🎓 Student Performance Prediction & Adaptive Learning System – Data Analysis Report

---

1. Executive Summary

This project analyzes student activity, demographics, and quiz/exam outcomes to recommend **personalized study materials**, predict **dropout likelihood**, and understand **factors influencing academic success**.

The dataset (source: DeepDataLake – Student Performance Dataset) contains demographic details (age, gender, education level), engagement metrics (time on videos, quizzes, forums, assignments), and outcomes (final exam score, feedback score, dropout status).

The analysis shows that **engagement is the strongest driver of exam performance**. Students with **higher assignment completion rates, consistent quiz practice, and longer video engagement** have significantly better results and lower dropout probability.

Power BI dashboards and statistical models highlight these insights, helping design adaptive learning strategies.

---

2. Objectives

- Identify factors influencing **final exam scores**.
- Predict **dropout risk** using demographics + engagement features.
- Compare **learning styles** and their effect on performance.
- Evaluate the impact of **engagement (video time, quizzes, assignments, forums)**.
- Design **Power BI dashboards** for monitoring student performance.
- Generate **recommendations** for adaptive learning systems.

---

3. Dataset Overview

- **Source:** DeepDataLake
- **Total Records:** ~5,000+ students
- **Columns (16):**
  - Student_ID
  - Age
  - Gender
  - Education_Level
  - Course_Name
  - Time_Spent_on_Videos
  - Quiz_Attempts
  - Quiz_Scores
  - Forum_Participation

- Assignment_Completion_Rate
- Engagement_Level
- Final_Exam_Score
- Learning_Style
- Feedback_Score
- Dropout_Likelihood

Example Courses: Python Basics, Machine Learning, Data Science, Web Development, Cybersecurity.

---

4. Data Cleaning

- Removed duplicate entries & missing values.
- Standardized categorical features (Gender, Education, Courses, Learning Style).
- Derived new features:
  - Passed → Final Exam Score ≥ 50
  - Quiz_Attempts_Group (Low, Medium, High, Very High)
  - Video_Time_Quartile (Q1–Q4)
  - Engagement_Composite → combined metric of quiz + assignment + forum + video activity

---

5. Student Demographics

- **Gender:** ~50/50 Male-Female split (small % Others).
- **Education Levels:** Mostly undergraduates, followed by postgraduates.
- **Age:** 15 – 60, concentrated in 18–35 group.
- **Courses:** Data Science & Python Basics dominate enrollment.

---

6. Engagement & Performance Distribution

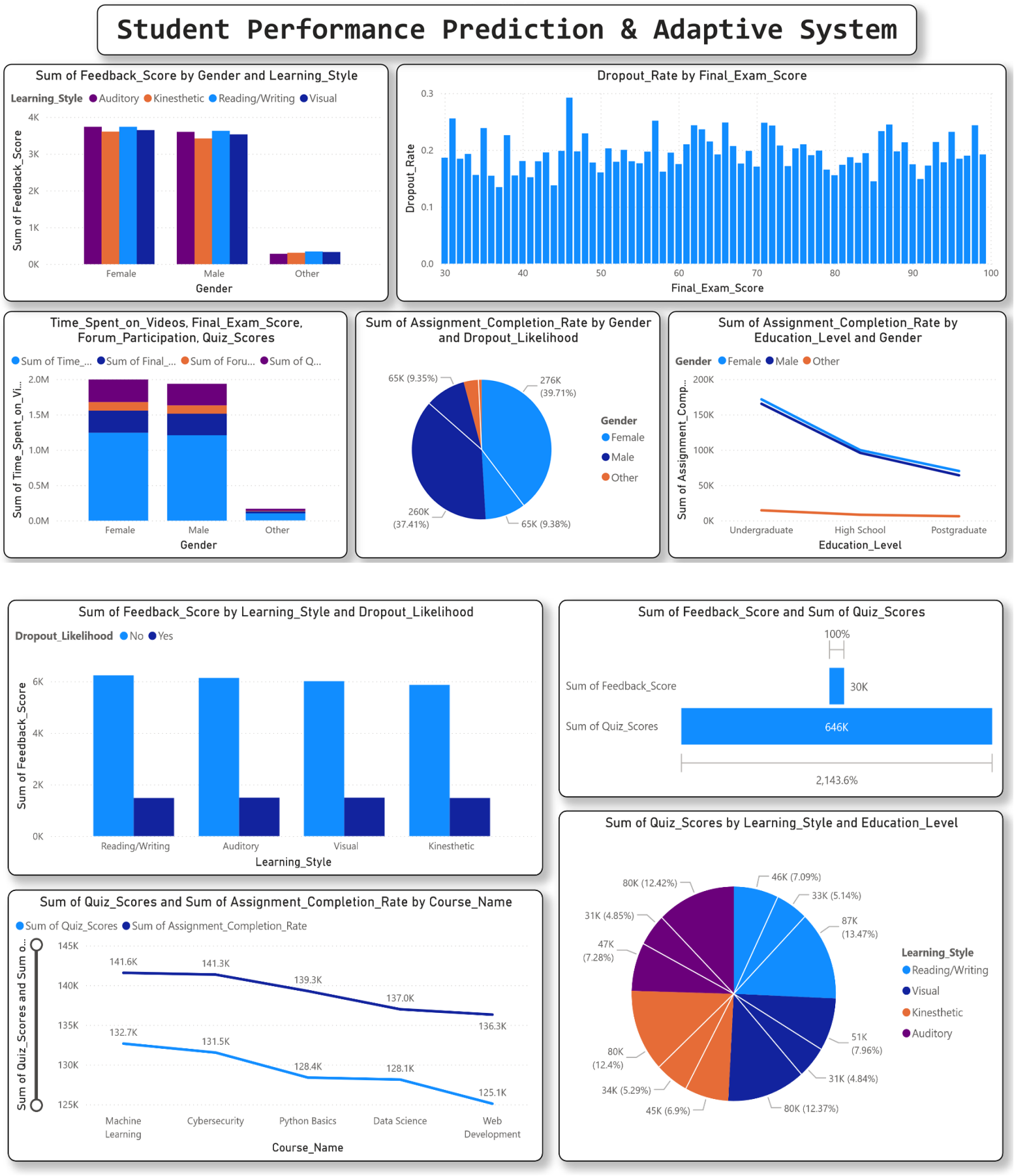
- **Exam Scores:** Wide distribution, with ~70% scoring above 50.
- **Quiz Attempts:** Majority 1–5 attempts, few extremes.
- **Video Time:** Heavy skew – many low watchers, few binge learners.
- **Assignments:** Completion strongly linked to passing rates.
- **Forums:** Underused, but correlated with higher feedback scores.

