DATA ANALYSIS

QUESTION PAPER ANALYSIS

Introduction

This report presents an analysis of the class\_marks.csv dataset, which contains examination performance data for 86 students. The objective is to examine student scores, identify performance patterns, and provide insights through data exploration and visualization. The findings are intended to inform academic evaluation and decision-making.

Dataset Description

The dataset includes 86 records, each with a Total score and marks for 11 questions: Q1aM4, Q1bM6, Q2aM6, Q2bM4, Q3aM5, Q3bM5, Q4aM3, Q4bM7, Q5M10, Q6aM4, and Q6bM6. The suffixes (e.g., M4, M6) indicate maximum marks per question, and missing values (NaN) denote unattempted questions. This structure enables a detailed assessment of overall and question-specific performance.

Data Exploration

1. Overview of Performance

The dataset reveals a range of total scores, from 17 to 37 in the sampled subsets, with frequent missing values across questions. This suggests varying levels of student engagement or difficulty with specific sections.

2. Bottom 10 Records

The bottom 10 students, with totals ranging from 17 to 37, exhibit a pattern of incomplete attempts. For instance, the student with a total of 17 scored 2/4 on Q1aM4 and 4/6 on Q2aM6 but left multiple questions unanswered, including Q4aM3, Q4bM7, and Q5M10. This indicates potential challenges in completing the exam fully.

3. Top 10 Records

The top 10 students, with totals from 20 to 37, demonstrate stronger performance. Notably, a student with a total of 36 achieved a perfect 10/10 on Q5M10, alongside 6/6 on Q1bM6 and 4/4 on Q2bM4. However, even among high performers, some questions (e.g., Q4aM3, Q4bM7) remain unattempted, suggesting selective prioritization.

Visualization: Q6 Performance Analysis

A line graph was developed to analyze the relationship between Q6 scores (assumed as the sum of Q6aM4 and Q6bM6) and average total marks. The graph features a clear line with data points, a shaded area for emphasis, and a grid for readability.

Findings:

A positive correlation exists: higher Q6 scores correspond to higher average total marks.

Students achieving full marks on Q6 (e.g., 10/10) consistently score above 35 overall, while lower Q6 scores align with totals below 25.

This suggests Q6, with a maximum of 10 marks, significantly influences overall performance.

Discussion

The analysis highlights distinct performance trends. The bottom 10 students struggle with incomplete submissions, potentially due to time constraints or lack of preparation. In contrast, the top 10 demonstrate greater consistency, though gaps persist. The Q6 analysis is the most significant finding, indicating its potential as a key determinant of success. The frequent missing values across the dataset underscore the need for further investigation into question difficulty or student readiness.

Conclusion

This analysis of the class\_marks.csv dataset provides actionable insights into student performance. The top performers excel through broader engagement, while the bottom group requires support to address incomplete attempts. The strong link between Q6 performance and total scores identifies it as a critical focus area. This report offers a clear, data-driven foundation for academic improvement strategies, presented with precision and professionalism.

End of Report