

Datathon @ UCI

Participation Guide



Last Updated: April 9th, 2024

This is a living document. Additions may be made before or during the datathon.

Datathon @ UCI	1
Welcome!	3
Rules	4
Overview	5
Participant Checklist	5
Datathon Resource Sheet	5
Schedule and ISEB Spaces	6
General Schedule	6
COVID Guidelines	6
Virtual & In-Person Spaces	7
Connecting to UCI Wifi	8
Data Challenges	9
Project Submission	9
Prizes and Judging	10
Prize Categories	10
Judging	11
Mentorships & Workshops	12
Mentors	12
Workshops	13
Sponsor Networking	14
Chat with Melissa Data	14
How to Talk To Sponsors	14
Frequently Asked Questions	15
Sponsors and Recognitions	17

Welcome!

Hello divers, and thank you for attending the second inaugural Datathon at UC Irvine! We're delighted to see familiar faces returning and to welcome new participants to our community. It's an honor to have you join us for another exciting weekend of exploration, creativity, and teamwork.

As we gather once again, both virtually and in spirit, we reflect on the incredible growth and achievements that have marked our journey since our first datathon, Atlantis. From the beautiful yet innovative projects to the lasting connections with friends and teammates, each datathon serves as a testament to the power of community and the boundless potential of data-driven insights.

This year's theme, Atlantis, serves as a beacon guiding us towards uncharted waters of discovery. Like the fabled city lost to the depths of time, our datasets harbor untold stories waiting to be unveiled. Together, we will unravel the mysteries, decode the patterns, and chart a course towards new horizons of knowledge and understanding.

In the spirit of exploration, let us embrace diversity of thought, celebrate varying perspectives of our team, and foster an environment where innovation flourishes. Whether you're a seasoned data scientist, a passionate novice, or somewhere in between, your contributions are invaluable, and your unique voice adds depth to our collective narrative.

The 2023-2024 Data@UCI Board 💙GINA💛

Jenny Tran, Ellie Lee, Emi Cervantes, Gordon Lin, Kaylie Chong, Nathan Samarasena, William Hou, Natalie Nguyen, Ashley Yung, Rosalind Guo, Katherine Fields, Azra Zahin, Shreya Shyam, Jake Silverman, Elane Elza Shane, Megan Le, Catelynn Chen

Rules

Data@UCI stands for inclusivity. We believe that every single person has the right to participate in a safe and welcoming environment.

Golden Rule: **Be respectful.** Harassment and abuse are never tolerated. If you are in a situation that makes you uncomfortable, or if the event itself is creating an unsafe or inappropriate environment, please inform an organizer. Should an event organizer be creating an unsafe space, please email our club advisor Dr. Mine Dogucu .

1. Be respectful to other participants, mentors, board members, sponsors, and other individuals involved with the event.
2. Please comply with organizer instructions.
3. You must do your OWN work. Please only submit work you did during this event and not anything you have done previously.
4. Data @ UCI reserves the right to disqualify participants for what we may deem as illicit or dishonest conduct.
 - a. Harassment includes but is not limited to offensive verbal or written comments related to gender, age, sexual orientation, disability, physical appearance, body size, race, religion, social class, economic status, veteran status. Additional cases of harassment include but not limited to sharing sexual images, deliberate intimidation, stalking, following, brigading, doxxing, harassing photography or recording, sustained disruption of talks or other events, inappropriate physical contact, and unwelcome sexual attention.
 - b. **If what you're doing is making someone feel uncomfortable, that counts as harassment and is enough reason to stop doing it.¹**
5. Do not use ISEB 5th and 6th floor, and do not use any room that's not either ISEB 1010 or ISEB 1200.
6. Feel free to utilize the UCI campus and its open study areas including the Student Center, Science Library, Langson Library, etc.
7. Respect your work areas and make sure to clean before leaving.
8. Be sure to keep your badge with you at all times.

If you have any questions, reach out to us on the [Atlantis Datathon Discord](#).

