

Student Outcomes From the 4TH Annual Data4Good Analytics Competition

BUTLER



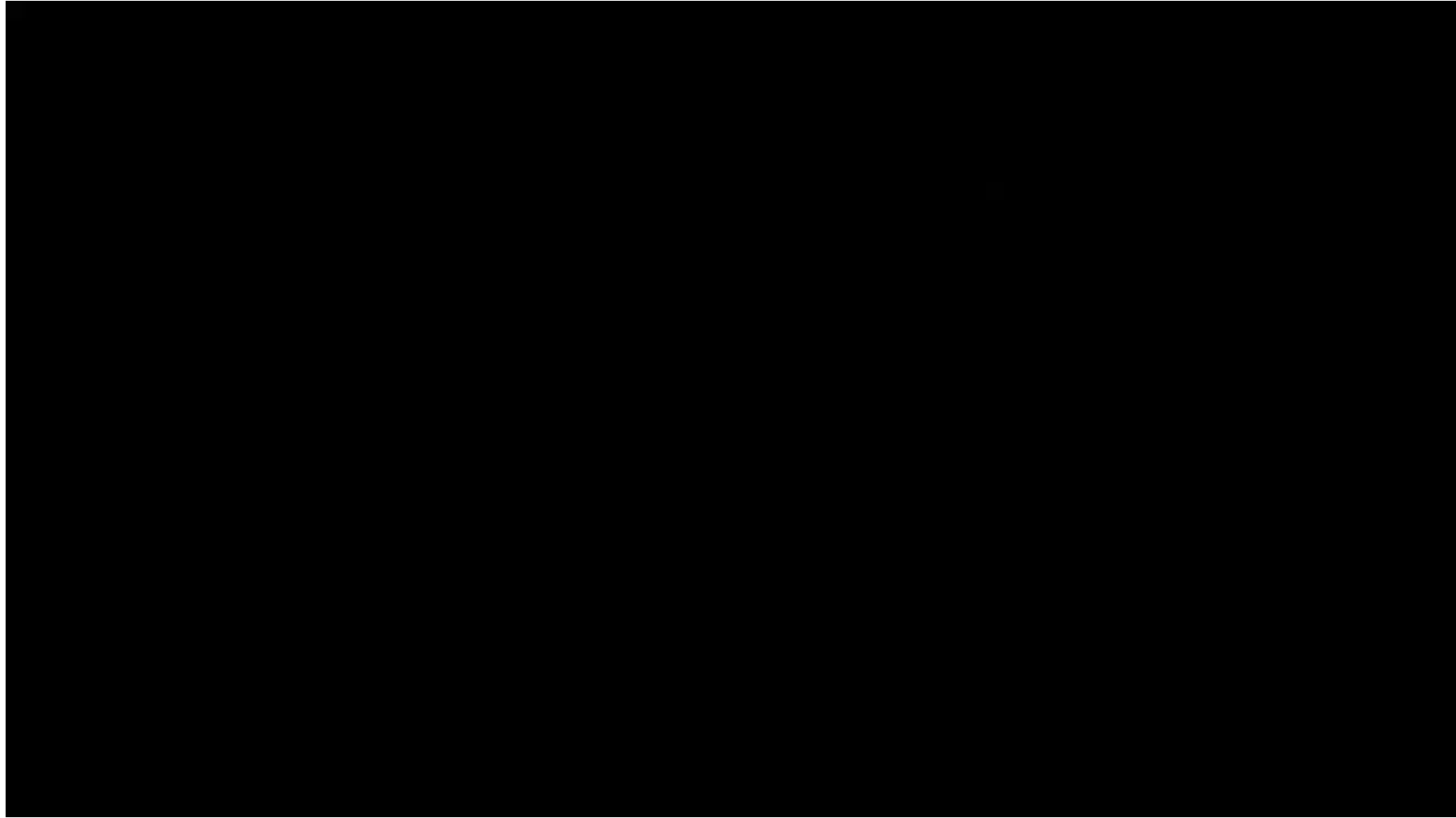
Matthew A. Lanham, Ph.D., CAP-X, ACE
Assistant Professor of Business Technology & Analytics
Butler University Lacy School of Business
mlanham1@butler.edu
MatthewALanham.com

ZETEC



BUTLER

Opportunities for internships at ZETEC



BUTLER

Mission and vision

Mission: Provide college students an opportunity to improve their knowledge, skills, and abilities of data analytics for good.

Vision: To be the premier data analytics for good competition in the United States.

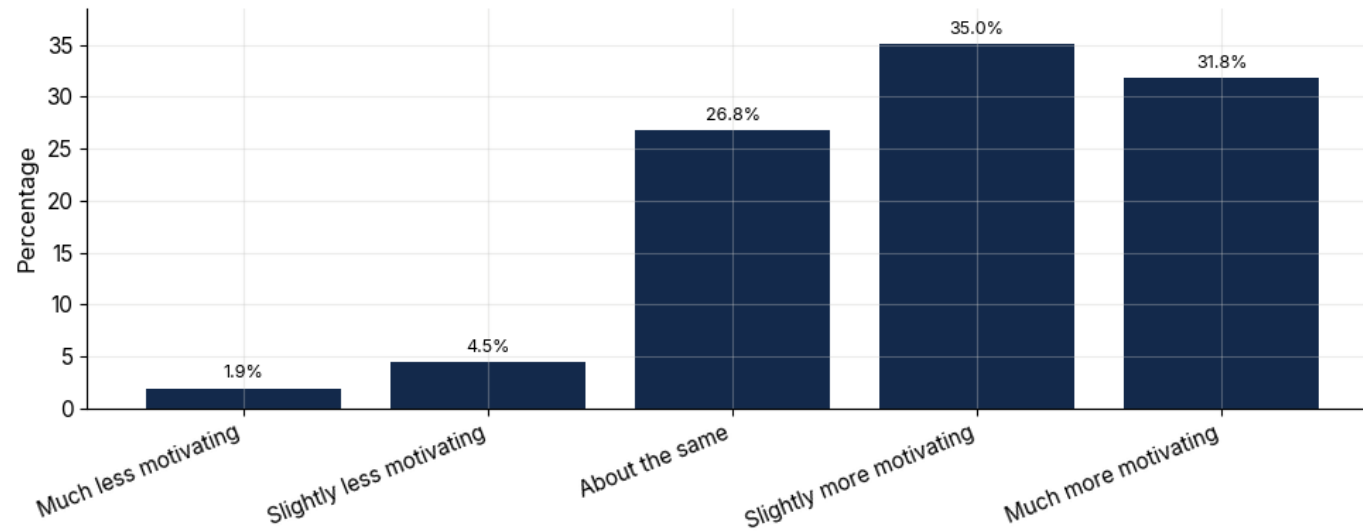
Mission and vision

Mission: Provide college students **and practitioners** an opportunity to improve their knowledge, skills, and abilities of data analytics for good.

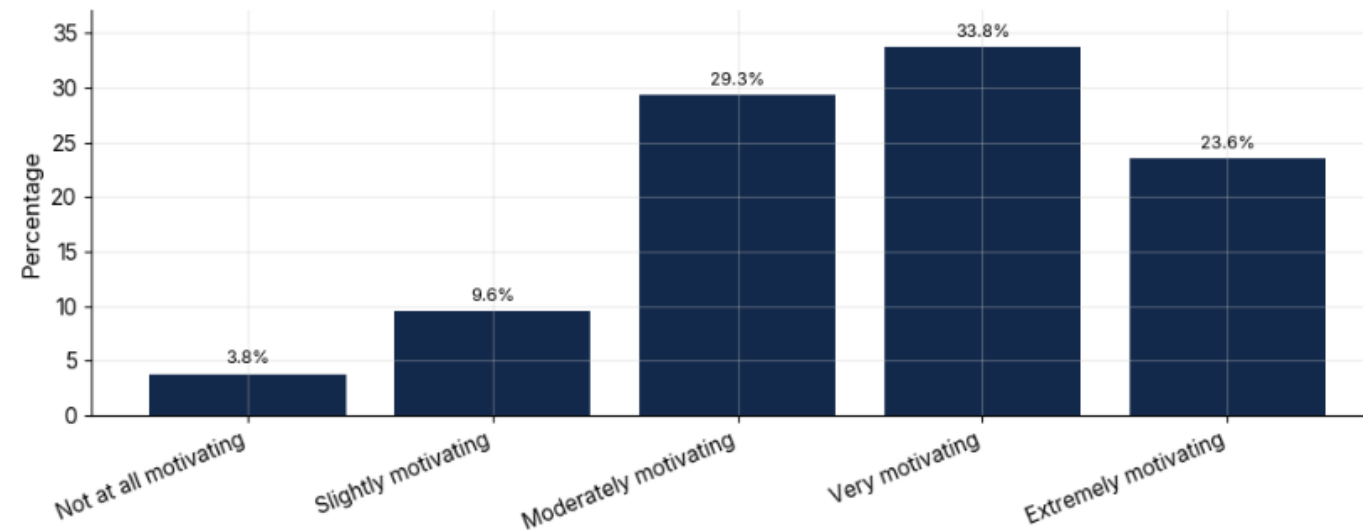
Vision: To be the premier data analytics for good competition in the ~~United States~~ **world**.

