

Módulo 4. Dados para a análise da pressão alta no futebol

Introdução

O principal objetivo deste módulo 4 é entender quais dados em nível tático defensivo são úteis para medir a capacidade de pressão alta de uma equipe. Neste módulo, pretendemos entender como os dados podem ajudar os treinadores a tomar decisões ótimas em nível defensivo.

Para isso, faremos uma breve introdução sobre o que é o big data aplicado ao futebol, quais fornecedores de dados existem, de quais deles podemos extrair dados defensivos de pressão, quais dados serão úteis para medir a capacidade de pressão alta de uma equipe e como podemos interpretá-los.

Vamos abordar os seguintes pontos:

- Introdução ao big data no futebol.
- Principais fornecedores de dados existentes no futebol. De onde podemos obter dados defensivos de pressão?
- Quais dados nos servirão para medir a capacidade de pressionar alto de uma equipe?
- Chaves para uma boa interpretação dos dados.

Unidade 4.1. Introdução ao big data no futebol

O uso de big data e inteligência artificial no futebol cresceu muito nos últimos anos, e ambos estão se tornando ferramentas muito importantes para obter melhores resultados e otimizar o desempenho de equipes e atletas profissionais.

A aplicação do big data esportivo no futebol busca gerar uma vantagem em relação ao adversário, tanto na tática e estratégia a serem utilizadas nas partidas quanto no mercado de transferências ou no cuidado com os jogadores.

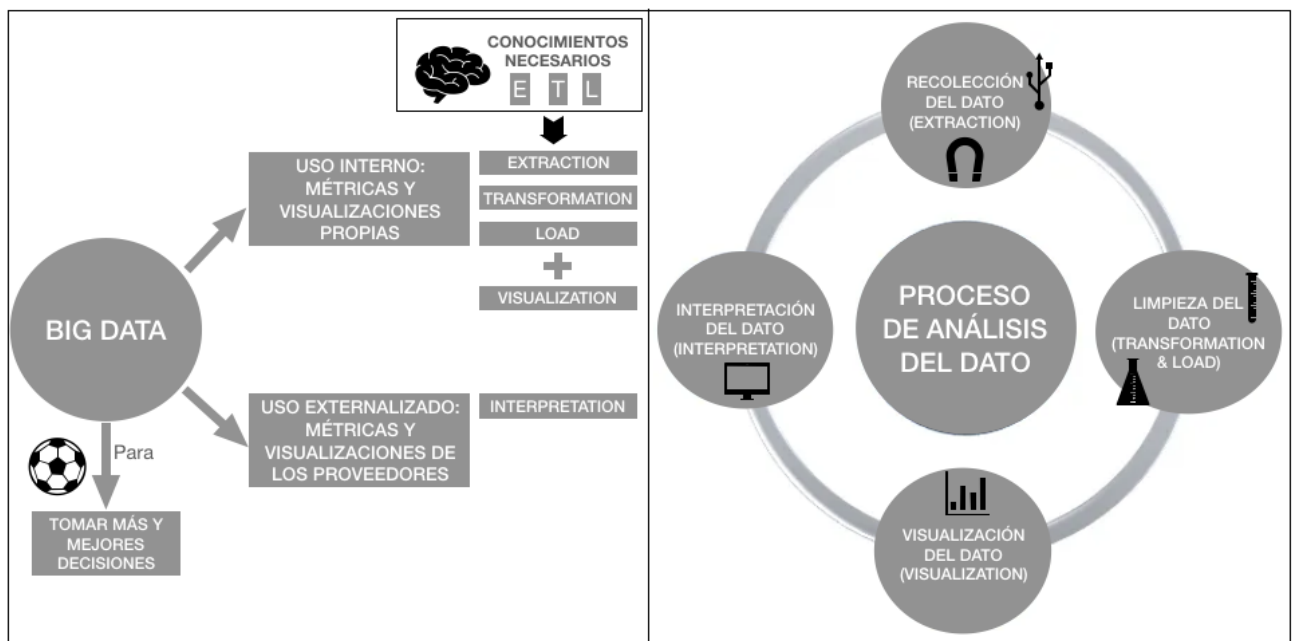
Em uma partida de futebol, são capturados cerca de oito milhões de dados sincronizados com o tempo de jogo, que registram informações geradas pelos jogadores e pela bola no campo. Essas informações — que não são visíveis a olho nu, pois o olho humano é capaz de reter apenas 30% do que acontece durante o jogo — são armazenadas e tratadas, através de ferramentas de big data, para serem utilizadas e consultadas para tomar decisões econômicas, táticas, físicas, etc.

Por isso, os clubes e treinadores usam essas tecnologias para melhorar a tomada de decisões. Isso é ainda mais comum no futebol profissional, que dispõe de mais recursos para acessar esse tipo de ferramenta e contratar pessoal capacitado para utilizá-las e aproveitá-las de forma mais eficiente. (Web-Futbol, s. f., <https://www.web-futbol.com/big-data-en-el-futbol/>).

Mas como uma equipe técnica pode usar o big data? Os treinadores e analistas podem utilizar os dados gerados pelo big data no futebol de duas maneiras:

- **Uso terceirizado:** contratam os serviços de um ou mais fornecedores de dados, como StatsBomb, Opta, Instat, Wyscout, etc., que, em suas plataformas de análise, oferecem métricas e visualizações de dados fechados e com seu próprio formato, que podem ser consultados, baixados e usados para análises e apresentações. Esse modo de uso dos dados não exige muitos conhecimentos em big data.
- **Uso interno:** gera suas próprias métricas e visualizações a partir de dados coletados de vários fornecedores. Os fornecedores mencionados anteriormente também oferecem a opção de baixar os dados brutos (raw data) por meio de suas APIs (Application Programming Interface) para depois tratá-los com programas de linguagem de programação (R ou Python) e criar uma visualização personalizada com programas como Power BI ou Tableau. Esse método de geração e visualização de informações exige um alto conhecimento dos processos relacionados ao big data e à ETL dos dados (extração, transformação e carga).

Figura 1: Uso do big data por parte de uma equipe técnica



<p>Tomar más y mejores decisiones - Uso interno: métricas y visualizaciones propias - Uso externalizado: métricas y visualizaciones de los proveedores - Interpretación - Conocimientos necesarios: Extracción, Transformación, Load + Visualización.</p> <p>Proceso de análisis del dato - Recolección del dato - Interpretación del dato - Visualización del dato - Limpieza del dato.</p>	<p>Tomar mais e melhores decisões - Uso interno: métricas e visualizações próprias - Uso externalizado: métricas e visualizações dos fornecedores - Interpretação - Conhecimentos necessários: Extração, Transformação, Carga (ETL) + Visualização.</p> <p>Processo de Análise de Dados - Coleta de dados - Interpretação dos dados - Visualização dos dados - Limpeza dos dados</p>
--	---

Fonte: elaboração própria.

As considerações a serem levadas em conta para as equipes técnicas que não têm tantos recursos financeiros para contratar os serviços de um fornecedor são:

Por um lado, existem plataformas de ligas profissionais e federações que facilitam o acesso gratuito a imagens e dados de tracking e eventing (raw data), desde que o clube em questão jogue em ligas profissionais como La Liga espanhola (Mediacoach), a Bundesliga alemã (Match Facts) ou a Premier League inglesa (DVMS).

Por outro lado, existem diversos sites de consulta (Understat, Fbref, WhoScored, Transfermarkt, etc.) que oferecem, de forma gratuita, acesso a dados de eventing que podemos baixar para trabalhá-los à nossa maneira através do web scraping.

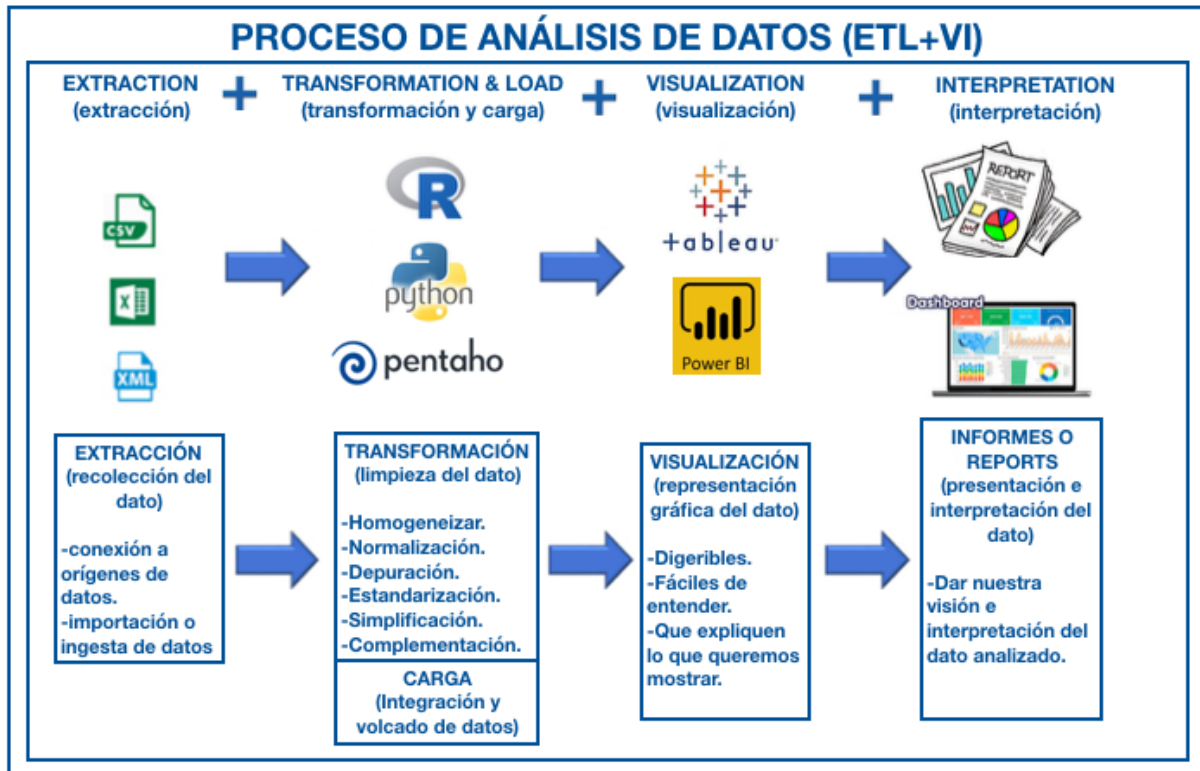
Em definitiva, o mais importante e o que vai fazer a diferença para um bom uso do big data aplicado ao futebol é ter tanto conhecimentos sobre futebol quanto sobre big data e os processos de tratamento de dados ETL, aos quais eu adicionaria a visualização e a interpretação dos dados:

Extraction (extração): coleta, importação ou ingestão dos dados.
Transformation (transformação): limpeza e homogeneização dos dados.
Load (carga): integração e upload dos dados.
Visualization (visualização): representação gráfica dos dados.
Interpretation (interpretação): apresentação e interpretação dos dados.

Para poder aplicá-los na nossa rotina de trabalho diário e para que seja realmente uma ferramenta útil para gerar um conhecimento aplicável ao nosso trabalho como treinadores e/ou analistas, devemos ter conhecimento sobre o processo de análise de dados. O big data:

Refere-se ao tratamento e análise de grandes quantidades de dados com o objetivo de extrair valor dessa informação. Essa gestão e análise de dados não podem ser tratadas por técnicas convencionais, mas sim por técnicas como o reconhecimento de padrões (machine learning) ou análises preditivas para prever mudanças nos volumes de dados. (Oviedo, 2020, p. 6).

Figura 2: Como usar o big data?



<p>Extracción</p> <ul style="list-style-type: none"> - Conexión a orígenes de datos - Importación o ingesta de datos <p>Transformación</p> <ul style="list-style-type: none"> - Homogeneizar - Normalización - Depuración - Estandarización - Simplificación - Complementación - Carga (Integración y volcado de datos) <p>Visualización</p> <ul style="list-style-type: none"> - Digeribles - Fáciles de entender - Que expliquen lo que queremos mostrar <p>Informes o reportes (presentaciones e interpretaciones del dato)</p> <ul style="list-style-type: none"> - Dar nuestra visión e interpretación del dato analizado 	<p>Extração</p> <ul style="list-style-type: none"> ● Conexão com fontes de dados ● Importação ou ingestão de dados <p>Transformação</p> <ul style="list-style-type: none"> ● Homogeneização ● Normalização ● Limpeza ● Padronização ● Simplificação ● Complementação ● Carga (Integração e Armazenamento de Dados) <p>Visualização</p> <ul style="list-style-type: none"> ● Compreensíveis ● Fácéis de entender
---	---

	<ul style="list-style-type: none"> • Que expliquem o que queremos mostrar <p>Relatórios (Apresentações e Interpretações dos Dados)</p> <ul style="list-style-type: none"> • Dar nossa visão e interpretação dos dados analisados
--	---

Fonte: elaboração própria.

Se queremos fazer um uso interno e próprio do processo de análise —mediante o filtrado da informação na forma de dados que vêm de diferentes fornecedores e dando-lhe a visualização que desejamos—, precisamos ter conhecimentos básicos sobre os processos de análise de big data e as ferramentas necessárias para sua aplicação:

Conhecimentos de extração de dados: conhecer e saber extrair das plataformas dos diferentes fornecedores os diversos formatos de arquivos que contêm a informação e que são importáveis. Esses formatos de arquivo podem ser: .csv, .excel, .xml e .json. É importante conhecer os fornecedores e os dados que nos interessam de cada um deles.

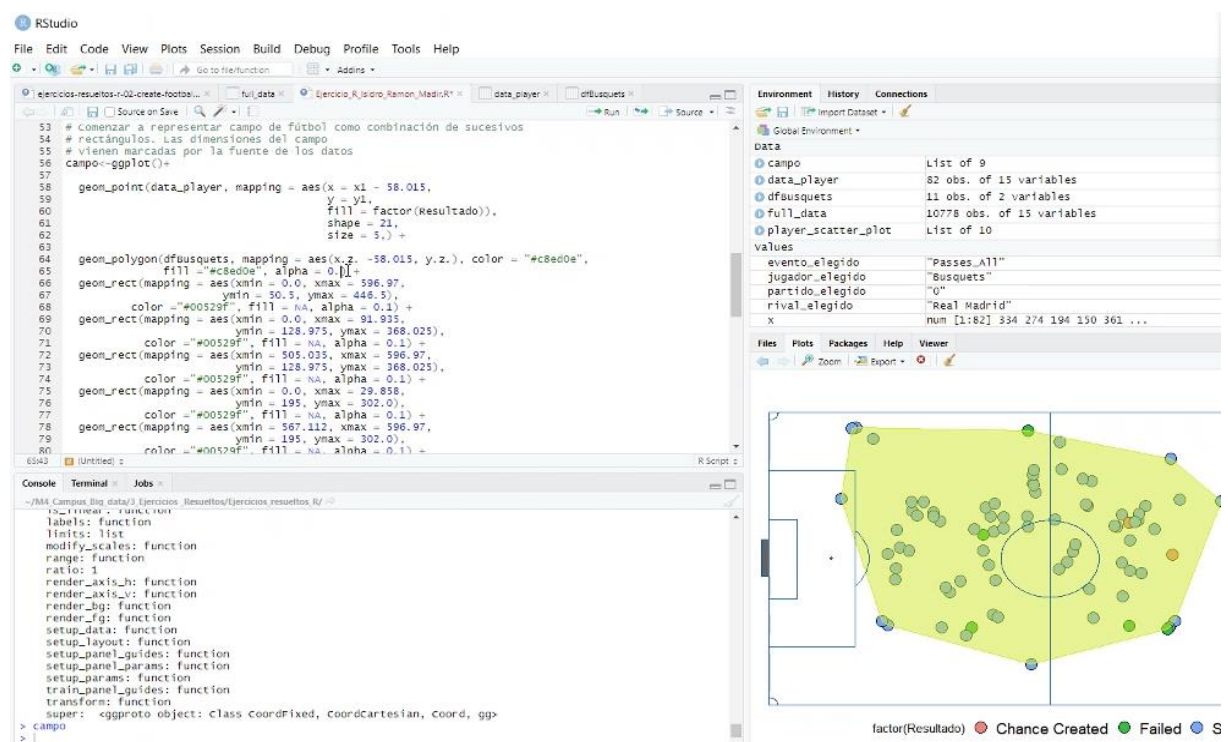
Conhecimentos de transformação e carga de dados: conhecer e saber usar as ferramentas de linguagem de programação, como R e Python, ou as ferramentas de integração de dados, como Pentaho, que são as mais utilizadas atualmente para extrair, transformar e carregar os dados desde o formato proposto pelos fornecedores para o formato que mais nos interessa:

2.1. R: é uma linguagem de programação gratuita. Este ambiente é comumente utilizado para computação estatística e gráfica e funciona nos sistemas Unix, Linux, Windows e MacOS.

As características e diferentes aplicações de R tornam-no uma ferramenta básica para os analistas de dados. Ele é utilizado em todas as fases da análise de dados:

- Aquisição dos dados das fontes disponíveis: bancos de dados, arquivos de texto, etc.
- Preparação dos dados: eliminação de duplicados, dados incorretos, valores extremos, etc.
- Análise dos dados: construção de modelos preditivos, de classificação, de agrupamento, etc.
- Comunicação dos resultados: elaboração de relatórios para apresentação dos resultados e conclusões.
- Aplicação dos resultados obtidos: por exemplo, utilização de modelos preditivos desenvolvidos para, com base em uma série de dados históricos (dados de treinamento e teste do modelo), prever certos resultados. (Unir, 2019, <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>).

Figura 3: Exemplo de interface em R



Fonte: captura de tela do software RStudio.

