MLB Home Field Advantage: Myth or Reality?

Orlando Marin, Tatiana Eng

Why This Topic?

- **Home Field Advantage is Often Assumed** – Fans, analysts, and players often assume that playing at home gives teams an edge, but the data may tell a different story.
- **MLB's Unique Factors** – Unlike other sports, MLB stadiums have varying dimensions, weather conditions, and crowd sizes, making home-field advantage more complex.
- **Accessible & Engaging** – Baseball data is widely available and can be relatively simple to understand, making this topic interesting for both casual fans and data enthusiasts.
- **Great Fit for Visualization** – Comparing home vs. away performance across teams lends itself well to engaging visualizations and dashboards.

Our Data Sources (2022 MLB Season)

- Kaggle (MLB Ballparks)
  - Physical and environmental dimensions of Major League Baseball stadiums
  - https://www.kaggle.com/datasets/paulrjohnson/mlb-ballparks
- ESPN (Game Statistics)
  - Team records - Home vs Away
  - Hitting and pitching stats - Home vs Away
  - https://www.espn.com/mlb/standings/_/season/2022/group/overall
- ESPN (Average Team Attendance)
  - Average team attendance - Home vs Away
  - https://www.espn.com/mlb/attendance/_/year/2022
- Data Description
  - 30 rows - 1 row per team
  - 25+ columns - will be finalized once data preparation is complete

Our Data Sources (Kaggle)

team_name MLB Team	ballpark Ballpark Name	# left_field Distance to Left Field	# center_field Distance to Center Field	# right_field Distance to Right Field	# min_wall_height Wall Height at Lowest Point	# max_wall_height Wall Height at Highest Point	# hr_park_effects Ballpark Effect on Home Runs	# extra_distance Extra Distance Added by Ballpark	# avg_temp Average Temperature
30 unique values	30 unique values	310 355	395 420	302 353	3 11.5	8 37	75 150	-7.7 22	63.8 80.8
ATL	Truist Park	335	400	325	11	15	99	1.8	79.2
AZ	Chase Field	328	407	335	7.6	25	84	13.2	80.8
BAL	Oriole Park at Camden Yards	333	400	318	7	21	107	-0.5	76.4

Our Data Sources (ESPN)

Los Angeles Dodgers Batting Stats 2022

More MLB Teams

Batting | Pitching | Fielding

2022 Regular Season | Home

Player Batting Stats - Home

NAME	GP	AB	R	H	2B	3B	HR	RBI	TB	BB	SO	SB	AVG	OBP	SLG	OPS	WAR
Trea Turner SS	80	315	46	92	23	1	9	44	144	22	66	13	.292	.341	.457	.798	0.0
Freddie Freeman 1B	79	292	53	90	18	1	9	45	137	43	48	3	.308	.394	.469	.863	0.0
Mookie Betts SS	73	284	60	79	20	2	16	39	151	23	48	7	.278	.340	.532	.871	0.0
Will Smith C	69	243	31	59	11	2	10	43	104	28	49	1	.243	.332	.428	.760	0.0

Our Data Sources (ESPN)

MLB Attendance Report - 2022

Year: 2022

Stadium: Select One

Attendance | Top Games | Beane Count | Closers | Cy Predictor | Streaks | Managers | Debuts | Park Factors | Homers

2022 Attendance		Home				Road			Overall		
RK	TEAM	GMS	TOTAL	AVG	PCT	GMS	AVG	PCT	GMS	AVG	PCT
1	LA Dodgers	81	3,861,408	**47,671**	0.0	81	31,104	0.0	162	39,388	0.0
2	St. Louis	81	3,320,551	**40,994**	0.0	80	27,895	0.0	161	34,485	0.0
3	NY Yankees	78	3,136,207	**40,207**	0.0	81	30,418	0.0	159	35,220	0.0

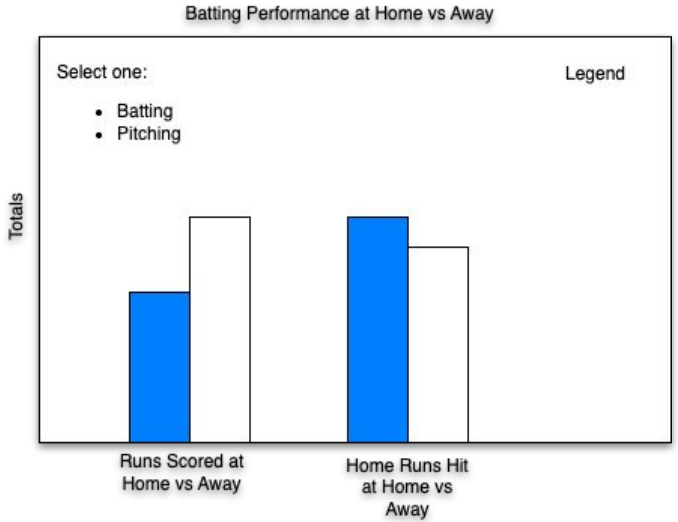
Key Questions We'll Answer

1. Do MLB teams perform better at home than away?
2. How do key in-game statistics (runs, homeruns, strikeouts, walks) differ at home vs away?
3. Which teams have the strongest and weakest home field advantage?
4. How does stadium location / elevation, dimensions, and attendance impact home-field advantage?
5. How do weather conditions (temperature, percentage of games played with roof) influence home-field advantage?

