ACES (ADMITTED CLASS EVALUATION SERVICE™)

ACES Retention Study for Sample University

ENTERING CLASS OF 2022 – SECOND YEAR

DATE: 2024-09-28

SUBMISSION ID: ABCDEFGH12345678

COLLEGEBOARD.ORG/ACES

Table of Contents

Introduction

This Admitted Class Evaluation Service (ACES) Retention Study is designed to inform your understanding of student retention at Sample University and assist you in predicting student retention outcomes. The student retention outcome you selected to study was 2nd Year Retention. Two prediction methods, logistic regression and decision tree modeling, are used to fit predictive models to your student data. This report includes a number of tables and graphs that describe your student data– both the data uploaded along with College Board data. It presents the predictive model results for all students as well as subgroups, and provides prediction equations that may be applied to future students to understand their likelihood of not returning.

In addition to the Retention studies, ACES makes available Admission Validity studies to examine relationships between College Board exam scores and college performance, and Year-over-Year studies to explore these relationships across entering cohort and year in college. Also available are Placement Validity studies to examine relationships between College Board exam scores and performance in particular courses, and Completion studies to examine relationships between College Board exam scores and student completion outcomes at your institution.

The Retention Study contains several sections:

- *Description of the Study Design for Your Institution* presents the report options selected and the variables to be included in the analyses.
- *Section 1: Background on Retention Modeling* discusses the types of prediction models that will be run and the results that will be presented.
- *Section 2: Descriptive Summary of the Study Measures* presents descriptive statistics (number of valid observations (N), mean, minimum, and maximum) for each numeric variable included in the analysis and frequencies for the retention measure in your student data.
- *Section 3: Evaluating Study Measures* assesses the strength of the relationship between the retention predictors and descriptors included in the study, both individually and in combination, and the student retention outcome measure you selected. These results appear in table and graph form and provide insight into which predictors are likely to be most useful. Information is also shared on the characteristics associated with non-returning students.
- *Section 4: Modeling Retention* applies several prediction models to your student data and presents useful summaries for evaluating the models.
- *Section 5: Using Prediction Models to Identify Students at Risk of Not Returning* includes a reference table that presents the estimated probability that students will return based on logistic regression models presented in Section 4. This section includes plots and tables relating SAT® scores to expected student retention probability.
- *Appendix A: Statistical Summaries for Subgroups* presents several student summaries by subgroups (if you included subgroups). For each subgroup, descriptive statistics, modeling results, and prediction evaluation information are presented.
- *Appendix B: Calculating Predicted Retention Probability* illustrates how to calculate a predicted retention probability for individual students from retention scores (described in Section 5).

A supplementary interactive graph file for this Retention Study can be downloaded from the ACES website. It contains dynamic versions of the tables and graphs in this study that can be viewed, manipulated, and exported using a browser. Instances in which the dynamic version of a table or graph contains more information than the version appearing in this study report are noted in the text.

Description of the study design for Sample University

Your Retention Study includes 2,808 students who entered Sample University in the fall of 2022. Each student's record included at least a retention indicator and SAT® scores. If you included ACT® scores in your submission, they were converted to SAT scores (using the published concordance tables) for those students without SAT scores in the College Board database and used in all retention analyses.

2nd Year Retention served as the criterion for student retention in your study.

ACES provided you with opportunities to customize your retention study to be most informative to your institution. You had the option of selecting which SAT scores to include in your study. You chose to use SAT Total score.

You requested that HS GPA, a measure of high school academic achievement, be included as a predictor. HS GPA was taken from the data file you submitted.

You requested that Cumulative GPA (through last term), a measure of college academic achievement, be included as a predictor.

You requested 3 additional predictors: NO_AP, AT_RISK, and RANKING. Of these predictors, there were 2 two-category (dichotomous (0, 1)) predictors: NO_AP and AT_RISK. The additional predictors will be referred to as "Add. Predictors" in tables and graphs displaying retention predictors, and any two-category (0, 1) predictors will be included in a table describing characteristics of non-returning students in Section 3.

You requested that the following 2 retention descriptor(s) be included in summaries describing characteristics of non-returning students: SAT Total RD with cut-point of 930 on SAT Total score and Cum GPA RD with cut-point of 1.75 on Cumulative GPA.

You requested the following subgroup analyses: SCHOOL.

Further information

- Visit: https://aces.collegeboard.org/
- Call: 800-439-8309
- E-mail: aces-collegeboard@norc.org

The complete statistical output for this report is available upon request by contacting ACES.

The College Board makes every effort to ensure that the information provided in this report and the accompanying data file are accurate. Inaccurate findings may be the result of missing or inaccurate data provided by the institution or discrepancies in matching the institution's data with the College Board database.

Section 1: Background on retention modeling

In this retention study, two modeling methods will be applied to your data in order to predict student retention.

Regression

The first modeling method is a form of regression analysis, called logistic regression, which is appropriate when the outcome measure is a dichotomy (0, 1), as is the case when predicting whether a student is a returning student (1) or non-returning student (0). All retention studies present a base regression model that includes SAT scores as predictors and, if you requested it, HS GPA. If any other predictors are included in your study design (for example, college GPA or additional predictors included in your student dataset), then regression results for a model with all predictors will also be presented. These two regression models will be referred to as the base model and the full model.

Results from each regression model include measures of how well the model fits your student data, the relative importance of each predictor, and weights that can be used to calculate the predicted probability of student retention for applicants and students early on in their college career, along with graphs that illustrate the relationship between key predictors and student retention. When reviewing model fit measures, keep in mind that all students with the retention outcome and predictor values are used both to estimate the model and to assess fit and that there may be differences in model performance when applied to new cohorts of students.

Types of predictor variables

This student retention study supports two types of predictor variables in regression analysis. Predictors can be on a numeric scale, for example, high school GPA or SAT Reading and Writing (RW) Section score. In addition, two-category predictor variables coded as 0 or 1 are supported. An uploaded predictor variable such as whether a student is in their family's first generation to attend college or not, coded as 1 or 0, is an example of a two-category predictor. Such predictors can be interpreted in the context of a regression analysis as how having the characteristic relates to the outcome measure.

Retention cut-off value and predicted retention probability

One consideration in building models to predict relatively rare events, such as students not returning to their institution, is that overall model accuracy can be very high by simply predicting that the more popular outcome will always occur. In other words, predicting that all students will be returning students at an institution with an annual 93 percent retention rate, will be 93 percent accurate overall. However, such a prediction model would be in error for all non-returning students and having the ability to identify such students is important to institutions.

Regression models developed to predict student retention will produce a predicted probability of retention for each student in the analysis dataset, along with prediction equations that can be applied to students in the future. Depending on model performance, the predicted probability of a student returning can be used to identify students at risk of not returning. The predicted probability of retention has a possible range of 0 through 1 and students are classified into two groups: predicted returning students and predicted non-returning students. They are classified into these groups using the cut-off probability value associated with the best model fit based on a measure often used to assess model performance in predictive classification tasks.

The model fit measure used is F1, which is a weighted average of the recall and precision of model predictions at a given predicted probability cut-point. Of those students who actually return (true positive and false negative predictions), recall records the percentage predicted to return (true positives). Of those students predicted to return (true positive and false positive predictions), precision records the percentage who actually return (true positives). The F1 measure thus incorporates both types of prediction errors: failure to correctly predict a returning student and failure to correctly predict a non-returning student. The predicted probability cut-off associated with the highest F1 value, is used when predicting students to return or not return based on the model.

This valuation, which takes into account the error of incorrectly predicting a non-returning student to return, is in line with greater institutional concern with students who are at risk of not returning. The predicted probability of retention cut-off applied will vary slightly from institution to institution and for different regression models. Similarly, the predicted probability of retention cut-off value may differ between regression and decision tree models; the latter models are discussed in the next section. The predicted probability cut-point off value will appear in the regression results.

This predicted probability of retention cut-off value will be used to classify students as returning students or non-returning students in summaries that involve predicted retention outcomes, such as the percent correctly classified and the counts and percentages in the regression model success tables..

Decision trees

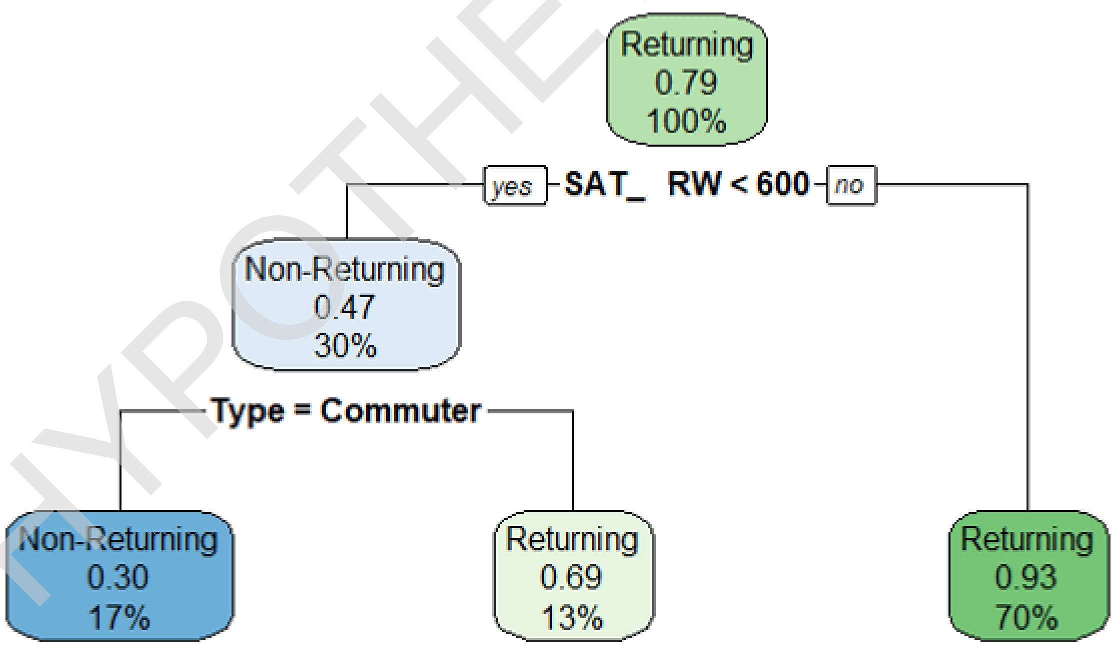
Decision tree modeling will also be applied to your data. It is a data mining method that operates by successively splitting the student dataset into smaller groups that tend to concentrate students with the same retention outcome (returning, non-returning) into the same group. Beginning with all students, the decision tree methodology will examine each predictor as it relates to the retention outcome for all students, then select the strongest predictor and use it to split the student data into two smaller groups, each of which is more uniform on the retention outcome. The process then repeats: for each of the two groups formed, the strongest predictor for that group is used to further split the students into two additional groups. This splitting process continues until a stopping criterion is reached (based on small group size, or no remaining or useful predictors). For a numeric predictor, such as the SAT RW score, the split is based on an optimal cut-point value, for example, students at or above 600 and students below 600. A two-category predictor would create a split based on its two categories; an example would be students living on campus or off campus. A numeric predictor, like an SAT score or college GPA, can be used at multiple levels within a decision tree: for example, first splitting into groups that separate high from low scores, then later splitting the high group into two additional groups, one with very high scores.

A decision tree provides a visual representation of your student data in terms of predictors that best separate returning from non-returning students. And the groups formed by a decision tree can be easily described since they are based on either cut-points of numeric predictors (e.g., students with SAT RW scores over 530) or two-category predictor values (e.g., first-generation students). Naturally, decision trees may capture interactive effects since all available predictors are considered when splitting a group. Thus, if student residence on or off campus affects retention for first-generation students but not for others, the decision tree would reflect this.

