Student Success Analytics Summit

AT BENTLEY

Bentley’s First Predictive Model for Student Success

Nathan Carter, Director, Center for Analytics and Data Science (CADS)
Gabrielle Friedman, Business Systems and Data Analyst, IT
Rahat Sabyrbekov, Data Scientist, Office of Student Success (OSS)
Gregory Vaughan, Chair, Mathematical Sciences Department

Student Success Analytics Summit
AT BENTLEY

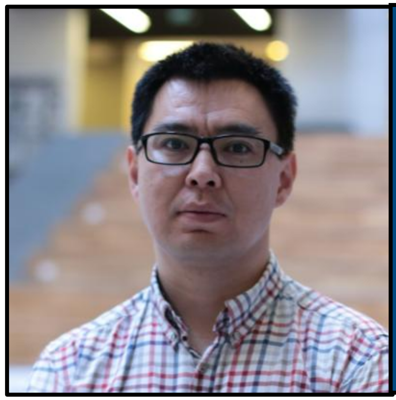
Outline

1. Project origins (Fall 2025)

Nathan Carter

2. Expanding the team (Spring 2026)

Greg Vaughan



3. Difficulties encountered

Rahat Sabyrbekov

4. Results and future work

Gabrielle Friedman

5. Discussion

Student Success Analytics Summit AT BENTLEY

Things you can look forward to

Lessons we learned

Activities

Student Success Analytics Summit AT BENTLEY

Outline

1. Project origins (Fall 2025)

Nathan Carter

3. Difficulties encountered

Rahat Sabyrbekov

2. Expanding the team (Spring 2026)

Greg Vaughan

4. Results and future work

Gabrielle Friedman

5. Discussion

Student Success Analytics Summit
AT BENTLEY

Problem origins

Jane De Leon Griffin
Associate Provost, Student Success
Professor, Modern Languages

Conversation in 2025 about collaborations between CADS and OSS

- CADS seeking internal consulting
- OSS eager to get more data-driven

<u>Resources OSS already had:</u>

- Jane’s position and mandate
- BaSSE project, including goals and funding
- New data scientist on team, Rahat
- Descriptive analytics dashboards

Problem statement

- Student attrition often becomes visible when it's already too late
- Can we augment our existing tools with early risk signals?

- Bentley's First Predictive Model for Student Success

Feasibility Study

- First predictive goal:
  Which students are on-track to graduate in four years?
- **Does OSS already have access to any variables that are correlated with graduating in four years?**
- Brought Greg Vaughan into the project
  - Statistics expertise
  - Department chair experience
- Answer: Yes, there are useful predictors, such as:
  - ACAD score
  - Grades in GHEFY courses

Lessons Learned

1. **On-campus partnerships have a lot of benefits**
   - Leveraging existing expertise
   - Making new connections
   - Using existing institutional knowledge
2. **Feasibility checks are worth it**
   - 5% of the cost of the whole project
   - Gave us confidence to proceed

Student Success Analytics Summit AT BENTLEY

Moving to the next phase

- The feasibility study said to proceed.

- **Goal for the full project:**
  - Create a dashboard for advisors
  - Alert them to which students are at risk for not graduating in four years
  - Advisors can then reach out with extra support

Outline

1. Project origins (Fall 2025)

Nathan Carter

3. Difficulties encountered

Rahat Sabyrbekov

2. Expanding the team (Spring 2026)

Greg Vaughan

4. Results and future work

Gabrielle Friedman

5. Discussion

Student Success Analytics Summit
AT BENTLEY

Identifying Tasks and Team

- Final Product: Dashboard
  - Data Analyst with PowerBI Expertise
- *Interpretable* Predictive Model under the hood
  - Statistician to perform model building and assessment
- Data Pipeline to Feed Model and Dashboard
  - Data Scientist (or two!) to build, refine, and maintain pipeline
- Optimize user experience
  - Experience Design team to identify optimal setup via end users

Full Team – Met every Thursday at 8:30

- **Data sourcing and oversight**
  - Steve Cross
  - Leela Munagapati
  - Rachel Penn
  - Annie Raunikar
  - Edlira Stefani
- **User Experience**
  - Chris Haas
  - Jason Reynolds
  - Atharv Wairagade
  - Peter Xu
  - Daisy Yin
- **Data cleaning and modeling**
  - Nathan Carter
  - Gabrielle Friedman
  - Rahat Sabyrbekov
  - Greg Vaughan
- **Project Leads**
  - Greg Coco
  - Jane Griffin

