The Five Pitching Archetypes - A Study on Their Performance in Every MLB Stadium

September 2, 2025

Written by: Emmanuel Viray III

University of California, San Diego - TritonBall Sports Analytics

1 Abstract:

This research study investigates how all 30 Major League Baseball (MLB) stadium environments will impact pitcher performance based on their archetype. By quantifying both the unique environmental variables and structural features of each MLB ballpark across North America, this research aims to classify MLB ballparks and determine how they align with five common pitcher archetypes: Balanced, Power, Groundball Specialist, Finesse, and Veteran. Using Principal Component Analysis (PCA) and KMeans clustering, stadiums and pitchers were projected into the same analytical space. These data science and machine learning engineering techniques enables a visual and quantitative comparison for each archetype in every MLB stadium. Various data visualizations were utilized to measure compatibility, revealing which pitcher types are best and worst suited for each unique ballpark profiles. The findings in this study provide valuable insight for roster optimization, pitcher development, and ballpark-specific game strategy for each MLB franchise.

2 Introduction:

Every MLB stadium possesses unique characteristics that influence gameplay. These factors are contributed to the city's environments all across North America. Some favor power hitters with short fences and thin air, while others reward pitching with deeper outfields and cooler climates. While hitting has long been analyzed in the context of park effects, pitcher performance in relation to ballpark architecture and environmental factors remains underexplored. This research seeks to bridge that gap by examining whether we can quantify the traits of each MLB stadium and determine how the types of pitchers thrive or suffer in each venue.

3 Research Question:

Can we quantify the environmental and structural characteristics of MLB ballparks to identify how the five pitcher archetypes perform in each stadium?

4 Hypothesis:

By accurately characterizing a pitcher's style, it is possible to identify the MLB stadium environments that best align with their strengths, thereby optimizing individual performance that contributes to overall team success.

5 Research Methods:

5.1 MLB Pitcher Data

Pitcher data was compiled into numerous structured CSV files containing performance statistics relevant to pitching style gathered from online sources (Baseball Reference, Baseball Savant, FanGraphs). Two pitchers were chosen from each MLB Franchise during the 2024 season and were split into two PCA graphs divided into each respective league (National League and American League). The initial framework for this study was inspired by a conversation with Collin Irwin, collegiate starting catcher at Western Oregon University. He introduced me to the concept of classifying pitchers into five primary archetypes based on their characteristics, tendencies, and pitch-mix. The WOU Wolves find great offensive success when being able to identify what kind of pitchers they are about to face to better prepare for their match-ups ahead of games.

By using Irwin's classification metric, I then categorized each pitcher into one of the five archetypes based on their statistical profile—Power, Finesse, Groundball Specialist, Balanced/Stock, and Veteran Pitcher. I chose to analyze each pitcher's average fastball velocity, strikeouts per nine innings (K/9), walks per nine innings (BB/9), groundball percentage (GB%), flyball percentage (FB%), home runs allowed per nine innings (HR/9), ERA, and FIP. These metrics were selected for their strong influence on pitcher outcomes and susceptibility to ballpark factors. For instance, HR/9 and FB% are key for assessing a pitcher's vulnerability in hitter-friendly parks, while GB% helps evaluate fit in stadiums with short fences where groundball pitchers are less punished by long balls.

5.2 MLB Stadium Data

To explore how MLB stadium characteristics impact pitcher performance by archetype, this study began with the collection of two core datasets: one representing MLB stadiums and their environmental/structural attributes, and another profiling notable pitchers across five established archetypes. The stadium dataset included variables such as altitude, humidity, wind speed and direction, average summer weather, fence height, and field dimensions (Gathered from NOAA National Center for Environmental Information and BallparkPal). These factors were chosen because they directly influence ball flight and contact outcomes. For example, altitude (like in Coors Field) reduces air resistance and increases ball carry, while fence height and field size affect whether deep fly balls are converted into home runs or outs. Similarly, wind patterns can significantly alter how a ball travels, especially in open-air stadiums.

5.3 Tools and Techniques

In this study, I used a variety of data science techniques to collect, clean, and analyze performance data for MLB pitchers and environmental factors across stadiums. I aggregated pitcher statistics from public sources like Baseball Reference, FanGraphs, and Baseball Savant, and compiled stadium-specific variables such as elevation, temperature, humidity, and wind direction gathered from NOAA and Ballpark Pal. This study was built on a Jupyter Notebook through UC San Diego's

DataHub research server for cognitive science. Using pandas, I cleaned and merged the datasets and created visualizations with matplotlib and seaborn to explore relationships between pitcher archetypes and ballpark conditions. I also employed scikit-learn's PCA to reduce multicollinearity among pitching metrics, allowing clearer interpretation of key performance contributors.

To categorize pitchers into distinct archetypes, I applied unsupervised machine learning techniques, beginning with Principal Component Analysis (PCA) to reduce dimensionality and identify the most influential performance features. PCA helps visualize these relationships in a two-dimensional space without losing significant information from the original dataset.PCA transformed a high-dimensional dataset—consisting of variables like fastball velocity, spin rate, strikeout rate (K/9), walk rate (BB/9), pitch usage percentages, and home run metrics—into two principal components that captured the majority of variance in pitcher profiles. These components helped visualize and interpret statistical tradeoffs (e.g., power vs. command) and provided a foundation for clustering pitchers by style. The resulting biplot clarified which features were most strongly associated with each axis, and allowed for simplified comparisons between pitchers across different roles and skillsets.

Following PCA, I used K-Means Clustering to group pitchers into five distinct archetypes—Power, Finesse, Groundball, Balanced, and Veteran—based on their PCA-transformed statistical vectors. K-Means assigned pitchers to the nearest cluster centroid in multi-dimensional space, allowing for objective classification without the need for labeled data. While PCA and clustering provided the structural foundation, I then applied a rule-based scoring system—informed by environmental and structural ballpark factors—to evaluate how each pitcher archetype matched to MLB stadiums displayed on a heatmap. The same numerical values that were showcased in the heatmap were then put onto a rank-based scatter plot for an alternative visualization to interpret the same statistics in a more digestable manner.

6 Coding Scheme

6.1 MLB Pitcher Archetypes:

6.1.1 Balanced Pitchers

Balanced pitchers maintain a well-rounded mix of pitch types and rely on sequencing and pitch diversity rather than pure velocity. Their fastball velocity tends to hover around league average, and they often throw all five pitches (FB, SL, CH, CB, etc.) with moderate usage. They aim to keep hitters off balance through variation in movement and speed. This archetype represents adaptability and consistency across different game situations.

6.1.2 Power Pitchers

Power pitchers thrive on high velocity and overpowering stuff, typically sitting in the upper 90s with their fastball. Their pitch mix heavily favors fastballs and sliders, maximizing swing-and-miss potential. They generate high strikeout rates (K/9) but may also allow more walks or home runs due to aggressive pitch locations. These pitchers dominate through raw physical tools, often challenging hitters in the zone.

6.1.3 Groundball Specialist

Groundball specialists are designed to minimize damage by inducing weak contact and ground balls with sinkers and two-seamers. They use a pitch mix heavy in low-spin, downward-moving pitches and often pair their sinker with sliders or changeups low in the zone. While they may not strike out many hitters, they excel in keeping the ball in the park and avoiding big innings. Their game plan is built around limiting fly balls and pitching to their infield defense.

6.1.4 Finesse Pitchers

Finesse pitchers excel through command, movement, and deception rather than velocity, often sitting below league average on the radar gun. They typically favor changeups, cutters, and curveballs, aiming to induce weak contact and minimize walks. These pitchers control the strike zone with precision and are highly effective at pitching to contact and staying efficient through innings. Their style emphasizes control and mental approach over physical power.

6.1.5 Veteran Pitchers

Veteran pitchers rely on experience, sequencing, and game awareness rather than elite velocity or pure stuff. Their pitch mix is refined over years of adaptation, often featuring a smart balance of off-speed pitches and location-based strategy. These pitchers know how to exploit hitter weaknesses and manage pitch counts, even if their velocity has declined.

6.2 MLB Pitcher Characteristics:

This table below characterizes the MLB pitching variables used as well as an explaination as to what type of statistic it is and how it is in relation to this research study.

Variable	**Definition**	**Type**	**Measurement**	**Rationale**
`Archetype`	Descriptive label for pitcher's style	Categorical	Archetypes	Used to classify pitching strategies and mechanics
`Avg FB Velocity`	Average 4-seam fastball velocity	Continuous	MPH	Power pitchers throw harder, while finesse pitchers rely less on velocity
`K/9`	Strikeouts per 9 innings	Continuous	Numeric	Reflects a pitcher's ability to miss bats
`BB/9`	Walks per 9 innings	Continuous	Numeric	Measures control and precision
`GB%`	Percentage of batted balls that are grounders	Continuous	Percent	Groundball pitchers thrive in HR-prone parks
`FB%`	Percentage of batted balls that are fly balls	Continuous	Percent	High flyball rates = more HR risk, especially in certain stadiums
`HR/9`	Home runs allowed per 9 innings	Continuous	Numeric	Key indicator of vulnerability to home runs
`ERA`	Earned Run Average	Continuous	Numeric	Reflects run suppression performance
`FIP`	Fielding Independent Pitching	Continuous	Numeric	Filters out defense; isolates pitcher-only results

6.3 MLB Stadium Characteristics:

This table below explains the MLB stadium variables used. I also added an explaination as to what type and measurement each variable is to corelate its signifance to the study.

