Attendance in Argentinian Football: Team and Weekday Effects in a Bayesian Model

Pedro Paiva

November 22th

Introduction

This project investigates how fan attendance varies across teams and days of the week in a professional sports league. We work with real ticket scan data from twelve teams, each with multiple home games. Some values are missing due to scanner failures. Our goals are to (1) identify how team popularity and game day affect attendance, and (2) predict attendance for games with missing data.

First, we fit a complete pooling model that assumes a shared attendance distribution across all games. We then build a hierarchical model that allows for variation by team and by day.

Preprocessing

We loaded the file 'sports-attendance-data.csv' and checked the main columns team, day, and attendance. We treated attendance as a count outcome, so we kept it in its original scale. No log or standardization was applied to the observed values.

Days of the week were ordered from Monday to Sunday and stored as an ordered categorical variable. This step makes the day index consistent with the calendar order and helps with later plotting and interpretation. Teams were also stored as a categorical variable. From these categories we created integer 'indices team_idx' and 'day_idx', starting at 0. These indices are what the Bayesian models use, while the category labels are kept so we can map results back to team names and day names.

We separated the dataset into two parts. The observed set contains all rows with attendance recorded and the missing set contains rows where attendance is NaN. In total, about twenty games are missing attendance, which matches the data note that scanner failures caused gaps. The missingness plot by game sequence shows that gaps are concentrated in a few teams, with no clear pattern across the season (Figure 1). This supports treating missingness as a data issue rather than a structural change in demand.

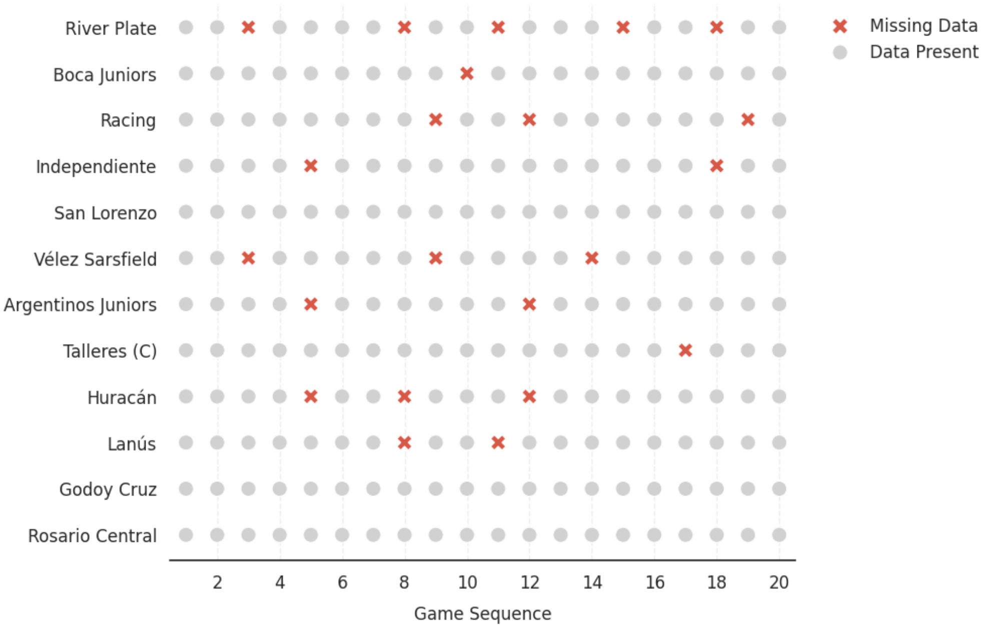


Figure 1: Missing attendance pattern by team and game order. Gray dots mark games with recorded attendance. Red crosses mark games where the scanner failed and attendance is missing.

We then explored attendance distributions. Histograms by team show clear differences in central tendency and spread (Figure 2). Large clubs such as River Plate, Boca Juniors, and Racing have higher typical attendance, while smaller clubs tend to cluster lower. Histograms by day show mild day effects, with weekend games tending to have higher medians and wider right tails (Figure 3). Across all groups, the distributions are right skewed and include some high attendance games.

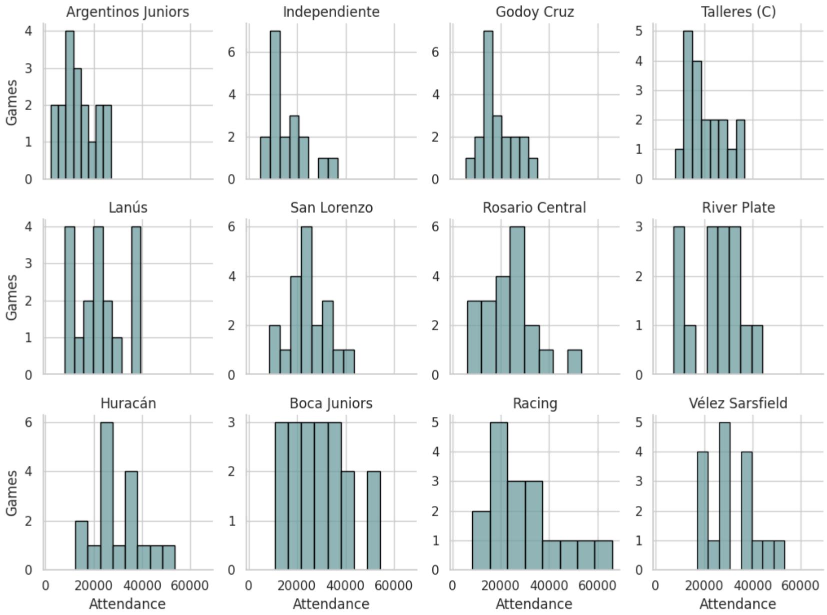


Figure 2: Observed attendance distributions by team. Each panel shows how draws differ across clubs. Large teams have higher typical attendance and wider right tails, while smaller teams cluster at lower values. The right skew in most panels suggests occasional very large crowds.

