

Data Analysis

Team 2

April 15, 2025

1 Introduction

This section presents an in-depth analysis of the cleaned data set. We observe key variables such as enrollment type, tuition payment behavior, gender, and age range, and aim to identify influential factors that contribute to student academic risk and enrollment decisions. These insights are essential to support data-driven decisions in student retention strategies and institutional planning.

2 Enrollment Trends

We analyze enrollment trends based on different factors such as gender, major, and study mode.

2.1 Gender

- Program 24: Very popular with both genders. Slightly more females (Gender 2) enrolled than males (Gender 1).
- Program 16, 27, 52: Good number of enrollments in both males (Gender 1) and females (Gender 2).
- Program 78, 4: Noticeably more females (Gender 2) enrolled than males (Gender 1).
- Program 54, 40, 33, 43: Strong enrollment of males (Gender 1) but very low enrollment of females (Gender 2).

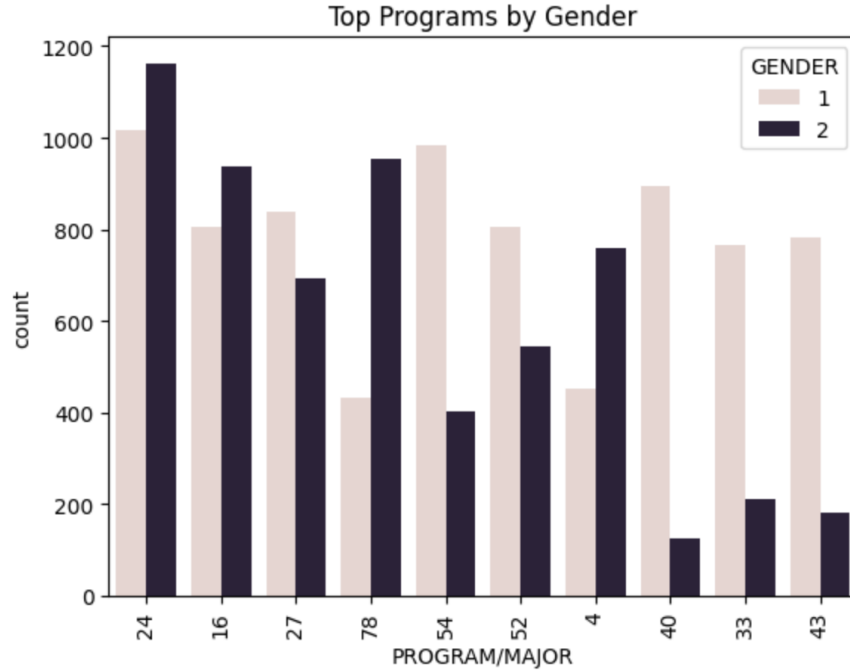


Figure 1: Top Programs vs Gender

2.2 Age Range

- Note:
 - 0 : ≤ 18
 - 1 : 19-20
 - 2 : 21-23
 - 3 : 24-29
 - 4 : ≤ 30
- Program 24: Attracts students from all age ranges. Particularly popular among students aged ≤ 30 .
- Program 16: Popularity keeps decreasing with age.
- Program 27, 54: Attracts students aged 24-29 and especially ≥ 30 . Not at all popular among younger students.
- Program 4, 40, 43: Not at all popular among older students aged ≥ 30 .
- Program 78, 52, 33: Fairly popular among students aged upto 24.

