

FootballMetric.AI

1. Introduction

Football (soccer) is the most popular sport in the world, and as the game evolves, so does the need for intelligent, data-driven analysis. Behind every pass, goal, and tackle lies a story told by numbers — and the ability to interpret those numbers brings a whole new dimension to both fans and professionals alike.

Football is no longer just a sport — it's a data-driven battlefield.

FootballMetric.AI is a comprehensive football analytics project powered by Python, machine learning, and visual storytelling. From comparing players and analyzing positional performance to building predictive models and discovering tactical insights, this project shows how data science can transform our understanding of the beautiful game.

2. General Exploratory Data Analysis

The dataset consists of 2,782 raw records and 267 columns, containing a wide range of variables. These include demographic, behavioral, and contextual features relevant to the analysis. To simplify the exploration and reduce complexity, a light version of the dataset with 160 selected columns was also prepared. This reduced version focuses on the most relevant or high-quality features, ensuring more efficient analysis while preserving the essential structure and information.

During this initial exploratory phase, we aim to:

- Understand the structure and types of data present
- Identify missing values and anomalies
- Examine basic statistical summaries (mean, median, standard deviation, etc.)
- Visualize distributions and relationships between key variables

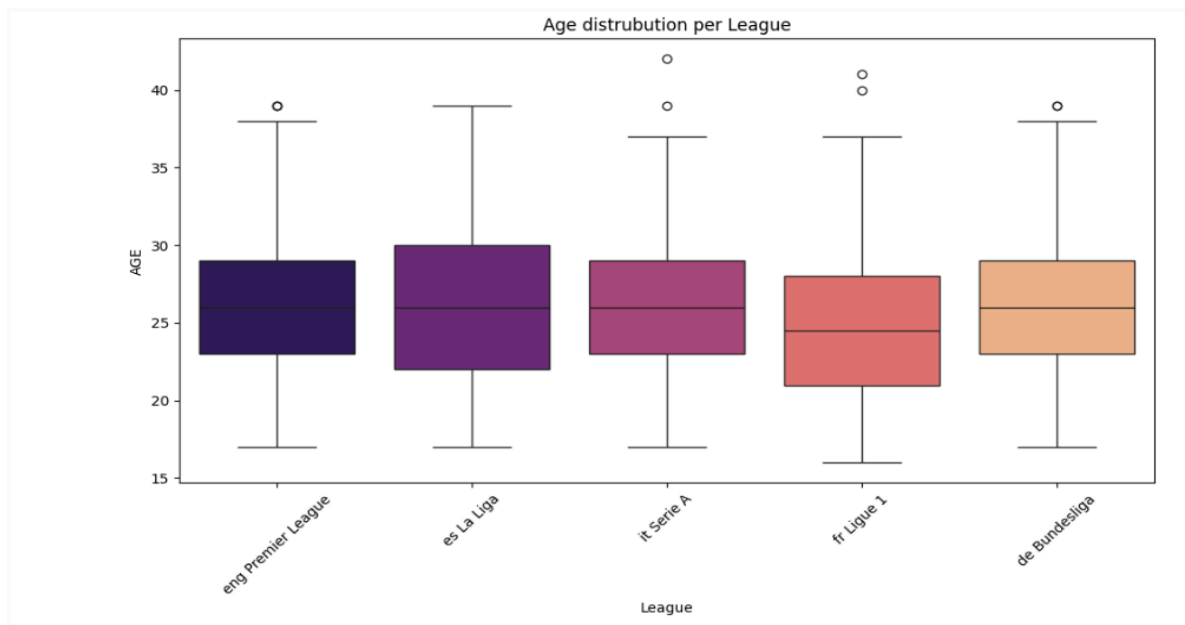
3. Comparative analysis of football leagues

A comparative study between the top 5 leagues in the world can show the points of power of each league by its player and the profile of its players

1. Age comparaison (youth players) :

The French Ligue 1 shows an interesting number of youth players with 270 players under the average age (25 y.o) and 153 under 22 y.o . (less than 75% of players).

