

General information	
Academic subject	Multivariate Data Analysis
Degree course	Strategie di Impresa e Management (SIM)
Academic Year	2° Year
European Credit Transfer and Accumulation System (ECTS)	8 CFU/ECTS
Language	Italian (English on demand)
Academic calendar (starting and ending date)	I° Semester (Lessons starting on September 28, 2022)
Attendance	Classroom attendance is not mandatory but recommended

Professor/ Lecturer	
Name and Surname	Massimo Bilancia
E-mail	massimo.bilancia@uniba.it, massi.bilancia@gmail.com
Telephone	NA
Department and address	Ionic Department - Taranto
Virtual headquarters	Microsoft Teams
Tutoring (time and day)	Tuesday 9.30 – 12.30 AM Thursday 9.30 – 12.30 AM

Syllabus	
Learning Objectives	The course provides the theoretical and practical foundations for Data Science techniques and their applications in the field of economics.
Course prerequisites	Basic knowledge of statistics and mathematics.
Contents	<p>Part I – Basic probability and statistical inference</p> <ol style="list-style-type: none"> 1. Introduction 2. Elementary probability 3. Discrete and continuous random variables 4. Double and multivariate random variables 5. The elements of statistical inference 6. Data matrices <p>Part II – Data Mining and knowledge discovery</p> <ol style="list-style-type: none"> 1. Supervised and unsupervised learning 2. Association rules and Market Basket Analysis 3. Naïve Bayes classification 4. Decision trees 5. Clustering I: hierarchical clustering 6. Clustering II: k-means clustering <p>Part III – Lab: Introduction to R and Data Mining libraries</p>
Books and bibliography	M. Bilancia (2020) <i>Dispense per il Corso di Metodi Statistici Multivariati – Versione 1.2 Febbraio 2020</i> . Freely available under Creative Commons 4.0 CC BY-NC-ND Licence.
Additional materials	

Work schedule			
Total	Lectures	Hands on (Laboratory, working groups, seminars, field trips)	Out-of-class study hours/ Self-study hours

Hours			
200	45 (1 CFU = 9h)	27 (1 CFU = 9h)	128
ECTS			
8	5	3	
Teaching strategy			
Lectures + PC-based lab instruction by R. Lab lectures may be delivered in blended learning mode, subject to applicable regulations.			
Expected learning outcomes			
Knowledge and understanding on:	The course aims to teach the basic elements of Data Science techniques.		
Applying knowledge and understanding on:	The student will learn to appreciate and apply in practice the models learned in the theoretical part of the course, using common data analysis software		
Soft skills	<ul style="list-style-type: none"> • <i>Making informed judgments and choices</i> <ul style="list-style-type: none"> ○ The student will be able to decide on the most appropriate model to gain new insights in the different intended application areas (business domain applications, business intelligence and marketing) • <i>Communicating knowledge and understanding</i> <ul style="list-style-type: none"> ○ The student will be able to produce reports on the new knowledge extracted by Data Mining techniques. • <i>Capacities to continue learning</i> <ul style="list-style-type: none"> ○ The course aims to provide the basic elements of Data Science techniques, knowledge on which is essential for the possibility of attending courses of a more advanced nature that prepare for the figure of Data Scientist. 		

Assessment and feedback	
Methods of assessment	<ol style="list-style-type: none"> 1. PRACTICAL TEST AT THE CALCULATOR (duration 2 hours), consisting of the analysis of a real data set using the methods explained in the laboratory lectures: The candidate must prepare a Microsoft™ Word™ document (or using Libre Office word processor) containing all R commands used, all textual and graphical output generated, and explanatory comments. You may refer to lab handouts 4-5-6-7, which are printed on paper, to complete the practical exam: You will have available for this purpose the handouts of these lectures, reproduced in a more practical format of 4 slides per sheet. These slides must be printed as you received them, i.e., you may not annotate them or add your notes to them (also, you may not use the PDF file of the handouts: Violation of these requirements will result in your inability to use them during the exam). Consultation of other texts or documents (in written or electronic form) is not allowed and will result in immediate exclusion from the exam. 2. WRITTEN EXAM TEST (duration 2 hours), centred on the preparation of a paper on two open questions (free form) chosen from all the theoretical topics covered in class. The first of the two questions relates to the first part of the course (set of slides on theory numbered 1-2-3-4-5), and the second relates to the second part (set of slides on theory numbered 6-7-8-9-10). No texts, documents or computer/telephone assistance of any kind may be used to complete the paper, as this will result in immediate disqualification from the exam. <p>It is possible to take the two tests in two separate attempts: However, the laboratory test must always precede the written test (note: exceptions to this rule are not possible). Passing the lab test will be retained for all appointments in the 2022/23 academic year. If the two tests are taken in the same appointment, both</p>

	<p>tests must be repeated if the lab test is not passed. There are no ongoing tests. Test results will be announced in an official Teams room designated for distribution of course materials.</p>
Evaluation criteria	<ul style="list-style-type: none"> • <i>Knowledge and understanding</i> <ul style="list-style-type: none"> ○ Ability to accrue knowledge in the field of probability theory and multivariate data analysis, with a view toward applications in economics and business. • <i>Applying knowledge and understanding</i> <ul style="list-style-type: none"> ○ Ability to apply learned theoretical skills in practice using specialized data analysis software. • <i>Autonomy of judgment</i> <ul style="list-style-type: none"> ○ Ability to select the appropriate data analysis technique for the nature of the proposed problem. • <i>Communicating knowledge and understanding</i> <ul style="list-style-type: none"> ○ Demonstrated ability to effectively explain the skills learned. • <i>Communication skills</i> <ul style="list-style-type: none"> ○ Ability to produce a meaningful report on the proposed problem.
Criteria for assessment and attribution of the final mark	Eligibility for laboratory examination, final grade expressed in 30/30 based on written examination.
Additional information	