

# Módulo 2. Análisis de datos

Para elaborar análisis sobre el juego con el mayor grado de profundidad posible, es indispensable tener como elemento amigo al dato. Importa mucho lo que percibimos y vemos, pero también es necesario contrastarlo y cuestionarlo con algo que sea objetivo y neutral. Porque el dato es siempre neutral, lo que varía, y aquí es donde nos detendremos la mayor parte de la explicación, es el uso que se le da, no el número. En este segundo módulo, el objetivo será el de dar a conocer el porqué de la importancia de la estadística avanzada en el análisis del juego, cómo usarlo para romper o generar relatos de opinión y cómo hibridar el dato con lo que uno ve, porque en el equilibrio está la clave de todo. Añadiremos ejemplos tanto colectivos como individuales que nos permitan generar contenido posteriormente.

Hay una célebre frase del entrenador español Juanma Lillo en una rueda de prensa cuando era entrenador de nacional que sirve como punto de partida, allí hacía referencia sobre el rol de la estadística en el fútbol y como su influencia puede resultar interesante respecto a la información que muestra pero que en definitiva la estadística, como ciencia dura, no logra plasmar lo más importante del juego. Esta analogía la defienden muchísimos entrenadores de primer nivel que rehúyen de las estadísticas porque las ven, en muchos casos, como una reducción que no se ciñe a la realidad, sino que es perversa y muestra algo que está sesgado. ¿Es realmente así?

La pasada temporada 2022/2023, Josep Guardiola hablaba sobre la elección de Bernardo Silva como extremo diestro en la vuelta de los octavos de final de la Champions League ante el RB Leipzig en marzo de 2023. Silva es uno de los jugadores más inteligentes que hay y Pep subrayaba lo siguiente: "Hay cosas que no aparecen en los datos. Bernardo es único. Nos hace jugar mejor, ser más agresivos, y esto no aparece en los datos". Como Guardiola, muchos técnicos hablan de cosas que no aparecen en los datos, así que primero tenemos que definir qué son. La mayoría de medios deportivos han usado el dato como un definidor: goles, asistencias, pases, centros, recuperaciones. Apartados con cifras absolutas que se arrojaban al lector. Con su frase sobre Bernardo Silva, Guardiola trata de alejar al futbolista de unos apartados estadísticos clásicos, porque no destaca ni por los goles, asistencias o recuperaciones. Pero este no es el debate que nos interesa.

En este módulo hablaremos de la estadística avanzada. El comúnmente conocido como *big data*, aunque me alejaré de este término para definirlo como estadística avanzada, se entiende como un conjunto de datos enorme, tanto que los sistemas tradicionales no suelen servir para medirlos, y que, por lo tanto, plantea nuevas preguntas y posibles respuestas para las que hace no tanto no había solución. En el mundo del fútbol todavía hay un largo camino por recorrer, estando en fase de construcción en muchísimas ligas y

países. A diferencia de ligas de otros deportes, como la NBA o la NFL, donde la estadística avanzada está integrada tanto en los distintos equipos que las conforman como, lo que nos interesa, en el espectro comunicativo que las rodea. Allí están más habituados a hablar de conceptos que mezclan el dato con el análisis, ofreciendo un contenido de mayor calidad.

## El dato para (re)descubrir al colectivo

Para el periodista, el dato tiene una utilidad distinta que para el que trabaja dentro de la industria del fútbol. Lo primero que hay que tener en cuenta es que, cuánto mejor sea la pregunta que se le haga al dato, mayor será el rendimiento que ofrecerá. En ese sentido, tiene una implicación plenamente activa con el periodista, pues depende en su mayor medida de la capacidad del que los analiza a la hora de poder sacar buenas historias.

Lo primero es conocer algunos de los conceptos más importantes a los que los periodistas se enfrentan una vez empiezan a usar la estadística avanzada para explicar lo que sucede. Es relevante entender el concepto, y saber que cada proveedor los define de una forma distinta o tiene unos parámetros particulares para medir cada dato, por lo que, dentro de este mundo, habrá pocos datos absolutos, ya que dependerán de muchos factores. Partir de esta premisa sirve para evitar establecer axiomas que no se ajusten a la realidad.

Los goles esperados (a partir de ahora les llamaremos los xG) se han vuelto particularmente conocidos en los últimos años en redes sociales dentro del mundillo *nerd* que disfruta usando las estadísticas. ¿Qué son? Cada modelo tiene sus particularidades, pero básicamente tiene en cuenta los siguientes factores: distancia del jugador a portería, ángulo del disparo, parte del cuerpo con la que se hace el disparo y tipo de pase previo a este y partiendo de todos estos conceptos, se le da un valor al disparo entre 0 y 1 que determina la probabilidad de que termine en gol. Básicamente, permite medir de cuánta calidad son los disparos. Son una herramienta muy útil para medir dinámicas colectivas o individuales de cara a portería. Lo desarrollaré más adelante. En este módulo, usaré datos de la empresa StatsBomb, referencia absoluta dentro de la industria y con un modelo respecto a los xG muy preciso.

Los xG se usan de forma errónea en muchos análisis, y en este módulo trataré de poner algunos ejemplos de cómo pueden ser útiles y cómo usarlos para generar contenido diverso.

