

Revenue growth divisions.



Projected sales of main products in 2013

Baseball Barrels

**OPS = OBP + SLG**

	TYU division			FRT division		
GHT	254	550	254	274	154	415
RDW	650	320	754	273	825	154
TRG	241	450	144	364	954	174
RTG	254	650	874	657	125	274
WEF	784	125	124	752	741	759
WRT	152	784	954	241	741	245



## C o n t e n t s

01 배럴 타구 선정

02 사용 데이터

03 사용 알고리즘

04 예측 결과

05 기대 효과

투표 결과를 참고하여 KBO만의 배럴기준 정의  
130km 이상의 타구 속도가 15 ~ 30 발사각  
150km 이상의 타구 속도가 30 ~ 45 발사각

응답	응답수
발사각 0~15도, 타구속도가 130km 이상인 타구	4 0%
발사각 0~15도, 타구속도가 150km 이상인 타구	18 8%
발사각 15~30도, 타구속도가 130km 이상인 타구	14 28%
발사각 15~30도, 타구속도가 150km 이상인 타구	25 50%
그 외	2 4%



# 사용 데이터 - 외부 데이터

STATIZ

검색...

스탯티즈+

스카이박스

기록실

경기일정

시즌정보

팀정보

선수정보

스페셜

추가정보

50

강백호

KT, 1루수

32

김재환

두산, 좌익수

22

김현수

LG, 좌익수

27

로맥

SSG, 1루수

이정후

키움, 중견수

37

박건우

두산, 중견수

25

양의지

NC, 지명타자

8

전준우

롯데, 좌익수

55

채은성

LG, 우익수

14

최정

SSG, 3루수

KBO 리그의 세이버메트릭스를 전문적으로 다루는 야구 통계 사이트  
-> 2018년 ~ 2021년도 선수별 필요 데이터 추출

Weka

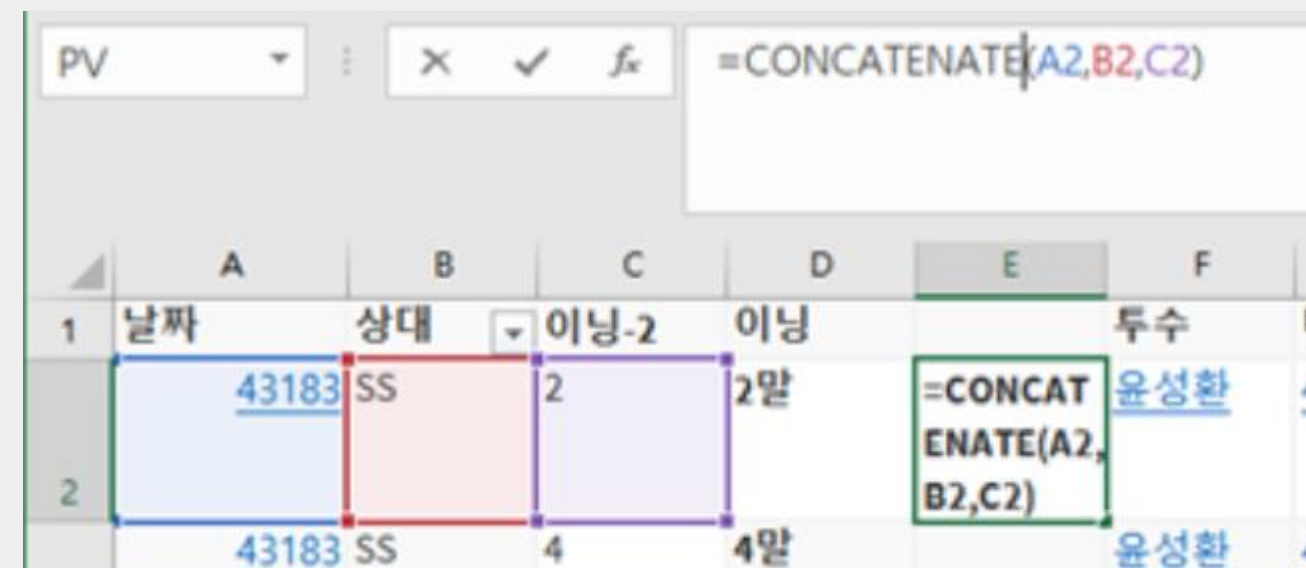
# Weka - 전처리 과정

해당 선수 Playlog를 2018년 ~ 2021년 엑셀 데이터로 저장



변경 전	변경 후
SSG	SK
삼성	SS
롯데	LT
두산	OB
