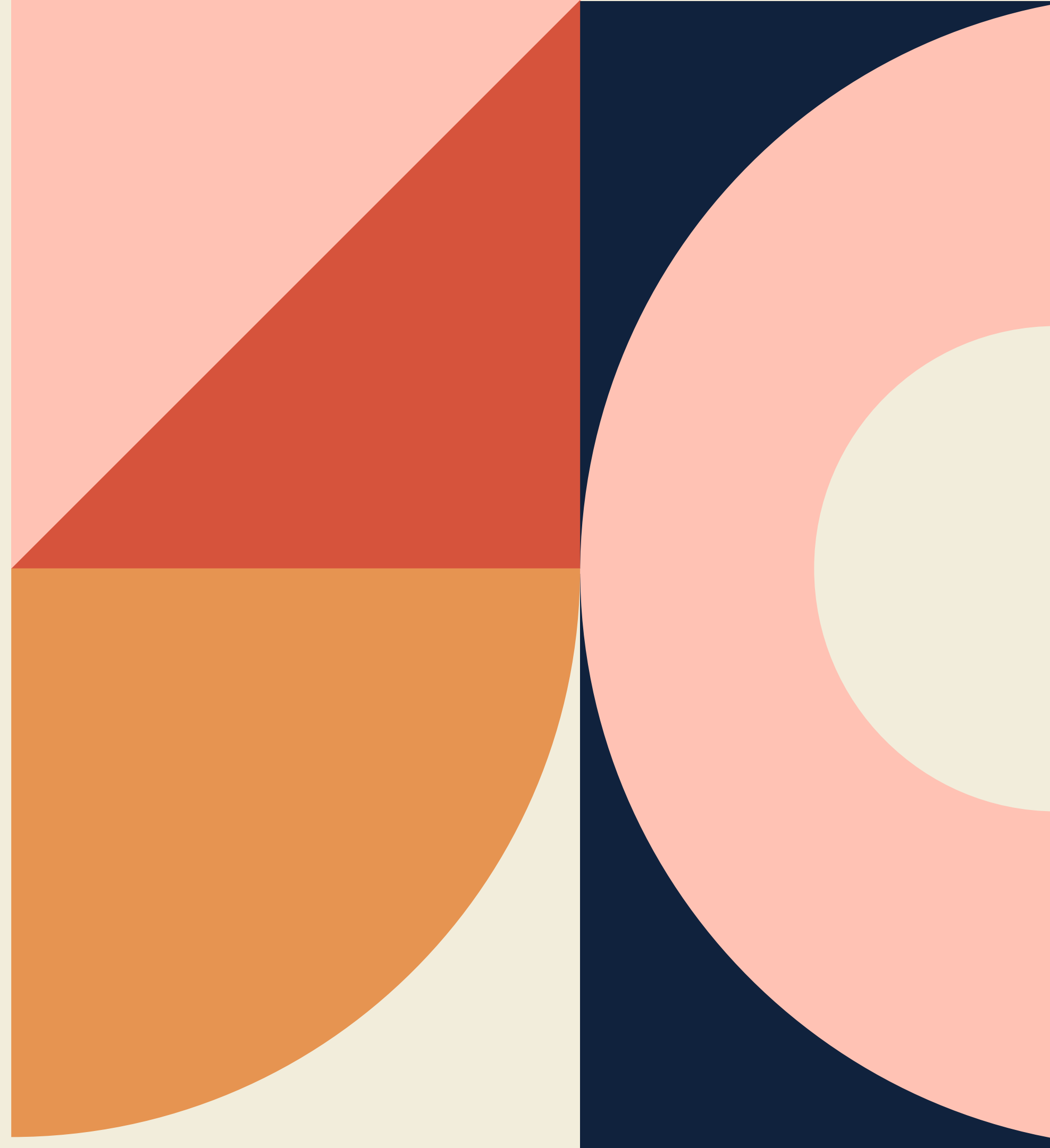




MACHINE LEARNING

By Mauboussin Tanguy
Lhuillier Benjamin



01 / INTRODUCTION

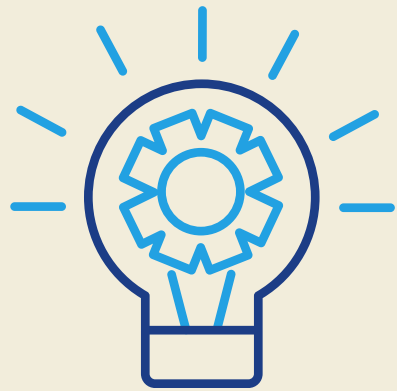
NOTRE PROBLÉMATIQUE

QUELS SONT LES VARIABLES LES PLUS PRÉDICTIVES EN TERMES
DE PERFORMANCE FOOTBALLISTIQUE?



