



빅데이터 **HADOOP ECOSYSTEM** 구축을 활용한 **KBO** 10개 구단 개선책 제시

2025. 04. 30

CONTENTS

I. 개요

1. 배경
2. 데이터 수집 / 요약
3. 환경 구성

II. 서론

1. HIVE를 이용한 SQL 처리 흐름
2. HIVE 쿼리 데이터 분석
3. KBO 공인구 반발계수

III. 분석

1. 20' ~ 24' 시즌 흐름도
2. 24' 구단 별 강점/약점

IV. 결론

1. 구단 별 25' 대비 방법

I. 개요

1. 배경

2. 데이터 수집 / 요약

3. 환경 구성

1. 배경 - 주제 선정

I. 개요

2024 시즌 KBO 총 관중 **1,088만 7,750명** 리그 출범 이후 최대 관중 수 기록함. 이를 통해 야구에 대한 대중의 관심이 크게 증가함. 팬층이 다양해진 만큼, 데이터 기반 분석을 통해 **팀과 선수에 대한 인사이트를 제공**하고자 이 주제를 선정

1000만 관중 시대 연 KBO리그, 누적 시청자 수는 2억5000만명

뉴스시스

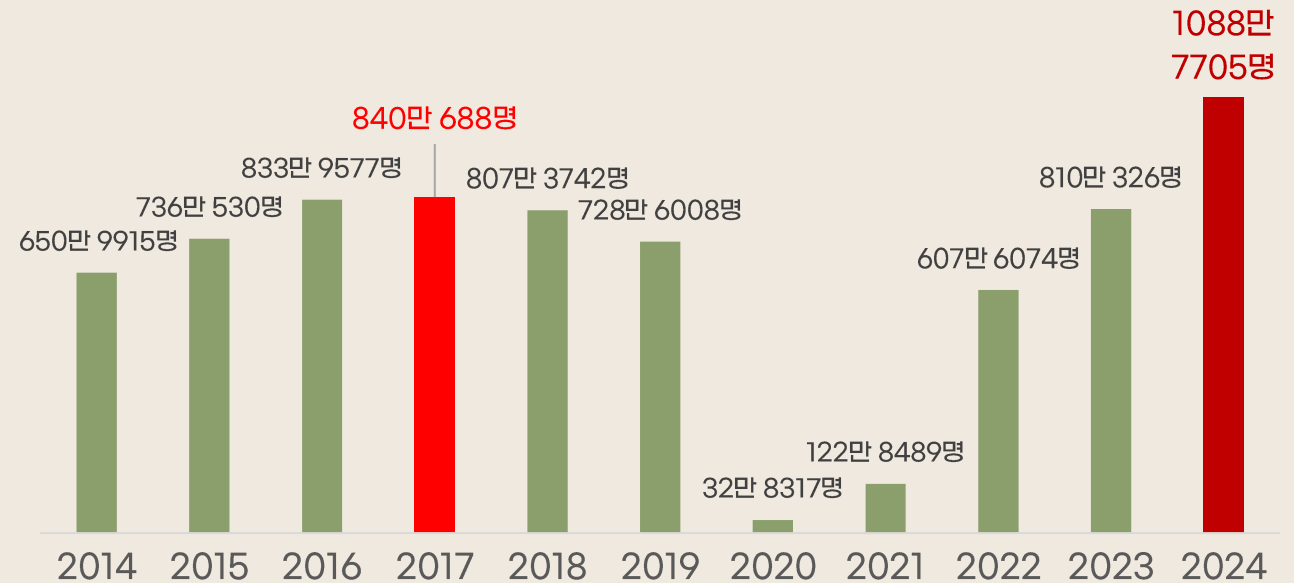
업데이트 2024.09.16. 18:13

가



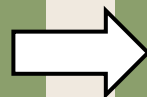
사상 첫 1000만 관중 달성한 프로야구

KBO 연도별 관중 수



*출처 : KBO 공식 홈페이지, 2025.04

신규 팬 유입 + 기존 팬 유지

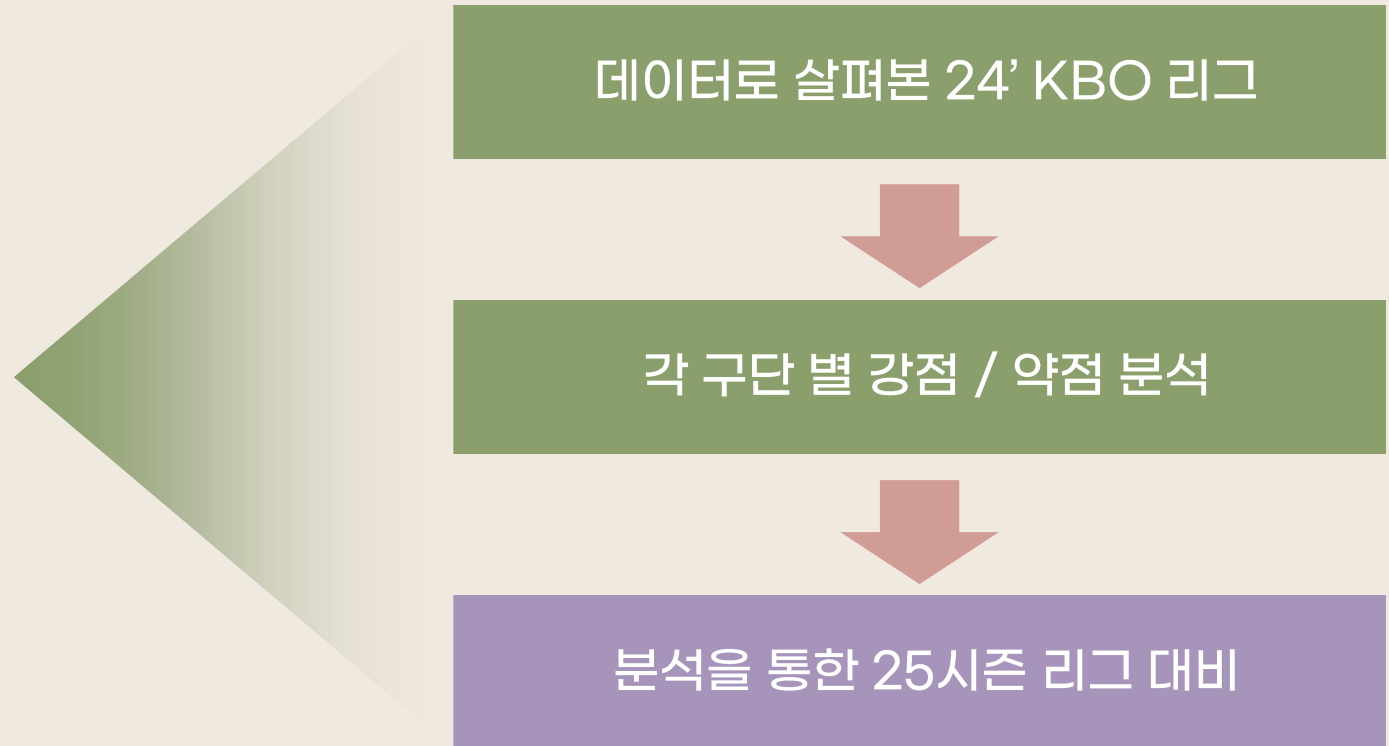


2025 시즌 KBO 관중 수 ↑ 예상



팬·구단 맞춤형 정보 제공 필요성 ↑

KBO 리그는 매 시즌 예측 불허의 판도 변화 존재, 단순 성적 비교만으로는 팀 전력 판단 불가



2. 데이터 수집 / 요약

I. 개요

2020 ~ 2024년 KBO 정규리그 순위 데이터 및 2024년 각 팀 성적 데이터 수집



5개년리그 데이터

필드명	설명	도메인
RANK	팀 순위	1 ~ 10
TEAM	팀명	KIA, 삼성, LG, 두산 ...
G	경기 수	144
WIN	승리 수	0 ~ 144
DRAW	무승부 수	0 ~ 144
LOSE	패배 수	0 ~ 144
GB	승차 수	0 ~ 144
WPCT	승률	0 ~ 1
YEAR	연도	20 ~ 24