Student Success Analytics Summit
AT BENTLEY

Defining Goals

- What do we mean when by predicting “student success?”
  - We debated between graduate in 4 years or 6 years.
- We ultimately settled on graduating in 4 years for two main reasons
  - An overwhelming majority of Bentley student graduate in 4, so looking at graduating in 6 years wouldn't change much.
  - We realistically had complete data for ~10 cohorts using grad in 4 years; using grad in 6 years would have reduced that by 20%.
- With this binary dependent variable, we used logistic regression

Defining Goals: One More Step

- Okay, we are modeling the *probability of graduating in 4* (**PG4**)! …but how does *when* we are making predictions play a role?
- Our ability to predict if someone will graduate in 4 will be different when they are Freshmen vs. Seniors
- **Our Solution: Different models for different points in time**
  - Model 1: predict PG4 before first semester
  - Model 2: predict PG4 after first semester
  - Model 3: predict PG4 after second semester
  - …
  - Model 8: predict PG4 after seventh semester

Identifying Predictors

- With this set up, we would look at the value of predictors at each given point in time for each respective model
- The variables we used came in two main categories
  - Academic performance
    - Cumulative GPA
    - Cumulative credits
    - Final Grades in a variety of foundational courses required for all students
  - Demographic
    - First Generation status
    - HS Academic (ACAD) score
    - Start semester

Activity 1

- There is a tension between the variables we want and those we have.
- Let's first consider the variables you want.
  1. **What variables would you consider ideal for predicting student success at your institution?**
  2. **Are there other measures of success you want to predict?**

Activity 1

- Some variables we wanted to include (but ultimately weren't able to):
  - Course level data such as homework or exam grades
  - LMS activity data
  - Progress reporting data

Student Success Analytics Summit AT BENTLEY

The Road Not Taken

- Some predictors we had to weigh against other factors
- Race and gender might have improved performance, but there were concerns about unintended implications.
    - E.g., “Student X is less likely to succeed because they are of race Y"
- We felt that any potential benefit these metrics would have brought could be gained in other places:
    - ACAD score
    - Initial college academic performance

Activity 2

- **Thinking back to the ideal metrics you identified in Activity 1, are there any complications that might make them difficult to use?**
- **Are there any alternatives at your institution that might “make up” for these if they weren’t included?**

Lessons Learned

1. **Alternate ideating and exploring**
   - It can be alluring to dream big, but building a model is an iterative process
2. **Be precise in your goals**
   - "Predicting success" can cover a wide range
   - Be clear about what you want to predict and when you want to predict it to help keep scope manageable
3. **Don't be afraid to start small**
   - Done is better than perfect
   - If you don't get everything in on the first attempt, once a model is built, it will be easier to add to later

