

투구 클러스터링을 활용한 불펜 투수의 전략적 운용

YSAL x sports2i 협업 프로젝트

김유경 윤석민 이주원 이준영 이지용



논문 및 24-1 프로젝트 소개

<논문 소개>

Using Clustering To Generate Bullpen Matchups

by Cam Rogers 
August 30, 2021

 `matchup("MIA", "Joc Pederson")`

Pitcher	eΔRE/100
Richard Bleier (LHP)	-3.22
Ross Detwiler (LHP)	-0.48

<24-1 프로젝트>

결과 검증

최종 결과의 유효성 확인

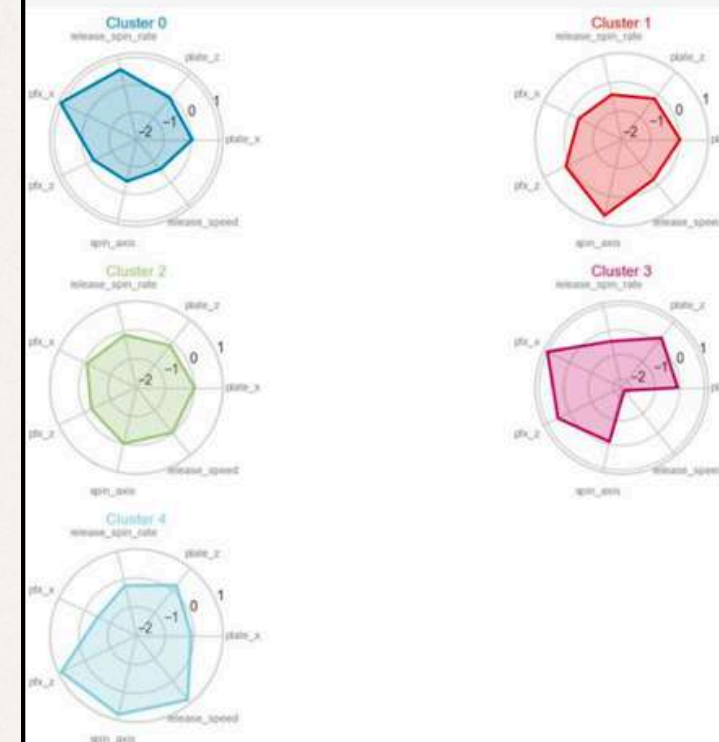
Bryan Shaw

<분석 결과>

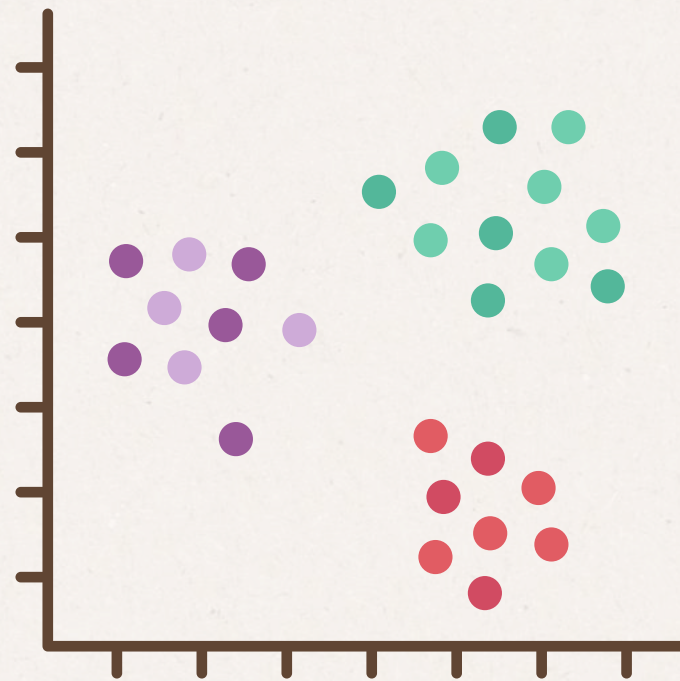
pitcher	<input checked="" type="checkbox"/> R-SL-0	R-SL-1	R-SL-2	R-SL-3	R-SL-4
Bryan Shaw	0.107535	0.002195	0.212143	0	0.562546

<실제 투구>

	pfx_x	pfx_z	release_speed	spin_axis	count
	mean	mean	mean	mean	
cluster					
R-SL-0	1.231	0.013	84.103	57.408	147
R-SL-1	0.370	0.560	94.367	153.000	3
R-SL-2	0.455	0.315	91.262	153.290	290
R-SL-4	0.366	0.610	93.411	176.636	769



프로젝트 단계 및 목표



1. 클러스터링을 통한
투구 유형 분류



2. 각 유형에 따른
타자들의 $e\Delta RE$ 계산



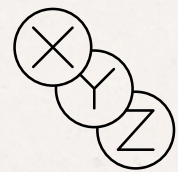
3. 상황 별 투수 선택

개선점

<기존 프로젝트>

<새로운 프로젝트>

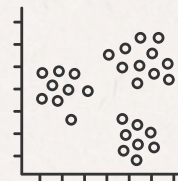
데이터 수 -26만 개



데이터 확대 및
클러스터링 변수 추가

데이터 수 -106만 개
release_pos_x, release_pos_z 추가

GMM 클러스터링 사용



클러스터링 방법
변화

Kohonen's SOM
클러스터링 사용

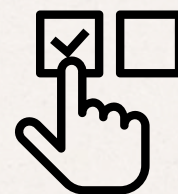
raw data의 delta_run_exp 열
활용해 간단하게 계산



새로운 eΔRE 계산

In_play와 out_play 상황을 나눠
상황 별 eΔRE 계산

X



Ball-Tree
Algorithm 사용

타자 별로 특정 상황 유도하는 데
유리한 투수 선택

목차

- 1 데이터 수집 및 전처리
- 2 모델 구현: 투수 클러스터링
- 3 $e\Delta RE$ 계산
- 4 Ball-Tree Algorithm
- 5 프론트 구현 및 사례 분석



데이터 수집 및 전처리

YSAL x sports2i 협업 프로젝트



데이터 수집

Baseball Savant STATCAST 2018-2023



- RP (계투) 투수 중 해당 기간 250 타석 이상 타자를 상대한 투수

- Columns (열)

: pitch_type, pitcher, batter, stand, p_throws, delta_run_exp, pfx_x, pfx_z, plate_x, plate_z, release_spin_rate, release_pos_x, release_pos_z, effective_speed, launch_angle, launch_speed, discription, events

- 동일 기간 좌/우 투수 상대로 각각 100 타석 이상 상대한 타자



데이터 전처리

비슷한 유형의 투구 그룹화 & 분류

FF->FF(Four-seam Fastball)

SI->MV(Moving Fastball)

ST, FC, SL -> SL(Slider)

CH, FS, SC -> OS(Off-Speed)

KC, CU, SV -> CV(Curveball)

CS, PO, 결측값 -> X(Nothing)

결측치 처리

동일 투수 내

동일 구종으로 그룹화

: release_spin_rate, spin_axis,
effective_speed 등

-> 평균값으로 대체

Column (열) 수정

Description 항목 (hit into play, swing strike 등)

In_play & not_in_play 항목으로 이분화

모델 구현 : 투수 클러스터링

YSAL x sports2i 협업 프로젝트



모델 구현 : 투수 클러스터링

GMM clustering

확률 기반 클러스터링

한 데이터가 여러 개의
정규 분포로부터 생성되었다
가정

데이터가 특정 분포에 속할
확률을 계산하여 클러스터링 수행
(가장 높은 likelihood 가질 분포 할당)

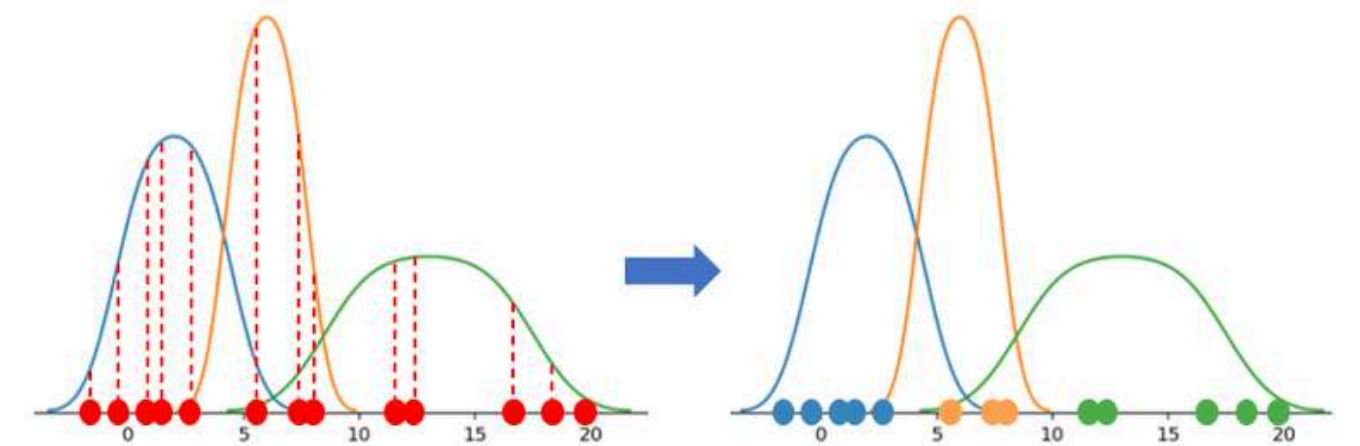
Kohonen's SOM Clustering

비지도 학습 기반 클러스터링

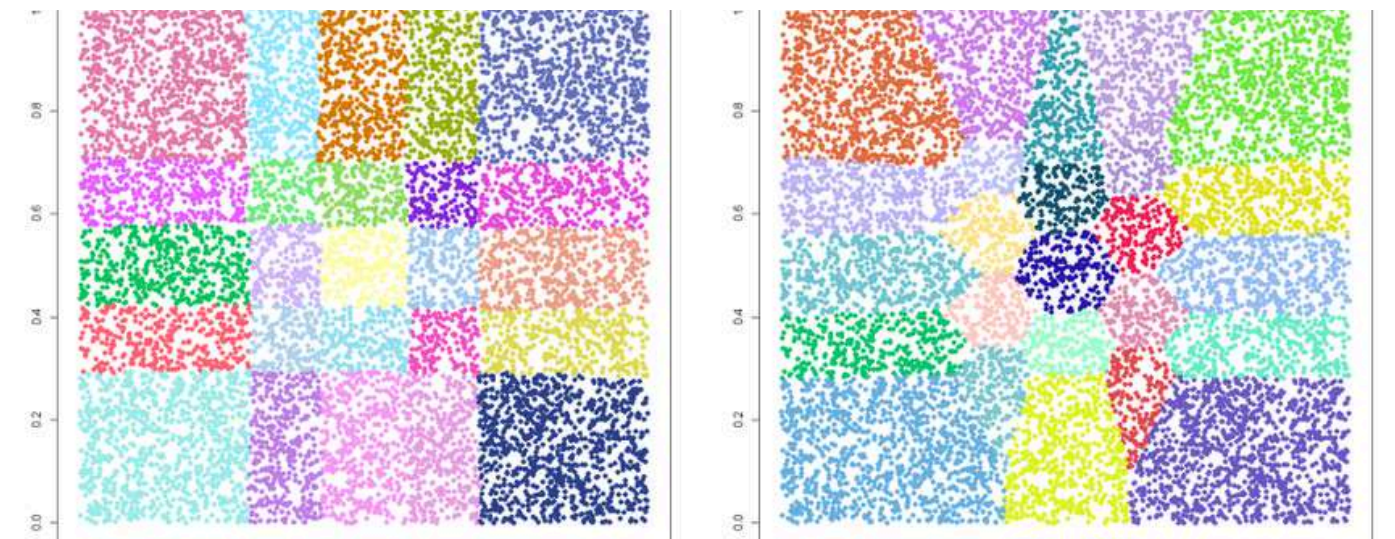
데이터의 고차원 공간을
저차원(보통 2차원)으로
시각화하고 군집화하는 기법