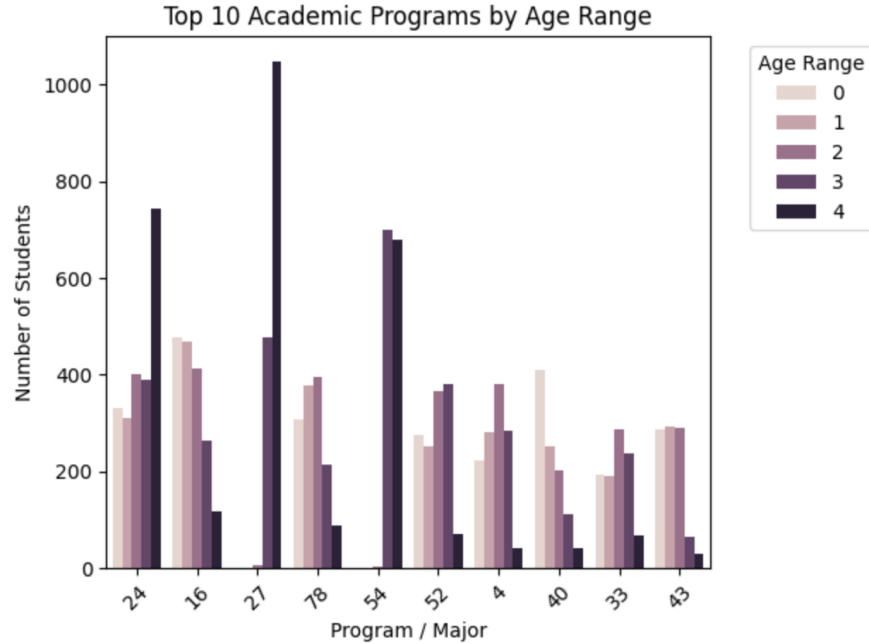


Figure 2: Top Programs vs Age Range

2.3 Study Mode

- Note:
 - 0 : Presencial
 - 1 : Virtual
 - 2 : Remote
- Program 24: Most popular program with presencial students. Least popular with virtual students.
- Program 27, 54: Least popular program with presencial students. Most popular with remote students.
- Program 78, 52, 40, 33: Fairly popular with presencial students. Not at all popular with virtual students.
- **Overall:** Presencial is the most popular study mode and Virtual is the least popular study mode. Programs which are popular with presencial have very low participation from virtual. Program 24 is the most popular program with presencial students. Program 54 is the most popular program with virtual students. Program 27 is the most popular program with remote students.

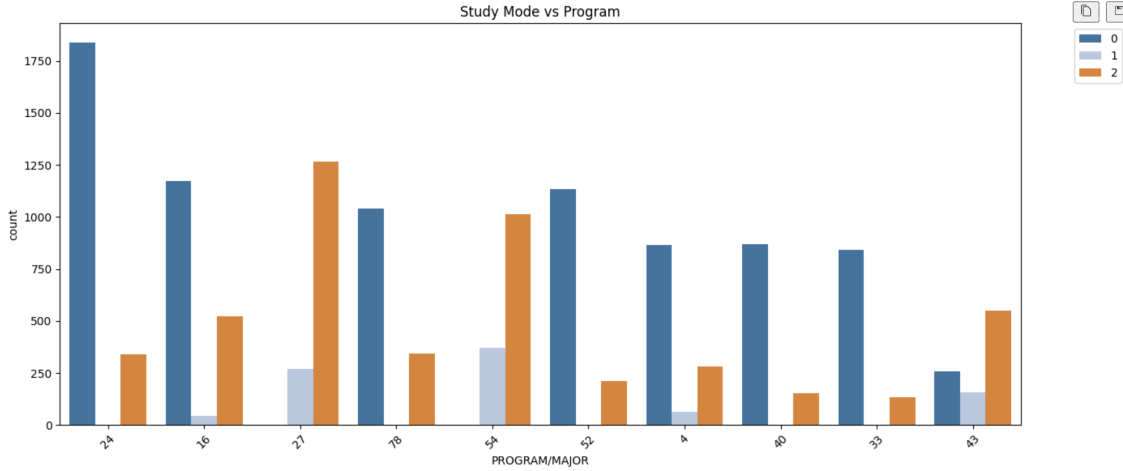


Figure 3: Top Programs vs Study Mode

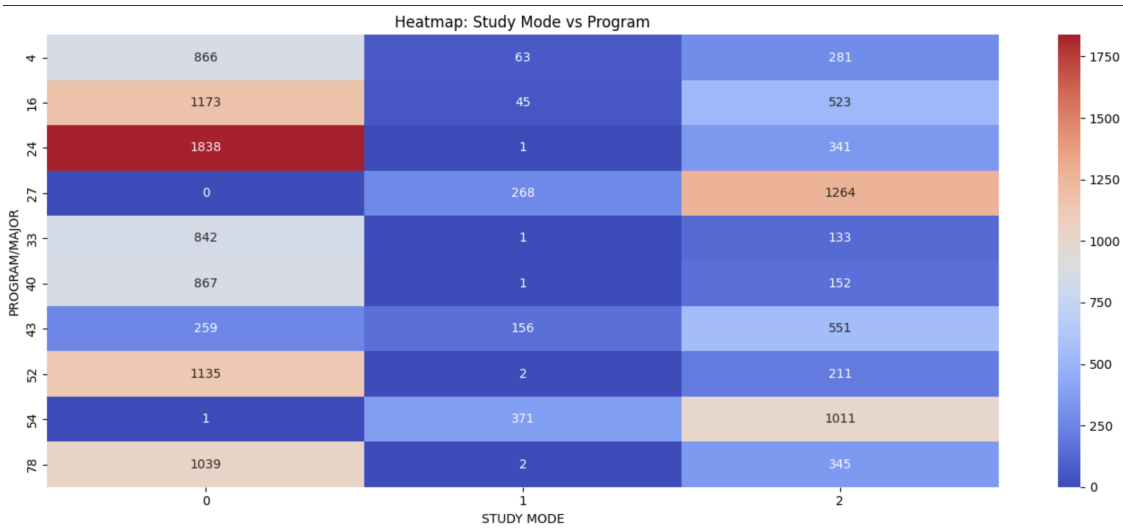


Figure 4: Heatmap: Top Programs vs Study Mode

3 Tuition Payment Behavior

- Negative trend in tuition payment behavior from 2022 to 2023.
- Indicates that the institution's efforts to improve financial literacy or support may not have been effective.
- The increase in the percentage of students not paying their tuition could be due to various factors, including economic conditions, changes in student demographics, or institutional policies.