¹

Rules adapted from the MLH code of conduct: <https://static.mlh.io/docs/mlh-code-of-conduct.pdf>

Overview

Participant Checklist

- Join our [Discord Server](#) and keep notifications on for important event announcements. Remember, all further communication will only take place on Discord.
- Form a team of up to 4 people. Check out our team formation resources below!
- Check-In on Saturday, April 13 @ 8:30 AM at ISEB
- Attend Opening Ceremony at 10 AM
- Network with our sponsors, check out our workshops, enjoy some good food, meet other participants, and work on your project!
- Submit your project by Sunday, April 14 by 10:30 AM by filling out our submission form.

Datathon Resource Sheet

We compiled a comprehensive list of resources and articles to help you and your team succeed at our datathon. Topics include, but are not limited to, **statistics**, **data visualization**, and **machine learning**. Feel free to take a look!

Datathon @ UCI	
Resource Sheet	
*Remember to check out our Datathon workshops!!	

Table of Contents	
Datathon @ UCI	1
Data @ UCI Workshops	2
Data Project Bootcamp	2
R Visualization	2
Intro to Python for Data Analysis	2
Beginner SQL	2
Intro to Tableau	2
Data Projects	2
Tips for a Good Data Project	2
Project Examples	2
Software & Applications	3
Exploratory Data Analysis (EDA)	3
Pandas: A Python Library	3
SQL	3
R	5
Time Series Analysis	5
Data Visualization	6
Plotly: Interactive Graphs!	6
Seaborn & Matplotlib	6
ggplot2: A R Library	6
Statistical Tests & Models	7
Hypothesis Testing (T-Test, Chi-Square)	7
ANOVA	7
Regression Models (Linear, Poisson, etc.)	8
Model Validation	9
Machine Learning	10
Common Terminology	10
ML Workflow	12
Model Diagnoses	12
ML Validation & Optimization	12
Sample Code	13
Statistical Models	13
Statistical Tests	13
ML Models	14
Misc. Resources	14
Resource Submission	15

[Datathon Resource Sheet](#)

Schedule and ISEB Spaces

General Schedule

Saturday: April 13th, 2024		
8:30 AM	Check-in & Breakfast	ISEB Front
9:15 AM	Team Formation Activity	ISEB 1200
10:00 AM	Opening Ceremony	ISEB 1010
10:45 AM	Workshop #1: Data Manipulation by Melissa Data	ISEB 1200
1:00 PM	Lunch	ISEB 1st Floor Lobby
2:00 PM	Workshop #2: Data Visualization with Ryan Yang	ISEB 1200
4:00 PM	Workshop #3: ML Workshop with Azra Zahinzra Zahin	ISEB 1200
6:00 PM	Dinner	ISEB 1st Floor Lobby
7:00 PM	Workshop #4: Presentation Tips by Hack@UCI	ISEB 1200
9:00 PM	Socials End	ISEB 1200
10:00 PM	Door closes	
Sunday: April 14th, 2024		
8:00 AM	Check-in & Breakfast	ISEB Front
10:30 AM	Submission Closes	
12:00 PM	Presentations	ISEB 1010
1:30 PM	Lunch	ISEB 1010
4:30 PM	Closing ceremony	ISEB 1010