유사한 특징의 데이터 포인트끼리
같은 클러스터로 분류

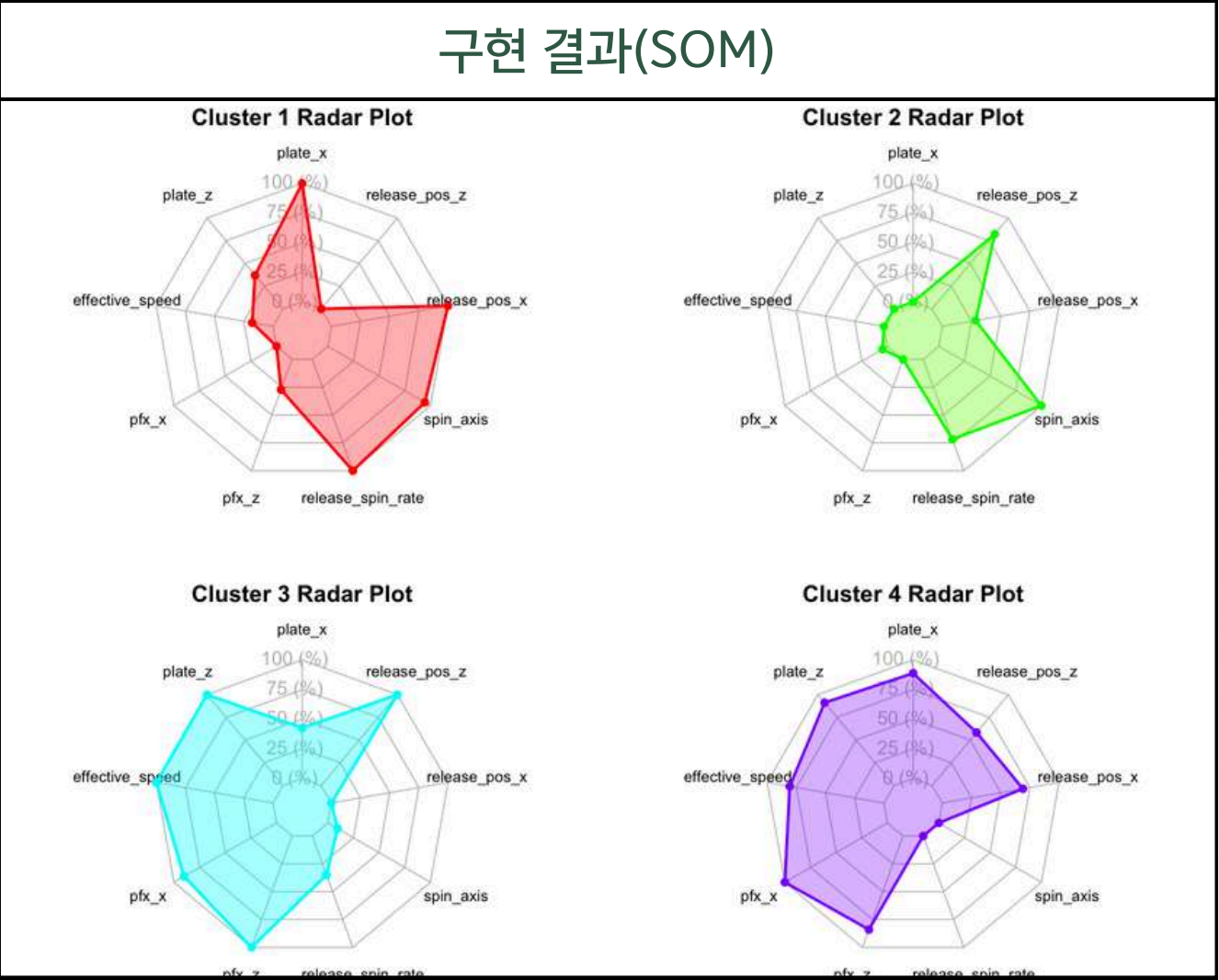
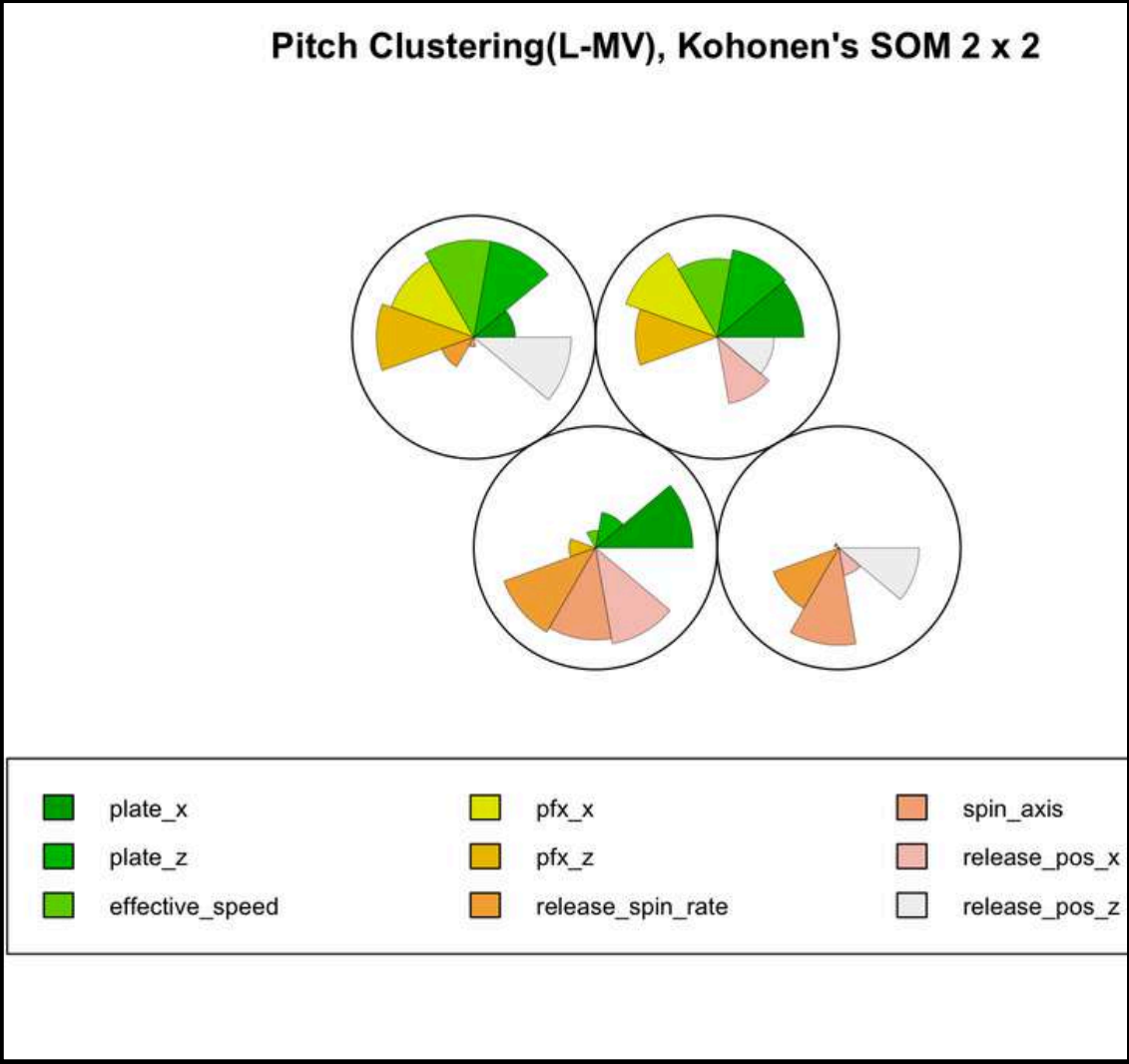
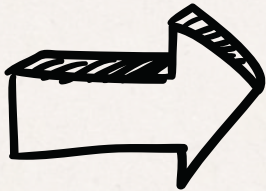
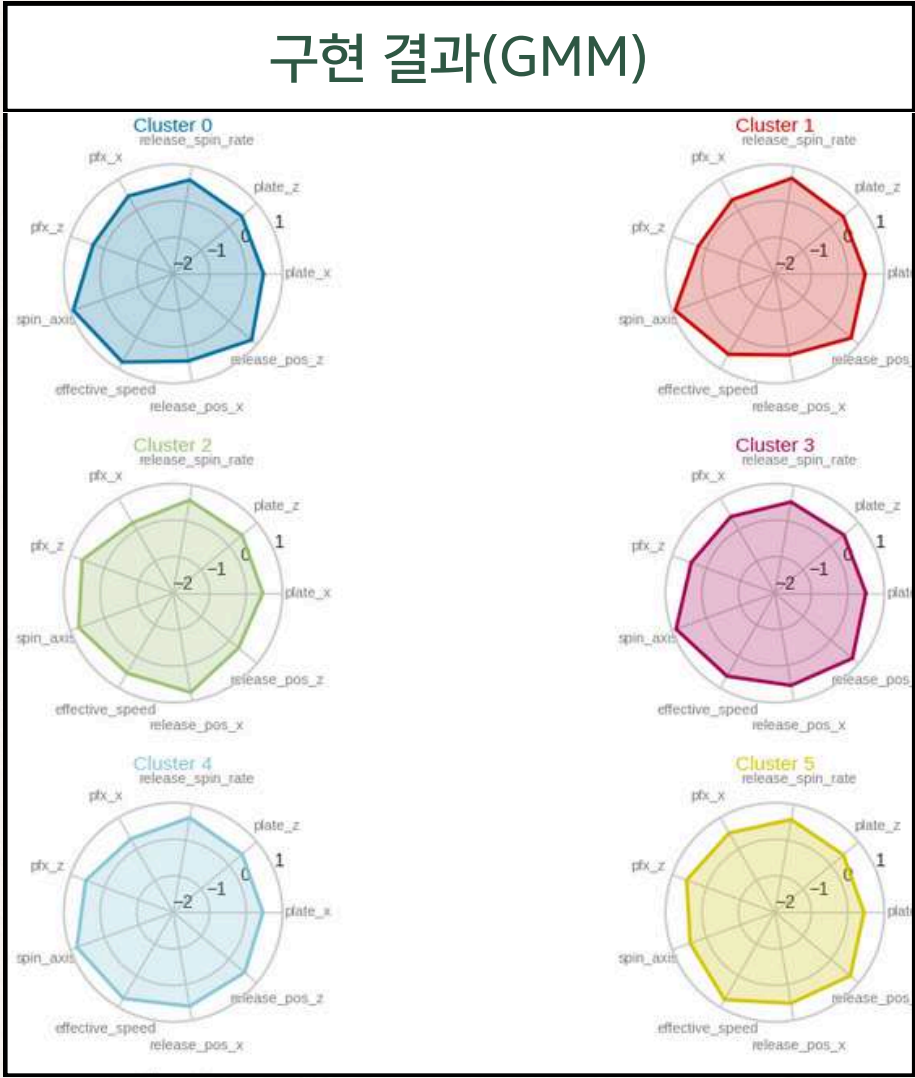
GMM clustering



SOM 클러스터링(uniform 분포, rectangle-hexagonal)