KIA	HT
한화	HH
넥센, 키움	WO
서군	삭제

제공 팀명과 동일하게 맞춤



제공 데이터와 VLOOKUP함수에 필요한 조건행 생성



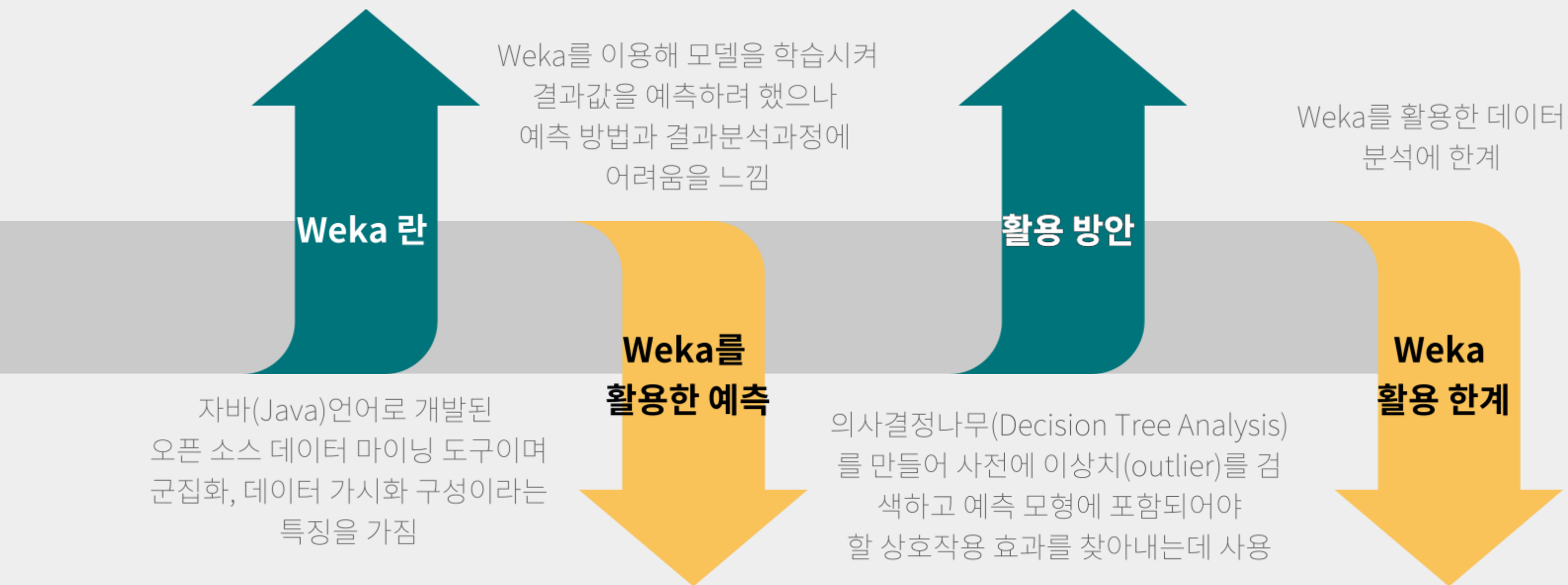
VLOOKUP함수 사용하여 통합

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	날짜	상대팀	타자	날짜	상대	상대투수	볼카운트	아웃카운트	HIT_VEL	HIT_ANGLE	HIT_RESULT_VEL	STADIUM	Target				
2	43183 SS		4 43183SS4 윤성환	0-0	2사		105.03	63.8	볼라이	114.22	정상	F					
3	43183 SS		7 43183SS7 윤성환	0-0	무사		154.38	19.7	볼라이	127.07	정상	T					
4	43184 SS		1 43184SS1 이명만	0-0	1사		151.37	12.6	1루타	136.03	정상	F					
5	43184 SS		5 43184SS5 이명만	1-0	2사		160.8		7 땅볼아웃	143.35	정상	F					
6	43184 SS		8 43184SS8 최정민	0-1	무사		155.16	35.7	볼라이	140.64	정상	T					
7	43186 LT		1 43186LT1 레일리	02월 01일	1사		152.85	32.1	피넌볼라이	128.04	정상	T					
8	43187 LT		3 43187LT3 김원호	0-0	1사		151.55	39.9	홈런	128.02	정상	T					
9	43187 LT		5 43187LT5 김원호	02월 01일	1사		101.78	0.7	땅볼아웃	129.53	정상	F					
10	43188 LT		3 43188LT3 송승준	01월 01일	2사		120.78	50.6	볼라이	141.42	정상	F					
11	43188 LT		6 43188LT6 송승준	02월 01일	무사		165.28	9	1루타	141.41	정상	F					
12	43188 LT		8 43188LT8 강명호	0-1	무사		127.96	40.4	볼라이	130.27	정상	F					
13	43189 KT		5 43189KT5 피터밴드	2-0	무사		103.47	-4.8	땅볼아웃	113.74	수행	F					