2.2. Python: é uma linguagem de programação gratuita de alto nível, que utiliza o conhecido “linguagem de script” e é funcional para quase tudo. A grande maioria das aplicações informáticas e plataformas web dependem de sua linguagem, pois é utilizada pela maioria dos desenvolvedores, cientistas de dados, engenheiros de software e até hackers, devido à sua versatilidade, flexibilidade e características orientadas a objetos. A linguagem Python pode ser executada em diferentes sistemas operacionais como Unix, Linux, MacOS e Windows. É importante aprender a manuseá-la, especialmente nos campos emergentes da ciência de dados, como análise de dados (big data), inteligência artificial e machine learning.

Através das diversas bibliotecas que conectam o Python com o mundo exterior, essa linguagem pode realizar as seguintes funções:

- Interagir com o mundo exterior, para o que será necessário ler e escrever dados em diferentes formatos (.csv, .xml, .excel, .json, .txt, .dat, binário, etc.).
- Preparar os dados por meio de limpeza e gerando dados estruturados a partir de dados não estruturados.
- Transformação dos dados aplicando operações matemáticas e/ou estatísticas a certos conjuntos de dados para obter como resultado novos dados (operações de agregação, segregação e filtragem).
- Apresentação dos dados e resultados obtidos por meio de gráficos estáticos e dinâmicos (com JavaScript), mapas de localização, grafos, etc.

- Modelos de aprendizado de máquina, nos quais os dados se conectam a algoritmos de aprendizado automático —supervisionados ou não— para a obtenção de modelos analíticos que podem prever —com maior ou menor grau— o futuro. (Fombella Pombal, 2020a, p.10).

