

Módulo 4. Datos para el análisis de la presión alta en el fútbol

Introducción

El objetivo principal de este módulo 4 es conocer qué datos a nivel táctico defensivo son útiles para medir la capacidad de presionar alto de un equipo. En este módulo, pretendemos conocer como los datos pueden ayudar a los entrenadores a tomar decisiones óptimas a nivel defensivo.

Para ello, vamos a hacer una breve introducción sobre en qué consiste el *big data* aplicado al fútbol, qué proveedores de datos existen, de cuáles de ellos podemos extraer datos defensivos de presión, qué datos nos van a servir para medir la capacidad de presionar alto de un equipo, y cómo podemos interpretarlos.

Para esto, vamos a comentar los siguientes puntos:

1. Introducción al *big data* en el fútbol.
2. Principales proveedores de datos existentes en fútbol. ¿De dónde podemos obtener datos defensivos de presión?
3. ¿Qué datos nos van a servir para medir la capacidad de presionar alto de un equipo?
4. Claves para una buena interpretación del dato.



Unidad 4.1. Introducción al *big data* en el fútbol

El uso del *big data* y la inteligencia artificial en el fútbol ha crecido mucho en los últimos años y ambos se están convirtiendo en herramientas muy importantes para obtener mejores resultados y optimizar el rendimiento de los equipos y deportistas profesionales.

La aplicación del *big data* deportivo en el fútbol busca generar una ventaja respecto al adversario, tanto en la táctica y la estrategia a utilizar en los partidos como en el mercado de fichajes o en el cuidado de los futbolistas.

En un partido de fútbol se capturan alrededor de ocho millones de datos sincronizados con el tiempo de partido, que registran la información generada por los jugadores y el balón en el espacio de juego. Esta información —que no se ve a simple vista, ya que el ojo humano solo es capaz de retener el 30 % de lo que pasa durante el juego— es almacenada y tratada, a través de las herramientas de *big data*, para ser utilizada y consultada para tomar decisiones económicas, tácticas, físicas, etc.

Por eso, los clubes y entrenadores usan estas tecnologías para mejorar la toma de decisiones. Y mucho más en el fútbol profesional, ya que dispone de mayores recursos para poder acceder a este tipo de herramientas y de la posibilidad de contratación de personal que sepa utilizarlas y aprovecharlas de una manera más eficiente. (Web-Fútbol, s. f., <https://www.web-futbol.com/big-data-en-el-futbol/>).

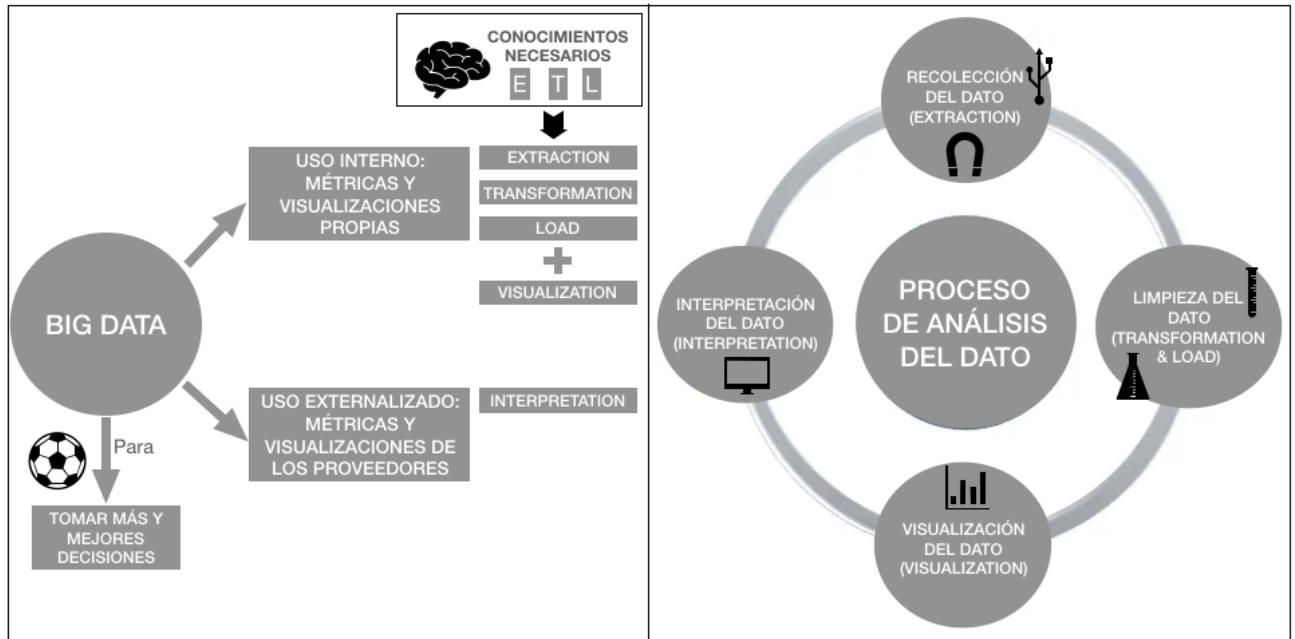
¿Pero cómo un cuerpo técnico puede usar el *big data*? Los entrenadores y analistas pueden utilizar los datos que genera el *big data* en fútbol de dos maneras:

- 1) Uso externalizado: se contratan los servicios de uno o varios proveedores de datos como StatsBomb, Opta, Instat, Wyscout, etc., los cuales, en sus plataformas de análisis, ofrecen métricas y visualizaciones de datos cerradas y con su propio formato, que se pueden consultar, descargar y utilizar para sus análisis y presentaciones. Esta manera de utilizar los datos no necesita de muchos conocimientos de *big data*.
- 2) Uso interno: genera sus propias métricas y visualizaciones a partir de datos recogidos de varios proveedores. Los proveedores antes mencionados también dan la opción de descargar los datos crudos que generan (*raw data*) desde sus API (*Application programming interface*), para luego tratarlos con programas de



lenguaje de programación (R o Python) y darle una visualización propia con programas como Power BI o Tableau. Esta manera de generar información y su visualización posterior demanda un alto conocimiento de los procesos relacionados con el *big data* y la ETL de los datos (extracción, transformación y carga).

Figura 1: Uso del *big data* por parte de un cuerpo técnico



Fuente: elaboración propia.

Las consideraciones a tener cuenta para los cuerpos técnicos que no tengan tantos recursos económicos para poder contratar los servicios de un proveedor son:

- a) Por un lado, existen plataformas de ligas profesionales y federaciones que facilitan el acceso gratuito a imágenes y datos de *tracking* y *eventing* (*raw data*), siempre y cuando el club en cuestión compita en ligas profesionales como La Liga española (Mediacoach), la Bundesliga alemana (Match Facts) o la Premier League inglesa (DVMS).
- b) Por otro lado, existen numerosos *websites* de consulta (Understat, Fbref, WhoScored, Transfermarkt, etc.) que ofrecen de manera gratuita el acceso a datos de *eventing* que podremos descargar para trabajarlos a nuestra manera mediante el *web scraping*.

En definitiva, lo más importante y lo que va a marcar la diferencia para hacer un buen uso del *big data* aplicado al fútbol es tener tanto conocimientos de fútbol como conocimientos



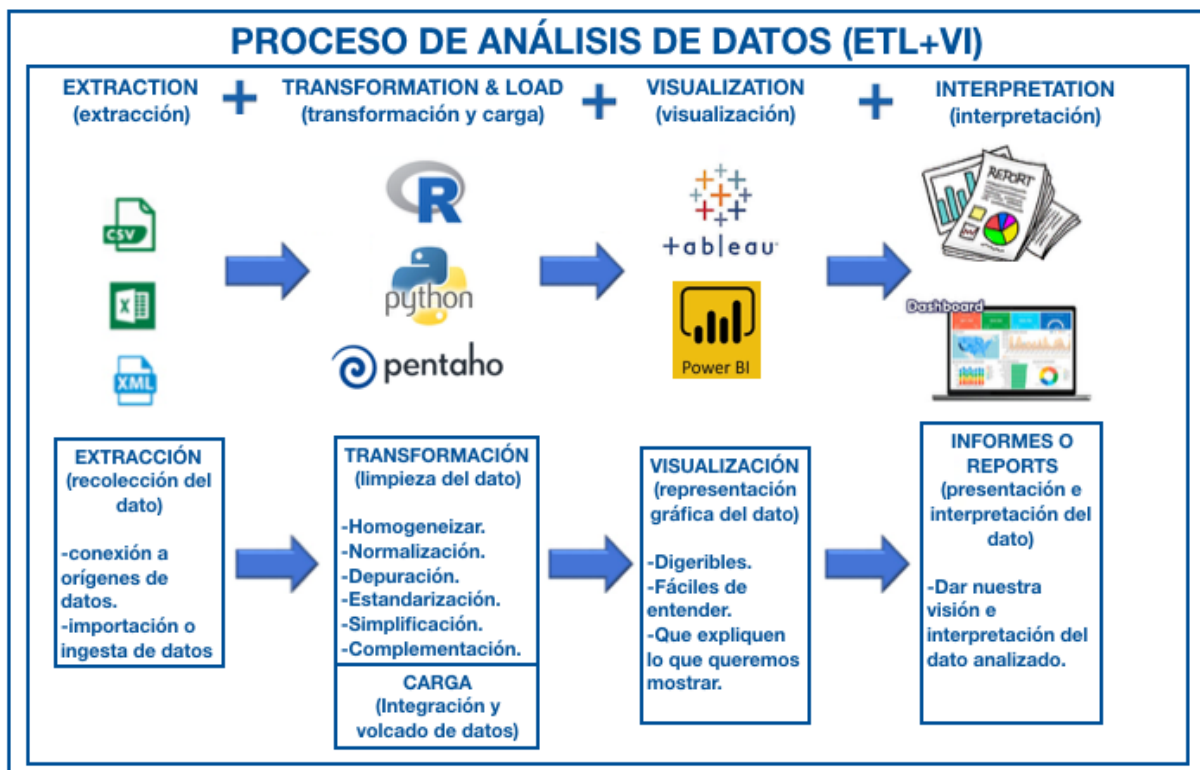
sobre el *big data* y los procesos de tratamiento del dato ETL, a los que yo añadiría la visualización y la interpretación del dato:

- 1) Extraction (extracción): recolección, importación o ingesta del dato.
- 2) Transformation (transformación): limpieza y homogeneización del dato.
- 3) Load (carga): integración y volcado de datos.
- 4) Visualization (visualización): representación gráfica del dato.
- 5) Interpretation (interpretación): presentación e interpretación del dato.

Para poder aplicarlo en nuestra rutina de trabajo diario y para que sea realmente una herramienta útil para generar un conocimiento aplicable a nuestra labor como entrenadores y/o analistas, debemos tener conocimientos sobre el proceso de análisis de datos. El *big data*:

Hace referencia al tratamiento y análisis de grandes cantidades de datos con el objetivo de extraer valor de esa información. Esta gestión y análisis de datos no pueden ser tratados mediante técnicas convencionales, sino mediante técnicas tales como el reconocimiento de patrones (*machine learning*) o análisis predictivos para vaticinar los cambios en los volúmenes de datos. (Oviedo, 2020, p. 6).

Figura 2: ¿Cómo usar el *big data*?



Fuente: elaboración propia.

Si queremos hacer un uso interno y propio del proceso de análisis —mediante el filtrado de la información en forma de datos que proceden de diferentes proveedores y dándole la visualización que queremos—, tenemos que tener unos mínimos conocimientos sobre los procesos de análisis del *big data* y las herramientas necesarias para su aplicación:

1. Conocimientos de extracción de datos: conocer y saber extraer de las plataformas de los diferentes proveedores los distintos formatos de archivos que contienen la información y que son importables. Estos formatos de archivo pueden ser: .csv, .excel, .xml, y .Json. Es importante conocer los proveedores y los datos que nos interesan de cada uno de ellos.
2. Conocimientos de transformación y carga de datos: conocer y saber usar las herramientas de lenguaje de programación, como R y Python, o las de integración de datos, como Pentaho, que son las que más se están utilizando actualmente para extraer, transformar y cargar los datos desde el formato propuesto por los proveedores al formato que más nos interese:

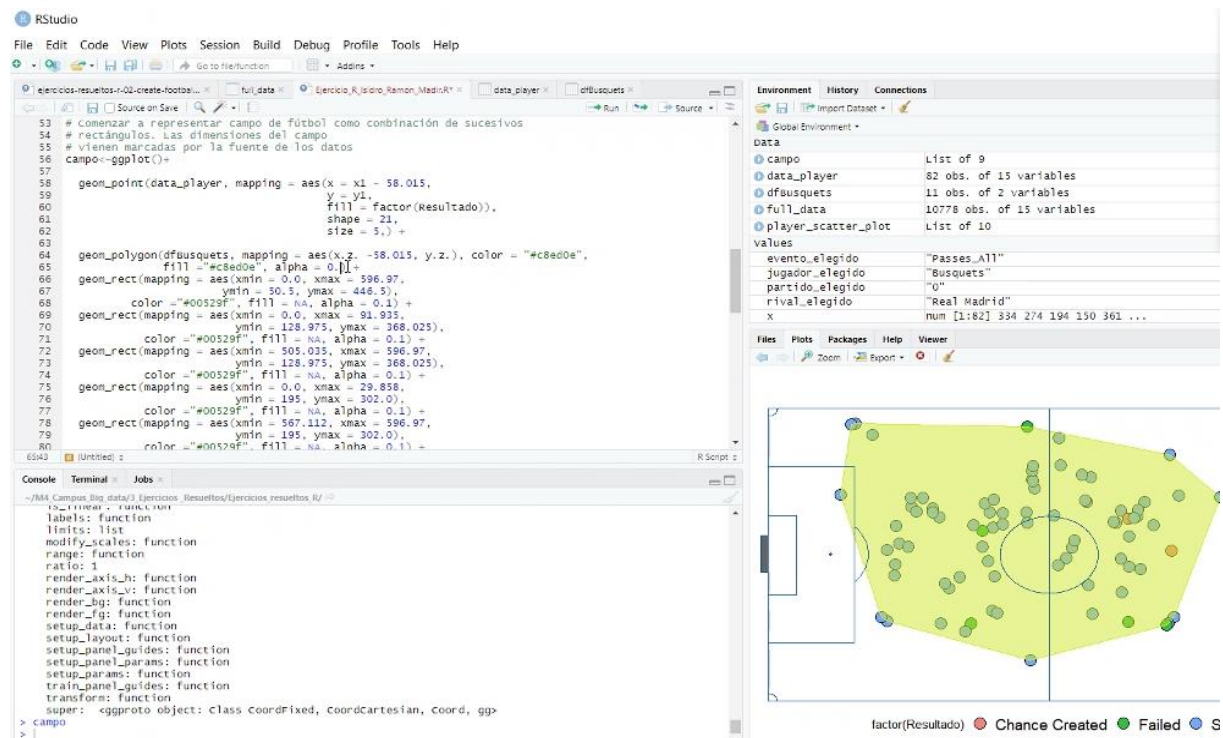
2.1. R: es un lenguaje de programación gratuito. Este entorno es comúnmente utilizado para la computación estadística y gráfica y funciona en sistemas Unix, Linux, Windows y MacOS.

Las características y diferentes aplicaciones de R lo convierten en una herramienta básica para los analistas de datos. Se utiliza en todas las fases de análisis de datos:

- Adquisición de los datos de las fuentes disponibles: bases de datos, archivos de texto, etc.
- Preparación de los datos: eliminación de duplicados, datos incorrectos, valores extremos, etc.
- Análisis de los datos: construcción de modelos predictivos, de clasificación, de agrupamiento, etc.
- Comunicación de los resultados: realización de informes para presentación de los resultados y conclusiones.
- Aplicación de los resultados obtenidos: por ejemplo, utilización de modelos predictivos desarrollado para, en función de una serie de datos históricos (datos de entrenamiento y test del modelo), predecir ciertas salidas. (Unir, 2019, <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>).



Figura 3: Ejemplo de *interface* en R



Fuente: captura de pantalla del *software* RStudio.

2.2. Python: es un lenguaje de programación gratuito de alto nivel, que utiliza el conocido "lenguaje de *scripting*" y que se funciona para casi todo. La gran mayoría de las aplicaciones informáticas y plataformas web dependen de su lenguaje, ya que es utilizado por la mayoría de desarrolladores, científicos de datos, ingenieros de *software* e incluso *hackers* debido a su versatilidad, flexibilidad y características orientadas a los objetos. El lenguaje Python puede ejecutarse en diferentes sistemas operativos como Unix, Linux, MacOS y Windows. Es importante aprender a manejarlo sobre todo en los campos emergentes de la ciencia de los datos, como el análisis de datos (*big data*), la inteligencia artificial y el *machine learning*.

A través de las diversas librerías que conectan a Python con el mundo exterior, este lenguaje puede llevar a cabo las siguientes funciones:

- Interactuar con el mundo exterior, para lo cual será necesario leer y escribir datos en diferentes formatos (.csv, .xml, .excel, .Json, .txt, .dat, binario, etc.).
- Preparar los datos a través de limpieza y generando datos estructurados desde datos no estructurados.
- Transformación de los datos a través de la aplicación de operaciones matemáticas y/o estadísticas a ciertos conjuntos de datos para obtener



como resultado nuevos datos (operaciones de agregación, segregación y filtrado).

- Presentación de los datos y resultados obtenidos haciendo uso de gráficos estáticos y dinámicos (con JavaScript), mapas de localización, grafos, etc.
- Modelos de aprendizaje automático en los que se conectan los datos con algoritmos de aprendizaje automáticos —supervisados o no— para la obtención de modelos analíticos que puedan predecir —con mayor o menor grado— el futuro. (Fombella Pombal, 2020a, p.10).

Figura 4: Ejemplo de *interface* en Python

The screenshot shows a Jupyter Notebook window titled 'Untitled' with the following content:

```

In [110]: def extraer_datos_debut(df):
           df['club_debut']=df['Debut deportivo'].map(str) #esta función es para que no falle el apply, pasando la columna a str.
           df['club_debut']=df['club_debut'].apply(lambda x: x.split('(')[-1].split(')')[0])
           df['club_debut']=df['club_debut'].apply(lambda x: x.replace('2013','9999').replace('2011','9999').replace('2009','9999'))
           df['club_debut']=df['club_debut'].apply(lambda x: x.replace('2002-04','C. D. Tenerife "B" '))
           df['club_debut'].replace('nan','9999',inplace = True)
           df['club_debut'].replace('9999',np.nan,inplace = True)

           df['fecha_debut']=df['Debut'].map(str)
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.split(' de ')[-1])
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('[1] ','').replace('[2] ','').replace('1 ','').replace('[A] ',''))

           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.split('1/')[1])
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('2018 (República Centroafricana)','2013'))
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('-', '9999').replace('nan', '9999'))
           df['fecha_debut'].replace('9999',np.nan,inplace = True)

           df['club_debut']=df['Debut deportivo'].map(str)
           |

           # No modifique las siguientes líneas
           assert(isinstance(df, pd.DataFrame))
           return df

In [111]: df_5=raw_data[['Debut','Debut deportivo']].copy()
           extraer_datos_debut(df_5)
           df_5

Out[111]:
           Debut                Debut deportivo  club_debut  fecha_debut
0           11 de noviembre de 2017                NaN         NaN         2017

```

Fuente: captura de pantalla del *software* Jupyter (Pérez, 2021).

