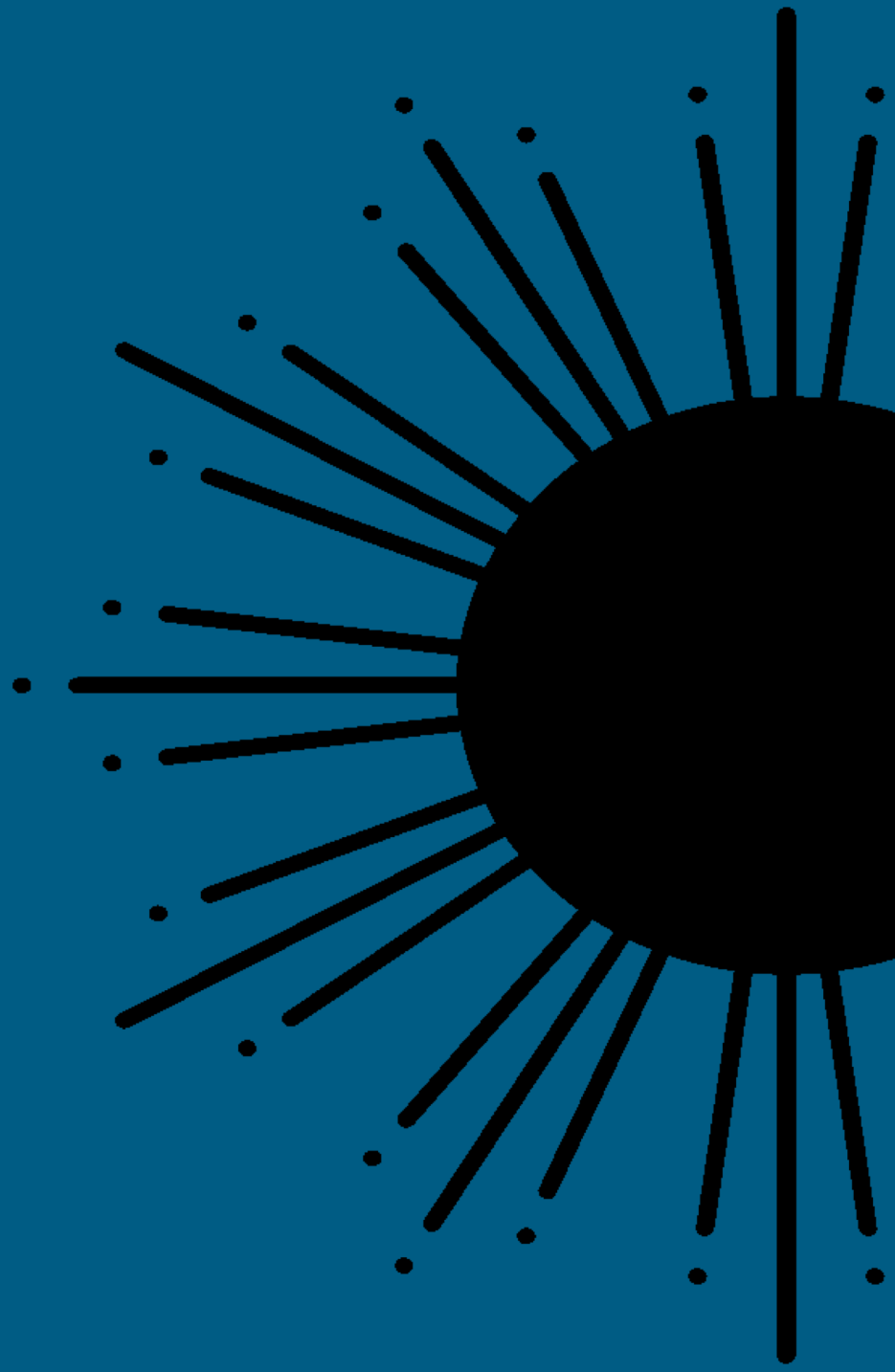
HelioCAMPUS
Better Data. Brilliant Insights.

Predictive Modeling Project Success Through Conversation, Process, and Transparency

R. P. Ruiz
Senior Data Scientist

CAIR 2021

Getting Started: Introductions

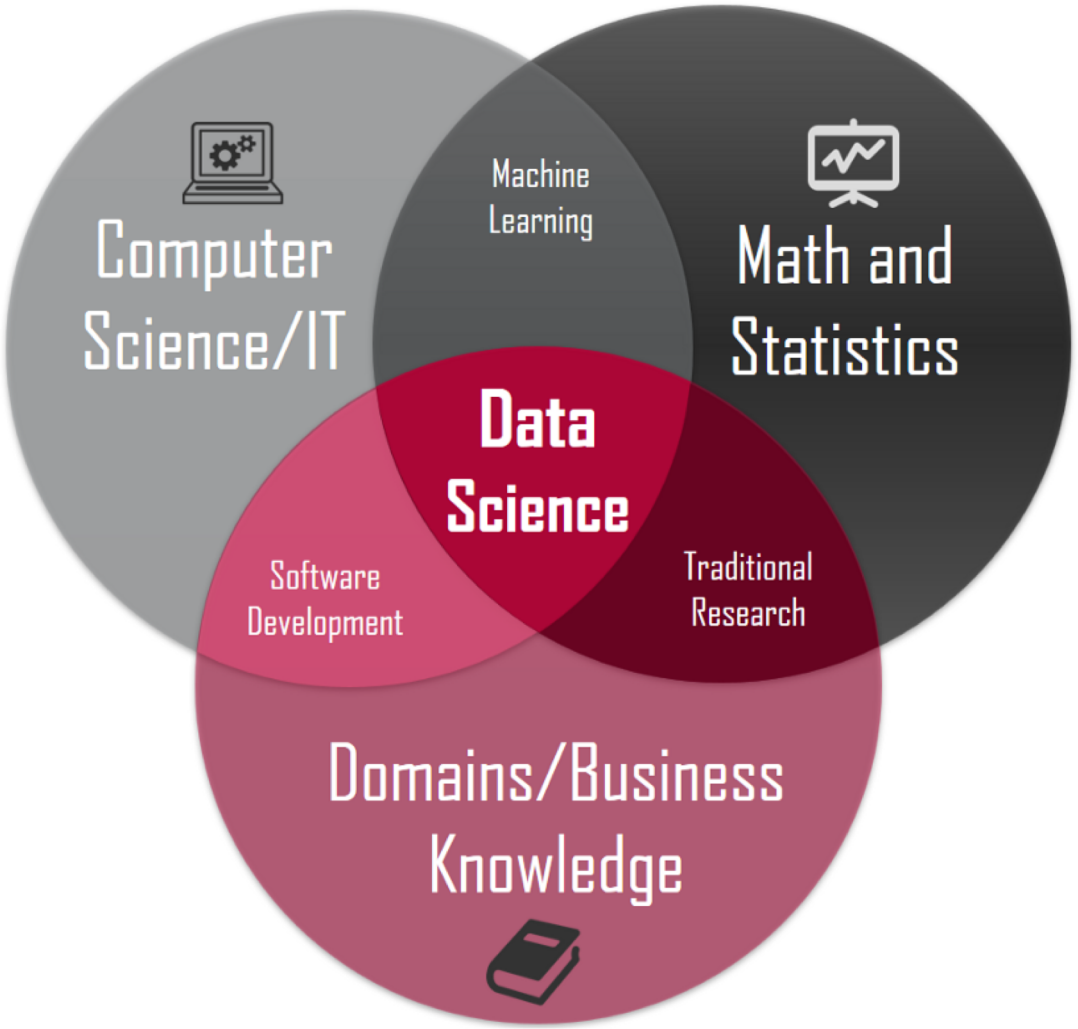


Plan for Today's Talk

- **Topics**
  - **Part I: Downside to Products/Black Boxes**
  - **Part II: Advantages to Process over Product**
  - **Part III: What's on your mind?**
  - **Part IV: Characteristics of Successful Predictive Modeling Projects**
- **Who Am I?**
- **What Are Your Predictive Modeling Project Experiences?**

Introductions: Who Am I?

- Hacker & Maker by Nature
- Linguist and Teacher by Training
- Data Scientist by Practice



Introductions: What Are Your Predictive Modeling Project Experiences?

How many of you (past, present, future)

- Have used predictive modeling in your work?
- Have formed a part of the stakeholders group?
- Have managed a predictive modeling project?
- How many of you are new to predictive modeling?

Part I: Downside to Black Box Models

Part I: Downside to Black Box Models

We are awash in black box algorithms (all social media)

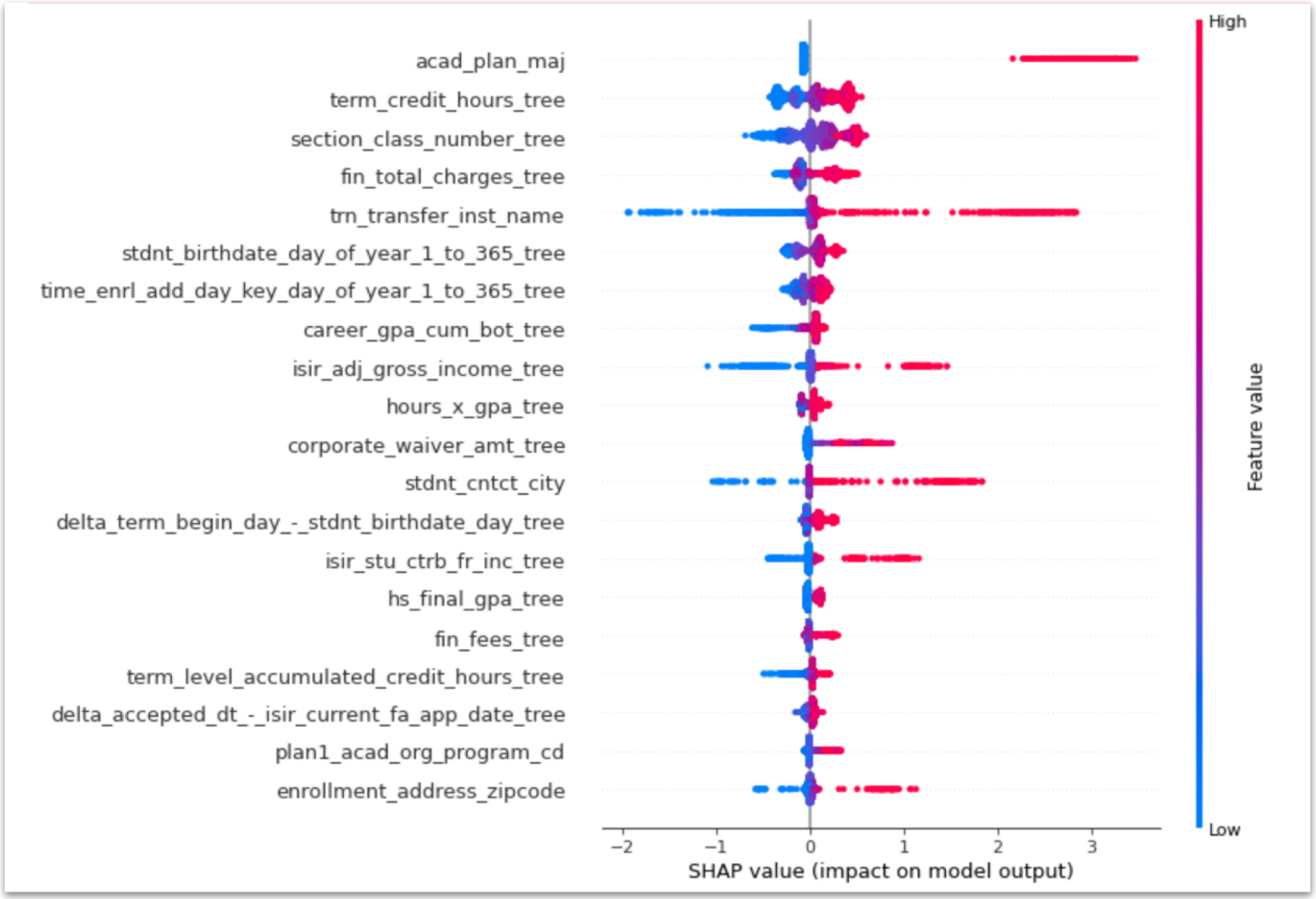
So, what's the problem? Lack of transparency:

- **How was a given prediction arrived at?**
- **Can't compare effects of including/excluding data**
  - **Pre-pandemic vs Pandemic**
- **Can't audit where it performs well, and where it doesn't**
- **Can't see who it advantages or disadvantages**

Part II: Advantages of Process over Product

Part II: Advantages of Process over Product

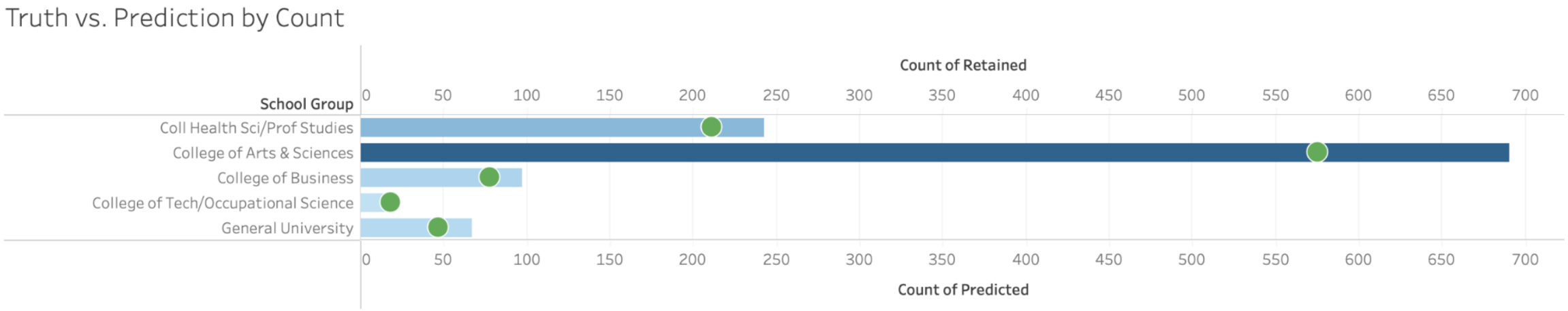
Is explainable



HelioCAMPUS

Part II: Advantages of Process over Product

Can be validated: where does it do better/worse?



HelioCAMPUS

Part II: Advantages of Process over Product

Can answer this question: “How was this score arrived at?” (Non-Standard)

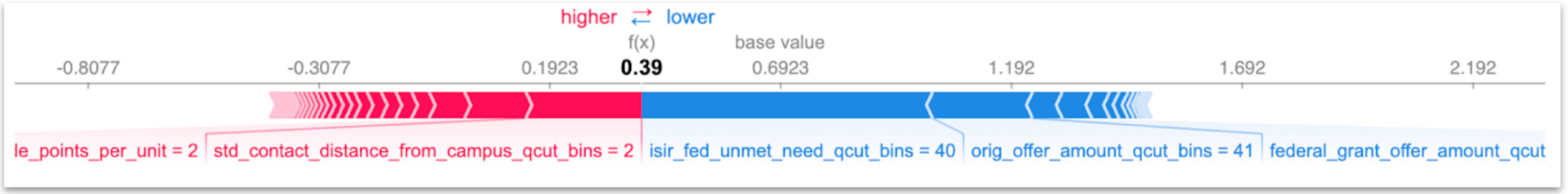
59% Likelihood to Retain: Student Has Financial Aid Problems

Top 3 features *lowering* their probability of retention *below* the average probability of 66%

Federal unmet need
Original student aid offer
Federal grant offer

Top 3 features *pushing* their probability of retention *up, towards* the average probability of 66%

Distance from campus
Grade points per unit
Financial aid application date



HelioCAMPUS

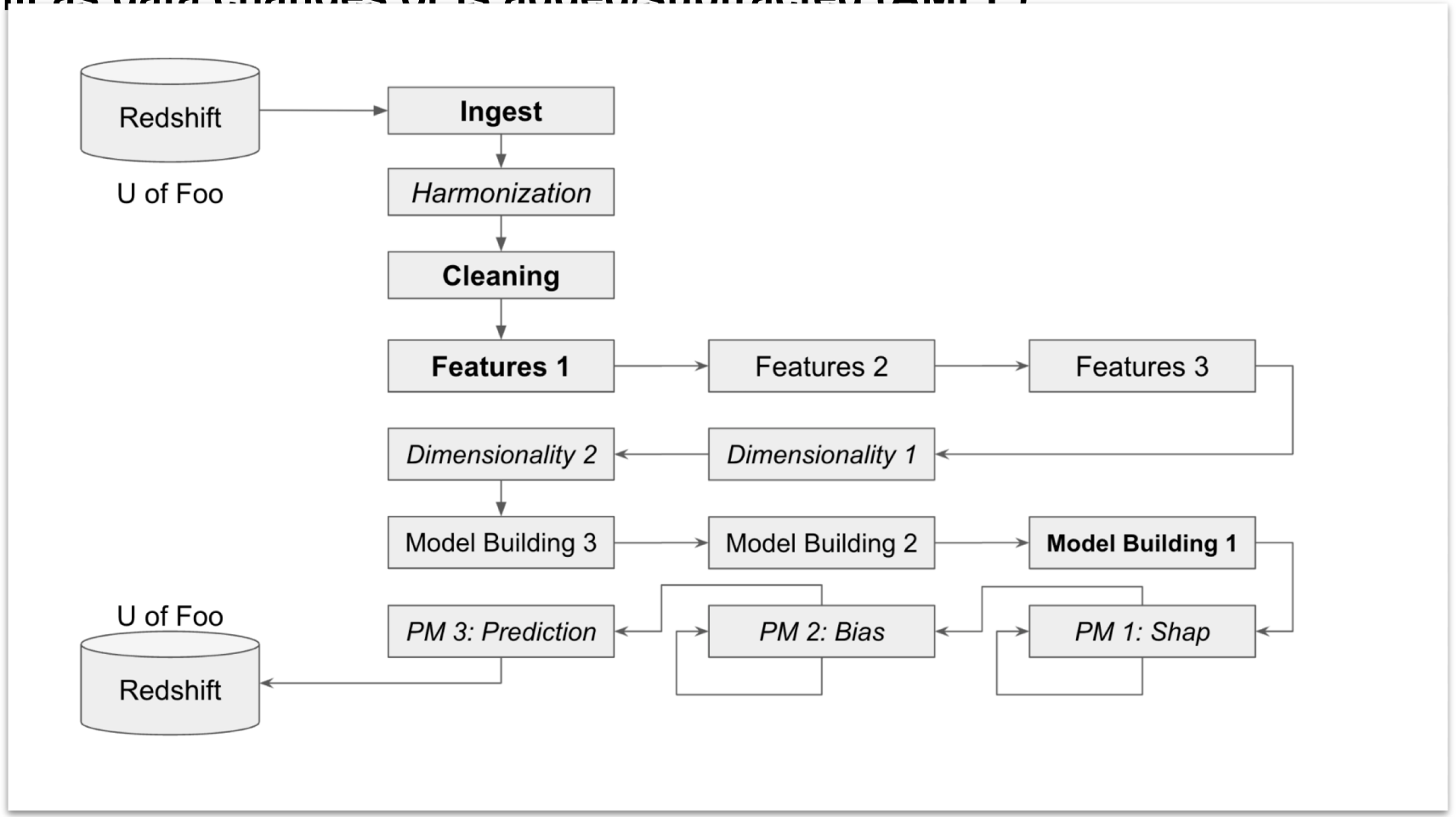
Part II: Advantages of Process over Product