Student Success Analytics Summit
AT BENTLEY

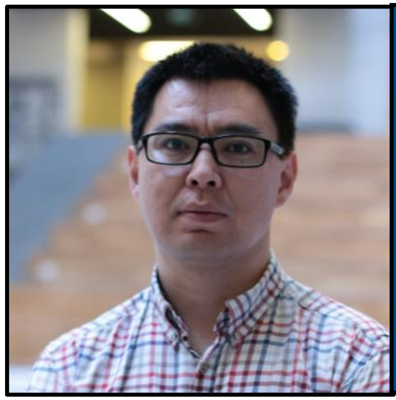
Outline

1. Project origins (Fall 2025)

Nathan Carter

2. Expanding the team (Spring 2026)

Greg Vaughan



3. Difficulties encountered

Rahat Sabyrbekov

4. Results and future work

Gabrielle Friedman

5. Discussion

Student Success Analytics Summit AT BENTLEY

Difficulties of various forms

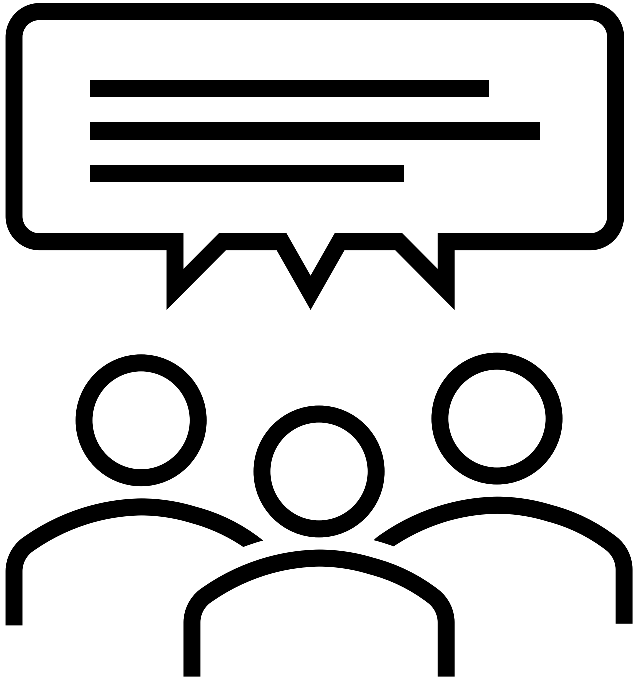
DIFFICULT DEFINITIONS

LOGISTICAL CHALLENGES

DATA QUALITY

Student Success Analytics Summit AT BENTLEY

Difficult definitions



- Edge cases like stop-outs must be defined before the training population can be set.
- Learn and follow the established institutional definitions (e.g. GHEFY, ACAD score).
- Counting "semesters" are complicated by summer and winter terms - > 'halves'.
- Combined BS/MS programs resist a single definition of success.
- The most predictive grade cutoff was not always DFW — sometimes it fell between B/C or A/B.
- D, F, W – may indicate different scenarios of student learning

Student Success Analytics Summit AT BENTLEY

Logistical challenges

Direct database queries were slow, so a cloud pipeline now stages the relevant data for efficiency.

Reusing existing data views saves time but risks outdated or mismatched data.

The university is transitioning to new platform - Microsoft Fabric – new for the whole team, adding a learning curve across the entire workflow

A large, complex pipeline introduces multiple points of failure to manage (e.g., 9 notebooks).

Data quality

- Thorough EDA surfaces inconsistencies that are expected in long-standing, multi-system institutional data. Some examples:
  - Cohort codes were sometimes missing or changed over time.
  - Credit totals occasionally decreased or stalled despite enrollment.
  - Some students had multiple academic scores or duplicate course grades, each requiring a documented resolution rule.
- Some data was not historically collected (e.g. First-Gen started in 2014) - > modeling tweaks

Student Success Analytics Summit AT BENTLEY

Activity 3

- **Imagine:**
  - You are an academic advisor. (Maybe you already are!)
  - You are about to open a dashboard covering all your advisees.
  - Your goal is to identify which ones to reach out to with an offer of extra support.
- **Consider:**
  - **What do you want the dashboard to show you at a glance?**
  - **What might you want to be able to drill down on?**

Student Success Analytics Summit AT BENTLEY

Activity 4

- **Imagine:**
  - You are the same academic advisor from Activity 3.
  - You have reached out to the advisees the dashboard flagged as most at risk.
  - Several have responded and scheduled appointments with you.
- **Consider:**
  - **What information do you want from the dashboard before you walk into those conversations with students?**

Lessons Learned

1. **Budget generously for EDA.** The uncertainty and de-obfuscation phase ran far longer than expected — plan for it up front.
2. **Partner with data professionals early and often.** Treat them as ongoing collaborators, not a one-time resource.
3. **Bad data is rarely the data team's fault.** Anomalies usually reflect institutional history, not current oversight.
4. **Documenting anomalies gives back.** Careful records of data issues are a real benefit the project returns to the data team.
5. **Check data constantly.** Validate against known truth (e.g., published grad-in-four rates) and basic reasonableness (e.g., GPA suddenly zero) to guard against pipeline failures.
6. **Data quality never "finishes."** It only gets better — proceed with good-but-imperfect data, documenting gaps for later follow-up.

