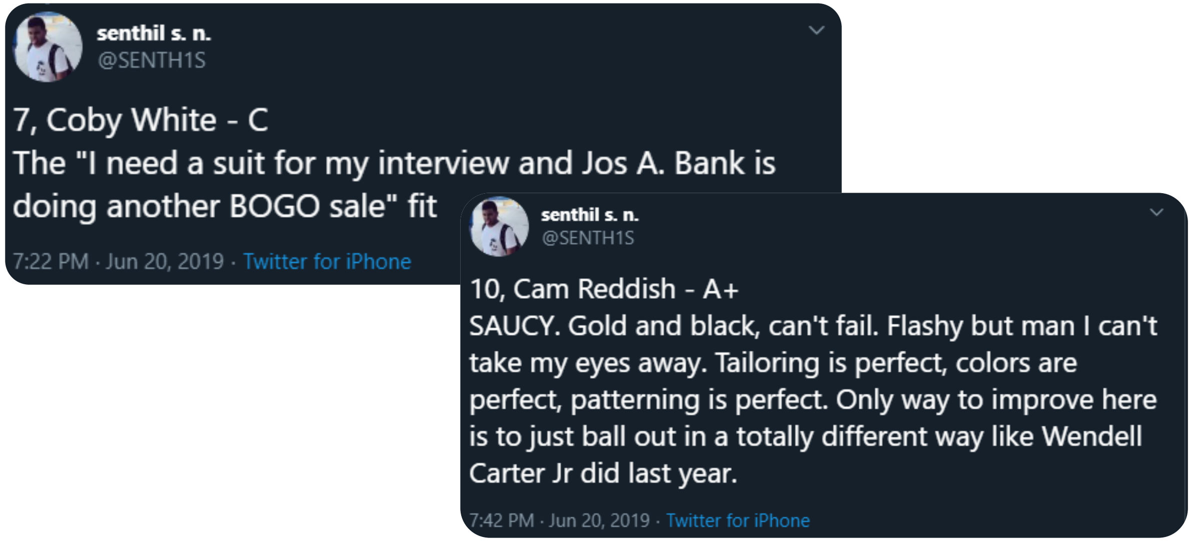
A MIXED EFFECTS MODELING APPROACH TO PREDICTING NBA FREE AGENCY

SENTHIL S. NATARAJAN | RITSAC 2019

I cook...

I offer unsolicited fashion advice...



senthil s. n.
@SENTH1S

7, Coby White - C
The "I need a suit for my interview and Jos A. Bank is doing another BOGO sale" fit

7:22 PM · Jun 20, 2019 · Twitter for iPhone

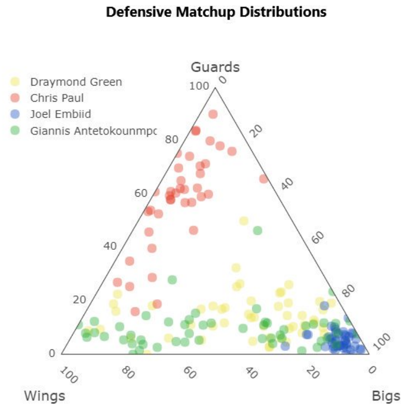
senthil s. n.
@SENTH1S

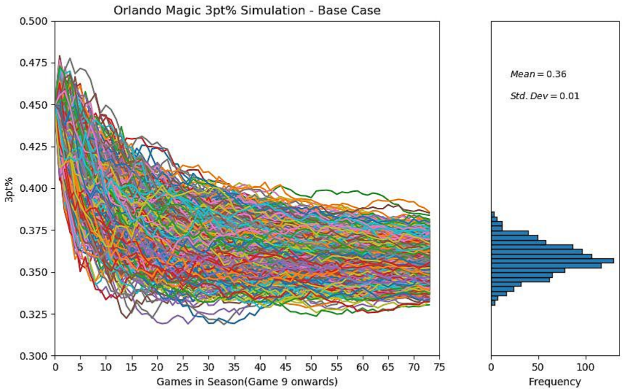
10, Cam Reddish - A+
SAUCY. Gold and black, can't fail. Flashy but man I can't take my eyes away. Tailoring is perfect, colors are perfect, patterning is perfect. Only way to improve here is to just ball out in a totally different way like Wendell Carter Jr did last year.

7:42 PM · Jun 20, 2019 · Twitter for iPhone

I draw things...

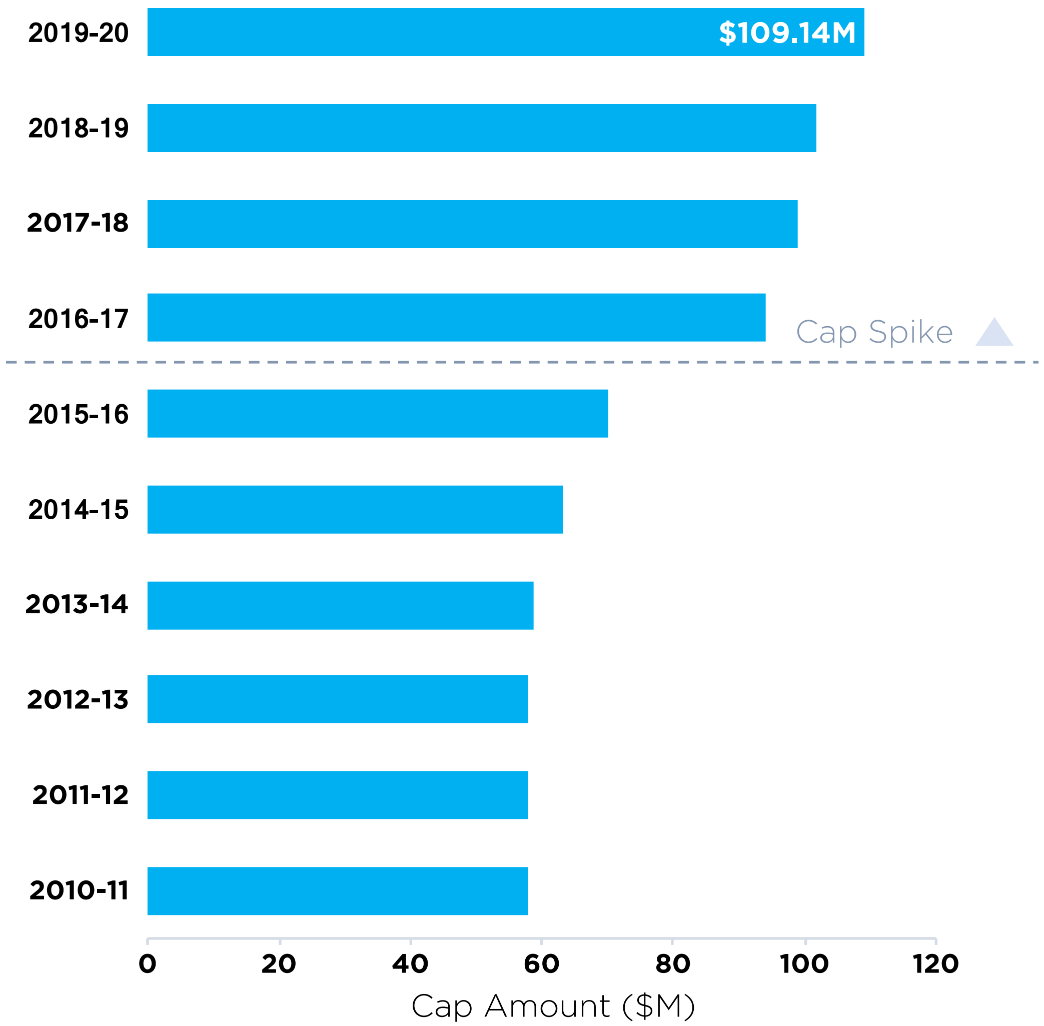
I do some basketball analytics also.



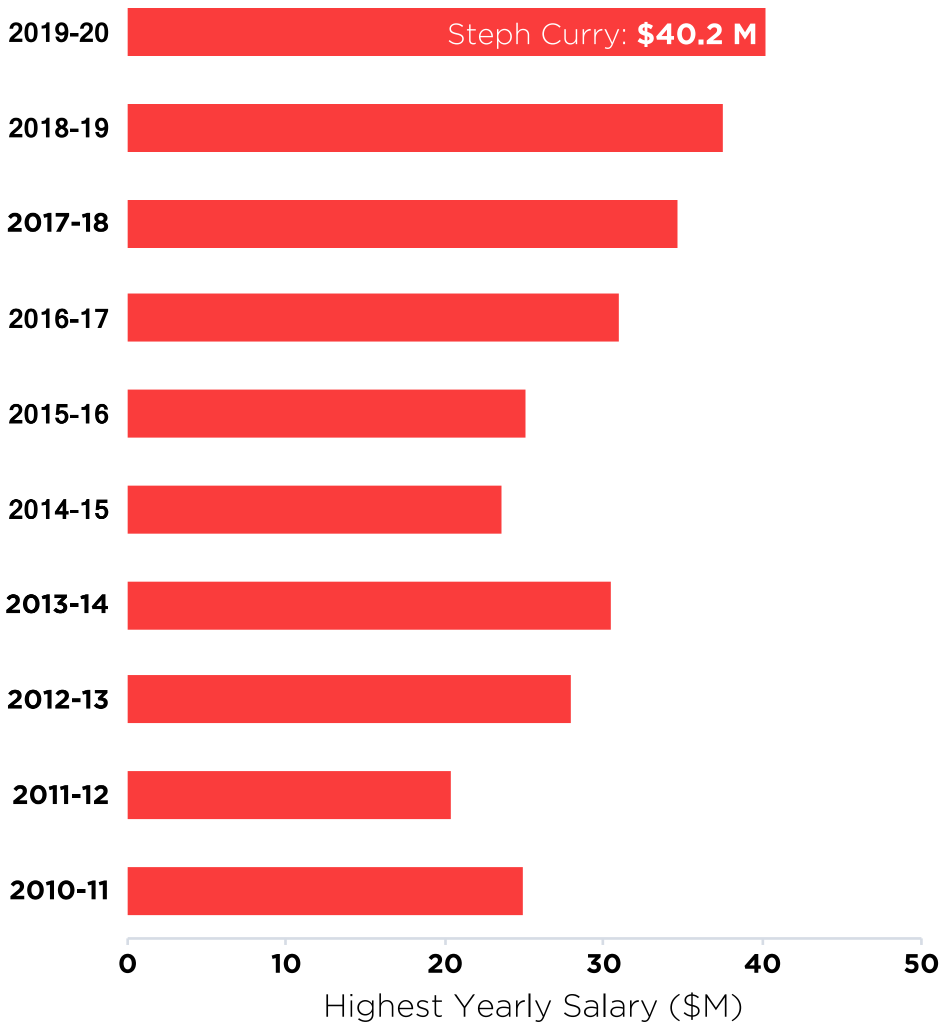


There's more money in the NBA than ever before… which makes properly utilizing that money more important than ever before

