Maximizing Roster Efficiency in the MLS

Oliver Daboo, Otis Birnbaum
Colby College, Williams College
External Advisor: Daniel Wicker, Charlotte FC

Motivation

- MLS is the only major professional soccer league in the world with a salary cap and has a lot parity.
- Lower budget teams have a chance to compete if they find players who contribute a lot relative to their wage

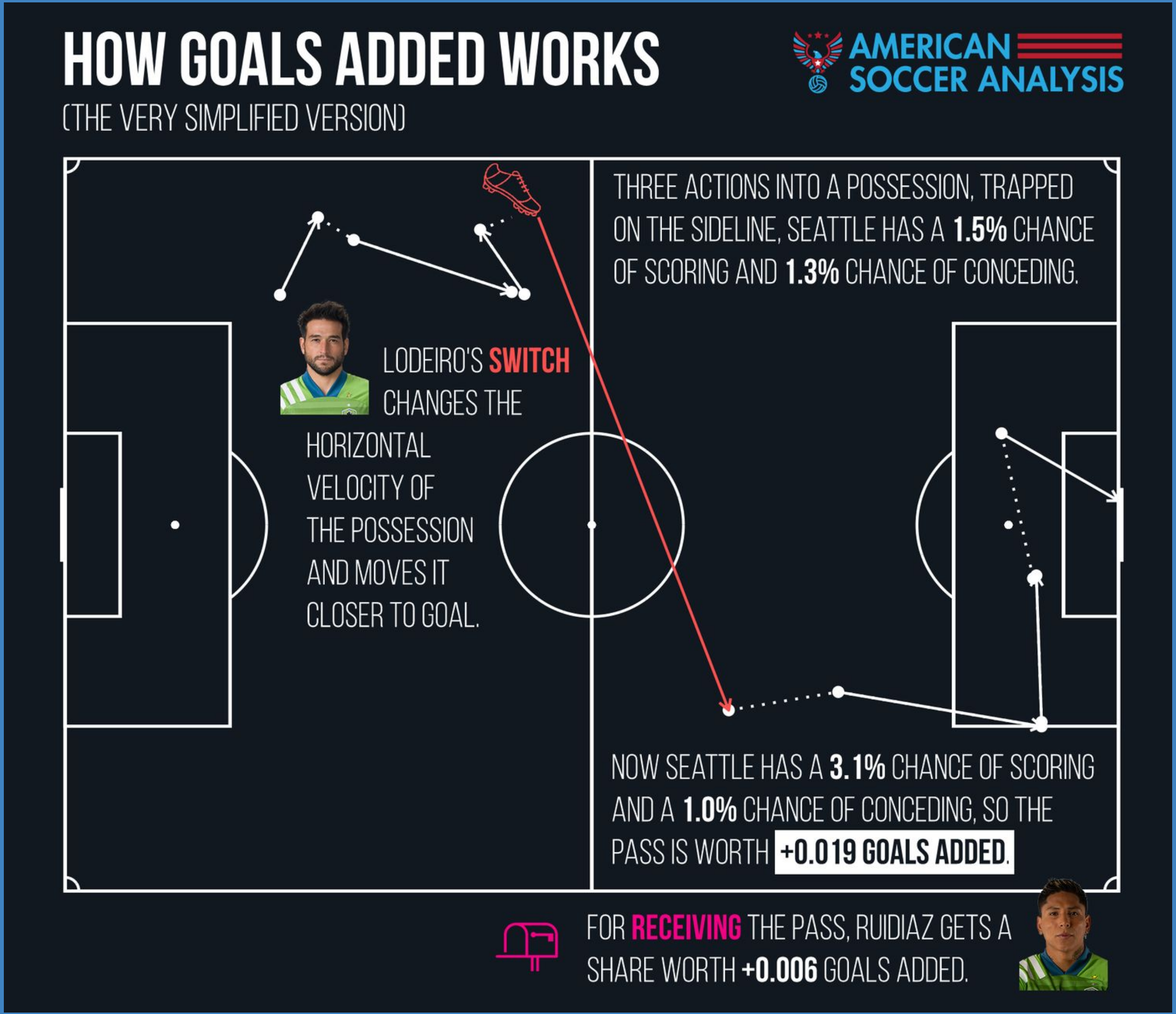
Main Question

- How should MLS teams construct their rosters and salaries spreads based on a players position, age, and nationality?