14	43189 KT		7 43189KT7 윤성환	1-0	1사		120.49	9.4	직간타	124.06	수행	F					
15	43190 KT		3 43190KT3 주경	02월 02일	무사		153.17	28.8	홈런	116.25	수행	T					
16	43191 KT		3 43191KT3 고영표	1-0	1사		141.58	-11.4	땅볼아웃	135.82	수행	F					
17	43191 KT		9 43191KT9 임상혁	01월 01일	2사		130.19	48	볼라이	150.47	수행	F					
18	43193 LG		2 43193LG2 소사	01월 02일	무사		122.1	6.3	땅볼아웃	137.88	정상	F					
19	43193 LG		5 43193LG5 소사	0-0	2사		168.27	12.8	1루타	134.09	정상	F					
20	43197 NC		4 43197NC4 박병	2-0	무사		162.63	28.3	홈런	141.98	정상	T					
21	43197 NC		5 43197NC5 박병	0-2	2사		138.13	44.9	볼라이	135.44	정상	F					
22	43198 NC		1 43198NC1 박준강	0-2	2사		128.29	29.5	2루타	123.46	정상	F					
23	43198 NC		5 43198NC5 박준강	01월 02일	무사		160.09	10.2	1루타	139.81	정상	F					
24	43198 NC		9 43198NC9 박준강	01월 01일	1사		138.8	6.7	1루타	128.34	정상	F					
25	43200 SS		1 43200SS1 박민태	0-1	2사		168.89	30.7	홈런	142.1	대구	T					
26	43200 SS		5 43200SS5 박민태	01월 01일	2사		158.84	-4.7	1루타	129.55	대구	F					
27	43201 SS		1 43201SS1 양정호	01월 02일	1사		105.78	-21.9	땅볼아웃	130.25	대구	F					
28	43201 SS		3 43201SS3 양정호	01월 01일	무사		157.87	9.6	2루타	138.38	대구	F					
29	43202 SS		2 43202SS2 박정현	01월 03일	무사		109.13	49.5	2루타	126.63	대구	F					

