Customer Orders Data

Cleaning Project

Prepared By: SOURABH ARYA

Date: 27 December 2025

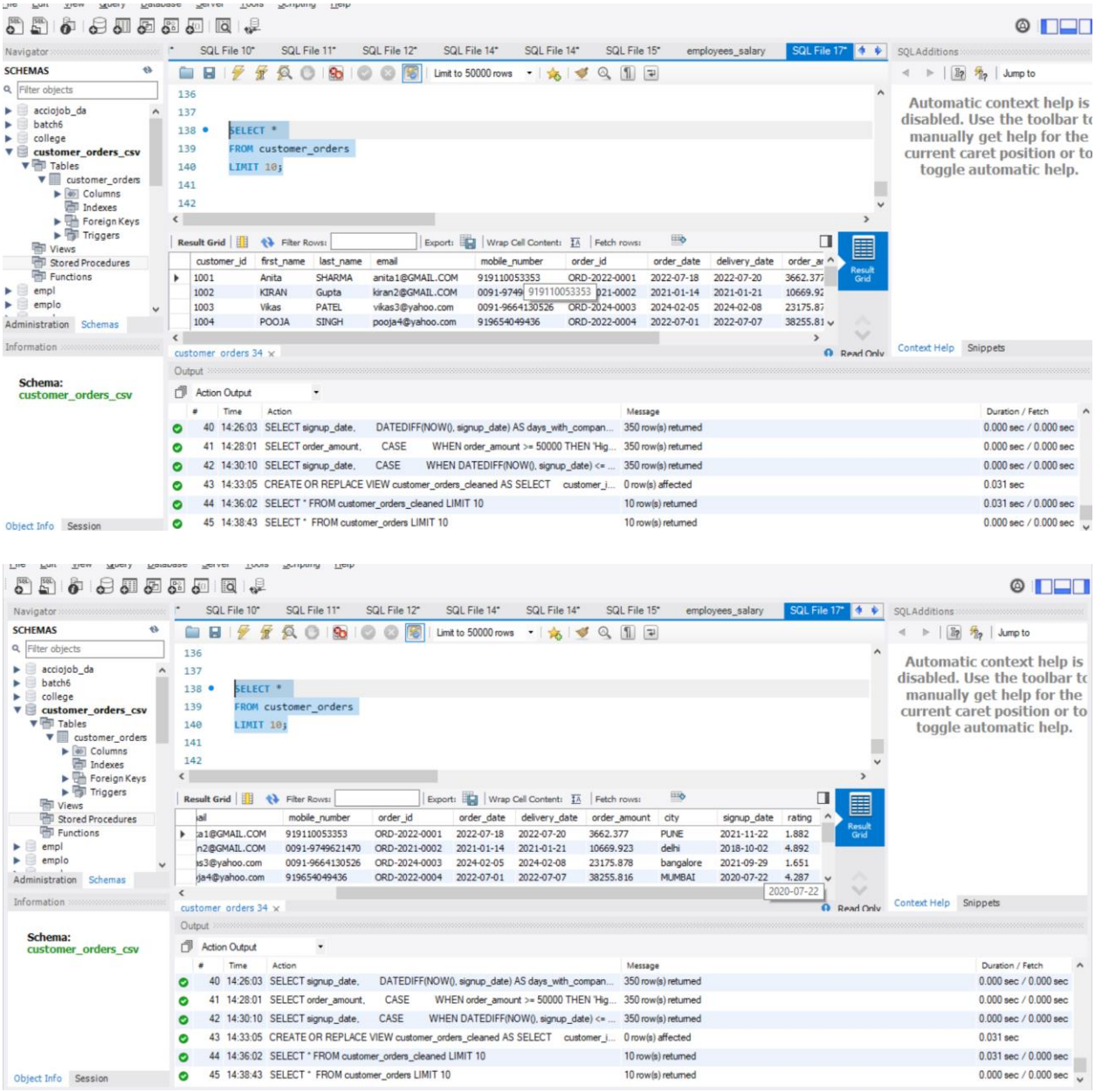
Tool Used: MySQL

Introduction

This project focuses on cleaning and standardizing a customer orders dataset. The goal was to ensure consistency across names, emails, mobile numbers, order details, ratings, and customer categories. A total of 15 steps were performed, and finally a cleaned view was created for analysis.

◆ STEP 0: Inspect Raw Data

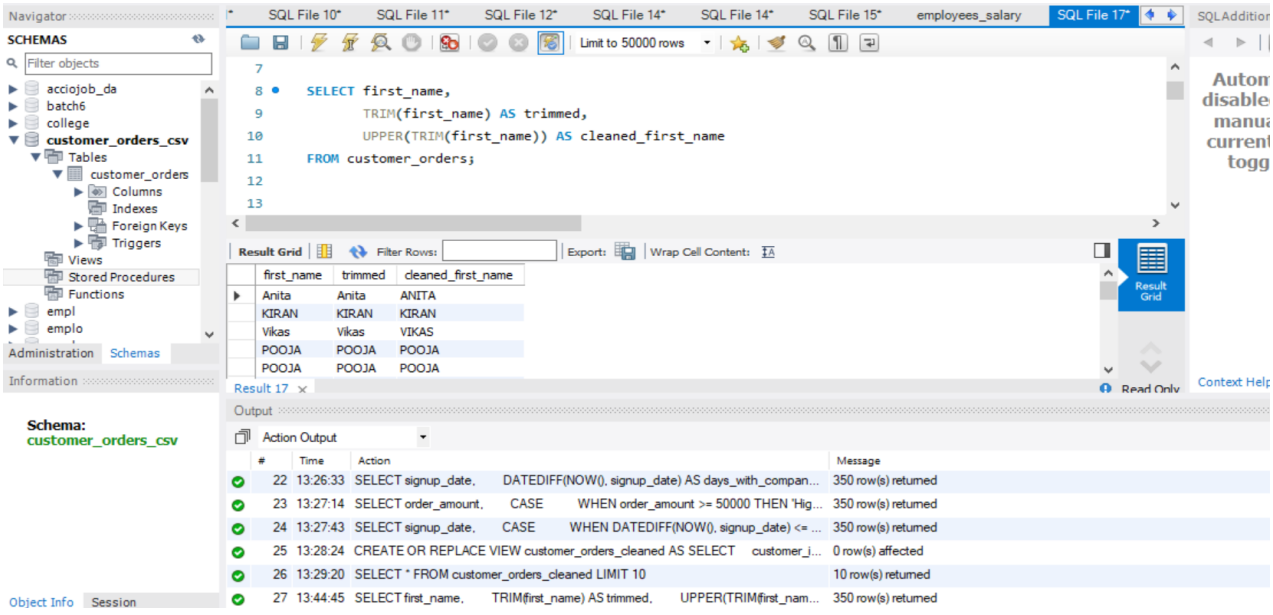
```
SELECT *
FROM customer_orders
LIMIT 10;
```



