

DASC2010 Calculus for Data Analytics in Science

Course Outline – Fall 2024

1. INSTRUCTOR

Prof. Shingyu Leung

Email: masyleung @ ust.hk

Office: 3436

Office hours: By appointment

2. TA

Mr. Young Kyu LEE

Email: ykleac @ connect.ust.hk

3. MEETING TIME AND VENUE

Time: Tuesday and Thursday 3PM-420PM

Venue: RM1409

4. COURSE DESCRIPTION

Credits: 3

Prerequisites: MATH 1014 OR MATH 1024 OR MATH 1020

Exclusions: MATH 2011, MATH 2023

A concise introduction to multivariable calculus using numerical computing software. Fundamental concepts from multivariable calculus with emphasis on applications and calculations using software. Topics from vectors, curves and parametric equations, differentiation in several variables, with applications in approximation and optimization, and integration in several variables.

5. INTENDED LEARNING OUTCOMES

Upon successful completion of this course, students should be able to

No.	ILOs
1	Explain the basic concepts and techniques of differential and integral calculus of functions of several variables.
2	Use a computer software to do calculation of the gradients, directional derivatives, arc length of curves, area of surfaces and volumes of solids.
3	Solve problems involving maxima and minima, line integral and surface integral, and vector calculus using a computer software

6. ASSESSMENT SCHEME

a. Final examination duration: 3 hours

b. Percentage of coursework, examination, etc.:

Assessment	Assessing Course ILOs
HW	1, 2, 3
Quiz	1, 2, 3
Midterm	1, 2, 3
Final Exam	1, 2, 3

Assessment marks for all assessed tasks will be released within two weeks of the due date.

One lowest HW score and one lowest Quiz score will be dropped.

The final exam is comprehensive, i.e., all materials taught throughout the semester will be tested, including those already tested in the midterm exam. However, the focus will be on topics not covered in the midterm.

Your final score for the course will be calculated based on the following formula,

$$\text{Scheme A} = 25\% \text{ HW} + 10\% \text{ Quiz} + 25\% \text{ MT} + 40\% \text{ Final}$$

$$\text{Scheme B} = 25\% \text{ HW} + 10\% \text{ Quiz} + 0\% \text{ MT} + 65\% \text{ Final}$$

$$\text{Grade} = \max(\text{Scheme A}, \text{Scheme B})$$

Grade	Short Description	Elaboration On Subject Grading Description
A	Excellent	Mastery of multivariable calculus concepts; consistently excels in problem-solving and software applications; demonstrates exceptional analytical skills.
B	Good	Strong understanding of calculus and its applications; effectively solves problems and uses software with minor errors; shows good analytical abilities.
C	Satisfactory	Adequate grasp of course material; meets basic requirements in problem-solving and software use; some inconsistencies in understanding and application.
D	Marginal Pass	Minimal understanding of key concepts; struggles with problem-solving and software applications; meets only the basic criteria for passing.
F	Fail	Fails to demonstrate understanding of multivariable calculus; unable to apply concepts or use software effectively; does not meet course requirements.

7. STUDENT LEARNING RESOURCES

Lecture Notes.

Reference

- Stewart, J. and Clegg, D. and Watson, S., Calculus – Early Transcendentals, Cengage. Chapters 12-15.

8. TEACHING AND LEARNING ACTIVITIES

Scheduled activities: 4 hours (Lecture + Tutorial)

9. COURSE SCHEDULE (26 lectures)

Week	Lecture(s)	Topic	Reading Materials
1	2	Introduction	Leung: 1
2	2	Python Basics	Leung: 2
3	2	Data Representation	Leung: 3 Stewart: 12.1-12.4
4	1	Classification Problem I	Leung: 4
4-6	3	Explicit Representation of the Decision Boundary	Leung: 5 Stewart: 12.5, 13.1-13.3

6-7	3	Implicit Representation of the Decision Boundary	Leung: 6 Stewart: 12.5-12.6, 14.1
8	1	Classification Problem II	Leung: 7
9	1	MIDTERM	Leung: 1-6
8-10	4	Partial Derivatives	Leung: 8 Stewart: 14.2-14.6
11	2	Optimization	Leung: 9 Stewart: 14.7-14.8
12	1	Classification Problem III	Leung: 10
12-13	2	Multiple Integrals	Leung: 11 Stewart: 15.1
13	1	Classification Problem IV	Leung: 12

10. COMMUNICATION AND FEEDBACK

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Students who have further questions about the feedback, including marks, should consult the instructor within five working days after receiving the feedback.

11. COURSE AI POLICY

Generative AI is permitted and requested to assist students in all assessments.

12. ACADEMIC INTEGRITY

Students are expected to adhere to the university's academic integrity policy. Students must uphold HKUST's Academic Honor Code and maintain the highest academic integrity standards. The University has zero tolerance for academic misconduct. Please refer to Academic Integrity | HKUST - Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

DASC2020 Applied Linear Algebra for Data Analytics in Science Spring 2024/25

Course Title: Applied Linear Algebra for Data Analytics in Science

Course Code: DASC2020

No. of Credits: 3

Any pre-requisites: MATH 1014 OR MATH 1020 OR MATH 1024

1. Instructor (s) – Name and Contact Details

Name: Dr. PANT, Nidhi

E-mail: pantnidhi@ust.hk

Office Hour: Tuesday (3 pm - 4 pm), Room no. 4451

2. Teaching Assistant (s) - Name and Contact Details

Name: YUE Cheuk Kan Kelvin

Email: ckkyue@connect.ust.hk

Office Hour: Wed 14:00 - 15:00 (By email appointment)

3. Meeting Time and Venue – Lectures, Tutorials/ Laboratory

Section	Date & Time	Venue
L1	Mon 4:30 PM – 5:50 PM Fri 12:00 PM - 1:20 PM	Rm 1527, Lift 22
T1	Th 04:00 PM - 04:50 PM	Rm 2463, Lift 25-26

4. Course Description - Credit Points, Pre-requisite, Exclusion, Brief Information/synopsis

This course is an introduction to linear algebra with applications in least square optimization. Topics include vectors, matrices, and eigenproblems. These ideas will be applied to different applications including data fitting, machine learning, pattern recognition, and finance, etc. Numerical computation is used throughout the course as a learning tool. No previous knowledge of linear algebra is assumed.

Credit Points: 3

Pre-requisite: MATH 1014 OR MATH 1020 OR MATH 1024

5. Intended Learning Outcomes

(State what the student is expected to be able to do at the end of the course according to a given standard of performance)

On successful completion of this course, students are expected to be able to:

1. Explain fundamental linear algebra concepts and techniques, including vector and matrix operations, determinants, matrix inverse, decomposition, eigen decomposition, and singular value decomposition.
2. Apply linear algebra techniques to solve least squares optimization problems in machine learning, including regression, classification, inversion, constrained problems, regularization, and model validation.
3. Organize and communicate the results of linear algebra computations in the context of machine learning applications, including visualizing data and identifying patterns and trends.
4. Design and analyze machine learning algorithms including dimensionality reduction techniques using singular value decomposition.

6. Assessment Scheme

Homework	10%
Tutorial	15%
Quiz	15%
Midterm	25%
Final Exam	35%

7. Criterion-Referenced

Grading

Grading in this course will be based on **criterion-referencing**, meaning students will be graded against a defined standard or rubric rather than being compared to other students.

- a. The difficulties of the assessment will be considered when determining the range of overall course grade in each final grade, such that the final grade reflects the criteria that students achieved in the course. Thus, the range of overall course grades of each final grade may vary semester to semester depending on the difficulties of the assessment, including homework, quizzes and exams.

8. Grading Rubrics

a. Grading rubrics for class exercises:

- i. 0 for no, or irrelevant, or failed attempt
- ii. 1 for correct attempt but with mistakes and incorrect answers
- iii. 2 for correct answer

b. Grading rubrics for homework, midterm and final exam questions.