Vamos a tomar los datos (StatsBomb, s. f.) de las cinco grandes ligas para medir cuántos xG se generan en cada una de media por partido esta temporada 22/23:

- Bundesliga (1,22 xG)



- Premier League (1,20 xG)
- Ligue 1 (1,18 xG)
- LaLiga (1,10 xG)
- Serie A (1,08 xG)

¿Qué señala esto? Los xG generados de media por cada equipo de cada una de las ligas, ofreciendo una visión de qué competiciones generan más situaciones potencialmente peligrosas. *A priori*, uno puede pensar que la Bundesliga y la Premier son las más divertidas, y el dato sirve para contrastar esta información: en ambas se generan más xG de media, lo que indica que hay más ocasiones o que son de mayor calidad. Para saberlo a ciencia cierta, habría que mirar equipo por equipo.

En cambio, si miramos la media de goles de cada equipo, vemos que la Serie A adelanta a LaLiga pese a tener menos xG producidos. ¿Qué nos dice esto? Que hay menor efectividad en LaLiga a la hora de materializar las ocasiones pese a generar más volumen. El resto de competiciones sí mantienen el mismo orden.

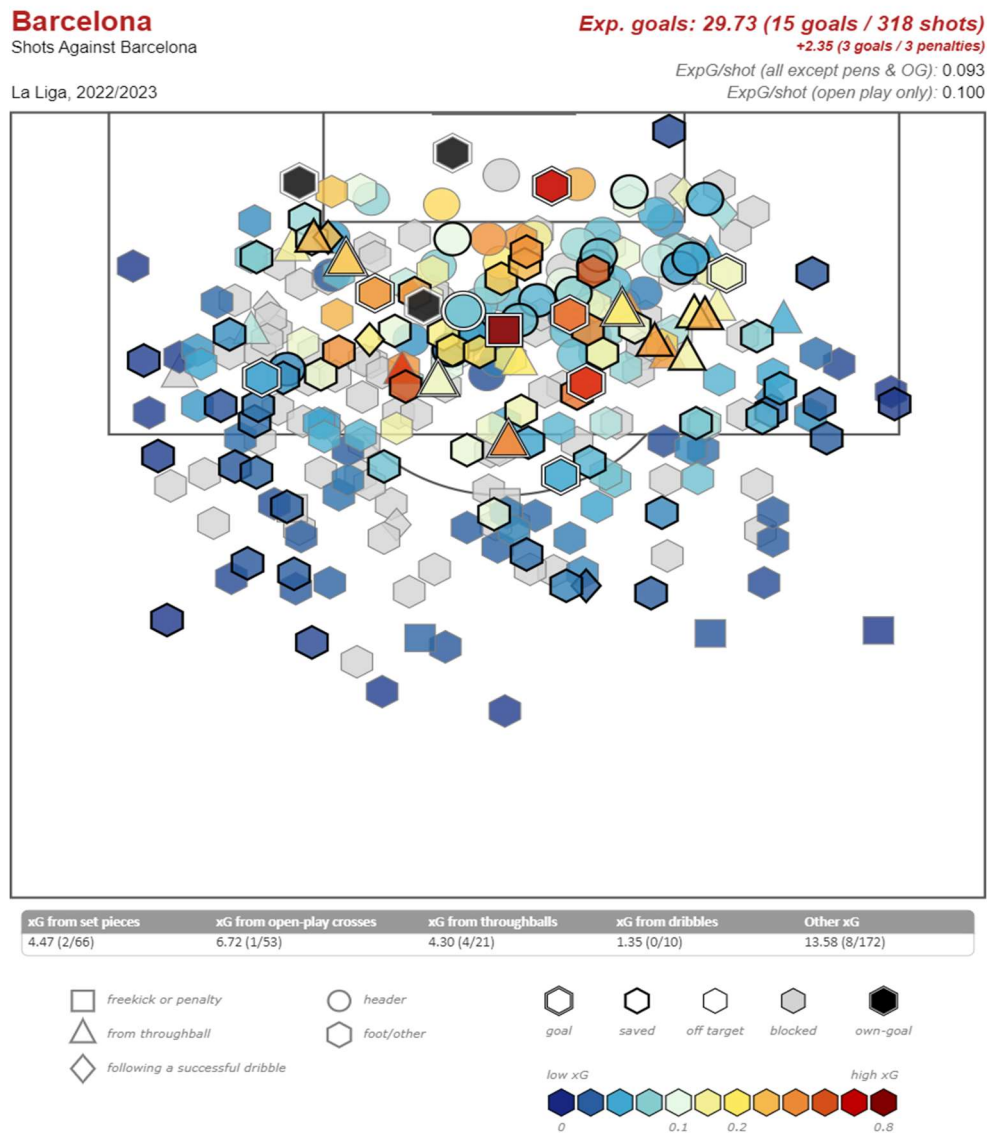
- Bundesliga (53)
- Premier League (50)
- Ligue1 (49)
- Serie A (44)
- LaLiga (43)

Solo partiendo de estos datos, que son superficiales, se daría pie a analizar cuestiones más profundas. ¿Cómo se generan ocasiones en cada Liga? ¿Existen patrones a nivel de juego para saber cómo se hacen los goles en cada competición? ¿Hay diferencias en cada una de las competiciones? Desde un dato muy simple, se puede tirar del hilo para ir generando debates alrededor de un mismo tema, ampliándolo o no dependiendo del interés de cada uno.

Básicamente, los xG nos interesan para dos temas: el estudio de los equipos y sus dinámicas y el rendimiento individual del futbolista a lo largo del tiempo. El primero nos permite saber cómo está rindiendo un colectivo con base en lo que genera y lo que termina materializando, lo que señala si el equipo está produciendo por encima de lo esperado o por debajo, siempre y cuando se tenga el suficiente contexto para analizar el dato. Es muy útil para señalar tendencias y que el periodista pueda saber si, según los datos, están o no correspondiendo con la realidad y si no lo están, ver qué grado de desviación existe para saber si es posible que cambien en poco tiempo. El del futbolista permite saber si el jugador es o no eficiente con base en el tipo de disparos que realiza, qué zonas de remate frecuente y si sobreproduce o no.