Outline

1. Project origins (Fall 2025)

Nathan Carter

3. Difficulties encountered

Rahat Sabyrbekov

2. Expanding the team (Spring 2026)

Greg Vaughan

4. Results and future work

Gabrielle Friedman

5. Discussion

Student Success Analytics Summit
AT BENTLEY

Probability to Graduate in 4 Years

PG4 Overview

Filters

Primary Program of Study: All

Academic Standing: All

Class Standing: All

First Generation: All

ACAD Score: All

Pell Grant: All

Legal Sex: All

Ethnicity / International: All

Housing: All

Athlete: All

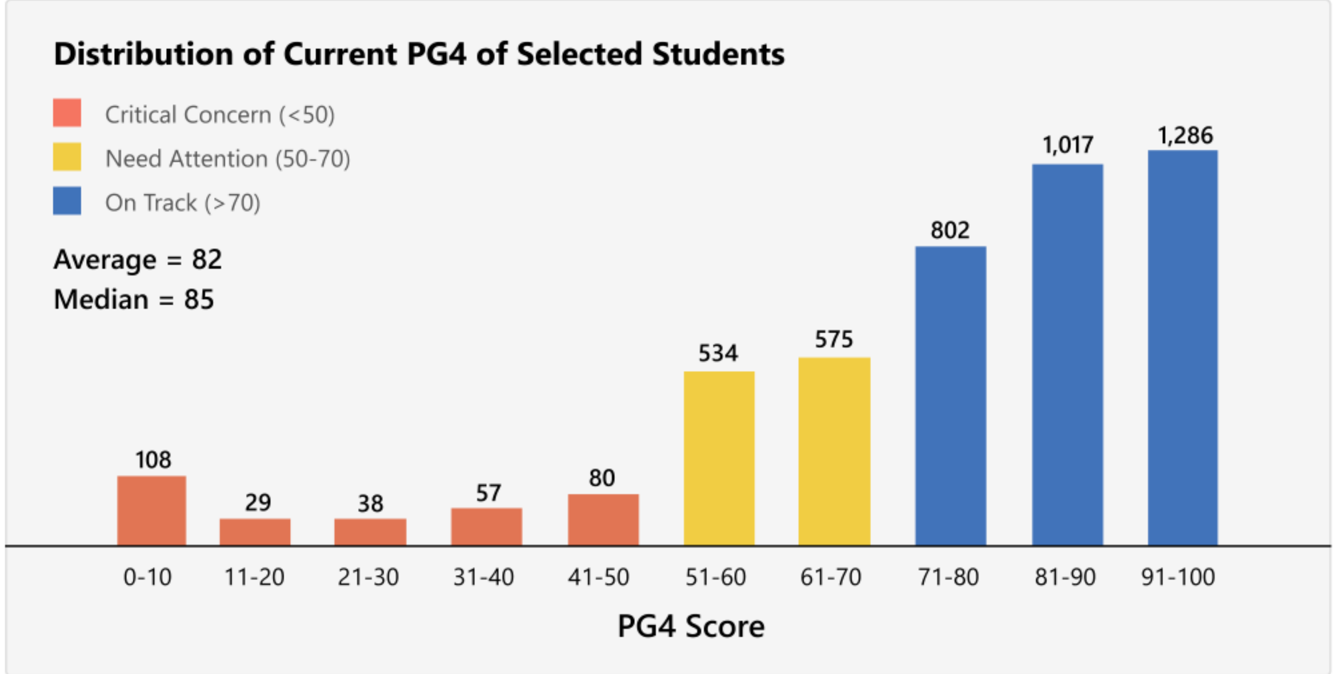
PG4 for Selected Students with Start Term Fall 2022 to Spring 2026

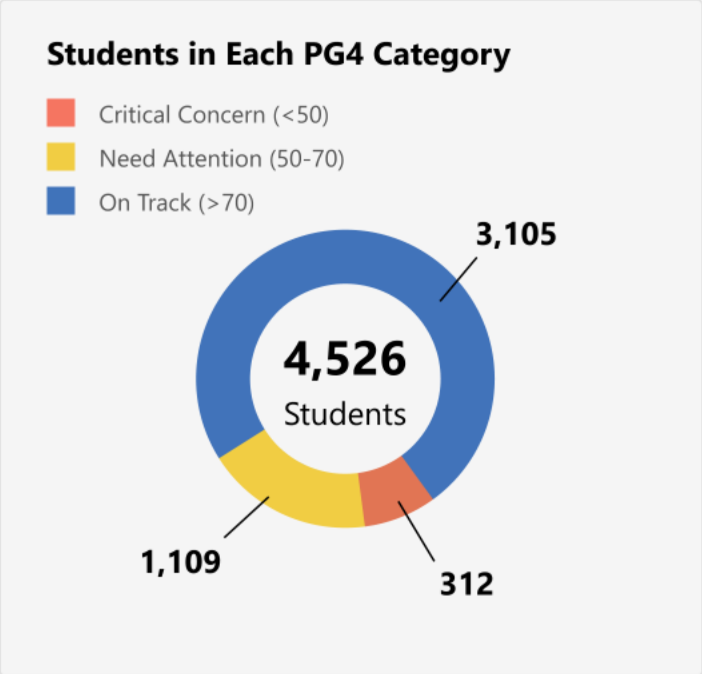
Start Term Range Begins: Fall 2022

Start Term Range Ends: Spring 2026

Prediction Date As Of: May 2026

Count / Percent





About the Distribution

4,526	00 - 100	Fall '22 - Spring '26	8
Selected Student Count	Range of PG4	Start Term Range	Cohorts Represented

