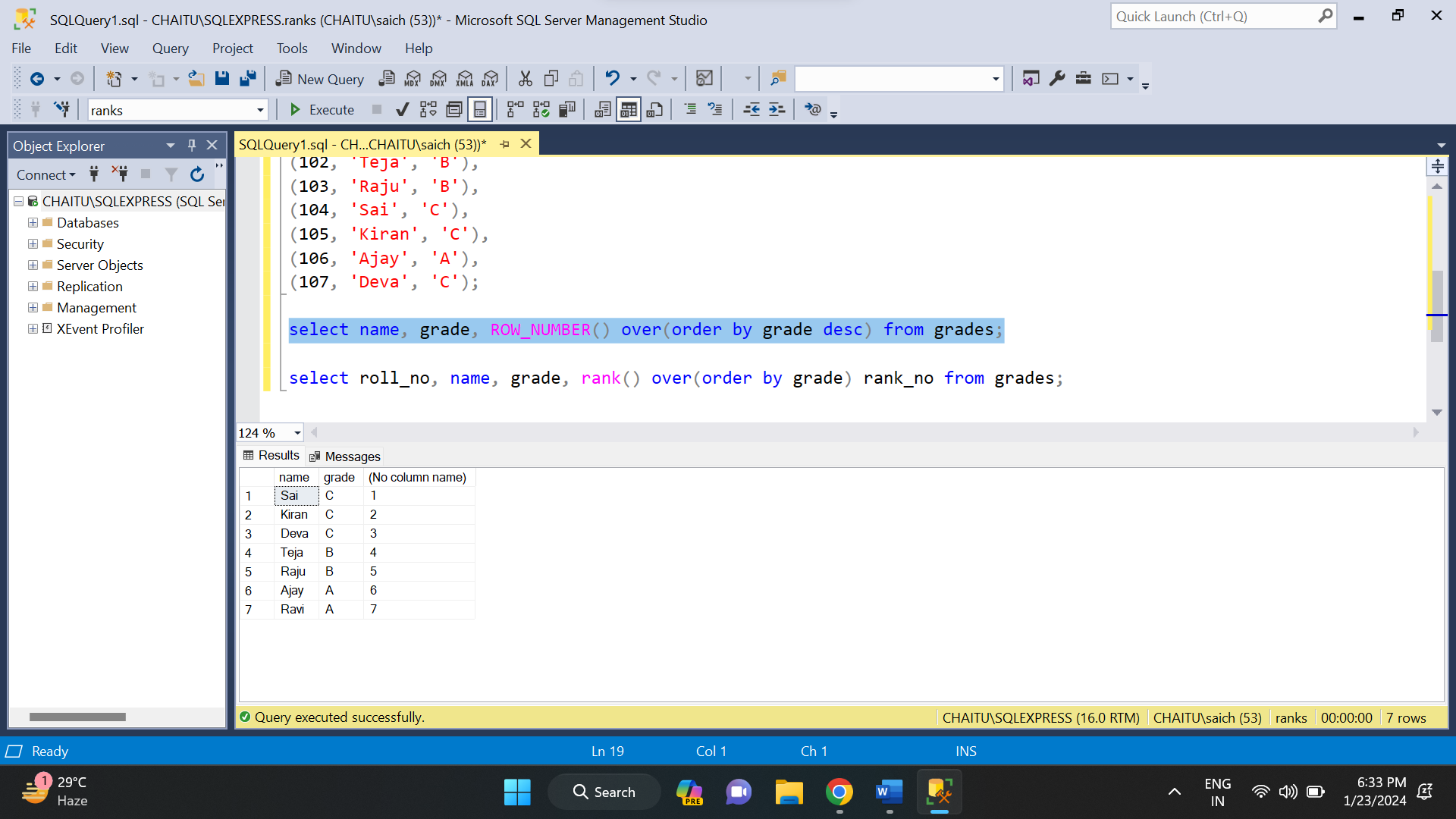
**Data Engineering Assignment – 5**

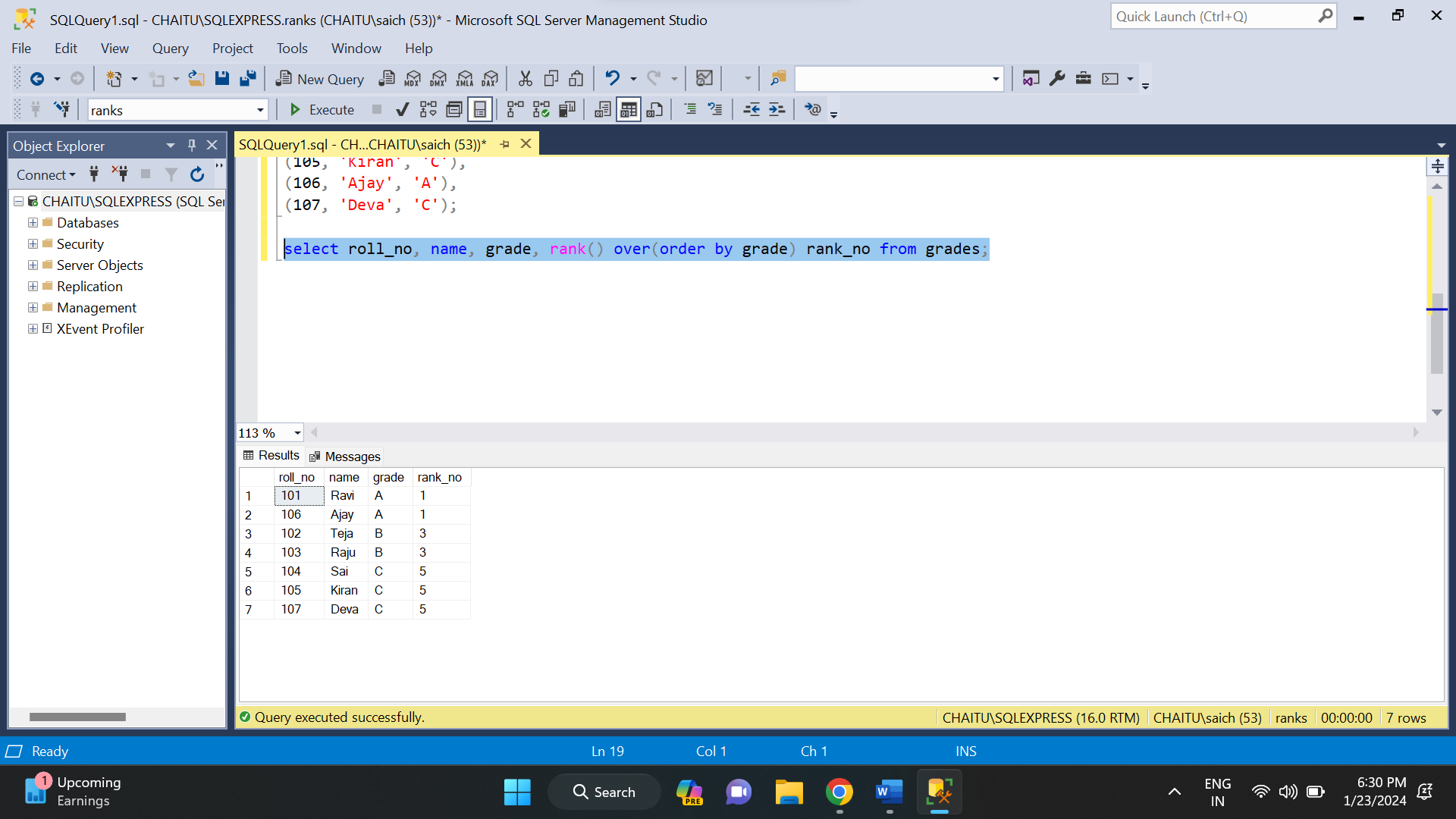
**Ranking Functions in SQL:**

There are various ranking functions available in SQL. They are:

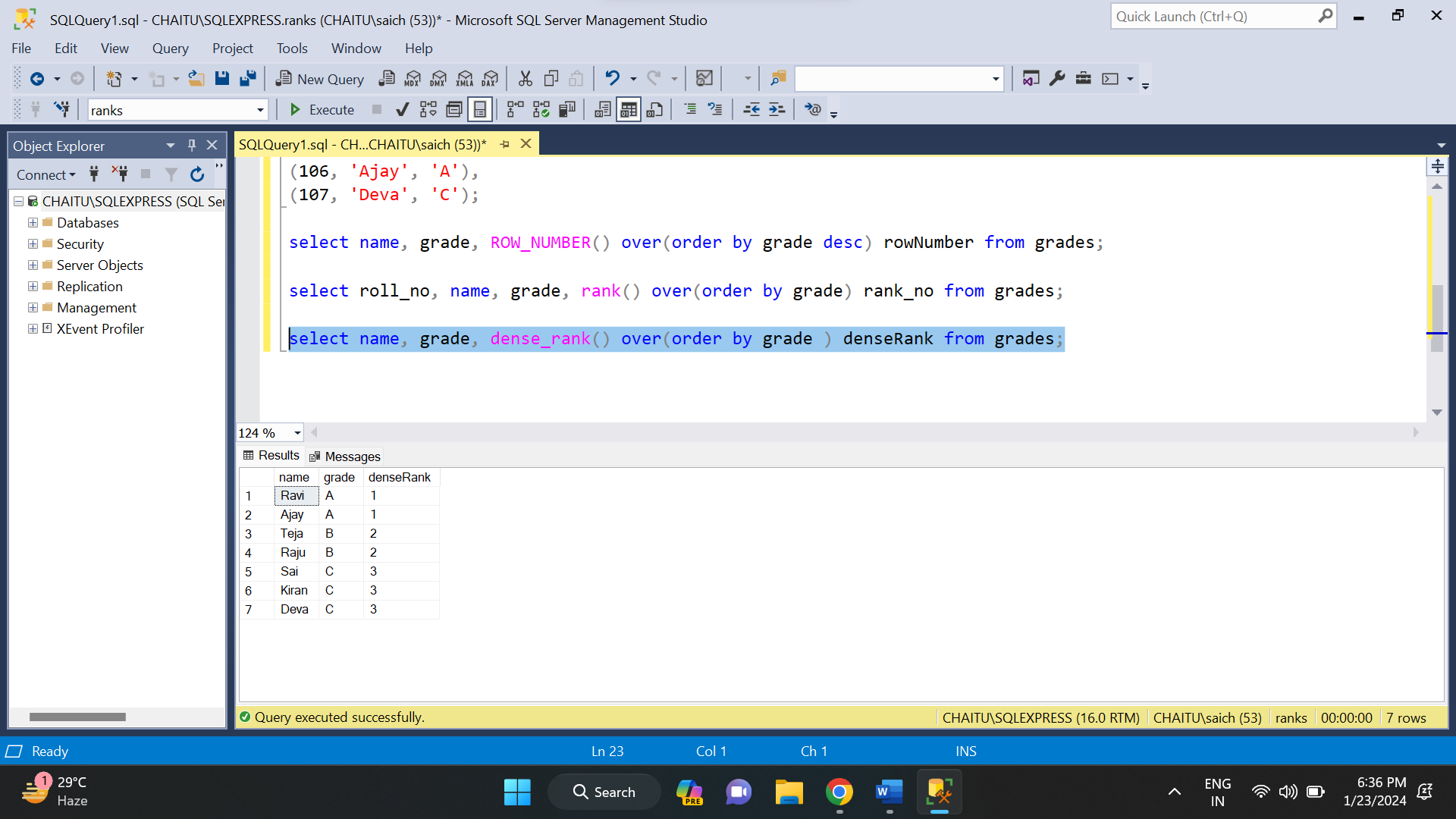
* **ROW\_NUMBER ():** This function assigns a unique number to each row starting with 1 and the order is specified by the order by clause.



