

TEMPORADA 2024/25



*ANÁLISIS EXPLORATORIO
DE DATOS DE EQUIPOS DE
FÚTBOL DE LA LIGA
ESPAÑOLA*



TABLA DE CONTENIDO

Introducción.....	
1. Hipótesis	
2. Análisis Univariante	
3. Análisis Bivariante.....	
4. Análisis Multivariante.....	
5. Visualizaciones Relevantes	
6. Conclusiones	
7. Recomendaciones	
8. Referencias	

INTRODUCCIÓN

El objetivo de este proyecto es realizar un análisis exploratorio de los datos de la Liga Española de Fútbol (LaLiga), con el fin de descubrir patrones y obtener insights relevantes sobre el rendimiento de los equipos a lo largo de la temporada 2024. Este análisis busca identificar variables clave que puedan influir en el desempeño de los equipos, como el número de goles marcados, los tiros a puerta, las faltas cometidas, las tarjetas, la posesión del balón, entre otros.

- Los datos utilizados en este proyecto provienen de una base pública de resultados y estadísticas detalladas de los partidos de la temporada 2024 de LaLiga. Estos datos incluyen información tanto de los partidos jugados en casa como fuera de casa por los equipos, lo que permite evaluar la influencia de diferentes factores en el rendimiento. La base de datos contiene registros de cada partido, con información sobre los goles, el resultado final (victoria, empate, derrota), y diversas métricas de desempeño de los equipos.
- El análisis se centrará en identificar patrones en el rendimiento de los equipos locales y visitantes, explorar las diferencias en el rendimiento de los equipos en casa frente a fuera de casa, y cómo los diferentes indicadores (como faltas, tarjetas y tiros a puerta) se correlacionan con el resultado de los partidos. También se explorará cómo ciertos equipos, como los más exitosos de la temporada, se comparan con aquellos que tienen un desempeño más modesto.

La base de datos utilizada contiene más de 300 registros, cada uno representando un partido jugado durante la temporada 2024. Cada registro incluye información detallada sobre el partido y las métricas de desempeño de ambos equipos (local y visitante). Entre las principales variables que se incluyen en el conjunto de datos se encuentran:

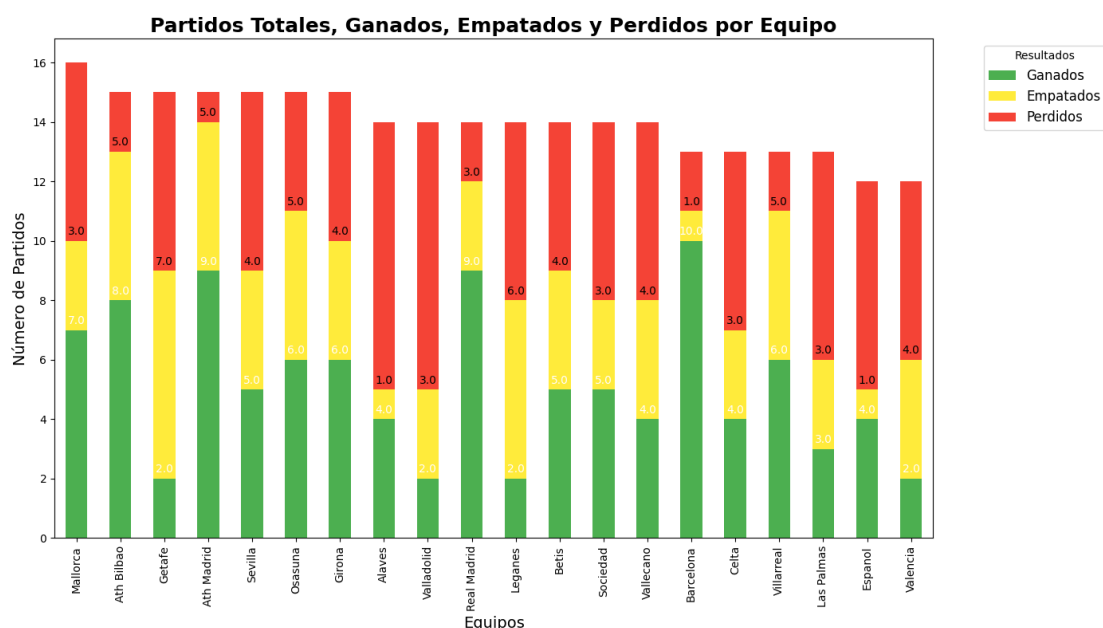
- **División:** La división o liga a la que pertenece el partido (por ejemplo, LaLiga).
- **Fecha y Hora:** La fecha y hora en la que se jugó el partido.
- **Equipos:** El equipo local y el equipo visitante (nombre de los equipos).
- **Resultado final:** Goles marcados por el equipo local (FTHG) y el equipo visitante (FTAG), y el resultado final del partido (FTR), que puede ser victoria para el equipo local (H), victoria para el visitante (A) o empate (D).
- **Goles en el primer tiempo:** Goles marcados en el primer tiempo para ambos equipos (HTHG para el local, HTAG para el visitante).
- **Estadísticas del partido:** Variables como tiros a puerta (HTS para el local, ATS para el visitante), faltas cometidas (TotalFoulsHome para el local, TotalFoulsAway para el visitante), tarjetas amarillas y rojas, entre otras.
- **Indicadores adicionales:** Variables adicionales como la posesión del balón (HomePossession, AwayPossession) y estadísticas de apuestas (como las probabilidades de apuestas de casas como Bet365 o Pinnacle).

A través del análisis de estas variables, se espera obtener una visión más clara de los factores que contribuyen al éxito de los equipos, y cómo esos factores pueden predecir el resultado de futuros partidos.

