Task 3: SQL for Data Analysis

Sample SQL Queries

1. Total Orders

SELECT COUNT(*) AS total_orders FROM orders;

2. Total Revenue

SELECT SUM(price + freight_value) AS total_revenue FROM order_items;

3. Average Order Value

SELECT AVG(price + freight_value) AS avg_order_value FROM order_items;

4. Orders by Payment Type

SELECT payment_type, COUNT(*) AS count FROM order_payments GROUP BY payment_type;

5. Top 5 States by Number of Customers

SELECT customer_state, COUNT(*) AS num_customers FROM customers GROUP BY customer_state ORDER BY num_customers DESC LIMIT 5;

6. Top 5 Selling Products

SELECT product_id, COUNT(*) AS units_sold FROM order_items GROUP BY product_id ORDER BY units_sold DESC LIMIT 5;

7. Create View: Monthly Revenue

```
CREATE VIEW monthly_revenue AS
    SELECT strftime('%Y-%m', orders.order_purchase_timestamp) AS month,
        SUM(order_items.price + order_items.freight_value) AS revenue
    FROM orders
    JOIN order_items ON orders.order_id = order_items.order_id
    GROUP BY month;
```

Interview Questions & Answers

1. What is the difference between WHERE and HAVING?

WHERE filters rows before aggregation, whereas HAVING filters after aggregation.

2. What are the different types of joins?

INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL OUTER JOIN.

3. How do you calculate average revenue per user in SQL?

Use GROUP BY to sum revenue per user, then take the average:

SELECT AVG(user_total) FROM (SELECT customer_id, SUM(price) AS user_total FROM order_items GROUP BY customer_id);

4. What are subqueries?

A subquery is a query nested inside another SQL query.

5. How do you optimize a SQL query?

Use indexes, avoid SELECT *, limit data scans, and analyze query plans.

6. What is a view in SQL?

A view is a virtual table created by a query; it does not store data itself.

7. How would you handle null values in SQL?

Use COALESCE() or IS NULL/IS NOT NULL conditions.