

WESTERN SYDNEY
UNIVERSITY



School of Computing, Engineering and Mathematics



Learning Guide

301044 Data Science
Spring 2019

Unit Details

Unit Code:	301044
Unit Name:	Data Science
Credit Points:	10
Unit Level:	7
Assumed Knowledge:	Basic Statistics, Computer Programming

Note: Students with any problems, concerns or doubts should discuss those with the Unit Coordinator as early as they can.

Unit Coordinator

Name: Liwan Liyanage
Phone: 4620 3467
Location: Room 26.1.16, Campbelltown Campus
Email: L.Liyanage@westernsydney.edu.au
Consultation Arrangement:
Thursday from 3pm-4pm in Room ER.1.07 Parramatta Campus (Hot Office) by appointment

Director of Academic Program

Name: Laurence Park
Phone: (02) 9685 9065
Location: Campus: Parramatta, Building: EN, Room: EN.1.32
Email: l.park@westernsydney.edu.au

Contents

1	About Data Science	2
1.1	An Introduction to this Unit	2
1.2	What is Expected of You	2
1.3	Changes to Unit as a Result of Past Student Feedback	2
2	Assessment Information	3
2.1	Unit Learning Outcomes	3
2.2	Approach to Learning	3
2.3	Contribution to Course Learning Outcomes	4
2.4	Assessment Summary	5
2.5	Assessment Details	6
2.5.1	Online Quizzes	6
2.5.2	Computer Based Assignment - Data Analysis Task	7
2.5.3	Lab Based Exam	8
2.6	General Submission Requirements	9
3	Teaching and Learning Activities	11
4	Learning Resources	15
4.1	Recommended Readings	15

Note: The relevant Learning Guide Companion supplements this document

1 About Data Science

1.1 An Introduction to this Unit

The explosion of data in the internet age opens up new possibilities for agencies and business to better serve and market to its customers. To take full advantage of these opportunities requires the ability to consolidate, manage and extract information from very large diverse data sets. In science, data sets are growing rapidly, with projects routinely generating terabytes of data. In this unit we examine the software tools and analytic methods that underpin a successful Data Science Project and gain experience in big data analytics.

1.2 What is Expected of You

Study Load

A student is expected to study an hour per credit point a week. For example a 10 credit point unit would require 10 hours of study per week. This time includes the time spent within classes during lectures, tutorials or practicals.

Attendance

It is strongly recommended that students attend all scheduled learning activities to support their learning.

Online Learning Requirements

Unit materials will be made available on the unit's vUWS (E-Learning) site (<https://vuws.westernsydney.edu.au/>). You are expected to consult vUWS at least twice a week, as all unit announcements will be made via vUWS. Teaching and learning materials will be regularly updated and posted online by the teaching team.

No E-Learning resources required for this Unit.

Special Requirements

Essential Equipment:

Not Applicable

Legislative Pre-Requisites:

Not Applicable

1.3 Changes to Unit as a Result of Past Student Feedback

The University values student feedback in order to improve the quality of its educational programs. The feedback provided helps us improve teaching methods and units of study. The survey results inform unit content and design, learning guides, teaching methods, assessment processes and teaching materials.

You are welcome to provide feedback that is related to the teaching of this unit. At the end of the semester you will be given the opportunity to complete a Student Feedback on Unit (SFU) questionnaire to assess the unit. You may also have the opportunity to complete a Student Feedback on Teaching (SFT) questionnaire to provide feedback for individual teaching staff.

As a result of student feedback, the following changes and improvements to this unit have recently been made:

- Motivation and the applications of data science will be included in the first lecture.
- The statistical background and theory behind R code and algorithms used in this unit will be provided.
- The Lab based exam will be held during Week 14. Students are required to show the exam answer paper to the supervisor after submitting to TurnItIn to ensure successful submission.
- Quizzes will be run during the later part of the lab classes.
- It is mandatory students upload their lab exercises at the end of each lab class via a TurnItIn link on the unit's vUWS site. Solutions will be available after submission is completed.

2 Assessment Information

2.1 Unit Learning Outcomes

	Outcome
1	Describe the issues (computational and social) in data science
2	Show when and how to apply the MapReduce paradigm to solve data analytics problems
3	Select and apply appropriate Machine learning and statistical algorithms to extract information from data
4	Evaluate and interpret the utility of information found using Data Science

2.2 Approach to Learning

This unit is one in which the later topics build on material covered earlier in the unit. As such, it is critical that students work consistently on this unit throughout the semester. Through the assessments and exams, it is intended to examine a student's understanding and ability to apply appropriate techniques.

Students who do not attend regularly and punctually for all practicals/tutorials may find that they are at risk of not passing their assessments.