Figura 4: Exemplo de interface em 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 funcion 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')[0])
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('11','').replace('2','').replace('1 ',''))
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('1/',''))
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.split('/')[0])
           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 modifiques 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

```

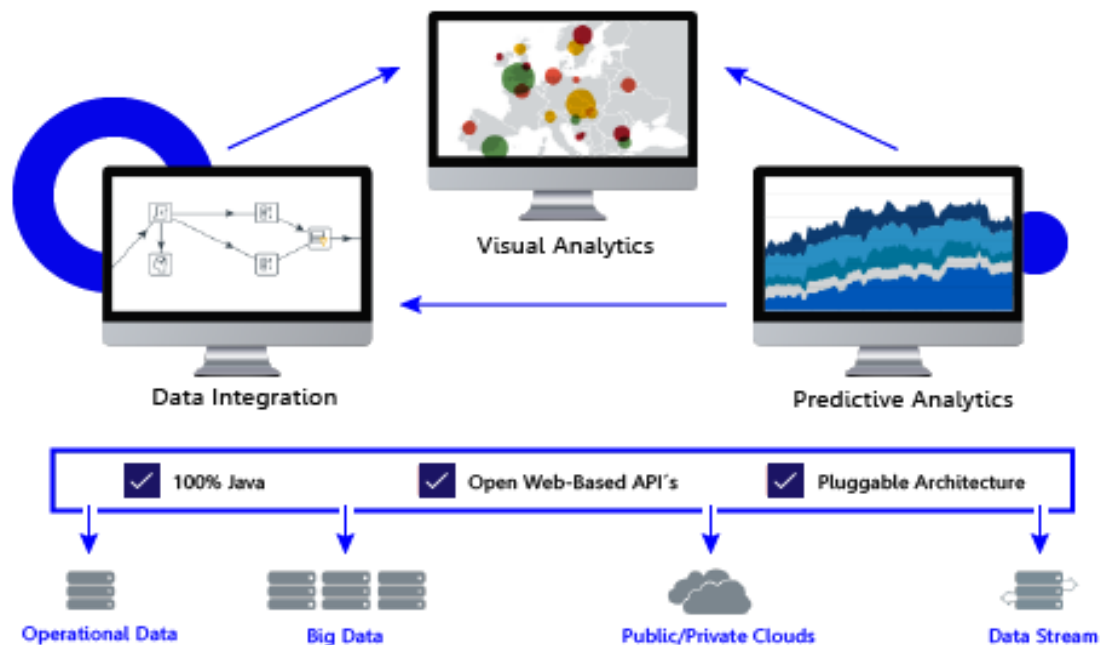
Fonte: captura de tela do software Jupyter (Pérez, 2021).

2.3. Pentaho: pertence à empresa Hitachi Vantara e é uma plataforma de business intelligence (BI) voltada e centrada na solução de processos, que inclui os componentes necessários para implementar soluções baseadas em processos como mineração de dados, ETL, geração de relatórios, etc. É muito utilizada para melhorar a capacidade de análise e tomada de decisões, pois oferece soluções compostas principalmente por uma infraestrutura de ferramentas de análise e relatórios integrados.

O Pentaho inclui em sua suíte ferramentas que cobrem as seguintes áreas:

- Pentaho Data Integration: ferramenta de integração de dados (ETL).
- Pentaho Reporting: criação de relatórios tradicionais.
- Pentaho Dashboard Designer: componente disponível apenas na versão Enterprise.
- Pentaho Community Dashboard Editor: plugin para criação de painéis de controle.
- O restante das ferramentas de criação de relatórios e visualização foi substituído em muitos casos por ferramentas de última geração, como Tableau ou Microsoft Power BI, que veremos mais adiante. (Fombella Pombal, 2020, pp. 42-44).

Figura 5: Ferramentas do Pentaho



Fonte: [Imagem sem título sobre ferramentas do Pentaho]. (s. f.). Recuperado de https://lh3.googleusercontent.com/4y-ZxvheJkzXhDva8cLQOAccgCwF1wFHRD9PP_SWnRQarn_OIyVU5Ktwc1cgrJY71bYvxQ=s130

Resumindo, o Pentaho é uma ferramenta muito útil para integrar e unificar dados provenientes de diferentes formatos e fornecedores em um formato único, permitindo uma ETL o mais rápida e eficaz possível, para posteriormente trabalhar com as ferramentas de visualização.

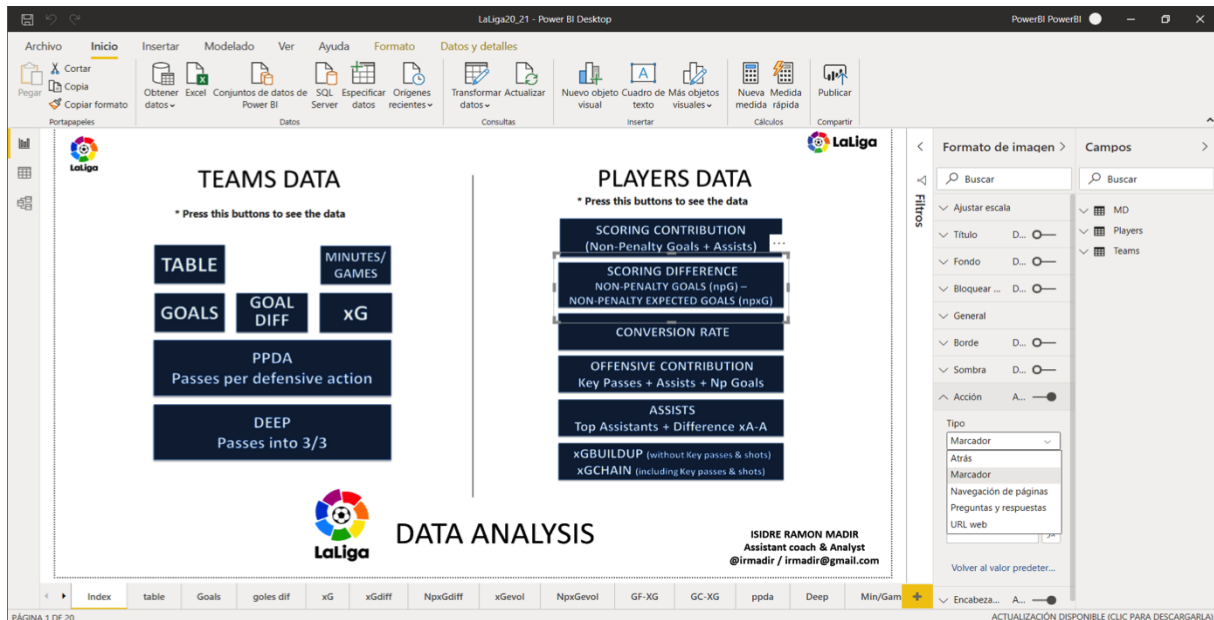
3) Conhecimento sobre a visualização de dados: conhecer e saber usar as ferramentas de visualização mais potentes do mercado. São elas Power BI e Tableau:

3.1. Power BI: o Microsoft Power BI é a principal ferramenta de visualização dentro do Microsoft Office 365. Está disponível apenas para sistemas operacionais Windows. Seu elevado número de usuários e grande aceitação no mercado se deve à facilidade de uso e ao baixo custo. Será possível criar painéis interativos de maneira muito ágil.

Em qualquer ambiente (empresas, clubes esportivos, etc.), é imprescindível unificar todas as informações disponíveis, tanto na nuvem quanto localmente, em torno de uma plataforma única. Para isso, o Power BI é uma ferramenta ideal para aplicar os conhecimentos de business intelligence (BI) na exploração de dados em tempo real a partir de diferentes fontes, além da criação de relatórios e painéis para oferecer, posteriormente, de maneira simples, visual e

intuitiva, os resultados de todos os elementos que desejamos analisar e compartilhar entre os profissionais de um mesmo clube.

Figura 6: Exemplo de visualização no Power BI



Fonte: captura de tela do software Power BI Desktop (Microsoft, 2021).

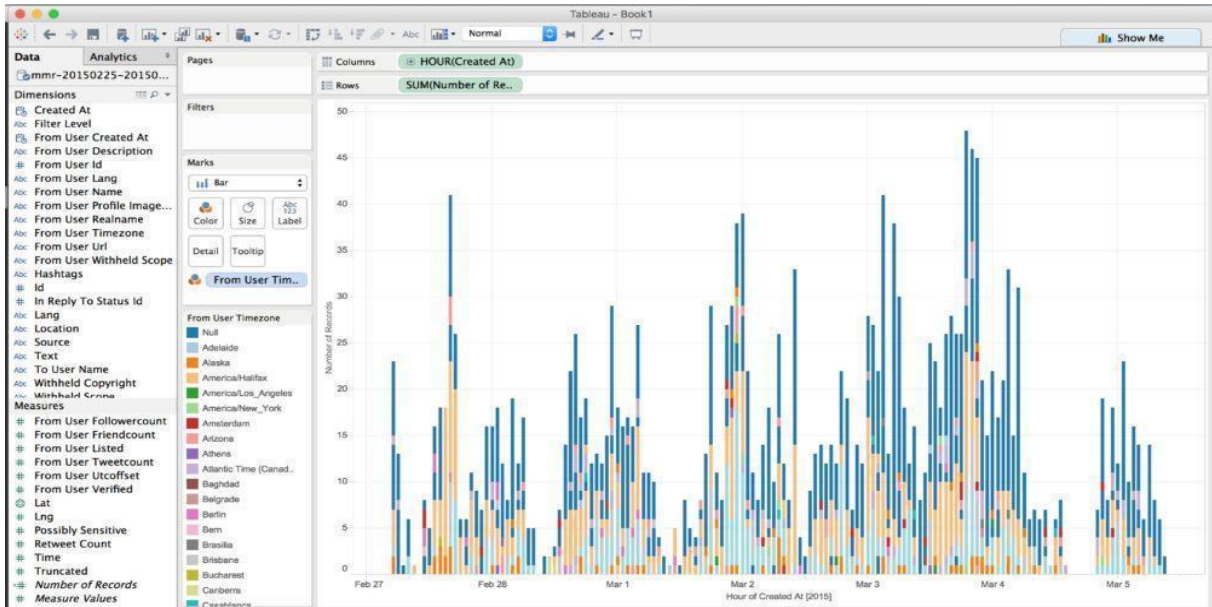
3.2. Tableau: é uma ferramenta poderosa de visualização de dados, utilizada na área de inteligência de negócios (business intelligence). Simplifica dados brutos em um formato muito fácil de entender.

A essência do Tableau é simples e, ao mesmo tempo, muito relevante: ajudar as pessoas e empresas a ver e entender todos os seus dados. Isso é alcançado oferecendo uma seleção de ferramentas úteis e intuitivas de inteligência de negócios:

- Escritório (Tableau Desktop): conecta e analisa os dados e informações.
- Servidor (Tableau Server): com o servidor, podemos colaborar de forma segura e compartilhar informações a partir dos dados carregados via Tableau Desktop (versão de desktop do software).
- On-line (Tableau Online): é uma versão do Tableau Server hospedada na nuvem. Assim, podemos acessar nossos dados sem a necessidade de um processo tedioso de instalação. (SoftwarePara, s. f., <https://softwarepara.net/tableau/>).

Através de funções simples, como arrastar e soltar, qualquer pessoa pode acessar e analisar dados de forma simples e até criar relatórios e compartilhar essas informações com outros usuários.

Figura 7: Exemplo de visualização no Tableau



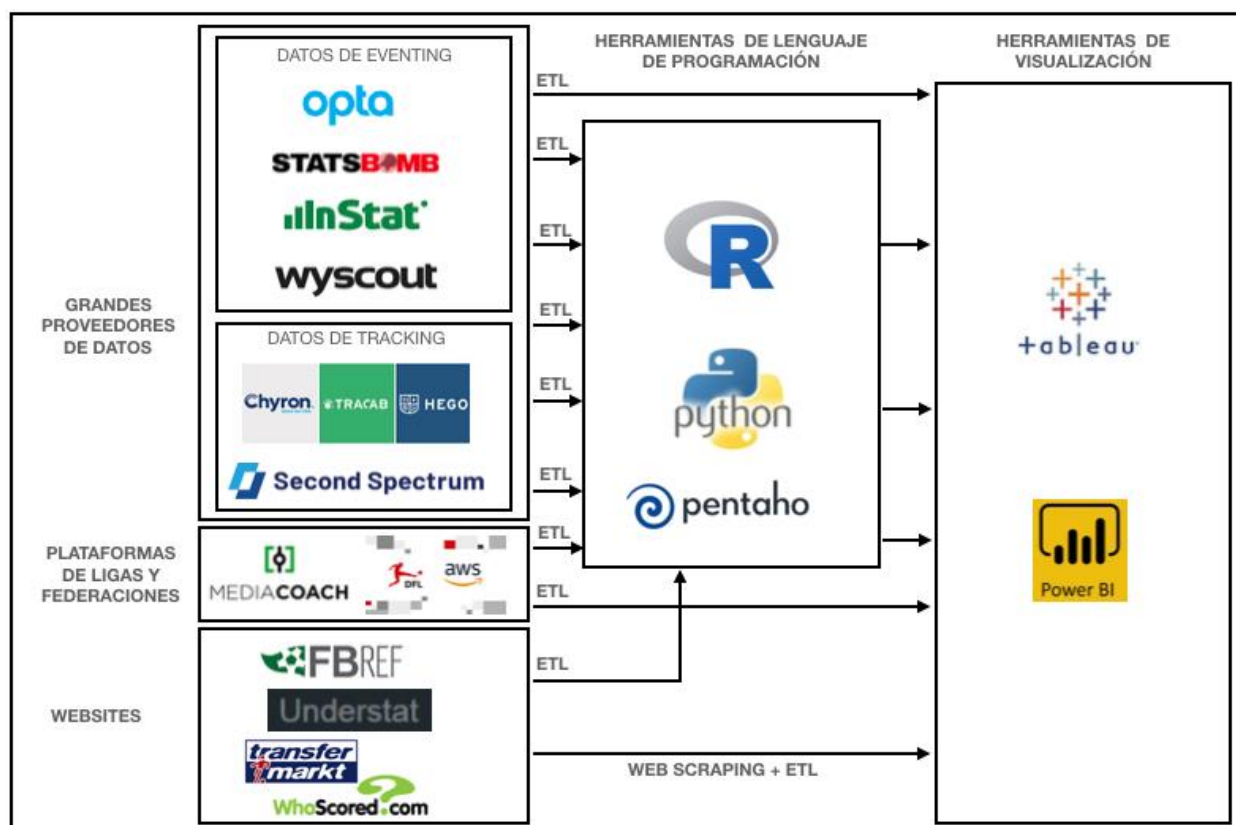
Fonte: De Juana, 2019, <https://www.muycomputerpro.com/2019/06/11/tableau-asi-es-la-empresa-por-la-que-salesforce-ha-pagado-mas-de-13-000-millones>

Unidad 4.2. Principais fornecedores de dados no futebol. Onde podemos obter dados defensivos de pressão?

Seguindo o professor do mestrado em big data esportivo da UCAM, David Fombella Pombal, vamos listar os principais fornecedores de dados e as diferentes ferramentas de programação e visualização que os analistas estão utilizando atualmente.

Existem dois grandes fornecedores de dados que atualmente monopolizam o fornecimento de dados de eventing para os clubes, plataformas de ligas nacionais europeias (como o Mediacoach, etc.) e até sites especializados em futebol. Estamos falando da Opta e StatsBomb. Além disso, existem outros fornecedores conhecidos como InStat, Wyscout, bem como Tracab e Second Spectrum (os dois últimos se dedicam a fornecer dados de tracking).

Figura 8: Principais fornecedores de dados no futebol



Grandes proveedores de datos -
Plataformas de ligas y federaciones -
Websites - Herramientas de lenguaje de programación - Herramientas de visualización

Principais Fornecedores de Dados -
Plataformas de ligas e federações -
Websites - Ferramentas de linguagem de visualização

Fonte: elaboração própria.

Opta: a marca Opta pertence ao grupo Perform, um grupo líder dedicado ao conteúdo e meios digitais de esportes. Esta empresa se dedica à coleta, recepção e armazenamento de dados de muitos e variados esportes, para distribuir essa informação aos clubes que trabalham com eles. No entanto, também atuam com outras vias de negócios (televisões, casas de apostas, jornalismo esportivo, etc.).

Por um lado, a Opta possui um conjunto de widgets para inserir em sites. É bastante comum ver esse tipo de conteúdo obtido em jornais digitais, pois inclui desde conceitos básicos até resultados, métricas e indicadores avançados. No que diz respeito ao futebol, alguns desses widgets disponíveis permitem acessar visualizações de métricas como os já conhecidos expected goals (xG) —gols esperados—, as zonas defensivas, os mapas de calor (heatmap), comparações de equipes e jogadores (player e team comparison), etc.

Figura 9: Widget da Opta 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)



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

Por outro lado, os feeds de dados da Opta, tanto de eventing quanto de tracking, com seus dados estatísticos baixáveis por meio de sua API, estão disponíveis para que os clubes os usem a seu critério, seja para análise de desempenho esportivo, scouting ou serviços de comunicação e imprensa. Por exemplo, o F24 (Event Detail Feed) "é um dos pacotes de dados

da Opta mais utilizados, pois inclui uma variedade de informações e nuances" (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
- **type_id** → 16 Gol
- **player_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

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

StatsBomb: é uma empresa do Reino Unido, recentemente criada, com uma equipe global de profissionais dedicados e experientes, com sede na Europa, Egito e Estados Unidos. Rapidamente, se consolidou junto à Opta como uma empresa líder no campo da análise de futebol.

A StatsBomb, que cobre mais de quarenta ligas em todo o mundo e possui os dados mais precisos e detalhados do mercado, fornece às equipes as ferramentas necessárias para competir e vencer no campo e no mercado de transferências.

Sua plataforma de análise, StatsBomb IQ, não só oferece aos analistas informações sobre equipes, jogadores e ligas, mas também auxilia diferentes meios de comunicação por meio de visualizações de dados personalizadas:

Dentro dos serviços de consultoria que oferecem, estão os seguintes:

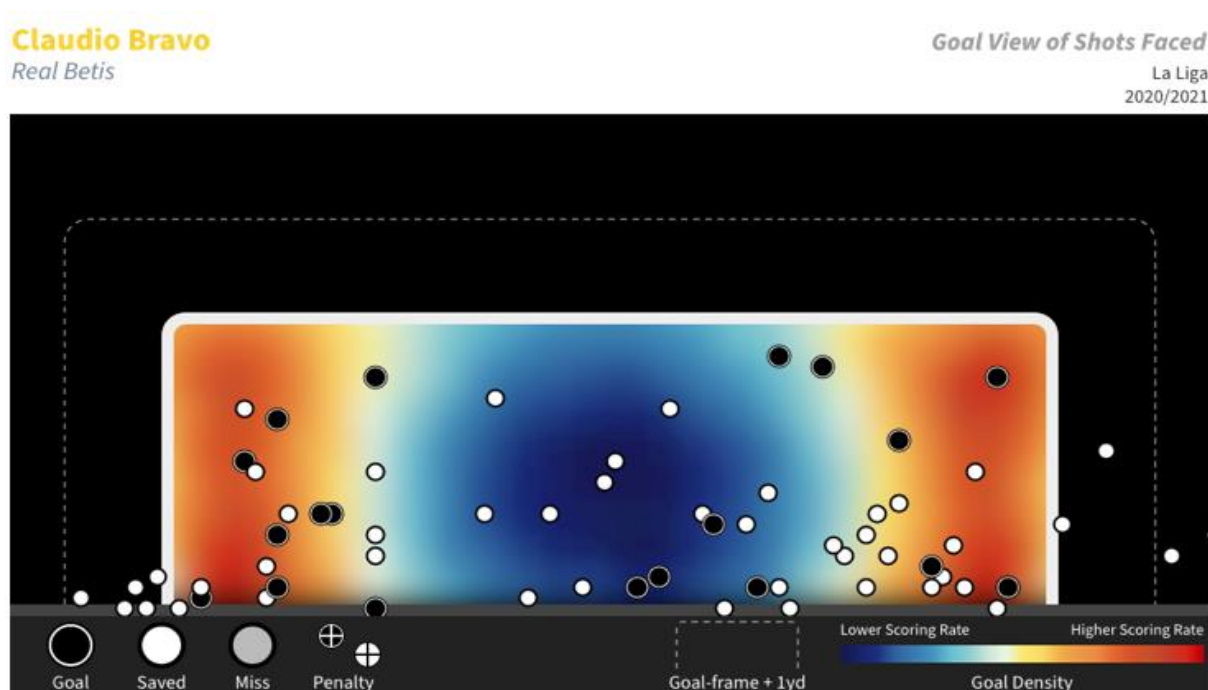
- Avaliação de jogadores (desempenho esportivo, informações pessoais, relatórios de recrutamento, pontos fortes e fracos do jogador, valores de mercado, etc.).
- Busca de treinadores de acordo com seu estilo de jogo.

- Análise de bola parada. (Fombella Pombal, 2020b, p. 138).

Destaca-se principalmente por suas avançadas métricas defensivas, como as pressões defensivas, contrapressões, recuperações após pressão, etc., e suas novas métricas para medir o desempenho do goleiro:

A StatsBomb inclui métricas e visualizações revolucionárias para a avaliação de goleiros. Esqueça coisas tão básicas como gols/chutes a gol para avaliar o desempenho do nosso goleiro. A StatsBomb inclui métricas de posicionamento do goleiro na linha de gol, desempenho ao pegar cruzamentos laterais, contribuição no ataque, atividade defensiva, passes completados sob pressão, etc. (Fombella Pombal, 2020b, p. 143).

Figura 11: Exemplo da métrica do goleiro em relação aos chutes recebidos na StatsBomb (shots faced)



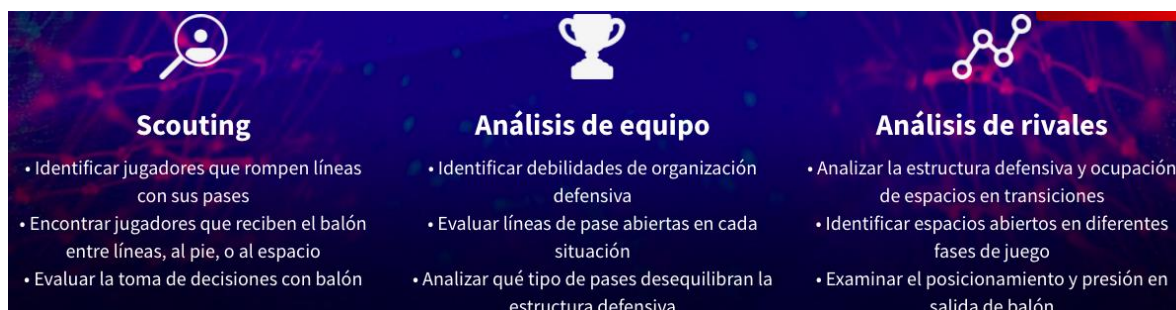
STATSBOMB

Fonte: 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>

A StatsBomb também possui uma API onde os clientes podem baixar todos os dados para tratá-los e fazer suas próprias visualizações, assim como um repositório no GitHub aberto, com informações de algumas competições, para visualizar seu formato de dados e realizar testes. Além disso, inclui uma biblioteca em R (<https://github.com/StatsBomb/StatsBombR>) que se conecta à sua API e obtém dados. (Fombella Pombal, 2020b, p. 144).

Por fim, a StatsBomb não para de inovar e avançar nas métricas, oferecendo dados e análises muito mais profundas do que aquelas que podemos alcançar com o eventing. Este é o chamado StatsBomb 360.

Figura 12: StatsBomb 360



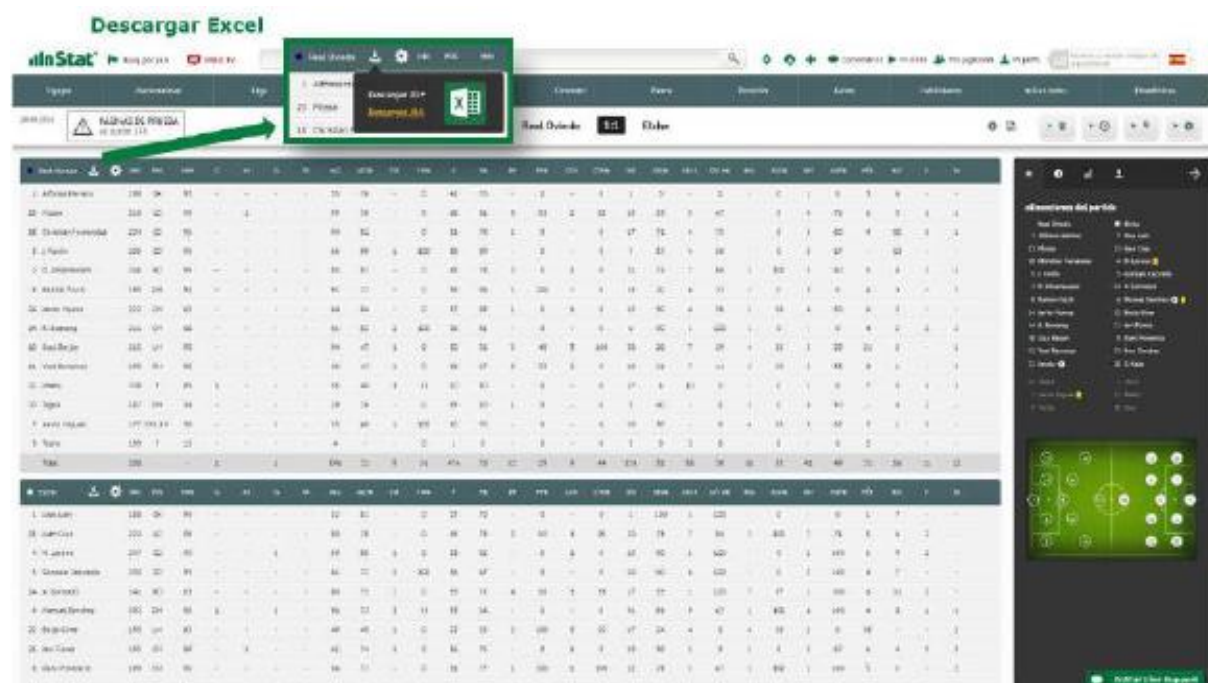
