What factors most influence NBA players' salaries?

Group 4:
Emma Cox
Allan Dontsis
Jason Reiver
Aston Welch
PJ Han

Backgrounds

Background

Some professional basketball players make millions of dollars a year, but what **performance factors** and statistics are the most impactful in determining a player's **future salary**?

Background

Basketball Reference, a website we found from the American Statistical Association, has a variety of statistics based on player performance.

These statistics are collected from the official NBA scorekeeping department.

We looked at per game season statistics for the 2017-2018 season to see how they determined player salaries for the 2018-2019 season.

Player	Salary	Pos	Age	G	GS	MP	FGP	3PP	2PP	eFGP	FTP	TRB	AST	STL	BLK	TOV	PF	PPG
Aaron Gordon	21590909	PF	22	58	57	32.9	0.434	0.336	0.497	0.5	0.698	7.9	2.3	1	0.8	1.8	1.9	17.6
Abdel Nader	1378242	SF	24	48	1	10.9	0.336	0.354	0.321	0.413	0.59	1.5	0.5	0.3	0.2	0.7	0.9	3
Al Horford	28928710	C	31	72	72	31.6	0.489	0.429	0.514	0.553	0.783	7.4	4.7	0.6	1.1	1.8	1.9	12.9
Al Jefferson	4000000	C	33	36	1	13.4	0.534	0	0.541	0.534	0.833	4	0.8	0.4	0.6	0.6	1.8	7
Alec Burks	11536515	SG	26	64	1	16.5	0.411	0.331	0.452	0.467	0.863	3	1	0.6	0.1	0.9	1.2	7.7
Alex Abrines	5455236	SG	24	75	8	15.1	0.395	0.38	0.443	0.54	0.848	1.5	0.4	0.5	0.1	0.3	1.7	4.7
Alex Len	4350000	C	24	69	13	20.2	0.566	0.333	0.568	0.567	0.684	7.5	1.2	0.4	0.9	1.1	2.3	8.5
Al-Farouq Aminu	6957105	PF	27	69	67	30	0.395	0.369	0.432	0.503	0.738	7.6	1.2	1.1	0.6	1.1	2	9.3
Alfonzo McKinnie	1349383	SF	25	14	0	3.8	0.533	0.333	0.833	0.633	0.667	0.5	0.1	0.1	0.1	0.2	0.6	1.5
Allen Crabbe	18500000	SG	25	75	68	29.3	0.407	0.378	0.461	0.529	0.852	4.3	1.6	0.6	0.5	1	2.2	13.2
Amir Johnson	2393887	C	30	74	18	15.8	0.538	0.313	0.57	0.558	0.612	4.5	1.6	0.6	0.6	0.7	2.6	4.6
Andre Drummond	25434263	C	24	78	78	33.7	0.529	0	0.536	0.529	0.605	16	3	1.5	1.6	2.6	3.2	15
Andre Iguodala	16000000	SF	34	64	7	25.3	0.463	0.282	0.567	0.514	0.632	3.8	3.3	0.8	0.6	1	1.5	6
Andre Roberson	10000000	SG	26	39	39	26.6	0.537	0.222	0.627	0.562	0.316	4.7	1.2	1.2	0.9	0.8	2.3	5
Andrew Wiggins	25467250	SF	22	82	82	36.3	0.438	0.331	0.475	0.481	0.643	4.4	2	1.1	0.6	1.7	2	17.7
Ante Zizic	1952760	C	21	32	2	6.7	0.731		0.731	0.731	0.724	1.9	0.2	0.1	0.4	0.3	0.9	3.7
Anthony Davis	25434263	PF	24	75	75	36.4	0.534	0.34	0.558	0.552	0.828	11.1	2.3	1.5	2.6	2.2	2.1	28.1
