UNIVERSITY of NICOSIA

Course Syllabus

Course Code	Course Title	ECTS Credits
COMP-240	Data Programming	6
Prerequisites	**Department**	**Semester**
COMP-111	Computer Science	Fall
Type of Course	**Field**	**Language of Instruction**
Required	Data Science	English
Level of Course	**Lecturer(s)**	**Year of Study**
1st Cycle	Dr. D. Trihinas	2nd
Mode of Delivery	**Work Placement**	**Corequisites**
Face-to-face	N/A	N/A

Course Objectives:

The main objectives of the course are to:

- Present fundamental programming concepts, data abstractions and computing techniques for Data Science.
- Introduce a problem-solving framework for “thinking with data” that targets the scientific treatment for data curation (e.g., data harvesting, preparation and manipulation).
- Present intuitive data types and effective data structuring techniques to speed data curation.
- Introduce programming paradigms (e.g., procedural, object-oriented, array-oriented, functional) suitable for Data Science projects and in which contexts are these paradigms embraced.
- Demonstrate how to leverage the Python programming language to develop well-structured and maintainable applications that apply data programming principles.
- Introduce the basics of statistical modelling and inference to simulate deterministic and stochastic phenomena found in Data Science projects.
- Present basic techniques for measuring the performance of data intensive applications and showcase methods to improve the efficiency of data manipulation.
- Demonstrate basic statistical modelling and plotting techniques to communicate insights.

UNIC

Learning Outcomes:

After completion of the course students are expected to be able to:

- Understand the basic concepts involved in computational thinking towards solving data-oriented problems.
- Distinguish between suitable data types and structuring techniques for different data programming challenges.
- Develop well-structured and maintainable programs in Python.
- Distinguish between suitable programming paradigms and evaluate strengths and weaknesses of each paradigm depending on the context of different Data Science tasks.
- Apply in practice the principles for pre-processing datasets and for indexing data to improve the efficiency of data trawling and extraction.
- Understand the algorithmic process for cleaning impure datasets and acknowledge the scientific treatments for dealing with missing data.
- Apply in practice the principles for combining, aggregating and summarizing datasets.
- Acknowledge how the data collection process influences the scope of inference.
- Comprehend basic methods towards plotting data in order to present insights extracted from datasets.
- Be familiar with the methods for measuring and optimizing the performance of data-intensive tasks to cope with large volumes of data

Course Content:

1. Programming for Data Science with Python
    a. Conditional Logic
    b. Iterative Programming
    c. Functional Decomposition
    d. Error Handling
    e. Data-Oriented Problem Solving
2. Associative Data Types and Structures
    a. Lists
    b. Dictionaries
    c. Tuples
    d. Sets
3. Interactive Programming Environments
    a. Jupyter Notebooks
    b. Markup
    c. Version Control
4. Arrays and Vectorized Computation
    a. Multi-dimensional Array Objects (e.g., NumPy)

b. Array-Oriented Programming
c. Vectorization for Arithmetic Operations
d. Array Broadcasting

5. Tabular and Heterogeneous Data
    a. Tabular Objects (e.g., DataFrames)
    b. Hierarchical Data Indexing
    c. Data Grouping, Aggregation and Summarization
6. Classes and Objects
    a. Data Abstraction and Encapsulation
    b. Inheritance
    c. Polymorphism
7. Dealing with Data Sources
    a. Strings and Encoding Schemes
    b. Regular Expressions
    c. File I/O
    d. Web Scraping
8. Data Pre-Processing
    a. Data Cleaning
    b. Lambda Transformations and Functional Programming
    c. Attribute Scaling
    d. Categorical Data Encoding
9. Data Wrangling
    a. Data Reshaping and Pivoting
    b. Sampling and Filtering
    c. Data Grouping and Aggregation
10. Statistical Modeling and Plotting
    a. Descriptive Statistics
    b. Simulation-Based Experimentation
    c. Basic Plotting
    d. Discovering Data Correlations

Learning Activities and Teaching Methods:

Lectures, In Class Exercises, Lab Sessions, Software Tool Tutorials, Case-Study Presentations, Discussions.

Assessment Methods:

Homework Assignments, Midterm Exam, Participation, Final Exam

Required Textbooks / Readings:

Title	Author(s)	Publisher	Year	ISBN
Python for Data Analysis* (3rd edition)	Wes McKinney	O'Reilly Media	2022	978-1-098-10403-0
Think Python** (2nd edition)	Allen B. Downey	Green Tree Press	2016	978-1-491-93936-9

* Also available as an open access e-book by the author https://wesmckinney.com/book/

** A free pdf version is available by the author https://greenteapress.com/wp/think-python-2e/

Recommended Textbooks / Readings:

Title	Author(s)	Publisher	Year	ISBN
Computational and Inferential Thinking: The Foundations of Data Science (2nd edition)	Ani Adhikari, John DeNero and David Wagner	OpenLibra	2022	9124423963
Python Data Science Handbook	Jake VanderPlas	O'Reilly Media	2016	978-1-491-91205-8
Doing Data Science	Cathy O'Neil and Rachel Schutt	O'Reilly Media	2014	978-1-449-35865-5