EXPERIMENT 0

Sanket patil D15A-37

AIM:-

To analyze the `student_performance.csv` dataset using **NumPy, Pandas, Matplotlib, and Seaborn** in order to:

- Compute statistical measures of the *Final_Score*
- Perform Min-Max normalization
- Handle and inspect dataset structure
- Create a categorical performance label
- Visualize relationships between study factors and final performance
- Derive meaningful insights from the analysis

1)Dataset Overview (Pandas Handling)

- Dataset Size: **20 rows × 5 columns**
- Columns:
    - Hours_Studied
    - Attendance
    - Assignment_Score
    - Midterm_Score
    - Final_Score

Observations:

- No missing values detected.
- All columns are numerical.
- Dataset is clean and ready for analysis.

2) NumPy Statistical Analysis on Final_Score

Computed Values:

- **Mean** = 68.95
- **Median** = 70.50
- **Standard Deviation** = 8.71

Interpretation:

- The average student score is around **69 marks**.
- The median is slightly higher than the mean, indicating mild left skewness.
- A standard deviation of 8.71 suggests **moderate variation** in performance.

3) Min-Max Normalization

Formula used:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Purpose:

- Scale values between 0 and 1
- Remove magnitude bias
- Prepare data for future ML models

Example normalized values:

```
[0.00, 0.16, 0.25, 0.38, 0.51]
```

4) Performance Categorization

Distribution:

- High → 2 students
- Medium → 14 students
- Low → 4 students

Insight:

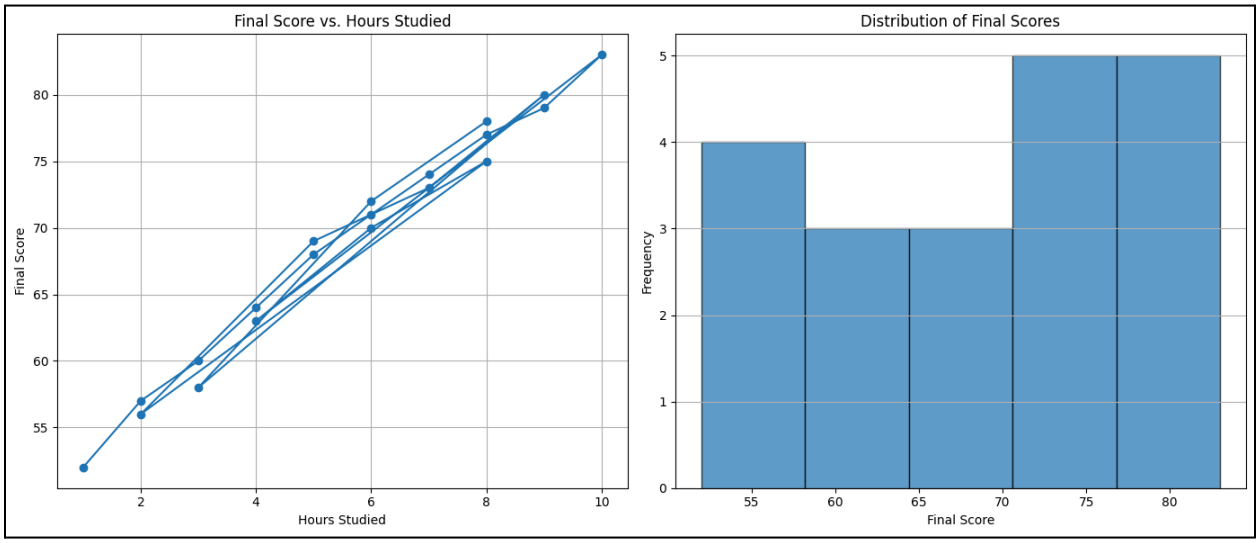
70% of students fall under the *Medium* performance group.

5) Visualization Insights

Line Plot & Scatter Plot

(Hours_Studied vs Final_Score)

- Clear positive trend observed.
- More study hours → Higher final score.
- Indicates possible linear relationship.



Histogram of Final_Score

- Scores cluster around 65–75 range.
- Distribution is approximately normal.
- Few students at extreme high/low ends.

Correlation Heatmap

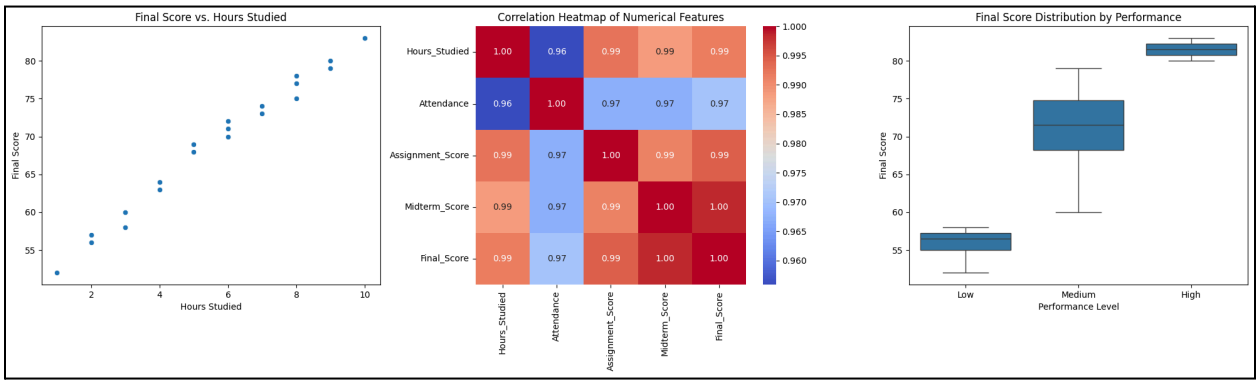
- Strong positive correlation between:
    - Midterm_Score & Final_Score

- Assignment_Score & Final_Score
- Hours_Studied & Final_Score

Boxplot (Final_Score by Performance)

- Clear separation between Low, Medium, and High groups.
- Confirms proper categorization logic.

No major outliers present.



Final Conclusion

The analysis successfully demonstrated how NumPy, Pandas, Matplotlib, and Seaborn can be used to:

- Compute statistical metrics
- Normalize numerical data
- Categorize performance levels
- Visualize relationships between academic factors

The results indicate a strong positive relationship between study-related variables and final performance, highlighting the importance of consistent academic engagement.