Pay to Play

Predicting NBA player contract values using player statistics and team information

The Problem

NBA Teams

How do I know how much to pay a player?

How much value is this player bringing to the team?

NBA Players

How do I know how much I should be paid?

Do I deserve more or less based on how I am playing?

Problem statement

How can NBA organizations and player agents predict player value? What features drive value?

Data

Per 36 Statistics

Stats adjusted for 36 min of game time

20 PTS, 36 mins... 20 PTS

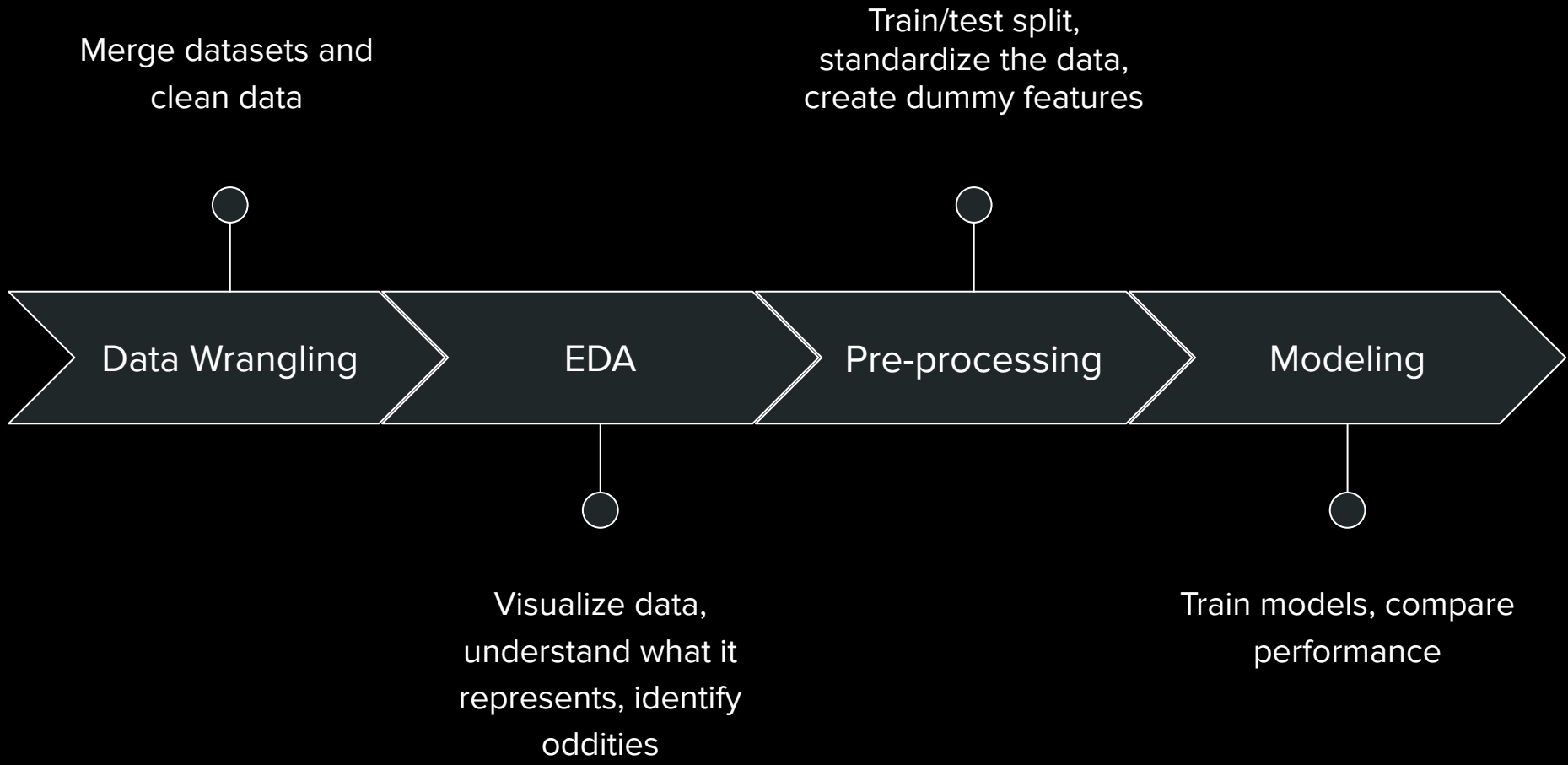
5 PTS, 9 mins... 20 PTS

Player Contracts

- Contract types
- Contract values

NBA Team Value

- Team value in billions of dollars
- Market size based on value


Merge datasets and clean data
Train/test split, standardize the data, create dummy features
Data Wrangling
EDA
Pre-processing
Modeling
Visualize data, understand what it represents, identify oddities
Train models, compare performance

