

Badanie statystyk ofensywnych najsukuteczniejszych piłkarzy w sezonie 2022/23 w top 5 ligach Europejskich

Wojciech Liberacki

2023-12-15

1. Dane

Dane przedstawiają statystyki 35 piłkarzy z największą liczbą bramek zdobytych w najlepszych pięciu ligach Europejskich (angielskiej, hiszpańskiej, włoskiej, niemieckiej i francuskiej) w sezonie 2022/23

topscorers <- read.csv("C:/Users/Wojte/OneDrive/Desktop/Projekty/Statystyczna_analiza_danych/topscorers.csv")
topscorers\$PKmiss<- topscorders\$PK-topscorders\$PKatt topscorders\$GlsPer90Min <- topscorders\$Gls/(topscorders\$Min/90)
topscorders\$Min <- NULL topscorders\$ast <- NULL topscorders\$PK <- NULL topscorders\$PKatt <- NULL topscorders
Player Gls G.Sh G.xG SoT Dist PKmiss GlsPer90Min ## 1 Erling Haaland 36 0.25 7.6 45.7 12.6 0 1.1700975 ## 2 Harry Kane 38 0.20 8.5 43.5 16.0 -1 0.7929515 ## 3 Kylian Mbappé 29 0.18 2.7 50.7 15.9 -2 0.9261888 ## 4 Alexandre Lacazette 27 0.20 2.8 43.9 14.4 -2 0.8296347 ## 5 Victor Osimhen 26 0.18 4.7 39.4 12.8 -1 0.9119252 ## 6 Jonathan David 24 0.15 0.4 46.3 15.8 -1 0.6822489 ## 7 Robert Lewandowski 23 0.17 -1.3 41.8 13.8 -1 0.7270811 ## 8 Folarin Balogun 21 0.12 -5.6 41.3 15.2 -1 0.6302101 ## 9 Karim Benzema 21 0.18 2.6 39.4 15.9 -1 0.7509570 ## 10 Lautaro Martinez 21 0.16 2.9 40.8 15.1 -1 0.7336957 ## 11 Habib Diallo 20 0.19 2.2 40.0 13.0 -1 0.6252171 ## 12 Ivan Toney 20 0.16 1.3 37.9 16.5 -1 0.6099627 ## 13 Elye Wahi 19 0.29 6.8 50.1 14.5 0 0.6804616 ## 14 Karim Benzema 19 0.12 2.5 30.4 15.9 -1 0.8390519 ## 15 Wissam Ben Yedder 19 0.21 1.6 54.5 14.2 -1 0.8089931 ## 16 Mohamed Salah 19 0.14 -2.7 37.2 15.1 -2 0.5197568 ## 17 Terem Moffi 18 0.22 2.5 46.8 14.3 -1 0.5969049 ## 18 Callum Wilson 18 0.21 6.5 44.3 12.4 0 0.8630944 ## 19 Marcus Rashford 17 0.16 1.6 45.4 16.4 0 0.5314545 ## 20 Boulaye Dia 16 0.32 7.2 44.7 15.0 0 0.5718824 ## 21 Niclas Füllkrug 16 0.15 3.2 43.2 16.0 0 0.5813484 ## 22 Joselu 16 0.13 1.5 37.9 16.1 0 0.4816054 ## 23 Christopher Nkunku 16 0.19 1.6 35.7 15.9 -1 0.7590933 ## 24 Lionel Messi 16 0.12 0.5 42.6 18.3 0 0.5075784 ## 25 Vedat Muriqi 15 0.15 2.5 35.2 15.3 -1 0.4596527 ## 26 Rafael Leão 15 0.16 4.2 31.5 17.9 0 0.5576208 ## 27 Gabriel Martinelli 15 0.19 5.7 38.0 15.4 0 0.4640445 ## 28 Vincenzo Grifo 15 0.14 4.7 43.1 22.6 -1 0.5693132 ## 29 Borja Iglesias 15 0.17 0.1 36.2 15.2 -1 0.5708245 ## 30 Martin Ødegaard 15 0.16 5.0 33.3 19.2 0 0.4317237 ## 31 Gaetan Laborde 15 0.18 1.5 52.9 15.4 -1 0.4900181 ## 32 Amine Gouiri 15 0.19 3.9 40.5 13.5 0 0.4910877 ## 33 Ollie Watkins 15 0.17 -1.5 53.6 12.6 -1 0.4317237 ## 34 Randal Kolo Muani 15 0.17 1.8 40.0 13.0 0 0.5131129 ## 35 Antoine Griezmann 15 0.14 2.7 39.6 17.6 0 0.4721931

