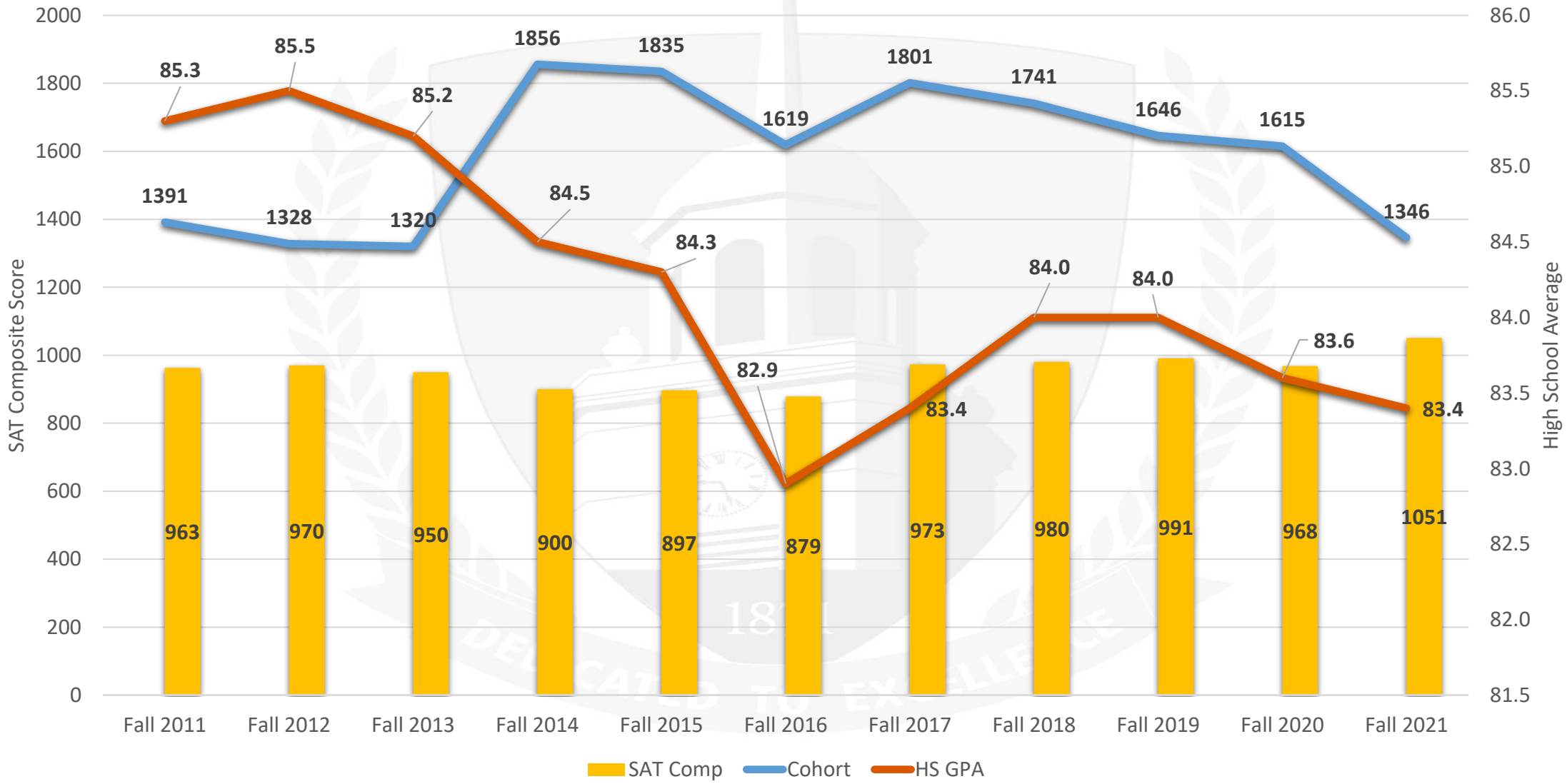
Mining and Modelling for Student Success