In the other hand the spanish league La Liga has more experienced players with 305 player above the average age (>25 y.o) and 147 player over 29 y.o (older than 75% of players)



1.Age distribution per league plot (boxplot)

Youngest and oldest players in each league :

Youngest Players:

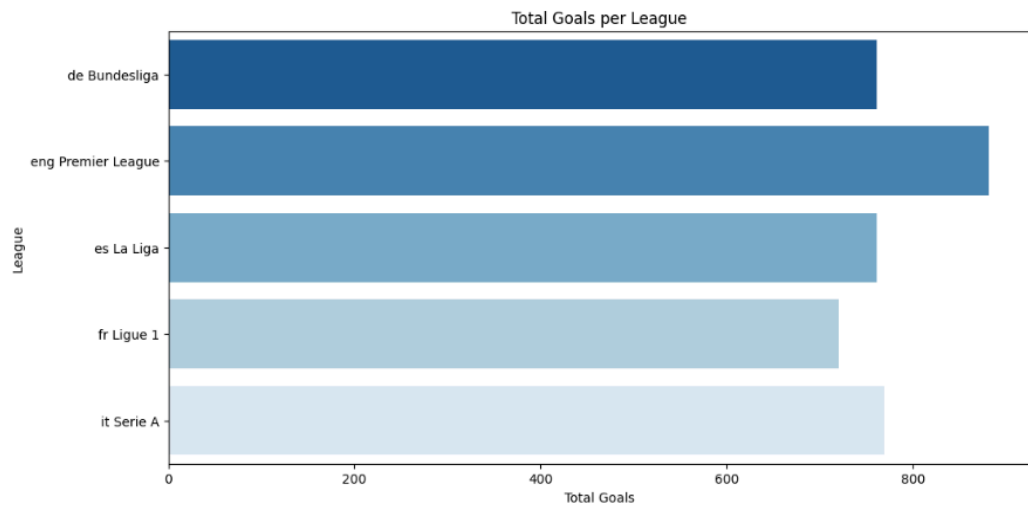
	Comp	Player	Age	Pos
962	de Bundesliga	Viggo Gebel	17.0	MF
1694	eng Premier League	Shumaira Mheuka	17.0	MF
307	es La Liga	Marc Bernal	17.0	MF
1805	fr Ligue 1	Djylan N'Guessan	16.0	FW
50	it Serie A	Honest Ahanor	17.0	DF

Oldest Players:

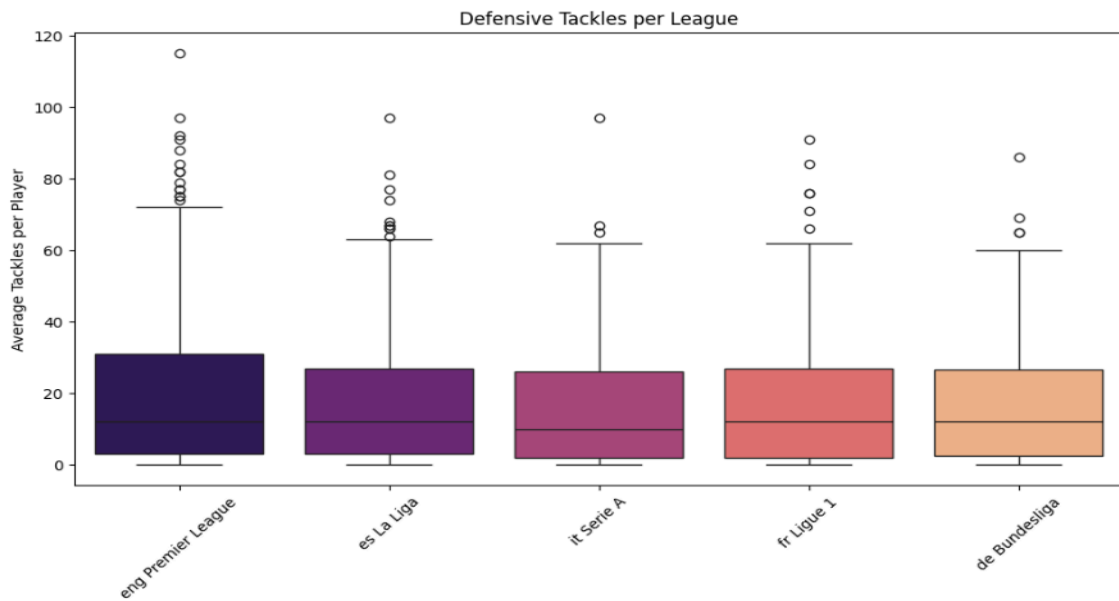
	Comp	Player	Age	Pos
1491	de Bundesliga	Anthony Losilla	39.0	MF
850	eng Premier League	Łukasz Fabiański	39.0	GK
75	es La Liga	Raúl Albiol	39.0	DF
649	fr Ligue 1	Dante	41.0	DF
2138	it Serie A	Pepe Reina	42.0	GK

2. Performance, Passing, and Defense Summary

The English Premier League leads the race when it comes to performance showing the good quality of its players .



