

SDS 3786 Laboratoire en sciences des données

Automne 2024 | Instructeur: P. Boily

DESCRIPTION DU COURS

La gestion et l'analyse des données. Un cours de laboratoire et projet incluant: la manipulation et le nettoyage de données; l'analyse exploratoire et la visualisation de données; méthodes de régression et de classification; évaluation de résultats et écriture de rapports.

NOTES DE COURS

P.Boily, **Data Understanding, Data Analysis, & Data Science**, DAL/QED.

ÉVALUATION ET COMMENTAIRES

Labos (60%); **projet** (30%); **participation** (10%).

Il y a 8 laboratoires – je compte les 7 meilleurs:

1. Les aspects non-techniques de l'analyse des données (9–23 sept)
2. La visualisation des données (24–30 sept)
3. Le traitement des données (1–7 oct)
4. La grammaire des graphiques (22–28 oct)
5. La mise en récit de données (29 oct–4 nov)
6. L'apprentissage automatique et les règles d'association (5–18 nov)
7. La classification/régression (19 nov–2 déc)
8. Le regroupement (19 nov–2 déc)

Vous travaillerez également sur un projet de science des données de votre choix:

- Choix d'équipe: 20 sept
- Proposition: 21 oct [5]
- Rapport de progrès: 11 nov [5]
- Présentation finale: à déterminer [10]
- Rapport final et portefeuille: à déterminer [10]

Je m'attends à ce que vous consacriez en moyenne **8-10 heures** par semaine à ce cours.

L'**engagement** et le **travail d'équipe productif** sont essentiels dans ce cours: vous devez être présents (dans les limites de la raison, nous en discuterons en classe) et actifs pour obtenir vos points de participation.

HORAIRE DU COURS

LAB: LUN 14:30-17:20 (LMX 241)

Les vidéos, les diapositives, et les énoncés de laboratoires sont disponibles sur le **site web du cours**.

IMPORTANT: Vous n'aurez pas le temps de réaliser les laboratoires si vous ne vous êtes pas préparé en regardant les vidéos et/ou en lisant les notes de cours avant le jour du laboratoire (c'est la raison pour laquelle je mets les PDF à disposition en premier lieu).

Vous devez vous présenter à chaque laboratoire avec un plan d'action.

Toutes les communications relatives au cours ont lieu sur **Slack** (pas de courriels, SVP!).

PRÉ-REQUIS

ITI 1520 + ((MAT 2771 + 2775) ou MAT 2777).

Je m'attends à ce que vous maîtrisiez leur contenu (ce cours sera beaucoup plus facile si vous êtes à l'aise avec la programmation).

NOTES

Vous devez avoir accès à un ordinateur sur lequel vous pouvez installer R, RStudio, Python, et/ou Jupyter, ainsi que divers paquets et modules au fur et à mesure des besoins.

Je vous invite à vous familiariser avec le **Règlement académique A-4 sur l'intégrité académique et inconduite académique**.