

MÁSTER EN BIG DATA DEPORTIVO

TRABAJO FINAL

MODULO 12

ESTADÍSTICAS LIGAS TOP 5 EUROPA TEMPORADA 2017-2018

GONZALO VARELA

La importancia de los datos.....	2
Estadísticas Temporada 2017-2018 Ligas Top 5 Europa.....	3
Procesamiento de datos y visualizaciones.....	4
Goles	4
¿Cuándo se convierten más goles?	5
¿Quiénes meten los goles?	6
Goles de pelota parada	7
Penales.....	9
Pases.....	11
Zona de pases.....	12
Matriz de pases	14
Dirección de pases	16
.....	17
Disparos.....	17
Expected Goals: xG vs xGa.....	20
Expected Goals: expectativa vs realidad	21
Expected Goals: xG acumulado por partido	22
Faltas.....	23
Tarjetas amarillas y rojas.....	24
¿Cuándo se amonesta?	25
¿Por qué se amonesta?	26
¿Dónde ocurren las faltas?	27
Duelos.....	27
Duelos ofensivos	28
Duelos defensivos	30
Duelos aéreos.....	30
Ejemplo Análisis en conjunto: Barcelona vs Real Madrid	32

La importancia de los datos

El análisis del fútbol a través de los datos pisa cada más fuerte. Cuando parecía que era algo meramente relacionado a deportes como el beisbol y el básquet, poco a poco ya es más habitual escuchar a periodistas y entrenadores mencionando términos más sofisticados como *expected goals*, algo que diez años atrás era totalmente impensado. Todavía hay muchos reticentes a utilizar los datos como una herramienta fundamental a la hora de analizar este deporte, sin embargo, hay otros pioneros que parecen llevar con orgullo esa impronta de *científicos del fútbol*.

Estamos viviendo en una época donde la información abunda y lo difícil es administrarla de manera adecuada. Cualquiera parece saber de todo y está al mismo nivel el comentario de una cuenta anónima en Twitter que la del entrenador del Liverpool. Las *fake news* son moneda corriente y nadie parece tener la necesidad de chequear la información. Por eso los datos son tan importantes. Son complementos fundamentales cuando debés fundamentar una postura y sin ellos sería imposible discernir de quién tiene razón.

Pongamos un ejemplo en el mundo del fútbol. ¿Cómo saber si conviene jugar de contraataque o intentar dominar el partido con la pelota? ¿Hacemos marcaje mixto o individual? ¿Los corners los tiramos al segundo palo o jugamos en corto? ¿Nos conviene atacar por derecha o por izquierda? ¿Quiénes son los jugadores rivales que más participan? ¿Hay alguna tendencia a que la pelota siempre pase por el mismo jugador? ¿Cómo medimos las chances claras que tiene un equipo? ¿Es suficiente hablar de posesión del balón o debemos tener en cuenta en qué parte del campo? ¿Es igual de relevante tomar la posesión cuando vas ganando a que cuando vas perdiendo? Preguntas hay demasiadas, el tema radica en si se pueden resolver y vamos a tratar de responderlas en lo que resta del documento, a través de los datos.

Estadísticas Temporada 2017-2018 Ligas Top 5 Europa

Para ello vamos a acudir una base de datos abierta utilizada por diversos especialistas del tema en un paper de la Revista Nature. [A public data set of spatio-temporal match events in soccer competitions \(nature.com\)](https://www.nature.com/datasets), cuyos autores son Luca Pappalardo , Paolo Cintia, Alessio Rossi, Emanuele Massucco, Paolo Ferragina, Dino Pedreschi y FoscaGiannotti.

Ésta cuenta con todos los eventos espacio-temporales que ocurrieron en los partidos jugados de las mejores competiciones de fútbol a nivel mundial. Más específicamente, se trata de los torneos disputados en la temporada 2017-2018 de la Primera División Inglesa, la Primera División Española, la Primera División Italiana, la Primera División Francesa y la Primera División Alemana. También contiene datos sobre la Copa Mundial de Fútbol 2018 y la Eurocopa 2016, sin embargo, opté en este caso por dejarlas afuera del análisis, ya que en mi percepción estas competiciones generalmente al ser tan cortas pueden llegar a mostrar análisis sesgados.

Los datos descriptos en el *paper* fueron obtenidos por Wyscout, una compañía líder en la industria del fútbol que conecta a los profesionales del fútbol a lo largo del mundo y trabaja con más de 50 asociaciones de fútbol y más de 1000 clubes a nivel mundial. En caso de que quieran interesarse de cómo se obtienen estos datos pueden o bien leer el *paper* o ingresar en la página oficial de Wyscout. A continuación, mostraré una tabla con las competiciones elegidas para analizar la cantidad de partidos, eventos y jugadores correspondiente a cada liga.

Competición	Cantidad de Partidos	Cantidad de eventos	Cantidad de jugadores
Liga Española	380	628.659	619
Liga Inglesa	380	643.150	603
Liga Italiana	380	647.372	686
Liga Alemana	306	519.407	537
Liga Francesa	380	632.807	629
Total	1826	3.071.395	3074

Tabla 1. Contiene información acerca de las distintas cantidades de partidos, eventos y jugadores que existe en la base de datos

Son datos más que suficientes para marcar ciertas tendencias que ocurren en el fútbol. A su vez, es de mi interés mostrar las diferencias fundamentales que existen entre las principales ligas europeas, como así también en los diferentes equipos involucrados. También se recopilarán, aunque en menor medida, datos individuales de los jugadores. No es mi objetivo en este Trabajo realizar un informe de Scouting acerca de jugadores potenciales de compra o revelaciones, sino que mi idea es ilustrar las características sobresalientes a nivel equipo y a escala nacional de clubes.

Procesamiento de datos y visualizaciones

Para limpieza y procesamiento de los datos, utilicé Python, mientras que para realizar las visualizaciones opté por Tableau en mayor medida y PowerBI, en menor.

Goles

Empezaremos con algo sencillo, pero no por eso despreciable. Los goles son la variable más importante, aunque tal vez la menos sofisticada. Sin embargo, no hay que subvalorarla por ello. Al fin y al cabo, los partidos se definen y se definirán por goles. En primer lugar, vamos a observar si existe una tendencia en la cantidad de goles por liga.

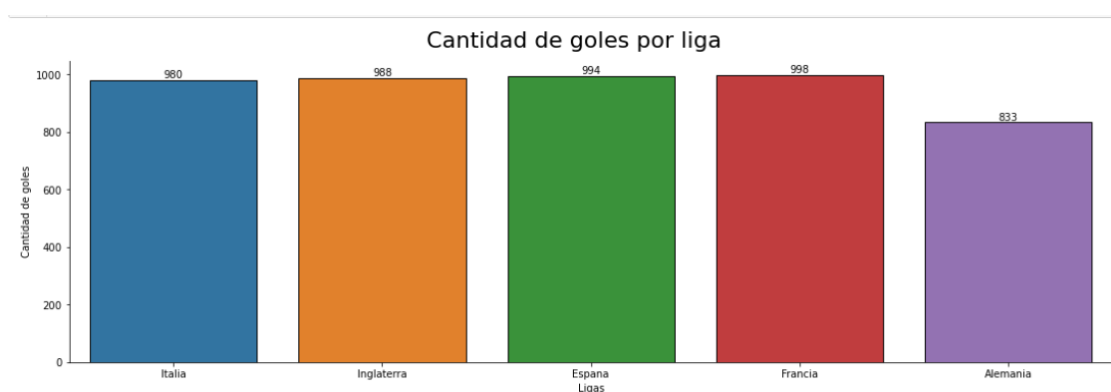


Figura 1. Muestra los goles que se convirtieron en las 5 ligas top de Europa

La **Liga Francesa** con 998 goles es la competición en la que más tantos se han marcado. Sin embargo, las demás ligas no se encuentran muy atrás, salvo la Alemana. Igualmente, hay que tener en cuenta que la Bundesliga cuenta con 74 partidos menos que las demás Ligas. Por lo tanto, este gráfico nos da una aproximación, pero no es definitivo, ya que está a distinta escala. Pero eso no es un problema, podemos solucionarlo si dividimos la cantidad de goles sobre la cantidad de partidos y de este modo obtendremos el promedio de goles por partido en cada liga.

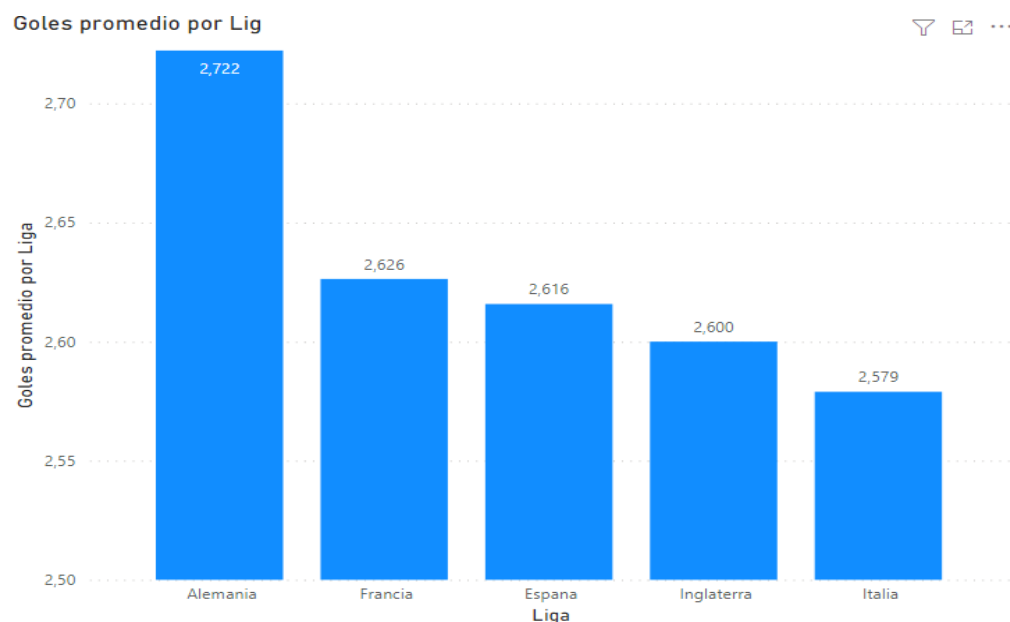


Figura 2. Indica los goles promedios que se marcaron en las cinco ligas top de Europa por partido

Nos damos cuenta de que en términos relativos es la **Liga Alemana** la que más goles convierte, con un promedio de 2,72 goles por partido, mientras que la **Liga de Italia** pasa a ser la que menos tantos se marcan por encuentro.

¿Cuándo se convierten más goles?

Por otro lado, también podemos saber en qué momento del partido se convierten los goles en cada liga. Y para eso lo que hice fue separar los 90 minutos en franjas de 15 y a su vez también aparté los goles que se convirtieron en tiempo de descuento. En este caso, por lo anterior mencionado, utilicé porcentajes y no cantidades, de esta manera, la Liga Alemana está en igual de condiciones para ser analizada. ¿Podemos sacar alguna conclusión observando los gráficos?

En primer lugar, algo que es bastante obvio: en el segundo tiempo se marcan más goles que en el primero. Alrededor de entre 40 y 42% en la primera parte y el restante en la complementaria. Esta tendencia se mantiene en todas las ligas por igual. En cuanto a los picos de cada gráfico, notamos que la **Liga Francesa** no tiene uno notorio, si no que se reparten de manera equitativa los goles, igualmente la franja de 75-90 predomina con un casi 17%. La **Liga Alemana**, a diferencia del resto, tiene su esplendor entre los 15-30 del primer tiempo. Si queremos observar goles en la **Liga Española** miremos directamente el final del partido.; si contamos los que se convierten en el adicionado del complemento, tenemos que 1 de cada 4 goles ocurre en los últimos 15 minutos. Mientras que por el lado de la **Liga Inglesa** y la **Liga Italiana** el apogeo ocurre entre los 60-75 minutos empezado el encuentro.

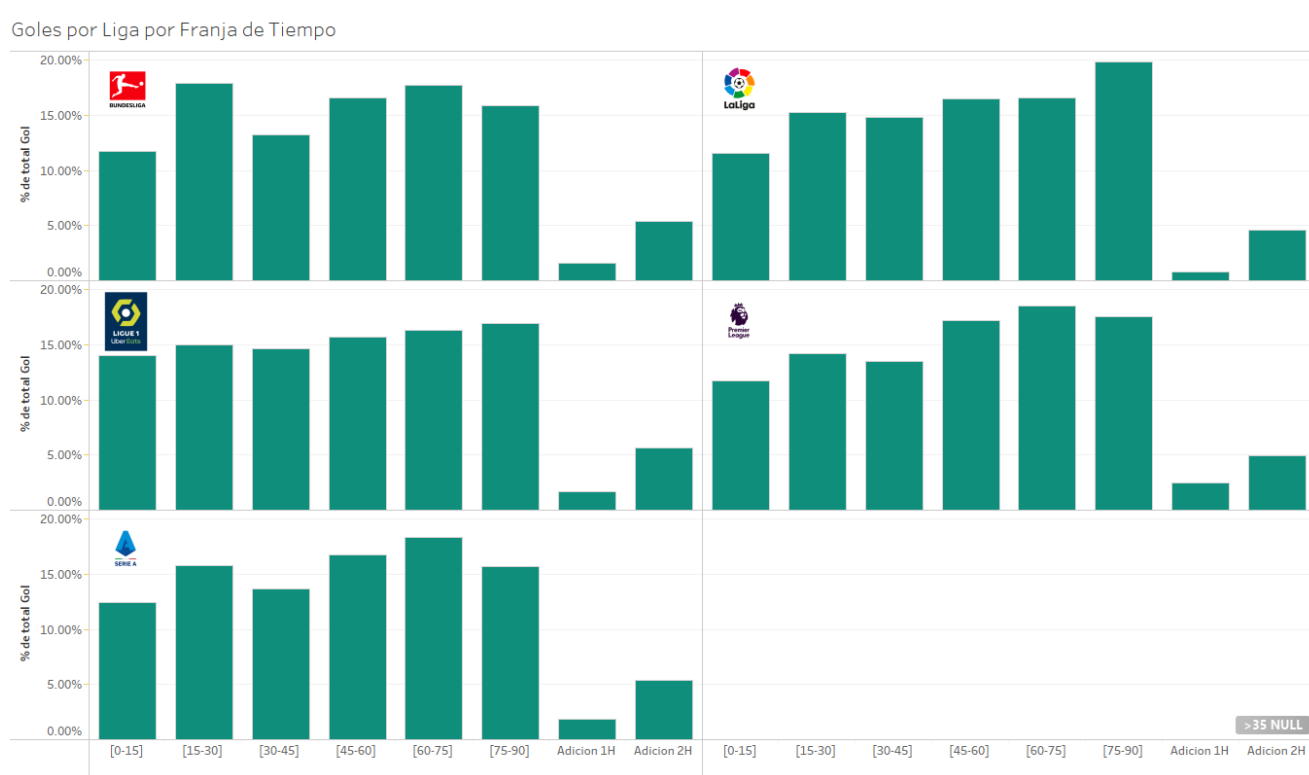


Figura 3. Cada barra señala el porcentaje de goles que se marcó en cada liga según la franja horaria indicada

Como hicimos con las Ligas, también lo podemos hacer con los equipos. En este caso, como vamos a analizar cada ligar por separado, opté por compararlos por las cantidades de goles por franja en vez de hacerlo por los porcentajes sobre el total.