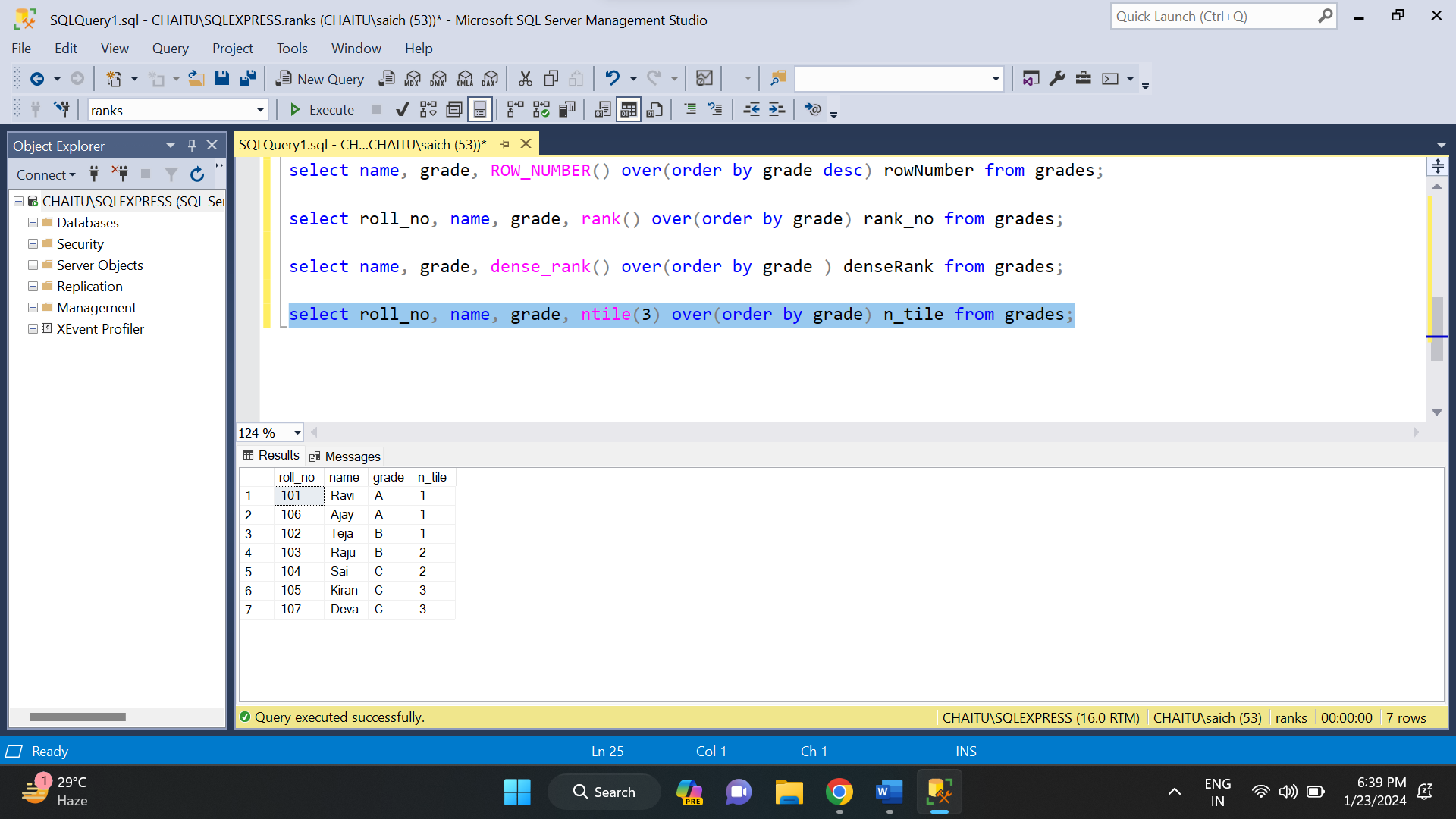
* **RANK ():** It assigns a unique rank to each distinct row, with ties receiving the same rank value and the order is specified by the order by clause.



