

# **Student Performance Analysis**

Data Cleaning, Statistics,  
Visualization, and Modeling  
Parts A–D

# Part A: Data Review

- Dataset: 500 rows, 6 columns
- No missing values or duplicates
- Categorical check
  - 'Sex' contains only Male & Female. Looks clean.
- Score anomalies:
  - Max scores (e.g., Test\_1\_Scores =118.2, Test\_3\_Score =119.4) might suggest some scaling or outliers, as 100 is a typical max.

# Part B: Summary Statistics

- University exam score by gender

Gender	Mean	Median	Std Dev
Female	71.525	69.9	16.84
Male	79.76	79.2	12.67

- Male students had a higher average university Exam score of 79.76

# Part B: Summary Statistics

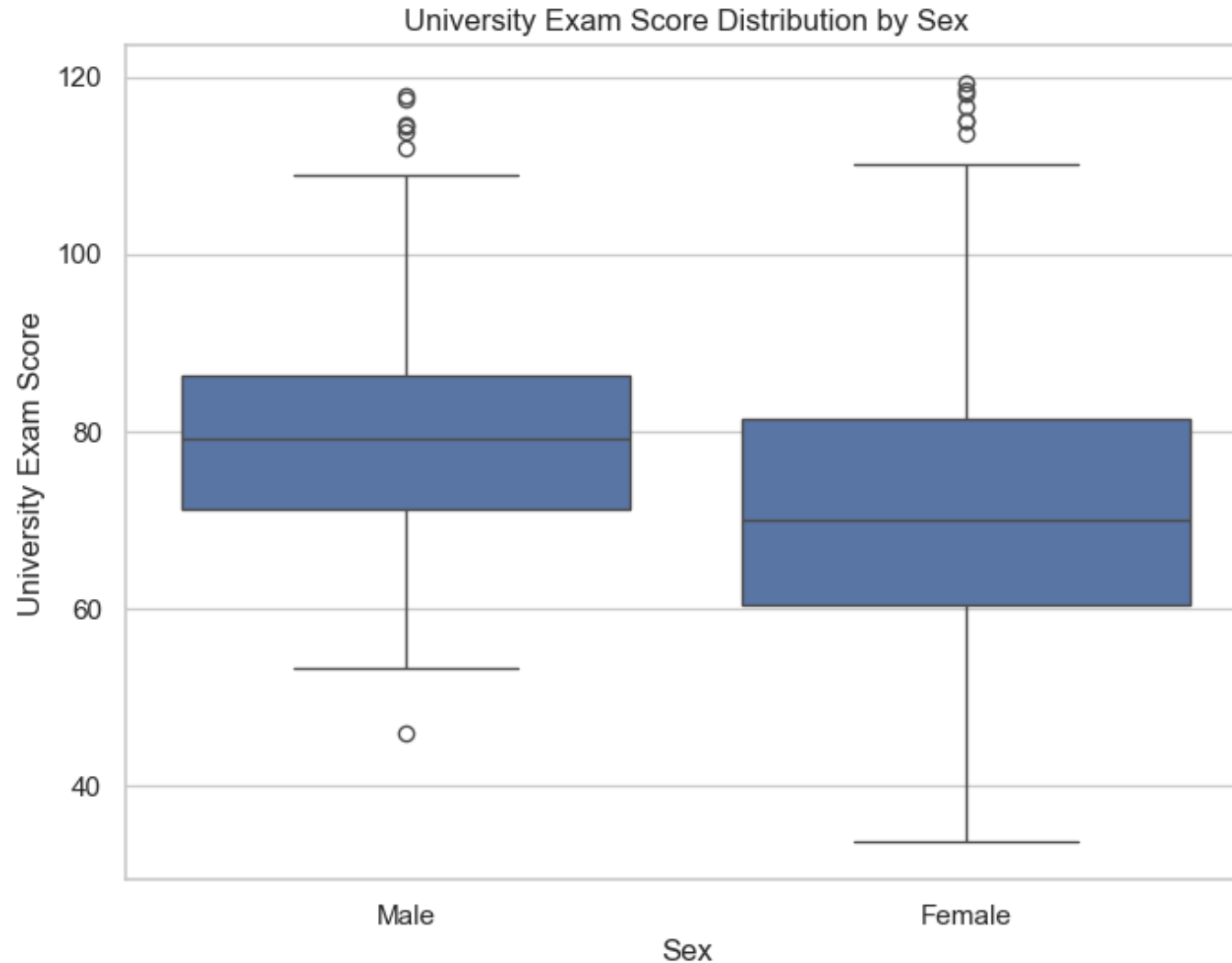
- Correlation Matrix

	Test_1_Score	Test_3_Score	University_Exam_Score
Test_1_Score	1.000000	0.193338	0.061521
Test_3_Score	0.193338	1.000000	0.047381
University_Exam_Score	0.061521	0.047381	1.000000

- Test\_1\_Scores and Test\_2\_Scores correlated by 0.193
- Still a weak correlation, but relatively the strongest among the three.