Facilitates discovery/serendipity

HelioCAMPUS

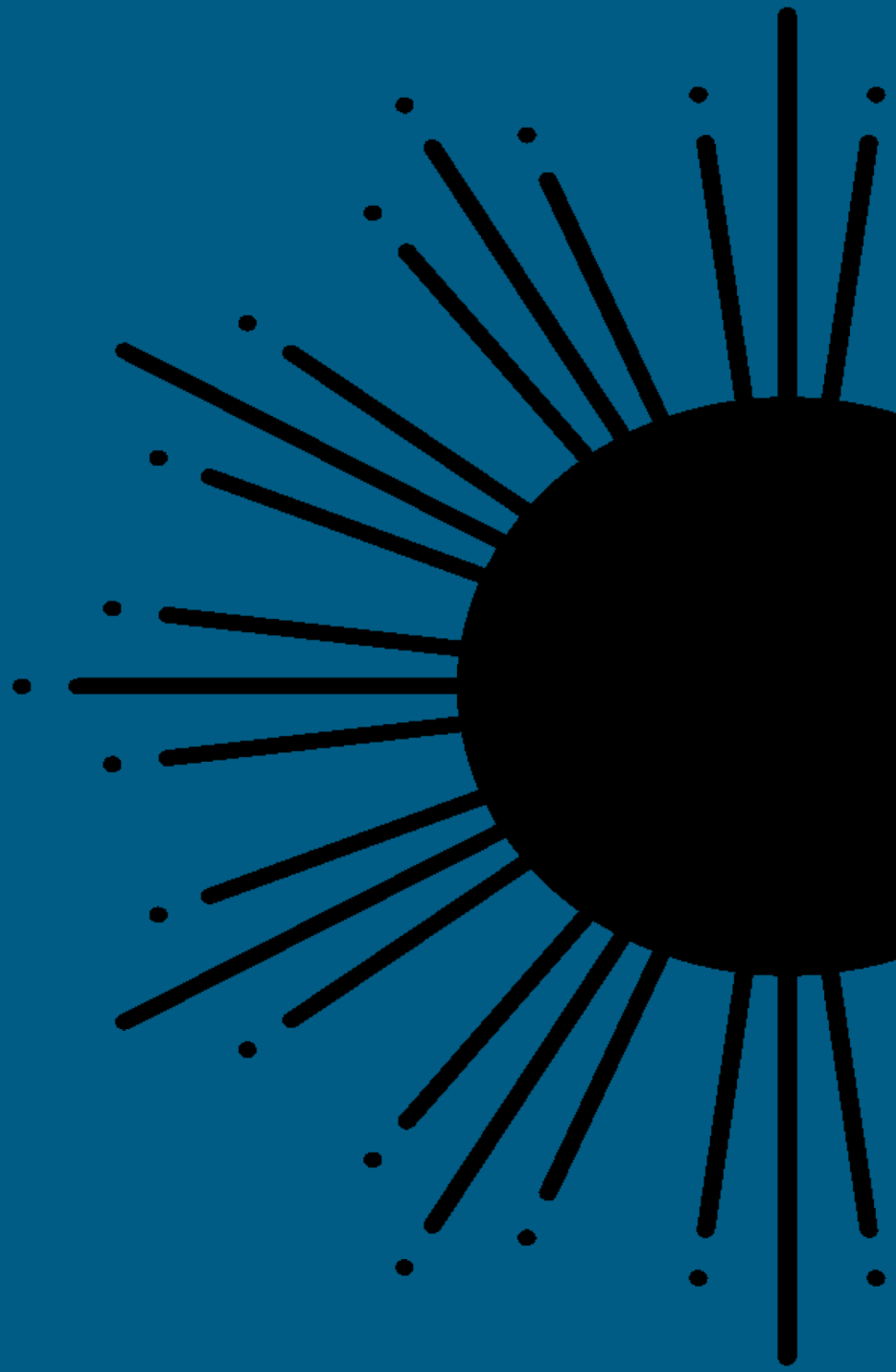
Part II: Advantages of Process over Product

Can be rebuilt as data changes or is added/subtracted (AMPE)



HelioCAMPUS

Part III: What’s on your mind?



Part IV: Characteristics of Successful Predictive Modeling Projects

Foundation

Part IV: Successful Predictive Modeling Projects

Seek common understanding through dialog

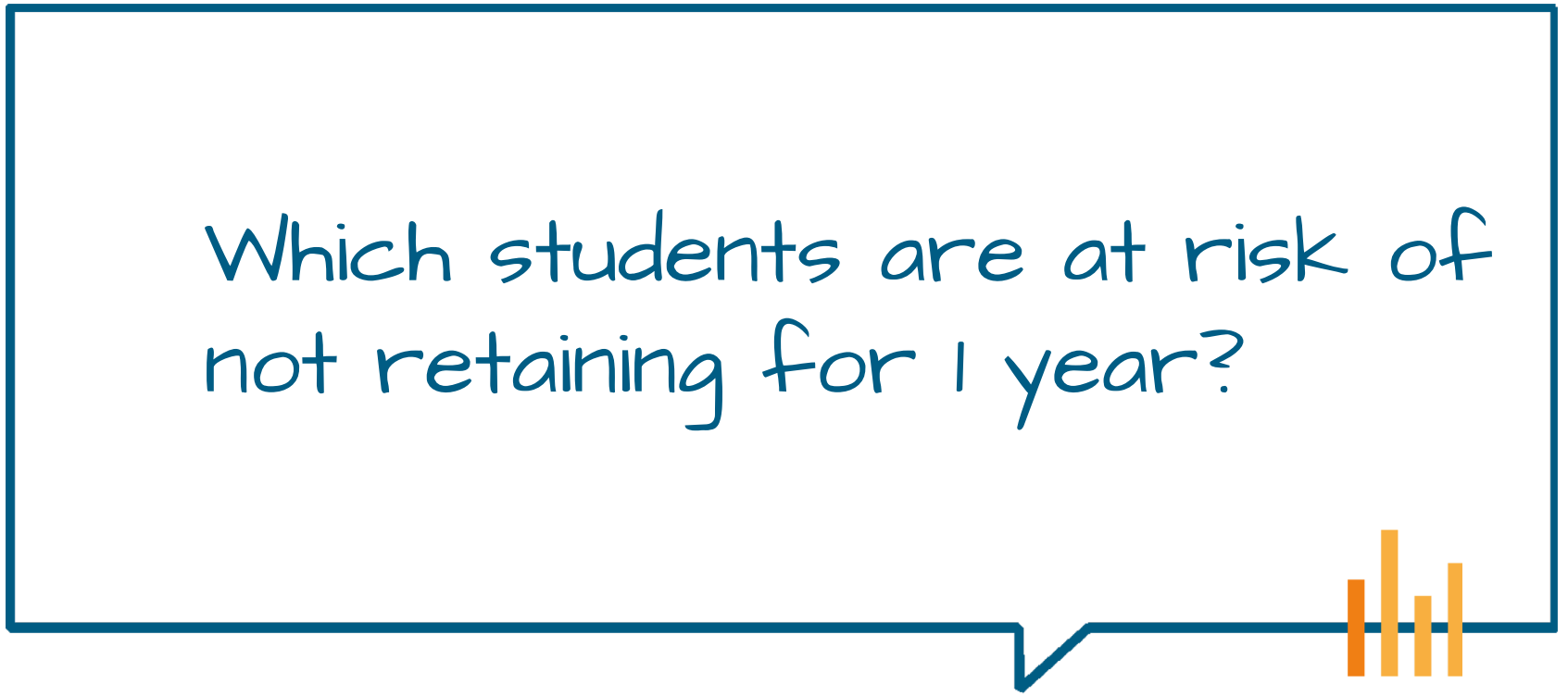
Applied Data Science 101

Step 1

Define the question

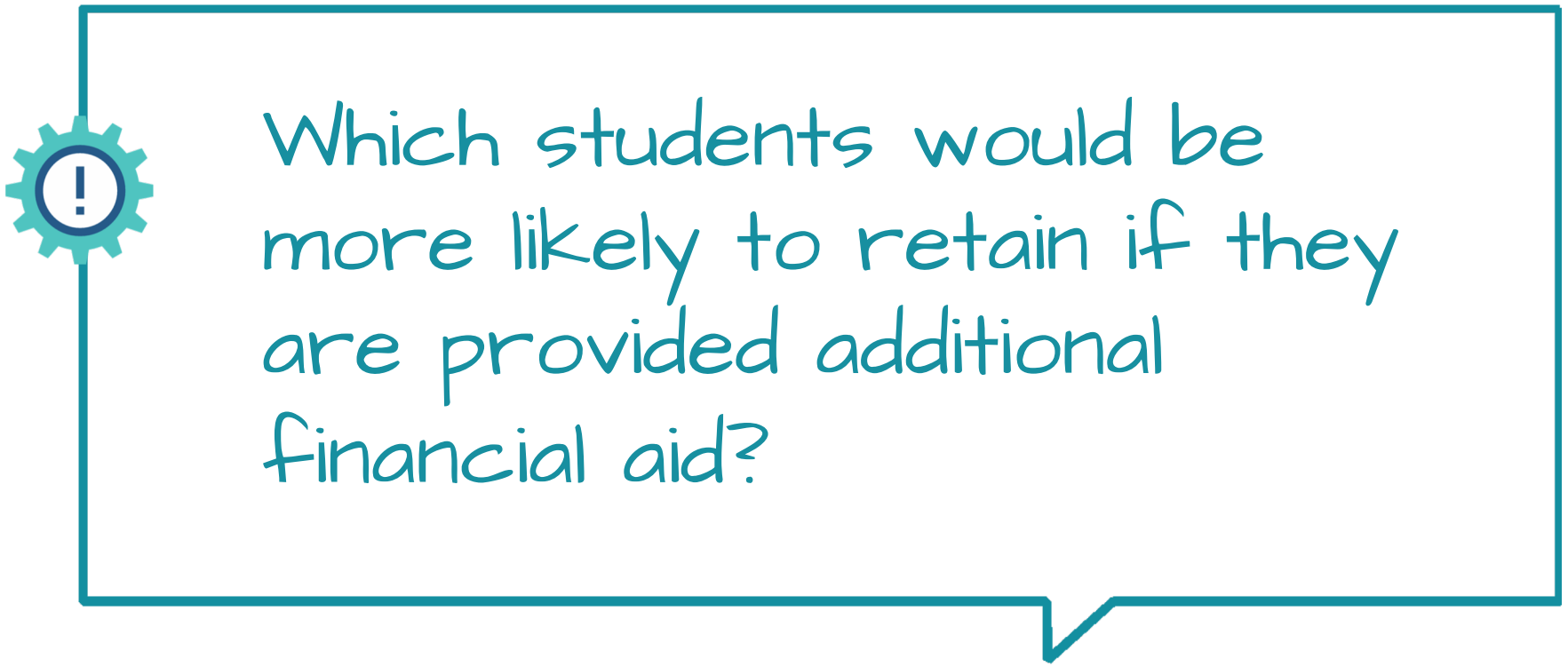
Part IV: Successful Predictive Modeling Projects

Seek to answer a specific question...



