



Bros

FINALL

PROJECT



목차

01 팀 소개

02 기획 및 설계

03 프로젝트 소개

04 데이터 분석

05 시연

06 QnA





01

팀 소개





팀 소개



Team Leader

임성래

각종 문서 작성

데이터 수집/가공

데이터 분석

순위 예측 ML모델 생성



Crew

강승원

데이터 수집

프론트 개발

백엔드 개발

UI,DB설계





02

기획 및 설계





추진 배경

교육 받은 기술을 활용한 프로젝트 구현

흥미 있는 주제

기록, 확률, 통계의 스포츠 => 야구!

예측 사이트의 부재





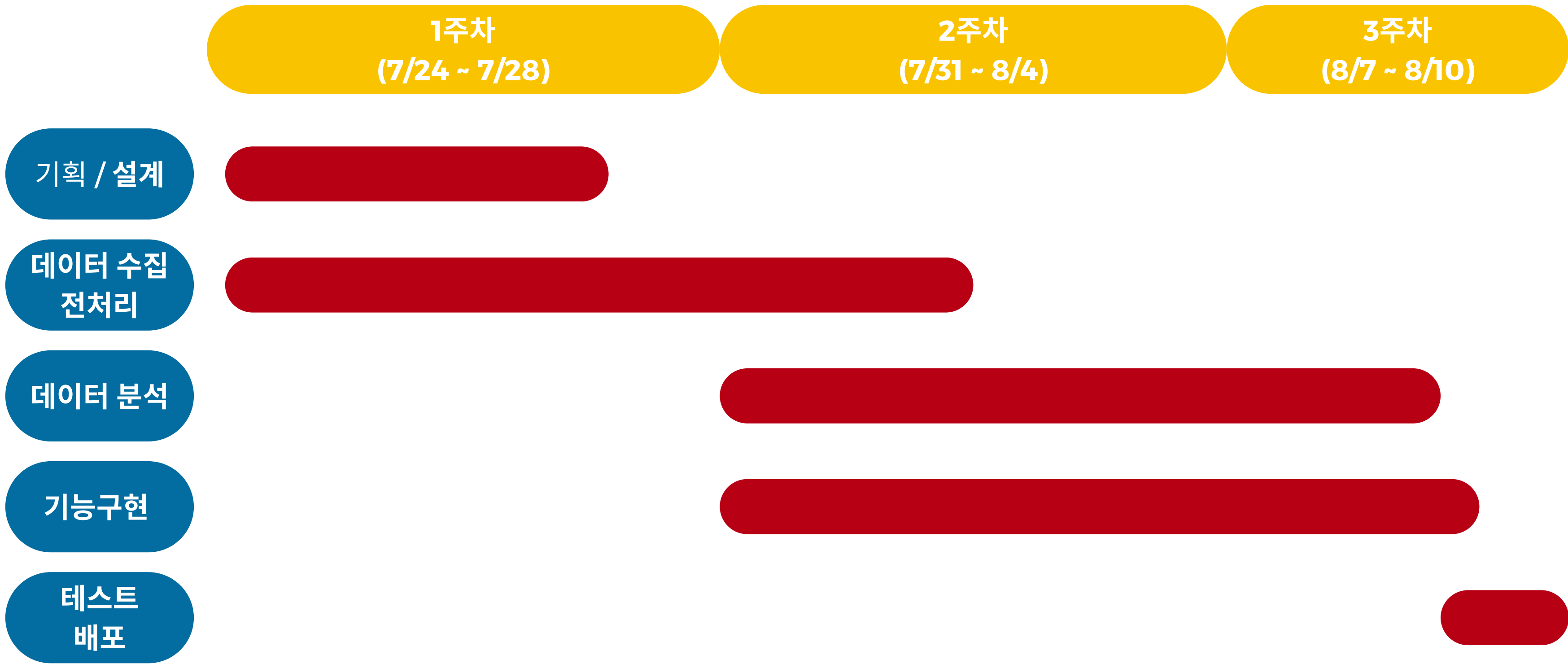
프로젝트 개요

- **기간 : 2023.07.24~2023.08.10**
- **목표: 야구 팬들을 위한 팀별 현황, 순위 예측 서비스 제공**





프로젝트 일정





DB 설계

투수정보

팀 정보

타자정보

투수정보 CopyOfCopyOfUntitled			
선수id	id	int	NOT NULL
team_id	team_id	int	NOT NULL
이름	name	varchar	NOT NULL
승리기여도	war	float	NOT NULL
안타	hits	float	NOT NULL
홈런	home_runs	int	NOT NULL
타율	batting_average	float	NOT NULL
OPS	on_base_plus_slugging	float	NOT NULL
출장수	games_played	int	NOT NULL
포지션	position	varchar	NOT NULL

팀정보 Untitled			
team_id	id	int	NOT NULL
팀이름	team	varchar	NOT NULL
창설연도	era	float	NOT NULL
평균타율	fielding_Percentage	float	NOT NULL
과중투율	whip	float	NOT NULL
타율	batting_average	float	NOT NULL
출루율	on_base_percentage	float	NOT NULL
OPS	on_base_plus_slugging	float	NOT NULL
장타율	slugging_percentage	float	NOT NULL