Anthony Tolliver	5750000	PF	32	79	14	22.2	0.464	0.436	0.561	0.632	0.797	3.1	1.1	0.4	0.3	0.7	1.8	8.9
Antonio Blakeney	1349383	SG	21	19	0	16.5	0.371	0.288	0.418	0.423	0.769	1.7	1.1	0.4	0.1	0.6	1	7.9
Aron Baynes	5193600	C	31	81	67	18.3	0.471	0.143	0.487	0.474	0.756	5.4	1.1	0.3	0.6	1	2.5	6
Austin Rivers	12650000	SG	25	61	59	33.7	0.424	0.378	0.461	0.508	0.642	2.4	4	1.2	0.3	1.8	2.5	15.1
Avery Bradley	12000000	SG	27	46	46	31.2	0.414	0.369	0.436	0.474	0.768	2.5	2	1.1	0.2	2.2	2	14.3
Bam Adebayo	2955840	C	20	69	19	19.8	0.512	0	0.523	0.512	0.721	5.5	1.5	0.5	0.6	1	2	6.9
Ben McLemore	5460000	SG	24	56	17	19.5	0.421	0.346	0.477	0.495	0.828	2.5	0.9	0.7	0.3	1.1	2.5	7.5
Ben Simmons	6434520	PG	21	81	81	33.7	0.545	0	0.551	0.545	0.56	8.1	8.2	1.7	0.9	3.4	2.6	15.8
Bismack Biyombo	17000000	C	25	82	25	18.2	0.52	0	0.521	0.52	0.65	5.7	0.8	0.3	1.2	1	1.9	5.7
Blake Griffin	32088932	PF	28	58	58	34	0.438	0.345	0.482	0.493	0.785	7.4	5.8	0.7	0.3	2.8	2.4	21.4
Boban Marjanovic	7000000	C	29	39	1	8.6	0.534		0.534	0.534	0.794	3.7	0.6	0.3	0.3	0.9	1.3	6
Bobby Portis	2494346	PF	22	73	4	22.5	0.471	0.359	0.514	0.52	0.769	6.8	1.7	0.7	0.3	1.4	1.8	13.2
Bogdan Bogdanovic	9000000	SG	25	78	53	27.9	0.446	0.392	0.485	0.529	0.84	2.9	3.3	0.9	0.2	1.6	2.2	11.8
Bojan Bogdanovic	10500000	SF	28	80	80	30.8	0.474	0.402	0.534	0.565	0.868	3.4	1.5	0.7	0.1	1.3	1.6	14.3
Bradley Beal	25434263	SG	24	82	82	36.3	0.46	0.375	0.507	0.527	0.791	4.4	4.5	1.2	0.4	2.6	2	22.6
Brandon Ingram	5757120	SF	20	59	59	33.5	0.47	0.39	0.483	0.497	0.681	5.3	3.9	0.8	0.7	2.5	2.8	16.1
Brook Lopez	3382000	C	29	74	72	23.4	0.465	0.345	0.549	0.536	0.703	4	1.7	0.4	1.3	1.3	2.6	13
Bryn Forbes	3125000	SG	24	80	12	19	0.421	0.39	0.446	0.51	0.667	1.4	1	0.4	0	0.5	1.4	6.9
Buddy Hield	3833760	SG	24	80	12	25.3	0.446	0.431	0.457	0.54	0.877	3.8	1.9	1.1	0.3	1.6	1.9	13.5
C.J. Miles	8333333	SF	30	70	3	19.1	0.379	0.361	0.434	0.516	0.835	2.2	0.8	0.5	0.3	0.6	1.9	10
Caleb Swanigan	1740000	PF	20	27	3	7	0.4	0.125	0.442	0.408	0.667	2	0.5	0.2	0.1	0.7	1.4	2.3
Cameron Payne	3263294	PG	23	25	14	23.3	0.405	0.385	0.419	0.486	0.75	2.8	4.5	1	0.4	1.4	1.8	8.8
Caris LeVert	1702800	SF	23	71	10	26.3	0.435	0.347	0.48	0.493	0.711	3.7	4.2	1.2	0.3	2.2	2.2	12.1
Carmelo Anthony	27928140	PF	33	78	78	32.1	0.404	0.357	0.437	0.476	0.767	5.8	1.3	0.6	0.6	1.3	2.5	16.2
Cedi Osman	2775000	SF	22	61	12	11	0.484	0.368	0.586	0.57	0.565	2	0.7	0.4	0	0.5	1.3	3.9

