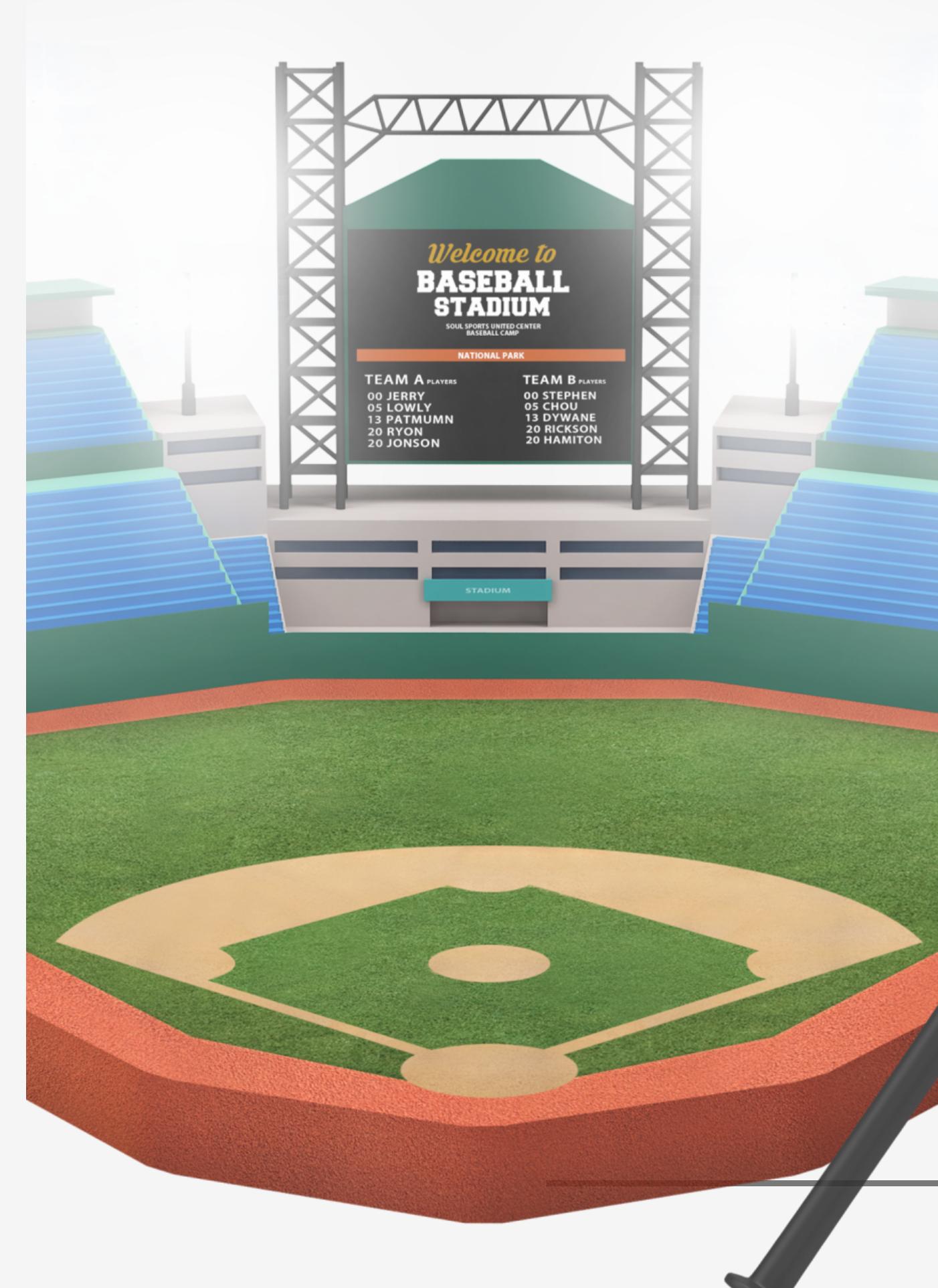


프로야구 관중수 예측

인기팀이 잘하면 관중이 늘어날까?

