



**Universidad  
Europea**

**UNIVERSIDAD EUROPEA DE MADRID**

**ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO**

**MASTER EN ANALISIS DE DATOS MASIVOS (BIG DATA)**

**TRABAJO FIN DE MÁSTER**

**Diseño e implementación de un sistema de  
apoyo a la toma de decisiones para el scouting  
de jugadores mediante herramientas Big Data.**

**Andrea Ruiz Naranjo**

**Dirigido por**

**DR. ALBERTO PARTIDA RODRIGUEZ**

**CURSO 2025-2026**

Diseño e implementación de un sistema de apoyo a la toma de decisiones para el scouting de jugadores mediante herramientas Big Data.



Andrea Ruiz Naranjo

---

**TÍTULO:** Diseño e implementación de un sistema de apoyo a la toma de decisiones para el scouting de jugadores mediante herramientas Big Data.

**AUTOR:** Andrea Ruiz Naranjo

**TITULACIÓN:** MÁSTER EN ANALISIS DE DATOS MASIVOS (BIG DATA)

**DIRECTOR/ES DEL PROYECTO:** Dr. Alberto Partida Rodríguez

**FECHA:** marzo 2026

## RESUMEN

Este trabajo aborda el diseño e implementación de una solución tecnológica orientada al análisis de jugadores de fútbol profesional desde una perspectiva de apoyo a la toma de decisiones. La propuesta surge de la necesidad de contar con herramientas más estructuradas, objetivas y reproducibles en los procesos de scouting, en un contexto donde el análisis deportivo ha comenzado a incorporar cada vez con mayor fuerza el uso de datos, métricas e inteligencia artificial. En lugar de depender únicamente de la observación subjetiva, el proyecto plantea una alternativa que organiza la información de rendimiento de los jugadores y la transforma en insumos útiles para su evaluación comparativa.

Para ello, se desarrolló una arquitectura de datos que integra fuentes abiertas, procesos de transformación, almacenamiento analítico y visualizaciones, permitiendo construir una herramienta capaz de comparar jugadores por posición mediante indicadores estandarizados y multicriterio. El principal aporte del trabajo demuestra que es posible combinar datos abiertos, analítica y visualización en una solución funcional aplicada al scouting deportivo, contribuyendo a una evaluación más clara, consistente y sustentada en evidencia dentro del entorno del fútbol profesional.

**Palabras clave:** Big Data, scouting, mercado de transferencias, sistema de apoyo a la decisión, Apache Hop, PostgreSQL, Apache Superset.

## ABSTRACT

This work addresses the design and implementation of a technological solution aimed at the analysis of professional football players from a decision-support perspective. The proposal arises from the need for more structured, objective, and reproducible tools in scouting processes, within a context in which sports analysis has increasingly incorporated the use of data, metrics, and artificial intelligence. Rather than relying solely on subjective observation, the project proposes an alternative that organizes player performance information and transforms it into useful inputs for comparative evaluation.

To achieve this, a data architecture was developed that integrates open data sources, transformation processes, analytical storage, and visualizations, making it possible to build a tool capable of comparing players by position through standardized indicators and multicriteria approaches. The main contribution of the study demonstrates that it is possible to combine open data, analytics, and visualization in a functional solution applied to sports scouting, contributing to a clearer, more consistent, and evidence-based evaluation within the environment of professional football.

**Keywords:** Big Data, scouting, transfer market, decision support system, Apache Hop, PostgreSQL, Apache Superset.

## **AGRADECIMIENTOS**

En el fútbol moderno, comprender el talento exige algo más que observar: exige analizar, comparar y decidir con evidencia.

## TABLA RESUMEN

|                                                                                                                                            | DATOS                                                                                                                                                     |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nombre y apellidos:</b>                                                                                                                 | Andrea Ruiz Naranjo                                                                                                                                       |
| <b>Título del proyecto:</b>                                                                                                                | Diseño e implementación de un sistema de apoyo a la toma de decisiones para el scouting de jugadores (Scouting tool v2.0) mediante herramientas Big Data. |
| <b>Directores del proyecto:</b>                                                                                                            | Alberto Partida Rodríguez                                                                                                                                 |
| <b>El proyecto se ha realizado en colaboración de una empresa o a petición de una empresa:</b>                                             | NO                                                                                                                                                        |
| <b>El proyecto ha implementado un producto:</b><br>(esta entrada se puede marcar junto a la siguiente)                                     | SI                                                                                                                                                        |
| <b>El proyecto ha consistido en el desarrollo de una investigación o innovación:</b><br>(esta entrada se puede marcar junto a la anterior) | SI                                                                                                                                                        |
| <b>Objetivo general del proyecto:</b>                                                                                                      | Diseñar por medio de herramientas de Big data para la toma de decisiones para el scouting de jugadores.                                                   |

## Índice

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| RESUMEN.....                                                                      | 3  |
| ABSTRACT .....                                                                    | 4  |
| TABLA RESUMEN .....                                                               | 7  |
| Capítulo 1. RESUMEN DEL PROYECTO .....                                            | 13 |
| 1.1 Contexto y justificación .....                                                | 14 |
| 1.2 Planteamiento del problema .....                                              | 16 |
| 1.3 Objetivos del proyecto .....                                                  | 17 |
| 1.4 Resultados obtenidos .....                                                    | 18 |
| 1.5 Estructura de la memoria.....                                                 | 19 |
| Capítulo 2. ESTADO DEL ARTE.....                                                  | 22 |
| 2.1 Estado del arte .....                                                         | 22 |
| 2.2 Contexto y justificación .....                                                | 29 |
| 2.3 Planteamiento del problema .....                                              | 30 |
| Capítulo 3. OBJETIVOS .....                                                       | 32 |
| 3.1 Objetivos generales .....                                                     | 32 |
| 3.2 Objetivos específicos .....                                                   | 32 |
| 3.3 Beneficios del proyecto.....                                                  | 32 |
| Capítulo 4. DESARROLLO DEL PROYECTO .....                                         | 34 |
| 4.1 Planificación del proyecto .....                                              | 34 |
| 4.2 Descripción de la solución, metodologías y herramientas empleadas .....       | 35 |
| 4.3 Recursos requeridos.....                                                      | 67 |
| 4.4 Presupuesto.....                                                              | 67 |
| 4.5 Viabilidad .....                                                              | 68 |
| 4.6 Resultados del proyecto .....                                                 | 69 |
| Capítulo 5. DISCUSIÓN .....                                                       | 73 |
| 5.1 Disponibilidad y acceso a fuentes de datos.....                               | 73 |
| 5.2 Pertinencia de la metodología adoptada .....                                  | 75 |
| 5.3 Evolución del enfoque: de la predicción al apoyo integral a la decisión ..... | 76 |
| 5.4 Limitaciones tecnológicas: el caso Apache Superset.....                       | 76 |

|             |                                                            |    |
|-------------|------------------------------------------------------------|----|
| 5.5         | Valor e impacto desde una perspectiva aplicada .....       | 77 |
| 5.6         | Guía de uso del sistema: Scouting tool v2.0 .....          | 78 |
| Capítulo 6. | CONCLUSIONES .....                                         | 81 |
| 6.1         | Conclusiones del trabajo.....                              | 81 |
| 6.2         | Conclusiones personales.....                               | 84 |
| Capítulo 7. | FUTURAS LÍNEAS DE TRABAJO.....                             | 86 |
| 7.1         | Ampliación de la cobertura de datos.....                   | 86 |
| 7.2         | Incorporación de nuevas fuentes de datos .....             | 86 |
| 7.3         | Mejora de la capa de visualización .....                   | 86 |
| 7.4         | Incorporación de funcionalidades analíticas avanzadas..... | 87 |
| 7.5         | Validación en entornos reales.....                         | 87 |
| Capítulo 8. | REFERENCIAS.....                                           | 89 |
| Capítulo 9. | ANEXOS .....                                               | 91 |
| 9.1         | Objetivos de desarrollo sostenible .....                   | 91 |
| 9.2         | Uso de la Inteligencia artificial en el TFM .....          | 91 |
| 9.3         | Entrega practica .....                                     | 92 |

## Índice de Figuras

|                                                                                   |    |
|-----------------------------------------------------------------------------------|----|
| Ilustración 1: Problema y causas del proyecto .....                               | 16 |
| Ilustración 2: Cuadro comparativo de scouting .....                               | 17 |
| Ilustración 3: Objetivos del proyecto.....                                        | 18 |
| Ilustración 4: Resultados obtenidos.....                                          | 19 |
| Ilustración 5: Estructura de la memoria .....                                     | 21 |
| Ilustración 6: Diagrama base de datos esquema gov .....                           | 36 |
| Ilustración 7: Diagrama base de datos esquema raw .....                           | 37 |
| Ilustración 8: Diagrama base de datos esquema stg .....                           | 38 |
| Ilustración 9: Diagrama base de datos esquema dwh.....                            | 39 |
| Ilustración 10: ETL P_PARSE_STATS_BOMB .....                                      | 50 |
| Ilustración 11: ETL P_PARSE_STATSBOMB_MATCHES .....                               | 50 |
| Ilustración 12: ETL W_INGEST_EVENTS_AUTO .....                                    | 51 |
| Ilustración 13: ETL W_EVENTS_ONE_BATCH.....                                       | 51 |
| Ilustración 14: ETL P_INGEST_RAW .....                                            | 51 |
| Ilustración 15: ETL P_CHECK_PENDING_EVENTS .....                                  | 52 |
| Ilustración 16: ETL P_PARSE_STATSBOMB_EVENTS .....                                | 52 |
| Ilustración 17: ETL P_INGEST_OPENFOOTBALL_RAW .....                               | 53 |
| Ilustración 18: ETL P_PARSE_OPENFOOTBALL .....                                    | 53 |
| Ilustración 19: ETL W_BUILD_FACT_PLAYER_MATCH_STATS.....                          | 53 |
| Ilustración 20: ETL W_SOCCER_MASTER.....                                          | 54 |
| Ilustración 21: Visualización Positional Performance Analysis.....                | 57 |
| Ilustración 22: Visualización Striker Dashboard (ST) .....                        | 58 |
| Ilustración 23: Visualización Attacking Midfielder / Winger Dashboard (AM/W)..... | 59 |
| Ilustración 24: Visualización Central Midfielder Dashboard (CM) .....             | 60 |
| Ilustración 25: Visualización Defensive Midfielder Dashboard (DM) .....           | 61 |
| Ilustración 26: Visualización Center Back Dashboard (CB) .....                    | 63 |
| Ilustración 27: Visualización Full Back / Wing Back Dashboard (FB/WB).....        | 65 |
| Ilustración 28: Limitaciones del proyecto.....                                    | 78 |
| Ilustración 29: Guía del uso del sistema .....                                    | 80 |
| Ilustración 30: conclusiones del proyecto .....                                   | 85 |

|                                                 |    |
|-------------------------------------------------|----|
| Ilustración 31: Conclusiones del proyecto ..... | 88 |
|-------------------------------------------------|----|

# Índice de Tablas

Tabla 1: Cronograma..... 35

Tabla 2: Conexiones lógicas del modelo de datos ..... 40

Tabla 3: Síntesis métricas por posición transfer score ..... 49

## Capítulo 1. RESUMEN DEL PROYECTO

El mercado de transferencias en el fútbol profesional se ha consolidado como un entorno altamente competitivo y complejo, en el que los clubes buscan maximizar tanto el rendimiento deportivo como la eficiencia económica de sus inversiones. En este contexto, el crecimiento en la disponibilidad de datos deportivos y el desarrollo de técnicas de Big Data han impulsado nuevas formas de análisis orientadas a apoyar los procesos de scouting y evaluación de jugadores (Fernández-Cortés Tolosa, 2024)<sup>1</sup>

Aunque, una parte significativa de la literatura existente se ha centrado en el desarrollo de modelos predictivos destinados a anticipar el rendimiento futuro o el valor de mercado de los jugadores. Si bien estos enfoques aportan valor analítico, su aplicación práctica en entornos reales de toma de decisiones se ve limitada por problemas de interpretabilidad, dependencia de datos propietarios y escasa alineación con las necesidades operativas de los decisores deportivos (Soria Polo, 2021)<sup>2</sup> (Vera Martínez, 2024)<sup>3</sup>.

Ante esta problemática, el presente trabajo propone el diseño e implementación de Scouting tool v2.0, un sistema de apoyo a la toma de decisiones (Decision Support System, DSS) orientado a la evaluación multicriterio de jugadores de fútbol profesional. A diferencia de enfoques predictivos, el sistema desarrollado tiene como objetivo estructurar y apoyar el proceso de scouting mediante la comparación objetiva de jugadores bajo criterios configurables en el proceso. Este planteamiento se alinea con la literatura sobre sistemas de apoyo a la decisión, que enfatiza el uso de herramientas analíticas para reducir la complejidad del proceso decisional sin sustituir el juicio experto (Muñoz Montoro, 2023)<sup>4</sup>.

La solución propuesta se basa en una arquitectura de Big Data compuesta exclusivamente por herramientas gratuitas y de código abierto, integrando procesos de extracción, transformación y carga (ETL), almacenamiento en bases de datos relacionales y visualización interactiva mediante técnicas de visual analytics. Este enfoque permite garantizar la reproducibilidad, transparencia y aplicabilidad práctica del

---

<sup>1</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>2</sup> Soria Polo, C. (2021). *Diseño y aplicación de técnicas de machine learning para optimizar el Scouting en clubes de fútbol*. Sevilla: Universidad de Sevilla.

<sup>3</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>4</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

sistema, facilitando la exploración de distintos escenarios de decisión y reforzando la comprensión de los resultados por parte del usuario (Fernández-cortés Tolosa, 2024)<sup>5</sup>.

En conjunto, el proyecto busca contribuir al ámbito del scouting deportivo mediante el desarrollo de una herramienta operativa que conecte la analítica avanzada con la toma de decisiones reales en el fútbol profesional, complementando el conocimiento experto y promoviendo decisiones más informadas y estructuradas.

## 1.1 Contexto y justificación

El fútbol profesional se ha transformado en una industria global caracterizada por una elevada competitividad, una creciente presión financiera y una constante búsqueda de ventajas deportivas sostenibles. En este contexto, las decisiones relacionadas con la identificación, evaluación y contratación de jugadores representan uno de los procesos estratégicos más relevantes para los clubes, ya que influyen directamente tanto en el desempeño deportivo como en la estabilidad económica de las organizaciones (Fernández-cortés Tolosa, 2024)<sup>6</sup>.

Tradicionalmente, el scouting de jugadores se ha basado en la experiencia y el criterio subjetivo de cuerpos técnicos, directores deportivos y demás personas interesadas. Si bien este conocimiento experto continúa siendo fundamental, la creciente disponibilidad de datos deportivos ha abierto la posibilidad de complementar dichos procesos mediante enfoques analíticos que permitan estructurar y comparar alternativas de forma más objetiva. No obstante, la complejidad inherente a la evaluación de jugadores, que involucra múltiples dimensiones técnicas, tácticas, físicas y contextuales, dificulta la adopción directa de modelos puramente cuantitativos en entornos reales de decisión (Muñoz Montoro, 2023)<sup>7</sup> (Vera Martínez, 2024)<sup>8</sup>.

En los últimos años, la analítica deportiva y las tecnologías de Big Data han sido ampliamente utilizadas para desarrollar modelos predictivos orientados a estimar el rendimiento futuro de los jugadores o su valor de mercado. Aunque estos enfoques han demostrado su utilidad desde una perspectiva académica, su aplicación práctica en los procesos de scouting presenta limitaciones relacionadas con la interpretabilidad de los resultados, la dependencia de datos propietarios y la escasa flexibilidad para adaptarse

---

<sup>5</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>6</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>7</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

<sup>8</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

a criterios específicos definidos por los responsables deportivos (Soria Polo, 2021)<sup>9</sup> (Vera Martínez, 2024)<sup>10</sup>.

Ante esta situación, surge la necesidad de herramientas que buscan apoyar y estructurar la toma de decisiones mediante la integración de datos heterogéneos y la evaluación multicriterio de alternativas. Los sistemas de apoyo a la toma de decisiones (DSS) proporcionan un marco adecuado para abordar este tipo de problemas complejos, ya que permiten combinar información cuantitativa, criterios configurables y visualización interactiva, facilitando una comprensión más clara de los compromisos que se deben aceptar al momento de tomar una decisión cuando se quiere mejorar un criterio (Muñoz Montoro, 2023)<sup>11</sup>.

En este sentido, el desarrollo de Scouting tool v2.0 se justifica como una respuesta a la necesidad de contar con una herramienta operativa, transparente y reproducible que permita comparar jugadores bajo criterios definidos por el usuario, sin depender de predicciones opacas ni de fuentes de datos propietarias. La adopción exclusiva de herramientas gratuitas y de código abierto refuerza, además, la viabilidad técnica y económica del sistema, alineándolo con los principios de reproducibilidad y accesibilidad promovidos en el ámbito académico y profesional (Muñoz Montoro, 2023)<sup>12</sup>.

Desde una perspectiva académica, el proyecto contribuye al campo de la analítica deportiva y los sistemas de apoyo a la decisión al proponer un enfoque alternativo al predominio de modelos predictivos, centrado en la estructuración del proceso decisional. Desde el punto de vista práctico, ofrece un marco aplicable al scouting profesional que puede adaptarse a distintos contextos competitivos y estrategias deportivas.

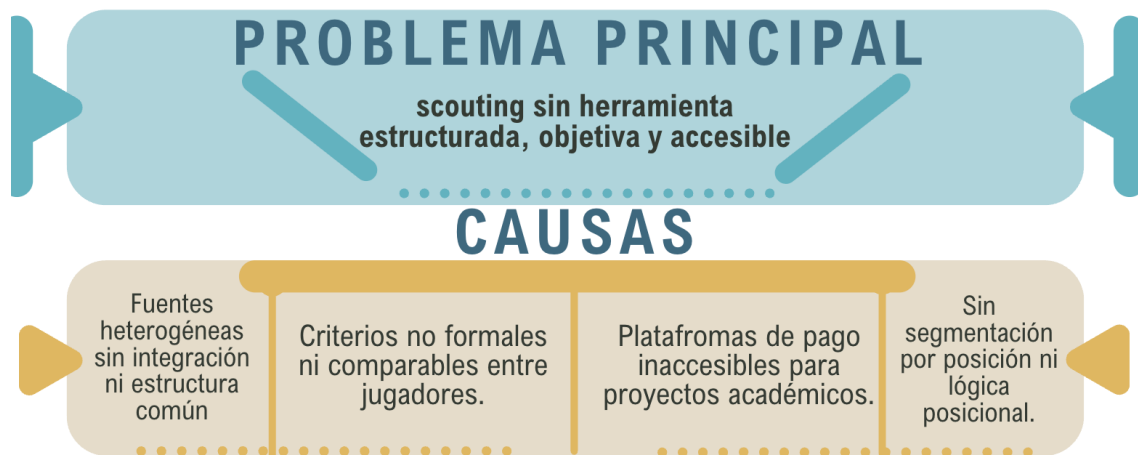
---

<sup>9</sup> Soria Polo, C. (2021). *Diseño y aplicación de técnicas de machine learning para optimizar el Scouting en clubes de fútbol*. Sevilla: Universidad de Sevilla.

<sup>10</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>11</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

<sup>12</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.



*Ilustración 1: Problema y causas del proyecto*

## 1.2 Planteamiento del problema

El proceso de scouting y selección de jugadores en el fútbol profesional constituye una problemática compleja que involucra múltiples criterios deportivos, económicos y contextuales. Los clubes deben evaluar simultáneamente variables relacionadas con el rendimiento individual, la edad, la experiencia competitiva, la versatilidad táctica y las restricciones presupuestarias, entre otros factores. Esta multiplicidad de dimensiones dificulta la comparación objetiva entre jugadores y aumenta el riesgo de decisiones subóptimas cuando el análisis se apoya exclusivamente en la intuición o en evaluaciones parciales.

A pesar del creciente volumen de datos disponibles en el ámbito del fútbol profesional, persisten dificultades para integrar información proveniente de fuentes heterogéneas y transformarla en conocimiento útil para la toma de decisiones. En la práctica, los datos suelen analizarse de forma aislada o mediante herramientas que no permiten una evaluación conjunta, flexible y sistemática de los distintos criterios relevantes. Como consecuencia, el proceso de scouting carece de una estructura formal que facilite la identificación de las mejores alternativas en función de los objetivos estratégicos del club.

Asimismo, muchas de las soluciones analíticas existentes se centran en modelos predictivos orientados a estimar resultados futuros, lo que limita su aplicabilidad en escenarios reales de decisión donde se requiere transparencia, interpretabilidad y capacidad de adaptación a criterios específicos definidos por los responsables deportivos. Estas limitaciones reducen su utilidad como herramientas de apoyo directo en los procesos cotidianos de scouting.

En este contexto, se identifica la ausencia de herramientas operativas que permitan estructurar el proceso de scouting mediante una evaluación multicriterio clara, configurable y basada en datos accesibles. Surge, por tanto, la necesidad de desarrollar un sistema que integre información deportiva relevante y facilite la comparación objetiva de jugadores, permitiendo al decisor explorar distintos escenarios y comprender los compromisos entre criterios sin sustituir el juicio experto.

¿Cómo puede diseñarse un sistema de apoyo a la toma de decisiones, basado en herramientas de Big Data y datos accesibles, que permita estructurar y mejorar el proceso de scouting de jugadores de fútbol profesional mediante una evaluación multicriterio configurable por el usuario?