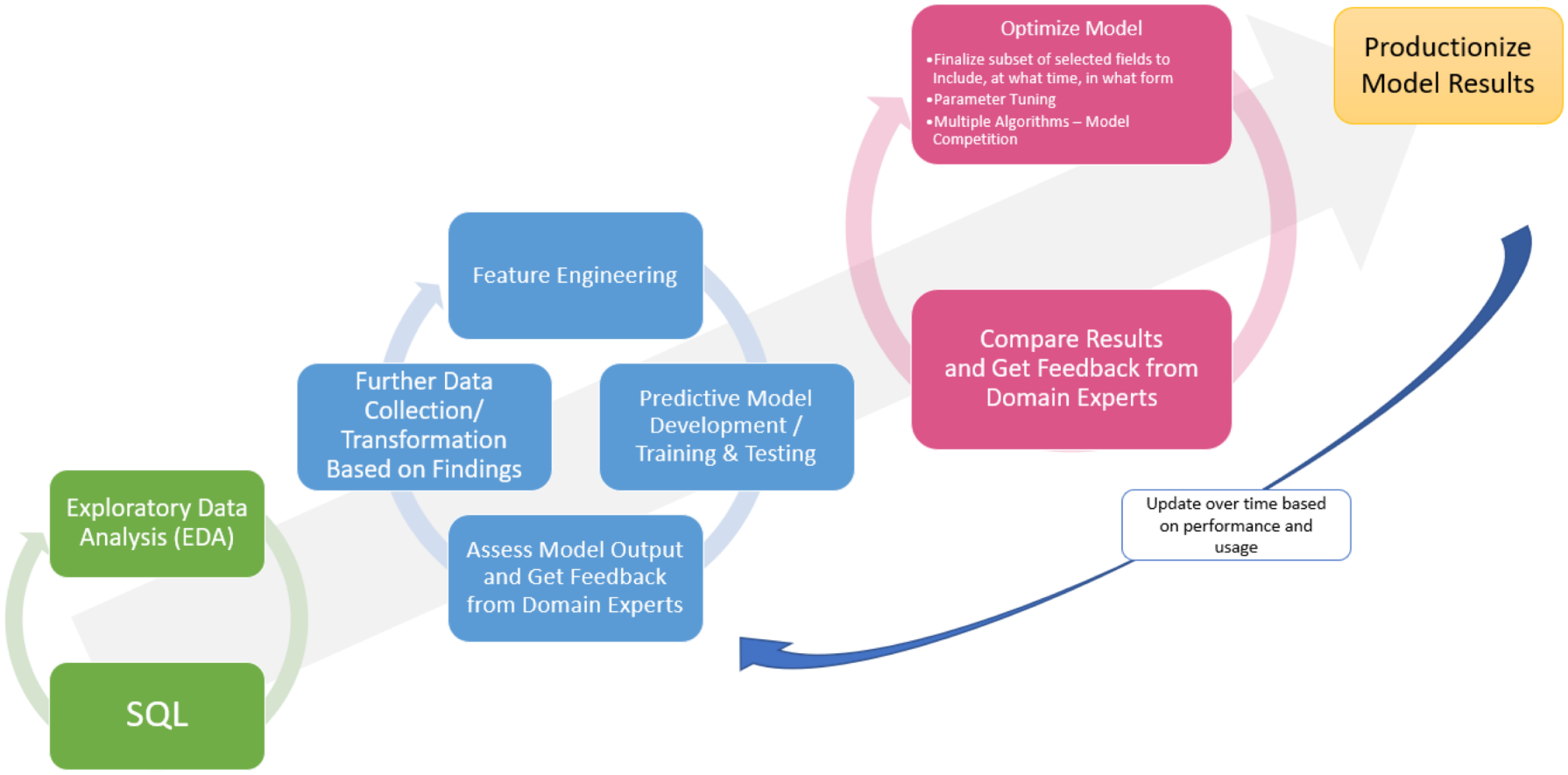
Part IV: Successful Predictive Modeling Projects

Seek to answer a specific question… which is *actionable*



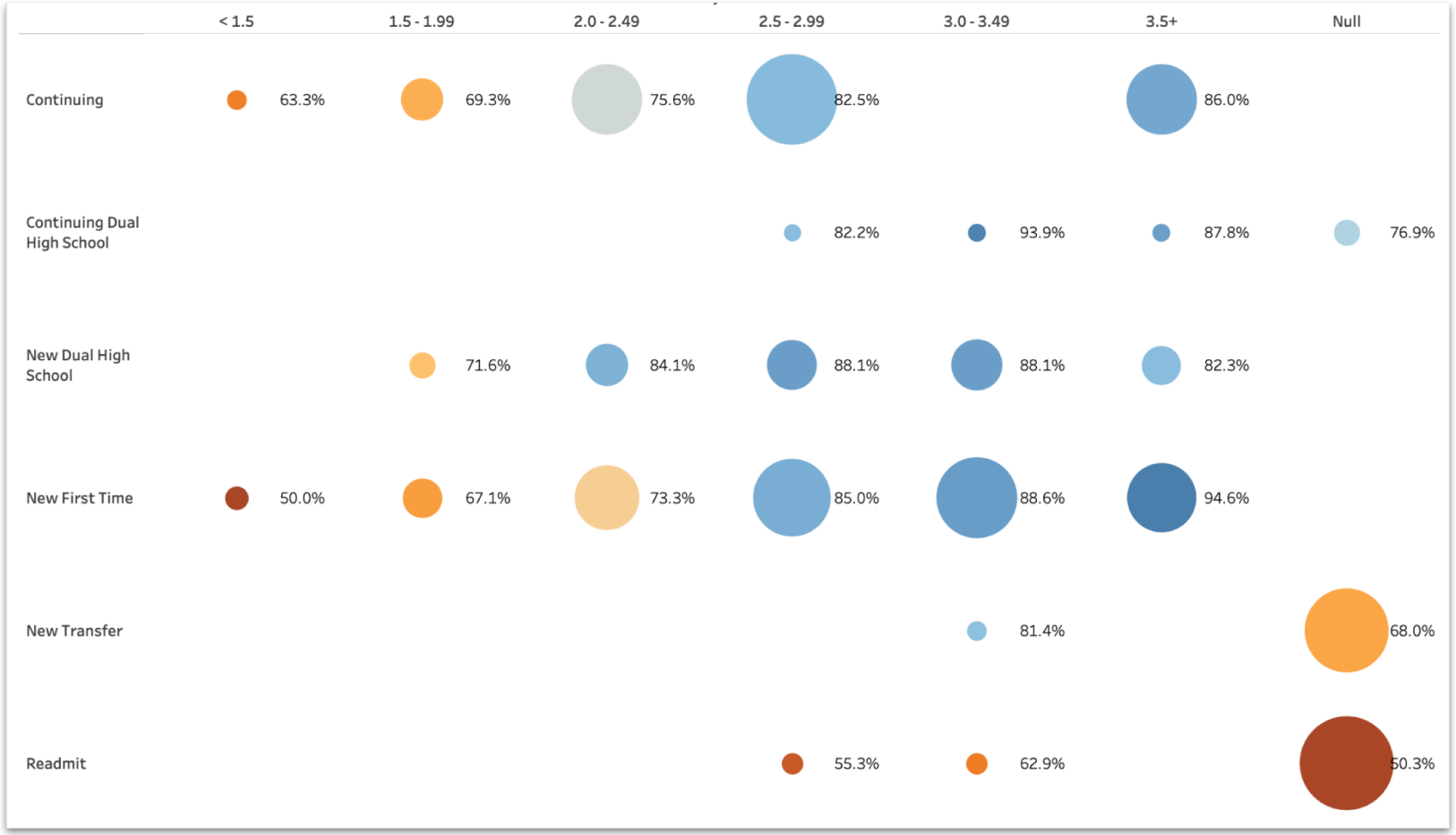
Part IV: Successful Predictive Modeling Projects

Are iterative in nature: outputs *and* concepts



Part IV: Successful Predictive Modeling Projects

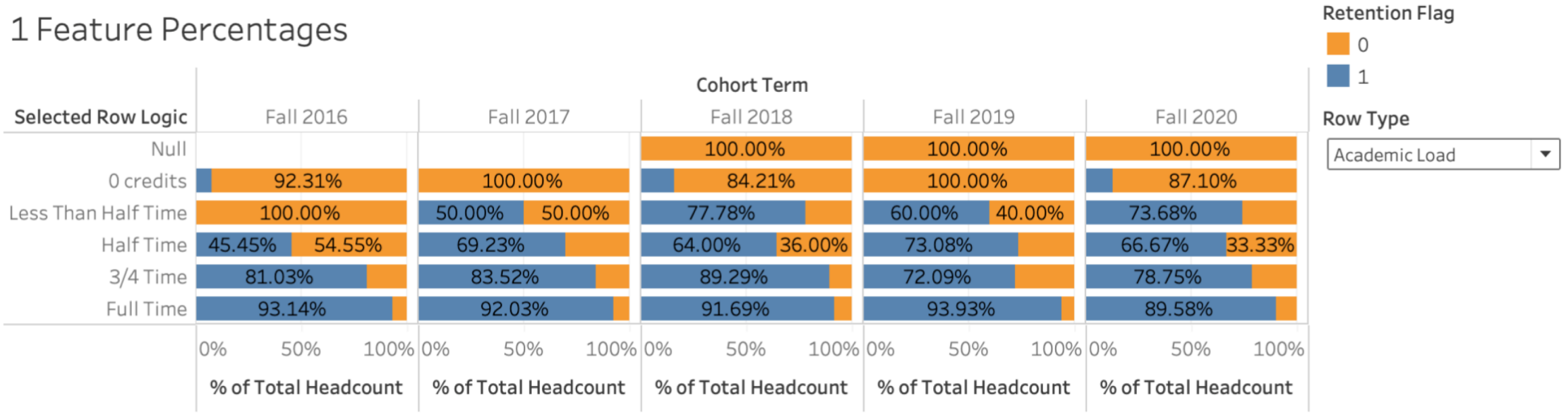
Explore the data before attempting to predict outcomes



HelioCAMPUS

Part IV: Successful Predictive Modeling Projects

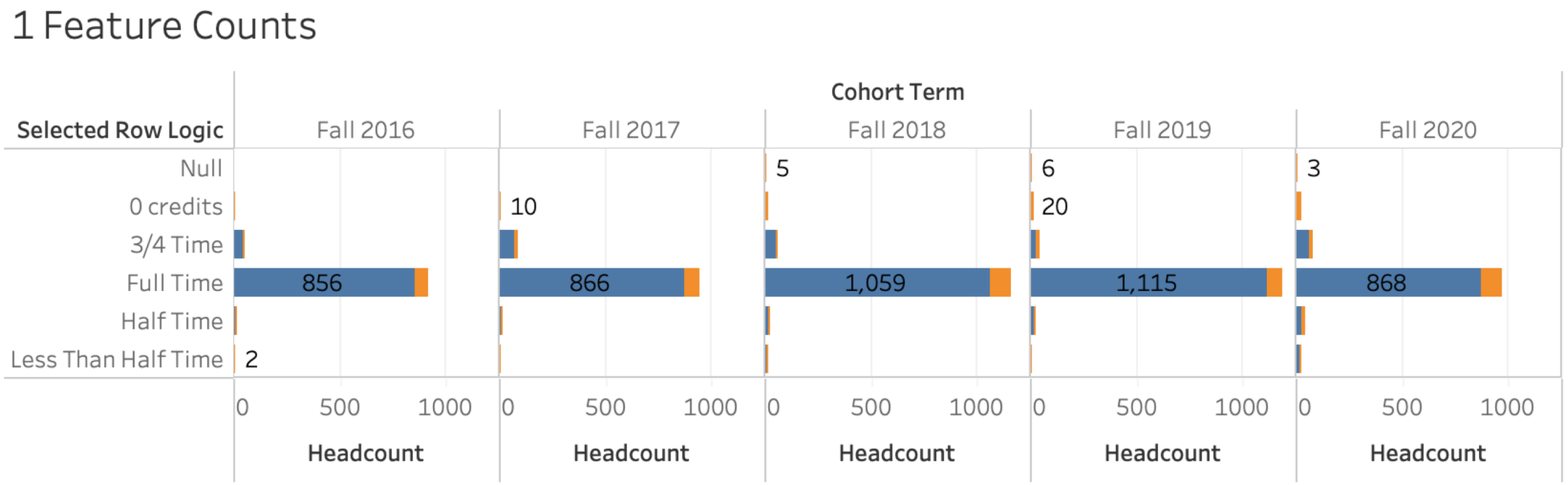
Explore the data before attempting to predict outcomes...



