

## Project Overview

This project focuses on analyzing transactional data from a retail store using SQL. The analysis ranges from data preparation to data exploration and business insights.

### 1. Table Creation

A table named 'retail\_sales' was created with the following columns:

- transactions\_id (Primary Key)
- sale\_date, sale\_time
- customer\_id, gender, age
- category, quanti (typo, should be 'quantity')
- price\_per\_unit, cogs, total\_sale (FIOAT should be FLOAT)

Recommendation: Add NOT NULL constraints and correct column spelling/types to ensure data accuracy.

### 2. Data Cleaning

Performed checks for missing values in critical fields like transaction ID, date, time, gender, etc. All rows with NULL values were identified and handled.

### 3. Data Exploration

- Total Sales Count: Counted all transactions.
- Unique Customers: Counted distinct customer\_id entries.
- Categories: Extracted unique product categories for segmentation.

### 4. Business Questions & Insights

- Q1: Sales on '2022-11-05' - Filtered transactions for that date.
- Q2: Clothing sales with quantity  $\geq 4$  in Nov 2022 - Filtered by date and category.
- Q3: Total sales & order count per category - Used SUM and COUNT.
- Q4: Average age of Beauty category customers - Used AVG(age).
- Q5: High-value transactions - total\_sale  $> 1000$ .
- Q6: Sales count by gender & category - Used GROUP BY.
- Q7: Best month per year by avg. sales - Used RANK() function.
- Q8: Top 5 customers by sales - SUM(total\_sale), ORDER BY DESC.
- Q9: Unique customers per category - COUNT DISTINCT.
- Q10: Shift-wise order distribution - CASE WHEN logic based on time.

### Key Findings

- Top Sales Categories can guide targeted marketing.
- Customer Demographics aid segmentation.
- Sales Timing insights help with staffing and inventory.
- Top Customers highlight loyalty program opportunities.

### Improvements & Next Steps

- Fix column names (quanti -> quantity) and types.
- Use DECIMAL for financial columns.
- Add indexes for performance.
- Consider advanced metrics like profit, CLV, and cohort analysis in future phases.