

# Noncognitive Skills & Traits Data in Student-level Predictive Analytics

CUPA 2022

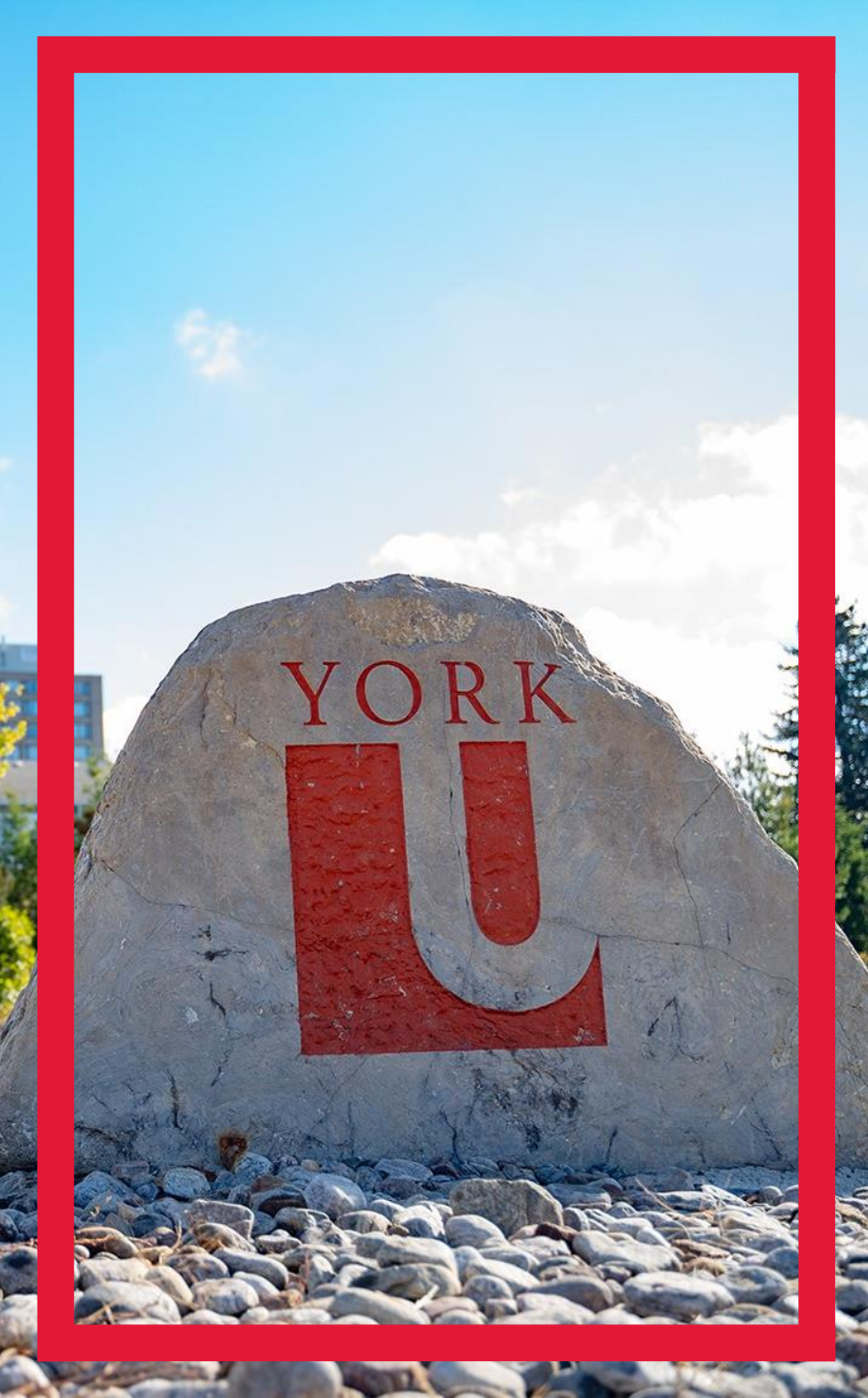
JONATHAN SCHNEIDER<sup>1</sup>, VIDUR KALIVE<sup>2</sup>, AND MARK CONRAD<sup>1</sup>

<sup>1</sup>OFFICE OF INSTITUTIONAL PLANNING AND ANALYSIS

<sup>2</sup>UNIVERSITY INFORMATION TECHNOLOGY

YORK 





## Outline

- What are noncognitive skills/traits/behaviors and why do they matter?
- Data about noncognitive skills/traits/behaviors
  - Student Self-Assessment
  - SAVY (York's Virtual Assistant)
- What changes when we leverage noncognitive datasets in predictive modelling?
  - Accuracy
  - Robustness to changing external factors
  - Interpretability



**What Do We Mean  
by Noncognitive  
Skills & Traits and  
Why Do They  
Matter?**



# What do we mean by noncognitive skills and traits?

- Patterns of thought, feelings and behaviours
- Behaviors, attitudes, beliefs, and social & emotional skills
  
- Many noncognitive skills and traits associated with academic performance...
  - Mindsets: sense of belonging, belief in one's abilities, belief in the value of academic work
  - Perseverance and Coping: grit, self-discipline, self-help, help-seeking
  - Social Skills: empathy, cooperation
  - Learning Strategies: study skills, self-regulated learning
  - Academic Behaviors: going to class, participating in class, doing homework, organizing materials and time  
*(not even close to an exhaustive list)*
  
- Noncognitive skills and traits develop over time
- Like habits, they are learned and can change

# Why do noncognitive skills and traits matter?

- As already mentioned, noncognitive skills and traits impact academic performance
- Paying attention to these skills and traits lets us...
  - Better understand individual students and the ways they approach and position themselves in relation to the academic experiences, how they make sense of and co-create those experiences
  - Notice behaviours that individual students are exhibiting and that typically lead to particular academic outcomes
  - Better understand exactly how to support an individual student
  - Help individual students understand their own academic experience and performance
- Paying attention to noncognitive skills and traits reinforces the idea that student-level analysis and outreach is a valuable approach
  - Net-new information
  - Weakly associated with cohort-based variables
- The idea is to augment the 'cohort-level' data (e.g. domestic/international; gender; 101s/105s) and get down to the level of individual students: their aspirations, goals, skills, behaviors.