Explanation: **This step was performed to quickly view the raw dataset before applying any cleaning. By selecting the first 10 rows, we can understand the structure of the table, the columns available (like first_name, last_name, email, mobile_number, order_date, etc.), and identify issues such as inconsistent formatting, extra spaces, mixed cases, or irregular values. This inspection helps plan the cleaning steps systematically.**

◆ STEP 1: Clean `first_name` (Spaces + Case)

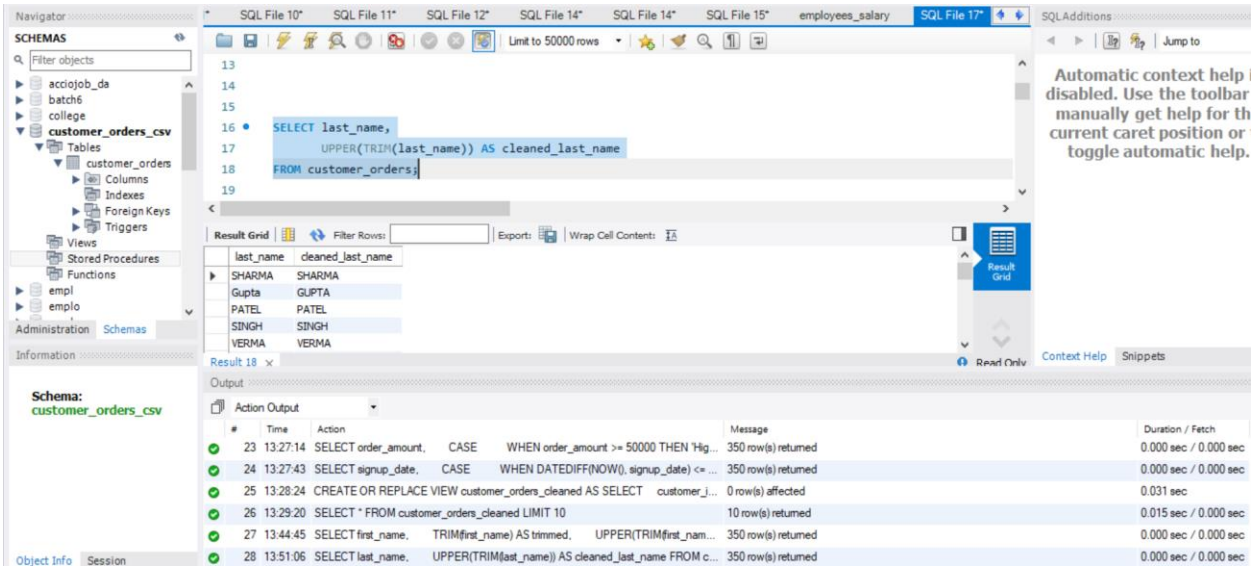
```
SELECT first_name, TRIM(first_name) AS trimmed,
UPPER(TRIM(first_name)) AS cleaned_first_name
FROM customer_orders;
```



Explanation: **Converted first names to uppercase and removed extra spaces**.

◆ STEP 2: Clean `last_name`

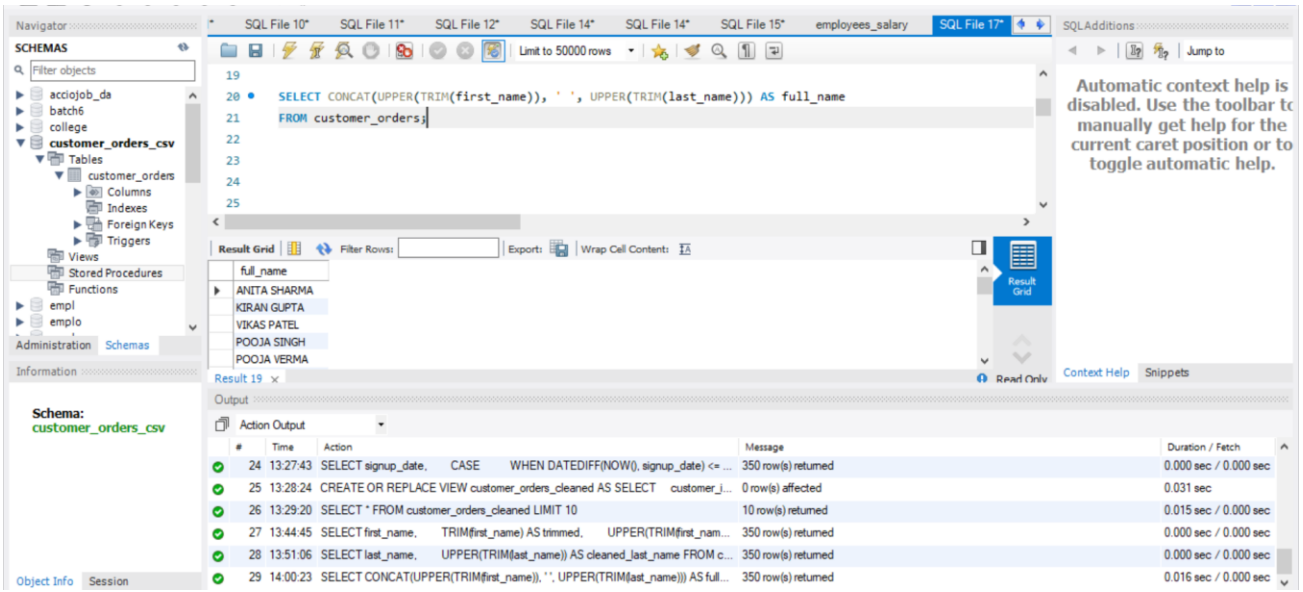
```
SELECT last_name,
       UPPER(TRIM(last_name)) AS cleaned_last_name
FROM customer_orders;
```



Explanation: **Converted last names to uppercase and removed extra spaces.**

◆ STEP 3: Create `full_name` (CONCAT)

