



Is Steph Curry Overpaid?

Predicting an NBA Player's Salary

Med C. Billian



Goals of this Project:

- To identify which game statistics are the best in predicting the Salary of NBA players?
- To predict the salary of an NBA player using a regression model and training it with NBA statistics from 2015 to 2020.
- Help determine if an NBA player is overpaid or underpaid.

Algorithms:

- Scrape data from:
 - <https://www.basketball-reference.com>
 - <https://hoopshype.com/salaries/>
- Train 6 models:
 - Multiple Regression
 - Multiple Regression + Standardized Data
 - Multiple Regression + Feature Engineering
 - **Ridge CV=5** + Feature Engineering
 - Lasso CV=5 + Feature Engineering
 - Elastic CV=5 + Feature Engineering
- Evaluate 6 models.
- Validate best model(s).
- Evaluate final model - Fine Tune.
- Test final model.



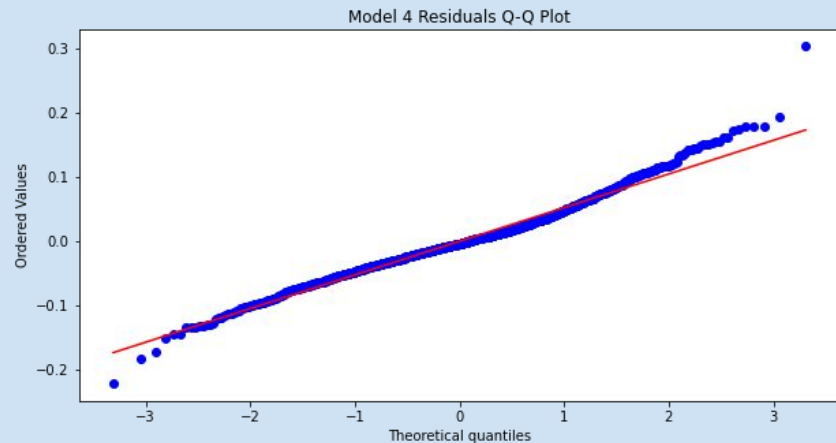
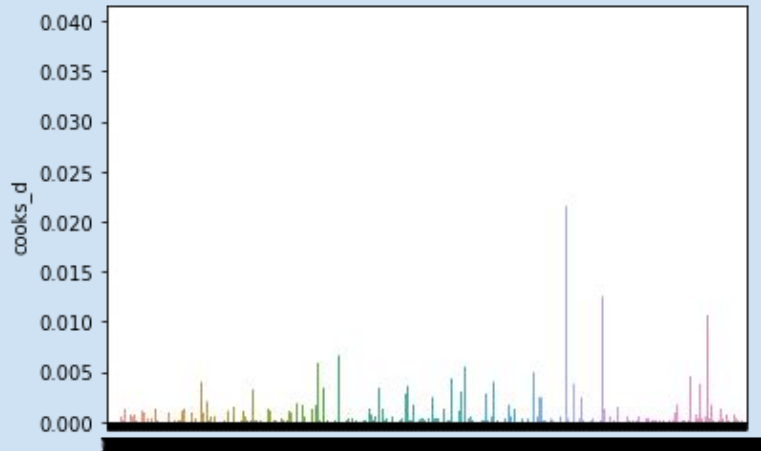
Data Set:

I combined the Players Game Statistics data and the Salary data from the 2 sites. I matched the salary data of a specific year with the game statistics data from the previous year.

1. Training Data
 - Salary from 2015 - 2018
 - Game Statistics from 2014-2017
2. Validation Data
 - Salary from 2019 - 2020
 - Game Statistics from 2018-2019
3. Test Data
 - Salary from 2021
 - Game Statistics from 2020

Target data was engineered to be the square root of the ratio of Salary and Salary Cap.



