

ENG-209

Data science pour ingénieurs avec Python

Cours 6 - Lab

Jean-Philippe Pellet, Éric Bouillet, Olivier Verscheure

27 Octobre 2025

Introduction à la modélisation: comprendre et visualiser les données

- **Objectif de cette séance pratique**

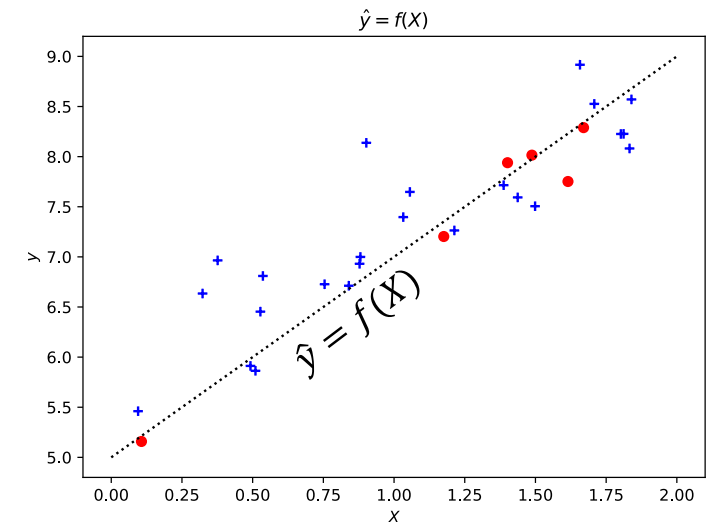
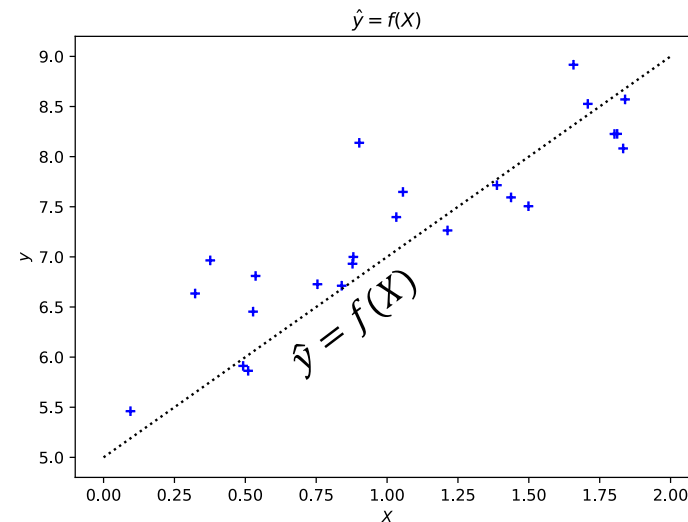
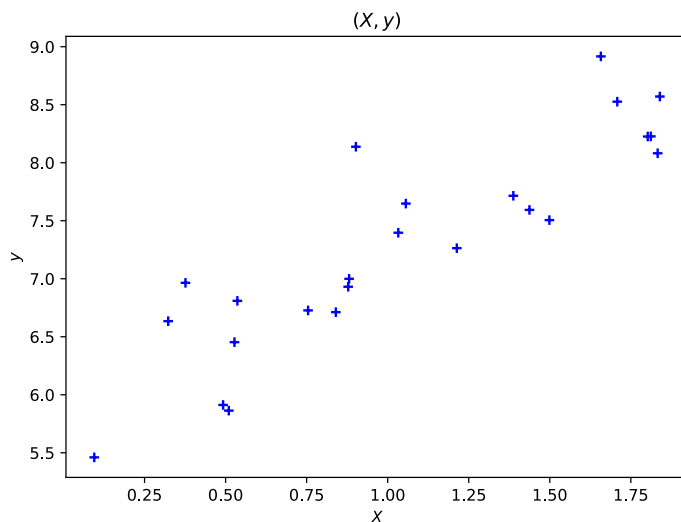
- ✦ Découvrir la démarche de la Data Science
- ✦ Comprendre le rôle central de la visualisation des données
- ✦ Se familiariser avec pandas, numpy, matplotlib

Sous réserve d'adaptations et modifications

L'apprentissage machine – en une phrase

- "Apprendre une fonction qui minimise l'erreur"

- ◆ On cherche une fonction $f(X)$ qui prédit y
- ◆ On ajuste f pour minimiser l'erreur sur les données d'entraînement
- ◆ Puis on teste sur de nouvelles données



