**Assignment 5**

**Name:** Jeevan Sai Badana

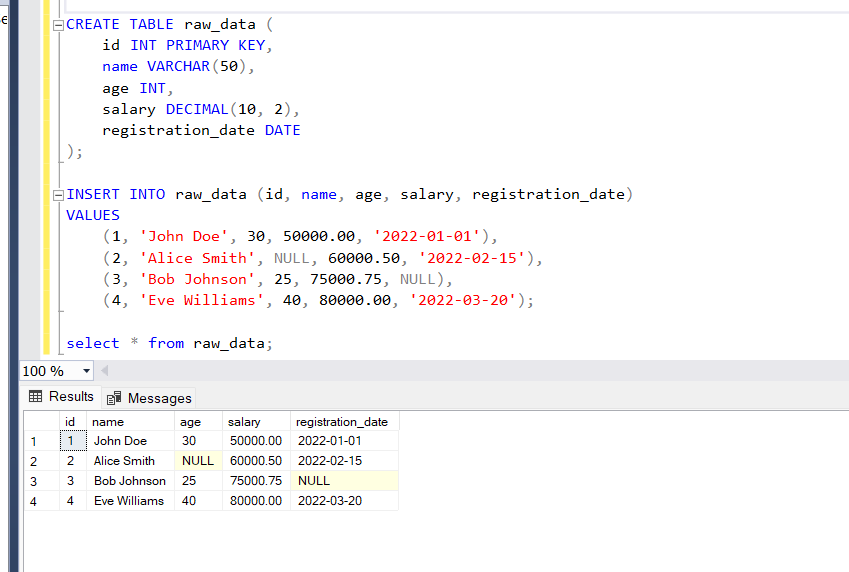
**Mail:** jeevansai100@gmail.com

**Date:** 23-01-2024

Topic: Data cleaning and transformation

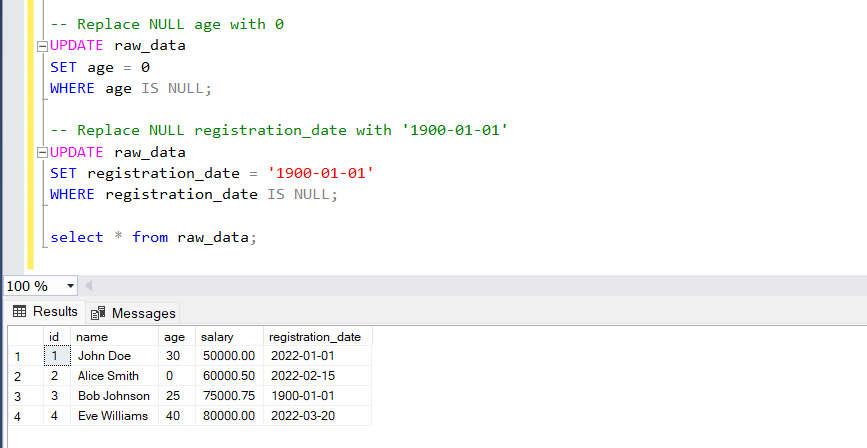
**Data cleaning & Transformation queries :**

For this purpose let’s create one raw table and insert some values into the table



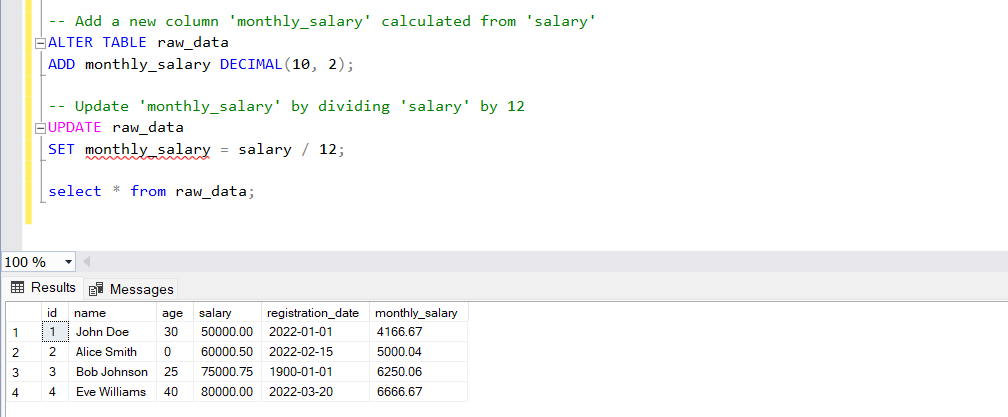
**1.Handling Missing Values:**

Replace NULL values with default values here in the table replace null with 0 and date with 1900-01-01