Variable	Definition	Type	Measurement	Rationale
`Elevation`	Height of stadium above sea level	Continuous	Feet	Affects air density and ball carry — key for flyball/power pitchers
`Avg Wind Speed`	Average in-game wind speed during season	Continuous	MPH	Influences ball flight; can assist or suppress batted balls
`Wind Direction`	Dominant wind direction (from home plate)	Categorical	Outward, Inward, Crosswind	Helps determine park's HR-friendliness
`Avg Temperature`	Mean temperature during games	Continuous	Fahrenheit	Warmer weather = more ball carry; affects air resistance and pitcher grip
`Humidity`	Average in-game humidity percentage	Continuous	%	Higher humidity lowers air density; ball carries farther slightly
`Left Field Distance`	Distance from home plate to LF wall	Continuous	Feet	Structural measure; short dimensions favor hitters
`Center Field Distance`	Distance from home plate to CF wall	Continuous	Feet	Reflects HR suppression potential
`Right Field Distance`	Distance from home plate to RF wall	Continuous	Feet	Indicates asymmetry or handedness advantages
`Fence Height`	Height of the outfield walls	Continuous	Feet	Fences can suppress HRs, especially for flyball pitchers
`HR Factor`	Statcast HR Park Factor	Continuous	Float	Outcome measure — how likely a park turns flyballs into HRs

6.4 Figure (1) MLB Pitchers:

Two starting pitchers were chosen from each MLB franchise. This makes a total sample size of sixty (60) MLB starting pitchers that will be displayed on the pitcher archetype PCA graph. Below, each pitcher is displayed with their 2024 team and statistics including their average fastball velocity, earned run average (ERA), K/9, BB/9, HR/9, Groundball %, and Flyball %.

	Pitcher	FB Velo	ERA	K/9	BB/9	HR/9	GB	FB
0	(ARI) Brandon Phaadt	95	4.71	9.2	2.1	1.2	41.5	28.8
1	(ARI) Ryne Nelson	94	4.24	7.5	2.0	1.0	42.5	24.1
2	(ATL) Chris Sale	94	2.38	11.4	2.0	0.5	45.1	24.0
3	(ATL) Max Fried	95	3.25	8.6	2.9	0.7	58.2	16.1
4	(BAL) Corbin Burnes	96	2.92	8.4	2.2	2.0	47.9	22.4
5	(BAL) Albert Suarez	93	3.70	7.3	2.9	1.1	34.1	31.6
6	(BOS) Kutter Crawford	94	4.36	8.6	2.5	1.7	36.2	32.3
7	(BOS) Tanner Houck	95	3.12	7.8	2.4	0.6	55.2	17.4
8	(CHC) Shota Imanaga	92	2.91	9.0	1.5	1.4	36.6	32.9
9	(CHC) Jameson Taillon	93	3.27	6.8	1.8	1.1	39.4	30.7
10	(CHW) Chris Flexer	92	4.95	6.9	3.5	1.4	35.5	27.6
11	(CHW) Garrett Crochet	97	3.58	12.9	2.0	1.1	44.5	25.9
12	(CIN) Hunter Greene	99	2.75	10.1	3.4	0.7	33.9	31.7
13	(CIN) Nick Martinez	92	3.72	7.4	3.4	1.6	36.5	27.9
14	(CLE) Tanner Bibee	94	3.47	9.7	2.3	1.1	34.6	31.6
15	(CLE) Ben Lively	90	3.81	7.0	2.9	1.4	42.2	28.3
16	(COL) Austin Gomber	90	4.75	6.3	2.1	1.6	37.5	28.7
17	(COL) Ryan Feltner	95	4.49	7.7	2.9	1.1	43.8	25.7
18	(DET) Tarik Skubal	94	2.39	10.7	1.6	0.7	45.5	24.7
19	(DET) Reese Olson	94	3.53	8.1	2.6	0.6	50.5	19.0
20	(HOU) Framer Valdez	93	2.91	8.6	2.8	0.7	59.8	15.0
21	(HOU) Hunter Brown	97	3.49	9.5	3.2	1.0	48.3	22.0
22	(KC) Seth Lugo	92	3.00	7.9	2.1	7.7	43.6	25.8
23	(KC) Cole Ragans	95	3.14	10.8	3.2	0.7	39.9	27.3
24	(LAA) Tyler Anderson	90	3.81	7.1	3.7	1.2	37.8	29.5
25	(LAA) Griffin Canning	94	5.19	6.8	3.5	1.6	40.6	30.0
26	(LAD) Tyler Glasnow	97	3.49	11.3	2.4	1.0	47.6	23.7
27	(LAD) Yoshinobu Yamamoto	95	3.00	10.5	2.2	0.7	47.5	20.0
28	(MIA) Trevor Rogers	93	4.53	7.3	3.9	1.0	46.7	25.4
29	(MIA) Edward Cabrera	96	4.95	10.0	4.7	1.4	46.9	21.9
30	(MIL) Freddie Peralta	95	3.68	10.4	3.5	1.3	36.2	29.5
31	(MIL) Colin Rea	93	4.29	7.2	2.3	1.6	38.6	30.4
32	(MIN) Pablo Lopez	94	4.08	9.6	2.0	1.3	43.4	26.6
33	(MIN) Bailey Ober	91	3.98	9.6	2.2	1.4	33.5	35.2
34	(NYM) Luis Severino	97	3.91	8.0	3.0	1.1	45.2	25.2
35	(NYM) Sean Manaea	92	3.47	9.1	3.1	1.0	37.0	29.1
36	(NYY) Carlos Rodon	94	3.96	10.0	2.9	1.6	33.8	35.4
37	(NYY) Nestor Cortes	92	3.77	8.4	2.0	1.2	30.7	32.1
38	(PHI) Zack Wheeler	96	2.57	10.1	2.3	0.9	42.5	26.0
39	(PHI) Aaron Nola	91	3.57	8.9	2.3	1.4	44.2	24.4

40	(PIT) Paul Skenes	100	1.96	11.5	2.2	0.7	51.6	20.3
41	(PIT) Mitch Keller	95	4.25	8.4	2.5	1.2	39.4	26.3
42	(OAK) JP Sears	92	4.38	6.8	2.4	1.4	36.8	32.0
43	(OAK) Mitch Spence	92	4.58	7.5	2.6	1.2	48.2	23.8
44	(SD) Dylan Cease	97	3.47	10.6	3.1	0.9	39.4	30.7
45	(SD) Michael King	94	2.95	10.4	3.3	0.9	40.0	23.7
46	(SF) Logan Webb	94	3.47	7.6	2.2	0.5	57.1	17.3
47	(SF) Blake Snell	93	3.12	12.5	3.8	0.5	41.7	26.8
48	(SEA) Logan Gilbert	95	3.23	9.5	1.6	1.1	44.1	24.4
49	(SEA) Bryce Miller	94	2.94	8.5	2.2	1.0	38.4	33.0
50	(STL) Sonny Gray	92	3.84	11.0	2.1	1.1	42.1	24.7
51	(STL) Andre Pallante	95	3.78	7.0	3.6	0.6	61.4	16.2
52	(TB) Zack Littell	92	3.62	8.1	1.8	1.3	33.7	27.8
53	(TB) Taj Bradley	95	4.11	10.0	3.1	1.4	41.0	27.9
54	(TEX) Nathan Eovaldi	94	3.80	8.8	2.2	1.2	48.4	24.5
55	(TEX) Andrew Heaney	92	4.28	8.9	2.3	1.3	33.5	30.6
56	(TOR) Jose Berrios	94	3.60	7.2	2.5	1.5	42.7	26.0
57	(TOR) Kevin Gausman	94	3.83	8.1	2.8	1.0	39.1	30.1
58	(WAS) Jake Irvin	92	4.41	7.5	2.5	1.4	44.0	25.5
59	(WAS) MacKenzie Gore	96	3.90	9.8	3.5	0.8	39.1	25.4

6.5 Figure (2) MLB Stadiums City Enviornment Factors:

All 30 MLB stadiums are displayed with their city specific environmental factors below:

	(Team) MLB Stadium	City, State	Altitude	Temperature	\
0	(ARI) Chase Field	Phoenix, AZ	1100	99	
1	(ATL) Truist Park	Atlanta, GA	304	82	
2	(BAL) Camden Yards	Baltimore, MD	20	87	
3	(BOS) Fenway Park	Boston, MA	20	87	
4	(CHC) Wrigley Field	Chicago, IL	600	80	
5	(CHW) Guaranteed Rate Field	Chicago, IL	594	83	
6	(CIN) Great American Ballpark	Cincinnati, OH	489	86	
7	(CLE) Progressive Field	Cleveland, OH	600	80	
8	(COL) Coors Field	Denver, CO	5200	88	
9	(DET) Comerica Park	Detroit, MI	600	83	
10	(HOU) Minute Maid Park	Houston, TX	40	90	
11	(KC) Kauffman Stadium	Kansas City, MO	900	88	
12	(LAA) Angel Stadium	Anaheim, CA	155	71	
13	(LAD) Dodgers Stadium	Los Angeles, CA	502	85	
14	(MIA) LoanDepot Park	Miami, FL	20	85	
15	(MIL) America Family Field	Milwaukee, WI	597	80	
16	(MIN) Target Field	Minneapolis, MN	840	80	
17	(NYM) City Field	New York, NY	46	85	
18	(NYY) Yankee Stadium	New York, NY	55	85	
19	(PHI) Citizens Bank Park	Philadelphia, PA	30	86	
20	(PIT) PNC Park	Pittsburgh, PA	700	84	

