

● Player Valuation

Modelling European football player valuation

<https://github.com/MarcoWong96/Capstone-Project.git>

January 2024
Marco Wong



“Doing a Leeds”





PROBLEM

In football's hyper competitive world, clubs face significant financial risks due to inaccurate player valuations.



SOLUTION

- 1. Utilize last season's player statistics to accurately determine player valuations.*
- 2. Find which features are most predictive of valuations*

Impact



Value Appreciation



Improve Fan Relations



Compliance Assurance



Long Term Sustainability



Financial Stability



Marketing Opportunities

Datasets

Player Statistics

2022 – 2023 Season

2921 rows x 143 columns

2021 – 2022 Season

2689 rows x 124 columns

Player Valuation

Transfermarkt

Player Valuation

440643 rows x 9 columns

Players

30302 rows x 23 columns

Proposed Solution:

Player X



Age:

Team:

Height:

Matches Played:

Passes Completed:

Goal Creating Actions:

Aerial Won%:



€ X mil



Player X



Player Statistics

Name	Age	Team	Matches	Passes Completed
X	20	MUN	180	690

Player Valuation

Player ID	Year	Valuation
121	2022	1,000,000

Player Name

Player ID	Player Name
121	X



Player X



Player Statistics

Name | Age | Team | Matches | Passes Completed

X 20 MUN 180 690

Player Name & Valuation Merged (On Player ID).

Player Name | Year | Valuation | Player ID

X 2022 1,000,000 121



Player X



Final Dataframe (*5422 rows × 128 columns*)

