**School of Economics and Management**

# DABN23, Data Analytics and Business Economics: Data Retrieval and Preparation, 7.5 credits

*Dataanalys och ekonomi: Insamling och förberedelse av data, 7,5 högskolepoäng*

**Second Cycle / Avancerad nivå**

## Details of approval

The syllabus was approved by The Board of the Department of Economics on 2025-11-06. The syllabus comes into effect 2025-11-06 and is valid from the spring semester 2026.

## General information

This is a single subject master course in data analytics and business economics. The course is an elective in the master programme Data Analytics and Business Economics. The course is optional within a number of master programmes at Lund University.

*Language of instruction:* English
(Teaching may be in Swedish if all registered students have a good knowledge of Swedish.)

| *Main field of study* | *Specialisation* |
|---|---|
| Data Analytics and Business Economics | A1N, Second cycle, has only first-cycle course/s as entry requirements |

## Learning outcomes

### Knowledge and understanding

Students shall have an understanding of:

- Python- and web-based Application Programming Interfaces (APIs),

- work flows for web scraping projects,

- popular Python libraries for web scraping and their functionality,

- exploratory analysis to guide feature engineering,

- encoding and transforming predictors,

- detecting and modeling interactions,

- handling missing data and outliers,

- feature selection goals, methods, and pitfalls.

## Competence and skills

Students shall have the ability to independently:

- use Python APIs for data collection,

- parse complicated HTML pages,

- write web crawlers in Python,

- plan and execute a workflow for feature engineering,

- engineer predictors for modeling performance,

- detect and model interactions using appropriate search strategies,

- handle missing data systematically.

## Judgement and approach

Students shall have the ability to pursue further studies in the subject and should be able to search for and evaluate subject related information with a high degree of independence. Students shall also have sufficient competence to write an empirical report where machine learning is an essential element.

## Course content

The first half of the course focuses on modern data retrieval from the web. Students learn the legal and ethical boundaries of scraping, the basics of HTTP and developer tools, and how to build robust scrapers in Python using HTML parsing, site-scale crawlers, and automation for forms, logins, and JavaScript-driven pages. We also cover reading common document formats (CSV, PDF, Word), working with public and undocumented APIs and storing results reliably.

The second half develops a rigorous workflow for data preparation and feature engineering in predictive modeling. We begin with problem framing and proper data partitioning, then move to preprocessing steps such as transformation, scaling, correlation filtering, and other diagnostics that are fit on training data and applied to assessment/test data. Building on that foundation, we emphasize exploratory visualization—of responses, predictors, and their relationships—to guide feature design and model diagnostics before and after initial fits. The core of this module is hands-on feature engineering and selection: encoding categorical predictors (from basic dummy variables to approaches for high-cardinality features), engineering numeric predictors via 1:1 transforms (e.g., Box-Cox/Yeo-Johnson), basis expansions and splines, many-to-many projections, detecting and creating interaction terms, and robust approaches to missing data and outliers. We conclude with a unifying overview of feature selection and its main classes (intrinsic, filter, wrapper) with an emphasis on avoiding overfitting.

## Course design

1. Teaching: Teaching consists of lectures and exercise classes.

## Assessment

1. Examination: The examination consists of a written exam, active participation in lectures and exercise sessions, and home assignments and other assignments during the course. Points gained on the home assignments count at exams during the current term. Other forms of examination may be used to a limited extent.

2. Limitations on the number of examination opportunities: –

The University views plagiarism and other academic dishonesty very seriously, and will take disciplinary action against students for any kind of attempted malpractice in connection with examinations and assessments. Plagiarism is considered to be a very serious academic offence. The penalty that may be imposed for this, and other unfair practices in examinations or assessments, includes suspension from the University for a specified period.

The examiner, in consultation with Disability Support Services, may deviate from the regular form of examination in order to provide a permanently disabled student with a form of examination equivalent to that of a student without a disability.

## Grades

Grading scale includes the grades: U=Fail, E=Sufficient, D=Satisfactory, C=Good, B=Very Good, A=Excellent
1. Grade (Definition), Points or percentage out of maximum points, Characteristic
A (Excellent) A distinguished result that is excellent with regard to theoretical depth, practical relevance, analytical ability, and independent thought.
B (Very good) A very good result with regard to the above-mentioned aspects.
C (Good) The result is of a good standard with regard to the above-mentioned aspects and lives up to expectations.
D (Satisfactory) The result is of a satisfactory standard with regard to the above-mentioned aspects and lives up to expectations.
E (Sufficient) The result satisfies the minimum requirements with regard to the above-mentioned aspects, but not more.
U (Fail) The result does not meet the minimum requirements with regard the above-mentioned aspects.
To pass the course, the student must have been awarded the grade of E or higher.

2. Weighting grades from different parts of the course: –

3. Grading scales for different parts of the course: –

## Entry requirements

Students admitted to the master programme Data Analytics and Business Economics are eligible for this course. Students admitted to the Master Programme in Economics with at least 30 ECTS-credits in economics at the advanced level including Advanced Econometrics are eligible for the course. For other students, a Bachelor degree including at least 30 ECTS-credits in statistics of which 7.5 ECTS-credits in econometrics or regression analysis, or a Bachelor degree in economics or business administration with at least 15 ECTS-credits in statistics of which 7.5 ECTS in econometrics or regression analysis is required.

# Further information

1. Transitional regulations: –

2. Limitations in the period of validity: –

3. Limitations: –

4. Similar courses: –

5. Limitations in renewed examination: –

This is a translation of the syllabus
which was established in Swedish.