21	(OAK) Oakland Coliseum	Oakland, CA	75	75
22	(SD) Petco Park	San Diego, CA	16	72
23	(SF) Oracle Park	San Francisco, CA	52	65
24	(SEA) T-Mobile Park	Seattle, WA	10	63
25	(STL) Busch Stadium	St. Louis, MO	436	88
26	(TB) Tropicana Field	St. Petersburg, FL	32	87
27	(TEX) Globe Life Stadium	Arlington, TX	278	96
28	(TOR) Rogers Centre	Toronto, ON	251	75
29	(WAS) Nationals Park	Washington, DC	19	90

	Humidity	Wind Speed (MPH)	Wind Direction (From Home Plate)
0	30	5	Outward
1	67	7	Crosswind
2	67	8	Outward
3	63	5	Outward
4	60	9	Inward
5	70	12	Outward
6	75	5	Outward
7	55	8	Outward
8	31	8	Outward
9	68	10	Crosswind
10	80	6	Neutral
11	70	10	Crosswind
12	65	7	Crosswind
13	50	6	Outward
14	80	8	Inward
15	72	10	Crosswind
16	65	10	Crosswind
17	65	6	Outward
18	68	3	Outward
19	67	9	Outward
20	70	7	Crosswind
21	71	10	Crosswind
22	74	6	Crosswind
23	80	10	Inward
24	66	7	Crosswind
25	67	5	Inward
26	75	7	Inward
27	57	8	Crosswind
28	65	8	Crosswind
29	72	8	Outward

6.6 Figure (3) MLB Stadiums Structural Factors:

All 30 MLB stadiums are displayed with their unique structural factors below measured in feet:

	(Team) MLB Stadium	Left Field	Center Field	Right Field	\
0	(ARI) Chase Field	330	407	335	
1	(ATL) Truist Park	335	400	330	
2	(BAL) Camden Yards	333	400	318	
3	(BOS) Fenway Park	310	390	302	
4	(CHC) Wrigley Field	355	400	353	
5	(CHW) Guaranteed Rate Field	330	400	335	
6	(CIN) Great American Ballpark	328	404	325	
7	(CLE) Progressive Field	325	410	325	
8	(COL) Coors Field	347	415	350	
9	(DET) Comerica Park	345	420	330	
10	(HOU) Minute Maid Park	315	406	326	
11	(KC) Kauffman Stadium	330	410	330	
12	(LAA) Angel Stadium	330	400	330	
13	(LAD) Dodgers Stadium	330	395	330	
14	(MIA) LoanDepot Park	344	407	335	
15	(MIL) America Family Field	344	400	345	
16	(MIN) Target Field	339	404	328	
17	(NYM) City Field	335	408	330	
18	(NYY) Yankee Stadium	318	408	314	
19	(PHI) Citizens Bank Park	329	401	330	
20	(PIT) PNC Park	325	399	320	
21	(OAK) Oakland Coliseum	330	400	330	
22	(SD) Petco Park	334	396	322	
23	(SF) Oracle Park	339	399	309	
24	(SEA) T-Mobile Park	331	401	326	
25	(STL) Busch Stadium	336	400	335	
26	(TB) Tropicana Field	315	404	322	
27	(TEX) Globe Life Stadium	332	400	325	
28	(TOR) Rogers Centre	328	400	328	
29	(WAS) Nationals Park	336	402	335	

	LF Wall	CF Wall	RF Wall
0	7.5	25.0	7.5
1	11.0	15.0	15.0
2	7.0	7.0	21.0
3	37.0	18.0	3.0
4	15.0	11.5	15.0
5	8.0	8.0	8.0
6	12.0	12.0	8.0
7	19.0	9.0	9.0
8	8.0	8.0	17.0
9	7.0	9.0	8.0
10	21.0	8.0	7.0

11	9.0	9.0	9.0
12	5.0	8.0	8.0
13	8.0	8.0	8.0
14	9.0	12.0	9.0
15	8.0	8.0	8.0
16	8.0	8.0	23.0
17	8.0	8.0	8.0
18	8.0	8.0	8.0
19	13.0	6.0	13.0
20	6.0	8.0	21.0
21	8.0	8.0	8.0
22	7.0	7.0	10.0
23	8.0	8.0	25.0
24	7.0	7.0	15.0
25	8.0	8.0	8.0
26	11.0	9.0	11.0
27	14.0	8.0	8.0
28	8.0	8.0	8.0
29	8.0	8.0	12.0

6.7 Figure (4) Home Runs Index in Each MLB Stadium:

Below shows a list of each MLB ballpark and a home run index score from the 2020-2024 MLB regular seasons. (100 = Average)

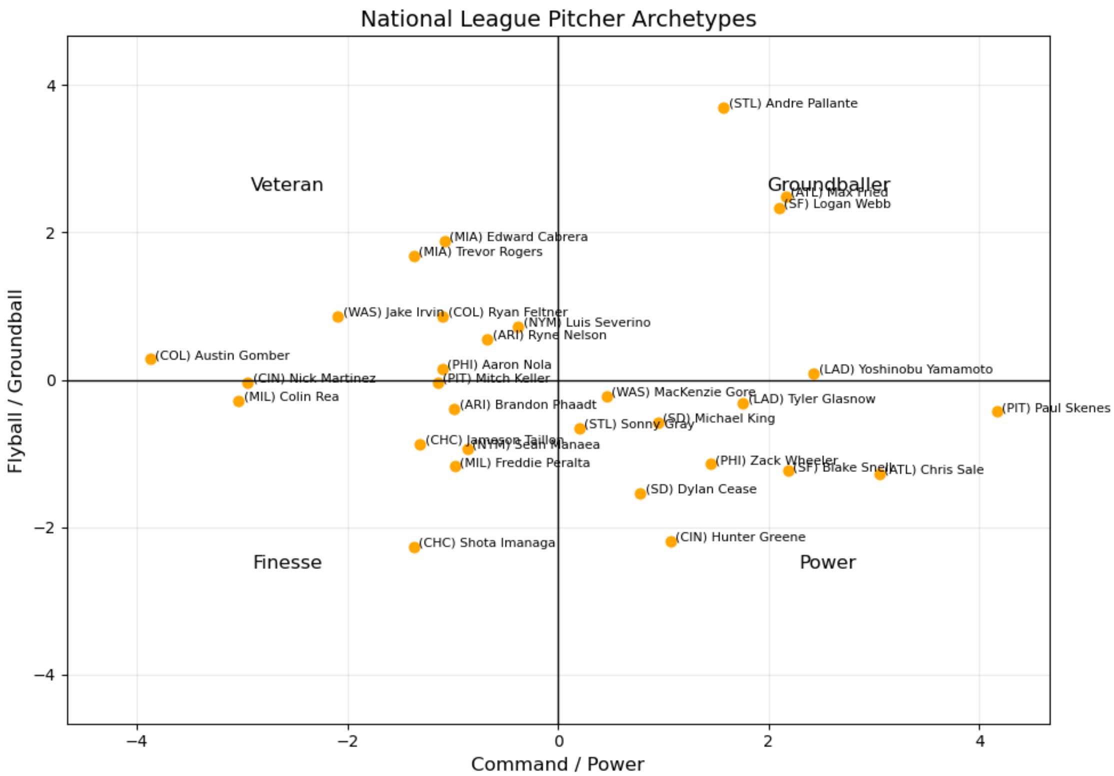
	(Team) MLB Stadium	2020	2021	2022	2023	2024
0	(ARI) Chase Field	94	86	82	79	92
1	(ATL) Truist Park	95	107	102	119	88
2	(BAL) Camden Yards	95	139	82	91	104
3	(BOS) Fenway Park	96	94	108	92	90
4	(CHC) Wrigley Field	91	124	90	105	85
5	(CHW) Guaranteed Rate Field	114	114	109	98	93
6	(CIN) Great American Ballpark	154	130	131	127	124
7	(CLE) Progressive Field	90	110	87	67	106
8	(COL) Coors Field	100	105	121	103	105
9	(DET) Comerica Park	97	72	74	90	92
10	(HOU) Minute Maid Park	81	99	112	92	116
11	(KC) Kauffman Stadium	92	76	80	95	82
12	(LAA) Angel Stadium	119	107	118	115	107
13	(LAD) Dodger Stadium	128	130	118	122	121
14	(MIA) LoanDepot Park	86	78	91	87	95
15	(MIL) American Family Field	108	96	121	107	111
16	(MIN) Target Field	86	110	97	107	113
17	(NYM) Citi Field	116	92	85	104	108
18	(NYY) Yankee Stadium	124	110	113	123	122
19	(PHI) Citizens Bank Park	117	104	113	114	114
20	(PIT) PNC Park	89	78	92	71	90
21	(OAK) Oakland Coliseum	83	86	76	88	82
22	(SD) Petco Park	112	94	95	89	123
23	(SF) Oracle Park	97	85	74	89	72
24	(SEA) T-Mobile Park	100	96	108	94	89
25	(STL) Busch Stadium	90	77	92	98	83
26	(TB) Tropicana Field	82	83	91	103	99
27	(TEX) Globe Life Field	94	96	113	133	105
28	(TOR) Rogers Centre	100	101	117	95	100
29	(WAS) Nationals Park	93	111	113	98	93

7 MLB Pitcher Archetypes:

Below shows Figures (5) and (6), a graph that presents a KMeans clustering analysis of MLB pitchers based on Principal Component Analysis (PCA) of their pitching metrics. Using the pitching statistics table in Figure (1), PCA was applied to reduce the dimensionality of features, capturing the most meaningful variance in just two components.

