DATA 100 Course Syllabus

DATA 100 001 - Introduction to Data Science in Python (3)

Course Description

DATA 100 (3) Introduction to Data Science in Python (3)

Fundamentals of data science with an emphasis on computational thinking, testing, debugging, and working with data sets. Real-world applications from disciplines in the sciences, humanities, medicine, engineering, social sciences, and others. No prior computing background is required.

Prerequisite: None

Equivalence: COSC 100

Learning Outcomes

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

1. develop the ability to "think computationally" using programming principles.
2. practice the creation of loops, conditionals, and functions to analyze data.
3. identify and use different data types to accomplish a variety of data science tasks.
4. apply common workflows to load, process, clean, and analyze data.
5. appraise the quality of data and assess its limitations in answering questions.
6. understand the role of testing and version control to writing sustainable code.
7. create reproducible, ethical, and sustainable data analyses.
8. apply the skills and techniques in this course to generate reproducible analyses.

Assessment

Item	Weight	Frequency
Learning Logs	10%	Weekly
Labs	30%	Weekly
Project	30%	Weekly
Tests	30%	Bi-weekly

Passing Criteria

All students must satisfy ALL conditions to pass the course:

1. Pass the Labs with an average grade of at least 50%, with no more than 4 missed labs.
2. Pass the Tests with an average grade of at least 50%.
3. Pass the Project with a grade of at least 40%.
4. Pass the Course overall with a grade of at least 50%.

If a student does not satisfy the appropriate requirements, the student will be assigned the **lower** of their earned course grade or, a maximum overall grade of 45 in the course.

Textbook

Portions of the following (open source) textbooks will be assigned as reading:

- Python Data Science Handbook, by Jake VanderPlas
- Python for Data Analysis, by Wes McKinney

Eventually, an open textbook will be developed using open resources.

Schedule

Wk	Starting	Topics	Project	Lab	Learning Logs	Tests
1	Week 1	Introduction to Data Science			LL 1	
2	Week 2	Terminal and Jupyter Notebook		L1	LL 2	Test 1
3	Week 3	Version Control with Git	PM1	L2	LL 3	
4	Week 4	Introduction to Python		L3	LL 4	Test 2
5	Week 5	Loading and working with data	PM2	L4	LL 5	
6	Week 6	Data Types: Lists and Dictionaries		L5	LL 6	Test 3
7	Week 7	Computation with `numpy`	PM3	L6		
8	Week 8	Controlling the flow			LL 7	Test 4
9	Week 9	Organizing your code	PM4	L7	LL 8	
10	Week 10	Objects in Python			LL 9	Test 5
11	Week 11	Data analysis with `scipy` and `pandas`	PM5	L8	LL 10	
12	Week 12	Data visualization		L9	LL 11	
13	Week 13	Releases and Reproducibility	PM6	L10	LL 12	