EDA ANALISIS DEPORTIVO – IBAI COSGAYA PRIETO

1. Introducción

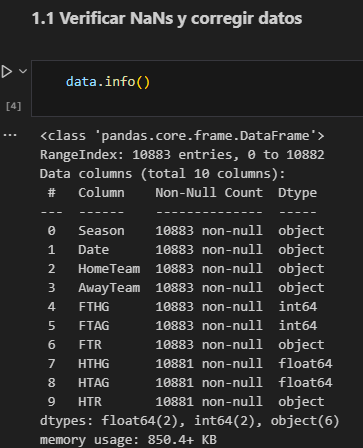
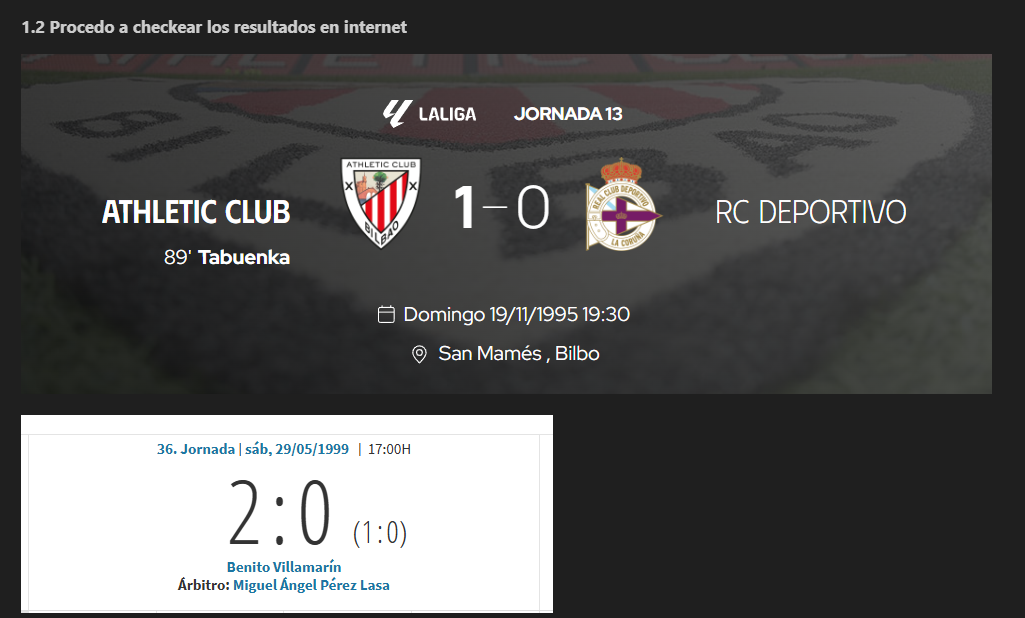
Mi EDA es un análisis deportivo sobre el Athletic Club. El deporte es una parte fundamental en mi día a día y es algo que me hace sentir pleno. Es por eso que he basado mi EDA en un deporte que suelo practicar a menudo: el futbol. Y unido con el fútbol, un equipo con el que me siento identificado, por los valores que transmite y el sentimiento de pertenencia que genera es el Athletic Club.  
  
He partido de una hipótesis: **“El Athletic Club tiende a remontar partidos cuando está perdiendo al descanso y juega de local”.** La idea viene de un sentimiento que tengo, y es que el Athletic es lo que es, en parte por el empuje del público de la Catedral. Que ha sido un equipo de conseguir remontadas. Ese es el punto de partida.

