Abhiram Basa

Assessment 7 of Data Engineering - 11/12/23 .

Today we discussed about

* **Cleaning the data set.**
  + Removing duplicate rows.
  + Removing null values.
  + Deleting the duplicate rows using row number().
  + Updating the data.
* **Schema diagrams** 
  + Star schema
  + Snowflake schema
* **Partition By**

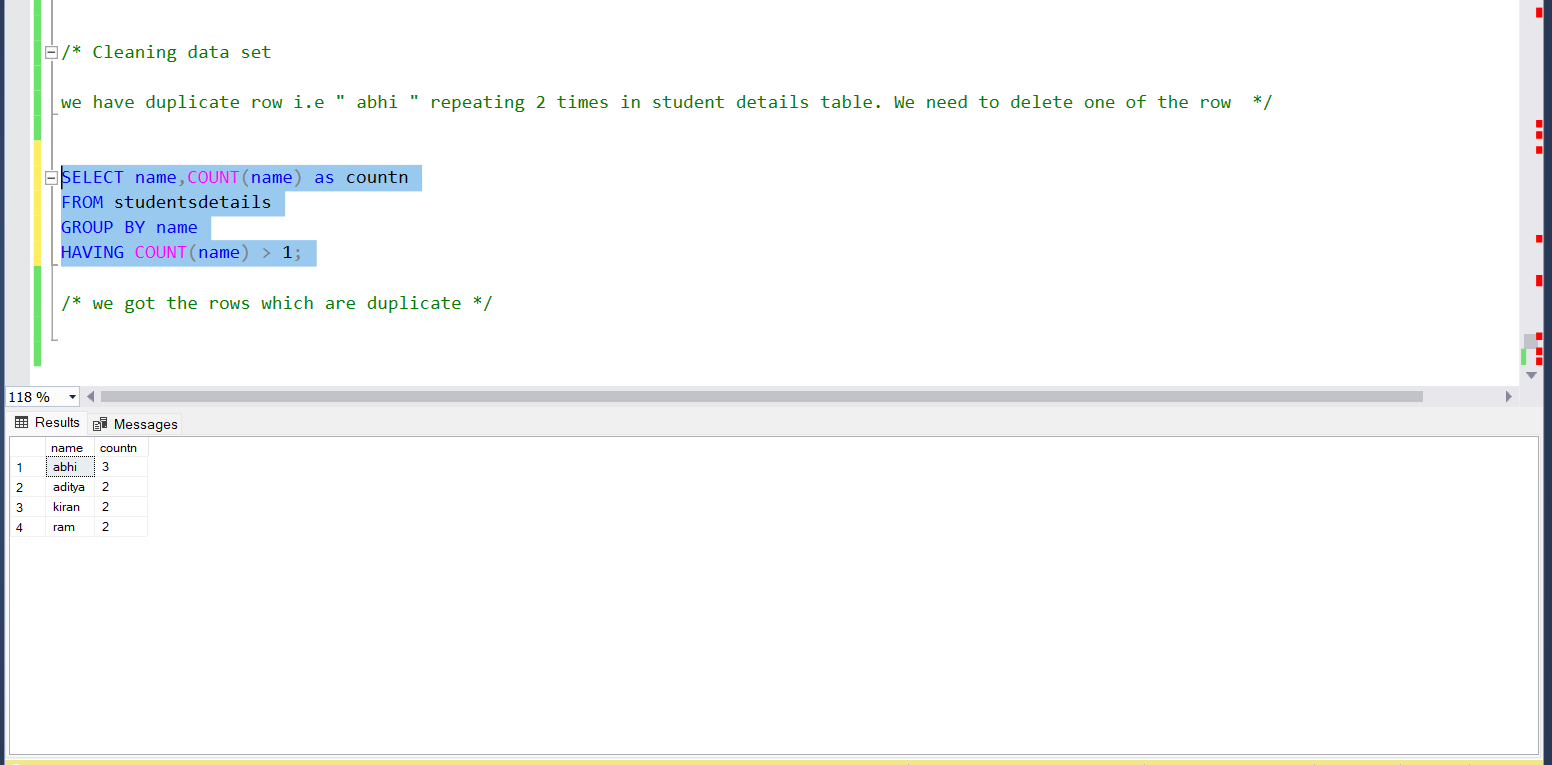
**Cleaning data set :**

To clean the data set we can use normal commands like delete, update, insert.