+



02

목차

01

분석 준비

- 문제 정의
- 데이터 수집

02

기초 통계 산출물

- 연도별 평균 관중수
- 월/요일별 평균 관중수
- 팀순위별 평균 관중수
- 팀별 평균 관중수

03

모델 적합

- 모델 적합시 주요 고려사항

04

분석 결과

- 모델 평가
- 관중수 예측 핵심 변수

01

분석 준비

- 문제 정의
- 데이터 수집

03

문제 정의

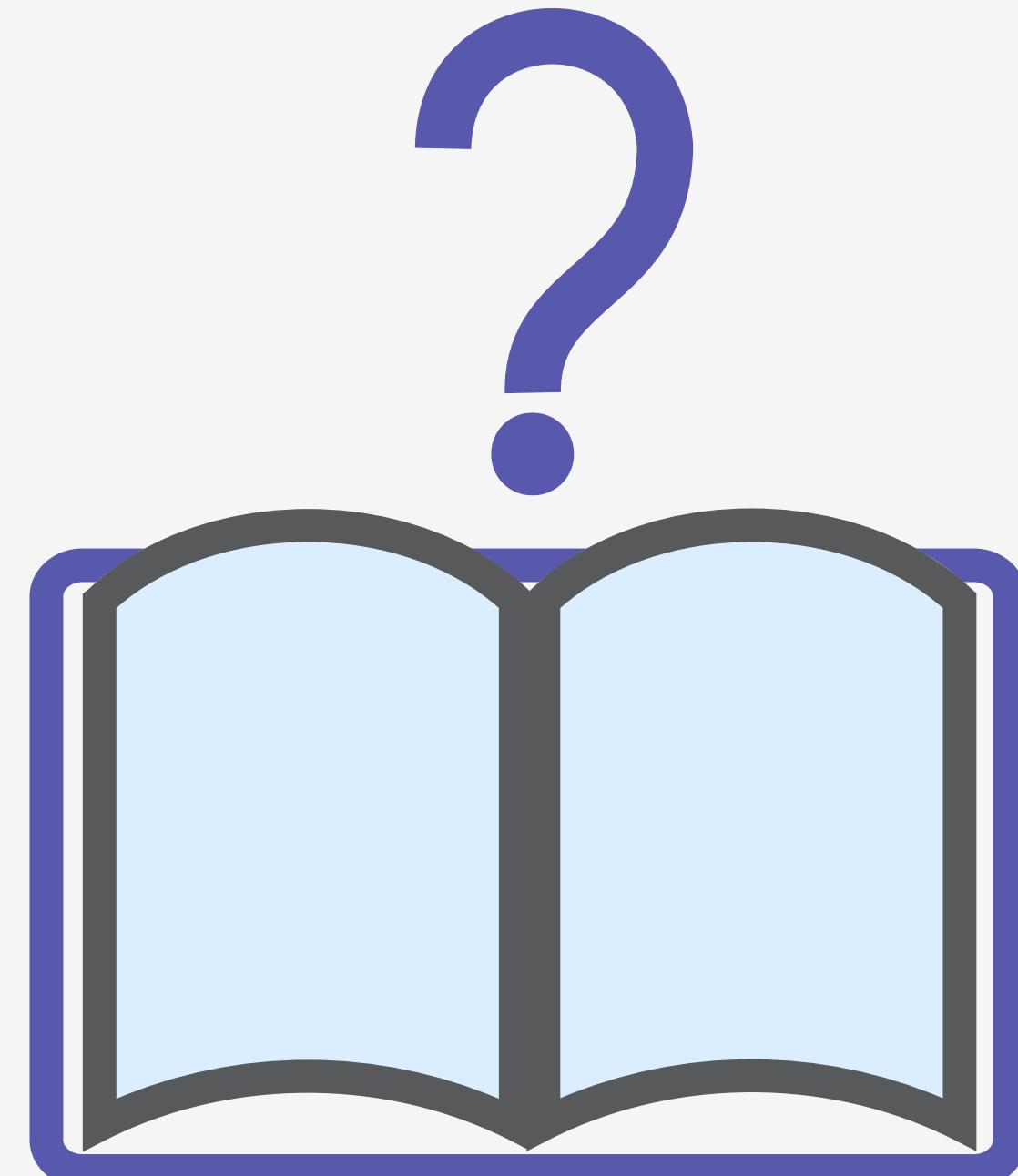
01

프로야구 관중수를 맞추는 좋은 모델 만들기

02

팀 순위, 날씨 등 어느 요인이 프로야구 관중
수를 잘 맞추는지 알아보기.

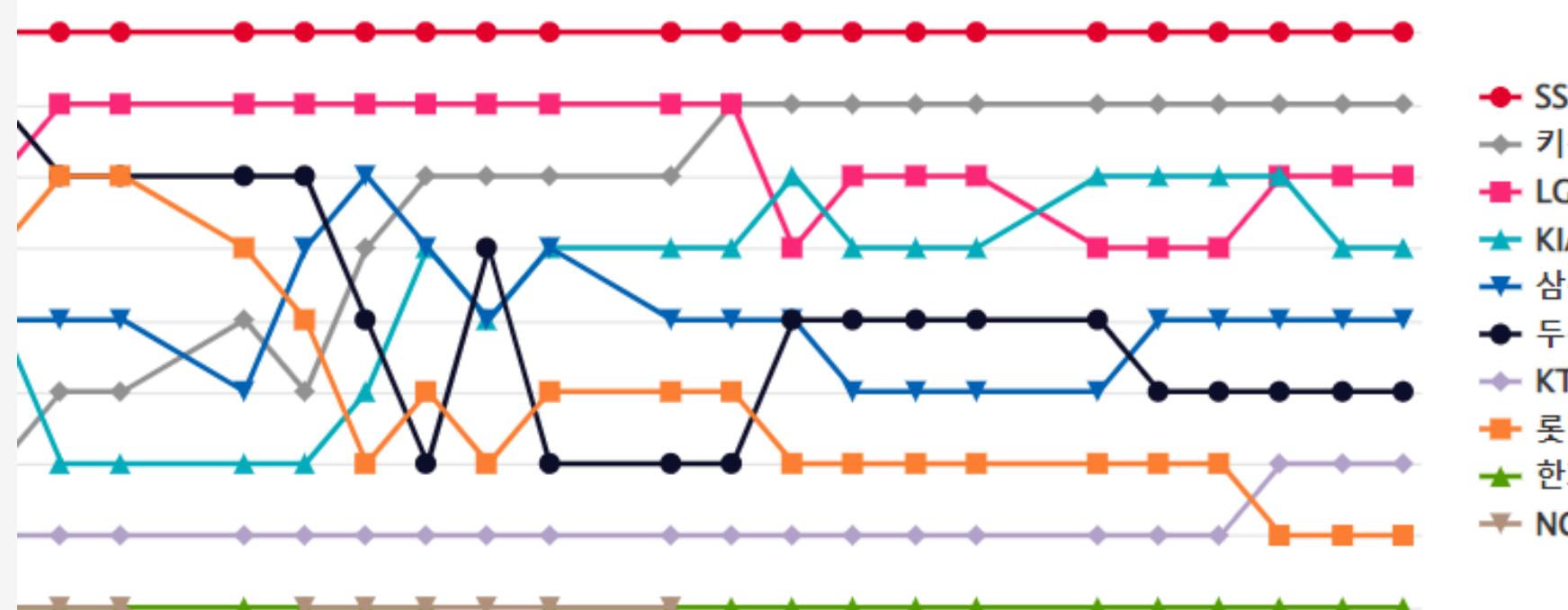
04





일자별 순위 그래프

오른쪽 팀명을 선택하시면 팀별로 그래프를 확인하실 수 있습니다.



데이터 수집

2017~2019년, 2022년 4/5월 데이터
2020, 2021년은 거리두기로 제외

○ KBO 홈페이지

한달 단위 일자별 팀 순위 그래프를 검색해
일일이 엑셀에 기록하여 수집.

○ 프로스포츠 정보광장

경기 날짜, 홈팀, 원정팀, 날씨,
관중 수 데이터 수집.
위에서 구한 팀 순위 데이터와 결합.

