Lingnan University

Department of Computing and Decision Sciences

Course Syllabus

Course Title	:	Introduction to Programming for Data Science
Course Code	:	CDS1001
Recommended Study Year	:	1
No. of Credits/Term	:	3
Mode of Tuition	:	Sectional Approach
Class Contact Hours	:	3 hours per week
Category in Major Prog.	:	Required
Prerequisite(s)	:	Nil
Co-requisite	:	Nil
Exclusion	:	Nil
Exemption Requirement	:	Nil

Brief Course Description:

This course is designed for students to provide an introduction to data science in the digital age. Data science concerns using data to understand and analyse actual phenomena. The course covers basic IT skills, such as computer programming to assist data manipulation, data analysis and data communication. Turing award winner Jim Gray predicts that data science will be a "fourth paradigm" of science, which is data-driven and can be differentiated from empirical, theoretical and computational paradigms. Students will learn the fundamentals and appreciate the importance of data science.

The first half of the course is about learning the programming language. The topics will include: the basic Python language syntax, variable declaration, basic operators, program flow and control, Python data structures, defining and using functions and recursion, file and operating system interface. In the second half of the course, basic Python packages designed for data science will be introduced, such as NumPy, SciPy, Pandas, and Matplotlib.

Aims:

The ability to analyse data using computer programs is essential in data science. This course aims at introducing the basic connects of data science and providing students with fundamental knowledge of programming in the Python programming language, which is one of the most popular programming languages for data science. This course assumes no prior experience with programming.

Learning Outcomes (LOs):

Upon the successful completion of this course, the student will be able to:

1. Identify the fundamental concepts of data science and the applications; (PLO1)
2. Describe the whole data science process; (PLO2)
3. Design, code, and test small Python programs;
4. Accurately use procedural statements — assignments, conditional statements, loops, function call, and arrays;

5. Manipulate and process data for solving problems using the logic of a programming language; (PLO6)
6. Retrieve and store data from/to different sources; (PLO6)
7. Plot data using appropriate Python visualization libraries. (PLO6)

Indicative Contents:

Data Science
Basic concepts of data science and the applications
Brief description of the data science process

Computer Programming
Language syntax, variable declaration, basic operators, program flow, program control
Definition of functions, function calls
File and operating system interface, web retrieval

Troubleshooting with Generative AI Tools
Visual Studio Code with Generative AI Extensions such as GitHub Copilot, ChatGPT, Genie AI, etc.

Data Science Computing
NumPy, SciPy, Pandas, n-dimensional array
Optimization, interpolation, sparse matrix

Data Visualization
Matplotlib, scatter plot, box plot, contour map, multidimensional chart

Teaching Method:

There are different teaching and learning activities including lectures and laboratories. The concepts, process, and applications of data science will be discussed in lectures. Students will also learn computer programming knowledge and the skills of manipulating, processing, retrieving, storing, and plotting data. Students will develop small programs and learn different software packages such as SciPy, NumPy, Pandas, and Matplotlib in laboratories.

Measurement of Learning Outcomes:

	Class Attendance and Participation	Assignments	Mid-term Examination	Examination
1. Identify the fundamental concepts of data science and the applications	x		x	
2. Describe the whole data science process	x		x	
3. Design, code, and test small Python programs		x	x	x
4. Accurately use procedural statements — assignments, conditional statements, loops, function call, and arrays		x	x	x
5. Manipulate and process data for solving problems using the logic of a programming language		x		x
6. Retrieve and store data from/to different sources		x		x
7. Plot data using appropriate Python visualization libraries		x		

1. There are a number of classroom activities to evaluate if the students can identify the basic concepts and the process of data science. Students are expected to identify applications in data science, especially to the domains of social science, arts, and business domains. In-class programming exercises will be assigned to ensure that students learn the skills in computer programming (LO1-2 and PLO1-2).
2. The assignments require students to apply programming skills to solve data science problems using Python (LO3-7 and PLO6).
3. The mid-term examination requires students to demonstrate their understanding of the basic concepts and the process of data science. Basic computer programming skills are also assessed (LO1-4 and PLO1-2).
4. The ability of computer programming as well as the skills of manipulating, processing, retrieving, and storing data are evaluated in the final examination (LO3-6 and PLO6).