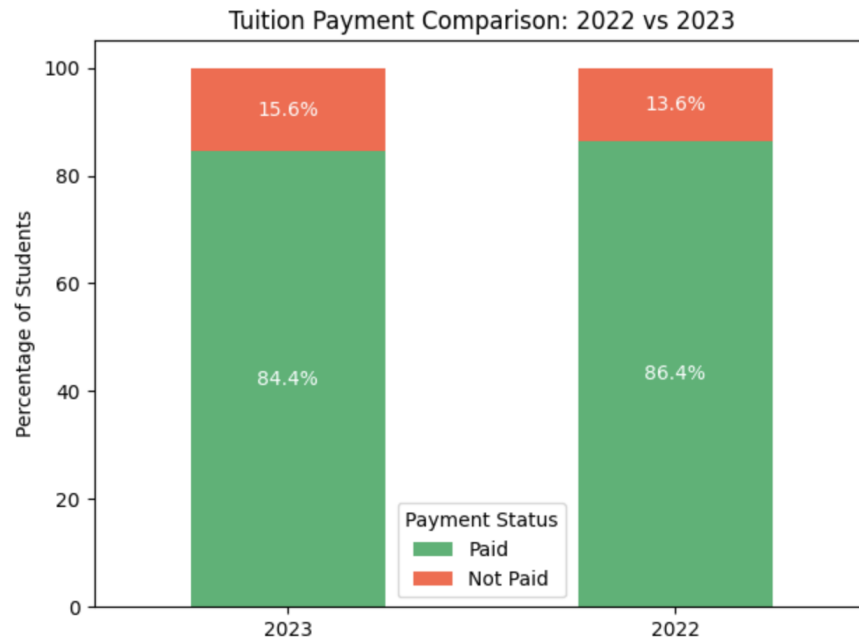


Figure 5: Tuition Payment: 2022 vs 2023

- The majority of students maintained their payment status, suggesting a stable financial situation for most students.
- A small number of students stopped paying their tuition fees, indicating a potential concern for retention and financial stability.

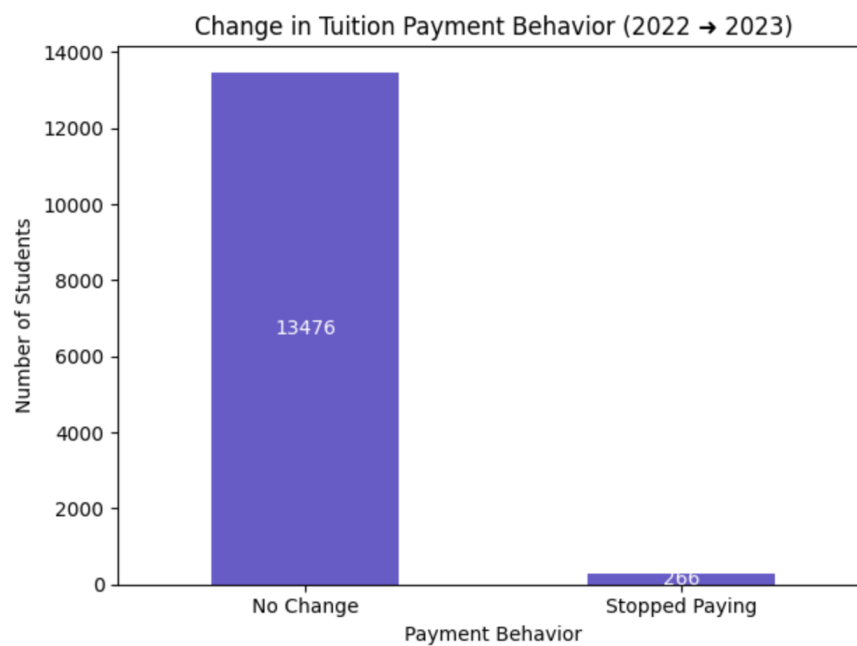


Figure 6: Tuition Payment Behavior

- Female students have a slightly higher percentage of tuition payment compared to male students.
- The difference in payment behavior between genders is minimal, indicating similar financial commitment across genders.
- Program 24 has the highest percentage of students who paid their tuition fees, indicating strong financial commitment.
- Programs 4 and 43 have the lowest percentage of students who paid, suggesting potential financial challenges or disengagement.
- The overall trend indicates that most programs have a higher percentage of students who paid their tuition fees compared to those who did not.