## CUADRO COMPARATIVO DE SCOUTING

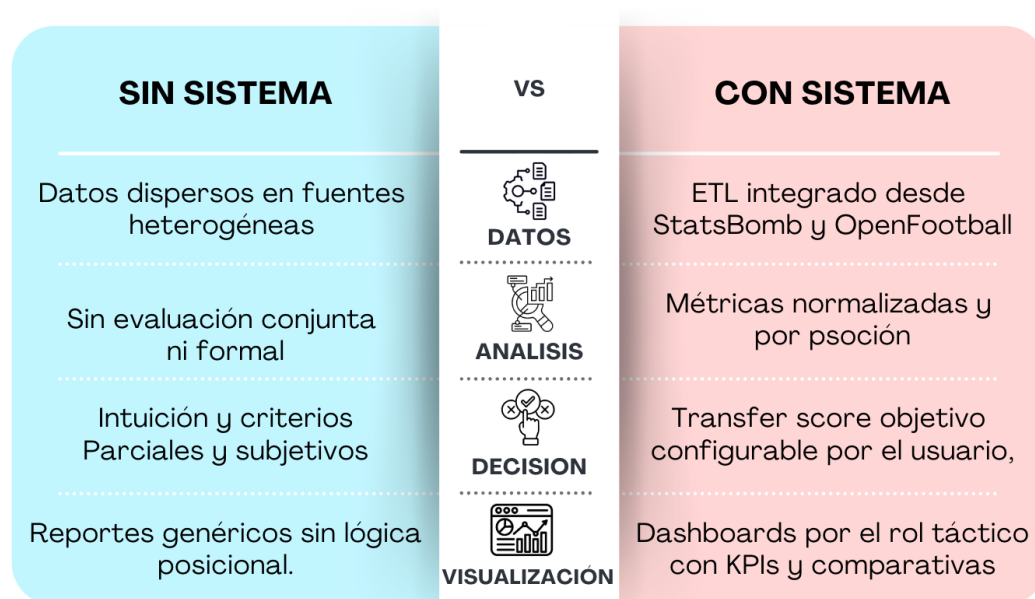


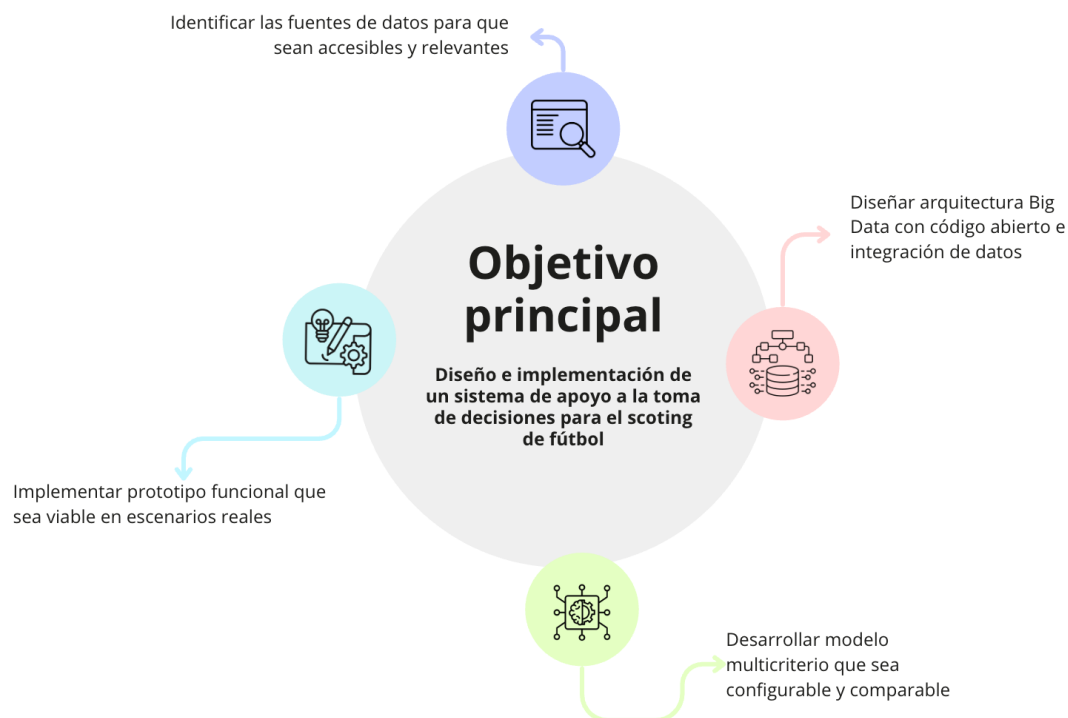
Ilustración 2: Cuadro comparativo de scouting

### 1.3 Objetivos del proyecto

Con la problemática planteada, el objetivo principal de este trabajo es el diseño e implementación de un sistema de apoyo a la toma de decisiones orientado al scouting de jugadores de fútbol profesional. Para ello, el proyecto se centra en la identificación de fuentes de datos accesibles y relevantes, así como en el diseño de una arquitectura

de Big Data basada en herramientas gratuitas y de código abierto que permita la integración, almacenamiento y visualización de la información.

Asimismo, se persigue el desarrollo de un modelo de evaluación multicriterio que facilite la comparación objetiva de jugadores a partir de criterios configurables por el usuario, permitiendo adaptar el análisis a distintos contextos y necesidades deportivas. Finalmente, el proyecto contempla la implementación y validación de un prototipo funcional del sistema propuesto, con el fin de demostrar su utilidad como herramienta de apoyo a la toma de decisiones en escenarios reales de scouting, manteniendo un alcance acorde con la duración y los objetivos del trabajo.



*Ilustración 3: Objetivos del proyecto*

## 1.4 Resultados obtenidos

Los resultados obtenidos en este proyecto fue obtener una arquitectura funcional para la integración, almacenamiento y visualización de datos sobre el fútbol provenientes de StatsBomb OpenFootball. Esta solución permitió construir una base de datos analítica en PostgreSQL, automatizar procesos ETL mediante Apache Hop y desarrollar dashboards en Apache Superset orientados al análisis de jugadores por posición. A nivel

metodológico, el principal resultado fue la construcción de una capa analítica basada en métricas por 90 minutos, indicadores de eficiencia, percentiles posicionales y un puntaje sintético de scouting por cada posición. En conjunto estos elementos dieron lugar a una herramienta de apoyo a la decisión que facilita la comparación objetiva de perfiles de jugadores y demuestra la viabilidad de implementar soluciones de scouting con tecnologías abiertas.



*Ilustración 4: Resultados obtenidos*

## 1.5 Estructura de la memoria

La presente memoria se organiza en varios capítulos que permiten comprender de manera progresiva el desarrollo del proyecto. En primer lugar, se presenta el estado del arte, en el que se recopilan y analizan investigaciones, enfoques y trabajos previos relacionados con la temática abordada, con el fin de ofrecer el contexto necesario y justificar la pertinencia del presente estudio.

A continuación, se expone el capítulo de objetivos, donde se define el objetivo general y los objetivos específicos que orientan el proyecto, así como los principales beneficios esperados de su desarrollo.

Como siguiente punto, es el desarrollo del proyecto, en el cual se describen la metodología utilizada, la planificación temporal, el paso a paso de la implementación, el presupuesto, el análisis de viabilidad y los resultados obtenidos. Este es el núcleo del trabajo, ya que explica cómo se llevó a cabo la propuesta y cuáles fueron sus principales productos y hallazgos.

También se presenta el capítulo de discusión, en el que se analiza e interpreta los resultados obtenidos durante la ejecución del proyecto, permitiendo contrastarlos con los objetivos planteados y valorar su alcance.

A continuación, se mencionan las conclusiones, donde se recogen tanto las conclusiones derivadas del trabajo realizado como las reflexiones surgidas a nivel personal a partir de los procesos de desarrollo. Con lo anterior propuesto, se definen las líneas futuras, orientadas a identificar posibles mejoras, ampliaciones o continuaciones del proyecto.

Finalmente, se incorporan las referencias bibliográficas utilizadas en el documento. Asimismo, en el apartado de anexos se incluyen tres subapartados: la relación del trabajo con los objetivos de desarrollo sostenible, el uso de herramientas de inteligencia artificial durante el desarrollo del proyecto y el enlace correspondiente al desarrollo implementado.

## ESTRUCTURA DE LA MEMORIA



*Ilustración 5: Estructura de la memoria*

## Capítulo 2. ESTADO DEL ARTE

### 2.1 Estado del arte

La nueva era digital en el deporte ha modificado la manera en que se analiza el rendimiento del fútbol profesional, en donde se identifican oportunidades de mejora en el mercado de jugadores. En el caso de este deporte, la disponibilidad creciente de datos abiertos, el desarrollo de herramientas analíticas y la incorporación de técnicas de aprendizaje automático han dado lugar a un campo de investigación que busca estructurar información útil para entrenadores, analistas, directores deportivos y scouts. Como resultado, el scouting ha dejado de ser una práctica de observación por la experiencia de una persona, para convertirse progresivamente en un proceso híbrido que integra juicio experto, con indicadores de rendimiento, procesamiento de datos y visualización (Aditya, 2024)<sup>13</sup> (Muñoz Montoro, 2023)<sup>14</sup> (Vera Martínez, 2024)<sup>15</sup>.

Basado en las investigaciones recientes se detalla que el cambio responde a la convergencia de varias líneas de trabajo. Por un lado, se ha fortalecido la discusión en torno al scouting y a la detección de talento con base en métricas comparables. Hacia otro punto, se ha ampliado un estudio de los indicadores de rendimiento y de las variables que ayudan a contextualizar el comportamiento de un jugador. Adicionalmente, se han desarrollado trabajos en los que emplean machine learning y enfoques cuantitativos para clasificar, estimar o predecir el rendimiento. En este aspecto, también, se observa una tendencia creciente a convertir estos resultados en aplicaciones, dashboards o arquitecturas de datos reproducibles orientadas al apoyo a la decisión (Fernández-Cortés Tolosa, 2024)<sup>16</sup> (Salas Saiz, 2025)<sup>17</sup> (Martínez Santana, 2025)<sup>18</sup>.

#### 2.1.1 Evolución del scouting: del juicio experto al análisis basado en datos

Durante muchos años el scouting se apoyó en la observación directa, la experiencia acumulada del scout y la valoración cualitativa del jugador en situaciones de competencia. Por tal motivo, la interpretación del talento dependía del criterio individual

---

<sup>13</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>14</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

<sup>15</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>16</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>17</sup> Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

<sup>18</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

y de la capacidad del observador para leer el juego. En los últimos años, este modelo empezó a transformarse a medida que los clubes y analistas tuvieran acceso a un volumen mayor de los datos deportivos, tecnologías de registro y herramientas computacionales para procesarlas. La revisión de Aditya (2024)<sup>19</sup> evidencia que la búsqueda de talento y la estandarización de datos de aptitudes físicas en los diferentes clubes de fútbol del mundo se han convertido en una línea de investigación relevante, en la que el big data, la tecnología y el análisis cuantitativo se consideran componentes estratégicos para la competencia entre los clubes.

Otro trabajo aplicado es el de Muñoz Montero (2023)<sup>20</sup>, en este se desarrolla una aplicación orientada al scouting de jugadores con el fin de optimizar el proceso de identificación de talentos, mientras que Vera Martínez (2024)<sup>21</sup>, propone una herramienta que combina el fútbol y ciencia de datos para predecir el rendimiento de futbolistas en distintos contextos. Para ambos trabajos, el scouting deja de entenderse como una actividad de observación y pasa a apoyarse en filtros, métricas comparables y mecanismos de análisis que permiten ampliar y sistematizar la búsqueda de jugadores.

La evolución de este proceso puede describirse como una transición desde el juicio experto hacia modelos híbridos de evaluación. En este escenario, la observación humana sigue siendo necesaria para interpretar aspectos tácticos, comportamentales o contextuales que no siempre quedan completamente capturados en una base de datos. No obstante, la analítica permite reducir parte de la subjetividad, y ampliar el universo de análisis y estructurar procesos de búsqueda más replicables. El valor de la tecnología en scouting radica en dotar al scout de una capa adicional de evidencia para sustentar la toma de decisiones (Vera Martínez, 2024)<sup>22</sup>.

Aunque estos avances han sido importantes, todavía falta una justificación la relevancia del análisis de datos sin desarrollar arquitecturas completas que articulen una captura de información, transformación analítica, comparación por posición y exposición final de resultados. Esta observación es importante porque muestra que la discusión se ha dedicado a resolver problemáticas de cómo convertir los datos en una herramienta

---

<sup>19</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>20</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

<sup>21</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>22</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

integral y operativa para evaluar perfiles de jugadores dentro de contextos reales de decisión (Vera Martínez, 2024)<sup>23</sup>.

### 2.1.2 Indicadores de rendimiento y comparabilidad de métricas en el fútbol

En esta segunda línea del estado del arte, se centra en los indicadores de rendimiento y en los criterios utilizados para evaluar jugadores. Fernández-Cortés Tolosa (2024)<sup>24</sup> ubica este problema al momento de analizar y evaluar indicadores de rendimiento en el fútbol, además de revisar sus diferencias, alcances y limitaciones. Esta orientación resulta clave, ya que muestra que el problema consiste en seleccionar, organizar e interpretar aquellos indicadores que realmente representan el comportamiento competitivo del jugador. En esta misma dirección, Salas Saiz (2025)<sup>25</sup> aborda el rendimiento en el fútbol profesional con apoyo del aprendizaje automático, lo que confirma el interés reciente por convertir eventos y estadísticas en variables útiles para análisis estructurados.

El rendimiento en el fútbol es multidimensional, es decir, no siempre puede resumirse con solo estadística, como goles o asistencias, también depende de conjuntos de variables técnicas, tácticas, físicas y situacionales. Martínez Sánchez (2024)<sup>26</sup>, se centra en los factores determinantes para el rendimiento deportivo, refuerza esta idea al estudiar la relación entre distintas variables y la producción ofensiva. De forma complementaria, varios trabajos recientes muestran que el rendimiento cambia según el contexto del partido, la posición del jugador, la competencia o la situación táctica, por lo que una lectura válida exige integrar más de una dimensión de análisis (Fernández-Cortés Tolosa, 2024)<sup>27</sup> (Martínez Santana, 2025)<sup>28</sup>.

En este apartado, la comparación de las métricas se convierte en una condición metodológica central. Las estadísticas solo resultan realmente útiles cuando pueden contrastarse entre jugadores de forma válida. Para ello, recientemente sugieren ajustar los indicadores por tiempo de juego, interpretar los resultados dentro de los grupos

---

<sup>23</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>24</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>25</sup> Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

<sup>26</sup> Martínez Sánchez, J. A. (2024). *Factores determinantes del rendimiento deportivo en fútbol*. Sevilla: Universidad Pablo de Olavide.

<sup>27</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>28</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

equivalentes. Aditya (2024)<sup>29</sup> resalta precisamente la importancia de la estandarización de los datos para los procesos de scouting, mientras Fernández-Cortés Tolosa (2024)<sup>30</sup> enfatiza la necesidad de revisar semejanzas y diferencias entre los indicadores empleados en la evaluación del rendimiento. Como consecuencia, este proyecto se apoya en esa discusión para construir una capa analítica que transforme eventos de partido en métricas comparables y útiles para scouting, en lugar de limitarse a estadísticas absolutas aisladas (Aditya, 2024)<sup>31</sup> (Salas Saiz, 2025)<sup>32</sup>.

### 2.1.3 Evaluación de jugadores con enfoques multicriterio y machine learning

En este apartado se desarrolla métodos cuantitativos, minería de datos y machine learning a la evaluación de jugadores. Aquí se detalla que el análisis del fútbol ha comenzado a desplazarse desde la simple descripción de estadísticas hacia la construcción de modelos capaces de clasificar, sintetizar o predecir el comportamiento del jugador. Vera Martínez (2024)<sup>33</sup>, realiza una herramienta con capacidad predictiva para estimar el rendimiento en distintos contextos. También se encuentra Salas Saiz (2025)<sup>34</sup> en donde se apoya en el aprendizaje automático para analizar el rendimiento profesional, mientras que Moro Portero (2025)<sup>35</sup> centra su trabajo en la predicción del potencial de jugadores mediante técnicas de machine learning. En conjunto, estos antecedentes evidencian una tendencia clara hacia la modelación cuantitativa del talento.

En muchos trabajos e investigaciones recientes se sitúa en el terreno de la predicción, como predecir el rendimiento, estimar el potencial o modelar ratings. El proyecto desarrollado en este texto se centra en un sistema de apoyo a la decisión basado en comparación multicriterio. Su objetivo principal es estructurar una evaluación interpretable y utilizable que permita comparar perfiles dentro del mercado de transferencias. Por esta razón, los antecedentes de machine learning son relevantes como marco metodológico y como contraste, pero también ayudan a delimitar el aporte

---

<sup>29</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>30</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>31</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>32</sup> Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

<sup>33</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>34</sup> Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

<sup>35</sup> Moro Portero, T. (2024). *Predicción del potencial de jugadores de fútbol mediante técnicas de machine learning*. Madrid: Universidad Politécnica de Madrid.

específico del trabajo: una solución que prioriza comparabilidad e interpretabilidad sobre predicción pura (Salas Saiz, 2025)<sup>36</sup> (Moro Portero, 2024)<sup>37</sup>.

La analítica avanzada y el machine learning han demostrado ser útiles para el fútbol profesional; sin embargo, sigue existiendo espacio para herramientas que permitan traducir el desempeño del jugador a métricas y puntajes comprensibles para el proceso de scouting. En otras palabras, el proyecto se posiciona en una zona intermedia entre el análisis descriptivo y modelos predictivos, articulando una estructura cuantitativa e interpretación práctica para la toma de decisiones (Vera Martínez, 2024)<sup>38</sup> (García González, 2025)<sup>39</sup>.

#### **2.1.4 Herramientas y dashboards para scouting y análisis del fútbol**

En este apartado se ubican trabajos que buscan trasladar las métricas, modelos y resultados analíticos a entornos visuales o aplicaciones concretas que pueden ser utilizadas por scouts, analistas o representantes deportivos. Muñoz Montoro (2023)<sup>40</sup>, orienta su trabajo al desarrollo de una aplicación de scouting para optimizar la identificación de talentos, mientras que Martínez Santana (2025)<sup>41</sup> diseña e implementa un dashboard para el análisis de la primera división española, integrando procesos de tratamiento de datos y explotación visual de indicadores. Estos antecedentes muestran que la investigación aplicada busca convertirlos en herramientas de consulta, para producir resultados analíticos.

En este mismo camino se encuentra Cuadrado Camarero (2025)<sup>42</sup>, quien combina análisis predictivo y visualización de datos en el fútbol usando inteligencia artificial. Su propuesta es relevante porque se evidencia que la capa mediante la cual el usuario ingresa, interpreta y compara información compleja ayuda a la toma de decisiones. De esta forma complementaria, Barrera León (2025)<sup>43</sup>, plantea el uso de la analítica de datos e inteligencia artificial para optimizar procesos de gestión deportiva en clubes del

---

<sup>36</sup> Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

<sup>37</sup> Moro Portero, T. (2024). *Predicción del potencial de jugadores de fútbol mediante técnicas de machine learning*. Madrid: Universidad Politécnica de Madrid.

<sup>38</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>39</sup> García González, J. J. (2025). *Evaluación y Modelado del Sistema de Rating en FIFA*. Madrid: Universidad Politécnica de Madrid.

<sup>40</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

<sup>41</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>42</sup> Cuadrado Camarero, E. (2025). *Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial*. Madrid: Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial.

<sup>43</sup> Barrera León, D. (2025). *Empleo de analítica de datos e IA para optimizar procesos de gestión deportiva en clubes de fútbol de Colombia*. Bogotá: Universidad de Ciencias Aplicadas y Ambientales.

fútbol colombiano, lo que se confirma que la aplicación práctica de estas herramientas ya se está extendiendo en distintos contextos institucionales del deporte. En conjunto, estos trabajos reflejan una tendencia clara hacia la construcción de productos analíticos orientados a la usabilidad.

Sin embargo, en los trabajos e investigaciones se revela que hay una limitante, en donde muchas propuestas se concentran en una sola capa del problema, ya sea la interfaz, otros el modelo predictivo, o la experiencia de consulta. Lo menos frecuente sigue siendo encontrar trabajos que integren, de forma coherente y documentada, la extracción automatizada de fuentes, el almacenamiento estructurado, la construcción de métricas comparables por posición y la exposición final de resultados en un dashboard orientado específicamente al scouting. Esta observación es un punto clave para el desarrollo de este proyecto, ya que, permite mostrar que su aporte está en la unión entre arquitectura de datos y la capa visual (Martínez Santana, 2025)<sup>44</sup> (Cuadrado Camarero, 2025)<sup>45</sup>.

La importancia de diseñar visualizaciones que sean consistentes con la lógica del análisis. Un dashboard útil es la suma de gráficos, ser una interfaz que organiza indicadores, permitir contrastes que facilitan una lectura comparativa del desempeño. Por ello, la construcción de dashboards especializados por posición, mostrar los KPI, indicadores en diferentes gráficas, se debe a una decisión del proceso de traducción del análisis a una herramienta de apoyo a la decisión (Martínez Santana, 2025)<sup>46</sup> (Barrera León, 2025)<sup>47</sup>.

### **2.1.5 Datos abiertos, arquitectura analítica y reproducibilidad**

Para este punto se debe mencionar a Santana Falcón (2023)<sup>48</sup> resulta relevante porque analiza los datos abiertos del fútbol aplicando la metodología CRIPS-DM y enfatiza la necesidad de transformar grandes volúmenes de datos de información en conocimiento útil. Su propuesta presenta atención al proceso completo: identificación de fuentes, preparación, transformación, análisis e interpretación. Este enfoque es pertinente para el presente trabajo, dado que una parte importante de su contribución radica tanto en

---

<sup>44</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>45</sup> Cuadrado Camarero, E. (2025). *Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial*. Madrid: Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial.

<sup>46</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>47</sup> Barrera León, D. (2025). *Empleo de analítica de datos e IA para optimizar procesos de gestión deportiva en clubes de fútbol de Colombia*. Bogotá: Universidad de Ciencias Aplicadas y Ambientales.

<sup>48</sup> Santana Falcón, F. (2023). *Análisis de datos abiertos de fútbol aplicando la metodología CRISP-DM*. Madrid: Universidad Politécnica de Madrid.

las métricas, en la visualización, como también lo hace en la arquitectura que hace posible el tratamiento sistemático del dato.

Muchos de los trabajos consultados que aplican la analítica deportiva, muestran dashboards, modelos o resultados finales, pero no siempre documentan con detalle la secuencia que conecta la fuente de datos con la capa analítica y visualización. Por eso los trabajos de Martínez Santana (2025)<sup>49</sup> y Enríquez Herrera (2025)<sup>50</sup> son útiles porque evidencian la importancia de procesos ETL, modelos de almacenamiento y dashboards en la consolidación de soluciones orientadas a la toma de decisiones. Aunque el trabajo de Enríquez Herrera (2023)<sup>51</sup> se sitúa en entender que una arquitectura analítica bien definida es condición para generar información oportuna y útil para los usuarios finales.

Los datos abiertos son un rasgo distintivo de la literatura reciente. La disponibilidad de fuentes accesibles ha permitido que estudiantes, investigadores y analistas construyan propuestas aplicadas sin depender exclusivamente de bases propietarias cerradas. Esto ha favorecido el desarrollo de herramientas, comparaciones y aplicaciones de scouting con mayor transferencias y posibilidad de réplica. En consecuencia, el uso de datos abiertos fortalece el carácter de reproducir las investigaciones y facilita su validación por terceros. (Santana Falcón, 2023)<sup>52</sup>.

Este enfoque permite diferenciar el aporte del presente proyecto frente a otros trabajos que se concentran únicamente en una parte del proceso, ya sea el análisis de datos, el modelado o la visualización final. En contraste, esta propuesta se formula de manera integral una arquitectura abierta, trazable y reproducible con una evaluación multicriterio de jugadores, lo que permite transformar datos abiertos en información comparable y útil para la toma de decisiones en scouting. De esta forma se responde tanto al problema técnico de organizar, procesar y visualizar la información, como a la necesidad de estructurar un mecanismo de evaluación más objetivo, consistente y sustentado en evidencia, contribuyendo así a resolver la pregunta de investigación desde una

---

<sup>49</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>50</sup> Enríquez Herrera, J. V. (2023). Business Intelligence & Data Analytics aplicado al proceso de seguimiento curricular en la universidad UPEC. *Minerva*, 9-20.

<sup>51</sup> Enríquez Herrera, J. V. (2023). Business Intelligence & Data Analytics aplicado al proceso de seguimiento curricular en la universidad UPEC. *Minerva*, 9-20.

<sup>52</sup> Santana Falcón, F. (2023). *Análisis de datos abiertos de fútbol aplicando la metodología CRISP-DM*. Madrid: Universidad Politécnica de Madrid.

perspectiva metodológica, analítica y aplicada. (Enríquez Herrera, 2023)<sup>53</sup> (Santana Falcón, 2023)<sup>54</sup>.

## 2.2 Contexto y justificación

La literatura reciente confirma que el análisis del fútbol ha avanzado notablemente en varias dimensiones. La revisión de Aditya (2024)<sup>55</sup>, la tesis de Fernández-Cortés Tolosa (2024)<sup>56</sup>, los desarrollos aplicados de Muñoz Montero (2023)<sup>57</sup>, Vera Martínez (2024)<sup>58</sup>, Salas Saiz (2025)<sup>59</sup>, Martínez Santana (2025)<sup>60</sup>, Santana Falcón (2023)<sup>61</sup>, así como los trabajos de machine learning y visualización revisados, muestran que el fútbol ya constituye a un dominio consolidado para la ciencia de datos, la evaluación cuantitativa y el diseño de herramientas analíticas. El jugador ha comenzado a ser entendido como un objeto de análisis medible, comparable y predecible. El scouting se ha enriquecido con métodos que permiten estructurar la búsqueda de talento con criterios más sistemáticos (Aditya, 2024)<sup>62</sup> (Fernández-Cortés Tolosa, 2024)<sup>63</sup>.

