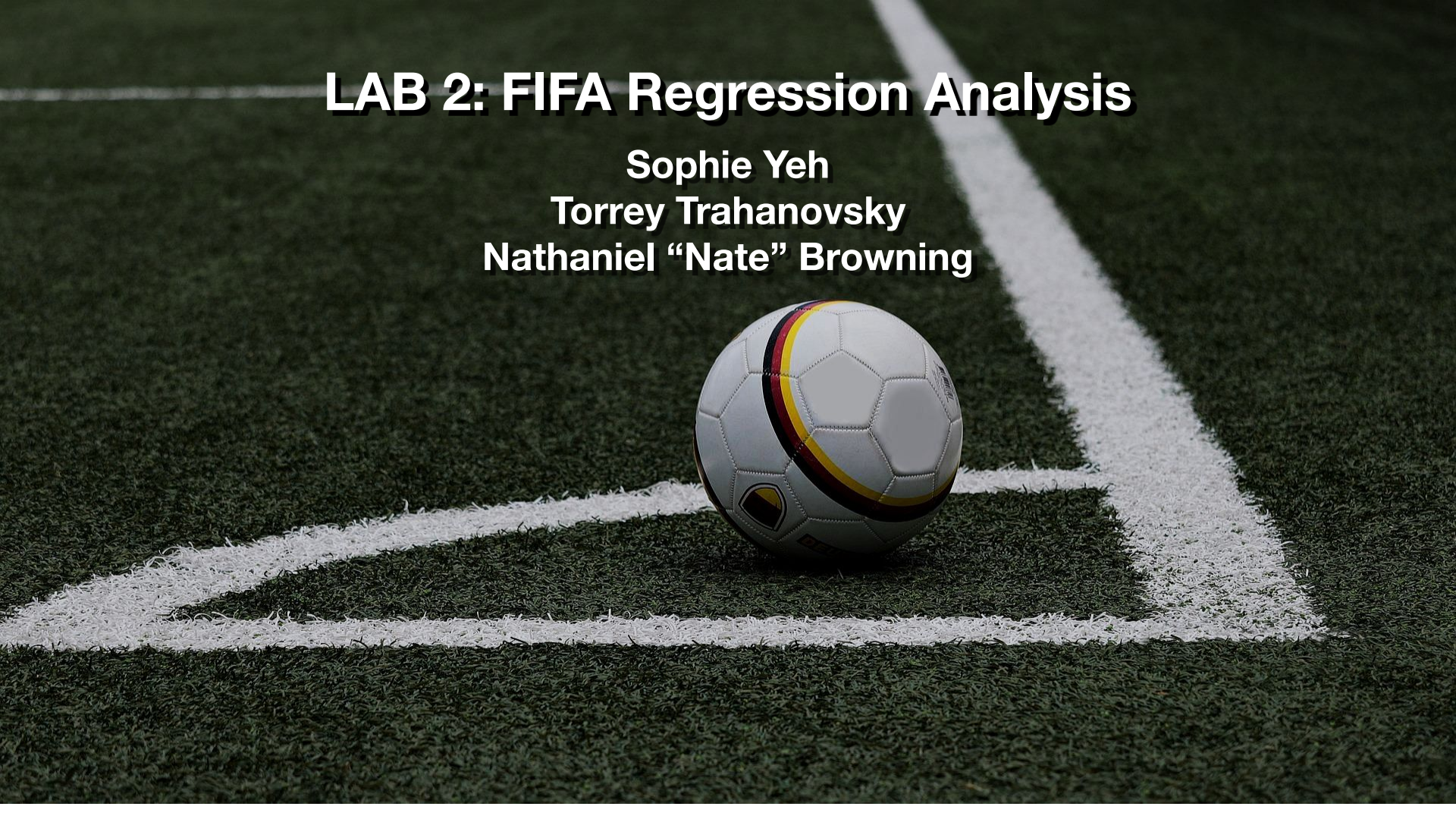
LAB 2: FIFA Regression Analysis

Sophie Yeh
Torrey Trahanovsky
Nathaniel “Nate” Browning



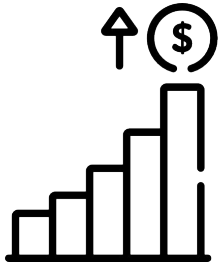
Intro

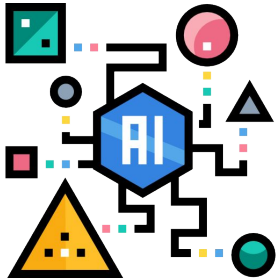
1

2

3

Research Question:
In 2022, how does a goal keeper's performance, background, and physical attributes affect their market value in football?