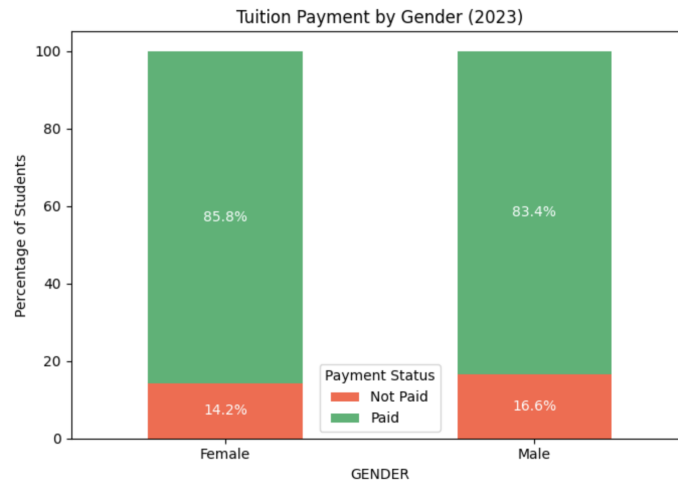


Figure 7: Tuition Payment vs Gender

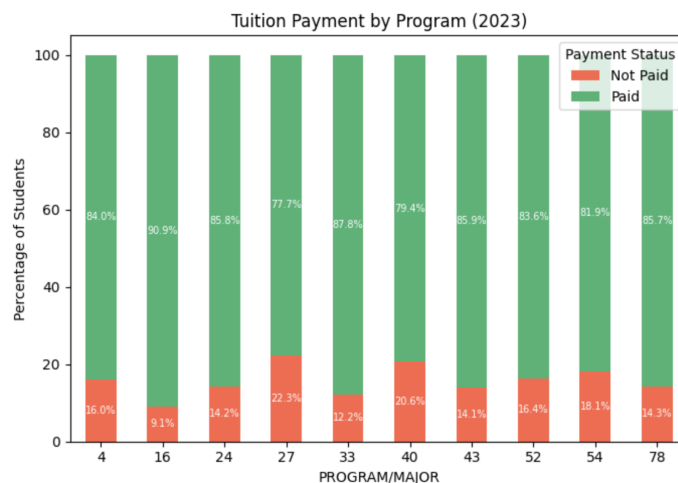


Figure 8: Tuition Payment vs Program

4 Popular Programs and Demographics

4.1 Shift/Schedule vs Program

- Note:
 - 0 : Mixto
 - 1 : Noche
 - 2 : Manana
 - 3 : Tarde
- Program 24: Popular across all shifts/schedules. Most popular program in the "Mixto" shift.
- Program 27, 54: Most popular programs in the "Noche" shifts.
- Program 78, 4: Almost equally popular in "Mixto", "Noche" and "Manana" shifts.
- **Overall:** "Noche" shift has the highest enrollment across most programs. "Tarde" shift has the least enrollment in general by a huge margin.