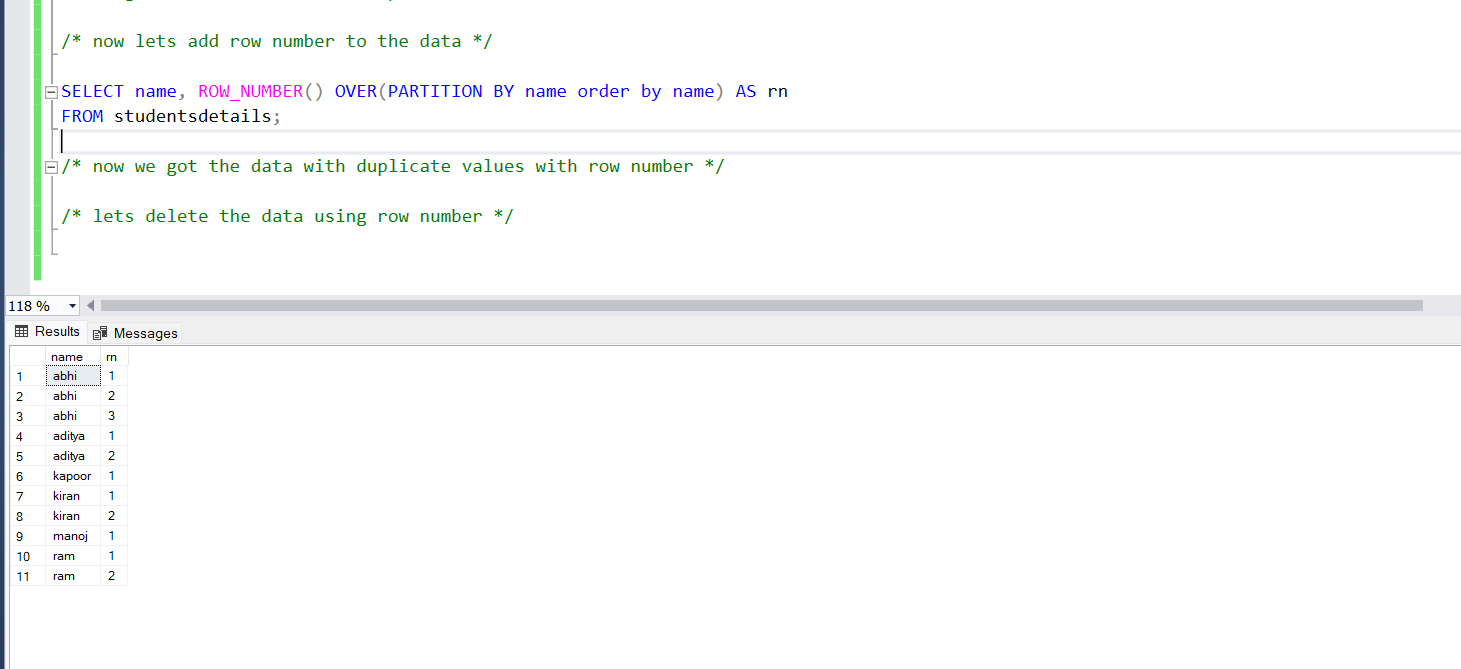
Data cleaning is very important step in data engineering. Other wise, the Analytics with improper data gives you in correct insights.

It is the very first step I.e to remove improper values, null values, etc.