Students prefer an altruistic “good” case

Compared to competitions focused primarily on commercial or profit-driven objectives, how motivating did you find this socially or altruistically oriented problem?



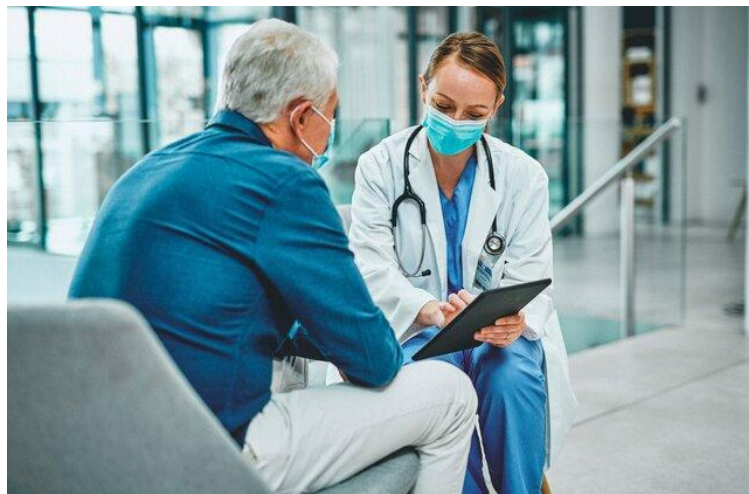
To what extent did the altruistic or ‘for good’ nature of the competition problem increase your motivation to engage deeply with the competition?



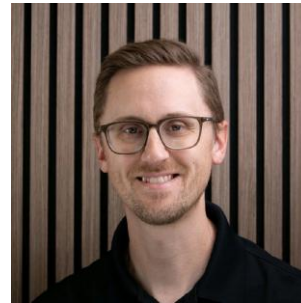
A “good” problem each year



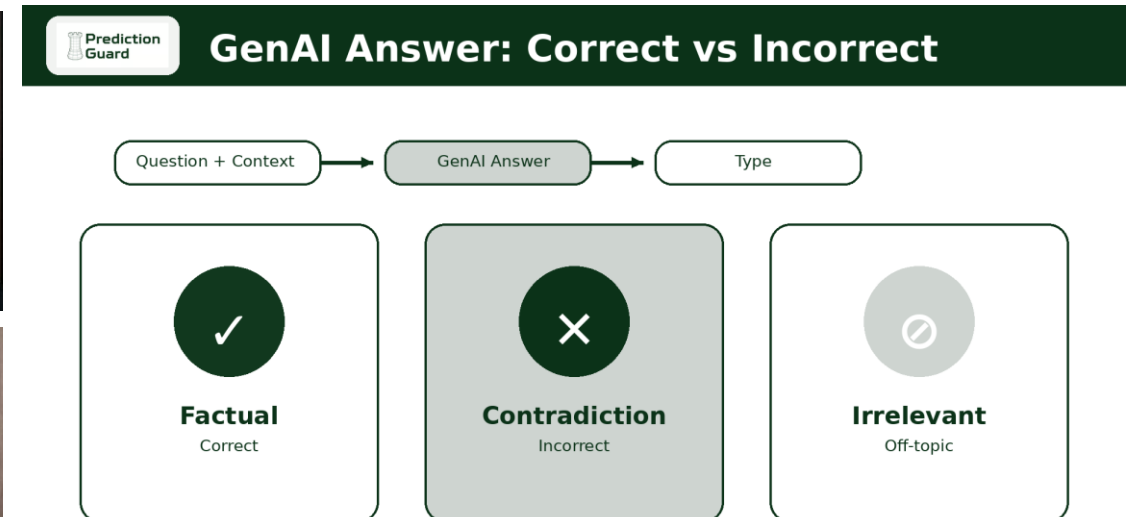
2022 – Captioning images in Children’s books
(SIL Global / Dan Whitenack)



2023 – LLMs in Healthcare (Prediction
Guard’s client / Dan Whitenack)



2024 – Understanding stages of grief
(TAPS.org)

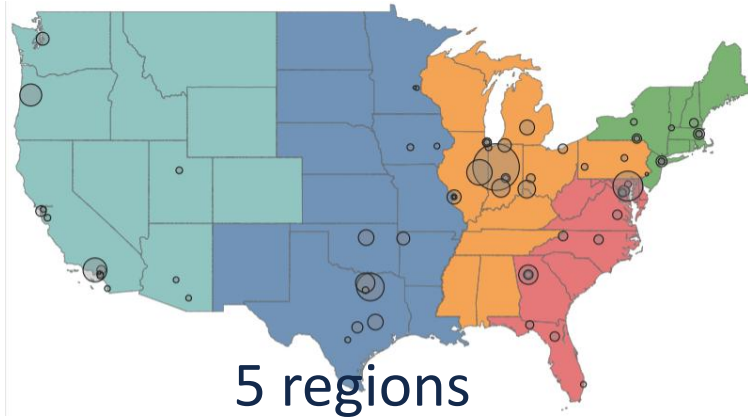


2025 – GenAI reliability (Prediction Guard / Dan
Whitenack / Todd Bordner)

Largest reach ever this year!

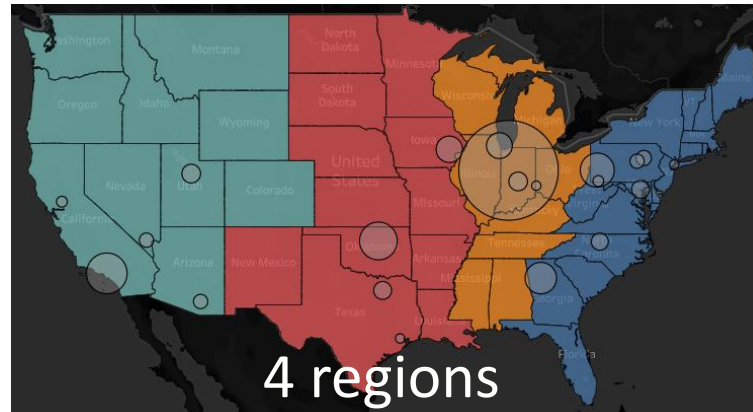
2023

- \$45,000 in prizes
- 95% passed AZ-900
- 290 teams
- 1031 students
- 72 universities



2024

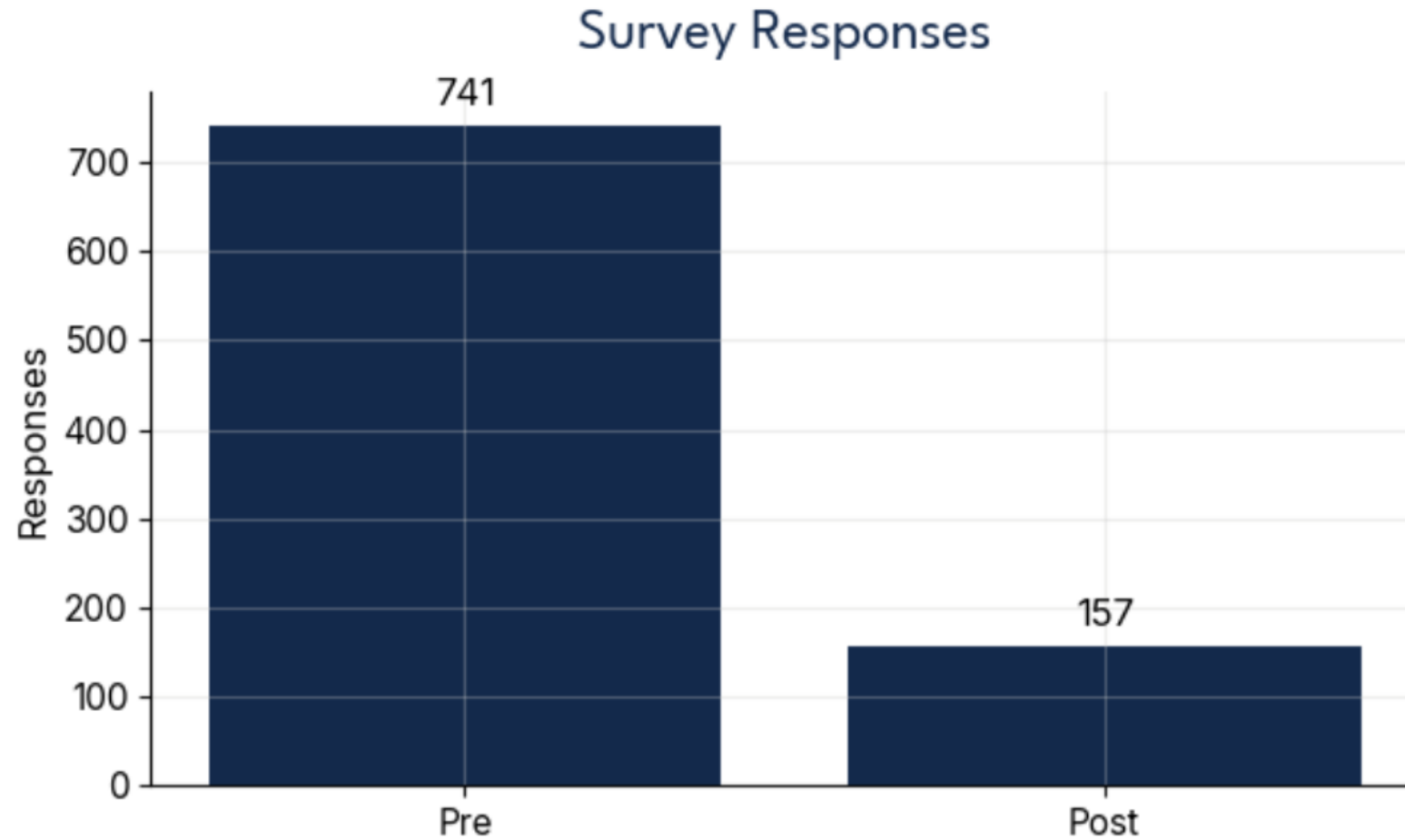
- \$40,000 in prizes
- 96% passed AI-900
- 243 teams
- 931 students
- 35 universities