2.Total Goals per league plot



4.Defensive Tackles per league plot (boxplot)

3. Players :

Top players from each league :

Top Scorers:

	Comp	Player	Gl	Age	Pos
1287	de Bundesliga	Harry Kane	23	31.0	FW
2243	eng Premier League	Mohamed Salah	27	32.0	FW
1449	es La Liga	Robert Lewandowski	25	36.0	FW
683	fr Ligue 1	Ousmane Dembélé	21	27.0	FW
2145	it Serie A	Mateo Retegui	22	25.0	FW

Top Assist Difference Players:

	Comp	Player	A-xAG	Age	Pos
116	de Bundesliga	Mohamed Amoura	4.6	24.0	FW
2166	eng Premier League	Antonee Robinson	6.0	27.0	DF
1858	es La Liga	Saúl Ñíguez	3.6	30.0	MF
1843	fr Ligue 1	João Neves	4.1	20.0	MF,DF
1502	it Serie A	Romelu Lukaku	4.9	31.0	FW

Top xG Players:

	Comp	Player	xG	Age	Pos
807	de Bundesliga	Hugo Ekitike	18.5	22.0	FW
2243	eng Premier League	Mohamed Salah	22.5	32.0	FW
1449	es La Liga	Robert Lewandowski	24.0	36.0	FW
823	fr Ligue 1	Emanuel Emegha	15.2	22.0	FW
1299	it Serie A	Moise Kean	15.6	25.0	FW

Top Tackling Players:

	Comp	Player	Tkl	Age	Pos
331	de Bundesliga	Tom Bischof	86.0	19.0	MF,FW
939	eng Premier League	Idrissa Gana Gueye	115.0	35.0	MF
1135	es La Liga	Omar El Hilali	97.0	21.0	DF
2285	fr Ligue 1	Andrey Santos	91.0	20.0	MF
914	it Serie A	Morten Frendrup	97.0	24.0	MF

4.Players comparative analysis

This analysis explores various aspects of player performance, including transfers, attacking efficiency, creativity, and defensive prowess.

1. Transferred Players

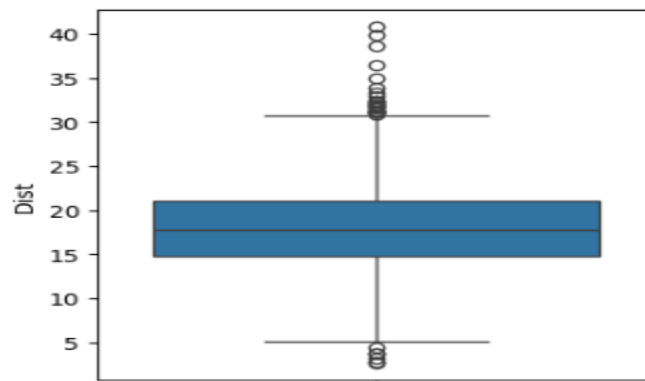
- there's 2631 different player in the dataset , 151 player has changed club in this period
- We observe that the transferred players are mostly from this nations : FRA,ITA,SPA,ENG
- French players had the biggest chance in the last market (ps:Ligue 1 has more youth players)

2. Attacking Performance

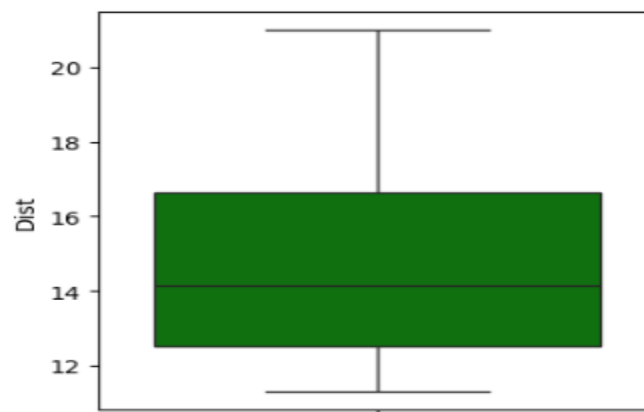
Scorers:

- Best 5 scorers (with penalties): Mo Salah >Lewandowski >Harry kane >Kylian Mbappé >Mateo Retegui.
- Best 5 scorers (without penalties): Lewandowski >Osmane Dembele >Erling Haaland >Mateo Retegui >Alexander Isak.