HelioCAMPUS

Part IV: Successful Predictive Modeling Projects

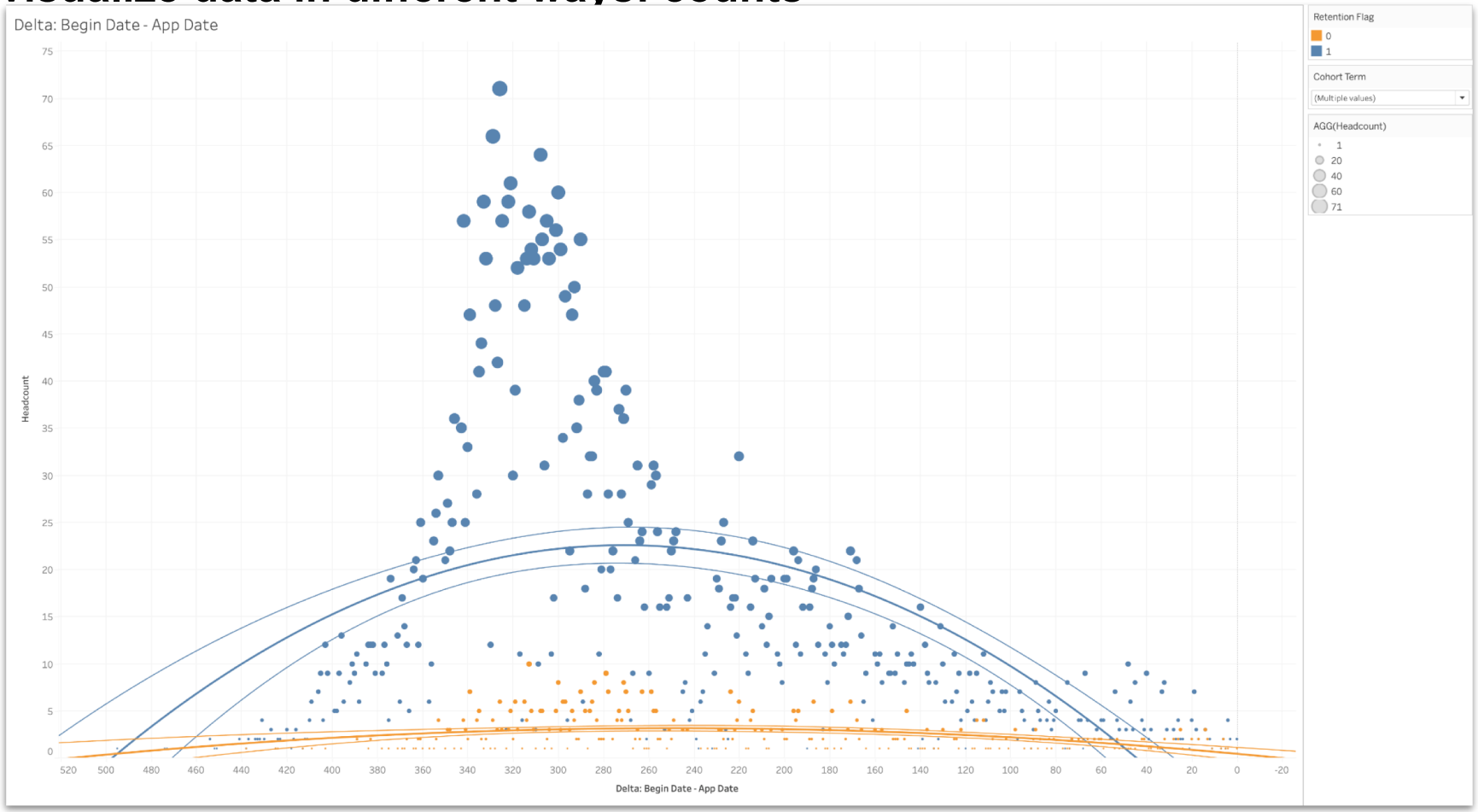
...And visualize that data different ways



HelioCAMPUS

Part IV: Successful Predictive Modeling Projects

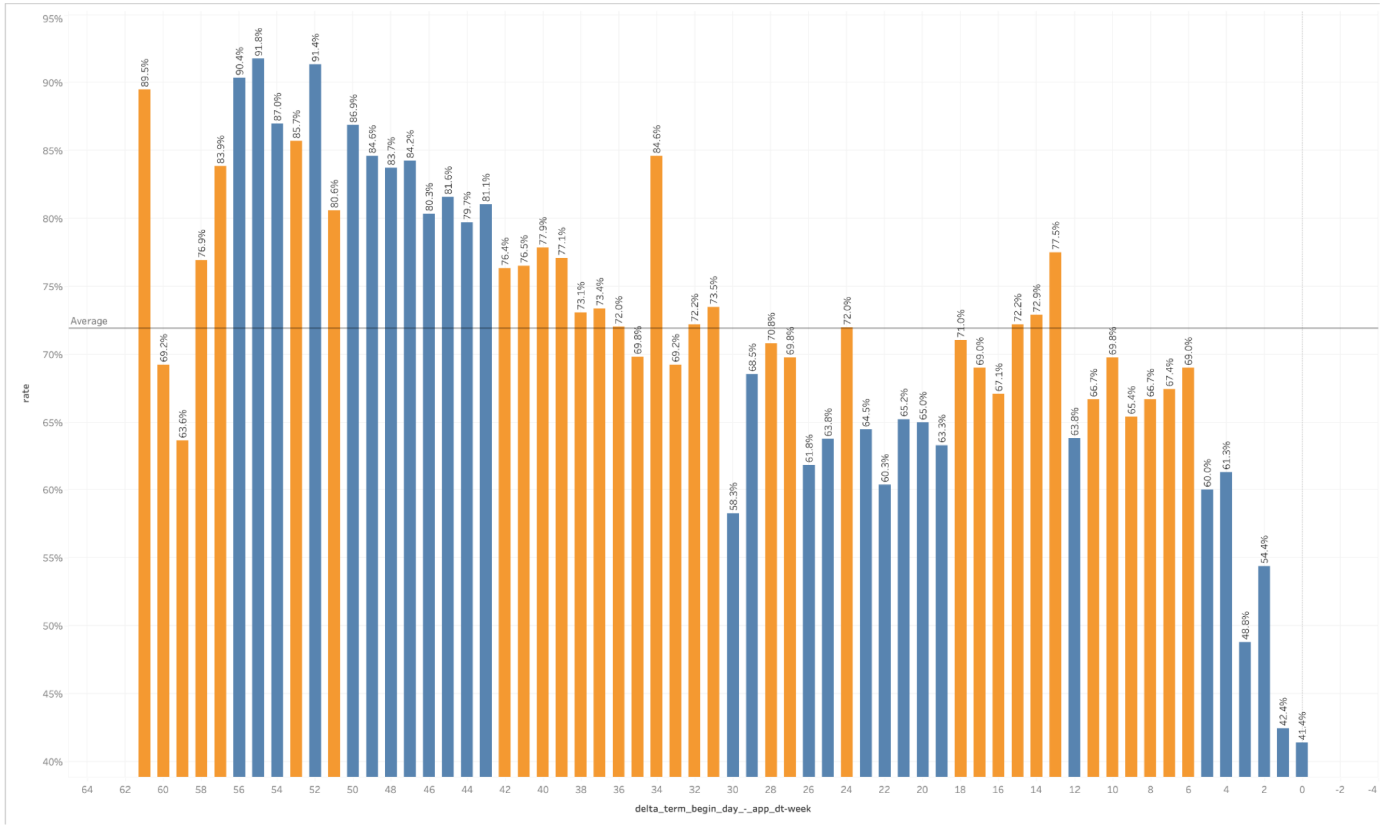
Same here: visualize data in different ways: counts



HelioCAMPUS

Part IV: Successful Predictive Modeling Projects

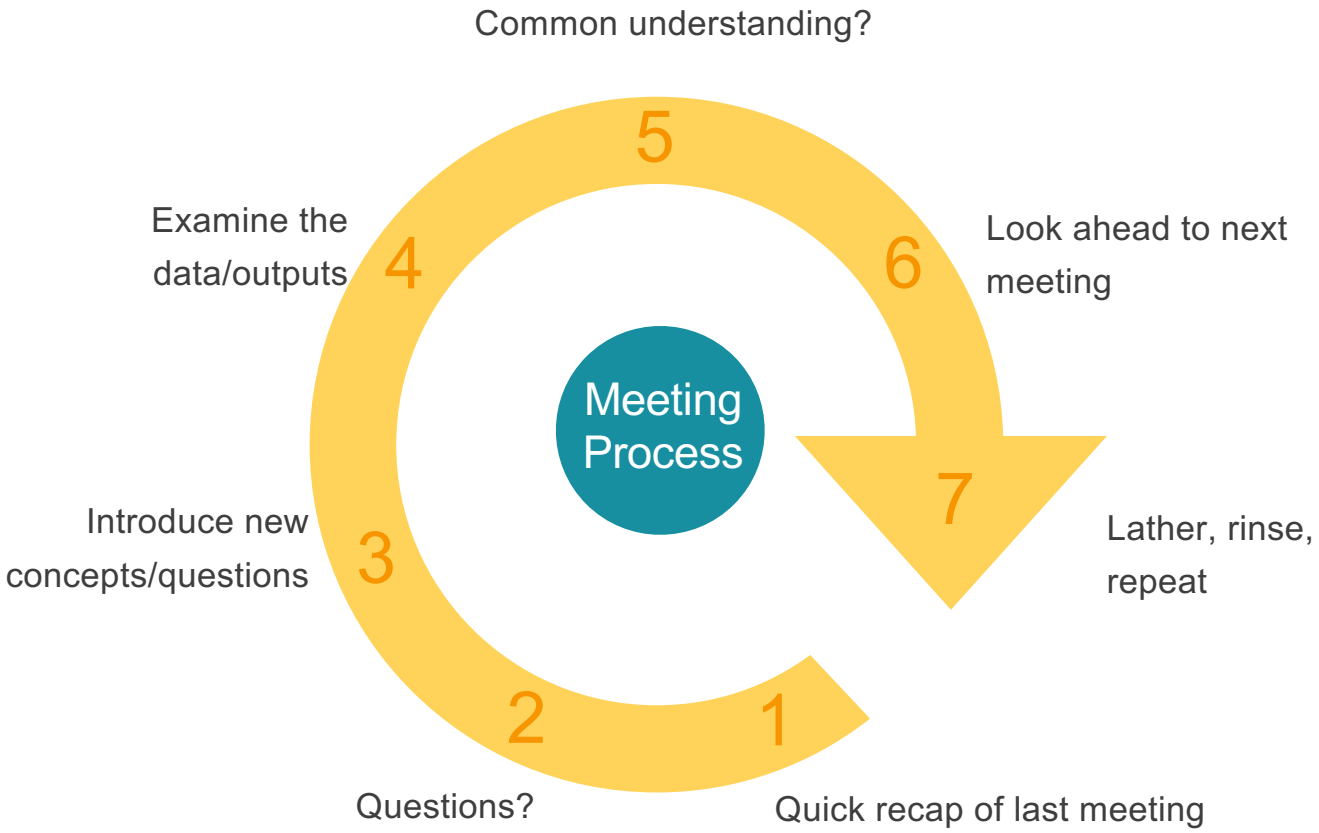
Same here: visualize data in different ways: percentages



