



# Informe de posibles fichajes

Septiembre 2023

—

Óscar Hernández

Departamento de Análisis,  
Club Atlético de Madrid SAD



<b>VISIÓN GENERAL.....</b>	<b>2</b>
<b>PREGUNTAR.....</b>	<b>2</b>
<b>PREPARAR.....</b>	<b>4</b>
<b>PROCESAR.....</b>	<b>6</b>
<b>ANALIZAR.....</b>	<b>6</b>
<b>COMPARTIR.....</b>	<b>7</b>
<b>ACTUAR.....</b>	<b>11</b>



## VISIÓN GENERAL

Este documento es parte del curso 8 del Curso de Análisis de Datos elaborado por Google que estoy realizando en Coursera. Este es un supuesto práctico donde demostraré todo lo que he aprendido en el curso y estará dividido en distintos informes relacionados con cada fase del proceso del análisis, los cuales se asemejan a los siguientes apartados:

- Una declaración clara de la tarea empresarial que elegiste investigar
- Una descripción de todas las fuentes de datos utilizadas
- Documentación de todas las limpiezas y manipulaciones de datos
- Un resumen de tu análisis
- Visualizaciones de respaldo y hallazgos clave
- En función de tu descubrimiento, una lista de entregables adicionales que crees que sería útil incluir para exploración posterior
- Tus conclusiones de alto nivel más importantes basadas en tu análisis

## PREGUNTAR

Para completar este apartado, se responderá a una serie de preguntas.

### **¿Qué tema estás explorando?**

Soy aficionado al fútbol y a los datos. Hasta hace relativamente poco, los datos sobre este deporte eran escasos y con poco valor, a diferencia de otros deportes como baloncesto o béisbol. Hoy en día ya existen multitud de empresas que recogen gran cantidad de datos de valor a tiempo real en partidos de fútbol y además se confía más en el valor de estos.

Debido a todo lo anterior, he decidido realizar esta actividad sobre datos relativos al fútbol, obtenidos de una de tantas páginas web con estos datos: *Understat*.

La empresa donde simulo estar trabajando será el equipo de fútbol Atlético de Madrid, en el departamento de análisis que tienen todos los clubes profesionales actualmente.

### **¿Cuál es el problema que intentas resolver?**

Los departamentos de análisis de un equipo de fútbol desarrollan multitud de tareas, pero me centraré en el mercado de fichajes. Durante los meses de julio y agosto, que son el periodo que separa las temporadas de clubes, se realizan distintos movimientos de jugadores entre equipos, en lo que se denomina mercado de fichajes. El trabajo de buscar jugadores ideales para el equipo se realiza durante todo el año, para tener posibles nombres encima de la mesa



durante el mercado. En ocasiones, equipos más fuertes económicamente consiguen fichar a jugadores con los que se quería seguir contando, por eso siempre se debe tener alternativas. El problema que intento resolver es este precisamente, conseguir una lista de jugadores que se pudieran fichar si de un día para otro de deja de contar con un miembro importante del equipo.

### **¿Qué métricas utilizarás para medir tus datos a fin de lograr tu objetivo?**

La gran mayoría de los datos sobre jugadores o equipos no son gratuitos. Por ejemplo, la web utilizada es la única que ofrece coordenadas de los tiros realizados por jugadores de forma gratuita. Esta web ofrece multitud de variables, pero las principales que serán la base del análisis serán las siguientes: goles esperados, asistencias esperadas y participación en goles esperados.

### **¿Quiénes son los interesados?**

La empresa es la principal interesada. Tener un plan b para posibles bajas de jugadores siempre será interesante tanto para la parte deportiva de la empresa como para la parte económica, puesto que mediante este análisis de minimizarían los gastos.

### **¿Cuál es tu audiencia?**

Por lo general, actualmente existe la figura de director deportivo, que es la figura de conexión entre lo económico y lo deportivo, puesto que sabe del presupuesto que se dispone y en qué se debe gastar (comprar, vender, mejorar contratos, ...) Es la persona a la que va dirigido este análisis, puesto que es la que debe dar el visto bueno a cualquier movimiento de jugadores. Pese a que en este caso se realiza de forma individual, este informe se suele realizar por el equipo que forma parte del departamento de análisis.