Weka를 통해 투수 분석 한계

아이티DUCK

# Weka - 결과





# 사용 데이터 - 전처리 과정

제공 데이터에서 타격 결과 데이터, 선수 정보 데이터를 통합한 후 배럴 기준에 맞는 Target 변수 생성

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	GYEAR	G_ID	PIT_ID	PCODE	NAME	MONEY	AGE_VA	T_ID	Rival	INN	HIT_VEL	HIT_ANG_V	HIT_RESULT	PIT_VEL	STADIUM	Target	
2	2018	20180324	180324_14	62797	양성우	10000만원	29 HH	WO		1	131.5	42.7	플라이	149.59	고척	F	
3	2018	20180324	180324_14	76753	송광민	24000만원	35 HH	WO		1	135.18	9.9	1루타	148.78	고척	F	
4	2018	20180324	180324_14	71752	김태균	140000만원	36 HH	WO		1	152.41	2.1	1루타	148.59	고척	F	
5	2018	20180324	180324_14	62700	하주석	12000만원	24 HH	WO		1	113.72	13.5	1루타	139.13	고척	F	
6	2018	20180324	180324_14	68730	호잉	400000달러	29 HH	WO		2	54.11	16.4	번트안타	134.07	고척	F	
7	2018	20180324	180324_14	78288	최재훈	8000만원	28 HH	WO		2	161.9	3.5	1루타	145.75	고척	F	

위 데이터 테이블의 컬럼 정의서

컬럼	컬럼명	비고
GYEAR	연도	
G_ID	경기코드	
PIT_ID	타구코드	각 타구가 가지는 고유의 코드(구분 값)
PCODE	선수코드	
NAME	선수명	
MONEY	연봉	
AGE_VA	나이	
T_ID	팀코드	선수 소속팀의 고유 코드 값
Rival	상대팀	
INN	이닝	
HIT_VEL	타구속도(km/h)	
HIT_ANG_VER	발사각도	
HIT_RESULT	타격결과	타격 결과 항목(총 17가지) 1> 1루타 2> 2루타 3> 3루타 4> 네아안타(1루타) 5> 땅볼아웃 6> 번트아웃 7> 번트안타 8> 병살타 9> 삼중살타 10> 야수선택 11> 인플레이플라이 12> 직선타 13> 파울플라이 14> 플라이 15> 홈런 16> 희생번트 17> 희생플라이
PIT_VEL	상대투수 투구속도(km/h)	
STADIUM	해당구장	
Target	배럴타구 여부	T : 두 기준 중 하나 이상 만족 130km 이상의 타구 속도가 15 ~ 30 발사각 150km 이상의 타구 속도가 30 ~ 45 발사각 F : 두 기준 모두 만족하지 않음

