유럽 5대 리그
선수 가격 예측 ML 개발

CONTENTS

01

분석 배경 / 목표

- 세계 축구 시장의 성장
- 스포츠 마케팅의 활성화
- 축구 구단 운영과 선수 영입

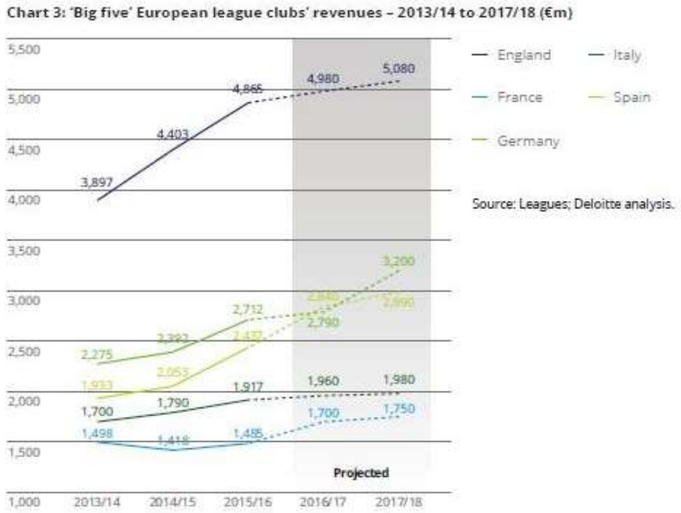
세계 축구 시장의 성장

'호황' 유럽 축구시장 규모 32조 돌파...잉글랜드 6조로 1위

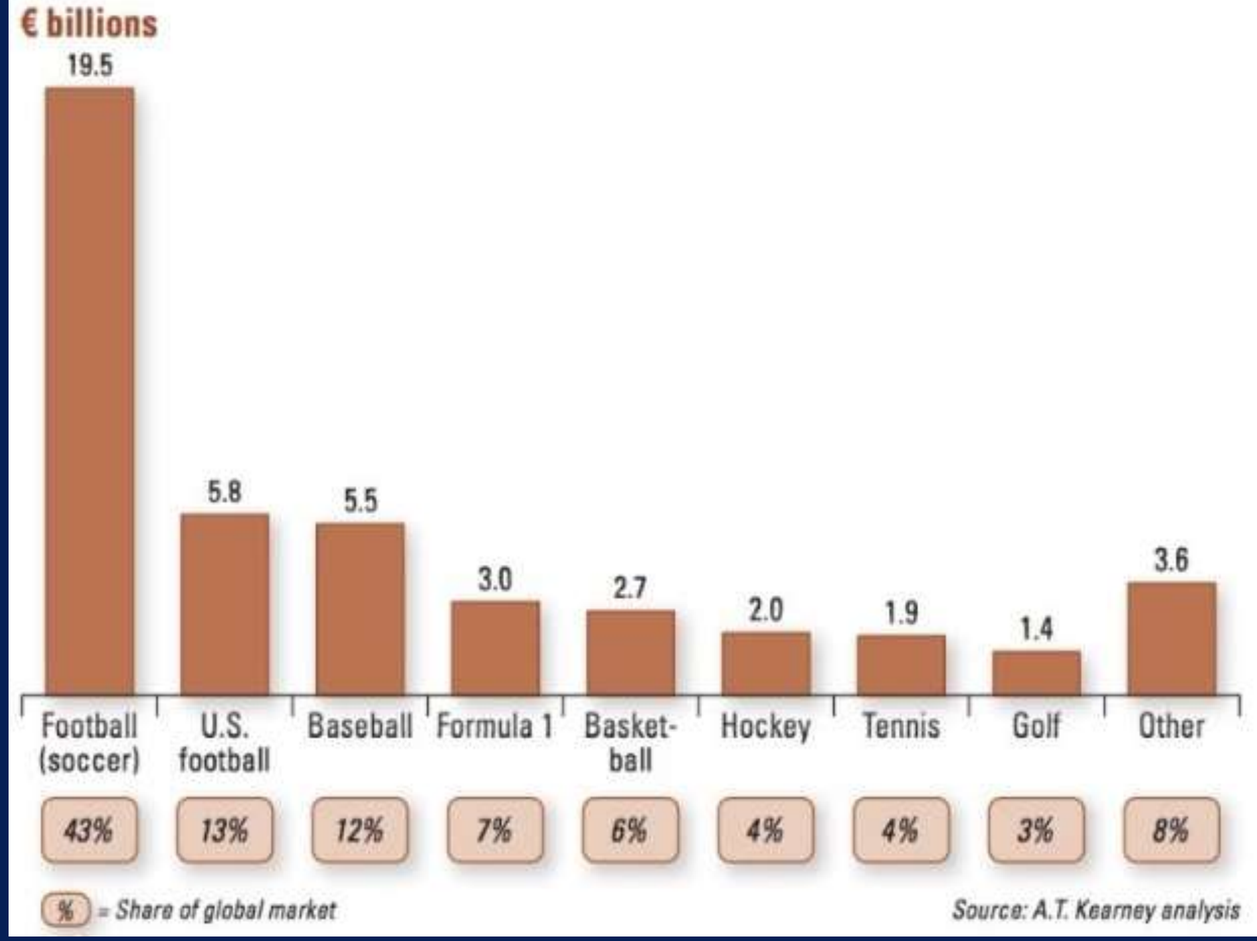
송고시간 | 2017-07-12 14:52

2015-2016시즌 기준, 전년 대비 12.8%↑증가...5대 빅리그가 54%

스페인, 2017-2018시즌 독일 따라잡고 2위 올라설 듯



유럽 5대 국가 축구시장 규모 증가 추이.[딜로이트 보고서 캡처]



스포츠 마케팅의 활성화


BARCLAYS
PREMIER LEAGUE
Premier
League


kt 5G
축구국가대표팀
이강인 선수 후원
kt


프리미어리그 TV중계권료 증가 추이
(단위:파운드)
5
4
3
2
1
0
1억9100만
1억7000만
12억
10억2400만
17억600만
17억7300만
30억1800만
51억3600만
1992~97
1997~01
2001~04
2004~07
2007~10
2010~13
2013~16
2016~19

스포츠 마케팅의 활성화

축구 성적과 직결되는 구단 수입

'상금만 1066억+α' 첼시...이래서 챔스는 돈 방석

첼시, 맨시티 꺾으며 구단 통산 두 번째 우승
우승 상금 및 중계권료 배분 등 2000억 수익

리버풀, 토트넘 꺾고 통산 6번째 챔스 우승
우승상금 252억...대회 총수익만 약 1000억

UCL 우승 상금은 얼마? 본선 진출만 해도 215억

기하급수적으로 상승하는 선수 몸값

MOST VALUABLE PLAYERS
IN THE WORLD

		AGE	MARKET VALUE
1	MBAPPÉ	22	€180M
2	NEYMAR	29	€128M
3	KANE	27	€120M
4	HAALAND	20	€110M
5	SALAH	28	€110M
6	SANCHO	21	€100M
7	A.-ARNOLD	22	€100M
8	STERLING	26	€100M
9	MANÉ	28	€100M
10	DE BRUYNE	29	€100M

transfermarkt

5/5

BIGGEST MARKET VALUE WINNERS
IN 2021

		MARKET VALUE	INCREASE
1	VLAHOVIĆ	€70M	+€54M
2	HAALAND	€150M	+€50M
3	VINICIUS JR.	€100M	+€50M
4	PEDRI	€80M	+€50M
5	BELLINGHAM	€75M	+€48M
6	WIRTZ	€70M	+€46M
7	MUSIALA	€55M	+€45M
8	GAVI	€40M	+€40M
9	SMITH ROWE	€38M	+€35.3M
10	GREALISH	€80M	+€30M

transfermarkt

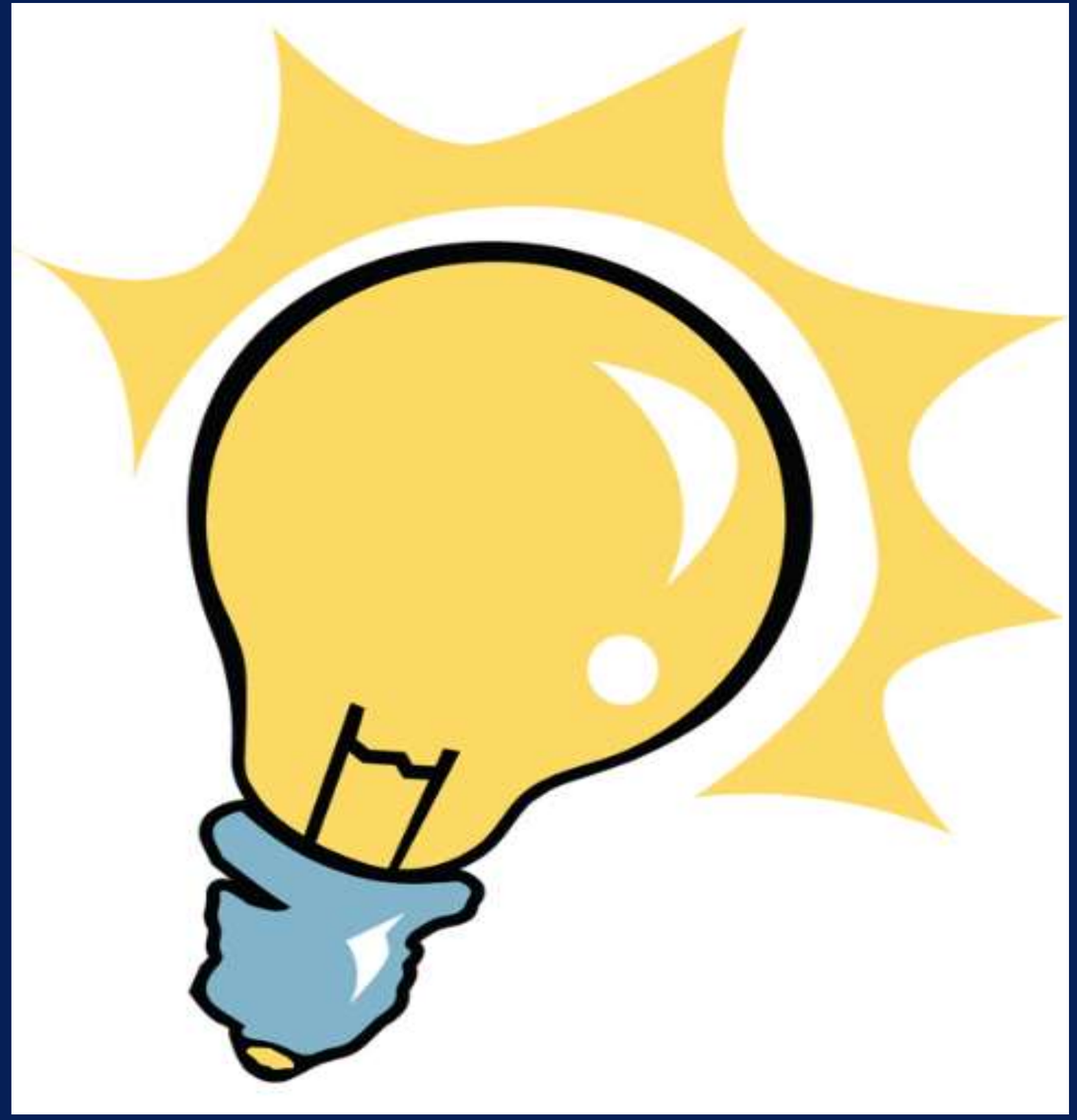
축구 구단 운영과 선수영입

	2016-2017
Real Madrid	419,3
FC Barcelona	390,7
Atlético de Madrid	182,8
Valencia CF	129,7
Sevilla FC	123,8
Villarreal	76,9
Athletic Club	61,4
Real Sociedad	56,7
RCD Espanyol	47,4
Real Betis	44,6
Málaga CF	43,1
Celta de Vigo	39,3
Granada CF	31,9
Leganés	30,2
Alavés	28,6
Deportivo de la Coruña	24,7
Las Palmas UD	24,6
SD Eibar	23,5
Sporting de Gijón	21,3
Osasuna	15,6
Total	1.816,3

구단 수입이 좋은
선수들의 몸값을
따라가지 못하고 있다.

구단 운영을 위한 전략적인 영입정책이 필요

분석을 통한 선수 가격 예측 ML



02

EDA

- 데이터 이해
- 데이터 전처리
- 데이터 EDA

데이터 소개

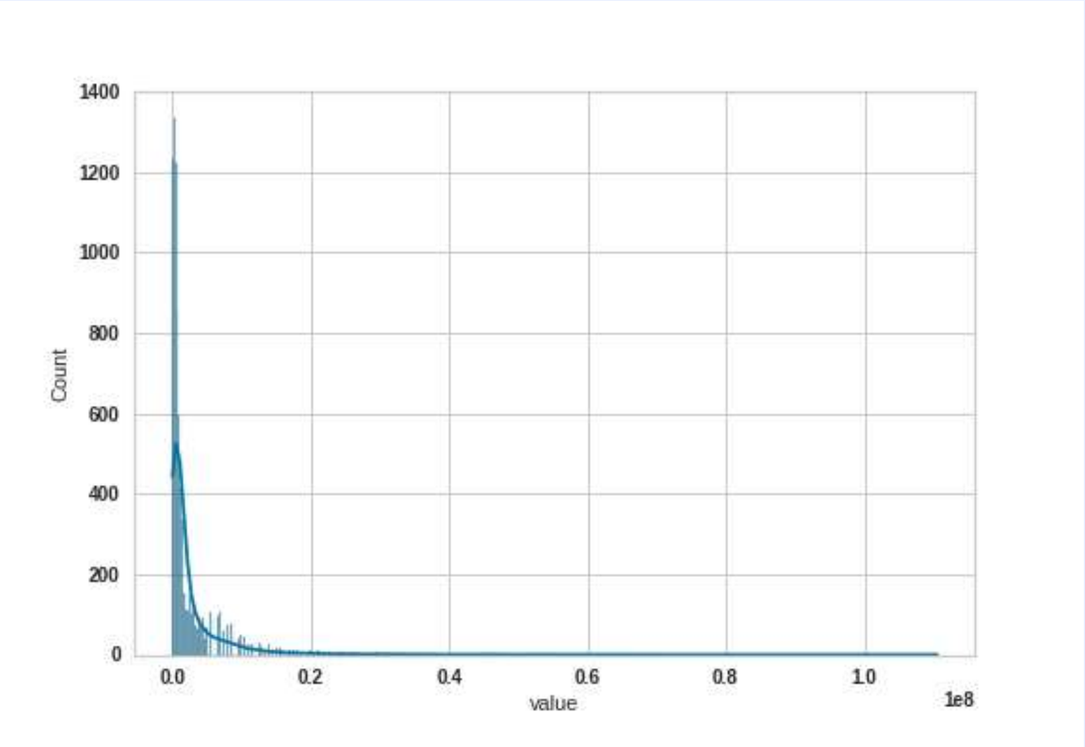
- 범주형 데이터

1. id
2. name
3. continent
4. contract_until
5. position
6. prefer_foot
7. reputation
8. stat_skill_moves

