FIFA19 Player Valuation

What’s a player’s worth?

Inspiration

“Moneyball”

Oakland Athletics built one of the most successful teams comprised of undervalued players by simply using player statistics.

Challenges

Inadequate pricing:

A club may overpay for a player or undersell a rising star.

Objective

Make accurate player value predictions:

Create a machine learning model to predict the market value of a player to help club managers obtain undervalued players at a bargain price

Most helpful for clubs with stricter budgets

Data Cleaning

More Data Cleaning

…

...

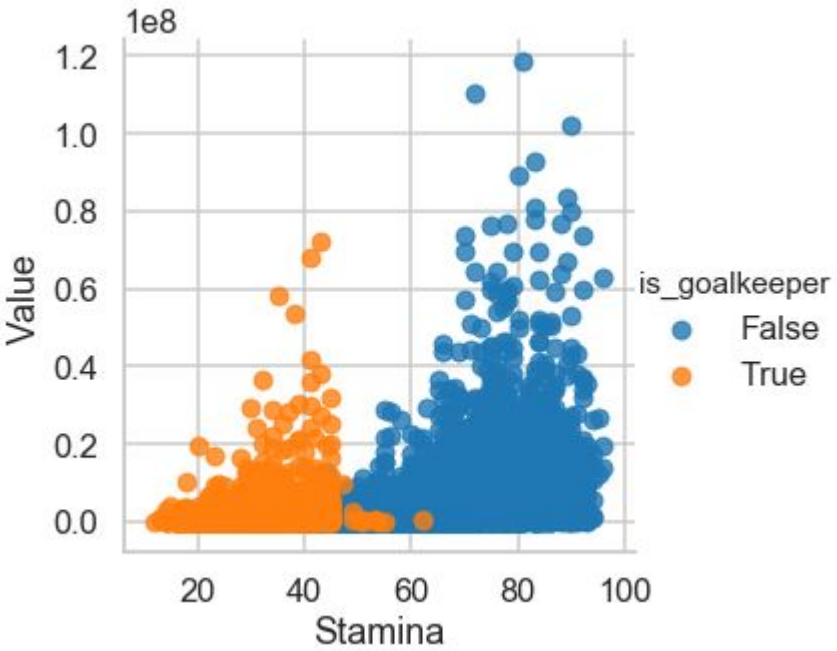
Data Cleaning Continues

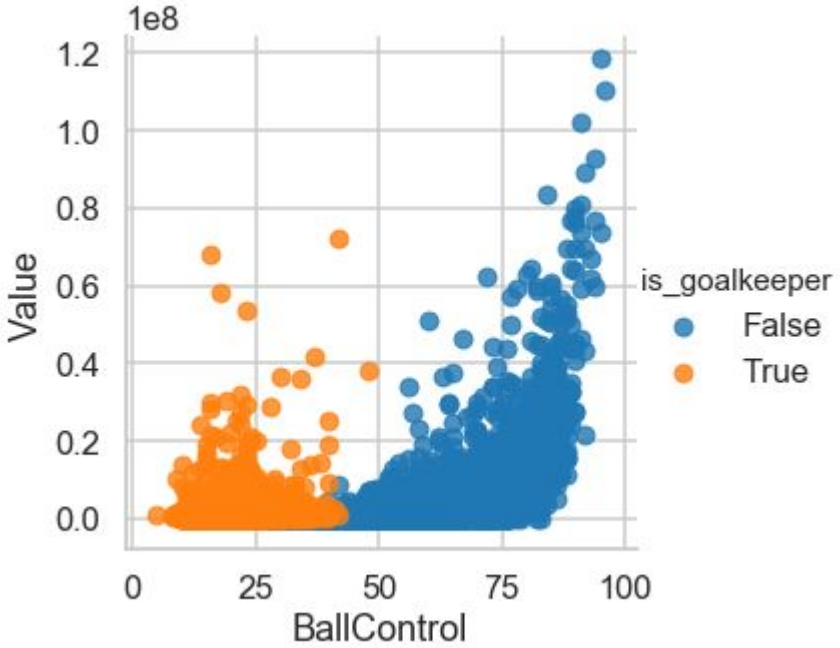
…