Sin embargo, se evidencia una limitación, la cual es que los estudios revisados abordan solo una dimensión específica del problema: algunos se enfocan en la predicción del rendimiento o del valor de mercado, otros en la construcción de indicadores, y otros en el diseño de aplicaciones o dashboards. En conclusión, son menos frecuentes las propuestas que integran de manera coherente y documentada fuentes de datos abiertas, ETL, almacenamiento relacional con enfoque analítico, métricas comparables por posición y visualización orientada explícitamente al apoyo a la toma de decisiones en scouting y mercado de transferencias. Esta brecha justifica el desarrollo del presente

---

<sup>53</sup> Enríquez Herrera, J. V. (2023). Business Intelligence & Data Analytics aplicado al proceso de seguimiento curricular en la universidad UPEC. *Minerva*, 9-20.

<sup>54</sup> Santana Falcón, F. (2023). *Análisis de datos abiertos de fútbol aplicando la metodología CRISP-DM*. Madrid: Universidad Politécnica de Madrid.

<sup>55</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>56</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>57</sup> Muñoz Montero, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

<sup>58</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>59</sup> Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

<sup>60</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>61</sup> Santana Falcón, F. (2023). *Análisis de datos abiertos de fútbol aplicando la metodología CRISP-DM*. Madrid: Universidad Politécnica de Madrid.

<sup>62</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>63</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

proyecto, cuyo aporte se restringe a la búsqueda de una solución tecnológica integral (Martínez Santana, 2025)<sup>64</sup> (Santana Falcón, 2023)<sup>65</sup>.

Este trabajo se sitúa en un punto intermedio entre las investigaciones centradas exclusivamente en modelos predictivos y aquellas orientadas únicamente a la visualización. Su propósito es integrar datos abiertos, una arquitectura analítica reproducible, métricas comparables por posición y un conjunto de dashboards especializados que conviertan el análisis en una herramienta operativa de apoyo a la decisión. De esta manera, el proyecto responde a una necesidad identificada en la literatura reciente: contar con soluciones que además de generar conocimiento analítico sobre el rendimiento de los jugadores, también permita utilizar dicho conocimiento de forma estructurada, interpretable y aplicable en contextos reales de scouting y evaluación del mercado de transferencias.

## 2.3 Planteamiento del problema

En el fútbol profesional, la toma de decisiones asociada al scouting y al mercado de transferencias exige cada vez más el uso de herramientas que permitan evaluar a los jugadores de manera objetiva, comparable y estructurada. La identificación del talento ya no es únicamente una observación objetiva, se debe tener soporte de la capacidad de integrar y transformar datos en información útil para apoyar las decisiones deportivas y económicas.

El análisis en los apartados anteriores sobre el estado del arte permite observar que, en los últimos años, el estudio del rendimiento del fútbol ha avanzado con propuestas centradas en la construcción de métricas, en la predicción del desempeño, en el uso de técnicas de aprendizaje automático y en el desarrollo de dashboards o herramientas visuales para el análisis de jugadores y equipos. Sin embargo, estos avances suelen presentarse de manera fragmentada, abordando solo una parte del problema.

En muchos casos, las soluciones existentes se concentran en una dimensión específica del problema, ya sea en el análisis estadístico, la predicción o en la visualización, pero no integran de manera completa el ciclo de adquisición del dato, la transformación, almacenamiento y explotación analítica Vera Martínez (2024)<sup>66</sup>, Cuadrado Camero

---

<sup>64</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>65</sup> Santana Falcón, F. (2023). *Análisis de datos abiertos de fútbol aplicando la metodología CRISP-DM*. Madrid: Universidad Politécnica de Madrid.

<sup>66</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

(2025)<sup>67</sup>. De igual forma, no siempre ofrecen criterios comparables por posición o una arquitectura reproducible que permita convertir datos abiertos en una herramienta operativa de apoyo en las decisiones Muñoz Montoro (2023)<sup>68</sup>, Martínez Santana (2025)<sup>69</sup>, Santana Falcón (2023)<sup>70</sup>. Esta situación genera una brecha entre el potencial analítico de los datos en el medio del fútbol y su aplicación práctica en contextos reales de las transferencias del fútbol. (Enríquez Herrera, 2023)<sup>71</sup> (Aditya, 2024)<sup>72</sup>.

Para mitigar estas necesidades, el problema que aborda este trabajo se centra en la falta de una solución integral que articule fuentes de datos abiertas, procesos ETL, almacenamiento relacional analítico, métricas comparables por posición y visualización, con el fin de apoyar de manera estructurada la evaluación multicriterio de futbolistas profesionales. Aún se identifican necesidades de una herramienta que conecte de forma coherente la ingeniería de datos con el proceso de scouting y traduzca el análisis cuantitativo en un soporte funcional para la toma de decisiones.

El presente trabajo busca responder a la necesidad de disponer de una solución tecnológica reproducible que permita organizar, procesar y visualizar información sobre el fútbol de manera estructurada, facilitando la comparación de jugadores según su posición y desempeño, para fortalecer los procesos de evaluación asociados al scouting.

---

<sup>67</sup> Cuadrado Camarero, E. (2025). *Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial*. Madrid: Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial.

<sup>68</sup> Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

<sup>69</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>70</sup> Santana Falcón, F. (2023). *Análisis de datos abiertos de fútbol aplicando la metodología CRISP-DM*. Madrid: Universidad Politécnica de Madrid.

<sup>71</sup> Enríquez Herrera, J. V. (2023). Business Intelligence & Data Analytics aplicado al proceso de seguimiento curricular en la universidad UPEC. *Minerva*, 9-20.

<sup>72</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

## Capítulo 3. OBJETIVOS

### 3.1 Objetivos generales

Diseñar un sistema de apoyo a la toma de decisiones basado en herramientas de Big Data y datos accesibles, que permita estructurar y mejorar el proceso de scouting de jugadores de fútbol profesional mediante una evaluación multicriterio.

### 3.2 Objetivos específicos

- Seleccionar fuentes de datos accesibles y relevantes para la evaluación de jugadores de fútbol profesional, garantizando su coherencia y viabilidad para el análisis.
- Diseñar una arquitectura de Big Data basada en herramientas gratuitas y de código abierto, que integre procesos ETL, almacenamiento en bases de datos y visualización.
- Desarrollar un modelo de evaluación multicriterio que permita comparar jugadores de manera objetiva según criterios definidos y ponderados.
- Implementar un prototipo funcional del sistema DSS (decision support system), demostrando su utilidad como herramienta de apoyo a la toma de decisiones en escenarios de scouting.

### 3.3 Beneficios del proyecto

Este proyecto aporta beneficios tanto nivel técnico como funcional. En primer lugar, permite estructurar y centralizar información del fútbol provenientes de fuentes abiertas mediante una arquitectura reproducible basada en procesos ETL, almacenamiento relacional y visualización. Esto facilita la transformación de datos dispersos en información organizada, comparable y útil para el análisis, reduciendo la dependencia de revisiones manuales y mejorando la disponibilidad de métricas para la evaluación de jugadores. Además, el uso de herramientas de código abierto favorece la accesibilidad, la transparencia del proceso y la posibilidad de replicar o ampliar la solución en otros contextos.

En segundo lugar, se ofrece un beneficio aplicado al scouting y al mercado de transferencias, al proporcionar una herramienta de apoyo a la decisión que permite comparar futbolistas de acuerdo con métricas específicas por posición. De esta manera, se fortalece la evaluación multicriterio de jugadores y se contribuye a que los procesos de análisis sean más objetivos, interpretables y consistentes. En conjunto, la propuesta genera una solución tecnológica funcional, que demuestra como la integración entre

Diseño e implementación de un sistema de apoyo a la toma de decisiones para el scouting de jugadores mediante herramientas Big Data.

Andrea Ruiz Naranjo

---

ingeniería de datos, analítica y visualización puede traducirse en un apoyo practico para la toma de decisiones en el ámbito del futbol profesional.

## Capítulo 4. DESARROLLO DEL PROYECTO

### 4.1 Planificación del proyecto

La planificación del proyecto se ha estructurado considerando una duración total de cinco meses, organizando las actividades de forma secuencial y parcialmente iterativa, de acuerdo con la evolución natural del planteamiento metodológico del trabajo.

En una primera fase, se llevó a cabo una exploración preliminar de los datos, con el objetivo de identificar la existencia de fuentes públicas relacionadas con el scouting y la evaluación de jugadores de fútbol profesional. Esta etapa permitió analizar la disponibilidad, calidad y accesibilidad de los datos, así como evaluar su potencial uso dentro del proyecto.

A partir de esta exploración inicial, se analizaron distintas alternativas para abordar el problema de investigación. En una primera instancia, se consideró una aproximación basada en técnicas de aprendizaje automático orientadas a la predicción. Sin embargo, tras evaluar sus implicaciones técnicas y su aplicabilidad práctica, se decidió cambiar el enfoque del proyecto hacia el uso de herramientas de Big Data y sistemas de apoyo a la toma de decisiones, con el fin de desarrollar una solución más transparente, interpretable y accesible para usuarios sin conocimientos técnicos avanzados.

Una vez definido el enfoque definitivo, se procedió al planteamiento del problema y a la definición de los objetivos del proyecto, delimitando claramente el alcance del sistema propuesto. Posteriormente, se abordó el desarrollo del producto, incluyendo el diseño de la arquitectura del sistema, la implementación del prototipo funcional y la definición de los mecanismos de validación.

De forma paralela al desarrollo técnico, se elaboró el estado del arte y los restantes apartados del documento, asegurando la coherencia entre los fundamentos teóricos y la solución propuesta. Finalmente, se realizó la validación del sistema mediante casos de uso representativos y se completó la redacción final de la memoria.

A continuación, se presenta el cronograma (Tabla 1) resumido de actividades del proyecto:

| FASE          | ACTIVIDADES PRINCIPALES                                           | DURACIÓN  |
|---------------|-------------------------------------------------------------------|-----------|
| <b>FASE 1</b> | Exploración de datos públicos y análisis de viabilidad            | 3 semanas |
| <b>FASE 2</b> | Análisis de alternativas metodológicas y redefinición del enfoque | 3 semanas |

Andrea Ruiz Naranjo

|               |                                                        |           |
|---------------|--------------------------------------------------------|-----------|
| <b>FASE 3</b> | Planteamiento del problema y definición de objetivos   | 2 semanas |
| <b>FASE 4</b> | Diseño de la arquitectura del sistema DSS              | 3 semanas |
| <b>FASE 5</b> | Desarrollo del prototipo y procesos ETL                | 5 semanas |
| <b>FASE 6</b> | Desarrollo de visualizaciones y validación del sistema | 3 semanas |
| <b>FASE 7</b> | Estado del arte y redacción de la memoria final        | 3 semanas |

Tabla 1: Cronograma

## 4.2 Descripción de la solución, metodologías y herramientas empleadas

La solución tecnológica se estructuró como una arquitectura de datos orientada a la integración, transformación y explotación analítica de información del fútbol proveniente de dos fuentes abiertas complementarias: StatsBomb y OpenFootball. La primera fuente, aporta datos de alta granularidad a nivel de partido y evento, mientras que OpenFootball suministra calendarios y resultados históricos organizados por liga y temporada. La implementación se realizó con Apache Hop como herramienta de orquestación ETL, PostgreSQL como sistema de almacenamiento relacional y motor de transformación analítica, y Apache Superset como capa de visualización. Bajo este enfoque, Hop coordina la ejecución de pipelines y workflows; PostgreSQL concentra la persistencia, limpieza, normalización y cálculo de métricas; y Superset convierte la capa analítica en dashboards orientados a la evaluación de jugadores. (Aditya, 2024)<sup>73</sup>

### 4.2.1 Diseño e implementación de la base de datos

#### 4.2.1.1 Propósito y criterios de diseño

La base de datos se implementó en PostgreSQL como núcleo técnico del sistema. Su diseño respondió a la necesidad de soportar el ciclo completo del dato, desde la captura inicial de respuestas externas hasta la consolidación de indicadores comparables para el análisis de jugadores. En este contexto, la base de datos se desarrolló como una arquitectura analítica progresiva capaz de transformar eventos de partidos en métricas de desempeño, medidas de eficiencia, comparaciones relativas y puntajes sintéticos de evaluación. (Enríquez Herrera, 2023)<sup>74</sup>

<sup>73</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>74</sup> Enríquez Herrera, J. V. (2023). Business Intelligence & Data Analytics aplicado al proceso de seguimiento curricular en la universidad UPEC. *Minerva*, 9-20.

Andrea Ruiz Naranjo

El diseño se orientó por cuatro criterios principales. El primero fue la trazabilidad, con el fin de conservar el origen de cada recurso descargado y el estado de cada proceso. El segundo fue la separación funcional, de modo que el dato crudo, el dato en preparación y el dato analítico no coexistieran dentro de una misma capa lógica. El tercero fue la reestructuración progresiva, es decir, la posibilidad de transformar el dato en distintas etapas hasta alcanzar una forma explotable para análisis. El cuarto fue la capacidad del sistema para ejecutarse en diferentes ocasiones sin generar duplicados ni inconsistencias estructurales. Estos criterios se materializaron en una arquitectura por capas organizada en los esquemas gov, raw, stg y dwh. (Pinargote Cárdenas & Alcívar Cedeño, 2025)<sup>75</sup>

**4.2.1.2 Arquitectura por capas**

El esquema gov como se evidencia en la ilustración 6 se diseñó como capa de gobierno y control operacional del proceso de ingesta. En este esquema se ubica la tabla gov.soccer\_driver, que funciona como tabla conductora de extracción. En ella se registran la fuente, la liga, la temporada, el endpoint a consumir, la URL específica, el estado de ejecución y los mensajes de error asociados. Su función principal es desacoplar la lógica de extracción de la lógica de transformación, de manera que cada recurso pueda ser tratado como una unidad independiente dentro del flujo.

| soccer_driver                                 | statsbomb_matches_driver |
|-----------------------------------------------|--------------------------|
| 123 driver_id                                 | 123 raw_id               |
| A-Z source                                    | 123 competition_id       |
| A-Z league_name                               | 123 season_id            |
| A-Z season_label                              | A-Z competition_name     |
| A-Z endpoint                                  | A-Z season_name          |
| A-Z url                                       | A-Z country_name         |
| A-Z format                                    | A-Z url                  |
| <input checked="" type="checkbox"/> is_active | A-Z status               |
| A-Z status                                    | 123 attempts             |
| 🕒 last_run_at                                 | 🕒 last_run_at            |
| A-Z error_msg                                 | A-Z last_error           |

Ilustración 6: Diagrama base de datos esquema gov

El esquema raw (Ilustración 7) se definió como la capa de persistencia de las respuestas obtenidas desde las fuentes externas. La tabla raw.http\_responses conserva el identificador del driver, la URL consultada, el código de respuesta, el contenido descargado y la fecha de ingesta. Su función es documental y operativa: almacenar la

<sup>75</sup> inargote Cárdenas, R. E., & Alcívar Cedeño, A. (2025). *Analítica de Datos e Indicadores de Rendimiento para optimizar la gestión Universitaria*. Sinapsis.

Andrea Ruiz Naranjo

---

evidencia de cada descarga para tener la auditoría del proceso y reprocesamiento de transformaciones sin necesidad de volver a consultar la fuente original.

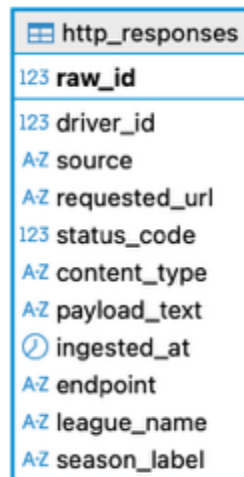


Diagrama base de datos esquema raw. El diagrama muestra una colección de campos con sus respectivos tipos de datos y atributos. El campo 'raw\_id' es el campo principal (PK) y tiene un tipo de dato de 123. Los campos 'driver\_id' y 'status\_code' también tienen un tipo de dato de 123. Los campos 'source', 'requested\_url', 'content\_type', 'payload\_text', 'endpoint', 'league\_name' y 'season\_label' tienen un tipo de dato de A-Z. El campo 'ingested\_at' tiene un tipo de dato de fecha y hora (🕒).

| http_responses    |
|-------------------|
| 123 raw_id        |
| 123 driver_id     |
| A-Z source        |
| A-Z requested_url |
| 123 status_code   |
| A-Z content_type  |
| A-Z payload_text  |
| 🕒 ingested_at     |
| A-Z endpoint      |
| A-Z league_name   |
| A-Z season_label  |

*Ilustración 7: Diagrama base de datos esquema raw*

El esquema stg (Ilustración 8) se empleó como zona intermedia de preparación. En esta capa se almacenan objetos semiestructurados ya asociados a un contexto lógico, por ejemplo, competiciones, partidos y eventos de StatsBomb, así como temporadas de OpenFootball. El propósito de stg es separar el parseo y la limpieza respecto del modelo final de consulta. Esta división permite revisar la calidad de las estructuras intermedias, corregir transformaciones y depurar el flujo antes de consolidar la información en el esquema analítico final.

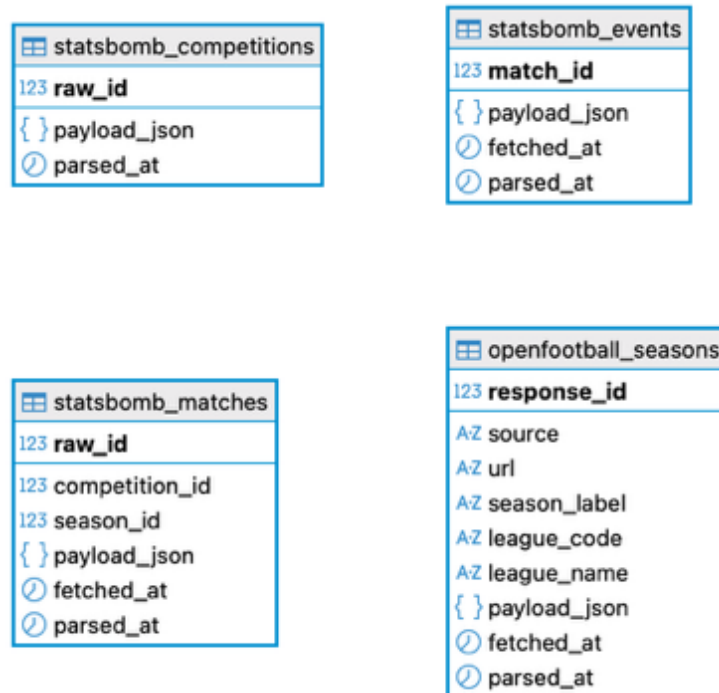


Ilustración 8: Diagrama base de datos esquema stg

El esquema dwh (Ilustración 9) concentra la capa analítica del sistema. En él se encuentran las tablas normalizadas y las vistas para el análisis y la visualización, entre ellas `dwh.statsbomb_matches_norm`, `dwh.statsbomb_events`, `dwh.openfootball_matches` y `dwh.fact_player_match_stats`. Esta capa reúne la versión final explotable del dato y constituye el soporte principal para la construcción de métricas, vistas comparativas y dashboards de scouting.



Ilustración 9: Diagrama base de datos esquema dwh

#### 4.2.1.3 Conexiones lógicas del modelo

| Tabla o esquema origen      | tabla o esquema destino     | Propósito                                       |
|-----------------------------|-----------------------------|-------------------------------------------------|
| gov.soccer_driver           | raw.http_responses          | Controlar y registrar la descarga de datos      |
| raw.http_responses          | stg.*                       | Clasificar y preparar la información descargada |
| stg.*                       | dwh.*                       | Normalizar los datos para análisis              |
| dwh.statsbomb_events        | dwh.fact_player_match_stats | Calcular estadísticas de jugador por partido    |
| dwh.fact_player_match_stats | vistas analíticas           | Resumir, comparar y evaluar jugadores           |

*Tabla 2: Conexiones lógicas del modelo de datos*

La arquitectura de datos del proyecto responde a una lógica secuencial de procesamiento en la que cada capa cumple una función específica dentro del ciclo de vida del dato. La capa gov concentra los mecanismos de control de la extracción, permitiendo definir y monitorear los recursos a consultar. La capa raw conserva las respuestas originales descargadas desde las fuentes, asegurando trazabilidad y la posibilidad de reprocesamiento. La capa stg actúa como una zona intermedia de preparación, donde la información se organiza y clasifica de acuerdo con su tipo de procedencia. Finalmente, la capa dwh constituye el núcleo analítico del sistema, al reunir tablas normalizadas y vistas derivadas orientadas a la generación de métricas, comparaciones y puntajes de evaluación.

En este contexto, las conexiones entre los objetos dependen de relaciones lógicas establecidas mediante identificadores comunes, procesos ETL y transformaciones SQL, mas no de claves foráneas. Esta estructura permite comprender la base de datos como una arquitectura integrada que soporta la transición desde el dato bruto hasta la información analítica utilizada en la evaluación comparativa de jugadores.

Asimismo, es importante señalar que la ausencia de conexiones físicas visibles en las tablas de cada esquema obedece a la forma en que fue diseñado el modelo. En las capas raw y stg, la prioridad fue garantizar la flexibilidad para la carga, reprocesamiento y transformación de datos provenientes de fuentes heterogéneas. En este tipo de arquitecturas, muchas relaciones se establecen de manera lógica a través de campos

compartidos como identificadores de respuesta, partidos o jugadores, y se materializan posteriormente en consultas SQL, agregaciones y vistas analíticas. De igual manera, en la capa dwh una parte importante de la estructura está compuesta por vistas, las cuales dependen funcionalmente de otras tablas y otras vistas, estas no se representan como relaciones físicas del mismo modo que una clave foránea. Por esta razón, aunque el diagrama no muestre líneas de conexión entre todos los objetos, si existe una secuencia relacional y funcional clara dentro del flujo del proyecto.

#### **4.2.1.4 Tablas principales del modelo**

**Tabla gov.soccer\_driver:** Esta tabla actúa como driver de extracción y constituye el punto de control de la arquitectura. Cada fila representa un recurso a consultar y contiene la fuente, la liga, la temporada, el endpoint, la URL, el estado de procesamiento y la traza de error, cuando aplica. Su función es gobernar el flujo de extracción y servir como referencia para la orquestación de las cargas.

**Tabla raw.http\_responses:** Almacena la respuesta HTTP asociada a cada recurso descargado. Contiene el driver\_id, la URL consultada, el código de respuesta, el tipo de contenido y el payload descargado, además de la fecha de ingesta. Esta tabla conserva la evidencia histórica de la descarga y permite reprocesar el pipeline sin depender nuevamente del origen.

**Tabla stg.statsbomb\_competitions:** Recibe la información de competiciones de StatsBomb en formato semiestructurado. Su función es servir como base para identificar las combinaciones válidas de competencia y temporada disponibles en la fuente.

**Tabla stg.statsbomb\_matches:** Almacena temporalmente los partidos descargados desde StatsBomb. En ella se conserva el JSON original antes de su normalización definitiva, lo que facilita el control de calidad y la depuración del proceso.

**Tabla stg.statsbomb\_events:** Concentra un registro por partido con el arreglo JSON completo de eventos. Esta tabla actúa como puente entre la persistencia cruda y la tabla final de eventos normalizados.

**Tabla stg.openfootball\_seasons:** Conserva los archivos descargados desde OpenFootball, preservando la información de temporada, liga y competición en un formato intermedio de staging.

**Tabla dwh.statsbomb\_matches\_norm:** Reúne la información normalizada de los partidos, incluyendo identificadores del encuentro, competencia, temporada, fecha, hora, equipos participantes, marcador, estadio y árbitro. Su función es aportar contexto competitivo y temporal al análisis.