02

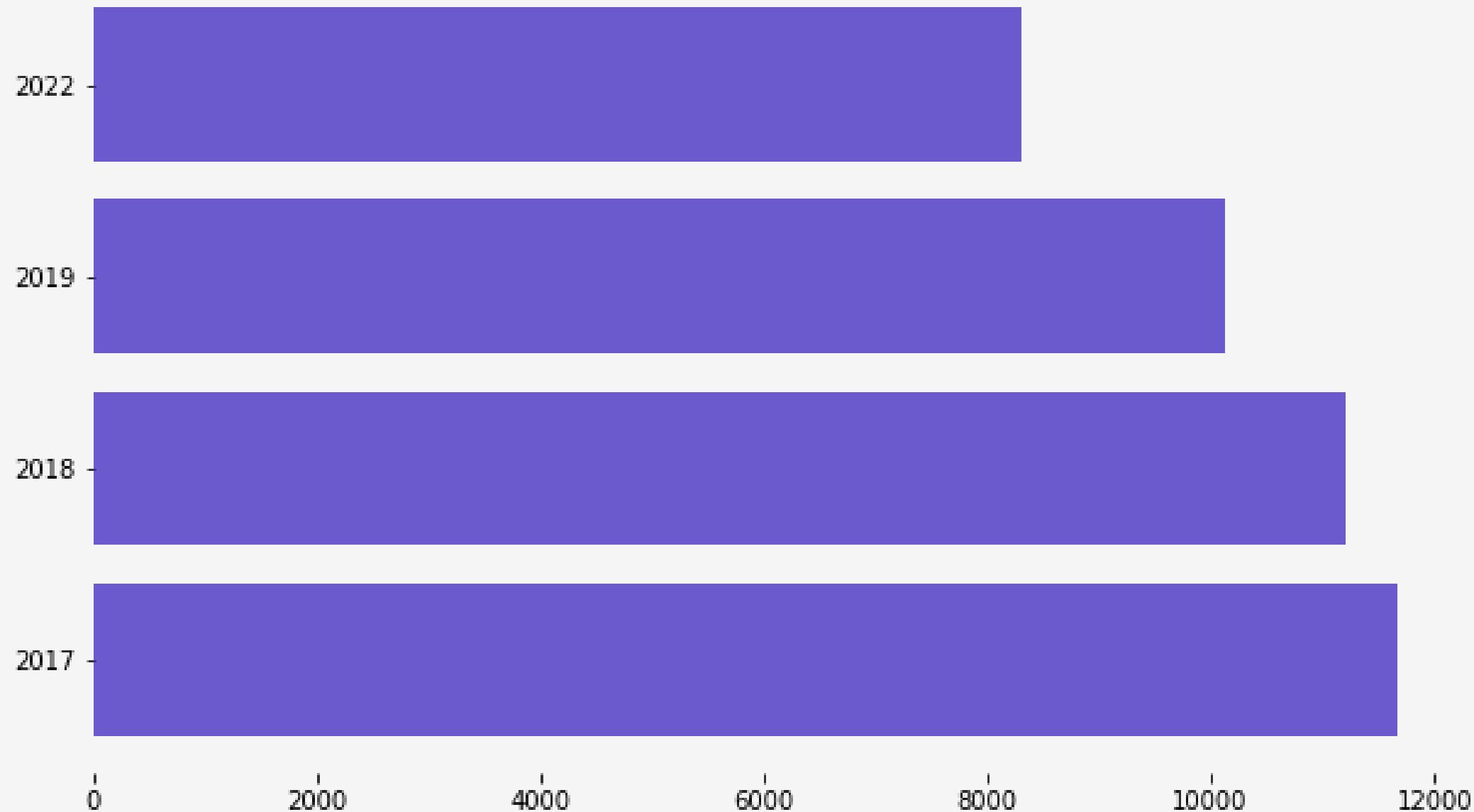
기초 통계 산출물

- 연도별 평균 관중수
- 월/요일별 평균 관중수
- 팀순위별 평균 관중수
- 팀별 평균 관중수

연도별 평균 관중수

2017년 부터 지속적으로
감소하는 모습이 보임.
다만 2022년 거리두기 영
향을 받은 점도 고려해야.

The number of spectators

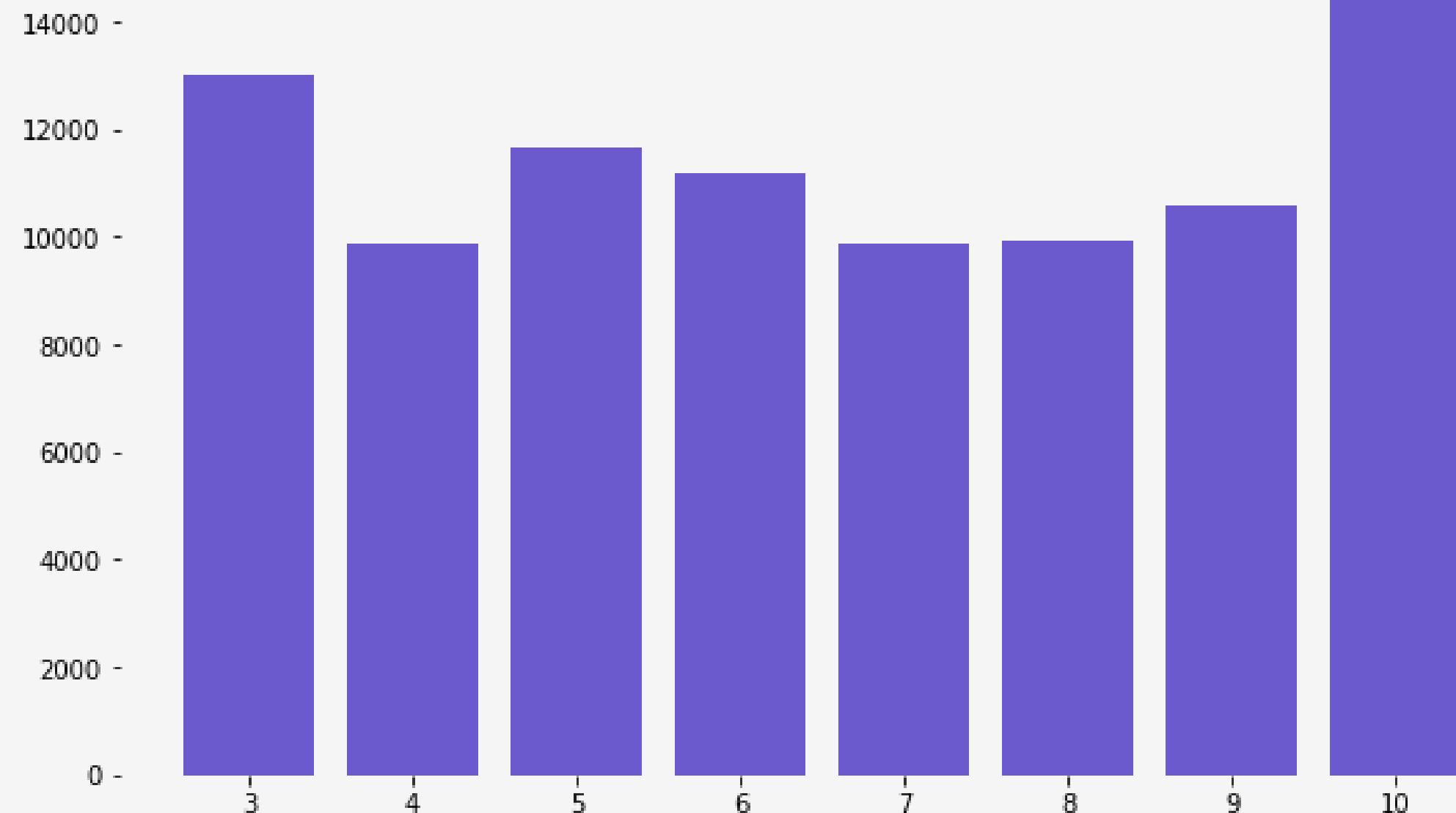


월별 평균 관중수

3월과 10월은 경기수가 적
으며 개막효과 고려.

