



Understanding First-Generation Graduation Rates at The College of Wooster: Applying Mathematical Models to Factors that Affect Retention.

Anna Truong, Advised by Dr. Drew Pasteur
Class of 2023, Mathematics Major

Abstract

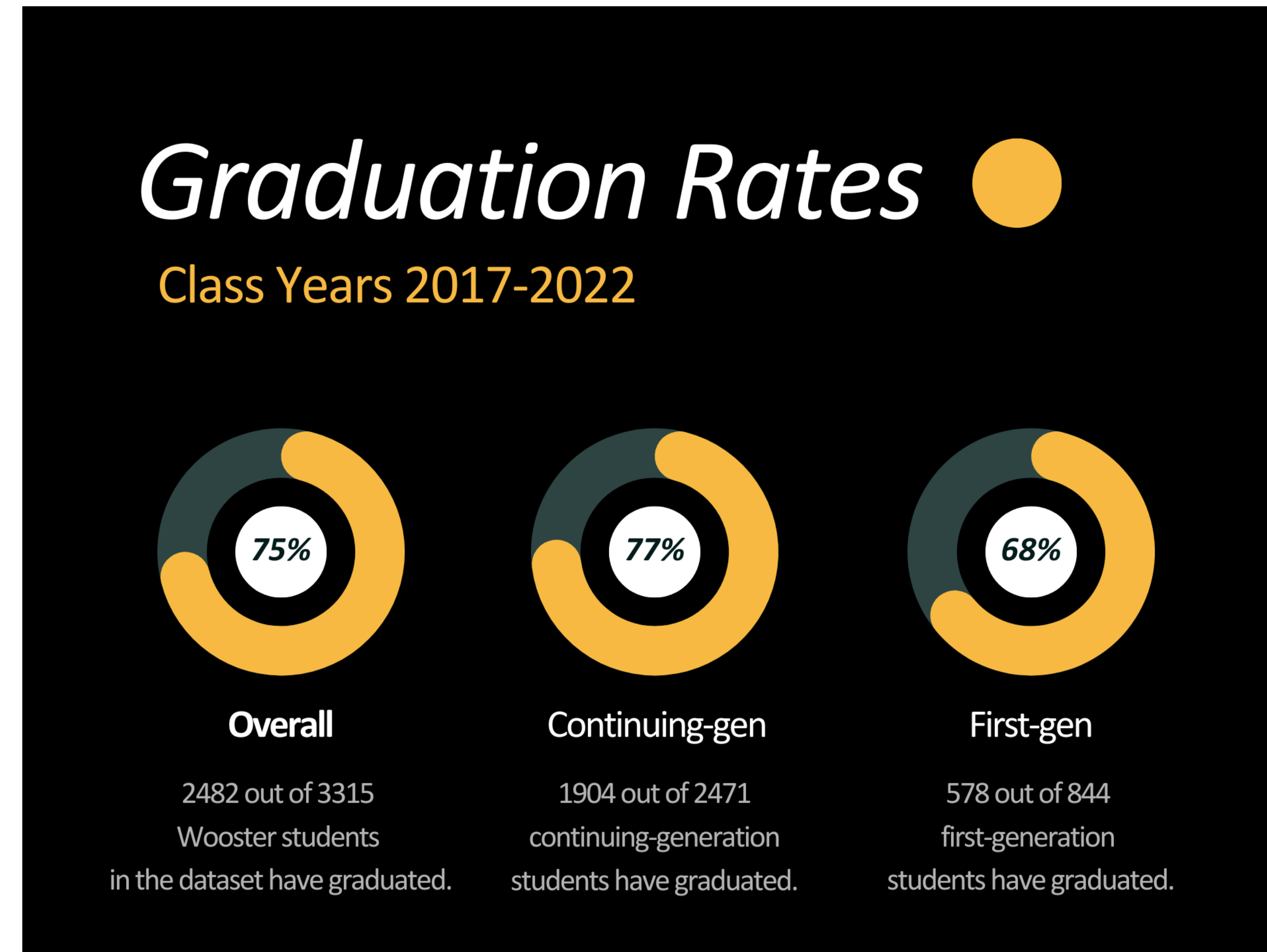
This study investigates the factors that affect retention and graduation rates for students at The College of Wooster, with a focus on first-generation students, using mathematical models like logistic and stepwise regression. First-generation students, defined as students whose parents/guardians have not completed a 4-year college/university, often struggle to graduate within the expected time frame. Due to a lack of parental experience and knowledge of the college environment, students struggle to successfully navigate their own college experience. If not properly addressed, this lack of support can result in attrition. The results demonstrate that first-generation students have lower graduation rates than continuing-generation students with pre-college, college, and demographic factors being significant predictors of graduation. The implications of this study can provide insight into areas of improvement for all students at The College of Wooster.

Methods

- EDA was performed to examine variables and identify outliers/missing values
- Logistic regression investigated the relationships between predictor variable(s) and the binary outcome of graduation.
- Stepwise logistic regression was used to identify important variables for predicting graduation based on their significance level and contribution to the model's overall fit.

Dataset

- Graduating classes of 2017 to 2022
- 3315 total students
- 844 first-generation students (25%)
- Additional Datasets
 - Admissions
 - Academic Resource Center
 - Conduct and Care Team Reports
 - Memberships to Clubs



Results

VARIABLE	First-gen	Continuing-gen
FIRSTGEN	- 0.435***	
STEMMAJOR	+ 3.91***	+ 3.93***
COLLEGE GPA	+ 2.11***	+ 1.99***
FYS GRADE	+ 0.746***	+ 0.777***
GENDER (female)	+ 0.647***	+ 0.588***
FINAID (yes)	+ 0.723**	+ 0.118
MAJORITY (race)	+ 0.0692	+ 0.268**
MINORITY (races)	- 0.0692	- 0.268**
CONDUCT	- 0.204*	+ 0.135*
STEM_INT	- 0.125	+ 0.120
INTERNATIONAL	+ 0.212	- 0.227
ARC	- 0.0103	- 0.0241
CARE TEAM	- 0.123	- 0.107
HSGPA	+ 0.0260	+ small
ACT	+ 0.0115	+ small
SAT	- small	+ small
HSRANK	+ small	- small
GLOMAD	+ 14.8	+ 0.437

Table 1: Bivariate Logistic Regression Results for the GRADUATED Response Variable. Each variable is analyzed separately as a predictor, and the coefficients are reported here. The significance levels are denoted by asterisks as follows: · for $p < 0.1$, * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

Highlights

- First-generation students had a 35.3% lower odds of graduating compared to continuing-generation students (OR = 0.647, 95% CI [0.548, 0.765])
- FYS GRADE, GENDER, and COLLEGE GPA
- STEMMAJOR may be due to survivorship bias
- Variables may vary by generational status:
 - CONDUCT (# of conduct reports)
 - STEM_INT (intending to be a STEM major)
 - FINAID (intending to file for financial aid)

Conclusions

- Increase resources for key groups:
 - First-gen students
 - Lower socioeconomic students
 - Students of color
- Centralized data system to improve information sharing
 - Consistent data collection
- Strengthen FYS (first-year seminar)
 - Building community
 - First-year advising

Limitations

- Method Limitations:
 - Limited knowledge of RStudio and applied statistics
 - Time constraints
- Data Limitations:
 - The accuracy of the models may have been limited by missing information, unused available variables, and low observations.
 - Unaccounted confounding variables.
 - Survivorship bias may have influenced the analysis, as only graduates were included.
 - Improper cleaning of datasets
- Generalizability Limitations:
 - May not be applicable to other populations or settings.