Steps	Maximum Points	Criteria for Grading	Performance Levels
1. Identify Key Principles	3	- Correct equations and concepts identified. -Logical reasoning and no irrelevant equations.	Excellent (3): All correct principles identified with no errors. Very Good (2.5): Minor mistakes (e.g., 1 irrelevant equation). Good (2): Some correct principles, with multiple minor mistakes. Fair (1): Many mistakes, with little relevance to the problem. Poor (0): No correct principles identified.
2. Apply Problem-Solving Steps	3	Correct steps followed in the algorithm. Logical flow of steps with no major errors.	Excellent (3): Logical steps with no errors. Very Good (2.5): Minor procedural errors. Good (2): Some correct steps, but major gaps in logic. Fair (1): Few correct steps or mostly irrelevant steps. Poor (0): No meaningful steps applied.
3. Solve Mathematical Problems	2	Correct mathematical procedures used. Relevant techniques applied, with no mistakes.	Excellent (2): No math errors; all intermediate steps shown. Very Good (1.75): Minor arithmetic or formatting errors. Good (1.5): Some mistakes, but major steps are correct. Fair (1): Frequent errors or major steps missing. Poor (0): No meaningful math applied.
4. Numerical Calculations	1	All required numbers are calculated correctly.	Excellent (1): All calculations correct and units properly used. Very Good (0.8): Minor calculation or unit errors. Good (0.6): Some calculation errors, but partial correctness. Fair (0.4): Frequent calculation errors

			or missing units. Poor (0): No correct calculations.
5. Interpret and Justify Results	1	Final answer is correct, clear, and justified.	Excellent (1): Clear, correct answer with complete justification. Very Good (0.8): Mostly correct but missing small details. Good (0.6): Correct answer but unclear justification. Fair (0.4): Some justification, but reasoning is incomplete. Poor (0): No justification or correct results.
Total: 10 points (100 %)			

9. Final Grade Descriptors:

[As appropriate to the course and aligned with university standards]

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Students with excellent performance in the course demonstrate a strong grasp of lecture materials, critical thinking ability, and excel in various course assessments.
B	Good Performance	Students with good performance in the course demonstrate a solid grasp of lecture materials, critical thinking ability, and good performance in various course assessments.
C	Satisfactory Performance	Students with satisfactory performance demonstrate an adequate understanding of lecture materials, critical thinking ability, and satisfactory performance in various course assessments.
D	Marginal Pass	Students with a marginal pass show limited understanding of lecture materials, critical thinking ability, and minimal performance in various course assessments.
F	Fail	Students who fail the course display a lack of understanding of lecture

10. DASC2020 policy on using Generative AI (GAI) tools including ChatGPT:

- a. You are encouraged to use any GAI tools in assisting your learning, including working on homeworks and tutorial exercises, just like we encourage you to work in groups with your classmates. But you **MUST** write up your own solution and acknowledge help you received from any source. Copying from GAI tools or any sources, or using them without proper acknowledgement is considered plagiarism.
- b. In our assessments I usually expect short and concise reasoning from you based on what you learned in this course. Reasons given by GAI tools are typically long and tedious, which I will not accept.
- c. No GAI tool is allowed in close-book assessments including midterm and final exams.

11. Student Learning Resources - Lecture Notes, Readings

Textbook:

Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares by
Stephen Boyd, Lieven Vandenbergh
Cambridge University Press; 1st edition (August 23, 2018)

12. Teaching and Learning Activities -

- a. Lectures:
 - Lecture slides provided courtesy of the authors of the textbook, together with a modified version of Python notebooks of the Python Companion to the textbook which will be published in the course JupyterHub.
 - Students are encouraged to bring their own device to class to practice with Python notebooks (No need NOT submit these notebooks).
- b. Tutorials:

A tutorial session consists of two parts, both are counted towards student's final grade:

 1. A **quiz** on Canvas. This will be counted in your final grade.
 2. A worksheet on JupyterHub that you have to submit on or before the next tutorial. This will be counted in your final grade.

13. Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include areas for improvement via canvas. Students who have further questions about the feedback including marks should consult the TA within one week after the assessment feedback is received.

14. Course Schedule (tentative)

Lecture	Topic	Textbook chapter
1	Vectors	1
2	Linear functions	2
3	Norm and distance	3
4	linear independence	5
5	Matrix	6
6	Matrix examples	7
7	Linear equations	8
8	Linear dynamic equations	9
9	Matrix multiplication	10

10	Matrix inversion I	11
----	--------------------	----

11	Matrix inversion II	11
12	Least square	12
13	Least squares data fitting I	13
14	Least squares data fitting II	13
15	Least squares classification	14
16	Multi-objective least squares and inversion	15.1, 15.3
17	Regularization	15.4
18	Constrained least squares	16

19	Portfolio optimization	17.1
20	Clustering	4
21	Determinant and eigenvalue problem	
22	Diagonalization and dynamical system	
23	Matrix decomposition	
24	Singular value decomposition (SVD) and dimensionality reduction	

The Hong Kong University of Science and Technology

Course Syllabus

Object-Oriented Programming for Data Analytics in Science

DASC2110

3 credits

Pre-requisites: COMP1021

Exclusions: COMP2012 & COMP2012H

Name: Wilson Woon

Email: wilsonwoon@ust.hk

Office Hours: By email appointment

Course Description

This course is designed to provide students with a solid understanding of Object-Oriented Programming (OOP) in Python and the skills to apply these concepts to practical data analytics scenarios. Through a structured, hands-on approach, students will progress from foundational Python programming to advanced OOP concepts and graphical user interface (GUI) development.

The course is divided into three key sections:

1. Core Python Refresher

Students will begin with a brief review of essential Python programming concepts they have previously learned. This section serves as a refresher to reinforce their knowledge and boost confidence before delving into OOP.

2. Object-Oriented Programming (OOP)

As the core focus of the course, this section introduces students to the principles and techniques of the OOP paradigm. Topics include classes, objects, inheritance, polymorphism, encapsulation, and other foundational OOP concepts. Students will also explore real-world applications of OOP in Python.

3. Graphical User Interface (GUI) Development

Students will apply their OOP knowledge to build interactive GUI applications using Python libraries. This section provides a practical understanding of how OOP principles are used in software interfaces.

In addition to these core sections, the course will cover other essential Python programming concepts such as exception handling, user input validation, menu design and implementation, and file I/O operations. By the end of the course, students will have the confidence and skills to design and develop data analytics applications using OOP techniques.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. Demonstrate proficiency in foundational and intermediate Python programming concepts.
2. Design and develop efficient solutions using Python Object-Oriented Programming (OOP), including the principles of classes, objects, inheritance, polymorphism, and encapsulation.
3. Apply OOP principles to solve real-world data analytics problems.
4. Build interactive Python applications with graphical user interfaces (GUIs) by leveraging OOP techniques and GUI libraries.
5. Implement robust Python programs with additional programming concepts such as exception handling, user input validation, menu-driven programming, and file I/O operations in the context of data analytics.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided in the assignment specification. Students must achieve a minimum score of 50% to obtain a marginal pass (D grade).

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
Problem Set 1	15%	9 March*
Problem Set 2	15%	23 March*
Project	15%	23 April*
Class Exercises	15%	Various dates*
Mid-Term Examination	15%	21 March*
Final examination	25%	27 May

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Problem Set 1	ILO1	This assignment will contain several programming questions that require students to apply their basic and intermediate knowledge of Python programming to solve data analytics problems.
Problem Set 2	ILO2, ILO3	In this assignment, students will apply OOP programming techniques to implement a simple application for collecting data.
Project	ILO1, ILO2, ILO3, ILO4, ILO5	Students will develop a GUI-based data collection application using Python OOP and intermediate Python programming techniques.

Class Exercises	ILO1, ILO2, ILO3, ILO4	Complete easy to intermediate programming exercises to reinforce understanding of course topics.
Mid-Term Examination	ILO1, ILO2, ILO3	A closed-book examination mainly covering the theoretical and practical aspects of core Python programming. Some questions on OOP will also be included.
Final Examination	ILO2, ILO3, ILO5	A closed-book examination covering the theoretical and practical aspects of intermediate Python programming and OOP.

Grading Rubrics

Please refer to the specifications of each assessment task.

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	• Demonstrate a solid understanding of key theoretical concepts. • Demonstrate a solid competency in solving problems effectively using Python Object Oriented Programming (POOP) concepts. • Demonstrate a solid competency in intermediate Python programming. • Demonstrate a solid competency in applying key concepts taught in class in data analytics workflow and applications. • Demonstrate a solid competency in designing and implementing a graphical user interface (GUI) application for scientific purposes.
B	Good Performance	The elaboration is the same as that of an A grade, but students who fall into this category demonstrate a good understanding and competency. In short, their outputs are not perfect, but they are good. Students may also demonstrate solid understanding or competency in only some parts of the assessment tasks.
C	Satisfactory Performance	Students are demonstrating adequate understanding and competence in all aspects of the assessment tasks.
D	Marginal Pass	Students are demonstrating minimal understanding and competence in all aspects of the assessment tasks.
F	Fail	Students are demonstrating little to no understanding and competence in all aspects of the assessment tasks.

Course AI Policy

In general, the use of Generative AI is **encouraged**. However, students must understand its limitations and drawbacks. Students are encouraged to use it to solve problems and improve their knowledge and competence. It should never be used to generate answers to assessment tasks. Generative AI is not permitted in closed-book examinations.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include comments on strengths and areas for improvement. Students who have further questions about the feedback, including marks, should consult the instructor within five working days after the feedback is received.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to [Academic Integrity | HKUST – Academic Registry](#) for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

Reading and Learning Resources

To excel in this course and gain a strong command of Python programming and Object-Oriented Programming (OOP) concepts, students are encouraged to make full use of the provided readings and resources. These materials have been carefully curated to complement the course content and include a mix of textbooks, online articles, and instructional videos. They serve as essential tools for reinforcing classroom learning, deepening your understanding of key concepts, and developing practical programming skills.