01 / INTRODUCTION

CHOIX DU JEU DE DONNÉES



- CRÉATION DE NOTRE PROPRE BASE DE DONNÉES
- MOINS DE DONNÉES MAIS PLUS DE FIABILITÉ
- AUCUNE DONNÉES MANQUANTES



CHOIX DES VARIABLES

	Variable Y	Variable X					
CLUB	Performance	Budget transfert	Joueurs formés club	Valeur club	Age équipe	Supporters	Prix équipe
Real Madrid	380	956	0,076	1080	27,6	0,731	884,65
Bayern	318	722	0,18	915	26,9	0,953	753,242
Barcelone	314	813	0,19	1060	27,3	0,713	966,036
Atlético Madrid	236	409	0,13	863	26,4	0,707	749,86
Chelsea	225	765	0,307	845	26,4	0,963	863,146
Juventus	215	630	0	762	28,8	0,878	665,198
Manchester U	211	1350	0,307	800	24,8	0,955	720,21
PSG	205	1272	0,137	1010	26,7	0,937	823,21
Porto	190	262	0,16	271	27	0,711	264,616
Manchester City	188	1297	0,083	1290	27,1	0,92	963,14
Arsenal	187	540	0,296	694	26,4	0,945	632,36
Benfica	184	180	0,32	333	25,2	0,77	286,696
Dortmund	178	374	0,153	616	25,4	0,93	572,05
Séville	176	286	0	293	27,1	0,732	299,09
Liverpool	170	542	0,041	1190	25,5	0,906	887,63
Shakhtar	164	225	0,25	152	26,9	0,091	142,426
Tottenham	156	604	0,125	889	25,5	0,75	753,09
Ajax	149	193	0,31	429	23,4	0,928	336,78
Naples	148	293	0,074	688	26,6	0,617	525,168
Zénit	146	483	0,043	204	24,3	0,496	188,478
Lyon	145	378	0,172	419	25,3	0,75	351,82
Schalke	144	334	0,16	242	25,5	0,935	208,326
Valence	137	304	0,2	529	24,5	0,736	370,2
Bâle	136	97	0,321	52	24,4	0,625	65,766
Leverkusen	134	263	0,125	434	26,5	0,895	366,426



POURQUOI ?

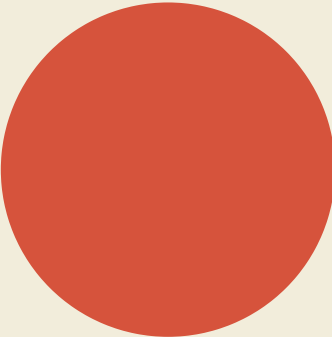
VARIABLE REPRÉSENTATIVE ET FIABLE

INDICATEUR CLÉS

A QUOI ÇA SERT ?

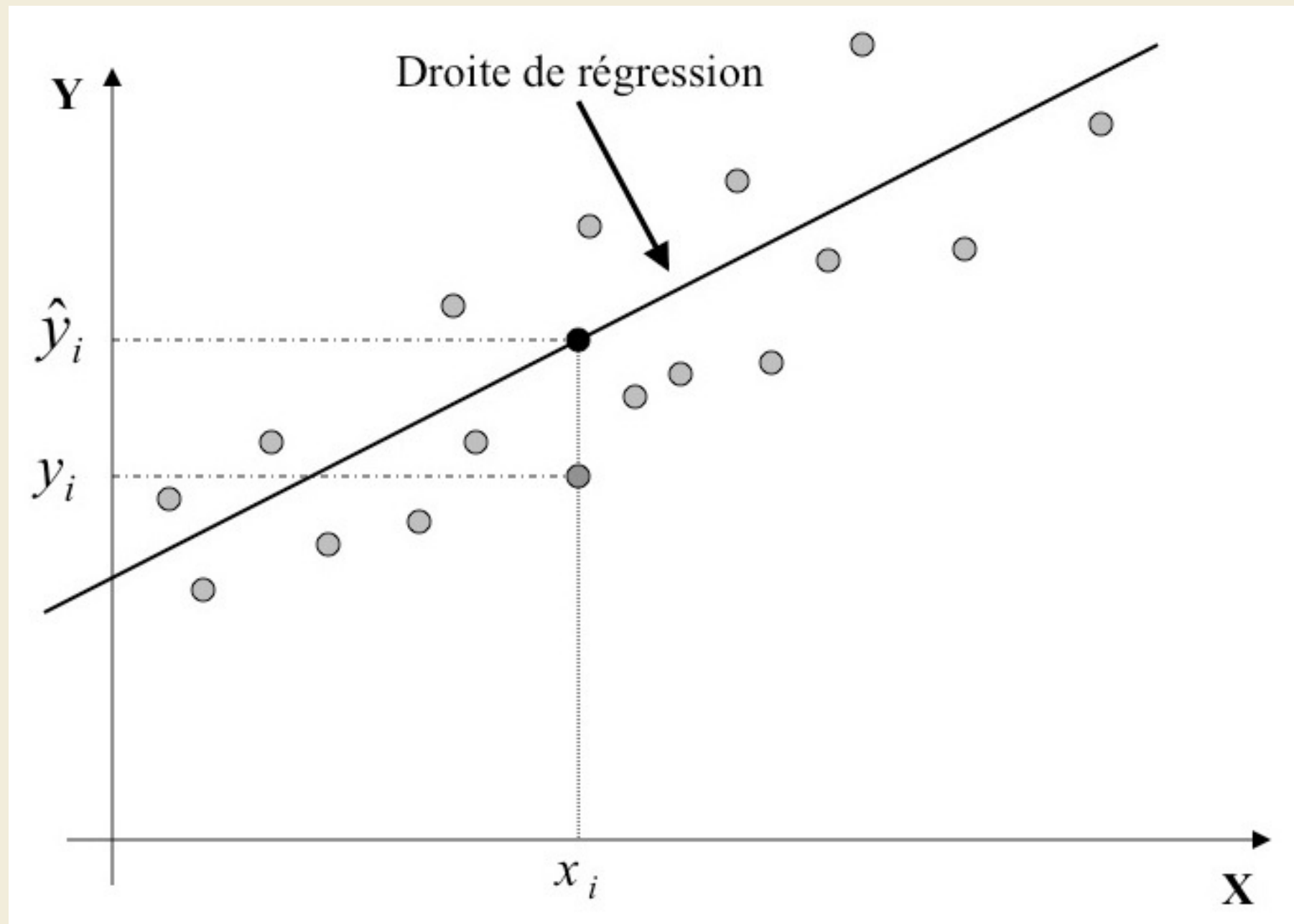
CONSTATER L'IMPACT SUR LA PERFORMANCE DES ÉQUIPES

