

UNIVERSITY OF NORTH CAROLINA SYSTEM - DATA PROPOSAL

PREPARED BY UPSWING, 2016

Scope: Upswing Data and Retention Services

Upswing is dedicated to improving student retention through partnerships for online tutoring and data analysis services. Although data report cards and preliminary reports on factors related to retention can be produced by Upswing without access to any owned information by the school, a more robust and prescriptive set of analysis is created when Upswing and the school enter into a data-sharing partnership as well. This proposal outlines the information and retention metrics that the UNC system could receive from Upswing on a continuous basis after giving Upswing access to key internal data points.

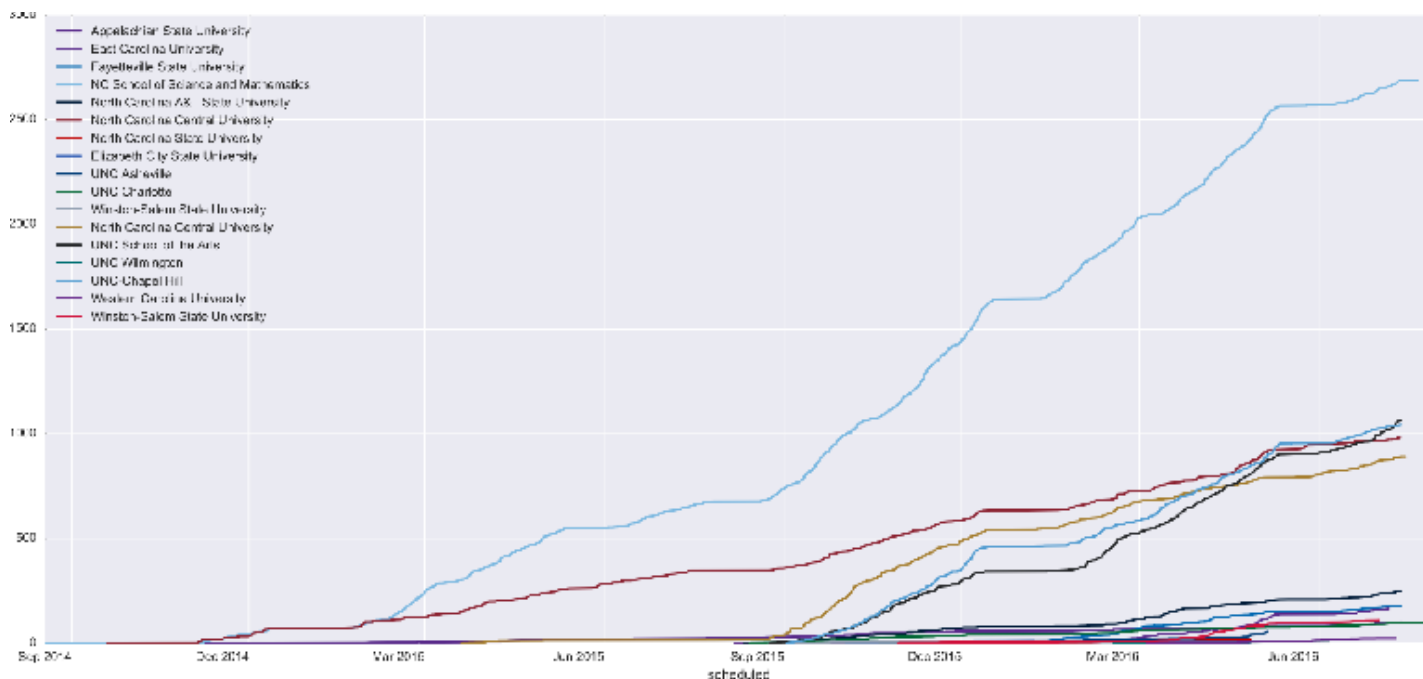
Upswing is committed to academic data integrity and security by strictly following guidelines for FERPA law and database management best practices.

All datasets and graphs that reference UNC-provided data in this proposal are based on hypothetical estimates unless otherwise specified.

Section 1: Basic Data Reporting

Without requiring additional data access from colleges and universities, Upswing can provide detailed usage and administrative statistics on a regular basis through timely report cards and a real-time admin dashboard.

Session counts and cumulative hours of usage represent the most basic insights available to school administrators. Upswing can provide usage comparisons across schools in the UNC system or in comparison to any selection of Upswing partners.