- 1) Object-Oriented Programming in Python. Download the e-book from <https://python-textbok.readthedocs.io/en/latest/index.html>
- 2) Python Full Course for Free by Bro Code. Available on YouTube via <https://www.youtube.com/watch?v=XKHETdghLK8>
- 3) A Survey of Popular Python IDEs in 2023. Available from <https://www.simplilearn.com/tutorials/python-tutorial/python-ide>
- 4) PyCharm Version Control w/ Git & GitHub | Basics in 12 Minutes | Updated 2024. Available from <https://www.youtube.com/watch?v=8ZEssR8VTKo>
- 5) Object-Oriented Programming in Python – Explained in Plain English by Tiago Monteiro. Available from <https://www.freecodecamp.org/news/object-oriented-programming-python/#where-to-go-from-here>
- 6) An Introduction to Object-Oriented Programming by NullPointerException. Available from <https://www.youtube.com/watch?v=3ycztbPPOlc>. This tutorial provides a high-level overview of OOP without any codes.
- 7) Python Object Oriented Programming (OOP) - For Beginners. Available from https://www.youtube.com/watch?v=JeznW_7DIB0

- 8) Advanced Exception Handling in Python by NeuralNine. Available from <https://www.youtube.com/watch?v=ZUqGMDppEDs>
- 9) Testing, Debugging, Exceptions, and Assertions by Ana Bell (MIT). Available from <https://www.youtube.com/watch?v=9H6muyZjms0>
- 10) How to Validate User Inputs in Python | Input Validation in Python by Fabio Musanni. Available from https://www.youtube.com/watch?v=LUWyA3m_r0
- 11) Make Menus In Python with While True Loop | Choose an Option in Python by Fabio Musanni. Available from <https://www.youtube.com/watch?v=ZBx7oWCJ4aY>
- 12) File Input/Output (I/O) by CS50P. Available from <https://www.youtube.com/watch?v=KD-Yoel6EVQ>

Programme	Bachelor of Science (Honours) in Data Analytics in Science		
Course Title	A Survey on Big Data in Science and Society	Course Code	DASC 2210
QF Level	5	Credit Units	1
Semester	Fall 2024-2025	Time / Week	Wed, 13:30 – 14:50
Instructor	Dr. Wilson Woon (wilsonwoon@ust.hk)	Venue	CYTG009B

Course Description: To give students a broad understanding of the theoretical and practical aspects of data analytics and its applications.

Pre-requisite: None.

Course Intended Learning Outcomes (CILOs)

After completing this course, students should be able to:

1. Understand the fundamentals of Data Science and Artificial Intelligence.
2. Understand the key phases in data analytics.
3. Apply data analytics skills in practical applications.
4. Understand data analytics applications in different areas of science and society.

Assessment and Grading

This course will be assessed using criterion-referencing, and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below.

Performance Evaluation

The course will be graded using the Distinction/Pass/Fail scheme. Students must achieve a minimum score of 50% to pass the course.

Assessment Task	Percentage of Overall Grade	Mapped CILOs
Class Assignments ¹	15%	ILOs 1, 2, and 4
Hands-on Activities ²	25%	ILO 3
Analyses & Visuals	25%	ILO 3
Written Report	15%	ILO 2, 3, and 4
Final Presentation	20%	ILO 2, 3, and 4
Total	100%	

Remarks:

1. At least one in-class assignment in each lesson on Weeks 3, 4, 5, 10, and 12.

2. At least one activity in each lesson on Weeks 6, 7, 8, 9, and 11.

Penalty for Late Submissions:

- 1 day: -20%
- 2 days: -40%
- 3 days or later: -100%

Bring Your Device (BYD)

Students must bring their laptops or tablets with a web browser and, most importantly, **Microsoft Excel 365 (English version)** to class. Students are responsible for ensuring their devices are fully charged and functional for class activities and assignments. The instructor will provide or recommend any additional software or tools required for the course.

It is important to note that all the course materials have been tested to run well on Windows 11. If you use another operating system or another version of Windows, the course materials may not work. You are responsible for finding a solution to this problem.

Readings/References:

- Marr, B. (2016) Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results, Wiley.
(Access via UST internal network: <https://onlinelibrary.wiley.com/doi/epub/10.1002/9781119278825>)
- Gutman, A.J. & Goldmeier, J. (2021) Becoming a Data Head, Wiley.
(Access via UST internal network: <https://app.knovel.com/kn/resources/kpBDHHTSU1/toc>)
- Mayer-Schonberger, V. & Cukier, K. (2013) A Revolution That Will Transform How We Live, Work, and Think, HarperCollins Publishers.
(Access via UST internal network: https://lbdiscovers.hkust.edu.hk/bib/cdi_askewsholts_vlebooks_9780544002937)
- Berman, J.J. (2018) Principles and Practice of Big Data, 2nd Edition. Academic Press. (Access via UST internal network: <https://www.sciencedirect.com/science/article/pii/B9780128156094099908>)

HKUST Code of Honour

Students must be aware of University policy and regulations on honesty in academic work and of the disciplinary guidelines and procedures applicable to breaches of such policies and regulations, as contained in <https://registry.hkust.edu.hk/resource-library/academic-integrity>.

Code of Conduct on the Use of Generative AI

The use of Generative AI is generally encouraged. However, students must understand its limitations and drawbacks. Proper acknowledgements about the use of this tool must be provided in each submitted assessment task.

Accommodation for Students with Special Needs

Students with special needs should inform the instructor of their needs at the beginning of the semester so that appropriate accommodations can be provided.

Tentative Teaching Plan

The following teaching plan is **subject to change**. Please check back from time to time.

Date	Topic	Notes/Due dates
Sept 4	- Course Outline - Introduction to Big Data and Data Science	
Sept 11	- Artificial Intelligence (AI): The Big Picture - The field of Data Science (different roles, expectations, and career prospects)	Class Assignment 0: Examine real-life job titles, technical skills required, and remunerations. Due: 23:55
Sept 18	Mid-Autumn Festival Holiday (No Class)	
Sept 25	Data Analytics Process (Part 1)	Class Assignments 1 & 2: A case study focusing on Ask & Prepare stages Due: 23:55
Oct 2	Data Analytics Process (Part 2)	Class Assignment 3: Understanding the characteristics of “dirty data” Due: 23:55 Class Assignment 4: Reflections on the Data Analytics Process Due: 23:55
Oct 9	Everything About Data	Class Assignment 5: Data Exploration Due: 23:55
Oct 16	Data Cleaning	Hands-on Activity 1: Data Cleaning Due: 23:55
Oct 23	Basic Microsoft Excel	Hands-on Activity 2: Data Cleaning Due: 23:55
Oct 30	Data Analysis & Visualization	Class Assignment 6: Data Analysis; Due: 23:55 Hands-on Activity 3: Data Analysis & Visualization; Due next day at 23:55
Nov 6	Best Practices for Communicating Visuals	Release Project Specifications
Nov 13	Predictive Analytics (Machine Learning)	Class Assignment 7: Training Regression Model; Due: 23:55
Nov 20	Predictive Analytics (Machine Learning)	Class Assignment 8: Training Machine Learning Models; Due: 23:55
Nov 27	Applications of Data Science & Wrapping Up	Project Submission is due on Nov 30
Nov 29	Project Consultation	

Overall Grading Rubrics

Grades	Short Description	Elaboration on subject grading description
Distinction	Excellent Performance	Demonstrate a passionate desire to learn and submit top-quality assessment tasks. Able to apply data analytics skills taught in class to solve problems. Demonstrate excellent written and spoken communication skills, especially in interpreting analytics results and proposing solutions. Possesses high proficiency in data analytics skills, including materials not covered in the course.
Pass	Satisfactory Performance	Demonstrate a passionate desire to learn and submit good-quality assessment tasks. Able to apply data analytics skills taught in class to solve problems. Demonstrate good written and spoken communication skills, especially in interpreting analytics results and proposing solutions. Possesses good proficiency in data analytics toolkits.
Fail	Do Not Meet Minimum Requirements	Demonstrate no desire to learn and submit quality deliverables. Not able to demonstrate data analytics and problem-solving skills. Unable to demonstrate satisfactory written and spoken communication skills.

Programme	Bachelor of Science (Honours) in Data Analytics in Science		
Course Title	Statistics & Probability for Data Analytics in Science	Course Code	DASC 2220
QF Level	5	Credit Units	3
Semester	Spring 2023-2024	Time / Week	Monday, 13:30 – 14:50 Friday, 9:00 – 10:20
Instructor	Dr. Wilson Woon (wilsonwoon@ust.hk)	Venue	Room 4579

Course Objectives

This course aims to provide students with a comprehensive understanding of statistics within data analytics and equip them with the skills to apply these concepts in practical data analytics tasks.

Pre-requisite (Official)

1. COMP1021 Introduction to Computer Science **and**
2. MATH1014 Calculus II **or** MATH1020 Accelerated Calculus **or** MATH1024 Honors Calculus II

Exclusions

1. MATH 2411 Applied Statistics
2. IEDA 2520 Probability for Engineers
3. IEDA 2540 Statistics for Engineers
4. ISOM 2500 Business Statistics
5. LIFS 3150 Biostatistics

Course Intended Learning Outcomes (CILOs)

After completing this course, students should be able to

1. Demonstrate a good understanding of statistics and probability concepts in data analytics.
2. Apply statistics and probability concepts using modern software such as Microsoft Excel.
3. Solve real-life data analytics problems using statistical concepts and tools.

Bring Your Device (BYD)

Students are required to bring their laptops or tablets to class. It should be equipped with Microsoft Excel, a web browser and access to the Internet. Students are responsible for ensuring their devices are fully charged and functional for class activities and assignments. The instructor will provide or recommend any additional software or tools required for the course.

