

Programme

Introduction à la science des données

Formateur: P. Boily

Cette formation est un survol de la discipline – les compétences en programmation et en statistique ne sont ni requises, ni essentielles. Les exercices se feront à l'aide d'un logiciel en ligne disponible gratuitement (aucune installation requise).

Chaque module dure 2 heures. On peut retrouver le matériel de courau <https://data-action-lab.com/isd>.

1^{er} module

Principes fondamentaux de l'analyse des données

Planification des analyses; données 101: notions de données de base; quelques définitions pratiques; flux de travail et sources; modèles et pensée systémique; considérations éthiques et meilleures pratiques; exercices.

2^e module

Collecte et gestion des données

Sources des données; théorie de l'échantillonnage et planification d'étude; moissonnage du web et collecte automatisée de données; modélisation des données et des connaissances; exercices.

3^e module

Principes de la visualisation des données

Visualisation des données avant l'analyse; visualisation des données post-analyse; catalogue des visualisations; règles de base de la conception et de la mise en page; tableau de bord; exercices.

4^e module

Traitement des données

Nettoyage des données; réduction et transformation des données; qualité de données et validation des données; exercices.

5^e module

Techniques de base d'analyse des données

Aperçus et calculs : concepts et techniques de base; descriptions de données; corrélation; analyse de régression; séries chronologiques et cartes de contrôle; exercices.

6^e module

Apprentissage statistique

Contexte d'apprentissage; règles d'association; classification; regroupement; enjeux et défis; exercices.