

MLB Home Field Advantage: Myth or Reality?

Tatiana Eng, Orlando Marin



Topics of Discussion

- Introductions
- Background Information and Project Goals
- Dataset Selection and Data Preparation
- Data Visualizations and Code
- Analysis Limitations
- Takeaways
- Questions





Background Information and Project Goals



Overview of Major League Baseball

- 30 teams total, located in the U.S. and Canada
- 162 regular season games per season from April to October
 - Approximately half of a team's games are played at home and the other half on the road
 - A team will play the same opponent for 3 days in a row – called a “series”
- Each team has a 25-player roster
- 9 innings in a game
 - Away team bats first (top half) each inning
 - Home team bats second (bottom half) each inning
 - While one team is hitting, the other team has 9 players in the field - pitcher, catcher, 4 infielders, 3 outfielders
 - Each team bats until the opposing team gets 3 outs
 - Team with the most runs scored at the end of the game wins
 - *Since the home team bats last, if they are winning after the top of the 9th inning, the game is over, and the home team will not bat in the bottom of the 9th



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.
- **Great Fit for Visualization** – Comparing home vs. away performance across teams lends itself well to engaging visualizations and dashboards.



Key Questions We'll Answer

1. Do MLB teams perform better at home than away, and why?
2. Which teams have the strongest and weakest home field advantage?
3. How do batting and pitching statistics differ at home vs away?
4. How does travel impact teams' performance?
5. Does environmental or park factors impact home field advantage?





Dataset Selection and Data Preparation



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>
- Kaggle (Sports Stadium Locations)
 - Latitude and longitude of each ballpark
 - <https://www.kaggle.com/datasets/logandonaldson/sports-stadium-locations>
- 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
- BetMGM
 - MLB Stadium Seating Capacity
 - <https://sports.betmgm.com/en/blog/mlb/biggest-mlb-stadiums-ranking-by-capacity-bm15/>
- Baseball Savant
 - Each team's number of miles traveled in 2022
 - <https://baseballsavant.mlb.com/visuals/map?team=&year=2022>



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.
- Added seating capacity into the data to calculate average home attendance rate
- Used the Home Wins / Losses, and Away Wins / Losses, to determine each team's win percentage at home vs away, and then used that information to calculate a Home Field Advantage score for each team, which is each team's home win percentage - away win percentage
- Added miles traveled into the data to see the relationship between travel distance and home field advantage



