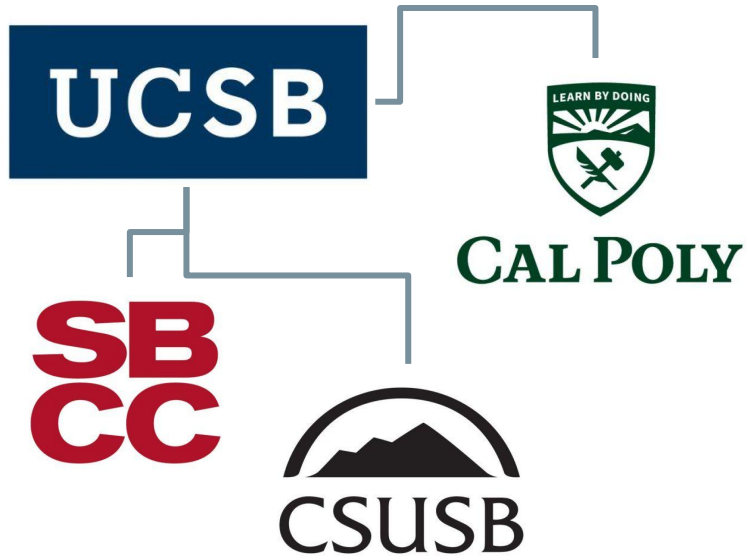# Central Coast Data Science Partnership: Training a New Generation of Data Scientists

Alexander Franks
UC Santa Barbara
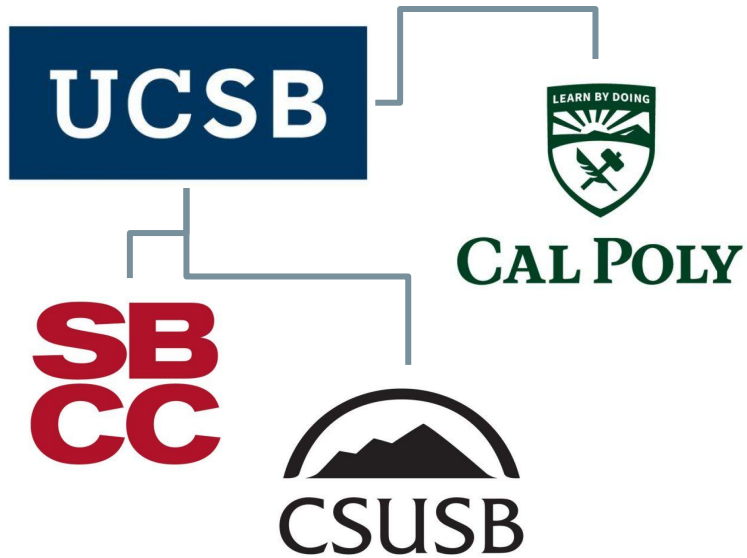
JSM 2023

Central Coast Data Science

- Data science training connecting three main public higher education institutions in California: UC, CSU, and Community Colleges
  (HDR DSC Awards #1924205 & #1924008)

- Information at:
  https://centralcoastdatascience.org/
- https://pstat197a.github.io

UC **SANTA BARBARA**

Central Coast Data Science
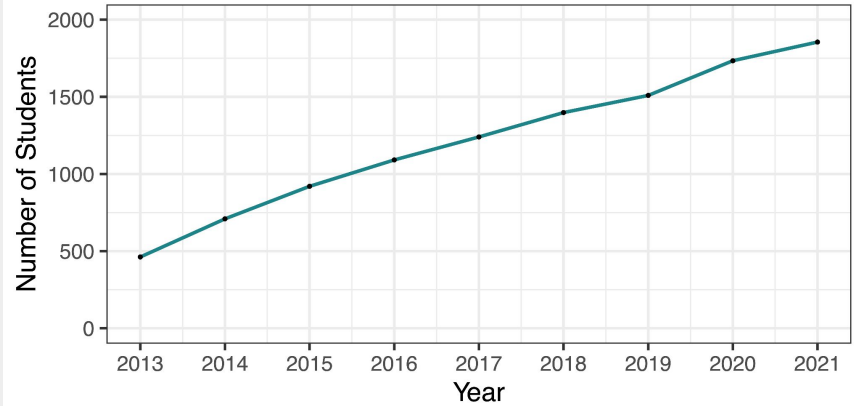
UCSB

CAL POLY

SBCC

CSUSB

- Close collaboration and connections between the institutions

- Stipends provide vehicle for UG training
  - UC and CSU recruiting at CC and High Schools
  - DS fellows as ambassadors