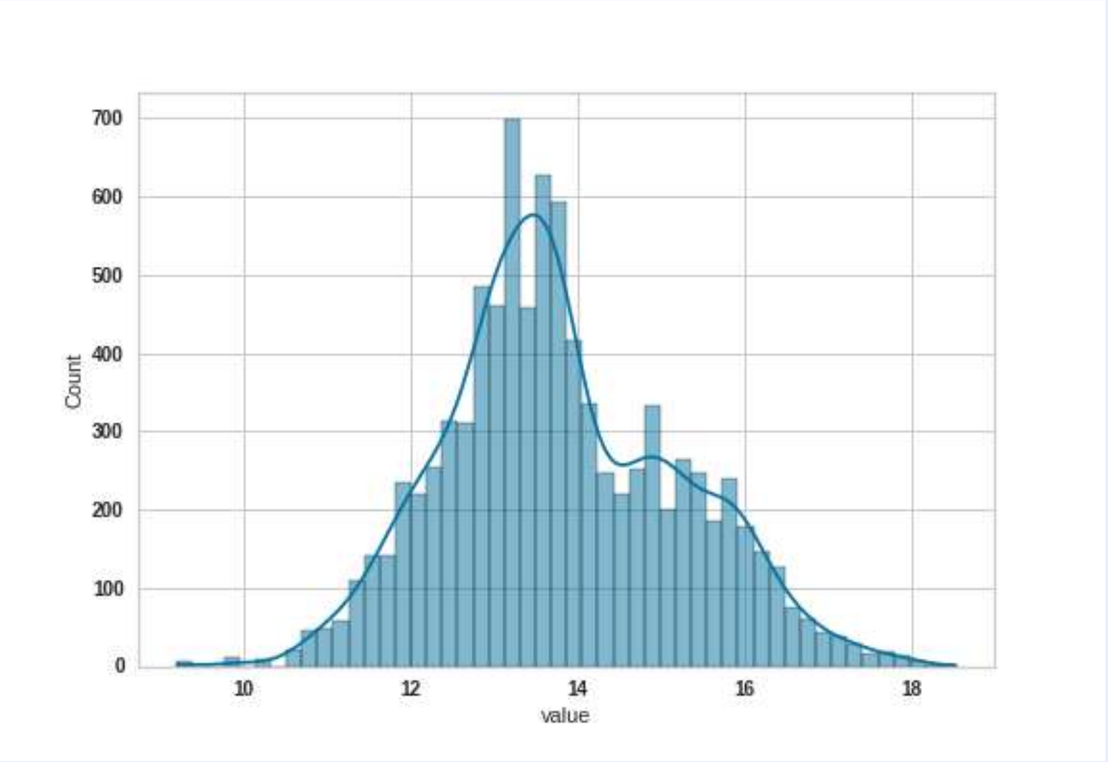
- 연속형 데이터

1. value

선수 가치 그래프

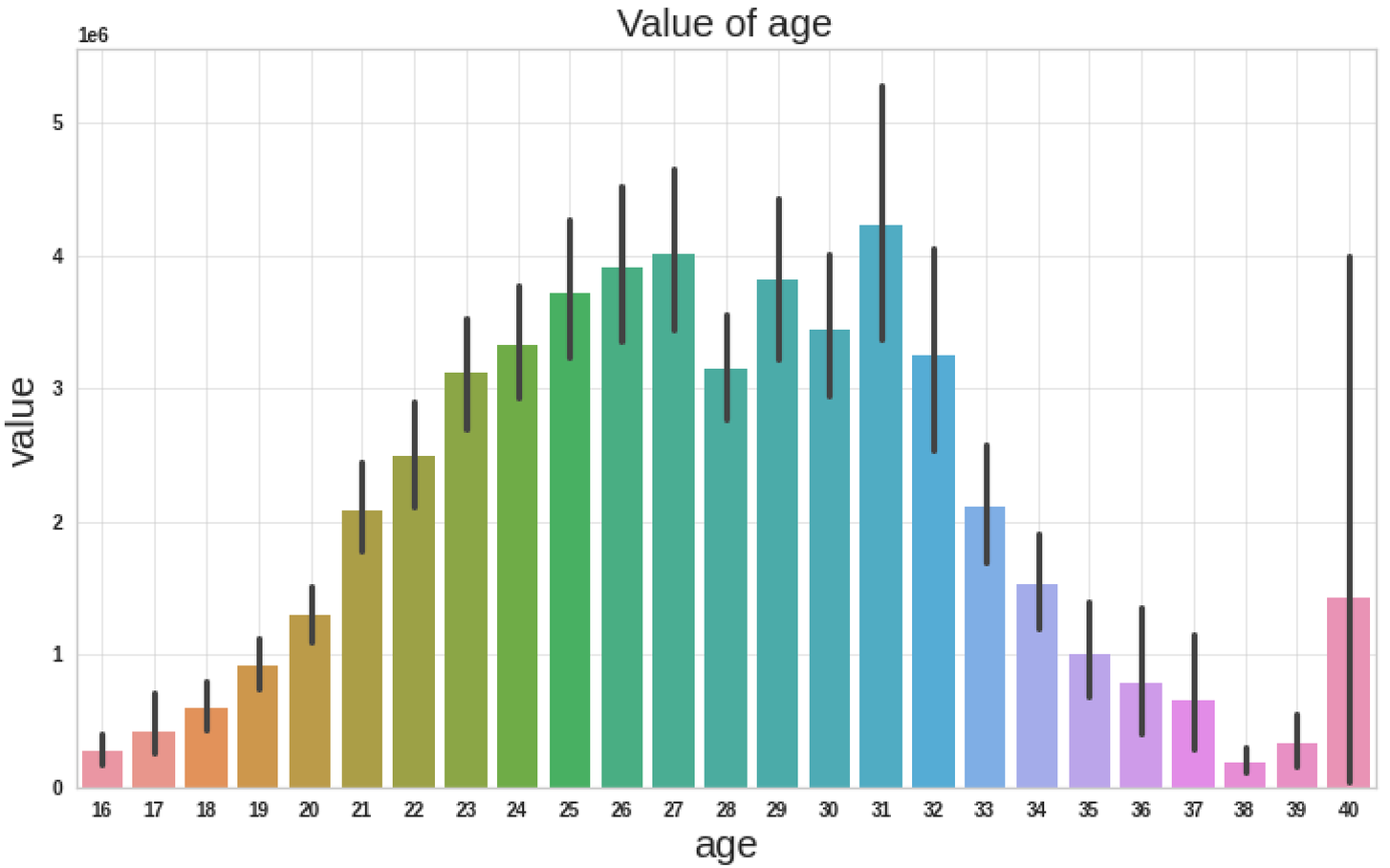


선수 가치 그래프

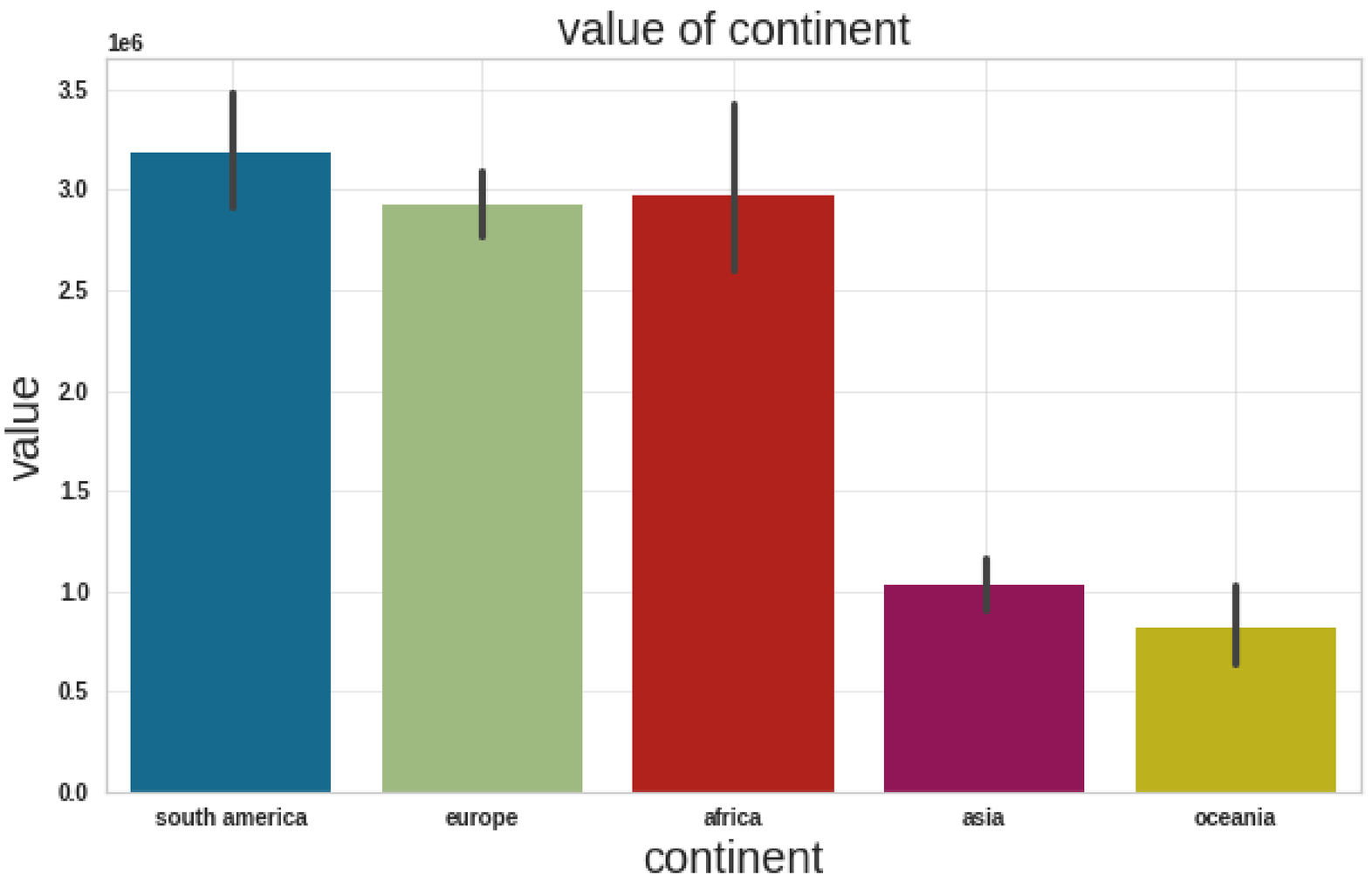


log 정규화를 이용

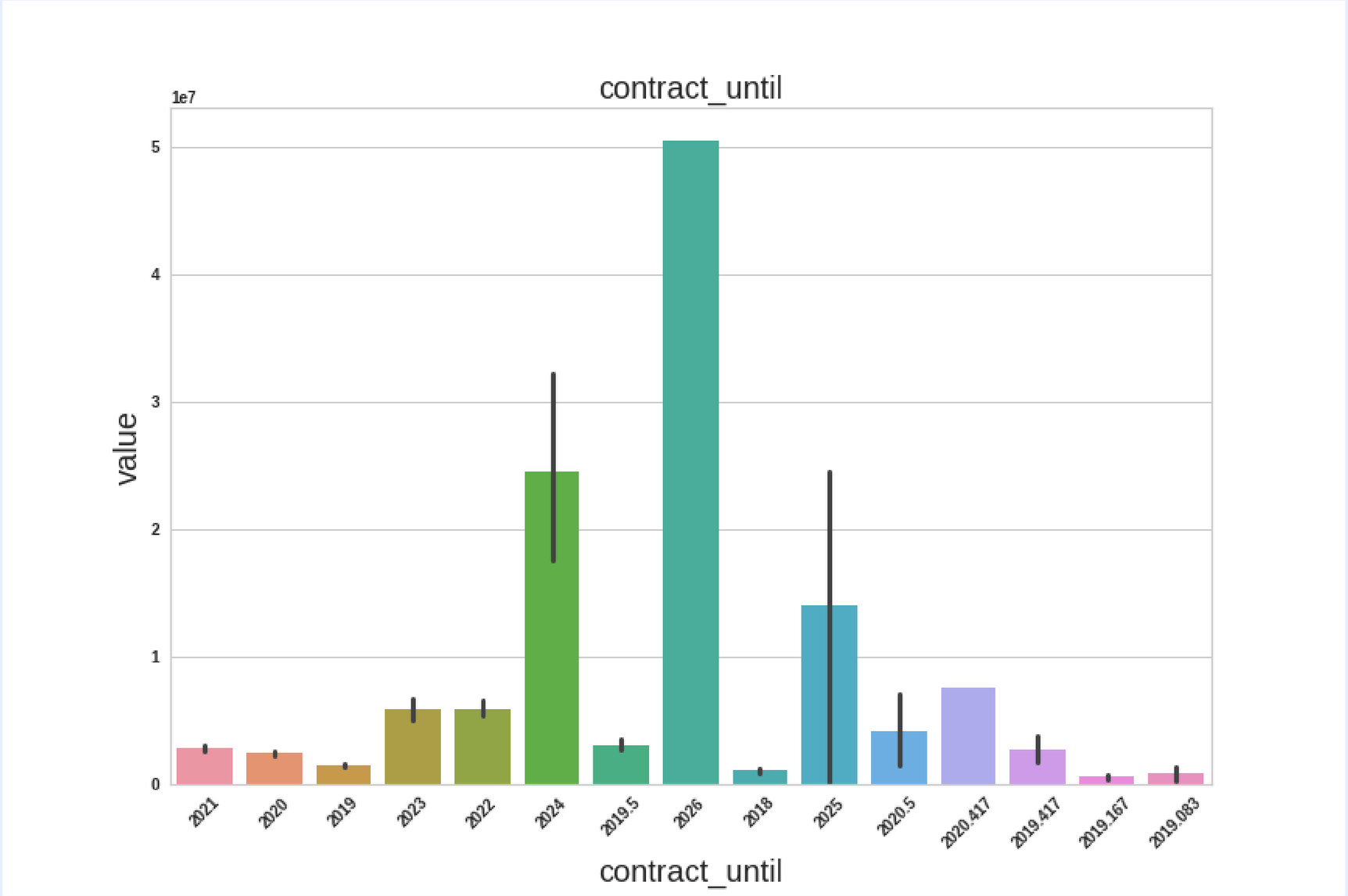
선수 나이 그래프


Value of age
1e6
5
4
3
2
1
0
value
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
age