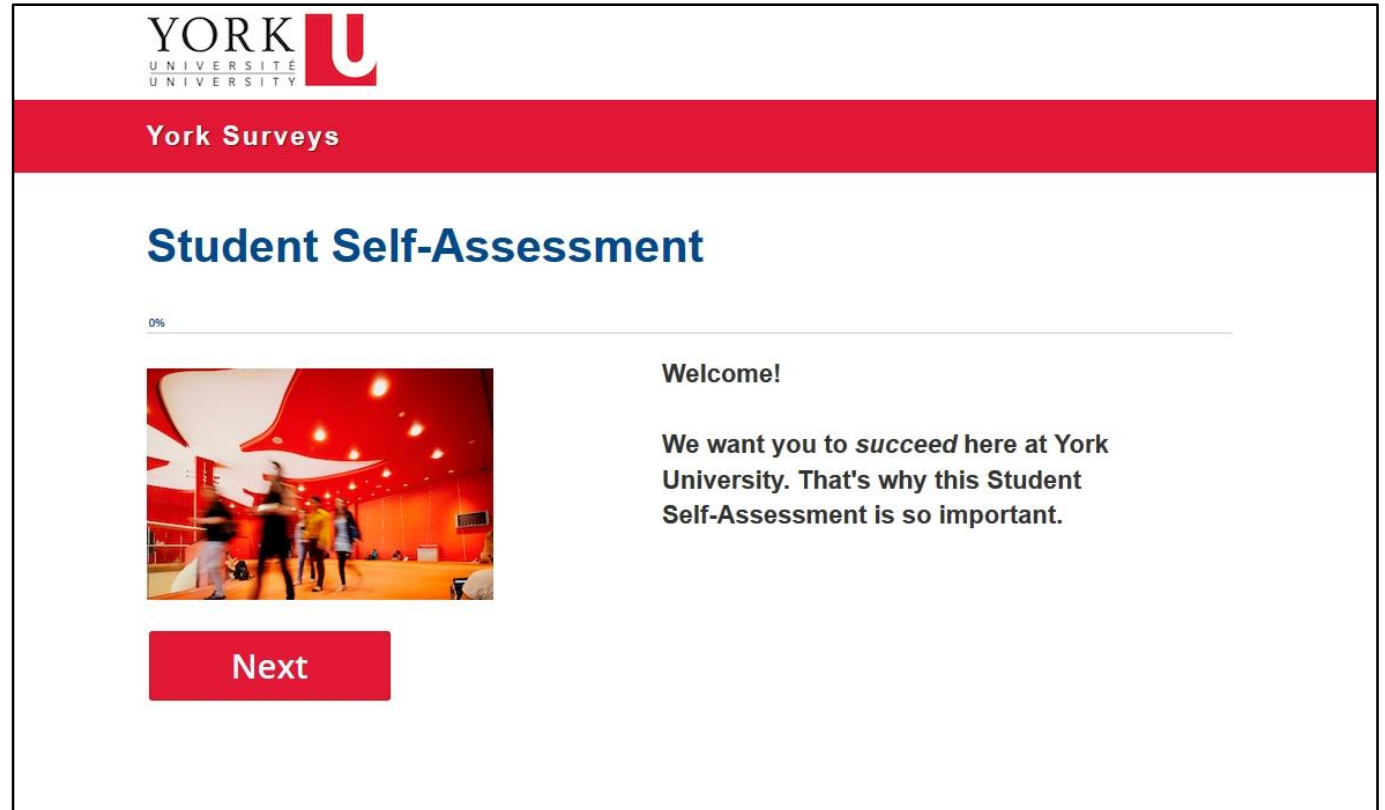
# Data About Noncognitive Skills and Traits



## Self-Report: Student Self-Assessment

An online survey instrument 'sent out' to all new, first-year undergraduate students.

We use it to gather insights about our students' non-cognitive skills and traits.






## Self-Report: Student Self-Assessment

## Insight is gained about our students'...

- internal & external motivations for attending university
- academic & career goal clarity
- academic self-efficacy
- coping skills
  - Personal resourcefulness (self-help)
  - Social resourcefulness (help-seeking)
  - Grit (persistence & passion for long-term goals)
- Uses existing measurement scales
- validated (except for the goal clarity scales)
- Also asks a small number of socio-demographic questions



**YORK**  
UNIVERSITY

York Surveys

## Student Self-Assessment

13%


### Part 1 of 4: Academic Goals and Motivations

How central were each of the following reasons to you?

	1 definitely not a central reason	2	3	4	5	6 definitely a central reason
Because of societal expectations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To help me make better decisions in life	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To access services available to students	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To delay finding a job	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Because higher education has intrinsic value	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To engage in thought-provoking discussions with peers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Because I have the intellectual ability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
To experience courses never offered in high school	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
For the general experience	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Back

Next



**YORK**  
UNIVERSITY

York Surveys

## Student Self-Assessment

46%

### Part 3 of 4: General Coping Skills...

How do you handle the following kinds of situations?

Indicate the degree to which each statement describes your thinking or behaviour. Some of the statements may be similar, but it is important that you read and try to answer each statement.

	1 extremely non-descriptive of me	2	3	4	5	6 extremely descriptive of me
When I do a boring job, I think about the less boring parts of the job and the reward that I will receive once I am finished.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I have to do something that makes me anxious, I try to visualize how I will overcome my anxiety while doing it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
By changing my way of thinking, I am often able to change my feelings about almost anything.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am feeling depressed, I try to think about pleasant events.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am faced with a difficult problem, I try to approach it in a systematic way.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When an unpleasant thought is bothering me, I try to think about something pleasant.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I am depressed, I try to keep myself busy with things I like.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I find it difficult to settle down and do a task, I look for ways to help me settle down.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Back

Next



# Actual Behaviours: Using SAVY



**SAVY**

Hi, how can I help you?

# About SAVY

## ➤ Who is SAVY?

- An AI virtual student assistant built using IBM Watson's Natural Language Processing.
- Trained with content to help students perform tasks related to their academic journey at York.

## ➤ Vast Knowledge Base

- SAVY covers over 500+ topics
- 1500+ conversational dialog flows
- Serves international and domestic undergraduate students across 10 schools at York, in both English and French

