DAPS&CO
Data Analytics for Politics, Society and Complex Organizations
Public and Corporate Communication
Università di Milano

Data Management and Database Concepts
AY 2022-2023

OVERVIEW

The lab covers the fundamental concepts and techniques of data storage, data management and databases in Python. Students will learn how to work with both relational and NoSQL databases, and gain hands-on experience using Python for data manipulation, cleaning, and analysis. Throughout the course students will gain an understanding of the importance of developing a data strategy and how it can add value to a research project.

LEARNING OBJECTIVES

Students will learn how to:

1. Understand the fundamental concepts and principles of data storage and management;
2. Be able to use Python for data manipulation, cleaning, and analysis;
3. Understand the differences between SQL and NoSQL databases and when and how to use each;
4. Be able to apply the concepts and techniques learned in the course to real-world data management challenges through hands-on projects and exercises.

COURSE DETAILS

The course will cover the following topics:

Unit 1: Introduction to data management and database concepts

- Overview of data management and database systems
- Basic database terminology and concepts
- Introduction to data storage options (e.g. local, remote, cloud)
- Basic concepts of remote computing and storage options (e.g. AWS, Google Cloud)

Unit 2: Basic Python programming for data manipulation and analysis

- Introduction to Python programming language
- Basic data types and operations in Python
- Working with data structures in Python (e.g. lists, dictionaries)
- Introduction to the Pandas library for data manipulation and analysis
- Basic Pandas concepts and data structures (e.g. Series, DataFrame)
- Working with and manipulating data in Pandas

Unit 3: Working with SQL databases in Python

- Introduction to SQL and relational database concepts
- Using the Python SQLite3 module to work with SQLite databases
- Querying and updating data in SQLite databases using Python

Unit 4: Working with MongoDB databases in Python

- Introduction to MongoDB and NoSQL database concepts
- Using the Python PyMongo module to work with MongoDB databases
- Storing, querying, and updating data in MongoDB databases using Python

Unit 5: Elaborating a data strategy for a research project

- Identifying the key data sources and types of data needed for the project
- Developing a plan for collecting, storing, and organizing the data
- Evaluating the quality and reliability of the data
- Determining the most appropriate methods and tools for analyzing and visualizing the data

COURSE SYLLABUS AND ORGANIZATION

Teaching methods

The course will be delivered through a combination of lectures and hands-on lab activities, with emphasis on the latter. Students will be expected to bring their own laptops to each class, and to have installed the required software (e.g. Python, a DBMS, etc.) prior to the start of the course. Instructions will be circulated before the start of the course. In the lab activities, the instructor will introduce key concepts and techniques, and provide examples and demonstrations to illustrate the material. Students will have the opportunity to apply the concepts and techniques learned in the lectures, and to work on hands-on projects and exercises. The instructor will provide guidance and support as needed. The lab will include a mix of individual and group activities, with a focus on collaboration and peer learning. All the datasets, replication files of the lab sessions and reference texts will be made available at a dedicated URL before the beginning of the course.

Professor

Giovanni Pagano

Course schedule and location

The course starts on Wednesday, January 18th, and concludes on Thursday, February 16th, 2023. Classes will be held in "Aula 26 – Conservatorio (edificio 1)" every Wednesday and Thursday from 14:30 to 16:30.