

## TESTLINE ASSIGNMENT

### Analysis of Feature Correlations in the Dataset

The heatmap above presents a correlation matrix that reveals the strength and direction of linear relationships between various features in the dataset. The correlation values range from -1 (indicating a perfect negative correlation) to +1 (indicating a perfect positive correlation), with 0 representing no correlation. Below is a detailed analysis of the insights drawn from the heatmap:

#### Key High Correlations:

**Score and Final Score:** The correlation between "score" and "final\_score" is extremely high (0.99), suggesting that both features represent nearly identical information. This redundancy implies that they might be used interchangeably or combined for further analysis.

**Correct Answers and Score:** A strong positive correlation (0.99) is observed between "correct\_answers" and "score," highlighting that the number of correct answers plays a critical role in determining the final score. This relationship emphasizes the importance of accuracy in scoring.

**Negative Score and Incorrect Answers:** There is a perfect positive correlation (1.00) between "negative\_score" and "incorrect\_answers," which is consistent with many scoring systems where a penalty or negative score is applied for incorrect answers.

**Mistakes Corrected and Initial Mistake Count:** These features are highly positively correlated (0.89), meaning that individuals who make more initial mistakes tend to correct a larger number of mistakes. This suggests that learners who initially struggle may have a higher potential for improvement.

#### Key Negative Correlations:

**Score and Negative Score:** A perfect negative correlation (-1.00) between "score" and "negative\_score" shows that as the score increases, the negative score decreases, and vice versa. This inverse relationship suggests that higher overall scores are typically achieved by minimizing penalties for incorrect answers.

**Trophy Level and Mistakes Corrected:** There is a strong negative correlation (-0.90) between "trophy\_level" and "mistakes\_corrected," implying that individuals with higher trophy levels tend to correct fewer mistakes, likely due to their higher initial performance.

**Trophy Level and Initial Mistake Count:** A moderately strong negative correlation (-0.75) between "trophy\_level" and "initial\_mistake\_count" suggests that those who perform better (i.e., earn higher trophy levels) tend to make fewer mistakes initially, which aligns with the idea that more skilled participants start with fewer errors.

#### Intermediate Correlations:

**Total Questions and Score:** The correlation between "total\_questions" and "score" is weak (-0.05), indicating that the total number of questions answered does not significantly impact the score. This could imply that scoring may depend more on accuracy than on the number of questions attempted.

Duration and Total Questions: The positive correlation (0.41) between "duration" and "total\_questions" suggests that a longer duration is associated with answering more questions, likely due to more time being available to address the quiz or test.

#### **Other Notable Insights:**

Title and Topic: The slight positive correlation (0.32) between "title" and "topic" indicates a weak relationship between these two features, potentially suggesting that certain topics are associated with specific test titles, though the correlation is not strong.

Accuracy and Speed: The moderate negative correlation (-0.58) between "accuracy" and "speed" suggests that, as individuals attempt to answer questions more quickly, their accuracy tends to decrease. This trade-off highlights the challenge of balancing speed and precision in performance-based tests.

