

Dscover 스포츠 분야 프로젝트

2022 프로야구 최종 순위 예측 시스템

김지환 손유승 최범규 임세민 임선주

INDEX

1

프로젝트 주제 및
선정 배경

2

데이터 수집

3

데이터 전처리

4

데이터 시각화

5

의의 및 한계

1 프로젝트 주제 및 선정 배경

2021 프로야구 한국시리즈 결과

2021 한국시리즈

7전 4선승제

- 4 1차전 2
- 6 2차전 1
- 3 3차전 1
- 8 4차전 4



KT위즈



두산베어스

지난 10년간 우승팀

2021	KT위즈
2020	NC다이노스
2019	두산베어스
2018	SK와이브스
2017	KIA타이거즈
2016	두산베어스
2015	두산베어스
2014	삼성라이온즈
2013	삼성라이온즈
2012	삼성라이온즈
2011	삼성라이온즈

자료/ 한국야구위원회(KBO)



2022 시즌 최종 순위 예측

내가 응원하는 팀이 플레이오프에 진출할 수 있을까?

과연 이번에는 어떤 팀이 우승할까?

우리 팀의 최종 순위는 어떻게 될까?

② 데이터 수집

1) 선수단 변화 목록

Team Batting		Bold indicates they played in the majors or are a HoFer																			Share & Export ▼		Glossary			
Rk	Name	Age	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	TB	GDP	HBP	SH	SF	IBB	Notes
1	Yong Hwan Baik	31	52	107	91	12	20	4	0	5	10	0	0	15	29	.220	.330	.429	.759	39	2	0	1	0	0	
2	Hyoung Woo Choi*	36	140	600	522	93	185	37	1	28	115	0	0	70	101	.354	.433	.590	1.023	308	9	5	0	3	5	
3	Jeong Yong Choi*	23	38	59	56	5	11	1	0	1	5	0	0	2	15	.196	.224	.268	.492	15	0	0	1	0	0	
4	Jung Min Choi*	31	24	11	9	12	1	0	0	0	2	1	0	2	1	.111	.273	.111	.384	1	0	0	0	0	0	
5	Won Jun Choi*	23	123	412	359	72	117	16	6	2	35	14	3	33	35	.326	.387	.421	.808	151	1	6	9	5	0	
6	Seung Taek Han	26	83	262	226	22	51	6	0	9	29	0	0	21	58	.226	.308	.372	.680	84	7	6	9	0	0	
7	Jong Pyo Hong*	20	40	88	80	7	20	3	0	0	7	0	2	4	13	.250	.282	.288	.570	23	1	0	3	1	0	
8	Dae In Hwang	24	63	138	116	14	32	4	0	4	16	0	0	21	29	.276	.384	.414	.798	48	5	0	0	1	0	
9	Yun Ho Hwang	26	25	73	70	5	17	3	0	0	5	2	0	3	10	.243	.274	.286	.560	20	3	0	0	0	0	
10	Young Suk Jang	30	11	37	31	1	4	0	0	0	2	0	0	4	7	.129	.243	.129	.372	4	3	1	0	1	0	
11	Ho Ryung Kim	28	89	159	138	36	32	5	2	4	16	3	3	14	55	.232	.314	.384	.698	53	2	3	3	1	0	
12	Joo Chan Kim	39	7	18	18	3	6	0	0	1	2	0	1	0	2	.333	.333	.500	.833	9	2	0	0	0	0	
13	Kyu Sung Kim*	23	103	182	157	21	28	2	1	3	11	2	1	13	40	.178	.246	.261	.507	41	1	1	11	0	0	
14	Min Sik Kim*	31	69	238	211	22	53	6	1	2	21	0	0	17	28	.251	.310	.318	.628	67	3	2	6	2	0	
15	Sun Bin Kim	30	85	351	303	42	100	19	0	1	37	1	1	40	43	.330	.406	.403	.808	122	12	2	1	5	0	
16	Jang Hyeok Ko#	30	7	13	13	0	3	0	0	0	0	0	0	0	5	.231	.231	.231	.462	3	0	0	0	0	0	
17	Chang Jin Lee	29	22	99	88	19	29	8	1	0	7	0	1	10	17	.330	.404	.443	.847	39	2	1	0	0	0	
18	Jeong Hun Lee*	25	3	4	4	0	0	0	0	0	0	0	0	0	2	.000	.000	.000	.000	0	0	0	0	0	0	
19	Jin Young Lee	22	32	27	22	4	5	1	0	0	5	1	0	5	9	.227	.370	.273	.643	6	1	0	0	0	0	
20	Woo Sung Lee	25	48	67	59	5	12	1	0	0	3	1	0	6	11	.203	.299	.220	.519	13	2	2	0	0	0	
21	Seon Jae Moon	30	15	6	6	1	1	0	0	0	0	0	0	0	2	.167	.167	.167	.333	1	0	0	0	0	0	
22	Ji Wan Na	35	137	556	468	73	136	19	1	17	92	0	0	64	114	.291	.392	.444	.837	208	6	18	0	6	1	
23	Ju Hwan Na	36	64	222	204	21	57	5	0	6	26	0	0	12	44	.279	.321	.392	.713	80	6	1	4	1	0	
24	Sun Woo Oh*	23	59	73	67	11	16	0	0	4	14	0	0	3	26	.239	.301	.418	.719	28	2	3	0	0	0	
25	Chan-Ho Park	25	141	531	480	63	107	16	0	3	36	15	8	36	87	.223	.276	.275	.551	132	14	0	12	3	0	
26	Min Park	19	6	2	1	1	0	0	0	0	0	0	0	1	1	.000	.500	.000	.500	0	0	0	0	0	0	
27	Preston Tucker*	29	142	631	542	100	166	40	0	32	113	0	2	76	67	.306	.398	.557	.955	302	12	9	0	4	3	
28	Min Sang Yoo*	31	126	456	391	38	96	23	0	8	65	0	0	53	80	.246	.341	.366	.707	143	14	6	2	4	1	

2017 – 2022 (5년치)