**Tabla dwh.statsbomb\_events:** Representa la capa más granular del sistema. Cada fila corresponde a un evento individual del partido e incluye atributos como evento, minuto, segundo, periodo, tipo de acción, jugador, equipo, posición, coordenadas y desenlace. Constituye la base elemental a partir de la cual se derivan las métricas del fútbol.

**Tabla dwh.openfootball\_matches:** Integra calendarios y resultados de OpenFootball. Aunque esta fuente no presenta el mismo detalle que StatsBomb, aporta una estructura histórica complementaria útil para el contexto competitivo del sistema.

**Tabla dwh.fact\_player\_match\_stats:** Es la tabla de hechos principal del modelo analítico. Resume, para cada combinación jugador y/o partido, un conjunto de métricas derivadas de los eventos normalizados, tales como pases, pases completados, tiros, tiros a puerta, goles, asistencias, conducciones, amagues, amagues exitosos, entradas, intercepciones, despejes y faltas cometidas o recibidas. También conserva el nombre del jugador, el equipo y una estimación de minutos jugados basada en el minuto máximo observado en los eventos.

#### **4.2.1.5 Construcción de la tabla de hechos analítica**

La construcción de dwh.fact\_player\_match\_stats se implementó mediante el workflow W\_BUILD\_FACT\_PLAYER\_MATCH\_STATS, utilizando una estrategia de truncado y reconstrucción completa. Esta decisión se justificó en que la tabla es totalmente derivada de dwh.statsbomb\_events y, por tanto, puede regenerarse íntegramente a partir del estado más reciente de la capa de eventos. La agregación se realiza mediante consultas SQL agrupadas por match\_id y player\_id, lo que permite sintetizar el comportamiento observado en miles de eventos elementales en una unidad analítica estable, comparable y apta para cálculo posterior de métricas.

#### **4.2.1.6 Construcción de vistas analíticas**

Sobre la tabla de hechos se configuró una secuencia de vistas orientadas a transformar métricas básicas de partido en indicadores comparables y útiles para el proceso de scouting. Estas vistas cumplen una función metodológica central, ya que organizan la lógica del análisis dentro de la base de datos y preparan una capa semántica estable para la visualización.

##### **4.2.1.6.1 Vista de enriquecimiento jugador por partido**

La primera vista de esta secuencia consiste en enriquecer los registros de jugador por partido con el contexto competitivo del encuentro. Su función es integrar las métricas de dwh.fact\_player\_match\_stats con la información de dwh.statsbomb\_matches\_norm, de modo que cada observación quede asociada a variables como competencia, temporada y fecha. Esto permite que las métricas individuales se interpreten dentro de un marco competitivo concreto.

En esta vista se calculan las primeras métricas derivadas, especialmente aquellas normalizadas por tiempo de juego. La lógica empleada fue expresar los conteos absolutos en tasas por 90 minutos, utilizando la fórmula general:

$$M90 = (M \times 90) / Min$$

A partir de esta fórmula se construyen variables como `goals_per90`, `assists_per90`, `shots_per90`, `shots_on_target_per90`, `passes_per90`, `carries_per90`, `dribbles_per90`, `tackles_per90`, `interceptions_per90` y `clearances_per90`. Estas métricas corrigen el sesgo asociado a tiempos de juego desiguales y permiten comparaciones más justas entre jugadores con distinto nivel de participación. (Fernández-Cortés Tolosa, 2024)

#### 4.2.1.6.2 Métricas de eficiencia y calidad de ejecución

Además de las tasas por 90 minutos, se definieron métricas de eficiencia porcentual orientadas a capturar la calidad de la acción. Entre las más importantes se encuentran el porcentaje de pases completados, el porcentaje de tiros a puerta y el porcentaje de amagues exitosos.

$pass\_completion\_pct = passes\_completed / passes$

$shots\_on\_target\_pct = shots\_on\_target / shots$

$dribbles\_won\_pct = dribbles\_won / dribbles$

Estas métricas cumplen una función complementaria frente a las tasas por 90. Mientras las tasas muestran la frecuencia de la acción, los porcentajes muestran su calidad o efectividad. La combinación de ambas dimensiones permite un análisis más robusto del rendimiento. (Soria Polo, 2021)<sup>76</sup>

#### 4.2.1.6.3 Vista de resumen por jugador, competencia y temporada

La siguiente etapa analítica consiste en consolidar las métricas por jugador, competencia y temporada. Esta vista resume la información acumulada de múltiples partidos y construye un perfil estadístico más estable del jugador dentro de un contexto competitivo específico. Aquí se agregan minutos totales, partidos jugados y todas las métricas base derivadas de la tabla de hechos.

Sobre estos acumulados se recalculan nuevamente las tasas por 90 minutos y los porcentajes de eficiencia, pero ya desde el desempeño agregado del jugador. Esta decisión reduce la sensibilidad del análisis frente a la variabilidad de una sola

---

<sup>76</sup> Soria Polo, C. (2021). *Diseño y aplicación de técnicas de machine learning para optimizar el Scouting en clubes de fútbol*. Sevilla: Universidad de Sevilla.

observación y mejora la representatividad del perfil analizado. (Fernández-Cortés Tolosa, 2024)<sup>77</sup>

#### 4.2.1.6.4 Determinación de la posición principal

Una vez consolidado el desempeño por jugador, se construyó una vista orientada a determinar la posición principal de cada futbolista. La lógica utilizada fue contabilizar la frecuencia con la que un mismo jugador aparece asociado a distintas posiciones dentro de `dwh.statsbomb_events` y seleccionar aquella con mayor recurrencia. Posteriormente, esa posición se agrupó en categorías funcionales más amplias: CB, FB/WB, DM, CM, AM/W y ST.

Esta clasificación fue esencial porque permitió adaptar el análisis a la lógica táctica de cada rol. En vez de comparar a todos los jugadores bajo un mismo criterio, la base de datos pasó a organizar el rendimiento en grupos funcionales comparables. (Paulis, 2025)<sup>78</sup>

#### 4.2.1.6.5 Vista de consumo posicional

Con las métricas acumuladas y la posición principal ya definida, se configuró la vista `v_player_metrics_summary_pos`. Esta vista funcionó como una de las principales estructuras de consumo para visualización, porque integra en una sola consulta el perfil estadístico del jugador, su contexto competitivo y su clasificación posicional. En términos metodológicos, representa la transición entre la capa de cálculo analítico y la capa de lectura visual.

#### 4.2.1.6.6 Percentiles por competencia, temporada y posición

La siguiente etapa consistió en transformar las métricas directas en métricas comparativas. Para ello se calculó el percentil de cada jugador dentro de su competencia, temporada y grupo posicional. Esta operación permite ubicar su desempeño en relación con sus pares dentro de un mismo entorno competitivo.

El percentil cumple una función clave: llevar métricas de distinta naturaleza a una escala común de comparación. Así, variables como goles por 90, intercepciones por 90 o

---

<sup>77</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>78</sup> Paulis, J. C. (2025). Uso de los sistemas de juego en las ligas profesionales del fútbol español con relación a las variables situacionales. *Revista Iberoamericana de Ciencias de la Actividad Física y el Deporte*, 424-443.

precisión de pase pueden interpretarse conjuntamente cuando se expresan como posición relativa dentro del grupo. El análisis deja entonces de depender exclusivamente del valor absoluto y pasa a responder qué tan alto o qué tan bajo se sitúa un jugador frente a sus comparables. (Aditya, 2024)<sup>79</sup>

#### 4.2.1.6.7 Puntaje sintético de scouting

Sobre la base de los percentiles se construyó la vista `v_transfer_targets_all_positions`, orientada a consolidar la comparación general de jugadores y a calcular un puntaje sintético denominado `transfer_score`. Esta vista representa la etapa final de la lógica analítica en base de datos.

El `transfer_score` responde a una lógica multicriterio y posicional. En todos los casos, el puntaje se construyó a partir de percentiles  $P(x)$ , lo que permite expresar cada métrica en términos relativos frente a jugadores comparables de la misma posición, competencia y temporada. Aplica la ponderación según la función táctica del rol, por lo que no aplica los mismos indicadores ni pesos a todos los jugadores. La justificación de esta decisión radica en que los criterios de evaluación en scouting no son homogéneos: un delantero debe ser priorizado por variables de finalización, mientras que un central requiere mayor peso en acciones defensivas y salida de balón. (García González, 2025)<sup>80</sup>

CB, son los defensores centrales, para ellos se diseñó la métrica donde se prioriza acciones asociadas a la solidez defensiva y a la capacidad de sostener la primera fase de construcción. Por esta razón, se asignó el mayor peso a las intercepciones (0.30), ya que reflejan anticipación, lectura de juego y la capacidad para cortar progresiones rivales antes de que se conviertan en situación de riesgo. En segundo lugar, los despejes (0.25) reciben una ponderación alta porque representan la capacidad del central para resolver situaciones de presión en el área o en zonas cercanas a la portería. Las entradas (0.20) complementan esta dimensión defensiva, ya que muestran la intervención directa del jugador en duelos y acciones de recuperación. Finalmente, la precisión de pase (0.15) y el volumen de pases (0.10) se incorporan para representar el papel del defensor central en la salida del balón, aunque con un peso menor al componente defensivo, dado que su función principal sigue siendo la contención y protección del bloque. (Fernández-Cortés Tolosa, 2024)<sup>81</sup> (Martínez Sánchez, 2024)<sup>82</sup>

---

<sup>79</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>80</sup> García, M. S. (2024). Influencia de las variables contextuales en las acciones técnicas en fútbol.

<sup>81</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>82</sup> Martínez Sánchez, J. A. (2024). *Factores determinantes del rendimiento deportivo en fútbol*. Sevilla: Universidad Pablo de Olavide.

Andrea Ruiz Naranjo

---

$$TS\_CB = 0.20 \cdot P(tackles\_per90) + 0.30 \cdot P(interceptions\_per90) + 0.25 \cdot P(clearances\_per90) + 0.15 \cdot P(pass\_completion\_pct) + 0.10 \cdot P(passes\_per90)$$

FB/WB, se conocen como los laterales o carrileros, la formula combina diferentes variables defensivas, de circulación y de progresión ofensiva, debido a la naturaleza híbrida de esta posición. Estos jugadores participan tanto en tareas de recuperación como en generación de amplitud, en la progresión por banda y en el apoyo ofensivo. Las entradas (0.20) y las intercepciones (0.15) capturan su aporte defensivo en duelos y coberturas. Por su parte, los pases (0.15) y la precisión del pase (0.10) representa su rol en la circulación y en la continuidad del juego. A esto se suman las conducciones (0.15) y amagues (0.15), que modelan la capacidad del jugador para progresar con el balón y romper líneas por banda. Finalmente, las asistencias (0.10) reflejan su contribución directa a la generación de ocasiones de gol. En conjunto, la ponderación busca capturar el equilibrio entre funciones defensivas y ofensivas propio del lateral moderno (Fernández-Cortés Tolosa, 2024)<sup>83</sup> (Paulis, 2025)<sup>84</sup>.

$$TS\_FB/WB = 0.20 \cdot P(tackles\_per90) + 0.15 \cdot P(interceptions\_per90) + 0.15 \cdot P(passes\_per90) + 0.10 \cdot P(pass\_completion\_pct) + 0.15 \cdot P(carries\_per90) + 0.15 \cdot P(dribbles\_per90) + 0.10 \cdot P(assists\_per90)$$

Para el medio centro defensivos (DM), se construyó bajo una lógica de equilibrio entre recuperación y circulación segura del balón. Esta posición actúa como filtro por delante de la defensa y, al mismo tiempo, como punto de apoyo en la salida del equipo. Por ello, las interceptaciones (0.25) reciben el peso más alto, ya que representan la capacidad del jugador para leer trayectorias, recuperar posesión y cortar secuencias rivales. Las entradas (0.20) refuerzan esta dimensión de contención en acciones de disputa directa. Al mismo tiempo, los pases (0.20) y la precisión del pase (0.20) adquieren una relevancia equivalente, porque el mediocentro defensivo debe ofrecer seguridad en circulación y continuidad en la fase de inicio. Finalmente, las conducciones (0.15) permiten incorporar la capacidad de progresar con el balón en transiciones o fases de reorganización ofensiva. La fórmula, en consecuencia, busca representar un perfil de

---

<sup>83</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>84</sup> Paulis, J. C. (2025). Uso de los sistemas de juego en las ligas profesionales del fútbol español con relación a las variables situacionales. *Revista Iberoamericana de Ciencias de la Actividad Física y el Deporte*, 424-443.

equilibrio entre destrucción, cobertura y soporte en la construcción (Fernández-Cortés Tolosa, 2024)<sup>85</sup> (Aditya, 2024)<sup>86</sup>.

$$TS_{DM} = 0.25 \cdot P(interceptions\_per90) + 0.20 \cdot P(tackles\_per90) + 0.20 \cdot P(passes\_per90) + 0.20 \cdot P(pass\_completion\_pct) + 0.15 \cdot P(carries\_per90)$$

CM, los mediocampistas centrales, se construyó la fórmula que orienta principalmente a medir la capacidad de articulación del juego. Este rol exige distribuir, enlazar sectores del campo y facilitar la progresión ofensiva, por lo que los pases (0.25) reciben el mayor peso. La precisión del pase (0.20) complementa esta dimensión, ya que en esta importa la cantidad de intervenciones, la calidad y la seguridad. Las conducciones (0.15) reflejan la capacidad de progresar con balón, mientras que las asistencias (0.15) incorporan el aporte del mediocampista central en la generación ofensiva. A diferencia de perfiles más adelantados, en esta posición también se considera la contribución sin balón, por eso intercepciones (0.10) y entradas (0.10) forman parte de la formula, aunque de menor peso. Finalmente, los amagues (0.05) se incluye como una dimensión complementaria que puede aportar desequilibrio en contextos específicos, pero no constituye el rasgo central del rol. En suma, la formula privilegia construcción, enlace y equilibrio (Fernández-Cortés Tolosa, 2024)<sup>87</sup> (Martínez Sánchez, 2024)<sup>88</sup>.

$$TS_{CM} = 0.25 \cdot P(passes\_per90) + 0.20 \cdot P(pass\_completion\_pct) + 0.15 \cdot P(carries\_per90) + 0.15 \cdot P(assists\_per90) + 0.10 \cdot P(interceptions\_per90) + 0.10 \cdot P(tackles\_per90) + 0.05 \cdot P(dribbles\_per90)$$

En media puntas y extremos, la formula se diseñó para capturar creación, desequilibrio y producción ofensiva. Las asistencias (0.20) y las entradas (0.20) reciben el mayor peso, estos representan dos dimensiones fundamentales de este tipo de jugador: la capacidad de generar ocasiones para terceros y de superar rivales en situaciones individuales. La efectividad de los amagues (0.10) complementa el volumen de intentos, añadiendo una dimensión de calidad. Las conducciones (0.15) reflejan la capacidad de progresar con balón, especialmente en transición o en ataques por fuera. Los goles

---

<sup>85</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>86</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>87</sup> Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

<sup>88</sup> Martínez Sánchez, J. A. (2024). *Factores determinantes del rendimiento deportivo en fútbol*. Sevilla: Universidad Pablo de Olavide.

(0.10) y los remates (0.10) permiten capturar la contribución directa a la finalización, mientras que los pases (0.10) y la precisión del pase (0.05) representan el vínculo del jugador con la circulación ofensiva. En conjunto, la formula busca perfilar futbolistas capaces de generar ventajas, romper estructuras defensivas y participar activamente en la construcción y definición del ataque (Aditya, 2024)<sup>89</sup> (Vera Martínez, 2024)<sup>90</sup>.

$$TS\_AM/W = 0.20 \cdot P(\text{assists\_per90}) + 0.20 \cdot P(\text{dribbles\_per90}) + 0.10 \cdot P(\text{dribbles\_won\_pct}) + 0.15 \cdot P(\text{carries\_per90}) + 0.10 \cdot P(\text{goals\_per90}) + 0.10 \cdot P(\text{shots\_per90}) + 0.10 \cdot P(\text{passes\_per90}) + 0.05 \cdot P(\text{pass\_completion\_pct})$$

ST, delanteros centro, para ellos se desarrolló la formula donde se concentra en la finalización, ya que esta posición tiene como responsabilidad principal transformar posiciones y ocasiones en gol. Por ello, los goles (0.35) reciben el peso más alto de todas las variables, al construir la expresión más directa de la efectividad ofensiva. Los remates (0.20) capturan el volumen de amenaza generado por el jugador, mientras que la precisión de tiro (0.15) introduce una dimensión de calidad en la finalización. Las asistencias (0.10) se incluyen para considerar delanteros que también participan en asociaciones y generación de juego. A esto se suman los amagues (0.10) y las conducciones (0.10), que permiten reflejar la capacidad del atacante para crear ventajas individuales, progresar con balón y construir sus propias situaciones de remate. En conjunto, la ponderación prioriza la eficacia goleadora, pero sin excluir la contribución complementaria del delantero a la dinámica ofensiva general (Aditya, 2024)<sup>91</sup> (Salas Saiz, 2025)<sup>92</sup>.

$$TS\_ST = 0.35 \cdot P(\text{goals\_per90}) + 0.20 \cdot P(\text{shots\_per90}) + 0.15 \cdot P(\text{shots\_on\_target\_pct}) + 0.10 \cdot P(\text{assists\_per90}) + 0.10 \cdot P(\text{dribbles\_per90}) + 0.10 \cdot P(\text{carries\_per90})$$

Desde el punto de vista metodológico, con este puntaje se propone definir el perfil estadístico del jugador en una medida comparable que facilite la identificación, filtrado y priorización de los perfiles dentro del mercado de transferencias.

---

<sup>89</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>90</sup> Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

<sup>91</sup> Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

<sup>92</sup> Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

En la siguiente tabla (Tabla 3) se realiza un resumen por posición sobre las variables tenidas en cuenta para obtener el transfer score.

| Posición | Variables principales del transfer score                                       | Lógica de priorización                        |
|----------|--------------------------------------------------------------------------------|-----------------------------------------------|
| CB       | Entradas, intercepciones, despejes, precisión de pase, pases                   | Recuperación, despeje y salida de balón       |
| FB/WB    | Entradas, intercepciones, pases, precisión, conducciones, amagues, asistencias | Rol híbrido entre defensa y progresión        |
| DM       | Intercepciones, entradas, pases, precisión, conducciones                       | Contención y circulación segura               |
| CM       | Pases, precisión, conducciones, asistencias, intercepciones, entradas, amagues | Construcción, enlace y equilibrio             |
| AM/W     | Asistencias, amagues, efectividad de amague, conducciones, goles, tiros, pases | Creación, desequilibrio y producción ofensiva |
| ST       | Goles, tiros, precisión de tiro, asistencias, amagues, conducciones            | Finalización y amenaza ofensiva               |

Tabla 3: Síntesis métricas por posición transfer score

## 4.2.2 Diseño e implementación del proceso ETL

### 4.2.2.1 Diseño del driver de extracción

El proceso ETL se implementó en Apache Hop y se estructuró alrededor de la tabla gov.soccer\_driver, que actúa como mecanismo de control de la extracción. Para no codificar manualmente las URLs dentro de cada pipeline, se definió crear una tabla conductora en la que cada fila representa un recurso concreto a consultar. El ciclo de vida de cada recurso se controla mediante estados como pending, ingested, parsed o error. Este diseño facilita la escalabilidad del sistema, mejora la trazabilidad y permite

implementar estrategias de procesamiento por lotes, particularmente necesarias en el caso de los eventos de StatsBomb. (Enríquez Herrera, 2023)<sup>93</sup>

En StatsBomb, el driver se alimenta de manera encadenada. Primero se cargan las competiciones disponibles, después se generan las URLs de partidos desde la combinación `competition_id` y `season_id`, y finalmente se construyen las URLs de eventos utilizando los `match_id` obtenidos en la etapa anterior. En OpenFootball, en cambio, el driver contiene directamente las URL de resultados por liga y temporada.

#### 4.2.2.2 ETL aplicado a StatsBomb

La primera etapa consistió en el tratamiento del endpoint de competiciones. El pipeline `P_PARSE_STATS_BOMB` (ilustración 10) toma las respuestas ya descargadas desde `raw.http_responses` y las carga en `stg.statsbomb_competitions`, preservando el contenido original en formato JSONB. Esta tabla se emplea como base para identificar las ligas y temporadas disponibles dentro del repositorio.

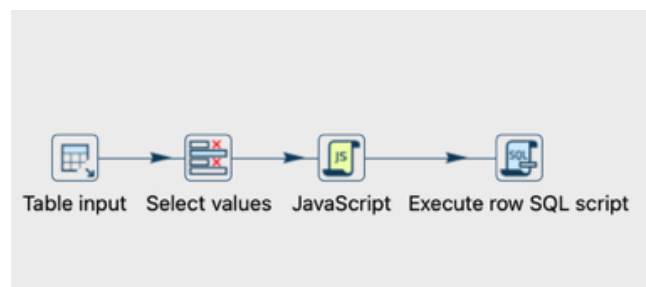


Ilustración 10: ETL `P_PARSE_STATS_BOMB`

La segunda etapa correspondió al tratamiento de partidos. El pipeline `P_PARSE_STATSBOMB_MATCHES` (ilustración 11) extrae la información descargada del endpoint de partidos, la almacena en `stg.statsbomb_matches` y posteriormente la normaliza en `dwh.statsbomb_matches_norm`. Durante esta etapa también se genera el driver de eventos, enlazando cada partido con su correspondiente recurso.



Ilustración 11: ETL `P_PARSE_STATSBOMB_MATCHES`

---

<sup>93</sup> Enríquez Herrera, J. V. (2023). Business Intelligence & Data Analytics aplicado al proceso de seguimiento curricular en la universidad UPEC. *Minerva*, 9-20.

Andrea Ruiz Naranjo

La tercera etapa fue la ingesta de eventos. Dado el tamaño de los archivos JSON y el elevado número de partidos, la descarga masiva produjo problemas de memoria. Para resolverlo se implementó un subflujo compuesto por W\_INGEST\_EVENTS\_AUTO (Ilustración 12), W\_EVENTS\_ONE\_BATCH (Ilustración 13), P\_INGEST\_RAW (Ilustración 14) y P\_CHECK\_PENDING\_EVENTS (Ilustración 15). La lógica consiste en descargar eventos en lotes, actualizar el estado del driver después de cada lote y repetir el proceso hasta completar todos los recursos pendientes.

