Home-Team Bias in MLB Umpiring

Phuong Vu, Yixin Cui, and Corinne Bish

Background & Context

Umpire bias is widely debated — some studies link it to player race, others to home-team advantage.

The true effect of umpire bias on game outcomes remains unclear.

Prior research shows that home advantage is not solely due to umpire bias or inconsistencies.

This project investigates whether significant home-team bias exists and how it impacts game results.

Findings can influence team strategies and fan perceptions of game fairness.

Research Question

Is there a statistically significant home team bias in MLB games umpiring?

Objective

We aim to investigate the presence/significance of home team bias in MLB games to analyze its potential impact on game outcomes.

Dataset Description

01 Source

https://umpscorecards.us/data/umpires

02 Relevance

Umpire scorecard data from the 2015 - 2022 MLB seasons

03 Main Variables

favor_home: The difference between the home team's run expectancy impact and the away team's run expectancy impact for a given game

total_run_impact: The sum of the favor of every missed call

04 Supporting Variables

incorrect_calls
accuracy
consistency
umpire
home
away

Methodology Overview

Preprocessing

- **Remove the rows containing missing values**
- **Convert columns to numeric:**
  home_team_runs
  away_team_runs
  pitches_called
  incorrect_calls
  correct_calls
  accuracy
  consistency
  favor_home
  total_run_impact

Supervised Learning

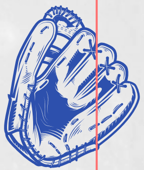
- **Regression:**
  - Models:
    - Simple Linear Regression
    - Multiple Linear Regression
  - Evaluation:
    - Validation set approach
    - K-Fold cross-validation
- **Classification:**
  - Models:
    - Logistic Regression
    - Decision Tree

Unsupervised Learning

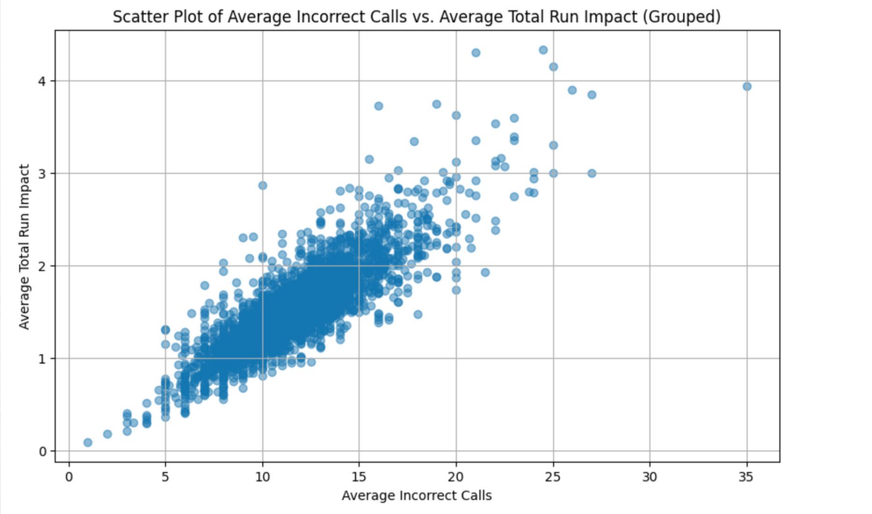
- **K-Means Clustering:**
  - Group decision patterns
  - Use a chi-squared test
- **Hierarchical Clustering:**
  - Agglomerative clustering
  - Dendrogram
  - Elbow plot
  - Define key clusters

