Teaching Guide				
Identifying Data				2023/24
Subject (*)	Data Management Technology		Code	614493106
Study programme	Mestrado Universitario en Técnicas Estadísticas (Plan 2019)			
Descriptors				
Cycle	Period	Year	Type	Credits
Official Master's Degree	1st four-month period	First	Optional	5
Language	Galician			
Teaching method	Face-to-face			
Prerequisites				
Department	Enxeñaría de ComputadoresMatemáticas			
Coordinador	López Taboada, Guillermo	E-mail	guillermo.lopez.taboada@udc.es	
Lecturers	López Taboada, Guillermo Oviedo de la Fuente, Manuel Teijeiro Paredes, Diego	E-mail	guillermo.lopez.taboada@udc.es manuel.oviedo@udc.es diego.teijeiro@udc.es	
Web	eio.usc.es/pub/mte			
General description	The objective of this course is that any student, regardless of his/her previous academic background, would be able to acquire a solid knowledge of database management, both relational and non-relational technologies. Likewise, familiarization with the main computational techniques for the practical management of massive data will be sought. This will give the student a great autonomy when it comes to processing and studying data, regardless of its format and origin.			

Study programme competences / results	
Code	Study programme competences / results
A16	CE1 - Coñecer, identificar, modelar, estudar e resolver problemas complexos de estatística e investigación operativa, nun contexto científico, tecnolóxico ou profesional, xurdidos en aplicacións reais.
A17	CE2 ? Desenvolver autonomía para a resolución práctica de problemas complexos surdidos en aplicación reais e para a interpretación dos resultados cara á axuda na toma de decisións.
A18	CE3 - Adquirir coñecementos avanzados dos fundamentos teóricos subxacentes ás distintas metodoloxías da estatística e a investigación operativa, que permitan o seu desenvolvemento profesional especializado.
A21	CE6 - Adquirir coñecementos teórico-prácticos avanzados de distintas técnicas matemáticas, orientadas específicamente á axuda na toma de decisións, e desenvolver a capacidade de reflexión para avaliar e decidir entre distintas perspectivas en contextos complexos.
A23	CE8 - Adquirir coñecementos teórico-prácticos avanzados das técnicas destinadas á realización de inferencias e contrastes relativos a variables e parámetros dun modelo estatístico, e saber aplicalos con autonomía suficiente nun contexto científico, tecnolóxico ou profesional.
A24	CE9 - Coñecer e saber aplicar con autonomía en contextos científicos, tecnolóxicos ou profesionais, técnicas de aprendizaxe automático e técnicas de análise de datos de alta dimensión (big data).
A25	CE10 - Adquirir coñecementos avanzados sobre metodoloxías para a obtención e o tratamento de datos derivados de distintas fuentes, como enquisas, internet, ou entornos ?na nube".
B1	CB6 - Posuír e comprender coñecementos que acheguen unha base ou oportunidade de ser orixinais no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación
B2	CB7 - Que os estudantes saiban aplicar os coñecementos adquiridos e a súa capacidade de resolución de problemas en ámbitos novos ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo
B3	CB8 - Que os estudantes sexan capaces de integrar coñecementos e enfrontarse á complexidade de formular xuízos a partir dunha información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas vinculadas á aplicación dos seus coñecementos e xuízos
B4	CB9 - Que os estudantes saiban comunicar as súas conclusións e os coñecementos e razóns últimas que as sustentan a públicos especializados e non especializados dun modo claro e sen ambigüidades
B5	CB10 - Que os estudantes posúan as habilidades de aprendizaxe que lles permitan continuar estudando dun modo que haberá de ser en gran medida autodirixido ou autónomo.

