Understanding Student Success in Higher Education

Prepared by: Nader Mohamed

Why This Matters

In modern universities, success is shaped by more than just grades.

Now we also analyze:

- Academic performance
- LMS engagement
- Risk level labels

> 📌 **So we asked:**
>
> - What makes a student high risk?
> - Is poor attendance enough to predict failure?
> - Can we help before it's too late?

! The Challenge

👨‍🏫 For Academic Advisors:

- Can we catch early signs of failure?
- Do quiet students struggle in silence?
- Is GPA enough?

🏛️ For University Leadership:

- Are some majors at higher risk?
- Do younger students drop out more?
- Is online engagement the key?

The Dataset

This dataset contains anonymized student records, including:

- 👨‍🎓 Demographics: Age, gender, major
- 📚 Academics: GPA, course load, attendance
- 💻 Behavior: LMS logins, videos, forums, submissions
- ⚠️ Risk Labels: Low, Medium, High

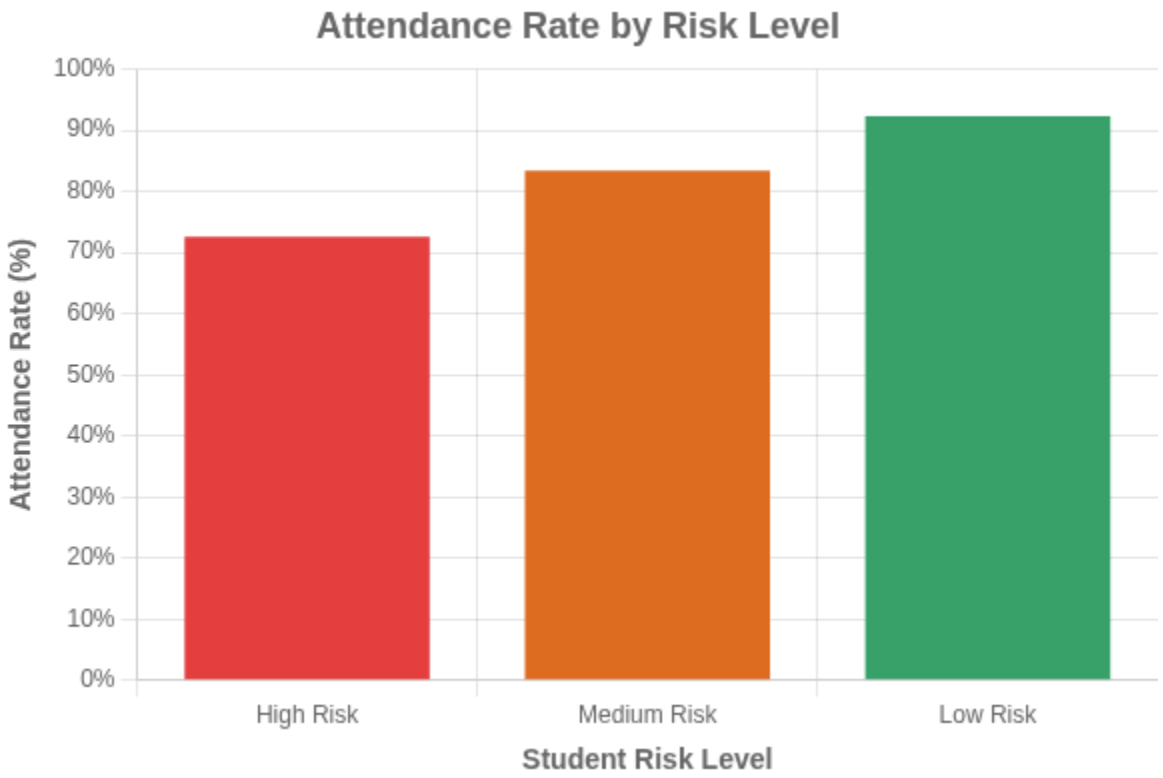
Why these data points matter:

- They help us identify struggling students early
- They reveal patterns across majors and demographics
- They connect online behavior to academic outcomes

Attendance Predicts Risk

Risk Level	Avg Attendance Rate
High	72.6%
Medium	83.4%
Low	92.3%

> **Insight:**
>
> **Students with <75% attendance are at risk**
>
> Early warning systems should monitor attendance



