Data Loading and Cleaning:

1. Why replace dashes with NaN when loading data?

Answer: Dashes often represent missing or invalid values. In data analysis and machine learning, we replace them with NaN for consistent handling and later cleaning.

2. Why clean and fill missing values in certain columns?

Answer: Data cleaning ensures model inputs are valid. Filling missing values helps prevent issues during model training and prediction.

Feature Engineering and Modeling:

3. Why choose these specific features for model input?

Answer: Features are chosen based on domain knowledge and their potential impact on predicting final scores and student learning patterns.

4. Why use different regression models for training?

Answer: Different models have different characteristics. Using multiple models allows us to compare their performance and choose the most suitable one.

5. Why convert scores to a numeric format?

Answer: Converting scores to numeric format is essential for mathematical operations and modeling. It helps in treating scores as numerical values rather than text.

6. Why drop certain rows based on specific conditions?

Answer: Dropping rows based on conditions, such as removing data with certain grades or zero values, helps filter out irrelevant or potentially problematic data, ensuring the quality of the dataset.

7. What is the significance of converting datetime columns?

Answer: Converting datetime columns is crucial for time-based analysis. It allows for effective handling of time-related operations and filtering.

8. How does handling invalid log data contribute to data quality?

Answer: Removing invalid log entries ensures that only reliable and meaningful data is used for analysis. This step improves the overall quality of the dataset.

9. Why drop rows with missing scores?

Answer: Dropping rows with missing scores ensures that we work with complete data, avoiding potential issues during analysis and modeling caused by incomplete information.

10. Why remove data with certain grades and specific conditions?

Answer: Removing data with specific grades or conditions focuses the analysis on relevant and meaningful information. It helps in maintaining the quality and accuracy of the dataset.

11. How does eliminating data with zero scores contribute to the analysis?

Answer: Eliminating data with zero scores removes instances where students may not have participated or completed activities. This ensures that the analysis reflects engaged and meaningful student interactions.

12. What's the rationale behind dropping log entries with invalid timestamps?

Answer: Dropping log entries with invalid timestamps ensures that time-related analyses are based on accurate and valid data. It contributes to the reliability of time-dependent insights.

13. Why filter out log entries beyond certain time constraints?

Answer: Filtering out log entries beyond certain time constraints ensures that only relevant and recent data is considered. This contributes to a more focused analysis, particularly in time-sensitive scenarios.

14. How does retaining data with maximum scores enhance the analysis?

Answer: Retaining data with maximum scores focuses the analysis on instances where students achieved their best performance. This helps in understanding and highlighting outstanding achievements within the dataset.

Model Evaluation and Comparison:

5. Why use MSE and MAE as performance metrics?

Answer: MSE and MAE are common metrics for regression problems, measuring the difference between predicted and actual values. Smaller values indicate better prediction accuracy.

6. Which model performed best according to MSE and MAE?

Answer: Results suggest the Random Forest

model performed best with the smallest MSE and MAE, indicating the most accurate predictions.

Project Goals and Future Work:

7. How does the model help identify students facing difficulties?

Answer: The model can identify students with lower-than-expected final scores or abnormal learning patterns, helping schools provide early support.

8. How will the project classify students based on typical learning patterns?

Answer: This can be achieved using clustering algorithms or classification models to group students with similar learning patterns into different categories.