Our Dataset (30 rows, 30 columns)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	team_name	team_abv	ballpark	ballpark_lat	ballpark_long	left_field_ft	center_field_ft	right_field_ft	min_wall_heig ht_ft	max_wall_hei ght_ft	avg_temp_f	elevation_ft	roof_pct	daytime_pct
2	Atlanta Braves	ATL	Truist Park	33.89	-84.468	335	400	325	11	15	79.2	1001	0	0.31
3	Arizona Diamondbacks	AZ	Chase Field	33.445278	-112.066944	328	407	335	7.6	25	80.8	1086	0.81	0.31
4	Baltimore Orioles	BAL	Oriole Park at Camden Yards	39.283889	-76.621667	333	400	318	7	21	76.4	33	0	0.35
5	Boston Red Sox	BOS	Fenway Park	42.34625	-71.09775	310	420	302	3	37	69.5	21	0	0.31
6	Chicago Cubs	CHC	Wrigley Field	41.948056	-87.655556	355	400	353	11.5	15	70.2	595	0	0.55
7	Cincinnati Reds	CIN	Great American Ball Park	39.0975	-84.506667	328	404	325	8	12	77.9	535	0	0.5
8	Cleveland Guardians	CLE	Progressive Field	41.495833	-81.685278	325	405	325	9	19	70.8	653	0	0.36
9	Colorado Rockies	COL	Coors Field	39.756111	-104.994167	347	415	350	8	17	75.4	5190	0	0.38
10	Chicago White Sox	CWS	Guaranteed Rate Field	41.83	-87.633889	330	400	335	8	8	71.6	595	0	0.38
11	Detroit Tigers	DET	Comerica Park	42.339167	-83.048611	345	420	330	7	9	73.3	600	0	0.61
12	Houston Astros	HOU	Minute Maid Park	29.756944	-95.355556	315	409	326	7	21	73	45	0.99	0.32
13	Kansas City Royals	KC	Kauffman Stadium	39.051	-94.48	330	410	330	9	9	77.4	856	0	0.49
14	Los Angeles Angels	LAA	Angel Stadium	33.800278	-117.882778	330	396	330	5	8	76.8	151	0	0.28
15	Los Angeles Dodgers	LAD	Dodger Stadium	34.073611	-118.24	330	395	330	4.5	8	74	515	0	0.24
16	Miami Marlins	MIA	LoanDepot Park	25.778056	-80.219722	344	407	335	9	12	72.2	10	0.98	0.27
17	Milwaukee Brewers	MIL	American Family Field	43.028333	-87.971111	344	400	345	8	8	73.4	597	0.45	0.36
18	Minnesota Twins	MIN	Target Field	44.981667	-93.278333	339	404	328	8	23	71.2	815	0	0.41
19	New York Mets	NYM	Citi Field	40.756944	-73.845833	335	408	330	8	8	73.9	10	0	0.33
20	New York Yankees	NY Yankees	Yankee Stadium	40.829167	-73.926389	318	408	314	8	8	72.8	55	0	0.39
21	Oakland Athletics	OAK	Oakland Coliseum	37.751667	-122.200556	330	400	330	8	8	70.6	3	0	0.49
22	Philadelphia Phillies	PHI	Citizens Bank Park	39.905833	-75.166389	329	401	330	6	13	76.6	20	0	0.34
23	Pittsburgh Pirates	PIT	PNC Park	40.446944	-80.005833	325	399	320	6	21	73.9	780	0	0.29
24	San Diego Padres	SD	Petco Park	32.7073	-117.1566	336	396	322	4	10	71.7	23	0	0.38
25	Seattle Mariners	SEA	T-Mobile Park	47.591	-122.333	331	405	327	7	15	63.8	10	0.16	0.38
26	San Francisco Giants	SF	Oracle Park	37.778611	-122.389167	339	399	309	8	25	64.4	0	0	0.49
27	St. Louis Cardinals	STL	Busch Stadium	38.6225	-90.193056	336	400	335	8	8	79.7	460	0	0.32
28	Tampa Bay Rays	TB	Tropicana Field	27.768333	-82.653333	315	404	322	9	11	72	15	1	0.41
29	Texas Rangers	TEX	Globe Life Field	32.747361	-97.084167	329	407	326	8	14	75.2	545	0.8	0.38
30	Toronto Blue Jays	TOR	Rogers Centre	43.641389	-79.389167	328	404	328	8	8	70.2	270	0.37	0.45
31	Washington Nationals	WSH	Nationals Park	38.872778	-77.0075	336	402	335	8	12	76.5	35	0	0.42