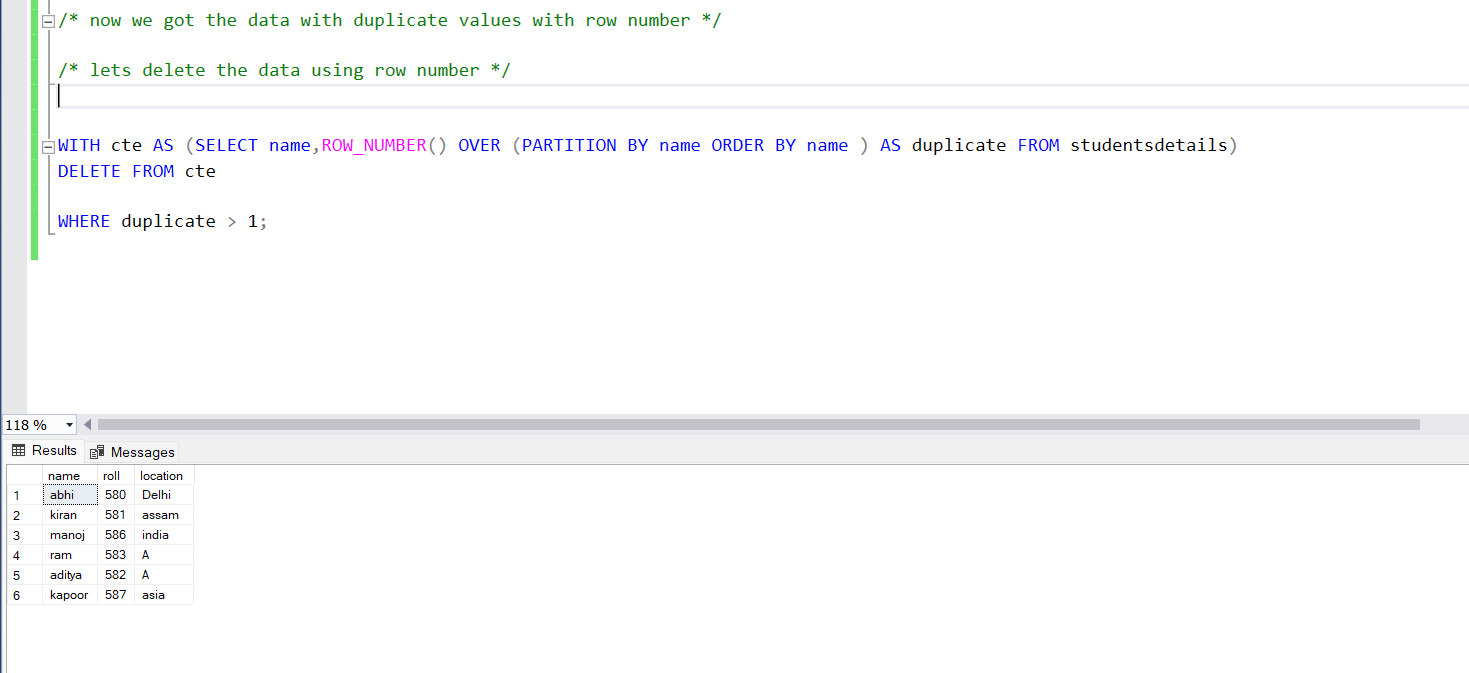
* Using row number to assign to each row and deleting the duplicate row using the row number.



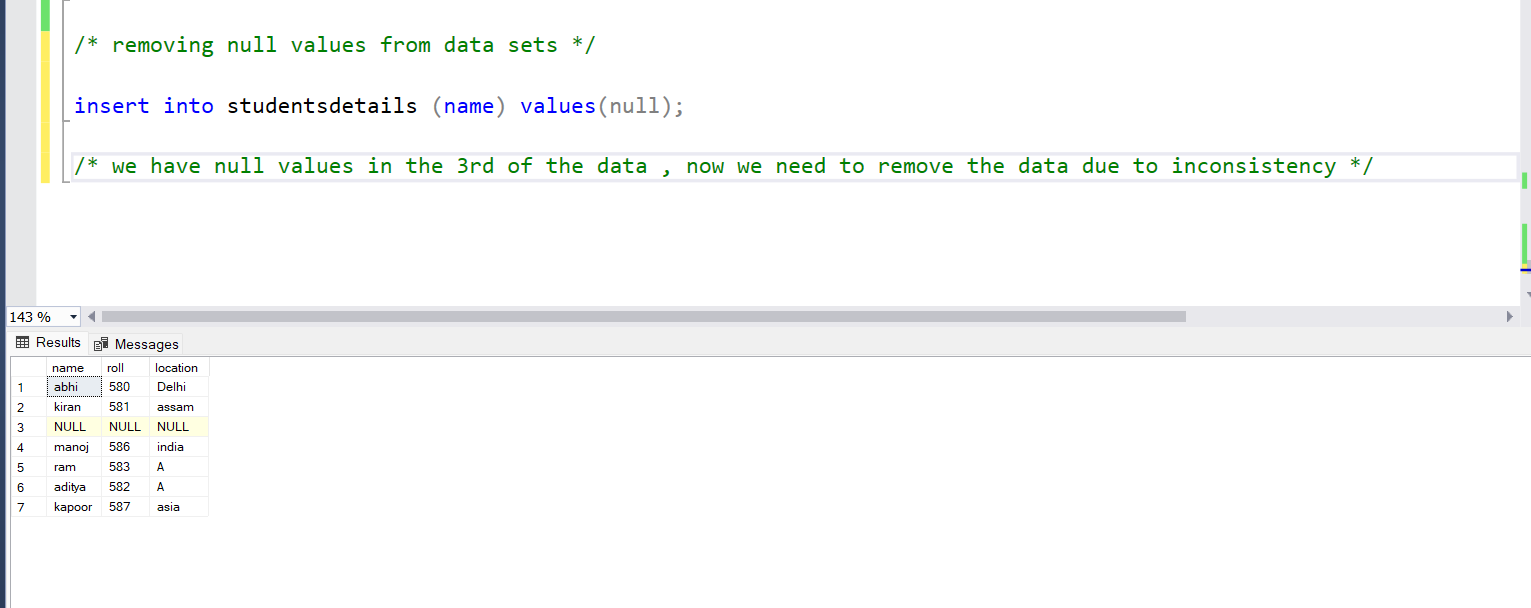
* we got the rows which are duplicate.
* now we got the data with duplicate values with row number.