In decision tree terminology, the primary node splits the data into additional groups. Additional nodes are created until a final split, the terminal node is created. The terminal nodes are of considerable interest since they define the groups used to produce decision tree predictions. They describe the characteristics of students who are more likely to return and students who are less likely to return. When examining decision trees, the size of the nodes should be taken into account since it informs whether a particular pattern applies to a small or large number of your students.

To illustrate, a simple decision tree is presented below.

Student Retention Decision Tree Example



Each node of the decision tree displays three pieces of information. The topmost item is the predicted outcome for students in that node, which is the most frequent outcome. The second item is the proportion of students with the outcome of interest (here, returning to the institution). And the last item in the node presents the percentage of all students in the analysis who are present in the node.

The root (top) node information indicates that before any splits are performed, the predicted outcome is (unsurprisingly) that students will return and that 0.79 (or 79 percent) of students in the analysis returned to the institution. Because the root node contains all students in the analysis, the percentage of students in the node is 100 percent.

In this decision tree, students are first split by SAT RW Section score: those with an SAT RW Section score below 600 are predicted to be non-returning students, while those who score 600 or higher are predicted to be returning students. Of the students with higher SAT RW scores, 93 percent are returning students, and this group constitutes 70 percent of all students in the analysis. Thirty percent of students have lower SAT RW scores, and only 0.47 (47 percent) of this group are returning students.

Students with SAT RW Section scores below 600 are further split into two groups based on whether they are residential or commuting students. There are no further splits in the decision tree in this example, so the decision tree contains two levels based on two splits, and there are three terminal nodes. Students with lower SAT RW scores and who are commuters are unlikely to return–only 30 percent (0.30) of students in this group return to the institution. In contrast, students with lower SAT RW scores and who are in residence are likely to return (69 percent).

For decision tree models a predicted probability of retention cut-off is selected based on the best F1 model performance value and is used when predicting students to return or not return. As is the case for regression models, this cut-off is used in constructing the decision tree percent correctly classified statistic and when producing the counts and percentages in the model success table. One consideration specific to decision trees, especially small trees with few terminal nodes, is that there may be a range of predicted probability cut-off values that produce the same predicted classification results and the same F1 measure. For example, in the decision tree illustrated above, 69% of students with an SAT RW score below 600 and who are in residence return while 93% of students with an SAT RW score at or above 600 return. Because of this gap between 69% and 93%, any probability cut-off value between 70% and 92% would make the same retention predictions and have the same F1 value. This contrasts with regression model results using SAT scores and GPA predictors, which produce near continuous predicted probability values across students. Because of this aspect of small decision trees, the predicted probability cut-off selected for a decision tree model should not be considered a unique value; generally any cut-off in a range of values is equivalent and this should be recognized. Instead, the decision tree model paths can be helpful in understanding your students and identifying combinations of characteristics that may be associated with students not returning to your institution.

When viewing model results, it should be kept in mind that decision trees and regression are different types of predictive models. Depending on the patterns in your student data, you may prefer one model over the other. Model predictions and predicted cut-off probability may vary across cohorts. A decision tree analysis will be attempted using all retention predictors (full model). Depending on the number of predictors, the strength of your predictors as they relate to student retention, the number of non-returning students (which can be small), and the volume of student data, a decision tree can be simple with just a few nodes or elaborate with many nodes. If the decision tree algorithm finds no worthwhile splits after examining all predictors as they relate to student retention, then a message will appear stating that a decision tree model was not applicable to the data.

Section 2: Descriptive summary of the study measures

This section presents a descriptive summary of the measures in your study.

For students included in this study–those with SAT scores (and ACT scores concorded to SAT scores) and a retention measure–the first table presents the basic counts and percentages of student retention at your institution.

Student 2nd Year Retention

Student Outcome	N (%)
Non-returning students	245 (9%)
Returning students	2,563 (91%)

The table below displays the mean, standard deviation (SD), minimum, and maximum of each individual measure selected for your study, and the number of students (N) with information available on each measure. Some measures may be available for all or nearly all of your students. Others may only be available for smaller groups of students. The table presents all measures with information available on 5 or more students.

Statistical Summaries of Study Measures for all Students

Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	2,808	3.77 (0.38)	2.30	4.30
SAT Test Score	SAT Total score	2,808	1183 (138)	610	1600
Add. Predictor	NO_AP	2,808	0.16 (0.37)	0.00	1.00
Add. Predictor	AT_RISK	2,808	0.05 (0.22)	0.00	1.00
Add. Predictor	RANKING	2,808	3.00 (1.41)	1.00	5.00
College GPA	Cumulative GPA	2,808	3.24 (0.56)	1.61	4.00

The next two tables present summaries of the study measures for returning students and for non-returning students.

Statistical Summaries of Study Measures for Returning Students

Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	2,563	3.78 (0.38)	2.30	4.30
SAT Test Score	SAT Total score	2,563	1192 (136)	610	1600
Add. Predictor	NO_AP	2,563	0.15 (0.36)	0.00	1.00
Add. Predictor	AT_RISK	2,563	0.04 (0.20)	0.00	1.00
Add. Predictor	RANKING	2,563	2.98 (1.41)	1.00	5.00
College GPA	Cumulative GPA	2,563	3.27 (0.54)	1.61	4.00

Statistical Summaries of Study Measures for Non-Returning Students

Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	245	3.66 (0.38)	2.70	4.30
SAT Test Score	SAT Total score	245	1081 (117)	750	1430
Add. Predictor	NO_AP	245	0.23 (0.42)	0.00	1.00

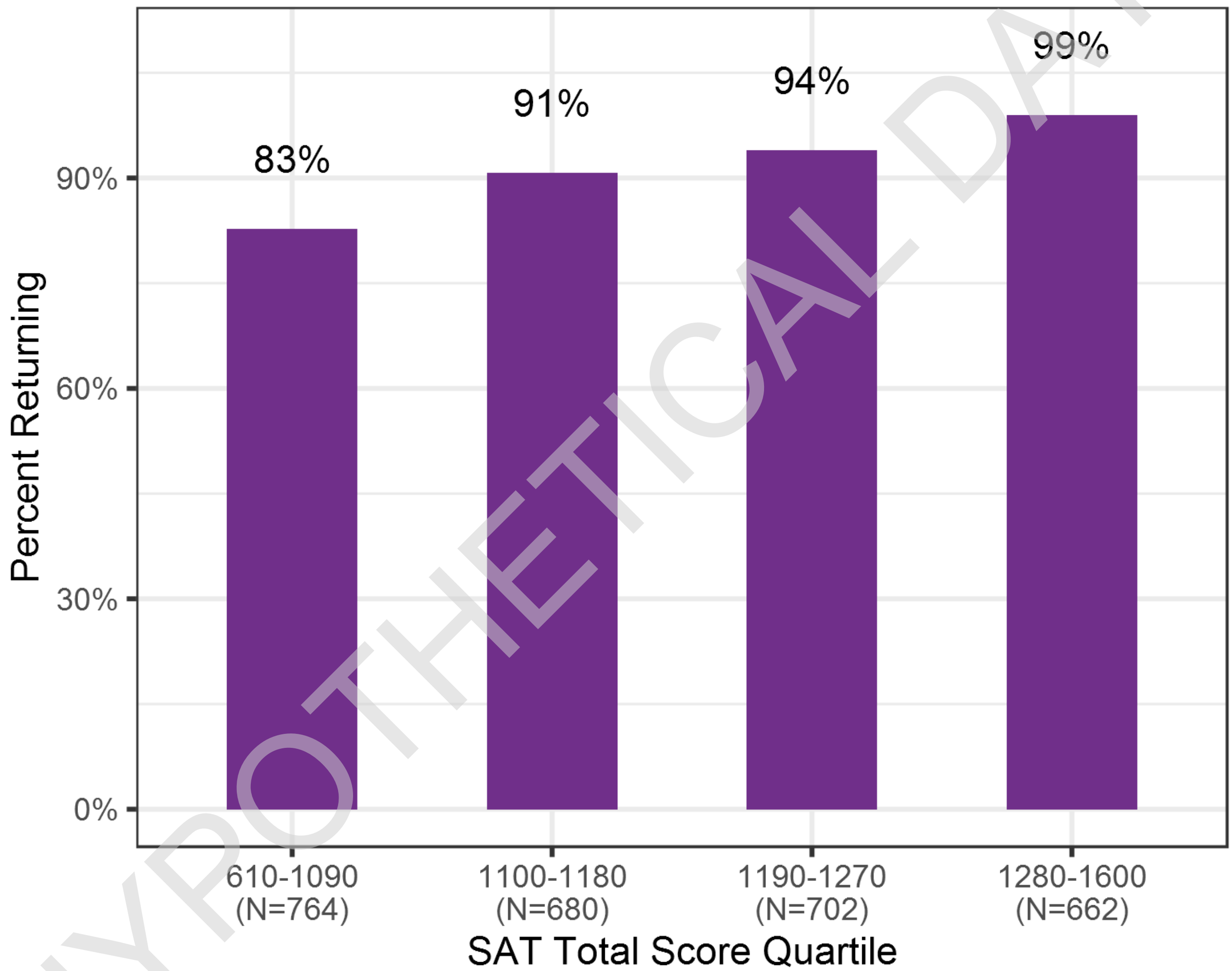
Type	Measure Name	N	Mean (SD)	Minimum	Maximum
Add. Predictor	AT_RISK	245	0.18 (0.38)	0.00	1.00
Add. Predictor	RANKING	245	3.22 (1.46)	1.00	5.00
College GPA	Cumulative GPA	245	2.88 (0.62)	1.62	4.00

Section 3: Evaluating study measures

This section presents several graphs that examine the relationship between retention predictors in your study and the retention outcome you chose: 2nd Year Retention. If two-category (0,1) predictors or retention descriptors were chosen in your study, then a table will display how often they occur. If there are between two and four descriptors and characteristics, a Venn diagram will present how often they occur in combination for students who did not return.

The first graph shows the average actual 2nd Year Retention Rate of your students for different SAT Total score quartiles, which illustrates the relationship between 2nd Year Retention and the SAT.