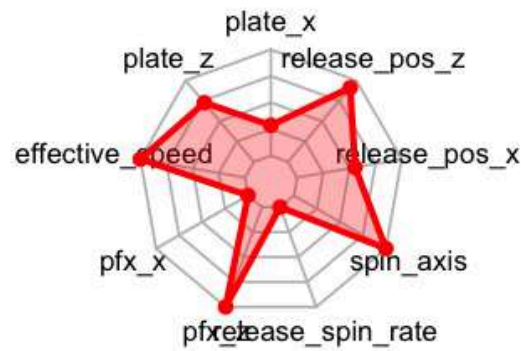


모델 구현 : 투수 클러스터링

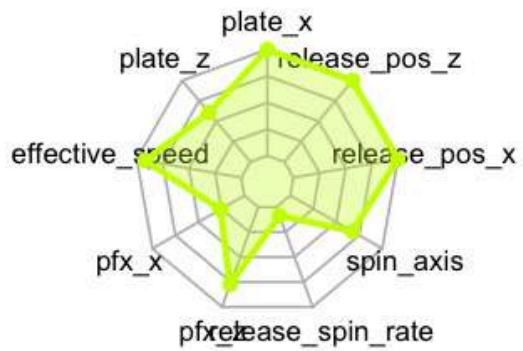


모델 구현 : 투수 클러스터링

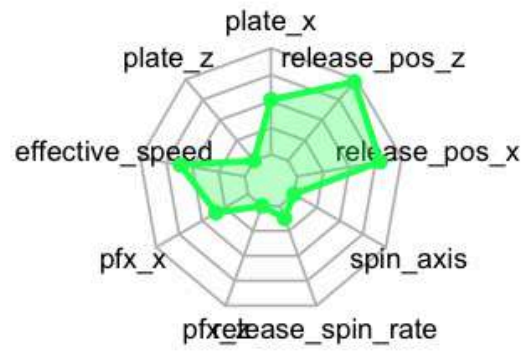
Cluster 1 Radar Plot



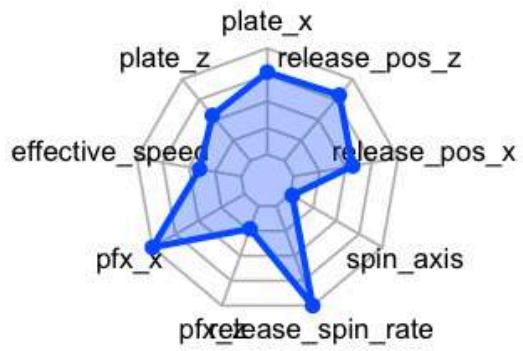
Cluster 2 Radar Plot



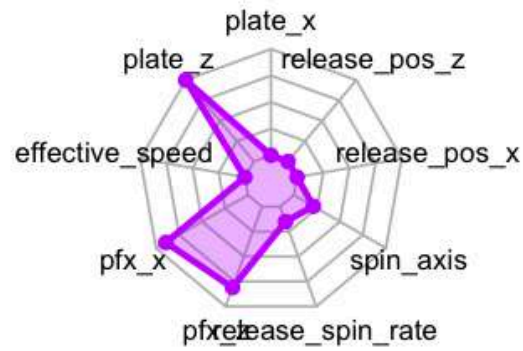
Cluster 3 Radar Plot



Cluster 4 Radar Plot

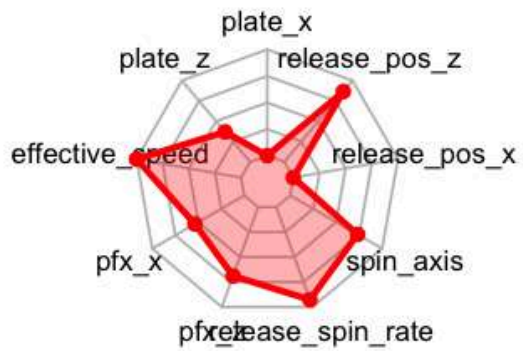


Cluster 5 Radar Plot

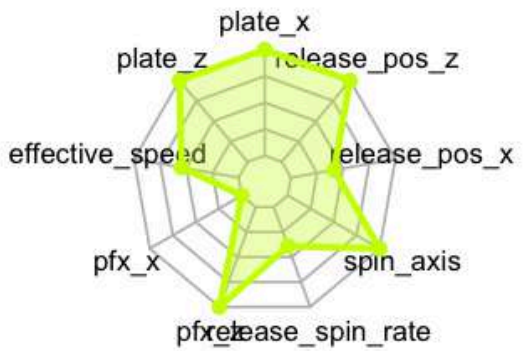


R-SL 클러스터링

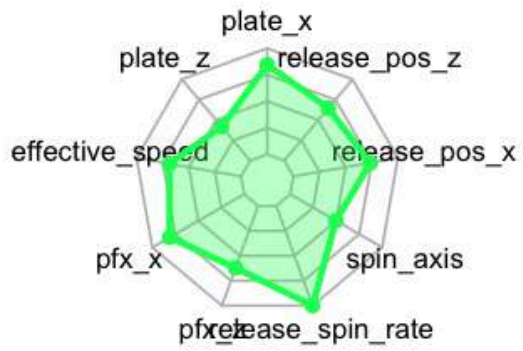
Cluster 1 Radar Plot



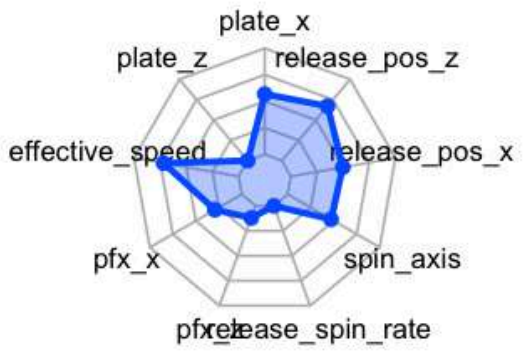
Cluster 2 Radar Plot



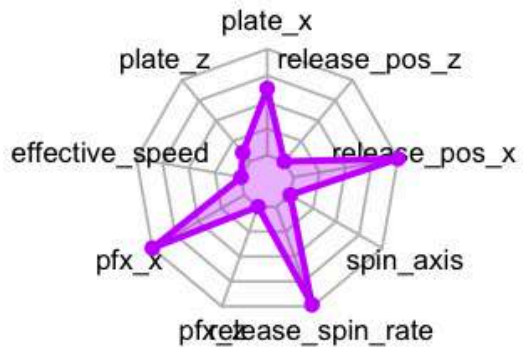
Cluster 3 Radar Plot



Cluster 4 Radar Plot



Cluster 5 Radar Plot



L-OS 클러스터링

eΔRE 계산

YSAL x sports2i 협업 프로젝트



RE(Run Expectancy)란?

특정 상황이 일어난 이후 이닝이 끝날 때까지 득점한 득점의 수

특정 상황이 발생한 횟수

“S-B-O 카운트 & 주자 상황에 따라 값이 다름!”

Δ RE(Delta Run Expectancy)란?

$$\Delta RE = (\text{After pitch}) \text{ Run Expectancy} + \text{Runs (득점)} \\ - (\text{Before pitch}) \text{ Run Expectancy}$$

각 Pitch 별 타격 창출 능력 평가!

e Δ RE 계산

기존 Δ RE

: S-B-O 카운트, 주자 '상황' 등이 반영됨

→ 타자의 순수 능력 평가 어려움



상황 요소와 독립적인
새로운 지표 e Δ RE 제안

e Δ RE 계산

In-Play 상황

타구 속도와 발사 각도를 독립변수로 하는 “다항 회귀”를 적용
→ Δ RE에서 타구 질에 대한 값만 추출

VS

Non In-Play 상황

Event 별로 분류하여 Δ RE의 평균값을 적용
(Called Strike, Ball, Foul 등 12가지)

eΔRE 계산

왜 타구 속도와 발사 각도인가?

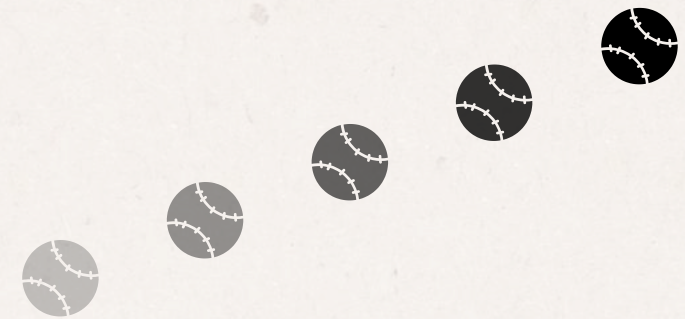
타구 속도(Exit Velocity).
: 배트에 맞은 공이
떠날 때의 속도

발사 각도(Launch Speed).
: 배트에 맞은 공이
날아가는 각도



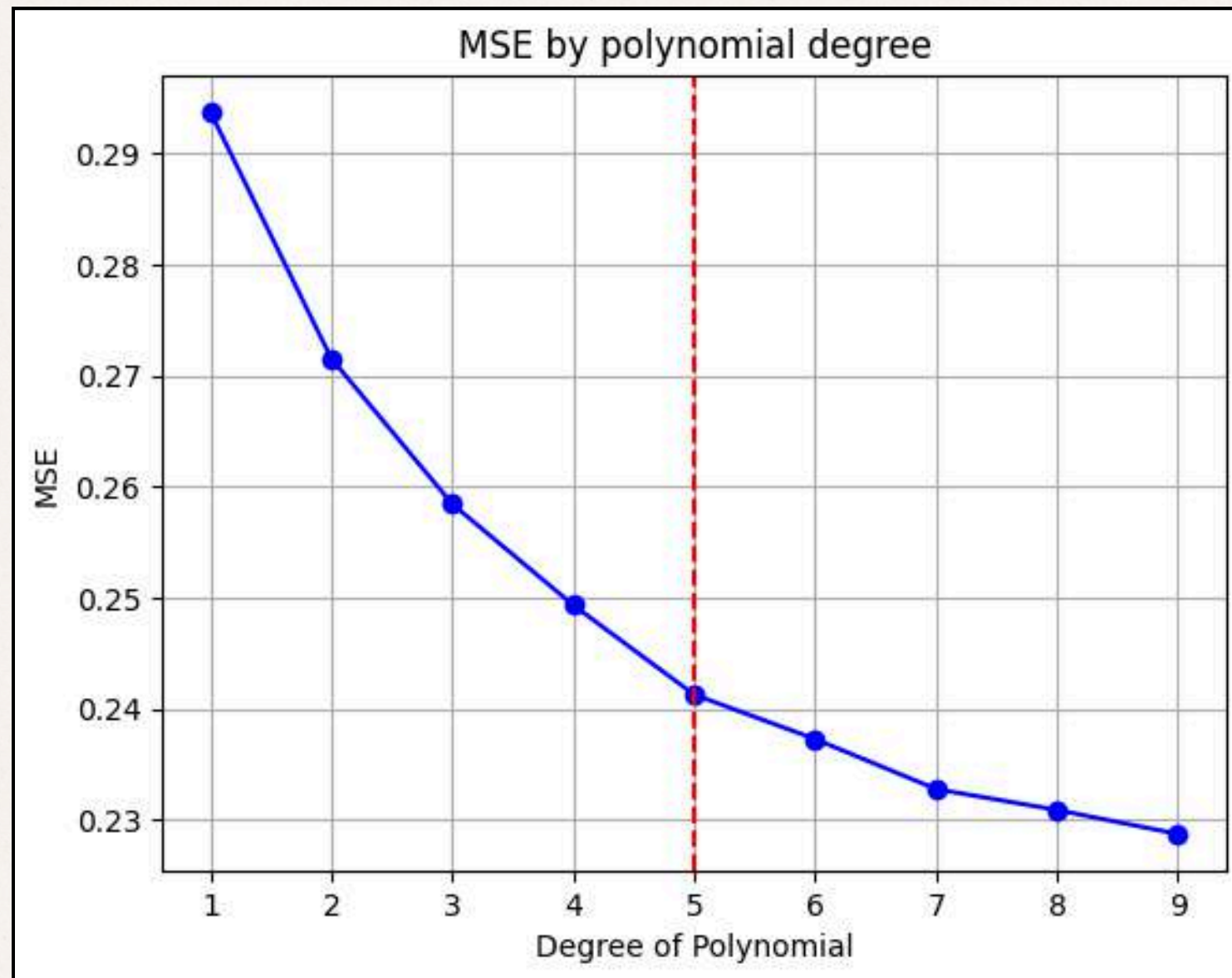
타구의 속도가 빠를수록,
발사 각도가 25도 근처일수록,
‘타구의 질’이 좋아짐

: 타자의 득점 생산성을
대표하는 지표!

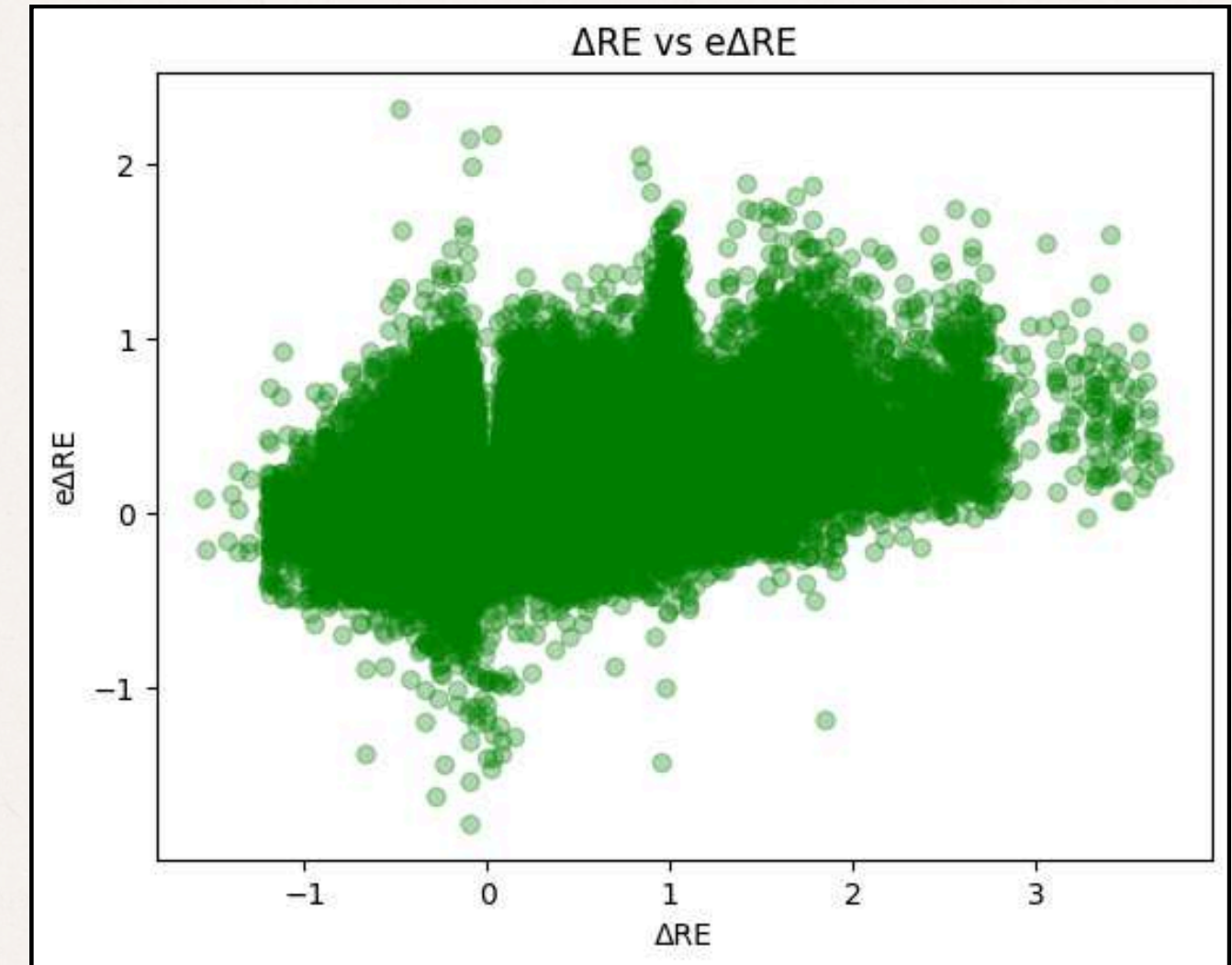


EXIT VELOCITY	111.6	[MPH]
LAUNCH ANGLE	22	[DEG]
PROJECTED DISTANCE	455	[FT]

e Δ RE 계산



Degree=5로 설정



Δ RE: 0.556, e Δ RE: 0.09409391301308334
 Δ RE: 0.32, e Δ RE: 0.3217461596457416
 Δ RE: -0.002, e Δ RE: -0.20809901556877008
 Δ RE: -1.136, e Δ RE: -0.05989302325473522
 Δ RE: -0.29, e Δ RE: -0.18080595070932404

