Data Exploration and Analysis Report

1. Overview of the Dataset

The dataset we analyzed consists of multiple CSV files, each representing different aspects of student learning, classroom activities, and engagement with a digital learning platform. The data is structured into different tables, each focusing on a specific component of the learning experience.

Datasets Available:

- Student Data: Contains student demographic details and classroom information.
- Classroom Analytics: Captures assignments and progress at the classroom level.
- Learning Analytics: Tracks students' learning progress, including text complexity and concept mastery.
- Motivation Metrics: Measures time spent on the platform and engagement levels.
- Performance Metrics: Records students' lesson completion rates, game scores, and reading comprehension scores.
- User Experience Metrics: Captures navigation patterns and help requests.
- Planets Analytics: Monitors classroom engagement with various planets in the platform.

2. Initial Thoughts Upon Seeing the Data

Upon first glance, the dataset appeared well-structured and contained valuable insights into student performance and engagement. Some key observations:

- The data is **highly relational**, with student_id and classroom_id acting as primary linking fields.
- Each dataset contributes a **different dimension** to student performance and learning.
- Some datasets, such as **Performance Metrics**, seem to be more central, as they reflect direct learning outcomes.
- The **mapping between datasets** is clear in most cases but requires merging on common keys for a holistic analysis.

3. Objective of the Analysis

Our primary goal was to **understand student performance** by evaluating the relationships between different datasets. Specifically, we aimed to:

- Identify **how classroom activities impact learning performance**.
- Examine **motivational factors** influencing student engagement.
- Assess **how user experience correlates with performance**.
- Determine which metrics are the most **influential predictors** of student success.

4. Relationship Between Datasets

- **Student Data** links to most datasets using student_id, making it foundational.
- **Classroom Analytics** connects to planets_analytics, indicating classroom-level performance at a higher level.
- **Learning, Motivation, and Performance Metrics** are directly student-based, helping us analyze individual learning patterns.
- **User Experience Metrics** provide insights into engagement patterns, which may affect performance outcomes.

5. Detailed Dataset Analysis

A. Student Data

- **Key Fields**: student_id, first_name, last_name, classroom, email_address
- **Purpose**: Serves as a base table to link all other datasets.
- **Insights**: Helps us track individual students across multiple datasets.
- **Concerns**: No direct performance indicators, only demographic details.

B. Classroom Analytics

- **Key Fields**: classroom_id, assignments_created, percentage_completed, planets_unlocked
- **Purpose**: Provides aggregated classroom-level data on assignments and progress.
- **Insights**:
  - Number of assignments could correlate with student performance.
  - Higher completion percentages indicate better classroom engagement.
  - More planets unlocked could mean higher interaction with the platform.

C. Learning Analytics

- **Key Fields**: student_id, text_complexity, student_performance, rate_of_improvement_in_reading_skills, concept_mastery_across_different_texts
- **Purpose**: Tracks learning progression on an individual level.
- **Insights**:
  - Text complexity may indicate a student's reading proficiency level.
  - Rate of improvement is a useful measure of learning progression.
  - Concept mastery can indicate depth of understanding.

D. Motivation Metrics

- **Key Fields**: student_id, total_time_spent_on_the_platform_mins, time_spent_on_each_planet_mins, time_spent_per_game_mins, frequency_of_logins
- **Purpose**: Measures engagement and time investment on the platform.
- **Insights**:
  - More time spent generally correlates with better learning outcomes.
  - Frequent logins suggest higher engagement and motivation.
  - Disproportionate time spent on games vs. reading could indicate a lack of focus on core learning activities.

E. Performance Metrics (Most Insightful Dataset)

- **Key Fields**: student_id, lesson_completion_rates, game_scores_for_mini_games, reading_comp_scores, frequency_of_reading_strategy_usage
- **Purpose**: This is the most critical dataset as it directly reflects student performance.
- **Why is it the most important?**
  - It provides **quantifiable success measures**.
  - It directly connects learning efforts (from Learning Analytics) to tangible results.
  - It allows us to track **progress over time**, unlike motivation or engagement, which may fluctuate without direct impact.
  - All other datasets contribute **indirectly** to performance but do not measure actual learning outcomes like this one does.

F. User Experience Metrics

- **Key Fields**: student_id, navigation_patterns_between_planets_and_activities, number_of_help_requests
- **Purpose**: Captures how students interact with the platform.
- **Insights**:
  - Frequent navigation might indicate **exploratory learning** or **difficulty finding content**.
  - A high number of help requests may correlate with **struggles in learning**.
  - Smooth navigation patterns may indicate **strong platform usability**.