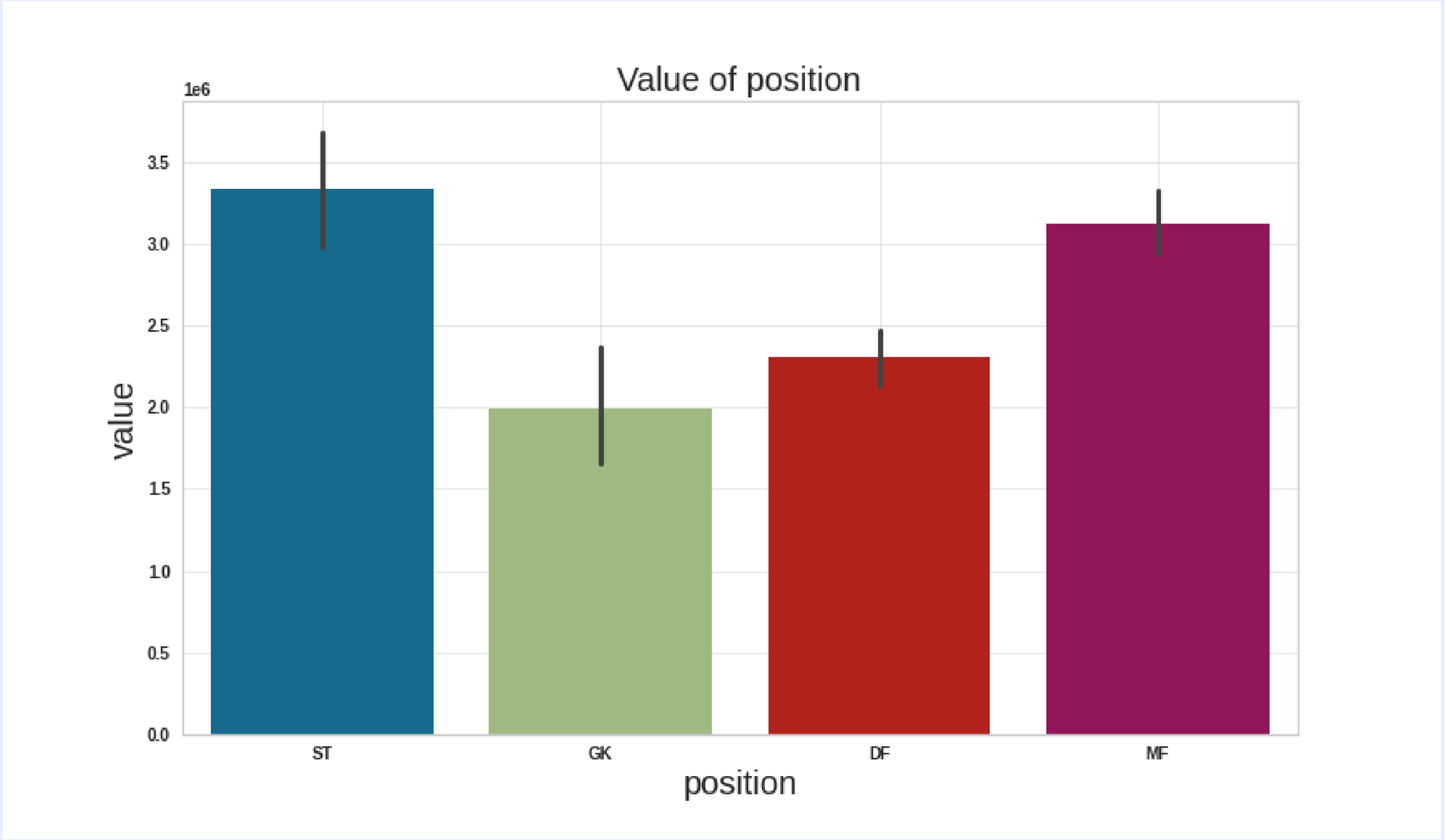
대륙별 선수 그래프


value of continent
1e6
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
value
south america
europe
africa
asia
oceania
continent

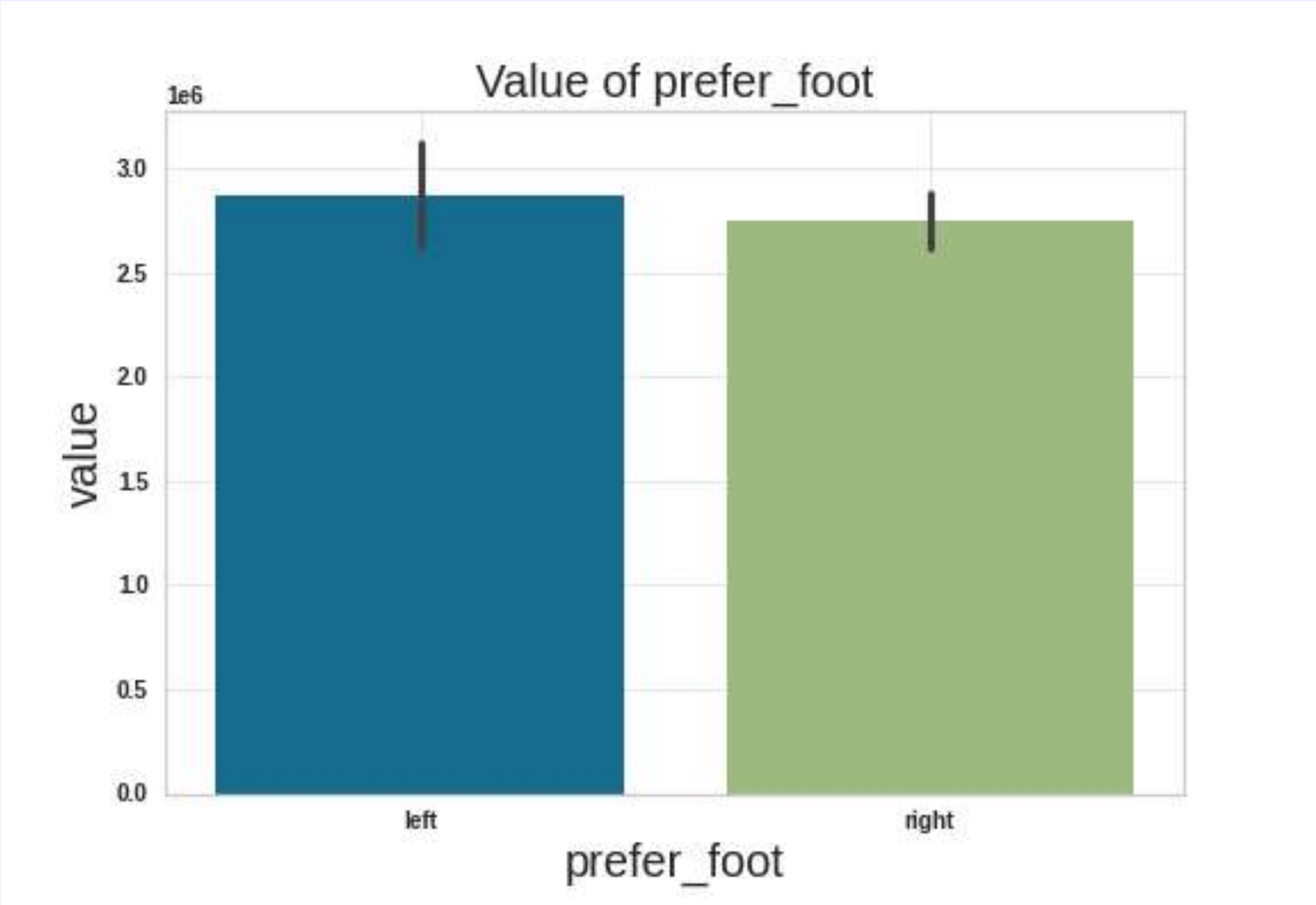
계약 기간별 선수 가치


contract_until
1e7
value
5
4
3
2
1
0
2021
2020
2019
2023
2022
2024
2019.5
2026
2018
2025
2020.5
2020.417
2019.417
2019.167
2019.083
contract_until