연도별 트레이드, FA, 은퇴, 데뷔,
용병, 군복무, 개명선수

2 데이터 수집

2) 팀별 성적 (시즌 기록)

ex) 2019 팀타자 데이터

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	순	이름	팀	WAR*	G	타석	타수	득점	안타	2타	3타	홈런	루타	타점	
2	1	키움		19	32.83	1637	5658	4991	780	1405	252	37	112	2067	741
3	2	두산		19	28.28	1738	5669	4912	736	1364	235	31	84	1913	692
4	3	NC		19	25.8	1783	5631	4968	674	1383	249	26	128	2068	639
5	4	KT		19	22.89	1887	5579	4962	650	1374	186	20	103	1909	615
6	5	LG		19	22.13	1825	5536	4928	641	1316	233	15	94	1861	604
7	6	SK		19	19.36	1731	5542	4919	655	1290	218	14	117	1887	622
8	7	삼성		19	17.45	1652	5533	4866	622	1245	230	26	122	1893	598
9	8	KIA		19	17.26	1807	5505	4874	605	1286	248	19	76	1800	566
10	9	한화		19	15.02	1738	5482	4882	607	1249	223	16	88	1768	562
11	10	롯데		19	10.91	1821	5488	4919	578	1231	214	22	90	1759	545

2017 - 2022 (5년치)

팀별 데이터, 팀별 타자 및
투수 성적

1	도루	도실	볼넷	사구	고4	삼진	병살	희타	희비	타율	홀루	장타	OPS	wOBA	wRC+	WAR*	WPA
2	110	33	514	69	13	943	122	36	48	0.282	0.354	0.414	0.768	0.352	114.2	32.83	9.03
3	102	38	562	73	16	799	117	39	83	0.278	0.355	0.389	0.745	0.344	109.9	28.28	6.12
4	86	45	456	83	14	880	126	64	60	0.278	0.345	0.416	0.762	0.346	106.6	25.8	-1.13
5	104	44	476	51	15	974	113	45	45	0.277	0.344	0.385	0.728	0.336	102.1	22.89	1.42
6	108	47	434	77	17	962	117	49	48	0.267	0.333	0.378	0.711	0.327	99	22.13	2.54
7	118	50	462	87	20	966	92	35	38	0.262	0.334	0.384	0.718	0.331	97.4	19.36	0.55
8	107	42	478	78	15	976	107	58	53	0.256	0.329	0.389	0.718	0.328	94.7	17.45	-4.93
9	87	42	496	59	13	919	108	31	45	0.264	0.336	0.369	0.706	0.328	95.1	17.26	-3.84
10	105	51	441	70	12	1090	112	44	45	0.256	0.324	0.362	0.686	0.317	92	15.02	-7.1
11	62	31	430	64	12	1085	108	35	40	0.25	0.316	0.358	0.674	0.312	85.8	10.91	-10.09

② 데이터 수집

3) 선수별 데이터

ex) 2020 투수 데이터 일부

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
순	이름	팀	WAR	출장	완투	완봉	선발	승	패	세	홀드	이닝	실점	자책	타자	안타	2타	3타	홈런	볼넷	고4	사구	삼진
1	데스파이너	20K		0	35	0	0	34	15	8	0	0	207.2	105	100	906	233	36	1	18	68	1	8
2	알칸타라	20두		0	31	1	0	31	20	2	0	0	198.2	58	56	795	174	38	3	12	30	0	9
3	스트레일	20롯데		0	31	0	0	31	15	4	0	0	194.2	61	54	777	148	27	2	10	51	0	7
4	루친스키	20N		0	30	0	0	30	19	5	0	0	183	67	62	770	173	12	3	14	57	0	11
5	뷰캐넌	20삼		0	27	1	0	27	15	7	0	0	174.2	71	67	728	172	19	2	16	50	0	7
6	컬리	20L		0	28	1	1	28	15	7	0	0	173.1	67	64	705	160	29	1	16	40	0	11
7	양현종	20K		0	31	0	0	31	11	10	0	0	172.1	99	90	746	180	40	4	13	64	0	5
8	서폴드	20한		0	28	1	1	28	10	13	0	0	165	107	90	734	203	30	3	19	42	0	13
9	핀토	20S		0	30	0	0	30	6	15	0	0	162	121	111	766	198	35	1	19	90	3	11
10	요키시	20키		0	27	0	0	27	12	7	0	0	159.2	53	38	640	144	26	3	6	25	0	6
11	가놈	20K		0	28	0	0	28	11	8	0	0	159.2	86	77	702	162	33	8	7	64	0	9
12	쿠에바스	20K		0	27	0	0	27	10	8	0	0	158	80	72	668	152	26	1	16	46	0	9
13	라이트	20N		0	29	0	0	29	11	9	0	0	157.2	87	82	696	164	27	2	12	63	0	10
14	박종훈	20S		0	29	0	0	28	13	11	0	0	157.1	90	84	691	146	21	3	14	78	1	22
15	브룩스	20K		0	23	1	1	23	11	4	0	0	151.1	43	42	586	131	18	0	4	24	0	3
16	임찬규	20L		0	27	0	0	26	10	9	0	0	147.2	76	67	641	143	29	2	14	65	0	5
17	박세웅	20롯데		0	28	0	0	28	8	10	0	0	147.1	85	77	658	177	27	2	20	47	0	8
18	최재홍	20삼		0	26	1	1	26	11	6	0	0	146	67	58	617	131	27	1	12	51	0	7
19	문승원	20S		0	25	1	0	25	6	8	0	0	145.2	64	59	610	136	29	0	13	45	0	6
20	월슨	20L		0	25	0	0	25	10	8	0	0	144.2	81	71	628	161	23	2	13	39	0	8
21	배제성	20K		0	26	0	0	26	10	7	0	0	141.1	67	62	619	130	30	1	12	76	0	4
22	원태인	20삼		0	27	0	0	26	6	10	0	0	140	84	76	626	162	32	1	19	56	2	5
23	유희관	20두		0	27	0	0	27	10	11	0	0	136.1	91	76	622	190	37	5	14	39	4	3
24	한현희	20키		0	25	0	0	25	7	9	0	0	135.2	77	75	587	141	20	1	10	37	2	14
25	소형준	20K		0	26	0	0	24	13	6	0	0	133	63	57	572	141	23	0	6	45	0	6
26	노경은	20롯데		0	25	1	0	24	5	10	0	0	133	76	72	574	139	31	1	19	40	2	3