7.1 Figure (5) National League Pitchers Archetype:

30 MLB pitchers from the 15 National League baseball teams are characterized on the Kmeans clustering graph below.



7.2 Analysis:

The X-axis (Command/Power) separates pitchers by their ability to dominate hitters via velocity and strikeouts versus those with control and finesse. The positive side of the X-axis reflects a higher velocity fastball, while the negative side of the X-axis emphasizes command and pitch-mix over power. The Y-axis reflects how well pitchers limit quality contact. A higher value in the Y-axis represents better contact suppression leading to more groundballs paired with a diverse pitch selection. A lower value on the Y-axis indicates the pitcher is susceptible to fly balls and home runs. Each quadrant is represented with the archetype that best categorizes the pitcher's style. The origin indicates the balanced cluster pitchers.

7.3 Focus Pitchers:

7.3.1 (46) Logan Webb: Groundball Specialist

Ranks high in the Contact Management & Pitch Mix dimension, indicating he is an elite groundball pitcher who excels at limiting hard contact and keeping the ball in the park. His arsenal is built around a heavy sinker and changeup combination that induces weak contact early in the count. His far-right position on the Command/Power axis also suggests he pairs movement and pitch sequencing with an above-average fastball velocity (94). This groundball-heavy profile makes him ideal for parks with smaller dimensions or high home run tendencies.

7.3.2 (27) Yoshinobu Yamamoto: Power/Balanced Hybrid

While still a groundball-focused pitcher, sits closer to the center of the PCA plot, indicating a more balanced profile that blends solid contact suppression with moderate power and command. His pitch mix features a mid-90s fastball, devastating splitter, and precise curveball, giving him tools from both power and finesse profiles. His positioning reflects a versatile arsenal and adaptability, making him effective in various pitching environments. This hybrid skillset allows him to adjust to different teams, stadiums, and matchups with consistency.

7.3.3 (8) Shota Imanaga: Finesse

Relies more on command and control rather than overpowering velocity, which aligns with a finesse pitching style focused on precision and pitch movement. His fastball plays up due to extension and deception, and is complemented by a splitter and slider that keep hitters off balance. His location lower on the y-axis suggests a tendency toward allowing more fly balls than ground balls.

7.3.4 (2) Chris Sale: Power

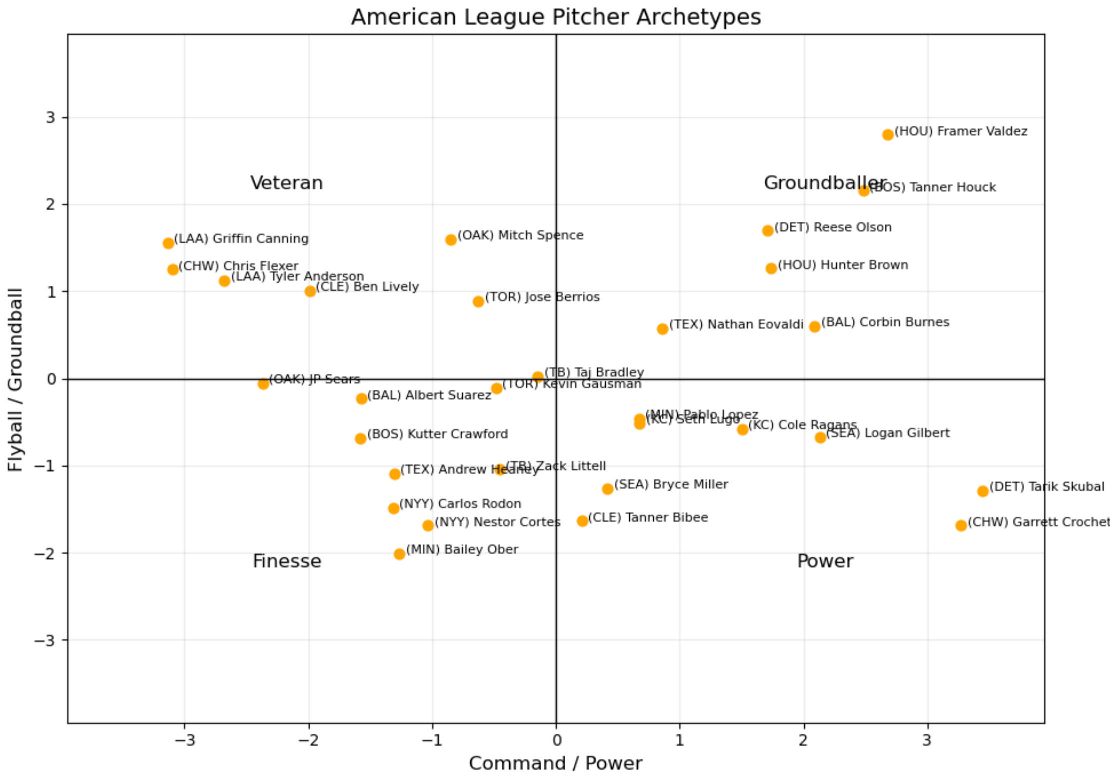
His elite fastball velocity and dominant slider, both of which generate high whiff rates, are the foundation of his power pitcher identity. His release mechanics are deceptive and aggressive, giving hitters minimal time to react. He consistently attacks hitters in the zone with high-velocity pitches and swing-and-miss stuff. These traits place him firmly in the power quadrant, even as he adjusts with age and injury history.

7.3.5 (16) Austin Gromber: Veteran

He relies more on experience, command, and pitch sequencing than on overpowering velocity or elite spin. His fastball velocity is below league average, but he uses a mix of sliders, curveballs, and changeups to keep hitters off balance and induce soft contact. Gomber's approach reflects a pitcher who adjusts his game based on matchup awareness and location precision rather than pure stuff, making him a strategic, innings-eating presence in a rotation.

7.4 Figure (6) American League Pitcher Archetypes:

30 MLB pitchers from the 15 American League baseball teams are characterized on the Kmeans clustering graph below.



7.5 Focus Pitchers

7.5.1 (43) Mitch Spence: Veteran

Strategic pitch usage, pitch-to-contact style, and reliance on sequencing rather than velocity. While his fastball velocity is modest, he mixes in cutters, changeups, and breaking balls with confidence and maturity beyond his experience level. He shows poise on the mound, working efficiently through innings by staying ahead in counts and minimizing walks. Spence's approach reflects a control-oriented style typical of veteran archetypes who rely on feel and execution.

7.5.2 (18) Tarik Skubal: Power

He combines upper-90s fastball velocity with elite strikeout ability and aggressive pitch execution. His arsenal includes a high-velocity four-seamer and a sharp slider, both designed to generate swings and misses at the top and bottom of the zone. Skubal's high K/9 rates and ability to challenge hitters with power pitches make him a clear example of this archetype.

7.5.3 (14) Tanner Bibee: Balanced

He combines solid fastball velocity with a well-rounded pitch mix and consistent command. He throws a mid-90s fastball, slider, changeup, and curveball with relatively even usage, keeping hitters guessing without leaning too heavily on any single pitch. Bibee generates both strikeouts and weak contact, showing the ability to adapt his approach depending on the situation. His blend of power, finesse, and control makes him a textbook example of a balanced pitcher.

7.5.4 (20) Framber Valdez: Groundball Specialist

Relies heavily on his sinker and curveball, both of which generate extreme downward movement and induce weak contact. He consistently leads the league in groundball rate, often exceeding 60%, by keeping the ball low in the strike zone and avoiding fly balls. His pitch mix and strategy are designed to minimize home run risk and rely on his infield defense to convert outs efficiently.

7.5.5 (33) Bailey Ober: Finesse

With exceptional command, low walk rates, and precise pitch placement rather than overpowering velocity, he embodies the finesse archetype. His fastball sits in the low 90s, but he pairs it with a well-located changeup and slider to disrupt timing and induce weak contact. Ober consistently works the edges of the strike zone and limits damage by avoiding mistakes over the plate.

7.6 Figure (7) Pitcher Archetype Averages:

By identifying all the starting pitchers into the 5 pitching archetypes in Figures (5) and (6), It allowed me to manually categorize all sixty (60) MLB starting pitchers into their own respective archetype. I then calculated the all the average statistics for each pitcher archetype. This data will be useful later on in the study.