<p>Scouting: Identificar jugadores que rompe líneas con sus pases, Encontrar jugadores que reciben el balón entre líneas, al pie o al espacio, Evaluar la toma de decisiones con balón</p>	<p>Scouting:</p> <ul style="list-style-type: none"> • Identificar jogadores que rompem linhas com seus passes. • Encontrar jogadores que recebem a bola entre linhas, nos pés ou em profundidade. • Avaliar a tomada de decisão com a bola.
<p>Análisis de equipo: Identificar debilidades de organización defensiva, Evaluar líneas de pase abiertas en cada situación, Analizar qué tipo de pases desequilibra la estructura defensiva</p>	<p>Análise de equipe:</p> <ul style="list-style-type: none"> • Identificar fragilidades na organização defensiva. • Avaliar as linhas de passe abertas em cada situação. • Analisar que tipo de passes desequilibram a estrutura defensiva
<p>Análisis de rivales: Analizar la estructura defensiva y ocupación de espacios en transiciones, Identificar espacios abiertos en diferentes fases de juego, Examinar el posicionamiento y presión en salida de balón</p>	<p>Análise de adversários:</p> <ul style="list-style-type: none"> • Analisar a estrutura defensiva e a ocupação de espaços nas transições. • Identificar os espaços abertos em diferentes fases do jogo. • Examinar o posicionamento e a pressão na saída de bola.

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

Instat e Wyscout: são outros provedores de dados que estão amplamente presentes nos clubes (90% dos clubes profissionais contratam seus serviços). São ferramentas muito semelhantes e extremamente úteis para clubes, diretores esportivos, treinadores, analistas, treinadores de goleiros, árbitros, os próprios jogadores e diversos meios de comunicação. São uma das melhores ferramentas para a seleção de talentos (scouting e recrutamento), pois oferecem tanto imagens em formato televisivo quanto visualizações de uma infinidade de dados que geram sobre equipes e jogadores. Além disso, oferecem a possibilidade de baixar seus dados brutos (raw data) para tratamento posterior, permitindo que os clientes gerem suas próprias visualizações:

Instat oferece estatísticas detalhadas da partida que incluem avaliações dos jogadores (índice Instat, gols, assistências, precisão de passes, cruzamentos e precisão nos cruzamentos, dribles, disputas, duelos, perdas, recuperações, faltas e cartões). A partir desta plataforma, é possível baixar um arquivo Excel com uma aba de equipes e outra de jogadores, permitindo trabalhar com essas informações em ferramentas de visualização, como Microsoft Power BI ou Tableau. (Fombella Pombal, 2020b, p. 67).

Figura 13: Instat: estatísticas detalhadas da partida



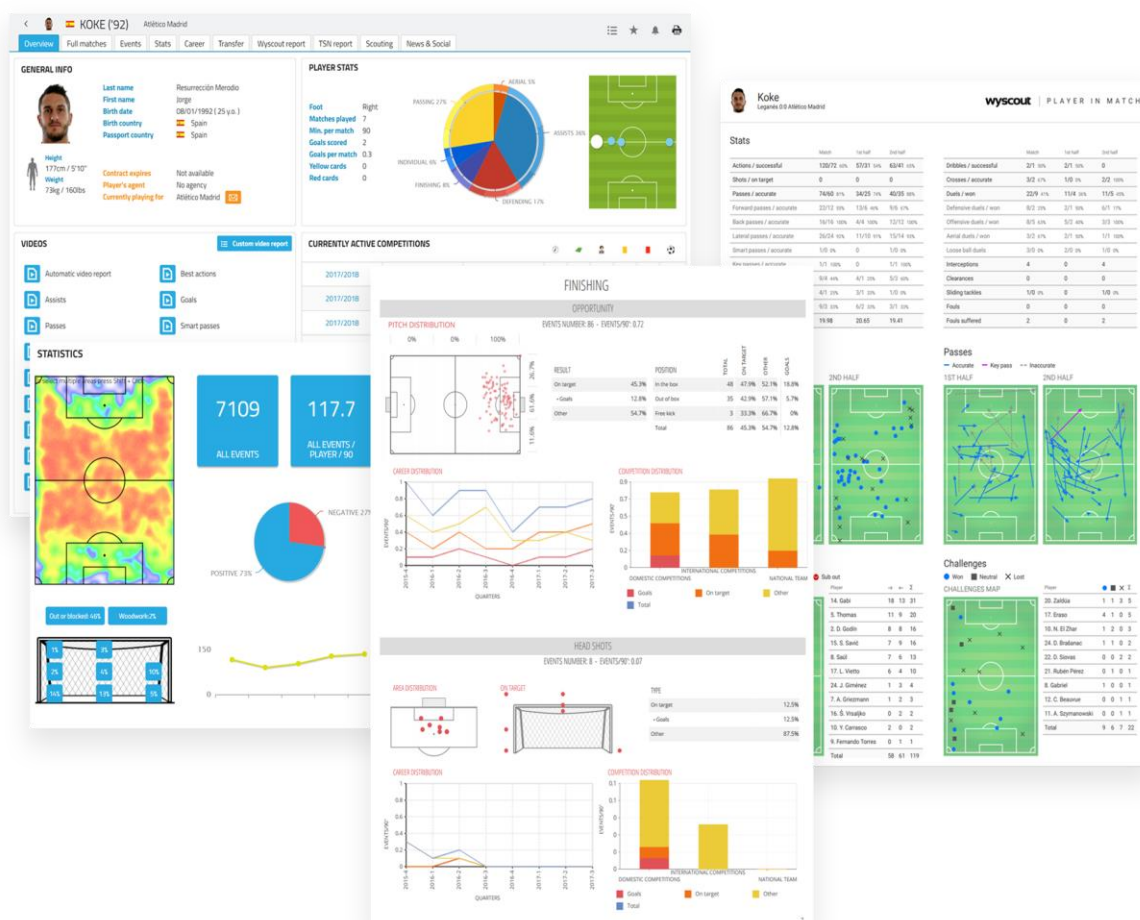
Fonte: Fombella Pombal, 2020b, p. 67.

Wyscout é uma ferramenta muito semelhante ao Instat. Em 2019, associou-se à Hudl e, juntamente com a ferramenta de análise SportsCode, possui duas das ferramentas mais utilizadas pelos clubes profissionais em nível mundial.

Alguns dos instrumentos avançados de análise estatística podem ajudar a acessar informações de todos os aspectos do jogo. Estas funcionalidades que a aplicação Wyscout oferece são:

- Relatórios estatísticos detalhados e completos para download (dados de eventos técnicos e táticos).
- Rankings de mais de 200 competições.
- Análise de jogadores: por meio da “Advanced search”, é possível descobrir novos talentos através de desempenhos estatísticos.
- Com a API do Wyscout, é possível baixar todos os dados necessários para integrar facilmente com ferramentas de BI e softwares de análise de vídeo como SportsCode ou Nacsport. (Fombella Pombal, 2020b).

Figura 14: Wyscout: estatísticas detalhadas da partida



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

"Os Feed Events (eventos) do Wyscout oferecem informações da partida com todos os eventos e sua posição x/y no campo de jogo... Esta base de dados histórica permite a análise matemática avançada e a detecção de padrões" (Fombella Pombal, 2020b, p. 107).

Até agora, falamos de fornecedores de eventing, mas existem outros fornecedores dedicados à geração de dados de tracking.

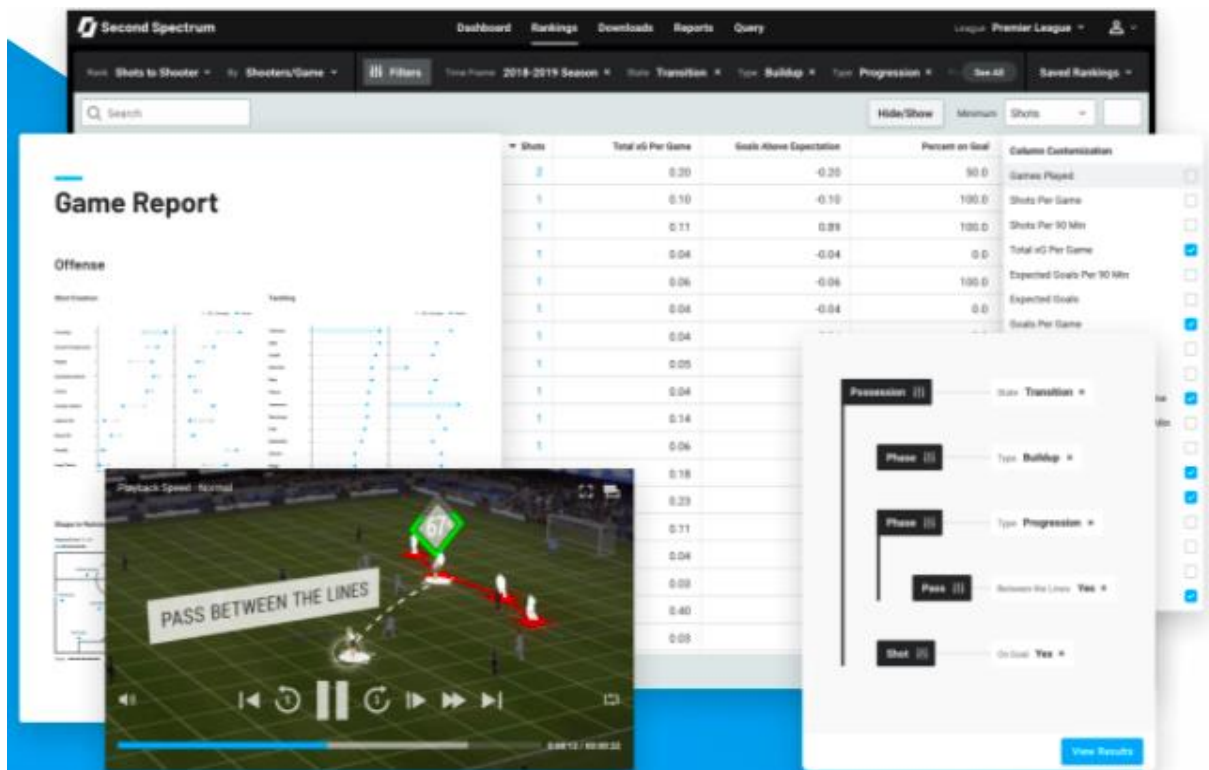
Fornecedores como Tracab e Second Spectrum capturam, por meio de câmeras robóticas inteligentes, o movimento dos jogadores e da bola, gerando arquivos de tracking raw data XYZ em tempo real.

Esses arquivos incluem informações XY dos 22 jogadores e XYZ da bola 25 vezes por segundo (25 Hz), gerando mais de 3 milhões de dados por partida: $(90 \text{ m} \times 60 \text{ s/m} \times 25 \text{ frames/s} \times 23 \text{ (jog. + bola)}) = 3\,105\,000$ dados/partida

Second Spectrum: é um dos sistemas de rastreamento de jogadores mais avançado do mundo. Aplica técnicas de visão computacional e aprendizado de máquina de última geração para produzir dados de localização rápidos e precisos para basquete, futebol, futebol americano e outros esportes. É o fornecedor oficial de dados da NBA e da Premier League inglesa.

Os dados de eventing e tracking que gera impulsionam novas formas para que as equipes e ligas compreendam, avaliem, melhorem e criem conteúdo sobre o seu jogo.

Figura 15: Tipo de relatório do Second Spectrum



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

Tracab: é um dos maiores fornecedores de tracking e eventing. Fornece dados para inúmeros clubes e federações de futebol ao redor do mundo (incluindo as ligas espanhola, alemã, holandesa, mexicana e japonesa; a Champions League, entre outras). Tracab oferece soluções em forma de dados esportivos, vídeos e visualizações, com informações valiosas para melhorar o desempenho das equipes profissionais, além de permitir uma gama cada vez maior de soluções inovadoras e em tempo real para envolver os torcedores de forma mais eficaz. Um exemplo claro está na liga espanhola: a LFP (Liga de Fútbol Profesional) oferece aos clubes a ferramenta Mediacoach, que se abastece de dados de tracking da Tracab junto com os dados de eventing da Opta:

Para elaborar o dossiê de cada partida, que geralmente ocupa entre 500 e 600 páginas com métricas e dados, o Mediacoach utiliza dois fornecedores. Por um lado, está a Tracab, responsável pelo tracking, ou seja, o acompanhamento contínuo de cada jogador em campo através de câmeras instaladas em todos os estádios da La Liga Santander e La Liga 1|2|3, financiadas pela La Liga. Este sistema fornece a distância percorrida, a velocidade máxima e média, o número de sprints, etc. Por outro lado, está o denominado eventing, no qual três pessoas da Opta registram os passes, perdas de bola, faltas, escanteios, etc. Tudo o que é mais objetivável. (Olmeda, 2020, <https://www.abc.es/contentfactory/post/eslaliga/laliga-apuesta-por-la-tecnologia-para-hacerse-mas-competitiva/>).

Figura 16: Exemplo da plataforma Mediacoach da liga espanhola

¿Qué es Mediacoach? Suite Productos Mediacoach

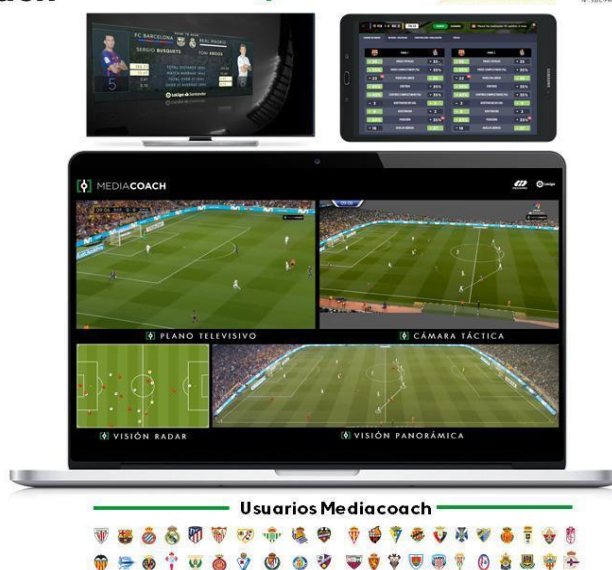
Desktop
Herramienta avanzada de vídeo análisis profesional integrada con datos físicos y técnico-tácticos

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 vídeos 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



Certificado Oficial
PROYECTO INNOVACIÓN
TECNOLÓGICA

ACIE
AGENCIA DE CERTIFICACIÓN
EN INNOVACIÓN TECNOLÓGICA

ENAC
CENTRO NACIONAL DE ASESORÍA
CERTIFICACIÓN
Nº 1103 PRODA

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

Essas plataformas são outra das fontes de dados que os clubes e as equipes técnicas das ligas profissionais que oferecem esse serviço possuem.

Plataformas de ligas e federações: as ligas e federações desempenham um papel crucial no acesso à informação e aos dados pelos clubes. A criação dessas ferramentas ou plataformas de acesso à informação, por parte de todas as equipes técnicas (treinadores, analistas e preparadores físicos) e todos os departamentos de análise dos clubes de suas respectivas ligas, democratizou o acesso à informação, pois anteriormente apenas os clubes mais poderosos podiam pagar a esses fornecedores para acessar seus dados. Isso acelerou muito mais a evolução da indústria de dados, melhorando seu desempenho e utilidade, ao mesmo tempo que facilitou o acesso a eles. Com o acesso e uso do conjunto de dados de tracking combinado com o eventing (raw data), os clubes conseguem reunir muito mais informações que podem trabalhar por conta própria. Isso é feito de forma terceirizada (consumindo as próprias métricas e visualizações que as plataformas oferecem) ou de forma mais interna (trabalhando e transformando os dados com processos de ETL que lhes permitem criar seu próprio conhecimento e suas próprias visualizações). Com o uso dessas plataformas, é possível melhorar o desempenho dos jogadores e reforçar o envolvimento dos torcedores com um conjunto simples de ferramentas que permite:

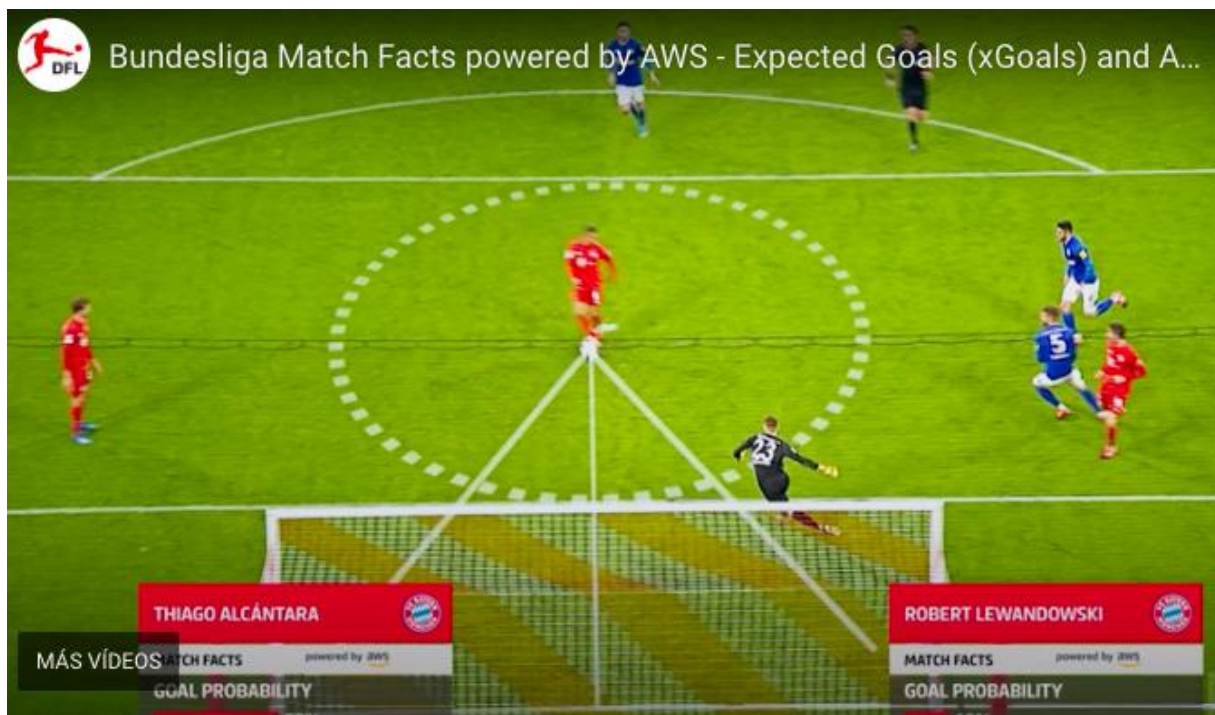
Unir dados e vídeo: as capacidades de dados e vídeo resultantes da combinação de tracking e eventing oferecem precisão incrível e fluxos de trabalho acelerados, permitindo que os analistas criem informações valiosas sobre o planejamento, a exploração e a gestão da carga dos jogadores para seus treinadores.

Melhorar as táticas das equipes: o acesso ao raw data dessas plataformas oferece tecnologia que proporciona dados de desempenho precisos e válidos através de vídeos informativos que sustentam a análise. Isso permite que a equipe técnica aprenda com o desempenho da equipe ao longo do tempo, desenvolva programas de melhoria específicos e elabore novas táticas para alcançar ainda mais sucesso.

Tomar decisões durante a partida: essas plataformas oferecem conteúdo em tempo real durante a partida para o monitoramento ao vivo do desempenho da equipe, além de dados de eventos com vídeo sincronizado, permitindo que os treinadores tomem mais e melhores decisões durante a partida, proporcionando uma vantagem competitiva crucial.

Otimizar o desempenho do jogador: a excelente tecnologia de rastreamento que esses fornecedores utilizam oferece uma precisão de posicionamento e uma qualidade de dados inigualáveis, que capturam o desempenho físico dos jogadores. Dessa forma, é um elemento-chave para o desenvolvimento de programas de treinamento especializado, minimização de lesões e manutenção de conjuntos de dados precisos e consistentes, tanto no treinamento quanto na partida.

Figura 17: Exemplo da plataforma da liga alemã Match Facts