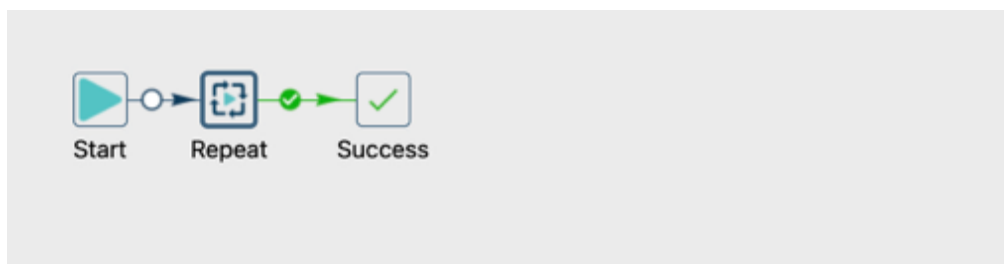


Ilustración 12: ETL W\_INGEST\_EVENTS\_AUTO

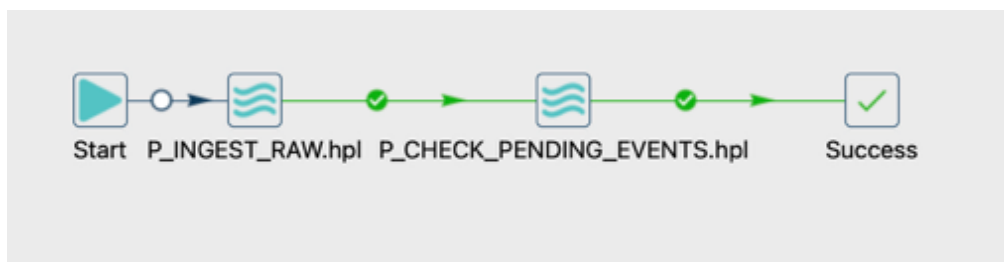


Ilustración 13: ETL W\_EVENTS\_ONE\_BATCH

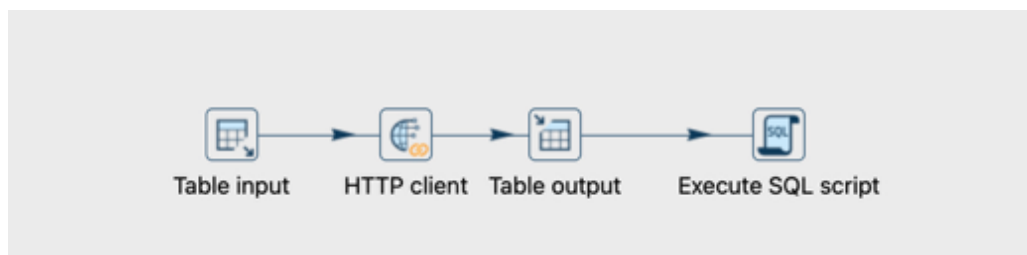
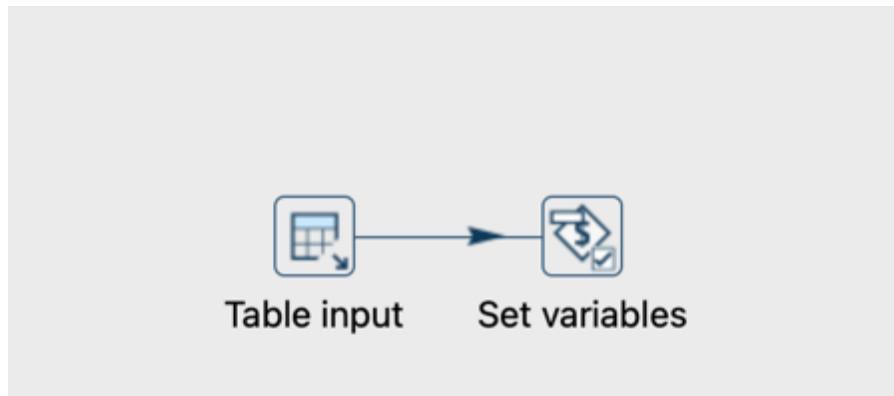
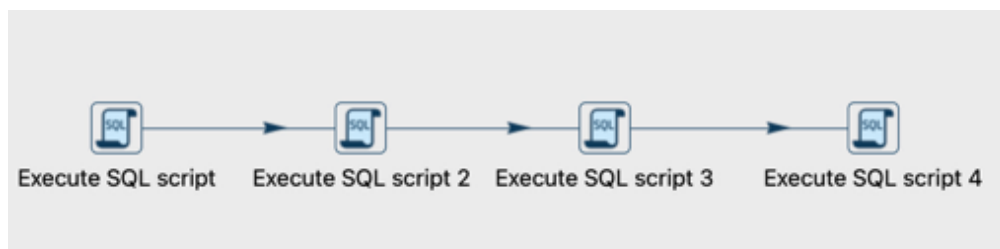


Ilustración 14: ETL P\_INGEST\_RAW



*Ilustración 15: ETL P\_CHECK\_PENDING\_EVENTS*

La cuarta etapa correspondió al parseo de eventos. El pipeline P\_PARSE\_STATSBOMB\_EVENTS (Ilustración 16) carga primero un registro por partido en stg.statsbomb\_events, donde cada fila conserva el arreglo JSON completo de eventos. Posteriormente, explota dichos arreglos mediante jsonb\_array\_elements y normaliza el detalle en dwh.statsbomb\_events. Este proceso incorpora limpieza previa por match\_id y controles frente a duplicados en la clave primaria compuesta por match\_id y event\_id, con el fin de asegurar procesos de ejecución reproducible.



*Ilustración 16: ETL P\_PARSE\_STATSBOMB\_EVENTS*

#### **4.2.2.3 ETL aplicado a OpenFootball**

OpenFootball se integró como fuente complementaria para disponer de calendarios y resultados por liga y temporada. La primera etapa se implementó mediante P\_INGEST\_OPENFOOTBALL\_RAW (Ilustración 17), que consulta el driver y almacena las respuestas en raw.http\_responses junto con el código HTTP y la meta data de contexto. La segunda etapa se desarrolló mediante P\_PARSE\_OPENFOOTBALL (Ilustración 18), que carga los archivos en stg.openfootball\_seasons y posteriormente reconstruye dwh.openfootball\_matches. Esta estrategia permite reprocesar la fuente manteniendo una salida consistente y libre de duplicados lógicos.

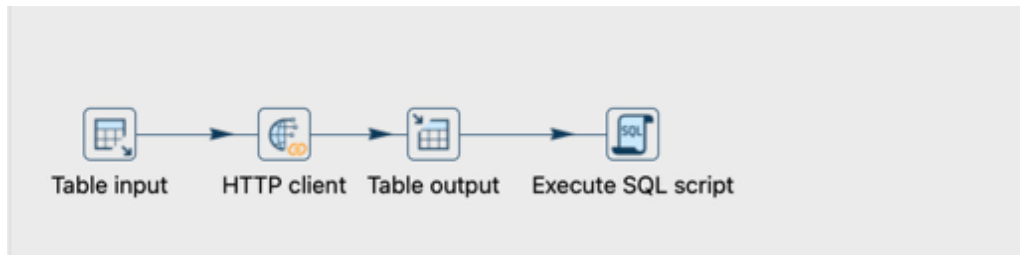


Ilustración 17: ETL P\_INGEST\_OPENFOOTBALL\_RAW



Ilustración 18: ETL P\_PARSE\_OPENFOOTBALL

#### 4.2.2.4 ETL carga estadísticas

Se ejecuta el workflow W\_BUILD\_FACT\_PLAYER\_MATCH\_STATS (Ilustración 19), en donde se hace un truncate a la tabla dwh.fact\_player\_match\_stats, después se inserta la información, en esta se encuentran las estadísticas de los jugadores y partidos.

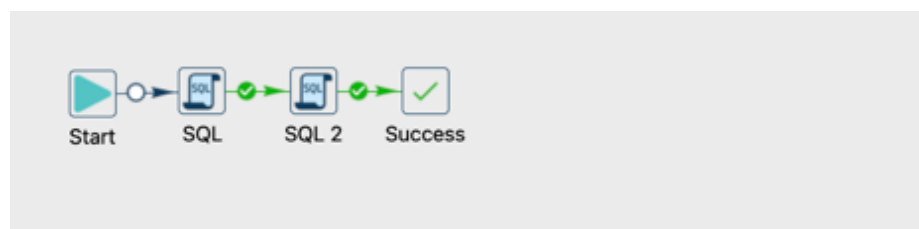


Ilustración 19: ETL W\_BUILD\_FACT\_PLAYER\_MATCH\_STATS

#### 4.2.2.5 Orquestación del flujo maestro

La ejecución integral del sistema se automatizó mediante el workflow W\_SOCCER\_MASTER (Ilustración 20), encargado de coordinar secuencialmente el parseo de competiciones, el tratamiento de partidos, la ingesta automática de eventos por lotes, el parseo de eventos, la transformación de OpenFootball y la construcción de dwh.fact\_player\_match\_stats. La orquestación distingue procesos principales y procesos auxiliares. Los principales generan tablas finales o disparan subflujos

completos; los auxiliares quedan encapsulados dentro de workflows específicos, reduciendo la complejidad visual del flujo global y mejorando su mantenibilidad.



Ilustración 20: ETL W\_SOCCER\_MASTER

#### 4.2.2.6 Validación, control de calidad e idempotencia

La validación del sistema se realizó en varios niveles. Primero se verificó la correcta descarga de recursos mediante el registro del status\_code y el seguimiento del estado de cada recurso en gov.soccer\_driver. Después se comprobó la estructura de las tablas de staging mediante conteos, revisión de tipos JSON y extracción de muestras representativas. Finalmente, se validó la consistencia de las tablas del datawarehouse contrastando conteos, claves primarias y la existencia efectiva de datos en las tablas analíticas resultantes.

Un aspecto central fue la ejecución controlada. Dado que el workflow maestro puede ejecutarse varias veces, fue necesario garantizar que una nueva ejecución no generara duplicados ni conflictos de integridad. En `dwh.statsbomb_events` se incorporó una lógica de limpieza previa de los partidos a reconstruir y controles ante inserciones duplicadas. En OpenFootball, la reconstrucción se controló por temporada y liga. En la fact table de jugador se optó por reconstrucción completa. Estas decisiones permitieron contar con un flujo robusto, repetible y coherente con los principios de una arquitectura ETL mantenible. (Pinargote Cárdenas & Alcívar Cedeño, 2025)<sup>94</sup>

### **4.2.3 Visualización de la información y construcción de dashboards**

#### **4.2.3.1 Construcción de la capa de visualización**

La visualización se desarrolló en Apache Superset como capa final de consumo analítico. Su propósito fue transformar la estructura consolidada en PostgreSQL en una herramienta útil para el análisis de jugadores. Para ello se diseñaron vistas analíticas en SQL que dejaran los datos listos para ser consultados en Superset. Las dos vistas principales utilizadas como fuente de los dashboards fueron `v_player_metrics_summary_pos` y `v_transfer_targets_all_positions`.

La primera se empleó para resumir métricas por posición y visualizar perfiles de rendimiento ya consolidados. La segunda se utilizó para integrar la comparación general de jugadores con el valor de transferencia y el puntaje sintético de evaluación. Esta decisión permitió que Superset operara sobre una capa semánticamente interpretada y no directamente sobre tablas operativas o cálculos dispersos dentro de cada gráfico. (Martínez Santana, 2025)<sup>95</sup>

#### **4.2.3.2 Lógica metodológica de los dashboards**

La lógica de los dashboards consistió en segmentar el análisis por posición. El supuesto metodológico detrás de esta decisión es que un delantero no debe evaluarse con los mismos indicadores que un defensor central, un mediocentro defensivo o un carrilero. A partir de esta premisa se definieron métricas específicas para cada grupo posicional y se construyó un dashboard para cada uno, además de uno general. (García, 2024)<sup>96</sup>

---

<sup>94</sup> Pinargote Cárdenas, R. E., & Alcívar Cedeño, A. (2025). *Analítica de Datos e Indicadores de Rendimiento para optimizar la gestión Universitaria*. Sinapsis.

<sup>95</sup> Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

<sup>96</sup> García, M. S. (2024). Influencia de las variables contextuales en las acciones técnicas en fútbol.

El proceso siguió una secuencia definida. En primer lugar, se consolidaron las vistas finales de consumo en PostgreSQL. En segundo lugar, se seleccionaron los indicadores más relevantes para cada rol. En tercer lugar, se diseñó la estructura visual de cada dashboard. En cuarto lugar, se construyó el dashboard general como visión panorámica del sistema. Finalmente, se desarrollaron los dashboards específicos por posición y se realizó una validación visual para garantizar consistencia entre ellos.

#### **4.2.3.3 Estructura visual**

Cada dashboard se estructuró con tres componentes principales: tarjetas KPI, utilizadas para resaltar indicadores clave y ofrecer una lectura rápida del comportamiento general de las métricas; diagramas de caja, útiles para observar la distribución de los indicadores y detectar valores atípicos dentro de cada grupo posicional; y una matriz comparativa de jugadores, con el fin de relacionar el desempeño deportivo. Esta estructura homogénea facilitó la comparación visual entre dashboards y mantuvo coherencia metodológica en toda la herramienta. (Contreras García & Cabrera Lozano, 2024)<sup>97</sup>

#### **4.2.3.4 Dashboards contruidos**

El primer dashboard desarrollado fue Positional Performance Analysis, concebido como visión global del sistema. En él se reunieron métricas generales y comparaciones entre posiciones, permitiendo observar cómo cambian los indicadores según el rol del jugador.

A partir de esta vista general se desarrollaron los dashboards específicos: ST, centrado en finalización y producción ofensiva, con énfasis en goles, remates, precisión de tiro y asistencias; AM/W, enfocado en creación y desequilibrio ofensivo mediante asistencias, amagues, efectividad de amagues, goles y volumen de pases; CM, orientado a perfiles de construcción y equilibrio, considerando pases, precisión, asistencias, conducciones e intercepciones; DM, centrado en recuperación, contención y seguridad en salida, a partir de intercepciones, entradas, pases y precisión de pase; CB, enfocado en acciones defensivas y salida de balón, como despejes, intercepciones, entradas y precisión de pase; y FB/WB, diseñado para evaluar amplitud, apoyo ofensivo y recuperación del balón mediante variables ofensivas y defensivas.

En el tablero conocido como Positional Performance Analysis (Ilustración 21), este ofrece una visión global del modelo analítico, permitiendo comparar las métricas más

---

<sup>97</sup> Contreras García, J. M., & Cabrera Lozano, R. (2024). Visualización de datos con Power BI. *Epsilon: Revista de la Sociedad Andaluza de Educación Matemática "Thales"*, 69-79.

Andrea Ruiz Naranjo

relevantes entre los distintos grupos posicionales. Su finalidad es mostrar cómo varían los indicadores de rendimiento según la función táctica del jugador y proporcionar una lectura general del comportamiento de la base de datos analizada, sirviendo como punto de partida para la exploración comparativa del sistema.

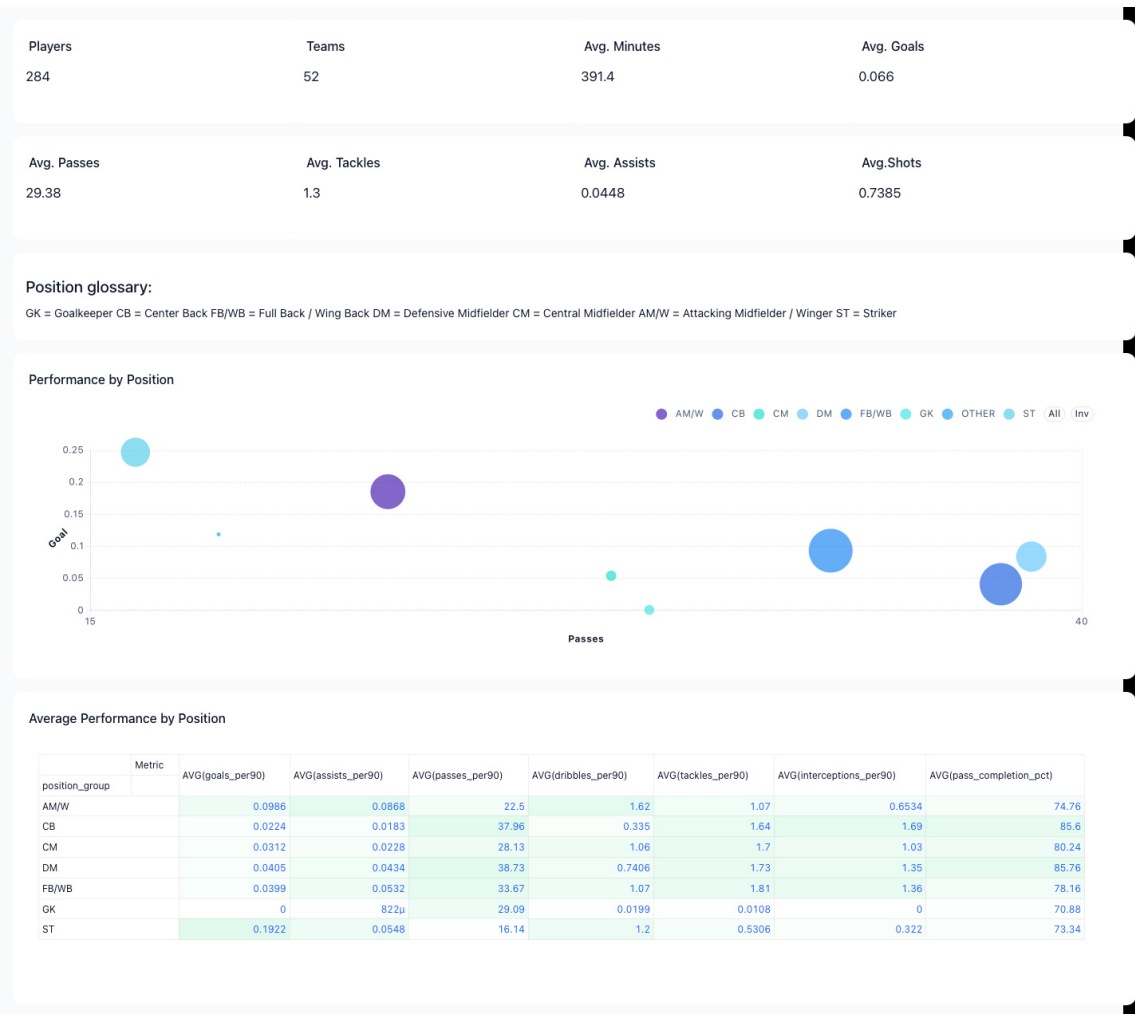


Ilustración 21: Visualización Positional Performance Analysis

El tablero llamado Striker Dashboard (Ilustración 22), se centra en el análisis de los delanteros centro, priorizando variables asociadas a la finalización y a la producción ofensiva. Para ello, incorpora métricas como goles, volumen de remates, precisión de tiro, asistencias, con el fin de evaluar la capacidad del jugador para generar peligro, concretar ocasiones y contribuir al rendimiento ofensivo del equipo.

Andrea Ruiz Naranjo

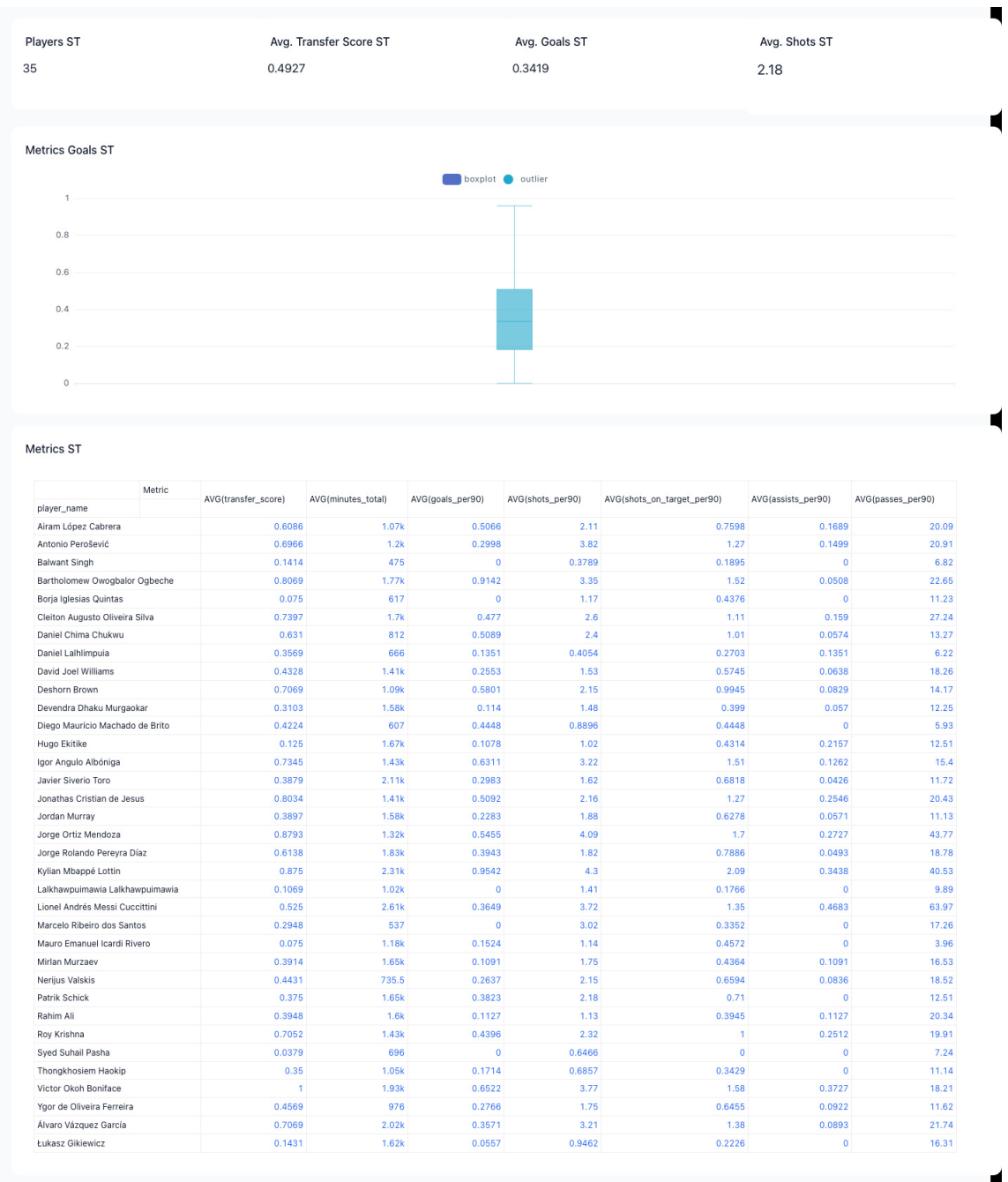


Ilustración 22: Visualización Striker Dashboard (ST)

En el dashboard de attacking Midfielder / Winger (Ilustración 23), se enfoca en jugadores con funciones de creación, desequilibrio y aporte ofensivo en zonas intermedias y exteriores del campo. El análisis se apoya en indicadores de asistencias, amagues, efectividad en los amagues, goles y volumen de pases, lo que permite identificar perfiles

Diseño e implementación de un sistema de apoyo a la toma de decisiones para el scouting de jugadores mediante herramientas Big Data.

Andrea Ruiz Naranjo

con capacidad para generar ventajas, romper líneas defensivas y participar activamente en la construcción de juego ofensivo.



Ilustración 23: Visualización Attacking Midfielder / Winger Dashboard (AM/W)

Andrea Ruiz Naranjo

Este tablero (Ilustración 24) está diseñado para evaluar perfiles de mediocampistas centrales cuya función principal combina construcción, enlace y equilibrio en el mediocampo. Para ello, considera métricas como volumen de pases, precisión de pase, asistencias, conducciones e interceptaciones, permitiendo analizar tanto la capacidad de distribución y progresión del balón como la contribución del jugador en tareas de recuperación.

