1 **What is the difference between WHERE and HAVING?**

Answer:

- WHERE filters **rows before grouping**
- HAVING filters **groups after GROUP BY**

Example from my task:

-- WHERE (filters rows)

SELECT *

FROM retail_sales_dataset

WHERE Gender = 'Female';

-- HAVING (filters grouped results)

SELECT `Product Category`, SUM(`Total Amount`)

FROM retail_sales_dataset

GROUP BY `Product Category`

HAVING SUM(`Total Amount`) > 1000;

WHERE works on rows, HAVING works on aggregated results.

---

2 **How does GROUP BY work?**

Answer:

GROUP BY groups rows that have the same value in a column and allows aggregate functions like SUM, AVG, and COUNT to be applied to each group.

Example:

SELECT `Product Category`,

    SUM(`Total Amount`) AS total_sales

FROM retail_sales_dataset

GROUP BY `Product Category`;

This groups all rows by **Product Category** and calculates **total sales per category**.

---

3 **When do you use ORDER BY?**

Answer:

ORDER BY is used to **sort the result set** in ascending (ASC) or descending (DESC) order.

Example:

SELECT *

FROM retail_sales_dataset

ORDER BY `Total Amount` DESC;

This shows **highest sales first**, which is useful for ranking and analysis.

4️⃣ What happens if a column has NULL values in SUM / AVG?

Answer:

- SUM() and AVG() **ignore NULL values**

- They only calculate using **non-NULL values**

- If all values are NULL, the result is NULL

This prevents calculation errors and ensures accurate results.

5️⃣ Write a query to find top 5 customers by total spend

✅ **Correct answer for your dataset** (uses Customer ID):

SELECT `Customer ID`,

    SUM(`Total Amount`) AS total_spent

FROM retail_sales_dataset

GROUP BY `Customer ID`

ORDER BY total_spent DESC

LIMIT 5;

This:

- Groups data by customer

- Calculates total spend

- Sorts highest spenders first

- Returns **top 5 customers**