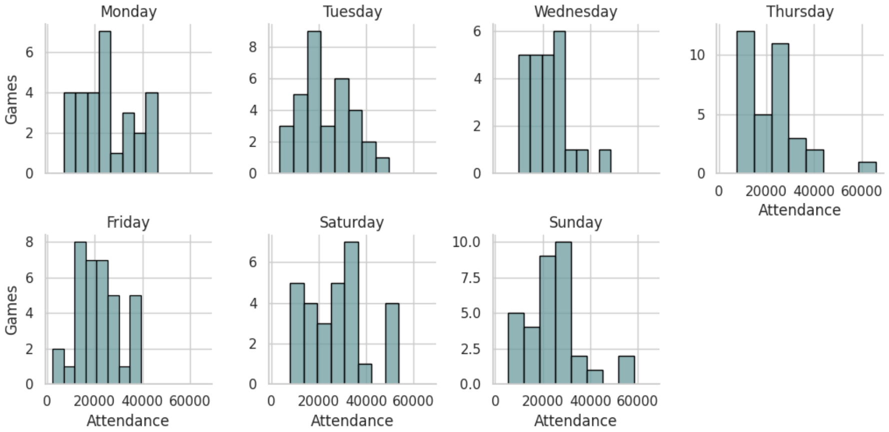


Figure 3: Observed attendance distributions by day of week. Weekend days show a small shift toward higher attendance and more high crowd games. Weekdays overlap more and sit closer to the league baseline.

Finally, we checked dispersion using the sample mean and variance of attendance (Code Cell 6). The variance is much larger than the mean, which shows overdispersion relative to a Poisson model. This diagnostic is one reason we later choose a Negative Binomial likelihood.

Model Description

Complete-pooling model

In the complete pooling model we assume every home game comes from one shared attendance distribution. This is the simplest baseline. It ignores which team is playing and which day the game occurs. The goal is to learn a league wide mean level of attendance and a league wide amount of extra variability, then check how far this simple story can go. Because attendance is a count and the exploratory check showed overdispersion, we use a Negative Binomial likelihood rather than a Poisson likelihood. A Poisson model forces the variance to equal the mean, but our data have variance far above the mean, so a Poisson model would understate uncertainty and fit the tail poorly. The Negative Binomial adds a dispersion parameter that allows wider spread, which matches the right skew and long tail seen in the histograms.

Formally, for game i

$$y_i \sim \text{NegBinomial}(\mu, \alpha)$$

where $\mu > 0$ is the shared mean attendance and $\alpha > 0$ controls dispersion. In the PyMC parameterization we use, the variance is

$$\text{Var}(y_i) = \mu + \frac{\mu^2}{\alpha}$$

so smaller α implies more overdispersion. This parameterization shows why we need α to capture the extra spread in the data.

Prior Checks

We choose priors that are weakly informative but anchored to the scale of the league. Professional football stadiums in Argentina vary a lot in size. Smaller Primera Division venues seat on the order of twenty to thirty thousand fans, for example Argentinos Juniors plays in a stadium of about 24,800 seats. At the other end, River Plate's Estadio Monumental holds about 84,000 to 85,000 spectators. Since attendance cannot exceed capacity and typical home crowds are some fraction of that capacity, it is reasonable to expect average attendances in the tens of thousands. To encode that scale while keeping $\mu > 0$, we put a LogNormal prior on μ by placing a Normal prior on $\log \mu$

$$\log \mu \sim \mathcal{N}(\log 20000,\ 0.4^2)$$

The center at 20,000 expresses a prior guess of a mid sized crowd for a professional league match. The standard deviation of 0.4 on the log scale is loose. It allows μ to vary over a wide range, covering small crowd scenarios and near sellout scenarios without forcing either (Figure 4).

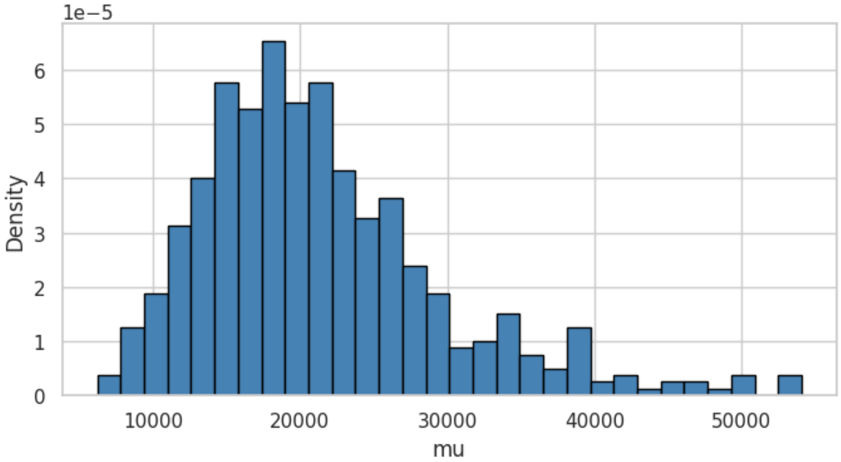


Figure 4: Prior distribution for the shared mean attendance μ in the complete pooling model. The prior places most probability on crowds in the tens of thousands, while still allowing lower and higher values.

For dispersion we use an Exponential prior

$$\alpha \sim \text{Exponential}(1/10)$$

This is because it enforces positivity and favors moderate values while still allowing smaller values. A mean of 10 implies a prior belief that overdispersion exists but is not extreme. In practice, α values near 2 to 15 cover a large range of plausible count variability for this setting. The prior histogram for α shows most mass in that region and a tail out to larger values, so the model can move toward lower or higher dispersion if the data require it (Figure 5).

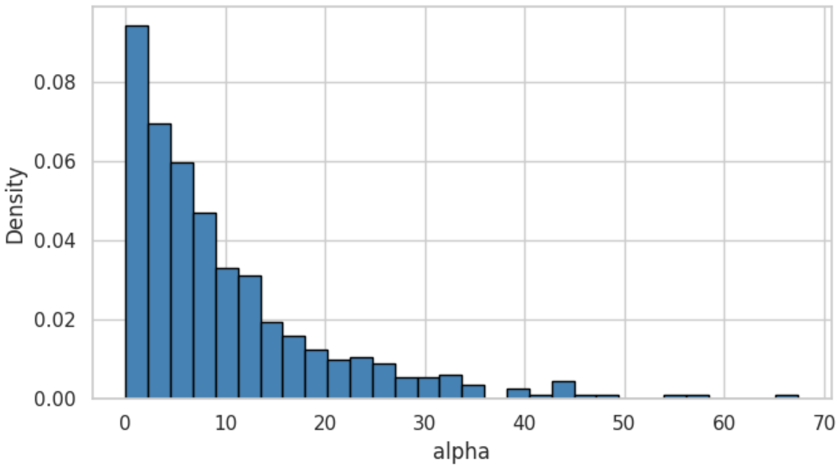


Figure 5: Prior distribution for the dispersion parameter α in the complete pooling model. The prior keeps most mass on moderate dispersion while allowing a long right tail.

