

ROUVEURE Kylian

VIENNE Soleymane

HADBI Souleimane

R 4.02

Rapport méthodologique

Sous le direction de **M.ROLLAND**

BUT Science des Données

2ème année - IUT Lumière

Année universitaire 2022/2023

SOMMAIRE

1. Introduction	3
2. Base de données	4
3. Corrélation	5
4. ACP	6
5. Sans la variable “Classement”	9
6. Conclusion & Ressenti	10
7. Annexes	11

1. Introduction

Afin de réaliser une étude statistique, nous avons choisi la Ligue 1 Uber Eats. Notre premier objectif est de comprendre quels critères de performances influent le plus sur la réussite d'une équipe dans le championnat.

Tout d'abord, la Ligue 1 est la plus haute ligue de football professionnels en France. Ce championnat est l'une des compétitives et l'une des plus populaires au monde. Cette compétitivité est due en partie à la qualité des équipes et des joueurs qui y jouent.

Dans ce rapport, nous voulons trouver l'équipe la plus complète du championnat et expliquer notre démarche.

Pour démontrer ceci, nous allons étudier la performance de toutes les équipes grâce aux données de possession, % tirs réussis, buts marqués, buts encaissés, % passes réussies, % dribbles réussis, % duels gagnés et le nombre de cartons jaunes.

Nous allons prendre comme problématique :

A partir des données de performances des équipes, y-a t'il un lien entre leurs performances et leurs résultats ?

2. Base de données

a. Les champs

La base de données provient de *Foot Mercato* sur le début de saison 2023. Elle contient comme variable :

- Equipe
- Classement
- Nb victoire
- % de possession moyen
- % de tir cadré
- % de passe réussi
- % de dribble réussi
- But marqué
- But encaissé
- Passe Décisive
- % de duels gagnés
- Carton Jaune

Voici un extrait de la base de données : **(Annexe 1)**

b. Traitement

```
# Change les virgules en points
data$X..TIR.CADRE <- gsub(",", ".", data$X..TIR.CADRE)
data$X..PASSE.REUSSI <- gsub(",", ".", data$X..PASSE.REUSSI)
data$X..DRIBBLE.REUSSI <- gsub(",", ".", data$X..DRIBBLE.REUSSI)
data$X..DUEL.GAGNE <- gsub(",", ".", data$X..DUEL.GAGNE)

# Convertir la variable de texte en pourcentages en nombres décimaux
data$X..TIR.CADRE <- as.numeric(gsub("%", "", data$X..TIR.CADRE)) / 100
data$X..PASSE.REUSSI <- as.numeric(gsub("%", "", data$X..PASSE.REUSSI)) / 100
data$X..DRIBBLE.REUSSI <- as.numeric(gsub("%", "", data$X..DRIBBLE.REUSSI)) / 100
data$X..DUEL.GAGNE <- as.numeric(gsub("%", "", data$X..DUEL.GAGNE)) / 100
```

Avant de faire les analyses, nous devons faire une dernière étape. Nous avons des variables en pourcentage de ce type : 12,13%. Rstudio traitera ces variables comme du texte et non des chiffres. Du coup, ce code permet de transformer les “,” en “.” et ensuite d’enlever les pourcentages et avoir des valeurs de ce type : 0.12. Après ces changements, nous pourrons faire nos analyses.

3. Corrélation

(Annexe 2)

Dans un premier temps, on fait un graphique de corrélation pour avoir un aperçu général des variables.

On remarque du coup qu'il y a des variables qui sont corrélées entre elles. Plus les variables sont corrélées, plus elles sont bleues, moins elles le sont, plus elles sont rouges. Du coup, on peut retenir que la variables "Nb Victoire" est corrélé avec "Possession moyenne", "Tir cadré" ou encore "But marqué". Elle ne l'ai pas du tout avec "but encaissé".

Ce graphique nous permet d'avoir un premier aperçu du jeu de données et d'avoir une première analyse. Nous allons donc rentrer dans les détails avec une analyse en Composante Principale.

4. ACP

Pour réaliser cette étude, nous avons réalisé un ACP afin de pouvoir observer les facteurs de performances qui influencent la réussite dans un championnat.

a. Dimension

Tableau des Dimensions : **(Annexe 3)**

Code R
resultats.ACP\$eig

Ci-dessus, nous trouvons le tableau des valeurs propres qui répertorie les contributions des dimensions de l'ACP. Ce tableau nous a permis de choisir les axes les plus pertinents à analyser.