It is important to note that all the course materials have been tested to run well on Windows 11. If you use

another operating system or another version of Windows, the course materials may not work. You are responsible for finding a solution to this problem.

Teaching and Learning Format

We will be using the face-to-face teaching mode. No attendance will be taken during the lessons. Students are responsible for their academic performance.

- 1) Lectures will be held at the venue indicated above. Besides lectures, students may be required to perform hands-on activities.
- 2) Tutorials: Irregular meetings. Please refer to the teaching plan below.
- 3) Post-lesson readings & activities: Students should expect some take-home readings, exercises, and other relevant activities.
- 4) Assignments. They will enhance your understanding of the course topics and provide valuable experience on how statistics is applied in real-life data analytics projects and applications.

The instructor will **try** to record each lecture session and upload them to Canvas within two days.

Performance Evaluation

The course will be graded using the conventional letter grading system found [HERE](#). To obtain a Marginal Pass (D grade) in the course, students must achieve a minimum score of 50%.

Assessment	Percentage of Grade	CILOs assessed
Assignment 1	8% (25 points)	1, 2, 3
Assignment 2	8% (25 points)	1, 2, 3
Assignment 3	8% (50 points)	1, 2, 3
Assignment 4	16% (109 points)	1, 2, 3
Review Quizzes	10% (50 Points)	1, 2
Mid-Term Exam*	20% (100 points)	1, 2
Final Exam*	30% (100 points)	1, 2, 3
Total	100% (459 points)	

* Closed-book examination

Penalty for Late Submissions

- within one day: -20%
- 2 days: -40%
- Three days or later: -100%

Note: No reminder will be sent.

Penalty for Multiple Submissions

Each assignment typically allows a maximum of two submissions, with only the latest attempt being graded. Any additional resubmissions must be pre-approved by the instructor and, if granted, will incur a 50% penalty unless otherwise stated. Submissions via alternative methods, such as email or external links, also require instructor approval and are subject to the same 50% penalty unless specified differently.

Readings and References (To be updated)

- 1) Excel for Beginners – The Complete Course. Available on YouTube via <https://www.youtube.com/watch?v=wbJcJCbCMg>
- 2) Using Excel's XLOOKUP Function. Available on YouTube via <https://www.youtube.com/watch?v=geceMWqY0Xw>
- 3) Kevin Stratvert, “Excel Charts and Graphs Tutorial”. Available on YouTube via <https://www.youtube.com/watch?v=eHtZrIb0oWY>
- 4) Calculating optimal number of bins in a histogram. Read the discussions: <https://stats.stackexchange.com/questions/798/calculating-optimal-number-of-bins-in-a-histogram>
- 5) Struggling with Histogram Bin Width and Bin Intervals? Read the solutions here: <https://www.qimacros.com/histogram-excel/how-to-determine-histogram-bin-interval/>
- 6) Quantitative Specialists, “Skewed Distributions and Mean, Median, and Mode (Measures of Central Tendency)”. Available on YouTube via <https://www.youtube.com/watch?v=xpbYKaEbcPA>.
- 7) StatQuest with Josh Starmer, “The Main Ideas behind Probability Distributions”. Available on YouTube via <https://www.youtube.com/watch?v=ol3hZJqXJuc>
- 8) StatQuest with Josh Starmer, “The Normal Distribution Clearly Explained!”. Available on YouTube via <https://www.youtube.com/watch?v=rzFX5NWoip0>
- 9) StatQuest with Josh Starmer, “The Central Limit Theorem, Clearly Explained!!”. Available on YouTube via <https://www.youtube.com/watch?v=YAlJCEDH2uY>
- 10) The Organic Chemistry Tutor, “Central Limit Theorem - Sampling Distribution of Sample Means - Stats & Probability”. Available on YouTube via <https://www.youtube.com/watch?v=4YLtvNeRIrg>
- 11) zedstatistics, “What are confidence intervals? Actually.” Available on YouTube via <https://www.youtube.com/watch?v=EJe3jiZNwUU>
- 12) StatQuest with Josh Starmer, “Confidence Intervals, Clearly Explained!!!”. Available from YouTube via <https://www.youtube.com/watch?v=TqOeMYtOc1w>
- 13) Data Demystified, “Statistical Significance and p-Values Explained Intuitively”. Available on YouTube via <https://www.youtube.com/watch?v=DAkJhY2zQ3c>

HKUST Code of Honour

Students must be aware of University policy and regulations on honesty in academic work and of the disciplinary guidelines and procedures applicable to breaches of such policies and regulations, as contained in <https://registry.hkust.edu.hk/resource-library/academic-integrity>.

Code of Conduct on the Use of Generative AI

In general, the use of Generative AI is **encouraged**. However, students must understand its limitations and drawbacks. Generative AI is **not** permitted in the mid-term and final exams. These examinations will be closed book.

Accommodation for Students with Special Needs

Students with special needs should inform the instructor of their needs at the beginning of the semester so that appropriate accommodations can be provided.

Tentative Teaching Plan

The following teaching plan is **subject to change**. Please check back from time to time.

Date	Topic	Readings	Notes/Due dates
Friday, 2 Feb	Course Outline & Introduction		
Monday, 5 Feb	Microsoft Excel Basics	1	
Friday, 9 Feb	Microsoft Excel Basics	2	
Monday, 12 Feb	Chinese New Year Holiday (No Class)		
Friday, 16 Feb	Data & Representations	3	
Monday, 19 Feb	Descriptive Statistics		Release Exercise 1
Friday, 23 Feb	Descriptive Statistics	4, 5, 6	- Release Assignment 1 - Release Exercise 2
Monday, 26 Feb	Descriptive Statistics		
Friday, 1 Mar	Descriptive Statistics		Release Review Quiz 1
Monday, 4 Mar	Modelling Data with Probability Distributions	7, 8	
Friday, 8 Mar	Modelling Data with Probability Distributions		Assignment 1 due
Monday, 11 Mar	Modelling Data with Probability Distributions		Release Assignment 2
Friday, 15 Mar	The Central Limit Theorem	9	Release Review Quiz 2
Monday, 18 Mar	The Central Limit Theorem	10	
Friday, 22 Mar	Mid-Term Consultation & Discussion		
Monday, 25 Mar	No Regular Class Mid-Term Exam		
Friday, 29 Mar	Mid-Term Break (No Class)		Assignment 2 due
Monday, 1 Apr	Mid-Term Break (No Class)		
Friday, 5 Apr	Mid-Term Break (No Class)		
Monday, 8 Apr	Making Estimates with Confidence Intervals	11, 12	
Friday, 12 Apr	Making Estimates with Confidence Intervals		
Monday, 15 Apr	Making Estimates with Confidence Intervals		Release Assignment 3 Release Review Quiz 3
Friday, 19 Apr	Drawing Conclusions with Hypothesis Tests	13	
Monday, 22 Apr	Drawing Conclusions with Hypothesis Tests		Assignment 3 due
Friday, 26 Apr	Drawing Conclusions with Hypothesis Tests		Release Assignment 4 Release Review Quiz 4

Monday, 29 Apr	Making Predictions with Regression Analysis		
Friday, 3 May	Making Predictions with Regression Analysis		Release Review Quiz 5
Monday, 6 May	Consultation and Assignment Work		Review Quiz 1 – 4 due
Friday, 10 May	Wrap-up & Final Exam Briefing		Assignment 4 due Review Quiz 5 due
	Study Break		
	Final Exam: Date: 22 May 2024 (Wednesday) Time: 12:30 – 3:30 p.m. Location: [SEAFRONT]TST Sports Ctr MP RM		

Performance Evaluation: Assessment & Grading Rubrics (**Tentative and subject to change**)

Overall Grading Rubrics

Outcome	Grade A: Excellent	Grade B: Good	Grade C: Satisfactory	Grade D: Marginal Pass	Fail
CILO 1	Demonstrate solid understanding. Went above and beyond in grasping the course concepts	Demonstrate good understanding.	Demonstrate adequate understanding.	Demonstrate minimal understanding.	Poor understanding.
CILO 2	Demonstrate expert-level competency	Demonstrate a good level of competency	Demonstrate some competencies	Minimal competencies	Poor competencies
CILO 3	Demonstrate creativity and a significantly high ability to analyse and evaluate cases critically.	Demonstrate good creativity and an ability to analyse and evaluate cases critically.	Demonstrate some creativity and an ability to analyse and evaluate cases critically.	Minimal competencies.	Poor competencies

Programme	Bachelor of Science (Honours) in Data Analytics in Science		
Course Title	Data Structures for Data Analytics in Science	Course Code	DASC 3120
QF Level	5	Credit Units	3
Semester	Fall 2024-2025	Time / Week	Mon, 15:00 – 16:30 Fri, 10:30 – 11:50
Instructor	Dr. Wilson Woon (wilsonwoon@ust.hk)	Venue	Room 2406

Course Objective

This course aims to provide students with an understanding of data structures and their applications in data analytics.

Pre-requisites

COMP1021 Introduction to Computer Science and DASC2110 Object-oriented Programming for Data Analytics in Science. Students must be proficient in the Python Object Oriented Programming (POOP) paradigm.

