



COS10022 – DATA SCIENCE PRINCIPLES

Dr Pei-Wei Tsai (Lecturer, Unit Convenor)
ptsai@swin.edu.au, EN508d

SWIN
BUR
NE

SWINBURNE
UNIVERSITY OF
TECHNOLOGY



Welcome to COS10022

Data Science Principles

Asynchronized
Short Videos

Live Online
Lectures

Laboratories

Online Tests
(week 5, 9, and 12)

Assignments
(Week 4 and 10)



- **Lecturer, Tutor, and Convenor**

- **Dr Pei-Wei Tsai**

- ptsai@swin.edu.au

- **Office: EN508d**

- **Main Research Fields:**

- **Intelligent Optimisation**
 - **Data Analytics**
 - **Machine Learning**



ABOUT DATA SCIENCE

Data Scientist

#8 in 100 Best Jobs

Data scientists use technology to glean insights from large amounts of data they collect. [Read More »](#)

Projected Jobs
59,400

Median Salary
\$103,500

Education Needed
Bachelor's

2024 Data Scientist Salary Guide - Australia

Published on January 19

Data Scientist Salary Estimate: Overview

The Annual Data Scientist Salary Report for Australia

Data scientists in Australia, you're in for some exciting data insights! In 2024, the [average salary for a data scientist is AU\\$93,760](#). This figure is more than just a number; it reflects the growing importance of data science in the business world.

Globally, data scientist salaries differ significantly, [as per various data sources](#). In the US, the average is \$156,717, contrasted by Australia's AU\$91,703. Canada offers C\$80,364, while India sees about INR 874,113. These disparities highlight the diverse valuation of data science skills

<https://datablokes.com.au/blog/data-science-salaries-in-australia>

Indeed's Best Jobs of 2024

Rank	Job Title	Average Annual Salary (\$75K minimum)	Jobs per 1M Total Jobs	% Change in Job Share, 2021 vs. 2024	% Containing Remote & Hybrid Phrases
1	Mental Health Technician	\$77,448	1,425	1%	18%
2	Loan Officer	\$192,339	1,044	3%	75%
3	Mental Health Therapist	\$76,140	865	132%	41%
4	Electrical Engineer	\$102,590	700	34%	19%
5	Construction Project Manager	\$103,431	662	37%	10%
6	Mechanical Engineer	\$96,091	552	24%	16%
7	Psychiatrist	\$258,440	552	36%	15%
8	Human Resources Manager	\$79,174	549	5%	13%
9	Senior Accountant	\$82,811	547	18%	24%
10	Data Engineer	\$130,135	532	2%	41%
11	Civil Engineer	\$93,967	524	44%	22%
12	Supply Chain Specialist	\$86,976	509	32%	13%

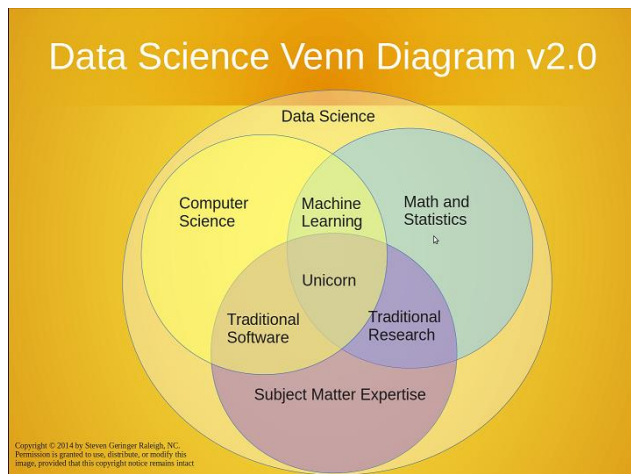
Source: Indeed's Best Jobs of 2024

About Data Scientist

- Data Scientist ranked in the top 8 in 100 best jobs.
- U.S.News
- Data Engineer ranked in the top 10 in Indeed's best job of 2024.
- Indeed's 2024 Workforce Insight Report
- Data Scientist's average salary in the US is AUD156,717 per year.
- Datablokes.