5,6월에 관중수가 많으며
더워지면 관중수가 감소하
는 것으로 보임.

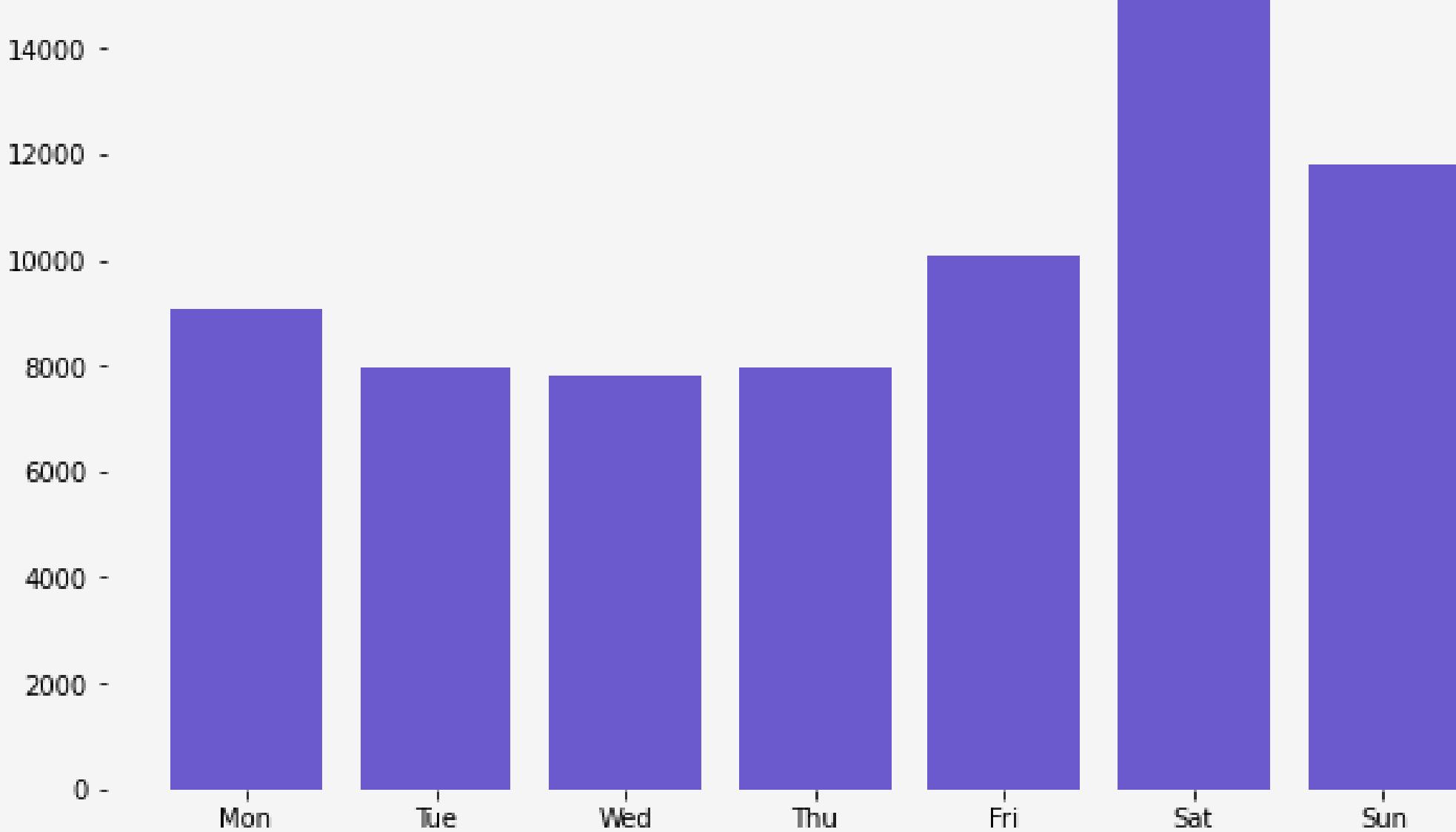
Spectators about month



요일별 평균 관중수

월요일은 대체경기여서 경
기수가 적어 표본이 흔듯.
평일(화/수/목) 관중이
적으며 토-일-금 순서로
관중수가 많음.

