

MODULE SPECIFICATION • CORE

Foundations of Data Science

This course offers a first overview of data science techniques at the graduate level. It takes participants through basic data types and properties, supervised learning techniques, and unsupervised learning methods. The course also addresses prediction and classification techniques using parametric, non-parametric and ensemble methods. Participants understand how to approach data analysis problems, which tools are available to them, and how to address common problems faced in data science applications.

20	25h	175h	200h
CREDITS	CONTACT HOURS	PRIVATE STUDY	TOTAL STUDY

COURSE CONVENOR

NAME

Dr Arthur Charpentier

OFFICE

By appointment — please email to arrange a convenient time

COURSE STRUCTURE

Econ 619 is a 20-credit course and students are expected to input approximately 200 hours of study. The total number of contact hours is 25 hours, leaving 175 hours for private study. Delivery is by lectures — 25 hours across the first five weeks of term, comprising 20 hours of lectures and 5 hours of tutorials. Optional clinics run on the last day of the course.

During private study, balance the course material (the primary source) against the recommended textbooks: think critically about how these fit into the body of knowledge, perform the exercises, complete coursework and revise. You can expect to perform well only by working consistently through the year.

Prerequisite. Must have successfully completed ECON 7901.

MODULE PROGRAMME

LECT.	TOPIC	SESSION CONTENT
01	Introduction to Data Science	- What is data science and why is it important? - The role of data in decision-making - The principles of statistical thinking
02	Data types	- Importing and cleaning data using popular tools and techniques - Basic data visualisation using graphs and plots - Introduction to the Pandas library for data manipulation
03	Statistics and Machine Learning Basics	- Supervised vs. unsupervised learning - Important data science methods for prediction and classification - Basic machine learning algorithms and their applications
04	Model Evaluation and Selection	- Measuring the performance of machine learning models - Choosing the appropriate model for a given problem - Avoiding common pitfalls in model evaluation and selection

KEY SKILLS

By the end of this course, students should have some knowledge and understanding of:

- Data fundamentals.** Handling diverse data types and understanding their properties.
- Machine learning.** Applying supervised and unsupervised learning algorithms.
- Prediction and classification.** Using various methods for accurate predictions and classifications.
- Analytical approach.** Systematic problem-solving for data analysis.
- Tool proficiency.** Competence in common data science tools for analysis and visualisation.
- Problem-solving.** Addressing challenges in real-world data science applications.
- Ensemble methods.** Application of ensemble methods for improved predictions.
- Communication skills.** Effectively conveying insights from data analysis.

DESIRED OUTCOMES

By the end of this course, students should be able to:

- Demonstrate a sound knowledge of applied econometric principles and basic quantitative techniques
- Demonstrate a sound knowledge of supervised and unsupervised learning techniques
- Present, interpret and analyse information in numerical form, and use econometric and other packages effectively
- Understand the relevance of different econometric approaches to specific applications in economics
- Select relevant information from large amounts of data

ASSESSMENT

WEIGHTING	100% coursework
ISSUED	End of week 6
DUE	End of week 10 — four weeks for completion



SUBMISSION	Electronically via the Moodle site for this course
FEEDBACK	Returned within 4 weeks of the deadline, as marker's notes appended to the PDF of your coursework

Marking criteria and penalties are set out in the Economics Undergraduate Handbook and the general course information paper, available via the Current Student page of the university website under Academic Regulations.

COURSE TEXT & RECOMMENDED READING

Book chapters and online material. Copies of the lecture slides are made available on the course web pages, together with additional material and course announcements.

- Kauermann, G., Küchenhoff, H. and Heumann, C. Statistical Foundations, Reasoning and Inference: For Science and Data Science. Springer Nature
- Phillips, J. M. Mathematical Foundations for Data Analysis. Springer Nature
- Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly