**RETAIL SALES**

**columns**

* transactions\_id – Unique transaction identifier
* sale\_date, sale\_time – Transaction time
* customer\_id – Unique customer
* gender, age – Customer demographics
* category – Product category
* quantiy – Number of units sold
* price\_per\_unit – Price of one unit
* cogs – Cost of goods sold
* total\_sale – Final transaction value

**Identified KPIs:**

* Total Sales (total\_sale)
* Revenue by Category
* Quantity Sold
* Customer Count
* Demographic-based sales (age, gender)

**Data Cleaning Performed**

Missing Values

* Removed rows with missing critical business values to ensure accuracy

Duplicate Records

* Removed all duplicate records

Incorrect Data Types

* Converted numeric columns to proper numeric data types

**Data Validation**

* No null values remaining
* No duplicate rows
* Numeric columns validated
* Dataset ready for KPI analysis and modeling

**Summary**

Issues Found:

* Missing values in age, quantity, pricing, and sales columns
* Duplicate records
* Improper data types

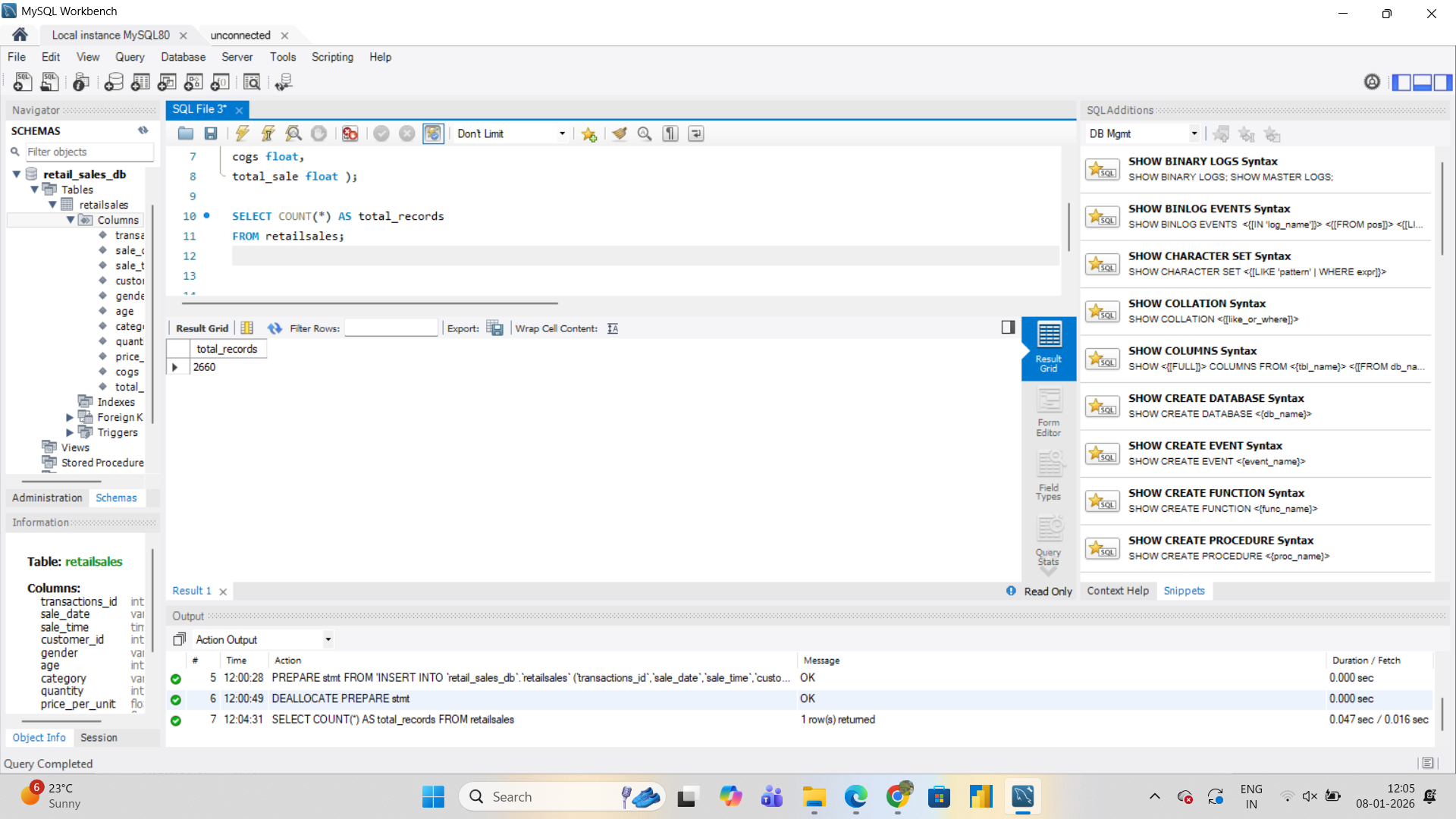
How They Were Fixed:

* Removed rows with missing critical values
* Eliminated duplicates
* Converted columns to correct data types

