ATER DENG MAYEN

SQL PROJECT WORDS REPORT

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

Order by transaction date to understand the chronological flow of

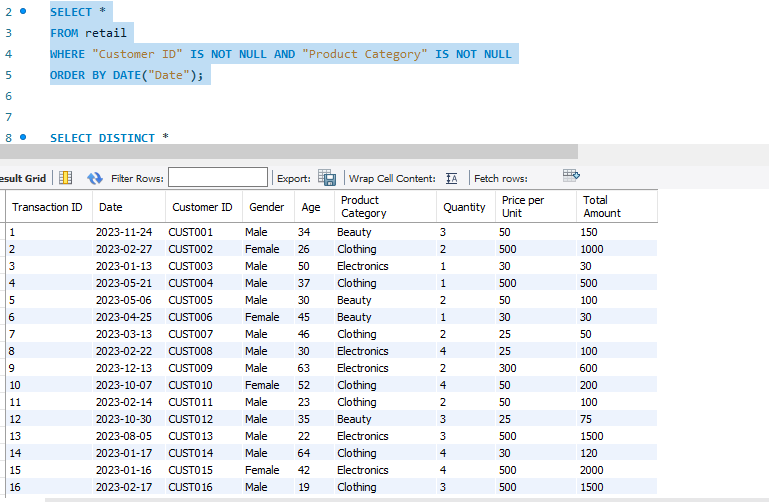
purchases.

SELECT \*

FROM retail

WHERE "Customer ID" IS NOT NULL AND "Product Category" IS NOT NULL

ORDER BY DATE("Date");



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

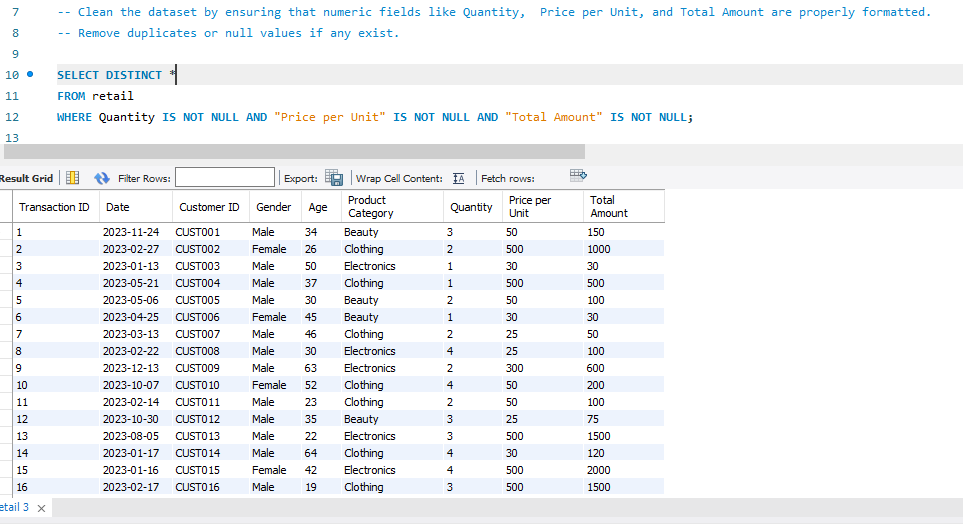
Price per Unit, and Total Amount are properly formatted.

Remove duplicates or null values if any exist.

SELECT DISTINCT \*

FROM retail

WHERE Quantity IS NOT NULL AND "Price per Unit" IS NOT NULL AND "Total Amount" IS NOT NULL;



3. Calculate the total and average revenue for each product

category.

Which categories bring in the most and least revenue?

SELECT "Product Category",

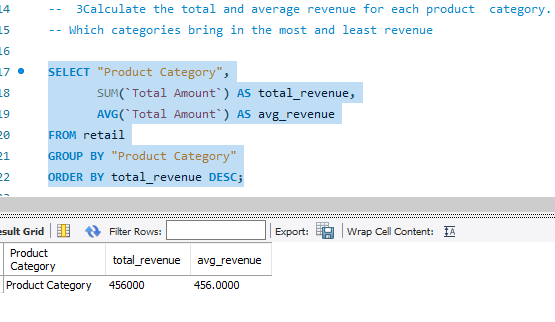
SUM(`Total Amount`) AS total\_revenue,

AVG(`Total Amount`) AS avg\_revenue

FROM retail

GROUP BY "Product Category"

ORDER BY total\_revenue DESC;



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

Summarize total revenue per month and order the results

chronologically.

