



POLITÉCNICA

INTERNATIONAL  
CAMPUS OF  
EXCELLENCE

COORDINATION PROCESS OF  
LEARNING ACTIVITIES  
PR/CL/001



E.T.S. de Ingenieros  
Informáticos

# ANX-PR/CL/001-01

## LEARNING GUIDE

### SUBJECT

**103000863 - Data Analysis**

### DEGREE PROGRAMME

10AZ - Master Universitario En Innovación Digital

### ACADEMIC YEAR & SEMESTER

2021/22 - Semester 1

## Index

---

### Learning guide

1. Description.....	1
2. Faculty.....	1
3. Prior knowledge recommended to take the subject.....	2
4. Skills and learning outcomes .....	2
5. Brief description of the subject and syllabus.....	4
6. Schedule.....	7
7. Activities and assessment criteria.....	10
8. Teaching resources.....	12
9. Other information.....	14

## 1. Description

---

### 1.1. Subject details

<b>Name of the subject</b>	103000863 - Data Analysis
<b>No of credits</b>	4.5 ECTS
<b>Type</b>	Optional
<b>Academic year of the programme</b>	Second year
<b>Semester of tuition</b>	Semester 3
<b>Tuition period</b>	September-January
<b>Tuition languages</b>	English
<b>Degree programme</b>	10AZ - Master Universitario en Innovación Digital
<b>Centre</b>	10 - Escuela Tecnica Superior De Ingenieros Informaticos
<b>Academic year</b>	2021-22

## 2. Faculty

---

### 2.1. Faculty members with subject teaching role

<b>Name and surname</b>	<b>Office/Room</b>	<b>Email</b>	<b>Tutoring hours *</b>
Antonio Jesus Diaz Honrubia	4302	antoniojesus.diaz@upm.es	Sin horario. contact the professor
Ernestina Menasalvas Ruiz (Subject coordinator)		ernestina.menasalvas@upm. es	Sin horario. contact the profesor
Alejandro Rodriguez Gonzalez	4302	alejandro.rg@upm.es	Sin horario. contact the professor

\* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

## 2.3. External faculty

Name and surname	Email	Institution
Mike Freeman	mikefree@uw.edu	U. Washington

## 3. Prior knowledge recommended to take the subject

---

### 3.1. Recommended (passed) subjects

- Data Processes

### 3.2. Other recommended learning outcomes

- big data basics
- data bases
- machine learning
- data bases, basics of machine learning

## 4. Skills and learning outcomes \*

---

### 4.1. Skills to be learned

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB08 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CG01 - Que los estudiantes sean capaces de predecir y controlar la evolución de situaciones complejas mediante el desarrollo de nuevas e innovadoras metodologías de trabajo adaptadas al ámbito científico/investigador,

tecnológico o profesional concreto, en general multidisciplinar, en el que se desarrolle su actividad.

CG02 - Que los estudiantes desarrollen la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro su ámbito temático explorando y generando nuevas ideas sistemáticamente, en contextos interdisciplinares y, en su caso, con una alta componente de transferencia del conocimiento.

CG03 - La capacidad de usar la lengua inglesa de manera competente, es decir, con capacitación para tareas complejas de trabajo y estudio.

CG07 - Capacidad de trabajar y comunicarse también en contextos internacionales.

## 4.2. Learning outcomes

RA35 - Ser capaz de procesar datos masivos.

RA17 - Acquire specialized knowledge from innovative fields of studies

RA18 - Apply techniques for modelling the context of use

RA34 - Conocer cómo se aplican las técnicas de computación científica en algún campo específico de ciencia o ingeniería.

RA50 - apply data analysis techniques on data streams

RA49 - apply IA techniques in real data scenarios

RA14 - Apply the acquired knowledge in real contexts

\* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

## 5. Brief description of the subject and syllabus

---

### 5.1. Brief description of the subject

In this course we will deepen on the importance of data for an organization. In fact, the course is centered on the process of extraction of knowledge from databases as a support for decision making.