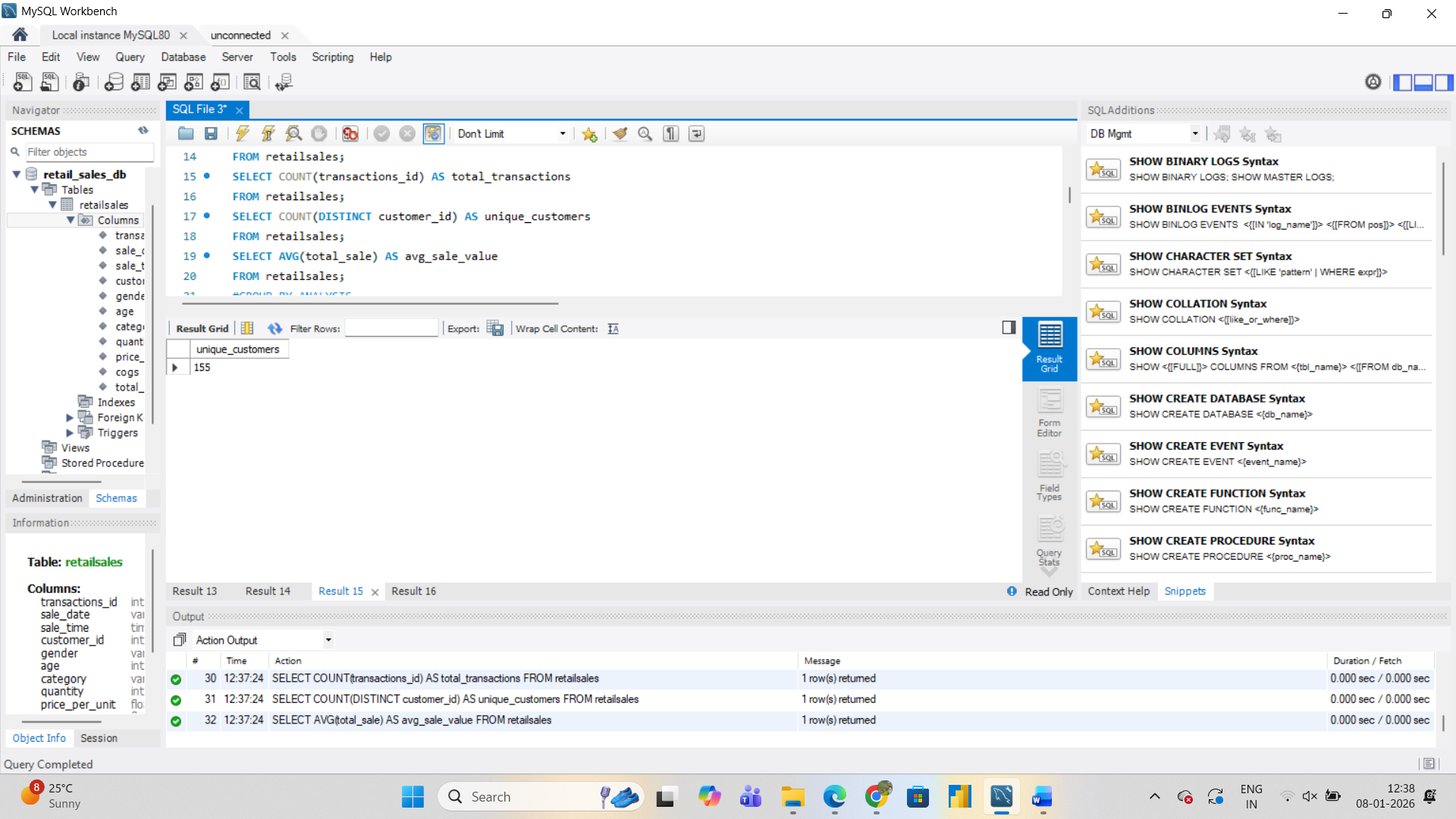
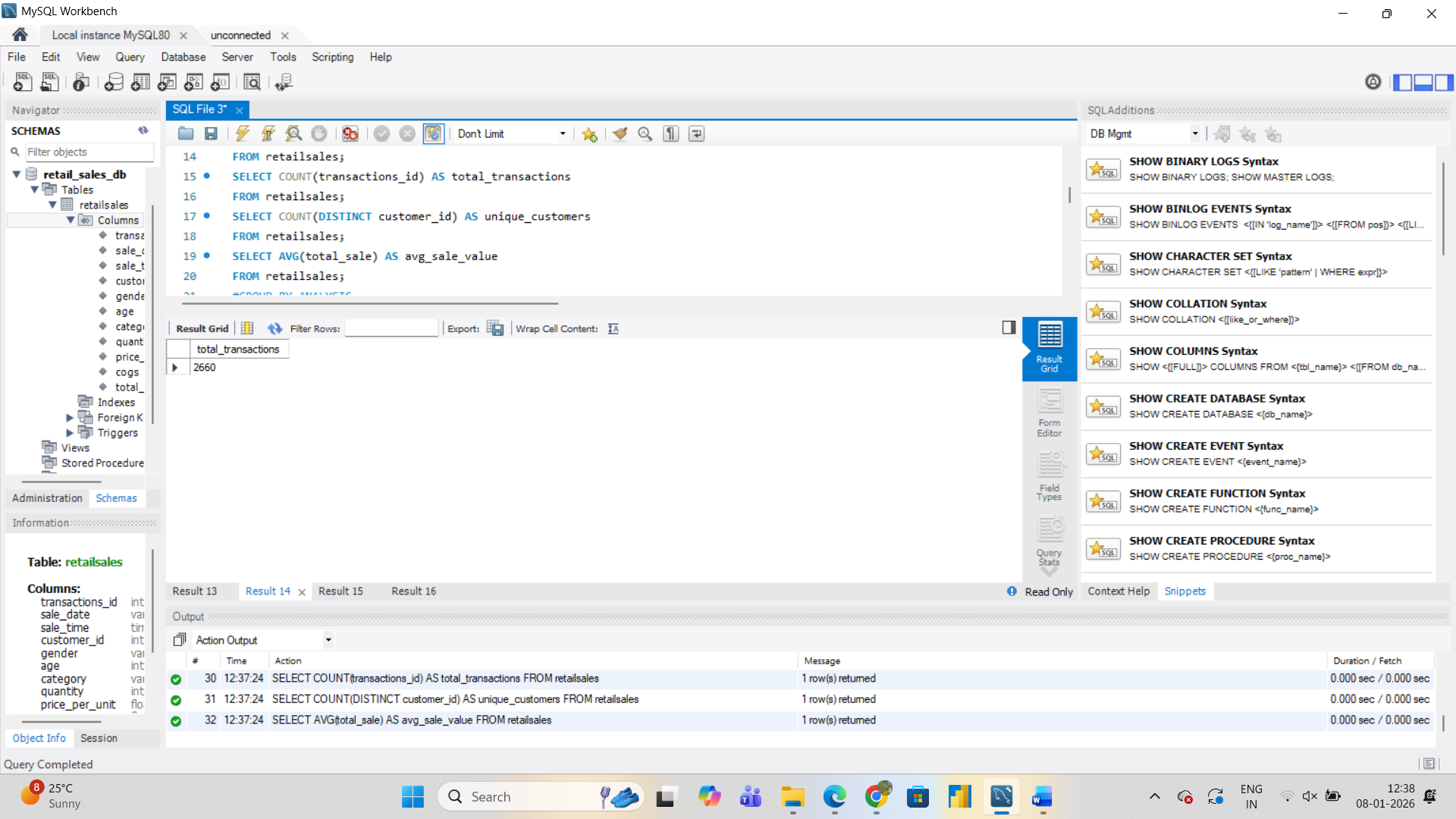