# Part C: Data Visualization

- Boxplot: Exam score by Sex

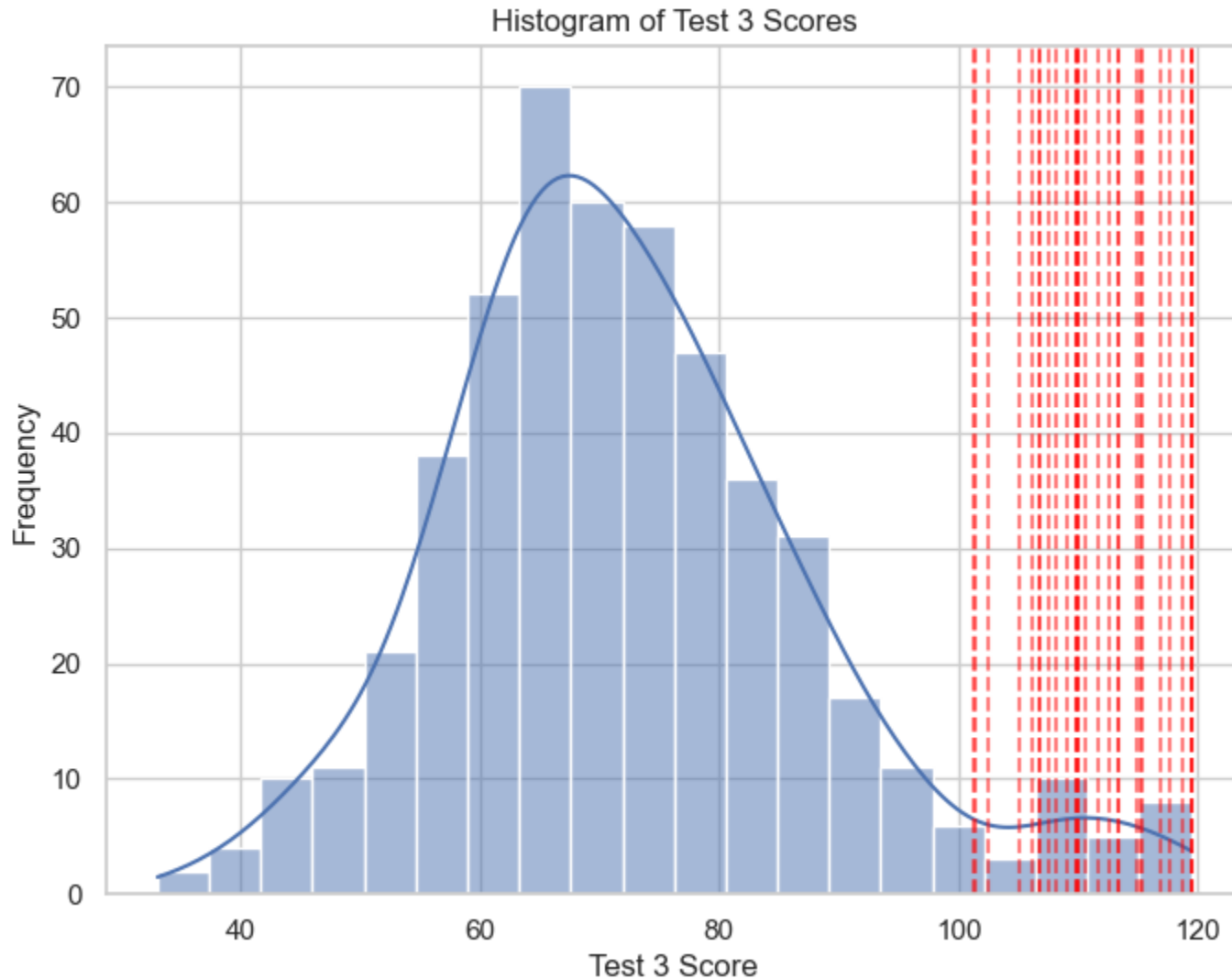


# Part C: Data Visualization

- Male students have a higher median university exam score.
- Female students show greater variation in the central 50% (wider box) and at the extremes (lower minimums, more outliers).
- Both groups have high-end outliers, but females also show low-end outliers (some below 40).