2017 - 2022 (5년치)
선수별로 정리한 타자 및 투수 성적

2 데이터 수집

4) 정규시즌 일정

ex) 2022 정규시즌 일정

04.26(화)	SSG	한화	3	1	1	0	05.07(토)	KIA	한화	13	2	1	0
	NC	두산	4	8	0	1		KT	1두산	8	1	1	0
	SSG	롯데	8	1	1	0		삼성	롯데	4	1	1	0
	LG	삼성	4	7	0	1		LG	NC	6	3	1	0
	KIA	KT	10	5	1	0		SSG	키움	3	0	1	0
04.27(수)	키움	한화	2	5	0	1	05.08(일)	KIA	한화	5	4	1	0
	NC	두산	5	6	0	1		KT	두산	5	0	1	0
	SSG	롯데	1	1	0	0		삼성	롯데	4	2	1	0
	LG	삼성	7	0	1	0		LG	NC	3	2	1	0
	KIA	KT	1	3	0	1		SSG	키움	6	2	1	0
04.28(목)	키움	한화	7	0	1	0	05.10(화)	KIA	한화	7	6	1	0
	NC	두산	9	5	1	0		한화	LG	1	9	0	1
	SSG	롯데	1	4	0	1		NC	롯데	0	7	0	1
	LG	삼성	6	4	1	0		SSG	삼성	3	1	1	0
	KIA	KT	3	5	0	1		KT	KIA	0	1	0	1
04.29(금)	키움	한화	3	1	1	0	05.11(수)	두산	키움	9	0	1	0
	롯데	LG	9	4	1	0		한화	LG	2	5	0	1
	두산	SSG	7	8	0	1		NC	롯데	5	6	0	1
	한화	NC	4	9	0	1		SSG	삼성	5	6	0	1
	삼성	KIA	4	3	1	0		KT	KIA	10	5	1	0
04.30(토)	KT	키움	6	4	1	0	05.12(목)	두산	키움	5	1	1	0
	삼성	KIA	5	4	1	0		한화	1LG	7	1	1	0
	롯데	LG	3	1	1	0		NC	롯데	4	2	1	0
	두산	SSG	1	2	0	1		SSG	삼성	2	4	0	1
	한화	NC	2	0	1	0		KT	KIA	4	5	0	1

② 데이터 수집

5) FA 및 트레이드 WAR 값

A	B	C	D	E	F	G	H	I	J	K
player	2018	2019	2020	2021	average	비고	before	after	position	WAR_diff
박병호	6.62	5.5	1.81	1.47	3.85	FA	키움	kt	1루수	1.648
박건우	3.42	4.68	3.78	4.62	4.125	FA	두산	nc	우익수	0.215
나성범	3.91	1.54	5.18	4.51	3.785	FA	nc	kia	우익수	0.541
손아섭	3.33	2.75	4.93	5.3	4.0775	FA	롯데	nc	우익수	0.1675
박해민	3.38	1.26	2.49	3.2	2.5825	FA	삼성	lg	중견수	-1.1555
허도환	0.11	-0.65	0.23	0.29	-0.005	FA	kt	lg	포수	-2.705
심창민	1.86		-0.54	0.16	#####	트레이드	삼성	nc	구원투수	-5.79467
김응민	-0.02	-0.02	-0.23	-0.04	-0.0775	트레이드	삼성	nc	포수	-2.6695
김태군		0.08	0.53	0.63	#####	트레이드	nc	삼성	포수	-1.93667
이학주		1.78	0.36	-0.04	0.7	트레이드	삼성	롯데	유격수	-0.874
최하늘		-0.34			-0.34	트레이드	롯데	삼성	구원투수	-6.842
김도현						트레이드	한화	kia		
이진영		-0.19	0.1	0.26	#####	트레이드	kia	한화	외야수	-1.76933
이민우	-0.36	-0.19	-0.75	-1.13	-0.6075	트레이드	kia	한화	선발투수	-6.6835
박동원	0.39	2.92	2.24	3.09	2.16	트레이드	키움	kia	포수	1.688
김태진	0.18	0.53	-0.65	0.64	0.175	트레이드	kia	키움	좌익수	-1.739
김민식	0.84	-0.14	0.08	0.62	0.35	트레이드	kia	ssg	포수	-1.152
김정빈			0.36	-0.72	-0.18	트레이드	ssg	kia	구원투수	-6.864
임석진						트레이드	ssg	kia		
정성곤	0.98	-0.67			0.155	트레이드	kt	ssg	구원투수	-5.667
이재호			-0.14		-0.14	트레이드	ssg	kt	구원투수	-7.57

③ 데이터 전처리

1) 기대 승률

실제승률 = 승수/(승수+패배 수)

기대승률 = 실제 승률이 아닌 예측 승률



③ 데이터 전처리

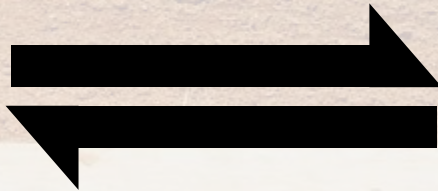
1) 기대 승률

기존 승률 계산법

피타고리안

기대 승률 =

$$\text{팀득점}^2 / (\text{팀득점}^2 + \text{팀실점}^2)$$



우리 조만의

기대 승률 =

$$0.0932 * (\text{팀평균득점} - \text{팀평균실점}) + 0.5$$

3 데이터 전처리

2) 지난 5개년 선형회귀분석 및 최종 승률

시즌	팀	X (평균팀득점 - 평균팀실점)	Y(최종승률)
2017년	KIA	1.131944444	0.608
	두산	1.1875	0.596
	NC	0.284722222	0.56
	넥센	0.173611111	0.486
	한화	-0.576388889	0.43
	롯데	0.291666667	0.563
	LG	0.152777778	0.489
	삼성	-1.069444444	0.396
	KT	-1.534722222	0.347
	SK	-0.041666667	0.524

```
In [10]: df = pd.read_csv(r'C:\Users\82107\Desktop\expectwin.csv')
df

Out[10]:
```