Our Dataset (30 rows, 30 columns)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	team_name	team_abv	ballpark	home_wins	home_losses	away_wins	away_losses	runs_scored_home	runs_scored_away	home_runs_h	home_runs_a	walks_home	walks_away	strikeouts_home	strikeouts_away	seat_capacity	avg_attendance_home	avg_attendance_away	miles_traveled											
2	Atlanta Braves	ATL	Truist Park	55	26	46	35	395	394	125	118	238	262	834	720	41084	38641	28216	30220											
3	Arizona Diamondbacks	AZ	Chase Field	40	41	34	37	353	349	71	102	250	254	633	583	48405	19817	29182	35924											
4	Baltimore Orioles	BAL	Oriole Park at Camden Yards	45	36	38	43	343	331	79	92	230	213	616	598	44970	17543	26132	27734											
5	Boston Red Sox	BOS	Fenway Park	43	38	35	46	392	343	86	69	263	263	669	677	37755	32408	26106	31572											
6	Chicago Cubs	CHC	Wrigley Field	37	44	37	44	324	333	73	86	263	277	722	661	41649	32305	29044	26828											
7	Cincinnati Reds	CIN	Great American Ball Park	33	48	29	52	367	281	89	67	298	314	750	664	43500	17447	30005	25574											
8	Cleveland Guardians	CLE	Progressive Field	46	35	46	35	325	373	50	77	228	207	716	674	34830	17050	25315	28771											
9	Colorado Rockies	COL	Coors Field	41	40	27	54	456	242	98	51	245	294	581	606	46897	32467	27801	29708											
10	Chicago White Sox	CWS	Guaranteed Rate Field	37	44	44	37	331	355	77	72	273	260	739	711	40615	24704	22118	30135											
11	Detroit Tigers	DET	Comerica Park	36	46	30	50	279	278	51	59	263	248	601	594	41083	19634	23612	34104											
12	Houston Astros	HOU	Minute Maid Park	55	26	51	30	368	369	116	98	217	241	834	690	41168	33197	25112	37977											
13	Kansas City Royals	KC	Kauffman Stadium	39	42	26	55	335	305	65	73	297	292	585	606	37903	15974	22474	28719											
14	Los Angeles Angels	LAA	Angel Stadium	40	41	33	48	330	293	106	84	261	279	745	638	45517	30339	25561	44038											
15	Los Angeles Dodgers	LAD	Dodger Stadium	57	24	54	27	422	425	106	106	209	198	775	690	56000	47671	31104	36694											
16	Miami Marlins	MIA	LoanDepot Park	34	47	35	46	296	290	71	73	253	258	739	698	37442	11203	27803	38852											
17	Milwaukee Brewers	MIL	American Family Field	46	35	40	41	362	363	110	109	238	283	824	706	41900	30155	27290	28557											
18	Minnesota Twins	MIN	Target Field	46	35	32	49	343	353	89	89	252	216	726	610	38544	22514	22827	31759											
19	New York Mets	NYM	Citi Field	54	27	47	34	376	396	81	90	221	207	828	737	41922	33308	29330	34686											
20	New York Yankees	NYY	Yankee Stadium	57	24	42	39	419	388	136	118	216	228	775	684	46537	40207	30418	32775											
21	Oakland Athletics	OAK	Oakland Coliseum	29	51	31	51	251	317	53	84	257	246	610	593	47170	9973	26031	42918											
22	Philadelphia Phillies	PHI	Citizens Bank Park	47	34	40	41	396	351	107	98	223	240	716	707	42901	28459	29193	34313											
23	Pittsburgh Pirates	PIT	PNC Park	34	47	28	53	313	278	74	84	284	302	627	623	38747	15524	28495	25306											
24	San Diego Padres	SD	Petco Park	44	37	45	36	404	404	77	76	230	238	734	717	40209	36931	29754	38460											
25	Seattle Mariners	SEA	T-Mobile Park	46	35	44	37	318	372	97	100	191	256	735	656	47929	28590	22973	46386											
26	San Francisco Giants	SF	Oracle Park	44	37	37	44	363	353	86	97	220	221	716	654	41265	30650	29337	35027											
27	St. Louis Cardinals	STL	Busch Stadium	53	28	40	41	392	380	98	99	220	269	612	565	44383	40994	27895	26772											
28	Tampa Bay Rays	TB	Tropicana Field	51	30	35	46	336	330	71	68	191	193	746	638	42735	19827	25027	34320											
29	Texas Rangers	TEX	Globe Life Field	34	47	34	47	342	365	101	97	300	281	703	611	40300	24831	22443	36945											
30	Toronto Blue Jays	TOR	Rogers Centre	47	34	45	36	359	416	102	98	206	218	708	682	41500	32763	26280	29880											
31	Washington Nationals	WSH	Nationals Park	26	55	29	52	297	306	76	60	283	275	651	569	41339	25017	27839	32999											