2.3. Pentaho: pertenece a la compañía Hitachi Vantara, y es una plataforma de *business intelligence* (BI) orientada y centrada en la solución de procesos, que incluye los componentes requeridos para implementar soluciones basadas en procesos como minería de datos, ETL, generación de informes, etc. Es muy utilizada para la mejora de la capacidad de análisis y de toma de decisiones, puesto que ofrece soluciones que se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrado.

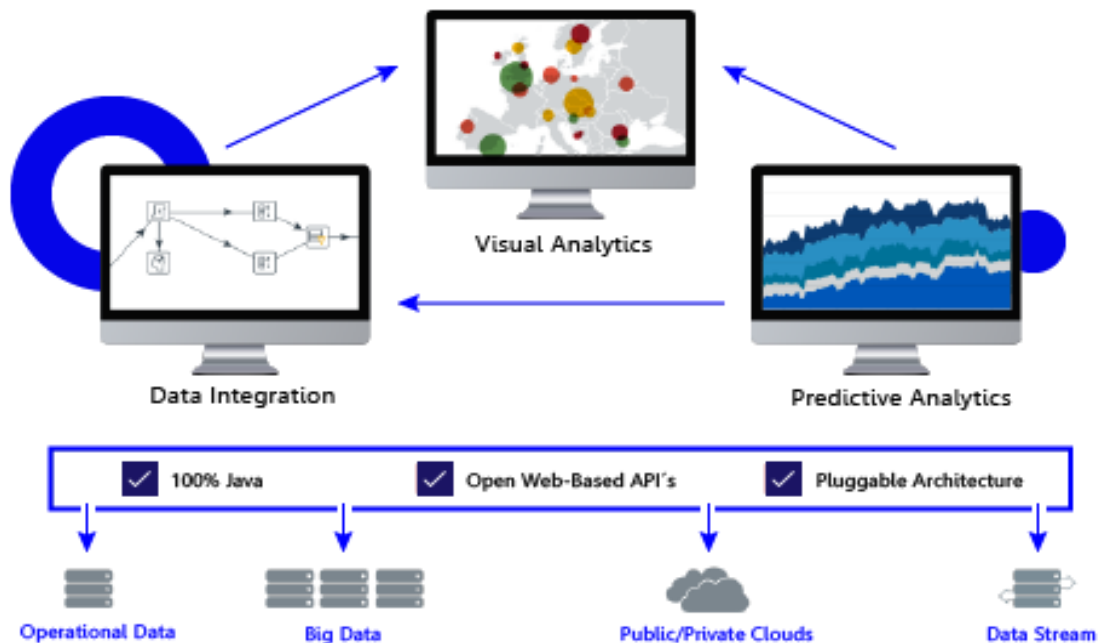
Pentaho incluye dentro de su *suite* herramientas que cubren las siguientes áreas:

- Pentaho Data Integration: herramienta de integración de datos (ETL).



- Pentaho Reporting: creación de Informes tradicionales.
- Pentaho Dashboard Designer: componente únicamente disponible en la versión Enterprise.
- Pentaho Community Dashboard Editor: *plugin* para la creación de cuadros de mando.
- El resto de las herramientas de creación de informes y visualización se han visto reemplazadas en muchos casos por herramientas de última generación, como Tableau o Microsoft Power BI, las cuales veremos más adelante. (Fombella Pombal, 2020, pp. 42-44).

Figura 5: Herramientas de Pentaho



Fuente: [Imagen sin título sobre herramientas de Pentaho]. (s. f.). Recuperado de https://lh3.googleusercontent.com/4y-ZxvheJkzXhDva8cLQOAccgCwF1wFHRD9PP_SWnRQarn_OIyVU5Ktwc1cgrJY71bYvxQ=s130

Resumiendo, Pentaho es una herramienta muy útil para integrar y unificar datos provenientes de diferentes formatos y proveedores en un formato único, que permite una ETL lo más rápida y eficaz posible, para proceder a trabajar con las herramientas de visualización.

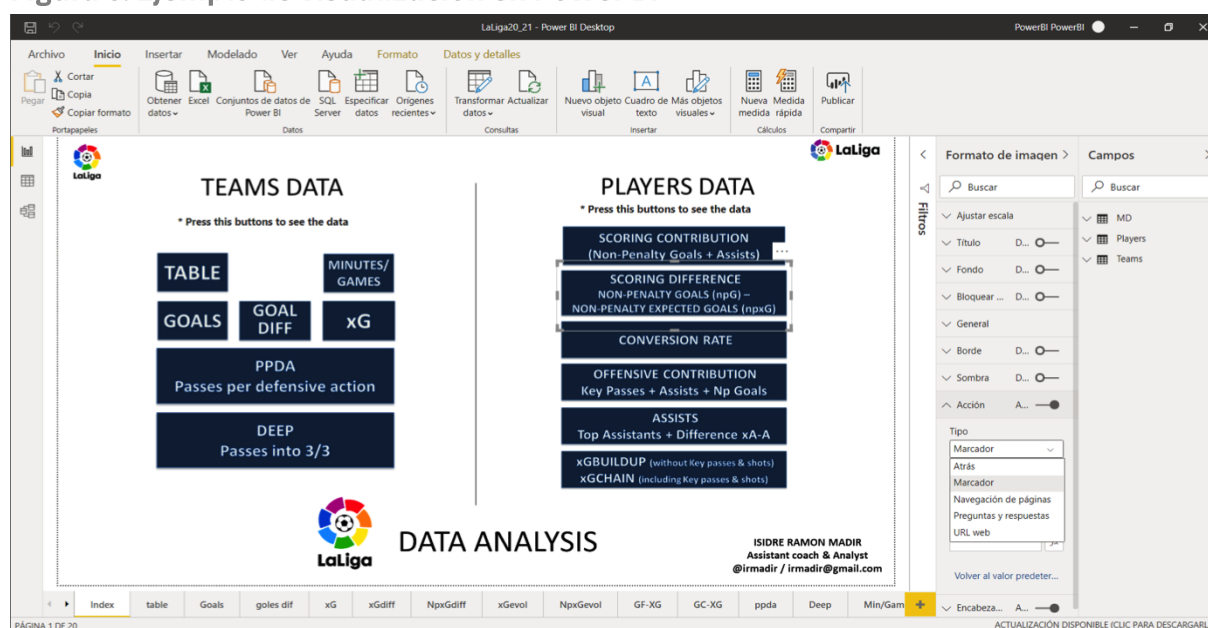
3) Conocimiento sobre la visualización de datos: conocer y saber usar las herramientas de visualización más potentes del mercado. Estas son Power Bi y Tableau:



3.1. Power BI: Microsoft Power BI es la principal herramienta de visualización dentro de Microsoft Office 365. Únicamente está disponible para sistemas operativos Windows. Su elevado número de usuarios y gran aceptación dentro del mercado radica en lo sencillo de su manejo y su bajo costo. Seremos capaces de realizar cuadros de mandos interactivos de forma muy ágil.

En cualquier entorno (empresas, clubes deportivos, etc.), resulta imprescindible unificar toda la información disponible, se encuentren en la nube o localmente, en torno a una misma plataforma. Para ello, Power BI es una herramienta ideal para aplicar los conocimientos del *business intelligence* (BI) sobre la explotación de datos en tiempo real a partir de diferentes fuentes, así como para la creación de informes y paneles y ofrecer, posteriormente y de una manera simple, visual e intuitiva, los resultados de todos aquellos elementos que nos interese analizar y compartir entre varios profesionales de un mismo club.

Figura 6: Ejemplo de visualización en Power Bi



Fuente: captura de pantalla del *software* Power BI Desktop (Microsoft, 2021).

3.2. Tableau: es una herramienta de visualización de datos potente, utilizada en el área de la inteligencia de negocios (*business intelligence*). Simplifica los datos en bruto en un formato muy fácil de entender.

La esencia de Tableau es simple y, a la vez, muy relevante: ayudar a las personas y empresas a ver y comprender todos sus datos. Lo consigue

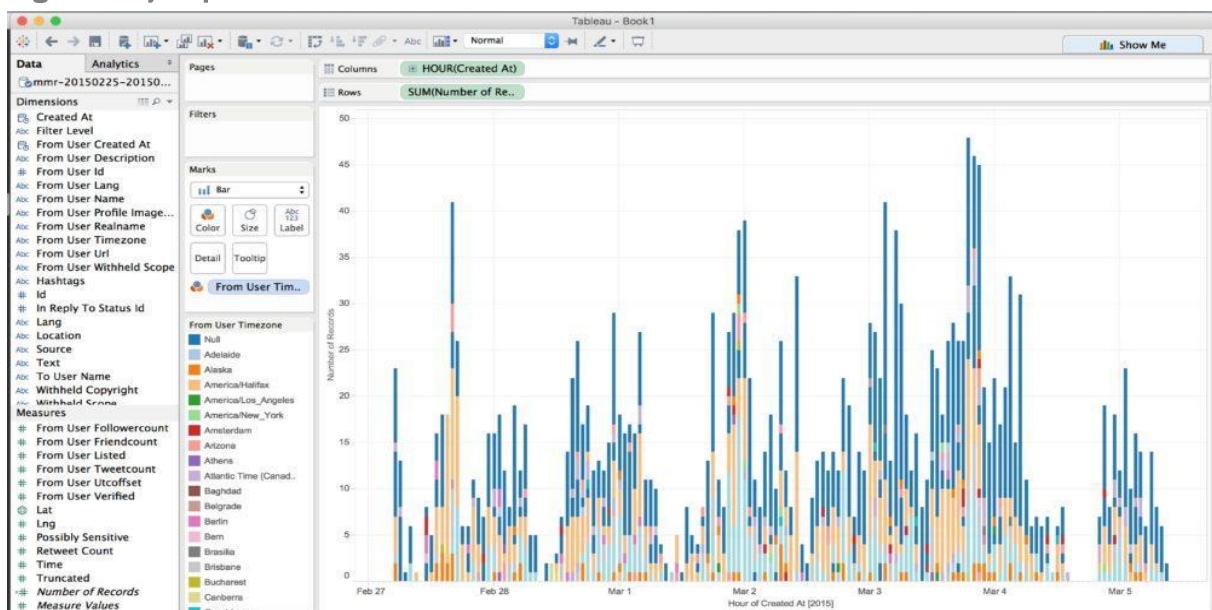


ofreciendo una selección de herramientas útiles e intuitivas de inteligencia de negocios:

- Escritorio (Tableau Desktop): conecta y analiza los datos e informaciones.
- Servidor (Tableau Server): gracias al servidor, podremos colaborar de forma segura y compartir la información a partir de los datos que hayamos subido a través de Tableau Desktop (la versión de escritorio del *software*).
- En línea (Tableau Online): se trata de una versión de Tableau Server alojada en la nube. De esta forma, podremos acceder a nuestros datos sin necesidad de tener que pasar por un tedioso proceso de instalación. (SoftwarePara, s. f., <https://softwarepara.net/tableau/>).

A través de funciones simples como la de arrastrar y soltar, cualquier persona puede acceder y analizar de forma sencilla datos e, incluso, crear informes y compartir esta información con otros usuarios.

Figura 7: Ejemplo de visualización en Tableau



Fuente: De Juana, 2019, <https://www.muycomputerpro.com/2019/06/11/tableau-asi-es-la-empresa-por-la-que-salesforce-ha-pagado-mas-de-13-000-millones>

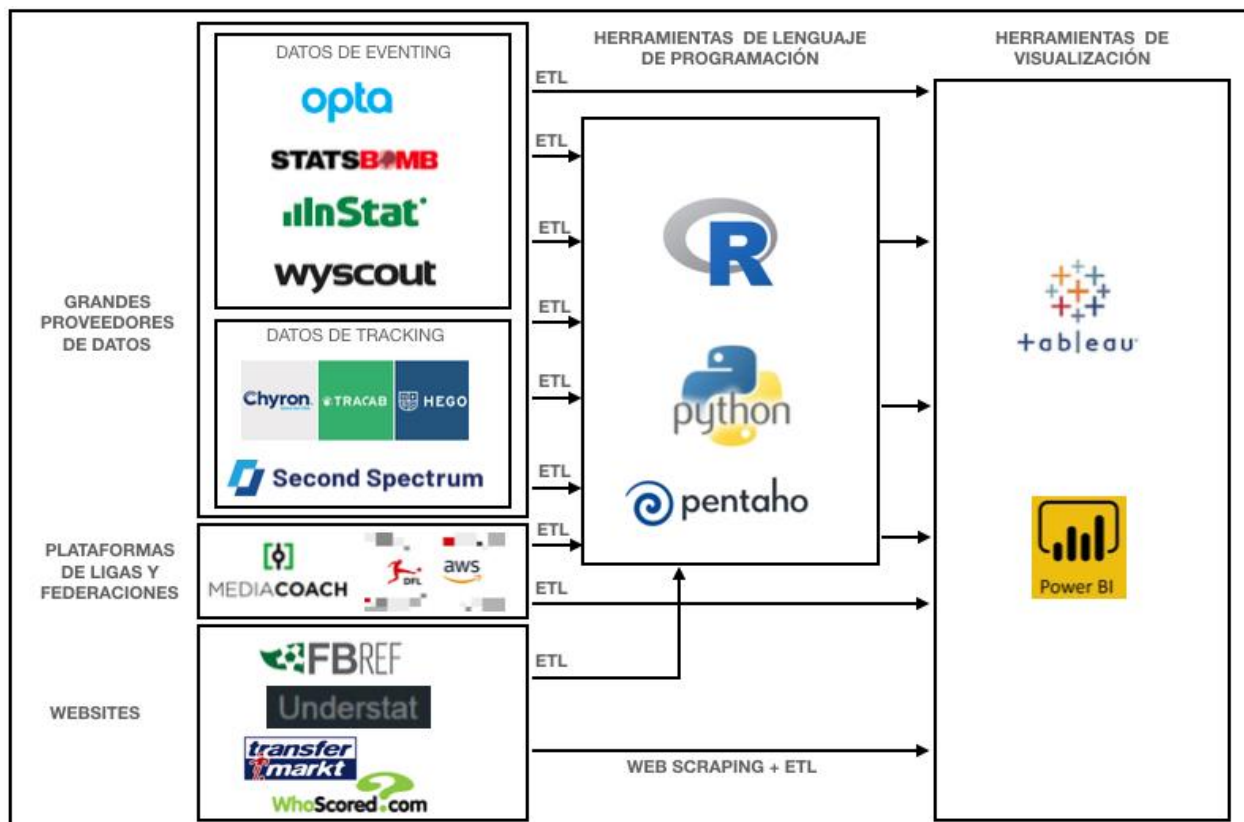


Unidad 4.2. Principales proveedores de datos en fútbol. ¿Dónde podemos obtener datos defensivos de presión?

Siguiendo al profesor del máster en *big data* deportivo de la UCAM, David Fombella Pomba, vamos a enumerar los principales proveedores de datos y las diferentes herramientas de programación y visualización que los analistas están utilizando en la actualidad.

Hay dos grandes proveedores de datos que, actualmente, monopolizan el provisionamiento de datos de *eventing* a los clubes, las plataformas de ligas nacionales europeas (Mediacoach, etc.) e incluso las webs especializadas en fútbol. Estamos hablando de Opta y StatsBomb. También, hay otros proveedores conocidos como InStat, Wyscout, así como Tracab o Second Spectrum (estos dos últimos se dedican a proporcionar datos de *tracking*):

Figura 8: Principales proveedores de datos en fútbol



Fuente: elaboración propia.

1. Opta: la marca Opta pertenece al grupo Perform, un grupo líder dedicado al contenido y medios deportivos digitales. Esta empresa se dedica a la recopilación, recepción y almacenamiento de datos de muchos y variados deportes, para distribuir esta información a los clubes que trabajan con ellos. Aunque, también, funcionan con otras vías de negocio (televisiones, casas de apuestas, periodismo deportivo, etc.).

Por un lado, Opta posee un conjunto de *widgets* para insertar dentro de página web. Es bastante frecuente ver este tipo de contenidos obtenidos en periódicos digitales pues incluyen desde conceptos básicos hasta resultados a métricas e indicadores avanzados. En lo que al fútbol se refiere, algunos de estos *widgets* disponibles permiten acceder a visualizaciones de métricas como los ya conocidos *expected goals* (xG) —goles esperados—, las zonas defensivas, los mapas de calor (*heatmap*), comparaciones de equipos y jugadores (*player and team comparison*), etc.

Figura 9: Opta *widget* de zonas defensivas (*defensive zones*)

Zonas defensivas incluye un polígono con la zona de las **acciones defensivas de cada jugador (Faltas, Entradas, Intercepciones, Tiros Bloqueados y Despejes)**



Fuente: Fombella Pombal, 2020b, p. 21.

Por otro lado, los *feeds* de datos de Opta, tanto de *eventing* como de *tracking*, con sus datos estadísticos descargables desde su API están disponibles para que los clubes los usen a su antojo, ya sea para análisis de rendimiento deportivo como para *scouting* o servicios de comunicación y prensa. Por ejemplo, el F24 (Event Detail Feed) “es uno de los paquetes de datos de Opta más utilizado, ya que incluye variedad de información y matices” (Fombella Pombal, 2020b, p. 31).

Figura 10: Opta Feed F24 (Event Detail Feed)

```
<Event id="1970142897" event_id="296" type_id="16" period_id="1" min="25" sec="14" player_id="20664"
team_id="43" outcome="1" x="78.4" y="38.2"
timestamp="2011-08-21T16:25:53.299" last_modified="2011-08-22T13:51:14">
  <Q id="357476634" qualifier_id="103" value="22.2" />
  <Q id="1145873833" qualifier_id="102" value="52.1" />
  <Q id="1577156338" qualifier_id="113" />
  <Q id="902661166" qualifier_id="72" />
  <Q id="174975923" qualifier_id="55" value="295" />
  <Q id="49122906" qualifier_id="29" />
  <Q id="97008017" qualifier_id="56" value="Center" />
  <Q id="934091399" qualifier_id="77" />
  <Q id="18684958" qualifier_id="136" />
  <Q id="65794542" qualifier_id="18" />
  <Q id="924818303" qualifier_id="160" />
  <Q id="994511953" qualifier_id="118" />
</Event>
```

Ahora nos toca descifrar la información incluida dentro del evento y sus calificadores:

Evento

- **id** → Evento número 296 del partido
- **player_id** → 16 Gol
- **team_id** → 20664 D. Silva en el **Feed 7** está la correspondencia entre id y nombre.
- **team_id** → 43 (Manchester City)
- **period_id** → 1 (1ª parte)
- **minuto y segundo** → 25:14
- **x e y del campo** → 78,4 y 38,2
- **outcome** → 1 Gol

Fuente: Fombella Pombal, 2020b, p. 32.