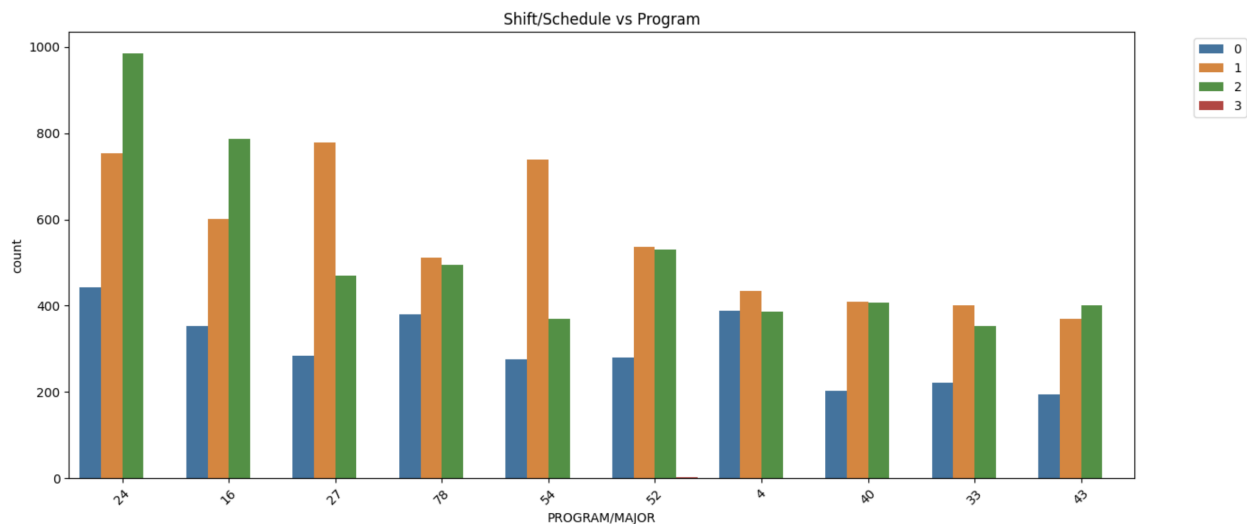


Figure 9: Shift/Schedule vs Program

4.2 Benefit Discounts vs Program

- Program 24: The most popular program across all benefit discount categories. Indicates that this program is accessible and attracts students with varying levels of financial support.
- Program 27, 54: Show significant participation from students availing benefit discounts. Suggests these programs are particularly appealing to students seeking financial aid.

- Program 78, 52: Moderate participation from students with benefit discounts. Indicates a balanced mix of students with and without financial support.
- Program 40, 33: Lower participation from students with benefit discounts compared to other programs. May indicate these programs are less reliant on financial aid for enrollment.
- **Overall:** Programs with higher participation from students availing benefit discounts are likely more affordable or offer better financial aid options. Programs like 24 and 27 demonstrate inclusivity and accessibility for students from diverse financial backgrounds.

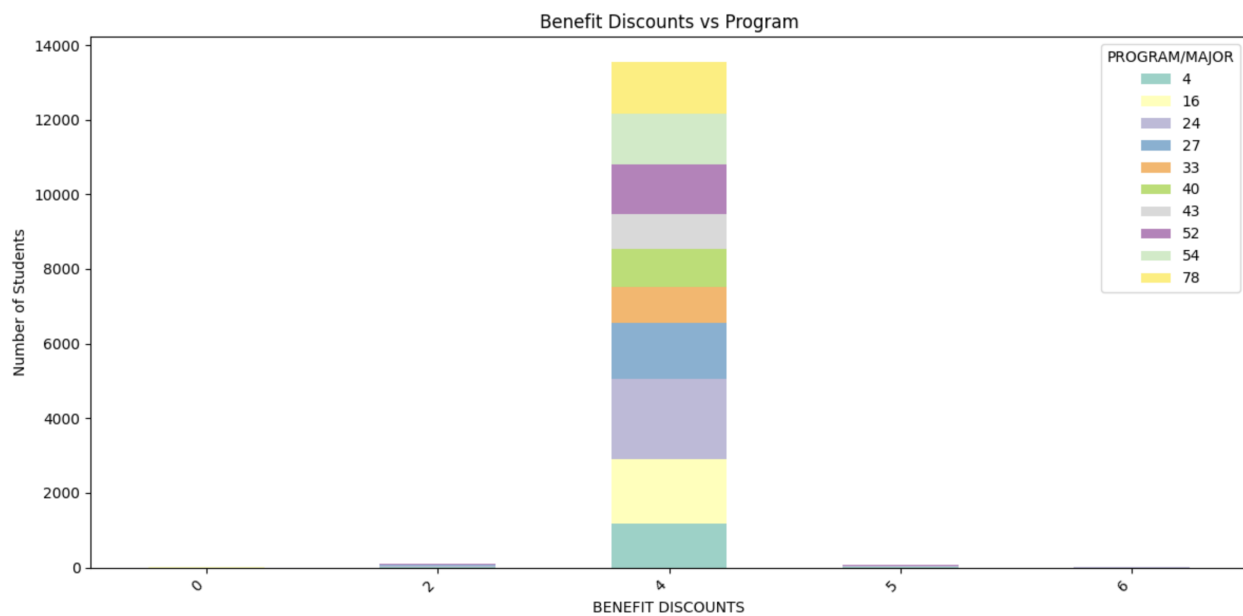


Figure 10: Benefit Discounts vs Program

4.3 Department vs Program

- Program 24: Strong representation across multiple departments. Indicates that this program is widely offered and popular across various departments.
- Program 43, 40, 33: Limited representation in fewer departments. Suggests these programs are niche and offered in select departments.
- **Overall:** Department 14 has the highest number of programs. Department 16 has the lowest number of programs. The distribution of programs across departments reflects the diversity and specialization of academic offerings.