Spectators about week

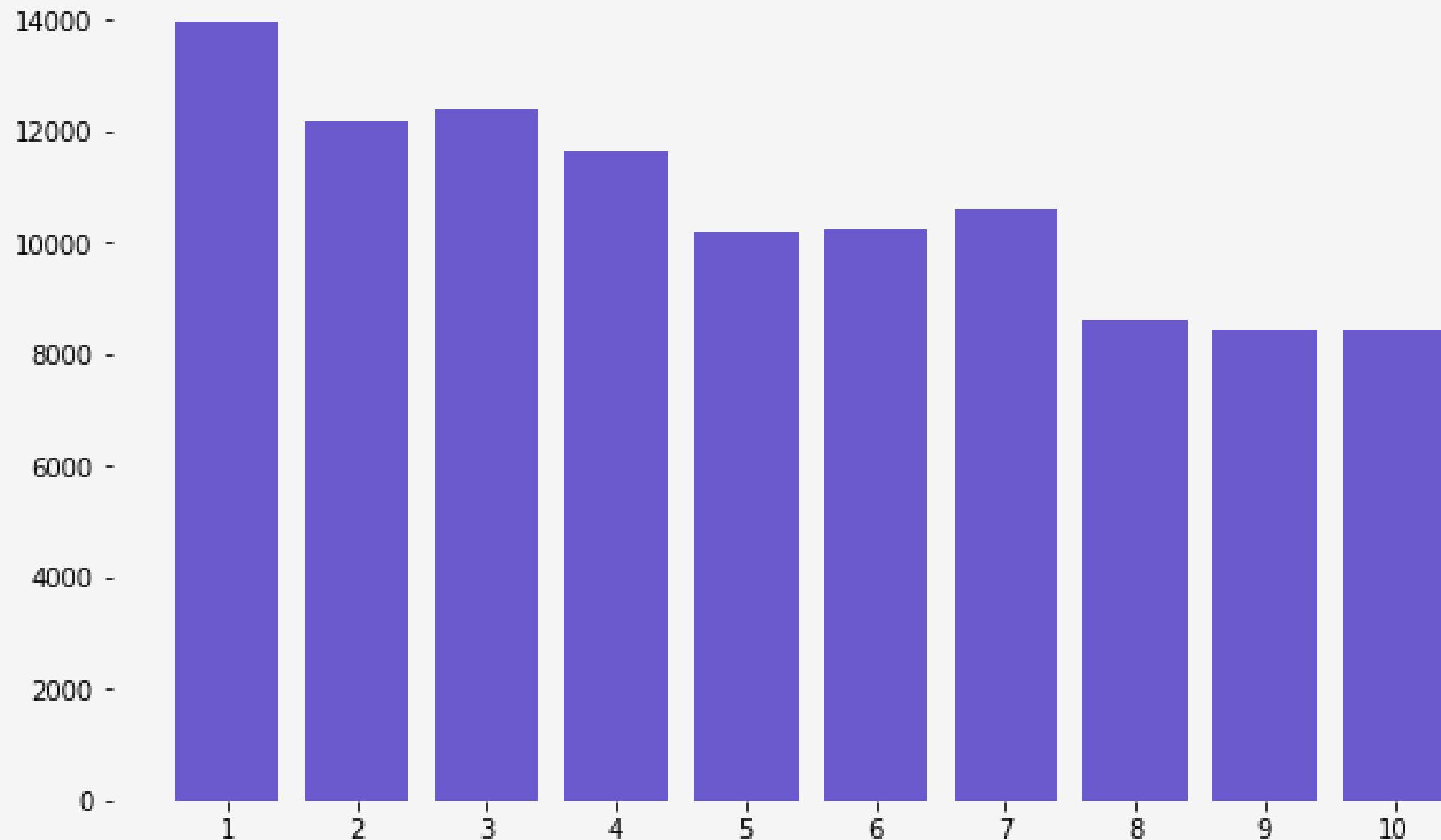


실시간 팀 순위별 평균 관중수

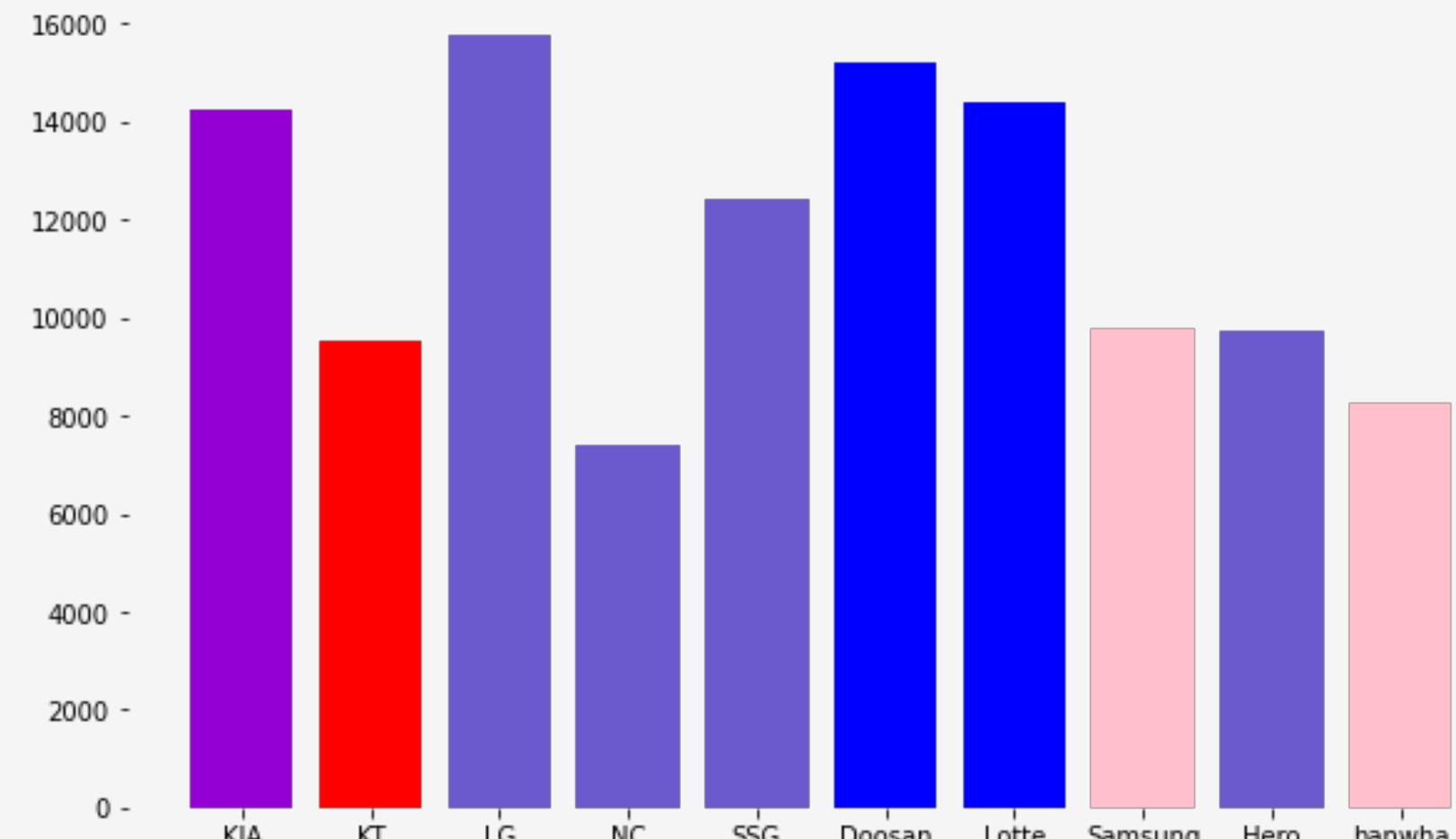
대체적으로 순위와
관중수가 비례하는 모습.
1등, 2~4등, 5~7등, 8~10등
4 그룹으로 분리해서
생각할 수도.

