

Guía 5

Introducción al razonamiento con datos

Vamos a trabajar con un conjunto de datos que recopila estadísticas de jugadores de fútbol de las 5 principales ligas europeas, que incluye información resumida por jugador de todos los partidos de las temporadas comprendidas entre 2014 y 2021. El dataset `players` abarca información resumida por jugador para todos los partidos disputados entre las temporadas 2014 y 2021, considerando el período en el que más tiempo ha competido el jugador en una determinada liga. Por otro lado, el dataset `shots` contiene las estadísticas detalladas de todos los tiros realizados en cada partido.

Las variables de los datasets son

`playersmatches`

- `player`: nombre del jugador
- `goals`: número de goles
- `shots`: número de tiros al arco
- `assists`: número de asistencias
- `keyPasses`: número de pases clave
- `yellowCard`: número de tarjetas amarillas
- `redCard`: número de tarjetas rojas
- `ligue`: nombre de la liga
- `position`: posición del jugador (se considera la posición en la que más ha jugado en todos los partidos)
- `games`: número de partidos
- `position_general`: agrupa la variable `position` en conjuntos más generales

`shots`

- `gameID`: identificador único de cada partido
- `minute`: minuto del tiro
- `lastAction`: última acción antes del tiro
- `shotType`: tipo de tiro
- `shotResult`: resultado del tiro
- `player`: nombre del jugador

1. ¿Cuáles son las unidades de cada dataset?
2. Para ambos conjuntos de datos, examinen cada variable de manera individual. Seleccionen el gráfico más adecuado para resumir la información de cada una de manera efectiva.

Parte 1. Fair play

3. ¿Qué posición (`position_general`) suele acumular más tarjetas amarillas? ¿Qué tanto peor es que la que posición que le sigue en cantidad de tarjetas amarillas?
4. Al abordar el punto anterior, ¿consideraron el número de partidos jugados por cada jugador? Ignorar este aspecto implicaría que una tarjeta amarilla recibida por un jugador que participó en un solo partido tenga el mismo peso que una recibida por un jugador que disputó 300 partidos. ¿Eso es razonable? Repitan el punto anterior pero ahora incorporen esta información para obtener un resultado más representativo.
5. ¿Cuáles son los 10 jugadores con más tarjetas amarillas? ¿Qué posición tienen en la cancha?
6. ¿Cuáles son los 10 jugadores con mayor proporción de tarjetas amarillas en relación al total de partidos disputados?
7. Entre todas las ligas, ¿cuál presenta la mayor proporción de tarjetas amarillas? ¿Y la de tarjetas rojas?
8. Dado que el punto anterior no considera la cantidad total de partidos jugados en cada liga, el resultado podría no reflejar con precisión cuál liga es la más "fair play". Entonces, ¿cómo podríamos ajustar el enfoque para responder a la pregunta: "*¿Cuál es la liga con el mejor fair play?*"?

Parte 2. ¿Gol o asistencia?

9. Como es de esperar, los delanteros son los jugadores que más goles convirtieron (verifiquen este dato). Sin embargo, ¿son ellos quienes tienen la mayor tasa de conversión de goles en relación con la cantidad de tiros al arco que realizan? Comparen la efectividad de los delanteros con la de la segunda posición más destacada en términos de goles anotados.
10. ¿Existe algún delantero que haya anotado menos goles que asistencias?
Respondan a esta pregunta utilizando un gráfico de dispersión (scatter plot) y sumando la recta $y = x$ con el siguiente código

```
geom_abline(slope = 1, intercept = 0, color = "black", size = 1)
```
11. ¿En qué momento del partido solía convertir Messi los goles? Dar una respuesta gráfica y otra escrita.
12. Los tiros al arco de Messi, ¿en qué proporción son realizados con el pie derecho? ¿Y los goles?

13. Queremos seleccionar a un delantero entre Neymar, Lionel Messi y Luis Suárez, basándonos en quién tiene la mayor proporción de marcación de goles en los tramos finales del partido, sin importar la cantidad de goles totales que tengan. ¿Con qué jugador deberíamos quedarnos?

Parte 3. Arranca en el banco

14. ¿Cuál es la posición con mayor proporción de suplentes?
15. Los suplentes, ¿convierten goles? ¿Cuántos?
16. ¿Cuántos partidos menos juegan los suplentes en comparación con el jugador que ha disputado la mayor cantidad de partidos en la misma liga?

Parte 4. Conferencia de prensa

Realizar los gráficos o calcular las métricas que permitan determinar si estas premisas son verdaderas o falsas:

- Los centrocampistas, en general, convierten más pases clave que goles.
- El pico de la distribución del minuto en el que Kylian Mbappé-Lottin convierte goles se encuentra aproximadamente a los 56 minutos.
- Cristiano Ronaldo suele marcar goles antes que Lionel Messi en los partidos.
- Los centrocampistas defensivos son más determinantes que los defensores centrales, en términos de pases clave por partido.

Bibliografía sugerida

<https://graphicsprinciples.github.io/cheatsheet.html>

<https://www.data-to-viz.com/caveats.html>

<https://www.atlassian.com/data/charts>