1. HIPÓTESIS

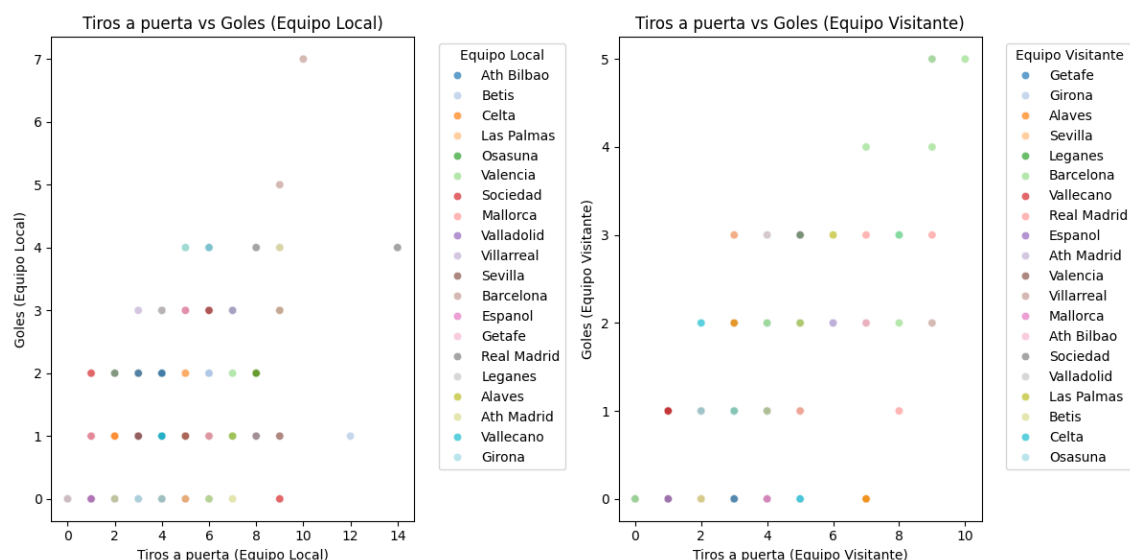
Hipótesis 1: Ventaja del equipo local

Los equipos locales tienen una mayor probabilidad de ganar debido a la ventaja de jugar en casa. Esto se puede explorar analizando los resultados de los partidos (ganados, empatados, perdidos) y comparando la frecuencia de victorias de los equipos locales frente a los visitantes.



Hipótesis 2: Relación entre tiros a puerta y goles marcados

Existe una correlación positiva entre la cantidad de tiros a puerta y los goles marcados. Esta hipótesis puede ser probada mediante un análisis bivalente, donde se investigue si los equipos que realizan más tiros a puerta tienden a marcar más goles.



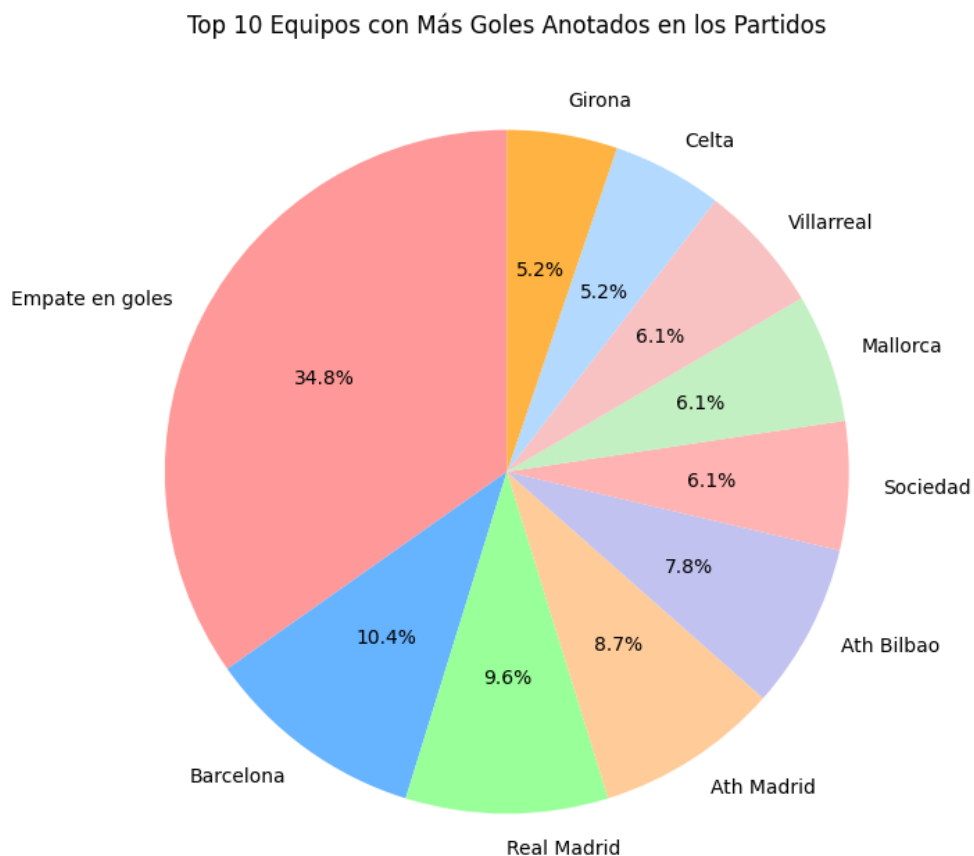
Hipótesis 3: Faltas y tarjetas amarillas

Los equipos que cometen más faltas tienden a recibir más tarjetas amarillas. Esto implica que el comportamiento agresivo de un equipo, medido por el número de faltas cometidas, puede estar relacionado con una mayor probabilidad de recibir tarjetas amarillas durante un partido.