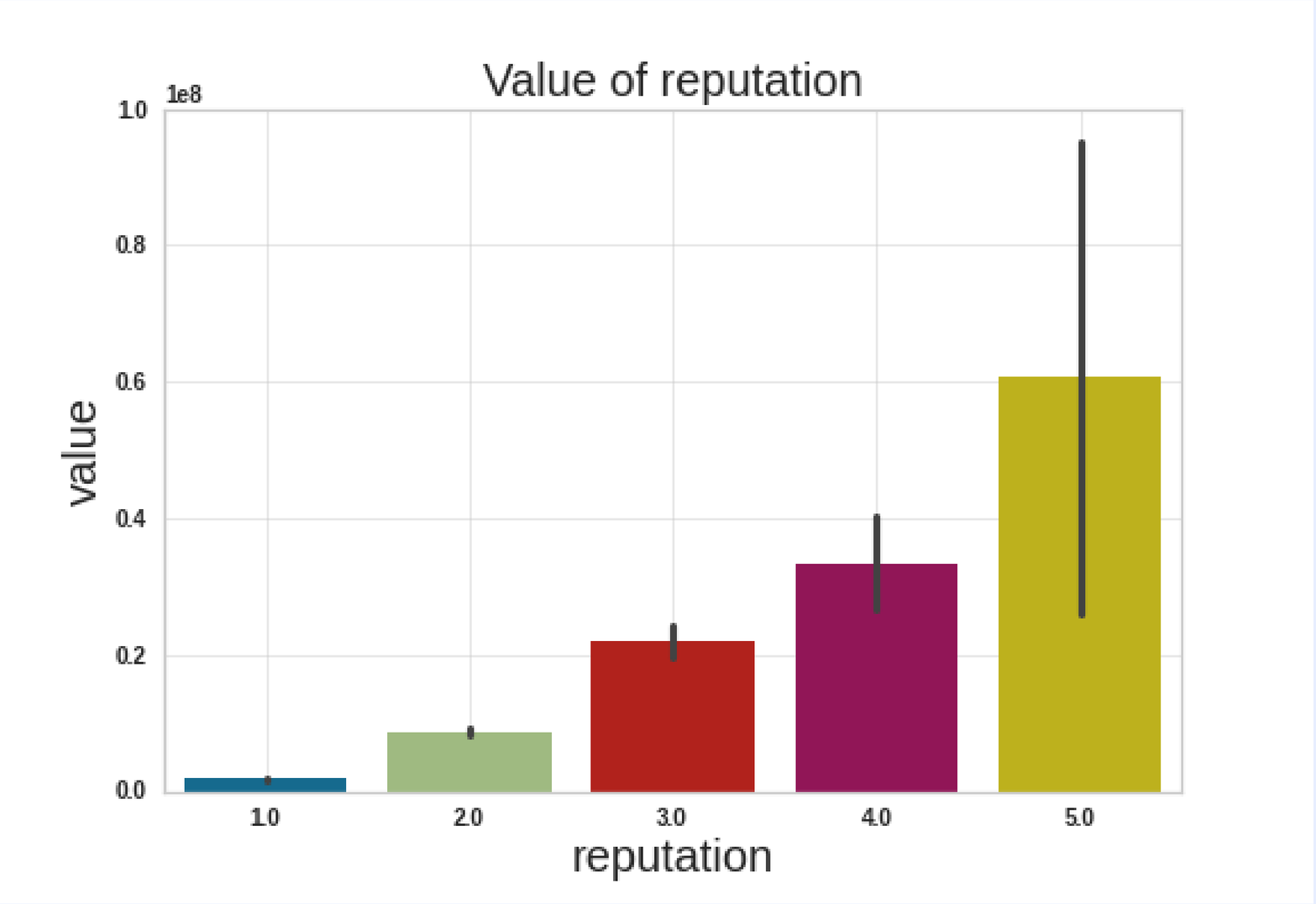
포지션 별 선수 가치


Value of position
1e6
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
value
ST
GK
DF
MF
position

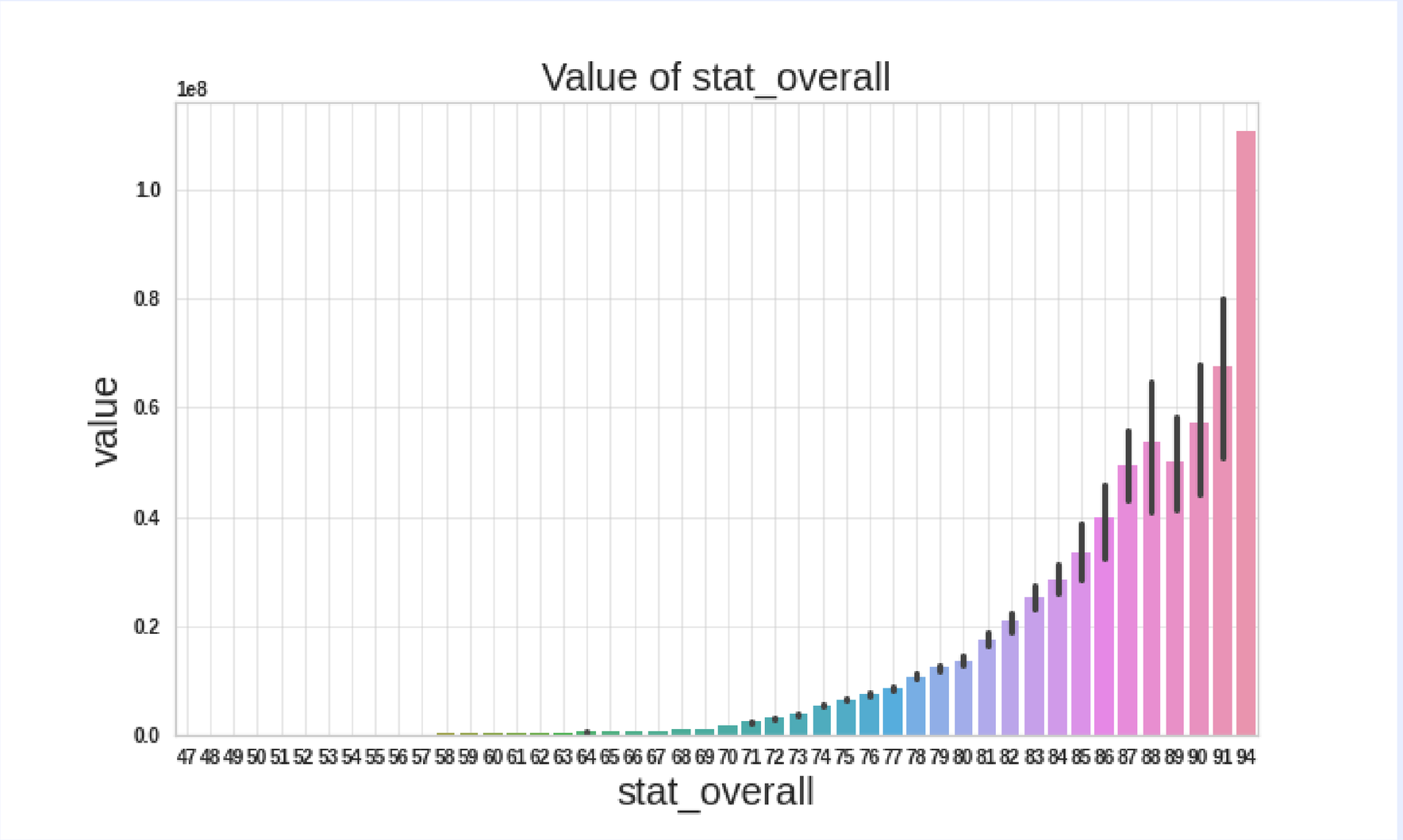
주발 별 선수 가치


Value of prefer_foot
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
value
left
right
prefer_foot