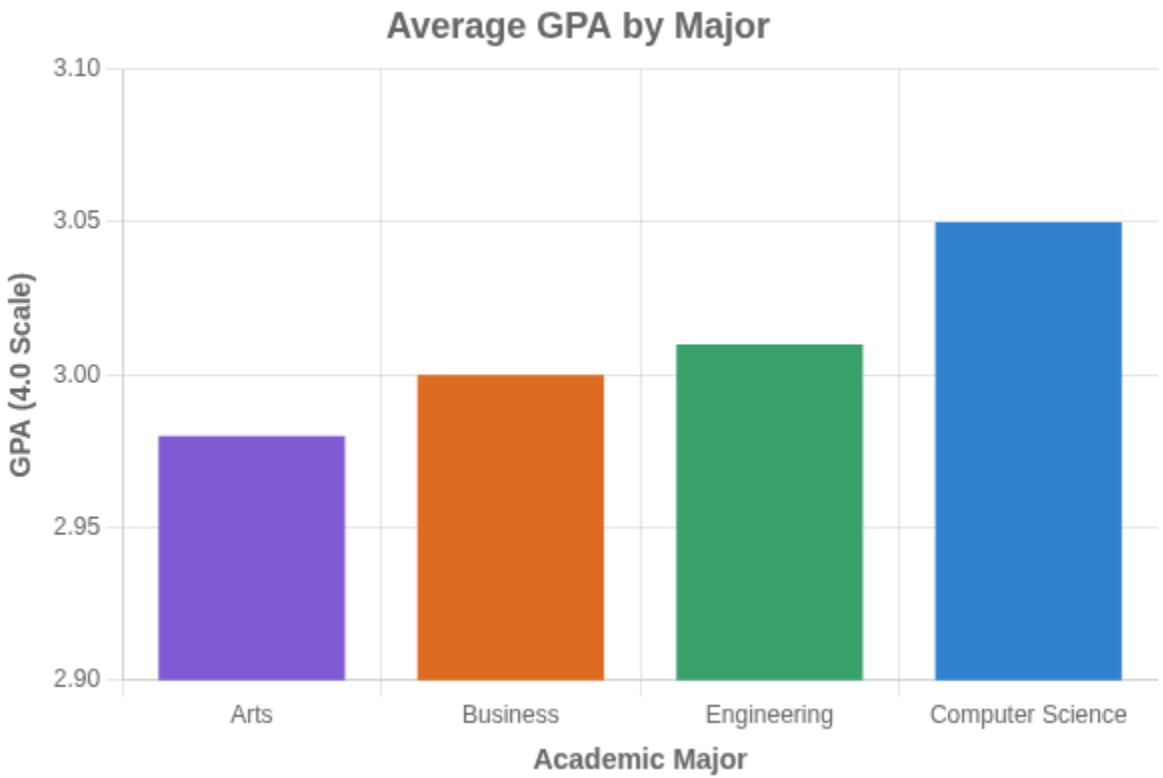
GPA by Major

Major	Avg GPA
Arts	2.98
Business	3.00
Engineering	3.01
Computer Science	3.05

📌 Insight:

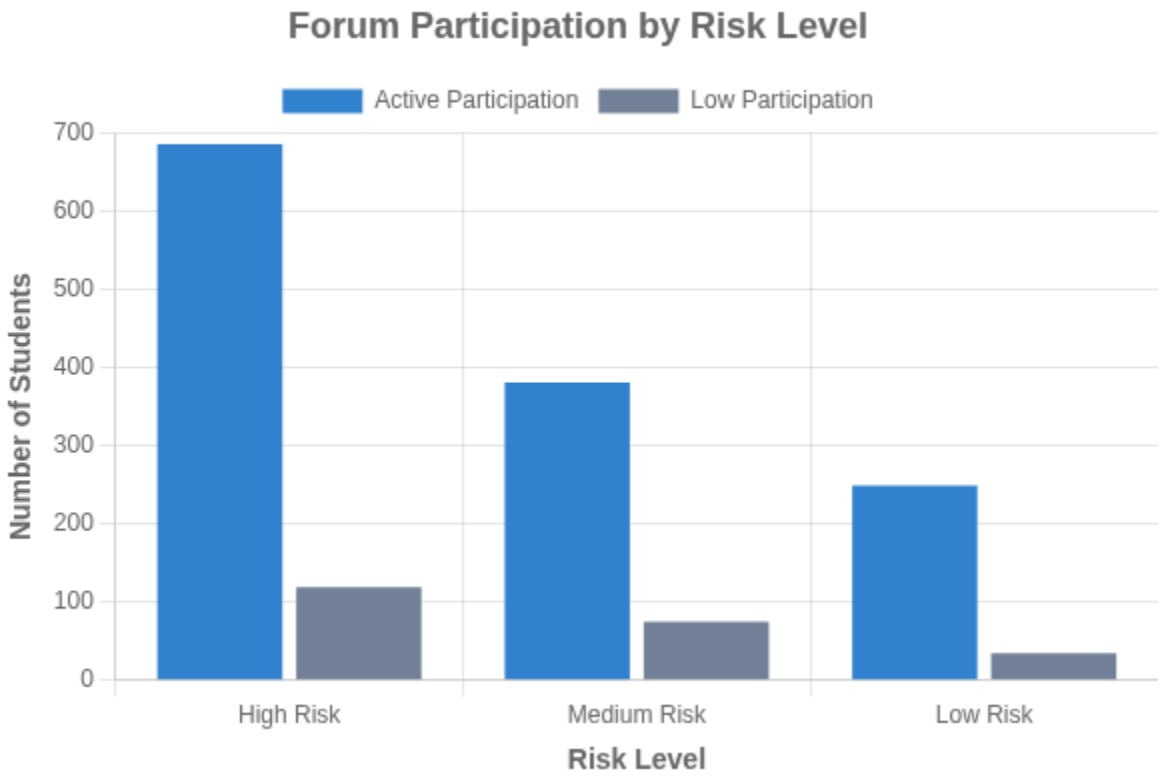
CS students slightly outperform

Arts students might need extra academic support



Forum Participation Signals Risk

Participation	Risk Level	Count
Active	High	686
Low	High	119
Active	Medium	381
Low	Medium	75
Active	Low	249
Low	Low	35



📌 Insight:

Low forum engagement = higher risk

LMS activity matters

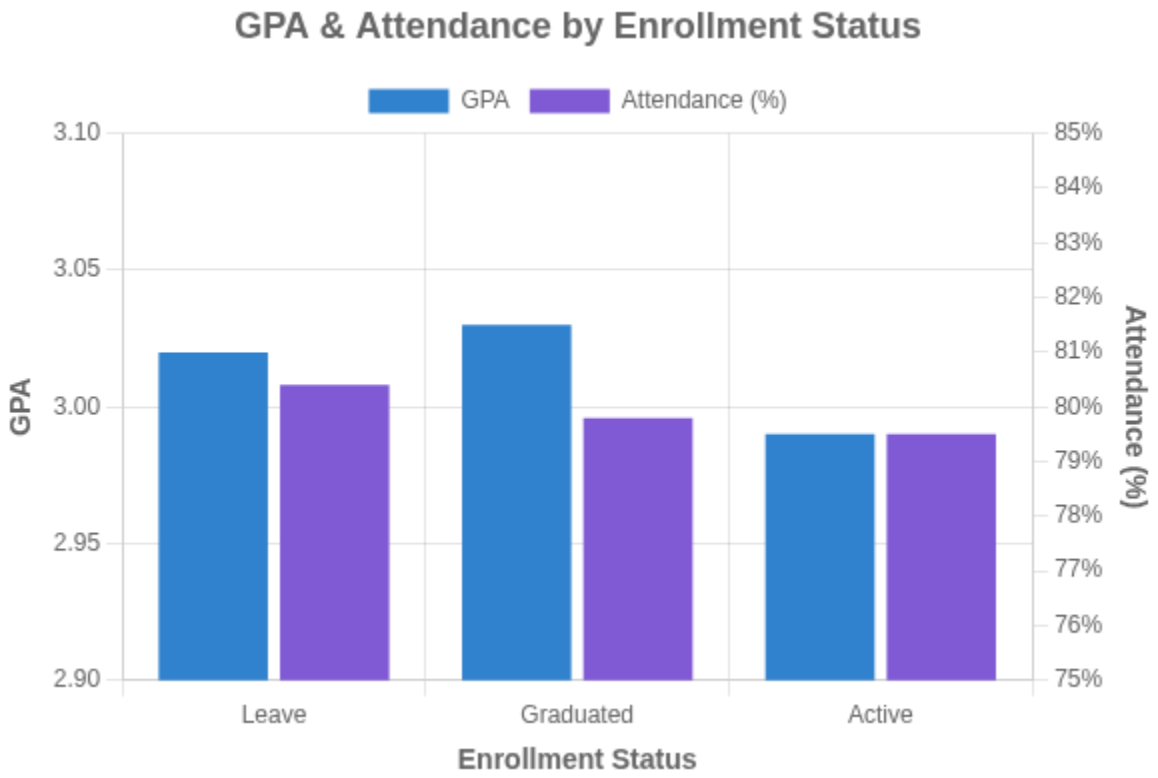
GPA & Attendance by Enrollment

Enrollment Status	GPA	Attendance (%)
Leave	3.02	80.4%
Graduated	3.03	79.8%
Active	2.99	79.5%

📌 Insight:

Active students slightly underperform

🎯 Possible signs of stress, burnout



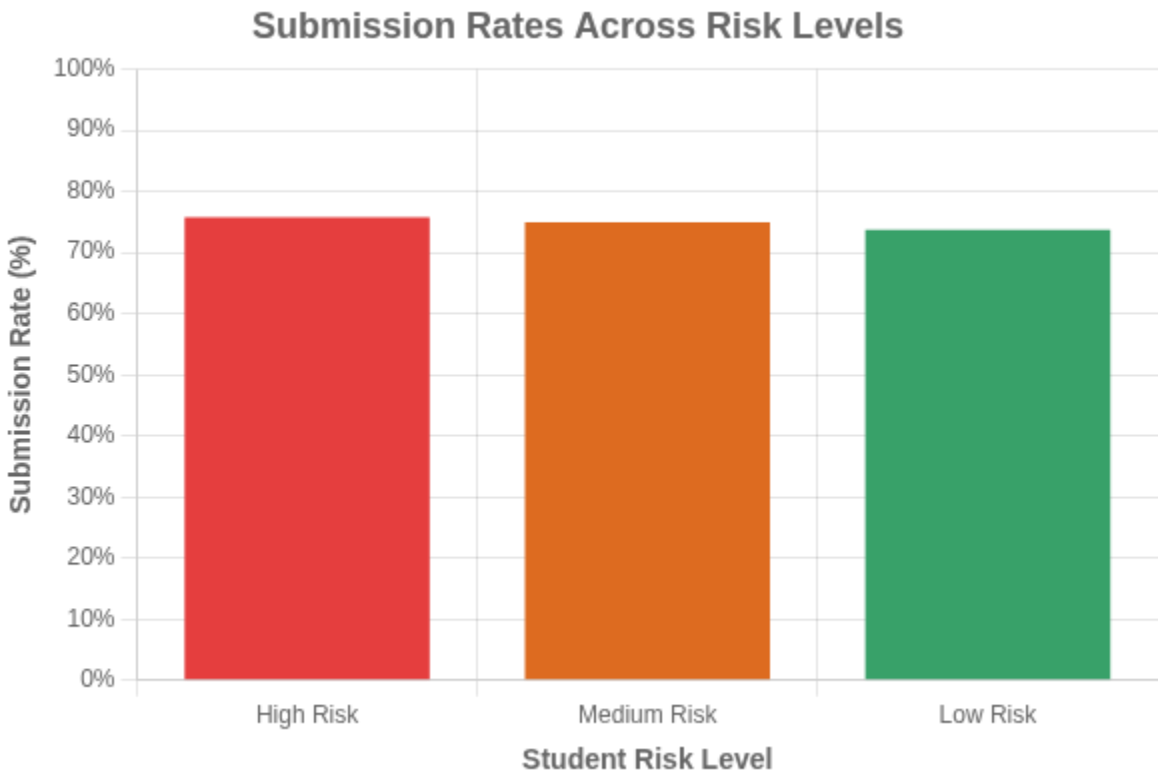
Assignment Submission Rate

Risk Level	Submission Rate
High	75.8%
Medium	75.0%
Low	73.7%

Insight:

Submission rate doesn't always reflect risk

Quality > Quantity?



