

# Case Studies in Data Science – Week 1

COSC2669/COSC2816

Damiano Spina & Johanne Trippas



## Acknowledgement of Country

RMIT University acknowledges the people of the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation on whose unceded lands we conduct the business of the University.

RMIT University respectfully acknowledges their Ancestors and Elders, past and present.

RMIT also acknowledges the Traditional Custodians and their Ancestors of the lands and waters across Australia where we conduct our business.



# Maculelê Song

*Eu sou um bom menino  
Minha mãe soube me educar  
Pisando em terras alheias  
Eu piso bem devagar...*

I am a good boy  
My mother knew how to educate me  
When stepping on foreign lands  
I step very slowly





# Womin Djeka

Boon wurrung\*\*

Pronounced **womin-jekaah** or **womin-checkkaah**

/Wo/ with a 'w' and not 'v', rhymes with 'go'; /min/ rhymes with 'tin';  
/jek/ or /djek/ rhymes with 'trek'; /a/ as the sound in 'aah'.

**Womin** Come? Or ask to come? **Djeka** What is your purpose? And/or what are your intentions?

This pronunciation comes from Boon wurrung and Woi wurrung as people of the Birrarung-ga; meaning 'country of the river'.

\* **Woi wurrung** (pronounced **woy-wur-rung**) - The Wurundjeri People of the Kulin Nation

\*\* **Boon wurrung** (pronounced as **bwoon-werrung**) - The Boon wurrung People of the Kulin Nation

☰ ▼ Reconciliation - Australian First Nations - On Country

☰ 📄 Our commitment to Reconciliation

☰ 📄 Welcome to Country and Acknowledgement of Country

☰ 📄 Womin Djeka Indigenous Orientation (student micro-credential)

☰ 📄 Truth-Telling: Racism and Reconciliation (student micro-credential)

☰ 📄 RMIT On Country City Walking Tours

☰ 📄 Significant Australian First Nations landmarks on RMIT Campuses

☰ 📄 Additional resources & student Indigenous electives



---

# Teaching Team

(E-mail for appointments)

**<casestudies@rmit.edu.au>**

## Course Coordinators

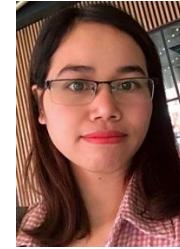
Dr Damiano Spina

Dr Johanne Trippas

## Tutors & Mentors

Lin Tian, H Ruda Nie, Akriti Verma,  
Saliha Muradoğlu, Chris Santosh John

**Guest presenters from industry**



# Why this course?

- Hands-on, industry-driven experience
  - What is Data Science?
  - What are the required and desirable skills?
  - How to plan, conduct, and present a Data Science project that is valuable to business?
  - What are the nuances of Data Science roles in different domains (e.g., banking, digital health, consultancy,...)
  - Data Science + Society





# Course Learning Outcomes (CLOs)

1. Obtain practical experience through applying data science concepts and techniques learnt in courses such as Data Science Professional by performing a data science project.
2. Develop a data science project to analyse, theorise and make conclusions about new situations in data science professional practice valuable to business and industry.
3. Contrast social impact and professional issues in the realm of different data science domains.
4. Analyse and evaluate professional practice case studies in teams, and critically assess the work of peers.
5. Communicate effectively to a variety of audiences through a range of modes and media, specifically, through written technical reports and oral presentations.





# Prerequisite Knowledge

You are expected to already have some technical knowledge of data science before commencing this course, hence prerequisite:

**COSC2670/COSC2738 - Practical Data Science**

COSC2669: This course is a preparation for working as a data scientist and is pre/co-requisite for the:

**Data Science Postgraduate Project**



# We Value Your Feedback

We need your feedback to keep improving the course!

**Course Experience Survey (CES)** is a confidential and anonymous 5-minute survey measuring student sentiment towards their course experience.

The results are used to enhance the teaching and learning outcomes at a subject level.



# CES Results from last years

## Aspects that needed improvement

- “It requires huge amount of self-learning. Some assignment tasks do not include material from any lectures.” -> More material to guide you through assignments (e.g., LaTeX tutorial today)
- Timetabling -> From starting at 7.30pm to a more friendly time: 10.30am on Tuesday
- There were some occasional instances where the staff's use of Canvas could be improved. An example was trying to locate the video for us to peer review, which was a bit difficult and unclear. -> Individual Task 1, Part 2 is now a standalone submission



# CES Results from last years

## Good aspects of the course

- Excellent guest speakers from industry
- Group project and presentations
  - “Working in groups. Doing self research to emulate an industry-like experience while developing the project”
- “Teaching team is very passionate”
- “The course wasn't too difficult, just had to allocate time wisely with group members to make it enjoyable.”







