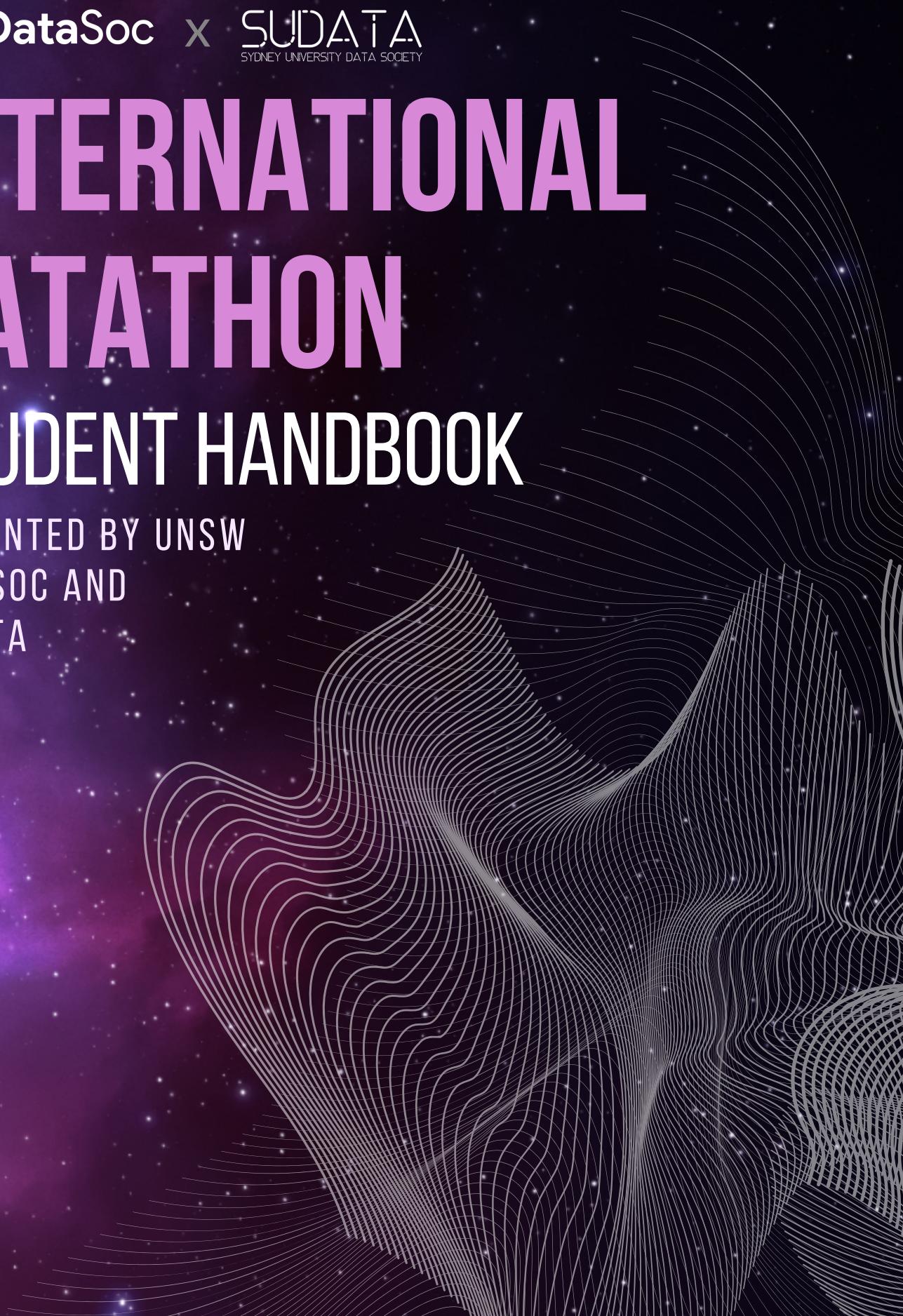




INTERNATIONAL DATATHON

STUDENT HANDBOOK

PRESENTED BY UNSW
DATASOC AND
SUDATA



CONTENTS

01 | President's Address

02 | Intro to DataSoc

03 | Intro to SUDATA

04 | Schedule

05 | Logistics

06 | Mentors

07 | Rules and Expectations

08 | Top Tips for Success

10 | Prizes

11 | Contact Us



PRESIDENT'S ADDRESS

Hello and welcome to DataSoc's Flagship Datathon!

The Datathon is YOUR opportunity to scope, solve and present a real-world problem from start to end. We've brought together universities across Australia and China, in an extension of the bilateral relationship that was established in 2019 between UNSW DataSoc and Tsinghua University Big Data Association.

For this Local Round, we have teams competing from both USYD and UNSW, so take this opportunity to meet other students and learn from each other. Whether you're a complete newbie or seasoned Datathon champion, be prepared to dive deep into a real-world issue over the next 24 hours - we're keen to see what you can come up with.

We are incredibly fortunate to have the support of the UNSW Data Science Hub, UNSW School of Maths and Stats and FASIC in making this a reality.