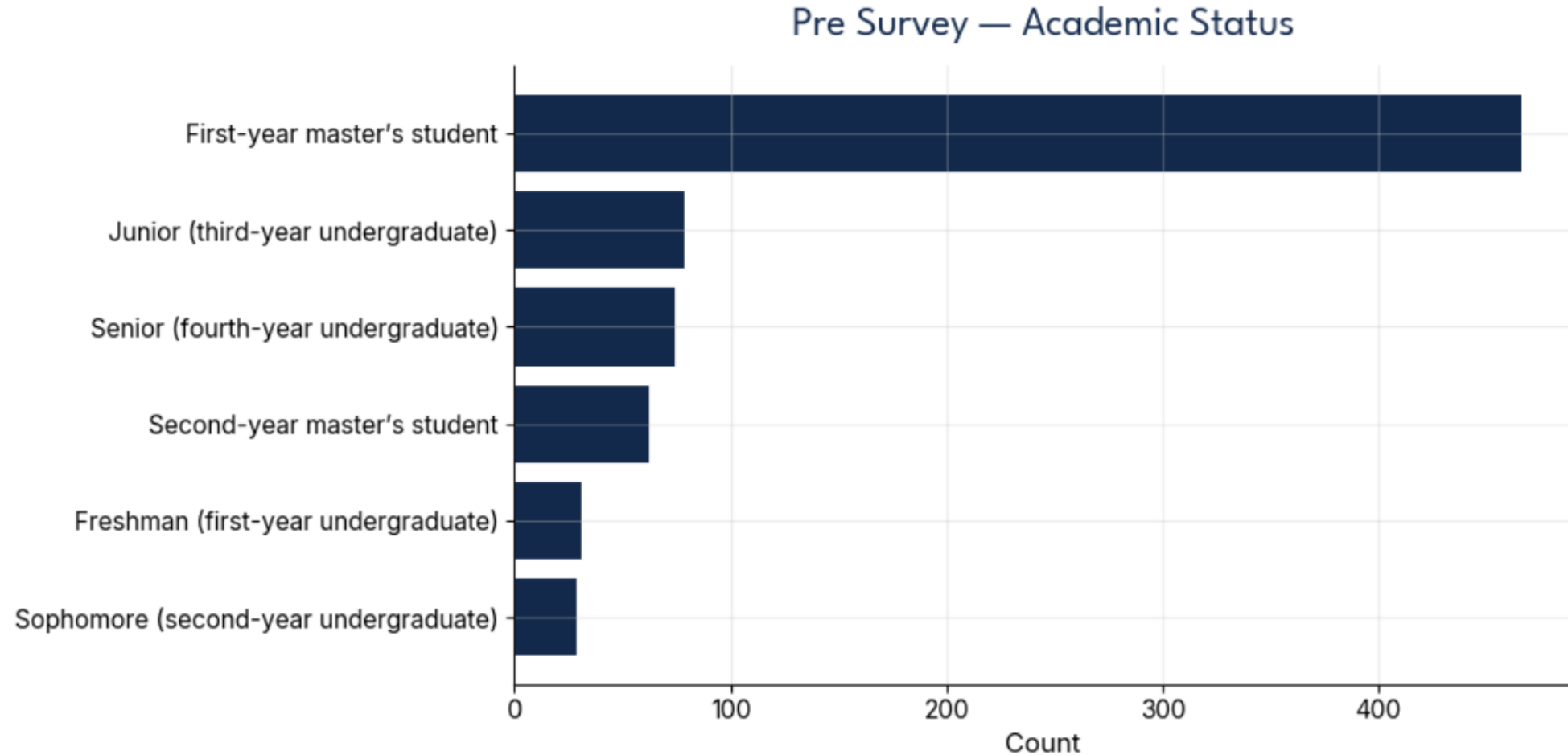
2025

- \$31,000 in prizes
- 96% passed AI-900
- 295 teams
- 1106 students
- 105 universities
- 15 non-US schools

