



THE SPORTS

Start of Sports Storytelling

인천 유나이티드 분석 프로젝트

오경택

목차

프로젝트 목적
문제 정의, 시장 분석

모델링 과정
데이터 파악, 특성 분석

비즈니스 인사이트
스포츠 시장 분석
멤버십, 회원 분석

앞으로 계획
모델링, 특성
데이터 파이프라인

01

02

03

04

01

프로젝트 목적

01 문제 정의

코로나19로 인한 관중 저하

- 코로나 이후에 시즌 축소, 무관중 경기, 관중 제한 등으로 프로스포츠 산업 위기에 직면

지속적인 K리그의 인기 저하

- K리그 시장의 흐름이 좋지 않다. 2012년 이후 계속 하락세였으며 2019년에 반등했지만, 코로나로 다시 위기에 직면

01 프로젝트 목표

예매현황 예측 및 분석

- 기업에 영향을 주는 특정 경기 예매율 예측

소비자 분석 및 마케팅 방향 제시

- 고객의 예매 기록과 특성에 따라 고객을 세분화하고, 이에 맞는 액션 아이템 고안
- 고객 그룹 별로 차별화 된 가격과 프로모션 도입과 고급 좌석에 대한 이용률을 통해 파악

02

모델링 과정

02 데이터 파악

예측 모델

일자별 경기 결과
데이터

날씨 데이터
(2010~2022)

전시즌 구단별 평균 관중 수

전시즌 최종 순위

비즈니스 인사이트

예매 데이터

설문조사 데이터

멤버십 데이터

회원 데이터

사용하지 않은 데이터

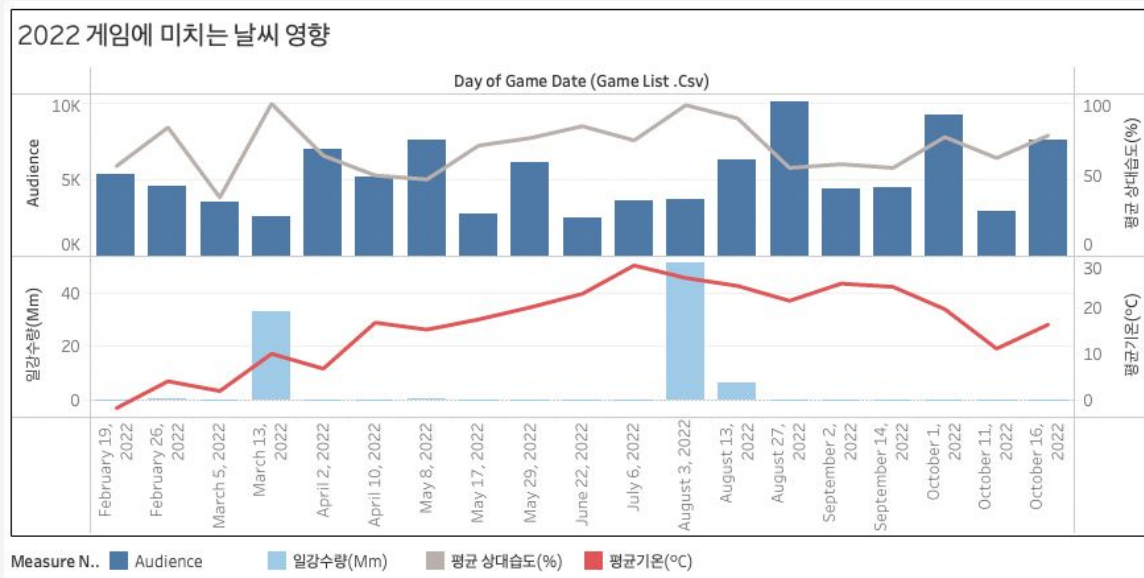
일자별 팀순위

날씨 데이터(2022,
인천)