Distribution of shooting distance among players



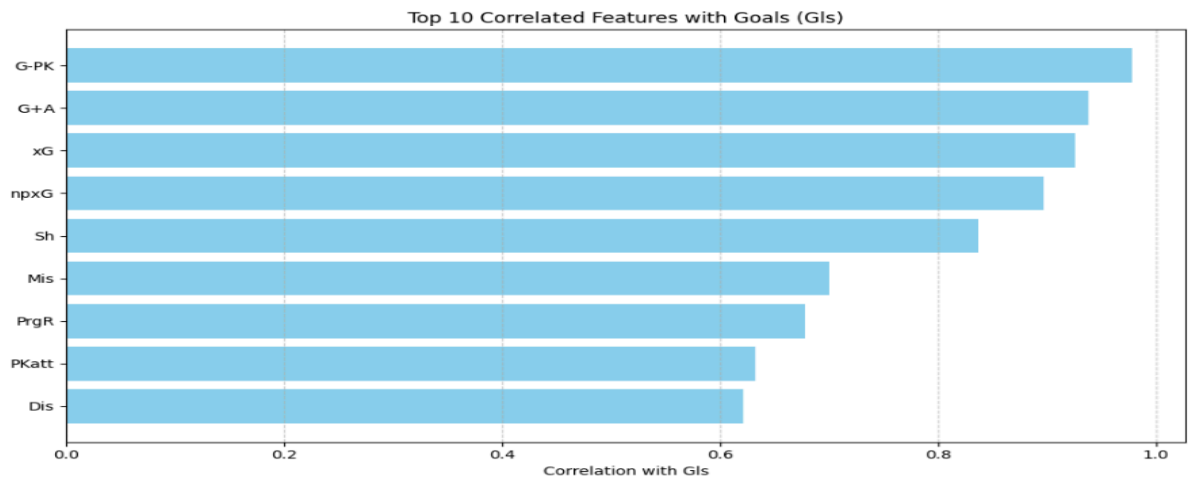
5.Average shooting distance for players plot (boxplot)



6.Average shooting distance for top scorers plot (boxplot)

Correlation of goals with other stats

to discover which conditions affects the most on scoring



7.Top 10 Correlated Features with goals (Gls) (barplot)

Assists:

Top assist makers: Mo Salah (17 ast) > Lamine Yamal (11 ast) >Mikkel Damsgaard (10 ast)
>Michael Olise (10 ast) > Bukayo Saka (10 ast)

Top Creators Based on Ast_stats_passing, xA, KP, and PPA

A creativity score calculated as a general factor to get best idea about a Player's chance creation

Player	Ast_stats_passing	xA	KP	PPA	CreativityScore
Michael Olise	10	10.5	64.0	95.0	179.5
Alex Baena	7	11.3	84.0	66.0	168.3
Mohamed Salah	17	7.2	70.0	74.0	168.2
Bruno Fernandes	9	6.3	75.0	60.0	150.3
Junya Ito	3	8.7	76.0	62.0	149.7
Lamine Yamal	11	8.8	51.0	75.0	145.8
Raphinha	8	8.3	75.0	54.0	145.3
Romano Schmid	4	6.9	56.0	75.0	141.9
Cole Palmer	7	7.7	71.0	51.0	136.7
Rayan Cherki	9	9.4	58.0	59.0	135.4

Players with Biggest Gap Between xA and Actual Assists

! Underperformers (unlucky or teammates missing chances):

Player	xA	Ast_stats_passing	xA_gap
Junya Ito	8.7	3	5.7
Maximilian Wittek	5.1	0	5.1
Alex Baena	11.3	7	4.3
Nadiem Amiri	8.0	4	4.0
Jamal Musiala	5.9	2	3.9
Mason Greenwood	6.9	3	3.9
Andrew Robertson	3.9	0	3.9
Pedri	7.6	4	3.6
Éderson	4.4	1	3.4
Jon Moncayola	3.3	0	3.3