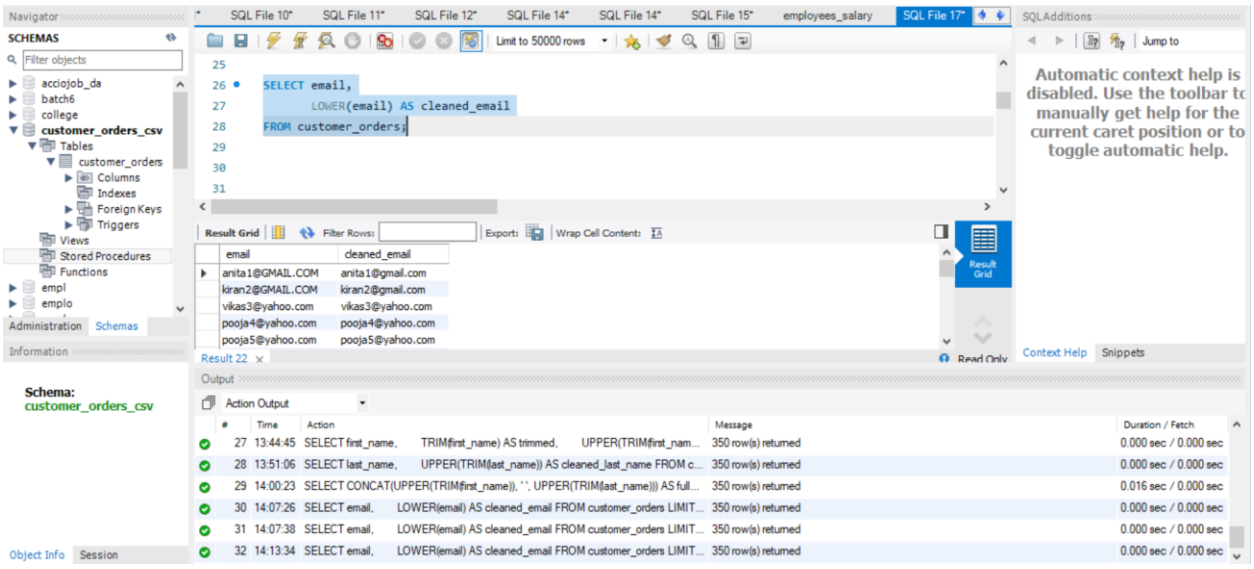
```
SELECT CONCAT(
UPPER(TRIM(first_name)),
' ',
UPPER(TRIM(last_name))
) AS full_name
FROM customer_orders;
```



Explanation: **Combined first and last names into a standardized full name.**

◆ STEP 4: Clean `email` (Standardization)

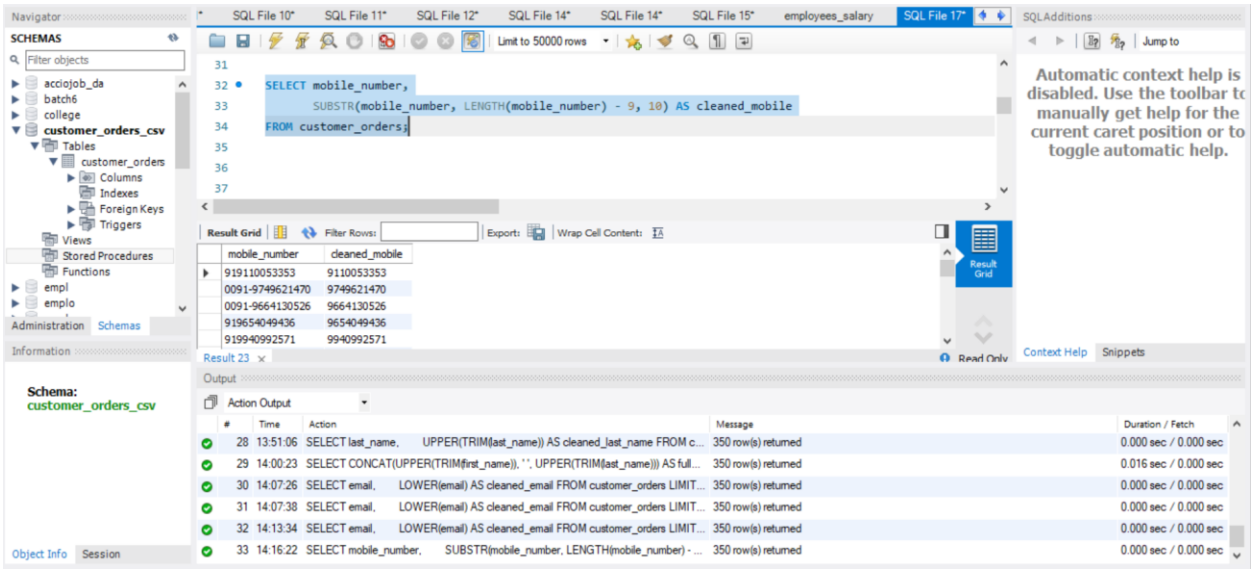
```
SELECT email,
       LOWER(email) AS cleaned_email
FROM customer_orders;
```



Explanation: **Converted all emails to lowercase.**

◆ STEP 5: Clean `mobile_number` (Extract last 10 digits)

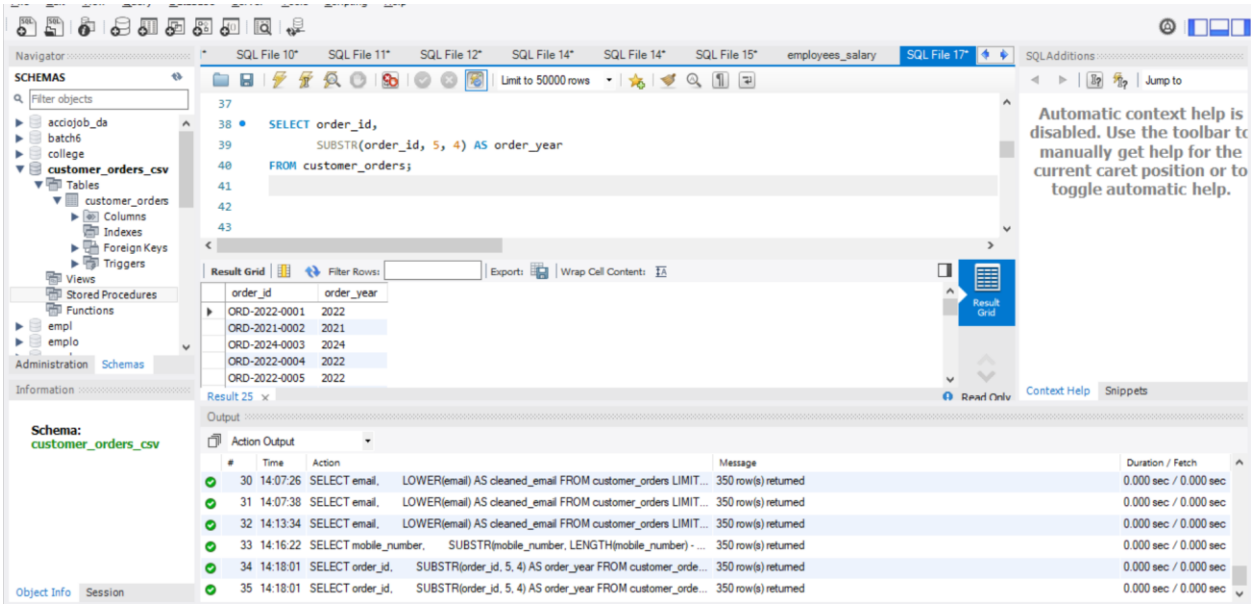
```
SELECT mobile_number,
       SUBSTR(mobile_number, LENGTH(mobile_number) - 9, 10)
AS cleaned_mobile
FROM customer_orders;
```