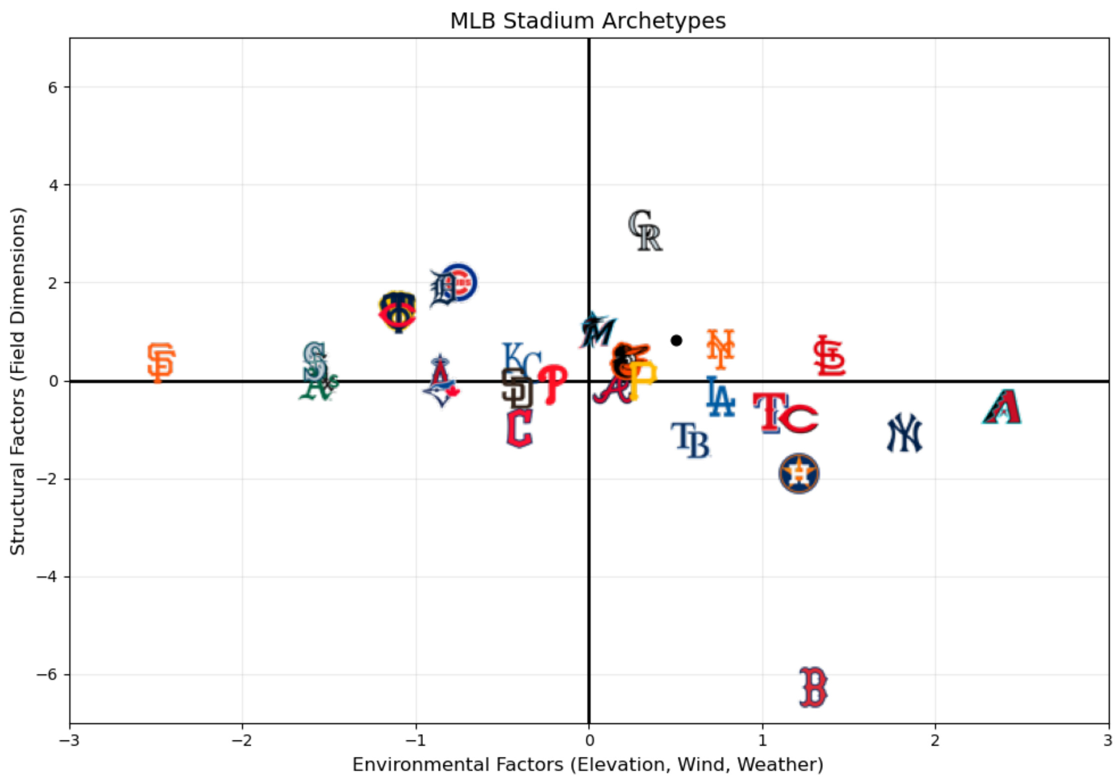
	Archetype	FB Velo	ERA	K/9	BB/9	HR/9	FIP	GB	FB
0	Balanced	94.88	3.84	8.74	2.66	1.42	3.88	40.6	27.3
1	Power	95.85	2.92	10.80	2.50	0.81	2.86	43.2	26.0
2	Groundballer	94.88	3.31	8.20	2.74	0.84	3.36	54.8	18.2
3	Finesse	92.33	3.87	8.85	2.20	1.40	4.10	34.4	32.8
4	Veteran	92.20	4.40	7.32	3.10	1.40	4.73	40.8	27.4

8 MLB Stadium Characteristics:

This scatter plot categorizes MLB stadiums using key environmental and structural factors. I used K-means clustering and PCA like in the pitcher archetypes which allows for a simple visualization despite the use of so many intersecting variables. The variables used include elevation, average summer temperature, humidity, wind speed and direction, and outfield dimensions such as fence height and field size.

8.1 Figure (8) MLB Stadium Archetypes:

This Kmeans clustering graph below will illustrate where each MLB stadium fairs based on their structural and environmental charactristics that was shown in Figures (2) and (3).



8.2 Analysis:

Figure (8) above categorizes the different types of MLB stadiums across the league.

X-axis: City-specific factors for the each stadium that takes humidity, temperature, and elevation into consideration. Left side of the axis represents colder average summertime weather and lower elevation. Right side displays hotter weather and higher city elevation.

Y-axis: Stadium specific ballpark factors that displays wall height and venue dimensions. Lower on the Y-axis shows a smaller field with lower fence sizes. Higher on the Y-axis represents larger MLB stadiums with higher fence sizes.

8.3 Focus Stadiums:

8.3.1 (SF) Oracle Park

San Francisco has a summertime weather average of 65° F and an elevation of 55ft which both rank amongst the lowest across all MLB stadiums. This is why it would be placed far left in the X-axis completely opposite to (AZ) Chase Field which has the highest average summertime temperature of 99° F and an elevation of 1100ft. However the field dimensions are about average, a little bigger than most ballparks so it sits postively, close to zero (0) on the Y-axis. This stadium greatly favors pitching with its larger dimensions, windy and colder conditions that suppresses ball flight.

8.3.2 (SD) Petco Park

Your typical MLB stadium is clustered closely towards the origin. San Diego is known for its amazing weather year round and Petco Park has about average built dimensions (330 ft LF/RF, 400 CF) that justifies why it sits in the middle. Stadiums that are also clustered around the center don't give a noticable advantage towards pitching or hitting.

8.3.3 (NY) Yankee Stadium

Yankee Stadium has one of the smallest fields in the MLB which is shown by it sitting on the negative side of the Y-Axis. With an average summertime weather of 85° F, the hotter climate gives baseballs longer airtime which grants hitting in New York an advantage over pitchers. This hitting advantage goes for the other MLB venues that also sits on the postive side of the X-Axis.

8.3.4 (COL) Coors Field

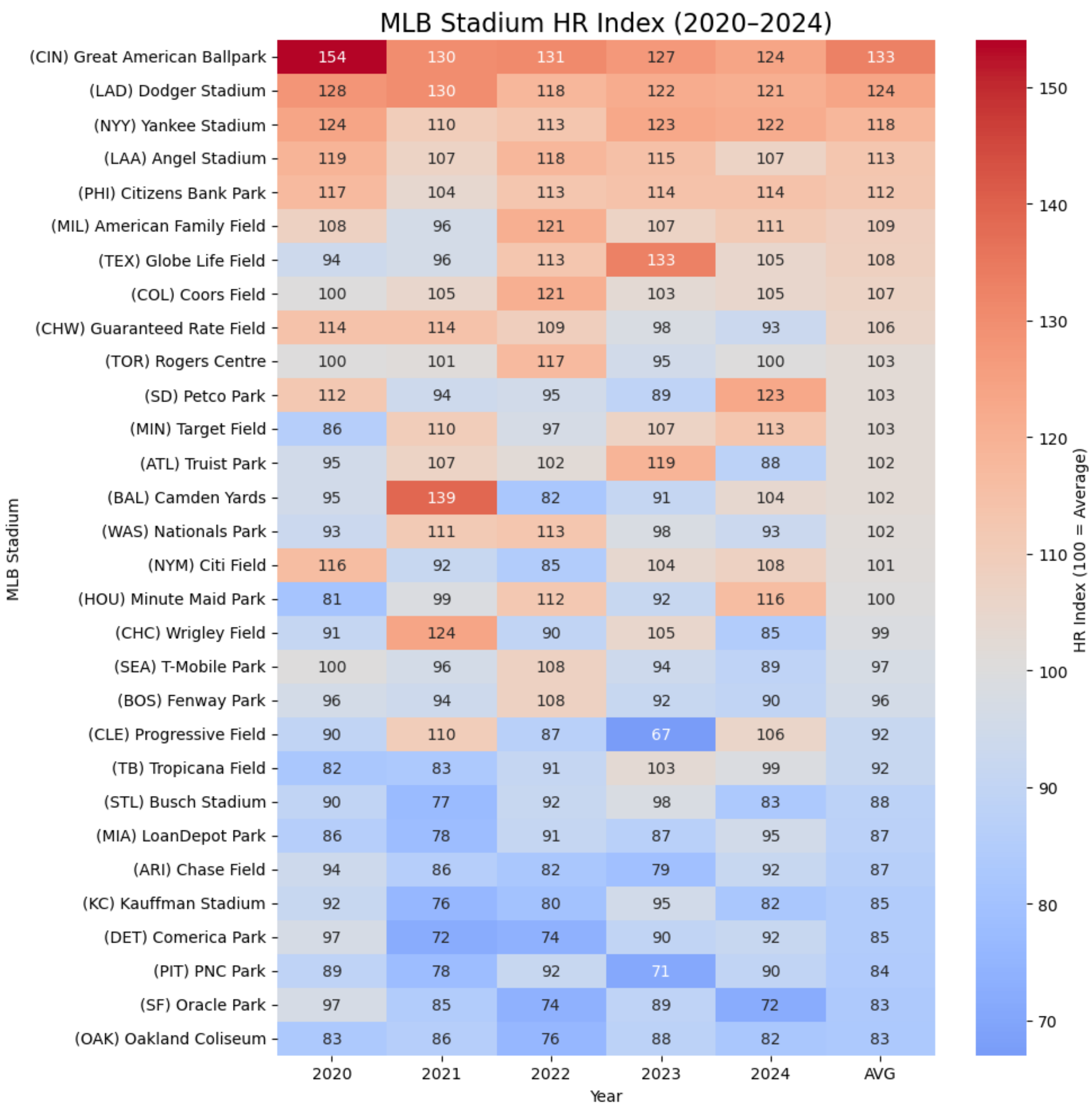
Coors Field is notorious for it being a hitter-friendly ballpark due to its high elevation (5200 ft) and short left-field fence height. However it is counteracted with it having one of the largest field dimensions (Largest: Wrigley Field) which is why its high elevation, warm weather (88°F) and large field dimensions has it rightfully sitting on the top-right quadrant of the graph.

8.3.5 (BOS) Fenway Park

Fenway is the most notable outlier displayed on this graph. Rightfully so, its unconventional field dimensions and the green monster (37 ft left field wall) makes it the true unicorn stadium of the MLB. I would say this is the toughest stadium to determine which side of the ball has an advantage due to its unqiue characteristics. However its uniqueness allows hitters to find gaps easily and the short left field porch and dimensions (302 ft RF, 3 ft wall) allows left handed hitters to thrive.

9 Figure (9) MLB Stadium Home Run Index (2020–2024):

The heat map below ranks the 30 MLB stadiums by the home run scoring index in past five MLB regular seasons (2020-2024) from least to greatest average score.