Course Intended Learning Outcomes (CILOs):

After completing this course, students should be able to:

1. understand data structures' fundamental concepts and principles.
2. demonstrate competency in advanced-level Python programming
3. produce effective and efficient programs using data structures to solve data analytics problems

Bring Your Device (BYD):

Students are required to bring their laptops or tablets to class, equipped with a web browser and access to the Internet. Additionally, students are expected to have **two** Python Integrated Development Environments (IDEs) available on their devices **after the first week**. Students are responsible for ensuring their devices are fully charged and functional for class activities. The instructor will provide or recommend any additional software or tools required for the course.

It is important to note that all the course materials have been tested to run well on Windows 11. If you use another operating system or another version of Windows, the course materials may not work. You are responsible for finding a solution to this problem.

Assessment and Grading

This course will be assessed using criterion-referencing, and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below.

Performance Evaluation

The course will be graded using the conventional letter grading system found [HERE](#). Students must achieve a minimum score of 50% to obtain a marginal pass (D grade).

Assessment	Percentage of Grade	CILOs assessed
Problem Set 1	5% (70 points)	1, 2
Problem Set 2	5% (80 points)	1, 2
Problem Set 3	15% (41 points)	1, 2
Problem Set 4	15% (55 points)	1, 2
Class Exercises	10% (9 points)	1, 2
Mid-Term Exam*	20% (100 points)	1, 2
Final Exam*	30% (107 points)	1, 2
Total	100% (462 points)	

* Closed-book examinations

Penalty for Late Submissions:

- 1 day: -20%
- 2 days: -40%
- 3 days or later: -100%

Readings/References:

1. A Survey of Popular Python IDEs in 2023. Available from <https://www.simplilearn.com/tutorials/python-tutorial/python-ide>
2. Baka, B. (2017) Python Data Structures and Algorithms: Implement Classic and Functional Data Structures and Algorithms Using Python, Packt Publishing.
(Access via UST internal network: https://lbdiscovers.hkust.edu.hk/bib/cdi_proquest_ebookcentral_EBC4868549)
3. Miller, B. N. and Ranum, D. L. (2011) Problem-Solving with Algorithms and Data Structures using Python, 2nd Edition, Franklin, Beedle & Associates.
(Access via <https://runestone.academy/ns/books/published/pythonds/index.html>). There is a wonderful [collection of YouTube videos](#) recorded by Gerry Jenkins to support all of the chapters in this text.
4. freeCodeCamp.org (2021) Data Structures - Computer Science Course for Beginners. Access via <https://www.youtube.com/watch?v=zg9ih6SVACc>. Note: This tutorial video provides an excellent introduction to data structures' principles and working modes. There is no code involved.
5. Python Simplified, Ultimate Guide to NumPy Arrays - VERY DETAILED TUTORIAL for beginners! Available on YouTube via <https://www.youtube.com/watch?v=ILRBYKwP8GQ>.
6. Bro Code, "Python lists, sets, and tuples explained". Available on YouTube via https://www.youtube.com/watch?v=gOMW_n2-2Mw.

7. Bro Code, “Python dictionaries are easy”. Available on YouTube via <https://www.youtube.com/watch?v=MZZSMaEAC2g&>

HKUST Code of Honour

Students must be aware of University policy and regulations on honesty in academic work and of the disciplinary guidelines and procedures applicable to breaches of such policies and regulations, as contained in <https://registry.hkust.edu.hk/resource-library/academic-integrity>.

Code of Conduct on the Use of Generative AI

The use of Generative AI is generally encouraged. However, students must understand its limitations and drawbacks. Proper acknowledgements about using this tool must be provided in each submitted assessment task.

Accommodation for Students with Special Needs

Students with special needs should inform the instructor of their needs at the beginning of the semester so that appropriate accommodations can be provided.

Tentative Teaching Plan

The following teaching plan is **subject to change**. Please check back from time to time.

	Topic	Readings	Notes/Due dates
Mon, 2 Sept	Course Outline & Introduction Introduction to Data Structures		
Fri, 6 Sept	Python IDEs Exploration	1	Class Exercise 0: 0%. Due same day
Mon, 9 Sept	Computational Complexity		
Fri, 13 Sept	Arrays (Part 1)		
Mon, 16 Sept	Arrays (Part 2)		Class Exercise 1: 0.5%. Due same day Release Problem Set 1
Fri, 20 Sept	NumPy (Part 1)	5	Class Exercises 2 & 3: 0.5%. Due same day
Mon, 23 Sept	NumPy (Part 2)	5	Class Exercise 4 & 5: 0.5%. Due same day
Fri, 27 Sept	Lists, Sets, Tuples, and Dictionaries	6, 7	Class Exercise 6 & 7: 0.5%. Due same day Problem Set 1 Due
Mon, 30 Sept	Stacks and Queues (Part 1)		Class Exercise 8 & 9: 0.5%. Due same day
Fri, 4 Oct	Stacks and Queues (Part 2)		Class Exercise 10 & 11: 0.5%. Due same day Release Problem Set 2
Mon, 7 Oct	Linked Lists (Part 1)		Class Exercise 12 & 13: 0.5%. Due same day
Fri, 11 Oct	Chung Yeung Festival Holiday – No Class		
Mon, 14 Oct	Mid-Term Consultation & Discussions		Problem Set 2 Due
Fri, 18 Oct	No Regular Class Mid-Term Exam		
Mon, 21 Oct	Linked Lists (Part 2) & Hash Tables (Part 1)		Class Exercise 14: 0.5%. Due same day
Fri, 25 Oct	Hash Tables (Part 2)		Class Exercise 15: 0.5%. Due same day

			Release Problem Set 3
Mon, 28 Oct	Search Algorithms		Class Exercise 16: 0.5%. Due same day
Fri, 1 Nov	Sort Algorithms (Part 1)		Class Exercise 17: 0.25%. Due same day Class Exercise 18: 0.25%. Due same day
Mon, 4 Nov	Sort Algorithms (Part 2)		Class Exercise 19: 0.5%. Due same day Problem Set 3 Due
Fri, 8 Nov	Trees (Part 1) & Sort Algorithms (Part 3)		Class Exercise 20: 0.5%. Due same day
Mon, 11 Nov	Trees (Part 2)		Class Exercise 21: 0.5%. Due on 15 Nov. Release Problem Set 4
Fri, 15 Nov	Trees (Part 3) & Sort Algorithms (Part 4)		Class Exercise 22: 0.5%. Due same day
Mon, 18 Nov	Trees (Part 4)		Class Exercise 23: 0.5%. Due same day
Fri, 22 Nov	Graphs		Class Exercise 24: 0.5%. Due same day Problem Set 4 Due
Mon, 25 Nov	Wrap-Up & Exam Briefing		
Fri, 29 Nov	Exam Consultation & Q&As		
	Study Break		
	Final Exam: 12 December Time: 12:30 – 3:30 Venue: CYTG009A		

Overall Grading Rubrics

Grades	Short Description	Elaboration on subject grading description
A+, A, & A-	Excellent Performance	Demonstrate a solid understanding of fundamental concepts and principles of data structures and algorithms. Able to apply suitable concepts and programming techniques to solve data analytics problems efficiently and effectively.
B+, B, & B-	Good Performance	Demonstrate a good understanding of fundamental concepts and principles of data structures and algorithms. Able to apply suitable concepts and programming techniques to solve data analytics problems.
C+, C, & C-	Satisfactory Performance	Demonstrate a satisfactory understanding of fundamental concepts and principles of data structures and algorithms. Able to solve data analytics problems, but the techniques can be improved.
D	Marginal Pass	Demonstrate a need for a greater understanding of fundamental concepts and principles of data structures and algorithms. Not able to solve all data analytics problems effectively.
F	Fail	No understanding of data analytics and algorithms concepts, principles, and implementations. Not able to solve any of the data analytics problems.

DASC 3230 Statistical Modeling and Machine Learning, Fall 2024

1. Instructor

Yi LIAO, Ph.D.

Assistant Professor, Division of Life Science

Email: liaoy@ust.hk

Office: Room 5471 (Lift 25/26)

Office Hours: Wed & Fri 3:00–3:45 PM, or by appointment if you have conflicting schedule

2. Teaching assistants

Weijin LIU (Tutorials 1–3, HW 1)

Email: weijin.liu@connect.ust.hk

Jooran LEE (Tutorials 4–6, HW 2)

Email: jleebu@connect.ust.hk

Wenjie ZHANG (Tutorials 7–9, HW 3)

Email: wzhangdk@connect.ust.hk

Zixiang LUO (Tutorials 10–12, HW 4)

Email: zluoaw@connect.ust.hk

3. Meeting Time and Venue

Lectures: Wed & Fri, 1:30 PM – 2:50 PM, Rm 4579, Lift 27-28

Tutorial: Fri, 5:00 PM – 5:50 PM, Rm 4402, Lift 17-18

4. Course description

This course introduces fundamental principles and techniques of statistical modeling for uncovering patterns and making predictions from data. Various statistical learning methods routinely used in the fields of data science and machine learning will be covered, including linear regression, classification, random forest, support vector machines, dimension reduction, clustering, graphical models, and neural networks. The course contents include both theoretical concepts and Python coding demonstrations to help students develop a solid understanding of the core principles and gain practical experience in algorithm implementation. Through conceptual and hands-on explorations, students will acquire essential knowledge and skills for careers in a society that is increasingly driven by data and machine intelligence.