Goles_Franja_Equipo

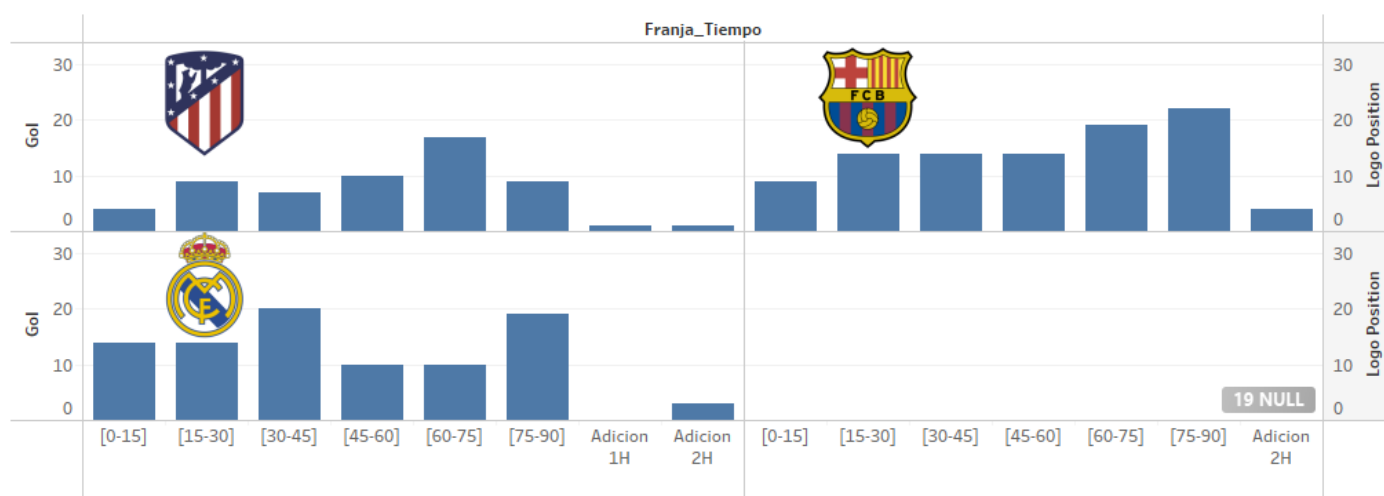


Figura 4. Cada barra señala el porcentaje de goles que se marcó en cada equipo según la franja horaria indicada

Analizamos a los tres equipos que terminaron en el podio de La Liga Española. Mientras que el Real Madrid marca con más frecuencia en los últimos 15 minutos del Primer Tiempo, el Colchonero lo hace entre los 60 y 75 minutos y el Barcelona entre los 75 y 90.

Otro dato para tener en cuenta sería en qué franja les convierten más a estos equipos. Dependiendo el resultado, podríamos sacar conclusiones como “entraron dormidos” o “se relajan al final”.

¿Quiénes meten los goles?

“El delantero vive del gol”, dicen seguido. Sin embargo, para que a un equipo le vaya bien debe sumar una buena cantidad de goles en las distintas líneas del campo. En la **Figura 5** observamos que en **España** 6 de cada 10 goles son de delanteros, el mayor porcentaje entre todas las ligas. **Alemania** tiene con un 13% a los defensores que en promedio del total convierten con más asiduidad. No hay nada mejor que los mediocampistas que llegan al gol o los denominados *box to box*, que como no podía ser de otra manera, se encuentran principalmente en la **Premier League**. De **Francia** podemos mencionar que es la liga en la cual los defensores tienen la menor cuota goleadora, mientras que la **Serie A** es la competición en la cual los goles están más repartidos según las posiciones.

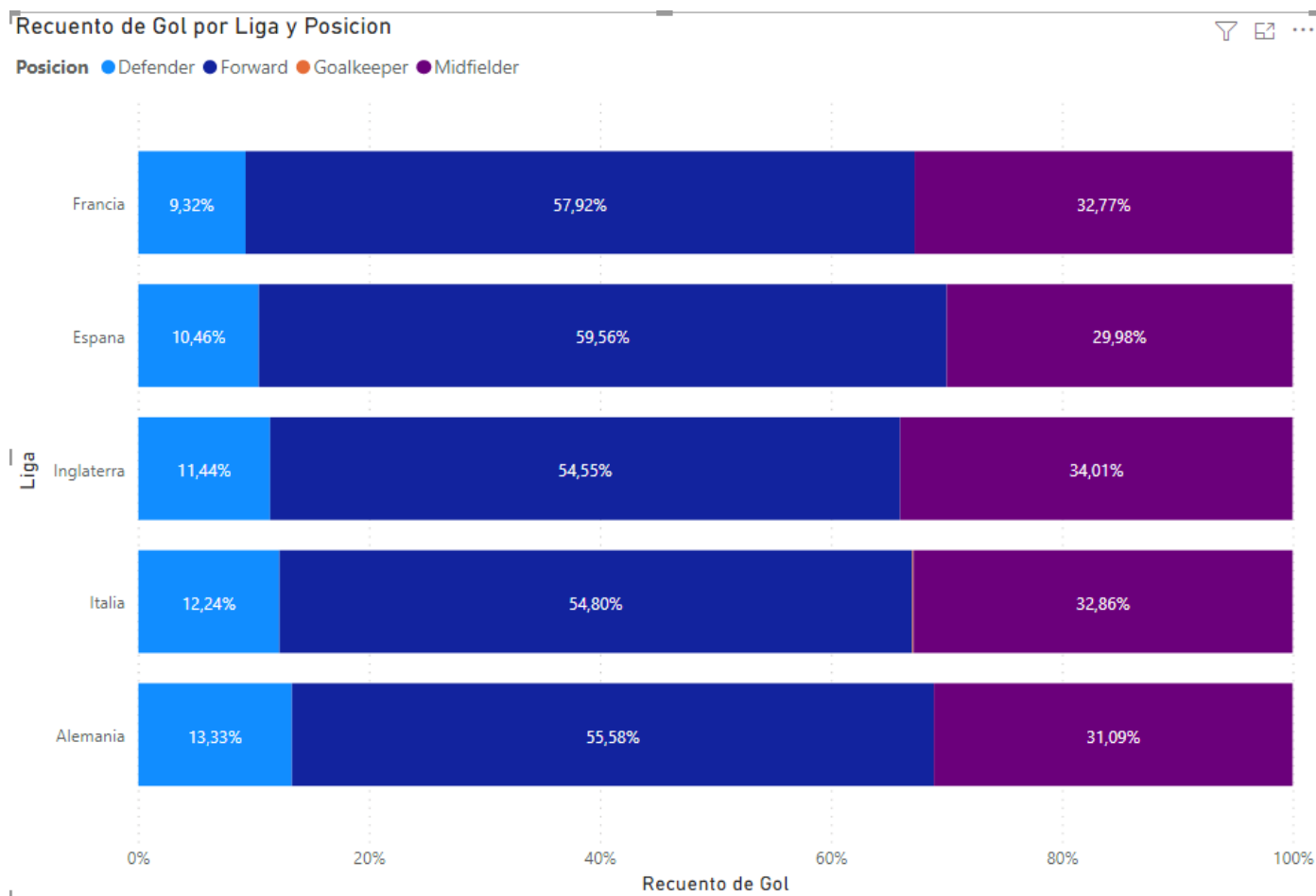


Figura 5. Muestra cuál es el porcentaje de los goles en cada liga según la posición que los jugadores ocupan en el campo de juego

Goles de pelota parada

En el Mundial 2018, el 43 % de los tantos vinieron por medio de la pelota parada, 73 de 169. 26 de esos goles se realizaron tras un tiro de esquina, 23 fueron de penal, 16 tras tiros libre, 6 por la ejecución de una falta y 2 se concretaron tras un lateral. La pelota parada es cada vez más importante y ya hay equipos que tienen entrenadores específicamente para ello. Es más, el Brentford tiene una persona a cargo de los laterales de los equipos. Sí. Leyó bien, cada vez que el equipo inglés tiene un lateral esto está a cargo de Thomas Grønnemark, un danés que revolucionó el fútbol y que según la BBC es el cargo más extraño en la historia del fútbol que tuvo alguien alguna vez.

[The Art of the Throw-In – An Interview with ‘The Throw-in Coach’ Thomas Grønnemark – Breaking The Lines](#)

Sin embargo, es difícil establecer con los datos qué goles vinieron mediante la pelota parada y cuáles no. En primer lugar, quiero aclarar que no tuve en cuenta los goles de pelota parada directo, ya sea penales, tiros libres, córners olímpicos, ya que en mi opinión eso tiene más que ver con la calidad del ejecutante que de un trabajo en equipo. Lo que quiero destacar en este apartado son los goles que provinieron de una pelota parada, pero terminaron con un disparo, ya sea de cabeza o con el pie. Lo que hice en este caso fue tomar los tiros libres que fueron centros en campo contrario, es decir, con la variable “X inicial” mayor a 50, y los córners. Generalmente, la siguiente acción a una pelota parada es un duelo. Cuando hay duelos siempre hay dos acciones: uno que lo gana y otro que la pierde.

Entonces, la secuencia sería así. Evento 1: pelota parada, Evento 2: duelo a; Evento 3; duelo b; Evento 4: gol. Esta es tan solo una de las posibilidades que puede suceder. Luego, de estudiarlo detenidamente establecí que iba a tomar a los eventos que sean disparos y hayan terminado en gol desde 1 a 5 eventos posteriores al córner o pelota parada. Claramente esta modalidad tendrá sus fallas, pero en general creo que funciona de la manera deseada.

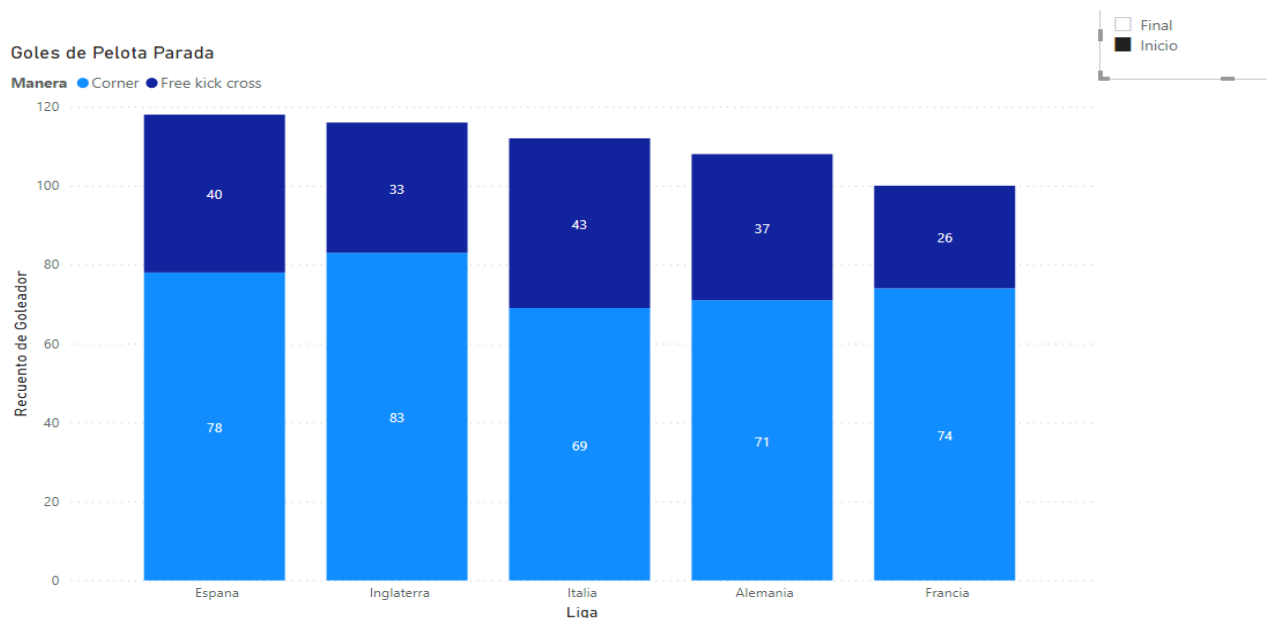


Figura 6. Indica la cantidad de goles a través de la vía del corner y de un centro tras un tiro libre en cada liga analizada

Tenemos a **España** como la Liga en la que más goles se convierten mediante la pelota parada como “asistencia”, con 78 goles mediante los córneres y 40 por la vía del centro. Lo de Alemania es más que destacable, ya que recordemos que tiene 76 partidos menos, por lo tanto, sería la liga en promedio por partido con más goles por estas vías. Lo cual tiene sentido si lo relacionamos con **Figura 5** que nos decía que la Bundesliga era la competición en la cual los defensores más marcaban en promedio.

Si observamos los equipos que marcaron más tantos luego de un córner o un tiro libre como centro, filtro los que hayan marcado al menos ocho goles por esta vía y calculo cuál es el porcentaje sobre el total de los tantos convertidos por el equipo me quedan:

Goles Pelota Parada y Total de Goles por Nombre_Equipo

● Goles Pelota Parada ● Total de Goles

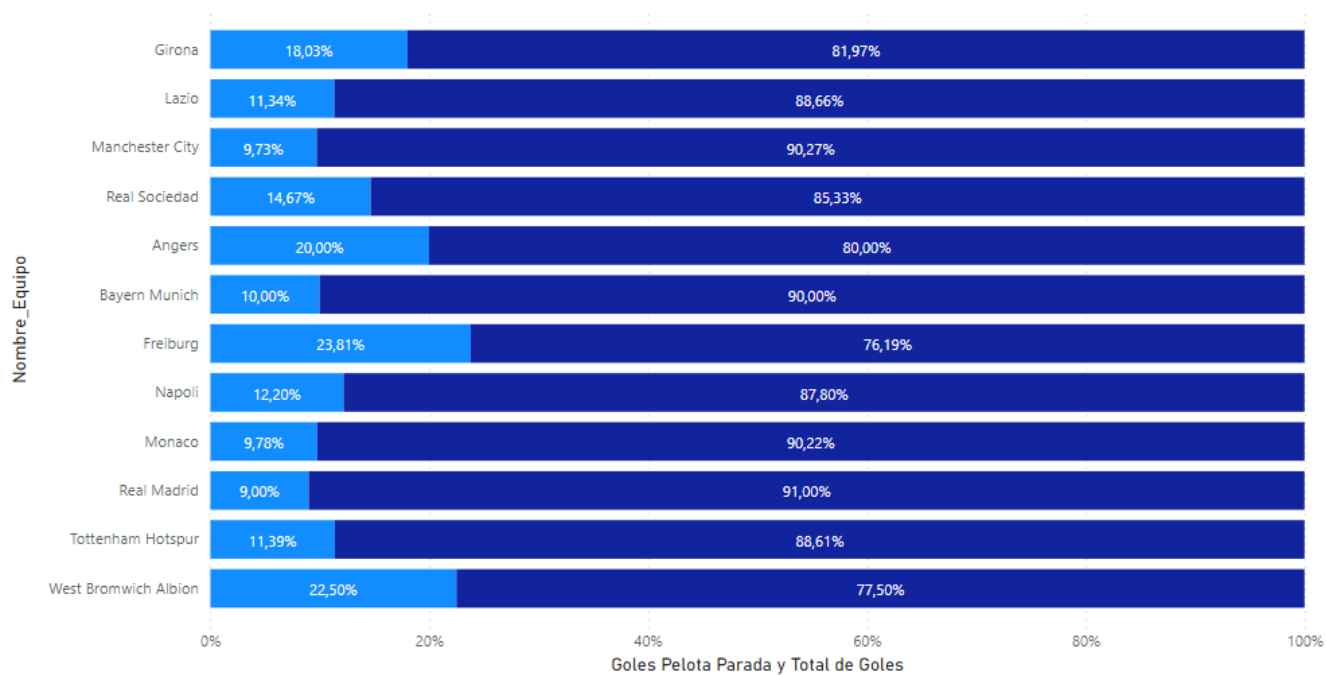


Figura 7. Muestra qué porcentaje representan los goles de pelota parada sobre el total de tantos de los equipos que al menos convirtieron 8 goles mediante esta modalidad

El Friburgo es el equipo que más se beneficia en proporción con los goles de pelota parada con 10 goles sobre 32 anotados. Para chequear nuestros cálculos fueron adecuados, coloqué en Google el siguiente texto: “Freiburg set piece” y en efecto me encontré con este artículo [Freiburg: The Bundesliga set-piece kings continuing World Cup legacy | Bundesliga](#) que reconfirma el predominio del Friburgo por esta vía. Lo mismo realicé con West Bromwich Albion y por suerte conseguí el mismo resultado: [Breaking down West Brom's set-piece routines, the best in the Premier League \(thesefootballtimes.co\)](#)

Penales

Y aquí llega mi punto favorito. Es probable que los tiros desde el punto de penal sean el evento en la cual las estadísticas más sirvan. Se trata de una jugada con pelota parada a 11 metros del portero, en la cual las chances de que se conviertan en gol son casi del 80 %. Hablamos de una acción predilecta para ser analizada. Las variables que hay acerca de esta jugada pueden llegar a ser muchas y algunas como las psicológicas que son casi imposible de cuantificar. Sin embargo, confío en que con los datos provenientes de WyScout se pueden sacar conclusiones contundentes acerca de los penales.