24' 팀 성적

필드명	설명	도메인
Rank	팀 순위	1~10
Team	팀명	KIA, 삼성, LG, 두산 ...
Year	연도	20 ~ 24
WAR	팀 총 WAR	실수
G	경기 수	144
oWAR	공격 WAR	실수
dWAR	수비 WAR	



총 32개 필드

AVG	팀 타율	0.000 ~ 1.000
OBP	팀 출루율	
SLG	팀 장타율	
OPS	팀 출루율 + 장타율	
R/ePA	타석 당 득점 비율	실수
wRC+	조정 득점 생산력	정수

2. 데이터 수집 / 요약

I. 개요

2024년 투수 및 타자 성적 데이터 수집



24' 투수 성적

필드명	설명	도메인
Rank	선수 순위	1~50
Name	선수 이름	네일, 광빈, 원태인 ...
Team	소속팀	KIA, 삼성, LG, 두산 ...
Year	연도	20 ~ 24
WAR	대체 수준 대비 승리 기여	실수



총 37개 필드

ERA	평균자책점	실수
RA9	9이닝 당 실점	
rRA9	조정 실점	
rRA9pf	구장 조정 실점	
FIP	수비 무관 자책점	
WHIP	이닝 당 주자 허용	



24' 타자 성적

필드명	설명	도메인
Rank	선수 순위	1~50
Name	선수 이름	김도영, 오스틴, 데이비슨 ...
Team	소속팀	KIA, 삼성, LG, 두산 ...
Year	연도	20 ~ 24
WAR	대체 수준 대비 승리 기여	실수



총 33개 필드

AVG	타율	실수
OBP	출루율	
SLG	장타율	
OPS	출루율 + 장타율	
R/ePA	기대 타석 당 득점	
wRC+	조정 득점 생산력 지표	

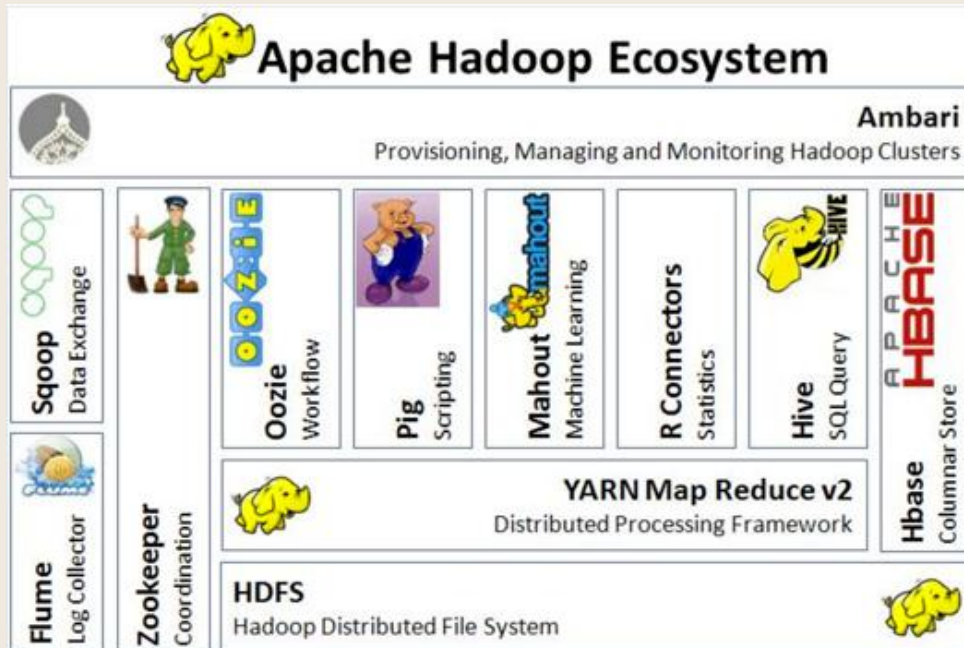
3. 환경 구성

I. 개요

Hadoop Ecosystem을 기반으로 데이터를 수집·저장하고, 논리적·물리적 모델링을 통해 데이터 웨어하우스에 적재함.

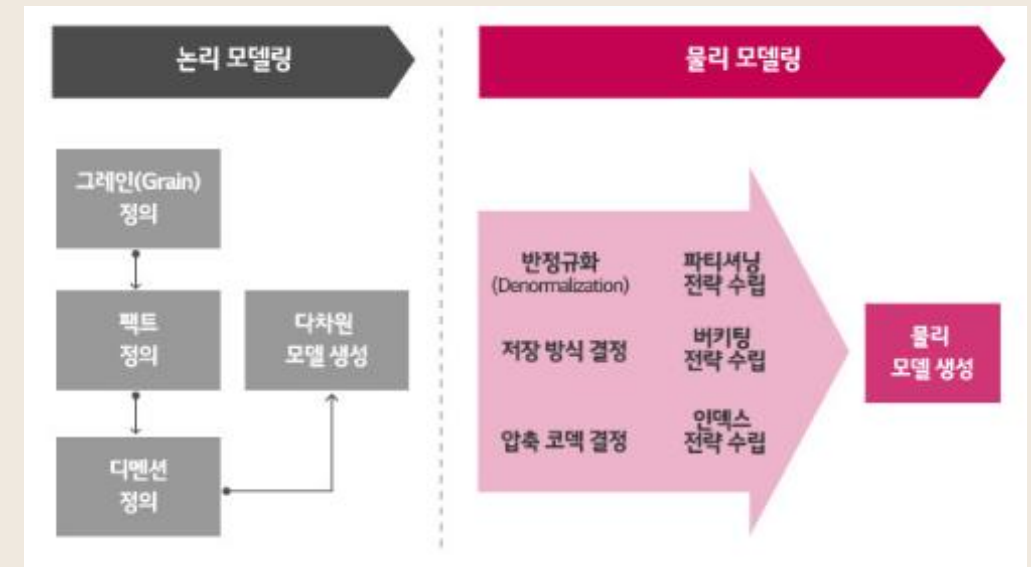
저장 환경 구성

빅데이터 분석을 위한 환경 구성
Hadoop Ecosystem



데이터 이관 및 적재

데이터 웨어하우스 모델링 적재



II. 처리

1. HIVE를 이용한 SQL 처리 흐름
2. HIVE 쿼리 데이터 분석
3. KBO 공인구 반발계수

1. HIVE를 이용한 SQL 처리 흐름

II. 서론

수집한 KBO 데이터를 Hive 테이블로 구성하고 적재하여 SQL 분석을 준비함

자료형 결정

RANK		INT
TEAM		STRING
G		INT
WIN		INT
DRAW		INT
LOSE		INT
GB		FLOAT
WPCT		FLOAT
YEAR		INT

테이블 생성

```
CREATE TABLE ranking(  
RANK INT,  
TEAM STRING,  
G INT,  
WIN INT,  
DRAW INT,  
LOSE INT,  
GB FLOAT,  
WPCT FLOAT,  
YEAR INT)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
LINES TERMINATED BY '\n'  
STORED AS TEXTFILE;
```

