Jennifer Fajardo

Post-Pandemic Play – Evaluating Home Advantage in European Soccer After COVID-19

Checkpoint 5

Question Snapshot

This project answers the question of whether the traditional home-field advantage in professional European soccer declined during and after the COVID-19 pandemic, particularly as stadium attendance restrictions disrupted the typical match environment. "Success" for this project means being able to quantify the shift clearly enough to say home-field advantage weakened, stayed the same, or rebounded. This insight matters for coaches, league officials, analysts, and bettors as identifying changes in home-field advantage can influence strategy, scheduling, and forecasting models.

Data Used

For the Exploratory Data Analysis, I used the cleaned file from my last checkpoint, "**Clean Data PL**" and "**Clean Data LL**". This dataset includes home and away results from the Premier League and La Liga across three seasons:

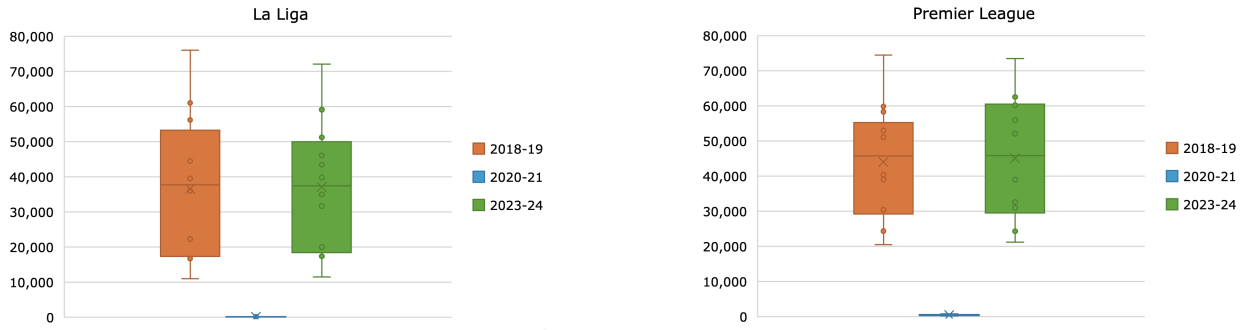
(1) Pre COVID: 2018-2019
(2) COVID: 2020-2021
(3) Post COVID: 2023-2024

Each row represents a single team in the specific league, capturing home and away performance for that season.

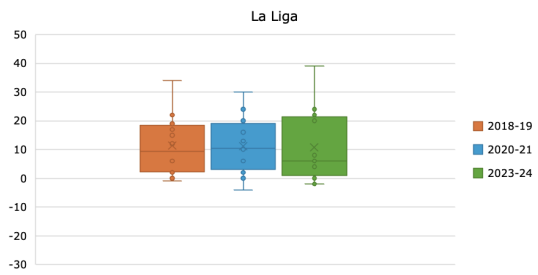
A major filter applied in this exploratory analysis was including only the teams that competed in all three seasons, ensuring that the comparisons across periods were consistent and not affected by promotions or relegations.

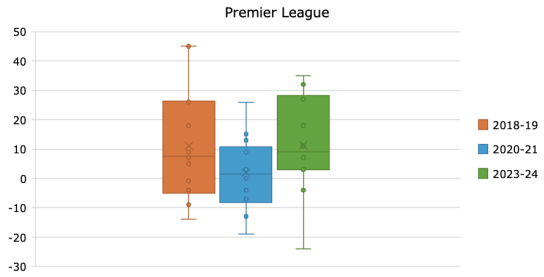
Univariate Distributions

Average Home Game Attendance: I only have attendance data for 3/12 teams in La Liga, but the boxplot looks similar to the Premier League, with very low numbers for the 2020-21 season.