Para comenzar enunciaré y explicaré los distintos filtros que me parecieron prudentes analizar:

- Pierna del ejecutante: hay una creencia popular que establece que los jugadores *cruzan* los penales, es decir, si patea un diestro lo hará a la derecha del arquero, mientras que un zurdo lo hará hacia la izquierda. Veremos si es cierto.
- Localía: al principio hablé de aspectos psicológicos y me pregunto si al ser local un jugador está más tranquilo o siente más presión a la hora de ejecutar un tiro desde los doce pasos.
- Resultado: tal vez el factor más importante en mi parecer. No es lo mismo patear un penal cuando tu equipo está ganando, está perdiendo o está empatando.
- Tiempo de juego: relacionado con el ítem anterior, se supone que hay más presión en patear un penalti al minuto 90 cuando tu equipo va perdiendo que hacerlo cuando tu equipo va ganando.
- Dirección del disparo: “Aseguralo al medio”, habrán escuchado más de una vez. ¿Cuántos penales que fueron al medio terminaron realmente en gol? ¿Existe algún lugar donde pateemos que nos minimice las chances de error?

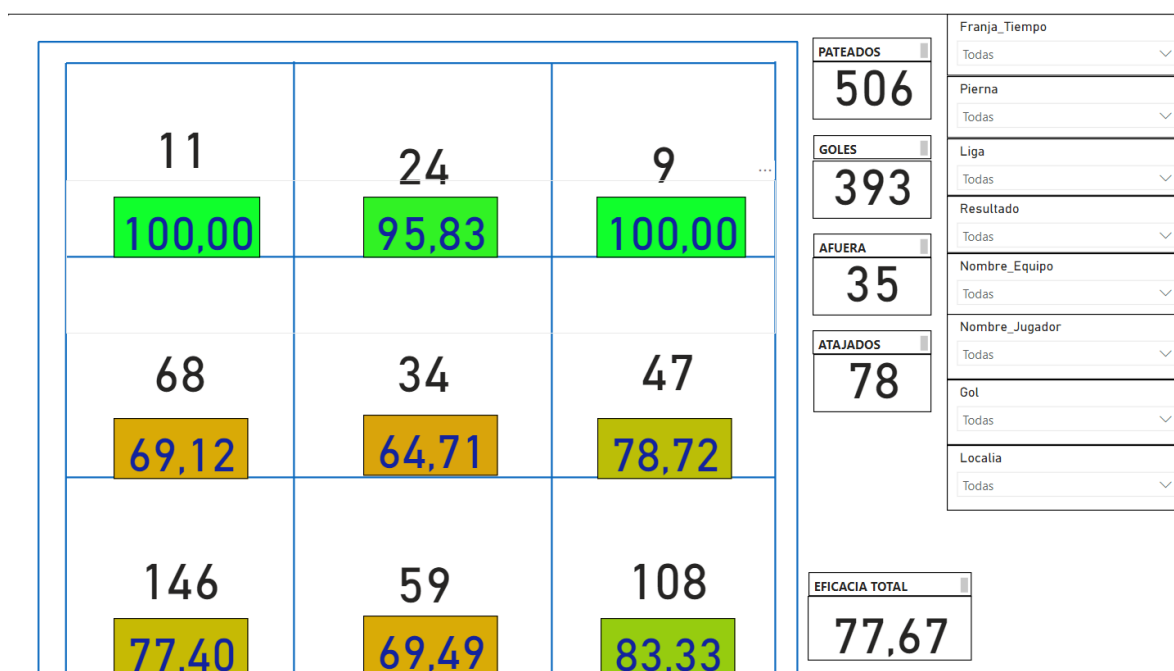


Figura 8. Indica la cantidad donde fue disparado cada penal y en color muestra el porcentaje de eficacia cuando el jugador disparó a ese sector del arco

Fueron analizados 506 penales pateados, de los cuales el 78 % fue convertido, el 7% fue desviado y el 15% fue atajado por el portero.

A continuación, evaluaremos el impacto de las variables elegidas en la eficacia para marcar un penalti.

- El **pie hábil del ejecutante** es un factor importante a la hora de elegir un pateador. El 80 % de los jugadores diestros convirtieron desde los doce pasos, mientras que los zurdos, el 66,67%.

- Los resultados en cuanto a la influencia de la **localía** del equipo del ejecutante pueden llamar la atención, ya que la eficacia para anotar un penal es mayor cuando vas de visita que cuando juegas en casa, de un 79,69 % a un 75,60%, respectivamente.
- Si analizamos el **resultado del encuentro** al momento de lanzar el tiro penal parecen bastante razonables. Cuando tu equipo gana el 83,44 % de los disparos terminan en la red, mientras que si no, las chances disminuyen al 75% en promedio.
- El **mejor momento para patear un penal** es entre los 15 y los 30 minutos de iniciado el encuentro, ya que con un 87,32 % lidera franja de los que más anotaron. El peor momento es minutos después, entre los 30 y los 45, con un 71,23%.
- En la **Figura 8** se observa perfectamente que **si pateas arriba y tu pelota va al arco tus chances aumentan** considerablemente para que termine en grito de gol. No se recomienda disparar a media altura y es más probable que conviertas si tu remate va hacia la izquierda del arquero.
- En cuanto a las Ligas, la **Alemana** con el 80,95 % es la que tiene más alto el porcentaje de eficacia y la **Italiana**, la que menor, con un 76,07 %.

Existen filtros de equipos y jugadores para ir probando cuál es el conjunto con mayor acierto a la hora de los penales y hacia donde patean con más frecuencia. Dos intenciones que me quedaron pendientes para otro trabajo son: cuando se repite un penal, ¿el ejecutante cambia la dirección o mantiene?; y la otra es hacia dónde se arrojan los arqueros con más frecuencia y analizar los penales más desde su perspectiva que la de los jugadores de campo.

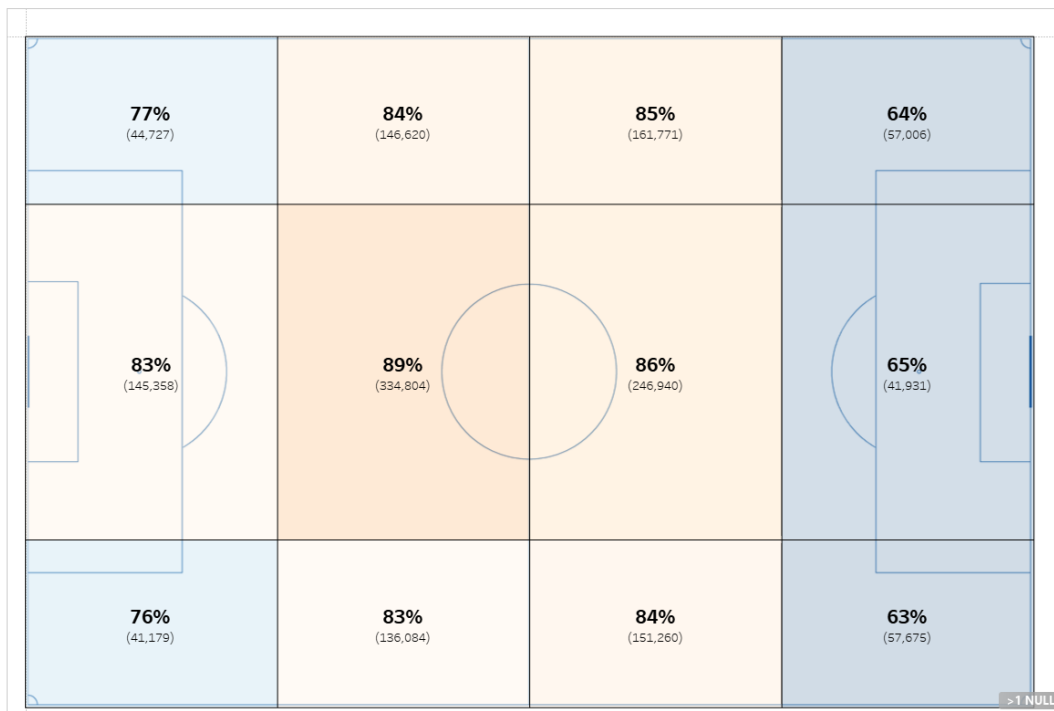
Pases

Los pases son el evento más habitual en un partido de fútbol. En general, el 50 % de los eventos en un encuentro terminan siendo pases. Habrá algunos equipos como el Barcelona y el Manchester City en el que la cantidad será mayor y otros como el Burnley o el Getafe en los cuales será menor. Sin embargo, la cantidad por sí sola no tiene mucho sentido de ser evaluada. Importa más el cómo, el cuándo y el dónde se producen. No es lo mismo tener el 80 % de la posesión, que la mayoría de los pases sean horizontales y en campo propio que tener el 40 % pero en campo contrario y siempre en dirección vertical. Otro elemento para tener en cuenta es la precisión, aunque para ello sí debemos tener en cuenta la cantidad, ya que capaz un jugador que realizó pocos pases y sencillos tiene una efectividad del 100% y sería un dato sacado de contexto que no nos aportaría valor.

Zona de pases

Por ello, el primer gráfico que vamos a mostrar expone un poco lo que venimos hablando. Más precisamente, verán un mapa de un campo de fútbol dividido en 12 zonas, en las cuales se indicarán la cantidad de pases en cada una de ellas y la precisión de estos. El análisis puede ser individual o a nivel grupal (ya sea equipos, ligas, etc). A su vez, podemos establecer filtros como dirección (horizontal, atrás o adelante), localía, rival, liga, tiempo, receptor, entre otras variables a elegir.

%Pases por Zona - Todo | Todo



Localia
(Todo)

Receptor
(Todo)

Franja Tiempo
(Todo)

Liga
(Todo)

Nombre Equipo
(Todo)

Posicion
(Todo)

Nombre Jugador
(Todo)

Rival
(Todo)

Direccion Pase
(Todo)

AGR(Porcentaje Pases Pr...
63% 100%

Figura 9. Muestra el porcentaje de eficacia de los pases en cada zona del campo

Como era de prever la mayor precisión se encuentra en la zona media del campo, mientras que en el último y primer cuarto del terreno la efectividad ya no es tan alta.

Podemos comparar dos equipos totalmente opuestos en su estilo y observar claramente sus diferencias en el mapa.

%Pases por Zona - Manchester City | Todo

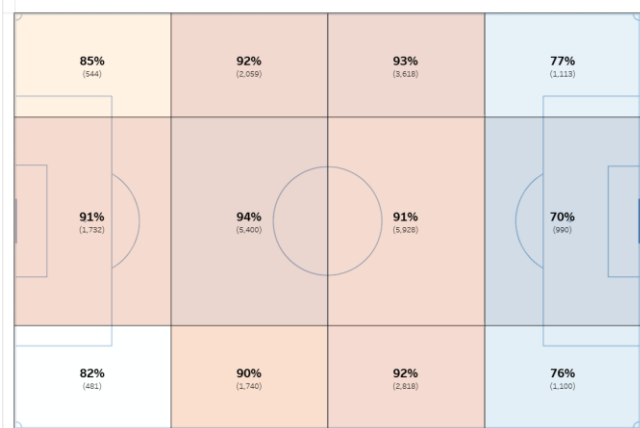


Figura 10. Muestra el porcentaje de eficacia de los pases en cada zona del campo del Manchester City

%Pases por Zona - Getafe | Todo

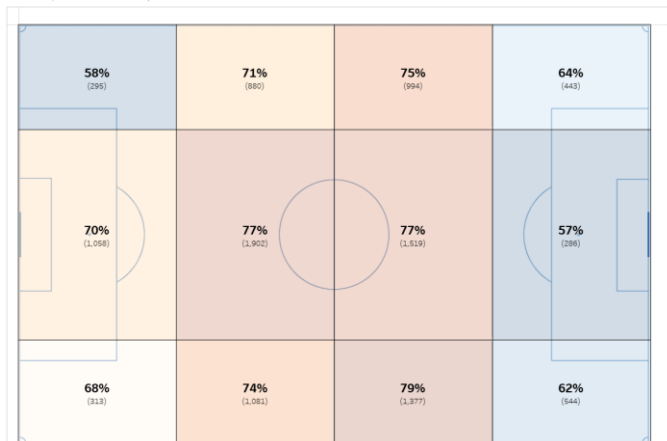


Figura 11. Muestra el porcentaje de eficacia de los pases en cada zona del campo del Getafe

Las diferencias en cantidad y efectividad de pases son sumamente amplias. Prácticamente el menor porcentaje que presenta el City en una de las zonas equivale al mayor del Getafe en alguna zona. Este gráfico nos sirve para darnos cuenta de cómo juega cada equipo sin siquiera haberlos visto alguna vez.

También podemos hacerlo a nivel individual. Por ejemplo, comparar a Messi con Griezmann.

