



What to pay a player?

Predicting Player Salary in the English Premier League

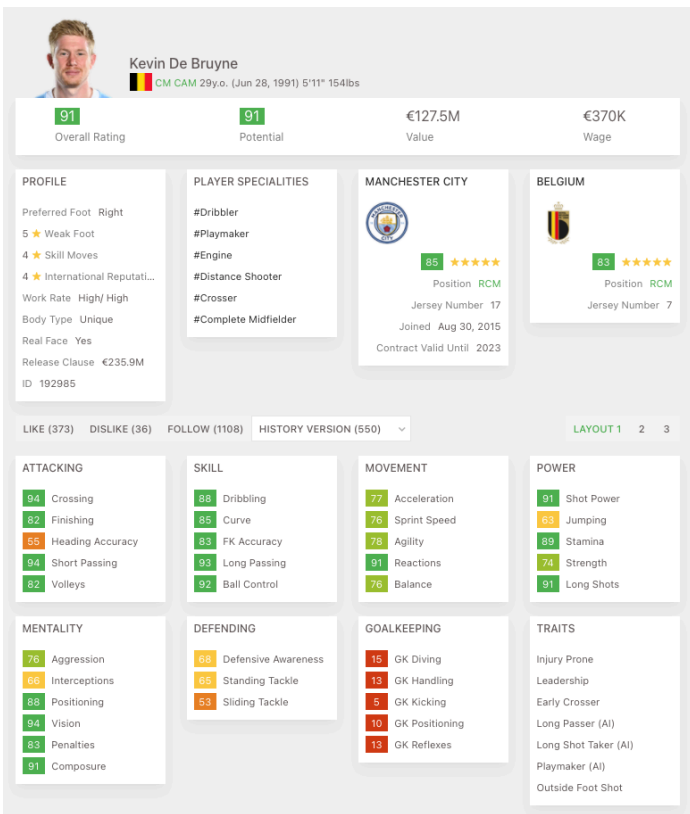
Rhys Carter
Metis, Spring 2021



Contents

- Project Background
- Salary (Wage) Modeling
- Findings
- Key Takeaways & Next Steps

What elements of a player's performance/profile impact their salary?



- Faster vs. slower players
- Position (e.g. forwards vs. midfielders)
- Age/experience

Used data from sofifa.com to create player profiles and model performance against the target, salary



1. Obtain Data

- Beautiful soup
- sofifa.com
- 649 English Prem League Players
- ~30 unique ratings



2. Begin Modeling

- Seaborn, Scikit-learn
- Simple & Multi-variate regression
- Feature down-selection

