ML TEST SCORES ANALYSIS REPORT

1. INTRODUCTION

This report presents an EDA of student test scores across three batches . The objective of this Analysis is to compare batch wise performance , understand score distributions , assess consistency and variability , identify outliers and derive key insights that can help improve academic outcomes

2. DATASET DESCRIPTION

It has 149 rows of data , which has the given attributes :
- Batch – Batch Identifier (AI_ELITE_4 , AI_ELITE_6 , AI_ELITE_7)
- User_ID – User ID of the student
- Score – student score out of 7

The 3 batches have almost equal number of students which allows for an unbiased comparison of their performance.

3. DATA PREPROCESSING

Before Analysis , the dataset was cleaned and validated :
- Column names were cleaned by removing leading and trailing whitespaces
- Score values were converted from string format (e.g '6/7') to integer values
- Data types were verified
- No missing or duplicate records were found.

4. DESCRIPTIVE STATISTICS

Descriptive statistical values such as mean, median, standard deviation, minimum, and maximum scores for batches were calculated

Key Observations:
- AI_ELITE_7 shows the highest average and median score, which reflects that it is performing best.
- AI_ELITE_4 has the lowest average score.
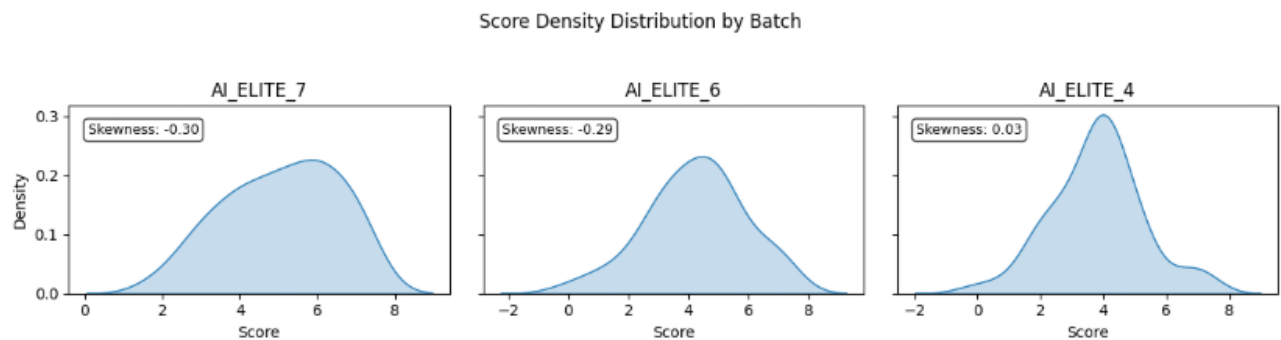- AI_ELITE_6 shows higher variability compared to other batches.

Batch	count	mean	median	std	min	max
AI_ELITE_4	48	3.791667	4.0	1.443376	0	7
AI_ELITE_6	48	4.229167	4.0	1.640441	0	7
AI_ELITE_7	53	5.056604	5.0	1.446682	2	7

5. SCORE DISTRIBUTION ANALYSIS (SKEWNESS)

Score distributions were analyzed using and KDE (density) plots to understand how scores are spread within each batch.

Observations:
- AI_ELITE_7 and AI_ELITE_6 show slight negative skewness, indicating a higher concentration of high scores.
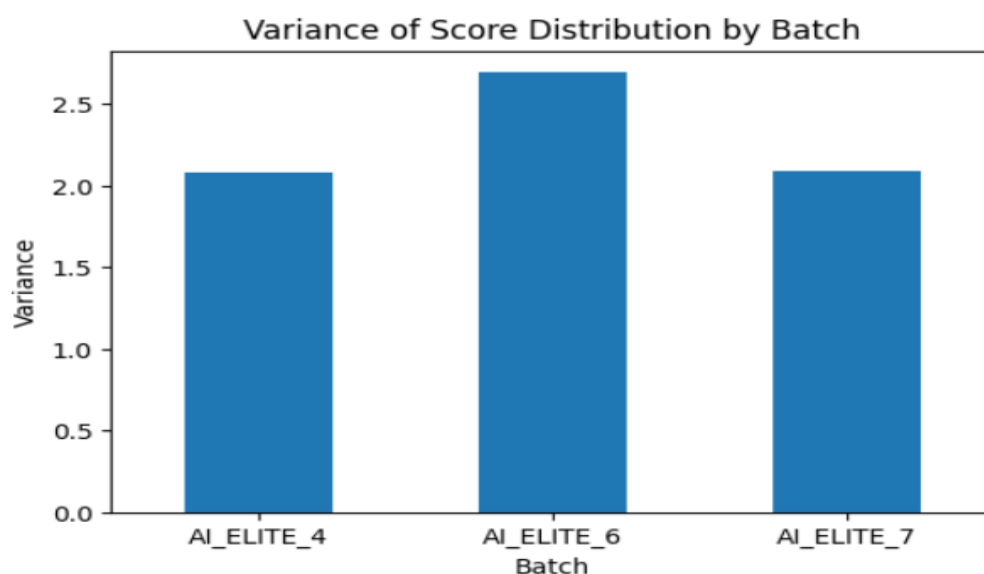- AI_ELITE_4 shows an almost symmetric distribution, suggesting balanced but moderate performance.