We also ran a prior predictive check. We sampled y from the model before conditioning on the data and compared these draws to observed attendance. The prior predictive distribution overlaps the observed range and reproduces the right skew. It slightly underweights the mid to high range near 25k to 35k, but still places probability there. This indicates the priors are not too tight and do not rule out realistic attendance patterns. Because the prior predictive covers the data scale, we keep these priors (Figure 6).

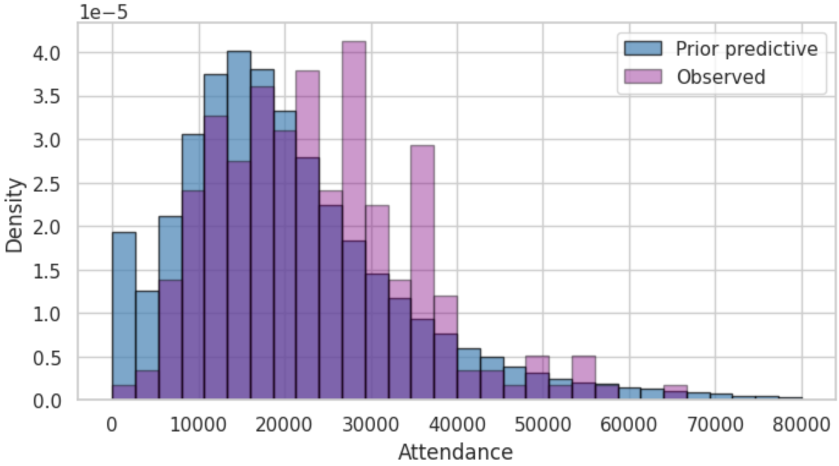


Figure 6: Prior predictive check for the complete pooling model. Blue bars show simulated attendance from the priors and purple bars show observed attendance.

Inference

To fit the model, we used 4 chains with 2000 tuning steps and 2000 posterior draws per chain. The tuning phase adapts the step size and mass matrix to the geometry of the posterior. We set `target_accept` to 0.9. This higher acceptance target is a standard choice for count models with heavy tails because it reduces the chance of divergent transitions. We also stored pointwise log likelihood so that later we can compare models with PSIS LOO.

Sampling diagnostics shows stable (Figure 7). The trace plots for μ and α show good mixing across chains and no sticking and the posterior densities from different chains overlap. The summary table gives $\hat{R} = 1.0$ for both parameters, which suggests convergence, and effective sample sizes are large for bulk and tail, which means the posterior is well explored. There were no divergences in this fit, so the sampler did not detect problematic curvature at the chosen tuning and acceptance settings.

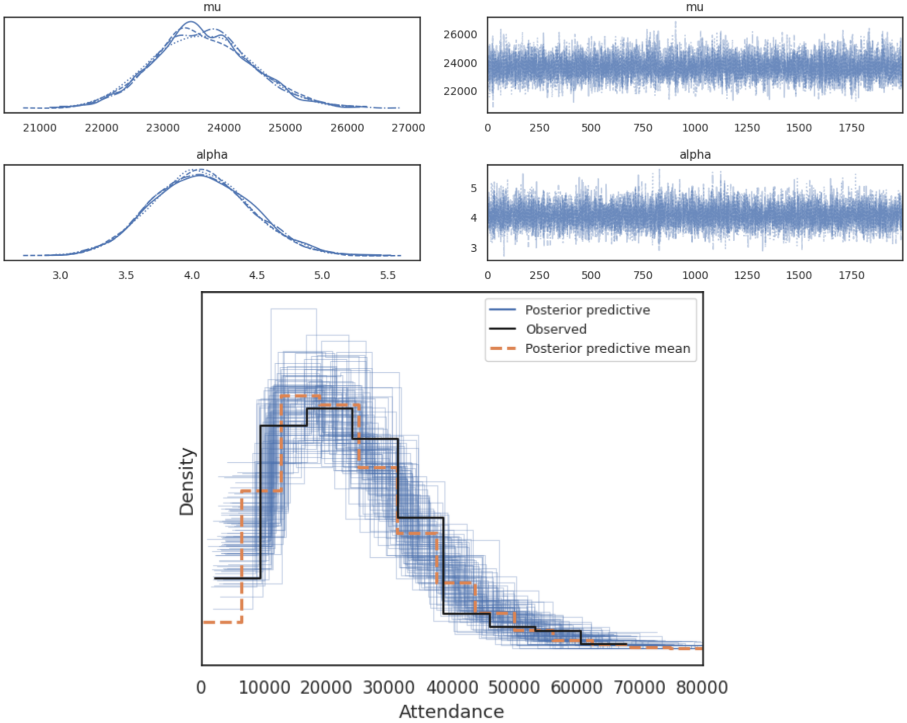


Figure 7: Trace plots and posterior densities for μ and α in the complete pooling model. Chains mix well and overlap, supporting convergence.

Posterior Checks

After fitting, the posterior for μ is centered near the middle of the data distribution (Table 1 and Figure 8). The posterior mean is about 23,650, and the 94 percent HDI is roughly 22,100 to 25,100. This is consistent with the histogram of observed counts, so the shared mean is plausible . The posterior for α is around 4.1 with a 94 percent HDI of about 3.37 to 4.78. Since smaller α implies higher variance, this confirms strong overdispersion relative to Poisson. Under the variance formula above, $\alpha \approx 4$ implies a variance many times the mean for games with μ in the tens of thousands. This matches the dispersion check done in preprocessing.

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS bulk	$\hat{R}$
μ	23649.9	799.6	22096.1	25113.8	7141	1.00
α	4.07	0.38	3.37	4.78	7954	1.00

Table 1: Posterior summary for the complete pooling model. μ is the league-wide mean attendance and α controls overdispersion.

