EAB

PRE-WORK

Retention Modeling Cohort: Session 1 Pre-Work Packet

July 12, 2023 | 1:00 p.m. – 2:30 p.m. Eastern Time

EAB Experiences RAPID INSIGHT

Hello and welcome to all our Rapid Insight Cohortians!

We are excited to kick off the beginning of our three-part Retention Modeling Cohort series. This pre-work packet is designed to help you gather the materials you will need for our first session together. Note that since the upcoming session is interactive, these *exercises are required so that you have content to work on during our working session.* You won't be graded on this assignment (we wouldn't be so cruel), however each of the following activities are necessary tools to help you **design your model**, **map out potential predictors, and gather relevant data**.

Instructions for each activity can be found prior to the section. Please don't hesitate to reach out if you get stuck or have any questions! You can always reach us at RI-Support@eab.com.

We can't wait to get started with you next month!

Sincerely,
James Cousins & Lily Brennan

Activity 1: Design Your Model

Instructions: Consider the following questions and determine the goal of your modeling.

1. What target population(s) do you hope to analyze? (e.g., Undergraduate, Degree-seeking, First Time Full Time, etc.)

2. What retention question(s) are you looking to answer? (e.g., Fall to Spring, Fall to Fall, First-year to Second-year, etc.)

3. How does your institution identify if a student has been or has not been retained? If you are unsure, our recommendation is to consider them retained once they have registered for the upcoming term.

Activity 2: Map Out Potential Predictors

Instructions: Below is a list of variables we suggest including when beginning a retention analysis. Not all of these variables are a requirement, and not all of these variables will be included in the final analysis. We do however suggest beginning with a "kitchen sink" approach – include everything you have, and let the model identify the valuable contributors.

You don't need to fill in the **Source System** and **Table Name** for *every* variable in this list. The goal is for you to be familiar with what information you have available to you and where it is stored so that you can find it during our first session. If using this table will be helpful for you to remember where your information is stored, then feel free to take notes and use this as reference sheet during session 1!

Suggested Higher Education Modeling Variables

Retention Model

Student – Demographic

Suggested Variable	Source System	Table Name
Identifier		
Date of birth or age at start of first term		
IPEDS Citizenship/Ethnicity/Race value		
Ethnicity		
Race (all categories, not restricted to IPEDS values)		
First generation status		
City of permanent residence		
Zip Code of permanent residence		
Student Residential Status (living on campus, off campus, commuter)		
Student employed flag		
Fee reduction/waiver requested		
Military background?		
Have children?		
Multiple flag fields identifying participation in various campus activities, campus life		

Institution - Academic Program Related

Suggested Variable	Source System	Table Name
Term identifier		
Starting student type (new, transfer)		
Number of credits registered for		
Current Student type code/desc (new, transfer, returning, continuing)		
Program level (Associate vs. Bachelors)		
Program type (online, face-to-face, hybrid)		
Degree sought type (assoc, bach, certificate) & subtype (arts, science), possibly as AA, AS, BA, BS, etc.		
Program code/desc		
Major code/desc		
Concentration code/desc		
Academic unit code/desc		
Department unit code/desc		
Transition course (how to go to college course) (y/n or actual if multiple types)		
Taking developmental courses?		
Previous semester Term GPA (if predicting for current students)		
Period GPA to period GPA for each period enrolled		
Hours attempted and completed period to period		
Degree hours completed		
Degree hours attempted		
Program changes		
Major change count		
Academic unit change Y/N (i.e., when formerly psych in Arts Sciences, now Economics in		
Business School)		
Retention Flag (0/1 telling us if the student was lost or not)		

Student - Academic / Online History

Suggested Variable	Source System	Table Name
High School class size		
High School Rank		
High School Rank in Class		
Advanced Placement (AP courses) articulated by count and hours		
Hours transferred from another institution		
Student support services (enrollment/admission counseling) available?		
Requested tutor, Assigned tutor		
Visited writing center		
Visited math center		

Financial Aid

Suggested Variable	Source System	Table Name
Receiving financial aid		
FAFSA yes/no		
EFC		
Pell eligibility (indicator of socio-economic status); Flag Pell awarded		
AGI Student; AGI Parent		
Orientation (session # if multiple)		
Flag parent/guardian attended orientation		
Flag application fee waived		
Any admissions index used for admit decision		
Gift Aid Y/N; Gift Aid Sum		
Scholarship amount offered original		
Loans amount		
Total financial aid amount offered original		
Awards by Federal, State, Private		
Workstudy Y/N		

Course – Related

Suggested Variable	Source System	Table Name
Total courses		
Total hours		
Total High DFW courses; Percent High DFW load		
Course registrations self or administrative		
Registered for classes online or in person?		
Course hours by instruction method		
Amount of text vs. interactive content in the course		
Lessons with immediate feedback?		
Teacher interaction with students		
Peer-to-peer forum?		
Courses at the 100/1000, 200/2000, 300/3000 level		
Ever attended a class on-campus		
Whether online learner plans to attend both online and on-campus classes		
Number of previous online courses taken		
Ever attended a class on-campus?		
Plans to attend both online and on-campus classes?		

Activity 3: Gather Relevant Data

Instructions: Consider the following questions and follow the steps to create data extracts.

Historical Data – When conducting a predictive analysis, it is important to use historical data so that the predictions are based on the significant patterns and behaviors of prior populations. We recommend gathering **3-5 years of historical data** to be included in your analysis.

1. How many years of historical student retention data do you have available?

2. Consider what the most recent completed cycle is and count back 3-5 years from there. (e.g., If you are interested in Fall to Spring retention, the most recent completed cycle is Fall 2022 to Spring 2023. So ideally include data from at least Fall 2020 to Spring 2023.)

   Which years of historical student retention data will you include in your analysis?

3. You will also need to include data from the upcoming (incomplete) cycle in which you want to predict the outcome of retention for. (e.g., If you are interested in Fall to Spring retention, you should include data on students enrolled in Fall 2023 so that we can predict their likelihood of retaining to Spring 2024).

   Which years/terms of students will you score for retention predictions?

Data Extracts – Now that you have identified what data you will include in your analysis, it will be useful to create data extracts for each table containing information you need to be used during our first session.

Note: Not all variables need to be in the same data extract at this phase (e.g. you may bring one data extract containing student demographic information from 2020-2023 and another extract containing course registration information from 2020-2023). You may include as many extracts as necessary!

1. Create data extracts for each table including relevant historical information for the time period specified in the last section.
2. Save data extracts as .**CSV** or **.XLSX** file types.
3. Make sure you have Construct and Predict installed on your computer.
4. Reach out to RI-Support@eab.com if you have any questions!