Key Factors in Student Retention: A Case Study

Avery Eddins

July 2024

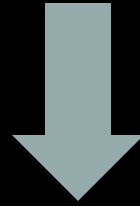# Dataset

- Student retention at Polytechnic Institute of Portalegre in Portugal
- Included information:
  - *Demographic*
  - *Financial (scholarship, current on tuition)*
  - *...*
  - *Retention: enrolled, graduated, or dropout*

Analysis

1. The data is mostly labels (non-numeric)
2. Which factors most strongly influence retention status?

χ^2 test + Cramer's V
(is there a relationship? + how strong is the relationship?)

Influential Factors

Two moderately strong relationships with retention status:

1. Tuition fees up to date
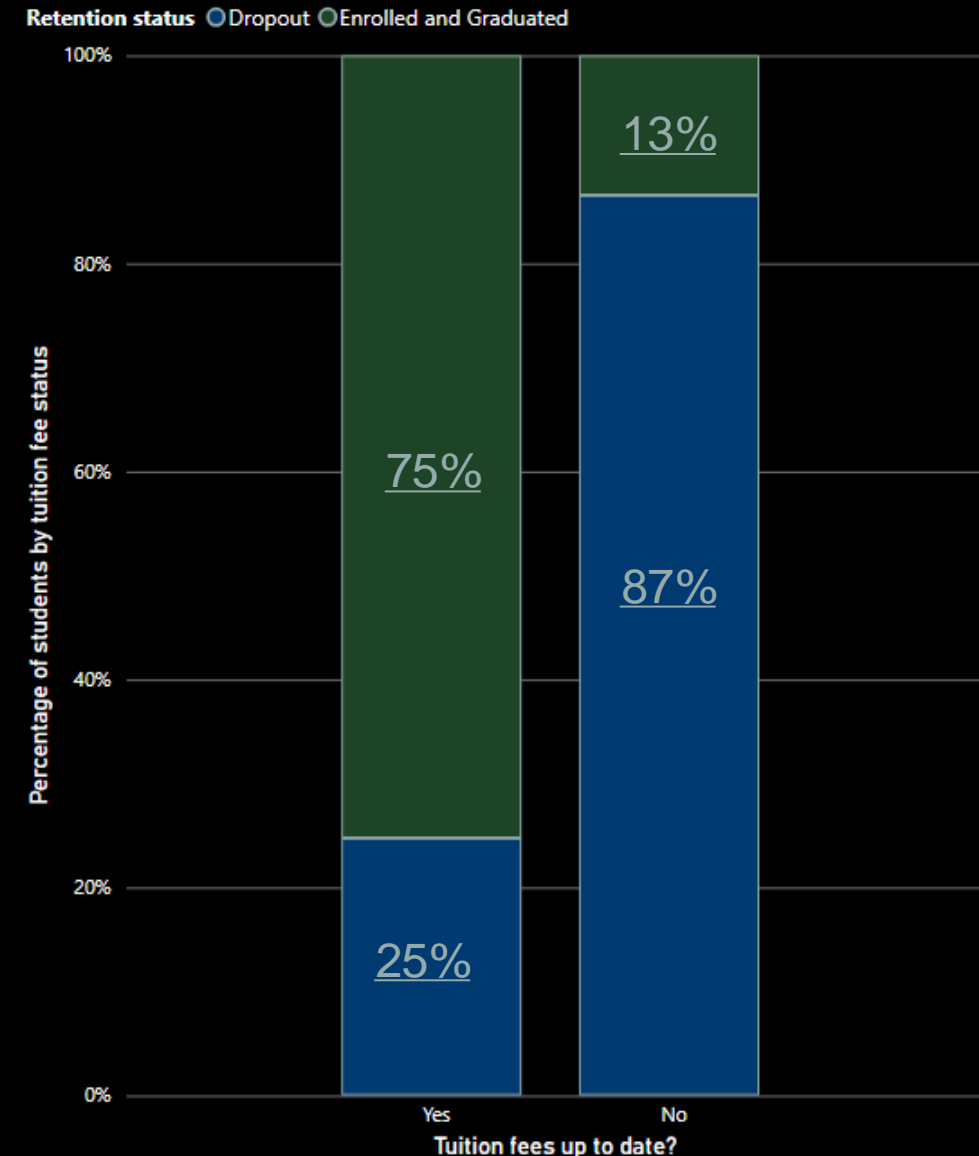2. Scholarship holder

Tuition Fees Up to Date

Of students up to date on tuition:
- 75% are enrolled/graduated students
- 25% are dropouts