Let’s just skip this step for the purpose of the presentation

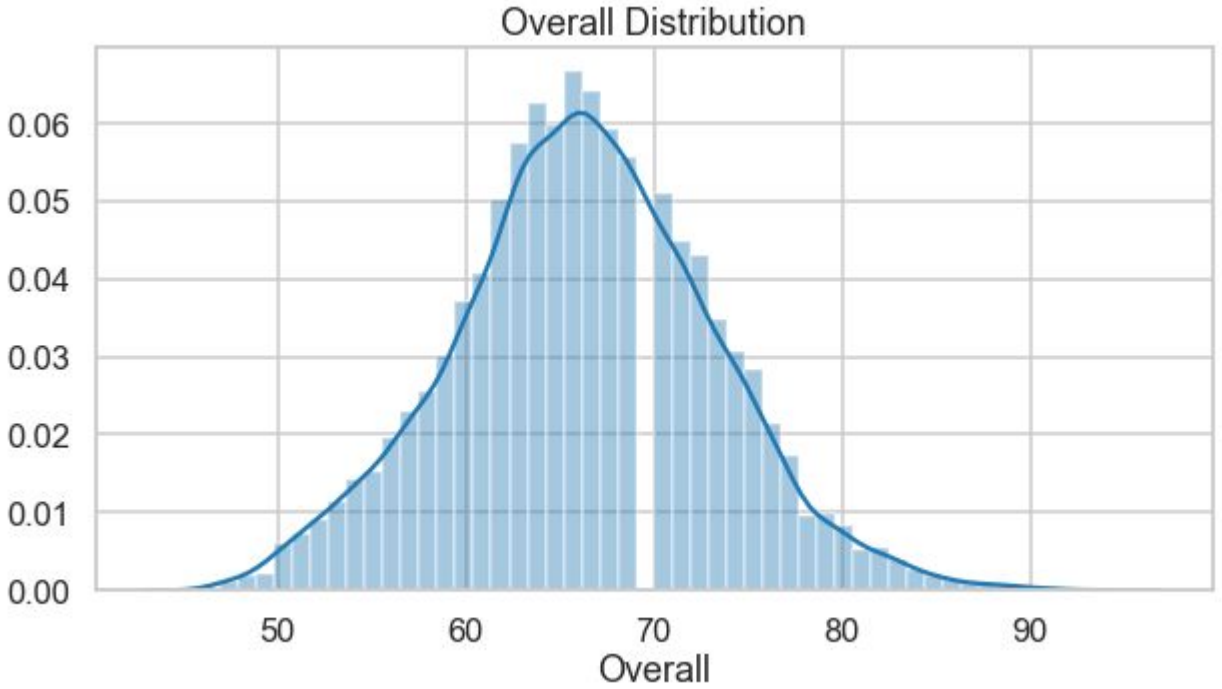
Exploratory Analysis

Separate Goalkeepers from Field Players

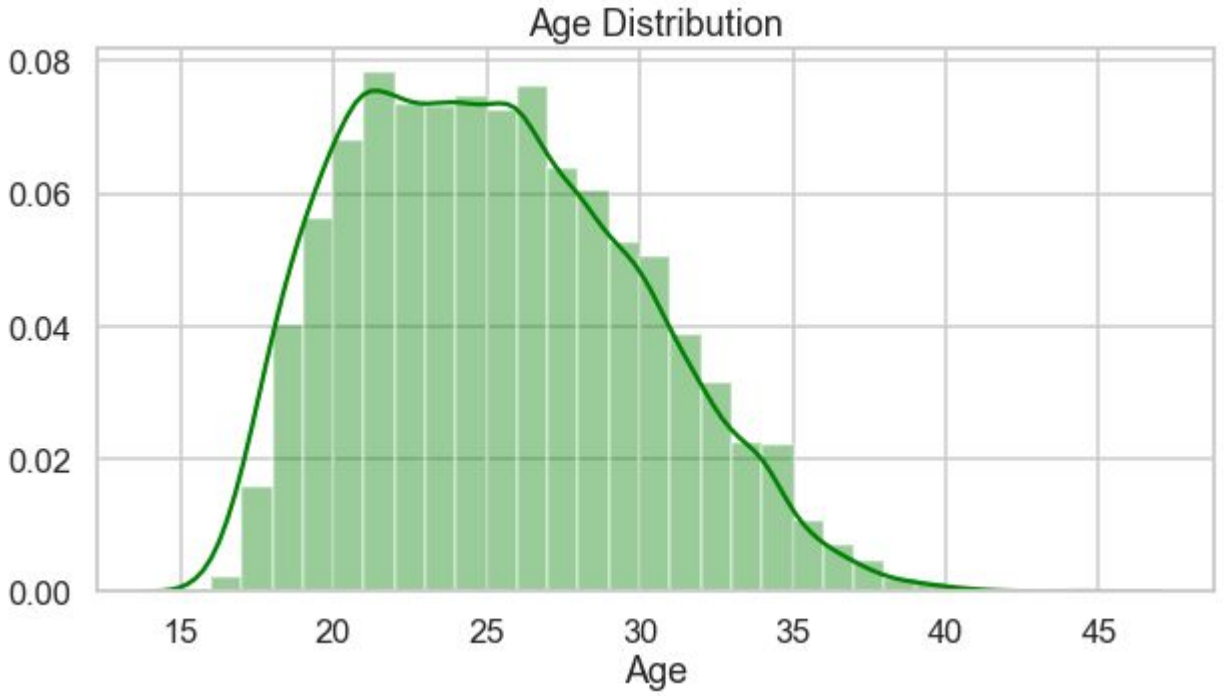




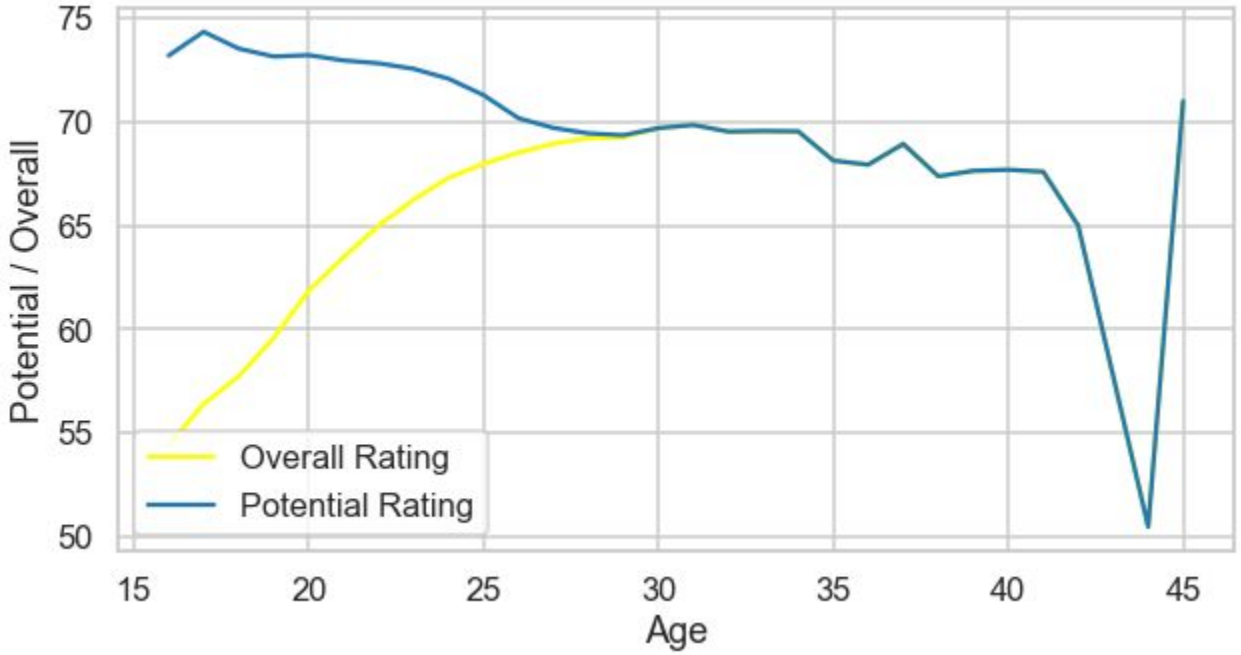
Players Overall Ratings Are Normally Distributed



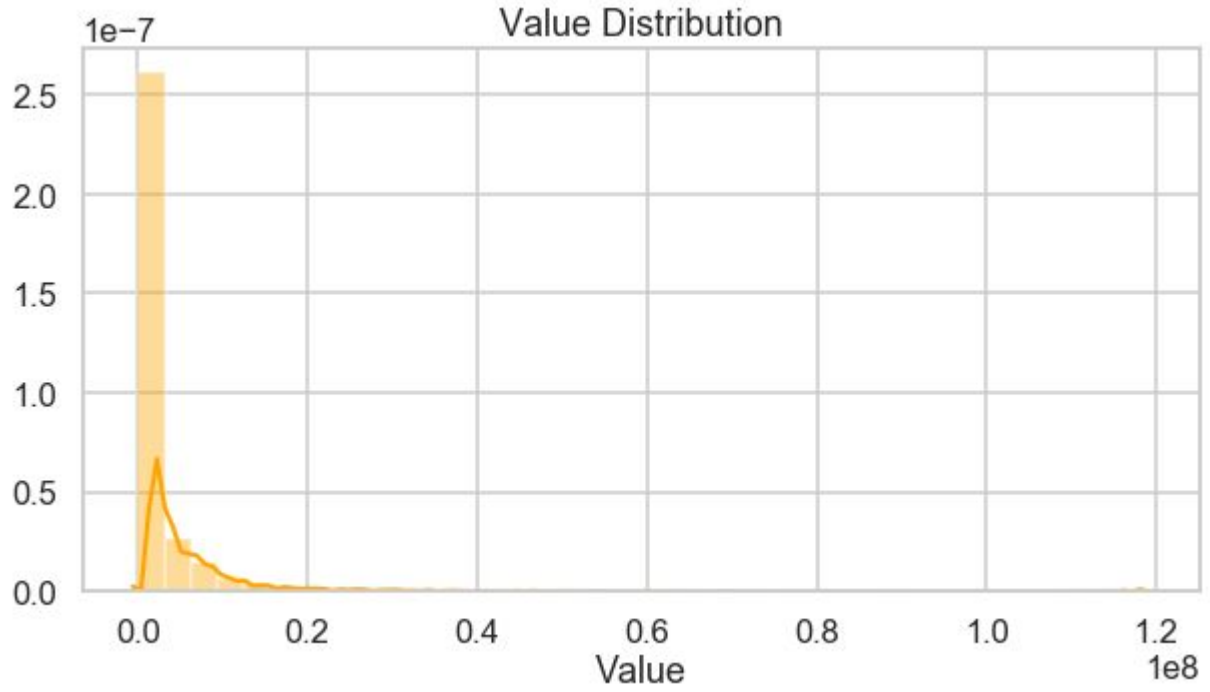
Player ages are NOT normally Distributed



