**Some Basic Questions based on MegaMart dataset**

**Basic Transformations**

* Convert order\_date from string to a DateType and extract year, month, day.
* Create a new column total\_price = quantity \* price\_per\_unit.
* Standardize categorical values (e.g., ensure order\_status has consistent formatting like "Delivered" vs "delivered").
* Filter all orders where order\_status = 'Delivered'.
* Replace nulls (if any in extended dataset) in payment\_method with "Unknown".

**Aggregations**

* Find the total revenue generated per product\_category.
* Count the number of orders per payment method.
* Find the top 5 products by sales (total\_price).
* Calculate the average order value per user\_id.
* Find the category with maximum cancelled orders.

**Joins & Grouping**

1. **Group orders by order\_status and payment\_method to see patterns (e.g., which payment mode has most cancellations).**
2. **Find users who purchased products from multiple categories.**
3. **Identify repeat customers (user\_id with more than one order).**

**Window Functions**

1. **Rank products within each product category by revenue.**
2. **For each user, calculate the running total spent over time.**
3. **Find the latest order per user using row number or rank.**

**Advanced Transformations**

1. **Calculate daily sales trend (group by order date).**
2. **Find the return rate = Returned Orders / Total Orders per category.**
3. **Detect high-value customers (users who spent more than 5000 total).**
4. **Pivot table: show categories as columns and count of delivered orders per month.**