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

Principais sites de dados no futebol: outra fonte onde podemos obter dados e informações relevantes sobre equipes, jogadores, partidas e métricas defensivas de pressão alta são os sites especializados em futebol. Além de serem públicos, gratuitos e de oferecerem até

mesmo algum tipo de visualização (como o site Understat), o mais interessante é a possibilidade de baixar seus dados (Fbref). Isso também pode ser feito por meio de web scraping, usando algumas das ferramentas de linguagem de programação mencionadas neste módulo (R ou Python), ou diretamente das ferramentas de visualização (Power BI ou Tableau). Sem dúvida, esses sites desempenham um papel fundamental para os clubes e equipes técnicas sem recursos econômicos. Vamos comentá-los:

Fbref: é um site gratuito com os dados mais avançados atualmente no mundo do futebol. Graças à sua associação com a StatsBomb, oferece métricas ainda mais avançadas do que sites como WhoScored. É o mais próximo do que oferecem Opta, StatsBomb, Wyscout e Instat, mas de forma pública e gratuita. Eles estão sempre melhorando o site, adicionando dados e métricas novas. Por exemplo, incluem dados como os xG (gols esperados), dados defensivos de pressão, dados avançados de goleiros, etc. Além disso, oferecem o dado total e seu cálculo por x90, organizando-os de forma muito atraente (métricas padrão, métricas de chutes, de passes, de criação de oportunidades, de posse de bola, de partidas e tempo jogado, de ações defensivas, etc.).

Figura 18: Exemplo do site 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](#).

Fonte: captura de tela do site Fbref, <https://fbref.com/en/comps/12/La-Liga-Stats>

Understat: neste site você pode encontrar informações para analisar equipes e jogadores das cinco grandes ligas da Europa, com métricas avançadas como xG (expected goals – gols esperados), NPxG (non-penalty expected goals – gols esperados sem penalidade), xPTS (expected points – pontos esperados), xA (expected assists – assistências esperadas), etc. E, no nível defensivo, destaca-se o já conhecido PPDA (passes per defensive action – passes por

ação defensiva). Assim como o Fbref utiliza os dados gerados pela StatsBomb, o Understat usa os dados da Opta, por isso é importante saber que há pequenas diferenças nos dados, pois cada fornecedor utiliza seus próprios algoritmos para gerá-los.

Figura 19: Exemplo do site Understat

Table		Charts		overall	home	away	Start date	End date													
Nº	Team	M	W	D	L	G	GA	PTS	xG	NPxG	xGA	NPxGA	NPxGD	PPDA	OPPDA	DC	ODC	xPTS			
1	Athletico Madrid	38	26	8	4	67	25	86	57.66	-9.34	52.31	35.87	+10.67	32.70	+19.61	10.32	12.38	310	165	68.41	-18.58
2	Real Madrid	38	25	9	4	67	28	84	65.07	-1.93	62.84	39.70	+10.70	32.75	+30.09	10.14	14.03	328	171	69.98	-14.02
3	Barcelona	38	24	7	7	65	39	79	83.98	-1.04	78.01	40.93	+2.93	37.22	+40.90	9.51	18.20	498	168	78.18	-0.82
4	Sevilla	38	24	5	9	53	33	77	55.88	+2.86	50.68	34.33	+1.33	29.84	+20.81	8.15	14.58	240	139	68.98	-10.04
5	Real Sociedad	38	17	11	10	59	38	62	63.15	+4.15	54.22	38.80	+0.80	35.08	+10.14	8.08	11.70	285	180	67.81	+5.81
6	Real Betis	38	17	10	11	50	50	61	49.77	-0.23	41.43	47.51	-2.49	40.08	+1.35	9.12	10.55	224	180	55.37	+5.63
7	Villarreal	38	15	13	10	60	44	58	57.86	-2.14	48.19	43.67	-0.33	38.98	+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	48.12	50.31	-0.69	44.37	+1.75	7.40	10.84	217	187	52.83	-0.17
10	Granada	38	13	7	18	47	65	48	41.05	-6.95	37.34	59.82	-8.18	50.81	-13.27	10.74	8.88	128	234	40.58	+5.42
9	Athletic Club	38	11	13	14	48	42	48	43.08	-2.94	39.35	38.48	-3.54	34.00	+5.34	9.75	9.80	202	187	54.77	+8.77
12	Cadiz	38	11	11	18	38	58	44	38.70	+0.70	33.73	58.91	+0.81	51.48	-17.75	18.57	5.93	77	274	39.01	+4.98
11	Osasuna	38	11	11	18	37	48	44	39.81	+2.81	35.15	48.83	+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	34.48	55.89	+2.89	49.74	-15.28	10.94	8.81	182	282	42.38	+0.82
14	Levante	38	9	14	15	48	57	41	40.83	-0.07	38.47	55.13	-1.87	52.18	-15.69	9.47	10.22	221	245	43.53	+2.53
16	Alaves	38	9	11	18	38	57	38	39.11	+3.11	33.17	51.55	-8.46	49.32	-18.15	12.25	7.88	119	212	43.23	+5.23
15	Oseta	38	9	11	18	28	43	38	35.82	+7.82	31.91	40.40	-2.60	34.32	-2.42	8.81	8.43	134	155	48.39	+10.39
17	Elche	38	8	12	18	34	55	38	30.40	-3.60	28.17	61.83	+6.63	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	16	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	35.84	48.84	-3.16	45.12	-9.48	8.99	10.58	149	158	48.82	+18.82

Fonte: captura de tela do site Understat, https://understat.com/league/La_liga

Transfermarkt: trata-se de um site que contém dados sobre equipes e seus jogadores, mas com foco em transferências e no valor de mercado dos jogadores. Nele, encontramos notícias sobre rumores de transferências; as últimas contratações e renovações; empréstimos; jogadores cujo contrato está para expirar e os que estão sem clube; as transferências mais caras; os jogadores mais valiosos; etc.

O site alcançou algo extremamente difícil: tornar-se uma referência em dados sobre a valorização econômica dos jogadores. Essa avaliação transcendeu o próprio site e, em muitos casos, é até utilizada por clubes (em geral, os menores). É possível buscar por posição ou por posições detalhadas, idade dos jogadores, etc. Nesta lista, encontramos o nome, a posição, a nacionalidade, a idade, o clube ao qual pertence e seu valor de mercado. Pode-se observar se o valor permanece o mesmo, se caiu ou, pelo contrário, se aumentou.

Figura 20: Exemplo do site Transfermarkt



Fonte: captura de tela do site Transfermarkt, <https://www.transfermarkt.es/bernardo-silva/profil/spieler/241641>

WhoScored: este portal de estatísticas de futebol não possui dados tão avançados quanto os mencionados anteriormente, mas oferece informações interessantes sobre a grande maioria das competições de futebol, apresentando-as em um site muito visual, intuitivo e que permite ajustar o nível de detalhe desejado. Contém dados avançados, por equipe ou jogador, e uma grande quantidade de dados por partida para as cinco grandes ligas. Inclusive, possui sua própria avaliação de desempenho de um jogador por partida. Além disso, oferece diversos tipos de estatísticas por países e por equipe em formato resumido (características ofensivas e defensivas, resultados ao longo do tempo, número de partidas invictas, jogos sem sofrer gols, partidas sem vitória ou sem marcar, etc.).

Figura 21: Exemplo do site 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

Fonte: captura de tela do site Whoscored, <https://www.whoscored.com>

Unidade 4.3. Principais métricas defensivas para medir a pressão alta

No mundo da análise de futebol, historicamente tem sido mais complicado medir a parte defensiva do jogo do que a ofensiva. Basicamente, porque enquanto no ataque o desempenho bem-sucedido acaba se traduzindo em realizar uma ação (finalizar, assistir, completar um drible, marcar gol etc.) e fazer isso mais vezes tende a ser melhor, na defesa nem sempre é assim. Na defesa, agir pouco pode ser melhor do que agir muito:

Um defensor que faça mais desarmes ou interceptações não é necessariamente melhor do que um que faça menos; esses dados medem mais a oportunidade do que a habilidade. Nas ações realizadas na defesa, há muito mais ruído do que sinal. Como podemos começar a dissipar esse ruído? (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

O objetivo defensivo é evitar que o time adversário progrida e crie oportunidades de gol e, portanto, muitas vezes o que não é feito é tão importante — ou até mais — do que o que é feito. Por exemplo, um posicionamento adequado ou um pressing coordenado de forma eficaz limitam as opções do adversário e, por isso, é menos provável que os jogadores mais recuados tenham que defender de maneira ativa com frequência.

No futebol, só recentemente começaram a ser estabelecidas as bases da análise estatística defensiva. Isso foi possível graças à contribuição de fornecedores de dados como StatsBomb e Opta, que, juntamente com as demandas dos clubes por melhorias nas estatísticas e dados existentes, e os avanços tecnológicos e conhecimentos de seus profissionais de diferentes áreas (informática, matemática, física, treinadores, analistas etc.), estão conseguindo um progresso nunca visto antes.

Para isso, vou me concentrar nas métricas defensivas fornecidas pelo provedor de dados StatsBomb, pois, na minha opinião, é o provedor que melhor coleta dados sobre as ações de pressão, tanto de equipes quanto de jogadores.

Seguindo Pablo Peña Rodríguez, chefe de Inovação Tática e Desenvolvimento de Negócios da StatsBomb, vamos comentar cada uma das métricas e visualizações defensivas que nos permitirão medir a capacidade de pressionar alto de uma equipe.

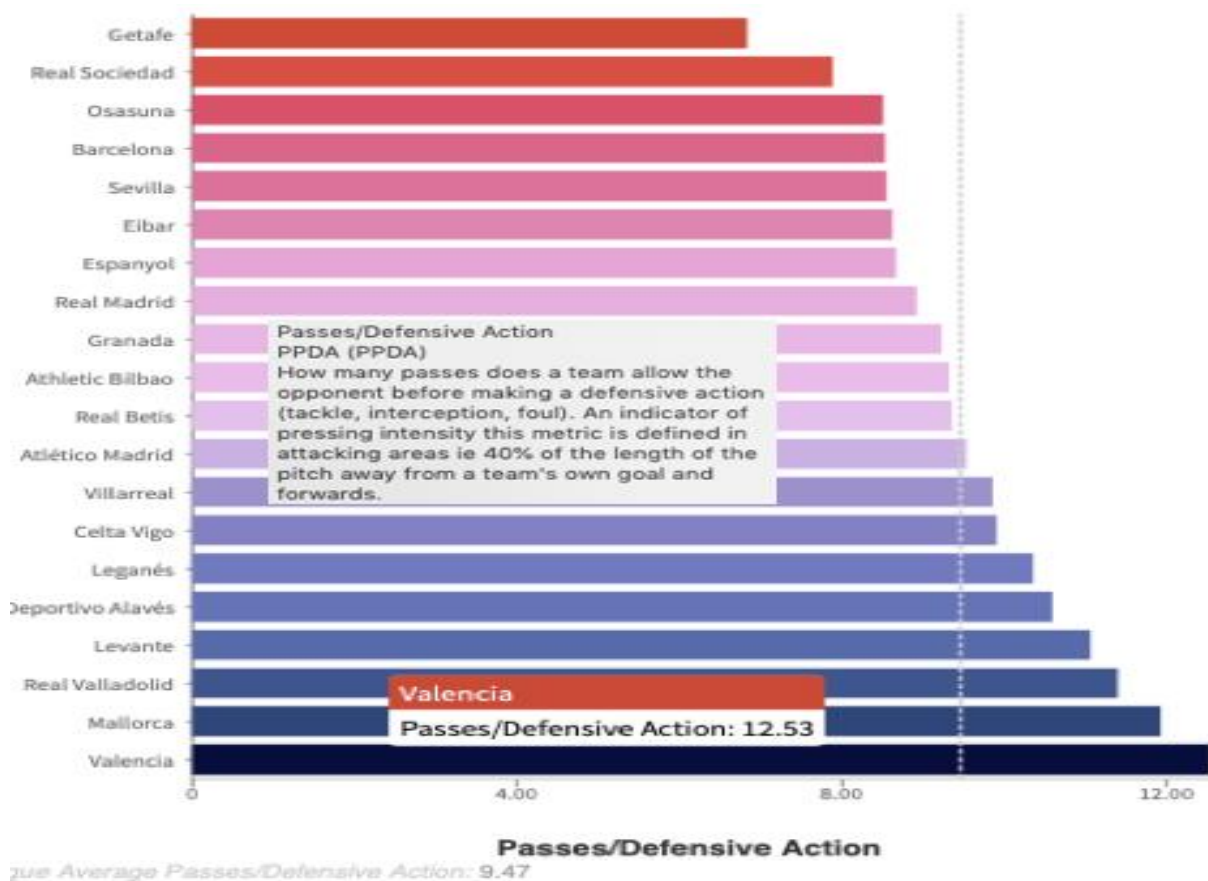
4.3.1. PPDA: Passes por cada ação defensiva

Essa é uma métrica criada por Colin Trainor em 2014. Ele estava realizando um trabalho em que tentava medir a capacidade de pressionar alto das equipes, tentando responder às preocupações dos treinadores sobre como, quando e onde as ações defensivas (desarmes, interceptações, faltas, duelos) eram realizadas pelas equipes.

Para isso, ele estabeleceu uma correlação entre os passes permitidos ao adversário e as ações defensivas da equipe que tentava recuperar a bola, com as seguintes conclusões: Quanto menos passes o adversário realiza por cada ação defensiva da minha equipe, maior é a intensidade defensiva da minha equipe. Ao contrário, quanto mais passes o adversário realiza por cada ação defensiva da minha equipe, menor é a intensidade defensiva exercida pela minha equipe.

Na Figura 22, vemos o ranking final da liga 2019/20 após as 38 rodadas do campeonato. Nela, vemos como o Getafe é a equipe que menos passes por cada ação defensiva realizada permite ao seu adversário na fase de início e progressão, e o Valencia, a equipe que mais passes por cada ação defensiva realizada permite ao seu adversário no campo rival.

Figura 22: PPDA em um ranking StatsBomb



Fonte: [Imagem sem título sobre PPDA em um ranking do StatsBomb]. (s. f.). Recuperado de https://lh3.googleusercontent.com/4QRdqGWNi0_bJ_Zs392Z0mH96LDHnlenpyl145mhA9-nOsTvqJTVuzR4NS2CSroCXb0U=s85

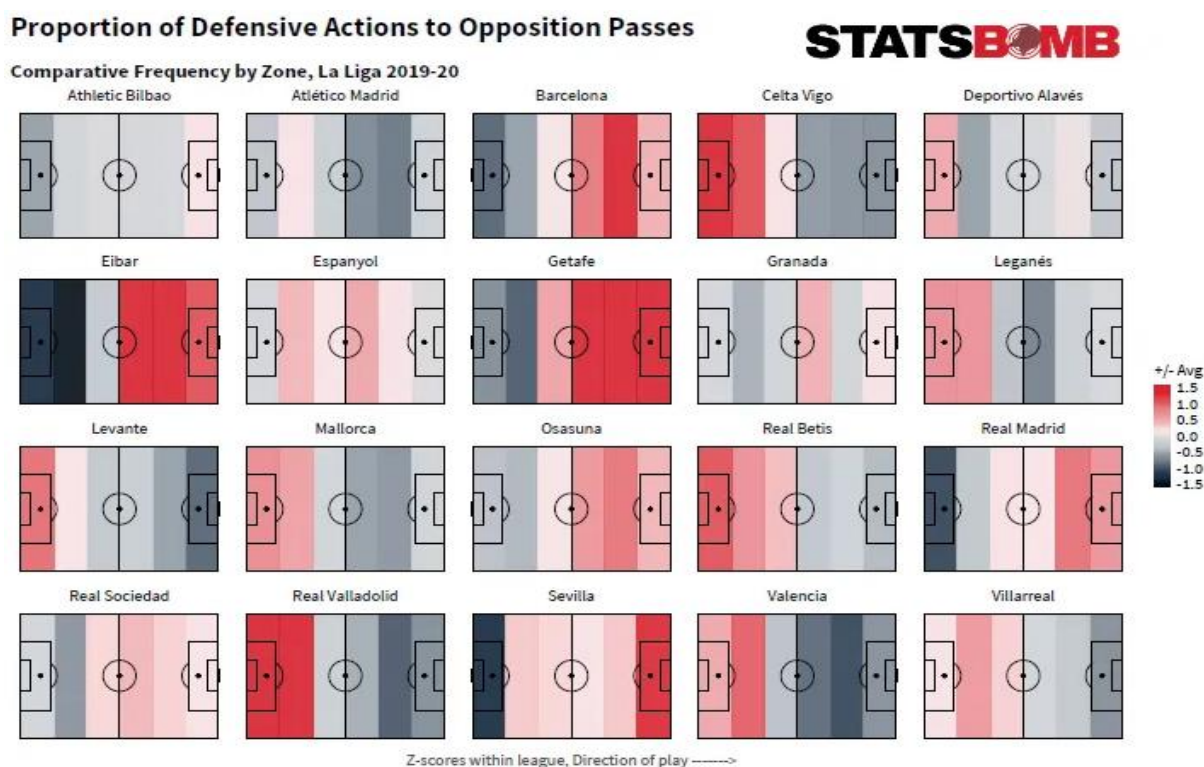
A partir desta visualização, podemos interpretar que o Getafe pressiona muito melhor que o Valencia no campo adversário, o que indica uma maior capacidade de exercer pressão alta e assim impedir o jogo associativo do oponente.