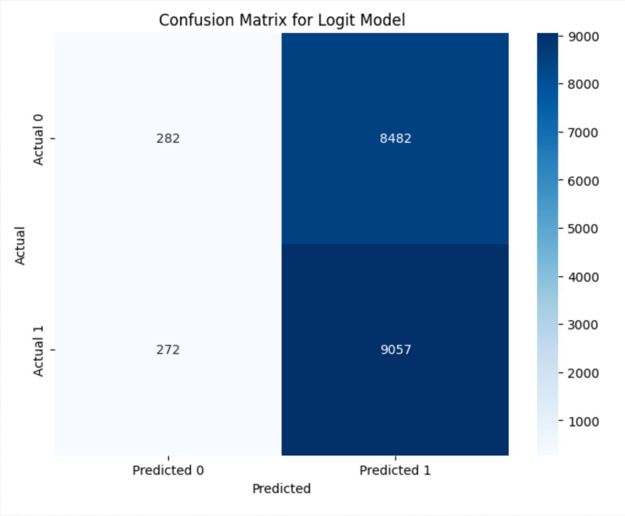
Supervised Learning

	Regression		Classification	
	Simple Linear Regression	Multiple Linear Regression	Logistic Regression	Decision Tree
Variables	**incorrect_calls** vs **favor_home** (Not Sig), **total_run_impact** (Sig +)	**incorrect_calls, accuracy, incorrect_calls * below_expected** vs **total_run_impact** (Sig +)	**incorrect_calls** vs **total_run_impact** (0 < median, else 1), **favor_home**	**incorrect_calls, umpire, home, away** vs **total_run_impact**
Evaluation Metrics	• **R-Squared:** 0.655 (TotRI) • **Validation MSE:** 0.2072 • **R-Squared:** 0.72 (Grouped) • **Average RMSE:** 0.2378	• **R-Squared:** 0.671 (TotRI) • **Validation MSE:** 0.443	• **R-Squared:** 0.43 (TotRI) • **Accuracy:** 0.516 (FavH)	• **Accuracy:** 0.77
Discussion	• Even though that incorrect calls favors the teams, it is unclear whether all incorrect calls directly favors only the home team but giving them a higher run expectancy. • Rather, incorrect calls favors the batter team.			

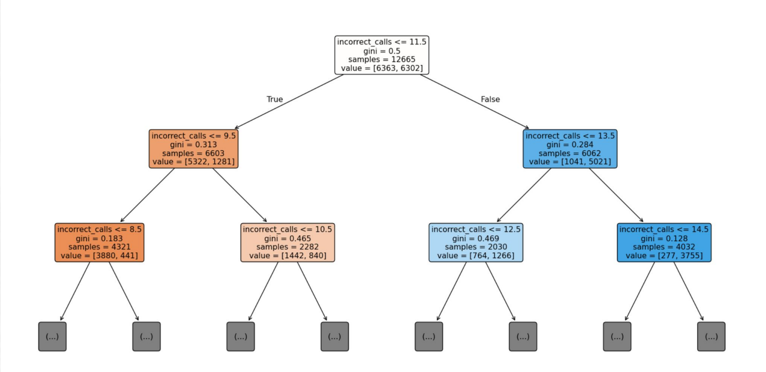


Results






```
                           Logit Regression Results
==============================================================================
Dep. Variable:     total_run_impact_category   No. Observations:          18093
Model:                                 Logit   Df Residuals:              18091
Method:                                  MLE   Df Model:                      1
Date:                       Mon, 28 Apr 2025   Pseudo R-squ.:            0.4361
Time:                               23:06:41   Log-Likelihood:          -7071.5
converged:                              True   LL-Null:                 -12541.
Covariance Type:                   nonrobust   LLR p-value:               0.000
==========================================================================
                     coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------
const             -6.7466      0.101    -67.089      0.000      -6.944      -6.549
incorrect_calls    0.5916      0.009     67.444      0.000       0.574       0.609
==========================================================================
```



Discussion & Limitations

Total Run Impact

Predictive Power

Strong correlation between incorrect calls and total run impact

↓

Support team strategy, especially performance or strategy during batter

Improve Accuracy

Incorporate umpire information and elevate accuracy

↓

Be better prepared when played as away team, do not want the home team gets the benefit from other elements and being the batter favored by umpire