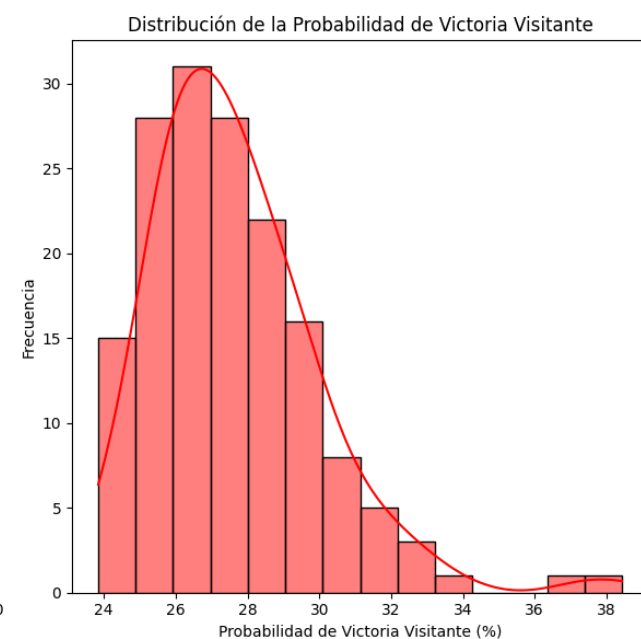
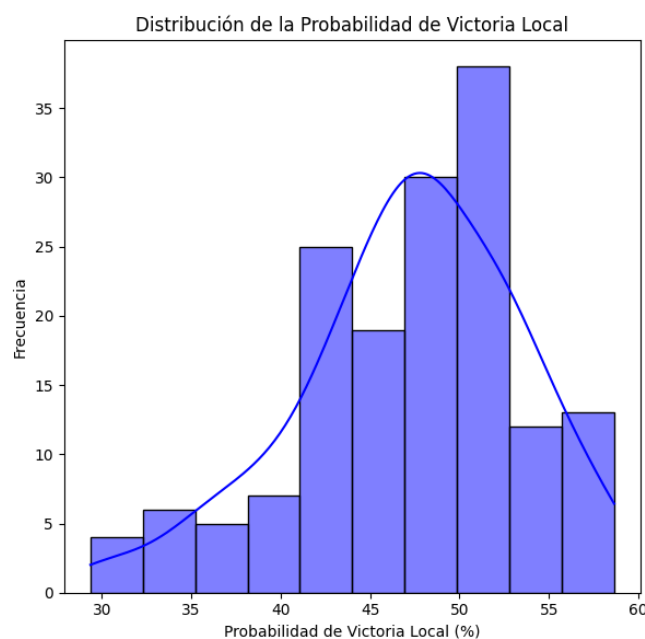
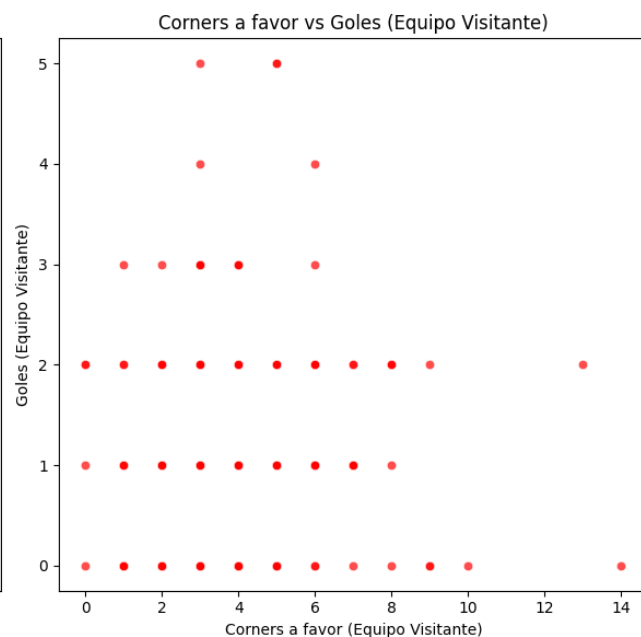
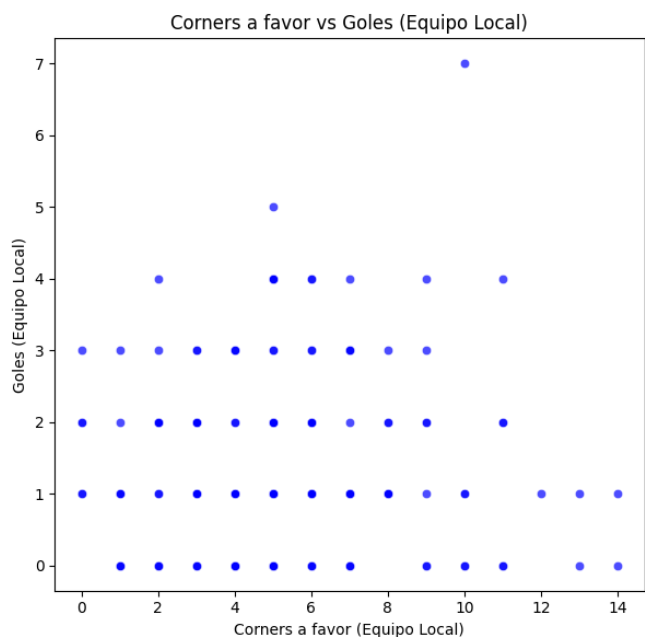
Hipótesis 4: Rendimiento de equipos con más goles a favor

Los equipos con un mayor promedio de goles marcados por partido tienen más probabilidades de terminar la temporada en una posición alta de la tabla. Esto puede analizarse observando la relación entre los goles marcados y la clasificación final de los equipos.



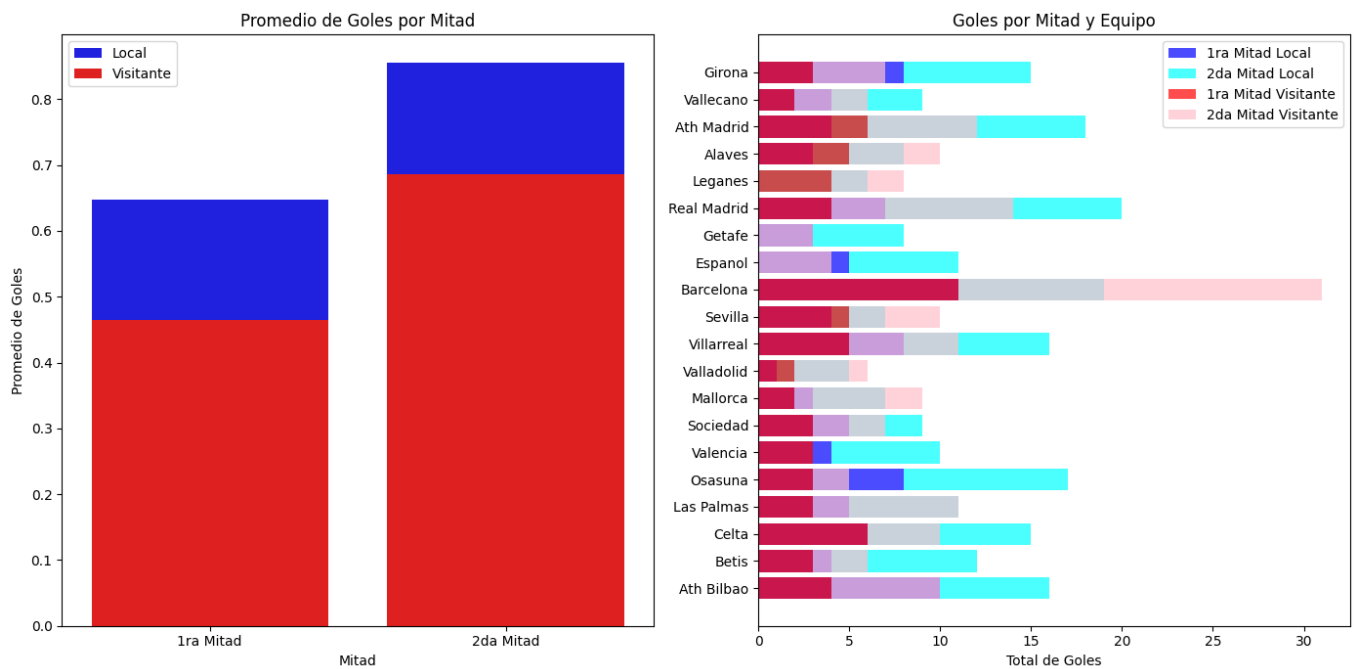
Hipótesis 5: Impacto de los corners en el resultado

