

Class 2. 클래식 스탯과 세이버 스탯(타자)

클래식 스탯과 세이버 스탯

스탯(stat): statistics 통계의 약자.

통계적인 수치로 나타낸 선수의 능력치.

스탯(stat) = 지표(indicator) = 지수(index)

1. 클래식 스탯 종류

타율, 홈런, 득점, 타점, 도루, 다승, 이닝, 평균자책점, 탈삼진, 볼넷, 출루율, 장타율 등

SABR(The Society for American Baseball Research)

- 1971년에 밥 데이비스 창시한 모임
- 야구를 통계학적/수학적으로 분석하는 방법론
- 빌 제임스: 세이버 메트릭스의 아버지

2. 세이버 스탯 종류

OPS(출루율+장타율), RC(출루율x장타율), BABIP(인플레이 타율), wOBA(가중 출루율), wRC+(조정 득점 생산력), FIP(수비 무관 평균자책점), DRS(디펜시스런세이브), WAR(승리 기여도), WHIP(이닝당 출루 허용), WPA(승리 확률 기여도) 등

참고1: 위키백과, 세이버메트릭스

참고2: <https://library.fangraphs.com/offense/offensive-statistics-list/>

타자 스탯

1. 타율(batting average, BA 또는 AVG)

타율 = 안타 / 타수
= 안타 / (타석 - 볼넷 - 사구 - 희생번트 - 희생플라이 - 타격방해 - 주루방해)

*타수: 타자가 타석에서 타격을 완료한 횟수

◎ 타자기록

기본기록		세부기록													
순위	선수명	팀명	AVG	G	PA	AB	R	H	2B	3B	HR	TB	RBI	SAC	SF
1	이정후	키움	0.360	123	544	464	78	167	42	6	7	242	84	0	9
2	전준우	롯데	0.348	144	619	552	88	192	46	0	7	259	92	1	8
3	강백호	KT	0.347	142	627	516	76	179	40	1	16	269	102	0	8
4	홍창기	LG	0.328	144	651	524	103	172	26	2	4	214	52	0	2
5	박건우	두산	0.325	126	525	458	82	149	31	2	6	202	63	0	6
6	양의지	NC	0.325	141	570	480	81	156	29	2	30	279	111	0	10
7	손아섭	롯데	0.319	139	610	542	88	173	29	2	3	215	58	3	1
8	페르난데스	두산	0.315	141	617	540	73	170	24	0	15	239	81	0	6
9	김선빈	KIA	0.307	130	564	501	55	154	32	0	5	201	67	1	5
10	안치훈	롯데	0.306	119	490	421	58	129	30	2	10	193	82	5	9

2021 KBO 타율 top 10 (출처: KBO 기록실)

2. 출루율(on base percentage, OBP)

출루율 = (안타 + 볼넷 + 사구) / (타수 + 볼넷 + 사구 + 희생플라이)
= (안타 + 볼넷 + 사구) / (타석 - 희생번트 - 타격방해)

◎ 타자기록

기본기록 세부기록														
순위	선수명	팀명	AVG	BB	IBB	HBP	SO	GDP	SLG	OBP	OPS	MH	RISP	PH-BA
1	홍창기	LG	0.328	109	4	16	95	6	0.408	0.456	0.864	57	0.343	0.500
2	강백호	KT	0.347	103	10	0	85	7	0.521	0.450	0.971	53	0.340	0.000
3	이정후	키움	0.360	62	1	9	37	8	0.522	0.438	0.960	51	0.341	0.000
4	양의지	NC	0.325	69	5	11	60	12	0.581	0.414	0.995	41	0.374	0.500
5	최경	SSG	0.278	84	3	22	102	9	0.562	0.410	0.972	28	0.281	0.600
6	추신수	SSG	0.265	103	4	12	123	8	0.451	0.409	0.860	34	0.274	0.222
7	정은원	한화	0.283	105	1	1	105	9	0.384	0.407	0.791	40	0.312	0.500
8	최재훈	한화	0.275	72	2	12	68	16	0.387	0.405	0.792	29	0.275	0.000
9	전준우	롯데	0.348	53	0	5	71	10	0.469	0.405	0.874	53	0.416	0.600
10	박건우	두산	0.325	50	2	11	72	16	0.441	0.400	0.841	44	0.325	0.250

2021 KBO 출루율 top 10 (출처: KBO 기록실)

이대호 선수 출루율 구하는 파이썬 코드

```
1  ## 이대호 선수 2022년 2022-05-11 현재 출루율 계산
2  # 출루율 = (안타 + 볼넷 + 사구)/(타수 + 볼넷 + 사구 + 희생)
3
4  H = 41
5  BB = 7
6  HBP = 1
7  AB = 117
8  SF = 2
9
10 OBP = round((H + BB + HBP) / (AB + BB + HBP + SF), 3)
11
12 print("이대호 선수 2022년 현재 출루율:", OBP)
13
14 # 항상 소숫점 아래 3자리까지 나오게 하려면?
15 print(f"이대호 선수 2022년 현재 출루율: {OBP:.3f}")
```

```
root@goorm:/workspace/kusf_study# python3 calc_obp.py
이대호 선수 2022년 현재 출루율: 0.386
이대호 선수 2022년 현재 출루율: 0.386
```

round 함수로 반올림

```
1  ## 허경민 선수 2022년 2022-05-11 현재 출루율 계산
2  # 출루율 = (안타 + 볼넷 + 사구)/(타수 + 볼넷 + 사구 + 희생)
3
4  H = 35
5  BB = 10
6  HBP = 3
7  AB = 105
8  SF = 2
9
10 OBP = round((H + BB + HBP) / (AB + BB + HBP + SF), 3)
11
12 print("허경민 선수 2022년 현재 출루율:", OBP)
13
14 # 항상 소숫점 아래 3자리까지 나오게 하려면?
15 print(f"허경민 선수 2022년 현재 출루율: {OBP:.3f}")
```

```
root@goorm:/workspace/kusf_study# python3 calc_obp1.py
허경민 선수 2022년 현재 출루율: 0.4
허경민 선수 2022년 현재 출루율: 0.400
```