Nous constatons que les 3 premiers axes sont les plus importants avec une représentation de 83% de l'information. Comme il est proche de 1, nous avons fait le choix de garder la dimension 3.

Tableau des Variables : **(Annexe 4)**

Code R
resultats.ACP\$var

À partir de ce tableau, nous pouvons voir la contribution des variables dans les différentes dimensions de l'ACP.

Les variables principales sont :

Dim 1 => Classement (14,3%), Nb victoire (14,0%), Passes Décisives (13,6%) et But marqué (13,1%)

Dim 2 => % Duels gagnés (45,9%), % Dribbles réussis (28%) & Cartons jaunes (16,4%)

Dim 3 => Cartons jaunes (57,7%), % Dribbles réussis (20,7%), % Tirs cadrés (7,6%) & % Possession moyenne (7,2%)

Dim 4 => Buts encaissés (41,5%), % Tirs cadrés (21,2%) et % Passes réussis (11%)

Tableau des Individus : **(Annexe 5)**

Code R
resultats.ACP\$ind

À partir de ce tableau, nous pouvons voir la contribution des équipes dans les différentes dimensions de l'ACP.

Mais pour les analyses détaillées en fonction des dimensions, on s'y penchera dans les prochaines parties cependant nous pouvons tout de même dire que dans l'ensemble le classement a logiquement un lien avec les buts encaissés et marqué mais pas forcément avec le % de passes réussis, dribbles réussis... avec des données de performances.

b. Dim 1 & 2

(Annexe 6 & 7)

Lorsque l'on fait un focus sur les variables des dimensions 1 et 2, on voit deux tendances :

- horizontale : à droite où il y a les statistiques principales des équipes (But marqué, tir cadré ...); à gauche où on a but encaissé et classement (Plus notre classement est élevé, plus on est dernier au classement).
- verticale : vers le haut on a duel gagné et dribble réussi; vers le bas la variable carton jaune.

A partir de ce graphique, on peut conclure deux choses : plus on est haut dans le classement, plus on a de but encaissé et moins on a de statistiques élevées (ce qui paraît logique). On voit aussi que les variables 'duel gagné' et 'dribble réussi' n'ont aucune voire peu aucune incidence sur la victoire d'une équipe.

Ce graphique des individus nous montre bien ce qu'on a dit précédemment. Par exemple, Reims 9eme a pris plus de but et plus de cartons jaunes que Lyon, mais sont devant eux au classement (Lyon 10eme). Ce graphique nous montre aussi que les individus les plus à gauche sont les plus haut au classement, et ceux à droite sont les plus bas.

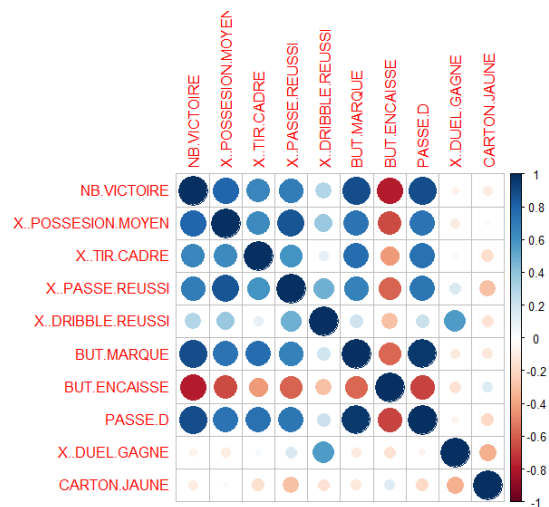
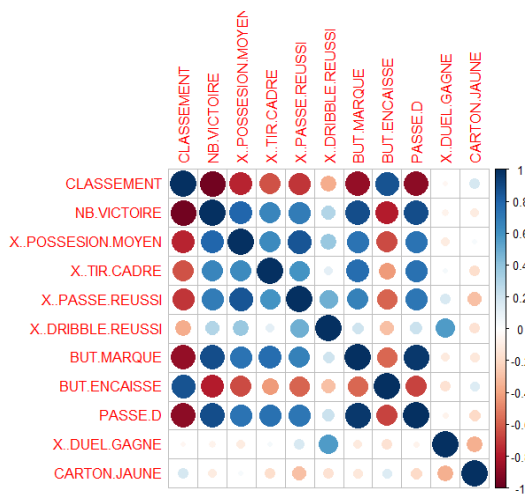
c. Dim 1 & 3

