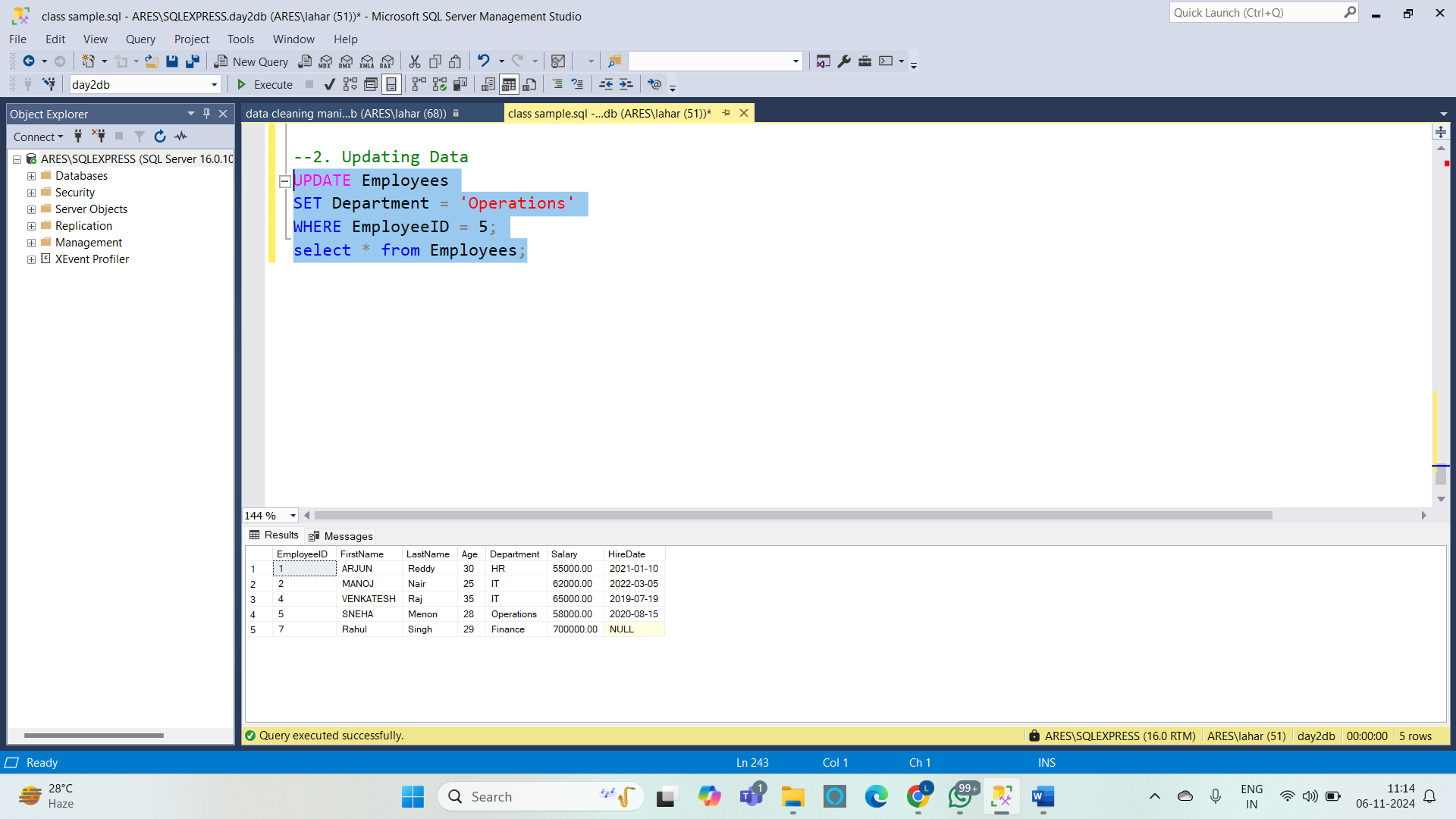
**2. Updating Data :**

UPDATE Employees

SET Department = 'Operations'

WHERE EmployeeID = 5;

select \* from Employees;