데이터 적재

RANK	TEAM	G	WIN	DRAW	LOSE	GB	WPCT	YEAR
1	NC	144	83	6	55	0	0.601	20
2	KT	144	81	1	62	4.5	0.566	20
3	LG	144	79	4	61	5	0.564	20
3	두산	144	79	4	61	5	0.564	20
5	키움	144	80	1	63	5.5	0.559	20
6	KIA	144	73	0	71	13	0.507	20
7	롯데	144	71	1	72	14.5	0.497	20
8	삼성	144	64	5	75	19.5	0.46	20
9	SSG	144	51	1	92	34.5	0.357	20
10	한화	144	46	3	95	38.5	0.326	20
1	삼성	144	76	9	59	0	0.563	21
1	KT	144	76	9	59	0	0.563	21
3	LG	144	72	14	58	1.5	0.554	21

```
load data local inpath  
"KBO20_24_팀순위.csv"  
overwrite into table ranking;
```

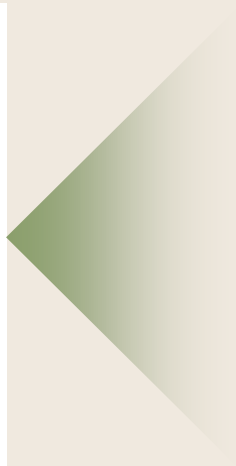
✓ STEP 1 완료 → 💡 Hive 분석

2. HIVE 쿼리 데이터 분석

II. 서론


SQL 작성 및 결과 예시

```
hive> SELECT
>   TRIM(team) AS team,
>   MAX(CASE WHEN year = 20 THEN rank END) AS `20`,
>   MAX(CASE WHEN year = 21 THEN rank END) AS `21`,
>   MAX(CASE WHEN year = 22 THEN rank END) AS `22`,
>   MAX(CASE WHEN year = 23 THEN rank END) AS `23`,
>   MAX(CASE WHEN year = 24 THEN rank END) AS `24`
> FROM
>   ranking
> WHERE
>   year IS NOT NULL AND rank IS NOT NULL AND TRIM(team) != 'TEAM'
> GROUP BY
>   team;
```



KIA	6	9	5	6	1
KT	2	1	3	2	5
LG	3	3	2	1	3
NC	1	7	6	4	9
SSG	9	6	1	3	5
두 산	3	4	9	5	4
롯데	7	8	8	7	7
삼성	8	1	7	8	2
키움	5	5	3	10	10
한화	10	10	10	9	8

```
hive> SELECT rank, team, g, win, draw, lose, gb, wpct
> FROM ranking
> WHERE year = 24
> ORDER BY rank ASC;
```



1	KIA	144	87	2	55	0.0	0.613
2	삼 성	144	78	2	64	9.0	0.549
3	LG	144	76	2	66	11.0	0.535
4	두 산	144	74	2	68	13.0	0.521
5	KT	144	72	2	70	15.0	0.507
5	SSG	144	72	2	70	15.0	0.507
7	롯데	144	66	4	74	20.0	0.471
8	한 화	144	66	2	76	21.0	0.465
9	NC	144	61	2	81	26.0	0.43
10	키 움	144	58	0	86	30.0	0.403

3. KBO 공인구 반발계수

II. 서론

반발계수 하락, 홈런 줄어든다?

스포츠 야구·MLB

KBO 공인구 반발계수, 작년보다 낮아졌다

김양희 기자

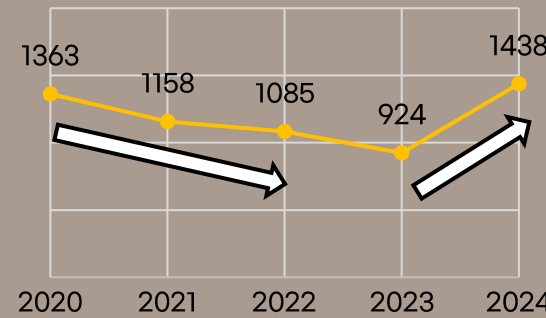
수정 2025-03-25 10:08 등록 2025-03-25 10:07

📈 2025 반발계수: 0.4123 (↓) (24': 0.4208)

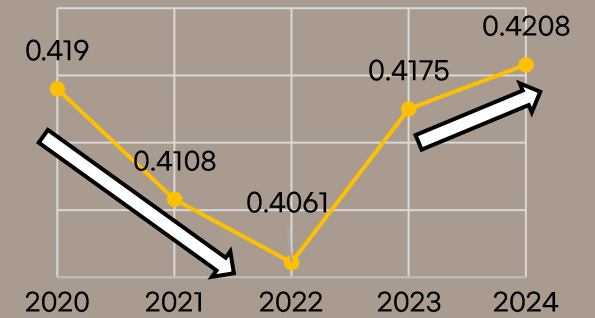
◆ -0.0085 → 예상 비거리 -1.7m

🏏 올해는 타·투 균형 예상

연도별 리그 홈런 개수



연도별 공인구 반발계수



2023년을 제외하면 반발계수와 홈런개수는 어느 정도 상관관계를 보임

예상 홈런 수 : **1100 ~ 1200개** 사이

(2023시즌은 제외하여 계산)

III. 분석

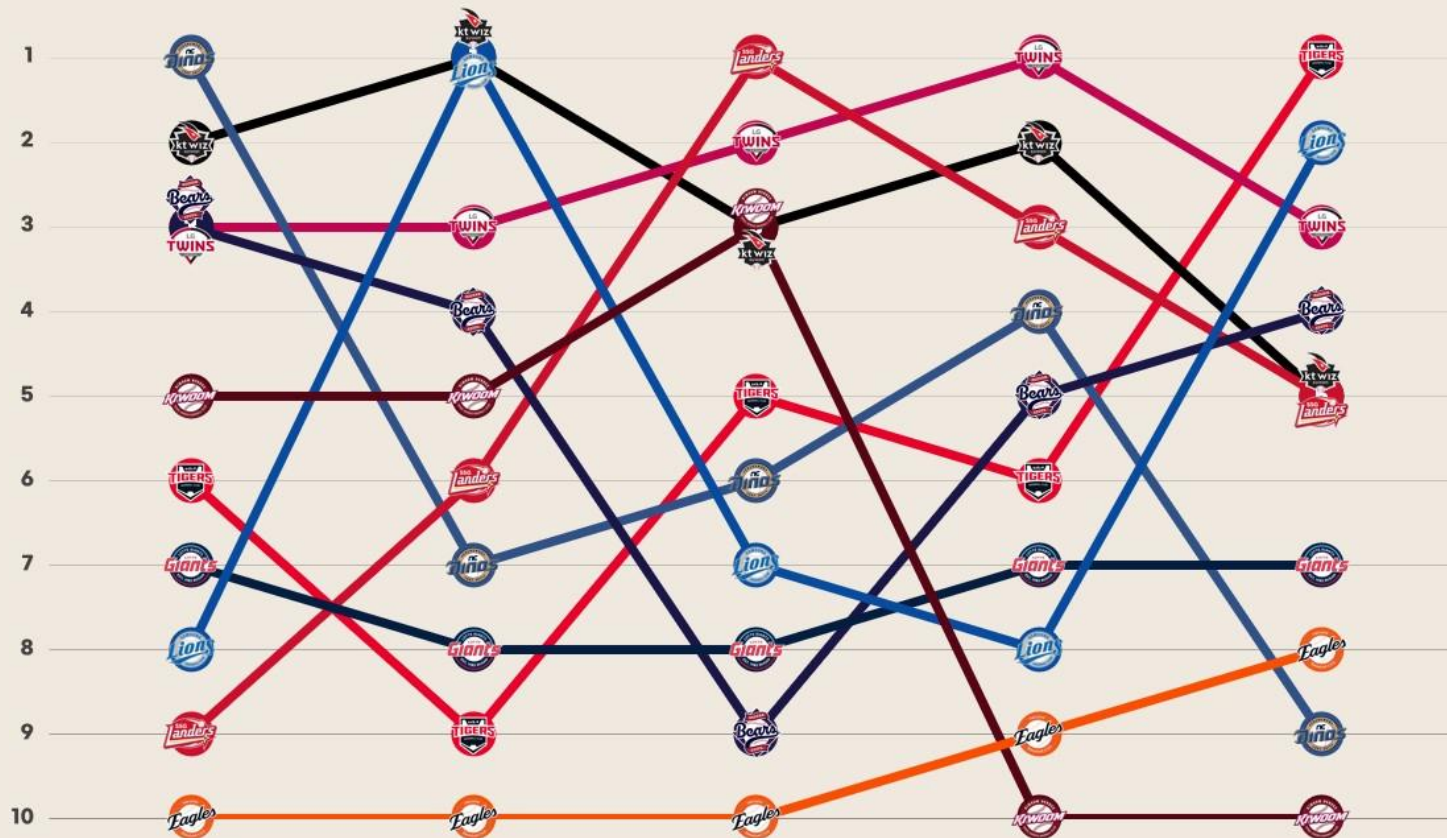
1. 20' ~ 24' 시즌 흐름도
2. 24' 리그 결산
3. 24' 구단 별 강점/약점