9.1 Analysis:

This data was gathered from baseballsavant.com using its statcast metrics. Statcast park effects show the observed effect of each displayed stat based on the events in the selected park. This heatmap visualizes the Home Run (HR) Index for every MLB stadium from 2020 to 2024, where a score of 100 represents league average. For example, the 154 HR mark for 2020 at Great American Ball Park means 54% more home runs were observed in this MLB stadium. Higher values (red) indicate parks that are more home run–friendly, while lower values (blue) represent parks that are more pitcher-friendly, suppressing home runs. By displaying data across multiple seasons, the chart highlights both consistency and variation in stadium environments, allowing us to identify which ballparks consistently inflate or suppress long balls.

9.2 Focus Stadiums:

9.2.1 (LAD) Dodger Stadium:

Dodger Stadium consistently shows high HR index values, often above 120, making it one of the most home run–friendly stadiums in MLB. The warm Southern California climate and favorable carry in the air give hitters a clear advantage. This makes it a dangerous park for pitchers, especially fly-ball pitchers who are more vulnerable to home runs. The Dodgers usually have some of the best offensive lineups in the MLB every year which may skew this graph. However the Dodgers organization may just be playing into their home field hitting advantage.

9.2.2 (ARI) Chase Field:

Even though Chase Field is located in Arizona's hot desert climate with thin, dry air that should boost ball flight, its HR Index of 87. Chase Field has deep outfield dimensions (407 feet to center, 334 to left, 335 to right), making it harder for hitters to clear the fences compared to smaller parks. Also the retractable roof keeps the ballpark climate controlled, preventing the extreme heat from exaggerating ball flight.

9.2.3 (LAA) Angel Stadium:

Angel Stadium hovers close to league average but trends slightly red, suggesting it leans hitter-friendly. The warm summer weather in Anaheim helps the ball carry, especially in night games when the air is still. While not extreme, it provides hitters with a modest advantage over pitchers.

9.2.4 (DET) Comerica Park:

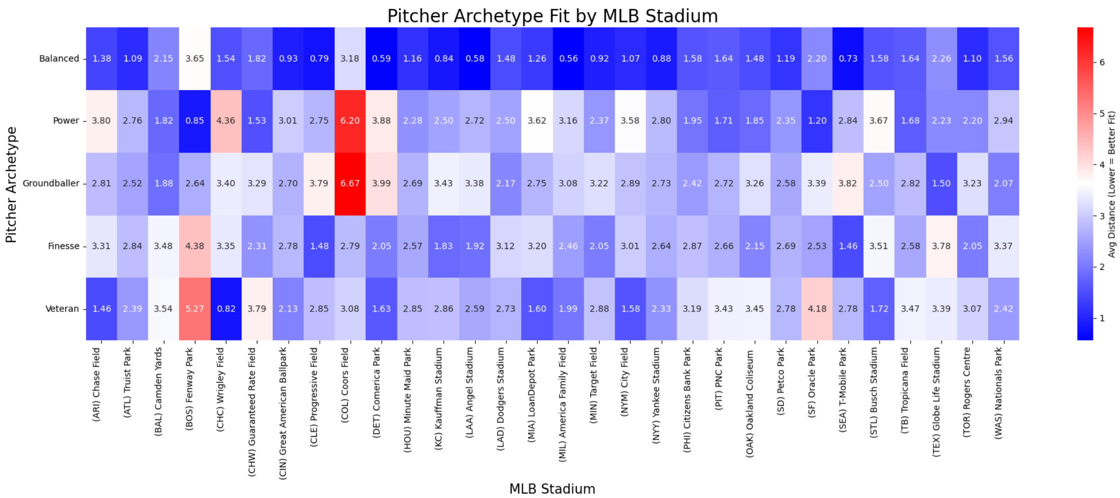
Comerica Park has HR indexes in the low 90s, categorizing it as a pitcher-friendly stadium. Its deep outfield dimensions, especially in center field, make it difficult for hitters to clear the fences. This environment often benefits pitchers and results in fewer home runs compared to league average parks.

9.2.5 (CHC) Wrigley Field:

Wrigley Field is one of the most unique ballparks in MLB because its HR index fluctuates heavily depending on wind conditions. On windy days blowing out, it becomes one of the most hitter-friendly parks, while wind blowing in makes it strongly pitcher-friendly. Overall, its averages place it near the middle, but the variability makes it unpredictable for both pitchers and hitters.

10 Figure (10) Pitcher Archetype Fit Across MLB Stadiums:

Using all the data previously displayed, we are now able to find out which MLB stadium best and worst fits the five pitcher archetypes. This PCA Euclidean Distance-Based Heatmap visualizes how well each of the five pitcher archetypes statistically aligns with the structural and environmental conditions of every MLB stadium. All the figures and pitcher statistics displayed previously in this study will play a factor in the heatmap below.



10.1 Graph Analysis:

Now that we have assessed all the necessary data needed, we have finally made it to heart of the study. This heatmap presents a comprehensive breakdown of how well each pitcher archetype aligns with various MLB stadiums based on proximity in PCA space. The rows represent the five pitcher archetypes while the columns represent each unique individual MLB stadium. The average stats for each archetype calculated in Figure (7) were used along with the city and stadium charactistics displayed on Figures (2) and (3) to build this PCA heatmap. The numerical values within each cell reflect the average distance between a pitcher archetype and the corresponding stadium's profile in a shared PCA space.

10.1.1 PCA Average Distance:

Blue, smaller average euclidean distance = better fit. A lower value (e.g., 0.58 - Balanced Pitchers in Angel Stadium) indicates that the characteristics of a particular pitcher archetype closely align with the environmental and structural features of that ballpark (like dimensions, elevation, wind, etc.).

Red, larger average euclidean distance = poor fit. A higher value (e.g., 6.67 - Groundballers in Coors Field) means the pitcher's profile and the stadium's conditions diverge significantly, implying the pitcher archetype may struggle in that environment.

10.2 Archetype Analysis:

We are now able to give a numeric measurement on how each of the five pitcher archetypes fair within each unique stadium environment. The colors on the heatmap make it easy to identify how well they would perform in each stadium. This would give us an accurate depiction on what archetype would dominate in a particular stadium and what archetype would get punished.

10.2.1 Balanced Pitchers

This archetype shows that they generate the best and most consistent results across all fields. They have the lowest fit score across the board with a 0.56 in the Milwalkee Brewers' home field America Family Field. They also have the lowest high fit score at 3.65 in Fenway Park, with no red fit scores across the league. There is no way for me to identify why this is the case with complete accuracy. However what I think is this archetype embodies every single redeeming quality that every other archetype has, but at an average and median level. This allows the archetype to perform with consistent results across all MLB stadiums.

10.2.2 Power Pitchers

The Power pitcher archetype shows a wide variance with a low of 0.85 and a high of 6.20. I would say this archetype is a double-edged sword. Pitchers who are categorized in this archetype have the ability to overpower pitchers with their high fastball velocity. Power pitchers with great command will dominate. However if a hitter makes solid contact, they are more likely to give up hard hit balls and home runs. So in smaller venues and hitter friendly stadiums, they are prone to get punished for their strengths.

10.2.3 Groundballer

Groundballers are great at keeping the baseball in the field. So in smaller MLB stadiums, they are able to thrive better than other archetypes simply because of their ability at suppressing fly balls and home runs. However they noticably have the highest fit score out of any other archetype with a 6.67 in Coors Field. This is due to the dimensions of Coors Field being so big that this this archetypes strengths would not be projected as strongly as other archetypes. But in a small stadium like Fenway Park or Camden Yards, their strengths are showcased at a noticable degree.

10.2.4 Finesse

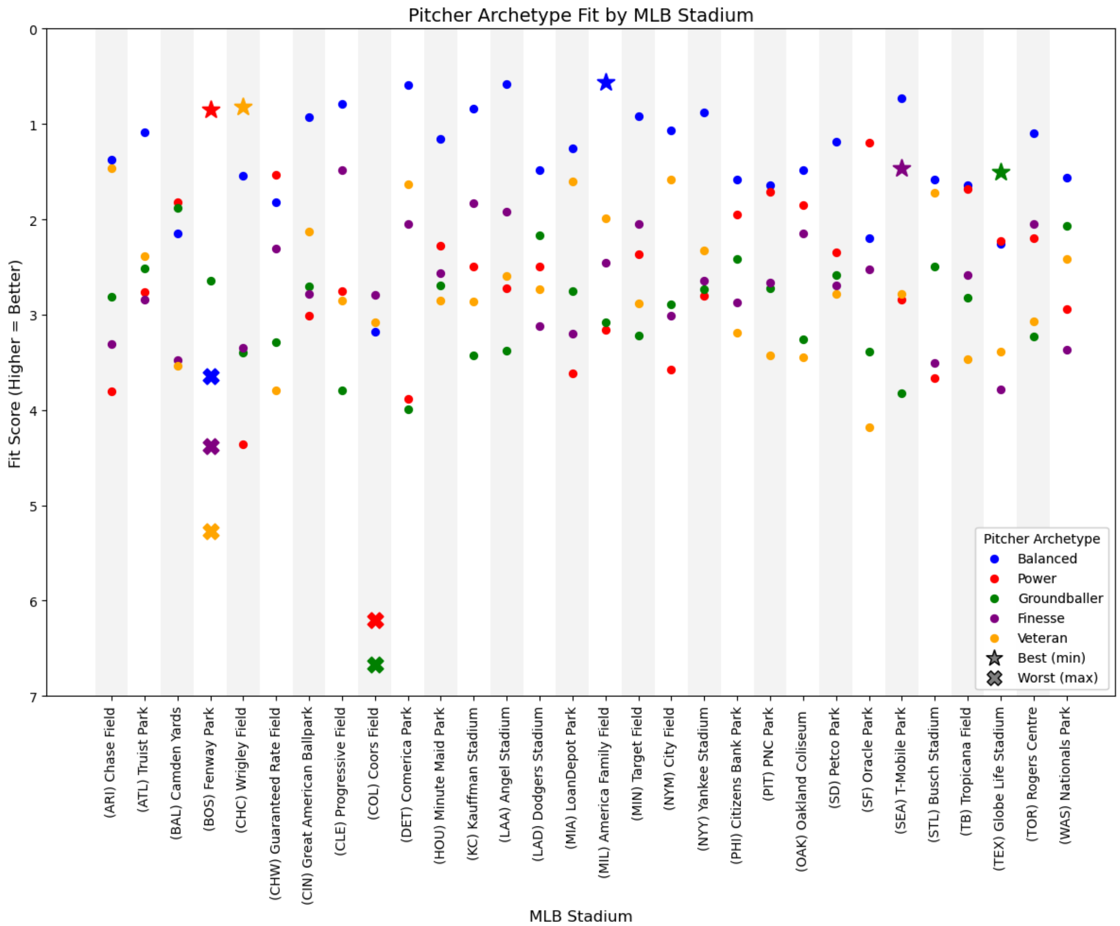
Finesse pitchers are in the opposite spectrum as power pitchers. Their high pitch-mix and reliability on offspead pitches gives them a wide variety in their arsenal. However a low sample size and graph variance makes this archetype consistent. The heatmap above shows that finesse pitchers are simply a safe option. It doesn't show signs of having a signficant advantage or disadvantage compared to the other archetypes, so their performance in each stadium should not include any shock value.

