

ECU33143 — Introduction to Big Data in Economics

Module Information	
Module Code	ECU33143
Module Title	Introduction to Big Data in Economics
Department	ECONOMICS
Year Group	Junior Sophister
Module Personnel	Dr. Jian Cao
Academic Year	2026-27
Credit Weighting	5 credits
Semester	Semester 1
Contact Hours	20 hours of lectures and 5 hours of tutorials

Module Content	
Module Description	This module introduces the field of Big Data, examining its main features, such as real-time availability, scale, types of variables and structure. It focuses on showing how different types of Big Data, including administrative datasets, high-frequency information, web-scraped datasets, can be used to answer questions in economics. Topics covered in this module will include: • Understanding Big Data in Economic Contexts: Key characteristics of "big data" • Challenges and opportunities specific to "Big data" such as: ○ Privacy & confidentiality ○ Data management practices ○ Scalable approach • Data Collection: Strategies & Challenges • Data Analysis: Techniques & Tools for Big Data • Applying Big Data: Case Studies in Economics

Module Content	
Module Learning Aims	This module will introduce students to the main topics that have been arising around the concept of “Big Data”. With increasing computing and storage capacity, new opportunities and challenges arise for researchers. The goal of this module is to provide an overview of different datasets, techniques and applications, with an emphasis on practical implementation.
Learning Outcomes	On successful completion of the module students should be able to: • Explain the challenges and opportunities arising from the “Big Data”. • Distinguish which type of data to collect based on the typology of empirical questions to tackle and how to efficiently collect them. • Apply the latest technique to analyze and summarize large amount of data. • Apply Big Data methodologies such as Data Mining and High-Dimensional Data Analysis in different research settings.
Indicative Reading / Resources	Taddy, Matt. Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions. McGraw Hill, 2019. Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani. An Introduction to Statistical Learning. Springer, 2021.

Assessment	
Assessment Components	Assignments 10%, Project Report 20%, Final Exam 70% (Module Concepts 35% + Project Defence 35%)
Assessment Details	• Assignments (10%): regular problem sets. • Project Report (20%): a self-selected project applying course methods. • Final Exam (70%), in two equally weighted halves: ○ Module Concepts (35%): questions on the concepts and techniques taught across the module. ○ Project Defence (35%): questions on your own project: the choices you made, the methods you used, and how you would interpret or extend your results. The aim is to confirm the work is your own.

Additional Information	
Prerequisites	ECU22031 & ECU22032 Mathematical and Statistical Methods
Co-requisites	None
Any other information	None