문자열 포매팅 기법: 항상 소수점 아래 3자리까지 출력되게 해 줌

기대득점표(run expectancy matrix)

특정 주자, 특정 아웃 상태에서 몇 득점을 기대할 수 있는지를 나타내는 표
시즌마다 조금씩 다름(투고타저, 공인구 등의 이슈)

지난 NC vs 롯데 경기 (2022년 5월 11일)

4회초 NC가 1:2로 뒤지고 있는 상황
무사 1, 2루 찬스
타석에는 NC 이명기, 투수는 롯데 나균안
주자들을 2, 3루로 보내기 위한 번트 시도

번트안타 성공시 기대득점 변화
1.4399 -> 무사 주자만루 **2.2654** ↑

희생번트 성공시 기대득점 변화
1.4399 -> 1사 주자 2,3루 **1.2831** ↓

작전대로 되었을 때

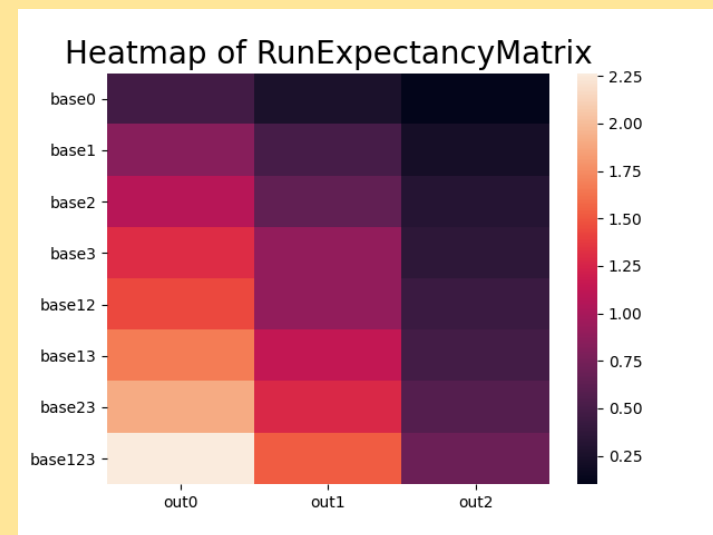
번트 실패 또는 삼진 등으로 주자를 못 보내고 아웃되었을 때 기대득점 변화
1.4399 -> 1사 주자 1,2루 **0.8919** ↓

희생번트를 성공해도 기대득점은 작아지는 상황

NC는 결과적으로 무사 1, 2루 상황에서 1점 뽑아냈음
무사 1,2루 상황의 기대득점에 근접한 결과

	0아웃	1아웃	2아웃
주자없음	0.4790	0.2572	0.0996
1루	0.8427	0.5004	0.2203
2루	1.0762	0.6487	0.3154
3루	1.3045	0.8916	0.3606
1,2루	1.4399	0.8919	0.4361
1,3루	1.6682	1.1348	0.4813
2,3루	1.9017	1.2831	0.5764
만루	2.2654	1.5263	0.6971

기대득점표 (2015 MLB 정규시즌 기준)



기대득점표(run expectancy matrix) 히트맵 그린 파이썬 코드

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

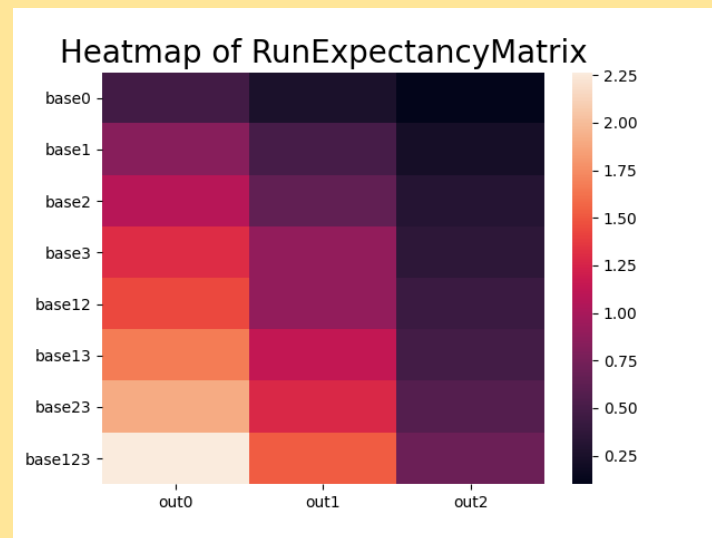
matrix = {'out0':[0.4790, 0.8427, 1.0762, 1.3045, 1.4399, 1.6682, 1.9017, 2.2654],
          'out1':[0.2572, 0.5004, 0.6487, 0.8916, 0.8919, 1.1348, 1.2831, 1.5263],
          'out2':[0.0996, 0.2203, 0.3154, 0.3606, 0.4361, 0.4813, 0.5764, 0.6971]}

df_run_expectancy_matrix = pd.DataFrame(matrix,
                                         index=['base0', 'base1', 'base2', 'base3', 'base12', 'base13', 'base23', 'base123'])

print(df_run_expectancy_matrix)