Mean 2nd Year Retention Rate by SAT Total Score Quartile



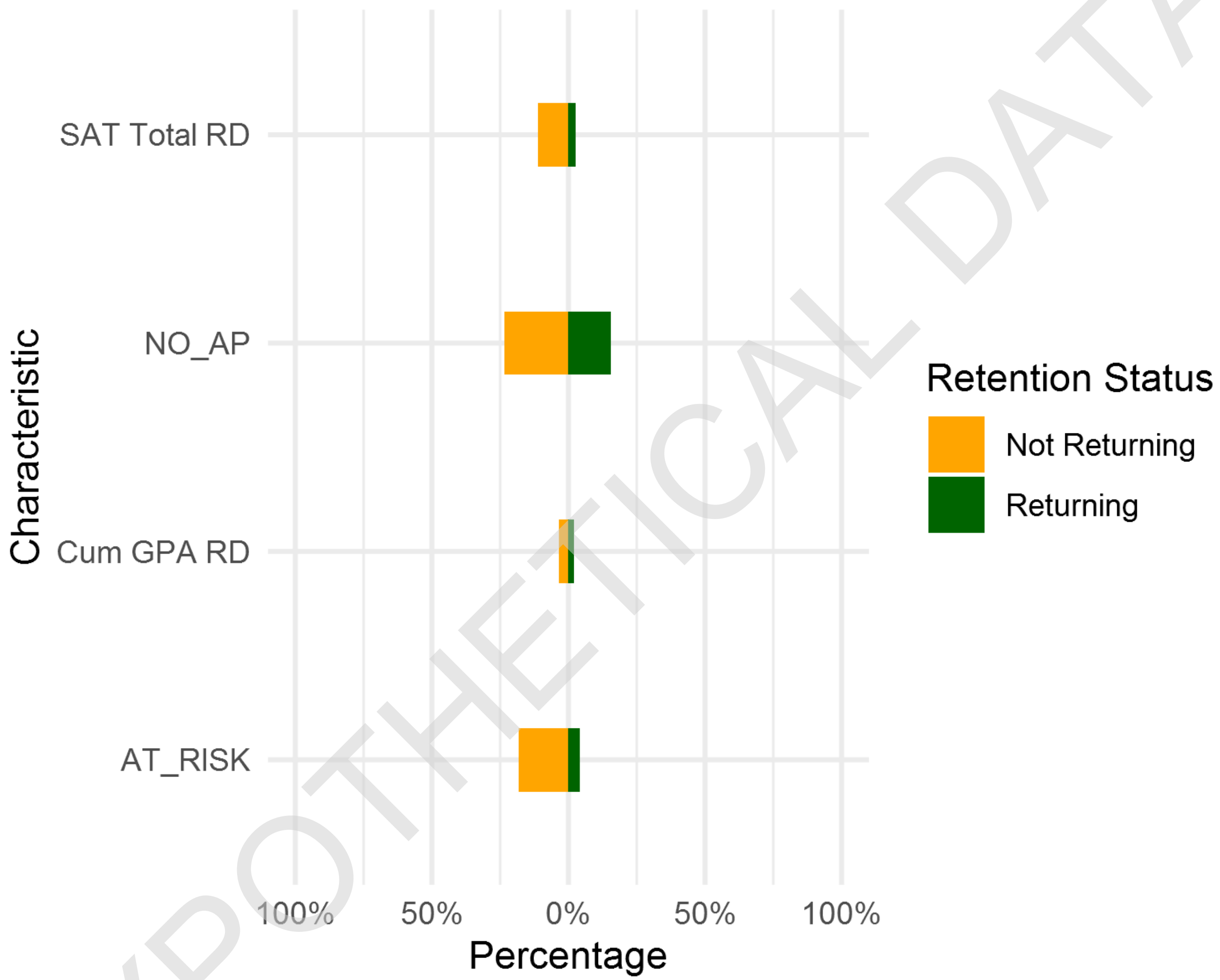
Notes:

- SAT Total score quartiles are based on the sum of the SAT RW and Math Section scores.
- Quartiles place students into four groups of approximately equal size based on the measure. Depending on the distribution of your students on the measure (e.g., no students with low measure values or a gap in the distribution of measure values), the quartile bands in the graph may not cover the full possible range of the measure; there may be gaps in values between the quartile bands, and there may be fewer than four bands.

The next graph presents a comparison between returning and non-returning students on any retention descriptors and two-category predictors of retention included in this study. For each characteristic, the two bars show the percent of returning students and the percent of non-returning students who exhibit the characteristic. Characteristics on which the student groups differ are of special interest and can help identify particular student characteristics that may be associated with not returning to your institution.

Retention descriptors are student characteristics that may provide insight into students who do not return. These descriptors can optionally be defined in the Retention Study Design and identify students who are at or below a user specified cut-point on such predictor variables as SAT tests, HS GPA, or college GPA. Any retention descriptors you defined in your Study Design will appear in the table below.

Characteristics of Returning and Non-Returning Students



Notes:

- Retention descriptors included in the graph are: SAT Total RD with cut-point of 930 on SAT Total score and Cum GPA RD with cut-point of 1.75 on Cumulative GPA.

The table below displays how frequently each retention descriptor or two-category predictor included in your study is present in non-returning students, which may be helpful in characterizing and understanding non-returning students at your institution.

Frequency of Retention Descriptors and Predictors for Non-Returning Students

Retention Descriptor or Predictor	Percentage (N)
SAT Total RD	11% (27)
Cum GPA RD	3% (8)

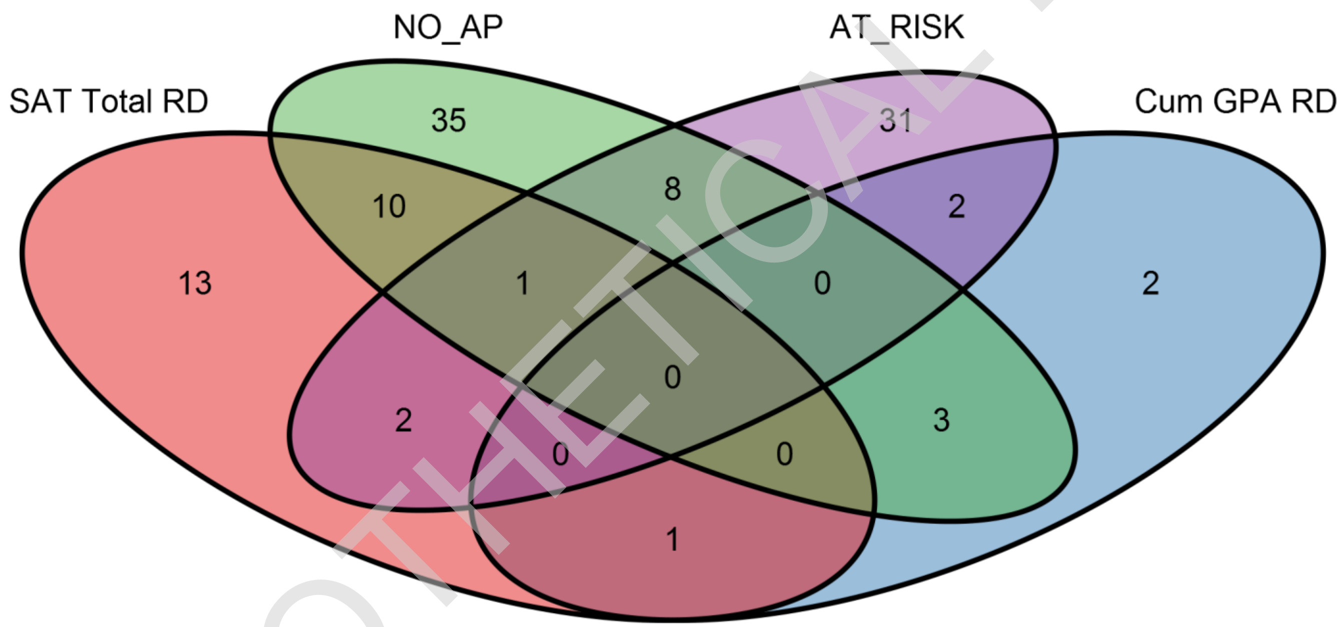
Retention Descriptor or Predictor	Percentage (N)
NO_AP	23% (57)
AT_RISK	18% (44)

Notes:

- Retention descriptors have “RD” added to their labels.

Below, the Venn diagram of the retention descriptors and two-category predictors shows how often these characteristics overlap or co-occur for non-returning students. Each ellipse represents a different characteristic, and the areas where ellipses overlap identify instances of non-returning students with multiple characteristics.

Venn Diagram of Retention Descriptors and Predictors for Non-Returning Students



Notes:

- Numbers within the regions of ellipse overlap are counts of non-returning students with specific combinations of retention descriptor or two-category predictor characteristics.
- Retention descriptors included in the graph are: SAT Total RD with cut-point of 930 on SAT Total score and Cum GPA RD with cut-point of 1.75 on Cumulative GPA.

Section 4: Modeling retention

This section will present the results from at least one, and as many as three, retention models applied to your students.

- A base logistic regression model will always be presented. It includes the SAT scores you selected and HS GPA if you selected it as a predictor in your model.
- A full logistic regression model with all predictors selected during the study design process will be presented if you selected more predictors than those appearing in the base model.
- A decision tree model will be run with all predictor variables, but results will only display if the decision tree expands beyond the first (root) node.

Assessing models

Percent correctly classified (all models)

A common way to measure the strength of prediction is to estimate the percentage of students correctly classified by the model. A student is considered to be correctly classified by the model if either: 1) it was predicted that the student would return, and they did return, or 2) it was predicted that the student would not return, and they did not return. The analyses reported here predict that a student will return if the student's estimated probability of returning is greater than or equal to the cut-off value presented in the results. Notice, however, that when nearly all of the students return, a predictor can have a high success rate even if it correlates very poorly with the criterion. For example, if 97 percent of the students return and the predictor simply predicts that all students will return, then the percent correctly classified will be 97.

Pseudo r-squared (regression models)

In ordinary regression, r-squared is a popular measure of model fit that estimates the proportion of variation in the outcome that can be predicted by the regression model. It has been extended to logistic regression in the form of pseudo r-squared measures which also attempt to estimate the proportion of variation in the outcome that can be predicted from the model. This study presents Nagelkerke's pseudo r-squared, which is scaled to range between 0 and 1, where 1 indicates the model perfectly predicts the outcome.

Note on composite predictors (regression models)

Predictor variables do not have to be used individually. Two or more predictors can be used together to form a composite predictor that may be stronger than either of the individual predictor variables alone. If you elected to use more than one predictor variable, the composite predictor is calculated by multiplying each individual predictor by a number that indicates its weight, or strength, in the prediction. The weighted predictors are added to a constant which results in a composite predictor score between approximately -3 to +3. This composite predictor score can be converted to a predicted probability of retention (see Appendix B).

Variable importance

Predictor variable importance in regression can be measured in various ways and there is no consensus on a single best measure in the context of logistic regression. In this study, two different measures of predictor importance for regression are presented.

Absolute Z and P Value

A measure of variable importance, used to order the predictors in a regression model, is the absolute Z statistic, calculated by dividing the coefficient by its standard error and taking its absolute value (so all values are positive). It is used in significance testing and measures how far each coefficient is from zero on a common scale. Predictors with larger absolute Z values are considered more important predictors and the predictors can be ranked on absolute Z. In addition, the statistical significance probability value (p value) for each predictor, calculated from the Z value, is also reported. While primarily used to assess the statistical significance of a predictor in the model, it can also provide a guide to variable importance: as absolute Z increases, the p value decreases, so predictors with smaller p values are considered more important.

Adequacy

The adequacy measure of variable importance in regression examines how much each predictor explains relative to the total amount explained by including all predictors, expressed as a percentage. It evaluates how well each predictor performs alone relative to all predictors. Alternatively, adequacy can be thought of as how well each single predictor can replace or act as a substitute for all predictor variables in the model. Predictors with higher adequacy percentages are more important and different predictors can be directly compared on this measure. The sum of the adequacy percentages across all predictors can exceed 100 percent due to overlap (correlations) among the predictors.

Note that in decision tree models, the best available predictor is selected to split a node. As a result, predictor variables selected to create early splits in a decision tree are considered more important predictors and predictors appearing in later splits or not appearing at all are considered less important.

Base regression model

The next two tables present information about the base regression model applied to your students.

Base Regression Model Fit Summary

Statistic	Base Model
N	2,808
Pseudo R-Square	0.12
% Correctly Classified	85.22%

Notes:

- In this model, to calculate the percentage correctly classified, students were predicted to return if their predicted probability of retention was 0.84 or greater.

Base Regression Model Summary

Predictor	Parameter estimate	Absolute Z	P - Value	Adequacy
(Intercept)	-5.23100			
SAT Total score	0.00639	10.88	0.000	99.85%
High School GPA	0.08668	0.48	0.629	12.16%

Notes:

- The absolute Z variable importance measure can be used to rank predictor variables in importance and the adequacy variable importance measure evaluates how much of the amount explained by all predictor variables can be accounted for by each individual predictor.