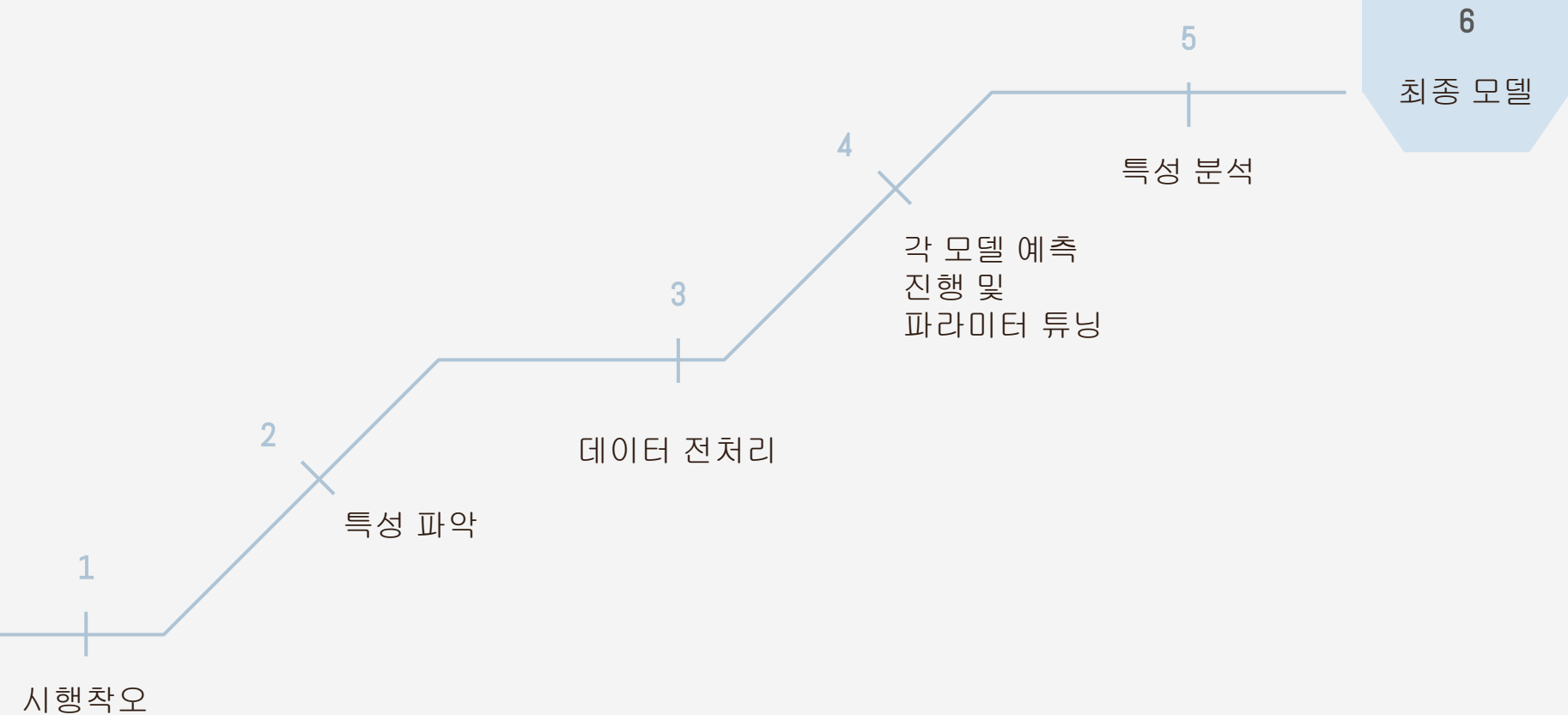
*핵심이 되는 주요 데이터

*외부에서 추가로 가져와 따로 제작한 데이터

02 데이터 파악



02 모델링 과정



02 시행 착오

1. 데이터 파악 명확성

- 모델예측에 필요한 데이터와 비즈니스 인사이트에 필요한 데이터 파악을 명확하게 나눌 필요성 느낌

2. 부족한 데이터의 한계

- 완성된 모델의 많이 부족한 결과 도출, 더 많은 데이터 양과 특성 재고려 필요성 느낌
- '평균 관객수'와 '전시즌 순위'등 꼭 필요한 특성 발견
- 다른 특성들 보다 각 구단의 인기도, 보유 팬 및 시즌 흥행 여부의 중요성 발견

3. 다시 시작

- 2012년부터 현재까지 모든 경기별 관객 수로 각 구단의 인기 규모를 파악
- 총 2400개 정도의 데이터를 가지고 분석, 전처리 부터 다시 시작함

02 주요 특성 파악



게임 날짜

게임 날짜, 년도, 월



경기 관련

홈경기 / 어웨이
게임 타입
일반 / 파이널 A / 파이널 B



주말 경기 여부

요일, 주말, 평일



일자별 날씨

해당 게임 날짜 기준의 날씨



전시즌 정보

전 시즌 순위
전 시즌 평균 관중 수



관중 수

결과 값(예측 해야하는 값)

02 전처리

01

결측치 및 이상치 제거

결측치 제거 &
'관중수' 2.5% 미만 97.5% 초과 데이터 삭제

02

각 특성 수정 및 보완

불필요 특성 제거 및 필요 특성 추가 제작
날씨 특성, 전시즌 데이터 조인

03

카테고리 특성 인코딩

'TargetEncoder' 및 'OrdinalEncoder'
사용해서 object값 int값으로 변경

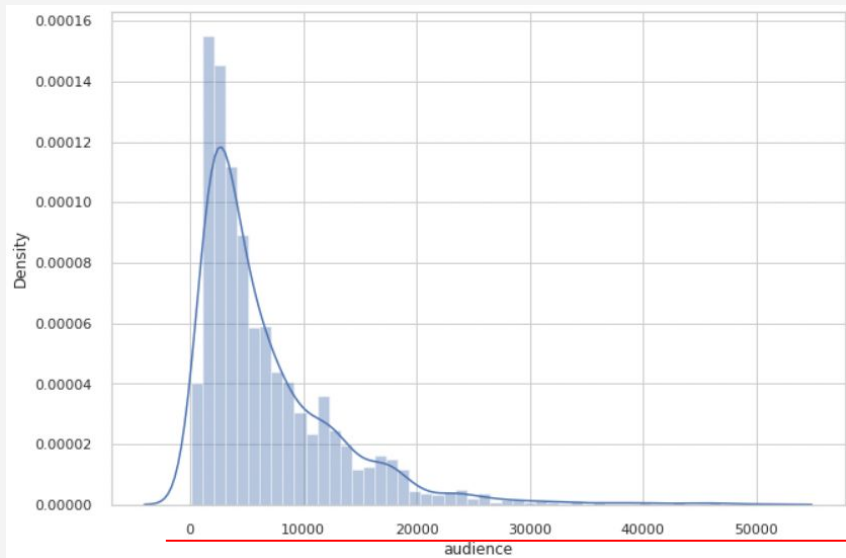
04

데이터 표준화

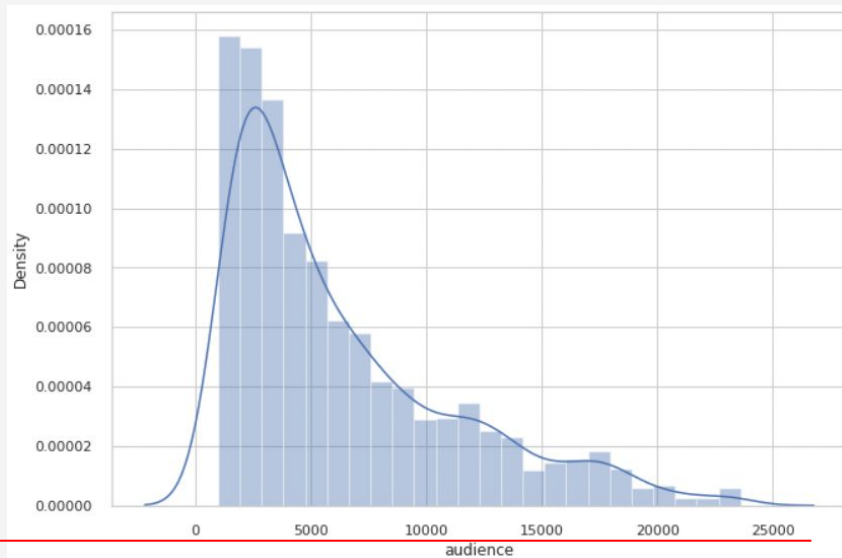
'StandardScaler' 이용하여
데이터 표준화

02 이상치 제거 비교

전



후



02 전처리 후 데이터

```
df_game.shape
```

```
(2154, 16)
```

데이터
크기

```
pd.options.display.float_format = '{:,.0f}'.format
```

```
df_game.sample(5)
```

Target

st_year	game_date	yoil	game_time	home_team_name	away_team_name	game_type_name	audience	weekend	month	temperature	precipitation	wind_speed	humidity	last_year_avg_audience	last_year_rank
2016	2016-07-16	토	19:00	수원fc	전남	스플릿일반	1512	1	7	21	49	4	100	1462	14
2017	2017-07-15	토	19:00	울산	광주	스플릿일반	5843	1	7	29	0	3	75	8744	4
2013	2013-08-28	수	19:00	서울	전북	스플릿일반	17516	0	8	26	0	3	59	20502	1
2015	2015-04-26	일	16:00	수원	대전	스플릿일반	11792	1	4	17	0	2	44	19608	2
2012	2012-06-24	일	17:00	전북	경남	스플릿일반	10777	1	6	25	0	2	56	16237	1