(Annexe 8 & 9)

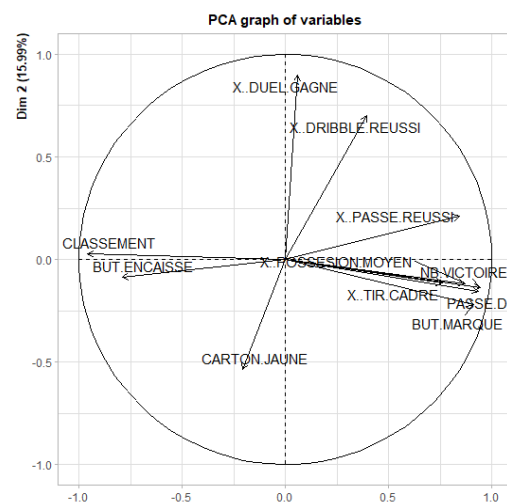
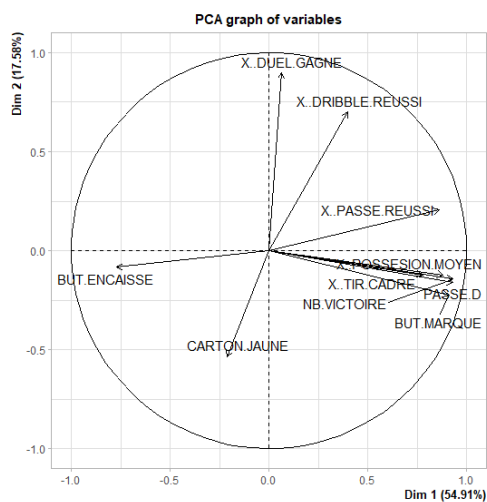
Lorsqu'on s'intéresse aux dimensions 1 et 3, on remarque que celui-ci n'est plus le même. La variable carton jaune est beaucoup plus représenté ici que dans la dimension 2. La variable dribble réussi est moins représentée et la variable duel gagné est presque inexistante sur cette dimension. On a toujours notre représentation horizontale comme on a toujours la dimension 1. Du côté du graphique, on voit immédiatement que Nice n'est plus du tout représenté, est Lille est surreprésenté par la variable Carton Jaune. A part quelques changements comme cela, il n'y a pas trop de changements.

On peut donc se dire que la variable classement influe sur nos résultats, et peut-être que sans cette variable on aurait quelques choses de plus représentative. Nous allons donc faire les mêmes analyses sur les dimensions 1 et 2 en enlevant la variable Classement.

5. Sans Classement



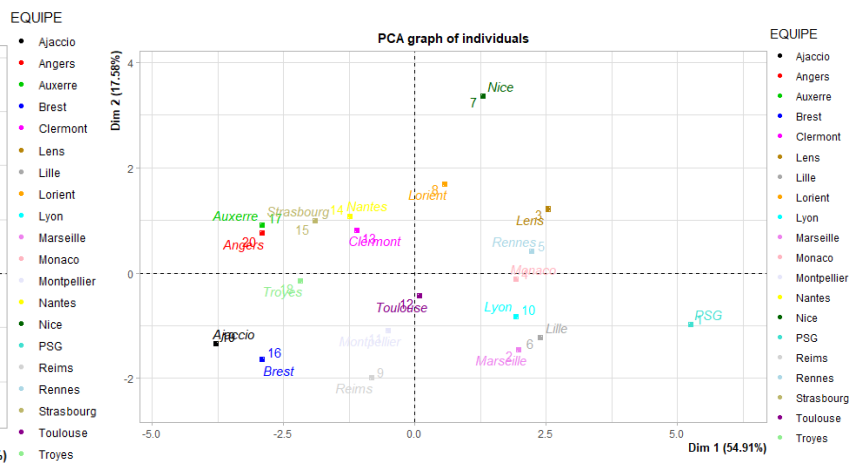
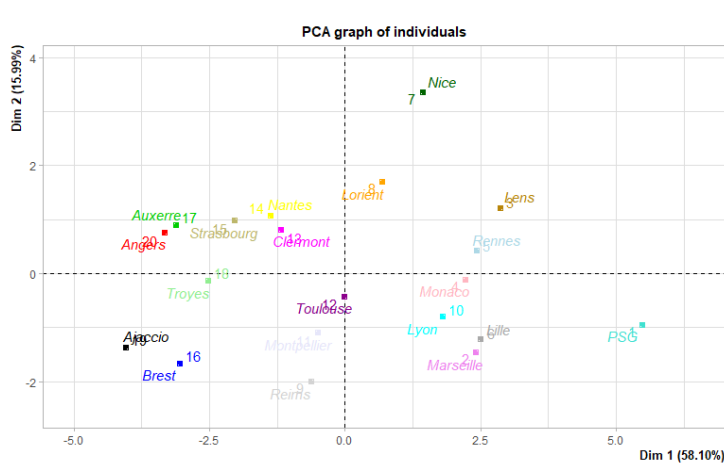
Lorsqu'on s'intéresse à la corrélation entre les variables, on ne notifie aucun à l'œil nu. Peut-être que l'ACP nous donnera plus d'informations.



