



# NBA Salaries

Predicting market value based on  
statistical performance

# King of the Market?

**Goal:** Who is the top free agent in the summer of 2021 based on their season this year? How much will they get paid?

## Kawhi Leonard

- 29 Years Old, Small Forward
- **2x** NBA Champion / Finals MVP
- **5x** NBA All Star, **6x** All NBA Defensive team
- **\$34.3 million** current salary (**11th highest**)

What should he be paid? Is he the best available?





# Model Setup

- Scraped with BeautifulSoup
- Data **2009-2021** / “Start of Analytical Era”
- 55 features** statistics
- Over **6,000** individual players
- Limited data to games played > 4 (5%)

**Target: Predicted Salary**

 Basketball Reference 

Statistics



 Hoops Hype 

Salary

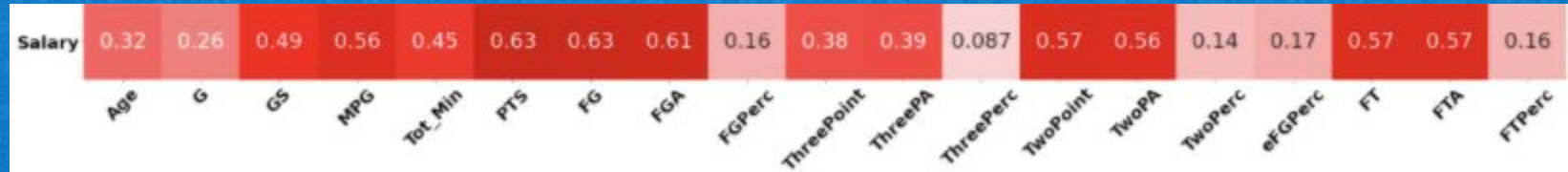


 Model 

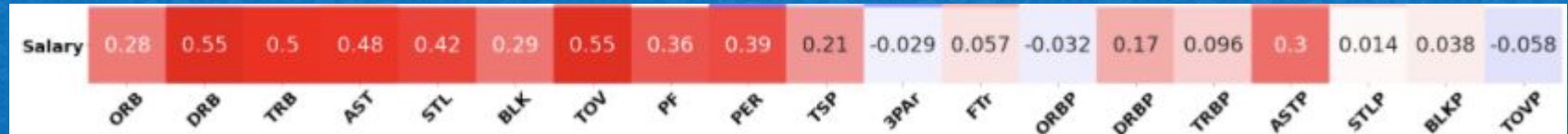


# Features - Points is the MVP

## Availability & Scoring



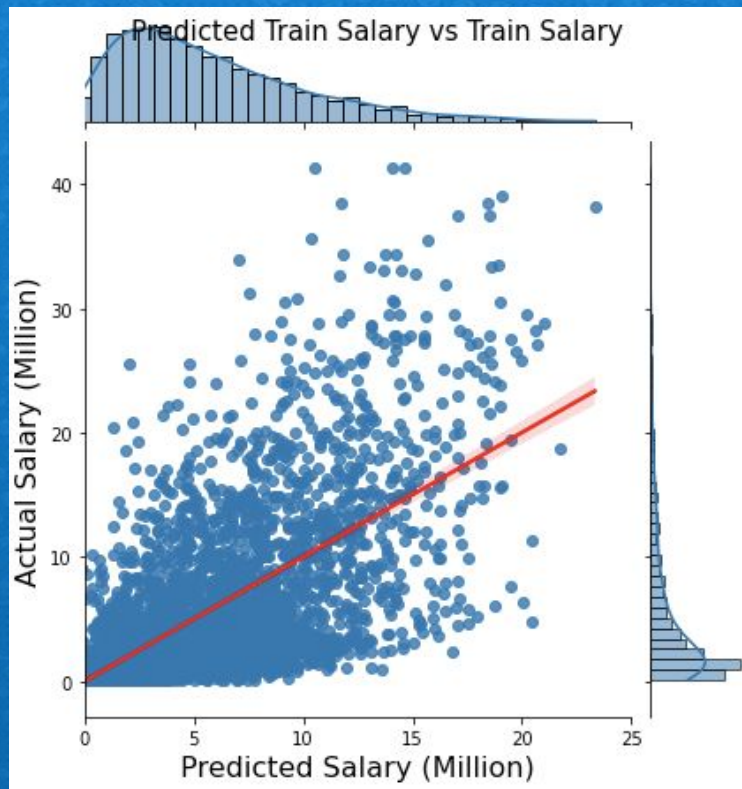
## Rebounding, Passing, Defense and Advanced