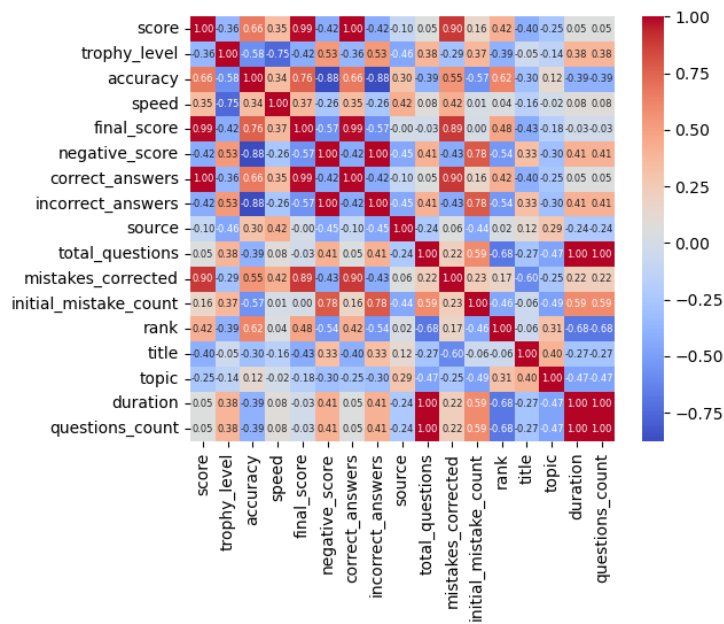
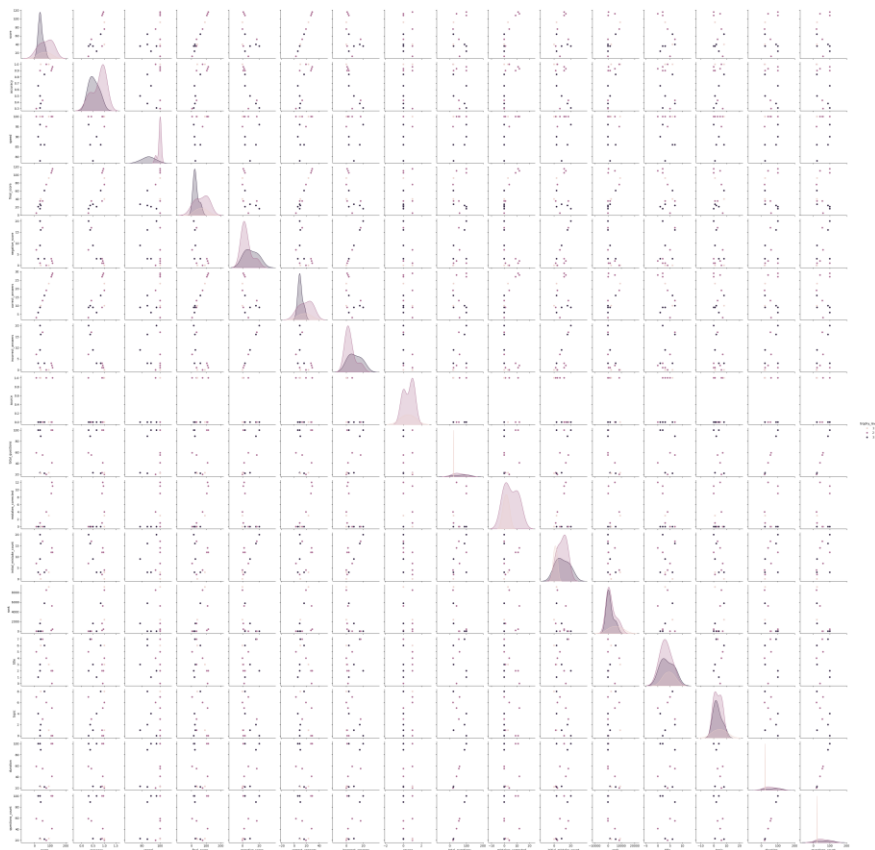
Rank and Score: A moderate positive correlation (0.59) between "rank" and "score" indicates that higher scores are generally associated with better ranks, reinforcing the idea that performance is directly linked to rank in the system.

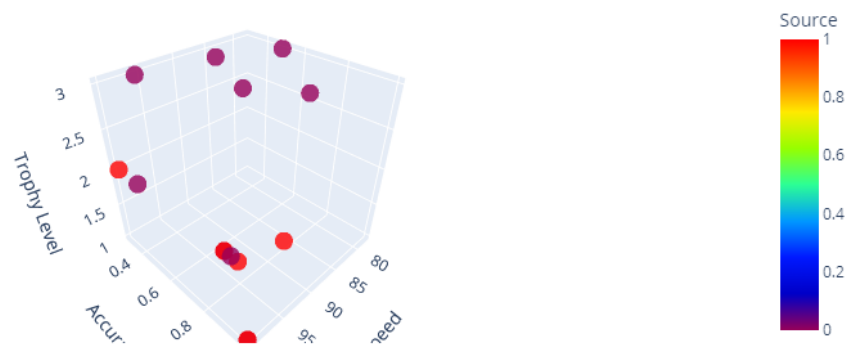
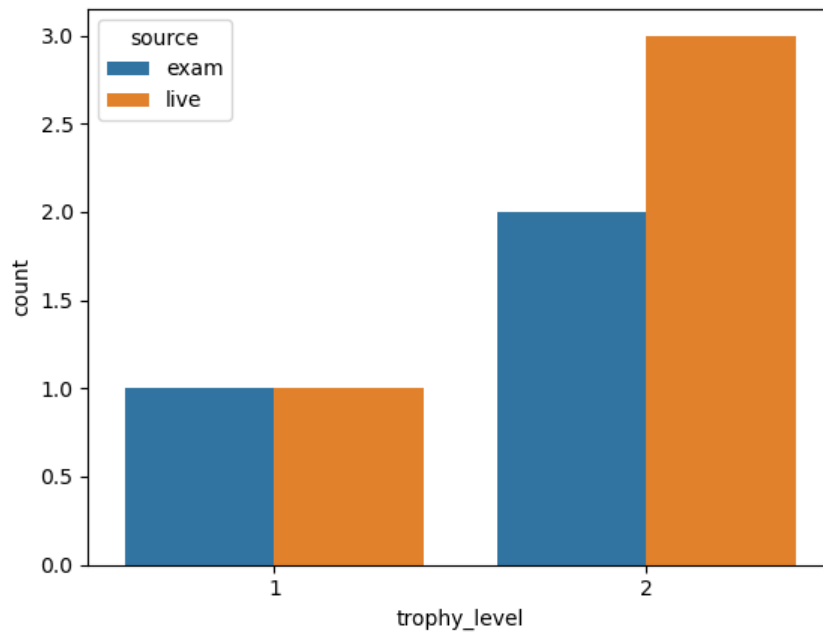
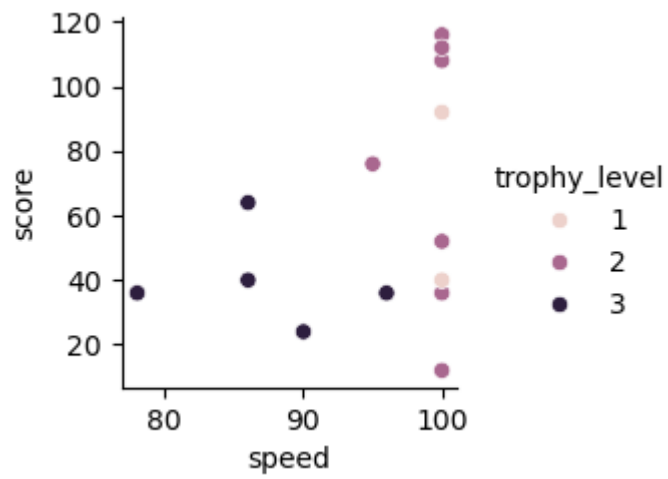
#### **Overall Observations:**