Sur ce graphique des variables, à gauche sans la variable Classement et à droite avec la variable Classement, on ne voit aussi aucun changement lorsque l'on enlève la variable Classement. On se doute que le graphique des individus nous donnera la même chose.

Avec Variable Classement

Sans Variable Classement



On voit aussi qu'il n'y a pas de changement direct. On remarque juste que toutes les variables sont tirées vers le bas, mais rien de plus.

6. Conclusion & ressenti

Pour répondre à notre question, il y a donc des variables qui influent sur la victoire d'une équipe comme la possession, tir cadré mais il y a aussi des variables qui n'influencent pas ou influencent très peu la victoire, comme les duel gagnés, dribbles réussis.

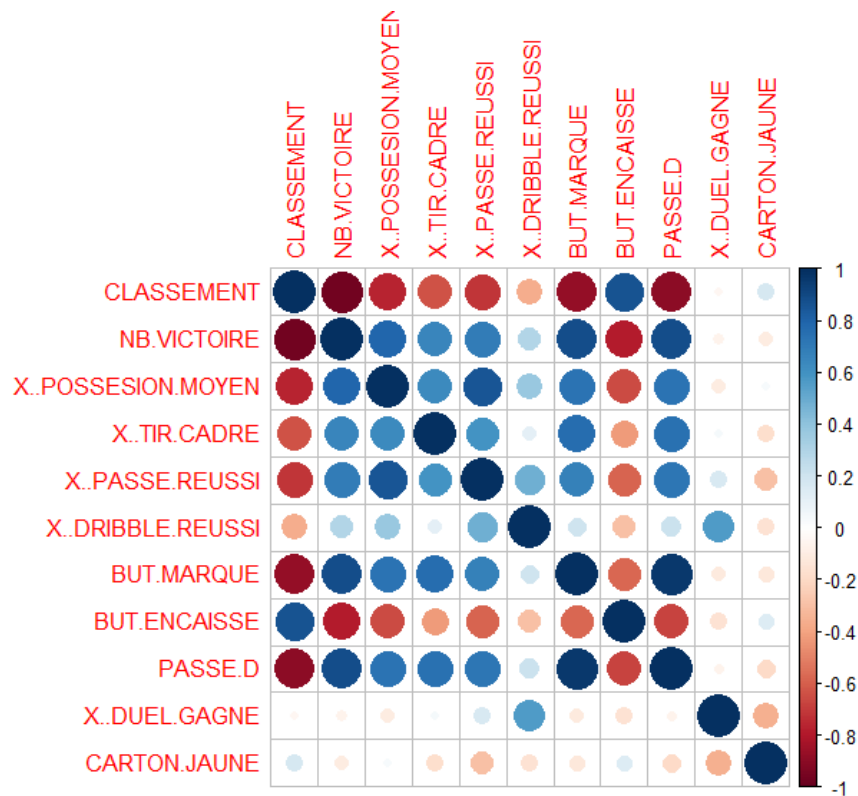
Dans ce projet, nos réelles difficultés ont été de trouver des données pertinentes et à réaliser le traitement afin de pouvoir afficher les % de passes, duels, dribbles qui ne sont pas forcément relevés explicitement dans les jeux de données. Nous pensons avoir pris les variables les plus pertinentes afin de juger et comprendre les facteurs de performances qui résulte sur le classement.