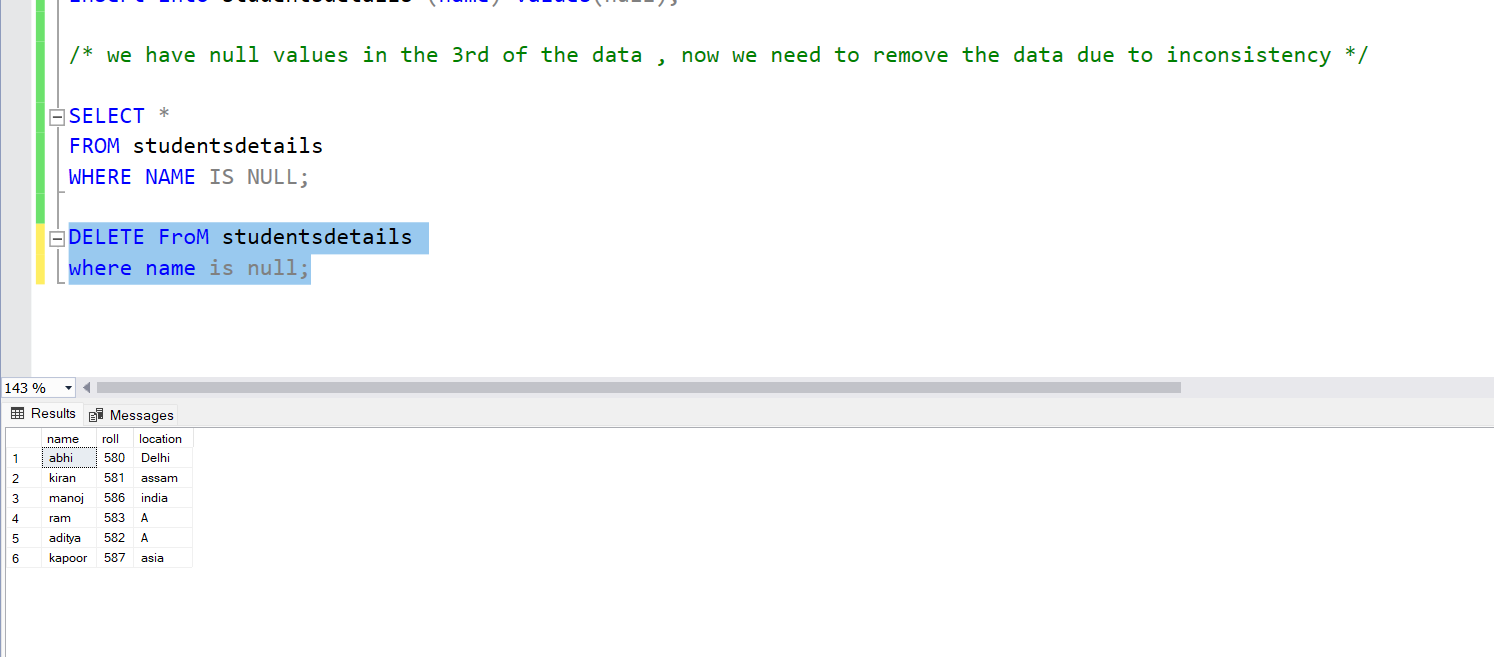
* lets delete the data using CTE, row number().