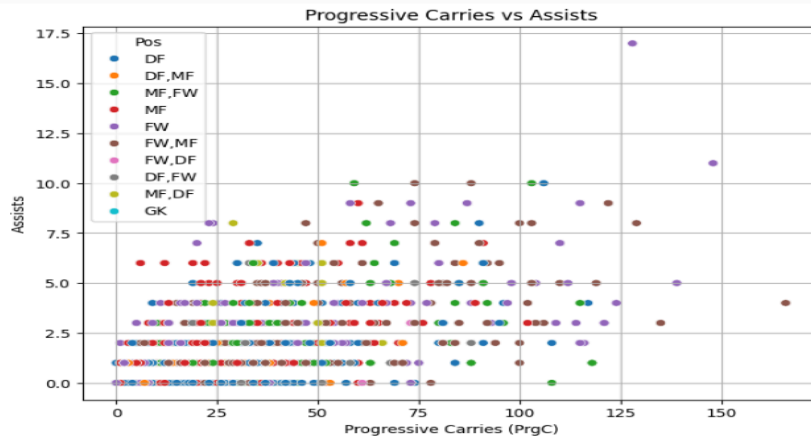
✨ Overperformers (possible luck or good finishers around):

Player	xA	Ast_stats_passing	xA_gap
Mohamed Salah	7.2	17	-9.8
Antonee Robinson	3.7	10	-6.3
Romelu Lukaku	1.9	8	-6.1
Oliver McBurnie	0.9	6	-5.1
Jacob Murphy	4.1	9	-4.9
Omar Marmoush	4.1	9	-4.9
Mohamed Amoura	3.2	8	-4.8
Bradley Barcola	4.4	9	-4.6
Eberechi Eze	3.4	8	-4.6
Ollie Watkins	1.6	6	-4.4

Relationship Between PrgP, PrgC and Assists

PrgP: progressive passes ,PrgC: progressive carries

Correlation PrgP ~ Assists: 0.473 | Correlation PrgC ~ Assists: 0.64



8.Progressive Carries vs Assists plot

Bad finishers :

these players should train on their finishing

Bad_finishing is factor to show the difference between expected goals and reel goals scored

Comp	Player	xG	Gls	Pos	Bad_Finishing
es La Liga	Abde Ezzalzouli	7.2	1	FW,MF	6.2
de Bundesliga	Hugo Ekitike	18.5	13	FW	5.5
fr Ligue 1	Breel Embolo	9.5	5	FW,MF	4.5
eng Premier League	Eberechi Eze	6.4	2	MF,FW	4.4
it Serie A	Tammy Abraham	7.0	3	FW	4.0
it Serie A	Mario Pašalić	5.8	2	MF	3.8
eng Premier League	Sávio	4.8	1	FW,MF	3.8
fr Ligue 1	Jim Allevinah	3.7	0	FW	3.7
es La Liga	Miguel	4.6	1	FW	3.6
fr Ligue 1	Marco Asensio	5.6	2	FW,MF	3.6

Great finishers :

players with goals more than their XG , goats or luck ?

Comp	Player	Gls	xG	Pos	Lucky_Finishing
it Serie A	Mateo Retegui	22	14.4	FW	7.6
eng Premier League	Chris Wood	18	11.0	FW	7.0
de Bundesliga	Patrik Schick	17	10.4	FW	6.6
eng Premier League	Matheus Cunha	13	6.7	MF,FW	6.3
de Bundesliga	Omar Marmoush	15	8.8	FW	6.2
fr Ligue 1	Ousmane Dembélé	21	14.9	FW	6.1
eng Premier League	Bryan Mbeumo	16	10.0	FW	6.0
de Bundesliga	Alexis Claude-Maurice	9	3.4	MF,FW	5.6
de Bundesliga	Harry Kane	23	17.9	FW	5.1
es La Liga	Sandro Ramírez	8	3.0	FW,MF	5.0

3. Defensive Performance

top 10 players with the most tackles won per 90 minutes

Player	Pos	TklW	90s	TklW_per_90
Yunus Emre Konak	FW,MF	4	0.3	13.333333
Asher Agbinone	MF	1	0.1	10.000000
Younes Lachaab	MF,FW	1	0.1	10.000000
Romain Esse	MF	1	0.1	10.000000
Rareș Ilie	MF	1	0.1	10.000000
Cheick Oumar Conde	MF	3	0.4	7.500000
Naci Ünüvar	MF,DF	2	0.3	6.666667
Vasilije Adžić	MF,FW	2	0.3	6.666667
Maxwel Cornet	FW	5	0.8	6.250000
Emi Buendía	MF,FW	6	1.1	5.454545

Who are the best all-around defenders (Tkl + Int)?