To calculate a predicted "retention score" for a student, you would multiply the student's value on each predictor variable by its prediction weight and then add the constant. Directions for using predicted retention scores will be found in Section 5. For example, the calculation of a predicted retention score for a student with an SAT Total score of 1180 and High School GPA of 3.8 would be -5.231 + 0.00639 * SAT Total score + 0.08668 * High School GPA or 2.54.

The table below examines the agreement between retention predicted by the model (classifying students with a predicted probability of returning of 0.84 or greater as predicted to return) and actual retention. Values in this table form the basis of the percent correctly classified reported earlier.

Base Model Success - Predicted Retention and Actual Retention

Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	80 (24%)	250 (76%)

Prediction	Actual Non-Returning Student	Actual Returning Student
Returning students	165 (7%)	2,313 (93%)

Full regression model

The next two tables present information about the full regression model applied to your students.

Full Regression Model Fit Summary

Statistic	Full Model
N	2,808
Pseudo R-Square	0.16
% Correctly Classified	86.57%

Notes:

- In this model, to calculate the percentage correctly classified, students were predicted to return if their predicted probability of retention was 0.83 or greater.

Full Regression Model Summary

Predictor	Parameter estimate	Absolute Z	P - Value	Adequacy
(Intercept)	-4.92813			
SAT Total score	0.00585	9.11	0.000	73.59%
High School GPA	-0.19167	1.00	0.316	8.96%
Cumulative GPA	0.58254	3.36	0.001	50.48%
NO_AP	0.20882	1.15	0.250	4.36%
AT_RISK	-0.69041	2.48	0.013	27.65%
RANKING	-0.13596	2.72	0.007	3.28%

Notes:

- The absolute Z variable importance measure can be used to rank predictor variables in importance and the adequacy variable importance measure evaluates how much of the amount explained by all predictor variables can be accounted for by each individual predictor.

To calculate a predicted “retention score” for a student, you would multiply the student’s value on each predictor variable by its prediction weight and then add the constant. Directions for using predicted retention scores will be found in Section 5. For example, the calculation of a predicted retention score for a student with an SAT Total score of 1180, High School GPA of 3.8, Cumulative GPA of 3.2, NO_AP of 1, AT_RISK of 1, and RANKING of 3 would be -4.92813 + 0.00585 * SAT Total score + -0.19167 * High School GPA + 0.58254 * Cumulative GPA + 0.20882 * NO_AP + -0.69041 * AT_RISK + -0.13596 * RANKING or 2.47.

The table below examines the agreement between retention predicted by the model (classifying students with a predicted probability of returning of 0.83 or greater as predicted to return) and actual retention. Values in this table form the basis of the percent correctly classified reported earlier.

Full Model Success - Predicted Retention and Actual Retention

Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	99 (30%)	231 (70%)
Returning students	146 (6%)	2,332 (94%)

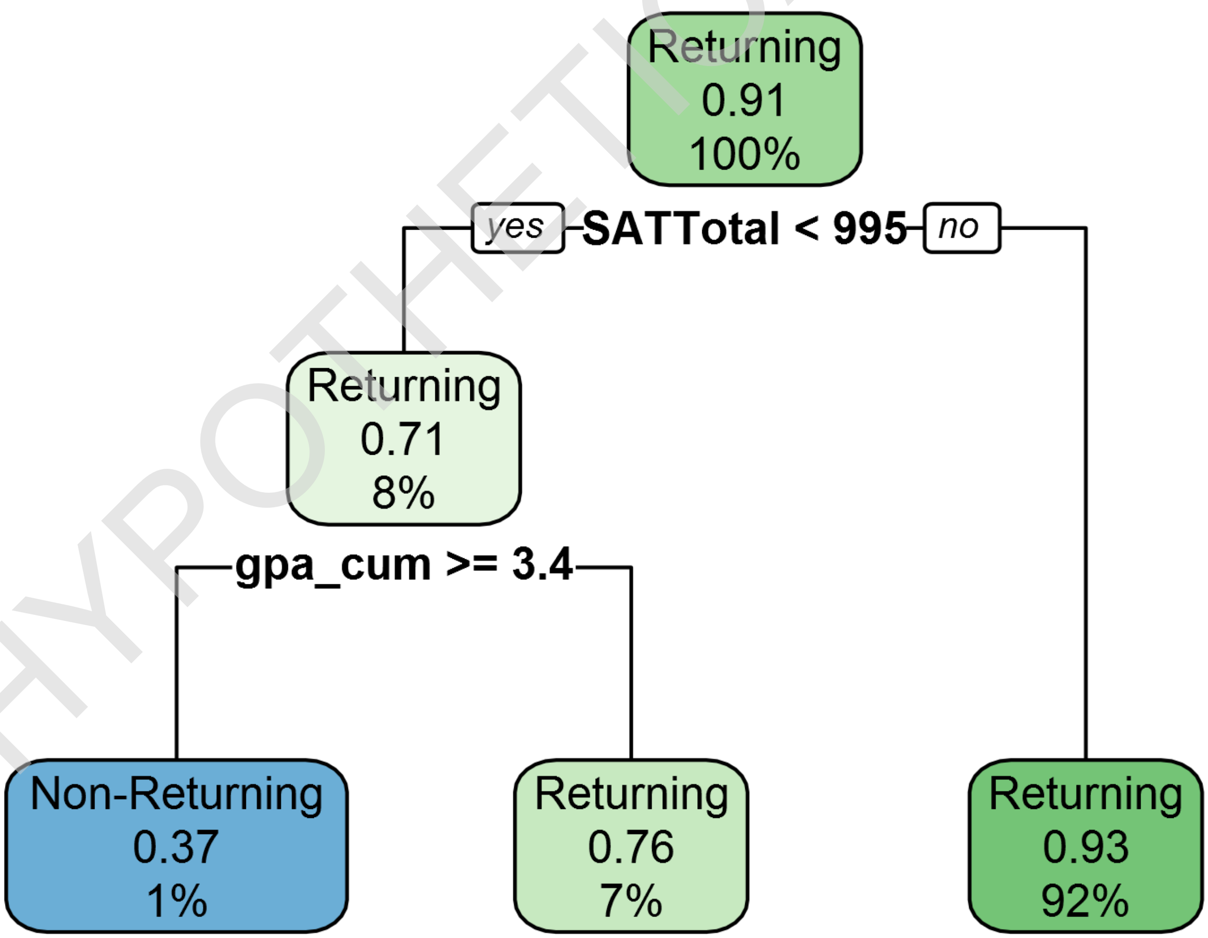
Decision tree model

As noted in Section 1, a decision tree builds a predictive model by successively splitting the data into subgroups that are homogeneous on the outcome measure, which in this case is retention. Since decision tree splits are built using student information as it relates to student retention, student characteristics that identify subgroups with relatively low retention rates are of interest and can be useful in thinking about students who may benefit from closer advisement to ensure they remain at the institution.

All predictor variables are eligible for inclusion in the decision tree, and those appearing higher in the tree tend to be more important in dividing the students into those with higher retention rates and those with lower retention.

Retention predictors included in the decision tree model are: SAT Total score, High School GPA, Cumulative GPA, NO_AP, AT_RISK, and RANKING.

Decision Tree Model



Notes:

- The uppermost node (root node) includes all students. To build the decision tree, students are further divided into groups based on predictors (cut-points for numeric predictors, category groupings for categorical predictors).
- The label information in each node includes three items: the decision tree retention prediction for students in that node (Returning, Non-Returning), the proportion of students within the node who were predicted to be returning, and the size of the node expressed as a percentage of students in the analysis.
- Below each node that splits is a label identifying the predictor and split criterion used to split the student data in that node into two groups that differ on retention.
- The two lines extending from each node lead to the child nodes resulting from the split.
- When a node is split, those students for whom the split condition is true follow the left branch below the node, and students for whom the split condition is false follow the right branch.
- The terminal nodes at the bottom of the tree are the end points of the decision tree model and represent final student groupings based on the predictors. These terminal nodes can be useful to better understand characteristics of non-returning students and in predicting retention for future students based on the decision tree model.

The table below examines the agreement between retention predicted by the decision tree model (classifying students with a predicted probability of returning of 0.76 or greater as predicted to return) and actual retention. Values in this table form the basis of the percent correctly classified. Using the decision tree model, the percent correctly classified is 88%.

Decision Tree Success-Predicted Retention and Actual Retention

Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	63 (29%)	151 (71%)
Returning students	183 (7%)	2,412 (93%)

It is important to keep in mind that the model success table aggregates over all terminal nodes in the decision tree. The decision tree itself should be examined to see which terminal nodes have higher proportions of non-returning students, since the characteristics of these groups can be useful in identifying and better understanding non-returning students.

Section 5: Using prediction models to identify students at risk of not returning

This section describes how to convert the predicted retention scores from the logistic regression model(s) in Section 4 to predicted retention probabilities. Also, several graphs and tables displaying predicted retention probability as a function of selected predictors will be presented.

Regression models

In Section 4, regression model weights were presented in a table with instructions on calculating the predicted retention score for a student. This predicted retention score is on a scale of approximately -3 to +3 and can be converted to a more easily interpretable retention probability to identify students at risk of not returning to the institution. The table below presents a list of retention score cutoff values, along with the predicted probability of retention. It can be used with the regression model(s) in Section 4.

Calculating Predicted Retention Probability from Retention Scores

Predicted Retention Score	Predicted Retention Probability
-2.94	0.05
-2.20	0.10
-1.73	0.15
-1.39	0.20
-1.10	0.25
-0.85	0.30
-0.62	0.35
-0.41	0.40
-0.20	0.45
0.00	0.50
0.20	0.55
0.41	0.60
0.62	0.65
0.85	0.70
1.10	0.75
1.39	0.80
1.73	0.85
2.20	0.90
2.94	0.95

Notes:

- To obtain a predicted retention probability from a retention score based on a logistic regression model, take the predicted score, calculated using both the regression weights and constant reported in Section 4 and the student's values on the predictors, and then locate that predicted retention score in the table to find the associated predicted retention probability.
- If your calculated retention score value doesn't match a value in the table, then round the calculated predicted retention value to two decimals, which should be near a table value. In this case the predicted probability of

retention will be accurate within 2.5 percent since the table is calibrated to 0.05 increments of predicted probability.

- Predicted retention probabilities can be calculated for all predicted retention scores (without requiring rounding and loss of accuracy) using a formula that can be implemented with a calculator or spreadsheet. Instructions for this appear in Appendix B.