시즌	year	int	NOT NULL
순위	rank	int	NOT NULL
승률	wln_percentage	float	NOT NULL

타자정보 CopyOfUntitled			
선수id	id	int	NOT NULL
team_id	team_id	int	NOT NULL
이름	name	varchar	NOT NULL
승리기여도	war	float	NOT NULL
안타	hits	float	NOT NULL
홈런	home_runs	int	NOT NULL
타율	batting_average	float	NOT NULL
OPS	on_base_plus_slugging	float	NOT NULL
출장수	games_played	int	NOT NULL
포지션	position	varchar	NOT NULL



03

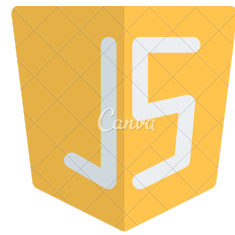
프로젝트 소개



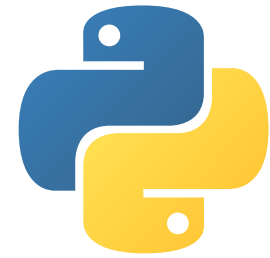


사용 기술 스택

Frontend



Backend



django

Data Analysis



Tool



DB



Server





프로젝트 소개

1. KBO 팀들의 현재 순위와 예측 순위 비교

<메인 페이지>

bros Community xxx xxx xxx admin			
팀 순위			
순위	팀	승률	업데이트 : 2023.08.08 18:08
1	LG	0.598	팀정보 선수정보
2	SSG	0.593	팀정보 선수정보
3	두산	0.543	팀정보 선수정보
4	NC	0.519	팀정보 선수정보
5	KT	0.494	팀정보 선수정보
6	롯데	0.488	팀정보 선수정보
7	KIA	0.474	팀정보 선수정보
8	한화	0.462	팀정보 선수정보
9	키움	0.449	팀정보 선수정보
10	삼성	0.388	팀정보 선수정보
예상 팀 순위			
순위	팀	업데이트 : 2023.08.08 18:08	
1	LG	팀정보	선수정보
2	NC	팀정보	선수정보
3	SSG	팀정보	선수정보
4	두산	팀정보	선수정보
5	KIA	팀정보	선수정보
6	키움	팀정보	선수정보
7	KT	팀정보	선수정보
8	한화	팀정보	선수정보
9	삼성	팀정보	선수정보
10	롯데	팀정보	선수정보