2. StatsBomb: es una compañía del Reino Unido, de reciente creación, con un equipo global de profesionales dedicados y experimentados con sede en Europa, Egipto y Estados Unidos. Rápidamente, se ha consolidado junto a Opta como una empresa líder en el campo de la analítica de fútbol.

StatsBomb, que cubre más de cuarenta ligas en todo el mundo y cuenta con los datos más precisos y granulares del mercado, proporciona a los equipos las herramientas necesarias para competir y ganar en el campo y en el mercado de fichajes.

Su plataforma de análisis, StatsBomb IQ, no solo proporciona a los analistas información sobre equipos, jugadores y ligas, sino que también ayuda a diferentes medios de comunicación a través de visualizaciones de datos personalizadas:

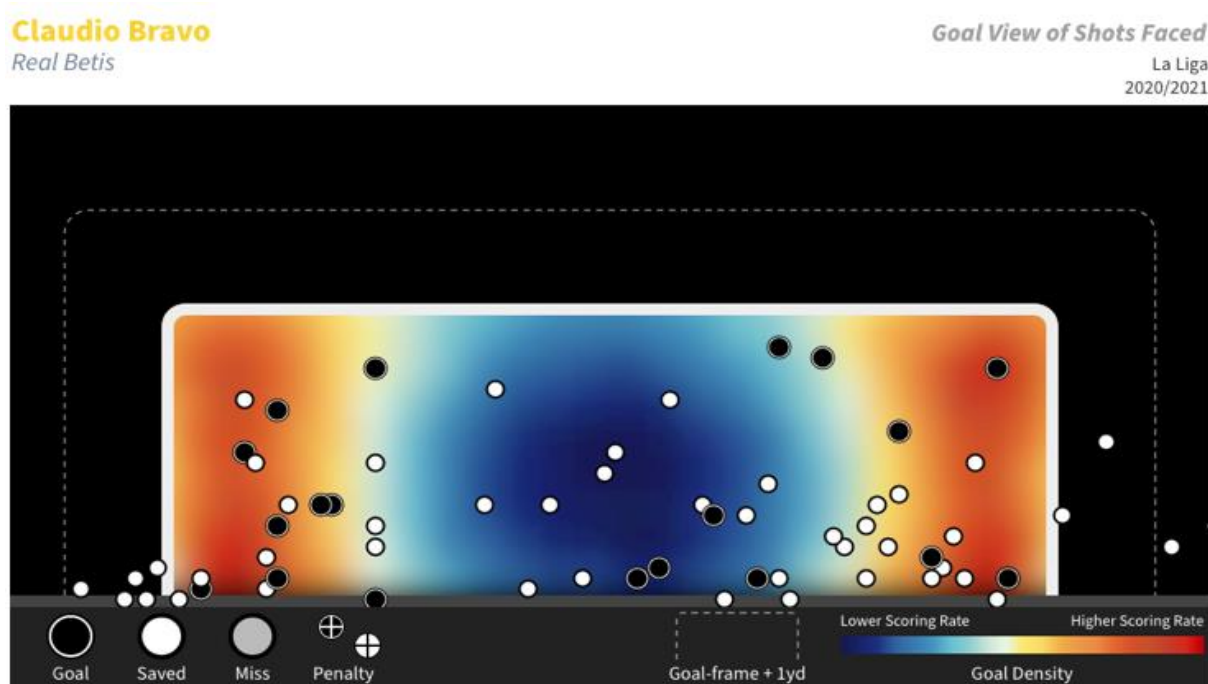
Dentro de los servicios de consultoría que ofrecen, están los siguientes:

- Evaluación de jugadores (rendimiento deportivo, información personal, informes de reclutamiento, debilidades y fortalezas del jugador, valores de mercado, etc.).
- Búsqueda de entrenadores según su estilo de juego.
- Análisis del balón parado. (Fombella Pombal, 2020b, p. 138).

Sobre todo, destaca por sus avanzadas métricas defensivas; como las presiones defensivas, contrapresiones, recuperaciones tras presión, etc., y sus nuevas métricas para medir el rendimiento del portero:

StatsBomb incluye métricas y visualizaciones revolucionarias para la evaluación de porteros. Olvidémonos de cosas tan básicas como goles/remates a puerta para evaluar el rendimiento de nuestro guardameta. StatsBomb incluye métricas de posicionamiento del portero en la línea de puerta, rendimiento atrapando centros laterales, contribución en ataque, actividad defensiva, pases completados bajo presión, etc. (Fombella Pombal, 2020b, p. 143).

Figura 11: Ejemplo de la métrica del portero en cuanto a tiros recibidos en StatsBomb (*shots faced*)



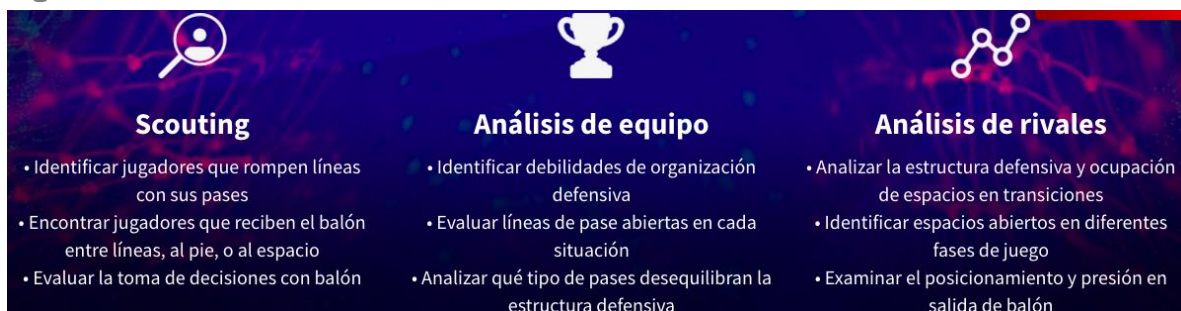
Fuente: Ovalle, 2021, <https://www.biobiochile.cl/noticias/futbol-internacional/chilenos-en-el-exterior/2021/05/27/que-dicen-los-datos-bravo-fue-el-tercer-arquero-que-mas-goles-evito-en-la-liga-la-temporada-pasada.shtml>

StatsBomb también posee una API donde los clientes pueden descargarse todos los datos para tratarlos y hacer sus propias visualizaciones, así como un repositorio en GitHub abierto y con información de algunas

competiciones para visualizar su formato de datos y realizar pruebas. También, incluye una librería de R (<https://github.com/StatsBomb/StatsBombR>) que se conecta a su API y obtiene datos. (Fombella Pombal, 2020b, p. 144).

Ya, por último, StatsBomb no para de innovar y avanzar en la métrica y ofrece datos y análisis mucho más profundos que los que podemos alcanzar con el *eventing*. Es el llamado StatsBomb 360.

Figura 12: StatsBomb 360



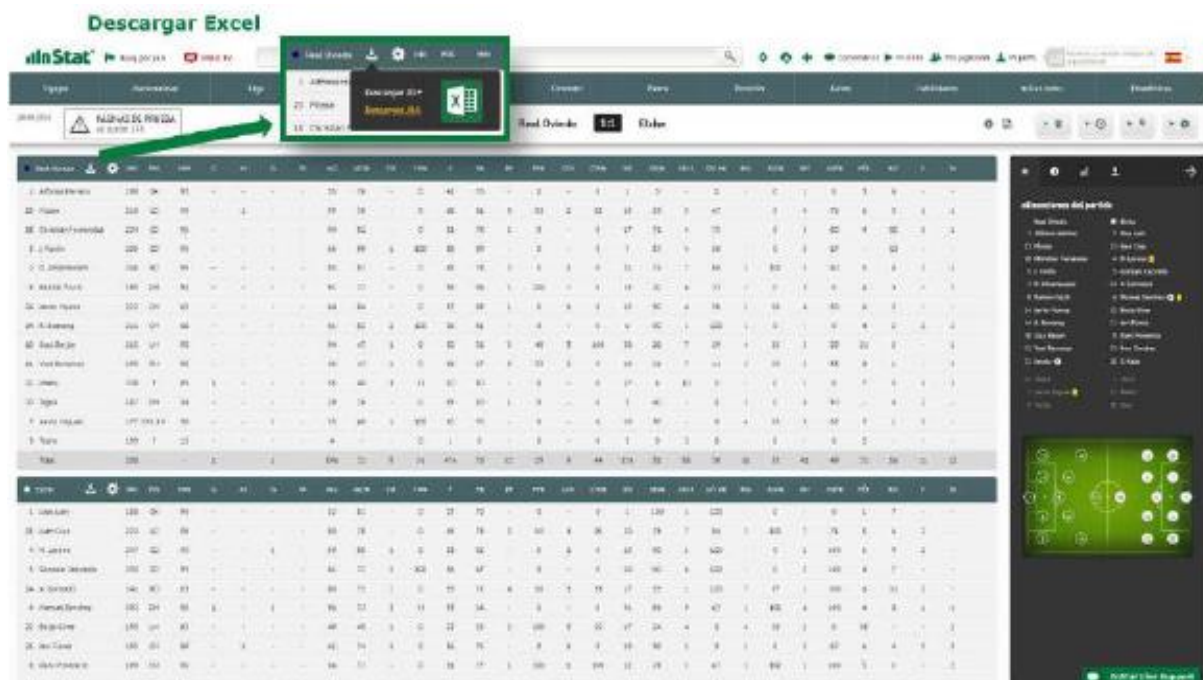
Fuente: StatsBomb, s. f., <https://StatsBomb.com/360-data/>

3. Instat y Wyscout: son otros proveedores de datos que están muy implantados en los clubes (el 90 % de los clubes profesionales tienen contratados sus servicios). Son herramientas muy parecidas y muy útiles para clubes, directores deportivos, entrenadores, analistas, entrenadores de porteros, árbitros, los propios jugadores y diversos medios de comunicación.

Son una de las mejores herramientas para la selección de talentos (*scouting and recruitment*), ya que ofrecen tanto imágenes en formato televisivo como visualizaciones de la infinidad de datos que generan sobre equipos y jugadores. También, ofrecen la posibilidad de descargar sus datos crudos (*raw data*) para su tratamiento posterior, con la posibilidad de generar sus propias visualizaciones a los clientes:

Instat ofrece estadísticas detalladas del partido en las que se incluyen las valoraciones de los jugadores (índice Instat, goles, asistencias, precisión de pases, centros y precisión en los centros, regates, disputas, duelos, pérdidas, recuperaciones, faltas y tarjetas). Desde esta página, podemos descargar un Excel con una pestaña de equipos y otra de jugadores, que nos permitirá trabajar con esta información con herramientas de visualización como Microsoft Power BI o Tableau. (Fombella Pombal, 2020b, p. 67).

Figura 13: Instat estadística detallada de partido



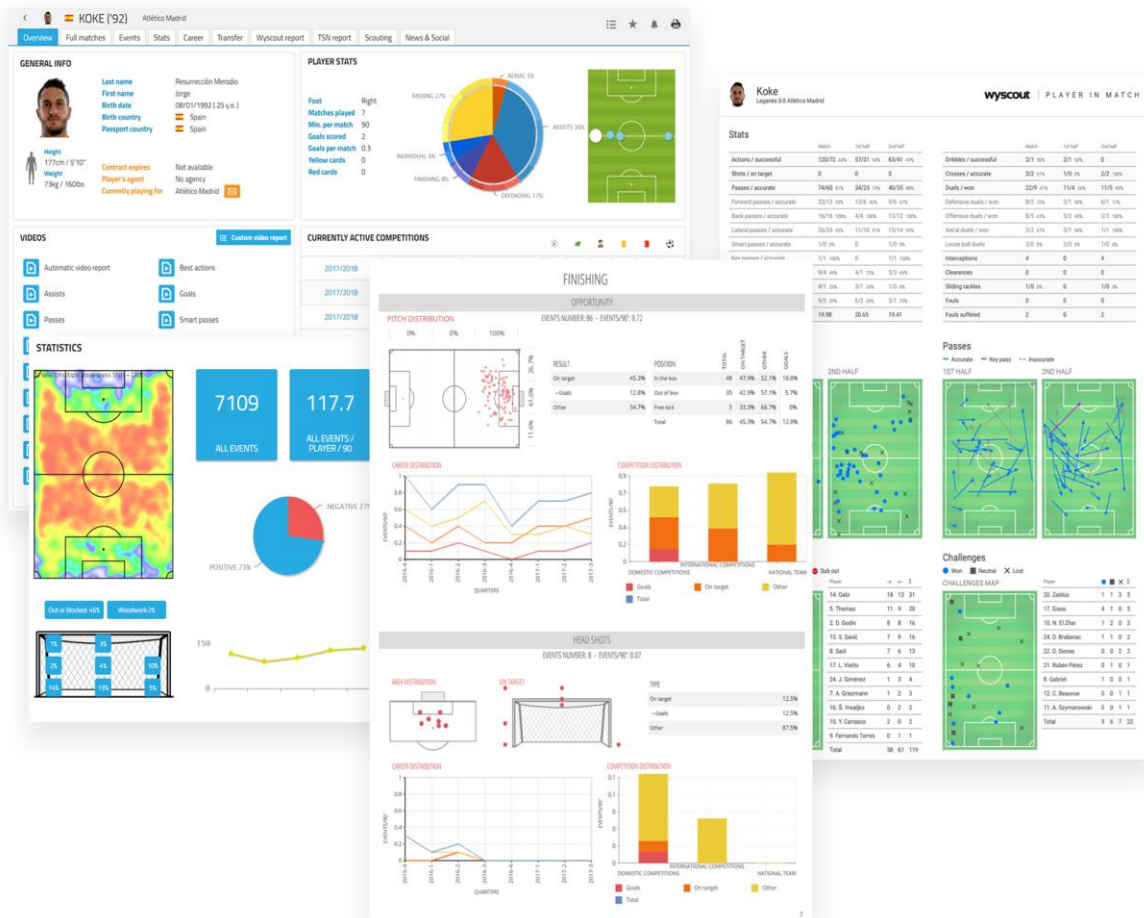
Fuente: Fombella Pombal, 2020b, p. 67.

Wyscout es una herramienta muy parecida a Instat. En 2019, se asoció con Hudl y, junto con la herramienta de análisis SportsCode, tiene dos de las herramientas más usadas por los clubs profesionales a nivel mundial.

Algunos de los instrumentos avanzados de análisis estadísticos pueden ayudar a acceder a información de todos los aspectos del juego. Estas funciones, que la aplicación Wyscout ofrece, son:

- Informes estadísticos detallados y completos descargables (datos de eventos técnicos y tácticos).
- *Rankings* de más de 200 competiciones.
- Análisis de jugadores: mediante la *“Advanced search”*, podremos descubrir nuevos talentos a través de prestaciones estadísticas.
- Con la API de Wyscout se pueden descargar todos los datos que necesite para integrar fácilmente con las herramientas BI y los *software* de análisis de video como SportsCode o Nacsport. (Fombella Pombal, 2020b).

Figura 14: Wyscout: estadística detallada de partido



Fuente: Wyscout, s. f., <https://wyscout.com/es/plataforma-futbol/>

“Los Feed Events (eventos) de Wyscout ofrecen información del partido con todos los eventos y su posición x/y en el terreno de juego... Esta base de datos histórica permite el análisis matemático avanzado y la detección de patrones” (Fombella Pombal, 2020b, p. 107).

Hasta ahora, hemos hablado de proveedores de *eventing*, pero existen otros proveedores que se dedican a la generación de datos de *tracking*.

Proveedores como Tracab y Second Spectrum capturan mediante cámaras robóticas inteligentes el movimiento de los jugadores y el balón, generando ficheros de *tracking raw data XYZ* en tiempo real.

Estos ficheros incluyen información XY de los 22 jugadores y XYZ del balón 25 veces por segundo (25 Hz), por lo que generan más de 3 millones de datos por partido:

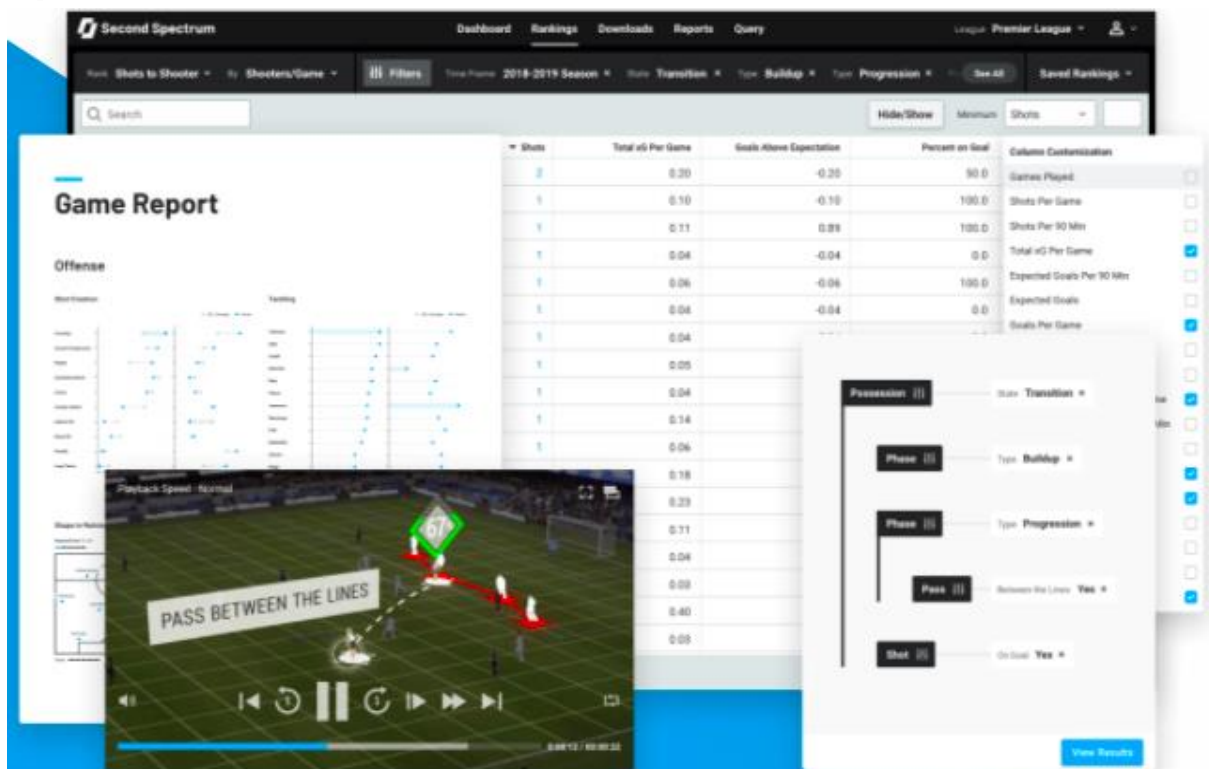
$$(90 \text{ m} \times 60 \text{ s/m} \times 25 \text{ frames/s} \times 23 \text{ (jug. + balón)}) = 3\,105\,000 \text{ datos/partido}$$



4. Second Spectrum: se trata de uno de los sistemas de seguimiento de jugadores más avanzado del mundo. Aplica técnicas de visión por computadora y aprendizaje automático de última generación para producir datos de ubicación rápidos y precisos para baloncesto, fútbol, fútbol americano y otros deportes. Es el proveedor oficial de datos de la NBA o la Premier League inglesa.