Data

- MLS Seasons 2021-2024 from **American Soccer Analysis**
- Used xG Difference to evaluate team performance
- Used Goals Added, a value added metric, to evaluate player performance
- Created a goals added per 90 minutes played per $10k (ga_per_90_per_10k) metric to find undervalued players
- Assigned players one of six Region Groups based on their nationality.

Player, Year	Club	Position	Age	Nationality	G+ Per 90 min.
Lionel Messi, 2024	Inter Miami	W	37	Argentina	.572
Cucho Hernandez, 2024	Columbus Crew	ST	25	Colombia	.471
Adam Buksa, 2021	New England Revolution	ST	25	Poland	.455
Riqui Puig, 2024	LA Galaxy	CM	25	Spain	.455
Cucho Hernandez, 2023	Columbus Crew	ST	24	Colombia	.451

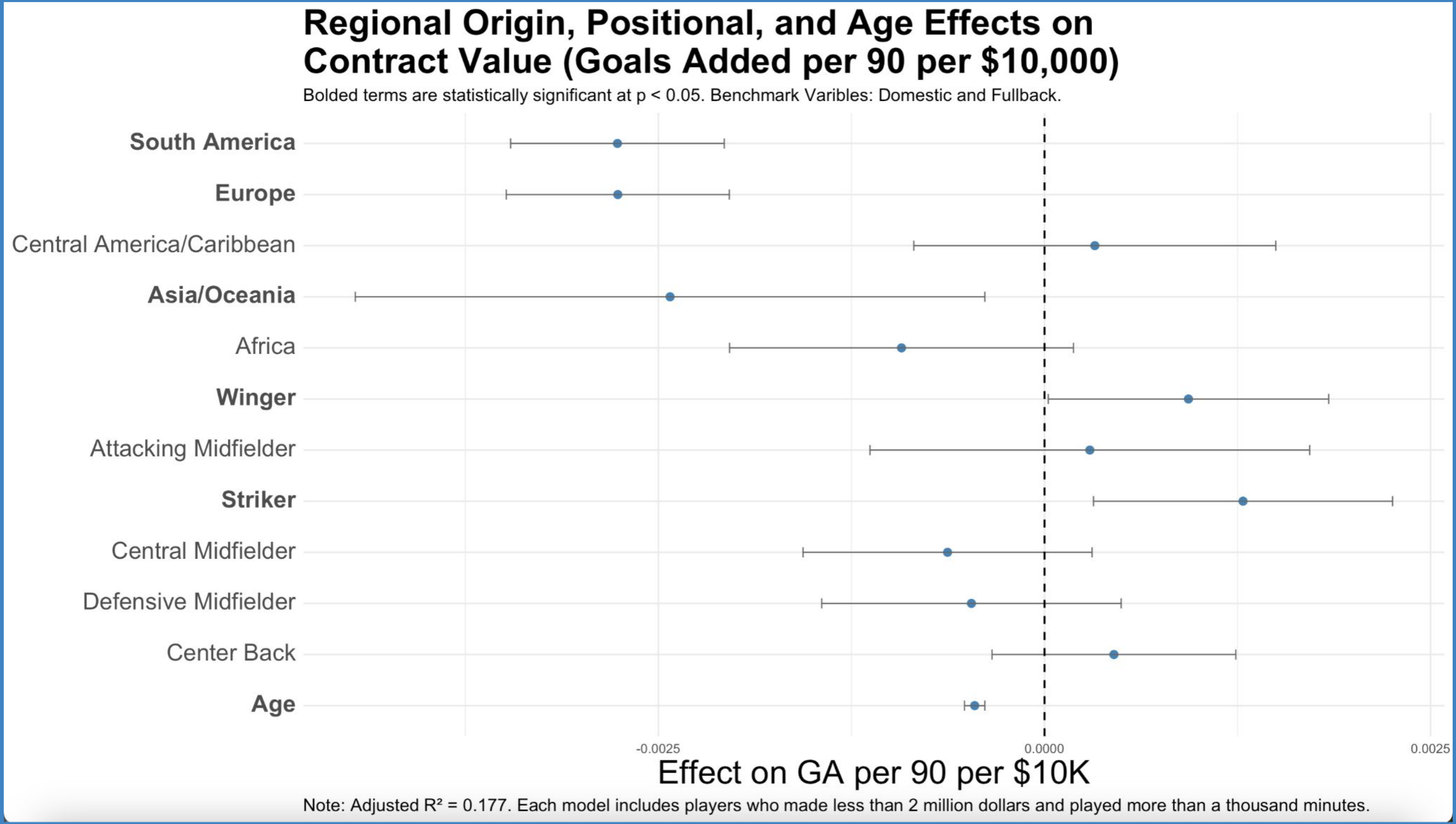

HOW GOALS ADDED WORKS
(THE VERY SIMPLIFIED VERSION)
AMERICAN SOCCER ANALYSIS
THREE ACTIONS INTO A POSSESSION, TRAPPED ON THE SIDELINE, SEATTLE HAS A 1.5% CHANCE OF SCORING AND 1.3% CHANCE OF CONCEDING.
LODEIRO'S SWITCH CHANGES THE HORIZONTAL VELOCITY OF THE POSSESSION AND MOVES IT CLOSER TO GOAL.
NOW SEATTLE HAS A 3.1% CHANCE OF SCORING AND A 1.0% CHANCE OF CONCEDING, SO THE PASS IS WORTH +0.019 GOALS ADDED.
FOR RECEIVING THE PASS, RUIDIAZ GETS A SHARE WORTH +0.006 GOALS ADDED.

Methods

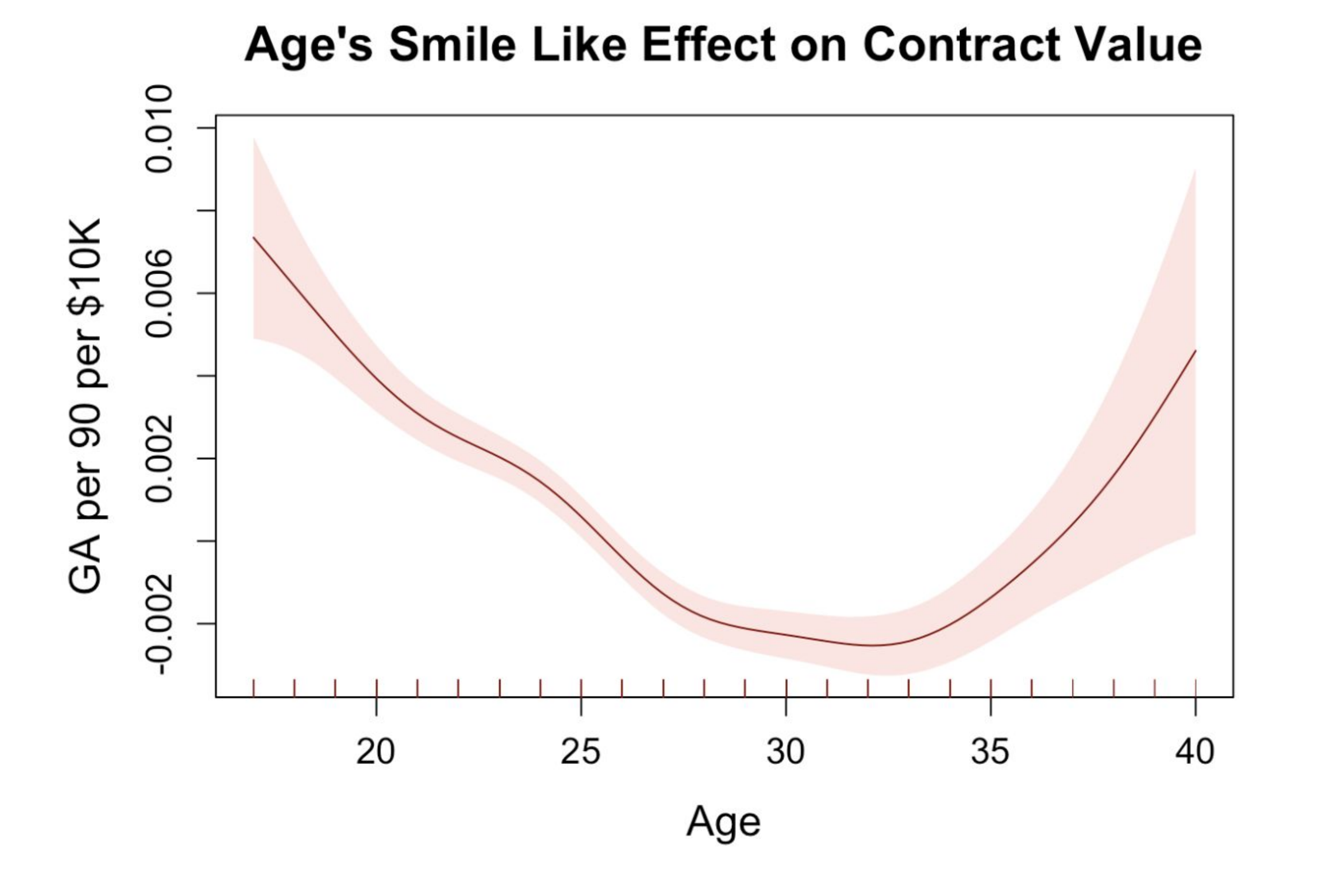
Modeling

- Filtered data by players who played more than 1,000 minutes and made less than $2 million
- Fit a linear model to predict ga_per_90_per_10k based on player ages, positions, and nationalities
- Used Generalized Additive Models to identify non-linear relationships between salary and G+ and age and G+ per 90 per $10,000

Player, Year	Club	Position	Age	Nationality	GA_Per_90_Per_10k
Patrick Agyemang, 2024	Charlotte FC	ST	23	USA	.0475
Célio Pompeu, 2023	St. Louis City	W	23	Brazil	.0462
Tani Oluwaseyi, 2024	Minnesota United	ST	24	Canada	.0411
Jacob Murrell, 2024	D.C. United	ST	20	USA	.0369
Fredy Montero, 2021	Seattle Sounders	ST	34	Colombia	.0350



Regional Origin, Positional, and Age Effects on
Contract Value (Goals Added per 90 per $10,000)
Bolded terms are statistically significant at p < 0.05. Benchmark Varibles: Domestic and Fullback.
South America
Europe
Central America/Caribbean
Asia/Oceania
Africa
Winger
Attacking Midfielder
Striker
Central Midfielder
Defensive Midfielder
Center Back
Age
-0.0025
0.0000
0.0025
Effect on GA per 90 per $10K
Note: Adjusted R² = 0.177. Each model includes players who made less than 2 million dollars and played more than a thousand minutes.



Age's Smile Like Effect on Contract Value
GA per 90 per $10K
0.010
0.006
0.002
-0.002
20
25
30
35
40
Age

There is a significant relationship for the more that teams spend on their offense proportionally to their defense, the less efficient they are

Discussion

Defense wins Championships (Maybe)

- While our team level models found that teams spending a higher proportion of their salary on defense did better, our G+ player level models found Strikers and Wingers to be slightly undervalued
- We lean towards defenders being undervalued given G+ bias towards attackers and that defenders are cheaper

Europeans, South Americans, Asian/Oceanic and Older Players are Overvalued (until they're really old)

- Being European, South American, Asian/Oceanic or being older all have negative effects on Goals Added per 90 minutes per 10,000 dollars spent

So what should MLS clubs do?

Younger, Domestic Players are Undervalued. Find Them.