Score Density Distribution by Batch

6. CONSISTENCY ANALYSIS (VARIENCE)

Variance was analyzed to assess consistency of performance within each batch.

Observations:
- AI_ELITE_6 has the highest variance, indicating unequal learning outcomes.
- AI_ELITE_4 has the lowest variance, showing consistent but lower performance.
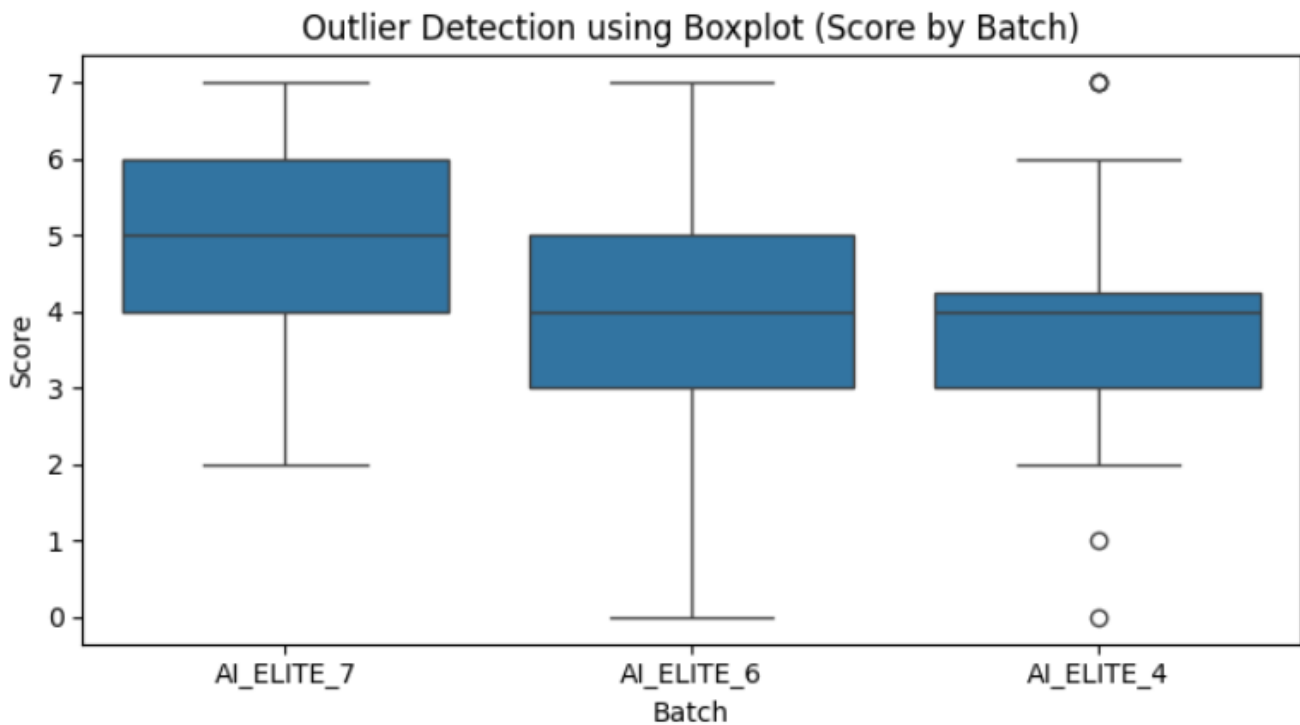- AI_ELITE_7 balances good average performance with moderate consistency.

7. OUTLIER ANALYSIS

To Analyze Outliers Interquartile Range (IQR) and box plots were used.
- AI_ELITE_4 and AI_ELITE_6 contain extremely low scores, indicating learning gaps.
- AI_ELITE_7 has fewer low-end outliers and a stronger baseline.
- High-end outliers (perfect scores) indicate top performers in all batches.

Outliers are treated as **valid student performances**, not data errors.

Outlier Detection using Boxplot (Score by Batch)

8. BATCH RANKING

Batches were ranked using multiple criteria , not a single metric

Rankings:
- Best Performing Batch: AI_ELITE_7 (highest mean and median).
- Most Consistent Batch: AI_ELITE_4 (lowest variance).
- Batch Needing Intervention: AI_ELITE_6 (high variability and low scorers).

9. Key Insights and Recommendations

Insights
- AI_ELITE_7 demonstrates strong overall performance with a solid baseline.
- AI_ELITE_6 shows significant performance inequality.
- AI_ELITE_4 is consistent but underperforms on average.
- Variability and outliers highlight where academic support is required.

Recommendations

- Introduce targeted remedial sessions for low-performing students in AI_ELITE_6.
- Replicate effective teaching practices observed in AI_ELITE_7.
- Use high-performing students for peer mentoring programs.
- Improve instructional strategies in AI_ELITE_4 to raise average performance.

10. CONCLUSION

This exploratory data analysis helped to make a complete comparison of the performance of the students of different batches. By using the descriptors of statistics, distribution, skewness, variance, and outliers, the performance of strengths, weaknesses, and areas of concern of the students of different batches have been identified.