# COMP6200

## Data Science

Session 1, Weekday attendance, North Ryde 2020

*Dept of Computing*

## Contents

- General Information  2
- Learning Outcomes  2
- General Assessment Information  3
- Assessment Tasks  3
- Delivery and Resources  6
- Unit Schedule  6
- Policies and Procedures  7
- Changes from Previous Offering  9

---

**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, Lecturer
Steve Cassidy
steve.cassidy@mq.edu.au
Contact via Email
Level 2, 4 Research Park Drive

Lecturer
Xuyun Zhang
xuyun.zhang@mq.edu.au

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

Credit points
10

Prerequisites

Corequisites

Co-badged status

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://students.mq.edu.au/important-dates

Learning Outcomes

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 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

<table>
<thead>
<tr>
<th>Name</th>
<th>Weighting</th>
<th>Hurdle</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshop Checkpoints</td>
<td>10%</td>
<td>Yes</td>
<td>Weekly</td>
</tr>
<tr>
<td>Data Science Portfolio</td>
<td>20%</td>
<td>No</td>
<td>Weeks 4, 6 &amp; 8 for feedback. Week 10 final.</td>
</tr>
<tr>
<td>Data Science Project</td>
<td>30%</td>
<td>No</td>
<td>Week 7, Week 13</td>
</tr>
<tr>
<td>Final Exam</td>
<td>40%</td>
<td>No</td>
<td>Final Exam Period</td>
</tr>
</tbody>
</table>

Workshop Checkpoints

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)
When you attend each weekly workshop you will be asked to complete or make a serious attempt at a task, either practical or involving discussion with the class.

There will be no opportunity to repeat work that you have missed unless you have a confirmed Special Consideration request for the day of your workshop (e.g. if you were ill). This means that you must attend at least 8 out of the 12 weeks of workshop and make a serious attempt at the set task each week.

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: 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 during the semester to get feedback and then submit a final version as an assignment.

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.

Data Science Project

Assessment Type: Project
Indicative Time on Task: 30 hours
Due: Week 7, Week 13
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.

Proposal and scoping document (week 7): 5% Final report (week 13): 15% Project presentation
video (week 13): 10%

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.

Final Exam
Assessment Type 1: Examination
Indicative Time on Task 2: 20 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:
• 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.

1 If you need guidance or support to understand or complete this type of assessment, please contact the Learning Skills Team

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 lecture each week and one two hour workshop in the computing laboratory. 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 3.6 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](#) 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.

<table>
<thead>
<tr>
<th></th>
<th>Overview of DS, Learning Python, Notebooks</th>
<th>SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Data formats, Python input and output</td>
<td>SC</td>
</tr>
<tr>
<td>3</td>
<td>Descriptive Statistics, simple visualisation</td>
<td>SS</td>
</tr>
</tbody>
</table>
Policies and Procedures

Macquarie University policies and procedures are accessible from Policy Central (https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central). 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 (Note: The Special Consideration Policy is effective from 4 December 2017 and replaces the Disruption to Studies Policy.)

Students seeking more policy resources can visit the Student Policy Gateway (https://students.mq.edu.au/support/study/student-policy-gateway). It is your one-stop-shop for the key policies you need to know about throughout your undergraduate student journey.
If you would like to see all the policies relevant to Learning and Teaching visit Policy Central (https://staff.mq.edu.au/work/strategy-planning-and-governance/university-policies-and-procedures/policy-central).

**Student Code of Conduct**

Macquarie University students have a responsibility to be familiar with the Student Code of Conduct: https://students.mq.edu.au/study/getting-started/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 improve your marks and take control of your study.

- Workshops
- StudyWise
- Academic Integrity Module for Students
- Ask a Learning Adviser

**Student Enquiry Service**

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

**Equity 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.

**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

The presentation component of the group project has been moved to a video presentation to help manage the logistics of a large number of students needing to present in the final week of semester.