HelioCAMPUS

Part IV: Successful Predictive Modeling Projects

Meet with stakeholders to discuss progress/deliverables/questions on a regular basis



Part IV: Successful Predictive Modeling Projects

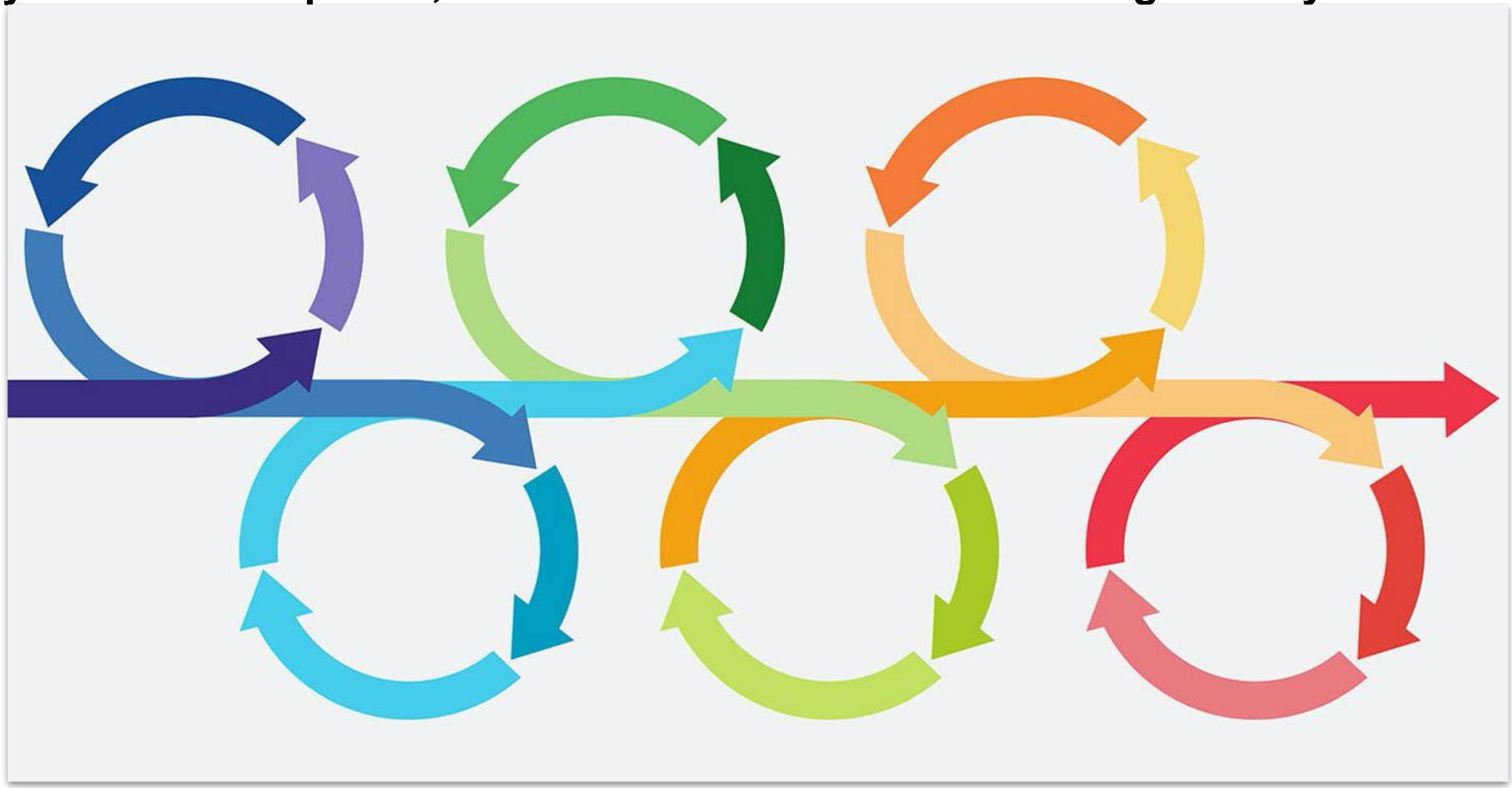
Have diverse roles & capacities across stakeholders

- **IR Specialists**
- **Advising**
- **Project management**
- **IT: Developers**
- **IT: Support**
- **Analysts**

Part IV: Characteristics of Successful Predictive Modeling Projects

Part IV: Successful Predictive Modeling Projects

Continually check assumptions, discoveries and conclusions along the way



Part IV: Successful Predictive Modeling Projects

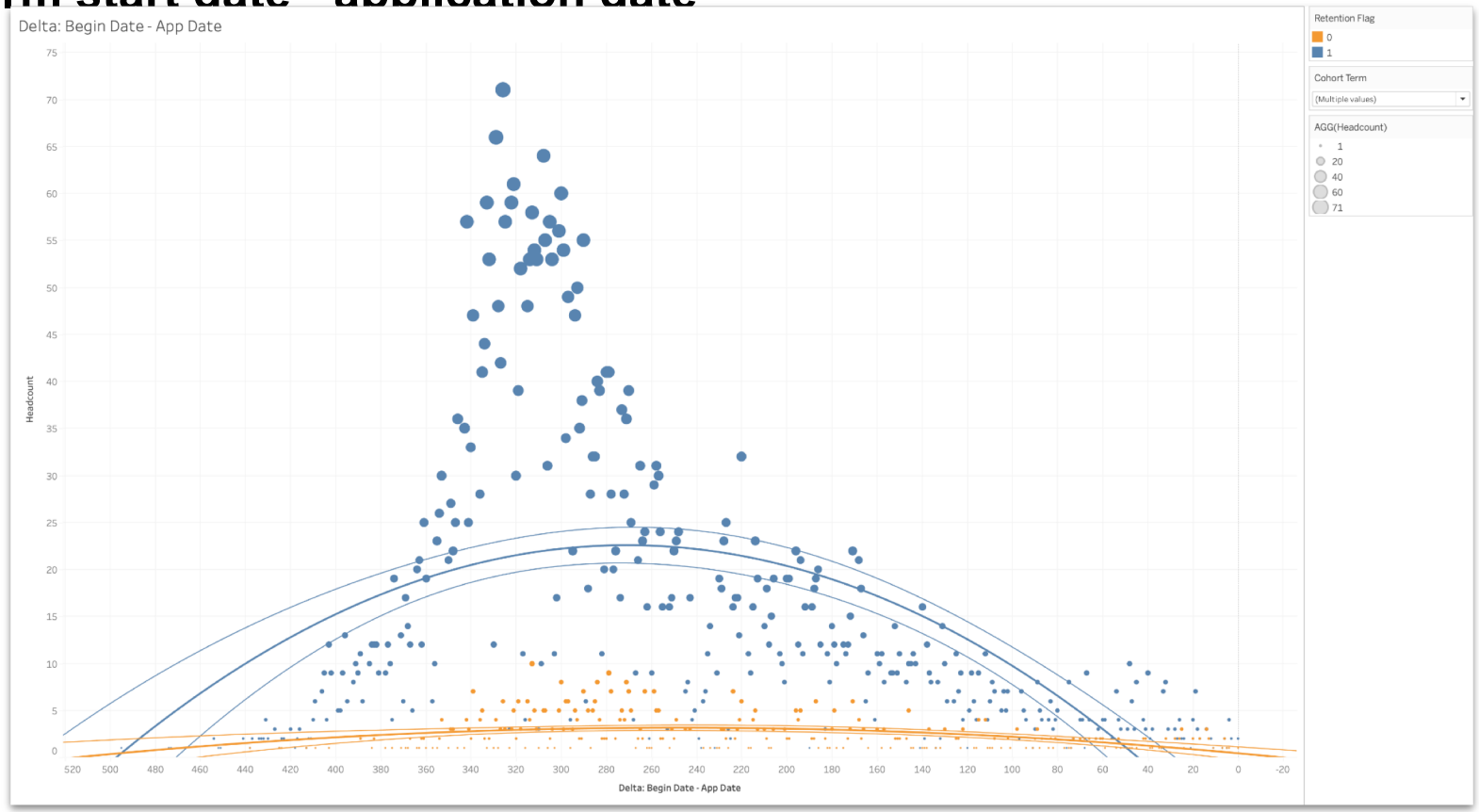
Start small & make incremental changes when training models

- **Basic model based on institutional knowledge: GPA, Hours, Financial Aid**
- **Hoover mode: Discovery**
- **Engineered features**
  - **Distance from campus**
  - **Hours vs GPA, Hours * GPA = power**
  - **Date transformations: Day of year, week of year, month of year**
  - **Date deltas**
  - **Can be augmented w/ new information after initial completion and delivery**