COVID Guidelines

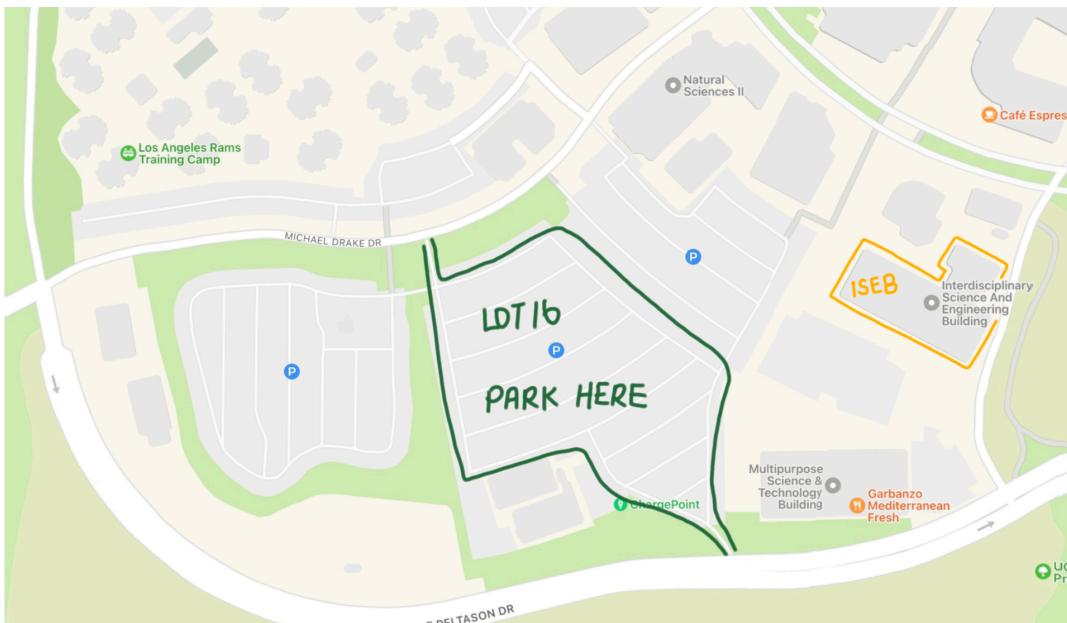
Masks are encouraged, but optional at this event. Please respect the personal choices of your fellow attendees. Please make sure to sanitize your hands frequently and practice good personal hygiene.

Virtual & In-Person Spaces

All the Datathon events will occur in the **Interdisciplinary Science and Engineering Building (ISEB)**. The building is located behind the Physical Sciences lecture hall, and it's next to the **Lot 16 parking lot**.

ISEB Address: [419 Physical Sciences Quad, Irvine, CA 92697](https://www.uci.edu/iseb)

Parking Lot Address: [Lot 16 UCI](https://www.uci.edu/parking/lots/lot-16)



Purchase Parking Permits

If you already have a UCI student parking pass, you do **not** need to purchase one. Your permit allows you to park in general unassigned and "AR" reserved stalls not marked "Enforced 24 Hours".

IMPORTANT: Unfortunately there has been an error with UCI Guest and Event Services preventing the parking link from working, so **you cannot pay for parking in Lot 16 online**. There will be parking attendants available at the entrance on Michael Drake Dr to give parking permits **cash only**.

However, there are alternative parking areas at **Lots 16H, 70 and Anteater Parking Structure (APS)** available for purchase with Park-by-Plate. I have provided a map on where the parking spots are located and will link their respective Google Maps locations. There will be Park-by-Plate signs that look *similar* to the example image provided, and you can enter the given codes of the designated area on the website <https://www.parking.uci.edu/parkbyplate> and have your parking spot secured. APS may have attendants available to pay for parking in-person (not guaranteed).

