**Name -Monika Gariya**

**Email-** [**monikagariya2023@gmail.com**](mailto:monikagariya2023@gmail.com)

**Data Engineering Batch 1**

**Date – 23-01-2024**

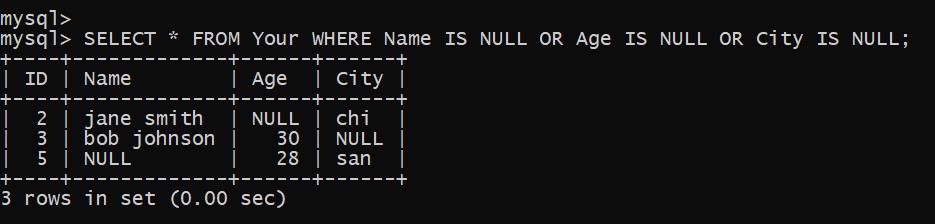
**Topic –** **Data Cleaning and Transformation ,Ranking Functions In SQL, Procedures**

**Data Cleaning and Transformation**

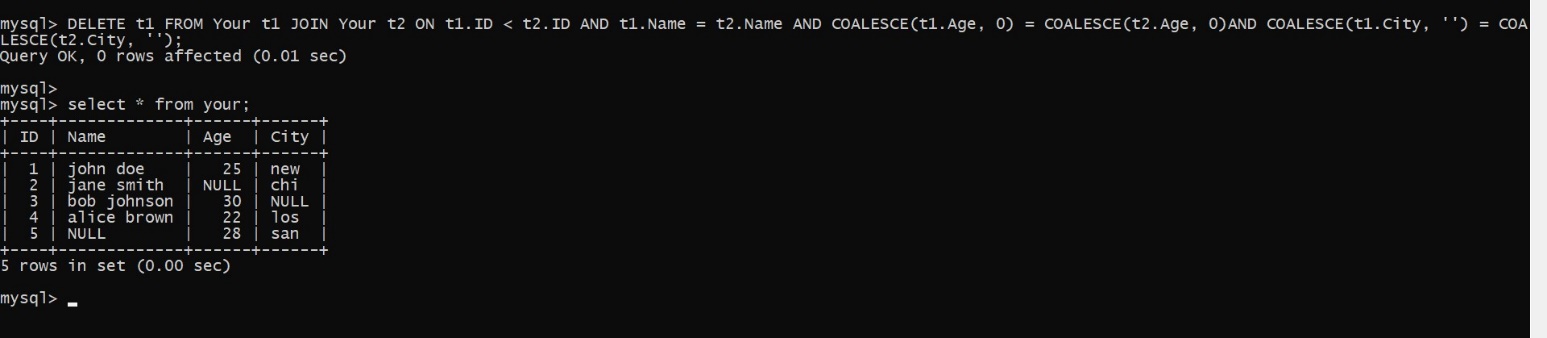
Data cleaning is a crucial process in any data . It involves identifying and correcting or removing inaccuracies, inconsistencies, and errors in the data.