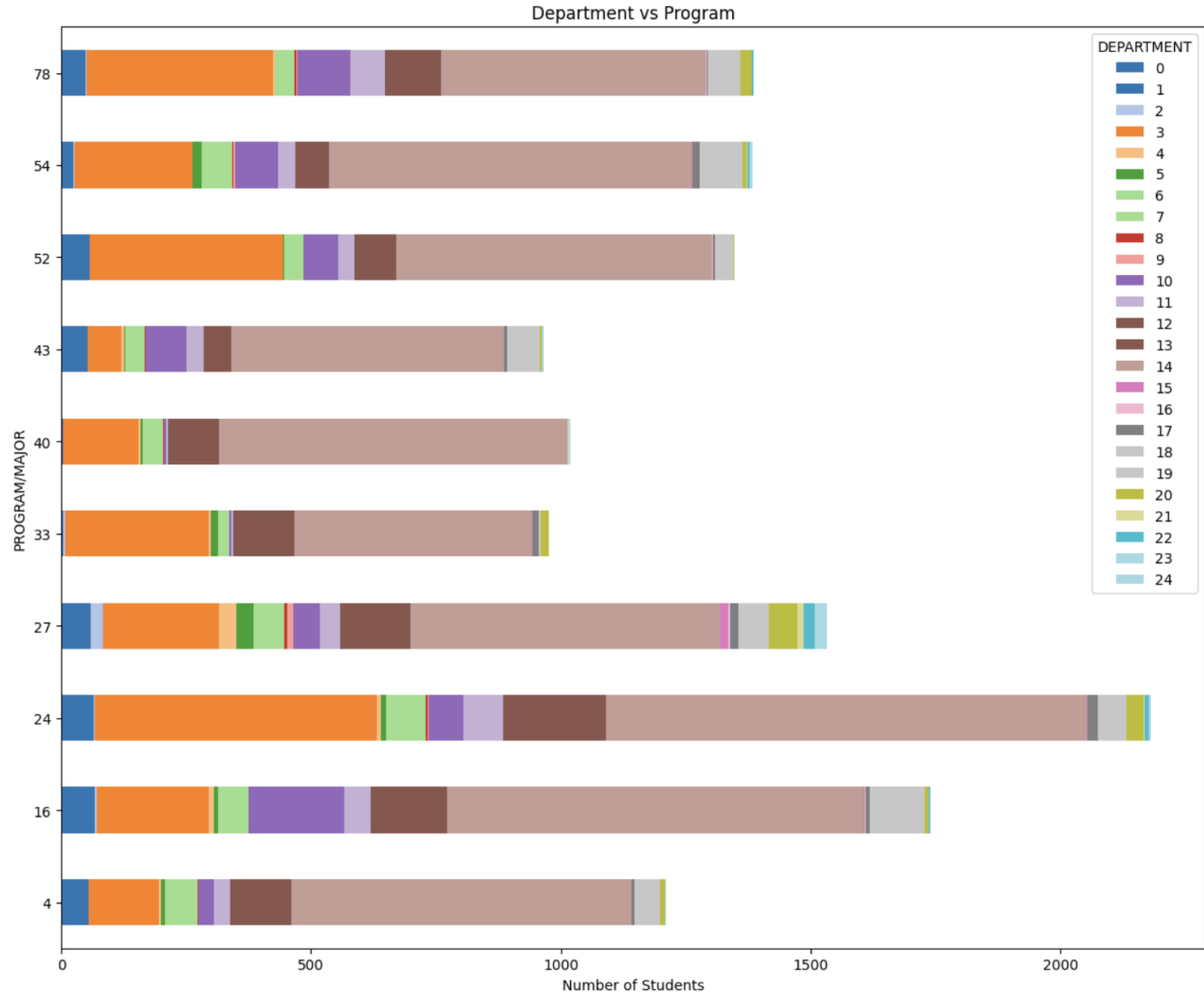


Figure 11: Department vs Program

4.4 Enrollment vs Program

- Program 24: The most popular program across all enrollment types. Indicates that this program is widely accessible and attracts a diverse range of students.
- Program 43: Least popular program across all enrollment types. Suggests that this program may have specific admission criteria or limited appeal.
- **Overall:** The "Reinstated" enrollment type has the highest number of students across most programs. The "New" enrollment type has the lowest number of students across all programs. The distribution reflects the diversity of academic offerings and student preferences.