1. 20~24년 리그 흐름

III. 분석

최근 5년간 리그 순위 흐름 분석을 통해 안정성과 변동성에 따른 성적 예측을 제시

20~24 시즌별 순위 흐름도



LG·한화·롯데·KT

5년간 순위 큰 변동 없음, 안정적이지만 예측 가능한 흐름

NC·삼성·키움

상위-하위 오가는 롤러코스터 흐름, 예측 어려움

SSG·두산

중위권 중심의 등락 반복, 고비 넘기면 상위권 가능성





KIA

최근 꾸준한 상승세, 올해도 기대치↑

2. 24' 리그 결산

III. 분석

24시즌 정규 리그 최종 결과 및 각 팀의 시즌 평가

순위	팀	경기 수	승	무	패	승차	승률
1	KIA 	144	87	2	55	0	0.613
2	삼성 	144	78	2	64	9	0.549
3	LG 	144	76	2	66	11	0.535
4	두산 	144	74	2	68	13	0.521
5	KT 	144	72	2	70	15	0.507
5	SSG 	144	72	2	70	15	0.507
7	롯데 	144	66	4	74	20	0.471
8	한화 	144	66	2	76	21	0.465
9	NC 	144	61	2	81	26	0.43
10	키움 	144	58	0	86	30	0.403

5위 순위 결정전
KT 승

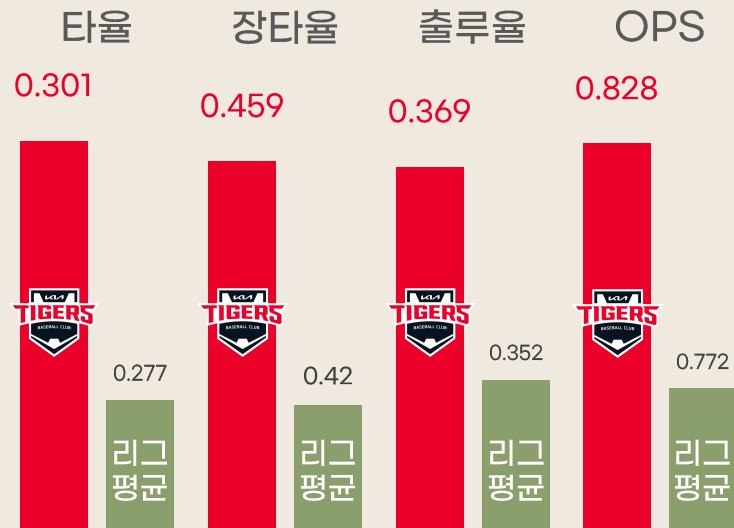
	시즌 내내 강력한 타선과 안정된 마운드 정규 우승 및 한국시리즈 우승
	시즌 초중반 불안했으나... 하반기 반등 하반기 구자욱의 미친 퍼포먼스
	디펜딩 챔피언 모습 보여주나 했는데... 선수진 부진으로 아쉽게 3위로 마무리
	시즌 초 불미스러운 사건으로 놀란 팬들 그래도 4위 안착 성공... 하지만 와카는...
	KBO 역사상 최초 5위 결정전 열세에도 불구하고 가을야구 진출 성공
	하반기 가을속 모드 발동... 막판 스퍼트 하지만 가을야구 문턱... 결정전 패배
	‘윤나고향’ 업고 포스트시즌 진출? 9월 불펜진 붕괴... 포스트시즌 탈락
	시즌 초반 7연승으로 분위기 땀는데... 결국 초반만 반짝... 8위 마무리
	시즌 초반 의심의 여지 없는 상위권... 부상 병동으로 인한 하락세... 9위 마감
	주축 선수 대거 이탈... 작년과 같은 흐름 그래도 승률 4할 넘겼다