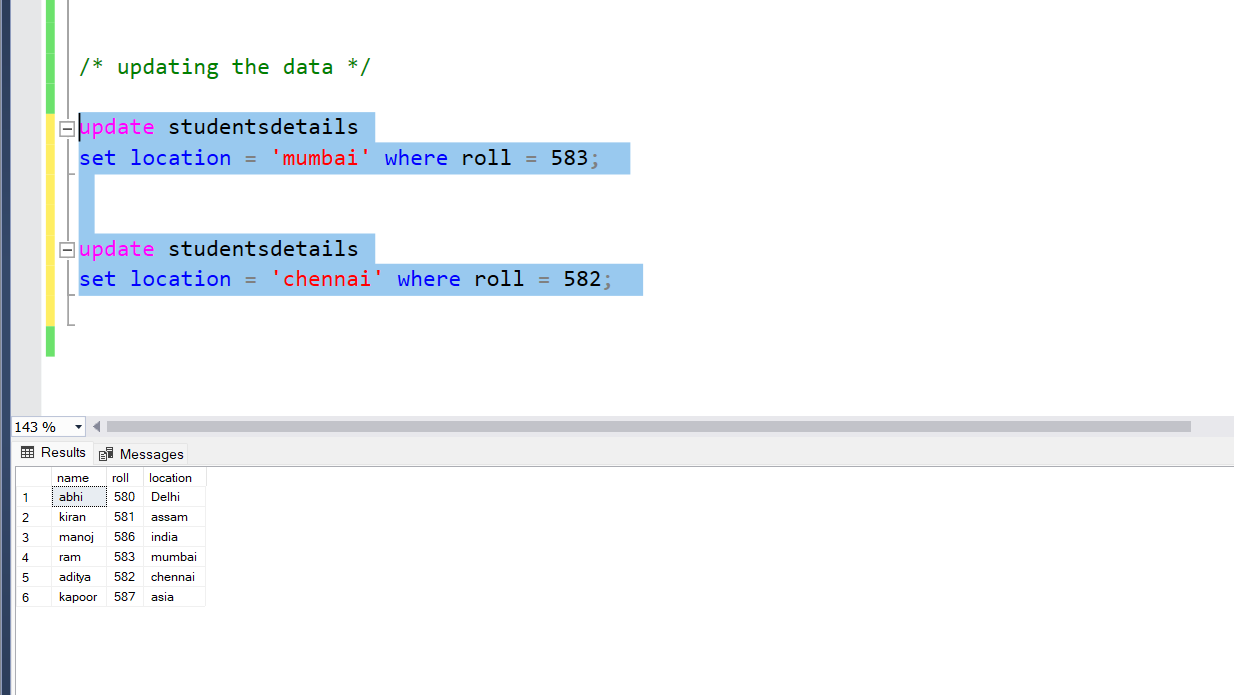
* removing null values from data sets.
* we have null values in the 3rd of the data , now we need to remove the data due to inconsistency.

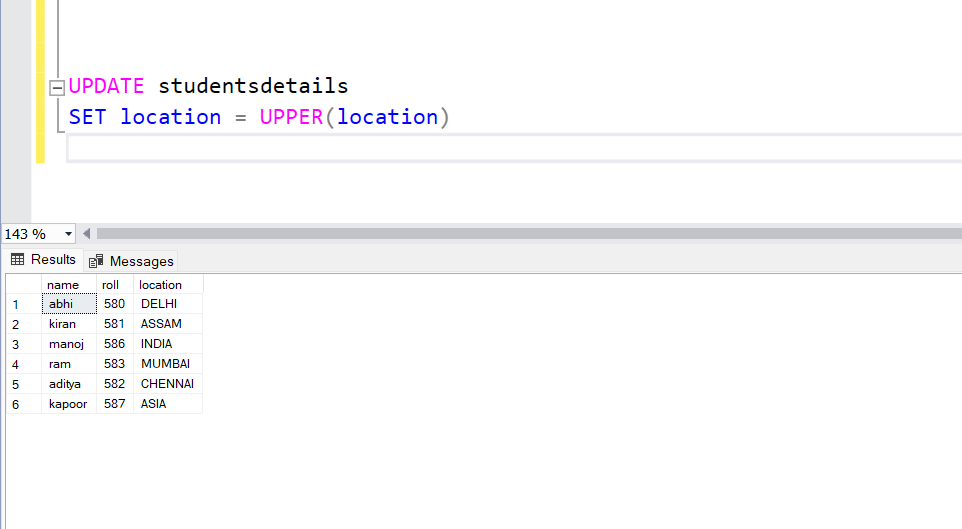


* We have deleted the data using delete command where the data is null.

