**SQL End Course Project Report**

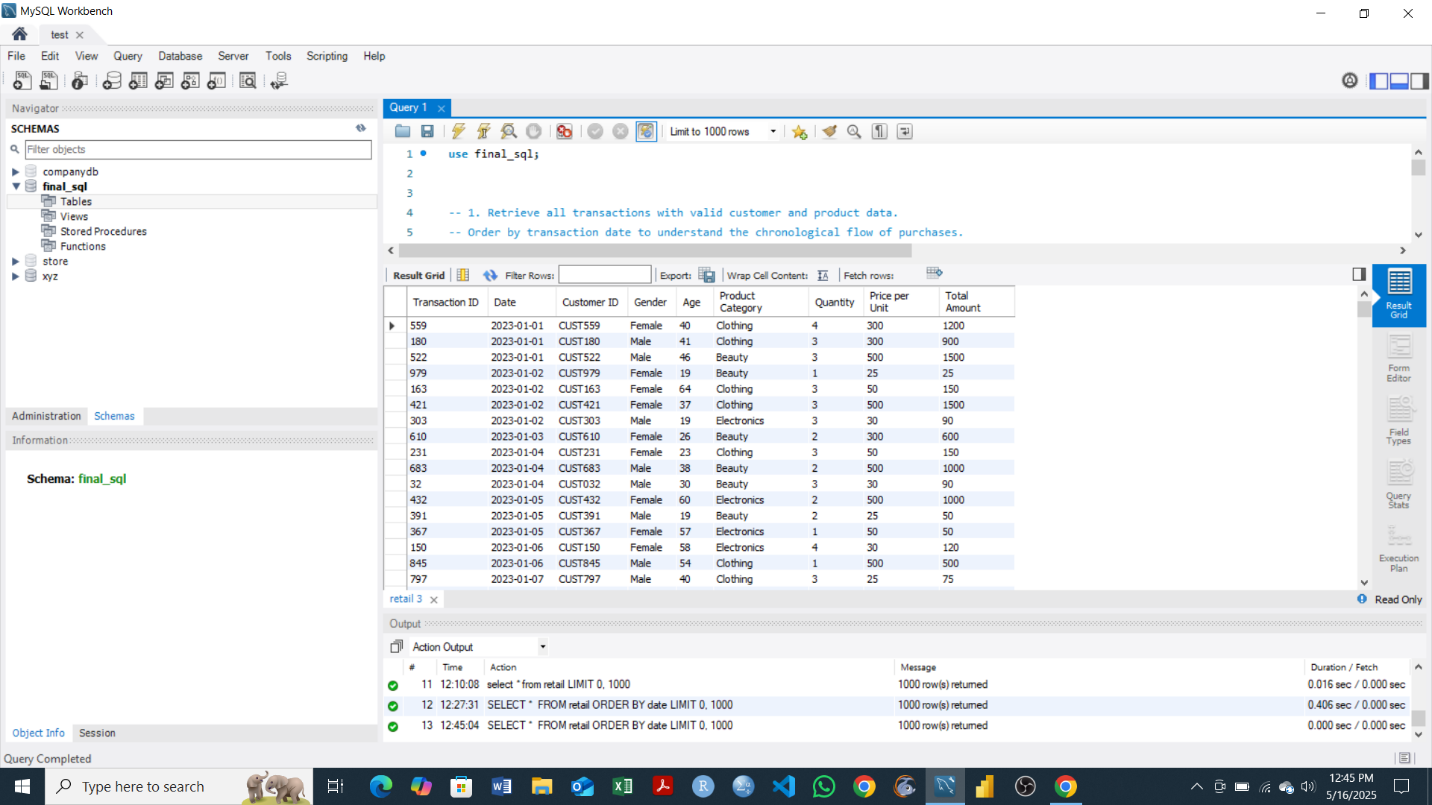
**-- 1. Retrieve all transactions with valid customer and product data.**

**-- Order by transaction date to understand the chronological flow of purchases.**

SELECT \*

FROM retail

ORDER BY date;

****

**Objective**

The query retrieves all transaction records from the retail table and sorts them by the date column in ascending order. This allows for the analysis of purchases over time.

**-- 2. Clean the dataset by ensuring that numeric fields like Quantity,**

**-- Price per Unit, and Total Amount are properly formatted.**

-- Remove duplicates or null values if any exist.

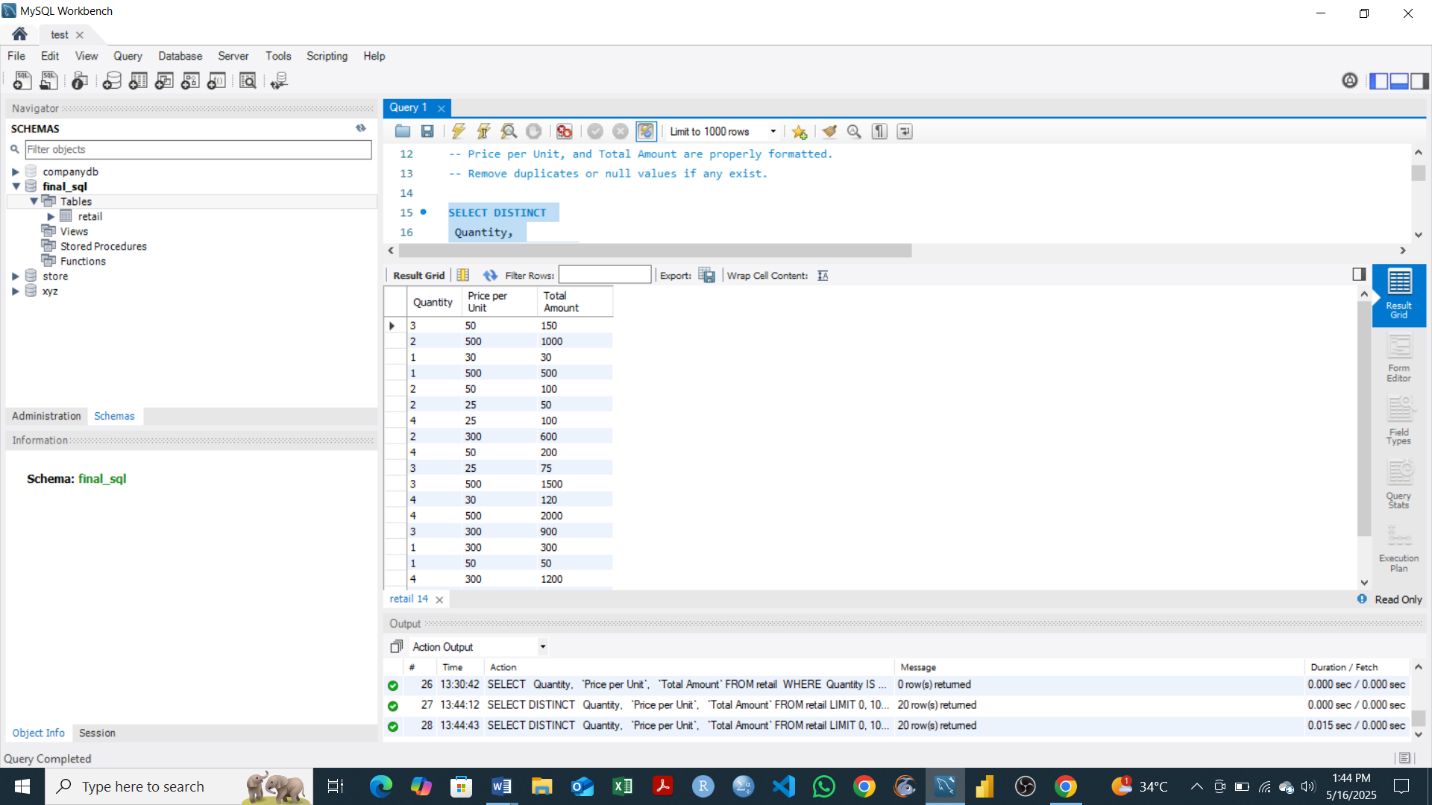
SELECT DISTINCT

Quantity,

`Price per Unit`,

`Total Amount`

FROM retail;