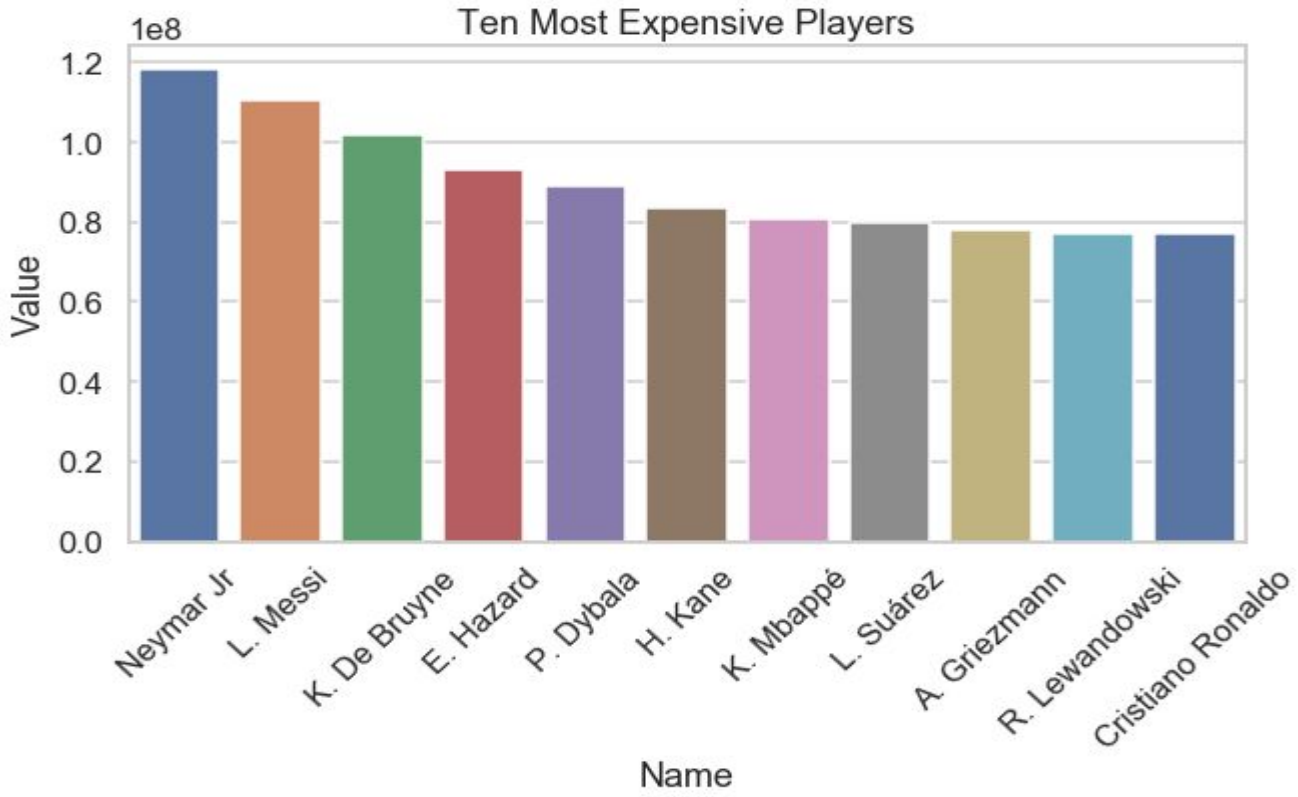
Peak age is reached at high 20s and low 30s



Player Valuations show a heavily positive skew



Heavily skewed because of superstars...