# Part C: Data Visualization

- Histogram of Test\_3\_Scores



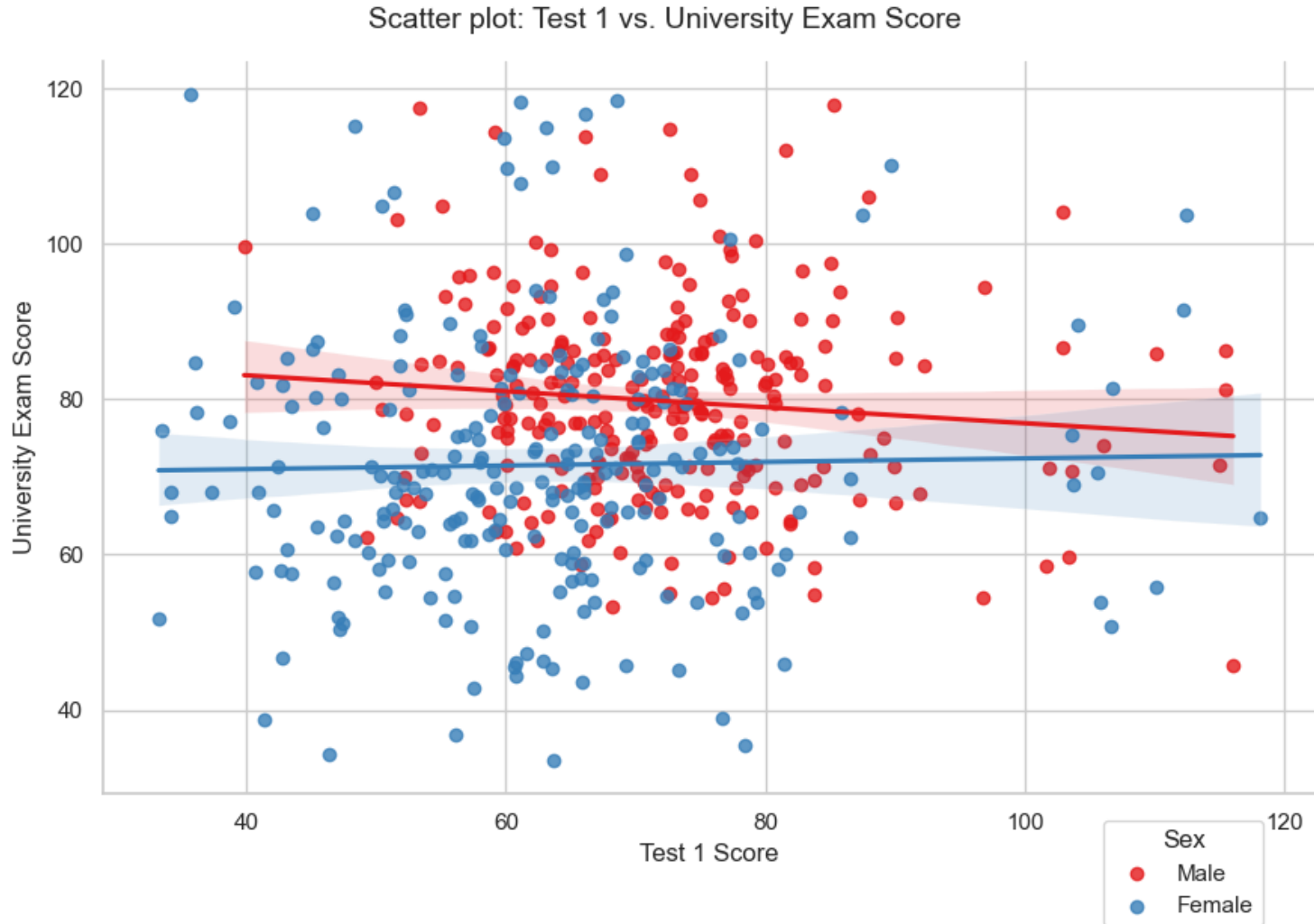
# Part C: Data Visualization

- Most students scored between 60 and 90 on Test\_3.
- The shape is slightly right-skewed:
  - A long tail stretches toward higher values.
  - This suggests that while most students performed around the middle range, a few scored unusually high.



# Part C: Data Visualization

- Scatter Plot – Test 1 vs. University Exam



# Part C: Data Visualization

- The scatter plot shows a weak positive relationship between Test\_1\_Score and University\_Exam\_Score.
  - Generally, students with higher Test\_1\_scores tend to have higher university exam scores, but it's not a strong trend, as confirmed by the low correlation (0.06).
- Both Male and Female students are spread across the range, but:

# Part C: Data Visualization

- Both Male and Female students are spread across the range, but:
  - Male students appear to dominate the higher score region of the university exam axis.
  - Female students are more concentrated in the lower to mid-range of university exam scores.
  - This aligns with earlier findings where males had a higher average university exam score.