Explanation: **Extracted last 10 digits of mobile numbers.**

◆ STEP 6: Extract Year from `order_id`

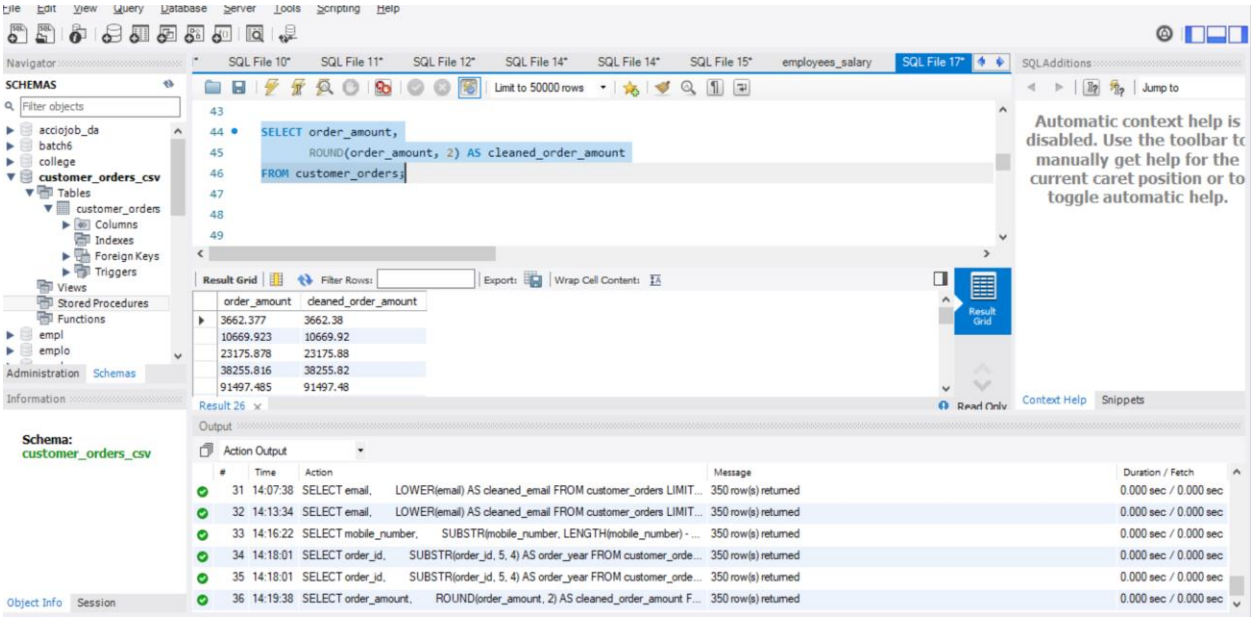
```
SELECT order_id,  SUBSTR(order_id, 5, 4) AS order_year
FROM customer_orders;
```



Explanation: **Extracted year from order ID.**

◆ STEP 7: Round `order_amount`

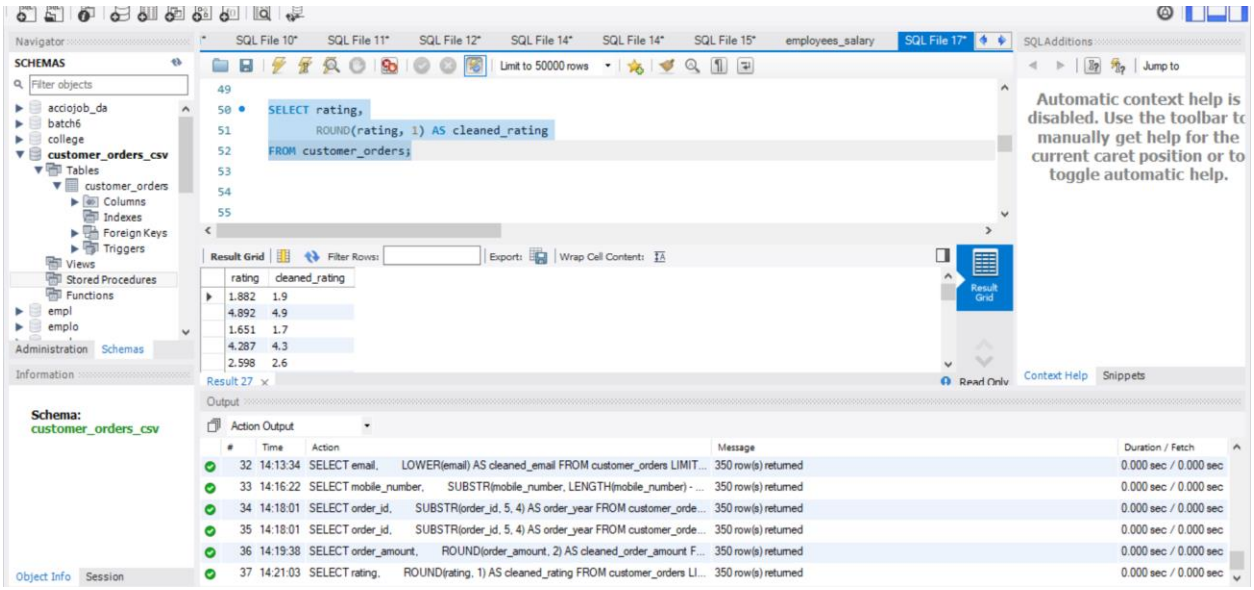
```
SELECT order_amount,
    ROUND(order_amount, 2) AS cleaned_order_amount
FROM customer_orders;
```