- Stack of best practices
  - Course development
  - Capstone organization
  - Seminars
  - Assessment

UC SANTA BARBARA

University of California Santa Barbara (UCSB)

Undergraduate PSTAT Majors (2013–2021)

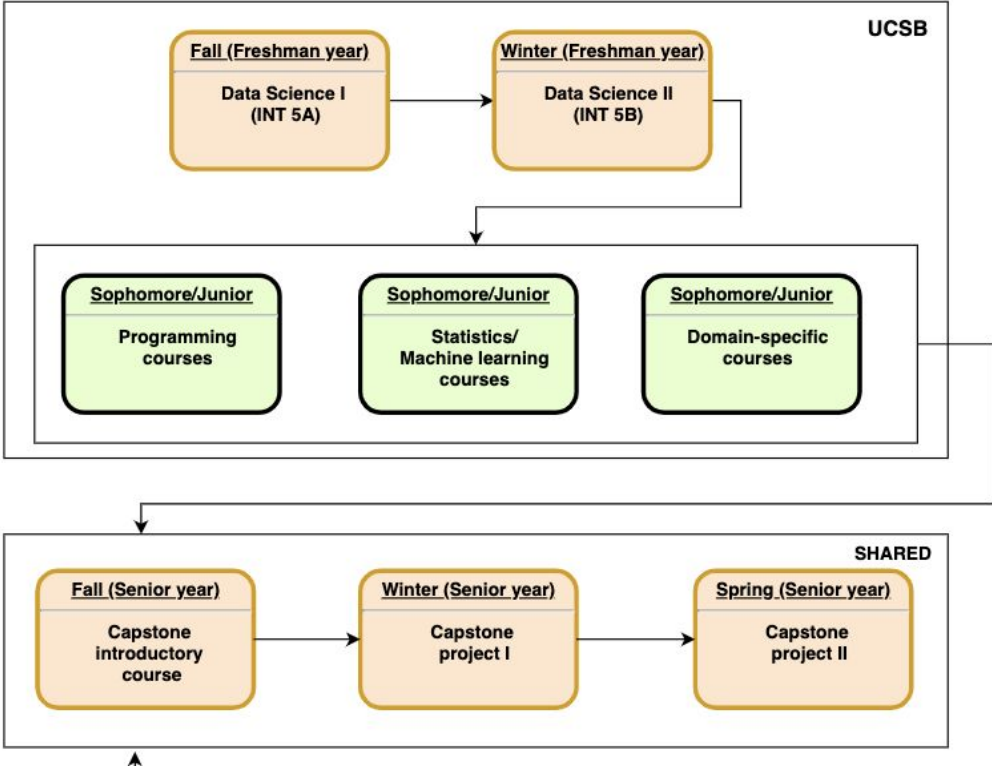Demand for DS skills exists outside of PSTAT!

- Campus Data Science Initiative
  - Working to unify research and educational activities related to DS across UCSB
  - DS Summit and faculty engagement

Santa Barbara City College (SBCC)

- SBCC is UCSB's main feeder college (17% of all transfers)

- Launched Intro to DS in Fall 2021
  - Based on Berkeley's Data 8
  - Looking to articulate DS courses at the UC and CSU

- Close collaboration with UCSB+SBCC faculty

- Data Science Fellows at UCSB transferred from SBCC
  - 2 SBCC Summer Fellows in 2021 transferred to UCSB in 2022
  - 2 UCSB Fellows in 2020-2021
  - 2 UCSB Fellows in 2021-2022

Data Science Pathways at UCSB

UCSB

Fall (Freshman year)
Data Science I
(INT 5A)

Winter (Freshman year)
Data Science II
(INT 5B)

Sophomore/Junior
Programming
courses

Sophomore/Junior
Statistics/
Machine learning
courses

Sophomore/Junior
Domain-specific
courses

SHARED

Fall (Senior year)
Capstone
introductory
course

Winter (Senior year)
Capstone
project I

Spring (Senior year)
Capstone
project II

UC SANTA BARBARA

New Introductory DS Courses at UCSB

- CS 5a/5b: Intro Data Science

  - Inspired by Berkeley's Data 8, but split over 2 quarters

  - CS 5a: focus on data science lifecycle, including question formulation, data collection and cleaning and visualization