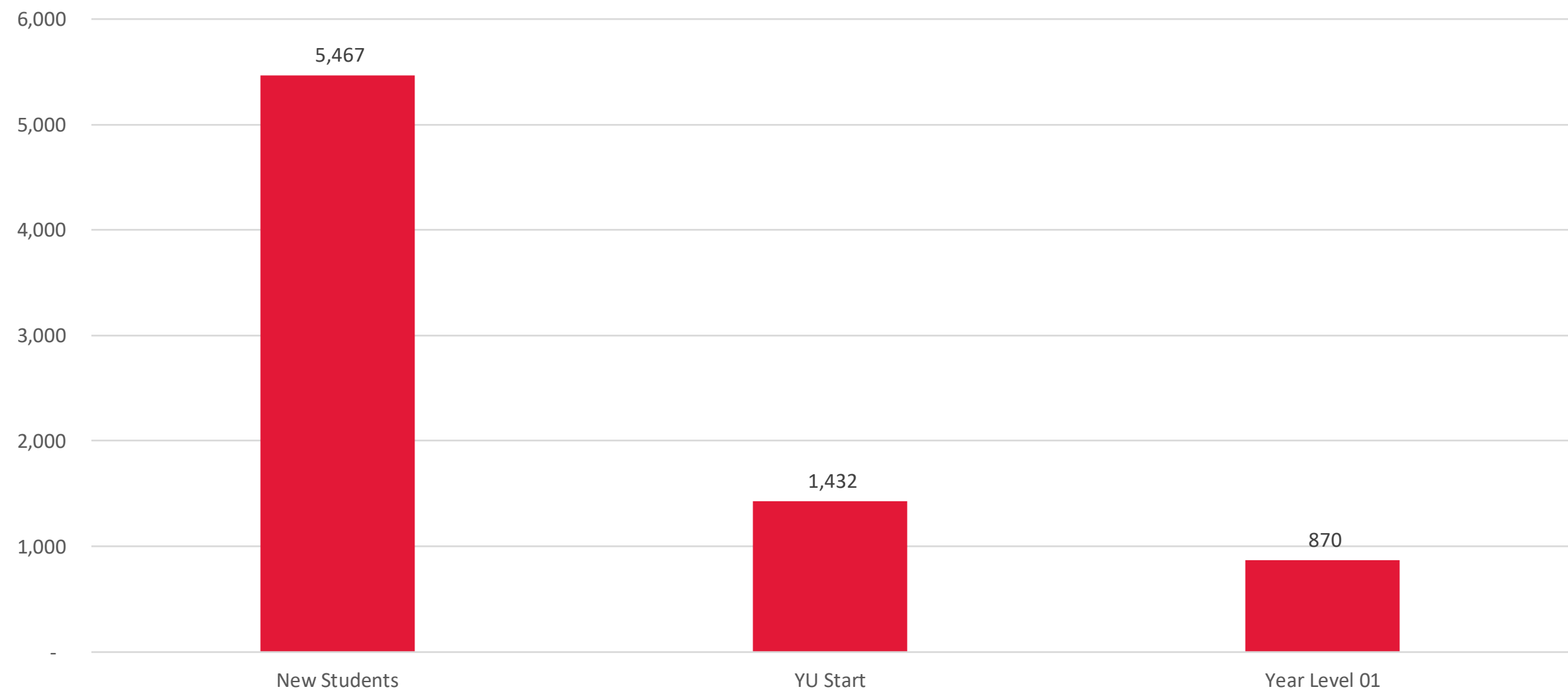
## ➤ What makes SAVY different?

- SAVY knows enough about individual students to provide them with more personalized content including reminders about upcoming deadlines and tasks relating specifically to them.



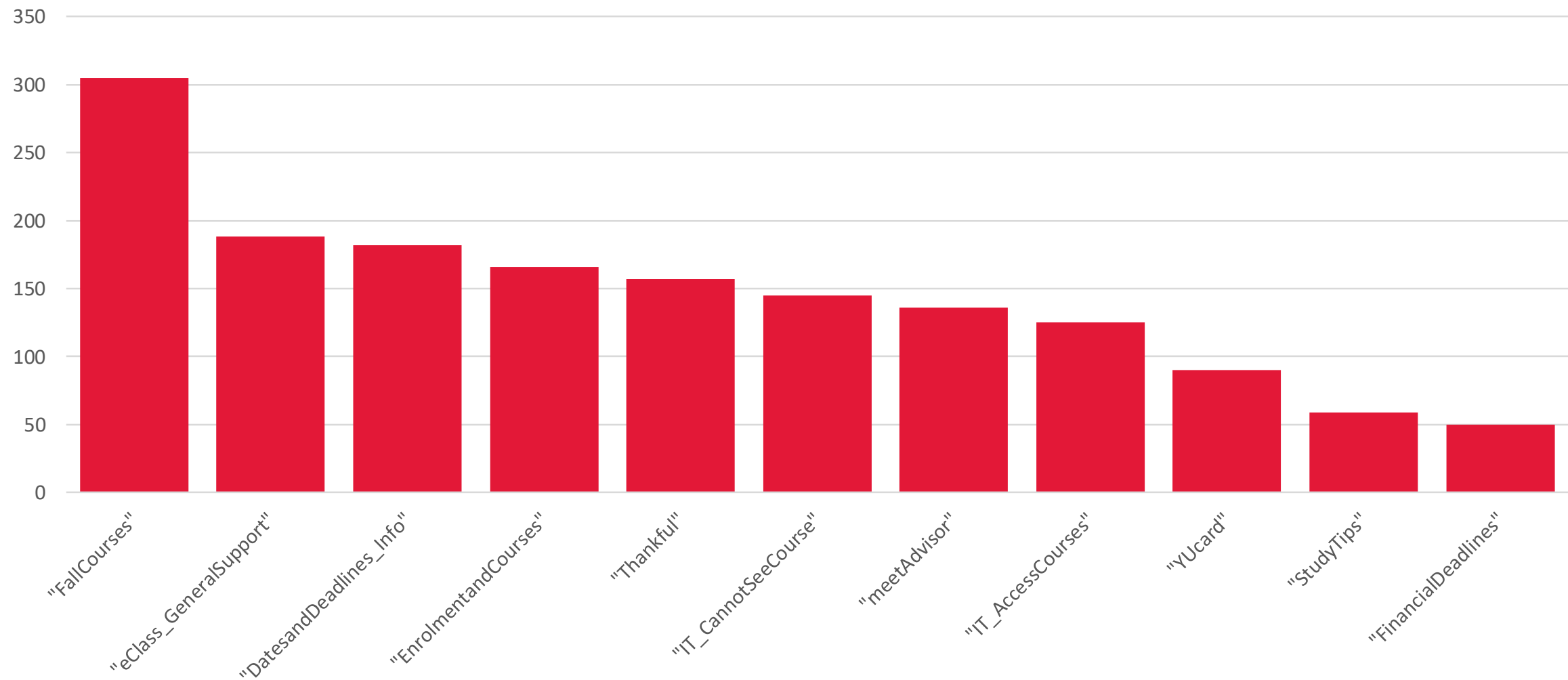
Hi, how can I help you?