Key insight: Schools can track how many hours from their contract are being used to ensure students are aware of the service and are utilizing the expected number of sessions. Comparing usage over time allows the identification of usage seasonality, and how usage rate compares (in unique numbers, sessions held, average duration, etc) to other schools of similar size or across the system.

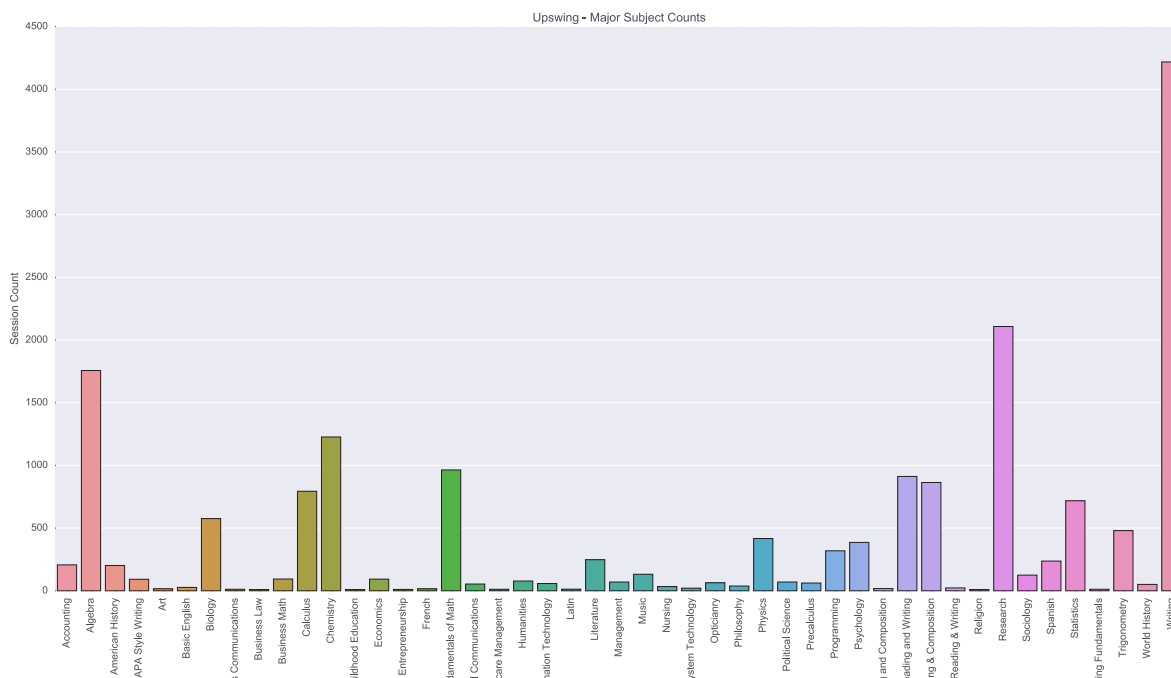
Individual students' usages are also tracked to identify those requiring the most help from Upswing. Upswing coaches are thoroughly trained to develop independent learning skills, so a student who requests a greater-than-average number of sessions can be flagged as needing additional academic support from administrators. The individuals with the highest number of abandoned sessions are also tracked to identify students or coaches who could benefit from greater oversight.

Power User	Total Sessions	Total Minutes
Jane Doe	17	988
Luke Ocean	24	925
Curi Georgia	19	814
Erick Lime	20	700
Iowa Jones	31	644

Abandoned Sessions Users	Abandoned Sessions	Late Sessions
John Smith	9	11
David Vador	5	6
Curi Georgia	5	4
Jill Jettison	4	11
Clark Kent	4	9

Key insight: Behavior and usage from individual coaches and students are tracked to give administrators insight on which students may be struggling and need to be placed on early alert for attrition. By identifying these areas of concern early, schools can create a timely response.