Name	Age	Team	Matches	Passes Completed	Valuation
X	20	MUN	180	690	1,000,000



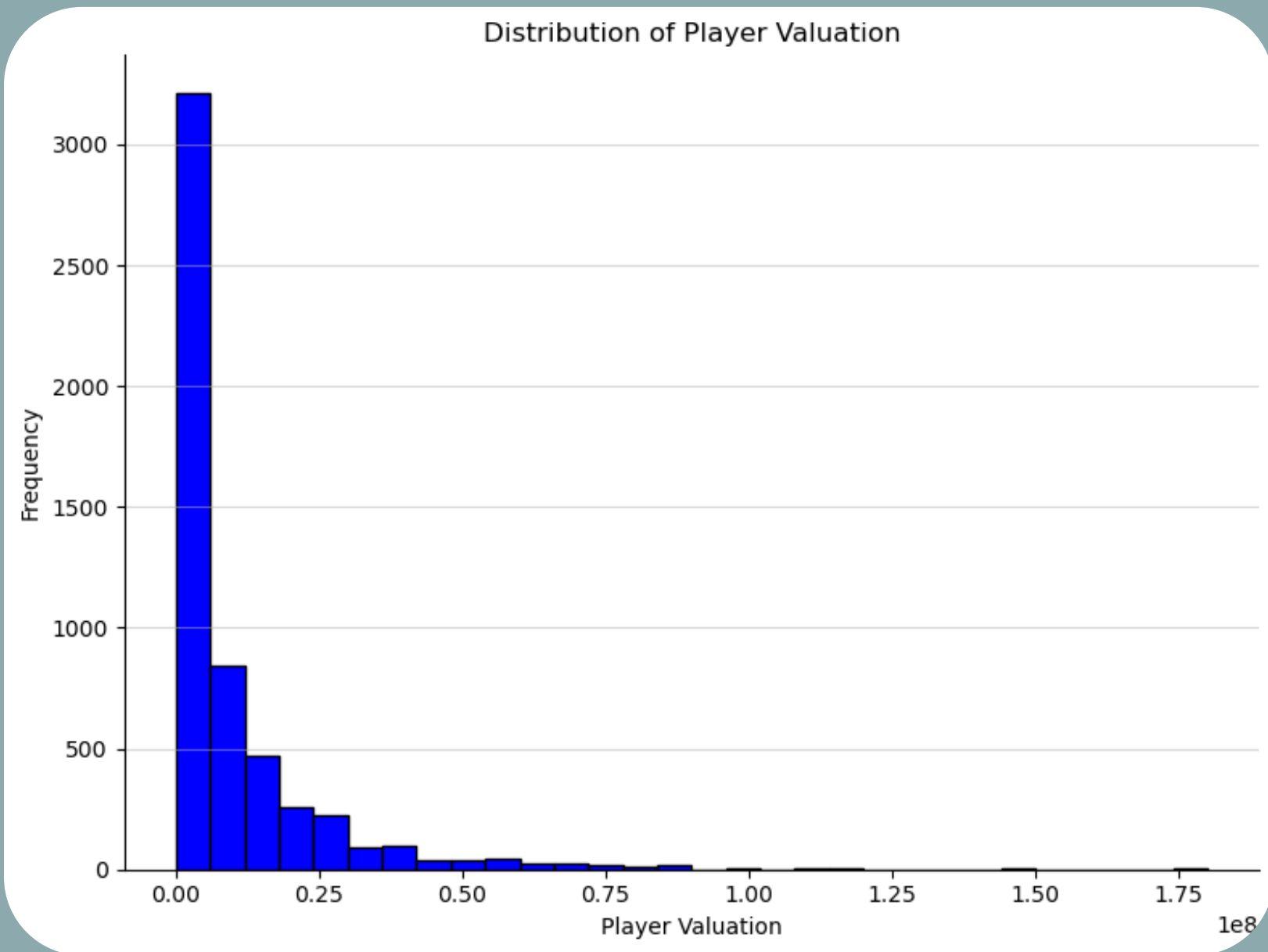
Name

Features

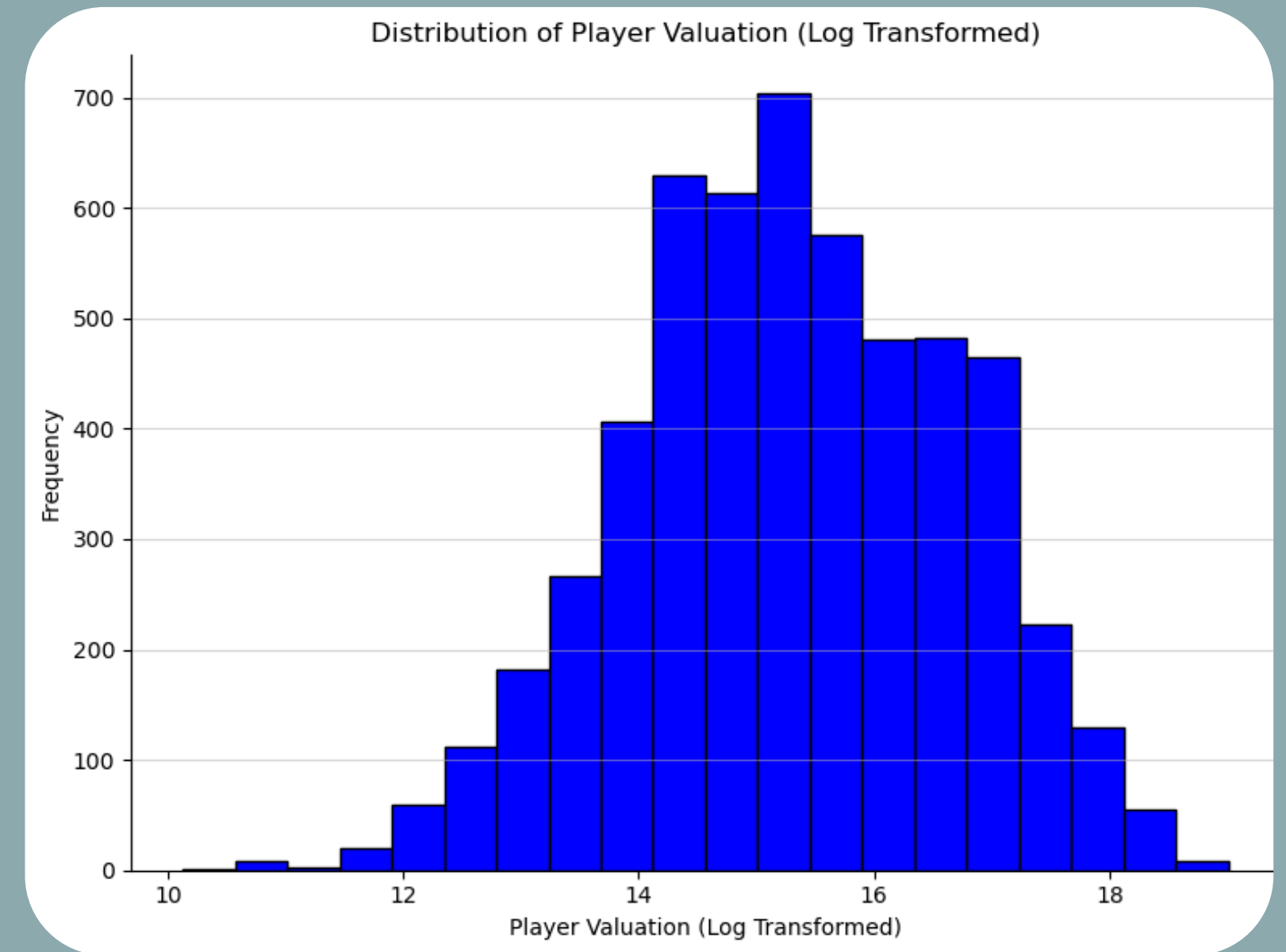
Target
(Valuation)