# SAVY Numbers – Incoming + Year Level 1

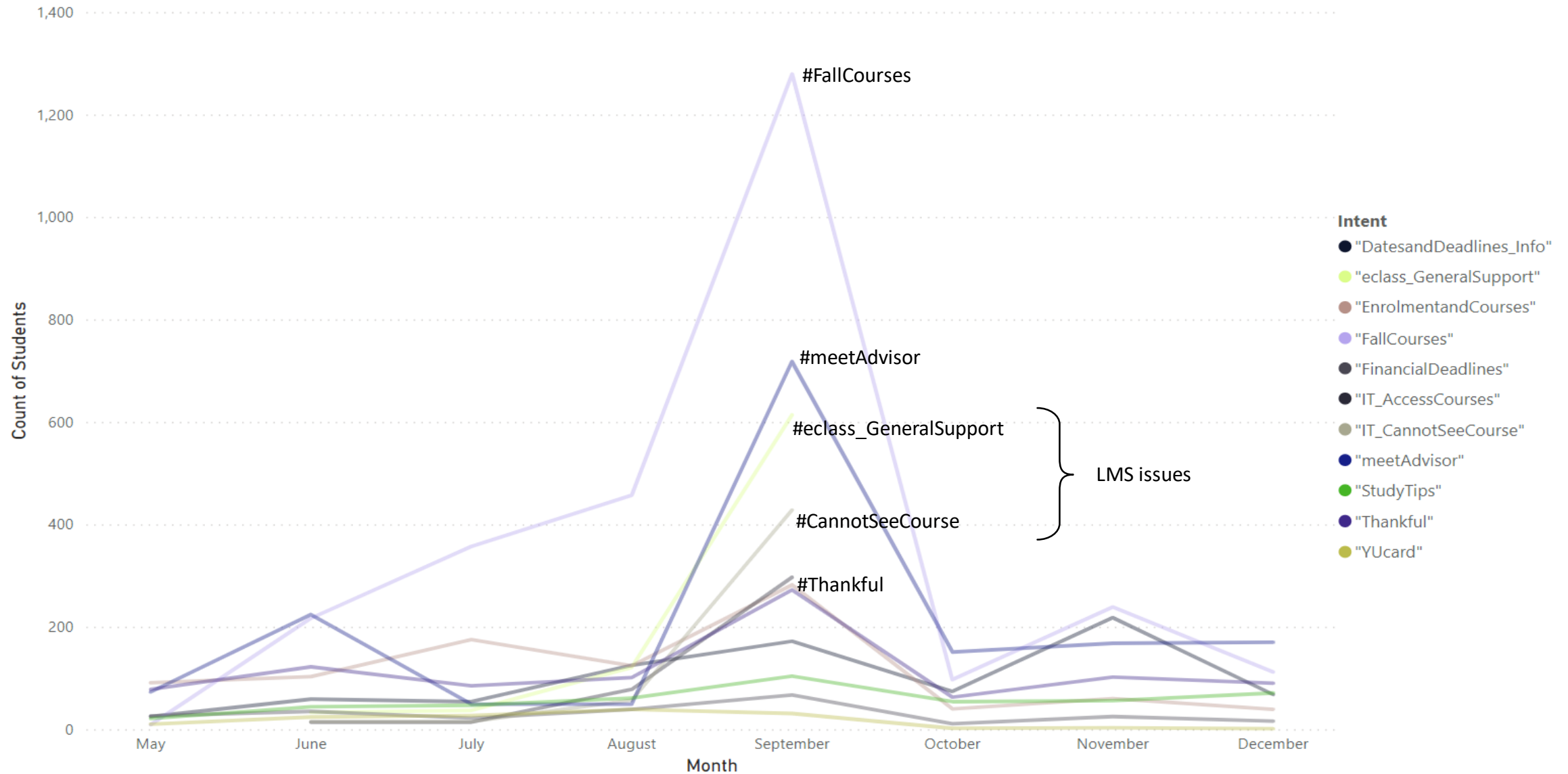




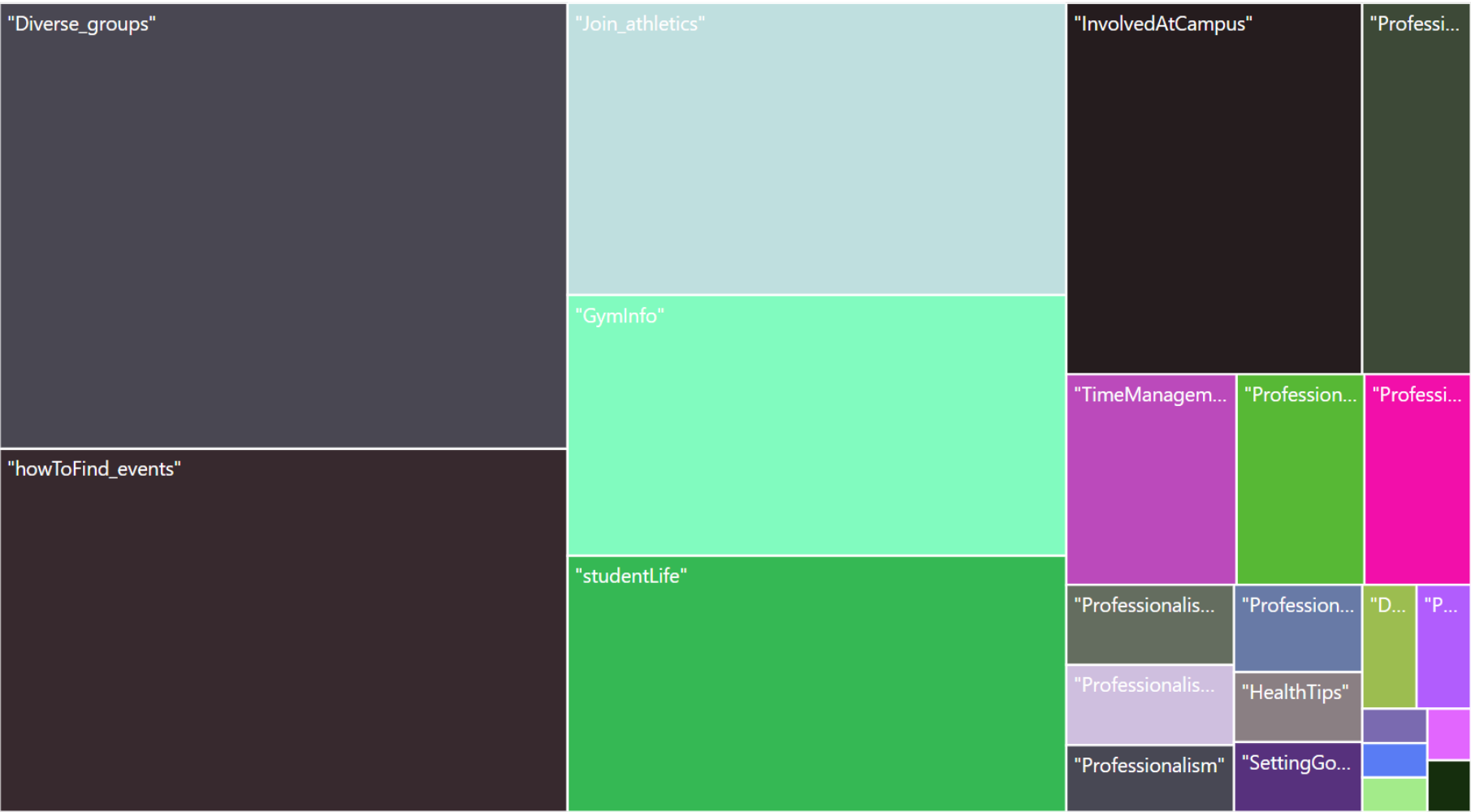
# SAVY Top Answers – Incoming (YU Start)



# SAVY – Temporal Dynamics



# Other Noncognitive data captured in SAVY



You. 7:42 AM  
How should I dress for an interview



Time to dress for success! Check out the Career Education & Development's [Want to Succeed On the Job?](#) info, which includes a dress-for-success section.