프로젝트 소개

2. 관심 있는 팀의 정보,예측 순위 구현

<팀 정보 페이지>

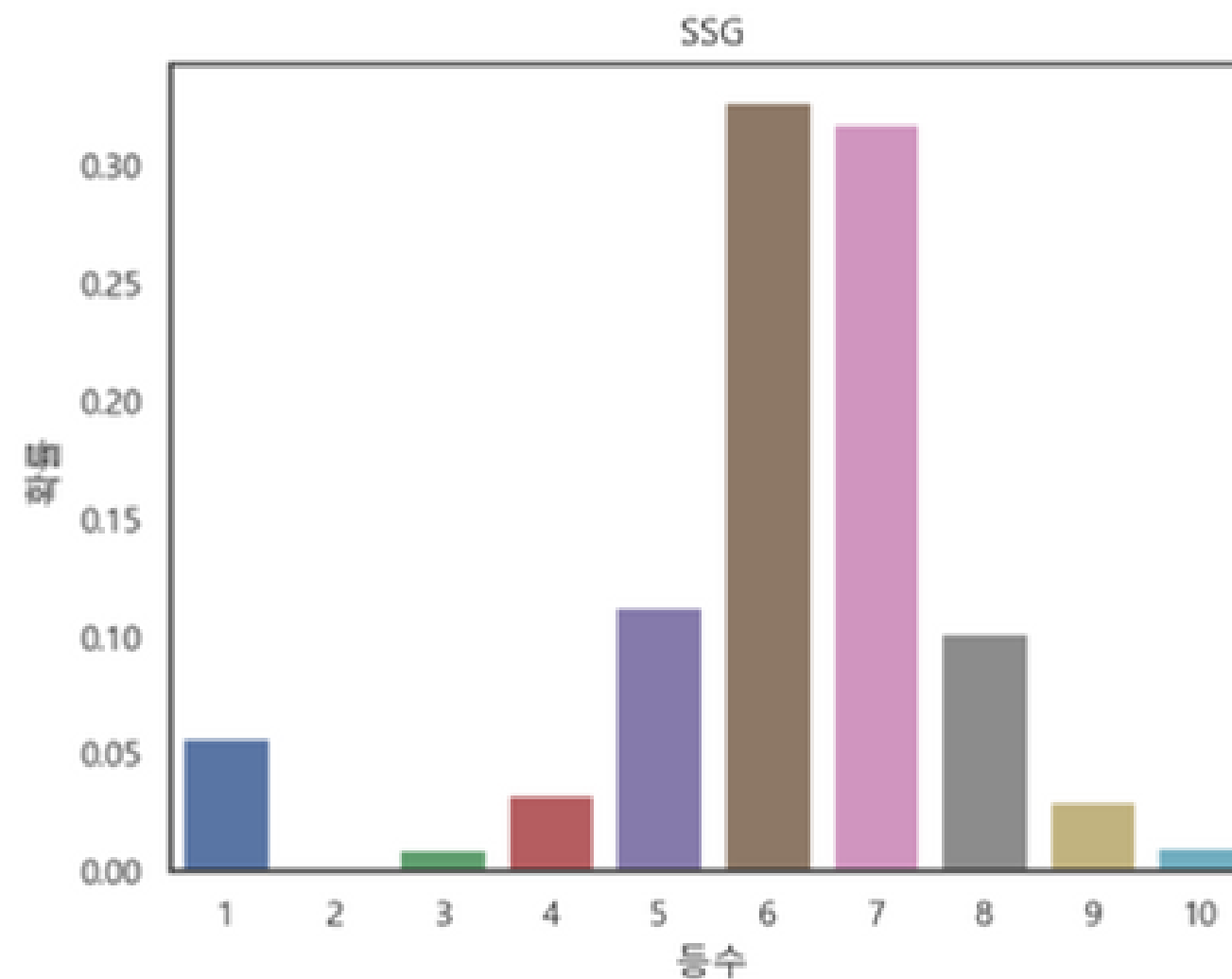
SSG 선수정보

SSG

시즌: 23
순위: 2
타율: 0.259
출루율: 0.336
장타율: 0.396
OPS: 0.732
수비율: 0.979
평균 자책점: 4.06
이닝당 출루율: 1.5