La réalité : le modèle, c'est 20 % du travail

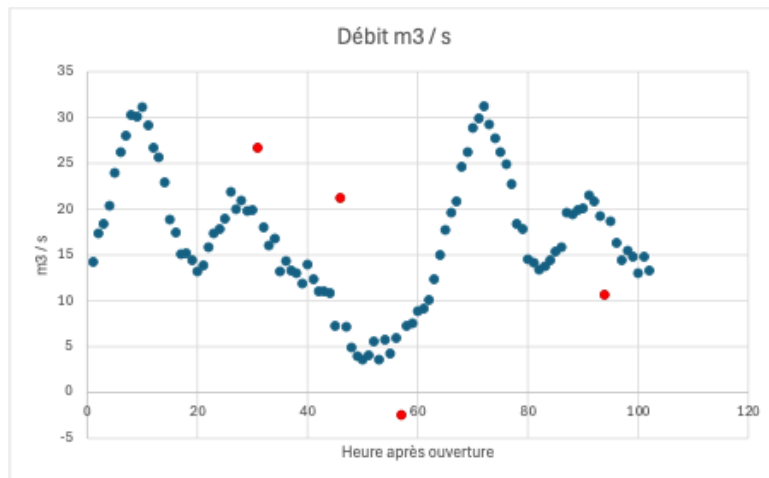
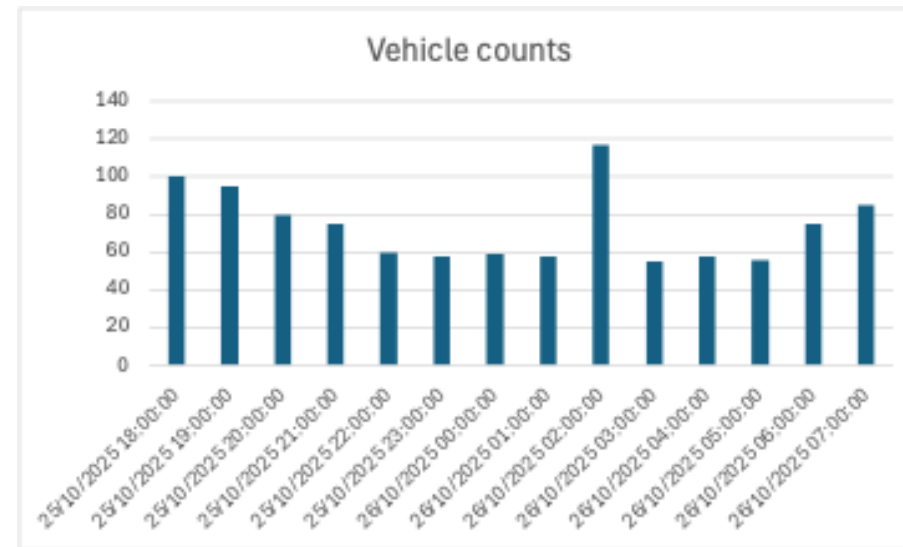
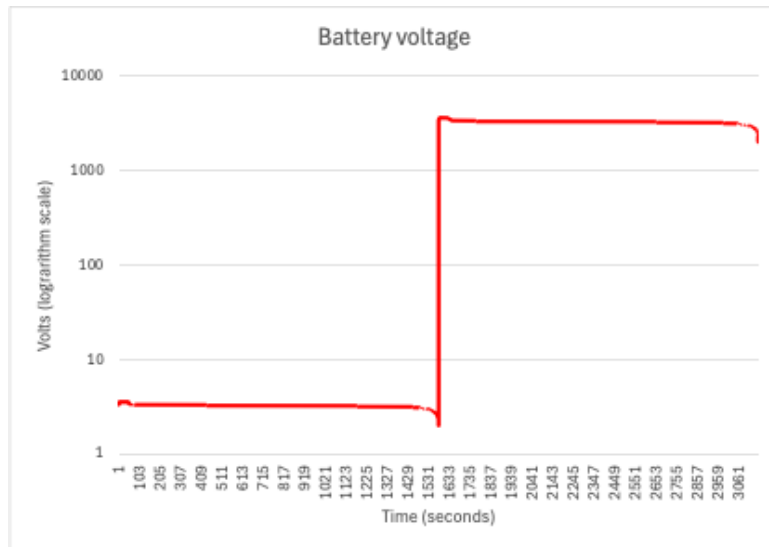
- **Le vrai travail commence avant le modèle**
 - ✦ Comprendre et nettoyer les données brutes = 70 % du temps
 - ✦ 20 % modélisation
 - ✦ 10 % validation, etc.