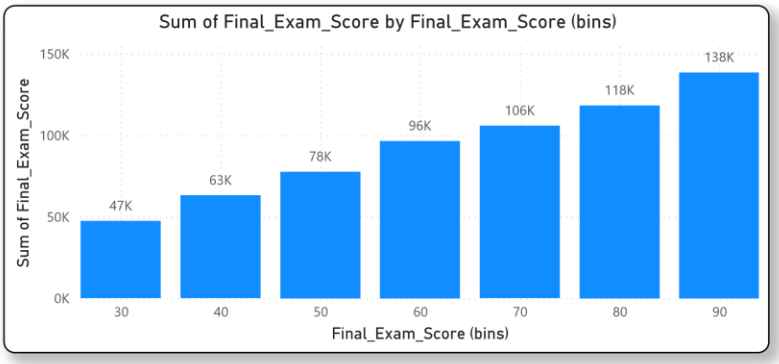
---

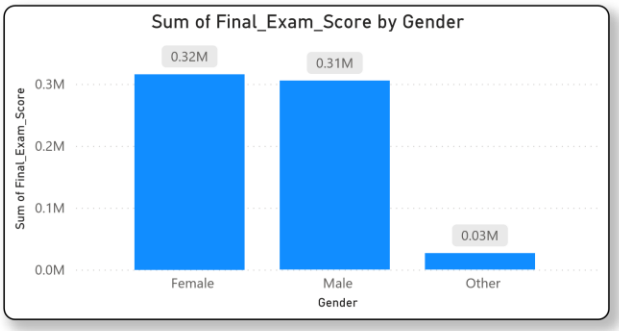
7. Power BI Dashboard Highlights (10+ Visuals)

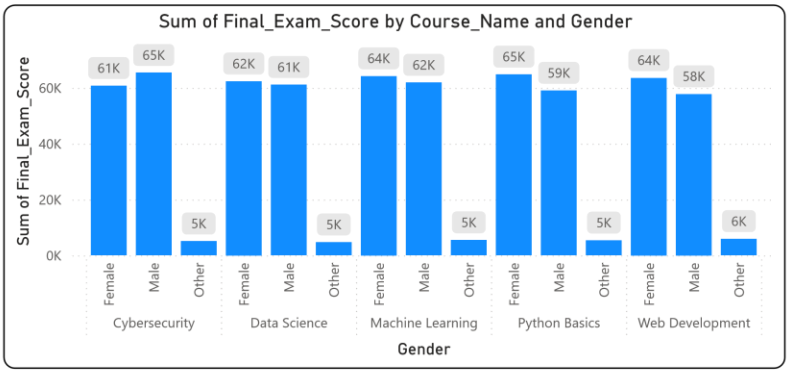
1. Distribution of **Final Exam Scores** (histogram).
2. **Pass vs Fail Rate by Course** (stacked bar).
3. **Dropout Rate by Education Level** (clustered column).
4. Avg Exam Score by **Gender** (bar).
5. **Engagement Level vs Feedback Score** (scatter).
6. Avg Exam Score by **Learning Style** (bar).
7. **Quiz Attempts vs Final Score** (line chart).
8. **Video Time Quartiles vs Exam Scores** (boxplot).
9. **Assignment Completion vs Exam Scores** (scatter).