EDA

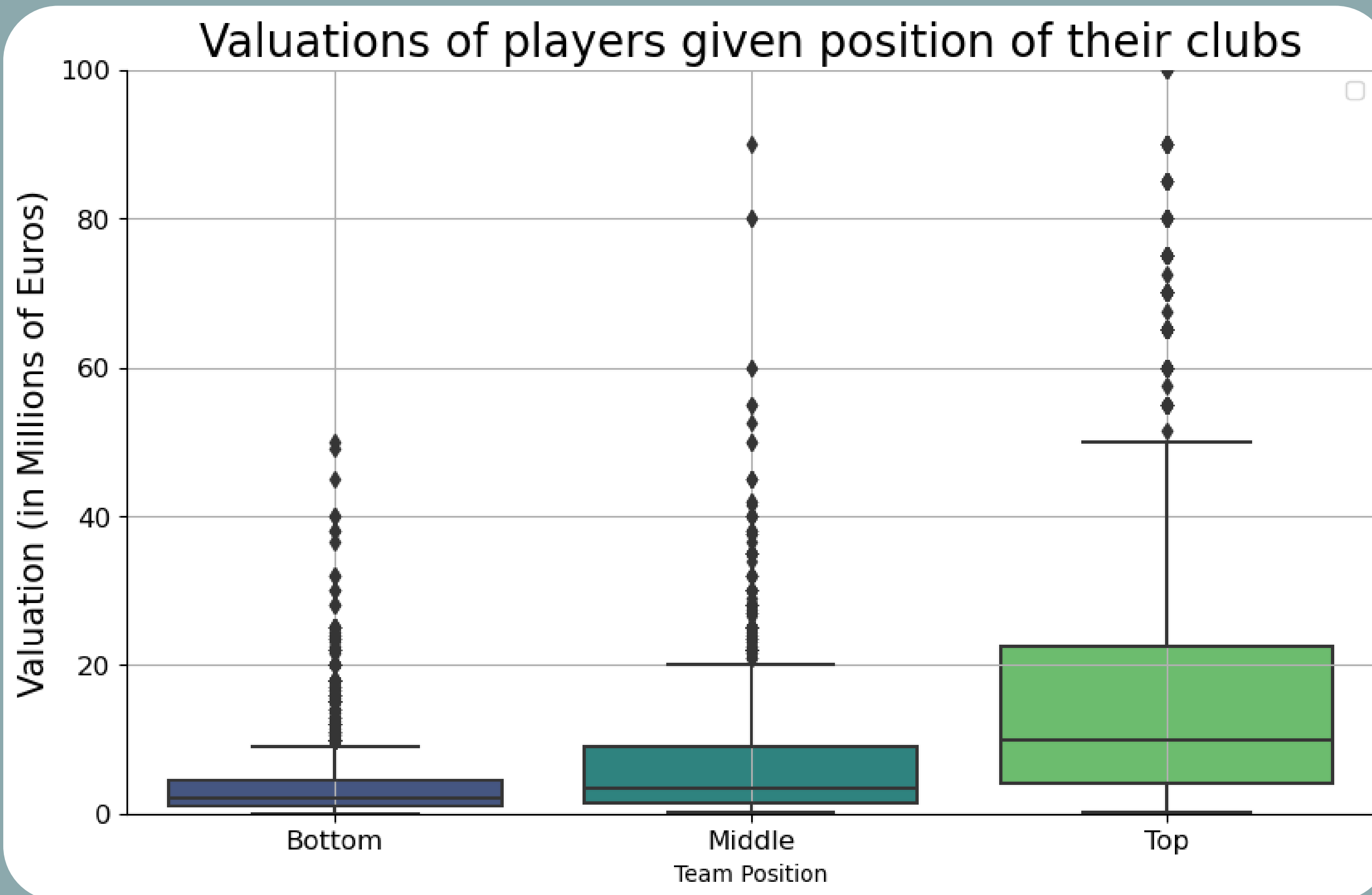
Original Valuation



Log Transformed



FURTHER EDA (Nations)



- Players in top teams have higher mean, median valuation
- Players in top teams have more high valuation outliers
- Anova Test:
P value: 0.00

Data Modeling

Linear Regression (Baseline)