Explanation: **Rounded order amounts to 2 decimal places.**

◆ STEP 8: Round `rating`

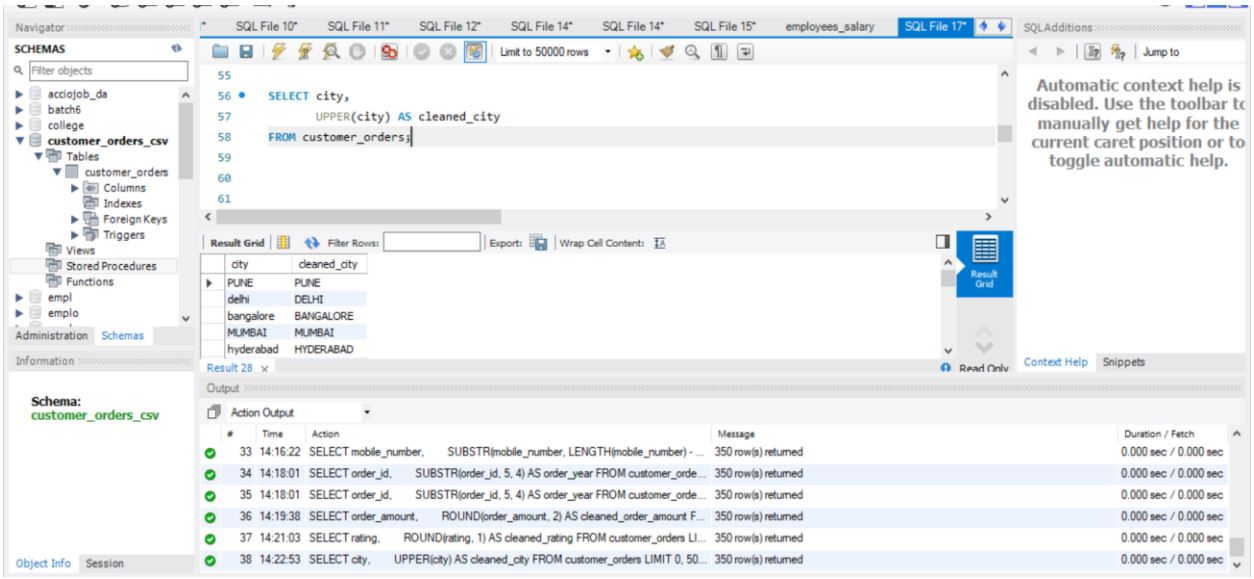
```
SELECT rating,
       ROUND(rating, 1) AS cleaned_rating
FROM customer_orders;
```



Explanation: **Rounded ratings to 1 decimal place.**

◆ STEP 9: Standardize `city`

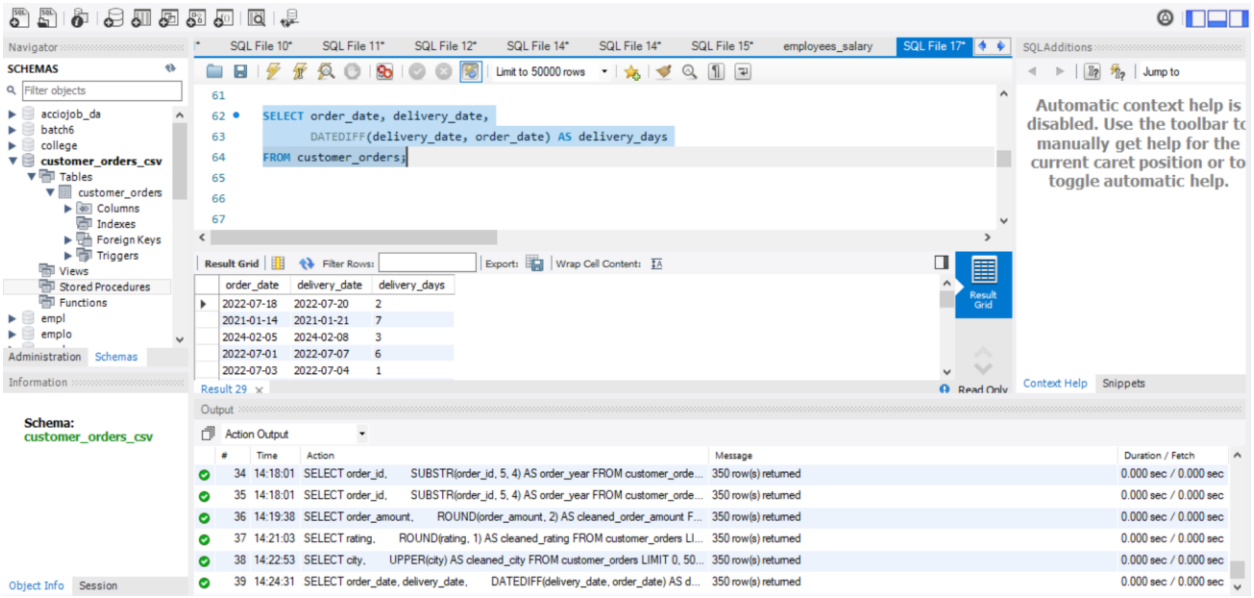
```
SELECT city,
       UPPER(city) AS cleaned_city
FROM customer_orders;
```



Explanation: **Converted city names to uppercase.**

◆ STEP 10: Delivery Time Calculation (DATEDIFF)

