

PROJET FINALE

But :

- 1) Analyser le jeu de données et faire ressortir les éléments marquant de votre analyse (des graphiques, des corrélations etc).
- 2) Après analyse et éventuelles modifications de la base de données (suppression sur **justification** des variables « inutiles » ou individus « inutiles », suppression des doublons, modifications de variables, création de variables ...)
- 3) Faire une ACP et **interpréter**.
- 4) Classification et **interprétation** de la classification (je vous conseille une classification mixte et aussi de visualiser vos clusters sur les plans factoriels de l'ACP)

Conseils pour le pré traitement des données :

- Supprimer les joueurs en doublon
- Ne garder que les joueurs ayant significativement joué sur la saison
- Re travaillé le poste des joueurs pour n'avoir que 4 modalités :
FW = attaquant, MF=Milieu de terrain, GK= goal, DF=défenseurs
- Mettre le nom des joueurs en row.names

Rendre un rapport PDF ou HTML (Rmarkdown) pour vendredi 24 mai, minuit.

Liste des variables :

- Player : Player's name
- Pos : Position
- Squad : Squad's name
- Age : Player's age
- Born : Year of birth
- MP : Matches played
- Min : Minutes played
- Goals : Goals scored or allowed
- Shots : Shots total (Does not include penalty kicks)
- SoT : Shots on target (Does not include penalty kicks)
- ShoDist : Average distance, in yards, from goal of all shots taken (Does not include penalty kicks)
- PasTotCmp% : Pass completion percentage
- PasTotDist : Total distance, in yards, that completed passes have traveled in any direction
- PasShoCmp% : Pass completion percentage (Passes between 5 and 15 yards) : short passes
- PasMedCmp% : Pass completion percentage (Passes between 15 and 30 yards) : medium passes
- PasLonCmp% : Pass completion percentage (Passes longer than 30 yards) : long passes
- PasAss : Passes that directly lead to a shot (assisted shots = passes D)
- PPA : Completed passes into the 18-yard box (surface de réparation)
- PasAtt : Passes attempted

- CK : Corner kicks
- SCA : Shot-creating actions
- GCA : Goal-creating actions
- Tkl : Number of players tackled
- Blocks : Number of times blocking the ball by standing in its path
- Int : Interceptions
- Clr : Clearances (Dégagement)
- Err : Mistakes leading to an opponent's shot
- Touches : Number of times a player touched the ball. Note: Receiving a pass, then dribbling, then sending a pass counts as one touch
- ToAtt : Number of attempts to take on defenders while dribbling
- Carries : Number of times the player controlled the ball with their feet
- CPA : Carries into the 18-yard box (surface de réparation)
- Rec : Number of times a player successfully received a pass
- CrdY : Yellow cards
- CrdR : Red cards
- Off : Offsides (hors-jeu)
- Crs : Crosses (centre : passe longue ou moyenne vers le centre du terrain vers le but de l'adversaire)
- Recov : Number of loose balls recovered