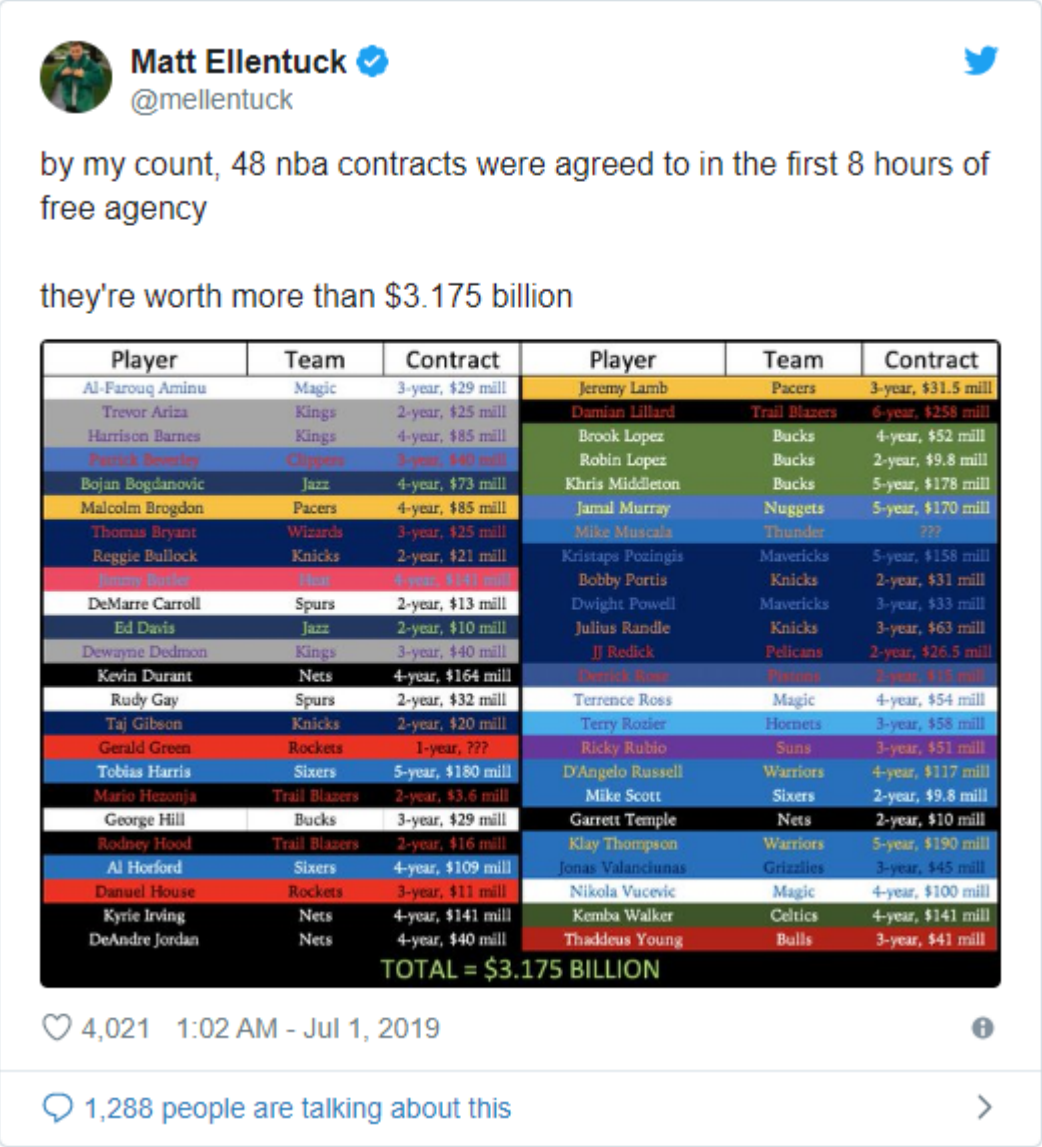
Rise in the NBA's Salary Cap



Rise in the NBA's Contract Values



Data per Sportrac



Matt Ellentuck @mellentuck

by my count, 48 nba contracts were agreed to in the first 8 hours of free agency

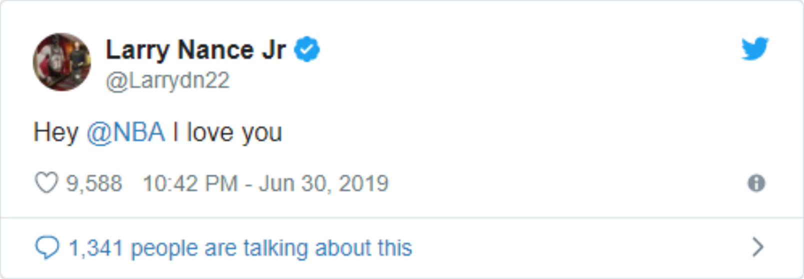
they're worth more than $3.175 billion

Player	Team	Contract	Player	Team	Contract
Al-Farouq Aminu	Magic	3-year, $29 mill	Jeremy Lamb	Pacers	3-year, $31.5 mill
Trevor Ariza	Kings	2-year, $25 mill	Damian Lillard	Trail Blazers	6-year, $258 mill
Harrison Barnes	Kings	4-year, $85 mill	Brook Lopez	Bucks	4-year, $52 mill
Patrick Beverley	Clippers	3-year, $40 mill	Robin Lopez	Bucks	2-year, $9.8 mill
Bojan Bogdanovic	Jazz	4-year, $73 mill	Khris Middleton	Bucks	5-year, $178 mill
Malcolm Brogdon	Pacers	4-year, $85 mill	Jamal Murray	Nuggets	5-year, $170 mill
Thomas Bryant	Wizards	3-year, $25 mill	Mike Muscala	Thunder	???
Reggie Bullock	Knicks	2-year, $21 mill	Kristaps Pozingis	Mavericks	5-year, $158 mill
Jimmy Butler	Heat	4-year, $141 mill	Bobby Portis	Knicks	2-year, $31 mill
DeMarre Carroll	Spurs	2-year, $13 mill	Dwight Powell	Mavericks	3-year, $33 mill
Ed Davis	Jazz	2-year, $10 mill	Julius Randle	Knicks	3-year, $63 mill
Dewayne Dedmon	Kings	3-year, $40 mill	JJ Redick	Pelicans	2-year, $26.5 mill
Kevin Durant	Nets	4-year, $164 mill	Derrick Rose	Pistons	2-year, $15 mill
Rudy Gay	Spurs	2-year, $32 mill	Terrence Ross	Magic	4-year, $54 mill
Taj Gibson	Knicks	2-year, $20 mill	Terry Rozier	Hornets	3-year, $58 mill
Gerald Green	Rockets	1-year, ???	Ricky Rubio	Suns	3-year, $51 mill
Tobias Harris	Sixers	5-year, $180 mill	D'Angelo Russell	Warriors	4-year, $117 mill
Mario Hezonja	Trail Blazers	2-year, $3.6 mill	Mike Scott	Sixers	2-year, $9.8 mill
George Hill	Bucks	3-year, $29 mill	Garrett Temple	Nets	2-year, $10 mill
Rodney Hood	Trail Blazers	2-year, $16 mill	Klay Thompson	Warriors	5-year, $190 mill
Al Horford	Sixers	4-year, $109 mill	Jonas Valanciunas	Grizzlies	3-year, $45 mill
Danuel House	Rockets	3-year, $11 mill	Nikola Vucevic	Magic	4-year, $100 mill
Kyrie Irving	Nets	4-year, $141 mill	Kemba Walker	Celtics	4-year, $141 mill
DeAndre Jordan	Nets	4-year, $40 mill	Thaddeus Young	Bulls	3-year, $41 mill

TOTAL = $3.175 BILLION

♡ 4,021 1:02 AM - Jul 1, 2019

1,288 people are talking about this

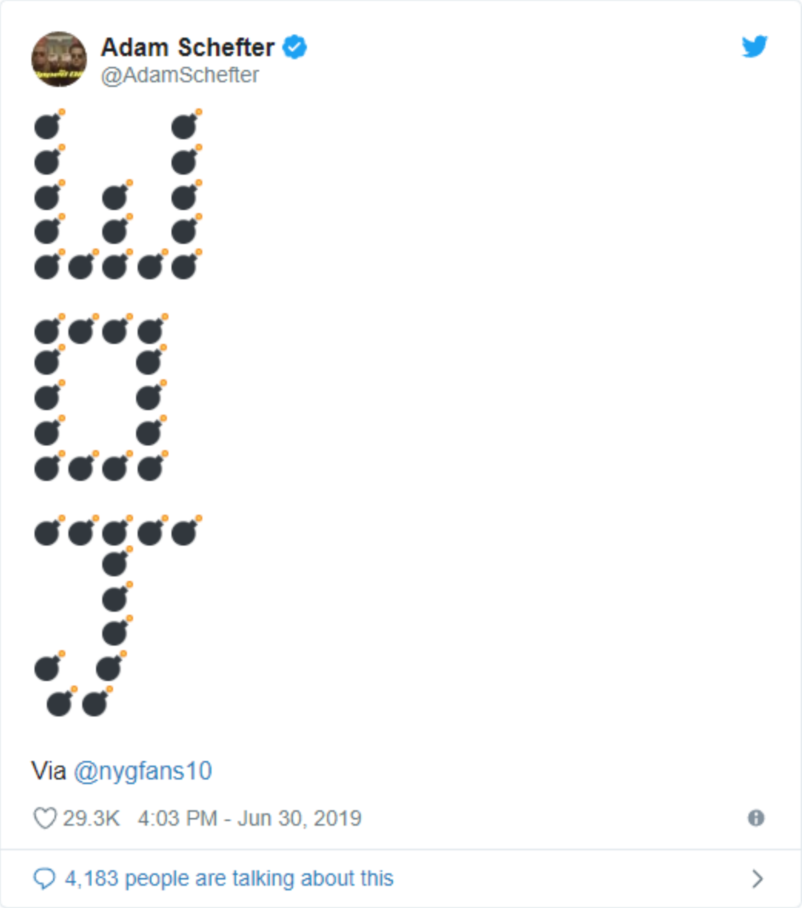


Larry Nance Jr @Larrydn22

Hey @NBA I love you

♡ 9,588 10:42 PM - Jun 30, 2019

1,341 people are talking about this



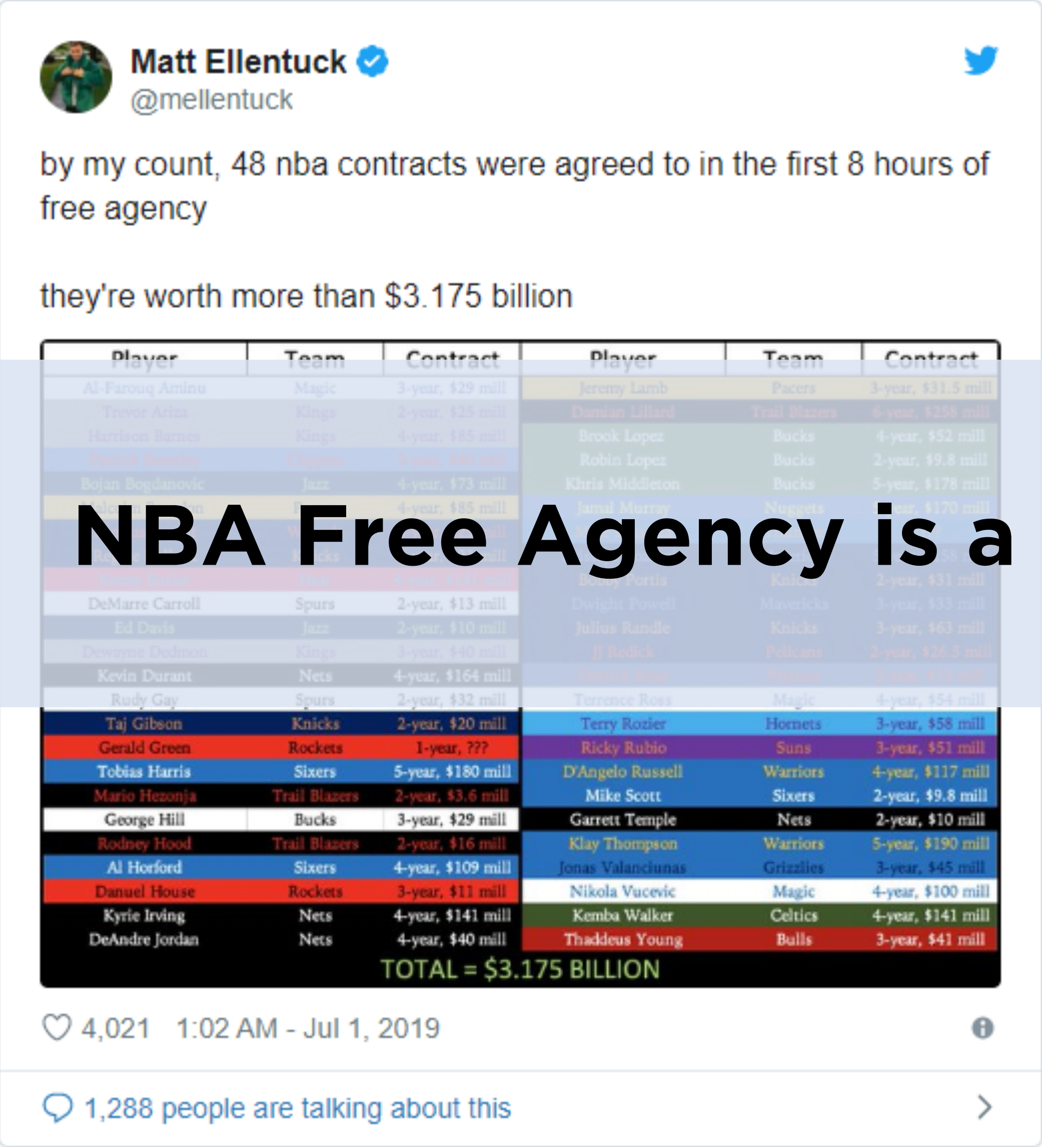
Adam Schefter @AdamSchefter

Via @nygfans10

♡ 29.3K 4:03 PM - Jun 30, 2019

4,183 people are talking about this

NBA Free Agency is a big (read: lucrative) deal.



Matt Ellentuck @mellentuck

by my count, 48 nba contracts were agreed to in the first 8 hours of free agency

they're worth more than $3.175 billion

Player	Team	Contract	Player	Team	Contract
Al-Farouq Aminu	Magic	3-year, $29 mill	Jeremy Lamb	Pacers	3-year, $31.5 mill
Trevor Ariza	Kings	2-year, $25 mill	Damian Lillard	Trail Blazers	6-year, $258 mill
Harrison Barnes	Kings	4-year, $85 mill	Brook Lopez	Bucks	4-year, $52 mill
Patrick Beverley	Clippers	3-year, $40 mill	Robin Lopez	Bucks	2-year, $9.8 mill
Bojan Bogdanovic	Jazz	4-year, $73 mill	Khris Middleton	Bucks	5-year, $178 mill
Malcolm Brogdon	Pacers	4-year, $85 mill	Jamal Murray	Nuggets	5-year, $170 mill
[illegible]	W[illegible]	[illegible] mill	[illegible]	[illegible]	[illegible]
Reggie Bullock	Knicks	[illegible] mill	[illegible]	[illegible]	[illegible] $158 mill
Jimmy Butler	Heat	4-year, $141 mill	Bobby Portis	Knicks	2-year, $31 mill
DeMarre Carroll	Spurs	2-year, $13 mill	Dwight Powell	Mavericks	3-year, $33 mill
Ed Davis	Jazz	2-year, $10 mill	Julius Randle	Knicks	3-year, $63 mill
Dewayne Dedmon	Kings	3-year, $40 mill	JJ Redick	Pelicans	2-year, $26.5 mill
Kevin Durant	Nets	4-year, $164 mill	Derrick Rose	Pistons	2-year, $15 mill
Rudy Gay	Spurs	2-year, $32 mill	Terrence Ross	Magic	4-year, $54 mill
Taj Gibson	Knicks	2-year, $20 mill	Terry Rozier	Hornets	3-year, $58 mill
Gerald Green	Rockets	1-year, ???	Ricky Rubio	Suns	3-year, $51 mill
Tobias Harris	Sixers	5-year, $180 mill	D'Angelo Russell	Warriors	4-year, $117 mill
Mario Hezonja	Trail Blazers	2-year, $3.6 mill	Mike Scott	Sixers	2-year, $9.8 mill
George Hill	Bucks	3-year, $29 mill	Garrett Temple	Nets	2-year, $10 mill
Rodney Hood	Trail Blazers	2-year, $16 mill	Klay Thompson	Warriors	5-year, $190 mill
Al Horford	Sixers	4-year, $109 mill	Jonas Valanciunas	Grizzlies	3-year, $45 mill
Danuel House	Rockets	3-year, $11 mill	Nikola Vucevic	Magic	4-year, $100 mill
Kyrie Irving	Nets	4-year, $141 mill	Kemba Walker	Celtics	4-year, $141 mill
DeAndre Jordan	Nets	4-year, $40 mill	Thaddeus Young	Bulls	3-year, $41 mill

