

# INTRO. A LA CIENCIA DE DATOS

## Entrega 5 Bases de datos relacionales

Julián Fraga  
fragajulian96@gmail.com

Septiembre 2022

## 1. Hay equipo - Ejercicios 1 a 5

En esta primera sección muestro el análisis de los goles y tiros al arco de los equipos de todas las ligas europeas. Hice dos gráficos de barras, uno para los goles por equipo rankeados de mayor a menor (*fig. 1*) y otro para los tiros al arco rival (*fig. 2*). En primera instancia se puede ver que los equipos con más goles no son necesariamente los que más tiros al arco tienen y esto quedará evidenciado más adelante.

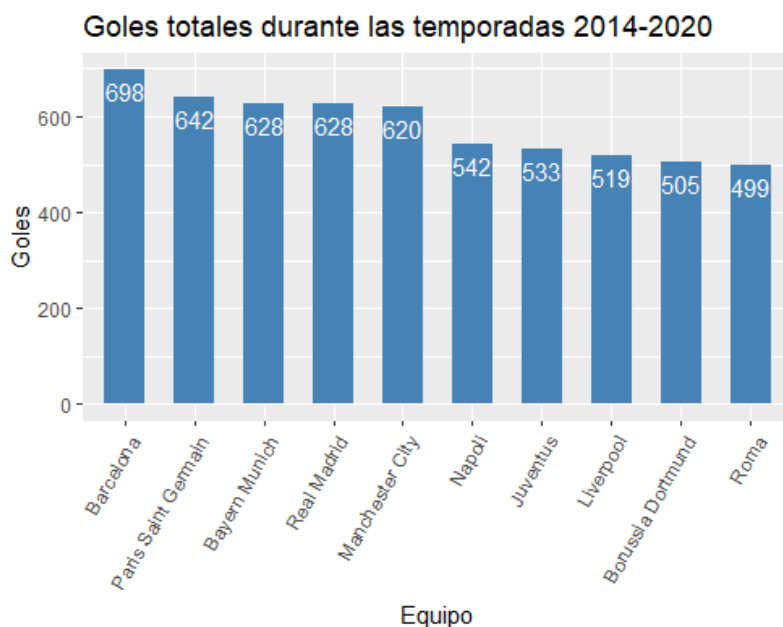


Figura 1: Suma de los goles convertidos por los diez equipos con más goles en las temporadas 2014 a 2020

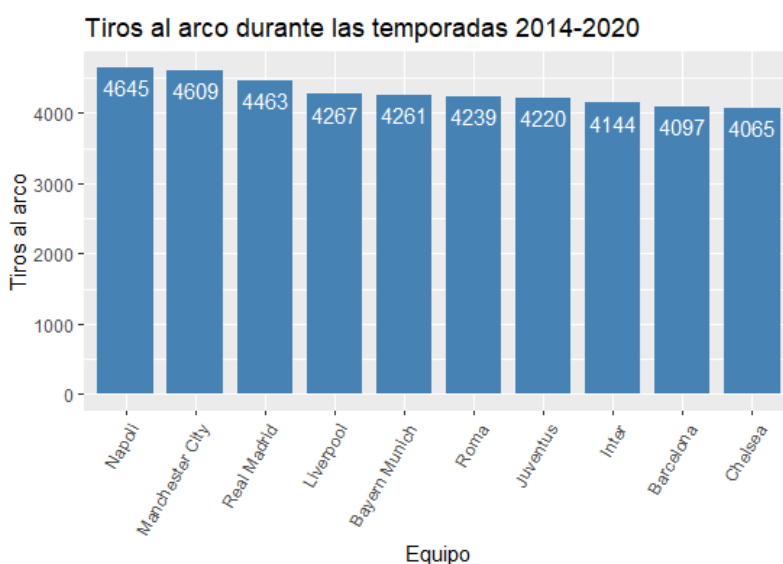


Figura 2: Suma de los tiros al arco por los diez equipos con más tiros en las temporadas 2014 a 2020

En la *figura 3* no se ve ninguna relación evidente entre goles y tiros al arco<sup>1</sup>. Sin embargo, al hacer *zoom out* y ver la totalidad de los equipos europeos se ve una relación bastante lineal (*fig. 4*).

<sup>1</sup>Una pregunta súper al margen que dejo abierta para análisis futuro (si es antes del Mundial capaz gano el Prode) es si existe una relación entre performance general de un dado equipo y el cociente *goles/tiros*, tal vez relacionarlo con la posesión de la pelota o los tiros libres bajo la hipótesis de que un equipo que juega a la contra o con jugadas de pelota parada tal vez tenga menos oportunidades de gol pero con mejores resultados por oportunidad.