2.3 Contribution to Course Learning Outcomes

3735: Master of Data Science

Course Learning Outcomes	ULO 1	ULO 2	ULO 3	ULO 4
1. Apply Data Science methods to problems in various disciplines (e.g. Business, Science, Social Science, Engineering, Education and the Humanities).	Introduced	Introduced	Developed	
2. Conduct and manage the formulation of problems and the use of data ethically and responsibly.	Introduced			
3. Design and conduct data gathering and analysis to provide information and advice that is reliable, valid, timely and relevant.	Introduced	Introduced	Developed	
4. Generate interpretive and predictive reports, working alongside professional colleagues in decision-making.			Developed	Developed
5. Provide expert advice to professional colleagues on the validity and reliability of interpretations and predictions based on analysis of large complex data sets.	Introduced		Developed	Developed

2.4 Assessment Summary

The assessment items in this unit are designed to enable you to demonstrate that you have achieved the unit learning outcomes. Completion and submission of all assessment items which have been designated as mandatory or compulsory is essential to receive a passing grade.

To pass this unit you must:

- submit all assessment items; AND
- achieve an overall mark of at least 50%.
- achieve at least 40% of the Final Lab Based Exam

Missing a mandatory assessment item will result in an FNS grade, unless an application for special consideration providing documentation of extenuating circumstances is approved in accordance with UWS policies. The special consideration form must be submitted within two (2) working days. If special consideration is granted, the weight of the missed item(s) will be added to the weight of the final exam.

Item	Weight	Due Date	ULOs Assessed	Threshold
Online Quizzes	20%	At the end of the practicals; Thursday in Weeks 4, 6, 8, 11, and 13	1, 2, 4	No
Computer Based Assignment - Data Analysis Task	40%	Week 11: Friday 4th October 11:59pm	3, 4	No
Lab Based Exam	40%	Week 14: Thursday 24th October during Lab session	3, 4	No

Feedback on Assessment

Feedback is an important part of the learning process that can improve your progress towards achieving the learning outcomes. Feedback is any written or spoken response made in relation to academic work such as an assessment task, a performance or product. It can be given to you by a teacher, an external assessor or student peer, and may be given individually or to a group of students. As a Western Sydney University student, it is your responsibility to seek out and act on feedback that is provided to you as a resource to further your learning.

Online Quizzes: In this unit you can expect feedback as soon as the quizzes are closed. You will receive all solutions to the questions.

Computer based Assignment: You can expect to receive marked assignments within two weeks after submission

2.5 Assessment Details

2.5.1 Online Quizzes

Weight:	20%
Type of Collaboration:	Individual
Due:	At the end of the practicals; Thursday in Weeks 4, 6, 8, 11, and 13
Submission:	Online in Lab class via vUWS. Enter answers into the online form. You have 20 minutes to complete the quiz once it is opened.
Format:	Each quiz consists of a series of short questions. There are 8 questions; each is worth 1 mark. Each quiz is worth 4% of the overall assessment for the unit. You will be required to use R Studio to analyse data for some questions in the quiz.
Length:	5 quizzes of 20 minutes each
Curriculum Mode:	Multiple Choice

The online quizzes will be held during the practical class in Weeks 4, 6, 8, 11 and 13 and are based on the material covered in previous week's lectures and practicals as listed in the Teaching and Learning Activity schedule given in this document. Quizzes will save and be submitted automatically when the time expires. Once started, the quiz must be completed in one sitting. Do not leave the test before clicking **Save and Submit**.

Resources:

Open Book: Students are allowed to use any material related to the subject. (Please note you have only 20 minutes to complete the quiz and hence time is limited.)

Resources include summaries and handy hints given in lectures; useful links or readings uploaded in vUWS; lecture slides and practical solutions done in class available in vUWS.

Marking Criteria:

Criteria	High Distinction	Distinction	Credit	Pass	Unsatisfactory
Online Quizzes	at least 85%	75% to 84%	65% to 74%	50% to 64%	less than 50%

2.5.2 Computer Based Assignment - Data Analysis Task

Weight:	40%
Type of Collaboration:	Both (Individual & Group)
Due:	Week 11: Friday 4th October 11:59pm
Submission:	Online via TurnItIn link in vUWS before due date
Format:	Word processed document in PDF format; mathematical equations may be hand written and scanned to be included in the document. Students will be using real data and the assignment will take the form of a mini research project.
Length:	To consist of 5 or so pages of text and computer output, equivalent to approx. 2000 words
Curriculum Mode:	Report

The assignment covers Data Science theory and applications. It involves a data analysis task that should be done using R Studio. The assignment uses real data which will be provided.

The report should consist of 5 or so pages of text and computer output, equivalent to approximately 2000 words excluding computer codes, data and R output. Students have the option of creating the document using Microsoft Word or R Markdown. Once created, save your report as a PDF file and submit it via the TurnItIn link on the unit's vUWS site.