Favor Home

Dataset

Look into a different dataset to see if it gives different result

Approach

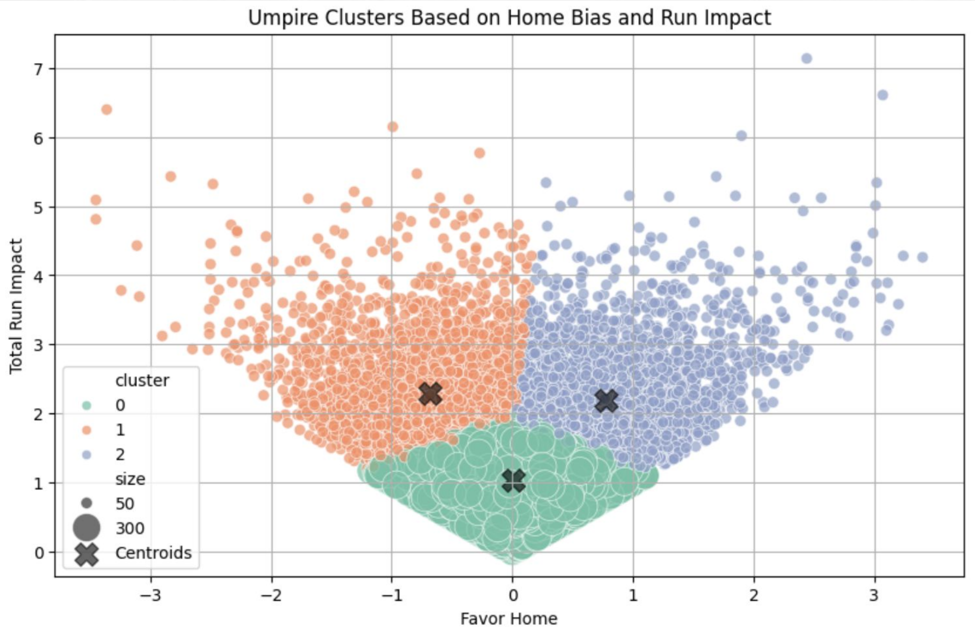
Try a different approach

K-Means Clustering

Overview

Used to **group decision patterns** and used a **chi-squared test** to assess whether bias toward home or away teams was statistically significant.

Results



Although most umpires are neutral, the chi-squared test (p = 0.022) shows **a significant bias overall for both home and away teams**. However, **home team bias is more common**, as shown by the larger cluster size.

Neutral	10577
Home Team Bias	3899
Away Team Bias	3617

K-Means Clustering

Discussion

Home team bias is most prevalent, which may impact perceptions of fairness among fans, teams, and officials.

Limitations

Relies only on two variables favor_home and total_run_impact. While these capture directional bias and impact, they may oversimplify the complexity of umpire behavior.

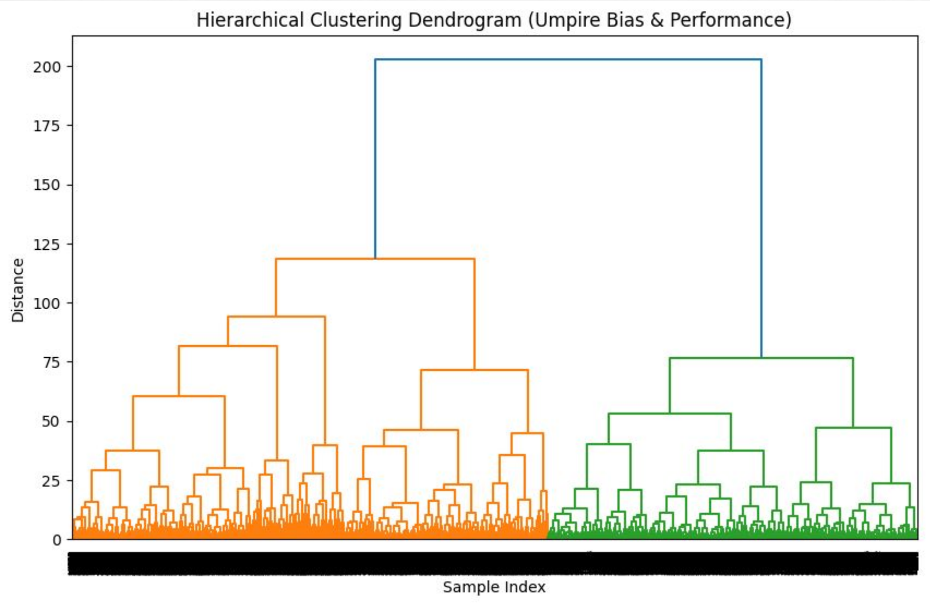
Does not account for important contextual factors like pitch type, inning pressure, umpire experience, or game intensity.

