

IDENTIFICATION

Course code : DAT502
ECTS : 2

DURATION

Lectures : 4
Tutorial classes : 4
Labs : 8
Total : 16

Project :
Personnal work :

ASSESSMENT

Continuous assessments and evaluated practical work.

TEACHING MATERIALS

Slides and guided practical work sheets.

LANGUAGE

English

CONTACT

Pierre Ugo TOURNoux
pierre.tournoux@univ-reunion.fr

Modifié le : 14 juillet 2026

UE : UE5-DATA

COURSE : Data : Description, Presentation, and Analysis

SCOPE OF THE COURSE

OBJECTIVES

Presenting data in an appropriate form is a challenging yet important task. For data scientists, this fundamentally enables them to display data in an attractive, easily and correctly understandable form for users, while also potentially providing valuable insights into the underlying structure of the data. This course aims to describe the tools, methods, and best practices for extracting information and visualizing it, whether in the context of exploratory analysis or the description of final results.

COMPETENCE AND SKILLS

Perform exploratory data analysis Visualize data

PROGRAM

Visualization approaches tailored to data types The role of visualization in data science. Illustrations, including historical and contemporary examples of visualization. Characteristics of effective visualization. Suitability of different techniques for different data and different users. Dashboards and interactive visualization. Software tools for visualization. Inference based on visualization. Preparing visualization—scaling, role of color. Types of charts—tables, scatter plots, pie charts, histograms, graphs, data maps, including pixel-based, glyph-based, graph-based, and map-based representations.

Working with different data types Data representation : numbers, text, images, data precision. Text data processing : bag of words, word counting, TF-IDF, n-grams, lexical analysis, syntactic analysis, semantic analysis, stopword filtering, stemming, basic applications. Image processing : data representation : multidimensional integer matrices, features, image operators, video operators. Object recognition. Extraction of higher-order features.

Information extraction Information Extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. It is an important technique for acquiring data from documents, web pages, and even multimedia content.

Data transformation Data transformation pipeline. Simple transformation function methods and their applications. Data normalization and its applications. Data encoding approaches and their applications. Data smoothing approaches and their applications.

Data cleaning Dimensions of data quality. Approaches to improving data quality. Data cleaning algorithms, including entity resolution, truth discovery, rule-based data cleaning. Different forms for data quality rules such as functional dependencies (FD), conditional functional dependencies (CFD), conditional inclusion dependencies (CIND), and matching dependencies (MD).

BIBLIOGRAPHY

REQUIREMENTS

- Basic knowledge of Python.
- DATA 051 : Databases.
- DEV 052 : Web Programming