# 사용 데이터 - 전처리 과정

- 볼넷, 사구, 고의4구, 출루율 데이터를 활용하여 OPS 예측에 사용

32 김재환 -  
두산, 좌익수

종합

연도별

그래프

날짜별

상황별

상대별

Playlog

상세분석

연봉

스탯위키

타자

투수

수비

정규

포스트

울스타

없음

100G

126G

133G

144G

162G

기본

확장

가치

클러치

타석

타구1

타구2

파워

팀배팅1

팀배팅2

도루

주루

구종가치

구종구사

연도	팀	나이	P	C	타석	타수	득점	안타	2타	3타	홈런	루타	타점	도루	도실	볼넷	사구	고의4	삼진	병살	회타	회비	비율					WAR*	WPA	
																							타율	출루	장타	OPS	wOBA	wRC+		
2008	두산	20	DH	14	21	21	0	3	0	0	0	3	0	0	1	0	0	0	6	0	0	0	0.143	0.143	0.143	0.286	0.136	-36.4	-0.34	
2011	두산	23	DH	30	56	54	4	10	3	0	2	19	8	1	1	1	1	0	14	2	0	0	0.185	0.214	0.352	0.566	0.258	46.7	-0.27	
2012	두산	24	C	13	43	39	7	5	1	0	1	9	1	0	0	4	0	0	9	0	0	0	0.128	0.209	0.231	0.440	0.215	23.7	-0.09	
2014	두산	26	C	52	96	85	12	26	6	0	3	41	13	1	0	11	0	1	12	3	0	0	0.306	0.385	0.482	0.868	0.386	127.2	0.81	-0.27
2015	두산	27	1B	48	180	153	24	36	8	0	7	65	22	4	1	22	2	1	39	6	1	2	0.235	0.335	0.425	0.760	0.336	100.9	0.48	0.59
2016	두산	28	LF	134	568	492	107	160	32	3	37	309	124	8	2	71	0	3	107	9	0	5	0.325	0.407	0.628	1.035	0.442	162.7	5.75	5.71
2017	두산	29	LF	144	636	544	110	185	34	2	35	328	115	4	1	81	7	10	123	8	0	4	0.340	0.429	0.603	1.032	0.440	170.7	7.45	7.45
2018	두산	30	LF	139	602	527	104	176	36	1	44	346	133	2	0	59	9	4	134	6	0	7	0.334	0.405	0.657	1.062	0.438	168.8	6.94	5.80
2019	두산	31	LF	136	573	494	76	140	20	5	15	215	91	3	1	63	5	5	113	7	0	11	0.283	0.363	0.435	0.798	0.362	124.0	3.56	2.53
2020	두산	32	LF	140	613	515	82	137	26	1	30	255	113	6	0	91	1	3	154	14	0	6	0.266	0.374	0.495	0.869	0.385	135.4	4.72	2.33
2021	두산	33	LF	97	403	337	59	95	15	1	20	172	79	2	2	57	5	1	91	4	0	4	0.282	0.390	0.510	0.900	0.408	151.9	3.70	2.73



# 사용 데이터 - 전처리 과정

- 각 선수별 배럴 타구 여부에 따른 평균 루타를 계산 - 추후 사용

## 강백호

T인 경우 (232개)

안타 22, 2루타 54, 3루타 3, 홈런 70개, 희생타 4, 범타 83 => 평균 루타 : 1.78루타

F인 경우 (941개)

안타 319, 2루타 59, 3루타 2, 홈런 5, 희생타 9, 범타 547  
+ 볼넷 245, 사구10, 고4 21 => 평균 루타 : 0.38루타

EX) 강백호 선수

T인 경우  $(22 + 54 \times 2 + 3 \times 3 + 70 \times 4) / (22 + 54 + 3 + 70 + 4 + 83) = 1.78$ 루타

F인 경우  $(319 + 59 \times 2 + 3 \times 2 + 5 \times 4) / (319 + 59 + 2 + 5 + 9 + 547 + 245 + 10 + 21) = 0.38$ 루타