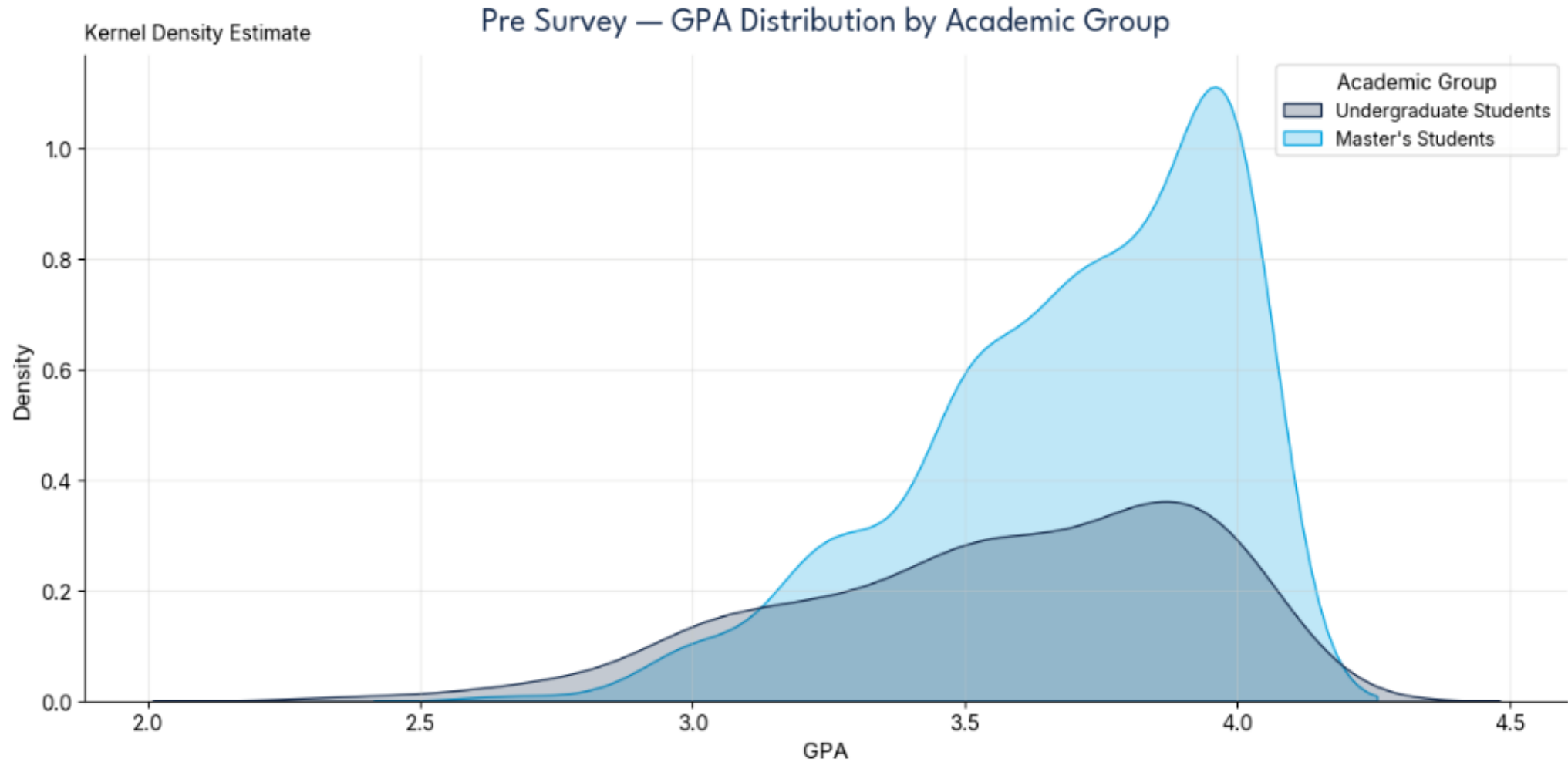
Student Feedback



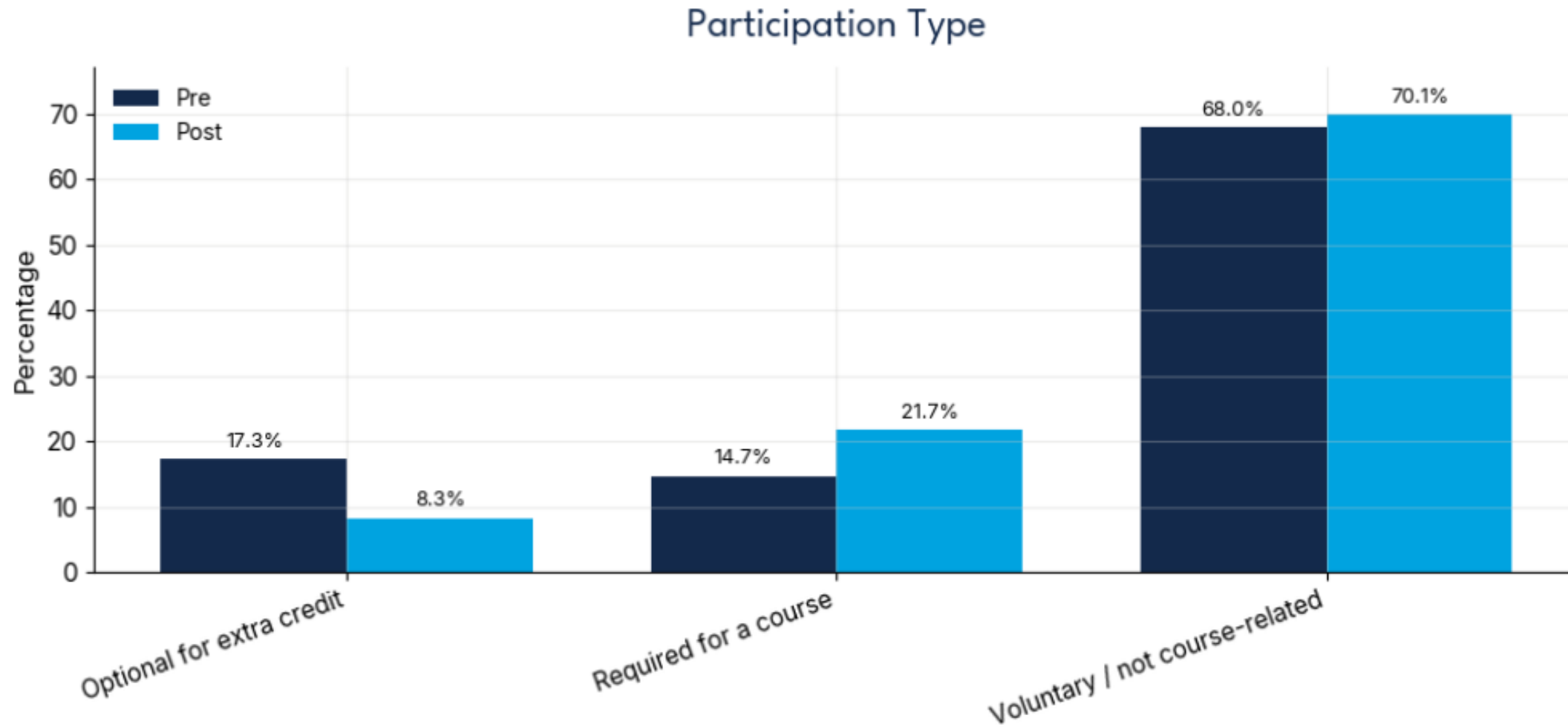
Strong turnout among masters' students



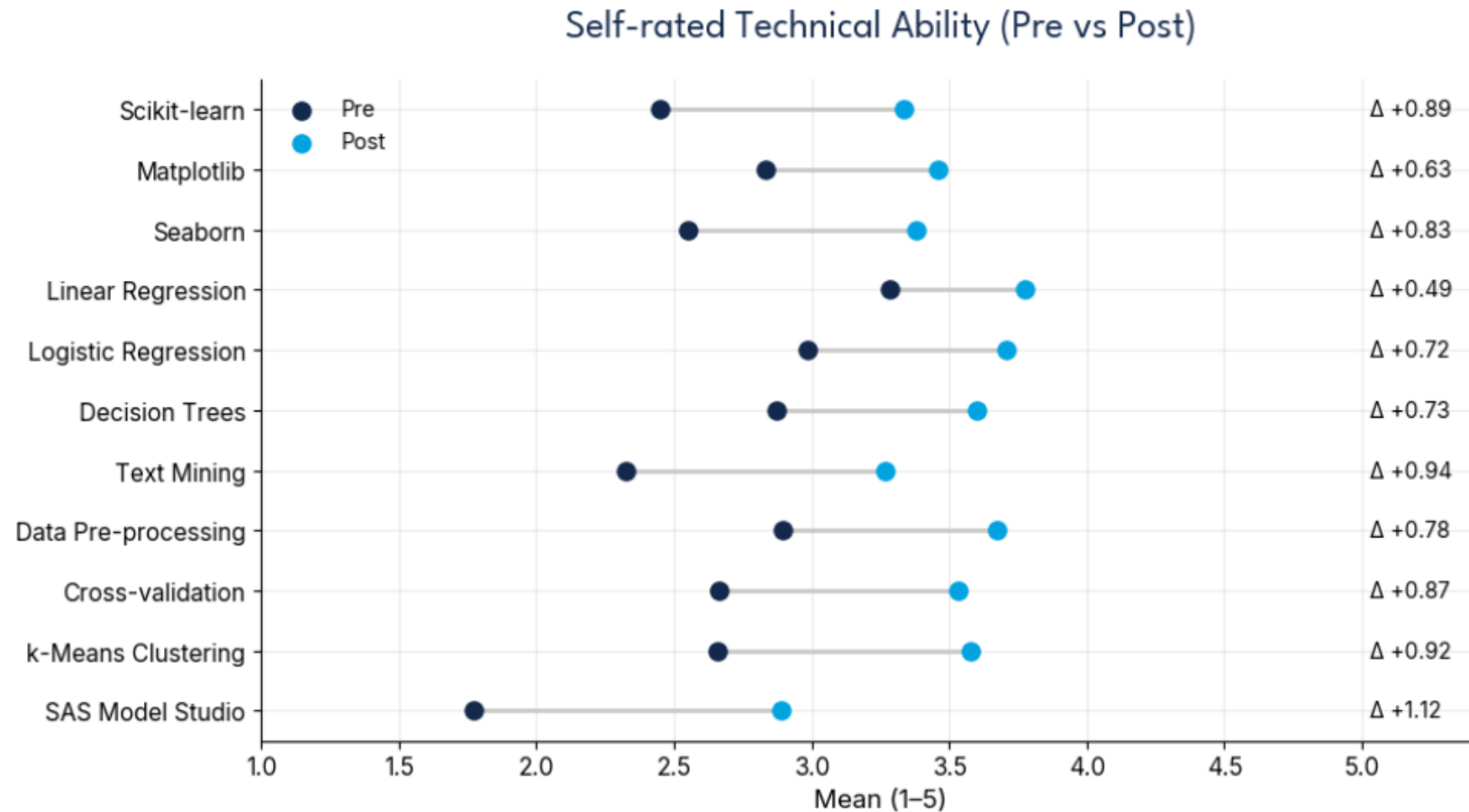
The competition is not just for the high performers