**1)Check missing value**

a) is null

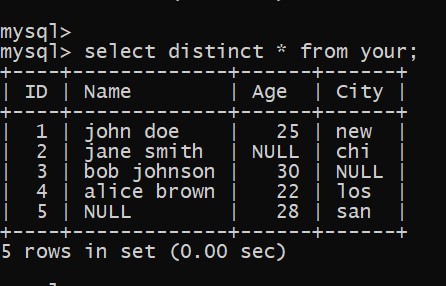


b) Colease



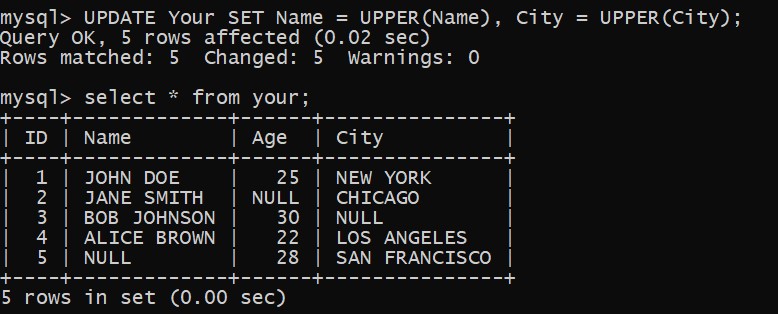
**2)Remove duplicate records**

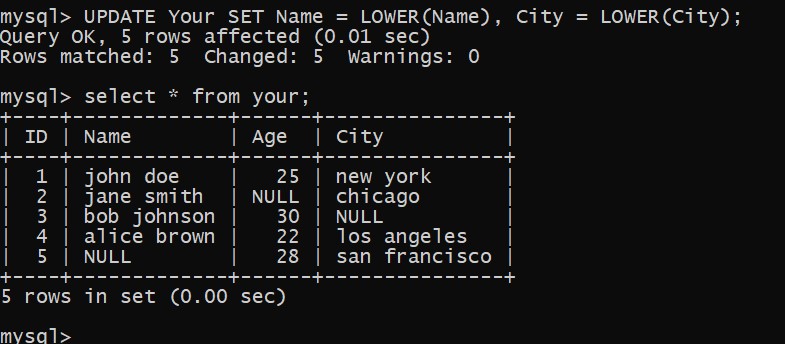
a) distinct



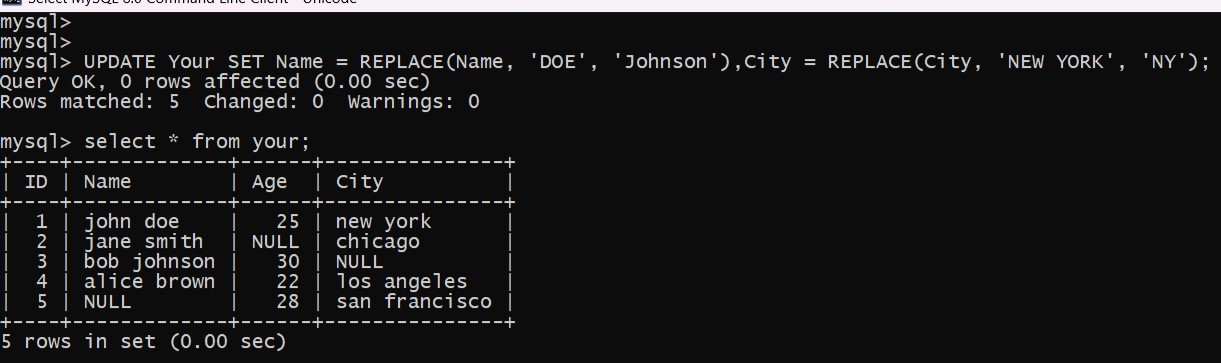
**3) Standardizing and transform data**

a) upper and lower

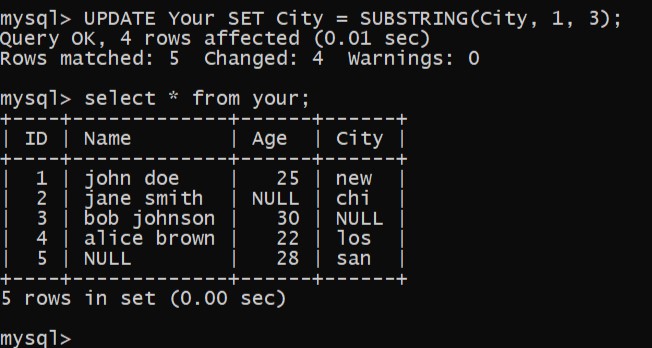




b) replace

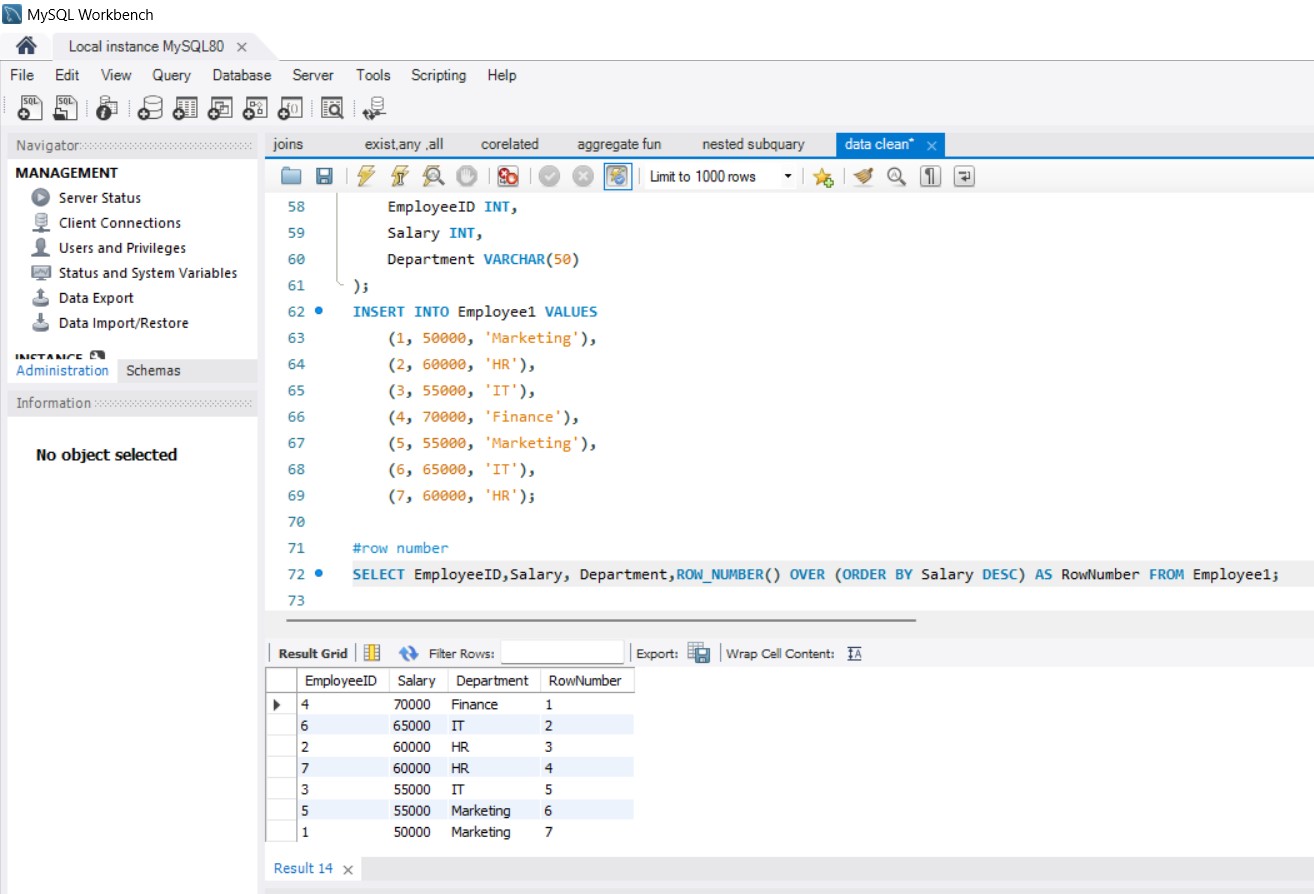


c)substring

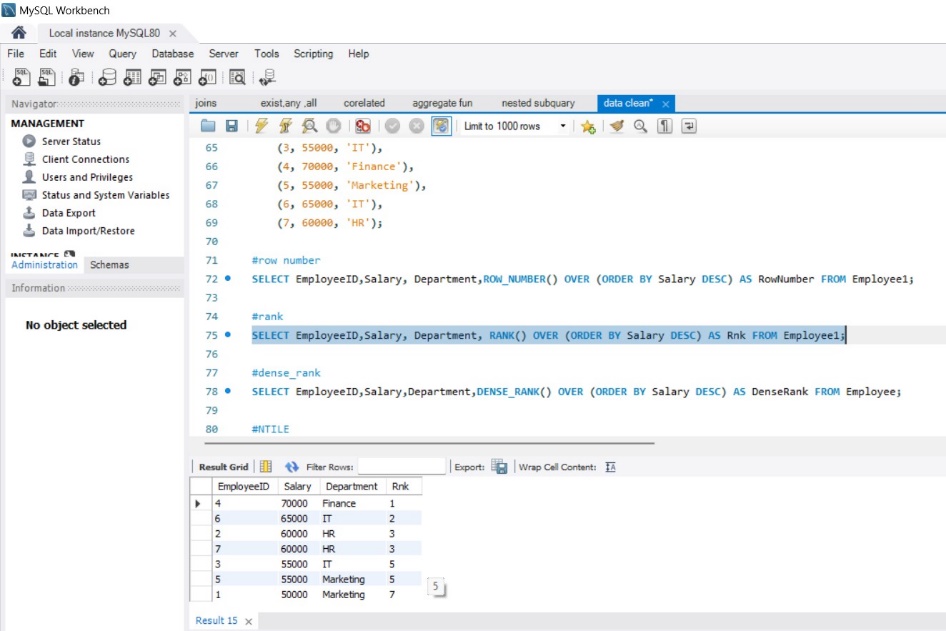