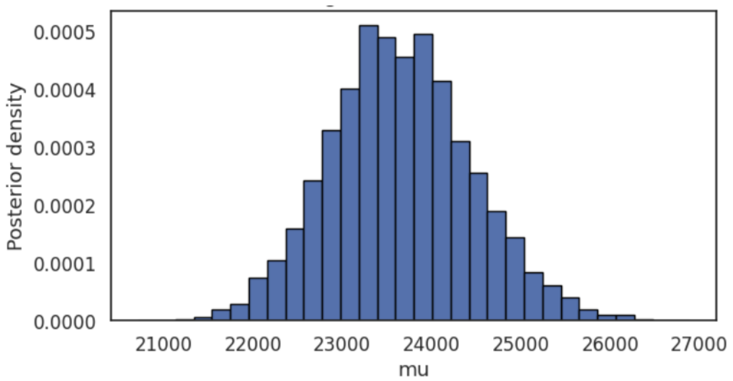


Figure 8: Posterior distribution of the league-wide mean attendance μ under the complete pooling model. Most mass is around 23k–24k, with moderate uncertainty.

We then generated replicated attendance datasets from the posterior and compared their density to the observed density. The PPC plot shows that the model captures the main right skew and the overall mass between about 10k and 35k. The predictive mean tracks the center of the observed distribution. The model also produces a long right tail, so it can generate high attendance games. Some mismatch remains. The pooled model smooths over visible differences between teams and days. In the PPC this shows up as draws that are too blunt, with less structure in the mid range than the observed data. This is expected because the model has no covariates to explain systematic shifts. The complete pooling model therefore serves as a baseline. It fits the league wide level and dispersion, but it cannot explain why some teams or days sit above or below the average.

Hierarchical model

We next allow attendance to vary by team and by day of week. A complete pooling model forces all games to share one mean, which ignores differences in fan bases and scheduling. A hierarchical model captures these differences while still sharing information across groups, which is useful because each team and each day has only a modest number of games, and partial pooling prevents extreme estimates driven by small samples.

Let y_{ij} be attendance for a home game played by team i on day j. We keep the same Negative Binomial likelihood used earlier because counts show extra dispersion beyond Poisson. The model is

$$y_{ij} \sim \text{NegBinomial}(\mu_{ij}, \alpha_{\text{disp}}),$$

with mean $\mu_{ij} > 0$ and dispersion $\alpha_{\text{disp}} > 0$. We model the mean on the log scale to ensure positivity and to make team and day effects additive in log space

$$\log \mu_{ij} = \alpha_0 + \beta_i + \gamma_j, \qquad \mu_{ij} = \exp(\alpha_0 + \beta_i + \gamma_j)$$

Here α_0 is the league level baseline log mean attendance. β_i is the team random effect and γ_j is the day random effect. In this parameterization, a positive β_i means team i draws more fans than the league baseline, and a positive γ_j means day j raises turnout relative to the baseline day.

Prior Checks

We place a prior on α_0 that reflects the scale of professional stadium attendance. Top division clubs in Argentina play in stadiums that range from medium venues to large arenas, as said before. Because α_0 is on the log scale, a Normal prior corresponds to a LogNormal prior on the mean attendance. We use

$$\alpha_0 \sim \mathcal{N}(\log 20000, 0.5^2)$$

Centering at $\log 20000$ encodes a baseline near twenty thousand tickets, which is a plausible middle point given common stadium sizes and ticket demand in the league. The standard deviation of 0.5 on log scale allows large variation, since $\exp(\alpha_0)$ can span from small crowds to sell out levels.

Team popularity differs because of history, market size, and fan culture. We model these differences with partial pooling

$$\beta_i \sim \mathcal{N}(0, \sigma^2_{\text{team}}), \qquad \sigma_{\text{team}} \sim \text{Exponential}(1)$$

The Normal prior on β_i says most teams are near the baseline, with some above and some below. The Exponential prior on σ_{team} favors modest variation but allows the data to support larger spread if needed. A mean of 1 on log scale implies that a one standard deviation team effect could change attendance by a factor of $\exp(1) \approx 2.7$, which is wide enough for a league where some clubs are national brands and others are local.

Days of week affect turnout through work schedules, travel time, and TV slots. Weekends might raise attendance, weekdays might lower it. We use the same hierarchical structure

$$\gamma_j \sim \mathcal{N}(0, \sigma_{\text{day}}^2), \qquad \sigma_{\text{day}} \sim \text{Exponential}(1)$$

This mirrors the team prior and expresses that day effects exist but are smaller than the baseline league scale. The prior again allows the data to decide the size of these effects.

We keep the dispersion prior from the first model

$$\alpha_{\text{disp}} \sim \text{Exponential}(1/10)$$

This gives prior mass on moderate overdispersion while allowing heavy tails. It matches the idea that attendance counts have extra variability from weather, rivalries, security limits, and late ticket sales.

To do a prior predictive check, we drew samples from the prior predictive distribution and compared them to plausible attendance ranges. The simulated counts cover small crowds and large crowds. The right tail is long, which is needed because sell out games and derby matches can push totals much higher than the baseline. This check supports that the priors are neither too narrow nor centered on an implausible scale (Figure 9).

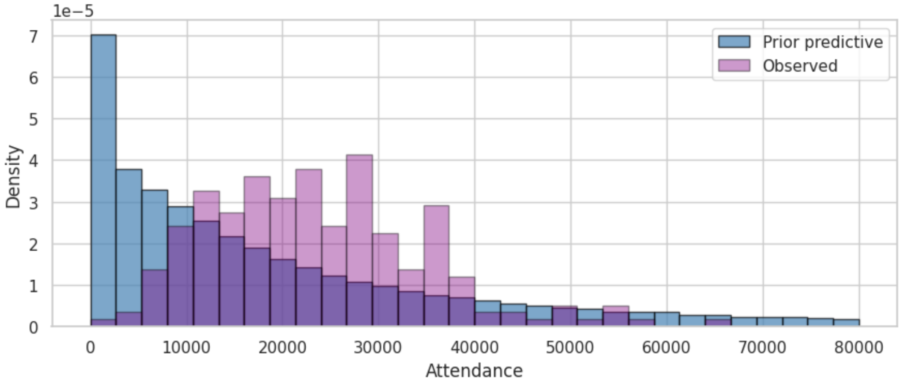


Figure 9: Prior predictive check for the hierarchical model. Blue bars are prior predictive draws and purple bars are observed attendance.

Inference

A hierarchical model introduces group effects for teams and days. In a centered form, we would write team effects as

$$\beta_i \sim \mathcal{N}(0, \sigma_{\text{team}}), \qquad \gamma_j \sim \mathcal{N}(0, \sigma_{\text{day}}).$$

This is the most direct way to express that teams and days vary around zero with unknown scales. The problem is that when a scale such as σ_{team} is small, the posterior has a funnel shape. Intuitively, small σ_{team} forces all β_i to live near zero, while large σ_{team} allows wide spread. These two regimes meet in a narrow region of parameter space. HMC must move through that narrow neck to explore both regimes, and it can lead to slow exploration and divergences.