-- Check if any NULLs remain

SELECT

Quantity,

`Price per Unit`,

`Total Amount`

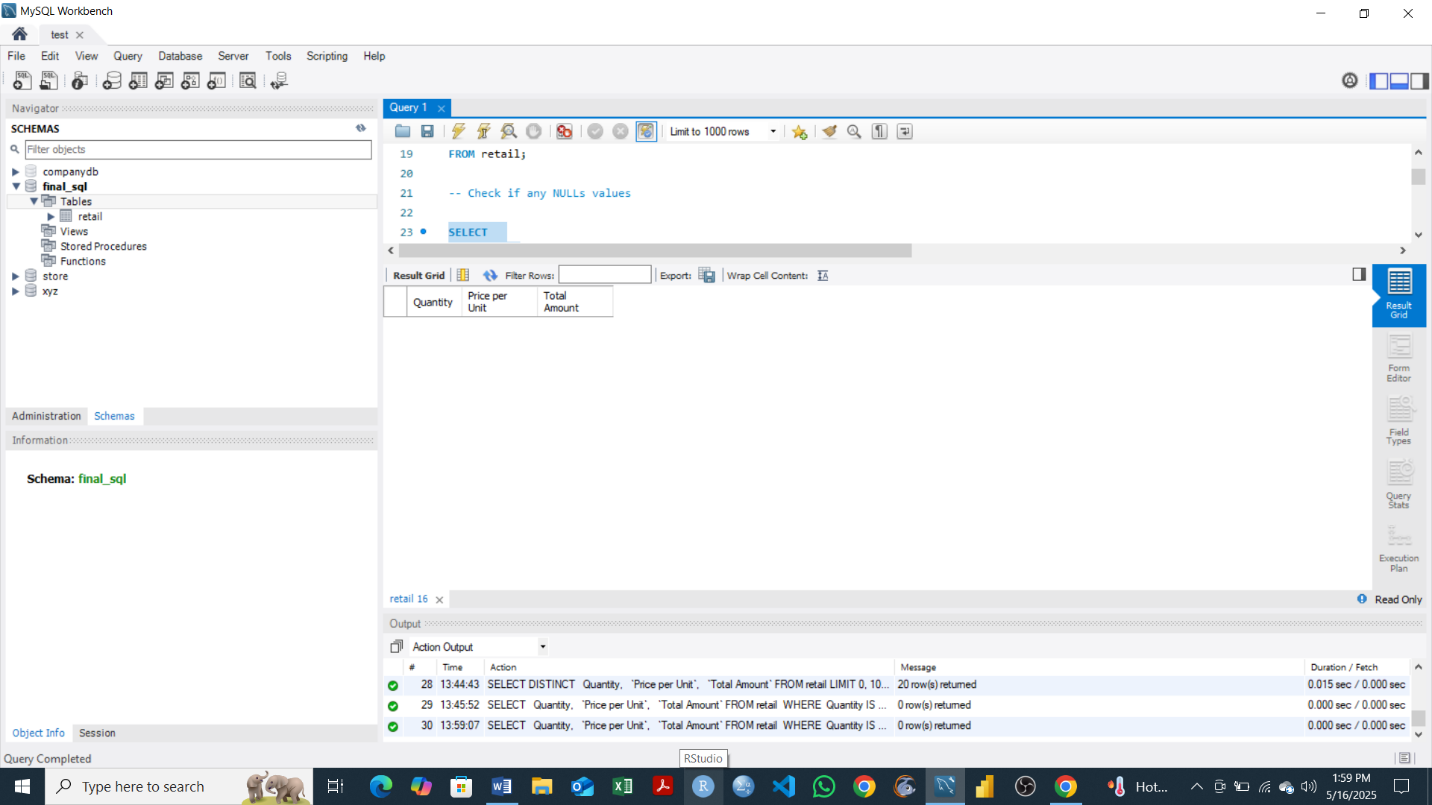
FROM retail

WHERE

Quantity IS NULL

OR `Price per Unit` IS NULL

OR `Total Amount` IS NULL;



**Objective**

To **identify and display any rows** in the retail table where the key numeric fields such as Quantity, Price per Unit, or Total Amount contain **null (missing) values**.

This helps ensure data integrity before performing further analysis or reporting, as null values can affect calculations, summaries, or visualizations. The output show that The dataset is **clean** and no null values remain in the specified columns.

**-- 3. Calculate the total and average revenue for each product category.**

**-- Which categories bring in the most and least revenue?**

SELECT

`Product Category`,

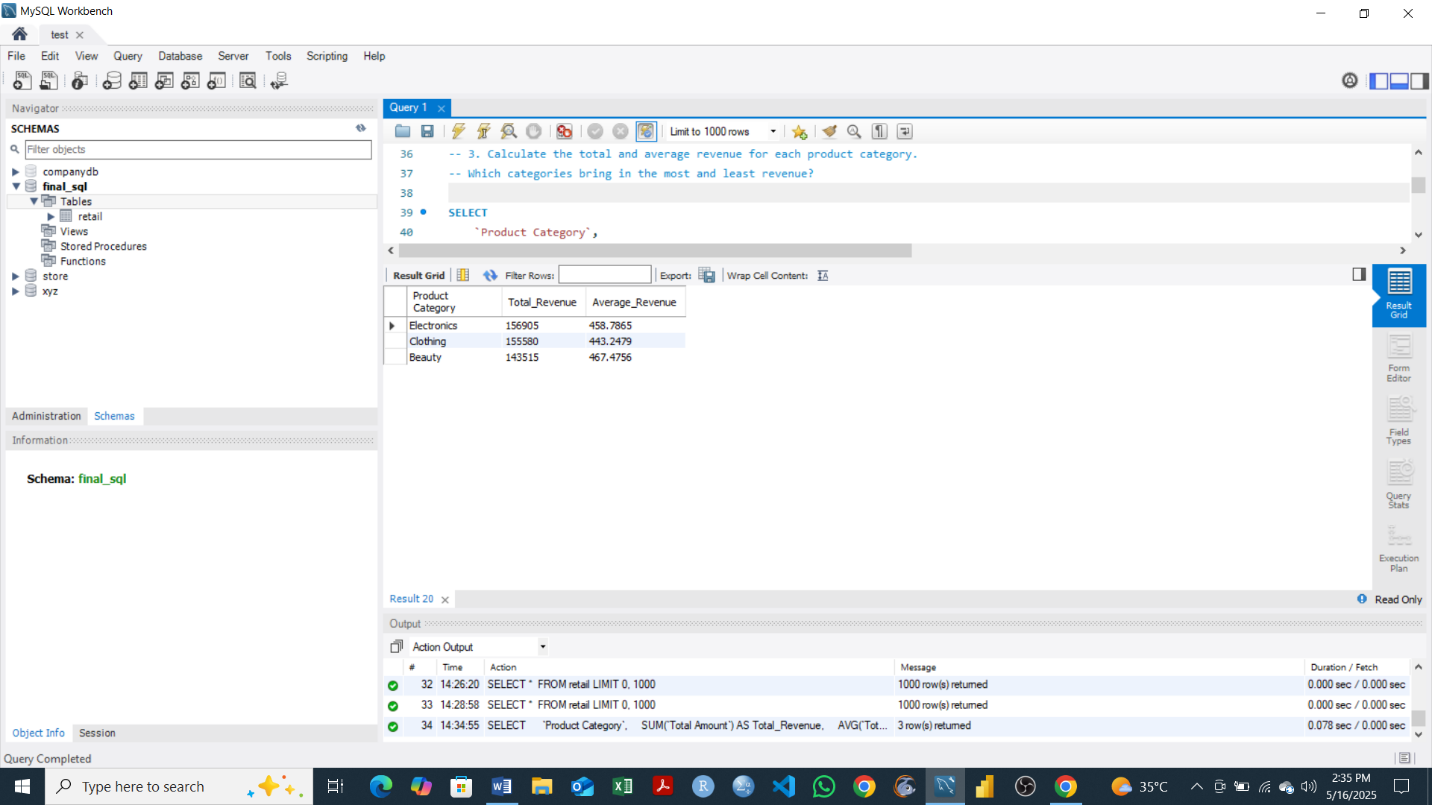
SUM(`Total Amount`) AS Total\_Revenue,

AVG(`Total Amount`) AS Average\_Revenue

FROM retail

GROUP BY `Product Category`

ORDER BY Total\_Revenue DESC;



**Objective**

The objective of this query is to analyze the financial performance of each product category by calculating both the total and average revenue generated from sales. By grouping the transactions by product category and summarizing their respective revenues, we can identify which categories contribute the most and the least to overall income. This insight helps in understanding customer spending patterns and can support strategic decisions such as inventory planning, marketing focus, and resource allocation.