Data Description

Predictor Variables:

Pos = Position [PG, SG, SF, PF, C]
Age = Age
G = Total number of games played
GS = Games started
MP = Minutes per game
FGP = Field goal percentage
3PP = 3 point field goal percentage per game
2PP = 2 point field goal percentage per game
eFGP = Effective field goal percentage
FTP = Free throw field goal percentage per game
TRB = Total rebounds per game
AST = Assists per game
STL = Steals per game
BLK = Blocks per game
TOV = Turnovers per game
PF = Personal fouls
PPG = Points per game

Methods

Methods

Process for arriving at the final model:

1. Choose the variables, build & clean the dataset by finding **outliers**
2. Build a **multiple regression model** with all 17 predictor variables
3. Remove variables that are statistically insignificant (high **p-values** & low **t-score** absolute values) and ensure that practical significance of the model was not negatively affected (**adjusted R-squared** of model) after removing them
4. Check for **multicollinearity** and ensure that no predictor variables were highly correlated with each other

Build and Clean
The data set

Build and Clean the dataset

Because we could not find a dataset with both player stats and salaries, we had to combine data from two different sources. We then had to remove players that are “rookies”, because there was not yet statistical data on their performance

NBA Stats.xlsx

+

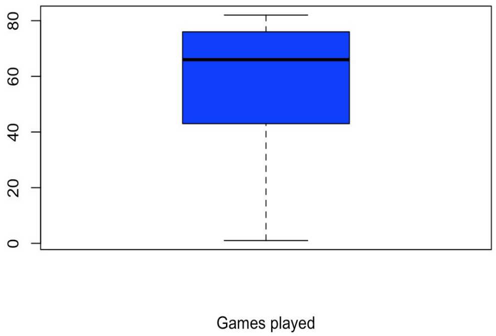
NBA Salaries

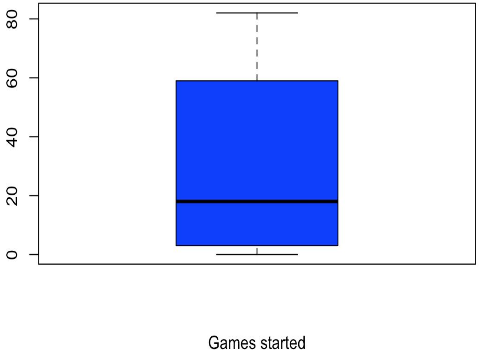
−

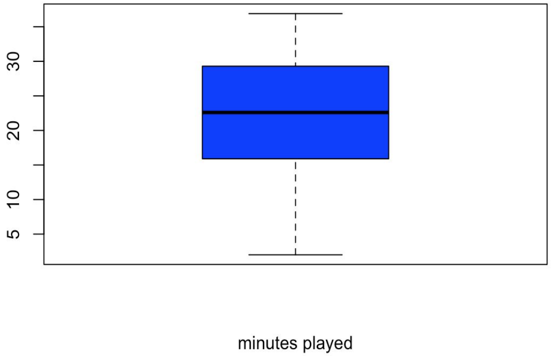
ROOKIE

Predictors Without Outliers

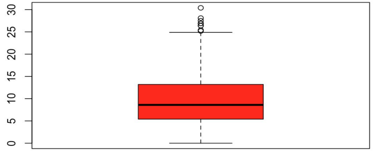
Due to a normal distribution in the data for that variable, outliers didn't exist in predictors, such as **"Games played"**, **"Games Started"**, and **"Minutes played"**



