

GRADO EN CIENCIA DE DATOS



VNIVERSITAT
ID VALÈNCIA

TRABAJO FIN DE GRADO

APLICACIÓN PARA LA VISUALIZACIÓN DE DATOS FUTBOLÍSTICOS

AUTOR:

ALEJANDRO HERNÁNDEZ BENEITO

TUTORES:

VALERO LAPARRA PÉREZ-MUELAS

JORGE VILA TOMÁS



VNIVERSITAT
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria **ETSE-UV**

TRABAJO FIN DE GRADO

APLICACIÓN PARA LA VISUALIZACIÓN DE DATOS FUTBOLÍSTICOS

AUTOR: ALEJANDRO HERNÁNDEZ BENEITO

**TUTORES: VALERO LAPARRA PÉREZ-MUELAS
JORGE VILA TOMÁS**

Declaración de autoría:

Yo, Alejandro Hernández Beneito, declaro la autoría del Trabajo Fin de Grado titulado “Aplicación para la visualización de datos futbolísticos ” y que el citado trabajo no infringe las leyes en vigor sobre propiedad intelectual. El material no original que figura en este trabajo ha sido atribuido a sus legítimos autores.

Valencia, 18 de julio de 2025

Fdo: Alejandro Hernández Beneito

Resumen:

Este trabajo tiene como objetivo desarrollar una prueba de concepto de una plataforma interactiva para el análisis de fútbol, orientada a cuerpos técnicos, analistas y departamentos de *scouting*. Se pretende mostrar cómo el análisis estadístico y la visión por computadora pueden integrarse en una herramienta práctica que apoye la evaluación de jugadores, el análisis de partidos y la toma de decisiones basada en datos.

La propuesta se basa en el uso de datos abiertos, técnicas de ciencia de datos y modelos de visión artificial preentrenados, todo ello integrado en una interfaz accesible mediante **Shiny for Python**. El resultado es una solución que, sin depender de software privativo, permite explorar métricas avanzadas, detectar patrones de juego y generar visualizaciones adaptadas a diferentes perfiles de usuario.

El proyecto se organiza en tres bloques principales:

- **Análisis de competiciones oficiales:** A partir de datos de **StatsBomb**, se analizan partidos reales (como los disputados en la Eurocopa 2024), con métricas colectivas y visualizaciones tácticas que permiten interpretar el rendimiento de selecciones y equipos de forma detallada.
 - **Evaluación de jugadores y perfiles:** Utilizando datos históricos de **FBref**, se comparan jugadores de las cinco grandes ligas europeas mediante gráficos personalizados, rankings por posición, modelos de predicción (goles/asistencias) y un sistema de búsqueda por perfil.
 - **Visión por computadora aplicada al vídeo:** Se aplican modelos preentrenados como **YOLOv8** para detectar jugadores en secuencias de vídeo futbolístico. El objetivo es explorar el potencial de esta tecnología para automatizar parte del análisis visual y ofrecer una capa adicional de comprensión táctica.
-

Abstract:

This project aims to develop a proof of concept for an interactive football analysis platform designed for coaching staff, analysts, and scouting departments. The goal is to demonstrate how statistical analysis and computer vision can be integrated into a practical tool that supports player evaluation, match analysis, and data-driven decision-making.

The solution is built using open data sources, data science techniques, and pre-trained computer vision models — all integrated into a user-friendly interface developed with **Shiny for Python**. The result is a flexible platform capable of exploring advanced metrics, identifying performance patterns, and generating customized visualizations.

The project is structured around three core components:

- **Analysis of official competitions:** Using **StatsBomb** data, real matches (e.g., UEFA Euro 2024) are analyzed with collective metrics and tactical visualizations to evaluate team performance in depth.
 - **Player profiling and evaluation:** With historical data from **FBref**, players from Europe's top five leagues are compared through radar plots, percentile rankings, predictive models (goals/assists), and a profile-based player search system.
 - **Computer vision applied to video:** Pretrained models such as **YOLOv8** are used to detect players in football match footage. This module explores how visual detection can automate and enrich tactical analysis.
-

Resum:

Aquest treball té com a objectiu desenvolupar una prova de concepte d'una plataforma interactiva d'anàlisi futbolística, orientada a cossos tècnics, analistes i departaments de *scouting*. Es pretén mostrar com l'anàlisi estadística i la visió per computador poden combinar-se en una eina pràctica per a l'avaluació de jugadors, l'anàlisi de partits i la presa de decisions basada en dades.

La proposta es fonamenta en l'ús de fonts obertes, tècniques de ciència de dades i models de visió artificial preentrenats, integrats en una interfície intuïtiva desenvolupada amb **Shiny for Python**. El resultat és una solució versàtil que permet explorar mètriques avançades, detectar patrons de joc i generar visualitzacions adaptades a cada necessitat.

El projecte s'estructura en tres blocs fonamentals:

- **Anàlisi de competicions oficials:** A partir de dades de StatsBomb, s'analitzen partits reals (com l'Eurocopa 2024) amb mètriques col·lectives i visualitzacions tàctiques que permeten avaluar el rendiment dels equips en profunditat.
 - **Avaluació de jugadors i perfils:** Amb dades històriques de FBref, es comparen jugadors de les cinc grans lligues europees mitjançant gràfics personalitzats, rànquings per posició, models predictius (gols/assistències) i un cercador per perfil.
 - **Visió per computador aplicada al vídeo:** Es fan servir models com YOLOv8 per detectar jugadors en seqüències de vídeo. L'objectiu és explorar com aquesta tecnologia pot automatitzar l'anàlisi visual i millorar la comprensió tàctica del joc.
-

Agradecimientos:

Quisiera expresar mi más sincero agradecimiento a todas las personas que han hecho posible la realización de este Trabajo de Fin de Grado.

En primer lugar y el más importante, a mi familia. Gracias no solo por estos 4 últimos años de apoyo y cariño, sino por los más de 21 que llevo a vuestro lado.

Gracias, papá y mamá, por aguantarme y apoyarme incondicionalmente.

Gracias, Lucas; yo también estoy orgulloso de ti, aunque no te lo diga.

También a mis tíos y primos, por estar siempre ahí.

Y, por último, gracias a mis abuelos, que como diría uno de ellos: he hincado los codos. Gracias tanto a los tres que siguen conmigo como al que me cuida allí donde esté.

También quiero agradecer a mis tutores Valero y Jorge, por su orientación, paciencia y valiosas sugerencias a lo largo de todo el desarrollo del proyecto. Su experiencia y apoyo han sido fundamentales para dar forma a este trabajo.

A mis compañeros y amigos, tanto a los de toda la vida como a los que he conocido durante la carrera, gracias por estar ahí en los buenos y en los malos momentos, y por hacer más llevadero el camino con vuestras conversaciones, consejos y compañía.

A todos, gracias.

Índice general

1. Introducción, Antecedentes y Motivación	11
1.1. Introducción	11
1.2. Antecedentes	11
1.3. Casos reales de aplicación del Big Data en el fútbol	12
1.4. Motivación	13
1.5. Objetivos del trabajo	13
2. Marco Teórico	16
2.1. ¿Qué es Big Data?	16
2.1.1. Breve historia del Big Data	17
2.1.2. Big Data en la actualidad: usos y beneficios	18
2.1.3. Principales beneficios del uso de Big Data	20
2.2. ¿Qué es el <i>Big Data</i> en el Deporte?	21
2.2.1. Orígenes	21
2.2.2. Evolución tecnológica en el deporte	24
2.2.3. El papel del <i>Big Data</i> en el deporte	24
2.2.4. Beneficios y desafíos del <i>Big Data</i> en el deporte	28
2.3. <i>Big Data</i> en el fútbol	30
2.3.1. Usos del <i>Big Data</i> en el fútbol	31
2.3.2. Tipos de datos utilizados en el fútbol	32
2.4. Principales proveedores de datos en el fútbol	36
2.5. Metodologías de análisis de datos en el fútbol	39
3. Materiales	46
3.1. Datos del estudio	46
3.2. Limpieza y preprocesamiento de datos	49
4. Modelos Predictivos	53
4.1. Modelos predictivos	53
4.2. Enfoque temporal	53

4.3.	Modelo de predicción de goles con XGBoost	53
4.3.1.	Preparación de los datos	54
4.3.2.	Variables predictoras	54
4.3.3.	Modelo y ajuste	54
4.3.4.	Evaluación del modelo	56
4.3.5.	Importancia de variables	56
4.4.	Modelo de predicción de asistencias con Random Forest	57
4.4.1.	Preparación de los datos	57
4.4.2.	Variables predictoras	58
4.4.3.	Modelo y ajuste	58
4.4.4.	Evaluación del modelo	59
4.4.5.	Importancia de variables	59
5.	Visión Por Computadora	61
5.1.	Visión por Computadora y Deep Learning aplicados al análisis futbolístico	61
5.2.	Objetivo del módulo y enfoque general	61
5.3.	Configuración inicial y dependencias	61
5.4.	Módulo de detección y clasificación de jugadores, balón y árbitros	62
5.5.	Módulo de seguimiento multiobjeto	63
5.6.	Módulo de detección y proyección de puntos clave del campo	63
5.7.	Módulo de proyección espacial al campo (homografía)	64
5.8.	Módulo de seguimiento del balón	65
5.9.	Funcionamiento de la función generadora de vídeos anotados	66
5.10.	Reflexión final	68
6.	Aplicación	69
6.1.	Análisis de Competiciones	70
6.1.1.	Análisis de eventos a nivel equipo	71
6.1.2.	Análisis de eventos a nivel jugador	83
6.2.	Análisis de Jugadores	89
6.2.1.	Top Jugadores	89
6.2.2.	Gráficos	93
6.2.3.	Buscador	100
6.3.	Visión por Computadora	102
6.4.	Conclusión	103
7.	Conclusiones y proyección futura.	104
7.1.	Conclusiones	104

7.2. Trabajo futuro	105
8. Anexos	107
Bibliografía	119

Capítulo 1

Introducción, Antecedentes y Motivación

1.1. Introducción

El fútbol, el deporte que levanta pasiones y conecta a millones de personas en todo el planeta, ha experimentado una transformación silenciosa, pero profunda, en los últimos años. Ya no se trata solo de lo que ocurre durante los 90 minutos en el campo. Y es que la tecnología, poco a poco, ha ido ganando terreno, cambiando nuestra manera de ver, entender y analizar el juego.

Hoy en día, los datos ocupan un lugar central en la toma de decisiones futbolísticas. Desde el cuerpo técnico hasta las oficinas de dirección deportiva, pasando por analistas y ojeadores, todos buscan lo mismo: comprender mejor el juego. El *Big Data*, el aprendizaje automático y las visualizaciones interactivas han abierto una nueva dimensión. Ahora es posible medir el impacto real de un jugador, anticiparse a su rendimiento o incluso detectar talento donde otros solo ven números sueltos.

En este trabajo se presenta una aplicación interactiva que une lo mejor de dos mundos: el análisis estadístico del rendimiento de jugadores y equipos, y la visión por computadora para interpretar lo que sucede en el terreno de juego a través de vídeo. La idea es clara: mostrar cómo estas tecnologías pueden convivir y complementarse para ofrecer una perspectiva más rica, más completa. Y, sobre todo, más útil.

1.2. Antecedentes

El fútbol genera una cantidad de datos abrumadora. Desde estadísticas tradicionales hasta métricas avanzadas extraídas de secuencias en vídeo. Y lo más fascinante es cómo se capturan.

“Las cámaras realizan el seguimiento de jugadores, árbitros y balón en el campo, registrando la posición de cada uno de ellos con gran precisión a un ritmo vertiginoso de 25 veces por segundo, para generar más de 3.500.000 de datos por partido que se traducen en unas 2.000 métricas por jugador.” (LALIGA, 2024) [1].

Con semejante volumen de información, no es de extrañar que el análisis basado en datos haya cobrado tanto protagonismo. Herramientas como **Opta Sports**, ampliamente utilizada en la Premier League, o **Beyond Stats**, desarrollada por LaLiga y Microsoft, permiten traducir todos esos datos en algo tangible: patrones de pase, zonas de presión, contribuciones al xG, y un largo etcétera.

Este proyecto toma como punto de partida ese mismo enfoque: combinar los datos más relevantes con técnicas avanzadas de análisis y visualización, y presentarlos de forma accesible mediante una aplicación construida con **Shiny**.

1.3. Casos reales de aplicación del Big Data en el fútbol

Para entender mejor hasta qué punto los datos están transformando el fútbol, basta con mirar algunos ejemplos concretos. Son historias reales, de clubes y jugadores que apostaron por la información y salieron ganando.

- **Kevin De Bruyne y su renovación con el Manchester City (2021):** lejos de lo habitual, De Bruyne decidió negociar su contrato sin agentes. ¿Su estrategia? Encargar un informe estadístico que demostraba, con datos en mano, su impacto real en el equipo. Métricas como la contribución al xG y comparativas con otros centrocampistas fueron clave para conseguir una renovación más que merecida, como así lo explica el artículo publicado por *The Athletic* (2024)[2].
- **Brentford FC, el club que desafió las reglas del juego:** sin el presupuesto de los grandes, el Brentford apostó por el análisis de datos como piedra angular. Detectaron talento infravalorado, como Ivan Toney u Ollie Watkins, y lo ficharon por precios bajos. ¿El resultado? Ventas millonarias y una plantilla competitiva. Incluso decidieron cerrar su cantera tradicional y crear el “Brentford B”, captando jugadores a través de algoritmos y métricas avanzadas. Este caso fue ampliamente cubierto por diversos medios, como se refleja en un reportaje de *Relevo* (2023)[3].
- **Liverpool y la estadística como brújula táctica:** cuando Klopp fue fichado en 2015, el club se apoyó en datos que demostraban que su mala temporada en el Dortmund había sido pura mala suerte. Desde entonces, el Big Data ha guiado sus decisiones. Así llegaron fichajes como Salah o Robertson, jugadores que no estaban en el radar de los grandes, pero brillaban en los modelos estadísticos del club. Este enfoque ha sido descrito en profundidad en un análisis de *Fintualist* (2022)[4].
- **LaLiga y Microsoft – Proyecto Beyond Stats:** como se comenta en un artículo publicado por *Microsoft* (2025)[5], esta colaboración ha llevado el análisis al siguiente nivel. Gracias a inteligencia artificial y visión artificial, se generan métricas en tiempo real que sirven tanto a técnicos como a aficionados. Una nueva forma de entender lo que sucede en el campo, basada en tecnología de vanguardia y desarrollos conjuntos entre LaLiga Tech y Microsoft.

Estos ejemplos no son simples anécdotas. Son la prueba de que los datos, cuando se usan con inteligencia, pueden marcar la diferencia. Este proyecto busca precisamente eso: acercar esas herramientas al mundo académico con una propuesta práctica, accesible y útil.

1.4. Motivación

La verdad es que este proyecto nace, sobre todo, de una motivación personal. De esa mezcla entre pasión por el fútbol y curiosidad por la tecnología. Porque no se trata solo de analizar por analizar, sino de entender mejor el juego que nos emociona cada fin de semana.

Hoy en día, cualquiera con acceso a Internet y ganas de aprender puede trabajar con datos reales. Plataformas como **StatsBomb** o **FBref** lo ponen al alcance de la mano, mientras que lenguajes como Python o librerías como **OpenCV** y **YOLO** hacen posible aplicar visión artificial sin grandes recursos.

Este trabajo no solo busca aprender y aplicar técnicas avanzadas, sino también crear una herramienta propia. Un espacio donde reunir toda esa información dispersa, visualizarla de forma clara y, por qué no, soñar con que pueda ser útil más allá del TFG.

1.5. Objetivos del trabajo

Detrás de este proyecto hay una idea que, desde el principio, ha latido con fuerza: aprovechar el poder del análisis de datos y la inteligencia artificial para entender el fútbol desde otro ángulo. Un ángulo más rico en matices, más justo con los detalles, más cercano a la realidad compleja que se vive sobre el césped. No se trata solo de generar visualizaciones llamativas o de entrenar modelos por curiosidad académica. La intención va mucho más allá: construir una prueba de concepto, un primer paso hacia una herramienta práctica, interactiva y accesible que pueda ayudar a cuerpos técnicos, analistas o incluso aficionados avanzados a interpretar el juego con datos. A descubrir patrones que no se ven a simple vista. A hacerse mejores preguntas. En la *Figura 1* podemos ver un esquema de la idea principal de la aplicación que nos ayudará a interpretar más fácilmente la organización de la aplicación y su funcionamiento.

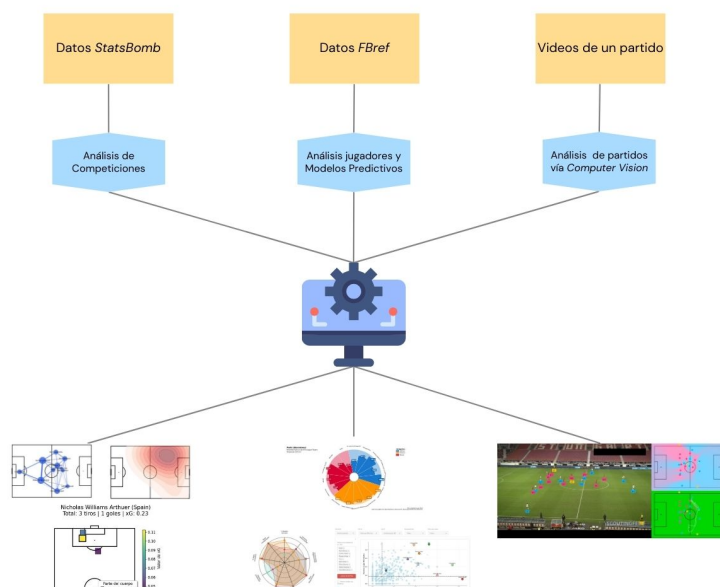


Figura 1.1: **Esquema de la aplicación.**

Fuente: Elaboración propia

Y con esa idea como guía, los principales objetivos de este trabajo han sido los siguientes:

1. **Diseñar un cuadro de mando interactivo con Shiny for Python:** Esta aplicación será, en cierto modo, el corazón del proyecto. A través de la librería **Shiny**, se buscará crear una interfaz intuitiva, visual y dinámica, que permita explorar información en tiempo real sin necesidad de conocimientos técnicos. El objetivo es que el usuario pueda filtrar jugadores, visualizar estadísticas clave, comparar perfiles o analizar equipos con apenas unos clics. La experiencia deberá ser fluida, cercana, casi como si uno pudiera “jugar” con los datos.
2. **Reunir y armonizar datos provenientes de fuentes abiertas distintas:** En el fútbol moderno los datos están por todas partes, pero rara vez vienen listos para ser utilizados de forma directa. En este proyecto se trabajará con dos fuentes principales: **StatsBomb**, centrada en eventos dentro del campo, y **FBref**, con estadísticas por temporada de miles de jugadores. Unir ambas requerirá limpieza, transformación y, sobre todo, encontrar puntos de conexión entre formatos muy distintos. El objetivo será construir una base sólida que permita comparar, cruzar y analizar sin trabas.

3. **Desarrollar modelos de predicción de goles y asistencias:** Uno de los retos más apasionantes será intentar anticipar lo impredecible: cuántos goles o asistencias puede aportar un jugador en la siguiente temporada. Para ello, se entrenarán modelos de aprendizaje automático que tomen como entrada los datos históricos y devuelvan estimaciones realistas. No se pretende sustituir la intuición humana, ni mucho menos, pero sí ofrecer un apoyo cuantitativo que pueda complementar la toma de decisiones en contextos de *scouting* o planificación.
4. **Incorporar visión por computadora para enriquecer el análisis en vídeo:** Porque el fútbol no son solo números, sino también movimiento, espacio, gestos y ritmo. Por eso se integrará un sistema básico de análisis visual mediante `OpenCV`, utilizando modelos preentrenados (como los disponibles en `Hugging Face` o `Roboflow`) basados en la arquitectura `YOLOv8`. Con ellos se buscará detectar jugadores en secuencias de vídeo, añadiendo una dimensión visual al análisis. Será un primer paso, pero ya permitirá vislumbrar las posibilidades de una observación más inteligente, casi automatizada.
5. **Crear una interfaz que permita analizar competiciones oficiales de forma interactiva:** Uno de los propósitos fundamentales será desarrollar una plataforma que facilite el análisis de partidos disputados en torneos reales, como ligas nacionales o competiciones internacionales. La interfaz deberá adaptarse con flexibilidad a distintos contextos competitivos, permitiendo explorar datos relevantes de cada encuentro, comparar equipos y visualizar métricas clave de forma accesible y visual. La herramienta, en este sentido, servirá como un punto de partida para estudios más profundos basados en datos reales de juego.
6. **Diseñar visualizaciones personalizadas según la posición del jugador:** No es lo mismo analizar a un portero que a un delantero. Cada posición tiene sus métricas clave, sus puntos fuertes y sus debilidades. Por eso se diseñarán gráficos como los *radar plots* o los *pizza charts*, adaptados cuidadosamente a cada rol en el campo. Estas visualizaciones no solo deberán ser atractivas visualmente, sino también reveladoras: permitirán ver de un vistazo qué hace especial a un jugador... y dónde podría mejorar.
7. **Crear un buscador por perfil de jugador:** A modo de “scouting digital”, se implementará un sistema que permita buscar jugadores según distintos filtros: edad, posición, rendimiento, métricas específicas, etc. La idea será sencilla, pero poderosa: encontrar el tipo de jugador que uno necesita sin tener que navegar entre miles de registros. Un lateral joven con buenos duelos defensivos, un mediocentro creativo con buena visión de pase, un delantero eficaz o más creador de juego. El buscador deberá convertir esos perfiles en algo tangible y fácil de encontrar.

Capítulo 2

Marco Teórico

El término *Big Data* puede sonar imponente a primera vista, casi como un concepto sacado de una novela de ciencia ficción. Pero en realidad, hace referencia a algo muy concreto y cotidiano: el manejo de conjuntos de datos tan enormes y complejos que las herramientas tradicionales ya no dan abasto para analizarlos de forma eficiente. Y es que, en los últimos años, la cantidad de datos que generamos ha explotado. Cada clic, cada publicación, cada sensor en un estadio o cada transacción deja una huella digital. Todo eso suma. Y mucho.

2.1. ¿Qué es Big Data?

Lo cierto es que no hay una única definición de Big Data. Cada institución lo aborda desde su propia mirada, dependiendo de qué aspecto quiera destacar. Una de las definiciones más reconocidas es la de la consultora tecnológica Gartner, que en 2020 lo describía de esta manera:

“Un gran volumen, velocidad o variedad de información que demanda formas costeables e innovadoras de procesamiento de información que permitan ideas extendidas, toma de decisiones y automatización del proceso” [6].

Por otro lado, Forbes (2021) opta por una visión más enfocada al mundo empresarial, con un tono casi entusiasta:

“Al análisis de información en cantidades industriales, todo un universo de oportunidades para las empresas aún por explorar” [7].

Y es que, más allá de las palabras elegidas, todas las definiciones coinciden en lo mismo: el Big Data no se trata solo de acumular datos por el gusto de tenerlos, sino de encontrarles un sentido. De convertir ese mar de información en conocimiento valioso.

Para entender mejor este concepto, suele hablarse de las famosas **5 V's**, que resumen sus principales características:

- **Volumen:** porque estamos hablando de cantidades inmensas de datos, que crecen a un ritmo abrumador.
- **Velocidad:** los datos se generan y deben procesarse casi en tiempo real, sin margen para perderse en el camino.
- **Variedad:** textos, imágenes, vídeos, audios, sensores, etc. Los datos vienen de todas partes y en todos los formatos imaginables.
- **Veracidad:** no todo lo que brilla es oro; hay que saber distinguir entre lo que es útil y lo que es ruido.
- **Valor:** y quizás esta sea la más importante: todo este esfuerzo solo tiene sentido si somos capaces de extraer algo que realmente nos sirva.

En resumen, el Big Data no es solo un reto técnico. Es una oportunidad enorme, compleja y fascinante de entender mejor el mundo que nos rodea. Ya sea para predecir tendencias, optimizar procesos o simplemente tomar decisiones más informadas, lo cierto es que los datos, bien utilizados, tienen mucho que decir.

2.1.1. Breve historia del Big Data

Hablar de Big Data hoy en día puede parecer algo moderno, casi futurista. Pero, si miramos un poco hacia atrás, descubrimos que sus raíces están mucho más lejos de lo que solemos imaginar. Ya en los años 60 y 70, con la llegada de los primeros ordenadores y los sistemas de almacenamiento digital, comenzó a gestarse esa idea, aún primitiva, de gestionar grandes volúmenes de datos de forma automatizada. Fue una época clave: se popularizaron las bases de datos relacionales, gracias a la propuesta de Edgar F. Codd en 1970 [8], lo que permitió organizar la información en tablas y consultarla de forma eficiente mediante lenguajes como *SQL*.

Después vinieron los años 80 y 90, un periodo marcado por el crecimiento imparable de internet y la informatización de las empresas. Y con ello, claro, una avalancha de datos. Empresas como *Amazon*, *Google* o *Walmart* no tardaron en darse cuenta del valor oculto en esos millones de registros que generaban a diario. Walmart, por ejemplo, ya utilizaba en los 90 sistemas avanzados para gestionar su inventario y analizar ventas en tiempo real [9]. Era una forma de afinar su cadena de suministro con una precisión que, hasta entonces, parecía ciencia ficción.

Pero el verdadero punto de inflexión llegó a mediados de los 2000. En 2005, se dio un paso gigante con la aparición de tecnologías diseñadas expresamente para procesar datos masivos de forma distribuida y escalable. Aquí es donde entra en escena *Hadoop*, una herramienta de código abierto basada en el modelo **MapReduce** que había sido ideado por Google [10]. ¿Qué lo hacía tan especial? Básicamente, permitía dividir los datos en partes más pequeñas y procesarlas en paralelo a través de múltiples nodos. Una revolución total si lo comparamos con las bases de datos clásicas.

A partir de ahí, todo fue acelerándose. Surgieron nuevas tecnologías que vinieron a complementar y potenciar aún más el ecosistema de Big Data:

- **NoSQL**, que trajo flexibilidad para trabajar con datos no estructurados, como documentos, grafos o pares clave-valor [11].
- **Spark**, que hizo posible procesar datos en memoria, reduciendo drásticamente los tiempos frente a *Hadoop*.
- **Sistemas de almacenamiento distribuido** como **HDFS**, **Amazon S3** o **Google BigQuery**, que facilitaron el acceso rápido y seguro a volúmenes de información descomunales.

Y así, poco a poco, lo que comenzó siendo un conjunto de herramientas para informáticos se convirtió en una palanca clave para múltiples industrias. Hoy el Big Data está en todas partes: desde la medicina personalizada y la logística inteligente, hasta el análisis político, el comercio electrónico o el deporte profesional. Empresas, gobiernos y organizaciones de todo tipo aplican modelos predictivos, minería de datos o algoritmos de aprendizaje automático sobre billones de datos. ¿Para qué? Para tomar mejores decisiones, anticiparse a lo que viene, evitar riesgos y descubrir oportunidades que, a simple vista, pasarían desapercibidas.

Eso sí, no todo ha sido un camino de rosas. Esta evolución ha traído consigo nuevos desafíos, muchos de ellos éticos y legales. La privacidad, la gobernanza de los algoritmos y el uso responsable de la inteligencia artificial son debates cada vez más presentes. Porque, aunque los datos nos abren puertas increíbles, también nos obligan a preguntarnos: ¿hasta dónde deberíamos llegar?

2.1.2. Big Data en la actualidad: usos y beneficios

Hoy en día, hablar de Big Data es casi inevitable cuando se trata de entender cómo funcionan las cosas a gran escala. Y es que, más allá de ser una palabra de moda, se ha convertido en una herramienta imprescindible para transformar el caos de millones de datos en decisiones certeras, en conocimiento útil. Lo que antes era un lujo reservado a grandes corporaciones, hoy es prácticamente una condición para mantenerse competitivo en un mundo hiperconectado y en constante cambio.

El impacto del Big Data se ha extendido como una red invisible que atraviesa todos los sectores. Desde la medicina hasta el deporte, pasando por la banca, la educación o el marketing, su presencia es cada vez más determinante. A continuación, repasamos algunos de los ámbitos donde su uso ha cambiado las reglas del juego:

- **Salud y biomedicina:** gracias al análisis masivo de datos ómicos, hoy es posible personalizar tratamientos médicos según el perfil genético del paciente, anticipar brotes epidemiológicos o incluso optimizar ensayos clínicos. También se ha vuelto común el seguimiento remoto de pacientes o la predicción de complicaciones antes de que ocurran [12].
- **Finanzas y banca:** en este sector, el Big Data no solo ayuda a detectar fraudes en tiempo real, sino que también permite evaluar riesgos crediticios con precisión quirúrgica. Los bancos ya no solo miran tu historial, también analizan tu comportamiento digital. Además, el análisis de sentimientos en redes sociales como X permite anticipar movimientos de mercado o detectar posibles crisis reputacionales [13].

- **Educación:** aquí, el cambio ha sido silencioso pero profundo. Plataformas basadas en inteligencia artificial, como Algor [14], ayudan a los estudiantes a transformar sus apuntes en esquemas visuales, resúmenes o incluso tarjetas de estudio de manera automática .
- **Comercio y marketing:** las grandes marcas ya no disparan a ciegas. Analizan cada clic, cada visita, cada segundo de atención. Gracias al Big Data, diseñan campañas publicitarias ajustadas al milímetro y ofrecen recomendaciones personalizadas que parecen adivinar lo que queremos incluso antes de que lo sepamos nosotros mismos [15].
- **Logística y transporte:** en este campo, los datos son como un mapa vivo que guía cada decisión. Las rutas se optimizan en tiempo real, se prevén atascos, se ajustan envíos según la demanda [16].
- **Medios de comunicación y redes sociales:** las plataformas analizan millones de interacciones por segundo para mostrarnos contenido relevante, detectar tendencias antes de que se viralicen y frenar la desinformación [17].
- **Deporte profesional:** sí, también aquí el Big Data es protagonista. Desde medir el esfuerzo físico de los jugadores hasta analizar movimientos tácticos o prevenir lesiones, los datos se han convertido en aliados estratégicos [18].

Todo esto es posible gracias a una serie de factores que han confluido casi como por arte de magia: la explosión de datos que generamos cada segundo (desde sensores, móviles, plataformas, etc.), el abaratamiento del almacenamiento, la potencia de la computación en la nube, y el avance brutal en algoritmos de análisis, como los de inteligencia artificial o aprendizaje automático.

2.1.3. Principales beneficios del uso de Big Data

Más allá de su versatilidad, lo que realmente hace especial al Big Data es cómo transforma la forma en que trabajamos, entendemos el mundo y tomamos decisiones. Aquí te resumo algunos de sus beneficios más destacados:

- **Decisiones más inteligentes, menos intuición:** con datos actualizados y objetivos en la mano, las decisiones estratégicas dejan de basarse en corazonadas. Se analizan patrones, se proyectan escenarios, se reducen los errores [19].
- **Eficiencia operativa al máximo:** al identificar cuellos de botella, errores repetidos o procesos ineficientes, se abren oportunidades para mejorar. Y no solo eso: muchas tareas rutinarias pueden automatizarse, liberando a las personas para que se enfoquen en lo que realmente importa [20].
- **Innovación que nace del dato:** a veces, entre millones de líneas de información, aparece una necesidad que nadie había visto antes. Es ahí donde surgen nuevas ideas, productos o servicios que marcan la diferencia y abren nuevos caminos [21].
- **Experiencias personalizadas, usuarios más felices:** ya no se trata solo de vender, sino de conectar. Y para eso, el Big Data permite diseñar experiencias únicas, adaptadas al gusto y al comportamiento de cada persona [19].
- **Detección precoz de riesgos:** los algoritmos de análisis pueden captar pequeñas anomalías que, de otro modo, pasarían desapercibidas. Esto es clave para prevenir fraudes, fallos técnicos, pérdidas millonarias o incluso crisis de reputación [20].

En definitiva, tener acceso a los datos ya no es suficiente. El verdadero desafío está en saber interpretarlos, extraerles valor y actuar con responsabilidad. Porque sí, el Big Data es poder, pero también implica una gran responsabilidad ética.

Conclusión

Podríamos decir que el Big Data ha cambiado el juego. No solo ha modernizado procesos o mejorado resultados, sino que ha transformado el modo en que las organizaciones y también las personas se relacionan con la información. Pasamos de tomar decisiones basadas en la intuición a estrategias construidas sobre certezas, números y patrones que hablan por sí solos. Su impacto es profundo, transversal y, lo más importante, todavía está en plena evolución. Y eso, sinceramente, emociona.

2.2. ¿Qué es el *Big Data* en el Deporte?

Se conoce el *Big Data* deportivo como el uso de grandes volúmenes de datos generados en el ámbito deportivo, procesados mediante herramientas tecnológicas avanzadas para extraer información valiosa. Estos datos provienen de múltiples fuentes, como sensores, dispositivos *wearables*, cámaras de alta velocidad, plataformas de *streaming* y redes sociales. Su análisis permite mejorar el rendimiento de los atletas, optimizar estrategias, prevenir lesiones y enriquecer la experiencia del aficionado [18].

Las aplicaciones del *Big Data* en el deporte abarcan desde la monitorización fisiológica (ritmo cardíaco, fatiga) hasta el análisis táctico (movimientos de jugadores, patrones de juego). Además, ayuda a impulsar decisiones empresariales en la gestión de clubes, como fichajes, planificación deportiva o campañas de marketing dirigido [22].

2.2.1. Orígenes

La relación entre deporte y tecnología no es algo nuevo, aunque a veces lo olvidemos. A lo largo de la historia, el ser humano ha buscado maneras de superarse, de ir un poco más allá, y en ese camino la tecnología ha sido una aliada poderosa. El deporte, como reflejo de nuestra pasión por el esfuerzo y la competencia, no podía quedarse al margen. Ya en el siglo XIX se empezaron a ver los primeros coqueteos entre ambos mundos. Por ejemplo, con la incorporación de cronómetros más precisos en disciplinas como la natación o el atletismo, se logró registrar los tiempos con mayor exactitud. Era el primer paso de muchos [23].

