

Utilisation & Exploitation des Données
Introduction to Data Science

Code cours *Course code:* **UED**

Crédits ECTS *ECTS Credits:* **1**

Département <i>Department</i>	: IA	Cours Lectures	:
Coordonnateurs <i>Lecturers</i>	: Brice CHARDIN	T.D. Tutorials	:
Période <i>Year of study</i>	: 1 ^{ère} année <i>1st year</i>	T.P. Laboratory sessions	:
Semestre <i>Semester</i>	: 2 ^{ème} semestre <i>2nd semester</i>	Projet <i>Project</i>	:
Evaluation <i>Assessment method(s)</i>	: 1 examen écrit, 3 TP <i>1 written exam, 3 practical work tests</i>	Non encadré <i>Homework</i>	:
Langue d'instruction <i>Language of instruction</i>	: Français <i>French</i>	Horaire global <i>Total hours</i>	: 18H30
Type de cours <i>Type of course</i>	: Obligatoire <i>Compulsory</i>		
Niveau <i>Level of course</i>	: <i>Undergraduate</i>		

Compétences attendues :

- Extraire de l'information de données structurées et semi-structurées.
- Automatiser et systématiser les traitements d'analyse des données.

Pré-requis :

- bases de la programmation (INF1)
- statistique descriptive

Contenu :

- structures de données (collections)
- transformation et préparation des données
- extraction d'information et analyse
- données structurées (CSV, relations) et semi-structurées (HTML, XML, JSON)
- implémentation en Python (pandas, Beautiful Soup) et en SQL

Bibliographie :

- Wes McKinney and the Pandas Development Team. pandas: powerful Python data analysis toolkit. Release 1.0.2. 2020.
- Wes McKinney. Python for Data Analysis. Second edition. O'Reilly Media, Inc. ISBN: 9781491957660. 2017.
- Philipp K. Janert. Data Analysis with Open Source Tools. O'Reilly Media, Inc. ISBN: 9780596802356. 2010.
- R. Lyman Ott and Michael Longnecker. An Introduction to Statistical Methods and Data Analysis. Fifth edition. Duxbury. ISBN: 0534251226. 2001.

Expected competencies:

- Discover knowledge from structured and semi-structured data.
- Design and implement data processing utilities.

Prerequisites:

- principles of computer programming (INF1)
- descriptive statistics

Content:

- data structures (collections)
- data wrangling
- knowledge discovery and data analysis
- structured (CSV, relational data) and semi-structured (HTML, XML, JSON) data
- Python (pandas, Beautiful Soup) and SQL implementations

Recommended reading:

- Wes McKinney and the Pandas Development Team. pandas: powerful Python data analysis toolkit. Release 1.0.2. 2020.
- Wes McKinney. Python for Data Analysis. Second edition. O'Reilly Media, Inc. ISBN: 9781491957660. 2017.
- Philipp K. Janert. Data Analysis with Open Source Tools. O'Reilly Media, Inc. ISBN: 9780596802356. 2010.
- R. Lyman Ott and Michael Longnecker. An Introduction to Statistical Methods and Data Analysis. Fifth edition. Duxbury. ISBN: 0534251226. 2001.