	Difference	Winrate
0	1.131944	0.608
1	1.187500	0.596
2	0.284722	0.560
3	0.173611	0.486
4	-0.576389	0.430
5	0.291667	0.563
6	0.152778	0.489
7	-1.069444	0.396
8	-1.534722	0.347
9	-0.041667	0.524
10	0.270833	0.486
11	1.005556	0.810

```
In [17]: df.shape

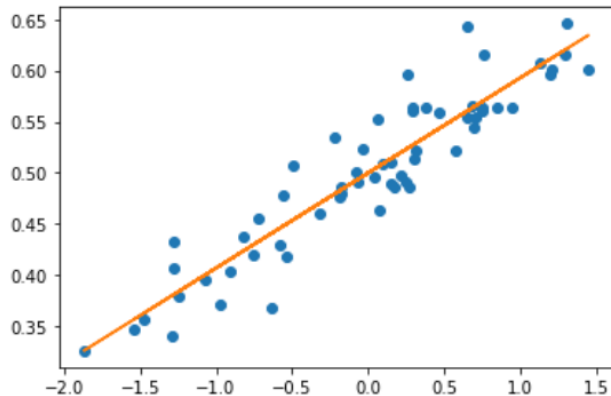
Out[17]: (60, 2)
```

17년 ~ 22년 현재까지의 경기 데이터 입력

3 데이터 전처리

2) 지난 5개년 선형회귀분석 및 최종 승률

```
In [12]: line_fitter = LinearRegression()  
line_fitter.fit(X.values.reshape(-1,1), y)  
  
Out[12]: LinearRegression()  
  
In [13]: plt.plot(X, y, 'o')  
plt.plot(X, line_fitter.predict(X.values.reshape(-1,1)))  
plt.show()
```



```
In [14]: line_fitter.coef_  
  
Out[14]: array([0.09322466])  
  
In [15]: line_fitter.intercept_  
  
Out[15]: 0.49993954260868106
```

$$\text{기대승률} = 0.0932 * (\text{팀평균득점} - \text{팀평균실점}) + 0.5$$

③ 데이터 전처리

3) 기대 승률 적용 결과 : 실제 승률과 기대 승률이 거의 일치

2022년	KIA	0.701754386	0.554	4	0.565404	2
	두산	0.25	0.491	6	0.5233	5
	NC	-0.637931034	0.368	10	0.440545	8
	넥센	0.25862069	0.596	2	0.524103	4
	한화	-1.24137931	0.379	9	0.384303	10
	롯데	-0.719298246	0.455	8	0.432961	9
	LG	0.74137931	0.561	3	0.569097	1
	삼성	-0.070175439	0.491	5	0.49346	7
	KT	0.068965517	0.464	7	0.506428	6
	SK	0.644067797	0.643	1	0.560027	3
			실제승률	실제 순위	기대승률	기대순위

③ 데이터 전처리

4) 시뮬레이션 원리 및 단계

구현 방법 :

파이썬 random, pandas 라이브러리 사용해 구현

3 데이터 전처리

4) 시뮬레이션 원리 및 단계

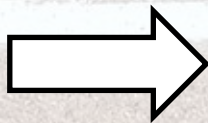
1. 각 팀의 기대승률에 따라
랜덤하게 승리 부여

ex)

한화 기대 승률 : 0.3(30%)

두산 기대 승률 : 0.6(60%)

한화가 두산 상대로 이길 확률은 $0.3 / (0.3 + 0.6) = 33.33\%$



#random.choice 함수로 구현
가중치를 기대 승률로 설정

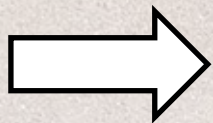
③ 데이터 전처리

4) 시뮬레이션 원리 및 단계

2-1)* 프로야구 경기 수에
맞춰 팀 당 144경기 진행

ex)

한화는 다른 9개 팀들을 16번씩 만남 ($9 \times 16 = 144$)



이긴 팀에 승리 횟수 1을 더하는 형식

* 5년치 데이터와 선수 이적을 반영한 시즌 Preview

3 데이터 전처리

4) 시뮬레이션 원

2022



년 06월06일 (2022년 06월05일 기준)

순위	팀명	경기	승	패	무	승률	게임차
1	SSG	56	36	18	2	0.667	0
2	키움	55	33	22	0	0.600	3.5
3	LG	56	31	24	1	0.564	5.5
4	KIA	55	30	24	1	0.556	6
5	삼성	54	26	28	0	0.481	10
6	두산	53	25	27	1	0.481	10
7	KT	55	25	29	1	0.463	11
8	롯데	54	24	28	2	0.462	11
9	한화	55	21	34	0	0.382	15.5
10	NC	55	19	36	0	0.345	17.5

이션

3 데이터 전처리

4) 시뮬레이션 원리 및 단계

```
df = pd.read_csv(r"C:\Users\W82103\Desktop\Wkbo.csv")
total_rank = [['doosan'], ['hanwha'], ['kia'], ['kiwoom'], ['kt'], ['lg'], ['lotte'], ['nc'], ['samsung'], ['ssg']]
rank_count = [['doosan'], ['hanwha'], ['kia'], ['kiwoom'], ['kt'], ['lg'], ['lotte'], ['nc'], ['samsung'], ['ssg']]
team = ['doosan', 'hanwha', 'kia', 'kiwoom', 'kt', 'lg', 'lotte', 'nc', 'samsung', 'ssg']

def simulation():
    ssg = [0.560, 36]
    lg = [0.569, 31]
    doosan = [0.523, 30]
    lotte = [0.433, 24]
    kiwoom = [0.524, 33]
    samsung = [0.493, 26]
    kt = [0.506, 25]
    kia = [0.565, 30]
    hanwha = [0.384, 21]
    nc = [0.440, 19]
    game = [doosan, hanwha, kia, kiwoom, kt, lg, lotte, nc, samsung, ssg]
    team = ['doosan', 'hanwha', 'kia', 'kiwoom', 'kt', 'lg', 'lotte', 'nc', 'samsung', 'ssg']
    game_win = []
```