Data Wrangling

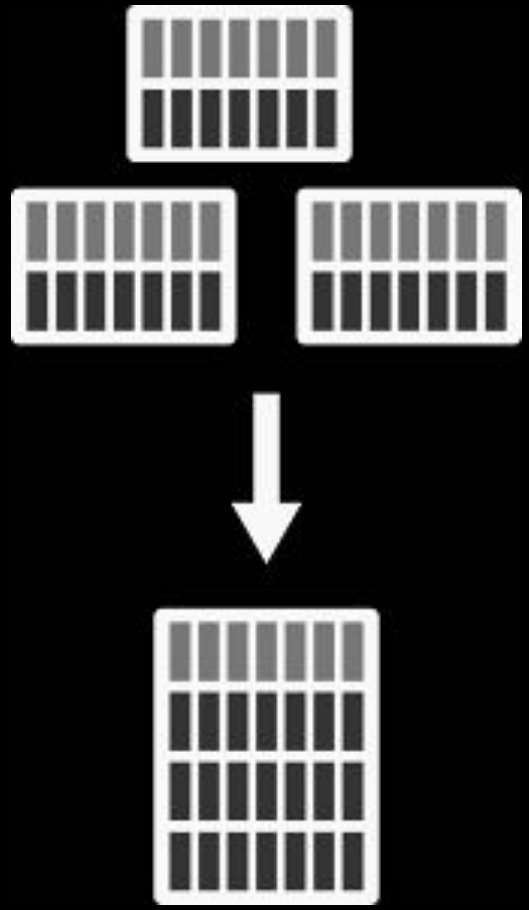
Merging data

Nulls

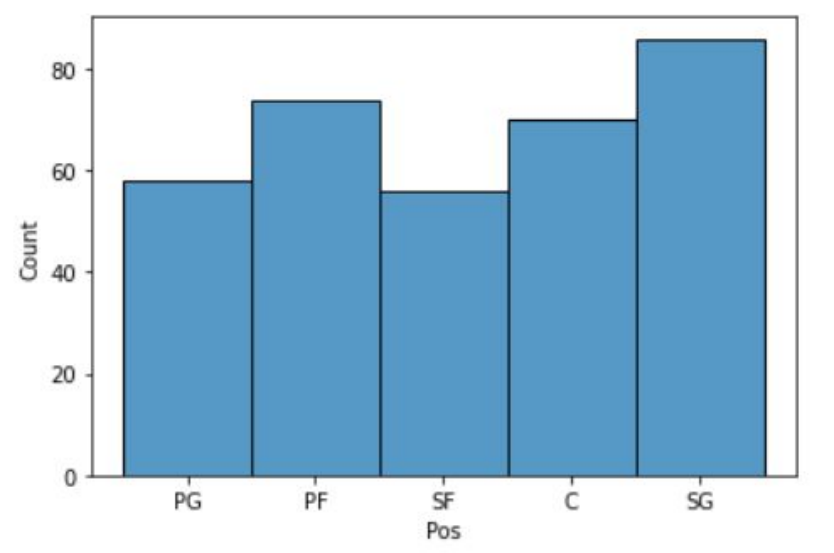
Repeated players

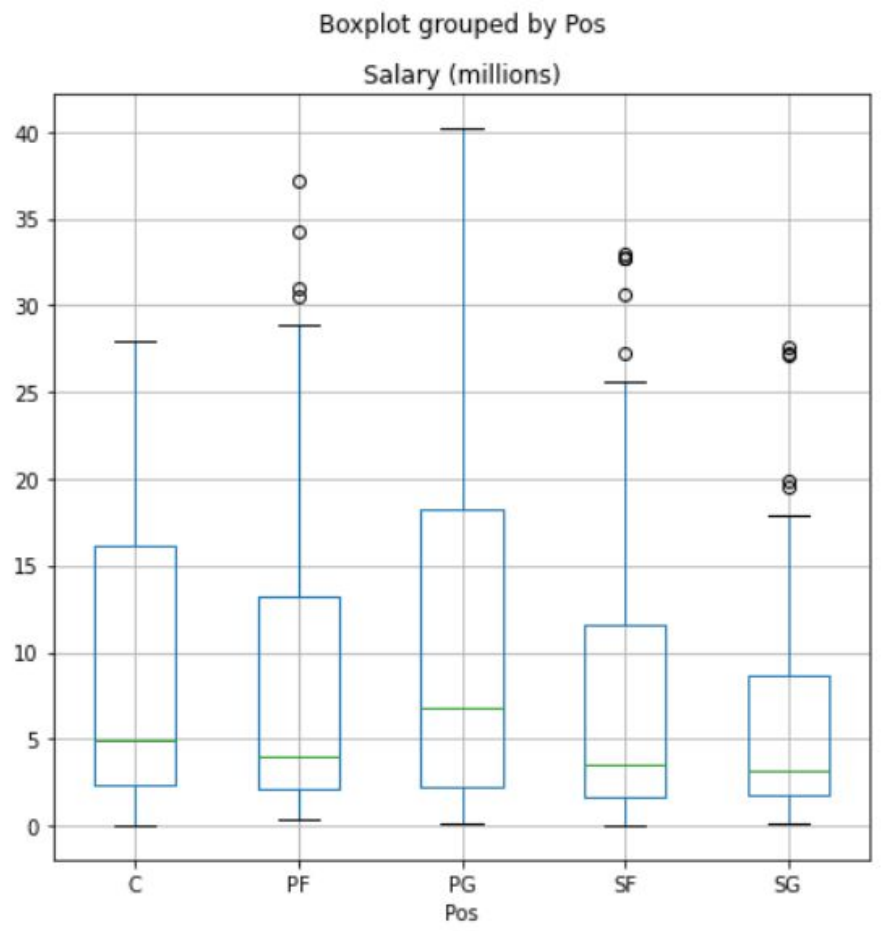
Inconsistent string values

Categorical data

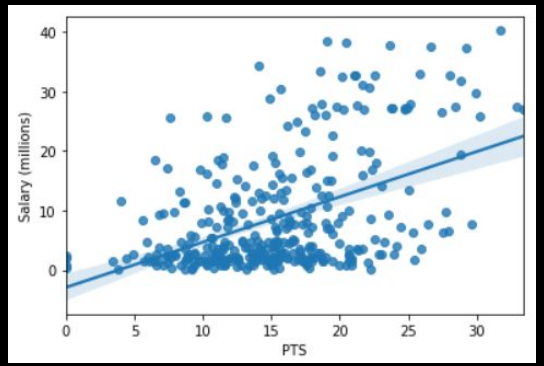


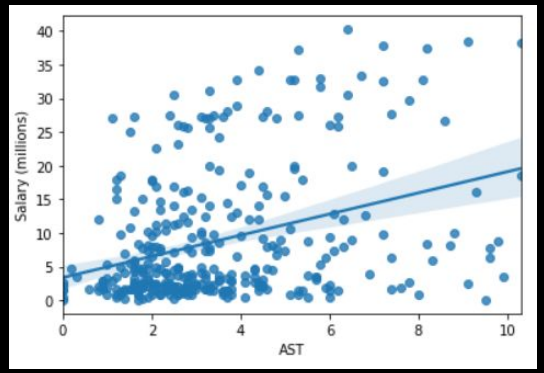
Exploratory Data Analysis

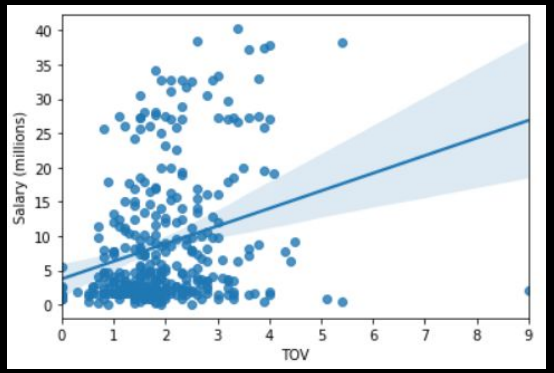




Exploratory Data Analysis