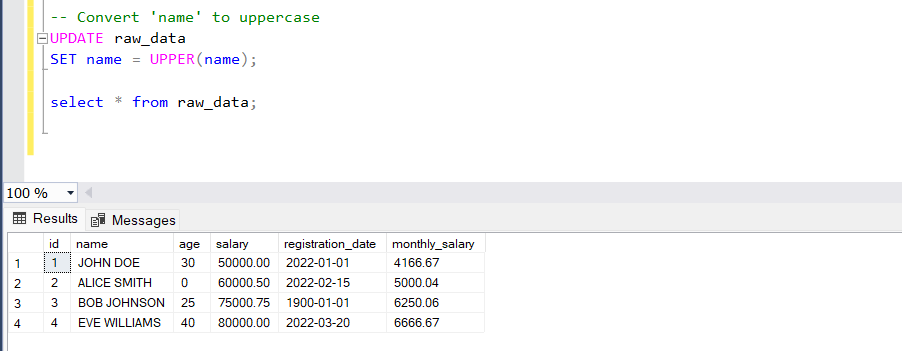


**2. Data Transformation:**

Calculating Derived Columns

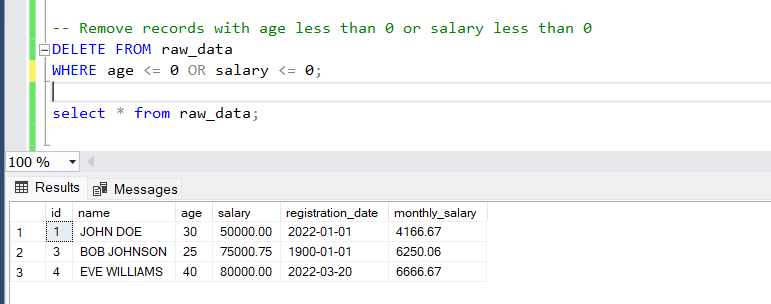


Standardizing Names to Uppercase



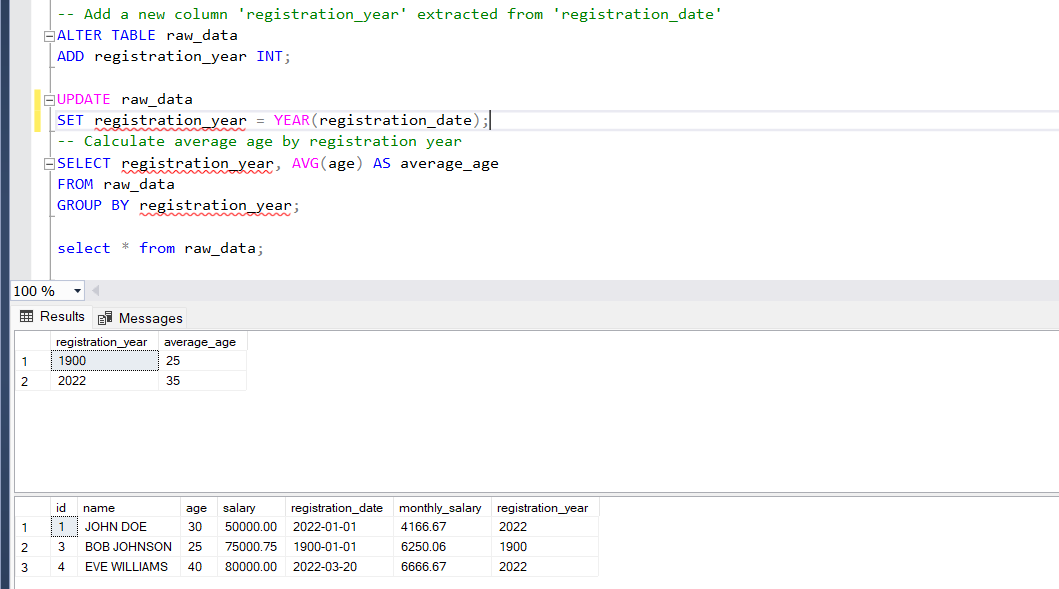
**3. Filtering Outliers:**

Removing Records with Invalid Data like age<=0 and salary<=0



4**. Aggregation and Grouping:**

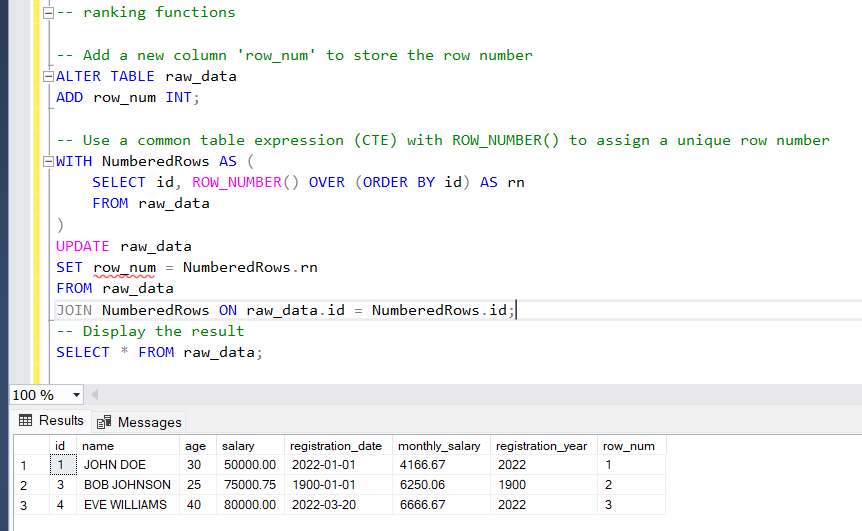
Calculate Average Age by Registration Year