<출력 예시>

타자에 따른 불펜 투수 Matching

Optimal Bullpen Pitcher =

argmin
Pitcher

\sum

Every Cluster
 \in Pitch Cluster

각각의 Pitch Cluster에 대한

투수의 투구 비율

X

타자의 eΔRE

타자에 따른 불펜 투수 Matching

	R-CV-1	R-CV-2	R-CV-3	R-CV-4	R-CV-5	R-CV-6	R-CV-7	R-FF-1	R-FF-2	R-FF-3	...	R-OS-1	R-OS-2	R-OS-3	R-OS-4	R-OS-5	R-SL-1	R-SL-2	R-SL-3	R-SL-4	R-SL-5
pitcher_name																					
Abreu, Albert (RHP)	0.000000	0.000000	0.000000	0.000862	0.000000	0.000862	0.000000	0.040966	0.000000	0.000431	...	0.0	0.000431	0.024580	0.110824	0.000000	0.183700	0.106511	0.016386	0.000000	0.000000
Abreu, Bryan (RHP)	0.032834	0.000714	0.000000	0.028551	0.017488	0.000000	0.001784	0.176303	0.008922	0.066024	...	0.0	0.000000	0.001784	0.000357	0.000000	0.001784	0.001071	0.241256	0.245539	0.000000
Adam, Jason (RHP)	0.000000	0.000000	0.000000	0.013403	0.011442	0.082707	0.000654	0.107225	0.014057	0.114090	...	0.0	0.016345	0.009480	0.207911	0.000981	0.000654	0.000654	0.042824	0.174240	0.000000
Adams, Austin (RHP)	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.022343	...	0.0	0.000000	0.000000	0.000000	0.000000	0.034877	0.000000	0.011444	0.764033	0.00218

	L-CV-1	L-CV-2	L-CV-3	L-CV-4	L-CV-5	L-CV-6	L-CV-7	L-FF-1	L-FF-2	L-FF-3	...	L-MV-5	L-OS-1	L-OS-2	L-OS-3	L-OS-4	L-OS-5	L-SL-1	L-SL-2	L-SL-3	L-SL-4
pitcher_name																					
Akin, Keegan (LHP)	0.000000	0.002125	0.000000	0.002657	0.000000	0.003188	0.0	0.259830	0.016472	0.000000	...	0.000000	0.015409	0.027099	0.136557	0.005845	0.000000	0.005845	0.099362	0.000000	0.149309
Alexander, Scott (LHP)	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.004067	0.001356	0.000000	...	0.000452	0.026661	0.001808	0.000000	0.028920	0.000000	0.000000	0.112969	0.017623	0.002711
Alexander, Tyler (LHP)	0.000000	0.015478	0.000000	0.000469	0.000000	0.000000	0.0	0.094278	0.003283	0.000000	...	0.017355	0.041276	0.071764	0.003283	0.047842	0.000000	0.001407	0.088649	0.128987	0.201220
Alvarado, José (LHP)	0.000000	0.000000	0.008391	0.000000	0.054903	0.000000	0.0	0.001439	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.005035	0.240230	0.000000

X

	L-CV-1	L-CV-2	L-CV-3	L-CV-4	L-CV-5	L-CV-6	L-CV-7	L-FF-1	L-FF-2	L-FF-3	...	R-OS-2	R-OS-3	R-OS-4	R-OS-5	R-SL-1	R-SL-2	R-SL-3	R-SL-4
batter_name																			
Wendle, Joey	-6.295963	-1.618051	-1.537611	-0.442490	-14.139381	-4.847900	-0.896839	1.791036	-1.129066	2.716224	...	-5.360523	-0.947940	0.156502	0.221441	0.917515	-0.603488	-3.080385	-3.946507
Carpenter, Matt	5.874938	-0.184511	0.000000	-1.457601	1.073543	-0.167267	6.306415	-0.185288	6.148050	5.874938	...	-1.707006	-1.351351	-0.250965	-1.380771	-3.391337	1.878252	-1.066873	-0.988129
Estrada, Thairo	-2.454737	2.674008	0.000000	10.305005	-1.523970	0.315526	-0.973859	1.889907	3.754108	9.327968	...	9.793909	-1.186361	1.120433	4.505596	-2.018591	-0.231073	0.295306	-2.349237
Smith, Pavin	-0.055861	-5.986660	0.000000	-2.228649	3.692124	-8.689737	0.000000	0.804526	-4.931936	1.155034	...	4.355643	3.838046	5.110726	-5.547026	-0.422895	-0.325938	-0.773624	2.118536

투수의

Pitch Distribution

타자의

eΔRE Distribution

타자에 따른 불펜 투수 Matching

<최종 결과>

```
matchup("NYM", "Trea Turner")
```

Pitcher	e_delta_RE
Brooks Raley	-2.06591
Phil Maton	-1.10372
Edwin Díaz	-0.843318

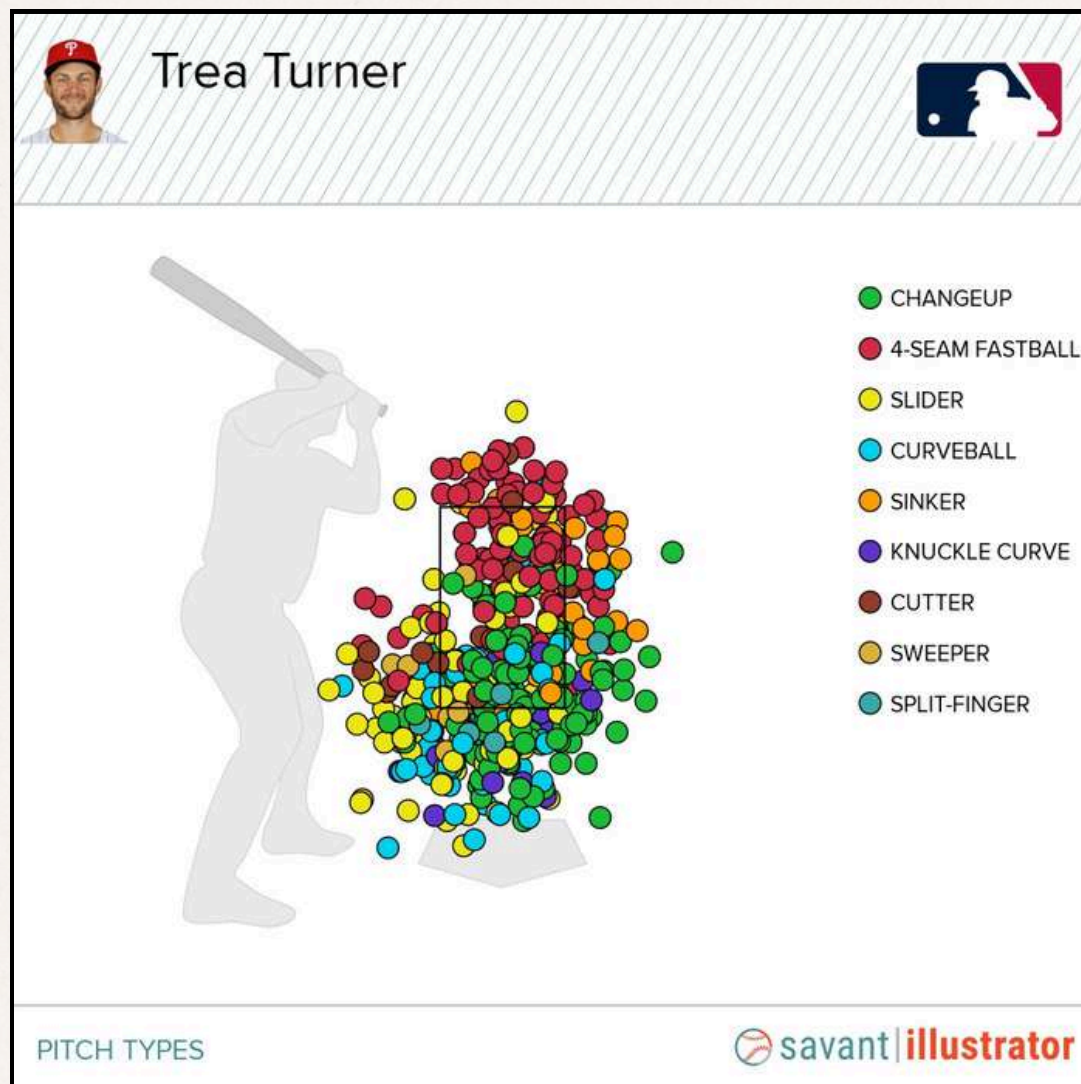
```
matchup("LAD", "Manny Machado")
```

Pitcher	e_delta_RE
Connor Brogdon	-0.839031
Evan Phillips	-0.240614
Blake Treinen	-0.102778

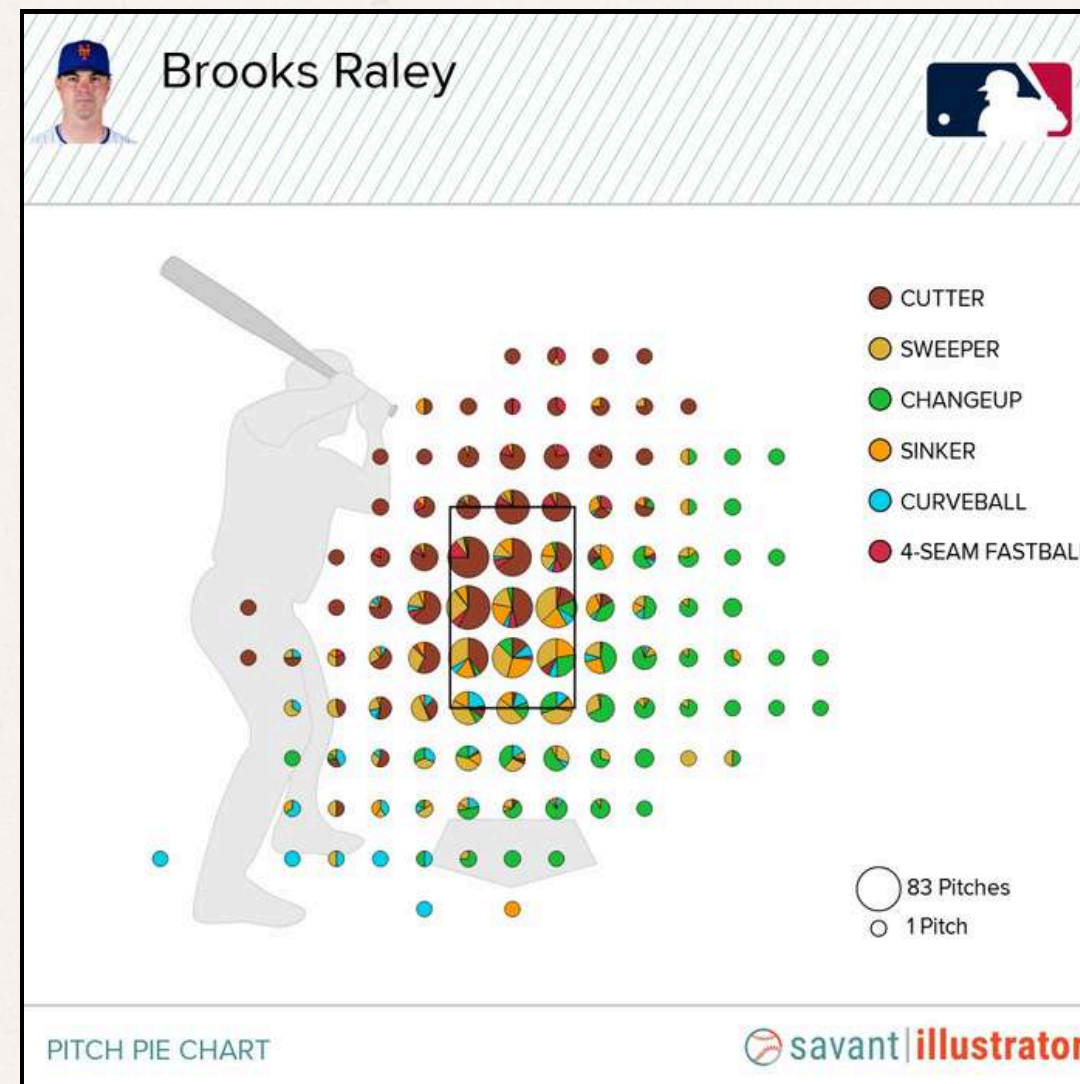
```
matchup("NYY", "Bobby Witt Jr.")
```

Pitcher	e_delta_RE
Tim Mayza	-0.98628
Phil Bickford	-0.667205
Tommy Kahnle	-0.564763

타자에 따른 볼펜 투수 Matching



스윙한 구종 데이터
(Trea Turner vs. LHP 2018~2024)



투구한 구종 분포 데이터
(Brooks Raley vs. RHB, 2020~2024)