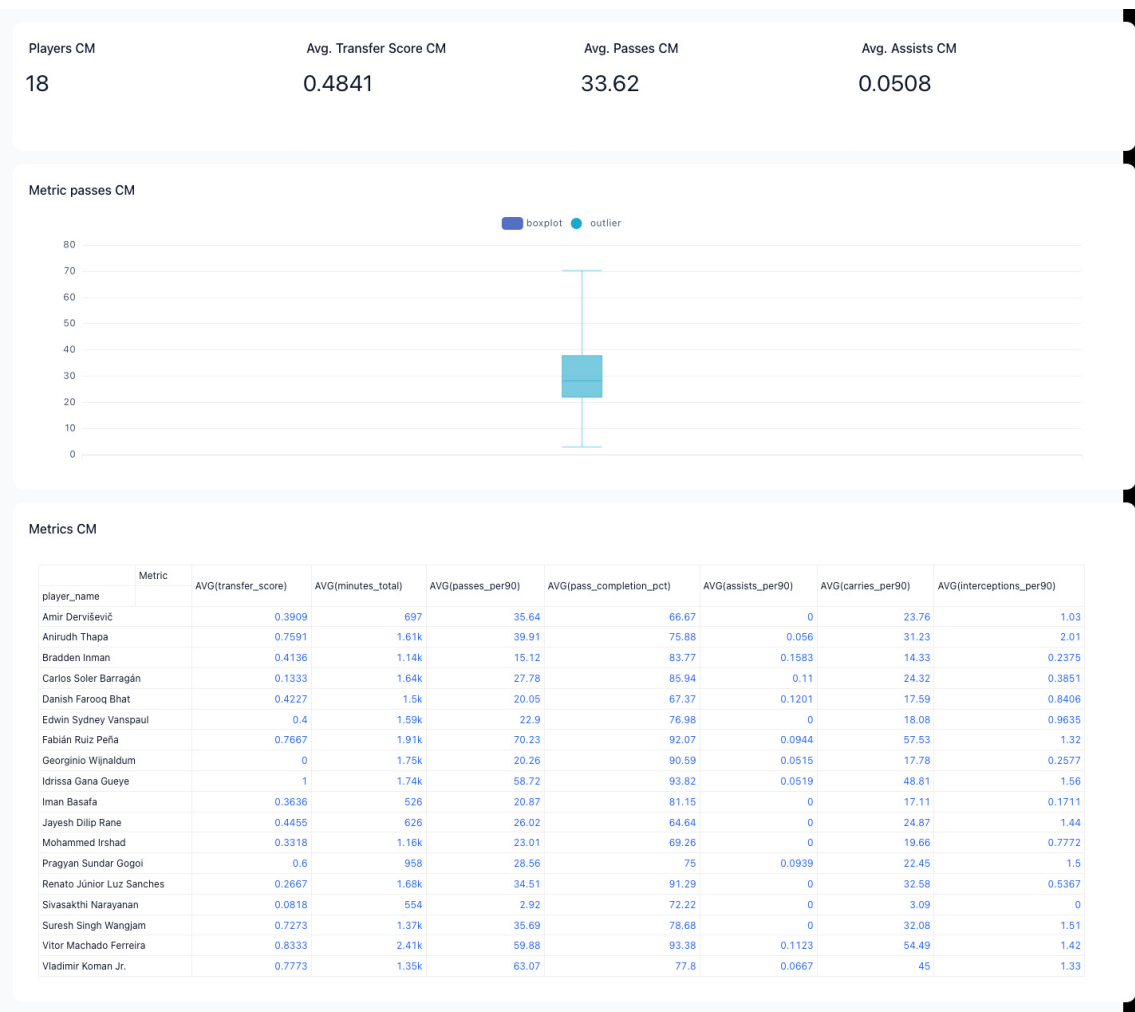


Ilustración 24: Visualización Central Midfielder Dashboard (CM)

En la siguiente imagen (Ilustración 25) se muestran los futbolistas cuya función principal está asociada a la recuperación, la contención y la seguridad en la salida del balón. El análisis incorpora métricas como interceptaciones, entradas, volumen de pases y precisión del pase, con el propósito de evaluar la capacidad del jugador para proteger la estructura defensiva del equipo y facilitar la circulación desde zonas retrasadas.

Andrea Ruiz Naranjo

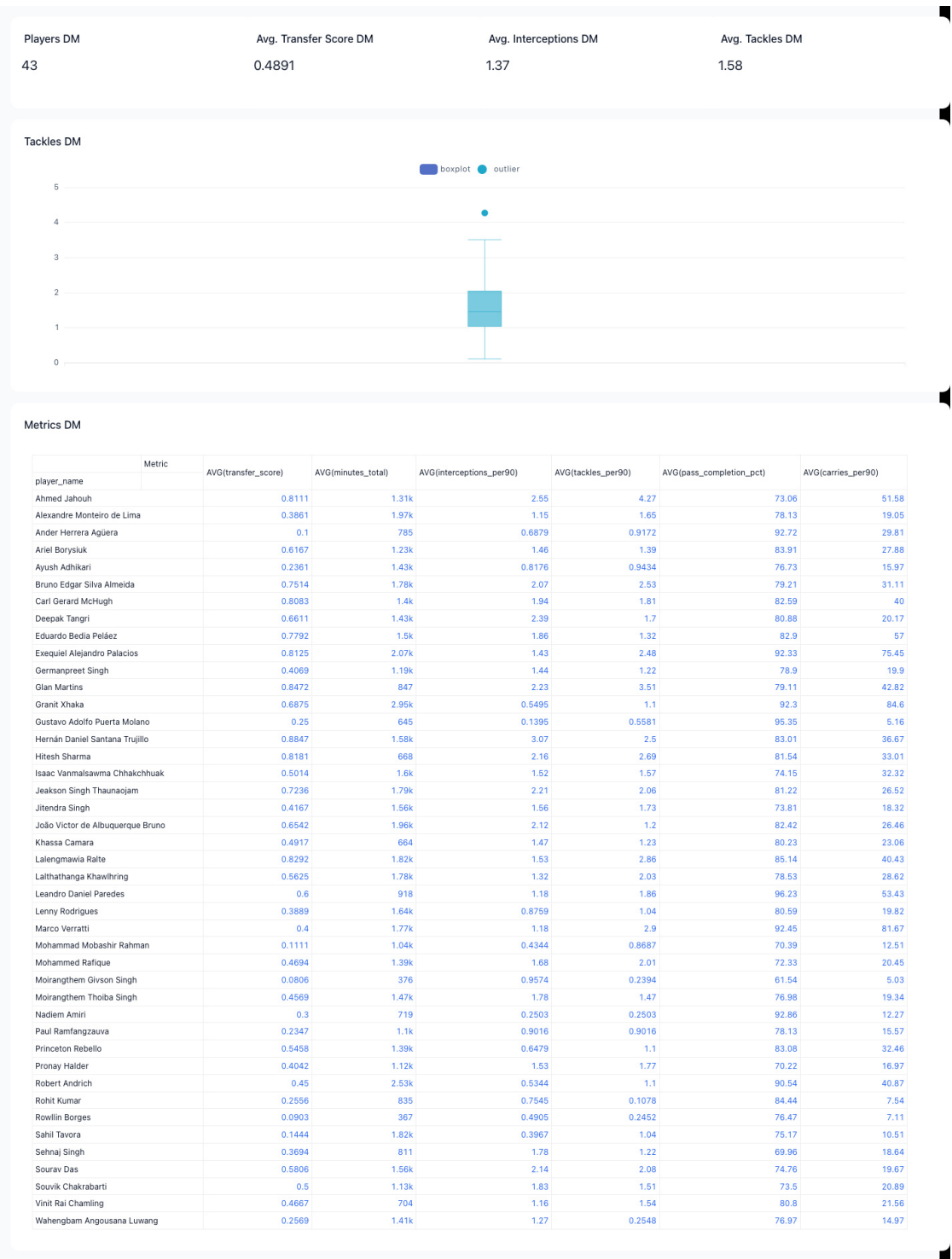


Ilustración 25: Visualización Defensive Midfielder Dashboard (DM)

Andrea Ruiz Naranjo

---

En la siguiente imagen (Ilustración 26) se analizan los defensores centrales a partir de acciones propias de su rol defensivo y de su participación en la salida del balón. Para ello, se emplean métricas como despejes, intercepciones, entradas y precisión de pase, que permiten valorar tanto la solidez defensiva del jugador como su capacidad para iniciar la construcción desde la línea de fondo.

Andrea Ruiz Naranjo

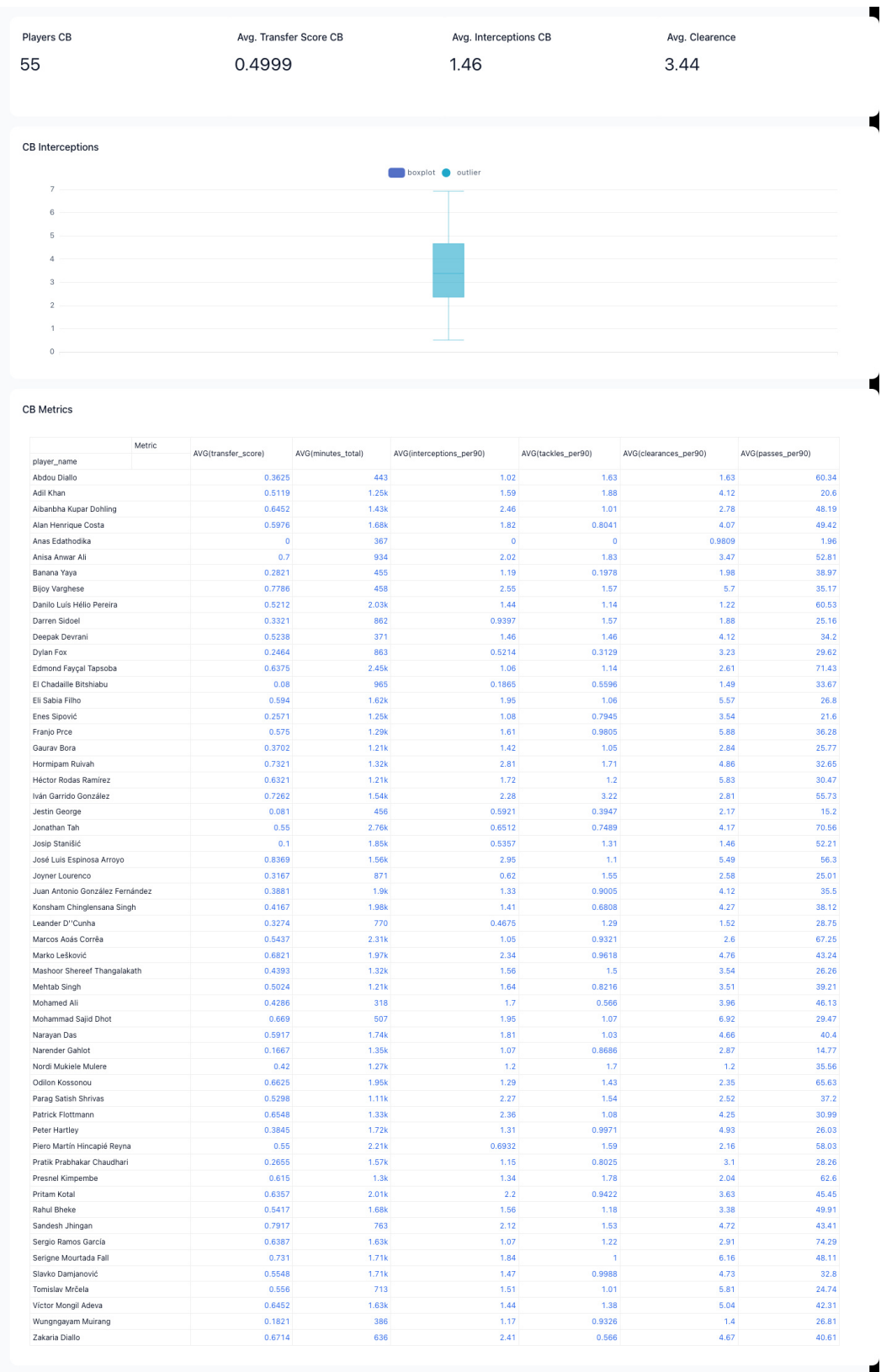


Ilustración 26: Visualización Center Back Dashboard (CB)

Para los jugadores con la posición de laterales y carrileros se combinan variables ofensivas y defensivas, ya que ellos cumplen funciones mixtas de amplitud, apoyo ofensivo y recuperación del balón por banda. A través de esta combinación de indicadores, se busca identificar perfiles capaces de contribuir tanto a la progresión del juego como a la cobertura y estabilidad defensiva del equipo (Ilustración 27).

Andrea Ruiz Naranjo

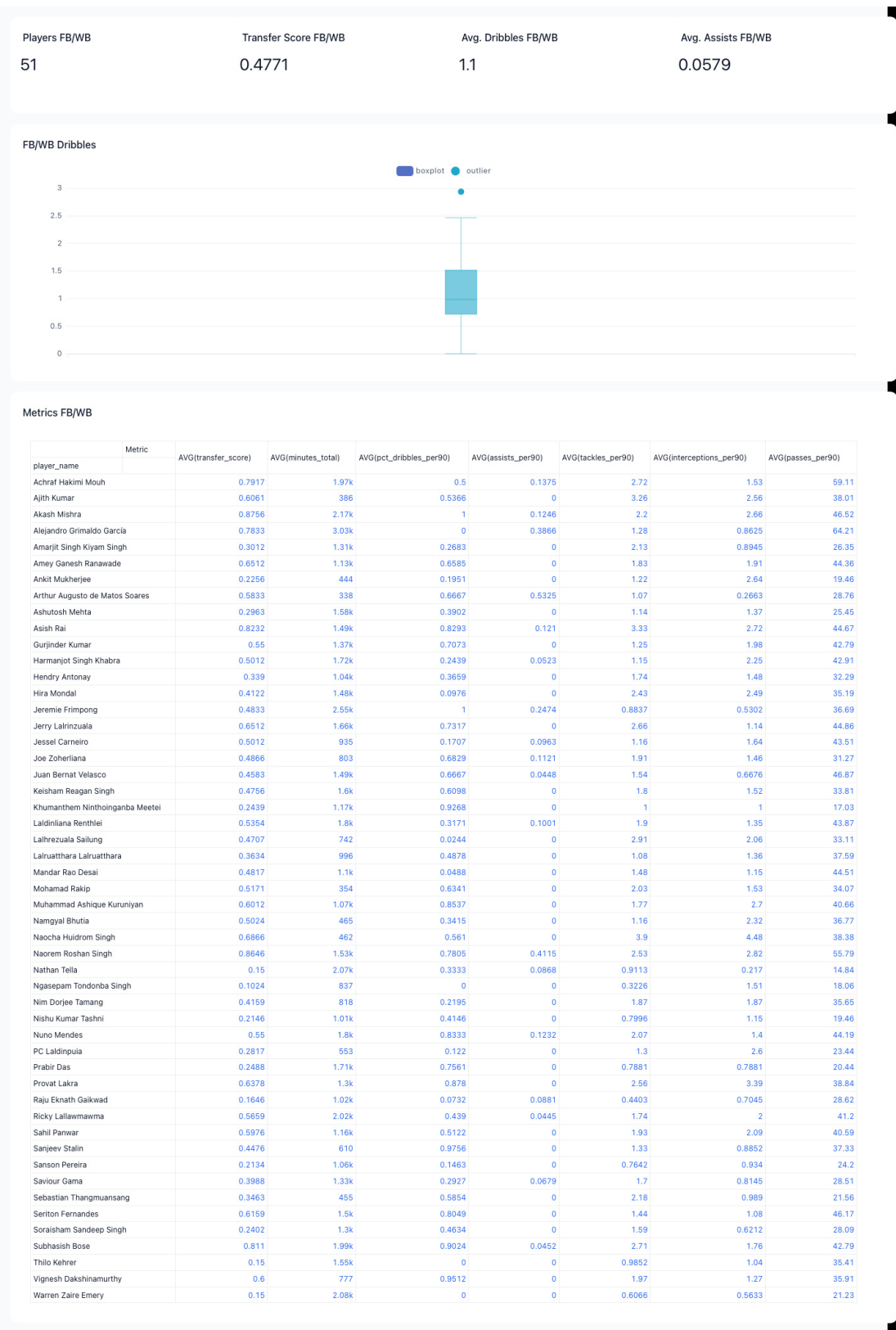


Ilustración 27: Visualización

Full Back / Wing Back Dashboard (FB/WB)

Los tableros desarrollados constituyen la capa final de explotación analítica del sistema y, en consecuencia, su utilidad va más allá de la simple presentación visual de estadísticas. Su propósito principal es convertir la información previamente integrada, transformada y estructurada en la base de datos es una herramienta de consulta, exploración y comparación orientada a la evaluación de jugadores. En términos prácticos, estos dashboards se diseñaron para ser un mecanismo de apoyo en distintas etapas del proceso de scouting: en una primera fase, permiten realizar una revisión general del comportamiento de los datos y de las diferencias de rendimiento entre posiciones; posteriormente, facilitan la preselección de futbolistas a partir de métricas clave según el rol táctico; y, en una etapa más avanzada, ofrecen una base objetiva para contrastar perfiles específicos y respaldar decisiones relacionadas con seguimiento, priorización o análisis de posibles incorporaciones. Desde esta perspectiva, su valor radica en hacer posible una lectura más estructurada y operativa de la información disponible.

El tablero general, positional performance analysis, cumple una función especialmente importante dentro del Sistema, ya que sirve como punto de partida para la interpretación del modelo analítico. Su utilidad consistente en proporcionar una visión panorámica del comportamiento de las principales métricas entre los distintos grupos posicionales, permitiendo identificar de manera rápida como cambian los indicadores de acuerdo con la función táctica del jugador. Este tablero puede emplearse para contextualizar el análisis antes de profundizar en posiciones específicas, para reconocer tendencias generales dentro de la base de datos y para comprender que dimensiones del rendimiento adquieren mayor relevancia según el rol en el campo. En otras palabras, su valor reside en ofrecer una lectura comparativa de alto nivel que ayuda al usuario a entender la lógica general del sistema y a orientar la exploración posterior hacia los dashboards personalizados.

Por su parte, los dashboards específicos por posición permiten un uso más focalizado y analíticamente más preciso. En ellos, la información deja de presentarse como una comparación amplia entre roles y pasa a organizarse según las exigencias concretas de cada perfil táctico. Esto significa que un analista puede utilizar el tablero de delanteros para examinar la capacidad de finalización y producción ofensiva, mientras que el tablero de defensores centrales podrá concentrarse en acciones de contención y salida de balón, y en el de mediocampistas centrales podrá evaluar construcción, enlace y equilibrio. Esta segmentación resulta variada porque evita comparaciones inadecuadas entre posiciones que cumplen funciones distintas y permite que la interpretación de métricas se realice dentro de un marco coherente. Así, los tableros específicos se convierten en una herramienta útil para comparar jugadores que compiten por un mismo rol, detectar perfiles con fortalezas particulares, identificar comportamientos

sobresalientes o atípicos y reconocer posibles compatibilidades entre el perfil estadístico del jugador y las necesidades tácticas del equipo.

Desde una perspectiva metodológica, otro aporte importante de estos dashboards es que fortalecen la trazabilidad del análisis. Dado que los indicadores visualizados provienen de una arquitectura previa de integración, transformación y consolidación de datos, cada resultado mostrado en el dashboard puede interpretarse como la salida de un proceso analítico estructurado. En consecuencia, el valor de los dashboards se expresa en tres niveles complementarios: el primer nivel, como mecanismo de visualización e interpretación; el Segundo, actúa como un instrumento de comparación entre perfiles de jugadores; y el tercer nivel, como soporte para decisiones más ordenadas, justificables y alineadas con la lógica posicional del scouting moderno. De esta manera, los dashboards desarrollados convierten la base analítica del Proyecto en una herramienta funcional para el análisis aplicado dentro del entorno del fútbol profesional.

### 4.3 Recursos requeridos

- Apache Hop, para el diseño y ejecución de ETLs.
- Apache Superset, para la construcción de los dashboards.
- PostgreSQL, como sistema de gestor de bases de datos.
- DBeaver, para la administración y consulta de la base de datos.
- Docker, para ejecución de Apache Superset.
- Microsoft Word, para la elaboración de la memoria.
- StatsBomb, como fuente de datos.
- OpenFootball, como fuente complementaria de datos.
- Java, Requisito para ejecutar Apache Hop
- GitHub, subir el proyecto
- Firefox, acceso a la interfaz de Apache Superset e investigación.

### 4.4 Presupuesto

| Tipo de coste                   | Valor        | Comentarios                                                                                                                                     |
|---------------------------------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| Horas de trabajo en el proyecto | 310<br>9300€ | Estimación de las 22 semanas de trabajo, con un promedio de 13 horas semanales. Adicionalmente 10 horas de personas externas en la revisión del |

Andrea Ruiz Naranjo

|                          |        |                                                                                 |
|--------------------------|--------|---------------------------------------------------------------------------------|
|                          |        | documento. El valor estimado por cada hora es de 30€.                           |
| Equipo técnico utilizado | 1200€  | Computador personal para el desarrollo del proyecto. MacBook Air 13" de Apple.  |
| Software utilizado       | 0€     | Las herramientas escogidas para el desarrollo de este proyecto son open source. |
| Estudios e informes      | 0€     | Las fuentes principales para el desarrollo de este proyecto fueron gratuitas.   |
| Materiales empleados     | 0€     | No se emplearon materiales físicos ni de laboratorios.                          |
| Total                    | 10500€ |                                                                                 |

## 4.5 Viabilidad

El proyecto se considera viable desde una perspectiva técnica, económica y operativa. A nivel técnico, su desarrollo fue posible gracias al uso de herramientas de código abierto y de amplia disponibilidad, como Apache Hop, PostgreSQL, Apache Superset y DBeaver, las cuales permitieron implementar los procesos de ETL, almacenamiento analítico y la visualización de resultados sin necesidad de infraestructura altamente especializada. Además, el acceso a fuentes de datos abiertas hizo posible la construcción del sistema sin depender de adquisiciones externas de información.

Centrándonos en el punto de vista económico, la propuesta presenta una viabilidad favorable, ya que la mayor parte de los recursos utilizados no implicó costos de licencia y el principal aporte correspondió al tiempo de trabajo invertido y al uso del equipo técnico disponible. En términos operativos, el proyecto también resulta viable porque la solución desarrollada integra de manera coherente la adquisición, transformación, almacenamiento y visualización de datos, demostrando que es posible construir una herramienta funcional de apoyo a la decisión con recursos accesibles y una arquitectura reproducible.

## **4.6 Resultados del proyecto**

### **4.6.1 Resultados de la integración y almacenamiento de datos**

Como resultado del proceso de desarrollo, se implementó una arquitectura de datos funcional para la integración de información del fútbol, procedente de las fuentes abiertas StatsBomb y OpenFootball. Esta arquitectura permitió estructurar el flujo completo del dato desde su extracción hasta su explotación analítica, utilizando Apache Hop para la orquestación ETL y PostgreSQL como sistema de almacenamiento y transformación.

La base de datos quedó organizada en cuatro esquemas (gov, raw, stg y dwh), lo que permitió separar el control operacional del proceso, el almacenamiento de respuestas originales, la preparación semiestructurada y la capa analítica final. Uno de los principales resultados de esta etapa fue la consolidación de una base de datos reproducible y trazable. La tabla gov.soccer\_driver permitió controlar de manera centralizada las fuentes, endpoints, ligas, temporadas, URLs y estados de ejecución, mientras que raw.http\_responses conservó la evidencia de las respuestas obtenidas. Esto hizo posible mantener un registro completo del proceso de ingesta y habilitó el reprocesamiento controlado de las transformaciones sin necesidad de depender nuevamente de las fuentes externas.

En términos de almacenamiento, el sistema logró consolidar la información en estructuras relacionales y semiestructuradas adecuadas para análisis posterior. En el esquema stg se prepararon competiciones, partidos, eventos y temporadas en un nivel intermedio, mientras que en el esquema dwh se normalizaron y organizaron las tablas analíticas principales del proyecto. Como resultado, se obtuvo una estructura de datos estable y escalable que soporta tanto el análisis estadístico como la visualización posterior.

El valor de esta etapa radica en que la arquitectura diseñada no depende de infraestructuras comerciales ni de acceso continuo a las fuentes originales. Cualquier equipo o analista que adopte este sistema puede reproducir, auditar y actualizar el proceso de integración de forma autónoma, lo que lo convierte en una solución viable para contextos con recursos tecnológicos limitados, como clubes de ligas menores o unidades de análisis de formación.