강백호 선수 T: 1.78루타, F: 0.38루타

김재환 선수 T: 1.64루타, F: 0.33루타

김현수 선수 T: 1.26루타, F: 0.36루타

로 맥 선수 T: 1.97루타, F: 0.30루타

박건우 선수 T: 1.17루타, F: 0.46루타

양의지 선수 T: 1.56루타, F: 0.36루타

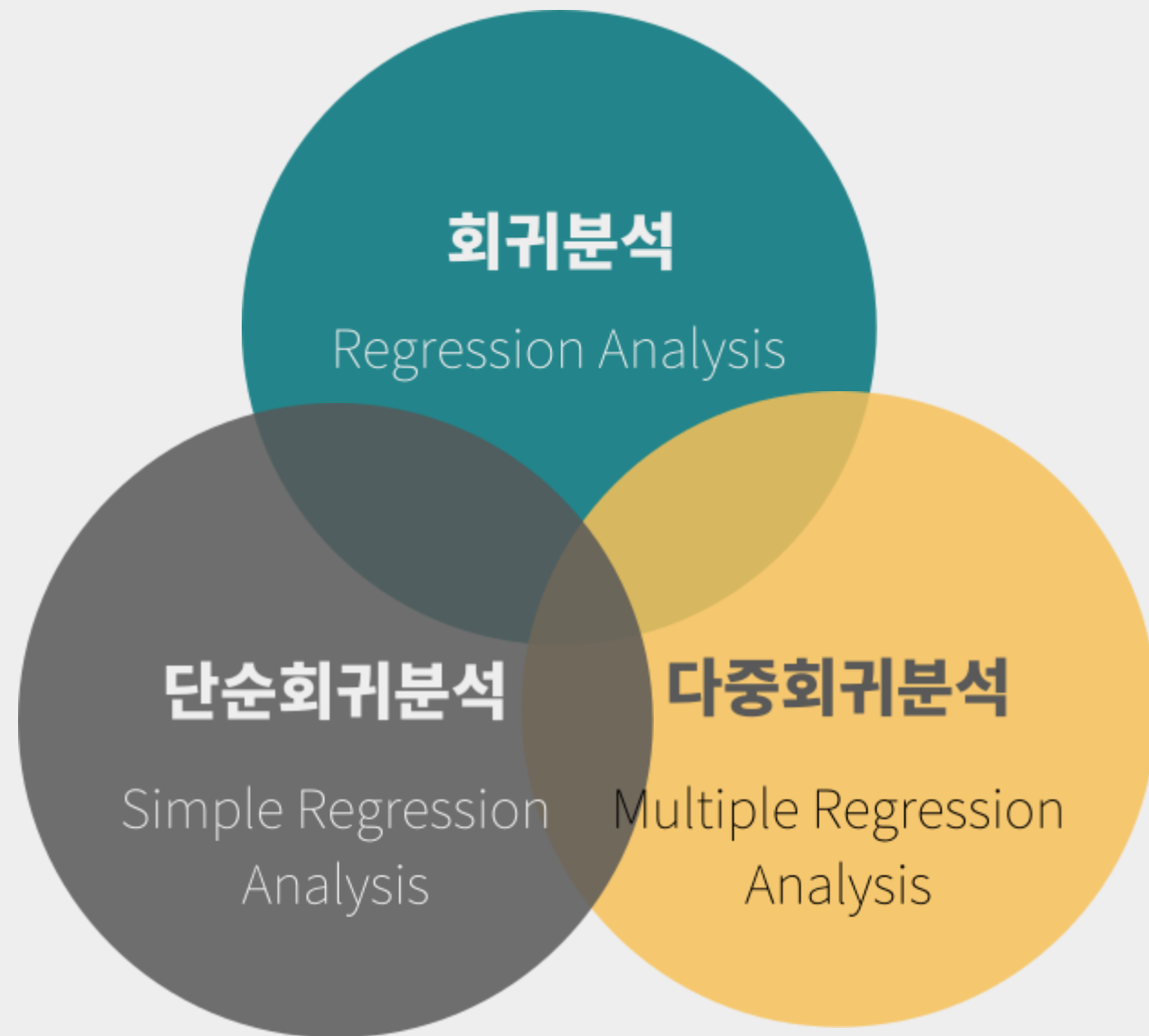
이정후 선수 T: 0.93루타, F: 0.48루타

전준우 선수 T: 1.57루타, F: 0.46루타

채은성 선수 T: 1.26루타, F: 0.36루타

최 정 선수 T: 1.91루타, F: 0.27루타

# 사용 알고리즘



## 회귀분석 선정 이유

01

### 데이터 야구의 주축

데이터 야구를 위한 주축이 결국은 과거의 데이터를 바탕으로 예측하는 것

02

### 상관관계 분석

타자들의 타구 결과를 예측하는데 영향을 끼치는 요소들의 상관관계를 쉽게 파악 가능

03

### 활용 방법 간단

데이터 분석을 하는데 있어 Excel을 활용하므로 높은 프로그래밍 지식을 요구하지 않음

# 예측 방법 - 장타율(SLG)

## 엑셀 피벗 테이블 활용

강백호	평균루타	T	1.78						
		F	0.38						
개수 : Target	열 레이블							해당 경기	
행 레이블	F	T	총합계	잔여경기	잔여타석	T 개수	F 개수	루타 합	
HH	78.18%	21.82%	100.00%	2					
대전	78.43%	21.57%	100.00%	0					
수원	77.97%	22.03%	100.00%	2	6.2	1.366	4.834	4.269	
HT	81.51%	18.49%	100.00%	2					
한주	79.69%	20.31%	100.00%	2	6.2	1.259	4.941	4.119	
수원	83.64%	16.36%	100.00%	0					
LG	81.29%	18.71%	100.00%	2					
수원	88.41%	11.59%	100.00%	2	6.2	0.719	5.481	3.362	
잠실	74.29%	25.71%	100.00%	0					
LT	78.77%	21.23%	100.00%	5					
사직	71.19%	28.81%	100.00%	3	9.3	2.680	6.620	7.286	
수원	83.91%	16.09%	100.00%	2	6.2	0.998	5.202	3.753	
NC	79.14%	20.86%	100.00%	4					
마산	88.89%	11.11%	100.00%	0					
수원	72.06%	27.94%	100.00%	2	6.2	1.732	4.468	4.781	