	Data Analyst	Machine Learning Engineer	Data Engineer	Data Scientist
Programming Tools	Very important	Very important	Very important	Very important
Data Visualization and Communication	Very important	Somewhat important	Somewhat important	Very important
Data Intuition	Somewhat important	Very important	Somewhat important	Very important
Statistics	Somewhat important	Very important	Somewhat important	Very important
Data Wrangling	Not that important	Not that important	Very important	Very important
Machine Learning	Not that important	Very important	Not that important	Very important
Software Engineering	Not that important	Somewhat important	Very important	Somewhat important
Multivariable Calculus and Linear Algebra	Not that important	Very important	Not that important	Somewhat important

● Not that important
 ● Somewhat important
 ● Very important



Data Scientist

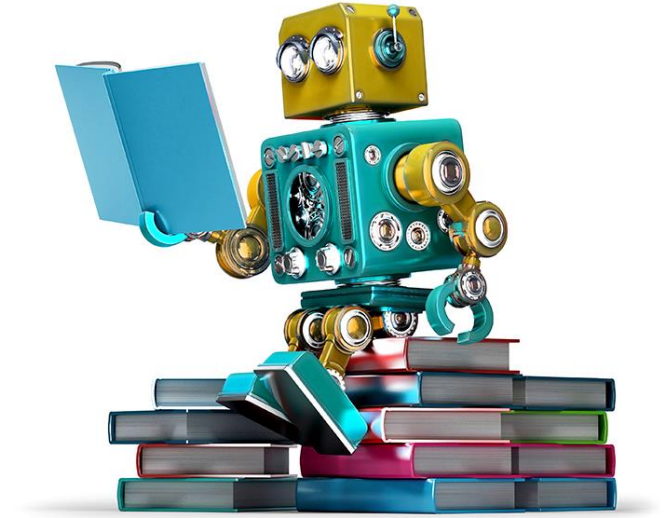
- Capable of analyzing and interpreting complex digital data to assist a business in its decision-making.



ABOUT THIS UNIT (COS10022)

Learning Outcomes

- After successfully completing this unit, you should be able to:
 1. Appreciate the roles of data science and Big Data analytics in organisational contexts.
 2. Compare and analyse the key concepts, techniques and tools for discovering, analysing, visualising and presenting data.
 3. Describe the processes within the Data Analytics Lifecycle.
 4. Analyse organisational problems and formulate them into data science tasks.
 5. Evaluate suitable techniques and tools for specific data science tasks.
 6. Develop and execute an analytics plan for a given case study.



colab



Lab Environment

Hardware

- Desktops
- You can also bring your own laptop

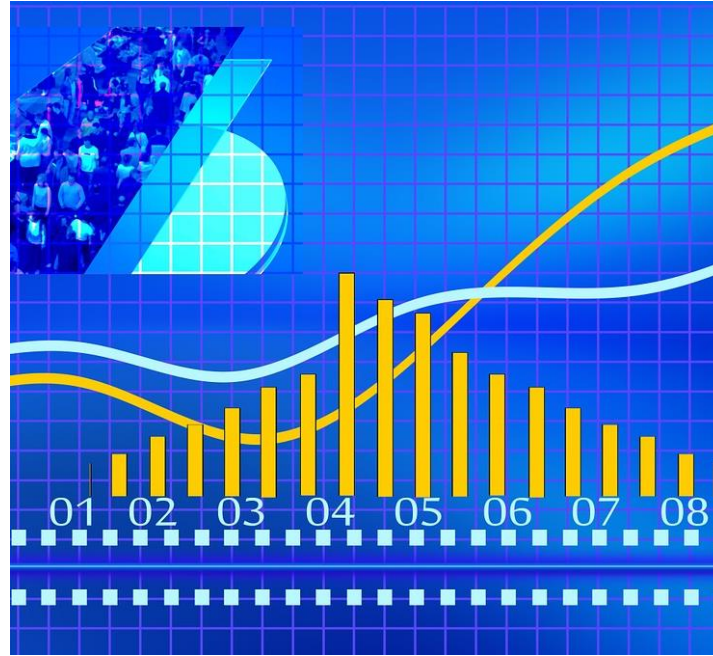
Software

- Google Colab
- KNIME (analytics libraries and tools)