1. Academies

2. SuperDraft

Case Study: Patrick Agyemang Drafted in 2023:

In 2024…
Salary: $71,401
Goals Scored: 10

In 2025…

Ryan Tolmich | Jul 15, 2025 18:50-04:00 0

Transfers | USA | P. Agyemang | Charlotte FC | Major League Soccer

USMNT striker Patrick Agyemang completes $8 million transfer to Derby County as Charlotte FC net club record fee

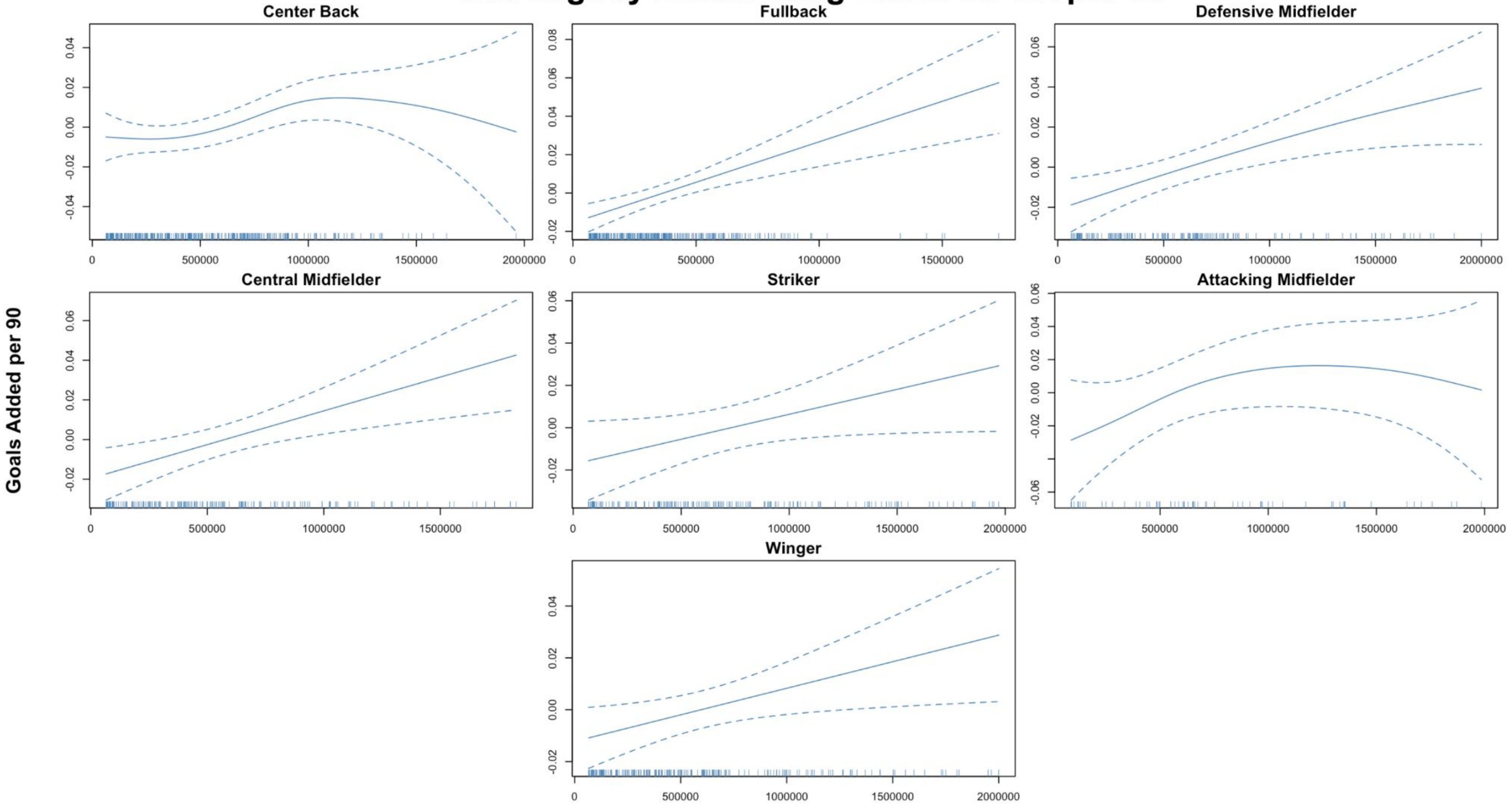
Coming off the USMNT's Gold Cup run, the striker will join the Championship club after breaking out in MLS and with the national team

Thank You!