Reparameterization means rewriting the same prior using a change of variables that separates the group effects from their scale. We introduce standardized raw effects with unit scale, then scale them up. For teams we set

$$\beta_i^{\text{raw}} \sim \mathcal{N}(0, 1), \qquad \beta_i = \beta_i^{\text{raw}}\,\sigma_{\text{team}}$$

For days we do the same

$$\gamma_j^{\text{raw}} \sim \mathcal{N}(0, 1), \qquad \gamma_j = \gamma_j^{\text{raw}}\,\sigma_{\text{day}}$$

This is called a non centered parameterization because the Gaussian prior is now placed on the raw effects, not on the final effects. The model is unchanged in meaning. If you integrate out β_i^{raw} you recover $\beta_i \sim \mathcal{N}(0, \sigma_{\text{team}})$, and the same for γ_j. What changes is the geometry the sampler sees. The raw effects are near independent of the scales in the prior, so the posterior looks more like a regular cloud instead of a funnel. That makes it easier for HMC to take stable steps.

After fitting the model with this form, we check diagnostics tied to the funnel issue. The pair plots of $(\alpha_0, \sigma_{\text{team}})$ and $(\alpha_0, \sigma_{\text{day}})$ show a compact shape without a narrow neck, and the sampler reports zero divergences. This matches the goal of the reparameterization, which is to improve sampling while keeping the same hierarchical assumptions.

The sampling report shows zero divergence (Figure 10). Trace plots mix well and the chains overlap. All reported $\hat{R}$ values equal 1.0 and effective sample sizes are large (Table 2) (Figure 11).

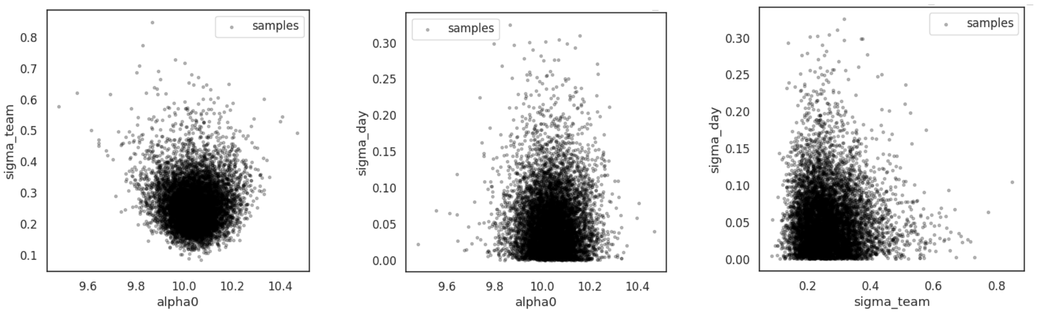


Figure 10: Joint posterior samples for hierarchical parameters. The clouds show stable geometry with no funnel shape, supporting the non-centered parameterization and the absence of divergences.

Parameter	Mean	SD	HDI 3%	HDI 97%	ESS bulk	$\hat{R}$
σ_{team}	0.267	0.078	0.135	0.405	2169	1.00
σ_{day}	0.054	0.044	0.000	0.130	3228	1.00
α_{disp}	4.990	0.481	4.070	5.867	6923	1.00
α_0	10.043	0.086	9.887	10.211	2198	1.00

Table 2: Posterior summary for the hierarchical model. α_0 is the baseline log mean, σ_{team} and σ_{day} are the scales of team and day effects, and α_{disp} controls dispersion.

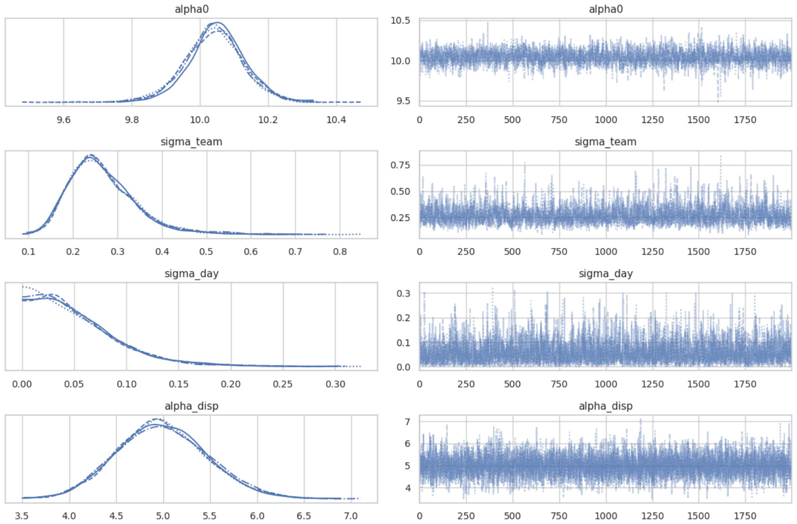


Figure 11: Trace plots and posterior densities for hierarchical parameters (α_0, σ_{team}, σ_{day}, and α_{disp}). Chains overlap and mix well, which supports convergence.

Posterior Checks

In a summary, the important posterior quantities are

- Baseline log mean α_0 has mean 10.043 and a 94 percent HDI from 9.887 to 10.211. On the original scale, $\exp(\alpha_0)$ gives a baseline attendance near 23 000 tickets for an average team on an average day.
- Team scale σ_{team} has mean 0.267 and 94 percent HDI 0.135 to 0.405. A one standard deviation team effect changes attendance by a factor of $\exp(0.267) \approx 1.31$. This implies sizable differences by team.
- Day scale σ_{day} has mean 0.054 and 94 percent HDI near 0 to 0.130. A one standard deviation day effect changes attendance by a factor of $\exp(0.054) \approx 1.06$. Day effects exist but are modest compared to team effects.
- Dispersion α_{disp} has mean 4.99 and 94 percent HDI 4.07 to 5.87. This supports overdispersion and justifies the Negative Binomial choice.

Overall, the posterior says team popularity is the main driver of variation, while day of week plays a secondary role.

We simulated draws from the posterior predictive distribution and overlaid them on the observed density (Figure 12). The hierarchical PPC shows strong overlap across the bulk of the distribution. The model reproduces the right skew and the long tail of large crowds. The predictive mean tracks the observed curve across low and medium values, with some remaining mismatch in the far right tail. This suggests there are still game specific shocks not captured by team and day alone, but the fit is much better than under complete pooling.