This course will be adapted depending on the students profile but main goal will be to deepen on the importance of data for an organization and deepen on the data lifecycle applying the methods learnt in real scenarios with special emphasis on techniques for data stream mining and complex network analysis

Consequently the course will start proposing a real scenario in which it will be required to proceed with the analysis of data sources in an organization and very briefly analyzed data base management systems.

Later students will understand the data value chain and will go deep into the process of knowledge extraction. At this stage CRISP-Dm methodology will be used.

The course will follow on the different phases of the process applied to the proposed scenario: i) business understanding, ii) data understanding, iii) data preparation, iv) modeling v) evaluation and vi) deployment.

Through all the phases the main emphasis will be on students getting hands on the different steps, techniques, algorithms and tools.

Before finishing the course will cover basic aspects of the GDPR and the implications on the process of knowledge extraction in a company

## 5.2. Syllabus

### 1. Introduction

- 1.1. Course description.
- 1.2. Data Science and Data Scientist Skills.
- 1.3. Proposal of a real case use.

### 2. Analysis of the available datasets to solve the problem

- 2.1. The BIG Data Value Chain.
- 2.2. Data Warehouse.
- 2.3. Data Lakes

### 3. The process of discovery applied to the use case

- 3.1. CRISP-DM

### 4. Business Understanding

- 4.1. Goal of BU.
- 4.2. Planning of a DataScience project.

### 5. Data Understanding

- 5.1. Understanding data.
- 5.2. Nulls and outliers detection.
- 5.3. Correlation analysis

### 6. Data Preparation

- 6.1. Preparing data for mining: dealing with problems encountered in understanding, transforming data, discretization, data reduction, agregation, ?.

### 7. Data mining/data modeling

- 7.1. Type of problems found: analysis of the aata nature , data problems and possible algorithms.
- 7.2. Classification, association and clustering

### 8. Evaluation and Deployment

- 8.1. Evaluation of the models.
- 8.2. Deployment of the models

### 9. Ethics

## 9.1. GDPR and implications in Data Science



## 6. Schedule

### 6.1. Subject schedule\*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1	<b>Unit 1</b> Duration: 02:00		<b>Unit 1</b> Duration: 02:00	
2	<b>Unit 2</b> Duration: 01:00		<b>Unit 2</b> Duration: 01:00  <b>Unit 2</b> Duration: 02:00	
3	<b>Unit 3</b> Duration: 01:00		<b>Unit 3</b> Duration: 01:00  <b>Unit 3</b> Duration: 02:00	
4	<b>Unit 4</b> Duration: 01:00		<b>Unit 4</b> Duration: 01:00  <b>Unit 4</b> Duration: 02:00	
5	<b>Unit 5</b> Duration: 01:00		<b>Unit 5</b> Duration: 01:00  <b>Unit 5</b> Duration: 01:00	
6	<b>Unit 6</b> Duration: 01:00		<b>Unit 6</b> Duration: 01:00  <b>Unit 6</b> Duration: 01:00	
7	<b>Unit 6</b> Duration: 01:00		<b>Unit 6</b> Duration: 01:00  <b>Unit 6</b> Duration: 01:00	

8	Unit 6 Duration: 01:00		Unit 6 Duration: 01:00  Unit 6 Duration: 01:00	
9	Unit 7 Duration: 01:00		Unit 7 Duration: 01:00  Unit 7 Duration: 01:00	
10	Unit 7 Duration: 01:00		Unit 7 Duration: 01:00  Unit 7 Duration: 01:00	<b>Evaluation First Assignment</b>  Continuous assessment Presential Duration: 02:00
11	Unit 7 Duration: 01:00		Unit 7 Duration: 01:00	
12	Unit 8 Duration: 01:00		Unit 8 Duration: 01:00  Unit 8 Duration: 01:00	
13	Unit 8 Duration: 01:00		Unit 8 Duration: 01:00  Unit 8 Duration: 01:00	
14	Unit 9 Duration: 02:00		Unit 9 Duration: 02:00	
15				<b>Evaluation Second Assignment</b>  Continuous assessment Presential Duration: 02:00
16				<b>Exam</b>  Final examination Presential Duration: 02:00  <b>Exam</b>  Continuous assessment Presential Duration: 02:00