### **4.6.2 Resultados del proceso ETL**

El proceso ETL permitió automatizar la descarga, transformación y carga de información del fútbol desde ambas fuentes. En el caso de StatsBomb, el flujo incluyó la carga de competiciones, la generación y tratamiento de partidos, la construcción dinámica del driver de eventos y la descarga secuencial de eventos por lotes. Este último componente

fue importante, ya que resolvió las limitaciones de memoria detectadas durante las pruebas iniciales de descarga masiva. Como resultado, se logró una carga completa y controlada de los eventos asociados a los partidos disponibles en la fuente, sin comprometer la estabilidad del entorno de ejecución. La lógica por lotes implementada en `W_INGEST_EVENTS_AUTO` permitió procesar recursos de gran volumen manteniendo seguimiento del estado de cada elemento descargado.

En términos de la metodología, se demostró la viabilidad de trabajar con datos de alta granularidad mediante una arquitectura ETL incremental y robusta. En OpenFootball, el proceso ETL permitió integrar calendarios y resultados por liga y temporada en una estructura homogénea. La estrategia de reconstrucción de la tabla final `dwh.openfootball_matches` a partir de las últimas respuestas válidas por liga y temporada garantizó consistencia en la salida y evitó duplicidad. En conjunto, los procesos ETL desarrollados demostraron ser repetibles y mantenibles, al permitir múltiples ejecuciones sin comprometer la integridad del modelo final.

El aporte de este componente va más allá de la automatización: el diseño incremental del proceso ETL garantiza que el sistema puede ser ejecutado en múltiples ocasiones sin generar inconsistencias, lo que reduce significativamente el costo operativo de mantenimiento. Esta característica es especialmente valiosa en entornos donde los datos se actualizan periódicamente y donde la confiabilidad del proceso es tan importante como la calidad del dato resultante.

#### **4.6.3 Resultados de la construcción de la capa analítica**

El resultado central del proyecto se encuentra en la construcción de la capa analítica sobre PostgreSQL. A partir de la tabla `dwh.fact_player_match_stats` fue posible resumir el desempeño de cada jugador por partido mediante variables como pases, tiros, tiros a puerta, goles, asistencias, conducciones, enganches, entradas, intercepciones, despejes y faltas. Esta tabla se constituyó en la base cuantitativa del sistema y permitió pasar del detalle de eventos individuales a una estructura comparable de rendimiento a nivel jugador por partido.

Sobre esta base se construyó una secuencia de vistas analíticas que transformó los datos en información útil para scouting. En primer lugar, se calcularon métricas normalizadas por 90 minutos, con el fin de corregir el sesgo generado por diferencias en tiempo de participación entre jugadores. En segundo lugar, se incorporaron métricas de eficiencia porcentual, tales como porcentaje de pases completados, porcentaje de tiros a puerta y porcentaje de enganches exitosos. En tercer lugar, se consolidaron perfiles estadísticos por jugador, competencia y temporada, reduciendo la sensibilidad del análisis frente a un único partido.

Un resultado relevante fue la identificación de la posición principal de cada jugador a partir de la frecuencia de aparición en los eventos. Esta clasificación permitió agrupar a los futbolistas en categorías funcionales comparables, tales como defensores centrales, laterales carrileros, mediocentros defensivos, mediocampistas centrales, media puntas extremos y delanteros. A partir de ello, el análisis dejó de ser genérico y pasó a responder a la lógica táctica de cada rol. Posteriormente, se calcularon percentiles por competencia, temporada y grupo posicional, lo que permitió ubicar a cada jugador en relación con sus pares dentro de un mismo contexto competitivo. Finalmente, se definió un puntaje sintético de scouting (`transfer_score`) construido mediante combinaciones ponderadas de métricas específicas por posición.

El valor diferencial de esta etapa es que transforma datos crudos en inteligencia deportiva accionable. El `transfer_score` es una herramienta de decisión objetiva que sintetiza el rendimiento multidimensional de un jugador en un indicador comprensible para tomadores de decisión sin formación técnica de datos. Además, al calcular las métricas de los jugadores en una tasa de 90 minutos, el sistema evita comparaciones injustas entre jugadores de ligas o roles distintos, dotando al análisis de una validez contextual que los enfoques tradicionales de scouting raramente alcanzan.

#### **4.6.4 Resultados de la visualización y construcción de dashboards**

En la etapa final se desarrolló la capa de visualización en Apache Superset, utilizando como base principal las vistas `v_player_metrics_summary_pos` y `v_transfer_targets_all_positions`. Como resultado, se construyó un conjunto de dashboards orientados al análisis de jugadores por posición, además de un dashboard general para comparación global entre roles.

El dashboard general permitió observar cómo cambian las métricas de rendimiento según la posición del jugador, ofreciendo una visión panorámica del sistema. Por su parte, los dashboards específicos para delanteros, media puntas extremos, mediocampistas centrales, mediocentros defensivos, defensores centrales y laterales carrileros, permitieron profundizar en métricas particulares de cada rol. Esta segmentación constituyó un resultado importante, ya que alineó la visualización con la lógica del scouting y evitó el uso de criterios homogéneos para posiciones tácticamente distintas. Desde el punto de vista visual, cada dashboard integró tarjetas KPI, diagramas de caja y matrices comparativas de jugadores. Esta estructura permitió resumir indicadores clave, analizar la distribución de las métricas y contrastar el desempeño de los futbolistas dentro de cada grupo posicional.

En esta etapa tiene como valor adicional en que cierra la brecha entre el análisis técnico y la toma de decisiones operativa. Al presentar la información de forma segmentada por posición, con indicadores clave visibles de manera inmediata, el sistema permite que un director técnico o analista deportivo pueda consultar perfiles de jugadores, identificar

candidatos para transferencias y comparar rendimientos sin necesidad de manipular datos directamente. En consecuencia, la visualización funciona como la interfaz de uso final de todo el sistema analítico construido, materializando su propósito práctico.

#### **4.6.5 Síntesis de los resultados obtenidos**

En síntesis, los resultados obtenidos muestran que fue posible construir una solución tecnológica completa para la integración, estructuración, análisis y visualización de datos del fútbol orientados al scouting. La arquitectura implementada permitió pasar de fuentes abiertas heterogéneas a una capa analítica estructurada, reproducible y visualmente interpretable.

El valor global del sistema desarrollado se puede entender en tres dimensiones. Desde la dimensión técnica, se demostró que es posible construir una solución de inteligencia deportiva completa utilizando exclusivamente herramientas de código abierto, lo que elimina barreras de costo para su adopción. Desde la dimensión analítica, el sistema supera la acumulación de estadísticas al producir métricas contextualizadas, perfiles posicionales y un puntaje sintético que refleja el rendimiento real de cada jugador dentro de su propio entorno competitivo. Desde la dimensión práctica, la capa de visualización traduce esa inteligencia en una interfaz accionable para perfiles no técnicos, acercando el sistema a un uso real en procesos de scouting y toma de decisiones deportivas.

De esta manera, este proyecto produjo un sistema de apoyo a la decisión sustentado en métricas comparables y dashboards especializados por posición, con potencial de ser adoptado, extendido y mantenido en contextos reales del fútbol profesional y semiprofesional.

## Capítulo 5. DISCUSIÓN

El desarrollo del presente proyecto permitió verificar que la construcción de una herramienta de apoyo a la toma de decisiones orientada al scouting y al mercado de transferencias es técnicamente viable a partir de fuentes de datos abiertas, procesos de integración ETL, almacenamiento relacional y visualización interactiva. Sin embargo, el proceso evidenció limitaciones concretas y decisiones de ajuste que resultaron determinantes para la configuración final de la propuesta. En consecuencia, la discusión se enfoca en cuatro ejes: la disponibilidad de datos, la pertinencia de la metodología adoptada, la evolución del enfoque de trabajo y las restricciones tecnológicas identificadas, además se muestra una guía de cómo se usa el sistema.

### 5.1 Disponibilidad y acceso a fuentes de datos

Una de las limitaciones más relevantes identificadas durante el proyecto fue la disponibilidad restringida de datos sobre el fútbol de calidad. Si bien existen plataformas y proveedores con información valiosa para procesos de scouting, una parte significativa de estos recursos opera bajo modelos de pago o presenta restricciones de acceso, descarga y uso que resultan incompatibles con el contexto de un proyecto académico desarrollado con recursos limitados. Esta situación condiciona la posibilidad de trabajar con conjuntos de datos más completos, variables más avanzadas o históricos de mayor profundidad temporal.

En respuesta a esta restricción, el proyecto optó por apoyarse en fuentes abiertas, principalmente StatsBomb y OpenFootball, lo cual permitió avanzar en el diseño e implementación del sistema. Esta decisión fue metodológicamente coherente con el objetivo de construir una arquitectura reproducible y accesible; sin embargo, implica reconocer que el alcance de la solución está condicionado por la cobertura y granularidad de dichas fuentes, las cuales, aunque ricas en detalle de eventos, no alcanzan la amplitud de los datos disponibles en entornos profesionales de scouting.

#### 5.1.1 Contexto del acceso a datos en el fútbol profesional

Una de las limitaciones más relevantes identificadas en el desarrollo del proyecto fue la disponibilidad restringida de datos sobre el fútbol de calidad. Si bien existen plataformas y proveedores con información valiosa para procesos de scouting, una parte significativa de estos recursos opera bajo modelo de pago o presenta restricciones de acceso, descarga y uso que resultan incompatibles con el contexto de un proyecto académico desarrollado con recursos limitados. Plataformas como Opta, Wyscout o InStat ofrecen datos detallados de rendimiento a nivel profesional, pero su acceso requiere contratos comerciales que exceden el alcance de este tipo de trabajo. Esta situación condiciona directamente el diseño del sistema, limitando la posibilidad de trabajar con conjuntos de

datos más amplios, variables más avanzadas o históricos de mayor profundidad temporal.

En respuesta a esta restricción, el proyecto optó por apoyarse en fuentes abiertas como alternativa viable y metodológicamente coherente con el objetivo de construir una arquitectura reproducible y accesible. Esta decisión implicó reconocer que el alcance de la solución estaría condicionado por la cobertura y granularidad de dichas fuentes, aunque estas fuentes tienen detalles de eventos y estadísticas, no alcanzan la amplitud de los datos disponibles en entornos profesionales de scouting.

### **5.1.2 Fuentes seleccionadas y criterios de elección**

La identificación de las fuentes utilizadas en el proyecto se realizó mediante una búsqueda activa en motores de búsqueda generales y foros especializados en analítica deportiva y ciencia de datos aplicada al fútbol. Durante este proceso se evaluaron diversas plataformas y repositorios, priorizando aquellas que cumplieron con dos criterios fundamentales: acceso completamente gratuito y ausencia de restricciones de uso que pudieran limitar su integración o reproducción por parte de terceros.

Bajo estos criterios se seleccionaron dos fuentes principales: StatsBomb y OpenFootball, ofrece datos de alta granularidad a nivel de evento por partido, incluyendo pases, tiros, conducciones, presiones, entradas y otras técnicas especiales y atributos detallados. StatsBomb es la fuente principal del trabajo con el análisis de rendimiento individual. OpenFootball, por su parte, proporciona calendarios de partidos, resultados y estructuras de competiciones, completando la información de StatsBomb con contexto competitivo e histórico. Ambas fuentes cuentan con documentación accesible, estructuras de datos consistentes y licencias abiertas que permiten su uso sin restricciones en proyectos de investigación y desarrollo.

### **5.1.3 Cómo encontrar y evaluar nuevas fuentes**

Para aquellos usuarios que deseen extender el sistema incorporando nuevas fuentes de datos, se recomienda seguir el siguiente proceso.

El primer paso es identificar fuentes mediante una exploración sistemática de recursos especializados. Para ello, se recomienda revisar repositorios públicos en plataformas como GitHub, donde la comunidad de analítica deportiva publica conjuntos de datos bajo etiquetas como football-data o soccer-analytics; consultar foros y comunidades especializadas en ciencia de datos aplicada al deporte, que suelen compartir referencias

a fuentes actualizadas y evaluar la calidad de forma colectiva; explorar plataformas como Kaggle, que albergan conjuntos de datos deportivos listos para su uso; y revisar las referencias bibliográficas de publicaciones académicas recientes sobre análisis del fútbol, ya que estas suelen citar explícitamente las fuentes utilizadas. Adicionalmente, algunos organismos oficiales como federaciones nacionales o ligas profesionales publican estadísticas de acceso libre, por lo que sus repositorios y sitios web constituyen también una fuente de consulta pertinente.

El segundo paso es evaluar la viabilidad de cada fuente considerando los siguientes criterios: tipo de licencia y condiciones de uso, formato de los datos y compatibilidad con la arquitectura existente, frecuencia de actualización y estabilidad del repositorio, nivel de granularidad disponible, y cobertura de ligas, temporadas y torneos.

El tercer paso es registrar la nueva fuente en la tabla de gobierno del sistema. Una vez seleccionada la fuente, el usuario debe añadir un nuevo registro en la tabla `gov.soccer_driver` especificando el nombre de la fuente, la URL o endpoint de acceso, la liga y temporada correspondiente, y el estado inicial `pending`. Esta tabla centraliza el control de todas las fuentes integradas y permite que la ETL las procese automáticamente sin modificar la lógica de los flujos existentes.

Como siguiente paso es adaptar el flujo ETL correspondiente en Apache Hop para gestionar la extracción, transformación y carga de los nuevos datos, siguiendo los patrones ya implementados para Statsbomb y OpenFootball, que pueden utilizarse como plantillas de referencia. En particular, los pipelines `P_INGEST_RAW` y `P_PARSE_STATSBOMB_EVENTS` ilustran la lógica de descarga y transformación que puede adaptarse a nuevas fuentes con estructura similar.

Seguir este proceso garantiza que cualquier extensión del sistema coherente con la arquitectura existente, mantenga la trazabilidad del dato desde su origen y preserve la reproducibilidad del conjunto de la solución.

## **5.2 Pertinencia de la metodología adoptada**

El enfoque estructurado por capas resultó especialmente adecuado para los objetivos del proyecto. La organización del sistema en etapas claramente diferenciadas, como la adquisición de datos, transformación, almacenamiento analítico y visualización, facilitó tanto el desarrollo técnico como la comprensión del problema como un proceso integral. Esta arquitectura respondió de forma satisfactoria a la necesidad de integrar las distintas fases del ciclo de vida del dato de una misma solución evitando la fragmentación del

análisis y garantizando la trazabilidad del proceso desde la fuente hasta la capa de decisión.

La metodología fue especialmente pertinente dado que el objetivo del proyecto fue la construcción de una herramienta funcional capaz de transformar datos de apoyo a la decisión. En este sentido, la arquitectura implementada demostró ser coherente con ese propósito, al integrar de manera estructurada los componentes técnicos necesarios para sostener un flujo de datos completo y reproducible.

### **5.3 Evolución del enfoque: de la predicción al apoyo integral a la decisión**

Durante las fases iniciales del proyecto, la propuesta contemplaba un enfoque más orientado hacia la predicción, con el objetivo de generar modelos capaces de anticipar comportamientos o resultados asociados al rendimiento de los jugadores. Sin embargo, a medida que avanzó la revisión del problema, el análisis de las fuentes de datos disponibles y la evaluación de la viabilidad técnica evidenciaron que centrar el trabajo de grado en una dimensión puramente predictiva podría reducir la aplicabilidad práctica de la propuesta y aumentar de forma significativa la complejidad metodológica, sin que ello se tradujera necesariamente en una mayor utilidad para el usuario final.

Por esta razón, el proyecto evolucionó hacia un modelo integral de apoyo a la decisión, orientado a unir diferentes fuentes de datos abiertas, procesos ETL, almacenamiento analítico y visualizaciones especializadas en una herramienta coherente y funcional. Este cambio de enfoque es una decisión metodológicamente pertinente que permitió alinear la propuesta con las condiciones reales del desarrollo y con una necesidad más concreta del problema identificado. La construcción de un sistema configurable, transparente e interpretable responde de manera más directa a las necesidades operativas del scouting que un modelo predictivo de caja negra con menor capacidad de explicación.

### **5.4 Limitaciones tecnológicas: el caso Apache Superset**

La elección de Apache Superset como herramienta de visualización fue coherente con el propósito de trabajar con soluciones de código abierto y de evitar dependencias de software propietario. No obstante, durante el desarrollo se identificó que, si bien Superset ofrece una base sólida para la construcción de dashboards interactivos, presenta limitaciones en la configuración y personalización de ciertas visualizaciones. Algunas opciones de diseño, el control detallado sobre elementos gráficos y la

flexibilidad visual resultan más restringidas en comparación con herramientas comerciales especializadas, como es el caso de Tableau, Power BI o Looker.

Esta limitación implica que, aunque Superset es adecuado para construir un prototipo funcional y reproducible, podría no ofrecer el mismo nivel de refinamiento visual o configuración avanzada requerido en escenarios más exigentes o con necesidades muy específicas de presentación. Cabe señalar, que esta restricción no compromete la validez funcional del sistema desarrollado, ya que las visualizaciones implementadas cumplen con los requisitos analíticos establecidos y permiten extraer conclusiones relevantes para el proceso de scouting.

## **5.5 Valor e impacto desde una perspectiva aplicada**

A pesar de las limitaciones señaladas, el resultado final del proyecto posee un valor relevante desde una perspectiva aplicada. La solución desarrollada demuestra que es posible construir una herramienta estructurada para apoyar el análisis multicriterio de jugadores utilizando recursos accesibles, fuentes de datos abiertas y una arquitectura tecnológica reproducible. Este aspecto es particularmente significativo porque traslada el valor del proyecto más allá del ejercicio académico: se trata de una propuesta que organiza el flujo de datos desde su captura hasta su interpretación final, contribuyendo a demostrar cómo la ingeniería de datos y la analítica avanzada pueden integrarse para generar herramientas útiles en procesos reales de scouting.

En síntesis, la discusión permite concluir que la metodología adoptada fue adecuada para el problema abordado, aunque su desarrollo estuvo condicionado por las limitaciones propias del acceso a datos y de las herramientas tecnológicas disponibles. El cambio de enfoque desde un modelo predictivo hacia un sistema integral de apoyo a la decisión constituyó una adaptación metodológicamente pertinente que fortaleció la coherencia y viabilidad de la propuesta. Finalmente, si bien la solución presenta restricciones frente a sistemas profesionales más robustos, su principal valor reside en haber demostrado que es posible diseñar una arquitectura funcional, accesible y reproducible capaz de transformar datos abiertos de fútbol en una herramienta útil para la evaluación comparativa de jugadores.



*Ilustración 28: Limitaciones del proyecto*

## 5.6 Guía de uso del sistema: Scouting tool v2.0

El sistema desarrollado en este proyecto, denominado Scouting tool v2.0, está diseñado para ser utilizado tanto por perfiles técnicos con conocimientos en ingeniería de datos como por analistas deportivos y scouts que necesiten comparar perfiles de jugadores de forma objetiva y estructurada. A continuación, se describe el flujo completo de uso de sistema, desde la configuración inicial de las fuentes hasta la interpretación de los resultados en los dashboards, con el fin de que cualquier usuario pueda replicar, extender o consultar la herramienta de forma autónoma.

### 5.6.1 Paso 1: Configuración de fuentes en la tabla del gobierno

El punto de entrada del sistema es la tabla `gov.soccer_driver`, que actúa como panel de control de toda la arquitectura. En ella, el usuario técnico debe registrar cada fuente de datos que desea incorporar el análisis, especificando la fuente, la liga, temporada, endpoint y URL correspondiente. El estado inicial de cada registro debe establecerse como pending, lo que indica al sistema que ese recurso está pendiente de procesamiento. Sin registros activos en esta tabla, la ETL no tiene instrucciones que ejecutar. Esta decisión de diseño permite añadir nuevas ligas, temporadas o fuentes sin modificar ningún pipeline, simplemente insertando nuevas filas en el driver, en este caso para las fuentes de StatsBomb o OpenFootball.

### 5.6.2 Paso 2: Ejecución del proceso ETL en Apache Hop

Una vez configurado el driver, el usuario técnico ejecuta el workflow maestro conocido como `W_SOCCER_MASTER`, que orquesta de forma automática y secuencial todos los procesos de extracción, transformación y carga. Para StatsBomb, el flujo descarga primero las competiciones disponibles mediante `P_PARSE_STATS_BOMB`, genera

después los partidos correspondientes con `P_PARSE_STATSBOMB_MATCHES` y finalmente descarga y normaliza los eventos asociados mediante `W_INGEST_EVENTS_AUTO` y `P_PARSE_STATSBOMB_EVENTS`. Esta última etapa opera por lotes para gestionar el alto volumen de eventos sin comprometer la estabilidad del entorno. Para OpenFootball, el proceso ejecuta mediante `P_INGEST_OPENFOOTBALL_RAW` y `P_PARSE_OPENFOOTBALL`, integrando calendarios y resultados por liga y temporada. Al finalizar la ejecución, todos los datos quedan disponibles en la capa analítica del esquema dwh, listos para su consulta en Superset.

### **5.6.3 Paso 3: Exploración de perfiles en Apache Superset**

Con los datos procesados, el usuario accede a Apache Superset, donde encontrará un conjunto de dashboards organizados por posición táctica: defensor central (CB), lateral o carrilero (FB/WB), mediocampista defensivo (DM), mediocampista central (CM), media punta o extrema (AM/W) y delantero (ST). Adicionalmente, existe un dashboard general de comparación entre todos los roles. El usuario selecciona la posición de interés y el sistema presenta automáticamente los jugadores de esa categoría con sus métricas normalizadas por 90 minutos, sus indicadores de eficiencia porcentual y su posición relativa en percentiles dentro de la misma competencia, temporada y grupo posicional.

Cada dashboard integra tres elementos visuales complementarios. Las tarjetas KPI muestran los indicadores clave del grupo de forma inmediata. Los diagramas de caja permiten analizar la distribución de cada métrica dentro del grupo posicional e identificar jugadores atípicos por encima o por debajo del rango habitual. Las matrices comparativas de jugadores permiten contrastar directamente el desempeño de varios futbolistas dentro de la misma categoría.

### **5.6.4 Paso 4: Interpretación del transfer score**

El transfer score es el indicador sintético central del sistema y constituye el principal mecanismo de priorización de candidatos. Se calcula a partir de combinaciones ponderadas de percentiles específicos para cada posición, lo que garantiza que la comparación sea siempre contextualizada y tácticamente coherente. Un delantero con transfer score alto destacará en variables de finalización como goles, tiros y precisión del tiro, mientras que un defensor central con puntuación elevada sobresaldrá en intercepciones, despejes y entradas. El usuario puede utilizar este indicador como primer filtro de preselección, identificando a los jugadores mejor posicionados dentro de su grupo comparativo antes de profundizar en sus métricas individuales.

Es importante tener en cuenta que el transfer score es una herramienta de apoyo a la decisión y no un sustituto del juicio experto. Su valor reside en estructurar una primera capa de análisis, reduciendo el tiempo invertido en la revisión manual de estadísticas dispersas y ampliando el universo de jugadores considerados más allá de los más conocidos.

### 5.6.5 Consideraciones prácticas para una interpretación fiable

Para obtener resultados representativos se recomienda trabajar con jugadores que tengan un volumen mínimo de participación registrada en la fuente, dado que las métricas se normalizan por 90 minutos y los percentiles se calculan dentro del grupo posicional. Jugadores con muy pocos minutos pueden presentar métricas infladas o percentiles poco fiables como consecuencia de la baja representatividad estadística de su muestra. Asimismo, las competiciones tienen mayor validez cuando se realizan dentro de la misma competencia y temporada, ya que el nivel competitivo y las características tácticas de cada liga condicionan directamente la distribución de las métricas y la interpretación de los percentiles.

