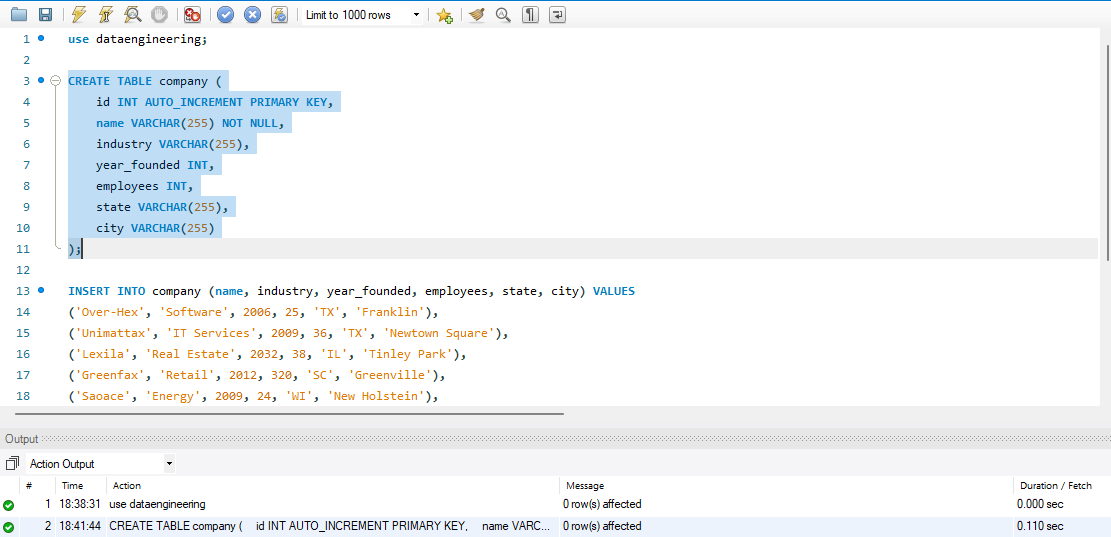
Name - Aparna Bharti

Batch - Data Engineering batch-1

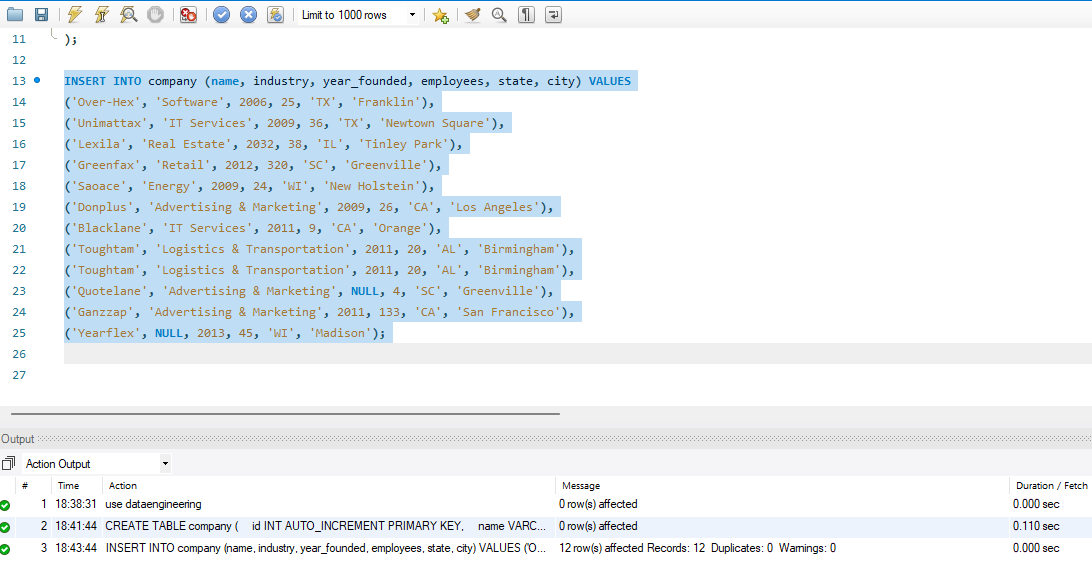
**Day -5**

**Data Cleaning**

**Create table company**

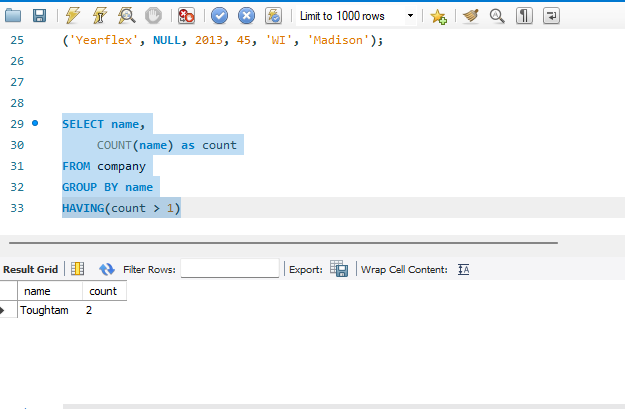
****

**Insert values**

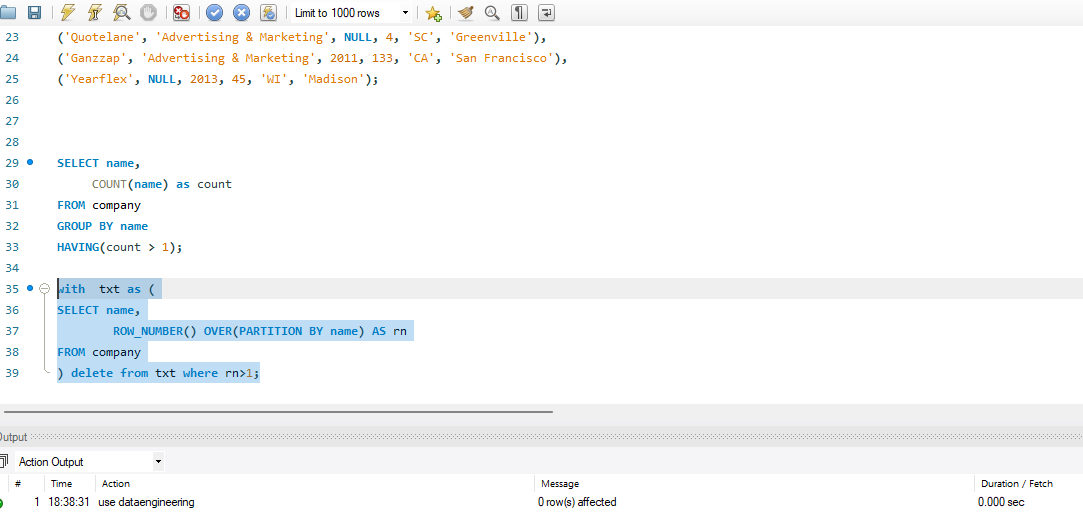
****

**Step1 :Deleting Duplicate data**

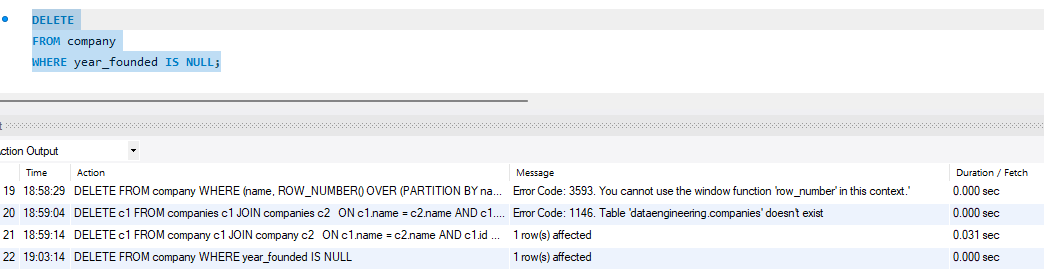
the GROUP BY and HAVING clauses to identify duplicate names. This query is going to count the number of times each name exists in the database using GROUP BY. Then it uses the HAVING clause to filter the results for only those names that exist more than once.

