



---

**Guide pédagogique**  
**Module « Collecte, manipulation et  
visualisation de données »**  
**8.1 (4 crédits ECTS)**

---

***Place du module et enjeux***

Ce module se focalise sur les données au sein d'un projet et plus particulièrement des étapes amont et aval d'un traitement informatique : leur acquisition (plan d'expérience, collecte) et leur visualisation (la partie gestion de base de données est, quant à elle, abordée dans le module 8.5). L'objectif est de sensibiliser les étudiants à l'importance de l'organisation de la collecte des données afin d'anticiper au maximum certains risques (échantillons trop petits, données imprécises, manquantes ou biaisées, informations redondantes, etc.).

---

**Teaching guide and syllabus**  
**« Data collection, manipulation and  
visualisation » module**  
**8.1 (4 ECTS credits)**

---

***Subject matter of importance and associated issues***

This module focuses on data within a project, specifically the upstream and downstream stages of data processing: data acquisition (experimental design, data collection), and their visualisation (database management is covered in module 8.5). The objective is to raise students' awareness of the importance of organising data collection and ensuring data quality, taking into account the treatments that will be used: issues such as sample size, imprecise or missing data and redundancy of information.

Responsable : Nicolas Sutton-Charani

Téléphone : 04 34 24 62 67

Courriel : [nicolas.sutton-charani@mines-ales.fr](mailto:nicolas.sutton-charani@mines-ales.fr)

<b>ENSEIGNEMENTS ACADÉMIQUES</b>	<b>Volume horaire</b>	Détail des coefficients	Crédits
<b>Collecte, manipulation et visualisation de données</b>	<b>47 h</b>		
○ Collecte des données et plan d'expériences	35	3	4
○ Validation, Visualisation, Restitution de données	12	1	

**Matière 1 :**

<b>Titre de la matière : Collecte des données et plan d'expériences</b>	
<b>Code : 2IA-8.1.1</b>	<b>Titre du module : «Collecte, manipulation et visualisation de données »</b>
<b>Semestre : S8</b>	<b>Cursus de rattachement : département 2IA</b>

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
35	45	9	0	12	13	1	10	3	3

<b>Résumé</b>	<p>Ce cours a pour objectif de poser les bases de la collecte de données. Depuis le plan d'expérience jusqu'à la collecte physique (questionnaires ou capteurs) ou virtuelle (Web) des données. Il y est, en particulier, question de qualité des données (imprécises, manquantes, aberrantes). La mise en application de ces enseignements dans le projet qui sera proposé dans la suite du module permettra de s'assurer que les étudiants les maîtrisent.</p>
---------------	--

<b>Responsable</b>	Nicolas Sutton Charani – CERIS/IMT Mines Alès
<b>Équipe enseignante</b>	Sébastien Harispe – CERIS/IMT Mines Alès Nicolas Sutton Charani – CERIS/IMT Mines Alès Andon Tchechmedjiev – CERIS/IMT Mines Alès Binbin Xu – CERIS/IMT Mines Alès

<b>Mots-clés</b>	Plan d'expérience, collecte de données, Web scraping, qualité des données
<b>Prérequis</b>	Régression linéaire, Python. Les étudiants seront amenés à utiliser leur ordinateur personnel et doivent donc le prendre à chaque séance.

<b>Contexte et objectif général :</b>	<p>Les enseignements du département 2IA, toutes options confondues, manipulent des données stockées de façon structurée ou non. La qualité des traitements informatiques sur ces données dépend de leur qualité. Cet enseignement pose les bonnes pratiques pour assurer cette qualité.</p>
---------------------------------------	---

<b>Programme et contenu :</b>	<ol style="list-style-type: none"> <li>1. Plan d'expérience</li> <li>2. Questionnaires</li> <li>3. Web scraping et services REST</li> <li>4. Collecte à partir de capteurs</li> <li>5. Qualité statistique des données</li> <li>6. Qualité du format des données / Pandas</li> <li>7. Projet</li> </ol>
-------------------------------	---