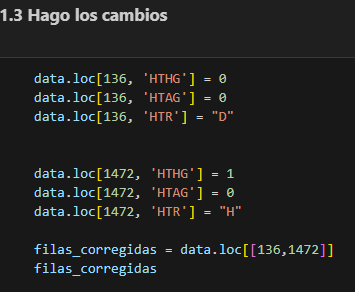
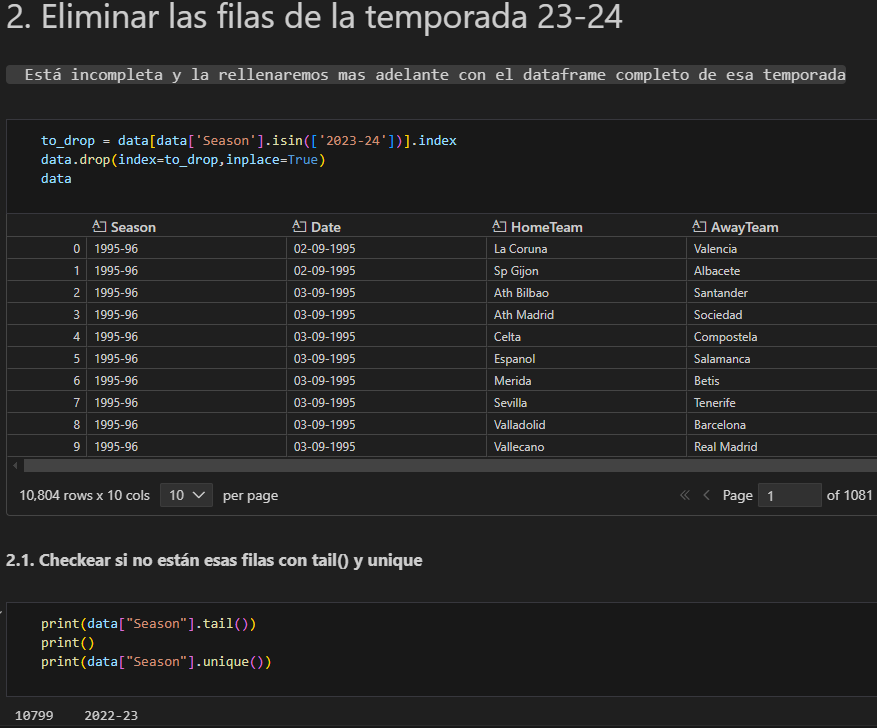
1. Obtención de datos y limpieza

Para analizar esta hipótesis he conseguido un Dataset de todos los partidos de La Liga desde la temporada 1995/96 a la 2023/24. El Dataset original de kaggle contenía información de métricas básicas como la temporada, la fecha, el equipo local, el visitante, resultado al descanso y resultado al final.

Tan solo había dos Nans. La limpieza en ese aspecto fue fácil porque tan solo tuve que mirar en internet los datos de esos partidos para

La temporada 23/24 estaba incompleta, es por eso que tuve que descargarme otro Dataset de esa temporada, quedarme con las mismas columnas que el Dataframe original y juntarlas:

Una vez que me quedo con las columnas necesarias, procedo a unirlo al main dataset, para así tener un dataset completo de la temporada 1995/96 a la 2023/24. Con esto ya estoy listo para poder empezar a filtrar lo que necesito.