Classroom Etiquette

- Usage of Personal Electronic Devices
 - Cell phones/PDAs (Set to Vibration mode or mute)
 - If your phone rings, answer it as you step out of the classroom.
- Bring your own rubbish with you when leaving the room.
- Inform the instructor and the lab partners (if any) of all absences from classroom sessions.
 - Excessive absences will be interpreted as non-attendance at the class.
- Although we encourage collaboration during the class, please treat the data files, code and lab as intellectual property of this unit and do not redistribute without the consent of the unit convenor.





```
78 // ...
79 $SESSION['_CAPTCHA']['config'] = serialize($captcha_config);
80
81 return array(
82     'code' => $captcha_config['code'],
83     'image_src' => $image_src
84 );
85
86 }
87
88 // ...
89 if ( !function_exists('hex2rgb') ) {
90     function hex2rgb($hex_str, $return_string = false, $separator = ',') {
91         $hex_str = preg_replace("/^#-9A-Fa-f$/", '', $hex_str); // Ge
92         $rgb_array = array();
93         if ( strlen($hex_str) == 6 ) {
94             $color_val = hexdec($hex_str);
95             $rgb_array['r'] = 0xFF & ($color_val >> 0x10);
96             $rgb_array['g'] = 0xFF & ($color_val >> 0x08);
97             $rgb_array['b'] = 0xFF & ($color_val >> 0x00);
98         } elseif ( strlen($hex_str) == 3 ) {
99             $rgb_array['r'] = hexdec(str_repeat(substr($hex_str, 0, 1), 2));
100             $rgb_array['g'] = hexdec(str_repeat(substr($hex_str, 1, 1), 2));
101             $rgb_array['b'] = hexdec(str_repeat(substr($hex_str, 2, 1), 2));
102         } else {
103             return false;
104         }
105         return $return_string ? implode($separator, $rgb_array) : $rgb_array;
106     }
107 }
108 // Draw the image
109 if ( !isset($_GET['code']) ) {
110     // ...
111 }
```

Expected Background

- Strong mathematical and quantitative capability.
- Basic programming skills.

Assessment and Task Details

Assessment Task	Individual or Group Task	Weighting	Submission Due Date
Assignment 1	Individual	20%	End of Week 4
Assignment 2	Individual	30%	End of Week 10
Online Test 1	Individual	10%	Week 5 Lab
Online Test 2	Individual	10%	Week 9 Lab
Online Test 3	Individual	30%	Week 12 Lab

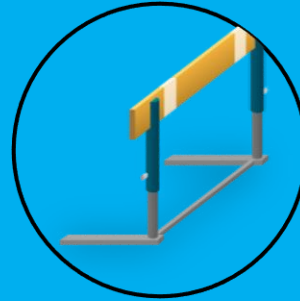


Requirements



Participation Requirement

- You are expected to attend all lectures and labs.
- Use the specified tool in the unit in all assignments and tests.



Minimum requirements to pass this unit

- Achieve an overall mark for the unit of 50% or more.
- Submit all assignments and participating in all tests.



Assignment Requirement

- Submit the assignment on time.
- Deductions will be applied for late submissions. 10% of the mark for that particular assignment will be deducted per day after the deadline.
- Maximum delay is 5 days.





**DEADLINE
* EXTENDED ***

Special Consideration?

If you encounter some difficulties such as medical issues, which affects your progress, during the study, you'll need to launch a **special consideration**.

The special consideration will only extend the deadline in a reasonable range for submitting the assignments. It will not be used to twist the evaluation criteria.



**DEADLINE
* EXTENDED ***

Special Consideration?

A special exam is offered to students who are granted special considerations, unless it is possible to accommodate the special considerations during the semester.



Before the end of the semester: deadline extended.
After the end of the semester: a special exam is required.

