



2024 INTER-UNI DATATHON

CONTENTS

02

WELCOME & INTRODUCTIONS

03

SCHEDULE

04

LOGISTICS

05

RULES AND EXPECTATIONS

06

TOP TIPS FOR SUCCESS

08

PRIZES

09

CONTACT US

WELCOME

Welcome to the Inaugural InterUni Datathon, bringing together Australia's top universities: UNSW, USYD, Macquarie University, the University of Melbourne, and Monash University.

This collaboration is a step towards fostering more opportunities like this in the future, where Australia's brightest minds in Data Science come together to compete, collaborate, and push the boundaries of innovation. We are proud to have Citadel Securities and H2O.ai as our sponsors, whose support makes this exciting event possible.

This datathon is a great opportunity for the specialists, experienced and those who are just beginning their data science journey to get their hands dirty with data. It's a great way to apply your knowledge and skills gained from the classroom in a real-world context.

We hope you are all excited for the weekend ahead. Be ready to channel your enthusiasm for data as we are ready to see some great data science-ing from you!

Kind Regards,
InterUni Datathon Team

SCHEDULE

SATURDAY September 14th

10:00 - 10:30 am	Opening Ceremony Kickoff, intros and question release
11:30 - 12:30 pm	Data Preparation Workshop Presenter: MDSS (Uni Monash)
12:30 - 1:30 pm	Recommended lunch break
2:30 - 3:30 pm	Machine Learning Workshop Modelling, Hyperparameter Optimisation Presenter: Julian Garrat (H2O.ai x UNSW)
4:00 - 5:00 pm	Open working time Students work on their respective Machine Learning Projects and slides.
5pm onwards	Closing Discord Help Channels Conclusion of the first day.

SUNDAY September 15th

9:00 - 10:00 am	Day 2 Welcome Students start polishing their slides.
10:00 - 11:00 am	Storytelling Workshop Presenter: James Hu (SUDATA)
12:00 - 1:00 pm	Slide Decks and Results due at 1pm
1:00 - 2:00 pm	Recommended lunch break and top 6 teams selected via Kaggle.
2:00 - 4:00pm	Finals Presentations 5 mins presentation + 2 mins Q&A
4:30pm - 5:30pm	Closing Ceremony Announcement of winners, acknowledgements and conclusion



LOGISTICS

Location

- **Sydney:**

Both days will be at **UNSW Colombo Theatre C** (Colombo Building K-B16-LG05).

- **Melbourne:**

Day 1: Room LG02-Woodside building for technology and Design, Monash University Clayton Campus (Uni Monash)

Day 2: Glyn Davis Building (Melbourne School of Design (Building 133) B121 **The Malaysian Theatre** (Uni Melb)

Kaggle Competition

Teams from Sydney are invited to participate in the Sydney Kaggle Competition, while teams from Melbourne will compete in the Melbourne Kaggle Competition.

The **top three teams from each city** will advance to the **final round**, where they will present their solutions to a panel of judges via Zoom.

The invitation to join Kaggle **will be released on the day of the InterUni Datathon.**

Submissions and Presentations

Your submissions should include a **presentation** and a **Kaggle submission**.

When submitting, submit into the **Google Drive folder provided with your team name** and upload files into it - these are time stamped and late submissions will not be accepted.

Finals will be in a Zoom webinar, recorded and open to the public. Please let us know in advance if you don't consent!

RULES & EXPECTATIONS

To ensure that all competitors have the best experience at this datathon, we ask that you follow these guidelines:



Teams must be comprised of 4-5 **people**



Show up on time and come prepared

- The datathon starts at 10 am sharp on the 14th of September.
- All team members must be present during the opening ceremony
- At least 1 member from each team must be present at each workshop
- Respect the helpers' and judges' time



Let everyone participate

- Respect each other's thinking and value everyone's contributions
- Value the learning from different inputs and listen to learn
- Stay open to new ways of doing things



Cheating will not be tolerated

- We expect that you complete this datathon to the best of your teams' ability
- All code that you have used that is not your own work must be quoted in comments with a reference to where it was found



TOP TIPS FOR SUCCESS

Simple is usually better

- Begin with a **strong understanding** of the context and scope of the problem, then start to formulate **simple hypotheses**.
- It's always exciting to run the most recent state-of-the-art model on your dataset, but that's not what we're looking for! Try starting from a **basic technical solution** and then slowly building up from there.