Más adelante, la aparición de materiales avanzados y equipamientos diseñados específicamente para mejorar el rendimiento marcó un antes y un después. Y es que, cuando un detalle puede cambiar el resultado de una competición, cualquier innovación cuenta [24].

Pero si hablamos de una revolución profunda, la verdadera transformación llegó con la digitalización. El *Big Data* ha irrumpido con fuerza en los últimos años, pero sus raíces se extienden mucho más atrás. Antes de los algoritmos y las plataformas inteligentes, hubo personas clave que se atrevieron a mirar los datos con otros ojos, intuyendo lo que estaba por venir. En el siglo XIX ya existía una «cultura de datos» en los deportes modernos, con resultados, clasificaciones y objetividad cuantitativa que sustentaron la futura era analítica [25].

Charles Reep

Como se comenta en un artículo de *ABC* (2020) [26], uno de esos pioneros fue Charles Reep (1904–2002), véase la *Figura 2.1*, quien, en los años 50, decidió que el fútbol no solo se jugaba con los pies, sino también con la cabeza y con un bolígrafo. Durante los partidos, Reep anotaba a mano los eventos que él consideraba importantes, buscando patrones y regularidades. Fue un visionario que supo ver el valor de los datos mucho antes que la gente de su época.

De sus apuntes nació la famosa *teoría de los tres pases*, una idea provocadora: según sus datos, la mayoría de los goles llegaban tras tres o menos pases desde el área propia. Esto llevó a Reep a defender un estilo de juego directo, veloz. Su enfoque, práctico y sin rodeos, rompía con la romántica creencia de que la posesión era sinónimo de éxito.

Durante años, este estilo marcó el ADN del fútbol inglés. Equipos enteros se moldearon bajo esa lógica vertical y física. No fue hasta la llegada de entrenadores como Pep Guardiola al *Manchester City*, en 2016, que se empezó a cuestionar ese paradigma con fuerza.

La influencia de Reep fue tan profunda que, aún hoy, se le reconoce como uno de los padres del análisis de datos en el fútbol. Lo suyo no eran estadísticas sofisticadas, pero sí una nueva manera de mirar el juego.



Figura 2.1: **Charles Reep**, pionero del análisis de datos en el fútbol.
Fuente: ABC

Billy Beane

Si hablamos de revolución, es imposible no mencionar a Billy Beane. Su historia se volvió mundialmente conocida gracias a la película *Moneyball* (2011), como se muestra en la *Figura 2.2*, protagonizada por Brad Pitt pero lo realmente impactante es cómo cambió para siempre el béisbol y de paso, la manera en que entendemos el deporte profesional.

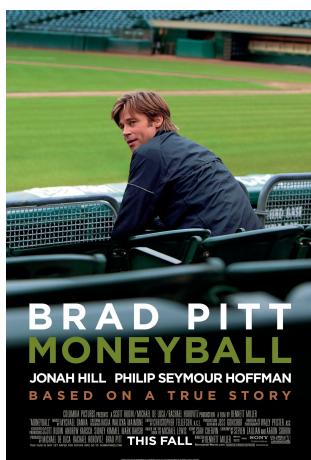


Figura 2.2: Película **Moneyball** (2011).
Basada en la historia de Billy Beane, la película muestra cómo el análisis estadístico cambió la forma de tomar decisiones en el béisbol profesional.
Fuente: IMDb

Como se menciona en un artículo de *ginvest* [27], Billy Beane (*Figura 2.3*), gerente general de los Oakland Athletics, se enfrentaba a un problema clásico: presupuesto limitado. Mientras otros equipos nadaban en millones, él tenía que hacer malabares. La

solución, inesperada y radical, fue apoyarse en el análisis estadístico avanzado para fichar jugadores infravalorados. Junto a su asistente Paul DePodesta, economista de formación, desarrollaron un modelo que se centraba en métricas concretas, más allá de la intuición o la “buena pinta” de un jugador.

El resultado fue espectacular. Con una plantilla humilde, los Athletics encadenaron 20 victorias consecutivas, batiendo un récord histórico. Pero lo más importante fue el mensaje: los números, bien utilizados, pueden contar verdades que los ojos no ven.



Figura 2.3: **Billy Beane.**

Gerente general de los Oakland Athletics, pionero en el uso de análisis de datos para la toma de decisiones en el béisbol.

Fuente: Wikipedia

Muthu Alagappan

Y del béisbol, damos el salto al baloncesto. Un artículo publicado por *Marca* (2023) [28] explica cómo Muthu Alagappan (*Figura 2.4*), un joven estudiante de Ingeniería Biomédica en Stanford, demostró que no hace falta ser un veterano para revolucionar un deporte. Usando análisis topológico de datos, una técnica matemática bastante compleja, Alagappan descubrió algo asombroso: la clásica división de jugadores en cinco posiciones no reflejaba la riqueza real del juego.

Según sus modelos, había al menos 13 estilos de jugadores distintos, cada uno con un perfil estadístico bien definido. Esta nueva visión, basada puramente en datos, ayudó a repensar cómo construir equipos y cómo valorar el impacto real de cada jugador.

Su trabajo llamó la atención de franquicias como los *Portland Trail Blazers* y los *Miami Heat*, que rápidamente lo sumaron a sus equipos de análisis. Su historia, a veces apodada “Muthuball”, es un ejemplo de cómo la creatividad, el conocimiento y la tecnología pueden cambiar el juego desde dentro.



Figura 2.4: **Muthu Alagappan.**

Pionero en el uso de análisis topológico de datos para redefinir las posiciones en el baloncesto. Su propuesta de 13 roles transformó la planificación táctica en la NBA.

Fuente: Medium

2.2.2. Evolución tecnológica en el deporte

Con el paso de las décadas, el cambio se aceleró. En los años 60, el análisis biomecánico comenzó a hacerse un hueco gracias a las cámaras de vídeo, que permitían grabar y estudiar los movimientos de los atletas con una precisión nunca vista. Esto permitió mejorar técnicas, ajustar entrenamientos y, en definitiva, rendir más [29].

En los 90, el videoanálisis se volvió indispensable en deportes colectivos. Los partidos comenzaron a dividirse en secuencias, jugadas clave, esquemas tácticos. La idea era clara: entender para mejorar [30].

A la vez, llegaron los primeros dispositivos de monitorización, como los pulsómetros. Y más adelante, avances en materiales como el *grafeno* o el *poliuretano* cambiaron la forma de fabricar raquetas, zapatillas o bañadores. Eran detalles, sí, pero decisivos [31].

Y entonces llegó el siglo XXI. Con él, una auténtica revolución. Sensores *GPS*, dispositivos *wearables*, análisis en tiempo real, plataformas de datos masivos, etc. Hoy en día se puede saber exactamente cuánto ha corrido un jugador, a qué velocidad, con qué esfuerzo, e incluso cuán cerca está de lesionarse [32].

La inteligencia artificial, por su parte, empieza a predecir lesiones, rendimientos e incluso resultados. Y si miramos hacia el futuro, tecnologías como la realidad aumentada o la realidad virtual prometen transformar los entrenamientos, simulando escenarios sin necesidad de pisar el campo [33].

2.2.3. El papel del *Big Data* en el deporte

Durante mucho tiempo, el deporte se apoyó casi exclusivamente en la intuición, la experiencia y la pasión. Sin embargo, en los últimos años, ha emergido una herramienta que está revolucionando todo: el *Big Data*. Su impacto va mucho más allá de lo técnico. Es una nueva forma de mirar el juego, de entenderlo, y sobre todo, de tomar decisiones que antes parecían inalcanzables.

Gracias al análisis masivo de datos, hoy es posible convertir montones de cifras aparentemente caóticas en información valiosa. Ya sea para mejorar el rendimiento de un jugador, ajustar una estrategia de equipo o prevenir una lesión antes de que ocurra, los datos se han convertido en aliados inseparables del deporte moderno.

Importancia de la estadística en el deporte

La estadística deportiva ha dejado de ser una curiosidad de nicho para convertirse en una brújula esencial. Nos permite traducir el caos del juego en patrones, tendencias y respuestas concretas. Con ella, es posible dejar atrás la mera intuición y ofrecer análisis objetivos basados en datos reales. En definitiva, se trata de observar el juego con una lupa mucho más precisa.

Aplicaciones actuales del *Big Data* en deportes

Si miramos alrededor, vemos cómo distintas disciplinas están abrazando el *Big Data* con entusiasmo. Algunas lo hacen por necesidad, otras por pura innovación, pero todas con el mismo objetivo: mejorar. A continuación, exploramos varios casos concretos que ilustran cómo esta revolución digital está dejando huella.

Fútbol

El fútbol, aunque a veces más conservador que otros deportes, no ha querido quedarse atrás. Poco a poco, ha ido integrando herramientas de análisis de datos e inteligencia artificial para tomar decisiones más fundamentadas.

- **VAR (Video Assistant Referee):** Ya es parte habitual de cualquier partido. Gracias a cámaras de alta precisión y repeticiones detalladas, el VAR ha cambiado por completo el arbitraje, ayudando a que las decisiones sean más justas y menos cuestionadas.
- **TacticAI:** Desarrollado por Google DeepMind en colaboración con el Liverpool F.C., este sistema utiliza aprendizaje automático para sugerir jugadas y ajustes tácticos. ¿Lo más interesante? Aprende constantemente gracias a la interacción humana y a millones de datos de partidos reales. Se explica más a fondo en el artículo publicado por *infobae* (2024) [34].
- **Scout Advisor (IBM y Sevilla F.C.):** Esta plataforma cruza datos biométricos, informes de ojeadores y modelos estadísticos complejos para identificar jóvenes promesas o refuerzos que encajen con la filosofía del club, como se explica en un artículo publicado por el propio *Sevilla F.C.* (2024) [35].

Baloncesto

En la NBA, el *Big Data* no es una opción, es parte del ADN de la competición. Cada jugada, cada pase, cada movimiento se mide y se analiza con un nivel de detalle asombroso.

- **NBA Advanced Stats:** Métricas como el *Player Efficiency Rating* (PER) permiten evaluar a los jugadores de forma global, ponderando cada segundo que pasan en pista. Es una forma más justa y completa de entender su impacto real.
- **Sistemas de tracking:** Gracias a tecnologías como *Second Spectrum*, se puede seguir en tiempo real el movimiento de cada jugador y del balón. Esto permite reconstruir jugadas, evaluar coberturas defensivas y diseñar jugadas más efectivas.
- **HomeCourt:** Esta app de entrenamiento basada en IA analiza en tiempo real el tiro, como se aprecia en la *Figura 2.5*, los desplazamientos y hasta la postura corporal. Es como tener un entrenador personal en el bolsillo. Se explica más a fondo a cerca de ella en una noticia publicada por *Viva Basquet* (2020) [36].

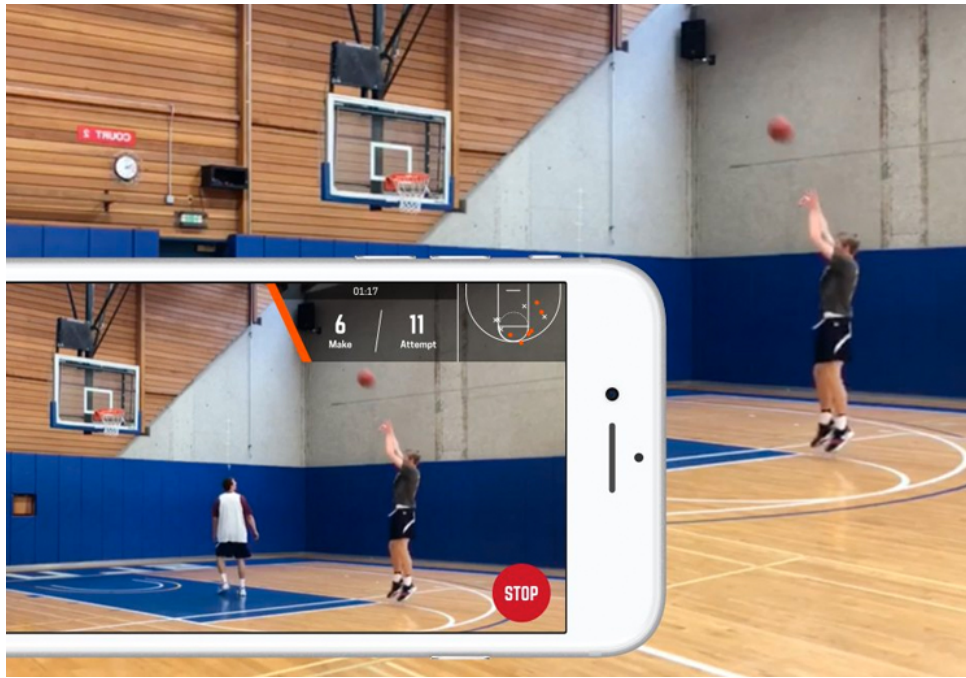


Figura 2.5: HomeCourt App

Fuente: Viva Basquet

Rugby

El rugby, deporte de equipo por excelencia, también ha visto en el *Big Data* una herramienta perfecta para afinar tanto en lo individual como en lo colectivo.

- **Federación Francesa de Rugby (FFR):** En colaboración con SAS, desarrollaron una inteligencia artificial capaz de analizar patrones colectivos basados en posiciones y acciones, no solo en estadísticas individuales como bien se explica en un artículo de la propia SAS [37].
- **All Blacks (Nueva Zelanda):** Uno de los equipos más exitosos del mundo también utiliza *SAS Visual Analytics* para cruzar datos físicos, técnicos y tácticos. Así, toman decisiones más acertadas sobre qué jugadores convocar y cómo prepararlos, también se comenta esta colaboración en otro artículo de SAS (2019) [38].

Tenis

- **Hawk-Eye (Ojo de halcón):** Este sistema, basado en cámaras de alta velocidad, determina con exactitud si una bola entró o no, la Figura 2.6 muestra un esquema de su funcionamiento. Es un ejemplo perfecto de cómo la tecnología ha traído transparencia y justicia al deporte.

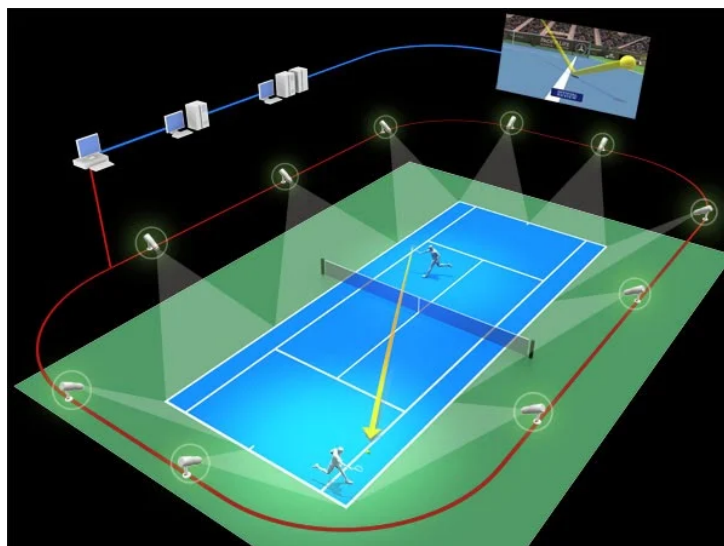


Figura 2.6: **Pequeño esquema** de como funciona el Ojo de halcón.

Fuente: Ertheo

Fórmula 1

Si hay un deporte donde el dato es casi una religión, ese es la Fórmula 1. Aquí, cada decisión se basa en simulaciones, predicciones y análisis milimétricos.

- **Telemetría en tiempo real:** Durante una carrera, cada coche genera más de 300 GB de datos como por ejemplo: velocidad, frenado, temperatura de los neumáticos, consumo de combustible, etc. Todo se analiza al instante para ajustar la estrategia, como se muestra en la *Figura 2.7*, se comenta más a fondo en un artículo publicado por MAPFRE (2020) [39] acerca del *Big Data* y la F1.

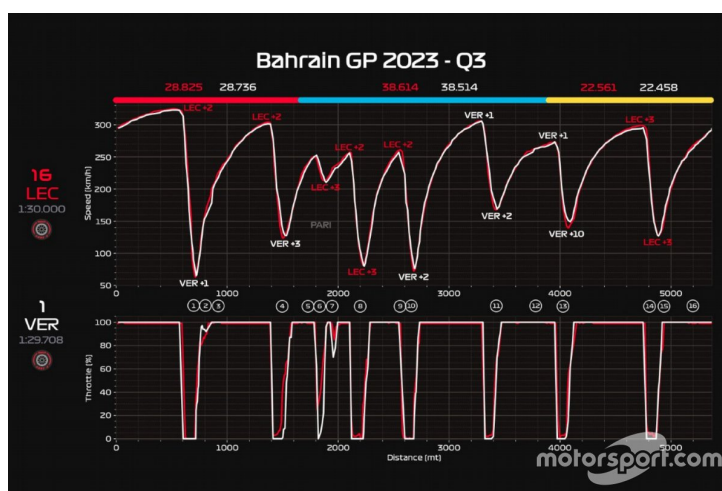


Figura 2.7: **Comparación de Telemetría** entre Max Verstappen y Charles Leclerc durante la Q3 del Gran Premio de Bahrein 2023

Fuente: Motorsport.com

- **Simulaciones predictivas:** Equipos como Mercedes utilizan gemelos digitales para probar modificaciones en el coche. Así, ahorran tiempo y recursos, y ganan ventaja sin necesidad de tocar la pista.

2.2.4. Beneficios y desafíos del *Big Data* en el deporte

El impacto del *Big Data* en el deporte ha sido claramente transformador. Decisiones que antes se basaban en intuiciones o en la experiencia subjetiva, hoy pueden apoyarse en datos objetivos, precisos y casi en tiempo real. Sin embargo, como en toda revolución tecnológica, junto con las oportunidades también emergen desafíos. Veamos ambos lados de esta moneda.

Beneficios que aporta

El uso del *Big Data* ha transformado no solo la forma de competir, sino también la preparación física, la planificación estratégica, la prevención de lesiones, la gestión organizativa y el marketing en una amplia variedad de disciplinas deportivas [22].

- **Optimización del rendimiento:** La monitorización continua de variables fisiológicas, biomecánicas y de carga de trabajo permite adaptar los entrenamientos a las necesidades individuales de cada deportista, mejorando su rendimiento y reduciendo riesgos asociados al sobreentrenamiento.
- **Ventaja táctica y estratégica:** El análisis de datos permite identificar patrones de comportamiento, rendimientos y resultados en distintas situaciones competitivas, facilitando la toma de decisiones más informadas para ajustar tácticas o estrategias en diferentes deportes.
- **Detección y desarrollo de talento:** La evaluación objetiva de habilidades y progresión de atletas desde etapas tempranas ayuda a planificar su desarrollo y potenciar su potencial en cualquier disciplina deportiva.
- **Prevención de lesiones:** El análisis predictivo basado en datos biométricos, cargas de entrenamiento y otros indicadores permite anticipar riesgos y diseñar programas de prevención, prolongando la carrera deportiva y mejorando la salud de los atletas.
- **Mejora de la experiencia del espectador:** A través del análisis en tiempo real y la personalización de contenidos digitales, el Big Data contribuye a enriquecer la interacción y el seguimiento de eventos deportivos por parte del público.
- **Estrategias de marketing y fidelización:** El conocimiento profundo de las audiencias y su comportamiento permite diseñar campañas más efectivas, segmentar públicos y mejorar la relación entre organizaciones deportivas, patrocinadores y aficionados.

Desafíos que no podemos ignorar

El uso del *Big Data* en el deporte plantea también retos técnicos, éticos y organizativos que deben ser abordados para maximizar su impacto positivo [22].

- **Gestión e interpretación de datos:** La cantidad masiva de información requiere de profesionales capacitados para extraer insights relevantes y aplicables en la práctica deportiva.

- **Privacidad y protección de datos:** La gestión responsable de datos personales y sensibles es fundamental para garantizar la confianza y el respeto hacia los deportistas y usuarios.
- **Resistencia al cambio:** La integración de tecnologías avanzadas puede encontrar barreras culturales y organizativas que dificultan su adopción y uso efectivo.
- **Desigualdad de recursos:** El acceso a tecnologías y capacidades analíticas puede ser limitado en ciertos entornos o disciplinas, generando brechas competitivas.
- **Dependencia excesiva de los datos:** Aunque valiosos, los datos deben complementarse con el conocimiento humano y la experiencia para tomar decisiones equilibradas y contextualizadas.

Conclusión

En resumen, el *Big Data* se ha consolidado como una herramienta esencial en el deporte moderno, facilitando mejoras en el rendimiento, la gestión, la experiencia del público y las estrategias comerciales. Su aplicación responsable y equilibrada es clave para aprovechar todo su potencial sin perder de vista los aspectos humanos que definen al deporte.

2.3. *Big Data* en el fútbol

En la última década, el fútbol ha experimentado una transformación notable gracias al uso intensivo de los datos. Estos han dejado de ser una herramienta auxiliar para convertirse en uno de los pilares fundamentales sobre los que se apoyan el rendimiento, la estrategia y la gestión en el fútbol moderno. Este cambio ha dado lugar a la aparición de nuevos perfiles profesionales especializados en análisis de datos, que hoy en día forman parte del cuerpo técnico en numerosos clubes de élite.

No obstante, la integración de estos perfiles varía significativamente según la filosofía de cada club. Mientras algunas entidades han apostado decididamente por una estructura sólida de analistas y científicos de datos, otras siguen relegando este rol a un segundo plano, como muestra la *Figura 2.8*, como se observa en las diferencias entre el FC Barcelona (con su Innovation Hub) y el Real Madrid [40].

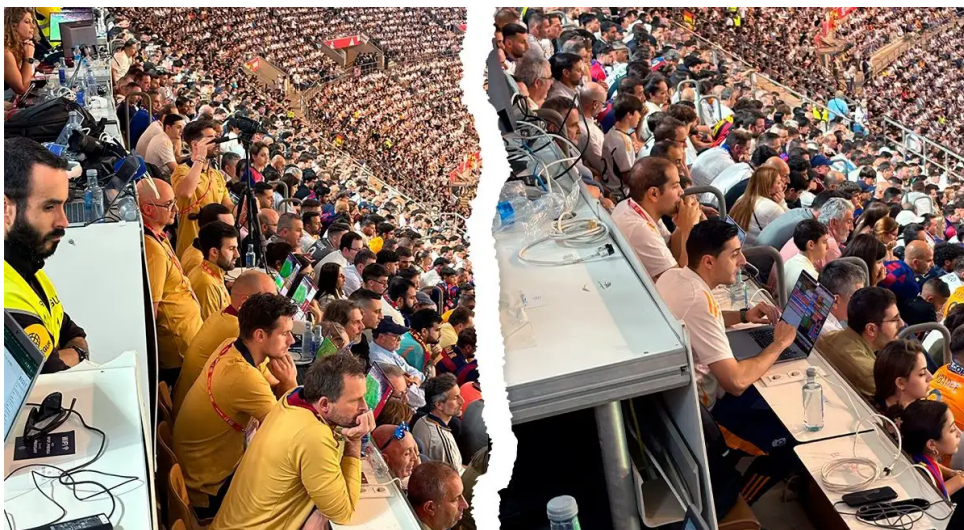


Figura 2.8: **Diferencias** entre la cantidad de analistas del Barcelona (izquierda) y Real Madrid (derecha) en el clásico de La Cartuja.

Fuente: Relevo

2.3.1. Usos del *Big Data* en el fútbol

Cuando pensamos en fútbol, solemos imaginar goles, celebraciones, estadios llenos. Pero detrás de todo eso hay algo menos visible, aunque igual de emocionante, que está cambiando la forma en que entendemos este deporte: el *Big Data*. Su aplicación es cada vez más extensa, y la verdad es que ya no se trata solo de números, sino de decisiones más inteligentes, estrategias mejor afinadas y hasta formas nuevas de disfrutar los partidos.

- **Análisis de rendimiento de jugadores:** Hoy en día, medir el impacto de un futbolista va mucho más allá de contar goles y asistencias. Métricas como los goles esperados (xG), las progresiones con balón, las presiones exitosas o incluso los movimientos sin balón nos ofrecen una visión más profunda y completa. Es como levantar el telón y descubrir todo lo que hay detrás de cada jugada. Nos ayuda a entender por qué un jugador puede ser decisivo aunque no aparezca en las portadas [41].
- **Táctica y estrategia:** Los cuerpos técnicos disponen de herramientas que, hace no mucho, parecían sacadas de la ciencia ficción. Mapas de calor que revelan zonas de influencia, modelos que predicen cómo se moverá un equipo, redes de pases que muestran quién organiza el juego desde la sombra. Con toda esta información, los entrenadores pueden ajustar su plan al detalle, ya sea para corregir errores o para aprovechar debilidades del rival en tiempo real [42].
- **Scouting y fichajes:** Aquí es donde el *Big Data* brilla como un faro en la niebla. Plataformas como *SciSports*, *TransferLab* o *Smarterscout* permiten detectar talento incluso en ligas menores o categorías juveniles. Gracias al análisis de miles de datos de diferentes competiciones, es posible anticipar el rendimiento futuro de un jugador. Es un cambio radical: lo que antes requería años de observación, ahora puede respaldarse con modelos predictivos sólidos. Eso sí, los datos nunca sustituyen el ojo experto ni la filosofía de un club, pero sí la complementan de forma muy valiosa [43].
- **Prevención de lesiones:** Los datos también cuidan del cuerpo. El análisis de la carga física, mediante sensores GPS, acelerómetros o incluso cuestionarios de percepción, permite detectar signos tempranos de fatiga o sobreentrenamiento. Esto da margen para actuar antes de que aparezca una lesión. En un calendario tan apretado como el del fútbol profesional, esta prevención puede marcar la diferencia entre competir o ver los partidos desde la grada [42].
- **Experiencia del aficionado:** Y, por supuesto, los datos también hacen que vivir el fútbol desde casa o desde la grada sea una experiencia mucho más rica. Aplicaciones interactivas, gráficos en tiempo real, estadísticas visuales o incluso realidad aumentada permiten al espectador entender el juego con otros ojos. Ya no solo se siente el partido: también se analiza, se interpreta, se comparte. Y eso crea una conexión mucho más profunda con el deporte [41].
- **Marketing deportivo:** Por último, pero no menos importante, está el mundo del marketing. Los clubes utilizan modelos predictivos para saber qué contenidos generan más interacción en redes, cuándo lanzar campañas o cómo segmentar mejor a su audiencia. En definitiva, usan los datos para emocionar mejor. Porque al final, todo esto también va de contar historias. Y los datos, bien utilizados, pueden contar algunas realmente poderosas [44].

2.3.2. Tipos de datos utilizados en el fútbol

La explosión del *Big Data* y el almacenamiento en la nube ha transformado por completo el modo en que entendemos el rendimiento en el fútbol. Hoy en día, tenemos acceso a cantidades impresionantes de información, precisa y muy detallada, que nos permite observar lo que antes pasaba desapercibido. Gracias a ello, podemos analizar el juego desde ángulos completamente nuevos y más enriquecedores.

Una de las áreas donde más se nota este cambio es en la evaluación del rendimiento de los jugadores. Lo interesante es que no se trata solo de contar goles o asistencias. Los datos que podemos recolectar abarcan desde lo físico hasta lo cognitivo, pasando por aspectos tácticos, emocionales o incluso personales [45].

Datos físicos

Estos datos nos hablan del cuerpo en acción. Reflejan el esfuerzo, la resistencia, la velocidad, etc. En definitiva, el motor del futbolista. Suelen obtenerse mediante GPS, sensores de inercia o cámaras especializadas, y ofrecen una radiografía bastante precisa del desgaste físico durante un partido o entrenamiento.

- Distancia total recorrida (en km)
- Número de sprints o aceleraciones por encima de los 20 km/h
- Distribución por zonas de velocidad (caminar, trotar, correr, sprint)
- Carga total de entrenamiento

Estos datos permiten a los preparadores físicos ajustar las sesiones con mucho más criterio. Así se logra ese equilibrio tan delicado entre mejorar el rendimiento y evitar lesiones, como se puede apreciar en la *Figura 2.9*. Porque sí, hay una delgada línea entre estar en forma y sobrecargarse.

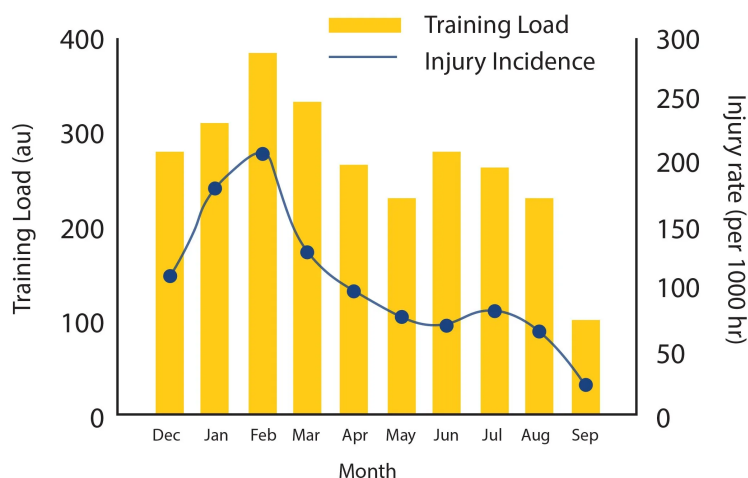


Figura 2.9: **Relación entre cargas de entrenamiento, forma física, riesgo de lesión y rendimiento.** Modelo propuesto por Orchard (2012).

Fuente: BARÇA INNOVATION HUB

Datos técnico-tácticos

Aquí entramos en el terreno de las decisiones con balón (y sin él). Son las estadísticas que recogen lo que un jugador hace sobre el césped: los pases que intenta, los que acierta, si regatea, recupera, intercepta, remata, asiste, etc. En otras palabras, su huella táctica durante el juego.

- Pases intentados y completados
- Regates exitosos
- Intercepciones o recuperaciones
- Goles, asistencias y otras acciones clave

Estos datos ayudan muchísimo en el análisis postpartido, pero también durante la semana para ajustar tácticas o trabajar puntos débiles. A veces, un simple porcentaje de acierto en pases puede explicar por qué un equipo se bloqueó en la salida de balón.

Datos espaciales

Si el fútbol es movimiento constante, estos datos son su mapa. Nos dicen exactamente dónde está cada jugador (y el balón) en cada instante del partido. Se obtienen mediante sistemas de tracking óptico o dispositivos GPS de alta frecuencia, y su potencial para entender el juego colectivo es enorme.

- Coordenadas x e y por cada frame del partido
- Distancias entre líneas o entre jugadores
- Áreas de influencia, presión o espacios ocupados

Gracias a estos datos, los analistas pueden ver cómo bascula una defensa, cuándo se genera una superioridad numérica o cómo se desactiva una presión rival. También se suelen utilizar para conocer las zonas de despliegue o de mayor presencia de un jugador en concreto, como se muestra en la *Figura 2.10*.

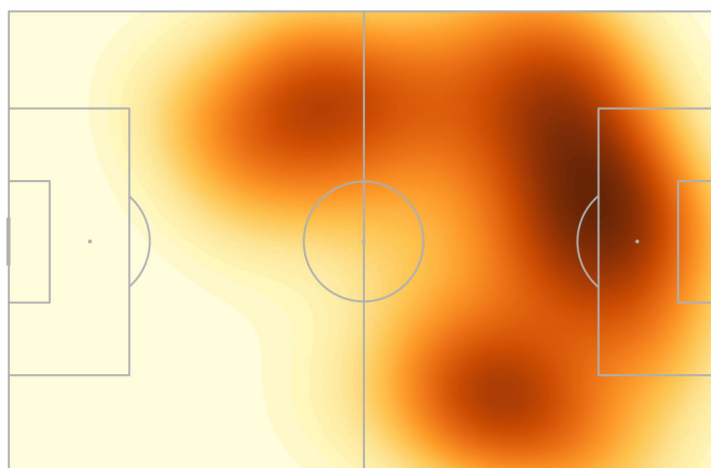


Figura 2.10: **Mapa de calor** de un jugador.

Fuente: *FÚTBOL CON DATOS*

Datos subjetivos

No todo se puede medir con sensores. Hay cosas que se perciben, que se intuyen, y que sólo el ojo humano es capaz de captar. Aquí entran los datos subjetivos, normalmente aportados por ojeadores o entrenadores, que valoran aspectos como el carácter, el liderazgo o la toma de decisiones bajo presión.

- Informes individuales o colectivos de observación
- Escalas de valoración (del 1 al 5, 10 o 100) en atributos técnicos o mentales
- Comentarios sobre actitud, madurez, estilo de juego o adaptabilidad

Estos informes suelen ser clave en procesos de scouting, ya que permiten hacerse una idea más completa del perfil del jugador más allá de lo numérico, la *Figura 2.11* muestra un ejemplo completo de este tipo de datos.