Of students not up to date on tuition:
- 87% are dropouts
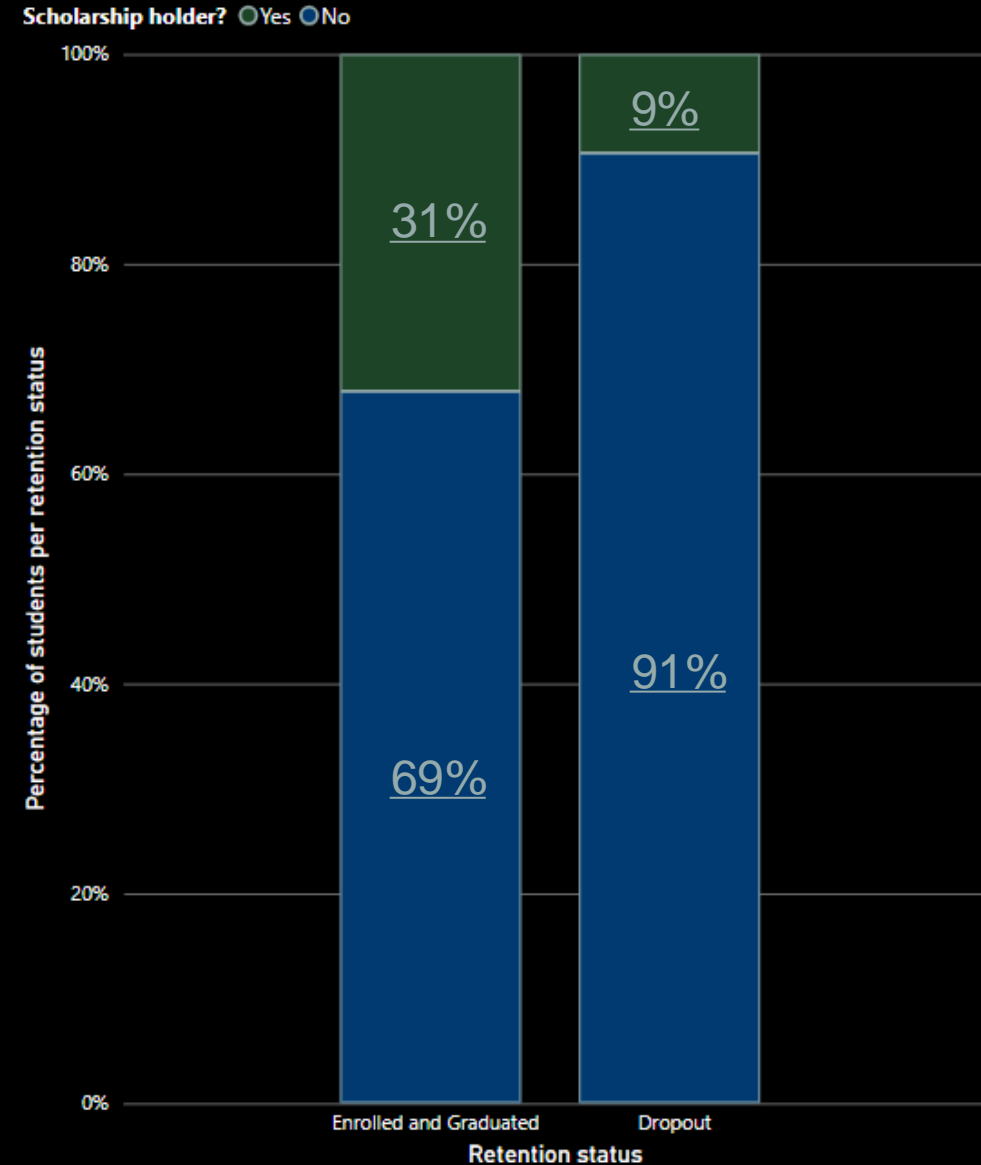- 13% are enrolled/graduated students

Scholarship Holder

Of enrolled/graduated students:
- 69% are not scholarship holders
- 31% are scholarship holders

Of dropouts:
- 91% are not scholarship holders
- 9% are scholarship holders

Interpretation

Correlation, not causation, has been determined

Tuition up to date possibilities:
- Not being up to date on tuition is a precursor to dropping out
- Dropouts fall behind on tuition because they dropped out
- Another factor (e.g. financial hardship) not present in dataset influencing both

Scholarship holder possibilities:
- Not having a scholarship is a predictor of dropping out
- Another factor (e.g. academic performance) not present in dataset influencing both

Analysis limitation

The analysis (χ^2 test + Cramer's V) has requirements:
- All cells in contingency table must be 5 or higher
- If condition not met between two factors, analysis skipped
- Attempted to group categories in certain factors together
  - *Age at enrollment → sort by decade*
- Bias: yes/no factors
  - *Smallest number of bins*
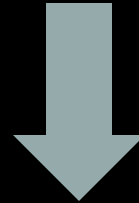 - *Don't have cells with less than 5 counts*

Tuition up to date? contingency table

	Enrolled	Graduated	Dropout
Yes	752	2180	964
No	42	29	457

Analysis limitation

Initial question: Which factors <u>most strongly</u> influence retention status?

5 or more counts

Question can answer: Which factors influence retention status?

Future Work

Limitations of analysis solutions:

- Regroup data points by context
  - *Age of enrollment by decade → 17-22, 23+*
- Machine learning to obtain feature importance scores
  - *Better for dynamic dataset, not this case study*

Recommendations

Analysis has shown that tuition and scholarship statuses affect retention

Actions for universities:
- Consider having broad financial information available to support staff
  - *Privacy/legal concerns?*
- Survey students who drop out
  - *Works towards individualized outreach and support*
  - *Values the students' stories*

Thank you!