

SDSC2001: PYTHON FOR DATA SCIENCE

Effective Term

Semester A 2024/25

Part I Course Overview

Course Title

Python for Data Science

Subject Code

SDSC - School of Data Science

Course Number

2001

Academic Unit

School of Data Science (DS)

College/School

School of Data Science (DS)

Course Duration

One Semester

Credit Units

4

Level

B1, B2, B3, B4 - Bachelor's Degree

Medium of Instruction

English

Medium of Assessment

English

Prerequisites

CS2311/CS2315 Computer programming / CS1315 Introduction to Computer Programming and SDSC1001 Introduction to Data Science*

* Pre-requisite SDSC1001 will be exempted for students who are enrolled in Minor in Data Science

Precursors

Nil

Equivalent Courses

Nil

Exclusive Courses

Nil

Part II Course Details

Abstract

This course provides students with an extensive exposure to the use of Python specifically for data science. Topics include Python language fundamentals, data analysis using Python libraries, applied machine learning and text mining in Python, and the practice of scientific computing. The students will acquire hands-on experience using Python and the popular packages related to data manipulation, processing and analysis, with the minimal backgrounds of theories in methodology aspects. The students are required to be able to independently develop Python codes from scratch in professional way for elementary algorithms in data sciences.

Course Intended Learning Outcomes (CILOs)

CILOs		Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Explain fundamentals of Python programming language and version-control tools including GitHub	20	x		
2	Utilize Python libraries for basic data representation, manipulation, pre-processing, analysis and visualization.	30	x	x	
3	Implement algorithms for fundamental methods in data analysis using Python libraries.	20	x	x	x
4	Demonstrate Python programming skills by conducting exploratory data analysis to solve practical problems in case studies.	30	x	x	x

A1: Attitude

Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability

Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: Accomplishments

Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

Learning and Teaching Activities (LTAs)

LTAs	Brief Description	CILO No.	Hours/week (if applicable)	
1	Lecture	Students will engage in lectures to apply basic programming via Python and practice methods in data analysis examples using computers.	1, 2, 3, 4	3 hours per week

2	Tutorial	Students will engage in tutorial activities to practice Python programming methods and implement algorithms in case studies.	2, 3, 4	1 hour per week
---	----------	--	---------	-----------------

Assessment Tasks / Activities (ATs)

ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)	
1	Assignments	2, 3, 4	25	Students will practice basic Python programming skills and utilize Python libraries to solve practical problems and demonstrate their understanding of concepts and methods for data science.
2	Course Project	2, 3, 4	25	In the course project, students will demonstrate their Python programming skills by implementing fundamental data analysis methods in real-world data analysis.

Continuous Assessment (%)

50

Examination (%)

50

Examination Duration (Hours)

2

Additional Information for ATs**Examination:**

The students will utilize Python libraries and write Python programming codes to perform basic data manipulation, analysis and visualization tasks and solve practical problems in order to demonstrate their programming skills.

Note: To pass the course, apart from obtaining a minimum of 40% in the overall mark, a student must also obtain a minimum mark of 30% in both continuous assessment and examination components.

Assessment Rubrics (AR)**Assessment Task**

Assignments

Criterion

1.1 Ability to discuss basic concepts and skills of Python language, and to write computer programs with the aid of common Python libraries for data representation, manipulation, analysis and visualization.

1.2 Capability to apply popular machine learning Python libraries for data analysis.

Excellent (A+, A, A-)

High

Good (B+, B, B-)

Significant

Fair (C+, C, C-)

Moderate

Marginal (D)

Basic

Failure (F)

Not even reaching marginal levels

Assessment Task

Course Project

Criterion

Ability to develop, validate and publish Python code to solve real problems by using Python libraries and implementing elementary data analysis methods.

Excellent (A+, A, A-)

High

Good (B+, B, B-)

Significant

Fair (C+, C, C-)

Moderate

Marginal (D)

Basic

Failure (F)

Not even reaching marginal levels

Part III Other Information

Keyword Syllabus

Basics of Python programming language:

Installation and setup of Python and Python Packages (shell and interactive environment); IPython and Jupyter Notebook; basics of Python language; data structures in Python; file operations; functions in Python; OOP; data processing; profiling and timing; version control skills (github)

Introduction to Python for scientific computing, data processing and plotting:

Scientific computing with NumPy (multidimensional array, indexing and slicing, array functions, pseudo-random generators); data manipulation with Pandas (series, dataframe, selection, filtering, sorting, ranking); visualization with Matplotlib.

Python for elementary machine learning (Scikit-Learn):

Pre-processing; feature extraction; simple linear regression analysis; binary classification analysis; decision tree and random forest; k nearest neighbour; model selection.

Reading List

Compulsory Readings

Title	
1	Python Data Science Handbook: Essential Tools for Working with Data, by Jake VanderPlas, O' REILLY 2016

Additional Readings

Title	
1	Data Analysis from Scratch with Python: Beginner Guide for Data Science, Data Visualization, Regression, Decision Tree, Random Forest, Reinforcement Learning, Neural Network and NLP using Python By Peter Morgan Publisher: AI Sciences LLC; 2 edition (August 13, 2018)
2	Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 2nd Edition By Wes McKinney O'Reilly Media; 2 edition (September 25, 2017)