APellidos	Nombre	Fecha nacimiento	Categoría	Fotografía		
HERNÁNDEZ	XAVI	25-1-80	INFANTIL "A"			
ERRORES DE CARRERA		EQUILIBRIO	COORDINACIÓN	REGULARIDAD		
- Corre de talones (sus movimientos son lentos y le cuesta arrancar)		REGULAR ↑	BUENA - Sus movimientos son correctos pero faltan de rapidez	REGULAR - No falta mucho trabajo en este sentido		
HABILIDAD						
BUENA - Sin ser su fuerte, tiene recursos en su juego para sacar partido de cualquier situación						
CONTEXTO ADAPTACIÓN	RESISTENCIA MUSCULAR	VELOCIDAD	ABRANCADA	FUERZA RÁPIDA		
BUENA Progres de todo buen centrocampista	BUENA (200mts)	REGULAR ↓ - Debe mejorar su agilidad y minutos de juego.	REGULAR	REGULAR (defensa horizontal)		
CONTROLES						
EXCELENTES - Supe este campo de control en sus movimientos con una excelente control de balón						
PASES	CENTROS	TIRO	CONDUCCIÓN	REGATES	ENSUADA EN EL CAMPO	RECONSTRUCCIÓN
MUY BUENOS - Se ve muy seguro en su juego, se puede decir que es un jugador de este tipo. - Muy buena técnica.	BUENOS - Con buena se de elegancia, a esta altura es muy buena.	ACEPTABLE - Debe mejorar para sus movimientos al ataque, al igual que su defensa.	MUY BUENA - Buen sentido de posición del balón. - Con mucha lo puede.	BUENOS - Sin ser su recurso habitual es correcto su nivel técnico en el regate.	EXCELENTE - Es su punto su mejor cualidad. Siempre está donde debe estar y siempre es un apoyo y una salida válida para sus compañeros.	REGULAR - Su juego constructivo es muy bueno, le cuesta la dificultad a veces este tipo de juego.

Figura 2.11: **Informe scout** completo de Xavi a los 14 años.
Fuente: MARCA

Datos cognitivos

Aquí nos adentramos en la mente del jugador. Hablamos de cómo percibe, decide, se concentra o gestiona la presión. Aunque suene menos tangible, cada vez hay más herramientas para medir estos aspectos: desde cuestionarios hasta tests de reacción o incluso análisis de lenguaje corporal.

- Niveles de concentración, ansiedad o fatiga mental
- Tests de percepción visual o tiempo de reacción
- Evaluaciones emocionales o motivacionales

Uno de los instrumentos más extendidos es el *Cuestionario Wellness*, que ofrece pistas sobre cómo se siente un jugador antes de entrenar o competir, en la *Figura 2.12* podemos encontrar un ejemplo de este tipo de cuestionario. Porque no solo es importante la salud del cuerpo, sino también cómo se encuentran los jugadores mental o anímicamente.

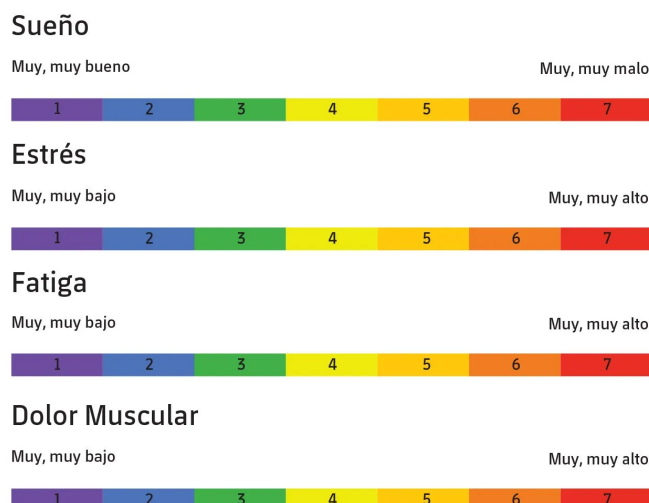


Figura 2.12: **Cuestionario wellness** aplicado en el FC Barcelona.
Fuente: BARÇA INNOVATION HUB

Datos personales

Por último, aunque no menos importantes, están los datos que rodean al jugador como persona: su contexto. Porque al final, un futbolista no es solo su número de goles, sino también su historia, sus motivaciones y su entorno.

- Edad, altura, historial médico, posición habitual
- Situación contractual, valor de mercado
- Idiomas que habla, entorno familiar, adaptación cultural

Páginas como *Transfermarkt* resumen este tipo de datos, como se muestra en la *Figura 2.13*.

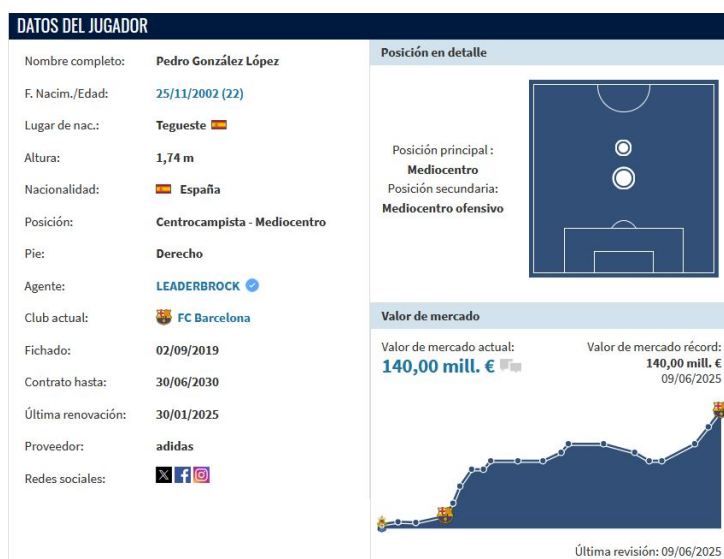


Figura 2.13: **Ficha resumen** del jugador del Barcelona Pedri.
Fuente: Transfermarkt

2.4. Principales proveedores de datos en el fútbol

En los últimos años, el *Big Data* ha transformado silenciosamente la forma en que entendemos el fútbol. Detrás de esta revolución están empresas especializadas en capturar hasta el más mínimo detalle de lo que ocurre en el campo. Desde cómo presiona un equipo hasta la cantidad de sprints de un lateral, estos proveedores son el motor invisible que alimenta el análisis moderno. A continuación, repasamos los actores más destacados en este ecosistema de datos que ya forma parte del día a día de clubes, analistas y aficionados.

Opta Sports (Stats Perform)

Hablar de datos en el fútbol sin mencionar a Opta sería quedarse a medias. Esta empresa, ahora integrada en *Stats Perform* [46], es una de las más veteranas y reconocidas del sector. Su especialidad es ofrecer estadísticas de eventos en tiempo real y con gran nivel de detalle: pases, tiros, duelos, posesión, etc., todo perfectamente etiquetado.

Y no son solo los clubes quienes confían en sus datos. Medios como la BBC o Sky Sports los utilizan para enriquecer retransmisiones, debates y análisis en directo. En la Premier League, por ejemplo, sus números ya forman parte del lenguaje habitual.

StatsBomb

StatsBomb [47] ha ganado terreno en poco tiempo gracias a un enfoque que va más allá del dato clásico. Sus estadísticas incluyen aspectos que antes pasaban desapercibidos, como la presión ejercida por el rival o el ángulo exacto de un pase.

Además, han hecho algo muy valioso: liberar parte de su base de datos para el uso académico y de desarrolladores. El dataset de la Eurocopa 2024, utilizado en este TFG, es un ejemplo de esa apuesta por abrir la puerta a la investigación.

Wyscout (Hudl)

Wyscout [48] combina lo mejor de dos mundos: estadísticas avanzadas y una gigantesca base de datos de vídeo. Su plataforma permite filtrar acciones muy concretas, como entradas exitosas o asistencias y observarlas directamente en vídeo. Esta funcionalidad lo convierte en una herramienta esencial para los departamentos de *scouting*.

Clubos de Serie A y LaLiga confían en Wyscout no solo para fichar, sino también para preparar partidos y analizar rivales.

FBref (Sports Reference)

FBref [49], desarrollado por Sports Reference, ha conseguido algo que no es fácil en este sector: poner al alcance de todo el mundo una base de datos detallada, gratuita y actualizada de las principales ligas del mundo. Desde estadísticas básicas como goles o asistencias, hasta métricas más avanzadas como acciones ofensivas creadas, pases progresivos o xGChain, su cobertura es impresionante.

En este TFG se ha utilizado como fuente principal tanto para recopilar estadísticas de jugadores de las cinco grandes ligas europeas como para alimentar los modelos de predicción. Más adelante se explicará con detalle cómo se han trabajado esos datos.

InStat

InStat [50] ofrece un enfoque más táctico y personalizable. Sus estadísticas se sincronizan con clips de vídeo, lo que permite revisar jugadas clave con total precisión. Entrenadores, agencias de representación y cuerpos técnicos lo usan para generar informes detallados y destacar patrones específicos.

Por ejemplo, se puede seguir el rendimiento defensivo de un central a lo largo de toda la temporada, acción por acción.

Tracab (ChyronHego)

Tracab [51] es sinónimo de *tracking* de alta precisión. Gracias a un sistema multicámara instalado en los estadios, este proveedor puede registrar la posición de todos los jugadores y del balón hasta 25 veces por segundo. Eso se traduce en datos espaciales muy valiosos: velocidad, distancia recorrida, presión ejercida, etc.

Ligas como la Bundesliga o la UEFA Champions League ya lo utilizan de forma habitual para seguir los movimientos de los jugadores en tiempo real.

Second Spectrum

Second Spectrum [52] lleva el análisis espacial un paso más allá. Es el proveedor oficial de datos para la Premier League y la MLS, y emplea visión por computador para reconstruir el partido en 3D. Esto permite estudiar desde cómo un equipo ocupa los espacios hasta cómo coordina una presión colectiva.

Sus modelos se han usado, por ejemplo, para analizar los movimientos sincronizados del Manchester City y su capacidad para generar superioridades en zonas clave.

SciSports

SciSports [53] aporta una capa de inteligencia predictiva. Utiliza modelos de *machine learning* para valorar el rendimiento y el potencial de los jugadores. Su métrica principal, *SciSkill*, estima cuánto puede aportar un futbolista en función de su encaje con el estilo del equipo.

Es especialmente útil para detectar talento oculto en ligas menos mediáticas o encontrar jugadores que podrían adaptarse a contextos tácticos muy concretos.

Sistemas GPS y sensores biométricos

Más allá de los proveedores anteriores, hay una dimensión física que también se analiza con detalle. Muchos clubes utilizan dispositivos GPS o sensores biométricos, integrados en petos o camisetas para monitorizar en vivo la carga de trabajo de los jugadores. Estos sistemas permiten registrar variables como:

- Distancia total recorrida.
- Número de sprints y aceleraciones.
- Ritmo cardíaco y carga fisiológica.
- Nivel de fatiga y riesgo de lesión.

Firmas como Catapult [54] o STATSports [57] han facilitado que los entrenamientos se personalicen casi al milímetro, reduciendo el riesgo de sobrecargas o lesiones innecesarias.

Conclusión

Cada proveedor tiene su propio enfoque y especialidad, y la elección de uno u otro depende de las necesidades concretas del club, cuerpo técnico o analista. A veces se busca entender cómo se mueve un bloque defensivo; otras, prevenir una lesión; otras, simplemente encontrar al próximo gran talento.

Pero lo realmente interesante es lo que ocurre cuando se combinan estos mundos: eventos, tracking y sensores. Al integrarlos, se consigue una imagen completa, casi tridimensional, del rendimiento en el fútbol. Una especie de radiografía que revela mucho más que lo que se ve en directo.

2.5. Metodologías de análisis de datos en el fútbol

Hoy en día, entender lo que pasa en un partido va mucho más allá de ver goles o contar posesiones. Gracias al análisis de datos, podemos capturar lo invisible, lo que a simple vista muchas veces se escapa. Y para hacerlo, existen distintas metodologías que han revolucionado la forma en la que entrenadores, analistas e incluso aficionados curiosos interpretan el juego.

Las más conocidas y utilizadas son el **eventing** y el **tracking**, aunque no son las únicas. Hay enfoques que complementan y enriquecen la visión general, como el análisis espacial, las redes de pases, los modelos predictivos o incluso simulaciones tácticas avanzadas.

A continuación, se explican las principales metodologías, sus herramientas y las aplicaciones prácticas que las han convertido en esenciales [55].

Eventing: cada acción cuenta

El *eventing* se basa en algo que puede parecer simple, pero que encierra un potencial enorme: registrar cada acción que ocurre durante un partido. Un pase, un disparo, una entrada, una pérdida, todo queda anotado, acompañado de datos como el jugador involucrado, la zona del campo, el minuto o el resultado de la acción.

Es como escribir una crónica detallada, jugada a jugada.

Eventos más comunes:

- Pases (exitosos o fallidos)
- Tiros (a puerta, bloqueados, desviados)
- Conducciones, regates y pérdidas
- Acciones defensivas (intercepciones, despejes, entradas)
- Faltas, penaltis, fueras de juego, etc.

Y lo más valioso es que no solo se registra, sino que se visualiza. A continuación, algunas de las representaciones más utilizadas:

Visualizaciones derivadas del eventing:

- **Mapa de tiros:** ideal para comprender el perfil ofensivo de un jugador o un equipo. Como se aprecia en la *Figura 2.14* el gráfico muestra desde dónde se disparó, con qué pierna, el valor de xG, y si fue gol o no.

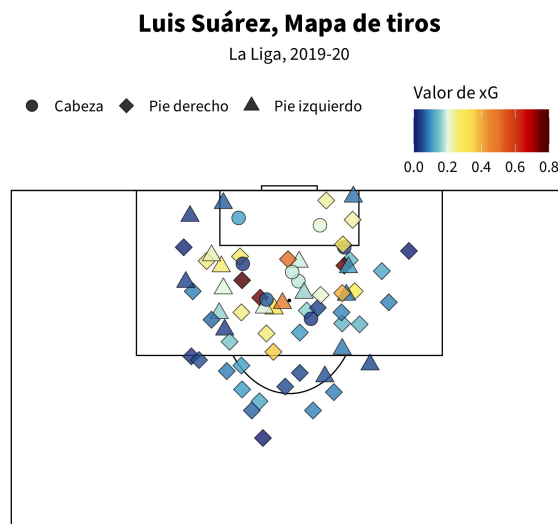


Figura 2.14: Mapa de tiros de Luis Suárez durante La Liga 2019-2020.

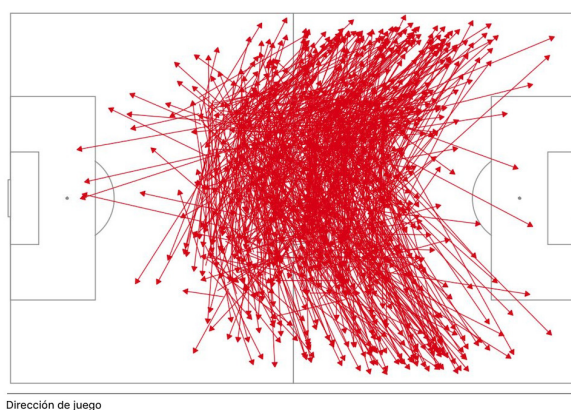
Fuente: hudlstatsbomb

- **Mapa de pases:** muestra cómo circula el balón, qué asociaciones destacan y qué zonas o tipos de pase se utilizan más, como se aprecia en la *Figura 2.15*.

Pases completados.

Desde el centro del campo.

Toni Kroos
Real Madrid
La Liga 2022-23



StatsBomb
Data Champions.

Figura 2.15: Mapa de pases completados de Toni Kroos durante La Liga 2022-2023.

Fuente: hudlstatsbomb

- **Mapa de conducciones:** visualiza trayectorias individuales con balón, detectando quién rompe líneas o genera peligro al avanzar, como bien muestra la *Figura 2.16*.

Carries ending in the box.

Nico Williams

Athletic Club
La Liga 2023/24

StatsBomb
Data Champions.

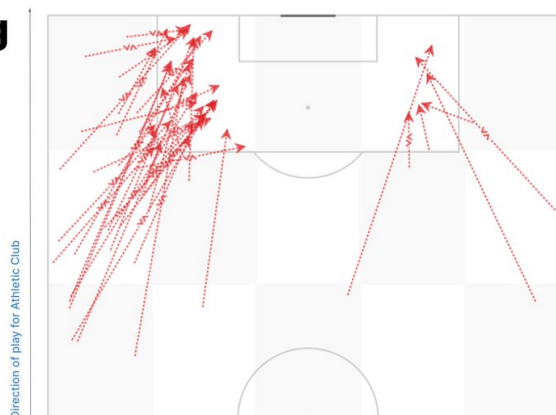


Figura 2.16: **Conducciones** hacia el área de Nico Williams durante La Liga 2023-2024.

Fuente: hudlstatsbomb

- **Mapa de acciones defensivas:** ubica entradas, despejes e intercepciones, ayudando a visualizar la intensidad y distribución defensiva, como se aprecia en la *Figura 2.17*.

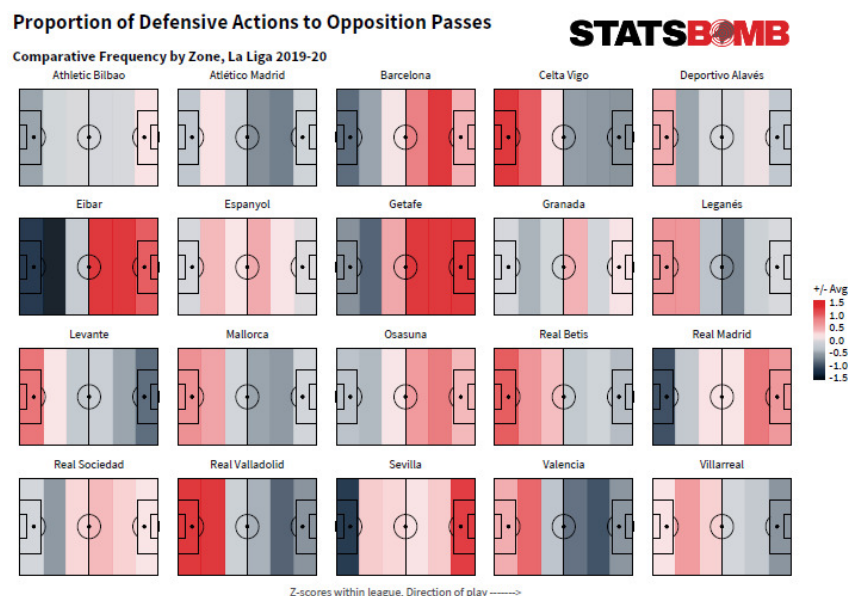


Figura 2.17: **Mapa de acciones defensivas** de todos los equipos durante La Liga 2019-2020.

Fuente: hudlstatsbomb

Aplicaciones prácticas del eventing:

- Cálculo de métricas avanzadas como *Expected Goals* (xG) y *Expected Assists* (xA).
- Análisis de patrones de juego a través de secuencias de eventos, como las jugadas previas a un gol.
- Comparación entre jugadores, ajustando estadísticas por posición o por cada 90 minutos.

Tracking: seguir cada movimiento, segundo a segundo

El *tracking* va más allá del registro de eventos. Es una metodología basada en el seguimiento automatizado y continuo del balón y los jugadores durante todo el partido. Se utilizan sistemas de cámaras (como TRACAB o Second Spectrum) o dispositivos GPS adheridos a los futbolistas.

Gracias a esto, obtenemos datos espaciales y temporales de altísima resolución, a veces hasta 25 veces por segundo.

¿Qué tipo de datos ofrece el tracking?

- Posición x e y del balón y cada jugador durante el encuentro.
- Velocidad, aceleración, cambios de ritmo.
- Distancia entre jugadores, amplitud del bloque, profundidad ofensiva.
- Movimientos sin balón, ayudas defensivas y coberturas.

Visualizaciones y modelos derivados del tracking:

- **Mapa de presión:** muy parecido al anterior, como se muestra en la *Figura 2.18*, identifica zonas donde un equipo ejerce mayor presión, detectando estructuras defensivas altas, medias o bajas.

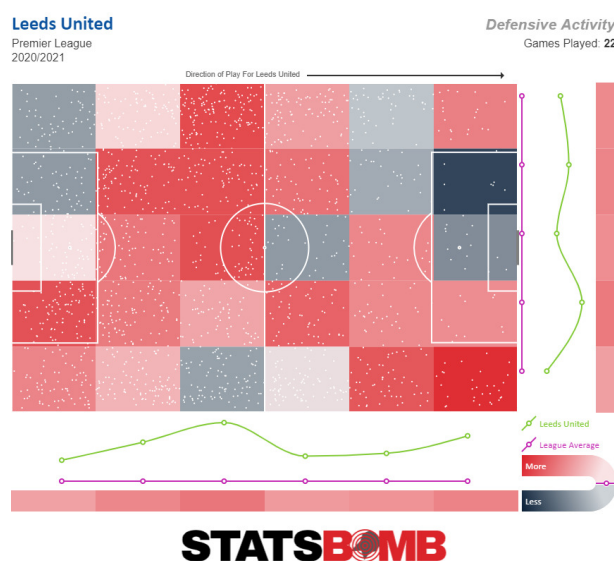


Figura 2.18: Mapa de presión de todos los equipos durante La Liga 2019-2020.

Fuente: hudlstatsbomb

- **Modelos de control espacial (Voronoi y Pitch Control):** calculan qué zonas domina cada jugador o equipo, basándose en su posición y velocidad. Ayudan a entender quién “controla” el espacio, algo fundamental para valorar la estructura táctica, como se aprecia en la *Figura 2.19*.

Dato curioso: Guardiola usaba este tipo de modelos durante su etapa en el FC Barcelona para ajustar presiones y posicionamientos.

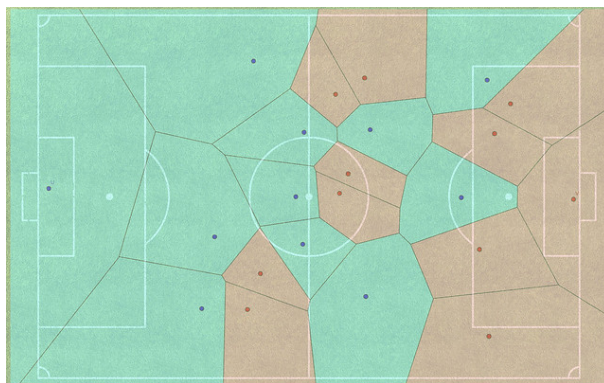


Figura 2.19: **Diagrama de Voronoi** aplicado al análisis futbolístico.

Fuente: LaSexta

- **Mapas de calor:** visualizan la intensidad y frecuencia del movimiento de un jugador en distintas zonas. Revelan su implicación posicional y recorrido habitual a lo largo de un encuentro, como se observa en la *Figura 2.20*.

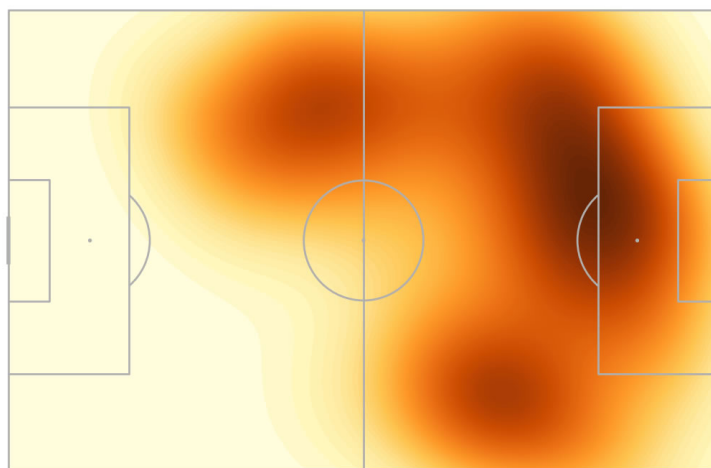


Figura 2.20: **Mapa de calor (HeatMap)**.

Fuente: FÚTBOL CON DATOS

Aplicaciones del tracking:

- **Análisis táctico detallado:** presiones, coberturas, repliegues, transiciones ofensivas y defensivas.
- **Evaluación del esfuerzo físico, organización colectiva y capacidad de recuperación del equipo.**

En definitiva, estas metodologías han abierto un universo nuevo en la forma de ver y analizar el fútbol. Permiten entender lo que ocurre sobre el césped con una profundidad que antes parecía imposible. Y, sinceramente, una vez te acostumbras a mirar el juego con estos datos, ya no hay vuelta atrás.

Integración de metodologías y miradas hacia el futuro

Hoy en día, analizar el fútbol va mucho más allá de contar goles o registrar estadísticas sueltas. La verdadera riqueza del análisis moderno está en cómo se entrelazan distintas metodologías, especialmente el *eventing* y el *tracking*, para ofrecer una visión más completa y real del juego. Y es que, por separado, cada una tiene su valor, pero cuando se combinan ahí es donde ocurre la magia.

Por ejemplo, no basta con saber que un jugador dio un pase clave. Lo interesante es entender cómo se movía antes de darlo, qué compañeros lo rodeaban, cuánta presión recibía o a qué velocidad corría. Ese tipo de contexto lo aporta el tracking, completando y enriqueciendo la lectura que ofrecen los eventos [56].

Cómo se complementan estas metodologías:

- **Análisis de transiciones:** al cruzar datos de eventos (como pérdidas o recuperaciones de balón) con variables de tracking (como velocidad o distancias recorridas), se pueden evaluar mejor las transiciones rápidas, los contraataques o cómo se organiza el equipo tras perder la posesión.
- **Evaluación integral de jugadores:** combinar métricas como goles esperados (xG), asistencias o acciones defensivas con datos físicos como sprints o carga de trabajo permite trazar un perfil mucho más completo y realista del rendimiento individual.
- **Simulaciones tácticas:** con registros históricos de eventos y movimientos, algunos equipos ya recrean situaciones de juego para ensayar distintos planteamientos. Es como tener una máquina del tiempo que te permite cambiar decisiones y ver qué hubiera pasado.

¿Y hacia dónde vamos?

- **Inteligencia Artificial y modelos predictivos:** cada vez más clubes utilizan algoritmos para anticipar lesiones, prever rendimientos futuros o incluso predecir resultados basados en patrones pasados.
- **Realidad Virtual e inmersión total:** se están desarrollando herramientas que permiten a jugadores y entrenadores meterse literalmente dentro del partido.

- **Datos en tiempo real:** ya no hay que esperar al post-partido. Muchos equipos están empezando a tomar decisiones tácticas durante el juego, ajustando el planteamiento en función de lo que realmente está pasando.

Reflexión final

Lo que antes eran tablas de goles y asistencias ahora se ha convertido en un universo de datos interconectados, donde cada pase, cada movimiento y cada decisión se puede medir, analizar y entender. El *eventing* nos cuenta qué pasó, el *tracking*, cómo y dónde ocurrió. Juntos, ofrecen un contexto mucho más enriquecido.

La incorporación de nuevas tecnologías como la IA o la realidad virtual no solo está revolucionando el trabajo de los cuerpos técnicos y analistas. También está empezando a cambiar cómo viven el fútbol los propios aficionados, que ahora tienen acceso a herramientas antes reservadas a los profesionales.

Y quién sabe quizás en unos años, cualquiera desde su móvil podrá analizar un partido como lo haría un analista de élite. Porque el fútbol, aunque cada vez más tecnológico, sigue siendo de todos.

Capítulo 3

Materiales

3.1. Datos del estudio

StatsBomb Open Data

Hablar de análisis avanzado en fútbol y no mencionar a StatsBomb sería quedarse corto. Esta empresa ha sabido ganarse un lugar privilegiado en el ecosistema del fútbol moderno, no solo por la calidad de sus datos, sino por su voluntad de compartirlos. En 2018 lanzaron su iniciativa *Open Data*[58], una plataforma abierta pensada para fomentar la investigación, el aprendizaje autodidacta y, por qué no, la experimentación de quienes buscamos ir un paso más allá con los datos.

Para este proyecto, se han utilizado los registros correspondientes a la Eurocopa 2024 (Competition ID: 55, Season ID: 282), accedidos mediante la librería oficial `statsbombpy`. Gracias a esta API, el acceso a los datos resulta sorprendentemente sencillo: partidos, eventos, alineaciones, esquemas tácticos; todo listo para ser analizado con unas pocas líneas de código.

Y es que uno de los grandes motivos para elegir esta fuente, además de tener datos abiertos, es precisamente eso: su facilidad de uso. No hay que pelearse con formatos raros ni con documentos mal estructurados. Todo viene limpio, ordenado y documentado. Además, existe una comunidad muy activa alrededor de StatsBomb, lo que significa que hay infinidad de recursos, ejemplos, notebooks y tutoriales que hacen que el proceso de aprendizaje sea mucho más accesible y motivador.

Lo más valioso de este dataset no es solo la cantidad de información, sino su nivel de detalle. Cada pase, disparo, recuperación o falta incluye coordenadas exactas del campo, el tipo de acción, el jugador implicado y hasta si había presión rival o no. Esto abre la puerta a análisis técnico-tácticos mucho más profundos, casi como si uno pudiera “leer” el partido desde los datos.

Kaggle y FBref

La segunda gran fuente de datos proviene de Kaggle, un sitio bien conocido por los que nos movemos en el mundo de la ciencia de datos. Allí se publicó el conjunto *Football Players Stats 2024-2025*[59], que reúne estadísticas de jugadores de las cinco grandes ligas europeas extraídas mediante técnicas de **web scraping** desde el portal `FBref.com`.

FBref, gestionado por *Sports Reference*, se ha convertido en un recurso imprescindible para quienes queremos ir más allá del gol y la asistencia. ¿Lo interesante? Que sus estadísticas provienen de Opta, uno de los gigantes del sector. Esto garantiza una base sólida y rigurosa, pero además incluye métricas avanzadas como los goles esperados (xG), asistencias esperadas (xA), pases progresivos, regates, acciones defensivas y un largo etcétera.

La elección de este dataset no fue casual. Además de ser completamente gratuito, está disponible en un formato limpio y estructurado, lo cual ahorra muchas horas de preprocesamiento. Y, al igual que con StatsBomb, hay una gran cantidad de contenido didáctico alrededor: notebooks, kernels públicos, vídeos y explicaciones paso a paso. Todo eso hace que trabajar con estos datos no sea solo viable, sino también fácil y práctico.

El dataset en sí es bastante amplio: más de 250 variables por jugador, organizadas por tipo de acción (ataque, defensa, posesión, físico, etc.). Aunque no fue construido por mí, el scraping fue realizado con mimo por su autor original y actualizado cada semana. Esa riqueza de columnas, mezclando variables numéricas y categóricas, lo convierte en una mina de oro para explorar el rendimiento individual a lo largo de una temporada completa.

Además, otra parte de los datos de FBref se ha recopilado específicamente para alimentar los modelos predictivos desarrollados en este trabajo. En este caso, se han utilizado estadísticas individuales de jugadores de las cinco grandes ligas europeas a lo largo de las últimas cinco temporadas. Esta recopilación más extensa se realizó mediante la librería *LanusStats*[61], lo que permitió automatizar el scraping de cientos de páginas y mantener una estructura coherente entre temporadas. Una tarea que, aunque técnica, fue tremendamente satisfactoria al ver cómo los datos cobraban vida a medida que avanzaba el análisis.

DFL - Bundesliga Data Shootout (Kaggle)

Cuando se lanzó la competición *DFL - Bundesliga Data Shootout* en Kaggle[60], se presentó una oportunidad relevante para la comunidad de análisis de datos aplicada al fútbol. Por primera vez, una liga profesional como la Bundesliga alemana puso a disposición pública un conjunto de datos de alto nivel, invitando a investigadores y analistas a trabajar con información detallada y estructurada.

Este dataset contiene datos oficiales de tracking y eventing organizados de forma rigurosa, incluyendo coordenadas espaciales, marcas temporales y descripciones detalladas de cada acción. Entre la información disponible se encuentran posesiones completas, posiciones de los jugadores en cada instante, velocidad y aceleración, datos que hasta hace poco eran accesibles únicamente para clubes o entidades con acuerdos exclusivos.

Estos son los datos sobre los cuales se aplican los modelos de visión por computadora desarrollados con la plataforma *Roboflow*, proporcionando así un contexto claro sobre su utilización en este proyecto.

Características de los datos

Tipos de datos

A lo largo del proyecto he trabajado con tres tipos de datos bastante distintos entre sí, aunque complementarios. Cada uno aporta una mirada diferente sobre el juego, y juntos permiten un análisis más completo y enriquecedor.

- **Datos de competiciones:** procedentes de *StatsBomb*. Recogen acciones específicas ocurridas durante los partidos, como pases, tiros, entradas o duelos, todo perfectamente codificado en formato JSON. Gracias a Python y su ecosistema de librerías, estos datos se transformaron en tablas manejables, listas para explorar. Lo bonito de este tipo de información es que permite seguir el hilo de una jugada, casi como si se rebobinara el partido, y analizar con lupa las decisiones que se toman sobre el césped.
- **Datos de rendimiento de los jugadores:** extraídos del dataset compartido en *Kaggle*, basado en estadísticas de *FBref*. Aquí, cada fila representa a un jugador, y cada columna resume su impacto en distintas áreas del juego.
- **Datos de tracking y eventing de la Bundesliga:** proporcionados por la competición *DFL - Bundesliga Data Shootout* disponible en *Kaggle*[60]. Este conjunto incluye datos oficiales de seguimiento posicional y eventos detallados, con coordenadas espaciales, tiempos exactos y descripciones precisas de cada jugada. Aunque no fue la fuente principal para los modelos predictivos de este proyecto, estos datos permiten experimentar con análisis visuales avanzados y técnicas de visión por computadora.

Para una mayor explicación de los dos primeros conjuntos de datos, revisar los **Anexos I y II**.

Volumen de datos

El volumen de información con el que se ha trabajado no es precisamente pequeño, y es que cuando hablamos de datos futbolísticos, cada detalle cuenta y se guarda.

- El dataset de eventos de *StatsBomb* recoge los 51 partidos disputados durante la Eurocopa 2024. Cada uno de ellos incluye entre 3500 y 4500 eventos aproximadamente, lo que se traduce en decenas de miles de filas en total. Cada evento viene acompañado de 77 columnas con todo tipo de detalles: desde la localización exacta en el campo hasta si hubo presión del rival o la altura de un pase.
- Por otro lado, el conjunto extraído desde *Kaggle* (basado en datos de *FBref*) incluye estadísticas de más de 2000 jugadores. Cada observación cuenta con más de 250 columnas, lo que ofrece un retrato muy completo del rendimiento de cada futbolista a lo largo de la temporada.
- Finalmente, el dataset de la *DFL - Bundesliga Data Shootout* contiene información detallada de múltiples partidos de la Bundesliga, con millones de registros entre

datos de tracking y eventos. Este volumen permite un análisis muy granular y es ideal para aplicar técnicas de visión por computadora y aprendizaje automático en contextos reales y exigentes.