Prerequisite: DASC 2020 and DASC 2220

Exclusion: None

5. Intended Learning Outcomes

Upon successful completion of this course, students should:

1. Develop the ability to formulate scientific questions into appropriate statistical models

2. Possess the knowledge to select the appropriate modeling approach(es) based on the nature of questions and the types of data
3. Be able to perform modeling and statistical analysis using the Python programming language
4. Be able to interpret modeling outcomes and draw conclusions from analysis results
5. Develop mechanistic understanding about the theories, potential pitfalls, and limitations of various modeling techniques introduced in the course

6. Assessment Scheme

Credit points: 3

<u>Assessment</u>	<u>Course ILOs</u>
Homework 40%	1,2,3,4,5
In-class Participation 10%	1,2,4,5
Midterm Exam 25%	1,2,4,5
Final Exam 25%	1,2,4,5

Homework: Homework assignments include derivations, proofs, and numerical and programming problems.

We will use Canvas to distribute and collect the homework. We will also send course announcements via Canvas. So please turn on the notification and check regularly. You are encouraged to post your questions under the Discussion tab on Canvas so others can see and offer help. The instructor will also check and answer the Discussion tab regularly.

You can discuss and collaborate with classmates on homework. However, each person must write and submit his/her own answers. Copy-and-paste is not allowed.

Late submissions: A 1% deduction will be applied for each homework assignment submitted within 24 hours after the deadline. No points will be awarded for assignments submitted more than 24 hours late.

In-class Participation: Attendance will be taken during each lecture. Students need to attend at least 50% of the lectures (12 out of 24 lectures) in order to earn a full (10%) participation score. Students attending 8–11 lectures will receive a 5% participation score, and those attending fewer than 8 lectures will not earn any participation score (0%). If a lecture session is cancelled due to unforeseen circumstances such as bad weather, it will be presumed that all students were present for that day's lecture.

Exams: The midterm exam is scheduled at the regular class time, 1:30 PM – 2:50 PM on Wednesday, October 30, and will be held in the regular classroom of Rm 4579. In case of class cancellation due to unforeseen circumstances, the exam will be postponed to Friday, November 1. The midterm exam will include all materials covered during the first 12 lectures of the course. The time and venue for the final exam will be arranged by the University, and will be announced once confirmed.

Both exams include multiple-choice and short-answer questions. Exams are closed-book, but each student is allowed to bring along one A4-size cheat sheet and a basic (non-graphing and non-programmable) calculator. In principle, the final exam focuses on materials covered after

the midterm exam only (lectures 13–24), but do note that certain concepts introduced later in the course do require a solid understanding of those covered earlier.

No make-up exam is allowed. If an exam is missed, it will automatically receive a score of zero, except for medically or legally justifiable reasons supported by signed documentation from a medical professional or an attorney. In such case, the missed midterm (or final) exam score will be substituted with the score from the final (or midterm) exam. Missing both the midterm and the final exams will result in zero scores for both, without exception.

7. Student Learning Resources

Lecture and tutorial slides and code snippets are the primary reference materials on which homework assignments and exams will be based, and will typically be uploaded to Canvas before each lecture.

Optional materials for reference:

An Introduction to Statistical Learning by James et al. [available online from HKUST library]

Pattern Recognition and Machine Learning by Bishop [reserved at HKUST library]

Machine Learning: A Probabilistic Perspective by Murphy [available online from HKUST library]

8. Teaching and Learning Activities

Scheduled weekly activities: 4 hours

Teaching Activities

Lecture

Tutorial

Course ILOs

1,2,3,4,5

3,4,5

9. Tentative Topics

The course is divided into 9 modules (with certain degree of interdependence), each focusing on one particular category of statistical modeling and machine learning techniques:

- (1) Course Introduction; Linear Regression
- (2) Model Selection
- (3) Linear Classification
- (4) Kernel Method and Support Vector Machines
- (5) Dimension Reduction
- (6) Clustering
- (7) Graphical Models
- (8) Combining Models*
- (9) Neural networks and Introduction to Deep Learning*

* Subject to change depending on course progress.

DASC 3240 Data Visualization in Science

Spring 2024-25

1. Instructor(s)

Name: Masayuki USHIO

Contact Details: Room CYT-2013 (Lift 35/36), ushio@ust.hk,

2. Meeting Time and Venue

Lectures: **Date/Time:** Monday and Wednesday (10:30–11:50)

Venue: CYT G009B

Tutorials: **Date/Time:** Thursday (18:00–18:50)

Venue: Room 2304 (Lift 17/18)

3. Course Description

Credit Points: 3 Pre-requisite: DASC 2220 Exclusion: COMP 4462

Data visualization is the graphical representation of applied data science. It can also provide us with a powerful way to communicate data-driven findings, motivate analyses, and detect flaws in an infographic or dashboard. This course illustrates how to use the techniques of data visualization and discovery tools to explore, visualize and analyze data. By the end of the course, students will be able to utilize tools and packages in R or Python to enhance their skills on science communication.

4. Intended Learning Outcomes

Upon successful completion of this course, students should be able to:

No.	ILOs
1	Elaborate the basic concept of visualization tools.
2	Do simple data processing given a variety of data formats with R/ Python.
3	Formulate a visualization solution to some real-data problems and interpret the results by graphs.
4	Implement some visualization techniques with R or Python.
5	Apply the conceptual and practical skills to interpret data in physics, chemistry, life science and ocean science as well as other disciplines with data.

5. Course Assessments

- Weekly Assignments (40%)
- Final Project (60%)

Assessment Task	Contribution to Overall Course (%)	Due Date
Weekly Assignments (Total 8 assignments)	5% each, 40% in total	Before the first class in each week
Final Project	60%	17:00, 16 May 2025

6. Major References

- Claus O. Wilke, “Fundamentals of Data Visualization”, StackAbuse. ISBN 9781492031086, <https://clauswilke.com/dataviz/>
- Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen, “ggplot2: Elegant Graphics for Data Analysis (3rd edition)”, <https://ggplot2-book.org/>
- Holz Yan, “from Data to Viz” https://github.com/holtzy/data_to_viz (MIT license)

7. Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Assignments	ILO1, ILO2, ILO3, ILO4	This task assesses students’ knowledge of and ability to implement basic data visualization (ILO1, ILO2, ILO3, and ILO4).
Final Project	ILO1, ILO2, ILO3, ILO4, ILO5	This task assesses students’ knowledge of and ability to implement basic data visualization (ILO1, ILO2, ILO3, and ILO4), and application of the conceptual and practical skills to interpret data (ILO5).

8. Final Grade Descriptors

Grades	Short Description	Elaboration on Subject Grading Description
A	Excellent Performance	Students demonstrate a deep and thorough understanding of the fundamental principles of data visualization, including layout, color usage, licensing, and tools for visualizing data. They are able to produce graphs that accurately and effectively represent the key features of raw data. Their code is correct, well-formatted, and easily understandable by others. They can accurately interpret graphs in any format.
B	Good Performance	Students demonstrate a strong understanding of the fundamental principles of data visualization, including layout, color usage, licensing, and tools for visualizing data. They are able to produce graphs that accurately represent the key features of raw data. Their code is generally correct and well-formatted, and it can be understood by others with careful review. They can accurately interpret graphs in most formats.
C	Satisfactory Performance	Students demonstrate a satisfactory understanding of the basic principles of data visualization, including layout, color usage, licensing, and tools for visualizing data. They are able to produce graphs that represent the key features of raw data, although they may overlook some important details. Their code is generally correct, but the graphs may not always be reproducible by others. They can accurately interpret graphs in standard formats.
D	Marginal Pass	Students demonstrate a minimal understanding of the basic principles of data visualization, including layout, color usage, licensing, and tools for visualizing data. They are able to produce graphs, but their graphs may lack some key features of the raw data. Their code may work, but it is difficult for others to read and often fails to reproduce the graphs. They can interpret graphs in common formats, but they often misinterpret graphs in less common or more complex formats.
F	Fail	Students have not met the minimum requirements for the course. They demonstrate a lack of understanding of the core concepts and tools in data visualization. They are unable to produce graphs in most formats. Their code does not work, and graphs cannot be generated. They are unable to interpret graphs in most formats.

9. Communication and Feedback

Marks for the assessments and Final Project will be communicated via Canvas within two weeks of the assessment date.

10. Course AI Policy

In this course, students are allowed to use generative AI to assist them in various ways. However, appropriate credit must be given for any use of generative AI. Additionally, students must review, analyze, and revise the output from AI to ensure it is better suited for the assessments or projects. Students must not copy, paste, and submit the output as if it is entirely their own work.

11. Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to [Academic Integrity – HKUST – Academic Registry](#) for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

DASC 3240 Spring 2024-25 Tentative Course Schedule**Lectures:** Monday and Wednesday (10:30–11:50)**Venue:** CYT G009B

Wk	L#	Date	Topic	Format	Assignments
1	1	3 Feb (Mon)	Course introduction Basics in figure and data presentation I	Lecture	
	2	5 Feb (Wed)	Basics in figure and data presentation II	Lecture	
2	3	10 Feb (Mon)	Basics in figure and data presentation – Exercise	Hands-on	
	4	12 Feb (Wed)	R and RStudio – I. Overview and installation	Lecture + hands-on	
3	5	17 Feb (Mon)	R and RStudio – II. R basics	Lecture + hands-on	
	6	19 Feb (Wed)	Data manipulation and ggplot2	Lecture + hands-on	
4	7	24 Feb (Mon)	Plotting data – Basics and more	Lecture + hands-on	
	8	26 Feb (Wed)		Assignments	Yes
5	9	3 Mar (Mon)	Scatterplot and regression	Lecture + hands-on	
	10	5 Mar (Wed)		Assignments	Yes
6	11	10 Mar (Mon)	Visualization of high dimensional data: PCA and others	Lecture + hands-on	
	12	12 Mar (Wed)		Assignments	Yes
7	13	17 Mar (Mon)	Visualization of spatial data: Maps and images	Lecture + hands-on	
	14	19 Mar (Wed)		Assignments	Yes
8	15	24 Mar (Mon)	Understanding variations and uncertainties in plots	Lecture + hands-on	
	16	26 Mar (Wed)	Understanding and interpreting plots	Lecture + hands-on	
9	17	31 Mar (Mon)		Assignments	Yes
		2 Apr (Wed)	Mid-term break	No class	
10	18	7 Apr (Mon)	Random topics in data visualization and programming	Lecture + hands-on	
	19	9 Apr (Wed)	plotly – Interactive plots	Lecture + hands-on	
11	20	14 Apr (Mon)	gganimation – Creating animation	Lecture + hands-on	
	21	16 Apr (Wed)		Assignments	Yes
12		21 Apr (Mon)	Easter Monday	No class	
	22	23 Apr (Wed)	Integrating R and Python	Lecture + hands-on	
13	23	28 Apr (Mon)	Preparing publication-ready plots	Lecture + hands-on	
	24	30 Apr (Wed)		Assignments	Yes
14		5 May (Wed)	Buddha's Birthday	No Class	
	25	7 May (Wed)	Course Review and Q&A session		

DASC 3240 Spring 2024-25 Tentative Schedule for Tutorials**Tutorials:** Thursday (18:00–18:50)**Venue:** Room 2304 (Lift 17/18)

Wk	T#	Date	Topic	
1	1	6 Feb (Thu)	Setting up R/RStudio environment	Technical support for the setup
2	2	13 Feb (Thu)	Setting up R/RStudio environment	Technical support for the setup
3	3	20 Feb (Thu)	R and basic statistics	
4	4	27 Feb (Thu)	R and basic statistics	
5	5	6 Mar (Thu)	Scatterplot and regression	Follow-up section for scatterplot
6	6	13 Mar (Thu)	Visualization of high dimensional data: PCA and others	Follow-up section for dimension reduction
7	7	20 Mar (Thu)	Maps and images	
8	8	27 Mar (Thu)	Understanding uncertainty	
9		3 Apr (Thu)	Mid-term break	No Class
10	9	10 Apr (Thu)	Setting up environment for animation	Technical support for the setup
11	10	17 Apr (Thu)	Setting up environment for Python	Technical support for the setup
12	11	24 Apr (Thu)	Setting up environment for Python	Technical support for the setup
13		1 May (Thu)	Labor Day	No Class
14	12	8 May (Thu)	Course Review and Q&A session	

Items of Course Outlines: DASC 3250, Fall 2025

1. Instructor (s) – Name and Contact Details

Zhichao Peng, pengzhic@ust.hk

2. Teaching Assistant (s) - Name and Contact Details

Zitong Chao, zchaoaa@connect.ust.hk

3. Meeting Time and Venue – Lectures, Tutorials/ Laboratory

Lectures: Tuesday, Thursday 15:00-16:20, Room 2302

Tutorials: Friday 15:30-16:20, Room 6602

4. Course Description - Credit Points, Pre-requisite, Exclusion, Brief Information/synopsis

Credit Points: 3

Pre-requisite: COMP 1021 And (DSAC 202/Math 2111/Math 2121/Math 2131/Math 2350)

Exclusion: MATH 3312, MECH 4740, PHYS 3142

Brief Information: This course introduces numerical methods for various data analytic tools. Topics include numerical algorithms for linear systems, eigenvalues and eigenvectors, nonlinear equations, interpolation and approximation, numerical integration and solution of ordinary differential equations, fundamental theory and techniques of constrained and unconstrained optimizations, fundamental techniques and software for machine learning. Examples are taken from various applications in both physical and life sciences.

5. Intended Learning Outcomes

(State what the student is expected to be able to do at the end of the course according to a given standard of performance)

1. (1) Know basic algorithms for numerical analysis, numerical linear algebra and data analysis. (2) Understand the basic mathematical concepts and intuition behind these algorithms. (3) Know when and how to apply/implement these algorithms in Python.

6. Assessment Scheme

<u>Assessment</u> (Percentage + assessment tasks)	<u>Assessing Course ILOs</u> (Respective course ILOs)
30% Homework	ILO (1)-(3)
30% Midterm exam	ILO (1)-(3)
40% Final exam	ILO (1)-(3)

7. Student Learning Resources - Lecture Notes, Readings

- (1) Lecture notes (JupyterNotebook with sample Python sample code and examples + corresponding PDF version)
- (2) Reading: Introduction to Linear Algebra, G. Strang (digital version freely available through HKUST library)
- (3) Numerical analysis: A Graduate Course, D.E. Stewart (digital version freely available through HKUST library)

8. Teaching and Learning Activities -

- a. Lectures: focus on mathematical aspects and intuition of algorithms, demo codes to demonstrate practical aspects on the implementation/application of algorithms
- b. Tutorials/Laboratory: focus on problem solving, implementation details of algorithms, solutions to homework/exam problems

9. Course Schedule

- a. *Topics taught*
- b. *Weeks or dates of teaching specific topics (optional)*

1. Direct and iterative algorithms for linear systems
2. Eigenvalues and eigenvectors, power method for eigenvalue computation and page-rank
3. SVD and principal component analysis
4. Least squares problem and QR decomposition
5. Numerical interpolation
6. Algorithms for nonlinear equations
7. Numerical integration
8. Numerical differentiation with finite difference method and numerical algorithms for ODEs
9. Basic optimization algorithms for constrained and unconstrained optimization
10. Brief introduction of neural networks and deep learning

Course Syllabus

[Jump to Today](#)

Practical Artificial Intelligence in Science

Course Description:

This course provides students with an overview of the field of Artificial Intelligence (AI) focusing on machine learning (ML) and its applications. It introduces several basic AI/ML Python software packages with emphasis on using existing AI techniques provided by the Python packages to solve problems in science. Throughout the course, students will gain hands-on experiences of using modern AI/ML software via solving practical problems.

Course ILOs:

1. Explain basic knowledge about the field of artificial intelligence (AI) and machine learning (ML).
2. Apply state-of-the-art Python AI/ML packages to solve realistic problems such as regression, optimization and neural networks.
3. Develop frameworks for using AI techniques to solve specific scientific problems.

Assessment Weightings:

Homework Assignment	40%
Midterm Project	20%
Final Report	30%
Final Presentation	10%

Course Schedule

Week	Topics	Briefly outline what this topic will cover (Include reading assignments if available)	Indicate which course ILOs this topic is related to (Write CILO-1, CILO-2, etc.)
------	--------	--	---

1	Overview of AI	Introducing basic concepts, historical development of ideas, branches, classic topics: search algorithms, knowledge representation, probabilistic reasoning	CILO-1
2	Introduction to Python AI packages	Python environment, AI packages such as Scikit-learn, PyTorch, TensorFlow, and example usages of the packages	CILO-2
3	Overview of Machine learning	Types of machine Learning: supervised, unsupervised, reinforcement learning	CILO-1
4	Regression, Generalized Linear Models	Linear regression, Logistic regression,	CILO-2
5	Basics for using ML	Data preparation, Performance Metrics and Validation	CILO-2
6	Model Selection and Optimization	Validation and optimize parameters, KNN classification, Naive Bayesian c, Support Vector Machine	CILO-2
7	Ensemble Learning	Decision Tree, Random Forest, Extremely Random Forest, Handling class imbalance, Boosting	CILO-1, CILO-2
8	Mid-Term Project	Work sessions for mid-term project	CILO-1, CILO-2, CLIO-3
9	Unsupervised Learning	Clustering data with K-means algorithm, Gaussian Mixture Models, Dimension Reduction	CILO-1, CILO-2
10	Neural Networks and Deep Learning	Perceptron, Neural Networks, Deep learning, Multiple layer perceptron,	CILO-1, CILO-2

Convolutional Neural Networks (CNN)

11	Applications in science research/topics for final project	Atmospheric science modeling, parameterization, forecasting, physics, astrophysics	CILO-1, CILO-3
12	Final Project	A research project work session	CILO-2, CILO-3
13	Final Project presentation		CILO-1-3

Course Summary:

Date	Details	Due
Tue Sep 17, 2024	 Knowledge of Artificial Intelligence (https://canvas.ust.hk/courses/58150/assignments/350752)	due by 11:59pm
Fri Oct 4, 2024	 Linear Regression Implementation using Gradient Descent (https://canvas.ust.hk/courses/58150/assignments/353496)	due by 11:59pm
Thu Oct 31, 2024	 Midterm Project (https://canvas.ust.hk/courses/58150/assignments/356173)	due by 11:59pm
Sat Nov 16, 2024	 Applications of Machine Learning Models (https://canvas.ust.hk/courses/58150/assignments/358612)	due by 11:59pm
Fri Dec 13, 2024	 Final Project (https://canvas.ust.hk/courses/58150/assignments/358763)	due by 11:59pm
	 Final project reflection summary (https://canvas.ust.hk/courses/58150/assignments/359803)	due by 11:59pm

The Hong Kong University of Science and Technology

Course Syllabus

SQL for Data Analytics

DASC4020

3 credits

Pre-requisite: COMP1021

Exclusions: ISOM3260 and COMP3311

Name: Wilson Woon

Email: wilsonwoon@ust.hk

Office Hours: By email appointment

Course Description

This course is designed to provide students with a solid foundation in database design principles and the practical skills to apply these concepts in data analytics. The course is divided into three key parts:

1. Database Design Principles: Students will learn the fundamentals of database design, including Entity-Relationship (ER) modelling, the Relational Model, Normalization, and the Dimensional Model for analytical databases.
2. Structured Query Language (SQL): Basic and Advanced SQL will be taught.
3. Applications in Data Analytics: The course will culminate in applying database concepts and SQL to real-world data analytics tasks, focusing on data preparation, transformation, and exploratory analysis.