Positions vs Valuations

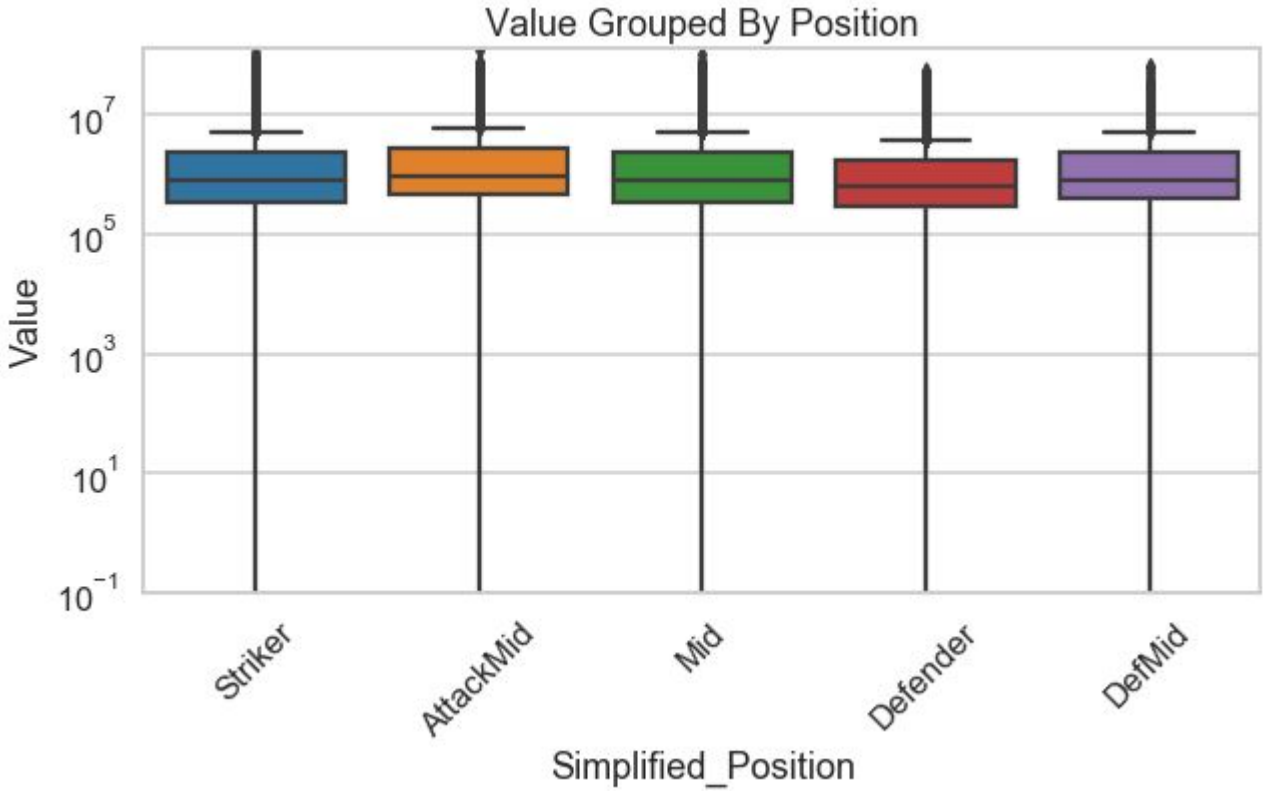
Attacking positions

Might cost you a

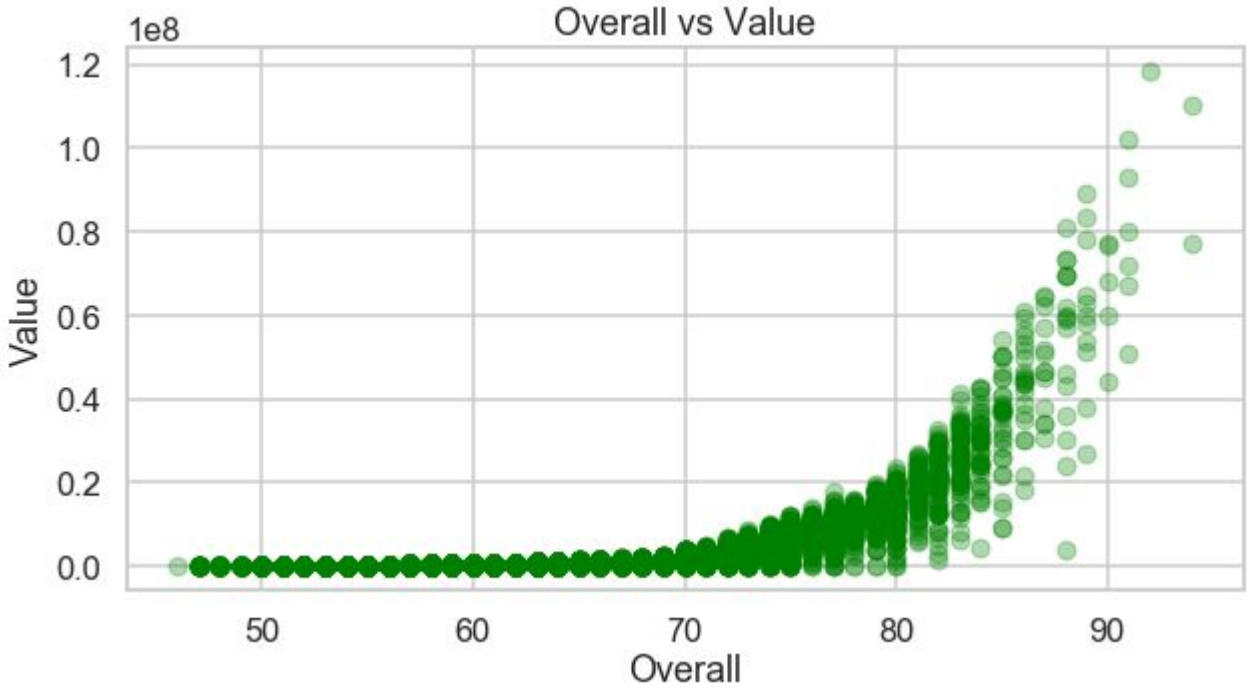
Little more than

Defenders and

Goalkeepers.