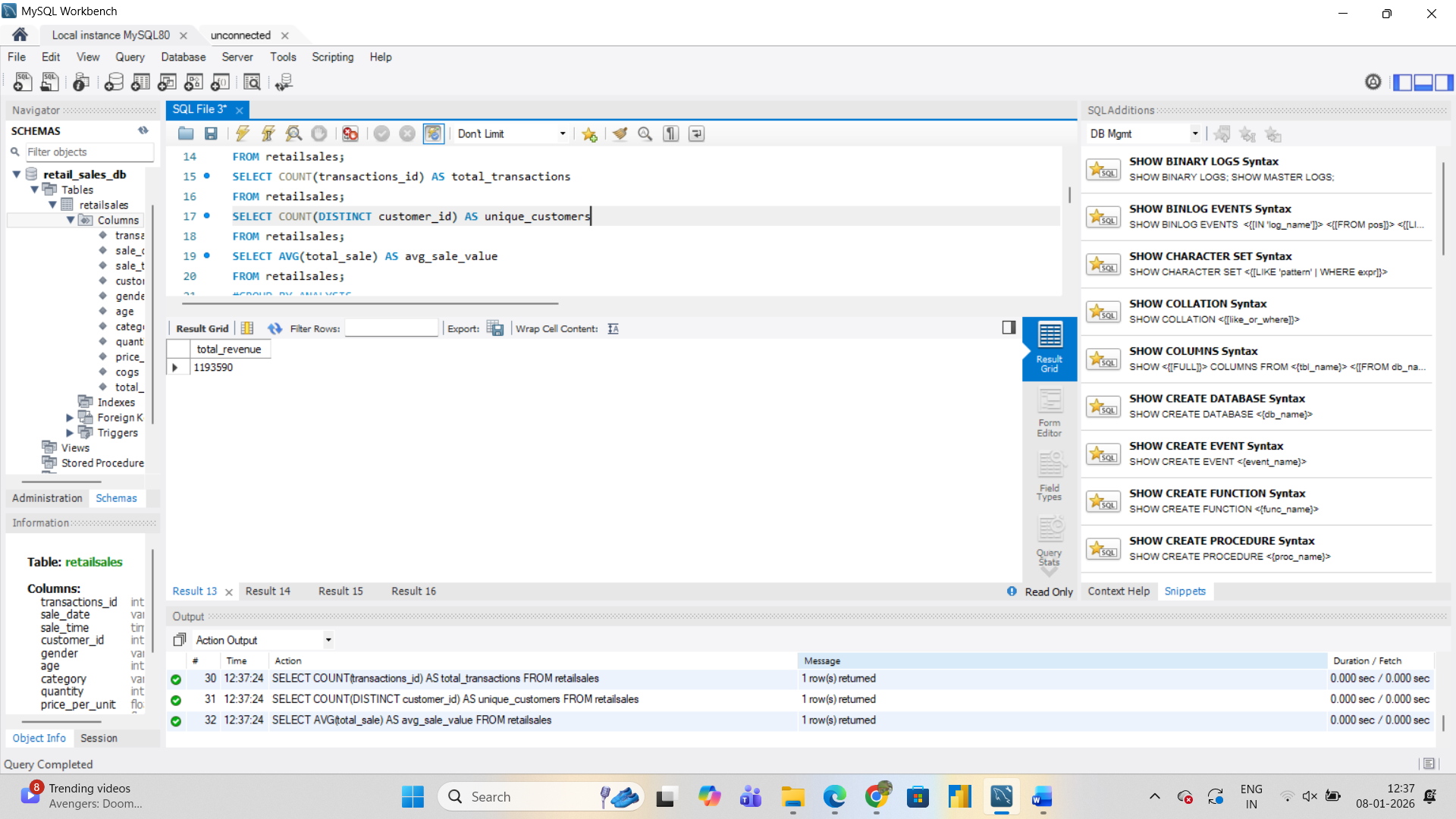
**SQL Analysis**