02 ML 모델

01

Baseline Model

관중수의 평균 값

02

Simple Linear Regression

'구단별 전시즌 평균 관중수' 특성 활용

03

Multiple Linear Regression

SelectKBest 특성 최적화

04

Ridge Regression

Ridge Alpha값 조절

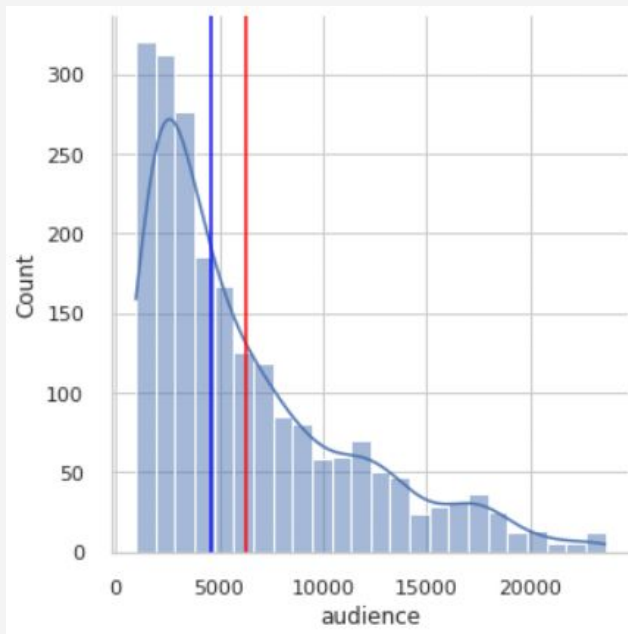
05

Random Forest

Random Search CV 파라미터 튜닝

02 Baseline Model

'관중 수'의 평균값



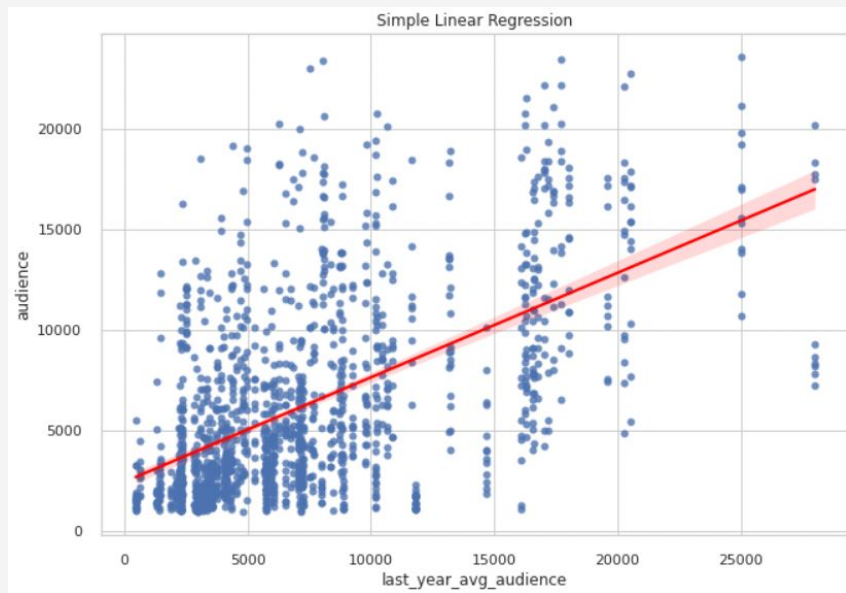
Mean: 6307.2

Median: 4613.0

	Metric	Score
0	MSE	24244053.8696
1	MAE	3838.7824
2	RMSE	4923.8251
3	R2	-0.0020

02 Simple Linear Regression

'지난 시즌 구단별 평균 관중 수' 특성 활용

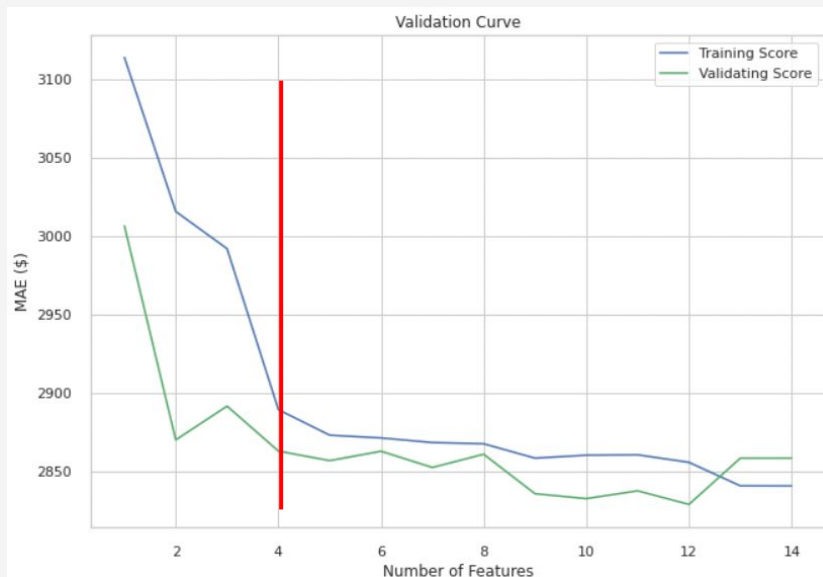


$$Y = 0.519x + 2458$$

	Metric	Score
0	MSE	15781683.6214
1	MAE	3000.5911
2	RMSE	3972.6167
3	R2	0.3478

02 Multiple Linear Regression

SelectKBest를 활용하여 최적의 특성 찾은 후 모델링



4 features: 홈팀, 주말, 전시즌 순위, 전시즌 평균 관중

	Metric	Score
0	MSE	14574320.867
1	MAE	2858.361
2	RMSE	3817.633
3	R2	0.406

02 Ridge Regression

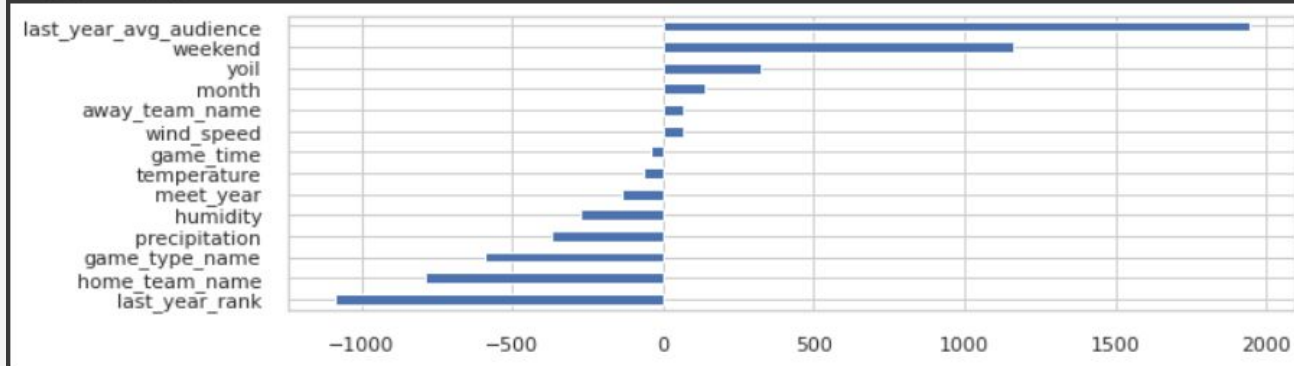
Alpha값을 조절하여 모델 최적화

Alpha = 0.02 일때 최적화

Ridge Regression, alpha=0.02

Validation MAE: \$2,782

Validation R2: 0.435



Metric

Score

0 MSE 13854204.480

1 MAE 2782.453

2 RMSE 3722.124

3 R2 0.435

02 Random Forest

Random Search CV를 통해 최적의 하이퍼 파라미터 탐색

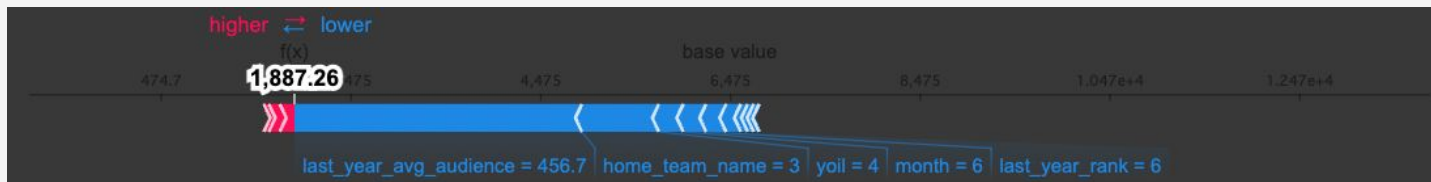
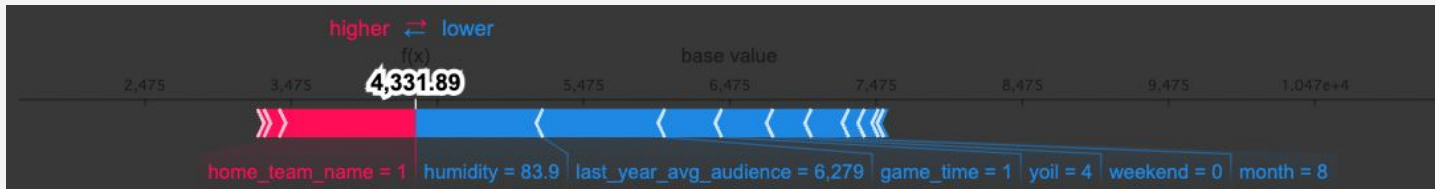
Hyper parameters:

- max_depth: 20
- max_features: 0.54
- n_estimators: 256
- simpleimputer__strategy: mean

	Metric	Score
0	MSE	9600585.543
1	MAE	2153.377
2	RMSE	3098.481
3	R2	0.609

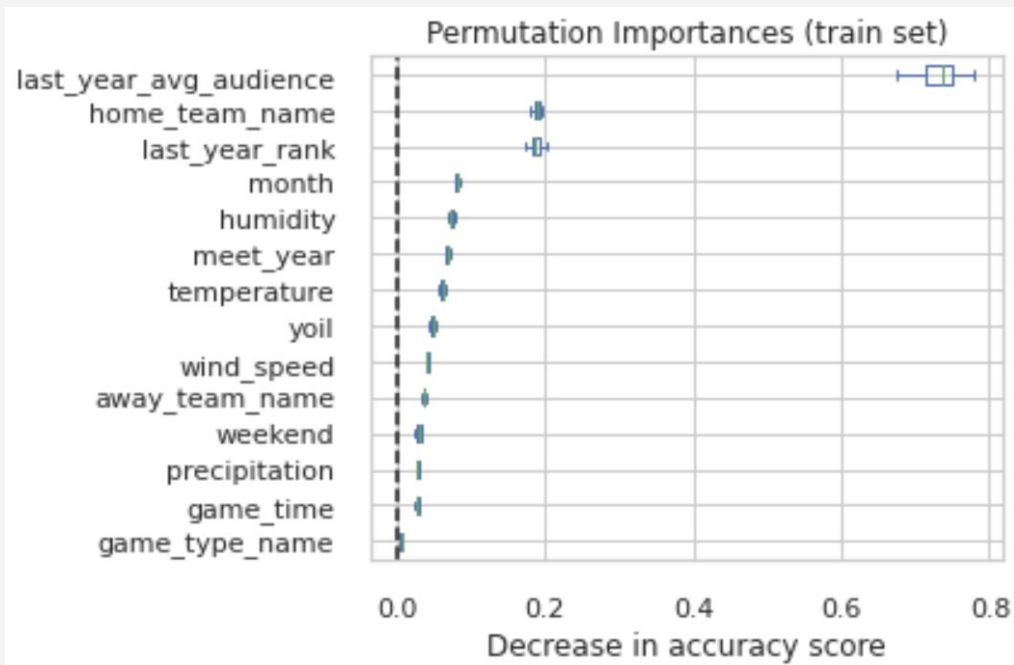
02 Shap

예제를 통한 각 특성들 그래프로 비교



02 Permutation Importance

랜덤 포레스트 모델의 각 특성을 모든 트리에 대해 평균 불순도 감소를 계산한 값



02 최종모델 Random Forest

하위 특성 4개 제거 후 최종 모델

Hyper parameters:

- max_depth: 15
- max_features: 0.38
- n_estimators: 252
- simpleimputer__strategy: mean
- Min_sample_leaf: 3
- smoothing : 60

	Metric	Score
0	MSE	7454247.829
1	MAE	2020.451
2	RMSE	2730.247
3	R2	0.666

02 최종모델 예시

인천 vs 서울 개막전 예상

- 년도: 2023
- 홈팀: 인천
- 어웨이팀: 서울
- 주말: 1
- 월: 2
- 온도: 10
- 풍속: 1
- 습도: 67
- 전 시즌 관중 수: 5261.5
- 전 시즌 랭킹: 4

6041.8명

03

비즈니스 인사이트

03 시장 분석

경쟁업체가 존재 하는가?