<b>Méthode et organisation pédagogique :</b>	<p>Les enseignements sont prévus pour 60 élèves, les séances de TD ou de TP sont réalisées sur ordinateurs personnels. Le découpage approximatif est prévu comme suit, des ajustements étant possibles :</p> <ul style="list-style-type: none"> <li>- Introduction (1h de Cours)</li> <li>- Plan d'expériences (1h de Cours, 1h de TP)</li> <li>- Questionnaires (1h de Cours, 1h de TP)</li> <li>- Web scraping et services REST (2h de Cours-TP)</li> <li>- Collecte à partir de capteurs (6h de Cours-TP)</li> <li>- Qualité statistique des données (1h de Cours, 1h de TP)</li> <li>- Qualité du format des données / Pandas (7h de Cours-TP)</li> <li>- Projet (10h de séance, 3h de soutenance)</li> </ul>
--	---

<b>Acquis d'apprentissage visés :</b>	<p>Mettre en application un plan d'expérience adapté à la problématique et au contexte pour collecter des données auprès d'utilisateurs, sur le Web ou à partir de capteurs. Vérifier et améliorer la qualité de ces données.</p>
---------------------------------------	---

<b>Évaluation :</b> Types d'épreuves et répartition des coefficients :	<ul style="list-style-type: none"> <li>- Un contrôle des connaissances acquises sera effectué via un QCM (coef 1).</li> </ul>
--	---

## Module « Collecte, manipulation et visualisation de données » – 2IA 8.1

<ul style="list-style-type: none"> <li>- Les notions vues dans ce cours seront mises en application dans le projet global du module dont l'évaluation se fera au travers de rapport (coef 2) et de soutenances (coef 1) visant à éclaircir certains points du rapport</li> <li>- Des évaluations non prévues à l'emploi du temps (contrôles surprise) peuvent avoir lieu.</li> </ul>
<b>Retour sur l'évaluation fait à l'élève :</b> 3 semaines après la soutenance : <ul style="list-style-type: none"> <li>- Retour des rapports et soutenances avec commentaires personnalisés</li> </ul>
<b>Support pédagogique et références :</b> Diapositives et références

### Matière 2 :

<b>Titre de la matière : Validation, Visualisation, Restitution de données</b>	
<b>Code : 2IA-iasd-8.1.2</b>	<b>Titre du module : «Collecte, manipulation et visualisation de données »</b>
<b>Semestre : S8</b>	<b>Cursus de rattachement : Département 2IA</b>

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
12	18	3		3	6		6	1	1

<b>Résumé</b>	Un pan important de l'analyse de données concerne la restitution à l'utilisateur des différents résultats obtenus après traitement. Ce cours permet de présenter les différents modes d'affichage visuel et d'outiller les étudiants avec la présentation de librairies logicielles dédiées.
---------------	--

<b>Responsable</b>	Nicolas Sutton-Charani – CERIS/IMT Mines Alès
<b>Équipe enseignante</b>	Nicolas Sutton-Charani – CERIS/IMT Mines Alès

<b>Mots-clés</b>	Visualisation, Restitution, ggplot2, R Markdown, Flexdashboard, Shiny.
<b>Prérequis</b>	Connaissance basique de R.