창원	84.91%	15.09%	100.00%	2	6.2	5.264	5.264	11.371	
OB	78.74%	21.26%	100.00%	3					
수원	81.67%	18.33%	100.00%	2	6.2	1.137	5.063	3.947	
잠실	76.12%	23.88%	100.00%	1	3.1	0.740	2.360	2.214	
SK	84.67%	15.33%	100.00%	2					
문학	80.56%	19.44%	100.00%	2	6.2	1.206	4.994	4.044	
수원	89.23%	10.77%	100.00%	0					
SS	79.86%	20.14%	100.00%	0					
대구	88.31%	11.69%	100.00%	0					
수원	69.35%	30.65%	100.00%	0					
WO	79.49%	20.51%	100.00%	2					
고척	82.14%	17.86%	100.00%	0					
수원	77.05%	22.95%	100.00%	2	6.2	1.423	4.777	4.348	
총합계	80.22%	19.78%	100.00%	22					
						예측	루타 합	53.494	
							타석 합	68.200	
							장타율	0.784	

예측해야할 선수별로 분리한 데이터셋을 활용하였습니다.

EX) 강백호 선수

1. 각 팀별 예측기간 동안의 잔여경기를 계산
2. 사전에 정의한 Target 변수를 활용
3. 과거 어떤 상대팀과 어느구장 경기인지에 따른 배럴타구 비율 도출
4. 한 경기당 타석은 규정타석(3.1)로 정의한 후 잔여 경기에 따른 잔여 타석 계산
5. (잔여 타석) \* (그 상황에 맞는 T, F 비율) 계산을 통해 각 경기당 T, F 개수 예측
6. (사전에 구한 T, F 상황별 루타) \* (T, F 예측 개수) 계산을 통해 해당 경기 루타 합 계산
7. SUM(루타 합) / SUM(잔여타석) 결과를 통해 장타율 예측



# 예측 결과 - 장타율(SLG)

강백호 선수(KT 위즈)

예측	루타 합	53.494
	타석 합	68.200
	장타율	0.784

김재환 선수(두산 베어스)

예측	루타 합	46.644
	타석 합	65.100
	장타율	0.716

김현수 선수(LG 트윈스)

예측	루타 합	38.303
	타석 합	68.200
	장타율	0.562

로맥 선수 (SSG 랜더스)

예측	루타 합	54.007
	타석 합	77.500
	장타율	0.697

박건우 선수(두산 베어스)

예측	루타 합	39.922
	타석 합	65.100
	장타율	0.613

양의지 선수(NC 다이노스)