  - CS 5b: more emphasis on statistical concepts (e.g. testing and regression)

  - General Education credits and pre-major credits in relevant domain areas

Data Science Fellows at UCSB

- Applicants are predominantly juniors and seniors
- Serve as "ambassadors" of data science at UCSB
- $5k stipend for DS fellows
- Participate in Capstone Projects
- "Committee" participation
  - Outreach
  - Education
  - Events
  - Infrastructure

Data Science Capstone

- 169 students completing full course sequence (52 in 2021, 60 in 2022, 57 in 23)

- 40 projects (12 in 2021, 14 in 2022, 14 in 2023)

- 11 different total industry partners

- 15 different academic partners (including external research centers, eg. sansum, stanford)

UC **SANTA BARBARA**

Data Science Capstone – Fall

- Tools and practices for collaboration
  - Version control with Git and GitHub
  - Design and systems thinking for data analysis and project management
- Series of data analysis modules
  - Chosen based on anticipated capstone projects
  - Lectures introduce questions, data, and methodology
  - Working groups of 4-5 students randomly assigned
  - Guided group projects extend in-class analysis
- Exposure to special topics
  - High dimensional data, multiple testing, natural language processing, image classification, advanced statistical and ML models, function approximation, correlated data
- Class site: https://pstat197.github.io/

Data Science Capstone - Winter/Spring

- Meet weekly for at least 1 hour with the project sponsor and DS mentor

- Meet once a week in class:

  - Project presentations and updates

  - Students provide feedback

- Class website: https://pstat197.github.io/

Projects

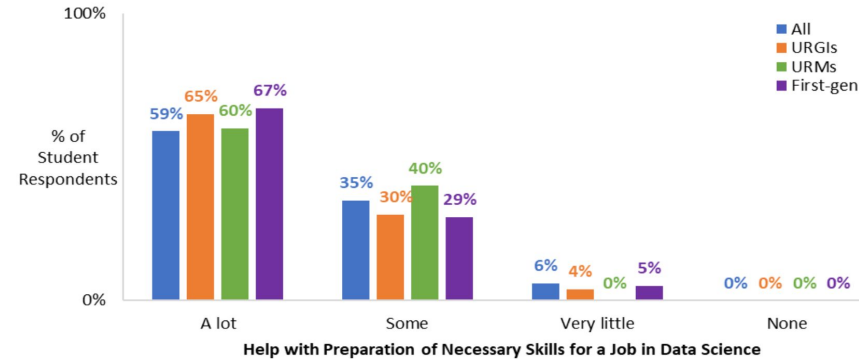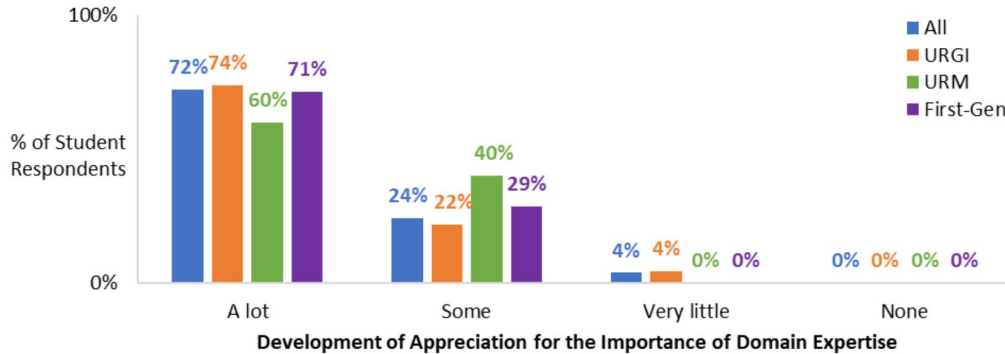|	
Quantifying Security - Visualizing Usability of Allthenticate Alltheniticate	**Stress indicators for frontline COVID-19 healthcare workers** Evidation Health
Demographic Predictors of Climate Opinion Energy and Environment Transitions (ENVENT) Lab at UCSB	**Product recommendation on scraped usage data** HG Insights
Modeling Neurodegeneration with Single Cell RNA Sequencing Neuroscience Research Institute (NRI)	**Deep Learning Recommendation System for Property Management Software** AppFolio
Combining computer vision and satellite imagery to map center pivot irrigation Earth Research Institute	**Compression of NLP-Domain Deep Learning Models** PwC (PricewaterhouseCoopers)
Deep image reconstruction from human brain activity Sprague Lab	**Wearables Predict Stress in Frontline Healthcare Workers** Evidation Health
Multi-class claims activity classification based on HTML data	**Big Bee: Hair Recognition & Quantification** Cheadle Center for Biodiversity and Ecological Restoration (CCBER)
Impacts of a documentary: The Social Dilemma - analyzing changes in social media usage behavior Center for Information Technology Research	**Interactive spatio-temporal visualization of long-term physical characteristics of California coastal waters** California Cooperative Oceanic Fisheries Investigations (CalCOFI)