El curso 2022/2023, el FC Barcelona ha sido la mejor defensa de Europa y una de las mejores de la historia de LaLiga: solo 18 goles encajados tras 37 jornadas. ¿Qué nos puede aportar los xG en este dato? Contexto.

**Figura 1: Estadísticas del FC Barcelona**



**StatsBomb**

Fuente: StatsBomb, 2019, <https://bit.ly/3sWn1Le>

Esto es un mapa de tiros típico de StatsBomb, en el que se muestran todos los remates recibidos por el FC Barcelona hasta la jornada 36 de LaLiga. Ha recibido 318 disparos (el que menos de toda LaLiga), pero lo interesante es en fijarnos en qué muestra la estadística avanzada- Los rivales han generado 29,7 xG (sin contar penaltis), pero lo increíble es que el Barça ha encajado (sin contar penaltis) ¡15 goles! La mitad de lo que, según este modelo, debería haber encajado el equipo. Esta diferencia le sitúa en lo más



alto de las cinco grandes ligas como el equipo con mayor diferencial entre lo esperado y lo encajado.

Aquí ya podemos empezar a usar este dato para hacernos preguntas. ¿Es sostenible en el tiempo esta diferencia entre lo que los rivales generan y lo que el equipo permite? Lo que muestra el dato del FC Barcelona es que el equipo ha contado con el mejor portero, uno que ha evitado muchos goles, y ha reducido las probabilidades al rival. ¿Es correcto hablar, por lo tanto, de la **mejor defensa**? Sería más acurado referirnos al Barça como la más **eficiente**, pero una que ha mostrado depender excesivamente de las individualidades. En cambio, el Manchester City sí ha minimizado más al rival, pero no ha contado con el plus individual.

Figura 2: Estadísticas del Manchester City

## Manchester City

Shots Against Manchester City

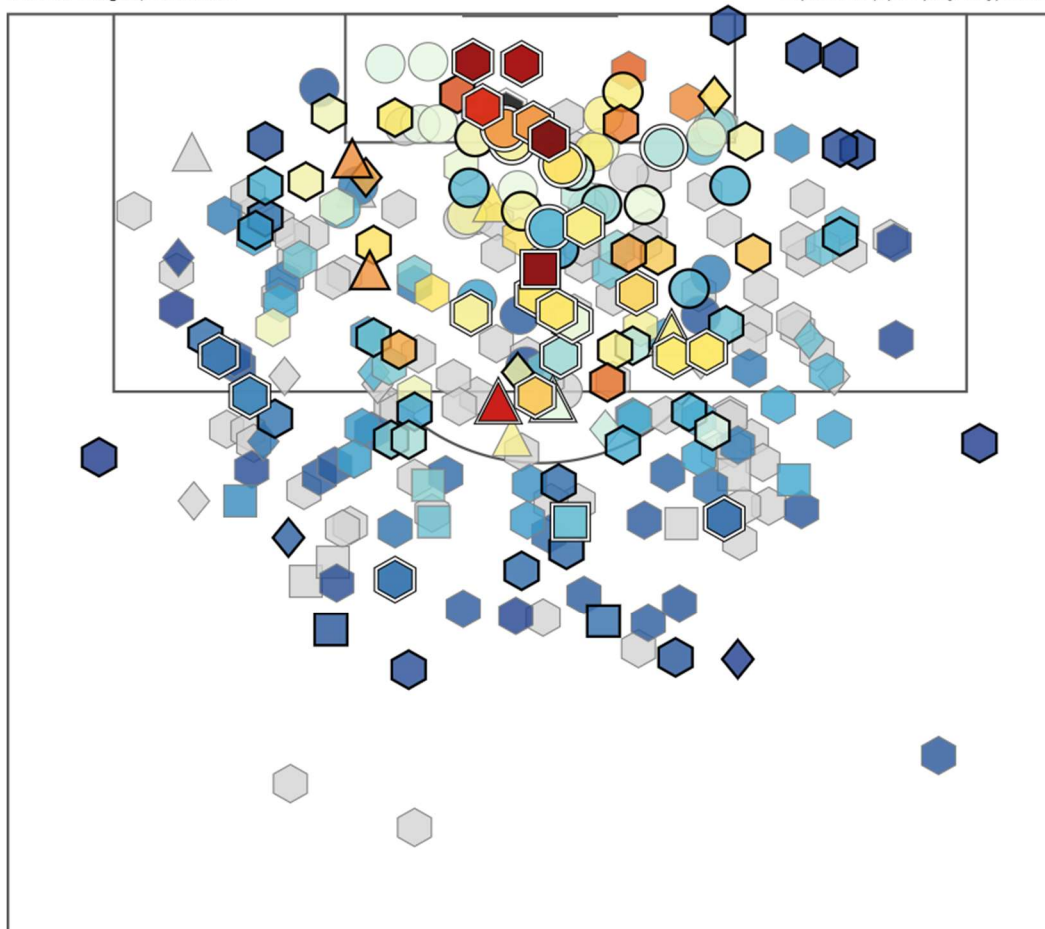
Premier League, 2022/2023

Exp. goals: 27.13 (29 goals / 283 shots)

+2.35 (3 goals / 3 penalties)