ax = sns.heatmap(df_run_expectancy_matrix)
plt.title('Heatmap of RunExpectancyMatrix', fontsize=20)
plt.savefig('heatmap of run expectancy matrix.png')
```

	0아웃	1아웃	2아웃
주자없음	0.4790	0.2572	0.0996
1루	0.8427	0.5004	0.2203
2루	1.0762	0.6487	0.3154
3루	1.3045	0.8916	0.3606
1,2루	1.4399	0.8919	0.4361
1,3루	1.6682	1.1348	0.4813
2,3루	1.9017	1.2831	0.5764
만루	2.2654	1.5263	0.6971



타자 스탯

3. 장타율(slugging percentage, SLG)

장타율 = (1루타 + 2*2루타 + 3*3루타 + 4*홈런)/타수

4. 순장타율(isolated power, IsoP)

순장타율 = 장타율 - 타율

= (1루타 + 2*2루타 + 3*3루타 + 4*홈런)/타수 - (1루타 + 2루타 + 3루타 + 홈런)/타수

= (2루타 + 2*3루타 + 3*홈런)/타수

*단타를 제거함으로 순수한 장타 능력을 알아보기 위한 스탯

장타율의 한계

장타없이 단타를 많이 쳐버려도 높게 나올 수 있음

10타수 8안타를 기록한 선수 A의 장타율 = $8/10 = 0.8 = 8\text{할}$

10타수 2홈런을 기록한 선수 B의 장타율 = $(4*2)/10 = 0.8 = 8\text{할}$

반면, 순장타율은?

선수A의 순장타율 = $0/10 = 0$

선수B의 순장타율 = $(3*2)/10 = 0.6 = 6\text{할}$

타자기록

기본기록

세부기록

◀

기록보기

▶

순위	선수명	팀명	AVG	BB	IBB	HBP	SO	GDP	SLG	OBP	OPS	MH	RISP	PH·BA
1	양익지	NC	0.325	69	5	11	60	12	0.581	0.414	0.995	41	0.374	0.500
2	최정	SSG	0.278	84	3	22	102	9	0.562	0.410	0.972	28	0.281	0.600
3	한유섭	SSG	0.278	60	4	10	96	11	0.534	0.373	0.907	33	0.304	0.429
4	이정후	키움	0.360	62	1	9	37	8	0.522	0.438	0.960	51	0.341	0.000
5	강백호	KT	0.347	103	10	0	85	7	0.521	0.450	0.971	53	0.340	0.000
6	구자욱	삼성	0.306	48	3	6	98	9	0.519	0.361	0.880	45	0.283	0.000
7	알테어	NC	0.272	57	2	11	156	6	0.514	0.358	0.872	31	0.248	0.000
8	오재일	삼성	0.285	58	8	0	106	5	0.512	0.366	0.878	36	0.276	0.000
9	나성범	NC	0.281	38	2	11	155	5	0.509	0.335	0.844	46	0.263	0.000
10	김재환	두산	0.274	81	3	5	127	9	0.501	0.382	0.883	40	0.338	0.500

2021 KBO 장타율 top 10 (출처: KBO 기록실)

타자기록

기본기록		세부기록												
순위	선수명	팀명	AVG	XBH	GO	AO	GO/AO	GW RBI	BB/K	P/PA	ISOP	XR	GPA	
1	최정	SSG	0.278	53	77	148	0.52	13	0.82	4.02	0.284	105.0	0.325	
2	양익지	NC	0.325	61	84	190	0.44	17	1.15	3.74	0.256	110.1	0.332	
3	한유섭	SSG	0.278	50	108	121	0.89	4	0.63	4.04	0.256	87.5	0.301	
4	알테어	NC	0.272	53	92	115	0.80	10	0.37	3.97	0.242	93.4	0.290	
5	나성범	NC	0.281	63	119	140	0.85	9	0.25	3.74	0.228	95.5	0.278	
6	김재환	두산	0.274	52	109	114	0.96	15	0.64	4.01	0.227	91.3	0.297	
7	오재일	삼성	0.285	45	99	102	0.97	9	0.55	3.82	0.227	79.9	0.293	
8	양석환	두산	0.273	50	79	147	0.54	12	0.31	3.98	0.217	80.1	0.274	
9	구자욱	삼성	0.306	62	122	169	0.72	9	0.49	3.87	0.214	103.8	0.292	
10	박동원	키움	0.249	43	105	94	1.12	11	0.46	4.06	0.211	65.6	0.269	

2021 KBO 순장타율 top 10 (출처: KBO 기록실)

이정후
장타율 4위 -> 순장타율 21위 ↓

강백호
장타율 5위 -> 순장타율 18위 ↓

박동원
장타율 16위 -> 순장타율 10위 ↑

박병호
장타율 26위 -> 순장타율 12위 ↑

장타율 순위에서는 상위 랭커였던 이정후, 강백호가 순장타율 순위에서는 10위에 들지 못했음

타율이 높았기 때문에 장타율 순위에서 높은 위치에 있었던 것
이정후 타율(0.360 1위), 강백호 타율(0.347 3위)

타자 스탯

5. OPS(on-base plus slugging)

OPS = 장타율 + 출루율

출루율의 값의 범위는 0~1이고 장타율의 값의 범위는 0~4이기 때문에 OPS는 값의 범위가 다른 것을 더한 꼴.
따라서 장타율이 높은 타자가 OPS에 있어서 무조건 유리

6. GPA(gross production average 또는 Gleeman production average)

GPA = (1.8*출루율 + 장타율) / 4

Aaron Gleeman이라는 사람이 만든 스탯

출루율이 장타율보다 80%만큼 더 중요하다는 연구 결과를 기초로 출루율에 1.8을 곱해줌
타율과 비슷한 범위가 되게 하기 위해서 4로 나눠줌

타자기록

기본기록

세부기록

◀

기록보기

▶

순위	선수명	팀명	AVG	BB	IBB	HBP	SO	GDP	SLG	OBP	OPS	MH	RISP	PH-BA