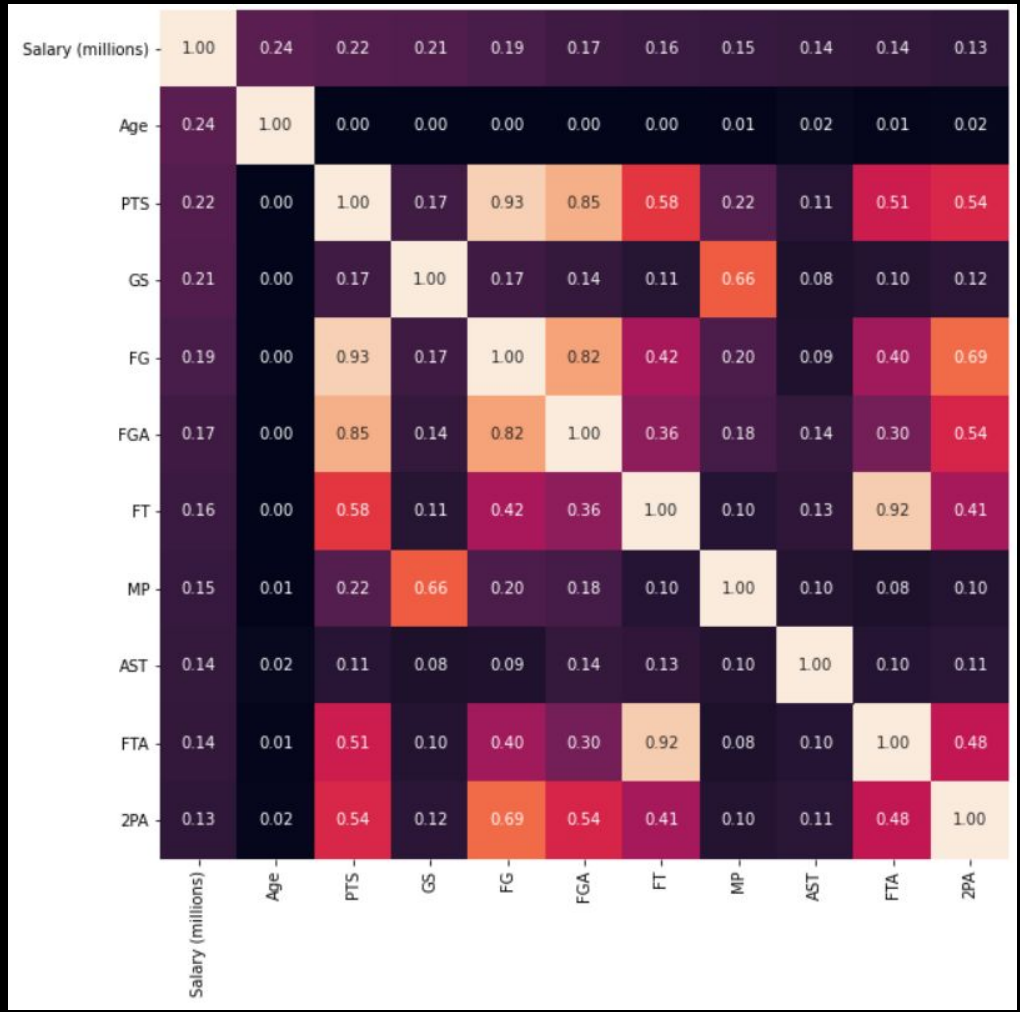


Exploratory Data Analysis

No standout features

Importance of Age

GS and MP



Pre-processing

Train/test split

Scale numerical data

Create dummy features

Combine data

	Age	G	GS	MP	FG	FGA	FG%
0	-1.470509	0.829981	1.345483	1.506218	2.123839	2.239854	0.113309
1	0.808674	0.002087	0.897741	0.336692	-0.058347	-0.308233	0.445095