[기대승률, 6월 6일 기준 승리 수]

#시뮬레이션 남은 경기 수

```
for i in range(10):
    for j in range(10):
        if df[team[i]][j] != 0:
            for x in range(df[team[i]][j]):
                a = game[i][0]
                b = game[j][0]
                win = random.choices(['a', 'b'], weights = (a, b))
                if win == ['a']:
                    game[i].append(1)
                    game[j].append(0)
                else:
                    game[i].append(0)
                    game[j].append(1)
```

③ 데이터 전처리

4) 시뮬레이션 원리 및 단계

3) 승리 수에 따라 순위 결정

모든 경기가 끝나면 승리 수에 따라 순위 부여



이 과정 10000번 반복

③ 데이터 전처리

4) 시뮬레이션 원리 및 단계

4) 해당 등수 횟수가 확률로 산출

ex)

한화가 1등을 10번 했다면 한화가 우승할 확률
 $= 10/10000 = 0.1\%$

한화가 10등을 9000번 했다면 한화가 꼴지할 확률
 $= 90\%$

3 데이터 전처리

5) 시뮬레이션 구현 - 지난 5개년 기준 시즌 preview

team	1	2	3	4	5	6	7	8	9	10	mean	expect_win_rate
doosan	3872	2058	1288	986	611	448	336	236	124	41	2.7305	0.597
kiwoom	1746	1785	1480	1248	1023	904	687	533	391	203	3.9909	0.548
lg	1138	1285	1337	1311	1174	1039	950	763	640	363	4.7211	0.526
nc	838	1071	1223	1234	1215	1137	1077	923	790	492	5.1435	0.516
kt	678	962	1124	1176	1181	1163	1158	1052	879	627	5.4264	0.502
kia	555	893	1031	1107	1241	1223	1120	1073	1002	755	5.6397	0.493
ssg	598	840	1014	1087	1143	1170	1162	1200	1009	777	5.6988	0.502
lotte	299	543	680	784	1004	1150	1288	1438	1499	1315	6.5642	0.472
samsung	224	403	578	725	913	1077	1238	1447	1703	1692	6.918	0.464
hanwha	52	160	245	342	495	689	984	1335	1963	3735	8.1669	0.415

3 데이터 전처리

5) 시뮬레이션 구현 - 6월 6일까지 기록 반영

team	1	2	3	4	5	6	7	8	9	10	mean	expect_win_rate
ssg	4594	2377	1417	888	465	182	58	17	1	1	2.1129	0.56
kiwoom	1649	2071	2049	1842	1330	694	292	64	9	0	3.2757	0.524
lg	1322	1896	2037	1832	1559	888	368	79	18	1	3.5056	0.569
kia	1238	1691	1985	1938	1734	924	355	103	29	3	3.6141	0.565
doosan	1079	1604	1777	1945	1807	1085	522	156	22	3	3.8073	0.523
samsung	67	187	377	740	1375	2487	2592	1487	534	154	6.2715	0.493
kt	49	154	320	669	1331	2495	2811	1494	540	137	6.3477	0.506
lotte	2	20	36	132	331	952	1990	3766	1912	859	7.7901	0.433
nc	0	0	1	8	42	172	533	1505	3661	4078	9.0777	0.44
hanwha	0	0	1	6	26	121	479	1329	3274	4764	9.1974	0.384

③ 데이터 전처리

6) 시뮬레이션 분석 후 시사점



스포츠는 강한 팀이 무조건 승리하는 것이 아니라, 승리할
확률이 높은 것

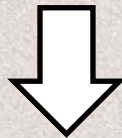
기대 승률에 의한 확률적 승리

③ 데이터 전처리

6) 시뮬레이션 분석 후 시사점



모든 팀은 잘하는 팀을 상대할 때보다 못하는 팀을
상대할 때 승리할 확률이 올라갈 것



승리 부여할 때 가중치로 이 부분을 구현

ex)

기대 승률이 0.6인 두산은 0.3인 한화를 상대하면
승리 확률 $66.66\%(0.6/(0.3+0.6))$

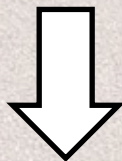
BUT 0.6인 ssg를 상대한다면 $50\%(0.6/(0.6+0.6))$

③ 데이터 전처리

6) 시뮬레이션 분석 후 시사점



우연 등 다양한 변수에 의해 못하는 팀도 우승할
가능성 존재



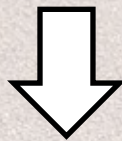
최대한 많은 횟수(10000번)의 반복으로 그 우연이
얼만큼의 확률로 발생하는지 확인 가능