SSG

1위 확률: 0.057
2위 확률: 0.0
3위 확률: 0.009
4위 확률: 0.033
5위 확률: 0.113
6위 확률: 0.327
7위 확률: 0.318
8위 확률: 0.102
9위 확률: 0.03
10위 확률: 0.01





프로젝트 소개

3. 관심 있는 팀의 현재 투수,타자 기록

<선수 정보 페이지>

타자 정보

이름	타율	팀 소속	안타	홈런	OPS	경기출전	승리기여도	포지션
최정	0.311	SSG	83	19	0.973	73	3.81	타자
에레디아	0.339	SSG	100	10	0.895	76	2.91	타자
박성한	0.267	SSG	73	5	0.72	77	2.0	타자
최지훈	0.269	SSG	75	2	0.685	67	1.45	타자
추신수	0.243	SSG	45	7	0.783	55	1.15	타자
최주환	0.238	SSG	58	14	0.771	73	1.15	타자
김민식	0.236	SSG	38	3	0.678	71	0.87	타자
김강민	0.321	SSG	17	2	0.865	27	0.65	타자
하재훈	0.342	SSG	13	2	1.024	14	0.64	타자
강진성	0.288	SSG	21	1	0.796	31	0.55	타자
김성현	0.278	SSG	30	0	0.685	47	0.42	타자
안상현	0.314	SSG	11	0	0.8	31	0.38	타자
최준우	0.317	SSG	13	0	0.719	20	0.21	타자
오태곤	0.226	SSG	35	4	0.64	75	0.12	타자
김정민	1.0	SSG	1	0	2.0	6	0.05	타자
조형우	0.167	SSG	9	1	0.515	32	-0.08	타자
김건웅	0.0	SSG	0	0	0.0	1	-0.08	타자
김찬형	0.188	SSG	3	0	0.423	11	-0.08	타자
고명준	0.0	SSG	0	0	0.0	2	-0.1	타자
최형	0.167	SSG	2	1	0.647	9	-0.11	타자

투수 정보

이름	평균 자책점	팀 소속	승	패	세이브	홀드	경기출전	이닝	승리기여도	포지션
서진용	1.21	SSG	1	0	24	0	37	37.1	2.37	투수
맥카티	2.52	SSG	7	3	0	0	13	75.0	1.91	투수
김광현	3.67	SSG	5	2	0	0	14	76.0	1.45	투수
고효준	2.56	SSG	3	1	0	7	38	31.2	1.07	투수
최민준	3.09	SSG	4	2	1	6	29	35.0	0.95	투수
노경은	3.95	SSG	5	3	2	18	39	41.0	0.8	투수
문승원	4.25	SSG	2	3	1	5	27	48.2	0.65	투수
열리아스	4.21	SSG	4	3	0	0	8	47.0	0.57	투수
박민호	0.9	SSG	0	0	0	0	10	10.0	0.27	투수
임준섭	3.86	SSG	0	0	0	2	26	23.1	0.25	투수
오원석	4.89	SSG	5	6	0	0	16	84.2	0.24	투수
신현민	2.16	SSG	0	0	0	0	7	8.1	0.21	투수
이건욱	3.6	SSG	0	0	0	0	8	15.0	0.21	투수
백승건	4.66	SSG	2	1	0	3	24	36.2	0.11	투수
한두솔	9.0	SSG	0	0	0	0	1	1.0	-0.01	투수
이기순	13.5	SSG	0	0	0	0	1	2.0	-0.06	투수
조성훈	6.43	SSG	0	1	0	0	2	7.0	-0.08	투수
정성곤	5.4	SSG	0	0	0	0	2	1.2	-0.15	투수
이로유후	6.21	SSG	3	0	0	4	25	33.1	-0.22	투수



프로젝트 소개

4. 게시물, 댓글, 좋아요 기능 구현

<커뮤니티 페이지>

bros Community xxx xxx xxx admin ▾

두산 파이팅 !!! 오늘 꼭 이겨주세요

ADMIN

조회수 2

4초 전

delete

Edit Post

두산

오늘은 홈런 빵빵 ~_~

♥ 1

🏷 홈런 두산



프로젝트 소개

5. 게시물 검색, 카테고리, 인기글 등 편의기능 구현

<커뮤니티 페이지>

[Go!](#)