👜 Career Tip: Less is more when it comes to accessorizing for an interview or business event – keep jewelry, fragrances, scarves and cosmetics to a minimum.



# Leveraging Noncognitive Data to Increase Predictive Accuracy at the Student Level



# Making Predictions That Are Actionable

- Timeline must be 'actionable':
  - A given outcome has to be predicted with enough foresight that an intervention could be staged and has enough time to be successful.
- Model must be:
  - Accurate
  - Robust
  - Interpretable (for downstream decisions and accountability)

## How can we support students on their academic journey?

---

We want to be able to provide relevant supports as early as possible in a student's unique academic journey.

# Example Analysis: Identify Students At-Risk of Not Completing a Course

## TASK

- Build a model:
  - Cohort: all incoming students to York
- Actionability:
  - Use early arriving data [prior to Oct 1<sup>st</sup>]
  - Predicts course drops at a student level [occurring after Oct 1<sup>st</sup>]
- Accuracy and robustness:
  - Train on one year and predict the next.
- Use a technique suitable for the data:
  - Handles missing and incomplete data
  - Good predictive power
  - In this case a boosted tree ensemble model (XGBoost)

## DATA AVAILABLE

- Registrarial:
  - Admission grades (High School) when available
  - Demographic information (age, gender, citizenship)
  - Enrolment transactions (enrolment date, course switches/drops)
- Noncognitive skills, traits, behaviours:
  - Moodle: Student activity in LMS logs
  - Inspire: Student-Advisor interactions
  - **Self-Assessment**: Scores on grit, goal clarity, motivation, and resourcefulness measurement scales
  - **SAVY**: Student-SAVY conversations

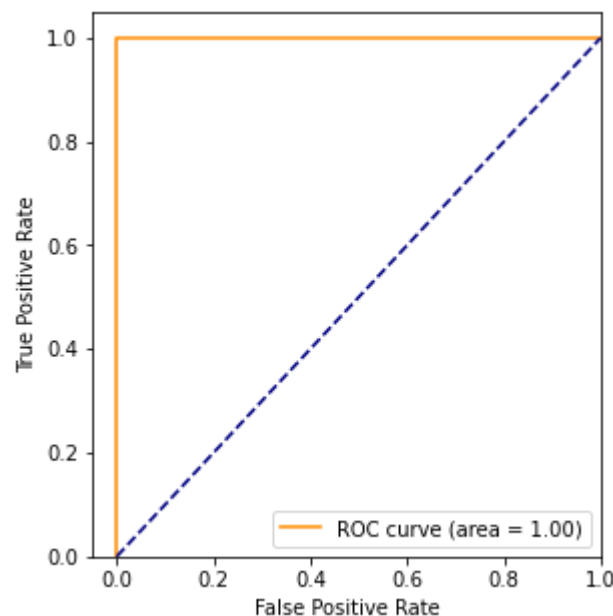


# Sidebar: Evaluating a Model with Unbalanced Data

## UNBALANCED DROP RATES

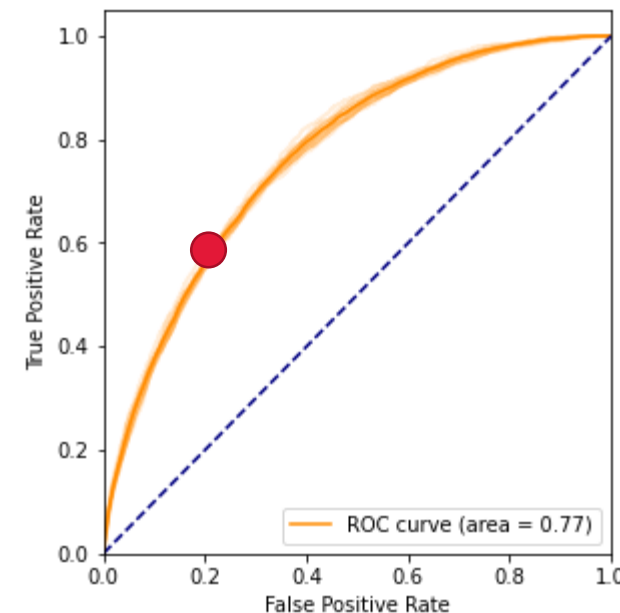
- Students are completing courses at a *much* higher rate than dropping them.
- A model that predicts no drops will do well (in terms of accuracy) on average.
- Receiver Operator Characteristic Curves shows the balance between the True Positive (TP) and False Positive (FP) rates.

## PERFECT PREDICTOR



- At any threshold, the TP rate is 100%, and the FP is 0%
- Blue dash indicates random predictions