****

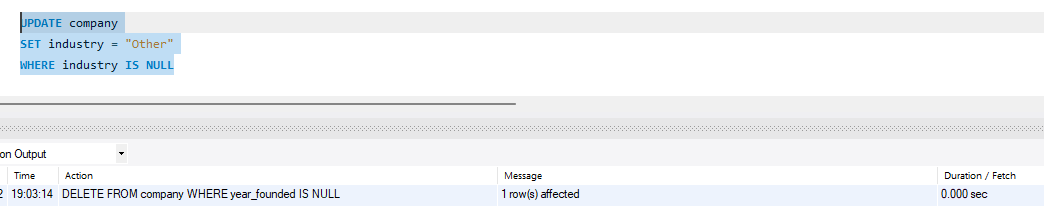
use a combination of ROW\_NUMBER() and DELETE



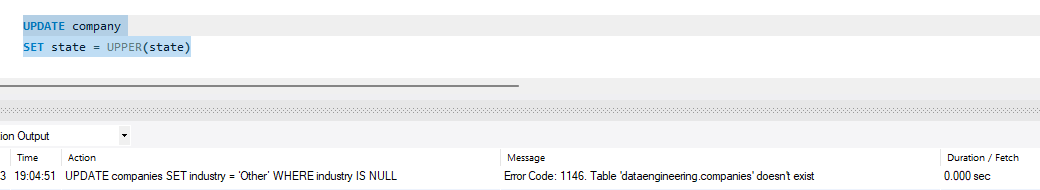
**Step 2:Removing NULL values**

****

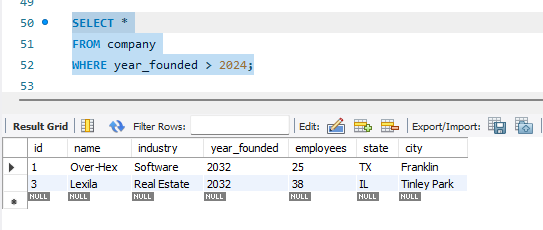
**Step 3:Put meaningful label for NULL values**

