COMP6200
Data Science
Session 1, In person-scheduled-weekday, North Ryde 2024
School of Computing

Contents

General Information ............................................. 2
Learning Outcomes ........................................... 3
General Assessment Information ....................... 3
Assessment Tasks .............................................. 5
Delivery and Resources .................................... 7
Unit Schedule .................................................. 8
Policies and Procedures .................................. 9
Changes from Previous Offering ...................... 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.
# General Information

**Unit convenor and teaching staff**
Convener and Teaching Staff
Yipeng Zhou
yipeng.zhou@mq.edu.au

Lecturer and Teaching Staff
Yang Zhang
yang.zhang@mq.edu.au

Teaching Staff
Behnaz Soltani
behnaz.soltani@mq.edu.au

Teaching Staff
Bilal Khan
bilal.khan@mq.edu.au

Teaching Staff
Munazza Zaib
munazza-zaib@mq.edu.au

Teaching Staff
Venus Haghighi
venus.haghighi@mq.edu.au

Teaching Staff
Siwei Luo
siwei.luo@mq.edu.au

Teaching Staff
Yao Deng
yao.deng@mq.edu.au

Teaching Staff
Linfeng Liang
linfeng.liang@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

Requirements to Pass this Unit
To pass this unit you must:

- Achieve a total mark equal to or greater than 50%, and
- Participate in, and undertake all hurdle activities for, a minimum of 8 of the 12 weekly workshops
Hurdle Assessments

Assessment 1: Practice-based task (10%)

Development of knowledge and skills requires continual practice at authentic problems in a laboratory-based setting. This unit has weekly laboratory classes and you must demonstrate your progress in developing and communicating knowledge and skills in a minimum of 8 of the 12 workshop practicals. This is a hurdle assessment meaning that failure to meet this requirement may result in a fail grade for the unit. It is allowed to take the second chance of each workshop hurdle assessment until they complete these workshops.

Late Assessment Submission Penalty

Unless a Special Consideration request has been submitted and approved, a 5% penalty (of the total possible mark of the task) will be applied for each day a written report or presentation assessment is not submitted, up until the 7th day (including weekends). After the 7th day, a grade of ‘0’ will be awarded even if the assessment is submitted. The submission time for all uploaded assessments is 11:55 pm. A 1-hour grace period will be provided to students who experience a technical concern.

For any late submission of time-sensitive tasks, such as scheduled tests/exams, performance assessments/presentations, and/or scheduled practical assessments/labs, please apply for Special Consideration.

Special Consideration

The Special Consideration Policy aims to support students who have been impacted by short-term circumstances or events that are serious, unavoidable and significantly disruptive, and which may affect their performance in assessment.

Written Assessments: If you experience circumstances or events that affect your ability to complete the written assessments in this unit on time, please inform the convenor and submit a Special Consideration request through ask.mq.edu.au.

Weekly practice-based tasks: To pass the unit you need to demonstrate ongoing development of skills and application of knowledge in 8 out of 12 of the weekly practical classes. If you miss a weekly practical class due to a serious, unavoidable and significant disruption, contact your convenor ASAP as you may be able to attend another class that week. If it is not possible to attend another class, you should still contact your convenor for access to class material to review in your own time.

Assessments where Late Submissions will be accepted

Assessment Portfolios, Workshops, Reflective Report and Critical Assessment Task– YES,
Assessment Tasks

<table>
<thead>
<tr>
<th>Name</th>
<th>Weighting</th>
<th>Hurdle</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examinations</td>
<td>40%</td>
<td>No</td>
<td>Week 6 and Exam Week</td>
</tr>
<tr>
<td>Weekly Tasks</td>
<td>10%</td>
<td>Yes</td>
<td>Every week</td>
</tr>
<tr>
<td>Critical Analysis Task</td>
<td>15%</td>
<td>No</td>
<td>Week 12</td>
</tr>
<tr>
<td>Data Science Portfolio</td>
<td>35%</td>
<td>No</td>
<td>Week 4, 6, 8, 10, 11</td>
</tr>
</tbody>
</table>

Examinations

Assessment Type 1: Examination
Indicative Time on Task 2: 10 hours
Due: Week 6 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.

Weekly Tasks

Assessment Type 1: Practice-based 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)

Development of knowledge and skills requires continual practice at authentic problems. During weekly workshops you will practice a range of tasks recording your progress on worksheets. To pass this hurdle assessment, you must be able to demonstrate your progress in developing and
communicating knowledge and skills in a minimum of 8 of the 12 weekly workshops.

On successful completion you will be able to:

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

**Critical Analysis Task**

Assessment Type: Report
Indicative Time on Task: 15 hours
Due: Week 12
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: Portfolio
Indicative Time on Task: 45 hours
Due: Week 4, 6, 8, 10, 11
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.

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 on campus lecture each week, and one two hour workshop in the computing laboratory. The on campus lecture would be in the form of live streaming or pre-recorded lecture videos as well. 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.

Week 1 Classes

In week 1, there is a two-hour lecture. Workshops will commence in week 1 as well. Every student should attend the workshop session which they have enrolled.

Methods of Communication

We will communicate with you via your university email and through announcements on iLearn. Queries to convenors can either be placed on the iLearn discussion board or sent to the unit convenor via the contact email on iLearn.

COVID Information

For the latest information on the University’s response to COVID-19, please refer to the Coronavirus infection page on the Macquarie website: https://www.mq.edu.au/about/coronavirus-faqs. Remember to check this page regularly in case the information and requirements change during semester. If there are any changes to this unit in relation to COVID, these will be communicated via iLearn.

Textbooks

We will refer to the following texts during the semester:
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, a 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**

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

<table>
<thead>
<tr>
<th></th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Overview of DS, Learning Python, Notebooks</td>
</tr>
<tr>
<td>2</td>
<td>Data formats, Python input and output</td>
</tr>
<tr>
<td>3</td>
<td>Descriptive Statistics, simple visualisation</td>
</tr>
<tr>
<td>4</td>
<td>Causality and correlation; Visualisation, Data Ethics</td>
</tr>
<tr>
<td>5</td>
<td>Predictive Modelling: Linear and Logistic Regression</td>
</tr>
<tr>
<td>6</td>
<td>Software Engineering for Data Science and Data Privacy</td>
</tr>
<tr>
<td>7</td>
<td>Feature sets and spaces; Unsupervised learning</td>
</tr>
<tr>
<td>8</td>
<td>Supervised Learning: K-Nearest Neighbours</td>
</tr>
<tr>
<td>9</td>
<td>Naive Bayes Classifiers</td>
</tr>
<tr>
<td>10</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>11</td>
<td>Learning Decision Trees</td>
</tr>
<tr>
<td>12</td>
<td>Summary</td>
</tr>
</tbody>
</table>
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://stu
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 Advocacy** provides independent advice on MQ policies, procedures, and processes

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/](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](http://www.mq.edu.au/about_us/offices_and_units/information_technology/help/).

The policy applies to all who connect to the MQ network including students.

Changes from Previous Offering
There are two changes.
First, we will include new materials on data ethics and data privacy in week 4 and week 6. Correspondingly, assessment tasks will involve questions related to these new topics.

Second, we do not run online workshop anymore. The mid and final exams are delivered as online quizzes on campus, and changed from open book to closed book.

Unit information based on version 2024.04 of the Handbook