

DESCRIPTIF DE L'UE

Code APOGEE à venir	<b>Formation de rattachement</b>	<b>Intitulé de l'enseignement</b>
	<b>L3 Université de Strasbourg – UE rattachée à l'UFR de Mathématique et Informatique</b>	<b>Initiation à la science des données</b>

**RESPONSABLES DE L'ENSEIGNEMENT**

<b>Nom/Prénom</b>	<b>Discipline</b>	<b>Composante / Courriel/Tél. (fixe)</b>
MAUMY-BERTRAND MYRIAM	26	UFR Mathématique et Informatique/ <a href="mailto:mmaumy@math.unistra.fr">mmaumy@math.unistra.fr</a> /03 68 85 02 30
MEILLIER CELINE	61	Télécom Physique Strasbourg/ <a href="mailto:meillier@unistra.fr">meillier@unistra.fr</a> / 03 68 85 45 70

**MODALITES D'ORGANISATION DE L'ENSEIGNEMENT**

Semestre 5 ou 6 Licence / 2 heures par semaine. - UE Libre ou intégrée dans le Semestre 5 ou 6 de chaque composante.

**LANGUE ET NIVEAU DE LANGUE DE L'ENSEIGNEMENT :** FRANÇAIS (SUPPORT EN FRANÇAIS)

## DESCRIPTION DE L'ENSEIGNEMENT

Dans un contexte scientifique, sociétal et économique où stockage, archivage, traitement et valorisation des données sont devenus des enjeux majeurs, cette U.E. permettra de découvrir les bases de la science des données, c'est-à-dire les méthodes et outils, théoriques et pratiques, pour comprendre les données : analyse, visualisation, apprentissage de modèles... Les cours seront faits en salle informatique, avec un poste de travail par étudiant et des manipulations fréquentes sur des exemples. Toutes les notions du cours seront illustrées à partir d'une base de données sur le cinéma.

### Prérequis :

Aucune expérience en programmation n'est requise, cependant est attendu des étudiants un intérêt et une curiosité pour les outils informatiques et les notions élémentaires de statistiques (moyenne, écart-type...).

### Plan du cours :

Initiation au notebook Python

Introduction – les différents types de données

Représentation statistique des données

- A) Représentation numérique des données
- B) Représentation graphique des données
- C) Importance de la représentation graphique

Exploration des données

- A) Recueil des données
- B) Prétraitements
- C) Relation entre les variables

Modélisation et prédiction

Mise en pratique des connaissances à travers un projet encadré permettant d'approfondir les connaissances introduites dans les supports de cours.

### Webographie :

Jeux de données et challenges : <https://www.kaggle.com>

Package Python dédié à la science des données : <http://scikit-learn.org> et <https://pandas.pydata.org>

Package Python dédié à la visualisation des données : <https://seaborn.pydata.org>

### Bibliographie :

Data science : fondamentaux et études de cas. Machine learning avec Python et R, Eric Biernat et Michel Lutz, Eyrolles.

Big data et machine learning - Le concept et les outils de la data science, 2ème édition, Pirmin Lemberger, Marc Batty, Médéric Morel, Jean-Luc Raffaelli, Dunod.

## COMPETENCES VISEES

### Objectifs en termes de connaissances

- Contrôler la qualité, la cohérence, la fiabilité, la représentativité des données à traiter.

- Savoir transformer des données pour les rendre utilisables par un algorithme de traitement de données.
- Expliquer et interpréter un jeu de données à partir des résultats fournis par les outils de prédiction, covariance, etc.
- Résumer et expliquer un jeu de données à ses pairs.

### Objectifs en termes de compétences

#### Compétences disciplinaires :

- Lire des données statistiques sous différentes formes, interpréter un tableau croisé, pour produire un document de synthèse de données statistiques.
- Se servir aisément des outils et méthodes de recueil, de traitement et d'analyse des données pour observer et analyser les phénomènes et/ou les comportements du sous domaine.
- Concevoir le traitement informatisé d'informations de différentes natures, telles que des données, des images et des textes.
- Exploiter des logiciels d'acquisition et d'analyse de données avec un esprit critique.
- Mettre en œuvre des techniques de programmation, par l'exemple, notamment pour développer des applications simples d'acquisition et de traitements de données.
- Analyser et interpréter les résultats produits par l'exécution d'un programme.
- Mobiliser les outils mathématiques nécessaires à la modélisation.
- Valider un modèle par comparaison de ses prévisions aux résultats expérimentaux et apprécier les limites de validité.

#### Compétences transversales :

- Analyser et synthétiser des données en vue de leur exploitation.
- Développer une argumentation avec esprit critique.

### ENSEIGNEMENT – ORGANISATION HORAIRE

Titre de l'enseignement	CM	CI	TD	TP
Initiation à la science des données		28h		
Eq. HETD= 35h				

### MODALITES D'EVALUATION DES ENSEIGNEMENTS

#### *En session principale*

Moyenne des évaluations écrites multiples (comptant pour 50% de la note finale) réalisées en séances au cours du semestre. Pas de convocation.

Évaluation (qui comptera pour 50% de la note finale) sous la forme d'un rendu de projet (notebook commenté permettant le traitement et l'analyse du jeu de données fourni en projet). Pas de convocation, rendu à la date fixée.

**Pas de session de rattrapage.**