**-- 4. Analyze the monthly sales trend over the entire dataset period.**

**-- Summarize total revenue per month and order the results chronologically.**

SELECT

YEAR(Date) AS Year,

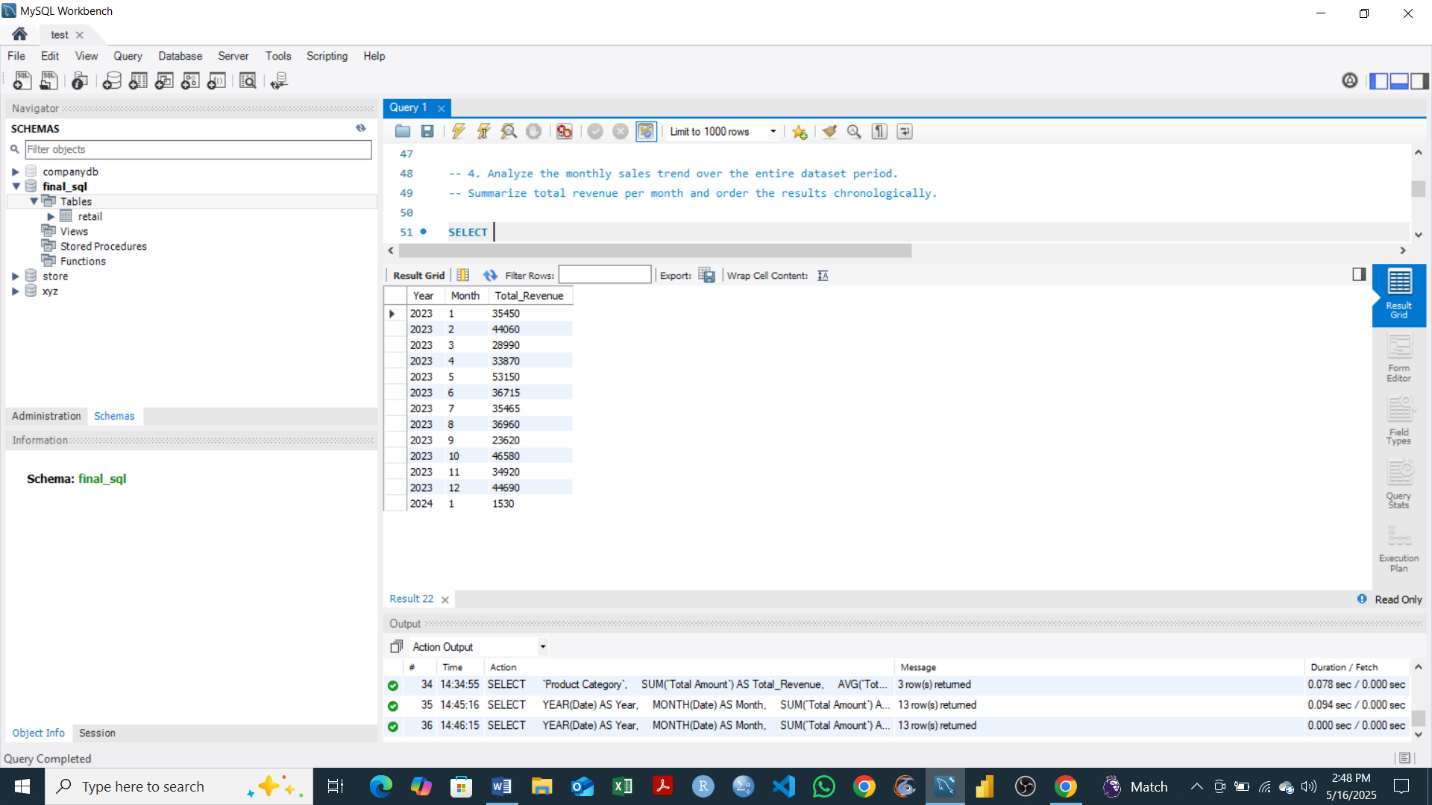
MONTH(Date) AS Month,

SUM(`Total Amount`) AS Total\_Revenue

FROM retail

GROUP BY Year, Month

ORDER BY Year, Month;



**Objective**

The goal of this query is to analyze the monthly sales trend by aggregating total revenue for each month across the dataset period. By grouping sales data by month and ordering it chronologically, the analysis helps uncover trends in sales performance over time—such as growth patterns, seasonal variations, or periods of high and low revenue. This insight is valuable for strategic planning, forecasting, and improving business decision-making based on sales cycles.

**-- 5. Identify the top 10 customers by total spending.**

**-- Rank customers based on how much they’ve spent across all transactions.**

SELECT

`Customer ID`,

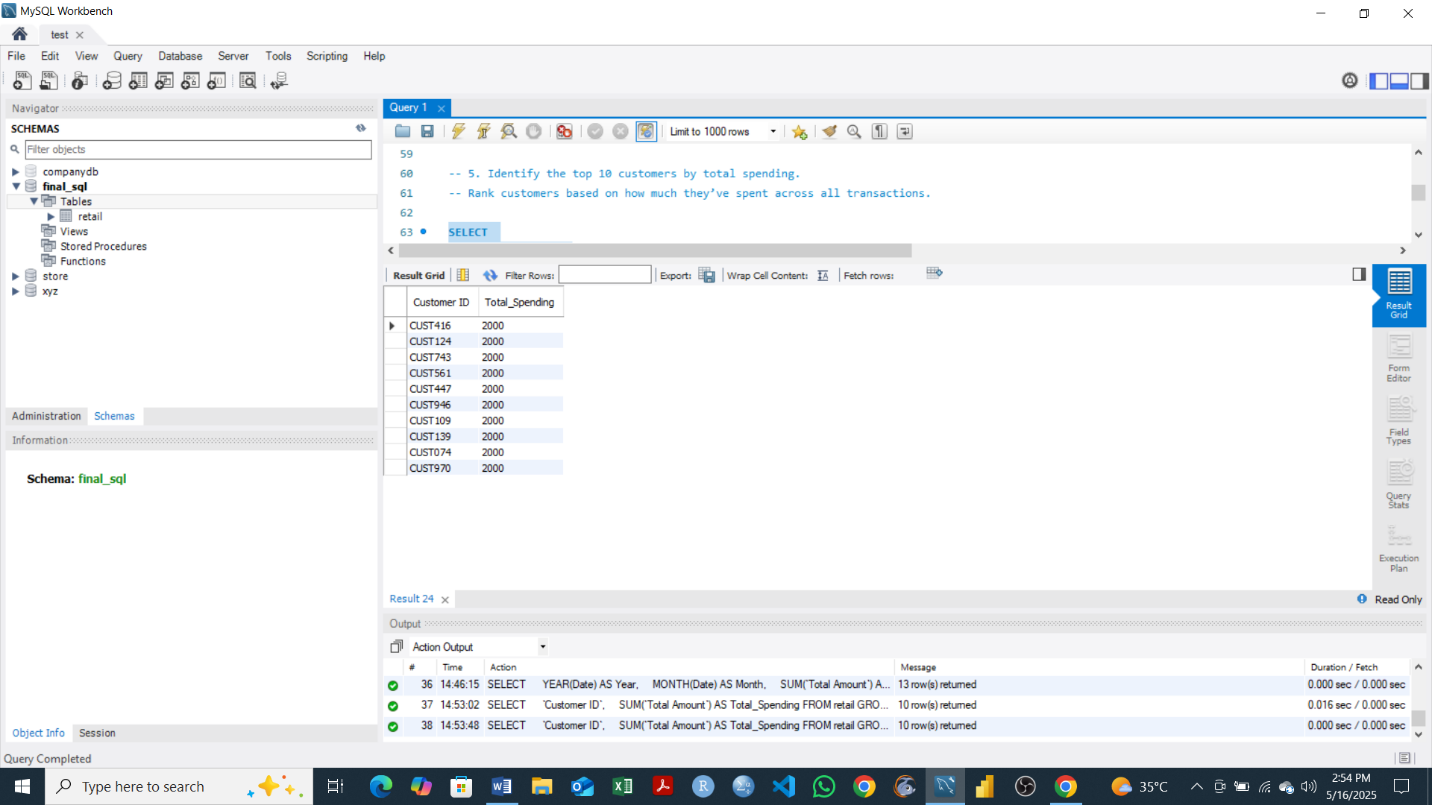
SUM(`Total Amount`) AS Total\_Spending

FROM retail

GROUP BY `Customer ID`

ORDER BY Total\_Spending DESC

LIMIT 10;



**Objective**

The objective of this query or an output is to identify the top 10 customers who have spent the most across all transactions in the dataset. By grouping the data by customer ID and calculating the total amount each customer has spent, the query ranks customers in descending order of their total spending. This helps in recognizing high-value customers, who can be targeted for loyalty programs, personalized marketing, or retention strategies to maximize business value.