Record Counts and Unique Values

* Calculated the total number of transactions to understand dataset size.
* Identified unique customers to measure customer reach.

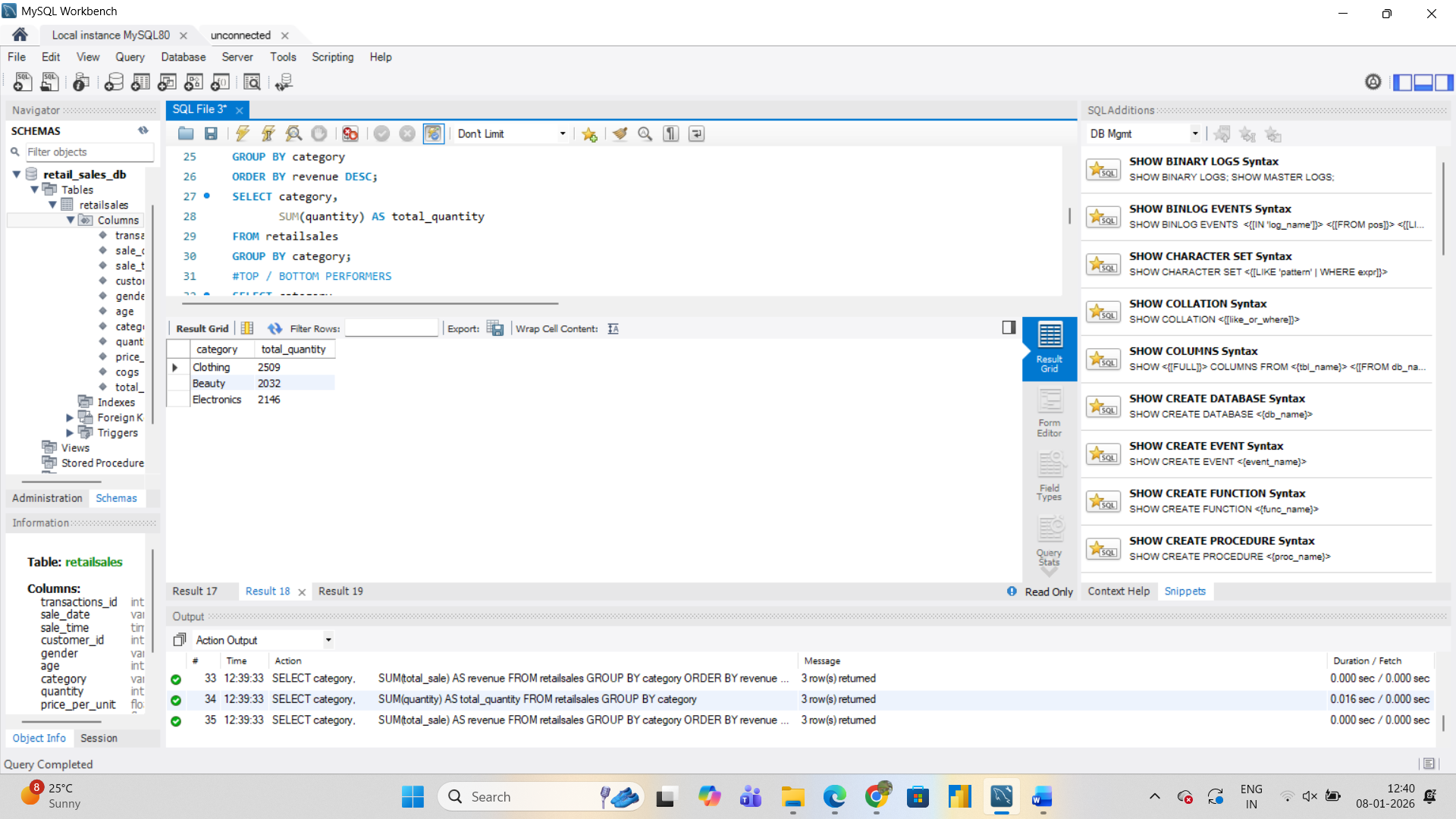


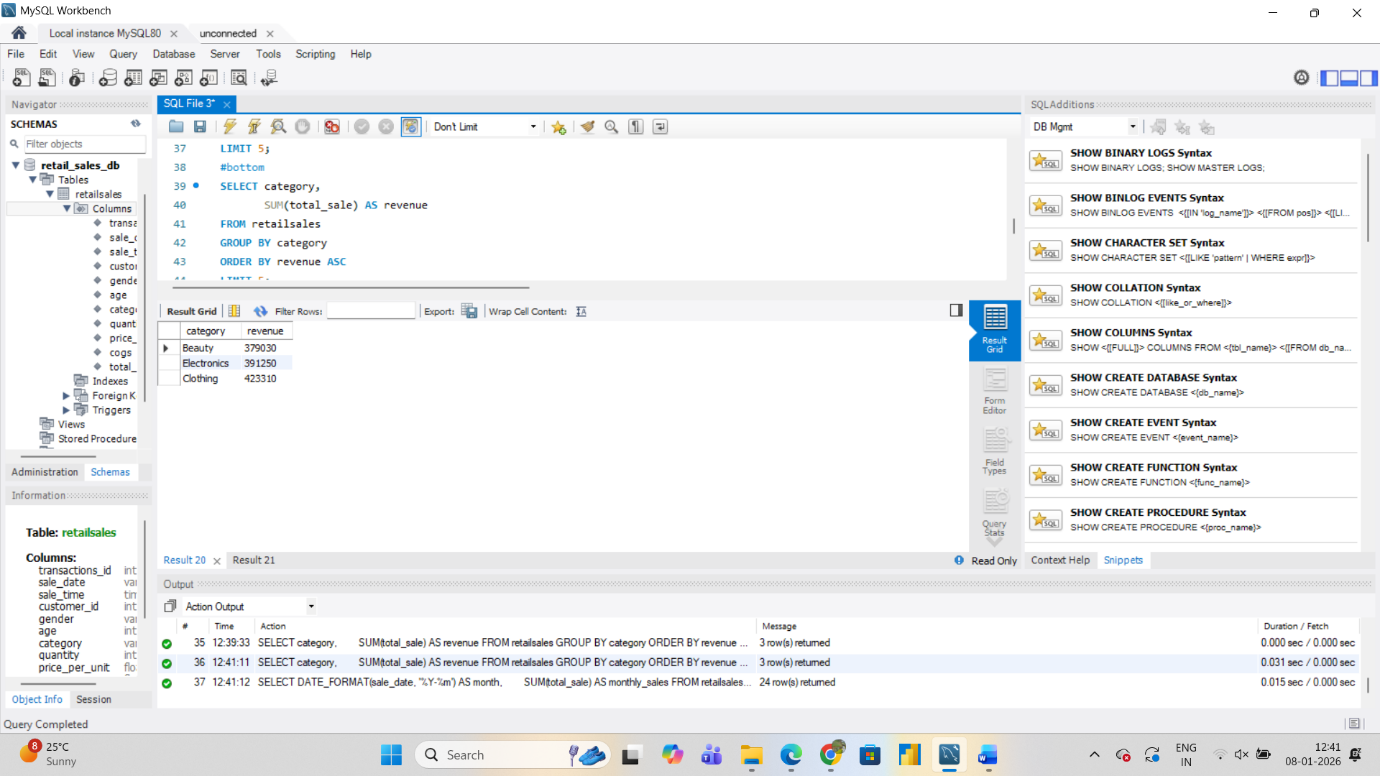
Aggregation Analysis

* Used SUM to calculate total revenue.
* Used AVG to find the average sale value.
* Used COUNT to measure transaction volume.  
    
