Task 4: SQL for Data Analysis

Objective: Use SQL queries to extract and analyze data from a database.
Tools: MySQL / PostgreSQL / SQLite

Example Queries:

```
1    SELECT customer_name, country FROM customers WHERE country = 'India' ORDER BY
     customer_name ASC;
2    SELECT country, COUNT(*) AS total_customers FROM customers GROUP BY country;
3    SELECT c.customer_name, o.order_id, o.order_date FROM customers c INNER JOIN
     orders o ON c.customer_id = o.customer_id;
4    SELECT product_name, price FROM products WHERE price > (SELECT AVG(price) FROM
     products);
5    CREATE VIEW sales_summary AS SELECT p.product_name, SUM(oi.quantity * p.price)
     AS total_sales FROM order_items oi JOIN products p ON oi.product_id =
     p.product_id GROUP BY p.product_name;
6    CREATE INDEX idx_customer_country ON customers(country);
```

GitHub Submission Structure:

■ dataset/ → Dataset file
■ screenshots/ → Output screenshots
■ data_analysis.sql → SQL queries
■ README.md → Explanation and outcomes

Interview Questions & Answers:

- **1. What is the difference between WHERE and HAVING?** WHERE filters rows before grouping, HAVING filters groups after aggregation.
- **2. What are the different types of joins?** INNER, LEFT, RIGHT, and FULL JOIN (if supported).
- **3. How do you calculate average revenue per user?** By summing revenue per user and taking the average using AVG().
- **4. What are subqueries?** Queries inside another query to fetch intermediate results.
- **5. How do you optimize a SQL query?** Use indexes, avoid SELECT *, limit results, and use EXPLAIN to analyze performance.
- **6. What is a view in SQL?** A virtual table created by saving a query.
- **7. How would you handle null values in SQL?** Use COALESCE(column, default_value) or IS NULL / IS NOT NULL.