3.2. Limpieza y preprocesamiento de datos

Una vez obtenidos los conjuntos de datos, el siguiente paso crucial fue preparar la información para su análisis y visualización. En el contexto de este trabajo, la limpieza y el preprocesamiento no solo implicaron cuestiones técnicas, sino también decisiones conceptuales: ¿qué queremos medir realmente?, ¿cómo comparamos a jugadores con diferentes contextos?, ¿qué métricas reflejan mejor el rendimiento en el campo?

Aunque las dos fuentes de datos ofrecían información rica y valiosa, sus características estructurales eran muy diferentes, como se explica en los **Anexos I y II**, lo que condicionó el enfoque y el tiempo dedicado al tratamiento de los datos.

StatsBomb

En el caso de *StatsBomb*, los datos fueron descargados mediante la librería oficial `statsbombpy`, que provee acceso directo y ordenado a los registros de partidos. Desde el primer momento, la calidad de estos datos fue evidente. Cada evento (ya fuera un pase, disparo, falta, recuperación o presión) venía descrito con una estructura clara, tipos de datos correctos y sin errores aparentes. Las variables estaban bien etiquetadas, y no fue necesario lidiar con valores nulos ni con inconsistencias entre tablas. Esta fiabilidad permitió trabajar casi de forma inmediata en la exploración visual de secuencias, mapas de calor, redes de pases y otros análisis tácticos.

Sin embargo, eso no significa que no hubiera que tomar decisiones. Por ejemplo, se realizó un filtrado de eventos no relevantes para ciertos análisis, se convirtieron coordenadas a escalas relativas (para permitir comparaciones entre equipos o jugadores) y se generaron métricas adicionales como la acumulación de acciones ofensivas por zona o la construcción de jugadas que terminan en disparo. Para una explicación más a fondo de los datos, consultar el **Anexo I**.

Kaggle/FBref

Por el contrario, el conjunto de datos obtenido de *Kaggle*, compilado a partir de estadísticas de FBref mediante `web scraping`, supuso un desafío mucho mayor. A primera vista, el dataset impresionaba por su volumen: más de 250 columnas por jugador, cada una con una métrica distinta, ya fuese bruta (como goles o minutos), relativa (por 90 minutos), o expresada en porcentaje. Pero esta riqueza también venía acompañada de cierta desorganización, errores menores y, sobre todo, duplicaciones y estructuras difíciles de manejar directamente. Para una explicación más a fondo de los datos, consultar el **Anexo II**.

A lo largo del proyecto, se realizó un proceso de limpieza y preprocesado sobre estos datos en Python. De esta forma, conseguimos una base de datos más sólida, coherente, completa y representativa del rendimiento real de cada jugador.

Variables finales

Selección de variables

El proceso de selección de variables fue una de las etapas más complejas y relevantes del trabajo. Partiendo de más de 250 columnas disponibles en el conjunto de datos original, se realizó una depuración exhaustiva para conservar únicamente aquellas estadísticas que aportaban valor real en términos analíticos y futbolísticos. El objetivo no era simplemente reducir, sino dar sentido: encontrar métricas que permitieran representar el rendimiento de los jugadores de forma justa, interpretable y comparativa.

Las variables finales se organizaron en bloques temáticos:

- **Identificación y contexto:** Player, Nation, Squad, Comp, Pos, Age, MP, Starts, Min, 90s.
- **Producción ofensiva:** Gls, Ast, G-PK, Gls_90, Ast_90, G+A, G+A_90, G-PK_90, xG, xAG, xG+xAG, xG_90, xAG_90, xG+xAG_90, npG, npG_90, G-xG, G-xG_90, np:G-xG, np:G-xG_90, Sh, SoT, Sh_90, SoT_90, G/Sh, G/SoT, Dist, npG/Sh.
- **Creación de juego y distribución:** xA, SCA, SCA90, GCA, KP, KP_90, Cmp, Att, Cmp_90, Cmp %, PPA, PPA_90, CrsPA, CrsPA_90, A-xAG, PassLive, PassDead.
- **Conducción, regate y progresión:** PrgP, PrgP_90, PrgC, PrgC_90, PrgR, Succ, SuccDrib_90, Carries, Touches, Touches_90, Att Pen, Rec, TotDist, PrgDist, 1/3, 1/3_90, Att_Act_90.
- **Métricas defensivas:** Tkl, Int, Tkl+Int, Tkl+Int_90, TklW, Tkl %, Att_stats_defense, Blocks_stats_defense, Blocks_90, Clr, Clr_90, Recov, Recov_90, Won %.
- **Rendimiento de porteros:** GA, GA90, PSxG, PSxG+/-, SoTA, Saves, Saves_90, Save %, Cs, Cs %.

Este conjunto de variables permitió capturar las dimensiones más relevantes del juego de cada futbolista, desde su contribución ofensiva y creativa hasta su rendimiento defensivo o bajo palos. El listado completo y la explicación detallada de cada variable pueden consultarse en el **Anexo II**.

Normalización de métricas por 90 minutos

Una de las primeras decisiones fue estandarizar las estadísticas por tiempo jugado. En el fútbol, un jugador que ha disputado 3000 minutos y otro que apenas ha jugado 900 no pueden compararse directamente si no se ajusta su producción al mismo marco temporal. Para ello, se creó una función que calculaba las métricas por 90 minutos, lo que permitió igualar el contexto de comparación.

Adicionalmente, se eliminaron del conjunto aquellos jugadores que no alcanzaron los 900 minutos disputados en la temporada, con el fin de garantizar comparaciones más estables y representativas. Todas las variables categóricas fueron convertidas a variables

numéricas mediante codificación para poder ser utilizadas en modelos de predicción y análisis estadístico posterior.

Este paso fue clave para variables ofensivas como goles, asistencias, disparos o regates, pero también para métricas defensivas y de construcción. Además, se añadieron nuevas columnas calculadas, como:

- **xG + xAG**: para evaluar la contribución ofensiva esperada total.
- **G+A_90**: suma directa de goles y asistencias, normalizada.
- **Tkl+Int_90**: acciones defensivas conjuntas por partido.
- **Att_Act_90**: indicador sintético de producción ofensiva total.

Este tipo de métricas compuestas permite capturar dimensiones complejas del rendimiento en un único valor más interpretativo.

Consolidación de jugadores con múltiples registros

Uno de los problemas más significativos del dataset era la presencia de jugadores con más de un registro. Esto se debía a que, durante una misma temporada, muchos futbolistas cambian de equipo (ya sea en el mercado de invierno, por cesión o traspaso), lo que generaba múltiples filas con estadísticas fragmentadas.

Para abordar esto, se diseñó un algoritmo de consolidación que:

- Agrupa jugadores por su nombre y edad.
- Suma las métricas acumulativas (como minutos, goles, tiros...).
- Calcula la media en aquellas métricas expresadas en porcentaje o por 90 minutos.
- Conserva los datos de identidad (equipo, nacionalidad, posición) del club más reciente, ya que este refleja mejor el contexto final del jugador.

Esta operación también evitó distorsiones en visualizaciones y rankings, donde un jugador podía aparecer dos veces con datos incompletos.

Cálculo de percentiles por posición

Otro paso fundamental fue la creación de percentiles por posición. En fútbol, comparar directamente a un defensa y a un delantero en métricas como regates o disparos no tiene mucho sentido: cada posición tiene su rol, y por tanto, sus expectativas estadísticas.

La solución fue crear grupos por posición (por ejemplo, porteros, defensas, centrocampistas, delanteros) y calcular para cada métrica su percentil relativo dentro del grupo. Esto permitió, por ejemplo, decir que un central estaba en el percentil 85 en despejes, o que un mediocentro tenía una frecuencia de pases progresivos superior al 90 % de sus pares.

Además, se tuvo en cuenta que no todas las métricas son mejor cuanto más altas. Para variables como “goles encajados por 90 minutos”, se invirtió el cálculo del percentil, de forma que un valor bajo se tradujera igualmente en un percentil alto, representando mejor la calidad del rendimiento.

Configuración personalizada de métricas por posición

Finalmente, se definieron configuraciones específicas por rol en el campo. No todas las estadísticas son igual de relevantes para todas las posiciones. Por ejemplo:

- En delanteros se priorizan métricas de finalización y creación ofensiva.
- En defensas, se valoran más las intervenciones, bloqueos y duelos ganados.
- En centrocampistas, cobran peso los pases progresivos, la posesión y la polivalencia.

Estas configuraciones permitieron alimentar gráficos como *radar plots* o *pizza plots*, adaptados a cada rol, y también construir rankings más realistas del rendimiento individual, ponderando la importancia relativa de cada métrica.

Índice de rendimiento ponderado por posición

Además de la visualización, se diseñó un **índice de rendimiento global** que combina múltiples estadísticas en una única métrica resumen. Para lograr una comparación justa y contextualizada, las estadísticas fueron ponderadas según la posición del jugador. Cada rol en el campo tiene funciones y contribuciones distintas, por lo que las métricas relevantes y su peso varían. Por ejemplo:

- En porteros, se priorizan métricas como PSxG+/-, porcentaje de paradas y goles encajados por 90 minutos.
- En defensas, tienen más peso los despejes, bloqueos, duelos ganados y pases largos completados.
- En centrocampistas, se valoran los pases progresivos, intervenciones defensivas y creación de oportunidades.
- En atacantes, destacan los goles, xG, asistencias, regates y tiros a puerta.

Las métricas fueron normalizadas en percentiles dentro de su posición para asegurar comparabilidad y evitar sesgos numéricos. Luego, se aplicaron pesos específicos para calcular una media ponderada, generando así un índice que refleja el impacto global del jugador según su rol.

Conclusión

El proceso de limpieza y preprocesamiento no solo fue una etapa técnica, sino una fase de reflexión profunda sobre cómo representar el fútbol con datos. A través de funciones diseñadas a medida, cálculos específicos y mucha iteración, se logró transformar un conjunto de datos crudo y diverso en una base coherente, contextualizada y lista para extraer valor real. Esta base de datos no solo permitió construir modelos predictivos y visualizaciones avanzadas, sino también contar historias. Historias de jugadores, de equipos, de roles, todo leído desde los números, pero con el fútbol como hilo conductor.

Capítulo 4

Modelos Predictivos

4.1. Modelos predictivos

Uno de los retos más emocionantes de este trabajo fue construir modelos capaces de anticipar el rendimiento ofensivo de los futbolistas. Tras muchas horas analizando datos, visualizando patrones y entendiendo cómo se relacionan entre sí ciertas estadísticas, surgió la pregunta inevitable: *¿podríamos predecir cuántos goles o asistencias hará un jugador la próxima temporada, basándonos en lo que hizo la anterior?*

4.2. Enfoque temporal

Se construyeron dos modelos predictivos independientes: uno para anticipar el número de **goles** que un jugador marcará en la próxima temporada, y otro para estimar sus **asistencias**. Ambos modelos comparten una estructura común: utilizan los datos de una temporada t para predecir lo que ocurrirá en la temporada siguiente $t+1$. De este modo, cada fila del conjunto de entrenamiento representa una temporada pasada de un jugador, y la variable objetivo corresponde a su producción ofensiva en la campaña siguiente.

Los datos abarcan las cinco últimas temporadas de las principales ligas europeas, lo que permite detectar patrones recurrentes y capturar la evolución del rendimiento a lo largo del tiempo.

4.3. Modelo de predicción de goles con XGBoost

El objetivo de este modelo es anticipar cuántos goles marcará un jugador en la próxima temporada, utilizando exclusivamente su rendimiento estadístico de la temporada anterior. Esta predicción es especialmente útil en contextos de análisis de rendimiento, scouting y planificación de plantillas, ya que permite estimar el potencial ofensivo de un futbolista con base en datos históricos.

4.3.1. Preparación de los datos

Para construir el modelo, se utilizaron datos multitemporada correspondientes a diferentes campañas consecutivas. Se reorganizó el dataset de forma que cada fila representa una combinación jugador-temporada, incorporando estadísticas de la temporada anterior como variables predictoras. Esta transformación temporal permite simular un entorno realista en el que únicamente se dispone de datos pasados para anticipar el rendimiento futuro.

4.3.2. Variables predictoras

Se consideraron múltiples grupos de variables, todas ellas derivadas de la temporada previa (`_prev`):

- **Edad y volumen de juego:** minutos jugados, partidos equivalentes a 90 minutos.
- **Remate:** goles, disparos totales, tiros a puerta, distancia media del disparo, `xG` (goles esperados), `npxG` (`xG` sin penaltis), y efectividad goleadora.
- **Creación de ocasiones:** asistencias, `xA` (asistencias esperadas), pases clave, centros, y acciones generadoras de disparos y goles.
- **Participación ofensiva:** conducciones progresivas, toques en el área, recepciones, y entradas al área rival.
- **Distribución y pase:** acierto de pase, distancia progresiva, y volumen de pases en zona ofensiva.
- **Variables categóricas:** posición, equipo y liga.

4.3.3. Modelo y ajuste

Para la predicción de la cantidad de goles en la siguiente temporada, se empleó el algoritmo **XGBoost Regressor**, una técnica de aprendizaje supervisado basada en árboles de decisión con boosting gradual. La elección de XGBoost se fundamenta en varios aspectos clave:

- **Capacidad para manejar relaciones no lineales y complejas:** XGBoost es capaz de capturar interacciones complejas entre variables, algo habitual en los datos futbolísticos donde múltiples factores afectan la capacidad goleadora.
- **Eficiencia y velocidad:** Gracias a su implementación optimizada, permite entrenar modelos rápidamente y con un menor consumo de recursos computacionales, lo que facilita la experimentación con diferentes configuraciones.
- **Robustez frente al sobreajuste:** Su mecanismo interno de regularización y el ajuste de hiperparámetros ayuda a prevenir el sobreajuste, mejorando la capacidad de generalización del modelo.

- **Éxito comprobado en problemas de regresión con datos tabulares:** XGBoost es un referente en competiciones y aplicaciones reales que trabajan con datos estructurados, mostrando un rendimiento sobresaliente en tareas similares.

Se procedió a ajustar los hiperparámetros mediante validación cruzada, optimizando parámetros como el número de árboles, profundidad máxima, tasa de aprendizaje y regularización.

El proceso de ajuste incluyó:

- Transformación de variables categóricas mediante *one-hot encoding*.
- Estandarización de las variables numéricas.
- Búsqueda de hiperparámetros mediante `GridSearchCV` con validación cruzada de 5 particiones.

Validación cruzada en el ajuste del modelo

Con el objetivo de evitar el sobreajuste y garantizar la robustez del modelo, se aplicó una técnica de validación cruzada durante el ajuste de hiperparámetros del algoritmo `XGBoostRegressor`. En concreto, se utilizó una **validación cruzada k -fold** con $k = 5$, implementada mediante la clase `GridSearchCV` de `scikit-learn`.

Este enfoque permite evaluar el rendimiento del modelo de forma más estable y menos dependiente de una única división entrenamiento-prueba.

Funcionamiento del proceso:

1. El conjunto de entrenamiento se divide en cinco subconjuntos (o “folds”) de tamaño similar.
2. Para cada combinación de hiperparámetros, se realizan cinco entrenamientos:
 - En cada uno, el modelo se entrena con cuatro folds (80 % de los datos) y se valida con el fold restante (20 %).
 - Se calcula el error absoluto medio (MAE) sobre el fold de validación.
3. Se promedian los errores obtenidos en las cinco iteraciones.
4. Se selecciona la combinación de hiperparámetros con menor error promedio.
5. Finalmente, el modelo se reentrena sobre todo el conjunto de entrenamiento utilizando los mejores hiperparámetros y se evalúa en el conjunto de prueba independiente.

Espacio de búsqueda de hiperparámetros:

```
param_grid_xgb = {  
    'model__n_estimators': [100, 200],  
    'model__max_depth': [3, 6, 9],  
    'model__learning_rate': [0.01, 0.1],  
    'model__gamma': [0, 0.1],  
    'model__subsample': [0.8, 1.0]  
}
```

Los hiperparámetros óptimos encontrados fueron:

- `n_estimators`: 200
- `max_depth`: 3
- `learning_rate`: 0.1
- `gamma`: 0
- `subsample`: 1.0

4.3.4. Evaluación del modelo

El modelo se evaluó sobre un conjunto de test independiente, obteniendo los siguientes resultados:

- **MAE (Error absoluto medio)**: 1.65 goles
- **R^2 (Coeficiente de determinación)**: 0.39

Esto indica que el modelo es capaz de explicar aproximadamente el 39% de la variabilidad en el número de goles futuros, con un error promedio de sólo 1.65 goles por jugador.

4.3.5. Importancia de variables

Tras el entrenamiento, se analizó la importancia relativa de las variables predictoras mediante el modelo XGBoost (*Figura 4.1*). Las tres características más influyentes en la predicción de goles fueron:

1. `shooting_xG_prev`: cantidad de goles esperados (xG) sin penaltis en la temporada anterior. Refleja la calidad y frecuencia de las oportunidades generadas por el jugador a partir de sus acciones ofensivas.
2. `possession_Att_Pen_prev`: número de conducciones que terminaron en el área rival. Este indicador mide la agresividad y capacidad del jugador para penetrar zonas de peligro mediante acciones individuales.
3. `shooting_SoT_prev`: total de tiros a puerta realizados en la campaña previa. Representa la precisión y volumen ofensivo del jugador en términos de finalización.

Estas tres variables combinan métricas de volumen, calidad de los tiros y capacidad de progresión hacia zonas peligrosas, capturando así distintos aspectos clave del perfil ofensivo de los jugadores.

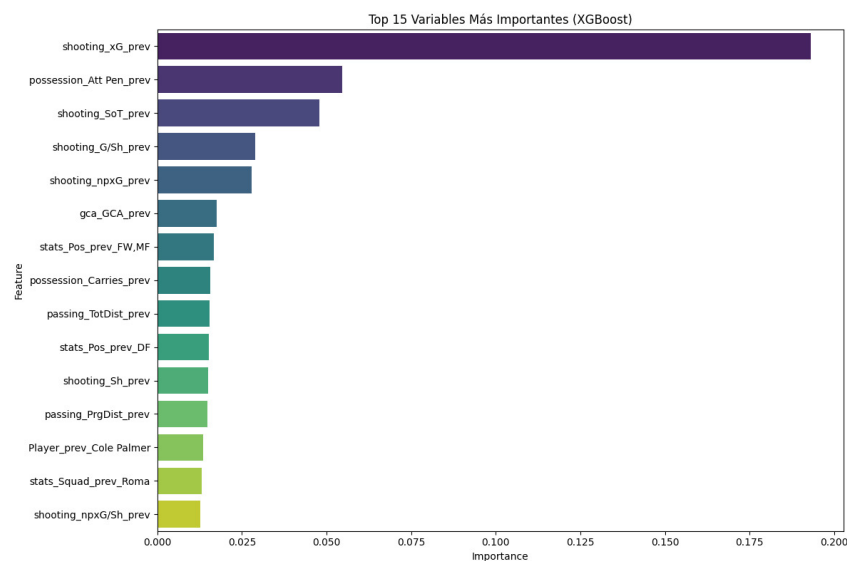


Figura 4.1: **Variables** más importantes del modelo de predicción de goles.

Fuente: Elaboración propia

4.4. Modelo de predicción de asistencias con Random Forest

El propósito de este modelo es predecir la cantidad de asistencias que un jugador registrará en la próxima temporada, empleando exclusivamente las estadísticas de su desempeño en la campaña anterior. Esta estimación resulta especialmente útil para el análisis de rendimiento y la planificación estratégica en contextos deportivos, ya que permite anticipar la capacidad creativa de un futbolista a partir de sus datos históricos.

Tras evaluar diversos algoritmos de regresión, como *XGBoost* o modelos lineales, se optó por un modelo basado en *Random Forest*, al ofrecer no solo los mejores resultados en términos de métricas de error, sino también predicciones más coherentes desde el punto de vista futbolístico.

4.4.1. Preparación de los datos

Se emplearon datos multitemporada en los que cada fila representa una combinación jugador-temporada, con variables predictoras correspondientes a las estadísticas de la temporada previa. Esta estructura permite un enfoque temporal realista para anticipar resultados futuros.

El preprocesamiento incluyó la eliminación de columnas redundantes y la transformación de variables categóricas como posición, equipo y liga mediante codificación *one-hot* para su correcta incorporación en el modelo.

4.4.2. Variables predictoras

Entre las variables utilizadas destacan las siguientes, todas correspondientes a la temporada anterior (`_prev`):

- **Edad y volumen de juego:** minutos jugados y partidos equivalentes a 90 minutos.
- **Creatividad ofensiva:** asistencias esperadas (`passing_xA_prev`), pases clave (`passing_KP_prev`) centros y precisión en el pase.
- **Participación ofensiva y progresión:** recepciones, entradas al área rival y conducciones progresivas.
- **Variables categóricas:** posición, equipo y liga.

4.4.3. Modelo y ajuste

Para la predicción del número de asistencias, se seleccionó un **Random Forest Regressor** debido a sus características particulares y adecuación a este tipo de variable:

- **Simplicidad y estabilidad:** Random Forest es un modelo robusto y relativamente sencillo de ajustar, que ofrece buenos resultados sin requerir una parametrización compleja.
- **Capacidad para evitar el sobreajuste:** Al combinar múltiples árboles entrenados sobre muestras aleatorias del conjunto de datos, reduce el riesgo de sobreajuste inherente a árboles individuales.
- **Interpretabilidad relativa:** Aunque menos transparente que modelos lineales, Random Forest permite analizar la importancia de variables de forma clara, facilitando la comprensión del modelo.
- **Resultados coherentes en el dominio de las asistencias:** En comparación con otros métodos probados, Random Forest proporcionó mejores predicciones y mayor consistencia para la variable de asistencias, donde las relaciones pueden ser menos lineales y más dispersas.

Como ya se ha comentado, se utilizó un **Random Forest Regressor** con los siguientes hiperparámetros por defecto:

- Número de árboles (`n_estimators`): 100
- Profundidad máxima (`max_depth`): sin límite (None)
- Número de variables a considerar por división (`max_features`): todas (1.0)
- Otros parámetros estándar: `min_samples_split=2`, `min_samples_leaf=1`, sin poda (`ccp_alpha=0.0`), entre otros.

Para garantizar una evaluación objetiva y evitar el sobreajuste, los datos se dividieron en dos conjuntos independientes: uno para entrenamiento y otro para prueba. Se utilizó la función `train_test_split` de la librería *scikit-learn* con una proporción del 80 % para entrenamiento y el 20 % restante para prueba. Esta división se realizó de forma aleatoria, fijando una semilla (`random_state=42`) para asegurar la reproducibilidad del proceso.

El conjunto de entrenamiento fue empleado para ajustar el modelo *Random Forest Regressor*. Antes del entrenamiento, se aplicó un preprocesamiento que incluyó la transformación de variables categóricas mediante codificación *one-hot* y el uso directo de variables numéricas. Este preprocesamiento se integró en un pipeline que permitió encadenar las transformaciones con el entrenamiento del modelo, facilitando un flujo ordenado y reproducible.

4.4.4. Evaluación del modelo

El desempeño se evaluó en un conjunto de prueba independiente, arrojando:

- **MAE (Error absoluto medio):** 1.19 asistencias
- **R² (Coeficiente de determinación):** 0.26

Estos resultados indican que el modelo explica aproximadamente el 26 % de la variabilidad en las asistencias futuras, con un error promedio menor a 1.19 asistencias por jugador.

4.4.5. Importancia de variables

El análisis de importancia (*Figura 4.2*) reveló que las dos variables más influyentes para la predicción fueron:

1. `passing_xA_prev`: asistencias esperadas en la temporada previa, que mide la calidad y cantidad de oportunidades creadas por el jugador.
2. `passing_KP_prev`: pases clave en la temporada previa, reflejando la capacidad para generar situaciones de peligro.

Estas métricas confirman que la producción creativa sostenida es fundamental para anticipar la capacidad asistidora futura.

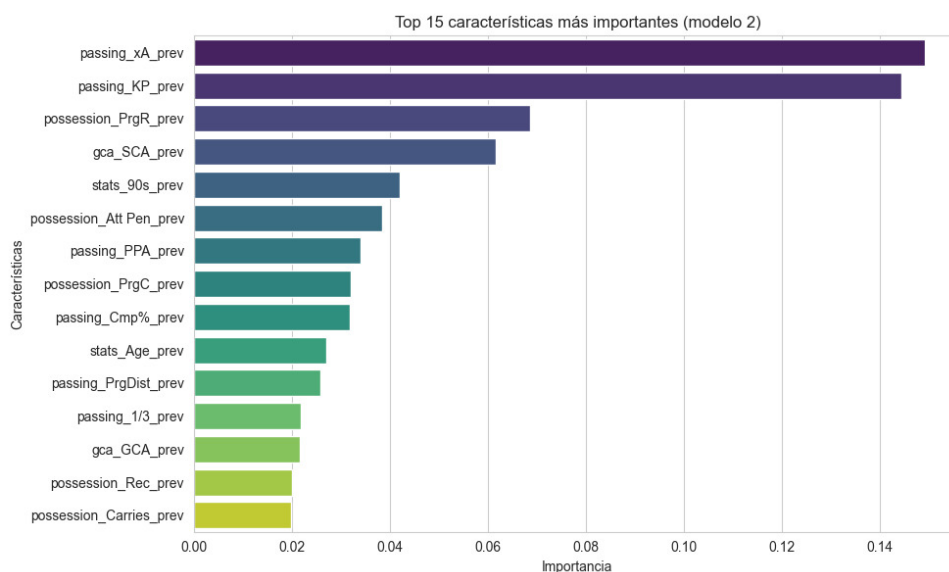


Figura 4.2: **Variables** más importantes del modelo de predicción de asistencias.

Fuente: Elaboración propia

Resultado

Los modelos desarrollados logran anticipar de forma razonable el rendimiento ofensivo de los jugadores, diferenciando claramente entre dos tareas muy distintas: marcar goles y asistir. Aunque el fútbol está lleno de factores impredecibles, desde decisiones tácticas hasta lesiones, rachas de forma o incluso la suerte, los datos permiten capturar señales valiosas que ayudan a estimar el potencial ofensivo futuro de un futbolista.

Este tipo de modelos no buscan ofrecer respuestas absolutas, sino aportar una mirada probabilística e informada sobre lo que podría suceder. Y lo cierto es que, con las variables adecuadas y un entrenamiento robusto, es posible acercarse bastante a esa predicción.

Los resultados obtenidos refuerzan esta idea. El error absoluto medio (*MAE*) fue de aproximadamente **1.62 goles** y **1.19 asistencias** por jugador. ¿Son estos errores mucho o poco? Para valorarlo, es fundamental ponerlos en contexto. En la temporada 2025 se anotaron un total de **3.705 goles** y **2.678 asistencias**, repartidos entre **1.784 jugadores**. Sin embargo, el reparto no fue equitativo: solo **965** jugadores marcaron al menos un gol y **940** ofrecieron al menos una asistencia.

Dicho de otro modo, el promedio global de goles por jugador fue de apenas **2.08**, y de **1.50** asistencias por temporada. En este contexto, un error medio de alrededor de **1.6** goles o **1.2** asistencias representa una desviación razonable, especialmente considerando que hay muchos jugadores con contribuciones muy bajas y unos pocos con cifras mucho más elevadas que distorsionan el promedio.

En la práctica, el modelo no siempre acertaba el número exacto, pero muchas veces detectaba correctamente la dirección del cambio: si un jugador estaba mejorando, o si, por el contrario, su impacto parecía decaer. Esa capacidad para anticiparse, aunque sea de forma aproximada, ya representa un valor importante en contextos de scouting, análisis o planificación de plantilla.

Capítulo 5

Visión Por Computadora

5.1. Visión por Computadora y Deep Learning aplicados al análisis futbolístico

El fútbol no se vive solo en el césped. Hoy, gran parte de su revolución ocurre en los píxeles de un vídeo. Y es que, con las herramientas adecuadas, una simple grabación de un partido se convierte en una fuente inagotable de información táctica, técnica y física. En este proyecto se ha explorado una implementación avanzada de visión por computadora centrada en el fútbol, partiendo de un enfoque publicado por *Roboflow*, una plataforma especializada en entrenamiento y despliegue de modelos de visión artificial.

5.2. Objetivo del módulo y enfoque general

El objetivo es claro: construir un sistema capaz de detectar jugadores, identificar a qué equipo pertenecen, seguir sus movimientos a lo largo del campo, y extraer información como la posesión del balón o la velocidad de los desplazamientos. Todo esto, únicamente a partir de un vídeo.

Para lograrlo, se combina el uso de modelos de detección de objetos, detección de puntos clave (*keypoints*) y algoritmos de seguimiento (*tracking*), apoyados por el uso de GPU en Google Colab y modelos de inferencia optimizados provistos por **Roboflow** y **HuggingFace**.

5.3. Configuración inicial y dependencias

Lo primero que se debe hacer para poder trabajar es seguir una serie de pasos de configuración muy relevantes antes de iniciar los módulos y modelos:

- Se cargan las claves de acceso a las APIs de HuggingFace y Roboflow, necesarias para usar modelos preentrenados y recursos computacionales.
- Se verifica la disponibilidad de GPU en Colab mediante el comando `nvidia-smi`, fundamental para ejecutar los modelos de inferencia en tiempo razonable.

- Se instalan librerías clave como `inference-gpu`, que permite correr modelos ligeros y rápidos de detección de objetos, y `supervision`, una librería especializada en visualización y lógica de seguimiento en vídeos deportivos.

Además, se descarga una serie de vídeos de ejemplo procedentes del dataset público de la Bundesliga (*DFL Data Shootout*), que servirán como base para el procesamiento posterior.

5.4. Módulo de detección y clasificación de jugadores, balón y árbitros

En este módulo se emplea un modelo de detección de objetos basado en una arquitectura preentrenada optimizada para deportes, cargado mediante la librería `supervision` [62]. Este modelo es capaz de identificar múltiples clases relevantes para el análisis futbolístico, tales como jugadores, porteros, árbitros y el balón.

La detección se realiza cuadro a cuadro, obteniéndose para cada objeto identificado una caja delimitadora (bounding box), una etiqueta de clase y una puntuación de confianza. Para mejorar la fiabilidad, se descartan detecciones con una confianza inferior a un umbral establecido (por ejemplo, 0.45) y se aplica supresión de no-máximos para eliminar solapamientos redundantes.

Una vez detectados los jugadores, se extraen los recortes correspondientes y se pasa cada uno a un clasificador secundario que asigna la pertenencia a uno de los dos equipos en juego, diferenciando así los conjuntos rivales en el partido. En el caso de los porteros, se determina su equipo asociándolos al conjunto correspondiente de jugadores detectados, basándose en la proximidad espacial y consistencia temporal, en la *Figura 5.1* se muestra un ejemplo.

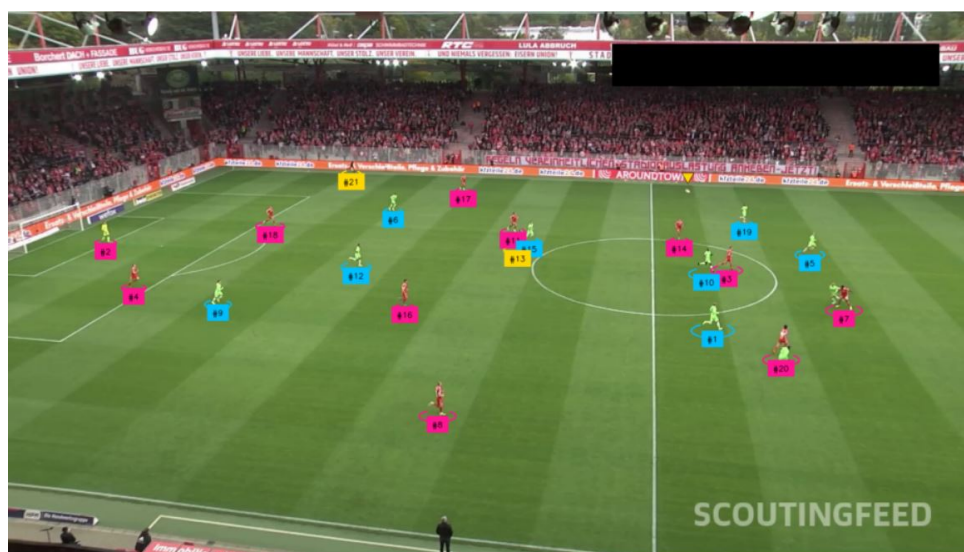


Figura 5.1: **Detección** de jugadores, árbitro y balón.

Fuente: *Roboflow*

5.5. Módulo de seguimiento multiobjeto

Para mantener la identidad de cada objeto a lo largo del vídeo y realizar un seguimiento continuo, se implementa un sistema de *tracking* multiobjeto utilizando el algoritmo ByteTrack [63]. Este método combina detección y asociación temporal para asignar identificadores únicos estables a cada jugador, árbitro y balón, incluso en situaciones de oclusión o cruce entre jugadores.

5.6. Módulo de detección y proyección de puntos clave del campo

Además de los objetos móviles, se emplea un modelo específico para la detección de *keypoints* estáticos del campo de juego, tales como esquinas, puntos de penalti y líneas de banda. Esta detección se realiza usando modelos basados en aprendizaje profundo accesibles a través de la plataforma Roboflow [64].

Los puntos clave detectados permiten calcular una homografía entre la imagen del vídeo y una vista aérea (plano bidimensional) del terreno de juego. Se aplica un suavizado temporal a esta transformación para evitar saltos bruscos en la proyección.