1	양의지	NC	0.325	69	5	11	60	12	0.581	0.414	0.995	41	0.374	0.500
2	최정	SSG	0.278	84	3	22	102	9	0.562	0.410	0.972	28	0.281	0.600
3	강백호	KT	0.347	103	10	0	85	7	0.521	0.450	0.971	53	0.340	0.000
4	이정후	키움	0.360	62	1	9	37	8	0.522	0.438	0.960	51	0.341	0.000
5	한유섬	SSG	0.278	60	4	10	96	11	0.534	0.373	0.907	33	0.304	0.429
6	김재환	두산	0.274	81	3	5	127	9	0.501	0.382	0.883	40	0.338	0.500
7	구자욱	삼성	0.306	48	3	6	98	9	0.519	0.361	0.880	45	0.283	0.000
8	오재일	삼성	0.285	58	8	0	106	5	0.512	0.366	0.878	36	0.276	0.000
9	전준우	롯데	0.348	53	0	5	71	10	0.469	0.405	0.874	53	0.416	0.600
10	알테어	NC	0.272	57	2	11	156	6	0.514	0.358	0.872	31	0.248	0.000

2021 KBO OPS top 10 (출처: KBO 기록실)

타자기록

기본기록 세부기록													
순위	선수명	팀명	AVG	XBH	GO	AO	GO/AO	GW RBI	BB/K	P/PA	ISOP	XR	GPA
1	강백호	KT	0.347	57	131	129	1.02	11	1.21	4.20	0.174	117.6	0.333
2	양의지	NC	0.325	61	84	190	0.44	17	1.15	3.74	0.256	110.1	0.332
3	이정후	키움	0.360	55	131	138	0.95	7	1.68	3.94	0.162	100.8	0.328
4	최정	SSG	0.278	53	77	148	0.52	13	0.82	4.02	0.284	105.0	0.325
5	홍창기	LG	0.328	32	138	121	1.14	8	1.15	4.17	0.080	106.4	0.307
6	한유섬	SSG	0.278	50	108	121	0.89	4	0.63	4.04	0.256	87.5	0.301
7	전준우	롯데	0.348	53	145	152	0.95	10	0.75	3.64	0.121	98.9	0.300
8	김재환	두산	0.274	52	109	114	0.96	15	0.64	4.01	0.227	91.3	0.297
8	추신수	SSG	0.265	42	118	102	1.16	4	0.84	4.11	0.187	93.4	0.297
10	오재일	삼성	0.285	45	99	102	0.97	9	0.55	3.82	0.227	79.9	0.293

2021 KBO GPA top 10 (출처: KBO 기록실)

출루율이 좋았던 홍창기(출루율 0.456 1위)와 추신수(출루율 0.409 6위)가 GPA에서는 높은 순위에 있음

타자 스탯

7. RC(run created, 득점 생산 또는 득점 창출력)

야구는 기본적으로 득점을 많이 해야 승리하는 종목
한 선수가 득점 생산에 얼마나 큰 기여를 했는지를 나타내고자 한 스탯
빌 제임스가 개발

기본 컨셉

$$RC = (A \times B) / C$$

여기서 A는 출루 능력, B는 진루 능력, C는 주어진 기회

베이직 RC

$$1) RC = ((\text{안타} + \text{볼넷}) \times \text{총루타}) / (\text{타수} + \text{볼넷})$$

*총루타: 만약 단타 100개, 2루타 25개, 3루타 5개, 홈런 20개를 쳤다면

$$\text{총루타} = 1 \times 100 + 2 \times 25 + 3 \times 5 + 4 \times 20 = 100 + 50 + 15 + 80 = 245 \text{루타}$$

$$2) RC = \text{출루율} \times \text{장타율} \times \text{타수}$$

$$3) RC = \text{출루율} \times \text{총루타}$$

도루 능력 반영 RC

$$1) RC = ((\text{안타} + \text{볼넷} - \text{도루실패}) \times (\text{총루타} + 0.55 \times \text{도루성공})) / (\text{AB} + \text{BB})$$

Technical 버전 RC

$$1) RC = ((\text{안타} + \text{볼넷} + \text{사구} - \text{도루실패} - \text{땅볼병살타}) \times (\text{총루타} + 0.26 \times (\text{볼넷} - \text{고의4구} + \text{사구})) + 0.52 \times (\text{도루} + \text{희생번트} + \text{희생플라이})) / (\text{타수} + \text{볼넷} + \text{사구} + \text{희생번트} + \text{희생플라이})$$

타자 스탯

7. RC(run created, 득점 생산 또는 득점 창출력)

2002 버전 RC

A: 안타 + 볼넷 + 사구 - 도루실패 - 땅볼병살타

B: $1.125 \times 1\text{루타} + 1.69 \times 2\text{루타} + 3.02 \times 3\text{루타} + 3.73 \times \text{홈런} + 0.29 \times (\text{볼넷} - \text{고의4구} + \text{사구}) + 0.492 \times (\text{희생번트} + \text{희생플라이} + \text{도루}) - 0.04 \times \text{삼진}$

C: 타수 + 볼넷 + 사구 + 희생번트 + 희생플라이

$$RC = (2.4C + A)(3C + B) / 9C - 0.9C$$

8. RC/27

1번부터 9번까지 같은 타자로 팀을 채웠을 경우

27개의 아웃카운트를 잡아 경기가 끝날 때까지 그 팀이 총 몇 점을 득점할 수 있는지를 나타낸 지표

$$RC/27 = RC / (\text{타수} - \text{안타} + \text{도루실패} + \text{희생번트} + \text{희생플라이} + \text{땅볼병살타}) \times 27$$

*RC/27도 RC와 마찬가지로 다양한 공식이 존재