**-- 6. Calculate the average transaction value for each customer.**

**-- How much does each customer spend per transaction on average?**

SELECT

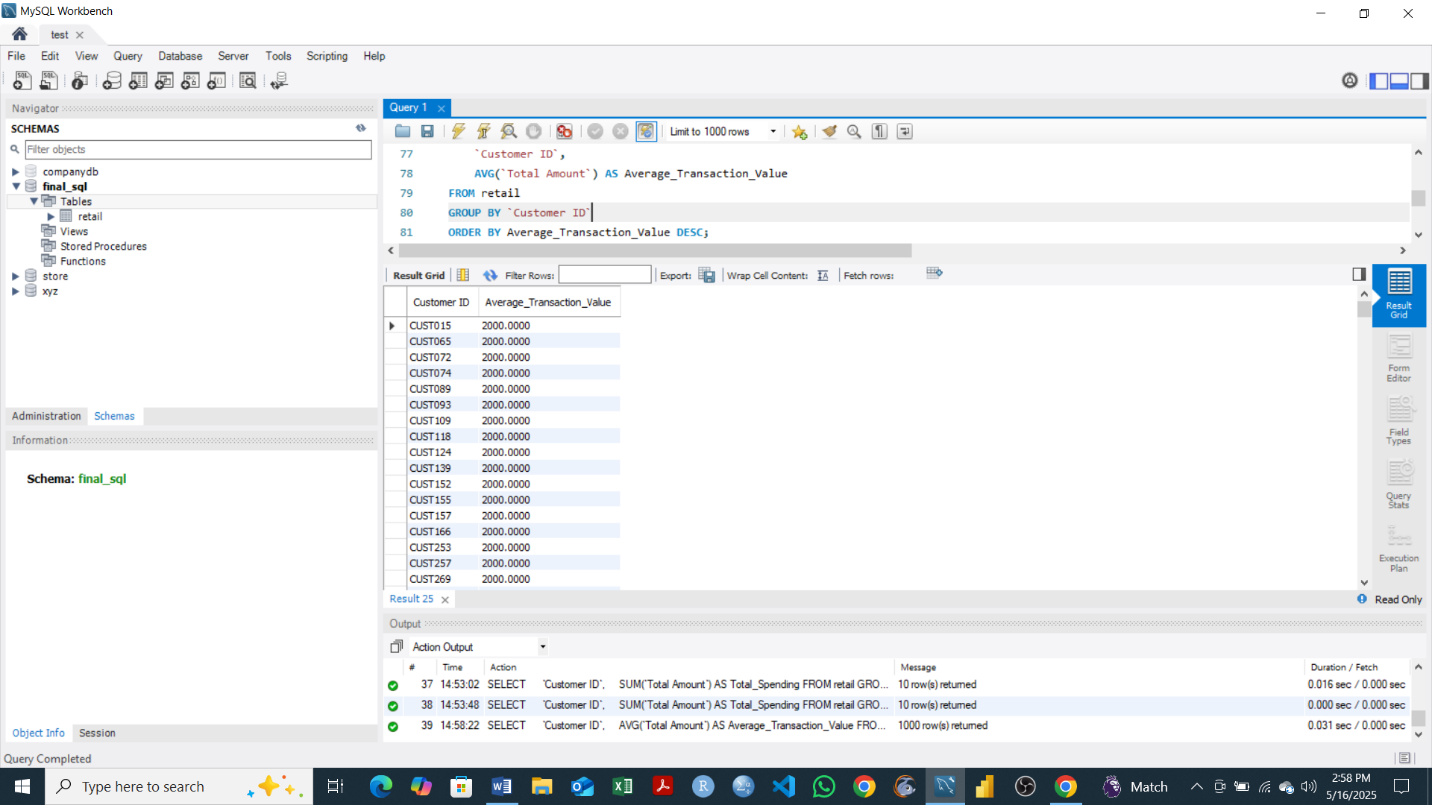
`Customer ID`,

AVG(`Total Amount`) AS Average\_Transaction\_Value

FROM retail

GROUP BY `Customer ID`

ORDER BY Average\_Transaction\_Value DESC;



**Objective**

The objective of this query is to calculate the average amount each customer spends per transaction. By grouping transactions by customer ID and computing the average of the Total Amount for each group, the query reveals individual spending behavior. This insight helps in understanding customer value, identifying big spenders, and segmenting customers based on their purchasing habits for better-targeted marketing and service strategies.

**-- 7. Group customers by gender and age brackets (e.g., 18–25, 26–35, 36–50, etc.).**

**-- Summarize total revenue and transaction count for each group.**

SELECT

`Customer ID`,

Gender,

Age,

CASE

WHEN Age BETWEEN 18 AND 25 THEN '18-25'

WHEN Age BETWEEN 26 AND 35 THEN '26-35'

WHEN Age BETWEEN 36 AND 50 THEN '36-50'

ELSE '51+'

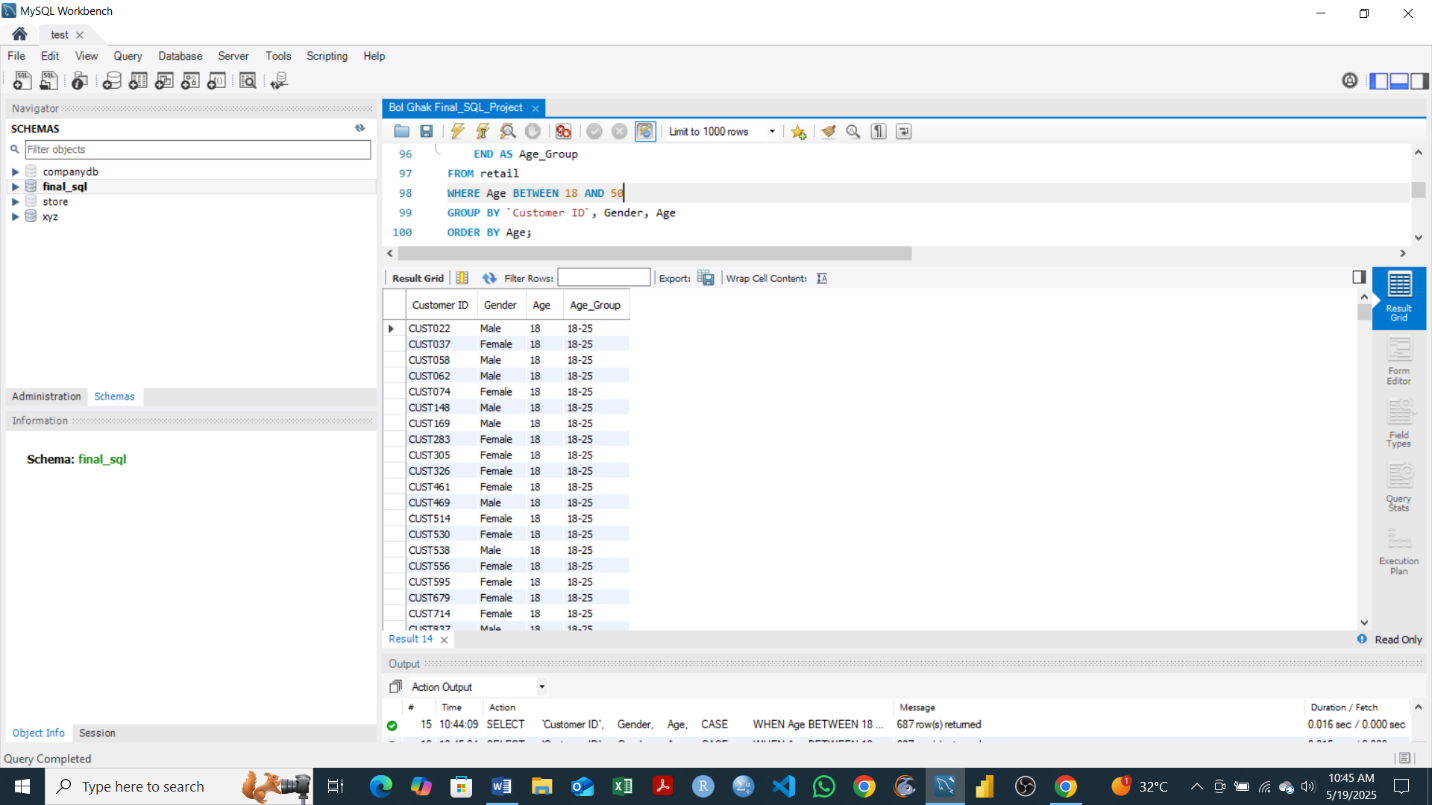
END AS Age\_Group

FROM retail

WHERE Age BETWEEN 18 AND 50

GROUP BY `Customer ID`, Gender, Age

ORDER BY Age;



**Objective**

The objective of this query is to segment customers into defined age groups between 18 and 50 years old and retrieve their demographic details, including gender and age. By grouping customers and categorizing them into age brackets (18–25, 26–35, 36–50), the analysis supports a better understanding of the customer base by age distribution. This can guide personalized marketing strategies, product offerings, and customer engagement based on age demographics.