기아 타이거즈 강점과 약점 분석

강점

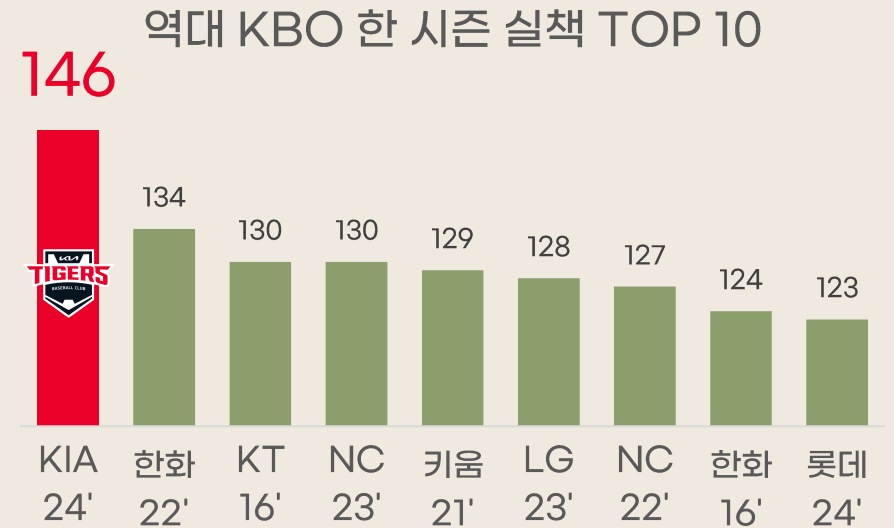


공격 4대 지표 싸슬이!
기아, 24시즌 리그 최강 타선 등극



약점

불명예 기록 추가...
기아, 한 시즌 역대 **최다 실책**



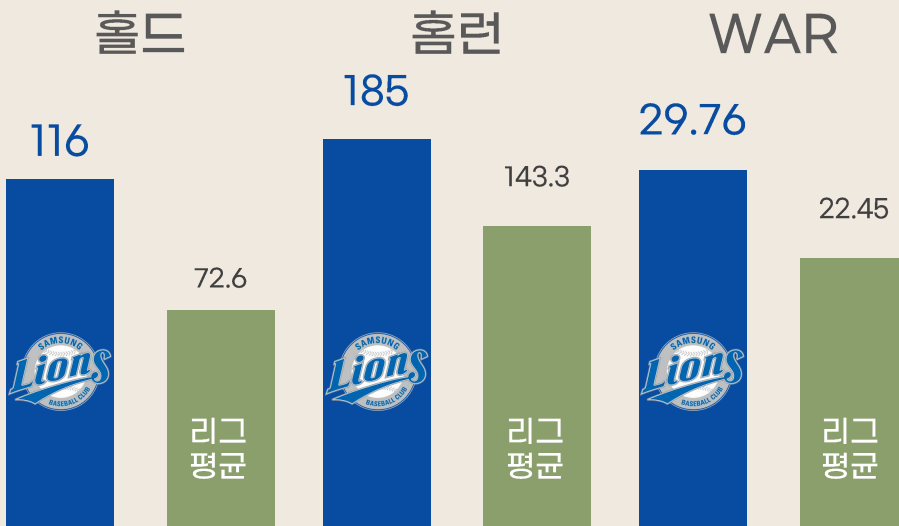
2. 삼성 라이온즈

III. 구단 분석

삼성 라이온즈 강점과 약점 분석

강점

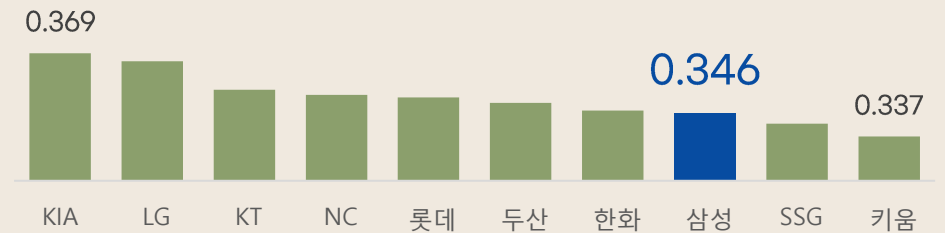
홈런·볼펜·수비 모두 리그 최강
완성도 높은 투타 밸런스를 갖춘 삼성



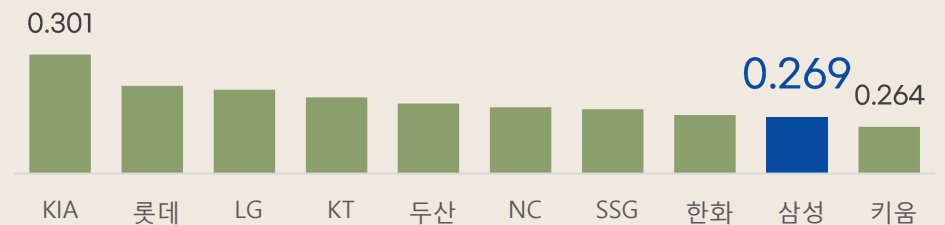
약점

홈런에 의존하는 빅볼 야구
낮은 타율과 출루율로 득점 생산의 한계 노출

24시즌 팀 출루율



24시즌 팀 타율



3. LG 트윈스

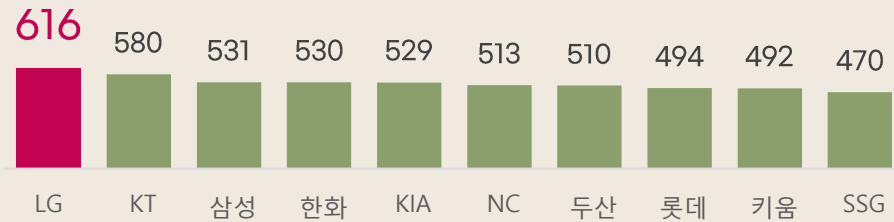
III. 구단 분석

LG 트윈스 강점과 약점 분석

강점

팀 최다 볼넷, 선구안 좋은 LG
공격/수비 상위권 밸런스형 팀

24시즌 팀 볼넷



dWAR

2.23



oWAR

26.87



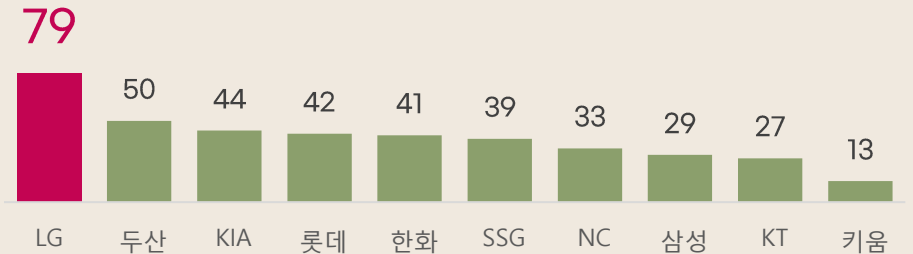
약점

도루 실패, 주루사 많아
출루 잘해도 공격 흐름 자주 끊김

24시즌 주루사



24시즌 도루 실패



4. 두산 베어스

III. 구단 분석

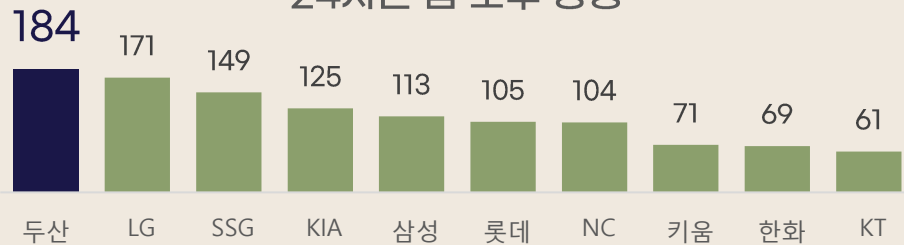
두산 베어스 강점과 약점 분석



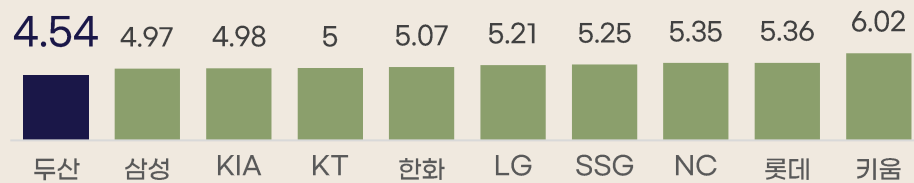
강점

도루 1위 + 강력한 불펜 자원,
빠르고 유연한 플레이

24시즌 팀 도루 성공



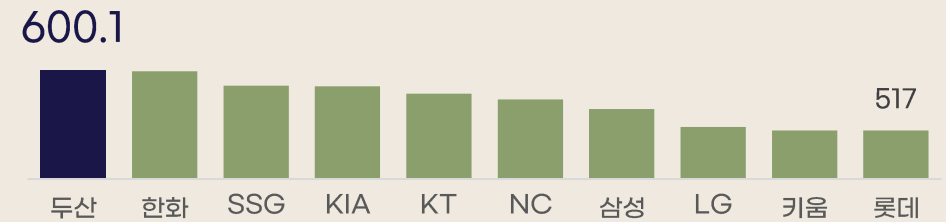
24시즌 구원 투수 평균자책점



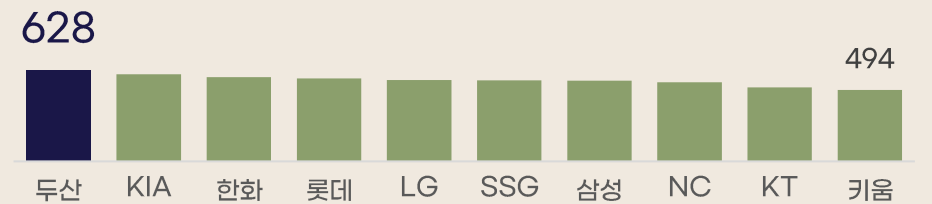
약점

구원 투수 최다 경기, 이닝 소화
일명 투마카세 선발 부상 + 부진