예측	루타 합	50.035
	타석 합	80.600
	장타율	0.621

이정후 선수(키움 히어로즈)

예측	루타 합	36.571
	타석 합	65.100
	장타율	0.562

전준우 선수(롯데 자이언츠)

예측	루타 합	38.946
	타석 합	74.400
	장타율	0.523

채은성 선수(LG 트윈스)

예측	루타 합	37.653
	타석 합	68.200
	장타율	0.552

최정 선수(SSG 랜더스)

예측	루타 합	49.823
	타석 합	77.500
	장타율	0.643

# 예측 방법 - 출루율(OBP)

- 배럴 타구를 통한 OPS 예측이 주제이지만 배럴 타구의 경우 타격을 하였을 때 적용되는 기준이라고 판단
- 볼넷, 사구, 고의 4구의 경우는 포함되지 못함
- 출루율의 경우 좋은 선구안을 가지거나 발이 빠른 선수들이 높은 경향을 가짐
- 선구안이나 달리기의 경우 그 선수의 고유 능력치라고 판단하여 해당 선수의 통산 출루율을 사용하기로 함

# 예측 결과 - 출루율(OBP)

강백호 선수(KT 위즈)

0.409

김재환 선수(두산 베어스)

0.386

김현수 선수(LG 트윈스)

0.401

로맥 선수 (SSG 랜더스)

0.377

박건우 선수(두산 베어스)

0.388

양의지 선수(NC 다이노스)

0.391

이정후 선수(키움 히어로즈)

0.404

전준우 선수(롯데 자이언츠)

0.360

채은성 선수(LG 트윈스)

0.358

최정 선수(SSG 랜더스)

0.390



# 예측 결과 - OPS

선수 이름	출루율(OBP)	장타율(SLG)	OPS
강백호 선수	0.409	0.784	1.193
김재환 선수	0.386	0.716	1.172
김현수 선수	0.401	0.562	0.963
로맥 선수	0.377	0.697	1.074
박건우 선수	0.388	0.613	1.001

선수 이름	출루율(OBP)	장타율(SLG)	OPS
양익지 선수	0.391	0.621	1.012
이정후 선수	0.404	0.562	0.966
전준우 선수	0.360	0.523	0.883
채은성 선수	0.358	0.552	0.910
최정 선수	0.390	0.643	1.033

# 기대효과

01

## 유망주 발굴

- 리그에서 필요로하는 장타형 유망주를 새로운 데이터 관점에서 비교 가능
- 장타형 유망주 발굴 용이

02

## 선수 훈련 방식 변화

- 기존 데이터에서 확인할 수 없었던 세부 데이터로 선수 유형을 다시 구분하여 맞춤형 훈련 적용
- 선수 유형에 맞는 훈련 적용

03

## 연봉 협상

- 선수 측이나 구단 측 입장을 대변할 수 있는 새로운 데이터 제시
- 선수에게 제시할 수 있는 새로운 데이터

# 기대효과

## 세이버매트릭스가 불러 온 야구계의 새바람

세이버 매트릭스의 창시자 빌 제임스는 편견에 사로잡히지 않고 각종 데이터를 분석 한 끝에 여러가지 지표들을 만들어 냈  
그 중 대표적인 피타고리언 기대 승률(Pythagorean Winning Percentage)은 오늘날 시즌이 진행됨에 따라 야구팬들이  
직접 계산해볼 정도로 유명한 지표

## 야구 팬들의 눈높이 향상에 따라 새 볼거리 제공

위와 같은 지표들로 인해 국내 야구팬들이 데이터를 보는 눈높이는 향상되었음  
매번 같은 데이터로 경기를 분석하는 것은 선수에게도 팬에게도 제자리 걸음일 것  
데이터 전문가들의 노력으로 만들어낸 새로운 지표로 야구 팬들에게 새로운 볼거리를 제공할 수 있을 것



Thank You

**감사합니다.**