Most people continue to participate outside of class



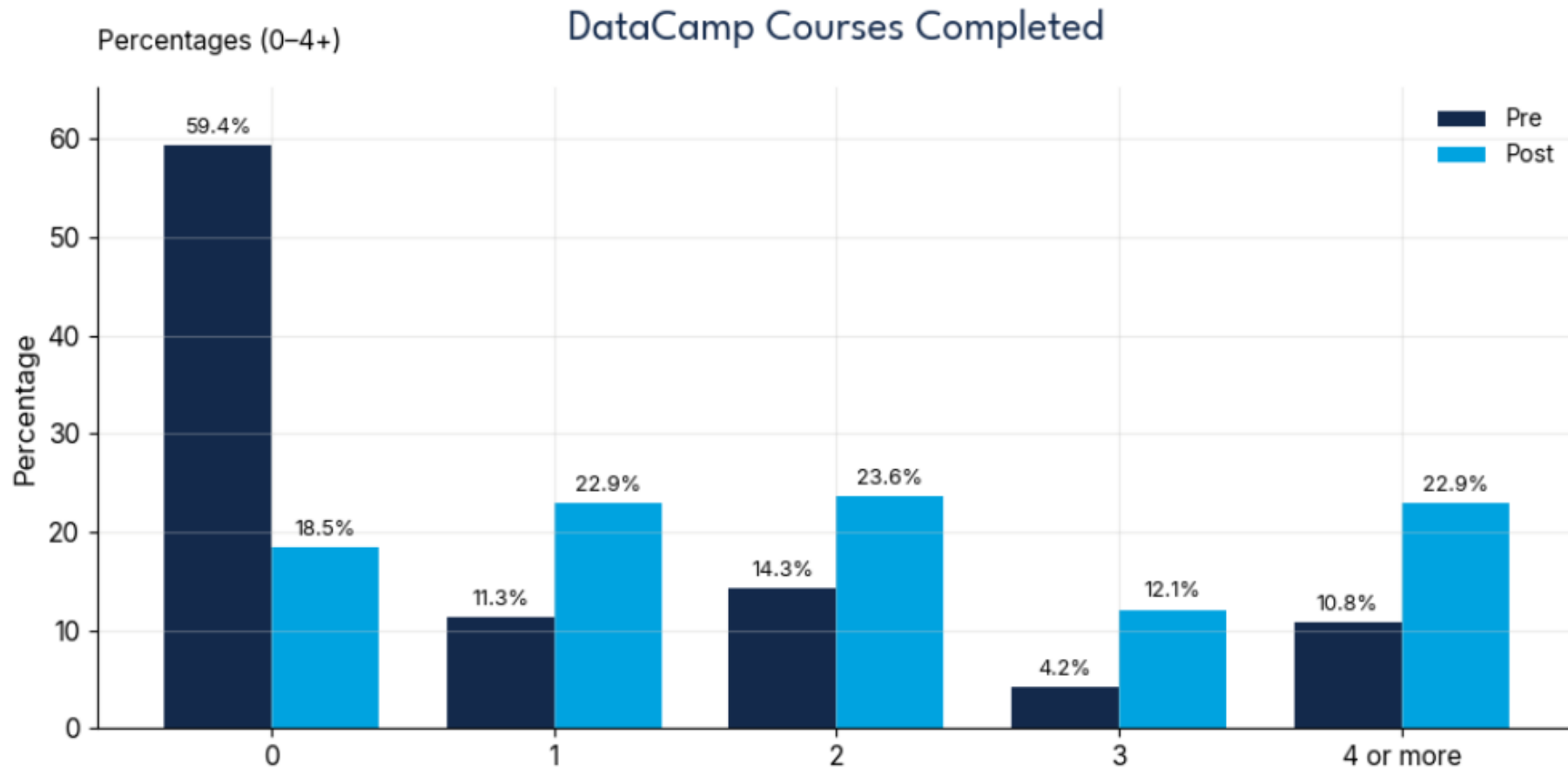
Most people continue to participate outside of class



Cronbach's α (skills battery) – Pre : 0.948

Cronbach's α (skills battery) – Post: 0.962

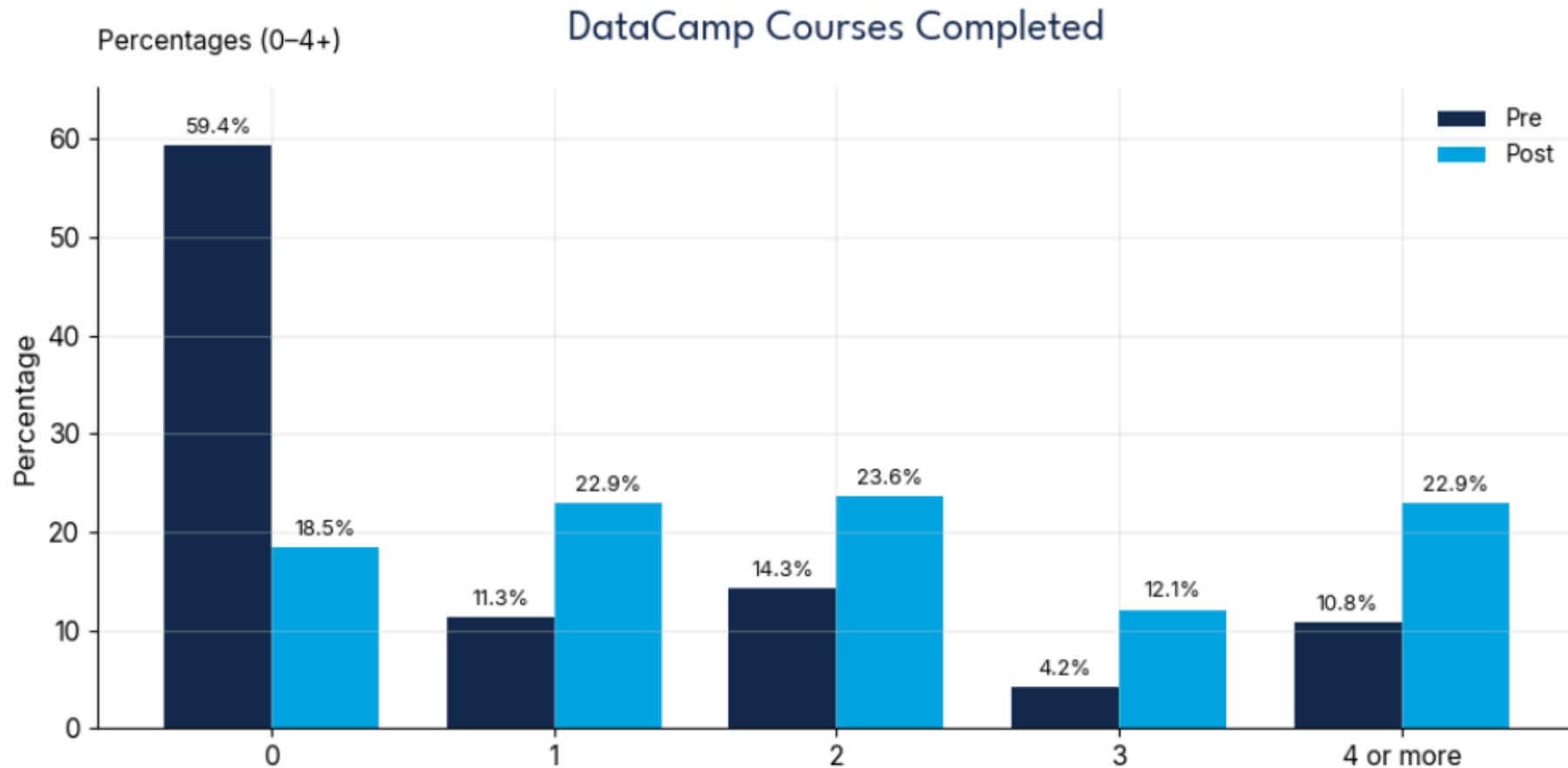
While not part of the competition points, participants completed DataCamp courses on their own



BUTLER

Note: Obtaining actual counts and course students in competition completed from DataCamp

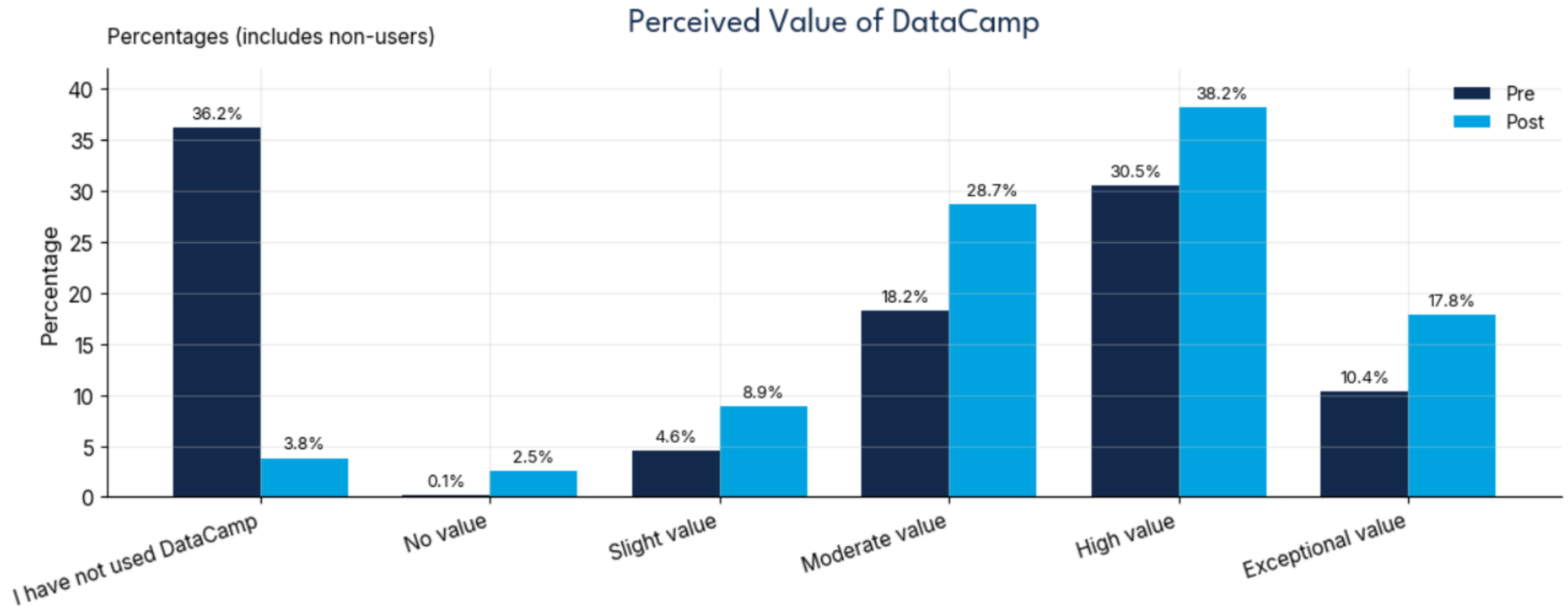
While not part of the competition points, participants completed DataCamp courses on their own