10

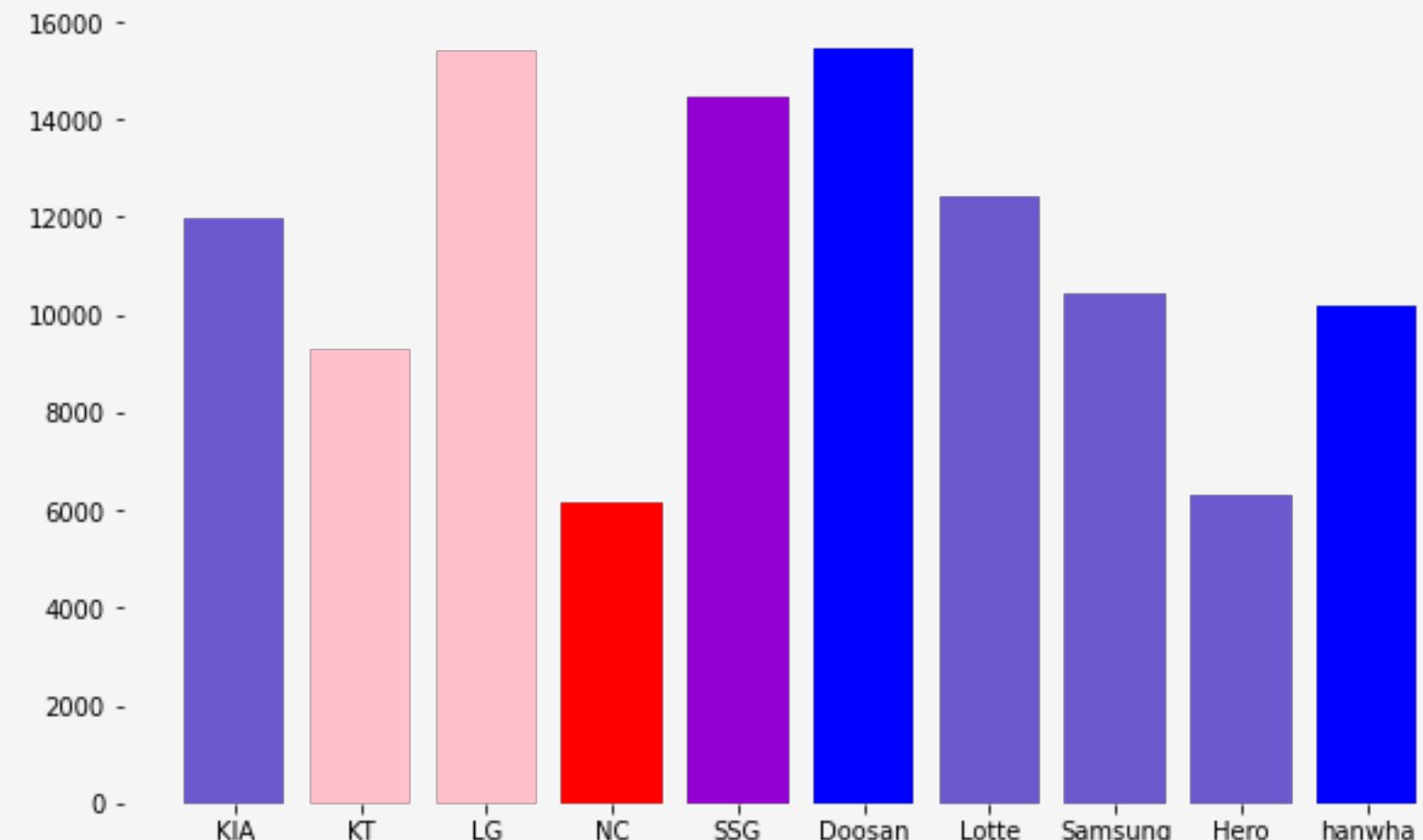
Spectators about team ranking



2017



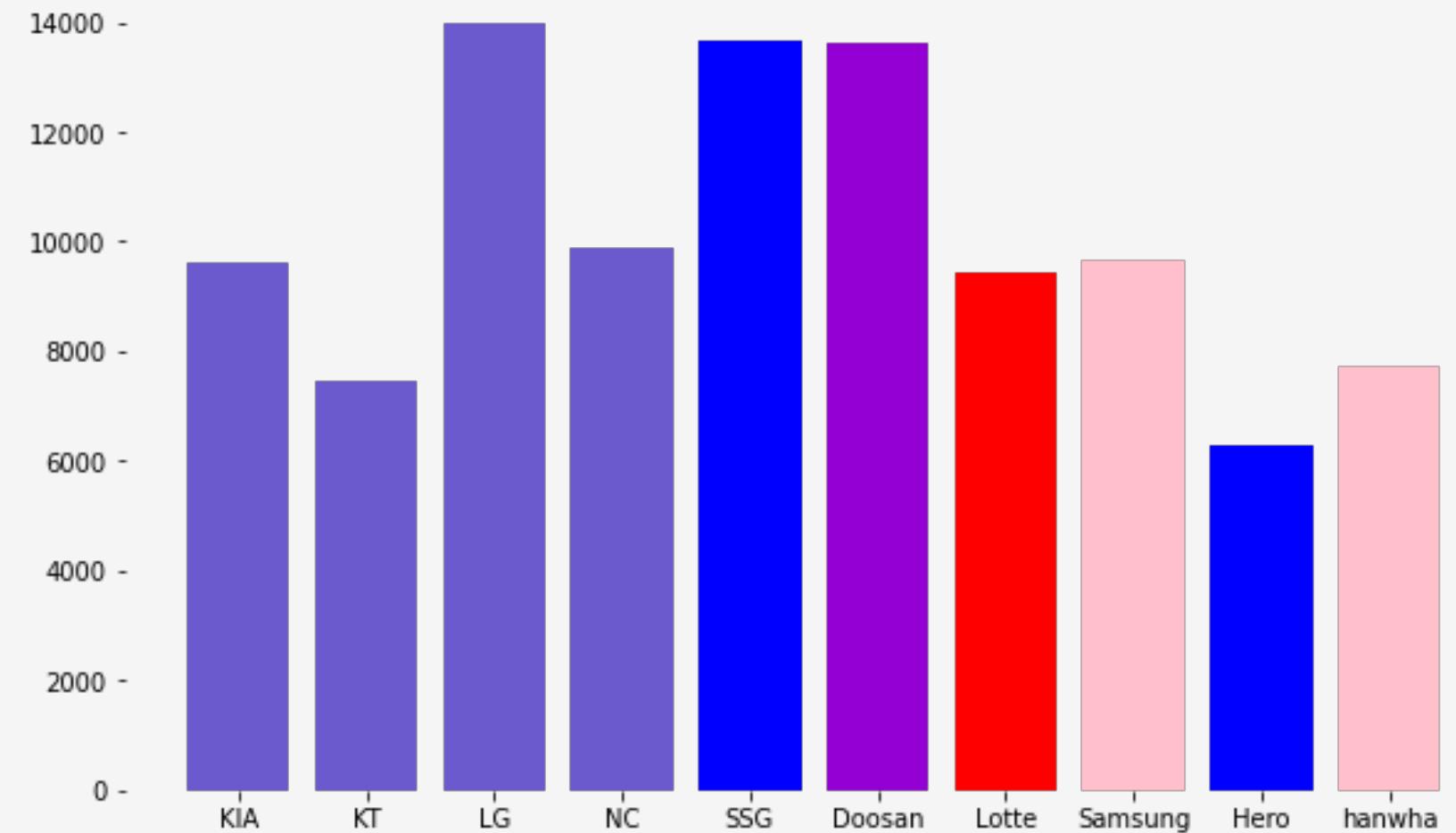