Part IV: Successful Predictive Modeling Projects: *Techniques*

Date transformations: Day of year, week of year, month of year

Date deltas: Term start date - application date

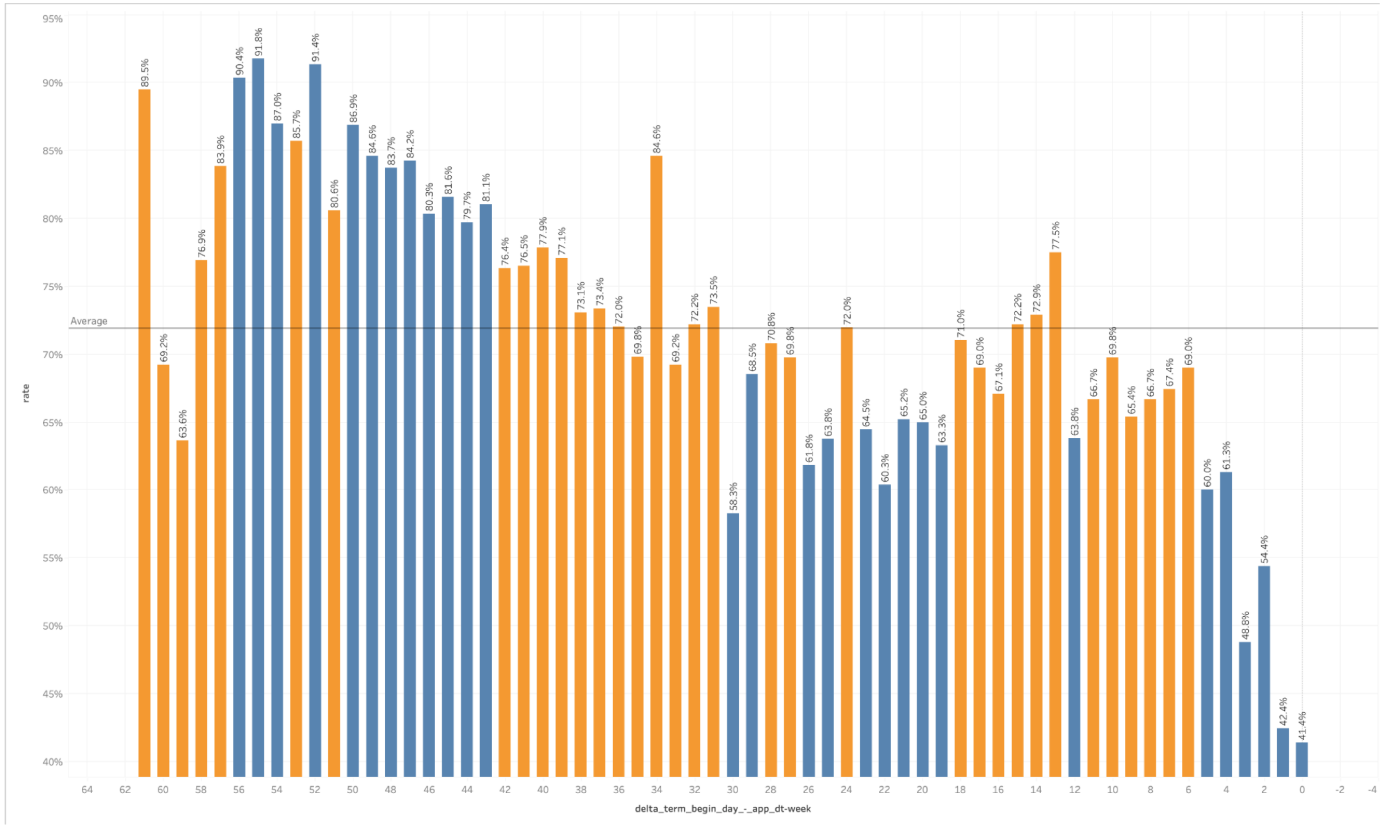


HelioCAMPUS

Part IV: Successful Predictive Modeling Projects: *Techniques*

Date transformations: Day of year, week of year, month of year

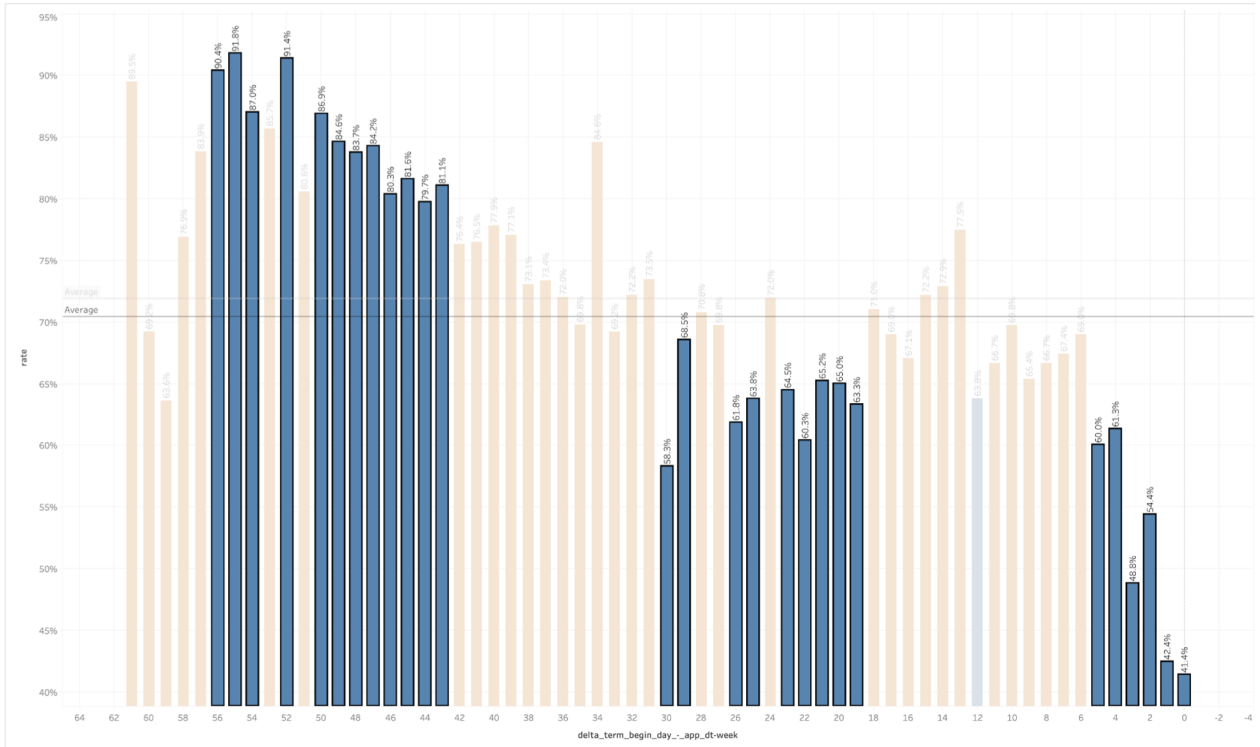
Date deltas: Term start date - application date



HelioCAMPUS

Part IV: Successful Predictive Modeling Projects: *Techniques*

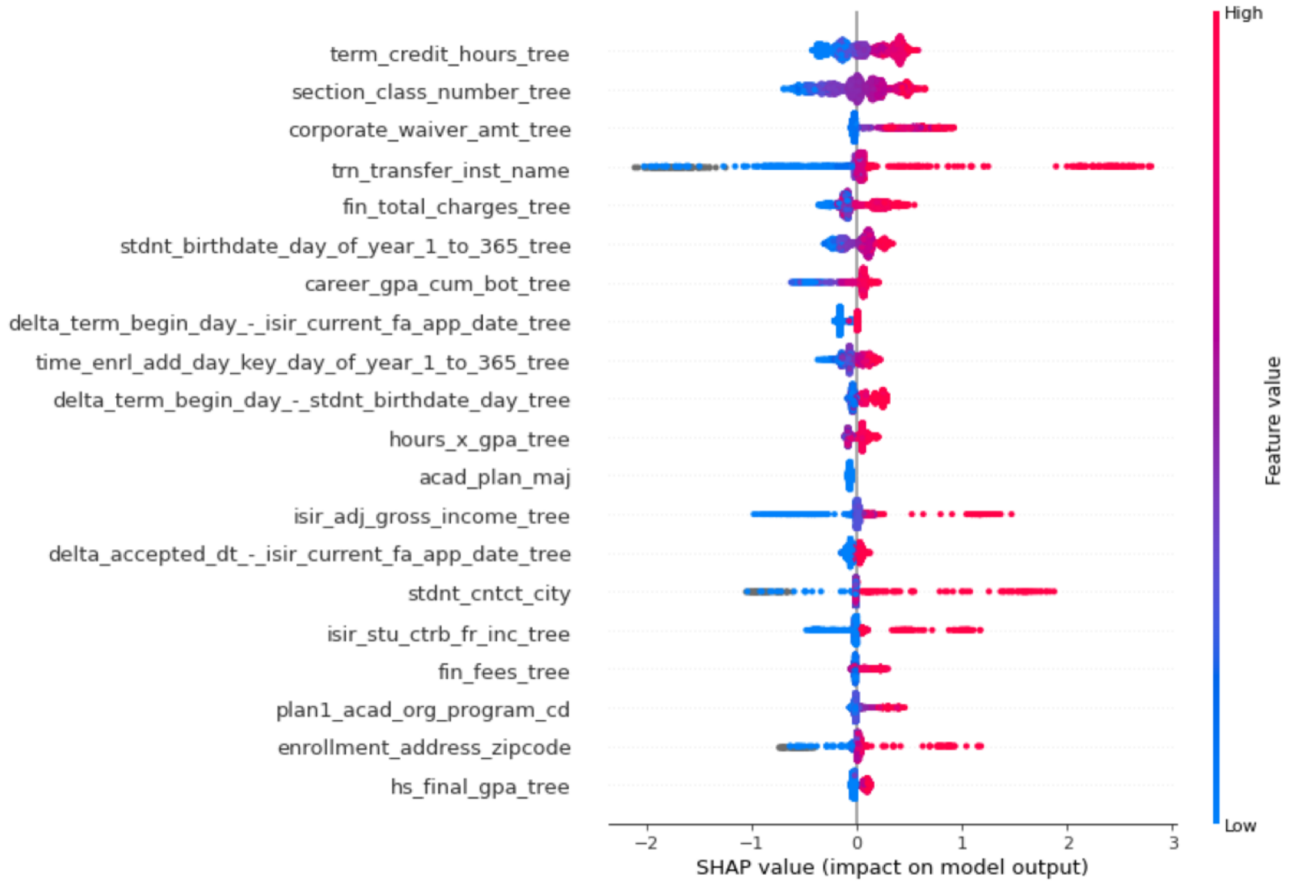
At one client's behest I performed special date analysis (non standard)