Los datos de *eventing* y *tracking* que genera impulsan nuevas formas de que los equipos y las ligas comprendan, evalúen, mejoren y creen contenido sobre su juego.

Figura 15: Tipo de informe de Second Spectrum



Fuente: Second Spectrum, s. f., <https://www.secondspectrum.com/ourwork/teams-leagues.html>

5. Tracab: es uno de los mayores proveedores de *tracking* y *eventing*. Este proporciona de datos a innumerables clubs y federaciones de fútbol del mundo (entre ellas están la liga española, alemana, holandesa, mexicana y japonesa; la Champions League; etc.). Tracab ofrece soluciones en forma de datos deportivos, videos y visualizaciones, con información muy valiosa para mejorar el rendimiento de equipos profesionales, además de permitir una gama cada vez mayor de soluciones en tiempo real altamente innovadoras y convincentes, para involucrar a los aficionados del deporte de manera más efectiva.

Un claro ejemplo lo tenemos en la liga española: la LFP (Liga de Fútbol Profesional) ofrece a los clubes la herramienta Mediacoach, que se provee de datos de *tracking* de Tracab junto con los datos de *eventing* de Opta:

Para elaborar el dossier de cada partido, que suele ocupar entre 500 y 600 páginas con métricas y datos, Mediacoach se nutre de dos proveedores. Por un lado, está Tracab, que se ocupa del *tracking*, es decir, del seguimiento continuo de cada jugador sobre el césped mediante unas cámaras instaladas en todos los estadios de La Liga Santander y La Liga 1|2|3, las cuales son sufragadas por La Liga. Este sistema se encarga de ofrecer la distancia recorrida, la velocidad máxima y media, el número de esprint, etc. Por otra parte, está el denominado *eventing*, en el que tres personas de Opta registran los pases, pérdidas, faltas, córners, etc. Todo lo que es más objetivable. (Olmeda, 2020, <https://www.abc.es/contentfactory/post/eslaliga/laliga-apuesta-por-la-tecnologia-para-hacerse-mas-competitiva/>).

Figura 16: Ejemplo de la plataforma Mediacoach de la liga española

¿Qué es Mediacoach?
Suite Productos Mediacoach

- Desktop**
Herramienta avanzada de video análisis profesional integrada con datos físicos y técnico-tácticos
- Live**
Aplicación que permite a los cuerpos técnicos monitorizar los datos de un partido en tiempo real
- Portal**
Repositorio de recursos online para analistas con la entrega de videos de partidos más rápida del mercado
- Reports**
Informes de rendimiento con métricas innovadoras validadas por analistas y expertos universitarios
- Broadcast**
Datos de tracking y eventing para mostrar en la señal de televisión antes, durante y después del partido

Usuarios Mediacoach

LaLiga

Fuente: López, 2019, <https://noesfutboleslaliga.elmundo.es/directos-al-futuro/posesion-pero-con-pegada-asi-es-el-adn-del-campeon-de-laliga-santander>

Estas plataformas son otra de las fuentes de datos que tienen los clubes y los cuerpos técnicos de las ligas profesionales que ofrecen este servicio.

6. Plataformas de ligas y federaciones: las ligas y federaciones juegan un papel importantísimo en cuanto al acceso a la información y al dato por parte de los clubes. La creación de estas herramientas o plataformas de acceso a la información, por parte

de todos los cuerpos técnicos (entrenadores, analistas y preparadores físicos) y todos los departamentos de análisis de los clubes de sus respectivas ligas, ha creado una democratización del acceso a la información, ya que anteriormente solo los clubes más potentes podían permitirse pagar a estos proveedores para acceder a sus datos.

Esto ha hecho evolucionar mucho más rápidamente a la industria del dato, mejorando sus prestaciones y utilidad, al mismo tiempo que se ha ido facilitando su acceso. Al tener acceso y utilizar el conjunto de datos de *tracking* combinado con el *eventing* (*raw data*), los clubes son capaces de aglutinar mucha más información que pueden trabajar por sí mismos. Esto se hace de una manera externalizada (consumiendo las propias métricas y visualizaciones que las plataformas les ofrecen) o de una manera más interna (trabajando y transformando el dato con procesos de ETL que les permitirán crear su propio conocimiento y sus propias visualizaciones).

Con el uso de estas plataformas, pueden mejorar el rendimiento de los jugadores y reforzar la participación de los aficionados con un conjunto simple de herramientas que permiten:

- a) Unir datos y video: las capacidades de datos y video resultantes de la combinación del *tracking* y *eventing*, brindan una precisión increíble y flujos de trabajo acelerados, lo que permite a los analistas crear información valiosa sobre la planificación, la exploración y la gestión de la carga de los jugadores para sus entrenadores.
- b) Mejorar las tácticas de los equipos: el acceso al *raw data* de estas plataformas proporciona tecnología que brinda datos de rendimiento precisos y válidos a través de videos informativos que respaldan su análisis. Esto permite al cuerpo técnico aprender del rendimiento del equipo a lo largo del tiempo, desarrollar programas de mejora específicos y elaborar nuevas tácticas para lograr un éxito aún mayor.
- c) Tomar decisiones durante el partido: estas plataformas ofrecen contenido en tiempo real durante el partido para el seguimiento en vivo del rendimiento del equipo, así como los datos de eventos con video sincronizado, que permiten a los entrenadores tomar más y mejores decisiones durante el partido, lo que les da una ventaja competitiva crucial.
- d) Optimizar el rendimiento del jugador: la excelente tecnología de seguimiento que utilizan estos proveedores proporciona una precisión de posicionamiento y una calidad de datos insuperable, que sirven para capturar el rendimiento físico de los jugadores. De esta forma, es un elemento clave para el desarrollo de programas de entrenamiento especializado, minimizar las lesiones y mantener conjuntos de datos precisos y consistentes, tanto del entrenamiento como del partido.

Figura 17: Ejemplo de la plataforma de la liga alemana Match Facts



Fuente: Deutsche Fussball Liga , 2020, <https://www.dfl.de/en/news/dfl-and-amazon-web-services-to-provide-new-real-time-match-analysis/>

7. Principales páginas web de datos en fútbol: otra de las fuentes de donde podemos obtener datos e información relevante sobre equipos, jugadores, partidos y métricas defensivas de presión alta son las páginas web especializadas en fútbol.

Además de ser públicas, gratuitas y de ofrecer incluso algún tipo de visualización (como la web Understat), lo más interesante es la posibilidad de descargar sus datos (Fbref). Esto puede hacerse también mediante *web scraping*, utilizando alguna de las herramientas de lenguaje de programación mencionadas en este módulo (R o Python), o directamente desde las herramientas de visualización (Power Bi o Tableau). Sin duda, estas páginas web juegan un papel fundamental para los clubes y cuerpos técnicos sin recursos económicos. Vamos a comentarlas:

7.1. Fbref: es una web gratuita con los datos más avanzados que existe actualmente en el mundo del fútbol. Es que, gracias a su asociación con StatsBomb, ofrece métricas todavía más avanzadas que las de páginas como WhoScored.

Es lo más cercano a lo que ofrecen Opta, StatsBomb, Wyscout e Instat, pero de forma pública y gratuita. No paran de mejorar su web, añadir datos y métricas novedosas. Por ejemplo, se incorporan datos como los xG (goles esperados), datos defensivos de presión, datos avanzados del portero, etc.

Además, ofrecen el dato total y su cálculo x90, y lo ordenan de una manera muy atractiva (métricas *standard*, métricas de tiro, de pase, de creación de ocasiones, de posesión, de partidos y tiempo jugado, de acciones defensivas, etc.).

Figura 18: Ejemplo de la página web Fbref

Squad Defensive Actions 2020-2021 La Liga [View Player Stats](#)

Squad Stats		Opponent Stats																							
Squad	#	PI	90s	Tackles				Vs Dribbles				Pressures				Blocks									
				Tkl	TklW	Def 3rd	Mid 3rd	Att 3rd	Tkl	Att	Tkl%	Past	Press	Succ	%	Def 3rd	Mid 3rd	Att 3rd	Blocks	Sh	ShSv	Pass	Int	Tkl+Int	Clr
Alavés	30	38.0	575	356	295	216	64	185	554	33.4	369	5447	1510	27.7	1752	2541	1154	568	105	4	463	432	1007	1065	9
Athletic Club	27	38.0	523	343	235	227	61	158	552	28.6	394	5250	1451	27.6	1494	2380	1376	553	86	2	467	348	871	897	4
Atlético Madrid	25	38.0	704	449	346	267	91	288	671	42.9	383	5124	1540	30.1	1665	2328	1131	579	93	0	486	395	1099	797	3
Barcelona	25	38.0	548	351	240	226	82	206	511	40.3	305	5075	1546	30.5	1404	2330	1341	519	82	0	437	354	902	474	11
Betis	25	38.0	605	378	289	240	76	196	557	35.2	361	5263	1514	28.8	1619	2495	1149	531	84	2	447	432	1037	855	10
Cádiz	34	38.0	611	385	337	218	56	222	623	35.6	401	5264	1433	27.2	2094	2304	866	612	118	4	494	411	1022	950	10
Celta Vigo	30	38.0	702	419	353	279	70	222	690	32.2	468	5997	1701	28.4	2063	2850	1084	610	97	1	513	375	1077	966	9
Eibar	30	38.0	572	362	249	236	87	189	564	33.5	375	5015	1602	31.9	1272	2251	1492	553	75	2	478	390	962	965	8
Elche	30	38.0	598	394	346	213	39	193	608	31.7	415	5168	1455	28.2	2068	2228	872	553	111	2	442	388	986	866	8
Getafe	31	38.0	617	383	261	278	78	199	558	35.7	359	6085	1689	27.8	1602	3055	1428	557	68	4	489	420	1037	1129	6
Granada	34	38.0	567	374	276	226	65	188	509	36.9	321	5027	1358	27.0	1616	2345	1066	611	142	5	469	406	973	955	9
Huesca	27	38.0	634	401	339	233	62	219	606	36.1	387	5451	1578	28.9	1951	2462	1038	551	103	1	448	348	982	775	6
Levante	29	38.0	573	368	289	216	68	200	602	33.2	402	5363	1499	28.0	1697	2350	1316	538	108	3	430	445	1018	991	12
Osasuna	28	38.0	569	354	255	236	78	204	536	38.1	332	5329	1550	29.1	1558	2487	1284	557	109	2	448	384	953	1027	3
Real Madrid	30	38.0	572	353	285	211	76	184	552	33.3	368	4880	1425	29.2	1562	2175	1143	535	100	4	435	406	978	667	9
Real Sociedad	30	38.0	560	342	251	235	74	172	529	32.5	357	5070	1487	29.3	1423	2340	1307	497	104	2	393	337	897	859	5
Sevilla	26	38.0	571	389	253	242	76	200	546	36.6	346	4716	1477	31.3	1392	2199	1125	503	95	3	408	334	905	767	8
Valencia	30	38.0	508	312	259	183	66	187	574	32.6	387	4855	1305	26.9	1794	2067	994	566	137	2	429	321	829	951	14
Valladolid	32	38.0	607	376	338	202	67	220	669	32.9	449	5839	1624	27.8	2051	2600	1188	601	117	2	484	390	997	989	8
Villarreal	29	38.0	566	340	253	242	71	177	580	30.5	403	4994	1425	28.5	1587	2226	1181	578	117	3	461	354	920	777	8

Totals may not be complete for all senior-level play, see [coverage note](#).
 Expected Goals ([xG explained](#)) and other Advanced Data provided by [StatsBomb](#), and is available for [these competitions](#).

Fuente: captura de pantalla de la web Fbref, <https://fbref.com/en/comps/12/La-Liga-Stats>

7.2. Understat: en esta página web podrá encontrar información con la cual analizar los equipos y los jugadores de las cinco grandes ligas de Europa, con métricas avanzadas como el xG (*expected goal*), NPxG (*non-penalty expected goal*), xPTS (*expected points*), xA (*expected assist*), etc. y, a nivel defensivo, destaca el ya conocido PPDA (*passes per defensive action*).

Así como Fbref utiliza los datos que genera StatsBomb, Understat utiliza los datos de Opta, por lo que debe saber que hay ligeras diferencias en los datos, pues cada proveedor utiliza sus propios algoritmos para generarlos.



Figura 19: Ejemplo de la página web Understat

Table		Charts		overall	home	away	Start date	End date													Settings
Nº	Team	M	W	D	L	G	GA	PTS	xG	NPxG	xGA	NPxGA	NPxGD	PPDA	OPPDA	DC	ODC	xPTS			
1	Atletico Madrid	38	26	8	4	67	25	88	57.88	-8.34	52.31	35.87	+10.67	32.70	+18.61	10.32	12.38	310	185	68.41	+8.59
2	Real Madrid	38	25	9	4	67	28	84	65.07	-1.83	62.84	38.70	+10.70	32.75	+30.09	10.14	14.03	328	171	69.98	+4.02
3	Barcelona	38	24	7	7	85	38	79	83.96	-1.04	78.01	40.93	+2.93	37.22	+40.80	9.51	16.20	499	188	78.18	-0.82
4	Sevilla	38	24	5	9	53	33	77	55.86	+2.86	50.68	34.33	+1.33	29.84	+20.81	8.15	14.58	240	139	68.96	+10.04
5	Real Sociedad	38	17	11	10	59	38	62	63.15	+4.15	54.22	38.80	+6.80	35.08	+19.14	8.08	11.70	285	180	67.61	+5.61
6	Real Betis	38	17	10	11	50	50	61	49.77	-0.23	41.43	47.51	-2.48	40.08	+1.35	9.12	10.55	224	180	55.37	+5.63
7	Villarreal	38	15	13	10	60	44	58	57.88	-2.14	48.19	43.67	-0.33	36.98	+11.21	9.85	11.43	291	182	63.77	+5.77
8	Celta Vigo	38	14	11	13	55	57	53	51.32	-3.88	46.12	50.31	-8.69	44.37	+1.75	7.40	10.84	217	187	52.83	-0.17
10	Granada	38	13	7	18	47	85	46	41.05	-5.85	37.34	59.82	-5.18	50.61	-13.27	10.74	8.88	128	234	40.58	-5.42
9	Athletic Club	38	11	13	14	46	42	46	43.06	-2.84	39.35	38.46	-3.54	34.00	+5.34	9.75	9.80	202	197	54.77	+5.77
12	Cadiz	38	11	11	16	38	58	44	38.70	+0.70	33.73	58.91	+6.81	51.48	-17.75	16.57	5.93	77	274	39.01	-4.99
11	Osasuna	38	11	11	16	37	48	44	39.61	+2.61	35.15	49.83	+1.83	43.88	-8.74	10.91	8.73	172	243	45.10	+1.10
13	Valencia	38	10	13	15	50	53	43	41.89	-8.11	34.46	55.69	+2.69	49.74	-15.28	10.94	8.81	182	282	42.38	-0.62
14	Levante	38	9	14	15	46	57	41	40.83	-5.07	38.47	55.13	-1.87	52.16	-15.69	9.47	10.22	221	245	43.53	+2.53
16	Alaves	38	9	11	18	36	57	38	39.11	+3.11	33.17	51.55	-5.46	49.32	-16.15	12.25	7.88	119	212	43.23	+6.23
15	Getafe	38	9	11	18	28	43	38	35.82	+7.82	31.91	40.40	-2.80	34.32	-2.42	8.81	8.43	134	155	48.39	+10.39
17	Elche	38	8	12	18	34	55	36	30.40	-3.80	28.17	61.63	+5.63	54.20	-28.03	13.23	8.87	78	280	30.31	-5.69
18	SD Huesca	38	7	13	18	34	53	34	37.98	+3.98	35.00	48.84	-4.36	44.92	-9.92	11.32	8.53	98	215	45.23	+11.23
19	Real Valladolid	38	5	18	17	34	57	31	37.92	+3.92	31.97	52.58	-4.44	47.38	-15.39	10.57	7.97	134	243	41.53	+10.53
20	Eibar	38	6	12	20	29	52	30	42.45	+13.45	35.64	48.84	-3.16	45.12	-9.48	8.99	10.58	149	168	48.62	+18.62

Fuente: captura de pantalla de la web Understat, https://understat.com/league/La_liga

7.3. Transfermarkt: se trata de una web que cuenta con datos acerca de los equipos y sus jugadores, pero que está enfocada en los fichajes y en el valor de los jugadores en el mercado. En ella, encontramos noticias sobre los rumores de traspasos; los últimos fichajes y renovaciones; cesiones; jugadores que acaban contrato y los que han quedado libres; los fichajes más caros; jugadores más valiosos; etc.

Ha conseguido algo extremadamente difícil: ser el referente sobre datos de la valoración económica de los jugadores. Esta valoración ha trascendido su propia web e incluso es utilizada en muchas ocasiones por clubes (pequeños, por lo general). Se puede buscar por posición o por posiciones detalladas, edad de los futbolistas, etc. En este listado, encontraremos el nombre, la demarcación, la nacionalidad, la edad, el club al que pertenece y su valor en el mercado. Se puede ver si el valor se mantiene, si ha bajado o si, por el contrario, ha subido.



Figura 20: Ejemplo de la página web Transfermarkt

#20 Bernardo Silva

Manchester City
Liga: EFL Primera División
Fichado: 01/07/2017
Contrato hasta: 30/06/2025

Nombre en país de origen: **Bernardo Mota Veiga de Carvalho e Silva**

Fecha de nacimiento: **10/08/1994**

Lugar de nacimiento: **Lisboa**

Edad: **26**

Altura: **1,73 m**

Nacionalidad: **Portugal**

Posición: **Medio campo - Mediocentro ofensivo**

Pie: **Izquierdo**

Agente: **Gestifute**

Club actual: **Manchester City**

Fichado: **01/07/2017**

Contrato hasta: **30/06/2025**

Última renovación: **13/03/2019**

Proveedor: **adidas**

Redes sociales: [Twitter](#) [Facebook](#) [Instagram](#)

70,00 mill. €

Posición en detalle

Posición principal: **Mediocentro ofensivo**

Posición secundaria: **Mediocentro Extremo derecho**

Valor de mercado

Valor de mercado actual: **70,00 mill. €**

Última revisión: 01/06/2021

Valor más alto: **100,00 mill. €**
13/06/2019

Ver página de valores detallados »

Premier League 20/21
38 partidos posibles

Partidos: **26** Tarjetas amarillas: **5**

Goles: **2** Segundas tarjetas amarillas: **-**

Asistencias: **6** Tarjetas rojas: **-**

63% Cuenta XI inicial

61% Minutos jugados

9% Participaciones de gol

Datos de rendimiento detallados

Fuente: captura de pantalla de la web Transfermarkt, <https://www.transfermarkt.es/bernardo-silva/profil/spieler/241641>

7.4. WhoScored: este portal de estadísticas de fútbol no cuenta con datos de tanto nivel como los anteriormente mencionados, pero tiene datos interesantes sobre la gran mayoría de competiciones de fútbol y los muestra en una página web muy visual, intuitiva y que permite jugar con el nivel de detalle que se quiera.