ALIMENTER L'ALGORITHME DE MACHINE LEARNING



01 - INTRODUCTION

ALGORITHME UTILISÉ



POURQUOI ?

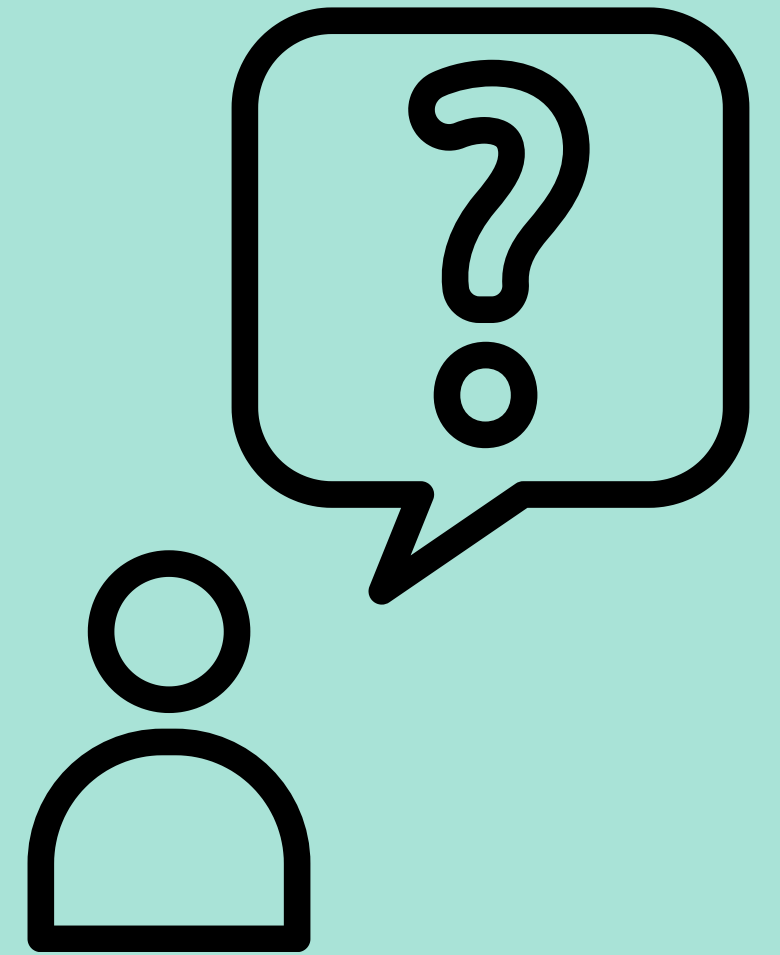
COMPRENDRE LE LIEN ENTRE LES VARIABLES

OBSERVER DES TENDANCES

TIRER DES CONCLUSIONS



02- PROBLÈMES RENCONTRÉS



02- NORMALISATION AVEC PYTHON

Définition de la variable à prédire:

```
GAnum = GA.iloc[:,1:8] ## On définit que X prend l'ensemble des variables de la colonne 1 à 8.
```

```
GAnum.head() ## On affiche le tableau avec les valeurs sélectionnées.
```