③ 데이터 전처리

6) 시뮬레이션 분석 후 시사점



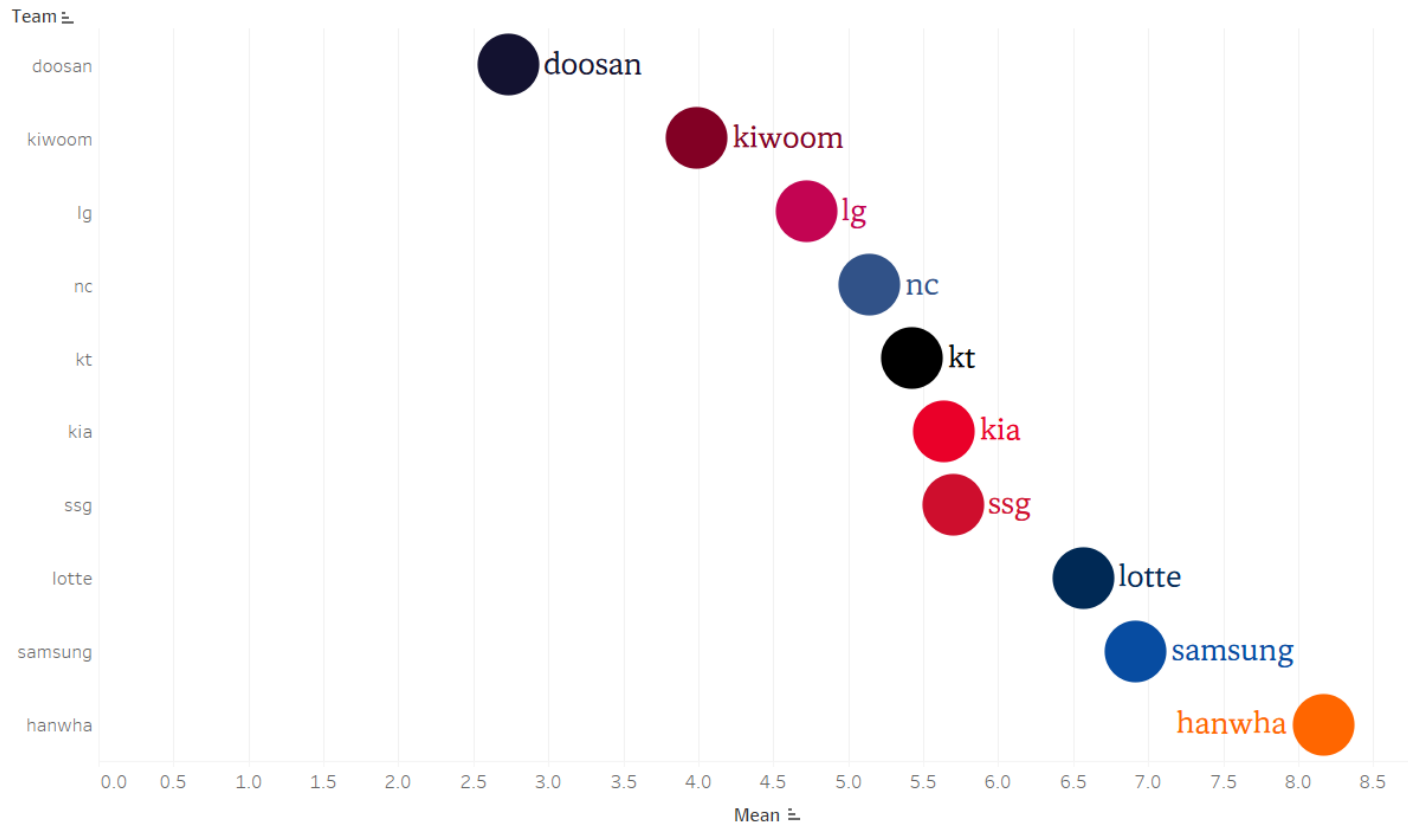
많은 횡수의 반복으로 그 팀의 실력에 따라
몇 등이 적합한지 = 확률이 제일 높은지 확인 가능



