8. Hive: Queries, aggregation

You are a data analyst working for a fictional company called "TechMart." TechMart operates several retail stores across the country and wants to analyze their sales data stored in Hive to gain insights into their business performance. Your task is to create a Hive table, populate it with dummy sales data, and perform some basic queries to analyze the sales trends.

Task 1: Creating the Hive Table

```
CREATE TABLE IF NOT EXISTS sales (
    transaction_id INT,
    product_name STRING,
    category STRING,
    price DOUBLE,
    quantity INT,
    transaction_date DATE
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';
```

Task 2: Inserting Dummy Data

```
INSERT INTO sales VALUES
(1, 'Laptop', 'Electronics', 1200.00, 10, '2024-01-01'),
(2, 'Smartphone', 'Electronics', 800.00, 15, '2024-01-02'),
(3, 'Headphones', 'Electronics', 50.00, 20, '2024-01-03'),
(4, 'T-Shirt', 'Apparel', 20.00, 50, '2024-01-04'),
(5, 'Jeans', 'Apparel', 40.00, 30, '2024-01-05');
```

Task 3: Find Total sales amount per category

```
SELECT category, SUM(price * quantity) AS total_sales_amount
FROM sales
GROUP BY category;
```

Task 4: Find top-selling products

```
SELECT product_name, SUM(quantity) AS total_quantity_sold
FROM sales
GROUP BY product_name
ORDER BY total_quantity_sold DESC
LIMIT 3;
```

Task 5: Get the monthly sales trend

```
SELECT
   MONTH(transaction_date) AS month,
   YEAR(transaction_date) AS year,
   SUM(price * quantity) AS total_sales_amount
FROM
   sales
GROUP BY
   YEAR(transaction_date), MONTH(transaction_date)
ORDER BY
   year, month;
```