2	1.061916	1.161138	-1.072323	-0.297877	-1.447012	-0.945255	-1.731420
3	0.048946	0.167666	0.987289	0.772036	1.875864	1.424466	0.883052
4	0.048946	-0.329070	-0.893226	-0.895552	-0.405513	0.252346	-1.227106

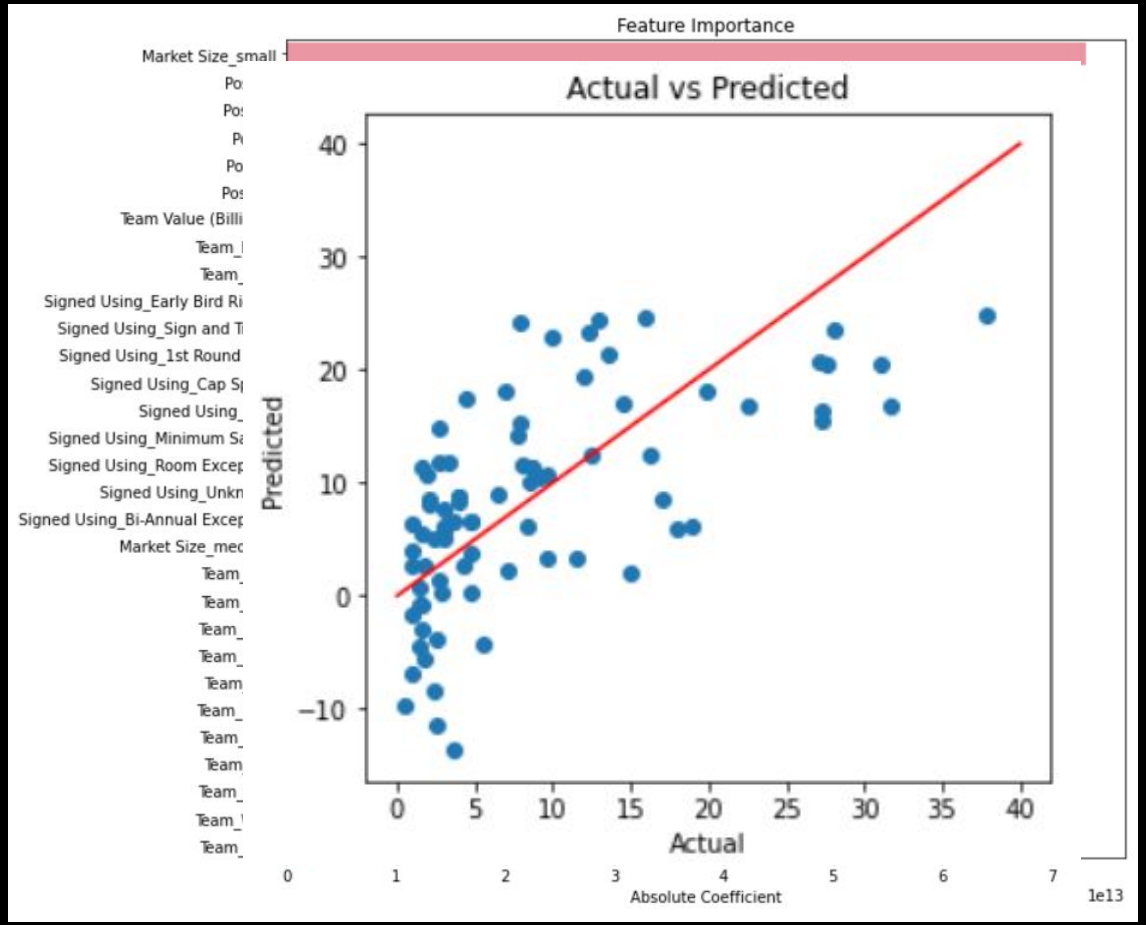
	Pos_C	Pos_PF	Pos_PG	Pos_SF	Pos_SG	Signed Using_1st Round Pick	Signed Using_Bi-Annual Exception	Signed Using_Cap Space	Signed Using_Early Bird Rights
0	0	0	1	0	0	1	0	0	0