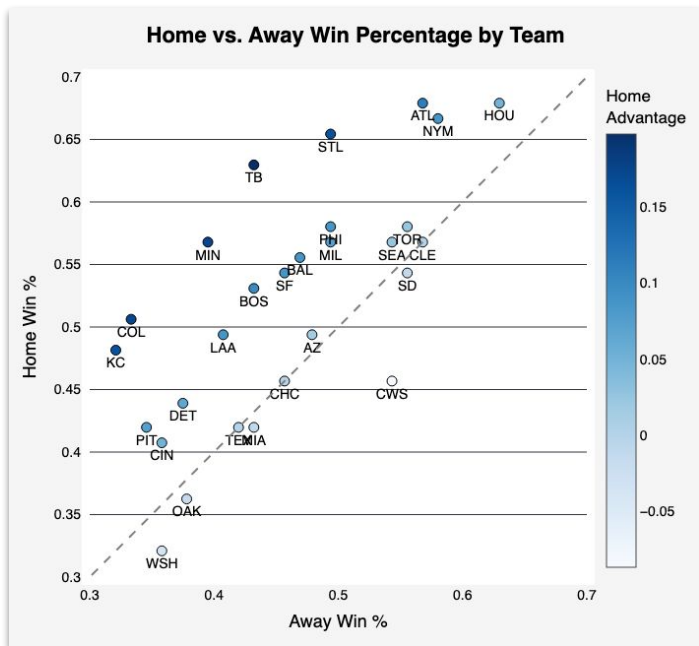


Data Visualizations and Code

(mlb-home-field-advantage.streamlit.app/)



Win Rates Show That Home Field Matters in Major League Baseball



```
# FIGURE 1: HOME VS AWAY WIN PERCENTAGE SCATTER PLOT

st.subheader("Win Rates Show That Home Field Matters in Major League Baseball")

with st.expander("How to read this scatter plot"):
    st.markdown("""
        This scatter plot shows each MLB team's home win percentage versus away win percentage.
        Each point represents a team, labeled by its abbreviation.

        - **Above the diagonal line** = better performance at home
        - **Below the line** = better performance on the road
        - **Color intensity** reflects the size of the team's home field advantage score
        - **Hover** over a point to view the team's full name, win percentages, and advantage score
        """)

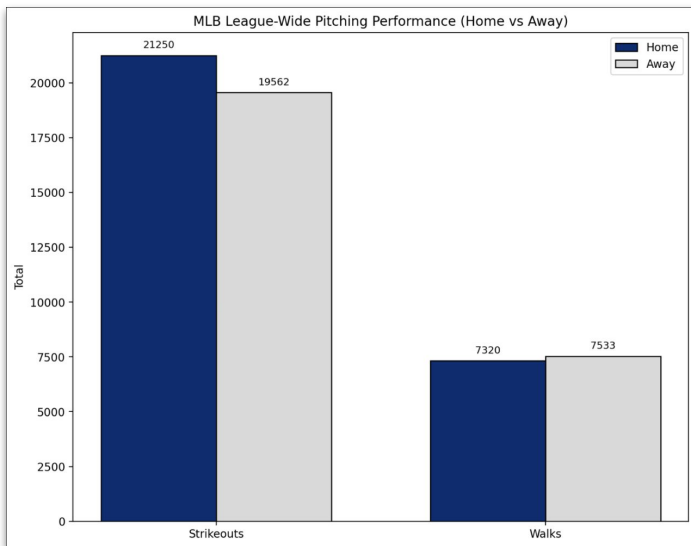
# Calculate win percentages
df["home_win_pct"] = df["home_wins"] / (df["home_wins"] + df["home_losses"])
df["away_win_pct"] = df["away_wins"] / (df["away_wins"] + df["away_losses"])
df["home_field_advantage"] = df["home_win_pct"] - df["away_win_pct"]

# Create the scatter plot
fig = go.Figure()

# Add team points
fig.add_trace(go.Scatter(
    x=df["away_win_pct"],
    y=df["home_win_pct"],
    mode="markers+text",
    text=df["team_abv"], # Shows abbreviations on the plot
    hovertext=df["team_name"], # Full name on hover
    textposition="bottom center",
    textfont=dict(size=14, color="black"),
    cliponaxis=False,
    marker=dict(
        size=10,
        color=df["home_field_advantage"],
        colorscale="Blues",
        colorbar=dict(title="Home<br>Advantage"),
        line=dict(width=1, color="black")
    ),
)))
```