****

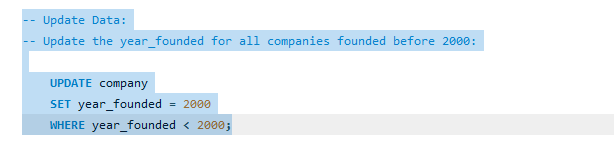
**Step 4 :Fix the capitalise value**

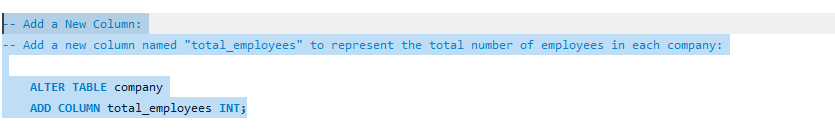
****

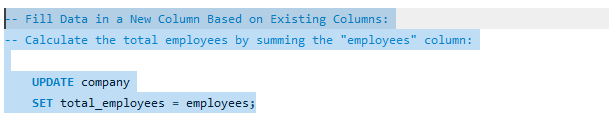
**Step 5:Correct logical error**

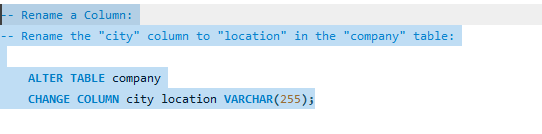
****

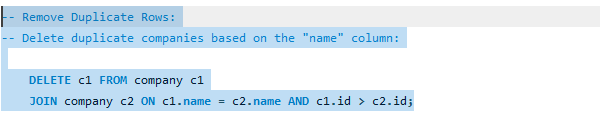
**Transformation queries**

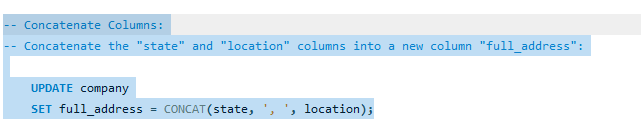
****

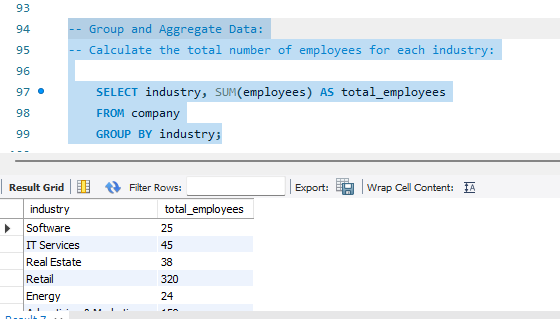
****

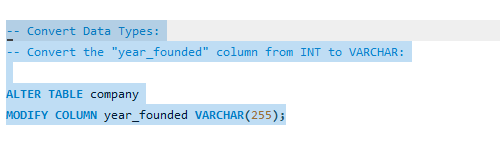
****

****

****

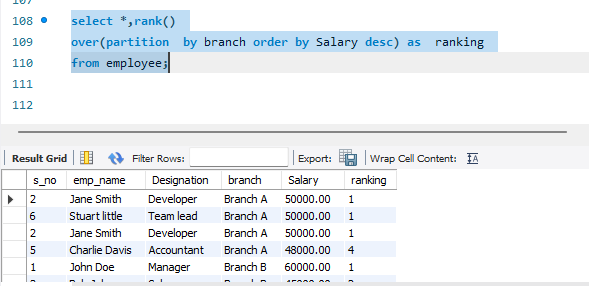
****

****

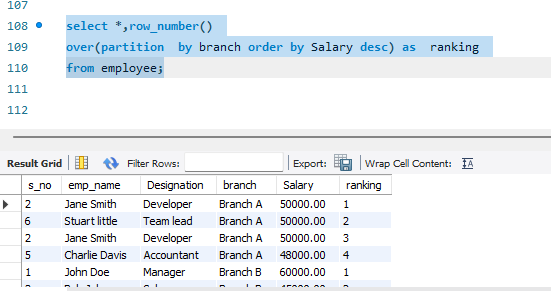
****

**Ranking Functions**

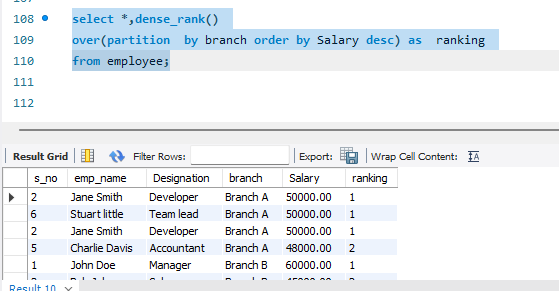
**Rank()**

****

**Row\_number()**

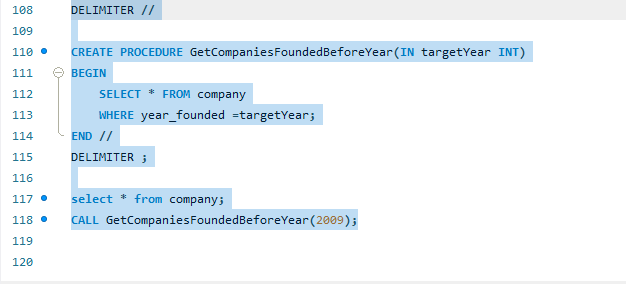
****

**Dense\_rank()**

****

**Stored Procedure**

A stored procedure is **a prepared SQL code that you can save, so the code can be reused over and over again**. So if you have an SQL query that you write over and over again, save it as a stored procedure, and then just call it to execute it.

****