Trea Turner는 바깥쪽과 가운데에서 떨어지는 체인지업,
몸쪽으로 꺾이는 커터,
가운데와 몸쪽에서 떨어지는 커브에 약점



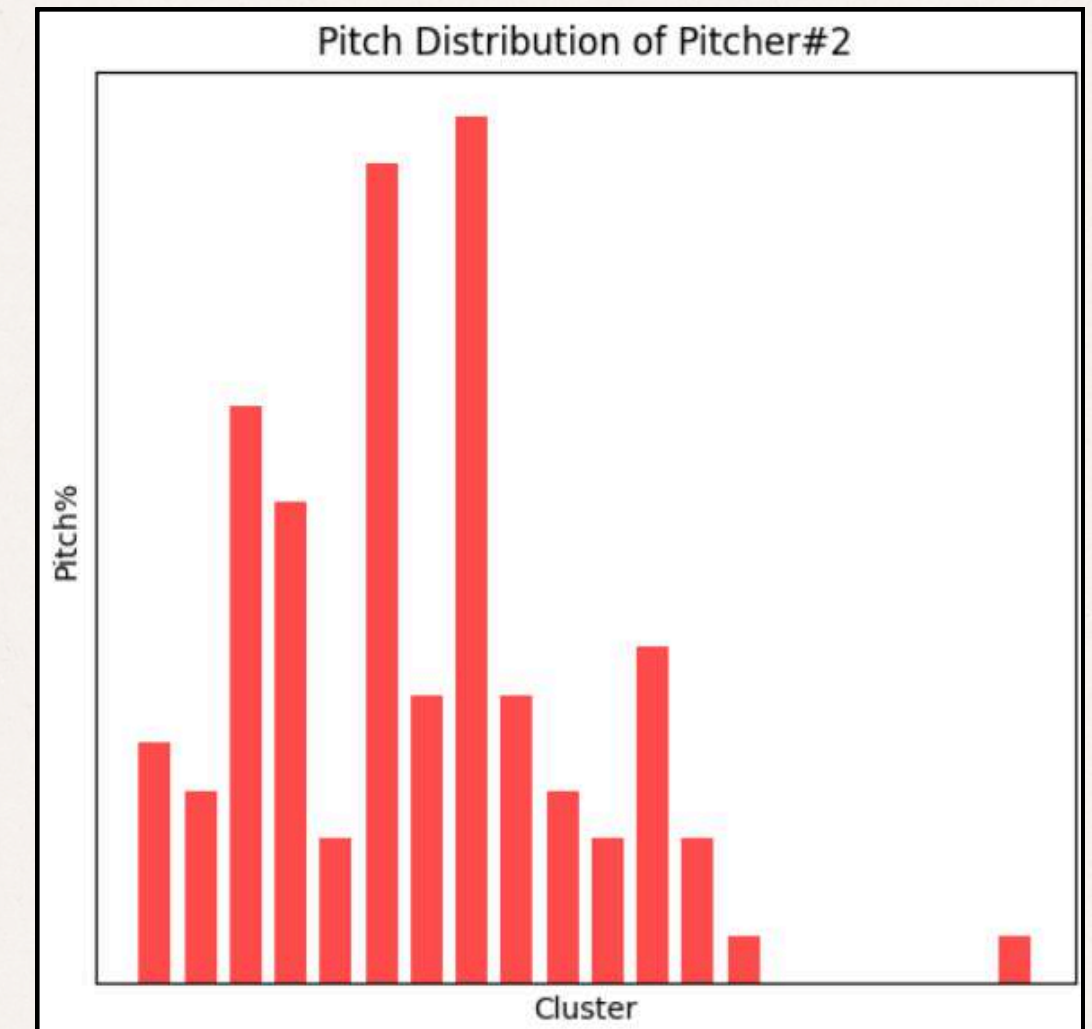
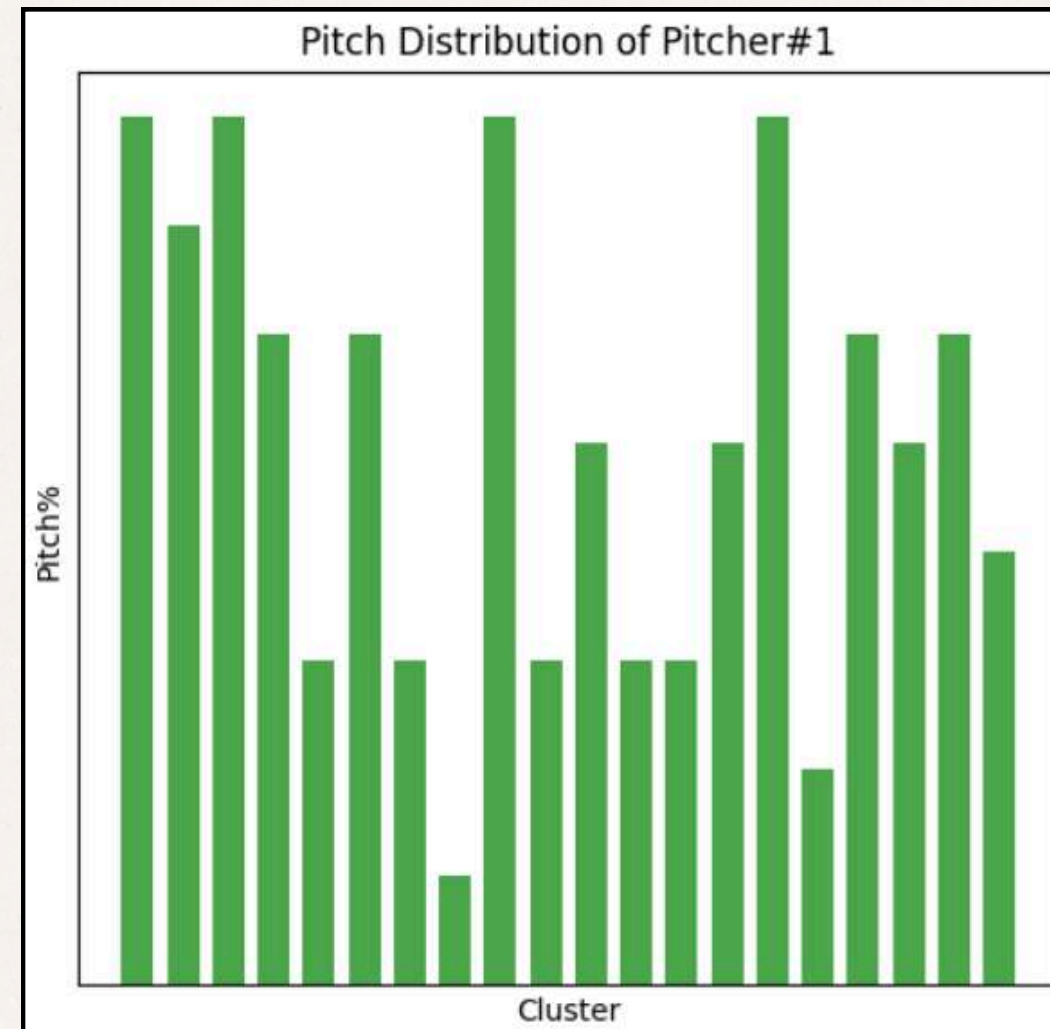
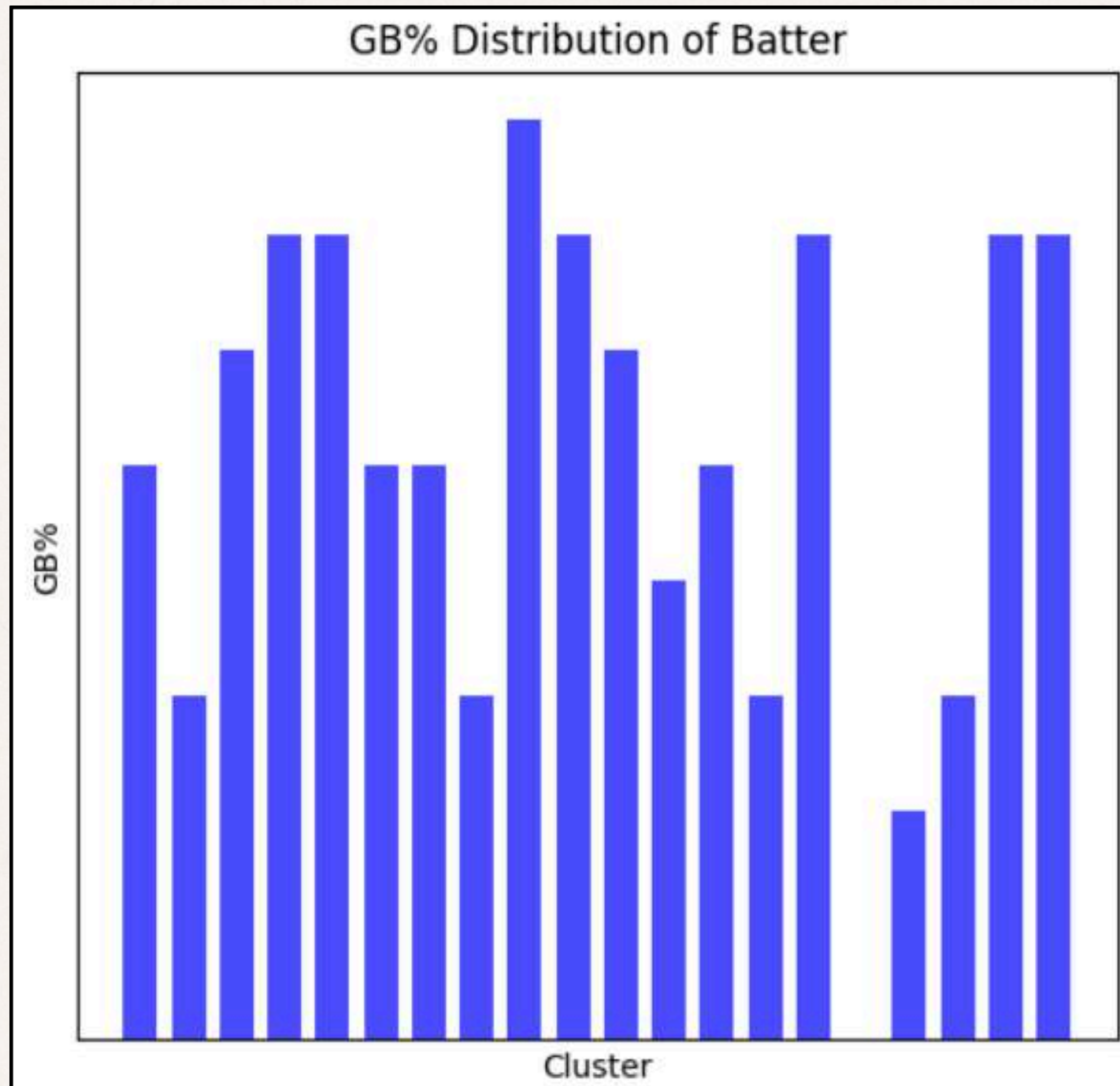
Brooks Raley의 투구 분포는 이에 딱 맞게 분포

Ball-Tree Algorithm

YSAL x sports2i 협업 프로젝트



접근 방식



Pitcher#1의 Pitch 분포가 Batter의 GB% 분포와 비교적 유사해 보임.
땅볼을 유도해야 하는 상황에서 Pitcher#1을 올리는게 이상적

접근 방식

	L-CV-1	L-CV-2	L-CV-3	L-CV-4	L-CV-5	L-CV-6	L-CV-7	L-FF-1	L-FF-2	L-FF-3	...	R-OS-1	R-OS-2	R-OS-3	R-OS-4	R-OS-5	R-SL-1	R-SL-2	R-SL-3	R-SL-4	R-SL-5
CJ Abrams	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	1.000000	0.000000	0.5	...	0.500000	0.500000	0.666667	0.571429	0.750000	0.454545	0.000000	0.500000	0.400000	1.000000
José Abreu	0.0	0.0	0.0	0.0	0.000000	0.0	1.0	0.272727	0.714286	0.0	...	0.428571	0.800000	0.166667	0.428571	0.250000	0.190476	0.480000	0.519231	0.517241	0.266667
Ronald Acuña Jr.	0.0	0.0	0.0	0.0	0.666667	1.0	0.0	0.000000	0.000000	1.0	...	0.000000	0.642857	0.333333	0.571429	0.500000	0.424242	0.379310	0.645161	0.461538	0.333333
Willy Adames	0.0	0.0	0.0	1.0	0.000000	0.0	1.0	0.250000	0.555556	0.0	...	1.000000	0.500000	0.000000	0.500000	0.428571	0.575758	0.459459	0.414634	0.411765	0.333333
Matt Adams	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	1.000000	0.333333	0.0	...	0.000000	0.250000	1.000000	0.555556	0.333333	1.000000	0.181818	0.500000	0.500000	0.000000
...

