

National Basketball Association Salary Predictor and Team Optimizer

Austin Cody

Springboard Data Science Capstone, September 2023 Cohort

Many thanks to Springboard mentor Vinit Koshti

The Problem

The NBA

467 active players in
2022-2023 season

Each varies in
skill/experience/salary

Can we estimate a player's
salary given their statistics?

An Optimal Team

Constraints include

- Budget
 - \$136,021,000
- Number of players
 - 15
- Balance of positions
 - min. 2 players per position

**What is the optimal
NBA team?**

Data Information

https://github.com/acody14/Springboard/blob/main/capstone2/capstone2.1_wrangling.ipynb

467 rows (for each player)

31 columns (individual statistics)

Data period: 2022-2023 NBA season

Data Cleaning Challenges

Irrelevant columns

Remove columns

The DataFrame contained a useless index column and a player id column which would not come in useful

Categorical columns

String Manipulation

Categorical columns contained values belonging to multiple classes. This required some string manipulation to make them usable for dummy variable creation.

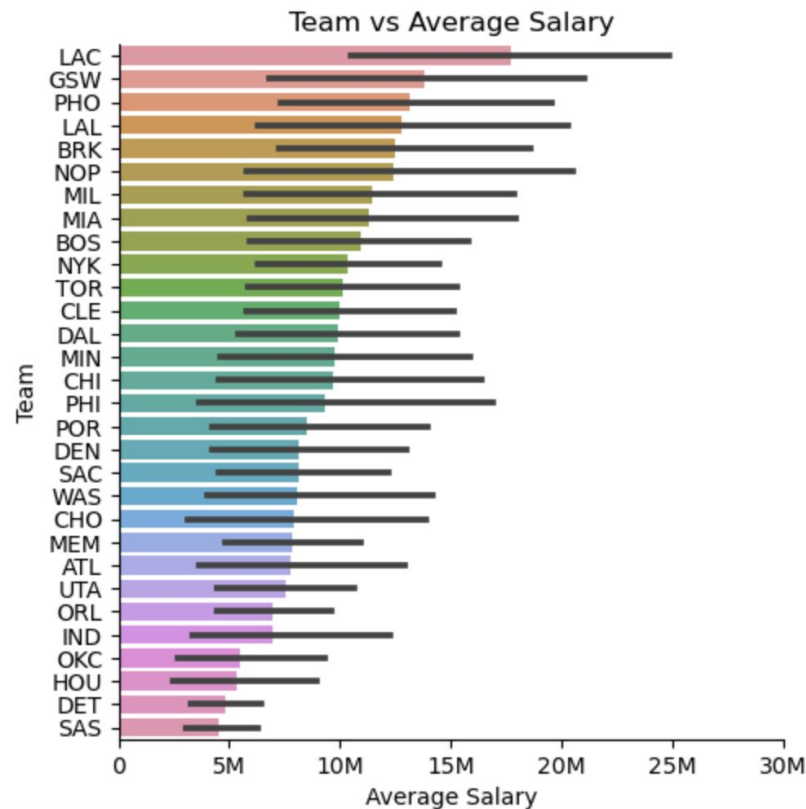
Missing values

Impute with 0

All percentage columns were missing values due to 0 division. First I imputed missing values with 0, and then removed rows for players who had played fewer than 10 games.

Exploratory Data Analysis: categorical

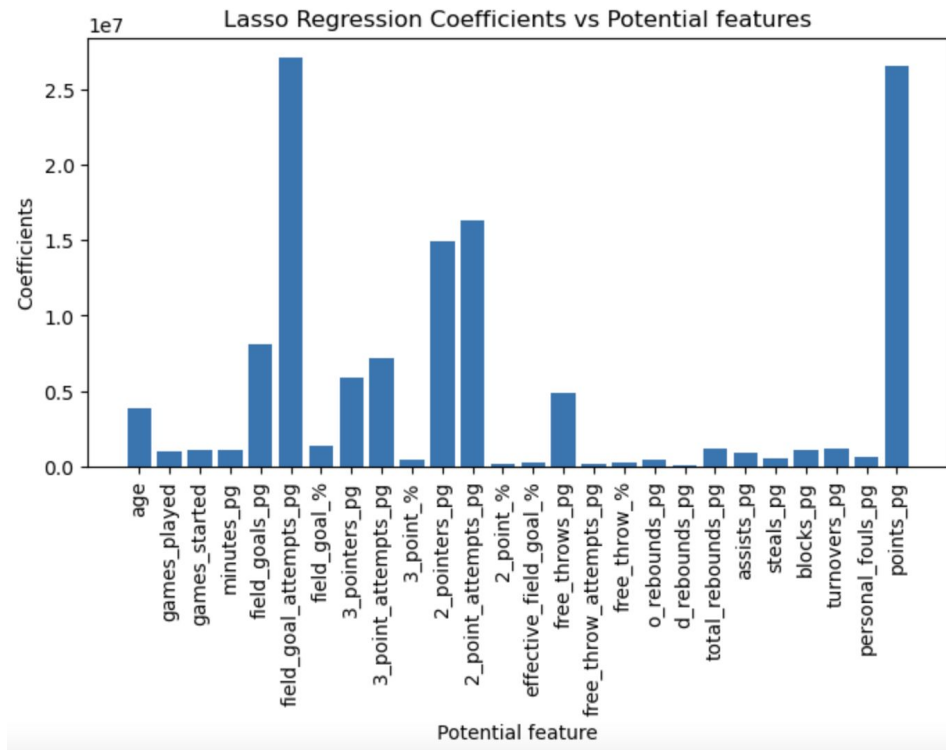
https://github.com/acody14/Springboard/blob/main/capstone2/capstone2.2_eda.ipynb