MLB Totals Reveal Stronger Pitching and Slight Offensive Boost at Home



```
# FIGURE 2: DOUBLE BAR GRAPHS

st.subheader("MLB Totals Reveal Stronger Pitching and Slight Offensive Boost at Home")

# Dropdown to select Batting or Pitching
option = st.selectbox(
    "Use the dropdown below to switch between batting and pitching metrics:",
    ("Pitching", "Batting")
)

# Calculate league-wide totals and create figure
if option == "Batting":
    total_runs_home = df["runs_scored_home"].sum()
    total_runs_away = df["runs_scored_away"].sum()
    total_hr_home = df["home_runs_home"].sum()
    total_hr_away = df["home_runs_away"].sum()

# Set up categories and totals
categories = ["Runs Scored", "Home Runs"]
home_totals = [total_runs_home, total_hr_home]
away_totals = [total_runs_away, total_hr_away]

# Set x positions for each category
x = range(len(categories)) # [0, 1]
width = 0.35 # width of each bar

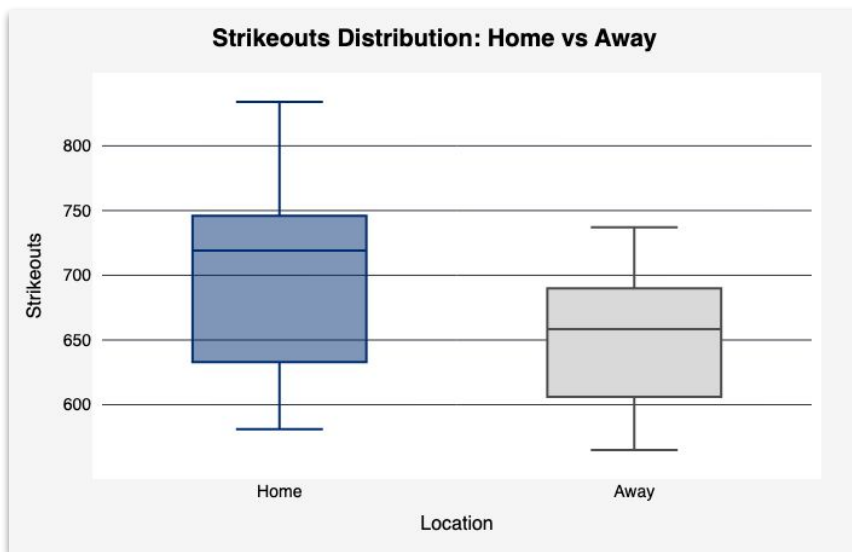
# Calculate x positions for the Home and Away bars
x_home = [pos - width/2 for pos in x] # Home bars shifted left
x_away = [pos + width/2 for pos in x] # Away bars shifted right

# Plot the bars with custom colors
fig, ax = plt.subplots(figsize=(10, 8))
bars_home = ax.bar(x_home, home_totals, width, label="Home", color="#002D72", edgecolor="black")
bars_away = ax.bar(x_away, away_totals, width, label="Away", color="#D9D9D9", edgecolor="black")

# Add value labels above each bar
for bar in bars_home + bars_away:
    height = bar.get_height()
    ax.annotate(f'{int(height)}',
                xy=(bar.get_x() + bar.get_width()/2, height),
                xytext=(0, 5), # 5 points vertical offset
                textcoords="offset points",
                ha='center', va='bottom', fontsize=9)
```