%Pases por Zona - Barcelona | L. Messi

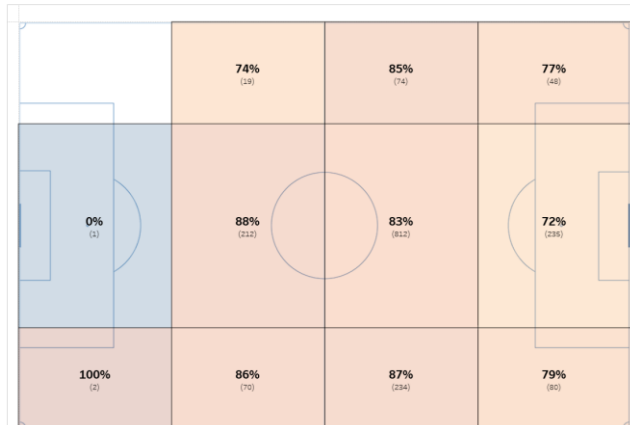


Figura 12. Muestra el porcentaje de eficacia de los pases en cada zona del campo de Messi en Barcelona

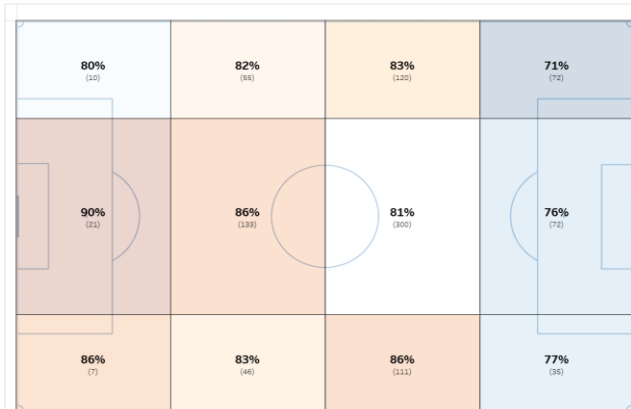


Figura 13. Muestra el porcentaje de eficacia de los pases en cada zona del campo de Griezmann en Atlético Madrid

Otra vez volvemos al estilo de juego de los equipos. Messi obviamente participa más en el armado debido a la intención del conjunto catalán en no rifar el balón y permanecer el mayor tiempo posible con él. Vemos que el argentino se posiciona más que nada en el medio, sin embargo, prefiere jugar más por la banda derecha que por la izquierda. En cambio, el francés no posee una banda preferida, pero a diferencia de Messi y seguramente influenciado por Diego Simeone, su DT en el Atlético, ha participado bastante en la fase defensiva del equipo madrileño.

Matriz de pases

La matriz de pases muestra la conexión entre jugadores de un mismo equipo en un partido. Para ello necesitamos la cantidad de pases entre los distintos jugadores del conjunto y la posición media de cada uno de ellos. Luego, utilizar un mínimo de pases en el cual nosotros creemos que a partir de ese número consideramos que existe una “conexión” entre dos jugadores del mismo equipo. En mi caso, elegí 5 pases como mínimo. Luego, cuanto más cantidad de pases exista entre dos miembros más oscura será la línea de representación de esa conexión. Esta red nos permite conocer en qué zona juega el equipo, quién o quiénes son los jugadores que más participan, qué sociedades se forman dentro del campo de juego, qué estilo de juego predomina, etc.

Matriz de Pases
 Equipo: Burnley
 Partido: Chelsea - Burnley, 2 - 3

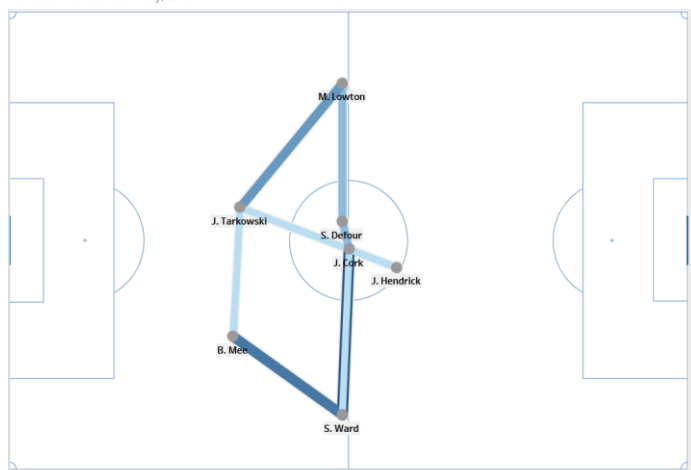


Figura 14. Matriz de pases del Burnley vs Chelsea

Matriz de Pases
 Equipo: Chelsea
 Partido: Chelsea - Burnley, 2 - 3

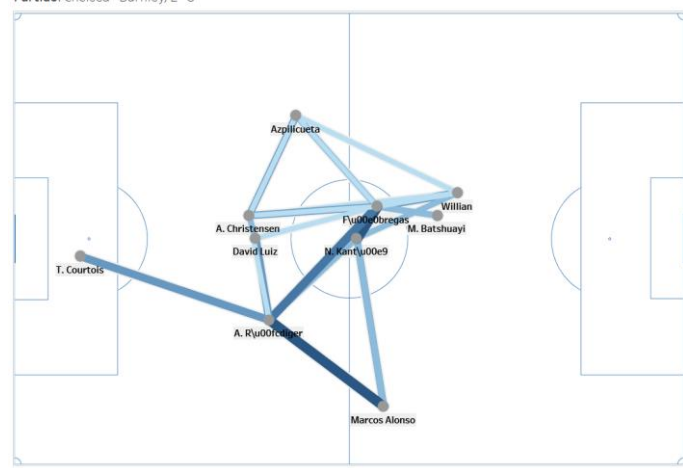


Figura 15. Matriz de pases del Chelsea vs Burnley

Acá tomamos un ejemplo de un partido disputado entre Burney y Chelsea, que terminó en victoria para los conducidos por Sean Dyche. El Burnley prácticamente no tuvo conexiones con sus delanteros, la mayoría de los pases fueron entre sus defensores centrales y los laterales, y las posiciones promedios de los jugadores son bastante defensivas. En cambio, el Chelsea tiene conexiones entre todas las líneas del equipo, se nota que Courtois salió en varias ocasiones jugando con Rudiger y Fábregas es, sin lugar a duda, el conductor del equipo y por donde pasa la pelota con más asiduidad. Con tan solo esta matriz ya somos capaces sin haber visto el encuentro de cómo fue.

Dirección de pases

Otra manera de comprender un partido de fútbol mediante los pases es observar en qué dirección y dónde se realizaron cada uno de ellos. Agregando variables como la longitud (corto o largo), dirección, (atrás, horizontal o adelante), el tipo (centro, pase sencillo, pase del arquero con la mano, etc) y la precisión, entre otras, el análisis se va a enriquecer cada vez más. En este caso, para ejemplificarlo elegí un partido de la Bundesliga entre Bayer Munich y Borussia Dortmund. Para que sea más claro, opté por tomar solo el Primer Tiempo del encuentro, en el que terminó ganando el conjunto de Munich por 4-0.

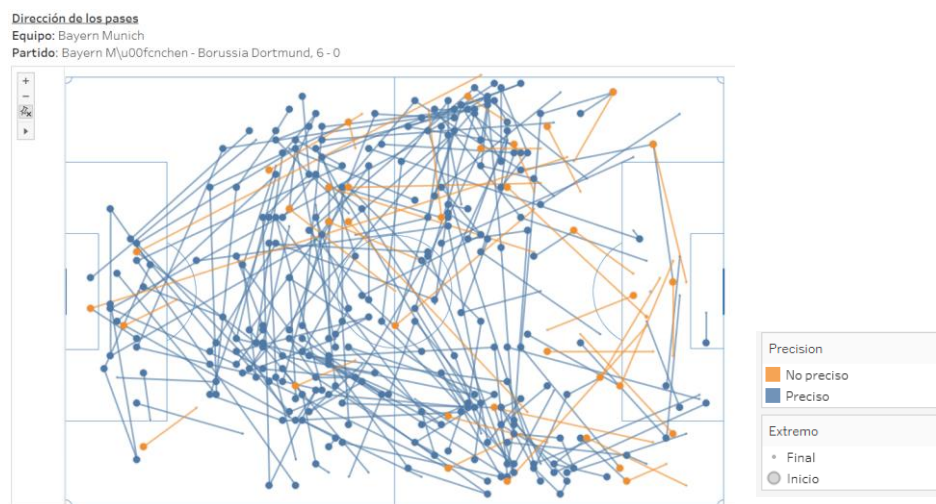


Figura 16. Muestra la dirección de los pases del Bayer Munich vs Borussia Dortmund durante el primer tiempo

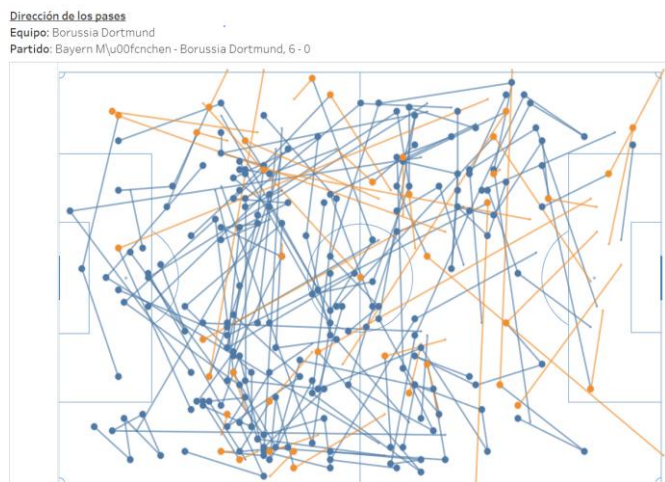


Figura 17. Muestra la dirección de los pases del Borussia Dortmund vs Bayer Munich durante el primer tiempo

Observamos un dominio del medio del equipo conducido Jupp Heynckes, con pocas equivocaciones en campo propio y arriesgando más en terreno rival, saliendo jugando desde abajo, mientras que el Dortmund con más equivocaciones, prácticamente sin llegar con precisión al arco contrario, sin embargo, intentando no perder su

dominio del balón. No se ve una diferencia de cuatro goles, pero por eso cada gráfico hay que observar en complemento con los demás para su correcto análisis.

Dirección de los pases
Equipo: Bayern Munich
Partido: Bayern München - Borussia Dortmund, 6 - 0
Jugador: F. Ribery

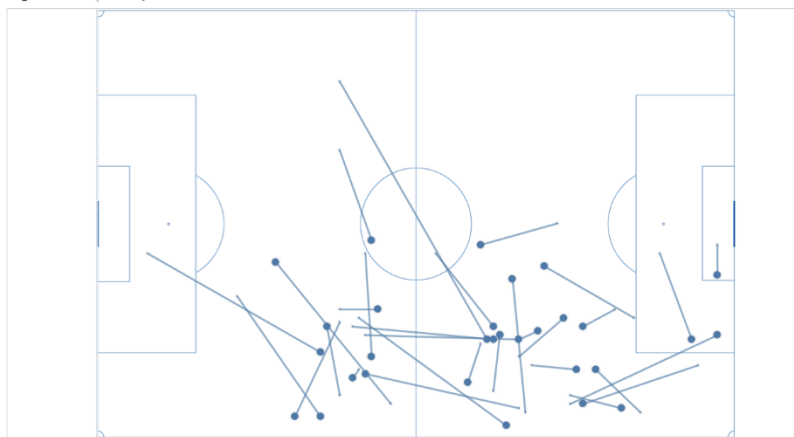


Figura 18. Muestra la dirección de los pases de Frank Ribery vs Borussia Dortmund durante el primer tiempo

Dirección de los pases
Equipo: Borussia Dortmund
Partido: Bayern München - Borussia Dortmund, 6 - 0
Jugador: C. Pulisic

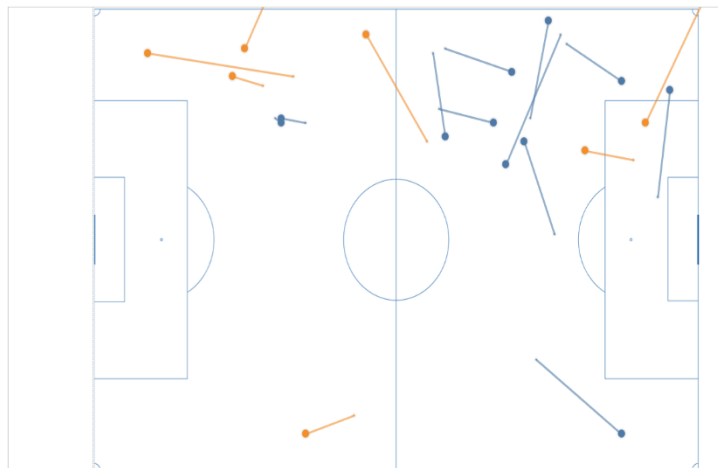


Figura 19. Muestra la dirección de los pases de C. Pulisic vs Bayer Munich durante el primer tiempo

Si analizamos los mapas de jugadores relevantes de cada equipo también podemos entender el resultado del partido. Frank Ribery no erró ningún pase en el Primer Tiempo frente al Borussia, mientras Cristian Pulisic se mostró más impreciso y con pases más cortos.

Disparos

El mapa de disparos es bastante habitual verlos en apps de fútbol o en programas de televisión. Principalmente, indica desde dónde fue el disparo (distancia), qué finalidad tuvo (desviado, bloqueado, a puerta) y no mucho más que eso. En esta ocasión, además de las variables que venimos usando traeremos la famosa y popular *expected goals* y la columna de cuadrantes, que muestra a qué parte del arco iba el remate, siendo 1 la parte superior izquierda y 9 la inferior derecha; si indica 0, significa que el tiro no fue al arco. Este evento no incluye penales.

En este caso analizaré los disparos del Barcelona, Atlético Madrid y Real Madrid, que en la Temporada 2017 - 2018 terminaron entre los tres primeros lugares de La Liga Española.

En primer lugar, comenzaremos analizando individualmente a los goleadores de cada equipo y los disparos de cada uno. Hablamos de Cristiano Ronaldo (Real Madrid), Lionel Messi (Barcelona) y Antoine Griezmann (Atlético Madrid).

Disparos
Jugador: L. Messi
Equipo: Barcelona

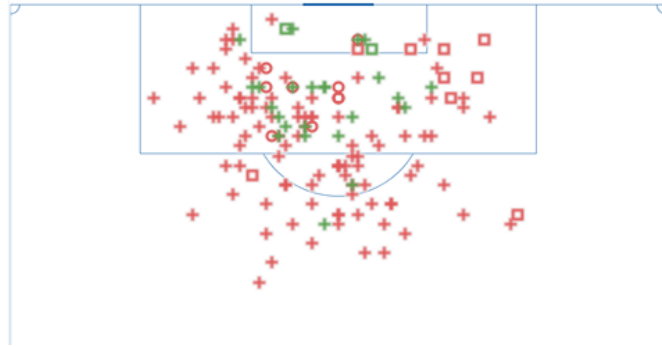


Figura 20. Calidad de los disparos de Messi en Liga Temporada 2017-2018

Disparos
Jugador: A. Griezmann
Equipo: Atletico Madrid

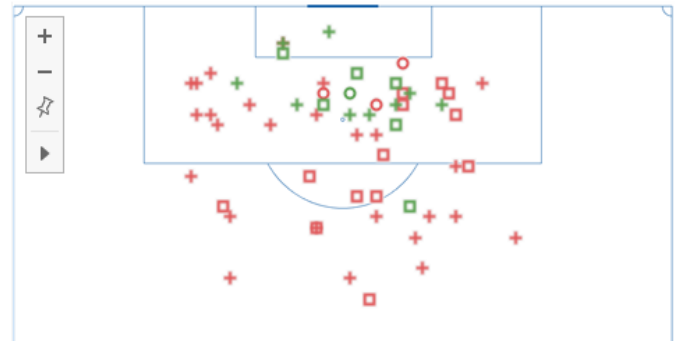
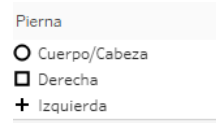
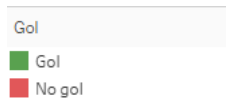


Figura 21. . Calidad de los disparos de Griezmanni en Liga Temporada 2017-2018



Disparos
Jugador: Cristiano Ronaldo
Equipo: Real Madrid

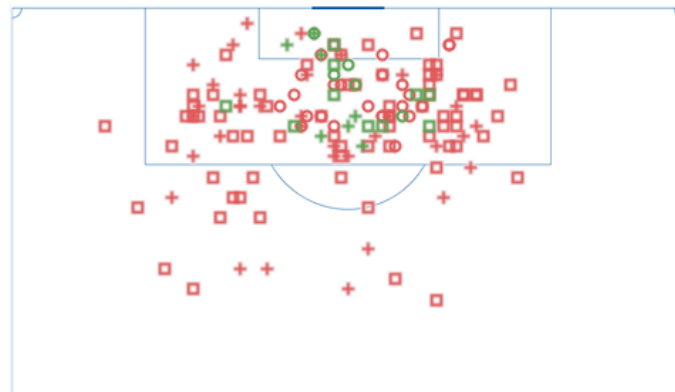


Figura 22. Calidad de los disparos de C. Ronaldo en Liga Temporada 2017-2018

En cuanto a **cantidad de disparos** observamos claramente que Griezmann es el que menos intenta y entre Cristiano y Messi están parejos. De **media distancia** el jugador que más prueba es el portugués, aunque solo tiene un gol fuera del área, al igual que Messi, mientras que el francés no tiene ninguno. Cristiano parecería ser el más versátil de los tres, ya que tiene goles de derecha, izquierda y cabeza en similar cantidad. El argentino prácticamente tiene solo goles con su zurda y a Griezmann no parece importarle para que pierna le queda la pelota a la hora de disparar.