https://centralcoastdatascience.org/projects

UC **SANTA BARBARA**

Regular Assessment

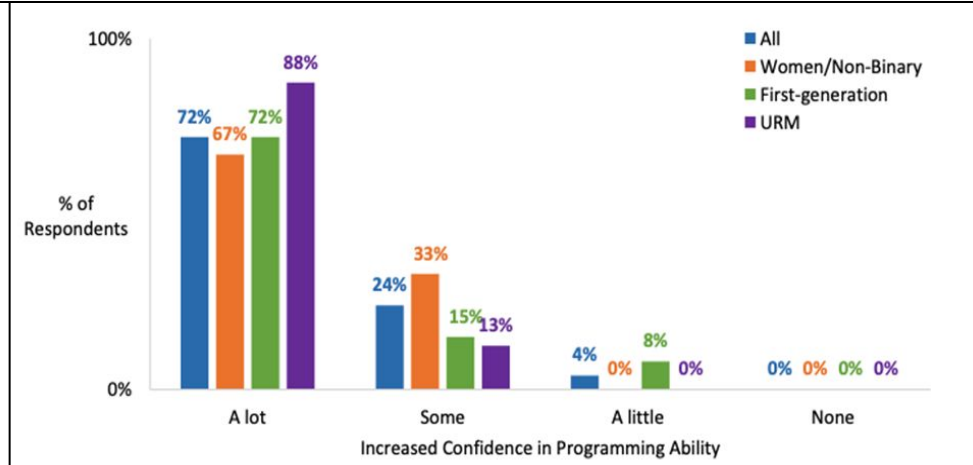Quarterly meetings with internal assessor allows us to track the impact of our DS classes across campus.

Assessment at UCSB and Cal Poly provided by Dr. Lubi Lenaburg, of UCSB's Center for Science & Engineering Partnerships (CSEP)

Development of Appreciation for the Importance of Domain Expertise

Help with Preparation of Necessary Skills for a Job in Data Science

Regular Assessment

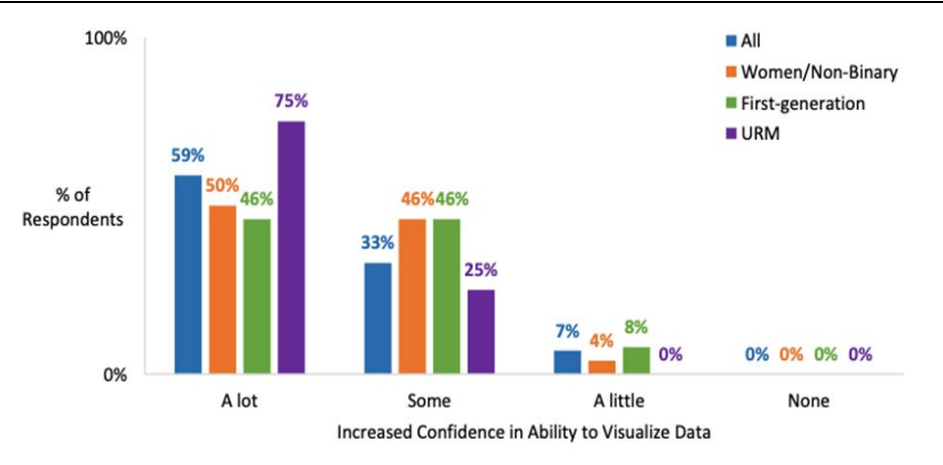Quarterly meetings with internal assessor allows us to track the impact of our DS classes across campus.

Assessment at UCSB and Cal Poly provided by Dr. Lubi Lenaburg, of UCSB's Center for Science & Engineering Partnerships (CSEP)

What next?