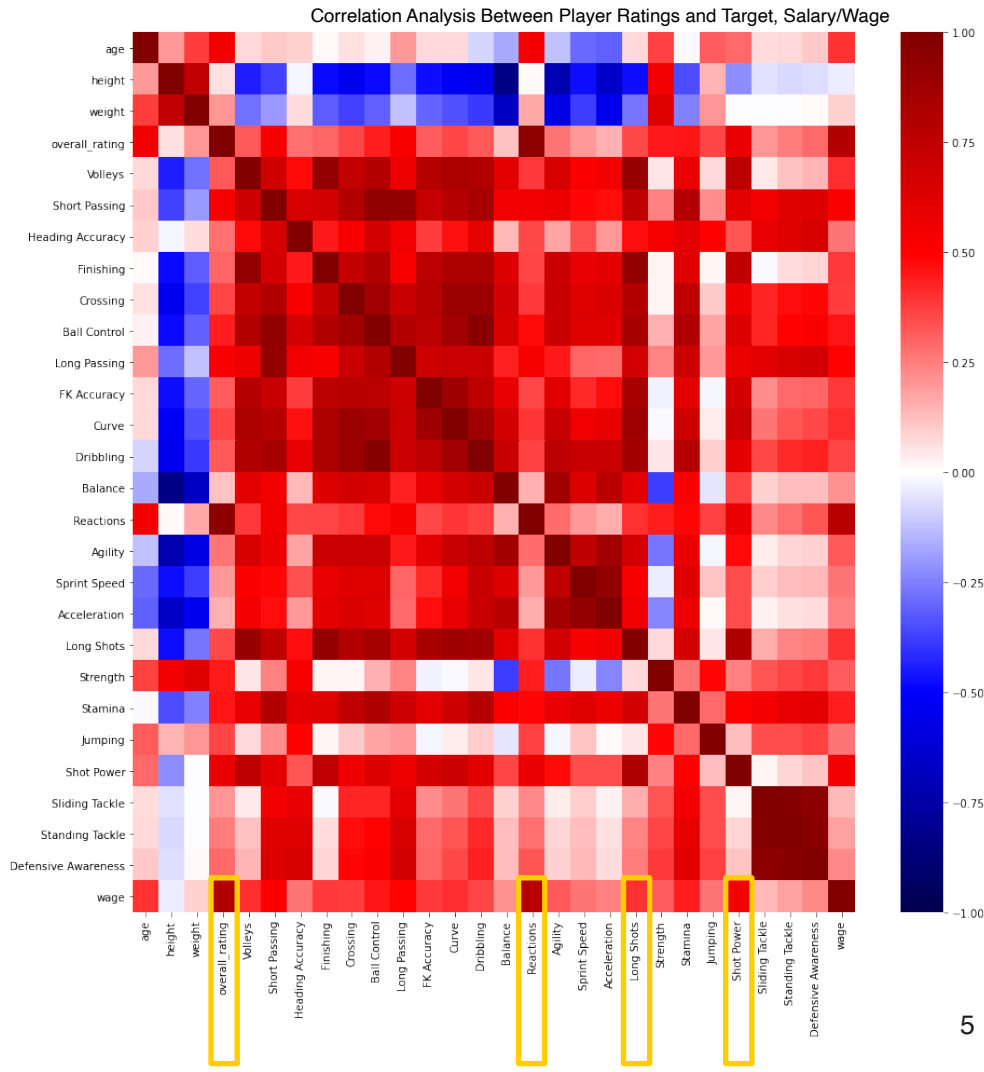


3. Findings

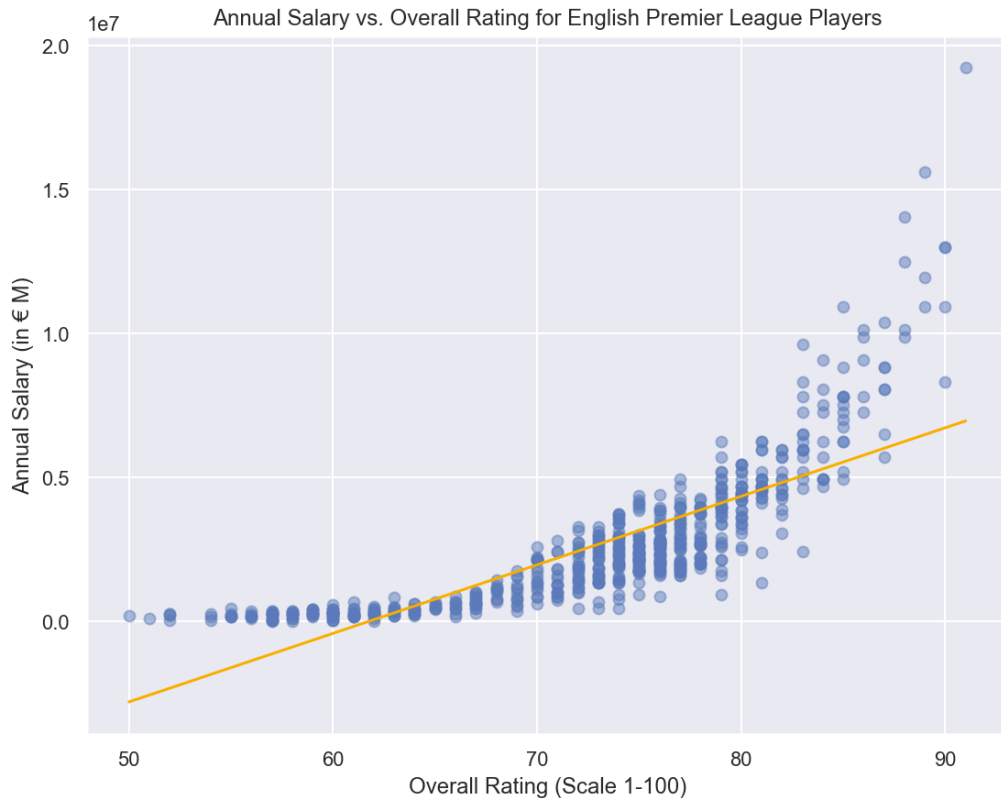
- Focus on offensive categories
- Outlier sensitive
- Overall player rating highly predictive