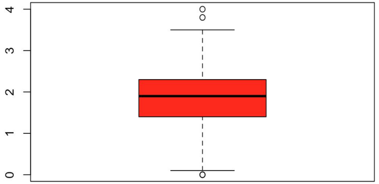




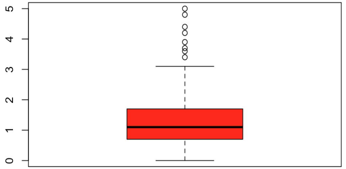
Predictors with outliers



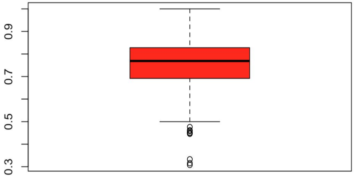
Points per game



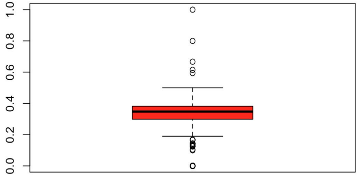
Personal Foul per game



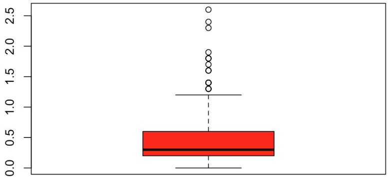
Total Turnover



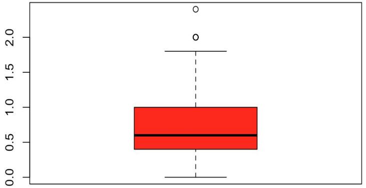
Free Throw %



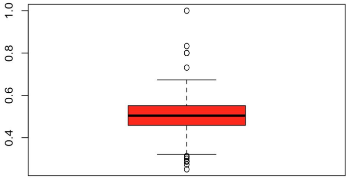
3 point field goal %



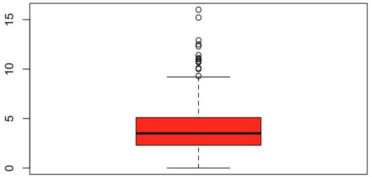
Total Block



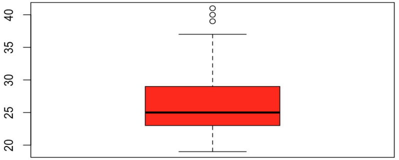
Total Steal



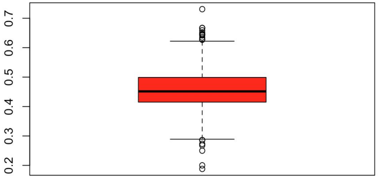
2 point field goal %



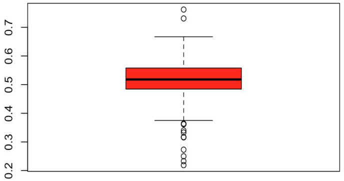
Total Rebound per game



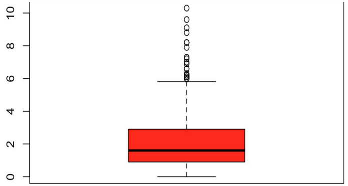
Age



Field goal %



Effective field goal %



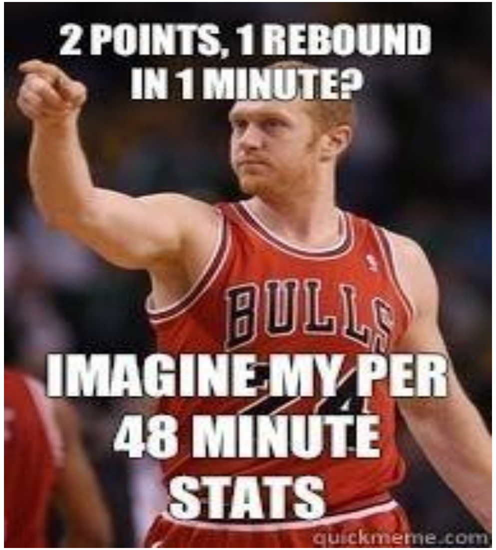
Total Assist

Identifying REAL Outliers

Exceptionally Better players?

VS

Wrong Averaging?



