

Universidad de Costa Rica

Escuela de Matemáticas

Departamento de Ciencias Actuariales

Profesor: Joshua Cervantes Artavia

Asignatura: CA0204 – Herramientas de ciencia de datos I

Grupo: 02

Trabajo escrito del proyecto Individual

Tema: Análisis comparativo del gasto y tendencias en los fichajes de las principales ligas europeas de fútbol.

Por: Robert Sivianny Meléndez Jiménez

Carnet: C34772

II Semestre 2025

10 de noviembre de 2025

### Tema de investigación:

Análisis comparativo del gasto y tendencias en los fichajes de las principales ligas europeas de fútbol.

### Objetivo General:

Analizar las estrategias de fichaje de las tres principales ligas europeas durante la temporada 2024-2025, identificando las diferencias en gasto total, nacionalidades predominantes, posiciones más cotizadas y la relación entre su rendimiento en la temporada anterior y su valor de transferencia.

### Explicación del código:

La metodología empleada en este proyecto se fundamenta en la obtención, depuración, análisis y visualización de datos relacionados al mercado de fichajes 2024-2025 de las tres principales ligas europeas: LaLiga, Premier League y Bundesliga. Vamos a explicar un poco el código, las conclusiones a las cuales se llegó y mis recomendaciones.

El código inicia con la carga de los datos obtenidos desde Transfermarkt utilizando la biblioteca worldfootballR, que permite extraer información deportiva de manera segura y efectiva. Esta biblioteca facilita la recopilación de variables importantes sobre jugadores, equipos, posiciones, valor de mercado, edad, minutos disputados y otros datos necesarios para el análisis. Sin embargo, no basta con descargar los datos, la primera dificultad sucedió al intentar unificar diferentes bases que contenían información en formatos ligeramente distintos, obligando a renombrar variables, eliminar inconsistencias y unificar criterios.

Para evitar tener que depender de la conexión con la página y disminuir el tiempo que duraba en compila el proyecto, se exportaron los datos a un archivo Excel y así ya no dependíamos de datos de terceros.

Posteriormente, se filtró el dataframe original para quedarse únicamente con las ligas de interés. La limpieza de datos se llevó a cabo con el uso extensivo de dplyr, mediante funciones de filtrado, ordenamiento, agrupación y resumen (como filter, arrange, group\_by, summarise). Se construyeron sub data frames que permitieron responder a las preguntas planteadas en el objetivo general de la investigación. Cada uno de estos subconjuntos se creó con una intención específica, por ejemplo: identificar a los fichajes más caros, encontrar los jugadores con mayor cantidad de minutos jugados, identificar a los goleadores, analizar distribución por edad, estudiar nacionalidades predominantes, calcular gasto total por liga y determinar gasto por equipo, por posición y por ventana de fichajes.

Una vez completada la fase de depuración, se pasó a la etapa de visualización. Aquí la librería principal fue ggplot2, Se emplearon gráficos de barras, gráficos de líneas, diagramas de caja, boxplots, y visualizaciones horizontales

mediante `coord_flip()` para asegurar que los nombres de los jugadores fueran legibles. Además, la librería `scales` permitió formatear valores numéricos, una tarea necesaria considerando que los montos se expresan en millones de euros.

Un punto clave fue asignar colores por liga tal que LaLiga se representó en azul, la Bundesliga en rojo y la Premier League en morado, escogidos según el color de su logo. Esta ayuda visual no solo facilita la lectura de cada gráfico por separado, sino que también permite hacer comparaciones intuitivas entre ellos a lo largo de todo el informe. El objetivo siempre fue comunicar información de manera clara, coherente y sencilla, incluso para personas que no sepan del tema.

En gráficos como el del gasto total por liga, por ventana o por zona del campo, se utilizaron barras verticales con etiquetas numéricas para facilitar la interpretación. En el caso de los top 10 fichajes más costosos, se aplicó coordenadas horizontales para mejorar su comprensión. Para el análisis de edad, se emplearon gráficos de línea que reflejan tendencias más allá de un único punto. Para comparar dispersión de precios, se utilizaron boxplots, ya que se puede observar variabilidad, mediana, cuartiles y outliers.