	Perf	Budget transfert	Joueurs formés club	Valeur club	Age équipe	Supporters	Prix équipe
0	380	956	0.076	1080	27.605000	0.731083	884.650
1	318	722	0.180	915	26.863333	0.952655	753.242
2	314	813	0.190	1060	27.258333	0.713076	966.036
3	236	409	0.130	863	26.426667	0.706702	749.860
4	225	765	0.307	845	26.438333	0.963210	863.146

```
#Ici, je standardise mon jeu de données.  
from sklearn.preprocessing import normalize
```

```
GAnorm = normalize(GAnum)
```

```
X = GAnorm.iloc[:,1:8]  
X.head()
```

	Budget transfert	Joueurs formés club	Valeur club	Age équipe	Supporters	Prix équipe
0	0.551203	0.000044	0.622698	0.015916	0.000422	0.510064
1	0.507028	0.000126	0.642564	0.018865	0.000669	0.528968
2	0.484381	0.000113	0.631542	0.016240	0.000425	0.575559
3	0.330578	0.000105	0.697527	0.021360	0.000571	0.606080
4	0.528455	0.000212	0.583718	0.018263	0.000665	0.596253

```
Y = GAnorm.Perf  
Y.head()
```

```
0    0.219097  
1    0.223317  
2    0.187079  
3    0.190749  
4    0.155428
```

Name: Perf, dtype: float64

Signification des coefficients
moins pertinente !!

02- NORMALISATION MANUELLE

```
X = GA.iloc[:,2:8] ## On définit que X prend l'ensemble des variables de la colonne 2 à 8.  
X.head() ## On affiche le tableau avec les valeurs sélectionnés.
```

	Budget transfert	Joueurs formés club	Valeur club	Age équipe	Supporters	Prix équipe
0	0.708148	0.076	0.837209	0.958507	0.731083	0.915749
1	0.534815	0.180	0.709302	0.932755	0.952655	0.779721
2	0.602222	0.190	0.821705	0.946470	0.713076	0.999996
3	0.302963	0.130	0.668992	0.917593	0.706702	0.776220
4	0.566667	0.307	0.655039	0.917998	0.963210	0.893489

```
Y=GA.Perf ## On définit que Y prend la colonne Perf du jeu de données GA.  
Y.head() ## On affiche le tableau avec les valeurs sélectionnés.
```

```
0    1.000000  
1    0.836842  
2    0.826316  
3    0.621053
```

CLUB	Perf	Budget transfert	Joueurs formés club	Valeur club	Age équipe	Supporters	Prix équipe
Real Madrid	1	0,708148	0,076	0,837209	0,959	0,731	0,91575
Bayern	0,836842	0,534815	0,180	0,709302	0,933	0,953	0,77972
Barcelone	0,826316	0,602222	0,190	0,821705	0,946	0,713	1
Atlético Madrid	0,621053	0,302963	0,130	0,668992	0,918	0,707	0,77622
Chelsea	0,592105	0,566667	0,307	0,655039	0,918	0,963	0,89349
Juventus	0,565789	0,466667	0,000	0,590698	1,000	0,878	0,68858
Manchester U	0,555263	1	0,307	0,620155	0,862	0,955	0,74553
PSG	0,539474	0,942222	0,137	0,782946	0,929	0,937	0,85215
Porto	0,5	0,194074	0,160	0,210078	0,939	0,711	0,27392
Manchester City	0,494737	0,960741	0,083	1	0,942	0,920	0,997
Arsenal	0,492105	0,4	0,296	0,537984	0,917	0,945	0,65459
Benfica	0,484211	0,133333	0,320	0,25814	0,875	0,770	0,29677
Dortmund	0,468421	0,277037	0,153	0,477519	0,881	0,930	0,59216

LES VALEURS GARDENT UNE COHÉRENCE PUISQUE NOUS TRAITONS LA NORMALISATION AVEC UNE
VALEUR DE RÉFÉRENCE PROPRE À CHAQUE COLONNE

02- NORMALISATION MANUELLE OU AVEC PYTHON ?

```
▶ print (lm1.coef_)#Interpreter la signification des coefficients
```

```
[-0.69352793  47.09052852 -0.73262488  0.7693153  -8.25297636 -0.89584751]
```

BUDGET TRANSFERT

JOUEURS FORMÉS
CLUB

VALEUR CLUB

AGE ÉQUIPE

SUPPORTERS

PRIX ÉQUIPE

```
▶ print (lm1.coef_)#Interpreter la signification des coefficients|
```

```
[ 0.06551946  0.54821782 -0.34436191 -0.3266426  0.15466845  0.88526347]
```

RÉSULTATS PLUS PROBANTS AVEC LA MÉTHODE **MANUELLE** CAR ON GARDE UNE CERTAINE COHÉRENCE ENTRE LES DONNÉES
LÀ OÙ PYTHON NORMALISE LA RANGÉE.