Probability to Graduate in 4 Years

PG4 by Program of Study

Filters

Primary Program of Study
All

Academic Standing
All

Class Standing
All

First Generation
All

ACAD Score
All

Pell Grant
All

Legal Sex
All

Ethnicity / International
All

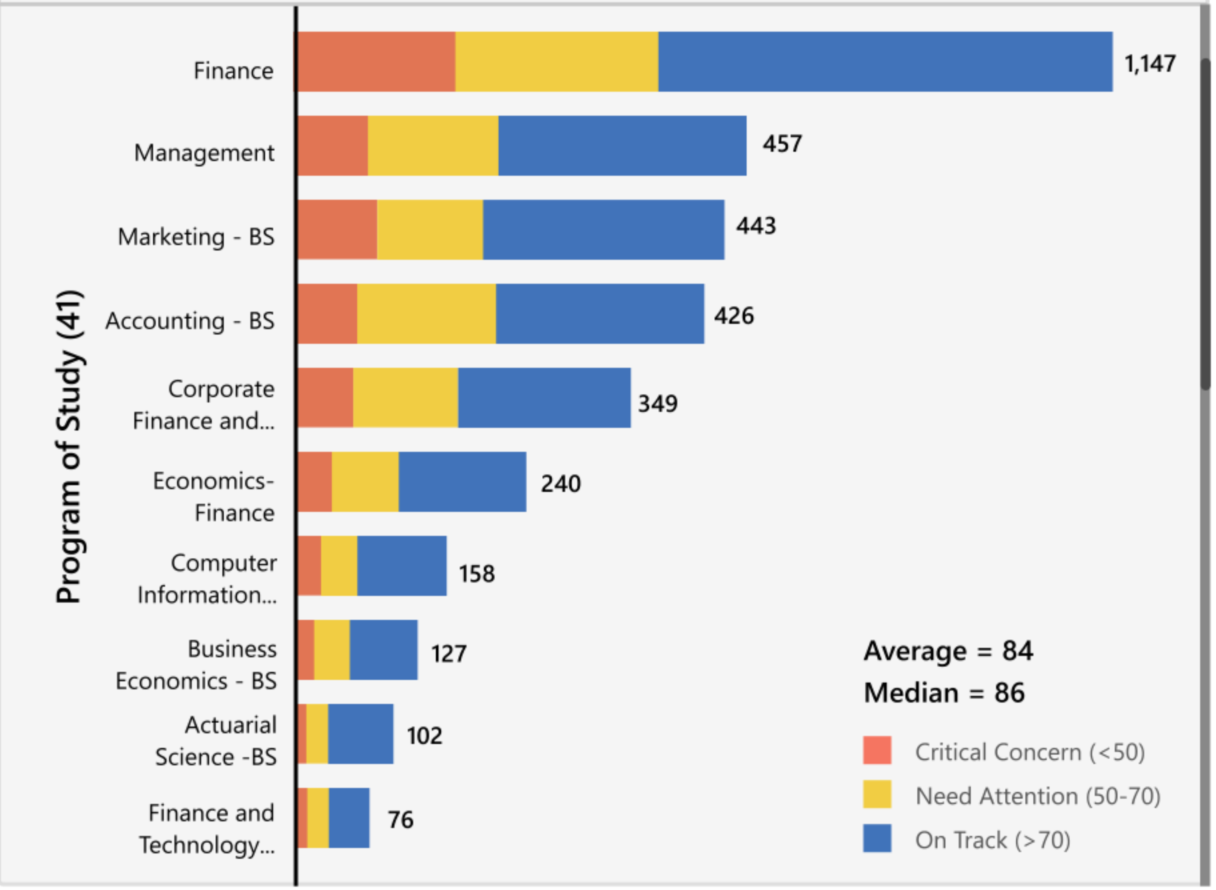
Housing
All

Athlete
All

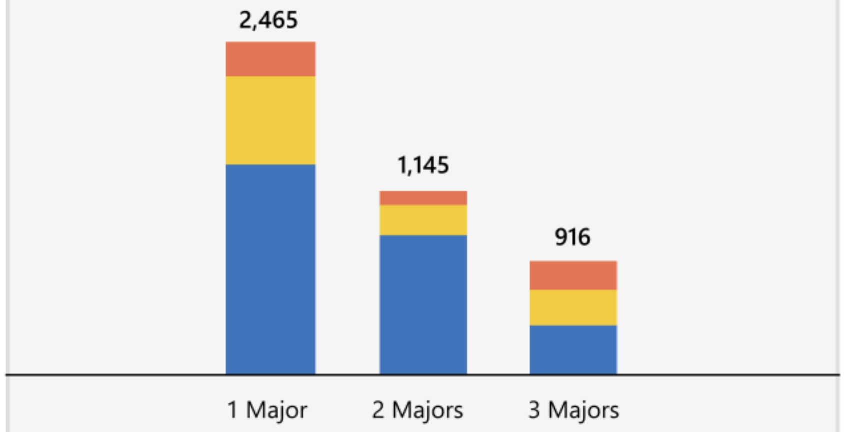
PG4 by Program of Study

Start Term Range Begins:	Start Term Range Ends:	Prediction Date As Of:	Student Program Load:
Fall 2022	Spring 2026	May 2026	All Majors and Minors