Zmienne:

- Gls: bramki zdobyte
- G.Sh: bramki zdobyte / ilość oddanych strzałów
- G.xG: bramki zdobyte - expected Goals(statystyka określająca bramki oczekiwane)
- SoT: średnia odległość oddawanego strzału
- PKmiss: rzuty karne niestrzelone (np. -1 - jeden karne niestrzelony)
- GlsPer90Min: bramki zdobyte na 90 minut piłkarza na boisku

2. Charakterystyka zmiennych

Statystyki opisowe

summary(topscorers)
Player Gls G.Sh G.xG SoT Dist PKmiss GlsPer90Min ## Length:35 Min. :15.0 Min. :0.120 Min. : -5.600 ## Class :character 1st Qu.:15.0 1st Qu.:0.150 1st Qu.: 0.900 ## Mode :character Median :18.0 Median :0.170 Median : 2.500 ## Mean :19.2 Mean :0.178 Mean : 2.303 ## 3rd Qu.:21.0 3rd Qu.:0.190 3rd Qu.: 3.750 ## Max. :36.0 Max. :0.320 Max. : 8.500 ## SoT. Dist PKmiss GlsPer90Min ## Min. :11.50 Min. :12.40 Min. : -2.0000 Min. :0.4217 ## 1st Qu.:37.95 1st Qu.:14.80 1st Qu.: -1.0000 1st Qu.:0.5103 ## Median :41.80 Median :15.20 Median : -1.0000 Median :0.5969 ## Mean :42.56 Mean :15.31 Mean : -0.6571 Mean :0.6455 ## 3rd Qu.:45.55 3rd Qu.:18.00 3rd Qu.: 0.0000 3rd Qu.:0.7550 ## Max. :58.10 Max. :22.60 Max. : 0.0000 Max. :1.1701

Współczynniki zmienności

coefficient_of_variation<- function(x){ cv <- sd(x) / abs(mean(x)) return(cv) } cvs <- sapply(topscorers[, -1], coefficient_of_variation) cvs
Gls G.Sh G.xG SoT Dist PKmiss ## 0.2685919 0.2444267 1.2094600 0.1445827 0.1367710 0.9724804 ## GlsPer90Min ## 0.2643011

Współczynniki zmienności dla każdej zmiennej są większe od 0.1, więc nie ma tu problemu w kontekście analizy skupień.

library(corrplot) cor_matrix <- cor(topscorers[, -1]) corrplot(cor_matrix, method = "number")
Gls G.Sh G.xG SoT Dist PKmiss GlsPer90Min ## 1.00 0.24 0.23 0.24 -0.32 -0.36 0.84 ## G.Sh 0.24 1.00 0.61 0.47 -0.42 0.18 0.33 ## G.xG 0.23 0.61 1.00 0.10 0.11 0.37 0.18 ## SoT 0.24 0.47 0.10 1.00 -0.30 -0.08 0.24 ## Dist -0.32 -0.42 0.11 -0.30 1.00 0.89 -0.37 ## PKmiss -0.36 0.18 0.37 -0.08 -0.84 1.00 -0.25 ## GlsPer90Min 0.84 0.33 0.18 0.24 -0.37 -0.25 1.00

Dla wszystkich par zmiennych korelacja na moduł mniejsza niż 0.9, więc nie ma powodu do odrzucenia do analizy skupień.

3. Cel projektu

Celem projektu jest analiza najlepiej strzelających zawodników top 5 lig piłkarskich, stworzenie rankingu najlepszych napastników (i graczy ofensywnych) w oparciu o statystyki definiujące jakość piłkarza i jego umiejętność strzelania bramek. Analiza ma też na celu wyłonic grup "nieoczywistych" zawodników z predyspozycjami na stanowienie samej czołówki ogólnej klasyfikacji strzelców(zawodnicy skuteczni z niekoniecznie najwyższą ilością bramek).