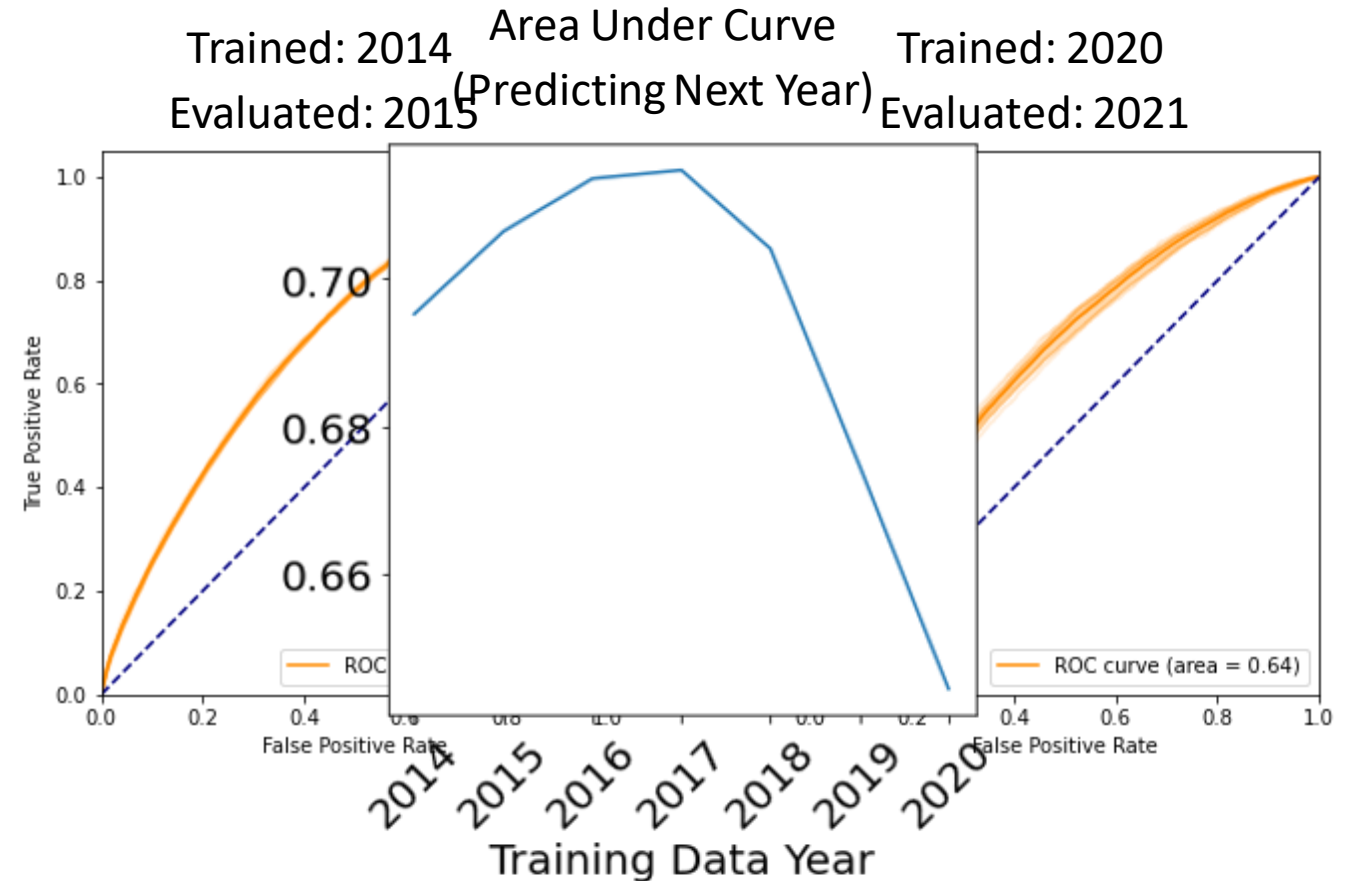
## “GOOD” PREDICTOR



- By altering the threshold, one can change the TP and FP trade-off
- At the indicated point:
  - TP  $\approx$  60%
  - FP  $\approx$  20%

# Using Only Registrarial Data (no Noncognitive Data)

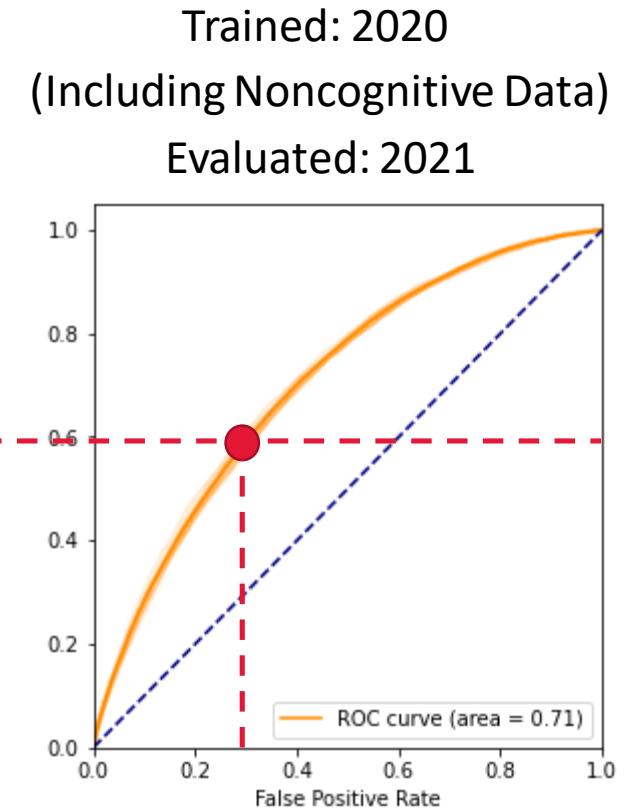
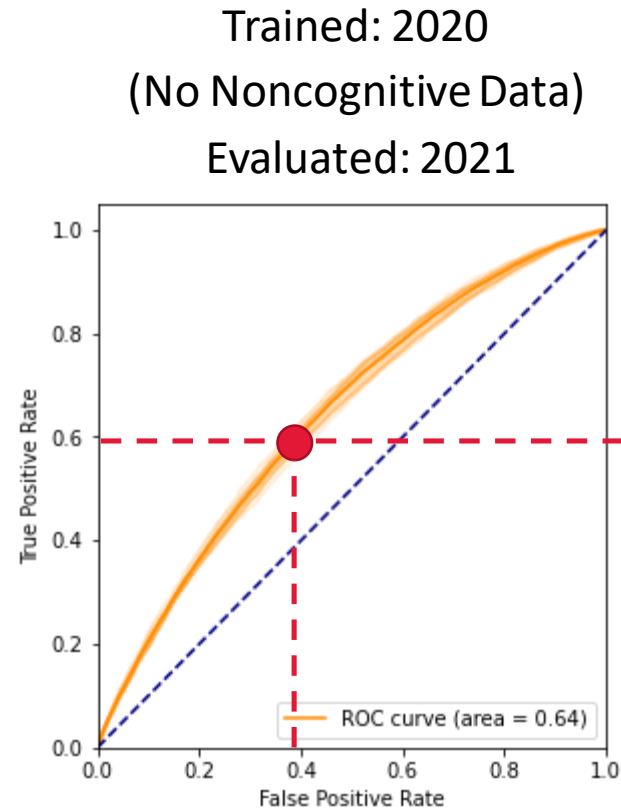
- Given the highly unbalanced dataset, we used an under-sampling approach (for the training data).
  - Under-sampling discards data and therefore all results presented are the average of 10 runs.
- Initially promising power, but inconsistent across years



... We need a way to increase year over year robustness

# Improving Predictive Performance with Noncognitive Data

- Goal is to increase overall accuracy and create a more robust model by adding other sources of data:
  - Non cognitive / Behavioral:
    - **Self assessment**
    - **SAVY interactions**
    - Advisor access
    - Moodle activity
- Also tested:
  - Training on previous years (with diminishing weights on previous years) – overall improvement
  - Normalizing by year – no consistent improvement