Inicialmente, a ideia deste estudo era identificar as equipes que melhor pressionavam no campo adversário, já que essa métrica se aplicava a 60% do total do campo mais distante da nossa baliza (o campo adversário + 10 metros do nosso próprio campo).

No entanto, no StatsBomb também utilizam essa métrica em todo o campo para estabelecer o perfil defensivo das equipes em cada um dos seis setores do campo. Além disso, os eventos defensivos estão normalizados de acordo com a posse de bola do adversário em cada setor:

Embora a ideia inicial fosse discriminar as equipes que apresentavam uma pressão mais alta, a métrica foi sendo desenvolvida e gradualmente adaptada para diferentes zonas, até se tornar uma das métricas descritivas mais utilizadas para mostrar as características estilísticas que definem a defesa das equipes. No StatsBomb, utilizamos essa métrica regularmente em nossas análises (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

Figura 23: Proporção de ações defensivas do StatsBomb



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

4.3.2. A pressão (pressure)

A chave das métricas avançadas para medir a ação e a intensidade defensiva de uma equipe está na coleta de dados sobre a pressão (pressure), definida pelo StatsBomb como a ação de acosar o possuidor rival:

"A pressão é a ação de atacar o possuidor de bola ou o receptor de um passe em um raio de cinco metros, dependendo da zona do campo, sem realizar um desarme, falta ou interceptar o passe (qualquer uma dessas ações pode ocorrer posteriormente, mas é um evento diferente)." (Peña Rodríguez, P. 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

A razão para coletar esse dado de pressão é a crença de que muitas das ações defensivas que permitem recuperar a posse de bola ou realizar uma pressão coletiva eficaz não se limitam às ações sobre a bola (desarmes, tackles, interceptações, duelos, despejos, etc.). Existem várias ações que contribuem para a recuperação da posse, como o ato de acosar para direcionar a pressão para uma área específica, fechar linhas de passe, acelerar a ação do jogador, forçar passes descontrolados ou fazer com que o rival se veja forçado a se livrar da bola.

Em termos coletivos, isso permite ter uma visão mais realista e completa dos mecanismos táticos utilizados pelas equipes na fase defensiva, respondendo a várias perguntas feitas por treinadores e analistas, como:

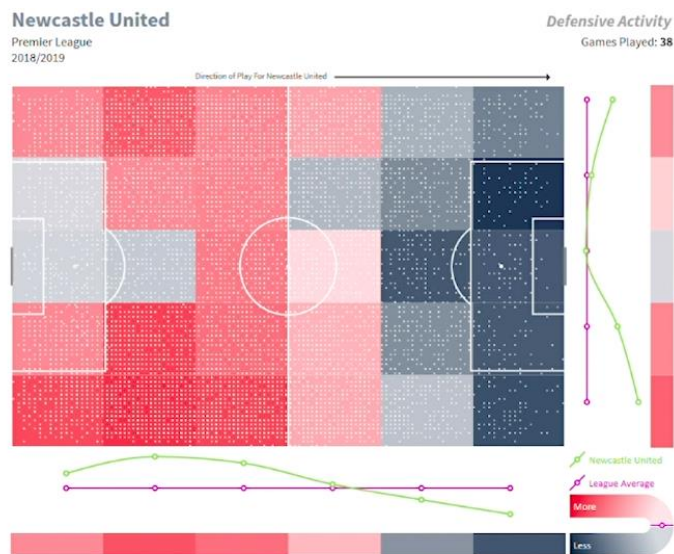
- Como podemos obter uma visão mais representativa da influência de determinados jogadores na defesa, além de recuperar a posse diretamente?
- Quais características estilísticas diferenciam as equipes que tentam pressionar alto das que defendem em blocos médios ou baixos?
- Em quais áreas os adversários começam a exercer pressão? Para qual lado eles os direcionam? Onde são mais fortes?
- Com que frequência pressionam em determinadas áreas do campo e quais jogadores estão mais envolvidos nisso?, etc. (Peña Rodríguez, 2020).

Para recuperar a posse o mais próximo possível do gol adversário durante a fase de pressão alta, as equipes tomam decisões estratégicas que definem seus perfis defensivos, os quais são claramente reconhecíveis nas visualizações dos dados defensivos analisados.

Mapa de atividade defensiva

De acordo com Peña Rodríguez (2020), o mapa de atividade defensiva mostra as ações defensivas realizadas pela equipe em cada quadrante. O StatsBomb divide o campo em cinco setores verticais e três horizontais (cada meio de campo). A cor vermelha indica maior atividade defensiva em comparação com a média da liga, enquanto a cor cinza escuro indica menor atividade defensiva.

Figura 24: Mapa de atividade defensiva do StatsBomb



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

STATSBOMB

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

Vermelho = Mais frequente
Azul = Menos frequente
Direção de jogo = Da esquerda para a direita

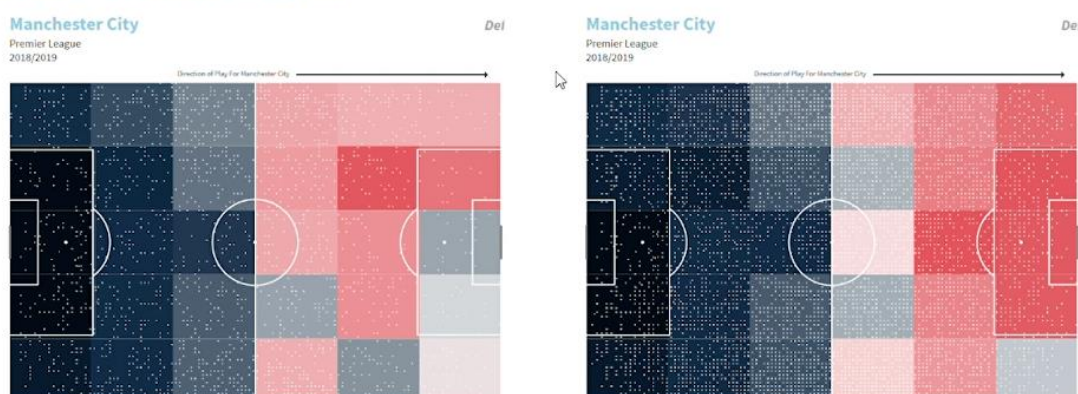
Fonte: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

Na figura 24, temos o Newcastle United da temporada 2018/19, sob o comando de Rafa Benitez. No mapa, podemos observar como a equipe está focada em defender em um bloco baixo em seu próprio campo, com maior atividade nas zonas exteriores e zonas irregulares no interior e no balcão da sua própria área. Também apresenta predominância defensiva no meio-campo, posicionando-se em um bloco médio, mas, à medida que avançamos em direção ao gol adversário, podemos perceber que não é uma equipe que se destaca por sua atividade defensiva de pressão alta (passamos de uma cor cinza para uma mais escura).

Na figura 25, temos um exemplo contrário, o Manchester City de Pep Guardiola, que é o exemplo de uma equipe com um perfil claro de pressão alta em campo adversário.

Figura 25: Comparação de mapas com métricas tradicionais e com a métrica de pressão.

Estilos: Presión alta



STATSBOMB

Las estadísticas defensivas tradicionales no abarcan todo el espectro

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

Derecha: Lo mismo pero incluyendo acciones de presión

Estilos: presión alta

Las estadísticas defensivas tradicionales no abarcan todo el espectro.

Izquierda: Man City 2018-19, tackles, interpretaciones, bloqueos, etc.

Derecha: lo mismo pero incluyendo acciones de presión

Estilos: Pressão alta

As estatísticas defensivas tradicionais não abrangem todo o espectro.

Esquerda: Man City 2018-19: Desarmes, intercepções, bloqueios, etc.

Direita: O mesmo, mas incluindo ações de pressão.

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

No gráfico à esquerda, vemos como seria o mapa com as estatísticas defensivas tradicionais, que medem apenas as ações defensivas em que há contato com a bola. Em contrapartida, no mapa à direita, são adicionadas as ações de pressão e perseguição da equipe sobre o adversário que possui a bola. A diferença é evidente. O dado de pressão nos oferece uma imagem mais representativa da parte defensiva do jogo e, mais especificamente, da capacidade de pressão alta das equipes.

Os eventos defensivos nos fornecedores de dados tradicionais foram entradas ou tackles, intercepções, duelos, despejos, etc. Embora esses dados sejam úteis até certo ponto, não oferecem uma visão representativa de todo o espectro de ações que impactam defensivamente a equipe, limitando a capacidade de extrair conclusões relevantes... Os dados da StatsBomb fornecem um novo evento que muda a maneira de abordar a análise defensiva: as pressões. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

A partir dessa afirmação, podemos concluir que o dado de pressão nos oferece uma visão melhor e mais ampla de como as equipes pressionam no campo adversário do que a métrica do PPDA.

Outras visualizações que podemos utilizar a nível coletivo são os rankings de equipes.

Rankings

Com os dados defensivos de pressão parciais (em uma determinada zona do campo) ou totais (em todo o campo), podemos diferenciar as diferentes fases defensivas.

Na figura 26, vemos um ranking retirado do site Fbref, no qual vemos em vermelho as pressões exercidas por cada equipe em cada terço do campo.

Def. 3.rd = bloco baixo.
 Mid. 3.rd = bloco médio.
 Att. 3.rd = bloco alto.

Em amarelo, no ranking à esquerda, vemos que o Sevilla ocupa a última posição em número de pressões totais exercidas durante a La Liga 2020/21.

Figura 26: Rankings de pressões totais exercidas e por terços 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	ShSv	Pass	
Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	
Coita 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	

Fonte: Adaptação própria de Fbref, s. f., <https://fbref.com/en/comps/12/La-Liga-Stats>

Outra visualização, neste caso da StatsBomb, podemos ver na figura 27. Nela, podemos observar um ranking que mede a capacidade de pressionar alto no campo adversário das equipes da La Liga 2020/21. Nesse ranking, vemos o Sevilla na 14ª posição.

Figura 27: Ranking de pressão no campo adversário da StatsBomb.



Fonte: [Imagem sem título sobre ranking de pressão no campo adversário]. (s. f.). Recuperado de <https://lh3.googleusercontent.com/CFilwr8zUC-i9Gk-kyX5x-NEZRYcqj2QH4WoN63PAWrea3qbb1Avq4JGoepUMUvt1feuA=s85>

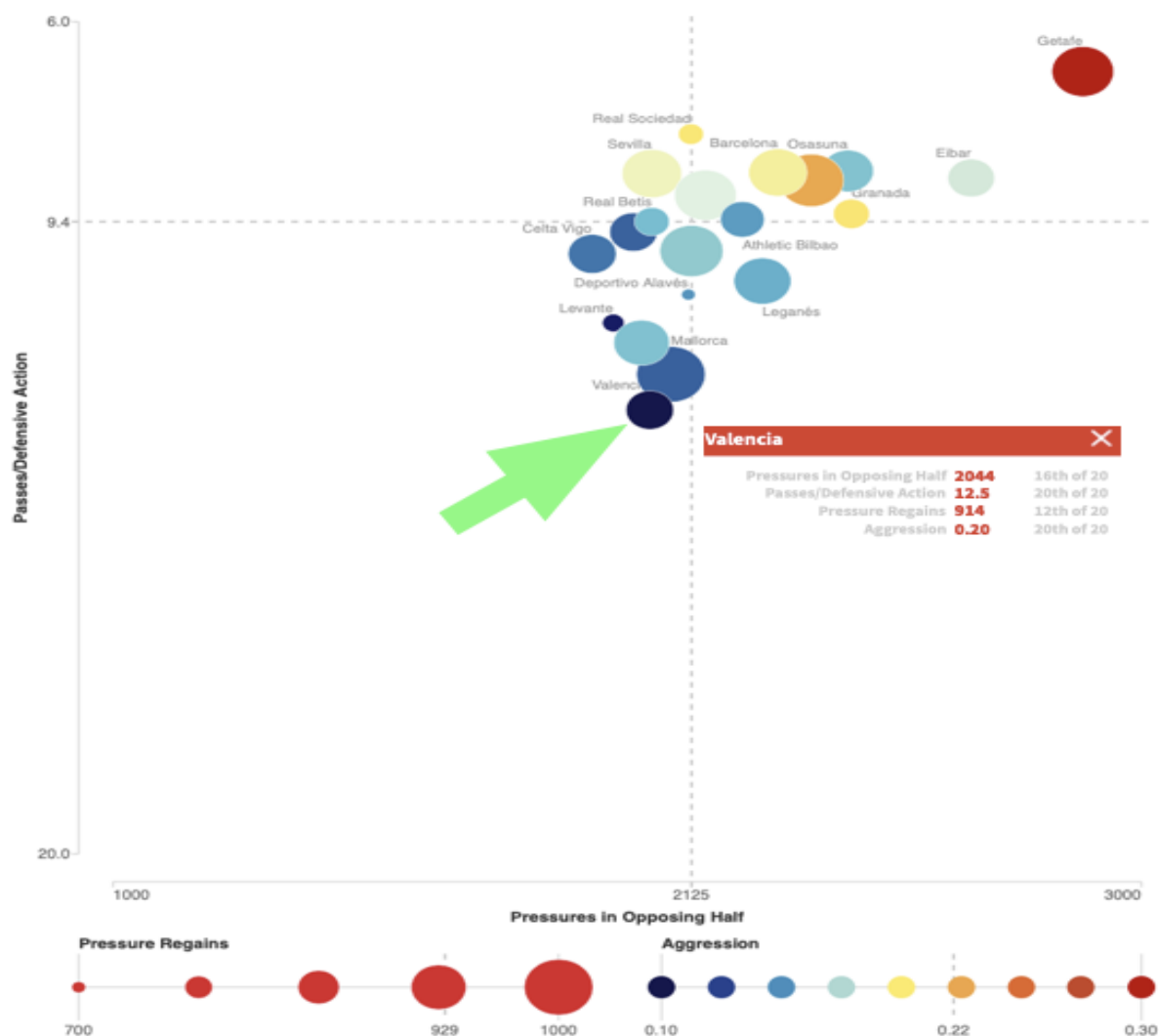
Gráficos de dispersão

Outra das visualizações mais utilizadas são os gráficos de dispersão, que permitem estudar as relações entre duas variáveis do jogo. Através deles, conseguimos posicionar uma equipe ou jogador em relação aos outros, verificar se estão acima ou abaixo da média, etc.

Na figura 28, podemos observar um exemplo de como medir a capacidade de pressão de uma equipe em um único gráfico. Nele, podemos observar até 4 variáveis de pressão. Assim estava o Valencia na temporada 2019/20 em termos de PPDA e pressões no campo adversário. Além disso, na plataforma IQ da StatsBomb, temos a possibilidade de adicionar duas variáveis:

- O tamanho da bolha: representa o dado de "recuperações pós-pressão".
- A cor da bolha: representa o dado de "agressividade" (nível de agressividade da equipe).

Figura 28: Capacidade de pressão de uma equipe em um gráfico de dispersão com 4 variáveis.



Fonte: [Imagem sem título sobre a capacidade de pressão de uma equipe]. (s. f.). Recuperado de https://lh3.googleusercontent.com/dilrueUqUvtmDf4Vyl_rTuVSwgSi_81ZkZbbaifEcEz6YIAGxBIQR04xmOKSxsmbllqUPw=s85

4.3.3. A pressão após perda (counterpressure)

Outra das métricas que devemos considerar para avaliar a capacidade de pressão no campo adversário é a pressão após perda (counterpressure ou gegenpressing), ou seja, as pressões exercidas nos cinco segundos seguintes à perda de posse de bola.

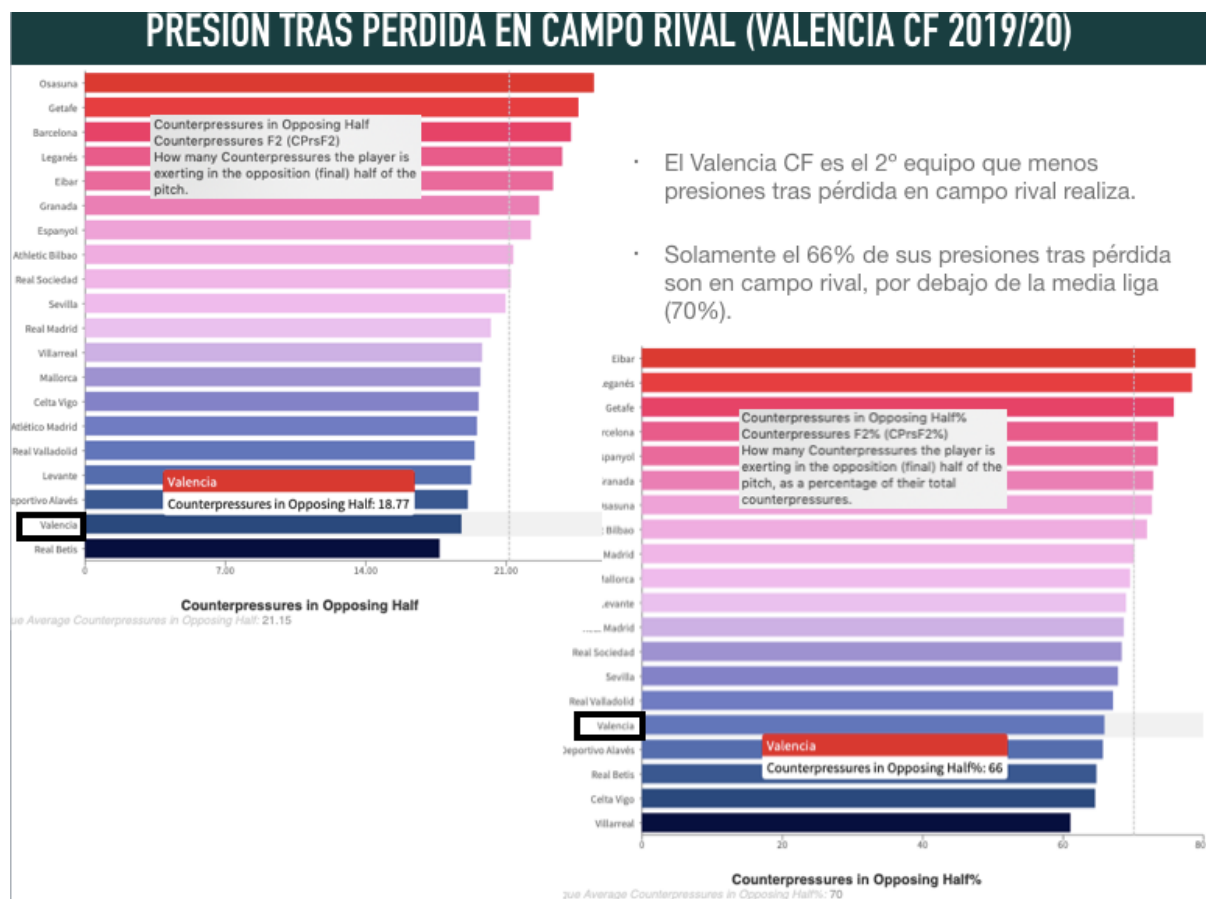
Essa pressão após perda (PTP) ocorre quando uma equipe pressiona o adversário durante os cinco segundos seguintes à perda da posse de bola. Normalmente, é aplicada no campo adversário, onde as perdas de bola costumam ocorrer, e a equipe está próxima o suficiente para que os jogadores ao redor possam realizar essa PTP ou contrapressão.

O objetivo inicial da PTP é interromper o contra-ataque do adversário durante a transição da fase de ataque para a defesa, fornecendo informações relevantes sobre quais jogadores têm maior influência nos momentos da transição ataque-defesa.