7. Annexes

Annexe 1 : Tableau des données

EQUIPE	CLASSEMENT	NB VICTOIRE	% POSSESSION MOYEN	% TIR CADRE	% PASSE REUSSI	% DRIBBLE REUSSI	BUT MARQUE	BUT ENCAISSE	PASSE D	% DUEL GAGNE
PSG	1	21	60	60,00%	90,09%	49,07%	68	28	49	49,85%
Lille	6	14	59	54,43%	85,93%	52,87%	51	36	31	48,61%
Lyon	10	11	58	58,27%	85,17%	47,92%	43	32	30	49,23%
Lens	3	16	57	41,67%	86,67%	54,42%	47	21	34	50,70%
Marseille	2	18	55	49,48%	83,14%	48,88%	53	28	30	48,67%
Rennes	5	15	55	53,18%	84,47%	51,75%	47	29	31	49,97%
Nice	7	11	51	48,09%	84,95%	54,24%	37	25	27	53,50%
Toulouse	12	10	51	51,34%	81,36%	49,14%	43	48	27	49,27%
Monaco	4	16	49	53,52%	81,06%	51,75%	57	37	39	50,04%
Clermont	13	10	48	49,08%	80,56%	50,32%	27	40	16	51,64%
Angers	20	2	47	43,38%	82,56%	50,67%	21	63	11	50,18%
Montpellier	11	11	47	51,05%	78,41%	48,40%	44	47	28	49,55%
Lorient	8	12	46	53,22%	82,65%	50,46%	41	37	25	51,28%

Annexe 2 : Tableau des corrélations entre les variables



Annexe 3 : Tableau de contribution des dimensions

	Contribution sur le nb de variable	Contribution en %	% cumulé
comp 1	6.391020857	58.10018961	58.10019
comp 2	1.758400710	15.98546100	74.08565
comp 3	0.980625783	8.91477984	83.00043
comp 4	0.648140122	5.89218293	88.89261
comp 5	0.517548039	4.70498217	93.59760
comp 6	0.369904593	3.36276903	96.96036
comp 7	0.181977695	1.65434268	98.61471
comp 8	0.078746811	0.71588010	99.33059
comp 9	0.043168529	0.39244117	99.72303
comp 10	0.025121579	0.22837799	99.95141
comp 11	0.005345282	0.04859347	100.00000

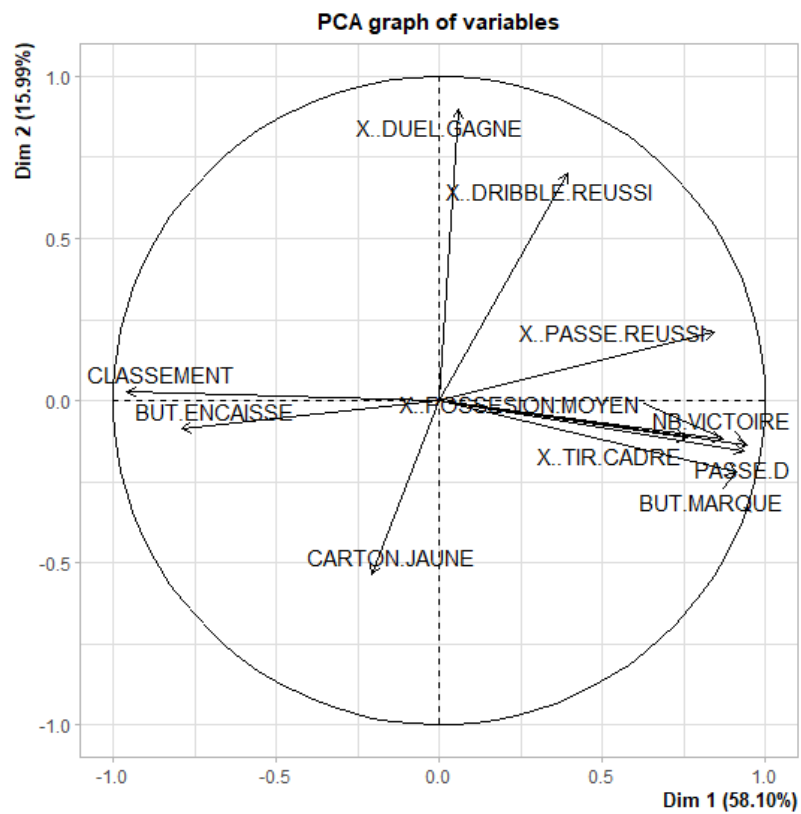
Annexe 4 : Tableau de contribution des variables