Lot 16H: <https://maps.app.goo.gl/Q31RdGckPgPghQxf9>

Lot 70: <https://maps.app.goo.gl/gjLXrkY47w8tdw3s6>

APS: <https://maps.app.goo.gl/1apeNpUp6YBLw95AA>

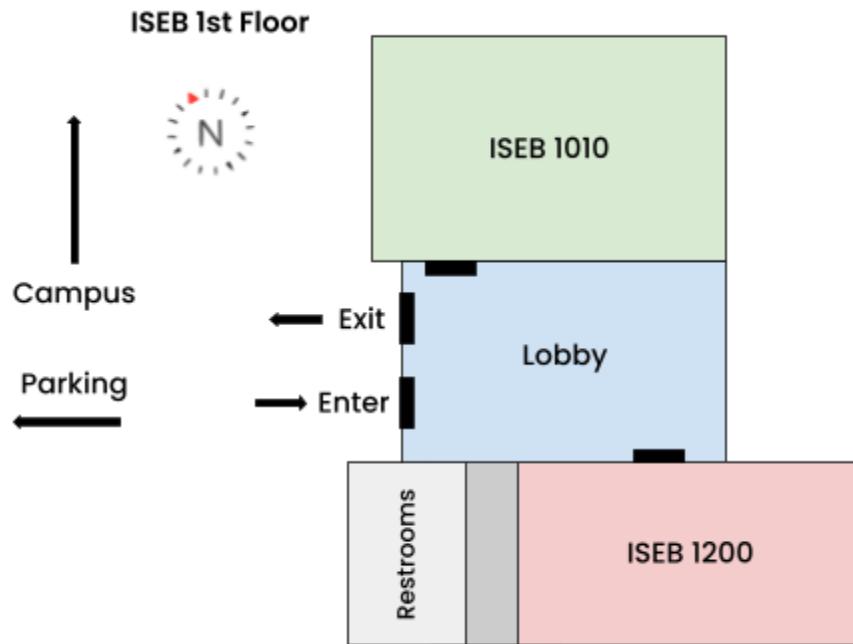




Overview of Datathon Spaces

Opening/Closing Ceremony	ISEB 1010
Workshops and Seminar	ISEB 1200
Project working space	ISEB 1010, 2nd and 3rd floors interaction space, libraries (Science Library and Langson Library are open 1 PM – 5 PM – make sure to come back for meals 😊)

You and your team can choose to work on your project anywhere on campus! However, if you choose to stay in ISEB to work on the project, **please do not use ISEB 5th and 6th floor, and do not use any room that's not ISEB 1010/1200.** Lastly, make sure to clean up after yourself and respect others!



Connecting to UCI Wifi

Current UCI Students

If you are currently a UCI Student you may access UCI wifi through [UCINet Mobile Access](#) or [eduroam](#).

Non-UCI Participant

If you are a participant not from UCI, you can access UCI through the [UCI-Guest Wifi](#) or you may connect through eduroam if you attend a participating institution.

Data Challenges

	Melissa API/Dataset Challenge [Sponsored by Melissa]	Create an application that utilizes one or more Melissa Cloud API and/or Data Sets. Judging will be based on creativity, originality, and best implementation. Access the Data Challenge here.
	Market Analysis in Dublin [Sponsored by StrataScratch]	Congratulations! Your team has been hired as a new city Airbnb city manager for Dublin, Ireland! As city manager(s), you will try to boost the number and quality of hosts in Dublin to fit the demands from guests. The goal of this challenge is to analyze, understand, visualize, and communicate the demand / supply in the market. Access the Data Challenge here.
	Automotive Bill of Materials Challenge [Sponsored by Quick Release]	You are a data analyst for Quick Release Automotive. We're trying to build an electric vehicle called the "QRV" and we need your help sorting out the data! The engineers have been releasing their parts into the Bill of Materials (BoM) without anyone checking the data quality first. All downstream stakeholders are worried we won't be able to order the parts we need on time. It's up to you to find out what's wrong and report back. We encourage you to use Excel, VBA, Python, or SQL- all languages that we use on a daily basis. Access the Data Challenge here.
	Choose your Own Data Adventure	Have your own idea? Feel free to choose your own adventure! Keep in mind that your project will be judged according to the same rubric as all other data projects. Possible Sources To Get You Started: Irvine Open Data Data.gov Data.world Kaggle

Project Submission

Please adhere to the following checklist to submit your project:

1. Create a 5-minute presentation that overviews your project. Be sure to discuss:
 - a. Why did you choose this project?

- b. What methods did you use? How did you explore the data?
 - c. What insights did you find? Why are they important?
 - d. **Be sure to keep your presentation no more than 5 minutes**
2. Create a submission on our Devpost: <https://atlantis-datathon-2024.devpost.com/>
 3. Submit your project on to our submissions google form
<https://forms.gle/RzMFM7qNJgtWSqfF6>
 4. Projects will be presented on Sunday in ISEB 1010, with a presentation preparation room available before your time slot in ISEB 1200.