Otra buena manera de compararlos es la **dirección** en la cual patean, es decir si lo hacen hacia la derecha abajo del portero o si prefieren hacerlo arriba a la izquierda. Lo veremos a continuación:

Disparos

Jugador: Cristiano Ronaldo

Equipo: Real Madrid

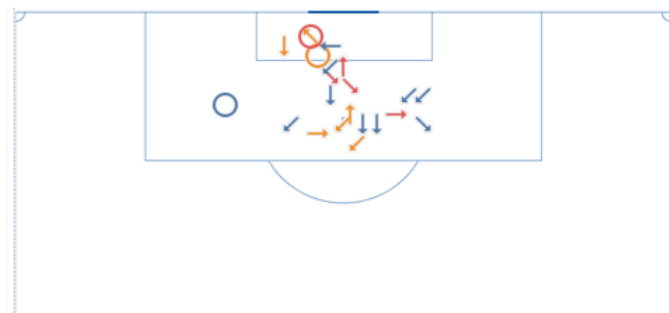


Figura 23. Dirección de los disparos de Messi en Liga Temporada 2017-2018

Disparos

Jugador: A. Griezmann

Equipo: Atletico Madrid

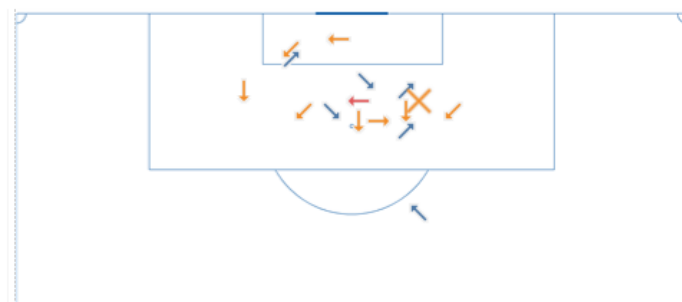


Figura 24. Dirección de los disparos de Griezmanni en Liga Temporada 2017-2018

■ Cuerpo/Cabeza
■ Derecha
■ Izquierda

Guardamises
↘ Cuadrante_1
↑ Cuadrante_2
← Cuadrante_4
○ Cuadrante_5
→ Cuadrante_6
✓ Cuadrante_7
↓ Cuadrante_8
↙ Cuadrante_9

Disparos

Jugador: L. Messi

Equipo: Barcelona



Figura 25.. Dirección de los disparos de C.Ronaldo en Liga Temporada 2017-2018

Empecemos por Lionel. Parecería no tener un patrón muy claro. La mitad de sus goles fueron hacia la esquina inferior izquierda. Lo único que puedes destacar es que suele patear a baja altura y a veces a media, casi nunca ejecuta remates altos.

Por el lado de Antoine, es capaz un poco más claro. Cuando patea con la derecha el balón suele ir a la derecha, mientras que cuando lo hace con su pie hábil la pelota lo hace hacia su izquierda.

Si vemos los goles de Cristiano, de sus goles con la pierna diestra, solo hay uno que se deposita en ángulo inferior derecho, los demás van en su mayoría a la izquierda, es decir, van cruzado. En cambio cuando dispara con la zurda es un poco más variado, pero el lado izquierdo, el derecho del arquero, en general, es el lugar por excelencia para los remates del portugués.

A nivel disparos creo que es mejor analizarlos a nivel individual que a nivel grupal, ya que cada jugador tiene sus preferencias a la hora de rematar.

Expected Goals: xG vs xGA

Debe ser la variable moderna más utilizada. Todos hablan de ella, pero pocos saben realmente lo que significa. Calcularla no es nada fácil, por eso los resultados que conseguí no me parecen contundentes, pero sirven para poder darse una idea de la peligrosidad de las llegadas de un equipo. “Los goles no se merecen, se hacen”, sin embargo, los xG llegaron un poco para darle validez al merecimiento de un equipo durante los noventa minutos. Básicamente, los *goles esperados* son un indicador estadístico que asigna una probabilidad de que una ocasión sea gol en función de las características de la jugada. Depende de la distancia, el ángulo, la cantidad de jugadores que hay entre la pelota y el arco, la potencia del disparo, con qué parte del cuerpo se golpeó el balón, cómo se creó la ocasión (si fue un centro, un contrataque) y todas las variables que creas importantes. Para ello, con un modelo de Machine Learning de regresión lo puedes hacer sin problema. En este caso solo calculé la distancia y el ángulo como variables y no realicé el modelo, saqué los números de [xG Model - Design and Implementation with R Tidymodels \(thesignificantgame.com\)](https://thesignificantgame.com), que indican cuanto peso tienen cada variable elegida. En consecuencia, los xG darán alto (como máximo 1) cuando el remate haya sido cercano al arco y en el centro y dará más cercano a 0 cuando el disparo sea de lejos y más lejano al centro.

Realicé tres gráficos de xG: dos con datos provenientes de FBRef y uno con los míos (de WyScout). Los de FBRef toman los datos directamente de StatsBomb, así que seguramente son más confiables que el modelo que traté de realizar. Empecemos.

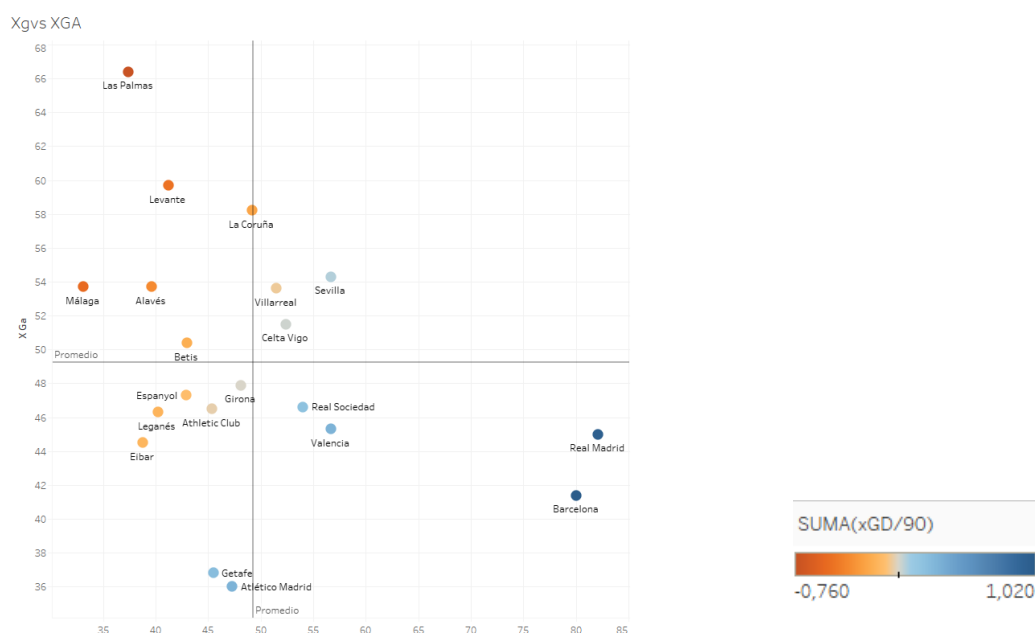


Figura 26.. Muestra los xG en eje y y los xGA en eje x y de los equipos de La Liga Española durante temporada 2017-2018

Gráfico bastante habitual donde se comparan los xG vs xGa, es decir, los expected goals contra los expected goals against, lo que equivale a los goles esperados del rival durante tu partido. Luego están los xGD, que son la diferencia entre xG y xGA. Podemos notar que cada equipo está con un color distinto y esto tiene que ver en si fue mayor su xG que su xGA va a estar en azul y en caso contrario estará en rojo.

Los equipos que están en cuadrante inferior derecho son aquellos que tienen un alto xG y un bajo xGA, es decir, llegan mucho y de manera clara al arco contrario y prácticamente ni le llegan a la suya: hablamos de Barcelona y Real Madrid, y en menor medida, Real Sociedad y Valencia. En el cuadrante inferior izquierdo, son equipos defensivos que le llegan poco y que crean pocas situaciones de gol: el Getafe y el Atlético Madrid son ejemplos claros de estos conjuntos. Luego tenemos equipos que crean mucho y quedan desprotegidos atrás, casos que se encuentran en la parte superior derecha, tales como Sevilla, Villarreal y Celta. Para finalizar están en la parte superior izquierda aquellos que le llegan mucho y crean pocas situaciones, donde el caso más representativo en esa temporada fueron Las Palmas, equipo que terminó en la posición 19 y terminó descendiendo.

Expected Goals: expectativa vs realidad

Otra manera de analizar los expected goals es compararlos con lo que sucedió en la realidad. Nos fijaremos si nuestros equipos metieron más goles de los que esperábamos y si les metieron menos o viceversa.

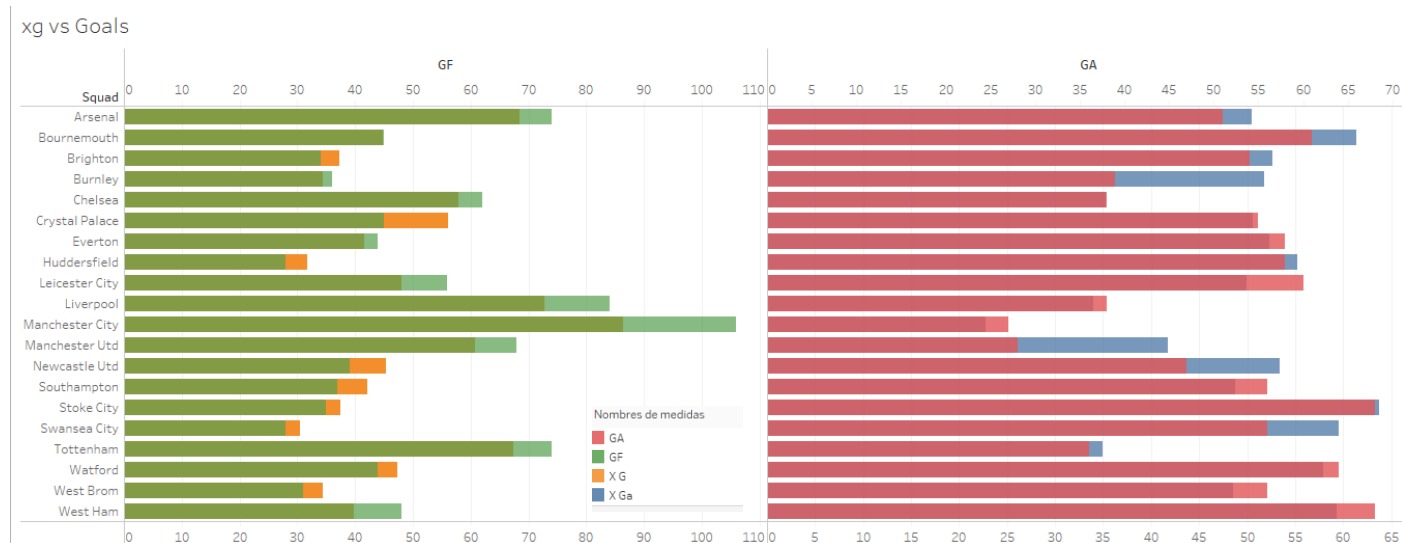


Figura 27. En la parte izquierda, muestra el contraste de los goles esperados contra los goles reales; en la parte derecha, lo observamos entre los goles en contra y los goles esperados en contra de los equipos de la Premier League

Vemos que en la parte izquierda se contraponen los goles contra los goles esperados y en la parte derecha los goles en contra Vs. los goles esperados en contra. En este caso analizamos la Premier League y observamos que en caso de que veamos el color naranja significa que los xG fueron mayores a nuestros goles en la realidad; si no vemos el color naranja, es lo contrario, es decir, marcamos más goles que los que se esperaban

de nosotros. En el lado derecho, si llegamos a ver el color azul en nuestros equipos equivale a decir que nos marcaron menos goles que lo que establecía el modelo, mientras que si no vemos el color azul es que nos marcaron más goles que lo que se esperaba.

Por lo tanto, vamos a ver casos significativos, como por ejemplo, el Crystal Palace que debería haber marcado más goles y por ende haber terminado en una mejor posición seguramente. Un caso contrario es el West Ham que marcó 10 goles más que lo que decían los xG, puede ser que haya llegado poco, pero siendo muy efectivo.

Del lado derecho podemos tomar al Manchester United, equipo al que le marcaron 15 goles menos de lo que debería haber pasado, lo que engrandece la performance del arquero, en este caso si vamos a las estadísticas reales, David De Gea esa temporada se convirtió en el arquero con más vallas invictas (18) en la competición. Contrariamente, al Leicester City le marcaron más goles de los que se esperaban, con lo cual podemos suponer que no habrá sido la mejor temporada de su arquero, Kasper Schmeichel.