Na figura 29, observamos como se pode medir a pressão após perda no campo adversário, com o exemplo do Valencia na temporada 2019/20:

- Quantificando a média de PTP no campo adversário por jogo.
- Quantificando o percentual do total de PTP no campo adversário.

Figura 29: A pressão após perda (counterpressure) no campo adversário.



<p>Presión tras pérdida en campo rival</p> <ul style="list-style-type: none"> - El Valencia CF es el 2º equipo que menos presiones tras pérdida en campo realiza. - Solamente el 66% de sus presiones tras pérdida son en campo rival, por debajo de la media liga (70%) 	<p>Pressão após perda no campo adversário</p> <ul style="list-style-type: none"> • O Valencia CF é o 2º time que menos pressiona após perda de bola no campo adversário. • Apenas 66% de suas pressões após perda ocorrem no campo adversário, abaixo da média da liga (70%).
--	--

Fonte: [Imagem sem título sobre a pressão após a perda]. (s. f.). Recuperado de https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKKC66O4tLENRDvJ7dDv1wYjnkHuYR_cmgkCmqNlv-TDYcxr7-HkM=s114

Até agora, falamos sobre dados para medir a capacidade de pressão das equipes, mas, como medir a capacidade de pressão dos jogadores individualmente?

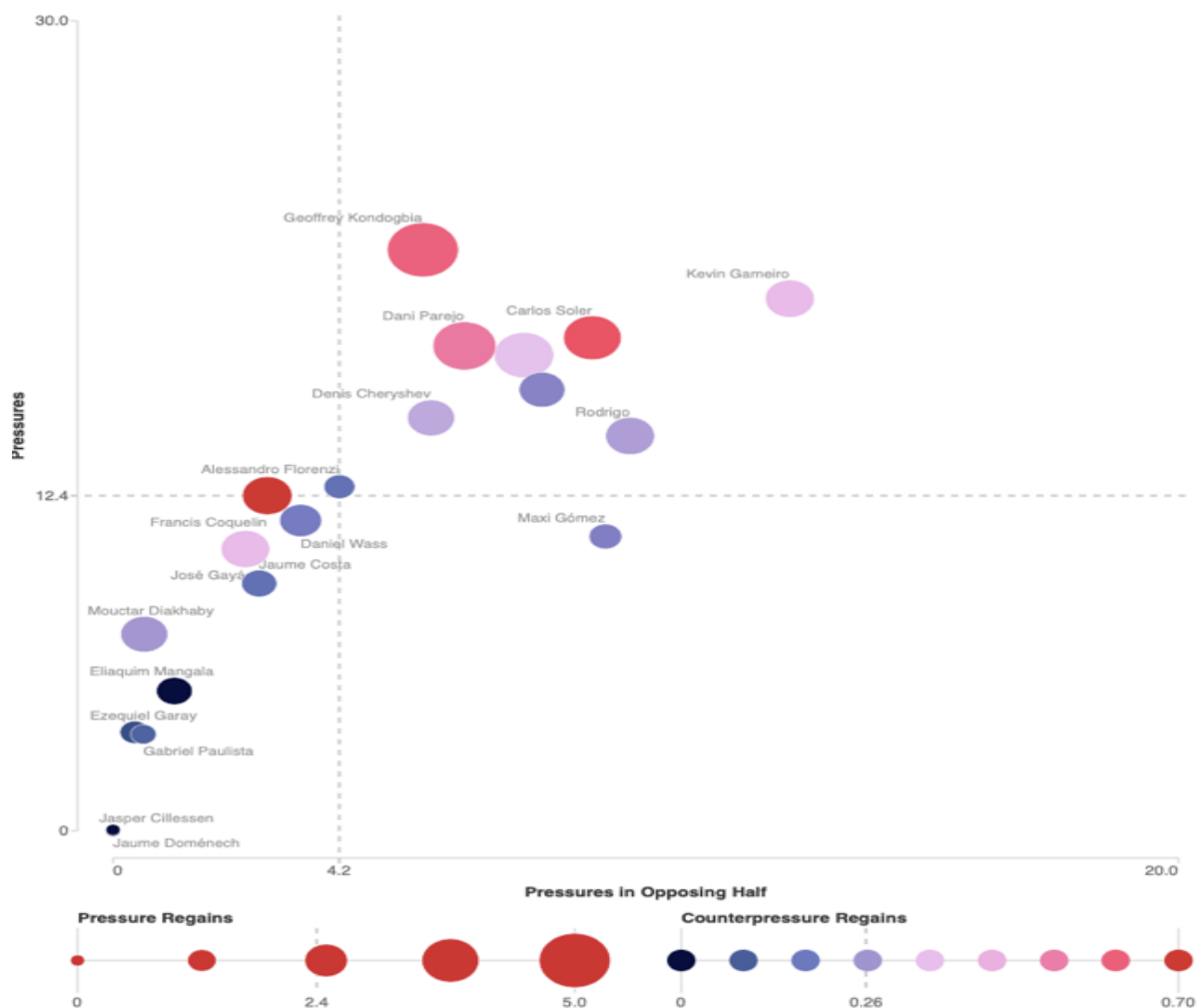
4.3.4. A pressão a nível individual: a recuperação pós-pressão (pressure regain)

A nível individual, o dado da pressão exercida é especialmente relevante para os atacantes e os meio-campistas (principalmente quando falamos de pressão alta). No entanto, também pode ser utilizado para os defensores, quando nos referimos à pressão em todo o campo de jogo.

Não vamos apenas registrar o número de vezes que os jogadores pressionam, mas também o nível de influência que isso tem na recuperação da posse de bola pelos seus companheiros e, em última instância, pela equipe.

E isso será alcançado utilizando a métrica pressure regains (recuperação pós-pressão). Se a equipe recuperar a bola nos cinco segundos seguintes a uma ação de pressão de um jogador, será atribuída uma recuperação pós-pressão ao jogador.

Figura 30: As 4 variáveis que medem a capacidade de pressão de um jogador.



Fonte: [Imagem sem título sobre as quatro variáveis que medem a capacidade de pressão de um jogador]. (s. f.). Recuperado de <https://lh3.googleusercontent.com/D56U0KRjGbQ4I1Y-E73kKoKHqs-wrtDfOtL6UhOOqERi7P8NwOIZpiYge1bGLg6QEML-t7M=s85>

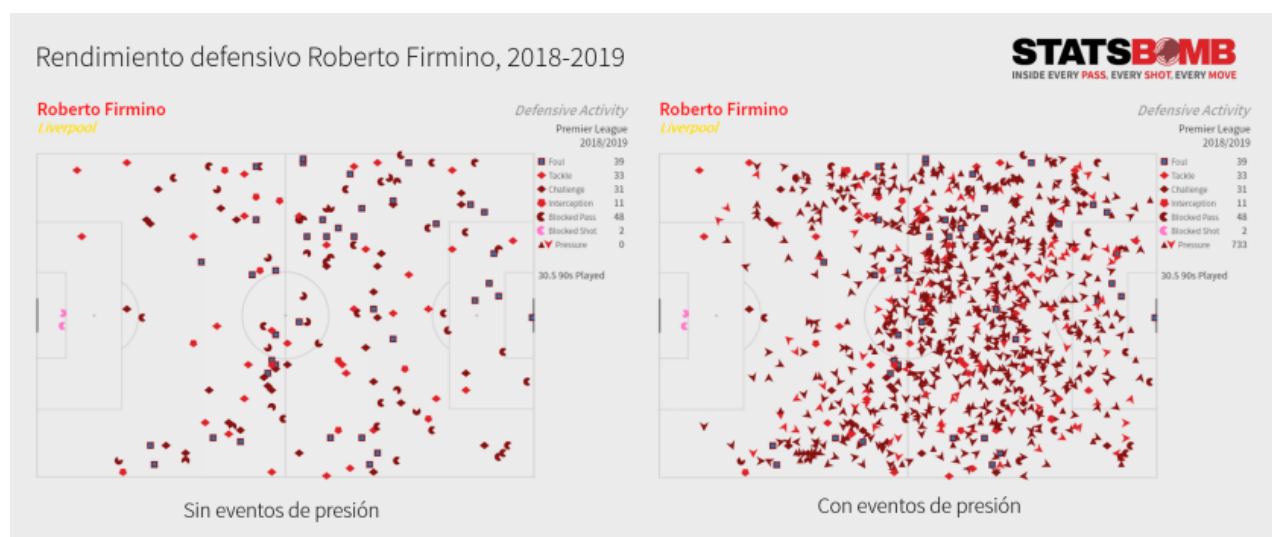
Com essa métrica, não se trata apenas de avaliar o desempenho defensivo dos atacantes e meio-campistas com ações sobre a bola (entradas ou interceptações), mas também de atribuir valor às ações defensivas que realizam para que a equipe recupere a posse de bola.

Ou seja, normalmente o atacante não pressiona para recuperar a bola por si mesmo, mas para cobrir linhas de passe específicas, orientar a saída de bola do adversário para áreas concretas (geralmente para as laterais), forçar os adversários a tomar ações descontroladas, cometer erros ou se livrar da bola. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

Em outros tipos de visualizações, como mapas de pressão, podemos observar cada uma das ações de pressão realizadas pelo jogador em questão, e de uma maneira mais visual, podemos ver qual área de pressão abrange o seu desempenho defensivo.

Assim, na Figura 31, com o objetivo de ver a importância desses eventos de pressão, podemos comparar a visão do desempenho defensivo de Roberto Firmino (2018-2019) excluindo as ações de pressão (gráfico à esquerda) e incluindo-as (gráfico à direita).

Figura 31: Exemplo de mapa de eventos de pressão individual de StatsBomb.



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

4.3.5. Vantagens e usos adicionais dos eventos de pressão

Além das avaliações estritamente defensivas sobre a pressão alta, os eventos de pressão nos permitem obter informações relevantes sobre a fase ofensiva das equipes, como por exemplo, como a pressão exercida por um adversário afeta uma equipe ou um jogador. Ou seja, é possível medir a capacidade de passe de um jogador quando ele está sob a pressão de um rival.

Isso pode ser feito com a métrica AUP ou "Actions under pressure" ou "Passes under pressure". Por exemplo, nas figuras 32 e 33, observamos como, diante da pressão alta forte e sufocante do Getafe, equipes como o Alavés e o FC Barcelona tiveram uma reação totalmente distinta no campo adversário.

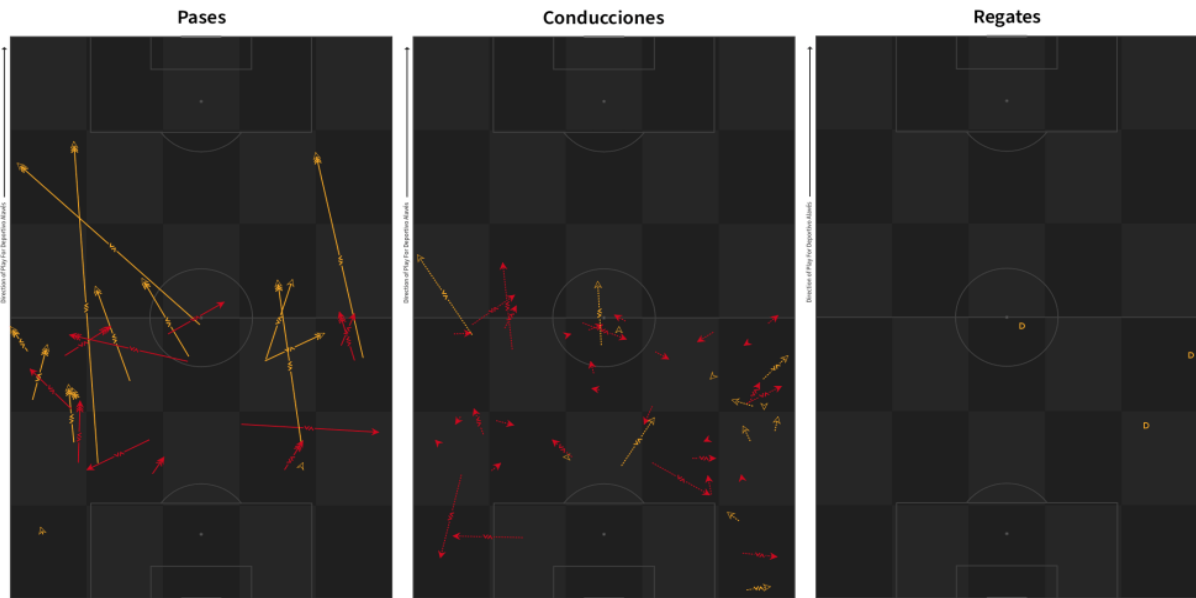
Podemos ver como o Alavés teve muitos problemas para combinar e conectar passes em campo rival:

Sofreu uma queda significativa em relação à sua média da temporada, tanto no percentual de passes completados quanto no percentual de passes completados sob pressão. Não completou nenhum dos três dribles que tentou (vermelho = ação completada; amarelo = ação falhada). (StatsBomb, 11 de fevereiro de 2021).

Figura 32: Mapas de ações realizadas sob pressão (StatsBomb).

Acciones realizadas bajo presión

Deportivo Alavés vs. Getafe, 13-07-2020



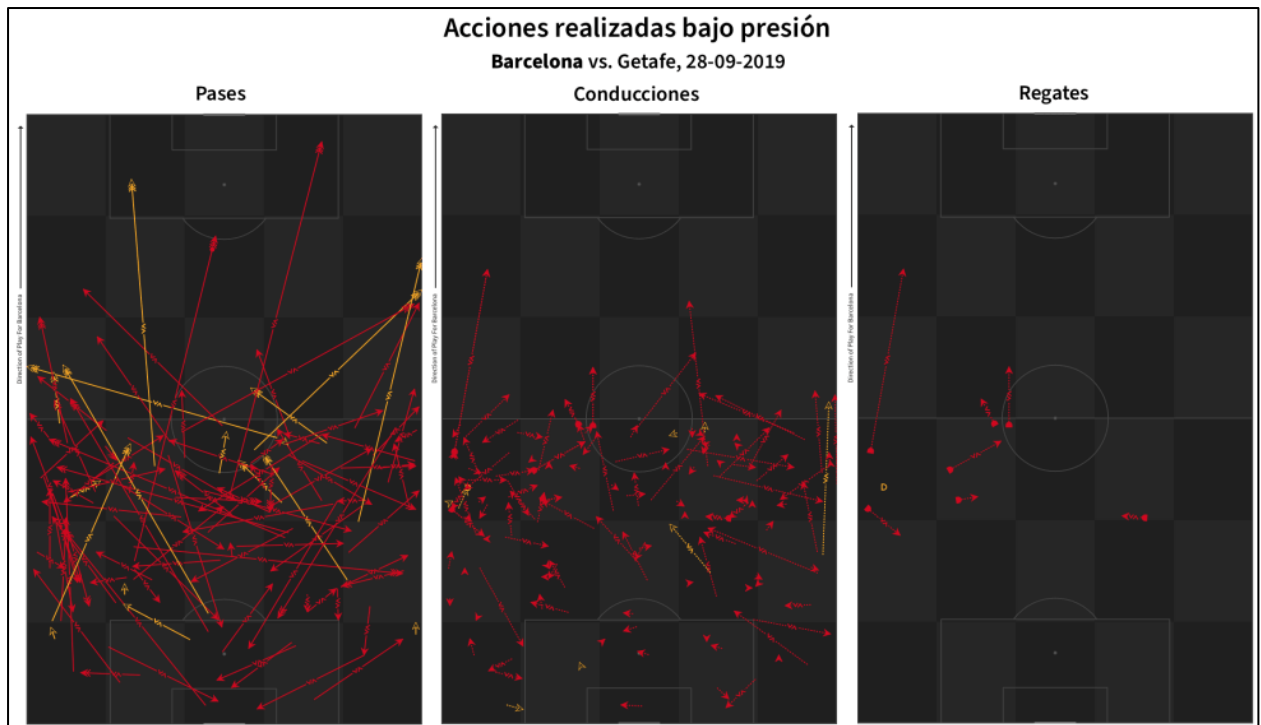
STATSBOMB

Acciones realizadas bajo presión
Pases - Conducciones - Regates

**Ações Realizadas Sob Pressão - Passes -
Conduções - Dribles**

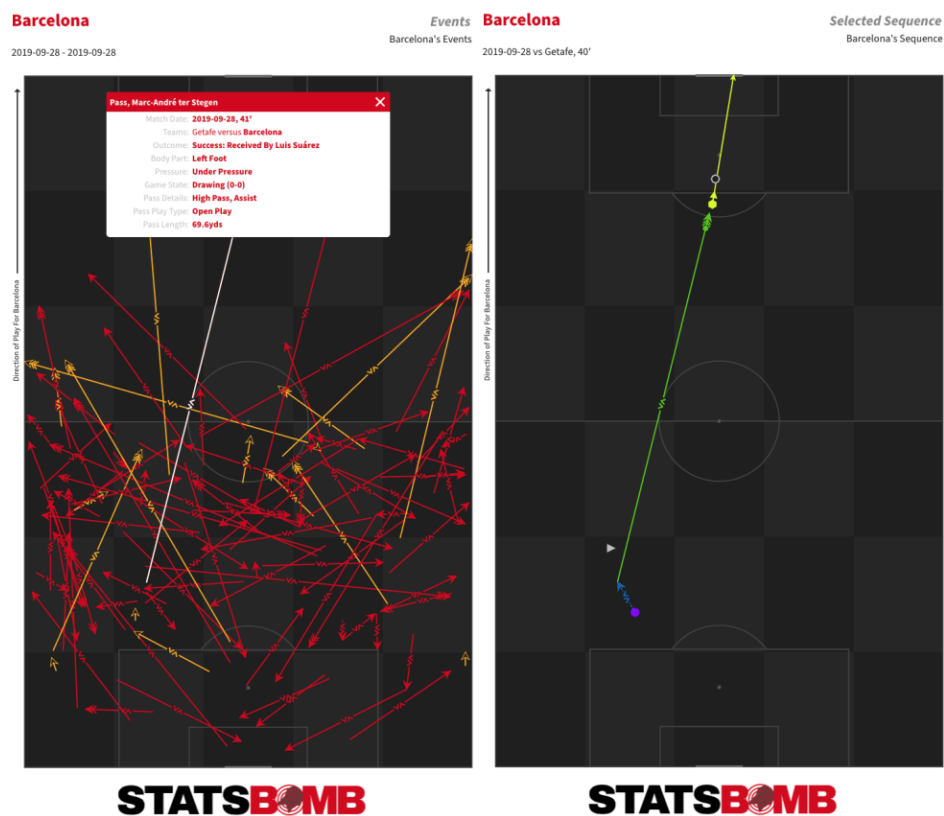
Fonte: StatsBomb, 11 de fevereiro de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

Figura 33: Mapas de ações realizadas sob pressão (StatsBomb).



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

Figura 34: Mapas de ações realizadas sob pressão (StatsBomb).

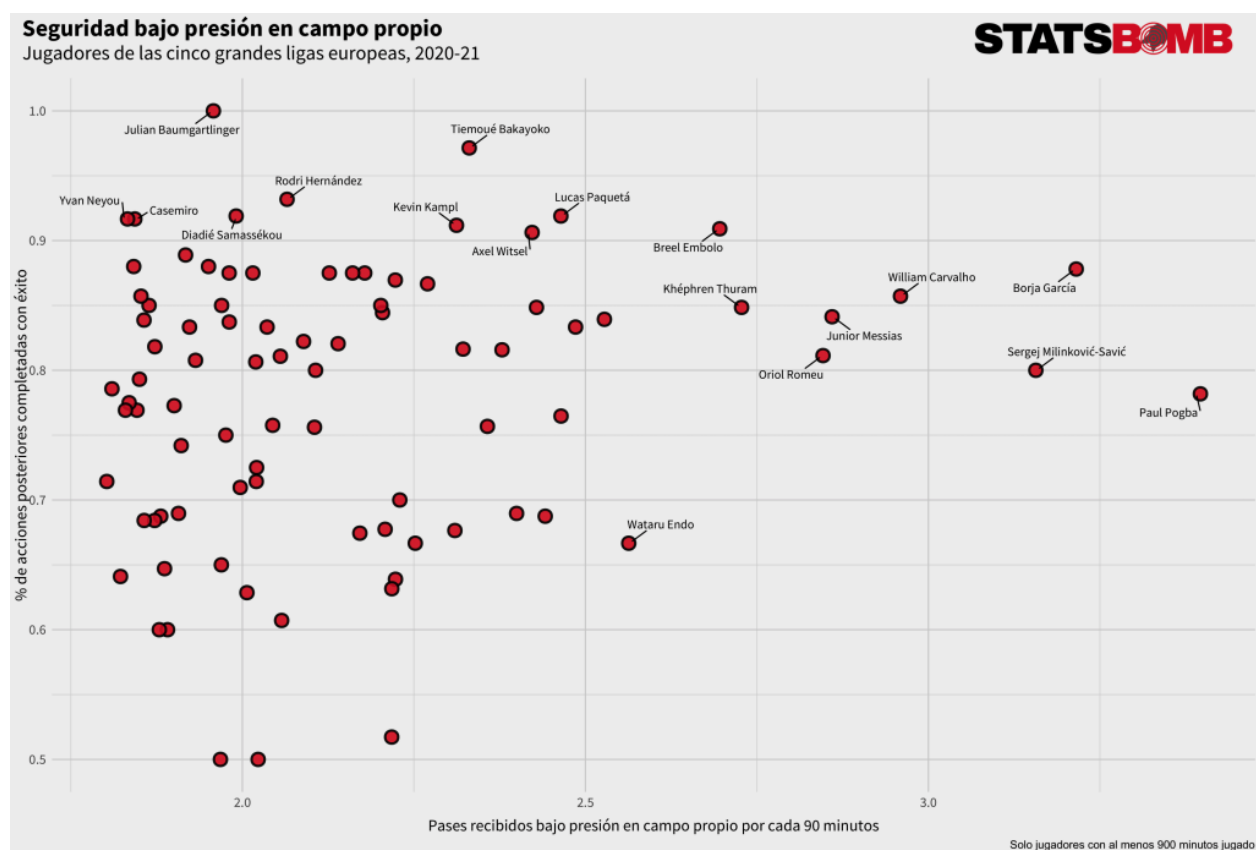


Fonte: StatsBomb, 11 de fevereiro de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

Em vez disso, como podemos observar na figura 34, o Barcelona conseguiu encontrar soluções para contrariar a pressão alta exercida pelo Getafe, conseguindo conectar com os companheiros no campo adversário a partir do seu próprio campo. Inclusive, marcou um gol em uma dessas conexões graças ao goleiro Ter Stegen, que assistiu com um passe no espaço desde seu próprio campo para Luis Suárez marcar.

Outra das métricas ofensivas derivadas dos eventos de pressão coletados pelo StatsBomb é a forma como os receptores de passes se comportam sob a pressão adversária ao receber o passe sob pressão no seu próprio campo. Assim, podemos medir sua capacidade de sair com a bola e dar continuidade ao jogo quando recebem o passe sob pressão em seu próprio campo. Podemos observar isso neste gráfico de dispersão (figura 35):

Figura 35: Recepção de passe sob pressão dos jogadores das cinco grandes ligas europeias.



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

Mas, o que podemos concluir a partir deste gráfico de dispersão? Aqui estão algumas observações sobre a capacidade que alguns desses jogadores têm de dar continuidade ao jogo quando recebem sob pressão:

Alguns nomes se destacam: Paul Pogba, do Manchester United, recebe mais passes sob pressão em seu campo do que qualquer outro jogador, embora perca a bola um pouco mais frequentemente do que jogadores como Borja García do Huesca ou William Carvalho do Real Betis, que também recebem muitos passes assim. Já Julián Baumgartlinger do Bayer Leverkusen recebe um volume menor desses passes (1,96 por 90), mas não perdeu a bola nenhuma vez até agora na temporada. (StatsBomb, 18 de fevereiro de 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-jugadores-de-las-cinco-grandes-ligas-europeas/>).

Unidade 4.3. Chaves para uma boa interpretação dos dados

Em nossas apresentações, usamos as visualizações que os provedores nos oferecem ou as visualizações que nós mesmos criamos. Mas o mais importante é como interpretamos os dados depois de vê-los nessas visualizações.

Aos olhos de diferentes pessoas e com diferentes pontos de vista, uma mesma visualização pode ter diferentes interpretações, tão válidas quanto as outras. O que dará maior ou menor validade a essas interpretações é o raciocínio sobre o motivo pelo qual fazemos essa interpretação, ou seja, as causas ou a origem dessa interpretação, que normalmente vêm de uma acumulação de fatos, ações ou situações de jogo que se repetem durante os jogos.

Por exemplo, na figura 36, no ranking à esquerda, vemos destacado em amarelo que o Sevilla está em último lugar em relação ao número total de pressões exercidas na La Liga 2020/21. No entanto, vemos que está em 2º lugar como a equipe com o maior percentual de pressões bem-sucedidas (com recuperação).

Se olharmos apenas o ranking à esquerda (o número de pressões exercidas), podemos interpretar que o Sevilla é uma das equipes que pior pressiona, mas precisamos dar contexto a esse dado. Fazemos isso da seguinte maneira:

- Explorando os dados ao máximo para uma maior compreensão: não devemos nos deixar levar pela pressa ou pelas primeiras impressões, mas devemos nos esforçar para entender o significado do valor que está sendo registrado (nesse caso, o número de pressões realizadas, sem considerar se o time recupera a bola ou não). O Sevilla pode realizar um número menor de pressões simplesmente porque tem maior posse de bola que a maioria das equipes.
- Contrastando os dados com outros valores relacionados ao dado em questão: isso nos dará um contexto adequado para avaliar e interpretar de forma mais válida aquele dado. Por exemplo, com a visualização do ranking à direita, estamos adicionando contexto ao ranking da esquerda. Isso nos mostra que o Sevilla é o time que menos pressiona, mas é o segundo time que mais recupera bolas após exercer a pressão. Portanto, estamos falando de uma das equipes mais eficazes no momento de pressionar.

Figura 36: Rankings de pressões totais exercidas e por terços de campo na web Fbref.

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	ShSv	Pass	
Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	
Celta 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	104	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	

Squad Stats		Successful Pressure Percentage									
Percentage of time the squad gained possession within 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	Int
Eibar	5015	1602	31.9	1272	2251	1492	553	75	2	478	390
Sevilla	4716	1477	31.3	1392	2199	1125	503	95	3	408	334
Barcelona	5075	1546	30.5	1404	2330	1341	519	82	0	437	354
Atlético Madrid	5124	1540	30.1	1665	2328	1131	579	93	0	486	395
Real Sociedad	5070	1487	29.3	1423	2340	1307	497	104	2	393	337