All projects must be submitted before Sunday, 4/14 at 10:30 AM PST to be eligible for prizes. A submission consists of all three above components.

Prizes and Judging

Prize Categories

	
Best Overall \$150 Amazon Gift Card	Runner-Up Pre-Built Keyboard
	
People's Choice Ikea DJUNGELSKOG	Data for Social Good Award Owala Water Bottle
	
Best Visualization Digital Camera	Best Presentation Mini-Projector



QUICK RELEASE

AN ALTEN COMPANY



stratascratch

Best Automotive Bill of Materials Analysis
[Sponsored by Quick Release]

*Mini Car Figurine +
\$50 Amazon Gift Card*

Best Analysis of Market Analysis in Dublin
[Sponsored by StrataScratch]

*Annual StrataScratch Subscription +
\$50 Amazon Gift Card*



Best Use of Melissa API or Data Sets
[Sponsored by Melissa]

*Exclusive Melissa Swag +
40K credits for Melissa lookups and
Cloud APIs +
Melissa Office Tour +
\$50 Amazon Gift Card*

Judging

Your projects will be evaluated by our panel of judges with backgrounds in academia and industry. Judging will occur in person on Sunday 4/14 and judges will be scoring your projects according to the rubric [here](#). Winners will be announced at the closing ceremony.

Mentorships & Workshops

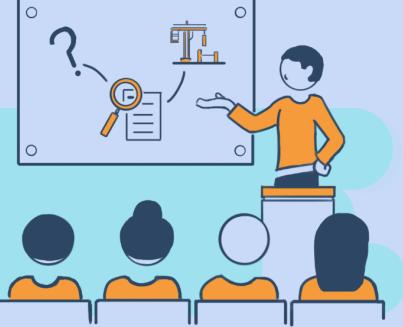
Mentors

Mentors will be available in person and mentors from Quick Release will be available on Discord in the **#quick-release** channel to assist anyone working on the Quick Release data challenge.

You may also message in the **#project-help** text channel anytime! But without further ado, here are our mentors and their availabilities:

Kyler Chin – In Person • Saturday, All Day	Grady Yu – In Person • Saturday, 10 AM – 10 PM
Randy Huynh – In Person • Saturday, All Day	Alicia Chuang – In Person • Saturday, 11 AM – 2 PM, 5 PM – 10 PM
Audrey Nguyen – In Person • Saturday, All Day	Quick Release – Discord (#quick-release) • Saturday, TBA
Sandra Nguyen – In Person • Saturday, All Day	William Hou – In Person • Saturday, TBA

Workshops

	
#1 🧑 Data Manipulation	#2 😎 Data Visualization
Time: Saturday, 10:45–11:45 AM Location: ISEB 1200 Topics: <ul style="list-style-type: none">• Data Quality Cycle• Melissa Education Portal<ul style="list-style-type: none">○ Lookup pages○ Cloud service APIs○ Programming tools Presenters: Melissa Data	Time: Saturday, 2:00–3:00 PM Location: ISEB 1200 Topics: <ul style="list-style-type: none">• How to translate your data visually• Making a dashboard• Communicate data-driven insights in graphical context Presenters: Ryan Yang
	
#3 🤖🧠 Intro to Tensorflow	#4 🚀 How To Submit & Present Your Datathon Project
Time: Saturday, 4:00–5:00 PM Location: ISEB 1200 Topics: <ul style="list-style-type: none">• Math & CS behind Tensorflow• Tensorflow Basics• Neural Network Implementation Prerequisite: Python, DeepNote Setup Presenters: Azra Zahin (Data @ UCI)	Time: Saturday, 7:00–8:00 PM Location: ISEB 1200 Topics: <ul style="list-style-type: none">• Project Submission• Presentation Tips• Mock Presentation Presenters: (Hack @ UCI)

Sponsor Networking

We are pleased to host representatives from two industry sponsors who are excited to meet you!