PG4 Risk Category by Primary Program of Study



Counts of Student Program Loads



About the Distribution

4,526
Selected Student Count

Fall '22 - Spring '26
Start Term Range

00 - 100
Range of PG4

41
Programs of Study Represented

Probability to Graduate in 4 Years

Advisor Cohort Details

Select Advisor Name: Amanda Curtis

Select PG4 Range: 0 50 70 100

Reset

Graphical View / Roster View

Filters

Primary Program of Study: All

Academic Standing: All

Class Standing: All

First Generation: All

ACAD Score: All

Pell Grant: All

Legal Sex: All

Ethnicity / International: All

Housing: All

Athlete: All

Filtered Result

	My Advisees	Total Students
Count	587 (↓ 12)	4,526
Share	12.9% of all students	100% of all students

Critical Concern

	My Advisees	Total Students
Count	42 (↑ 2)	314
Share	7.2% of all students	6.9% of all students

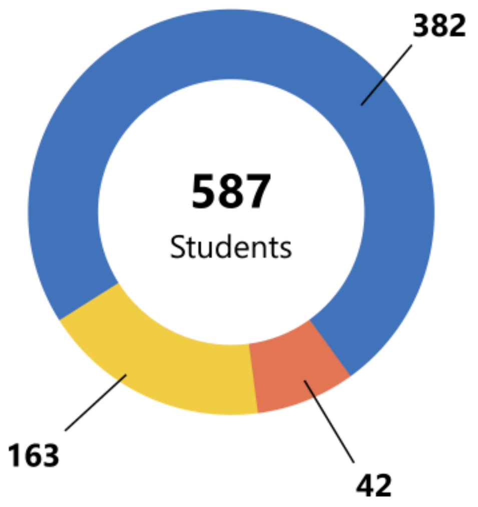
Needs Attention

	My Advisees	Total Students
Count	163 (↓ 4)	1,108
Share	27.8% of all students	24.5% of all students

Advisor View Details

- Critical Concern (<50)
- Need Attention (50-70)
- On Track (>70)

Count / Percent



Biggest PG4 Changes

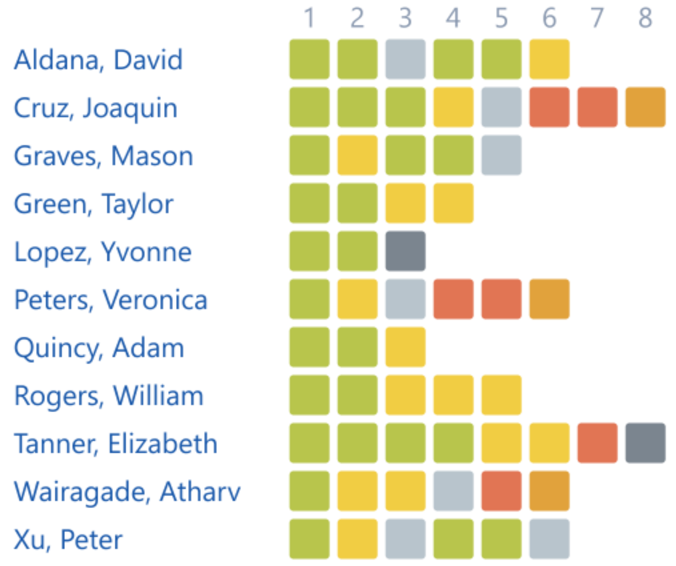
Compared to December 2025

-24	McCarthy, Michael
-19	Lardis, Skye
-18	Smith, Nathan
-16	Flores, Patience
-14	Peterson, Ashley
+12	Scott, Elliot
+13	Campbell, Jacob
+16	Brooks, Sunday
+19	Mbeki, Kwame
+21	Whitestone, Emily

Students Reviewed by CAS

☑ Students with Current Good Standing Included

Sort By: Alphabetical



Good · Concern · Recovery · Separation · Recovery following separation · Review

Probability to Graduate in 4 Years

Individual Student Profile

Search by student name or Bentley ID...