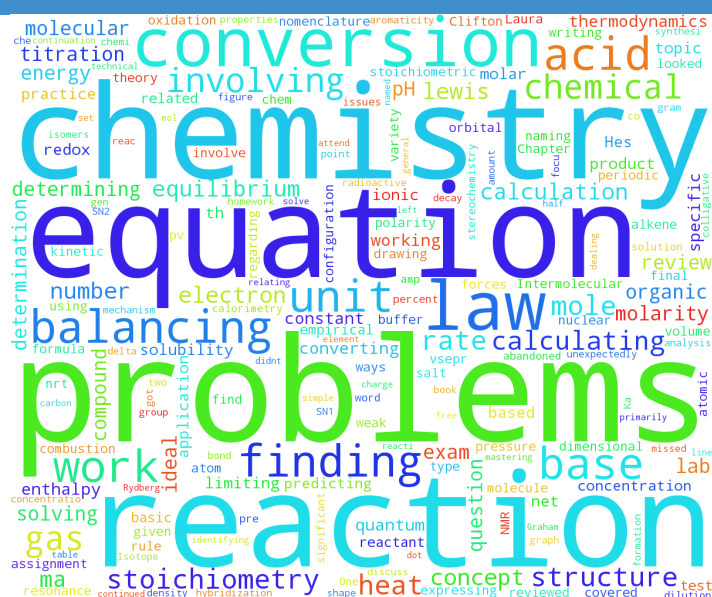
In addition to the amount of time spent on the online tutoring platform by individuals, data is also provided on what subjects and classes are most frequently requested. This data can be aggregated for all Upswing users across all partners, or broken down by schools or individual student data provided by schools with CIP codes to allow for direct comparison.



Key insight: By recognizing patterns in the subjects students seem to have the most difficulty in, school leaders can focus additional outreach on students in those classes.

Coaches provide feedback on students that can be mined with natural language processing. The word frequencies include “subject focus areas” of topics within the subject that the student sought help and “subject misunderstanding comments” which record what topics the student is having difficulty in understanding.

The example on the following page considers coach feedback from chemistry sessions and shows that students in this school had the most difficulty with Lewis Acid and Base Reactions, chemical equation balancing, and pH-pOH calculation.

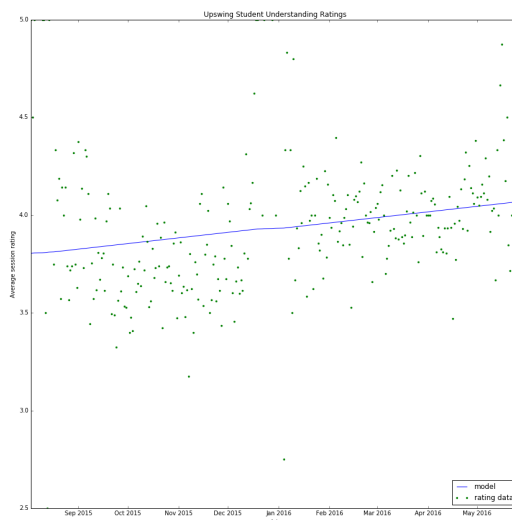


Key insight: Administrators can view these detailed analyses to determine areas of difficulty by subject or students. This feedback can be provided back to faculty members for additional in-class instruction.

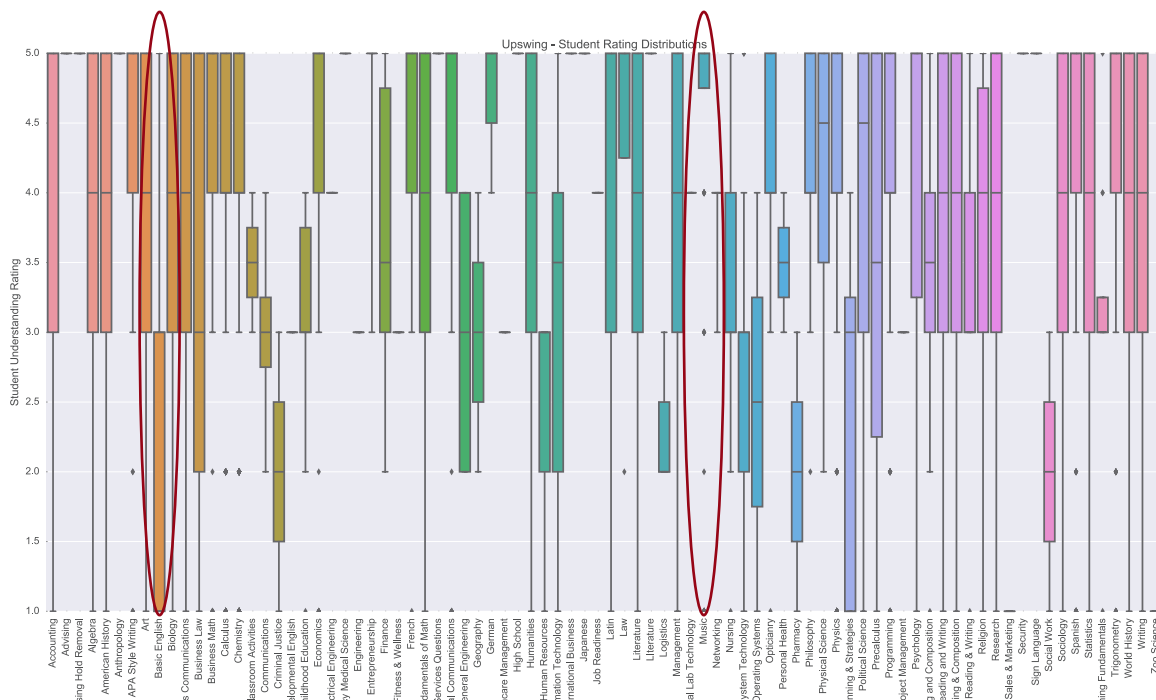
With the online tutoring platform, schools expect to see improvement in student understanding throughout the academic year. The efficacy of Upswing tutoring is tracked with the metric “student understanding ratings” as provided by coach feedback every time a session concludes. The correlation between tutoring hours and student understanding can be analyzed for individual or aggregate students, subjects, or schools.