Note: Mathematical equations may be hand written and scanned to be included in the document. The assignment will be available in Week 5 and you will receive the marked assignments within 2 weeks of submission.

Resources:

Open Book: Students are allowed to use any material related to the subject.

Resources include lecture slides and practical solutions done in class available in vUWS; summaries and handy hints given in lectures; useful links or readings uploaded in vUWS;

Marking Criteria:

Criteria	High Distinction	Distinction	Credit	Pass	Unsatisfactory
Assignment	at least 85%	75% to 84%	65% to 74%	50% to 64%	less than 50%

2.5.3 Lab Based Exam

Weight:	40%
Type of Collaboration:	Individual
Due:	Week 14: Thursday 24th October during Lab session
Submission:	In Lab Class via TurnItIn Link in vUWS
Format:	Word processed document in PDF format. Will use R Studio for analysis. Logically present answers to each question incorporating R/R Studio outputs including graphs and charts with text.
Length:	2 hours
Curriculum Mode:	Numerical Problem Solving

Students are expected to use the software R Studio and carry out necessary modelling tasks and answer the examination questions in a Word document connecting theory and applications. Students need to create the document using Microsoft Word. Once created, save your report as a PDF file and submit it via the TurnItIn link on the unit's vUWS site.

Students are required to obtain at least 40% of the Final Lab Based Exam to pass this unit.

Resources:

Open Book: Students are allowed to use any material related to the subject.

Resources include lecture slides and practical solutions done in class available in vUWS; summaries and handy hints given in lectures; useful links or readings uploaded in vUWS;

Marking Criteria:

Criteria	High Distinction	Distinction	Credit	Pass	Unsatisfactory
Lab Based Exam	at least 85%	75% to 84%	65% to 74%	50% to 64%	less than 50%

2.6 General Submission Requirements

Submission

- All assignments must be submitted by the specified due date and time.
- Complete your assignment and follow the individual assessment item instructions on how to submit. You must keep a copy of all assignments submitted for marking.

Turnitin

- The Turnitin plagiarism prevention system may be used within this unit. Turnitin is accessed via logging into vUWS for the unit. If Turnitin is being used with this unit, this means that your assignments have to be submitted through the Turnitin system. Turnitin from iParadigms is a web-based text-matching software that identifies and reports on similarities between documents. It is also widely utilised as a tool to improve academic writing skills. Turnitin compares electronically submitted papers against the following:
 - Current and archived web: Turnitin currently contains over 24 billion web pages including archived pages
 - Student papers: including Western Sydney University student submissions since 2007
 - Scholarly literature: Turnitin has partnered with leading content publishers, including library databases, text-book publishers, digital reference collections and subscription-based publications (e.g. Gale, Proquest, Emerald and Sage)
- Turnitin is used by over 30 universities in Australia and is increasingly seen as an industry standard. It is an important tool to assist students with their academic writing by promoting awareness of plagiarism. By submitting your assignment to Turnitin you will be certifying that:
 - I hold a copy of this assignment if the original is lost or damaged
 - No part of this assignment has been copied from any other student's work or from any other source except where due acknowledgement is made in the assignment
 - No part of the assignment has been written for me by any other person/s
 - I have complied with the specified word length for this assignment
 - I am aware that this work may be reproduced and submitted to plagiarism detection software programs for the purpose of detecting possible plagiarism (which may retain a copy on its database for future plagiarism checking).

Self-Plagiarising

- You are to ensure that no part of any submitted assignment for this unit or product has been submitted by yourself in another (previous or current) assessment from any unit, except where appropriately referenced, and with prior permission from the Lecturer/Tutor/Unit Co-ordinator of this unit.

Late Submission

- If you submit a late assessment, without receiving approval for an extension of time, (see next item), you will be penalised by 10% per day for up to 10 days. In other words, marks equal to 10% of the assignment's weight will be deducted from the mark awarded.
- For example, if the highest mark possible is 50, 5 marks will be deducted from your awarded mark for each late day.
- Saturday and Sunday are counted as one calendar day each.
- Assessments will not be accepted after the marked assessment task has been returned to students.
- This is consistent with Clause 51 of the Western Sydney University's Assessment Policy - Criteria and Standards-Based Assessment.

Extension of Due Date for Submission

Extensions are only granted in exceptional circumstances. To apply for an extension of time, locate an application form via the Western Sydney University homepage or copy the following link:

https://www.westernsydney.edu.au/currentstudents/current_students/forms

Application forms must be submitted to the Unit Coordinator/Convenor. Requests for extension should be made as early as possible and submitted within policy deadlines. Appropriate, supporting documentation must be submitted with the application. An application for an extension does not automatically mean that an extension will be granted. Assessments will not be accepted after the marked assessment task has been returned to students.

Resubmission Resubmission of assessment items will not normally be granted if requested.