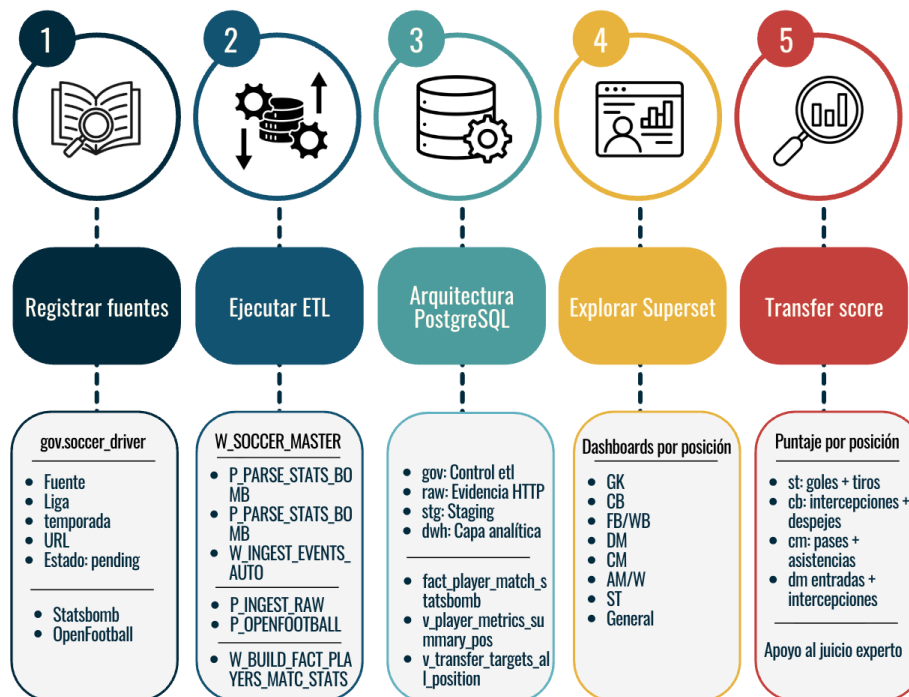


Ilustración 29: Guía del uso del sistema

## Capítulo 6. CONCLUSIONES

### 6.1 Conclusiones del trabajo

El presente trabajo permitió demostrar que la evaluación de jugadores en el fútbol profesional puede fortalecerse mediante una aproximación más estructurada, objetiva y reproducible, sustentada en el uso de datos abiertos, procesos de analítica avanzada y visualización interactiva. El principal aporte del proyecto consistió en el desarrollo integral de una arquitectura de datos funcional, métricas comparables por posición y un sistema de apoyo a la toma de decisiones aplicado al scouting deportivo.

Desde una perspectiva académica, la investigación contribuye a reducir una brecha identificada en la literatura especializada: la escasez de propuestas que integren de forma coherente la adquisición de datos, su transformación mediante procesos ETL, el almacenamiento analítico, la evaluación multicriterio y la visualización orientada al decisor. En este sentido, el proyecto amplía el conocimiento aplicado en la intersección entre el fútbol profesional, la analítica de datos y los sistemas de apoyo a la decisión, al demostrar que es posible transformar fuentes abiertas en una herramienta útil y accesible para comparación objetiva de perfiles de jugadores.

Asimismo, el desarrollo del proyecto evidenció limitaciones concretas asociadas a la disponibilidad de datos y al alcance de las herramientas tecnológicas implementadas, lo cual confirma que este campo de investigación presenta un amplio margen de desarrollo. En consecuencia, el proyecto sienta las bases para futuras líneas de trabajo relacionadas con la incorporación de nuevas fuentes de datos, la ampliación y refinamiento de métricas, el ajuste dinámico de ponderaciones por posición la inclusión de variables económicas vinculadas al mercado de transferencias y la validación práctica de herramientas en contextos reales del fútbol profesional.

En conclusión, este trabajo demuestra que la ingeniería de datos, la analítica y la visualización pueden integrarse para generar valor concreto en procesos deportivos de alta complejidad decisional, y que dicha integración es viable incluso en entornos con recursos y acceso a datos limitados, siempre que exista una metodología sólida y coherente con el problema abordado.

#### 6.1.1 Ejemplo de uso: Evaluación de un defensor central con scouting tool v2.0

##### 6.1.1.1 Escenario planteado

El cuerpo técnico de un club de fútbol profesional necesita reforzar la posición de defensor central antes del próximo mercado de transferencias. El director deportivo le encarga al analista de datos identificar perfiles defensivos sólidos con buena capacidad de salida del balón, dentro del universo de jugadores disponibles en el sistema. El

analista no cuenta con el tiempo para revisar manualmente las estadísticas individuales de los diferentes jugadores a nivel mundial para esa posición, y necesita una primera preselección objetiva y justificada que pueda presentar al cuerpo técnico en la siguiente reunión.

#### **6.1.1.2 Acceso al sistema y contexto del grupo analizado**

El analista accede al dashboard de defensores centrales en Apache Superset. El sistema ha procesado automáticamente los datos de StatsBomb y OpenFootbal, este presenta de forma inmediata un resumen del grupo posicional: 55 defensores centrales analizados, con un transfer score promedio de 0.4999, una media de 1.46 intercepciones por 90 minutos y 3.44 despejes por 90 minutos. Estos valores actúan como referencia comparativa del grupo: cualquier jugador que los supere significativamente está rindiendo por encima de sus pares posicionales en el mismo contexto competitivo.

Sin necesidad de abrir ninguna hoja de cálculo ni cruzar fuentes manualmente, el analista ya tiene en pantalla el ranking completo de los 55 defensores centrales, junto con el diagrama de caja que permite identificar visualmente quienes se sitúan en el cuartil superior.

#### **6.1.1.3 Identificación del candidato**

Al revisar el ranking, el sistema destaca a Bijoy Varghese con un transfer score de 0.7786, uno de los más altos del grupo y claramente por encima de la media poblacional. Esto significa que, ponderando conjuntamente sus métricas defensivas y de construcción, este jugador supera aproximadamente el 78% de los defensores centrales de su grupo comparativo. Para el analista, este dato es el punto de entrada, es una señal objetiva que justifica profundizar en el perfil.

Al descomponer el score en sus métricas individuales, el sistema revela lo siguiente:

- Intercepciones por 90 minutos: 2.55, frente a una media de 1.46. Esto indica una capacidad notable para leer el juego y cortar líneas de pase antes de que se conviertan en peligro, una de las cualidades más valoradas en un defensor central.
- Entradas por 90 minutos: 1.57, dentro del rango normal del grupo, lo que indica que su aporte defensivo se base tanto en la anticipación como en el duelo físico.
- Minutos totales: 458, un volumen de participación suficiente para que las métricas normalizadas por 90 minutos sean estadísticamente representativas y no estén infladas por una muestra pequeña.

#### **6.1.1.4 Cómo calcula el sistema transfer score**

El valor 0.7786 surge de una lógica multicriterio estructurada. El sistema calcula el percentil de cada métrica de Bijoy Varghese dentro de su grupo posicional, es decir, que proporción de los 55 defensores centrales rinde menos que el en cada variable. Después aplica la formula ponderada para CB:

$$TS_{CB} = 0.20 \cdot P(tackles\_per90) + 0.30 \cdot P(interceptions\_per90) + 0.25 \cdot P(clearances\_per90) + 0.15 \cdot P(pass\_completion\_pct) + 0.10 \cdot P(passes\_per90)$$

Las intercepciones reciben el mayor peso con 0.30, porque son el indicador más representativo de la anticipación y la lectura defensiva en esta posición. Los despejes 0.25 y las entradas 0.20, completan la dimensión defensiva, mientras que la precisión de pase 0.15 y el volumen de pases 0.10, incorporan la capacidad del jugador en la fase de construcción. El resultado es un indicador único, comparable y técnicamente coherente con el rol del defensor central.

#### **6.1.1.5 ¿Qué puede hacer el analista con esta información?**

Con estos datos, el analista puede presentar al cuerpo técnico una preselección justificada y objetiva. Bijoy Varghese aparece en la lista por las métricas normalizadas y percentiles que su rendimiento defensivo está por encima del 75% de sus pares comparables en el mismo contexto competitivo. Esta justificación basada en datos complementa, sin sustituir, la observación experta del scout, que podrá centrar su atención en un perfil ya filtrado y fundamentado.

Adicionalmente, el analista puede comparar directamente Bijoy Varghese con otros defensores del ranking, como Jose Luis Espinosa Arroyo con un 0.8369 o Sandesh Jhigan de 0.7917, para contrastar perfiles y decidir a cuáles merece la pena hacer seguimiento más detallado.

#### **6.1.1.6 Valor real demostrado**

Este ejemplo demuestra de forma concreta que Scouting tool v2.0 aporta valor real en tres dimensiones. En eficiencia, porque reduce de horas a segundos el proceso de identificación y comparación de defensores centrales dentro de un universo de 55 jugadores. En objetividad, porque el ranking depende de las métricas estandarizadas, normalizadas por tiempo de juego y calculadas dentro del mismo contexto posicional y competitivo. En interpretabilidad, porque cualquier resultado puede descomponerse en

sus métricas individuales, haciendo el análisis transparente y explicable ante el cuerpo técnico, algo que los modelos predictivos de caja negra no siempre permiten.

En definitiva, este caso de uso confirma que es posible construir una herramienta de scouting funcional, accesible y de valor real utilizando exclusivamente datos abiertos y herramientas de código abierto, lo que constituye el aporte central de este proyecto.

## **6.2 Conclusiones personales**

A nivel personal, este proyecto representó un proceso de aprendizaje significativo tanto en el plano técnico como en el metodológico. Su desarrollo permitió fortalecer competencias vinculadas a la integración de datos, el diseño e implementación de procesos ETL, la organización de información en capas analíticas y la construcción de visualizaciones orientadas a la toma de decisiones. Más allá de los aspectos técnicos, el proyecto también contribuyó a comprender con mayor profundidad cómo la tecnología puede aplicarse de forma estructurada para resolver un problema real en el ámbito del fútbol profesional.

Una de las enseñanzas más relevantes del proceso fue la importancia de la adaptación metodológica durante el desarrollo de una investigación. A lo largo del proyecto fue necesario replantear ideas iniciales, ajustar el alcance y redefinir el enfoque para alinearlos con la viabilidad técnica, la disponibilidad real de los datos y el propósito concreto de la propuesta. Esta experiencia permitió comprender que la flexibilidad y la capacidad de reformular decisiones que constituyen a una condición inherente al desarrollo riguroso y honesto de cualquier proyecto de investigación aplicada.

Finalmente, esta experiencia aportó a la formación académica y crecimiento profesional, al posibilitar la integración de conocimientos de analítica, bases de datos, ingeniería de datos y visualización en una propuesta concreta y funcional. Más allá del resultado final obtenido, el proyecto dejó aprendizajes valiosos en torno a la disciplina, la resolución sistemática de problemas y la construcción de soluciones con sentido práctico y aplicabilidad real.

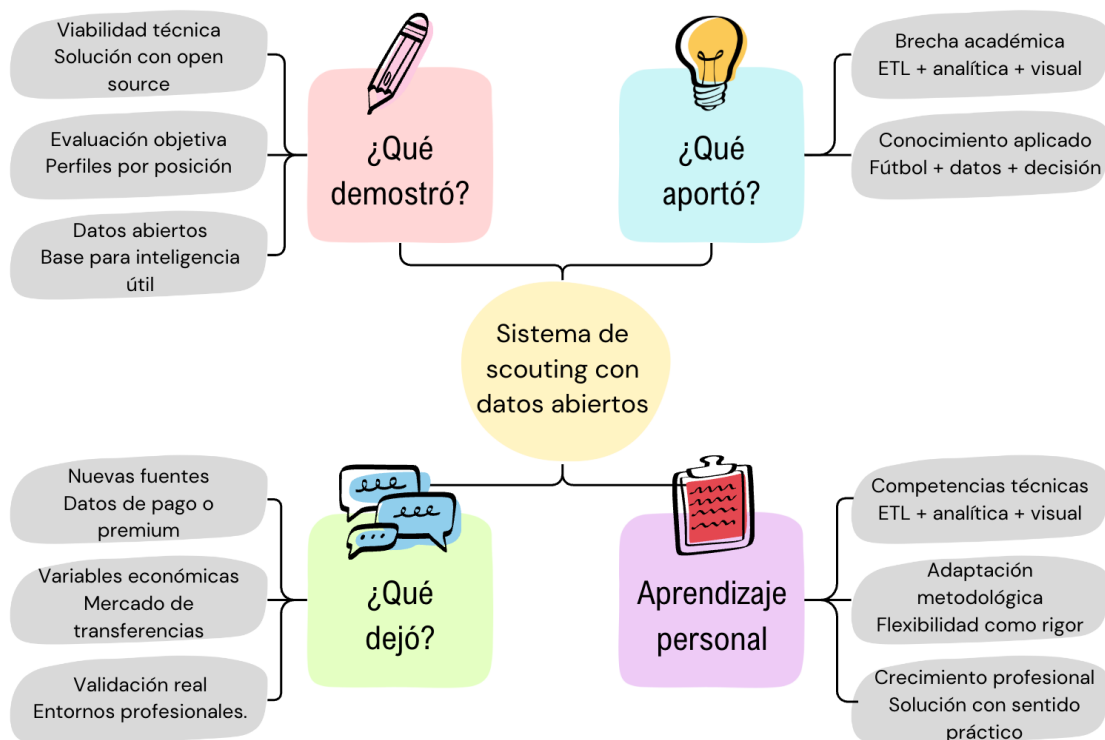


Ilustración 30: conclusiones del proyecto

## **Capítulo 7. FUTURAS LÍNEAS DE TRABAJO**

El desarrollo del presente proyecto es la base para diversas extensiones que permitirían fortalecer tanto el alcance analítico del sistema como su utilidad práctica en contextos reales de scouting y mercado de transferencias. A continuación, se describen las principales líneas de trabajo identificadas a partir de las limitaciones encontradas y de las oportunidades de mejora observadas durante el proceso de implementación.

### **7.1 Ampliación de la cobertura de datos**

Una extensión natural y prioritaria de la propuesta consiste en incrementar el número de ligas, competiciones y temporadas integradas en el sistema. Esta ampliación permitiría enriquecer la base histórica de análisis, aumentar la representatividad estadística de los perfiles construidos y realizar comparaciones más robustas entre jugadores que operan en distintos contextos competitivos. Un mayor volumen de datos históricos contribuiría además a mejorar la estabilidad de los percentiles posicionales y a así obtener un mejor indicador de transfer score al tener un mayor poder discriminatorio entre perfiles similares.

### **7.2 Incorporación de nuevas fuentes de datos**

Otro punto para tener en cuenta en un futuro es integrar fuentes de datos complementarias, incluyendo plataformas comerciales especializadas, con el propósito de acceder a variables más avanzadas, mayor profundidad histórica y métricas de rendimiento que actualmente no se encuentran disponibles en las fuentes abiertas utilizadas. De manera paralela, la incorporación de datos relacionados con el valor mercado y el costo de transferencia de los jugadores, proveniente de la fuente de TransferMarket, ampliaría significativamente la utilidad del sistema, al permitir que la evaluación contemple la dimensión deportiva del rendimiento y la económica, que resulta determinante en los procesos reales de toma de decisiones en el mercado de fichajes.

### **7.3 Mejora de la capa de visualización**

Aunque Apache Superset permitió construir dashboards funcionales dentro de un entorno de código abierto, una línea de mejora relevante consistente en evaluar la migración hacia herramientas de visualización con mayores capacidades de personalización e interactividad. Plataformas como Tableau, Power Bi, Looker o soluciones basadas en frameworks de desarrollo web permitirían diseñar interfaces flexibles, intuitivas y orientadas a la experiencia del usuario final. Este avance resultaría

valioso en contextos de uso real, donde la claridad visual y la facilidad de exploración de los datos son tan importantes como la calidad analítica del sistema subyacente.

## **7.4 Incorporación de funcionalidades analíticas avanzadas**

Finalmente, el sistema podría evolucionar hacia una propuesta más completa mediante la integración de técnicas de aprendizaje automático que amplíen sus capacidades más allá del análisis descriptivo y comparativo actual. Entre las extensiones más relevantes se encuentran la implementación de modelos de recomendación basados en similitud de perfiles mediante algoritmos de clustering como K-means, el desarrollo de módulos predictivos orientados a anticipar eventos de alto valor deportivo, como la detección temprana de jugadores con potencial excepcional, la predicción de riesgo de lesión a partir de indicadores de carga física, o la identificación de carga física, o la identificación de caídas sostenidas de rendimiento mediante series temporales. La construcción de sistemas de ranking dinámicos que ajusten automáticamente las ponderaciones del transfer score según el perfil táctico del equipo evaluador.

Estas extensiones permitirían que el sistema transite progresivamente desde su rol actual descriptivo y comparativo hacia uno verdaderamente predictivo y prescriptivo, respondiendo a la evolución natural de la analítica deportiva donde la ventaja competitiva se desplaza cada vez más hacia la capacidad de anticipar en lugar de simplemente describir, y consolidándose, así como una herramienta de apoyo a la decisión de mayor alcance para el scouting y la evaluación de futbolistas profesionales.

## **7.5 Validación en entornos reales**

Una línea transversal a todas las anteriores, es la validación del sistema en contextos reales de uso. Colaborar con analistas deportivos, directores técnicos o departamentos de scouting de clubes profesionales o semiprofesionales permitiría contrastar la utilidad práctica de la herramienta, identificar necesidades no contempladas en el diseño actual y orientar las mejoras futuras hacia criterios concretos de aplicabilidad. Esta validación externa haría que el proyecto fortaleciera su impacto real en el ámbito deportivo.

Andrea Ruiz Naranjo

---



*Ilustración 31: Conclusiones del proyecto*

## Capítulo 8. REFERENCIAS

Aditya, R. S. (2024). Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática. *Retos*, 1382-1389.

Fernández-Cortés Tolosa, J. A. (2024). *Análisis y evaluación de los indicadores de rendimiento en fútbol*. Badajoz: Universidad de Extremadura.

Muñoz Montoro, F. (2023). *Desarrollo de una aplicación para Scouting de jugadores en el mundo del fútbol: Optimización del proceso de identificación de talentos*. Valencia: Universitat Politècnica de València.

Soria Polo, C. (2021). *Diseño y aplicación de técnicas de machine learning para optimizar el Scouting en clubes de fútbol*. Sevilla: Universidad de Sevilla.

Vera Martínez, R. (2024). *Predicción y rendimiento: innovación en el scouting del fútbol con machine learning*. Barcelona: Universitat Pompeu Fabra.

Salas Saiz, G. (2025). *Análisis de rendimiento en el fútbol profesional usando aprendizaje automático*. Madrid: Universidad Politécnica de Madrid.

Enríquez Herrera, J. V. (2023). Business Intelligence & Data Analytics aplicado al proceso de seguimiento curricular en la universidad UPEC. *Minerva*, 9-20.

Martínez Santana, Y. (2025). *Diseño e implementación de un cuadro de mando para el análisis de la Primera División española de fútbol*. Valencia: Universitat Politècnica de València.

Grisales Aguirre, A. M., Espinosa Villa, L. M., & Alzate, P. M. (2024). La evolución de la visualización de datos: Campos de aplicación y herramientas de Business Intelligence. *Revista de Ingenierías Interfaces*.

Martínez Sánchez, J. A. (2024). *Factores determinantes del rendimiento deportivo en fútbol*. Sevilla: Universidad Pablo de Olavide.

Pinargote Cárdenas, R. E., & Alcívar Cedeño, A. (2025). *Analítica de Datos e Indicadores de Rendimiento para optimizar la gestión Universitaria*. Sinapsis.

Contreras García, J. M., & Cabrera Lozano, R. (2024). Visualización de datos con Power BI. *Epsilon: Revista de la Sociedad Andaluza de Educación Matemática "Thales"*, 69-79.

García, M. S. (2024). Influencia de las variables contextuales en las acciones técnicas en fútbol.

Paulis, J. C. (2025). Uso de los sistemas de juego en las ligas profesionales del fútbol español con relación a las variables situacionales. *Revista Iberoamericana de Ciencias de la Actividad Física y el Deporte*, 424-443.

García González, J. J. (2025). *Evaluación y Modelado del Sistema de Rating en FIFA*. Madrid: Universidad Politécnica de Madrid.

Santana Falcón, F. (2023). *Análisis de datos abiertos de fútbol aplicando la metodología CRISP-DM*. Madrid: Universidad Politécnica de Madrid.

Moro Portero, T. (2024). *Predicción del potencial de jugadores de fútbol mediante técnicas de machine learning*. Madrid: Universidad Politécnica de Madrid.

Cuadrado Camarero, E. (2025). *Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial*. Madrid: Análisis Predictivo y Visualización de Datos en Fútbol usando Inteligencia Artificial.

Barrera León, D. (2025). *Empleo de analítica de datos e IA para optimizar procesos de gestión deportiva en clubes de fútbol de Colombia*. Bogotá: Universidad de Ciencias Aplicadas y Ambientales.

## Capítulo 9. ANEXOS

### 9.1 Objetivos de desarrollo sostenible

El presente proyecto tiene relación con tres objetivos de desarrollo sostenible (ODS), los cuales se describen a continuación.

El vínculo más estrecho se establece con el ODS 9, industria, innovación e infraestructura, dado que el proyecto propone una solución tecnológica orientada a modernizar un proceso de análisis y toma de decisiones dentro de la industria del deporte profesional. La construcción de una arquitectura de datos basada en herramientas de código abierto, la implementación de procesos ETL automatizados, el almacenamiento analítico en capas y el desarrollo de dashboards interactivos representan una contribución concreta a la innovación tecnológica aplicada, demostrando que es posible construir infraestructuras de datos funcionales y reproducibles con recursos accesibles.

En segundo lugar, es el ODS 10, reducción de las desigualdades, ya que el uso de métricas estandarizadas, percentiles posicionales y un puntaje sintético configurable favorece criterios de evaluación más objetivos e igualitarios entre jugadores. Esta aproximación reduce el peso de factores subjetivos en los procesos de selección de talento, contribuyendo a que jugadores de distintos contextos competitivos pueden ser comparados en condiciones equitativas. Al utilizar herramientas de código abierto y fuentes de datos gratuitas, el sistema es accesible para clubes con recursos limitados, lo que contribuye a reducir la brecha tecnológica entre organizaciones deportivas de diferentes tamaños.

Finalmente, se relaciona el ODS 8, trabajo decente y crecimiento económico, en tanto que contribuye a la modernización de los procesos de evaluación y selección de talento en la industria del fútbol profesional. Al ofrecer una herramienta estructurada y basada en datos para el scouting, el proyecto apoya las profesiones vinculadas al análisis deportivo, favoreciendo decisiones más fundamentadas en la gestión de recursos humanos dentro del sector.

### 9.2 Uso de la Inteligencia artificial en el TFM

A lo largo del proyecto se recurrió al uso de herramientas de inteligencia artificial como apoyo en las siguientes :

- Arreglar los errores arrojados en las ETLs.

Andrea Ruiz Naranjo

---

- Ayudar a encontrar paginas donde se encontrará la data de jugadores y partidos gratuitamente.
- Mejorar las fórmulas o métricas del transfer score por posición.
- Ayudar a mejorar la redacción de algunos párrafos en la memoria.

### **9.3 Entrega practica**

En el siguiente link de GitHub se encuentra el proyecto práctico que se realizó:  
[https://github.com/AndreaRuizNaranjo/bigdata\\_analytics-football.git](https://github.com/AndreaRuizNaranjo/bigdata_analytics-football.git)

Diseño e implementación de un sistema de apoyo a la toma de decisiones para el scouting de jugadores mediante herramientas Big Data.



**Universidad  
Europea**

Andrea Ruiz Naranjo

---