UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

# COURSE SYLLABUS

# Data Science

**2627-2-E1808M012**

## Learning objectives

The objective of this course is to introduce students to programming and provide them with an overview of the data science and artificial intelligence techniques employed in a business environment. It covers the fundamentals of programming (Python), databases (SQL), and machine learning. **All classes and course materials will be in English.**

The course can be taken as an alternative to 'Informatica Generale e Laboratorio Informatico', which focuses instead on providing a general understanding of the issues related to technological innovation in financial companies and society, data and information management, digital business models, business information systems and their organisational and management implications.

**Learning Objectives (according to the Dublin Descriptors)**

**1. Knowledge and understanding**
By the end of the course, students will have acquired:

- knowledge of the fundamental principles of imperative programming;
- understanding of the basics of relational databases and the SQL language;
- an overview of the main data science, machine learning, and artificial intelligence techniques applicable in a business context;
- understanding of the role of data and digital innovation in business processes, particularly in the financial sector.

**2. Applying knowledge and understanding**
By the end of the course, students will be able to:

- design and implement simple programs in Python;
- query and process data using SQL;
- propose operational solutions for data management and analysis within interdisciplinary teams.

**3. Making judgements**
By the end of the course, students will be able to:

- critically assess the adequacy and limitations of IT solutions adopted in business contexts;
- identify the most relevant emerging technologies and evaluate their strategic impact on companies;
- integrate technical, organizational, and managerial knowledge to support data-driven business decision-making processes.

**4. Communication skills**
By the end of the course, students will be able to:

- use appropriate technical language to communicate effectively with IT specialists, data scientists, and business stakeholders;
- present analysis results and project proposals to both technical and non-technical audiences.

**5. Learning skills**
By the end of the course, students will have developed:

- the ability to continuously update their knowledge of emerging technologies in the field of applied business computing;
- the skills to autonomously and proactively learn new programming and data analysis tools and techniques.

## Contents

The course "Data Science" aims to acquaint students with programming and give them a comprehensive understanding of the data science and artificial intelligence methods utilized within a business setting. The program includes a brief introduction to artificial intelligence, an introductory overview of programming using the Python language, learning the basic concepts of databases using SQL, and gaining practical skills in machine learning techniques. **All classes and course materials will be in English.**

## Detailed program

*Introduction to Business Information Systems*: architectures and applications supporting the operational, tactical, and strategic needs of enterprises.

*Digital trends*: digital transformation processes in companies and technological evolution in the information society.

*Artificial Intelligence*: historical overview and future prospects of AI, with a focus on the latest developments, including large language models and their applications.

*Programming Basics*: basic concepts of imperative programming and development of simple programs in Python.

*Methods for Data Management and Analysis*: relational databases, SQL, data warehousing, and knowledge graphs.

*Machine Learning and Data Mining*: supervised and unsupervised solutions, with a specific focus on neural networks.

*Computer Lab*: the lab component aims to provide students with intermediate-level proficiency in using Microsoft

Excel.

## Prerequisites

Mathematical and logical knowledge acquired in high school. Proficiency in the English language.

## Teaching methods

Data Science - 42 hours

- 5 lecture-based sessions, all held in person.
- 11 sessions primarily focusing on Python and SQL, combining an initial lecture-based section with subsequent interactive activities, all held in person.

Computer Lab (Excel) - 12 hours

- 6 interactive laboratory sessions of 2 hours each, delivered live online.

## Assessment methods

The assessment of learning includes a written test and, optionally, an additional oral examination. The written test, which includes multiple-choice questions and open-ended questions, will take place in the teaching labs (laboratori didattici) in order to evaluate the student's skills in using the Python language and the Microsoft Excel application.

The final exam consists of two sections.

**Data Science Section:**

- 16 multiple-choice questions (1 correct answer out of 4), each worth 1 point (0 points for incorrect or missing answers). These questions will be randomly selected from a predefined list. Specifically, there will be three types of questions: 1) Data Science and Computer Science theory, 2) Python theory, 3) Python code interpretation (where a small code snippet is provided, and the output is requested).
- 1 Python exercise worth a maximum of 8 points (0 points for a substantially or mostly incomplete response). The exercise requires writing a code snippet to perform a specific task (e.g., iterating over a list and removing all even numbers, finding the minimum in a numeric list).
- 1 SQL query worth a maximum of 8 points (0 points for a substantially or mostly incomplete response; please note that the query must be executable). The exercise (similar to the one presented in the "SQL Exercises - Example SQL question in the final exam" section) will present a database consisting of several tables and will require formulating a specific query (usually involving multiple tables).

This section awards the maximum achievable score in the exam.

**Computer Lab Section (Excel):**

- 3 additional multiple-choice questions (1 correct answer out of 4) on the contents of the Computer Lab (answering these questions requires the use of Excel). The questions will cover both theory and practice.

Errors or omissions in the Computer Lab section result in the following penalties applied to the score obtained in the General Computer Science section:

- 3 correct answers: no penalty
- 1 incorrect or missing answer: -1 point
- 2 incorrect or missing answers: -3 points
- 3 incorrect or missing answers: -5 points

The final grade is determined by the Computer Science section score minus any penalties from the Computer Lab section (e.g., 27 points in the General Computer Science section, -3 points in the Computer Lab section - i.e., 2 incorrect or missing answers - resulting in a total score of 24/30).

## Textbooks and Reading Materials

- The slides used in the classroom and other study materials available on e-Learning platform.
- W3C Python tutorial (https://www.w3schools.com/python/)
- Downey, A. B. (2015). Think Python: How to Think Like a Computer Scientist, Version 2.4. Green Tea Press. (Freely available online with Creative Commons licence, https://greenteapress.com/thinkpython2/thinkpython2.pdf)

Optional textbook

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. Aurélien Géron. O'Reilly, 2019.

## Semester

First semester.

## Teaching language

English.

## Sustainable Development Goals

INDUSTRY, INNOVATION AND INFRASTRUCTURE