12/7/2021

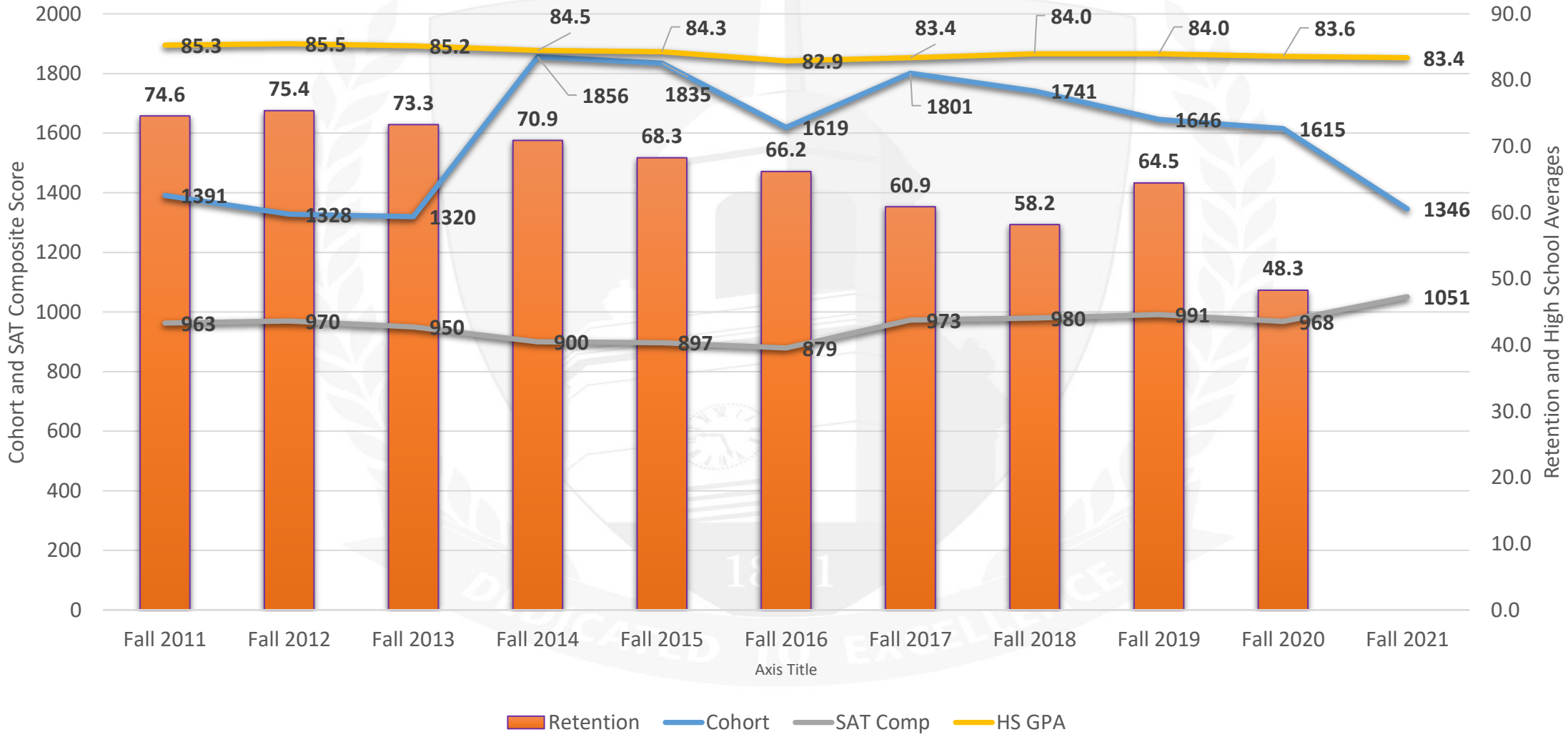
Mission Statement

Buffalo State is a diverse and inclusive college committed to the intellectual, personal, and professional growth of its students, faculty, staff, and alumni. Our mission is to empower students to succeed and to inspire a lifelong passion for learning. Buffalo State is dedicated to excellence in teaching, research, service, scholarship, creative activity, and cultural enrichment.