4. Porządkowanie liniowe

Metoda Hellwiga z rangami

1. Wszystkie zmienne są skumulowanymi dlatego je standardyzuje

topscorers_scaled <- topscorders
variables_to_standardize <- c("Gls", "G.Sh", "G.xG", "SoT.", "Dist", "PKmiss", "GlsPer90Min")
topscorers_scaled[, variables_to_standardize] <- scale(topscorers_scaled[, variables_to_standardize])
2. Wartość maksymalna z każdej kolumny
pattern <- apply(topscorers_scaled[,2:8],2,max) pattern
Gls G.Sh G.xG SoT Dist PKmiss ## 3.258822 3.263772 2.070099 2.525417 3.478414 1.028289 ## GlsPer90Min ## 3.074494
3. Odległości obiektów od wzorca(uwzględniono rangi):
Gls - 0.125 G.Sh - 0.25 G.xG - 0.3 SoT - 0.025 Dist - 0.025 PKmiss - 0.1 GlsPer90Min - 0.175
wages <- c(0.125,0.25,0.25,0.05,0.05,0.1,0.1,0.175) distances <- apply(topscorers_scaled[,2:8], 1, function(row) {sqrt(sum(wages*(row-pattern)^2))}) distances
[1] 1.418821 1.954376 2.418375 2.476569 2.480052 3.007887 3.116560 3.951687 ## [8] 2.742520 2.964943 2.986249 3.162640 2.934770 3.690351 2.775573 3.033232 ## [17] 2.882767 2.967009 3.292028 2.585969 3.331162 3.646869 3.174228 3.696608 ## [25] 3.616418 3.368976 3.189061 3.387957 3.557453 3.393795 3.375620 3.301757 ## [33] 3.745072 3.488347 3.550543
4. Odległość "możliwie daleka"
max_dist <- mean(distances) + 2*sd(distances) max_dist
[1] 4.229464
5. Obliczone współczynniki:
hellwig <- c(1- distances/max_dist) result<- data.frame(topscorers[,1], hellwig) names(result) <- c("name","hellwig") result[order(result\$hellwig, decreasing = TRUE),]
name hellwig ## 1 Erling Haaland 0.6649806 ## 2 Harry Kane 0.53791406 ## 13 Elye Wahi 0.46956155 ## 3 Kylian Mbappé 0.42820776 ## 4 Alexandre Lacazette 0.41444048 ## 5 Victor Osimhen 0.41302500 ## 20 Boulaye Dia 0.38859449 ## 9 Lois Openda 0.35156604 ## 15 Wissam Ben Yedder 0.34375313 ## 17 Terem Moffi 0.31848040 ## 10 Lautaro Martinez 0.30343550 ## 18 Callum Wilson 0.29849050 ## 11 Habib Diallo 0.29157741 ## 6 Jonathan David 0.28883034 ## 7 Robert Lewandowski 0.26331332 ## 12 Ivan Toney 0.25218096 ## 23 Christopher Nkunku 0.24949634 ## 27 Gabriel Martinelli 0.24598942 ## 19 Marcus Rashford 0.22164423 ## 32 Amine Gouiri 0.21050352 ## 21 Niclas Füllkrug 0.21239147 ## 26 Rafael Leão 0.20345079 ## 31 Gaetan Laborde 0.20167987 ## 28 Vincenzo Grifo 0.18900294 ## 30 Martin Ødegaard 0.19758282 ## 34 Randal Kolo Muani 0.17522720 ## 35 Antoine Griezmann 0.16052187 ## 29 Borja Iglesias 0.15880805 ## 25 Vedat Muriqi 0.14046549 ## 22 Joselu 0.13774087 ## 24 Lionel Messi 0.12740537 ## 14 Karim Benzema 0.12557454 ## 33 Ollie Watkins 0.11452795 ## 16 Mohamed Salah 0.09368847 ## 8 Folarin Balogun 0.06567670

Interpretacja

Bez zaskoczenia Erling Haaland okazał się najlepszym napastnikiem. Jednak analiza przyniosła porządną efek, pokazała ona dużo większą jakość strzelczą takich zawodników jak Wahi (miejsce w rankingu: 13 -> 3) czy Dia (20 -> 7) i niższą niż ta na pierwszy rzut oka widoczną, takich zawodników jak Lewandowski (7 -> 15).