Expected Goals: xG acumulado por partido

A medida que va transcurriendo un partido uno siente que uno de los dos equipos debería estar ganando y esto lo podemos medir con el xG acumulado por partido. Recordemos que en este caso el modelo fue creado por mí, dejando de lado algunas variables importantes, por lo que los resultados pueden no ser del todo correctos.

xG acumulado

Equipo: Manchester United

Partido: Manchester United - Everton, 4 - 0

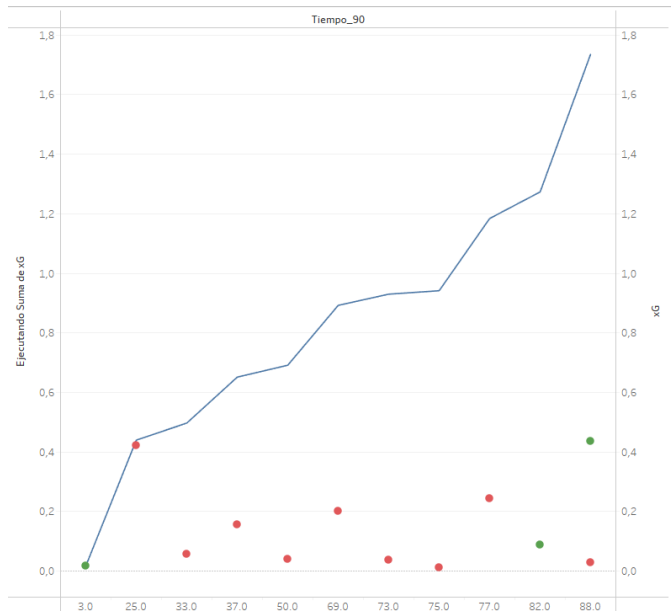


Figura 28. Indica la evolución de los xG del Manchester United durante el partido Vs. Everton

xG acumulado

Equipo: Everton

Partido: Manchester United - Everton, 4 - 0

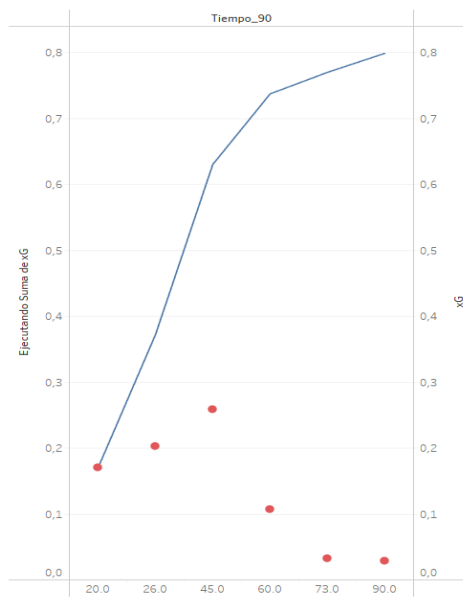


Figura 29. Indica la evolución de los xG del Everton durante el partido Vs. Manchester United

Analizamos un partido de la Premier League entre Manchester United (lado izquierdo) y Everton (lado derecho). Los remates están marcados con puntos y cuanto más claro haya sido el disparo más arriba se

encontrará el punto; si fue gol es verde y si no es rojo. El partido terminó 4 a 0 en favor de los Diablos Rojos. Podemos ver como a los 3 minutos un remate con chances bastantes bajas de ser gol terminó al fondo de la red. Para cerciorarse, pueden ver el golazo de Antonio Valencia aquí [Manchester United 4-0 Everton \(17-18\) | Premier League Classics | Manchester United - YouTube](#), la verdad es que había muy pocas chances de que sea gol. Luego si siguen viendo hay una clara muy chance para Lukaku que es la que tiene más de 0.4 de xG. Y así podemos ir viendo todo el video y corroborando que tan probable que sea gol es la chance y comparando con el modelo. Por lo que vi de este partido, el índice de los xG resultó mejor de lo que pensaba, aunque hay muchas cosas por mejorar.

Faltas

Cuando decimos que un equipo es aguerrido seguramente la cantidad de faltas sea un factor importante para evaluarlo de esa manera. Puede ser una manera para que el rival no pueda entrar en ritmo y no es lo mismo hacer una infracción en la puerta del área grande que en mitad de cancha. También vamos a analizar por qué le sacan amarillas los jugadores y si hay equipos más proclives a recibirlas por “hacer tiempo” o protestarle al árbitro.

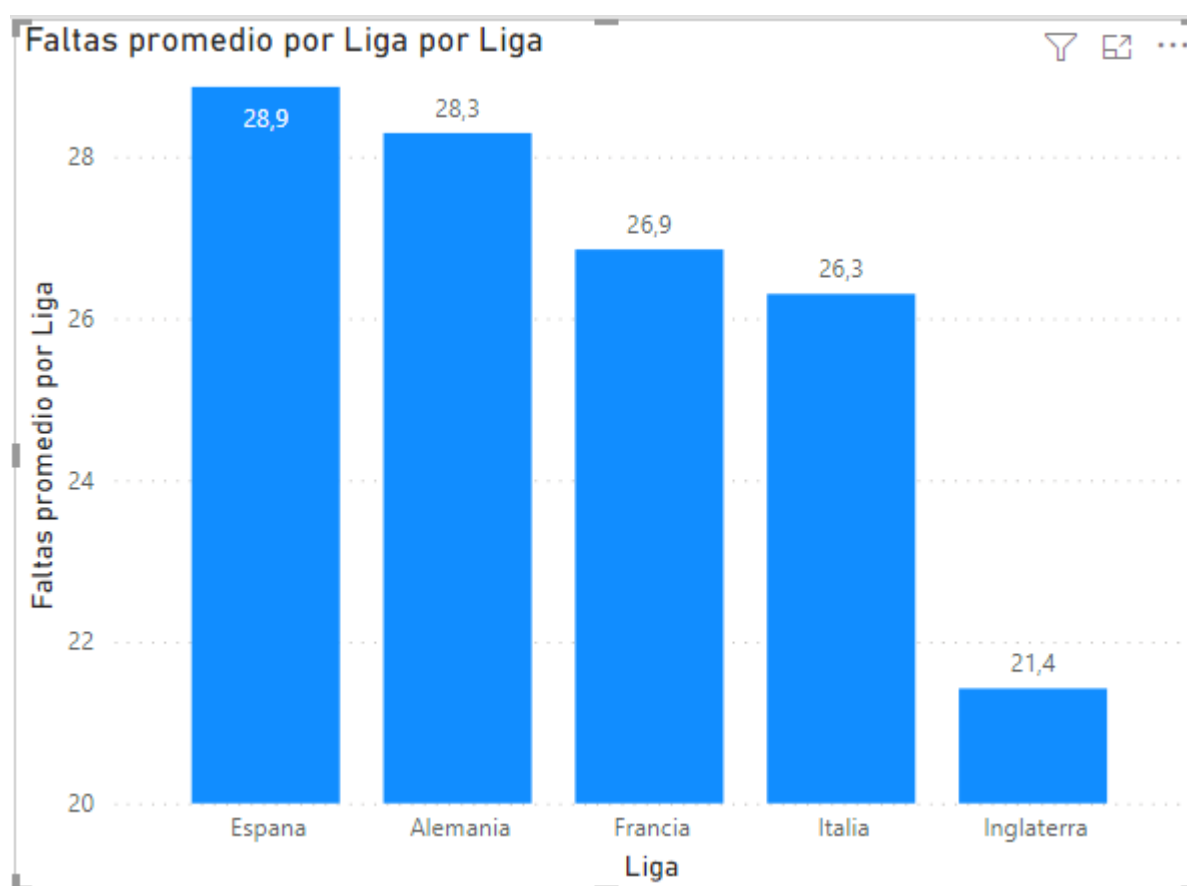


Figura 30. Indica las faltas promedios que se marcaron en las cinco ligas top de Eruopa por partido

Al igual que lo hicimos con los goles lo hacemos con las faltas. La Premier League, tal la vez más conocida por el “siga o siga” o “play on” (en inglés), dando a entender que los árbitros son más permisivos y no cobran con

facilidad un *foul*. Por otro lado, tenemos a la Española y Alemana con 28,9 y 28,3 faltas en promedio por partido, como las Ligas en la cuales se pitan más.

Tarjetas amarillas y rojas

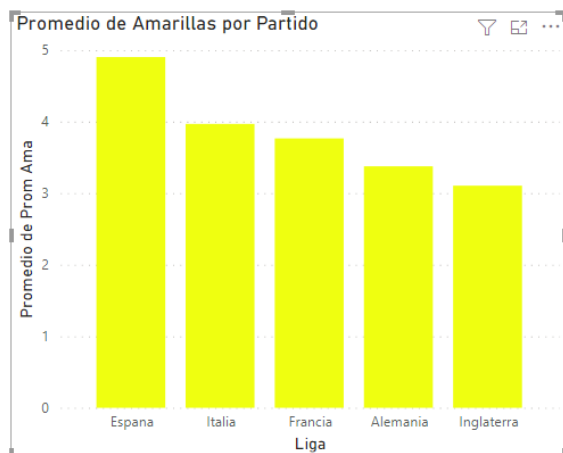


Figura 31. Promedio de amarillas por partido en cada liga

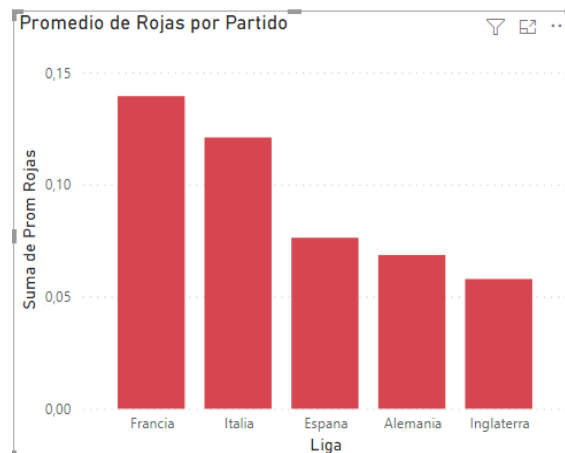


Figura 32. Promedio de rojas por partido en cada liga

Tiene sentido entonces que España sea el país en el cual más tarjetas amarillas se muestran por partido, con un promedio de 5 y que Inglaterra sea el país con menos amonestaciones por encuentro, con un promedio de 3. En las rojas las cosas cambian un poco. Francia es la Liga en la cual más rojas se sacan por partido con un 0,15 aprox, es decir, cada 6 partidos hay una roja, mientras que en la Premier League para ver una expulsión tenemos que esperar 16 partidos o lo que es lo mismo cada tres fechas completas hay solamente dos rojas.

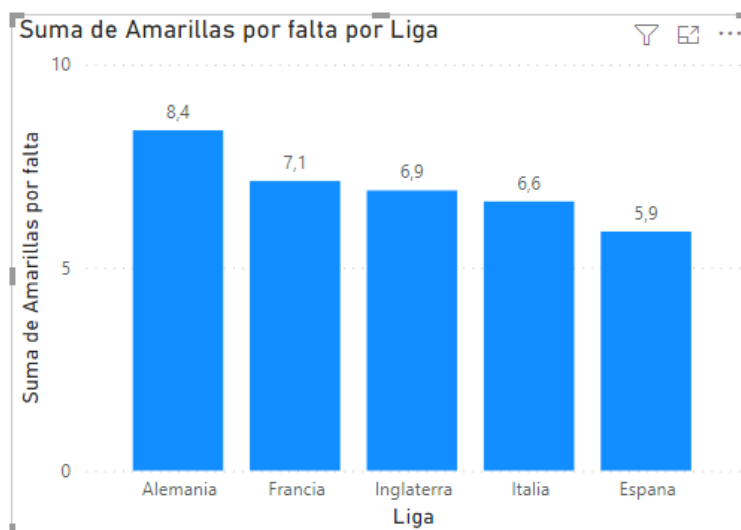


Figura 33. Cada cuantas faltas se saca amarilla en promedio por Liga

Sin embargo, cuando parecía que los datos nos decían que la Premier League era la más permisiva un nuevo gráfico nos hace replantear ese pensamiento. En la **Figura 27** observamos cada cuantas faltas un árbitro amonesta a un jugador y llegamos a que la Bundesliga termina siendo la menos rigurosa. Cada 8 faltas, saca

una amarilla, aproximadamente. Los árbitros españoles parecen no querer que el partido se les vaya de las manos, por lo que a la sexta falta amonestan al jugador.

¿Cuándo se amonesta?

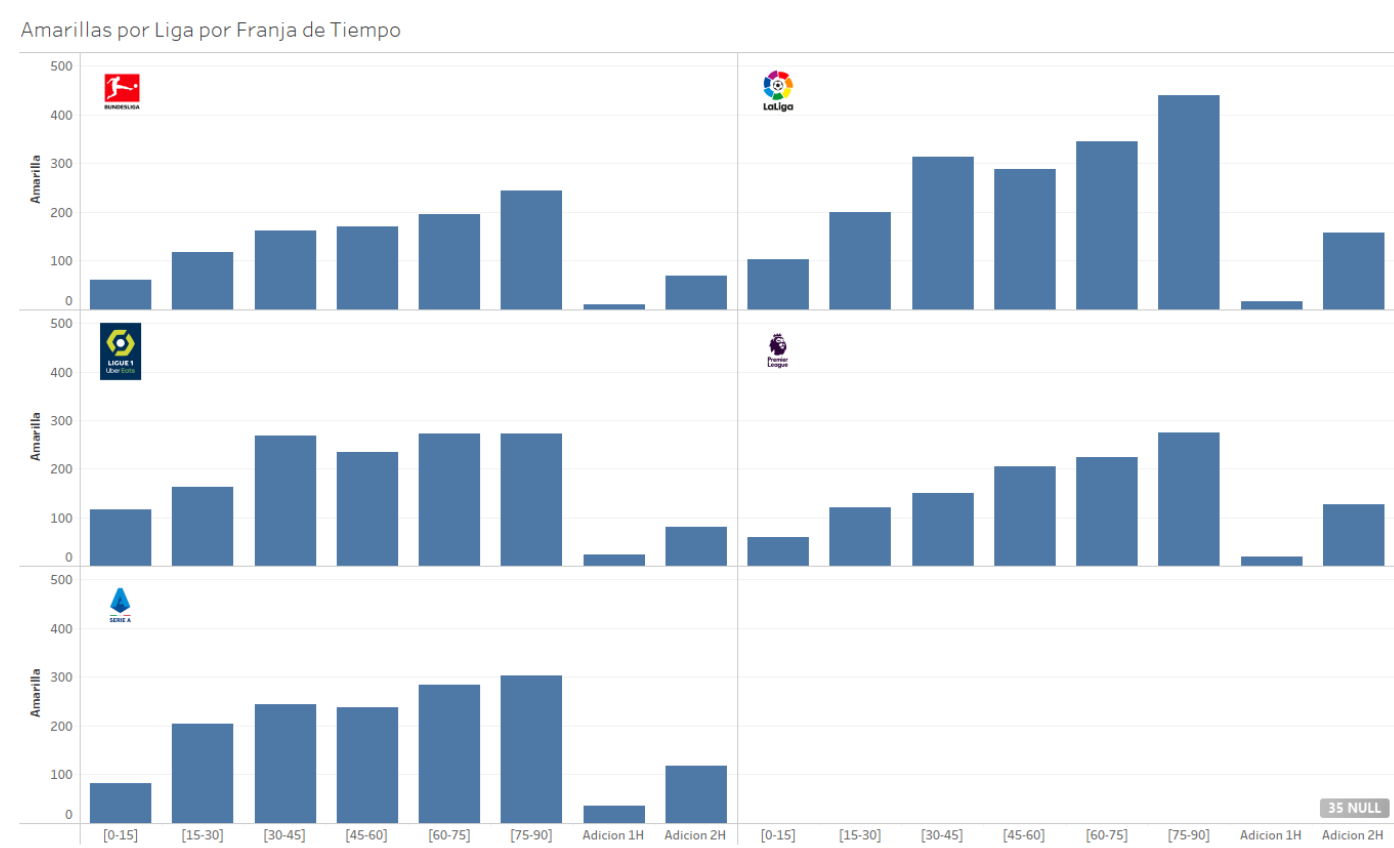


Figura 34. Cada barra señala el porcentaje de amarillas que se sacaron en cada liga según la franja horaria indicada

Siempre se dice “hay que hacer las faltas al principio porque no te amonestan” y los datos nos establecen que ese dicho popular es cierto. La franja de 0 a 15 minutos es en la que menos amarillas sanciona el referí. Después depende de la Liga, pero la mayoría es en los últimos 15 minutos del encuentro donde tiene la tarjeta “fácil”.