All the best for the Datathon, and most importantly, we hope this is the start of a lifetime of exploring patterns and possibilities.

Laurel Lu and Cherelle Johannes

President, DataSoc and SUDATA.



INTRO TO DATASOC

Since DataSoc was founded in 2017 by a group of 5 friends, it has grown to a community of over 3300+ students. Our society now captures a diverse range of skill sets, university stages and degrees - including but not limited to actuarial studies, business analysts, computer scientists and of course, data scientists.

DataSoc is now not just a UNSW society, but also the largest constituent Data Science Society in Australia. We bring together a community of data science professionals and enthusiasts who are driven to discover how data shapes the world around us. We aim to offer an open platform to connect, educate and empower students with the skills and knowledge to unlock the power and potential of data science.

Today, we have grown to be a team of 50 and we look forward to bringing students more opportunities to advance their data science skills and to connect to the industry.

Get involved and keep up to date with what we get up to!

Signup: <https://forms.gle/U7gXM87okE7ojfdDA>

SpArc: <http://bit.ly/datasoc-sparc>

Facebook: <https://www.facebook.com/DataSoc/>

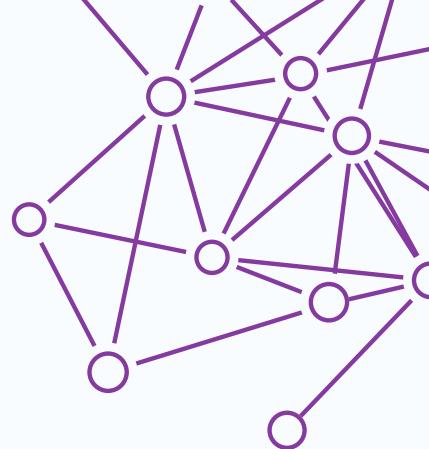
LinkedIn: <https://au.linkedin.com/company/datasoc>

Instagram: <https://www.instagram.com/unswdatasoc>

Website: unswdata.com

General Enquiries: hello@unswdata.com

INTRO TO SUDATA



Sydney University Data Society (SUDATA) was established in 2020 as the de-facto community for all students with an interest in data science. We are dedicated to providing our students with connections to industry opportunities and a platform to widen your network and enrich your data science experience.

Our mission is to empower like-minded students seeking a data-driven yet fun community by providing the resources to further your passions, and the environment to build lifelong bonds. With members ranging from engineering to arts, we aim to unite students through a shared hatred of R and cultivate an accessible social environment for students of any discipline looking for a place to share, learn and make friends.

Become a member: <https://sudata.getqpay.com/>

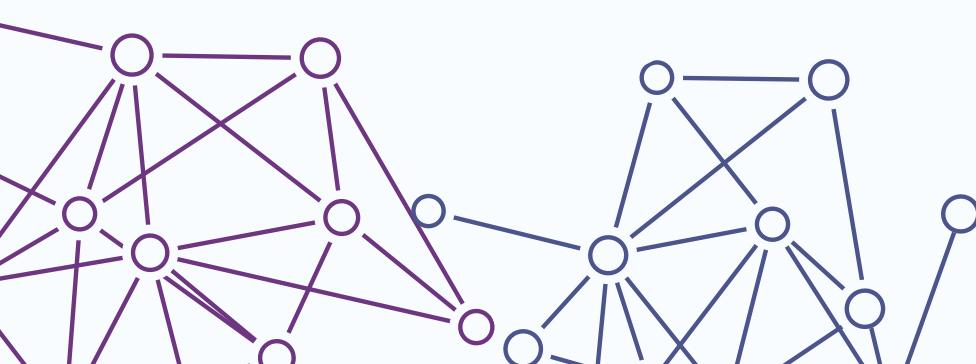
Join us at our events!

Facebook: <https://www.facebook.com/usyd.sudata/>

Instagram: <https://www.instagram.com/usyd.sudata/>

Come talk to us <3

Discord: <https://discord.gg/csRnUsaEbM>





SCHEDULE

SATURDAY OCTOBER 30

9:00 - 10:00 AM

Case Brief Kickoff

Kickoff, intros and case release

2:00 - 5:00 PM

(Your team will be invited to a one hour time slot)

Mentor Consultation Session

SUNDAY OCTOBER 31

10:00 AM

Submissions Due

10:30 AM - 12:00 PM

Heats Round

2:00 - 4:00 PM

Finals Round

4:00 - 4:30 PM

Closing

LOGISTICS

COMMUNICATION & ADMIN

We will be using Slack and Google Drive for the Local Round.

- Be sure to join the **International Flagship Datathon** Slack workspace.
Some important channels include:
 - **#announcements** - important announcements, such as case release
 - **#general** - general updates and additional information
 - **#faq** - common questions that have been answered across different team channels
 - **#q-and-a** - ask questions to the organisers (DataSoc)
- Make sure that you're added to your private team channels. If you haven't been added, reach out to us in the **#logistics** channel.