24시즌 구원투수 이닝 수



24시즌 구원 등판 수



5. KT 위즈

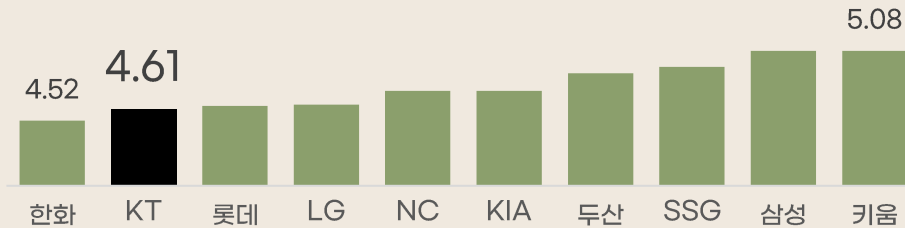
III. 구단 분석

KT 위즈강점과 약점 분석

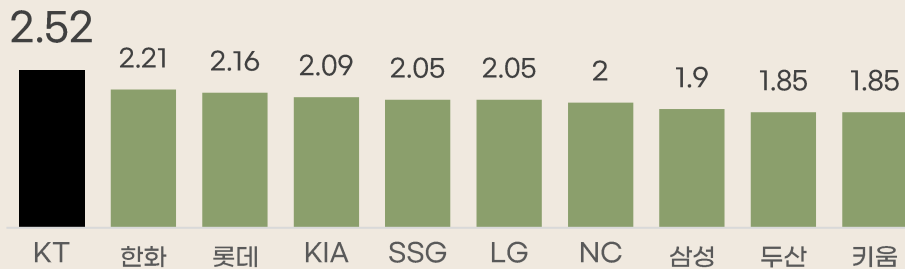
강점

올해 부상으로 약해진 투수진...
그래도 강점은 **투수**

24시즌 *FIP



24시즌 *K/BB



약점

노쇠화된 내야진
세대 교체 필요성 대두



김상수
35



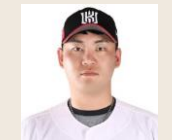
황재균
38



오윤석
33

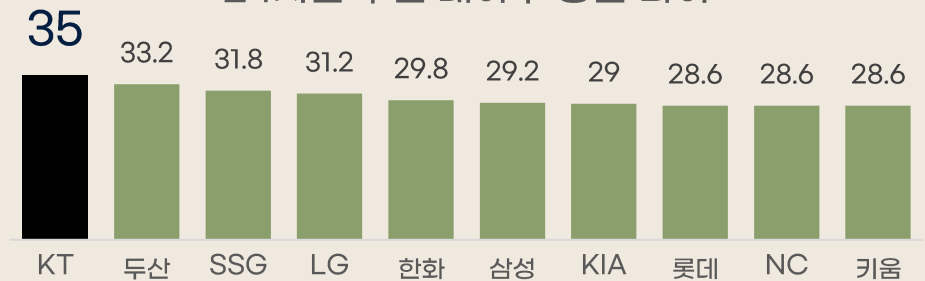


문상철
34



장성우
35

24시즌 주전 내야수 평균 나이



*FIP: 수비와 상관없는 독립된 피칭. 투수에게만 책임이 있는 평균자책점

*K/BB: 볼넷 삼진 비율. 볼넷 1개에 삼진 개수

6. SSG 랜더스

III. 구단 분석

SSG 랜더스 강점과 약점 분석

강점

거의 최악의 시즌을 보낸 24시즌
그래도 신인 덕에 웃다



박지환

타율 0.276
신인 최초 7연타석 안타
잠재성 높은 긍정 평가

정준재

타율 0.307
데뷔시즌 주전 2루수
차세대 테이블 세터

조병현

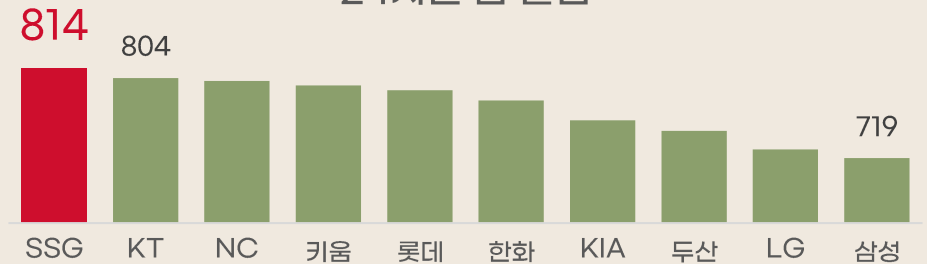
첫 풀타임 시즌
10타자 연속 삼진 기록
핵심 불펜 필승조 활약



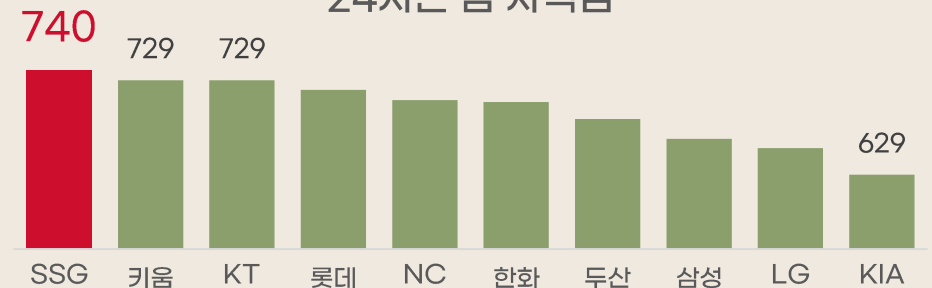
약점

리그 **최다 자책점 + 최다 실점**
망가진 투수 + 수비

24시즌 팀 실점



24시즌 팀 자책점



7. 롯데 자이언츠

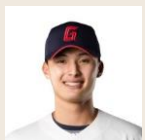
III. 구단 분석

롯데 자이언츠 강점과 약점 분석

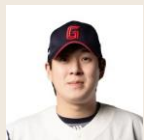


강점

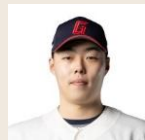
젊은 타선 '윤나고황+손' 리빌딩 성공
리그 2위 타율로 타격 활발