Categories

한화 (4)
두산 (2)
롯데 (0)
KT (0)
LG (0)
NC (0)
SSG (0)
삼성 (0)
키움 (0)
KIA (0)
미분류 0

조회수 많은 글

잉이잉이잉이잉이...	조회 251	댓글 4
ㅂㅈㄷㄱ	조회 144	댓글 4
23	조회 133	댓글 0
ㅂㅈㄷㄱ	조회 132	댓글 1



04

데이터 분석





메인 화면에 구상할 팀 기본 데이터 수집 (팀명, 현재 순위, 승률 데이터)

	팀 순위									
순	팀	G	승	패	무	승차	승률	득점	실점	
1	LG	81	49	30	2	-	0.620	437	348	
2	SSG	79	46	32	1	2.5	0.590	371	355	
3	두산	79	42	36	1	6.5	0.539	333	342	
4	NC	78	39	38	1	9.0	0.507	367	326	
5	롯데	77	38	39	0	10.0	0.494	329	368	
6	KIA	76	36	39	1	11.0	0.480	341	319	
7	KT	80	37	41	2	11.5	0.474	364	379	
8	한화	78	34	40	4	12.5	0.460	325	334	
9	기움	86	38	46	2	13.5	0.452	342	361	
10	삼성	80	31	49	0	18.5	0.388	325	402	



데이터 수집

2. 선수 데이터

약 2,300개의 타자 데이터 수집

(이름, WAR, 안타, 홈런 등 타자 데이터)

약 3,100개의 투수 데이터 수집

(이름, WAR, 승, 패 등 투수 데이터)

기본 확장 가치 클러치 타석 타구1 타구2 파워 팀배팅1													
순위	이름	팀	정렬 WAR*	G	타석	타수	득점	안타	2타	3타	홈런	루타	E
1	노시환	23 한 3B	4.53	78	354	309	51	98	18	0	19	173	
2	김혜성	23 키 2B	3.87	85	375	339	64	110	19	3	5	150	
3	양의지	23 두 C	3.82	75	299	248	31	83	16	0	8	123	
4	최정	23 S 3B	3.81	73	314	267	64	83	14	0	19	154	
5	이정후	23 키 CF	3.57	83	378	324	49	101	23	2	6	146	
6	박동원	23 L C	3.31	77	303	254	38	69	13	1	15	129	
7	홍창기	23 L RF	3.13	80	373	301	59	100	25	2	0	129	
8	예레디아	23 S LF	2.91	76	324	295	43	100	17	0	10	147	
9	강민호	23 삼 C	2.79	70	284	251	35	77	11	0	11	121	

선수별 기록



데이터 수집

3. 팀 세부 데이터

훈련 데이터 수집
2020년도~2022년 팀 데이터
(팀타격,수비, 투수 데이터)

테스트 데이터 수집
2023년 현재까지 팀 데이터
(팀타격,수비, 투수 데이터)

	SEASON	TEAM	AVG	OBP	SLG	ERA	WHIP	GRA
1	20	LG	0.277	0.349	0.428	4.390	1.420	3.0
2	20	KT	0.284	0.358	0.436	4.560	1.460	2.0
3	20	키움	0.269	0.355	0.408	4.390	1.390	5.0
4	20	NC	0.291	0.366	0.462	4.590	1.420	1.0
5	20	삼성	0.268	0.338	0.394	4.780	1.470	8.0
6	20	SSG	0.250	0.329	0.383	5.600	1.600	9.0
7	20	롯데	0.276	0.354	0.408	4.640	1.410	7.0
8	20	두산	0.293	0.365	0.427	4.400	1.460	3.0

