DSAT Score Improvement Analysis Report

July 30, 2025

Abstract

This report presents the results of the DSAT Score Improvement Analysis Tool, developed for HighScores.ai, analyzing student performance on Digital SAT (DSAT) practice tests. The tool processes student attempt data, identifies weak topics, performs what-if analysis to simulate score improvements, and generates visualizations. Two student datasets are analyzed, with performance metrics, high-impact questions, and suggested score improvements for Math and Reading and Writing sections. Limitations include missing Module 2 data and potential placeholder scoring.

1 Introduction

The DSAT Score Improvement Analysis Tool processes student performance data from JSON files (`67f2aae2c084263d16dbe462user_attempt_v2.json`, `66fece285a916f0bb5aea9c5use` and scoring maps (`scoring_DSAT_v2.json`), using MongoDB for data storage. It extracts insights from `What-if-analysis.docx`, performs what-if analysis by simulating two additional correct answers in Module 1, and generates visualizations with Matplotlib and Chart.js. This report summarizes performance, score improvements, and high-impact questions for two students.

2 Methodology

The tool uses Python 3.13 with `pymongo`, `matplotlib`, and `python-docx`. Data is imported into a MongoDB database (`dsat_analysis`), with collections `sat_scoring` and `student_results`. Performance is analyzed by subject (Math, Reading and Writing) and module (Module 1: Static, Module 2: Hard/Easy). What-if analysis simulates correcting two additional Module 1 questions to estimate score gains. An optimal threshold for Module 2 difficulty is calculated, though limited by incomplete data.

3 Results

3.1 Student 1 (v2.json)

Weak Topics (Math): Linear equations in two variables (33.33%), Circle Equations (0.00%), Area (0.00%), etc.
Weak Topics (Reading and Writing): Subject-verb agreement (0.00%), Pronoun-antecedent agreement (0.00%).

Table 1: Performance and What-If Analysis for Student 1 (v2.json)

Metric	**Math**	
Module 1	14/22 (63.64%)	
Module 2	0/0 (hard)	
Scaled Score	410	
Total Score		840
What-If: +2 Correct in Module 1		
Current Score	410	
New Score	450	
Score Gain	40	
Per-Question Gain (est.)	∼20	
High-Impact Questions	`659e991004e80b72d57ac8c2` (Nonlinear functions, moderate)	`65b69f5`
	`65b8bcea117be8dcc30c340e` (Circle Equations, moderate)	

Table 2: Performance and What-If Analysis for Student 2 (v3.json)

Metric	**Math**	
Module 1	20/22 (90.91%)	
Module 2	0/0 (hard)	
Scaled Score	520	
Total Score		1020
What-If: +2 Correct in Module 1		
Current Score	520	
New Score	550	
Score Gain	30	
Per-Question Gain (est.)	∼15	
High-Impact Questions	`65ae7e3e836c92e3d55d8cf3` (Area, moderate)	`659041da1d`
	`65aa7f51ea9130b9997676b6` (Nonlinear equations, hard)	`65901ab81`

3.2 Student 2 (v3.json)

Weak Topics (Math): Area (0.00%), Circle Theorems (0.00%).
Weak Topics (Reading and Writing): None.

3.3 Visualizations

The tool generated a bar chart (`what_if_analysis.png`) comparing current and what-if scores, and a Chart.js configuration (`chartjs_config.json`) for interactive visualization. Due to missing Module 2 data, scores may rely on placeholder scaling.

4 Limitations

- **Missing Module 2 Data**: Both JSON files lack proper `Hard` or `Adaptive` sections, limiting scoring accuracy. A full DSAT practice test on Bluebook is recommended.
- **Placeholder Scoring**: If `scoring_DSAT_v2.json` is incomplete, scores use placeholders (e.g., 200 + raw * 15 for Math).

- **MongoDB Query Issue**: Queries for `student_results` fail, forcing fallback to local JSON files. Re-importing with correct `_id` fields resolves this.
- **Optimal Threshold**: The reported 30.00% threshold is inaccurate due to placeholder data; actual data suggests ∼65%.

5 Recommendations

- **Complete Module 2 Data**: Take a full DSAT practice test on https://bluebook.collegeboard.org to generate JSON files with Module 1 and Module 2 data.
- **Verify Scoring Maps**: Ensure `scoring_DSAT_v2.json` contains complete mapping for raw to scaled scores.
- **Fix MongoDB Queries**: Re-import data with unique `_id` fields combining `question_id` and `source_file`.
- **Focus Areas**: For Student 1, prioritize Math topics like Linear equations and Circle Equations, and Reading and Writing topics like Subject-verb agreement. For Student 2, focus on Math topics like Area and Circle Theorems.

6 Conclusion

The DSAT Score Improvement Analysis Tool successfully processes student data, identifies high-impact questions, and estimates score gains (e.g., 30–40 points in Math, 20–30 points in Reading and Writing per student). Despite limitations, it provides actionable insights for score improvement. Future work includes integrating complete Module 2 data and refining MongoDB queries.