* **DENSE\_RANK ():** This function returns the rank of each row within a result set partition, with no gaps in the ranking values. The rank of a specific row is one plus the number of distinct rank values that come before that specific row.

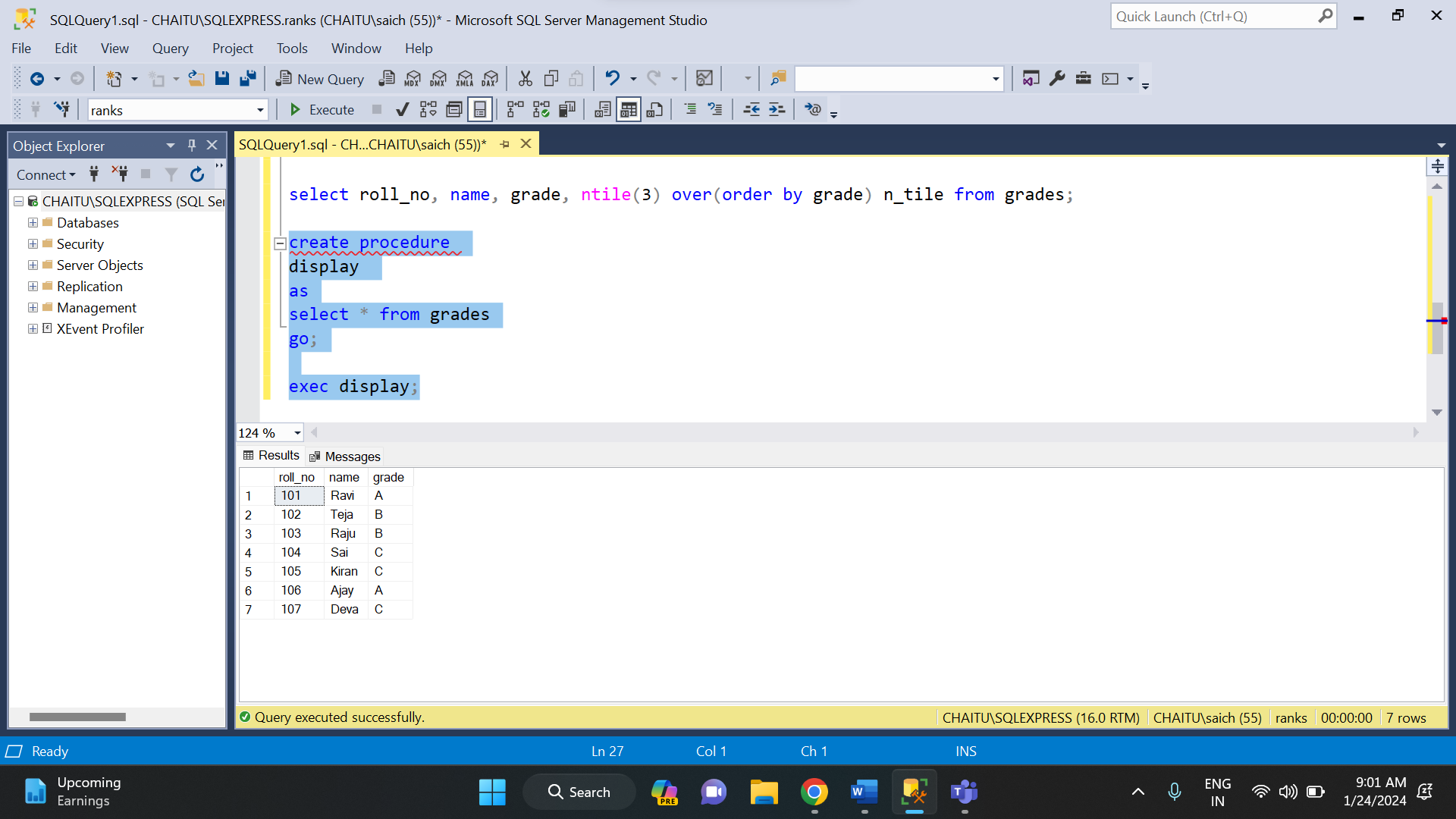


* **NTILE ():** It divides the result set into a specified number of roughly equal parts(tiles). The number of tiles is specified as an argument to the NTILE function.



**STORED PROCEDURES IN SQL:**

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. You can also pass parameters to a stored procedure, so that the stored procedure can act based on the parameter value(s) that is passed.



**DATA CLEANING AND TRANSFORMATION:**

Data cleaning and transformation is very essential. It is a process of identifying and removing duplicate and incorrect data. This help in making data more accurate and can improve the decision-making process and other applications related to data.

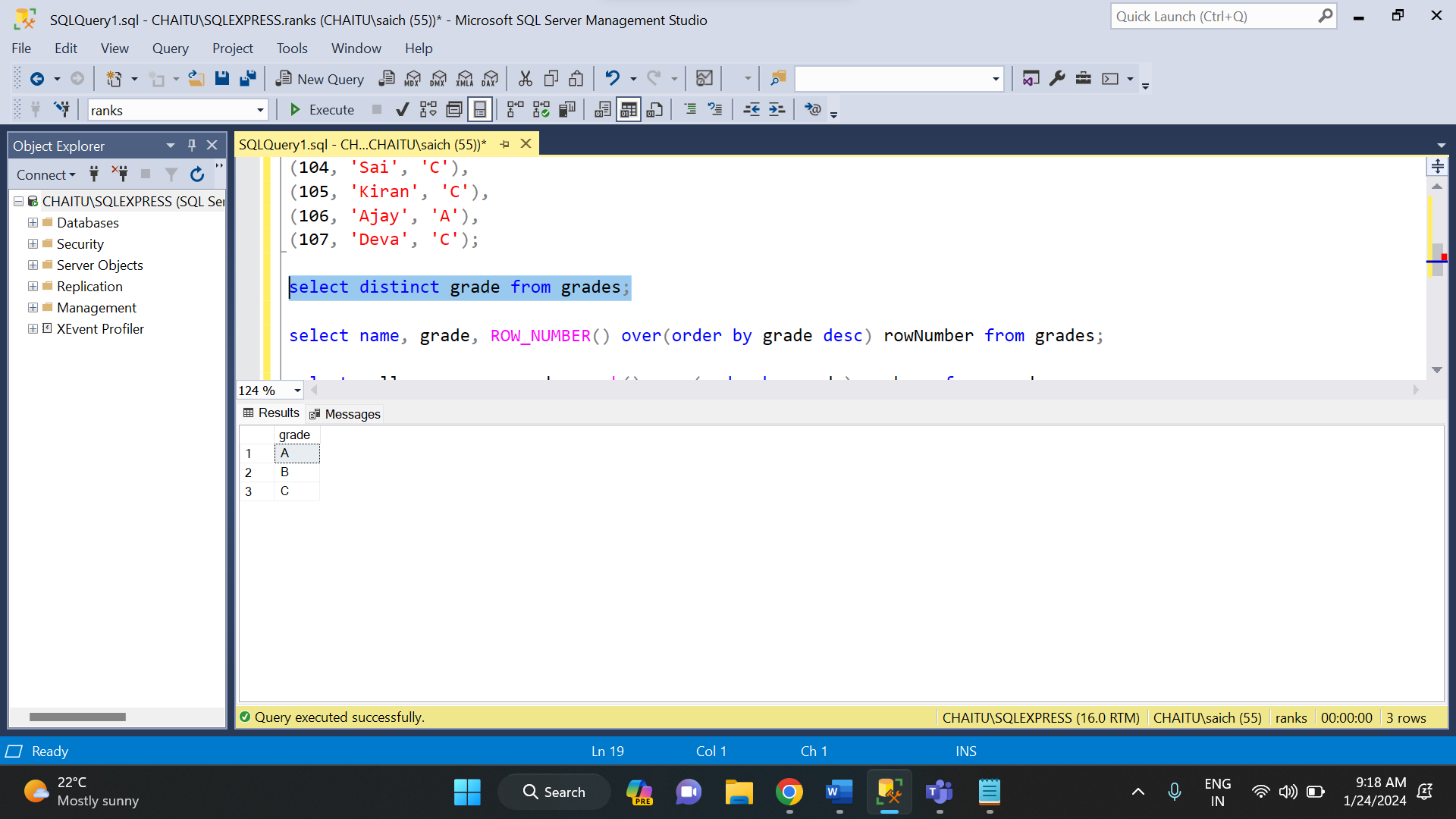
There are various steps involved in this data cleaning and transformation. They are:

* Identify null values
* Removing duplicate values
* Handling missing data
* Normalize data