# (Tentative) Schedule

1. **Course introduction and LaTeX workshop**
2. Bias in data science Hatch
3. Data science in the e-commerce domain marqo.ai
4. Doing a PhD in data science CHAI & ADM+S
5. Iterative data science
6. Data science in design Canva
7. Data science in the biomedical domain and government policy MCRI & Solar Victoria
8. Data science in the energy domain EnergyAustralia
9. Data storytelling ANZ
10. Data science in the logistics domain Australia Post
11. Project Presentations (you, breakout rooms on Teams)
12. Project Presentations (you)



# Assignments

# Assessments

## **Assessment 1: Individual Assessments (40%)**

Two written individual assessment tasks during the semester.

Individual Task 1 – Week 4 (20%)

Individual Task 2 – Week 8 (20%)

## **Assessment 2: Work Integrated Learning Project (50%)**

This project will be undertaken in teams of 4-6 members.

Milestone – Week 5 (5%)

Oral presentation – Week 10 (20%)

Written report – Week 12 (25%)

## **Assessment 3: Individual Reflective Portfolio (10%)**

The portfolio will consist of fortnightly reflections on the learning outcomes of Assessments 1 and 2.

Weeks 2, 4, 6, 8, 10 – 2% each

**Check Canvas for more details!**



# Setting Expectations

- You don't "lose" marks, you earn them!
  - High Distinction (HD) ~ Above Expectations
- **Detailed rubrics** > Feedback in class > Comments in assignments
- Questions?
  - **FAQ** -> Use the **Discussion Forum** in Canvas
  - Keep in mind the silence policy (48hr before the deadline of assignments)

## Frequently asked questions

[It is expected that this document will grow during the course. We encourage you to check it regularly.]

## Contacting the course coordinators or teaching team

Q: I have a question. Shall I send an email to the course coordinators/teaching team?

A: No, unless you want to discuss something personal.

Most of the questions related to the course are not personal, and it is likely that (i) have already been answered in these FAQs or in the discussion forums on Canvas and/or (ii) the answers to your questions are informative to other students. Therefore, Canvas forums will be the primary communication channel to ask questions to the teaching team.

Q: I have a personal question; who and how should I email?

A: Email [casestudies@mit.edu.au](mailto:casestudies@mit.edu.au) and start your email with the subject [COS2669/COS2186] *subject matter*. E-mails or any messages (including Teams since we do not use Teams for this course) without the [COS2669/COS2186] tag and COSC-email address will not be monitored. An email could look like this

> Send

To Case Studies in Data Science X

Cc

[COS2669/COS2186] ELP extension for John Doe (s1234567890)

Hi Damiano and Johanne,

I have an approved ELP and would like...

[EMAIL BODY]

Best wishes,

John Doe  
s1234567890





# Marking

- You start with a blank page and 0 marks
- You earn marks for what you do
- **You get rewarded for what you do**
- You don't get penalised for what you don't do
- Full marks is possible, but rare and needs to be earned



# Marking

- Distinction:
  - “a difference or contrast between similar things or people”
  - “**excellence that sets someone or something apart from others**”  
e.g., a grade in an exercise denoting **excellence**.
- Excellence:
  - “the quality of being outstanding or extremely good.”

NN	PA	CR	DI	HD
Fail	Pass	Credit	<i>Distinction</i>	<i>High Distinction</i>
0-49	50-59	60-69	<i>70-79</i>	<i>80-100</i>



# Marking

- Grades measure **performance**, not **ability** (!!)
- **You have the ability** because you are in this course (!!)
- “You play how you train”
  - Your performance is determined by your training
  - (Include your favourite sporting analogy here: e.g., Capoeira 😊)
- Our job is to get you to **perform** to the best of your **ability**
- Completing your work each week is the best way to ensure good performance ...



# Academic Integrity

Academic integrity is about honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge and ideas.

You should take extreme care that you have:

- Acknowledged words, data, diagrams, models, frameworks and/or ideas of others **you have quoted (i.e., directly copied)**, summarised, paraphrased, discussed or mentioned in your assessment through the appropriate referencing methods,
- Provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites.

