

## Topics in Computational Statistics: Data Engineering

**Instructor:** *K.P.PANOUSIS*

Course Code: 61235

Course Type: Elective of Course Group 2

Course Level: Graduate (MSc)

Year of Study: A'

Semester: 2<sup>nd</sup>

ECTS: 3

Language: English

### Course Description

The course lays proper foundations in Data Engineering with emphasis on Statistical and Data Science applications. We will begin with basic SQL concepts such as queries, joins and aggregation as well as fundamental databases and in-memory analytics notions. In addition, we will introduce standard Python syntax, modules, data types and structures, operations control flow and input/output operations and database connectivity. Then, topics from the entire data lifecycle will be covered including data ingestion, transformation, loading, visualization, modeling, deployment, update, monitoring, maintenance, and documentation. More advanced or modern subjects with software engineering flavor such as parallelism, cloud computing, error handling, testing and version control will also be sampled for exposition.

### Prerequisites

None

### Target Learning Outcomes

Upon completion of the course, students will

- Understand the principles of data engineering and their significance on statistical analysis.
- Have a solid foundation in SQL and Python programming.
- Comfortably perform standard data engineering tasks with focus on statistical modeling.
- Have hands on experience on in-memory computing technologies.

### Recommended Bibliography

- McKinney, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.", 2012.
- Fangoehr, Hans. Python for Computational Science and Engineering. Online Resource 2022

- Crickard, Paul. *Data Engineering with Python: Work with massive datasets to design data models and automate data pipelines using Python*. Packt Publishing Ltd, 2020.
- Beaulieu, Alan. *Learning SQL: master SQL fundamentals*. " O'Reilly Media, Inc.", 2009.
- Tanimura, Cathy. *SQL for Data Analysis*. " O'Reilly Media, Inc.", 2021.

### **Teaching and Learning Activities**

The course is delivered in six weekly 3-hour lectures. Every week homework will be assigned (some to be submitted). A team project will also be assigned.

### **Assessment and Grading Methods**

The final grade is the weighted average of the final examination grade (50%) and the assignment/projects (50%).