🎯 📝 Problem Statement

“You are a Data Scientist for an educational store chain that tracks student purchases across regions. Your job is to analyze student sales data to gain insights into buying patterns, sales trends, and performance of different product categories. You are also expected to clean, explore, and visualize the data to support business decisions and build predictive models.”

🚀 ✅ Tasks / Features / Actions for Students to Perform

---

1 Data Understanding + Cleaning

- 🟢 Load the dataset using Pandas.
- 🟢 Check for missing values and handle them if any.
- 🟢 Verify data types (e.g., dates, numbers).
- 🟢 Remove duplicates if they exist.
- 🟢 Ensure no negative prices, quantities, or totals.

---

2 Exploratory Data Analysis (EDA)

- 📊 Show how many sales per product type.
- 📊 Plot sales trend over months (PurchaseDate).
- 📊 Find average FinalTotal by region.
- 📊 Show top-selling products overall.
- 📊 Compare average FinalTotal across grades (do higher grade students spend more?).
- 📊 Visualize preferred payment methods.

---

3 Sales Performance Insights

- 💡 Which region generated the highest sales?
- 💡 Which product has the highest revenue?
- 💡 Which payment method is most popular?
- 💡 What is the average discount given across products?

4️⃣ Visualization tasks

- 📈 Line plot: daily or monthly sales trends.
- 🛒 Bar plot: product vs total sales.
- 🌍 Pie chart: sales distribution by region.
- 🟪 Heatmap: correlation between UnitPrice, Quantity, Total, Discount, FinalTotal.
- 🟠 Boxplot: FinalTotal by payment method.

---

5 Feature Engineering

- 🛠️ Create a Month column from PurchaseDate.
- 🛠️ Create a DiscountRate = Discount / Total.
- 🛠️ Flag high-value purchases (FinalTotal > X).