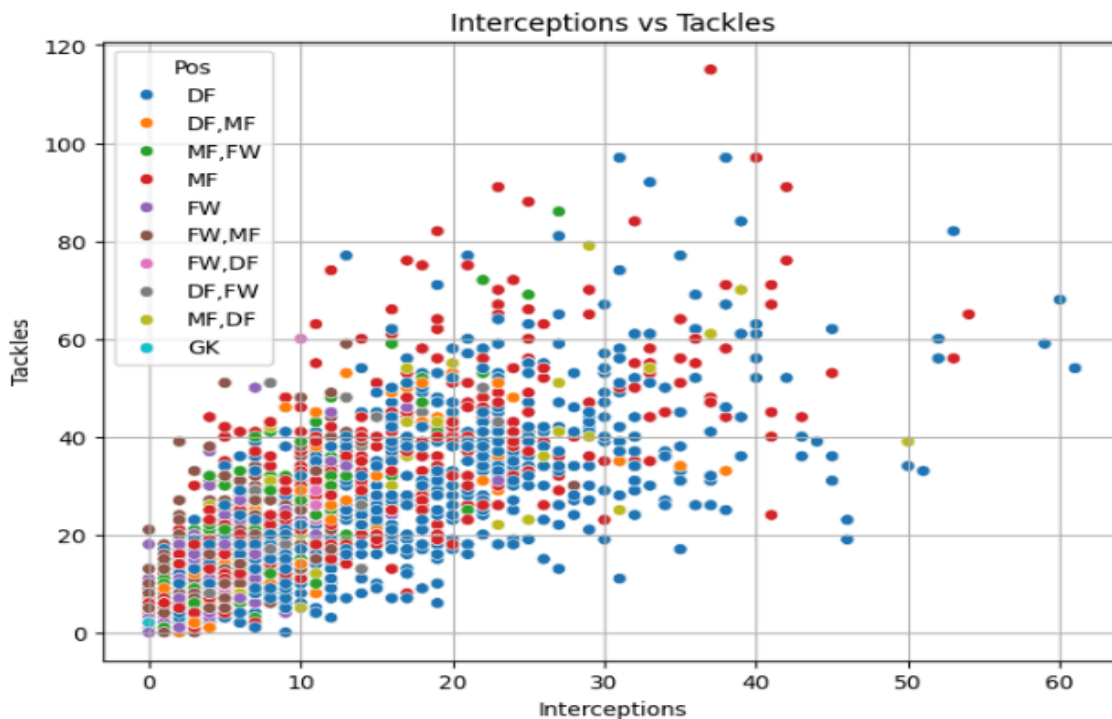
Player	Pos	Tkl	Int	Tkl+Int
Idrissa Gana Gueye	MF	115.0	37	152.0
Morten Frendrup	MF	97.0	40	137.0
Antonee Robinson	DF	82.0	53	135.0
Omar El Hilali	DF	97.0	38	135.0
Moisés Caicedo	MF	91.0	42	133.0
Carmona	DF	68.0	60	128.0
Daniel Muñoz	DF	97.0	31	128.0
Noussair Mazraoui	DF	92.0	33	125.0
Victor Bernth Kristiansen	DF	84.0	39	123.0
Senne Lynen	MF	65.0	54	119.0

Which players are most involved defensively (Tkl, Int, Blocks, Clr) – combined total?

Def_Activity calculated factor to get a better rating to defender's activity

Player	Pos	Tkl	Int	Blocks	Clr	Def_Activity
James Tarkowski	DF	61.0	39	12.0	202.0	314.0
Patrick Mainka	DF	54.0	61	14.0	164.0	293.0
Antonee Robinson	DF	82.0	53	44.0	109.0	288.0
Maxence Lacroix	DF	62.0	45	12.0	159.0	278.0
Ardian Ismajli	DF	23.0	46	3.0	205.0	277.0
Murillo	DF	44.0	29	6.0	196.0	275.0
Omar El Hilali	DF	97.0	38	23.0	114.0	272.0
Diego Coppola	DF	59.0	59	4.0	146.0	268.0
Aaron Wan-Bissaka	DF	60.0	52	49.0	107.0	268.0
Diogo Leite	DF	56.0	52	17.0	141.0	266.0

Relationship between Interceptions and Tackles?



9.Interception relation with Tackles

Correlation between Interceptions and Tackles: 0.776

Defensive errors leading to goal

Players with Most Errors Leading to Goals:

Player	Pos	Err
Joško Gvardiol	DF	8.0
Calvin Bassey	DF	7.0
Mario Martín	MF	7.0
Dario Essugo	MF	7.0
Mickael Nade	DF	6.0
Taylor Harwood-Bellis	DF	6.0
Ainsley Maitland-Niles	DF	6.0
Mateus Fernandes	MF	5.0
William Saliba	DF	5.0
Modibo Sagnan	DF	5.0