International Federation of Association Football (FIFA)

Research Question Created with measurement goals, model specs, and EDA deciding models

Use model to optimize "product" and provide tooling to FIFA

Research Question Addressing Process

1. Measurement Goals
2. Testing the research question via a null-hypothesis and various tests/stargazers
3. Design:
    a. **Primary design:** Causal and explanatory:
        i. Attempting to analyze collinearity between various x variables and see how they might indicate whether the y variable or other variables are affected by variances.
    b. **Secondary Design:** Exploratory:
        i. In which during exploratory data analysis, various graphs, tables, stargazer, tests, and other visualization techniques are used to understand the data.
4. Minimal bias models that increase in efficiency as developed to a final model.

Data

- **Limited Scope:** Goalkeepers
  - Nans are eliminated, data is cleaned, ready to EDA.

- **Related:** Outcome, Potential, Value, Skills
  - Outcome is made up of various other dataset values
  - Skills encompass other attributes
  - Potential is closely related to outcomes underlying values
  - Value already accounts for some pre-existing data

- **Peculiar yet Useful:** Special, Wage, Contract Years, Reputation
  - Special is a string of #’s yet indicates various attributes
  - Wage is monthly making conclusions on pay clearer
  - Contract_Years is custom and calculates years remaining
  - Reputation is noteworthy and potentially needs to contrast with nationality

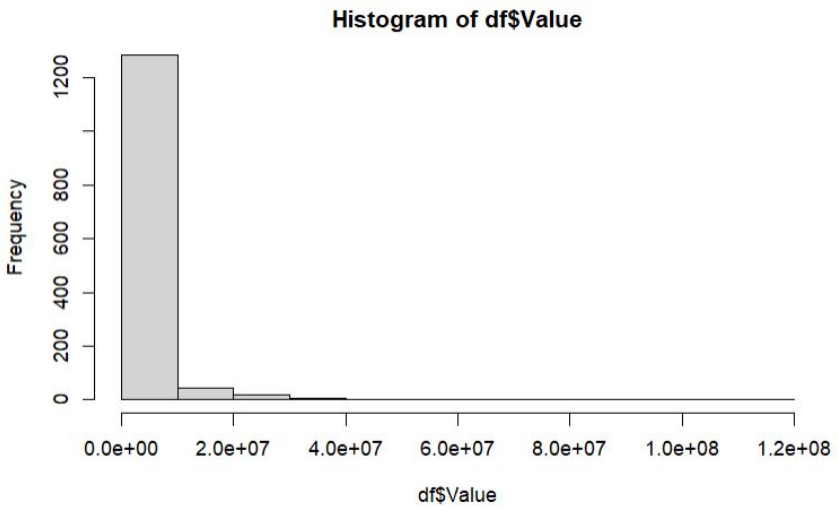
Key Covariates

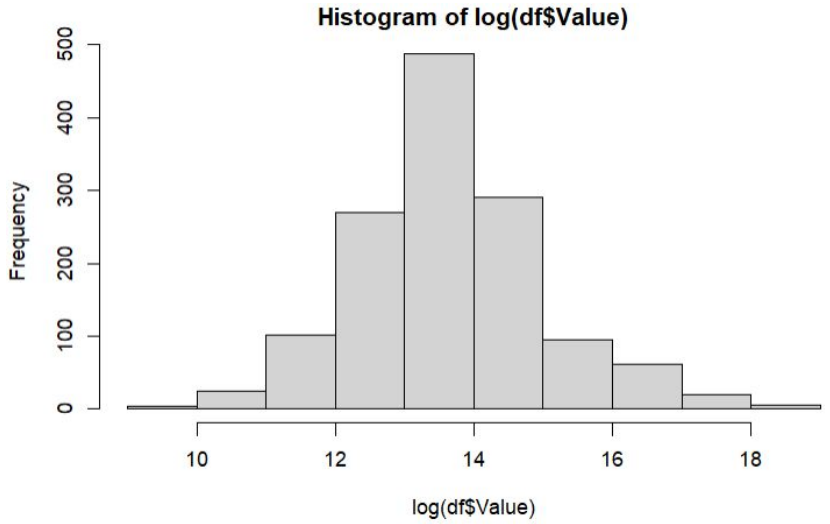
The dataset provided includes all players included in the FIFA 2022 video game including data on their physical description, position, and ratings of skill. The following covariates were considered key to our study and are included below:

- **Age:** The age of the player in the respective row
- **Height:** The height of the player in centimeters
- **Weight:** The weight of the player in kilograms
- **Special:** A rating metric to measure skills of a player
- **Value:** The market value for trading of a player measured in Euros

Covariate Transformations

- Age covariate has been made a quadratic to avoid a non-linear correlation to the success measure
- Value covariate has been logged to adjust for normality.