### **¿Cómo pueden ayudar tus conocimientos a que el cliente tome decisiones?**

No todos los equipos de fútbol acaban de fiarse del todo de los datos para fichar jugadores. Pese a que los datos son una variable más a tener en cuenta para el fichaje, suelen ponderar más las sensaciones de los expertos al ver jugar a los jugadores. Como se ha aprendido en el curso, se debe utilizar el razonamiento analítico y no dejarse llevar por sensaciones para la toma de decisiones empresariales. Este análisis y su correspondiente información-conocimiento es lo que puede ayudar al director deportivo a tomar decisiones basadas en datos.



## PREPARAR

Los datos escogidos son los que proporciona la página web *Understat*: <https://understat.com/>. Esta página ofrece datos por partido o por jugador y contiene información de las 5 grandes ligas( Inglaterra, España, Alemania, Francia e Italia) además de la liga rusa.

Se ha escogido esta fuente de datos porque es de las más completas de las que están disponibles de forma gratuita.

La recolección de datos ha sido mediante web scrapping, modificando y actualizando el código creado por @douglasbc en <https://github.com/douglasbc/scraping-understat-dataset>

Existen varias formas más de obtener estos datos, como son:

- Están disponibles en varias fuentes de datos de Kaggle como en <https://www.kaggle.com/datasets/abrarhossainhimself/understat-data-for-teams-players-2014-present>
- Existe una librería de R creada por *Understat* para tener acceso a estos datos.

Los datos sobre jugadores están guardados en archivos csv por temporada y liga. Los datos sobre partidos están guardados en un único archivo csv. Los datos se obtienen de esta forma, pero para prepararlos de forma adecuada se eliminarán temporadas antiguas, solo quedándonos con las 3 últimas además de eliminar la liga rusa, puesto que los datos obtenidos allí por los jugadores no son comparables con el nivel del resto de ligas.

Respecto la credibilidad de los datos, existen métricas más simples como minutos jugados, pases completados, etc que son fiables y exactas. La duda viene sobre otras métricas basadas en la esperanza como los goles esperados. En la vista principal de Understat hablan sobre como se calculan estas métricas. Resumiendo, es mediante una base de datos inmensa y utilizando redes neuronales. Como estos datos se crean mediante modelos, cada empresa que ofrece datos de goles esperados da valores diferentes dependiendo de su modelo, pero las diferencias no son significativas.

Este análisis se realiza de forma individual. Si se realizara en equipo la opción ideal sería utilizar la librería de R para obtener solo los datos necesarios y no modificarlos.

Como en cualquier empresa, para la elaboración del informe y del análisis, se quiere invertir el menor tiempo posible, por tanto, para la descripción de la base de datos se copiará aquí la descripción realizada por @douglasbc en su Github.

**Players columns:**

- id - Unique identifier number for every player
- player\_name - Player name
- games - Total amount of games played
- time - Total amount of minutes played
- goals - Total amount of goals scored
- xG - Sum of Expected Goals of every shot, calculated by the Understat xG model
- assists - Total amount of assists
- xA - Sum of Expected Goals of shots from a player's key passes
- shots - Total amount of shots
- key\_passes - Total amount of key passes
- yellow\_cards - Total amount of yellow cards
- red\_cards - Total amount of red cards
- position - Every position played during the season
- team\_title - Name of the team
- npg - Non penalty goals
- npxG - Sum of Expected Goals of every shot, excluding penalties
- xGChain - Total xG of every possession the player is involved in
- xGBuildup - Total xG of every possession the player is involved in without key passes and shots

**Shots columns:**

- id - Unique identifier number for every shot
- minute - Time of the shot
- result - Goal, MissedShots, SavedShot, BlockedShot, ShotOnPost, OwnGoal
- X - X coordinate of the shot, between 0 and 1
- Y - Y coordinate of the shot, between 0 and 1
- xG - Number Between 0 and 1, describing the Expected Goals of the shot, calculated by the Understat xG model
- player - Player that made the shot
- h\_a - Home/away team
- player\_id - Unique identifier number for every player
- situation - OpenPlay, FromCorner, SetPiece, DirectFreekick, Penalty
- season - The year when the season started, so season 2014/2015 is 2014, season 2015/2016 is 2015
- shotType - RightFoot, LeftFoot, Head, OtherBodyPart
- match\_id - Unique identifier number for every match
- h\_team - Home team
- a\_team - Away team
- h\_goals - Goals scored by home team
- a\_goals - Goals scored by away team
- date - Date of the match
- player\_assisted - Player that assisted the shot
- lastAction - Last action before the shot (None, Pass, Cross, Aerial, Standard, TakeOn, Rebound...)