Assessment:

Class Attendance and Participation	10%
Assignments	40%
Mid-term Examination	20%
Examination	30%
Total	100%

Required/Essential Readings:

1. Guttag, John V., *Introduction to Computation and Programming Using Python: With Application to Understanding Data,* MIT Press, 2016.
2. VanderPlas, Jake, *Python Data Science Handbook: Essential Tools for Working with Data,* O'Reilly Media, 2016.

Recommended/Supplementary Readings:

1. Grus, Joel, *Data Science from Scratch: First Principles with Python*, O'Reilly Media, 2015.
2. Hill, Christian, *Learning Scientific Programming with Python*, Cambridge University Press, 2016.
3. McKinney, Wes, *Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython. 2nd Edition*, O'Reilly Media, 2017.
4. NumPy User Guide (https://docs.scipy.org/doc/numpy/user/index.html)
5. Ramalho, Luciano, *Fluent Python: Clear, Concise, and Effective Programming*, O'Reilly Media, 2015.
6. Severance, Charles, *Python for Informatics: Exploring Information,* CreateSpace Independent Publishing Platform, 2013.
7. SciPy Lecture Notes (http://www.scipy-lectures.org/)

Important Notes:

(1) Students are expected to spend a total of 9 hours (i.e. 3 hours of class contact and 6 hours of personal study) per week to achieve the course learning outcomes.

(2) Students shall be aware of the University regulations about dishonest practice in course work, tests and examinations, and the possible consequences as stipulated in the Regulations Governing University Examinations. In particular, plagiarism, being a kind of dishonest practice, is "the presentation of another person's work without proper acknowledgement of the source, including exact phrases, or summarised ideas, or even footnotes/citations, whether protected by copyright or not, as the student's own work". Students are required to strictly follow university regulations governing academic integrity and honesty.

(3) Students are required to submit writing assignment(s) using Turnitin.

(4) To enhance students' understanding of plagiarism, a mini-course "Online Tutorial on Plagiarism Awareness" is available on https://pla.ln.edu.hk/.

Rubric for Examination of CDS1001 – Introduction to Programming for Data Science

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Design small Python programs (LO3)	The designs are relevant, elegant, and effective. Moreover the program logics are correct.	The designs are not elegant or ineffective. But the logics of programs are correct.	The designs are basically ineffective. The logics of programs are incorrect.
Code small Python programs (LO3)	The program codes are nearly completely correct.	The program codes have some minor errors.	The program codes have some significant errors.
Accurately use procedural statements (LO4)	More than 80% of procedural statements are used correctly.	50% to 80% of procedural statements are used correctly.	Less than 50% of procedural statements are used correctly.
Manipulate and process data for solving problems using the logic of a programming language (LO5)	The right packages and methods are selected to handle the problems. They are used correctly to manipulate and process data.	The right packages and methods are selected to handle the problems. There are some minor issues in applying these packages and methods.	The student cannot select the right packages and methods to handle the problems, or , or uses them incorrectly.
Retrieve data from different sources (LO6)	The right packages and methods are used. They are used correctly to retrieve data.	The right packages and methods are used. But there are some minor issues in applying these packages and methods.	The student cannot select the right packages and methods, or , or uses them incorrectly.
Store data to different sources (LO6)	The right packages and methods are used. They are used correctly to store data.	The right packages and methods are used. But there are some minor issues in applying these packages and methods.	The student cannot select the right packages and methods, or , or uses them incorrectly.