Modeling - Linear Regression

$R^2 = 0.23$

RMSE = $7.70 million

MAE = $6.41 million



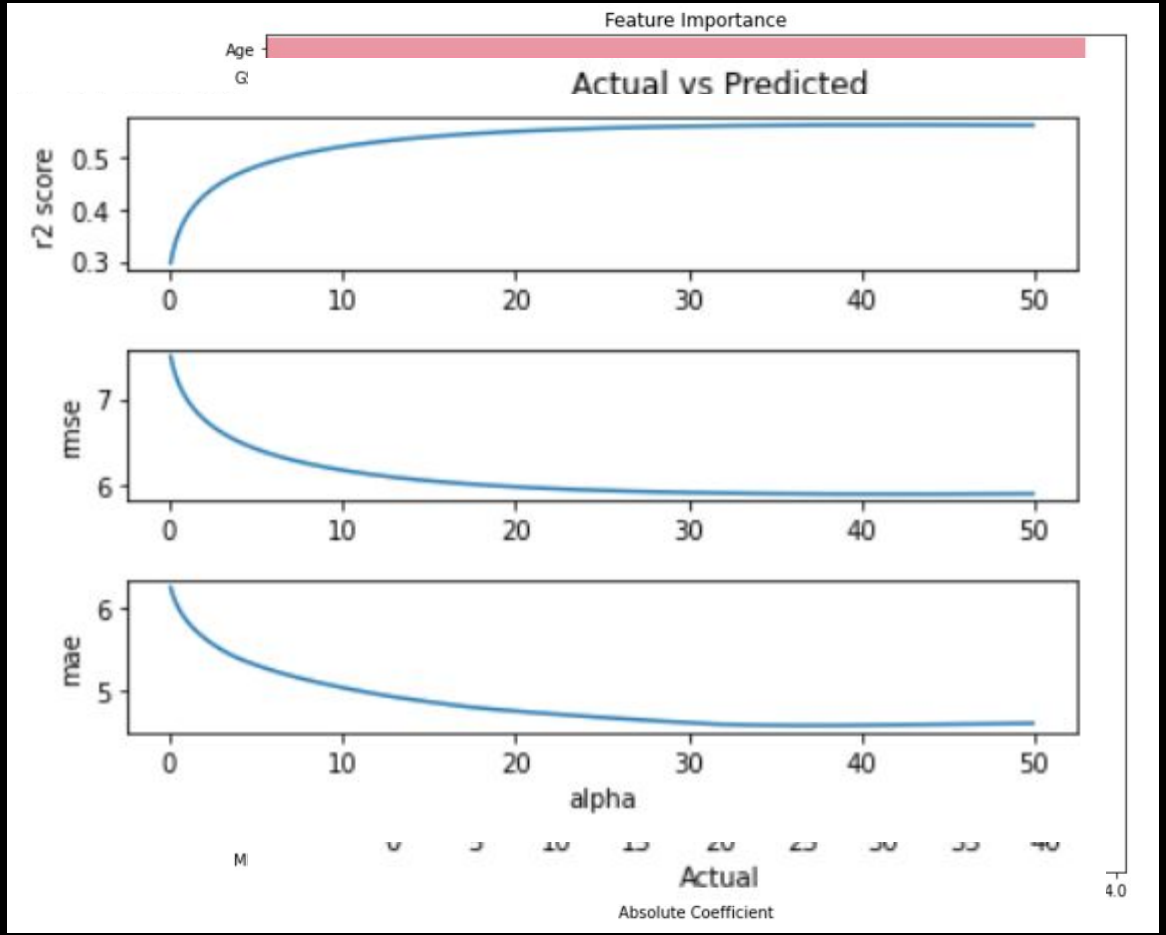
Modeling - Ridge

Alpha = 42.1

R^2 = 0.52

RMSE = $6.06 million

MAE = $4.72 million



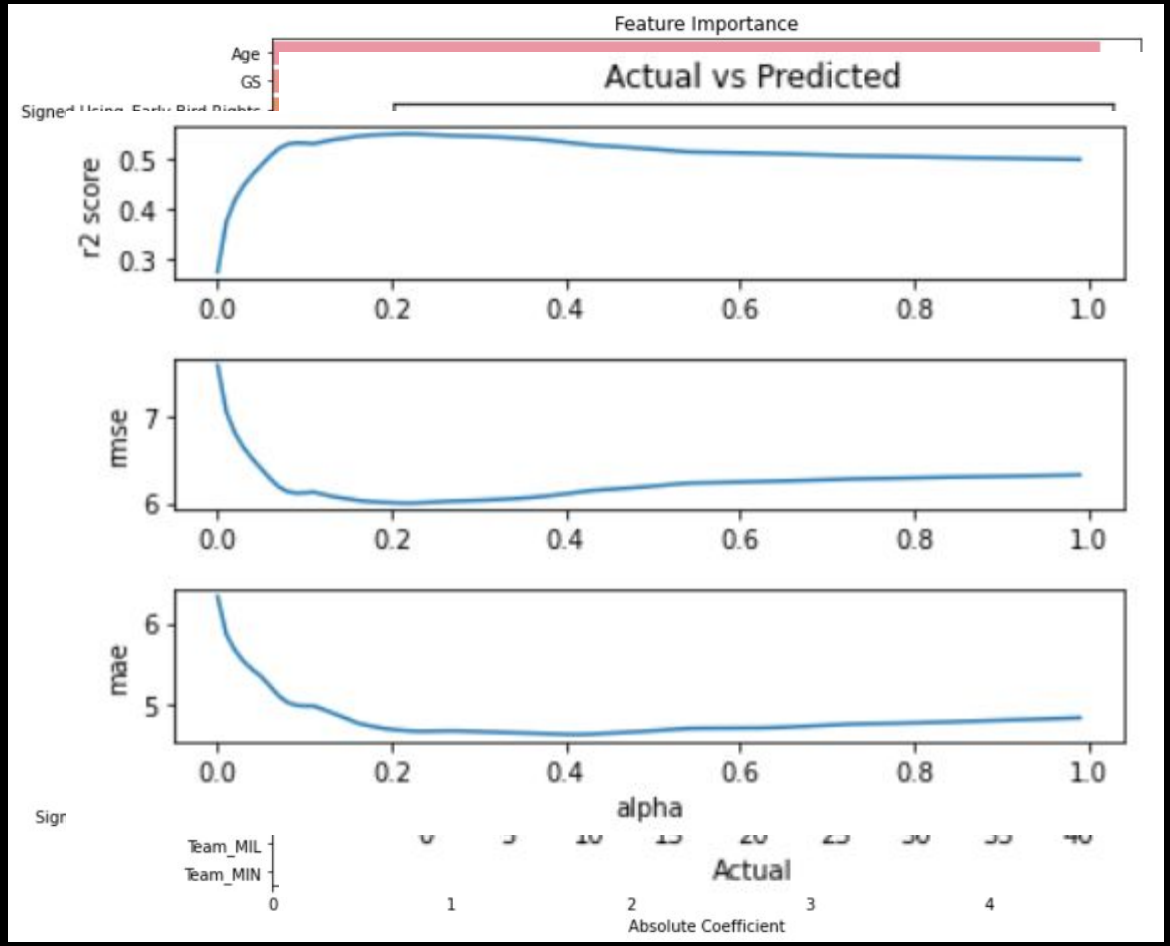