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
CLASSEMENT	14.32468847	0.03808616	0.12975585	9.354563729	2.740051e-01
NB. VICTOIRE	13.95389412	1.10178874	0.18654559	4.342814344	1.192759e-01
X. . POSSESSION.MOYEN	11.87021980	0.77559476	7.22209311	6.684641753	9.407619e+00
X. . TIR.CADRE	9.04248669	0.74701459	7.59355198	21.236611770	2.298964e+01
X. . PASSE.REUSSI	11.16471449	2.57347732	0.20208498	10.952318072	1.759086e+01
X. . DRIBBLE.REUSSI	2.44763464	28.00756813	20.72617404	3.780364965	1.197065e+00
BUT.MARQUE	13.06221471	2.71066532	1.74910849	1.065208467	3.391437e+00
BUT.ENCAISSE	9.77966540	0.42344370	1.84608752	41.487186102	5.881906e-05
PASSE.D	13.63535234	1.36223048	2.55243916	0.003132757	1.225550e+00
X. . DUEL.GAGNE	0.05437763	45.89297971	0.04968759	0.114916012	2.712095e+01
CARTON.JAUNE	0.66475171	16.36715110	57.74247169	0.978242029	1.668355e+01

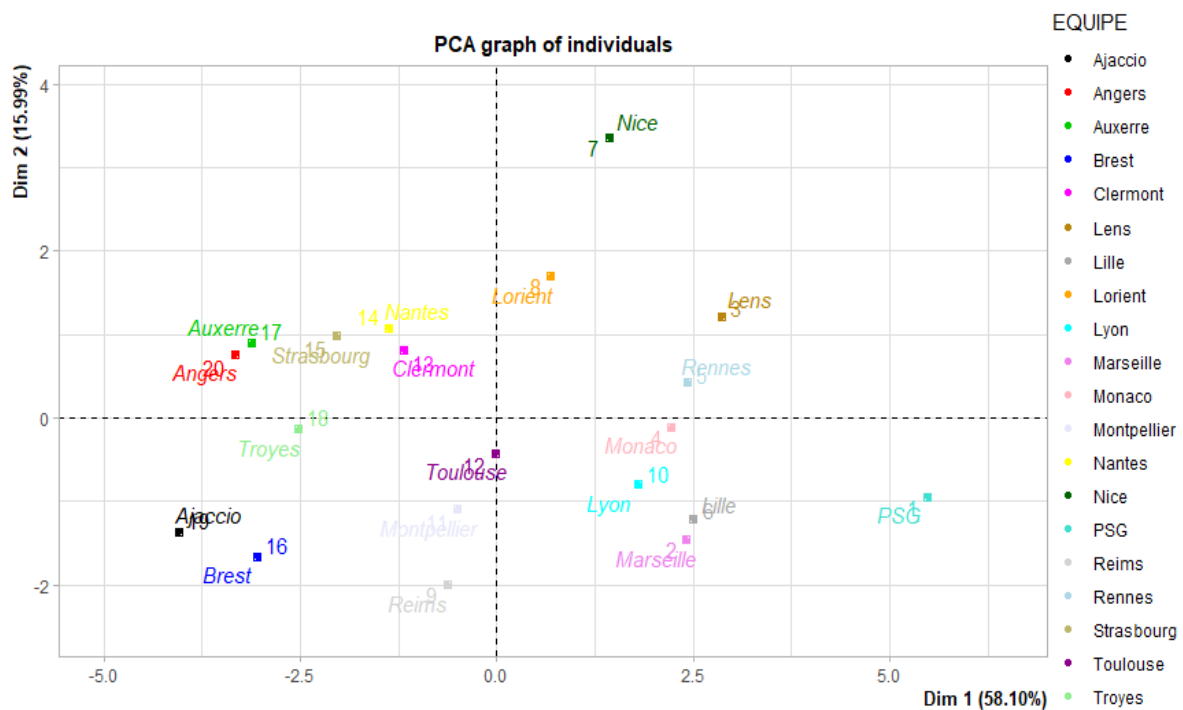
Annexe 5 : Tableau de contribution des individus