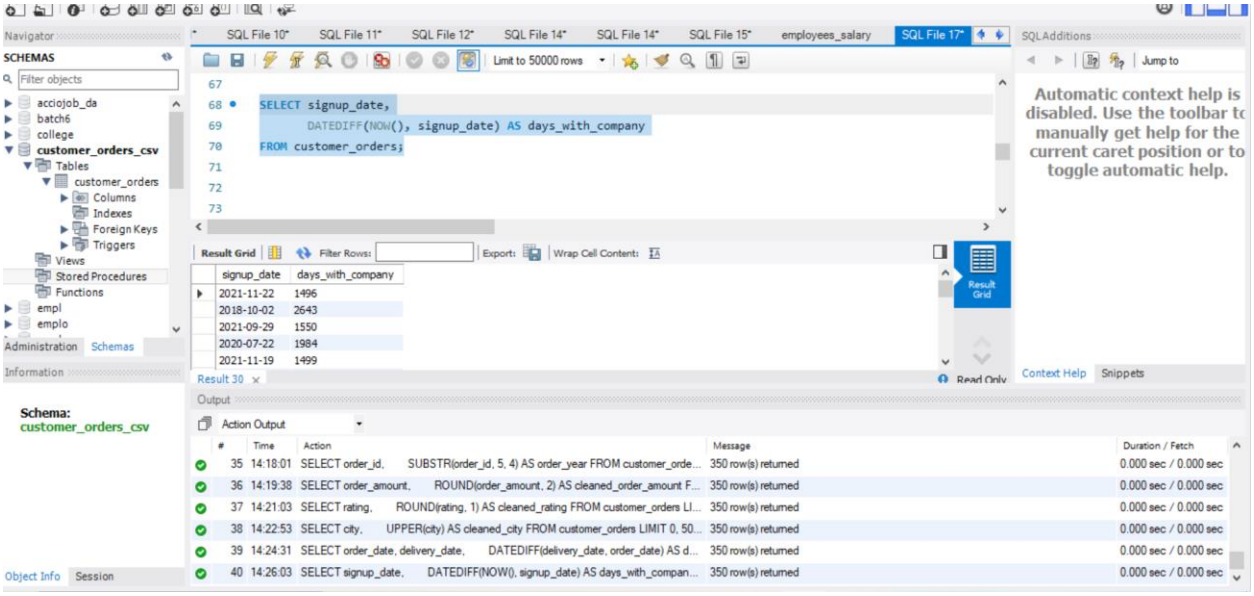
```
SELECT order_date, delivery_date,
       DATEDIFF(delivery_date, order_date) AS delivery_days
FROM customer_orders;
```



Explanation: **Calculated delivery days between order and delivery dates.**

◆ STEP 11: Customer Tenure Calculation

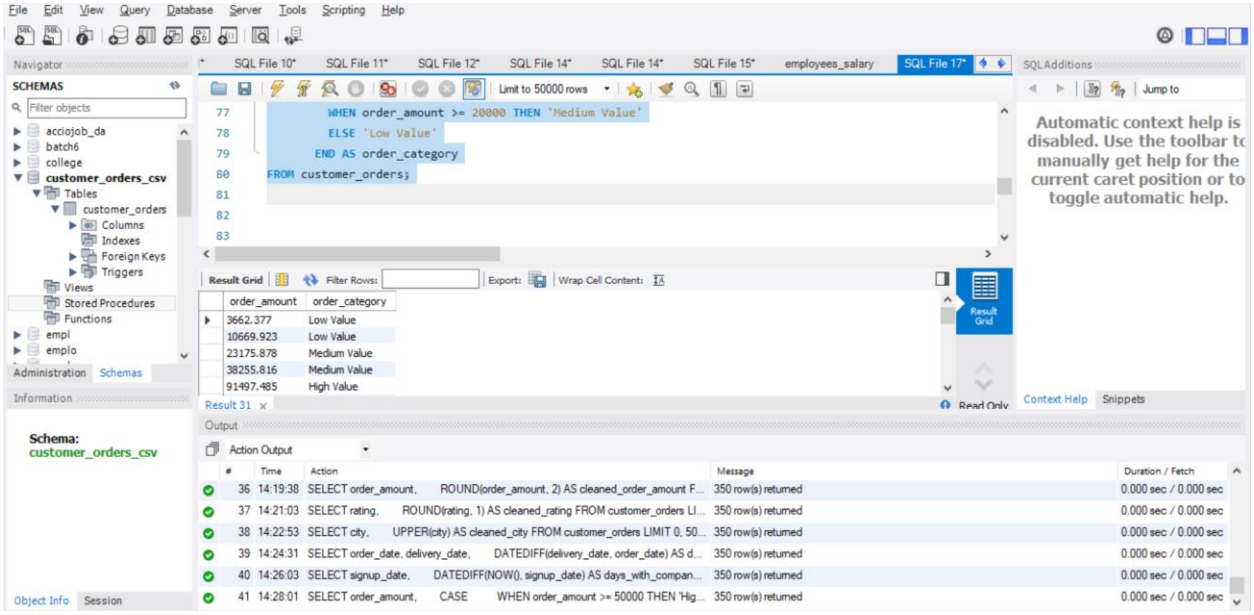
```
SELECT signup_date,
    DATEDIFF(NOW(), signup_date) AS days_with_company
FROM customer_orders;
```



Explanation: **Calculated how many days each customer has been with the company.**

◆ STEP 12: `CASE WHEN` – Order Value Categor

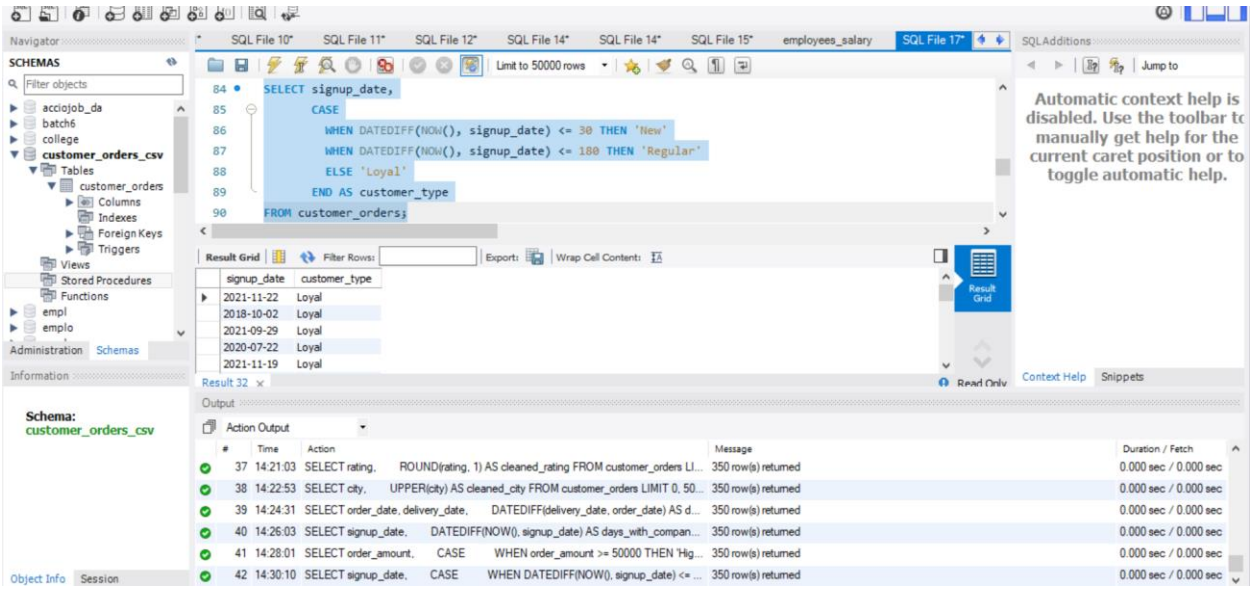
```
SELECT order_amount,
    CASE
        WHEN order_amount >= 50000 THEN 'High Value'
        WHEN order_amount >= 20000 THEN 'Medium Value'
        ELSE 'Low Value'
    END AS order_category
FROM customer_orders;
```