The following graph shows daily average student understanding ratings for students that have scheduled more than 5 sessions. Given a p-value < 0.01 and a dataset exceeding 10,000 sessions, the increase in student understanding over the academic year is statistically significant. The average student understanding increased from 3.7 to 4.1 out of 5.

Rating	Meaning
1	Poor
2	Fair
3	Average
4	Very Good
5	Excellent

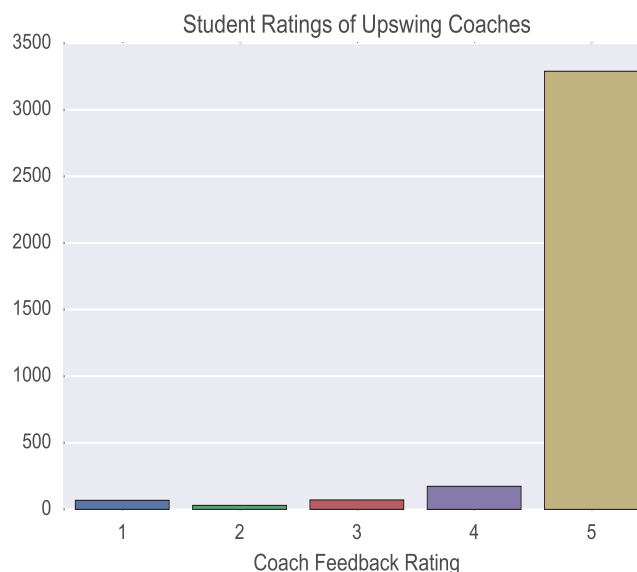


Student understanding ratings can be further broken down by subjects. See Appendix A: Box Plots for a detailed explanation. The highest and lowest understanding by subjects are circled below - Basic English for low, Music for high.



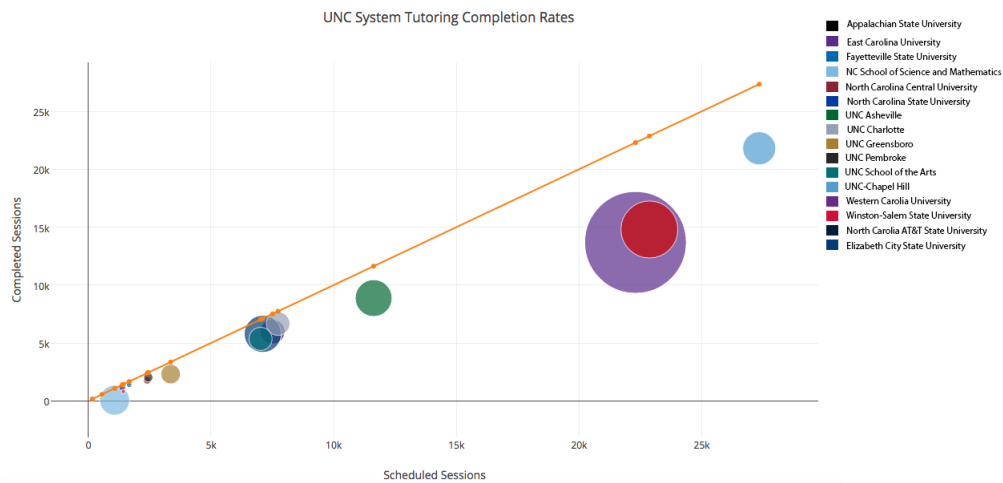
Key insight: School leaders can use this data to check the efficacy of their online tutoring platform and identify key areas by subject, school, or student that may be of concern.