2018



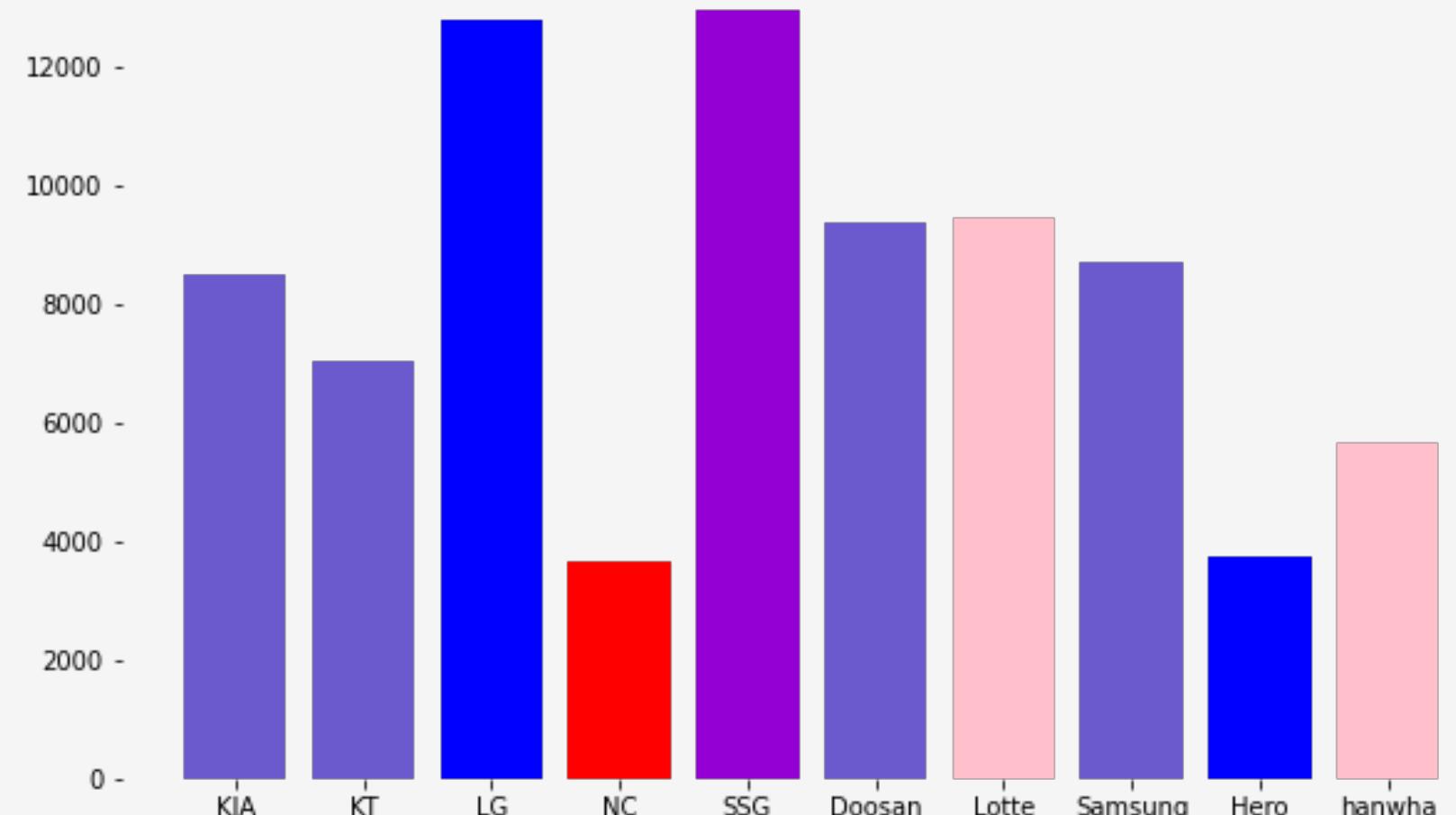
팀별 평균 관중수

보라색이 1등,
파란색이 2~3등,
연한 파란색이 4~7등,
핑크색이 8~9등,
빨간색이 10등.