Gracias a esta calibración, las posiciones detectadas de jugadores, balón y árbitros se proyectan con precisión sobre el plano del campo, facilitando la interpretación táctica espacial, como podemos ver en la figura 5.2 se resaltan con círculos rojos los puntos claves detectados.



Figura 5.2: **Detección** de puntos clave del campo.

Fuente: Roboflow

5.7. Módulo de proyección espacial al campo (homografía)

Para transformar las coordenadas de detección de imagen (2D en píxeles) a posiciones reales sobre el campo (2D en metros), se requiere una estimación precisa de la transformación geométrica entre la imagen y el plano del terreno. Esto se consigue mediante una **homografía**, calculada a partir de puntos clave (*keypoints*) estáticos del campo.

Este módulo se basa en los siguientes pasos:

1. **Detección de puntos clave:** Se utiliza un modelo de detección especializado, cargado desde Roboflow [64], entrenado para reconocer elementos fijos del terreno de juego como líneas, esquinas, círculos centrales y áreas de penalti.
2. **Emparejamiento con plantilla:** Los puntos detectados se emparejan con una plantilla estática de campo (una vista aérea del terreno reglamentario) mediante un algoritmo de correspondencia basado en distancia y coherencia geométrica.
3. **Cálculo de la homografía:** A partir de los puntos emparejados se estima la matriz de homografía H utilizando el algoritmo de RANSAC, que permite descartar correspondencias erróneas.
4. **Suavizado temporal:** Para evitar saltos entre fotogramas consecutivos (por variaciones en los puntos detectados), se aplica un filtro temporal sobre la matriz H , interpolando con homografías anteriores.
5. **Proyección de coordenadas:** Finalmente, se transforman las posiciones de los objetos móviles (jugadores, balón, árbitros) desde coordenadas de imagen a coordenadas del campo usando $H \cdot (x, y, 1)^\top$.

Este módulo permite construir visualizaciones en vista aérea, generar mapas de calor realistas y calcular métricas espaciales como la ocupación de zonas, distancias recorridas o presión sobre áreas del campo.

Además, se convierte en un componente clave para futuras extensiones del sistema, como el análisis de formaciones, detección de líneas defensivas o cálculo de fuera de juego basado en perspectiva corregida.

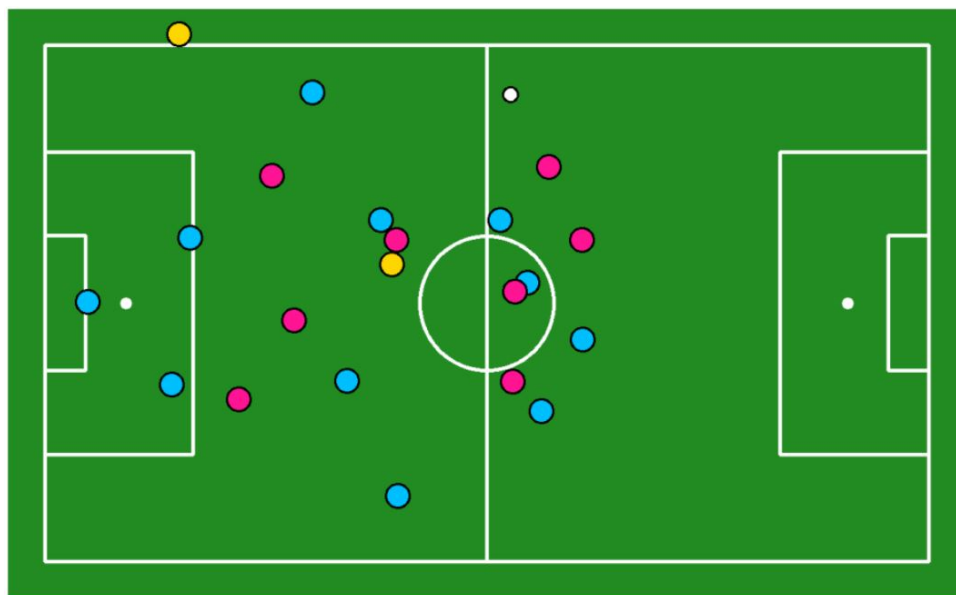


Figura 5.3: **Detección y proyección 2d** de jugadores, árbitro y balón.

Fuente:Roboflow

5.8. Módulo de seguimiento del balón

El seguimiento del balón representa un reto particular en visión por computadora, debido a su reducido tamaño, alta velocidad y frecuentes oclusiones. Para resolver esta tarea se emplea una combinación de detección cuadro a cuadro y lógica de asociación temporal, similar a la aplicada en el tracking de jugadores.

Tras detectar la posición del balón en cada fotograma mediante el modelo de detección multiclase, se emplea nuevamente el algoritmo ByteTrack [63] para asignar un identificador único al balón. Sin embargo, a diferencia del caso de los jugadores, solo se espera una instancia por fotograma, lo que simplifica la asociación, pero también la vuelve más susceptible a errores por falsos negativos.

Para paliar esto, se introduce una lógica adicional:

- Si el balón no es detectado durante algunos fotogramas consecutivos, se asume su continuidad y se extrapola su posición basada en la trayectoria previa.
- Se aplica un filtro temporal (media móvil o suavizado exponencial) sobre la trayectoria del balón para reducir fluctuaciones abruptas debidas a errores de detección.
- En caso de múltiples detecciones, se elige la de mayor confianza o la más próxima a la trayectoria reciente.

Este módulo es fundamental para calcular métricas de posesión (quién está más cerca del balón), velocidad del balón, momentos clave como disparos o pases o la propia trayectoria del balón. Su correcta implementación es clave para la fiabilidad del análisis global.

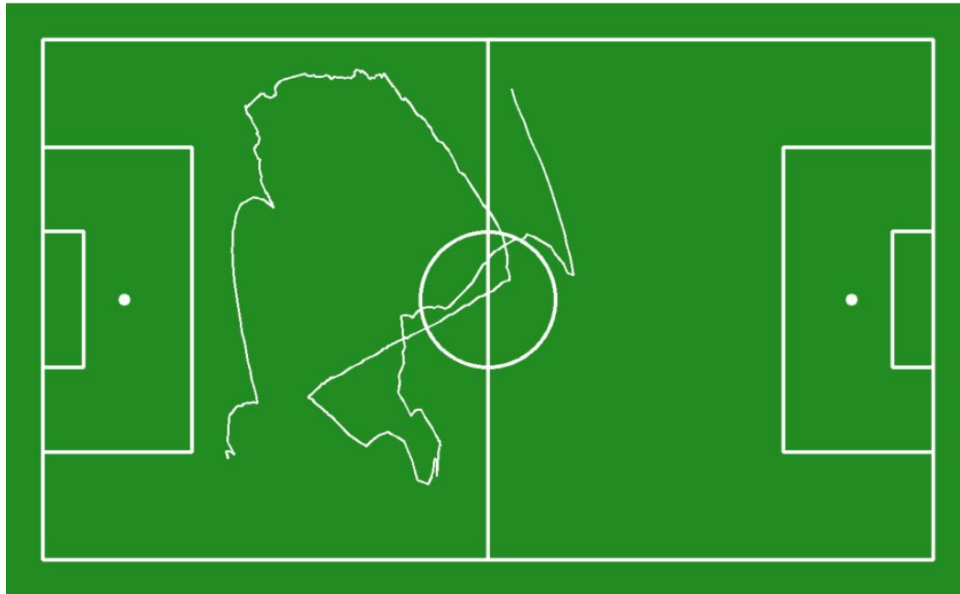


Figura 5.4: **Seguimiento** de la trayectoria balón.

Fuente: Roboflow

5.9. Funcionamiento de la función generadora de vídeos anotados

La función encargada de transformar el vídeo original en tres salidas visualmente enriquecidas, detección y seguimiento, diagrama de Voronoi y vista aérea proyectada, se organiza como un pipeline modular, en el que cada etapa aporta una pieza clave del análisis. Muchos de los módulos utilizados ya han sido explicados en detalle en apartados anteriores (detección de objetos, clasificación por equipos, tracking y detección de puntos clave), así que en esta sección nos centraremos en cómo se combinan para lograr un resultado coherente y visualmente informativo.

1. **Carga y lectura del vídeo:** Todo comienza con la lectura del archivo de vídeo, que se realiza fotograma a fotograma mediante una librería especializada como `opencv-python`. Este enfoque permite aplicar el procesamiento en tiempo real o por bloques, según se necesite.
2. **Pipeline de detección y seguimiento:** Para cada fotograma, se invocan los módulos previamente descritos: detección de objetos (jugadores, balón, árbitros), clasificación secundaria por equipo, y seguimiento multiobjeto mediante `ByteTrack`. El objetivo aquí no es solo detectar, sino dar continuidad a las identidades a lo largo del tiempo. Gracias a esta integración, obtenemos una estructura sólida que permite saber no solo quién está en el campo, sino también cómo se mueve cada entidad en cada instante.
3. **Proyección espacial a vista aérea:** En fotogramas clave, se activa el módulo de detección de puntos de referencia del terreno de juego, ya explicado en secciones anteriores, que permite calcular la homografía y proyectar todos los elementos del campo a una vista en planta. Para evitar que esta vista sufra de saltos o temblores visuales, se aplica un suavizado temporal que estabiliza la transformación.

4. **Visualización y renderizado de salidas:** Esta es la fase más visible del proceso, donde se plasman todos los datos recogidos. Se generan tres vídeos diferentes:

- **Vídeo anotado con detección y seguimiento:** Sobre el vídeo original, se dibujan las cajas delimitadoras con información como clase, ID y equipo, todo ello codificado por colores. El balón, al ser un elemento especialmente relevante, se destaca con un estilo visual distinto.
- **Vídeo con diagramas de Voronoi:** A partir de las posiciones proyectadas de los jugadores sobre el plano del campo, se calcula un diagrama de Voronoi dinámico [65], que muestra visualmente el espacio dominado por cada jugador. Esta representación ayuda a entender cómo se distribuye el control territorial en cada jugada.
- **Vídeo con vista aérea proyectada:** Aquí se construye una versión limpia del campo de fútbol desde arriba, con las posiciones exactas de jugadores, balón y árbitros. Este tipo de visualización es especialmente útil para un análisis táctico sin distracciones, donde lo importante es la ocupación del espacio y las relaciones posicionales.

Esta representación visual facilita la comprensión de aspectos tácticos, tales como control espacial, posesión y dinámica de juego. La *Figura 5.5* nos da una idea de cuál es el resultado final.

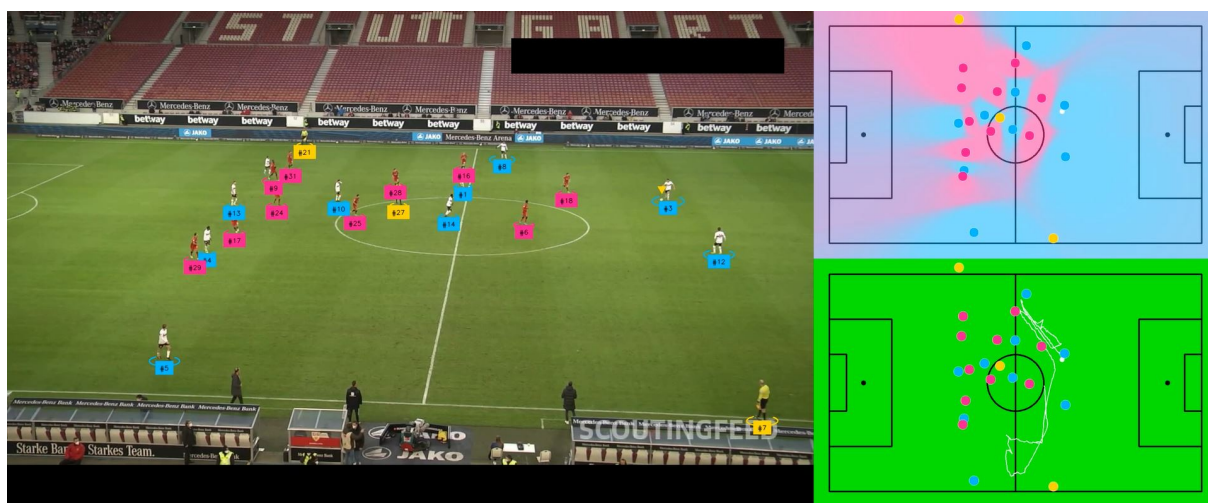


Figura 5.5: **Video Final** tras aplicar la función.

Fuente: Elaboración Propia (Roboflow)

5.10. Reflexión final

Trabajar con estos modelos, módulos y funciones ha sido toda una experiencia. Combina lo mejor de la visión por computadora moderna con un diseño modular y accesible, y deja muy claro que esta tecnología ya no es exclusiva de grandes clubes o empresas tecnológicas. Hoy, cualquier apasionado del fútbol con ganas de aprender puede construir un sistema inteligente que observe el juego desde una nueva perspectiva.

A pesar de las limitaciones como errores en la clasificación de equipos o detecciones fallidas en condiciones de baja calidad, los resultados son prometedores. Este tipo de herramientas puede integrarse en el análisis de partidos amateur, ayudar en entrenamientos formativos o incluso aportar valor en retransmisiones y contenidos interactivos.

Capítulo 6

Aplicación

El desarrollo de esta aplicación nace de una inquietud genuina: acercar el análisis avanzado de datos y la inteligencia artificial al mundo del fútbol profesional, sobre todo, útil para quienes aman este deporte. Y es que detrás de cada pase, cada carrera y cada gol hay montañas de datos esperando ser transformadas en conocimiento que inspire decisiones.

Para lograrlo, ha sido necesario tejer un entramado de metodologías que, en conjunto, dan vida a la aplicación. Desde el tratamiento y limpieza de grandes volúmenes de datos, hasta la implementación de modelos predictivos que, en cierto modo, intentan asomarse al futuro. Además, se han diseñado herramientas interactivas de exploración visual que buscan no solo mostrar números, sino contar historias.

En este apartado se desgranar las estrategias, técnicas y herramientas que han hecho posible convertir datos en algo casi tangible. Se explica, por ejemplo, cómo se decidió abordar el preprocesamiento de datos estadísticos de jugadores y competiciones, la normalización de métricas según la posición de cada futbolista, el uso de percentiles para dar contexto a cada cifra, o los escalados personalizados que permiten comparar realidades distintas. También se detallan los modelos de predicción que tratan de anticipar el rendimiento futuro de los jugadores, un reto que siempre deja lugar a la emoción.

Por otro lado, se habla del diseño de visualizaciones como los gráficos radar y los pizza plots, integrando librerías especializadas como `mplsoccer` [68], y de la construcción de toda la arquitectura de la aplicación en Shiny para Python, un espacio donde el usuario puede perderse, explorar, comparar y, en definitiva, disfrutar.

Porque, en el fondo, este enfoque metodológico busca algo más que precisión técnica: pretende democratizar el acceso a herramientas de análisis y visualización que antes estaban reservadas a unos pocos, y devolver al aficionado, o al analista que comienza, la capacidad de ver el fútbol con otros ojos.

Todo el código utilizado para desarrollar esta aplicación está disponible públicamente en un repositorio de GitHub [69]. En la *Figura 6.1* podemos encontrar cómo se organiza el menú de la aplicación y qué contiene la aplicación a grandes rasgos.

Figura 6.1: **Menú** de la aplicación*Fuente: Elaboración Propia*

6.1. Análisis de Competiciones

En la primera sección de la aplicación encontramos una ventana para el análisis de competiciones, como bien se muestra en la *Figura 6.1*. Esta se divide en dos secciones, la primera para visualizar los eventos del partido y otra sección para visualizar los eventos de jugadores concretos.

El fútbol moderno demanda un enfoque analítico que trascienda la mera observación para adentrarse en el estudio sistemático de los patrones de juego. Este sistema de análisis integral ofrece una metodología estructurada para examinar tanto el desempeño colectivo como individual, sustentado en datos objetivos y visualizaciones interpretativas.

La plataforma se fundamenta en un diseño reactivo que adapta dinámicamente su interfaz según las selecciones del usuario. Al elegir un equipo, se despliegan automáticamente sus partidos en el torneo, ordenados cronológicamente y contextualizados según la fase de la competición, permitiendo acceder a gráficos y estadísticas del equipo a lo largo del partido.

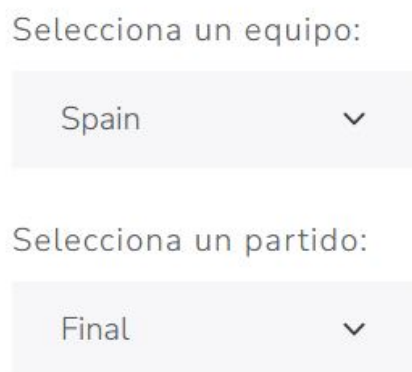
Selección dinámica de partidos y jugadores

Para ofrecer una experiencia interactiva y adaptada al usuario, el sistema implementa un mecanismo reactivo que actualiza las opciones disponibles según las selecciones previas.

- **Filtrado de partidos por equipo:** Cuando el usuario escoge un equipo, se filtran automáticamente todos los partidos en los que dicho equipo ha participado durante el torneo. Estos partidos se ordenan cronológicamente para facilitar la exploración temporal.
- **Asignación contextual de fases:** Cada partido recibe una descripción dinámica que indica la fase del torneo a la que pertenece (fase de grupos, octavos, cuartos, semifinal, final), facilitando así la comprensión del contexto competitivo sin necesidad de referencias externas.
- **Actualización de la selección de partidos:** La lista de partidos se actualiza en la interfaz con las descripciones generadas, permitiendo una navegación intuitiva.
- **Carga y filtrado de jugadores:** Al seleccionar un partido específico, el sistema recupera automáticamente los eventos correspondientes, identificando a los jugadores que participaron. Se extrae la lista única de futbolistas involucrados y se actualiza el selector para que el usuario pueda elegir un jugador particular para análisis detallado.

- **Configuración del equipo para posteriores análisis:** Además, se obtienen y presentan los nombres del equipo local y visitante del partido, facilitando la selección de conjuntos para análisis específicos posteriores.

En la *Figura 6.2* podemos ver que esta metodología reactiva asegura que la navegación sea coherente y fluida, y que el usuario siempre disponga de opciones relevantes en función de su contexto de análisis.



The image shows two vertically stacked dropdown menus. The first menu is titled 'Selecciona un equipo:' and has 'Spain' selected with a downward arrow. The second menu is titled 'Selecciona un partido:' and has 'Final' selected with a downward arrow. Both menus are light gray with a white border.

Figura 6.2: **Selección** de equipos y partidos
Fuente: Elaboración Propia (Datos: Statsbomb)

6.1.1. Análisis de eventos a nivel equipo

Supongamos que nos encontramos en la ventana donde se muestran las estadísticas del equipo. Lo que encontramos es bastante completo: un resumen general del partido, gráficos de tiros, gráficos de pases, la formación y la red promedio de pases, un mapa de recuperaciones y, para rematar, un selector de goles que permite explorarlos en detalle. Comenzaremos hablando acerca del resumen estadístico del partido, el cual se muestra en la *Figura 6.3*.

Resumen estadístico del partido

Cuando el usuario elige el equipo y el partido, la aplicación se pone manos a la obra y genera un resumen estadístico muy detallado. Este resumen compara el rendimiento del equipo seleccionado con el de su rival, ofreciendo así una visión rápida y bastante clara de cómo se desarrolló el encuentro, basándose en las métricas más importantes.

- **Recuperación de eventos:** Gracias al `match_id` asociado al partido, la aplicación carga todos los eventos registrados con la función `parser.event()`. El resultado es un `DataFrame` bien estructurado, que contiene la secuencia completa y cronológica de lo que sucedió en el partido.
- **Identificación automática de rivales:** A partir del equipo seleccionado, la aplicación detecta al rival cruzando la información guardada en el conjunto `euro_2024`.
- **Cálculo de estadísticas clave:** Luego, los eventos se filtran por equipo y se calculan las métricas que realmente cuentan, por ejemplo:

- *Goles totales y goles en prórroga* (considerando desde el minuto 90 en adelante).
 - *Tiros totales y tiros a puerta*, porque no es lo mismo disparar a las nubes que poner el balón entre los tres palos.
 - *Porcentaje de posesión*, que se estima como la proporción del tiempo que cada equipo tuvo el control del balón respecto al total del partido.
 - *Número de pases, faltas, tarjetas (amarillas y rojas) y fuera de juego*, para tener un panorama completo del juego.
- **Estructura del resumen:** Toda esta información se ordena en una tabla comparativa donde, para cada estadística, aparecen los valores correspondientes tanto al equipo elegido como a su adversario. En el caso de los goles, si hubo prórroga, se muestran en formato `Goles(normal) (prórroga)` para no dejar lugar a dudas.
 - **Orden lógico y amigable:** Las estadísticas no aparecen al azar, sino que siguen un orden pensado (`orden_estadisticas`) que prioriza las métricas más determinantes, como goles y tiros, facilitando así la lectura y comprensión rápida.

ESTADÍSTICA	SPAIN	ENGLAND
Goles	2	1
Tiros	16	9
Tiros a puerta	6	3
Posesión (%)	62.75	37.25
Pases	593	324
Faltas	12	7
Tarjetas amarillas	0	0
Tarjetas rojas	0	0
Fueras de juego	0	0

Figura 6.3: **Resumen de estadísticas** del partido España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

En definitiva, este resumen le permite al usuario tener en un solo vistazo una síntesis clara y precisa del desempeño colectivo. Y es que, para analizar tácticas, preparar informes o simplemente seguir el torneo con un poco más de profundidad, esta visión rápida resulta indispensable.

Visualización del xG acumulado y mapas de tiros

Entender cómo se desarrolla un partido de fútbol más allá del marcador no siempre es fácil. Por eso, contar con herramientas que reflejen no solo cuántas ocasiones hubo, sino también la calidad de esas oportunidades, es fundamental. Aquí entran en juego dos visualizaciones que nos cuentan una historia más profunda: el gráfico de xG acumulado y los mapas de tiros de cada equipo.

¿Y qué es exactamente el xG o “goles esperados”?

Como bien se explica en un artículo de *hudlstatsbomb* [70], el xG es una métrica que, a primera vista, puede sonar muy técnica, pero la verdad es que es una nueva forma de entender el rendimiento cara a puerta de un equipo o jugador. Básicamente, nos dice la probabilidad de que un disparo termine en gol, teniendo en cuenta detalles como desde dónde se lanzó el tiro (distancia y ángulo respecto a la portería), qué tipo de jugada fue, o con qué parte del cuerpo impactó el balón. Por ejemplo, un disparo cercano, sin defensas cerca, tendrá un xG alto, porque es mucho más probable que acabe en gol. En cambio, un disparo lejano o muy forzado tendrá un xG bajo. Así, esta cifra nos ayuda a ver qué equipo realmente generó peligro, incluso cuando el marcador parece no contar toda la verdad.

Gráfico de xG acumulado

Este gráfico se despliega minuto a minuto para mostrar cómo cada equipo va acumulando sus oportunidades de gol a lo largo del partido. Refleja con claridad quién generó más peligro real y cuándo ocurrieron los goles, para poder comparar la calidad de las ocasiones con el resultado final.

Además, el gráfico incorpora un filtro muy útil: permite decidir si se quieren incluir o no los penaltis que se producen durante el juego (excluyendo siempre la tanda de penaltis). Y es que un penalti durante el partido, al ser una oportunidad muy clara, puede inflar la sensación de peligro generado por un equipo, cuando en realidad fue un único momento aislado. Así, esta opción ayuda a analizar el desarrollo ofensivo con más precisión y sin dejarse llevar por esos eventos puntuales que podrían distorsionar la lectura general.

Por último, el gráfico se adapta automáticamente para cubrir partidos que se extienden hasta la prórroga, abarcando hasta 120 minutos de juego. En la *Figura 6.4* se nos muestra la funcionalidad del gráfico.

☒ Incluir penaltis en el xG

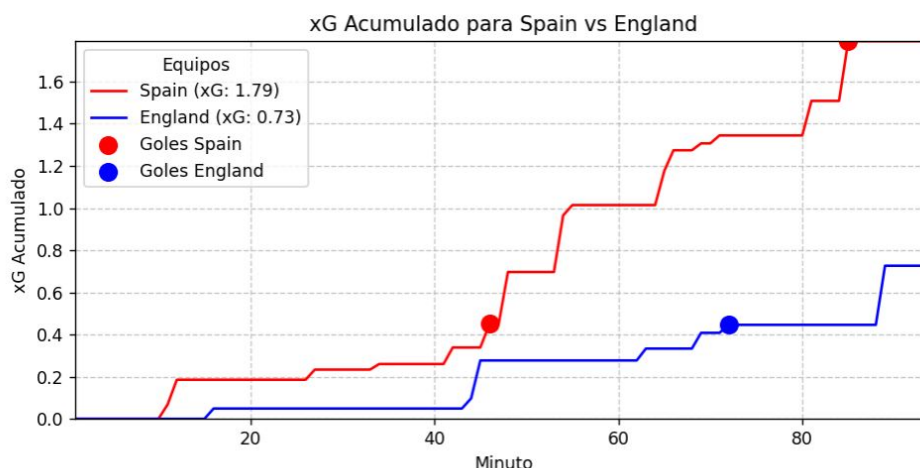


Figura 6.4: **Gráfico del xG acumulado** del partido España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Mapas de tiros por equipo

Por otro lado, los mapas de tiros nos permiten “ver” el campo de juego desde otra perspectiva: cada disparo queda plasmado en su ubicación exacta dentro del campo rival (gracias a las coordenadas x e y con las que cuenta cada disparo), con símbolos que indican si el tiro fue con el pie, la cabeza u otra parte del cuerpo. El tamaño y el borde de cada punto muestran si el disparo acabó en gol o no, y los colores de estos están calibrados para comparar con justicia la calidad de las ocasiones entre los dos equipos. Visualizando los gráficos que muestra la *Figura 6.5*, no solo entendemos dónde se hicieron los tiros, sino también qué tan peligrosos fueron realmente.

Al final, estas dos visualizaciones se complementan para ofrecernos una visualización mucho más enriquecedora y emocionante del partido, algo que va más allá del simple marcador.

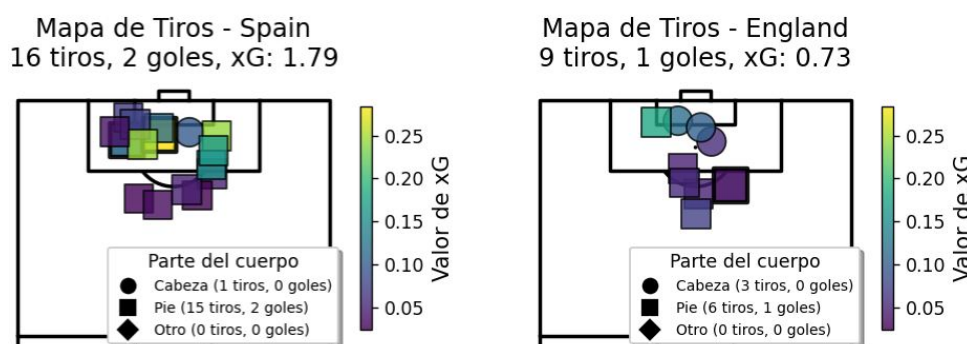


Figura 6.5: **Mapa de tiros por equipo** del partido España-Inglaterra

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas para los gráficos

Para construir estas visualizaciones se utiliza principalmente la librería `matplotlib` junto con `seaborn` para facilitar el manejo estético y funcional de los gráficos. En el caso del gráfico de xG acumulado, se procesan los eventos del partido minuto a minuto, acumulando el xG generado por cada equipo, lo que permite una representación continua y clara del desarrollo ofensivo a lo largo del tiempo. Se implementan filtros específicos para incluir o excluir penaltis, evitando así que momentos puntuales distorsionen la narrativa visual.

En cuanto a los mapas de tiros, se aprovecha la información espacial de los disparos mediante coordenadas que se traducen en posiciones sobre un campo de fútbol a escala real, generado también con `matplotlib`. Se cuida especialmente la diferenciación visual mediante el uso de distintos símbolos, colores y tamaños para reflejar características clave de cada disparo (tipo de disparo, éxito o no, calidad del tiro). Para evitar la saturación y superposición, se ajusta el tamaño y la transparencia de los puntos, asegurando una visualización clara y equilibrada.

Además, se ha prestado atención a la escalabilidad del gráfico para adaptarse a diferentes tamaños de pantalla y resoluciones, así como a la inclusión de leyendas descriptivas que facilitan la interpretación incluso para quienes no estén familiarizados con las métricas avanzadas. En resumen, el enfoque técnico busca que el análisis no solo sea riguroso, sino también accesible y visualmente atractivo, para conectar mejor con el usuario y potenciar la comprensión del juego.

Análisis y visualización de pases

En el fútbol, el pase es mucho más que un simple gesto técnico; es la columna vertebral del juego. Es la manera en que un equipo arma jugadas, crea ocasiones y marca el tempo del partido. Por eso, entender quién distribuye el balón y cómo lo hace nos abre una ventana para apreciar la esencia del juego colectivo.

Para acercarnos a esta realidad, presentamos dos gráficos que reflejan distintos aspectos del rendimiento en pases durante un encuentro.

Top 10 jugadores por porcentaje de pases completados

Este gráfico muestra a aquellos futbolistas que, con al menos 10 pases realizados, destacan por su precisión. Ordenados de mayor a menor según su porcentaje de pases exitosos, nos permiten identificar a los jugadores que aportan confianza y estabilidad en la circulación del balón. Además, los valores aparecen dentro de las barras, facilitando una lectura rápida y directa, sin complicaciones. Podemos ver un ejemplo a continuación con la *Figura 6.6*.

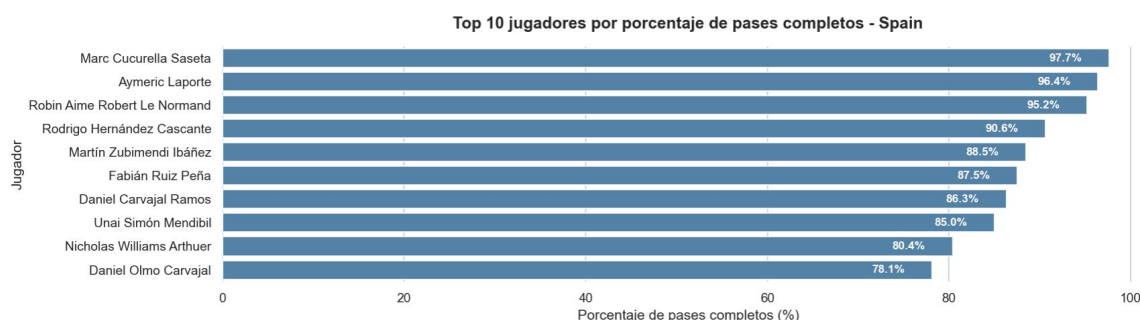


Figura 6.6: **Gráfico del Top 10 pasadores de España por porcentaje de acierto** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Pases completos e incompletos

Como complemento, el segundo gráfico desglosa el volumen total de pases de cada jugador (*Figura 6.7*), diferenciando claramente entre los que fueron completados y los que no. Esta visión apilada aporta una dimensión extra: no solo importa la precisión, sino también cuánto participa cada futbolista en la construcción del juego.

La función encargada de extraer estos datos filtra exclusivamente los pases del equipo seleccionado, clasifica cada pase según su éxito y calcula estadísticas clave para que estas visualizaciones reflejen fielmente el desempeño de los jugadores en esta faceta.

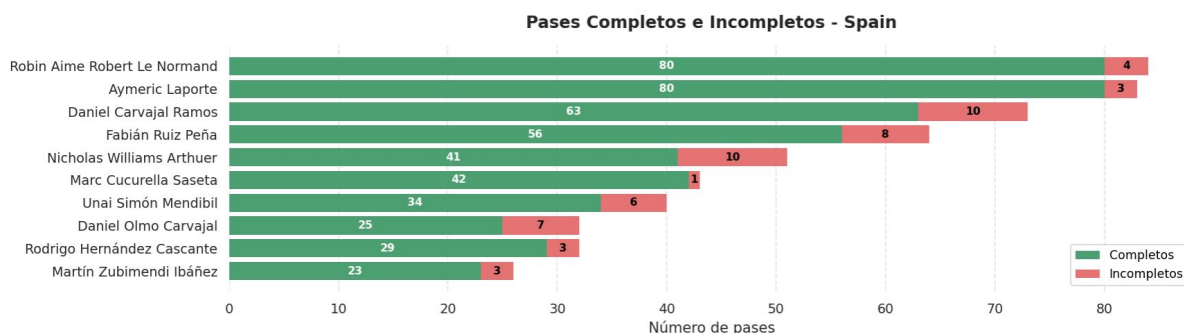


Figura 6.7: **Proporción y número de pases del Top 10 pasadores de España** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas para los gráficos

Desde un punto de vista técnico, estas visualizaciones se generan utilizando principalmente la librería `matplotlib`, que permite crear gráficos de barras personalizados y visualmente claros. Se presta especial atención a la legibilidad, con colores diferenciados para pases completados y fallados, y valores numéricos sobre las barras para que el usuario pueda captar rápido el desempeño sin perderse en detalles.

Los datos se procesan para incluir solo pases relevantes, descartando aquellos sin sentido estadístico o fuera del contexto del partido (por ejemplo, pases de saque de banda o fuera de juego). Esto garantiza que las gráficas reflejen con precisión la dinámica real del

equipo.

Además, se cuida la disposición visual para evitar saturación o solapamiento, ajustando el tamaño y el espaciado de los elementos. Todo ello con la idea de que el análisis sea tan fluido y natural como ver el partido, pero con la ventaja de contar con una mirada analítica que revela lo que a simple vista se escapa.

