

2020 DATATHON STUDENT HANDBOOK

PRESENTED BY UNSW DATASOC AND ATLASSIAN



DataSoc



ATLASSIAN

CONTENTS

02

WELCOME & INTRODUCTIONS

05

SCHEDULE

06

LOGISTICS

07

RULES AND EXPECTATIONS

08

TOP TIPS FOR SUCCESS

10

PRIZES

11

JUDGING CRITERIA

12

CONTACT US

WELCOME

Welcome to the UNSW Data Science Society's annual datathon! This year, we are pleased to announce our partnership with Atlassian to bring you this event.

Alongside mentors from Atlassian, BCG, Accenture, EY, KPMG, Westpac, Deloitte and UNSW School of Maths and Stats, we look forward to delving into the many powerful possibilities that data presents in solving problems.

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.

I 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!

Ellen Wang
President, DataSoc 2020



INTRO TO DATASOC

Since DataSoc was founded in 2017 by a group of 5 friends, it has grown to a community of over 2800+ 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!

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

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

Contact us: hello@unswdata.com

INTRODUCTION TO

ATLASSIAN

Throughout Atlassian's 18 years, our trajectory has taken a variety of unexpected twists and turns, but all the while, our goal has remained simple and direct: To help unleash the potential of every team. And because we've continued to use that goal to inform our vision, we've contributed to everything from medical advancements and space travel, to disaster response and pizza deliveries — our products help empower teams worldwide, and advance humanity through the power of software.

As we've grown and our customer base has continued to expand and diversify, we've continued to focus on making tooling and processes that are equally sustainable and equitable for our customers and their communities. We understand that high performing teams are anchored in people with varied perspectives and diverse approaches to problem-solving. In order to support those teams and individuals to do their best work, they need the ability to be authentic at work and feel they're empowered to make Atlassian a better place for everyone.

We're a team of 4000+ Atlassians supporting an international group of 170,000+ customers. We build tools like Jira, Confluence, Bitbucket, and Trello to help teams across the world become more nimble, creative, and aligned - collaboration is the heart of every product we dream of, design, and ship at Atlassian.

Data driven-by-design

Atlassian is over a decade old, but our Data & Analytics team is much younger. We'll have to blaze new trails to enable important growth decisions, so we're constantly growing, learning, and trying to do things differently. You'll be joining a team that is crazy smart and very direct. We ask hard questions and challenge each other to constantly improve our work. We are self-driven but team oriented. We're all about enabling growth by delivering the right data and insights in the right way to partners across the company.



SCHEDULE

SATURDAY OCTOBER 3

9:00 - 10:00 am	Opening Ceremony Kickoff, intros and question release
10:00 - 11:00 am	Data Preparation Workshop Presenter: Geoff Pidcock (Atlassian)
12:00 - 1:00 pm	Recommended lunch break
2:00 - 3:00 pm	Mentoring session 1 Data cleaning, feature engineering and making assumptions
3:00 - 4:00 pm	Open consultation session Ask our industry reps anything!
4:00 - 5:00 pm	Mentoring session 2 Suitable modelling techniques and choosing effective visualisations
6pm onwards	Closing Slack #open-consultation and #mentoring-sessions channels

SUNDAY OCTOBER 4

9:00 - 9:10 am	Day 2 Welcome Submit Code at 10am
10:00 - 11:00 am	Storytelling Workshop Presenter: Perry Wang (Atlassian)
12:00 - 1:00 pm	Slide Decks due at 1pm Recommended lunch break
1:00 - 2:00 pm	Heats Presentations 5 mins presentation + 1 min Q&A
3:00 - 3:15 pm	Announcement of Finalists
4:00 - 5:00 pm	Finals Presentations 7 mins presentation + 2 mins Q&A
5:30 - 6:00 pm	Closing Ceremony Announcement of winners, acknowledgements and conclusion

LOGISTICS

Communication

Join our Slack workspace: <https://bit.ly/3jpJxmf>

- **#general** - all event-wide announcements + Zoom links
- **#qanda** - reach out to DataSoc organisers
- **#mentoring-sessions** - topical questions for mentors
- **#open-consultation** - reach out to industry reps

Mentoring and Consultation

Mentoring sessions are for you to get help on the datathon problem. We have recommended topics which reflect the analysis process and reps' areas of expertise.

The **open consultation** session can be used to connect with industry reps in any way you like, such as:

- Continuing to ask technical questions
- Getting to know their career journey
- Finding out more about the companies they work at

Mentors will use:

- if they are considering your question
- to mark your question as answered

Feel free to privately message specific mentors on Slack given availability.

For easier discussion, you can:

- Use the call feature on Slack
- Jump back into our Zoom link and we'll put you in a breakout room

Refer to the **Industry Representatives Booklet** for their profiles!