Large Sample Assumptions

- Two Large Sample Assumptions:
  - I.I.D. Data
  - A unique BLP exists

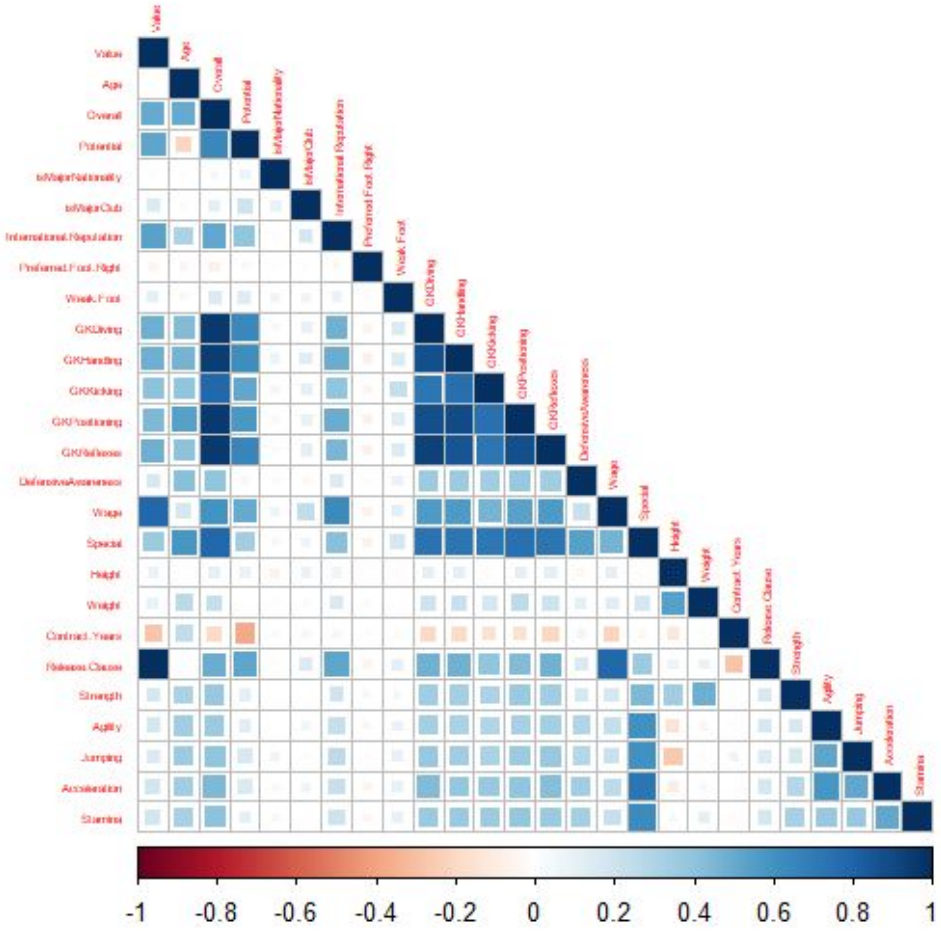
	Value	Age	International.Reputation	Contract.Years	Height	Weight
Value	1.936189e-17	8.966355e-12	-1.572446e-10	2.042353e-11	2.466207e-12	-1.117534e-12
Age	8.966355e-12	4.522965e-05	-9.757234e-05	-4.026555e-05	3.951699e-06	-7.548542e-06
International.Reputation	-1.572446e-10	-9.757234e-05	5.405479e-03	-8.134411e-05	-7.173825e-07	-1.348264e-05
Contract.Years	2.042353e-11	-4.026555e-05	-8.134411e-05	5.768788e-04	4.410155e-06	-3.079784e-06
Height	2.466207e-12	3.951699e-06	-7.173825e-07	4.410155e-06	6.341614e-06	-1.171131e-05
Weight	-1.117534e-12	-7.548542e-06	-1.348264e-05	-3.079784e-06	-1.171131e-05	3.165123e-05
Special	-5.678725e-13	-1.463929e-06	-1.948137e-06	1.465920e-06	-3.826358e-07	9.712513e-08
Weak.Foot	-5.408108e-13	2.316212e-05	-2.791212e-05	-1.950185e-05	-1.319097e-05	1.812295e-06
Agility	-1.526464e-13	-1.008096e-06	1.088026e-06	-1.923391e-06	-4.854470e-07	2.030005e-06
Strength	-2.548573e-13	1.652363e-07	5.427484e-06	-8.117177e-07	1.038322e-06	-6.484298e-06
Jumping	1.369207e-12	1.031512e-06	-1.455716e-05	-5.745417e-06	-1.714009e-07	4.357575e-08
Acceleration	1.786416e-12	4.589359e-06	8.835660e-06	-2.748215e-06	1.795620e-06	-5.071582e-08
Stamina	6.430750e-13	1.609595e-06	1.192106e-05	-1.382219e-06	3.352939e-07	8.166150e-07