**-- 8. Compare the number of one-time buyers versus repeat buyers.**

**-- Group customers by purchase frequency to determine repeat behavior.**

SELECT

Purchase\_Count,

COUNT(\*) AS Number\_of\_Customers

FROM (

SELECT

`Customer ID`,

COUNT(\*) AS Purchase\_Count

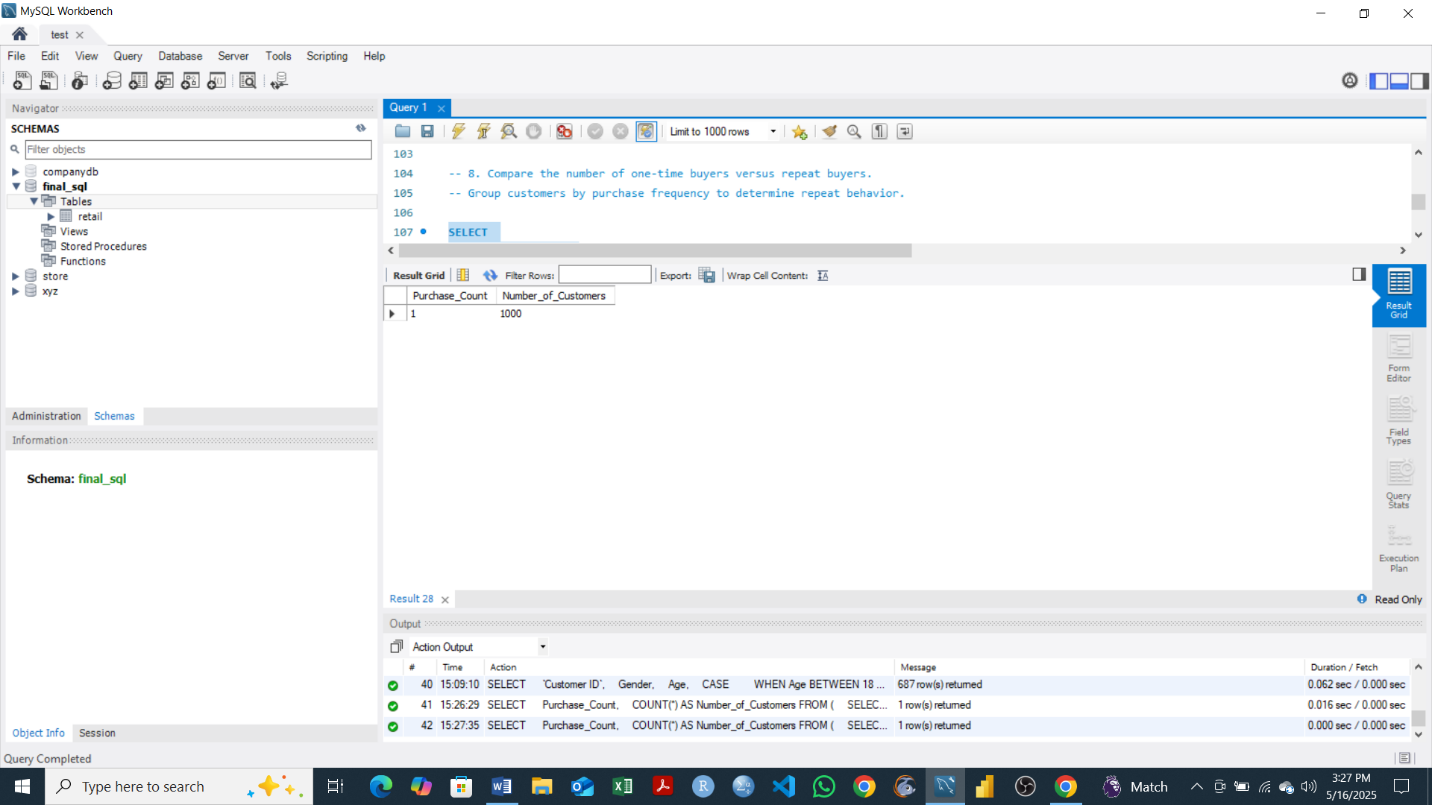
FROM retail

GROUP BY `Customer ID`

) AS Customer\_Purchases

GROUP BY Purchase\_Count

ORDER BY Purchase\_Count;



**Objective**

The goal of this query is to analyze customer purchasing behavior by comparing the number of one-time buyers to repeat buyers. It first counts how many purchases each customer has made and then groups customers based on their purchase frequency. This breakdown reveals how many customers only purchased once versus those who made multiple transactions, offering valuable insight into customer retention and loyalty trends.

**-- 9. Identify inactive customers who have not made a purchase in the last 6 months.**

**-- Use the most recent date in the dataset as the reference point.**

SELECT

`Customer ID`,

MAX(Date) AS Last\_Purchase\_Date

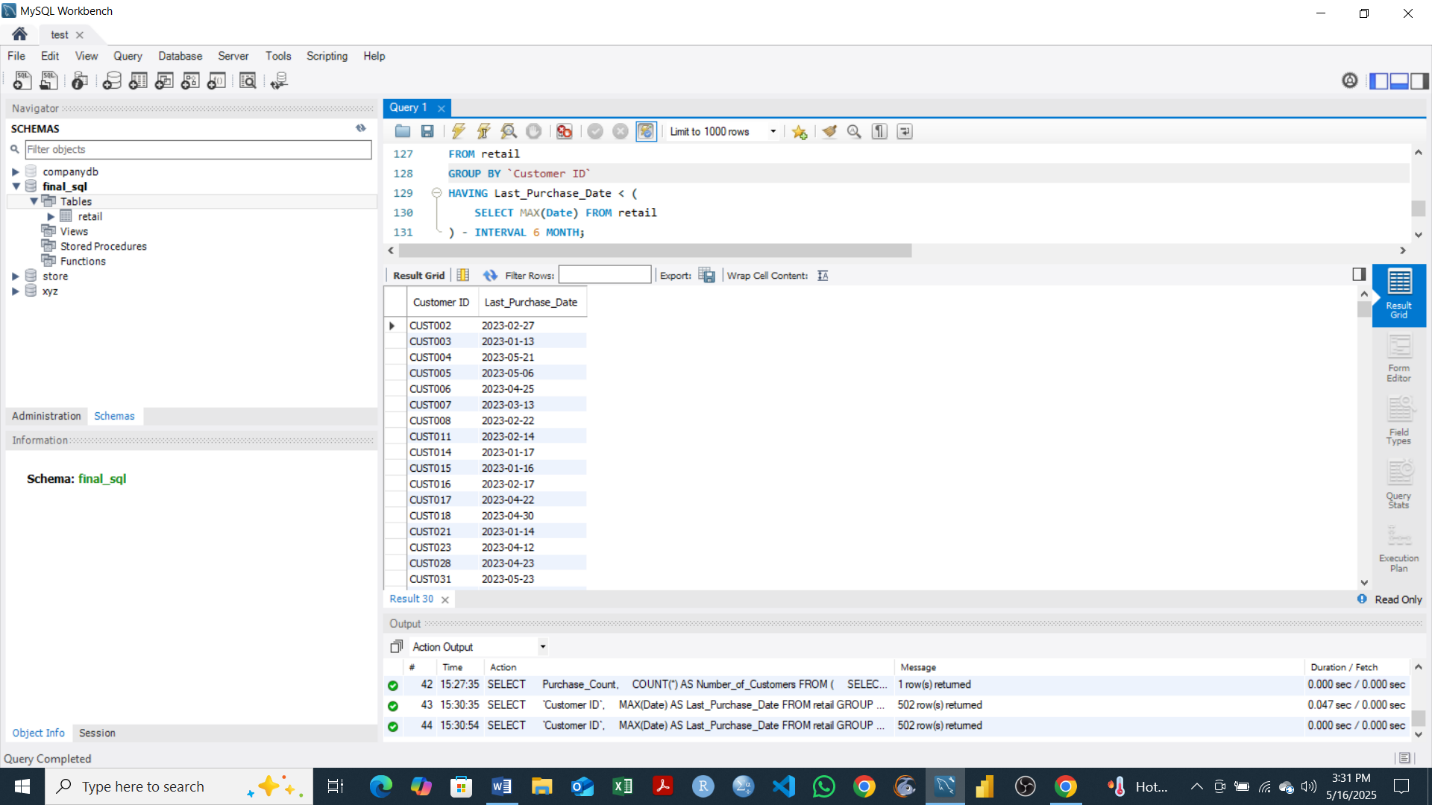
FROM retail

GROUP BY `Customer ID`

HAVING Last\_Purchase\_Date < (

SELECT MAX(Date) FROM retail

) - INTERVAL 6 MONTH;



**Objective**

The objective of this query is to identify inactive customers—those who have not made any purchases in the last 6 months. By first finding each customer's most recent purchase date and comparing it against the latest transaction date in the dataset, we can filter out those who have not engaged with the business recently. This insight is valuable for re-engagement campaigns, retention strategies, and understanding customer churn trends.