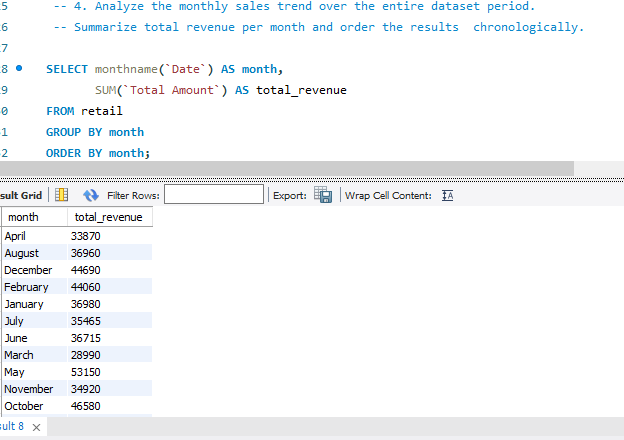
SELECT monthname(`Date`) AS month,

SUM(`Total Amount`) AS total\_revenue

FROM retail

GROUP BY month

ORDER BY month;



5. Identify the top 10 customers by total spending.

Rank customers based on how much they’ve spent across all

transactions.

SELECT "Customer ID",

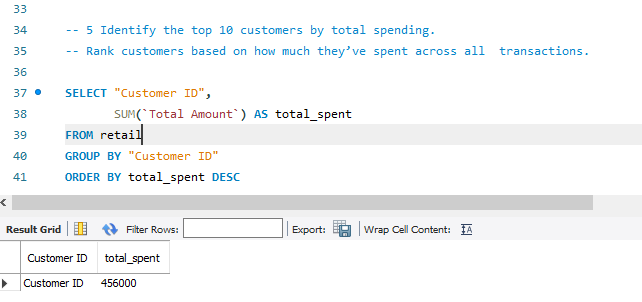
SUM(`Total Amount`) AS total\_spent

FROM retail

GROUP BY "Customer ID"

ORDER BY total\_spent DESC

LIMIT 10;



6. Calculate the average transaction value for each customer.

How much does each customer spend per transaction on

average?

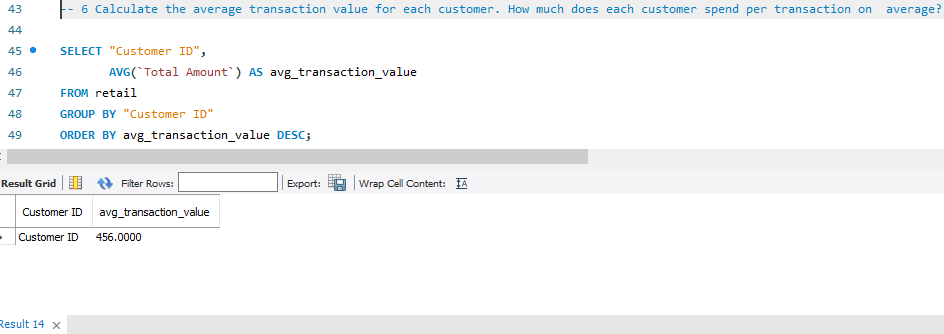
SELECT "Customer ID",

AVG(`Total Amount`) AS avg\_transaction\_value

FROM retail

GROUP BY "Customer ID"

ORDER BY avg\_transaction\_value DESC;



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 Gender,

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+'

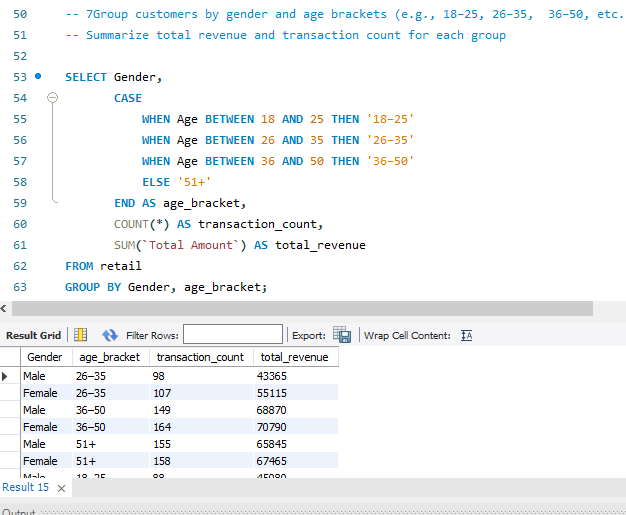
END AS age\_bracket,

COUNT(\*) AS transaction\_count,

SUM(`Total Amount`) AS total\_revenue

FROM retail

GROUP BY Gender, age\_bracket;