홈런왕 최정 9명 vs 타격왕 이정후 9명 vs 도루왕 김혜성 9명 (2021년 기준)

최정

$$A = \text{안타} + \text{볼넷} + \text{사구} - \text{도루실패} - \text{땅볼병살타} \\ = 121 + 84 + 22 - 6 - 9 = 212$$

$$B = 1.125 \times 1\text{루타} + 1.69 \times 2\text{루타} + 3.02 \times 3\text{루타} + 3.73 \times \text{홈런} + 0.29 \times (\text{볼넷} - \text{고의4구} + \text{사구}) + 0.492 \times (\text{희생번트} + \text{희생플라이} + \text{도루}) - 0.04 \times \text{삼진} \\ = 1.125 \times 68 + 1.69 \times 17 + 3.02 \times 1 + 3.73 \times 35 + 0.29 \times (84 - 3 + 22) + 0.492 \times (1 + 12 + 8) - 0.04 \times 102 = 274.922$$

$$C = \text{타수} + \text{볼넷} + \text{사구} + \text{희생번트} + \text{희생플라이} \\ = 436 + 84 + 22 + 1 + 12 = 555$$

$$RC = (2.4C + A)(3C + B) / 9C - 0.9C \\ = (2.4 \times 555 + 212)(3 \times 555 + 274.922) / (9 \times 555) - 0.9 \times 555 = 100.148$$

$$RC/27 = 100.148 / (436 - 121 + 6 + 1 + 12 + 9) \times 27 = \mathbf{7.883}$$

이정후

$$A = 167 + 62 + 9 - 3 - 8 = 227$$

$$B = 1.125 \times 112 + 1.69 \times 42 + 3.02 \times 6 + 3.73 \times 7 + 0.29 \times (62 - 1 + 9) + 0.492 \times (0 + 9 + 10) - 0.04 \times 37 = 269.378$$

$$C = 464 + 62 + 9 + 0 + 9 = 544$$

$$RC = (2.4 \times 544 + 227)(3 \times 544 + 269.378) / (9 \times 544) - 0.9 \times 544 = 105.590$$

$$RC/27 = 105.590 / (464 - 167 + 3 + 0 + 9 + 8) \times 27 = \mathbf{8.993}$$

김혜성

$$A = 170 + 65 + 1 - 4 - 8 = 224$$

$$B = 1.125 \times 144 + 1.69 \times 20 + 3.02 \times 3 + 3.73 \times 3 + 0.29 \times (65 - 3 + 1) + 0.492 \times (0 + 10 + 46) - 0.04 \times 97 = 257.992$$

$$C = 559 + 65 + 1 + 0 + 10 = 635$$

$$RC = (2.4 \times 635 + 224)(3 \times 635 + 257.992) / (9 \times 635) - 0.9 \times 635 = 90.077$$

$$RC/27 = 90.077 / (559 - 170 + 4 + 0 + 10 + 8) \times 27 = \mathbf{5.917}$$

2022 버전 RC 사용했음

김혜성 9명 < 최정 9명 < 이정후 9명

5.917점 < 7.883점 < 8.993점

타율 0.304 0.278 0.360

이정후와 최정의 타율은 거의 1할이나 차이가 나지만 득점 생산에 있어서는 약 1점 차이
타율 맹신은 금물

Quiz. 이정후 9명보다 더 센 팀은?

김혜성 9명 < 최정 9명 < 이정후 9명 < ??? 9명

5.917점 < 7.883점 < 8.993점

타자 스탯

9. XR(extrapolated runs, 추정득점)

*짐 퍼테이도가 개발
*RC와 목적성은 거의 같음. 실제로 득점에 얼마나 기여했는지 추정.

$$XR = 0.5*1루타 + 0.72*2루타 + 1.04*3루타 + 1.44*홈런 + 0.38*(사구+볼넷-고의4구) + 0.25*고의사구 + 0.18*도루 - 0.32*도루실패 - 0.09*(\underline{타수-안타-삼진}) - 0.098*삼진 - 0.37*병살타 + 0.37*희생플라이 + 0.04*희생번트$$

‘범타’를 의미

타자기록

기본기록 세부기록													
순위	선수명	팀명	AVG	XBH	GO	AO	GO/AO	GW RBI	BB/K	P/PA	ISOP	XR	GPA
1	강백호	KT	0.347	57	131	129	1.02	11	1.21	4.20	0.174	117.6	0.333
2	양의지	NC	0.325	61	84	190	0.44	17	1.15	3.74	0.256	110.1	0.332
3	홍창기	LG	0.328	32	138	121	1.14	8	1.15	4.17	0.080	106.4	0.307
4	최정	SSG	0.278	53	77	148	0.52	13	0.82	4.02	0.284	105.0	0.325
5	구자욱	삼성	0.306	62	122	169	0.72	9	0.49	3.87	0.214	103.8	0.292
6	이정후	키움	0.360	55	131	138	0.95	7	1.68	3.94	0.162	100.8	0.328
7	전준우	롯데	0.348	53	145	152	0.95	10	0.75	3.64	0.121	98.9	0.300
8	피렐라	삼성	0.286	56	143	166	0.86	10	0.64	3.76	0.210	96.3	0.285
9	나성범	NC	0.281	63	119	140	0.85	9	0.25	3.74	0.228	95.5	0.278
10	알테어	NC	0.272	53	92	115	0.80	10	0.37	3.97	0.242	93.4	0.290

2021 KBO XR top 10 (출처: KBO 기록실)

팀 선수들의 XR을 모두 더하면 실제 팀 득점과 거의 같다?

TRY

각자 좋아하는 팀의 2021시즌 XR 총합을 구해보자!