Decision Tree

Random Forest

Neural Network

Bagging

Gradient Boosting





Model Evaluation



R Squared



Adj R Squared



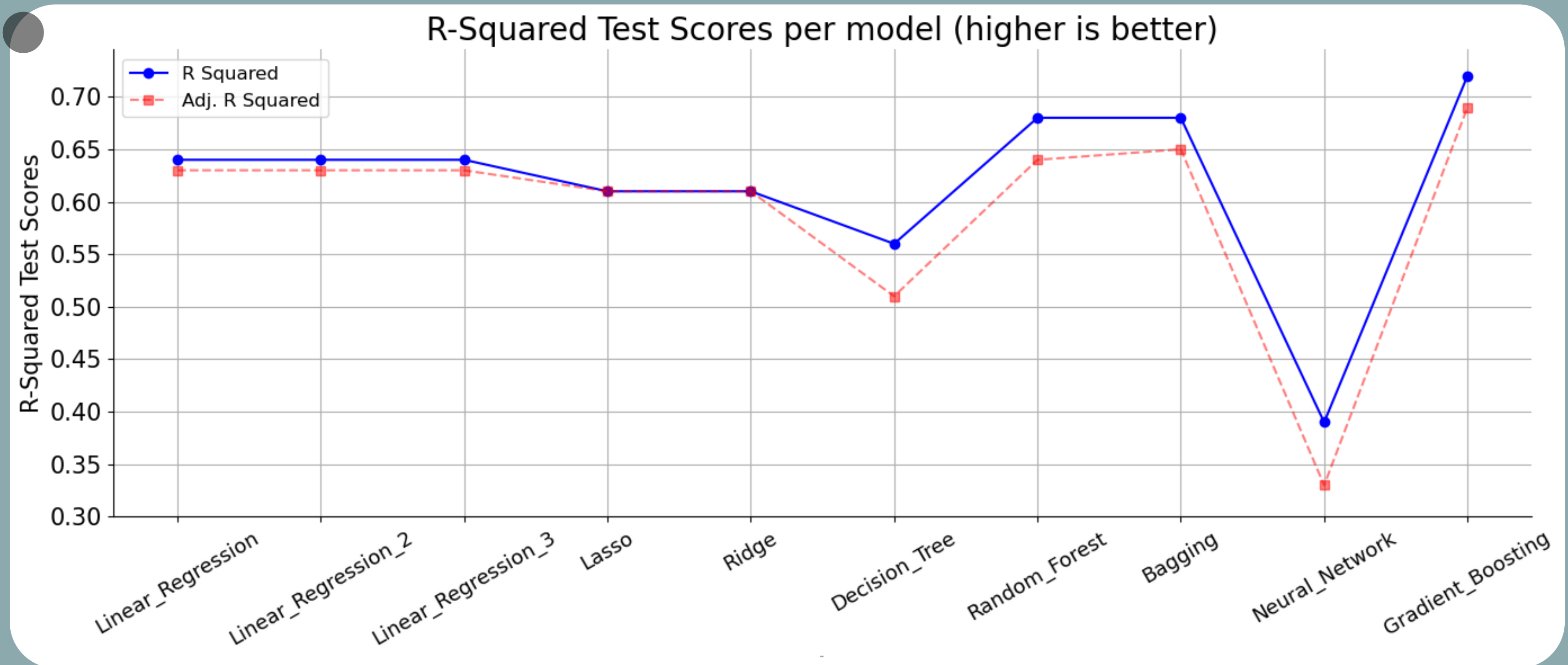
Mean Absolute Error (MAE)



Percentage Mean Absolute
Error (PMAE)