Filters

Primary Program of Study: All

Academic Standing: All

Class Standing: All

First Generation: All

ACAD Score: All

Pell Grant: All

Legal Sex: All

Ethnicity / International: All

Housing: All

Athlete: All

Student PG4 Details

Marcus A. Rivera (Marc)

Start Term: Fall 2023

Recovery after Separation

Early Expected Completion: No

Most Recent PG4 Score

31 ⚠

Critical Concern

-7 pts since Fall 2025

36th percentile

−34 vs. avg (65)

−27 vs. med (58)

Top Factors Increasing PG4

Business Statistics

First Year Math

First Year CIS

Top Factors Decreasing PG4

Credits Earned

Cumulative GPA

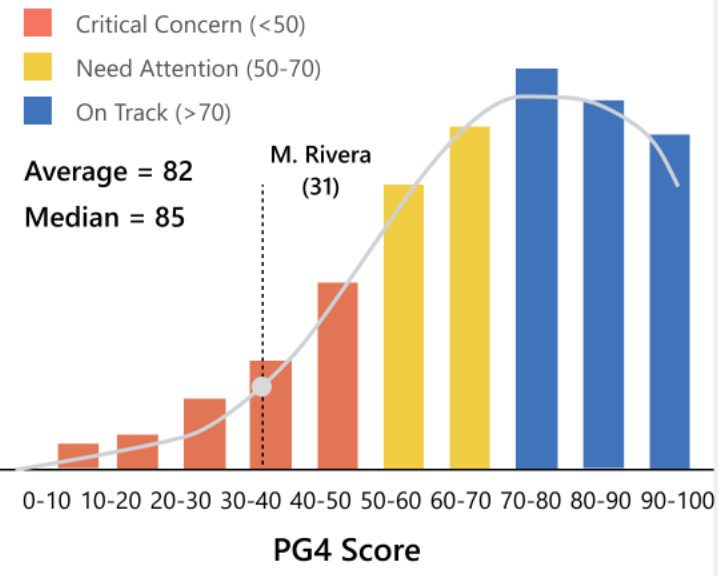
First Year Writing

Academic Record

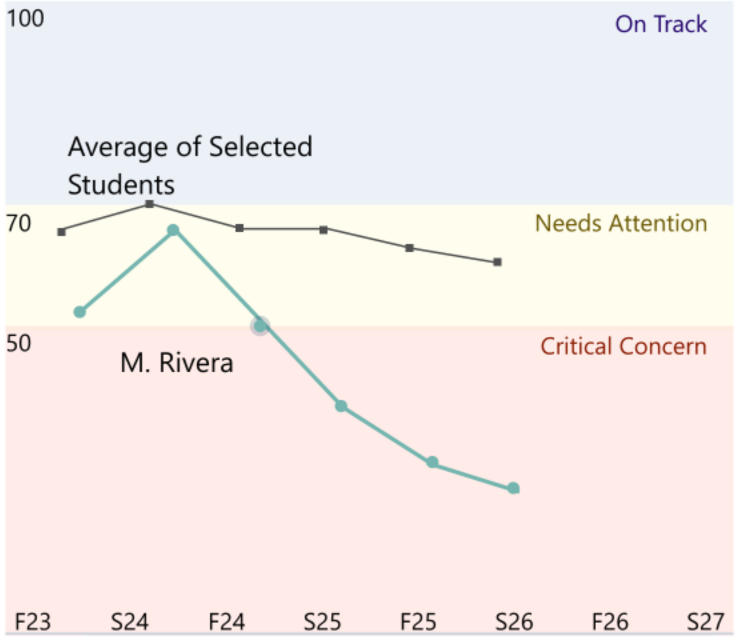
Cumulative GPA	Cumulative Credits Earned
2.11	72/120
	+15 since Dec 2025

Academic Standing History

PG4 Distribution: All Students



Student PG4 by Semester with Reference Groups



PG4 by Academic Semester

Show reference group data

- [x] Average
- [ ] Median

Reset

First Generation

- (●) All students
- ( ) Same as student

Legal Sex

- (●) All
- ( ) Same as student

Ethnicity / International

- (●) All
- ( ) Same as student

Program of Study

- (●) All students
- ( ) Same as student

Future work

- Iterating and <u>refining the model</u> over time
- <u>New types of variables</u> we didn't have access to for this model
  - Financial
  - Academic engagement
  - Advisor interactions
- <u>When you intervene</u> for students' good, that makes the model's predictions get worse over time, but for a good reason.

Student Success Analytics Summit AT BENTLEY

Outline

1. Project origins (Fall 2025)

Nathan Carter

3. Difficulties encountered

Rahat Sabyrbekov

2. Expanding the team (Spring 2026)

Greg Vaughan

4. Results and future work

Gabrielle Friedman

5. Discussion

Student Success Analytics Summit
AT BENTLEY

Probability to Graduate in 4 Years

Advisor Cohort Details

Select Advisor Name: Amanda Curtis

Select PG4 Range: 0 50 70 100

Reset

Graphical View | Roster View

Filters

Primary Program of Study: All

Academic Standing: All

Class Standing: All

First Generation: All

ACAD Score: All

Pell Grant: All

Legal Sex: All

Ethnicity / International: All

Housing: All

Athlete: All

Filtered Result

My Advisees	Total Students
587 ↓ 12	4,526
12.9% of all students	100% of all students