Rubric for Mid-Term Examination of CDS1001 – Introduction to Programming for Data Science

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Identify the fundamental concepts of data science (LO1)	The student demonstrates a clear understanding of the fundamental concepts of data science. The student can correctly identify most of these concepts.	The student demonstrates a reasonable understanding of the fundamental concepts of data science. The student can correctly identify some or most of these concepts.	The student demonstrates a limited understanding of the fundamental concepts of data science. The student can correctly identify few of these concepts.
Identify the applications of data science (LO1)	The student demonstrates a clear understanding of the applications of data science. The student can correctly list most of these applications.	The student demonstrates a reasonable understanding of the applications of data science. The student can correctly list some or most of these applications.	The student demonstrates a limited understanding of the applications of data science. The student can correctly list few of these applications.
Describe the whole data science process (LO2)	The student can describe all procedures in the whole process correctly.	The student can describe most procedures in the whole process correctly.	The student can only describe 0-3 procedures in the whole process correctly.
Design small Python programs (LO3)	The designs are relevant, elegant, and effective. Moreover the program logics are correct.	The designs are not elegant or ineffective. But the logics of programs are correct.	The designs are basically ineffective The logics of programs are incorrect.
Code small Python programs (LO3)	The program codes are nearly completely correct.	The program codes have some minor errors.	The program codes have many errors.
Accurately use procedural statements (LO4)	More than 80% of procedural statements are used correctly.	50% to 80% of procedural statements are used correctly.	Less than 50% of procedural statements are used correctly.

Rubric for Assignments of CDS1001 – Introduction to Programming for Data Science

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Design small Python programs (LO3)	The designs are relevant, elegant, and effective. Moreover the program logics are correct.	The designs are not elegant or ineffective. But the logics of programs are correct.	The designs are basically ineffective. The logics of programs are incorrect.
Code small Python programs (LO3)	The program codes are nearly completely correct.	The program codes have some minor errors.	The program codes have many errors.
Test small Python programs (LO3)	The programs are tested thoroughly. The testing procedure is correct	The programs are not thoroughly tested. There are minor errors in the testing procedure	The programs are not being tested, No results have been generated. There are major errors in the testing procedure.
Use procedural statements (LO4)	More than 90% of procedural statements are used correctly.	70% to 90% of procedural statements are used correctly.	Less than 70% of procedural statements are used correctly.
Manipulate and process data for solving problems using the logic of a programming language (LO5)	The right packages and methods are selected to handle the problems. They are used correctly to manipulate and process data.	The right packages and methods are selected to handle the problems. There are some minor errors in applying these packages and methods.	The student cannot select the right packages and methods to handle the problems, or , or uses them incorrectly.
Retrieve data from different sources (LO6)	The right packages and methods are used. They are used correctly to retrieve data.	The right packages and methods are used. But there are some minor issues in applying these packages and methods.	The student cannot select the right packages and methods, or , or uses them incorrectly.
Store data to different sources (LO6)	The right packages and methods are used. They are used correctly to store data.	The right packages and methods are used. But there are some minor errors in applying these packages and methods.	The student cannot select the right packages and methods, or , or uses them incorrectly.
Plot data using Python visualization libraries (LO7)	The right packages and methods are used. They are used correctly to visualize data.	The right packages and methods are used. But there are some minor errors in applying these packages and methods.	The student cannot select the right packages and methods, or , or uses them incorrectly.

Rubric for Class Attendance and Participation of CDS1001 – Introduction to Programming for Data Science

Criteria	Very good (4-6)	Satisfactory (2-4)	Unsatisfactory (0-2)
Attendance	Full, punctual attendance in class and mandatory seminars.	Occasional absences or lateness from class or mandatory seminars.	Frequent or recurring absence or lateness from class or mandatory seminars.
Behaviour	Displays mature behavior; often goes beyond expectations; exemplary adherence to rules.	Rarely displays improper or disruptive behavior; respects rules and boundaries.	Displays improper or disruptive behaviour; ignores rules and boundaries.
Participate in class discussion	The students is always being active during class discussions.	The student is sometimes being active during class discussions.	The student keep silent during class discussions.
Enthusiasm in the subject	The student always respond to teacher's questions, or raise questions or provide further examples related to the topic of interest.	The student sometimes respond to teacher's questions, or raise questions, or provide further examples related to the topic of interest.	The student seldom or never respond to teacher's questions, nor raise questions, nor provide further examples related to the topic of interest.
Quality of comments	Comments are always insightful & constructive. Comments balanced between general programming knowledge, opinions & specific, thoughtful criticisms or contributions.	Comments are sometimes constructive, with occasional signs of insight. Comments not always relevant to the discussion.	Comments are uninformative. Heavy reliance on personal opinion & personal feeling