Higher Ratings Cost More Money



Which positions are skilled in which attributes ?

AttackMid

DefMid

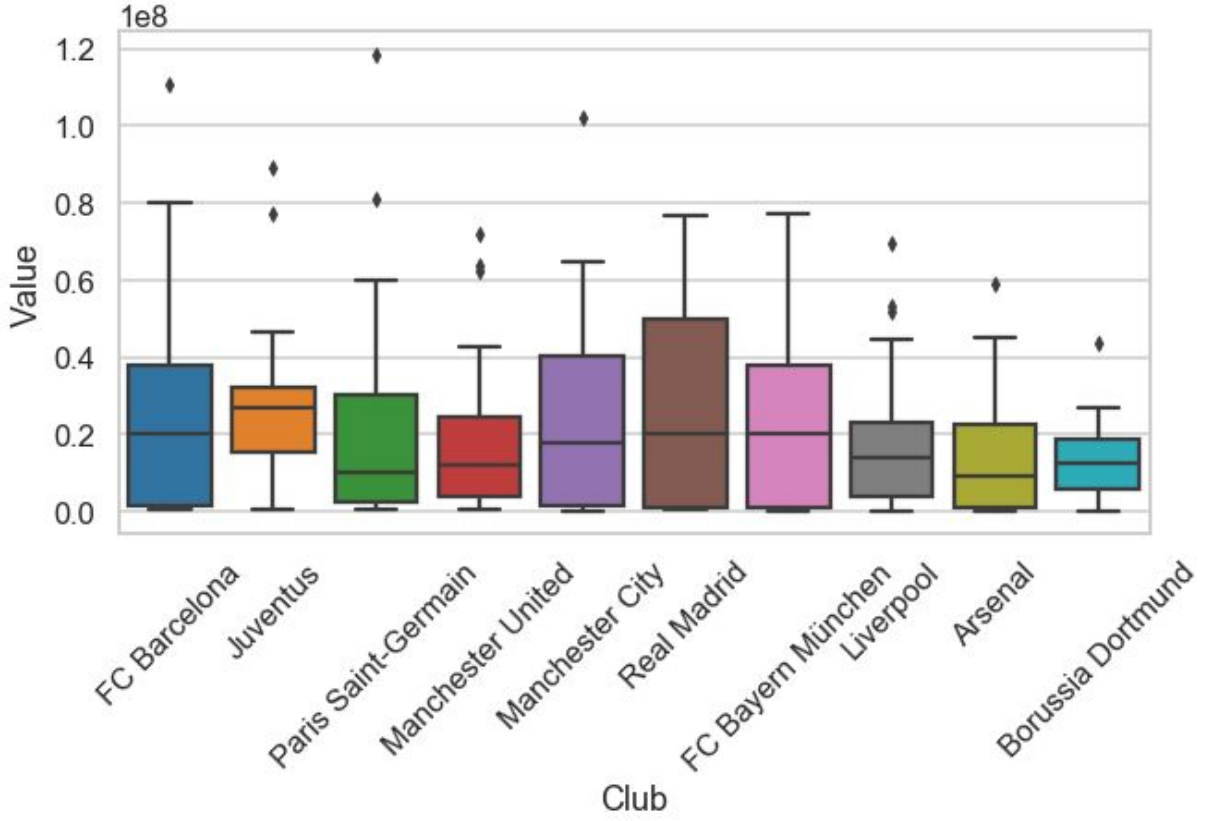
Defender

GK

Mid

Striker

Where are the most valuable players?



