

DataFest

September 14-21, 2018 Davis, CA

<https://goo.gl/forms/hW hrAGjLKK3AcbD02>

The DSI is holding a DataFest! We'll provide the problem and the data, and you'll work in small teams to explore interesting approaches and solutions. UC Davis students, postdocs, and staff with all skill levels are encouraged to participate. Alumni, community members, and other UC and CSU students should contact datascience@ucdavis.edu.

YOUR CHALLENGE

Sherry-Lehmann is a well-known, established wine and spirits seller. The UC Davis Library has over 200 scanned "wine" catalogs created by Sherry-Lehmann spanning a 30 year period and over 4,000 pages. How can we best recover information about wine prices over time from these scans so it can be used by researchers, archivists, wine historians and enthusiasts? The Data Science Initiative is actively helping with this project and we challenge *you* to find creative ways to solve this problem and correctly extract the data from the catalogs.

This data recovery project combines:

- | | | |
|--|---|--|
| <input type="checkbox"/> optical character recognition (OCR) | <input type="checkbox"/> image processing | <input type="checkbox"/> statistics & machine learning |
| <input type="checkbox"/> computer vision | <input type="checkbox"/> text processing | <input type="checkbox"/> common sense & problem solving! |

DETAILS

The event kicks off on Friday, September 14th with an info session to frame the problem and form teams. Come and work with your friends, or we'll help you form a team. Teams are expected to be between 2-4 people. Group work days are September 18-21 (10am-2pm each day). Your entire team is expected to work in the DSI during those days and times, and we'll provide you with coffee, tea and snacks to keep you going. But you aren't restricted to working on the project only during those times - you can stay and work in the DSI or library. All projects must be completed by noon on September 21, and teams will present their results that afternoon.

SHOULD I PARTICIPATE?

This DataFest is a great opportunity to:

- ☐ work on an important, relevant Data Science project
- ☐ develop a portfolio project
- ☐ work in teams, develop teamwork skills
- ☐ learn about OCR
- ☐ use workflow best-practices
- ☐ problem solving experience

Everyone's welcome and useful. There are many ways to contribute - across different subtasks and at all skill levels - so don't be shy. Subtasks include developing truth-sets, testing results, integrating data, identifying patterns in the images, developing ideas for extraction approaches, programming, and fitting models. And there are lots of smaller sub-projects, so "finishing" your project doesn't necessarily mean recovering all the data!

SCHEDULE

Sept. 14th	Sept. 18-20	Sept 21
11:30am – 12:30pm	10am-2pm	10am-2pm
Info session, form teams	DSI work sessions	Project wrap-up & presentations

AWARDS

Each team must present a short summary of their findings to the other teams and a panel of judges comprised of staff and faculty from the DSI and Library. Teams will be judged based on their creativity, insight, and communication. Separate prizes will be awarded for technical solutions and conceptual approaches, so everyone - regardless of incoming skills - has a chance to succeed. Awards will be announced after the products are reviewed by the DSI and Library.