팀 데이터



데이터 전처리

X 값 전처리

리그 평균 데이터, 팀 데이터를 정리하고,
팀 통계 데이터를 정규화하여
새로운 데이터프레임을 생성

```
4 # X 전처리 함수
5 def preprocess_X(X):
6     db = defaultdict(list)
7     n_db = defaultdict(list)
8     for item in X.itertuples():
9         if item[2] == 'LEAGUE':
10             db[item[1]] = [item[3], item[4], item[5], item[6], item[7]]
11         else:
12             n_db['AVG'].append(item[3]/db[item[1]][0])
13             n_db['OBP'].append(item[4]/db[item[1]][1])
14             n_db['SLG'].append(item[5]/db[item[1]][2])
15             n_db['ERA'].append(item[6]/db[item[1]][3])
16             n_db['WHIP'].append(item[7]/db[item[1]][4])
17     return pd.DataFrame(n_db)
```

Y 값 전처리

결측값을 처리하여 데이터의 무결성을 유지

```
# y 전처리 함수
def preprocess_y(y):
    y.dropna(inplace=True)

    return y
```

전처리 코드



데이터 분석

데이터 스케일링 작업

표준화 된 X_{train} ,

평균 값 mean ,

그리고 사용된 표준편차 값 std 를 반환하여

일관된 방식으로 모델 학습에 사용

```
5 # 스케일링 진행해주는 함수
6 def standard_X(X, **kwargs):
7     mean, std = 0, 0
8     if not kwargs:
9         mean = X.mean(axis=0, skipna=False)
10        std = X.std()
11    else:
12        mean = kwargs["mean"]
13        std = kwargs["std"]
14    for col in X.columns:
15        X[col] = (X[col] - mean[col]) / std[col]
16
17    return X, mean, std
```



데이터 분석

모델 학습 과정

최근 3년 팀 데이터

X

팀 타율, 출루율, 장타율, 방어율, 피출루율

Y

순위

신경망 모델인

neural_network 학습 모델 생성

```
import seaborn as sns
import neural_network
import data_preprocessing
import warnings

warnings.filterwarnings(action='ignore')
%config Completer.use_jedi = False
plt.rc("font", family="Malgun Gothic")
sns.set(font="Malgun Gothic", rc={"axes.unicode_minus": False}, style='white')

# np & pd 관련 설정
np.random.seed(42)
np.set_printoptions(precision=2, suppress=True)
pd.options.display.float_format = '{:.3f}'.format

# 훈련 데이터 로드
train_data = pd.read_csv("data/KBO_TRAIN.csv")

# 전처리
# 3년치 각팀, 리그 평균 주요 팀 지표
# 승패를 가르는 기본 지표
X_train, y_train = train_data.drop('GRADE', axis=1), train_data['GRADE']
X_train = data_preprocessing.preprocess_X(X_train)
y_train = data_preprocessing.preprocess_y(y_train)

# 스케일링
```