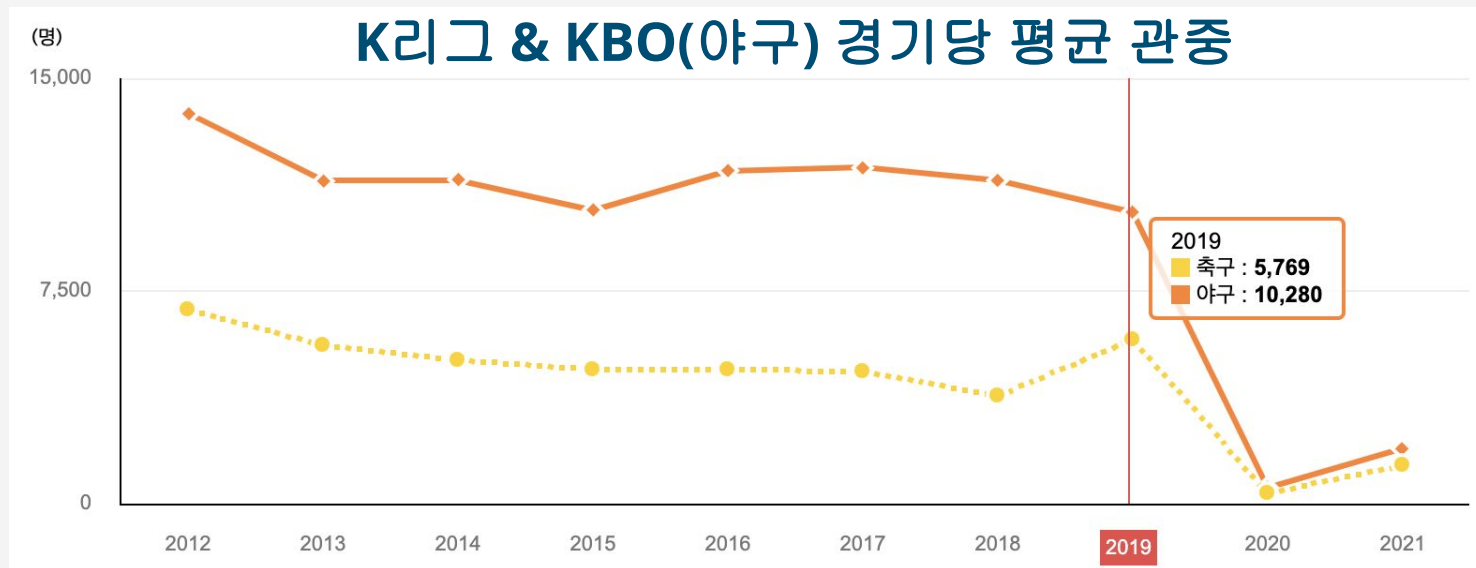
- 현재 더스포츠가 동종 업계 **유일**의 **CRM**, 데이터 마케팅 솔루션을 제공

각 구단들이 원하는것이 무엇인가?

- 데이터 기반으로 타겟 마케팅을 통해 최소한의 비용으로 **최적의 마케팅 효과**를 도출하길 원함
- 기본 보유 **팬 이탈 방지**를 위한 마케팅, 손해를 최소화 하길 원함

03 스포츠 시장 분석

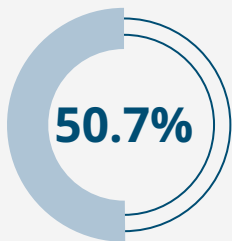
10년간 축구와 야구 관중 비교



03 스포츠 시장 분석

야구보다 축구를 좋아하는 사람이 더 많다

- Q. 귀하께서는 야구와 축구 중 무엇을 더 좋아하십니까?



축구



야구



잘 모름

*출처
조사기관: 리얼미터
조사기간: 2021년 1월 8일(1일간)
조사대상: 전국 18세 이상 남녀
표본수: 500명

03 스포츠 시장 분석

왜 K리그에 관심이 없을까?

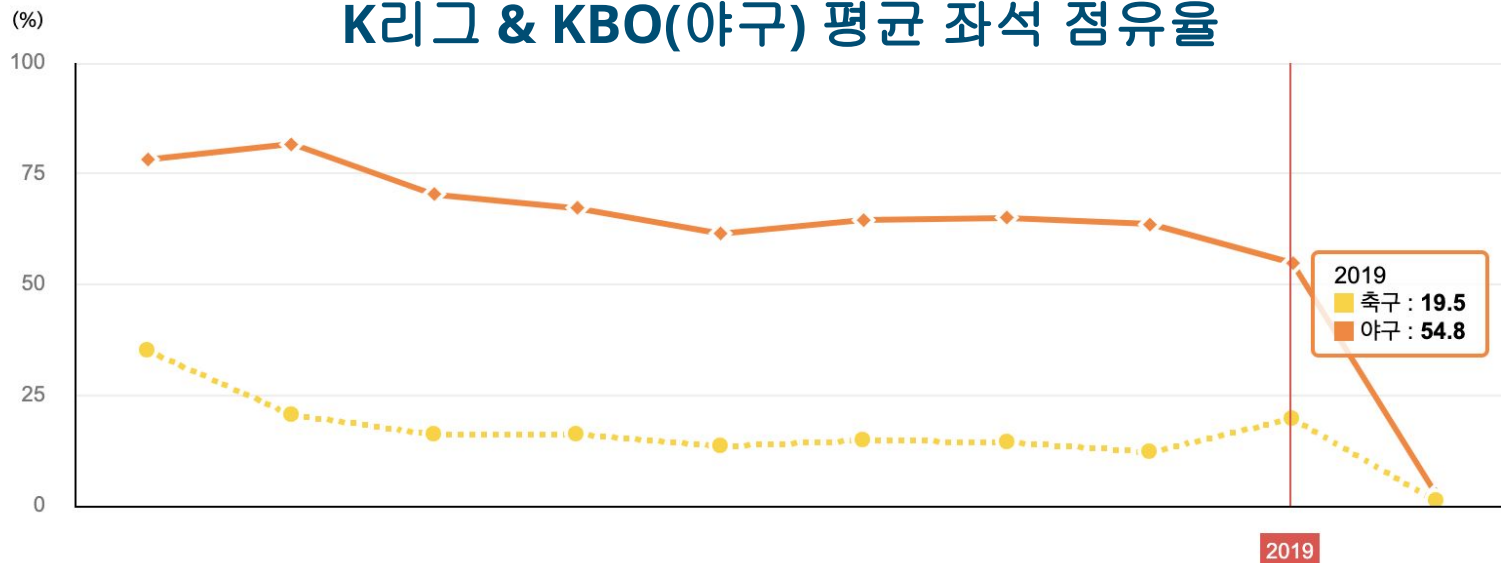
- 재미없다는 인식
- 선수들 기량
- 조작 매수 이력
- 탑 스타 부재
- 카메라 워크
- 선수간의 선후배 관계