Trend in Avg SAT Scores and High School Averages for First-time, Full-time Matriculated



Trend in Retention Rates, Avg SAT Scores and High School Averages for First-time, Full-time Matriculated



What is Academic Analytics

Definition

Measurement, collection, analysis and reporting of data about learners and their contexts for the purposes of understanding and optimizing learning and the environments in which it occurs

- 1st Annual Conference on Learning Analytics and Knowledge (2011

Types of Analytics

- Descriptive Analytics
  - Leverage aggregate data and data mining to provide insight into past events
  - “What Happened?”
- Predictive Analytics
  - Utilize statistical models and forecasting to understand potential future events
  - “What is Likely to Happen?”
- Prescriptive Analytics
  - Utilize optimization and simulation algorithms to advise and inform strategies to achieve a desired outcome
  - “What Must We Do to Achieve…”

Analytics @ Buffalo State

Working to build reliable Predictive Analytics model to inform a Prescriptive Analytics approach

Goals:

1. Identify additional transactional metrics available within the SIS (Banner) that are highly predictive of student success/struggle
2. Correlate factors with Intervention strategies to support First-time Student Success and Student Attrition
3. Deliver a 'Risk Score' predicted to have high potential of supporting student success – Students – Advisors – Administrator

Developing Predictive Model

ETL

- Extract – Transform – Load
- SQL, Python, SPSS, Veera Analytics

Queries (Extract)

- **Banner Tables** (Some of the Banner Tables):
  - SFRSTCR (Registration)
  - SPRIDEN, SPBPERS, GORPRAC, GORRACE (Demographic)
  - SHRTGPA (Term), SHRLGPA (Cumulative)
  - SGBSTDN (School, Department, Program)
  - RORSTAT, RCRAPP1,RCRAPP2, RCRAPP3, RCRAPP4, RCRESAR (Financial Aid)
  - SARADAP (Admission Info)
  - SOBYSDS (Frozen file)

Developing Predictive Model

Sample of Admission Fields Included:

- Admitted term,
- Admit desc,
- gender,
- address,
- city,
- zip,
- state,
- county,
- country,
- citizen,
- veteran,
- Income level
- School Applied,
- Degree Applied,
- Major Applied,
- Major Applied,
- Residence,
- Student Status (F/P)
- Department Applied,
- Program Applied

Sample of Financial Fields Included:

Income and Assets

Tax Information

Plan and Educational Background

- Degree type
- Yr in College
- Diploma Type

Demographic

- Residency
- Citizenship
- Marital Status (Parents and Students)
- Dependency
- Parents Education

Developing Predictive Model

Sample of Admission Fields Included:

Academic Profile

- High School Average
- SAT Scores
- ACT SCORE
- Former High School

Developing Predictive Model

Analyzing Data (Load for Persisting after 1 Term)

- **Model specifics – fields included:**
  - Tap Received (+)
  - High School Average (+)
  - Permanent Residence, Unkn (-)
  - Year in USA (-)
  - EFC (+)
  - Financial Aid Tap (+)
  - General Freshman (-)
  - Remedial Courses (+)
  - Housing, Off-Campus (-)