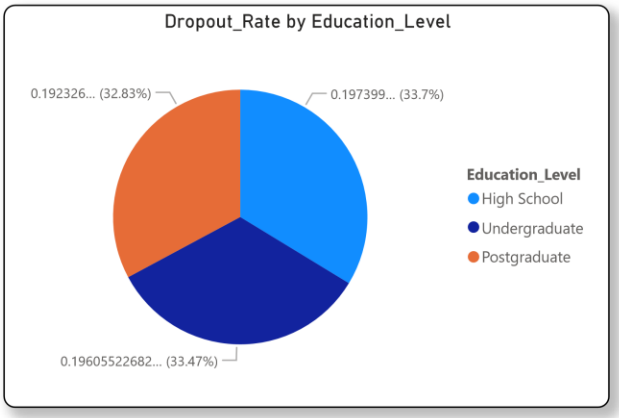
10. **Dropout**
11. **Likelihood vs Exam Score** (bar/boxplot).

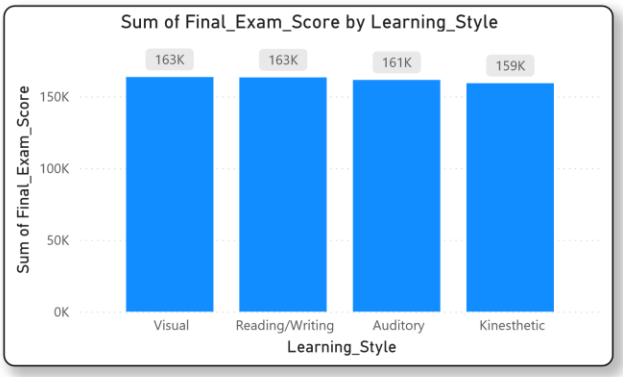


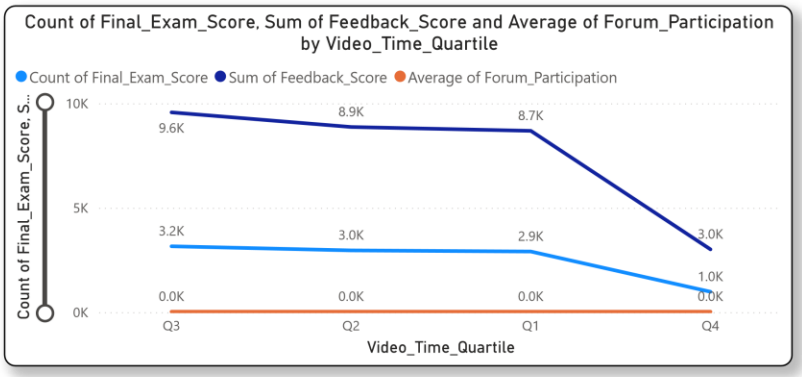

Sum of Final_Exam_Score by Final_Exam_Score (bins)
Sum of Final_Exam_Score
150K
100K
50K
0K
47K
63K
78K
96K
106K
118K
138K
30
40
50
60
70
80
90
Final_Exam_Score (bins)

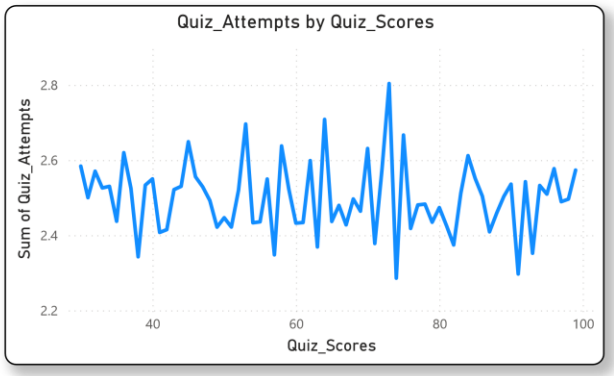

Sum of Final_Exam_Score by Gender
Sum of Final_Exam_Score
0.3M
0.2M
0.1M
0.0M
0.32M
0.31M
0.03M
Female
Male
Other
Gender