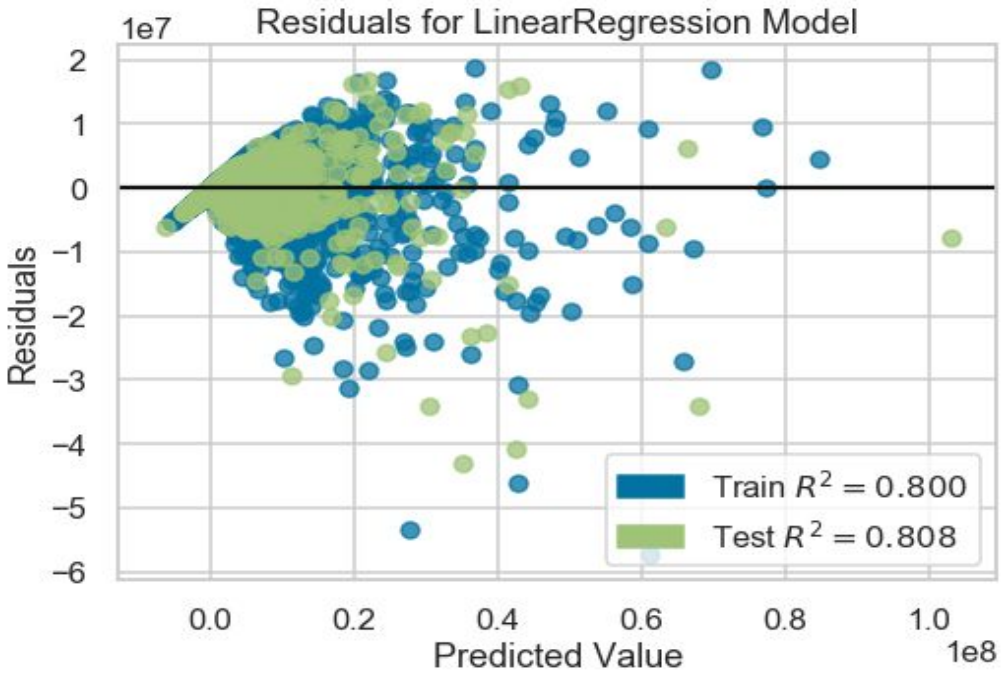
Machine Learning

Evaluation Metric :

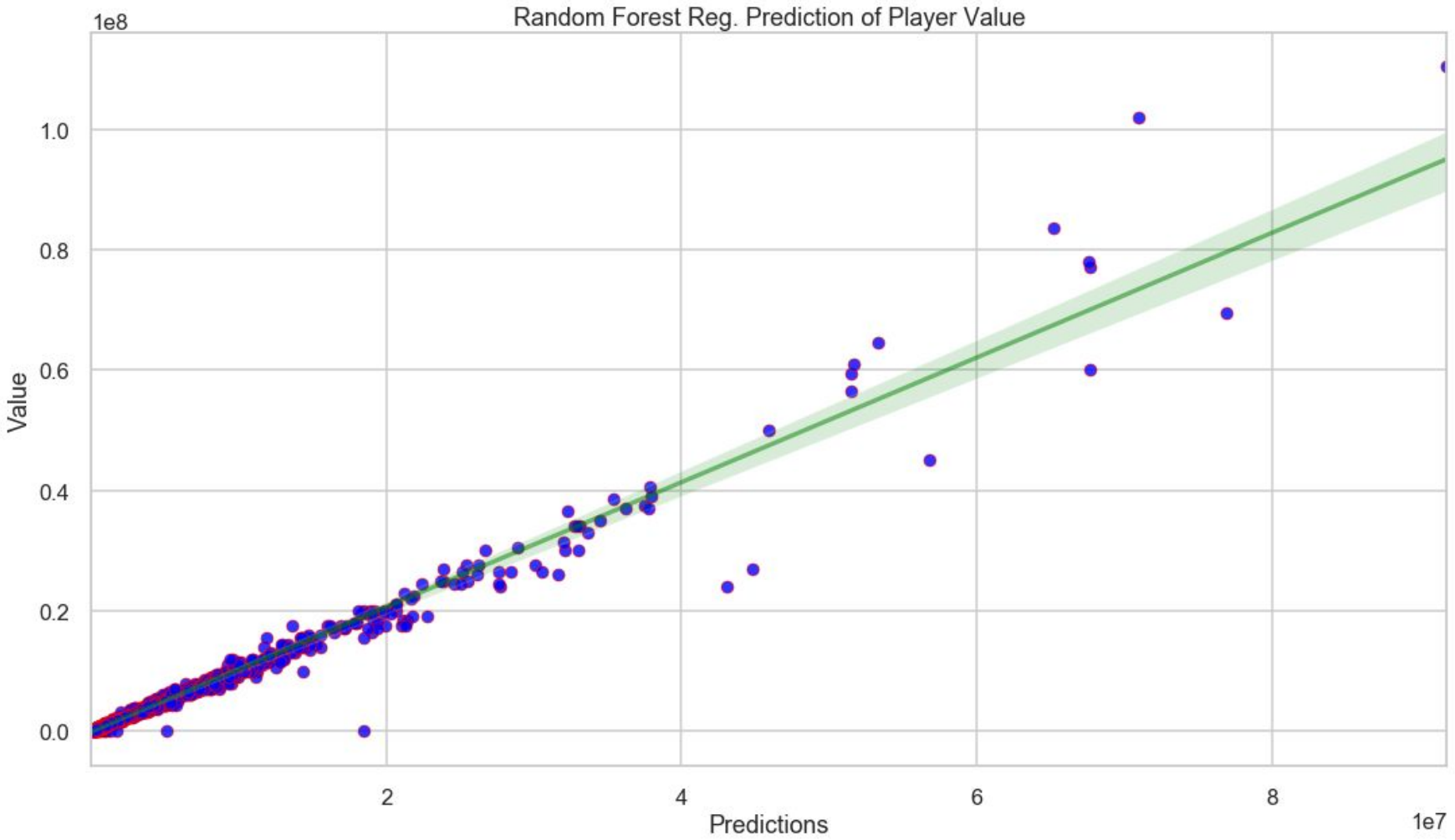
$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Linear Models Performed well but...