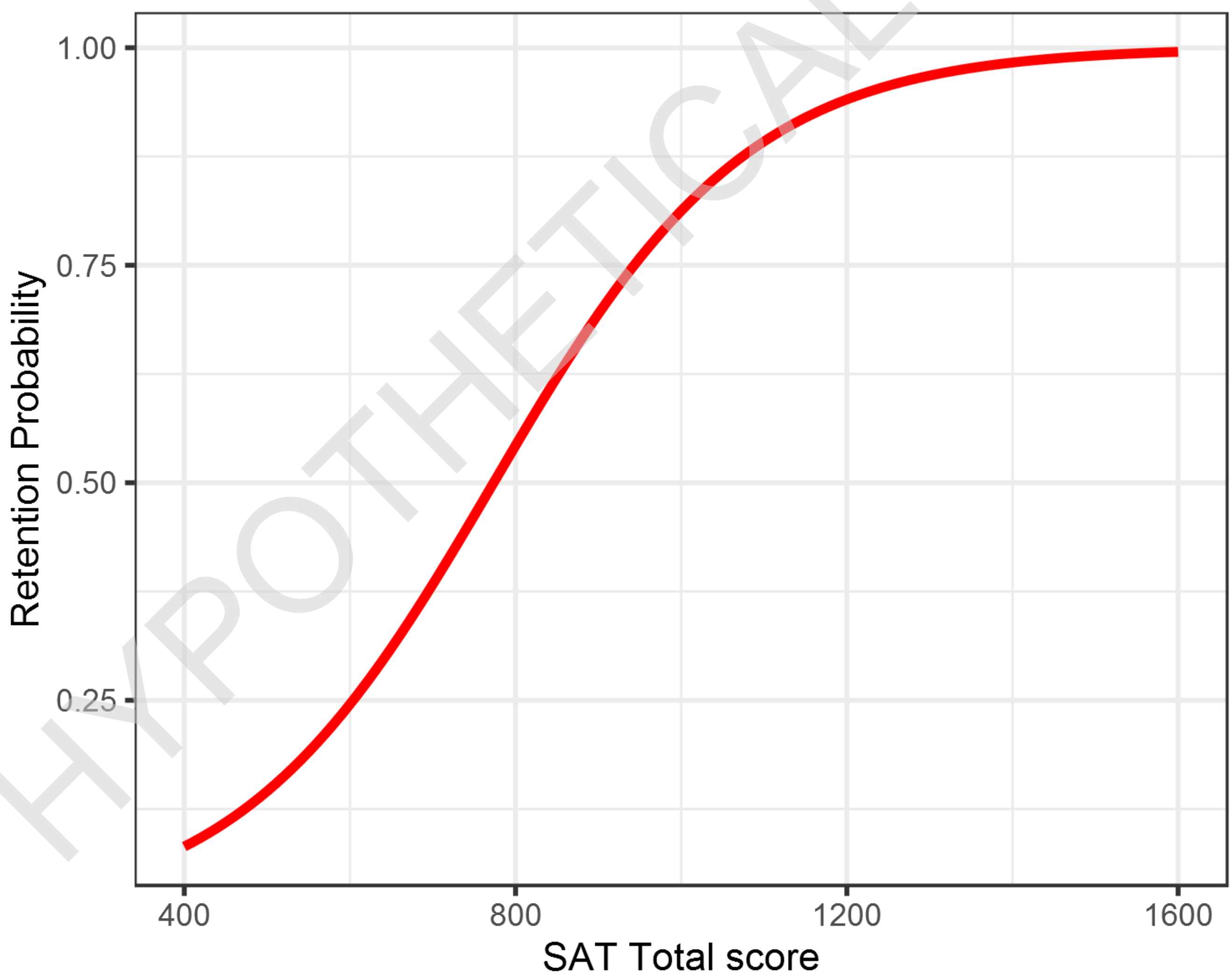
For example, if the predicted retention score for a student calculated using the student's values and the regression weights and constant from Section 4 is -1.39, the predicted probability of retention for the student is 0.20, and this student would have a relatively low predicted probability of retention.

Predicted retention probability related to select predictors

Some plots and tables are presented that relate predictor variables (from the predictors you requested) to predicted probability of retention.

The next graph illustrates the relationship between SAT Total score and the predicted probability of retention based on your student data.

Predicted 2nd Year Retention Probability Based on SAT Total score

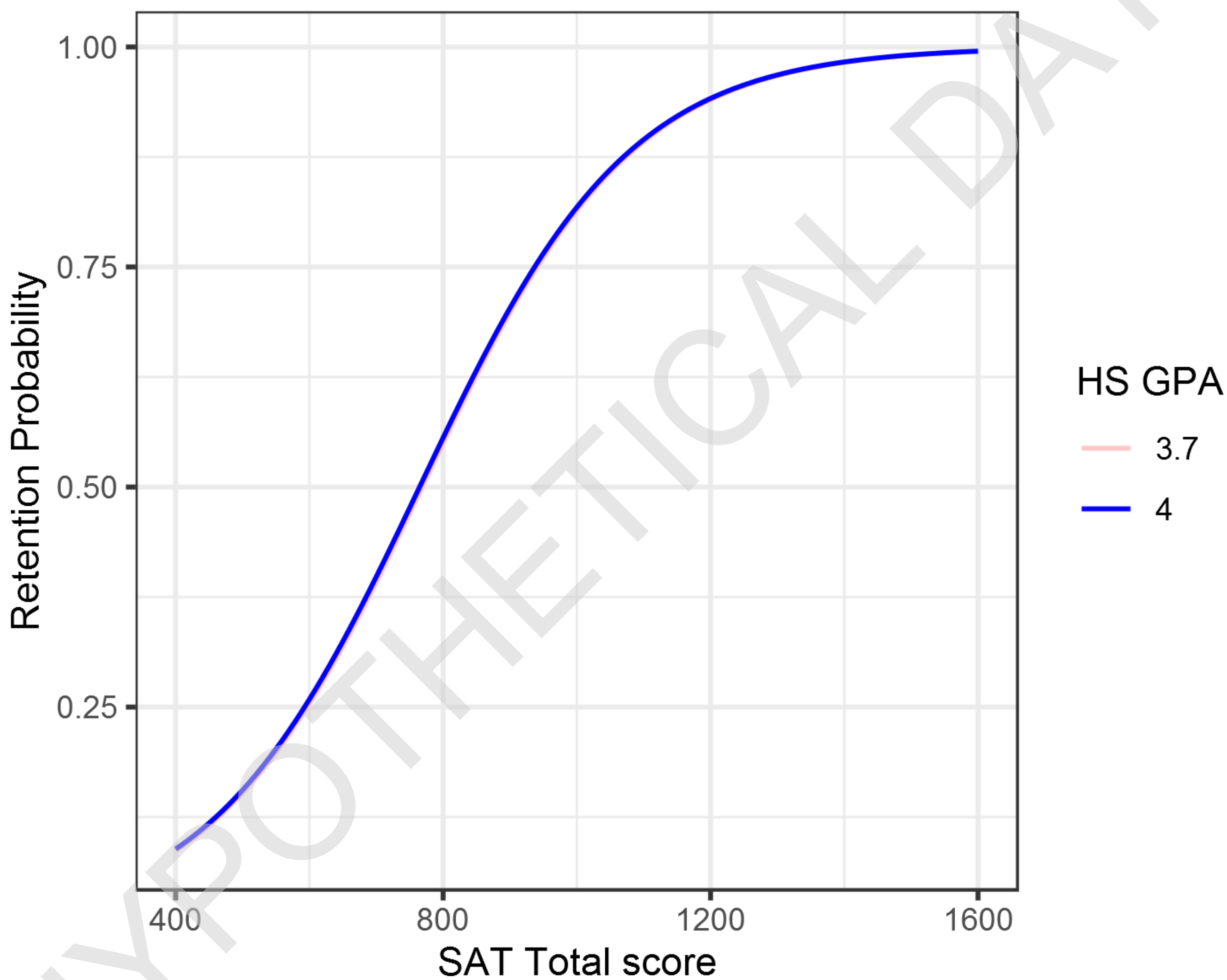


Notes:

- The predicted retention probability in this plot is based only on SAT Total score (i.e., if you selected other predictor variables, they are not reflected in this plot).
- The plot illustrates the relationship between SAT Total score and predicted retention probability. You can use this plot to better understand how particular SAT Total score values relate to particular probabilities of returning to your institution.

The next graph illustrates the relationship between SAT Total score and the predicted probability of retention for the 50th and 75th percentile HS GPA values based on your student data. This plot helps you to understand differences in predicted probabilities by different HS GPA values for students with similar SAT scores and vice versa.

Predicted 2nd Year Retention Probability Based on SAT Total score and HS GPA



Notes:

- The predicted retention probability in this plot is based only on the predictor variables listed in the plot (i.e., if you selected other predictor variables, they are not reflected in this plot).
- If only a single line appears in the plot then the 50th and 75th percentile values of HS GPA are the same for students in the analysis.
- This plot illustrates the relationship between SAT Total score and predicted retention probability for the 50th and 75th percentile values of HS GPA. This plot helps you to understand differences in predicted probabilities by different HS GPA values for students with similar SAT scores and vice versa.

The table below presents predicted completion probabilities for your students by averaging the predicted completion probabilities over SAT Total score quartiles and HS GPA categories formed by rounding or grouping HS GPA values.

Estimated Predicted 2nd Year Retention Probabilities by SAT Total score and HS GPA

SAT Total score	HS GPA A Mean (N)	HS GPA B Mean (N)	HS GPA C or below Mean (N)
610-1090	0.84 (553)	0.79 (203)	0.66 (8)
1100-1180	0.92 (551)	0.91 (129)	
1190-1270	0.95 (599)	0.95 (100)	0.95 (3)
1280-1600	0.98 (599)	0.97 (63)	

Appendix A: Statistical summaries for subgroups

This appendix contains descriptive statistics for each student subgroup. In addition, modeling results will be presented for each of the subgroups. Subgroup specific models may be useful if the prediction model for all students does not fit a specific subgroup well.

To better understand how subgroups might differ from the full student sample with regard to retention, the tables right below present the percent correctly classified from Section 3, calculated separately for each subgroup. In other words, the common logistic regression model (fit to all students in the study) is applied to each subgroup and evaluated.

Subgroup Results Predicting Retention - Base Regression Model

Subgroup	Number of Students Modeled	N (%) Correctly Classified
SCHOOL: College	2,287	1,958 (86%)
SCHOOL: Engineering	521	435 (83%)

Subgroup Results Predicting Retention - Full Regression Model

Subgroup	Number of Students Modeled	N (%) Correctly Classified
SCHOOL: College	2,287	1,988 (87%)
SCHOOL: Engineering	521	443 (85%)

SCHOOL: College

This section presents a descriptive summary of the study measures students in SCHOOL: College.

Student 2nd Year Retention for SCHOOL: College

Student Outcome	N (%)
Non-returning students	182 (8%)
Returning students	2,105 (92%)

The table below displays the mean, standard deviation (SD), minimum, and maximum of each individual measure selected for your study, and the number of students (N) with information available on each measure. Some measures may be available for all or nearly all of your students. Others may only be available for smaller groups of students. The table presents all measures with information available on 5 or more students.

Statistical Summaries of Study Measures for all Students in SCHOOL: College

Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	2,287	3.76 (0.39)	2.30	4.30
SAT Test Score	SAT Total score	2,287	1185 (137)	610	1600
Add. Predictor	NO_AP	2,287	0.17 (0.38)	0.00	1.00
Add. Predictor	AT_RISK	2,287	0.05 (0.22)	0.00	1.00
Add. Predictor	RANKING	2,287	3.00 (1.41)	1.00	5.00
College GPA	Cumulative GPA	2,287	3.24 (0.55)	1.61	4.00

The next two tables present summaries of the study measures for returning students and for non-returning students.