Contiene datos avanzados, por equipo o jugador, y una enorme cantidad de datos por partido para las cinco grandes ligas. Incluso, tiene su propio *rating* de rendimiento de un jugador por partido. Además, tiene todo tipo de estadísticas por países y por equipo en modo resumen (características ofensivas y defensivas, resultados a lo largo del tiempo, número de partidos invicto, portería a cero, partidos sin ganar o sin marcar, etc.).

Figura 21: Ejemplo de la página web WhoScored

Top Player Statistics

View: Overall Home Away

R	Player	Team	Player Goals/Team Goals		Contribution
1	R. Lewandowski	Bayern	41	99	41%
2	L. Messi	Barcelona	30	85	35%
3	C. Ronaldo	Juventus	29	77	38%
4	A. Silva	E. Frankfurt	28	69	41%
5	K. Mbappé	PSG	27	86	31%
6	E. Haaland	B. Dortmund	27	75	36%
7	R. Lukaku	Inter	24	89	27%
8	H. Kane	Tottenham	23	68	34%
9	K. Benzema	R. Madrid	23	67	34%
10	G. Moreno	Villarreal	23	60	38%

Aggression		Shots per Game				
P. Schiattarella	Benevento	14	1	L. Messi	Barcelona	5.6
W. Faes	Reims	11	2	C. Ronaldo	Juventus	5.1
M. Cassamá	Reims	8	3	R. Lewandowski	Bayern	4.7
S. Savic	Atletico	15	0	Z. Ibrahimovic	AC Milan	4.3
D. Suárez	Getafe	12	1	L. Insigne	Napoli	4.1

Tackles per Game		Dribbles per Game				
L. Abergel	Lorient	3.8		L. Messi	Barcelona	4.5
W. Ndidi	Leicester	3.7		A. Traoré	Wolves	4.1
V. Rongier	Marseille	3.6		A. S-Maximin	Newcastle	3.8
A. Tchouameni	Monaco	3.5		J. Doku	Rennes	3.7
T. Mitchell	Crystal Palace	3.4		R. de Paul	Udinese	3.4

Fuente: captura de pantalla de la web Whoscored, <https://www.whoscored.com>



Unidad 4.3. Principales métricas defensivas para medir la presión alta

En el mundo de la analítica de fútbol, históricamente siempre ha sido más complicado tratar de medir la parte defensiva del juego que la ofensiva. Básicamente, porque mientras que en el ataque el rendimiento exitoso acaba traducándose en realizar una acción (rematar, asistir, completar un regate, meter gol etc.) y hacerlo más veces suele ser mejor, en la defensa no siempre sucede lo mismo. En defensa, actuar poco puede ser mejor que actuar mucho:

Un defensa que realice más entradas o interceptaciones no es necesariamente mejor que uno que realice menos, estos datos miden más la oportunidad que la habilidad. En las acciones que se realizan en defensa, hay mucho más ruido que señal. ¿Cómo podemos empezar a disipar ese ruido? (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

El objetivo defensivo es evitar que el equipo rival progrese y cree ocasiones de gol y, por tanto, muchas veces lo que no se hace es tan importante —incluso más— que lo que sí se hace. Por ejemplo, un posicionamiento adecuado o un *pressing* coordinado de manera eficaz limitan las opciones del rival y, por tanto, es menos probable que los jugadores más retrasados tengan que defender de manera activa frecuentemente.

En el fútbol no ha sido hasta hace muy poco tiempo que se han puesto las bases del análisis estadístico defensivo. Esto ha sido gracias a la contribución de proveedores de datos como StatsBomb y Opta, que, junto con las demandas provenientes de los clubes en cuanto a mejoras estadísticas y datos existentes y los avances tecnológicos y conocimientos de sus profesionales de diferentes ámbitos (informáticos, matemáticos, físicos, entrenadores, analistas, etc.), están consiguiendo un avance nunca antes visto.

Para ello, voy a centrarme en las métricas defensivas que ofrece el proveedor de datos StatsBomb, ya que, según mi punto de vista, es el proveedor que mejor recoge los datos de las acciones de presión, tanto de equipos como de jugadores.

Siguiendo a Pablo Peña Rodríguez, jefe de Tactical Innovation and Business Development de StatsBomb, vamos a comentar cada una de las métricas y visualizaciones defensivas que nos van a permitir medir la capacidad de presionar alto de un equipo.



4.3.1. PPDA: Pases por cada acción defensiva

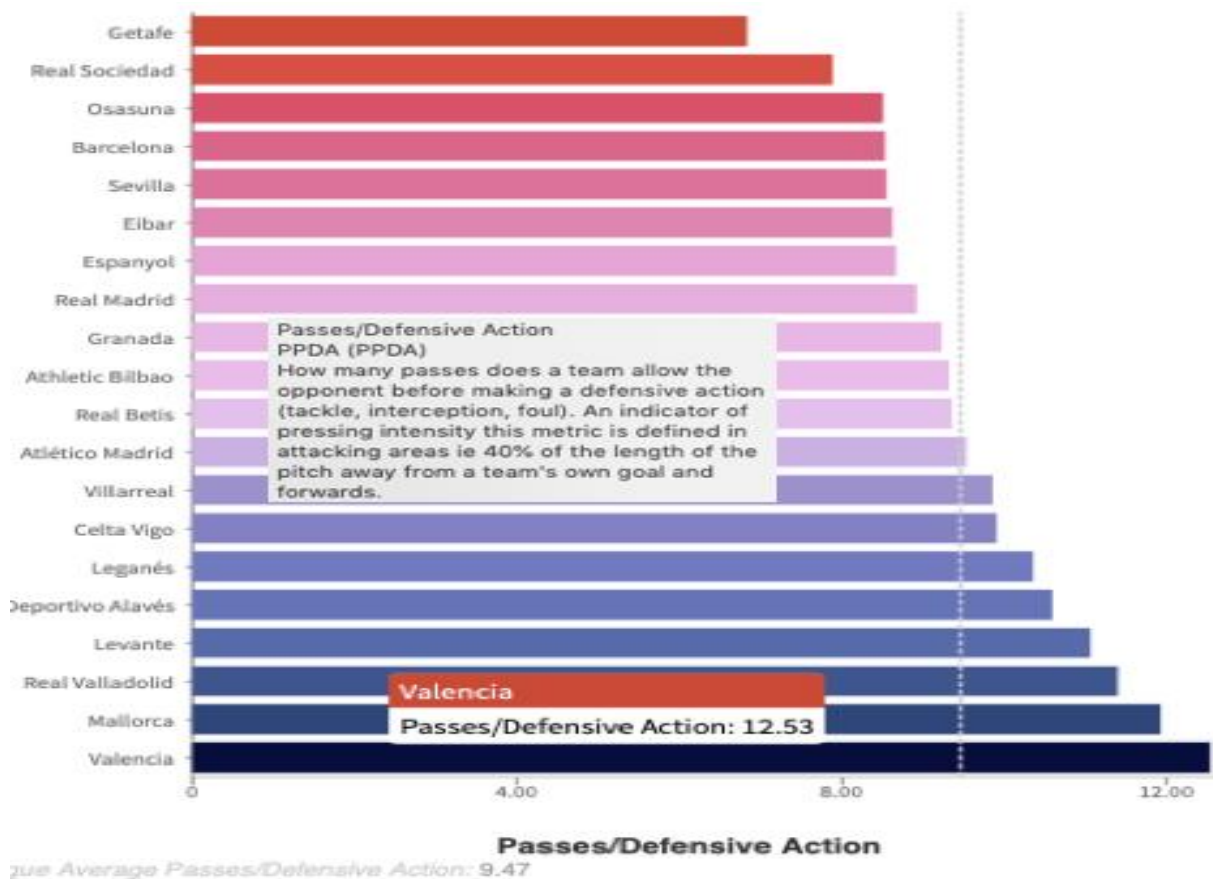
Esta es una métrica que creó Colin Trainor en 2014. Se encontraba realizando un trabajo en el que intentaba medir la capacidad de presionar alto de los equipos, intentando responder a las inquietudes de los entrenadores sobre cómo, cuándo y dónde realizaban las acciones defensivas los equipos (entradas, interceptaciones, faltas, duelos).

Para ello, estableció una correlación entre los pases permitidos al rival y las acciones defensivas del equipo que intentaba recuperar el balón, con las siguientes conclusiones:

- a) Cuántos menos pases realiza el rival por cada acción defensiva de mi equipo, más intensidad defensiva tiene mi equipo.
- b) Al contrario, cuántos más pases realiza el rival por cada acción defensiva de mi equipo, menor intensidad defensiva ejerce mi equipo.

En la figura 22, vemos el *ranking* final de la liga 2019/20 tras las 38 jornadas de campeonato. En este, vemos como el Getafe es el equipo que menos pases por cada acción defensiva realizada permite en campo rival a su oponente en la fase de inicio y progresión y el Valencia, el que más pases por cada acción defensiva realizada permite en campo rival.

Figura 22: PPDA en un *ranking* StatsBomb



Fuente: [Imagen sin título sobre PPDA en un *ranking* de StatsBomb]. (s. f.). Recuperado de https://lh3.googleusercontent.com/4QRdqGWNi0_bJ_Zs392Z0mH96LDHnIenpyl145mhA9-n0sTvqJTvUzR4NS2CSroCXb0U=s85

De esta visualización, podemos interpretar que el Getafe presiona mucho mejor que el Valencia en el campo rival, con lo que tiene una mayor capacidad de ejercer presión alta para impedir el juego asociativo del oponente.

En un primer momento, la idea inicial de este estudio era la de identificar a los equipos que mejor presionaban en campo rival, ya que esta métrica se aplicaba en el 60 % del total del campo más alejado de nuestra portería (el campo rival + 10 metros de nuestro propio campo).

Pero en StatsBomb también la utilizan en todo el campo para establecer el perfil defensivo de los equipos en cada uno de los seis sectores del campo. Además, los eventos defensivos están normalizados en función de la posesión del rival en cada sector:

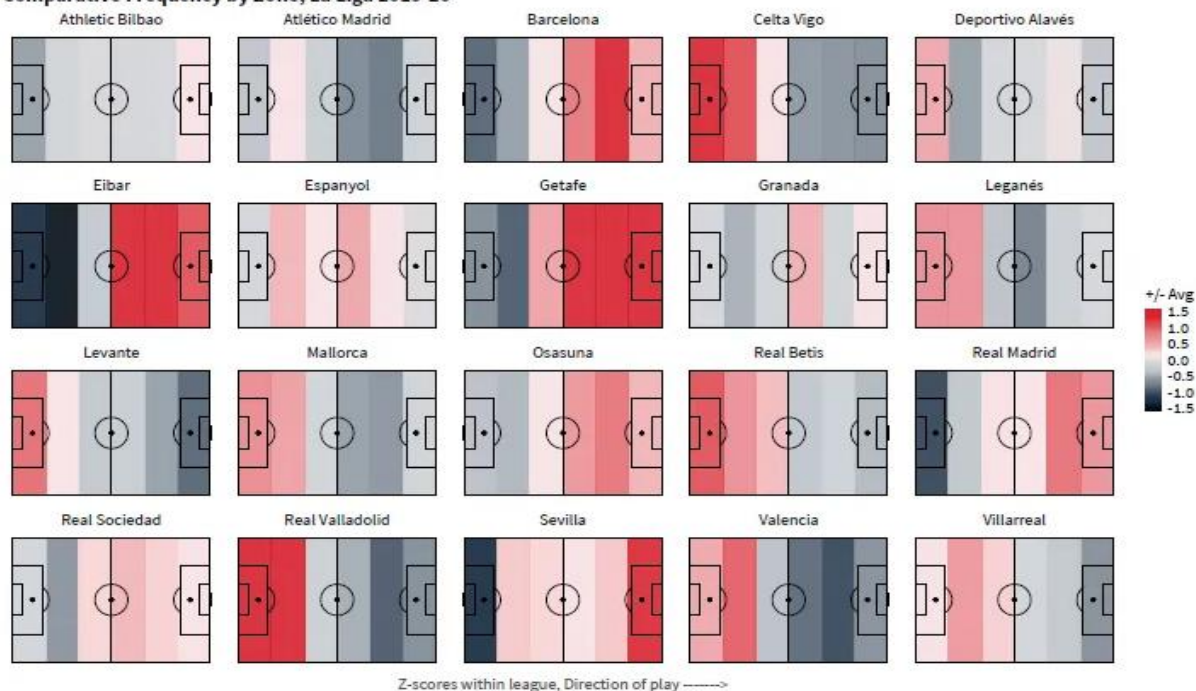
Pese a que la idea inicial era discriminar a los equipos que mostraban una presión más alta, la métrica se fue desarrollando y, gradualmente, adaptando a diferentes zonas hasta ser una de las métricas descriptivas más empleadas para mostrar los rasgos estilísticos que definen la defensa de los equipos. En StatsBomb, la utilizamos regularmente en los análisis. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

Figura 23: Proporción de acciones defensivas de StatsBomb

Proportion of Defensive Actions to Opposition Passes



Comparative Frequency by Zone, La Liga 2019-20



Fuente: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

4.3.2. La presión (*pressure*)

En mi opinión, la clave de todas las métricas avanzadas para medir la acción y la intensidad defensiva de un equipo es la toma del dato de presión (*pressure*), que para StatsBomb es la acción de acosar al poseedor rival:

La presión es la acción de atacar al poseedor de balón o al receptor de un pase en un radio de cinco metros en función de la zona del campo, sin llegar a realizar una entrada, falta o interceptar el pase (cualquiera de estas acciones puede ocurrir posteriormente, pero es un evento diferente). (Peña Rodríguez, P. 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

La razón detrás de recoger este dato de presión se refuerza por la creencia acerca de que muchas de las acciones defensivas que permiten que, posteriormente, un equipo recupere la posesión o lleve a cabo una presión colectiva eficaz no se limitan a las acciones que ocurren sobre el balón (las entradas o *tackles*; interceptaciones; duelos; despejes; etc.). Sino que hay toda una serie de acciones que llevan a ello, tales como acosar para orientar la presión hacia una determinada zona, cerrar líneas de pase,



acelerar la acción del jugador, forzar pases descontrolados o hacer que el rival poseedor se quite el balón de encima.

A nivel colectivo, esto nos permite tener una visión más realista y completa de los mecanismos tácticos que los equipos usan en fase defensiva y podemos dar respuesta a las preguntas que muchos entrenadores y analistas se hacen cuando estudian las capacidades defensivas del rival:

- a) ¿Cómo podemos tener una visión más representativa de la influencia de determinados jugadores en la defensa, más allá de recuperar la posesión directamente?
- b) ¿Qué rasgos estilísticos diferencian a los equipos que tratan de presionar alto respecto de los que defienden en bloques medios o bajos?
- c) ¿En qué zonas los rivales comienzan a ejercer presión? ¿Hacia qué lado dirigen a los rivales? ¿Dónde son más fuertes?
- d) ¿Con qué frecuencia presionan en determinadas zonas del campo y qué jugadores son los más involucrados en ello?, etc. (Peña Rodríguez, 2020).

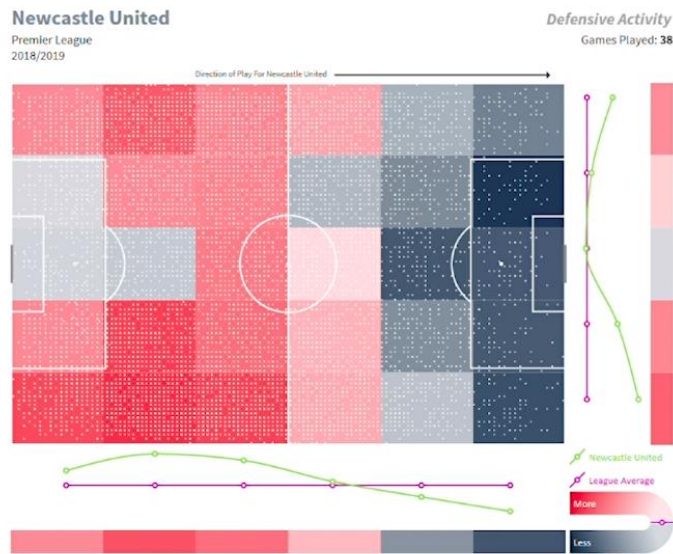
Para recuperar la posesión lo más cerca posible de la portería rival durante la fase de presión alta, los equipos toman decisiones estratégicas que marcan sus perfiles defensivos y que son claramente reconocibles en las visualizaciones de los datos defensivos analizados.

La primera visualización del dato de presión la tenemos en los mapas de actividad defensiva.

Mapa de actividad defensiva

Siguiendo a Peña Rodríguez (2020), en el mapa de actividad defensiva vemos las acciones defensivas realizadas por el equipo en cada cuadrante. StatsBomb divide el campo en cinco sectores verticales y tres sectores horizontales (cada medio campo). El color rojo indica una mayor actividad defensiva respecto a la media de la liga en cuestión y el color gris oscuro nos indica una menor actividad defensiva respecto a la media.

Figura 24: Mapa de actividad defensiva de StatsBomb



- Rojo= más frecuente
- Azul=Menos frecuente
- Dirección de juego = De izquierda a derecha

STATSBOMB

Fuente: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

En la figura 24, tenemos al Newcastle United de la temporada 2018/19, dirigido por Rafa Benitez. En el mapa, vemos cómo el equipo está enfocado en defender en un bloque bajo en campo propio, con mayor actividad en las zonas exteriores y con zonas irregulares en el interior y en el balcón del área propia. También, tiene predominancia defensiva en medio campo, posicionándose en un bloque medio, pero, a medida que avanzamos hacia la portería rival, podemos apreciar que no es un equipo que destaque por su actividad defensiva de presión alta (vamos pasando de un color gris a uno más oscuro).

En la figura 25, tenemos un ejemplo de lo contrario, el Manchester City, de Pep Guardiola, es el ejemplo de un equipo que tiene un claro perfil defensivo de presión alta en campo rival.

Figura 25: Comparación de mapas con métricas tradicionales y con la métrica de presión

Estilos: Presión alta



Las estadísticas defensivas tradicionales no abarcan todo el espectro

Izquierda: Man City 2018-19, tackles, interceptaciones, bloqueos, etc

Derecha: Lo mismo pero incluyendo acciones de presión

STATSBOMB

Fuente: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

En el gráfico de la izquierda, tenemos cómo sería el mapa solamente con las estadísticas defensivas tradicionales, que solo miden las acciones defensivas en las que hay contacto sobre el balón. En cambio, en el mapa de la derecha, se añaden las acciones de presión y acoso del equipo sobre el rival poseedor del balón. La diferencia es evidente. El dato de presión nos ofrece una imagen más representativa del apartado defensivo del juego y, más concretamente, de la capacidad de presión alta que tienen los equipos.

