COMP8240

Applications of Data Science

Session 2, Special circumstances 2021

School of Computing

Contents

General Information 2

Learning Outcomes 3

General Assessment Information 3

Assessment Tasks 3

Delivery and Resources 7

Unit Schedule 8

Policies and Procedures 9

Assessment Standards 10

Disclaimer

Macquarie University has taken all reasonable measures to ensure the information in this publication is accurate and up-to-date. However, the information may change or become out-dated as a result of change in University policies, procedures or rules. The University reserves the right to make changes to any information in this publication without notice. Users of this publication are advised to check the website version of this publication [or the relevant faculty or department] before acting on any information in this publication.

Session 2 Learning and Teaching Update

The decision has been made to conduct study online for the remainder of Session 2 for all units WITHOUT mandatory on-campus learning activities. Exams for Session 2 will also be online where possible to do so.

This is due to the extension of the lockdown orders and to provide certainty around arrangements for the remainder of Session 2. We hope to return to campus beyond Session 2 as soon as it is safe and appropriate to do so.

Some classes/teaching activities cannot be moved online and must be taught on campus. You should already know if you are in one of these classes/teaching activities and your unit convenor will provide you with more information via iLearn. If you want to confirm, see the list of units with mandatory on-campus classes/teaching activities.

Visit the MQ COVID-19 information page for more detail.

General Information

Unit convenor and teaching staff
Lecturer, Convenor
A/Prof Mark Dras
mark.dras@mq.edu.au
4 Research Park Drive, level 2, office 208
By appointment.

Lecturer
Dr Rolf Schwitter
rolf.schwitter@mq.edu.au
4 Research Park Drive, level 3, office 359
By appointment.

Tutor
Asim Adnan Eijaz
asimadnan.eijaz@mq.edu.au
By appointment

Credit points
10

Prerequisites
COMP6200 or COMP6210 or ITEC657

Corequisites

Co-badged status
COMP7860

Unit description
This unit deals with the application of Data Science techniques to the analysis of data in a research context. Topics covered include the management of data and keeping track of intermediate results, small and large scale data processing techniques, scripting experiments, version control for source code and data, the problem of replication of research results, data publication and presentation of results in various forms. Students will complete a significant data analysis project that will use available data sets to address a research question and present results to a well defined target audience.

Important Academic Dates

Information about important academic dates including deadlines for withdrawing from units are

available at https://www.mq.edu.au/study/calendar-of-dates

Learning Outcomes

On successful completion of this unit, you will be able to:

ULO1: Define and manage a project involving empirical research.

ULO2: Apply a knowledge of programming and/or use of appropriate applications (for e.g. data gathering, curation, cleaning or analysis) in the context of practical work relevant to an empirical research project.

ULO3: Articulate clearly a coherent argument in written and oral form to a variety of audiences.

ULO4: Apply a knowledge of the principles of ethical conduct of research, including an examination of the role of open access to data and publications.

ULO5: Demonstrate best practice in document preparation and management in research.

General Assessment Information

Submission of assignments will be for the most part via iLearn; presentations associated with assessments will be given and assessed during class time.

Late Submission

No extensions will be granted without an approved application for Special Consideration. There will be a deduction of 20% of the total available marks made from the total awarded mark for each 24 hour period or part thereof that the submission is late. For example, 25 hours late in submission for an assignment worth 10 marks – 40% penalty or 4 marks deducted from the total. No submission will be accepted after solutions have been posted.

Assessment Tasks

Name	Weighting	Hurdle	Due
Practical Assignments	30%	No	throughout the semester
Project Proposal Presentation	5%	No	week 6
Project Update Presentation	5%	No	week 10
Final Presentation	5%	No	week 13
Final Report	35%	No	week 13
Exam	20%	No	exam period

Practical Assignments

Assessment Type [1]: Practice-based task
Indicative Time on Task [2]: 30 hours
Due: **throughout the semester**
Weighting: **30%**

There will be some small practical assignments during the semester, linked to the lecture material and weekly practical exercises.