## Team, Efficiency & Year



# More To Life Than Scoring



Points to Predict Salary

$$R^2 = .39$$

Mean Error: \$3.5 Million

Next Step: Add **ALL** Features





# Mo' Money Mo' Problems

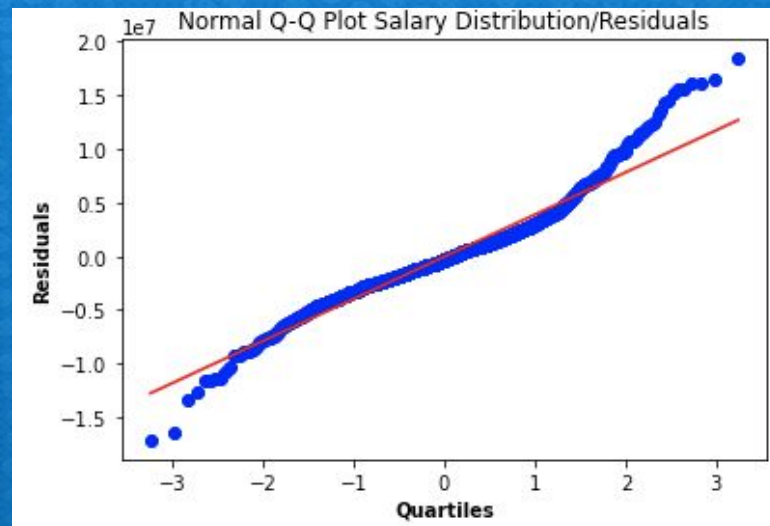
- Linear Regression w all features

$$R^2 = .591$$

Mean Error: \$2.98 Million

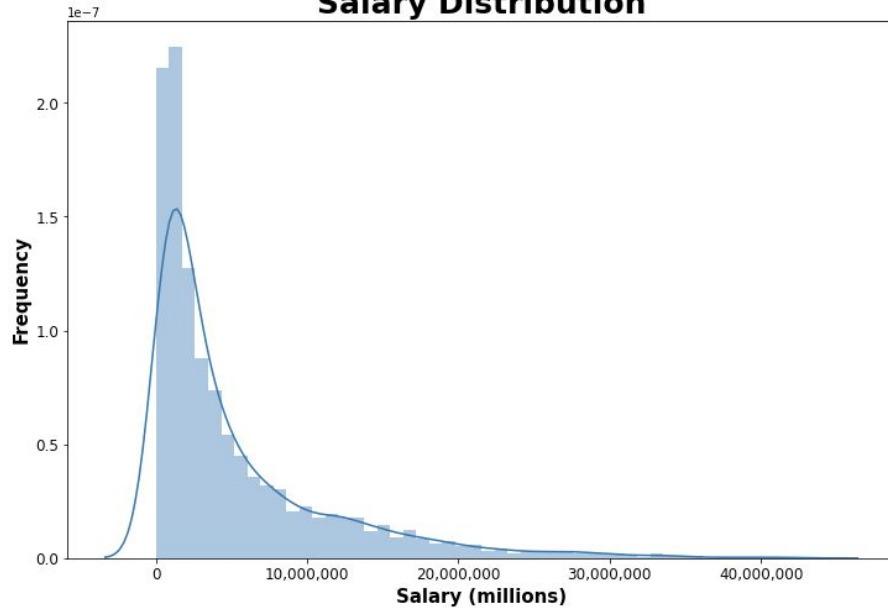
- Salaries at Extremes less accurate