We simulated draws from the posterior predictive distribution and overlaid them on the observed density.

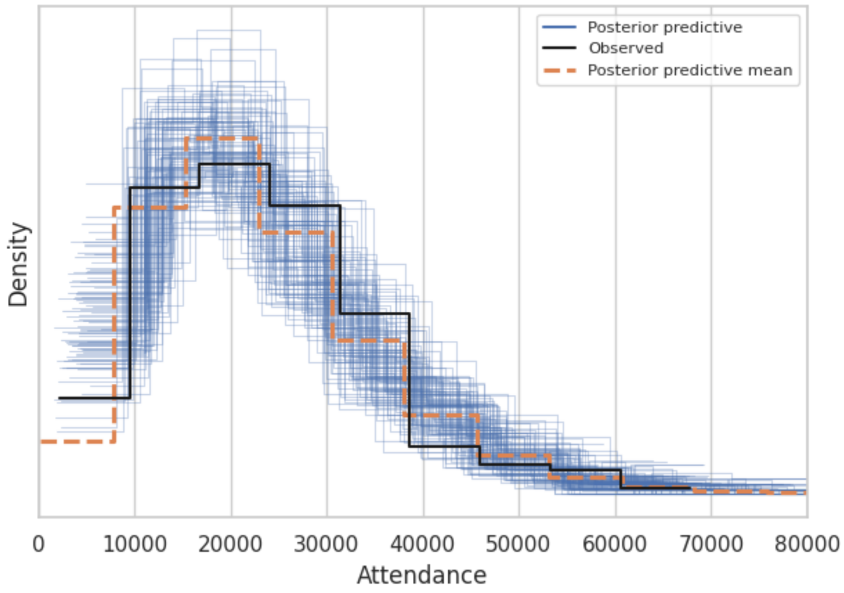


Figure 12: Posterior predictive check for the hierarchical model. The black curve is the observed density, blue curves are posterior predictive draws, and the dashed orange curve is the predictive mean. The model matches the main mass and the right tail of attendance.

After, we compare posterior predictive densities to observed attendance for each team (Figure 13). For teams with stable fan bases, the predicted distribution centers near the observed histogram and matches its spread. For teams with missing games, the predictive distribution remains wide, which is expected because the model expresses uncertainty rather than forcing a single point. The model also captures that large clubs tend to shift the distribution right, while smaller clubs shift it left. This pattern aligns with the positive and negative β_i draws in the posterior.

We also compared predictions by day (Figure 14). Weekend days show a mild right shift in predicted density relative to weekdays, matching the observed pattern. The overlap by day is strong, which fits the posterior result that σ_{day} is small but not zero.

Model Comparison

To decide which model to trust for prediction, we compare the complete pooling model and the hierarchical model using PSIS LOO cross validation. It estimates how well each model would predict a new game that was not used in fitting. It does this by leaving out each observation in turn and approximating the out of sample log predictive density.

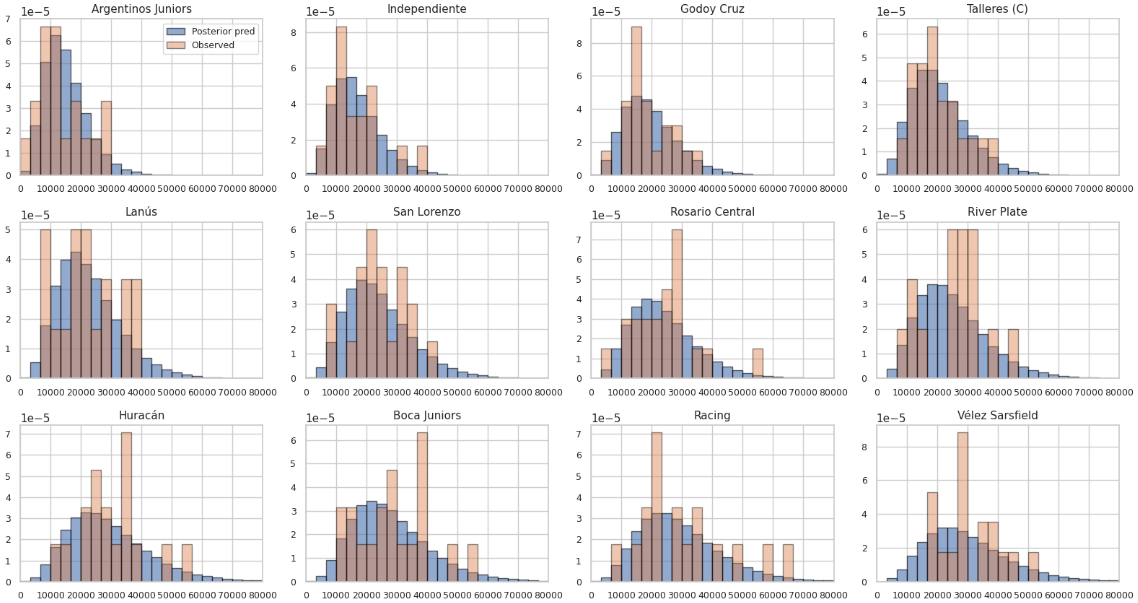


Figure 13: Posterior predictive versus observed attendance by team. Blue histograms are posterior predictive draws and orange histograms are observed games. The model captures the shift between large and small clubs and matches most team-level spreads.

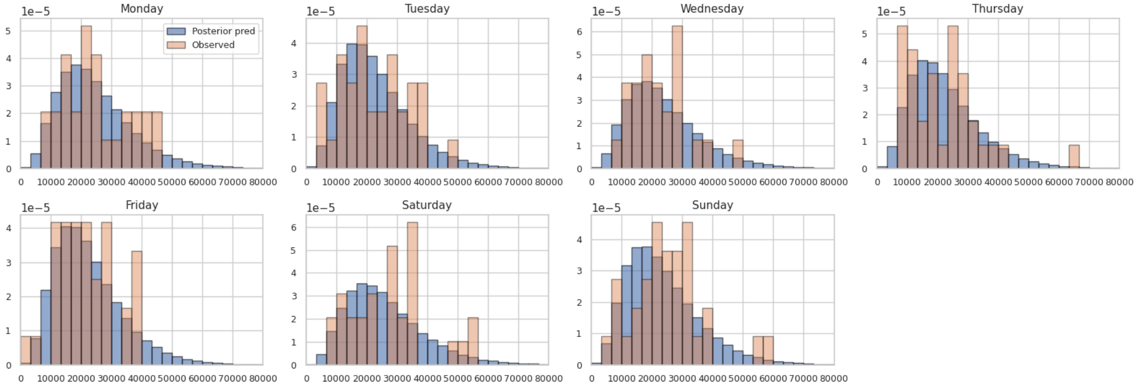


Figure 14: Posterior predictive versus observed attendance by day of week. Blue histograms are posterior predictive draws and orange histograms are observed games. Weekend days show a small right shift, while weekdays stay close to the baseline.