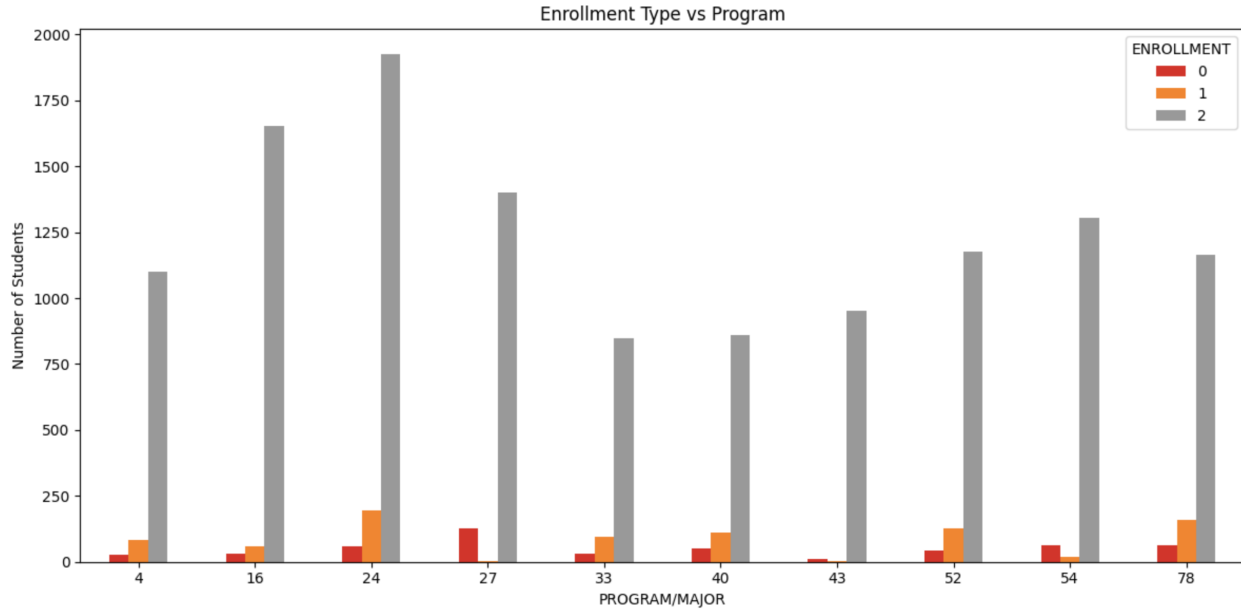


Figure 12: Enrollment vs Program

5 Academic Risk Analysis

5.1 Enrolled Courses vs Academic Risk

- Academic risk level 0 (not at risk): Wide variation in the number of enrolled courses (ranging from 0 to 6). Median is around 2 courses. Indicates high diversity in course load among students not at risk.
- Academic risk levels 1 to 5: As risk increases, the number of enrolled courses becomes more consistent and increases linearly (1 to 5 courses). No variation at each risk level — students with higher risk seem to be taking exactly as many courses as their risk level. Suggests a potential direct mapping between number of courses and risk level from 1 to 5 (e.g., 3 enrolled courses → 3 at-risk courses).

- **Overall:**

→ The violin plot gives similar information as the box plot but also shows the density/distribution of enrolled courses:

- For risk level 0, there are multiple peaks at 1, 2, and 3 courses.
- Suggests several common course loads among students not at risk.
- For risk levels 1–5, the density collapses into a single point at each risk level (e.g., risk level 3 = 3 enrolled courses).

→ Again, implies no variation in course loads for higher-risk students, which might be due to:

- Data preprocessing (e.g., risk level might be equal to number of courses).
- An actual policy (e.g., assigning risk based on the number of courses).