Explanation: **Categorized orders into High, Medium, or Low value.**

◆ STEP 13: `CASE WHEN` – Customer Type

```
SELECT signup_date,
    CASE
        WHEN DATEDIFF(NOW(), signup_date) <= 30 THEN 'New'
        WHEN DATEDIFF(NOW(), signup_date) <= 180 THEN 'Regular'
        ELSE 'Loyal'
    END AS customer_type
FROM customer_orders;
```

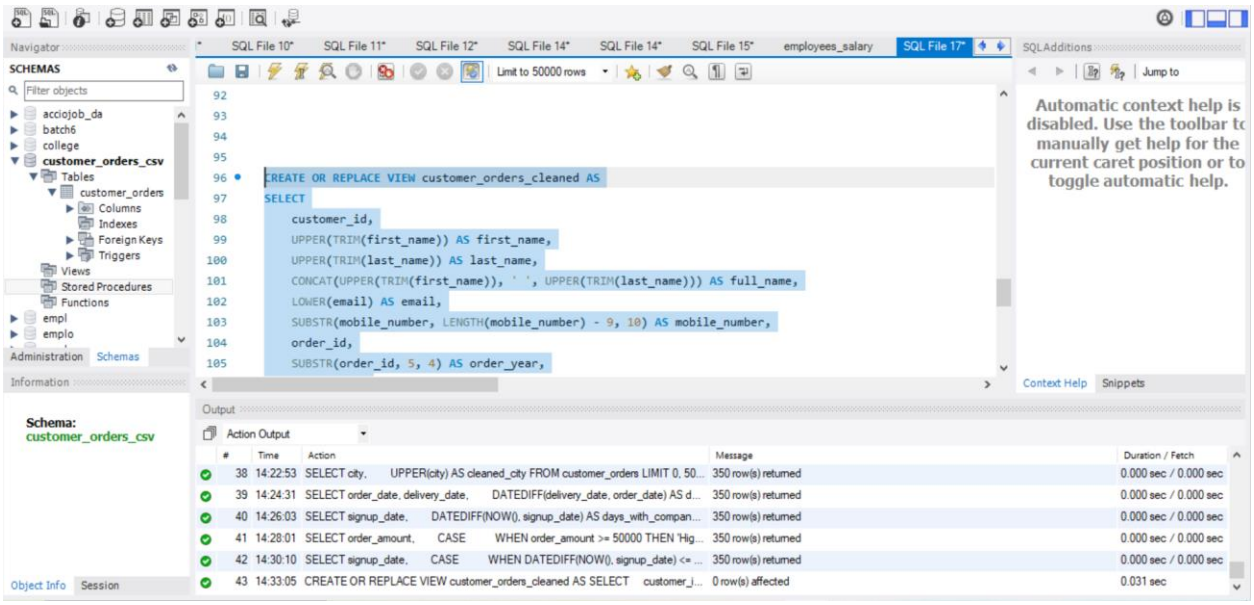


Explanation: **Classified customers as New, Regular, or Loyal.**

◆ STEP 14: FINAL CLEANED VIEW (Industry Practice)

```
CREATE OR REPLACE VIEW customer_orders_cleaned AS

SELECT

  customer_id,

  UPPER(TRIM(first_name)) AS first_name,

  UPPER(TRIM(last_name)) AS last_name,

  CONCAT(UPPER(TRIM(first_name)), ' ', UPPER(TRIM(last_name))) AS full_name,

  LOWER(email) AS email,

  SUBSTR(mobile_number, LENGTH(mobile_number) - 9, 10) AS mobile_number,

  order_id,

 SUBSTR(order_id, 5, 4) AS order_year,

  order_date,

  delivery_date,

  DATEDIFF(delivery_date, order_date) AS delivery_days,

  ROUND(order_amount, 2) AS order_amount,

  UPPER(city) AS city,

  signup_date,

  DATEDIFF(NOW(), signup_date) AS customer_tenure_days,
```

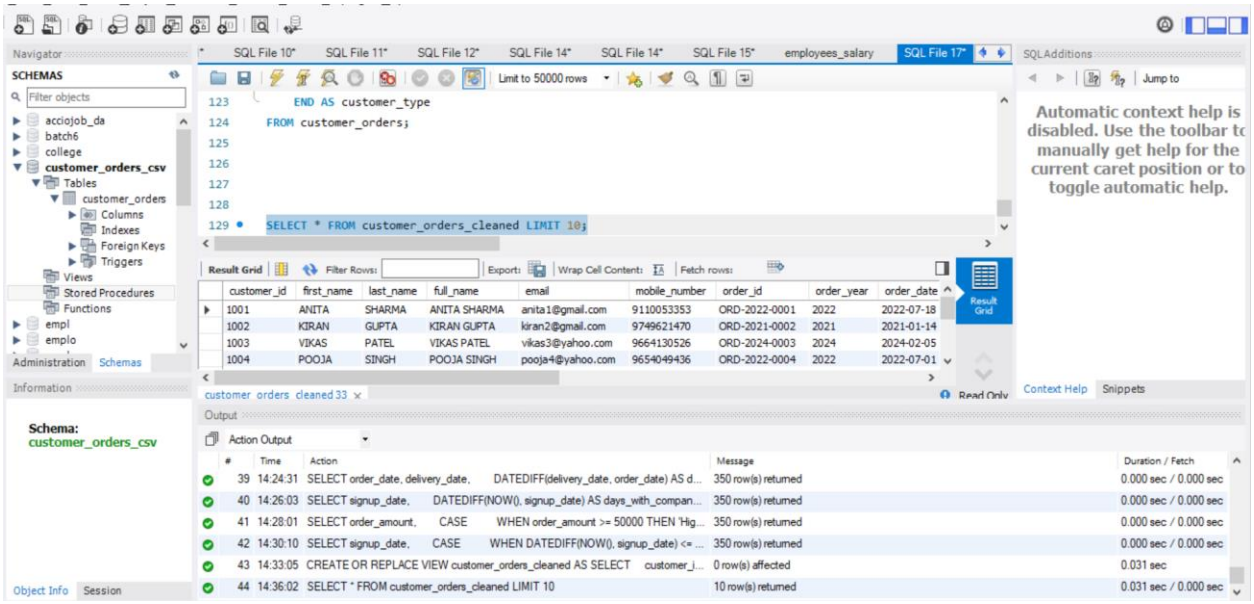
```
    CASE
        WHEN order_amount >= 50000 THEN 'High Value'
        WHEN order_amount >= 20000 THEN 'Medium Value'
        ELSE 'Low Value'
    END AS order_category,
    ROUND(rating, 1) AS rating,
    CASE
        WHEN DATEDIFF(NOW(), signup_date) <= 30 THEN 'New'
        WHEN DATEDIFF(NOW(), signup_date) <= 180 THEN 'Regular'
        ELSE 'Loyal'
    END AS customer_type
FROM customer_orders;
```

