Assessment Task:

1. **Explore this dataset and highlight key insights in terms of aggregate and drill down of this sales data.**

2. **Share SQL scripts used to explore data and generate metrics & Key insights in bullet points via GitHub link**

I'll first explored the dataset to understand its structure, including column names, data types, and missing values. Then, I'll extract key insights and provide SQL scripts for analysis.

Dataset Overview:

- **Total Records:** 500,000
- **Columns:** 19
- **Missing Values:** Present in CustomerID, TransactionDate, PaymentMethod, StoreType, CustomerAge, CustomerGender, ProductName, and Region.

Next Steps:

1. Convert TransactionDate to a proper datetime format.
2. Check for duplicate records.
3. Perform key aggregations and insights extraction.

Data Cleaning Observations:

- **No Duplicate Records Found.**
- **Missing Values:**
    - CustomerID, TransactionDate, PaymentMethod, StoreType, CustomerAge, CustomerGender, ProductName, and Region have missing values (up to 50,000 in some cases).

Key Insights:

- **Total Sales:** ₹10,202,662,960.19
- **Average Sales per Transaction:** ₹20,405.33
- **Total Quantity Sold:** 3,747,755 units
- **Average Discount Applied:** 25%
- **Most Sold Product:** Notebook
- **City with Highest Sales:** Kolkata

Now, I'll generate SQL queries to derive these insights.

CODE:

```
-- Total Sales
SELECT SUM(TransactionAmount) AS Total_Sales FROM sales_data;

-- Average Sales per Transaction
SELECT AVG(TransactionAmount) AS Average_Sales FROM sales_data;

-- Total Quantity Sold
SELECT SUM(Quantity) AS Total_Quantity FROM sales_data;

-- Average Discount Applied
SELECT AVG(DiscountPercent) AS Average_Discount FROM sales_data;

-- Most Sold Product
SELECT ProductName, COUNT(*) AS Product_Count
FROM sales_data
GROUP BY ProductName
ORDER BY Product_Count DESC
LIMIT 1;

-- City with Highest Sales
SELECT City, SUM(TransactionAmount) AS Total_Sales
FROM sales_data
GROUP BY City
ORDER BY Total_Sales DESC
LIMIT 1;
```