Modeling - Lasso

Alpha = 0.22

R^2 = 0.48

RMSE = $6.31 million

MAE = $5.12 million



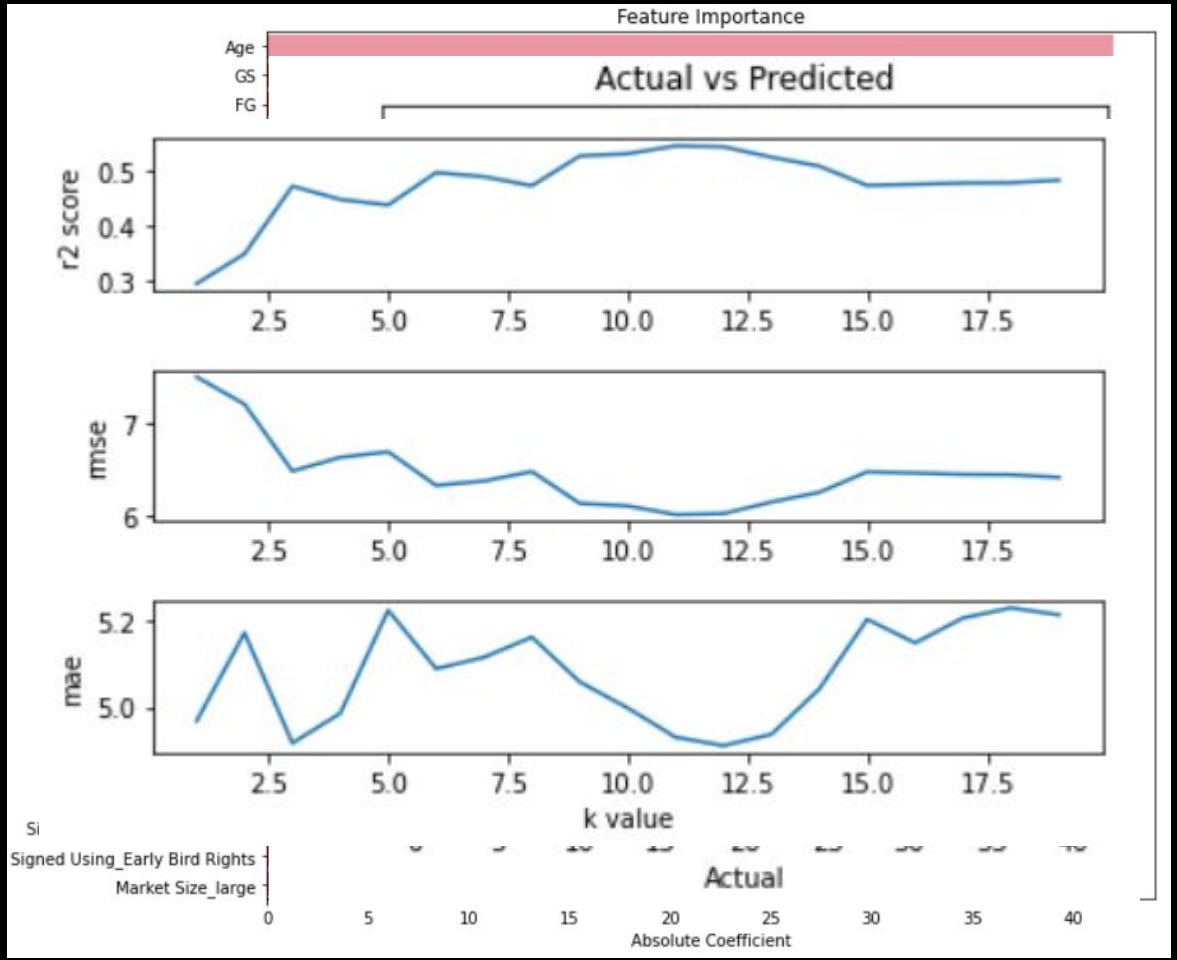
Modeling - KNN

k_neighbors = 11

R^2 = 0.41

RMSE = $6.71 million

MAE = $4.94 million



Modeling - Random Forest