- Skewed Distribution with less sample of high paid players

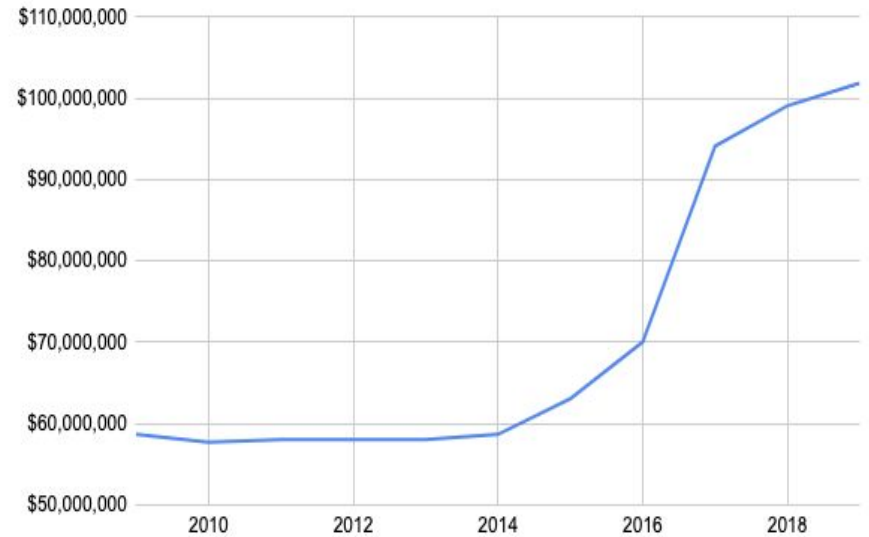


# Growing Revenues

**Salary Distribution**



**NBA Salary Cap**



# Modeling - Ridge Wins At the Buzzer

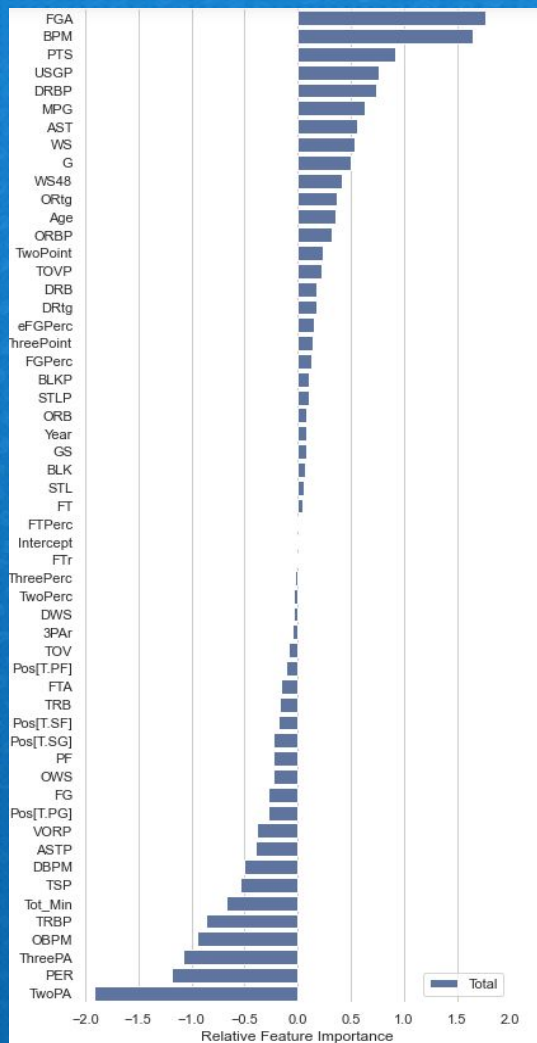
Methodology	$R^2$ (Train/Val)	$R^2$ (Cross Val)
Linear Regression	.5917	.5926
🏆 Ridge Regression 🏆	.5918	.5929
Lasso Regularization	.5916	.5915
Elastic Net	.591	.5912