02- AUGMENTATION DE LA BASE DE DONNÉES

Club	Perf	Budget transfert	Joueurs formés club	Valeur club	Age équipe	Supporters	Prix équipe
Real Madrid	1	0,708148148	0,076	0,837209302	0,958507	0,731	0,915752622
Bayern	0,836842105	0,534814815	0,180	0,709302326	0,932755	0,953	0,779724565
Barcelone	0,826315789	0,602222222	0,190	0,821705426	0,946470	0,713	1
Atlético Madrid	0,621052632	0,302962963	0,130	0,668992248	0,917593	0,707	0,77622366
Chelsea	0,592105263	0,566666667	0,307	0,65503876	0,917998	0,963	0,893492582
Juventus	0,565789474	0,466666667	0,000	0,590697674	0,999595	0,878	0,688585104
Manchester U	0,555263158	1	0,307	0,620155039	0,862211	0,955	0,745531222
PSG	0,539473684	0,942222222	0,137	0,782945736	0,928530	0,937	0,852152508
Porto	0,5	0,194074074	0,160	0,210077519	0,939120	0,711	0,273919398
Manchester City	0,494736842	0,960740741	0,083	1	0,941609	0,920	0,997002182
Arsenal	0,492105263	0,4	0,296	0,537984496	0,917072	0,945	0,654502582
Benfica	0,484210526	0,133333333	0,320	0,258139535	0,874884		
Dortmund	0,468421053	0,277037037	0,153	0,47751938	0,880729		
Séville	0,463157895	0,211851852	0,000	0,227131783	0,941609		
Liverpool	0,447368421	0,401481481	0,041	0,92248062	0,884896		
Shakhtar	0,431578947	0,166666667	0,250	0,117829457	0,933738		
Tottenham	0,410526316	0,447407407	0,125	0,689147287	0,885938		
Ajax	0,392105263	0,142962963	0,310	0,33255814	0,811169		
Naples	0,389473684	0,217037037	0,074	0,533333333	0,924884		
Zénit	0,384210526	0,357777778	0,043	0,158139535	0,842072		
Lyon	0,381578947	0,28	0,172	0,324806202	0,880150		
Schalke	0,378947368	0,247407407	0,160	0,187596899	0,884838		
Valence	0,360526316	0,225185185	0,200	0,410077519	0,849653		
Bâle	0,357894737	0,071851852	0,321	0,040310078	0,847569		
Leverkusen	0,352631579	0,194814815	0,125	0,336434109	0,920197		

CLUB	Perf	Budget transfert	Joueurs formés club	Valeur club	Age équipe	Supporters	Prix équipe
Real Madrid	1	0,708148	0,076	0,837209	0,959	0,731	0,91575
Bayern	0,836842	0,534815	0,180	0,709302	0,933	0,953	0,77972
Barcelone	0,826316	0,602222	0,190	0,821705	0,946	0,713	1
Atlético Madrid	0,621053	0,302963	0,130	0,668992	0,918	0,707	0,77622
Chelsea	0,592105	0,566667	0,307	0,655039	0,918	0,963	0,89349
Juventus	0,565789	0,466667	0,000	0,590698	1,000	0,878	0,68858
Manchester U	0,555263	1	0,307	0,620155	0,862	0,955	0,74553
PSG	0,539474	0,942222	0,137	0,782946	0,929	0,937	0,85215
Porto	0,5	0,194074	0,160	0,210078	0,939	0,711	0,27392
Manchester City	0,494737	0,960741	0,083	1	0,942	0,920	0,997
Arsenal	0,492105	0,4	0,296	0,537984	0,917	0,945	0,65459
Benfica	0,484211	0,133333	0,320	0,25814	0,875	0,770	0,29677
Dortmund	0,468421	0,277037	0,153	0,477519	0,881	0,930	0,59216
Séville	0,463158	0,211852	0,000	0,227132	0,942	0,732	0,3096
Liverpool	0,447368	0,401481	0,041	0,922481	0,885	0,906	0,91883
Shakhtar	0,431579	0,166667	0,250	0,117829	0,934	0,091	0,14743
Tottenham	0,410526	0,447407	0,125	0,689147	0,886	0,750	0,77956
Ajax	0,392105	0,142963	0,310	0,332558	0,811	0,928	0,34862
Naples	0,389474	0,217037	0,074	0,533333	0,925	0,617	0,54363
Zénit	0,384211	0,357778	0,043	0,15814	0,842	0,496	0,1951
Lyon	0,381579	0,28	0,172	0,324806	0,880	0,750	0,36419
Schalke	0,378947	0,247407	0,160	0,187597	0,885	0,935	0,21565
Valence	0,360526	0,225185	0,200	0,410078	0,850	0,736	0,38321
Bâle	0,357895	0,071852	0,321	0,04031	0,848	0,625	0,06808
Leverkusen	0,352632	0,194815	0,125	0,336434	0,920	0,895	0,37931
Milan	0,342105	0,166667	0,13	0,393798	0,8934028	0,736	0,44512
Roma	0,286842	0,140741	0,215	0,30155	0,928125	0,698	0,40391
PSV	0,242105	0,066667	0,31	0,143411	0,8541667	0,546	0,19541
Salzburg	0,239474	0,059259	0,34	0,113178	0,8020833	0,67	0,15743
Marseille	0,221053	0,103704	0,09	0,173643	0,9652778	0,459	0,21855
Braga	0,210526	0,056296	0,267	0,089147	0,8611111	0,47	0,10269
Lazio	0,205263	0,1	0,217	0,257364	0,9166667	0,678	0,32374
Celtic	0,192105	0,062963	0,469	0,071318	0,8645833	0,532	0,09786
Astoria Bilbao	0,192105	0,07037	0,284	0,162016	0,9305556	0,719	0,22508
Galatasaray	0,181579	0,088889	0,218	0,075194	0,9583333	0,754	0,10705
Monaco	0,163158	0,159259	0,123	0,264341	0,8194444	0,44	0,39292
Florentina	0,144737	0,085185	0,14	0,149612	0,9618056	0,591	0,26185

