

Consignes pour le TD Clustering

Lecture des données

Lire sous forme de dataframe le fichier csv suivant

https://raw.githubusercontent.com/sebastienQNL/IRIAF_kmeans_2021/main/CC%20GENERAL.csv . Le séparateur est la virgule.

Traitement des valeurs manquantes

Identifier les valeurs manquantes et mettre en oeuvre une stratégie de traitement.

Réalisation d'une ACP

- standardiser les données
- afficher le screeplot pour définir le nombre d'axes à retenir
- afficher les variables les plus contributrices sur chaque axe
- afficher le graphe des individus sur les axes 1 et 2

Clustering

- à partir des coordonnées des individus sur les axes, réaliser un clustering (méthode à votre convenance)
- choisir le nombre de cluster
- afficher le cluster sur les deux premières dimensions de l'ACP
- expliquer les clusters