We focus on the expected log predictive density, elpd_{loo}. A higher value means better predictive fit because the model assigns more probability to data it has not seen (Figure 15) (Figure 16).

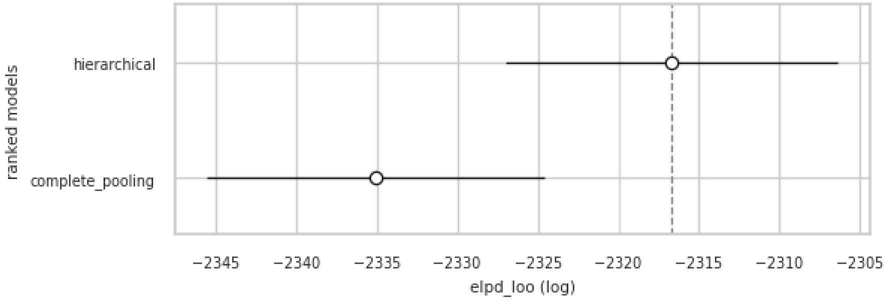
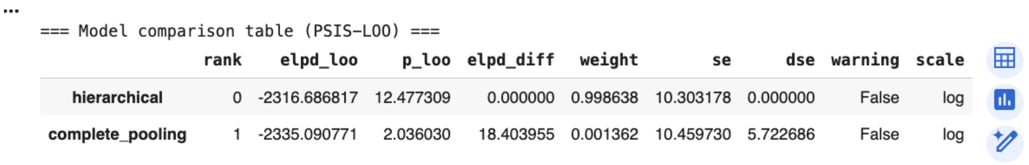


Figure 15: PSIS-LOO comparison plot. Points show elpd_loo estimates and horizontal bars show their standard errors. Higher elpd_loo (less negative) indicates better predictive fit, favoring the hierarchical model.



...

```
=== Model comparison table (PSIS-LOO) ===
```

	rank	elpd_loo	p_loo	elpd_diff	weight	se	dse	warning	scale
hierarchical	0	-2316.686817	12.477309	0.000000	0.998638	10.303178	0.000000	False	log
complete_pooling	1	-2335.090771	2.036030	18.403955	0.001362	10.459730	5.722686	False	log

Figure 16: PSIS-LOO model comparison table for the complete pooling and hierarchical models. The hierarchical model has a higher (less negative) expected log predictive density and almost all of the model weight, showing better out-of-sample predictive performance.

The results favor the hierarchical model. The hierarchical model has $\text{elpd}_{\text{loo}} = -2316.87$ while the complete pooling model has $\text{elpd}_{\text{loo}} = -2335.09$. The difference is about 18.40 points in elpd. This gap is large compared to the reported standard error for the difference, so it is not explained by Monte Carlo noise. The PSIS LOO weights also place almost all mass on the hierarchical model, with weight about 0.998 for the hierarchical model and about 0.002 for the pooling model. This means the hierarchical model is the clear winner in terms of out of sample prediction.

This outcome matches what we would expect. The hierarchical model adds team and day effects with partial pooling. Each team and each day gets its own adjustment to the baseline log mean, but these adjustments are tied together by shared group level priors. In practice, the model learns that games for high draw teams shift the mean upward while games for low draw teams shift it downward.

Given this, we use the hierarchical model for the rest of the analysis.

Prediction of Missing Attendance

Method for imputation

We refit the hierarchical model on the full dataset, keeping the missing attendance entries as missing in the likelihood. In PyMC, passing an array with `NaN` values to the observed node activates Bayesian imputation. The sampler treats those entries as latent variables drawn from the same generative process as the observed games. So, for each game g with team i and day j we use

$$y_g \sim \text{NegBinomial}(\mu_{ij}, \alpha_{\text{disp}}), \qquad \mu_{ij} = \exp(\alpha_0 + \beta_i + \gamma_j)$$

During sampling, the model produces posterior draws for y_g whenever y_g is missing. We then summarize these draws with a posterior mean prediction and a 94% highest density interval. The mean gives a single best guess. The interval shows uncertainty from both limited data and remaining game level noise.

Predicted values

Table 3 reports the imputed attendance for each missing game. The values line up with the patterns learned earlier. High draw teams such as River Plate, Boca Juniors, Racing, and Vélez Sarsfield receive predicted counts in the upper range of the league. Lower draw teams such as Argentinos Juniors and Independiente receive lower predictions. Weekend games tend to be above the same team midweek baseline when the day effect is positive, which is visible in the weekend predictions for Racing, Vélez Sarsfield, Talleres, and Argentinos Juniors (Figure 16).

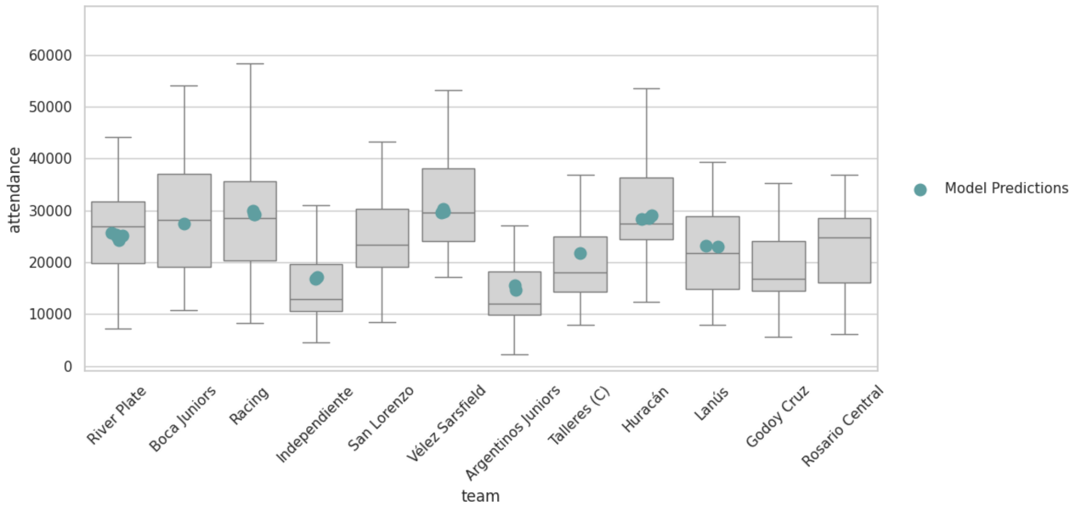


Figure 17: Imputed attendance by team. Gray boxplots summarize observed home-game attendance for each club. Teal dots are posterior mean predictions for missing games. Predictions sit within each team's observed spread, with higher values for large clubs and lower values for smaller clubs.