Finalmente, el código finaliza con un análisis conjunto de las dos partes de la investigación que serían el rendimiento deportivo y el mercado de pases. Esto permitió generar una visión más completa, que no se limita únicamente a temas económicos, sino que se busca relacionarlo con la gloria deportiva. Para ello se calcularon correlaciones lineales con el coeficiente de Pearson, evaluando la relación entre el gasto total y puntos, así como entre fichajes y goles marcados. Estas mediciones cuantitativas permiten afirmar, pero ya con números y no por intuición que tan fuerte es la relación entre inversión y rendimiento en la temporada.

## **Resultados:**

Los resultados que arroja la investigación muestran un comportamiento claro del mercado de fichajes en la temporada 2024-2025. Una de las conclusiones más evidentes del análisis es que la Premier League se posiciona como la liga con mayor poder adquisitivo. Sus clubes realizan fichajes de mayor valor, poseen una mayor dispersión en sus montos son líderes en los rankings de fichajes más caros. Esto no es ninguna casualidad, la estructura económica del fútbol inglés, impulsada por ingresos televisivos y la marca global que es la Premier League, permite inversiones significativas incluso en clubes de media tabla o en equipos que posteriormente terminan descendiendo.

En contraste, LaLiga muestra un comportamiento discreto, con montos considerables, pero más equilibrados. Sus clubs tienden a realizar menos inversiones y por costos menores, enfocándose en posiciones que consideran prioritarias. La Bundesliga, por su parte, presenta un gasto más regular, con una

tendencia marcada hacia jugadores jóvenes y promesas de proyección. Estas ligas optan mas por la cantera que por la cartera ya que la tendencia en estos últimos años es que los equipos de la liga española y alemana no pueden luchar económicamente contra los equipos de la Premier league ya que sus presupuestos son inalcanzables hasta para los mejores equipos de otras ligas, como se puede notar cuando se compara el presupuesto de los 3 peores equipos contra LaLiga.

Los análisis por posición demostraron que delanteros y defensas son las zonas del campo donde se desembolsa mayor cantidad de dinero. Esto se explica porque son los mayores protagonistas del juego. Los delanteros son los encargados de los goles y en ellos recae el rendimiento del equipo, ya que sin la emoción que le dan el futbol no sería el mismo, por otro lado, los defensas centrales son los encargados en evitarlos. Los mediocampistas también representan un gasto importante, aunque más equilibrado, mientras que los porteros se consolidan como la posición menos costosa en el promedio. Aunque los porteros aparecen como la posición menos costosa en promedio, esta estadística puede resultar engañosa. A diferencia de otras posiciones, el arquero ocupa un rol único en el campo, solo hay uno por equipo. En cambio, posiciones como mediocampistas, defensas o delanteros incluyen múltiples variantes tácticas y roles específicos, lo que implica un mayor número de jugadores por plantilla. Esta diferencia genera una mayor dispersión en los costos individuales y puede inflar el promedio de otras posiciones, haciendo que el costo relativo del arquero parezca más bajo de lo que realmente es.

El análisis por nacionalidades muestra que las ligas prefieren talento local, aunque la Premier exhibe una notable globalización en los movimientos de mercado. LaLiga también demuestra presencia importante de jugadores latinoamericanos y europeos, mientras que la Bundesliga mantiene un equilibrio entre jugadores alemanes y extranjeros jóvenes.

Los gráficos relacionados con la edad evidencian que el mayor gasto se concentra entre los 24 y 28 años, rango en el cual un jugador combina experiencia, madurez táctica y pleno de condiciones físicas. A partir de los 30 años, las ligas tienden a reducir su inversión, en línea con el declive natural del rendimiento físico y la devaluación acelerada de este tipo de jugadores.

En contraste, las contrataciones de jugadores muy jóvenes son casos atípicos. Aunque su número es menor, estas promesas suelen implicar apuestas estratégicas de alto riesgo y potencial, donde el gasto no responde a rendimiento inmediato sino una apuesta por fichar a la futuras estrellas a un precio menor, pero con mucho riesgo ya que estas jugadas no siempre salen bien.

El análisis de los minutos jugados indica que los jugadores con mayor cantidad de minutos tienden a tener un valor de mercado superior, aunque este patrón no es del todo cierto ya que se pueden presentar excepciones debido a lesiones, descanso y decisiones técnicas.