2) Calcul du score du modele:

```
score1=lm1.score(X_test,Y_test)
print (score1)
## Le R2 est de 34%
```

0.1586221392661451

AVOIR UN SCORE DE MODÈLE PLUS ÉLEVÉ

2) Calcul du score du modele:

```
score1=lm1.score(X_test,Y_test)
print (score1)
## Le R2 est de 18%
```

0.1821980846428548

AVOIR UN ÉCHANTILLON DE
DONNÉES PLUS CONSÉQUENT
POUR DES INTERPRÉTATIONS
FACILITÉES !

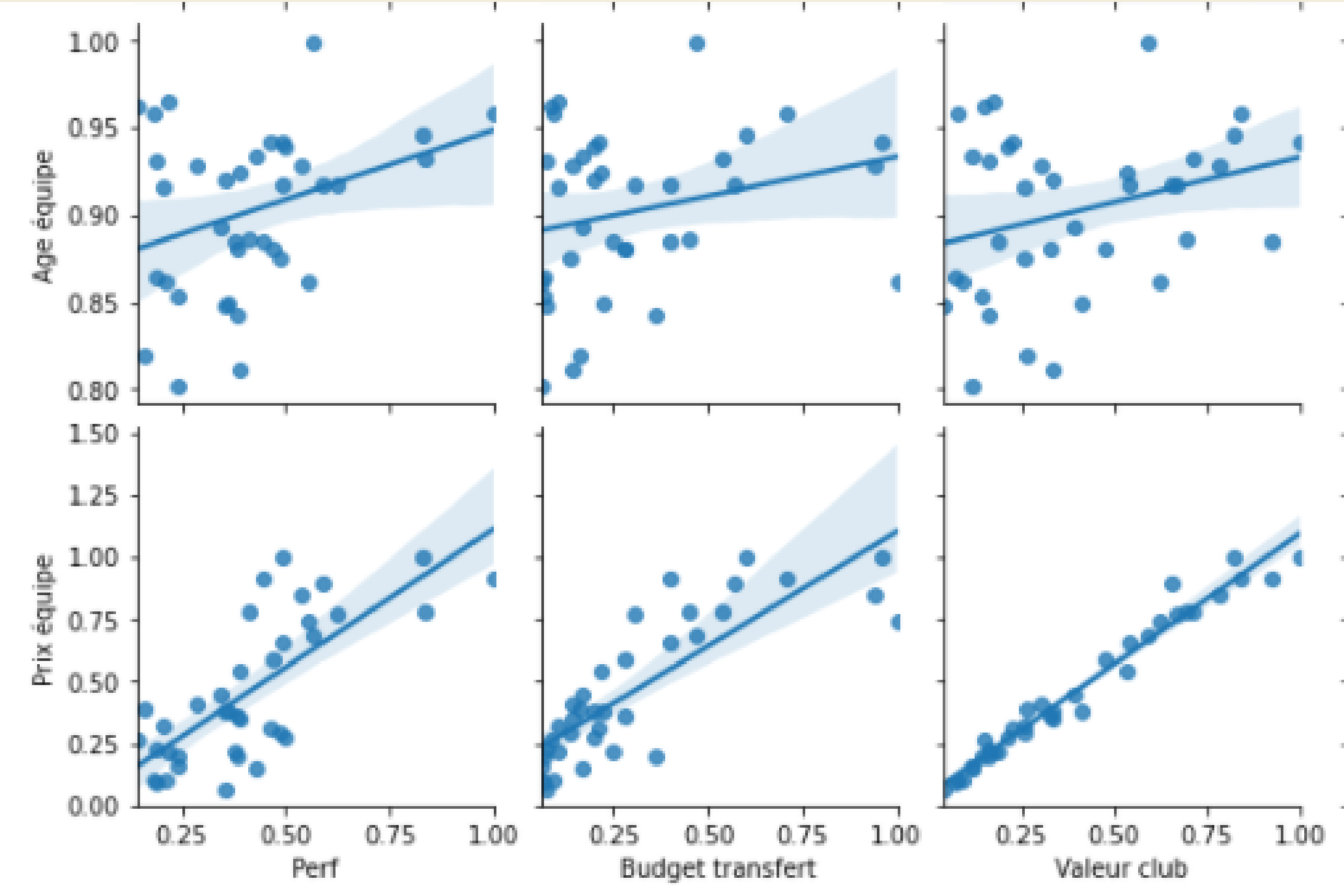
03- SUPPRESSION DE VARIABLES

POURQUOI ?