ExpG/shot (all except pens & OG): 0.096

ExpG/shot (open play only): 0.098



xG from set pieces	xG from open-play crosses	xG from throughballs	xG from dribbles	Other xG
7.44 (6/82)	6.36 (4/34)	2.11 (3/10)	1.21 (0/16)	10.23 (16/143)



# StatsBomb

Fuente: StatsBomb, 2020, <https://bit.ly/3sQHCAU>

Este es el mapa de tiros en contra del Manchester City, que ha encajado (sin contar penaltis) cerca del doble de goles que el FC Barcelona en la Premier League: 28. Son 13 más que los azulgrana. Pero, si nos detenemos en la carta de tiros, vemos que el City ha mostrado a lo largo de la temporada una mayor capacidad para minimizar disparos al rival (ha recibido 283 tiros, 35 menos que el Barça) y ha permitido menos xG que los de

Xavi: 27,1. Aun así, la tabla clasificatoria señala que el City ha recibido más goles, por lo que se genera una asociación directa: ha sido peor defensa.

Por otro lado, los xG no tienen tanta relevancia cuando hablamos de un partido en concreto, para determinar merecimientos o establecer la diferencia entre ambos conjuntos, partiendo de cuánto han generado. Normalmente, los equipos con jugadores de calidad suelen estar por encima de los xG porque tienen a futbolistas capaces de transformar situaciones más complejas en goles con mayor asiduidad. Un equipo puede ganar 2-0 habiendo generado 0,8 xG y que el rival haya generado más de 1 xG y, sin embargo, no haya anotado ninguno. Esto puede tener varias lecturas que, solo quedándonos con el dato, no se resuelven. ¿Cuántos disparos ha necesitado cada equipo? ¿Más lejos o cerca de la portería? ¿Con más o menos tiempo para ejecutarlos? Se puede generar 0,8 xG con muchos disparos de baja calidad y un 1xG con apenas dos tiros delante del portero. Los números no reflejarán el dominio en el encuentro.

¿Y qué sucede si un centro no es rematado pese a que el futbolista se ha lanzado en plancha y ha estado a dos centímetros de impactar con la pelota? ¿Cuenta como ocasión? ¿Se mide de alguna forma? No computa ni como disparo ni los xG lo miden, por lo que tiene un valor de cero. Por eso el dato, por muy valioso que pueda ser, tiene lagunas, y deberá estar complementado por nuestra mirada.

Hay que tener en cuenta todo el contexto y no guiarse solo por el dato; el dato por sí solo no suele contar nada *per se*, sino que necesita un envoltorio que le dé sentido y contenga una narrativa, un porqué. La hibridación entre dato y análisis es lo que enriquece el contenido y suma valor añadido al trabajo, por eso lo más complejo es saber elegir qué datos son relevantes, y eso solo se encuentra realizando las preguntas pertinentes.

### **El dato para generar una narrativa**

Cuando hablamos de jugadores y sus rendimientos, las estadísticas suelen tener siempre un peso enorme en la conversación y en la generación de narrativas concretas y estados de opinión. Se acude a ellas para acreditar o desacreditar según convenga. El dato se torna peligroso cuando se usa mal, porque suele generar futbolistas alejados de lo que realmente son, y narrativas que se instalan en el imaginario colectivo del aficionado. Al igual que cuando se habla de equipos, con el jugador el análisis de datos debe partir de preguntas muy concretas, teniendo siempre en cuenta el contexto en el que ese futbolista juega, la liga y sus características, el rol que tiene en plantilla, etc. El jugador no puede trasladarse a otro contexto sin atender a las variables a las que está sometido. Es, quizás, la parte más compleja del análisis de datos a la hora de generar contenido.

Algo que me interesa es cómo usar el dato para desmontar prejuicios, que suelen estar en el día a día de los medios, encasillando y etiquetando jugadores de forma errónea, algo

que denota cierta pereza analítica. Para el que narra la realidad, siempre será más sencillo contar con espacios amables en los que uno sabe lo que encontrará, y etiquetar al jugador genera seguridad; el dato, el que se usa de forma correcta, ha de servir para desdibujar fronteras y analizar en profundidad, evitando quedarte en la superficie de las cuestiones.

Esta temporada (2022/2023), Pedri González ha repartido una asistencia en todas las competiciones. Jugando como interior, y normalmente haciéndolo en zonas adelantadas, es una cifra muy baja para un jugador de su perfil. Se puede hacer un tema hablando de esta cifra absoluta y seguir repitiendo frases que acompañan al jugador desde su debut en el Barça (que solo da pases hacia atrás, que le falta valentía, etc.) o mirar más allá y ver qué hay detrás de esa estadística. ¿Por qué solo lleva una? ¿Hacia dónde van sus pases? ¿Rompe líneas? ¿Deja al compañero en ventaja? Analizar el juego para que el dato cuente cosas y no al revés, usando lo primero como base y lo segundo como un complemento que puede potenciar al juego y su desarrollo.

Si analizamos a todos los centrocampistas de las cinco grandes ligas esta temporada veremos un dato curioso: Pedri es el tercer jugador (superado por De Bruyne y Rodrigo de Paul) con más xA de todos, con 0,21 por cada 90 minutos, según StatsBomb (s. f.). Las *expected Assists* miden el número de asistencias que un futbolista tendría que dar en función de la calidad de los disparos que genera, es decir, a mayor calidad del disparo, mayor probabilidad de dar una asistencia. No es una estadística definitiva, pues un pase de la muerte tendrá un alto valor, pero no nos dirá la dificultad del envío ni cómo se ha gestado ese envío. Pero este dato es muy útil para entender el impacto del jugador a la hora de generar situaciones de gol, un medidor de calidad, aunque no de efectividad, porque el último gesto depende del compañero.