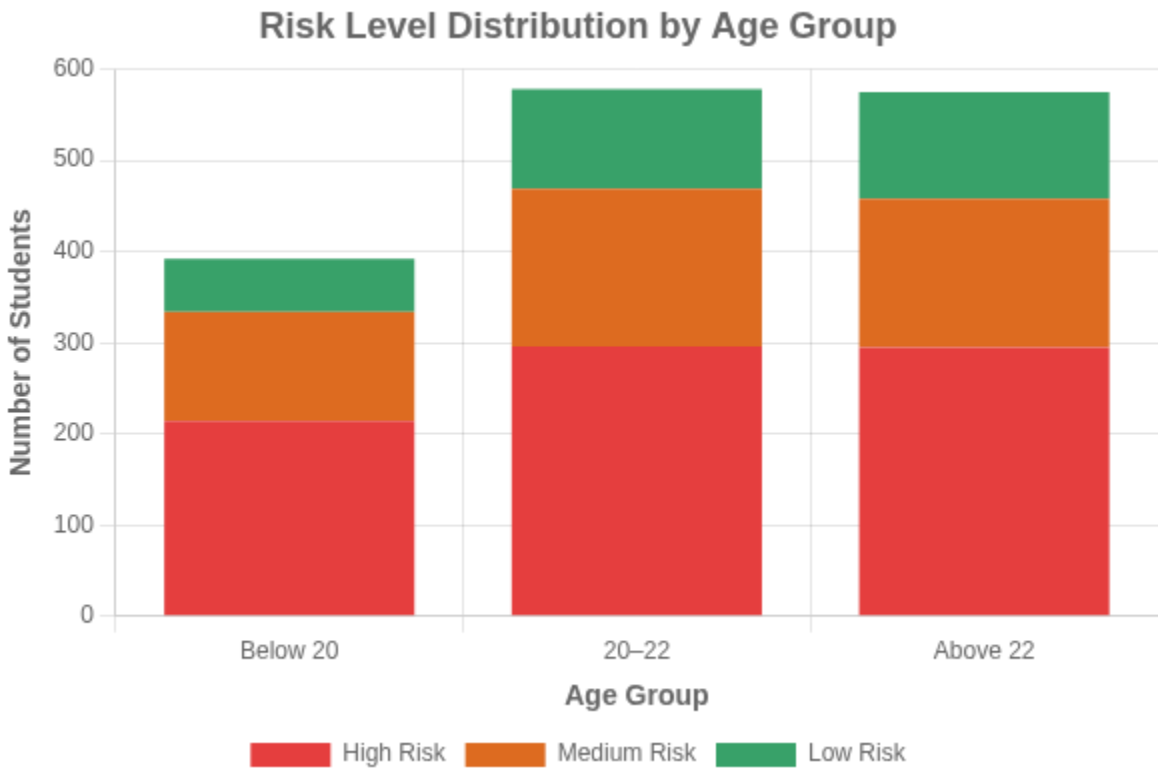
Younger = Riskier

Age Group	High Risk	Medium	Low
Below 20	214	120	58
20–22	296	173	109
Above 22	295	163	117

📌 Insight:

Students under 20 face higher risk

Focus support on new/young students



Smart but Silent

Some students had:

- High GPA (e.g. 3.9+)
- 0 forum participation
- 0 video completion
- Minimal LMS use

> 📌 **Insight:**
>
> - GPA ≠ engagement
> - Silent students may be succeeding — or struggling silently