TOTAL = $3.175 BILLION

♡ 4,021 1:02 AM - Jul 1, 2019

1,288 people are talking about this

Larry Nance Jr @Larrydn22

Hey @NBA I love you

♡ 9,588 10:42 PM - Jun 30, 2019

1,341 people are talking about this

Adam Schefter @AdamSchefter

Via @nygfans10

♡ 29.3K 4:03 PM - Jun 30, 2019

4,183 people are talking about this

Some shout-outs before we get started…

For the inspiration:

+ The Younggren twins for their NHL Free Agency Model
+ Andrew Johnson, Nylon Calculus, for his research on predictive factors of contract value
+ U.C. Berkeley Sports Analytics Group for their contract classification model

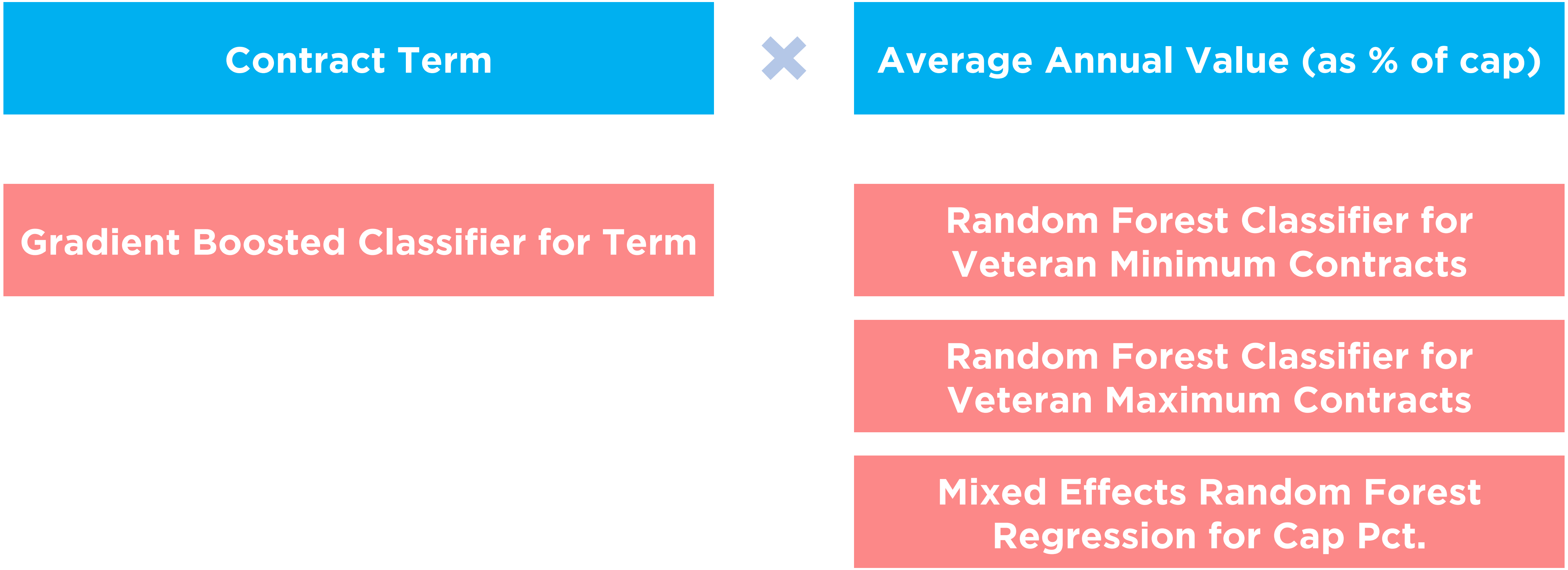
For the data:

+ Basketball Reference for the player stats
+ RealGM and Sportrac for the salary cap data
+ [Unnamed NBA team] for the historical database of NBA contracts

For the models:

+ Hajjem, Bellavance, Larocque et al. for their research on Non-Linear Mixed Effects modeling
+ The data science team at Manifold for their Python implementation of Mixed Effects Random Forests

There's two primary aspects to prediction of a player contract: Term and Money



Contract Term × Average Annual Value (as % of cap)

Gradient Boosted Classifier for Term

Random Forest Classifier for Veteran Minimum Contracts

Random Forest Classifier for Veteran Maximum Contracts

Mixed Effects Random Forest Regression for Cap Pct.

Data preparation for all models involved creating a weighted vector of stats per player

Contract Term	
Year N	60%
Year N-1	30%
Year N-2	10%

Average Annual Value (as % of cap)	
Year N	80%
Year N-1	10%
Year N-2	10%

+ Training data set: All non-rookie player contracts between 2009 and 2019 *(n=2870)*
+ OOS data set: Player contracts in 2019 Free Agency, w/ 2018-19 Win Shares >= 1 *(n=109)*

Contract Term Model

Input dataset for Contract Term model

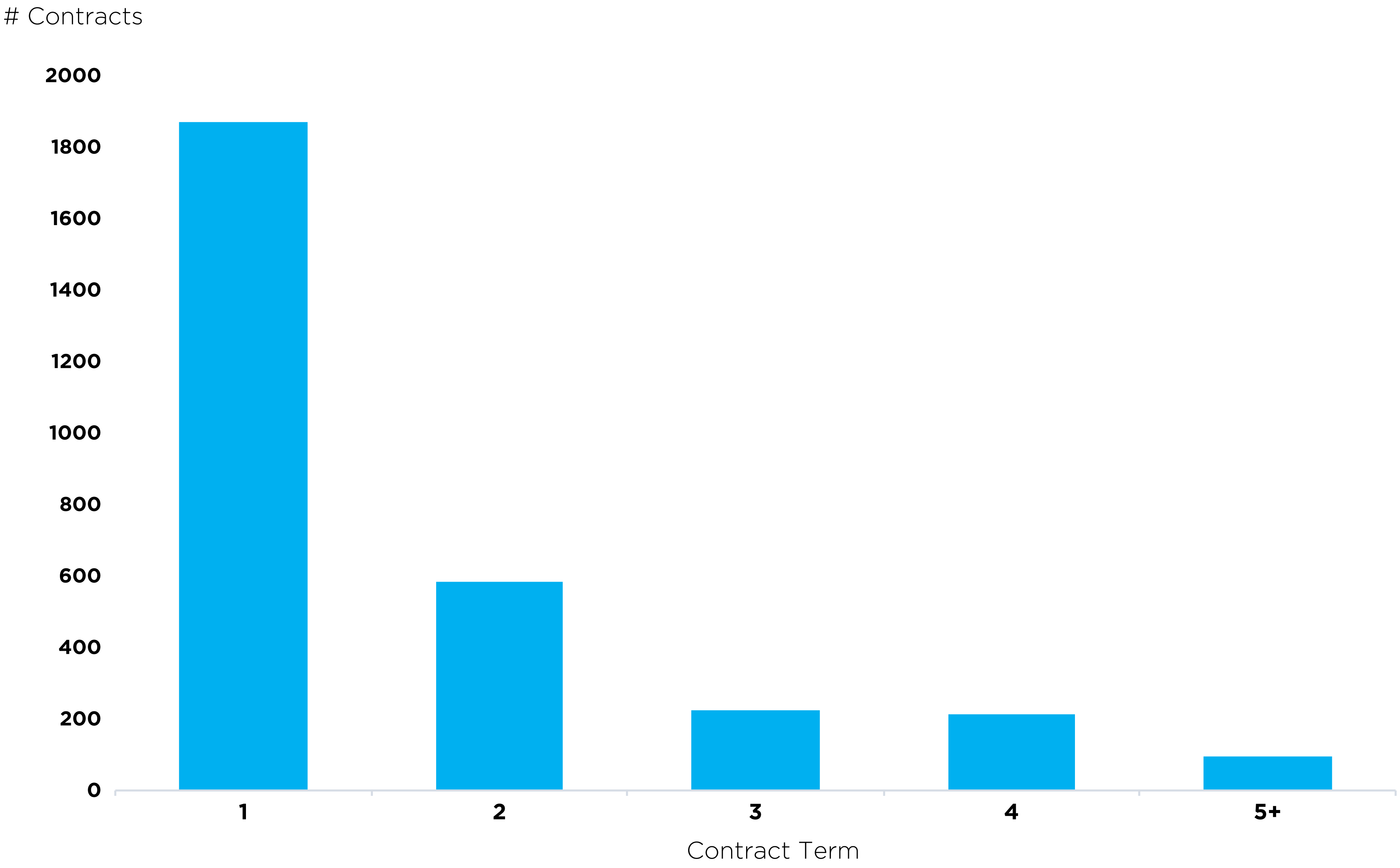
Biographical Variables:

+ Height
+ Weight
+ Draft Position
+ Age
+ Position (one hot encoded)
+ Contract Type (UFA, RFA, Extension)

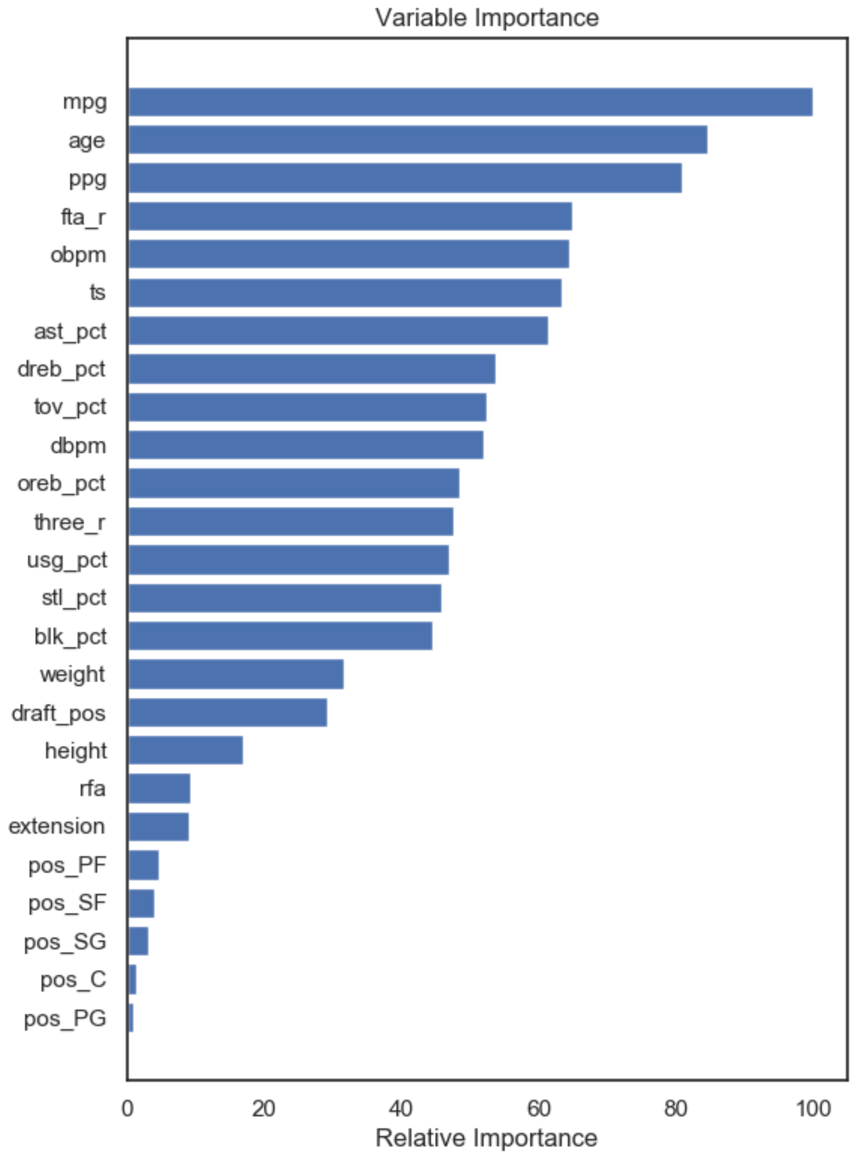
Player Statistics:

+ True Shooting %
+ Free Throw Attempt Rate
+ Three Point Attempt Rate
+ Offensive Rebound Pct.
+ Defensive Rebound Pct.
+ Assist Rate
+ Steal Rate
+ Block Rate
+ Usage Rate
+ Turnover Rate
+ Offensive Box Plus-Minus
+ Defensive Box Plus-Minus
+ Points per Game
+ Minutes per Game