Preparing the Data

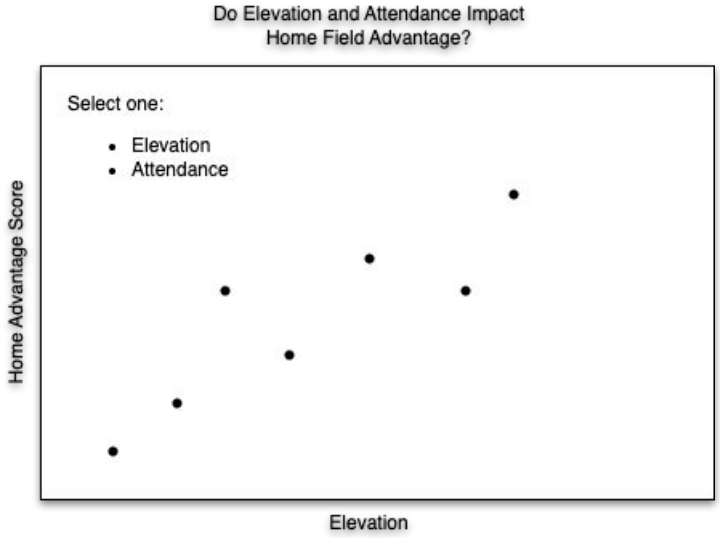
- Merge datasets to consolidate all necessary team, stadium, and performance data into a single source.
- Add stadium latitude & longitude to enable interactive map visualizations. Might also add stadium capacity.
- Calculate key metrics for each team:
  - Home Win Percentage = home_wins / (home_wins + home_losses)
  - Away Win Percentage = away_wins / (away_wins + away_losses)
  - Home Advantage Score = Home Win Percentage - Away Win Percentage
  - Home Percentage Seats Filled = average_attendance_home / stadium_capacity
- Standardize units for field dimensions and wall heights (yards or feet) to ensure consistency.

Dashboard Design and Visualizations



- Interactive bar chart that compares a team's offensive performance home vs away or a team's pitching performance home vs away
- User selects "batting" or "pitching", and based on the user's selection, the statistics displayed will change

Dashboard Design and Visualizations

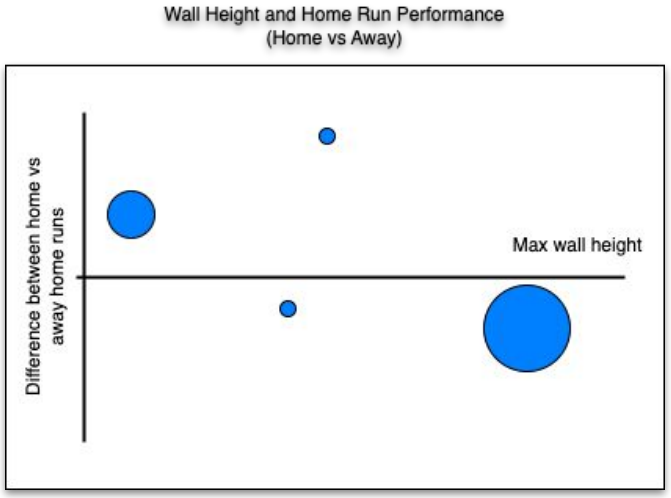


- Interactive scatter plot that shows the correlation between stadium elevation and home advantage score
- Users can toggle between “Elevation” and “Attendance” to also see the correlation between attendance and home advantage score

Dashboard Design and Visualizations

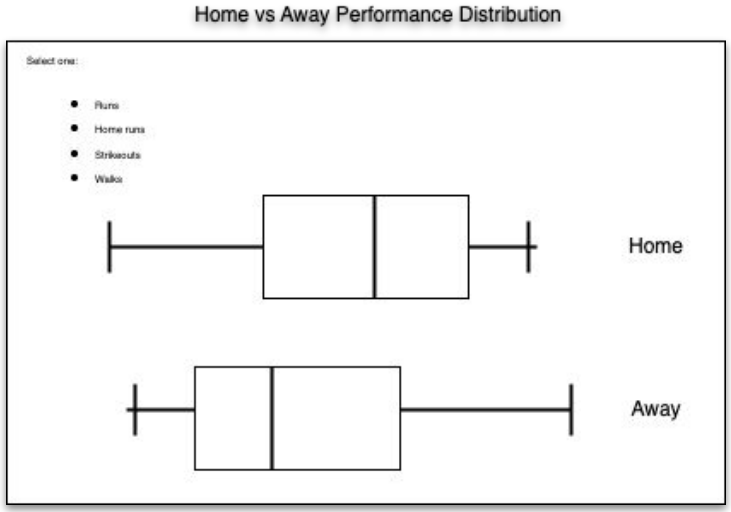
- Map to show the relationship between stadium location and Home Field Advantage
- Use stadium coordinates to plot the location of each stadium on a map
- The color of the bubbles will represent the Home Advantage Score
- The size of the bubbles will represent the average Home Attendance
- Users can hover over bubbles to see details about the team's stadium location and Home Advantage Score

Dashboard Design and Visualizations



- Interactive bubble chart that compares a stadium's max wall height, teams' differences in number of home runs hit at home vs away, and percentage of seats filled
- User hovers over a bubble to see statistics for that team

Dashboard Design and Visualizations



- Interactive box plot that shows home vs away performance (for all teams) for the selected statistic
- User has the option to select runs, home runs, strikeouts, or walks