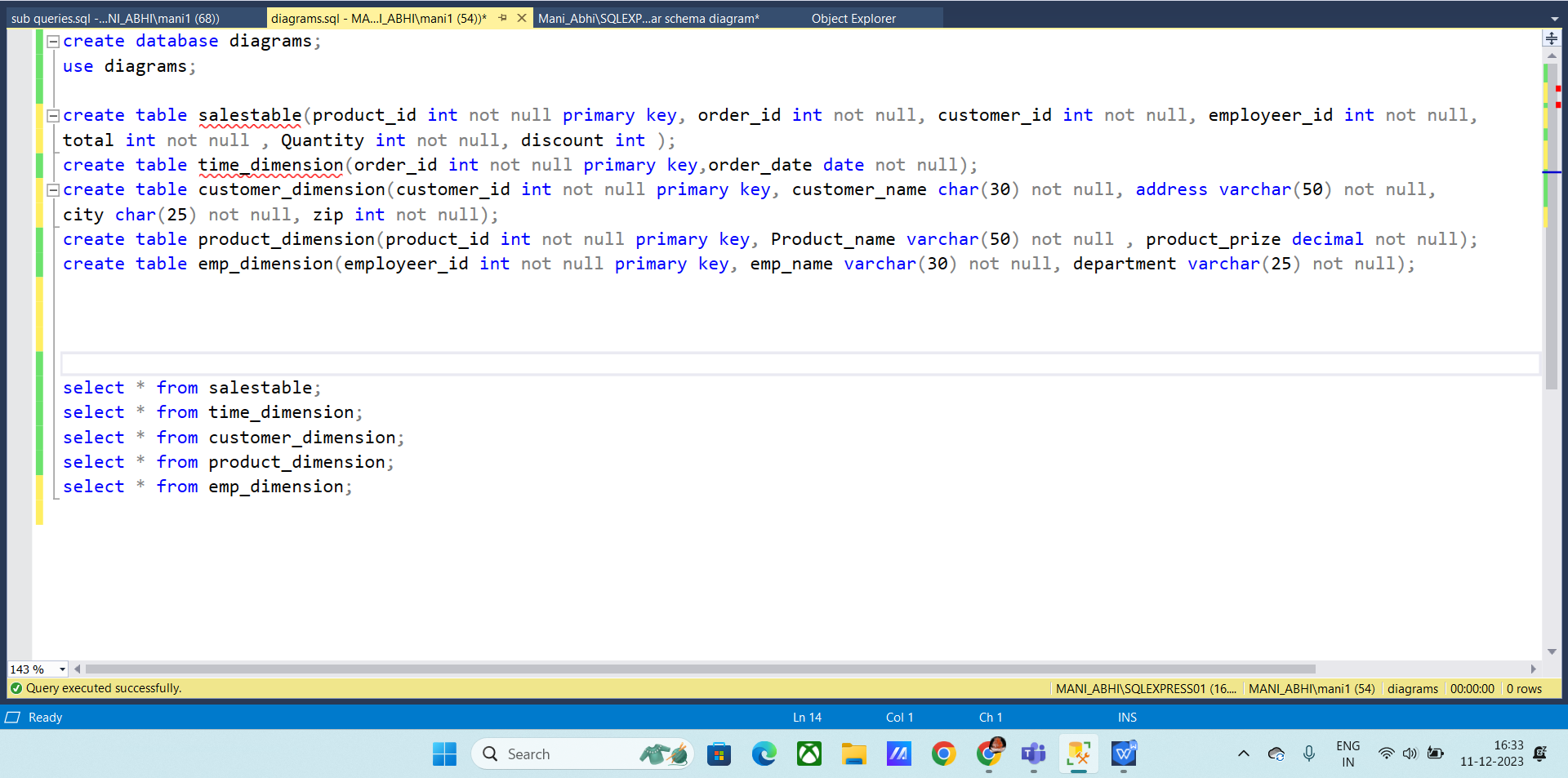


**Updating the data**





**Diagrams, star and snowflake schema diagrams**