Como enlaces de interés:

- Artículo explicando los *Goles Esperados*.  
<https://es.coachesvoice.com/cv/goles-esperados-o-xg-en-que-consisten/>
- Explicación de otras métricas calculadora mediante modelos.  
<https://objetivoanalista.com/tag/metricas/>

## PROCESAR

- ¿Qué herramientas eliges y por qué?
- ¿Has garantizado la integridad de los datos?
- ¿Qué pasos seguiste para garantizar que tus datos están limpios?
- ¿Cómo puedes verificar que tus datos están limpios y listos para analizar?
- ¿Documentaste tu proceso de limpieza para poder revisar y compartir estos resultados?

Documentación de todas las limpiezas y manipulaciones de datos

Para el procesamiento de datos, se ha escogido la herramienta RStudio, puesto que es la más completa de las estudiadas. Dentro de este programa se realiza el procesamiento, la limpieza, el análisis y la visualización de datos. Puesto que la recopilación de datos se ha realizado correctamente y los datos de *Understat* son fiables, la integridad de datos para el análisis individual está garantizada.

Ya se ha comentado la organización de los datos por liga y temporada en archivos csv, así como de los tiros realizados en todos los partidos estudiados.

Respecto la limpieza de datos, solo se ha tenido que separar la columna *Posición* en tres columnas, para facilitar el análisis posterior. Este único cambio ha sido aclarado en el archivo para posibles revisiones.

Para la limpieza de datos del Data Frame de tiros, se han eliminado valores nulos creados por el proceso de web scrapping y se seleccionan los tiros más recientes.

Para cada subselección de datos correspondiente a los jugadores de una posición, se han realizado cambios concretos que se describen en el propio análisis de datos.

## ANALIZAR

El proceso de análisis y del procesamiento propio de cada subconjunto de datos se ha realizado mediante R Markdown. El informe correspondiente donde se ve todo el proceso y se explica estará también disponible.



## COMPARTIR

Para realizar la presentación de los aspectos importantes y las conclusiones de este estudio, se realizará mediante Google Slides. Se incorporarán aquí como imágenes las diapositivas creadas más relevantes.



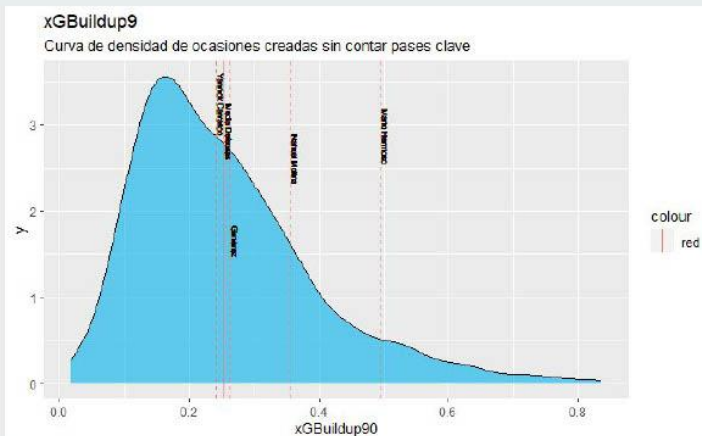




## DEFENSAS

Joshua Kimmich  
Raphael Guerreiro  
Alphonso Davies  
Noussair Mazraoui  
Pascal Groß  
Kaoru Mitoma  
Trent Alexander-Arnold  
Kieran Trippier  
Estupiñán  
Andrew Robertson  
Jordi Alba  
Edon Zhegrova  
Nuno Mendes  
Mário Rui

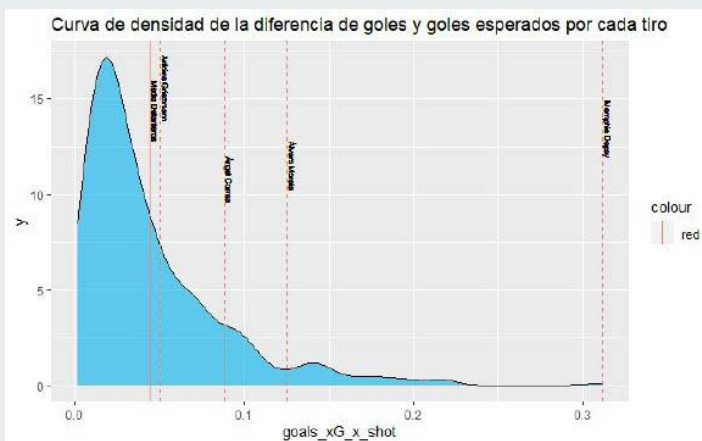
## INFORME DE POSIBLES FICHAJES DEL ATLÉTICO DE MADRID



## DELANTEROS

Munas Dabbur  
Roberto Firmino  
Neymar  
Vincenzo Grifo  
Serge Gnabry  
Kylian Mbappe-Lottin  
Romelu Lukaku  
Christopher Nkunku  
Mohamed Salah

