# **Exercice 1**

L'objectif de cet exercice est de réaliser un modèle de machine learning (IA) basé sur la régression logistique capable de prédire si un patient à une maladie cardiaque ou non. Pour cela, nous mettons à votre disposition un jeu de données de 918 patients .

1. Charger le fichier de données **Coeur.xlsx** dans un objet **data**.
2. Copiez l’objet **data** dans un nouvel objet **df** avec lequel vous allez travailler pour le reste de l’exercice.
3. Vérifiez s’il y a des doublons dans l’objet **df** ; si c’est le cas, supprimez-les.
4. Vérifier s’il y a des constantes (variable avec une seule modalité) dans **df** ; si c’est le cas, supprimez les.
5. Vérifier s’il y a des données manquantes dans **df** ; si c’est le cas, supprimez les.
6. Faite une normalisation de toutes les variables quantitatives en divisant les valeurs de chaque colonne par la moyenne de celle-ci. Attention, la variable cible (**COEUR**) n'est pas à normaliser.
7. Faite un recodage de toutes les variables qualitatives.
8. On souhaite séparer la variable cible **COEUR** des autres variables de votre objet **df**. Pour cela, stocker la variable **COEUR** dans un objet appelé **Y** et les autres variables dans un objet appelé **X**.
9. En utilisant la fonction **train\_test\_split** du module **model\_selection** du package **sklearn**, subdiviser vos données **X** et **Y** en respectivement **X\_train**, **X\_test** et **Y\_train**, **Y\_test**. **X\_train** et **Y\_train** doivent contenir respectivement 80 % des données de **X** et **Y**.
10. En utilisant la classe **LogisticRegression** du module **linear\_model** du package **sklearn**, construisez un modèle de régression logistique en utilisant les données **X\_train** et **Y\_train** et la méthode **fit** de la classe **LogisticRegression**.
11. En utilisant les données de **X\_test** et la méthode **predict\_proba** de la classe **LogisticRegression** affichez la probabilité d'appartenance à chacune des classes de la variable **COEUR**.
12. En utilisant les données de **X\_test** et la méthode **predict** de la classe **LogisticRegression** prédisez la classe d'appartenance de chaque individu de **X\_test**.
13. En utilisant les données de **X\_test** et **Y\_test** et les méthodes **confusion\_matrix, accuracy\_score**, **recall\_score**, et **precision\_score**, du module **metrics** du package **sklearn**, affichez respectivement la matrice de confusion, le taux succès, la sensibilité et la précision du modèle.