LUISS

Syllabus

Academic Year	2021/2022
Program	Data Science and Management
course	Python and R for Data Science (lab)
Term	I semester
Year	1
SSD	ING-INF/05 - Sistemi di elaborazione delle informazioni
Credits	6

INSTRUCTIONAL GOALS	The course aims at providing technical skills about coding aspects for data analysis. The Python programming language and the R environment are illustrated with a specific focus on those libraries, modules and functions that allow the students to manage data effectively. This provides an in-depth understanding of the approaches to be used to preprocess, clean, visualize and analyze data related to a plethora of different contexts. Students in this course will mainly acquire practical skills, necessary to analyze real data.
INTENDED LEARNING OUTCOMES They describe what a learner is expected to know, understand and be able to demonstrate after completion of a learning path.	**Knowledge and understanding**: The course will offer key techniques to manage and analyze data programmatically, in order to extract useful insights. The course will provide a good understanding of the fundamental issues in data analysis, along with the knowledge of all those programming libraries that are needed to analyze data in an effective way. **Applying knowledge and understanding**: On successful completion of this course students will be able to: ● Organize, visualize, and analyze large, complex datasets by means of Python and R programming language. ● Extract knowledge from data. ● Make use of data science tools and techniques effectively. **Making judgements**: Students are expected to analyze complex datasets. Working both independently and in small groups, students will be required to complete a project work. The project work will allow students to make their own judgements about the most appropriate data science techniques to be applied in a given use case. This will, in turn, allow the students to assess critically the advantages and disadvantages of the different approaches. **Communications Skills**: The course will give the students the possibility to understand terms and concepts related to data science, along with the main concepts, libraries and abstractions of R and Python programming in this context. Communication in small groups of students (that is, two or three students) will play a key role in the development of the course project work. Small groups are important communication units in academic, professional, civic, and personal contexts; the students will be able to communicate their ideas and analyses in the most effective way, along with being able to write appropriate technical reports of the data analyses carried out. **Learning skills**: This course will empower students with the capability to enhance their technical skills in order to manage data effectively by means of the Python and the R programming languages. A strong emphasis will be given to the practical application of data science libraries to real use cases.

Pre-requisites	Basic computer programming skills are required.
Course content	The course will cover the following topics: • Data Loading, Storage, and File Formats. • Data Cleaning and Preparation. • Data Manipulation with Pandas. • Data Wrangling: Join, Combine, and Reshape. • Plotting and Visualization. • Data Aggregation and Group Operations. • Advanced Numpy. • Data Import with readr. • Data Visualization with ggplot2. • Data Transformation with dplyr. • Relational Data with dplyr.
Reference Books	Lecture notes, research papers and course material will be made available on the e-learning platform. Recommended reading: • "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", 2nd Edition by Wes McKinney.Publisher: O'Reilly Media, Inc. Release Date: October 2017. ISBN: 9781491957660. • "R for Data Science: Import, Tidy, Transform, Visualize, and Model Data" by Garrett Grolemund and Hadley Wickham. Release Date: December 2016. ISBN: 978-1491910399.
Teaching Methods	The course consists of practical lab sessions and group project works, complemented by lectures in order to explain the underlying concepts.
Assessment	There will be a written midterm exam and a project. In the midterm exam students are required to demonstrate that: • they understand the fundamental issues in data analysis, and • they know and are able to use effectively the main software libraries needed to analyze data; • they are able to organize, visualize, and analyze large, complex datasets (in Python and R); • they can effectively deploy basic data science tools and techniques. The midterm will count for 50% of the grade. Students that will not take the midterm during the course are required to take an oral exam after the course, where they are required to demonstrate the same skills described above. In the project students are required to demonstrate that they are able to: • analyze and assess critically advantages and disadvantages of different basic data science techniques; • communicate their ideas, findings and analyses effectively; • write appropriate technical reports about the data analyses carried out. The project will count for 50% of the grade. The overall assessment will take into account the level of knowledge and understanding of fundamental issues in data analysis acquired by the students; their capacity for thinking analytically, logically and critically; their capacity to design and evaluate solutions for basic data science problems; their capacity to present effectively findings and conclusions and to write detailed technical reports and documentation about their project work.