Just as tracking student understanding provides important insights, reviews by students for their coaches are also monitored. These ratings and reviews are available for students to view when selecting a coach, and are also sent to administrators for oversight on Upswing coaches and the school's own tutors. Ratings by aggregate or individual coaches, subjects, and schools can also be generated. With Upswing data to date, the average coach rating is 4.81 out of 5, with a median of 5.

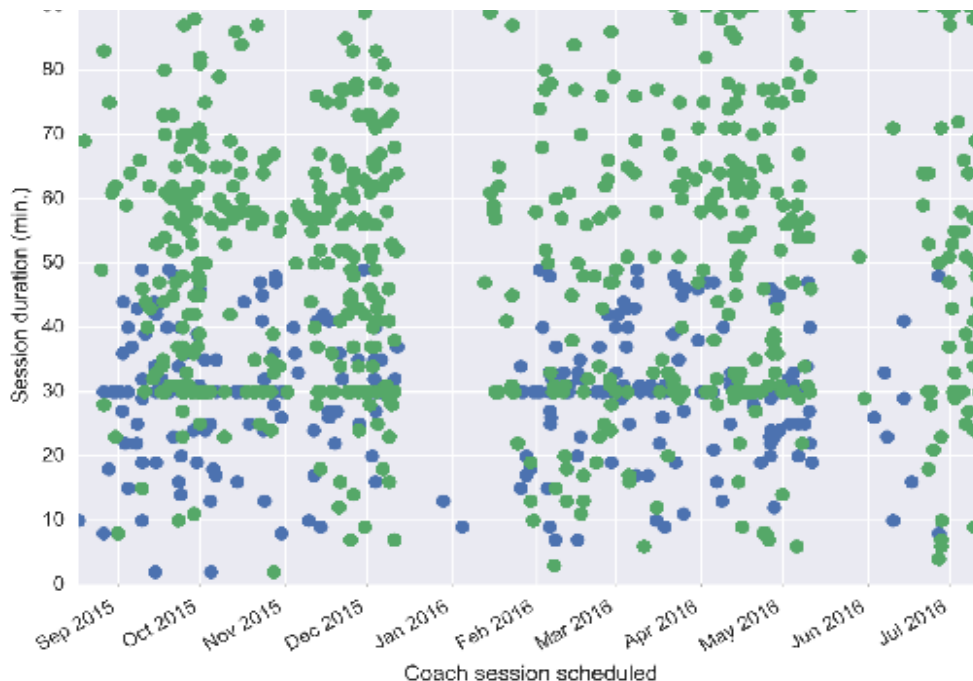


Key insight: Coach performance is monitored based on student ratings to quickly identify performance risk factors.

Any usage indicators can be compared across schools as well. In this example, the number of completed sessions to the number of scheduled sessions are graphed by schools. The vertical distance from the diagonal line indicates the level of abandoned or cancelled sessions, and the size of the circle represents the total volume.



Any binomial variable can be isolated for further analysis by comparing sessions where the variable is true to sessions where the variable does not exist. As an example of this capability, the graph below highlights sessions that occurred with students taking remedial courses (green) compared to sessions with students who are not taking remedial courses (blue). It can be observed that the remedial students are far more likely to have session go above 50 minutes in duration.