Predicting: Persists				
Variable	Coef	S.E.	Wald chi-sqr	p-value
Intercept	-2.866	0.7371	15.12	0.000101
SOBYSDS_TAP_RECEIVED	0.4060	0.1065	14.52	0.000139
HIGH_SCHOOL_AVG	0.03746	0.00557	45.18	2.0006e-13
Binary(PERMRESD,U)	-0.5320	0.08054	43.64	2.0006e-13
Cube(YEARSUSA)	-0.000112	0.000018	38.25	2.0006e-13
LOGe(EFC)	0.04085	0.01146	12.70	0.000365
FinancialAid_TAP	0.7454	0.1189	39.28	2.0006e-13
Binary(ADMIT_DESC,General Freshman)	-0.5770	0.1234	21.86	2.938e-6
LOGe(REMIDY_COURSES)	1.024	0.2631	15.15	0.000099
EFC	0.000018	4.6091e-6	15.91	0.000067
Binary(HOUSING_1,Off Campus)	-0.5010	0.1339	14.01	0.000182
Binary(MOTHER_HI_GRADE,Middle school)	-0.5303	0.1498	12.54	0.000399
Binary(BANR_YR_IN_COLLEGE,Sophomore)	1.426	0.4638	9.457	0.00210
Binary(DEPARTMENT_APPLIED,AAD)	-0.6422	0.2099	9.359	0.00222
Binary(MAJOR,Elec Eng Tec, Smart Grid)	-1.202	0.4358	7.613	0.00579
Binary(MAJOR,Early Childhood and Childhood)	2.285	1.018	5.035	0.02484
Binary(RACE,IB)	-0.7800	0.3188	5.986	0.01442
Binary(DEPARTMENT,Health, Nutrition & Dietetics)	-0.4145	0.1958	4.482	0.03426
Binary(RACE,BW)	-0.4398	0.2035	4.670	0.03069
SAT_COMPOSIT	0.000675	0.000291	5.382	0.02034
Binary(RESIDENCE_CODE,0)	0.2786	0.1227	5.154	0.02320
LOGe(COUNTY)	-0.2690	0.1263	4.535	0.03320

Developing Predictive Model

Analyzing Data (Load for Persisting after 1 Term)

Logistic Regression Model

- Stepwise regression –Starts with the current model and continues to add the next best variable and will continue to test the remaining variables for Significant contributors until No significant contributors remain. It is at that point the model will stop to add variables

- For this model Step1 Tap Received and at Step 22 Unaccomp_YouthSchool

Candidate Variables	Score Chi-Square
SOBYSDS_TAP_RECIEVED	105.98
FinancialAid_TAP	86.96
HIGH_SCHOOL_AVG	72.50
TAP_Applied	57.97
Binary(PERMRESD,U)	45.35
Cube(YEARSUSA)	32.84
YEARSUSA	31.89
Binary(SIGNED_BY,Applicant and Parent)	21.99
Binary(HOUSING_1,Off Campus)	20.28
LOGe(REMIDY_COURSES)	19.00
LOGe(F_CLASSES)	19.00
SOBYSDS_PELL_RECIEVED	18.65
Binary(PAR_STATE_RES,NY)	17.90
REMIDY_COURSES	16.71
F_CLASSES	16.71
Binary(DEPENDENCY,D)	15.72
Binary(ADMIT_DESC,General Freshman)	15.28
FinancialAid_PELL	15.26
Binary(DEGREE_TYPE,1st bachelors)	14.79
Binary(ADMIT_DESC,EOP Freshman)	14.32
Binary(YR_IN_COLLEGE,Sophomore)	13.89
Binary(BANR_YR_IN_COLLEGE,Sophomore)	13.89

Candidate Variables	Score Chi-Squa
UNACCOMP_YOUTH_SCHOOL	3.337
Binary(SIGNED_BY,Applicant Only)	3.120
Binary(DEGREE_TYPE,Grad/prof)	3.080
Binary(YR_IN_COLLEGE,Freshmen, Prior College)	2.987
Binary(BANR_YR_IN_COLLEGE,Freshmen, Prior College)	2.987
Binary(PAR_TX_RET_FILED,Will file)	2.844
Binary(MAJOR_APPLIED_DESC,Graphic Design)	2.633
Binary(GENDER,F)	2.531
LOGe(PAR_EDUC_CREDITS)	2.529
Binary(MAJOR_APPLIED_DESC,Childhood Education)	2.487
Binary(DEPARTMENT,History and Social Studies Edu)	2.378
Binary(WANTS_FWS,Don't know)	2.373
PAR_EDUC_CREDITS	2.248
Binary(MAJOR_APPLIED_DESC,Mathematics 7-12)	2.126
Binary(MAJOR_APPLIED_DESC,History)	2.114
Binary(DEGREE_TYPE,1st bachelors)	1.777
Binary(DEPARTMENT_APPLIED,HIS)	1.562

