Student Exam Score Analysis Report

M Srineela Reddy
2211CS010335

1. Introduction:

The purpose of this project is to analyze a student exam score dataset using PySpark. By applying distributed data processing techniques, we explored the impact of various factors such as study hours, sleep, attendance, and prior performance on student exam results. The results are visualized through bar, line, and scatter charts to derive insights into student performance trends.

2. Dataset Overview:

- Entries: Multiple student records
- Columns:
  - hours_studied – Number of study hours per day
  - sleep_hours – Average sleep duration per day
  - attendance_percent – Attendance percentage in class
  - previous_scores – Previous exam scores
  - exam_score – Current exam score

- The dataset is clean and complete, with no missing values, making it reliable for analysis.

3. Key Findings:

a) Correlation Analysis:

- Study hours and attendance show a strong positive correlation with exam scores.
- Sleep hours exhibit a moderate positive effect, suggesting a balanced lifestyle helps performance.
- Previous scores are a reliable indicator of current exam results, confirming consistent learning patterns.

b) Study Hours and Exam Performance:

- Students who study more hours generally achieve higher scores.
- A steady upward trend is observed in performance with increased study time.

c) Attendance and Performance:

- Higher attendance percentages are linked with improved scores.
- Attendance reinforces the importance of class participation.

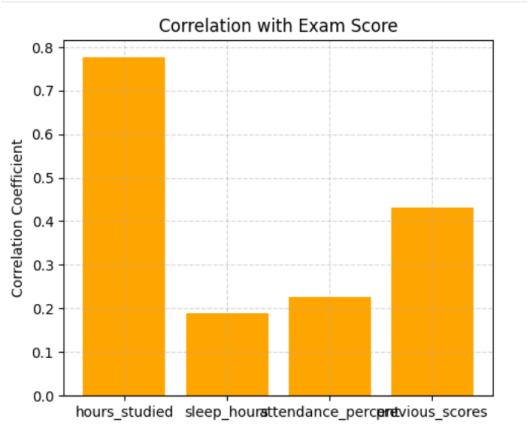
d) Improvement Analysis:

• The difference between current and previous exam scores indicates that most students showed positive improvement.
• Average improvement was notable, with some students achieving significant gains.

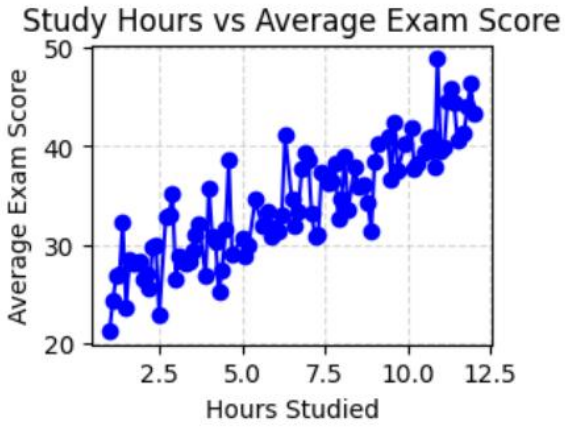
4. Data Visualization (DV):

Using PySpark with Matplotlib, the following charts were generated:

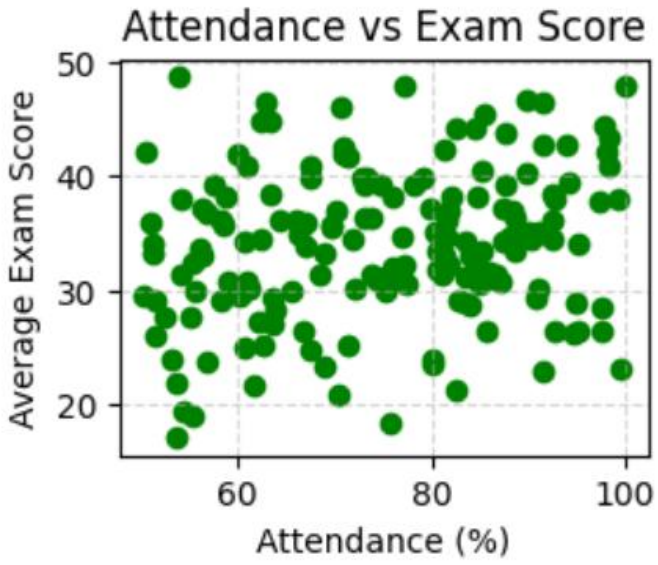
• Correlation with Exam Scores (Bar Chart): Feature-wise correlation strength.



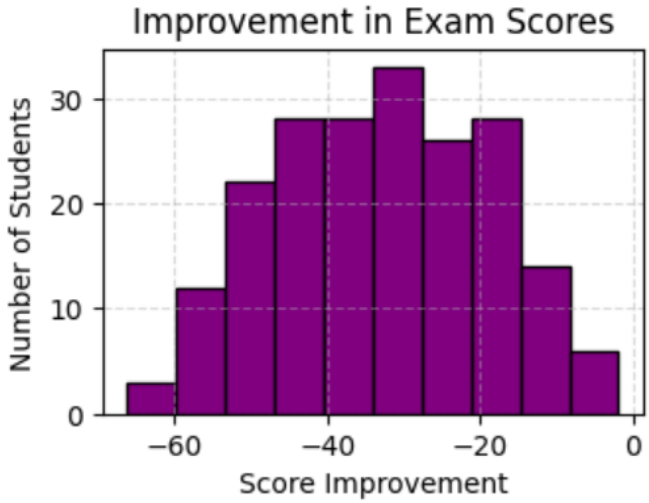
• Study Hours vs Average Exam Score (Line Chart): Trend of study hours on performance.



• Attendance vs Exam Score (Scatter Plot): Effect of attendance on performance.



• Improvement Distribution (Histogram/Boxplot): Variations in score improvement.



These visualizations provide clear, data-driven insights into academic performance trends.

5. Conclusion:

• Study hours, attendance, and previous scores are the strongest determinants of performance.
• Balanced sleep contributes positively to exam outcomes.
• Students with consistent study habits and class participation achieve higher scores.
• Overall, the analysis demonstrates that academic performance can be improved by focusing on study discipline, healthy routines, and regular attendance.