¿Por qué se amonesta?

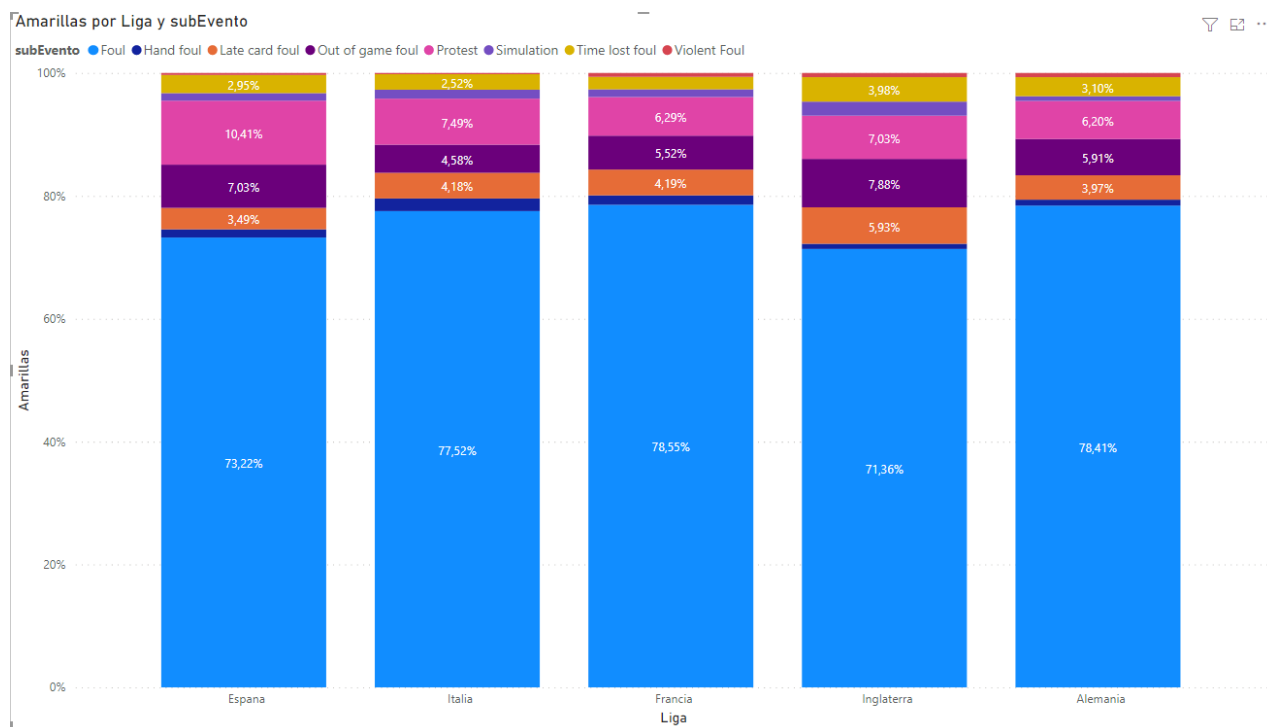


Figura 35. El porcentaje de los distintos motivos por cuales se amonestó en cada liga

Para terminar con los análisis de las Ligas, veremos cuales son los motivos por los cuales se sacan amarillas en cada una de ellas.

Out of game foul (o faltas fuera del partido): son aquellas que se dan principalmente mediante agarrones en situaciones de pelota parada y el referí se cansa y le saca amarilla a los dos. Una hipótesis posible es que en la Liga Española es donde más ocurren este tipo de acciones.

Simulación: los ingleses, amantes del “fair play”, ven con muy malos ojos cuando un jugador simula, por lo tanto, es la Premier League la que más sanciona estas jugadas.

Hacer tiempo: cuando un equipo va ganando y está con lo justo, lo más probable es que empiece a “hacer tiempo”, es decir, demorando el inicio de las jugadas y tirándose al suelo por presuntas dolencias. La Liga Inglesa es la que más lo sanciona, pero seguramente que no es la que más lo hace.

Protestar: una de cada diez amarillas que se sacan en España es por protestar. En ese tipo de acciones los árbitros españoles tienen pocas pulgas.

¿Dónde ocurren las faltas?

Por último, como con todos los demás eventos, existe el mapa del campo de juego en el cual vemos en dónde se produjeron las faltas, si hubo sanciones (amarilla, roja) y qué tipo de infracción. En este ejemplo son las faltas de Real Madrid vs Barcelona.

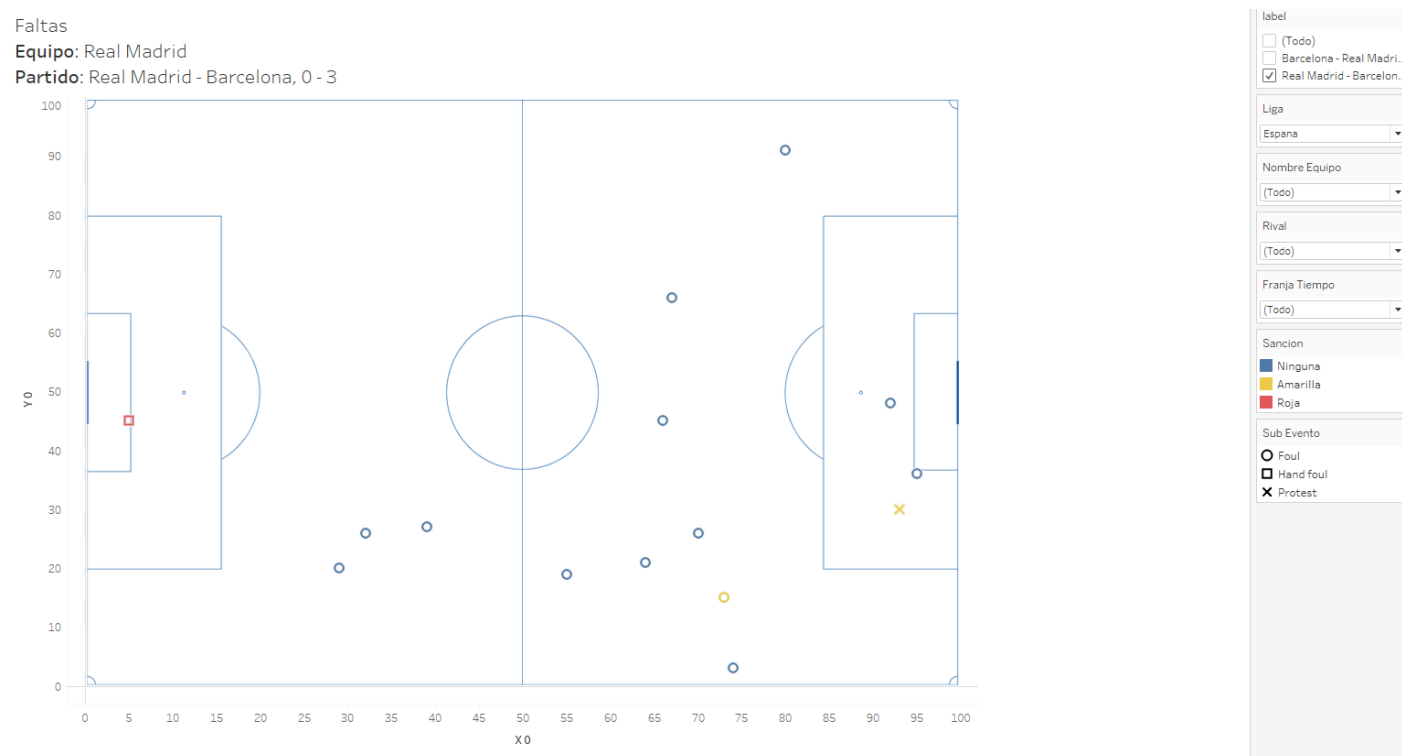


Figura 36. Mapa de las faltas que hizo el Real Madrid contra Barcelona

Duelos

Duelos en un partido hay todo en el tiempo. Pueden ser ofensivos o defensivos. Aéreos o en el suelo. Lo podés ganar, perder o ninguna de ellas. Es una buena manera para cuantificar, según corresponda, el desempeño de un jugador durante un partido. Si analizamos a Messi vamos a analizar los duelos ganados ofensivos; en cambio, con Casemiro lo haremos con los defensivos, mientras que con Pique nos fijaremos en los aéreos también.

Duelos ofensivos

Los duelos ofensivos son la manera para entender si un jugador es bueno driblando o haciendo regates. Sin fijarse en las estadísticas imaginamos a Messi y Neymar en los primeros puestos. Sin embargo, con este evento hay que tener cuidado de una cosa: **importa más el porcentaje que la cantidad.**

Hay veces que escuchamos que Vinicius es el jugador que más duelos perdió en La Liga y es bastante obvio que así suceda. Es que el brasilero cada pelota que agarra quiere encarar a los rivales y muchas veces va a perder. Es mucho más probable que esa etiqueta de perder duelos la tenga Vinicius que la tenga otro jugador que no encare tanto. Por eso, es más valorable la eficacia que tiene que la cantidad. Tampoco es lo mismo ponderar a alguien por haber ganado el 60% si fueron 40 duelos los que disputó que si fueron 500, en consecuencia, hay que fijarse bien y tratar de encontrar criterios adecuados para un buen análisis.

Elegí mínimo que haya tenido 200 duelos, un aproximado de entre 5 y 6 duelos ofensivos por partido, según la Liga que corresponda.

Porcentaje Duelos Ganados

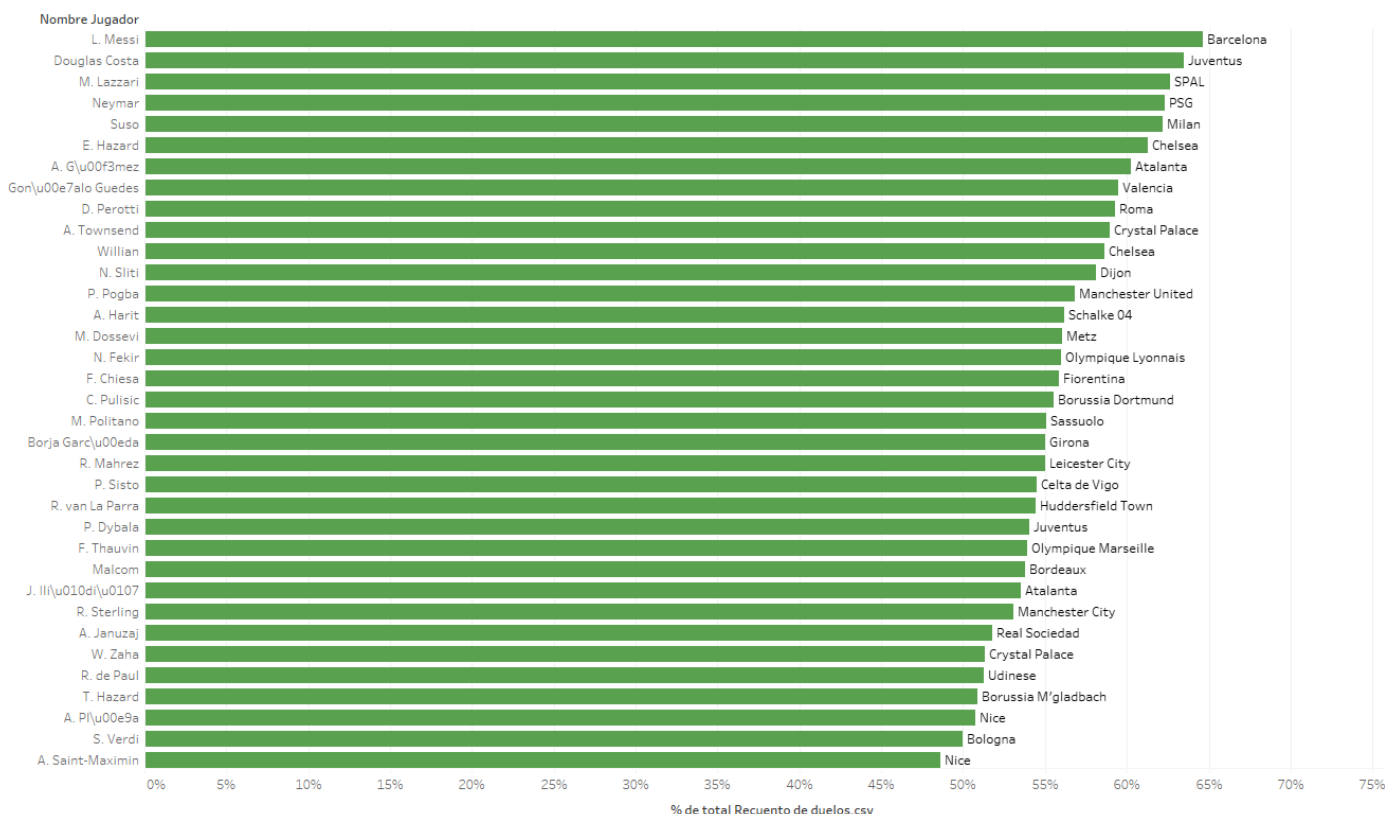


Figura 37. Muestra los mejores jugadores con mayor porcentaje de duelos ganados ofensivos con un mínimo de 200 duelos ofensivos disputados

Lionel Messi, como sospechábamos, es el jugador con al menos 200 duelos ofensivos, que más gana y entre los primeros 10 no vemos ninguno de la Bundesliga.

Si aumentamos como mínimo a 300 duelos, cada vez son menos los jugadores que quedan.

Porcentaje Duelos Ganados

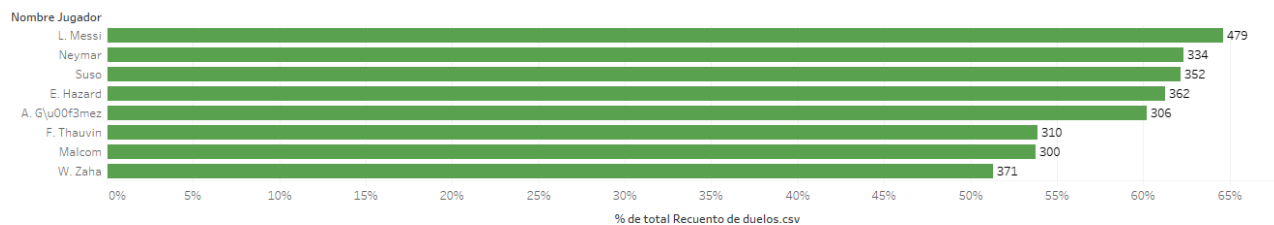


Figura 38. Muestra los mejores jugadores con mayor porcentaje de duelos ganados ofensivos con un m nimo de 300 duelos ofensivos disputados

 Si comparamos los duelos ofensivos de Neymar con lo Messi?