총 54개의 Cluster

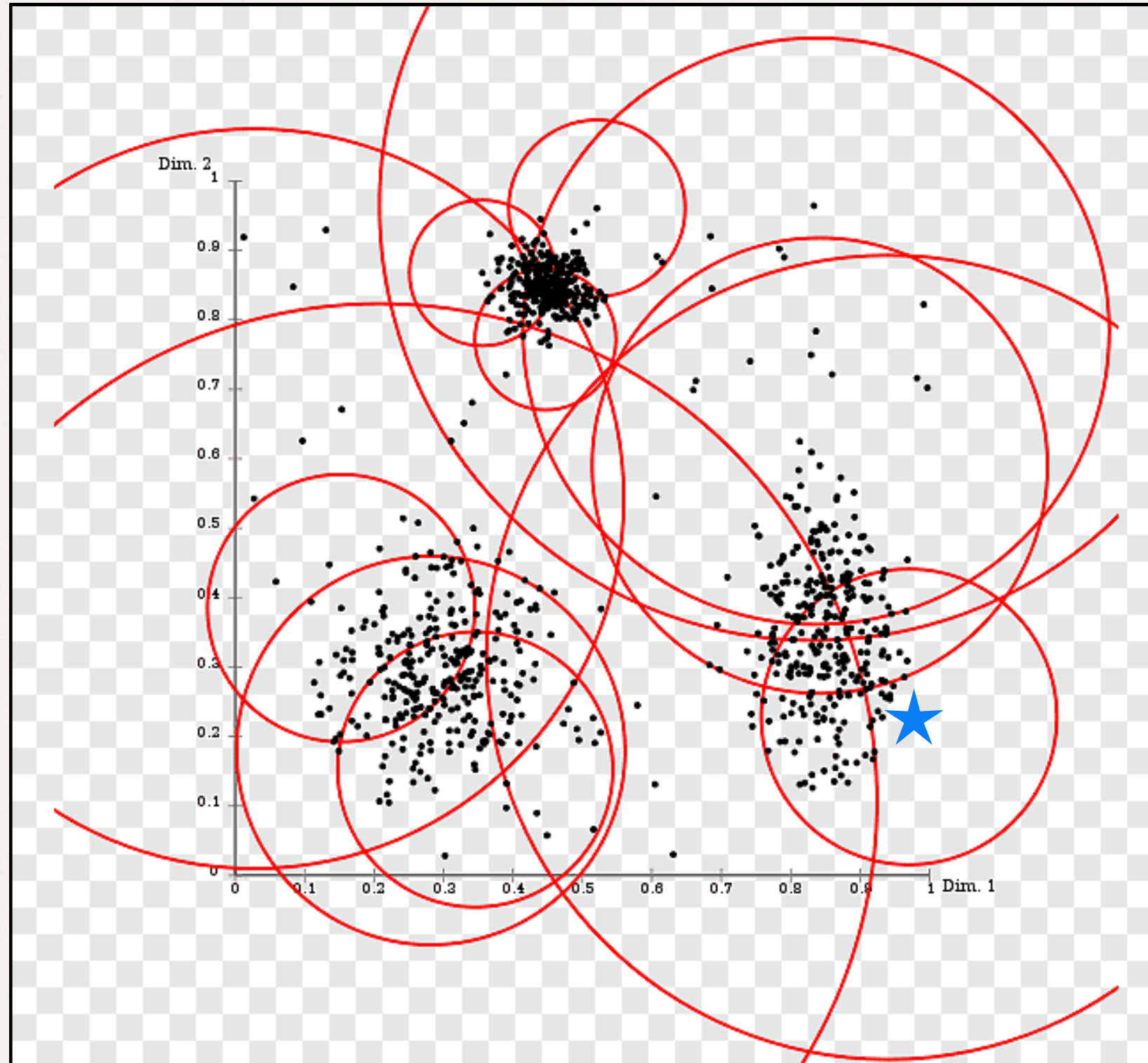
타자의 Cluster별 GB%(or Whiff%) & 투수의 투구 분포 데이터 추출



54차원 공간에 투수 및 타자를 표현

해당 타자와 가장 가까운 거리에 있는 상위 3명의 투수를 출력!

Ball-Tree Algorithm



이진 트리 기법을 활용하여
여러 개의 **hyperspheres**로 데이터 공간을 분할

다른 알고리즘 대비
고차원 데이터에서 매우 효율적인 데이터 구조를 만들어냄.



54차원 공간 상에서
NN(Nearest Neighbor)을 찾기에 용이!

Ball-Tree algorithm 전처리

GB%

[illegible]

Whiff%

[illegible]

Ball-Tree algorithm 전처리

Pitch Dist.

[illegible]

Ball-Tree algorithm 결과

▶	matchup("SD", "Shohei Ohtani", model_type="GB")	
↔		
	Pitcher	NN Distance
	Wandy Peralta	3.58946
	Austin Davis	3.63401
	Jason Adam	3.67388

땅볼 유도 상황

▶	matchup("KC", "Aaron Judge", model_type="WHIFF")	
↔		
	Pitcher	NN Distance
	Chris Stratton	2.86014
	Michael Lorenzen	2.86942
	Seth Lugo	2.87473






헛스윙 유도 상황

결과가 NN-distance로 출력됨!

Ball-Tree Algorithm Matching : GB

Batted Ball Profile														
Season	GB %	FB %	LD %	PU %	Pull %	Straight %	Oppo %	Weak %	Topped %	Under %	Flare/Burner %	Solid %	Barrel %	Barrel/PA
2016	46.2	15.4	30.8	7.7	42.3	30.8	26.9	7.7	30.8	19.2	30.8	7.7	3.8	2.6
2017	55.8	20.4	18.8	5.0	35.9	41.4	22.7	8.3	40.9	22.1	20.4	5.5	2.2	1.5
2018	47.9	15.3	29.4	7.4	30.1	41.1	27.6	1.8	38.0	20.9	32.5	3.1	3.1	2.2
2019	52.5	26.2	20.5	0.8	40.2	36.9	23.0	7.4	40.2	17.2	16.4	7.4	9.0	6.4
2020	47.4	21.1	28.9	2.6	31.6	39.5	28.9	7.9	35.5	21.1	26.3	3.9	3.9	2.6
2021	57.8	14.9	22.7	4.5	39.6	34.4	26.0	4.5	42.9	16.9	23.4	1.9	7.8	5.5
2022	55.7	20.9	15.8	7.6	34.8	36.7	28.5	7.6	42.4	23.4	19.0	0.6	4.4	3.1
2023	57.1	22.1	17.1	3.6	46.4	36.4	17.1	7.1	40.0	17.9	22.9	4.3	7.1	4.4
2024	53.3	18.9	22.1	5.7	42.6	36.9	20.5	6.6	41.8	17.2	20.5	5.7	8.2	6.2
Player	53.7	19.6	21.7	5.0	37.7	37.8	24.3	6.3	40.3	19.7	22.9	4.0	5.5	3.8
MLB	44.4	23.8	24.7	7.1	37.2	37.5	25.2	3.9	32.6	24.8	24.4	5.9	7.0	4.8

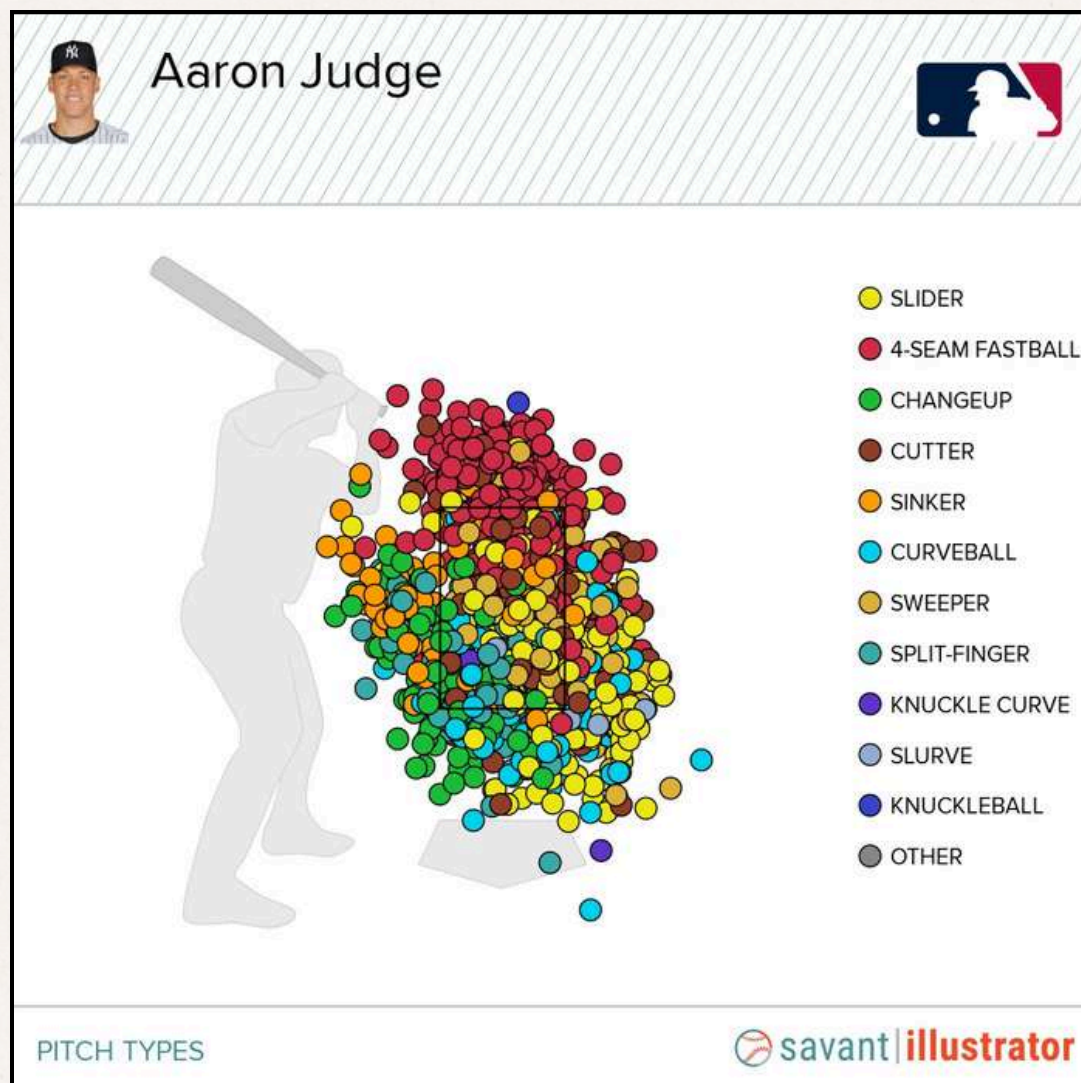
Note: All figures in this table cover the period 2015-present.

Similar Pitchers to Wandy Peralta:  2020 - Phillips Valdez  2020 - Max Fried  2023 - Touki Toussaint  2017 - Steve Cishek  2020 - Yusei Kikuchi --- Compare All Similar Pitchers

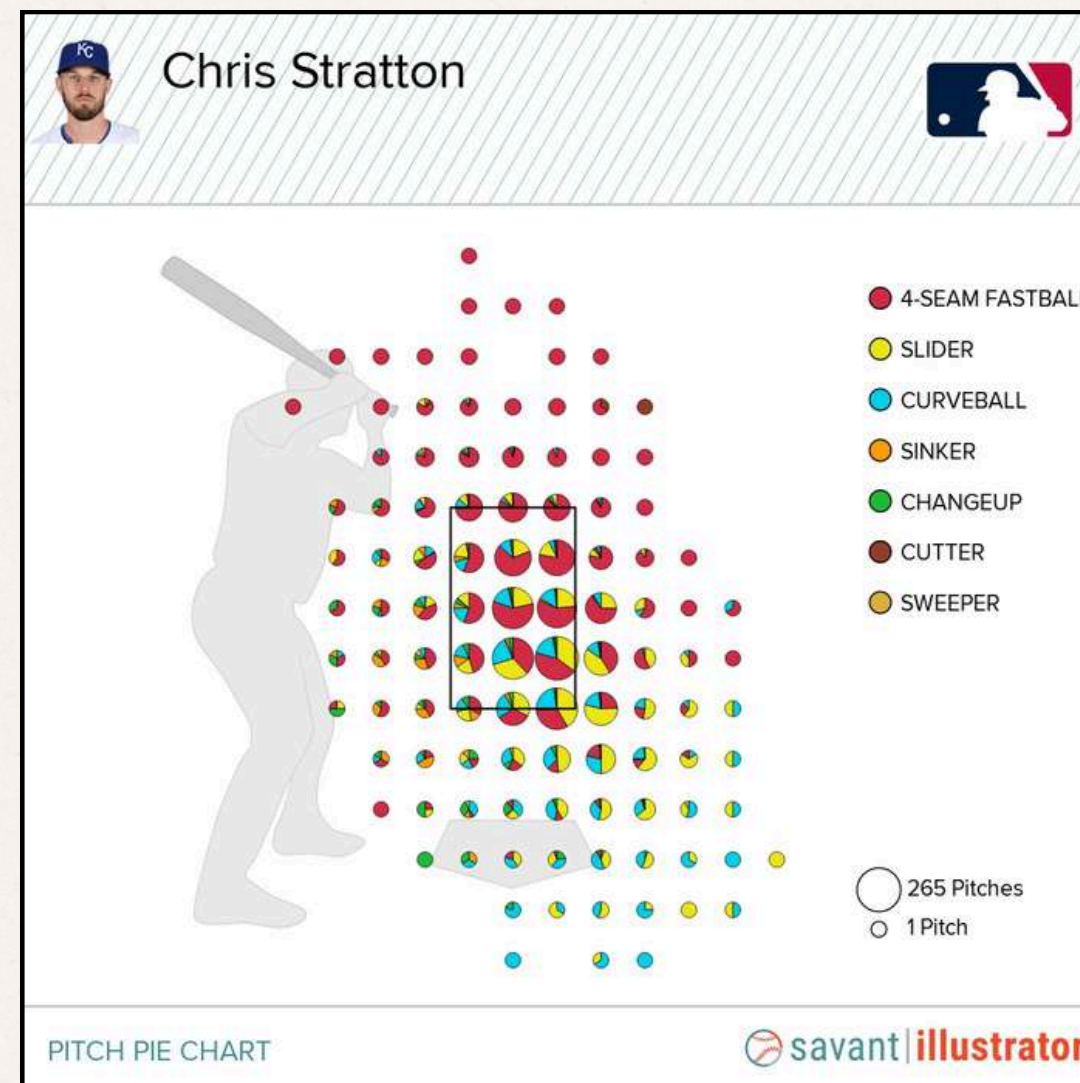
- Wandy Peralta의 MLB 통산 GB%는 53.7%로, MLB 전체 평균 GB%인 44%를 크게 웃뎀.
- 또한, 2018년~2024년의 기간 동안 MLB 좌완 투수 중 전체 9위의 GB%(52.9%)*를 가지고 있음.
- 좌완 투수에게 비교적 약한 Ohtani Shohei를 상대로 적절한 선택지라 할 수 있음.