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 customer\_count

FROM (

SELECT "Customer ID", COUNT(\*) AS purchase\_count

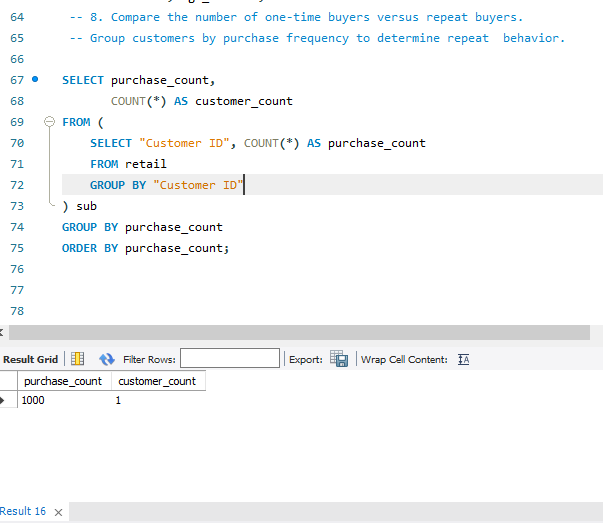
FROM retail

GROUP BY "Customer ID"

) sub

GROUP BY purchase\_count

ORDER BY purchase\_count;



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.

WITH LatestDate AS (

SELECT MAX(Date) AS MaxDate FROM retail

),

LastPurchase AS (

SELECT `Customer ID`, MAX(Date) AS LastPurchaseDate

FROM retail

GROUP BY `Customer ID`

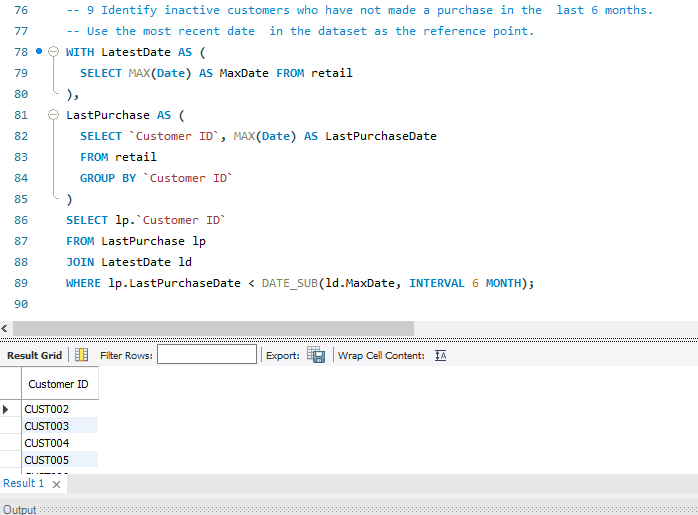
)

SELECT lp.`Customer ID`

FROM LastPurchase lp

JOIN LatestDate ld

WHERE lp.LastPurchaseDate < DATE\_SUB(ld.MaxDate, INTERVAL 6 MONTH);



10. Perform RFM (Recency, Frequency, Monetary) analysis for

customer segmentation.

Recency: Days since last purchase; Frequency: Number of

purchases; Monetary: Total amount spent.

WITH LatestDate AS (

SELECT MAX(Date) AS MaxDate FROM retail

),

CustomerStats AS (

SELECT

`Customer ID`,

MAX(Date) AS LastPurchaseDate,

COUNT(\*) AS Frequency,

SUM(`Total Amount`) AS Monetary

FROM retail

GROUP BY `Customer ID`

)

SELECT

cs.`Customer ID`,

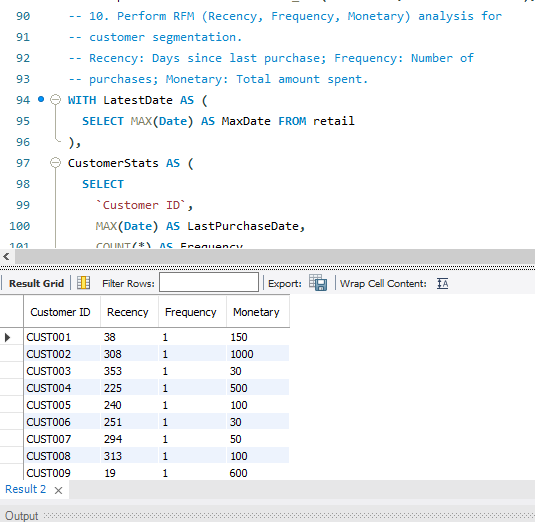
DATEDIFF(ld.MaxDate, cs.LastPurchaseDate) AS Recency,

cs.Frequency,

cs.Monetary

FROM CustomerStats cs

JOIN LatestDate ld;



11. Find the product categories with the highest average quantity per

transaction.