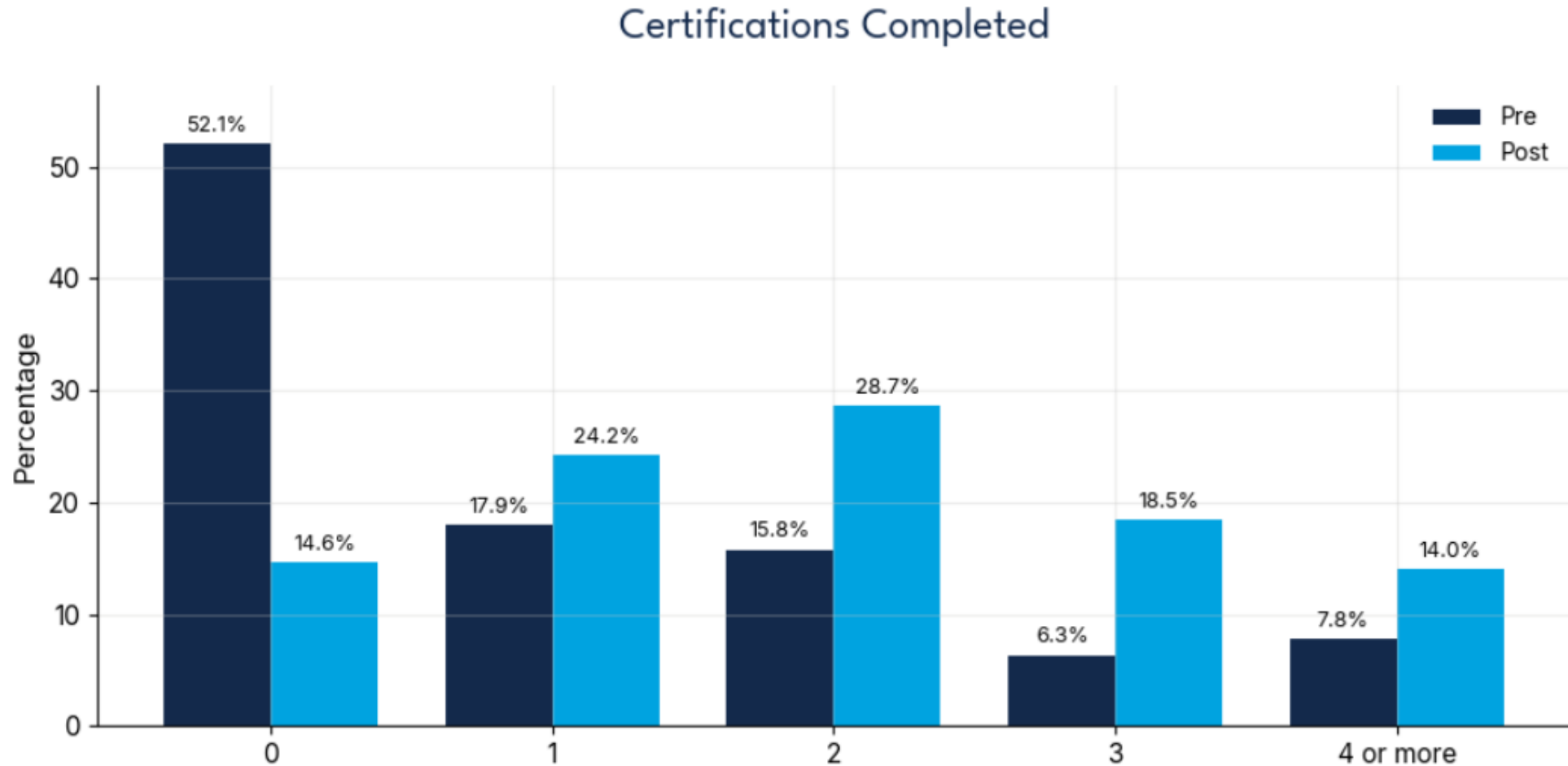
BUTLER

Note: Obtaining actual counts and course students in competition completed from DataCamp

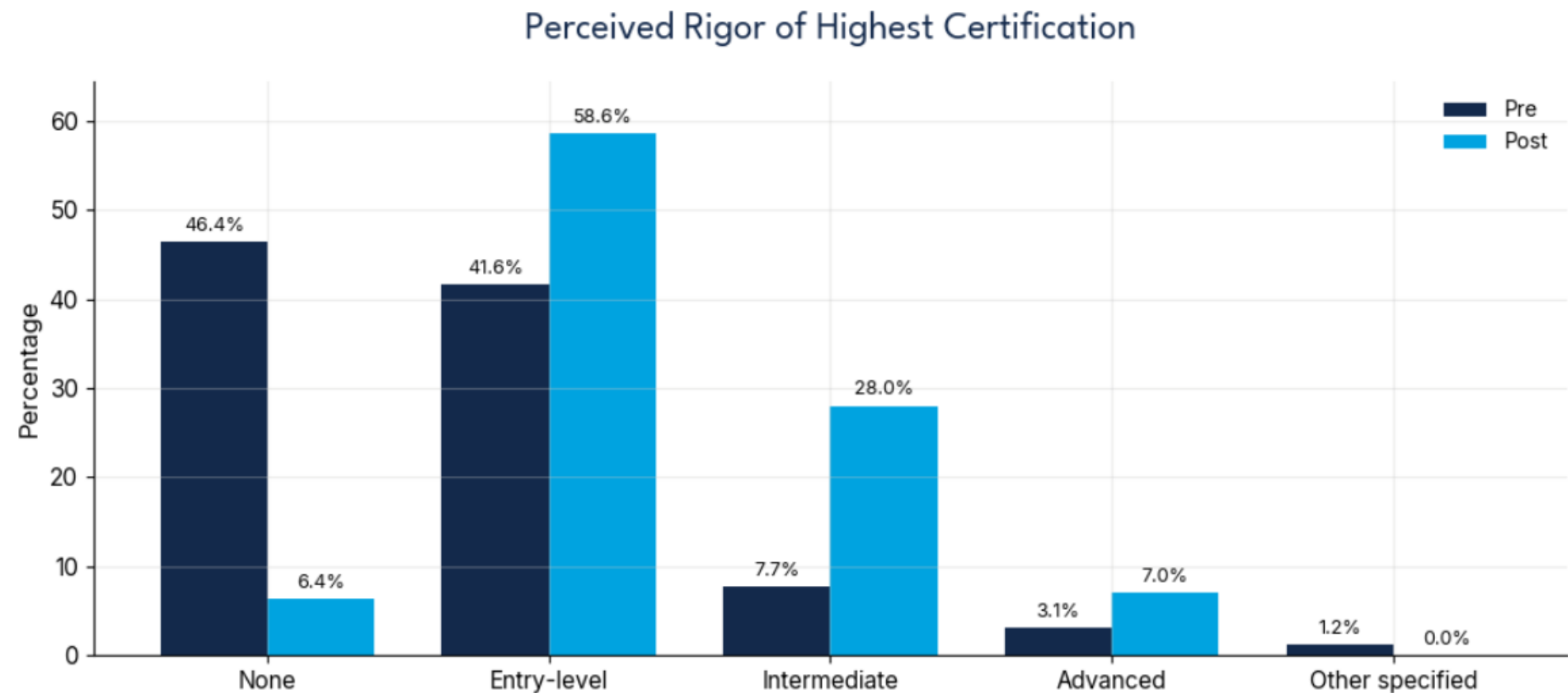
Participants perception of DataCamp improved



Participants are earning more certifications



Participants are earning more certifications



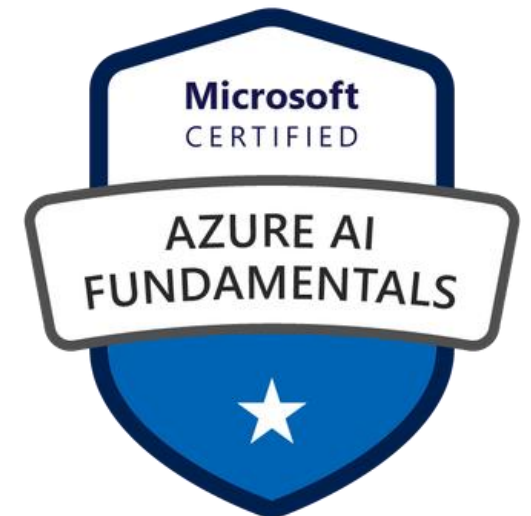
Note: AI-900 was anchored as an “Entry-level” certification for both pre- and post- questions

96% pass rate again on AI-900!

