

TP : Généralités

Vous pouvez faire les exercices dans le langage de votre choix.

1 Exercice 1 : Nettoyage et exploration de données

1. Télécharger le jeu de données suivant : [lien](#)
2. Nettoyer le jeu de données. En particulier, on s'intéressera aux points suivant :
 - encodage des valeurs manquantes ;
 - gestion des valeurs extrêmes ;
 - gestion des valeurs abberantes ;
 - gestion des doublons ;
 - harmonisation des dates ;
 - suppression des tirets, points, ... dans les numéros de téléphone.
 - gestion des titres (Mr., Ms., Dr., ...) dans les noms ;
 - harmonisation des pays ;
 - conversion des Oui/Non en TRUE/FALSE ;
 - etc.
3. Faire une exploration unidimensionnelle/bidimensionnelle. En particulier, on pourra faire :
 - graphiques des données quantitatives ;
 - calcul de corrélations ;
 - tableaux de fréquences pour les variables qualitatives ;
 - etc.