Un mayor número de corners a favor de un equipo está relacionado con un mayor rendimiento ofensivo y, por lo tanto, con una mayor probabilidad de ganar el partido. Se puede investigar si los equipos que tienen más corners tienden a ganar más partidos o a marcar más goles.



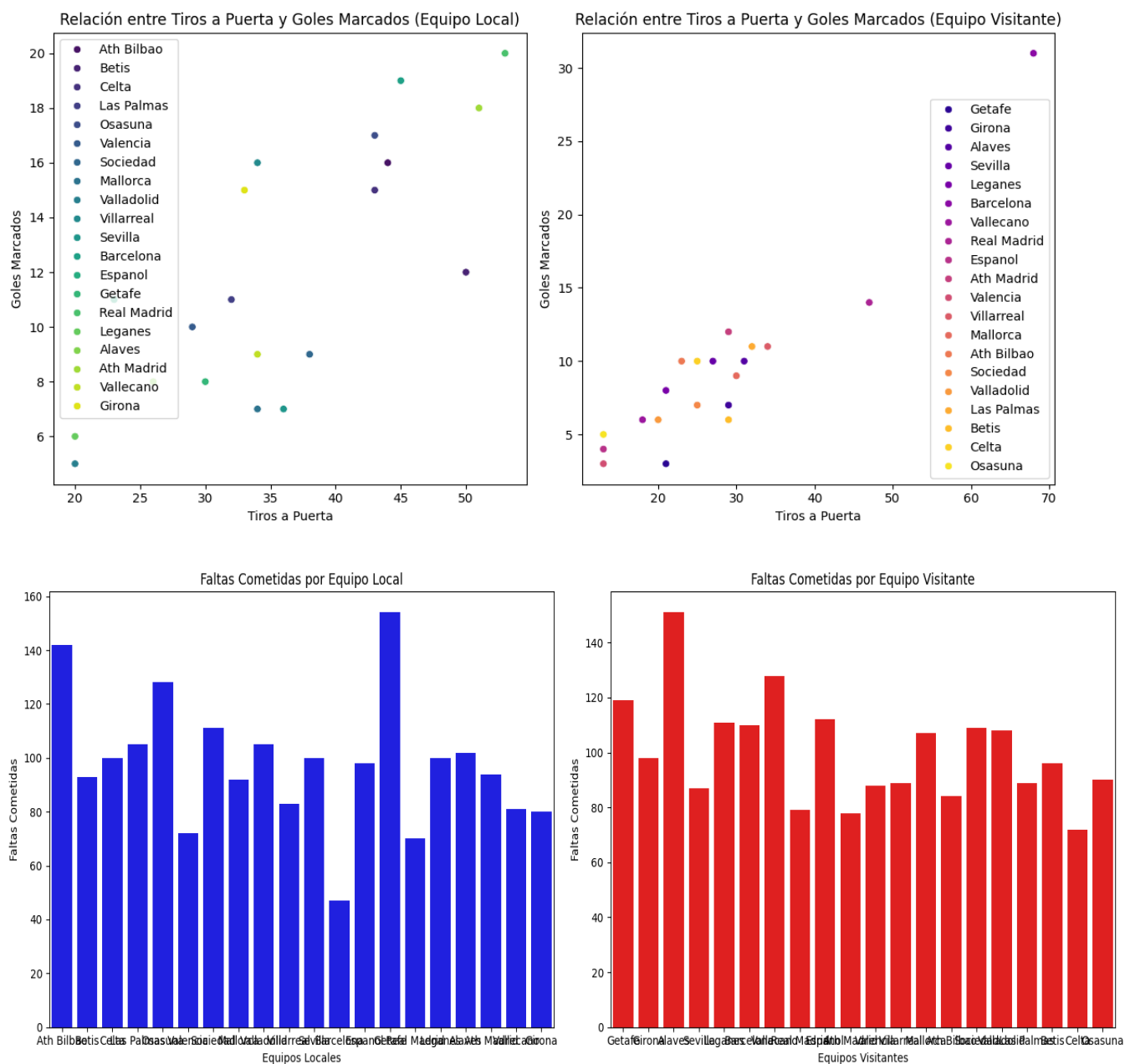
Hipótesis 6: Tendencia en el rendimiento por tiempo de juego