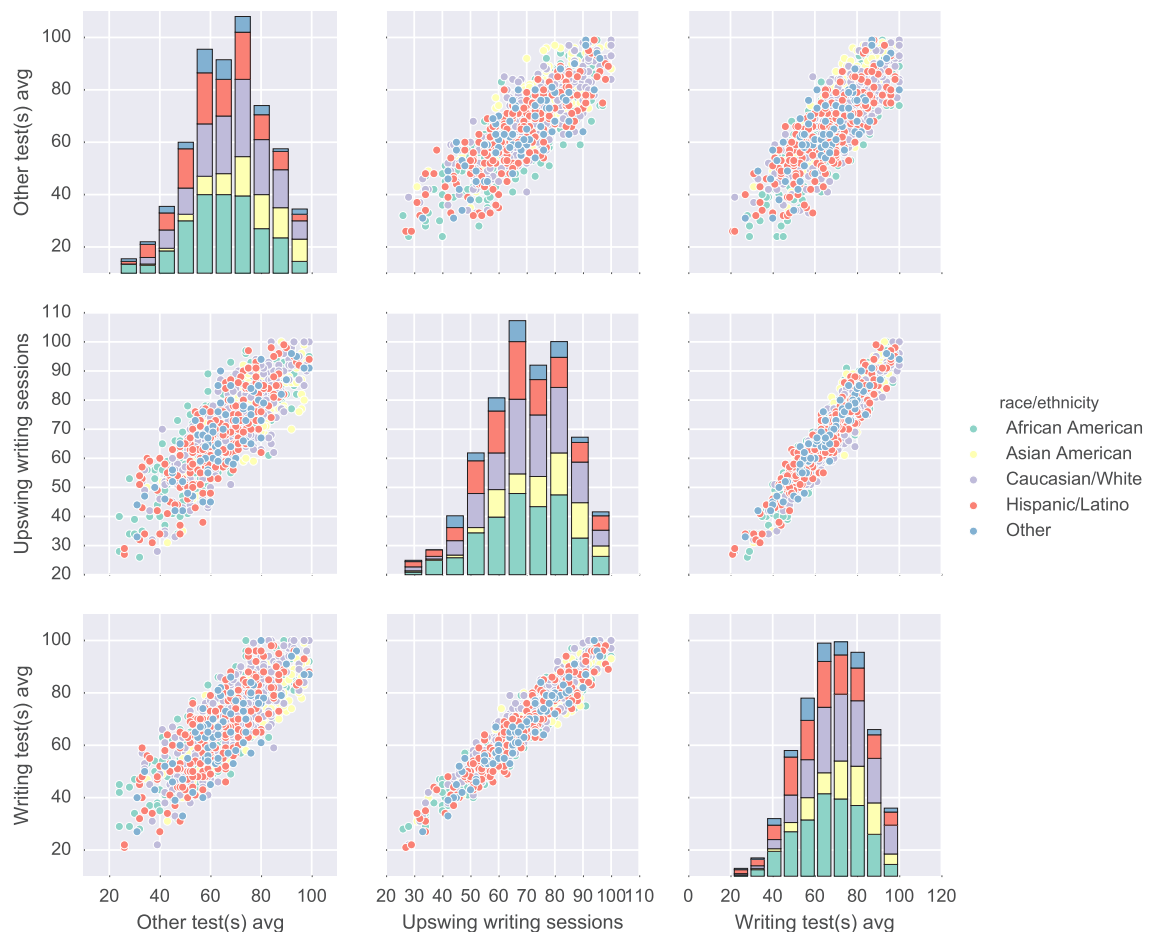
Key insight: Areas of interest across schools can be tracked for level of student involvement and usage.

Section 2: Advanced Retention Analytics

The following sample data presents the advanced retention reporting capabilities Upswing can provide upon delivery of UNC-specific information. Although the personal information of the students is encouraged to be used to filter data in the first section for additional research and insights, it is here that the data will be most useful for an overall retention strategy. These datasets will be used to analyze performance in terms of key risk factors for attrition. As enrollment figures are provided over time, the efficacy of retention strategies can be proven.

Upswing usage can be correlated with student academic performance for analogous subject-matters. These correlations can be further filtered by demographic data provided by each school.