240 participants passed the AI-900
at a 96% pass rate!

Regional breakdown					
Region	Candidates	Passed	Failed	Pass rate	Avg score (Passed)
Region 1	89	86	3	96.6%	872
Region 2	133	127	6	95.5%	891
Region 3	4	4	0	100.0%	900
Region 4	25	23	2	92.0%	890

Top schools (by Passed)	Count
Purdue University	59
Butler University	43
Johns hopkins university	23
American University	9



BUTLER

Nearly 200 SAS Knowledge Badges earned!



Rochelle Fisher  • 1st

Higher Education | Strategy & Innovation | Ethical AI

2w • 

SAS was a proud sponsor of the Data4Good2025 competition, empowering students to fight misinformation in education using data and AI. Over the course of the competition:

- 200 students received training in SAS software and academic resources
- Nearly 200 SAS Knowledge Badges earned

Thank you to everyone who joined us in shaping a future where analytics makes a difference! We look forward to connecting with finalists **The Johns Hopkins University** in February.

Lori Downen Data4Good Analytics Competition **Matthew A. Lanham**

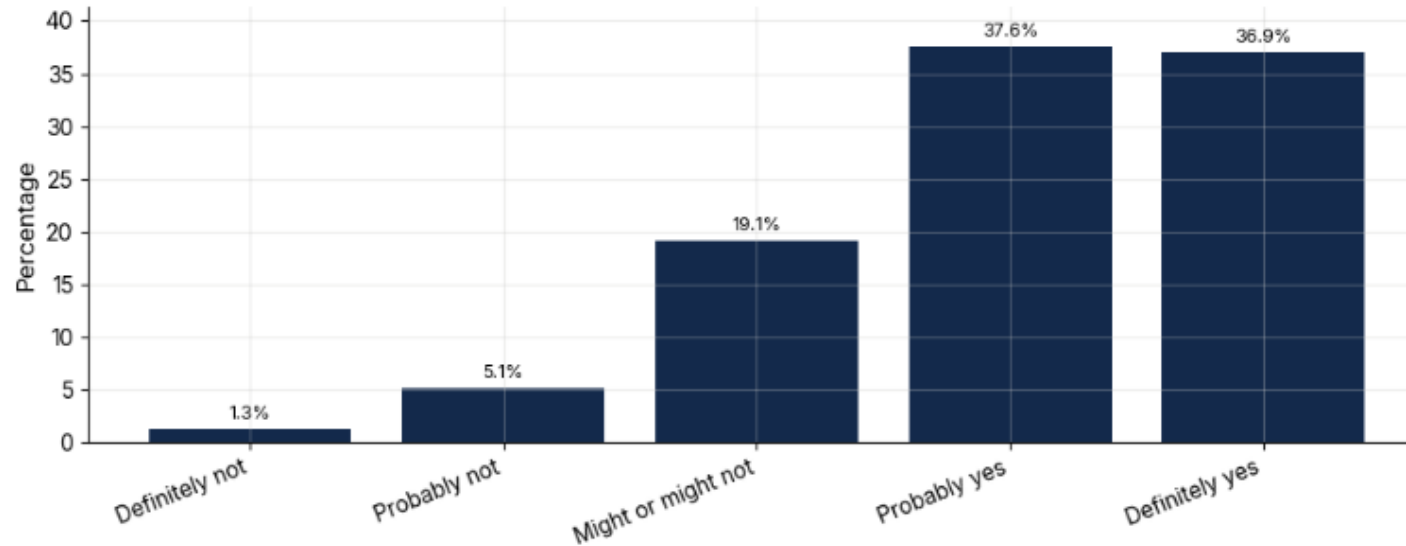
<https://lnkd.in/eK5VrEvy>

#AnalyticsEducation #AI #DataForGood #HigherEd



How can we get faculty to incorporate competition into their classroom or program?

Do you believe you would have spent more time on this competition had an instructor at your university made the competition as a graded course component?



Everyone is a winner when they participate!

Region 3 – Undergraduate						
index	Rank	Team Name	School Name	Total Points	AI-900 Training [Sunday, December 10th 2025]	Business Problem Framing [Sunday, November ...
91	1	Dawghouse9	Butler University	11565	750	750
87	2	ButlerBusiness5	Butler University	11382	750	750
97	3	Observatory8	Butler University	10427	750	750
90	4	Data Dawgs	Butler University	9665	750	750
94	5	Holcomb3	Butler University	9649	750	500
92	6	Esports14	Butler University	9599	750	750
83	7	BackTable	Butler University	9455	0	500
93	8	Hinkle10	Butler University	9349	750	750
96	9	Midtown1	Butler University	9282	750	750
130	10	jet2 holiday	Purdue University	9224	750	750
88	11	Colonnade7	Butler University	8999	750	750
89	12	Commons13	Butler University	8982	750	750
84	13	Butler Bias	Butler University	8732	750	750
85	14	Butler Blue	Butler University	8104	750	500
99	15	Traditions16	Butler University	7149	250	750
95	16	Invitation2	Butler University	6800	750	500
Rows: 49 						
					 Expand	



My course: 94% (50 / 53) passed AI-900 with average of 85%

Every student/team could have their own poster...

Data Dawgs

Data4Good Competition



Katie Mihelic, Luisa Hungria, Italy Lewis, Jinna Hyatt, Dr. Matthew Lanham (Advisor)
Butler University Lacy School of Business
kmihelic@butler.edu; hungria@butler.edu; itlewis@butler.edu; jhyatt@butler.edu



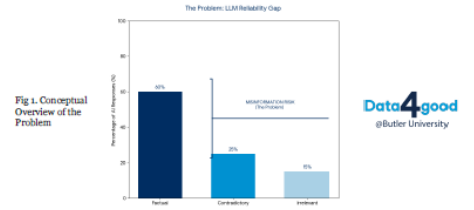
BUTLER
LACY SCHOOL
OF BUSINESS

ABSTRACT

This project is the result of the 2025 Data4Good Competition and aims to leverage data analytics and machine learning capabilities to solve a real-world problem that has significant positive impact on society. The purpose of this project is to derive valuable insights from the given dataset and build a model that can enable informed decision-making. The importance of this project is that it has the ability to make the solution more efficient, accurate, and effective with the use of data analytics. On the technical side, we used techniques such as data cleaning, analysis, feature development, and modeling with the use of Python and the DataCamp DataLab.

BUSINESS PROBLEM FRAMING

- The project addresses the problem of misinformation generated by large language models (LLMs) in educational settings. The central question is whether an AI-generated answer to a question is factual, contradictory, or irrelevant, given the question and supporting context.
- As AI tools are increasingly used by students, incorrect or irrelevant AI responses can mislead learners and undermine trust in educational technology. Accurately identifying unreliable AI-generated content is essential for responsible and safe use of LLMs in education.
- Primary stakeholders include educational institutions, students, educators, and developers of AI-powered learning tools who rely on accurate and trustworthy AI outputs.
- The project uses a labeled dataset of question-answer pairs and is constrained by data quality, class balance, and the need for interpretable and generalizable models. Performance is evaluated using a weighted confusion matrix that equally prioritizes all classes.
- The solution supports improved AI reliability, reduces the risk of misinformation, and enables safer integration of LLMs into educational platforms by flagging incorrect or irrelevant responses.



ANALYTICS PROBLEM FRAMING

Problem Statement: This is a multi-class classification problem. Given the question-answer pair, the goal is to classify the type of relationship between them into one of the predefined categories: "factual", "contradiction", or "irrelevant".

Clear Assumptions: The clear assumptions highlight that the data quality is representative, embedding models effectively capture semantic meaning, preprocessing enhances performance, features are sufficient for classification, and class imbalance is noted as an area for potential improvement.