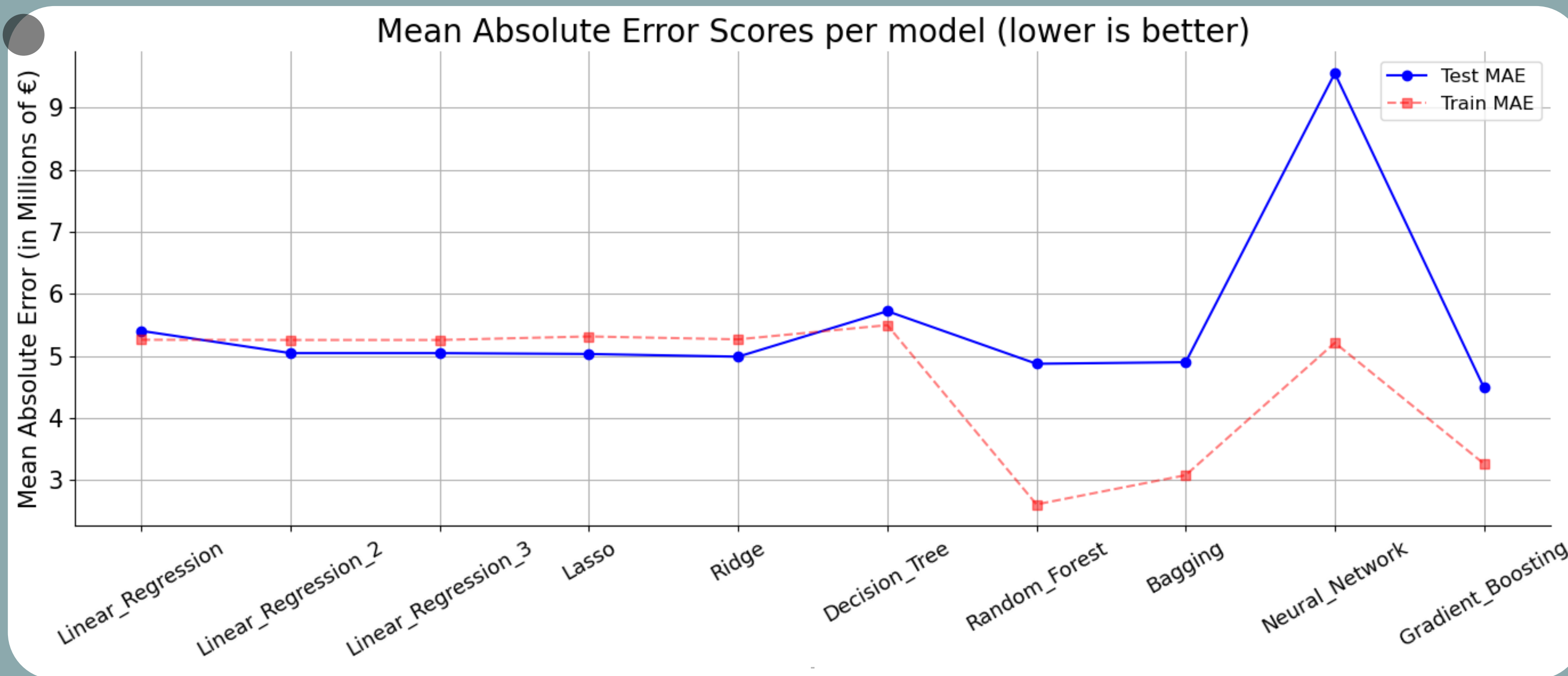


R Squared



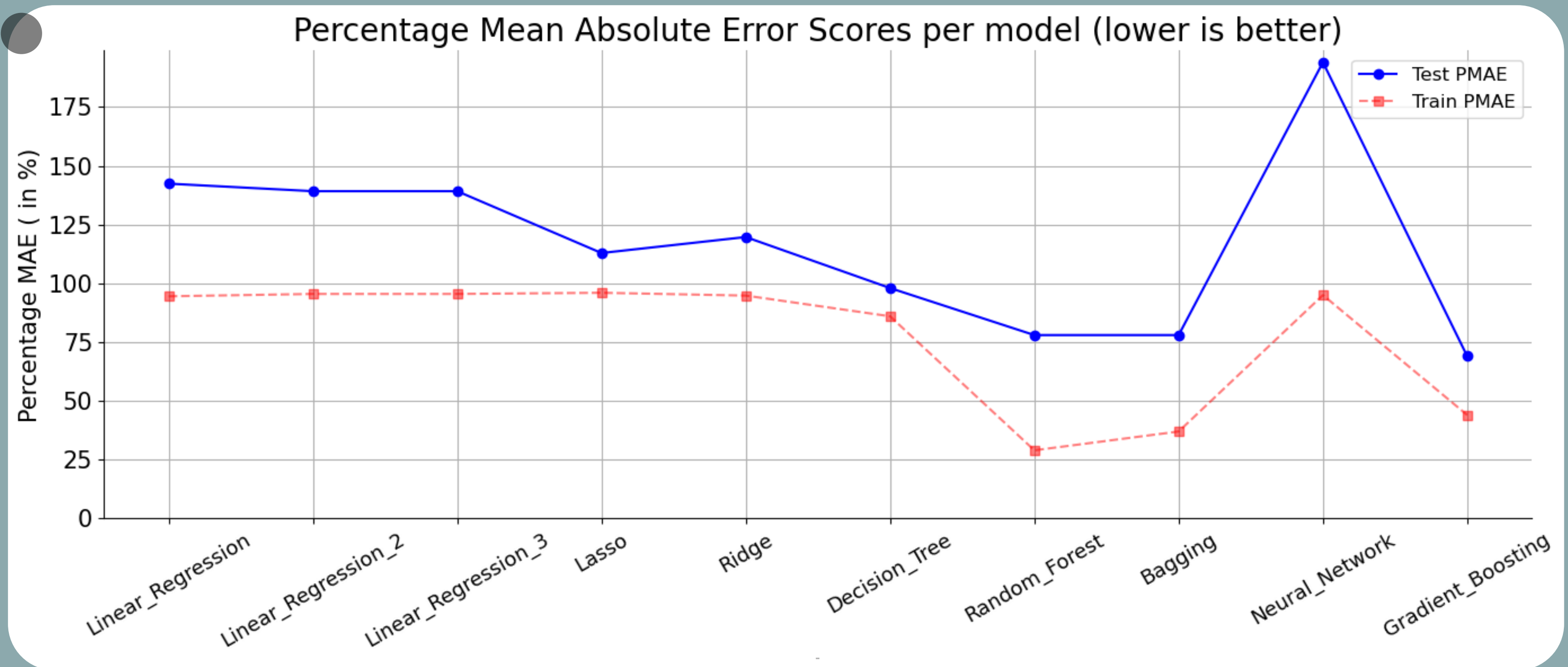


Mean Absolute Error