17				
----	--	--	--	--

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

\* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

## 7. Activities and assessment criteria

### 7.1. Assessment activities

#### 7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
10	Evaluation First Assigment		Face-to-face	02:00	25%	4 / 10	CG03 CG07 CB07 CB06 CB08 CG01 CG02
15	Evaluation Second Assigment		Face-to-face	02:00	25%	4 / 10	CG03 CG07 CB07 CB06 CB08 CG01 CG02
16	Exam		Face-to-face	02:00	50%	4 / 10	CG03 CG07 CB07 CB06 CB08 CG01 CG02

#### 7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
16	Exam		Face-to-face	02:00	100%	5 / 10	CB07 CB06 CB08 CG03 CG07 CG01 CG02

### 7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Exam in July		Face-to-face	02:00	100%	5 / 10	CB07 CB06 CB08 CG01 CG02 CG03 CG07

## 7.2. Assessment criteria

The "only final exam" evaluation will only be offered if the UPM "Normativa Reguladora de los Sistemas de Evaluación" requires so in the academic year 2015-2016, and the procedure to opt in will be as stated by the head of studies. **Continuous evaluation:** The course will be evaluated by:

- 2 practical assignments,
- 1 exam

.Attendance is mandatory (It is allowed not to attend up to 6 hours without proper justification)

Practical assignments will be done in groups among those enrolled in the course at the beginning of the academic year (nature and number of components will be established at the beginning of the course, depending on the number of students enrolled).

In order to pass the course in the fall semester the requirements are:

- 1. To obtain a minimum of 50 points out of 100 in the added evaluation.
- 2. It is MANDATORY to do the exam and do the practical assignment.
- 3. In the exam and on the practical assignments students must obtain a minimum of 40%.

Final score will be calculated as follows:

- 50% Practical assignment (divided between the two assignments)

- 50% Final exam

**Final exam evaluation** Those students whose extraordinary circumstances cannot perform the continuous evaluation, and having done the final exam evaluation written petition during the first 15 days of the course, will perform the final exam evaluation without having the opportunity to do the continuous evaluation;

Those students failing to attend 85% of of the lectures will also do the final evaluation.

In these premises, the final exam evaluation will consist of an exam as stated by the head of studies.

Measures against copies and fraud Rights and duties of college students are gathered on the statutes of the Universidad Politécnica de Madrid (BOCM de 15 de noviembre de 2010) and in the statutes of the college student (RD 1791/2010 de 30 de diciembre). Article 124 a) of EUPM fixes the duty of the student... "to follow with responsibility and taking advantage of the learning process, knowledge acquisition correspondent to its condition of college student"... and the article 13 of the statutes of the college student in its point d) also specifies as duty of the college student "abstain from the use or cooperation in fraudulent procedures in the evaluation assessments, in the assignments developed or in the official documents of the university". In the case that in the development of the evaluation assessments it is appreciated a breach in the duties as college student, the subject coordinator may communicate the headmaster as established in the article 74 (n) of EUPM to have the competences to "propose the initiation of a disciplinary procedure to any College member, by its own initiative or as instance from the "Comisión de Gobierno"" to the Rector, pursuant to the statutes and rules of application.

## 8. Teaching resources

---

### 8.1. Teaching resources for the subject

Name	Type	Notes
Moodle	Web resource	<a href="http://moodle.upm.es">http://moodle.upm.es</a>
Data Mining book	Bibliography	Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann, ISBN 978-0-12-374856-0, 2011.

Smart Machines book	Bibliography	Smart Machines: IBM's Watson and the Era of Cognitive Computing. Columbia University Press (October 15, 2013)
Database Systems: The Complete Book (DS:CB), by Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom	Bibliography	
"MySQL Administrator's Bible". Sheeri K. Cabral and Keith Murphy. Wiley	Bibliography	
Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Addison Wesley (May, 2005).	Bibliography	
Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann, ISBN 978-0-12-374856-0, 2011.	Bibliography	

## 9. Other information

---

### 9.1. Other information about the subject

teaching will be online.

Platform of the university will be used.

Sessions will be recorded for students to access them anytime.