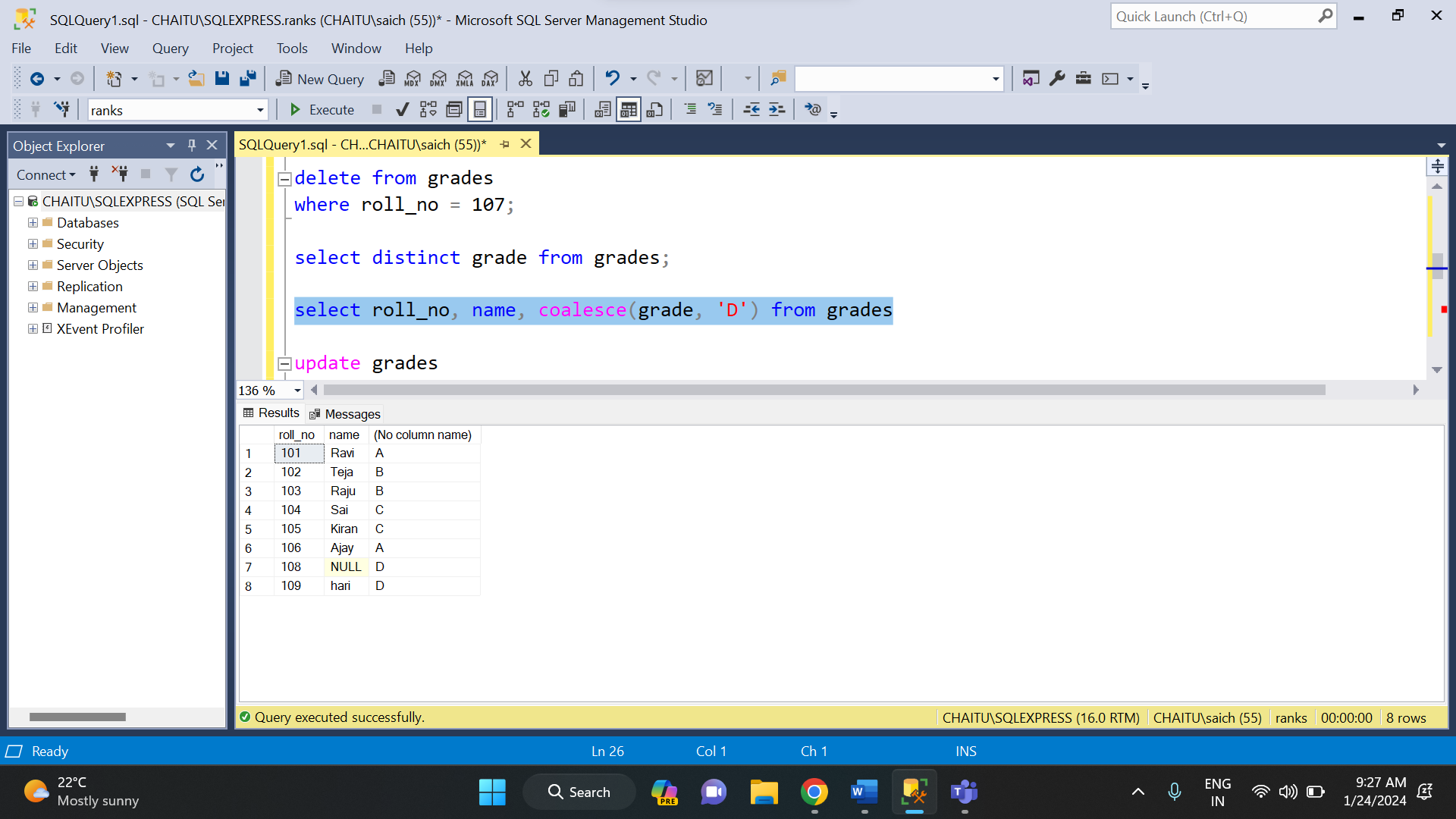
There are also several SQL statements related to this data cleaning and transformation. They are:

* DISTINCT
* COALESCE
* CASE WHEN
* UPDATE
* DELETE
* STRING FUNCTIOINS (CONCAT, TRIM, SUBSTRING etc.)