Por último, la correlación entre gasto y rendimiento demuestra una relación positiva pero moderada. Los equipos que más invierten suelen tener mejores resultados, pero no de manera absoluta. Hay casos de equipos con gasto elevado que no logran posiciones altas, y casos de equipos con menor inversión que muestran estabilidad competitiva.

### **Recomendaciones:**

La investigación permite extraer varias recomendaciones y hasta un posible continuación del proyecto, pero se van a dar recomendaciones tanto desde la parte metodológica además de la perspectiva deportiva.

Desde el punto de vista de continuar con el proyecto, es recomendable ampliar su base de datos ya que con más información se podrán ir viendo los comportamientos típicos de forma más rápida y concreta. Adicionalmente se podría haber implementado un mayor número de variables, así de esta forma se podrían realizar comparaciones entre variables y conseguir métricas valiosas para el análisis.

Por otro lado, desde un punto de vista de clubes y gestión deportiva, el análisis demuestra que el gasto debe ir acompañado de planificación y méritos deportivos. Invertir una suma grande de dinero no te asegura alcanzar la gloria deportiva, pues el dinero el dinero es una herramientas mas dentro de un proyecto que involucra muchas más variables. Los equipos realmente exitosos no se basan únicamente en los fichajes más costosos, sino en una estructura solida que los respalda, una idea clara de juego, un entrenador que sepa sobrellevar las adversidades, un gran ambiente entre los jugadores y una visión deportiva clara a largo plazo.

Para mi los clubes antes de querer fichar los jugadores más caros, deberían mejorar su planificación interna de la temporada, construir un gran equipo competitivo. Un plantel bien organizado y con una idea de juego es lo primordial. Una vez esta base existe si es hora de realizar fichajes puntuales estratégicos que esperamos que generen un gran impacto. Los fichajes deberían ser la última pieza del rompecabezas y no la base de un proyecto ganador.

Un equipo que ficha poco, pero ficha bien, puede tener mejores resultados que uno que gasta en exceso sin un plan deportivo consistente. Un gran ejemplo fue el PSG que durante mucho años se dedicaba a ser protagonista en los mercados de fichajes, pero sin tener resultados hasta que decidieron por armar un equipo y aunque ya no tenían estrellas de renombre lograron algo más complicado formar un

equipo ganador y fuerte en todas sus líneas. Reforzar una o dos posiciones clave es una estrategia mucho más eficiente que renovar media plantilla cada temporada sin un criterio definido.

En conclusión, la eficiencia en el gasto depende más de la capacidad de planificación que de la cantidad invertida. Construir primero un equipo competitivo y después fichar piezas específicas es la estrategia más sostenible y eficaz. Un fichaje de calidad puede marcar la diferencia, pero solo si se integra en un equipo que ya tiene un proyecto bien definido.

Finalmente, para futuras versiones del análisis, sería recomendable enriquecer la base con información de rendimiento avanzado, así como automatizar la actualización de los datos, permitiendo replicar el estudio para otras temporadas y realizar comparaciones históricas más sólidas.

#### Referencias:

Ziv, J. (n.d.). *worldfootballR: Extract and clean world football (soccer) data* [R package]. GitHub. <https://jaseziv.github.io/worldfootballR/>

Transfermarkt. (n.d.). *Fichajes de fútbol, rumores, valores de mercado y noticias*. <https://www.transfermarkt.es/>

ESPN. (2025, mayo 7). *El cambio estructural que llevó a PSG a la final de la Champions League*. [https://www.espn.co.cr/futbol/champions-league/nota/\\_id/15150173/cambio-estructural-que-llevo-a-psg-a-final-de-champions-league](https://www.espn.co.cr/futbol/champions-league/nota/_id/15150173/cambio-estructural-que-llevo-a-psg-a-final-de-champions-league)

CIMEC. (n.d.). *Coeficiente de correlación de Pearson: Qué es, cómo se calcula y ejemplos*. <https://www.cimec.es/coeficiente-correlacion-pearson/>

Marca. (2023, julio 1). *LaLiga pierde valor: solo cinco clubes superan los 100 millones en plantilla*. <https://www.marca.com/futbol/primera-division/2023/07/01/649994bf46163f20b98b45d3.html>