

Étude de cas

Projet de fin de module : Créer un modèle statistique de
A à Z

A DEPOSER ICI : <https://forms.gle/hxsLcszy6sX5ppky8>

Date de fin de rendu : 05/05/2025

-3 points par jour de retard.

PROJET A FAIRE EN BINOME

Contexte – Ton rôle

Tu es analyste apprenant en intelligence artificielle.

Une entreprise t'a confié une mission : **comprendre et prédire l'absentéisme des salariés** à partir de données RH.

Tu vas devoir structurer ta réflexion comme un chef de projet IA :

- Depuis le besoin métier
- Jusqu'à l'interprétation des résultats d'un modèle statistique

Objectif du projet



Tu dois construire un **modèle statistique complet** (régression ou arbre) pour :

- **Prédire l'absentéisme** ou sa probabilité,
- **Comprendre les facteurs principaux** qui influencent les absences,
- Et **formuler des recommandations claires** à la direction RH.

Données à utiliser

Fourni avec le projet

Livrables attendus

Livrable	Format	Objectif
 Notebook Python	.ipynb	Analyse + modélisation + commentaires
 Mini rapport (2 pages)	.pdf ou .docx	Résumé du projet, résultats clés, recommandations

Étapes à suivre (obligatoires)

1. Compréhension du problème

- Reformule le besoin métier avec tes mots : pourquoi l'entreprise veut-elle ce modèle ?
- Précise ce que tu vas prédire (variable cible) et quel type de problème statistique c'est (régression ou classification ?).

2. Analyse exploratoire

À l'aide de Python et pandas/seaborn :

- Calcule les **statistiques descriptives** (moyenne, médiane, min, max, etc.),
- Fais des **histogrammes, boxplots et nuages de points** sur les variables clés,
- Si possible, fais une **ACP** pour observer les profils globaux,
- Rédige une courte synthèse : que remarques-tu ? Quelles relations possibles vois-tu ?

3. Sélection des variables

- Choisis entre **3 et 5 variables explicatives** pertinentes,
- Justifie ton choix avec les données, les corrélations, ou le bon sens métier.

4. Modélisation

- Entraîne un ou plusieurs modèles :
 - **Régression linéaire** si tu veux prédire le nombre d'absences,
 - **Régression logistique** si tu veux prédire si quelqu'un s'absentera ou non,
 - **Arbre de décision** si tu veux expliquer facilement.
- Commente ton code : que fait chaque étape ?

5. Évaluation du modèle

- Choisis les bons indicateurs (MAE, Accuracy, courbe ROC...) selon ton modèle.
- Explique ce que veulent dire ces scores.
- Compare plusieurs modèles si possible.

6. Interprétation et recommandations

- Quelles **variables sont les plus importantes** ?
- Donne **des exemples concrets** pour la RH :

« Les salariés en CDD avec peu d'ancienneté et un long trajet sont les plus à risque. »

- Fais une **synthèse visuelle claire** (graphique, tableau, phrase forte).