Correlation analysis highlighted key attacking attributes

- Overall Rating
- Reactions
- Shot Power, etc.



Simple Regression: Overall Rating achieved a high R^2 , even without transformations to the variable

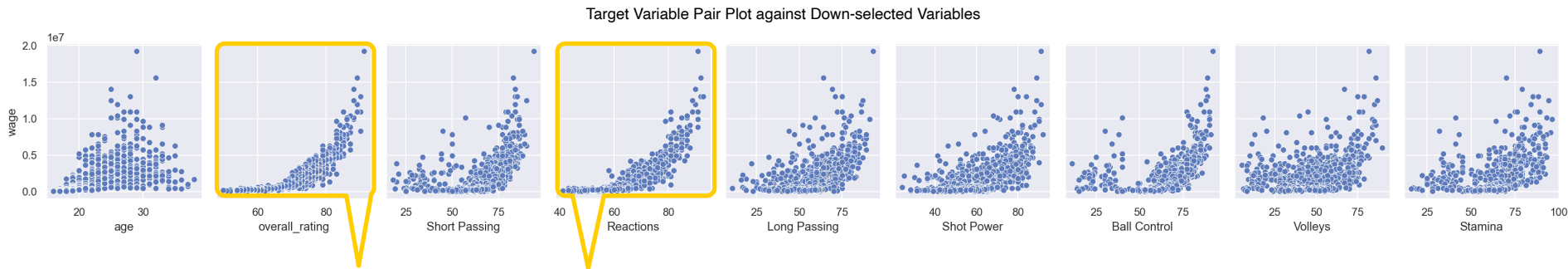


R^2 : 0.6472

Slope: 238091.64

y-intercept: -14696667.86

Down-selected variables show exponential relationship while approaching rating limit of 100



Relationships to adjust at minimum

Adjusting exponential-appearing relationships and adding interaction terms improves on baseline R^2

Model	Description	R^2
Simple Linear Regression	<ul style="list-style-type: none">Target (wage) against Overall Rating	0.647216
Multi-Variate Regression	<ul style="list-style-type: none">Add Overall Rating SquaredStandardize variable scaleAdd dummy variables for Position (e.g. Forward, Midfield, Defender)	0.853717
LASSO	<ul style="list-style-type: none">5 fold Cross-ValidationAuto-chosen Alpha: ~91.16	0.853914
Multi-Variate Regression	<ul style="list-style-type: none">Adding Overall Rating Inverse TransformAdding Reactions Squared	0.880734
Multi-Variate Regression	<ul style="list-style-type: none">Adding interaction term for Forward Shot Power	0.881305

Premier League model needs to be augmented with other leagues for broader application globally



- Overall-Rating strong
- Reactions and Volleys (Awareness Factors)
- Position Relevance (Forwards vs. Others)



- **Additional Leagues**
English Championship, others
- **Expanded Interaction Terms**
Position-dependent
- **Additional Experience Data**
Minutes Played, etc.

Appendix

Down-select Feature Pair Plot