Statistical Summaries of Study Measures for Returning Students in SCHOOL: College

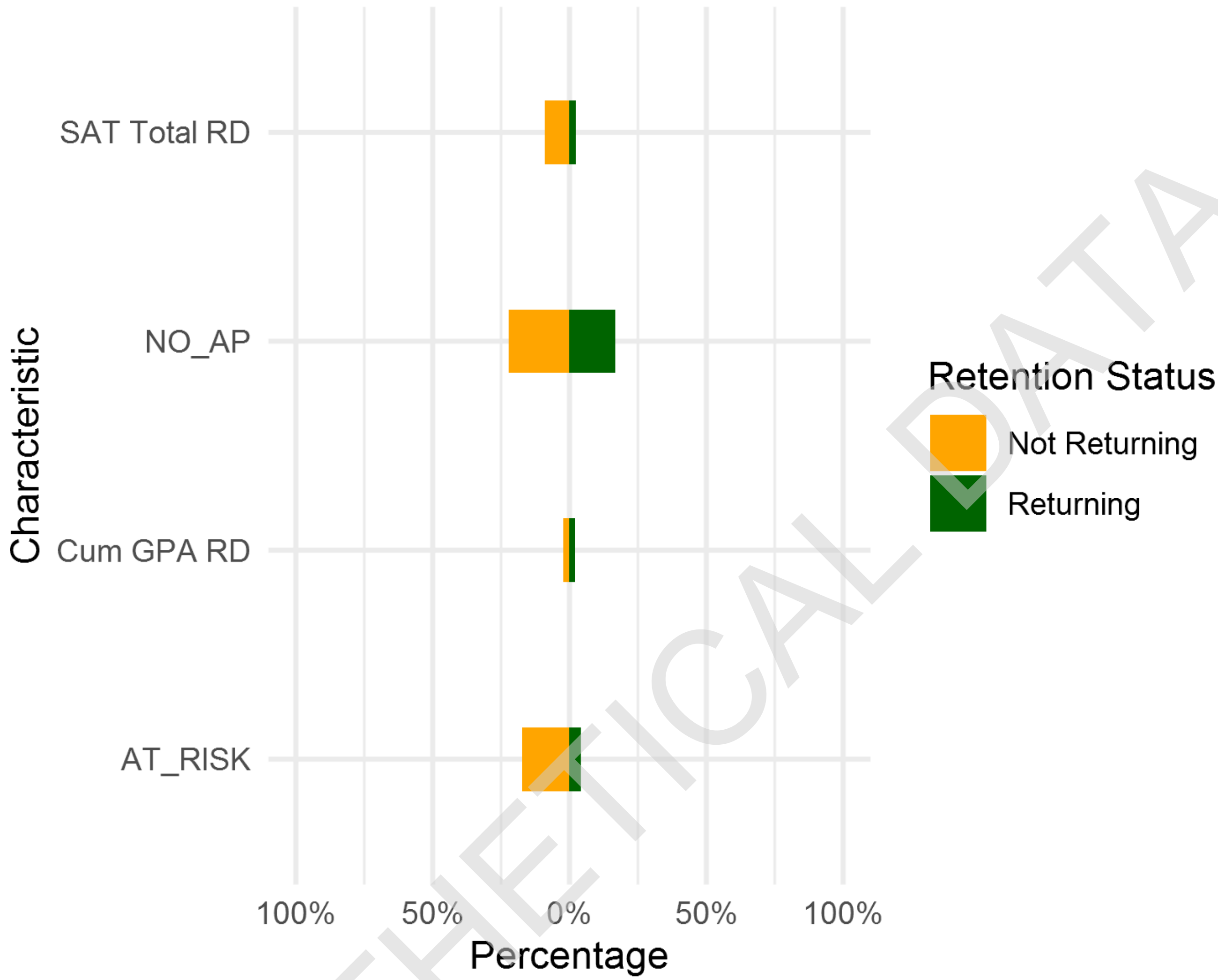
Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	2,105	3.77 (0.39)	2.30	4.30
SAT Test Score	SAT Total score	2,105	1193 (135)	610	1600
Add. Predictor	NO_AP	2,105	0.17 (0.37)	0.00	1.00
Add. Predictor	AT_RISK	2,105	0.04 (0.20)	0.00	1.00
Add. Predictor	RANKING	2,105	2.97 (1.41)	1.00	5.00
College GPA	Cumulative GPA	2,105	3.27 (0.54)	1.61	4.00

Statistical Summaries of Study Measures for Non-Returning Students in SCHOOL: College

Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	182	3.67 (0.37)	2.70	4.30
SAT Test Score	SAT Total score	182	1088 (114)	750	1410
Add. Predictor	NO_AP	182	0.22 (0.42)	0.00	1.00
Add. Predictor	AT_RISK	182	0.17 (0.38)	0.00	1.00
Add. Predictor	RANKING	182	3.29 (1.47)	1.00	5.00
College GPA	Cumulative GPA	182	2.91 (0.63)	1.62	4.00

The graph below presents, for returning and non-returning students in SCHOOL: College the percentage with characteristic measured by each retention descriptor or two-category predictor of retention.

Characteristics of Returning and Non-Returning Students for SCHOOL: College



Notes:

- Retention descriptors included in the graph are: SAT Total RD with cut-point of 930 on SAT Total score and Cum GPA RD with cut-point of 1.75 on Cumulative GPA.

The table below displays how frequently each retention descriptor or two-category predictor included in your study is present in non-returning students, which may be helpful in characterizing and understanding non-returning students at your institution.

Frequency of Retention Descriptors and Predictors for Non-Returning Students in SCHOOL: College

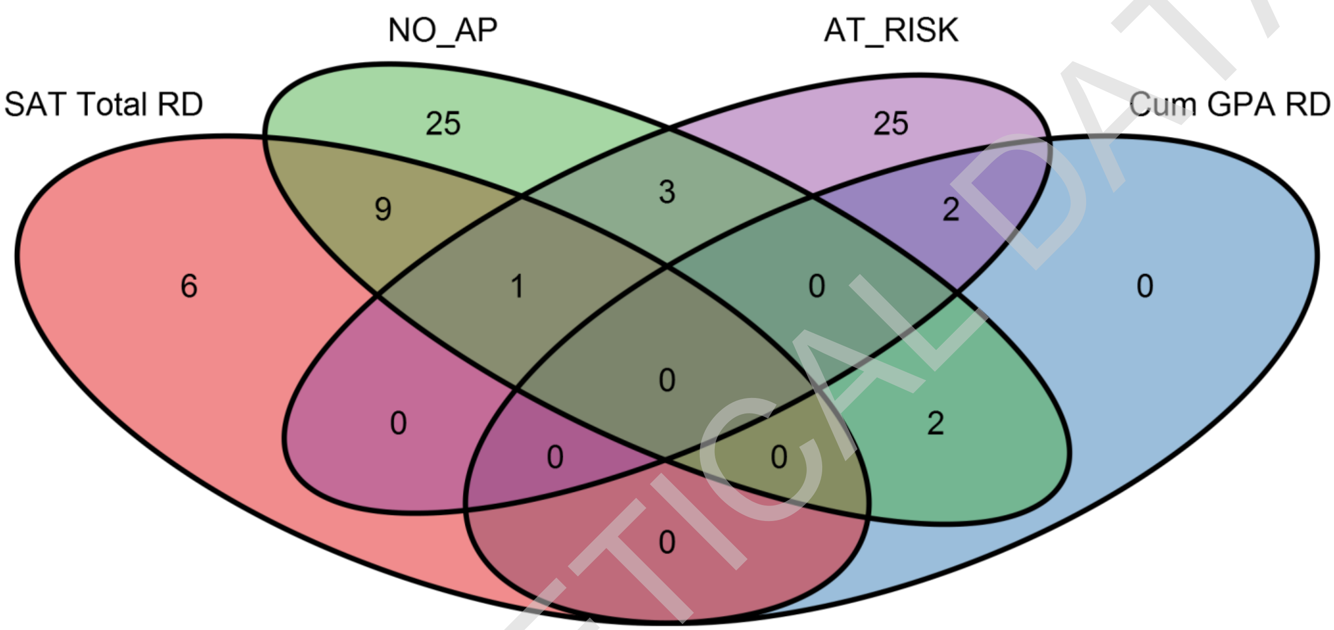
Retention Descriptor or Predictor	Percentage (N)
SAT Total RD	9% (16)
Cum GPA RD	2% (4)
NO_AP	22% (40)
AT_RISK	17% (31)

Notes:

- Retention descriptors have “RD” added to their labels.

Below, the Venn diagram of the retention descriptors and two-category predictors shows how often these characteristics overlap or co-occur for non-returning students. Each ellipse represents a different characteristic, and the areas where ellipses overlap identify instances of non-returning students with multiple characteristics.

Venn Diagram of Retention Descriptors and Predictors for Non-Returning Students in SCHOOL: College



Notes:

- Retention descriptors included in the graph are: SAT Total RD with cut-point of 930 on SAT Total score and Cum GPA RD with cut-point of 1.75 on Cumulative GPA.

The two tables below present information about the base regression model for students in subgroup SCHOOL: College.

Base Regression Model Fit Summary for SCHOOL: College

Statistic	Base Model
N	2,287
Pseudo R-Square	0.10
% Correctly Classified	85.83%

Notes:

- In this model, to calculate the percentage correctly classified, students were predicted to return if their predicted probability of retention was 0.85 or greater.

Base Regression Model Summary for SCHOOL: College

Predictor	Parameter estimate	Absolute Z	P - Value	Adequacy
(Intercept)	-4.53049			
SAT Total score	0.00608	9.08	0.000	100.00%
High School GPA	0.01326	0.06	0.949	9.93%

Base Model Success - Predicted Retention and Actual Retention for SCHOOL: College

Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	52 (21%)	194 (79%)
Returning students	130 (6%)	1,911 (94%)

The two tables below present information about the full regression model for students in subgroup SCHOOL: College.

Full Regression Model Fit Summary for SCHOOL: College

Statistic	Full Model
N	2,287
Pseudo R-Square	0.15
% Correctly Classified	88.02%

Notes:

- In this model, to calculate the percentage correctly classified, students were predicted to return if their predicted probability of retention was 0.83 or greater.

Full Regression Model Summary for SCHOOL: College

Predictor	Parameter estimate	Absolute Z	P - Value	Adequacy
(Intercept)	-4.34072			
SAT Total score	0.00592	7.95	0.000	70.57%
High School GPA	-0.23188	1.06	0.289	7.01%
Cumulative GPA	0.46856	2.36	0.018	44.44%
NO_AP	0.39105	1.83	0.067	2.18%
AT_RISK	-0.79473	2.45	0.014	26.69%
RANKING	-0.17303	3.01	0.003	5.82%

Full Model Success - Predicted Retention and Actual Retention for SCHOOL: College

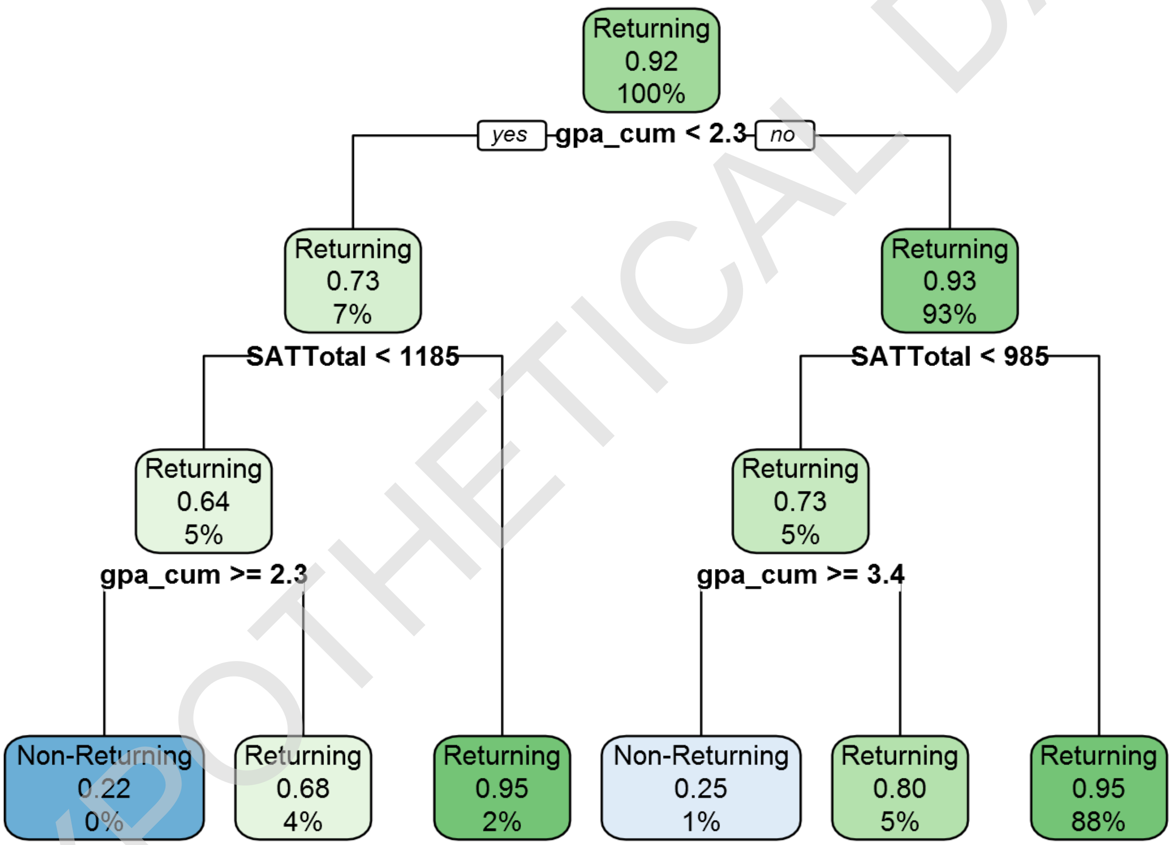
Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	62 (29%)	154 712%)
Returning students	120 (6%)	1,951 (94%)