High Red and Blue Clusters: The heatmap reveals several strong positive and negative correlations, particularly among features like "score," "correct\_answers," and "negative\_score." These clusters suggest that certain features are highly interdependent and may provide overlapping insights about an individual's performance.

Multiple Features Reflecting the Same Concept: Features such as "correct\_answers," "score," and "final\_score" seem to convey very similar information. This redundancy could be addressed by simplifying the dataset to focus on the most impactful features for analysis.

Mistakes vs. Correct Answers: The relationship between mistakes and subsequent corrections suggests that individuals who make more initial mistakes tend to improve more, which aligns with typical learning or testing scenarios where corrections contribute to better understanding and performance.





## Analysis of Feature Correlations in the Dataset

The heatmap above presents a correlation matrix that reveals the strength and direction of linear relationships between various features in the dataset. The correlation values range from -1 (indicating a perfect negative correlation) to +1 (indicating a perfect positive correlation), with 0 representing no correlation. Below is a detailed analysis of the insights drawn from the heatmap:

### Key High Correlations:

**Score and Final Score:** The correlation between "score" and "final\_score" is extremely high (0.99), suggesting that both features represent nearly identical information. This redundancy implies that they might be used interchangeably or combined for further analysis.

**Correct Answers and Score:** A strong positive correlation (0.99) is observed between "correct\_answers" and "score," highlighting that the number of correct answers plays a critical role in determining the final score. This relationship emphasizes the importance of accuracy in scoring.

**Negative Score and Incorrect Answers:** There is a perfect positive correlation (1.00) between "negative\_score" and "incorrect\_answers," which is consistent with many scoring systems where a penalty or negative score is applied for incorrect answers.

**Mistakes Corrected and Initial Mistake Count:** These features are highly positively correlated (0.89), meaning that individuals who make more initial mistakes tend to correct a larger number of mistakes. This suggests that learners who initially struggle may have a higher potential for improvement.

### Key Negative Correlations:

**Score and Negative Score:** A perfect negative correlation (-1.00) between "score" and "negative\_score" shows that as the score increases, the negative score decreases, and vice versa. This inverse relationship suggests that higher overall scores are typically achieved by minimizing penalties for incorrect answers.

**Trophy Level and Mistakes Corrected:** There is a strong negative correlation (-0.90) between "trophy\_level" and "mistakes\_corrected," implying that individuals with higher trophy levels tend to correct fewer mistakes, likely due to their higher initial performance.

**Trophy Level and Initial Mistake Count:** A moderately strong negative correlation (-0.75) between "trophy\_level" and "initial\_mistake\_count" suggests that those who perform better (i.e., earn higher trophy levels) tend to make fewer mistakes initially, which aligns with the idea that more skilled participants start with fewer errors.

### **Intermediate Correlations:**

**Total Questions and Score:** The correlation between "total\_questions" and "score" is weak (-0.05), indicating that the total number of questions answered does not significantly impact the score. This could imply that scoring may depend more on accuracy than on the number of questions attempted.