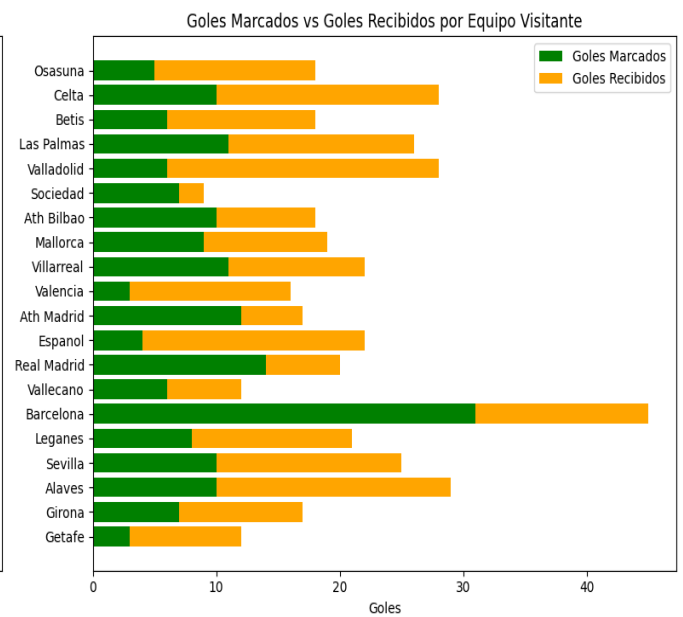
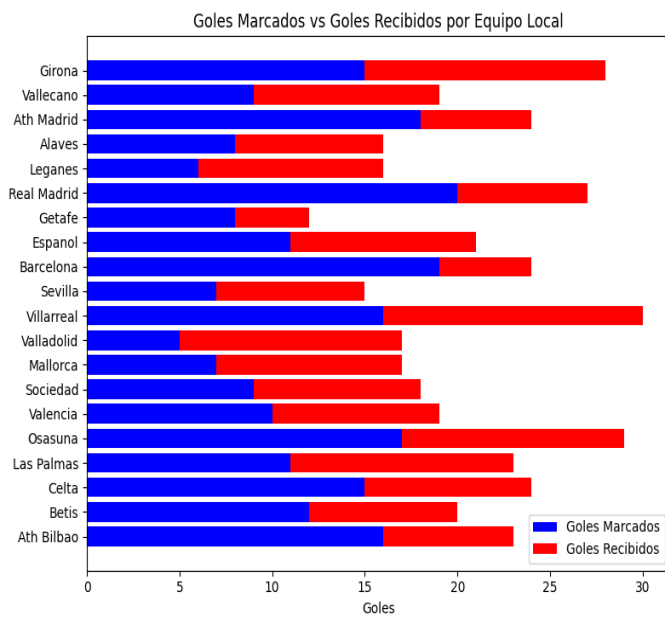
Los equipos que anotan más goles en la segunda mitad de los partidos tienen una mayor probabilidad de ganar debido a una mejor adaptación táctica o física en la segunda mitad. Esta hipótesis puede probarse al comparar los goles anotados en la primera y segunda mitad de los partidos.



Hipótesis 7: Estrategias de ataque y defensiva

Existen diferencias en el rendimiento defensivo y ofensivo de los equipos en función de su estrategia de juego, que puede medirse a través de estadísticas como los tiros a puerta, los goles marcados, las faltas cometidas, y los goles recibidos.



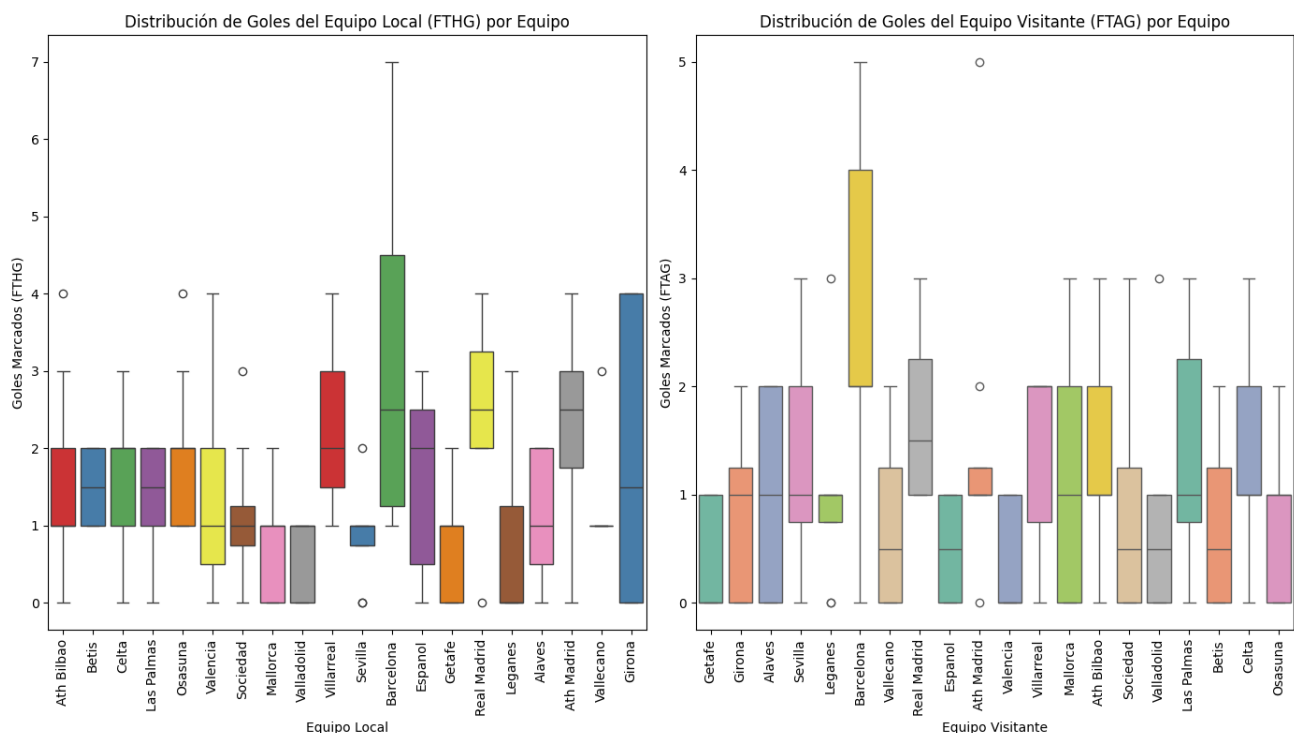


2. ANÁLISIS UNIVARIANTE

El análisis univariante realizado proporciona una visión detallada de las distribuciones de goles y probabilidades de victoria para los equipos, permitiéndonos identificar patrones de rendimiento y tomar decisiones informadas sobre las estrategias de los equipos.

Boxplots: Para cada una de estas variables, se generaron **boxplots** (diagrama de cajas) para visualizar la distribución de los goles, con énfasis en:

- **Mediana:** Representada por la línea dentro de cada caja, muestra el valor central de los goles marcados por los equipos.
- **Cuartiles:** Las cajas representan los cuartiles, que dividen la distribución en cuatro partes. Esto nos da una idea de la variabilidad de los goles de los equipos.
- **Outliers:** Los puntos fuera de las "cajas" indican valores atípicos, es decir, equipos que tuvieron un rendimiento excepcionalmente alto o bajo en términos de goles marcados.



