#### **07 – Relational data: Football dataset**

##### **Autor: Peitsch, Pablo mail:** [preynosopeitsch@estudiantes.unsam.edu.ar](mailto:preynosopeitsch@estudiantes.unsam.edu.ar)

##### **Github:**[@PPeitsch](https://github.com/PPeitsch) **Fecha: 2023-05-15**

##### Se realiza un estudio entre los mejores jugadores de una liga de fútbol europea. Para ello se utiliza el dataset de las mejores 5 ligas de futbol en Europa de 2014 a 2020 [1]. En primer lugar, se contabilizan los goles totales en todas las ligas, tanto de local como visitante, Tabla 1, a. Luego, se selecciona la liga con mayor número de goles (suma de los goles de visitante y local), dando como resultado la liga Serie A. Luego. de esta liga se contabilizan los goles totales de todos los jugadores y se seleccionan los 5 jugadores con mayor número, Tabla 1, b.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | leagueID | name | awayGoals\_l | homeGoals\_l | sumGoals | | <int> | <chr> | <int> | <int> | <int> | | 2 | Serie A | 3363 | 4113 | 7476 | | 1 | Premier League | 3211 | 4002 | 7213 | | 4 | La Liga | 3035 | 4037 | 7072 | | 5 | Ligue 1 | 2898 | 3758 | 6656 | | 3 | Bundesliga | 2793 | 3531 | 6324 |   **a.** | |  |  |  | | --- | --- | --- | | playerID | name | total\_goals | | <int> | <chr> | <int> | | 1209 | Ciro Immobile | 128 | | 1293 | Gonzalo Higuaín | 108 | | 1230 | Fabio Quagliarella | 102 | | 1513 | Mauro Icardi | 102 | | 1186 | Andrea Belotti | 97 |   **b.** |

**Tabla 1.** **a.** Goles totales en las 5 mejores ligas europeas. **b.** Los 5 jugadores más goleadores de la liga Serie A.

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 y, a su vez, cuál fue el resultado de esas asistencias y tiros al arco. 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 Tabla 2, debajo.

IMAGEN

Figura 1.

| total | player\_name | rojas | amarillas |
| --- | --- | --- | --- |
| <dbl> | <chr> | <int> | <int> |
| 22.00 | Álvaro González | 4 | 72 |
| 22.00 | Dani García | 2 | 80 |
| 21.75 | Damián Suárez | 4 | 71 |
| 21.25 | Granit Xhaka | 7 | 57 |
| 20.50 | Yannick Cahuzac | 9 | 46 |

**Tabla 2.** Total de tarjetas rojas, amarillas y la suma de estas dos, teniendo en cuenta cada 4 amarillas 1 roja.

**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>