Make sure you use in-text citations and appropriate referencing style:

<https://www.rmit.edu.au/library/study/referencing>





# Academic Integrity in the Context of Generative AI

Read carefully the terms and conditions in the assignment descriptions on Canvas

**Declaration of Use or Not Use of Generative AI is mandatory.  
The use of Generative AI for Individual Reflective Portfolio is not allowed.**



---

# WIL Project: Start Creating Your Groups Now!

## Teams of 4-6 members

1. Introduce yourself on the **Virtual Café on Canvas**
  - Describe the **technical and/or professional skills** you bring to the data science team
  - Describe **complementary skills** you are looking for
2. Assign members to the same group on Canvas  
(pick one from 1...80 already available, don't create new ones!)
3. Register group filling out the form: <https://forms.office.com/r/UMcGZPi3P6>
  - **Reflection Portfolio in Week 4 asks for a Group ID.**
  - **First WIL Project Milestone due on Week 5**



---

# WIL Project Theme

**Test-driven Retrieval-Augmented Generation (RAG) for a domain of your interest:**

**- Education, Finance, Architecture, Engineering, Health, Science, Art**

You can use any tool to build it, but you need to use a quantitative approach to measure performance -> start here: <https://bit.ly/walert>



# WIL Project: Why RAG?

Goal: Create chatbots that can answer user questions in various contexts by cross-referencing authoritative knowledge sources

- Every organization would like to have a customised solution to answer questions from their (heterogenous, sparse, distributed) knowledge base
- IR + LLM -> RAG

RMIT is an international leader in the field

1<sup>st</sup> place at SIGIR'25 LiveRAG Challenge

<https://liverag.tii.ae/>



<https://www.admscentre.org.au/arc-centre-of-excellence-team-wins-global-liverag-challenge-at-sigir-2025/>





# Assignments Already Available on Canvas

- Individual Task 1 (Due Week 4)
- Reflective Portfolio 1 (Due Week 2)



# Writing for Data Science

Hands on activity

# Have you heard of LaTeX?



# What is LaTeX?

- **LaTeX** (pronounced '*Lay-Tek*' or '*Lah-Tek*') is a typesetting system
- You write a plain text file (`.tex`) using commands to describe your document's structure and content
  - You define the *structure* and *content*; LaTeX creates the *style*
- You compile it to create a professionally formatted PDF
  - It provides superior quality for mathematics, tables, and scientific documents





# Advantages: Never hand in sloppy documents!

- **Reproducibility and version control**
  - Plain text (`.tex`) file
  - Track changes, collaborate, and manage versions using Git
- **Automate**
  - LaTeX handles the numbering of figures, tables, and equations, and compiles your bibliography and table of contents
- **Professional**
  - Create quality reports, academic papers, and theses that are formatted to industry standards







# LaTeX templates

- **LaTeX** templates are pre-designed document structures
- Templates usually include
  - The main `.tex` file
  - A class (`.cls`) or style (`.sty`) file
  - A bibliography (`.bib`) and a bibliography styling (`.bst`) file
  - Often a `README` file
  - A folder for images and one for tables





# (A few) LaTeX best practices

## Organisation and structure

- Maintain clean and organised code
- Use clear and consistent naming conventions for labels and files
- Comment on complex code to ensure clarity
- Divide large documents into smaller files using `\input` or `\include`

## Compilation and debugging

- Compile the code successfully before stopping work each session
- Limit the use of `\usepackage` to the preamble to avoid compilation issues
- Learn to interpret LaTeX error messages and utilise log files for debugging

## Efficiency and maintenance

- Avoid unnecessary packages and commands to prevent conflicts or document bloat
- Establish a consistent workflow: compile frequently and address minor errors early
- Label copies periodically, especially before making major changes

## Documentation and reference

- Use reference managers or BibTeX for handling citations efficiently



# What is Overleaf?

6

My Paper on Astronomy and Computing

ReviewShareSubmitHistoryChat

Code EditorVisual Editor

Normal text

Recompile

```
\begin{equation}
\frac{d}{dt} \left( \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) - \frac{\partial \mathcal{L}}{\partial q_i}
\end{equation}
```

\section{Case Study: Image Analysis of Galactic Structures}

```
\begin{figure}[h]
\centering
\includegraphics[width=0.8\linewidth]{galaxy.jpg}
\caption{View of the Milky Way Galaxy}
\label{fig:galaxy-figure}
\end{figure}
```

\subsection{Image Processing Algorithms}

## Exploring the Nexus of Astronomy and Computing

Dr. Aurora Celestia Starlight  
*Department of Astrophysics, Stellar University*

---

### Abstract

This paper delves into the intricate relationship between astronomy and computing, exploring the impact of computational techniques on astronomical research. A case study is presented, highlighting the application of advanced algorithms to analyze astronomical data. The study includes an equation, an image, and a data table to illustrate key concepts.

