COURSE DESCRIPTION – ACADEMIC YEAR 2022/2023

Course title	**Programming and Visualisation for Data Science**
Course code	76086
Scientific sector	ING-INF/05
Degree	Master in Software Engineering for Information Systems (LM-18)
Semester	1
Year	1
Credits	12
Modular	Yes

Total lecturing hours	80
Total lab hours	40
Attendance	Not compulsory. Non attending students have to agree with the lecturer on the modalities of independent study at the beginning of the course.
Prerequisites	Basic programming concepts
Course page	https://ole.unibz.it/ and https://teams.microsoft.com/

Specific educational objectives	The course belongs to the type "caratterizzanti – discipline informatiche". **Module 1: Data Visualization and Exploration** The course is designed to acquire professional skills and knowledge useful when exploring datasets. In particular, the student will be able to visualize datasets choosing the most appropriate technique for the data at hand, and will be able to get insights from the data supported by the visualizations, using basic statistical tools. The student will also learn to avoid the common pitfalls in visualization that can mislead the analysis. Visualization and data handling are done using the R programming language, following the best practices of reproducible research. **Module 2: Programming for Data Science** The course is designed to provide specific professional skills. The students will learn how to organize and analyze data by writing programs, using the Python programming language. More specifically, the students will practically learn to import, manipulate, analyze, visualize, and model a dataset. The students will also get familiar with libraries that can be effectively used for data analytics.

Module 1	**Data Visualisation and Exploration**
Module code	73062A
Module scientific sector	INF/01
Lecturer	Matteo Ceccarello
Contact	mceccarello@unibz.it
Scientific sector of lecturer	INF/01
Teaching language	English
Office hours	Arranged beforehand by email
Lecturing assistant (if any)	--

Contact LA	--
Office hours LA	--
Credits	6
Lecturing hours	40
Lab hours	20
List of topics	• Reproducible analysis practices • Human perception for effective visualization • Data types and visual encodings • Visualization idioms • Exploratory data analytics, data exploration, and feature engineering • Advanced libraries for data visualization
Teaching format	Frontal lectures, lab assignments, project work.

Module 2	**Programming for Data Science**
Module code	73062B
Module scientific sector	ING-INF/05
Lecturer	Antonio Liotta
Contact	antonio.liotta@unibz.it
Scientific sector of lecturer	ING-INF/05
Teaching language	English
Office hours	Arranged beforehand by email
Lecturing assistant (if any)	--
Contact LA	--
Office hours LA	--
Credits	6
Lecturing hours	40
Lab hours	20
List of topics	• Languages for programming data and data visualization • Integrated Development Environments for Data Science • Data wrangling, cleaning, and preprocessing • Advanced libraries for linear algebra and statistics • Data science pipelines, from data ingestion to models and analysis • Model tuning, validation, and testing
Teaching format	Frontal lectures, lab assignments, project work.

Learning outcomes	Knowledge and understanding: • D1.1 - Possess sound knowledge of both the fundamentals and the application aspects of the various fundamental areas of computer science • D1.3 - Acquire thorough knowledge of the scientific method of investigation applied to even complex systems and innovative technologies that support information technology and its applications • D1.8 - Obtain ability to read and understand specialised scientific documentation, such as conference proceedings, articles in scientific journals, technical manuals Applying knowledge and understanding: • D2.1 - Know how to apply the fundamentals of empirical analysis of ICT data for the construction of mathematical

	models for evaluating and predicting characteristics of software applications and systems Making judgments • D3.1 - Independent ability to select documentation from various sources, including technical books, digital libraries, technical scientific journals, web portals or open source software and hardware tools. Communication skills • D4.2 - Ability to present one's work in a clear and comprehensible way in front of an audience, including non-specialists • D4.3 - Ability to structure and draft scientific and technical documentation Learning skills • D5.3 – In the context of a problem-solving activity, ability to extend even incomplete knowledge taking into account the project's final objective
	•

Assessment	The exam modalities are the same for both the attending and the non-attending students. Project work (70% of the final grade) and oral exam (30% of the final grade). All project works must have been submitted, at the very latest, 15 days ahead of the oral exam. In case of a positive mark, the projects will count for all 3 regular exam sessions.
Assessment language	English
Assessment Typology	Collegial
Evaluation criteria and criteria for awarding marks	70% project work, 30% oral exam. • Relevant for project work: clarity of presentation, ability to gain useful and novel insights from data, creativity, critical thinking, ability to adhere to reproducible research best practices • Ability to use R software to perform basic data preparation tasks, ability to properly use R plotting facilities, ability to summarize the concepts of the Grammar of Graphics and of human perception, ability to choose the best type of graphical representation for different types of data, correct usage of basic statistical tools • Ability to use Python to employ (understand, recall and use) data analytics methods in practical settings, from data collection and curation, to data analysis and visualization.

Required readings	• *Data Visualization. A practical introduction.* Haley. Available online • *R for Data Science*. Wickham. Available online • *A layered grammar of graphics*. Wickham. Available online •

	• *Python Data Science Handbook*, by Jake VanderPlas. O'Reilly Media (1st Edition, 2016). Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it
Supplementary readings	• *Fundamentals of Data Visualization*. Wilke. Available online • *Visualization Analysis and Design*. Munzer. Amazon • *Data Visualization: Charts, Maps, and Interactive Graphics.* Grant. Amazon • *Doing Data Science*. Cathy O'Neil, Rachel Schutt. O'Reilly, 2013, https://www.oreilly.com/library/view/doing-data-science/9781449363871/ • *Python for Data Analysis.* By Wes McKinney. O'Reilly, 2nd Edition, 2017, https://www.oreilly.com/library/view/python-for-data/9781491957653/
Software used	• Rstudio https://www.rstudio.com/ • Jupyter Notebook (for Python programing)