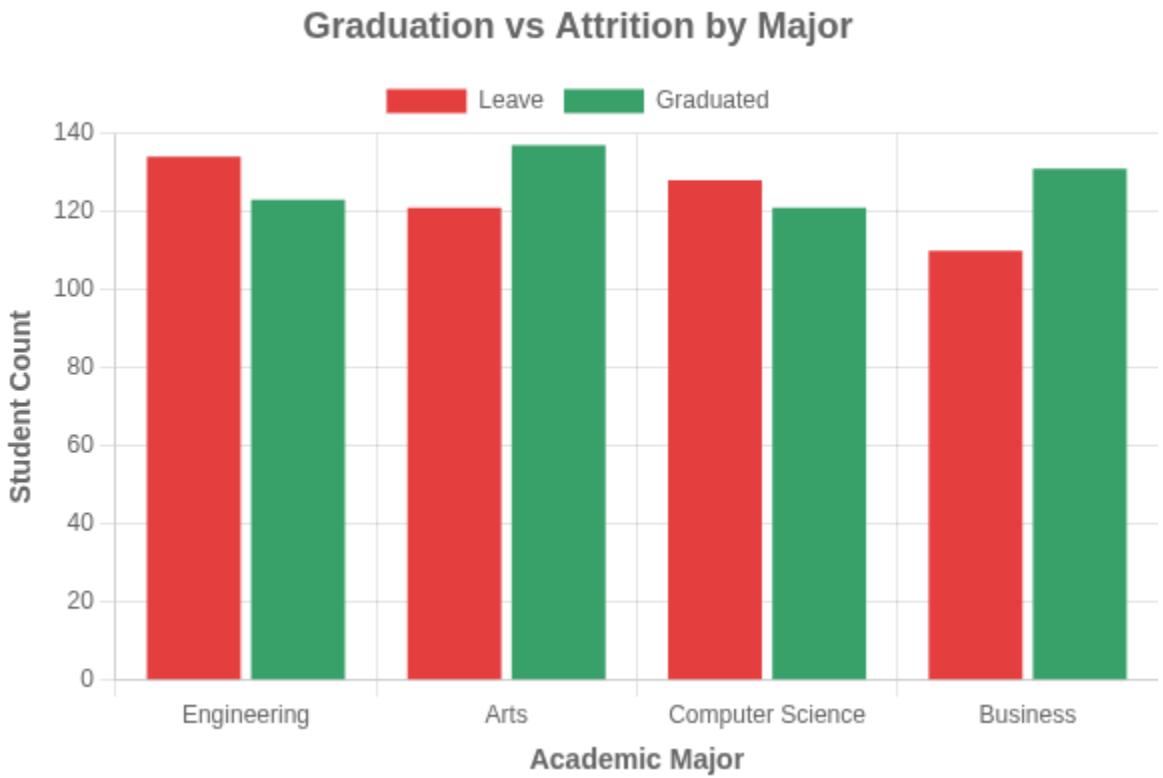
Leave vs Graduate by Major

Major	Leave	Graduated
Engineering	134	123
Arts	121	137
Computer Science	128	121
Business	110	131

Insight:

Engineering & CS have high attrition

These programs may need redesign or more support



Key Takeaways

For Academic Advisors:

- Watch attendance + forum use
- Younger students need more care
- GPA ≠ everything

For Leadership:

- Support high-risk majors (CS, ENG)
- Design early alert systems using behavior
- Quiet students may need outreach

Conclusion

This isn't just about data
it's about using insights to change real lives.

> With the right signals, we can help students **before they fall behind**.