  ****

GROUP BY Analysis

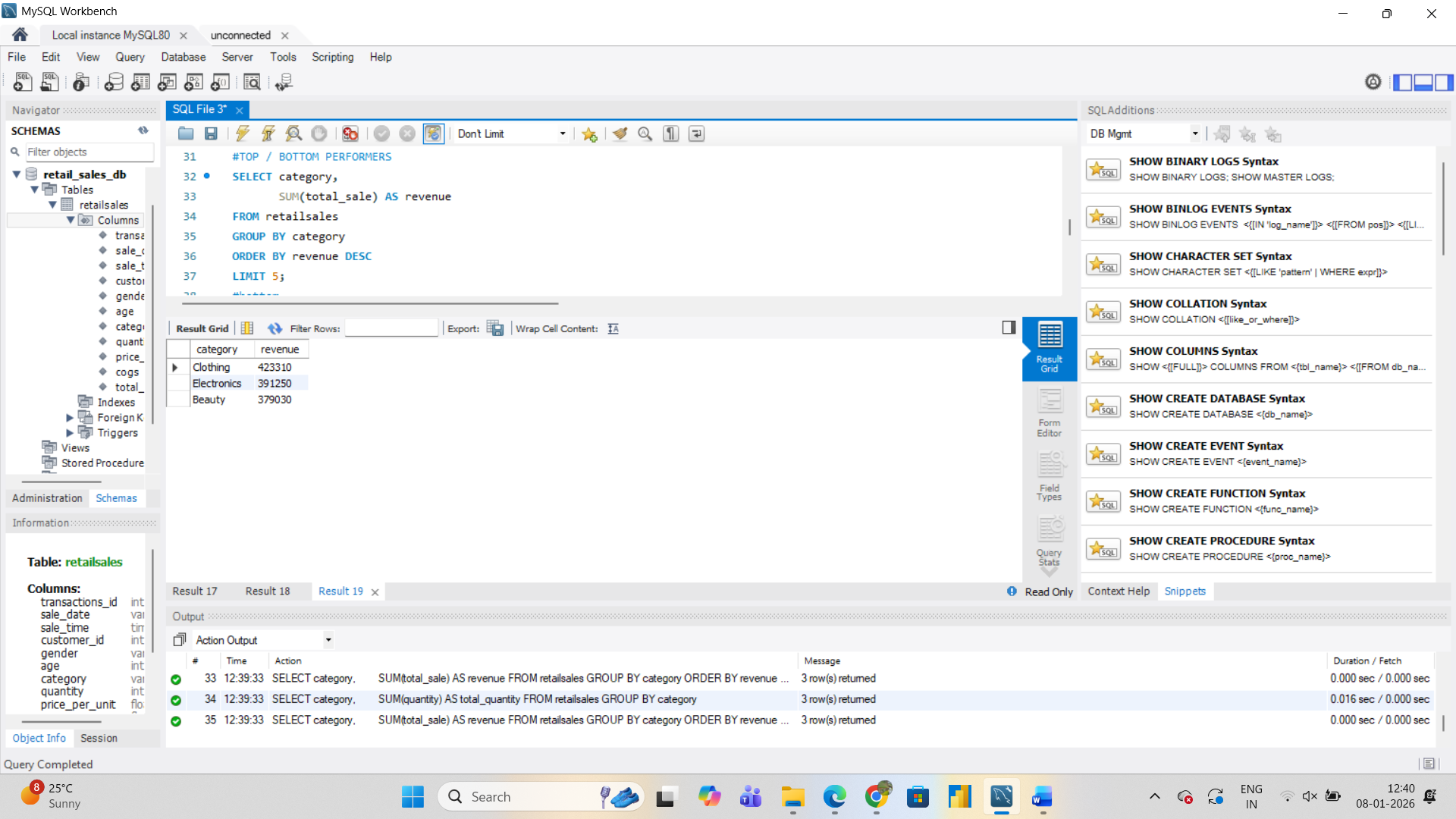
Grouped data by product category to analyze: Quantitysold per category and Revenue contribution by category