**-- 10. Perform RFM (Recency, Frequency, Monetary) analysis for customer segmentation.**

**-- Recency: Days since last purchase; Frequency: Number of purchases; Monetary: Total amount spent.**

-- Step 1: Calculate RFM metrics for each customer

SELECT

`Customer ID`,

-- Recency: Days since last purchase

DATEDIFF(

(SELECT MAX(Date) FROM retail),

MAX(Date)

) AS Recency,

-- Frequency: Number of purchases

COUNT(\*) AS Frequency,

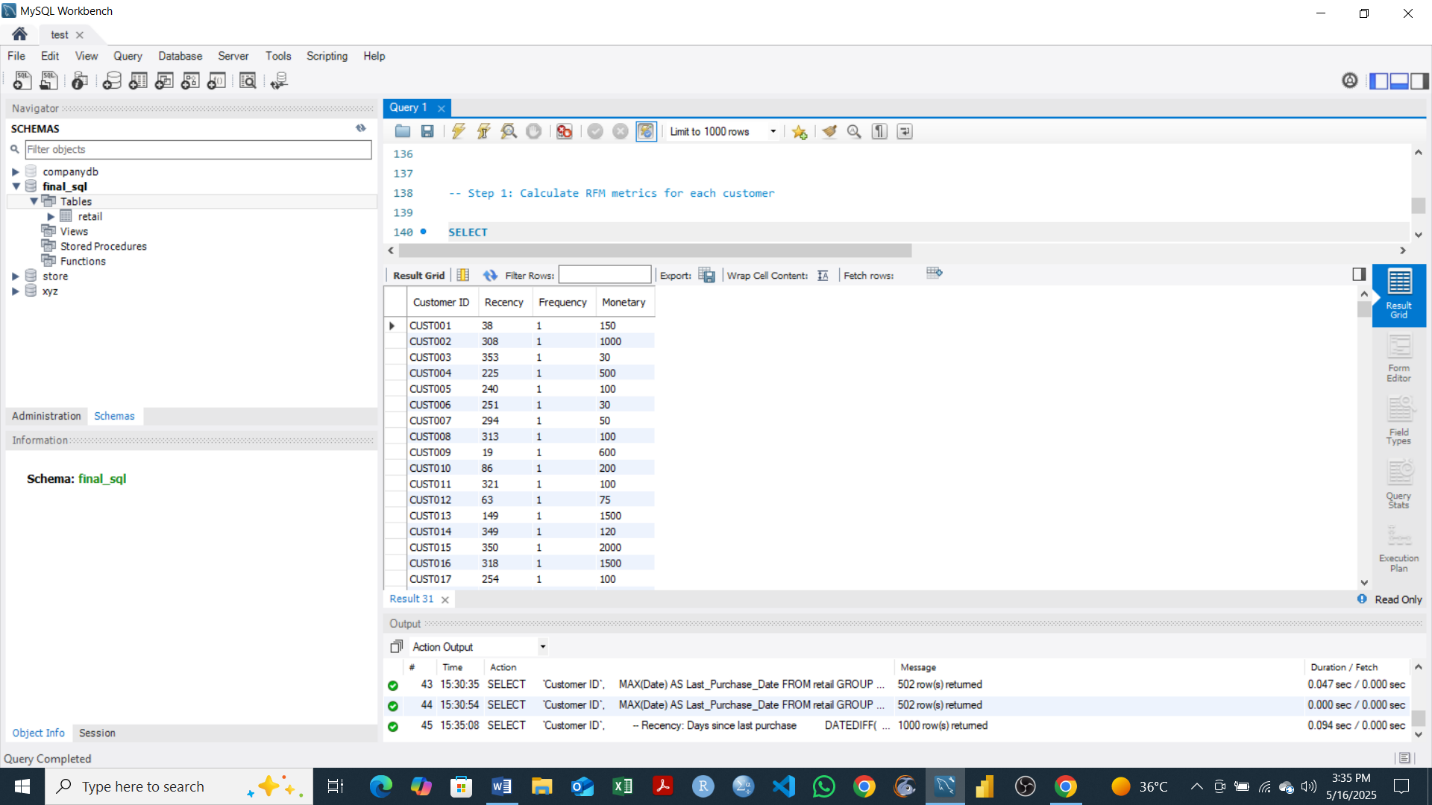
-- Monetary: Total amount spent

SUM(`Total Amount`) AS Monetary

FROM retail

GROUP BY `Customer ID`

ORDER BY `Customer ID`;



**Objective**

The purpose of this query is to perform an RFM (Recency, Frequency, Monetary) analysis for customer segmentation. RFM is a marketing technique used to evaluate and rank customers based on their transaction history. Recency measures how long it has been since a customer's last purchase, Frequency indicates how often the customer has purchased, and Monetary reflects how much they have spent in total. This analysis helps identify valuable customer segments such as loyal buyers, potential churners, and high-spending individuals, enabling targeted marketing and customer retention strategies.

**-- 11. Find the product categories with the highest average quantity per transaction.**

**-- Which product types are purchased in bulk?**

SELECT

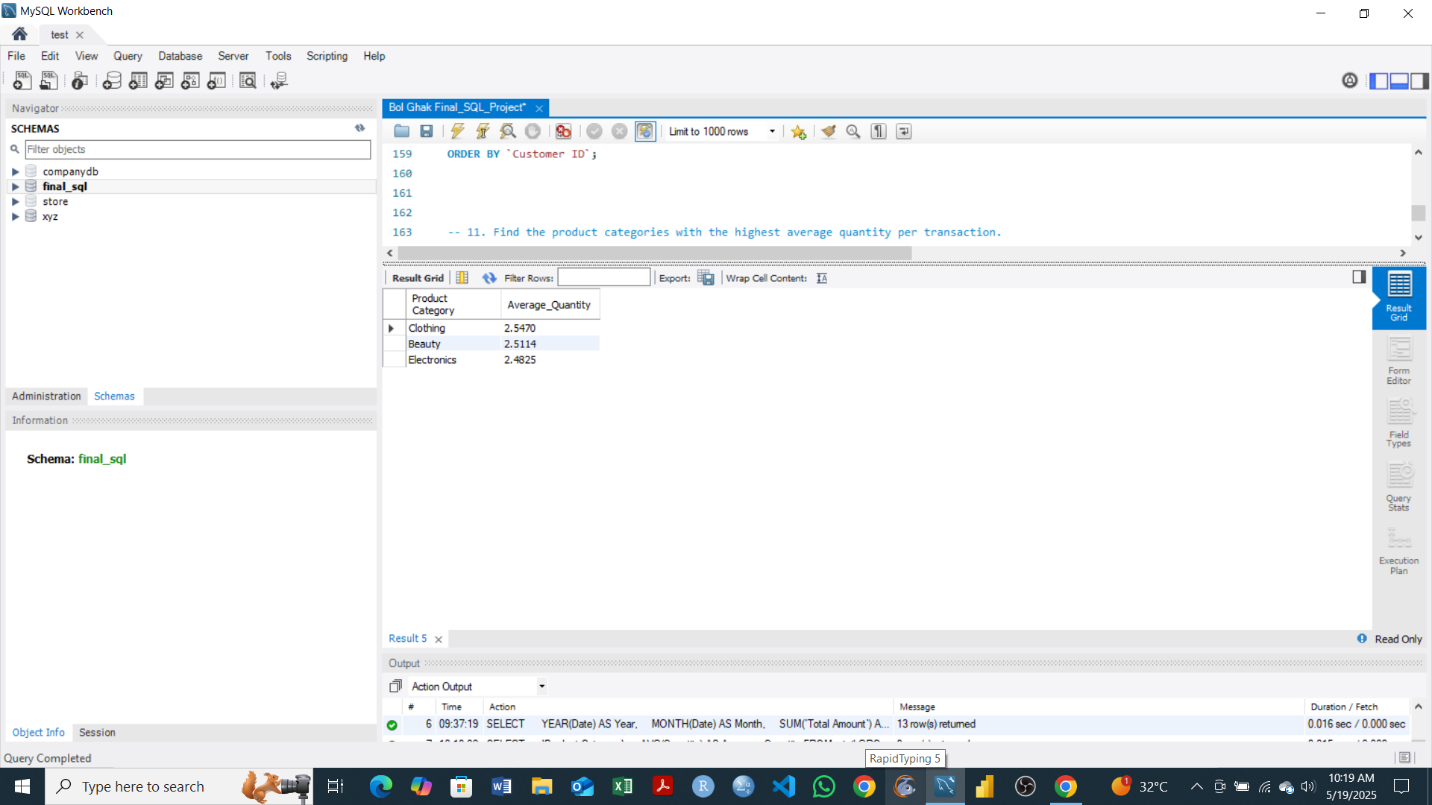
`Product Category`,

AVG(Quantity) AS Average\_Quantity

FROM retail

GROUP BY `Product Category`

ORDER BY Average\_Quantity DESC;



**Objective**

The goal of this query is to determine which product categories are typically purchased in larger quantities per transaction. By calculating the average quantity purchased for each product category and ordering the results in descending order, the query highlights which items are most commonly bought in bulk. This insight can support inventory planning, pricing strategies, and marketing efforts focused on volume-based promotions.