ex)
한화가 꼴등 횡수가 다른 팀들에 비해 많다면,
한화의 실력은 꼴등이 적합

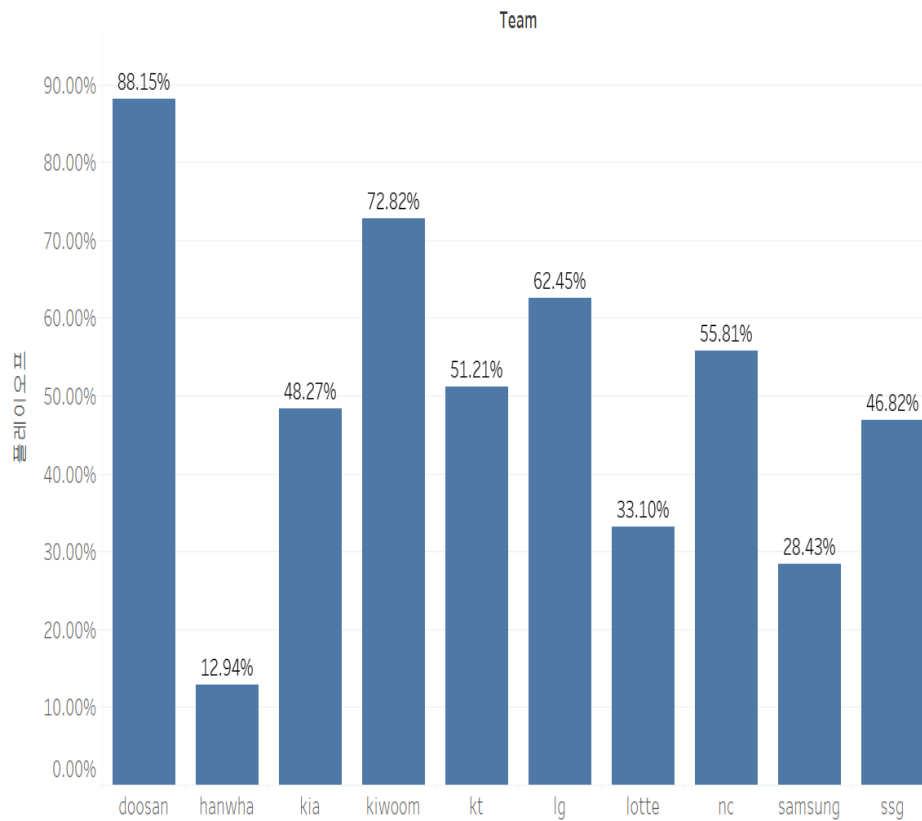
4 데이터 시각화 – 지난 5개년 기준 시즌 preview

지난 5개년치 시뮬레이션 결과

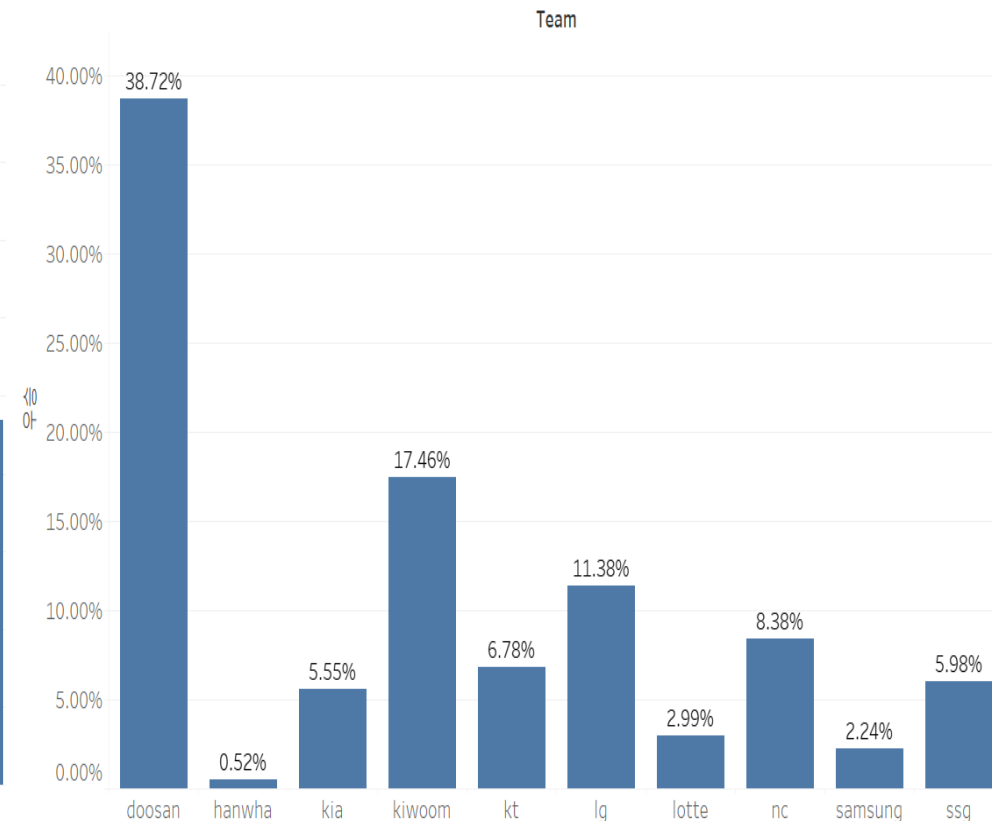


4 데이터 시각화 - 2022 최종 순위 예측

<플레이오프 진출 확률 - 5개년>

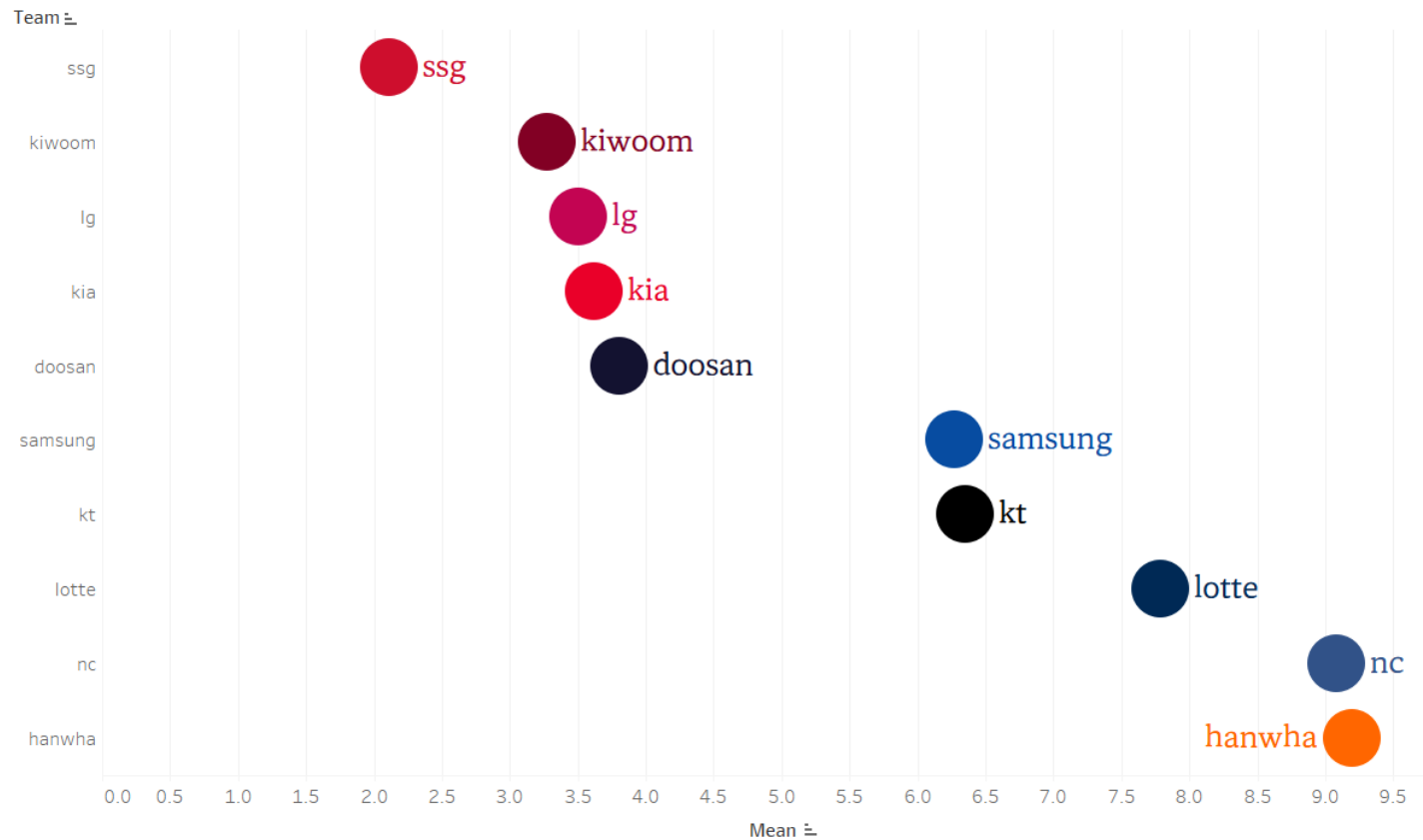


<우승 확률 - 5개년>



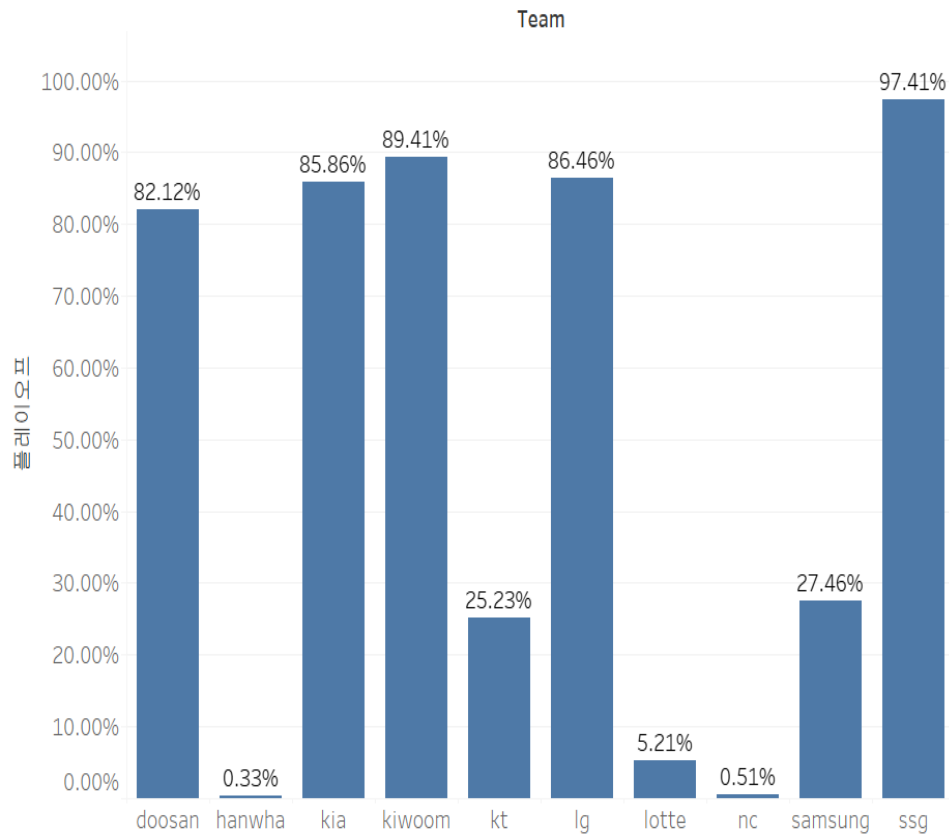
4 데이터 시각화 - 2022 최종 순위 예측

2022 시뮬레이션 결과