Chat with Melissa Data

Melissa Data is a global leader in data quality, address verification and identity resolution, helping over 10,000 businesses worldwide harness accurate data for a more compelling customer view. Representatives will be tabling in the lobby of ISEB on Saturday (4/13) morning. We highly recommend saying hello after you check in!

How to Talk To Sponsors

It can be scary talking to industry professionals! To help you navigate the process, we've put together the following checklist to help guide your interaction. Be confident!! You deserve to be here and sponsors are here to talk to YOU!!!

1. Before approaching the table, we recommend thinking of a few questions you would like to ask! Are you on the job hunt? If so, what sorts of positions are you looking for? Are you just interested in the company?
2. There may be a line to speak with the sponsors. If so, please line up and wait for your turn!
3. When it is your turn, introduce yourself with your name, major, and year! It may also help to state what you are hoping to get out of the interaction.
 - a. Ex.) Hello! My name is Petr Anteater and I'm a fourth-year student studying school spirit. I've never heard of [SPONSOR NAME] before and would love to learn more about what you do.
4. Close the conversation by thanking them for their time!

Sponsors may also have swag items and handouts for you to take – if offered, feel free to grab them :)

Frequently Asked Questions

Q: How many people can we have on a team?

Teams are capped at four people due to prize limitations.

Q: What's up with the liability waiver I signed on the form?

Since the datathon is held in-person on the UCI campus, we are required to have all attendees sign a liability waiver so that all parties are protected in the event of an unexpected incident. A similar waiver would be signed to be eligible for participation in any other event hosted by Data@UCI or otherwise. Though they are possible, **there are no risks anticipated at this event.**

Q: Can I come in late?

You are free to come and go as you please during the event after you have checked in. Please just make sure to have your badge on at all times when you are at ISEB. If you are checking in outside of the normal check-in window, please message on Discord so we may welcome you.

Q: Can I bring a friend who was not accepted into the Datathon?

Unfortunately, due to liability considerations and capacity restrictions, **only those that have been accepted to participate in the event are allowed in the venue.** We will be checking badges during the entire event and anyone not able to verify their identity will be asked to leave.

Q: I'm a beginner. Can I still have fun at the event?

Of course! Our Atlantis Datathon is beginner friendly and we welcome those with all levels of experience and backgrounds :) As a beginner, we highly recommend that you take advantage of as many resources as possible. Mentors are available throughout the entire event for questions and assistance. Feel free to approach or message them as often as you need! You are not bothering them, they are here for you!! Additionally, we recommend checking out all of the workshops offered at the event. Finally, please review the datathon resource sheet in detail! The document provides a wealth of information and is one of the most valuable resources we offer. If you are having trouble getting started on a project, we highly recommend checking out a recording of the "Data Projects Bootcamp" workshop! Remember, everyone here at the event is here to help! Take advantage of the mentors, workshop hosts, sponsors, and fellow hackers :))

Q: Can I use ChatGPT in my submission?

Creative use of generative AI tools to assist in your project in appropriate settings is allowed. Please keep in mind that a project designed and implemented largely with generative AI is certainly not going to score high on our judging rubric.

Q: Is my project still eligible to win Best Overall or Runner Up if I submit to a specific data challenge prize category?

Yes! Your project will still be eligible even if you submit to a specific prize category. For example, if my project is on one of Melissa Data's challenges and it is particularly outstanding, you might win Best Overall instead of the specific Melissa Data challenge.

Q: I have another question that is not found here

Please reach out to an organizer on Discord! We are here for you :)

Sponsors and Recognitions



We would like to thank Melissa, StrataScratch and Quick Release for providing amazing data challenges for our attendees to tackle.

We would like to thank Melissa, UCI ODIT, UCI MSBA, StrataScratch, Celsius, Red Bull, and That's It for providing us funding and merchandise that we can distribute at our datathon. We would also like to thank ASUCI's Student Programming Funding Board for their support in covering event costs.

Finally, we would love to thank all the mentors, workshop hosts, and attendees for making our datathon possible. Without you, there would be no datathon.



Data @ UCI, 2024
Irvine, California