이러한 이유들로 해외 축구리그와 비교가 될 수는 있지만, K리그의 가장 큰 메리트는 **직관**

03 스포츠 시장 분석

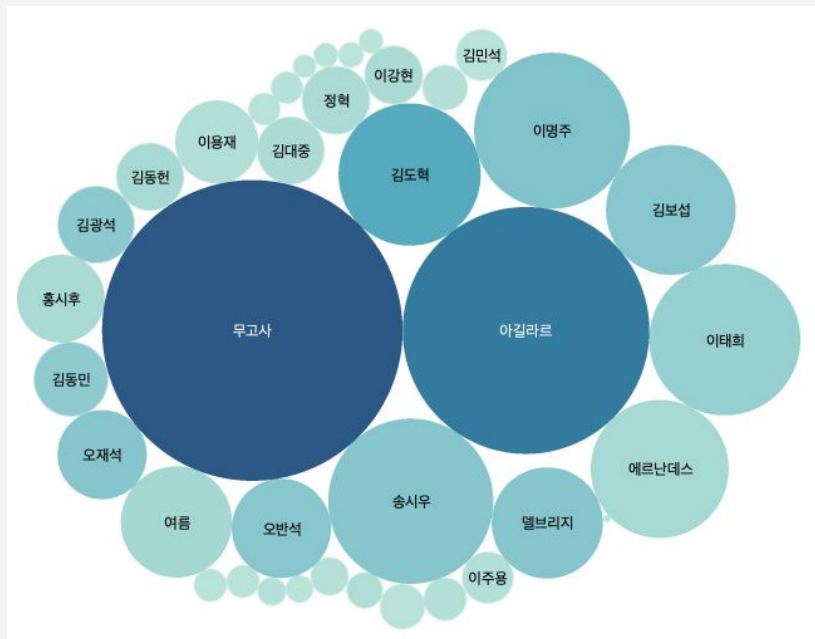
경기장 점유율을 높일 필요 있음

K리그 & KBO(야구) 평균 좌석 점유율



03 인천 팬 분석

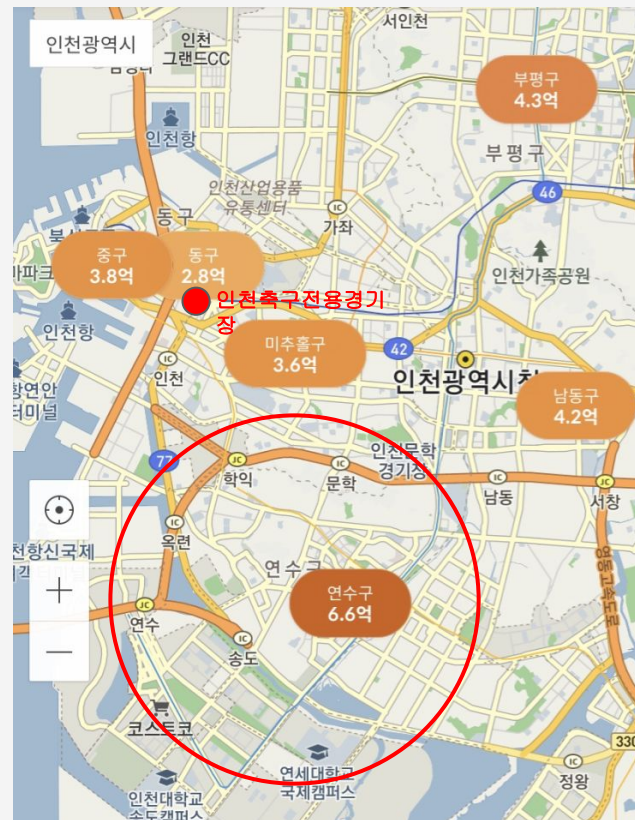
2022 인천 팬들이 좋아하는 선수



1. 무고사 현재 이적
2. 아길라르
3. 송시우
4. 이명주
5. 이태희
6. 김도혁
7. 에르난데스
8. 김보섭
9. 여름
10. 델브리지

2022 인천 팬 거주지역 상위 10개

21986	연수구 송도동	22188	미추홀구 용현동	21561	남동구 구월동	22003	연수구 송도동
		21982	연수구 송도동	22002	연수구 송도동	22160	미추홀구 용현동
22009	연수구 송도동	22001	연수구 송도동	22018	연수구 송도동		