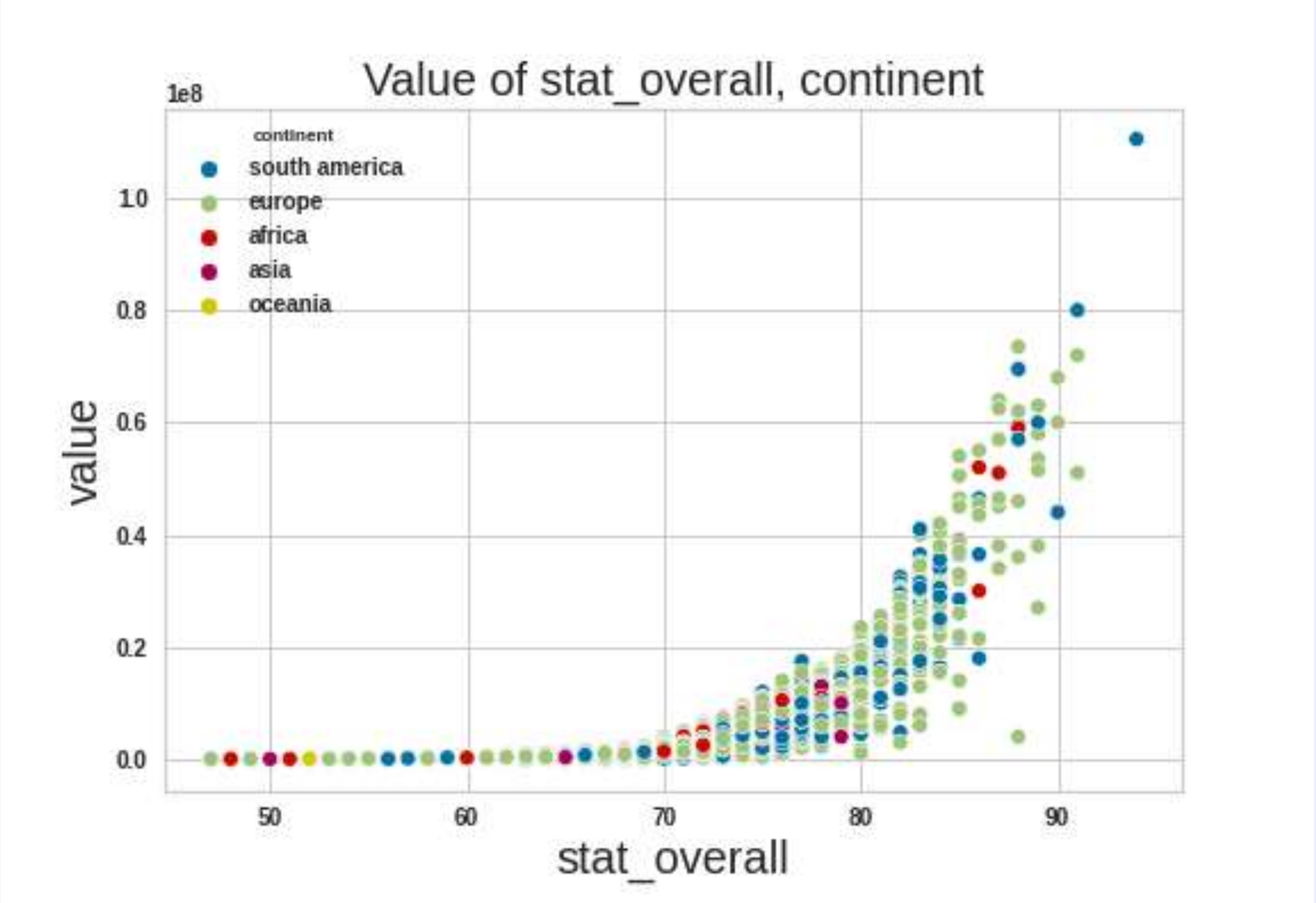
명성 별 선수 가치


Value of reputation
1e8
1.0
0.8
0.6
0.4
0.2
0.0
value
1.0
2.0
3.0
4.0
5.0
reputation

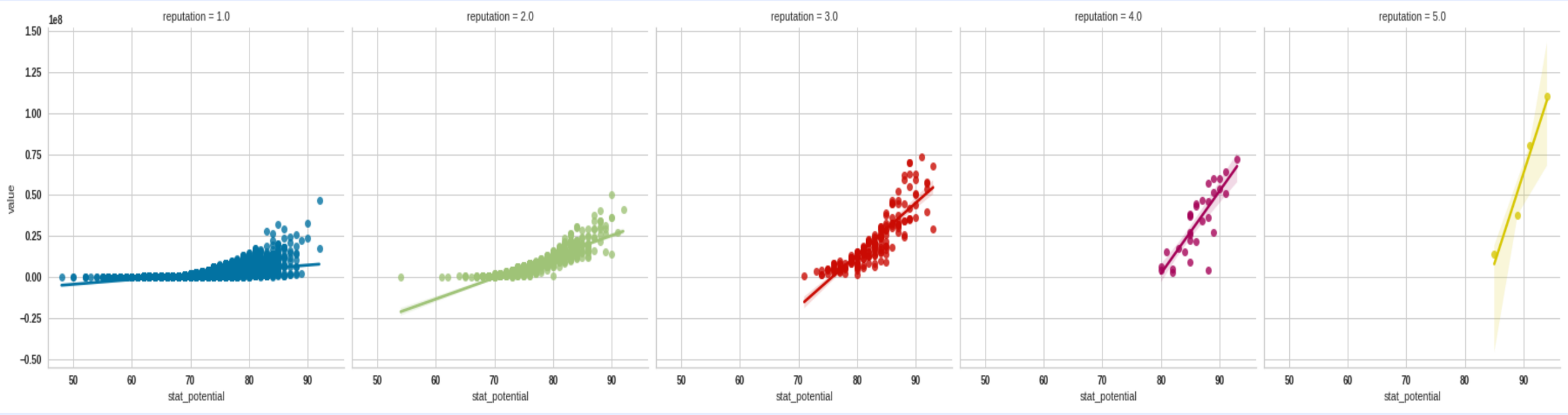
개인 스탯 별 선수 가치


Value of stat_overall
1e8
1.0
0.8
0.6
0.4
0.2
0.0
value
47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 94
stat_overall

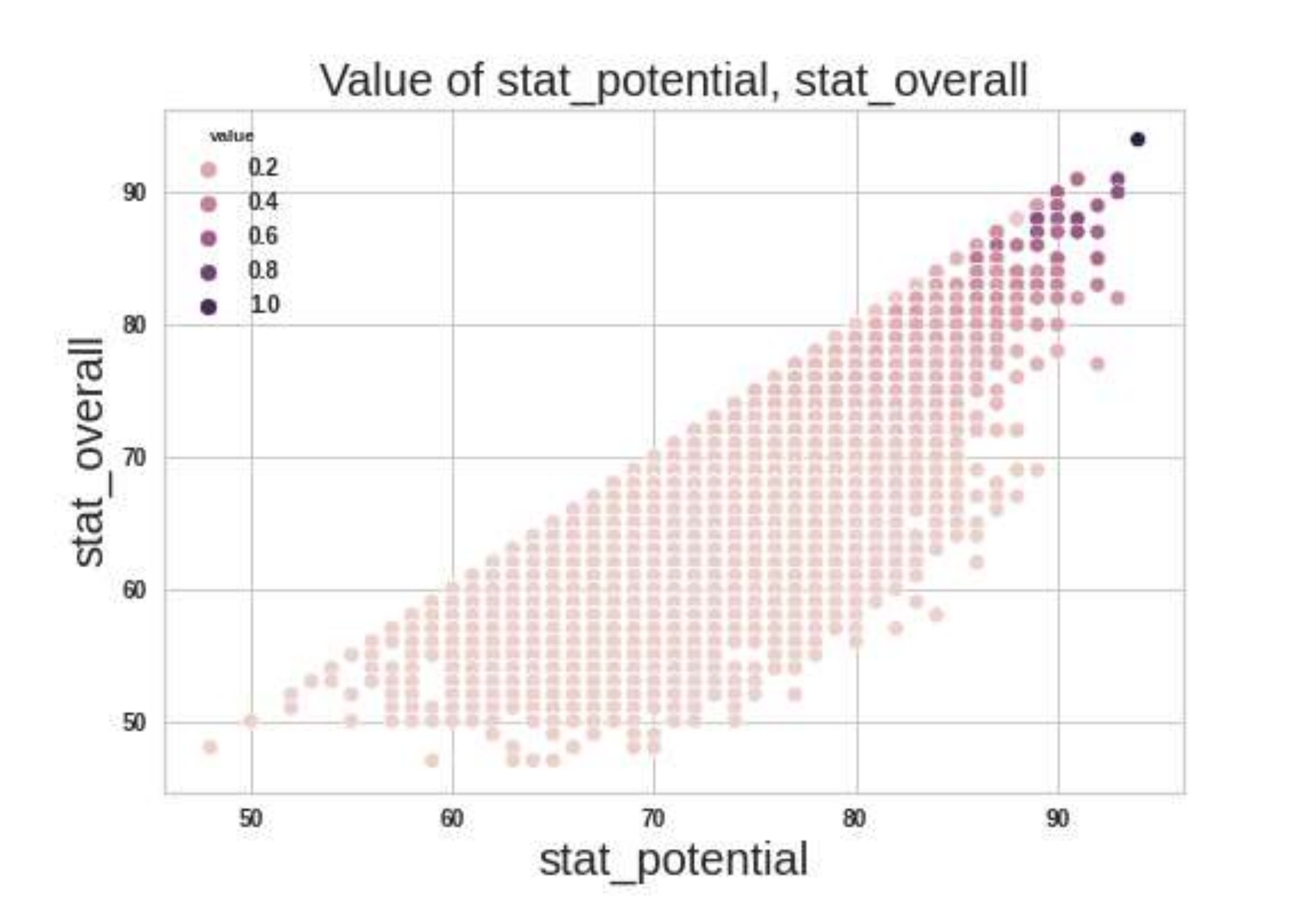
선수 스탯과 국가별 선수 가치



명성과 잠재능력 사이의 선수 가치



잠재 능력과 개인 스탯 사이의 선수 가치



선수 스킬 별 가치


Value of stat_skill_moves
1e7
2.5
2.0
1.5
1.0
0.5
0.0
value
1.0
2.0
3.0
4.0
5.0
stat_skill_moves