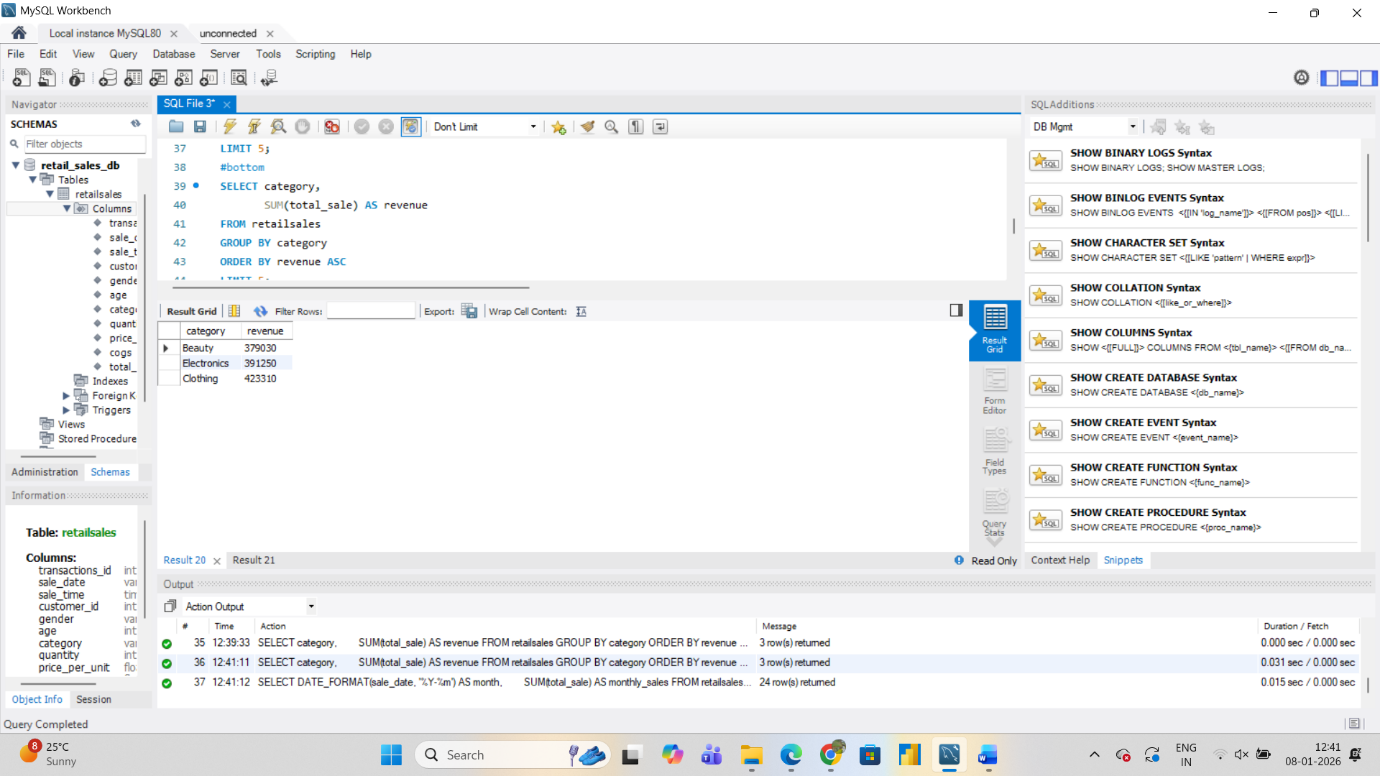
****

****

Top and Bottom Performers

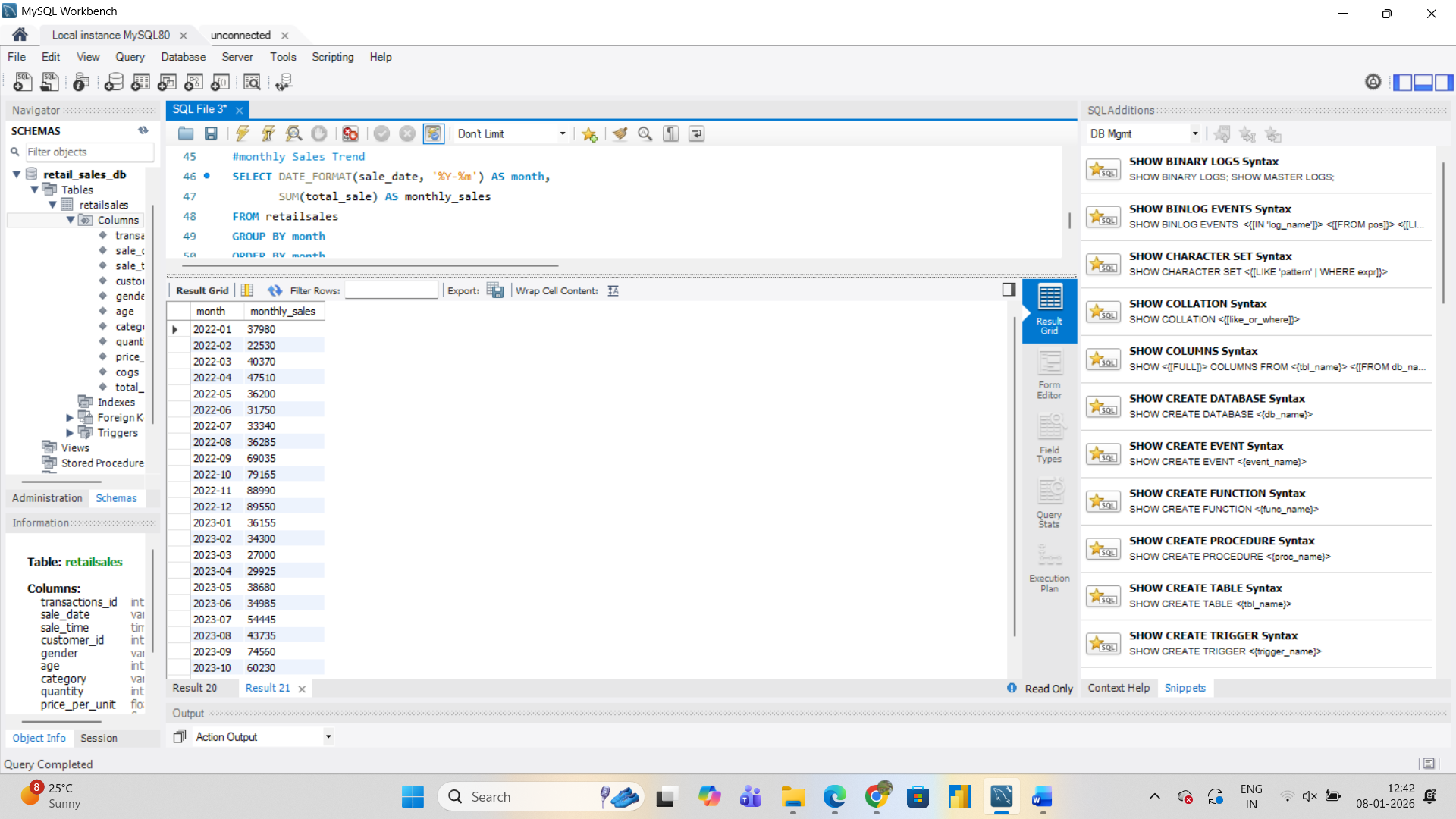
* Identified top 5 categories by revenue to understand high-performing products.
* Identified bottom 5 categories by revenue to detect underperforming segments.