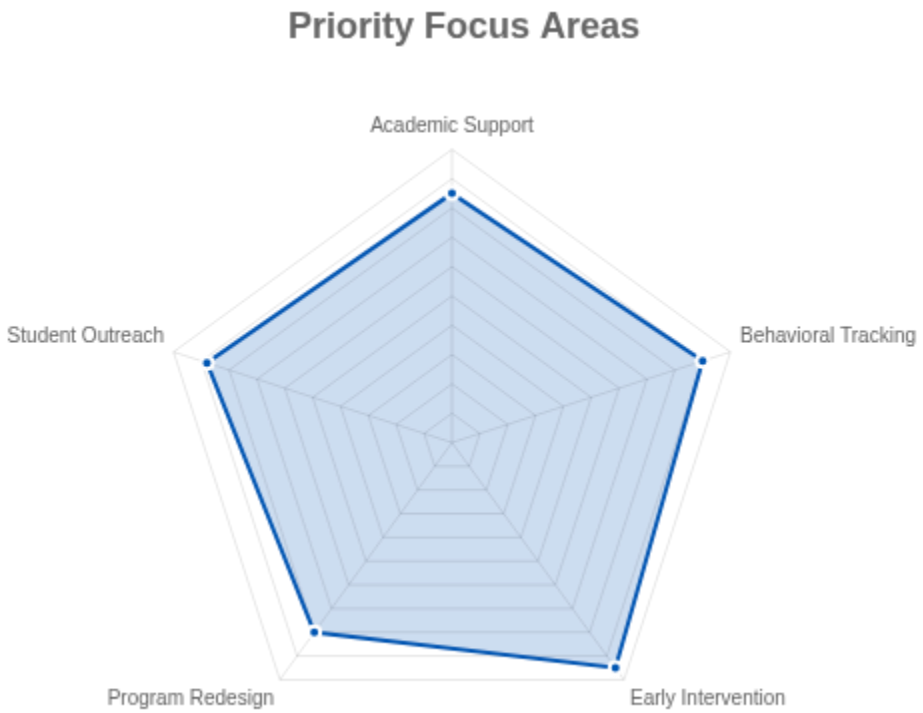
Recommendations

Audience	Recommendation
🎓 Students	Stay active, ask for help early
🧑 Advisors	Use engagement + GPA, not GPA alone
🏛️ Faculty/Admins	Redesign high-risk programs, mentor younger students
🧠 Data Teams	Build dashboards for behavior + academics

💡 Implementation Strategy:

Focus on integrated approach combining behavioral data with academic metrics

Prioritize early intervention systems for at-risk students



Final Thought

> **Academic success is multi-dimensional.** With the right insights, we can support **every student**, not just the ones who ask for help.
>
> *—Nader Mohamed*

SECTION 02

SQL Queries – Technical Deep Dive

The following slides show SQL queries underpinning the insights. Each is used to analyze key aspects of student success from your data warehouse.

</> SQL Q1 – Average Attendance by Risk Level

students.sql

```sql
SELECT
    risk_level,
    AVG(attendance_rate) AS avg_attendance FROM
students
GROUP BY risk_level
ORDER BY CASE
    WHEN risk_level = 'High' THEN 1
    WHEN risk_level = 'Medium' THEN 2 ELSE 3
END;
```

What This Query Does

This SQL query calculates the **average attendance rate** for students grouped by their risk level classification (High, Medium, Low).

It then sorts the results in a custom order, placing high-risk students first, medium-risk second, and low-risk last, allowing for immediate comparison of attendance patterns across risk categories.

📈 Purpose: Shows how attendance varies across risk levels

</> SQL Q2 – GPA by Major

students.sql

```sql
SELECT major,
    AVG(GPA) AS avg_gpa
FROM students
GROUP BY major
ORDER BY avg_gpa ASC;
```

What This Query Does

This SQL query calculates the **average GPA** for each academic major in the student database.

Results are ordered by ascending GPA values, showing which majors have lower average GPAs first. This helps identify which programs might need additional academic support resources.

🎓 Purpose: Identify majors with lower GPA to support students

</> SQL Q3 – Forum Participation & Risk

student_engagement.sql

```
SELECT
    CASE
        WHEN forum_participation_count ELSE < 3 THEN 'Low Participation'
        'Active Participation'
    END AS participation_group,
    risk_level,
    COUNT(*) AS student_count FROM
students
GROUP BY
    CASE
        WHEN forum_participation_count < 3 THEN 'Low Participation' ELSE 'Active
        Participation'
```

What This Query Does

This SQL query categorizes students into two forum participation groups (**Active** vs **Low**) based on their forum activity, then counts students in each participation-risk category combination.

Purpose: Highlights the relationship between engagement and risk

</> SQL Q4 – GPA & Grades by Enrollment

students.sql

```sql
SELECT
    enrollment_status,
    AVG(GPA) AS avg_gpa,
    AVG(attendance_rate) AS avg_attendance
FROM students
GROUP BY enrollment_status;
```

What This Query Does

This SQL query **compares academic performance metrics** (GPA and attendance rates) across different enrollment statuses (Active, Leave, Graduated).