<b>Contexte et objectif général :</b> Les méthodes d'Intelligence Artificielle et de Science des Données sont toutes basées sur une compréhension de la donnée. Utiliser des méthodes de visualisations appropriées est une compétence indispensable pour mener ce travail à bien, en particulier savoir restituer les données via le bon type de graphique, transmettre les informations pertinentes de façon visuellement efficace, s'adapter à son public (néophyte ou expert), produire des rapports visuels et concis.
<b>Programme et contenu :</b> <ul style="list-style-type: none"> <li>- Introduction à la visualisation de données (3h de cours, 3h de TP)             <ul style="list-style-type: none"> <li>o ggplot2</li> <li>o R Markdown</li> <li>o Interactive plots</li> <li>o Mapping</li> <li>o Flexdashboard</li> <li>o Shiny apps</li> <li>o Projet (6h de TP)</li> </ul> </li> </ul>
<b>Méthode et organisation pédagogique :</b> Les enseignements sont prévus pour 60 élèves en promo complète. Les TP seront réalisés sur les ordinateurs personnels de ces derniers.
<b>Acquis d'apprentissage visés :</b> Connaître les bonnes pratiques de visualisation de données et les différents types de graphique selon les données à présenter et le type d'analyse visée. Savoir les construire à l'aide de R et en particulier grâce à la librairie ggplot2. Être capable de les intégrer dans un tableau de bord ou une application Web.
<b>Évaluation :</b> L'évaluation se fera par un projet: développement d'un tableau de bord ( <i>flexdashboard</i> ) ou d'une application R-shiny mettant en jeu différents types de visualisation de données et à déployer sur le Web.
<b>Retour sur l'évaluation fait à l'élève :</b> 3 semaines après la dernière séance.
<b>Support pédagogique et références :</b> Supports de cours et références (livres, tutoriels).

## Méthode et organisation pédagogique:

Il s'agit d'un enseignement relativement classique avec une partie réalisée en cours magistral et une partie appliquée au travers de TP et projets.

## Modalité d'évaluation

Le niveau d'acquisition des compétences sera évalué selon les exigences suivantes :

N° indicateur	Indicateur
1	connaître les savoirs formels et pratiques du socle des fondamentaux
2	Exploiter les savoirs théoriques et pratiques
3	Analyser, interpréter, modéliser, émettre des hypothèses, et résoudre

## Répartition

Matière	Contrôle	Coefficients	Type de notation	Indicateurs évalués	Chapitres
Plans d'expérience et collecte des données	QCM	1	Individuelle	Tous	Tous
Plans d'expérience et collecte des données	Projet (rapport + soutenance)	3	En trinôme (projet)	Tous	Tous
Visualisation de données	Projet	2	En binôme	Tous	Tous

## Engagement de l'étudiant, éthique et professionnalisme

La démarche éthique est définie dans le règlement intérieur de l'établissement. Chaque étudiant s'engage à en prendre connaissance et à la respecter.

**Obligation des cours** : présence obligatoire pour tous à chaque séance.

**Nombre d'heures estimées de travail personnel** : pour acquérir les compétences demandées, il est nécessaire que l'étudiant consacre minimum 45 min de travail personnel de compréhension et d'approfondissement par séance de cours.  
30h de travail personnel pour le projet.

**Nombre d'heures estimées de préparation aux travaux dirigés (TD)** : 45 min

**Pénalité pour retard** : Tout travail remis en retard sans motif valable peut être pénalisé de 1 point par jour de retard.

## Équipe enseignante

Nom	Domaine d'expertise	Courriel/Téléphone
Nicolas SUTTON-CHARANI	Science de la Donnée, Apprentissage Automatique et Théories de l'Incertain	<a href="mailto:nicolas.sutton-charani@mines-ales.fr">nicolas.sutton-charani@mines-ales.fr</a> 04 34 24 62 67
Andon TCHECHMEDJIEV	Apprentissage automatique, IA et développement informatique	<a href="mailto:andon.tchechmedjiev@mines-ales.fr">andon.tchechmedjiev@mines-ales.fr</a> 04 34 24 62 16
Sébastien HARISPE	Apprentissage Automatique, IA et Représentation des Connaissances	<a href="mailto:sebastien.haripse@mines-ales.fr">sebastien.haripse@mines-ales.fr</a> 04 34 24 62 82
Binbin Xu	Machine learning, AI, and Software development	<a href="mailto:binbin.xu@mines-ales.fr">binbin.xu@mines-ales.fr</a> 04 34 24 62 78

<b>ACADEMIC TEACHING</b>	<b>Teaching hours</b>	Coefficients	Crédits
<b>Data collection, manipulation and visualisation</b>	<b>47 h</b>		
○ Data collection and experimental design	35	3	4
○ Data validation, visualisation, reporting	12	1	