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
1	2.352705e+01	2.57017658	2.490407061	3.742074e+00	0.01221336
2	4.540372e+00	6.07006086	0.470382907	5.690618e+00	1.10384429
3	6.382845e+00	4.21012009	14.440832501	7.277117e+00	11.02837420
4	3.850731e+00	0.03859717	1.149313477	3.182109e-01	7.57912445
5	4.598801e+00	0.50261779	0.151706234	6.890530e-02	1.50046141
6	4.889565e+00	4.17528989	20.694816554	1.173802e+01	0.34690192
7	1.610471e+00	32.10924318	0.636218957	2.618483e+00	1.27898159
8	3.734158e-01	8.18178885	19.603194677	6.575421e-01	0.11730639
9	3.141628e-01	11.41847397	0.001117195	1.804196e+01	0.73186401
10	2.531385e+00	1.82095844	3.141269746	5.122876e+00	1.84387204
11	1.915495e-01	3.41093313	0.507944275	1.586018e-01	9.87146275
12	3.580068e-04	0.50440519	5.164849480	3.919277e+00	0.89888389
13	1.093291e+00	1.88845310	4.727192669	5.990141e-05	10.34596728
14	1.468832e+00	3.29597730	1.092416025	1.164382e+00	0.08605309
15	3.276575e+00	2.73115586	4.296614986	4.205557e-01	19.49351299
16	7.263686e+00	7.83044496	5.296503172	5.462958e+00	0.05764861
17	7.585811e+00	2.27083305	1.449301848	1.152308e+00	27.79724375
18	4.988529e+00	0.05234025	2.980460189	9.281707e+00	0.40833449
19	1.280473e+01	5.27927189	10.323575567	1.583867e+00	0.40739322
20	8.707847e+00	1.63885847	1.381882481	2.158048e+01	5.09055626

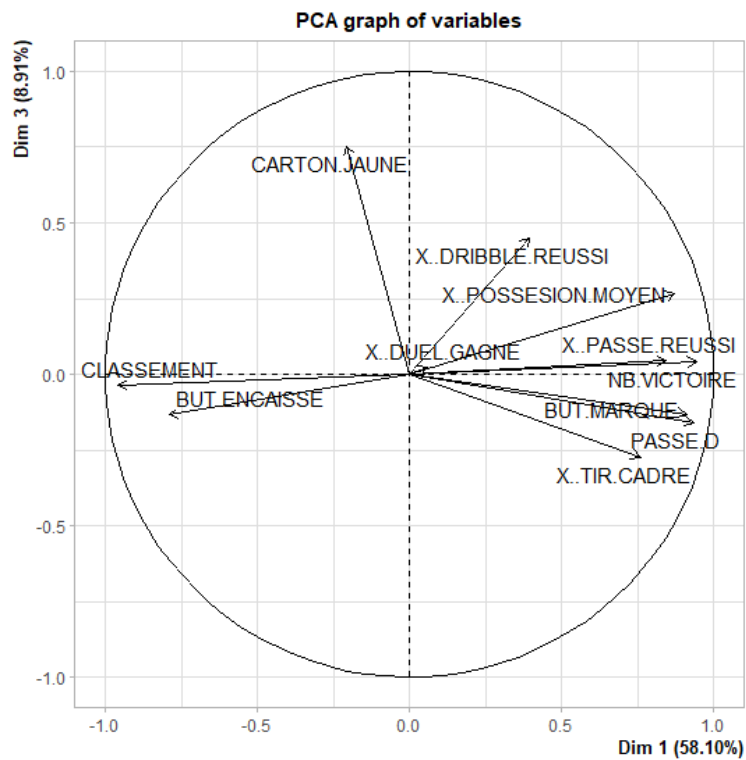
Annexe 6 : ACP Variables Dim 1 & 2



Annexe 7 : ACP Individus Dim 1 & 2



Annexe 8 : ACP Variables Dim 1 & 4



Annexe 9 : ACP Individus Dim 1 & 3