Los eventos defensivos en los proveedores de datos tradiciones han sido las entradas o *tackles*, interceptaciones, duelos, despejes, etc. Aunque estos datos son útiles hasta cierto punto, no llegan a dar una visión representativa del espectro completo de acciones que tienen impacto a nivel defensivo en el equipo, limitando así la capacidad de extraer conclusiones relevantes... Los datos de StatsBomb proporcionan un nuevo evento que cambia la manera de afrontar el análisis defensivo: las presiones. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

De esta afirmación podemos extraer la conclusión de que el dato de presión nos ofrece una visión mejor y más amplia sobre como presionan los equipos en campo rival que la métrica del PPDA.

Otras visualizaciones que podemos utilizar a nivel colectivo son los *rankings* de equipos.

Rankings

Con los datos defensivos de presión parciales (en una determinada zona del campo) o totales (en todo el campo), podemos diferenciar las diferentes fases defensivas.

En la figura 26, vemos un *ranking* tomado de la web de Fbref, en el que vemos en color rojo las presiones ejercidas por cada equipo en cada tercio del campo.

- Def. 3.rd = bloque bajo.
- Mid. 3.rd = bloque medio.
- Att. 3.rd = bloque alto.

En amarillo, en el *ranking* de la izquierda, vemos que el Sevilla es último en número de presiones totales ejercidas durante La Liga 2020/21.

Figura 26: Rankings de presiones totales ejercidas y por tercios de campo

Squad Stats		Number of times applying pressure to opposing player who is receiving, carrying or releasing the ball									
Squad	Press ▼	Succ	%	Def 3rd	Mid 3rd	Att 3rd	Blocks	Sh	Sh5v	Pass	
Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	
Ceuta Vigo	5997	1701	28.4	2063	2850	1084	610	97	1	513	
Valladolid	5839	1624	27.8	2051	2600	1188	601	117	2	484	
Huesca	5451	1578	28.9	1951	2462	1038	551	103	1	448	
Alavés	5447	1510	27.7	1752	2541	1154	568	105	4	463	
Levante	5363	1499	28.0	1697	2350	1316	538	108	3	430	
Osasuna	5329	1550	29.1	1558	2487	1284	557	109	2	448	
Cádiz	5264	1433	27.2	2094	2304	866	612	118	4	494	
Betis	5263	1514	28.8	1619	2495	1149	531	84	2	447	
Athletic Club	5250	1451	27.6	1494	2380	1376	553	86	2	467	
Eibar	5168	1455	28.2	2068	2228	872	553	111	2	442	
Atlético Madrid	5124	1540	30.1	1665	2328	1131	579	93	0	486	
Barcelona	5075	1546	30.5	1404	2330	1341	519	82	0	437	
Real Sociedad	5070	1487	29.3	1423	2340	1307	497	104	2	393	
Granada	5027	1358	27.0	1616	2345	1066	611	142	5	469	
Eibar	5015	1602	31.9	1272	2251	1492	553	75	2	478	
Villarreal	4994	1425	28.5	1587	2226	1181	578	117	3	461	
Real Madrid	4880	1425	29.2	1562	2175	1143	535	100	4	435	
Valencia	4855	1305	26.9	1794	2067	994	566	137	2	429	
Sevilla	4716	1477	31.3	1392	2199	1125	503	95	3	408	

Fuente: adaptación propia de Fbref, s. f., <https://fbref.com/en/comps/12/La-Liga-Stats>

Otra visualización, en este caso de StatsBomb, la vemos en la figura 27. En ella podemos observar un *ranking* que mide la capacidad de presionar alto en campo rival de los equipos de La Liga 2020/21. En este *ranking* vemos al Sevilla 14.^a posición.

Figura 27: *Ranking* de presión en campo rival de StatsBomb



Fuente: [Imagen sin título sobre *ranking* de presión en el campo rival]. (s. f.). Recuperado de <https://lh3.googleusercontent.com/CFilwr8zUC-i9Gk-kyX5x-NEZRYcgj2QH4WoN63PAWrea3qbb1Avq4JGoepUMUvt1feuA=s85>

Gráficos de dispersión

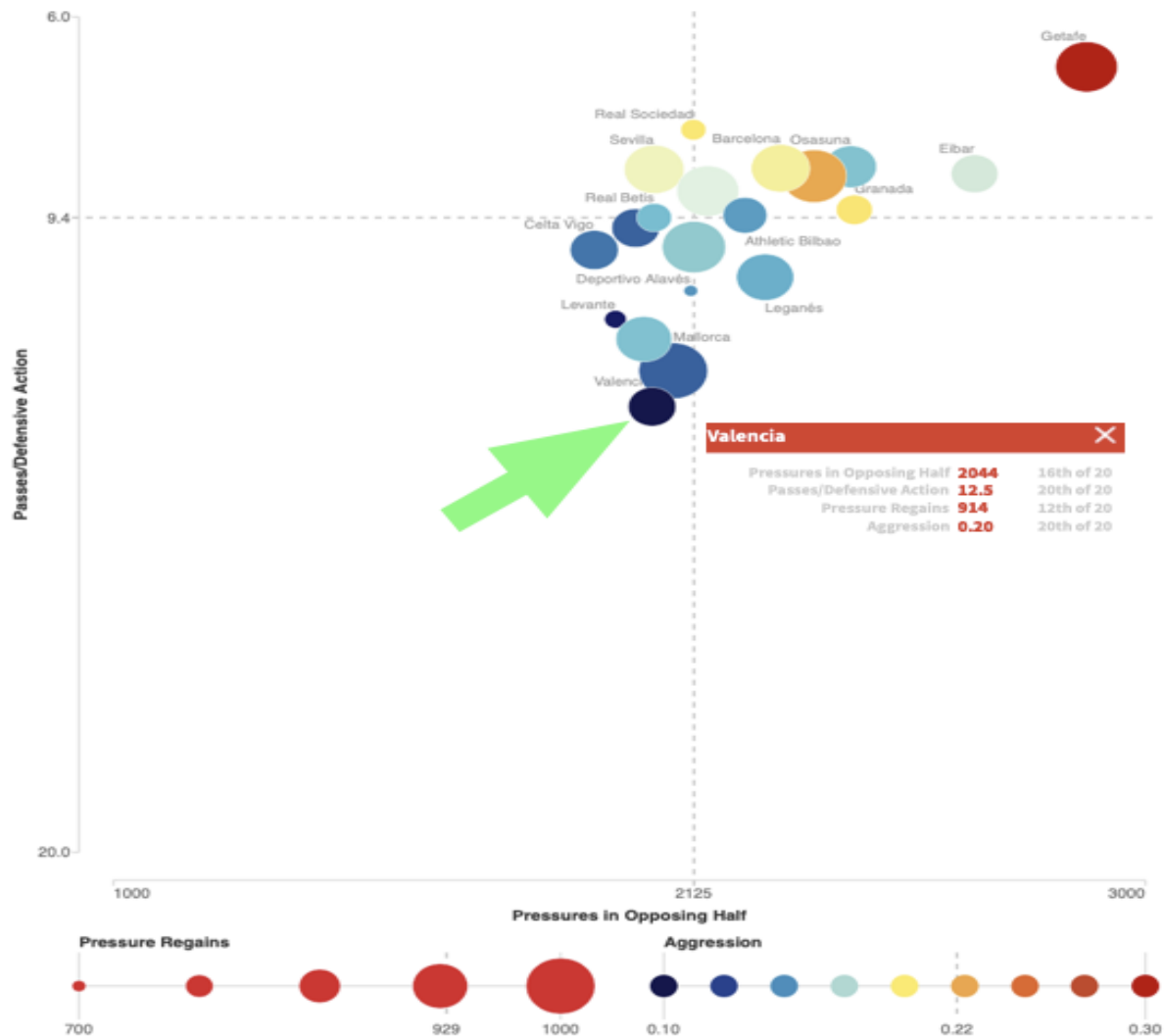
Otra de las visualizaciones más utilizadas son los gráficos de dispersión, que permiten estudiar las relaciones entre dos variables del juego. Gracias a ellos logramos situar a un equipo o jugador respecto del resto, ver si está por encima o por debajo de la media, etc.

En la figura 28, podemos observar un ejemplo de cómo se puede medir la capacidad de presión de un equipo en un solo gráfico. En él podemos observar hasta 4 variables de presión. Así estaba el Valencia de la temporada 2019/20 en cuanto a PPDA y presiones en campo rival. Pero, además, en la plataforma IQ de StatsBomb nos dan la posibilidad de añadirle 2 variables:



- a) El tamaño de la burbuja: representa el dato de "pressure regains" (recuperaciones pospresión ejercida).
- b) El color de la burbuja: representa el dato de "aggression" (el nivel de agresividad del equipo).

Figura 28: Capacidad de presión de un equipo en un gráfico de dispersión de 4 variables



Fuente: [Imagen sin título sobre la capacidad de presión de un equipo]. (s. f.). Recuperado de https://lh3.googleusercontent.com/dilrueUqUvtmDf4VyI_rTuVSwgSi_81ZkBzbaifEcEz6YIAGxBIQR04xmOKSxsmbllqUPw=s85



4.3.3. La presión tras pérdida (*counterpressure*)

Otra de las medidas que tenemos que tener en cuenta para valorar la capacidad de presión en campo rival es la presión tras pérdida (*counterpressure* o *gegenpressing*), es decir, las presiones ejercidas en los cinco segundos posteriores a la pérdida.

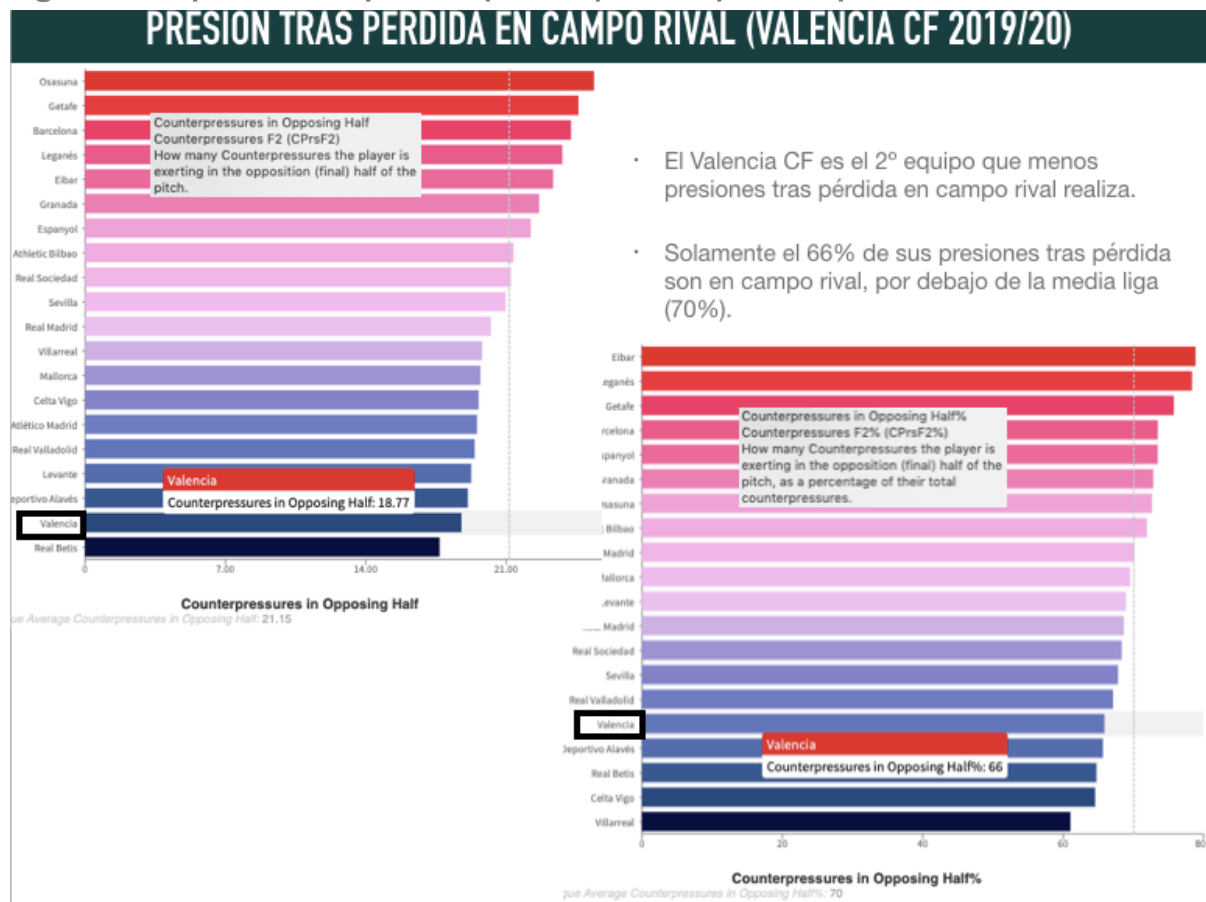
Esta presión tras pérdida (PTP) se da cuando un equipo presiona al rival durante los cinco segundos posteriores a la pérdida de la posesión del balón. Generalmente, se aplica en campo rival, en donde normalmente se producen las pérdidas de balón y el equipo está lo suficientemente junto como para que los jugadores cercanos puedan realizar esta PTP o contrapresión.

El objetivo inicial de la PTP es interrumpir el contraataque del oponente durante la fase de transición de ataque a defensa. Por lo que nos dará información relevante sobre qué jugadores tienen mayor influencia en los momentos de la transición ataque-defensa.

En la figura 29, observamos cómo se puede medir la presión tras pérdida en campo rival, con el ejemplo del Valencia en la temporada 2019/20:

- Cuantificando la media de PTP en campo rival por partido.
- Cuantificando el porcentaje del total de PTP en campo rival.

Figura 29: La presión tras pérdida (*counterpressure*) en campo rival



Fuente: [Imagen sin título sobre la presión tras la pérdida]. (s. f.). Recuperado de https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKKC6604tLENRDvJ7dDv1wYjnkHuYR_cmgkCmqNiv-TDYcyr7-HkM=s114

Hasta ahora, hemos hablado de datos para medir la capacidad de presión de los equipos, pero, ¿cómo medir la capacidad de presión de los jugadores de manera individual?

4.3.4. La presión a nivel individual: la recuperación pospresión (*pressure regain*)

A nivel individual, el dato de la presión ejercida es especialmente relevante para los atacantes y los centrocampistas (sobre todo si hablamos de la presión alta). Aunque, también, se puede emplear para los defensores, cuando hacemos referencia a la presión en la totalidad del campo del juego.

No solamente vamos a recoger el número de veces que presionan los jugadores, sino que también vamos a recoger el nivel de influencia que tiene en la recuperación de sus compañeros y, en definitiva, de su equipo.

Y esto lo vamos a conseguir utilizando la métrica *pressure regains* (recuperación pospresión). Si el equipo recupera la pelota en los cinco segundos posteriores a una acción de presión de un jugador, se le asignará una recuperación pospresión al jugador.

Figura 30: Las 4 variables que miden la capacidad de presión de un jugador



Fuente: [Imagen sin título sobre las cuatro variables que miden la capacidad de presión de un jugador]. (s. f.). Recuperado de <https://lh3.googleusercontent.com/D56U0KRjGbQ4I1Y-E73kKoKHqs-wrtDf0tL6UhOQcERi7P8NwOIZpjYge1bGLg6QEML-t7M=s85>

Con esta métrica, ya no solo se tratará de evaluar el rendimiento defensivo de los delanteros y centrocampistas con las acciones sobre el balón (entradas o interceptaciones), sino que se tratará de dar valor a las acciones defensivas que realizan para que el equipo recupere el balón:

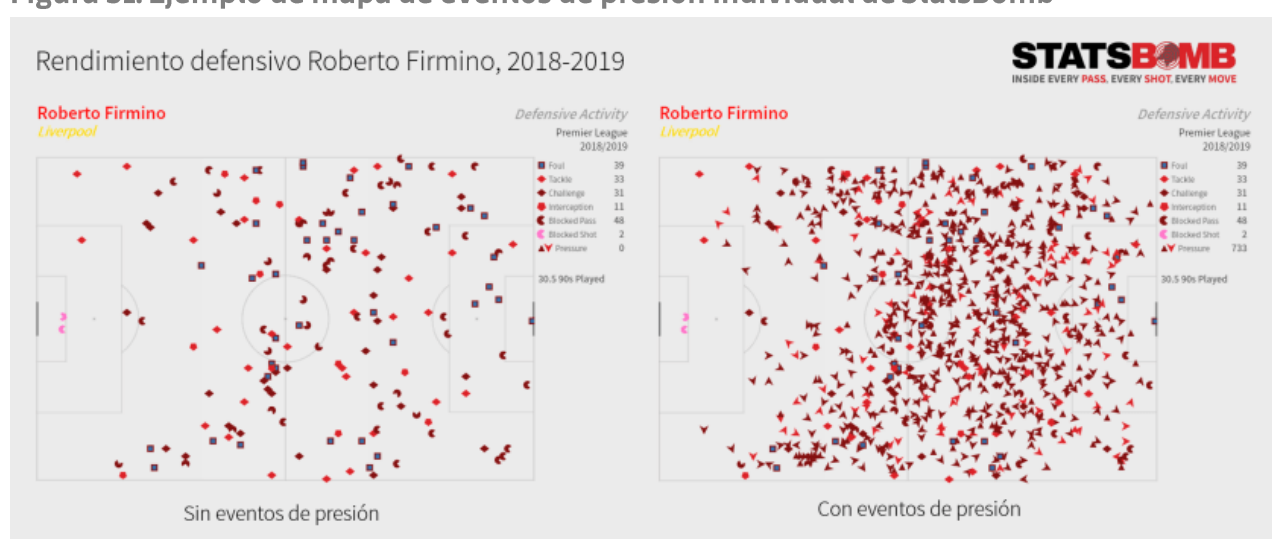
Es decir, normalmente el delantero no presiona para recuperar la pelota por sí mismo, sino para tapar líneas de pase concretas, orientar la salida de balón del rival hacia zonas concretas (normalmente exteriores), forzar a los rivales a realizar acciones descontroladas, cometer errores o a quitarse el balón de encima. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).



En otro tipo de visualizaciones, como los mapas de presión, podemos ver cada una de las acciones de presión realizadas por el jugador en cuestión, y, de una manera más visual, podemos observar qué campo de presión abarca su rendimiento defensivo.

Así, en la figura 31, con el objetivo de ver la importancia de estos eventos de presión, podemos ver la comparación de la visión del rendimiento defensivo de Roberto Firmino (2018-2019) excluyendo las acciones de presión (gráfico izquierdo) e incluyéndolas (gráfico derecho).

