

On considère le dataset `titanic.csv`, qui contient des informations sur les passagers, en particulier s'ils ont survécu au naufrage.

L'objectif de ce travail est double: visualisation et modélisation des données en rapport avec la survie ou non du passager.

Exercice 1 Dans un premier temps, vous nettoierez les données:

- *élimination des colonnes pour lesquelles beaucoup de données sont manquantes, ou peu pertinentes.*
- *élimination des lignes contiennent encore des données manquantes après élimination des colonnes*

Exercice 2 Dans un second temps, vous proposerez différents types de visualisation qui vous semblent appropriés, l'objectif étant de repérer visuellement des motifs de corrélation entre les différentes informations.

Exercice 3 Enfin, vous proposerez un modèle permettant de prédire la survie ou non d'un passager en fonction des données présentes dans la table. Vous prendrez soin de séparer vos données en échantillon d'apprentissage et échantillon de test pour évaluer votre modèle. Vous donnerez ainsi une estimation de la précision de votre modèle (proportion de bonnes prédictions dans l'échantillon de test).