**Figura 3: Momento del pase de Dani Ceballos a Rodrygo Goes en el gol del brasileño ante el Rayo Vallecano**



Fuente: Wyscout, <https://shre.ink/2f8G>

Imagen del momento del pase de Dani Ceballos a Rodrygo Goes en el gol del brasileño ante el Rayo Vallecano. Un envío que tiene lugar a más de 35 metros y que, aun rompiendo una línea de presión, deja lejos a su compañero. La xA de este pase es de 0,04. Este pase contabiliza como asistencia porque el disparo de Rodry acaba dentro. Pero el trabajo del periodista es ir más allá. A final de temporada, a la hora de contabilizar goles y asistencias, los números pueden presentar sorpresas. Las xA permiten entender si el jugador produce situaciones de peligro y, por lo tanto, acerca a su equipo al gol, más allá de las asistencias que pueda dar. ¿Son trasladables a otro equipo y contexto? ¿Cuánta variabilidad ofrecen? Si en vez de Rodrygo el que recibe el pase es un jugador de menos calidad, ¿cuántas veces sería asistencia este pase?

En la otra cara de la moneda tenemos a Pedri. Mientras que Ceballos ha acabado LaLiga con cuatro asistencias, el canario solo ha repartido una. ¿Qué dice la xA? Que Pedri (0,21) suma casi el doble que Ceballos (0,11) por cada 90 minutos. Este dato permite poner en perspectiva el impacto de los futbolistas, alejando al jugador de anomalías estadísticas. Que Pedri lleve solo una asistencia en Liga no es algo lógico, porque no señala una carencia del jugador, sino del que finaliza.

**Figura 4: Momento del pase de Pedri ante el Betis, en una jugada que terminaría en nada**



Fuente: Wyscout, <https://shre.ink/2f8G>

Momento del pase de Pedri ante el Betis, en una jugada que terminaría en nada, sin disparo del receptor, pese a la calidad del envío (rompiendo dos líneas de pase). Esto significa que este envío ni siquiera contabiliza como **pase clave** (pase que termina en un

disparo). Se pierde en el limbo. Pero son precisamente este tipo de jugadas las que el analista debe rescatar para ponderar algo que se obvia en el debate y que queda soterrado bajo el mal nombre de **las estadísticas**.

**Figura 5: Lewandowski controla la pelota, evitando que el pase llegue a De Jong**



Fuente: Wyscout, <https://shre.ink/2f8G>

Lewandowski controla la pelota, evitando que el pase llegue a De Jong, que se encontraba en posición legal y mucho más buena para poder ejecutar el disparo. Como esta acción, ha habido muchas a lo largo del curso, y ahí el dato aporta valor porque señala una tendencia que, para el análisis, es muy relevante: que lleve o no asistencias no implica que el canario sea uno de los mejores últimos pasadores de Europa. De ahí la paradoja: el dato en bruto, descontextualizado, sirve para generar narrativas dependiendo de intereses ajenos, pero no explica ni cuenta lo que sucede.

Otro dato importante para medir el valor de un pasador son los *through balls* (balones al espacio que ponen a correr al compañero por detrás de la última línea defensiva del rival): cualquier pase que rompe la última línea del rival para que un compañero corra al espacio. En un fútbol en el que cada vez se hace más hincapié en los espacios a explotar a la espalda de la línea de presión, con equipos que van a buscar muy arriba al contrario y otros que limitan las recepciones entre líneas, el contar con pases que se salten toda esa estructura es enormemente valioso. Desde la temporada 2016/2017 miramos qué jugadores aparecen en el *TOP 15* en este apartado para ver qué tipo de futbolista suele dominar esta estadística:

- Messi (5 temporadas)
- Neymar (3 temporadas)
- Di María (2 temporadas)

- Verratti (2 temporadas)
- De Bruyne (1 temporada)
- Dani Alves (1 temporada)
- Jonathan Viera (1 temporada)

No hace falta comentar los nombres para ver al tipo de futbolista al que nos referimos, ni su nivel técnico y creativo. Aparecen algunos de los mejores jugadores del último lustro en el último tercio del campo. Analizando únicamente a los centrocampistas de este curso 22/23, el tercer jugador que más *through balls* realiza por cada 90 minutos es Pedri (0,51), solo por detrás de De Bruyne y Verratti. Es decir, el canario aparece también en un apartado que no suele aparecer en una hoja estadística normal, pero que sirve para ver qué tipo de pases suele realizar, chocando frontalmente con el discurso de que Pedri no tiene tanta influencia en el último gesto.

Seguro que muchos habréis leído o visto análisis que usan el porcentaje de acierto en el pase como un valor estadístico absoluto para analizar el rendimiento del futbolista. En la mayoría de medios se sigue confundiendo el acierto con un valor positivo, cuando por sí mismo no indica nada, o muy poca cosa. Para dotar de valor real a este tipo de dato, es indispensable formular las preguntas idóneas para que muestre datos útiles para el análisis.

El dato no muestra, sino que orienta. Ofrece una perspectiva amplia, y es trabajo de quien lo interroga cuestionarlo para que la imagen se vaya acotando. Si un central ha dado 111/113 pases en un partido, ¿Qué nos dice este dato? Realmente, muy poca cosa. Que su equipo ha tenido mucha posesión y, por lo tanto, el tipo de pase que ha tenido que realizar ha sido, en su mayoría, de seguridad.