**Final Ridge Test Score: .5947**

Mean Error: \$2.8 Million







# Top Ridge Features

Field Goal Attempts

Box Plus Minus

Points

Usage

Defensive Rebounding

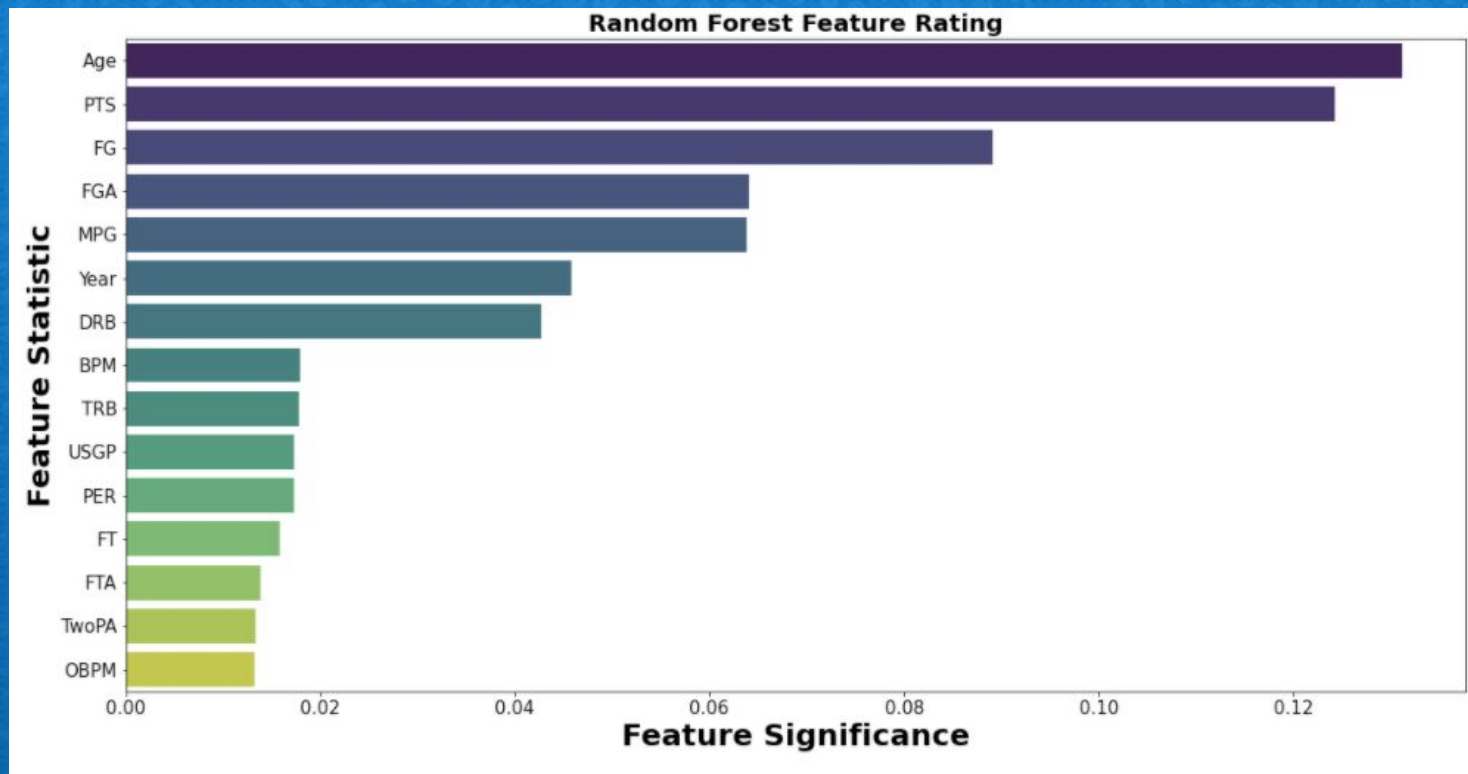
Minutes Per Game

Assists

Win Share

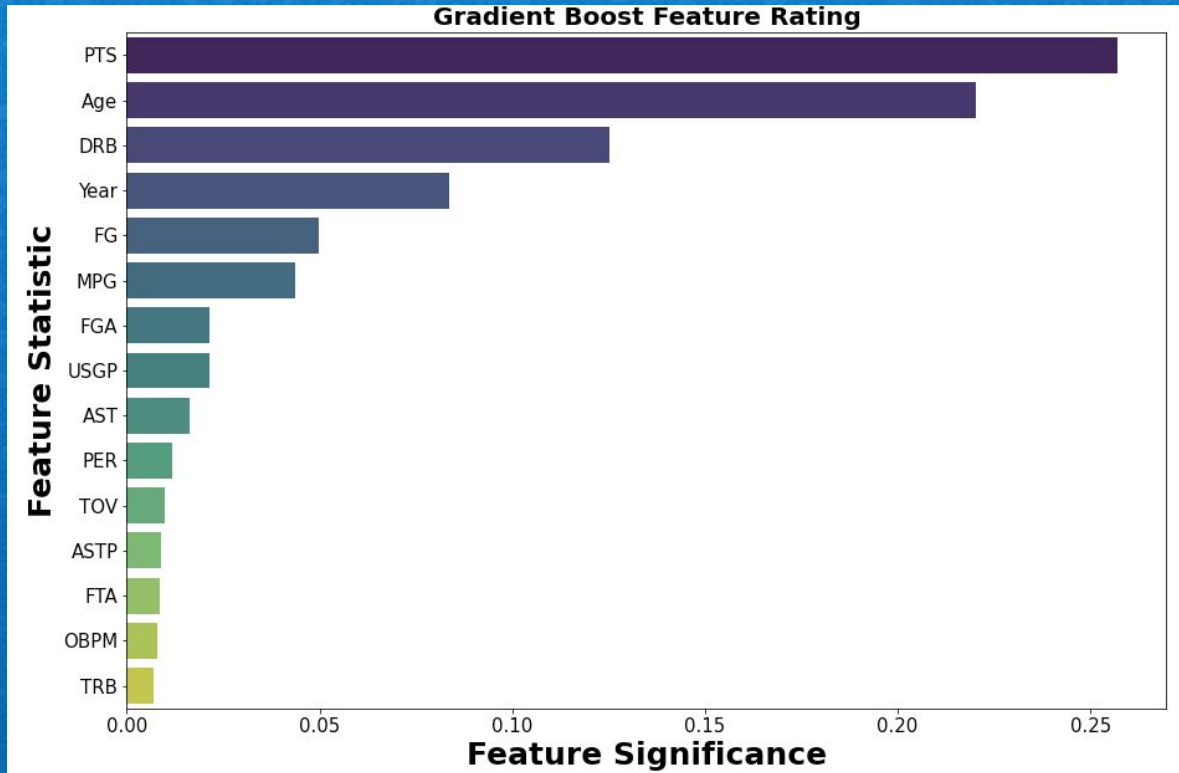
# Random Forest - Feature Importance

Using up to 15 variables Random Forest able to achieve .66  $R^2$  / Error: \$2.42 M



# Gradient Boosting- Feature Importance

Gradient Boost able to achieve the best of .67  $R^2$  / Error: \$2.39 M





# Show Me The Money

## Khwai Model Value

\$24.8 Million +/- 2.8 M vs 34 Current  
+\$2 million vs all possible free agents