En definitiva, estos gráficos no solo muestran números, sino que nos permiten sentir el latido del equipo, con su paciencia, su eficacia y el riesgo que asumen en cada pase. Porque, al final del día, el fútbol es un juego de detalles, y cada pase cuenta.

Visualización de la red de pases y análisis espacial

En el fútbol moderno, la comprensión del juego pasa necesariamente por analizar no solo quién pasa el balón, sino también cómo se mueven esos pases en el espacio del campo. Para ello, se utilizan gráficos que representan la red de pases entre jugadores y la ubicación promedio en la que realizan esas conexiones, aportando una visión clara y detallada del control y fluidez del equipo.

En nuestra aplicación, la visualización de la red de pases se construye a partir de los datos de eventos del partido, donde cada pase está registrado con su punto de inicio y final en coordenadas dentro del campo. Estas coordenadas se procesan cuidadosamente para representar fielmente la posición media desde la que cada jugador distribuye y recibe el balón.

Tratamiento y cálculo de las coordenadas

Para cada jugador, se extraen las posiciones de todos los pases realizados y recibidos durante el primer tiempo del partido (minutos 0 a 45), para definir un punto medio que refleje su área de influencia en la circulación del balón, obteniendo el esquema táctico del 11 inicial. Esto implica combinar los valores x e y de los pases que originan y reciben, calculando un promedio que sitúa al jugador en un lugar representativo del campo. Este método nos ayuda a evitar distorsiones debidas a pases muy esporádicos o desplazamientos puntuales.

Construcción de la red de pases

Con las posiciones promedio de los jugadores definidas, se dibujan puntos proporcionales al volumen total de pases realizados por cada uno, lo que destaca la importancia relativa de cada futbolista en el entramado colectivo. Además, se traza una línea entre pares de jugadores que han intercambiado pases en más de dos ocasiones, con un grosor proporcional a la frecuencia de dichos pases. En la *Figura 6.8* podemos ver cómo esta representación gráfica nos permite identificar fácilmente los dúos con mayor conexión y el peso de cada vínculo en el sistema de juego.

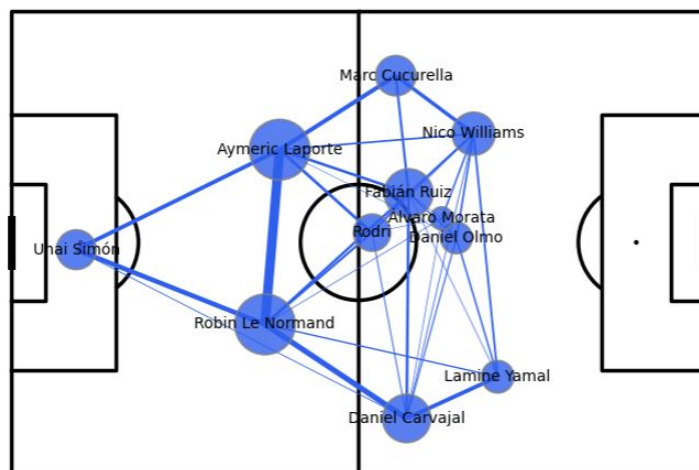


Figura 6.8: **Red de pases y posición promedio de España** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Visualización de todos los pases del equipo

Junto a la red de pases, se genera un gráfico complementario que muestra todos los pases realizados por el equipo durante el primer tiempo, independientemente de quién los ejecuta. En esta visualización, cada pase se representa con una flecha sobre el campo que indica el punto de inicio y de finalización del pase. Además, se destacan visualmente aquellos pases que avanzan hacia el último tercio del campo rival, resaltando la capacidad ofensiva del equipo.

Para garantizar la fiabilidad del análisis, se aplican filtros que eliminan pases irrelevantes o fallidos, asegurando que solo se visualicen los movimientos que reflejan un patrón real de circulación del balón. Como en la red de pases, la orientación, escala y proporción del campo se mantienen constantes para permitir comparaciones coherentes entre diferentes partidos o equipos.

El gráfico de la *Figura 6.9* facilita un análisis más profundo de la estrategia de posesión y la construcción de jugadas del equipo, permitiendo observar tanto la cantidad como la dirección de los pases.

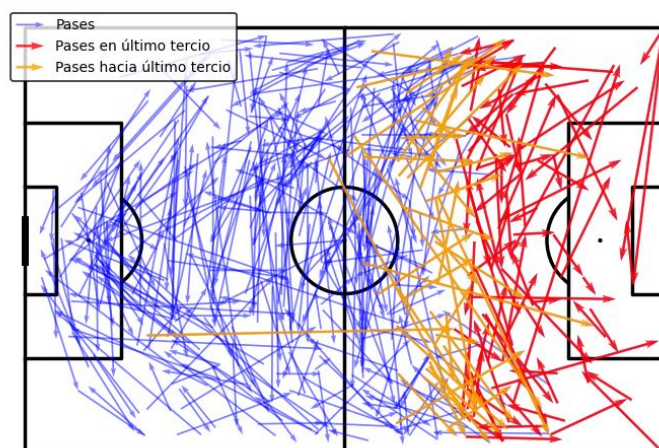


Figura 6.9: **Pases totales de España** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas para los gráficos

Para crear estas visualizaciones se emplea la librería `mplsoccer`, que facilita la creación de un campo de fútbol a escala real, respetando las dimensiones oficiales y permitiendo un control preciso sobre el diseño y los elementos gráficos. El manejo de las coordenadas es esencial para que tanto la red de pases como el gráfico de todos los pases sean coherentes y fáciles de interpretar.

En el caso de la red de pases, se verifica que las posiciones promedio calculadas para cada jugador se encuentren dentro del área del campo, ajustando si es necesario para evitar inconsistencias o distorsiones en la representación táctica. Para los pases totales, también se asegura que las coordenadas de inicio y finalización estén bien registradas y dentro del rango esperado, permitiendo una visualización fiel de las trayectorias de balón.

Se aplican filtros previos para eliminar pases fallidos, irrelevantes o fuera del rango temporal definido (minutos 0 a 45), asegurando que ambas visualizaciones reflejen únicamente patrones de circulación válidos y significativos.

El tamaño de los elementos gráficos, como los marcadores de los jugadores, las líneas de pase o las flechas direccionales, se ajusta dinámicamente en función del volumen de datos representado, manteniendo una proporción visual adecuada. Esto incluye el grosor de las líneas entre jugadores en la red de pases y la intensidad visual de los pases progresivos en el gráfico de trayectorias.

Además, se incorpora un sistema de *alias* para sustituir los nombres reales de los jugadores por versiones abreviadas o adaptadas, lo que mejora la legibilidad del gráfico y evita que las etiquetas largas saturen la visualización. Esta asignación de alias se realiza de forma automática mediante una lógica de búsqueda y reemplazo consistente en todos los partidos, garantizando claridad en la presentación de los datos.

Por último, se mantiene una configuración gráfica uniforme para todos los partidos y equipos, lo que permite realizar comparaciones entre visualizaciones sin que las diferencias de escala, orientación o diseño interfieran en el análisis táctico.

Mapa de calor y distribución por zona de recuperaciones

El momento en que un equipo recupera el balón puede cambiar por completo el rumbo de un partido. Estas acciones defensivas no solo detienen ataques rivales, sino que muchas veces marcan el inicio de una jugada ofensiva. Por ello, visualizar dónde y con qué frecuencia ocurren estas recuperaciones puede ofrecer una perspectiva clave sobre la organización táctica y el posicionamiento defensivo de un equipo.

Para ilustrar este aspecto, se han desarrollado dos visualizaciones que se complementan: un mapa de calor de recuperaciones sobre el campo de juego y un gráfico circular que resume la proporción de acciones en zonas ofensivas y defensivas.

Mapa de calor de recuperaciones

Esta visualización muestra de forma intuitiva las zonas del campo donde el equipo recupera más balones. Utilizando las coordenadas proporcionadas por los datos de *StatsBomb*, se representa un mapa de calor que indica la densidad de recuperaciones, con tonos más intensos en las áreas de mayor frecuencia.

Cada recuperación cuenta con una coordenada x e y que refleja el punto exacto dentro del campo donde se produce la acción, medida respecto al sistema de referencia de *StatsBomb*, donde x varía de 0 a 120 (a lo largo del campo) y y de 0 a 80 (a lo ancho). Esto permite graficar la posición real de cada recuperación y, al combinar todas ellas mediante técnicas de suavizado de densidad, obtener un panorama general de las zonas de mayor actividad defensiva.

Sobre el campo también se grafican puntos individuales que indican cada recuperación específica, permitiendo ver la distribución exacta y no solo el agregado general, como se muestra en la *Figura 6.10*. Esta combinación ofrece una lectura tanto cuantitativa como espacial del comportamiento defensivo del equipo.

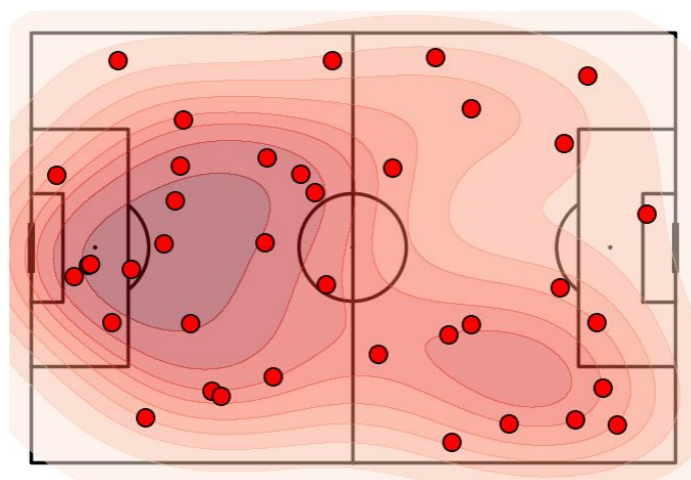


Figura 6.10: **Mapa de recuperaciones de España** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Distribución zonal (pie chart)

Para simplificar e interpretar rápidamente la orientación táctica del equipo, se divide el campo en dos mitades: zona ofensiva y zona defensiva. En este caso, se considera el valor de x para determinar la zona: aquellas recuperaciones con x mayor a 60 se etiquetan como recuperaciones en zona ofensiva, mientras que las demás corresponden a la mitad defensiva.

El gráfico circular, *Figura 6.11*, resume la proporción de recuperaciones en cada mitad del campo. Además, se incorporan etiquetas personalizadas que muestran tanto el porcentaje como el número total de acciones en cada zona, facilitando así una comprensión rápida del balance defensivo del equipo.



Figura 6.11: **Distribucion por zona de las recuperaciones de España** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas para los gráficos

Ambos gráficos se generan combinando herramientas especializadas en visualización deportiva y análisis estadístico. El mapa de calor se construye sobre una maqueta de campo usando la clase `Pitch` de `mlsoccer`, mientras que la densidad se calcula mediante `seaborn.kdeplot`, que permite generar gradientes suaves y visualmente intuitivos.

Cada punto de recuperación se añade al gráfico mediante la función `scatter` para resaltar eventos individuales, usando colores contrastantes para una mejor legibilidad. Las coordenadas x e y provienen directamente de la columna `location` de los datos de *StatsBomb*, procesadas para separar los valores en columnas independientes y facilitar su uso.

Por otro lado, el gráfico circular se genera con `matplotlib`, incorporando etiquetas personalizadas mediante una función auxiliar que calcula tanto el porcentaje como la cantidad absoluta de recuperaciones por zona. Los colores y bordes de cada sección están cuidadosamente elegidos para garantizar claridad visual y distinción entre categorías.

Antes de visualizar, los datos pasan por un filtrado específico para asegurarse de que solo se consideren las recuperaciones del equipo seleccionado en el partido correspondiente, descartando cualquier ruido o acciones no relevantes.

En conjunto, estas visualizaciones permiten interpretar de forma clara y visual cómo

y dónde el equipo impone su presencia defensiva, transformando datos numéricos en una narrativa táctica comprensible.

Selector y visualización de goles

En el fútbol, cada gol representa mucho más que un simple número en el marcador. Este módulo ha sido diseñado precisamente para capturar y explorar esos momentos de forma visual, detallada y personalizada.

El funcionamiento comienza con un efecto reactivo que permanece atento a cualquier cambio en la selección de partido o equipo. Cuando el usuario realiza una nueva elección, el sistema filtra los datos de *StatsBomb*, localizando únicamente aquellos eventos de tipo **Shot** cuyo resultado fue **Goal** para el equipo seleccionado, gracias a la columna `outcome_name`. Este filtrado, aunque aparentemente discreto, resulta fundamental para garantizar la coherencia y la relevancia de la información mostrada.

En caso de que no existan goles para el equipo elegido, el sistema informa de manera clara y directa. Si los hay, genera dinámicamente un selector que permite al usuario escoger cuál desea analizar en mayor profundidad. Este proceso abre la puerta a una exploración visual precisa de cada acción de gol.

La visualización se lleva a cabo mediante la función `plot_shot_freeze_frame`, que produce un “freeze frame” del disparo. En esta imagen estática se muestra la disposición exacta de todos los jugadores en el instante previo al gol, combinando los datos de alineación (`lineup`) y de congelamiento (`freeze frame`). Cada jugador se ubica en el campo según su posición real en ese momento, ofreciendo al usuario la posibilidad de comprender el contexto táctico de la jugada. Podemos entenderlo de una manera más visual con la *Figura 6.12*.



Selecciona un gol para ver:

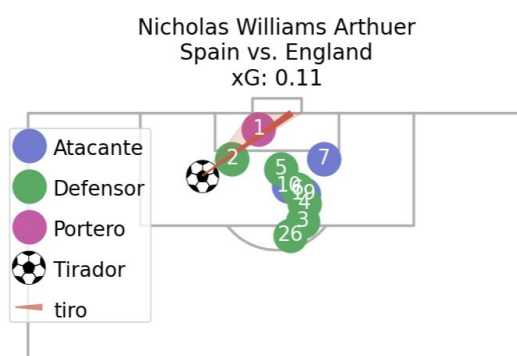
Gol 1 

Figura 6.12: Gol de *Nico Williams* en la final de la *Eurocopa 2024*, España-Inglaterra.
Fuente: *Elaboración Propia (Datos: Statsbomb)*

Consideraciones técnicas del gráfico.

Este módulo se estructura en torno a tres componentes principales:

- Un efecto reactivo que detecta cambios de entrada y actualiza el selector de goles utilizando `get_shot_ids_for_team()`, encargado de devolver los identificadores de disparos convertidos en gol por el equipo seleccionado.
- Una función de interfaz de usuario (`selector_goles()`) que construye y actualiza las opciones del selector en función del partido y equipo seleccionados.
- Una función de visualización (`grafico_goles()`) que genera el gráfico personalizado o muestra un mensaje adecuado en caso de no haber goles disponibles.

La lógica reactiva y la comunicación entre componentes se gestionan mediante la librería `shiny` para Python. La parte gráfica se basa en `mplsoccer`, partiendo de uno de sus ejemplos oficiales de visualización de “freeze frames”, pero adaptándolo para dotarlo de interactividad y conexión directa con las entradas del usuario. Además, se emplean otras librerías de apoyo como `pandas` para la manipulación de datos, `numpy` para operaciones numéricas y `matplotlib` para la generación final de gráficos.

En conjunto, este módulo transforma datos de eventos en representaciones visuales detalladas, ofreciendo un recurso riguroso y eficaz para el análisis de goles dentro del flujo de un partido.

Con esto, cerramos la parte dedicada a la visualización de eventos a nivel de equipo. Ahora, toca cambiar el foco. Vamos a sumergirnos en un análisis más concreto: el del jugador individual.

6.1.2. Análisis de eventos a nivel jugador

En esta sección ponemos la lupa sobre el rendimiento de un jugador específico, explorando cada acción con detalle para entender no solo lo que hizo, sino cómo y dónde lo hizo. Veremos sus pases, sus conducciones, sus disparos y hasta su influencia en el campo a través del mapa de calor. Porque, al final, el fútbol también se construye en esos pequeños gestos individuales que, combinados, terminan decidiendo partidos. Y es que cada pase cuenta una historia, cada conducción revela intenciones y cada disparo guarda la emoción de lo imprevisible.

Extracción y organización dinámica de jugadores por equipo

Para identificar con precisión a los jugadores que participaron en un partido concreto, se extraen inicialmente los nombres de los equipos local y visitante utilizando el identificador único del encuentro dentro del conjunto de datos de la Eurocopa 2024. A partir de esta información, se filtran los eventos correspondientes al partido para obtener la lista exclusiva de futbolistas que intervinieron en cada equipo, descartando valores nulos o datos inconsistentes que podrían afectar la calidad del análisis.

Estos nombres se organizan en un diccionario que agrupa a los jugadores según su equipo, lo que permite una actualización dinámica y eficiente del selector de jugadores en la interfaz de la aplicación. De este modo, el usuario puede seleccionar de manera ágil y precisa al futbolista deseado para realizar un análisis detallado de sus acciones y desempeño durante el encuentro.

Este proceso no solo aporta claridad y orden, sino que también garantiza que el análisis se centre exclusivamente en los protagonistas reales del partido, mejorando así la exactitud de los resultados y facilitando una experiencia de usuario más fluida y satisfactoria. Podemos echar un vistazo a la selección de jugador si prestamos atención a la *Figura 6.13*.

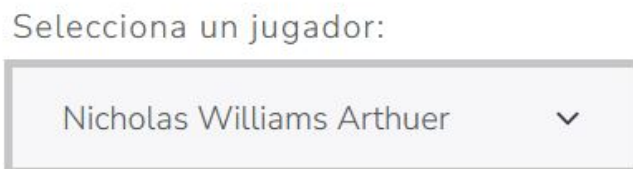


Figura 6.13: **Selección de Jugadores**
Fuente: Elaboración Propia (Datos: Statsbomb)

Visualización de pases por jugador

Para analizar con detalle la contribución de cada futbolista en la circulación del balón, se implementa una visualización que representa todos los pases realizados por un jugador específico durante el partido seleccionado. Esta función toma como punto de partida los eventos registrados en el encuentro y filtra exclusivamente aquellos que corresponden a pases ejecutados por el jugador de interés, excluyendo acciones como los saques de banda para focalizarse en jugadas de juego abierto.

En la práctica, se extraen las coordenadas de inicio y fin de cada pase, procesándolas para asegurar que las posiciones estén correctamente interpretadas y representadas sobre el campo. Esto es fundamental, ya que en ocasiones los datos originales contienen estructuras complejas que requieren un manejo cuidadoso para evitar errores en la visualización.

Para facilitar la comprensión y resaltar aspectos tácticos relevantes, los pases se codifican mediante colores según su localización y dirección: en azul se muestran todos los pases, en rojo aquellos que se realizan dentro del último tercio ofensivo y en naranja los pases que avanzan hacia esa zona clave del campo (como en la visualización de pases totales dado el equipo). Este detalle aporta una lectura más profunda sobre el rol y la forma de jugar del jugador, como podemos ver en la *Figura 6.14*, siendo **Nico Williams** un extremo muy ofensivo y abierto a banda.

De este modo, esta visualización no solo aporta datos numéricos, sino que invita a interpretar el juego desde la perspectiva individual, conectando la técnica del jugador con la estrategia colectiva de su equipo.



Figura 6.14: **Pases de Nico Williams** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas

Para la creación de estos gráficos se utiliza la biblioteca `mlsoccer`, que facilita el trazado de campos de fútbol a escala real y permite un control exhaustivo sobre los elementos gráficos. Es especialmente útil para representar trayectorias y posiciones con precisión, respetando las dimensiones oficiales del terreno de juego.

Un reto importante en este proceso es el manejo de las coordenadas de los pases, ya que algunos registros contienen valores en formatos complejos (como listas o arrays), que deben ser convertidos correctamente a valores numéricos simples para evitar errores visuales. Por ello, se aplican funciones específicas que extraen el primer elemento válido de cada coordenada, garantizando que la visualización sea fiel a los datos reales.

Finalmente, se considera la experiencia del usuario: cuando no existen pases registrados para un jugador en particular, la aplicación no muestra un gráfico vacío, sino un mensaje claro y directo que informa de esta situación, evitando confusión o malinterpretación.

Este equilibrio entre precisión técnica y usabilidad es clave para ofrecer una herramienta que realmente conecte con el análisis táctico y facilite la interpretación del rendimiento individual sobre el campo.

Mapa de calor

Para profundizar en la distribución espacial de los pases de un jugador durante un partido, se implementa un mapa de calor que refleja las zonas del campo donde dicho futbolista ha tenido mayor protagonismo en la circulación del balón. Esta visualización va más allá de simplemente contar pases; busca mostrar dónde se concentra la actividad del jugador, ofreciendo así una perspectiva más táctil y visual de su influencia en el juego.

La función filtra los eventos de pase del jugador seleccionado, excluyendo aquellos con resultados negativos o que corresponden a saques de banda, para garantizar que el mapa refleje solo acciones efectivas y relevantes dentro del juego abierto. Posteriormente, se eliminan datos incompletos o inconsistentes, asegurando la calidad y fiabilidad del

análisis.

Con esta base, se genera el mapa de calor que permite identificar las áreas con mayor concentración de actividad. Se utiliza una escala de colores rojizos que va de tonos suaves a intensos, facilitando la interpretación visual intuitiva y atractiva. En la *Figura 6.15* encontramos un ejemplo de este gráfico.

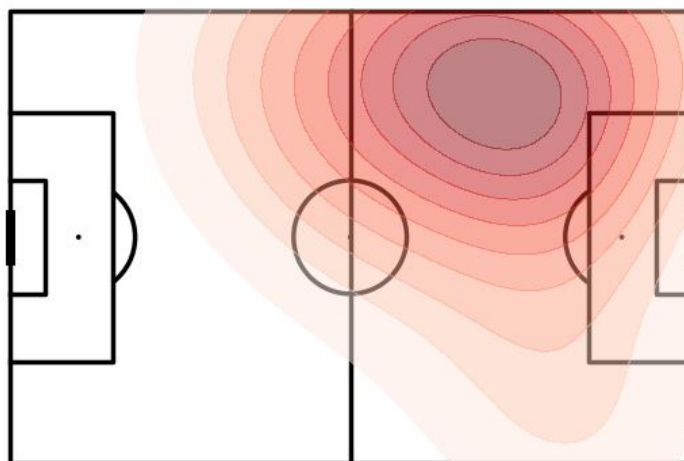


Figura 6.15: Mapa de Calor de *Nico Williams* en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas

La creación de este mapa de calor se realiza con la biblioteca `mplsoccer`, que ofrece herramientas especializadas para trazar campos de fútbol y representar datos espaciales de forma precisa y visualmente atractiva.

Se emplea un estimador de densidad de núcleo `seaborn.kdeplot` para suavizar los datos puntuales de los pases, lo que permite identificar zonas con mayor concentración de actividad. Para mejorar la legibilidad y evitar posibles errores derivados de conjuntos de datos pequeños o poco distribuidos, se ajusta el parámetro de suavizado (`bw_adjust`) y se establecen niveles explícitos para la representación cromática.

Un aspecto fundamental en la implementación técnica es el manejo cuidadoso de la cantidad y calidad de los datos: cuando el número de pases es muy bajo, la visualización puede no resultar representativa o incluso fallar al intentar generarse. Por ello, se incluye un control que verifica que existan suficientes datos para crear el mapa, y en caso contrario se informa claramente al usuario para evitar confusiones.

Visualización de conducciones por jugador

Para comprender mejor el comportamiento de un jugador con el balón en los pies, se ha desarrollado una visualización que representa todas las conducciones realizadas durante un partido. En este contexto, una conducción es cualquier acción en la que el futbolista avanza con el balón controlado, sin importar si el movimiento es hacia adelante, lateral o incluso hacia atrás.

El proceso consiste en filtrar los eventos correspondientes a las conducciones de un jugador específico, obteniendo así la posición inicial y final de cada acción. Estas trayectorias se plasman mediante flechas sobre un campo de fútbol esquematizado, permitiendo observar con claridad los desplazamientos realizados.

Este gráfico aporta una perspectiva espacial esencial para entender cómo un jugador se mueve y contribuye en el juego, permitiendo identificar patrones y estilos de conducción que pueden ser clave en el análisis táctico, como se muestra en la *Figura 6.16* estaríamos hablando de un jugador con un perfil muy ofensivo y al que le gusta encarar.

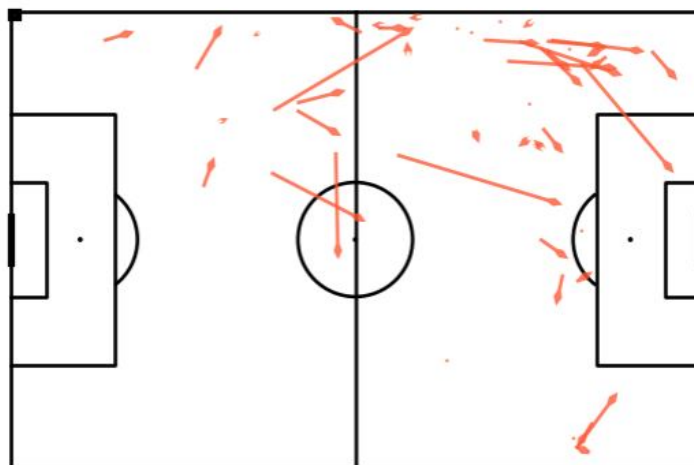


Figura 6.16: **Conducciones de Nico Williams** en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas

La representación gráfica se realiza utilizando la biblioteca `mlsoccer`, que ofrece herramientas especializadas para la visualización futbolística, facilitando una precisión y estética adecuadas.

Las coordenadas de inicio y fin de cada conducción se extraen y adaptan para asegurar su compatibilidad con las funciones de dibujo, incluso cuando los datos provienen en formatos complejos como listas o tuplas.

Para equilibrar legibilidad y sutileza, las flechas se dibujan con un color anaranjado suave y un grosor moderado, acompañadas de una transparencia que permite distinguir zonas de mayor concentración sin saturar la imagen.

En caso de que no se registren conducciones para el jugador en el encuentro seleccionado, la función lo comunica explícitamente, manteniendo la transparencia y claridad en el análisis.

Mapa de tiros por jugador

Para analizar la capacidad ofensiva y las características de finalización de un jugador, se ha desarrollado un mapa de tiros que visualiza todas las acciones de disparo realizadas durante un partido, excluyendo los penales en tanda de penales. Esta representación

espacial permite observar desde qué zonas del campo se originan los disparos, la calidad esperada de gol (xG) de cada intento y el resultado final. Se trata de la función ya usada para visualizar el mapa de tiros de cada equipo adaptada a un único jugador.

Cada disparo se marca con un símbolo que varía según la parte del cuerpo utilizada (pie, cabeza u otro), diferenciándose entre tiros que terminaron en gol y aquellos que no, mediante el tamaño, borde y transparencia del marcador.

Además, el valor de xG de cada disparo se representa mediante un gradiente de color que ayuda a identificar rápidamente la peligrosidad relativa de cada intento. Para facilitar la interpretación, se incluye una leyenda con los símbolos usados y un recuento de tiros y goles por cada parte del cuerpo.

El título del gráfico incorpora información relevante, incluyendo el nombre del jugador, el equipo al que pertenece, el total de tiros efectuados, los goles anotados y la suma del valor esperado de gol, proporcionando un resumen compacto y claro del desempeño ofensivo. Podemos ver un ejemplo de este gráfico en la *Figura 6.17*

Con esto daríamos por finalizada la sección de **Análisis de Competiciones** dentro de la aplicación e iniciaríamos con la segunda sección, **Análisis de Jugadores**.

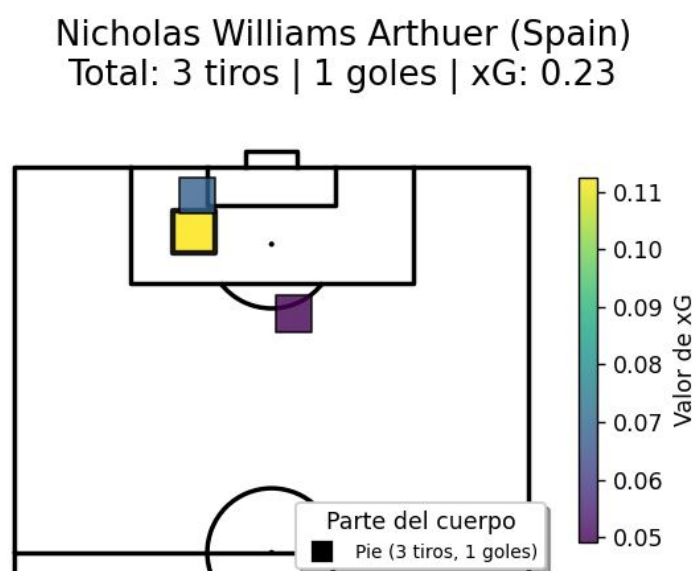


Figura 6.17: Mapa de tiros de *Nico Williams* en la final de la *Eurocopa 2024*, España-Inglaterra.

Fuente: Elaboración Propia (Datos: Statsbomb)

Consideraciones técnicas

El mapa se genera utilizando la biblioteca `mplsoccer`, que permite representar campos de fútbol y superponer datos de eventos con alta precisión.

Las coordenadas de los disparos se extraen directamente de los eventos, desglosando las posiciones en ejes x e y para su correcta ubicación en el campo.

Se emplea un esquema de colores basado en la métrica de xG, utilizando un mapa de colores `viridis` para mejorar la perceptibilidad y accesibilidad visual.

Los marcadores se personalizan según la parte del cuerpo utilizada para el disparo,

asignando formas específicas (círculo, cuadrado, diamante) para facilitar la distinción visual.

La leyenda incluye tanto los símbolos como el conteo de tiros y goles, integrando información cuantitativa y cualitativa.

En caso de que el jugador no haya realizado ningún disparo durante el partido, la visualización muestra un mensaje explícito indicando esta circunstancia para mantener la claridad en el análisis.

6.2. Análisis de Jugadores

Entramos ahora en la segunda gran sección de la aplicación, como bien muestra el esquema de la *Figura 6.1*, en este caso se trata de una ventana centrada en el análisis individual de jugadores.

En el fútbol actual, el análisis del rendimiento individual se ha vuelto una herramienta esencial para evaluar a los jugadores de manera objetiva y con mayor profundidad. La verdad es que ya no basta con mirar solo goles o asistencias; ahora queremos entender el juego desde muchos ángulos, y para eso esta sección se propone una metodología que combina técnicas estadísticas avanzadas con visualizaciones interactivas como *radar plots*, *pizza plots*, etc. Haciendo posible una evaluación mucho más completa del desempeño futbolístico.

Cabe destacar que esta parte de la aplicación toma como referencia la filosofía visual de plataformas como *DataMB* [71], que han contribuido a popularizar este tipo de representaciones gráficas. No obstante, las visualizaciones, las métricas seleccionadas y la metodología implementada han sido desarrolladas de manera original y adaptadas específicamente al enfoque de este proyecto.

6.2.1. Top Jugadores

Comenzaremos la primera ventana de navegación dentro de esta sección de la aplicación, llamada **Top Jugadores**, la cual se basa en tablas en las que visualizar jugadores en base a una estadística concreta.

Visualización de los mejores Jugadores

En este apartado se muestran tablas dinámicas con los **Top 10 goleadores**, **Top 10 asistentes** y **Top 10 porteros** según las estadísticas filtradas. Los usuarios pueden personalizar la vista seleccionando liga, posición y rango de edad, lo que permite enfocar el análisis en diferentes grupos de jugadores.

Las tablas se actualizan automáticamente según los filtros, ordenando a los jugadores por sus principales métricas ofensivas o defensivas, y mostrando indicadores relevantes como goles, asistencias, goles esperados (xG), paradas o diferencia de goles esperados para porteros. Esta presentación facilita la identificación rápida de los jugadores más destacados en cada categoría. Podemos encontrar el selector de filtros en la *Figura 6.18* y la tabla de goleadores correspondiente en la *Figura 6.19*.

Selecciona liga:

Todas



Filtrar por posición:

Todos



Filtrar por edad

Todos



Figura 6.18: Filtros aplicables a las tablas.

Fuente: Elaboración Propia

Top 10 Goleadores							
JUGADOR	EQUIPO	GOLES	XG	G-XG	XG NO PENAL	GOLES SIN PENALTIS	GOLES/90
Kylian Mbappé	Real Madrid	31	25.9	5.1	18.6	24	0.96
Mohamed Salah	Liverpool	29	25.2	3.8	18.2	20	0.77
Robert Lewandowski	Barcelona	27	27.1	-0.1	24	24	0.91
Harry Kane	Bayern Munich	26	20.3	5.7	13.3	17	0.98
Mateo Retegui	Atalanta	25	18.9	6.1	14.9	21	0.94
Alexander Isak	Newcastle Utd	23	20.3	2.7	17.2	19	0.75
Omar Marmoush	Manchester City	22	14.3	7.7	12.8	20	0.73
Erling Haaland	Manchester City	22	22	0	18.8	19	0.72
Ante Budimir	Osasuna	21	18.3	2.7	12.3	13	0.64
Serhou Guirassy	Dortmund	21	22.7	-1.7	19.6	18	0.73

Figura 6.19: Tabla de Top Goleadores.

Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

Se utiliza un dataframe base (`df_jugadores_filtered`) (al que, como se explicó en el capítulo anterior, se le ha aplicado una limpieza y procesado) que contiene los datos estadísticos y demográficos de los jugadores.

Los filtros aplicados son:

- **Liga:** se puede filtrar por competición o mostrar todas.
- **Posición:** se traduce la selección del usuario a un conjunto de códigos estándar de posición (por ejemplo, “Delantero” mapea a [FW, ST, CF, ...]). Se filtra usando la primera posición listada para cada jugador.
- **Edad:** permite limitar el rango a jugadores menores de 25, 23, 21 o 19 años.

Para goleadores y asistentes, tras filtrar, se ordena el dataframe por goles (**Gls**) o asistencias (**Ast**) y se seleccionan los primeros 10 jugadores.