FTHG: Goles del equipo local en tiempo completo

FTAG: Goles del equipo visitante en tiempo completo

B365CAHH: Probabilidad de victoria del equipo local (según Bet365) para el equipo local

B365CAHA: Probabilidad de victoria del equipo visitante (según Bet365) para el equipo visitante.

PCAHH: Probabilidad de victoria del equipo local (según Pinnacle) para el equipo local.

PCAHA: Probabilidad de victoria del equipo visitante (según Pinnacle) para el equipo visitante.

MaxCAHH: Probabilidad máxima de victoria del equipo local (según varias casas de apuestas) para el equipo local.

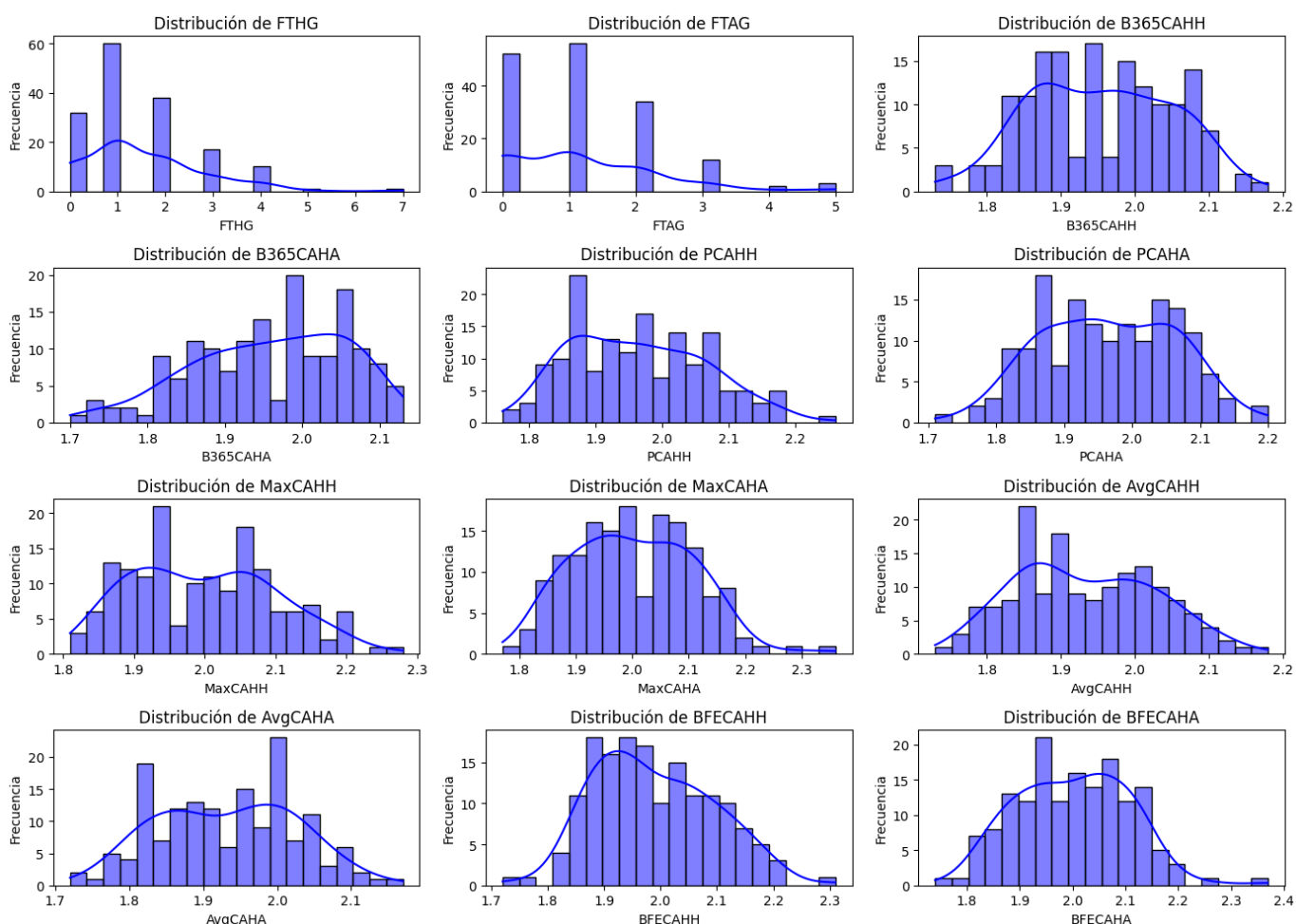
MaxCAHA: Probabilidad máxima de victoria del equipo visitante (según varias casas de apuestas) para el equipo visitante.

AvgCAHH: Probabilidad promedio de victoria del equipo local (según varias casas de apuestas) para el equipo local.

AvgCAHA: Probabilidad promedio de victoria del equipo visitante (según varias casas de apuestas) para el equipo visitante.