## Why So Low?

- Playoff data not in sample/Player Awards
- Defense not valued in model
- Khwai plays limited games to prevent injury
- Brand Value not factored in



# Next Steps

- Bring in postseason data/award elements
- Utilize team data
- Better incorporate Salary cap data/time element
- Explore player revenue off court (i.e. Followers)
- Further explore eliminating lower salaries/outliers

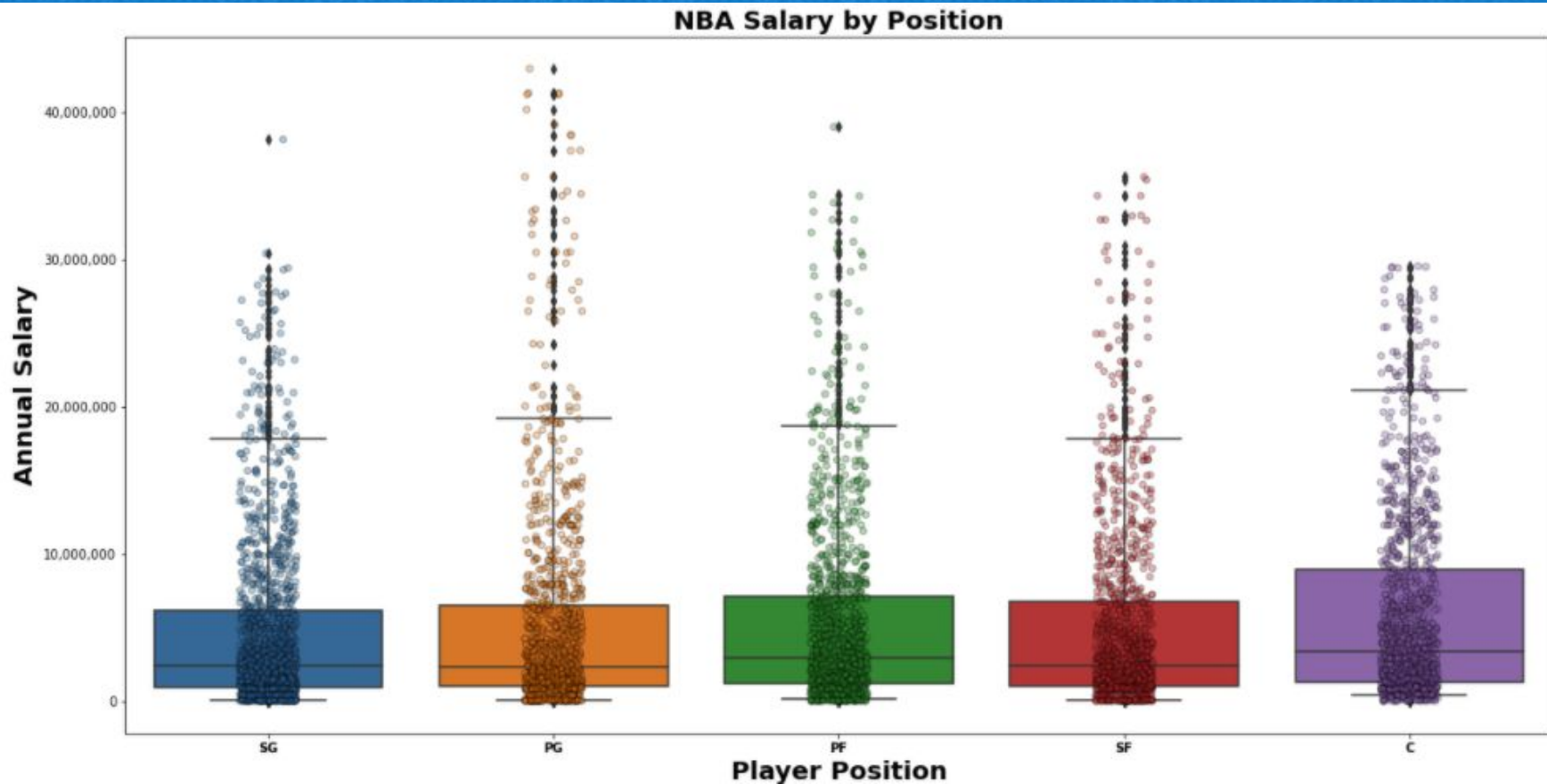


# Appendix



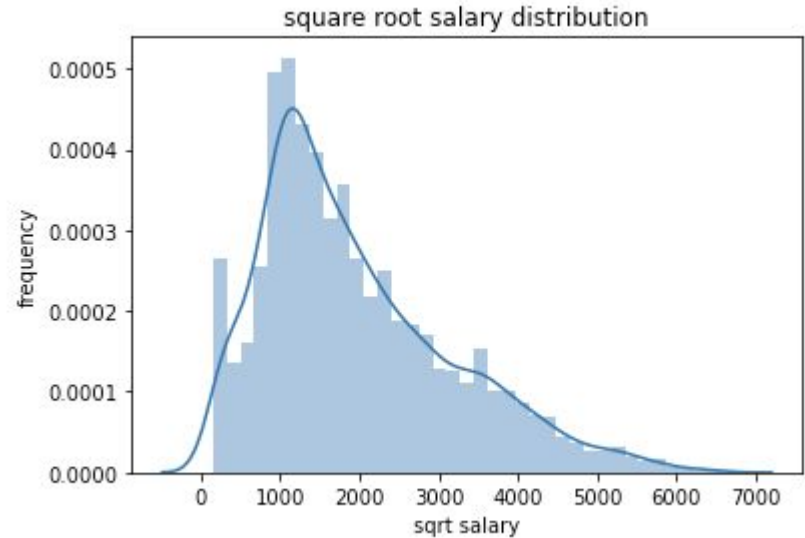
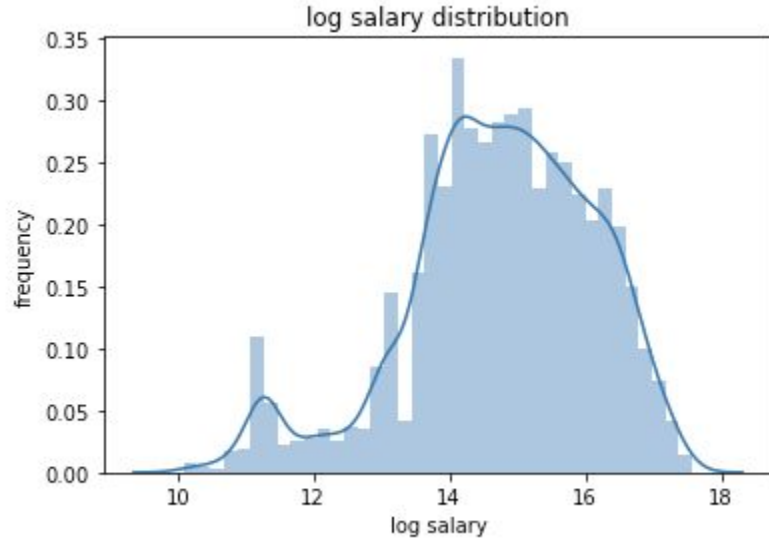