Plan your strategy

- Attend the data preparation session: This session is crucial for learning practical tips on data munging and cleaning, which will save you time later.
- Spend time getting familiar with the data: Dedicate an hour or so to explore the data, understand its structure, and identify key areas of focus for your analysis.
- Collaborate on defining questions or hypotheses: Work with your team to decide on the specific questions or hypotheses you want to investigate. This will ensure everyone is aligned and working toward the same goals.
- Frame your problem statement: Use a clear problem statement to guide your analysis. It's fine to revisit and refine your problem as you discover more through the data.
- Leverage workshops for feedback: Take advantage of workshops to ask questions, discuss any challenges you're facing, and get feedback on your modeling approach and assumptions.
- Conduct your analysis and apply modeling: Once you've gathered the data, dive into your analysis, testing your hypotheses with appropriate models. Create visualizations to help illustrate your findings.
- Focus on consolidation and code cleanup: On the second day, refine your findings, streamline your code, and make sure everything is clear and presentable.
- Prepare for your presentation: Attend the pitching session to learn effective ways to present your findings. Rehearse your pitch to ensure smooth delivery during the actual presentation.

TOP TIPS FOR SUCCESS

Teamwork

- Be **complementary** and take the time to understand your peers' skill sets. Your team may have diverse range of skills - communicate your strengths and get to know what you bring to the team!
- Be **comfortable** in your communications - set up a messenger chat or discord channel or anything else that works for your team!
- Take frequent and effective **breaks** in between working sessions and look out for your teammates.

Online collaboration tools

- You have the option of using **online collaboration tools** or doing the datathon **locally**. We highly recommend using or learning Git when participating in the datathon.
- **Git** is a version control system that lets you manage and keep track of your source code history. These online services allow you to manage repositories, share code and build on each other's work. [Find out more here](#)
- Some frequently used platforms and their guides are shared below:
 - **Bitbucket:** <https://support.atlassian.com/bitbucket-cloud/docs/tutorial-learn-bitbucket-with-git/>
 - **GitHub:** <https://opensource.com/article/18/1/step-step-guide-git>

Making the most of the workshops

- You're not going through this datathon without support! The workshops are designed to help guide you through the entire process. Expert facilitators will be available during these sessions to provide valuable insights and answer any questions you may have.
- Make sure to come prepared with specific questions or challenges you've encountered, so that you can make the most of the time and ensure all participants get the opportunity to learn.

Don't forget to have fun

- Remember that the datathon is as much of a **learning experience** as it is a competition!
- It will be a steep learning curve getting used to collaborating on code, writing up a proposal and thinking about all the implications of a given problem.

PRIZES

FIRST PLACE



Prize Pack

Citadel Securities sponsored 5x backpack and 5x jumper



Merchandise

H2O.ai Sponsored Merchandise



Prize

Citadel Securities sponsored 5x JBL Speakers

SECOND PLACE



Prize Pack

Citadel Securities sponsored 5x backpack



Merchandise

H2O.ai Sponsored Merchandise



Prize

Citadel Securities merchandise

THIRD PLACE



Cash Prize

Cash Prize awarded to best presentation



Cash Prize

Cash Prize awarded to best model with Kaggle metrics

Prizes will be mailed to respective Data Science societies.

CONTACT US

InterUni ORGANISING COMMITTEE

UNSW DataSoc

hello@unswdata.com

USYD SUData:

hello@sudata.org

Macquarie Uni MACS

macs.exec@gmail.com

Uni Melb DSCUBED

hello@dscubed.org.au

Uni Monash MDSS

mdss.monash@gmail.com