- VARIABLES NON PERTINENTES AU VU DE LEURS NON-LINÉARITÉS.
- AVOIR UNE INTERPRÉTATION PLUS AISÉE DES VARIABLES LES UNES COMPARÉES AUX AUTRES.

GA.head(10)

	CLUB	Perf	Budget transfert	Valeur club	Age équipe	Prix équipe
0	Real Madrid	1.000000	0.708148	0.837209	0.958507	0.915749
1	Bayern	0.836842	0.534815	0.709302	0.932755	0.779721
2	Barcelone	0.826316	0.602222	0.821705	0.946470	0.999996
3	Atlético Madrid	0.621053	0.302963	0.668992	0.917593	0.776220
4	Chelsea	0.592105	0.566667	0.655039	0.917998	0.893489
5	Juventus	0.565789	0.466667	0.590698	0.999595	0.688582
6	Manchester U	0.555263	1.000000	0.620155	0.862211	0.745528
7	PSG	0.539474	0.942222	0.782946	0.928530	0.852149
8	Porto	0.500000	0.194074	0.210078	0.939120	0.273918
9	Manchester City	0.494737	0.960741	1.000000	0.941609	0.996998



03- ALGORITHME RETENU - PARAMÈTRES

MÉTHODE DE NORMALISATION MANUELLE
SIGNIFICATION DES COEFFICIENTS PLUS PERTINENTE

PASSAGE DE LA BDD DE 25 ÉQUIPES À 37
ECHANTILLON PLUS COMPLET

SUPPRESSION DE 2 VARIABLES PRÉDICTIVES
VARIABLES NON PERTINENTES LORS DE L'EXPLORATION DES DONNÉES

```
Remarquez aussi split, cela signifie que l'on va diviser notre jeu de donnée en deux: une partie pour la premiere phase qui correspond a la phase d'entrainement de la machine et l'autre partie pour tester si ce que la machine aura appris est fiable.
```