그리고 실제 팀 득점과 비교해보자!

2021

KBO 정규시즌

두산

포지션 선택

경기상황별1

경기상황별2

◎ 타자기록

기본기록 세부기록													
순위	선수명	팀명	AVG	XBH	GO	AO	GO/AO	GW RBI	BB/K	P/PA	ISOP	XR	GPA
1	김재환	두산	0.274	52	109	114	0.96	15	0.64	4.01	0.227	91.3	0.297
2	페르난데스	두산	0.315	39	170	171	0.99	4	1.86	3.48	0.128	87.0	0.287
3	양석환	두산	0.273	50	79	147	0.54	12	0.31	3.98	0.217	80.1	0.274
4	박건우	두산	0.325	39	111	132	0.84	3	0.69	3.70	0.116	78.9	0.290
5	김인태	두산	0.259	24	92	100	0.92	3	0.93	3.88	0.119	51.8	0.262
6	허경민	두산	0.278	30	175	131	1.34	6	1.03	3.76	0.088	51.8	0.243
7	박계범	두산	0.267	18	87	84	1.04	7	0.65	4.12	0.090	45.3	0.255
8	정수빈	두산	0.259	26	112	72	1.56	2	0.58	3.82	0.115	37.4	0.240
9	강승호	두산	0.239	25	76	78	0.97	2	0.28	3.79	0.136	34.6	0.229
10	안재석	두산	0.255	12	49	51	0.96	2	0.30	3.94	0.090	23.4	0.229
11	장승현	두산	0.232	12	51	36	1.42	3	0.38	3.81	0.090	21.0	0.230
12	박세혁	두산	0.219	12	77	68	1.13	1	0.51	4.02	0.051	19.9	0.201
13	김재호	두산	0.209	9	56	80	0.70	3	0.85	3.91	0.052	18.6	0.203
14	조수형	두산	0.286	4	21	13	1.62	0	0.62	4.08	0.083	16.9	0.280
15	최용제	두산	0.279	3	27	28	0.96	1	0.52	3.89	0.029	13.1	0.237
16	안권수	두산	0.238	1	20	9	2.22	0	1.00	3.47	0.048	4.3	0.202
17	오재원	두산	0.167	3	19	24	0.79	2	0.33	3.90	0.042	2.8	0.155
18	박지훈	두산	0.333	1	3	2	1.50	0	0.33	3.57	0.083	2.4	0.297
19	신성현	두산	0.182	0	2	1	2.00	0	0.50	4.60	0.000	1.5	0.226
20	권민석	두산	0.214	2	3	1	3.00	0	0.14	4.20	0.143	1.2	0.209
21	전민재	두산	0.250	0	0	2	0.00	0	0.00	3.20	0.000	0.6	0.153
22	김민혁	두산	0.000	0	3	1	3.00	0	1.00	4.25	0.000	0.1	0.113
23	권휘	두산	-	0	0	0	-	0	-	-	-	0.0	0.000
23	김강률	두산	-	0	0	0	-	0	-	-	-	0.0	0.000
23	김명신	두산	-	0	0	0	-	0	-	-	-	0.0	0.000
23	김민규	두산	-	0	0	0	-	0	-	-	-	0.0	0.000
23	남호	두산	-	0	0	0	-	0	-	-	-	0.0	0.000
23	문대원	두산	-	0	0	0	-	0	-	-	-	0.0	0.000
23	박종기	두산	-	0	0	0	-	0	-	-	-	0.0	0.000
23	박치국	두산	-	0	0	0	-	0	-	-	-	0.0	0.000

팀 선수들의 XR을 모두 더하면 실제 팀 득점과 거의 같다?

$$\%error = |approx - exact|/exact * 100$$

🕒 팀기록(타자)

팀 득점

◀ 기록보기 ▶

순위	팀명	AVG	G	PA	AB	R	H	2B	3B	HR	TB	RBI	SAC	SF
1	SSG	0.261	144	5736	4899	755	1278	204	13	185	2063	712	55	52
2	두산	0.268	144	5645	4900	738	1314	235	17	110	1913	698	46	53
3	롯데	0.278	144	5726	5009	727	1393	266	10	107	2000	697	38	46
4	키움	0.259	144	5653	4874	722	1262	244	28	91	1835	671	35	56
5	KT	0.265	144	5627	4810	719	1276	219	11	106	1835	673	69	44
6	삼성	0.267	144	5552	4836	712	1292	204	18	133	1931	669	48	68
7	NC	0.261	144	5532	4808	702	1254	203	17	170	2001	670	59	43
8	LG	0.250	144	5533	4750	654	1188	210	11	110	1750	621	52	45
9	한화	0.237	144	5495	4685	599	1111	213	17	80	1598	554	54	49
10	KIA	0.248	144	5577	4831	568	1198	183	21	66	1621	546	41	37
합계		0.260	720	56076	48402	6896	12566	2181	163	1158	18547	6511	497	493

팀 XR 합계

백분율 오차

765.2

1.35%

683.0

7.45%

695.5

4.33%

642.0

11.1%

710.3

1.21%

699.2

1.80%

707.4

0.77%

678.9

3.81%

574.1

4.16%

579.7

2.06%

3.80%

팀 XR 합계는 팀 득점과 매우 유사하다

2021년 KBO 팀 XR 총합 구하는 파이썬 웹 크롤링(web crawling) 코드 - 1

웹 크롤링(web crawling): 정해진 규칙에 따라 복수 웹 페이지를 브라우징하는 행위

웹 스크래핑(web scraping): 웹 사이트에서 원하는 부분에 위치한 데이터를 추출하여 수집하는 기술

<https://minmong.tistory.com/182>

<https://www.snugarchive.com/blog/python-web-scraping/>