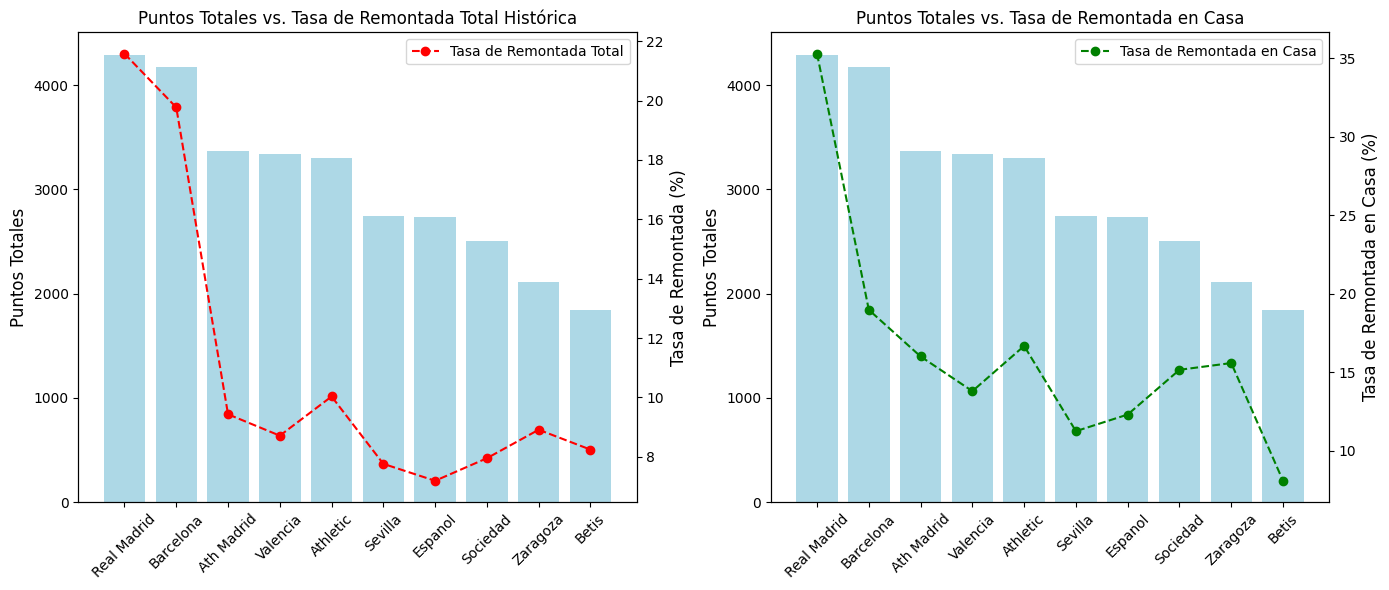


1. Filtrado, operaciones, comparaciones y gráficos

Los datos que disponía son de carácter general, es decir, son datos básicos que no muestran o analizan las dinámicas del juego y otros factores que influyen en él, como las tarjetas amarillas, rojas, faltas, fueras de juego, jugadores, entrenador, plantilla. Al final hay muchos factores que influyen en el desarrollo del juego y para un primer análisis de datos no quería aventurarme a ser demasiado pretencioso y luego no saber manejar todos los datos que tenía. Por eso, iba a basar mi hipótesis en tres ideas generales:

1. **Comparar la puntuación histórica de los 10 mejores equipos de La Liga con su tasa de remontadas.**

*¿Más remontadas suponen históricamente más puntos?* Para responder a esa pregunta que nos va a ayudar a responder a la hipótesis, he decidido comparar la puntuación histórica de los equipos de la liga con más puntos y compararlos con la situación del Athletic, con el fin de saber si hay una relación entre los puntos históricos y la tasa de remontadas. Para ello, descargué un dataset de la esa clasificación y procedí a hacer las comparaciones y gráficos.

La clave aquí es la generación de un bucle-función que me permite, mediante una lista, acceder a esos 10 equipos y generar las tasas de remontadas históricas (en casa y fuera) y la tasa de remontadas en casa. Tras realizar todas las agrupaciones, nos queda este gráfico:

**Se trata de un gráfico combinado (bar + line), con dos ejes Y (doble eje) en cada subgráfico**

1. **Puntos Totales vs. Tasa de Remontada Total (izquierda)**

Aquí vemos que los equipos con más puntos históricos (Real Madrid, Barcelona, Athletic) tienen, en general, una mayor tasa de remontadas. Sin embargo, hay una caída pronunciada en la tasa de remontada a partir del Athletic, lo que indica que **no siempre hay una relación proporcional entre la cantidad de puntos y la capacidad de remontar**. Equipos con menos puntos, como Zaragoza o Valencia, tienen una tasa de remontada relativamente estable pero menor que la de los líderes.

2. **Puntos Totales vs. Tasa de Remontada en Casa (derecha)**

**Aquí se observa un comportamiento más irregular** en la tasa de remontadas en casa.

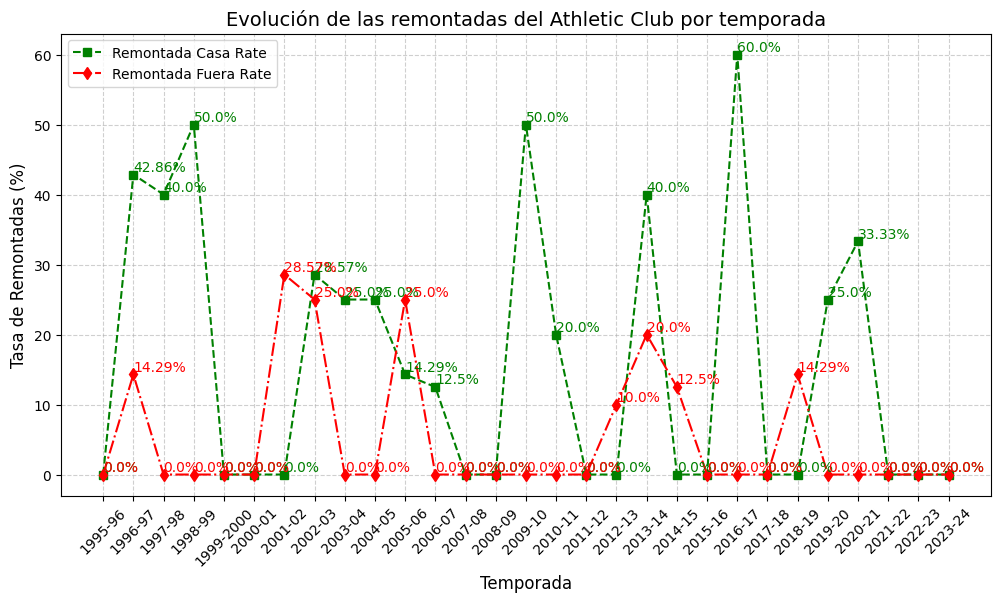
Algunos equipos como Real Sociedad tienen un pico de remontadas en casa a pesar de no estar en la cima de puntos históricos. La tasa de remontada en casa varía más que la total, lo que sugiere que el factor local puede influir en la capacidad de remontar.