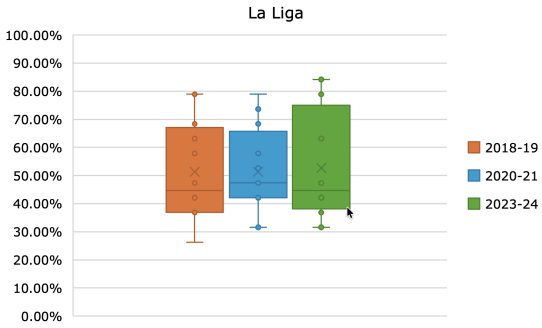


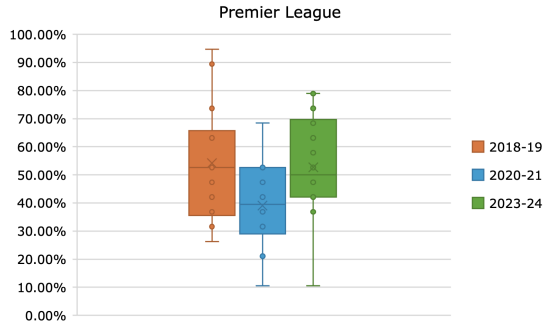
Home Goal Differential: La Liga's distribution is what I was expecting more compared to the Premier League.





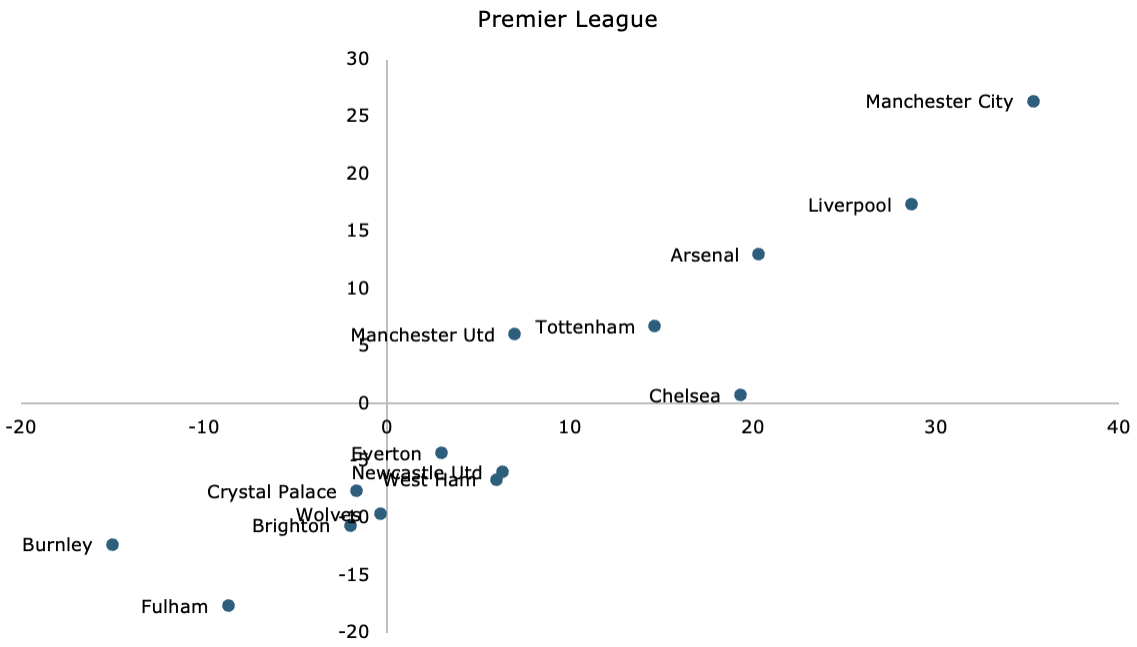
Home Win Percentage: Opposite of home goal differential, the Premier League's distribution is what I was expecting more compared to La Liga.