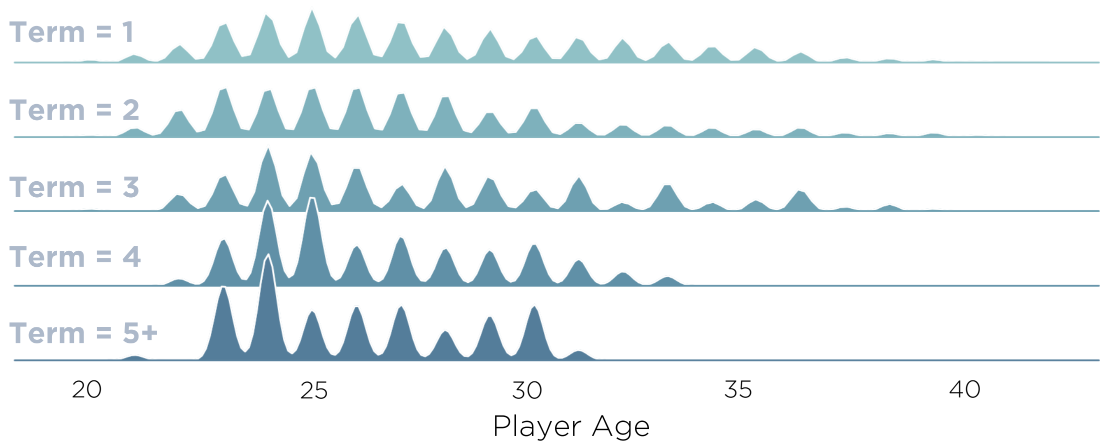
The distribution of player contracts is skewed towards short term deals



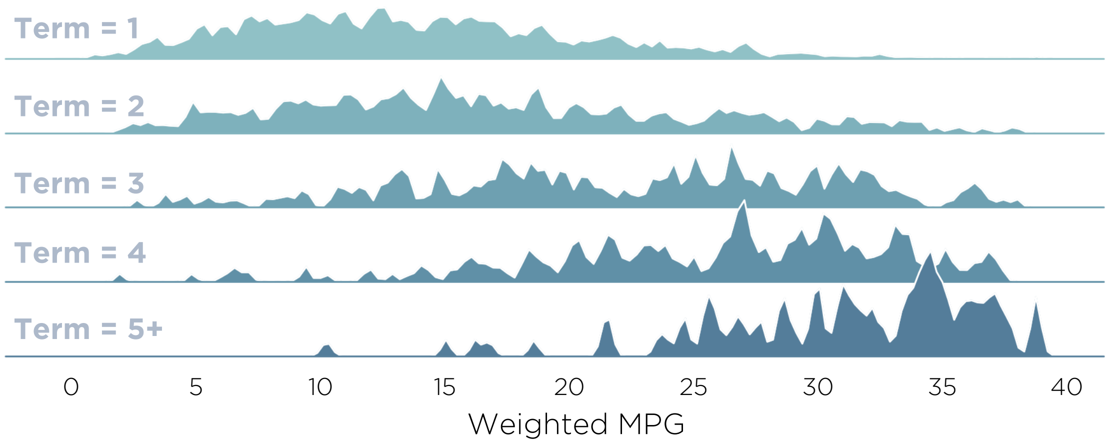
A player's usage and age are major factors in determining the length of their contracts



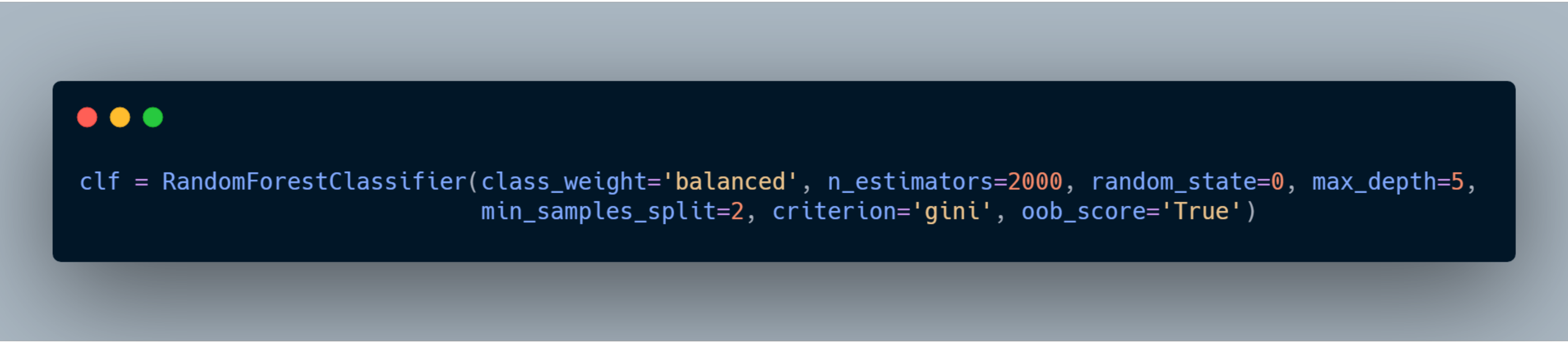
Distribution of Player Age by Contract Length



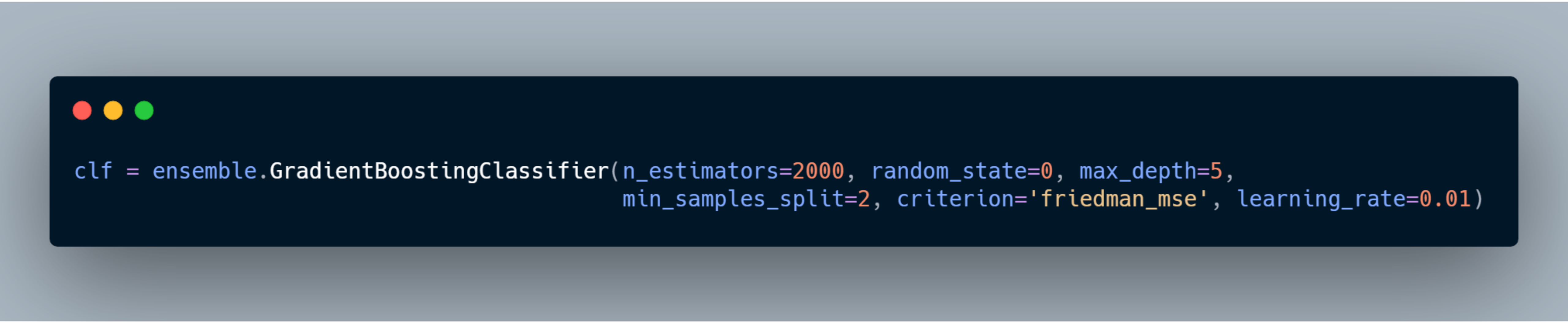
Distribution of Weighted MPG by Contract Length



The skewness of contract length led to first exploring using a class-weighted random forest, but a GBM produced decisively better results, both in and out of sample

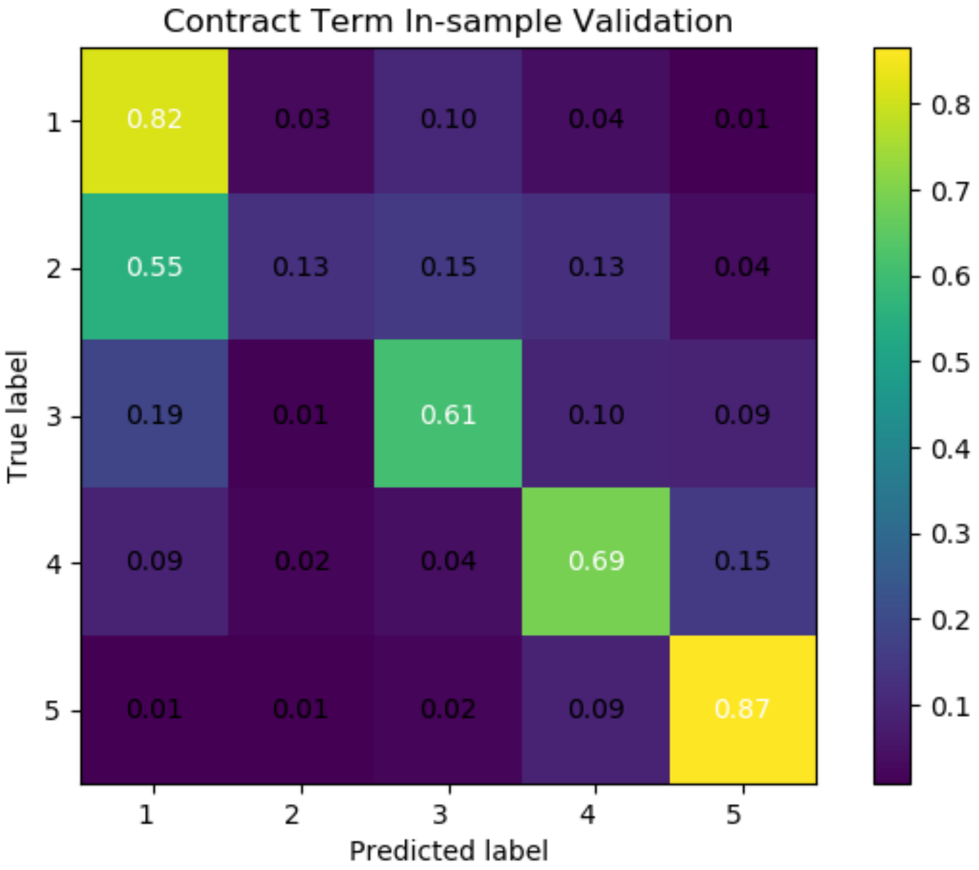


```
clf = RandomForestClassifier(class_weight='balanced', n_estimators=2000, random_state=0, max_depth=5,
                             min_samples_split=2, criterion='gini', oob_score='True')
```

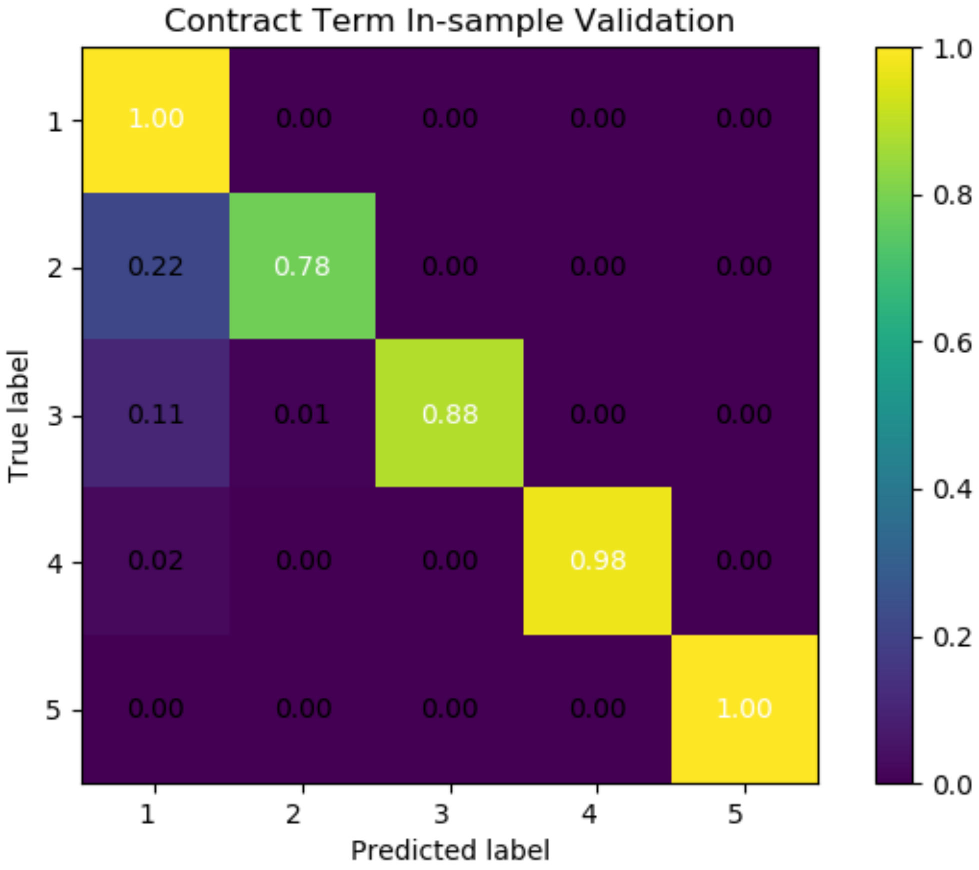


```
clf = ensemble.GradientBoostingClassifier(n_estimators=2000, random_state=0, max_depth=5,
                                          min_samples_split=2, criterion='friedman_mse', learning_rate=0.01)
```

The skewness of contract length led to first exploring using a class-weighted random forest, but a GBM produced decisively better results, both in and out of sample