```
from selenium import webdriver
from selenium.webdriver.support.ui import Select
from selenium.webdriver.common.keys import Keys
from selenium.common.exceptions import UnexpectedAlertPresentException

import time

team_name = input("\n\n팀 XR 총합을 구해볼 팀 이름을 입력하세요(두산, 롯데, 삼성, 키움, 한화, KIA, KT, LG, NC, SSG):")

chromedriver = './chromedriver.exe'
driver = webdriver.Chrome(chromedriver)

url = 'https://www.koreabaseball.com/Record/Player/HitterBasic/Basic1.aspx'

driver.get(url)
driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_ddlSeason_ddlSeason [value='2021']").click()

time.sleep(1)

team_select = Select(driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_ddlTeam_ddlTeam"))
team_select.select_by_visible_text(team_name)

time.sleep(1)

driver.find_element_by_xpath('//*[@id="cphContents_cphContents_cphContents_udpContent"]/div[2]/div[1]/ul/li[2]/a').click()

time.sleep(2)

table_tbody = driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_udpContent > div.record_result > table > tbody")
```

(뒤 페이지에 이어서)

```

xr_list = []
for tr in table_tbody.find_elements_by_tag_name("tr"):
    xr = tr.find_element_by_css_selector("td:nth-child(13)")
    xr_list.append(float(xr.text))

print(xr_list)

driver.find_element_by_css_selector('#cphContents_cphContents_cphContents_ucPager_btnNo2').click()

time.sleep(1)

table_tbody = driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_udpContent > div.record_result > table > tbody")

xr_list1 = []
for tr in table_tbody.find_elements_by_tag_name("tr"):
    xr = tr.find_element_by_css_selector("td:nth-child(13)")
    xr_list1.append(float(xr.text))

print(xr_list1)

xr = xr_list + xr_list1

xr_sum = sum(xr)

print("\n\n-----")
print(team_name, "팀 XR 합계:", xr_sum)
print("-----\n\n")
driver.quit()

```

팀 XR 총합을 구해볼 팀 이름을 입력하세요(두산, 롯데, 삼성, 키움, 한화, KIA, KT, LG, MC, SSG):SSG

SSG 팀 XR 합계: 765.2

2021년 KBO 팀 XR 총합 구하는 파이썬 웹 크롤링(web crawling) 코드 - 2

이번에는 팀 이름 입력 받지 않고 알아서 10개 팀 XR 총합 한번에!

```
from selenium import webdriver
from selenium.webdriver.support.ui import Select
from selenium.webdriver.common.keys import Keys
from selenium.common.exceptions import UnexpectedAlertPresentException

import time

team_name = ['두산', '롯데', '삼성', '키움', '한화', 'KIA', 'KT', 'LG', 'NC', 'SSG']

chromedriver = './chromedriver.exe'
driver = webdriver.Chrome(chromedriver)

team_xr_sum = dict()

for name in team_name:

    url = 'https://www.koreabaseball.com/Record/Player/HitterBasic/Basic1.aspx'

    driver.get(url)
    driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_ddlSeason_ddlSeason [value='2021']").click()

    time.sleep(1)

    team_select = Select(driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_ddlTeam_ddlTeam"))
    team_select.select_by_visible_text(name)

    time.sleep(1)

    driver.find_element_by_xpath('//*[@id="cphContents_cphContents_cphContents_udpContent"]/div[2]/div[1]/ul/li[2]/a').click()

    time.sleep(2)

    table_tbody = driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_udpContent > div.record_result > table > tbody")
```

(뒤 페이지에 이어서)

```

xr_list = []
for tr in table_tbody.find_elements_by_tag_name("tr"):
    xr = tr.find_element_by_css_selector("td:nth-child(13)")
    xr_list.append(float(xr.text))

print(xr_list)

driver.find_element_by_css_selector('#cphContents_cphContents_cphContents_ucPager_btnNo2').click()

time.sleep(1)

table_tbody = driver.find_element_by_css_selector("#cphContents_cphContents_cphContents_udpContent > div.record_result > table > tbody")

xr_list1 = []
for tr in table_tbody.find_elements_by_tag_name("tr"):
    xr = tr.find_element_by_css_selector("td:nth-child(13)")
    xr_list1.append(float(xr.text))

print(xr_list1)

xr = xr_list + xr_list1

xr_sum = sum(xr)

print("\n\n-----")
print(team_name, "팀 XR 합계:", xr_sum)
print("-----\n\n")

team_xr_sum[name] = xr_sum

print("\n\n-----")
print("KBO 팀 XR 합계")
print(team_xr_sum)
print("-----\n\n")