Figura 31: Ejemplo de mapa de eventos de presión individual de StatsBomb



Fuente: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

4.3.5. Ventajas y usos adicionales de los eventos de presión

Además de las valoraciones estrictamente defensivas sobre la presión alta, los eventos de presión nos permiten tener información sobre la fase ofensiva de los equipos, como, por ejemplo, saber cómo le afecta a un equipo o a un jugador determinado la presión ejercida por un rival. Es decir, se puede medir la capacidad de pase de un jugador cuando está bajo la presión de un adversario.

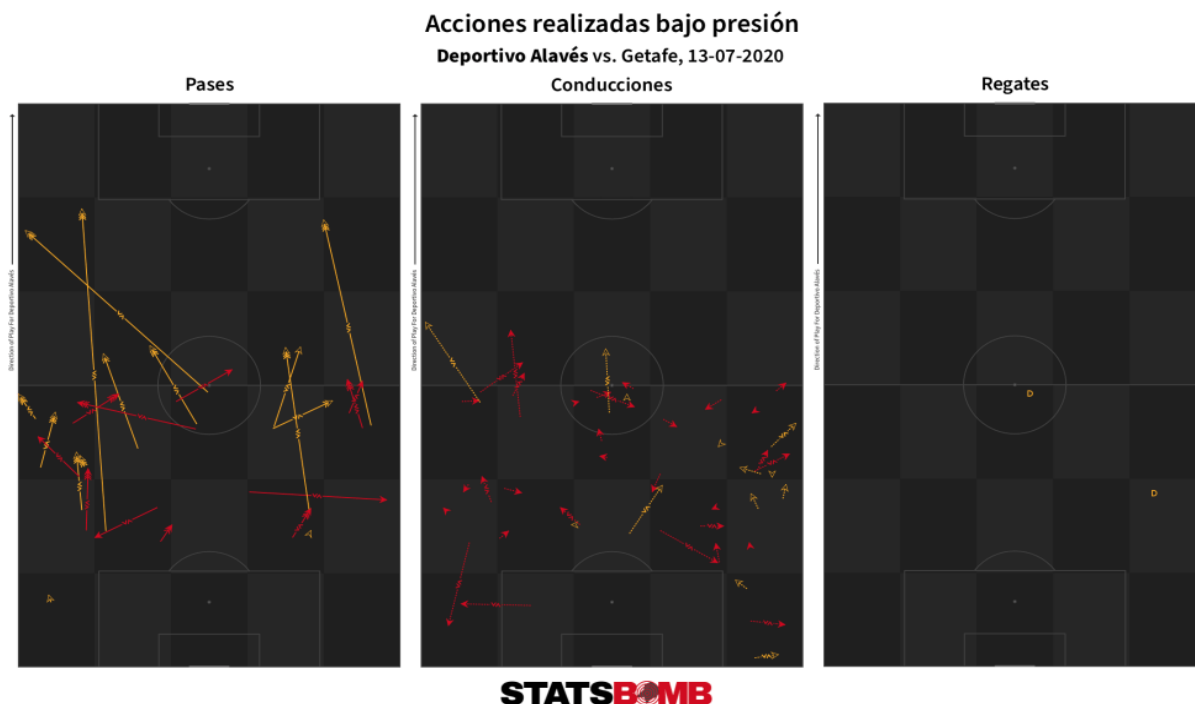
Esto lo podemos conseguir con la métrica AUP o "Actions under pressure" o "Passes under pressure". Por ejemplo, en las figuras 32 y 33, vemos cómo ante la presión alta fuerte y asfixiante del Getafe, equipos como el Alavés y el FC Barcelona tuvieron una reacción totalmente distinta en el campo contrario.

Podemos ver como el Alavés tuvo muchos problemas para combinar y enlazar pases en campo rival:



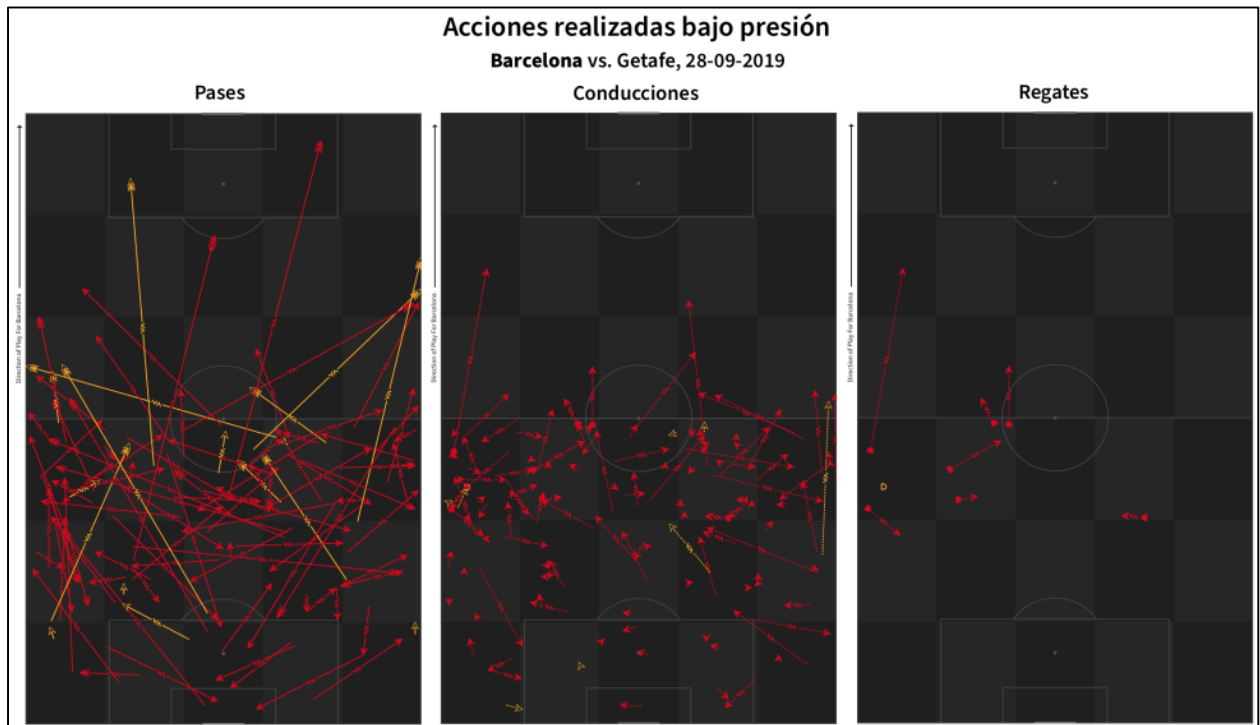
Sufrió un bajón notable respecto a su media de la temporada, tanto en porcentaje de pases completados como en el porcentaje de pases completados bajo presión. No completó ninguno de los tres regates que intentó (rojo = acción completada; amarillo = acción fallida). (StatsBomb, febrero 11, 2021).

Figura 32: Mapas de acciones realizadas bajo presión (StatsBomb)



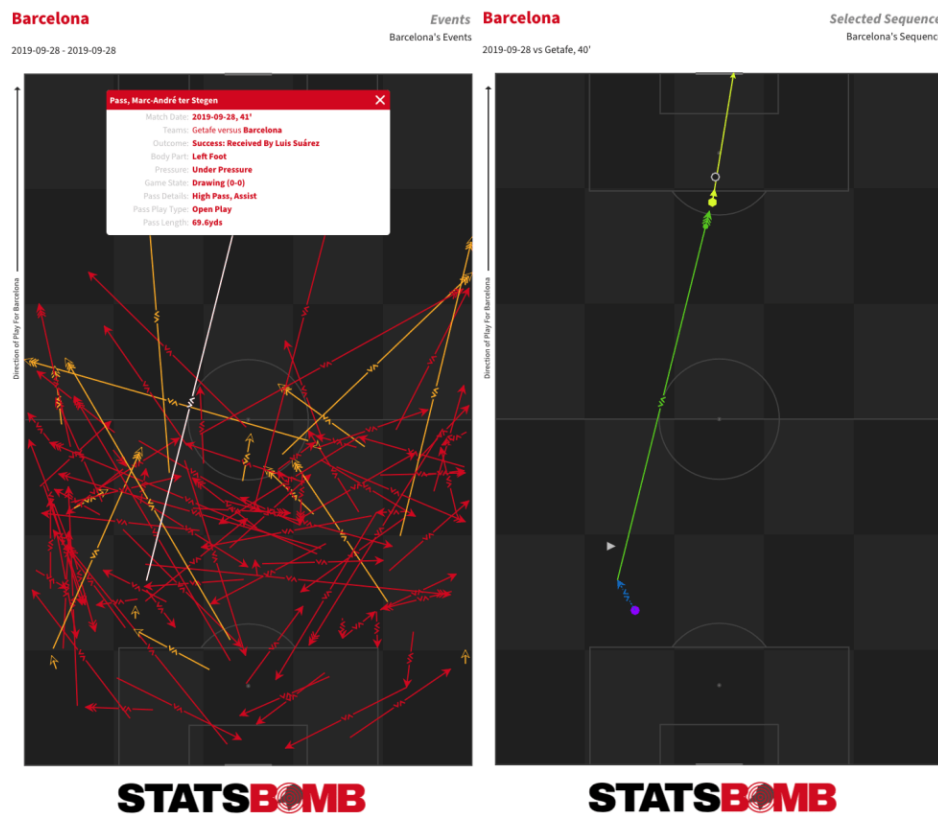
Fuente: StatsBomb, 11 de febrero de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

Figura 33: Mapas de acciones realizadas bajo presión (StatsBomb)



Fuente: StatsBomb, 11 de febrero de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

Figura 34: Mapas de acciones realizadas bajo presión (StatsBomb)



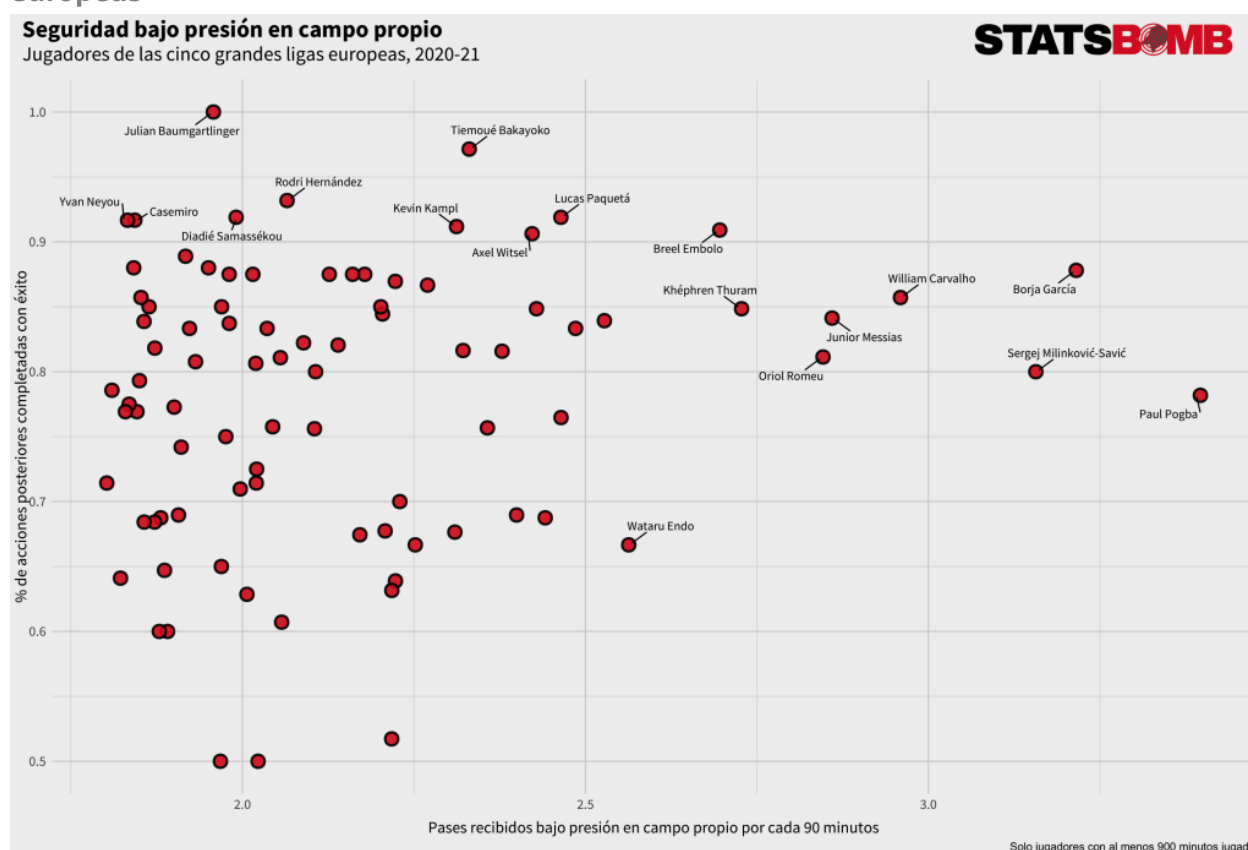
Fuente: StatsBomb, 11 de febrero de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>



En cambio, como podemos observar en la figura 34, el Barcelona sí pudo encontrar soluciones para contrarrestar la presión alta ejercida por el Getafe, pudiendo enlazar con los compañeros en el campo rival desde el campo propio. Incluso, pudo marcar un gol en una de estas conexiones gracias al portero Ter Stegen, quien asistió con un pase al espacio desde su propio campo a Luís Suarez para marcar.

Otra de las métricas ofensivas derivadas de los eventos de presión recogidos por StatsBomb es la de cómo se desenvuelven los receptores de pase ante la presión rival cuando reciben el balón bajo presión en campo propio. Así, podemos medir su capacidad de salida de balón y de darle continuidad al juego cuando lo reciben bajo presión en el campo propio. Lo podemos observar en este gráfico de dispersión (figura 35):

Figura 35: Recepción de pase bajo presión de los jugadores de las cinco grandes ligas europeas



Fuente: StatsBomb, 18 de febrero de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-jugadores-de-las-cinco-grandes-ligas-europeas/>

Pero, ¿qué tipo de conclusiones podemos sacar de este gráfico de dispersión? Ahí van algunas menciones sobre la capacidad que algunos de estos jugadores tienen para darle continuidad al juego cuando reciben bajo presión:



Algunos nombres resaltan: Paul Pogba, del Manchester United, recibe más pases bajo presión en campo propio que cualquier otro jugador, aunque pierde el balón con un poco más frecuencia que futbolistas como Borja García del Huesca o William Carvalho del Real Betis, quienes también reciben muchos pases así. Julián Baumgartlinger del Bayer Leverkusen recibe un menor volumen de estos pases (1.96 por 90), pero no ha perdido ni una vez el balón en lo que va de temporada. (StatsBomb, 18 de febrero de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-jugadores-de-las-cinco-grandes-ligas-europeas/>).

Unidad 4.4. Claves para una buena interpretación del dato

En nuestras presentaciones, hacemos uso de las visualizaciones que los proveedores nos ofrecen o de las visualizaciones que nosotros hemos creado. Pero lo más importante es cómo nosotros interpretamos el dato tras verlo en esas visualizaciones.

A ojos de diferentes personas y con diferentes puntos de vista, una misma visualización puede tener diferentes interpretaciones, tan válidas las unas como las otras. Lo que les va a dar mayor o menor validez es el razonar acerca de por qué hacemos esa interpretación, es decir, las causas o el origen de esa interpretación que, normalmente, vienen de una acumulación de hechos, acciones o situaciones de juego que se repiten durante los partidos.

Por ejemplo, en la figura 36, en el *ranking* de la izquierda vemos resaltado en amarillo que el Sevilla está último en cuanto al número de presiones totales ejercidas de La Liga 2020/21, pero, en cambio, lo vemos en 2.ª posición como el equipo que mejor porcentaje de presiones exitosas tiene (con recuperación).

Si solamente vemos el *ranking* de la izquierda (el del número de presiones ejercidas) podemos interpretar que el Sevilla es uno de los equipos que peor presiona, pero tenemos que darle contexto al dato. Lo hacemos de la siguiente manera:

- a) Explorando los datos al máximo para una mayor comprensión: no dejarnos llevar por las prisas y las impresiones a primera vista, sino esforzarnos por entender el significado del valor que se está registrando (en este caso, se toman el número de presiones realizadas, sin tener en cuenta si se recupera el balón o no). El Sevilla puede realizar una menor cantidad de presiones, simplemente porque tiene una mayor posesión que la mayoría de equipos.
- b) Contrastando el dato con otros valores relacionados con el dato en cuestión: esto nos va a dar un contexto adecuado para valorar y darle una interpretación más válida a ese dato. Por ejemplo, con la visualización del *ranking* de la derecha, le estamos aportando contexto al *ranking* de la izquierda. Viene a decirnos que el Sevilla es el equipo que menos presiones realiza, pero el segundo equipo que más balones recupera tras ejercer esa presión. Por tanto, estamos hablando de uno de los equipos que se muestra más eficaz al momento de presionar.