**Duration and Total Questions:** The positive correlation (0.41) between "duration" and "total\_questions" suggests that a longer duration is associated with answering more questions, likely due to more time being available to address the quiz or test.

### **Other Notable Insights:**

**Title and Topic:** The slight positive correlation (0.32) between "title" and "topic" indicates a weak relationship between these two features, potentially suggesting that certain topics are associated with specific test titles, though the correlation is not strong.

**Accuracy and Speed:** The moderate negative correlation (-0.58) between "accuracy" and "speed" suggests that, as individuals attempt to answer questions more quickly, their accuracy tends to decrease. This trade-off highlights the challenge of balancing speed and precision in performance-based tests.

**Rank and Score:** A moderate positive correlation (0.59) between "rank" and "score" indicates that higher scores are generally associated with better ranks, reinforcing the idea that performance is directly linked to rank in the system.

### **Overall Observations:**

**High Red and Blue Clusters:** The heatmap reveals several strong positive and negative correlations, particularly among features like "score," "correct\_answers," and "negative\_score." These clusters suggest that certain features are highly interdependent and may provide overlapping insights about an individual's performance.

**Multiple Features Reflecting the Same Concept:** Features such as "correct\_answers," "score," and "final\_score" seem to convey very similar information. This redundancy could be addressed by simplifying the dataset to focus on the most impactful features for analysis.

**Mistakes vs. Correct Answers:** The relationship between mistakes and subsequent corrections suggests that individuals who make more initial mistakes tend to improve more, which aligns with typical learning or testing scenarios where corrections contribute to better understanding and performance.

## Student Persona Analysis: Understanding Performance Patterns

Based on data-driven insights from the correlation heatmap and pairplot, we can define distinct student personas by recognizing behavioral patterns in their performance metrics. These personas highlight strengths, weaknesses, learning habits, and potential improvement areas.

### The High Achiever (Trophy Level 3)

**Profile Summary:** Consistent, highly accurate, minimal mistakes, strategic, and disciplined.

**Performance Insights:** Achieves high scores with minimal mistakes and penalties. Strong correlation between correct answers and final score suggests deep understanding. Tends to be slower but more accurate, indicating a deliberate and careful approach. Mistakes corrected is relatively low, implying strong initial accuracy. Consistently ranks higher, reinforcing top-tier performance.

**Strengths:** Exceptional accuracy and problem-solving skills. High consistency, rarely affected by incorrect answers. Strategic in answering questions, focusing on quality over speed.

**Weaknesses:** May take more time, requiring a balance between accuracy and speed. Less need for improvement, but could benefit from additional challenges.

**Recommended Strategy for Growth:** Participate in time-based challenges to improve efficiency without losing accuracy. Attempt harder problems to avoid stagnation. Engage in mentoring to reinforce learning by guiding lower-level students.

### The Intermediate Performer (Trophy Level 2 )

**Profile Summary:** Hardworking, improving steadily, above average but not elite yet.

**Performance Insights:** Consistently achieves moderate to high scores, showing steady improvement. Positive correlation between mistakes corrected and score improvement indicates active learning. Negative scores exist but at a reduced level compared to lower-tier students. Some inconsistency remains in the speed vs. accuracy trade-off.

**Strengths:** Strong grasp of concepts with room for improvement. Actively corrects mistakes, showing learning progress. Balanced approach to time and accuracy.

**Weaknesses:** Inconsistencies in performance; some questions are still incorrectly answered. Still prone to negative scores, suggesting occasional carelessness. Needs greater confidence in applying knowledge.

**Recommended Strategy for Growth:** Focus on minimizing careless mistakes to further improve scores. Increase exposure to different question types for greater adaptability. Track speed to ensure efficiency without compromising accuracy.

### The Struggling Learner (Trophy Level 1 )

**Profile Summary:** Inconsistent, experimental, making frequent mistakes, and lacking stability.

**Performance Insights:** High variability in correct and incorrect answers, suggesting inconsistent understanding. Negative scores are significantly higher, indicating frequent errors. Strong correlation between mistakes corrected and score improvement, but the starting point is much lower than other levels. Less concentration in higher ranks, meaning performance needs stabilization.

**Strengths:** Open to learning, as seen in attempts to correct mistakes. Shows potential for improvement when guided properly. Can transition to higher levels with structured practice.

**Weaknesses:** High mistake frequency, leading to penalty accumulation. Score fluctuation suggests a lack of deep conceptual understanding. Struggles with balancing speed and accuracy, often rushing through questions.

**Recommended Strategy for Growth:** Practice fundamental concepts rigorously to reduce incorrect responses. Take guided step-by-step exercises before attempting complex problems. Start slow, then gradually increase speed while maintaining accuracy.