Exploratory Data Analysis: numerical

https://github.com/acody14/Springboard/blob/main/capstone2/capstone2.2_edu.ipynb

Lasso Regression Machine Learning Technique identifies useful features. Non-useful features will converge to 0



Data Preprocessing

https://github.com/acody14/Springboard/blob/main/capstone2/capstone2.3_preprocessing.ipynb

1. Dummy Variable Creation
2. Split data into train and test sets
3. **Scale Data ->**

Power Transformer

Tested with Linear Regression model

RMSE: \$7,939,877

r^2 : 0.53

Standard Scaler

Tested with Linear Regression model

RMSE: \$7,464,872

r^2 : 0.59

Modeling

https://github.com/acody14/Springboard/blob/main/capstone2/capstone2.4_modeling.ipynb

Lasso Regression

Hyperparameter Tuning
with GridSearchCV

Train r^2 : 0.661

Test r^2 : 0.701

Cross Validated r^2 : 0.643

Random Forest

Hyperparameter Tuning
with Bayesian Optimization

Train r^2 : 0.921

Test r^2 : 0.755

Cross Validated r^2 : **0.720**

XGBoost

Hyperparameter Tuning
with Bayesian Optimization

Train r^2 : 0.956

Test r^2 : 0.724

Cross Validated r^2 : 0.719

Best Model

Random Forest

Hyperparameter Tuning
with Bayesian Optimization

Train r^2 : 0.921

Test r^2 : 0.755

Cross Validated r^2 : 0.720



Hyperparameters:

max_depth: 7.046122532763487

max_features: 4.887569511250316

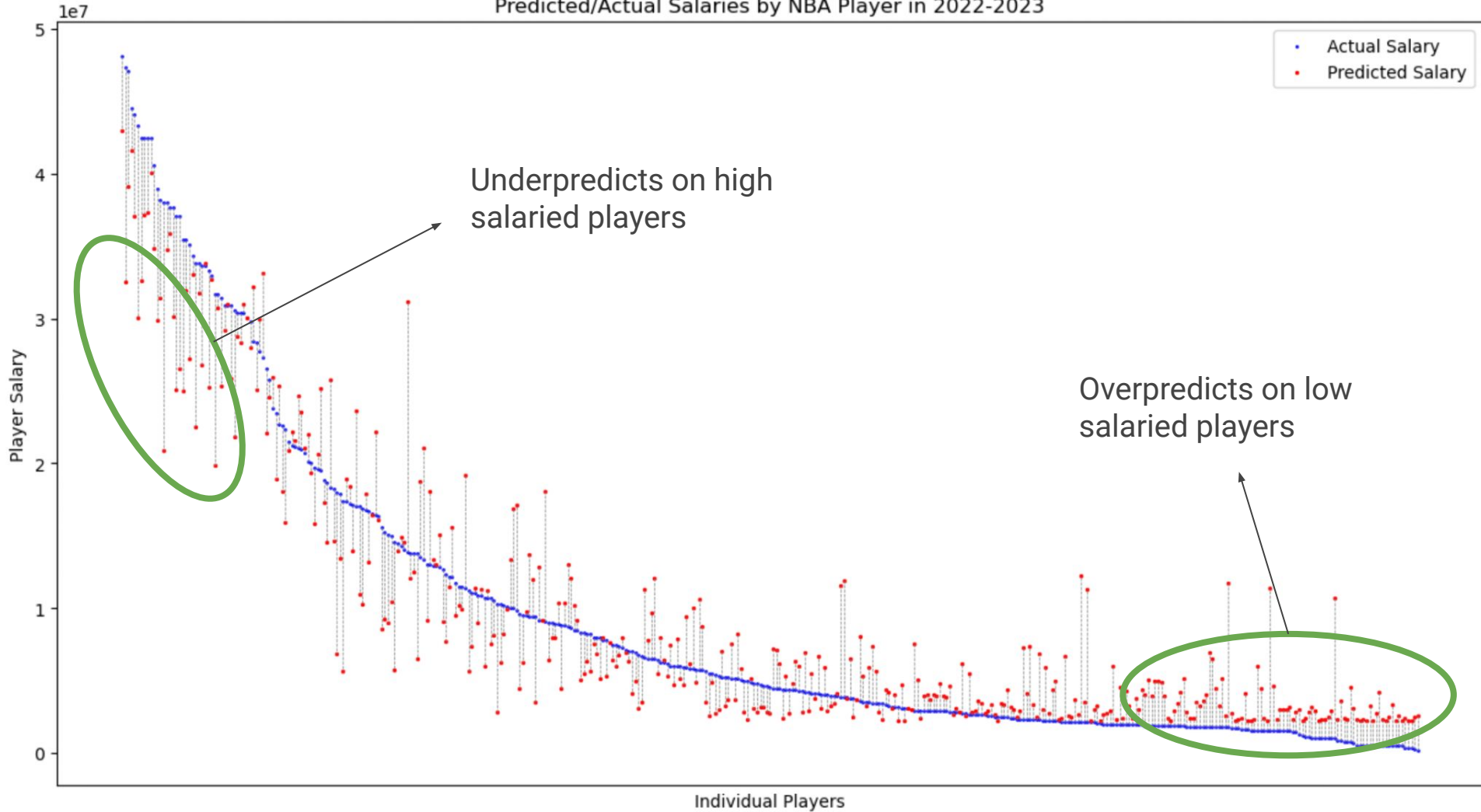
max_leaf_nodes: 29.644614290474742

min_samples_leaf: 1.6204560131132242

min_samples_split: 2.1333262991437074

n_estimators: 495.6940265294837

Predicted/Actual Salaries by NBA Player in 2022-2023



Team Optimizer

Uses Linear Programming from PuLP library

Constraints:

- Budget under \$136,021,000 (salary cap for subsequent NBA season)
- 15 players per team
- A balance of positions on the team - no fewer than 2 players per position

Player: Harrison Barnes
Position: Power Forward
Actual Salary: 18352273
Pred Salary: 25733256

Player: Alec Burks
Position: Shooting Guard
Actual Salary: 10012800
Pred Salary: 16860113

Player: Damion Lee
Position: Shooting Guard
Actual Salary: 2133278
Pred Salary: 12221791

Player: Lauri Markkanen
Position: Power Forward
Actual Salary: 16475454
Pred Salary: 22170036

Player: De'Andre Hunter
Position: Small Forward
Actual Salary: 9835881
Pred Salary: 17151845

Player: Desmond Bane
Position: Shooting Guard
Actual Salary: 2130240
Pred Salary: 11294308

Player: Brook Lopez
Position: Center
Actual Salary: 13906976
Pred Salary: 31123767

Player: Mason Plumlee
Position: Center
Actual Salary: 9080417
Pred Salary: 18079033

Player: Tre Jones
Position: Point Guard
Actual Salary: 1782621
Pred Salary: 11742386

Player: Jordan Clarkson
Position: Shooting Guard
Actual Salary: 13340000
Pred Salary: 21091805

Player: Jordan Poole
Position: Point Guard
Actual Salary: 3901399
Pred Salary: 11555161

Player: Austin Reaves
Position: Shooting Guard
Actual Salary: 1563518
Pred Salary: 11392986

Player: Dillon Brooks
Position: Small Forward
Actual Salary: 11400000
Pred Salary: 19185552

Player: Keldon Johnson
Position: Small Forward
Actual Salary: 3873024
Pred Salary: 11888570

Player: Kris Dunn
Position: Point Guard
Actual Salary: 1000001
Pred Salary: 10702841

Optimal NBA Team

- Team stays within all constraints
- Model estimates that this team is undervalued by over \$130M

Total Prediction Error: -133405568
Total Spent on Salaries: 118787882
Budget: 136021000
Number of Centers: 2
Number of Point Guards: 3
Number of Shooting Guards: 5
Number of Power Forwards: 2
Number of Small Forwards: 3

Conclusion

This model could be generalized to players entering the NBA draft, or could be used to optimize open spots on a team while keeping others constant. To improve the accuracy of the model I could involve an injury likelihood predictor. Even forgoing the optimization section, this model can be used to simply identify how undervalued or overvalued a particular player is, indicating that a player might be a good target for a trade.

Many thanks to Springboard mentor Vinit Koshti and to the Springboard team