On successful completion you will be able to:

- Apply a knowledge of programming and/or use of appropriate applications (for e.g. data gathering, curation, cleaning or analysis) in the context of practical work relevant to an empirical research project.
- Demonstrate best practice in document preparation and management in research.

Project Proposal Presentation

Assessment Type [1]: Presentation
Indicative Time on Task [2]: 10 hours
Due: **week 6**
Weighting: **5%**

Part of the assessment for the unit will be built around a single project you will devise, as part of a small group. This initial presentation is to pitch the idea to the audience (lecturers and students): explain the data you'll be using, give any relevant background, and outline a plan for tackling the project. The workload for the task includes the time spent on the project needed for the presentation, as well as the presentation itself.

On successful completion you will be able to:

- Define and manage a project involving empirical research.
- Articulate clearly a coherent argument in written and oral form to a variety of audiences.

Project Update Presentation

Assessment Type [1]: Presentation
Indicative Time on Task [2]: 10 hours
Due: **week 10**
Weighting: **5%**

This presentation will give an update on the state of the project. The workload for the task includes the time spent on the project needed for the presentation, as well as the presentation itself.

On successful completion you will be able to:

- Define and manage a project involving empirical research.
- Articulate clearly a coherent argument in written and oral form to a variety of audiences.

Final Presentation

Assessment Type [1]: Presentation
Indicative Time on Task [2]: 10 hours
Due: **week 13**
Weighting: **5%**

This presentation will describe to an audience the results of your project. Feedback from the presentation can be incorporated into the final report. The workload for the task includes the time spent on the project needed for the presentation, as well as the presentation itself.

On successful completion you will be able to:

- Define and manage a project involving empirical research.
- Apply a knowledge of programming and/or use of appropriate applications (for e.g. data gathering, curation, cleaning or analysis) in the context of practical work relevant to an empirical research project.
- Articulate clearly a coherent argument in written and oral form to a variety of audiences.

Final Report

Assessment Type [1]: Report
Indicative Time on Task [2]: 20 hours
Due: **week 13**
Weighting: **35%**

This report will describe the completed project as a whole: what the goals were, what data was used, how it was processed, and what the results were relative to the goals. It may also include any related programs written as part of the project, etc.

On successful completion you will be able to:

- Define and manage a project involving empirical research.
- Apply a knowledge of programming and/or use of appropriate applications (for e.g. data gathering, curation, cleaning or analysis) in the context of practical work relevant to an empirical research project.
- Articulate clearly a coherent argument in written and oral form to a variety of audiences.
- Apply a knowledge of the principles of ethical conduct of research, including an examination of the role of open access to data and publications.
- Demonstrate best practice in document preparation and management in research.

Exam

Assessment Type [1]: Examination
Indicative Time on Task [2]: 13 hours
Due: **exam period**
Weighting: **20%**

Final examination.

On successful completion you will be able to:

- Apply a knowledge of programming and/or use of appropriate applications (for e.g. data gathering, curation, cleaning or analysis) in the context of practical work relevant to an empirical research project.
- Articulate clearly a coherent argument in written and oral form to a variety of audiences.
- Apply a knowledge of the principles of ethical conduct of research, including an examination of the role of open access to data and publications.

---

[1] If you need help with your assignment, please contact:

- the academic teaching staff in your unit for guidance in understanding or completing this type of assessment
- the Writing Centre for academic skills support.

[2] Indicative time-on-task is an estimate of the time required for completion of the assessment task and is subject to individual variation

Delivery and Resources

CLASSES

Each week consists of a formally designated two hours of lecture and one hour practical session, although the lecture session may involve some practical aspects as well. For details of days, times and rooms, consult the University timetables webpage (http://www.timetables.mq.edu.au). It will be co-taught with COMP7860.

REQUIRED AND RECOMMENDED TEXTS AND/OR MATERIALS

There is no set text for the unit. We will be providing pointers to reading material over the course of the unit.