10.2.5 Veteran

Veterans capitolize on their experience and command over velocity and spin rate. The heatmap shows they have a larger variance with a low of 0.85 in Wrigley Field and a high of 5.27 in Fenway Park making this archetype pretty difficult to measure. Same as Fenway Park being the wildcard baseball stadium, I would classify the veterans as the wildcard pitching archetype. So depending on the stadium, this archetype's performance greatly varies.

11 Figure (11) Pitcher Fit Scatter Plot:

A scatter plot displaying the pitcher archetypes fit in each MLB stadium was added to show a different visualization and interpretation of the heatmap of Figure (10) to illustrate the same numerical measurements.



11.1 Analysis:

Using the numbers displayed in Figure (10), each archetype is plotted by a different colored dot on the graph based on their fit. You can see that the archetypes can be ranked based on each stadium. For example, Angel Stadium's worst to best performing archetypes go from groundballer, to power, to veteran, to finesse, to balanced pitchers being the best performers in this stadium. Stars and X's are meant to represent each archetypes best on worst fit out of all 30 ballparks. For veteran pitchers, their archetype is best suited in the conditions of Wrigley Field, however their tendencies would not fair well in Fenway Park. Notice how the Y-Axis is in decending order from 7 to 0, this due to the nature of the PCA heatmap in Figure (10). Remember that the lower the number is, the greater fit for the given archetype. So the higher the colored dot is on the graph above shows a better fit for the archetype in the respective venue even though the numeric measurement is smaller.

11.2 Each Archetype's Best and Worst Pitching Environment

Balanced Pitchers:

Best - Milwaukee, Wisconsin (MIL) America Family Field, home of the Milwaukee Brewers

Worst - Boston, Massachusetts (BOS) Fenway Park, home of the Boston Red Sox

Power Pitchers:

Best - Boston, Massachusetts (BOS) Fenway Park, home of the Boston Red Sox

Worst - Denver, Colorado (COL) Coors Field, home of the Colorado Rockies

Groundball Specialists:

Best - Arlington, Texas (TEX) Globe Life Field, home of the Texas Rangers

Worst - Denver, Colorado (COL) Coors Field, home of the Colorado Rockies

Finesse Pitchers:

Best - Seattle, Washington (SEA) T-Mobile Park, home of the Seattle Mariners

Worst - Boston, Massachusetts (BOS) Fenway Park, home of the Boston Red Sox

Veteran Pitchers:

Best - Chicago, Illinois (CHI) Wrigley Field, home of the Chicago White Sox

Worst - Boston, Massachusetts (BOS) Fenway Park, home of the Boston Red Sox

11.3 Focus Stadiums:

We will focus on the same five stadiums previously highlighted in Figure (8) Section 8.3.

11.3.1 (SF) Oracle Park

I find the Oracle Park archetype fit scores to be interesting to say the least. Giants ace Logan Webb is one of the most efficient groundball pitchers in the MLB who finds great success pitching at home in San Francisco. However groundball pitchers like him are ranked second to last when compared to the other archetypes in Oracle Park. My hypothesis is that Oracle Park is simply THE premire pitching destination across all Major League Baseball stadiums hands down. Groundball pitchers greatly benefit in smaller venues as seen in Dodger Stadium and Globe Life Field. However Oracle Park's high wind speeds (10 mph) and inward direction combined with a cooler climate, and above average field dimensions gives ANY pitcher the advantange over hitters. Taking these factors into consideration, this explains why power pitchers are highly successful in San Francisco. Their above average fastball velocity will not get punished as much as compared to other venues. Fly balls and hard hit contact will be greatly suppressed due to both the environmental and structural advantages presented for pitchers in San Francisco. This is also supported in Figure (9) which shows Oracle Park having the lowest average home run index (83).

11.3.2 (SD) Petco Park

It's cool to see how Petco Park has the lowest variance compared to all other MLB stadiums. The variance between all 5 archetypes is small, and balanced pitchers are favored in this venue.

This reinforces my claim in Section 8.3.2, that San Diego's perfect weather conditions and typical field dimensions leads to this venue being extremely average. With an about average home run index score of 103 in Figure (9), there is no noticable advantage granted for neither pitchers nor hitters. Petco Park can be seen as the perfect mediator ballpark when judging pitching and hitting performance.

11.3.3 (NY) Yankee Stadium

Yankee Stadium also has a noticably small variance compared to other MLB stadiums. However unlike Petco Park, this is a hitter friendly stadium. Power pitchers are ranked last among all five pitching archetypes in Yankee Stadium. This insight is supported in Figure (9) as Yankee Stadium has a top 3 average home run index score in the past five years (118). Figure (3) shows New York's above average summer temperature of 85 and outward wind direction justifies this stat. In cities with high humidity like Yankee Stadium (68) and Minute Maid Park (80), it allows baseballs to travel in the air longer than low humidity environments. Yankee Stadium also has a short left field porch of 314 ft and a low fence height of 8 feet which is nice for left handed power hitters. So pitchers who rely on overpowering hitters can get punished if a hitter creates solid contact with their high velocity pitches, as they are at high risk in batters generating flyballs into sneaky homeruns.

11.3.4 (COL) Coors Field

The Colorado Rockies have the worst pitching venue across all MLB stadiums. Although Figure (10) shows Coors Field having one of the biggest ballparks, it still does not make up for the stadium being so poor for pitching. It has the highest average fit score of 4.38 and the worst fit for both power and groundball pitchers at a signficant margin. But why is this? Looking back at Figure (2), the stadium's characteristics favor hitters significantly. Higher altitude influences ball flight and Coor Field in Colorado has an elevation of 5200 ft, which is by far the highest built stadium in the MLB (2nd: Chase Field - 1100 ft). With an average summer time weather of 88 and an outward wind speed of 8 mph, it makes pitching in Colorado such a challenge. This is why groundball pitchers suffer the most. Their ability to keep in the ball in the field does not fit well with the large dimensions of Coors Field. Hitters are able to find larger gaps and defenders are forced to cover a larger space.

11.3.5 (BOS) Fenway Park

Figure (11) shows how unique Fenway Park truly is. It is the only stadium that has multiple X's and a star. With its large variance, quantifying pitching performance in Fenway may soley just depend on who is pitching. Balanced pitchers severely underperform in Boston in comparison to other stadiums. This outlier may be due to its unique dimensions, their strengths may not be projected well as in comparison to your typical MLB stadium. The same can be said about the finesse and veteran archetype. Power pitchers thrive the most in Fenway possibly due to the high wall heights. Having the green monster helps in suppressing home runs dispite the left field porch only being 310 ft. The high center field wall height of 18 ft also makes it hard for hitters to hit it out of the middle part of the field as well. Like in Yankee Stadium, left handed power hitters gain an advantage because the right field fence is 3 feet. So all in all, Figure (11) showcases MLB's true unicorn stadium in Boston's Fenway Park.

11.4 (36) Carlos Rodon

I would like to highlight Carlos Rodon, starting pitcher for the New York Yankees. He signed with them during the 2023 off-season after pitching for the San Francisco Giants in 2022. With the Giants, he posted a 2.88 ERA, 6.6 H/9, 0.6 HR/9, and a career low 2.25 FIP. However after moving to New York in 2023, his stats sky rocketed with a 6.85 ERA, 9.1 H/9, 2.1 HR/9, and a career high 5.79 FIP. He was able to better adjust to the new stadium conditions of Yankee Stadium in 2024, but not to the same extent as for when he was with the White Sox or Giants, who both play in pitcher friendly venues as shown in Figure (8). Carlos Rodon's career is a good example of how influencial stadium environments really are when accounting for pitching performance. He was able to find great success in San Francisco due to the pitcher friendly stadium conditions, however it did not carry over to the same extent with the New York Yankees.