MENTOR FORUM

- A mentor and a few DataSoc representatives will be added to your private team slack channel, and will respond to questions that you post in the channel.
- You can expect semi-frequent responses.
- Please be sure to:
 - Summarise your team's question
 - Include all and only relevant materials within a self-contained post, e.g. results or code snippets, slides, and files
 - Be conscious of your mentor's time and efforts by prioritising important questions
 - Treat them with respect and thank them for their time!
- There will be a **1hr consultation session**, held on zoom, with a mentor that your team will be invited to attend. This is a valuable opportunity to receive live advice from a professional and we highly encourage you all to utilise it.

SUBMISSIONS

- You are allowed a 7 minute presentation, which will be directly followed by a 3 minute Q&A on your presentation by judges.
- Submit your slide deck and code into your allocated Google Drive folder with the correct team number by 10.00am Sunday 31st October. You may also be required to submit other relevant case material. No late submissions will be accepted - folders will be locked strictly at 10:00am.
- Heats will take place simultaneously on Zoom, in different breakout rooms. Only one person needs to share their screen while their team presents. Make sure that you have been added to your private heat channel on Slack!
- Finals will also take place on Zoom, but will be recorded and made publicly available. If you don't consent to being recorded, please let us know prior!

MENTORS

CBA

JOE ZHANG



AUTUMN
COMPASS

ESTHER LAM



RAY RUIYUAN
ZHANG



THOMAS CRAWLEY



RELEVANCE

AI

JACKY WONG



RULES AND 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 3 or 4 people, with a minimum of 1 non-cis male.



Show up on time and prepared

- **All team members must attend** the case brief kick-off at 9:00am on Saturday 30th October
- **All team members must attend** their allocated heats round to present on Sunday 31st Oct between 10:30am and 12:00pm
- Respect the mentors' time by succinctly asking only important and relevant questions on Slack
- Check Slack regularly for announcements and updates



Let everyone participate

- Respect each other's thinking and value everyone's contributions
- Stay open to new ways of approaching problems



No cheating allowed

- We expect that you complete this datathon to the best of your teams' ability
- For code that is not your own, clearly reference the source you obtained it from and place the code snippet in a comment above where you used it



Do your best

- An unfinished product is better than nothing at all; you will still be able to receive judges' feedback afterward
- Make the most of a rare international competition that provides mentors!
- Every team who submits will be able to present at the heats



If you do need to drop out of the competition

- Please email competitions@unswdata.com directly so we can remove your Slack channel and reallocate your mentors

TOP TIPS FOR SUCCESS



SIMPLE IS USUALLY BETTER

- Keep your answers and solutions to the problems simple and make sure it properly answers the question. Remember, “Less is more”.



PLANNING IS YOUR BEST FRIEND

- Have an eye for detail and take time to fully understand the intricacies of the dataset.
- Perform data cleaning and exploratory analysis before doing predictive analytics.
- Plan out deadlines and delegate tasks among team members. One of the biggest reasons why teams do poorly is not having planned out deadlines. This will help spread out the workload evenly among your time.
- Have a contingency plan in place in the event of any technical difficulty or member changes.



THE SLACK WORKSPACE IS A POWERFUL TOOL

- **Your team channel:** Put your team specific questions here to be answered by your DataSoc representatives.
- **FAQ channel:** Refer to this channel to see if your question has already been addressed. DataSoc will be collating frequently asked questions across the team channels.

TOP TIPS FOR SUCCESS



VISUALISE YOUR DATA

- The old saying "A picture is a thousand words" is vital when it comes to communicating your findings.
- Some popular visualisation tools are: R, Tableau, Excel.



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.

Some frequently used platforms and their guides are shared below:

- **GitHub:** <https://opensource.com/article/18/1/step-step-guide-git>
- **Bitbucket:** <https://www.atlassian.com/git/tutorials/learn-git-with-bitbucket-cloud>



DON'T FORGET TO HAVE FUN

Don't forget to have fun with your team members while doing the Datathon! After all, the datathon is as much of a **learning experience** as it is a competition.

This is a fantastic opportunity to collaborate on code, write up a proposal and think about all the implications of a real-world problem.

The biggest tip in being successful at the Datathon is by having fun in the process!

PRIZES

LOCAL ROUND PRIZES (ACROSS UNSW AND USYD)

Top 6 teams advance to International Round (3 from UNSW, 3 from USYD)

1ST PLACE



DataSoc Merch



\$400 Cash



Coffee Chat

with an industry professional

2ND PLACE



\$300 Cash



Coffee Chat

3RD PLACE



\$200 Cash



Coffee Chat

4TH TO 6TH PLACE



\$100 Cash

LUCKY DRAW
\$100 Uber Eats
Voucher
ANY PARTICIPATING TEAMS

INTERNATIONAL ROUND PRIZES

1ST PLACE



\$2000 Cash



2ND PLACE



\$1500 Cash

3RD PLACE



\$1000 Cash

CONTACT US

If you have any inquiries about the competition, message us on Slack in the q-and-a channel. Otherwise please contact us via our respective society emails.

UNSW DataSoc

competitions@unswdata.com

Sydney University Data Society

sydney.uni.data.society@gmail.com