Submissions and Presentations

When submitting, make a **Google Drive folder with your team name** and upload files into it - these are time stamped and late submissions will not be accepted. Make sure that hello@unswdata.com has edit access.

Heats will be held in simultaneous breakout rooms on Zoom. One member of each team can share their screen as you all present.

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 **1-4 people**



Show up on time and come prepared

- The datathon starts at 9am sharp on the 3rd of October
- 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 mentors' and judges' time (15min max per consultation)



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

Written by Arik Friedman, Principal Data Scientist at Atlassian

- Attend the **data prep session** for data munging tips.
- Allocate an hour to get familiar with the data sources, explore the data, decide on specific aspects you'd like to focus on.
- Agree with the group on a specific set of questions/hypotheses you'd like to study. Form your problem statement to frame your analysis - it's ok to reframe your problem at a later stage as you learn more through your data analysis
- Pull together the specific data you need for your analysis. Join various data sources if needed, create new features if needed to support your analysis.
- The **first mentoring session** is a great opportunity to bring up any questions or challenges you face through the data munging process, and ask for feedback on your approach and assumptions.
- Explore the data set you put together and apply modeling as needed to carry out your analysis and test your hypotheses. Put together visualisations that demonstrate your findings.
- The **second mentoring session** provides a chance to consult on modeling and visualisation approaches. Throughout the analysis you may find that new hypotheses emerge, or that additional data is needed to dig deeper into some aspects, so you may go back and forth between those stages.
- On the second day, focus on consolidating your findings and cleaning up the code.
- Attend **the pitching session** for tips on delivering your findings effectively.
- Form your pitch and rehearse your 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 your mentors

- You're not doing this datathon alone! With a dedicated team of mentors manning the Slack workspace during mentoring and open consultation hours, you will be well supported.
- Come prepared to mentoring sessions with **clear questions in mind** so that we can give all contestants an equal opportunity to learn from our talented mentors.

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



Coffee Chat

with a Senior Atlassian Representative
Potential fast track (subject to eligibility and roles)



\$900

Prize money from DataSoc and UNSW Maths and Stats



Merchandise

from Atlassian and UNSW Maths and Stats

SECOND PLACE



Resume Review

with an Atlassian senior campus recruiter



\$600

Prize money from DataSoc and UNSW Maths and Stats



Merchandise

from Atlassian and UNSW Maths and Stats

THIRD PLACE



\$600

Prize money from DataSoc and UNSW Maths and Stats



Merchandise

from Atlassian and UNSW Maths and Stats

JUDGING CRITERIA

Criteria	Considerations	Marks
Answer Quality	<p>1. Hypothesis</p> <ul style="list-style-type: none"> Defined a clear problem statement and hypothesis that is used to frame the analysis <p>2. Research - translating the question into a data problem</p> <ul style="list-style-type: none"> Extensive detailed research/usage of country-specific characteristics based on problem statement or hypothesis Usage of multiple reputable sources to support findings Summarises findings and hypotheses concisely Clear referencing to data sources <p>3. Analysis</p> <ul style="list-style-type: none"> Usage of clear quantitative analysis as evidence to support theories Efficient and hypothesis-driven usage of datasets (provided and/or external) to support answer Uses compelling visualisations to support thesis/hypotheses <p>4. Insights</p> <ul style="list-style-type: none"> Conclusions are clearly communicated and interpreted The conclusions and insights address the problem statement and answer the problem/question effectively. 	5 10 20 15
Presentation Quality	<ul style="list-style-type: none"> Effectively conveys the answer through appropriate visualisations Demonstrates an understanding of the context and impact of the analysis <ul style="list-style-type: none"> Concise visual presentation Clear delivery and message throughout the presentation Engaging and easy to follow 	20 10
Creativity of Investigation	<ul style="list-style-type: none"> Degree of creativity and original thinking in identifying novel questions to ask about the data Innovative method of analysis to generate useful insights 	10
Code Quality	<ul style="list-style-type: none"> Readability and interpretability, effective documentation Modelling detail and accuracy Reproducibility of analysis, transparency and accountability 	10
Total		100

CONTACT US

DATASOC ORGANISING COMMITTEE

Primary: Laurel Lu

Vice-President (External)

external@unswdata.com

Secondary: Ellen Wang

President

hello@unswdata.com

Marketing: Nigel Lin

Vice-President (Internal)

internal@unswdata.com

Question/Criteria: Ivan Hou

Education Director

resources@unswdata.com

Corporate Outreach:

Vincent Chen

Sponsorships Director

Corporate Outreach:

Winston Sun

Sponsorships Director

ATLASSIAN ORGANISING COMMITTEE

Question/Criteria:

Arik Friedman

Principal Data Scientist

Logistics:

Zuli Posada

Senior Campus Recruiter

DESIGN CREDITS

Vincent Chen :3

Ellen Wang

Nigel Lin



DataSoc



ATLASSIAN