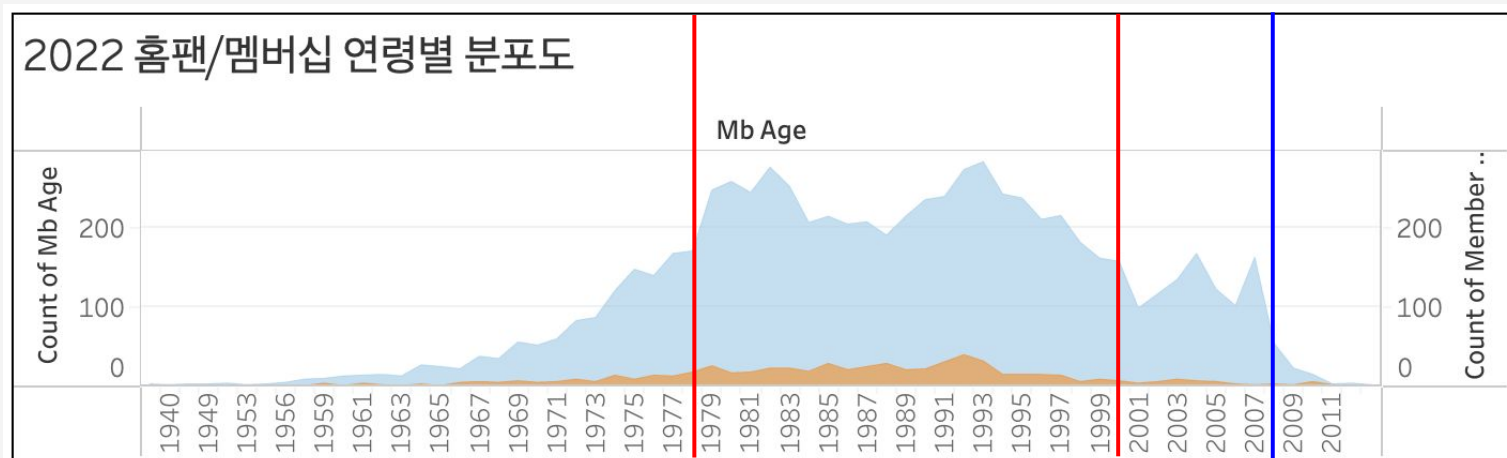


03 인천 팬 분석

2022 인천 팬 연령

99년생 - 24세

79년생 - 44세

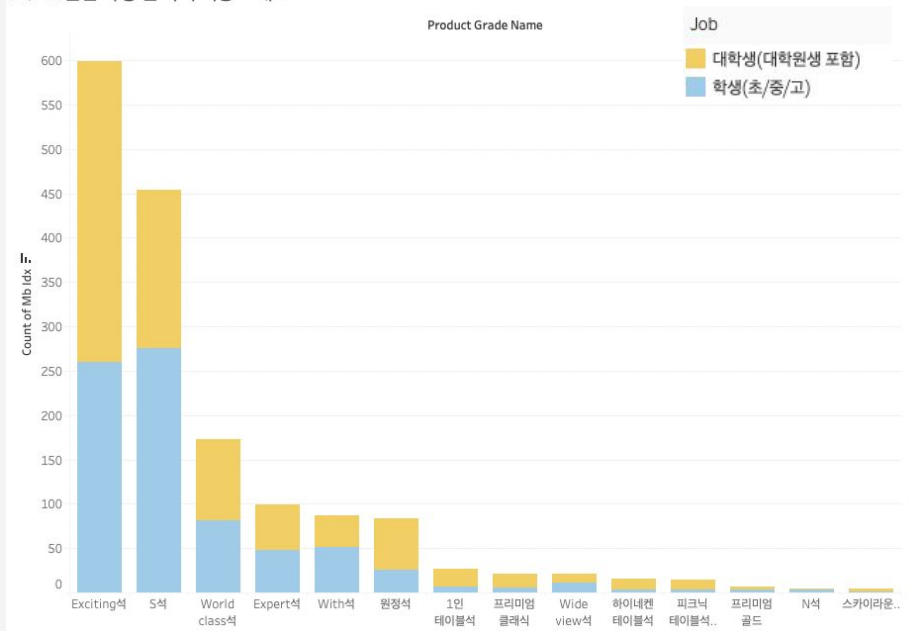


03 인천 팬 분석

2022 인천 팬 직업군

Job	Null	블루	블랙
회사원(사무직/연구직)	1,618	130	78
회사원(제조/건설/기술직)	666	31	24
대학생(대학원생 포함)	609	30	6
학생(초/중/고)	613	17	12
자영업자	372	16	15
전업주부	217	5	12
프리랜서	211	12	4
공무원	195	13	3
무직	182	9	4
서비스업 종사자	166	20	6
교사	149	16	7
보건 및 의료종사자	116	12	4
군인(소방/경찰/교정직 포함)	76	2	4
금융 및 보험업 종사자	50	5	8
예술 및 디자인 종사자	54	4	2
방송 및 언론 종사자	25	8	2
법률 종사자	16	2	2

2022 인천 학생 팬 좌석 이용 그래프



03 멤버십 분석

학생들을 위한 프로모션이 무엇이 있을까?

멤버십

03 멤버십 분석

2022 인천 멤버십 등급 및 개선 방안

2022시즌 멤버십 등급

1개 아이디 당 멤버십 블루, 블랙, 골드 중 1개만 구매 가능



2022 멤버십
블루

70,000원
(선착순 110개 한정 판매)



2022 멤버십
블랙

200,000원
(선착순 50개 한정 판매)



2022 멤버십
골드

3,000,000원
(선착순 10개 한정 판매)

팬들이 이용하는 가장 큰 이유는 **선예매**

개선 필요한 부분

1. 현재 단 하루 일찍 선 예매 가능
2. 멤버십 간의 차별화 부족
3. 골드 멤버십 회원 없음
4. 상품권 단일화

03 멤버십 제안

2022 인천 멤버십 등급 차별화

예매 할인권 → 예매 티켓당 할인

	블루	블랙	골드
선예매 티켓 수	3	5	10
예매 할인권	5천 x 2	5천 x 10	1만 x 10

	블루	블랙	골드
선예매 티켓 수	3	5	10
예매 티켓당 할인 (가능좌석)	S Exciting Export	S/E/E + World With Wide	프리미엄 포함 전좌석

고급 좌석 이용률 증가 효과

03 멤버십 제안

2022 인천 멤버십 등급 개선 필요

2023 멤버십
학생

선예매 2장 가능

S석, Exciting석
티켓 할인

현재 골드 등급 회원 없음 (*개선 필요)

대상: 인천광역시 내 재학중인 학생 모두

가격: 1~2만원

* 인원 제한 필요

04

앞으로 계획

FUTURE PROJECTS

01

모델링 개선

현재 모델 성능이 매우 낮으므로
신경망모델(NN) 모델을 만들 필요 있음

02

특성 개선 및 도출

더 확실한 특성을 찾을 필요 있음
예) 날짜별 순위 / K1,K2의 차이 등

03

데이터 축적
파이프라인 개선

설문조사 등 현재 쌓고 있는 데이터
파이프라인을 좀 더 활용 가능성 있도록
개선 필요 (현재 데이터 매우 복잡)

04

무기 및 강점

스포츠 구단들이 더스포츠를
필요로 할 수 밖에 없는 "이유" 찾기

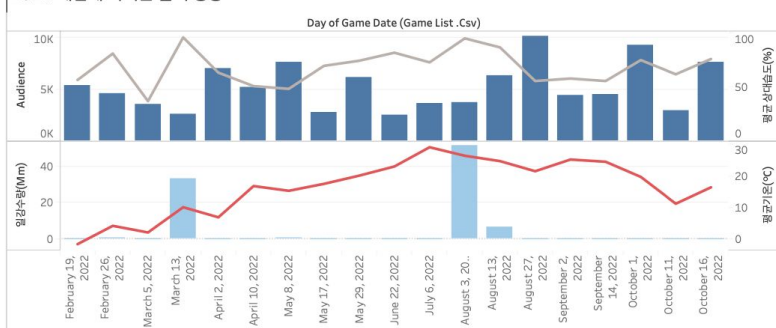
RESOURCES

Data & Website

- [\[더스포츠\] 더스포츠 관련 정보](#)
- [\[기상청\] 날씨 데이터\(2010~2022\)](#)
- [\[프로스포츠\] 관중 정보 데이터](#)
- [\[K리그 데이터 포털\] 관중 정보 데이터](#)
- [\[네이버\] 순위 정보 데이터](#)
- [\[e-나라지표\] 프로스포츠 운영 현황](#)
- [\[유튜브/리얼미터\] 축구&야구 조사](#)
- [\[공식 홈페이지\] 인천 유나이티드 멤버십 관련](#)

