PREDICTIVE ANALYSIS OF ONE-YEAR RETENTION

Jorge Martinez Director of Research & Reporting, Enrollment Services

Caroline Neary Senior Research Analyst, Undergraduate Student Success



SESSION OUTCOMES

- Identify observable student characteristics related to one-year retention
- Create a predictive model used to identify students at risk of not being retained
- Make this predictive data actionable on campus



UH: INSTITUTIONAL OVERVIEW

- Large, 4-year, public, urban university
 - 38k undergraduate students
 - 72% attend full-time
 - 46% first-generation
 - 37% Hispanic
 - 38% receive Pell grant



BACKGROUND

- Retention and Graduation Task Force
- Plateaued retention rates
- Actionable predictive model





STUDY OVERVIEW

Objective:

Review student characteristics that are related to one-year retention and create a predictive model that can be used to identify students at risk of not being retained.

Population:

Full-time, degree-seeking, first-time in college (FTIC) cohorts from Fall 2016, Fall 2017, and Fall 2018 (N=13,927).



STUDY OVERVIEW

COLLECT

> Compiled predictors expected to be related to retention

ANALYZE

➢ Explored models to identify significant predictors

Established a parsimonious model to predict retention

INTERPRET

➤ Calculated probabilities to define students' risk levels

REPORT

Reported findings through HTML report and incorporated them into the advising platform



41 VARIABLES CONSIDERED

Demographic Admissions		Financial	Academic		
High School Rank	Gender	Estimated Family Contribution	Hours Taken		
Transfer Credits	Race/Ethnicity	Unmet Financial Need	Hours Passed		
SAT/ACT Score	Age	Scholarship Recipient	Term GPA		
Application Date	First Generation	Lost Scholarship	Cumulative GPA		
Orientation Date	Commute Distance	Financial Aid Award Amount	Academic Standing		
First Choice College	Residency Status	Total Loans	D/F/W Grades		
	Region of Residence	FAFSA Verification Selection	College		
		Independent FAFSA	Full-time/Part-time		
		Pell Eligibility	STEM Major		
		Financial Delinquency	UHin4		
		Payment Deferment Plan	Honors		
			College Change		
			Math Core Credit		
			English Core Credit		
			First Year Math Level		
			Term Withdrawal		
			CORE 1101		

UNIVERSITY of **HOUSTON**

DESCRIPTIVE 1-YEAR RETENTION RATES: DEMOGRAPHIC





County of residence





DESCRIPTIVE 1-YEAR RETENTION RATES: ADMISSIONS







DESCRIPTIVE 1-YEAR RETENTION RATES: ACADEMIC







DESCRIPTIVE 1-YEAR RETENTION RATES: ACADEMIC

Variable	Min	Max	Std. Dev.	Retained Average	Not Retained Average
First Term Hours Taken	0	27	1.6	15	14.5
Second Term Hours Taken	0	23	2.2	14.9	13.6
First Term GPA	0	4	1.00	3.1	2
Second Term GPA	0	4	1.1	3.1	1.7
First Term DWF	0	8	1.3	0.5	1.9
Total DWF	0	12	2.1	1.2	3.5



DESCRIPTIVE 1-YEAR RETENTION RATES: FINANCIAL

Variable		May	Avorago	Std Dov	Retained Not Retained		
Valiable	IVIIII	IVIAX	Average	Slu. Dev.	Average	Average	
Estimated Family Contribution (EFC)	\$0	\$999,999	\$16,512	\$39,889	\$17,082	\$13,310	
Total financial aid (excluding loans)	\$0	\$53 <i>,</i> 324	\$7,479	\$6,991	\$7,636	\$6,600	
Total amount in loans (all sources)	\$0	\$42,550	\$2,367	\$4,657	\$2,263	\$2,951	
Unmet need	\$0	\$44,728	\$7 <i>,</i> 976	\$7,453	\$7 <i>,</i> 651	\$9,827	



RESEARCH QUESTION

How is retention mediated by the combined effects of each predictor?

