

Projet : séance 1 et séance 2

Pour la séance 2, vous devez :

- Proposer un dataset contenant suffisamment de données et le faire valider par l'enseignant
- Vous devez faire la description des données contenues dans le dataset :
 - Le nombre de lignes, de colonnes
 - Les informations dans le dataset (sens des colonnes...)
 - La nature des données
 - Le format (wide ou long)
 - Faites un premier script de nettoyage des données :
 - identifiez les données manquantes et proposez des solutions ;
 - vérifiez les colonnes avec des catégories ;
 - analysez les données numériques pour trouver et corriger les données aberrantes ;
 - enlevez les données dupliquées ;
 - vérifiez les intervalles des valeurs par rapport à la cohérence du contenu ;
 - vérifiez les colonnes de textes ;
 - vérifiez le contenu des colonnes liées ;
 - normalisez les valeurs numériques si besoin.

Une fois cela effectué, proposez les variables clés que vous souhaitez analyser et les graphiques/tables que vous souhaitez mettre en œuvre.