Para porteros, el filtrado fuerza la posición GK y la ordenación es por PSxG+/- o, si no está disponible, por número de paradas (Saves). Se calcula el porcentaje de paradas si no existe en los datos.

Las tablas muestran sólo columnas relevantes para cada tipo de jugador, renombradas para mejorar su comprensión y con valores redondeados para facilitar la lectura.

El diseño es flexible para adaptarse a cambios en las columnas o a la disponibilidad de datos, garantizando robustez y usabilidad.

Predicciones

Una de las funcionalidades más emocionantes y reveladoras de la aplicación es la capacidad de anticipar lo que podría suceder en la próxima temporada. Porque sí, por muy interesante que resulte analizar lo que ya ocurrió, hay algo especialmente estimulante en tratar de imaginar lo que está por venir. ¿Quién liderará la tabla de goleadores? ¿Qué jugador se convertirá en el mayor asistente? A través de modelos de machine learning explicados en el **Capítulo 5** entrenados con datos históricos, se generan predicciones individuales que, como muestra la *Figura 6.20*, alimentan dos tablas comparativas: una centrada en goles y otra en asistencias.

JUGADOR	GOLES TEMPORADA ACTUAL	GOLES TEMPORADA PRÓXIMA	JUGADOR	ASISTENCIAS TEMPORADA ACTUAL	ASISTENCIAS TEMPORADA PRÓXIMA
Kylian Mbappé <i>Real Madrid</i>	31	27	Mohamed Salah <i>Liverpool</i>	18	14
Harry Kane <i>Bayern Munich</i>	26	24	Lamine Yamal <i>Barcelona</i>	13	10
Mohamed Salah <i>Liverpool</i>	29	21	Florian Wirtz <i>Leverkusen</i>	12	9
Serhou Guirassy <i>Dortmund</i>	21	20	Bruno Fernandes <i>Manchester Utd</i>	10	9
Erling Haaland <i>Manchester City</i>	22	19	Bradley Barcola <i>Paris S-G</i>	10	8
Alexander Sørloth <i>Atlético Madrid</i>	20	17	Julian Brandt <i>Dortmund</i>	10	8
Alexander Isak <i>Newcastle Utd</i>	23	17	Bukayo Saka <i>Arsenal</i>	10	8
Ollie Watkins <i>Aston Villa</i>	16	16	Vincenzo Grifo <i>Freiburg</i>	11	8
Ante Budimir <i>Osasuna</i>	21	16	Pascal Groß <i>Dortmund</i>	9	8
Lois Openda <i>RB Leipzig</i>	9	15	Jacob Murphy <i>Newcastle Utd</i>	12	8

Figura 6.20: **Tablas de Top Goleadores y Asistentes**, temporada 25/26.

Fuente: Elaboración Propia (Datos: Fbref)

Estas tablas no sólo muestran cifras frías, sino que permiten observar la evolución potencial de cada jugador. Al contrastar el rendimiento actual con las proyecciones futuras, se facilita una lectura más contextualizada, más útil y, por qué no, también más humana. Al fin y al cabo, el fútbol es un deporte de historias, trayectorias y expectativas.

Consideraciones técnicas

Para la generación de estas tablas se parte de los DataFrames `top_goleadores_pred` y `top_asistentes_pred`, que contienen tanto las estadísticas reales de la temporada actual como las predicciones para la siguiente.

Antes de mostrarlas al usuario, se aplican los siguientes pasos de transformación:

- **Conversión a valores enteros:** Las predicciones generadas por el modelo pueden contener decimales, pero se redondean a enteros para facilitar su lectura y evitar interpretaciones erróneas.
- **Formateo del nombre del jugador:** Se presenta una columna combinada que incluye el nombre en negrita y, justo debajo, el equipo del jugador en color gris y estilo cursivo. Este pequeño detalle mejora la estética y facilita la identificación rápida del contexto de cada futbolista.
- **Selección de columnas clave:** Se conservan únicamente tres columnas: nombre del jugador (y su equipo), goles o asistencias en la temporada actual, y predicción para la próxima. Esta elección busca reducir el ruido visual y centrar la atención en la comparación esencial.
- **Conversión a tabla HTML con estilo personalizado:** El DataFrame se transforma en una tabla HTML con una hoja de estilos propia. Se definen tipografías, colores de fondo, bordes y espaciado para lograr una visualización limpia, uniforme y agradable.

El resultado final es una tabla clara, con diseño responsivo y orientada a la comparación directa. Además, su estructura es lo suficientemente flexible como para adaptarse a futuros cambios de datos o columnas, sin comprometer la funcionalidad. Esta capacidad predictiva no pretende ser una certeza, sino una herramienta más para el análisis y la exploración, apoyada en el rigor de los datos y la intuición que siempre acompaña al fútbol.

Rankings y exploración de métricas individuales

En este apartado se construyó un panel que permite visualizar de forma interactiva los jugadores con mejores valores en distintas métricas específicas (goles, asistencias, pases, acciones defensivas, estadísticas de portero, etc.), o bien consultar un índice de rendimiento agregado; se trata de una versión más completa de los Tops anteriores, pero solo permitiendo visualizar una estadística.

No se trata solo de calcular un índice único, sino de habilitar la exploración de rankings según la métrica seleccionada, ofreciendo así una herramienta flexible para comparar rendimiento en diferentes aspectos del juego. Además, se añadió la posibilidad de ajustar el número de jugadores mostrados (por ejemplo, top 5, top 15, top 50), así como filtrar por posición, liga o edad máxima. Permitiendo de esta forma poder encontrar a los jugadores más determinantes en cada métrica, como bien se muestra en la *Figura 6.21*.

Con esto daríamos por finalizada la ventana **Top Jugadores** .

Seleccionar Métrica: Índice de Rendimiento Número de jugadores a mostrar: 10

JUGADOR	EQUIPO	POSICIÓN	EDAD	ÍNDICE DE RENDIMIENTO
Ousmane Dembélé	Paris S-G	FW	27	94.97
Michael Olise	Bayern Munich	FW	22	94.91
Mohamed Salah	Liverpool	FW	32	93.49
Raphinha	Barcelona	FW	27	93.01
Florian Wirtz	Leverkusen	MF	21	90.67
Bradley Barcola	Paris S-G	FW	21	90.19
Yehvann Diouf	Reims	GK	24	90.17
Leroy Sané	Bayern Munich	FW	28	89.15
Yahia Fofana	Angers	GK	23	89.14
Pedri	Barcelona	MF	21	88.6

Figura 6.21: **Rankings** en base al Índice de Rendimiento.

Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

Para construir los rankings se partió del dataframe filtrado según la posición, la liga y el rango de edad seleccionados. Posteriormente, se ordenó el dataframe por la métrica elegida y se mostraron los primeros N jugadores, según la cantidad definida en la interfaz.

En métricas de tipo porcentual, se incorporó un filtro adicional para garantizar un mínimo de intentos o acciones (por ejemplo, al menos 10 entradas para el porcentaje de entradas ganadas, o 10 tiros a puerta enfrentados para el porcentaje de paradas en porteros), evitando así rankings distorsionados por muestras demasiado pequeñas.

6.2.2. Gráficos

Ahora accedemos a la segunda ventana de navegación dentro de esta sección de la aplicación. En este apartado, tal como sugiere su nombre, se presentan tres gráficos distintos que nos permiten evaluar, visualizar y analizar el rendimiento de diferentes jugadores de forma dinámica e intuitiva.

Cada gráfico ofrece una perspectiva complementaria, pensada para facilitar la interpretación de los datos. Además, al ser completamente interactivos, invitan al usuario a explorar con mayor profundidad, comparando rendimientos y descubriendo detalles que podrían pasar desapercibidos en un análisis más tradicional.

Radar

Tabla comparativa de estadísticas

Junto al gráfico radar, se presenta una tabla con las estadísticas utilizadas, mostrando tanto el valor bruto como el percentil alcanzado por cada jugador. Esta tabla permite un análisis numérico detallado complementario al gráfico.

Incluye además información contextual como el equipo, edad y minutos jugados. Cada columna representa a un jugador y cada fila a una estadística, permitiendo comparaciones

directas y ordenadas.

Como podemos ver en la *Figura 6.22* los percentiles aparecen entre paréntesis junto al valor correspondiente para una lectura rápida. La cabecera de cada jugador utiliza el mismo color asignado en el radar.

ESTADÍSTICA	PEDRI	JOSHUA KIMMICH	LUKA MODRIĆ
Equipo	Barcelona	Bayern Munich	Real Madrid
Edad	21	29	38
% Duelos Ganados	34.30 (14)	60.00 (86)	43.80 (41)
Acciones Defensivas/90'	2.72 (60)	2.28 (43)	2.86 (63)
Conducciones Progresivas/90'	2.88 (79)	1.61 (48)	3.05 (82)
Pases Progresivos/90'	11.25 (99)	11.46 (100)	10.25 (99)
Pases Clave/90'	2.19 (92)	2.63 (97)	3.25 (99)
xAG (Asistencias)/90'	0.23 (87)	0.23 (88)	0.34 (98)
Regates Exitosos/90'	1.25 (73)	0.38 (16)	0.69 (42)
% Pases Completos	87.45 (86)	89.51 (94)	87.56 (87)
Pases Completos/90'	75.12 (98)	103.45 (100)	84.93 (99)
Minutos	2879	2847	1827

Figura 6.22: **Tabla de estadísticas** comparativas entre *Pedri*, *Kimmich* y *Modric*.

Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

La tabla se genera como un componente HTML personalizado, usando funciones de **shiny** y estilos CSS para lograr una presentación limpia.

Los datos se extraen del mismo conjunto de entrada que alimenta el gráfico radar, y se ordenan de manera consistente. Se emplean funciones auxiliares para redondear, colorear cabeceras y formatear los valores con percentiles.

Gráfico radar por jugador

Para representar de forma comparativa el perfil técnico de uno o varios jugadores, se ha implementado un gráfico tipo radar que visualiza distintas estadísticas normalizadas según percentiles por posición. Este gráfico permite detectar fortalezas y debilidades relativas de cada jugador frente al resto de la base de datos en su misma posición.

Cada eje del radar corresponde a una métrica específica (como pases progresivos, duelos defensivos o tiros por 90 minutos), y su escala se define en función de valores mínimo y máximo esperados para la posición seleccionada. Los valores se representan como percentiles para garantizar comparabilidad.

Como podemos apreciar en la *Figura 6.23*, el gráfico permite superponer hasta tres jugadores simultáneamente, diferenciándolos por color. Se trazan polígonos rellenos y líneas exteriores, facilitando la visualización de similitudes o contrastes entre los perfiles seleccionados.

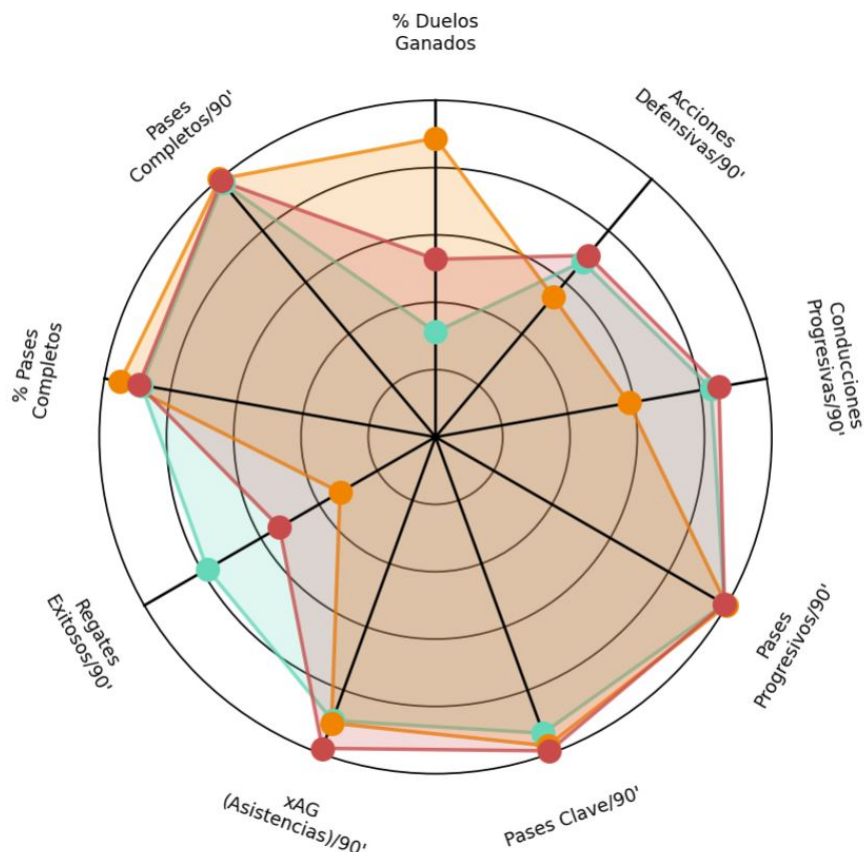


Figura 6.23: **Gráfico de radar** comparativo de *Pedri* (Verde), *Kimmich* (Naranja) y *Modric* (Rojo).

Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

El gráfico se genera con la clase `Radar` de la biblioteca `mplsoccer`, utilizando parámetros personalizados por posición, incluyendo:

- Columnas y etiquetas específicas por posición (GK, DF, MF, FW).
- Rangos de valores máximos y mínimos ajustados al percentil 0 y 100.
- Indicadores de si un valor más bajo es mejor (`lower_is_better`).

Los percentiles utilizados fueron calculados durante la etapa de limpieza y preprocesamiento de los datos.

Cuando no hay jugadores seleccionados, se muestra un mensaje neutral solicitando la selección de al menos uno.

Jugadores similares

Para identificar futbolistas con perfiles estadísticos similares al jugador seleccionado, se ha implementado una herramienta basada en la similitud del coseno.

Se normalizan las estadísticas numéricas (excluyendo edad, minutos, percentiles y datos categóricos) mediante un escalado robusto, y se calcula la similitud respecto al primer jugador seleccionado. El resultado muestra los cinco jugadores más parecidos, ordenados por porcentaje de similitud.

Como se aprecia en la *Figura 6.24*, esta funcionalidad permite descubrir alternativas comparables o talentos con características cercanas al perfil de interés, útil para análisis de scouting o comparativas de rendimiento.

PLAYER	SQUAD	POS	AGE	SIMILITUD
Nicolò Barella	Inter	MF	27	91.2
Bruno Guimarães	Newcastle Utd	MF	26	90.9
Éderson	Atalanta	MF	25	87.9
Rodrigo De Paul	Atlético Madrid	MF	30	87.4
Angelo Stiller	Stuttgart	MF	23	86.9

Figura 6.24: Jugadores similares a Pedri.
Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

Se utiliza la clase `RobustScaler` de `sklearn.preprocessing` para estandarizar los datos, minimizando el efecto de valores extremos.

La similitud se calcula con la función `cosine_similarity` de `sklearn.metrics.pairwise`, comparando vectores de características entre jugadores.

Los resultados se visualizan en una tabla simple que incluye nombre, equipo, edad, posición y el porcentaje de similitud con el jugador seleccionado.

Se eliminan previamente columnas categóricas, identificadores, variables redundantes y cualquier columna con valores nulos o no numéricos.

Pizza Plot

Gráfico pizza por jugador

Para representar de forma visual e intuitiva el perfil de un único jugador, se ha implementado un gráfico tipo *pizza plot*, basado en la clase `PyPizza` de la biblioteca `mplsoccer`. Este tipo de visualización permite destacar los puntos fuertes y débiles del jugador seleccionado mediante la representación radial de múltiples métricas, normalizadas en forma de percentiles frente a otros jugadores de su misma posición.

A diferencia del gráfico radar, que permite comparar varios jugadores, el gráfico pizza está diseñado para enfocar la atención en un único perfil, utilizando colores temáticos por categoría (ataque, posesión, defensa) y etiquetas claras en cada segmento. El resultado es una visualización atractiva, informativa y de fácil interpretación.

Cada métrica se representa como un segmento del gráfico, con un valor entre 0 y 100 que indica el percentil del jugador respecto al resto de futbolistas en su posición. Cuanto mayor sea el relleno de un segmento, mejor es el rendimiento relativo en esa métrica.

Como se aprecia en la *Figura 6.25*, el gráfico que se muestra en incluye también el nombre del jugador, su equipo y temporada, además de una leyenda que indica qué métricas han sido invertidas para mantener la lógica visual.

Pedri (Barcelona)
Percentile Rank vs Top-Five League Players
Temporada 2024-25

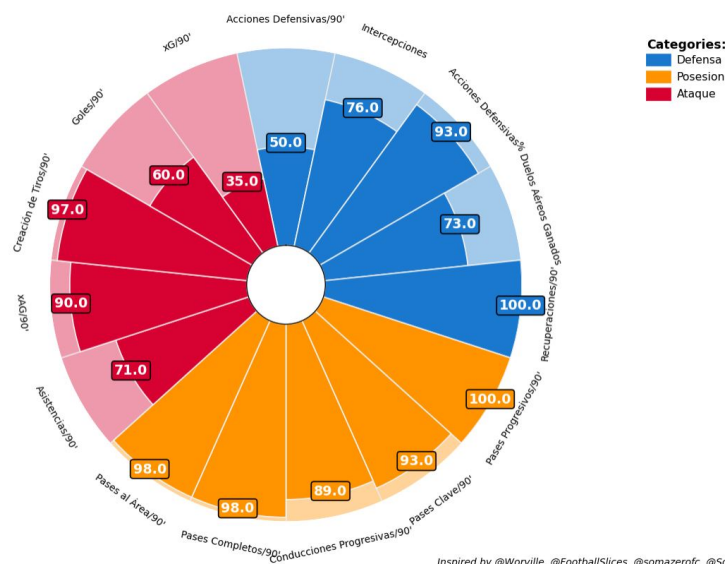


Figura 6.25: **Gráfico de Pizza de Pedri.**
Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

El gráfico se genera dinámicamente según la posición del jugador (GK, DF, MF, FW), utilizando configuraciones predefinidas que incluyen:

- Columnas y etiquetas específicas para cada posición, definidas en un diccionario externo.
- Cálculo de percentiles sobre jugadores de la misma posición, filtrando casos con datos nulos.
- Inversión de métricas donde un valor menor implica mejor rendimiento (por ejemplo, goles encajados por 90 minutos).

Visualmente, se utilizan esquemas de color diferenciados:

- Para jugadores de campo: azul (defensa), naranja (posesión) y rojo (ataque).
- Para porteros: una única tonalidad temática para las métricas propias del puesto.

Los datos empleados provienen de las base de datos de *FBref*. La implementación del gráfico pizza está inspirada en el trabajo de @Worville, @FootballSlices, @somazeroFc, @Soumyaj15209314 y la biblioteca *mplsoccer* para Python.

Tabla de fortalezas y debilidades

De forma complementaria al gráfico, se genera una tabla resumen que destaca las cinco estadísticas en las que el jugador presenta un mejor rendimiento relativo (mayor percentil), y las cinco en las que se sitúa en el nivel más bajo frente a sus pares de posición.

Este análisis permite sintetizar los aspectos en los que el jugador sobresale y aquellos donde podría mejorar, ofreciendo un perfil claro y enfocado desde el punto de vista del rendimiento estadístico.

Cada fila incluye el nombre de la métrica y el percentil alcanzado, ordenado de mayor a menor (en fortalezas) o de menor a mayor (en debilidades), como podemos apreciar en la *Figura 6.26*. Las métricas invertidas se tratan correctamente para que el análisis sea coherente en todos los casos.



Figura 6.26: **Puntos fuertes y débiles** de *Pedri*.
Fuente: *Elaboración Propia (Datos: Fbref/ Kaggle)*

Consideraciones técnicas

La lógica de cálculo sigue los siguientes pasos:

- Filtrado del conjunto de datos para incluir únicamente jugadores de la misma posición.
- Eliminación de columnas no numéricas o irrelevantes (por ejemplo, edad, equipo, minutos, identificadores).
- Cálculo del ranking para cada métrica, ajustando el orden cuando `lower_is_better` es verdadero.
- Selección de las cinco mejores y cinco peores métricas para el jugador, mostrando el nombre de la estadística y su percentil.

La tabla resultante se presenta junto al gráfico, destacando visualmente los nombres de las métricas y utilizando colores consistentes con el gráfico principal. Se añade una aclaración en caso de que alguna de las métricas haya sido invertida, para asegurar una correcta interpretación por parte del usuario.

Scatterplot

Gráfico de dispersión (scatterplot)

Para facilitar la exploración y comparación simultánea de dos métricas estadísticas de los jugadores, se ha implementado un gráfico de dispersión interactivo que permite visualizar correlaciones y distribuciones dentro del conjunto de datos.

Este scatterplot cuenta con múltiples filtros para refinar la selección de jugadores:

- **Posición:** Permite filtrar jugadores según su rol en el campo (portero, defensa, centrocampista, delantero).
- **Liga o competición:** Selección de una o varias ligas para analizar jugadores de determinados campeonatos.
- **Edad:** Restricción por rangos etarios (por ejemplo, menores de 25, 23, 21 años, etc.).

Además, se puede seleccionar qué dos estadísticas representar en los ejes X e Y, y optar por mostrar los nombres de los jugadores directamente en el gráfico para identificar puntos relevantes. También se incorpora la opción de destacar hasta 20 jugadores concretos con colores personalizados y marcadores más grandes para una mejor visualización.

El tamaño de cada punto está ponderado según los minutos jugados, de modo que los jugadores con mayor participación aparecen con un marcador más visible, aportando contexto adicional a la interpretación.

Opcionalmente, el gráfico puede mostrar las líneas de mediana en ambos ejes, facilitando la referencia respecto a la distribución general de los datos.

Si observamos la *Figura 6.27*, podemos ver que esta herramienta interactiva permite un análisis profundo y flexible del rendimiento de los futbolistas en función de distintas métricas y filtros, mejorando la comprensión y la toma de decisiones basadas en datos.



Figura 6.27: **Scatter-Plot**, eje x (Goles por 90 minutos) y eje y (Asistencias por 90 minutos), de algunos jugadores.

Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

La lógica de filtrado se implementa mediante funciones reactivas que:

- Aplican los filtros de posición, liga y edad sobre el conjunto de datos completo.
- Actualizan dinámicamente las opciones de jugadores disponibles para resaltar en función de los filtros aplicados, de manera que un jugador que juegue en la *Premier League* no estará disponible si se aplica un filtro de *LaLiga*.
- Normalizan el tamaño de los marcadores en función de los minutos jugados para reflejar la participación relativa.
- Generan el gráfico usando Matplotlib, controlando colores, transparencia, y etiquetas para facilitar la interpretación visual.

Con esto concluimos la segunda ventana de navegación y ahora nos enfocamos en la última dentro de esta sección, el **Buscador**.

6.2.3. Buscador

Buscador avanzado por perfil estadístico

Esta ventana presenta una herramienta inversa al *Radar* o *Pizza Plot*. En lugar de partir de un jugador para ver sus estadísticas, se definen las características deseadas para encontrar a los jugadores que mejor encajan en ese perfil.

Este buscador facilita la detección de talento y la selección basada en métricas específicas, permitiendo explorar la base de datos según los atributos estadísticos buscados.

El buscador permite filtrar por:

- **Posición:** Selección del rol principal en el campo (portero, defensa, centrocampista, delantero).
- **Liga o competición:** Filtro para acotar el conjunto de jugadores a una o varias ligas específicas.
- **Edad:** Restricción por rangos de edad predefinidos (por ejemplo, menores de 25, 23, 21 años, etc.).
- **Métricas estadísticas:** Selección flexible de métricas clave por posición y, de forma destacada, la posibilidad de añadir métricas adicionales que no se incluyen en los perfiles estándar. Esto permite al usuario personalizar y afinar su búsqueda, explorando perfiles muy específicos o poco comunes.

Mediante sliders que ajustan los percentiles mínimos requeridos en cada métrica, el usuario puede definir umbrales claros que filtran jugadores que superen esos valores en las estadísticas seleccionadas, facilitando la identificación de talentos con características particulares.

Si nos fijamos en los resultados que se muestran en la *Figura 6.28*, la tabla dinámica que incluye datos básicos y los valores de las métricas con su correspondiente percentil, facilitando la interpretación del rendimiento relativo de cada jugador.



Figura 6.28: **Buscador** para jugadores que jueguen como centrocampistas.
Fuente: Elaboración Propia (Datos: Fbref/ Kaggle)

Consideraciones técnicas

El buscador se implementa con las siguientes características técnicas:

- Se aplican filtros de posición, liga, edad y métricas seleccionadas sobre una copia del dataset original para preservar la integridad de los datos.
- Se calcula el percentil de cada métrica si no está precalculado, asegurando comparaciones homogéneas.
- Se aplican mínimos de intentos para métricas porcentuales para garantizar la fiabilidad de los datos.
- La interfaz genera controles dinámicos (sliders) para ajustar los percentiles mínimos requeridos en cada estadística, incluyendo las métricas adicionales seleccionadas por el usuario, brindando gran flexibilidad y personalización.
- Los resultados se ordenan por minutos jugados para priorizar jugadores con mayor participación y relevancia.
- Se gestionan casos sin resultados mostrando mensajes claros para el usuario.
- La actualización y renderizado son reactivos, proporcionando una experiencia interactiva y en tiempo real.

Con esta funcionalidad finaliza la sección dedicada a la exploración avanzada y búsqueda de perfiles dentro del conjunto de datos de jugadores, destacando la posibilidad de adaptar la búsqueda a necesidades muy específicas mediante la inclusión de métricas adicionales.

6.3. Visión por Computadora

Nos adentramos en la última de las 3 secciones descritas en la *Figura 6.1*, en este caso, se centra en la visión por computadora.

Esta sección de la aplicación permite al usuario visualizar fragmentos de vídeo de partidos de fútbol a los que se les han aplicado técnicas avanzadas de visión artificial.

Al ingresar al panel, se muestra un título y un breve mensaje introductorio que da contexto al módulo. A continuación, aparece un selector desplegable que permite elegir entre distintos vídeos procesados previamente (por ejemplo: “Partido 1”, “Partido 2”, “Partido 3”). En un futuro, lo ideal sería que el usuario pudiese subir el vídeo a la plataforma y este fuese procesado por el sistema. Este selector está diseñado para ser simple e intuitivo, facilitando la navegación entre diferentes partidos disponibles.

El diseño de esta ventana se organiza en dos partes:

- **A la izquierda:** se encuentra el componente de reproducción del vídeo seleccionado. Este reproductor muestra directamente los videos a los que se les ha aplicado un procesamiento de visión por computadora, incluyendo detección de jugadores, estimación de poses clave, seguimiento de movimientos y anotaciones visuales superpuestas.
- **A la derecha:** se presenta una explicación detallada de las funcionalidades del módulo, junto con una lista de las fases principales del procesamiento visual aplicado. Estas fases incluyen: detección de jugadores y balón en cada fotograma, estimación de posturas, seguimiento multiobjeto con asignación de ID, análisis de posesión y trayectorias, y visualización enriquecida del resultado.

Esta ventana, la cual podemos apreciar en la *Figura 6.29*, no solo ofrece una herramienta visual poderosa, sino que también busca acercar al usuario a las posibilidades que ofrece la inteligencia artificial aplicada al análisis de partidos, abriendo nuevas formas de interpretar el juego desde una perspectiva técnica y dinámica.



Figura 6.29: **Visualización** de la pestaña de *Computer Vision*.

Fuente: Elaboración Propia

6.4. Conclusión

Y con esta sección damos por concluida la presentación detallada de la aplicación desarrollada. Un proyecto que, a lo largo de sus tres grandes apartados, ha perseguido un objetivo claro y ambicioso: acercar el poder del *Big Data* y la inteligencia artificial al mundo del fútbol de manera accesible, visual y funcional.

Lejos de ser una herramienta más, esta aplicación representa una ventana hacia el futuro del análisis deportivo, donde la combinación de datos masivos, modelos predictivos y visualizaciones intuitivas permite no solo entender mejor el rendimiento de los jugadores, sino también descubrir nuevas narrativas que enriquecen la experiencia de cualquier aficionado o analista.

En conjunto, este proyecto es una pequeña muestra de lo que la tecnología puede ofrecer cuando se alinea con la pasión por el deporte rey. Porque detrás de cada estadística hay una historia esperando ser contada, y detrás de cada jugada, una oportunidad para ver el juego desde una perspectiva diferente, más profunda y, sin duda, más fascinante.

Así, la aplicación no solo invita a explorar datos, sino a conectar con el fútbol desde una dimensión donde la ciencia y la emoción se entrelazan, abriendo camino a un análisis inteligente que, a su vez, alimenta el amor por este deporte universal.

Capítulo 7

Conclusiones y proyección futura.

7.1. Conclusiones

Este trabajo ha tenido como objetivo principal el desarrollo de una prueba de concepto de una plataforma interactiva para el análisis de datos en el fútbol, combinando técnicas de análisis estadístico avanzado con herramientas de visión por computadora. A lo largo del proyecto, se han definido una serie de objetivos específicos que, en su mayoría, han sido alcanzados satisfactoriamente. A continuación, se repasan brevemente, junto a su grado de cumplimiento:

1. **Diseñar un cuadro de mando interactivo con Shiny for Python:** Este objetivo se ha cumplido plenamente. Se ha desarrollado una interfaz funcional, visual y dinámica que permite explorar datos filtrando por jugador, equipo o métrica, con gráficos interactivos y comparaciones personalizadas.
2. **Reunir y armonizar datos provenientes de fuentes abiertas distintas:** Este objetivo también se ha alcanzado con éxito. Se han combinado datos de eventos (StatsBomb) con estadísticas individuales (FBref, Kaggle), lo que ha exigido tareas complejas de limpieza, normalización y transformación para asegurar su compatibilidad y utilidad conjunta.
3. **Desarrollar modelos de predicción de goles y asistencias:** Se ha cumplido este objetivo mediante la implementación de modelos de aprendizaje automático, como `XGBoost` o `Random Forest`, capaces de predecir el rendimiento futuro de los jugadores en función de su historial. Los resultados han mostrado una precisión aceptable, sobre todo a la hora de detectar tendencias.
4. **Incorporar visión por computadora para enriquecer el análisis en vídeo:** Aunque se trata de un objetivo más exploratorio, se ha logrado una implementación básica utilizando modelos preentrenados (YOLOv8) y vídeos etiquetados. Se ha conseguido detectar y seguir jugadores en secuencias reales, lo que abre la puerta a futuras mejoras en tracking, posicionamiento y generación automática de métricas visuales.
5. **Crear una interfaz para analizar competiciones oficiales de forma interactiva:** Este objetivo se ha completado. La aplicación permite cargar y explorar partidos de competiciones como la Eurocopa 2024, visualizar eventos en el campo y

comparar el rendimiento de distintas selecciones, lo cual aporta un valor importante en contextos competitivos reales.

6. **Diseñar visualizaciones personalizadas según la posición del jugador:** Objetivo conseguido. Se han implementado tanto *radar plots* como *pizza charts* adaptados a diferentes posiciones (portero, defensa, centrocampista, delantero), mostrando percentiles y rankings en función de métricas relevantes según el rol.
7. **Crear un buscador por perfil de jugador:** Este objetivo también se ha cumplido satisfactoriamente. Se ha desarrollado un filtro interactivo que permite identificar jugadores según criterios definidos, facilitando procesos de *scouting* o comparación por perfiles de rendimiento.

En conclusión, los resultados del proyecto no solo cumplen los objetivos planteados, sino que también evidencian el potencial del análisis de datos y la inteligencia artificial como herramientas de apoyo en el ámbito futbolístico. Más allá del componente académico, se ha construido una base funcional que podría ser útil en entornos reales para cuerpos técnicos, analistas o aficionados con perfil técnico. El camino recorrido demuestra que es posible acercar el fútbol a un análisis más profundo, riguroso y visualmente accesible.

Este trabajo, por tanto, representa un primer paso hacia soluciones más avanzadas, en las que el dato y la imagen conviven para ofrecer una nueva mirada sobre el juego. Las líneas futuras podrían incluir la mejora de los modelos predictivos, una mayor automatización del análisis visual y la integración de datos de tracking más precisos.

7.2. Trabajo futuro

Este proyecto ha sido solo el primer paso de un camino mucho más amplio. Y, la verdad, es que mientras lo desarrollaba, no dejaban de surgir ideas y posibilidades que me gustaría explorar en el futuro.

Una de las primeras ampliaciones que tengo en mente es extender el análisis de competiciones. Hasta ahora, el foco ha estado en la Eurocopa 2024 gracias a los datos detallados que ofrece StatsBomb. Sin embargo, el fútbol no se detiene ahí. Sería muy valioso incluir más torneos para poder comparar estilos de juego, niveles de intensidad o rendimiento colectivo bajo distintos contextos. Al final, el fútbol se vive de muchas formas y cuantas más capas podamos analizar, mejor entenderemos el juego.

Algo parecido ocurre con los jugadores. Actualmente, el análisis se centra en futbolistas de las cinco grandes ligas europeas y solo de la temporada actual. Pero, ¿y si ampliamos la base de datos? Integrar datos de más temporadas permitiría estudiar la evolución de los jugadores, detectar trayectorias ascendentes (o declives sutiles) y construir modelos predictivos más sólidos. Además, incluir ligas menos mediáticas, como la Eredivisie, las ligas de Sudamérica, o incluso segundas divisiones no solo enriquecería el análisis, sino que abriría la puerta a descubrir talentos ocultos o entender cómo varían ciertos patrones según el contexto competitivo.

Por último, uno de los desafíos más emocionantes y ambiciosos es llevar el análisis por visión artificial a otro nivel. Pensé en crear una plataforma accesible y funcional donde cualquier usuario pueda subir un vídeo de un partido o jugada, y obtener resultados automáticamente en el mismo entorno sin necesidad de conocimientos técnicos. Es un objetivo muy complejo, sí, pero perfectamente posible si se integran herramientas de visión por computadora, interfaces amigables y un poco de ingenio.

