1. BASIC INFORMATION

Course	Data Processing
Degree program	Masters in Big Data Analytics
School	Escuela de Arquitectura, Ingeniería y Diseño
Year	1
ECTS	6
Credit type	Compulsory
Language(s)	English
Delivery mode	Face-to-face / Online
Semester	1
Academic year	2025-2026
Coordinating professor	Oscar Marbán Gallego

2. PRESENTATION

This compulsory 6 ECTS course prepares students to work on data analysis projects using large volumes of data. Students will use foundational Python programming skills to deepen their understanding of this language in various analysis and processing tasks, as well as to work with industry-standard tools.

Students will be trained in data ingestion and processing, and will acquire knowledge of the entire data preprocessing lifecycle for machine learning—from ingestion to processing and preparation—along with examples of its application using machine learning models.

3. LEARNING OUTCOMES

Knowledge
CON04. Understand the fundamentals of distributed processing and storage of large volumes of data.

Skills
HAB05. Design a data extraction, transformation, and loading (ETL) process using parallel processing and scalable resources.

Competences
CP04. Design and execute a complete knowledge discovery process, including the phases of data storage, processing, and visualization.
CP05. Apply data analysis techniques and both supervised and unsupervised machine learning models, even when the available information is incomplete or limited.

4. CONTENT

The course is organized into learning units, which are further divided into topics:

Unit 1. Introduction to Data Analysis
- Introduction to Data Analysis
- Data in the Real World
- Data Formats
- Data Preparation as a Process
- Stages in Data Preparation

Unit 2. Data Ingestion
- Discovery and Characterization
- Sampling and Variability
- Confidence
- Data Ingestion with Pandas

Unit 3. Data Preparation
- Missing Values
- Anomalies and Noise
- Data Normalization and Transformation
- Time Series Preparation

Unit 4. Exploratory Data Analysis
- Analytical Visualization
- Dataset Dimensionality and Sampling
- Data Anonymization

Unit 5. Feature Engineering
- Feature Engineering (I)
- Feature Engineering (II)
- Feature Importance
- Feature Selection

Unit 6. Document and Text Analysis
- Introduction to Text Analysis
- Text Representation and Access
- Document Preprocessing

5. TEACHING-LEARNING METHODOLOGIES

The following types of teaching-learning methodologies will be applied:

- MD1: Face-to-face lecture or, if applicable, through the virtual campus
- MD2: Cooperative learning
- MD3: Problem-Based Learning (PBL)
- MD4: Group research (a) and/or (b) group problem-solving
- MD6: Simulation environments and field experiences

6. LEARNING ACTIVITIES

Listed below are the types of learning activities and the number of hours the student will spend on each one:

Campus-based mode:

Learning activity	Number of hours
Masterclasses	20
Individual Activities	35
Integrative Activities	15
Lab activities	10
Autonomous work	68
Exams and tests	2
TOTAL	**150**

Virtual mode:

Learning activity	Number of hours
Masterclasses	20
Individual Activities	35
Integrative Activities	15
Lab activities	10
Autonomous work	68
Exams and tests	2
TOTAL	**150**

7. ASSESSMENT

Listed below are the assessment systems used and the weight each one carries towards the final course grade:

Campus-based mode:

Assessment system	Weight
SE1: Exams	50%
SE2: Report writing	30%
SE3: Practical Use Cases	20%

Virtual mode:

Assessment system	Weight
SE1: Exams	60%
SE2: Report writing	25%
SE3: Practical Use Cases	15%

When you access the course on the *Campus Virtual*, you'll find a description of the assessment activities you have to complete, as well as the delivery deadline and assessment procedure for each one.

7.1. First exam period

To pass the course in the first exam period, you must obtain a final course grade of at least 5 out of 10 (weighted average).

In any case, you will need to obtain a grade of 4.0 in the final exam in order for it to count towards the final grade along with all the grades corresponding to the other activities.

7.2. Second exam period

To pass the course in the second exam period, you must obtain a final grade of at least 5 out of 10 (weighted average).

In any case, you will need to obtain a grade of at 4.0 in the final exam in order for it to count towards the final grade along with all the grades corresponding to the other activities.

The student must deliver the activities not successfully completed in the first exam period after having received the corresponding corrections from the professor, or those that were not delivered in the first place.

8. SCHEDULE

This table shows the delivery deadline for each assessable activity in the course:

Assessable activities	Deadline
Activity 1	Week 5
Activity 2	Week 7
Activity 3	Week 9
Activity 4	Week 11

Ordinary call	At the end of subject
Extraordinary call	First weekend of September

This schedule may be subject to changes for logistical reasons relating to the activities. The student will be notified of any changes as and when appropriate.

9. BIBLIOGRAPHY

The main reference work for this subject is:

- García, S. Herrera, F. Luengo, J. (2015). Data Preprocessing in Data Mining. Springer.
- McKinney, W. (2017). Python for Data Analysis. Data Wrangling with Pandas, NumPy, and IPython. USA. O'Reilly.
- Kuhn, M. Johnson, K. (2019). Feature Engineering and Selection: A Practical Approach for Predictive Models. Chapman & Hall.
- Kuhn, M. Johnson, K. (2013). Applied Predictive Modeling. Springer.
- James, G. Witten D. Hastie T. y Tibshirani, R. (2017) Introduction to Statistical Learning: with applications in R. Springer.
- Towards data science. https://towardsdatascience.com/
- Saul Dobilas. https://solclover.com/
- Python.org. https://docs.python.org/es/3.8/tutorial/index.html

10. EDUCATIONAL GUIDANCE, DIVERSITY AND INCLUSION UNIT

From the Educational Guidance, Diversity and Inclusion Unit we offer support to our students throughout their university life to help them reach their academic achievements. Other main actions are the students' inclusions with specific educational needs, universal accessibility on the different campuses of the university and equal opportunities.

From this unit we offer to our students:

1. Accompaniment and follow-up by mean of counselling and personalized plans for students who need to improve their academic performance.
2. In terms of attention to diversity, non-significant curricular adjustments are made in terms of methodology and assessment for those students with specific educational needs, pursuing an equal opportunity for all students.
3. We offer students different extracurricular resources to develop different competences that will encourage their personal and professional development.
4. Vocational guidance through the provision of tools and counselling to students with vocational doubts or who believe they have made a mistake in their choice of degree.

Students in need of educational support can write to us at:

orientacioneducativa@universidadeuropea.es

11.   ONLINE SURVEYS

Your opinion matters!

The Universidad Europea encourages you to participate in several surveys which help identify the strengths and areas we need to improve regarding professors, degree programs and the teaching-learning process.

The surveys will be made available in the "surveys" section on virtual campus or via e-mail.

Your assessment is necessary for us to improve.

Thank you very much for your participation.