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.5, random_state=42)
```

GA.head(10)

	CLUB	Perf	Budget transfert	Valeur club	Age équipe	Prix équipe
0	Real Madrid	1.000000	0.708148	0.837209	0.958507	0.915749
1	Bayern	0.836842	0.534815	0.709302	0.932755	0.779721
2	Barcelone	0.826316	0.602222	0.821705	0.946470	0.999996
3	Atlético Madrid	0.621053	0.302963	0.668992	0.917593	0.776220
4	Chelsea	0.592105	0.566667	0.655039	0.917998	0.893489
5	Juventus	0.565789	0.466667	0.590698	0.999595	0.688582
6	Manchester U	0.555263	1.000000	0.620155	0.862211	0.745528
7	PSG	0.539474	0.942222	0.782946	0.928530	0.852149
8	Porto	0.500000	0.194074	0.210078	0.939120	0.273918
9	Manchester City	0.494737	0.960741	1.000000	0.941609	0.996998

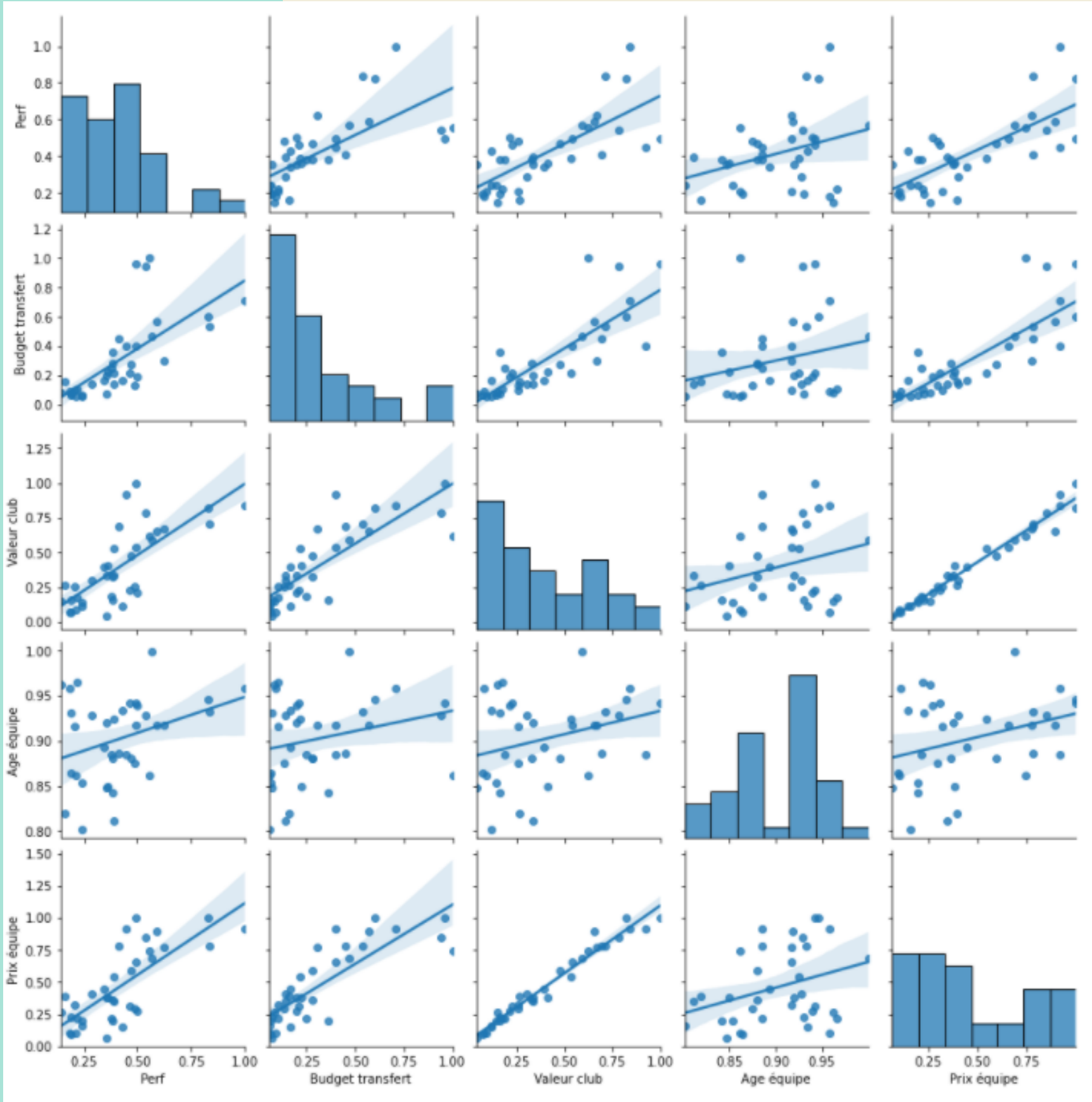
#Quelle est la taille du jeu de données?
#Ici, nous avons 200 rangées et 48 colonnes.
GA.shape

(37, 6)

#Ici, je vérifie que je n'ai pas de données manquantes.
GA.isnull().sum()

CLUB	0
Perf	0
Budget transfert	0
Valeur club	0
Age équipe	0
Prix équipe	0
dtype:	int64

03- ALGORITHME RETENU - INTERPRÉTATION



```
print (lm1.coef_)#Interpreter la signification des coefficients|
[ 0.16825282 -0.85592884 -0.88500894  1.27349209]
```

```
print (lm1.intercept_)
# sans aucune influence des facteurs budget transfert, valeur club etc...
0.9271417710127352
```

GA.corr()

	Perf	Budget transfert	Valeur club	Age équipe	Prix équipe
Perf	1.000000	0.689876	0.729949	0.328402	0.744565
Budget transfert	0.689876	1.000000	0.828066	0.249812	0.829243
Valeur club	0.729949	0.828066	1.000000	0.298025	0.984828
Age équipe	0.328402	0.249812	0.298025	1.000000	0.324580
Prix équipe	0.744565	0.829243	0.984828	0.324580	1.000000

03- ALGORITHME RETENU - RÉSULTATS

1) Calcul de l'erreur du modele:

```
print (np.sqrt(mean_squared_error(Y_test, pred1)))  
0.16802079333084055
```

2) Calcul du score du modele:

```
score1=lm1.score(X_test,Y_test)  
print (score1)  
## Le R2 est de 34%  
0.1821980846428548
```

3) Calcul de l'erreur du modele de base:

```
#baseline prediction  
baseline=np.mean(Y_test)  
baseline
```

```
0.4221606648199446
```

```
baseline_pred=np.empty(len(Y_test))# cree un nouveau tableau  
baseline_pred.fill(baseline) #rempli avec predictions  
print(np.sqrt(mean_squared_error(Y_test,baseline_pred)))  
## Correspond à la marge d'erreur du modèle
```

```
0.18579711634613183
```

04 / CONCLUSION

QUELS SONT LES VARIABLES LES PLUS PRÉDICTIVES EN TERMES DE PERFORMANCE FOOTBALLISTIQUE?

LIEN EXISTANT ENTRE LA PERFORMANCE ET CERTAINES VARIABLES

- BUDGET DE TRANSFERT
- VALEUR DU CLUB
- PRIX DE L'ÉQUIPE

**INTERPRÉTATION DES GRAPHIQUES
LES DIFFÉRENTS COEFFICIENTS**

BESOIN D'UN ÉCHANTILLON DE DONNÉES PLUS CONSÉQUENT.