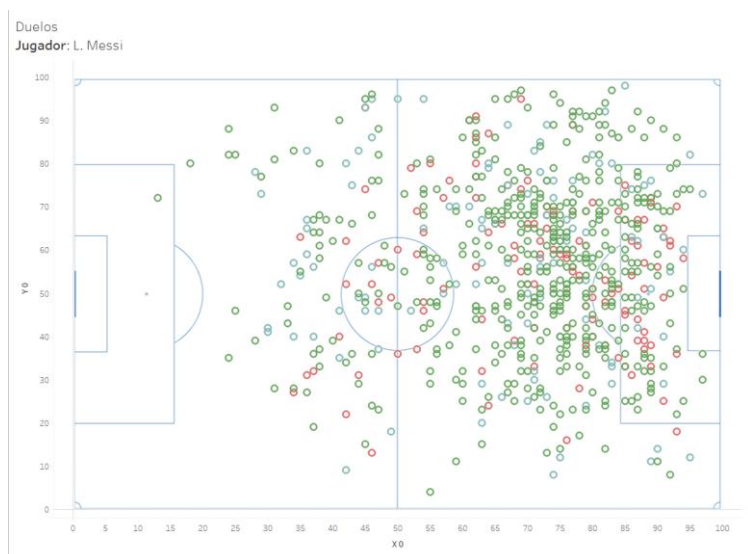


Figura 39. Mapa de los duelos ofensivos de Messi durante la Liga en la temporada 2017-2018

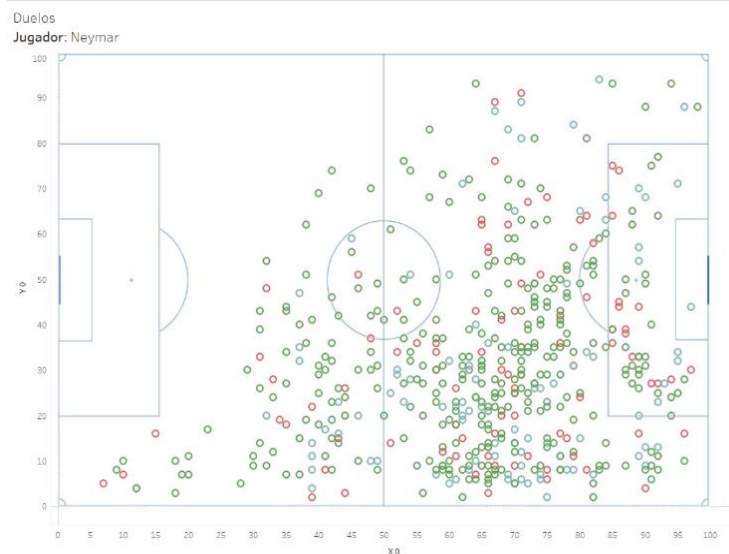


Figura 40. Mapa de los duelos ofensivos de Neymar durante la Liga en la temporada 2017-2018

Duelos defensivos

Con filtro de al menos 80 duelos defensivos

Porcentaje Duelos Ganados

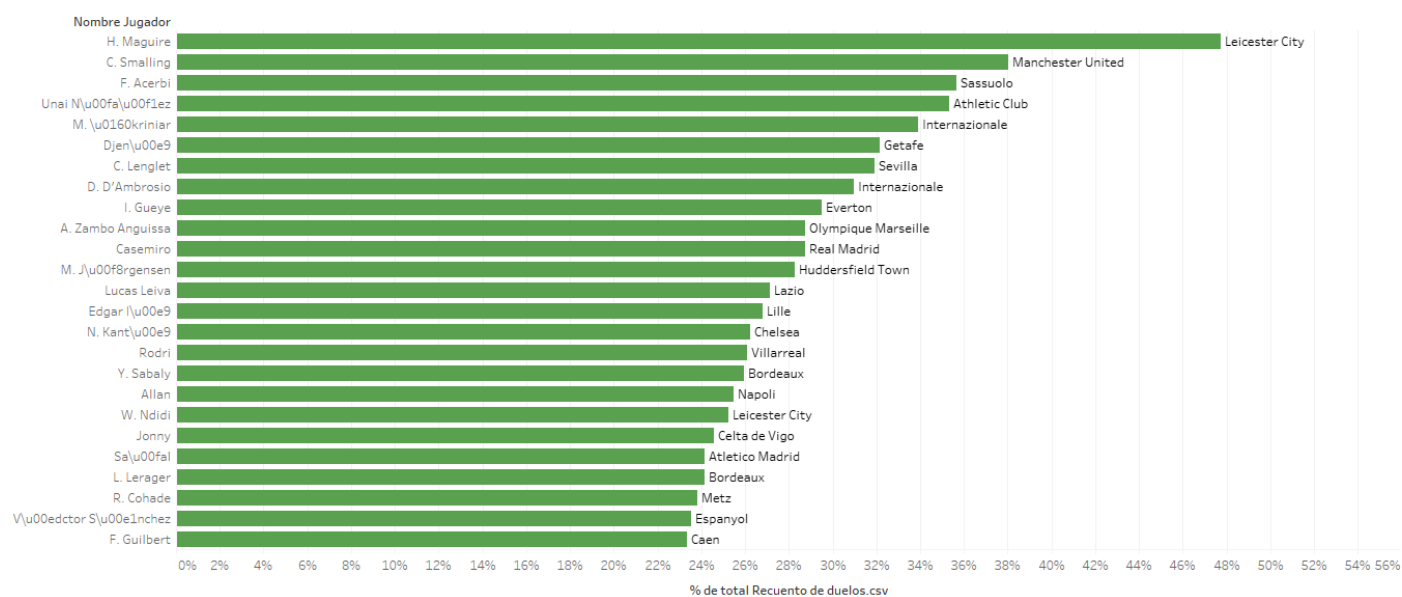


Figura 41.. Muestra los mejores jugadores con mayor porcentaje de duelos ganados defensivos con un m\u00ednimo de 80 duelos defensivos disputados

Duelos aéreos

Con al menos 100 duelos aéreos disputados

Porcentaje Duelos Ganados

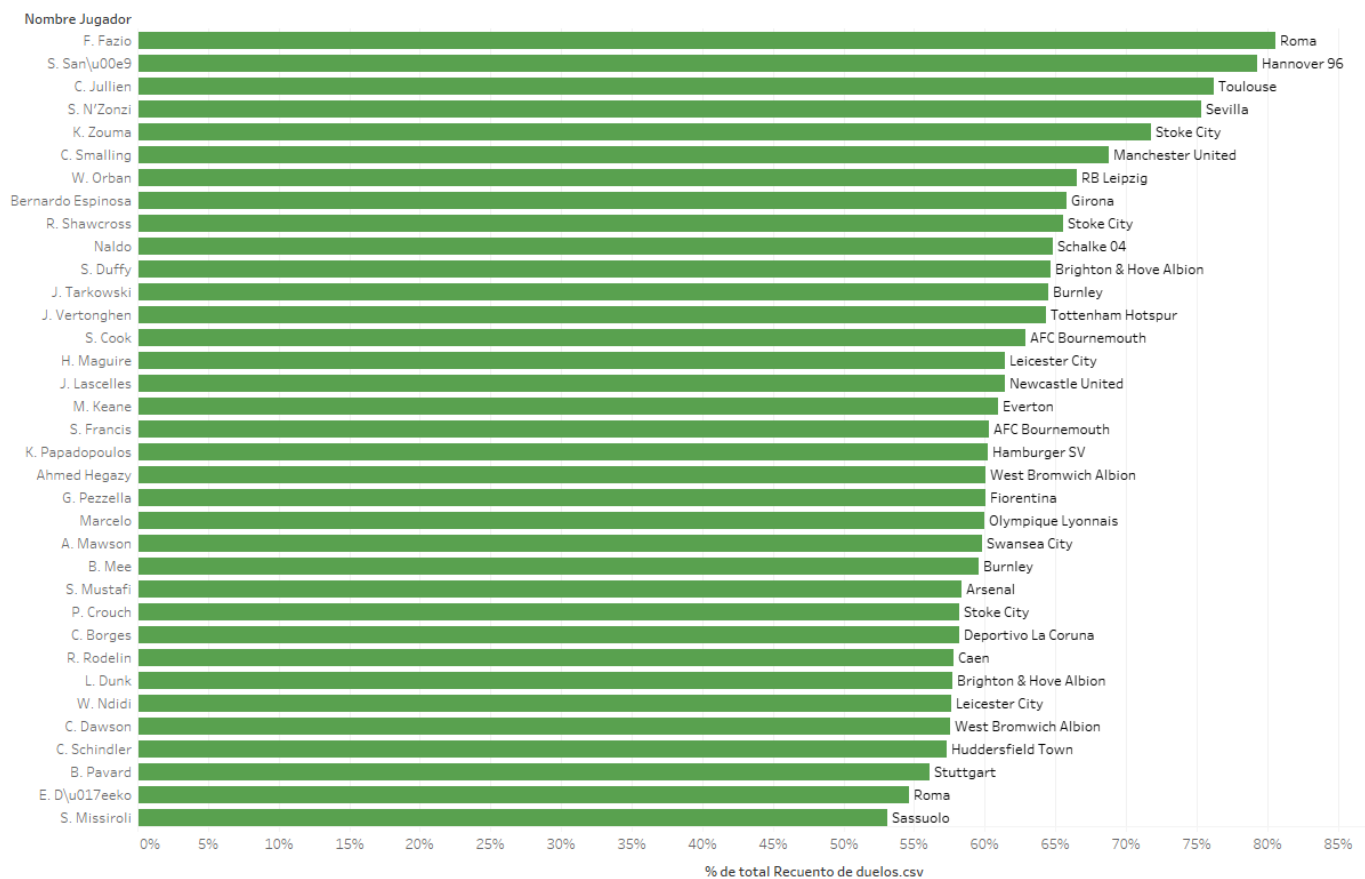


Figura 42. Muestra los mejores jugadores con mayor porcentaje de duelos ganados aéreos con un mínimo de 100 duelos aéreos disputados

Ejemplo Análisis en conjunto: Barcelona vs Real Madrid

Lo enriquecedor que tiene Tableau es luego unificar todos los gráficos en un solo Dashboard para analizar conjuntamente la performance de un equipo o jugador. En este caso, se mostrarán los datos del Barcelona en un partido disputado contra Real Madrid.

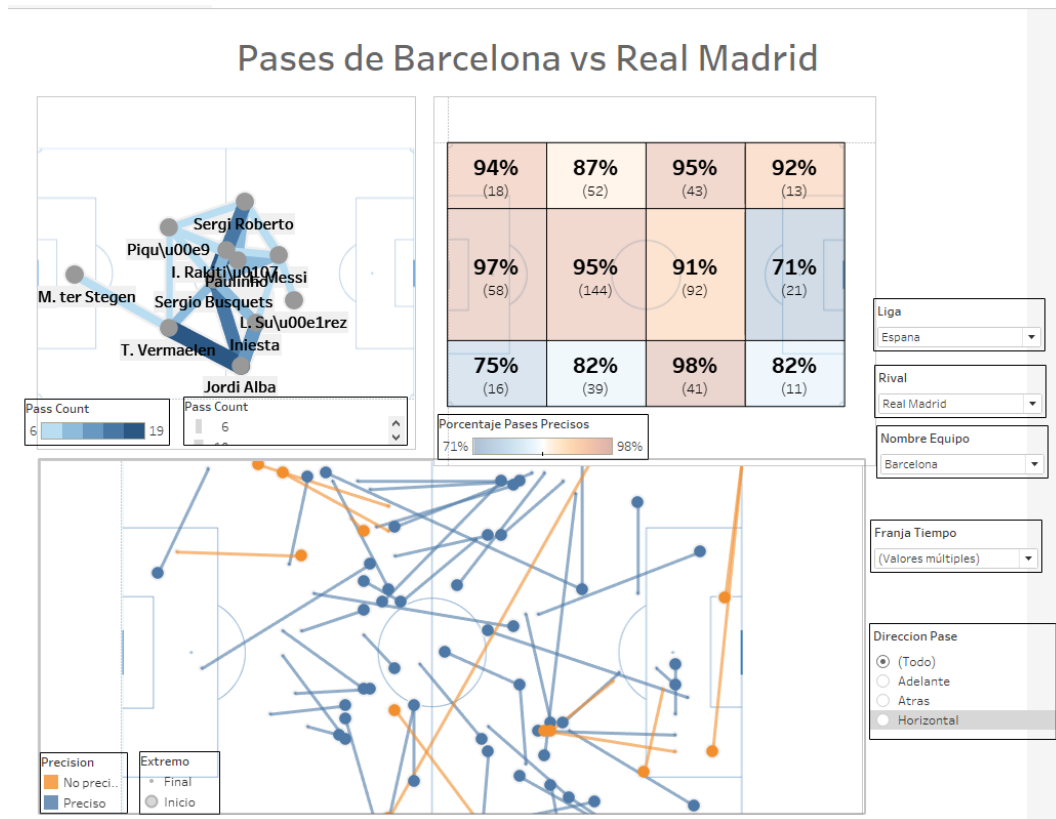


Figura 43. Resumen de los distintos gráficos de pases para Real Madrid enfrentando a Barcelona

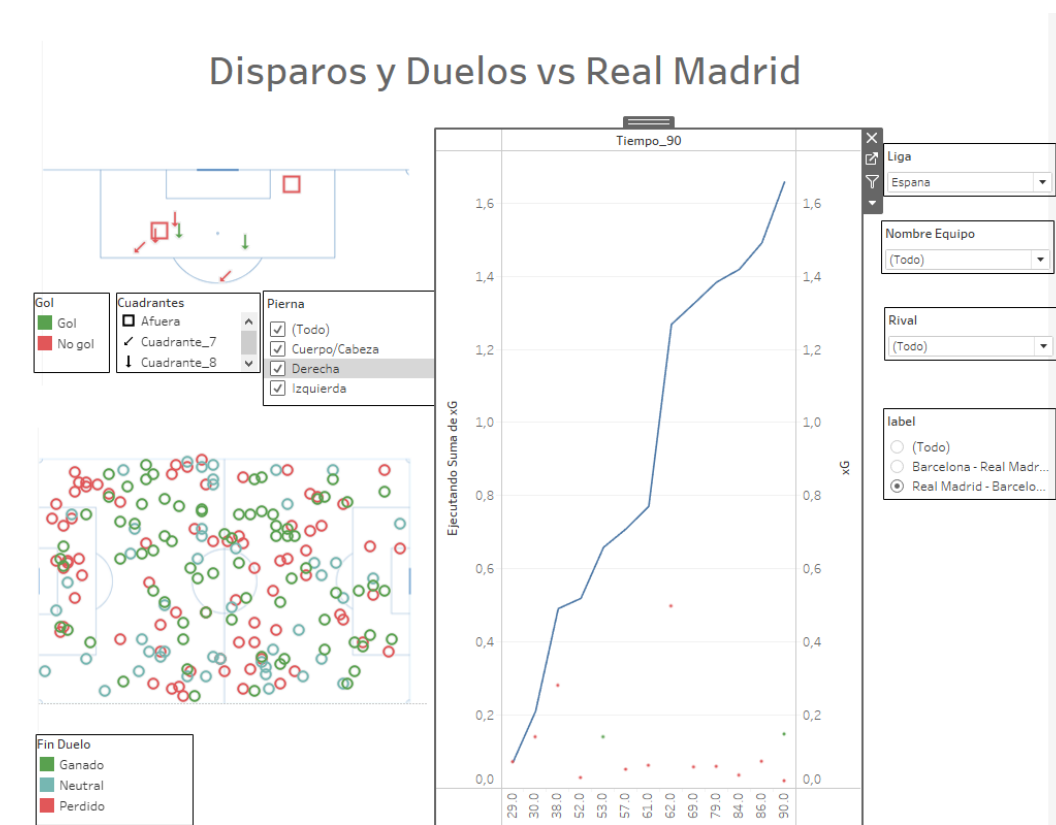


Figura 44. Resumen de los distintos gráficos de disparos y duelos para Real Madrid enfrentando a Barcelona