Success Metrics:

- Overall Accuracy: Giving a general measure of the model's effectiveness.
- F1-Score (per class): Balanced measure of model's precision and recall.
- Confusion Matrix: Breakdown of each model's strengths and weaknesses.

Justification: This analytical problem is justified by the critical need for robust models that can accurately classify nuanced text relationships in real-world applications, despite significant class imbalance.

Personal Development & Outcomes

- Became Microsoft AI-900 AI Fundamentals Certified.
- Completed 5 DataCamp courses on Python Programming, Scikit-learn, Natural Language Processing, and Cluster Analysis.
- Learned the INFORMS Business Problem Framing Course.
- Earned a SAS Badge on SQL Essentials Knowledge.
- Earned a SAS Badge on AI Foundations Knowledge.
- Developed a much greater understanding of Python and AI/ML overall.

DATA

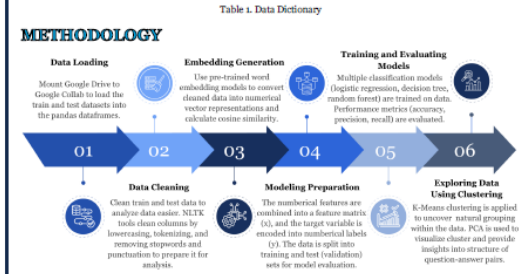
Important Data Relationship:

- The "type" label is determined by collective semantic meaning and coherence derived from "question", "answer", and "context" fields.

Business Problem Framing Reflection:

- Addresses a critical business need to automate question-answer classification by framing it as a multi-class text classification problem, leveraging NLP and machine learning.

Feature	Priority	Description	Role
type	High (target variable)	Categorical label: factual, contradiction, or irrelevant.	Primary outcome variable to predict.
question	High (feature)	Text string representing the user's query.	Essential input for intent and context understanding.
answer	High (feature)	Text string containing the provided response.	Crucial for evaluating correctness/relevance against the question.
context	Medium (feature)	Text string with additional background information.	Provides deeper semantic understanding.

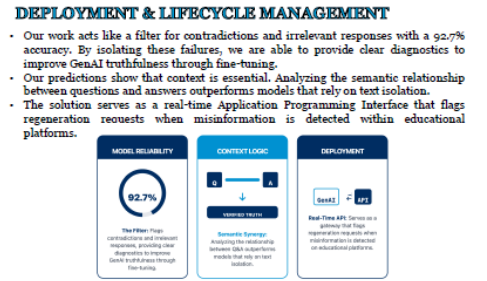
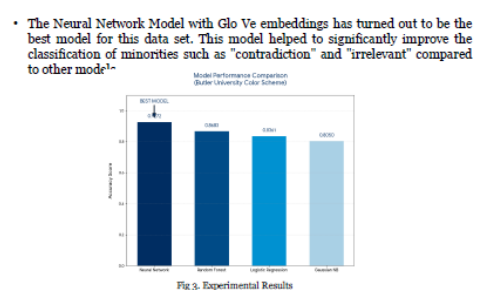


MODEL BUILDING & EXPERIMENTAL RESULTS

Rubric

- This table summarizes the performance of various predictive model experiments on different feature sets. It highlights models that generalize well to unseen data, shows where overfitting occurs, and provides insight into the relative effectiveness of embeddings, TF-IDF features, and hybrid approaches.

Features	Predictive Model	Train Accuracy	Test Accuracy
Glove-wiki-gigaword-go to 2 PCAs	Logistic Regression	0.8327	0.8361
Glove-wiki-gigaword-go to 2 PCAs	Decision Tree	1.0000	0.7827
AI MiniLM-L6-v2 embeddings (aka dlm) -> scaled -> PCA features -> 2 PCAs for visualization + K-Means	K-Clustering	Silhouette: 0.0395	Silhouette: 0.0399
AI MiniLM-L6-v2 embeddings (aka dlm) -> scaled -> PCA features -> 2 PCAs for visualization + Random Forest	Random Forest	1.0000	0.8683
AI MiniLM-L6-v2 embeddings (aka dlm) -> scaled -> PCA standardized features -> 2 PCAs for visualization + Neural Network	Neural Network	0.9458	0.9279
Hybrid TF-IDF + GloVe Features (20) -> scaled -> combined features -> Naive Bayes Classification	Naive Bayes Model	V1: 0.4259 V2: 0.4259 V3: 0.4259	V1: 0.4259 V2: 0.4259 V3: 0.4259
Character-level TF-IDF (4-grams, max 5000 features, derived from preprocessed Q&A)	Multiclass Naive Bayes on Character-Level TF-IDF	0.8950	0.8927
Word-level TF-IDF (4-grams, max 5000 features) from question + answer text	Extra Trees Classifier	0.8925	0.8925



KEY TAKE-AWAYS

What went well:

- Neural Networks with MiniLM embeddings — strong performance / good generalization.
- Logistic Regression and TF-IDF — stable and consistent.
- Character-level TF-IDF + Naive Bayes — surprisingly effective for simpler models.

What didn't go well/key considerations:

- Decision Trees — severe overfitting, especially with PCA + GloVe.
- K-Means — poor clustering (silhouette < 0).
- High-capacity models risk overfitting: PCA may limit learning of subtle patterns.
- Feature engineering is critical: naive hybrid approaches can hurt performance.
- MiniLM embeddings outperform GloVe for semantic tasks.

Competition Outcome

- Data Dawgs placed 26th in Region 3 in the Data for Good competition.
- An invalid Case Challenge file meant no official score was recorded.
- Ranking reflects participation in other components, not Case Challenge performance.







Student perceptions – “career readiness”

89%

believed this competition helped fill a knowledge, skill, or ability gap that they could showcase to employers

81%

said they would “probably” or “definitely” discuss this competition work in a future interview to highlight their analytics skills

Student perceptions – “process, process, process”

8%

have heard of the INFORMS CAP JTAs/IAF prior to this competition

80%

perceived knowledge of the INFORMS IAF was moderately to advanced

Note: 17% at beginning of competition

89%

plan on using the INFORMS CAP JTAs in the future when designing and delivering analytical solutions.

Note: 80% last year

Engagement with Business Problem Framing course

Status	Count
Complete!	145
Pending	186
(blank)	15
Grand Total	346



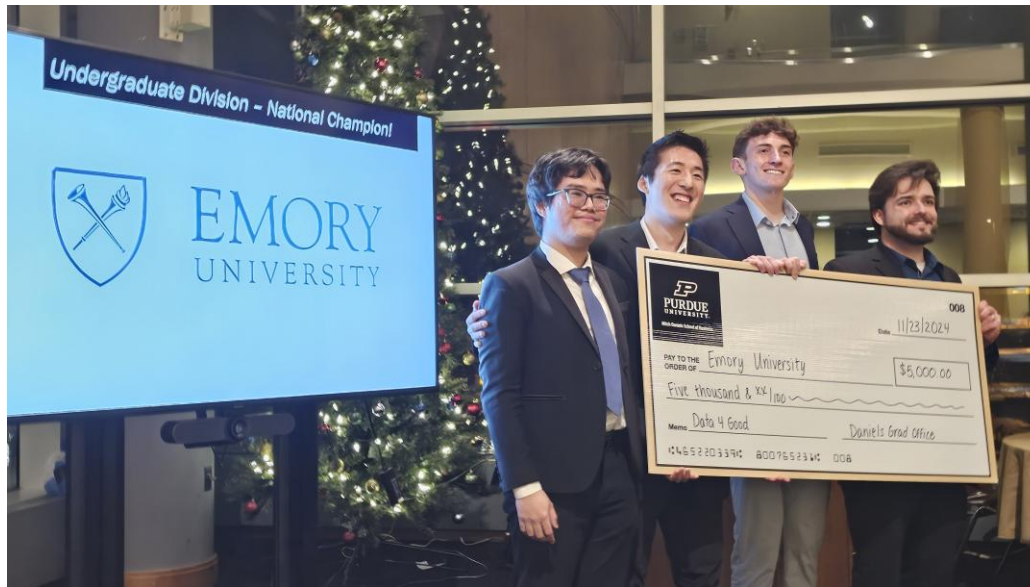
Student perceptions – “process, process, process”

89%

believed this competition was worth the time and money invested

89%

would recommend students to participate in the competition next year!



Future

- ✓ More/better student outcomes!
- ✓ Better engagement/promotion/understanding of the benefits of our key analytics community – INFORMS!
- ✓ Provide value for our INFORMS practitioner members – could there be engagement there – maybe different competition tracks/certifications?
 - CAP-E?
 - AI-900 + another partner certification (Andrew coming up with ideas)
- ✓ DataCamp / SAS
 - Continue SAS badge offerings
 - Can we better bridge their platform capabilities
 - Improved real-time scoring leaderboard
 - Better/faster follow up
 - Might bring back quizzes instead of video capturing to award points