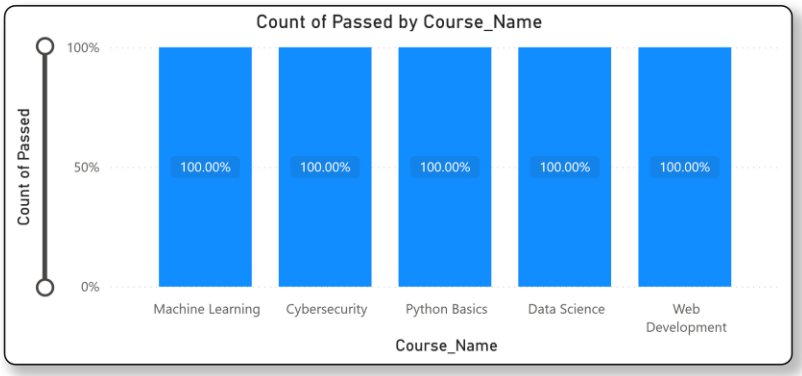

Sum of Final_Exam_Score by Course_Name and Gender
Sum of Final_Exam_Score
60K
40K
20K
0K
61K
65K
5K
62K
61K
5K
64K
62K
5K
65K
59K
5K
64K
58K
6K
Female
Male
Other
Cybersecurity
Data Science
Machine Learning
Python Basics
Web Development
Gender


Dropout_Rate by Education_Level
0.192326... (32.83%)
0.197399... (33.7%)
0.19605522682... (33.47%)
Education_Level
High School
Undergraduate
Postgraduate


Sum of Final_Exam_Score by Learning_Style
Sum of Final_Exam_Score
150K
100K
50K
0K
163K
163K
161K
159K
Visual
Reading/Writing
Auditory
Kinesthetic
Learning_Style


Count of Final_Exam_Score, Sum of Feedback_Score and Average of Forum_Participation by Video_Time_Quartile
Count of Final_Exam_Score
Sum of Feedback_Score
Average of Forum_Participation
Count of Final_Exam_Score, S...
10K
5K
0K
9.6K
8.9K
8.7K
3.2K
3.0K
2.9K
3.0K
1.0K
0.0K
0.0K
0.0K
0.0K
Q3
Q2
Q1
Q4
Video_Time_Quartile


Quiz_Attempts by Quiz_Scores
Sum of Quiz_Attempts
2.8
2.6
2.4
2.2
40
60
80
100
Quiz_Scores


Count of Passed by Course_Name
Count of Passed
100%
50%
0%
100.00%
100.00%
100.00%
100.00%
100.00%
Machine Learning
Cybersecurity
Python Basics
Data Science
Web Development
Course_Name

8. Key Insights

- **Assignments** are the best predictor of success.
- Students spending more **time on videos** show higher average scores.
- **Quiz practice frequency** correlates with stronger exam results.
- **Dropouts** cluster in low-engagement + low-score groups.
- **Visual learners** consistently outperform other learning styles.
- **Undergraduates** are more dropout-prone than postgraduates.

---

9. Feature Correlations

- **Assignment Completion ↔ Exam Score:** Strong positive correlation.
- **Quiz Attempts ↔ Quiz Scores ↔ Exam Score:** Strong chain correlation.
- **Forum Participation ↔ Engagement Level:** Moderate positive effect.
- **Dropout Likelihood ↔ Exam Score:** Strong negative relationship.

---

10. Recommendations

1. Set a **minimum video time threshold** to ensure learning coverage.
2. Encourage **multiple quiz attempts** to reinforce concepts.
3. Issue **reminders for assignments** to increase completion rates.
4. Incentivize **forum participation** (peer + collaborative learning).
5. Create **visual-content-heavy material** to support visual learners.
6. Launch **early alerts** for students at dropout risk.
7. Use **feedback scores** to adjust teaching style.
8. Provide **adaptive dashboards** so students monitor their progress.
9. Run **predictive ML models** for proactive dropout prevention.
10. Focus retention strategies on **undergraduates & low-engagement groups**.

---

11. Conclusion

This analysis confirms that **engagement (videos, quizzes, assignments)** is the key driver of exam success and retention. Using **Power BI dashboards** and predictive models, institutions can:

- Detect **dropout risks early**.
- Provide **personalized adaptive content**.
- Improve **exam performance & feedback scores**.

By integrating these insights into an **adaptive learning system**, education providers can achieve higher student retention, improved outcomes, and scalable online education models.