**Class 1**

<b>Class title : Data collection and experimental design</b>	
<b>Code : 2IA-8.1.1</b>	<b>Module title : « Data Collection, manipulation and visualisation»</b>
<b>Semester: S8</b>	<b>Classification : CSAI department</b>

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
35	45	9	0	12	13	1	10	3	3

<b>Summary</b>	<p>This course aims to lay data collection basis. From the design of experiments to the physical (questionnaires or sensors) or virtual (Web) data collection. A special focus is given on data quality (imprecise, missing, outliers).</p> <p>An implementation project, proposed further in the module, will ensure that students are able to put into practice the knowledge provided in this course.</p>
----------------	--

<b>Head</b>	Nicolas Sutton Charani – CERIS/IMT Mines Alès
<b>Teaching team</b>	Sébastien Harispe – CERIS/IMT Mines Alès Nicolas Sutton Charani – CERIS/IMT Mines Alès Andon Tchechmedjiev – CERIS/IMT Mines Alès Binbin Xu – CERIS/IMT Mines Alès

<b>Key words</b>	Design Of Experiment, data collection, web scraping, data quality, outliers
<b>Prerequisites</b>	Linear regression, Python. Students will be brought to use their personal laptops and therefore have to bring them to each session.

**Context and general objective:**  
 The courses in the CSAI department, all options combined, handle data potentially stored in a structured way. The data processing quality depends on the data quality. This course offers the good practice to achieve this quality.

- Programme and contents:**
1. Design Of Experiment (DOE)
  2. Questionnaires
  3. Web-scraping and REST services
  4. Data acquisition from sensors
  5. Statistical data quality
  6. Computer (format) data quality
  7. Project

- Method and pedagogic organisation:**  
 The teachings are planned for 60 students divided into 2 groups for the workshops or Lab session performed on the students' personal laptops.  
 The organisation is planned as follows:
- Introduction (1h lecture)
  - Design of Experiments (DOE) (1h lecture, 1h lab)
  - Questionnaires (1h lecture, 1h lab)
  - Web-scraping and REST services (2h lecture/lab)
  - Sensor collect (6h lecture/lab)
  - Statistical quality of data (1h lecture, 1h lab)
  - Data format quality / Pandas (7h lecture/lab)
  - Project (10h tutored sessions, 3h defences)

**Targeted skills or knowledge :**  
 Implement a design of experiments adapted to a given problem and to its context in order to collect data from users, from the web, or from sensors. Check and improve the resulting data's quality.

- Evaluation :** Testing procedures used and associated coefficient system :
- multiple-choice questionnaire (coef 1)
  - Project: report (coef 2) and defence (coef 1)
  - Unscheduled assessments (surprise checks) may occur.

<p><b>Feedback made to the student :</b> 3 weeks after the defence: i) making MCQ corrections available and ii) reports and defence feedback with personalised comments</p>
<p><b>Teaching material and references :</b> Slides and references</p>

**Class 2**

<b>Class title :</b> Data validation, visualisation, reporting	
<b>Code :</b> 2IA-8.1.2	<b>Module title :</b> « Data collection, manipulation and visualisation»
<b>Semester:</b> S8	<b>Classification :</b> CSAI department

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
12	18	3		3	6		6	1	1

<b>Summary</b>	An important aspect of data analysis is the display of the results obtained after processing. This course introduces the various visual display modes and provides students with the tools they need, with the presentation of dedicated software libraries.
----------------	--

<b>Head</b>	Nicolas Sutton-Charani – CERIS-I3A/IMT Mines Alès
<b>Teaching team</b>	Nicolas Sutton-Charani – CERIS-I3A/IMT Mines Alès

<b>Key words</b>	Data visualisation, Restitution, ggplot, R Markdown, Flexdashboard, Shiny
<b>Prerequisites</b>	Basic knowledge of R