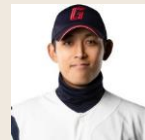
윤동희
0.293



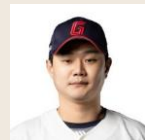
나승엽
0.312



고승민
0.308

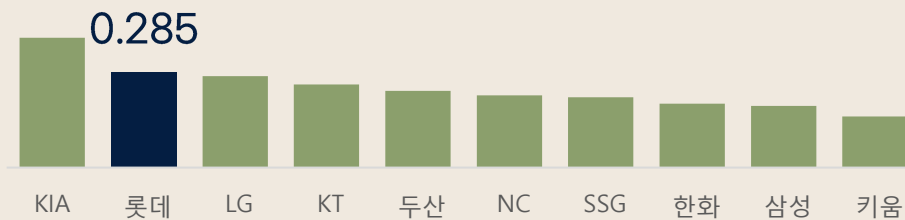


황성빈
0.320



손호영
0.318

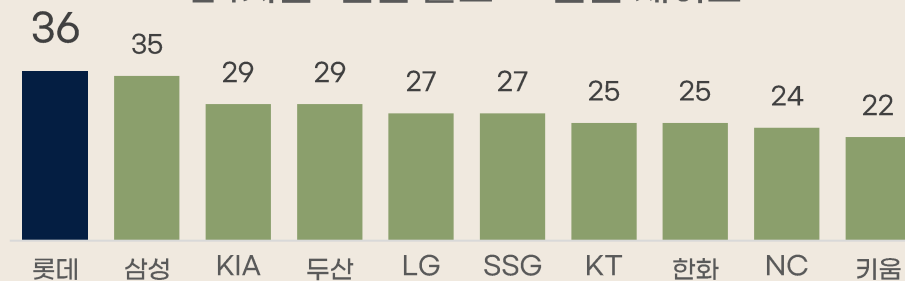
24시즌 팀 타율



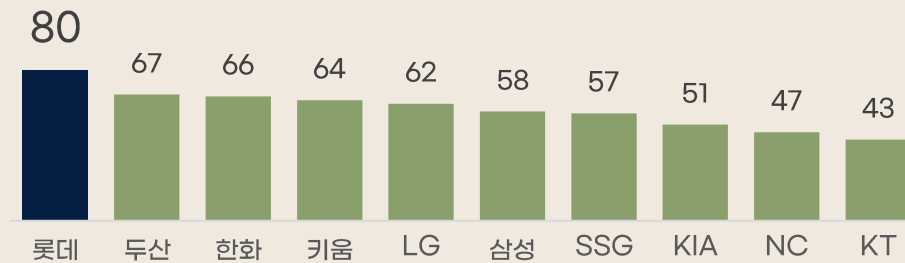
약점

포수·불펜 불안정
실점이 쉽게 나는 구조

24시즌 *블론 홀드 + *블론 세이브



24시즌 포수 폭투 허용



*블론 홀드 : 홀드 상황에 마운드에 올라왔지만 리드를 지키지 못하고 동점이나 역전을 허용한 경우

*블론 세이브 : 세이브 상황에서 등판했으나 리드를 지키지 못하고 동점이나 역전을 허용한 경우

8. 한화 이글스

III. 구단 분석

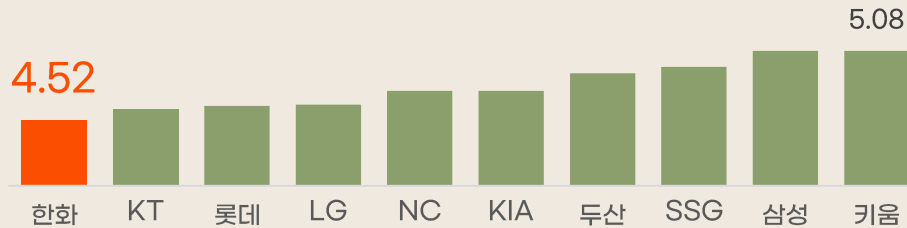
한화 이글스 강점과 약점 분석



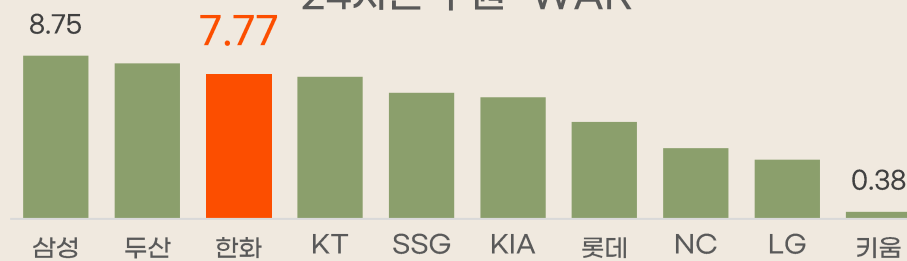
강점

선발 투수 성적은 아쉽지만
구원 WAR 3위, FIP 1위

24시즌 *FIP



24시즌 구원 *WAR



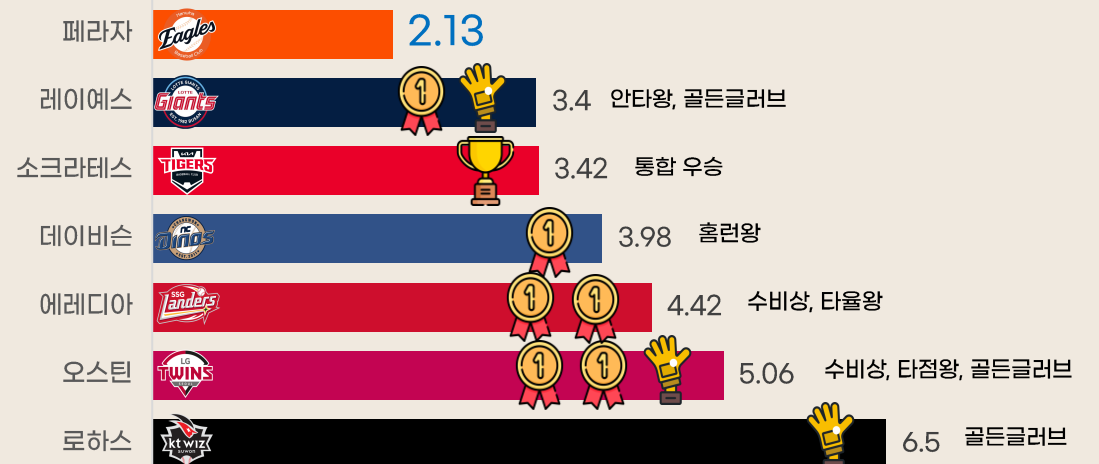
*FIP: 수비와 상관없는 독립된 피칭, 투수에게만 책임이 있는 평균자책점

*WAR: 대체 수준 대비 승리 기여, 선수가 팀에 기여하는 전체적인 가치

약점

계속된 용병 실패
타 구단 대비 전력 약화

24시즌 용병 타자 WAR 및 수상



*규정 타석 소화한 용병만 포함

NC 다이노스 강점과 약점 분석



강점

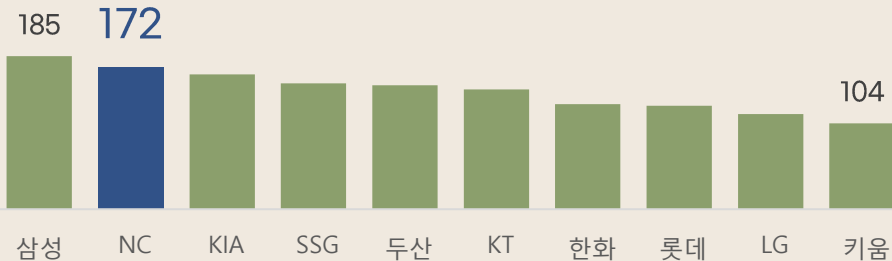
외인 성공률 높고 **도루 저지율 1위**
홈런도 잘치는 NC



외인 골든글러브

24' 하트(투수)
23' 페디(투수)
16' 테임즈(1루수)
15' 해커(투수), 테임즈(1루수)