5. Analiza skupień

Odleglosci miedzy obiektami(metoda Euklidesowa):

library(clusterSim) data_distances <- dist(topscorers_scaled[,2:8], method = "euclidean")
Metoda Warda
data_groups <- hclust(data_distances, method = "ward.D") data_groups
Call: ## hclust(d = data_distances, method = "ward.D") ## ## Cluster method : ward.D ## Distance : euclidean ## Number of objects: 35

Dendrogram

dendrogram <- plot(data_groups, labels = topscorders[,1], xlab="Footballers", hang = -1)
Cluster Dendrogram
Height
Footballers
hclust ("ward.D")

Sprawdzam, która liczba grup (2 lub 3) jest bardziej opytymalna przy pomocy indeksów:

Silhouette

index.S(data_distances, cutree(data_groups, k=2))
[1] 0.1762526
index.S(data_distances, cutree(data_groups, k=3))
[1] 0.1919911

Huberta i Levine'a

index.G3(data_distances, cutree(data_groups, k=2))
[1] 0.3736044
index.G2(data_distances, cutree(data_groups, k=3))
[1] 0.3847736

Bakera i Huberta

index.G2(data_distances, cutree(data_groups, k=2))
[1] 0.2564772
index.G2(data_distances, cutree(data_groups, k=3))
[1] 0.5810385

Każdy z indeksów "sugeruje" podział na 3 grupy.

Interpretacja

g1<- data.frame(topscorders[c(20,20,30,27,32,34,10,21,24,22),]) g2 <- data.frame(topscorders[c(13,33,9,18,6,7,15,11,17,25,29,13,10,12,6,14),]) g3 <- data.frame(topscorders[c(13,29,1,3,4,2,5),])
Grupa nr.1 - Grupa z zawodnikami, którzy mają najmniej bramek w porównaniu do innych grup, ale są bardzo skuteczni (średnia G.xG = 3.04, dla gr. 2 = 0.35)
g1

Player Gls G.Sh G.xG SoT Dist PKmiss GlsPer90Min ## 26 Vincenzo Grifo 15 0.14 4.7 43.1 22.6 -1 0.5686312 ## 28 Rafael Leao 15 0.16 4.2 31.5 17.9 0 0.5576208 ## 30 Martin Ødegaard 15 0.16 5.0 33.3 19.2 0 0.4317237 ## 27 Gabriel Martinelli 15 0.19 5.7 38.0 15.4 0 0.4840445 ## 32 Amine Gouiri 15 0.19 3.9 40.5 13.5 0 0.4910877 ## 34 Randal Kolo Muani 15 0.17 1.8 40.0 13.0 0 0.5131129 ## 18 Callum Wilson 18 0.21 6.5 44.3 12.4 0 0.8630974 ## 21 Niclas Füllkrug 16 0.15 3.2 43.2 16.0 0 0.5813484 ## 24 Lionel Messi 16 0.12 0.5 42.6 18.3 0 0.5075784 ## 22 Joselu 16 0.13 1.5 37.9 16.1 0 0.4816054
summary(g1[,2:8])

Gls G.Sh G.xG SoT ## Min. :15.0 Min. :0.120 Min. :0.500 Min. :31.50 ## 1st Qu.:15.0 1st Qu.:0.1425 1st Qu.:1.575 1st Qu.:37.92 ## Median :16.00 Median :0.1600 Median :3.250 Median :40.25 ## Mean :15.6 Mean :0.1620 Mean :3.040 Mean :39.44 ## 3rd Qu.:16.0 3rd Qu.:0.1850 3rd Qu.:14.575 3rd Qu.:42.98 ## Max. :18.0 Max. :0.2100 Max. :5.700 Max. :44.30 ## Dist PKmiss GlsPer90Min ## Min. :12.40 Min. : -1.0 Min. :0.4317 ## 1st Qu.:13.95 1st Qu.: 0.0 1st Qu.:0.4858 ## Median :16.05 Median : 0.0 Median :0.5103 ## Mean :16.44 Mean : -0.1 Mean :0.5472 ## 3rd Qu.:18.20 3rd Qu.: 0.0 3rd Qu.:0.5599 ## Max. :22.60 Max. : 0.0 Max. :0.8631

Grupa nr.2 - Grupa zawodników strzelających średnio więcej bramek, najsukuteczniejszych i stwarzających największe zagrożenie