3. **¿Confirma el gráfico mi hipótesis?**

El Athletic tiene una tasa de remontada en casa relativamente alta, pero no es la más alta en comparación con otros equipos. Si bien los equipos con más puntos tienden a tener tasas de remontada más altas en general, la variabilidad en la gráfica sugiere que otros equipos también remontan mucho en casa (por ejemplo, la Real Sociedad tiene un pico alto).

Para afirmar la hipótesis con más certeza, habría que comparar la tasa de remontadas del Athletic en casa vs. fuera de casa, en lugar de compararlo con otros equipos.

1. **Comparar las remontadas del Athletic en San Mamés con las que consigue fuera.**

*¿El Athletic remonta más en San Mamés o fuera de casa?* Para ello comparo la tasa de remontadas del Athletic en casa y fuera y también la histórica, para saber si el “factor casa” es determinante a la hora de remontar:

Se trata de un **gráfico de barras dobles** (uno para partidos en casa y otro para fuera), comparando la **tasa de remontadas del Athletic Club** según la condición de local o visitante a lo largo de las temporadas. Es una visualización comparativa que resalta las diferencias de rendimiento según el estadio.

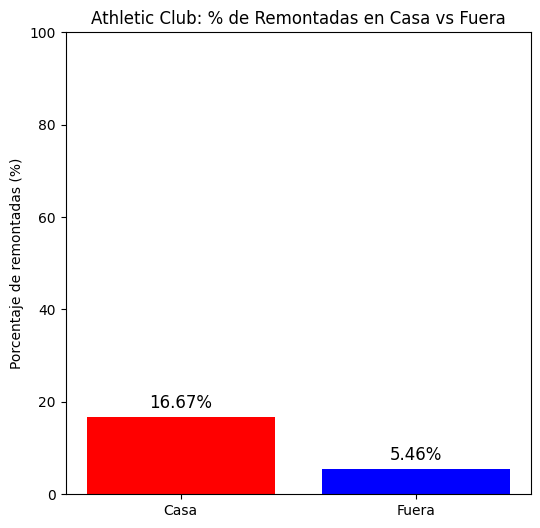
1. **Argumentos a favor de la hipótesis:**

* Se observa que el Athletic tiene **más remontadas en casa que fuera** en la mayoría de las temporadas.
* Esto apoya la idea de que **el factor San Mamés** y el empuje de la afición pueden influir en el equipo.
* Algunas temporadas muestran **cero remontadas fuera**, mientras que sí hay varias en casa.

1. **Argumentos en contra:**

* No todas las temporadas siguen esa regla: hay algunas con **nula diferencia o incluso más remontadas fuera**.
* La muestra es limitada a partidos remontados (no al total jugado), lo que puede generar sesgo.
* No se han tenido en cuenta **factores externos** como el rival, el estado físico, el once titular, etc.

1. **Conclusión breve:**  
   La gráfica da indicios de que el Athletic Club tiene más capacidad de remontada cuando juega en casa, lo que va en línea con la hipótesis. Sin embargo, la variabilidad en algunas temporadas sugiere que no es una regla inquebrantable, y se necesitaría más contexto para afirmarlo con total rotundidad.