Pitching Fuels MLB Home Field Advantage More Than Hitting



```
# FIGURE 3: BOX PLOTS

st.subheader("Pitching Fuels MLB Home Field Advantage More Than Hitting")

# Dropdown for metric selection
metric = st.selectbox(
    "Use the dropdown below to switch between runs, home runs, strikeouts, and walks:",
    ("Runs", "Home Runs", "Strikeouts", "Walks")
)

# Map user-friendly labels to actual column names in the dataset
metric_columns = {
    "Runs": ("runs_scored_home", "runs_scored_away"),
    "Home Runs": ("home_runs_home", "home_runs_away"),
    "Strikeouts": ("strikeouts_home", "strikeouts_away"),
    "Walks": ("walks_home", "walks_away")
}

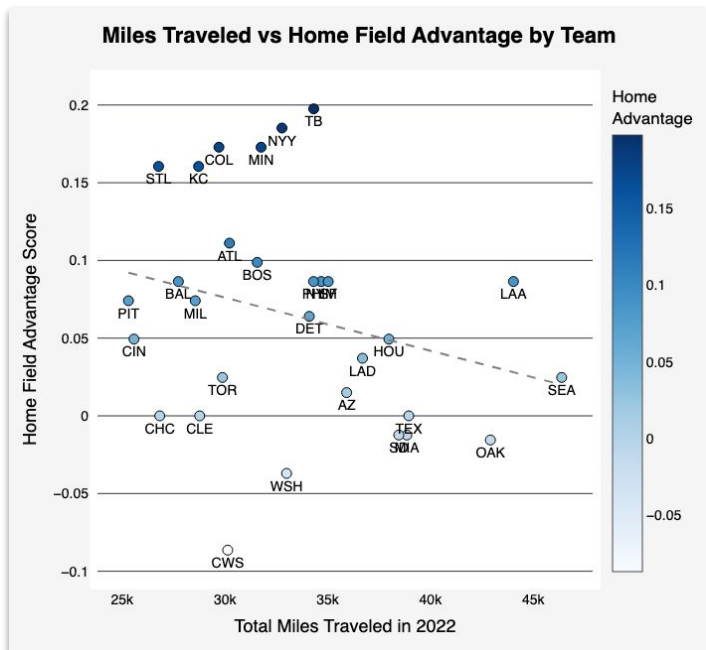
home_col, away_col = metric_columns[metric]

# Prepare the data in long format for box plot
data = pd.DataFrame({
    "Performance": df[home_col].tolist() + df[away_col].tolist(),
    "Location": ["Home"] * len(df) + ["Away"] * len(df)
})

# Create box plot with custom colors
fig = px.box(
    data,
    x="Location",
    y="Performance",
    color="Location",
    color_discrete_map={
        "Home": "#002D72",  # Navy
        "Away": "#D9D9D9"  # Light gray
    }
)
```



More Travel, More Trouble? Road Fatigue and Home Field Advantage



```
# FIGURE 4: MILES TRAVELED SCATTER PLOT

st.subheader("More Travel, More Trouble? Road Fatigue and Home Field Advantage")

with st.expander("How to read this scatter plot"):
    st.markdown("""
        This scatter plot shows the relationship between the **total miles traveled** by each MLB team and their **home field advantage score** in 2022.

        - **Higher up = stronger home field advantage**
        - **Farther right = more miles traveled**
        - Each point represents a team (abbreviation shown)
        - Hover over a point to view full team name, travel miles, and advantage score
        - A *trendline* has been added to show the overall direction of the relationship
        """)

# Prepare data
x = df["miles_traveled"]
y = df["home_field_advantage"]

# Fit a linear regression model
slope, intercept = np.polyfit(x, y, 1)
line_x = np.linspace(x.min(), x.max(), 100)
line_y = slope * line_x + intercept

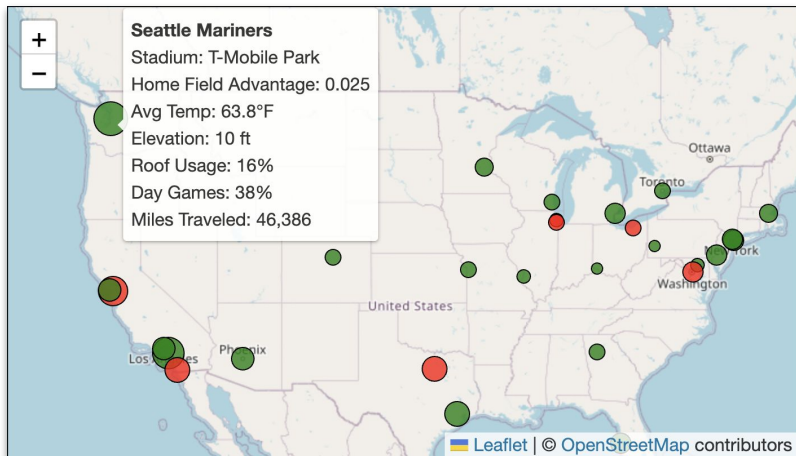
# Create the scatter plot
fig = go.Figure()

# Team data points
fig.add_trace(go.Scatter(
    x=x,
    y=y,
    mode="markers+text",
    text=df["team_abv"],
    hovertext=df["team_name"],
    textposition="bottom center",
    textfont=dict(size=14, color="black"),
    clipmaxis=False,
    marker=dict(
        size=10,
        color=y,
        colorscale="Blues",
        colorbar=dict(title="Home Field Advantage",
            line=dict(width=1, color="black"))
    )
))
```