Decision tree model

As noted in Section 1, a decision tree builds a predictive model by successively splitting the data into subgroups that are homogeneous on the outcome measure, which in this case is retention. Since decision tree splits are built using student information as it relates to student retention, student characteristics that identify subgroups with relatively low retention rates are of interest and can be useful in thinking about students who may benefit from closer advisement to ensure they remain at the institution.

All predictor variables are eligible for inclusion in the decision tree, and those appearing higher in the tree tend to be more important in dividing the students into those with higher retention rates and those with lower retention.

Retention predictors included in the decision tree model are: SAT Total score, High School GPA, Cumulative GPA, NO_AP, AT_RISK, and RANKING.



Notes:

- The uppermost node (root node) includes all students. To build the decision tree, students are further divided into groups based on predictors (cut-points for numeric predictors, category groupings for categorical predictors).
- The label information in each node includes three items: the decision tree retention prediction for students in that node (Returning, Non-Returning), the proportion of students within the node who were predicted to be returning, and the size of the node expressed as a percentage of students in the analysis.
- Below each node that splits is a label identifying the predictor and split criterion used to split the student data in that node into two groups more homogeneous on retention.

- The two lines extending from each node lead to the child nodes resulting from the split.
- When a node is split, those students for whom the split condition is true follow the left branch below the node, and students for whom the split condition is false follow the right branch.
- The terminal nodes at the bottom of the tree are the end points of the decision tree model and represent final student groupings based on the predictors. These terminal nodes can be useful to better understand characteristics of non-returning students and in predicting retention for future students based on the decision tree model.

The table below examines the agreement between retention predicted by the decision tree model (classifying students with a predicted probability of returning of 0.80 or greater as predicted to return) and actual retention. Values in this table form the basis of the percent correctly classified. Using the decision tree model, the percent correctly classified is 88%.

Decision Tree Success–Predicted Retention and Actual Retention for SCHOOL: College

Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	74 (316%)	161 (69%)
Returning students	108 (5%)	1,944 (95%)

It is important to keep in mind that the model success table aggregates over all terminal nodes in the decision tree. The decision tree itself should be examined to see which terminal nodes have higher proportions of non-returning students, since the characteristics of these groups can be useful in identifying and better understanding non-returning students.

SCHOOL: Engineering

This section presents a descriptive summary of the study measures students in SCHOOL: Engineering.

Student 2nd Year Retention for SCHOOL: Engineering

Student Outcome	N (%)
Non-returning students	63 (12%)
Returning students	458 (88%)

The table below displays the mean, standard deviation (SD), minimum, and maximum of each individual measure selected for your study, and the number of students (N) with information available on each measure. Some measures may be available for all or nearly all of your students. Others may only be available for smaller groups of students. The table presents all measures with information available on 5 or more students.

Statistical Summaries of Study Measures for all Students in SCHOOL: Engineering

Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	521	3.79 (0.37)	2.70	4.30
SAT Test Score	SAT Total score	521	1172 (142)	740	1600
Add. Predictor	NO_AP	521	0.12 (0.33)	0.00	1.00
Add. Predictor	AT_RISK	521	0.06 (0.24)	0.00	1.00
Add. Predictor	RANKING	521	3.00 (1.41)	1.00	5.00
College GPA	Cumulative GPA	521	3.21 (0.57)	1.62	4.00

The next two tables present summaries of the study measures for returning students and for non-returning students.

Statistical Summaries of Study Measures for Returning Students in SCHOOL: Engineering

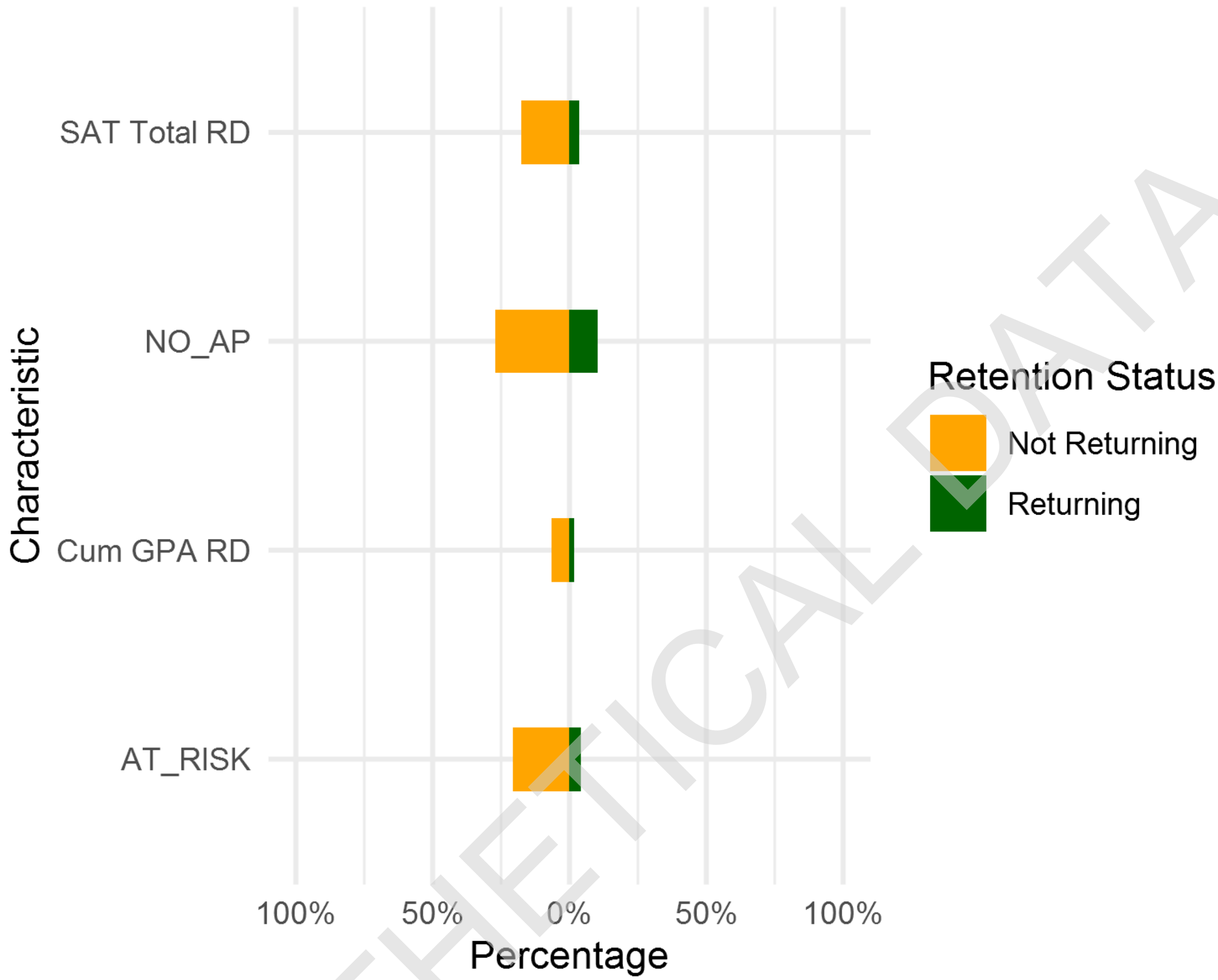
Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	458	3.82 (0.35)	2.70	4.30
SAT Test Score	SAT Total score	458	1188 (138)	740	1600
Add. Predictor	NO_AP	458	0.10 (0.30)	0.00	1.00
Add. Predictor	AT_RISK	458	0.04 (0.20)	0.00	1.00
Add. Predictor	RANKING	458	3.00 (1.41)	1.00	5.00
College GPA	Cumulative GPA	458	3.28 (0.54)	1.67	4.00

Statistical Summaries of Study Measures for Non-Returning Students in SCHOOL: Engineering

Type	Measure Name	N	Mean (SD)	Minimum	Maximum
High School GPA	High School GPA	63	3.64 (0.42)	2.70	4.30
SAT Test Score	SAT Total score	63	1060 (122)	770	1430
Add. Predictor	NO_AP	63	0.27 (0.45)	0.00	1.00
Add. Predictor	AT_RISK	63	0.21 (0.41)	0.00	1.00
Add. Predictor	RANKING	63	3.03 (1.44)	1.00	5.00
College GPA	Cumulative GPA	63	2.77 (0.59)	1.62	3.92

The graph below presents, for returning and non-returning students in SCHOOL: Engineering the percentage with characteristic measured by each retention descriptor or two-category predictor of retention.

Characteristics of Returning and Non-Returning Students for SCHOOL: Engineering



Notes:

- Retention descriptors included in the graph are: SAT Total RD with cut-point of 930 on SAT Total score and Cum GPA RD with cut-point of 1.75 on Cumulative GPA.

The table below displays how frequently each retention descriptor or two-category predictor included in your study is present in non-returning students, which may be helpful in characterizing and understanding non-returning students at your institution.