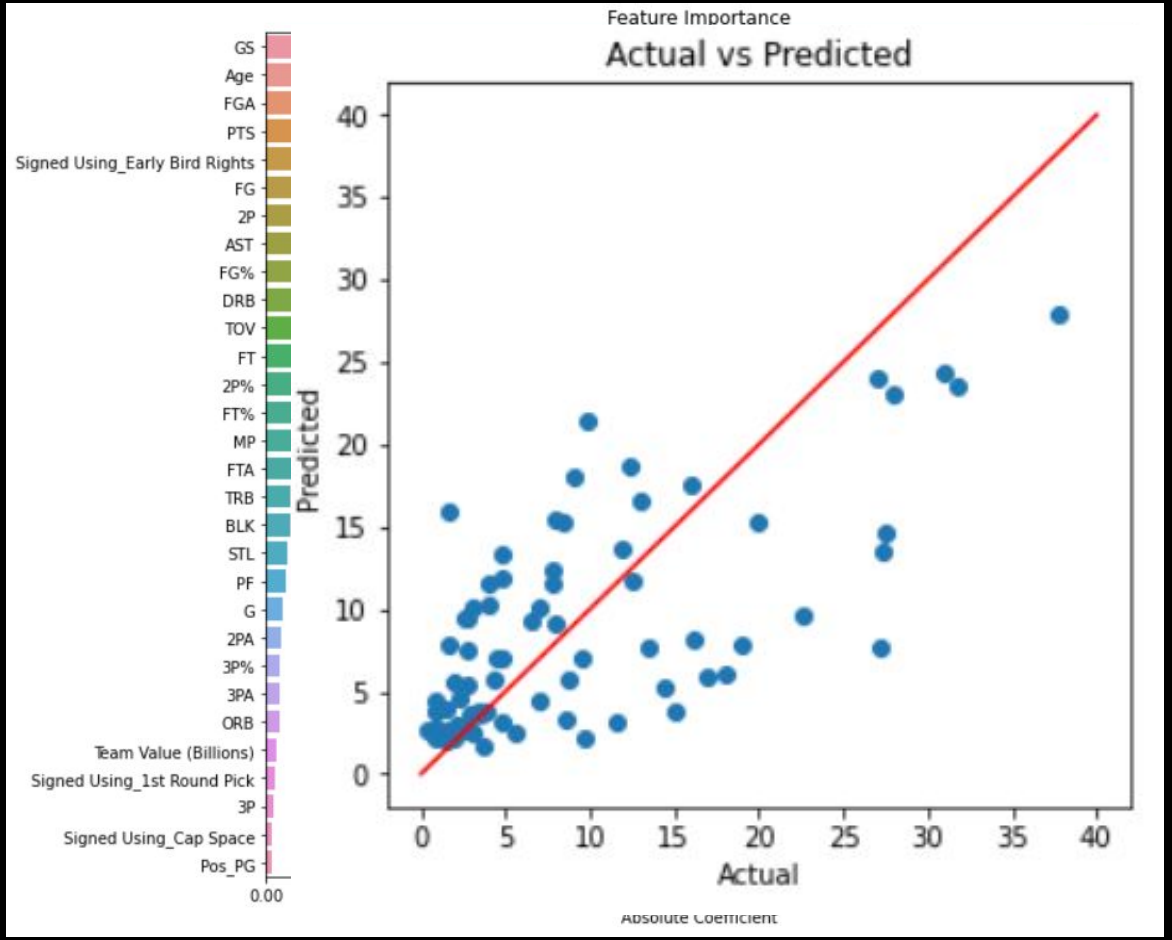
Parameter optimization

3 folds, 100 candidates, totalling 300 fits

$R^2 = 0.48$

RMSE = $6.34 million

MAE = $4.76 million



Model Performance

	Linear Regression	Ridge	Lasso	KNN	Random Forest
r^2	0.23	0.52	0.48	0.41	0.48