B17	CG1 - Coñecer, comprender e saber aplicar os principios, metodoloxías e novas tecnoloxías na estatística e a investigación operativa en contextos científico/académicos, tecnolóxicos ou profesionais especializados e multidisciplinares, así como adquirir as destrezas e competencias descritas nos objectivos generales do título.
B18	CG2 - Desenvolver autonomía para identificar, modelar e resolver problemas complexos da estatística e da investigación operativa en contextos científico/académicos, tecnolóxicos ou profesionais especializados e multidisciplinares.
B19	CG3 - Desenvolver a capacidade para realizar estudos e tarefas de investigación e transmitir os resultados a públicos especializados, académicos e xeneralistas.
B20	CG4 - Integrar coñecementos avanzados e enfrontarse á toma de decisións a partir de información científica e técnica.
B21	CG5 - Desenvolver a capacidade de aplicación de algoritmos e técnicas de resolución de problemas complexos no eido da estatística e a investigación operativa, manexando o software especializado axeitado.
C11	CT1 - Desenvolver firmes capacidades de razoamento, análise crítica e autocrítica, así como de argumentación e de síntese, contextos especializados e multidisciplinais.
C12	CT2 - Desenvolver destrezas avanzadas no manexo de Tecnoloxías da Información e a Comunicación (TIC), tanto para a obtención de información como para a difusión do coñecemento, nun ámbito científico/académico, tecnolóxico ou profesional especializado e multidisciplinar.
C13	CT3 - Ser capaz de resolver problemas complexos en novos escenarios mediante a aplicación integrada dos coñecementos.
C14	CT4 - Desenvolver unha sólida capacidade de organización e planificación do estudo, asumindo a responsabilidade do seu propio desenvovemento profesional, para a realización de traballos en equipo e de xeito autónomo.
C15	CT5 - Desenvolver capacidades para o aprendizaxe e a integración no traballo en equipos multidisciplinais, nos ámbitos científico/académico, tecnolóxico e profesional.

Learning outcomes			
Learning outcomes	Study programme competences / results		
Manage autonomously and solvently the software necessary to access datasets in professional environments and / or in the cloud.	AC16 AC17 AC23 AC24 AC25	BJ5 BJ17 BJ19 BJ20 BJ21	CJ12
Know how to manage massive data sets in a multidisciplinary environment that allows participation in complex professional projects that require the use of statistical techniques.	AC18 AC21	BJ1 BJ2 BJ3 BJ4 BJ18	CJ11 CJ13 CJ14 CJ15
Knowing how to relate the design and management database software with the especifically implemented data analysis software.	AC16 AC17 AC21 AC24 AC25	BJ17 BJ18 BJ21	CJ12 CJ13

Contents	
Topic	Sub-topic
1. Introduction to the SQL language	1.1 Relational Databases 1.2 SQL Sintax 1.3 Connecting R to databases
2. Introduction to NoSQL technologies	2.1 Concepts and types of NoSQL Databases (documentary, colum-oriented, key/value & graph) 2.2 Connecting R to NoSQL

3. Technologies for massive data handling	3.1 Big Data Technologies 3.2 Visualization and implementation of dashboards 3.3 Introduction to the analysis of massive data

Planning				
Methodologies / tests	Competencies / Results	Teaching hours (in-person & virtual)	Student?s personal work hours	Total hours
Guest lecture / keynote speech	A16 A18 A21 A23	20	30	50
Laboratory practice	A16 A18 A24 A25 B2 B3 B5 B20 C11 C12 C13 C14	13	26	39
Problem solving	A16 A17 B2 C13	0	28	28
Seminar	A17 A24 A25 B1 B17 B19 B21 C15	2	3	5
Mixed objective/subjective test	A16 A18 A21 A23 A24 A25 B1 B2 B3 B4 B17 B18 C13	2.5	0	2.5
Personalized attention		0.5	0	0.5
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	Description
Guest lecture / keynote speech	Expository sessions, in which they will present concepts and / or procedures, providing basic information necessary to understand a theoretical perspective or a practical procedure, promoting the participation of students.
Laboratory practice	Interactive sessions of laboratory (computer lab) practices or problem solving, where the teachers will support and supervise how the students put in practice the acquired knowledge.
Problem solving	Personal work of the students in the realization of practical exercises and problem solving from the lectures and laboratory practices.
Seminar	Group work technique aimed at in-depth exploration of given topic, consisting of group discussion, individual engagement, preparation of texts and collective conclusions.
Mixed objective/subjective test	Mixed test consisting of essay-type and objective test questions. Essay section consists of open (extended answer) questions; objective test may contain multiple-choice, ordering and sequencing, short answer, binary, completion and/or multiple-matching questions.