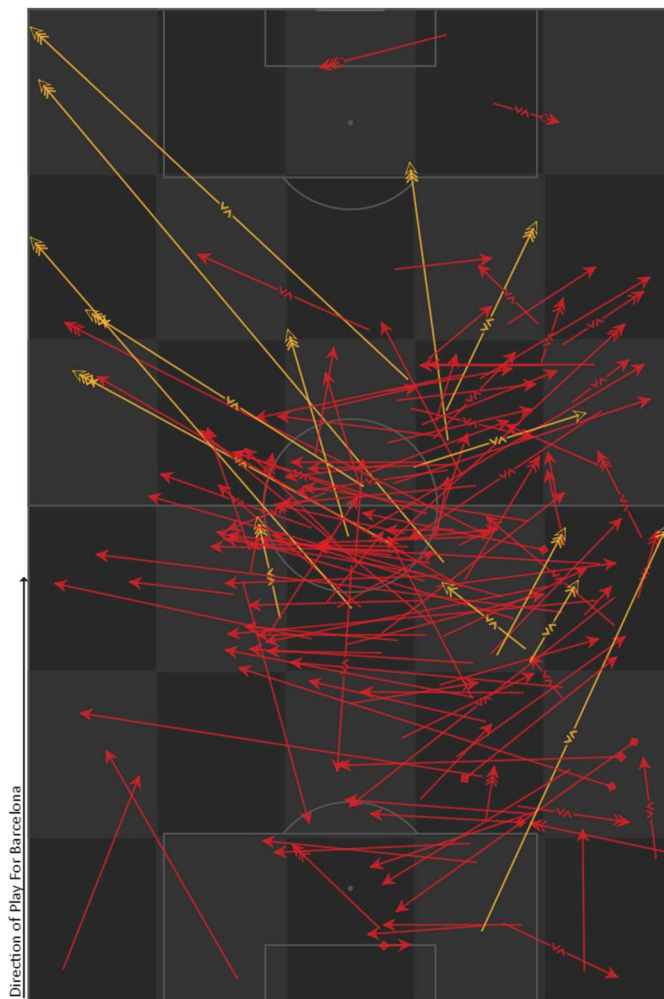


## Figura 6: Ronald Araujo

Ronald Araújo  
Barcelona

Events

2023-04-16 - 2023-05-02



**StatsBomb**

Fuente: StatsBomb, <https://shre.ink/2f8b>

Este mapa de pases muestra todos los envíos de Ronald Araujo en el partido de LaLiga del curso 22/23 ante el Getafe (en rojo los acertados, en amarillo los fallados). El central uruguayo completó 122/130 pases, una cifra altísima. ¿Se puede usar como argumento para una posible mejoría del central del Barça con balón? Para ello, habría que ver distintos puntos que son claves para analizar el dato y darle un sentido. De esos 130 pases intentados, ¿cuántos fueron verticales? ¿Cuántos batieron líneas rivales? ¿Cuántos le dieron una ventaja al poseedor? ¿Qué opciones tenía el pasador?

Para eso es importante subrayar dos conceptos clave: *eventing* y *tracking*. El primero son todos los eventos que tienen lugar en un partido, como los pases, disparos, recuperaciones... y el *tracking* se encarga de saber dónde están en todo momento los 22 integrantes del partido y la pelota, aportando un contexto indispensable para la

interpretación de los eventos. Will Spearman es astrofísico y lleva desde 2018 trabajando en el Liverpool, en uno de los departamentos de análisis de datos más preparados de todo el mundo. Desde febrero de 2023, tras la marcha de Ian Graham, Spearman es jefe de departamento: *director of research*. El astrofísico norteamericano diseñó, en 2016, un modelo que revolucionó el fútbol y que tiene que ver con el *tracking*.

Spearman diseñó el *Pitch Control*, que como él mismo explicó en la página web del Liverpool, "al ser un juego de espacios, nuestra idea es, mediante el *tracking*, saber qué zonas están bajo control de un equipo en todo momento, para saber cómo cerrar espacios o cómo abrirlos. Es la noción básica para saber cómo de valioso es el espacio. Lo llamamos oportunidades para marcar sin balón". El pitch control, que se puede encontrar en herramientas de análisis de vídeo como MetricaSports o LongoMatch, permiten ver la relevancia del espacio en todo momento, con y sin balón.

Si regresamos a lo comentado anteriormente sobre los pases de Ronald Araujo (que eran datos de *eventing*), podremos empezar a extraer el verdadero valor de cada envío, porque al saber la localización exacta de cada futbolista en el momento exacto de la jugada, permite saber si el pase dado era la mejor opción posible. ¿Cómo? Teniendo en cuenta la disposición de sus compañeros, sabiendo si el jugador estaba o no presionado en el momento de dar el pase, si estaba de espalda a la portería o de cara y un montón de datos más que aportan contexto y, en definitiva, arrojan luz al dato que se nos presenta muchas veces como el más relevante.

StatsBomb lanzó hace dos años un modelo llamado On Ball Value (OBV, a partir de ahora) que se encarga de medir el impacto de distintas acciones (pases, conducciones, tiros y acciones defensivas) a la hora de marcar o conceder un gol de forma separada. Es un modelo muy útil para ver realmente el impacto de cada acción y que ahonda en los detalles, que son los que terminan marcando la diferencia a la hora de hablar de los datos.

Los dos jugadores de la Premier League que, según OBV, añaden más valor desde el pase son dos laterales: Kieran Trippier y Trent Alexander-Arnold. Curiosamente, estos dos jugadores están por debajo del 75 % de acierto en el pase, cuando la media de la Premier entre todos los laterales es del 78 %. Es decir, hablamos de dos jugadores de mucho volumen de intervenciones, pero con un acierto por debajo de la media, algo que es altamente positivo tanto para Newcastle como para Liverpool.