11.5 Takeaways

Using the numeric measurements in Figure (10), Figure (11) allows us to visualize how each archetype fair against each other in all MLB stadiums. It's easier to interpret than the heat map and the stars and X's shows where we should expect each archetype's best and worst performances. It also supports my claims back in Section 8.3, but provides a visual analysis with comparisons to other MLB stadiums.

12 Discussion

This study provides a data-driven framework for understanding the relationship between pitcher archetypes and MLB ballpark characteristics. By combining Statcast pitching metrics with detailed environmental and structural stadium data, by creating meaningful clusters of both pitchers and parks. The use of PCA allowed us to distill complex multivariate information into interpretable visualizations, while KMeans clustering and Euclidean distance measures helped quantify fit between pitcher styles and stadium environments with numeric values. This approach acknowledges that not all pitcher performance is purely skill-based—context, such as park factors, matters significantly in evaluating outcomes like home run suppression, contact rates, and ERA.

I like how the heatmap stands out as a key tool, it gives a grade on how well each archetype fairs in each stadium which makes it easy to identify if a pitching matchup is expected to do good in each given stadium. It is the backbone of Section 11 which is simple to read with in-depth analysis of various stadiums and archetype characteristics. The scatter plot gives a different viewpoint on the same statistical metrics provided for us in the Figure (12) heatmap. Through these graph we are able to see each archetype's best and worst performing venues as well as an easily interpretable ranking system in each stadium.

A noticable example of a franchise that plays into their home field advantage very well are the San Francisco Giants. They found great success in the early 2010's as their dynasty won three championships in five years simply off of the backs of phenomenal pitching. Those championship winning teams consistently ranked in the lower quartile offensively in hits and runs produced. However their power starting pitchers like Tim Lincecum and Madison Bumgarner kept them afloat just enough to find great playoff success. Their "Core-Four" bullpen also provided meaningful late-game support, suppressing opposing hitters and sealing numerous one-run playoff wins. In the present day, free agent signings like Jung Hoo Lee and Matt Chapman showcase their incredible fielding ability and range which gives support to their pitchers. Young gold-glove catcher Patrick

Bailey consistently leads the league in numerous defensive catching metrics which reinforces the San Francisco Giants as being a pitching and defensive focused team due to their pitcher-friendly home field environment. However since hitters struggle in Oracle Park, this research reveals a contributing factor as to why the Giants continue to be a mediocre offensive ball club year in and year out.

I will admitt that it was incredibly satisfying to see that my claims and hypothesis is correctly backed up using real statistics on MLB pitchers and stadiums. I had a few doubts that there wasn't a signficant correlation to pitcher performance and venue characteristics, however the findings showed to be extremely revealing. Prior to building this project, my father and I had numerous conversations about how pitching performance is in fact influenced by city and venue environments. Being able to quantify our insights through research using data science and machine learning engineering is work that I am proud to produce and showcase.

I hope my readers learned a lot about the game of baseball as much as I did while making this research paper. Although I've played this sport for over 12 years, having Irwin explain the five pitching archetypes gave me a new perspective on how the game is played. Baseball is an art and every pitcher is an artist with their own creative style. Every single baseball field is unique and the distinct city conditions across North America influence every single game which demonstrates the true beauty of this sport.

12.1 Limitations

While this study offers a novel framework for matching pitcher archetypes to MLB stadium profiles, it comes with several limitations. First, the PCA and clustering analyses are based on a fixed set of input variables, meaning the results are highly dependent on the chosen metrics (e.g., HR/9, GB%, wind speed, fence height). If important factors were omitted—such as pitch movement data, batter tendencies, or in-game strategy adjustments—the model may not capture the full picture of performance variability. Additionally, some variables (like wind direction or temperature) fluctuate throughout a season, and my dataset assumes summertime weather averages which may not reflect real-time game conditions and postseason baseball.

Pitching archetypes were also manually assigned by me basing my classifications only through my personal interpretations of their statistics and my findings in Figures (5) and (6). So this gives a loose bias on how a small sample size of sixty (60) MLB starting pitchers perform and shapes this study on how I personal identify each pitcher. An MLB franchise general manager or pitching coach may have different classifications as me which can slightly alter the findings on pitcher archetypes. However I trust my judgement enough to claim that the findings in this study is accurate and creditable.

Another limitation lies in the assumption that proximity in PCA space equates directly to performance success. While PCA distance gives a statistically sound estimate of similarity between pitcher types and stadiums, it doesn't guarantee causation. Other contextual factors like team defense, bullpen support, and game pressure can and will influence outcomes beyond what my model accounts for. Furthermore, pitcher archetypes are not always rigid; players evolve over time, and some may blend multiple styles or adjust based on ballpark conditions, something a static clustering method like KMeans does not fully capture.

12.2 Future Research

Future research can build upon this study by incorporating more dynamic, real-time variables to increase the accuracy and predictive power of pitcher-park compatibility models. Instead of relying solely on seasonal averages, researchers could use game-by-game data that accounts for changing wind directions, temperatures, humidity levels, and even crowd noise—each of which may subtly influence pitcher performance. Additionally, integrating Statcast-level pitch tracking data (e.g., spin rate, pitch movement, release point consistency) could allow for a more granular definition of pitcher archetypes, potentially identifying subtypes or hybrid styles not captured in this study.

From a strategic standpoint, this research can directly influence team decision-making. Front offices could reference this study for such archetype-stadium compatibility insights to guide free agent signings, bullpen management, or rotation planning during road trips. This reaserch can also greatly aid playoff readiness for any given series, as teams are able to isolate one stadium and cultivate a pitching game-plan based on the opposing city and MLB stadium characteristics. For player development, it offers a framework for refining skill sets in harmony with home park features. Franchises are able to look at this study and cultivate a roster that plays into their strengths and weaknesses giving each stadium their own unique home field advantage. Ultimately, this research paper highlights the value of integrating advanced analytics, environmental science, and machine learning in baseball strategy—and opens the door for even more granular matchup modeling for the future of baseball.

Another promising direction is to extend the analysis to batter-pitcher-stadium three-way matchups, which would account for how a specific pitcher fares against certain hitters in specific environments. This could help teams make more precise game-day decisions and optimize matchup-based strategies. Machine learning models such as random forests or gradient boosting could also be used to predict game outcomes or individual pitcher performance given the interaction between pitcher traits, stadium context, and opponent characteristics.

Lastly, this framework could be applied beyond MLB to minor league parks, international leagues, or college baseball, where environmental variability and structural asymmetry are often even more pronounced. Over time, refining this approach could assist front offices in not only managing current rosters but also in scouting, drafting, and developing players whose skill sets align with the unique features of their home ballparks—bridging the gap between player talent and environmental fit for a competitive advantage.

13 Conclusion

This study supports the hypothesis that by accurately characterizing a pitcher's style, it is indeed possible to identify MLB stadium environments that best align with their strengths. Through the use of PCA, KMeans clustering, distance-based heatmapping, and a ranking plot graph, we demonstrated that different pitcher archetypes perform better in specific ballpark conditions based on structural and environmental fit. These findings highlight the value of matching player traits with contextual park factors, reinforcing the idea that optimizing individual performance through strategic deployment can contribute directly to overall team success. This data-driven approach not only validates the hypothesis but also provides a practical framework for decision-making in modern baseball analytics.

14 Works Cited

Ballpark Pal
"MLB Park Factors and Environmental Data." *BallparkPal.com*, https://ballparkpal.com/. Accessed 4 Aug. 2025.

Baseball Reference
"2024 Pitching Stats." *Baseball-Reference.com*, Sports Reference LLC, https://www.baseball-reference.com/leagues/majors/2024-standard-pitching.shtml. Accessed 17 Jul. 2025.

Baseball Savant
"Statcast Search and Player Pitching Profiles." *Baseball Savant*, MLB Advanced Media, https://baseballsavant.mlb.com/. Accessed 12 Aug. 2025.

FanGraphs
"2024 Pitcher Leaderboards." *FanGraphs Baseball*, https://www.fangraphs.com/leaders.aspx?pos=all&stats=pit. Accessed 2 Aug. 2025.

Irwin, Collin
"The 5 Pitching Archetypes." *Western Oregon University*, Personal interview. Accessed 18 Jul. 2025

MLB.com
"Ballpark Dimensions and Park Factors." *MLB Advanced Media*, https://www.mlb.com/. Accessed 23 Jul. 2025.

NOAA (National Oceanic and Atmospheric Administration)
"Climate Normals and Weather Averages by City." *National Centers for Environmental Information*, https://www.ncei.noaa.gov/. Accessed 14 Aug. 2025.

StatMuse
"Weather, Ballpark Stats, and MLB Performance Data." *StatMuse*, https://www.statmuse.com/mlb. Accessed 7 Aug. 2025.

WeatherSpark
"Average Weather in Major US Cities." *WeatherSpark*, https://weatherspark.com/. Accessed 19 Aug. 2025.

JupyterLab
"Project Jupyter." *Jupyter.org*, https://jupyter.org/. Accessed 15 Jul. 2025.

UC San Diego DataHub
"Data Science Computing Platform for UCSD Students." *datahub.ucsd.edu*, https://datahub.ucsd.edu/. Accessed 15 Jul. 2025.

Inquiries:

Email: jhunerryviray3@gmail.com

Linked-In: https://www.linkedin.com/in/emmanuel-viray/

GitHub: https://github.com/emjviray