G. Planets Analytics

- **Key Fields**: classroom_id, planet_id, number_of_active_students
- **Purpose**: Tracks student participation in different learning environments.
- **Insights**:
  - Active student counts per planet help us **understand platform engagement**.
  - Comparing planets across classrooms might reveal **which environments are more effective**.
  - If planets_unlocked is high, it suggests that students are actively progressing through the platform.

6. Conclusion and Next Steps

Key Takeaways:

- **Performance Metrics is the most insightful dataset** as it quantifies learning outcomes.
- **Motivation, Learning, and User Experience metrics** contribute to performance but are **indirect indicators**.
- **Classroom Analytics and Planets Analytics provide macro-level insights** but need to be linked to individual student data for deeper analysis.

Key Insights:

Performed Correlation matrix for the "Performance Metrics" columns to identify which columns relate to each other and the correlation coefficient.

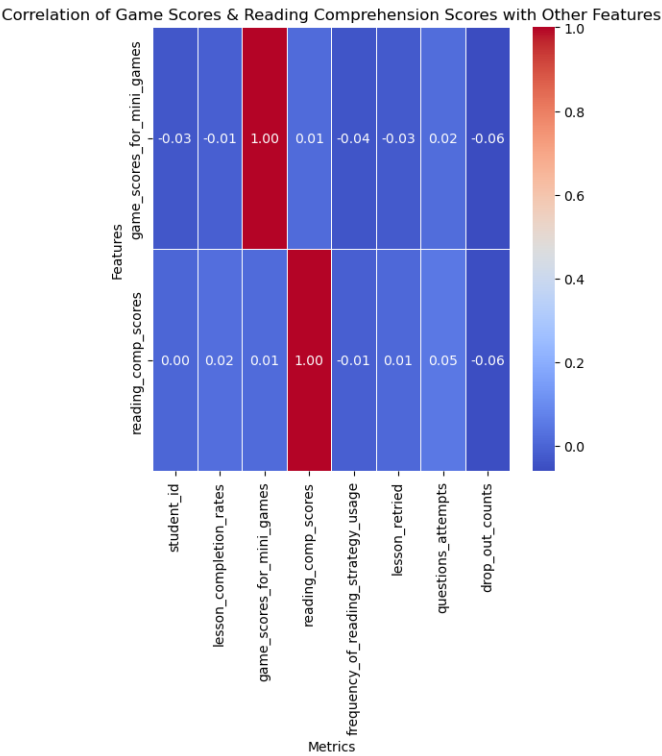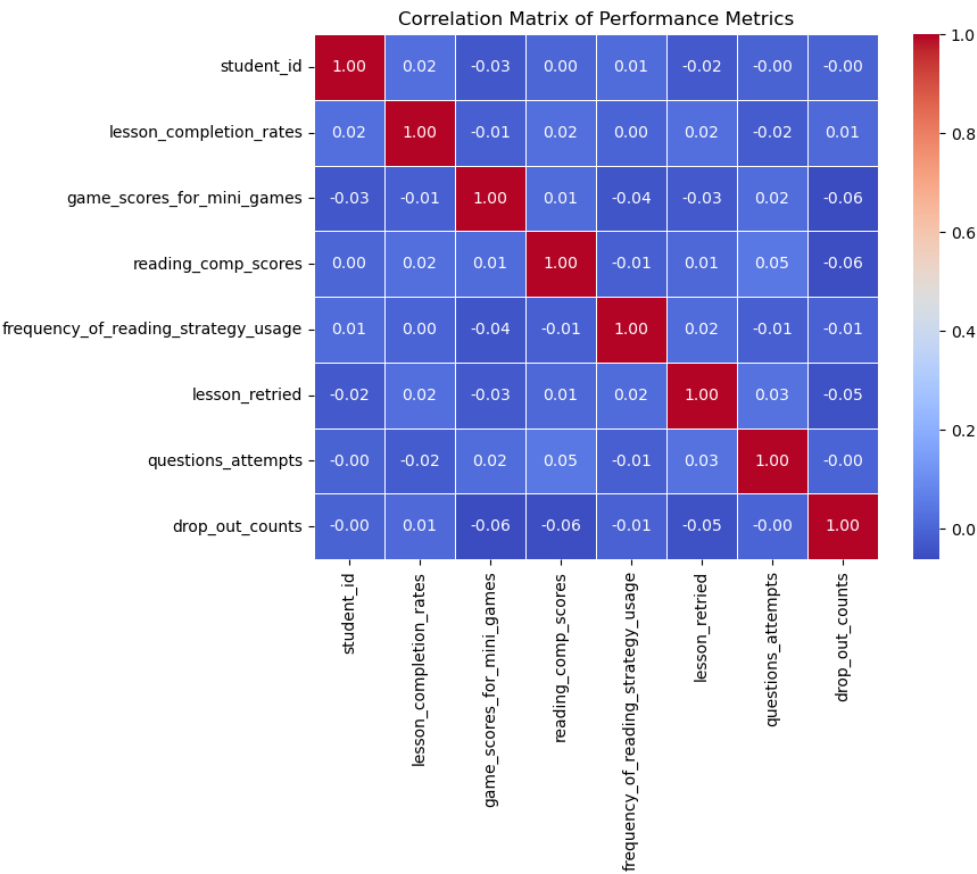1. Correlation Matrix before adding new columns
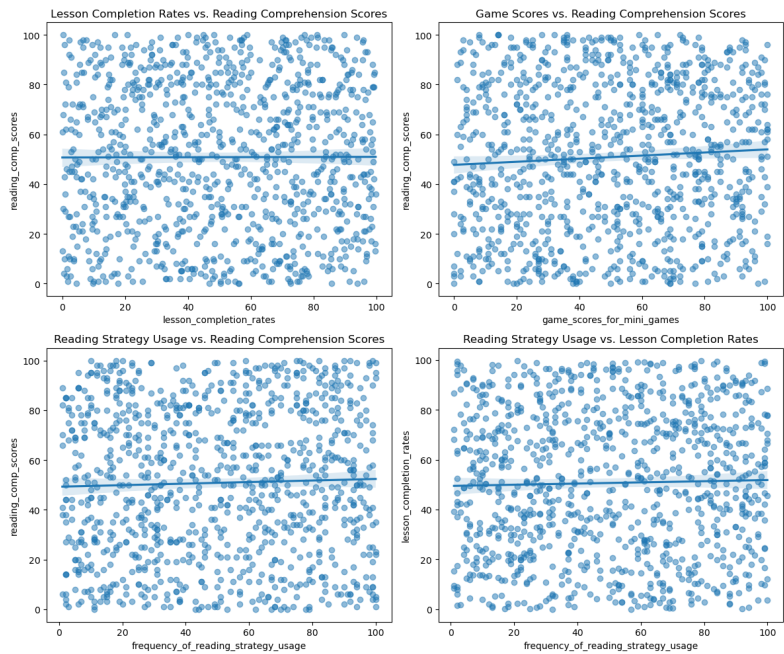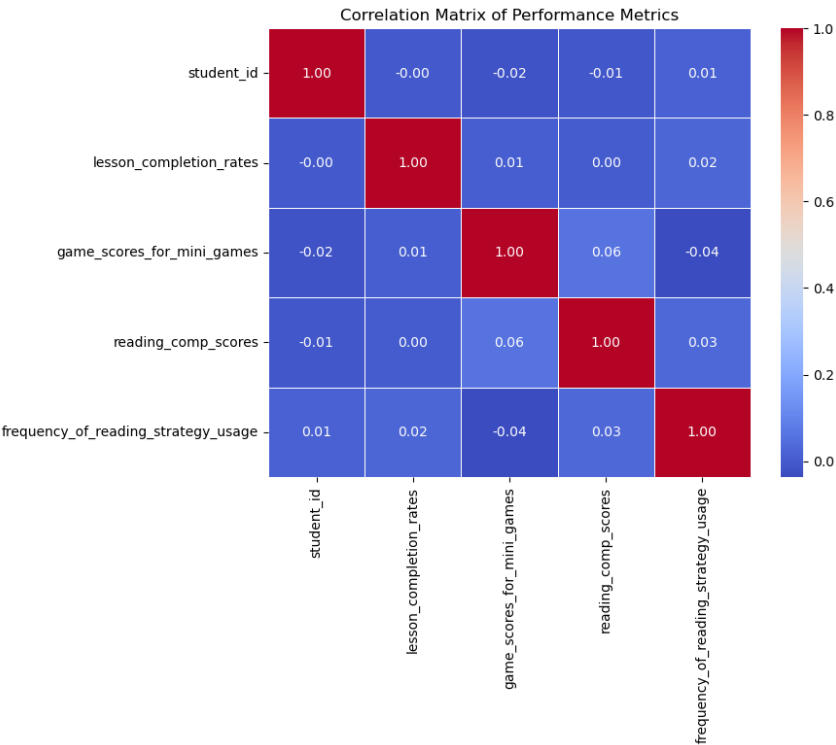
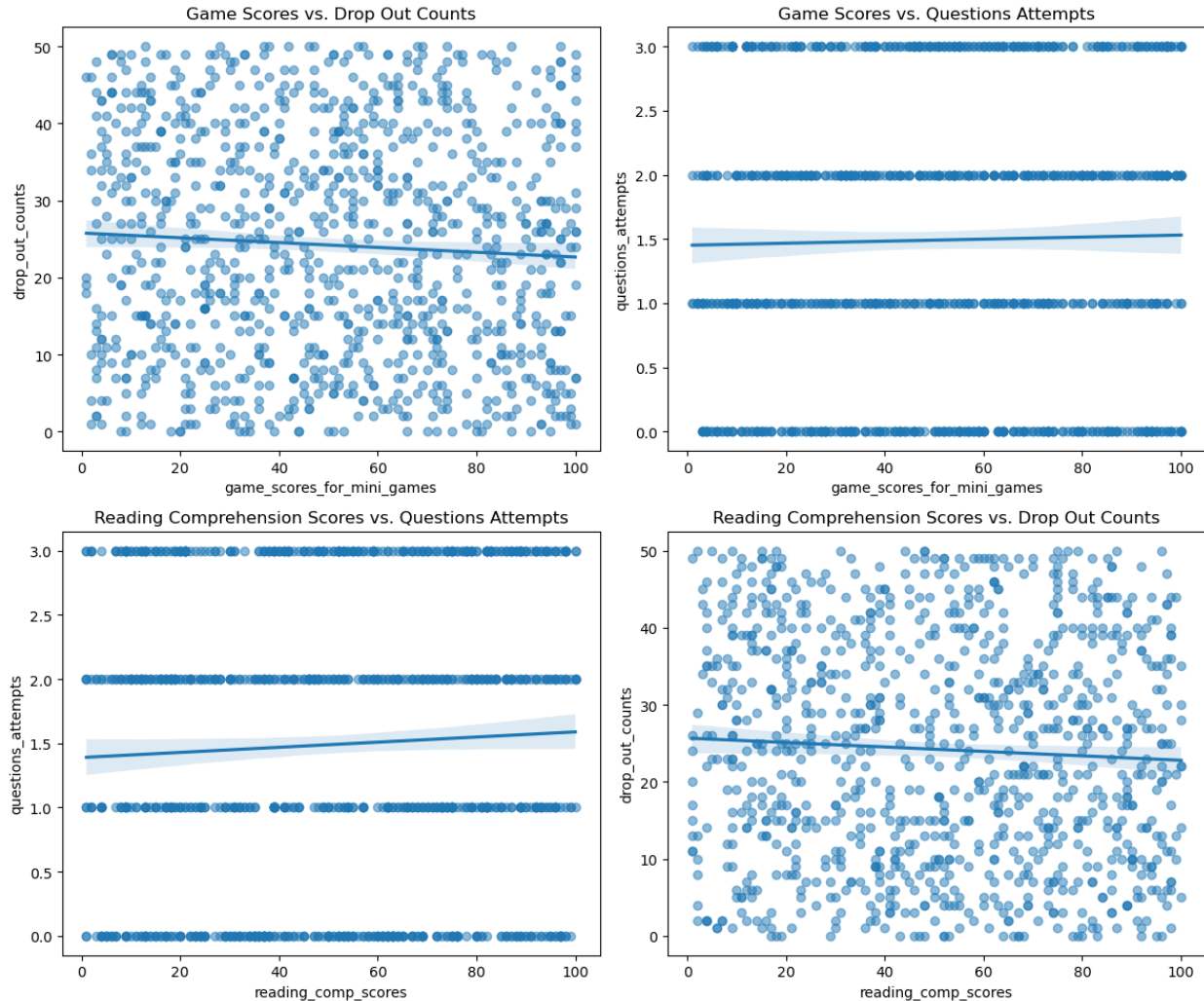Few Insights that we derived from the correlation matrix:

