

07 – Relational data: Football dataset

Autor: Peitsch, Pablo

mail: preynosopeitsch@estudiantes.unsam.edu.ar

Github: @PPeitsch

Fecha: 2023-05-15

Se realiza un estudio entre los mejores jugadores de una liga de fútbol europea. Para esto se utiliza el dataset de las mejores 5 ligas de fútbol en Europa de 2014 a 2020 [1]. Se desea saber, por un lado, cuáles son los 5 jugadores más influyentes en los partidos y, por otro, los 5 jugadores con peor fairplay. Para el primer caso, se contabilizan el número de asistencias y tiros al arco de cada jugador en todas las ligas; para ello se puntúan a los jugadores según: $\text{score} = (\text{total_goals} + \text{total_asistencias}) / \text{total_players}$. Se pueden observar los resultados en la Tabla 1, y en la Figura 1 respecto al resto de jugadores. En el segundo caso, se evalúan las tarjetas amarillas y rojas y se cuenta el total de tarjetas, ponderando 1 roja cada 4 amarillas para el total de tarjetas. Se puede ver el resultado de este último en la Figura 2, debajo.

playerID	name	total_asistencias	total_goals	score	total_players
2371	Cristiano Ronaldo	74	215	0.05434374	5318
2098	Luis Suárez	96	168	0.04964272	5318
227	Robert Lewandowski	56	203	0.04870252	5318
2099	Neymar	119	114	0.04381346	5318
647	Harry Kane	57	163	0.04136894	5318

Tabla 1. Total de tarjetas rojas, amarillas y la suma de estas dos, teniendo en cuenta cada 4 amarillas 1 roja.

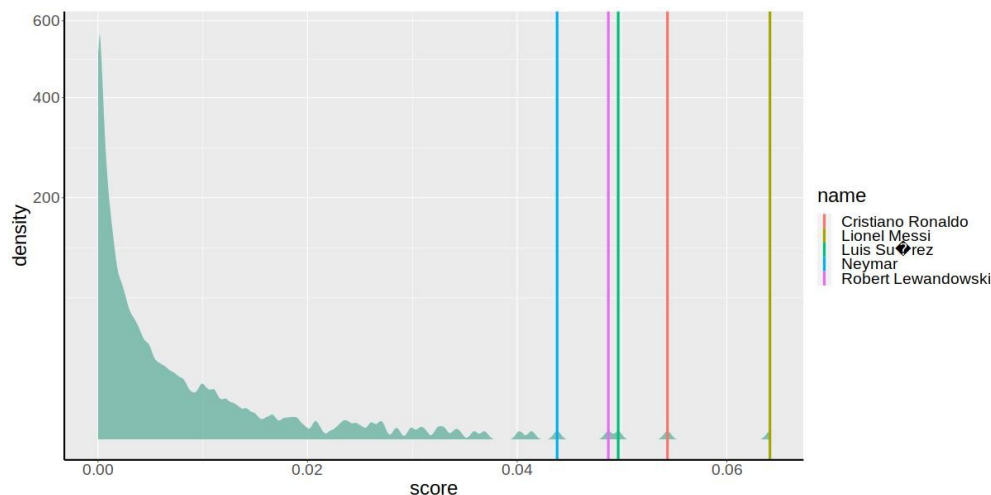


Figura 1. Score de los mejores 5 jugadores respecto del resto.

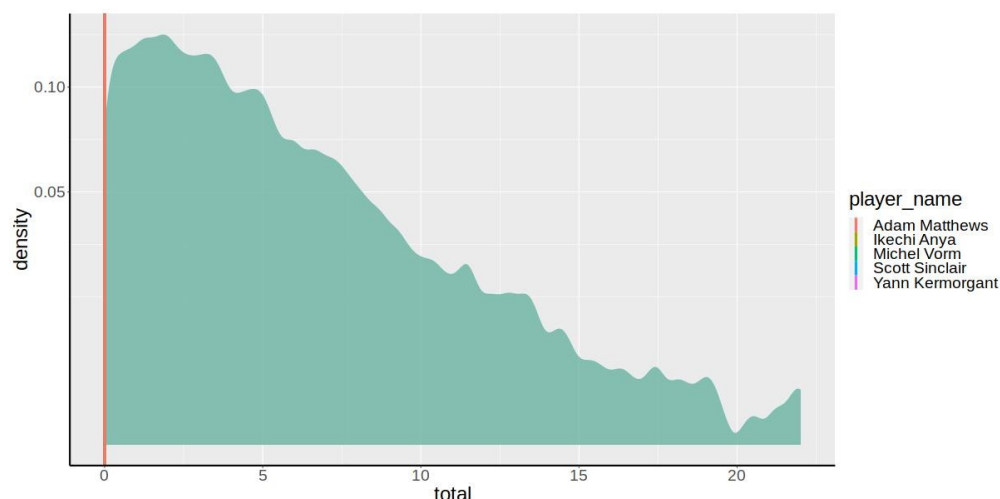


Figura 2. Score de los mejores 5 fairplayers respecto del resto.

REFERENCIAS.

1. Football Dataset, <https://www.kaggle.com/datasets/technika148/football-database>
2. Xstats, <https://understat.com/>
3. Expected Goals (xG), <https://onefootball.com/es/noticias/que-son-los-expected-goals-xg-30199741>