Linear model assumptions were not met.



Random Forest Regressor


Random Forest Reg. Prediction of Player Value
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Value
2
4
6
8
1e7
Predictions

Random Forest Model Performed Better without having to meet model assumptions

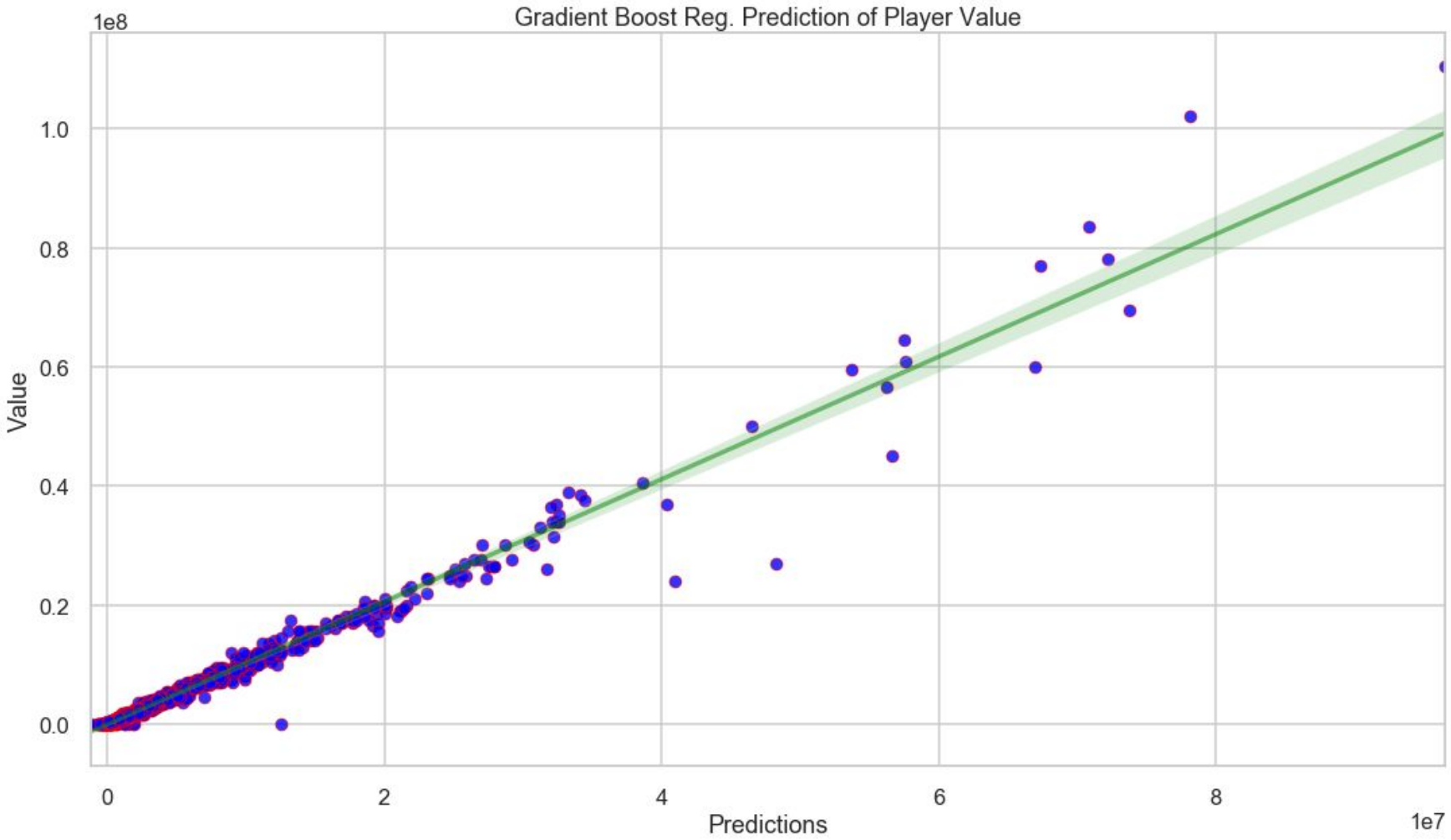
RMSE : 1127037 (initial)

RMSE : 115881 (after tuning)

Feature Importance with Random Forest Model

Most important features were “Age’ and “Overall” rating.

Gradient Boosting Regressor


Gradient Boost Reg. Prediction of Player Value
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Value
0
2
4
6
8
Predictions
1e7

Performance Similar to RF regressor

RMSE : 931869 (initial)

RMSE : 922059 (post tuning)

Feature Importance for GB model

In addition to “**Overall**” and “**Age**” attributes, the gradient boosting regressor also gave some significant weight to “**Dribbling**”, “**Wage**” and “**Ball Control**” features

Future Implementation

Create models based on real football player data

Help club managers with strict budgets find “best value” players that will make the most impact on the field without breaking the bank.