# SQL Techniques & Skills Demonstrated

This project demonstrates a comprehensive application of SQL for data analysis using MySQL Workbench. The following SQL concepts and techniques were applied throughout the project to ensure data integrity, reusability, and meaningful business insights.

**Database Design & Data Handling**

- Created a structured relational database and defined a sales table with appropriate data types and **NOT NULL constraints** to prevent missing or invalid data.
- Imported transactional data from a CSV file into MySQL and validated successful data ingestion using aggregate and sample queries.

**Data Wrangling & Cleaning**

- Performed data inspection to ensure data consistency and completeness.
- Handled potential data quality issues at the schema level using constraints rather than post-import filtering.

**Feature Engineering**

- Engineered new analytical columns using SQL expressions and conditional logic:
    - time_of_day: Categorized sales into Morning, Afternoon, and Evening using **CASE statements** and time-based conditions.
    - day_name: Extracted weekdays from transaction dates using **date functions**.
    - month_name: Extracted month names to enable monthly sales and profit trend analysis.
- These derived features enabled deeper time-based and behavioral analysis without modifying the raw data source.

**Exploratory Data Analysis (EDA)**

- Conducted extensive EDA using SQL to answer business-driven questions related to:
    - Product performance
    - Customer behavior
    - Sales trends
    - Branch and city-level revenue distribution
- Applied **aggregate functions** such as SUM(), AVG(), and COUNT() combined with GROUP BY and HAVING clauses to summarize and compare data.

**Advanced SQL Analysis**

- Utilized **Common Table Expressions (CTEs)** to structure complex queries for improved readability and modularity.

- Applied **window functions** such as RANK() with PARTITION BY to identify top-performing days per branch based on customer ratings.
- Implemented **CASE expressions** for conditional classification of product performance (e.g., Good vs Bad based on average sales).

**Views & Query Optimization**

- Created reusable **SQL views** to simplify repeated analytical queries and enhance maintainability.
- Views were used to abstract complex aggregations and present summarized insights in a clean and reusable format.
- Implemented **indexes** on frequently queried columns to demonstrate awareness of query performance optimization.

**SQL Functions & Concepts Used**

- **Aggregate Functions:** SUM(), AVG(), COUNT(), MAX()
- **Conditional Logic:** CASE WHEN
- **Date & Time Functions:** DAYNAME(), MONTHNAME(), MONTH(), HOUR()
- **Window Functions:** RANK() OVER (PARTITION BY … ORDER BY …)
- **Joins & Subqueries:** Used for comparative and conditional analysis
- **CTEs (WITH clause):** For structured and readable complex queries
- **Views:** For reusability and abstraction

This project showcases practical SQL proficiency by applying real-world data analysis techniques rather than focusing solely on query syntax.

# Questions Answered:

**Generic Question**

1. How many unique cities does the data have?
2. In which city is each branch?

**Product**

1. How many unique product lines does the data have?
2. What is the most common payment method?
3. What is the most selling product line?
4. What is the total revenue by month?
5. What month had the largest COGS?
6. What product line had the largest revenue?
7. What is the city with the largest revenue?
8. What product line had the largest VAT?
9. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales
10. Which branch sold more products than average product sold?
11. What is the most common product line by gender?
12. What is the average rating of each product line?

**Sales**

1. Number of sales made in each time of the day per weekday
2. Which of the customer types brings the most revenue?
3. Which city has the largest tax percent/ VAT (**Value Added Tax**)?
4. Which customer type pays the most in VAT?

**Customer**

1. How many unique customer types does the data have?
2. How many unique payment methods does the data have?
3. What is the most common customer type?
4. Which customer type buys the most?
5. What is the gender of most of the customers?
6. What is the gender distribution per branch?

7. Which time of the day do customers give most ratings?
8. Which time of the day do customers give most ratings per branch?
9. Which day fo the week has the best avg ratings?
10. Which day of the week has the best average ratings per branch?