Random Forest

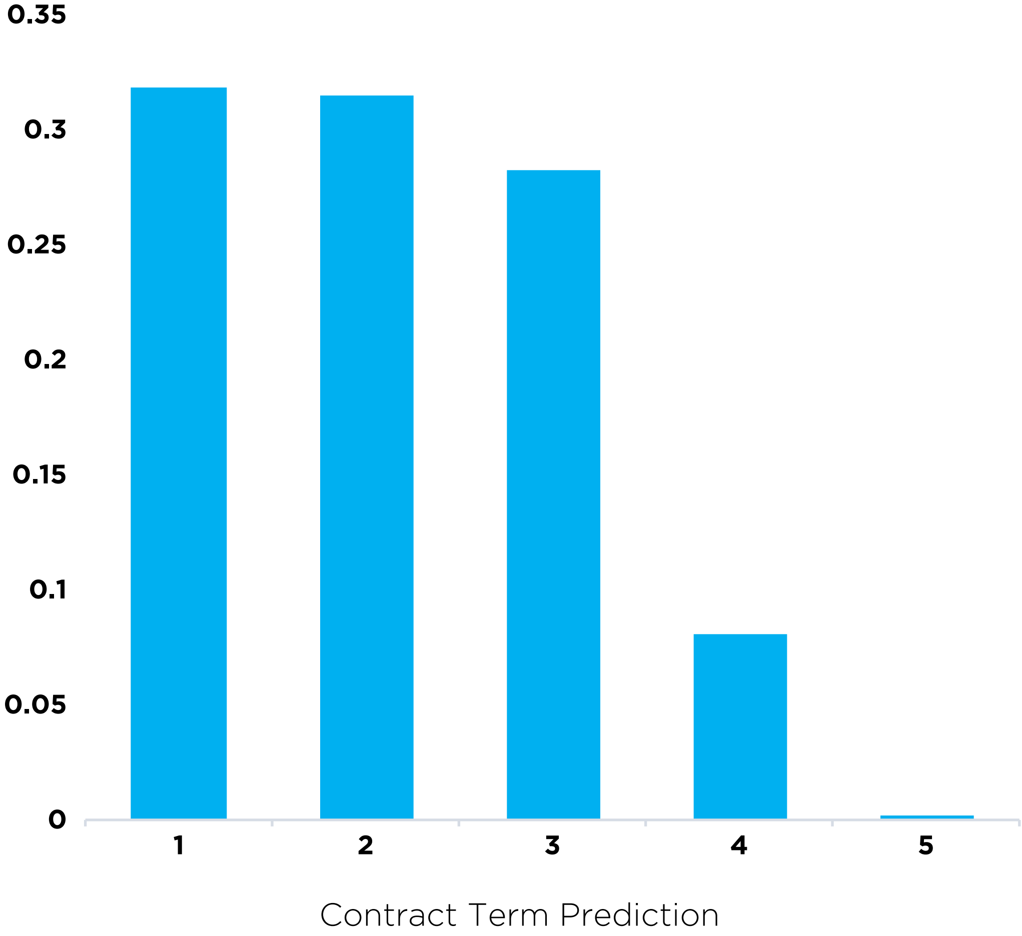


Gradient Boosted Decision Tree

In order to improve the out of sample prediction, I dug into the class probability predictions and created an overlay

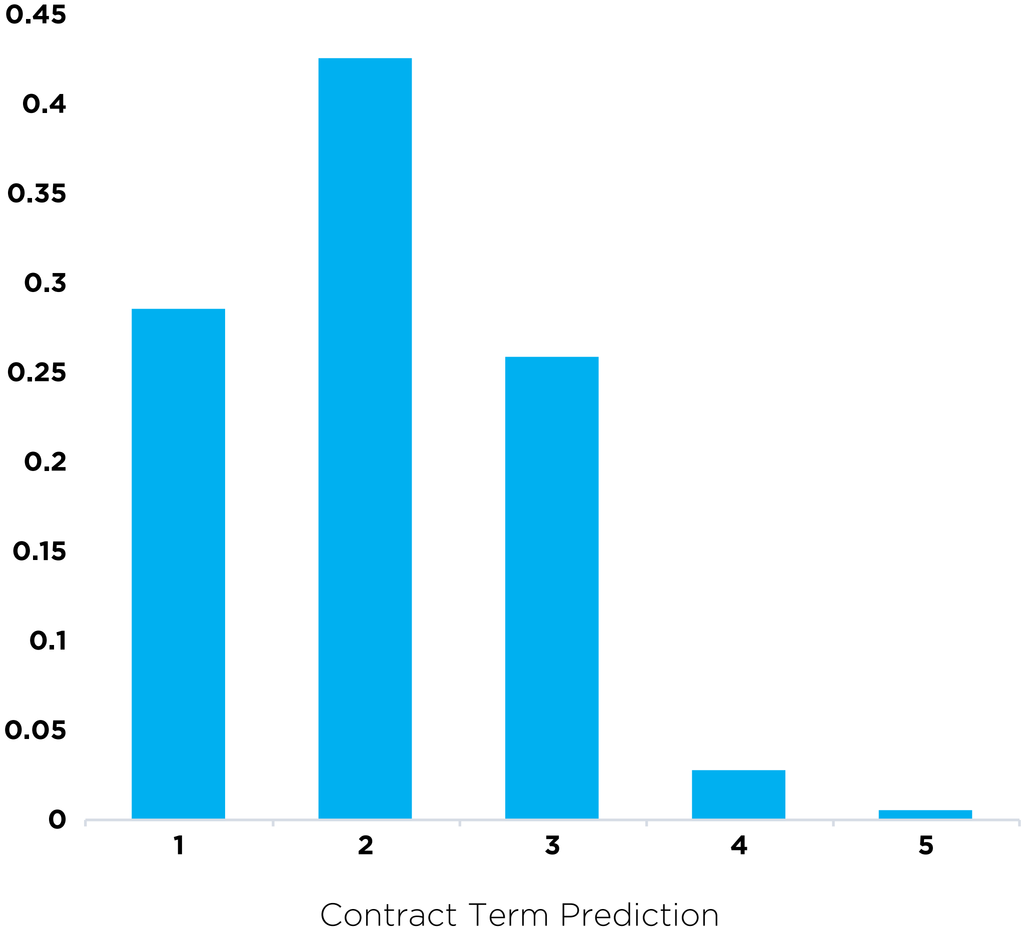
Luke Kornet Prediction Probabilities

Actual Term: 2

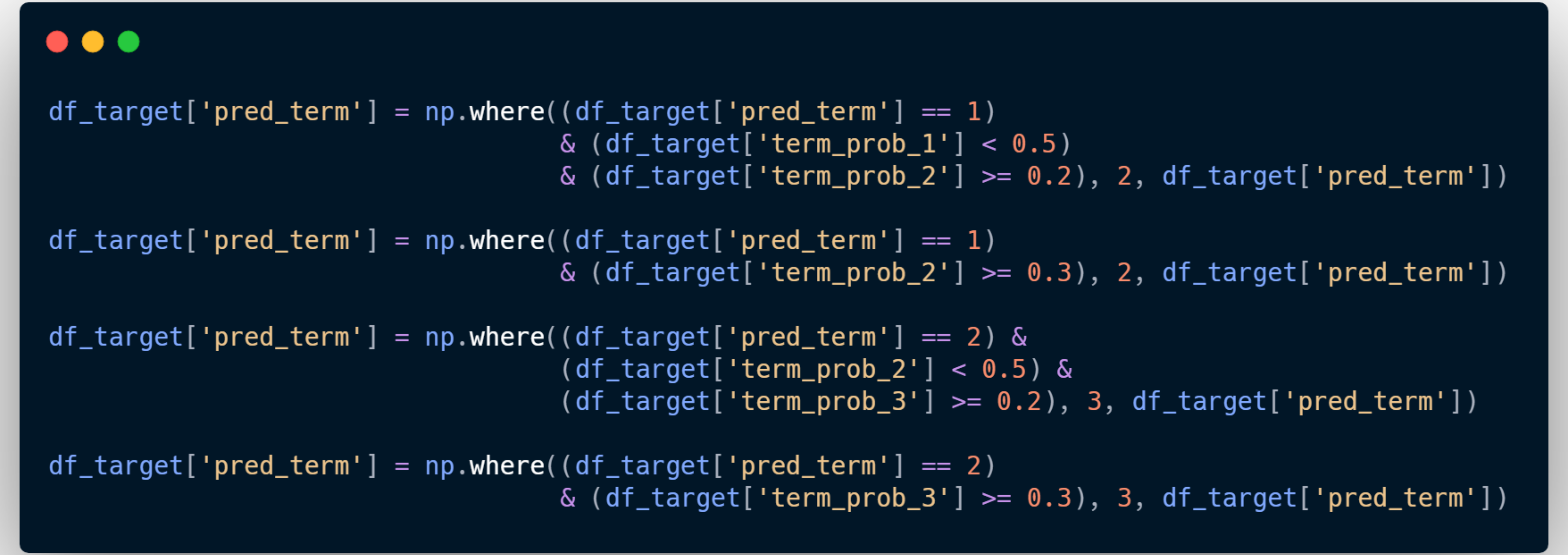


Tomas Satoransky Prediction Probabilities

Actual Term: 3



In order to improve the out of sample prediction, I dug into the class probability predictions and created an overlay



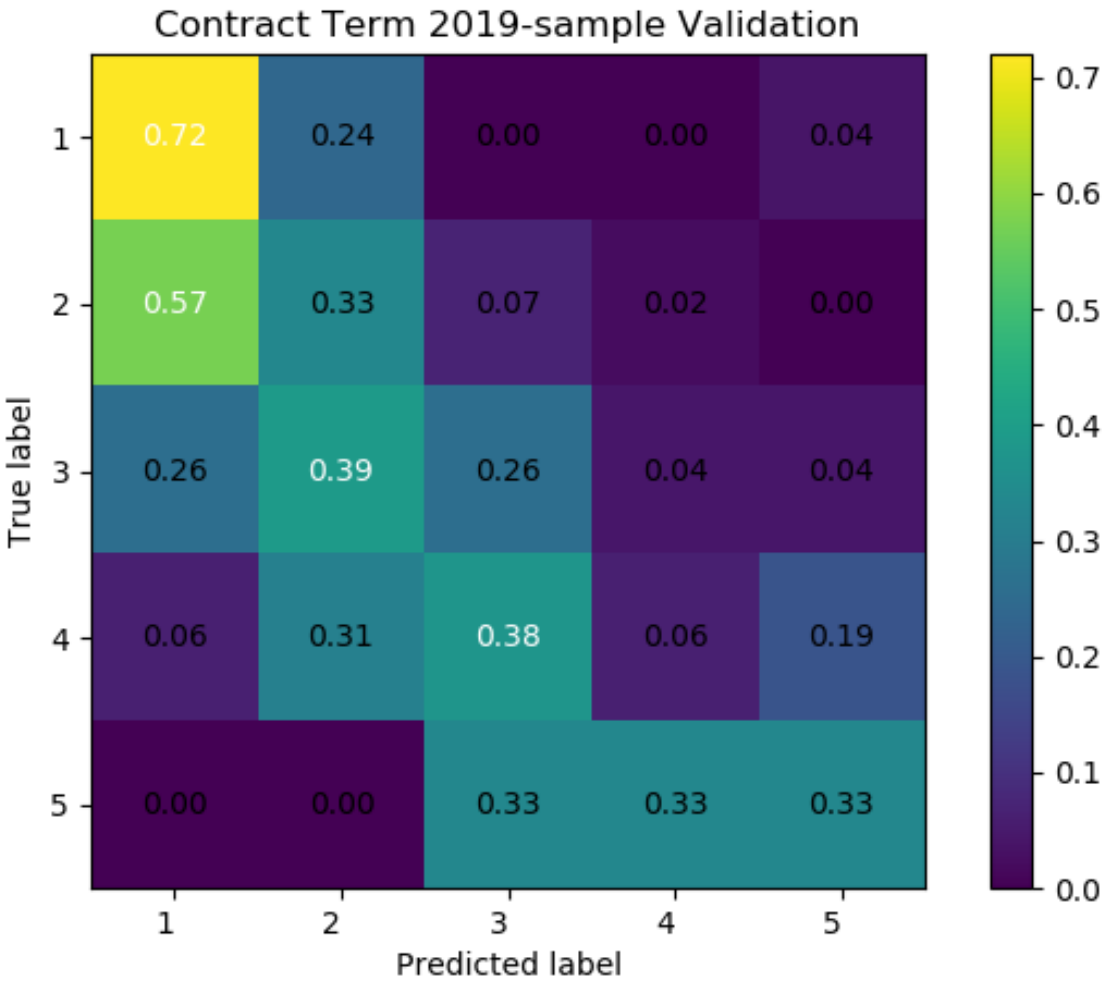
```
df_target['pred_term'] = np.where((df_target['pred_term'] == 1)
                                  & (df_target['term_prob_1'] < 0.5)
                                  & (df_target['term_prob_2'] >= 0.2), 2, df_target['pred_term'])

df_target['pred_term'] = np.where((df_target['pred_term'] == 1)
                                  & (df_target['term_prob_2'] >= 0.3), 2, df_target['pred_term'])

df_target['pred_term'] = np.where((df_target['pred_term'] == 2) &
                                  (df_target['term_prob_2'] < 0.5) &
                                  (df_target['term_prob_3'] >= 0.2), 3, df_target['pred_term'])

df_target['pred_term'] = np.where((df_target['pred_term'] == 2)
                                  & (df_target['term_prob_3'] >= 0.3), 3, df_target['pred_term'])
```

After the overlay, predictions are robust but still skew towards shorter contracts… however, this might be okay

