COMP6200

Data Science

Session 1, Online-scheduled-weekday 2022

School of Computing

Contents

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.

General Information

Unit convenor and teaching staff
Convener and Lecturer
Yipeng Zhou
yipeng.zhou@mq.edu.au

Lecturer
Zhu Sun
z.sun@mq.edu.au

Tutor
Subhash Sagar
subhash.sagar@mq.edu.au

Tutor
David Warren
david.warren@mq.edu.au

Tutor
Yao Deng
yao.deng@mq.edu.au

Tutor
Julius Lu
jianchao.lu@mq.edu.au

Tutor
Jiwei Guan
jiwei.guan@mq.edu.au

Credit points
10

Prerequisites

Corequisites

Co-badged status
COMP2200

> Unit description
> 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

Late submissions **will be accepted but will incur a penalty** unless there is an approved Special Consideration request. A 12-hour grace period will be given after which the following deductions will be applied to the awarded assessment mark: 12 to 24 hours late = 10% deduction; for each day thereafter, an additional 10% per day or part thereof will be applied until five days beyond the due date. After this time, a mark of zero (0) will be given. For example, an assessment worth 20% is due 5 pm on 1 January. Student A submits the assessment at 1 pm, 3 January. The assessment received a mark of 15/20. A 20% deduction is then applied to the mark of 15, resulting in the loss of three (3) marks. Student A is then awarded a final mark of 12/20.

Assessment Tasks

Name	Weighting	Hurdle	Due
Weekly Submissions	10%	Yes	every week

Name	Weighting	Hurdle	Due
Critical Analysis Task	15%	No	week12
Data Science Portfolio	35%	No	week 4, 6, 8, 10
Examinations	40%	No	week7 and exam week

Weekly Submissions

Assessment Type [1]: Participatory task
Indicative Time on Task [2]: 0 hours
Due: **every week**
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.

Critical Analysis Task

Assessment Type [1]: Report
Indicative Time on Task [2]: 15 hours
Due: **week12**
Weighting: **15%**

You will be given a sample notebook describing the analysis of a dataset. You will provide a critical analysis of this notebook and suggest improvements in the way that data is analysed and results are presented.

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.

Data Science Portfolio

Assessment Type [1]: Portfolio
Indicative Time on Task [2]: 45 hours
Due: **week 4, 6, 8, 10**
Weighting: **35%**

The portfolio assessment will consist of a number of data analysis problems that you will be given through the semester. These will involve writing code to analyse one or more data sets. These will be marked individually through the semester and then as an overall portfolio at 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.

Examinations

Assessment Type [1]: Examination
Indicative Time on Task [2]: 10 hours
Due: **week7 and exam week**
Weighting: **40%**

Examinations 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.

---

[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.

Unit Schedule

Unit Schedule

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

1	Overview of DS, Learning Python, Notebooks	SS
2	Data formats, Python input and output	SS
3	Descriptive Statistics, simple visualisation	SS
4	Causality and correlation; Visualisation	SS
5	Predictive Modelling: Linear and Logistic Regression	SS
6	Software Engineering for Data Science	SS
7	Feature Engineering; Unsupervised Learning	SS/XZ
8	K-Nearest Neighbours Classifiers	XZ
9	Naive Bayes Classifiers	XZ
10	Artificial Neural Networks	XZ
11	Decision Tree Models	XZ
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
- Assessment Procedure
- Complaints Resolution 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

Academic Integrity

At Macquarie, we believe academic integrity – honesty, respect, trust, responsibility, fairness and courage – is at the core of learning, teaching and research. We recognise that meeting the expectations required to complete your assessments can be challenging. So, we offer you a range of resources and services to help you reach your potential, including free online writing and maths support, academic skills development and wellbeing consultations.

Student Support

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

The Writing Centre

The Writing Centre provides resources to develop your English language proficiency, academic writing, and communication skills.

- Workshops
- Chat with a WriteWISE peer writing leader
- Access StudyWISE
- Upload an assignment to Studiosity
- Complete the 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

Macquarie University offers a range of Student Support Services including:

- IT Support
- Accessibility and disability support with study
- Mental health support
- Safety support to respond to bullying, harassment, sexual harassment and sexual assault
- Social support including information about finances, tenancy and legal issues

Student Enquiries

Got a question? Ask us via AskMQ, or contact Service Connect.

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

Change of Assessment Tasks:

 The group project has been removed due to the growing size of the unit enrollment and the difficulty managing groups in such a large class. It has been replaced by an additional task in the portfolio which will be more open ended than in previous years. We also add a Critical Analysis task which will ask you to reflect on a notebook that we have written to help gain some insight into what makes a good analysis.