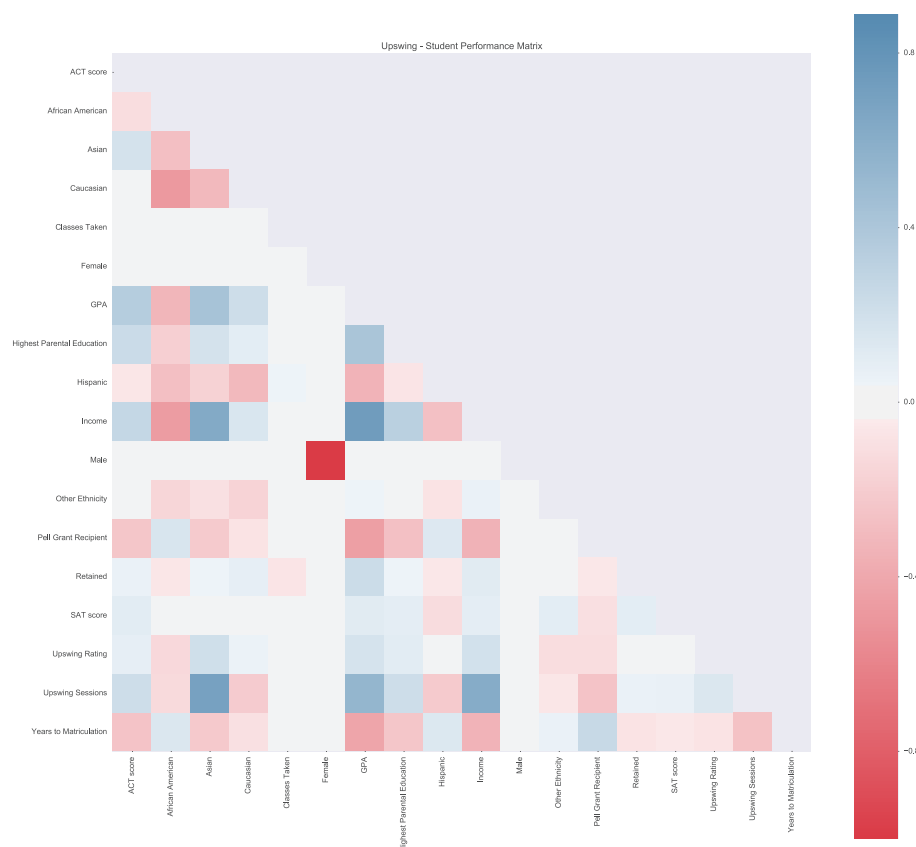
Each point on the following graph represents a single student. In this example, the number of Upswing writing sessions per student are plotted against the student's average writing test scores and non-writing test scores. The color of the dots represent a demographic grouping of interest - shown here as ethnicity. See Appendix B: Subject-Matter Correlation for a detailed explanation.



Key insight: School leaders can visualize academic performance across dimensions of Upswing usage, subject performance, and demographics. The areas of stronger correlation are areas to address or consider in a retention strategy, and can be observed over time to look for improvement.

In addition, correlations beyond academic performance can be analyzed for retention factors. Pertinent variables provided by both Upswing and the school can be examined for likelihood of correlation, pointing administrators to addressable factors that could affect retention.

The following correlation plot demonstrates a sample of variables and their relationships to each other. The intersection of factors that are positively correlated are blue and negatively correlated intersections are red. The darker the color, the stronger the positive or negative correlation. Based on this example dataset, the most positive correlations to GPA are income and number of Upswing sessions. See Appendix C: Correlation Matrix for further detail.



Key insight: These correlations tables answer the most important questions for creating a retention strategy by showing the predicted effect on variables of interest. This can prove the efficacy of learning center initiatives and identify non-intuitive student retention factors.

For example, some of the questions that could be answered with the correlation table are:

- Do African American students using Upswing show greater improvement in math relative to African American students who don't use Upswing?
- Do students on financial aid utilize Upswing more than those who do not receive aid?
- Do students who have scheduled more than 5 Upswing sessions have a higher GPA than students who have visited the on-campus learning center more than 5 times?

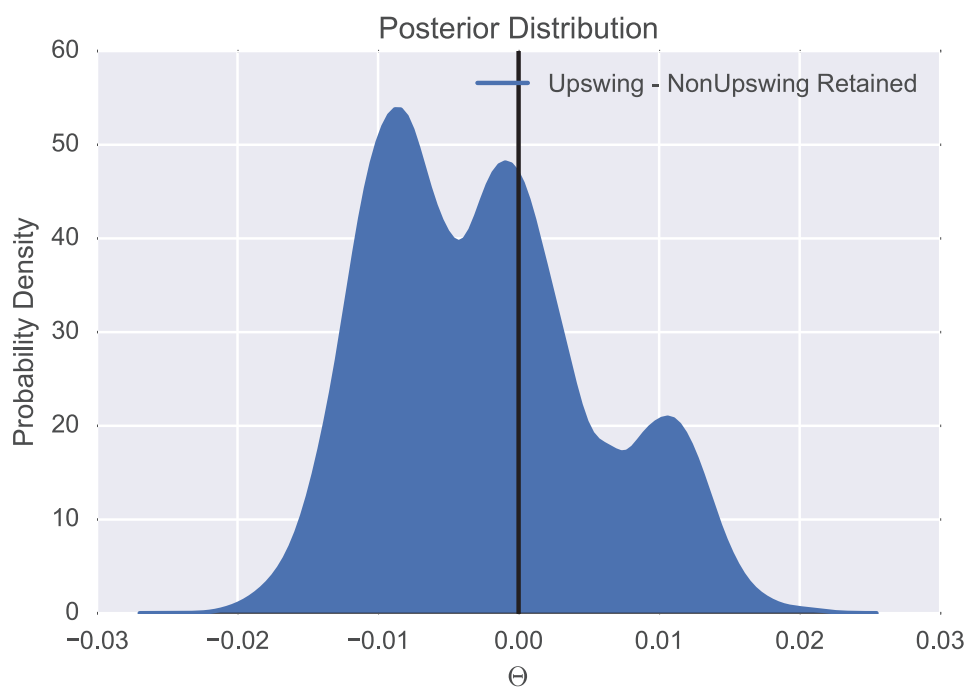
Section 3: Predictive Retention Analytics