**-- 12. Identify the busiest sales day of the week.**

**-- Which day(s) consistently have the highest transaction volume or revenue?**

SELECT

DAYNAME(Date) AS Day\_of\_Week,

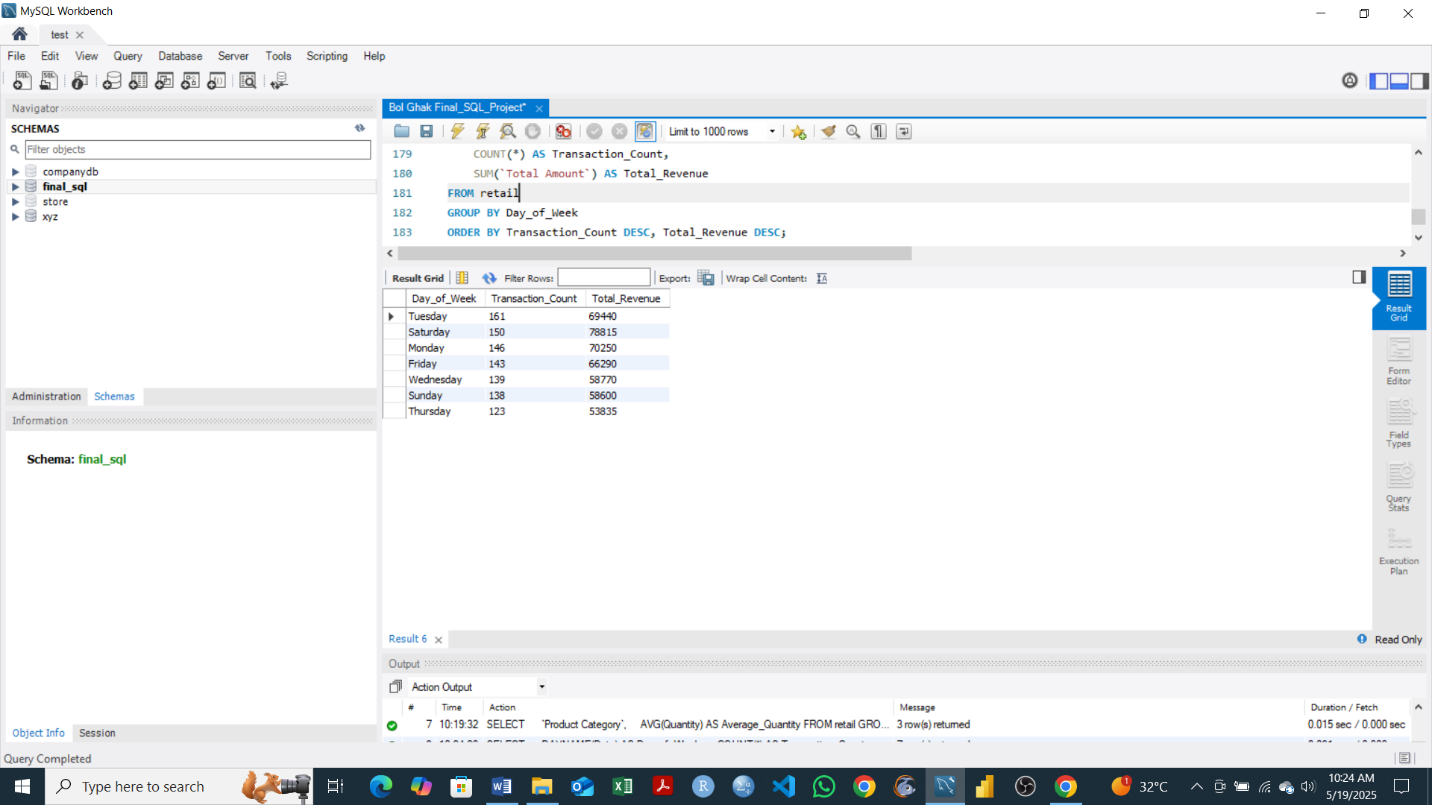
COUNT(\*) AS Transaction\_Count,

SUM(`Total Amount`) AS Total\_Revenue

FROM retail

GROUP BY Day\_of\_Week

ORDER BY Transaction\_Count DESC, Total\_Revenue DESC;



**Objective**

The objective of this query is to identify which day of the week experiences the highest sales activity. By analyzing both the number of transactions and the total revenue generated each day, the query helps pinpoint the busiest sales day(s). Understanding weekly sales patterns supports staffing decisions, marketing timing, and operational planning, ensuring resources are aligned with peak business periods.

**-- 13. Calculate total revenue and average spend per transaction by gender.**

**-- Are there differences in spending patterns across genders?**

SELECT

Gender,

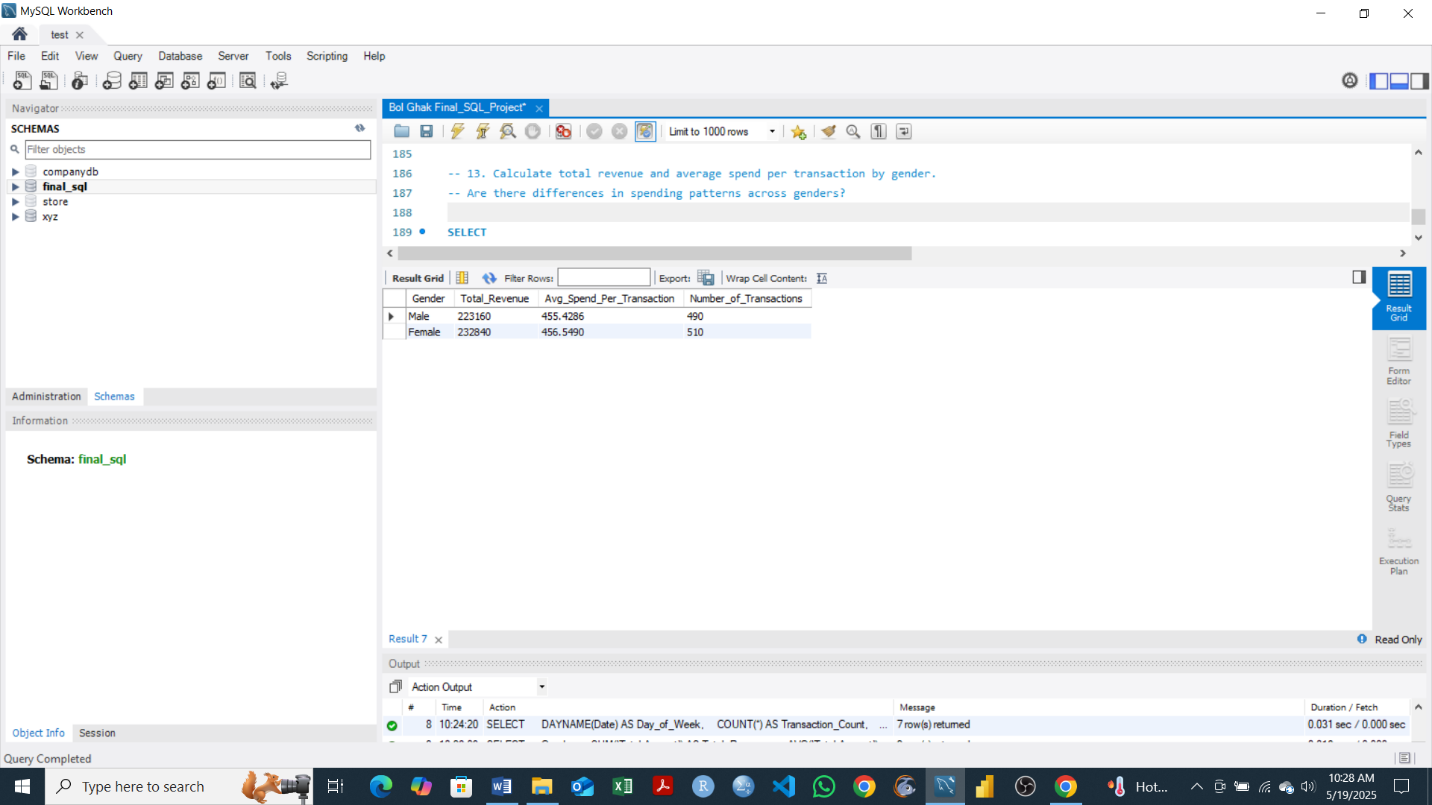
SUM(`Total Amount`) AS Total\_Revenue,

AVG(`Total Amount`) AS Avg\_Spend\_Per\_Transaction,

COUNT(\*) AS Number\_of\_Transactions

FROM retail

GROUP BY Gender;



**Objective**

This query aims to compare spending behavior across different genders by calculating both the total revenue and the average spend per transaction for each gender. By grouping transactions by gender, the analysis reveals whether there are significant differences in how much male and female customers typically spend. These insights can help tailor marketing strategies, promotions, and customer engagement initiatives based on spending patterns.

