COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI ICADE CIHS

**Syllabus**
**2024 - 2025**

# GENERAL INFORMATION

| Data of the subject | |
|---|---|
| **Subject name** | Data Acquisition |
| **Subject code** | DTC-IMAT-211 |
| **Mainprogram** | Bachelor's Degree in Mathematical Engineering and Artificial Intelligence |
| **Involved programs** | Grado en Ingeniería Matemática e Inteligencia Artificial [Second year] |
| **Level** | Reglada Grado Europeo |
| **Quarter** | Semestral |
| **Credits** | 4,5 ECTS |
| **Type** | Obligatoria (Grado) |
| **Department** | Department of Telematics and Computer Sciencies |

| Teacher Information | |
|---|---|
| **Teacher** | |
| **Name** | Juan Ignacio Carnicero Tresca |
| **Department** | Department of Telematics and Computer Sciencies |
| **EMail** | jicarnicero@icai.comillas.edu |
| **Teacher** | |
| **Name** | Francisco Gómez Martín |
| **EMail** | fgmartin@icai.comillas.edu |
| **Teacher** | |
| **Name** | Ignacio Villanueva Romero |
| **Department** | Department of Telematics and Computer Sciencies |
| **EMail** | ivillanueva@icai.comillas.edu |

# DESCRIPTION OF THE SUBJECT

## Contextualization of the subject

### Prerequisites

Python programming

## Course contents

## Contents

**Block 1: Data Extraction and Transformation**

- Automata Theory
- Regular Expressions

**Block 2: Data Cleaning and Quality**

- Data Integrity and Quality
- Cleaning and Normalization
    - Encoding Management, Date Manipulation, etc.
- Data Imputation

**Block 3: Data Organization**

- Description of Primary File Types (txt, JSON, XML, CSV)
- Data Conversion Processes

**Block 4: Information Storage Files**

- Analysis of Various File Types (PDF, XLS, DOC)
- Import/Export Processes

**Block 5: The Web as a Data Source**

- HTML Language: Tags and DOM Structure
- Web Scraping Tools and Libraries
- Introduction to HTTP
- Introduction to APIs
- Automation of Data Retrieval Processes from the Web

## EVALUATION AND CRITERIA

| Evaluation activities | Evaluation criteria | Weight |
|---|---|---|
| **Exams:**<br>• Intersemester Exam (15%)<br>• Final Exam (50%) | • **Intersemester Exam (15%)**: Assessment of the knowledge acquired in automata theory, regular expressions, and data cleaning through problem-solving using programming.<br>• **Final Exam (50%)**: Evaluation of computational and abstract thinking for problem-solving in data extraction, cleaning, organization, and storage through programming. | 65 % |
| **Final practice (20%)** | The knowledge acquired will be assessed weekly through an individual practical case that must be solved through programming. | 20 % |
| **Practice sessions (15%)** | • **Collaborative Methodology and Best Practices** Teamwork, following the use of | |

| | | |
|---|---|---|
| • Attitude, participation, and completion of the problems posed in collaborative and individual sessions.<br>• Teamwork<br>• Oral communication | collaborative best practices.<br>• **Classroom Work and Oral Communication**: Participation and completion of problems posed in sessions. Ability to orally present results obtained in practical cases. | 15 % |

## Grading

The final grade for the **regular** and **extraordinary** exam sessions for this course will depend on the evaluation of the following activities:

**Final Grade** = 15% Intersemester Exam + 50% Final Exam + 10% Weekly Assignments + 20% Final Project + 15% Collaborative Work and In-Class Attitude

This final grade will only be applied if a minimum grade of 5.0 is obtained in the Final_Exam.

Failure to attend 15% or more of the in-person hours for this course may result in the inability to participate in both the regular and extraordinary exam sessions.

## BIBLIOGRAPHY AND RESOURCES

### Basic References

Slides and code provided by the course instructors.