Frequency of Retention Descriptors and Predictors for Non-Returning Students in SCHOOL: Engineering

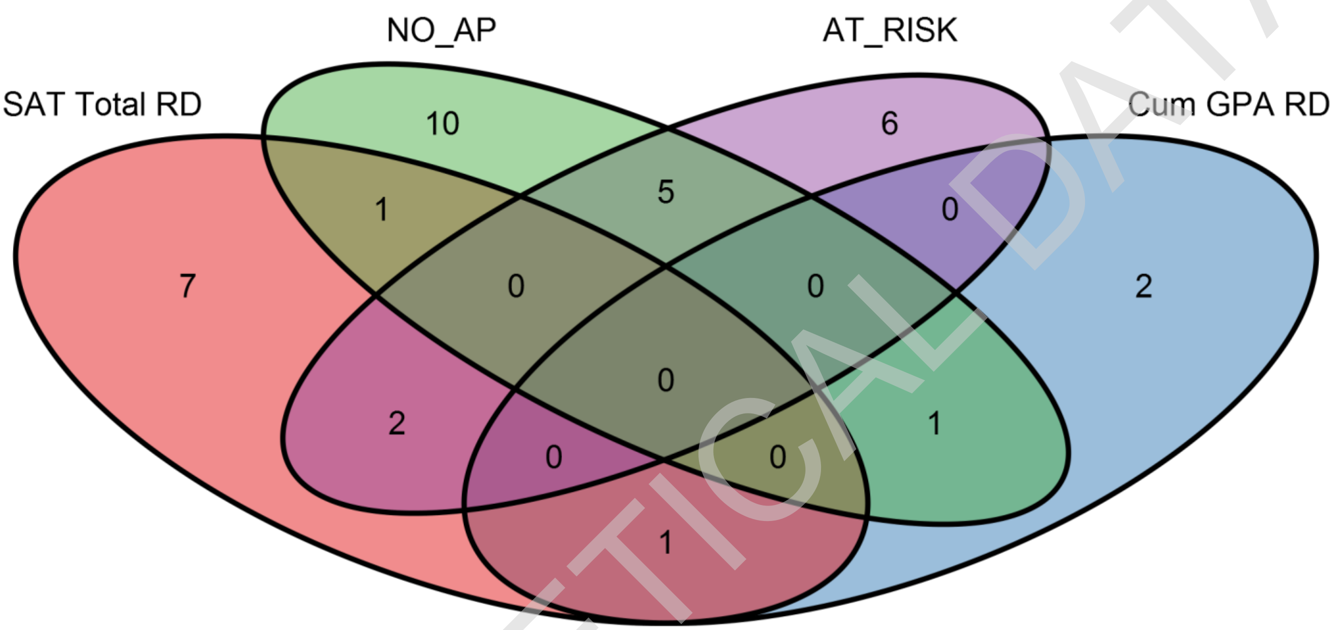
Retention Descriptor or Predictor	Percentage (N)
SAT Total RD	17% (11)
Cum GPA RD	6% (4)
NO_AP	27% (17)
AT_RISK	21% (13)

Notes:

- Retention descriptors have "RD" added to their labels.

Below, the Venn diagram of the retention descriptors and two-category predictors shows how often these characteristics overlap or co-occur for non-returning students. Each ellipse represents a different characteristic, and the areas where ellipses overlap identify instances of non-returning students with multiple characteristics.

Venn Diagram of Retention Descriptors and Predictors for Non-Returning Students in SCHOOL: Engineering



Notes:

- Retention descriptors included in the graph are: SAT Total RD with cut-point of 930 on SAT Total score and Cum GPA RD with cut-point of 1.75 on Cumulative GPA.

The two tables below present information about the base regression model for students in subgroup SCHOOL: Engineering.

Base Regression Model Fit Summary for SCHOOL: Engineering

Statistic	Base Model
N	521
Pseudo R-Square	0.18
% Correctly Classified	82.15%

Notes:

- In this model, to calculate the percentage correctly classified, students were predicted to return if their predicted probability of retention was 0.79 or greater.

Base Regression Model Summary for SCHOOL: Engineering

Predictor	Parameter estimate	Absolute Z	P - Value	Adequacy
(Intercept)	-7.66405			
SAT Total score	0.00715	5.74	0.000	97.36%
High School GPA	0.43554	1.17	0.243	22.64%

Base Model Success - Predicted Retention and Actual Retention for SCHOOL: Engineering

Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	25 (31%)	55 (69%)
Returning students	38 (9%)	403 (91%)

The two tables below present information about the full regression model for students in subgroup SCHOOL: Engineering.

Full Regression Model Fit Summary for SCHOOL: Engineering

Statistic	Full Model
N	521
Pseudo R-Square	0.24
% Correctly Classified	85.41%

Notes:

- In this model, to calculate the percentage correctly classified, students were predicted to return if their predicted probability of retention was 0.77 or greater.

Full Regression Model Summary for SCHOOL: Engineering

Predictor	Parameter estimate	Absolute Z	P - Value	Adequacy
(Intercept)	-6.86352			
SAT Total score	0.00565	4.25	0.000	71.98%
High School GPA	-0.01010	0.03	0.980	16.74%
Cumulative GPA	0.91479	2.50	0.012	61.18%
NO_AP	-0.51733	1.42	0.156	17.13%
AT_RISK	-0.34953	0.61	0.539	26.74%
RANKING	-0.03034	0.29	0.774	0.05%

Full Model Success - Predicted Retention and Actual Retention for SCHOOL: Engineering

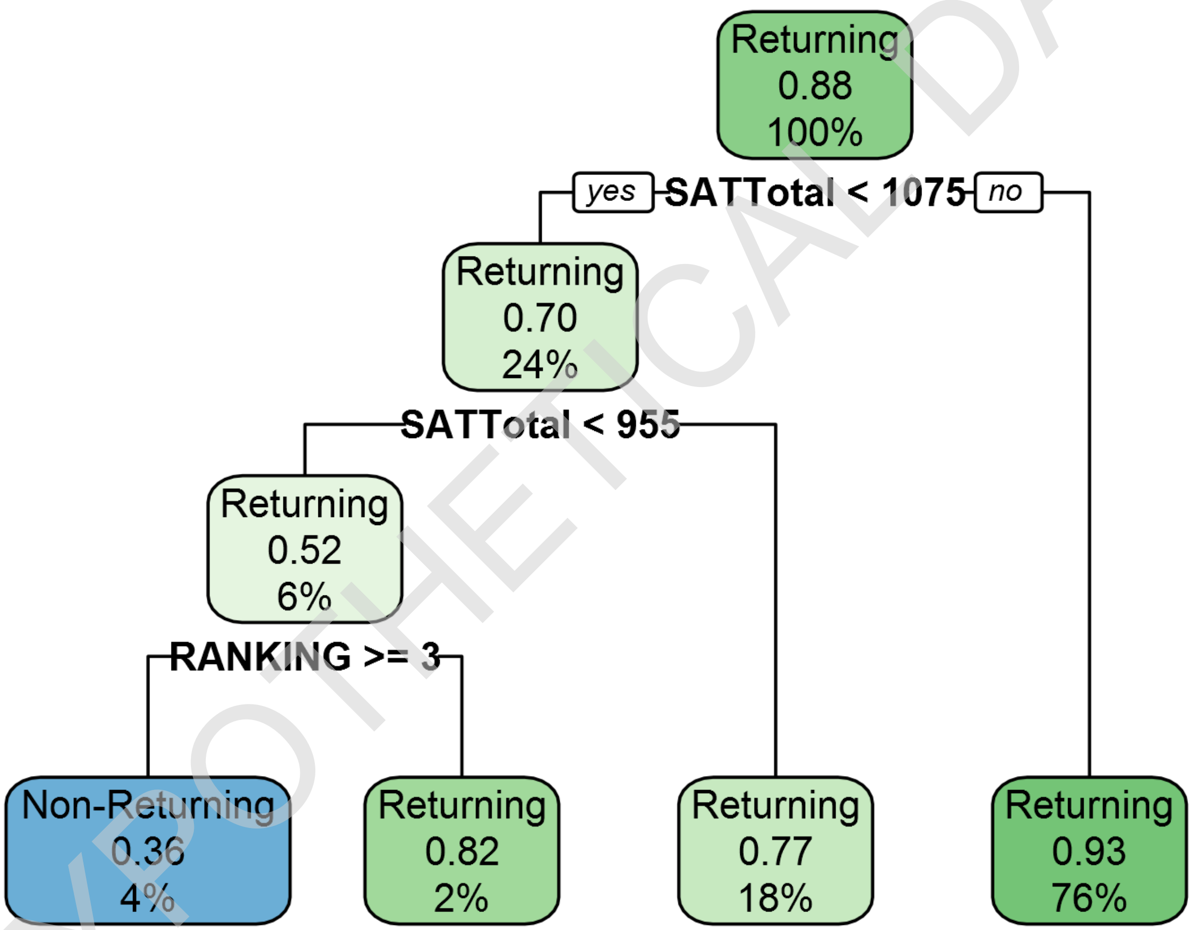
Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	32 (42%)	45 (58%)
Returning students	31 (7%)	413 (93%)

Decision tree model

As noted in Section 1, a decision tree builds a predictive model by successively splitting the data into subgroups that are homogeneous on the outcome measure, which in this case is retention. Since decision tree splits are built using student information as it relates to student retention, student characteristics that identify subgroups with relatively low retention rates are of interest and can be useful in thinking about students who may benefit from closer advisement to ensure they remain at the institution.

All predictor variables are eligible for inclusion in the decision tree, and those appearing higher in the tree tend to be more important in dividing the students into those with higher retention rates and those with lower retention.

Retention predictors included in the decision tree model are: SAT Total score, High School GPA, Cumulative GPA, NO_AP, AT_RISK, and RANKING.



Notes:

- The uppermost node (root node) includes all students. To build the decision tree, students are further divided into groups based on predictors (cut-points for numeric predictors, category groupings for categorical predictors).
- The label information in each node includes three items: the decision tree retention prediction for students in that node (Returning, Non-Returning), the proportion of students within the node who were predicted to be returning, and the size of the node expressed as a percentage of students in the analysis.
- Below each node that splits is a label identifying the predictor and split criterion used to split the student data in that node into two groups more homogeneous on retention.

- The two lines extending from each node lead to the child nodes resulting from the split.
- When a node is split, those students for whom the split condition is true follow the left branch below the node, and students for whom the split condition is false follow the right branch.
- The terminal nodes at the bottom of the tree are the end points of the decision tree model and represent final student groupings based on the predictors. These terminal nodes can be useful to better understand characteristics of non-returning students and in predicting retention for future students based on the decision tree model.

The table below examines the agreement between retention predicted by the decision tree model (classifying students with a predicted probability of returning of 0.78 or greater as predicted to return) and actual retention. Values in this table form the basis of the percent correctly classified. Using the decision tree model, the percent correctly classified is 79%.

Decision Tree Success–Predicted Retention and Actual Retention for SCHOOL: Engineering

Prediction	Actual Non-Returning Student	Actual Returning Student
Non-returning students	35 (31%)	79 (69%)
Returning students	28 (7%)	379 (93%)

It is important to keep in mind that the model success table aggregates over all terminal nodes in the decision tree. The decision tree itself should be examined to see which terminal nodes have higher proportions of non-returning students, since the characteristics of these groups can be useful in identifying and better understanding non-returning students.

Appendix B: Calculating predicted retention probability

Section 4 presented logistic regression weights (coefficients) that can be used to calculate predicted retention scores, and Section 5 included a table mapping predicted retention scores to predicted retention probabilities. However, this table displayed predicted retention probabilities in increments of 0.05, which places limits on its accuracy and usefulness.

As an alternative, predicted retention probabilities can be calculated from predicted retention scores through functions available on calculators or via spreadsheet formulas.

The relation between predicted retention probability (prob) and the predicted retention score (score) is:

prob = exp(score)/(1+exp(score))

Where *exp* is the exponential function, which is available on calculators, statistical programs, and spreadsheets.

To illustrate using the base logistic model results from Section 4, the calculation of a predicted retention score for a student with SAT Total score of 1180 and High School GPA of 3.8 would be -5.231 + 0.00639 * SAT Total score + 0.08668 * High School GPA or 2.54. And the predicted retention probability would be exp(2.54)/(1 + exp(2.54)) or 0.93. So the predicted retention probability for this student would be 93%.

In Excel, if the predicted retention score were the first entry in column A, the formula below could be used to calculate the predicted probability of retention:

= EXP(A1)/(1+EXP(A1))

In this way, predicted retention probabilities can be calculated for new students using the regression weights in Section 4.

Please feel free to contact the ACES team at aces-collegeboard@norc.org or 800-439-8309 for assistance in interpreting this report and making use of the information included.