- Next year's capstone class is fully enrolled

- Need to identify project sponsors

- Need to identify UCSB project mentors

- Can we scale the capstone beyond 60? Who gets into the capstone?

- New grants…

Southern California Consortium for Data Science

- Newly funded by CA Learning Lab (2023+)
- Data Science Pathways Development project:
  - Inter-segmental team of 8 institutions
  - Develop multi-lateral articulation agreements
  - Support incoming Transfers
  - Coordinate curriculum development
  - Professional development for DS instructors
  - Build open-source instructional materials for DS Ethics and inclusive teaching
  - Regional Datathons and DS Career Days
- Will reach 10K+ students and 200+ faculty

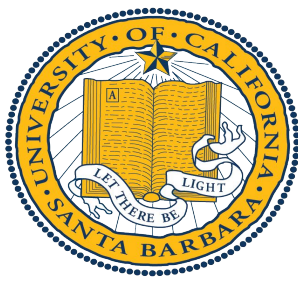# Pacific Alliance for Low-income Inclusion in Statistics and Data Science

- New multi-institutional consortium funded by the National Science Foundation
- Focuses specifically on **Statistics & Data Science** (SDS) programs
- **Goals:**
  - Increase SDS degree completion by high-achieving undergraduates with demonstrated financial need
  - Bridge between undergraduate and graduate programs in Statistics and Data Science
  - Prepare students for well-paying data-rich careers
- Partnership of **7** institutions across California and Washington:

  - UC Santa Barbara
  - U of Washington
  - UC Irvine
  - CSU Monterey Bay
  - Cal Poly SLO
  - CSU East Bay
  - San Diego State U

Thanks!

- Ambuj Singh (PI, CS UCSB)
- Mike Ludkovski (STAT, UCSB)
- Trevor Ruiz (STAT, UCSB)
- Nathalie Guebels (CS, SBCC)
- Sang-Yun Oh (STAT UCSB)
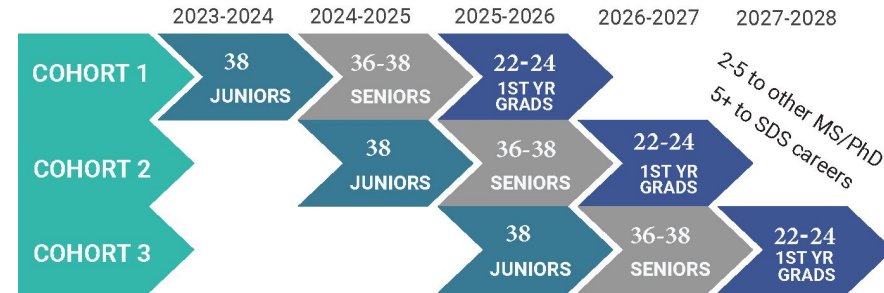- Industry sponsors and participating academic labs!

UC **SANTA BARBARA**

Pacific Alliance for Low-income Inclusion in Statistics and Data Science (PALiISaDS)

PALiISaDS: Timeline

- Cohort-based, three-year bridge program that engages sophomore / transfer students and follows them during their junior and senior undergraduate years
- Continued support during first year of graduate program if remain within the Alliance
- Inaugural cohort begins in **Fall 2023**
- Project lasts for 6 years through 2028
  - A total of **300** scholarships to be awarded
  - Need-based scholarships of up to **$10,000** a year for up to three years (junior, senior, 1sr year grad)
  - About 38 students to be admitted per year