Real Madrid	4880	1425	29.2	1562	2175	1143	535	104	4	435	406
Osasuna	5329	1550	29.1	1558	2487	1284	557	109	2	448	384
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	432
Villarreal	4994	1425	28.5	1587	2226	1181	578	117	3	461	354
Celta Vigo	5997	1701	28.4	2063	2850	1084	610	97	1	513	375
Eibar	5168	1455	28.2	2068	2228	872	553	111	2	442	388
Levante	5363	1499	28.0	1697	2350	1316	538	108	3	430	445
Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	420
Valladolid	5839	1624	27.8	2051	2600	1188	601	117	2	484	390
Alavés	5447	1510	27.7	1752	2541	1154	568	105	4	463	432
Athletic Club	5250	1451	27.6	1494	2380	1376	553	86	2	467	348
Cádiz	5264	1433	27.2	2094	2304	866	612	118	4	494	411
Granada	5027	1358	27.0	1616	2345	1066	611	142	5	469	406
Valencia	4855	1305	26.9	1794	2067	994	566	137	2	429	321

Fonte: adaptação própria do site Fbref.com, s. f., <https://fbref.com/en/comps/12/La-Liga-Stats>

Na figura 37, Fredi Martín nos oferece uma série de chaves — com as quais estamos totalmente de acordo — para tratar e dar a correta interpretação aos dados.

Figura 37: Chaves para uma correta interpretação dos dados.

- Comprender el deporte del que estoy tratando
- Explorar lo 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

<p>Tratar los datos</p> <ul style="list-style-type: none"> • Comprender el deporte del que estoy tratando • Explorar los datos para una mejor comprensión 	<p>Tratamento dos Dados</p> <ul style="list-style-type: none"> • Comprender o esporte sobre o qual estou analisando
---	---

<ul style="list-style-type: none"> ● 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 espacios de reflexión 	<ul style="list-style-type: none"> ● Explorar os dados para uma melhor compreensão ● Contextualizar os dados ● Normalizar os dados ● Definir a mensagem que se deseja transmitir com os dados ● Fazer uma apresentação/visualização correta dos dados ● Ser conciso ● Compreensão de Métricas Avançadas: xG, VAEP, xT ● Minhas conclusões podem ser diferentes das de outros destinatários, deixando espaço para reflexão.
---	---

Fonte: [Imagem sem título sobre as chaves para uma correta interpretação dos dados]. (s. f.). Recuperado de https://lh3.googleusercontent.com/D3OigKfmY4FIR3Zv_0IF5HEm-PM2akwwbRIOR3d9me5JMY8UKXVB9U_v4W0rmnWZLiyUxv4=s109

De todas essas chaves, vamos comentar algumas delas. Podemos concluir que, para oferecer uma boa interpretação dos dados em nossas apresentações, é necessário:

1. Compreender o futebol como esporte em toda sua dimensão: quanto mais soubermos, mais contexto podemos dar ao dado e, portanto, melhor será a interpretação.
2. Explorar os dados ao máximo: para ter uma visão mais ampla do espectro a ser estudado (as pressões, etc.). Isso nos permitirá uma melhor compreensão do dado e, assim, proporcionar uma melhor interpretação.
3. Contextualizar o dado: é necessário buscar e encontrar as razões e os motivos por trás da interpretação que estamos dando ao dado. Por exemplo, os dados mostram que Marcos Llorente tem mais gols e assistências no Atlético de Madrid do que quando jogava no Real Madrid. Para dar contexto a esse ou esses dados, precisamos fazer as perguntas e hipóteses adequadas: Por que Llorente, quando jogava no Real Madrid, não chegava tanto à área adversária e não marcava ou assistia tanto quanto no Atlético de Madrid sob o comando de Simeone? Uma explicação pode ser que o treinador pedia coisas diferentes (que oferecesse mais equilíbrio, que não fosse tanto ao ataque porque outros jogadores finalizavam as jogadas, etc.) ou talvez o jogador tenha evoluído ofensivamente para ter mais presença na área. Podem existir várias respostas para essas perguntas. O importante é analisá-las para chegar a uma interpretação mais próxima da realidade.

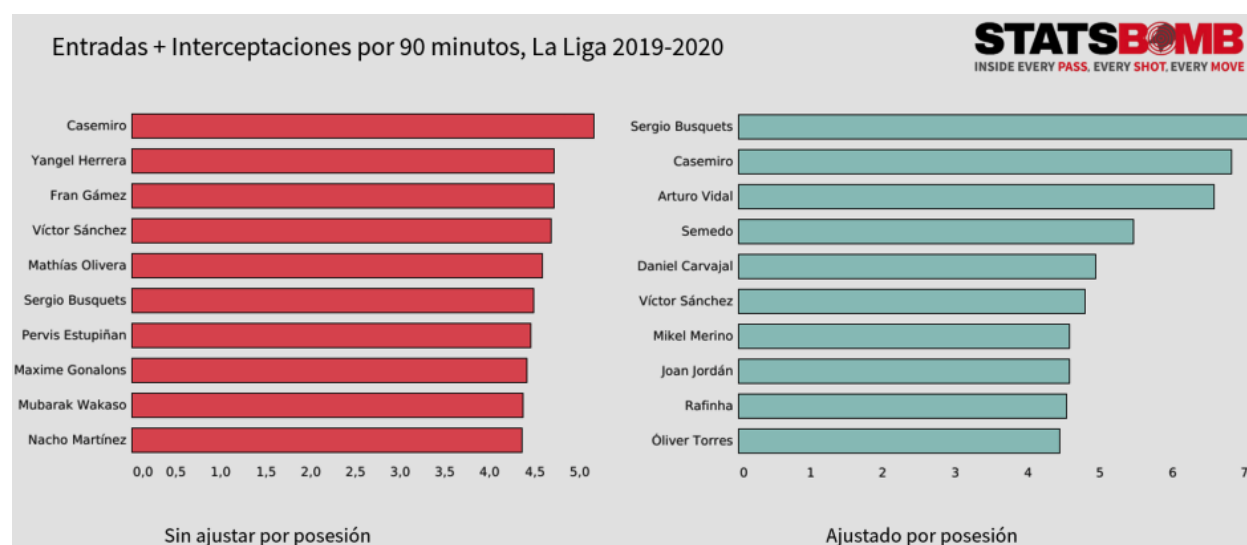
- Normalizar os dados: quando nos depararmos com visualizações de dados defensivos de jogadores ou equipes, como interceptações, desarmes, duelos, etc., é crucial normalizá-los. Ou seja, ajustar essas métricas defensivas de acordo com a posse de bola. O estilo de jogo das equipes afeta os dados registrados em valores absolutos. Os jogadores só podem realizar ações defensivas quando sua equipe não tem a posse de bola. Portanto, dependendo do tempo em que a equipe está com a posse, eles terão mais ou menos oportunidades de realizar ações defensivas. Isso significa que equipes com maior posse de bola tendem a ter menos ações defensivas, e os números absolutos de desarmes, interceptações (ou pressões) podem ser distorcidos.

O ajuste com base na posse é uma solução prática para padronizar e normalizar os valores, permitindo comparações entre diferentes jogadores ou equipes, corrigindo fatores externos ao seu estilo de jogo ou habilidade. No entanto, é importante ressaltar que, uma vez ajustados pela posse, os números não correspondem mais diretamente às ações reais observadas em vídeo. Portanto, os números absolutos ainda são úteis para avaliar o desempenho dos jogadores (por exemplo, para definir limites).

Na Figura 38, vemos um ranking das interceptações e desarmes (por 90 minutos) dos jogadores da La Liga. À esquerda estão os dados absolutos, enquanto à direita estão os dados ajustados com base na posse de bola de suas equipes.

Como podemos observar, jogadores como Casemiro não aparecem no ranking de valores absolutos simplesmente porque têm menos oportunidades de somar ações defensivas, já que possuem mais posse de bola do que jogadores de equipes com menos posse, que são forçados a defender por mais tempo, aumentando suas chances de realizar ações defensivas.

Figura 38: Normalização dos dados por ajuste de posse.



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

Realizar uma boa apresentação/visualização dos dados, bem como concretizar a mensagem e ser conciso: já sabemos que a simplicidade das visualizações que apresentamos, assim como a brevidade e clareza das nossas opiniões, facilita o entendimento por parte do receptor.

Deixar espaço para a reflexão, sabendo que as opiniões dos outros podem ser diferentes: a subjetividade que o futebol carrega em seu DNA também se reflete nos dados que ele gera (se uma mesma ação no jogo pode ser avaliada de maneira diferente por cada um dos observadores, como não aconteceria com os dados gerados pelo próprio jogo?). Por isso, devemos ser respeitosos com as opiniões alheias, desde que essas opiniões sejam fundamentadas da mesma forma bem trabalhada que a nossa, após termos realizado todos os processos mencionados anteriormente e aplicado todas essas chaves para uma boa interpretação.

Conclusões

É evidente que o futebol evoluiu ao longo de sua história e continuará a evoluir. Porém, na área da aplicação do big data e da inteligência artificial (IA), o futebol entrou em uma espiral de modernização que obriga treinadores, assistentes, diretores esportivos e todos que querem fazer parte do seu entorno a se formar para se manter atualizados e compreendê-la. Isso está se tornando cada vez mais importante nos processos de melhoria do desempenho dos jogadores e do jogo das equipes.

Dito isso, é importante ter conhecimentos sobre o futebol, tanto em sua dimensão técnico-tática condicional quanto no processo de treinamento e, também, na análise dos dados gerados. Os dados são cada vez mais utilizados pelos clubes e equipes técnicas, que contratam profissionais especializados para realizar análises úteis e proveitosas, melhorando a tomada de decisões diária nos diversos aspectos mencionados anteriormente.

Por isso, conhecer esse mundo emergente dos dados é essencial para não ficar para trás, aproveitar ao máximo essa ferramenta e melhorar nossas habilidades profissionais, além do desempenho das equipes.

O conhecimento dos processos de tratamento e visualização dos dados defensivos apresentados neste módulo 4 ajudará a melhorar nossas habilidades como analistas e/ou treinadores ao desvendar o jogo. Aprendemos a fazer isso por meio da análise de vídeo e agora precisamos incorporar os dados como um elemento adicional para ampliar nosso ponto de vista, tornando-o bidirecional:

- **Ir do jogo para o dado:** apoiar-se nas visualizações dos dados para reforçar as opiniões que formamos ao analisar a jogada por meio do vídeo-análise e observar os comportamentos, tanto próprios quanto do adversário.
- **Ir do dado para o jogo:** obter informações relevantes sobre o jogo por meio dos dados que dificilmente conseguiríamos com a análise de vídeo (por falta de tempo,

atenção ou simplesmente devido ao grande volume de partidas que precisamos comparar). Por exemplo, obter todos os passes ao espaço realizados por uma equipe desde seu próprio campo em todas as partidas da La Liga, etc.

Como despedida deste curso, quero apenas dizer que foi um grande desafio para mim compartilhar tudo o que aprendi sobre o tema da pressão alta e dividir com todos vocês. Me considero muito afortunado por ter aprendido com treinadores e colegas tão capacitados. Para mim, é um prazer compartilhar com vocês todo o conhecimento que é fruto da minha formação e das minhas experiências anteriores.

Espero que este curso tenha sido do seu agrado e útil para o seu desenvolvimento profissional. Espero vê-los em breve, trabalhando em uma comissão técnica profissional, caso ainda não estejam fazendo isso.

Referências

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-ZxvheJkzXhDva8cLQOAccgCwF1wFHRD9PP_SWnRQarn_OlyVU5Ktwc1cgrJY71bYvxQ=s130

[Imagen sin título sobre la capacidad de presión de un equipo]. (s. f.). Recuperado de https://lh3.googleusercontent.com/dilrueUqUvtmDf4Vyl_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-wrtDfOtL6UhOQcERi7P8NwOIZpjYge1bGLg6QEML-t7M=s85>

[Imagen sin título sobre la presión tras la pérdida]. (s. f.). Recuperado de https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKKC66O4tLENRDvJ7dDv1wYjnkHuYR_cmgkCmqNlv-TDYcxr7-HkM=s114

[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

[Imagen sin título sobre PPDA en un *ranking* de StatsBomb]. (s. f.). Recuperado de https://lh3.googleusercontent.com/4QRdqGWNi0_bJ_Zs392Z0mH96LDHnlenpyl145mhA9-nOsTvqJTvUzR4NS2CSroCXb0U=s85

[Imagen sin título sobre *ranking* de presión en el campo rival]. (s. f.). Recuperado de <https://lh3.googleusercontent.com/CFilwr8zUC-i9Gk-kyX5x-NEZRYcqj2QH4WoN63PAtWrea3qbb1Avq4JGoepUMUvt1feuA=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/>