Which product types are purchased in bulk?

SELECT

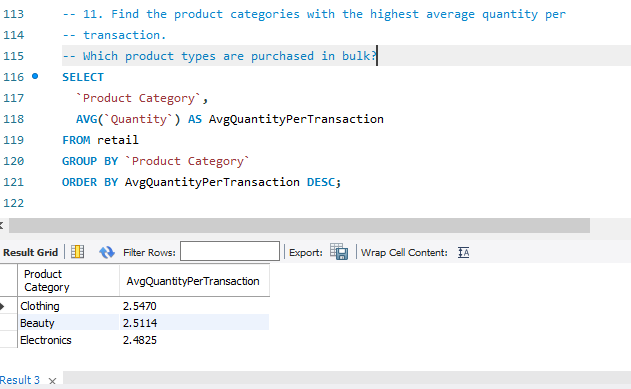
`Product Category`,

AVG(`Quantity`) AS AvgQuantityPerTransaction

FROM retail

GROUP BY `Product Category`

ORDER BY AvgQuantityPerTransaction DESC;



12. Identify the busiest sales day of the week.

Which day(s) consistently have the highest transaction volume or

revenue?

SELECT

DAYNAME(`Date`) AS DayOfWeek,

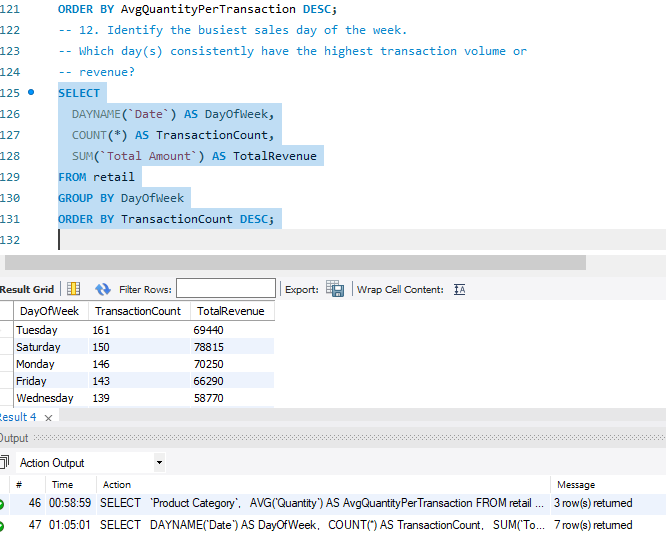
COUNT(\*) AS TransactionCount,

SUM(`Total Amount`) AS TotalRevenue

FROM retail

GROUP BY DayOfWeek

ORDER BY TransactionCount DESC;



13. Calculate total revenue and average spend per transaction by

gender.

Are there differences in spending patterns across genders?

SELECT

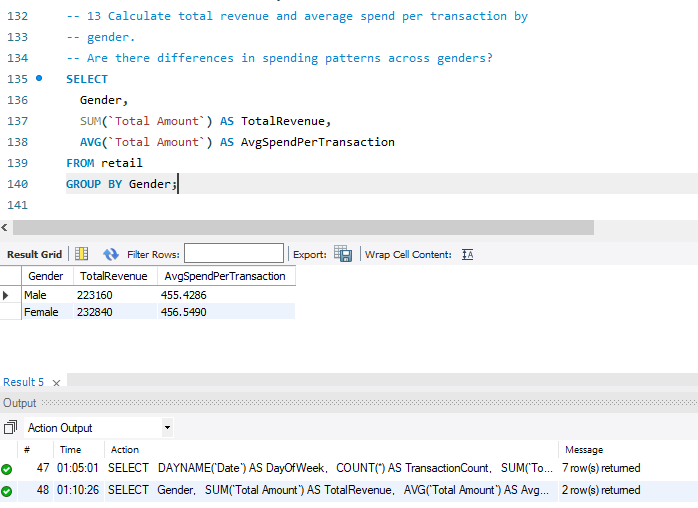
Gender,

SUM(`Total Amount`) AS TotalRevenue,

AVG(`Total Amount`) AS AvgSpendPerTransaction

FROM retail

GROUP BY Gender;



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

Based on number of transactions involving each category.