**Keywords:** Astronomy & Computing, Astronomy, Computing, Interdisciplinary Research, Data Analysis

---

### 1. Introduction

Astronomy and computing have become inseparable companions, with computational techniques playing a pivotal role in advancing our understanding of the cosmos. This paper provides a comprehensive study of the intersection between these two fields, emphasizing the transformative effects of computing on astronomical research.

The application of computational techniques in astronomy has revolutionized data analysis, simulations, and modeling. One fundamental equation capturing the essence of computational modeling in astrophysics is:

---

### 3. Case Study: Image Analysis of Galactic Structures

To demonstrate the practical application of computational techniques, we present a case study involving the analysis of galactic structures using advanced image processing algorithms.

---

### 4. Case Study: Image Analysis of Galactic Structures

To quantitatively assess the impact of computational methods, we present a summary table of key parameters derived from the image analysis.

---

*Preprint submitted to Journal of Computational Astrophysics and Data Science*

July 1, 2024




Figure 1: View of the Milky Way Galaxy





# LaTeX templates

- All written tasks are complete in LaTeX (except WIL Project Proposal)
- Templates and LaTeX tutorial are provided
  - Individual Task 1 will have much details on LaTeX
  - Progressively, you need to be pro-active and look things up if you want to use LaTeX well



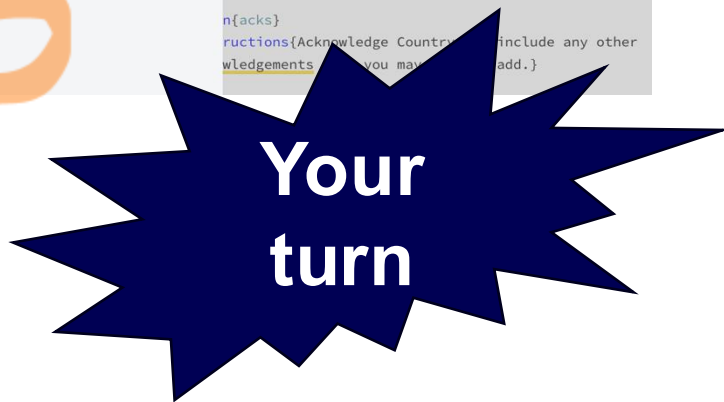
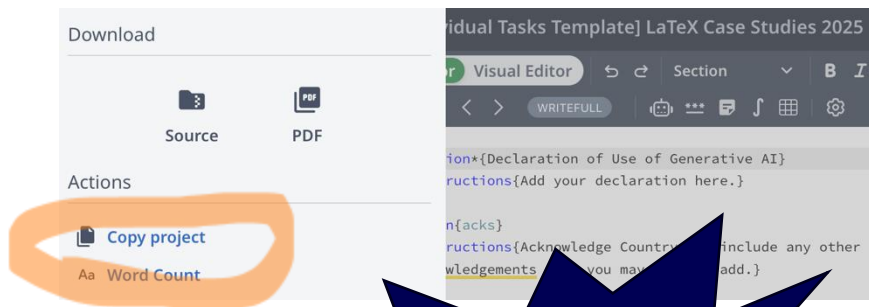
**Quick  
demo**





# <https://www.overleaf.com/register>

- Create an Overleaf account with your email
- <https://www.overleaf.com/read/hjqhmfnnvwtgd#d4945b>

A screenshot of the Overleaf registration page. At the top is the Overleaf logo. Below it is the text 'Create an account'. There are two buttons: 'Continue with Google' and 'Continue with ORCID'. Below these is the text 'OR'. A green box highlights the registration form, which includes fields for 'Email' and 'Password', and a 'Create account' button. Below the form is the text 'OR' and a link for 'Work/university single sign-on'. A red 'X' is drawn over the 'Log in with SSO' button. At the bottom, there is a link for 'Already have an account? Log in' and a footer with legal notices.



# References and DOIs

- Use a `.bib` file to store your references in BibTeX format. Each entry includes metadata like author, title, journal, year, and DOI
- To cite a reference in your text:  
`\cite{key}`  
`\cite{trippas2020towards}`

```
@article{trippas2020towards,  
  author = {Johanne~R. Trippas and Damiano Spina and Paul  
and Hideo Joho and Sanderson, Mark and Cavedon, Lawrence},  
  title = {Towards a Model for Spoken Conversational Search},  
  journal = ipm,  
  year = {2020},  
  volume = {57},  
  number = {2},  
  issn = {0306-4573},  
  url = {https://doi.org/10.1016/j.ipm.2019.102162},  
  doi = {10.1016/j.ipm.2019.102162},  
  pages = {1--19}  
}
```