****

****

Date-Based Analysis

* Performed monthly sales trend analysis using the sale date.

****

BUSINESS INSIGHTS

The analysis shows that a few product categories contribute the majority of revenue. Monthly sales follow a consistent trend, indicating stable customer demand. These insights can help optimize inventory and marketing strategies.

**Descriptive statistics**

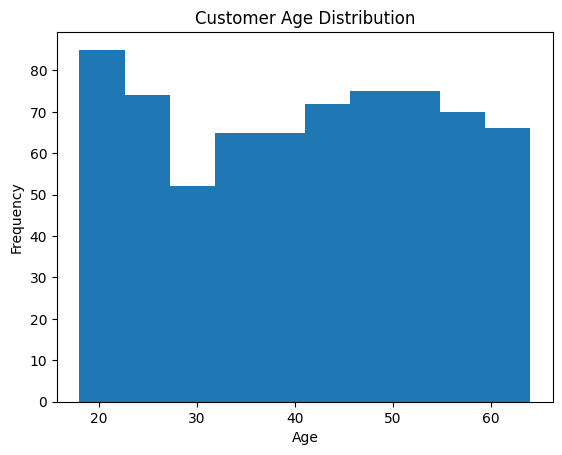
Data shows realistic retail behavior with middle-aged customers, low quantities per purchase, and high variation in total sales.

**Correlation analysis**

Total sales are strongly correlated with price per unit and cost, while age and IDs have negligible influence.

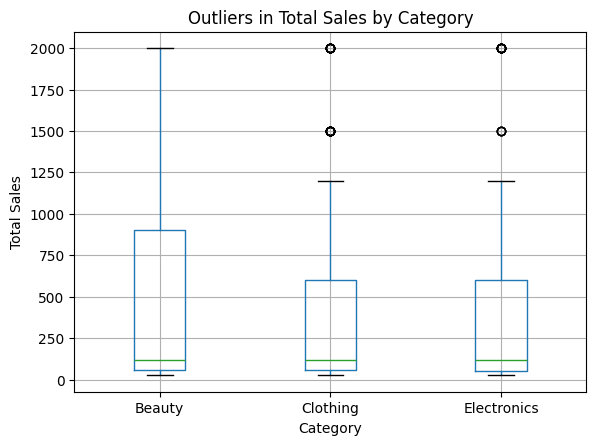
**Distribution analysis**

Sales and customer spending are right-skewed, indicating many small purchases and few high-value transactions.

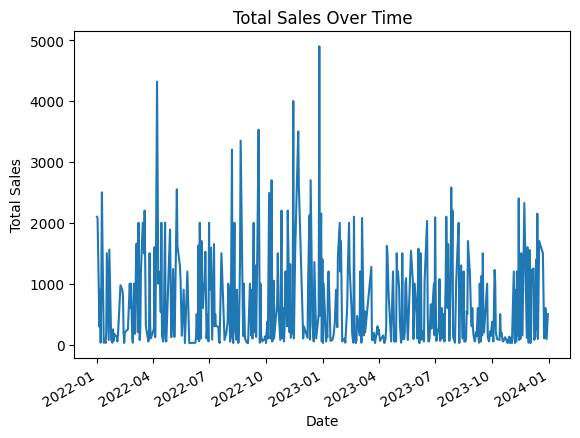


Histogram – Age distribution  
The age distribution shows most customers fall in the middle-aged group, indicating a dominant adult customer base.

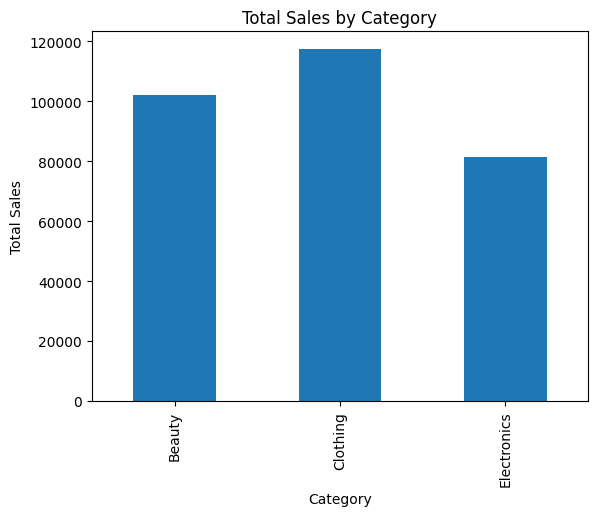
Box plot – Outliers in total sales by category  
The box plot reveals significant high-value outliers across categories, representing bulk or premium purchases.



Line chart – Sales trend over time  
The sales trend fluctuates over time, suggesting seasonal demand or promotional impacts.