데이터 분석 및 모델 예측

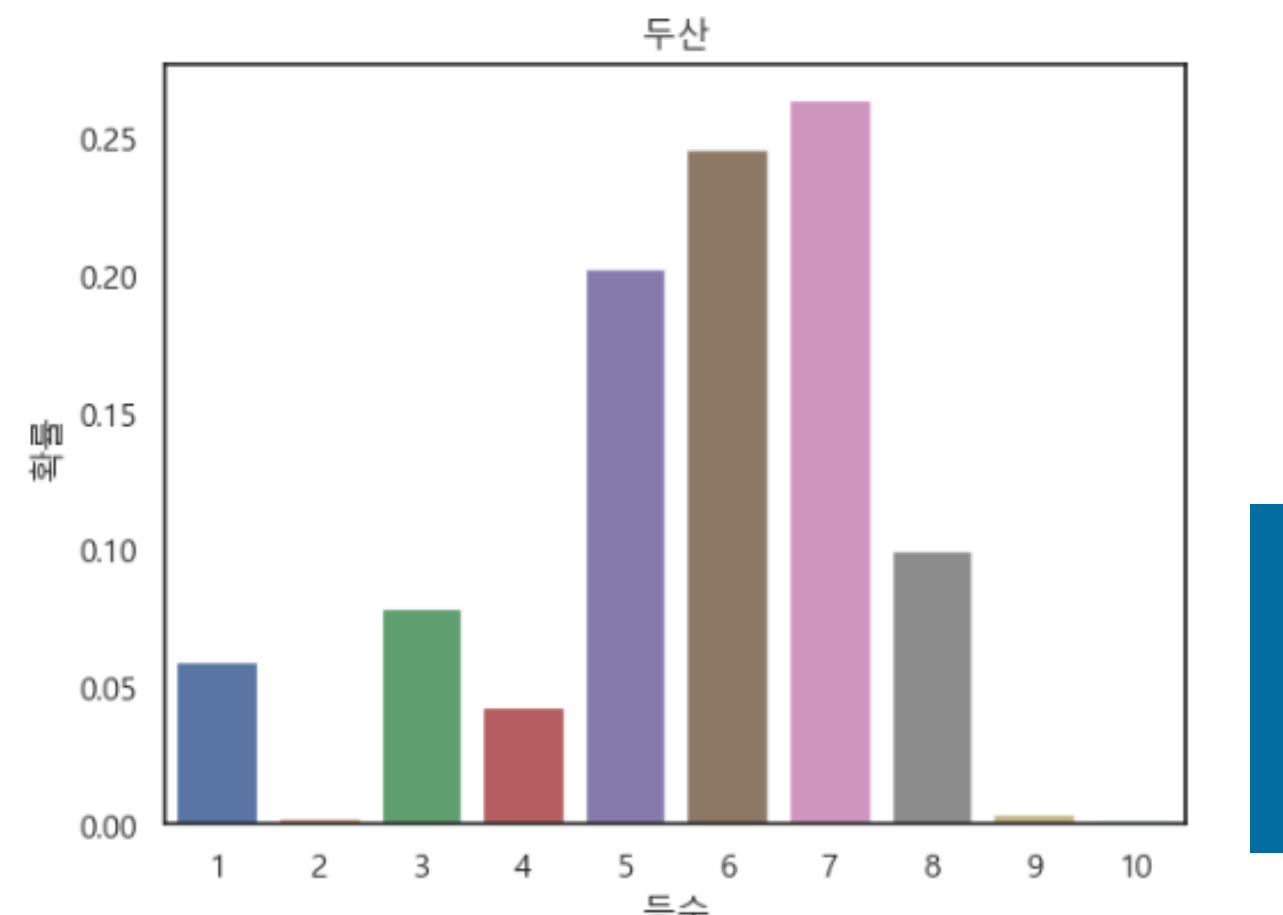
예측 수행

23년 중반까지의 10개팀
X_test 도 같은 방식으로
데이터 INPUT

각 팀별

순위 예측, 확률 구현 완료

	LG	NC	SSG	두산	KIA	키움	KT	한화	삼성	롯데
1	0.947	0.003	0.057	0.059	0.000	0.000	0.001	0.007	0.000	0.001
2	0.000	0.850	0.000	0.002	0.026	0.024	0.000	0.000	0.000	0.000
3	0.010	0.143	0.009	0.079	0.974	0.975	0.001	0.003	0.000	0.000
4	0.042	0.004	0.033	0.043	0.000	0.000	0.001	0.006	0.000	0.001
5	0.001	0.000	0.113	0.202	0.000	0.000	0.030	0.030	0.001	0.005
6	0.000	0.000	0.327	0.246	0.000	0.000	0.266	0.273	0.038	0.148
7	0.000	0.000	0.318	0.264	0.000	0.000	0.232	0.099	0.008	0.040
8	0.000	0.000	0.102	0.100	0.000	0.000	0.465	0.094	0.012	0.067
9	0.000	0.000	0.030	0.004	0.000	0.000	0.004	0.287	0.478	0.449
10	0.000	0.000	0.010	0.001	0.000	0.000	0.001	0.201	0.464	0.290





05

시연





<http://3.35.170.65:8000/>