## INFORME DE POSIBLES FICHAJES DEL ATLÉTICO DE MADRID



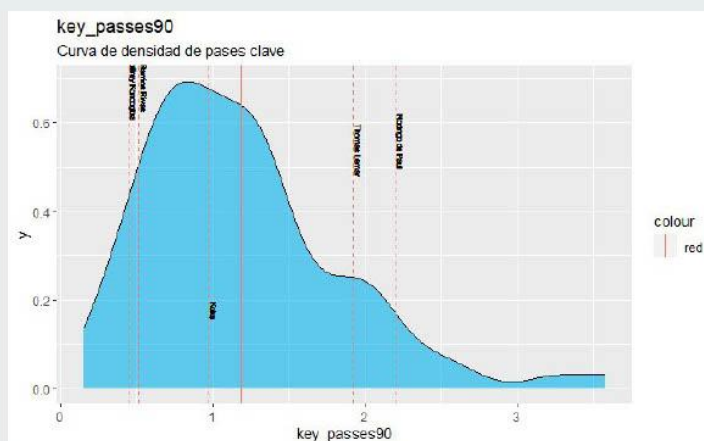


## CENTROCAMPISTAS

Jamal Musiala  
Kingsley Coman  
Bruno Fernandes  
Kevin De Bruyne  
Fábio Vieira  
NGolo Kanté  
Luka Modric  
Rodrigo de Paul\*  
Toni Kroos  
Joaquín  
Maxence Caqueret

\*Actual jugador del Atlético de Madrid

## INFORME DE POSIBLES FICHAJES DEL ATLÉTICO DE MADRID



## INFORME DE POSIBLES FICHAJES DEL ATLÉTICO DE MADRID

### CONCLUSIONES

- Se ha situado a **nuestros jugadores** respecto al resto para poder estudiar sus **aspectos a mejorar**.
- Se han **encontrado nuevos jugadores** que igualan o mejoran a los nuestros.
- Se han encontrado grandes jugadores desconocidos que pueden ser una gran **oportunidad de mercado**.



## INFORME DE POSIBLES FICHAJES DEL ATLÉTICO DE MADRID

### OBJETIVOS

- Estudiar las características de nuestros jugadores y compararlos con el resto.
- Buscar jugadores adecuados para cada puesto mediante el uso de datos.



## INFORME DE POSIBLES FICHAJES DEL ATLÉTICO DE MADRID

Óscar Hernández Gallego

Departamento de Análisis,  
Atlético de Madrid.

Septiembre 2023



# FIN

Óscar Hernández Gallego

Departamento de Análisis,  
Atlético de Madrid.

Septiembre 2023

## ACTUAR

Tras realizar todas las fases del proceso del análisis de datos, queda el último paso. En este caso, el objetivo del estudio era prevenir un posible suceso, por lo que no se debe actuar por el momento.

En el hipotético caso de que hubiese que fichar un nuevo jugador, dependiendo de:

- ¿En qué posición ha sido la baja?
- ¿Cuáles son las cualidades concretas del jugador que causa baja?
- ¿Cuál es el presupuesto del fichaje y salario del nuevo jugador?
- Posibles oportunidades de negocio como jugadores por los que no haya que pagar traspaso o no estén dispuestos a seguir en su equipo.
- Preferencias del entrenador.
- Posibles beneficios económicos. Por ejemplo, fichar a un jugador de un país con poca cultura futbolera es una táctica interesante a nivel de ampliación de marca.

se fichará a un jugador, pero a nivel deportivo, se ha proporcionado con anterioridad una lista de buenos sustitutos gracias a este trabajo.

Generalmente, la figura que pone en común todas las situaciones mencionadas anteriormente y pondera lo más importante, junto con la ayuda del entrenador, es el director deportivo.