Ultimately, the goal of Upswing is to predict the extent to which the online tutoring and data analysis services improve retention rates, and to identify the student populations where the methodology is - and is not - working. The correlation table in the previous section is used as a foundation for a hierarchical model where prospective weights are placed on each variable based on its effect on the dependent variable of interest (GPA in this example).

For example, an African American, female, Pell grant recipient, with X number of Upswing sessions would be assigned a certain probability of being retained based on past and present datasets. The model makes a prediction whether that student is retained, and continually corrects itself based on how far new, incoming data deviates from the prediction.

In the beginning when historical data is sparse however, school leaders still want to receive face-value predictions about whether the Upswing student population has a statistically significant difference in retention relative to the whole school population. To create such a model, a Bayesian binomial proportionality test is used to make a confidence interval. The confidence interval states the probability that the difference in retention rates between the two populations (Upswing users vs. non-Upswing users) is statistically significant given the sample size.

See Appendix D: Probability Distribution for additional detail.



 Key insight: This distribution shows the probability that Upswing had a positive change on retention rates.

Appendix A: Box Plots

The box plot graph isolates student understanding ratings by specific subjects that can be customized by school or across CIP classifications. The colored boxes represent where the majority of students fall in terms of their understanding: the 25th percentile to the 75th percentile. The line in the middle of each colored box represents the median value.

The narrow lines extending up and down from the boxes show the students who fall below the 25th percentile in understanding and above the 75th percentile. By looking at where the majority of students are in their understanding by subjects, the box plot shows trends in which subjects have higher or lower understanding and how evenly the distribution is spread.

Appendix B: Subject-Matter Correlation

This correlation table filters academic performance between subjects by any demographic or other factor of choice. The graph at each intersection represents the correlation between the x- and y-axis. A tighter line indicates that the subjects are closely related, while a wider distribution indicates that the subjects are more loosely related.

The bar graphs across the diagonal line display the demographic factors within the intersections. By examining the relative size of the sections, conclusions can be made about which group makes up the majority of the high-performance or low-performance students relative to the other demographic groups in each subject.

Appendix C: Correlation Matrix

The correlation matrix plots any factor requested for further study against each other to view how closely the factors are correlated. A strong positive correlation (dark blue) indicates that the two factors move together - i.e., if one increases, the other increases in a similar proportion. A strong negative correlation (dark red) indicates that the two factors move in opposite directions - i.e., if one increases, the other will decrease, or they are binomial opposites. The matrix is created by statistically calculating the exact coefficient of correlation between each possible combination of factors, and can be customized based on the data provided for individual or aggregate schools, students, and subjects to answer the specific questions each school requests. To examine the effects to one factor, read up and down or across to see every intersection to that factor and determine which ones are the strongest.

Appendix D: Probability Distribution

This distribution examines the difference in retention percentage of Upswing users compared to non-Upswing users to determine if the difference is statistically significant and can be directly correlated to the use of Upswing instead of some other factor (i.e., pure chance). The area under the graph on the left side of the vertical line is the probability that there was a positive effect relative to the control group. With historical data and machine learning, the distribution will be able to move more to the left side as the platform becomes better at making predictions and identifying which students can benefit from the platform.

Appendix E: School Data Request

For the highest quality of data and reporting capabilities, Upswing would request historical and present values by each school for specified criteria and factors that the schools wish to have examined. Moving forward, receiving updated values at the end of each academic year allows the Upswing data platform to learn from itself, make better predictions, and provide more complete reports and analysis.

For example, some of the data that could be requested for further study are:

- Enrollment numbers
- Student demographic information
- Student admissions information (for example, SAT scores)
- Student quiz and exam scores
- Student GPA
- Students completing remedial coursework
- Counts of visits to on-campus learning centers

Contact Us

For more information about the data and analysis provided by Upswing, contact your sales representative or email speak@upswing.io