📘 Problem Definition: Student Performance Analysis and Prediction

🎯 Objective:

You are given a dataset named **Marksheet.csv** containing student names and their marks in four class tests: `TEST-1_MARKS`, `TEST-2_MARKS`, `TEST-3_MARKS`, and `TEST-4_MARKS`. Your task is to clean the dataset, perform total mark calculation, assign grades based on the total, and then build a **simple linear regression model** to predict total marks using only the first test (`TEST-1_MARKS`).

📄 Dataset Description:

Each row in the dataset corresponds to a student. The columns are as follows:

Column Name	Description
Name	Name of the student
TEST-1 MARKS	Marks obtained in Test 1 (out of 25)
TEST-2 MARKS	Marks obtained in Test 2 (out of 25)
TEST-3 MARKS	Marks obtained in Test 3 (out of 25)
TEST-4 MARKS	Marks obtained in Test 4 (out of 25)

🧹 Perform Required Data Cleaning

📊 Total Marks Calculation:

- Calculate and create new column for the **Total Marks** as the sum of marks from all four tests:

  $$\text{Total Marks} = T1 + T2 + T3 + T4$$

🏅 Grade Assignment Rules:

Assign grades based on the following total marks thresholds:

Grade	Criteria (Total Marks)
A	85 and above
B	75 to 84.99
C	65 to 74.99
D	50 to 64.99
E	35 to 49.99
F	Below 35

Note: The total marks are out of 100 (each test is out of 25).

📈 Regression Task:

Build a **Simple Linear Regression Model** using:

- **Independent Variable (X):** `TEST-1_MARKS`
- **Dependent Variable (y):** `Total_Marks`

The goal is to:

1. Fit a model to predict `Total_Marks` based on `TEST-1_MARKS`.
2. Output the model's **intercept** and **slope**.
3. Generate predictions for total marks.
4. Compare the predicted and actual total marks and find MSE and R2 score
5. **Visualize** the data points and regression line using a scatter plot.

🧠 Deliverables:

1. Cleaned and processed DataFrame with:
   - `Total_Marks` column.
   - `Grade` column.
   - `Predicted_Total` column (from the regression model).
2. Regression model parameters (intercept and coefficient).
3. Scatter plot showing actual vs predicted values with a regression line

---

Extra:

Objectives

1. **Search** through random_state values in the range 0–50.
2. For each random_state:
    - Split the data (80% train / 20% test).
    - Train a linear regression model on the training set.
    - Evaluate its R^2 on the test set.
3. **Identify** the random_state that yields the **maximum R^2**.
4. **Plot** a line graph of R^2 score vs. random_state for all values in [0,1,…,50].