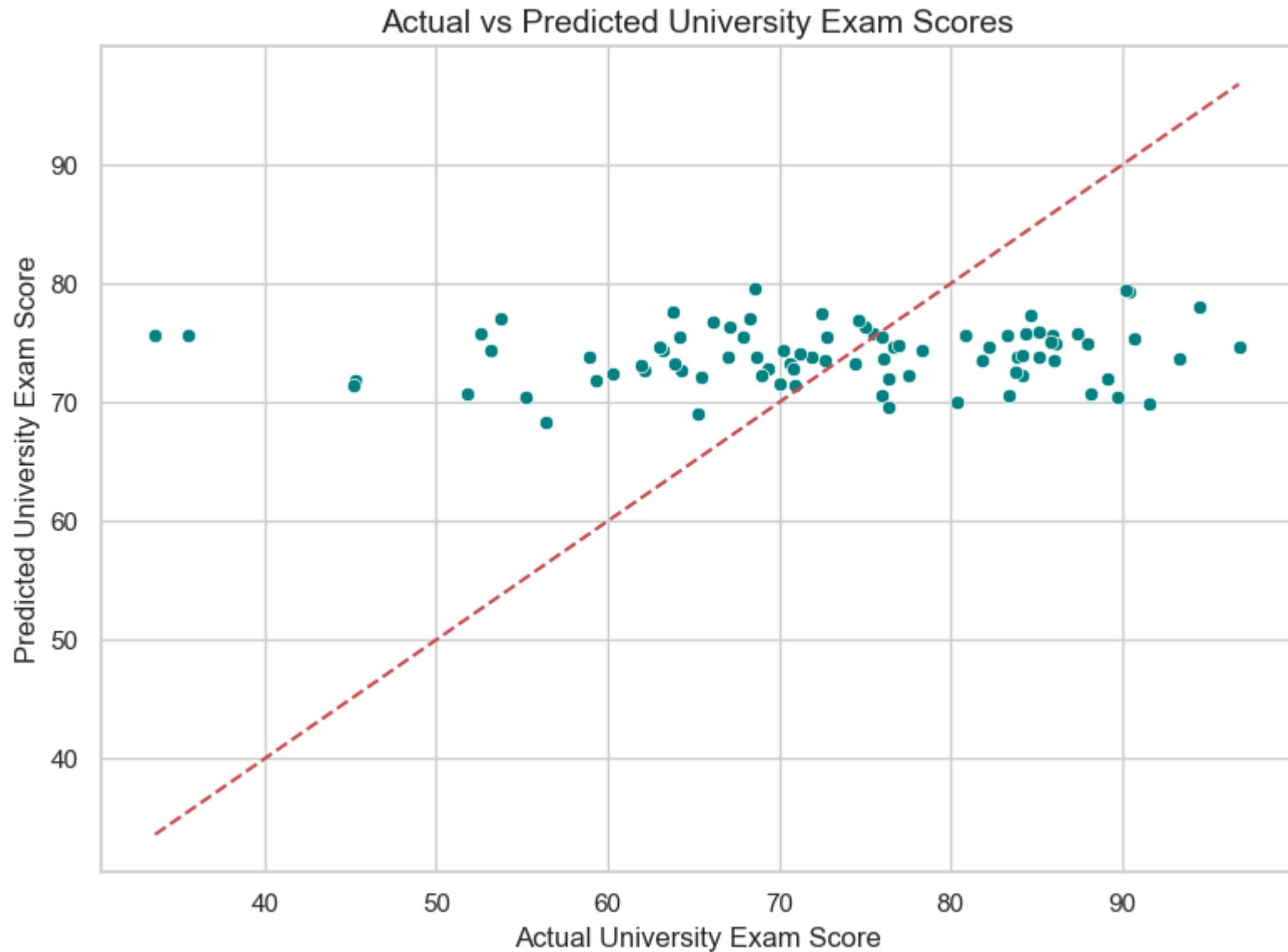
# Part C: Data Visualization

- **Linearity**
  - There's no clear linear trend, confirming that Test\_1\_Score alone is not a strong predictor of final exam performance.

# Part D: Modeling

- Outliers removed (scores >100)
- Model Equation:  
***UES=57.23+0.086(Test\_1\_Score)+0.156(Test\_3\_Score).***
- $R^2 = 0.0076 \rightarrow$  weak predictive power
  - (Most of the variation in exam scores is caused by other factors not included in the model.)
- $RMSE \approx 13 \rightarrow$  average prediction error
  - For exam scores that range from 33 to 100, an error of 13 is relatively high.

# Part D: Modeling



*Thank you!*