Stadium Environment and Home Field Advantage



```
# FIGURE 5: MAP VISUALIZATION WITH TABS

# Calculations
df["home_win_pct"] = df["home_wins"] / (df["home_wins"] + df["home_losses"])
df["away_win_pct"] = df["away_wins"] / (df["away_wins"] + df["away_losses"])
df["home_field_advantage_score"] = df["home_win_pct"] - df["away_win_pct"]
df["attendance_rate"] = df["avg_attendance_home"] / df["seat_capacity"]

# Normalize miles traveled for marker sizing (shared across both tabs)
max_miles = df["miles_traveled"].max()
min_miles = df["miles_traveled"].min()

# Create tabs for two different map views
tab1, tab2 = st.tabs(["🌿 Environmental Factors", "🏟️ Park Factors"])

# ----- TAB 1: ENVIRONMENTAL FACTORS -----
with tab1:
    st.subheader("🌿 Stadium Environment and Home Field Advantage")

    st.markdown("
    **Marker color** = Home Advantage Score (green = positive, red = neutral or negative)**
    **Marker size** = Total Miles Traveled in 2022**")

    with st.expander("Show caption and interpretation"):
        st.markdown("
        This map highlights each stadium's environmental conditions and their home field advantage in 2022.

        ~ **Green markers** = stronger home performance
        ~ **Red markers** = equal or worse performance at home
        ~ **Larger markers** = more miles traveled by the team in 2022
        ~ Hover to explore temperature, elevation, roof usage, day games, and travel distance
        ")

    m_env = folium.Map(location=[39.8283, -98.5795], zoom_start=4, tiles="OpenStreetMap")

    for _, row in df.iterrows():
        lat = row["ballpark_lat"]
        lon = row["ballpark_long"]
        team = row["team_name"]
        stadium = row["ballpark"]
        hfa = row["home_field_advantage_score"]
        temp = row["avg_temp_f"]
        elev = row["elevation_ft"]
        roof = row["roof_pct"]
        day = row["daytime_pct"]
        miles = row["miles_traveled"]
```





Analysis Limitations



Analysis Limitations

- **Scope of 2022 Season**
 - Our analysis is based on data from the 2022 MLB season. Patterns may shift in across multiple seasons due to roster changes, ballpark updates, or new league policies.
 - Future work could expand the dataset to include multiple seasons to identify long-term trends and better understand year-over-year changes.
- **Team-Level Aggregation**
 - The dataset summarizes statistics at the team level, which shows us general trends but limits visibility into game-to-game variation, weather impacts, and which players are injured or sitting out that game.
 - Future analyses could incorporate game-level or player-level data to capture more granular effects and improve the precision of performance insights.
- **Focus on Tangible Factors**
 - Our analysis looks at measurable, game-level variables like location, stadium characteristics, and win/loss records. Psychological and emotional factors—such as morale, momentum, or pressure—are not captured in the available data but remain important context for performance.
 - Future research could explore qualitative data such as player interviews, team morale reports, and social media sentiment to assess the impact of psychological and emotional factors.





Takeaways



Takeaways

- MLB teams perform better at home vs away.
 - 73% of MLB teams had a higher win percentage at home in 2022, confirming a league-wide home field advantage trend.
 - Better pitching performance at home vs away
- Teams with the 3 best and 3 worst home field advantage scores:
 - Best: Tampa Bay Rays (0.198), New York Yankees (0.185), Minnesota Twins/Colorado Rockies (tied at 0.173)
 - Worst: Oakland Athletics (-0.016), Washington Nationals (-0.037), White Sox (-0.086)
- Pitching drives home field advantage.
 - Teams struck out 8.63% more batters and issued 2.8% fewer walks at home, suggesting that pitchers benefit more than hitters from playing at home.
- Traveling fewer than 35,000 miles is ideal:
 - Teams that traveled more than 35,000 miles during the 2022 season averaged a home field advantage score of 0.026, compared to 0.084 for those under 35k — a 223% difference.
- The effects of environmental or park factors on home field advantage vary by team.
 - Higher elevation, large wall dimensions, and higher attendance rates correlate with higher home field advantage scores in some cases — but no single factor explained the trend across the league.





Thank you!