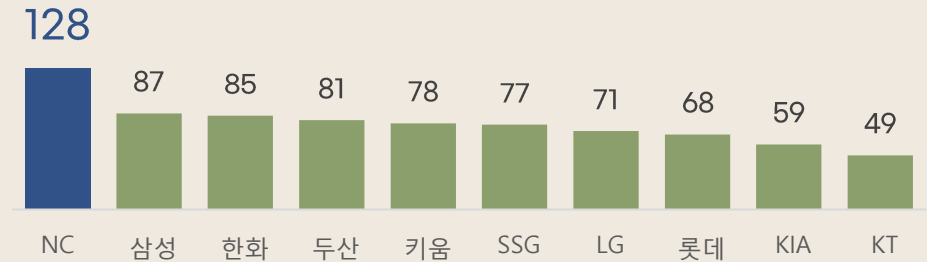
24시즌 팀 홈런



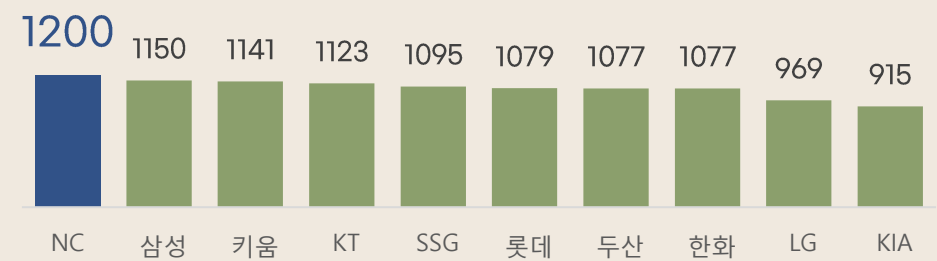
약점

삼진 많은 타선 + 몸에 맞는 공 1등
누적된 충격이 부상으로 이어짐

24시즌 몸에 맞는 공



24시즌 팀 삼진



10. 키움 히어로즈

III. 구단 분석

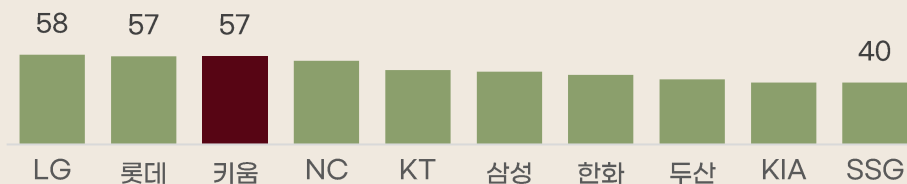
키움 히어로즈 강점과 약점 분석



강점

리빌딩 시즌임에도 불구하고
준수했던 선발, *QS, *QS+ 리그 2위

24시즌 QS 수



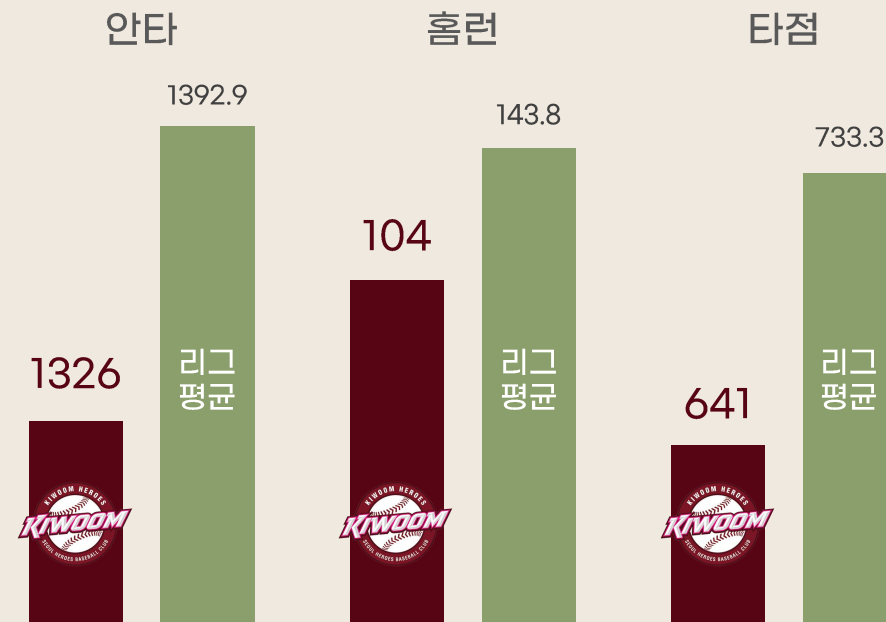
24시즌 QS+ 수



*QS: 60이닝 3자책점 이하 기록 | *QS+: 70이닝 3자책점 이하 기록

약점

실망스러운 타격 지표
안타, 홈런, 타점 모두 최하위



IV. 결론

1. 구단 별 25' 대비 방법

1. 구단 별 25' 시즌 대비 방법

IV. 결론

기아 타이거즈, 삼성 라이온즈, LG 트윈스, 두산 베어스, KT 위즈의 2024 시즌 문제점과 2025 시즌 개선 방향을 도출함.



1. 실책이 많고 병살이 잦아 공/수 부분 흐름이 끊김
2. 부상 관리만 잘하면 올해도 좋은 시즌 보낼 것이라 예상



수비 강화
+
부상 관리



1. 빅볼 스타일 보완을 위해 단타 출루도 섞는 편이 좋아 보임
2. 주전 내야수 부상 시 대체자 부족



밸런스 있는 빅볼 야구
+
내야 뎁스 강화



1. 작전 야구를 하기에 도루 실패가 많아 효율이 매우 낮음
2. 불펜 강팀에서 몇년간 혹사로 인해 불펜이 매우 약화됨



스몰볼 최소화
+
불펜 강화



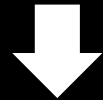
1. 베테랑 타자들의 부진이 매우 아쉬움
2. 불펜 혹사로 인해 올해 불펜 핵심 이탈이 관건. 관리가 시급



베테랑 부진 탈출
+
불펜 관리



1. 내야 포지션 고령화로 체력 부담
2. 시즌 초 선발 투수들의 부상이 성적에 큰 영향



내야 뎁스 강화
+
부상 관리

1. 구단 별 25' 시즌 대비 방법

IV. 결론

SSG 랜더스, 롯데 자이언츠, 한화 이글스, NC 다이노스, 키움 히어로즈의 2024 시즌 문제점과 2025 시즌 개선 방향을 도출함.



1. 이승용 감독의 현재의 결과에 집중인지
신인 육성으로 미래의
집중인지 운영 방향성
정립이 시급

2. 구장 친화적인 빅볼
야구 부활 필요



운영 방향성 정립
+
빅볼 야구

1. 야수진에 비해 리필
딩이 대체로 안 된 투
수진

2. 포수, 유격수 자리
에 대한 아쉬움



포수, 유격수 보완
+
불펜 보완

1. 계속되는 외인 성적
부진

2. 좋은 투수진을 갖고
있는 팀이기에 수비에
서 실수를 줄여주면
질 경기도 이길 것이
라고 생각



좋은 용병
+
수비 강화

1. 손아섭, 박건우, 신
민혁 등 부상 자원이
많아서 아쉬움

2. 확실한 토종 선발이
부족하다고 생각 구창
모가 복귀하면 좀 더
나은 상황이 올 거라
생각



부상 관리
+
선발 야구

1. 김혜성 이탈로 다시
덱스 강화

2. 선수 육성은 잘하는
팀이기에 유망주들의
성장세가 팀 성적의
큰 역할



덱스 강화
+
선수 육성

감사합니다.