Critical Concern

My Advisees	Total Students
42 ↑ 2	314
7.2% of all students	6.9% of all students

Needs Attention

My Advisees	Total Students
163 ↓ 4	1,108
27.8% of all students	24.5% of all students

Advisor View Details

Search by Name or Bentley ID

Last Name	First Name	Bentley ID	PG4	PG4 Change	Progress	Academic Standing	Start Term	Program of Study	GPA	Legal Sex
Rivera	Marcus	10016842	21	-3	Paused	Recovery	Fall 2023	Program of Study	2.43	Male
Dattachaudhuri	aminolpopihell...	10017246	24	-8	Progress	Recovery	Spring 2023	Extra Long Program of S...	2.47	Male
Ruiz Cabrera	Maria Alejandra	10017234	47	-4	Progress	Recovery	Winter 2023	Program of Study	2.34	Female
Xu	Lou	10016226	58	-13	Paused	Concern	Fall 2024	Long Program of Study	2.71	Male
Montgomery	Kelly	10016947	62	-11	Paused	Good	Fall 2025	Program of Study	2.58	Male
Banner	Marc...	10016784	64	-9	Progress	Concern	Spring 2023	Extra Long Program of S...	2.72	Female
Aldana	Josephine	10017048	68	-12	Paused	Good	Fall 2023	Program of Study	2.81	Male
Jimenez-Cort...	Pedro	Bentley ID	74	-7	Progress	Concern	Fall 2024	Extra Long Program of S...	2.84	Male
Hollingsworth	Mallory	Bentley ID	77	+5	Progress	Concern	Fall 2023	Program of Study	2.96	Female

Probability to Graduate in 4 Years

PG4 by Start Term

Filters

Primary Program of Study: All

Academic Standing: All

Class Standing: All

First Generation: All

ACAD Score: All

Pell Grant: All

Legal Sex: All

Ethnicity / International: All

Housing: All

Athlete: All

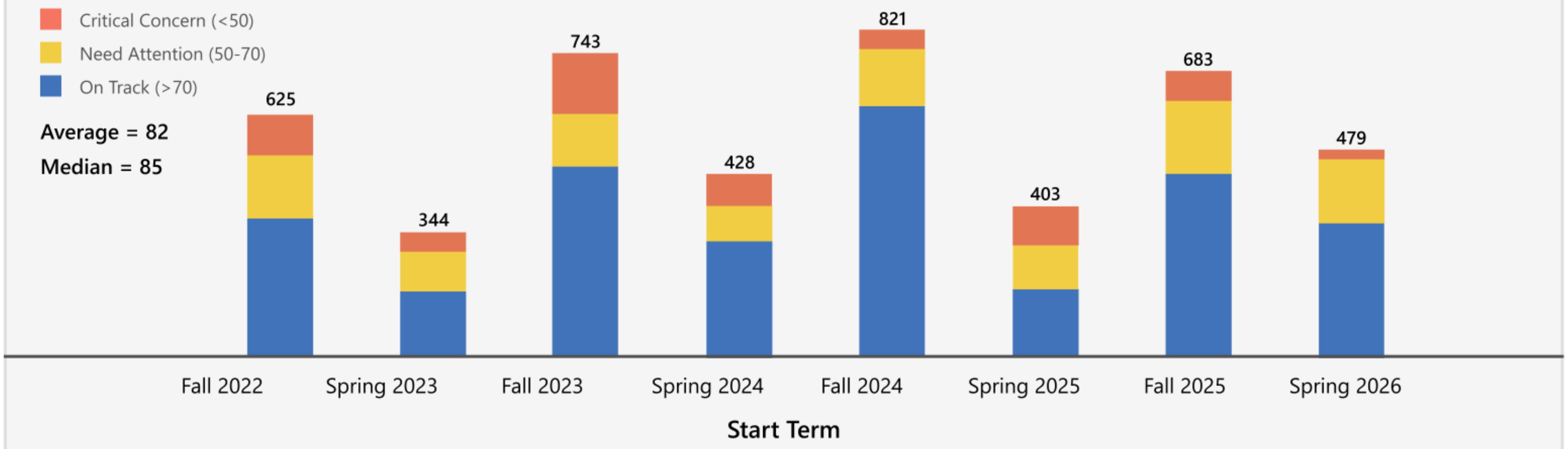
PG4 by Start Term

Start Term Range Begins: Fall 2022

Start Term Range Ends: Spring 2026

Prediction Date As Of: May 2026

PG4 Risk Category by Start Term



About the Distribution

4,526	00 - 100	Fall ’22 - Spring ’26	8
Selected Student Count	Range of PG4	Start Term Range	Cohorts Represented