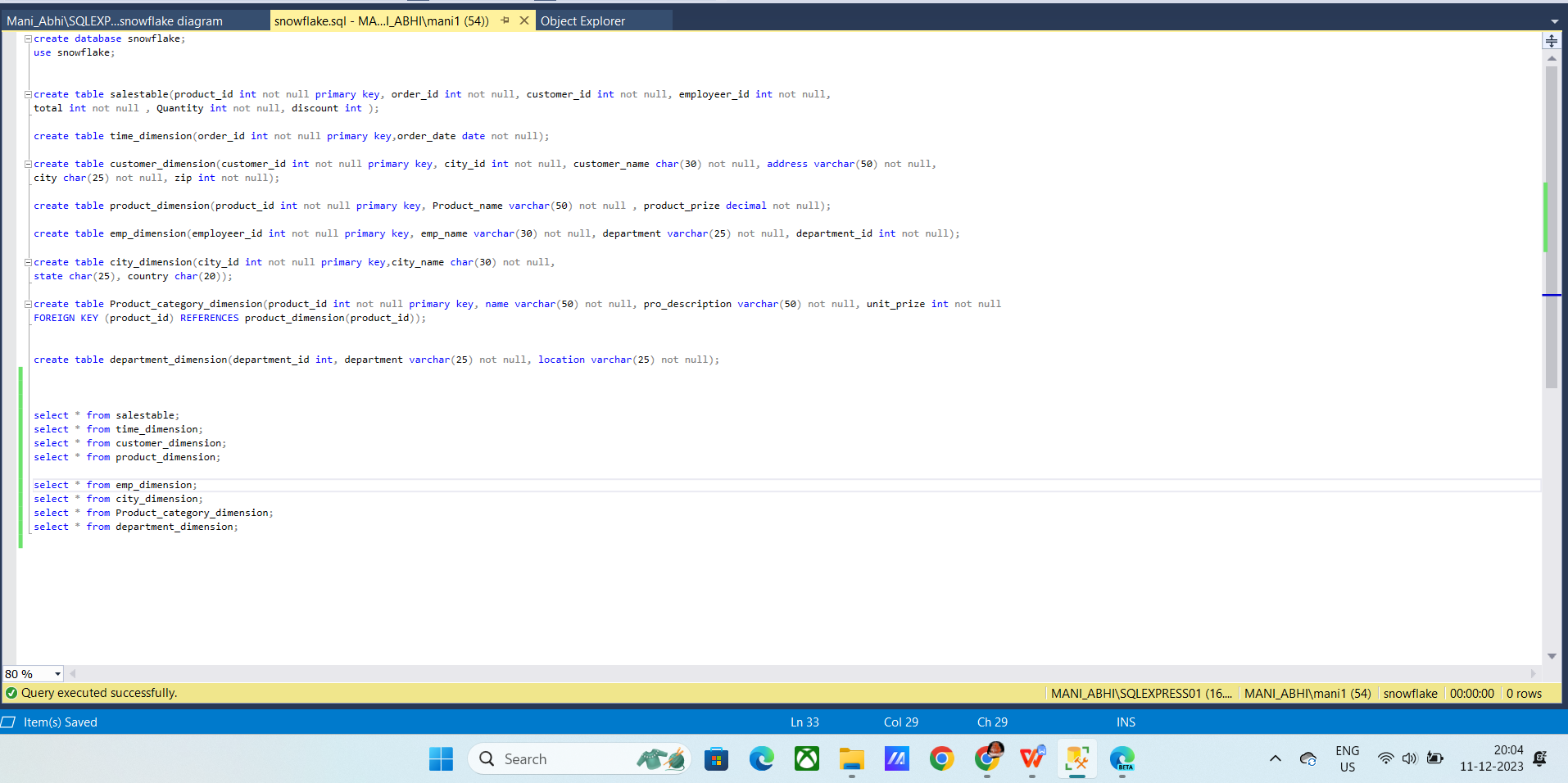




Normal diagrams for tables

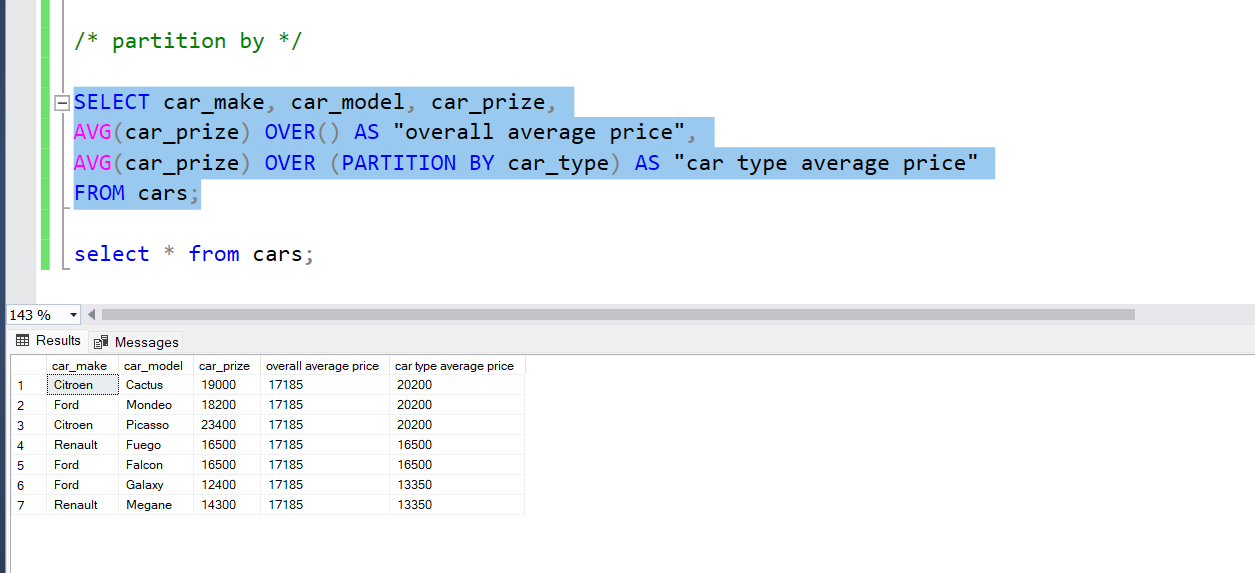


**Snow flake schema diagram:**

