COMP6200

Data Science

Session 1, Special circumstances 2021

School of Computing

Contents

Notice

As part of Phase 3 of our return to campus plan, most units will now run tutorials, seminars and other small group activities on campus, and most will keep an online version available to those students unable to return or those who choose to continue their studies online.

To check the availability of face-to-face activities for your unit, please go to timetable viewer. To check detailed information on unit assessments visit your unit's iLearn space or consult your unit convenor.

General Information

Unit convenor and teaching staff
Convenor/Lecturer
Jia Wu
jia.wu@mq.edu.au
Contact via Email
283, 4 Research Park Drive
By Appointment

Lecturer
Amin Beheshti
amin.beheshti@mq.edu.au
Contact via Email
365, 4 Research Park Drive
By Appointment

Tutor
Sonit Singh
sonit.singh@mq.edu.au

Credit points
10

Prerequisites

Corequisites

Co-badged status
COMP2200

Unit description
This unit has an online offering for S2 which is **synchronous**, *meaning there will be set times to attend online lectures and tutorials.*

This unit introduces students to the fundamental techniques and tools of data science, such as the graphical display of data, predictive models, evaluation methodologies, regression, classification and clustering. The unit provides practical experience applying these methods using industry-standard software tools to real-world data sets. Students who have completed this unit will be able to identify which data science methods are most appropriate for a real-world data set, apply these methods to the data set, and interpret the results of the analysis they have performed.

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: 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.

General Assessment Information

Weely Submissions [Hurdle]

This is a hurdle requirement (see assessment policy for more information on hurdle assessment tasks). This means that you must complete at least 8 out of the 12 weeks of workshop and make a serious attempt at the set task each week, otherwise, you will fail to pass this unit immediately without considering other assessment components.

Late Submission

No extensions will be granted without an approved application for Special Consideration. There will be a deduction of 10% 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 – 20% penalty or 2 marks deducted from the total. No submission will be accepted after solutions have been posted.

Supplementary Exam

If you receive special consideration for the final exam, a supplementary exam will be scheduled after the normal exam period, following the release of marks. By making a special consideration application for the final exam you are declaring yourself available for a resit during the supplementary examination period and will not be eligible for a second special consideration approval based on pre-existing commitments. Please ensure you are familiar with the policy prior to submitting an application. Approved applicants will receive an individual notification one week prior to the exam with the exact date and time of their supplementary examination.

Assessment Tasks

Name	Weighting	Hurdle	Due
Data Science Portfolio	20%	No	Weeks 4, 6 & 8 for feedback. Week 10 final.
Final Exam	40%	No	Final Exam Period
Data Science Project	30%	No	Week 6, Week 12
Weekly Submissions	10%	Yes	Weekly

Data Science Portfolio

Assessment Type [1]: Portfolio
Indicative Time on Task [2]: 30 hours
Due: **Weeks 4, 6 & 8 for feedback. Week 10 final.**
Weighting: **20%**

The portfolio assessment will consist of three small data analysis problems that you will be given through the semester. These will involve writing code to analyse one or more data sets. You will show the versions in the workshops for feedback and then submit a final version towards the end of semester.

On successful completion you will be able to:

- Identify the appropriate Data Science analysis for a problem and apply that method to the problem.
- Interpret Data Science analyses and summarise and identify the most important aspects of a Data Science analysis.
- Present the results of their Data Science analyses both verbally and in written form.

Final Exam

Assessment Type [1]: Examination
Indicative Time on Task [2]: 10 hours
Due: **Final Exam Period**
Weighting: **40%**

The exam will assess your knowledge and understanding of the data analysis and machine learning methods covered in the semester.

On successful completion you will be able to:

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

- Discuss the broader implications of Data Science analyses.

Data Science Project

Assessment Type [1]: Report
Indicative Time on Task [2]: 40 hours
Due: **Week 6, Week 12**
Weighting: **30%**

In groups of 3-4, students will be given or will find one or more datasets and are asked to develop an analysis of this data and present a report. This project should include using more than one dataset, cleaning and analysing the data, training at least two different predictive models and using the model to make some conclusions. The report should be reproducible, all methods not only documented but available as an executable archive along with the data.

On successful completion you will be able to:

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

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

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

- Discuss the broader implications of Data Science analyses.

Weekly Submissions

Assessment Type [1]: Participatory task
Indicative Time on Task [2]: 0 hours
Due: **Weekly**
Weighting: **10%**
This is a hurdle assessment task (see assessment policy for more information on hurdle assessment tasks)

A submission of a small task based on the workshop each week. This may be a short quiz or the result of a practical task.

On successful completion you will be able to:

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

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

[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

There will be one two hour online lecture each week, and one two hour workshop in the computing laboratory or online. The online lecture would be in the form of live streaming or pre-recorded lecture videos. You are expected to attend both classes as they provide complimentary learning activities each week. In practical classes you will write code and experiment with various data sets; in lectures we will discuss the methods you are learning and how the results of your analysis can be interpreted.

Textbooks

We will refer to the following texts during the semester:

Introduction to Data Science A Python Approach to Concepts, Techniques and Applications **Igual**, Laura, **Seguí**, Santi (electronic edition available via MQ Library)

Computational and Inferential Thinking: The Foundations of Data Science By Ani Adhikari and John DeNero (available on GitBooks)

You will be given readings from these and other sources each week.

Technology Used and Required

We will make use of Python 3 for data analysis, including a range of modules such as *scikit-learn, pandas, numpy* that provide additional features. These can all be installed via the Anaconda Python distribution. We will discuss this environment and the installation process in the first week of classes.

We will use Jupyter Notebook as a way of developing and presenting the analysis results. This

is included in the full Anaconda distribution.

Project Work

A major part of the assessment in this unit is based on a project that you will complete in groups. This will allow you to explore the techniques you are learning in class in a real-world data analysis exercise.

Unit Schedule

Unit Schedule

The indicative list of topics is shown here, this is subject to change based on feedback from the class.

1	Introduction to Data Science	AB
2	Data Structures	AB
3	From Privacy to Causality and correlation	AB
4	Visualisation	AB
5	Predictive Modelling - Regression	AB
6	Software Engineering for Data Science	AB
7	Supervised Learning	JW
8	Unsupervised Learning	JW
9	Naive Bayes	JW
10	Artificial Neural Networks	JW
11	Decision Tree Models	JW
12	Advanced Topics / Guest Lecture	Guest
13	Summary	All

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.

Changes from Previous Offering

This year the S1 semester break starts from Week 6 instead of Week 7. Thus, we make a minor change to the due dates of the project. The new deadline of the project proposal is Week 6 and the final project report deadline is Week 12.