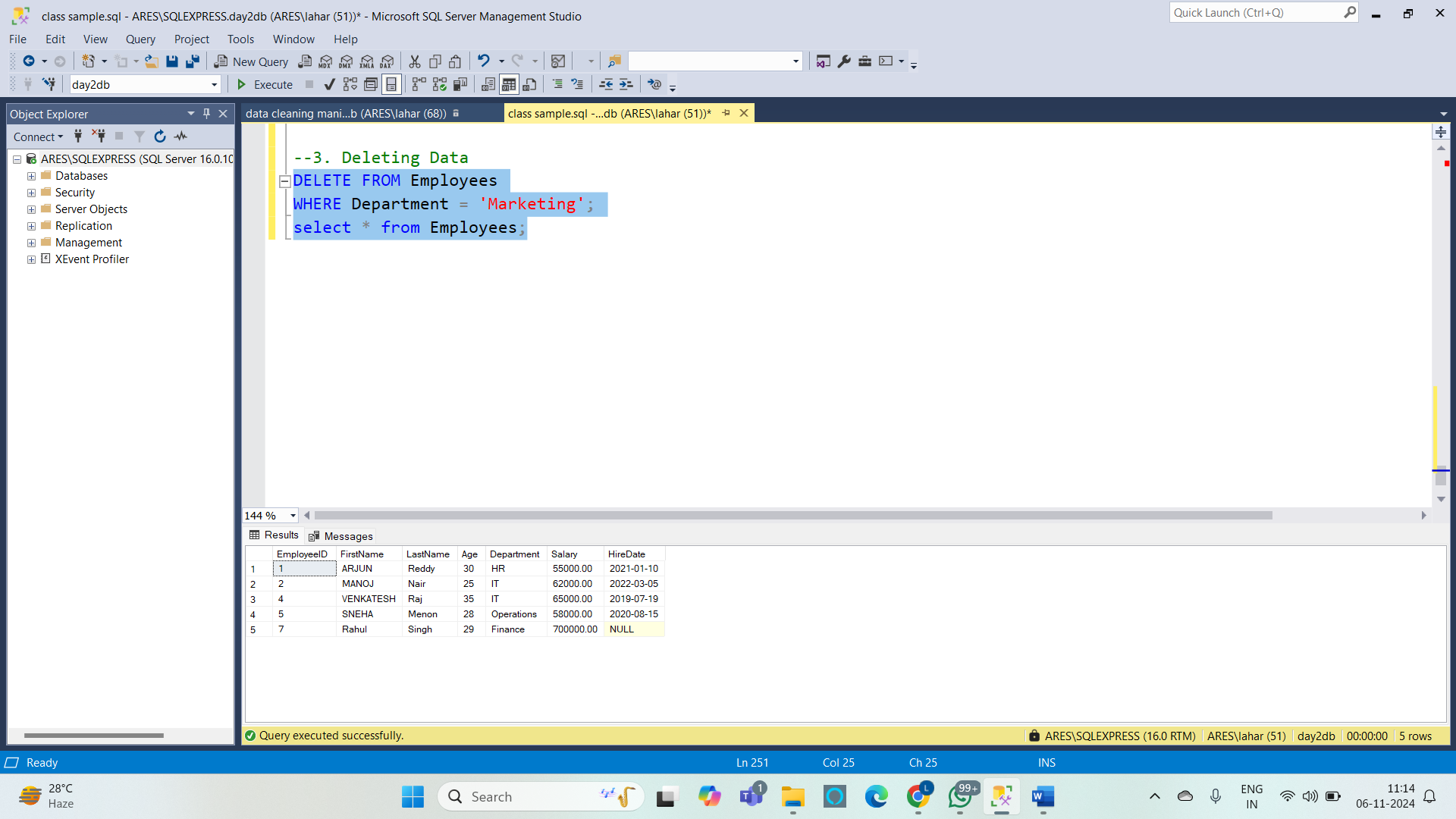


**3. Deleting Data :**

DELETE FROM Employees

WHERE Department = 'Marketing';

select \* from Employees;