driver.quit()

```

```

-----
KBO 팀 XR 합계
{'두산': 683.0, '롯데': 695.5, '삼성': 699.1999999999999, '키움': 642.0000000000001, '한화': 574.1000000000001, 'KIA': 579.7, 'KT': 710.3, 'LG': 678.9, 'NC': 707.3999999999999, 'SSG': 765.2}
-----

```

타자 스탯

10. 가중출루율(weighted on-base average, wOBA)

팬그래프 2013버전:

$$\text{wOBA} = \frac{(0.69 \times \text{고의4구제외한볼넷} + 0.72 \times \text{사구} + 0.89 \times 1\text{루타} + 1.27 \times 2\text{루타} + 1.62 \times 3\text{루타} + 2.10 \times \text{홈런})}{(\text{타수} + \text{고의4구제외한볼넷} + \text{희생플라이} + \text{사구})}$$

스탯티즈 2022 버전:

$$\text{wOBA} = \frac{(0.72 \times \text{고의4구제외한볼넷} + 0.75 \times \text{사구} + 0.9 \times 1\text{루타} + 0.92 \times \text{실책출루} + 1.24 \times 2\text{루타} + 1.56 \times 3\text{루타} + 1.95 \times \text{홈런})}{(\text{타석} - \text{고의4구})}$$

*툼 탕고 개발

*가중치를 나타내는 계수는 시즌마다 득점 환경에 따라 바뀜

11. 가중득점창출력(weighted runs created, wRC)

구장별 조정(park factor)을 거친 형태의 wOBA

*파크팩터란?

투수의 무덤 콜로라도를 홈구장으로 쓰는 타자가 다른 구장을 홈으로 쓰는 타자보다 유리 구장에 의한 영향을 최소화하기 위해 기록에 조정이 필요함

2021 KBO wOBA top 10 (출처: 스탯티즈)

기본	확장	가치	클러치	타석	타구1	타구2	파워	팀배팅1	팀배팅2	도루	주루	구종가치	구종구사									
순	이름	팀	생산력	타석	HR%	BB%	K%	BB/K	IsoP	IsoD	BABIP	Spd	PSN	타격 생산력				타격 생산력 (파크팩터 조정)				
			wOBA											wOBA	wRC	wRC _{/27}	wRAA	wOBA	wRC	wRC _{/27}	wRAA	wRC+
1	양의지	21 N DH	.443	570	5.26	12.1	10.5	1.15	.256	.089	.315	3.0	3.75	.443	117.6	9.15	47.5	.439	115.9	9.02	45.8	165.4
2	강백호	21 K 1B	.443	627	2.55	16.4	13.6	1.21	.174	.103	.385	3.7	12.31	.443	129.3	9.78	52.2	.439	127.6	9.65	50.5	165.5
3	이정후	21 키 CF	.441	543	1.29	11.4	6.8	1.68	.162	.078	.374	5.7	8.24	.441	111.4	9.52	44.6	.440	110.7	9.46	43.9	165.8
4	최정	21 S 3B	.431	555	6.31	15.1	18.4	0.82	.284	.132	.277	3.5	13.02	.431	109.3	8.60	41.0	.425	106.4	8.38	38.2	155.9
5	홍창기	21 L CF	.417	651	0.61	16.9	14.6	1.16	.080	.128	.393	5.2	6.82	.417	120.6	8.85	40.6	.420	122.2	8.96	42.1	152.6
6	전준우	21 롯데 LF	.406	619	1.13	8.6	11.5	0.75	.121	.057	.384	3.6	6.46	.406	109.2	7.72	33.1	.402	106.8	7.55	30.7	140.4
7	한유섬	21 S RF	.404	519	5.97	11.6	18.5	0.63	.256	.094	.287	2.5	1.94	.404	90.8	7.27	27.0	.399	88.4	7.09	24.6	138.6
8	김재환	21 두 LF	.401	566	4.77	14.3	22.4	0.64	.227	.108	.316	3.3	3.72	.401	97.2	7.25	27.6	.413	103.1	7.69	33.5	148.2
9	추신수	21 S DH	.400	580	3.62	17.8	21.2	0.84	.187	.144	.315	5.0	22.83	.400	99.6	7.47	28.2	.397	98.0	7.35	26.7	137.4
10	박건우	21 두 RF	.394	525	1.14	9.5	13.7	0.69	.116	.075	.371	4.9	8.21	.394	87.3	7.12	22.8	.399	89.4	7.30	24.9	138.5

2021 KBO wRC+ top 10 (출처: 스탯티즈)

기본	확장	가치	클러치	타석	타구1	타구2	파워	팀배팅1	팀배팅2	도루	주루	구종가치	구종구사									
순	이름	팀	생산력+	타석	HR%	BB%	K%	BB/K	IsoP	IsoD	BABIP	Spd	PSN	타격 생산력				타격 생산력 (파크팩터 조정)				
			wRC+											wOBA	wRC	wRC ₂₇	wRAA	wOBA	wRC	wRC ₂₇	wRAA	wRC+
1	이정후	21 키 CF	165.8	543	1.29	11.4	6.8	1.68	.162	.078	.374	5.7	8.24	.441	111.4	9.52	44.6	.440	110.7	9.46	43.9	165.8
2	강백호	21 K 1B	165.5	627	2.55	16.4	13.6	1.21	.174	.103	.385	3.7	12.31	.443	129.3	9.78	52.2	.439	127.6	9.65	50.5	165.5
3	양의지	21 N DH	165.4	570	5.26	12.1	10.5	1.15	.256	.089	.315	3.0	3.75	.443	117.6	9.15	47.5	.439	115.9	9.02	45.8	165.4
4	최정	21 S 3B	155.9	555	6.31	15.1	18.4	0.82	.284	.132	.277	3.5	13.02	.431	109.3	8.60	41.0	.425	106.4	8.38	38.2	155.9
5	홍창기	21 L CF	152.6	651	0.61	16.9	14.6	1.16	.080	.128	.393	5.2	6.82	.417	120.6	8.85	40.6	.420	122.2	8.96	42.1	152.6
6	김재환	21 두 LF	148.2	566	4.77	14.3	22.4	0.64	.227	.108	.316	3.3	3.72	.401	97.2	7.25	27.6	.413	103.1	7.69	33.5	148.2
7	전준우	21 롯데 LF	140.4	619	1.13	8.6	11.5	0.75	.121	.057	.384	3.6	6.46	.406	109.2	7.72	33.1	.402	106.8	7.55	30.7	140.4
8	한유섬	21 S RF	138.6	519	5.97	11.6	18.5	0.63	.256	.094	.287	2.5	1.94	.404	90.8	7.27	27.0	.399	88.4	7.09	24.6	138.6
9	박건우	21 두 RF	138.5	525	1.14	9.5	13.7	0.69	.116	.075	.371	4.9	8.21	.394	87.3	7.12	22.8	.399	89.4	7.30	24.9	138.5
10	추신수	21 S DH	137.4	580	3.62	17.8	21.2	0.84	.187	.144	.315	5.0	22.83	.400	99.6	7.47	28.2	.397	98.0	7.35	26.7	137.4

투수 친화적 구장에서 뛰고 있는
김재환, 박건우의 경우
파크 팩터를 고려한 wRC+ 순위가
wOBA 순위보다 높다