03

알고리즘 개발

- 회귀 성능 지표(RMSE)
- 축구 시장의 동향
- 산업의 잠재시장 파악

RMSE 회귀 모델 성능 지표

```
1  best=compare_models(sort='RMSE')
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
gbr	Gradient Boosting Regressor	2.372584e+05	6.188890e+11	7.588892e+05	0.9820	0.3105	0.2304	0.628
et	Extra Trees Regressor	1.788255e+05	8.348724e+11	8.641612e+05	0.9747	0.1100	0.0575	2.074
knn	K Neighbors Regressor	2.379928e+05	9.391368e+11	8.711445e+05	0.9749	0.1263	0.0884	0.082
rf	Random Forest Regressor	1.823307e+05	9.305258e+11	8.882605e+05	0.9728	0.1037	0.0577	2.003
lightgbm	Light Gradient Boosting Machine	1.880413e+05	1.221815e+12	1.006854e+06	0.9646	0.1518	0.0949	0.134
dt	Decision Tree Regressor	2.371025e+05	1.725184e+12	1.257910e+06	0.9480	0.1293	0.0626	0.044
ada	AdaBoost Regressor	1.740430e+06	4.582206e+12	2.129471e+06	0.8588	1.5708	5.1230	0.436
lasso	Lasso Regression	1.851095e+06	1.156727e+13	3.361485e+06	0.6608	1.3561	4.9596	0.202
ridge	Ridge Regression	1.854057e+06	1.156986e+13	3.361941e+06	0.6607	1.3607	4.9758	0.040
llar	Lasso Least Angle Regression	1.851109e+06	1.157079e+13	3.361991e+06	0.6607	1.3556	4.9595	0.626
br	Bayesian Ridge	1.853414e+06	1.157140e+13	3.362154e+06	0.6606	1.3631	4.9720	0.032
lr	Linear Regression	1.851543e+06	1.157738e+13	3.363002e+06	0.6604	1.3559	4.9598	0.728
omp	Orthogonal Matching Pursuit	1.880502e+06	1.256965e+13	3.514480e+06	0.6273	1.3457	4.9237	0.019
en	Elastic Net	2.113650e+06	1.494127e+13	3.816823e+06	0.5640	1.4750	6.4946	0.174
huber	Huber Regressor	1.706364e+06	1.898068e+13	4.303276e+06	0.4466	1.0339	2.1844	0.302
par	Passive Aggressive Regressor	2.008029e+06	2.935585e+13	5.367762e+06	0.1357	1.0277	1.5908	0.173
dummy	Dummy Regressor	3.081626e+06	3.375861e+13	5.767832e+06	0.0014	1.7502	5.9142	0.016

01 RMSE 란?

평균 제곱근 편차 또는 평균 제곱근 오차로서,
추정 값 또는 모델이 예측한 값과 실제 환경에서 관찰되는 값의 차이를 다룰 때 흔히 사용되는 척도 입니다.

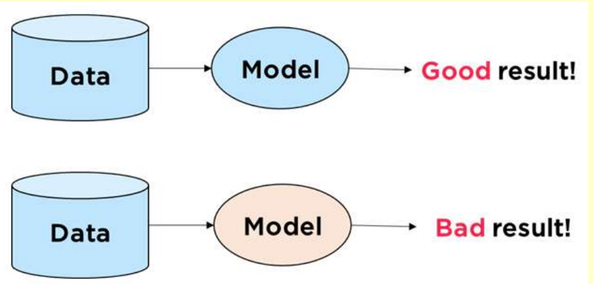
하이퍼 파라미터 튜닝

```
# 하이퍼파라미터 튜닝
tuned_gbr = tune_model(gbr, optimize = 'RMSE', n_iter = 3)
tuned_lightgbm = tune_model(lightgbm, optimize = 'RMSE', n_iter = 3)
tuned_knn = tune_model(knn, optimize = 'RMSE', n_iter = 3)
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	244199.1544	7.605253e+11	8.720810e+05	0.9791	0.1321	0.0866
1	224243.7640	1.324847e+12	1.151020e+06	0.9591	0.1127	0.0806
2	217186.4962	4.477635e+11	6.691513e+05	0.9828	0.1194	0.0815
3	223847.3562	5.533085e+11	7.438471e+05	0.9770	0.1246	0.0824
4	180533.3617	1.778199e+11	4.216870e+05	0.9918	0.1283	0.0895
5	214004.5260	4.506237e+11	6.712851e+05	0.9872	0.1288	0.0849
6	252256.8436	9.433645e+11	9.712695e+05	0.9690	0.1202	0.0813
7	234012.8628	5.695207e+11	7.546659e+05	0.9861	0.1180	0.0800
8	316614.7881	4.724456e+12	2.173581e+06	0.9041	0.1260	0.0843
9	214765.3587	5.232783e+11	7.233798e+05	0.9871	0.1116	0.0766
Mean	232166.4512	1.047551e+12	9.151968e+05	0.9723	0.1222	0.0828
Std	33672.4598	1.261345e+12	4.582200e+05	0.0246	0.0066	0.0035

02 하이퍼 파라미터

사용자의 입력값, 혹은 설정 가능한 옵션입니다.
모든 데이터와 문제에 대해 가장 좋은 하이퍼 파라미터 값이 있으면 좋겠으나, 데이터에 따라 좋은 하이퍼 파라미터의 값은 다릅니다.



앙상블 학습

```
# 앙상블

blender = blend_models(estimator_list=compare_models(n_select=3
                                sort='RMSE'))
```

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	190420.2136	4.433436e+11	6.658405e+05	0.9878	0.1848	0.1051
1	174058.2397	6.008064e+11	7.751170e+05	0.9814	0.1374	0.0834
2	189711.1296	3.121025e+11	5.586614e+05	0.9880	0.1798	0.1018
3	201537.6036	8.338475e+11	9.131525e+05	0.9654	0.1608	0.1046
4	150936.6570	1.398294e+11	3.739377e+05	0.9936	0.1748	0.1083
5	173958.6480	3.409053e+11	5.838710e+05	0.9903	0.1270	0.0918
6	214821.9587	6.938727e+11	8.329902e+05	0.9772	0.1778	0.1022
7	184867.0927	5.325901e+11	7.297877e+05	0.9870	0.2220	0.0943
8	244449.6231	2.335068e+12	1.528093e+06	0.9526	0.1549	0.0917
9	165746.4269	3.228595e+11	5.682072e+05	0.9921	0.2406	0.1217
Mean	189050.7593	6.555224e+11	7.529658e+05	0.9815	0.1760	0.1005
Std	25213.5853	5.925195e+11	2.975986e+05	0.0125	0.0331	0.0102

03 앙상블 학습

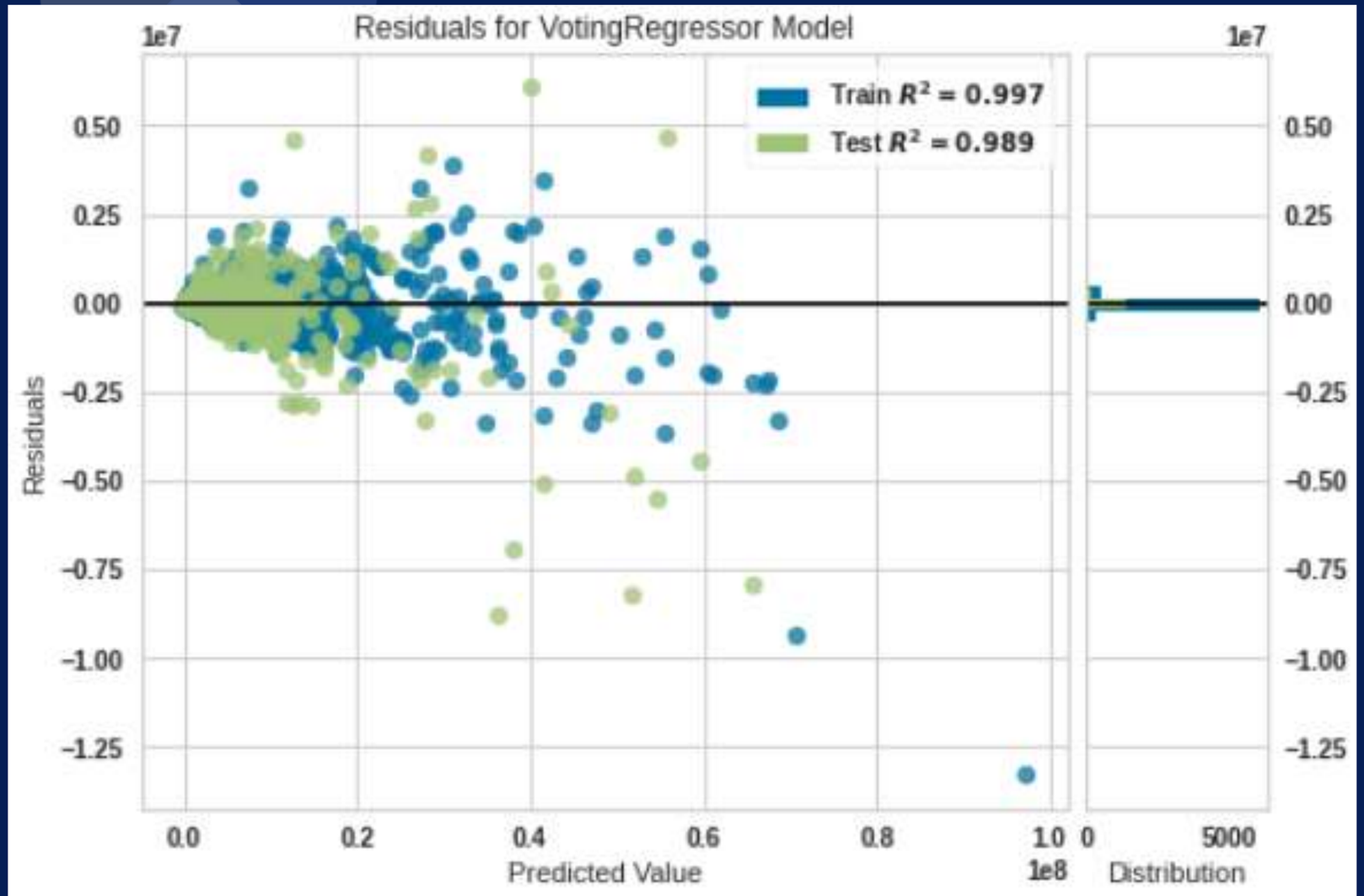
앙상블 학습(Ensemble Learning)은 여러 개의 분류기를 생성하고, 그 예측을 결합함으로써 보다 정확한 예측을 도출하는 기법을 말합니다.