The unit has some parallels with the freely available Software Carpentry course. We'll be using those resources as supplementary ones for the unit.

UNIT WEBPAGE AND TECHNOLOGY USED AND REQUIRED

Web Home Page

The unit will make extensive use of the iLearn course management system, including for delivery of class materials, discussion boards, submission of work and access to marks and feedback. Students should check the iLearn site (https://ilearn.mq.edu.au) regularly for unit updates.

Questions and general queries regarding the content of this unit, its lectures or mixed classes, or its assignments should be posted to the discussion boards on the iLearn site. In particular, any questions which are of interest to all students in this unit should be posted to one of these discussion boards, so that everyone can benefit from the answers. Questions of a private nature should be directed to the unit teaching staff.

Technology Used and Required

The practical work in this unit involves programming in the Python language (http://www.python.org/) which is widely used for the sorts of scripting purposes covered in this unit. Python can be downloaded free of charge for a range of operating systems from the Python website.

Note that as this is a master's unit, there will be some self-directed learning. We do not expect that you will know Python before the unit starts, but will pick up the necessary elements in the first few weeks of the unit; we will give pointers to resources for learning Python, and will include snippets of Python in lecture notes where relevant to computational experiments. We will generally (but not always) use Jupyter Notebooks for Python examples, and will use Google Colab as the environment for running them. (Google Colab is a free environment that can used for some sorts of data analysis relevant to practical assignments and the major project.)

The unit will also use various other tools for e.g. data gathering and annotation. Much of the work that involves cloud computing will be carried out in the Amazon AWS framework; students can get access to a free, twelve-month educational licence for its use.

Unit Schedule

The focus of this unit is understanding the notions of open science and reproducible research. Much work in both academia and industry is driven by the free availability of papers, code and data that allow the replication and extension of existing work. In this unit, your major project will involve getting access to some of these resources, reproducing some existing work with the original data, and then investigating whether e.g. the replication works with new data. To engage fully with these freely available resources, competence with a range of techniques and tools is necessary. This project will be carried out in small groups.

Below is a tentative schedule. The weekly topics are intended to cover useful techniques and tools for carrying out your data-oriented project, and may change depending upon chosen student projects, etc.

Week 1	Philosophy of (computer) science Tools for empirical research: Jupyter Notebooks
Week 2	Introduction to cloud computing and virtual machines Discussion of data-based projects
Week 3	Version control and the linux shell Discussion of data-based projects
Week 4	Introduction to data gathering and curation
Week 5	Latex and document typesetting
Week 6	Project proposal presentations
Week 7	Data analysis tools and Python
	RECESS
Week 8	Handling messy data Data management
Week 9	Data annotation
Week 10	Project update presentations
Week 11	Databases
Week 12	Additional topics
Week 13	Final presentations

Policies and Procedures

Macquarie University policies and procedures are accessible from Policy Central (https://policies.mq.edu.au). Students should be aware of the following policies in particular with regard to Learning and Teaching:

- Academic Appeals Policy
- Academic Integrity Policy
- Academic Progression Policy
- Assessment Policy
- Fitness to Practice Procedure
- Grade Appeal Policy
- Complaint Management Procedure for Students and Members of the Public
- Special Consideration Policy

Students seeking more policy resources can visit Student Policies (https://students.mq.edu.au/support/study/policies). It is your one-stop-shop for the key policies you need to know about throughout your undergraduate student journey.

To find other policies relating to Teaching and Learning, visit Policy Central (https://policies.mq.edu.au) and use the search tool.

Student Code of Conduct

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/admin/other-resources/student-conduct

Results

Results published on platform other than eStudent, (eg. iLearn, Coursera etc.) or released directly by your Unit Convenor, are not confirmed as they are subject to final approval by the University. Once approved, final results will be sent to your student email address and will be made available in eStudent. For more information visit ask.mq.edu.au or if you are a Global MBA student contact globalmba.support@mq.edu.au

Student Support

Macquarie University provides a range of support services for students. For details, visit http://students.mq.edu.au/support/

Learning Skills

Learning Skills (mq.edu.au/learningskills) provides academic writing resources and study strategies to help you improve your marks and take control of your study.