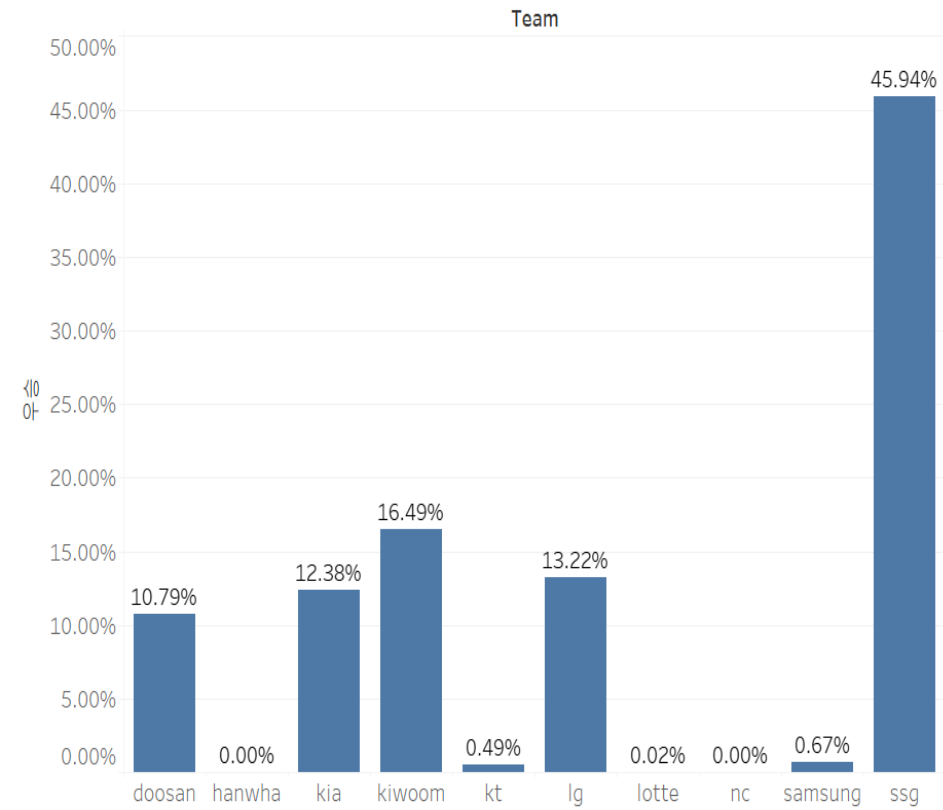


4 데이터 시각화 - 2022 최종 순위 예측

<플레이오프 진출 확률 - 2022>



<우승 확률 - 2022>



4 데이터 시각화 - 올해 최종 순위 예측



5 의의 및 한계

의의

우리만의 방법으로 기대 승률과 최종 순위 예측

득점과 실점이라는 두가지 변수에 의해
최종 승률과의 상관관계 도출

실제 순위와 기대 승률이 거의 일치

한계

외국인 선수 영입, 신인 선수의 데뷔 등
기준 데이터가 없는 선수의 성적을 반영하지 못함

팀 성적에는 주축 선수의 부상 및 부진, 감독의 변화 등
다양한 요인이 있지만 이를 수치화하여 반영하기 어려움

Dscover 스포츠 분야 프로젝트

감사합니다.

김지환 손유승 최범규 임세민 임선주