The results reveal slight but meaningful variations in performance patterns between students who have graduated, those on leave, and currently active students, helping identify potential trends or concerns.

🎓 Purpose: Compares GPA and performance across enrollment statuses

</> SQL Q5 – Submission Rate by Risk

students.sql

```sql
SELECT
    risk_level,
    AVG(assignment_submission_rate) AS avg_submission
FROM students
GROUP BY risk_level
ORDER BY risk_level;
```

What This Query Does

This SQL query calculates the **average assignment submission rate** for each risk level category (High, Medium, Low).

Interestingly, the data shows that submission rates don't always correlate directly with risk levels, suggesting that the **quality** of submissions may be more important than quantity in determining student success.

Purpose: Evaluates submission patterns across risk levels

</> SQL Q6 – Engagement for GPA > 3.5

students.sql

```sql
SELECT
    AVG(lms_logins_past_month) AS avg_logins,
    AVG(avg_session_duration_minutes) AS avg_session,
    AVG(video_completion_rate) AS avg_video_rate
FROM students
WHERE GPA > 3.5;
```

What This Query Does

This SQL query analyzes **online engagement patterns** specifically for high-achieving students (GPA > 3.5).

It calculates three key metrics: average LMS logins in the past month, average session duration in minutes, and average video completion rate among top performers, providing insights into how academically successful students engage with online learning resources.

📈 Purpose: See how high-performing students engage online

</> SQL Q7 – Age & Risk Level

students.sql

```
SELECT
    CASE
        WHEN age < 20 THEN 'Below 20' WHEN age'20-22'
        BETWEEN 20 AND 22 THEN
        ELSE 'Above 22'
    END AS age_group,
    risk_level, COUNT(*)
    AS count
FROM students
GROUP BY


    CASE
        WHEN age < 20 THEN 'Below 20'
```

What This Query Does

This SQL query categorizes students into three age groups (**Below 20**, **20-22**, and **Above 22**) and counts how many students in each age group fall into each risk level category.

The results reveal which age demographics are most vulnerable to academic risk, showing that younger students (below 20) have a higher proportion of high-risk cases compared to older students.

Purpose: Identify which age groups face higher academic risk

</> SQL Q8 – High GPA, Low Engagement

students.sql

```sql
SELECT
    student_id,
    GPA,
    lms_logins_past_month,
    assignment_submission_rate
FROM students
WHERE GPA > 3.5
  AND (lms_logins_past_month < 5
       OR assignment_submission_rate < 0.5);
```

What This Query Does

This SQL query identifies **academically successful students** (GPA > 3.5) who show minimal engagement with the learning platform, either through few LMS logins or low assignment submission rates.

These students may be "silent achievers" who succeed without much online interaction, or they could be at risk of future performance decline despite current good grades.

🎓 Purpose: Detect "silent achievers" or hidden strugglers

</> SQL Q9 – Leave vs Graduated by Major

students.sql

```sql
SELECT major, enrollment_status, COUNT(*) AS count
FROM students
WHERE enrollment_status IN ('Leave', 'Graduated')
GROUP BY major, enrollment_status
ORDER BY major, enrollment_status;
```

What This Query Does

This SQL query analyzes **student retention and attrition** by comparing the number of students who have left versus graduated across different majors.

By filtering for only 'Leave' and 'Graduated' enrollment statuses and grouping by major, the query reveals which academic programs have higher dropout rates relative to graduation rates - a key metric for identifying programs that may need intervention.

Purpose: Shows which majors have highest attrition rates

SQL Q9 – High Risk Engagement Metrics

high_risk_engagement.sql

```sql
SELECT
    AVG(lms_logins_past_month) AS avg_logins,
    AVG(video_completion_rate) AS avg_video,
    AVG(assignment_submission_rate) AS avg_submission,
    AVG(forum_participation_count) AS avg_forum
FROM students
WHERE risk_level = 'High';
```

What This Query Does

This SQL query calculates **average engagement metrics** specifically for students classified as high risk. It examines four key behavioral indicators:

- LMS login frequency
- Video content completion rates
- Assignment submission rates
- Forum participation activity

⚠ Purpose: Measures typical online behavior patterns of struggling students