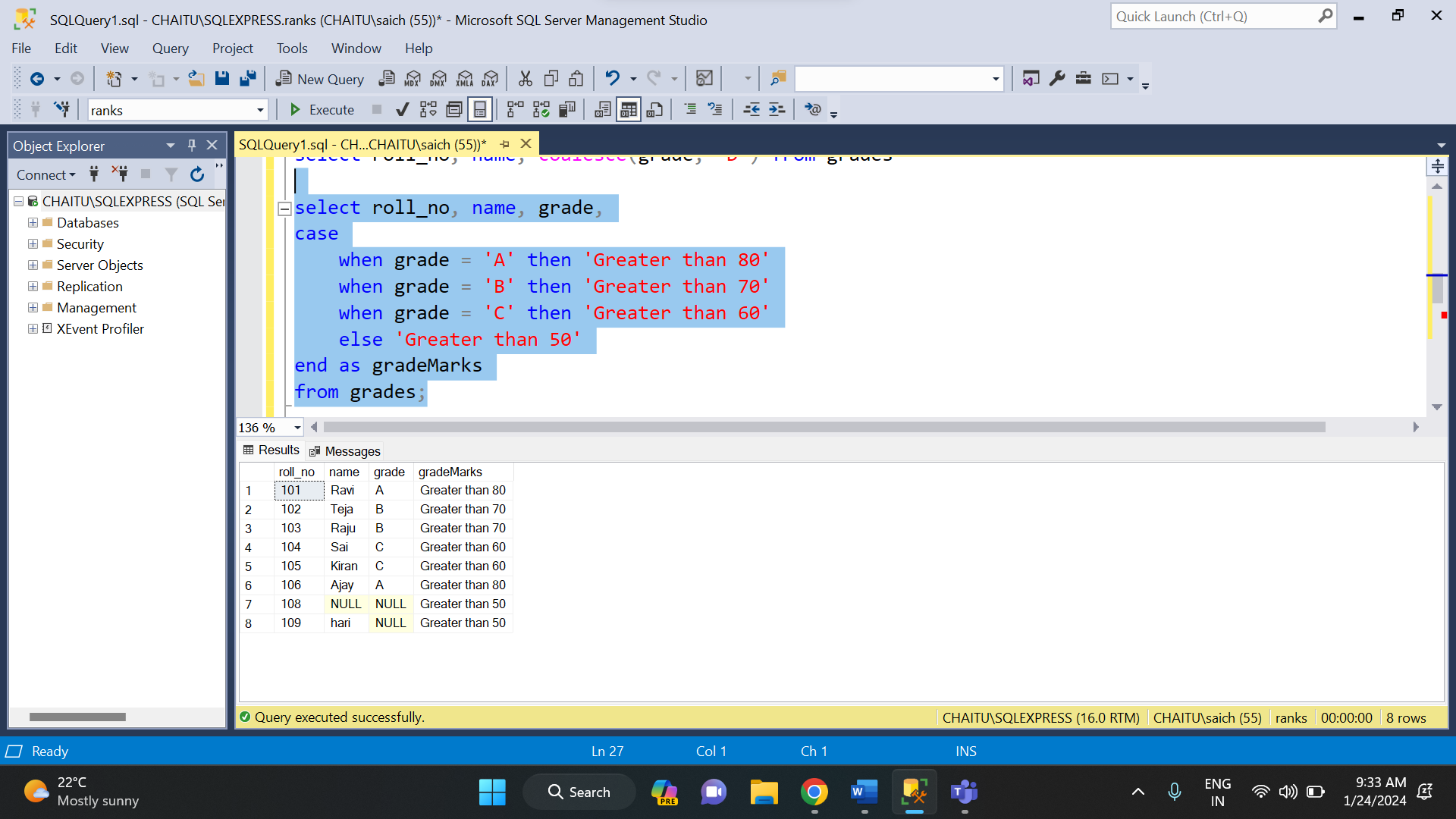
**DISTINCT:** This is very useful in identifying and removing duplicates.



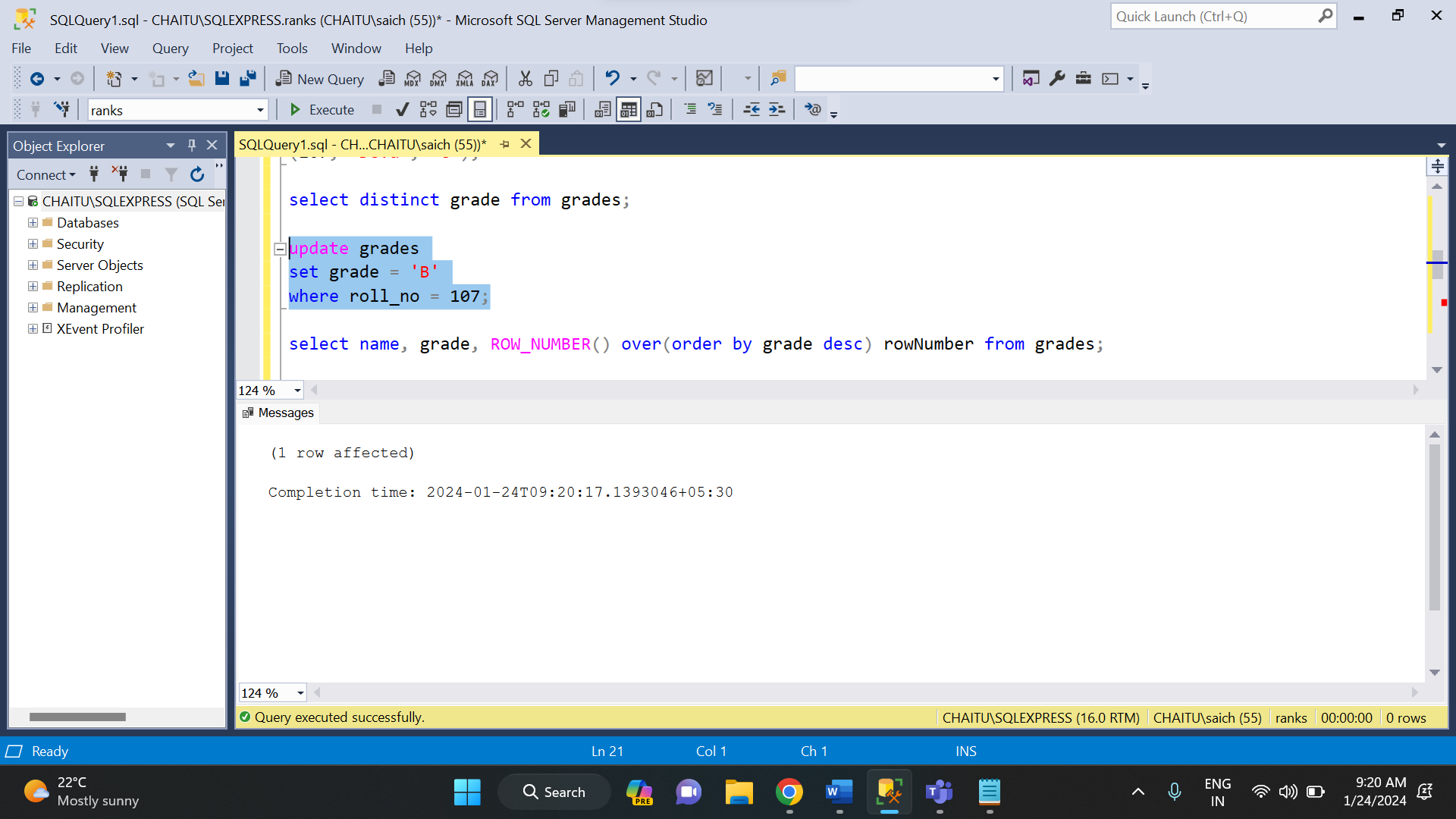
**COALESCE:** The COALESCE function in SQL is used to return the first non-null expression among its arguments. It is particularly useful for handling null values in a result set.