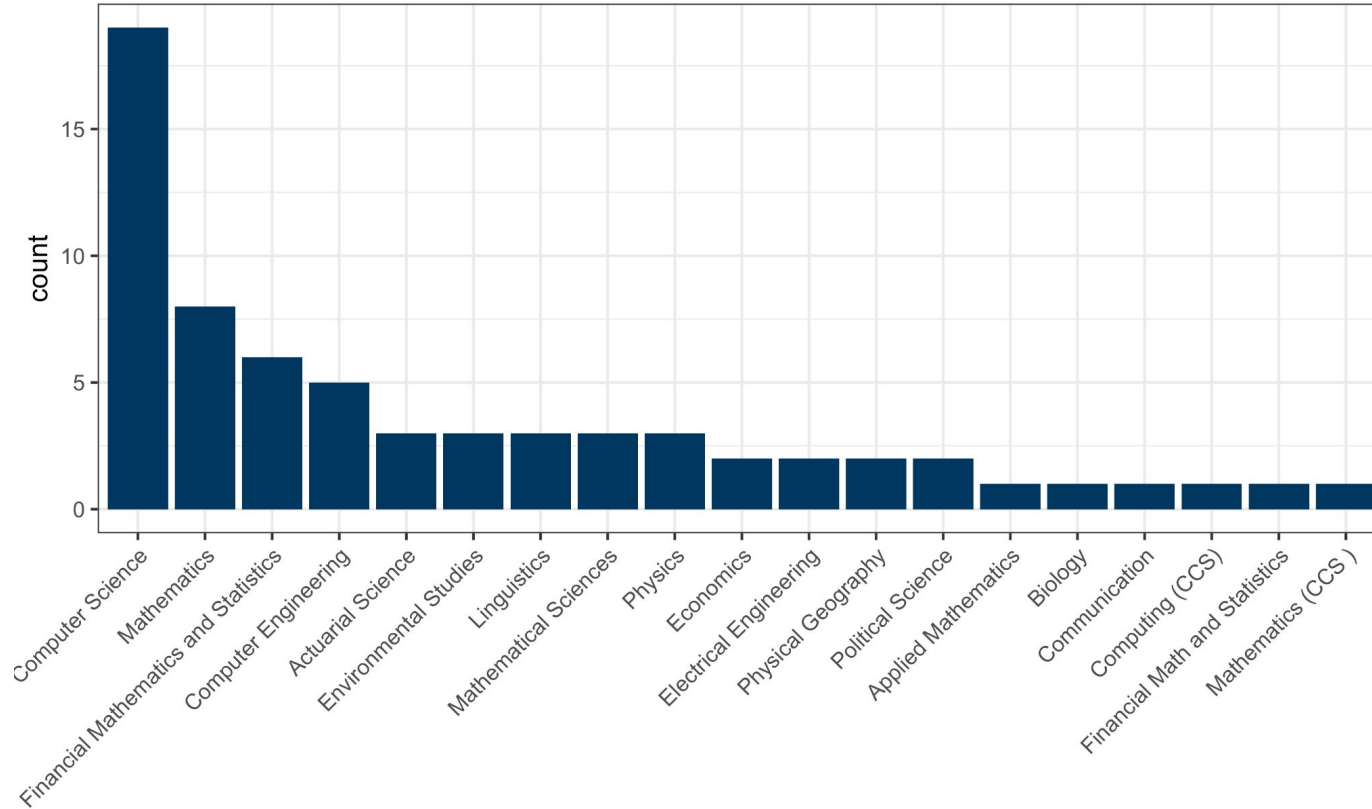|	2023-2024	2024-2025	2025-2026	2026-2027	2027-2028
COHORT 1	38 JUNIORS	36-38 SENIORS	22-24 1ST YR GRADS		2-5 to other MS/PhD 5+ to SDS careers
COHORT 2		38 JUNIORS	36-38 SENIORS	22-24 1ST YR GRADS	
COHORT 3			38 JUNIORS	36-38 SENIORS	22-24 1ST YR GRADS

Pacific Alliance for Low-income Inclusion in Statistics & Data Science

- Cohort-based, three-year bridge program that engages sophomore / transfer students and follows them during their junior and senior undergraduate years
- Continued support during first year of graduate program if remain within the Alliance
- 7 participating institutions: UCSB, UCI, SDSU, CP SLO, UW, etc.
- Inaugural cohort begins in Fall 2023: **7 students just short-listed**
- Project is NSF-funded and lasts for 6 years through 2028
- Need-based scholarships of up to $10,000 a year
- Eligibility: SDS Major + rising Junior + Pell-eligible
- Wrap-around support activities; graduate and faculty mentors
- Several event series:
  - Seminar series on SDS careers/programs
  - Workshops on grad school preparation
- paliisads.org

Non-Stats Majors in Capstone

Community Outreach

"For the longest time, I had been hesitant about taking interest in majors similar to mathematics or STEM, but after hearing your presentation, I am starting to reconsider my feelings. I appreciated how you gave us examples about the modern uses of Data Science. It was important for me to see the broad range of areas it was used for, which broke the negative stereotype I originally held toward these majors." -High school student

"One of my students' father emailed to say thank you because his son took personal initiative to sign up for a coding class immediately after your presentation. The father was very happy to see his son excited for something after a school year where it looked like the son lost motivation in school and the future" -High school teacher

UC **SANTA BARBARA**