**SQL ranking functions**

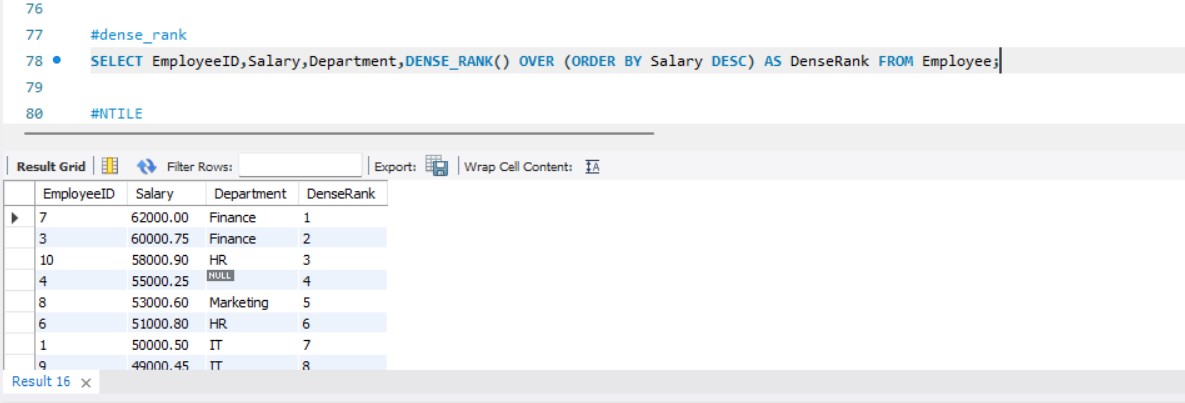
**1)ROW\_NUMBER** () – returns serial number of the row with possibility of division into parts (partitions) and sorting. It starts from 1 for every group.



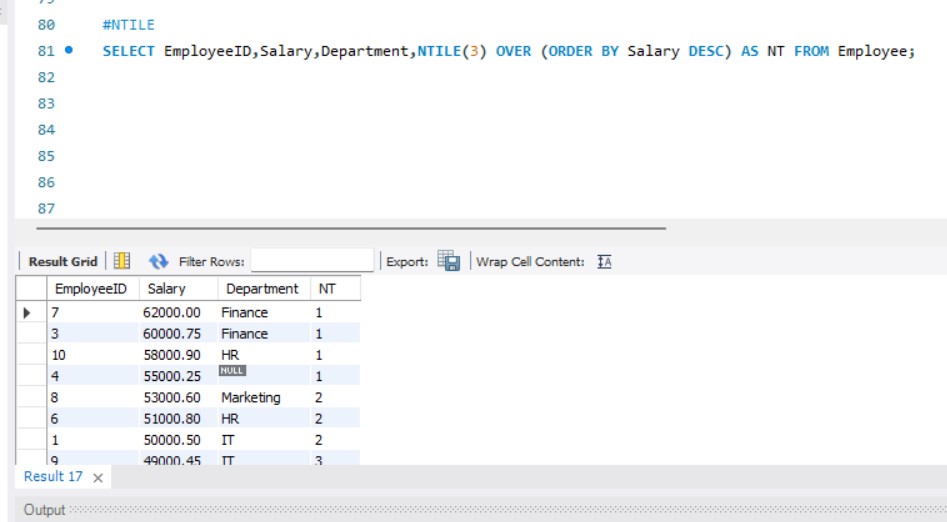
**2)RANK**() – returns the rank of every record part by part (partitions)



**3)DENSE\_RANK**() – returns the rank of every record part by part (partitions) without gaps between ranks



**4)NTILE** () – divides rows into n groups depending on the argument



**STORED PROCEDURE**

A stored procedure is a **group of one or more pre-compiled SQL statements**into a logical unit. It is stored as an object inside the database server. It is a subroutine or a subprogram in the common computing language that has been created and stored in the database. Each procedure in SQL Server always **contains a name, parameter lists, and Transact-SQL statements**.