By the end of this course, students will be able to design efficient databases, write simple and complex SQL queries, and integrate database techniques into data analytics workflows.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. demonstrate an understanding of database design principles by developing Entity-Relationship Diagrams (ERDs), applying the Relational Model, and designing databases using techniques such as Normalization and Dimensional Design.
2. write and execute SQL queries to perform data retrieval, manipulation, and aggregation, including advanced operations such as joins, nested queries, and subqueries
3. apply database design and SQL skills to real-world data analytics tasks, including data preparation, transformation, and exploratory analysis, to derive meaningful insights

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided in the assignment specification.

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
Assignment 1	15%	19 March *
Assignment 2	15%	13 April *
Assignment 3	15%	30 April *
Class Exercises	15%	continuous *
Final examination	40%	17 May

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Assignment 1	ILO1	Real-life questions on designing effective databases using various design principles.
Assignment 2	ILO2	Write SQL statements to solve real-life data analytics problems.
Assignment 3	ILO3	Design an effective database for a real-life dataset, utilising suitable design principles, and write SQL statements to answer data analytics questions.
Class Exercises	ILO1, ILO2, ILO3	Complete easy to intermediate exercises to reinforce understanding of course topics.
Final Examination	ILO1, ILO2	A closed-book examination covering the theoretical and practical aspects of database design and SQL programming.

Grading Rubrics

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided in the assignment specification. Students must achieve a minimum score of 50% to obtain a marginal pass (D grade).

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	• Demonstrate a solid understanding of key theoretical concepts • Demonstrate a solid competency in technical drawing of ER Diagrams and Relational Model. • Demonstrate a solid competency in creating efficient databases through Normalisation and Dimensional Modelling • Demonstrate a solid competency in writing simple and complex SQL statements. • Demonstrate a solid competency in applying key concepts taught in class in data analytics workflow and applications.
B	Good Performance	The elaboration is the same as for the A grade, but students who fall into this category are demonstrating good understanding and competency. In short, their outputs are not perfect, but they are good. Students may also demonstrate solid understanding or competency in only some parts of the assessment tasks.
C	Satisfactory Performance	Students are demonstrating adequate understanding and competency in all aspects of the assessment tasks.
D	Marginal Pass	Students are demonstrating minimal understanding and competency in all aspects of the assessment tasks.
F	Fail	Students are demonstrating little to no understanding and competency in all aspects of the assessment tasks.

Course AI Policy

In general, the use of Generative AI is **encouraged**. However, students must understand its limitations and drawbacks. Students are encouraged to use it to solve problems and improve their knowledge and competence. It should never be used to generate answers to assessment tasks. Generative AI is not permitted in closed-book examinations.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include comments on strengths and areas for improvement. Students who have further questions about the feedback, including marks, should consult the instructor within five working days after the feedback is received.

Required Texts and Materials

1. Modern database management by Jeffrey A. Hoffer, V. Ramesh, Heikki Topi. Pearson Education Limited. 2020 Edition.
2. Star schema: the complete reference by Christopher Adamson. McGraw-Hill. 2010 Edition.
3. The data warehouse toolkit: the definitive guide to dimensional modelling by Ralph Kimball. John Wiley & Sons. 2013 Edition

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to [Academic Integrity | HKUST – Academic Registry](#) for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

The Hong Kong University of Science and Technology

Course Syllabus

Capstone Project for Data Analytics in Science

DASC4300

3 credits

Name: Wilson Woon

Email: wilsonwoon@ust.hk

Office Hours: By email appointment

Course Description

This course provides final-year data analytics students in the School of Science with a comprehensive platform to apply and integrate their acquired knowledge, technical skills, and problem-solving abilities. By completing this capstone project in a chosen area of interest, students will gain practical experience in managing real-world data analytics challenges. The course also aims to prepare students for career opportunities or advanced academic pursuits by enabling them to produce a professional-quality project that demonstrates their competencies to potential employers or graduate school supervisors.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. demonstrate proficiency in data collection, cleaning, and preparation to ensure data quality and reliability.
2. design and create an insightful and visually compelling dashboard that effectively communicates key findings and major insights from a dataset.
3. solve complex data analytics problems by critically analysing data, selecting appropriate tools and techniques, and applying domain knowledge to propose well-reasoned solutions.
4. utilize data analytics tools and technologies confidently and efficiently to address a wide range of analytical tasks.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided in the assignment specifications.

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
Class Exercises	5%	Week 1 – 3*
Deliverable 1: Project Outline and Data Cleaning	10%	10 March*
Deliverable 2: Data Collection Methodology	25%	24 March*
Deliverable 3: Project Dashboard	30%	9 April*
Deliverable 4: Data Analyses	30%	30 April*

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Class Exercises	ILO1, ILO2, ILO3, ILO4	Complete class exercises to reinforce understanding of course topics.
Deliverable 1	ILO1	Identify a data-intensive area of interest, find a suitable dataset, clean it and describe it.
Deliverable 2	ILO1, ILO4	Describe and implement the data collection methodology.
Deliverable 3	ILO2, ILO4, ILO5	Create a dashboard that outlines the key insights derived from the dataset. It should focus on what the data reveals on the surface, such as trends, comparisons, summaries, and key metrics, and why these patterns occur. Finally, produce a video presentation.
Deliverable 4	ILO3, ILO4, ILO5	Perform data-driven analyses on the dataset with the aim of identifying problems, limitations, and/or areas for improvement in the domain area. Utilise domain knowledge to propose suitable solutions or course of actions. Finally, produce a video presentation.

Grading Rubrics

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided in the assignment specification. Students must achieve a minimum score of 50% to obtain a marginal pass (D grade).

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrate a passionate desire to learn and submit top-quality deliverables. Made impeccable contributions to group submissions (if any). Demonstrate excellent written and spoken communication skills. Possesses high proficiency in data analytics toolkits. Demonstrate a strong ability in selecting and describing the most suitable data collection methodology for the selected dataset and domain. The Dashboard in Deliverable 3 is visually appealing, highly interactive, conforms to standard data visualization principles and presents important insights into the dataset and domain. Finally, the last deliverable is data-driven and highly impactful.
B	Good Performance	Demonstrate a passionate desire to learn and submit good-quality deliverables. Made impeccable contributions to group submissions (if any). Demonstrate good written and spoken communication skills. Possesses good proficiency in data analytics toolkits. Demonstrate a good ability in selecting and describing the most suitable data collection methodology for the selected dataset and domain. The Dashboard in Deliverable 3 is visually appealing, highly interactive, conforms to standard data visualization principles and presents important insights into the dataset and domain. Finally, the last deliverable is data-driven and impactful.
C	Satisfactory Performance	Demonstrate a passionate desire to learn and submit good-quality deliverables. Made substantial contributions to group submissions (if any). Demonstrate satisfactory written and spoken communication skills. Possesses satisfactory proficiency in data analytics toolkits. Demonstrate a satisfactory ability in selecting and describing the most suitable data collection methodology for the selected dataset and domain. The Dashboard in Deliverable 3 is visually appealing, interactive, conforms to standard data visualization principles and presents important insights into the dataset and domain. Finally, the last deliverable is not entirely data-driven and somewhat impactful.
D	Marginal Pass	Demonstrate an adequate desire to learn and submit quality deliverables. Made adequate contributions to group submissions (if any). Demonstrate adequate written and spoken communication skills. Possesses adequate proficiency in data analytics toolkits. The deliverables are largely substandard and lack impact.
F	Fail	Demonstrate a lack of desire to submit adequate quality work in most assessment tasks. The instructor is not convinced that the student wants to do well academically.

Course AI Policy

In general, the use of Generative AI is **encouraged**. However, students must understand its limitations and drawbacks.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include comments on strengths and areas for improvement. Students who have further questions about the feedback, including marks, should consult the instructor within five working days after the feedback is received.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to [Academic Integrity | HKUST – Academic Registry](#) for the University's definition of plagiarism and ways to avoid cheating and plagiarism.