• Game Scores for mini games has a strong correlation with reading comprehension scores which indicates if the student has commensurate scores for both game scores and reading comprehension scores.

• Game Scores for mini games has also a strong correlation with frequency of strategy use but with a negative magnitude which indicates the scores are bound to increase if the strategy use is kept at minimum which indicates the student has good mastery over the concepts.

• Surprisingly, lesson completion rates and reading comprehension scores had no correlation.

• Frequency of reading strategies had almost no effect on performance.

We added new columns based on on the business requirements and further analysis:

Lessons Re-tried(0-50)	New Feature (Performance)	Measure persistence
Question Attempts(0-3)	New Feature (Performance)	Understand difficulty levels & effort
Dropout Counts(0-50)	Dropout Counts(0-50)	Identify students who start but don't finish

2. Correlation Matrix after adding new columns

Correlation Matrix of Performance Metrics

Correlation of Game Scores & Reading Comprehension Scores with Other Features

Few Insights that we derived from the correlation matrix:

- Game Scores for mini games has a strong correlation with drop out counts but with a negative magnitude, indicating the scores increased when the drop out counts reduced.
- Reading comprehension scores also has a strong correlation with drop out counts but with a negative magnitude, indicating the scores increased when the drop out counts reduced.
- Reading comprehension scores has a positive correlation with questions attempt indicating the scores increased with the number of questions attempted.