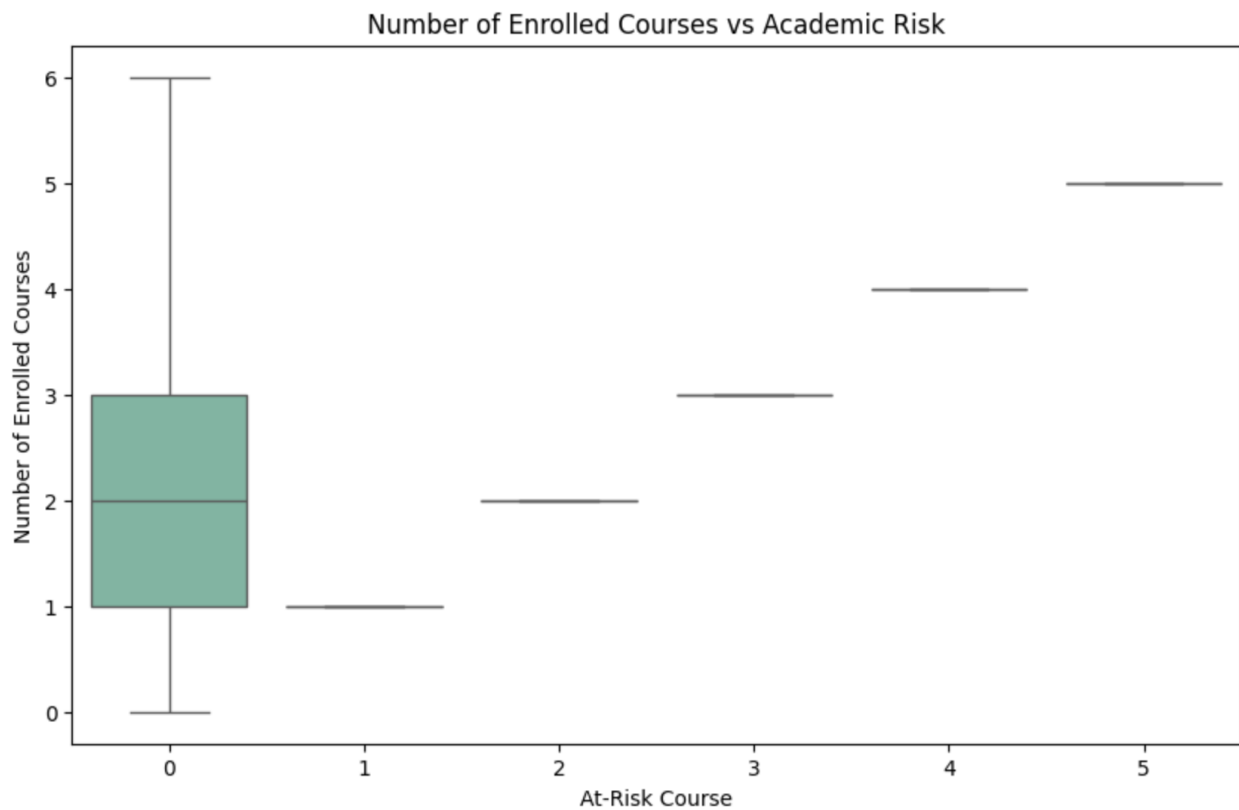


Figure 13: Number of Enrolled Courses vs Academic Risk

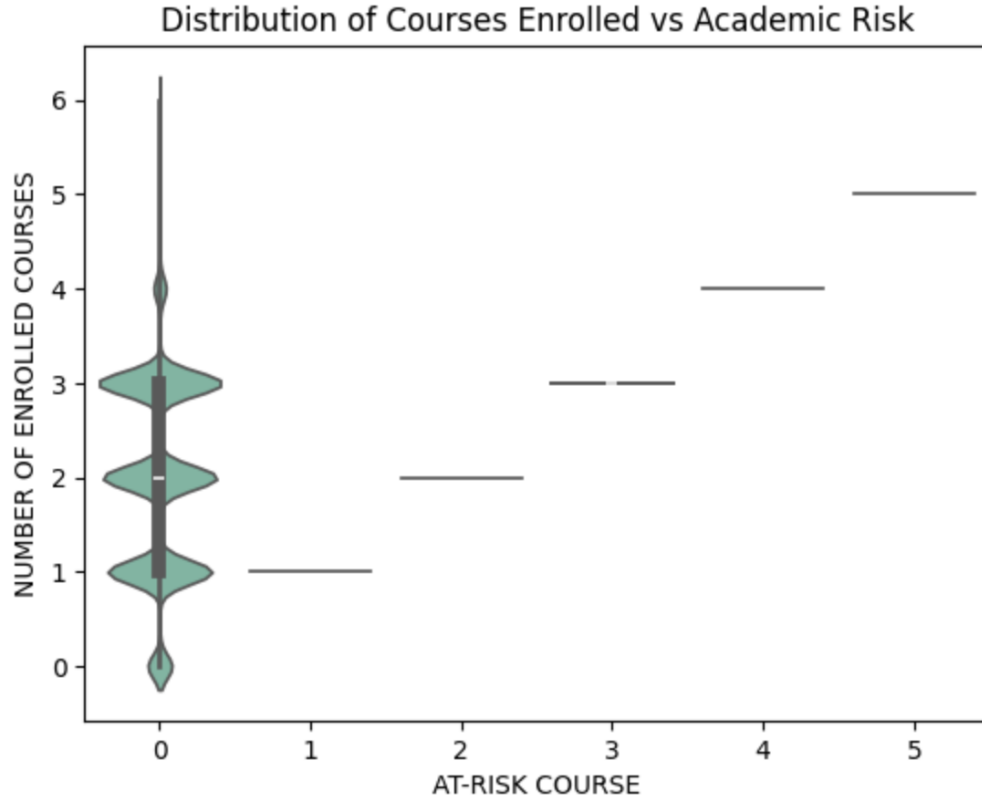


Figure 14: Distribution of Enrolled Courses vs Academic Risk

6 Conclusion

We explored and analyzed student enrollment data from a Peruvian university, highlighting the key trends. The findings show that enrollment patterns are influenced by gender, study mode, age range, while academic risk is closely associated with course load. Overall, this analysis demonstrates the value of data-driven insights in supporting institutional decisions, improving student outcomes, and optimizing academic resource allocation.