SELECT

`Product Category`,

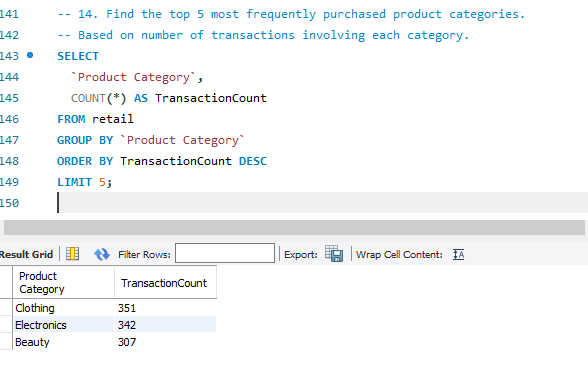
COUNT(\*) AS TransactionCount

FROM retail

GROUP BY `Product Category`

ORDER BY TransactionCount DESC

LIMIT 5;



15. Determine the percentage of total revenue contributed by each

age group.

Which customer age brackets are most valuable to the business?

WITH AgeBracketed AS (

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 BETWEEN 51 AND 65 THEN '51–65'

ELSE 'Other'

END AS AgeGroup

FROM retail

),

GroupRevenue AS (

SELECT AgeGroup, SUM(`Total Amount`) AS Revenue

FROM AgeBracketed

GROUP BY AgeGroup

),

TotalRevenue AS (

SELECT SUM(`Total Amount`) AS Total FROM retail

)

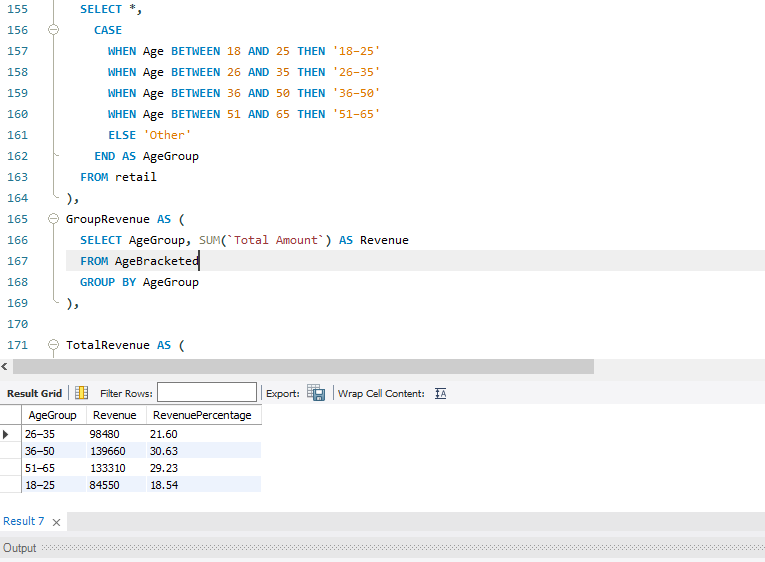
SELECT

gr.AgeGroup,

gr.Revenue,

ROUND((gr.Revenue / tr.Total) \* 100, 2) AS RevenuePercentage

FROM GroupRevenue gr, TotalRevenue tr;



This project provided valuable insights into customer behavior, product trends, and sales patterns through SQL-based analysis of a retail sales dataset. By performing targeted queries in MySQL Workbench, I uncovered several key findings:

**Inactive Customers**: A significant number of customers have not made purchases in the last 6 months, highlighting opportunities for re-engagement campaigns.

**RFM Analysis**: Segmenting customers based on recency, frequency, and monetary value helped identify high-value and loyal customers, as well as those at risk of churn.**Bulk Purchases**: Certain product categories showed high average quantities per transaction, suggesting they are commonly purchased in bulk—ideal for discount or bundle strategies

**Sales Patterns**: The busiest day of the week was identified, providing insight for resource planning and promotional timing.

**Spending by Gender**: Revenue and average transaction amounts were analyzed by gender, revealing subtle differences in purchasing behavior that can support targeted marketing.

**Top Product Categories**: The five most frequently purchased categories were determined, allowing the business to focus on high-demand inventory.

**Revenue by Age Group**: Younger and middle-aged customer segments contributed the largest share of total revenue, helping refine customer targeting and engagement strategies and many others.

Overall, this analysis supports data-driven decision-making in areas such as marketing, inventory management, and customer relationship management. Future work could incorporate time-series analysis or predictive models to forecast trends and personalize offers more effectively thank.