Uncertainty varies across cases. Games for teams with fewer observed matches or higher within team spread have wider intervals. For example, River Plate has several missing games but also a wide attendance range in observed data, so its intervals remain broad. In contrast, Independiente has lower predicted means and somewhat tighter bands, showing a more stable low to mid range draw in the season.

Team	Day	Predicted	HDI lower	HDI upper
River Plate	Thursday	24321	7271	46475
River Plate	Tuesday	25144	6365	47469
River Plate	Friday	24901	5170	45249
River Plate	Monday	25726	5042	46467
River Plate	Tuesday	25315	6325	46805
Boca Juniors	Wednesday	27498	6674	49658
Racing	Tuesday	29218	7200	53185
Racing	Sunday	30034	7026	54071
Racing	Tuesday	29183	7156	54773
Independiente	Tuesday	17237	4502	32651
Independiente	Thursday	16723	4194	30174
Vélez Sarsfield	Wednesday	29605	7338	54897
Vélez Sarsfield	Saturday	30390	8418	57018
Vélez Sarsfield	Monday	29794	7789	54552
Argentinos Juniors	Tuesday	14648	3979	28149
Argentinos Juniors	Saturday	15506	3714	28911
Talleres (C)	Saturday	21828	5567	40748
Huracán	Wednesday	28488	7089	53048
Huracán	Friday	29120	6397	54984
Huracán	Wednesday	28348	6504	52475
Lanús	Monday	23013	5561	41736
Lanús	Sunday	23293	5739	42887

Table 3: Predicted attendance for games with missing data. Predictions are posterior means with 94% highest density intervals.

Interpretation

These imputations should be read as model based reconstructions, not as exact recovered counts. Each prediction blends the global baseline α_0 which sets a league wide level, the team effect β_i shifts that level up or down depending on popularity and home support, and the day effect γ_j adjusts for typical differences in scheduling. The dispersion parameter α_{disp} leaves room for game specific shocks such as weather or competing events. Because the model pools information across all teams and days, predictions for sparse cases are stable while still reflecting real differences. This is the main practical advantage of partial pooling for filling gaps.

Consultant Summary

We built a model to understand what drives stadium attendance and to fill in the games where the scanner count is missing (Figure X). The main idea is that some teams attract more fans than others, and some days of the week bring bigger crowds. We measured both patterns at the same time, and we used a count model that allows extra spread in attendance, since game to game variation is large.

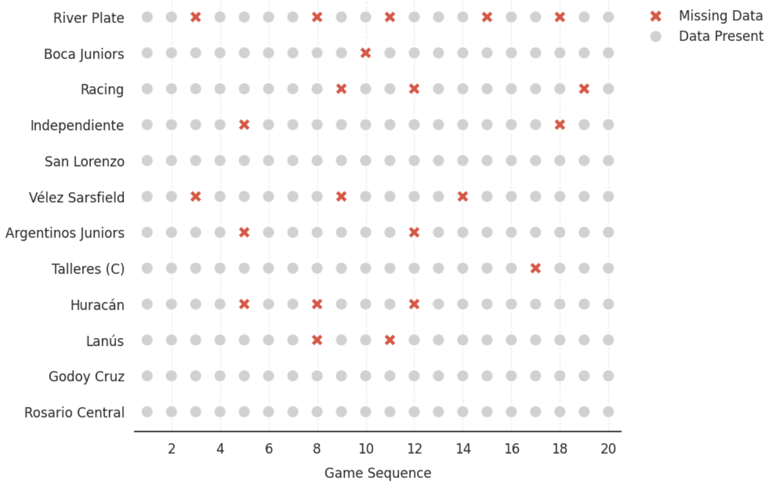


Figure 18: Missing attendance pattern by team and game order. Gray dots mark games with recorded attendance. Red crosses mark games where the scanner failed and attendance is missing.

The strongest result is that team popularity is the main driver of turnout. After controlling for day of week, teams still differ a lot in expected attendance. The posterior spread for team effects is much larger than the spread for day effects, so who is playing matters more than when they play.

Day of week matters, but the size of the effect is smaller. The model finds a modest weekend uplift compared with weekdays. The day effects are close to zero on the log scale and are tightly estimated, which means that day differences exist but do not reshape attendance as much as team differences do. This matches the predicted versus observed plots by day, where the distributions overlap a lot across Monday through Sunday, with only a gentle shift upward on Saturday and Sunday.

We have some reasons to trust these results. First, the hierarchical model predicts better than the complete pooling model. In the PSIS LOO comparison it has the higher expected log predictive density and almost all of the model weight, so it is the better tool for forecasting new games. Second, the posterior predictive checks show that simulated attendance from the model looks like the real data. The global PPC reproduces the right skew and wide spread, and the by team and by day PPCs show that the model captures high drawing and low drawing clubs without overreacting to a few unusual matches. Third, the hierarchical structure improves stability through shrinkage. Raw team averages can jump around when a club has only a few games. The model pulls extreme averages toward the global mean when data are limited, while keeping clear differences when the signal is strong.

We then used the hierarchical model to impute missing attendance counts. For each missing game we report a predicted mean and a 94% credible interval. The predicted values fall in the same range as the observed games for that team and day. Some intervals are wide because attendance has large natural variation from match to match. The filled gaps plot confirms that the predicted points sit inside the observed team level distributions, which supports the credibility of the imputations.

For planning purposes, team identity should be the first input when forecasting turnout or setting staffing and security levels. Day of week is still useful, mainly to flag a small weekend boost, but it does not offset the effect of a high profile or low profile home team. The same modeling approach can be updated each season to keep these baselines current and to improve planning for games with limited historical data.