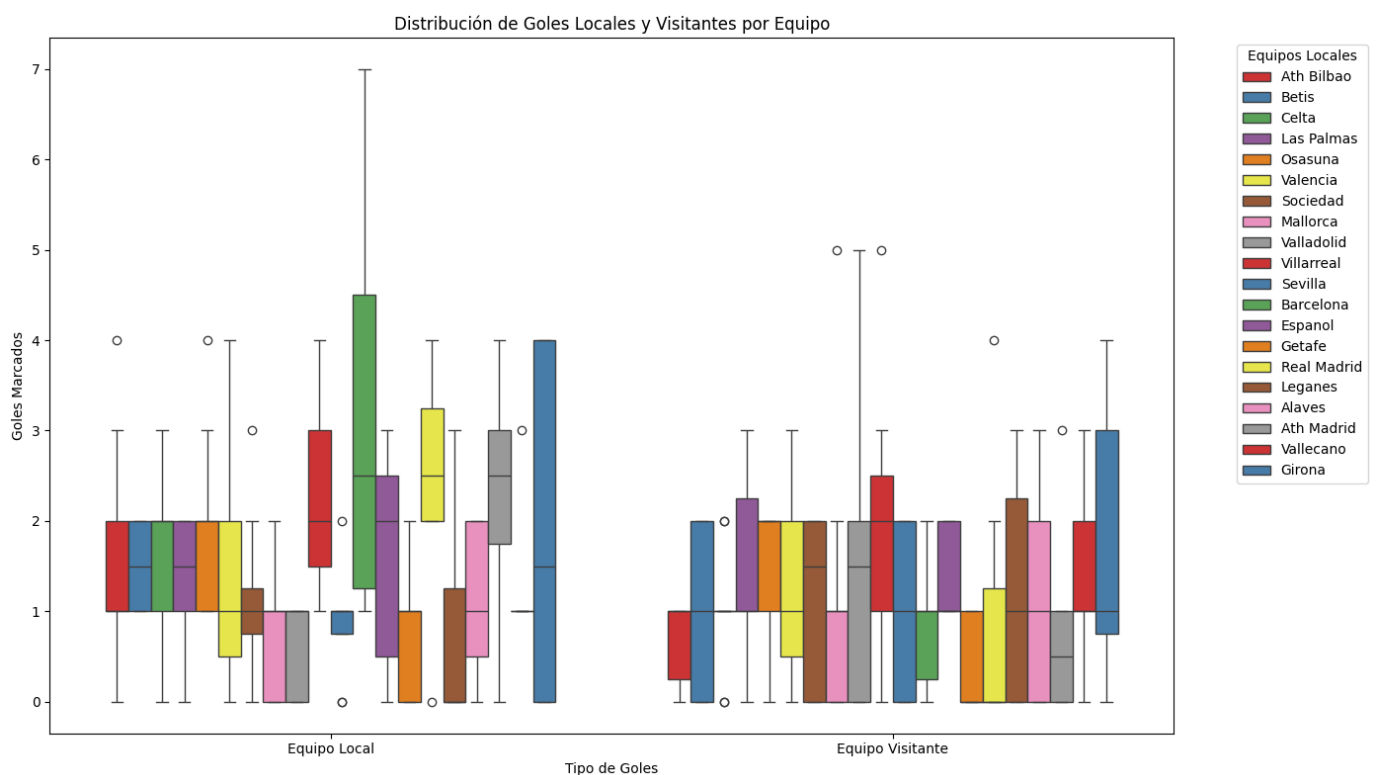
BFECAHH: Probabilidad de victoria del equipo local según Betfair (para el equipo local).

BFECAHA: Probabilidad de victoria del equipo visitante según Betfair (para el equipo visitante).



3. ANÁLISIS BIVARIANTE

En este análisis bivalente, hemos explorado la relación entre los **goles marcados por el equipo local** (FTHG) y los **goles marcados por el equipo visitante** (FTAG) a lo largo de diferentes partidos, utilizando un **boxplot**. El objetivo de este análisis es observar cómo se distribuyen los goles de ambos equipos (local y visitante) y cómo estas distribuciones varían según los equipos que participan en los partidos.

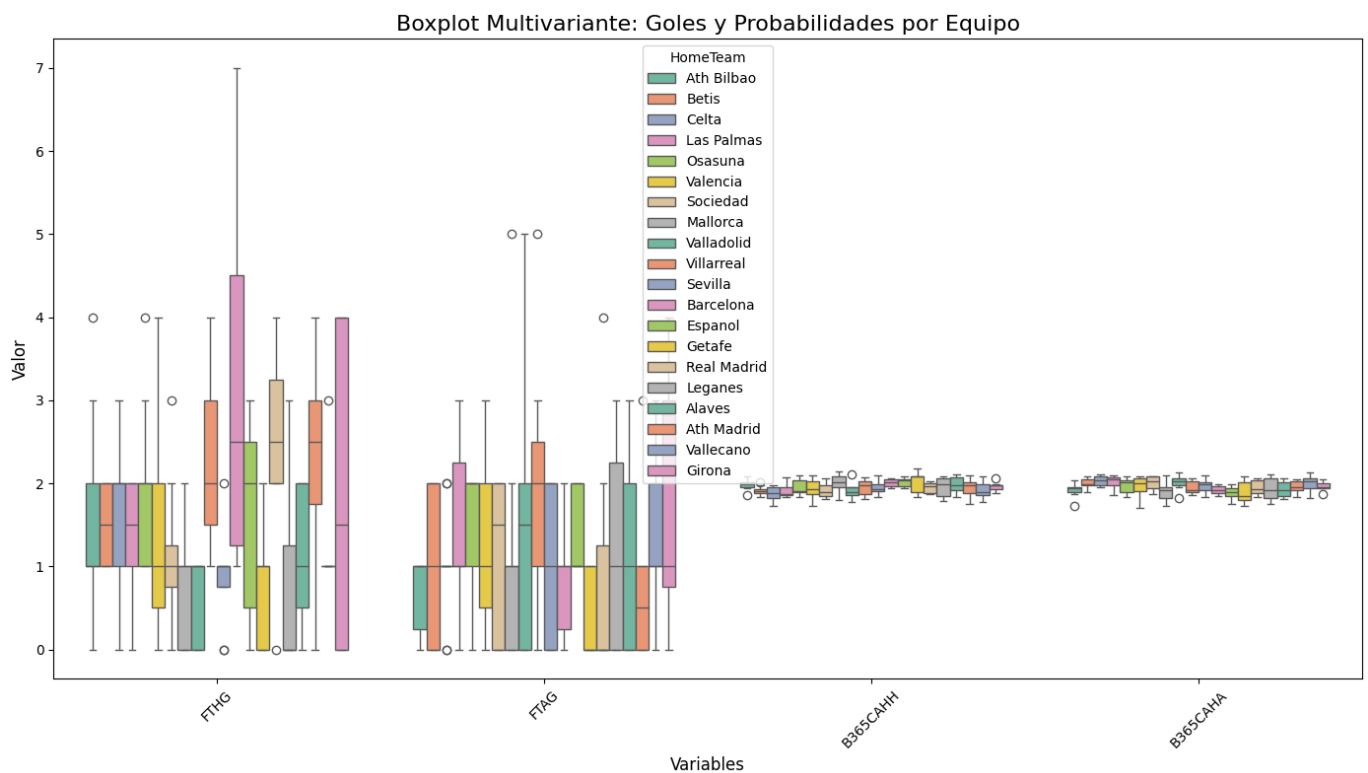


El análisis bivalente realizado con el boxplot proporciona una visión clara sobre la variabilidad de los goles marcados por los equipos locales y visitantes. Además, muestra la dispersión de estos goles para cada equipo, lo que puede ser útil para identificar patrones de rendimiento en casa y fuera de casa. Este tipo de análisis puede ser útil para comprender cómo la ubicación de un equipo (local o visitante) afecta a su rendimiento ofensivo en términos de goles marcados.