**-- 14. Find the top 5 most frequently purchased product categories.**

**-- Based on number of transactions involving each category.**

SELECT

`Product Category`,

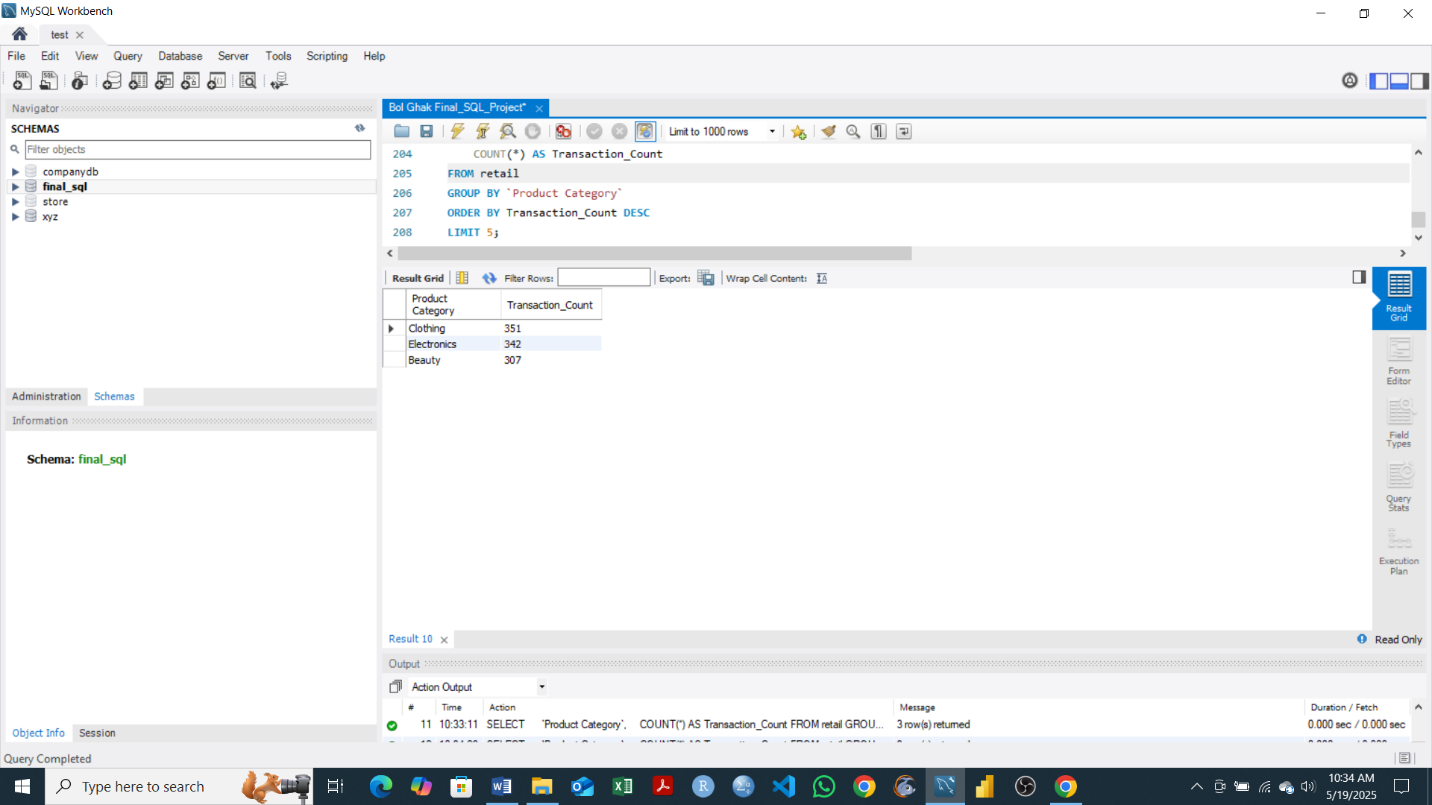
COUNT(\*) AS Transaction\_Count

FROM retail

GROUP BY `Product Category`

ORDER BY Transaction\_Count DESC

LIMIT 5;



**Objective**

The goal of this query is to identify the top 5 product categories that are most frequently purchased, based on the number of transactions associated with each category. This helps reveal customer preferences and popular items, which is useful for inventory management, targeted marketing, and optimizing product offerings to meet demand.

**-- 15. Determine the percentage of total revenue contributed by eachage group.**

**-- Which customer age brackets are most valuable to the business?**

-- Step 1: Categorize customers into age brackets and calculate revenue per group

SELECT

CASE

WHEN Age BETWEEN 18 AND 25 THEN '18-25'

WHEN Age BETWEEN 26 AND 35 THEN '26-35'

WHEN Age BETWEEN 36 AND 50 THEN '36-50'

WHEN Age > 50 THEN '51+'

ELSE 'Unknown'

END AS Age\_Group,

SUM(`Total Amount`) AS Group\_Revenue,

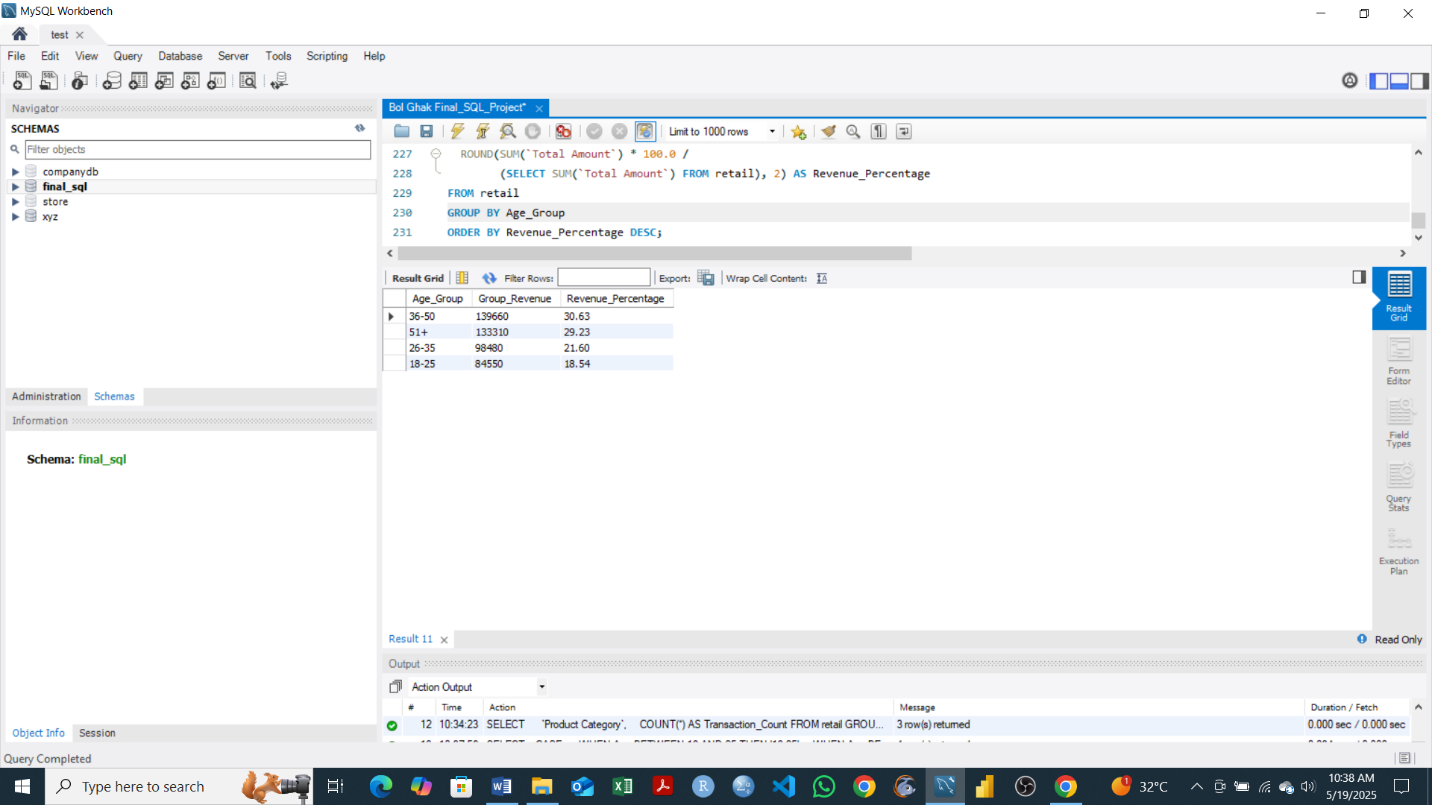
ROUND(SUM(`Total Amount`) \* 100.0 /

(SELECT SUM(`Total Amount`) FROM retail), 2) AS Revenue\_Percentage

FROM retail

GROUP BY Age\_Group

ORDER BY Revenue\_Percentage DESC;



**Objective**

This query calculates the percentage of total revenue generated by different customer age groups. By segmenting customers into defined age brackets and computing each group's contribution to overall sales, the analysis highlights which age demographics are most valuable to the business. These insights support more focused marketing, product targeting, and customer relationship strategies based on age-based spending trends.