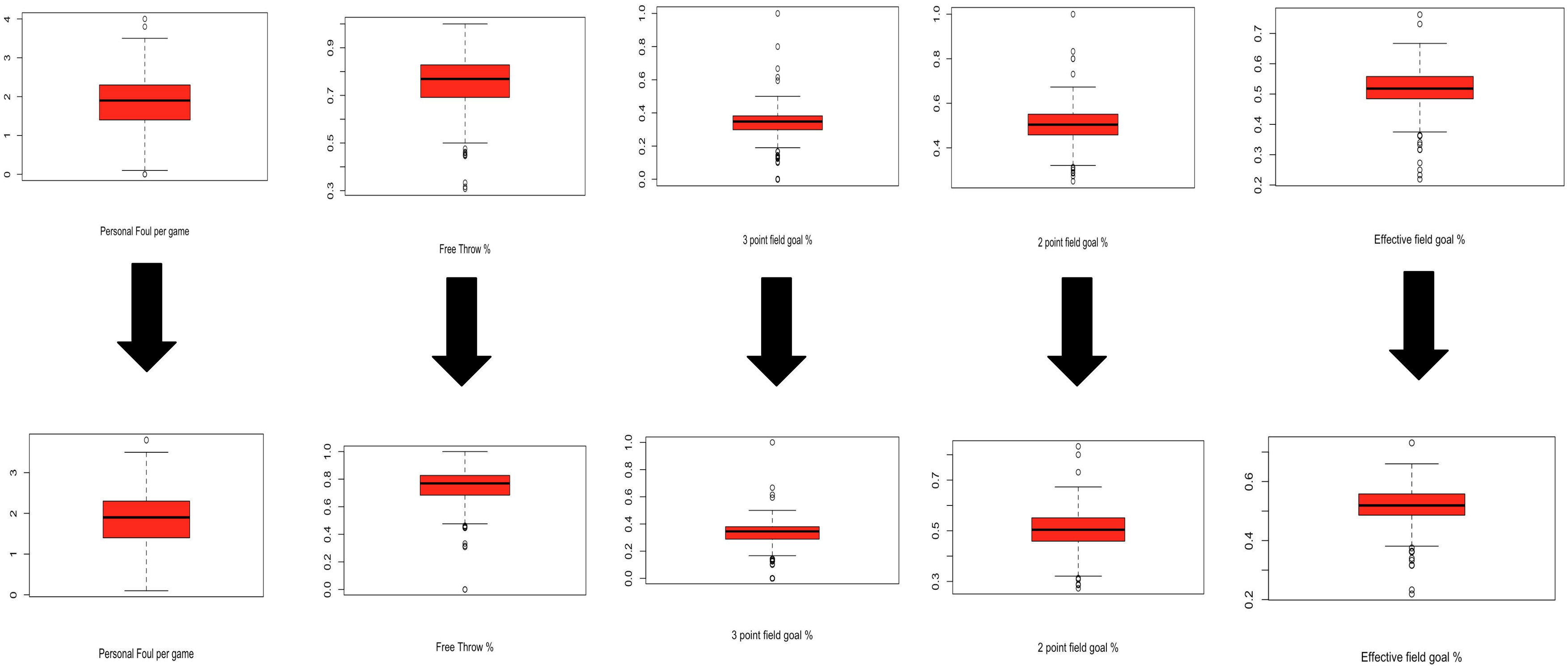
REAL Outliers

Identifying valid outliers:

Outliers due to abnormally high or low stats of players who played less than **10** games

Player	Salary	Pos	Age	G	GS	MP	FGP	3PP	2PP	eFGP	FTP	TRB	AST	STL	BLK	TOV	PF	PPG
Tony Bradley	1679520	C	20	9	0	3.2	0.273	0	0.3	0.273	1	1.2	0.1	0	0	0	0.1	0.9
Chinanu Onu	1544951	C	21	1	0	22	0.4	0	0.4	0.4	0	4	1	0	0	3	4	4
Justin Patton	2667600	C	20	1	0	4	0.5	0	0.5	0.5	0	0	0	1	0	0	1	2
Josh Smith	5331729	PF	32	3	0	4	0.25	0	0.25	0.25	0	1.3	0	0	0	0	1	0.7
Tyler Lydon	1874640	PF	21	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0
Jon Leuer	10002681	PF	28	8	0	17	0.417	0	0.455	0.417	0.867	4	0.6	0.1	0.4	0.9	2	5.4
Wade Baldw	1544951	PG	21	7	0	11.4	0.667	0.8	0.625	0.762	0.6	1.1	0.7	0.3	0.1	0.6	1.7	5.4
Monte Morri	1349383	PG	22	3	0	8.3	0.667	0	0.8	0.667	1	0.7	2.3	1	0	0.3	0.3	3.3
Jeremy Lin	13768421	PG	29	1	1	25	0.417	0.5	0.4	0.458	1	0	4	0	0	3	3	18
Luol Deng	16747954	SF	32	1	1	13	0.5	0	0.5	0.5	0	0	1	1	0	1	0	2
Daniel Hamil	1349383	SG	22	6	0	4.7	0.455	0.4	0.5	0.545	0	0.8	1.3	0.2	0	0.3	0.3	2

Changes in Predictor Variables



Multiple Regression

Building the Multiple Regression

First Step

Erased 4 Least Statistically Significant Variables:

- STL[Steals]
- FTP [Free-Throw Percentage]
- X2PP [Two-Point Field Goal Percentage]
- MP [Minutes per games

Building the Multiple Regression

Second Step

Erased variables with multicollinearity

- eFGP (GVIF=6.682006)

Why eFGP?