****

**Ranking functions in sql**

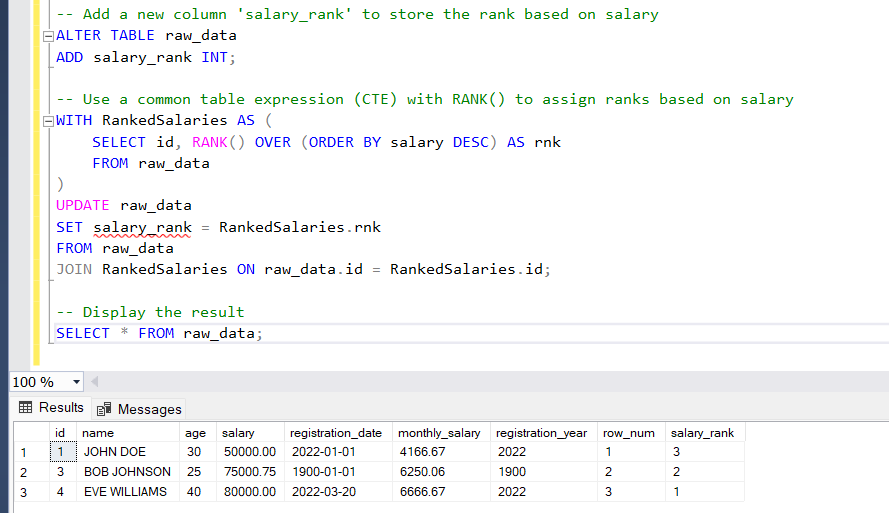
**ROW\_NUMBER():**

Definition: The ROW\_NUMBER() function is a window function that assigns a unique number to each row within a result set based on the specified ordering.



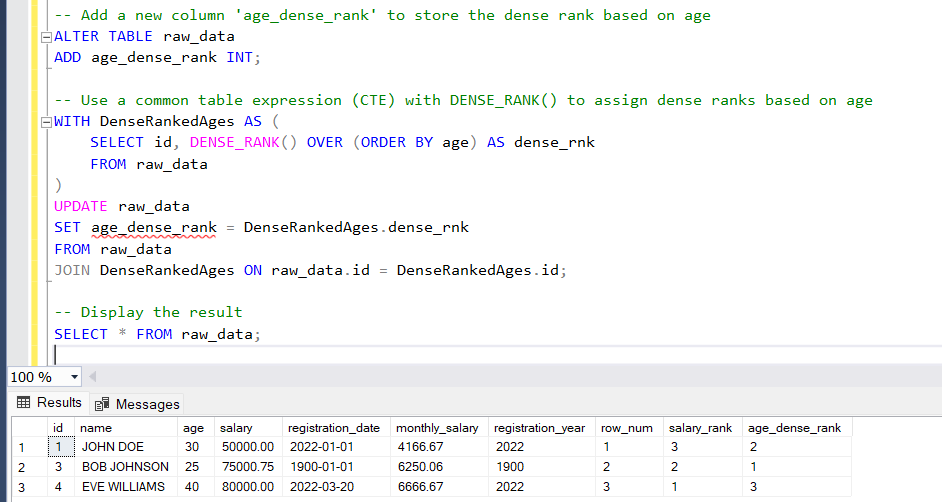
**RANK():**

Definition: The RANK() function is a window function that assigns a rank to each row within a result set based on the values in one or more columns.



DENSE\_RANK():

Definition: The DENSE\_RANK() function is a window function that assigns a dense rank to each row within a result set based on the values in one or more columns.



**Stored Procedure:**

Definition:

The ProcessRawData stored procedure is designed to perform data cleaning and transformation tasks on the raw\_data table. It encapsulates a series of SQL statements to handle missing values, transform data, filter outliers, and perform aggregation based on the characteristics of the raw data.

Benfits:

* Code Reusability: The stored procedure encapsulates a set of actions, making it reusable whenever similar data processing is required.
* Encapsulation: Complex data operations are encapsulated within the stored procedure, promoting cleaner and more modular code.
* Consistency: By centralizing data processing logic, the stored procedure ensures consistent application of data cleaning and transformation rules.
* Maintenance: Changes to the data processing logic can be made in one centralized location (the stored procedure) rather than scattered throughout the application code.
* Performance: Precompiled stored procedures can provide performance benefits compared to executing individual SQL statements separately.