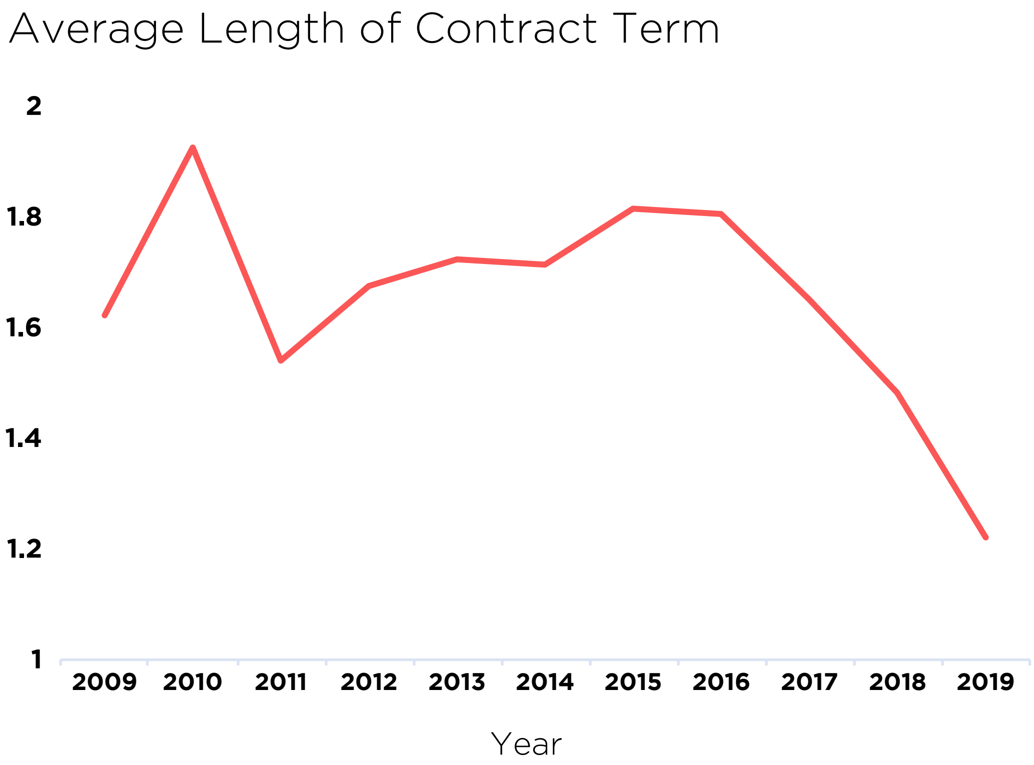

Contract Term 2019-sample Validation

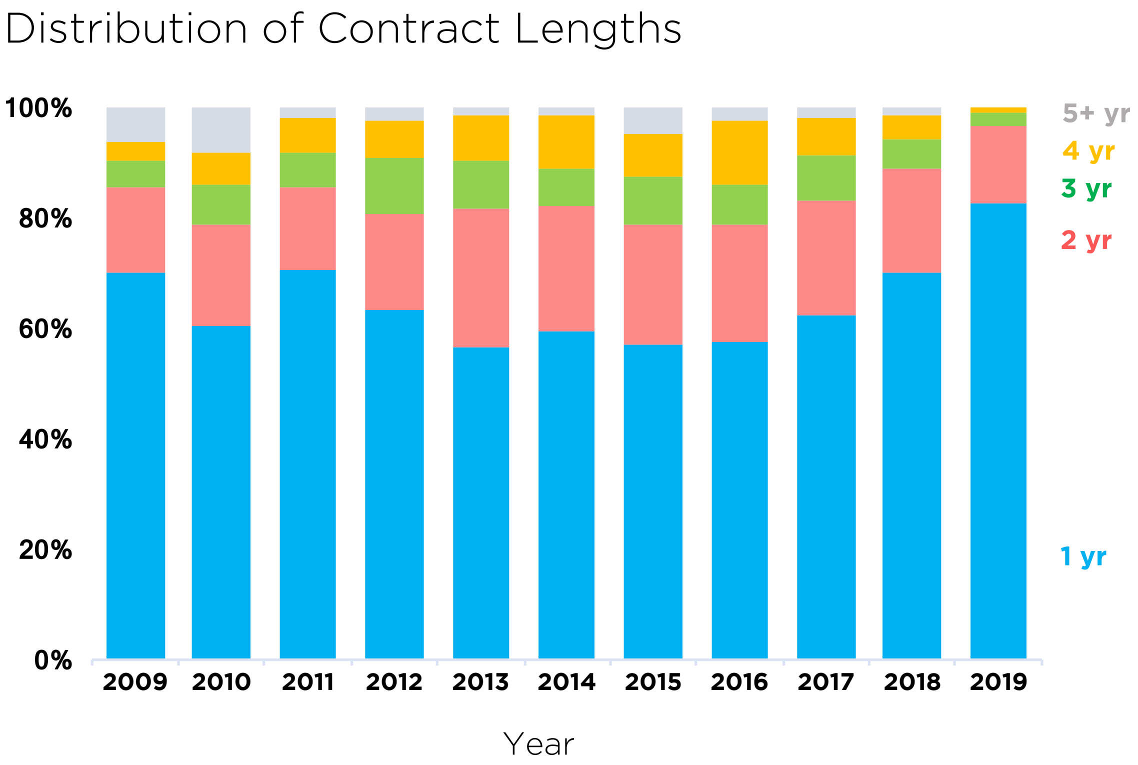
True label \ Predicted label	1	2	3	4	5
1	0.72	0.24	0.00	0.00	0.04
2	0.57	0.33	0.07	0.02	0.00
3	0.26	0.39	0.26	0.04	0.04
4	0.06	0.31	0.38	0.06	0.19
5	0.00	0.00	0.33	0.33	0.33

MAE:

0.8 years

After the overlay, predictions are robust but still skew towards shorter contracts… however, this might be okay





Contract Average Annual Value Model

Input dataset for Contract AAV model

Biographical Variables:

+ Height
+ Weight
+ Draft Position
+ Age
+ Contract Type (UFA, RFA, Extension)
+ Player Position

+ Player Position will actually be the grouping variable for our Mixed Effects Model

+ Contract Term prediction from the Term Model is an input into the AAV model

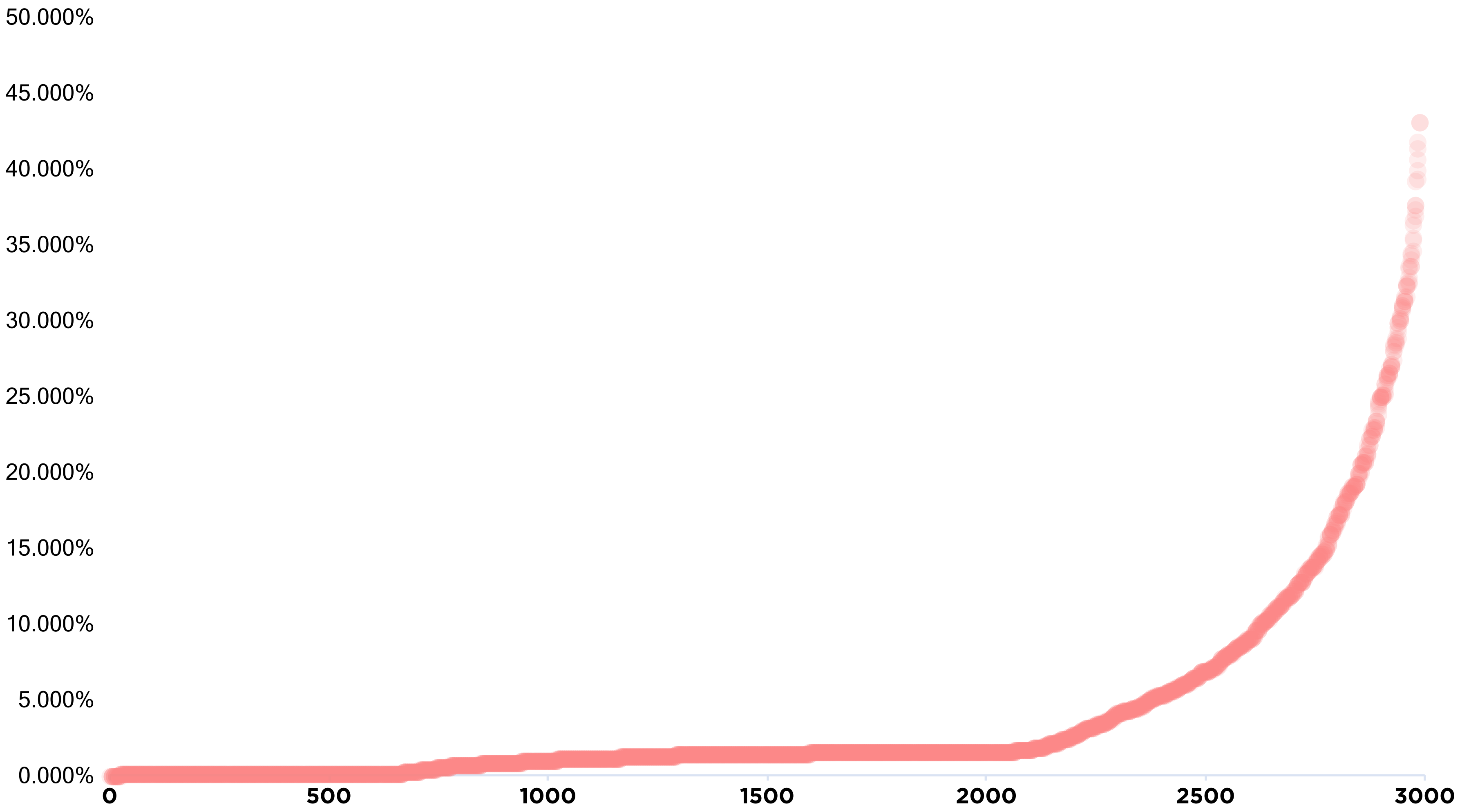
+ Contract AAV will be modeled as cap pct.

Player Statistics:

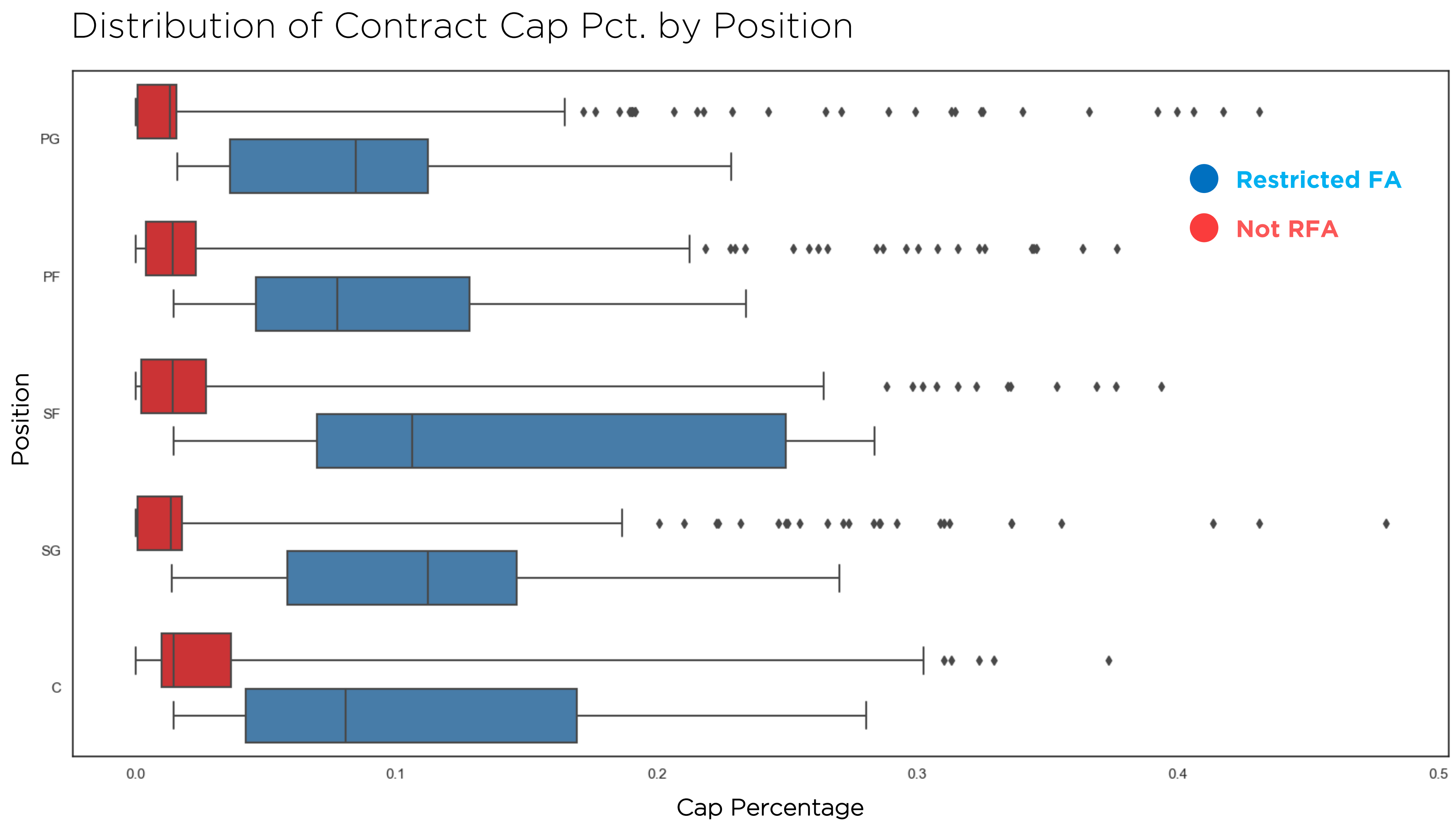
+ True Shooting %
+ Free Throw Attempt Rate
+ Three Point Attempt Rate
+ Offensive Rebound Pct.
+ Defensive Rebound Pct.
+ Assist Rate
+ Steal Rate
+ Block Rate
+ Usage Rate
+ Turnover Rate
+ Defensive Box Plus-Minus
+ Points per Game
+ Minutes per Game
+ Contract Term

The overall distribution of contract AAV is non-linear

Distribution of Ordered Average Annual Contract Cap Pct.