Used regression modeling to predict retention at three points in time:

- Start of fall term
- Start of spring term
- End of academic year

Focus on early intervention



COLLECT & ANALYZE

- Data collection
 - Census/official reporting day
 - GD2/FGD
- Data sources
 - UHIR Data Warehouse
 - People Soft
- Dependent variable: Enrolled next fall (0/1)



COLLECT & ANALYZE

- Tool
 - R statistical programming language
 - Libraries
 - tidyverse: data wrangling
 - jtools: analysis and presentation of scientific data
 - caret: classification and regression training
 - ROCR: visualizing classifier performance
 - Rmarkdown: literate programming and presentation



ANALYZE

- Bivariate models
- Conceptual models
 - Exploratory modeling around three themes
 - High school preparedness
 - Financial characteristics
 - UH academic performance
- Feasible Solutions Algorithm
 - rFSA
 - Searches data space for models with user-specified form that are statistically optimal under a measure of model quality



PREDICTIVE MODELS

Predictor	Fall Model	Spring Model	Full Year Model
Race/Ethnicity	х	х	х
First Generation	х	х	х
County of Residence	х	х	х
High School Rank	х	х	х
SAT Score	х	х	х
Test Credits	х	х	х
Transfer Credits	х	х	х
Orientation Month	х	х	х
UHin4 Member	х	х	Х
Fall Hours Enrolled	Х	х	х
Total Loan Amount	х	х	х
Unmet Need Amount	х	х	х
Fall Payment Deferment	х	х	х
Spring Enrollment		х	х
Scholarship Status		х	х
Fall GPA		х	х
Spring Payment Deferment		х	х
Cumulative Hours Taken			Х
Cumulative GPA			Х
College Change			х

UNIVERSITY of **HOUSTON** OFFICE OF THE PROVOST

INTERPRET & REPORT

Model Tables

¥7	T!	Odds	Std.	P- Sig.
variable	Estimate	Ratio	Error	Value level
Intercept	-0.1709	0.8429	0.4057	0.6736
African-American (ref. White)	0.5503	1.7338	0.0980	0.0000 ***
Asian/PI	0.7633	2.1453	0.0864	0.0000 ***
Hispanic	0.1568	1.1698	0.0746	0.0356 *
Other	0.1482	1.1597	0.1295	0.2523
Male (ref. Female)	-0.2022	0.8169	0.0546	0.0002 ***
First-Generation (ref. Not First Gen.)	-0.1050	0.9003	0.0612	0.0863
Generation unknown	-0.1986	0.8199	0.1120	0.0761
Adjacent counties (ref. Harris County)	0.0657	1.0679	0.0688	0.3396
Other Texas counties	-0.4637	0.6290	0.0728	0.0000 ***
Out-of-state	-0.4759	0.6213	0.1535	0.0019 **
International	0.4977	1.6449	0.2707	0.0660
High school rank 0-19% (ref. Top 10%)	-0.7992	0.4497	0.1937	0.0000 ***
20-39%	-0.5914	0.5536	0.1318	0.0000 ***
40-59%	-0.3959	0.6731	0.0998	0.0001 ***
60-79%	-0.3017	0.7396	0.0794	0.0001 ***
80-89%	-0.1830	0.8328	0.0799	0.0220 *
Not ranked	-0.3030	0.7386	0.0957	0.0016 **
SAT score	0.0001	1.0001	0.0026	0.9675
Test credits at entry	0.0244	1.0247	0.0036	0.0000 ***
Transfer credits at entry	0.0106	1.0107	0.0019	0.0000 ***
April orientation (ref. May-June)	0.5861	1.7970	0.1869	0.0017 **
July orientation	-0.2745	0.7600	0.0639	0.0000 ***
August orientation	-0.6772	0.5080	0.0832	0.0000 ***
UHin4 (ref. Not UHin4)	0.1482	1.1597	0.0612	0.0154 *
Fall hours taken	0.1444	1.1553	0.0189	0.0000 ***
Total loans (thousands)	-0.0115	0.9886	0.0051	0.0245 *
Unmet need (thousands)	-0.0161	0.9840	0.0037	0.0000 ***
Deferment fall (ref. No deferment)	-0.5128	0.5988	0.0673	0.0000 ***

Fall Model Logistic Regression Table.