g2
Player Gls G.Sh G.xG SoT Dist PKmiss GlsPer90Min ## 31 Gaetan Laborde 15 0.18 1.5 52.9 15.4 -1 0.4900181 ## 33 Ollie Watkins 15 0.17 -1.5 53.6 12.6 -1 0.4317237 ## 9 Lois Openda 21 0.19 2.6 47.6 13.1 0 0.7508939 ## 18 Callum Wilson 18 0.21 6.5 44.3 12.4 0 0.8630974 ## 6 Jonathan David 20 0.15 0.4 46.3 15.8 -1 0.6822489 ## 7 Robert Lewandowski 23 0.17 -1.3 41.8 13.8 -1 0.7270811 ## 15 Wissam Ben Yedder 19 0.21 1.6 54.5 14.2 -1 0.8089931 ## 11 Habib Diallo 20 0.19 2.2 40.0 13.0 -1 0.6252171 ## 17 Terem Moffi 18 0.22 2.5 46.8 14.3 -1 0.5969049 ## 25 Vedat Muriqi 15 0.15 2.5 35.2 15.3 -1 0.4596527 ## 29 Borja Iglesias 15 0.17 0.1 36.2 15.2 -1 0.5708245 ## 23 Christopher Nkunku 16 0.19 1.6 35.7 15.9 -1 0.7590933 ## 10 Lautaro Martinez 21 0.16 2.9 40.8 15.1 -1 0.7336957 ## 12 Ivan Toney 20 0.16 1.3 37.9 16.5 -1 0.6099627 ## 16 Mohamed Salah 19 0.14 -2.7 37.2 15.1 -2 0.5197568 ## 8 Folarin Balogun 21 0.12 -5.6 41.3 15.2 -1 0.6302101 ## 14 Karim Benzema 19 0.12 -2.5 36.4 15.9 -1 0.8390579
summary(g2[,2:8])

Gls G.Sh G.xG SoT ## Min. :15.00 Min. :0.1200 Min. : -5.6000 Min. :35.20 ## 1st Qu.:16.00 1st Qu.:0.1500 1st Qu.:1.3000 1st Qu.:37.20 ## Median :18.00 Median :0.1700 Median :1.3000 Median :41.30 ## Mean :18.76 Mean :0.1706 Mean :1.7571 Mean :46.57 ## 3rd Qu.:21.00 3rd Qu.:0.1900 3rd Qu.:2.2000 3rd Qu.:46.80 ## Max. :24.00 Max. :0.2200 Max. :2.9000 Max. :54.50 ## Dist PKmiss GlsPer90Min ## Min. :12.40 Min. : -2.0000 Min. :0.4317 ## 1st Qu.:13.80 1st Qu.: -1.0000 1st Qu.:0.5708 ## Median :15.10 Median : -1.0000 Median :0.6302 ## Mean :14.64 Mean : -0.9412 Mean :0.6528 ## 3rd Qu.:15.40 3rd Qu.: -1.0000 3rd Qu.:0.7599 ## Max. :16.50 Max. : 0.0000 Max. :0.8631

Grupa nr.3 - Grupa zawodników strzelających najwięcej bramek, najsukuteczniejszych i stwarzających największe zagrożenie

g3
Player Gls G.Sh G.xG SoT Dist PKmiss GlsPer90Min ## 12 Erling Haaland 36 0.20 7.6 45.7 12.6 0 0.6804616 ## 20 Boulaye Dia 16 0.32 7.2 44.7 15.0 0 0.5718824 ## 1 Erling Haaland 36 0.25 7.6 45.7 12.6 0 1.1700975 ## 3 Kylian Mbappé 29 0.18 2.7 50.7 15.9 -2 0.9261888 ## 4 Alexandre Lacazette 27 0.20 2.8 43.9 14.4 -2 0.8296347 ## 2 Harry Kane 38 0.20 8.5 43.5 16.0 -1 0.7929515 ## 5 Victor Osimhen 26 0.18 4.7 39.4 12.8 -1 0.9119252
summary(g3[,2:8])

Gls G.Sh G.xG SoT ## Min. :16.00 Min. :0.1800 Min. :2.7000 Min. :39.40 ## 1st Qu.:22.50 1st Qu.:0.1900 1st Qu.:3.7500 1st Qu.:43.70 ## Median :27.00 Median :0.2000 Median :6.8000 Median :44.70 ## Mean :26.14 Mean :0.2314 Mean :5.7571 Mean :46.57 ## 3rd Qu.:29.50 3rd Qu.:0.2700 3rd Qu.:7.4000 3rd Qu.:48.20 ## Max. :36.00 Max. :0.3200 Max. :8.5000 Max. :58.10 ## Dist PKmiss GlsPer90Min ## Min. :12.60 Min. : -2.0000 Min. :0.5719 ## 1st Qu.:13.60 1st Qu.: -1.5000 1st Qu.:0.7367 ## Median :14.50 Median : -1.0000 Median :0.8296 ## Mean :14.46 Mean : -0.8571 Mean :0.8404 ## 3rd Qu.:15.45 3rd Qu.: 0.0000 3rd Qu.:0.9191 ## Max. :16.00 Max. : 0.0000 Max. :1.1701
--