	Special	Weak.Foot	Agility	Strength	Jumping	Acceleration	Stamina
Value	-5.678725e-13	-5.408108e-13	-1.526464e-13	-2.548573e-13	1.369207e-12	1.786416e-12	6.430750e-13
Age	-1.463929e-06	2.316212e-05	-1.008096e-06	1.652363e-07	1.031512e-06	4.589359e-06	1.609595e-06
International.Reputation	-1.948137e-06	-2.791212e-05	1.088026e-06	5.427484e-06	-1.455716e-05	8.835660e-06	1.192106e-05
Contract.Years	1.465920e-06	-1.950185e-05	-1.923391e-06	-8.117177e-07	-5.745417e-06	-2.748215e-06	-1.382219e-06
Height	-3.826358e-07	-1.319097e-05	-4.854470e-07	1.038322e-06	-1.714009e-07	1.795620e-06	3.352939e-07
Weight	9.712513e-08	1.812295e-06	2.030005e-06	-6.484298e-06	4.357575e-08	-5.071582e-08	8.166150e-07
Special	2.062228e-07	-2.912201e-06	-2.766056e-07	-3.173245e-07	-4.482648e-07	-7.903201e-07	-6.929067e-07
Weak.Foot	-2.912201e-06	1.580790e-03	2.907927e-06	6.156373e-06	-1.094476e-06	8.092737e-06	4.002910e-06
Agility	-2.766056e-07	2.907927e-06	1.009712e-05	4.830901e-07	-1.895367e-06	-2.769290e-06	5.161993e-07
Strength	-3.173245e-07	6.156373e-06	4.830901e-07	1.007810e-05	9.240762e-07	-2.046880e-09	-1.153295e-06
Jumping	-4.482648e-07	-1.094476e-06	-1.895367e-06	9.240762e-07	9.173139e-06	-5.254134e-07	1.680742e-07
Acceleration	-7.903201e-07	8.092737e-06	-2.769290e-06	-2.046880e-09	-5.254134e-07	1.464574e-05	-1.770330e-06
Stamina	-6.929067e-07	4.002910e-06	5.161993e-07	-1.153295e-06	1.680742e-07	-1.770330e-06	2.168487e-05

- IID Data exists with each player being a unique individual
- We are able to find the inverse of $E[X^TX]$
- No perfect collinearity exists within the data being used
- We can assume the data meets the large sample assumptions

Correlation Results

- Goalkeeper skills highly correlated with one another
- Player skills all highly correlated to the “Special” covariate
- “Special” VIF is close to 2, not higher than 4 so we will use in model



Linear Regression Model 1: Base Model

	Dependent variable:		
	log(Value)		
	(1)	(2)	(3)
Age	0.306***	0.447***	0.742***
	(0.043)	(0.039)	(0.043)
I(Age2)	-0.008***	-0.010***	-0.015***
	(0.001)	(0.001)	(0.001)
Height	0.030***	0.016***	0.012
	(0.007)	(0.006)	(0.007)
Weight	0.002	0.0003	-0.006
	(0.005)	(0.004)	(0.005)
Special	0.007***	0.006***	
	(0.0002)	(0.0002)	
Contract.Years		-0.193***	-0.263***
		(0.019)	(0.021)
Agility			0.003
			(0.003)
Strength			0.013***
			(0.003)
Jumping			0.014***
			(0.003)
Acceleration			0.009***
			(0.003)
Stamina			0.008*
			(0.004)
Weak.Foot			0.121***
			(0.036)
International.Reputation		0.847***	1.142***
		(0.053)	(0.059)
Constant	-2.395**	-0.867	-0.889
	(1.149)	(1.023)	(1.279)
Observations	1,358	1,358	1,358
R2	0.557	0.655	0.543
Adjusted R2	0.555	0.653	0.539
Residual Std. Error	0.909 (df = 1352)	0.803 (df = 1350)	0.926 (df = 1345)
F Statistic	339.813*** (df = 5; 1352)	465.862*** (df = 7; 1350)	133.052*** (df = 12; 1345)

model1 msr
0.8229954

Linear Regression Model 2

Anova Test:
Model 2 has a
p-value < 2.2e-16.