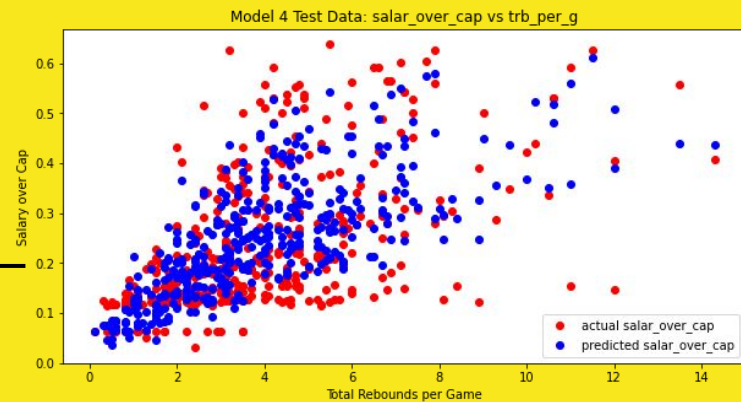
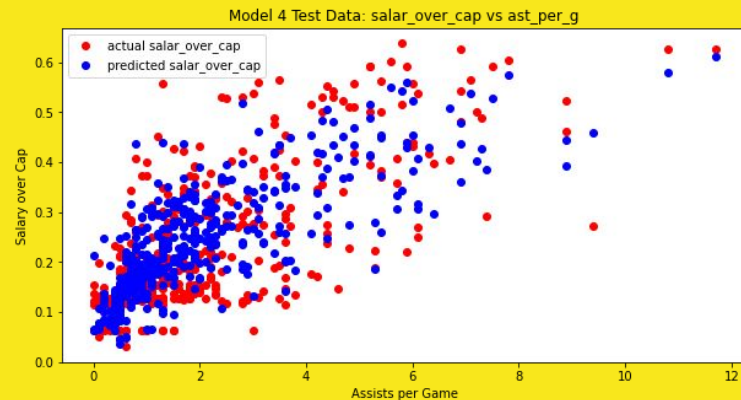
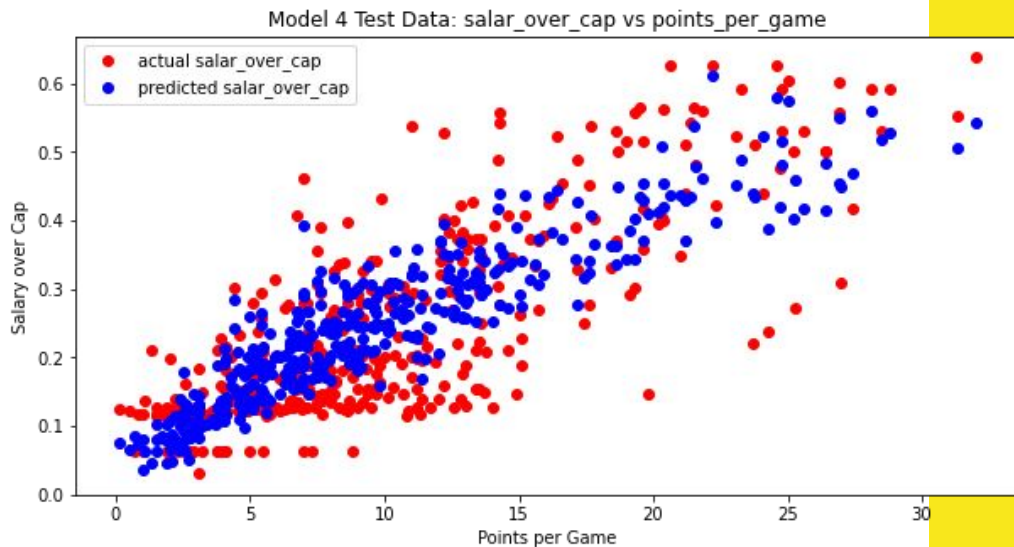


Model Evaluation & Selection

- Cook's Distance for Influential Points, Variance Inflation Factor for Collinearity, and QQ Plots for Residual Independence.
- Multiple regression with the feature engineering model, RidgeCV, and LassoCV models displayed the highest performance using the training data from 2014-2018. $\sim R^2 = .59$
- But the RidgeCV model showed a slightly higher $R^2 = 0.63$ with validation data from 2018 to 2020.
- Training and Validation data scored $R^2=.61$ on the refitted final model.
- *Final Ridge CV model scored $R^2=0.69$ with the test data from 2020 to 2021*

KEY TAKEAWAY

- The game statistics that can be used to predict the salary of an NBA player:
 - Average Points per Game
 - Total Rebounds per Game
 - Assists per Game are



KEY TAKEAWAY???

- According to this model, Steph Curry's 2021 Salary should be \$33,050,197
- but his actual 2021 salary is \$45,780,966.



MORE QUESTIONS!

- Should I use Time Series as my model for this project?
What about cross sectional data?
- What other data features can I use to improve my model?




Data Sources:

<https://www.basketball-reference.com>

<https://hoopshype.com/salaries/>

Thank You

A close-up photograph of a silver and black ballpoint pen writing the words "Thank You" in cursive on a piece of light-colored, textured paper. The pen is positioned at the end of the word "You", with its tip touching the paper.