Application for Special Consideration

It is strongly recommended that you attend all scheduled learning activities to support your learning. If you have suffered misadventure, illness, or you have experienced exceptional circumstances that have prevented your attendance at class or your completion and submission of assessment tasks, you may need to apply for Special Consideration via the Western Sydney University website. http://www.westernsydney.edu.au/currentstudents/current_students/services_and_facilities/special_consideration2 or the Student Centre/Sydney City Campus Reception. Special Consideration is not automatically granted. It is your responsibility to ensure that any missed content has been covered. Your lecturer will give you more information on how this must be done.

3 Teaching and Learning Activities

Weeks	Topic	Lecture	Prac/Lab	Assessments Due
Week 1 22-07-2019	Introduction to Data Science	Introduction to Data Science: Motivation, Applications and Overview		
Week 2 29-07-2019	Supervised Learning: Linear Models 1	Linear Regression: - Simple Linear Regression - Estimating the Coefficients - Assessing the Accuracy of the Coefficient Estimates - Assessing the Accuracy of the Model - Estimating the Regression Coefficients	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	
Week 3 05-08-2019	Supervised Learning: Linear Models 2	Multiple Regression: - Estimating the Coefficients - Assessing the Accuracy of the Coefficient Estimates - Assessing the Accuracy of the Model - Estimating the Regression Coefficients - Polynomial Regression - Qualitative Predictors	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	
Week 4 12-08-2019	Supervised Learning: Classification	Classification: - An Overview of Classification - Why Not Linear Regression? - Logistic Regression - The Logistic Model	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	- Online Quizzes

Weeks	Topic	Lecture	Prac/Lab	Assessments Due
		<ul style="list-style-type: none"> - Estimating the Regression Coefficients - Making Predictions - Multiple Logistic Regression 		
Week 5 19-08-2019	Supervised Learning: Tree Based Methods	Decision Trees: <ul style="list-style-type: none"> - The Basics of Decision Trees - Regression Trees - Classification Trees - Trees Versus Linear Models - Advantages and Disadvantages of Trees 	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	
Week 6 26-08-2019	Supervised Learning: Classification	Support Vector Machines: <ul style="list-style-type: none"> - SVM Motivation - Slack Variables - Kernel Functions 	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	- Online Quizzes
Week 7 02-09-2019	Guest Lecture	Data Science Building Blocks and Industry Experience	Data Science Building Blocks and Industry Experience	
Week 8 09-09-2019	Unsupervised Learning: Principal Component Analysis	Principal Component Analysis: <ul style="list-style-type: none"> - The Challenge of Unsupervised Learning - Principal Components Analysis - What are Principal Components? - Another Interpretation of Principal Components 	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	- Online Quizzes

Weeks	Topic	Lecture	Prac/Lab	Assessments Due
		<ul style="list-style-type: none"> - More on PCA - Other Uses for Principal Components 		
Week 9 16-09-2019	Intra Session Break	Intra Session Break	Intra Session Break	
Week 10 23-09-2019	Unsupervised Learning: Clustering	Clustering Methods: <ul style="list-style-type: none"> - K-Means Clustering - Hierarchical Clustering - Practical Issues in Clustering 	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS. Computer based assignment is based on real data and takes a format of a research project.	
Week 11 30-09-2019	Resampling Method: Resampling and Error Estimation	<ul style="list-style-type: none"> - Cross-Validation - The Validation Set Approach - Leave-One-Out Cross-Validation - k-Fold Cross-Validation - Bias-Variance Tradeoff - Cross-Validation - Cross-Validation on Classification Problems 	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	<ul style="list-style-type: none"> - Online Quizzes - Computer Based Assignment - Data Analysis Task
Week 12 07-10-2019	Guest Lecture	Visualisation and Visual Analytics Overview by example	Application to demonstrate techniques learned in lectures using R. Submit lab exercises via TurnItIn link in vUWS.	
Week 13 14-10-2019	Revision	Revision Summary of the unit content	Presentations	- Online Quizzes
Week 14 21-10-2019	Final Computer Based Exam	Final Computer Based Exam	Lab based exam	- Lab Based Exam

Weeks	Topic	Lecture	Prac/Lab	Assessments Due
Week 15 28-10-2019				
Week 16 04-11-2019				
Week 17 11-11-2019				

The above timetable should be used as a guide only, as it is subject to change. Students will be advised of any changes as they become known on the unit's vUWS site.

4 Learning Resources

4.1 Recommended Readings

Essential Reading

- Dalgaard, Peter. (2008). Introductory statistics with R (2nd ed.). New York: Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning : data mining, inference, and prediction (2nd ed.). New York: Springer.
- James, G, Witten, D, Hastie, T, & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R (Vol. 103). New York: Springer.
- Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. New York: Springer.