Questions?

Appendix

Salary Has a Positive Linear, but Slightly Diminishing Effect on GA per 90



Guaranteed Compensation

Note: Model includes players who made less than 2 million dollars and played more than a thousand minutes.

Table 10. Linear Model Coefficients Predicting Guaranteed Compensation

Term	Estimate	SE	t	p	95% CI Low	95% CI High
(Intercept)	-556924.45	63942.642	-8.70975	0.00000	-682350.462	-431498.44
age	32072.86	2326.887	13.78359	0.00000	27508.579	36637.14
general_positionCB	90841.52	27886.941	3.25749	0.00115	36140.187	145542.86
general_positionCM	204584.75	33064.797	6.18739	0.00000	139726.844	269442.65
general_positionDM	216203.72	34260.971	6.31050	0.00000	148999.476	283407.97
general_positionST	323932.68	34220.838	9.46595	0.00000	256807.153	391058.20
general_positionW	232919.43	32080.337	7.26050	0.00000	169992.587	295846.28
general_positionAM	354967.26	50307.313	7.05598	0.00000	256287.483	453647.04
region_groupSouth America	179660.95	24445.009	7.34960	0.00000	131711.101	227610.80
region_groupCentral America/Caribbean	-36228.70	41423.290	-0.87460	0.38193	-117482.112	45024.72
region_groupEurope	248329.42	25515.469	9.73250	0.00000	198279.820	298379.02
region_groupAfrica	92337.88	39352.968	2.34640	0.01908	15145.482	169530.28
region_groupAsia/Oceania	138940.39	72030.606	1.92891	0.05393	-2350.481	280231.26

Table 11. Model Fit Statistics for Compensation Model

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance
0.25863	0.25272	363048.9	43.80985	0	12	-21609.74	43247.49	43322.06	1.986294e+14

Table 12. Linear Model Coefficients Predicting Guaranteed Compensation per 90 minutes

Term	Estimate	SE	t	p	95% CI Low	95% CI High
(Intercept)	0.11462	0.00896	12.78784	0.00000	0.09704	0.13221
age	0.00036	0.00033	1.10531	0.26920	-0.00028	0.00100
general_positionCB	0.01104	0.00391	2.82340	0.00481	0.00337	0.01871
general_positionCM	0.02092	0.00464	4.51241	0.00001	0.01182	0.03001
general_positionDM	0.01349	0.00480	2.80803	0.00505	0.00407	0.02291
general_positionST	0.10028	0.00480	20.90496	0.00000	0.09087	0.10969
general_positionW	0.06296	0.00450	14.00058	0.00000	0.05414	0.07178
general_positionAM	0.08546	0.00705	12.11784	0.00000	0.07162	0.09929
region_groupSouth America	0.02050	0.00343	5.98336	0.00000	0.01378	0.02723
region_groupCentral America/Caribbean	0.01178	0.00581	2.02833	0.04270	0.00039	0.02317
region_groupEurope	0.01319	0.00358	3.68892	0.00023	0.00618	0.02021
region_groupAfrica	0.00045	0.00552	0.08154	0.93502	-0.01037	0.01127
region_groupAsia/Oceania	0.02227	0.01010	2.20572	0.02755	0.00247	0.04208

Table 13. Model Fit Statistics for Predicting Guaranteed Compensation per 90 minutes

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.r
0.3408	0.33555	0.05089	64.92593	0	12	2376.352	-4724.704	-4650.134	3.90325	

Table 14. Linear Model Coefficients Predicting Guaranteed Compensation per 90 minutes per $10k