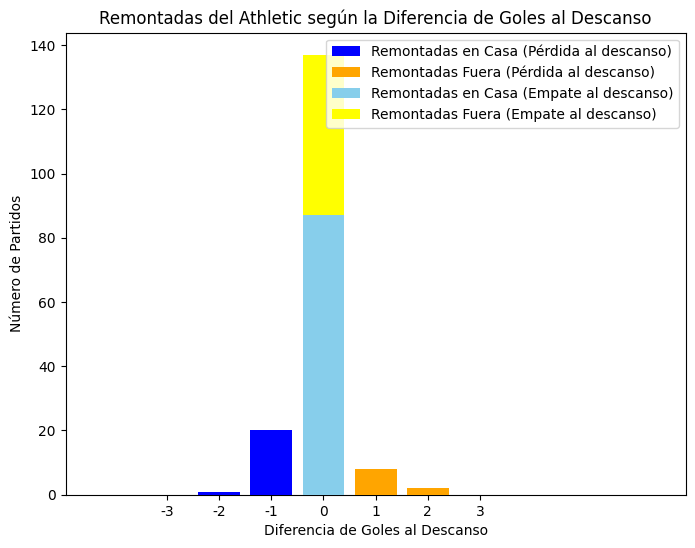


Este último grafico sirve también para hacernos una idea del factor de que el Athletic juegue de local. Cuando va perdiendo en casa al descanso es más probable que remonte, en comparación con hacerlo fuera. De hecho, remonta tres veces mas en San Mamés que fuera.

1. **Comparar las remontadas en función de la diferencia de goles al descanso**

“¿MENOS GOLES EN CONTRA == MÁS OPCIONES DE REMONTAR?”

Saber si la diferencia de goles con la que se va a perdiendo al descanso influye en la capacidad de remontar en casa y fuera



Este es un **gráfico de barras apiladas**, que muestra el número de partidos en los que el Athletic Club ganó el partido tras ir perdiendo o empatando al descanso, diferenciando si fue jugando **en casa o fuera**.

* 1. **Lo que muestra:**
* La mayoría de las **remontadas totales** del Athletic ocurren cuando el partido estaba **empatado al descanso** (barra central, valor 0), especialmente **en casa**.
* Cuando el equipo va **perdiendo por 1 gol al descanso**, también logra remontar a veces, pero en menor medida.
* Las **remontadas tras ir perdiendo por más de 1 gol** son extremadamente raras (casi nulas).
* Prácticamente **no hay remontadas fuera de casa** cuando el equipo va perdiendo por más de un gol.
  1. **Argumentos a favor de la hipótesis**
* Claramente, el Athletic tiene **más capacidad de remontada cuando juega en San Mamés**, lo que **respalda la hipótesis del factor campo**.
* El equipo es más **capaz de remontar empates o desventajas mínimas** en casa.
* Se aprecia un **mayor número de partidos ganados tras empates al descanso en casa**, lo que puede indicar que **aprovechan la segunda parte** para imponerse.
  1. **Argumentos en contra**
* El número de **remontadas desde desventajas reales (más de un gol)** es muy bajo, incluso en casa, lo que **limita la fuerza de la hipótesis** si se quiere afirmar que remontan “a menudo”.
* Las **remontadas fuera de casa** son pocas, incluso cuando el partido está igualado, lo que **reduce la consistencia** del patrón.
* No se tienen en cuenta **otros factores** como el rival, expulsiones, dinámica del partido, etc., que podrían influir en el resultado.

1. Conclusiones

En cuanto a mi hipótesis:

En base al análisis realizado, podemos decir que la hipótesis se **cumple parcialmente**:

* Se demuestra que **el Athletic Club tiene una mayor capacidad de remontada en casa** que fuera, especialmente en situaciones de empate al descanso o desventajas mínimas.
* Sin embargo, **no se puede afirmar que sea un equipo que remonta con frecuencia cuando va claramente perdiendo**, ya que los datos muestran que solo en casos concretos ha sido capaz de hacerlo.
* Por tanto, la hipótesis se **respalda parcialmente**, pero **requiere matices importantes**. No se puede generalizar que el Athletic sea un equipo netamente remontador, sino que **bajo ciertas condiciones específicas (jugar en casa, empates o 1 gol abajo)**, tiene más posibilidades de acabar ganando.

En cuanto al trabajo en general:

Este análisis exploratorio es solo **la punta del iceberg**. En el fútbol intervienen miles de variables: desde los jugadores disponibles, el entrenador, las decisiones arbitrales, hasta aspectos intangibles como la confianza, la moral o el estilo de juego. Este EDA es un **primer acercamiento general** que ayuda a **formarse una idea amplia** sobre ciertos patrones del Athletic Club, pero no pretende ser concluyente.

Para poder sacar afirmaciones más sólidas haría falta incorporar datos más complejos como **posesión, tiros a puerta, tarjetas, alineaciones, cambios tácticos, entre otros**. Aun así, este trabajo sirve como base para comprender cómo se comporta un equipo a lo largo del tiempo, especialmente en su estadio, y cómo se podrían seguir desarrollando análisis más ambiciosos en el futuro.