

Foundations of Data Science

3 ECTS

Overview and Objectives

This is an intensive 20-hour course based on a hands-on approach using Jupyter notebooks, all material is motivated by specific information retrieval and data analysis questions and each thematic unit concludes with a small project.

The course will be delivered by Omiros Papaspiliopoulos, ICREA Research Professor and Director of the BGSE Data Science Center, in collaboration with Data Scientists affiliated to the Data Science Center.

Course Outline

The course provides training in information retrieval, data analysis and machine learning with Python and R and evolves along the thematic following units:

1. Programming with Python

Keywords: data types, functions, objects

2. Data analysis with Python

Keywords: pandas, database management, seaborn

3. Scraping data

Keywords: scraping, API's, cloud storage

4. Supervised learning

Keywords: sklearn, linear models, cross validation, regularisation, lasso, trees, ensembles, boosting

4. Unsupervised learning

Keywords: clustering, factors, independent component analysis and matrix factorization

5. Crash course in R