강력한 하나의 모델을 사용하는 대신 보다 약한 모델 여러개를 조합하여 더 정확한 예측에 도움을 주는 방식입니다.

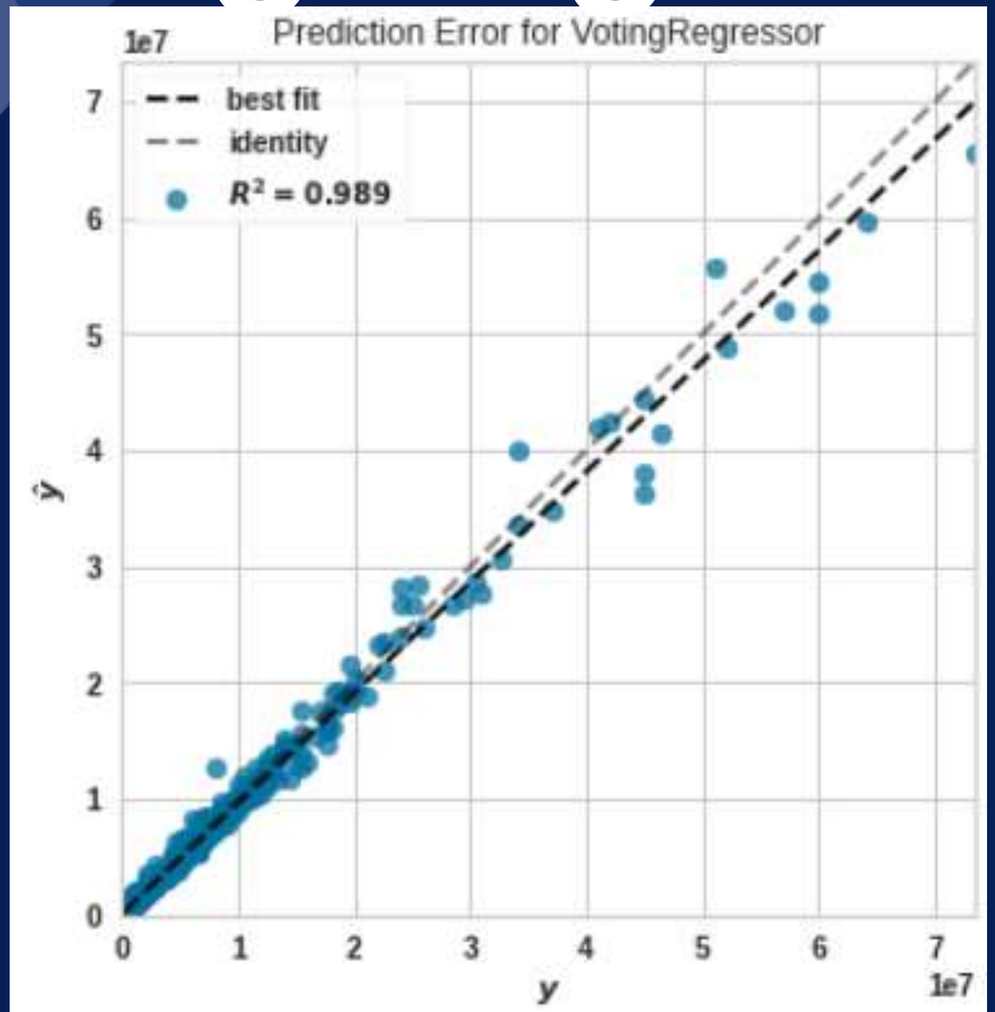
XGBoost, LightGBM과 같은 앙상블 알고리즘이 머신러닝의 선도 알고리즘으로 인기를 모으고 있습니다.

앙상블 학습은 일반적으로 보팅(Votting), 배깅(Bagging), 부스팅(Boosting) 세 가지의 유형으로 나눌 수 있습니다.

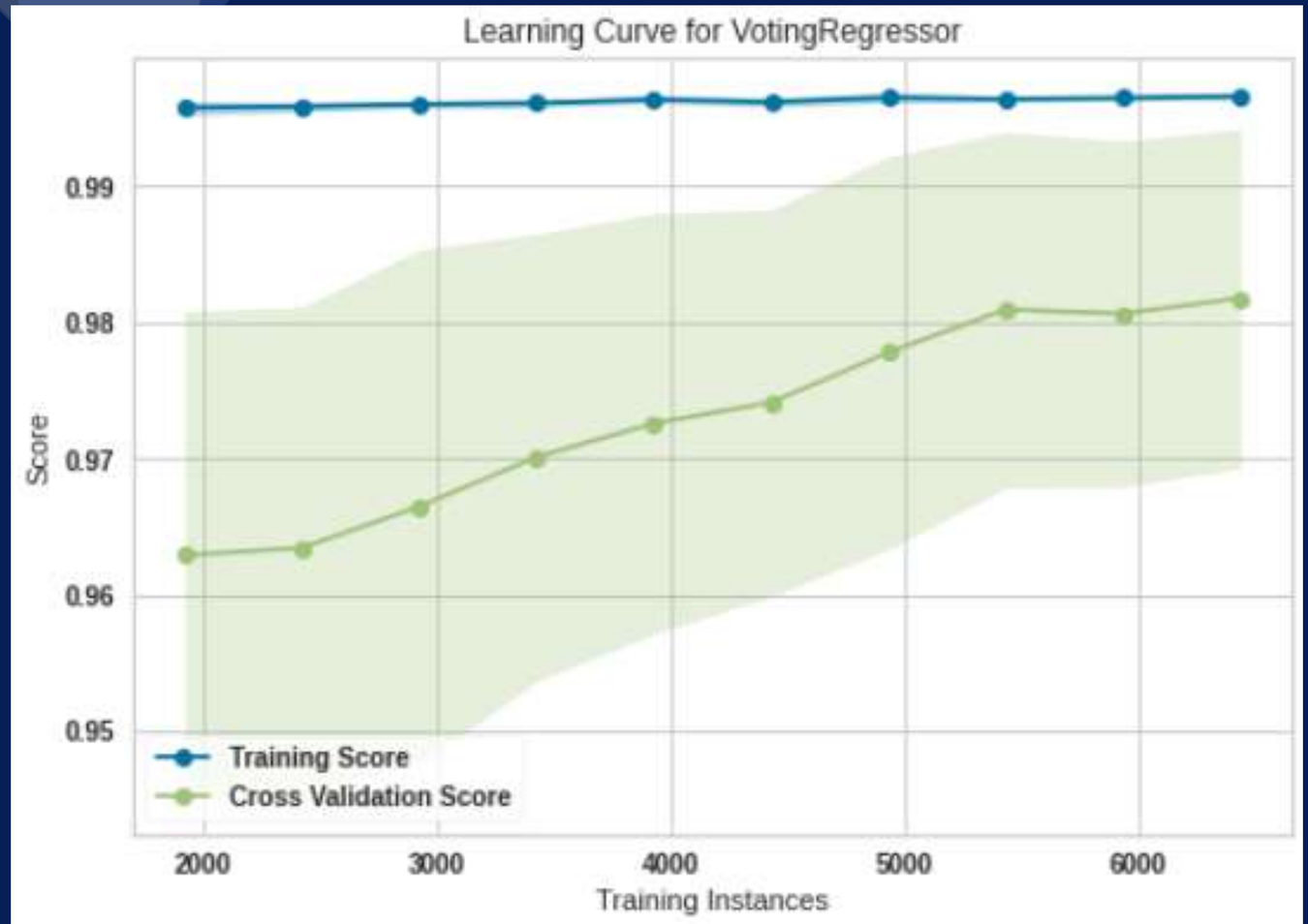
Residuals for VotingRegressor Model



Prediction Error for VotingRegressor



Learning Curve for VotingRegressor



선수 이적 시장의 동향


PREMIER LEAGUE CLUBS
2021 SUMMER TRANSFER WINDOW
RANKED
1 Manchester United
2 Chelsea
3 Tottenham
4 West Ham
5 Aston Villa
6 Crystal Palace
7 Brighton
8 Leicester City
9 Manchester City
10 Southampton
11 Wolves
12 Norwich City
13 Brentford
14 Watford
15 Burnley
16 Leeds United
17 Everton
18 Arsenal
19 Liverpool
20 Newcastle
90 MIN


HIGHEST SPENDING LEAGUES
2021/2022 - SO FAR
TOTAL SPENDING
1 PREMIER LEAGUE €1.02BN
2 SERIE A €446.6M
3 BUNDESLIGA €340.3M
4 LIGUE 1 €301.9M
5 LALIGA €152.3M
6 PREMIER LIGA €90.6M
7 JUPILER PRO LEAGUE €74.3M
8 LIGA PORTUGAL BWIN €71.6M
9 SAUDI PROFESSIONAL LEAGUE €50.2M
10 SERIE B €50.0M

산업의 잠재시장 파악


EA SPORTS
FIFA ONLINE 4
FIFA


20:46 PM
SPORTS BETTING LIVE
MATCH 1 BET NOW
MATCH 2 BET NOW
MATCH 3 BET NOW


THE AGENTS

04

프로세스 결과

- 활용 방안 / 기대 효과
- 개선 방향

활용 방안 / 기대효과


InterFootball
gettyimagesKorea


DISABLED
9
37
TURK


WHAT
MAKES
A GOOD TRADE?

개선 방향

노트북 이슈로 인해 좋은 모델들을
다양하게 앙상블 학습 하지
못하였습니다.

훨씬 많은 양의 데이터가 필요합니다.