En resumen, lo que hoy es una base sólida podría convertirse, con tiempo y dedicación, en una herramienta mucho más completa, versátil y abierta. Porque el fútbol sigue evolucionando, y nosotros también tenemos que estar preparados para ir siempre un paso más allá.

Capítulo 8

Anexos

Anexo I – Estructura de los datos de eventos de StatsBomb

Una de las características más potentes y desafiantes del dataset de StatsBomb es su complejidad estructural. Lejos de ofrecer una única tabla plana con todos los eventos, la información de cada partido se distribuye en distintos bloques de datos que se complementan entre sí. Esto permite representar la riqueza del juego con mucho más detalle, aunque también requiere cierta comprensión técnica para poder manejarlos correctamente.

A continuación, se describen las cuatro tablas principales que conforman los datos de eventos obtenidos mediante la API `statsbombpy` para cada partido.

La tabla de eventos principal `df`

Esta es la tabla base donde se registra, evento a evento, lo que sucede durante un partido: pases, tiros, faltas, recuperaciones, sustituciones, etc. Cada fila representa una acción concreta. Esta tabla contiene, entre otras, las siguientes variables destacadas:

- **id, index:** identificadores únicos del evento.
- **period, minute, second, timestamp:** momento exacto en el que ocurre la acción.
- **team_name, player_name, position_name:** protagonistas del evento.
- **type_name, sub_type_name:** tipo y subtipo de acción (por ejemplo, “Pass” o “Shot”).
- **x, y, end_x, end_y:** coordenadas de inicio y fin del evento en el campo.
- **pass_recipient_name, pass_length, pass_height_name:** detalles sobre el pase, si es el caso.
- **shot_statsbomb_xg, shot_first_time, shot_one_on_one:** información específica para tiros.
- **goalkeeper_position_name:** posición del portero en eventos defensivos.

- **under_pressure, counterpress, aerial_won, foul_won_penalty, etc.:** indicadores adicionales que enriquecen cada acción.

Esta tabla es la columna vertebral del análisis: desde ella se cruzan datos con los otros bloques y se extraen la mayoría de las estadísticas del partido.

Eventos relacionados `df_related`

Esta tabla actúa como un mapa de relaciones entre eventos. Por ejemplo, permite conectar un pase con el disparo que genera o una recuperación con la acción anterior. Incluye variables como:

- **id, index:** evento original.
- **id_related, index_related:** evento relacionado.
- **type_name, type_name_related:** tipos de eventos que se relacionan (por ejemplo, un pase y un disparo).

Gracias a esta tabla es posible reconstruir secuencias encadenadas de juego y analizar, por ejemplo, qué jugadas generaron una ocasión clara o cómo se encadenaron errores defensivos antes de un gol.

Fotograma congelado `df_freeze`

Este bloque es particularmente útil para análisis tácticos, ya que ofrece un “congelado” del campo en el instante exacto de un disparo. Se puede ver dónde estaban posicionados todos los jugadores en ese momento. Las variables principales son:

- **player_name, position_name:** nombre y posición del jugador representado.
- **x, y:** coordenadas del jugador en el campo en el instante del disparo.
- **teammate:** indica si el jugador era compañero del que disparó o no.

Este tipo de información permite hacer análisis espaciales muy valiosos, como detectar si un delantero estaba rodeado de defensas o si un portero tenía visión clara del balón.

Esquemas y formaciones `df_tactics`

Por último, la tabla de táctica registra cómo estaba organizada la alineación del equipo en el momento en que se produce un cambio táctico o se inicia el partido. Algunas de sus variables clave son:

- **player_name, position_name:** jugadores titulares y su rol inicial.
- **jersey_number:** número de camiseta del jugador.

- **event_tactics_id**: identificador del evento táctico.

Esta tabla se cruza generalmente con la información del equipo y el tipo de formación para estudiar patrones colectivos, como cambios de esquema o sustituciones estratégicas.

En conjunto, estas tablas forman una estructura robusta que permite analizar el juego desde múltiples perspectivas: técnica, táctica, espacial y temporal. Aunque puede parecer abrumadora al principio, una vez comprendida, se convierte en una herramienta poderosa para explorar el fútbol con una profundidad que rara vez se alcanza solo con los ojos.

Anexo II. Descripción de variables y procesamiento del dataset FBref

El conjunto de datos de *Kaggle/Fbref* cuenta con 267 columnas y 2793 observaciones. Este anexo recoge todas las variables extraídas del conjunto de datos de FBref. Se han agrupado temáticamente para facilitar su interpretación. Las variables repetidas entre secciones (por ejemplo, goles o asistencias) no se han eliminado para reflejar fielmente la estructura original del dataset antes de la limpieza y filtrado. Además, incluye un desglose de qué variables se eligieron y qué nuevas variables se crearon para el dataset final.

1. Variables Generales (Overview)

- **Rk:** Número de ranking en la tabla.
- **Player:** Nombre del jugador.
- **Nation:** Nacionalidad del jugador (código de 3 letras).
- **Pos:** Posición principal (ej. DF, MF, FW).
- **Squad:** Club actual del jugador.
- **Comp:** Competición en la que participa (ej. Premier League).
- **Age:** Edad expresada en años y días.
- **Born:** Año de nacimiento.
- **MP:** Partidos jugados.
- **Starts:** Partidos como titular.
- **Min:** Minutos jugados.
- **90s:** Minutos jugados dividido entre 90 (equivalente a partidos completos).
- **Gls:** Goles anotados.
- **Ast:** Asistencias.
- **G+A:** Goles más asistencias.
- **G-PK:** Goles sin contar penaltis.
- **PK:** Penaltis convertidos.
- **PKatt:** Penaltis intentados.
- **CrdY:** Tarjetas amarillas.
- **CrdR:** Tarjetas rojas.
- **xG:** Goles esperados (Expected Goals).
- **npG:** Expected Goals sin contar penaltis.

- **xAG**: Asistencias esperadas.
- **npxG+xAG**: Suma de npxG y xAG.
- **PrgC**: Conducciones progresivas.
- **PrgP**: Pases progresivos.
- **PrgR**: Pases progresivos recibidos.
- **G+A-PK**: Goles más asistencias sin penaltis.
- **xG+xAG**: Expected Goals + Expected Assisted Goals.

2. Estadísticas de Tiro (Shooting)

- **Sh**: Tiros totales.
- **SoT**: Tiros a puerta.
- **SoT %**: Porcentaje de tiros a puerta.
- **Sh/90**: Tiros por cada 90 minutos.
- **SoT/90**: Tiros a puerta por cada 90 minutos.
- **G/Sh**: Goles por tiro.
- **G/SoT**: Goles por tiro a puerta.
- **Dist**: Distancia promedio de los tiros.
- **FK**: Tiros de falta directa.
- **npxG/Sh**: xG sin penaltis por tiro.
- **G-xG**: Diferencia entre goles y xG.
- **np:G-xG**: Diferencia entre goles y npxG.
- **Gls_stats_shooting, PK_stats_shooting, PKatt_stats_shooting, xG_stats_shooting, npxG_stats_shooting, 90s_stats_shooting**: Copias redundantes de métricas ya listadas en *Overview*

3. Estadísticas de Pase (Passing)

Generales

- **Cmp**: Pases completados.
- **Att**: Pases intentados.
- **Cmp %**: Porcentaje de precisión.
- **TotDist**: Distancia total de pases.

- **PrgDist:** Distancia progresiva.
- **xA:** Expected Assists.
- **A-xAG:** Diferencia entre asistencias reales y esperadas.
- **KP:** Pases clave.
- **1/3:** Pases al tercio final.
- **PPA:** Pases al área.
- **CrsPA:** Centros al área.
- **Ast_stats_passing, xAG_stats_passing, PrgP_stats_passing:** Variables duplicadas.

Tipos de Pase

- **Live:** Pases en juego.
- **Dead:** Pases de balón parado.
- **FK:** Faltas.
- **TB:** Pases filtrados.
- **Sw:** Cambios de orientación.
- **Crs:** Centros.
- **TI:** Saques de banda.
- **CK:** Saques de esquina.
- **Off:** Pases en fuera de juego.
- **Blocks:** Pases bloqueados.
- **Cmp_stats_passing_types, Att_stats_passing_types, FK_stats_passing_types:** Repetidos de otras secciones.

4. Creación de Juego (SCA/GCA)

- **SCA:** Acciones que generan tiros.
- **SCA90:** SCA por 90 minutos.
- **GCA:** Acciones que generan goles.
- **GCA90:** GCA por 90 minutos.
- **PassLive, PassDead, TO:** Origen de las acciones (pases, pérdidas).
- **Sh:** Tiros que generan otra acción.
- **Fld:** Faltas recibidas que generan tiros.
- **Def:** Recuperaciones que generan tiros.

5. Defensa

- **Tkl:** Entradas realizadas.
- **TklW:** Entradas ganadas.
- **Def/Mid/Att 3rd:** Zonas del campo donde se realizan entradas.
- **Tkl %:** Porcentaje de éxito en entradas.
- **Lost:** Entradas perdidas.
- **Blocks:** Bloqueos de tiros o pases.
- **Sh:** Bloqueos de tiro.
- **Pass:** Bloqueos de pase.
- **Int:** Intercepciones.
- **Tkl+Int:** Suma de entradas e intercepciones.
- **Clr:** Despejes.
- **Err:** Errores que terminan en tiro rival.

6. Posesión

- **Touches:** Toques totales.
- **Def/Att Pen:** Toques en el área propia o rival.
- **Succ:** Regates exitosos.
- **Succ %:** Éxito en regates.
- **Tkld:** Regates fallidos.
- **Tkld %:** Porcentaje de fallos.
- **Carries:** Conducciones.
- **TotDist, PrgDist:** Distancia total y progresiva.
- **PrgC:** Conducciones progresivas.
- **CPA:** Conducciones al área.
- **1/3:** Conducciones al tercio final.
- **Mis:** Pérdidas de balón.
- **Dis:** Recuperaciones tras pérdida.
- **Rec:** Pases recibidos.

7. Tiempo de Juego

- **MP, Min, Starts:** Repetidos.
- **Mn/MP:** Minutos por partido.
- **Min %:** Porcentaje de minutos jugados.
- **Mn/Start:** Minutos por titularidad.
- **Subs:** Veces entrando como suplente.
- **Mn/Sub:** Minutos por suplencia.
- **unSub:** Veces sustituido.
- **PPM:** Puntos por partido.
- **onG/onGA:** Goles a favor/en contra con el jugador en cancha.
- **+/-:** Diferencia de goles.
- **On-Off:** Diferencia de rendimiento con/sin el jugador.

8. Misceláneas

- **2CrdY:** Dobles amarillas.
- **Fls:** Faltas cometidas.
- **Fld, Off, Crs, Int, TklW:** Repetidos.
- **PKwon:** Penaltis provocados.
- **PKcon:** Penaltis cometidos.
- **OG:** Goles en propia puerta.
- **Recov:** Recuperaciones.
- **Won, Lost, Won %:** Duelos y éxito en los mismos.

9. Porteros (Goalkeeping)

Básico

- **GA:** Goles recibidos.
- **GA90:** Goles recibidos por 90 minutos.
- **SoTA:** Tiros enfrentados.
- **Saves:** Atajadas.
- **Save %:** Porcentaje de atajadas.

- **W/D/L**: Victorias, empates, derrotas.
- **CS**: Porterías imbatidas.
- **CS %**: Porcentaje de porterías a cero.

Avanzado

- **PSxG**: Post-shot xG.
- **PSxG+/-**: Diferencia con goles reales.
- **Cmp**: Pases completados (repetido).
- **Launch %**: Porcentaje de saques largos.
- **AvgLen**: Longitud media de saque.
- **Stp**: Tiros parados.
- **#OPA**: Acciones ofensivas (salidas fuera del área).

La verdad es que lo primero que salta a la vista al trabajar con los datos de FBref es su nivel de detalle. Es impresionante. Cada hoja de cálculo parece contar una historia completa sobre cada jugador: desde lo más básico, como los minutos que ha jugado o los goles que ha marcado, hasta aspectos mucho más complejos y matizados, como cuántos pases ha recibido en zonas peligrosas o cuánta diferencia hay entre lo que debería haber marcado y lo que realmente marcó. Y claro, todo esto abre un mundo de posibilidades: desde analizar el rendimiento con lupa hasta construir modelos que intenten predecir el futuro.

Ahora bien, esa riqueza tiene un precio. A medida que te adentras en el dataset, empiezas a notar algo: muchas variables se repiten, aparecen bajo distintos nombres o prefijos, o simplemente cambian la unidad (por partido completo, en total o por 90 minutos). Al principio puede parecer una ventaja, pero enseguida se vuelve un pequeño caos. Y es que, si no limpias y reorganizas bien toda esa información, puedes terminar con más ruido que claridad.

Por eso, antes de lanzarse a cualquier análisis serio, hay que hacer un trabajo paciente de selección y depuración: eliminar duplicados, unificar columnas y quedarse solo con lo que realmente aporta valor. Solo así se consigue una base de datos que no solo sea rica, sino también clara, manejable y, sobre todo, útil.

10. Selección final y variables creadas

Tras un proceso riguroso de limpieza, transformación y generación de métricas, se seleccionó un conjunto definitivo de variables que permiten analizar el rendimiento futbolístico desde diferentes ángulos: exploratorio, descriptivo y predictivo. A partir de más de 250 columnas originales del dataset de FBref, se depuraron aquellas redundantes o poco representativas. Las variables finales se agrupan a continuación por temáticas, diferenciando entre originales y nuevas (calculadas específicamente para este trabajo).

Identificación y contexto del jugador

■ Variables originales:

- Player: Nombre del jugador.
- Nation: Nacionalidad principal.
- Squad: Club en el que juega durante la temporada.
- Comp: Competición (liga o torneo).
- Pos: Posición principal (DF, MF, FW, GK).
- Age: Edad en años.
- MP, Starts, Min, 90s: Partidos jugados, titularidades, minutos totales y número de partidos equivalentes a 90 minutos.

Producción ofensiva y definición**■ Variables originales:**

- Gls: Goles totales.
- G-PK: Goles sin penaltis.
- xG: Goles esperados.
- npxG: Goles esperados sin penaltis.
- G-xG: Diferencia entre goles y goles esperados.
- np:G-xG: Variante sin penaltis de la anterior.
- Sh: Tiros totales.
- SoT: Tiros a puerta.
- G/Sh: Goles por tiro.
- G/SoT: Goles por tiro a puerta.
- Dist: Distancia media de los disparos.
- npxG/Sh: Goles esperados sin penaltis por tiro.

■ Variables nuevas:

- Gls_90: Goles por 90 minutos.
- G-PK_90: Goles sin penaltis por 90 minutos.
- xG_90: Goles esperados por 90 minutos.
- npxG_90: Goles esperados sin penaltis por 90 minutos.
- G-xG_90: Diferencia entre goles y xG por 90 minutos.
- np:G-xG_90: Variante sin penaltis por 90 minutos.
- Sh_90: Tiros totales por 90 minutos.
- SoT_90: Tiros a puerta por 90 minutos.

Asociación y pase**■ Variables originales:**

- Ast: Asistencias totales.
- xAG: Asistencias esperadas.
- xG+xAG: Goles esperados más asistencias esperadas.
- A-xAG: Diferencia entre asistencias reales y esperadas.
- Cmp, Att: Pases completados e intentados.
- PrgP: Pases progresivos.
- KP: Pases clave.
- PPA: Pases al área.
- 1/3: Pases al último tercio.
- CrsPA: Centros al área.
- SCA, SCA90: Acciones que preceden un tiro (totales y por 90).
- GCA: Acciones que preceden un gol.

■ **Variables nuevas:**

- Ast_90: Asistencias por 90 minutos.
- xAG_90: Asistencias esperadas por 90 minutos.
- xG+xAG_90: Suma de xG y xAG por 90 minutos.
- G+A: Goles más asistencias totales.
- G+A_90: Goles más asistencias por 90 minutos.
- Cmp_90: Pases completados por 90 minutos.
- Cmp %: Porcentaje de pases completados.
- KP_90: Pases clave por 90 minutos.
- PrgP_90: Pases progresivos por 90 minutos.
- PPA_90: Pases al área por 90 minutos.
- 1/3_90: Pases al último tercio por 90 minutos.
- CrsPA_90: Centros al área por 90 minutos.

Conducción, regate y posesión

■ **Variables originales:**

- Carries: Conducciones con balón.
- Touches: Toques de balón.
- PrgC: Conducciones progresivas.
- PrgR: Carreras progresivas.
- Succ: Regates exitosos.
- Att Pen: Acciones ofensivas en el área.
- Rec: Recepciones de pase.
- TotDist, PrgDist: Distancia total y progresiva recorrida con balón.

■ Variables nuevas:

- PrgC_90: Conducciones progresivas por 90 minutos.
- SuccDrib_90: Regates exitosos por 90 minutos.
- Touches_90: Toques de balón por 90 minutos.
- Att_Act_90: Métrica compuesta de actividad ofensiva (tiros, pases clave, regates) por 90 minutos.

Defensa y recuperación**■ Variables originales:**

- Tkl: Entradas totales.
- TklW: Entradas ganadas.
- Tkl %: Porcentaje de entradas ganadas.
- Int: Intercepciones.
- Tkl+Int: Suma de entradas e intercepciones.
- Att_stats_defense: Acciones defensivas intentadas.
- Blocks_stats_defense: Bloqueos.
- Clr: Despejes.
- Recov: Recuperaciones de balón.
- Won %: Porcentaje de duelos aéreos ganados.

■ Variables nuevas:

- Tkl+Int_90: Entradas + intercepciones por 90 minutos.
- Blocks_90: Bloqueos por 90 minutos.
- Clr_90: Despejes por 90 minutos.
- Recov_90: Recuperaciones por 90 minutos.

Porteros**■ Variables originales:**

- Saves: Paradas realizadas.
- Save %: Porcentaje de paradas.
- SoTA: Tiros a puerta enfrentados.
- GA, GA90: Goles encajados (totales y por 90).
- PSxG: Goles esperados tras disparo.
- PSxG+/-: Diferencia entre PSxG y goles recibidos.
- Cs, Cs %: Porterías a cero (totales y en porcentaje).

■ Variables nuevas:

- Saves_90: Paradas por 90 minutos.

Conclusión:

El proceso de depuración no solo sirvió para reducir ruido informativo, sino también para identificar y conservar las variables más representativas del rendimiento de los jugadores, eliminando aquellas redundantes o poco relevantes. De este modo, se logró simplificar la estructura del dataset sin perder su valor analítico, lo que permitió trabajar con una base de datos más enfocada, eficiente y directamente orientada a los objetivos del proyecto.

Bibliografía

- [1] LaLiga, “LaLiga se convierte en epicentro en julio del Big Data en el fútbol,” *LaLiga Noticias*, 8 de julio de 2024. Disponible en: <https://www.laliga.com/noticias/laliga-epicentro-en-julio-del-big-data-en-el-futbol>
- [2] G. Evans, “How data is changing football contract negotiations and transfers,” *The Athletic*, 3 de enero de 2024. Disponible en: <https://www.nytimes.com/athletic/5166636/2024/01/03/football-contract-negotiation-data-analytics-transfers>
- [3] F. Fraile, “Brentford, el recolector de diamantes de la Premier League,” *Relevo*, 13 de febrero de 2023. Disponible en: <https://www.relevo.com/futbol/premier-league/brentford-recolector-diamantes-20230210141439-nt.html>
- [4] Fintualist, “El Big Data llegó al fútbol al estilo Moneyball,” *Fintualist*, consultado en julio de 2025. Disponible en: <https://fintualist.com/chile/columnas/el-big-data-llego-al-futbol-al-estilo-moneyball>
- [5] Microsoft News Center, “LaLiga transforma la experiencia futbolística y su gestión interna con la IA de Microsoft,” *Microsoft News*, 14 de febrero de 2025. Disponible en: <https://news.microsoft.com/es-es/2025/02/14/laliga-transforma-la-experiencia-futbolistica-y-su-gestion-interna-con-la-ia-de-microsoft>
- [6] Gartner, “Big Data Definition,” *Gartner IT Glossary*, 2020. Disponible en: <https://www.gartner.com/en/information-technology/glossary/big-data>
- [7] I. Álvarez, “¿Qué es el Big Data?,” *Forbes España*, 10 de febrero de 2016 (actualizado el 12 de mayo de 2021). Disponible en: <https://forbes.es/empresas/3238/que-es-el-big-data/>
- [8] Codd, E. F. (1970). “A Relational Model of Data for Large Shared Data Banks.” *Communications of the ACM*, 13(6), 377–387.
- [9] CIO Staff (2007). “45 Years of Wal-Mart History: A Technology Time Line.” *CIO.com*. Disponible en: <https://www.cio.com/article/274537/infrastructure-45-years-of-wal-mart-history-a-technology-time-line.html>
- [10] George Firican (2021). “The History of Big Data.” *LightsOnData.com*. Disponible en: <https://www.lightsondata.com/the-history-of-big-data/>
- [11] Piesco, N. (2023). “The Evolution of Big Data: From SQL Databases to Interactive Analytics.” *Medium – DET Life*. Disponible en: <https://blog.det.life/the-evolution-of-big-data-from-sql-databases-to-interactive-analytics-bafdb3a9c481>
- [12] E.S. García. “Manual de instrucciones para dejar atrás la medicina de talla única.” *El País*, 25 de junio de 2025. Disponible en: <https://elpais.com/economia/branded/fundaciones/2025-06-25/manual-de-instrucciones-para-dejar-atras-la-medicina-de-talla-unica.html>

- [13] S. Domínguez. “Impacto del Big Data en la gestión de riesgos en el sector financiero.” *OpenWebinars Blog*, 26 de septiembre de 2023. Disponible en: <https://openwebinars.net/blog/impacto-del-big-data-en-la-gestion-de-riesgos-en-el-sector-financiero/>
- [14] Algor Education, “Algor: plataforma de aprendizaje adaptativo con IA”. Disponible en: <https://www.algorededucation.com/es>
- [15] Okorie, G. N., Egieya, Z. E., Ikwue, U., Udeh, C. A., Adaga, E. M., DaraOjimba, O. D., & Oriekhoe, O. I. (2024). *Leveraging Big Data for Personalized Marketing Campaigns: A Review*. ResearchGate. Disponible en: https://www.researchgate.net/publication/378293643_LEVERAGING_BIG_DATA_FOR_PERSONALIZED_MARKETING_CAMPAIGNS_A_REVIEW
- [16] J. Wang. “Real-time Optimization Algorithm for Intelligent Logistics Transportation Routes.” *ACM Digital Library*, 3 de agosto de 2024. Disponible en: <https://dl.acm.org/doi/10.1145/3662739.3672306>
- [17] Shahzad, K., Khan, S. A., Ahmad, S. and Iqbal, A. (2022). “A Scoping Review of the Relationship of Big Data Analytics with Context-Based Fake News Detection on Digital Media in Data Age.” *Sustainability*, 2 de noviembre de 2022. Disponible en: <https://doi.org/10.3390/su142114365>
- [18] The Impact Lawyers (2023). “The impact of new technologies in sport: big data and data analysis to improve sports performance.” Disponible en: <https://theimpactlawyers.com/articles/the-impact-of-new-technologies-in-sport-big-data-and-data-analysis-to-improve-sports-performance>
- [19] S. Mcauley. “What is Big Data? Explore 6 Key Benefits and Challenges.” *TMGVoice*, 26 de septiembre de 2024. Disponible en: <https://tmgvoice.com/2024/09/26/what-is-big-data-explore-6-key-benefits-and-challenges/>
- [20] IndustryTrends. “Leveraging Big Data Analytics to Drive Smarter Decisions.” *Analytics Insight*, 12 de diciembre de 2024. Disponible en: <https://www.analyticsinsight.net/data-analytics/leveraging-big-data-analytics-to-drive-smarter-decisions>
- [21] E. Matias. (2024). “8 Benefits of Enterprise Big Data Analytics for Decision Making.” *O Maringá*, 15 de diciembre de 2024. Disponible en: <https://omaringa.com.br/en/column/diversities/8-benefits-of-big-data-analytics-for-business-decision-making/>
- [22] DataTas (2024). “How Big Data is Changing Sports Analytics.” Disponible en: <https://datatas.com/how-big-data-is-changing-sports-analytics/>
- [23] “The Evolution of Sports Technology: A Historical Perspective,” *HillV*, 24 de abril de 2024. Disponible en: <https://hillv.com/article/the-evolution-of-sports-technology%3A-a-historical-perspective-471/>
- [24] Guttmann, A. (2000). “Development of Modern Sports.” En: J. Coakley and E. Dunning (Eds.), *Handbook of Sports Studies* London: Sage Publications. *OmniLogos*. Disponible en: <https://omnilogos.com/development-of-modern-sports/>
- [25] M. Stauff. “A Culture of Competition: Sport’s Historical Contribution to Datafication,” *TMG Journal for Media History*, 1 de noviembre de 2018. Disponible en: <https://tmgonline.nl/articles/10.18146/2213-7653.2018.365>

- [26] Hughes , “Charles Reep, el pionero estadístico contra la posesión,” *ABC Deportes*, 29 de marzo de 2020. Disponible en: https://www.abc.es/deportes/futbol/abci-raros-futbol-charles-reep-pionero-estadistico-contra-posesion-202003290822_noticia.html
- [27] A. Vinaixa, “El economista que cambió los deportes,” *ginvest*, 10 julio de 2023. Disponible en: <https://ginvest.es/leconomista-que-va-canviar-els-esports-2/>
- [28] L. Cordero Ramírez, “Llegó la era del baloncesto sin posiciones,” *Marca Blogs – Máster Big Data Deportivo*, 10 de julio de 2023. Disponible en: <https://www.marca.com/blogs/master-big-data-deportivo/2023/07/10/la-era-del-baloncesto-sin-posiciones.html>
- [29] E. Hood. “6 Historical Development – Biomechanics emerged from physical education as a specialized area of study in the mid-1960s and 1970s,” *SlidePlayer (s.f.)*. Disponible en: <https://slideplayer.com/slide/12744719/>
- [30] Ariel Dynamics (2007). “History of Performance Analysis Research.” Disponible en: <https://123dok.com/id/article/history-of-performance-analysis-research.10219357>
- [31] Wikipedia (2025). “Polar Electro.” Consultado en julio de 2025. Disponible en: https://en.wikipedia.org/wiki/Polar_Electro
- [32] Sports Universitas (2024). “¿Qué es la tecnología wearable en el deporte?”, 20 de agosto de 2024. Disponible en: <https://sportsuniversitas.com/blog/que-es-la-tecnologia-wearable-en-el-deporte/>
- [33] Victor R. A. Cossich, D. Carlgreen, R. J. Holash and L. Katz “Technological Breakthroughs in Sport: Current Practice and Future Potential of Artificial Intelligence, Virtual Reality, Augmented Reality, and Modern Data Visualization in Performance Analysis”, *applied sciences* 4 de diciembre de 2023. Disponible en: <https://www.mdpi.com/2076-3417/13/23/12965>
- [34] J. Pepinosa, “Liverpool y Google crean un asistente con inteligencia artificial para ganar los partidos,” *Infobae*, 19 de marzo de 2024. Disponible en: <https://www.infobae.com/tecnologia/2024/03/19/liverpool-y-google-crean-un-asistente-con-inteligencia-artificial-para-analizar-los-partidos/>
- [35] Sevilla F.C. e IBM, “El Sevilla FC transforma el proceso de scouting de jugadores con la IA generativa de IBM” *Sevilla F.C. Noticias*, 23 de enero de 2024. Disponible en: <https://sevillafc.es/actualidad/noticias/sevilla-fc-ibm-watsonx-ia-generativa-2024>
- [36] P. Riquelme, “HomeCourt, la interesante app para que mejores tu juego” *Viva Basquet*, 2 de marzo de 2020. Disponible en: <https://vivabasquet.com/104180-homecourt-la-interesante-app-para-que-mejores-tu-juego/>
- [37] SAS, “French national rugby team boosts performance with AI and analytics,” *SAS Press Release*, actualizado en abril de 2023 (colaboración desde 2021). Disponible en: https://www.sas.com/en_is/customers/french-rugby-federation.html
- [38] SAS, “SAS insights drive All Blacks as ‘Official Analytics Supplier’ to New Zealand Rugby,” *SAS Noticias*, 3 de septiembre de 2019. Disponible en: https://www.sas.com/es_mx/customers/french-rugby-federation.html
- [39] Redacción MAPFRE, “El análisis de datos en la Fórmula 1, la diferencia entre la victoria y la derrota,” *Mapfre Innovación*, 28 de febrero de 2020. Disponible en: <https://www.mapfre.com/actualidad/innovacion/formula-1-y-analisis-de-big-data/>

- [40] AnalyiSport (2025). “How Is Data Used in La Liga?”, 1 de diciembre de 2022. Disponible en: <https://analysisport.com/insights/how-is-data-used-in-la-liga/>
- [41] D.Jaja. (2025). “The power of big data in transforming football analytics.” *Sportmonks Blog*, 2 de julio de 2025. Disponible en: <https://www.sportmonks.com/blogs/the-power-of-big-data-in-transforming-football-analytics/>
- [42] S. Srivastava. (2025). “Data Analytics in Sports: Use-Cases, Examples, and Costs.” *Appinventiv Blog*, 17 de junio de 2025. Disponible en: <https://appinventiv.com/blog/data-analytics-in-sports-industry/>
- [43] SciSports (2023). “Optimising the scouting workflow.” *SciSports Track Record*, mayo de 2023. Disponible en: <https://www.scisports.com/track-record/optimising-the-scouting-workflow/>
- [44] R. Freemansonico “The Role of Data Analytics in Transforming Sports Marketing Strategies.” *Globally for Sport* Disponible en: <https://globallyfactorsport.com/the-role-of-data-analytics-in-transforming-sports-marketing-strategies/>
- [45] Springbord (2023). “Types of Sports Data to Measure in 2023,” *Springbord Blog*. Disponible en: <https://www.springbord.com/blog/types-of-sports-data-to-measure-in-2023/>
- [46] Stats Perform, “Sports Data and Analytics Solutions”. Disponible en: <https://www.statsperform.com>
- [47] StatsBomb, “Football Data and Analytics”. Disponible en: <https://statsbomb.com>
- [48] Wyscout, “Scouting y análisis de rendimiento para clubes de fútbol”. Disponible en: <https://wyscout.com>
- [49] FBref, “Fútbol: estadísticas de jugadores y equipos”. Disponible en: <https://fbref.com/es>
- [50] InStat, “InStat Sport: Performance Analysis”. Disponible en: <https://instatsport.com>
- [51] ChyronHego, “TRACAB: Tracking Systems”. Disponible en: https://chyron.com/?__hstc=70687055.ed3fe574a5399ccd436c74e161d8137d.1752313078455.1752313078455.1752313078455.1&__hssc=70687055.1.1752313078455&__hsfp=3790421854
- [52] Second Spectrum, “The Next Way of Watching Sports”. Disponible en: <https://www.secondspectrum.com>
- [53] SciSports, “Football Intelligence and Data Analytics”. Disponible en: <https://www.scisports.com>
- [54] Catapult Sports, “Tecnología de rendimiento para equipos deportivos”. Disponible en: <https://www.catapult.com/es/>
- [55] Macwilliam. “Understanding Data Providers: Eventing vs. Tracking,” *Football Analytics*. Disponible en: <https://favoredoutdoorlife.com/understanding-data-providers-eventing-vs-tracking/>
- [56] SciSports (2021). “The power of combining tracking and event data.” **SciSports Blog**, 9 de abril de 2021. Disponible en: <https://www.scisports.com/the-power-of-combining-tracking-and-event-data/>
- [57] STATSports, “GPS Performance Monitoring”. Disponible en: <https://eu.shop.statsports.com/es>

-
- [58] StatsBomb, “statsbombpy: A Python Wrapper for the StatsBomb API,” GitHub, 2025. Disponible en: <https://github.com/statsbomb/statsbombpy>
- [59] H. Sidorowicz, “Football Players Stats 2024–2025,” Kaggle, 2025. Disponible en: <https://www.kaggle.com/datasets/hubertsidorowicz/football-players-stats-2024-2025>
- [60] Kaggle y Deutsche Fußball Liga (DFL). *DFL - Bundesliga Data Shootout*. Disponible en: <https://www.kaggle.com/competitions/dfl-bundesliga-data-shootout>.
- [61] F. Rabaños, “LanusStats: librería de Python para scraping y visualización de datos de fútbol,” GitHub repository, 2024. Disponible en: <https://github.com/federicorabanos/lanusStats>
- [62] J. Pérez et al., *Supervision: A Python Library for Video Annotation and Object Detection*, GitHub repository, 2023. <https://github.com/roboflow/supervision>
- [63] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, “ByteTrack: Multi-Object Tracking by Associating Every Detection Box”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.
- [64] Roboflow, *Roboflow Platform for Computer Vision*, 2024. <https://roboflow.com>
- [65] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, Wiley, 1995.
- [66] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, 2019.
- [67] B. Xiao, H. Wu, and Y. Wei, “Simple Baselines for Human Pose Estimation and Tracking”, in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
- [68] A. Durgapal and A. Rowlinson (2025). “Examples — mplsoccer gallery.” *mplsoccer documentation*. Disponible en: <https://mplsoccer.readthedocs.io/en/latest/gallery/index.html>
- [69] Alejandro Hernández Beneito. *TFG - Código fuente de la aplicación*. GitHub repository, 2025. <https://github.com/Alex-h-b/TFG>
- [70] Hudl, “¿Qué son los goles esperados (xG)?”, *Hudl Blog – Performance Analysis*. Disponible en: <https://www.hudl.com/blog/que-es-el-xg>
- [71] DataMB (2025). “Football Visuals and Statistics”. Disponible en: <https://datamb.football/>