*출처: <https://www.fangraphs.com/leaders/major-league?pos=all&stats=pit&season=2024&season1=2018&ind=0&team=0&sortcol=13&sortdir=default&hand=L>

Ball-Tree Algorithm Matching : WHIFF



스윙한 구종 데이터
(Aaron Judge vs. RHP 2018~2024)



투구한 구종 분포 데이터
(Chris Stratton vs. RHB, 2018~2024)

Aaron Judge는 높은 존에 형성되는 포심 패스트볼,
바깥쪽에 떨어지는 슬라이더와 커브에 스윙하는 약점



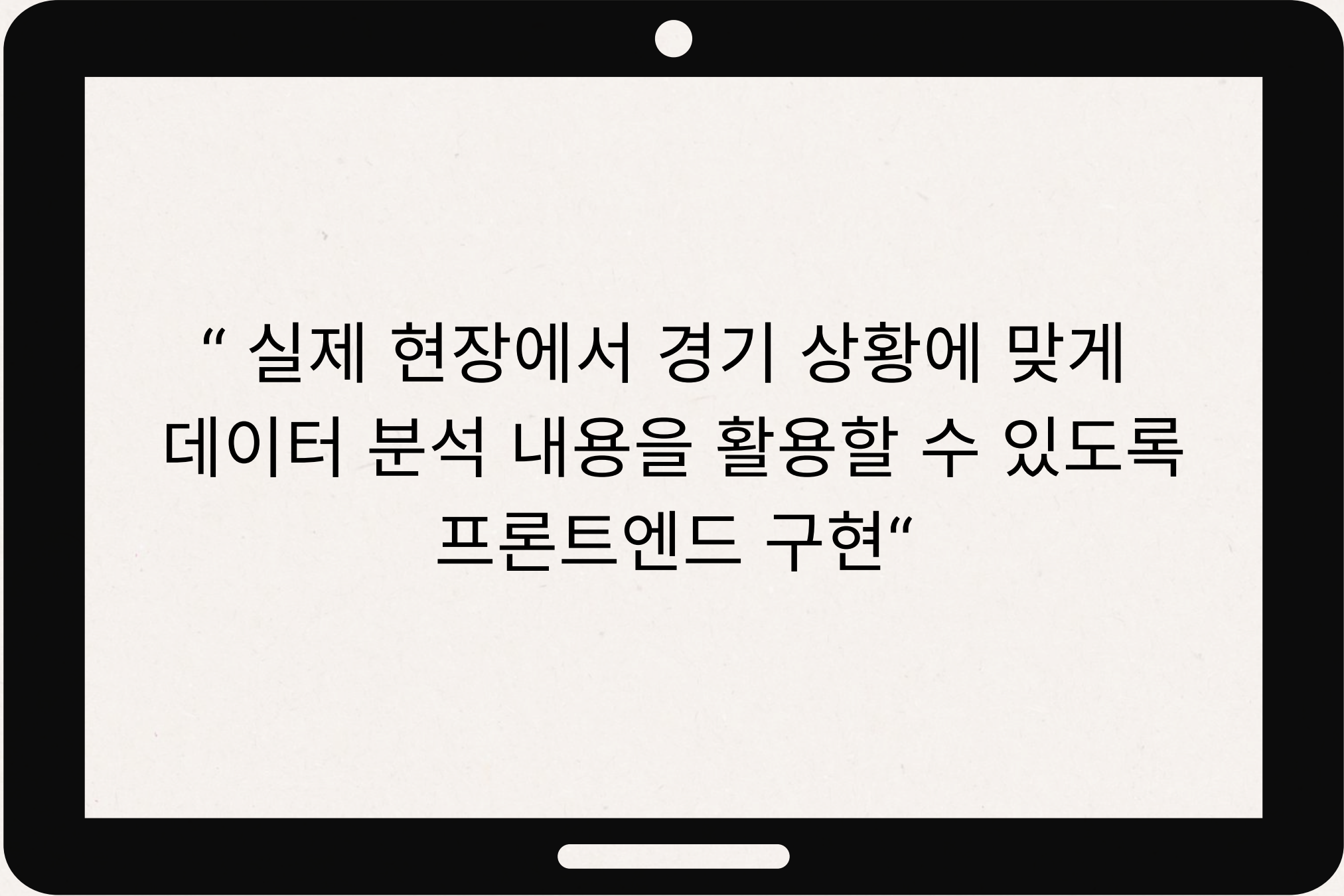
Chris Stratton의 투구 분포는 이에 가깝게 분포

프론트엔드 구현

YSAL x sports2i 협업 프로젝트



프론트엔드 구현 목적



“ 실제 현장에서 경기 상황에 맞게
데이터 분석 내용을 활용할 수 있도록
프론트엔드 구현”

프론트엔드 구현 방법

초기 구상

최적의 불펜 투수 운용

YSAL x sports2i

시작하기

투수의 팀을 선택하세요

YSAL x sports2i

상대 타자를 선택하세요

전체 A B C D E F G

[illegible]