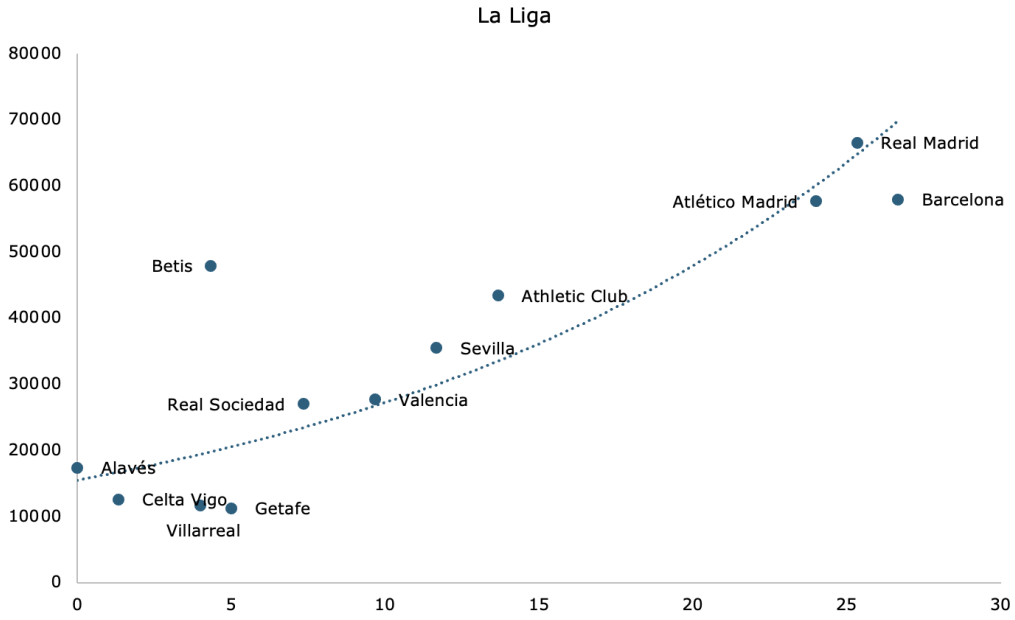
Relationships (Pairs)

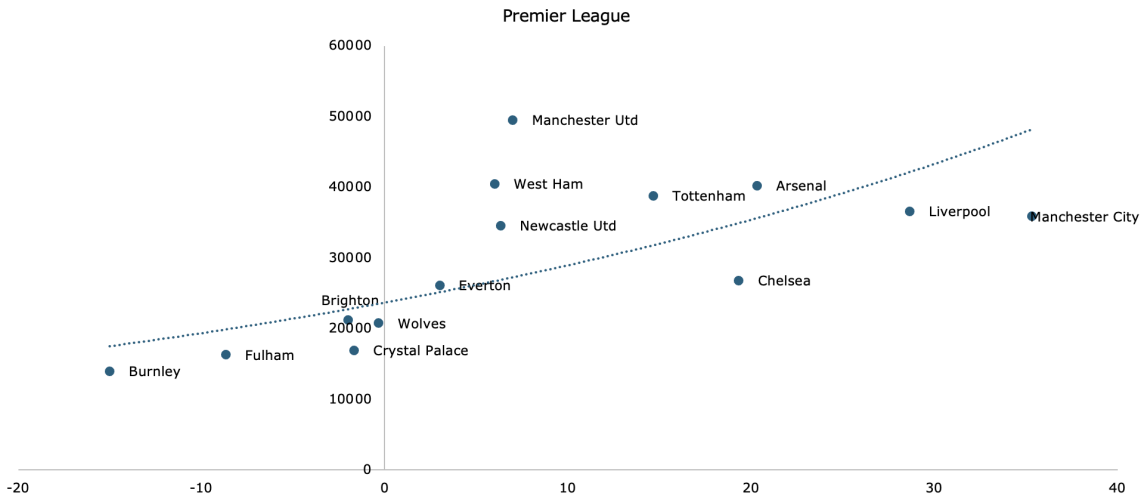
Home GD vs. Away GD **average across the three seasons: 2018-19, 2020-21, 2023-24*



Plotting home goal differential against away goal differential shows a positive relationship, as teams that perform well at home also tend to perform well on the road. The strongest clubs in each league cluster in the upper right area of the plot, reflecting consistent performance in both environments. Lower table teams often have negative away goal differentials and only moderate home goal differentials, suggesting that weaker squads rely more heavily on home advantage. This pattern aligns with common sports intuition, where strong teams perform well regardless of location while weaker teams experience the largest drop in performance away from home.