4. ANÁLISIS MULTIVARIANTE

Este análisis multivariante proporciona una visión profunda de cómo las variables de goles y probabilidades se distribuyen entre los diferentes equipos. Los boxplots son útiles para visualizar la dispersión de los datos, ayudando a identificar equipos con un rendimiento más homogéneo frente a aquellos con mayor variabilidad. Además, las probabilidades de victoria ofrecieron una perspectiva adicional sobre cómo los equipos son percibidos por las casas de apuestas, lo que puede ser útil para predecir el desempeño de los equipos en futuros partidos.

Este análisis permite avanzar en la comprensión de los factores que influyen en los resultados de los partidos, y puede ser un primer paso para desarrollar modelos predictivos más complejos basados en estas y otras variables.



5. VISUALIZACIONES RELEVANTES

Se generaron varias visualizaciones para comprender mejor los datos:

1. **Boxplot de goles locales por equipo:** Muestra cómo se distribuyen los goles anotados por los equipos locales, ayudando a identificar patrones en su rendimiento.
2. **Boxplot de goles visitantes por equipo:** Similar al anterior, pero para los equipos visitantes.
3. **Boxplot de probabilidades de victoria local y visitante:** Muestra cómo las probabilidades de que un equipo gane varían entre las casas de apuestas.
4. **Comparación entre goles locales y goles visitantes:** Compara los goles anotados por los equipos locales y visitantes en los partidos.
5. **Histograma de resultados de partidos:** Muestra la frecuencia de victorias, empates y derrotas.
6. **Correlación entre goles y probabilidades de victoria:** Analiza si existe una relación entre los goles marcados y las probabilidades de victoria

6. CONCLUSIONES

En este análisis, se observó que los equipos locales suelen marcar más goles, lo cual puede explicarse por factores como el apoyo de los aficionados y la familiaridad con el estadio. Además, encontramos que los tiros a puerta están fuertemente relacionados con los goles, lo que significa que los equipos que crean más oportunidades de gol tienen más probabilidades de ganar.

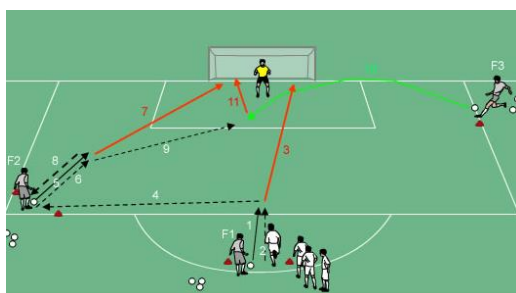
Sin embargo, algunos equipos, como "Valladolid", no aprovechan bien sus oportunidades y, a pesar de realizar muchos tiros a puerta, no logran marcar tantos goles. Esto sugiere que hay otros factores, como la calidad de los tiros o la defensa rival, que también influyen en el rendimiento.

Por último, las probabilidades de victoria dadas por las casas de apuestas a veces no coinciden con los resultados reales, ya que algunos equipos superan las expectativas. En resumen, aunque las estadísticas ayudan a predecir los resultados, siempre hay elementos impredecibles en el fútbol que pueden influir en el desempeño de los equipos. Cabe destacar, la BBDD descargada debe ir actualizándose en base a los nuevos partidos jugados para un mejor pronóstico, en este EDA obtuvimos los datos a través de Football-Data.co.uk el miércoles 11/12/2024.

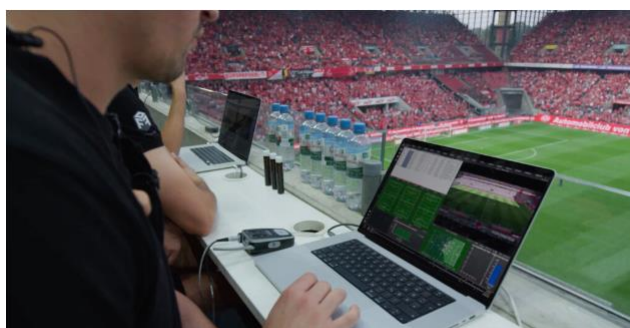
7.RECOMENDACIONES

Basado en los hallazgos, se sugieren algunas recomendaciones:

- **Mejorar la precisión de los tiros:** Dado que los tiros a puerta se correlacionan con los goles, los equipos deben trabajar en la técnica de finalización.



- **Análisis de equipos atípicos:** Equipos como el "Valladolid" deben revisar su efectividad al convertir tiros en goles.
- **Optimización de la defensa:** Analizar equipos con un buen rendimiento defensivo para mejorar la capacidad de evitar goles.



LOS ANALISTAS DE RENDIMIENTO DEL FC KOLN UTILIZAN MATCHTRACKER PARA RECOPIRAR Y ANALIZAR DATOS DURANTE UN PARTIDO EN VIVO, GARANTIZANDO UNA PLANIFICACIÓN ÓPTIMA DEL RENDIMIENTO.

- **Estudio del impacto de las faltas y tarjetas:** Es recomendable investigar cómo las sanciones afectan el rendimiento de los equipos.
- **Aprovechar los goles en la segunda mitad:** Algunos equipos muestran mejor rendimiento ofensivo en la segunda mitad. Ajustar tácticas durante el partido puede ser crucial.
- **Tomar decisiones basadas en datos:** Los equipos deberían usar análisis de datos más avanzados para mejorar sus tácticas

8.REFERENCIAS

- Este análisis se ha realizado con herramientas como **Python**, **Pandas**, **Matplotlib**, **Seaborn**, **Numpy**, **statsmodels**, **scikitlearn**, utilizando base de datos de **Football-Data.co.uk**, una fuente confiable para estadísticas de fútbol. Además, se apoyó en metodologías de análisis exploratorio de datos y análisis estadístico.