Differences between
Model 1 and 2 are
significant.

	Dependent variable:		
	log(Value)		
	(1)	(2)	(3)
Age	0.306***	0.447***	0.742***
	(0.043)	(0.039)	(0.043)
I(Age2)	-0.008***	-0.010***	-0.015***
	(0.001)	(0.001)	(0.001)
Height	0.030***	0.016***	0.012
	(0.007)	(0.006)	(0.007)
Weight	0.002	0.0003	-0.006
	(0.005)	(0.004)	(0.005)
Special	0.007***	0.006***	
	(0.0002)	(0.0002)	
Contract.Years		-0.193***	-0.263***
		(0.019)	(0.021)
Agility			0.003
			(0.003)
Strength			0.013***
			(0.003)
Jumping			0.014***
			(0.003)
Acceleration			0.009***
			(0.003)
Stamina			0.008*
			(0.004)
Weak.Foot			0.121***
			(0.036)
International.Reputation		0.847***	1.142***
		(0.053)	(0.059)
Constant	-2.395**	-0.867	-0.889
	(1.149)	(1.023)	(1.279)
Observations	1,358	1,358	1,358
R2	0.557	0.655	0.543
Adjusted R2	0.555	0.653	0.539
Residual Std. Error	0.909 (df = 1352)	0.803 (df = 1350)	0.926 (df = 1345)
F Statistic	339.813*** (df = 5; 1352)	365.862*** (df = 7; 1350)	133.0*** (df = 12; 1345)

model2 msr
0.6410832

Linear Regression Model 3

	Dependent variable		
	log(Value)		
	(1)	(2)	(3)
Age	0.306***	0.447***	0.742***
	(0.043)	(0.039)	(0.043)
I(Age2)	-0.008***	-0.010***	-0.015***
	(0.001)	(0.001)	(0.001)
Height	0.030***	0.016***	0.012
	(0.007)	(0.006)	(0.007)
Weight	0.002	0.0003	-0.006
	(0.005)	(0.004)	(0.005)
Special	0.007***	0.006***	
	(0.0002)	(0.0002)	
Contract.Years		-0.193***	-0.263***
		(0.019)	(0.021)
Agility			0.003
			(0.003)
Strength			0.013***
			(0.003)
Jumping			0.014***
			(0.003)
Acceleration			0.009***
			(0.003)
Stamina			0.008*
			(0.004)
Weak.Foot			0.121***
			(0.036)
International.Reputation		0.847***	1.142***
		(0.053)	(0.059)
Constant	-2.395**	-0.867	-0.889
	(1.149)	(1.023)	(1.279)
Observations	1,358	1,358	1,358
R2	0.557	0.655	0.543
Adjusted R2	0.555	0.653	0.539
Residual Std. Error	0.909 (df = 1352)	0.803 (df = 1350)	0.926 (df = 1345)
F Statistic	339.813*** (df = 5; 1352)	365.862*** (df = 7; 1350)	133.052*** (df = 12; 1345)

model3 msr
0.8491938

Final Linear Regression Model: Model 2

- Lowest MSR
- Highest R^2
- Added variables improved model

	Dependent variable:
	log(Value)
Age	0.447***
	(0.039)
I(Age2)	-0.010***
	(0.001)
Height	0.016***
	(0.006)
Weight	0.0003
	(0.004)
Special	0.006***
	(0.0002)
Contract.Years	-0.193***
	(0.019)
International.Reputation	0.847***
	(0.053)
Constant	-0.867
	(1.023)
Observations	1,358
R2	0.655
Adjusted R2	0.653
Residual Std. Error	0.803 (df = 1350)
F Statistic	365.862*** (df = 7; 1350)

Conclusion

Key Findings from Model:

- As the **Age** of a Goalkeeper is increased by 1 year, their market value is **increased** by **55.4%**
- As the **Height** of a Goalkeeper is increased by 1 kilogram, their market value is **increased** by **1.61%**
- As the "**Special**" ranking of a Goalkeeper is increased by 1 rating, their market value is **increased** by **0.57%**
- As a Goalkeeper's **contract years** increase by 1 year, their market value **decreases** by **17.56%**
- As a Goalkeeper's **international reputation** is increased by 1 rating, their market value **increases** by **133.23%**

Using the above, we are able to gain insight on how soccer clubs can maximize the market value that their Goalkeeper holds. There is reason to consider these values and how they might directly impact a team's defense on the market side as time passes. Regardless, the information is beneficial for both clubs and goalkeepers to ensure everyone is able to predict how certain factors will affect their market value.

Q&A