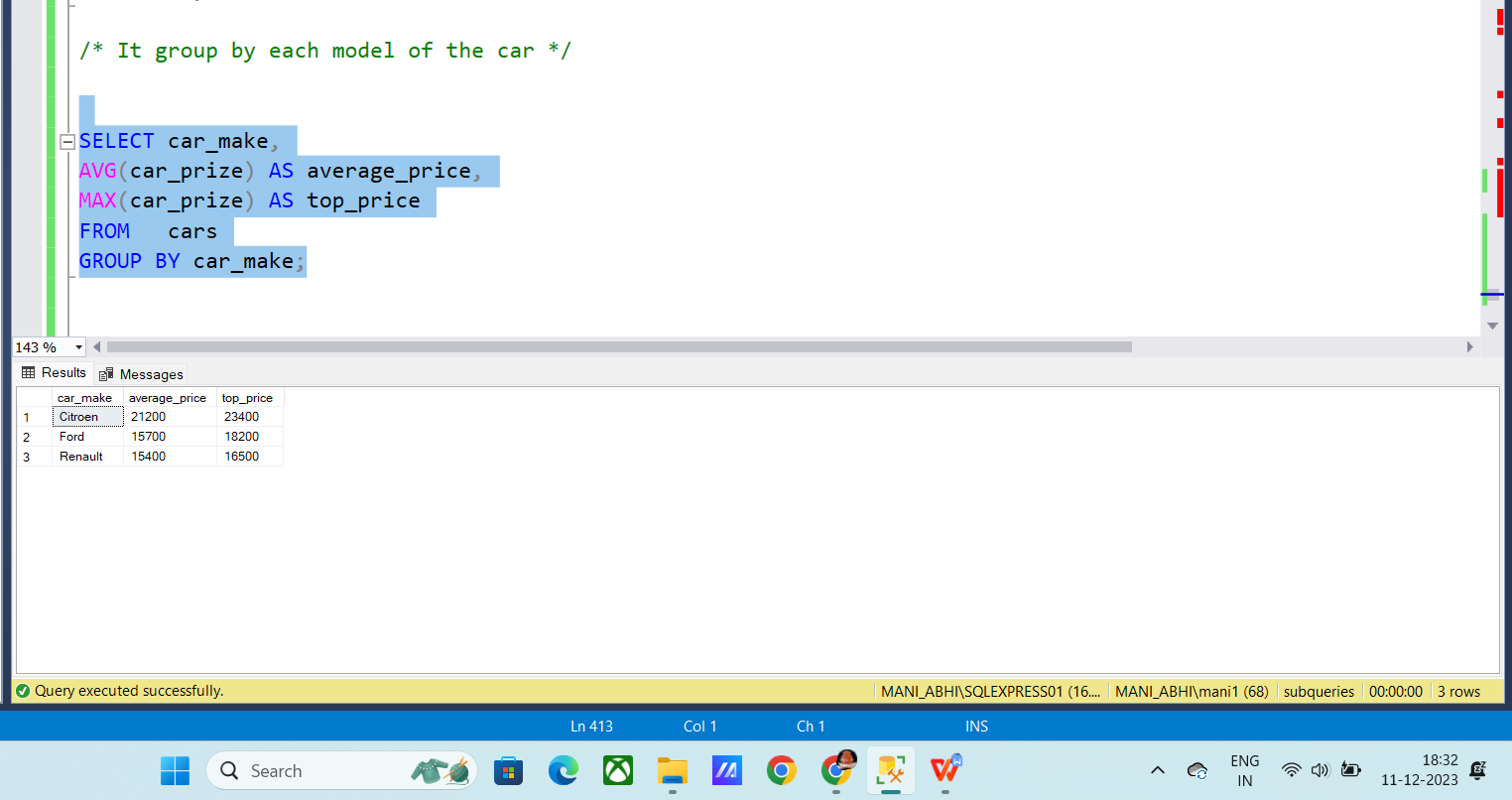




**Partition By :**



It group by each car\_make attribute of the car.



It will make groups of each mode.