Figura 36: Rankings de presiones totales ejercidas y por tercios de campo en la web Fbref

Squad Stats										Successful Pressure Percentage														
Number of times applying pressure to opposing player who is receiving, carrying or releasing the ball										Percentage of time the squad gained possession with five seconds of applying pressure Minimum 6.44 pressures per squad game to qualify as a leader														
Squad	Press	Succ	%	Def 3rd	Mid 3rd	Att 3rd	Blocks	Sh	ShSv	Pass	Squad	Press	Succ	%	Def 3rd	Mid 3rd	Att 3rd	Blocks	Sh	ShSv	Pass	Int	Tkl	
Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	Eibar	5015	1602	31.9	1272	2251	1492	553	75	2	478	390		
Celta Vigo	5997	1701	28.4	2063	2850	1084	610	97	1	513	Sevilla	4716	1477	31.3	1392	2199	1125	503	95	3	408	334		
Valladolid	5839	1624	27.8	2051	2600	1188	601	117	2	484	Barcelona	5075	1546	30.5	1404	2330	1341	519	82	0	437	354		
Huesca	5451	1578	28.9	1951	2462	1038	551	103	1	448	Atlético Madrid	5124	1540	30.1	1665	2328	1131	579	93	0	486	395		
Alavés	5447	1510	27.7	1752	2541	1154	568	105	4	463	Real Sociedad	5070	1487	29.3	1423	2340	1307	497	104	2	393	337		
Levante	5363	1499	28.0	1697	2350	1316	538	108	3	430	Real Madrid	4880	1425	29.2	1562	2175	1143	535	100	4	435	406		
Osasuna	5329	1550	29.1	1558	2487	1284	557	109	2	448	Osasuna	5329	1550	29.1	1558	2487	1284	557	109	2	448	394		
Cádiz	5264	1433	27.2	2094	2304	866	612	118	4	494	Huesca	5451	1578	28.9	1951	2462	1038	551	103	1	448	348		
Betis	5263	1514	28.8	1619	2495	1149	531	84	2	447	Betis	5263	1514	28.8	1619	2495	1149	531	84	2	447	432		
Athletic Club	5250	1451	27.6	1494	2380	1376	553	86	2	467	Villarreal	4994	1425	28.5	1587	2226	1181	578	117	3	461	354		
Eibar	5168	1455	28.2	2068	2228	872	553	111	2	442	Celta Vigo	5997	1701	28.4	2063	2850	1084	610	97	1	513	375		
Atlético Madrid	5124	1540	30.1	1665	2328	1131	579	93	0	486	Eibar	5168	1455	28.2	2068	2228	872	553	111	2	442	388		
Barcelona	5075	1546	30.5	1404	2330	1341	519	82	0	437	Levante	5363	1499	28.0	1697	2350	1316	538	108	3	430	445		
Real Sociedad	5070	1487	29.3	1423	2340	1307	497	104	2	393	Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	420		
Granada	5027	1358	27.0	1616	2345	1066	611	142	5	469	Valladolid	5839	1624	27.8	2051	2600	1188	601	117	2	484	390		
Eibar	5015	1602	31.9	1272	2251	1492	553	75	2	478	Alavés	5447	1510	27.7	1752	2541	1154	568	105	4	463	432		
Villarreal	4994	1425	28.5	1587	2226	1181	578	117	3	461	Athletic Club	5250	1451	27.6	1494	2380	1376	553	86	2	467	348		
Real Madrid	4880	1425	29.2	1562	2175	1143	535	100	4	435	Cádiz	5264	1433	27.2	2094	2304	866	612	118	4	494	411		
Valencia	4855	1305	26.9	1794	2067	994	566	137	2	429	Granada	5027	1358	27.0	1616	2345	1066	611	142	5	469	406		
Sevilla	4716	1477	31.3	1392	2199	1125	503	95	3	408	Valencia	4855	1305	26.9	1794	2067	994	566	137	2	429	321		

Fuente: adaptación propia de la web de Fbref.com, s. f., <https://fbref.com/en/comps/12/La-Liga-Stats>

En la figura 37, Fredi Martín nos aporta una serie de claves —con las que estamos muy de acuerdo— para tratar y darle al dato una interpretación correcta.

Figura 37: Claves para una correcta interpretación de los datos

- Comprender el deporte del que estoy tratando
- Explorar los datos para una mejor comprensión
- Contextualizar los datos
- Normalizar los datos
- Concretar el mensaje que se quiere transmitir con los datos
- Realizar una correcta presentación / visualización de los datos
- Ser conciso
- Comprender las métricas avanzadas: xG, VAEP, xT
- Mis conclusiones pueden ser diferentes a la de otros destinatarios dejar espacio a la reflexión

Fuente: [Imagen sin título sobre las claves para una correcta interpretación de los datos]. (s. f.). Recuperado https://lh3.googleusercontent.com/D3OigKfmY4FIR3Zv_0IF5HEm-PM2akwwbRIOR3d9me5JMY8UKXVB9U_v4W0rmnWZLiyUxv4=s109



De todas estas claves, vamos a comentar algunas de ellas. Podemos concluir que, para ofrecer una buena interpretación del dato en nuestras presentaciones, hay que

- a) Comprender fútbol como deporte en toda su dimensión: cuanto más sepamos, más contexto podemos darle al dato y, por tanto, una mejor interpretación.
- b) Explorar los datos al máximo: para tener una visión más amplia del espectro a estudiar (las presiones, etc.). Esto nos dará la posibilidad de tener una mejor comprensión del dato y, por ende, le podremos dar una mejor interpretación.
- c) Contextualizar el dato: hay que buscar y encontrar las razones y los porqués de la interpretación que le estamos dando al dato. Por ejemplo, los datos demuestran que Marcos Llorente tiene más goles y asistencias en el Atlético de Madrid que cuando jugaba en el Real Madrid. Para darle contexto a ese o esos datos, tenemos que hacer las preguntas y las hipótesis adecuadas: ¿Por qué Llorente cuando jugaba en el Real Madrid no pisaba tanto el área rival y no marcaba ni asistía tanto como lo hace en el Atlético de Madrid bajo las órdenes del Cholo Simeone? Una razón puede ser porque el entrenador de turno le pide cosas diferentes (que dé más equilibrio, que no vaya tanto al ataque, porque ya tiene otros jugadores que finalizan la jugada, etc.) o, quizás, porque el jugador demuestra otra mentalidad y ha evolucionado ofensivamente para tener más presencia en área. Puede haber varias respuestas a las preguntas. Se trata de cotejarlas y tenerlas en cuenta para dar con una interpretación lo más cercana a la realidad que se pueda.
- d) Normalizar los datos: cuando tengamos en frente visualizaciones de datos defensivos de algunos equipos o jugadores; tales como las interceptaciones, entradas, duelos, etc., hay que tener bien presente que, para hacer una buena interpretación de esos datos, necesitamos normalizarlos. Es decir, hacer un ajuste de esas métricas defensivas en función de la posesión. El estilo de juego de los equipos afecta a los datos que se registran en valores absolutos:

Los jugadores solo pueden realizar acciones defensivas cuando su equipo no tiene la posesión. Por tanto, en función del tiempo que el equipo esté en posesión, tendrán más o menos oportunidades de sumar acciones defensivas. Esto hace que los equipos con volúmenes elevados de posesión tengan menor número de acciones defensivas y, por tanto, los números absolutos de *tackles*, interceptaciones (o presiones) tienen más ruido que lo deseable. (Peña Rodríguez, P. 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

El ajuste en función de la posesión es una solución práctica para estandarizar y normalizar los valores de modo que se puedan comparar entre diferentes



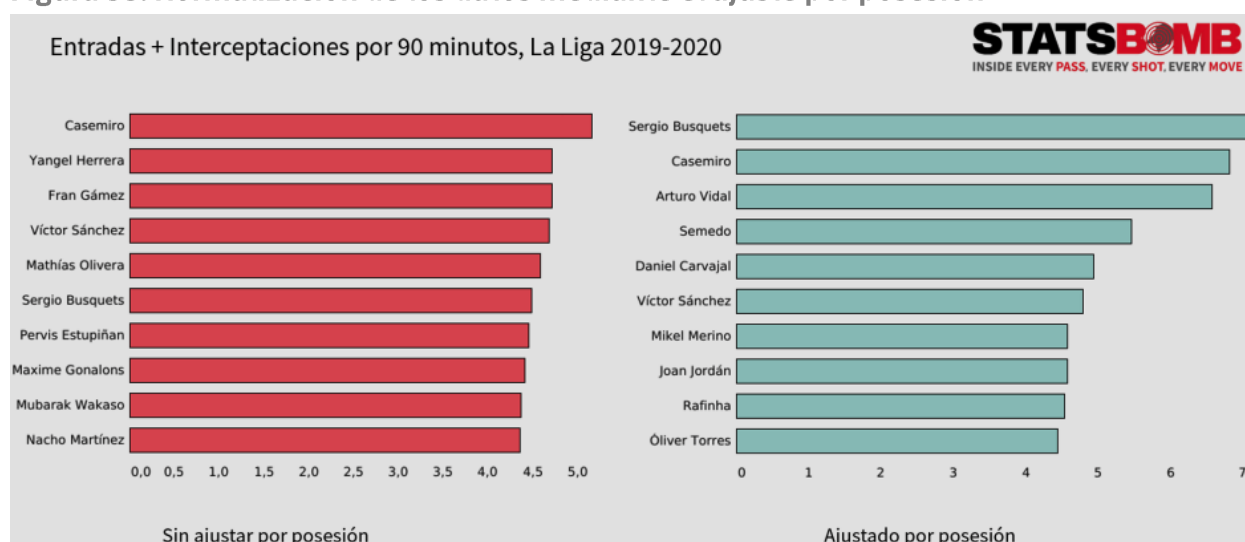
jugadores o equipos, corrigiendo aspectos ajenos a su habilidad o estilo de juego respectivamente:

Sin embargo, es importante remarcar que, una vez ajustada por posesión, ya no existe una correspondencia verídica entre el valor de la métrica y las acciones reales que podamos ver en vídeo. Es por ello que los números absolutos siguen teniendo utilidad a la hora de evaluar el rendimiento de los jugadores (por ejemplo, para definir los límites). (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

En la figura 38, vemos un *ranking* de las entradas e interceptaciones (x90 minutos) de los jugadores de La Liga. En la izquierda, tenemos los datos absolutos, mientras que, en la derecha, tenemos los datos ajustados en función de la posesión de sus equipos.

Como podemos observar, jugadores como Casemiro ni aparecen en el *ranking* de valores absolutos, simplemente por el hecho de que tienen menos oportunidades de sumar acciones defensivas, ya que tienen mucho más tiempo el balón en su poder que jugadores de otros equipos que tienen menos posesión y se ven obligados a defender mucho más tiempo, con lo que tienen más opciones de sumar eventos defensivos.

Figura 38: Normalización de los datos mediante el ajuste por posesión



Fuente: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

- e) Realizar una buena presentación/visualización de los datos, así como concretar el mensaje y ser conciso: ya sabemos que la sencillez de las visualizaciones que



presentemos, así como la brevedad y la claridad de nuestras opiniones, van a facilitar su entendimiento por parte de del receptor.

- f) Dejar espacio a la reflexión, sabiendo que las opiniones de los demás pueden ser diferentes: la subjetividad que el fútbol lleva implícita en su ADN también se refleja en los datos que genera (si una misma acción del juego puede ser juzgada de manera diferente por cada uno de los interlocutores que ven esa acción, ¿cómo no va pasar con los datos que el propio juego genera?). Por esa razón, debemos ser respetuoso con las opiniones de los demás, siempre y cuando esas opiniones tengan un fundamento tan trabajado y conseguido como el nuestro, tras haber llevado a cabo todos los procesos anteriormente mencionados y habiendo aplicado todas estas claves para su buena interpretación.

Conclusiones

Es evidente que el fútbol ha evolucionado a lo largo de su historia y siempre lo hará. Pero en la faceta de la aplicación del *big data* y la inteligencia artificial (AI), el fútbol ha entrado en una espiral de modernización que nos obliga a entrenadores, asistentes, directores deportivos, y todo aquel que quiera formar parte de su entorno a formarnos para poder estar al día y entenderla. Esta cada vez cobra mayor importancia en los procesos de mejora del rendimiento de los futbolistas y del juego de los equipos.

Dicho esto, es importante tener nociones de fútbol, tanto en su dimensión técnico-táctica condicional como en la del proceso de entrenamiento y, también, en la del análisis de los datos que genera. Los datos se utilizan cada vez más por los clubes y cuerpos técnicos, quienes dan cabida a profesionales para que puedan realizar un análisis útil y provechoso para mejorar la toma de decisiones diaria que afecta a todos los ámbitos comentados con anterioridad.

De ahí la importancia de conocer este mundo emergente del dato, para no quedar fuera de juego, para sacarle provecho y mejorar tanto nuestras prestaciones profesionales como el rendimiento del equipo.

El conocimiento de los procesos de tratamiento y visualización de los datos defensivos presentados en este módulo 4 servirá para mejorar nuestras habilidades como analistas y/ o entrenadores al descifrar el juego. Lo aprendimos a hacer mediante el video-análisis y ahora tenemos que incorporar el dato como un elemento más en el que apoyarnos para ampliar nuestro punto de vista y dotarlo de bidireccionalidad:

- a) Ir del juego al dato: apoyarnos en las visualizaciones del dato para reforzar las opiniones que hemos conseguido al analizar la jugada mediante el video-análisis y observar las conductas, tanto propias como del rival.

- b) Ir del dato al juego: conseguir información relevante sobre el juego mediante los datos que difícilmente podríamos conseguir con el video-análisis (por falta de tiempo, de atención o, simplemente, por el volumen alto de partidos que queremos cotejar). Por ejemplo, conseguir todos los pases al espacio realizados por un equipo desde su propio campo en todos los partidos de La Liga, etc.

Como despedida de este curso, solo quiero decirles que ha sido todo un reto para mí volcar en estos apuntes todo el conocimiento que tengo sobre el tema de la presión alta y compartirlo con todos ustedes.

Me considero muy afortunado de haber aprendido de entrenadores y compañeros de profesión tan buenos y preparados. Para mí, es todo un placer compartir con ustedes todos estos conocimientos, que son fruto de mi formación y de mis experiencias previas.

Espero que este curso les haya gustado y les sea útil a nivel profesional.

Espero verlos a todos y todas próximamente, trabajando en un cuerpo técnico profesional, si es que no lo están haciendo todavía.



Referencias

De Juana, R. (2019). Tableau: así es la empresa por la que Salesforce ha pagado más de 13.000 millones [Publicación en el blog Muy Computer Pro]. Recuperado de <https://www.muycomputerpro.com/2019/06/11/tableau-asi-es-la-empresa-por-la-que-salesforce-ha-pagado-mas-de-13-000-millones>

Deutsche Fussball Liga (DFL). (2020). DFL and Amazon Web Services to provide new real-time match analysis [Publicación en la web de Deutsche Fussball Liga]. Recuperado de <https://www.dfl.de/en/news/dfl-and-amazon-web-services-to-provide-new-real-time-match-analysis/>

Fbref. (s. f.). League Table [Publicación en la web de Fbref]. Recuperado de <https://fbref.com/en/comps/12/La-Liga-Stats>

Fombella Pombal, D. (2020a). Módulo 4. Análisis de datos deportivos con R y Python [Apunte de clase]. Máster en *Big Data* Deportivo, Universidad Católica San Antonio de Murcia.

Fombella Pombal, D. (2020b). Módulo 7. Proveedores de datos deportivos – Opta, InStat, Wyscout, Mediacoach, STATS, StatsBomb [Apunte de clase]. Máster en *Big Data* Deportivo, Universidad Católica San Antonio de Murcia.

[Imagen sin título sobre herramientas de Pentaho]. (s. f.). Recuperado de https://lh3.googleusercontent.com/4y-ZxvheJkzXhDva8cLQ0AccgCwF1wFHRD9PP_SWnRQarn_OIyVU5Ktwc1cgrJY71bYvxQ=s130

[Imagen sin título sobre la capacidad de presión de un equipo]. (s. f.). Recuperado de https://lh3.googleusercontent.com/dilrueUqUvtmDf4VyI_rTuVSwgSi_81ZkBzbaifEcEz6YIAGxBIQR04xmOKSxsmbllqUPw=s85

[Imagen sin título sobre las cuatro variables que miden la capacidad de presión de un jugador]. (s. f.). Recuperado de <https://lh3.googleusercontent.com/D56U0KRjGbQ4I1Y-E73kKoKHqs-wrtDf0tL6UhOQcERi7P8NwOIZpjYge1bGLg6QEML-t7M=s85>

[Imagen sin título sobre la presión tras la pérdida]. (s. f.). Recuperado de https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKKC6604tLENRDvJ7dDv1wYjnkHuYR_cmgkCmqNiv-TDYcyr7-HkM=s114

[Imagen sin título sobre las claves para una correcta interpretación de los datos]. (s. f.). Recuperado de https://lh3.googleusercontent.com/D30igKfmY4FlR3Zv_OIF5HEm-PM2akwwbRIOR3d9me5JMY8UKXVB9U_v4W0rmnWZLiyUxv4=s109

[Imagen sin título sobre PPDA en un *ranking* de StatsBomb]. (s. f.). Recuperado de https://lh3.googleusercontent.com/4QRdqGWNi0_bJ_Zs392Z0mH96LDHnIenpyl145mhA9-n0sTvqJTvUzR4NS2CSroCXb0U=s85

[Imagen sin título sobre *ranking* de presión en el campo rival]. (s. f.). Recuperado de <https://lh3.googleusercontent.com/CFilwr8zUC-i9Gk-kyX5x-NEZRYcgj2QH4WoN63PAWrea3qbb1Avq4JGoepUMUvt1feuA=s85>

López, L. (2019). Posesión, pero con pegada: así es el ADN del campeón de La Liga Santander. *El mundo* [Versión *online*]. Recuperado de <https://noesfutboleslaliga.elmundo.es/directos-al-futuro/posesion-pero-con-pegada-asi-es-el-adn-del-campeon-de-laliga-santander>

Microsoft. (2021). Power BI Desktop [*software* de análisis de datos]. Redmond, US: Microsoft.

Olmeda, M. (2020). La tecnología que ayuda a los equipos a preparar los partidos. *Diario ABC* [Versión *online*]. Recuperado de <https://www.abc.es/contentfactory/post/eslaliga/laliga-apuesta-por-la-tecnologia-para-hacerse-mas-competitiva/>

Ovalle, C. (2021). Qué dicen los datos: Bravo fue el tercer arquero que más goles evitó en La Liga la temporada pasada. *Diario BioBio* [Versión *online*] Recuperado de <https://www.biobiochile.cl/noticias/futbol-internacional/chilenos-en-el-exterior/2021/05/27/que-dicen-los-datos-bravo-fue-el-tercer-arquero-que-mas-goles-evito-en-la-liga-la-temporada-pasada.shtml>

Oviedo Fernández, F. (2020). Módulo 1. Introducción al *big data* y teorías: teoría de los juegos y teoría de la toma de decisiones [Apunte de clase]. Máster en *Big Data* Deportivo, Universidad Católica San Antonio de Murcia.

Peña Rodríguez, P. (2020). Introducción a las métricas defensivas en fútbol. Recuperado de <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

Pérez, F. (2021). Proyecto Jupyter [*software* de lenguajes de programación]. Berkeley, US: Proyecto Jupyter.

RStudio Inc. (2021). RStudio [*software* para lenguaje de programación]. Boston, US: RStudio Inc.

Second Spectrum. (s. f.). Win more games [Publicación en la web de Second Spectrum]. Recuperado de <https://www.secondspectrum.com/ourwork/teams-leagues.html>

SoftwarePara. (s. f.). Tableau [Publicación en el blog *SoftwarePara*]. Recuperado de <https://softwarepara.net/tableau/>



StatsBomb. (11 de febrero de 2021). ¿Cómo se desenvuelven ante presión rival los equipos de la liga? [Publicación en la web de StatsBomb]. Recuperado de <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

StatsBomb. (18 de febrero de 2021). ¿Cómo se desenvuelven ante presión rival los jugadores de las cinco grandes ligas europeas? [Publicación en la web de StatsBomb]. Recuperado de <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-jugadores-de-las-cinco-grandes-ligas-europeas/>

StatsBomb. (s. f.). StatsBomb 360 [Publicación en la web de StatsBomb]. Recuperado de <https://StatsBomb.com/360-data/>

Unir. (2019). Lenguaje R, ¿qué es y por qué es tan usado en *Big data*? [Publicación en el blog Unir]. Recuperado de <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>

Web-Futbol. (s. f.). *Big data* en el fútbol [Publicación en el blog Web-Futbol]. Recuperado de <https://www.web-futbol.com/big-data-en-el-futbol/>

Wyscout. (s. f.). Datos, videos y productos: una combinación perfecta [Publicación en la web de Wyscout]. Recuperado de <https://wyscout.com/es/plataforma-futbol/>