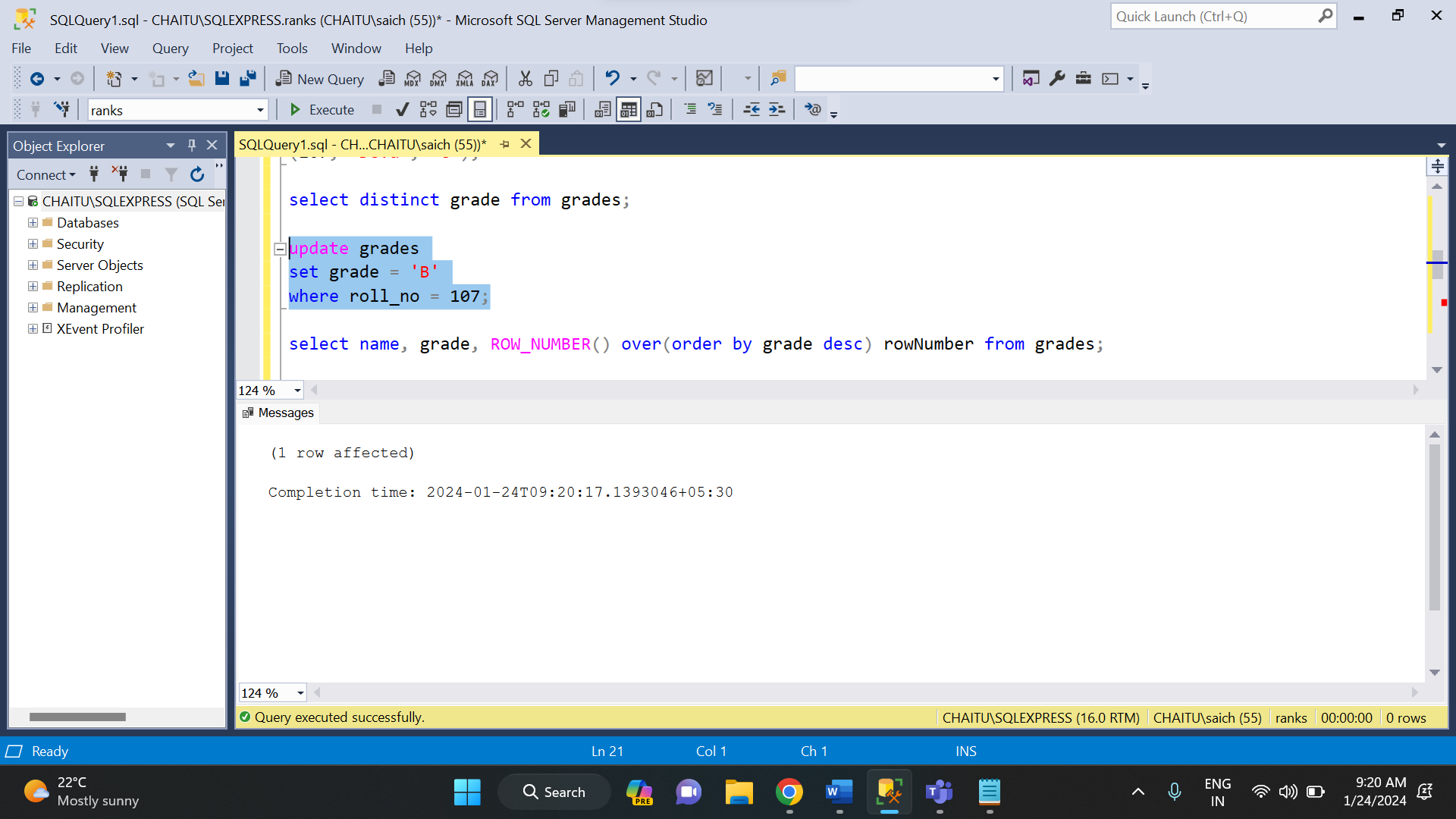
**CASE WHEN:** The CASE WHEN statement in SQL is used for conditional logic within a query. It allows you to perform different actions based on specified conditions.



**UPDATE:** This statement is used to update the values in the table.

****

**DELETE:** The delete statement is used to delete a one or more rows from a table.

****

All of the string have been specified in the day - 4 assignment.