



GENERAL INFORMATION

Data of the subject	
Subject name	Data Acquisition
Subject code	DTC-IMAT-211
Main program	Grado en Ingeniería Matemática e Inteligencia Artificial
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 Sciences

Teacher Information	
Teacher	
Name	Ignacio Villanueva Romero
Department	Department of Telematics and Computer Sciences
E-Mail	ivillanueva@icai.comillas.edu
Teacher	
Name	Juan Ignacio Carnicero Tresca
Department	Department of Telematics and Computer Sciences
E-Mail	jjcarnicero@icai.comillas.edu

DESCRIPTION OF THE SUBJECT

Contextualization of the subject
Prerequisites
Python programming

Course contents

Contents
Block 1: Data Extraction and Transformation <ul style="list-style-type: none">Automata TheoryRegular 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
<p>Exams:</p> <ul style="list-style-type: none"> • Intersemester Exam • Final Exam 	<ul style="list-style-type: none"> • Intersemester Exam (15%): Assessment of the knowledge acquired in automata theory, regular expressions, and data cleaning through problem-solving using programming. • Final Exam (45%): Evaluation of computational and abstract thinking for problem-solving in data extraction, cleaning, organization, and storage through programming. 	60 %
<p>Labs</p> <ul style="list-style-type: none"> • Weekly Assignments (10%) • Final Project (20%) 	The knowledge acquired will be assessed weekly through an individual practical case that must be solved through programming.	30 %
<p>Practice sessions</p> <ul style="list-style-type: none"> • Attitude, participation, and completion of the problems posed in collaborative and individual sessions. • Teamwork • Oral communication 	<ul style="list-style-type: none"> • Collaborative Methodology and Best Practices (5%): Teamwork, following the use of collaborative best practices. • Classroom Work and Oral Communication (5%): Participation and completion of problems posed in sessions. Ability to orally present results obtained in practical cases. 	10 %



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

Syllabus
2023 - 2024

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 + 45% Final Exam + 10% Weekly Assignments + 20% Final Project + 10% 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.

In compliance with current regulations on the **protection of personal data**, we would like to inform you that you may consult the aspects related to privacy and data [that you have accepted on your registration form](#) by entering this website and clicking on "download"

<https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>