YSAL x sports2i

상황 옵션을 선택하세요

eRE 값 ☒

땅볼 유도 상황

삼진 유도 상황

YSAAL x sports2i

결과



투수	eRE 값

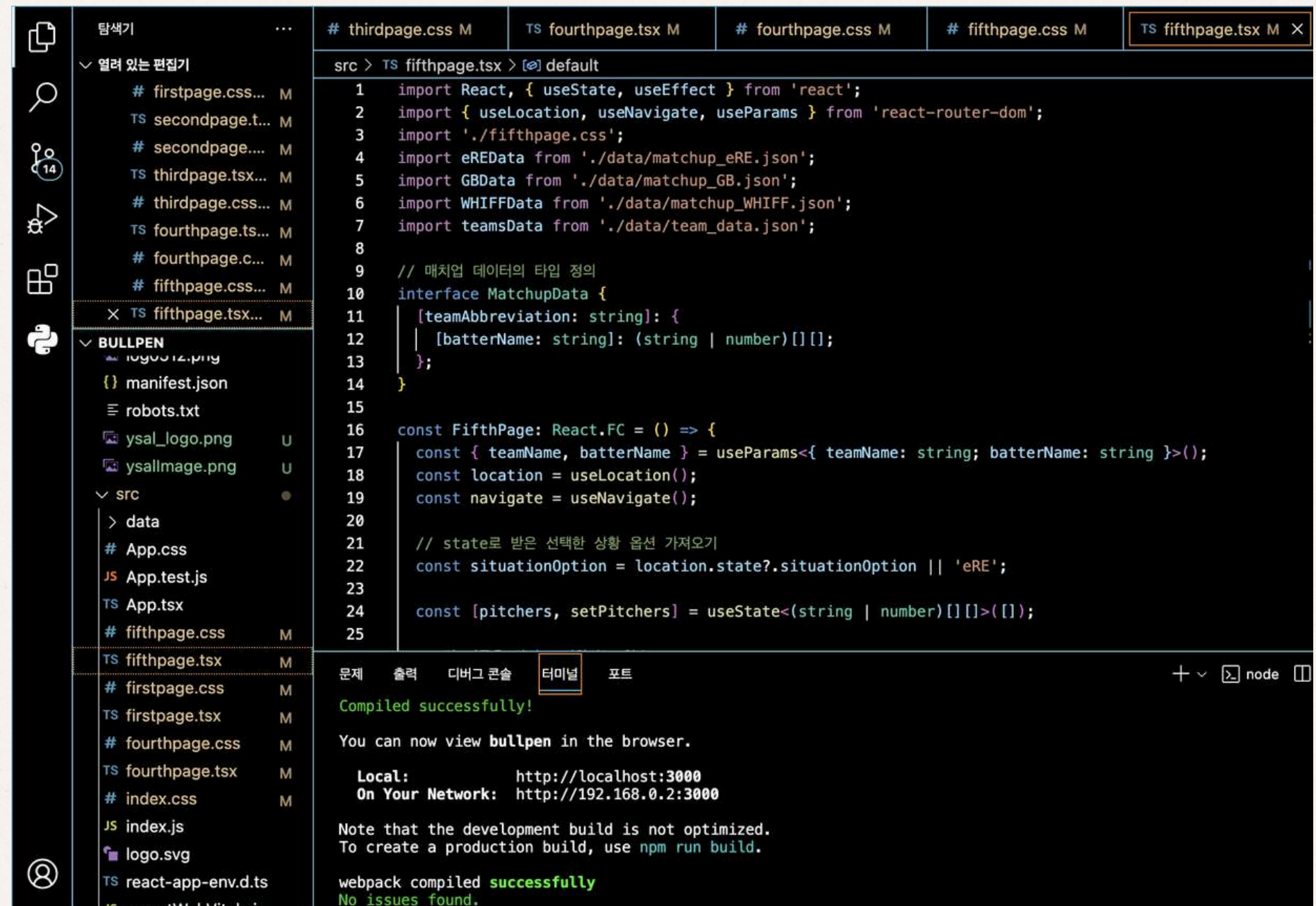
YSAL x sports2i

프론트엔드 구현 방법

실제 구현

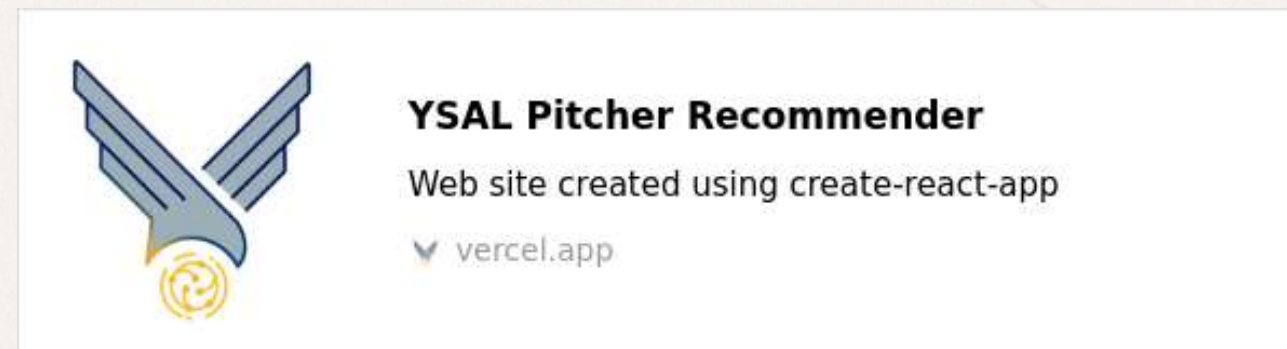
REACT 사용

1. 홈 화면 생성
2. 투수의 팀 선택
3. 상대 타자 선택
4. 구체적인 상황 옵션 선택
5. 최종 결과 출력



프론트엔드 구현

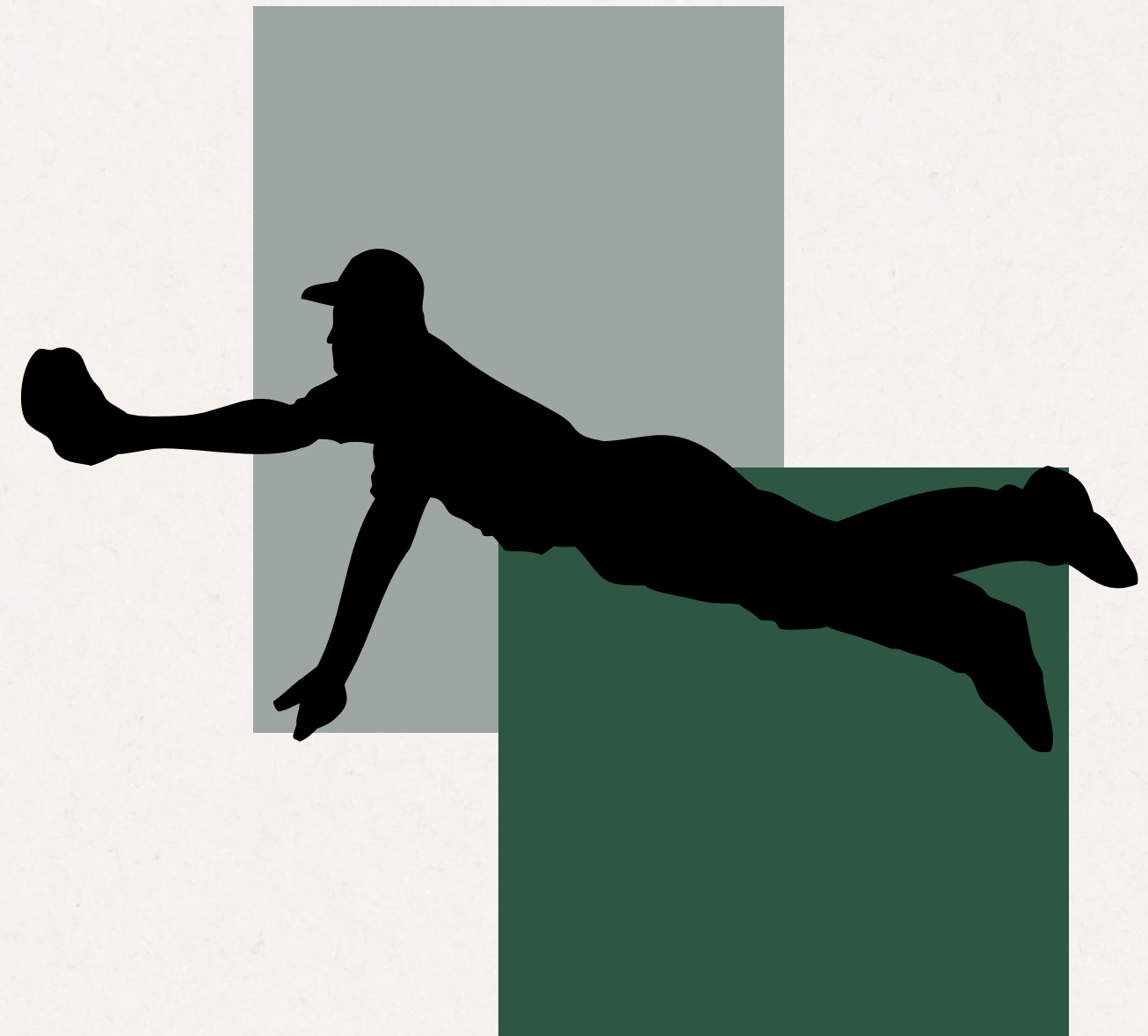
실제 구현



<https://ysal-bullpen-recommend.vercel.app/>

실제 사례 분석

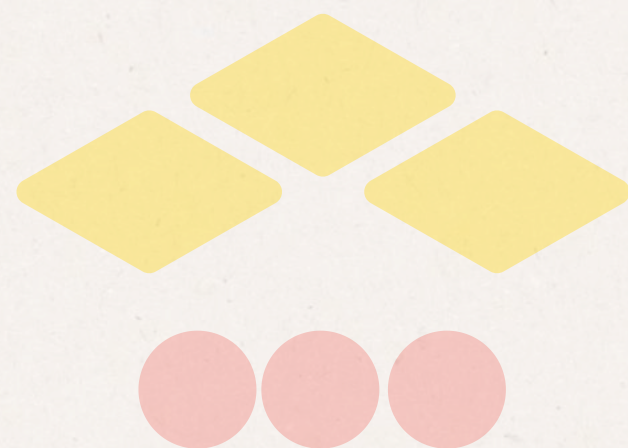
YSAL x sports2i 협업 프로젝트



eΔRE case

2024 NLDS1 **NYM** vs. **PHI** 1차전

Game Situation



- 8회말 0out **PHI** 공격
- **NYM**가 6 대 2로 앞서고 있는 상황
- **PHI**의 중심타선 상대로 기대득점 억제 필요성

+

Batter Up

Kyle Schwarber
Trea Turner
Bryce Harper

Model's Recommendation : **Phil Maton !**

vs. Kyle Schwarber

최종 결과 (eΔRE)	
Alex Young	-1.5193254577687512
Brooks Raley	-1.1775489831752393
Jake Diekman	-0.7730353327123393

vs. Trea Turner

최종 결과 (eΔRE)	
Brooks Raley	-2.065912142360885
Phil Maton	-1.1037210558439223
Edwin Díaz	-0.8433183013792372

vs. Bryce Harper

최종 결과 (eΔRE)	
Alex Young	0.5505634781452635
Ryne Stanek	0.8452148330976035
Phil Maton	0.9649946842734488

*Alex Young, Brooks Raley, Jake Diekman NLDS 로스터 제외

eΔRE case

2024 NLDS1 NYM vs. PHI 1차전

NYM's Choice & Result

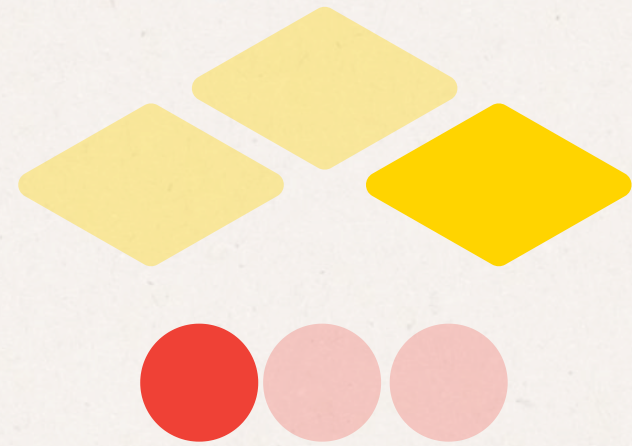
63	▼ 8	K Schwarber	P Maton	0	___	1-5	Kyle Schwarber struck out looking.	0.60	0.50	2.7%	-.015	-0.24
64	▼ 8	T Turner	P Maton	1	___	1-5	Trea Turner struck out swinging.	0.35	0.26	1.8%	-.009	-0.16
65	▼ 8	B Harper	P Maton	2	___	1-5	Bryce Harper doubled to right (Liner).	0.15	0.10	2.7%	.009	0.22
66	▼ 8	N Castellanos	P Maton	2	_2_	1-5	Nick Castellanos singled to left (Grounder). Bryce Harper advanced to 3B.	0.45	0.32	4.1%	.013	0.18
67	▼ 8	A Bohm	P Maton	2	1_3	1-5	Alec Bohm reached on fielder's choice to third. Nick Castellanos out at second.	0.94	0.50	1.4%	-.026	-0.50

- 이 날 홈런을 기록했던 Kyle Schwarber를 삼진으로 처리하는 등 무실점으로 이닝을 마무리
- 상대적으로 높은 eΔRE를 기록했던 Bryce Harper를 상대로는 2루타를 허용
- 그러나, 모델의 제안에 걸맞게 PHI의 기대 득점을 최소화하는데 성공

Ground Ball case

2024 NLWC2 ATL vs. SD 1차전

Game Situation



- 2회말 1out SD 공격
- SD가 3 대 0으로 앞서고 있는 상황
- 땅볼 유도를 통한 더블 플레이가 목표

+

Batter Up

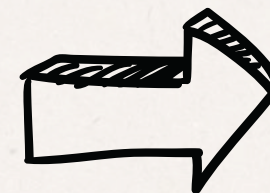
Luis Arraez

(2024시즌 삼진율 3.4%)

Model's Recommendation

최종 결과 (NN Distance)

Aaron Bummer	3.8897031416437278
Jesse Chavez	3.913338139660593
Joe Jiménez	3.917510863051566



Aaron Bummer !

Ground Ball case

2024 NLWC2 ATL vs. SD 1차전

ATL's Choice & Result

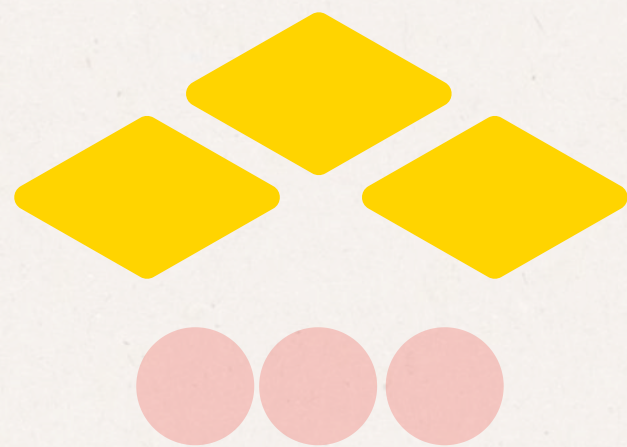
Bottom of the 2nd, Padres Batting, Ahead 2-0, Braves' AJ Smith-Shawver facing 7-8-9											
b2	2-0	0	---	1,(0-0)		SDP	Jake Cronenworth	AJ Smith-Shawver	2%	77%	Hit By Pitch
b2	2-0	0	1--	5,(2-2)		SDP	Donovan Solano	AJ Smith-Shawver	6%	84%	Single to RF (Line Drive to RF Line); J. Cronenworth to 3B
b2	2-0	0	1-3	2,(0-1)	RO	SDP	Kyle Higashioka	AJ Smith-Shawver	-1%	82%	Flyball: CF/Sacrifice Fly (Deep CF-RF); J. Cronenworth Scores
									Aaron Bummer replaces AJ Smith-Shawver pitching		
b2	3-0	1	1--	2,(0-1)	OO	SDP	Luis Arráez	Aaron Bummer	-3%	80%	Ground Ball Double Play: 2B-1B (SS-2B)
1 run, 1 hit, 0 errors, 0 LOB. Braves 0, Padres 3.											

- 추가 실점 시, 경기 초반인 2회부터 경기 운영이 어렵게 될 수 있는 상황
- 더블 플레이 유도 통해 위기 탈출
- TMI) Aaron Bummer의 지난 6년간 GB%는 65.7%로 MLB 좌완 전체 1위이다.

Strike Out case

20240506 SEA vs. MIN

Game Situation



- 7회초 0out SEA 공격
- MIN이 1대 0으로 앞서고 있는 상황
- 절대적으로 삼진을 잡아 실점을 억제해야하는 상황

Batter Up
Ty France

Model's Recommendation

최종 결과 (NN Distance)	
Griffin Jax	1.7603150637127531
Diego Castillo	1.7855462775069402
Trevor Richards	1.7864181308781633



Griffin Jax !

Strike Out case

20240506 SEA vs. MIN

MIN's Choice & Result

Top of the 7th, Mariners Batting, Behind 0-1, Twins' Griffin Jax facing 3-4-5												
									Griffin Jax replaces Simeon Woods Richardson pitching			
t7	0-1	0	---	7,(3-2)		SEA	Jorge Polanco	Griffin Jax	-7%	62%	Single to RF (Line Drive to Short CF-RF)	
t7	0-1	0	1--	6,(3-2)		SEA	Mitch Haniger	Griffin Jax	-17%	45%	Single to CF (Ground Ball); J. Polanco to 3B	
t7	0-1	0	1-3	7,(3-2)		SEA	Cal Raleigh	Griffin Jax	-7%	39%	Walk; M. Haniger to 2B	
t7	0-1	0	123	4,(1-2)	O	SEA	Ty France	Griffin Jax	12%	51%	Strikeout Swinging	
t7	0-1	1	123	3,(1-1)	RO	SEA	Mitch Garver	Griffin Jax	0%	51%	Flyball: CF/Sacrifice Fly (Deep CF); J. Polanco Scores	
t7	1-1	2	12-	4,(0-2)	O	SEA	Luke Raley	Griffin Jax	8%	59%	Strikeout Swinging	
									1 run, 2 hits, 0 errors, 2 LOB. Mariners 1, Twins 1.			

- 2연속 안타 및 볼넷을 허용해 만루의 위기에 처함
- But, 모델의 추천 결과에 맞게 Ty France를 삼진 처리
- 결과적으로 무사 만루(평균 2.304점의 기대 득점*)의 상황에서 1점만 허용하며 실점 최소화

*출처: <https://tbt.fangraphs.com/bases-loaded-nobody-out-what-happens-next/>

한계점 및 발전과제

YSAL x sports2i 협업 프로젝트



한계점 및 발전과제

Data Bias

데이터 개수를 유의미하게 증가시켰음에도 불구하고 희소한 투구 데이터의 경우, GB%, Whiff% 값이 극단적으로 표현됐다.

>>> 이를 보정할만한 방법론 고민

Pitcher's Condition & Form

투수 분류에 컨디션이나 현재 시즌의 폼과 같은 정보가 반영되지 않았다.

>>> 투수의 컨디션과 폼을 수치화할 수 있을까?

Front-end

투수 매칭 프로그램에서 eΔRE값이나, NN distance 값 뿐만 아니라, 실제 상대전적 등 유의미한 데이터들을 추가로 제시해 더욱 효과적인 결과 도출

>>> 상대 전적 등 매일 수시로 바뀌는 데이터를 프론트엔드에 반영해야 할 필요성

How About KBO?

MLB에 비해 비교적 세이버매트릭스 데이터가 부족한 KBO에서 이를 적용할 수 있을까?

>>> KBO형 투수 매칭 프로그램 제작



감사합니다

Thank You