VIFs of **POS**, **FGP**, **eFGP**, **AST**, and **TOV** are higher than 5

3 Ways to calculate eFGP (Effective field goal %)

1) (FG+(0.5*3P))/FGA
2) (PPG–FT/2)/FGA
3) (2FG+(1.5*3FG))/FGA

BASKETBALL REFERENCE

Erasing the Multicollinear variable

- Removing eFGP from model helped out adjusted R-Squared increase from **0.5609** to **0.5625**
- Number of variables with high **VIF** decreased from **5** to **2**

Building the Multiple Regression

Last Step

3) Erased 3 Least Statistically Significant Variables once again

- FGP [Field goal percentage] + BLK [Block] + X3PP [3 points percentage]
- Removing these variables increased out adjusted R^2 from **.557** to **.5645**
- Pos [Position] was also insignificant, however removing it brought our R^2 fro, **.5645** to **5616**

Results

Final Model: Equation

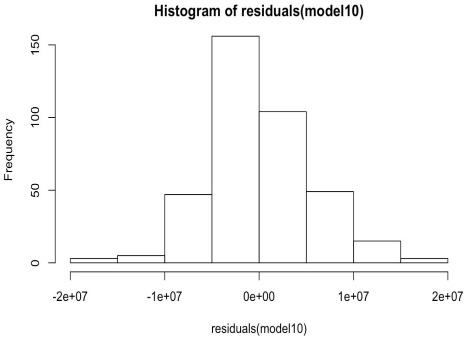
Predicted Salary =

-6,340,676 + -587,743PosPF + -3,025,079PosPG + -616,389PosSF + -956,144PosSG + 406,488Age + -66,437G + 76,171GS + 793,605TRB + 1,428,030AST -2,466,672TOV + -1,269,613PF + 616,871PPG

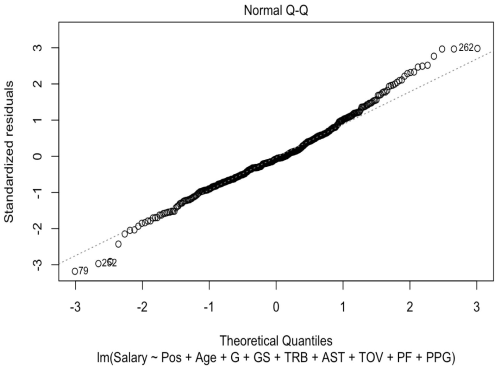
Interpretation

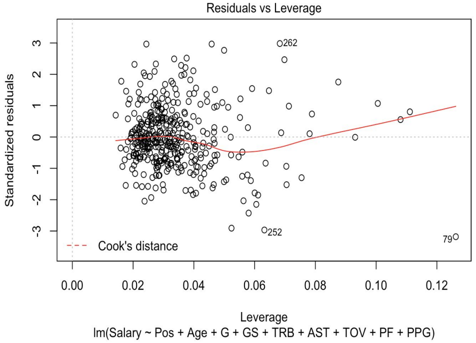
- Adjusted R^2: **56.45%**
  - 56.45% of the variability in NBA salaries in 2018-2019 can be explained by our predictor variables of player statistics in 2017-2018
- P-value of the model is less than 2.2e^-16
  - Model is statistically significant
- Residual Standard Error:
  - Measures the spread around the line. The less scatter around the line, the smaller the residual standard deviation and the stronger the relationship between x and y

Multiple Regression Assumptions



- ☑ Linearity
- ☑ Normality
- ☑ Homoscedasticity(not perfect)





Interpretation

Predicted Salary = -6,340,676 + -587,743PosPF + -3,025,079PosPG + -616,389PosSF + -956,144PosSG + 406,488Age + -66,437G + 76,171GS + 793,605TRB + 1,428,030AST -2,466,672TOV + - 1,269,613PF + 616,871PPG

“**Age** = Prime of the NBA player (30s)”

“**G** = More games played, Less potential”

“High competition in **Guard Positions** due to Modern Basketball’s **fast pace** Game Style.”

Positive factor for Salaries: **Age, G, TRB, AST**, and **PPG**

Negative factor for Salaries: **G, TOV**, and **PF**

Interpretation of **Pos:**

Ranking of the position with highest value is

Center > **Power forward** > **Small Forward** > **Shooting guard** > **Point Guard**

Model Test-Prediction

Prediction Intervals:

- We used Draymond Green's salary, close to the mean for a NBA Salary, and used his statistics to predict a salary using our model
  - Actual: $17,469,565, Prediction: $17,515,342
  - Confidence Interval: With 95% confidence, a player with Draymond Green's exact statistics will make between $14,266,769 and $20,763,915

Model Test-Prediction

Prediction Intervals:

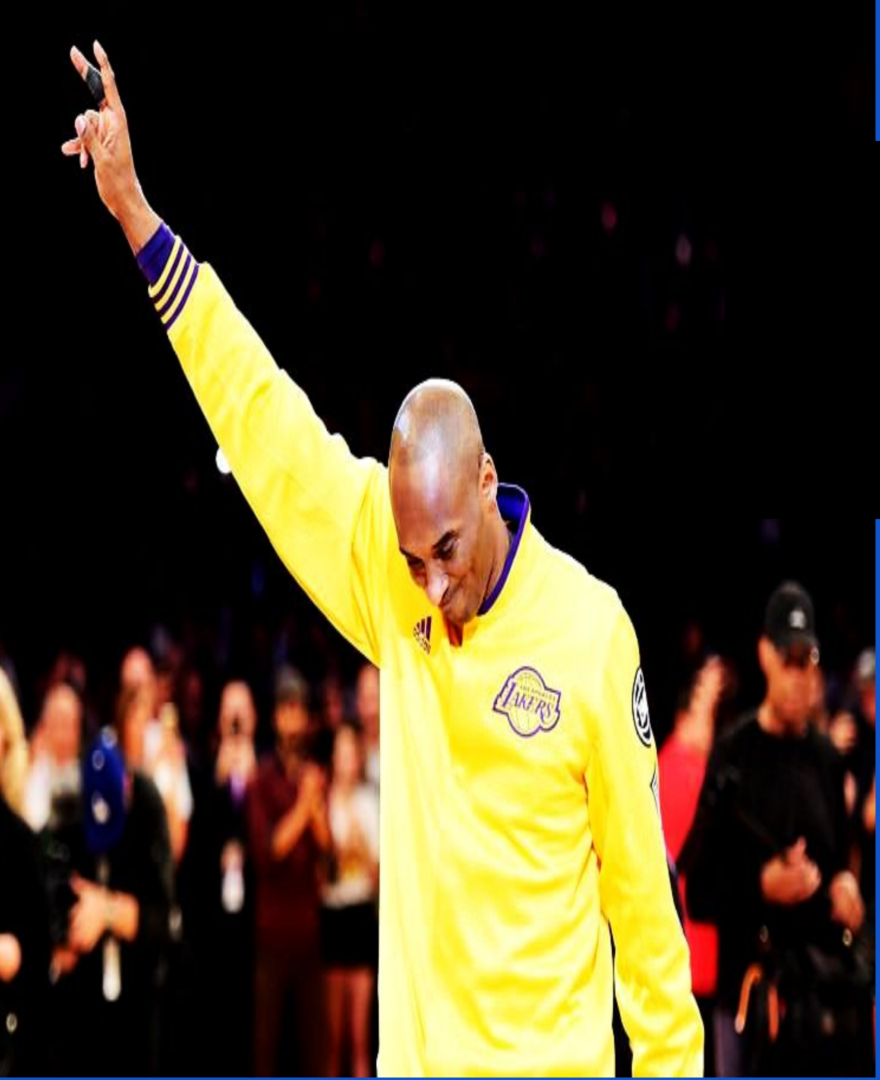
- We used Ben Simmons' salary, last year's NBA Rookie of the Year, and used his statistics to predict a salary he should be getting instead of his rookie contract using our model
    - Actual: $6,434,520
    - Prediction: $16,155,818
    - Confidence Interval: With 95% confidence, a player with Ben Simmons' rookie year statistics means he deserve a new contract between $13,523,119 and $18,788,517
- Will see if his contract will be similar as our prediction

Limitation of data set

1) **Multi-year Contract**
   a) many players in our dataset are locked in to multi-year contracts where the previous year's performance does not affect the following year's salary in this case.
2) **Removal of Rookie data**
   a) we had to remove a few players from the dataset due to their "rookie" status in the league (not having 2017-2018 statistics data)
3) **Sacrifice of players for team**
   a) Lastly, our dataset doesn't consider the fact that some players are willing to take salary cuts to stay with their team due to team salary cap restraints.

For the extension of the study

- **Advanced statistics**
  - The Use of advanced NBA statistics such as usage and efficiency rating might improve the quality of the model.
- **Data reflecting Reputation of the player**
  - Often, the reputation of the player cause increase or decrease of the salary, so inclusion of the data reflecting players' reputation will help the future extension of the study
- **College performance data for Rookie**
  - Most NBA players participate in at least one year of NCAA basketball before going to the NBA, and the performance at the college reflected as the potential for Rookie player so the performance in the NCAA will be helpful for the extension of the data



Thank you