Personalized attention	
Methodologies	Description
Laboratory practice Seminar Problem solving	Personalized attention (or in a small group / very small group) in these methodologies, both in the classroom and in the tutoring hours, to solve any doubts that may arise in the development of the teaching/learning process and that have not been resolved effectively previously.

Assessment			
Methodologies	Competencies / Results	Description	**Qualification**
Laboratory practice	A16 A18 A24 A25 B2 B3 B5 B20 C11 C12 C13 C14	Evauation of the laboratoy practices carried out by the students.	40

Mixed objective/subjective test	A16 A18 A21 A23 A24 A25 B1 B2 B3 B4 B17 B18 C13	The final exam will evaluate the following aspects: Theoretical Concepts regarding the subject: Mastering the theoretical and operational knowledge of the subject. Practical expertise: Handling and Understanding of the operational and theoretical knowledge of the subject	60

Assessment comments
To pass the course in the 1st opportunity call, it is mandatory to obtain at least on the one hand a 30% of the maximum possible grade in the laboratory practices, and on the other hand, at least a 30% of the maximum possible grade of the written mixed objective/subjective test performed during the term and to have a final grade (lab practices + mixed objective/subjective test) at least a 50% of the maximum possible grade.Only the final mixed objective/subjective test is remediable in the 2nd opportunity call. The other parts of the grade are the ones obtained during the term. Students taking advantage of the ahead December call will be evaluated using their grades (lab practices) from the previous term. The only condition to pass the course in this call is to obtain a final grade equal to or greater than 50% of the maximum possible grade.Once a student gets an evaluation for a laboratory practice implies he/she will be graded. Thus, "not graded" mark is not possible once an exercise has been evaluated.The fraudulent performance of tests or assessment activities, once proven, will directly result in the grade of suspension in the call in which it is committed: the student will be graded with "suspension" (numerical grade 0) in the corresponding call for the academic year, whether the commission of the offense occurs in the first opportunity or in the second. For this, your rating will be modified in the first opportunity report, if necessary.

Sources of information	
Basic	- López-Taboada, G. y Fernández-Casal, R. (2021). Prácticas de Tecnologías de Gestión y Manipulación de Datos. . https://gltaboada.github.io/tgdbook - Daroczi, G. (2015). Mastering Data Analysis with R. Packt Publishing - Grolemund, G. y Wickham, H. (2016). R for Data Science. https://r4ds.had.co.nz/ & O'Reilly - Silberschatz, A., Korth, H. y Sudarshan, S. (2014). Fundamentos de Bases de Datos. Mc Graw Hill
Complementary	- Wes McKinney (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly (2ª ed.) - Tom White (2015). Hadoop: The Definitive Guide. O'Reilly (4ª ed.) - Alex Holmes (2014). Hadoop in practice. Manning (2ª ed.) - Centro de Supercomputación de Galicia (2019). Servicio de Big Data del CESGA. https://bigdata.cesga.es/ - Rubén Fernández Casal (2019). Ayuda y Recursos para el Aprendizaje de R. https://rubenfcasal.github.io/post/ayuda-y-recursos-para-el-aprendizaje-de-r/

Recommendations
Subjects that it is recommended to have taken before
Subjects that are recommended to be taken simultaneously
Subjects that continue the syllabus
Other comments

Recommendations for studying this subjectDue to the large practical component of the subject, it is advisable to be up-to-date with practices during the semester. The software tools used in this course are generally open-source or have free license for students.Gender Perspective-According to the different application
regulations for university teaching, the gender perspective will be
incorporated in this subject (non-sexist language will be used, bibliography
from authors of both sexes will be used, students will be encouraged to
participate in class...)- Work will be done to identify and modify
prejudices and sexist attitudes and influence the environment to modify them
and promote values of respect and equality.-Situations of discrimination based on
gender must be detected and actions and measures will be proposed to correct
them.