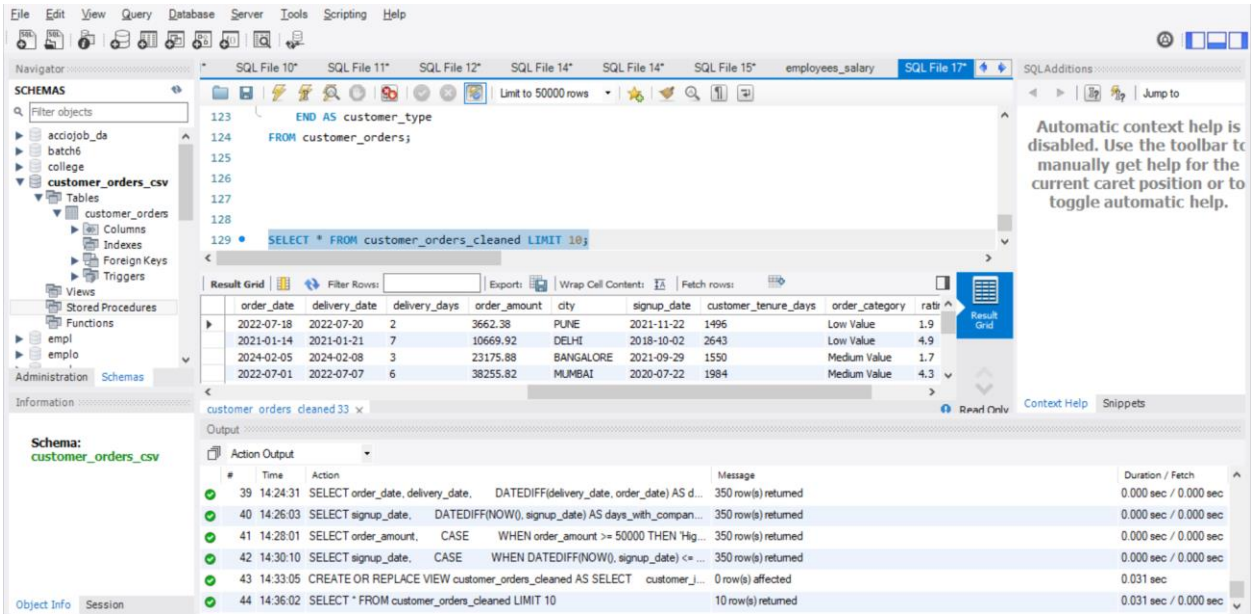


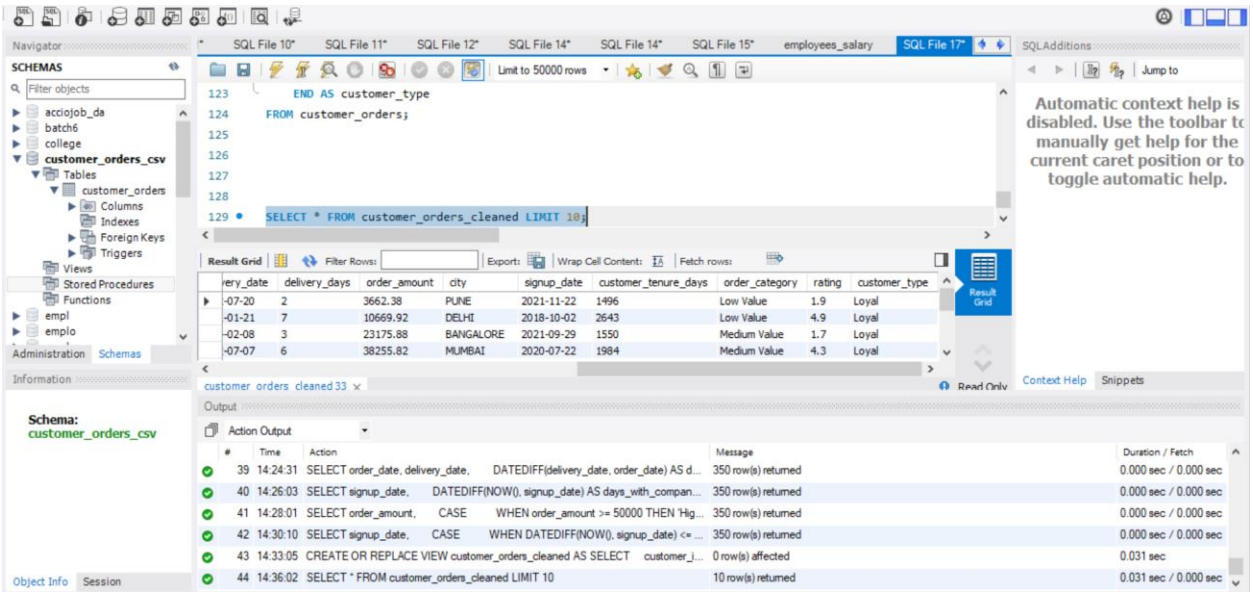
Explanation: **Created a view combining all transformations**.

◆ STEP 15: Validate Cleaned Data

```
SELECT * FROM customer_orders_cleaned LIMIT 10;
```







Explanation: **Verified that all cleaned columns are correctly displayed.**