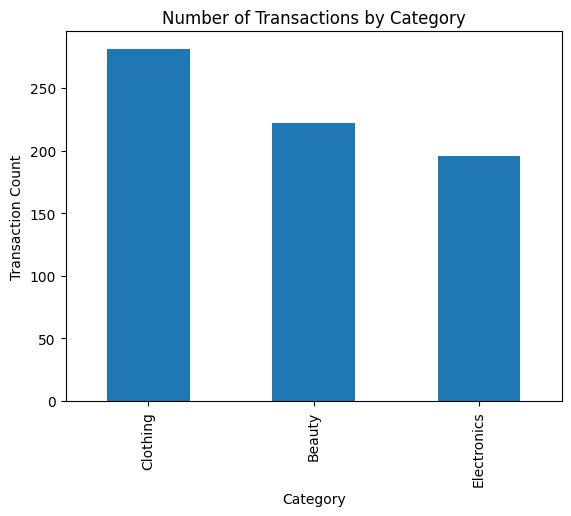


Bar chart – Sales by category  
The bar chart shows that the Clothing category generates the highest total sales among all categories.

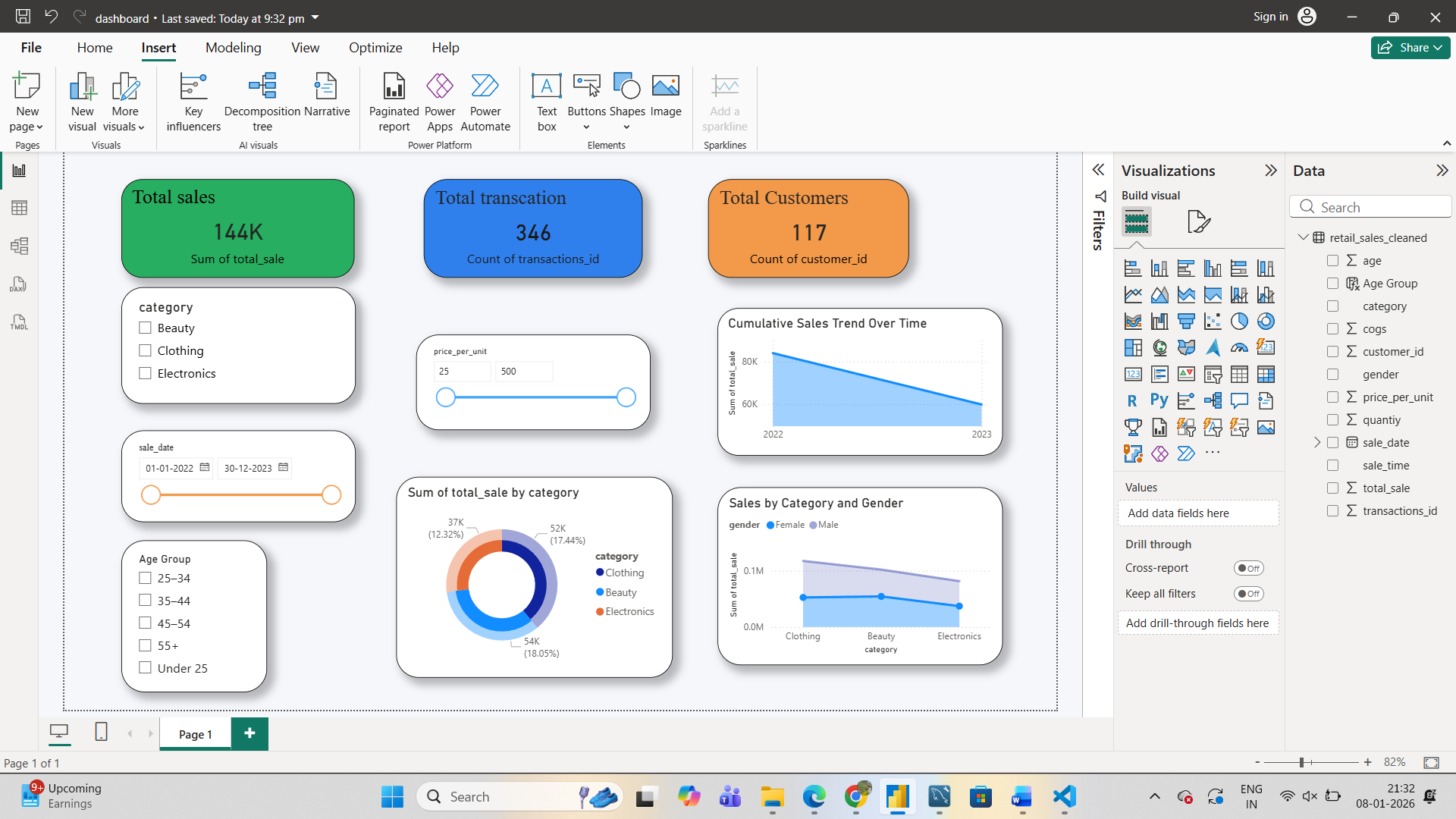


Bar chart – Average quantity purchased by gender

Average quantity purchased varies slightly by gender, indicating similar buying behavior across genders.



Dashboard



The dashboard was developed to analyze retail sales performance and extract meaningful insights using interactive visualizations and key performance indicators. It combines KPI cards, trend analysis, category-wise comparisons, and outlier detection to provide a comprehensive view of the dataset.

The analysis shows that sales fluctuate over time, indicating variations in demand that may be influenced by seasonal patterns or promotional activities. Among the product categories, Clothing consistently generates the highest total sales, highlighting it as the primary revenue contributor. Customer age distribution indicates that middle-aged customers form the majority of the customer base.

Spending behavior analysis reveals that total sales are strongly driven by product pricing rather than customer demographics. Correlation and visual analysis confirm that higher-priced products contribute significantly to revenue. Most transactions involve small purchase quantities, while a limited number of high-value transactions account for a disproportionate share of total sales.

Outlier analysis using box plots highlights the presence of unusually high sales values across categories, which are likely due to bulk or premium purchases. Despite these anomalies, overall purchasing patterns remain consistent across genders, showing only minor variation in average quantity purchased.

Based on the insights, business efforts should focus on strengthening high-performing categories such as Clothing while introducing targeted strategies during low-sales periods. Monitoring premium transactions and customer segments can further support data-driven decision-making. Overall, the dashboard effectively converts raw sales data into actionable insights through clear and interactive visual storytelling.