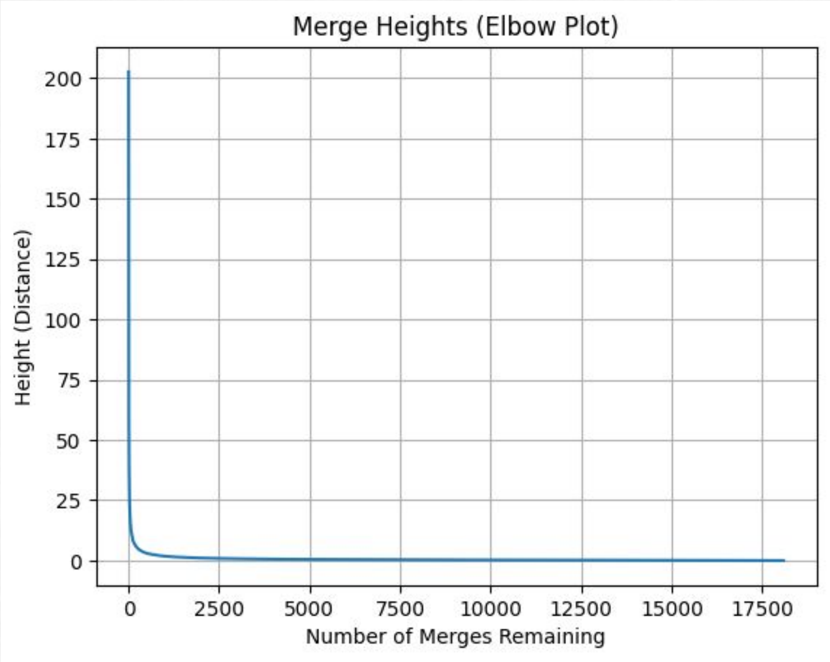
Hierarchical Clustering

Overview

Used **agglomerative clustering** to build a **dendrogram**, assess merge heights with an **elbow plot,** and define **key clusters**.

Results





Hierarchical Clustering

Results

cluster	favor_home	total_run_impact	accuracy	consistency
1	0.125919	0.976895	94.589246	94.445686
2	-0.582561	1.743164	91.673700	92.192913
3	0.369925	2.126765	90.043539	92.172754

- **Cluster 1** = Reliable umpires, small home bias, high accuracy
- **Cluster 2** = Away bias, medium errors
- **Cluster 3** = High home bias + worst accuracy → risk

Discussion

- Use Cluster 1 umpires in high-stakes games
- Monitor and retrain Cluster 3 umpires
- Promote transparency and fairness in officiating
- Support long-term performance tracking

Limitations

- Sensitive to outliers and scale
- Results may vary with linkage method
- Computationally intensive for large datasets

Future Directions

Modeling Enhancements

Explore non-linear models (Random Forest, XGBoost) for improved prediction of run impact or bias

Feature Expansion

Include contextual features like inning, score differential, or umpire experience for richer modeling

Bias Monitoring

Build a monitoring tool to track umpire cluster assignment over time to flag potential bias trends

Real-time Application

Incorporate this analysis into umpire assignment strategies for playoff or high-stakes games

Conclusion & Novelty

- **Supervised learning models** confirmed that incorrect calls increase run expectancy, although bias toward home teams is not automatic for every missed call.
- **Hierarchical and K-Means clustering** revealed clear umpire groups based on bias and game impact, highlighting a small but meaningful segment with strong home-team bias and lower performance.
- These insights can help leagues **assign more reliable umpires to critical games and proactively monitor officiating quality**.
- Future work should **incorporate more contextual variables** (e.g., pitch location, game tension) to refine bias detection and prediction.

References

Hsu, M. (2023). Umpire Home Bias in Major League Baseball. Journal of Sports Economics, 25(4), 423-442. https://doi.org/10.1177/15270025231222631 (Original work published 2024)

Flannagan, K.S., Mills, B.M. & Goldstone, R.L. The psychophysics of home plate umpire calls. Sci Rep 14, 2735 (2024). https://doi.org/10.1038/s41598-024-52402-y

Tainsky, Scott & Mills, Brian & Winfree, Jason. (2013). Further Examination of Potential Discrimination Among MLB Umpires. Journal of Sports Economics. 16. 10.1177/1527002513487740.

Kim, Jerry & King, Brayden. (2014). Seeing Stars: Matthew Effects and Status Bias in Major League Baseball Umpiring. Management Science. 60. 2619-2644. 10.1287/mnsc.2014.1967.