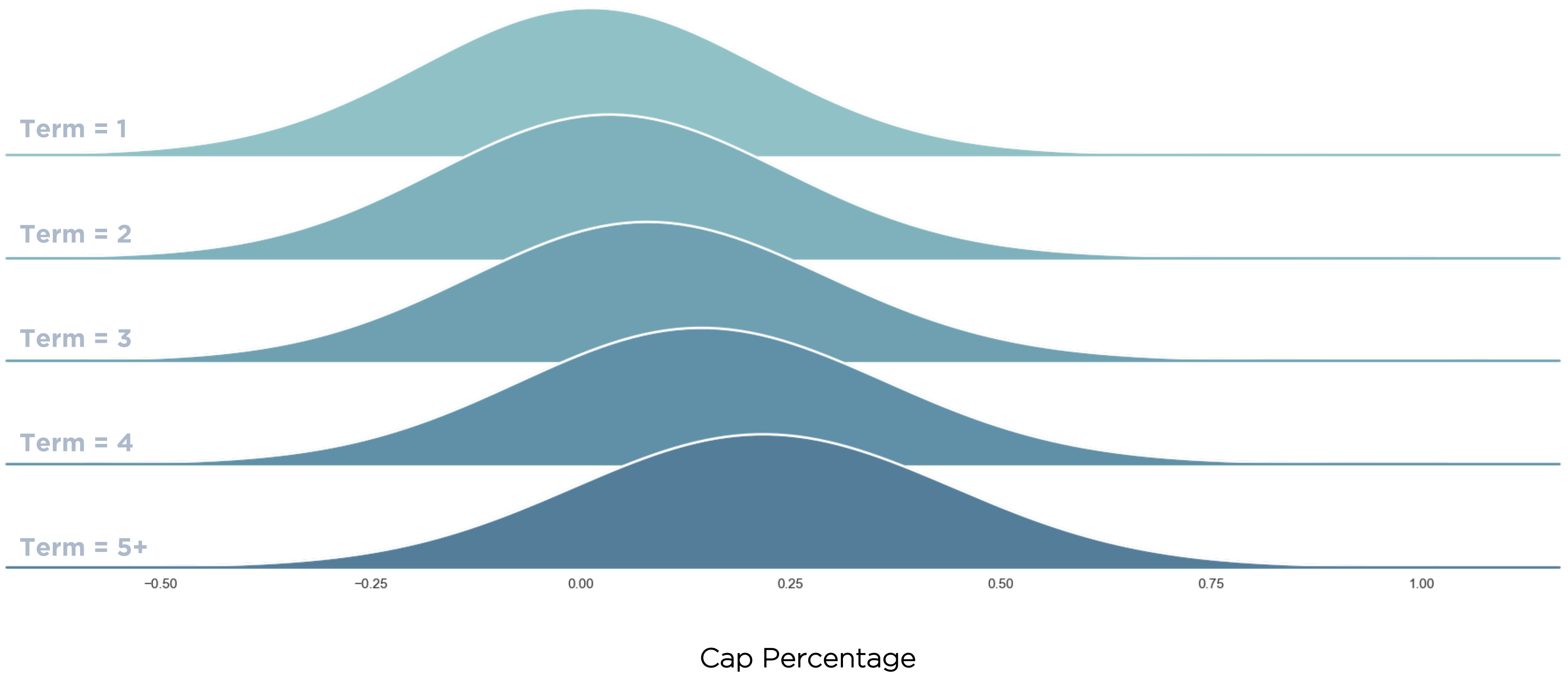


The position of NBA players significantly impacts the distribution of contract AAV

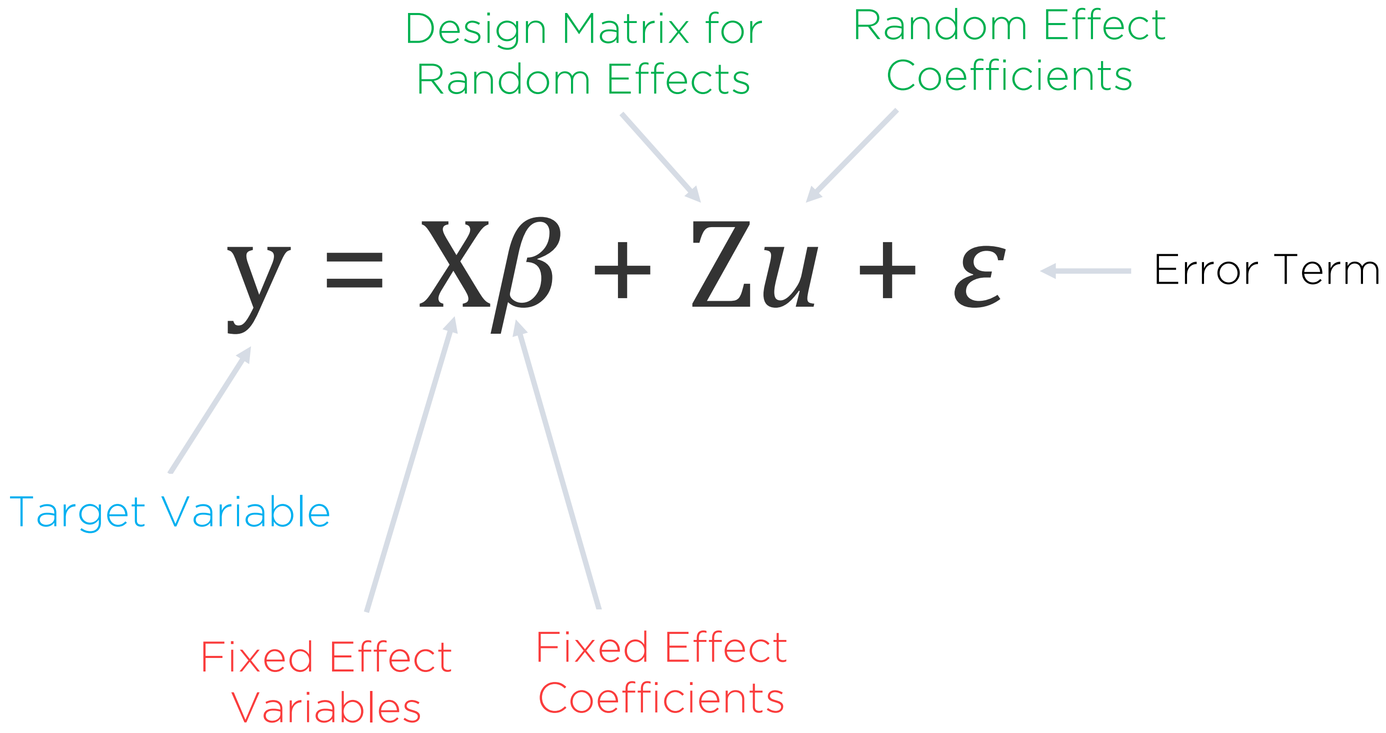


Contract term is a significant predictor of contract AAV, with greater terms positively correlated against contract value

Distribution of Contract Cap Pct. by Term Length



The grouping of positions naturally lends itself to a Mixed Effects model structure



$$y = X\beta + Zu + \varepsilon$$

Design Matrix for Random Effects

Random Effect Coefficients

Error Term

Target Variable

Fixed Effect Variables

Fixed Effect Coefficients

The grouping of positions naturally lends itself to a Mixed Effects model structure

Player Position
(PG, SG, SF, PF, C)

$$y = X\beta + Zu + \varepsilon$$

Cap Pct.

PPG, TS%, USG%,
Defensive BPM, etc.

Most popular implementations of Mixed Effects models are linear, but data scientists at Manifold developed an implementation of a Mixed Effects Random Forest

Linear Mixed Effects

$$y = X\beta + Zu + \varepsilon$$

Implementation: Python Statsmodels MixedLM

In-Sample RMSE: 0.032 ($3.5M under 2019-20 salary cap)

Non-Linear Mixed Effects

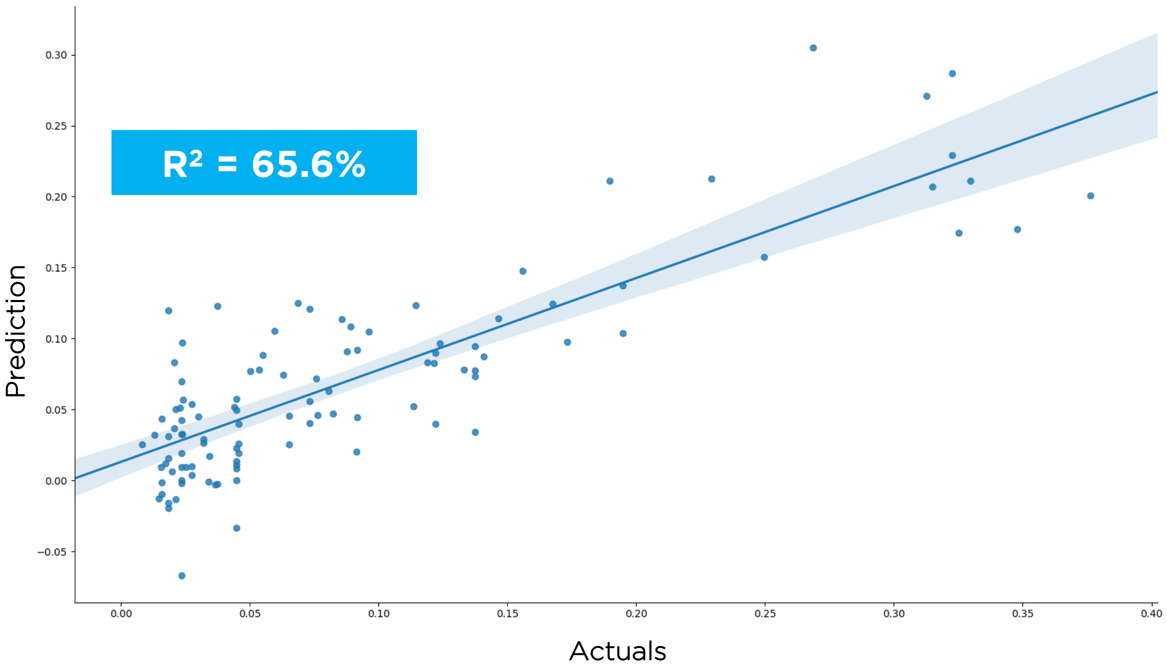
$$y = f(X) + Zu + \varepsilon$$

Implementation: Python MERF

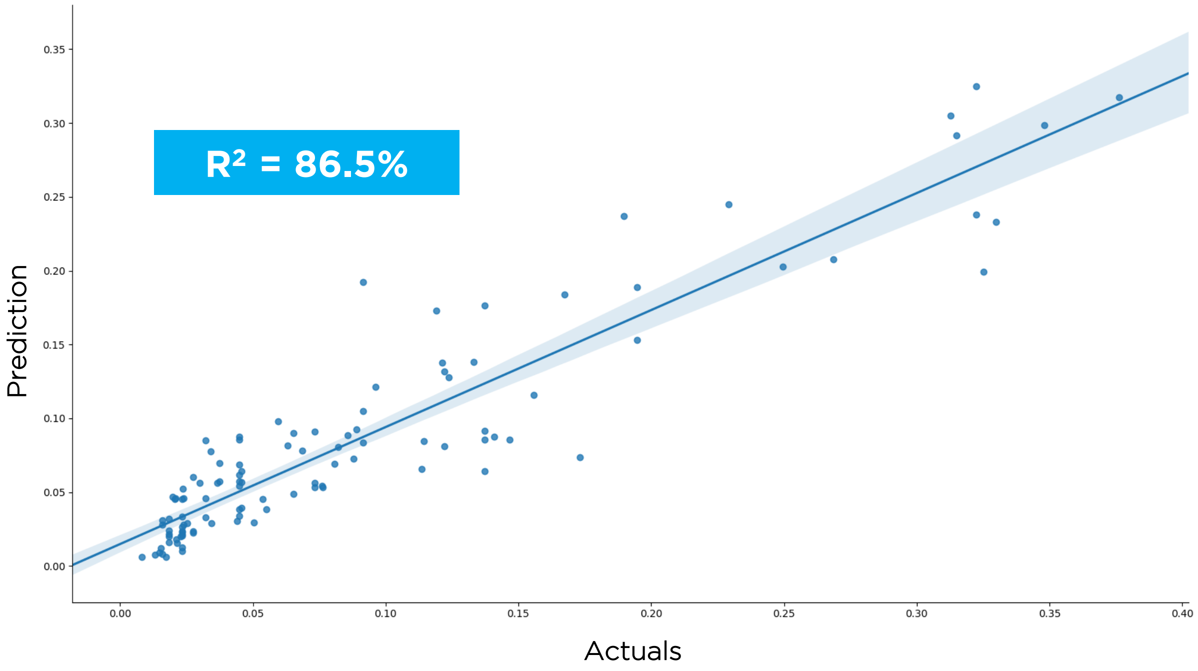
In-Sample RMSE: 0.009 ($0.98M under 2019-20 salary cap)