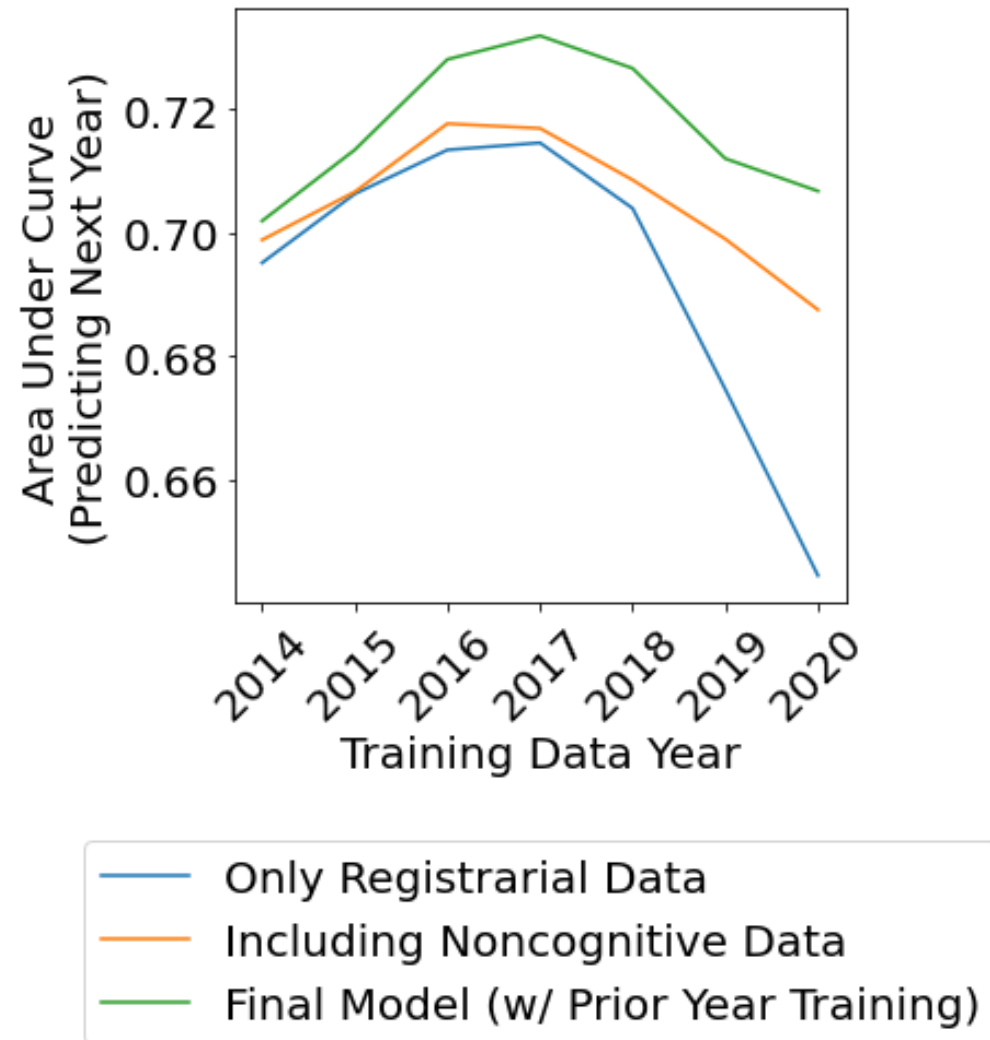
Adjusting the threshold for TP to 60%, FP decreases  $\approx$  10%

# Improving Predictive Performance with Noncognitive Data

**Increased predictive accuracy and robustness to external variables**

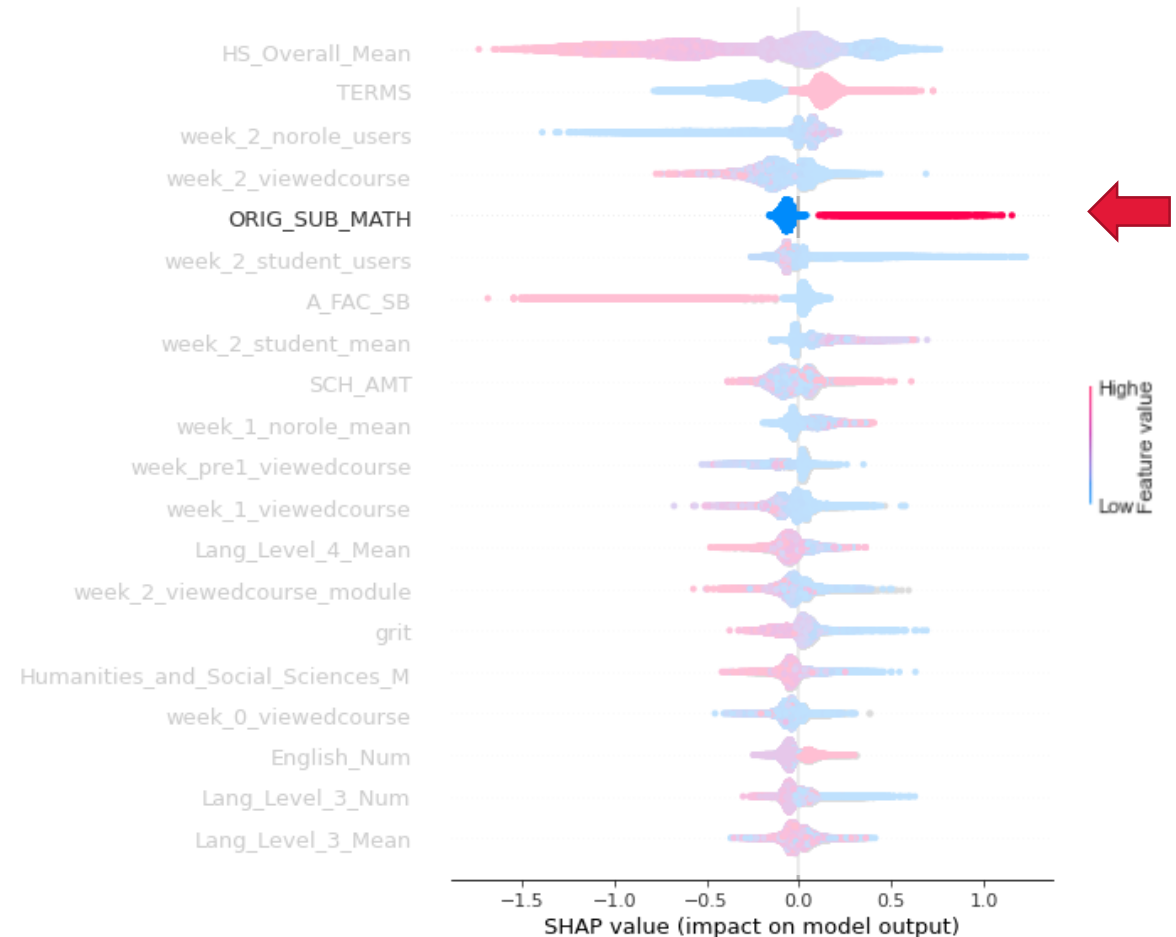
Model Benefits from multiple sources of data:

- Behavioral
- Noncognitive



## Sidebar: Interpretability

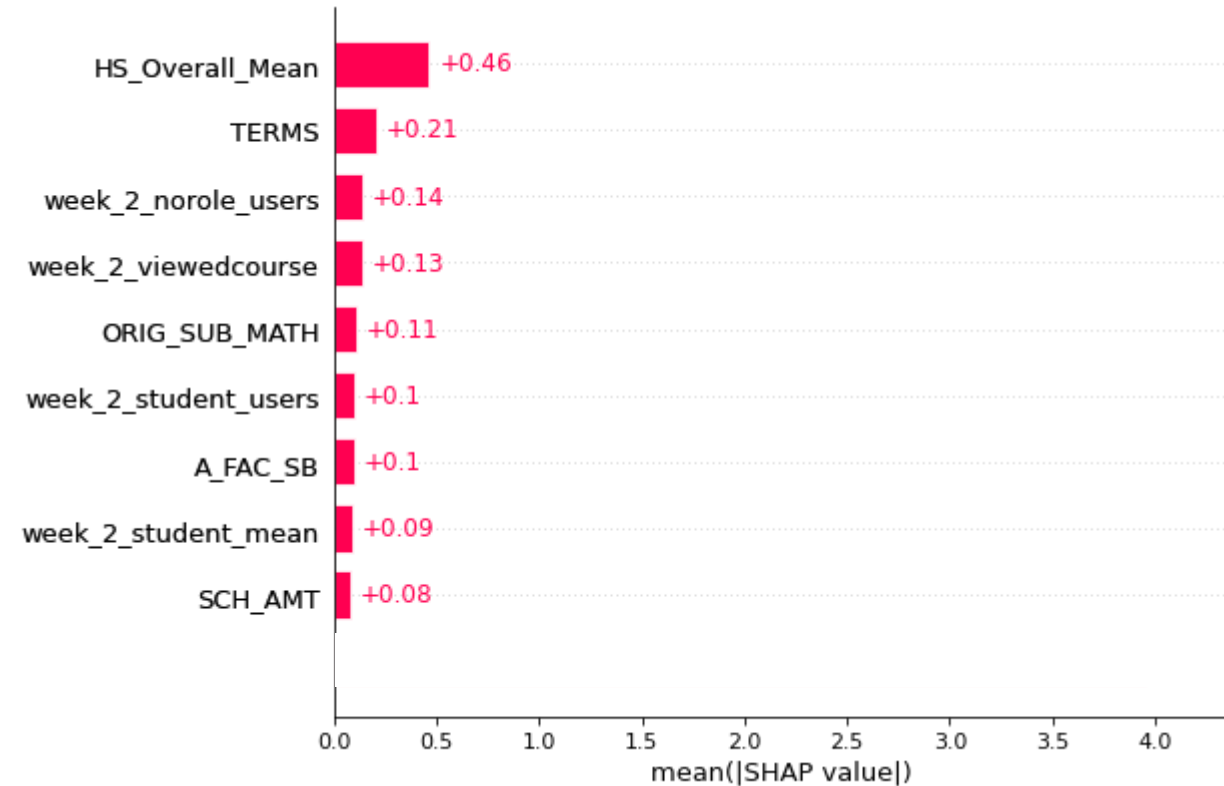
- There are ethical considerations when any model is 'deployed'
  - Ideally, we could understand the *exact* rules used (such as with Logistic Regression)
    - But... as the models get more complex, it becomes more and more difficult to parse (so called 'black boxes')
- Many advancements have been made, including:
  - **SHAP** (SHapley Additive exPlanations)
- While not perfect, they attempt to explain the model (not the data).





# Making Inferences at the Cohort Level

- How important, *on average*, a feature is for predicting
- Is the model *fair*?
- Usually concerned with the 'Top-N' features...
  - But that misses a lot of importance when predicting at the student-level

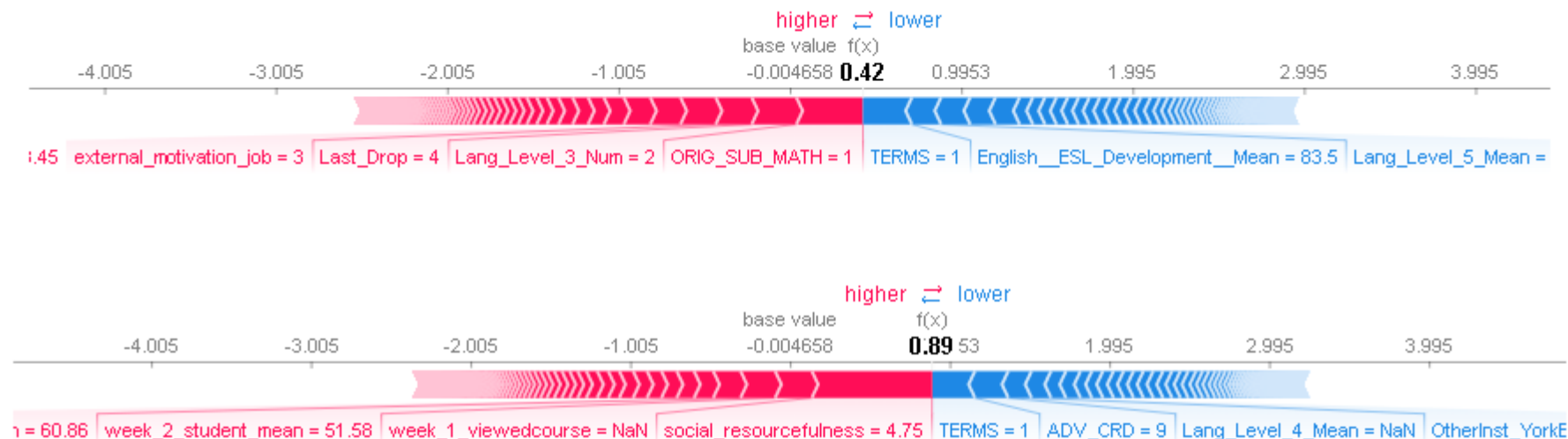


# Making Inferences at the Student Level

Can lead to different inferences about what's important

Can we “trust” a student-level model?

- *Why* is the model predicting what it is?
- *Should* it be acted on?



# Final Thoughts and Observations

- SAVY – Promising new data source
  - 2020 was the first year data are available
  - SAVY was still 'learning' at that point
- Increased predictive ability by incorporating noncognitive skills and traits
  - Increases accuracy
  - Increases robustness
- Cohort-level vs student-level predictions
  - Tools available to interpret models at both levels
  - Downstream intervention approaches can be based on either cohort-level or student-level inferences
  - Inferences, and thus interventions, can differ depending on level of analysis



Questions?