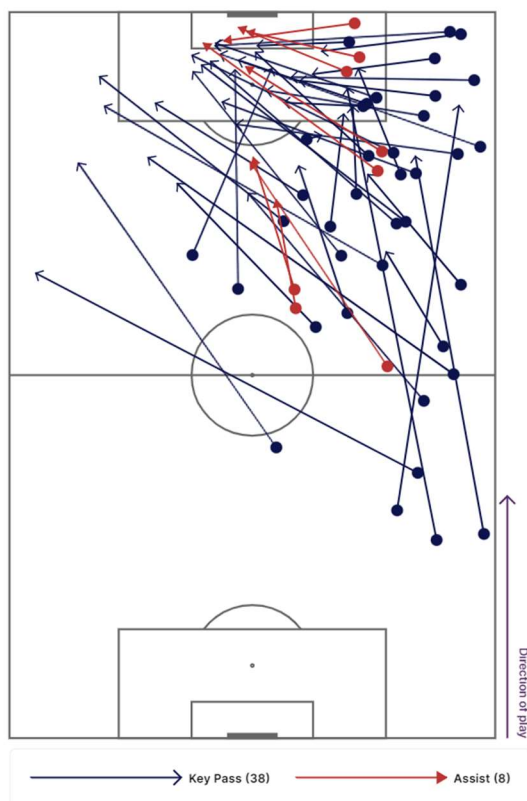
Figura 7

### Key Passes and Assists

Open Play

Passes

46 Passes



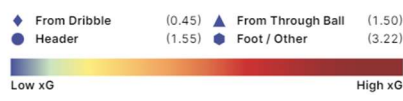
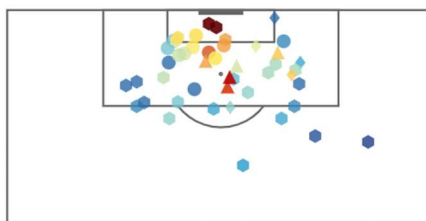
### Trent Alexander-Arnold

Liverpool

Premier League, 2022/2023

Resulting Shots

46 Shots, 8 Goals (6.73 xG, 0.15 xG/shot)



StatsBomb

Fuente: StatsBomb, <https://shre.ink/2f8b>

El tipo de envíos que suele realizar Arnold, los que más valor añaden, son pases verticales, muchos de los cuales desde zonas muy alejadas de la portería rival, que están alrededor del 50 % de acierto. Klopp ha explicado muchas veces que prefiere un pase fallado a uno acertado, especialmente desde los que nacen desde el flanco y van hacia dentro, porque un pase fallado implica una segunda jugada, y el Liverpool de Jürgen Klopp se prepara enormemente para este tipo de situaciones.

Sin tener en cuenta el porqué de este tipo de pases y su utilidad, pese a que sean envíos poco efectivos, el análisis basado en un porcentaje, aunque sea muy elevado el error comparándolo con otros jugadores de su misma posición, carecerá de una mirada profunda. Lo mismo se puede aplicar a Kevin de Bruyne, uno de los mejores centrocampistas de los últimos años que tiene, pese a jugar en un equipo de mucha posesión y efectividad como el Manchester City, uno de los porcentajes más bajos de acierto en el pase.



Desde que Pep Guardiola es entrenador del Manchester City, su equipo ha sido siempre el que ha liderado la tabla en el porcentaje de acierto en los pases entre todos los equipos de la Premier League. En un equipo que promedia un 88 % de acierto, De Bruyne suma apenas el 74 %, uno de los registros más bajos para un centrocampista. En la Premier, solo hay tres centrocampistas que en el curso 22/23 hayan tenido peor porcentaje que el belga: Soucek, Dramsgaard y Jensen, jugadores de West Ham y Brentford respectivamente, dos equipos con poco acierto en el pase, algo que tiene correlación directa con los nombres propios. Pero no en De Bruyne.

El propio jugador ha explicado públicamente que él nunca se fija en ese dato, ni siquiera le da importancia. De Bruyne ha liderado, en cinco de los siete cursos en los que ha estado Guardiola, la tabla de xA en su equipo, es decir, de asistencias esperadas: es el futbolista que más volumen genera, el más valioso. Por lo tanto, tomar el dato de su porcentaje de acierto en pases en frío no sirve de nada sin el contexto que le rodea, sino que genera una narrativa negativa que no se ciñe a la realidad del futbolista.

Otro apunte interesante para analizar a un futbolista es tener en cuenta que el dato esté ajustado a la posesión de su equipo. ¿Por qué? No es lo mismo que un futbolista realice 2 entradas buenas por partido si su equipo tiene el 47 % de posesión que realizar 1,6 con el 64 %. El segundo futbolista tiene un registro más alto si tenemos en cuenta que su equipo está mucho más tiempo con el balón. Lo mismo sucede con otro tipo de eventos como las presiones, intercepciones o despejes, estadísticas sin balón que, sin ese contexto, no dicen mucho. Es lógico que, defensivamente, los jugadores que están en equipos de mucha menos posesión tengan más volumen en estos datos, pero no implica que sean mejores en estos apartados, solo que tienen más tiempo para realizarlos.

