

# Fiche d'auto évaluation

## Traiter

### Apprentissages critiques développés :

- Définir des critères pertinents pour le nettoyage des données.
- Gérer les valeurs manquantes avec des méthodes adaptées en fonction du contexte.
- Manipulation plus avancée de certaines bibliothèques python tel que Numpy, Pandas, Matplotlib et Seaborn.

**Auto-évaluation :** Sur cette compétence, je pense avoir acquis une bonne maîtrise. J'ai su appliquer un nettoyage rigoureux et justifié, en m'appuyant sur des critères clairs comme la suppression des colonnes avec plus de 80 % de valeurs manquantes, ou l'usage de la médiane plutôt que la moyenne pour éviter les effets des outliers. Mais j'ai également affiné ma démarche de travail qui consiste à tester différentes méthodes et garder la plus robuste lorsqu'on ne sait pas quel effet aura cette action sur nos résultats. Cette étape, m'a permis de mieux comprendre l'importance d'un nettoyage bien réfléchi.

## Analyser

### Apprentissages critiques développés :

- Utilisation de la régression linéaire pour détecter et éliminer les variables fortement corrélées.
- Réalisation et interprétation d'ACP pour la réduction de dimension.
- Comparaison critique entre deux approches analytiques.

**Auto-évaluation :** J'ai fait de réels progrès sur cette compétence. La phase d'analyse a été très intéressante, notamment lors de l'analyse en composantes principales et de la comparaison des deux approches. J'ai appris à comparer deux ACP en évaluant la variabilité expliquée, la représentation des variables, et la qualité de représentation des individus. J'ai compris que l'élimination a priori des variables corrélées, que je pensais bénéfique, c'est finalement avéré néfaste.

## Valoriser

### Apprentissages critiques développés :

- Capacité à interpréter des résultats.
- Explication des limites et des choix.
- Mise en valeur des résultats pertinents grâce à des visuels (graphiques) au lieu de tableaux qui sont illisibles
- Rédaction d'un notebook « propre » afin qu'il soit lisible et compréhensible
- Mise en ligne du travail : repos Git Hub

**Auto-évaluation :** Je pense avoir acquis de solides compétences en interprétation et valorisation des résultats en data science. Tout au long du projet, j'ai su analyser les résultats obtenus et en tirer des conclusions pertinentes. Par exemple, j'ai su expliquer pourquoi l'ACP1 apportait une meilleure

visualisation des données, en m'appuyant sur des indicateurs concrets. De plus, j'ai été capable d'identifier les limites de certaines méthodes, comme l'instabilité du K-Means en présence d'outlier. J'ai également pu mettre en valeur les résultats importants à l'aide de graphiques, préférés aux tableaux bruts souvent difficiles à lire. Le notebook final a été rédigé de manière claire et structurée, dans le but d'être facilement compréhensible par un lecteur extérieur et tout le code est commenté, afin de le rendre compréhensible. Enfin, j'ai mis le projet en ligne (sur mon Git Hub).

## **Modéliser**

### **Apprentissages critiques développés :**

- Mise en œuvre de modèles de machine learning (régression multiple, K-Means, Random Forest, CAH).
- Choix justifié du modèle en fonction des résultats et des contraintes.
- Évaluation des performances à l'aide de métriques appropriées.

**Auto-évaluation** : Je pense avoir beaucoup progressé sur cette compétence. J'ai expérimenté plusieurs modèles (K-Means, Random Forest, Classification Automatique Hierarchisée), et j'ai appris à choisir le plus pertinent en fonction du contexte et des résultats. J'ai aussi été confronté aux limites des modèles, notamment face aux données déséquilibrées, ce qui m'a poussé à ajuster notre stratégie. Cette approche itérative, qui consiste à essayer plusieurs solutions et ajuster les paramètres de nos modèles, a renforcé ma compréhension des enjeux de la modélisation. J'ai également appris à utiliser le package Scikit-learn en python, qui est un super outil pour faire du Machine Learning.

D'un point de vue personnel, j'ai vraiment apprécié ce projet. Il m'a permis d'explorer différentes idées et approches, de tester, comparer, puis ajuster mes choix en fonction des résultats. Au-delà de l'aspect technique, ce projet m'a aussi montré à quel point la qualité des données est importante : elles ne sont pas toujours propres, et il est donc important de bien comprendre leur contexte avant de les nettoyer ou de les exploiter. Cette phase de réflexion en amont est tout aussi importante que la modélisation elle-même, et c'est ce qui rend ce projet super intéressant.