RMSE	7.70	6.06	6.31	6.71	6.34
MAE	6.41	4.72	5.12	4.94	4.76
Negative Values	Yes	Yes	Yes	No	No

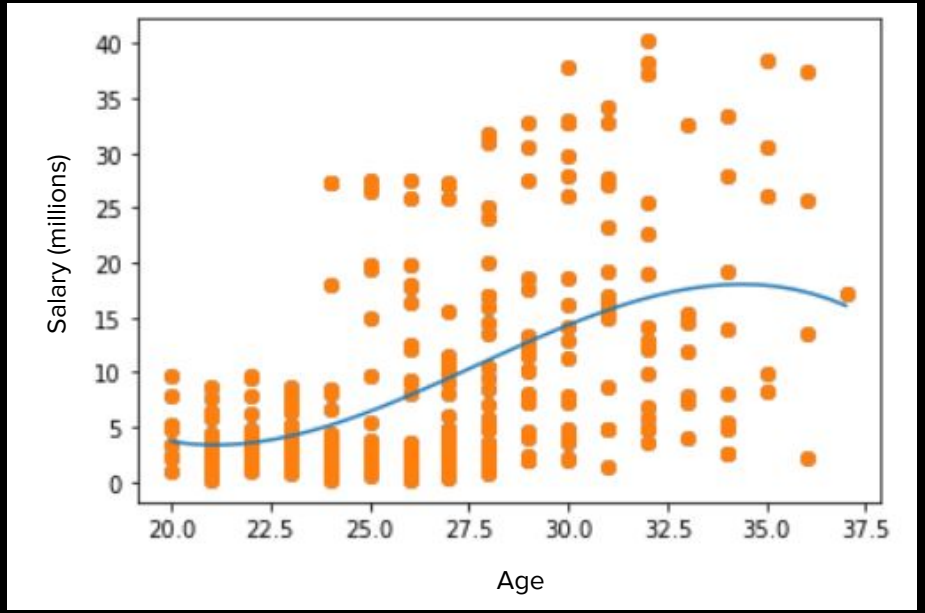
Takeaways and Future Improvements

Importance of Age

Not all stats are created equal

More data

Raw data



Questions?