Observations: 12,406 Log-Likelihood: -4,775.782 AIC: 9,609.563 Pseudo-R²: 0.086 AUC: 0.678

* p<0.05; ** p<0.01; *** p<0.001

UNIVERSITY of **HOUSTON** OFFICE OF THE PROVOST

FALL MODEL

When controlling for all other variables in the model:

- African American, Asian/PI, and Hispanic students were all more likely to be retained relative to white students.
- SAT score was not significant
- Test and transfer credit was significant
- Loans, unmet need, and deferred payments were significant

African American (ref. White) Asian/PI Hispanic Other Male (ref. Female) First-generation (ref. Not first gen.) Generation unknown Adjacent counties (ref. Harris County) Other Texas counties Out-of-state International High school rank 0-19% (ref. Top 10%) 20-39% 40-59% 60-79% 80-89% Not ranked SAT score Test credits Transfer credits April orientation (ref. May-June) July orientation August orientation UHin4 (ref. Not UHin4) Fall hours taken Total loans (thousands) Unmet need (thousands) Deferment fall (ref. No deferment)





SPRING MODEL

When fall academic performance and spring enrollment factors were added to the model:

- High school predictors were no longer significant
- Gender, UHin4 participation, and loan amount were also no longer significant
- Part-time spring enrollment was significant







FULL YEAR MODEL

- Cumulative hours was significant
- Loans, unmet need, and payment deferment were no longer significant
- Lost scholarship was significant
- GPA was significant



UNIVERSITY of HOUSTON

PREDICTED PROBABILITY SCORES

We can operationalize the models to identify at-risk students for intervention:

- Calculate predicted probability score for each student
- Stratify students by risk of attrition
 - High risk (up to 5th percentile)
 - Medium risk (above 5th through 30th percentile)
 - Low risk (above 30th percentile)

Thresholds based on EAB's population health management model





MODEL EVALUATION

Model	Pseudo-R2	True Positives	True Negatives	False Positives	False Negatives	AUC
Fall	0.086	85.0%	0.4%	13.9%	0.7%	67.8%
Spring	0.294	84.2%	5.3%	9.0%	1.5%	81.3%
Full-Year	0.348	84.0%	6.3%	8.0%	1.7%	85.1%





CONCLUSIONS

When controlling for the variables in each model, some student groups which were less likely to be retained included:

- Students from outside Harris County and its adjacent counties
- Students enrolled in a payment deferment plan
- Students who attended orientation later
- Students with greater unmet financial need
- Students who lost a scholarship
- Students with lower GPA
- Students who take fewer hours
- Students enrolled part-time or not enrolled in spring



OPERATIONALIZATION

- Created tags in EAB Navigate
 - *Retention: Outreach* (moderate risk)
 - *Retention: Intervention* (high risk)
- Applied tags to first-year cohort in early fall and updated in early spring
- Introduced to advisors and Student Success staff



OPERATIONALIZATION

- Fall 2020 cohort
 Pilot
- Fall 2021 cohort
 - Advising appointment campaign with tagged students in spring 2022
- Fall 2022 FTIC Cohort
 - Advising appointment campaign with tagged students in fall 2022 and spring 2023



EVALUATION

- Accuracy of model with Fall 2020 and 2021 cohorts
- Using Fall 2020 pilot cohort as a comparison group, see if tags increased advising interactions for students tagged as high/moderate risk
- Look for difference in retention outcomes for students tagged as high/moderate risk by their frequency of advising interactions



CONSIDERATIONS

- Incorporate data on student engagement and sense of belonging into the model
- How does this model interact with the EAB Navigate model currently in place?



CONTACT INFORMATION

Jorge Martinez jxm@uh.edu

Caroline Neary csneary@uh.edu

Scan with camera for report