2019



2022



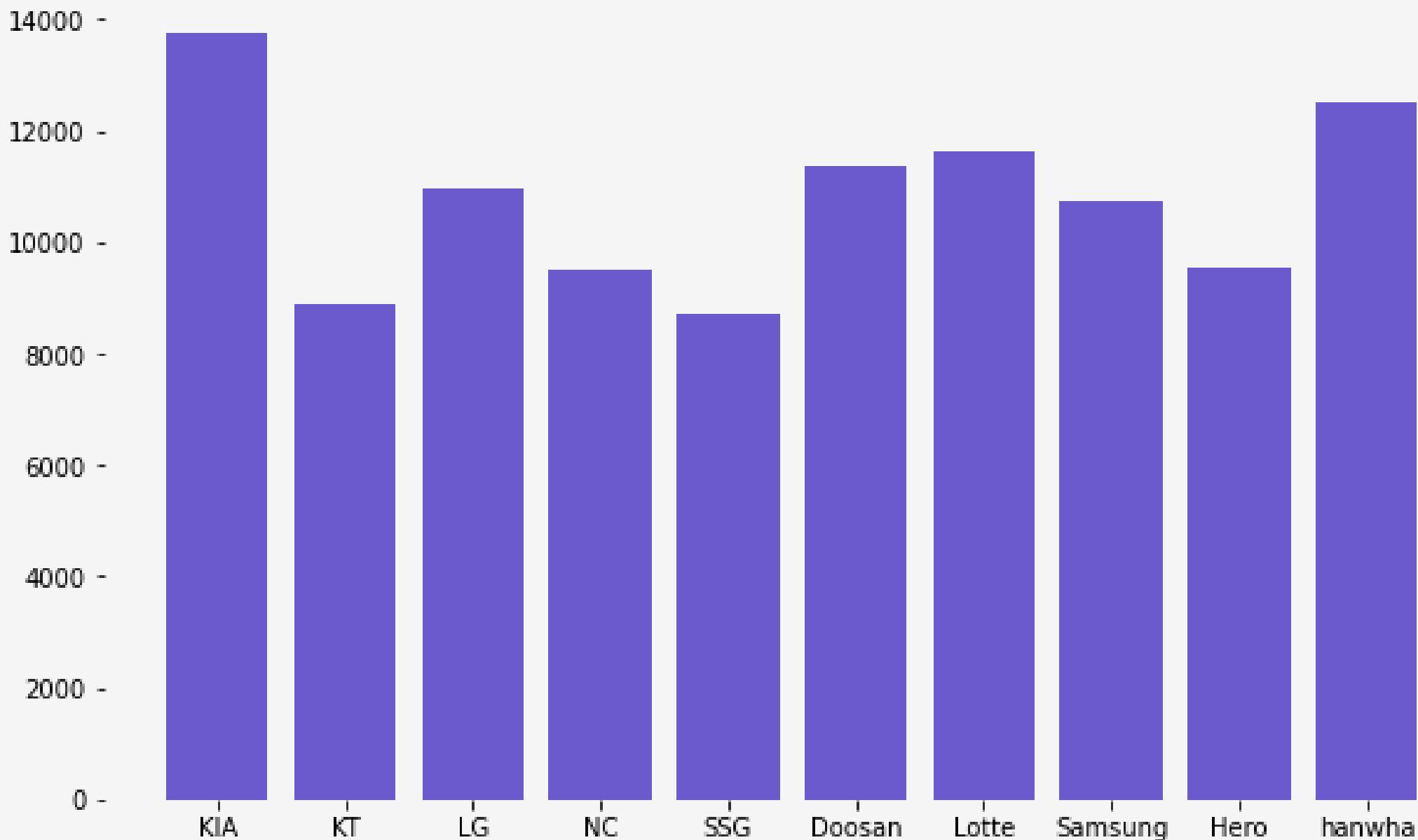
팀별 평균 관중수

대부분 순위가 높아지면
관중수가 늘어나는 모양새.
다만 키움, KT등 비 인기
구단은 순위와 관중수의
연관성이 떨어짐.

원정 팀별 평균 관중수

홈팀 관중수는 수도권 팀이
인기도 대비 큰 값을 가짐.
보다 객관적으로 전국구
인기팀을 판단하는 지표.
기아-한화-롯데-두산-엘지
순으로 집계됨.

Spectators to the visiting team



03

모델 적합

- 모델 적합시 주요 고려사항

모델 적합시 주요 고려사항



- 사용한 설명변수

연도, 월, 요일, 홈팀, 원정팀, 날씨,
평균기온, 홈팀순위, 원정팀순위

- Label Encoding

범주형 변수(요일, 홈팀, 원정팀, 날씨)를
숫자와 맵핑시킴

- Light GBM

가벼우면서 성능이 우수해 많이 쓰이는
Light GBM 회귀 모형 사용.

04

분석 결과

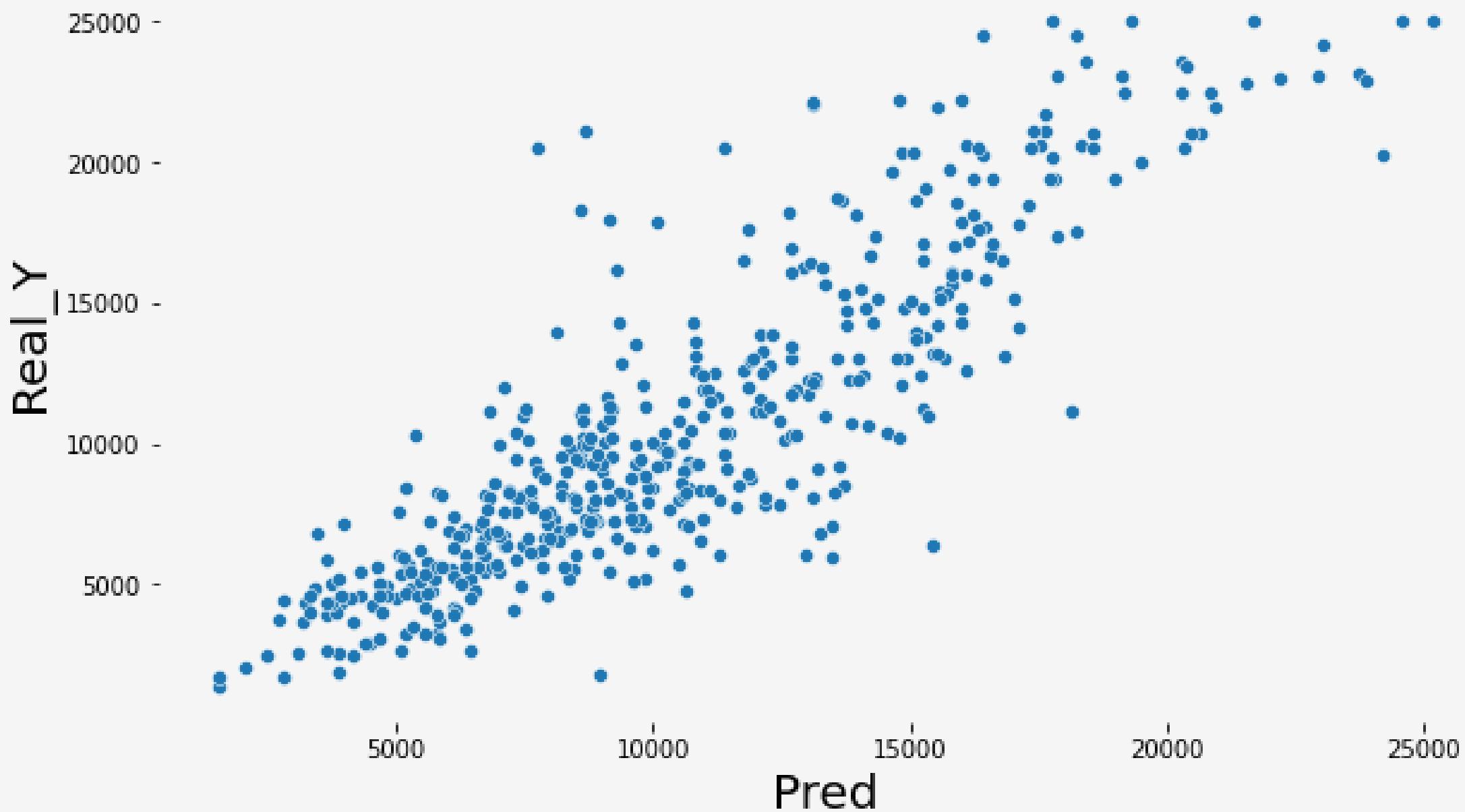
- 모델 평가
- 관중수 예측 핵심 변수

모델 평가

실제값과 예측값 그래프

예측값이 실제값과 정확하게 일치하진 않아도 꽤 비슷하게 따라감.
MAE가 약 2000으로, 평균적으로 2000명 정도 오차가 있는 모델.

Pred & Read_Y



관중수 예측 핵심 변수

변수 중요도를 이용하여

변수 중요도 값이 클 수록 관중수 예측 시 핵심적인 역할을 하는 변수임.
기온이 가장 중요하며, 홈팀-월-원정팀-홈팀순위-요일 변수 순.
프로야구 관중수는 요일에 민감한데,
홈팀순위가 그보다 더 중요한 변수.





발표를 들어주셔서
감사합니다 :)