The MERF model performed demonstrably better in OOS testing on 2019 free agents

Actual Cap Pct. vs **MixedLM** Predictions



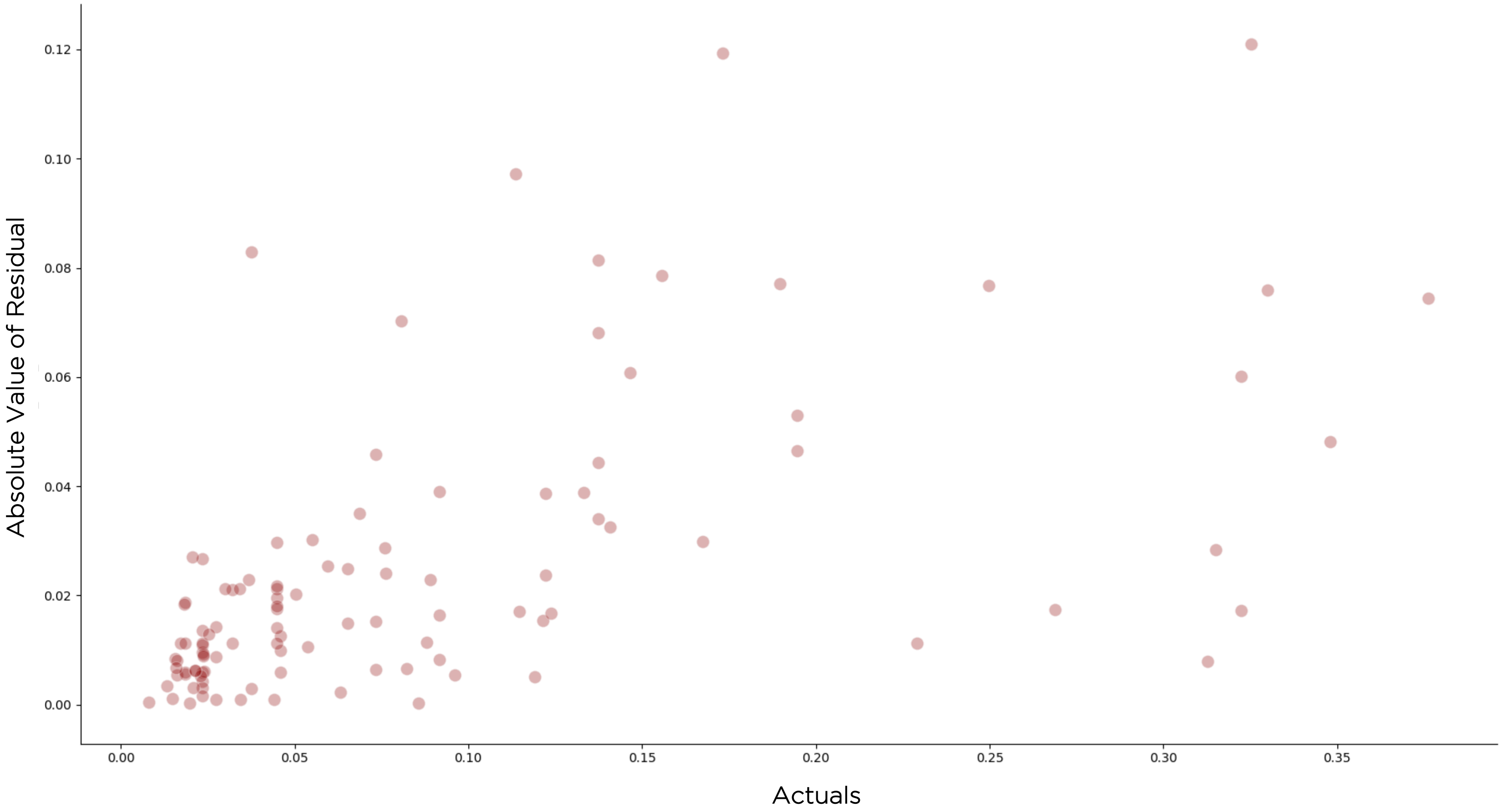
Actual Cap Pct. vs **MERF Version 1** Predictions



Note: DeMarcus Cousins was held out from the target dataset due to the highly anomalous circumstances around his injury situation.

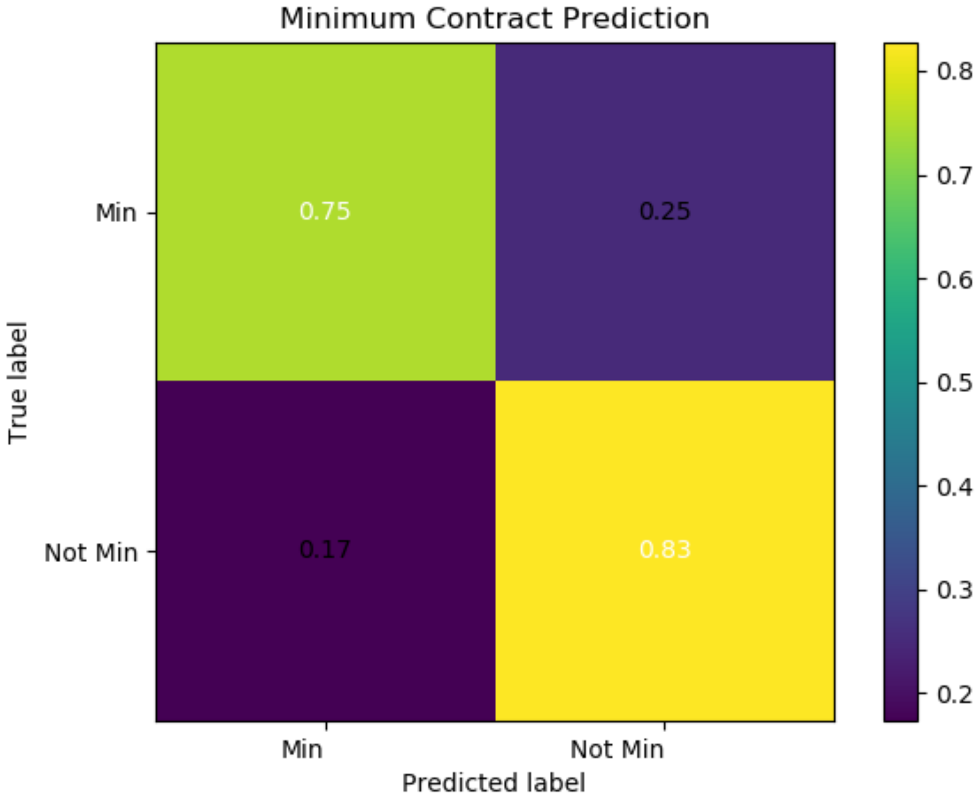
The MERF model held up well in out of sample testing, but we could still tweak it further along the tails

Actual Cap Pct. vs **MERF Version 1** Residuals

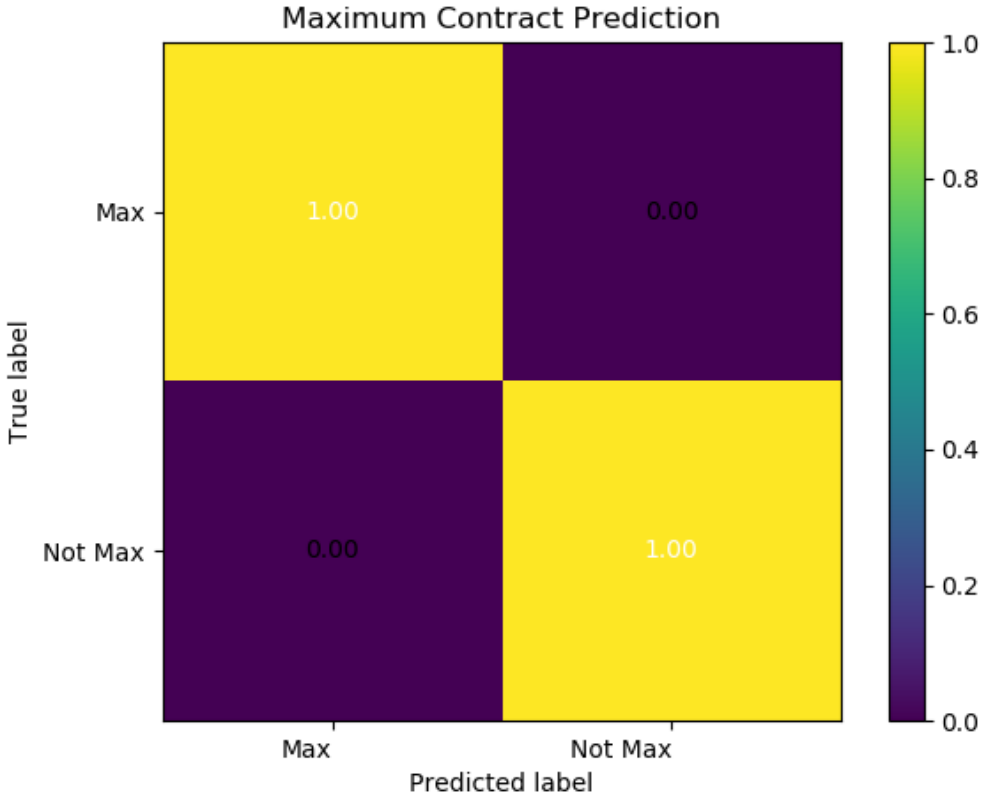


Note: DeMarcus Cousins was held out from the target dataset due to the highly anomalous circumstances around his injury situation.

A class-weighted random forest classifier is effective out of sample at predicting which players will receive minimum and maximum contracts

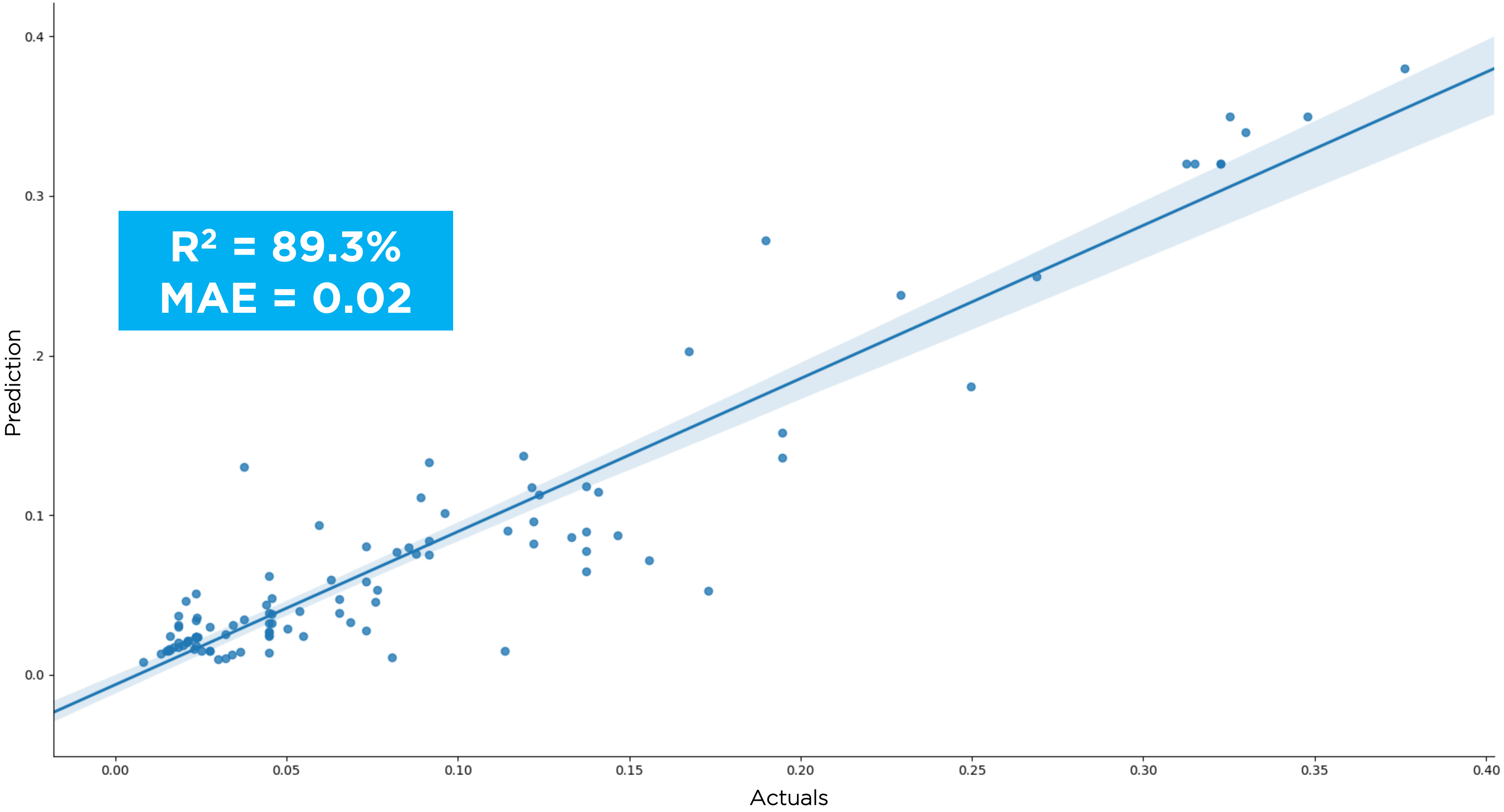


If the probability of Min >= 0.3, predict Min



If the probability of Max >= 0.5, predict Max

After applying the minimum and maximum contract overlays, the out of sample performance of the MERF model is further improved



R^2 = 89.3%
MAE = 0.02

Note: DeMarcus Cousins was held out from the target dataset due to the highly anomalous circumstances around his injury situation.

Lessons Learned: Biggest Misses

TERRY ROZIER (PG, 24 YEARS OLD)

Predicted Cap AAV: 5.5%
Actual Cap AAV: 17.3%
Residual: -11.8 bps

Key Stats, 2018-19 season:
9 ppg | 23 mpg | 0.5 TS% | 0.14 FTr

JULIUS RANDLE (PF, 24 YEARS OLD)

Predicted Cap AAV: 26.3%
Actual Cap AAV: 19%
Residual: +7.3 bps

Key Stats, 2018-19 season:
21.4 ppg | 30.6 mpg | 0.6 TS% | 0.45 FTr

+ Career years (the concept of a "contract year") is influential

+ Same as with the term model, utilization, efficiency, and scoring-related statistics play a big role

+ Model can't parse the context in "good stats, bad team" from "bad stats, good team"

+ Injuries can alter contract offers

+ A team's situation affects the opportunity cost of a contract offer (i.e. *who else would Team X pay, anyway?)*

+ More "bargain" contracts are signed deeper into free agency among the non max-level players, so timing also plays a part

But on a positive note, biggest hits!

Quinn Cook (PG, 26 years old)
Predicted Cap AAV: 2.6%
Actual Cap AAV: 2.7%
Residual: -0.1 bps

- Los Angeles Lakers

Nikola Vucevic (C, 28 years old)
Predicted Cap AAV: 23.4%
Actual Cap AAV: 22.9%
Residual: +0.5 bps

- Orlando Magic

JJ Redick (SG, 35 years old)
Predicted Cap AAV: 11.5%
Actual Cap AAV: 12.1%
Residual: -0.5 bps

- New Orleans Pelicans

SHAKLACKATY! We can go to break now.

@SENTH1S on Twitter

threesenths@gmail.com

Sometimes on Nylon Calculus, sometimes on Nightingale, sometimes on my personal blog, every time on Red Pants Friday though.