El analista de datos Jesús Lagos (@Vdot\_spain en Twitter) es uno de los grandes referentes a la hora de generar contenido desde el dato dentro del mundo del fútbol. En un hilo de Twitter (2022), se cuestionaba un dato que había aparecido en televisión durante un partido del FC Barcelona: Gavi había corrido 13 km durante el choque, y Jesús se cuestionaba si esto era mucho o no, dependiendo de muchos factores. ¿A cuánta intensidad los corría? ¿Cuál es el contexto del partido? A partir del dato, se va al partido y se analiza, viendo cada acción y entendiendo si lo recorrido era o no necesario, si los kilómetros derivan de un mal posicionamiento previo, etc. Pero, de nuevo, el dato por sí solo muestra algo que brilla y es goloso, pero que una vez lo absorbes no te dice mucho si no lo acompaña la pregunta pertinente.



Figura 8: Tuit de José Lagos



Fuente: captura de pantalla de Twitter (<https://bit.ly/3PvOLiS>)

En este módulo estamos abordando la necesidad de cuestionar siempre al dato, y darle un contexto para que pueda generar un discurso y no solo un titular. Lo segundo probablemente venda más, pero lo primero es lo que nos debe preocupar. En un universo repleto de titulares, de datos que llaman la atención que se usan para alabar o sentenciar a discreción, cuestionar al dato se vuelve una necesidad imperiosa para generar un contenido de mayor profundidad. Entender que el dato no genera un valor añadido por sí solo, y que es misión de cada uno aportar el valor añadido que merece, o darle un contexto para que su contenido sí responda cuestiones básicas.

El análisis táctico, sumado al uso de la estadística avanzada le da otra dimensión al trabajo, porque permite responder distintas preguntas, además de ahondar en lo que el ojo detecta. Es un complemento, nunca puede sustituir la dieta, sino ayudarla para lograr objetivos más ambiciosos, tanto a la hora de analizar a un equipo como a un jugador en concreto. ¿Dónde se pueden conseguir los datos? StatsBomb es una herramienta profesional de pago que trabaja para muchísimos clubes de primer nivel y con distintos medios de comunicación para ofrecer mayor profundidad en su contenido. Pero, tiene datos en bruto en abierto, que pueden ser usados como campo de pruebas. La mayoría

de datos que incluyen *tracking* sí son de pago, pues son muy valiosos, pero la mayoría de datos de *eventing* se pueden encontrar en distintas plataformas, una de ellas es Whoscored.com. También está Sofascore.com, que ofrece estadísticas de *eventing* de muchas competiciones y es gratuita.

Figura 9: Whoscored.com

Estadísticas de Fútbol

Estadísticas del Campeonato Detalladas | Todas las Estadísticas de Equipo | Comparación

Estadísticas de Equipo

Resumen | Defensivo | Ofensivo | Detallado

General | Local | Visitante

Equipo	Campeonato	Goles	Tiros pp	Disciplina	Posecion%	AciertoPase%	Aéreos	Rating
1. Bayern Munich	Bundesliga	92	18.5	45 <b>3</b>	64.3	87.1	13.9	6.94
2. Manchester City	Premier League	94	15.8	44 <b>1</b>	65.2	89.2	11.6	6.90
3. Borussia Dortmund	Bundesliga	83	16.6	66 <b>0</b>	58.3	84.4	12.6	6.88
4. Paris Saint-Germain	Ligue 1	89	15	53 <b>5</b>	60.9	90.5	6.6	6.88
5. Real Madrid	LaLiga	75	17	68 <b>3</b>	61.2	90.0	8.9	6.86
6. Barcelona	LaLiga	70	15.1	78 <b>6</b>	64.8	88.1	12.3	6.85
7. Arsenal	Premier League	88	15.6	52 <b>0</b>	59.7	85.4	12.9	6.81
8. Napoli	Serie A	77	16.2	47 <b>1</b>	62.1	87.8	12.8	6.81
9. Newcastle	Premier League	68	15	62 <b>1</b>	52.2	79.8	14.7	6.79
10. Manchester United	Premier League	58	15.6	78 <b>2</b>	53.8	82.3	12.3	6.77
11. Lyon	Ligue 1	65	13.7	59 <b>3</b>	58.0	85.1	12	6.77
12. Liverpool	Premier League	75	15.9	57 <b>1</b>	60.6	84.2	12.7	6.75

Fuente: captura de pantalla de Whoscored (www.whoscored.com)

Cada año, el mundo del análisis de datos añade nuevas herramientas y modelos, perfecciona los que ya están instaurados y deslegitima algunos que se creían intocables. En una industria que avanza tan rápido, es importante estar al día y ver cuáles son las últimas novedades y tendencias, ya que aunque no seas analista de datos profesional, sí son una herramienta potente que te van a servir en tu día a día, a la hora de generar contenido.



## Referencias

**Lagos, J.** [Vdot\_Spain]. (11 de octubre de 2022). *Se está hablando del partido de Gavi contra el Celta donde corrió más de 13km. El dato por sí solo no aporta mucho, necesitamos más información para saber si "es mucho o poco" Veamos unos datos de Gavi que son más importantes que esos 13km.* <https://t.co/JJdbTxJACN> [publicación de Twitter]. [https://twitter.com/Vdot\\_Spain/status/1579842579101016066](https://twitter.com/Vdot_Spain/status/1579842579101016066)

**Statbomb** (2022). *Los mejores datos de fútbol.* <https://statsbomb.com/es/>

**StatsBomb** (2019). *La historia de los goles esperados (xG) y aplicaciones prácticas.* <https://statsbomb.com/es/articulos/futbol/que-son-los-goles-esperados-xg/>

**StatsBomb** (2020). *Why aren't Manchester City better?* <https://statsbomb.com/articles/soccer/why-arent-manchester-city-better/>