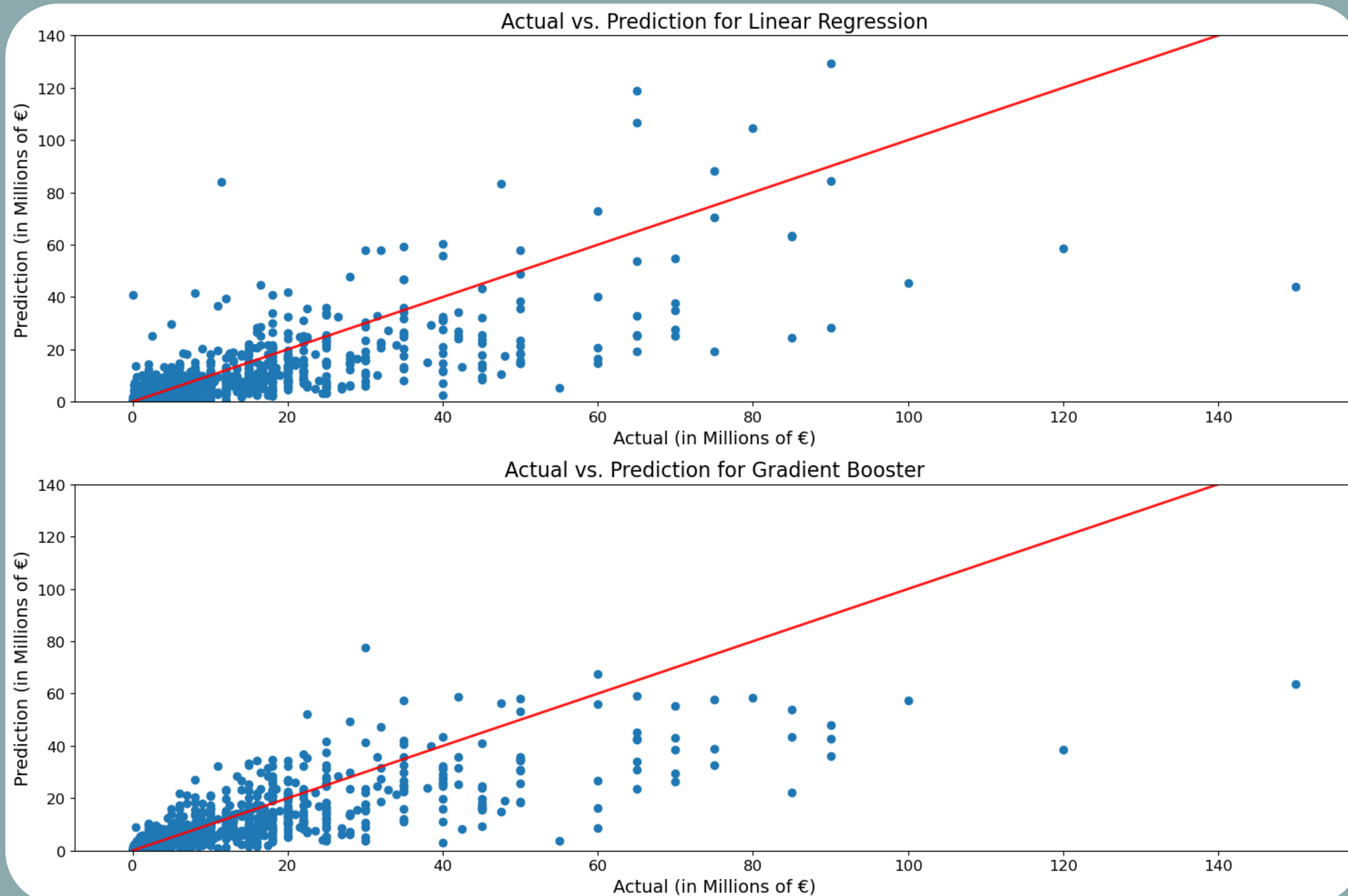




% Mean Absolute Error



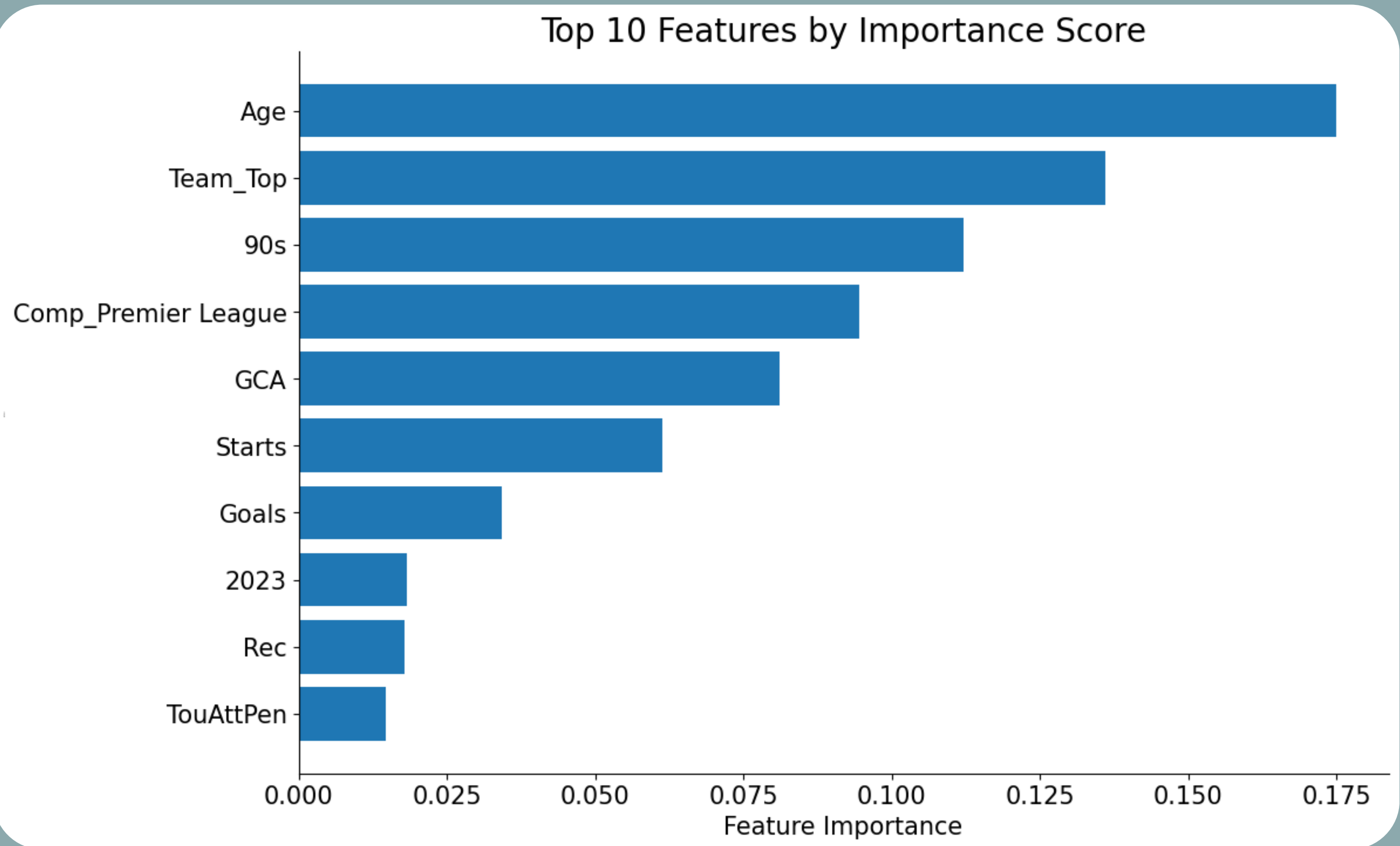
Actual vs Predicted Values



- Red line is perfect valuation
- Linear Regression more spread out
- Gradient Booster largely undervalues