# Positional Distributions

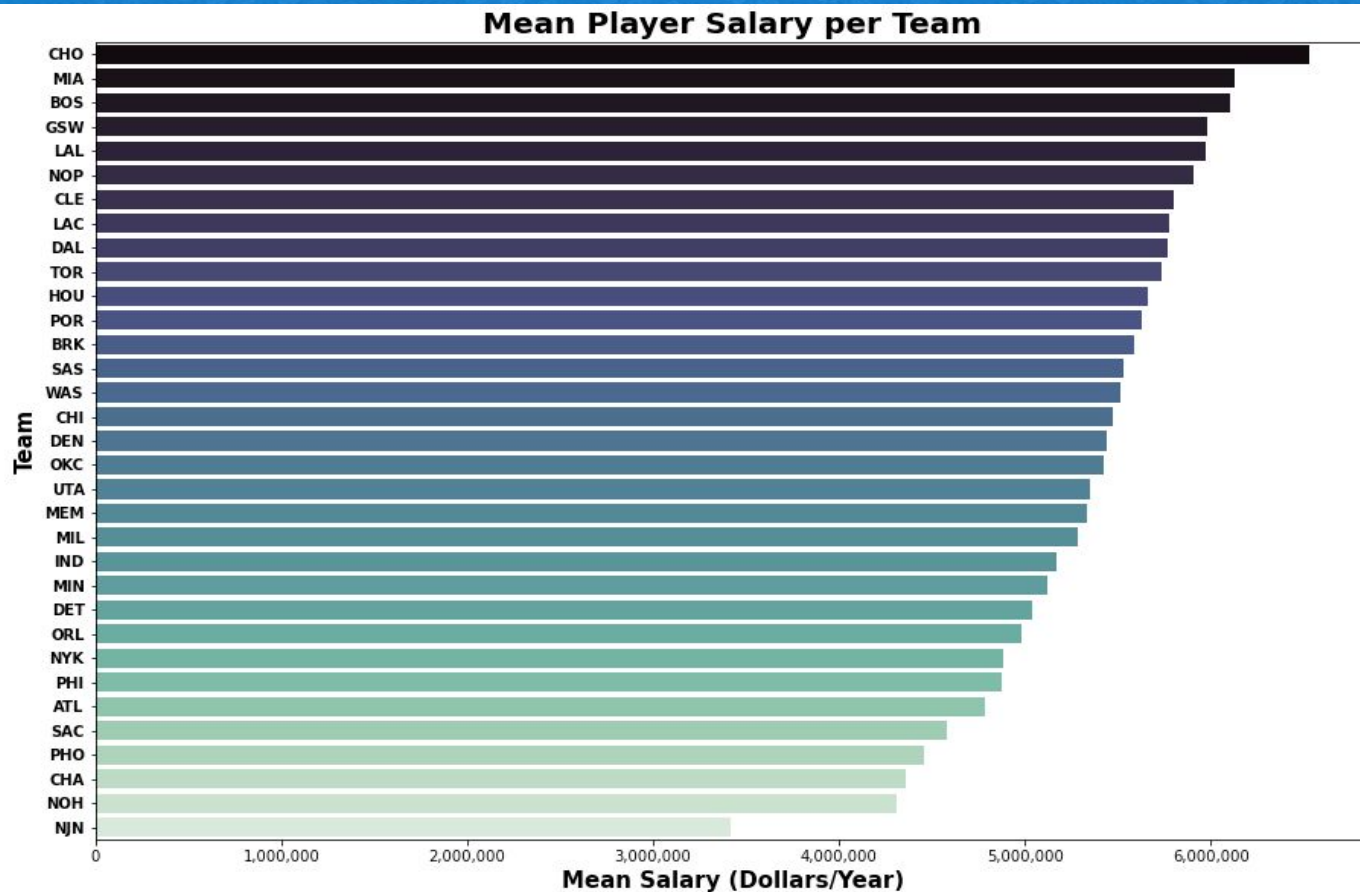


# Log and Square Root Target Distributions

Square root achieved  $R^2$  of .612 (Ridge)



# EDA - Teams





# Prediction vs Current - Free Agents

Player	Predicted	Current
John Collins	\$12.9	\$2.7
Mike Conley	\$18.9	\$34
Victor Oladipo	\$19.3	\$21
Demar Derozan	\$22.7	\$27.7
Serge Ibaka	\$14.1	\$9.2

\*\*\* Values in Millions



# Example Distributions From Data