Anzac Day Alternative Classes

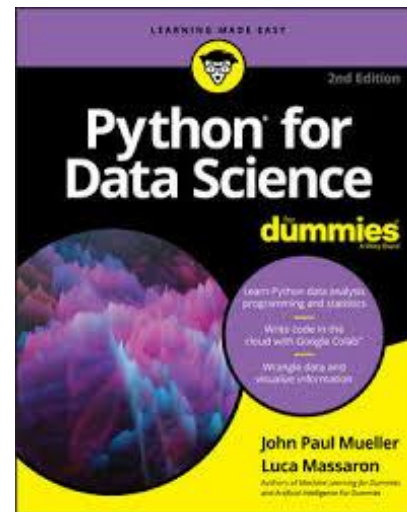
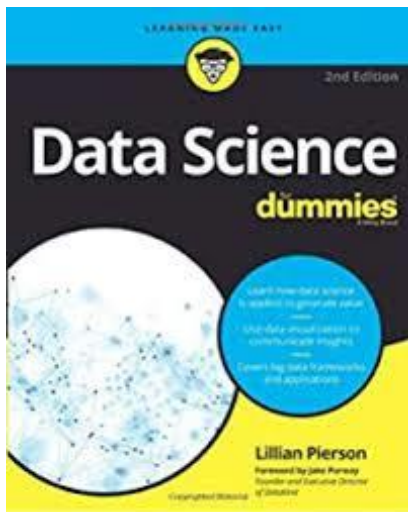
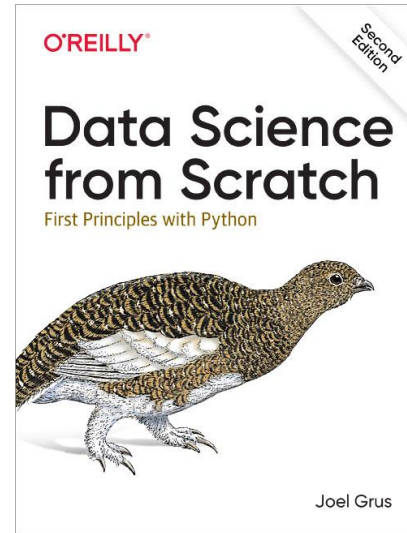
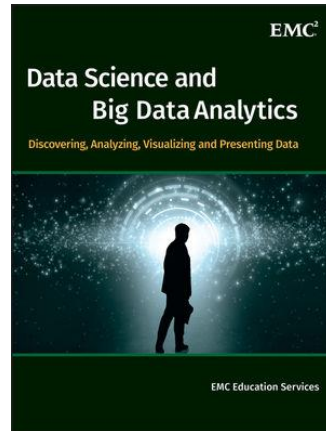
**25/04 (Thursday) is Anzac Day,
which is a public holiday.**

**CL1/03: 10:30 – 12:30 → 12:30 –
14:30 on 22/04 (Monday) @BA404.**

**CL1/04: 14:30 – 16:30 → 18:30 –
20:30 on 22/04 (Monday) @BA404.**

**CL1/06: 12:30 – 14:30 → 10:30 –
12:30 on 22/04 (Monday) @BA404.**

**CL1/07: 08:30 – 10:30 → 14:30 –
16:30 on 22/04 (Monday) @BA404.**



Texts and Resources

- Unless stated otherwise, the materials presented in this lecture are taken from:
 - Dietrich, D. ed., 2015. *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. EMC Education Services.
 - Schutt, R. and O'Neil, C., 2013. *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.
 - Pierson, L., *Data Science for Dummies, 2nd Edition*, John Wiley & Sons © 2017
 - Mueller, J. P. and Massaron, L., *Python for Data Science for Dummies, 2nd Edition*, John Wiley & Sons © 2019 (432 pages), ISBN:9781119547624
 - Joel Grus, 2019. *Data Science from Scratch – First Principles with Python, 2nd Edition*, O'Reilly Media, Inc.



Unit Summary

- Unit structure
- Requirements, Expectations, and Rules
- About Data Science
- Big Data Ecosystem

Summary