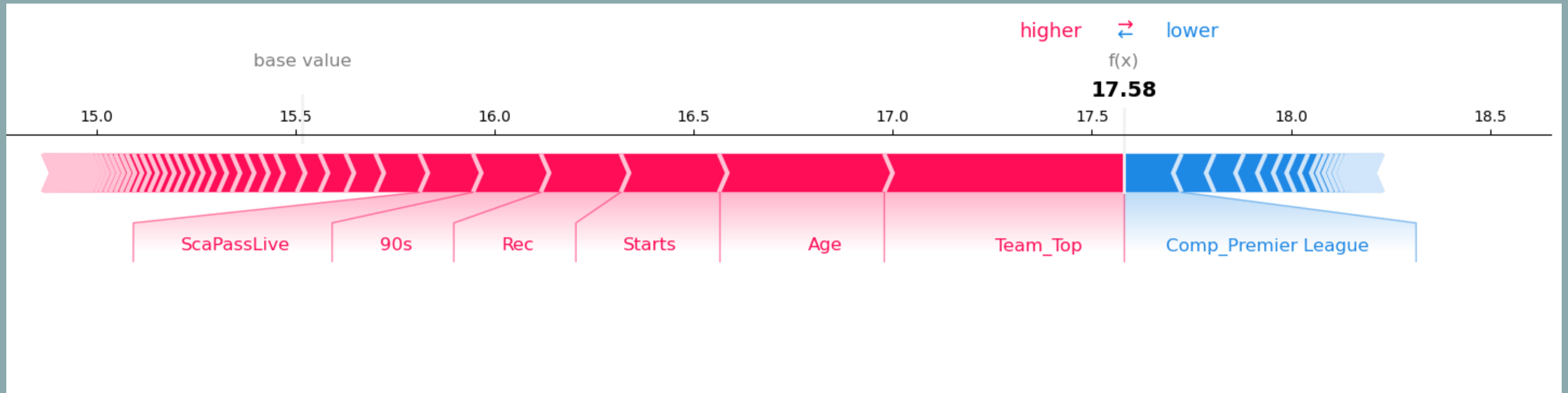
Feature Importance





Alphonso Davies

Valuation: €70 million



Positive Features:

- Top Team
- Age
- Starts

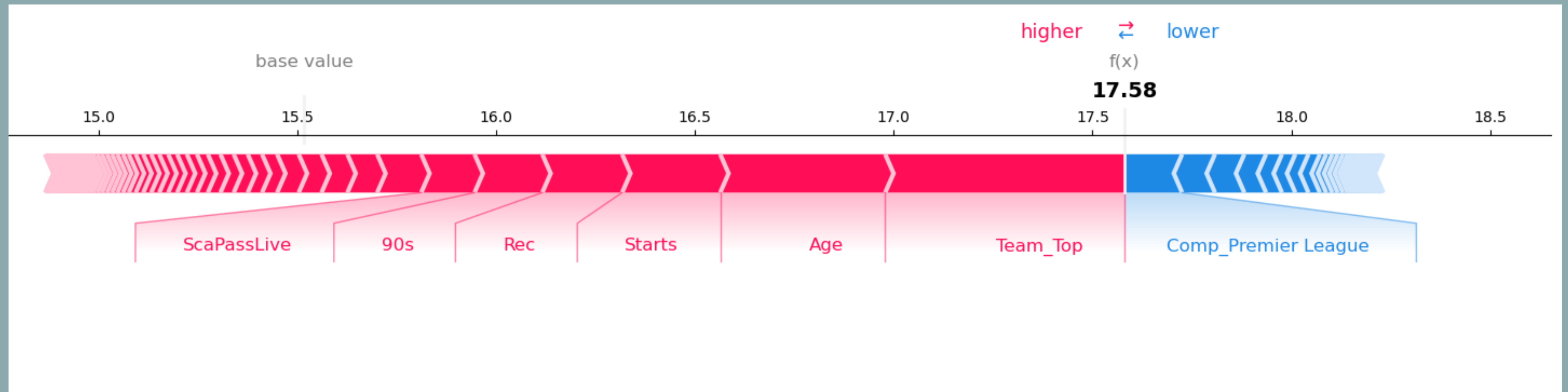
Negative Features:

- Doesn't play in Premier League



Alphonso Davies

Valuation: €70 million



Shap Prediction:
43 Million

Booster Prediction:
43 million

Product Demo





Future Steps

- Look at transfer fees in past two years and evaluate model
- Find methods to improve model: Contract length, wages, other statistics
- Create more robust application
- Consider ethical implications of model/ project



Thank you!

LinkedIn Profile



<https://github.com/MarcoWong96/Capstone-Project.git>