<p><b>Context and general objective:</b> Artificial Intelligence and Data Science methods are all based on an understanding of data. Using appropriate visualisation methods is an essential skill for any analysis, in particular knowing how to render data using the right graphic type, convey relevant information in a visually effective way, adapt to your audience (whether neophytes or experts), and produce visual and concise reports.</p>
<p><b>Programme and contents:</b></p> <ul style="list-style-type: none"> <li>• Introduction to data visualisation (3h lecture, 3h lab)             <ul style="list-style-type: none"> <li>○ Ggplot2</li> <li>○ R Markdown</li> <li>○ Mapping</li> <li>○ Flexdashboards</li> <li>○ Shiny apps</li> <li>○ Project (6 hours)</li> </ul> </li> </ul>
<p><b>Method and pedagogic organisation:</b> The course will be organised as follows for 60 students. Practical work will be carried out on the students' personal computers.</p>
<p><b>Targeted skills or knowledge :</b> Knowledge of the best practices for data visualisation and the different types of graphs depending on the data to be presented and the type of analysis targeted. Know how to build them using R, and in particular the ggplot2 library. Integrate them into a dashboard or Web application.</p>
<p><b>Evaluation :</b> Evaluation will be carried out through a project: development of a dashboard (flexdashboard) or an R-shiny application featuring different types of data visualisation, to be deployed on the Web.</p>
<p><b>Feedback made to the student :</b> 3 weeks after the defence</p>
<p><b>Teaching material and references :</b> books and tutorials</p>

**Method and teaching organisation:**

This is a classical course containing a theoretical part with standard courses and a practical one through lab sessions and a project.

## Testing procedures

The student's level of knowledge acquisition will be evaluated according to the following points:

N° Indicator	Indicator
1	To know the formal and practical knowledge constituting the foundation of a given field
2	Exploit theoretical and practical knowledge
3	Analyse, interpret, model, hypothesise and solve problems

*Grading scheme:*

Class	Exam	Coefficients	Administration mode	Evaluated Indicators	Chapters
Data collection and experimental design	multiple-choice questionnaire	1	Individual (multiple-choice questionnaire)	all	all
Data collection and experimental design	Project evaluation (report + defence)	3	Groups of three (project)	all	all
Data visualisation	Graded lab	2	Groups of two	all	all

## Student commitments, ethics and professionalism

Expectations concerning ethics are defined in the establishment's code of conduct. Each student is expected to know and respect the code of conduct.

**Obligatory presence in class:** Students must attend all courses, seminars and labs.

**Estimated hours of personal study:** in order to acquire the required learning level, the student is expected (must) to spend a minimum of 45 min of personal study time per hour spent in class. 30h of personal work for the project

**Estimated hours of preparation required for labs/Work Shop:** 45 min

**Late penalties:** Late works are subject to penalties as follows: 1 point per day (ratings between 0 and 20).

## Teaching team

Name	Field of expertise	Email/phone
Nicolas SUTTON-CHARANI	Data Science, Machine Learning and soft computing	<a href="mailto:nicolas.sutton-charani@mines-ales.fr">nicolas.sutton-charani@mines-ales.fr</a> 04 34 24 62 67
Andon TCHECHMEDJIEV	Machine Learning, AI and programming	<a href="mailto:andon.tchechmedjiev@mines-ales.fr">andon.tchechmedjiev@mines-ales.fr</a> 04 34 24 62 16
Sébastien HARISPE	Machine Learning, AI and Knowledge Representation	<a href="mailto:sebastien.haripse@mines-ales.fr">sebastien.haripse@mines-ales.fr</a> 04 34 24 62 82
Binbin XU	Machine learning, AI, and Software development	<a href="mailto:binbin.xu@mines-ales.fr">binbin.xu@mines-ales.fr</a> 04 34 24 62 78

## Approbation

Ce guide pédagogique entré en vigueur à compter du 02 janvier 2024. Il est porté à la connaissance des élèves par une publication sur le site de l'école

Rédaction	Vérification	Validation
L'enseignant responsable du module : Nicolas Sutton-Charani	La responsable de département : Sylvie Ranwez	Pour le directeur et par délégation, Le directeur des études : Michel Ferlut