Les données brutes

- **D'où viennent les données ?**

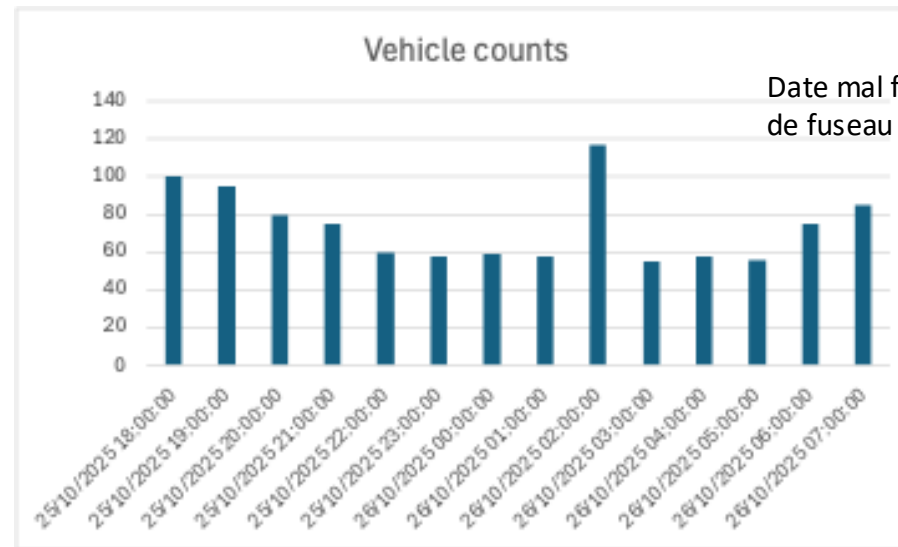
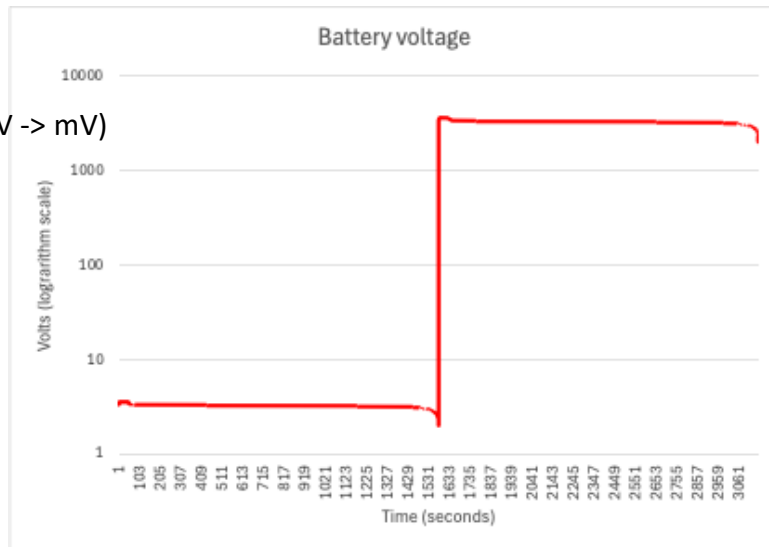
- ✦ Capteurs, logs, bases SQL, fichiers Excel, APIs...
- ✦ Souvent : incohérences, unités différentes, valeurs manquantes

Exemples typiques d'erreurs ou anomalies

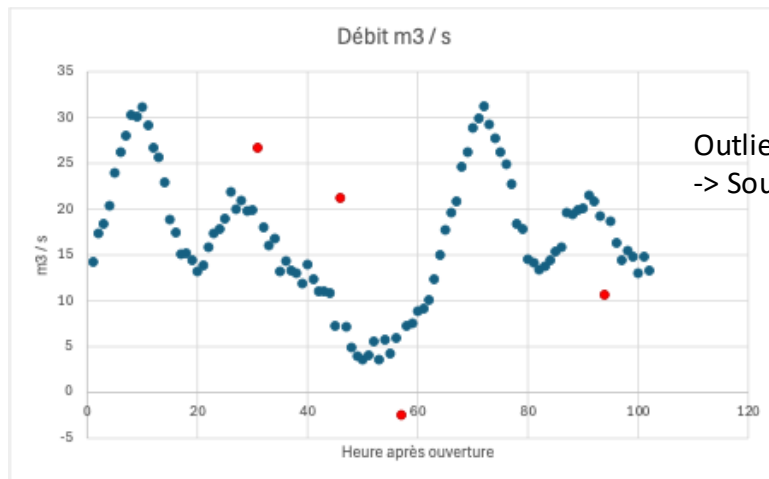


Exemples typiques d'erreurs ou anomalies

Capteur réinitialisé,
changement d'unité (V -> mV)



Date mal formatée / décalage
de fuseau horaire



Outliers extrêmes : erreur ou phénomène réel ?
-> Souvent il faut l'expertise métier pour trancher.

Explorer les données

- **Première étape : comprendre la structure**
 - ✦ Statistiques descriptives, e.g. `pandas.DataFrame.describe()`
 - ✦ Types de variables : numériques, catégorielles, temporelles
 - ✦ **Visualiser** pour comprendre les distributions

Visualiser les relations

- **Identifier corrélations et interactions**

- ✦ Scatter plots, heatmaps de corrélation
- ✦ Détecter dépendances, non-linéarités, redondances
- ✦ Base de la **sélection et de la transformation** des caractéristiques X

Interpréter, pas décorer !

- **Chaque graphique doit répondre à une question**
 - ✦ Ne pas tracer “pour faire joli”
 - ✦ Chaque figure doit **avoir un message clair**
 - ✦ Si vous ne pouvez pas l’interpréter, vous perdez votre temps

Démarche type d'un projet

- **Les 5 étapes clés**

- ✦ Comprendre le problème métier
- ✦ Explorer et visualiser les données
- ✦ Nettoyer et transformer
- ✦ Sélectionner les variables
- ✦ Construire et tester le modèle