HelioCAMPUS

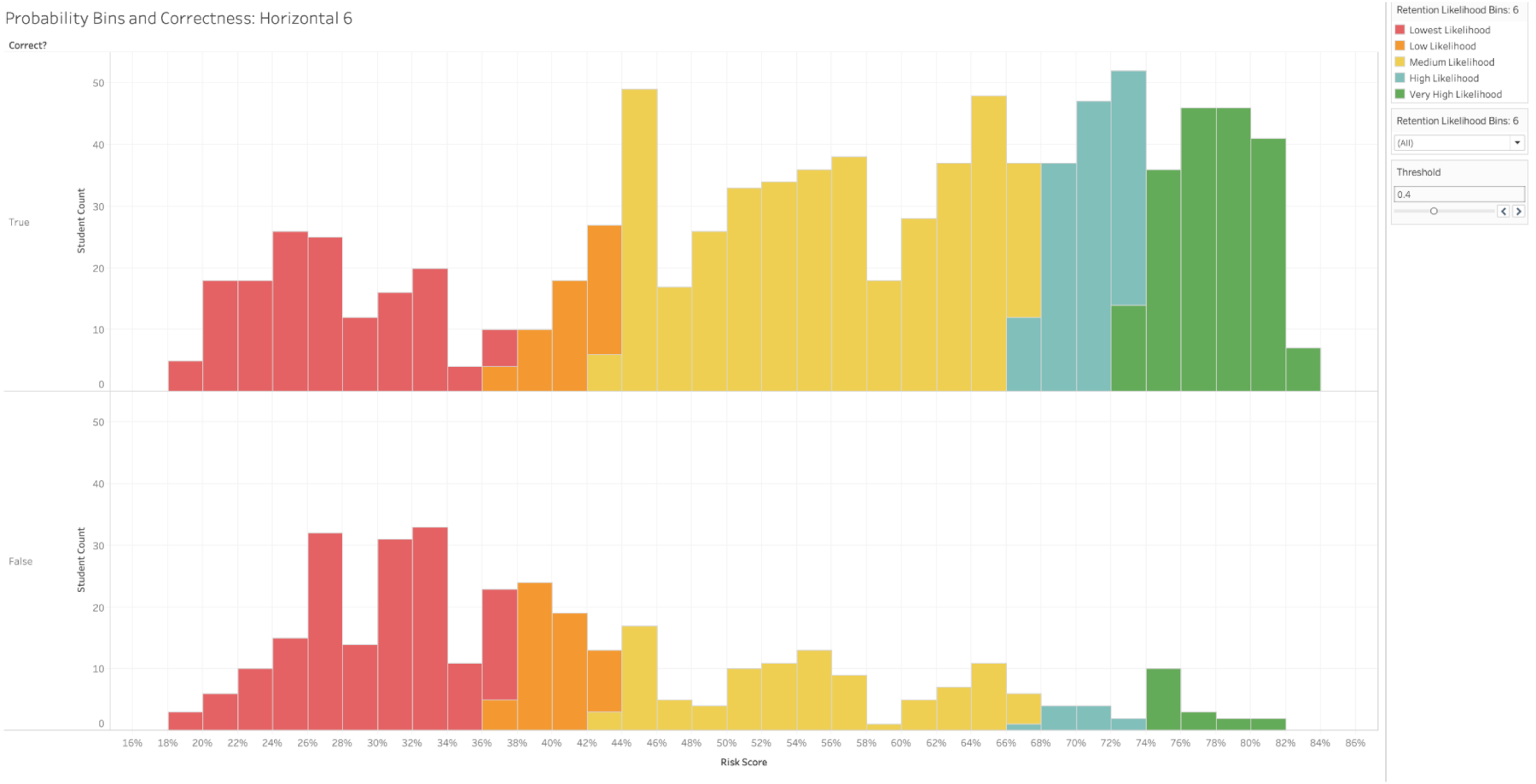
Part IV: Successful Predictive Modeling Projects

Have explainable transformations: Hours vs GPA vs Hours * GPA



Part IV: Successful Predictive Modeling Projects

Validate against known historical outcomes



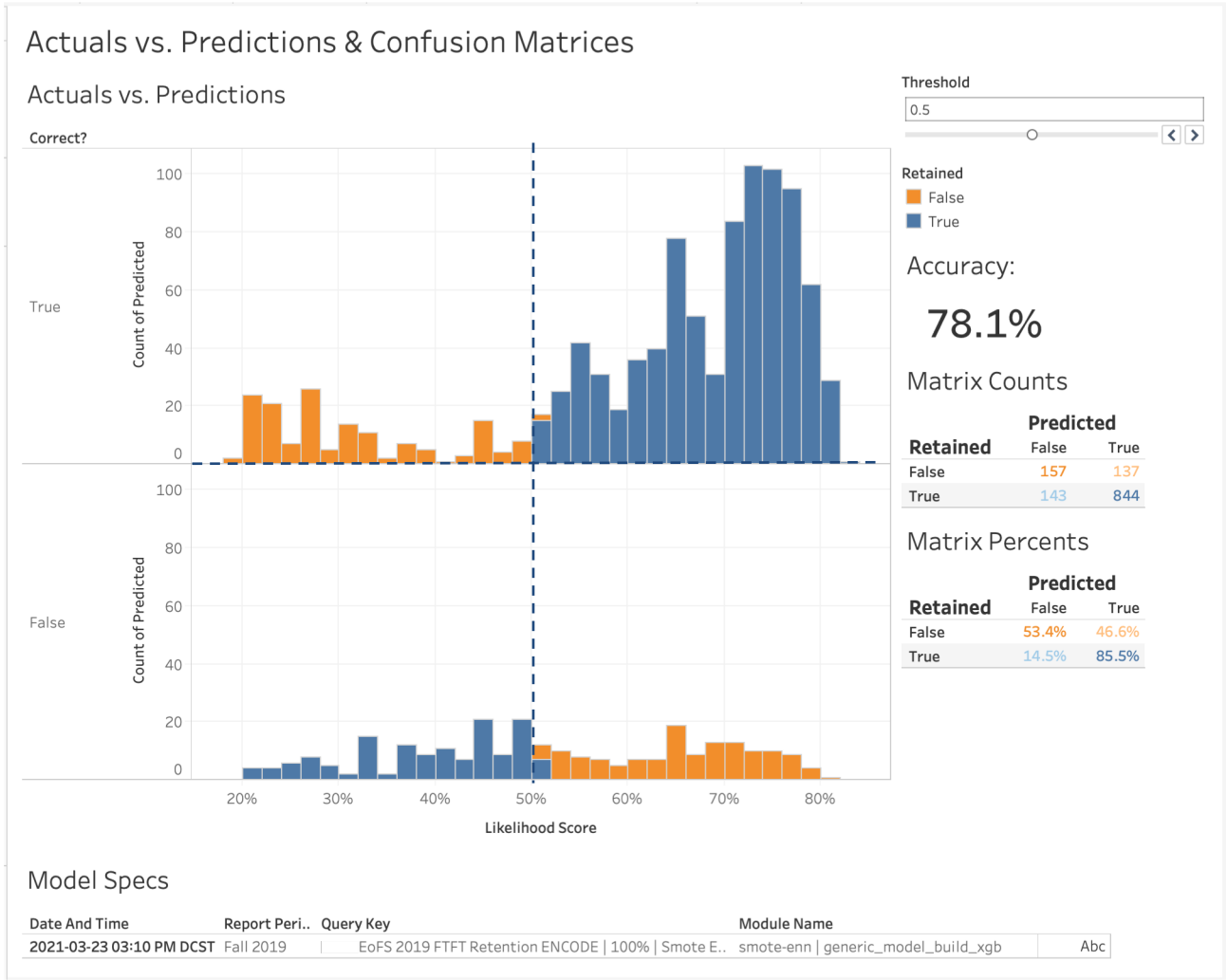
HelioCAMPUS

Part IV: Successful Predictive Modeling Projects

- Acknowledge that there's no free lunch
  - Noise: predictions are only as good as your data
  - Cost/benefit: Which errors are least/most costly?
- Are clear about what they're trying to predict:
  - Retention vs. stopping out = majority vs minority classes

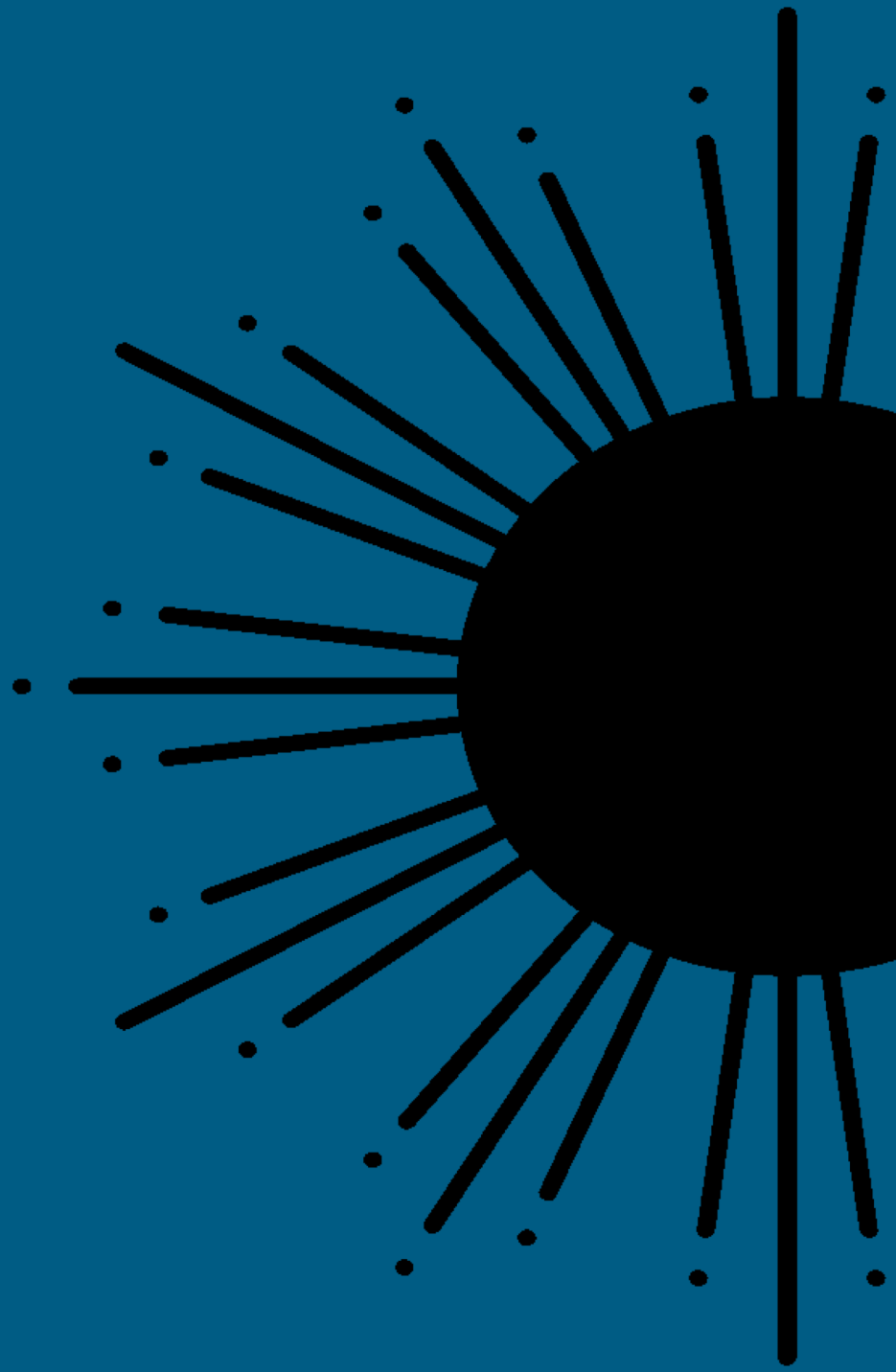
Part IV: Successful Predictive Modeling Projects: *Concepts*

No free lunch



HelioCAMPUS

Q&A



Thank you

HelioCAMPUS