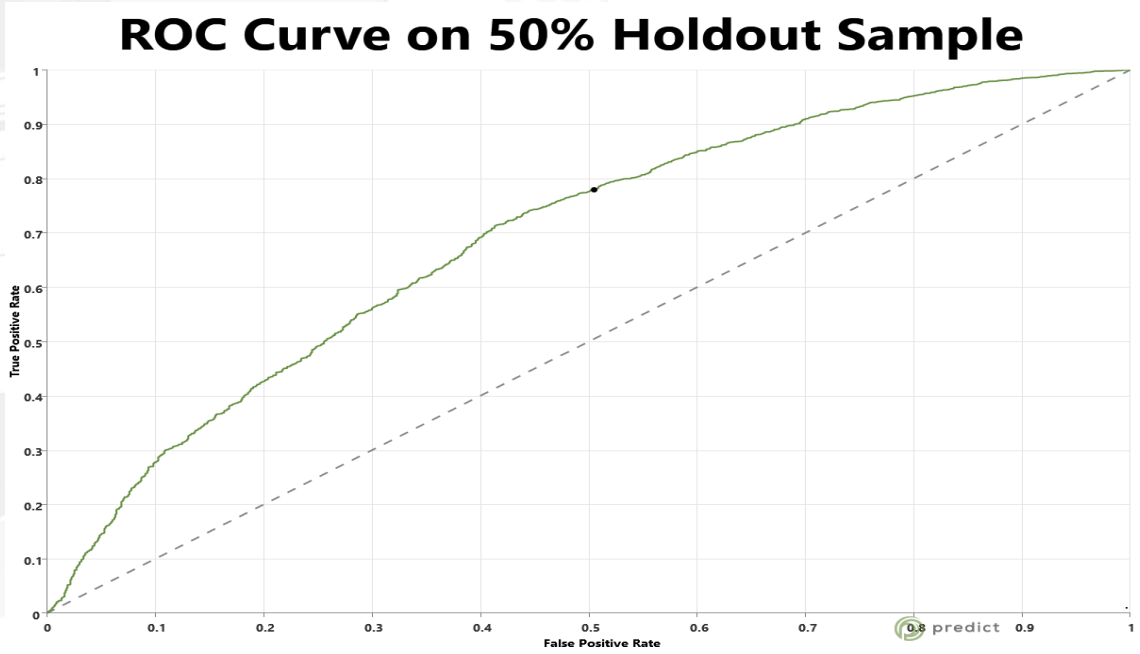
Developing Predictive Model

Analyzing Data (Load for Persisting after 1 Term)

Logistic Regression Model

- Concordance
  - 5.7K Pairs (1 + 0)
  - 69.06% Concordant (Prob. 1 > Prob. 0)
  - 30.37% Discordant (Prob. 1 < Prob. 0)
  - ROC Curve – Receiver Operating Characteristic
  - True Positive vs. False Positive
  - 0.72 Area Under Curve

Percent Concordant	69.06%
Percent Discordant	30.37%
Percent Tied	0.57%
Total # of Pairs	5,797,806



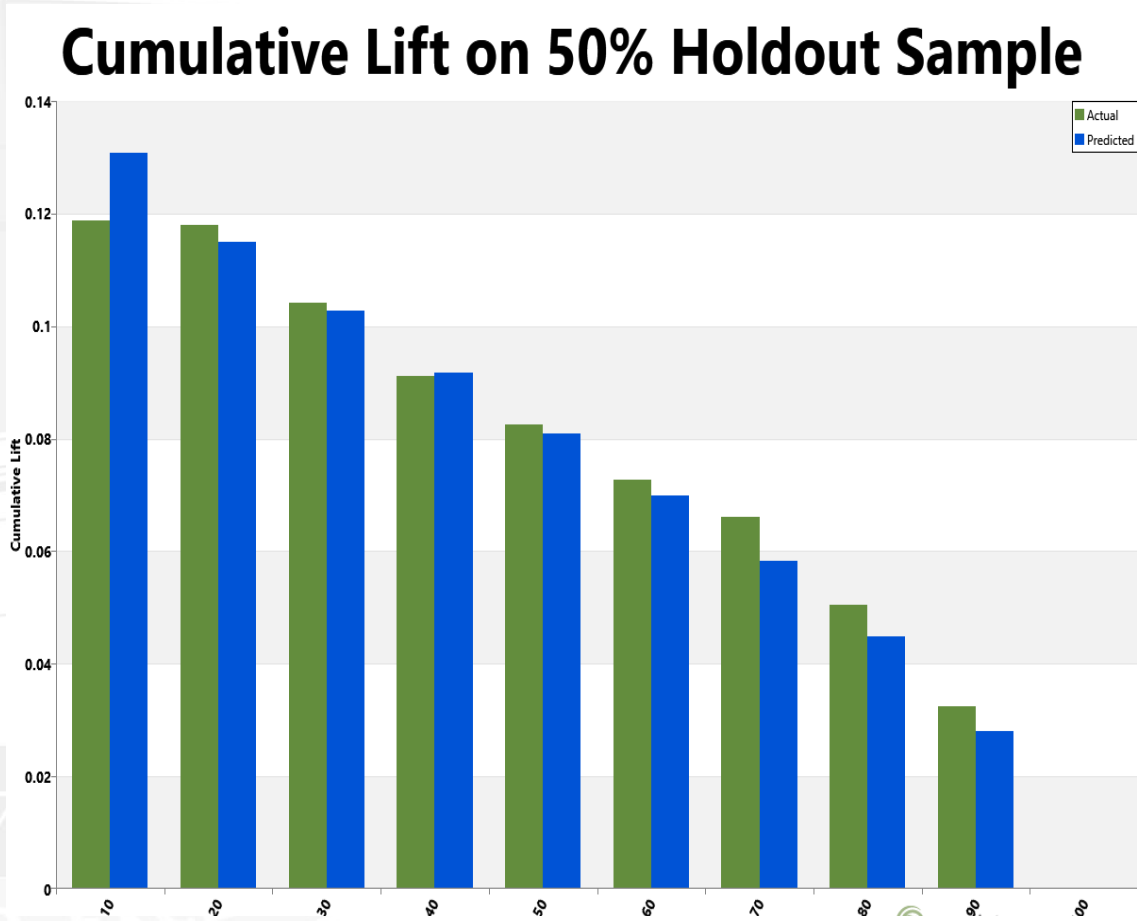
Developing Predictive Model

Analyzing Data (Load for Persisting after 1 Term)

Logistic Regression Model

- Decile Analysis
  - Actual vs. Predicted

- Results indicate potential utility for model



Developing Predictive Model
Model Contribution

Variable Contribution	
Variable	Percentage Model Contribution
FinancialAid_TAP	10.48%
EFC	7.79%
Cube(YEARSUSA)	7.40%
HIGH_SCHOOL_AVG	7.19%
Binary(PERMRESD,U)	6.47%
Binary(MAJOR,Early Childhood and Childhood)	6.46%
SOBYSDS_TAP_RECIEVED	6.13%
Binary(BANR_YR_IN_COLLEGE,Sophomore)	6.12%
Binary(ADMIT_DESC,General Freshman)	6.01%
LOGe(EFC)	5.37%
LOGe(REMIDY_COURSES)	4.32%
Binary(HOUSING_1,Off Campus)	3.59%
Binary(MOTHER_HI_GRADE,Middle school)	3.54%
SAT_COMPOSIT	2.78%
Binary(DEPARTMENT_APPLIED,AAD)	2.78%
Binary(RESIDENCE_CODE,0)	2.71%
LOGe(COUNTY)	2.36%
Binary(MAJOR,Elec Eng Tec, Smart Grid)	2.30%
Binary(RACE,IB)	2.15%
Binary(DEPARTMENT,Health, Nutrition & Dietetics)	2.03%
Binary(RACE,BW)	2.02%

Developing Predictive Model

- Next Steps
  - Continue to Refine this Model
  - Work Simultaneous Model for an Attrition Profile
  - Predicting First Semester GPA, First-Year Retention
- Develop a Scoring Model
  - Apply predictive model to independent data set (Scoring)
  - Single semester of data (Fall 2021)
  - Key: excluded from model data set
  - Used to compare predicted vs. actual result
  - Generate Scoring File
  - Apply model to Scoring File • Generate Predicted Score