<u>Attendance vs. Home GD</u> **average across the three seasons: 2018-19, 2020-21, 2023-24*





A scatterplot of attendance vs. home GD shows a clear positive relationship in both La Liga and the Premier League. Teams with larger home crowds generally have stronger home goal differentials. The upward sloping trendlines align with typical sports intuition, since larger stadiums often correspond to stronger clubs with more resources, deeper squads, and more intimidating home environments. In both leagues, the pattern begins to flatten slightly among the very highest attendance teams (Manchester United, Real Madrid, Barcelona) indicating mild diminishing returns once clubs reach elite performance levels. The relationship is directionally strong but not perfectly linear, suggesting that attendance contributes to home performance but does not solely determine it.

Subgroup Comparisons

LA LIGA **teams that played in all 3 seasons (12/20)*

	Average Home GD	Average Away GD
2018-19	11.33	-2.58
2020-21	11.17	1.17
2023-24	10.75	-0.92

PREMIER LEAGUE **teams that played in all 3 seasons (14/20)*

	Average Home GD	Average Away GD
2018-19	11.00	-1.93
2020-21	1.93	3.07
2023-24	11.29	-2.21

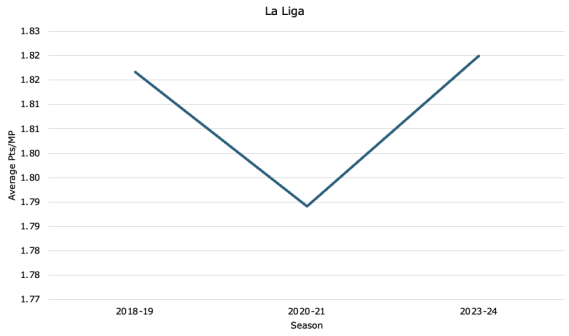
The average home goal differential is much higher than the average away GD in 2018-19 and 2023-24 for both leagues, showing a clear and meaningful home-field advantage where teams score more and concede less. The exception in both leagues is 2020… away goal differential increases sharply and even surpasses home performance in the Premier League, suggesting a temporary reduction in home advantage. This pattern is unlikely to be noise, as the shift appears in both leagues and aligns with known disruptions during the pandemic season.

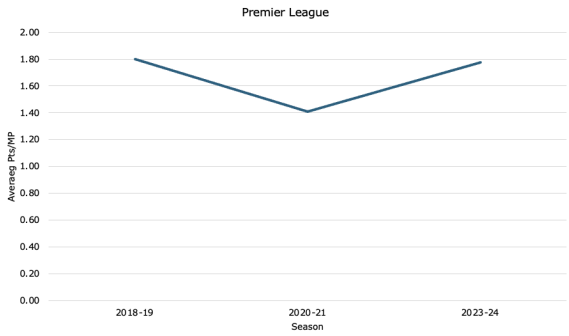
Overall, home performance tends to be stronger with 2020-21 standing out as a structural outlier rather than random variation.

Time View

Average Pts/MP each season

Both leagues show a clear drop in average home points per match during the 2020-21 season, followed by a recovery in 2023-24. This suggests a temporary COVID effect rather than streakiness or back-to-back schedule effects.





Outliers & Gaps

The most notable outlier in the data is the 2020-21 season, where home performance metrics decline sharply across both leagues. This pattern reflects a legitimate disruption caused by COVID stadium closures and limited capacity rather than data errors. I will retain these observations and treat them as a distinct COVID period instead of capping or removing values to assess how reduced fan presence impacted home-field advantage.

Average home attendance data is largely missing for La Liga during the 2020-21 season. After looking through both La Liga and FBref sites, this appears to be a structural issue rather than a data error as stadiums were closed for most of the season. Because this missingness affects nearly all teams in that year, it limits meaningful attendance based comparisons. **The remainder of the analysis will focus on the Premier League where attendance data is complete and consistent.**

Early Takeaways / Next Steps

Early Takeaways:

- Home advantage is clearly present in 2018-19 and 2023-24, with stronger home goal differentials and points per match.
- The 2020-21 season stands out as a meaningful break, with home advantage weakening or reversing during COVID.
- Higher attendance is positively associated with stronger home performance, though with diminishing returns at elite clubs.

Next Steps:

- For Checkpoint 6, restrict the dataset to the 14 Premier League teams that played all three seasons to ensure clean, consistent comparisons.

- For Checkpoint 7, expand the analysis to all 20 Premier League teams to evaluate how promotion and relegation affect home-field advantage dynamics.