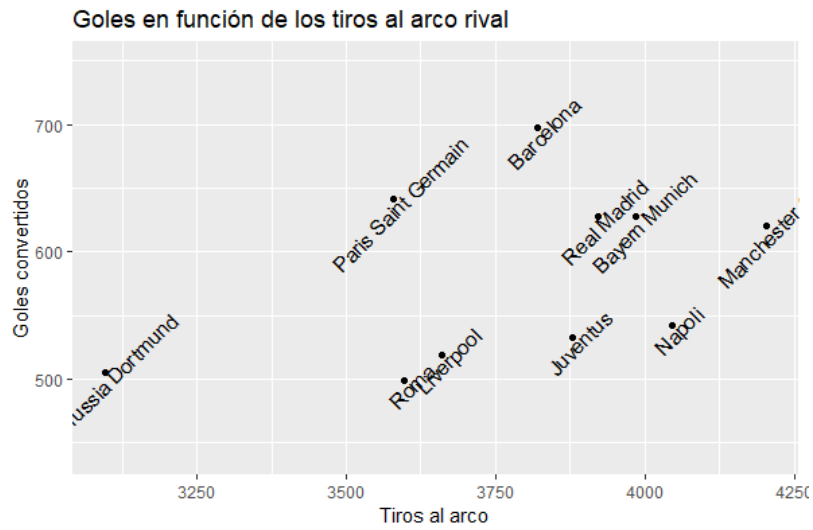


Figura 3: Gráfico de dispersión entre goles convertidos y tiros al arco de los diez equipos goleadores

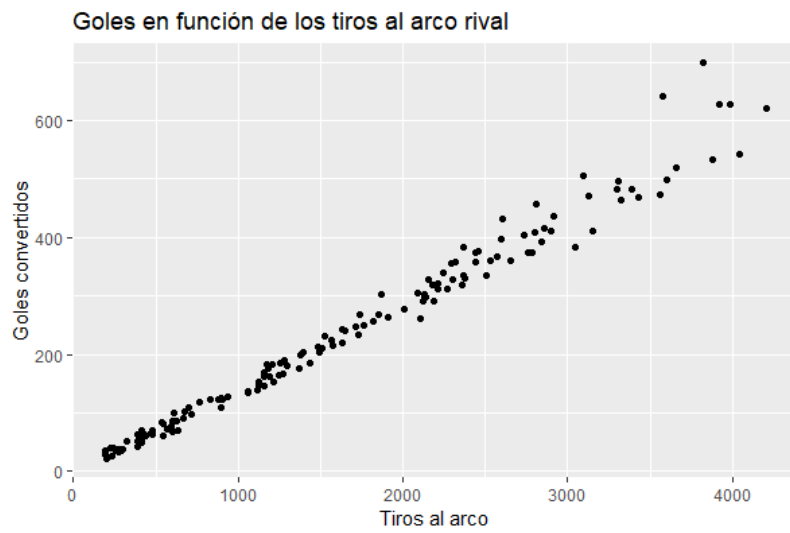


Figura 4: Gráfico de dispersión entre goles convertidos y tiros al arco de todos los equipos europeos. Se aprecia una tendencia lineal

## 2. Notti magiche

Para el análisis de jugadores elegí la Serie A de Italia. Fue más interesante para mí el proceso de cruzar datos que los resultados obtenidos por cómo hubo que ingeniárselas con los *joins* para obtener el *dataframe* adecuado. Del gráfico 5 no tengo mucho por comentar, sólo que se ve alguna diferencia entre, por ejemplo, Higuaín e Icardi en el sentido de que los goles de Higuaín parecieran estar ubicados mayormente en la mitad de cada tiempo, mientras que Icardi tiene una distribución más uniforme, tirando para la contribución sobre el final de los partidos.

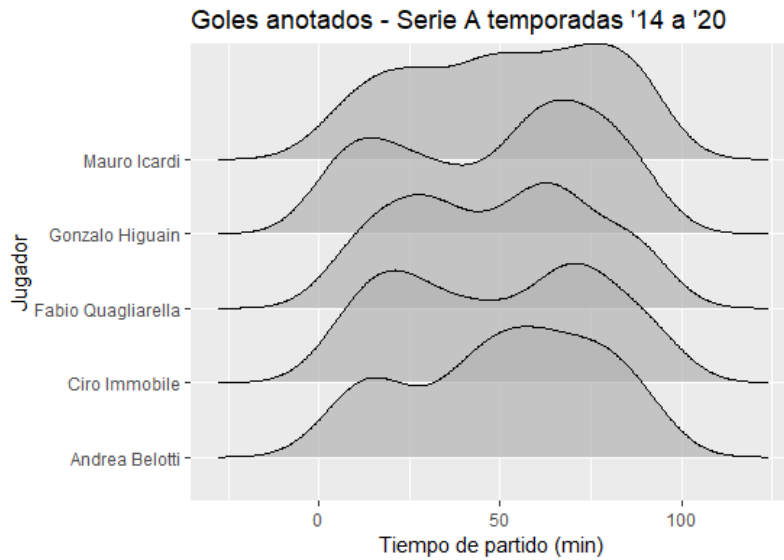


Figura 5: Gráfico de densidad estimada a partir de la distribución de goles por minuto de cada jugador. Los jugadores listados son el top 5 goleadores de la liga