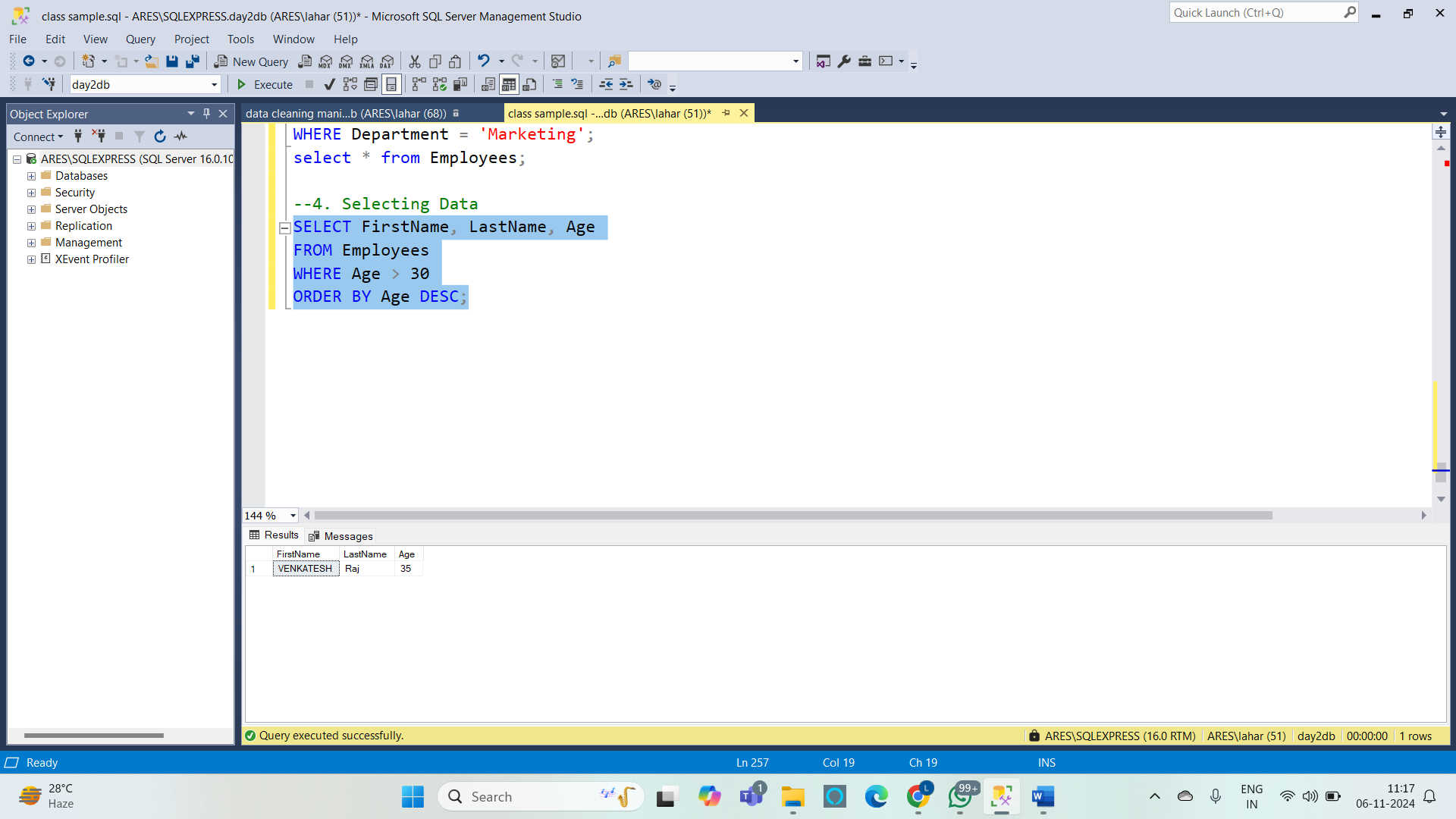


**4. Selecting Data :**

SELECT FirstName, LastName, Age

FROM Employees

WHERE Age > 30

ORDER BY Age DESC;

**5. Filtering Data with Additional Conditions :**

SELECT FirstName, LastName, Department

FROM Employees

WHERE Department IN ('HR', 'Sales') AND Age BETWEEN 25 AND 45

ORDER BY Department, Age;

