COMP2200/COMP6200 Lecture 1a – Who's who

Greg Baker

23rd February 2026

Welcome to COMP2200/COMP6200

Introduction to Data Science

Today's lecture topics

- Who we all are
- Assessments
- Learning objectives
- Spare: data science jobs
- What data scientists do
- Acquiring data
- Jupyter/Colab notebooks (and loading a CSV in pandas)
- k-means clustering

Meet the Teaching Team

Dr Usman Naseem (Unit convenor)

- Contact for unit-related problems
- Reviews special considerations (e.g. illness, misadventures)

Greg Baker (Lecturer)

- Ask me about lecture content
- Worked at places like Google, Atlassian, Daisee
- Research areas: foundations of machine learning, digital humanities, low resource languages

Benjamin Pope

- Oxford PhD physicist and astronomer
- NASA Sagan Postdoctoral Fellow
- Scientific consultant for *War of the Worlds*
- `https://www.linkedin.com/in/benjamin-pope-1656811a2/`
- `https://benjaminpope.github.io/`

Other people you will want to know

Siwei Luo (Super Tutor)

- Answers questions on iLearn

Your prac demonstrator

- Will run the 2-hour practical
- First person to speak to if you have questions
- Put their name in your phone or computer

Assessments

Assessment	Weight	Due Date
Machine Learning Project	30%	Friday 3rd April 2026
Python Project	30%	Friday 22nd May 2026
Exam	40%	Exam Period

- Everything is submitted and marked on iLearn.
- **To pass: 50% overall.**
- Late penalty: 5% per day.
- iLearn will assign a time-of-day deadline for each assessment; late penalties start at that time.
- Release dates: Machine Learning Project released end of Week 3 (Friday 13th March 2026); Python Project released end of Week 7 (Friday 24th April 2026). You only need to finish by the due date (Week 6 and Week 11).

To do well in this unit

- Listen to the lecture (in person or online)
- Do the practice assessment during your prac session
- The assessments are the same as the weekly (optional) quizzes, and are broken up into weeks as well.
- Ask your prac demonstrator for help on questions that you got wrong
- If you need additional questions there are other banks of questions
- Lock in a time each week after your prac to do the actual assessment

Learning objectives for the unit

ULO1 Identify the appropriate Data Science analysis for a problem and apply that method to the problem.

ULO2 Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.

ULO3 Present the results of their Data Science analyses both verbally and in written form.

ULO4 Discuss the broader implications of Data Science analyses.

Learning objectives (less formally)

ULO1 Supervised / unsupervised / network analysis / recommendation engines. Code things with the `scikit-learn` library and the Orange tool

ULO2 Read what other data scientists have done, and win arguments against them

ULO3 Not everyone can read code, so we have to use human languages

ULO4 Think about the bigger picture of Data Science—how it shapes decisions, changes industries, and (sometimes) can be used for evil

One more thing

Greg thrives on chaos and disorder.
He's at his best when you interrupt and ask questions.