- Getting help with your assignment
- Workshops

- StudyWise
- Academic Integrity Module

The Library provides online and face to face support to help you find and use relevant information resources.

- Subject and Research Guides
- Ask a Librarian

Student Services and Support

Students with a disability are encouraged to contact the Disability Service who can provide appropriate help with any issues that arise during their studies.

Student Enquiries

For all student enquiries, visit Student Connect at ask.mq.edu.au

If you are a Global MBA student contact globalmba.support@mq.edu.au

IT Help

For help with University computer systems and technology, visit http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/.

When using the University's IT, you must adhere to the Acceptable Use of IT Resources Policy. The policy applies to all who connect to the MQ network including students.

Assessment Standards

The unit will be graded according to the following general descriptions of the letter grades as specified by Macquarie University. In the course of the unit, these grade descriptions will be discussed with respect to example projects.

High Distinction (HD, 85-100): provides consistent evidence of deep and critical understanding in relation to the learning outcomes. There is substantial originality and insight in identifying, generating and communicating competing arguments, perspectives or problem solving approaches; critical evaluation of problems, their solutions and their implications; creativity in application as appropriate to the discipline.

In the context of this unit, the project has a good design, and has used some data that is interesting or non-obvious, or has required some effort to obtain or use. It involves a good analysis of the data, and fairly extensively draws on the techniques and tools presented in the unit and possibly on others discovered independently by the student. The project is described in a report and a presentation that are well-structured and essentially free from errors; these would be of a standard that could be presented at a conference with little or no polishing.

Distinction (D, 75-84): provides evidence of integration and evaluation of critical ideas, principles and theories, distinctive insight and ability in applying relevant skills and concepts in

relation to learning outcomes. There is demonstration of frequent originality in defining and analysing issues or problems and providing solutions; and the use of means of communication appropriate to the discipline and the audience.

In the context of this unit, the project has a good design, and has used some data that is interesting or non-obvious, or has required some effort to obtain or use. It involves a good analysis of the data, and fairly extensively draws on the techniques and tools presented in the unit. The project is described in a report and a presentation that are well-structured and mostly free from errors; these would be of a standard that could be presented at a conference with some polishing.

Credit (Cr, 65-74): provides evidence of learning that goes beyond replication of content knowledge or skills relevant to the learning outcomes. There is demonstration of substantial understanding of fundamental concepts in the field of study and the ability to apply these concepts in a variety of contexts; convincing argumentation with appropriate coherent justification; communication of ideas fluently and clearly in terms of the conventions of the discipline.

In the context of this unit, the project has a sound design, and demonstrates some thought in the choice of data. It involves a good analysis of the data, and uses a reasonable number of the techniques and tools presented in the unit. The project is described in a report and a presentation that are well-structured and mostly free from errors.

Pass (P, 50-64): provides sufficient evidence of the achievement of learning outcomes. There is demonstration of understanding and application of fundamental concepts of the field of study; routine argumentation with acceptable justification; communication of information and ideas adequately in terms of the conventions of the discipline. The learning attainment is considered satisfactory or adequate or competent or capable in relation to the specified outcomes.

In the context of this unit, the project has a satisfactory design and uses some easily accessible data. It involves a successful, or nearly successful, analysis of data, and shows some familiarity with tools or techniques presented in the unit. The project is described in a satisfactory report and presentation.

Fail (F, 0-49): does not provide evidence of attainment of learning outcomes. There is missing or partial or superficial or faulty understanding and application of the fundamental concepts in the field of study; missing, undeveloped, inappropriate or confusing argumentation; incomplete, confusing or lacking communication of ideas in ways that give little attention to the conventions of the discipline.