CDSS Data Discovery Partner Orientation: Fall 2025

UC Berkeley College of Computing, Data Science, and Society

Agenda

1. Program overview
2. Fall 2025 dates and timeline
3. Expectations for you as a mentor
4. Student expectations
5. Course goals
6. Symposium
7. NDA and confidentiality
8. Computational resources
9. Contact us

Thank you!

- Thank you for being here today!
- Thank you for being a partner
  - This transformative experience is deeply valued by students and only possible with your participation
  - Your support makes the program possible and has a direct impact on your students and the Berkeley community

Program team

George Obaido, PhD
Program Director



Emily Remirez, PhD
Associate Director



You may also hear from CDSS Associate Dean **Narges Nourouzi**, Professor **Deb Nolan**, or CDSS Organization Development Director **Sierra J. White-Koutsoubos**, who advise the program

Data Discovery Program Overview

Project: Working with you

- Hands-on data practicum
- Work with a project partner/mentor and student teammates
- Meet weekly and complete project assignments

CDSS 198: Working with us

- Weekly in-person classes:
  - Student team meeting
  - Professionalization, guest lectures and group discussions
  - Short written assignments
- Passing or non-passing grade

Fall 2025 dates and timeline

September 3	Deadline to form your team (Please use the DDMatch platform to send offers)
September 8	Students' first day of Data Discovery course
October 13 - 20	Mid-term student progress report
December 1 - 5	Last week of class/work on project
December 10	Data Discovery Symposium, program evaluation
Student holidays	September 1, November 11, November 26-28

Reminders for extending offers

- All offers should be sent via the DDMatch platform you used to review applications
  - An official record of offers is crucial for enrolling students in
- **Limit offers to your desired team size**
  - You can expect a high yield from positions offered
    - Make the 5 offers first, then extend additional offers if not accepted
  - It will be much easier to recruit students “late” than deal with an excess of accepted offers
- There is a cap on team size (5 students) and overall program size

Rescinding offers and closing applications

- Offers should expire after 2-3 days
  - **Please rescind offers that don’t have a response**
  - Invite an alternate student
- **Close your project to new applications when your team is full**
  - Signals that your team is finalized **via email** so we can begin enrolling students
  - Protects students’ time and hopes with their applications

Ensuring an educational experience for students

- Data Discovery is part of students' curriculum:
  - Their learning and mentorship should be a primary goal
- Every task should have some kind of educational or professional development value, even if it's not glamorous
- Include students in discussions about the larger context
  - They should understand why they're doing what they're doing
- Talk to students about their individual goals and share your experience or advice

Expectations for students

- Students should work with you to establish expectations for your specific projects during a first assignment, the mentoring agreement (more on next slide)
- We ask everyone to
  - Maintain their weekly hourly commitment, on average
  - Proactively communicate with you and their teammates
  - Remain intellectually honest in their work and not use generative AI without talking to you explicitly

Mentoring Agreement topics

1. Contact information and policies
2. Project description and goals
    a. Check their understanding of your project
    b. Share their personal goals and look for connections
3. Meeting schedules and expectations
4. Expectations for working together
5. Integrating Data Discovery program assignments
6. Anything else you want to cover

Data Discovery course

Introducing an in-person course component for students

Why the course?

- Increase student engagement
- Provide students with a dedicated time to meet
- Promote social interaction among students
- Teach “soft skills” to support project work

Course goals

1. **Contextualizing their work** within the larger framework of your project and integrating the experience with their studies and larger goals
2. **Communicating** clearly and genre-appropriately about the data, methods, and interim results with different audiences
3. **Working as part of a team** towards collaborative learning, project management, and problem-solving
4. **Developing personal processes** around time management, resourcefulness, and independent work

Some course topics

- Working together and setting expectations as a team
- Understanding the data science life cycle and the characteristics of your dataset
- Time management and workflow
- Problem-solving, resourcefulness, and perseverance
- Exploratory data analysis and visualization
- Designing an effective poster presentation

Data Discovery Symposium: December 10

- We welcome you to join us on campus for student poster presentations!
  - **Wednesday, Dec 10, 1:30-3pm**
  - Hearst Memorial Mining Building lobby (Berkeley campus)
  - Detailed invitation to follow
- Your presence improves the symposium for all students



More about the Symposium

- Students are required to deliver a final presentation for the course
  - Presenting a poster at the Symposium satisfies this requirement
- Concerns about a public poster?
  - Talk to your students ASAP so they can make alternative plans
- Requiring a presentation of your own?
  - Students can use this for their assignment if you provide confirmation
- You should approve students' poster for accuracy and are welcome to support or advise its creation, but it should be student-authored

Poster printing

- **Please plan to cover the cost of printing for your student team this semester ($30-$50)**
  - If you are unable to do so, please email us as soon as possible
- A link for our optional group order will be sent closer to the date
  - Advantage: Convenient for students and discounted
  - Disadvantage: Early submission deadline
  - You’re welcome to use any service you prefer

Non-disclosure agreements & confidentiality

- NDAs or confidentiality agreements may be used to protect your project while maintaining an open, supportive environment
  - Refer to this example for a UC Berkeley aligned format https://tinyurl.com/DD-NDA-example
- NDAs should:
  - Clearly define what info is confidential and how it can be used
  - Avoid overly broad restrictions that can interfere with students' work
  - **Avoid financial penalties** that can cause anxiety and detract from the educational focus of the program
- Required trainings should be factored into project work time

Computational resources

- Data Discovery provides high-performance computing resources for students and partners.
- Students can request access to **Savio**, Berkeley's supercomputing infrastructure, for Python and R workloads.
- Also, students and partners can request access to the **CDSS DataHub** for large-scale data processing, AI, and machine learning tasks.
- Both platforms include pre-installed Python and R data science libraries, with options for additional software.
- After teams are formed, email cdss-datadiscovery@berkeley.edu to request access to Savio and the CDSS DataHub

Questions?

Email: cdss-datadiscovery@berkeley.edu