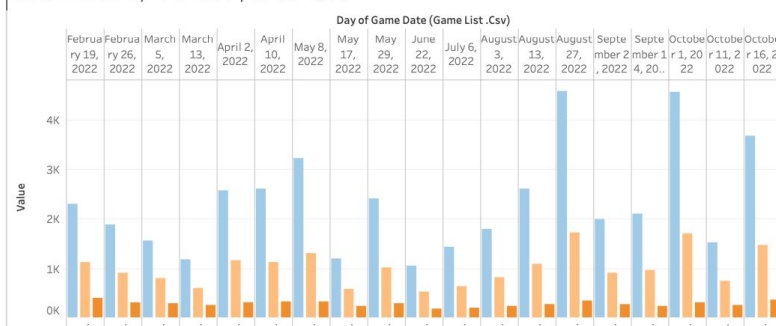
Dashboard 1

2022 게임에 미치는 날씨 영향



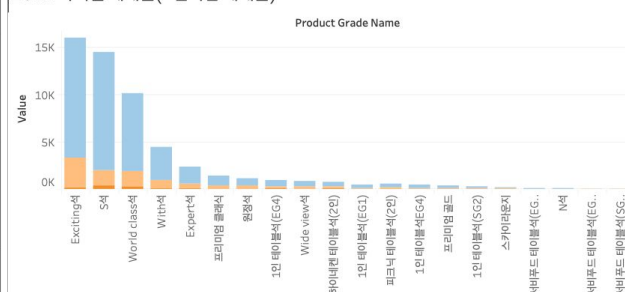
Measure N.. Audience 일강수량(Wm) 평균 상대습도(%) 평균기온(°C)

2022 전체관중 수 / 예매 티켓 수 / 멤버십 티켓 수

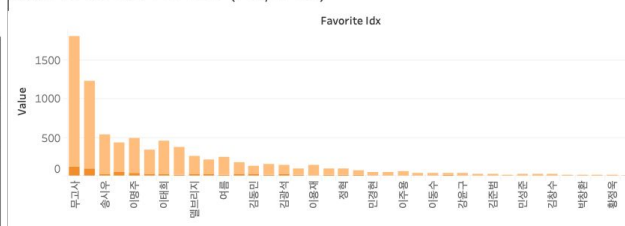


Measure N.. Count of Product Count of Mb idx Count of Memb.

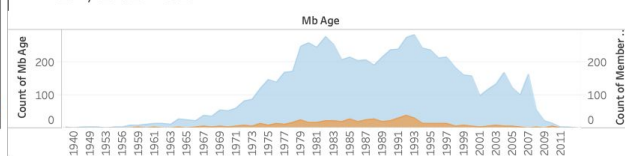
2022 작식별 예매율(+멤버십 예매율)



2022 인천 팬이 좋아하는 선수 (일반/멤버십)

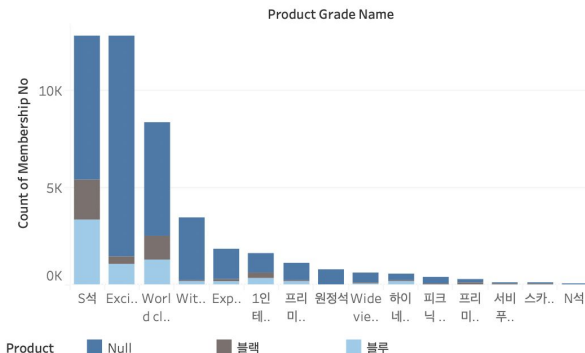


2022 홈팬/멤버십 연령별 분포도



Dashboard 2

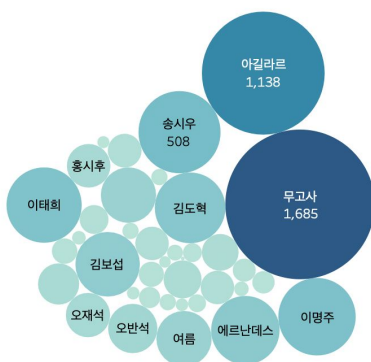
2022 인천팬(+멤버십) 좌석 이용률



2022 인천 팬들 좌석

Product Grade Na..	Null	블루	블랙
S석	7,359	3,352	2,064
Exciting석	11,312	1,033	430
World class석	5,868	1,281	1,211
With석	3,222	148	70
Expert석	1,522	183	104
1인 테이블석	970	317	307
프리미엄 클래식	887	145	91
원정석	788	2	
Wide view석	508	63	52
하이네켄 테이블석	314	144	88
피크닉 테이블석(2인)	304	28	30
프리미엄 골드	167	18	78
서비푸드 테이블석	79	13	23
스카이라운지	77	19	11
N석	42	6	

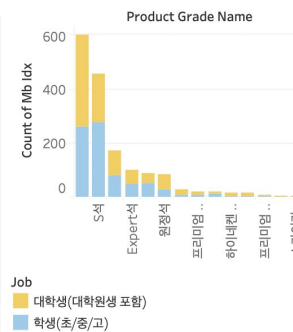
2022 인천 팬이 좋아하는 선수



2022 인천 팬 직업군

Job	Null	블루	블랙
회사원(사무직/연구직)	1,618	130	78
회사원(제조/건설/기술직)	666	31	24
대학생(대학원생 포함)	609	30	6
학생(초/중/고)	613	17	12
자영업자	372	16	15
전업주부	217	5	12
프리랜서	211	12	4
공무원	195	13	3
무직	182	9	4
서비스업 종사자	166	20	6
교사	149	16	7
보건 및 의료종사자	116	12	4
군인(소방/경찰/교정직 포함)	76	2	4
금융 및 보험업 종사자	50	5	8
예술 및 디자인 종사자	54	4	2

2022 인천 학생 팬 좌석 이용 그래프



감사합니다

오경택

010-2127-9718

kyungtaek95@gmail.com