Term	Estimate	SE	t	p	95% CI Low	95% CI High
(Intercept)	0.01859	0.00092	20.15950	0.00000	0.01678	0.02040
age	-0.00045	0.00003	-13.46425	0.00000	-0.00052	-0.00039
general_positionCB	0.00045	0.00040	1.11769	0.26388	-0.00034	0.00124
general_positionCM	-0.00063	0.00048	-1.31558	0.18851	-0.00156	0.00031
general_positionDM	-0.00047	0.00049	-0.95713	0.33866	-0.00144	0.00050
general_positionST	0.00129	0.00049	2.60438	0.00929	0.00032	0.00225
general_positionW	0.00093	0.00046	2.01465	0.04412	0.00002	0.00184
general_positionAM	0.00029	0.00073	0.40469	0.68577	-0.00113	0.00172
region_groupSouth America	-0.00276	0.00035	-7.84113	0.00000	-0.00346	-0.00207
region_groupCentral America/Caribbean	0.00033	0.00060	0.54555	0.58546	-0.00085	0.00150
region_groupEurope	-0.00276	0.00037	-7.50566	0.00000	-0.00348	-0.00204
region_groupAfrica	-0.00093	0.00057	-1.63059	0.10319	-0.00204	0.00019
region_groupAsia/Oceania	-0.00242	0.00104	-2.33306	0.01978	-0.00446	-0.00039

Table 15. Model Fit Statistics for Predicting Guaranteed Compensation per 90 minutes per $1

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.r
0.18316	0.17665	0.00524	28.15897	0	12	5832.871	-11637.74	-11563.17	0.04133	

Methods

Team Modeling

- Split up each teams roster into the percentage of their total salary they pay different slots of players
  - Split up into 6 groups of 3 of the top 18 paid players
- Predicted xG difference using linear regression, regularization models, and XGBoost
- Trained the models on data from 2021-2023, tested it on 2024 and used bootstrapped cross-validation to evaluate model performances

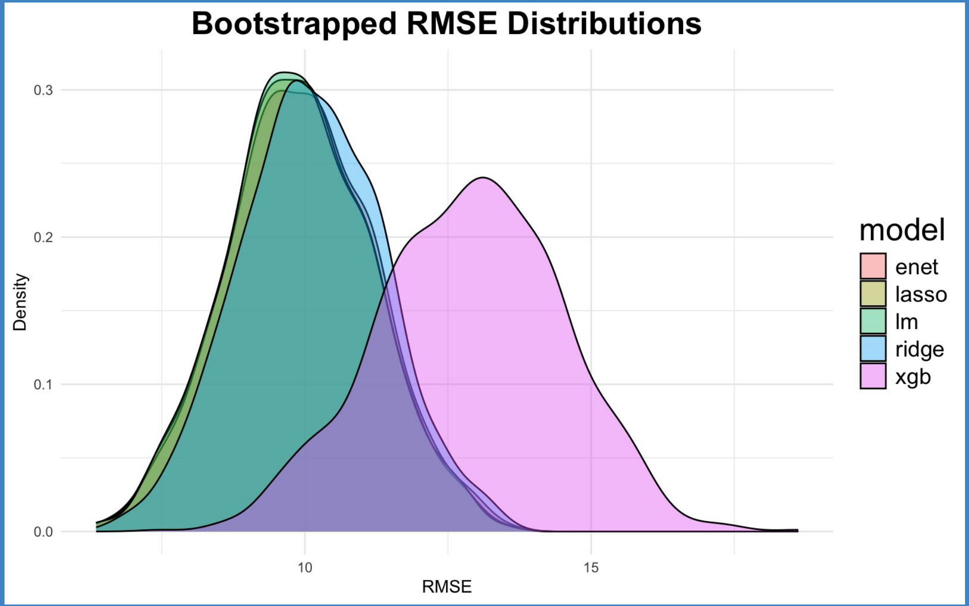


Table 3. Linear Model Summary Predicting each Team's xG Difference by their Average Salary

Term	Estimate	SE	t	p
(Intercept)	0.90351	7.65199	0.11808	0.90688
avg_guaranteed_compensation	0.00000	0.00001	-0.12225	0.90361