



NBA Player Statistics

By: Brooks Jones

Why are we performing this analysis?

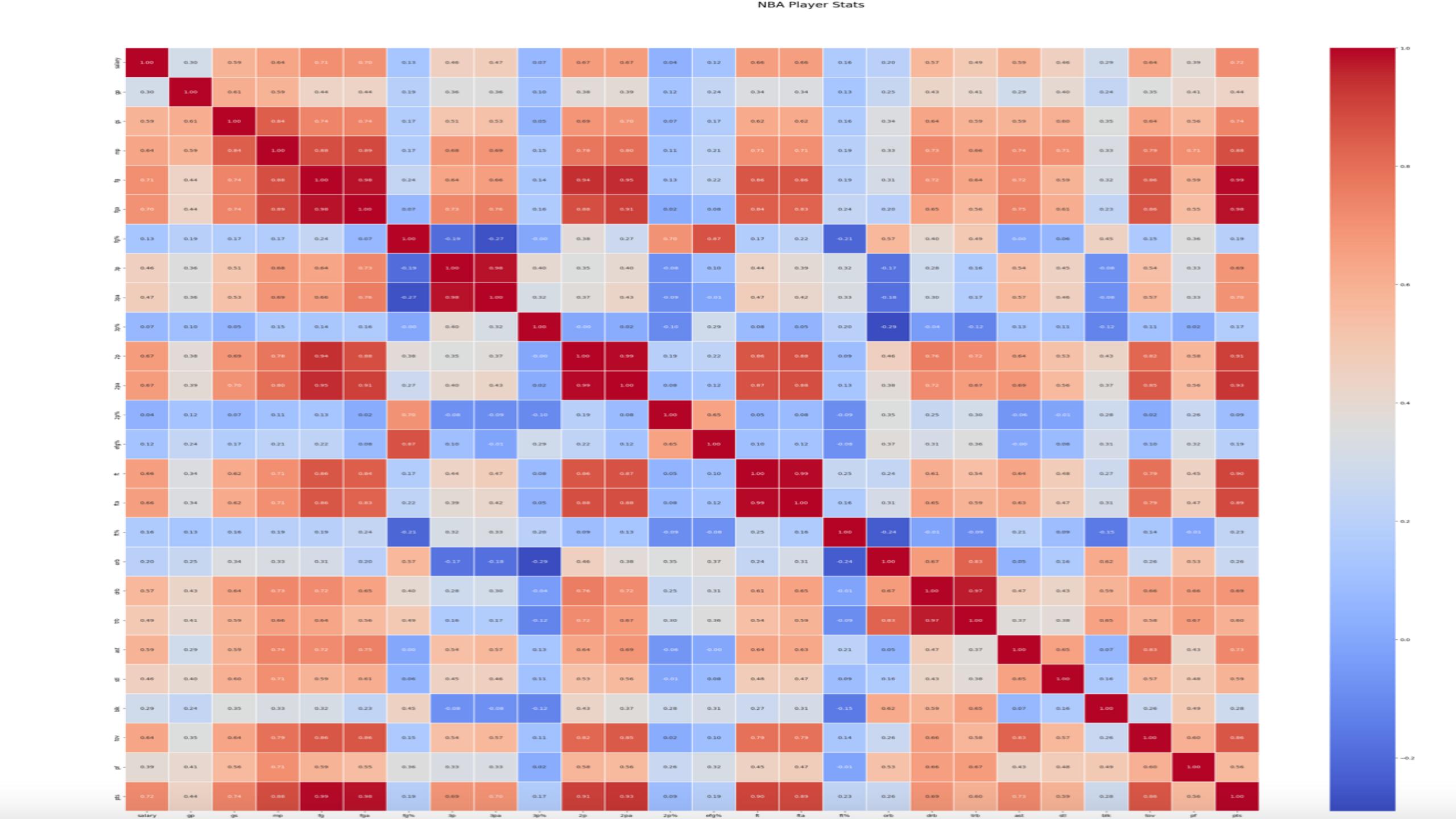
- NBA players salary are calculated on performance metrics.
- Which stats were the most important when calculating a players salary?
- How can a team use this data to calculate a players salary?
- How can a team track the performance of their players?

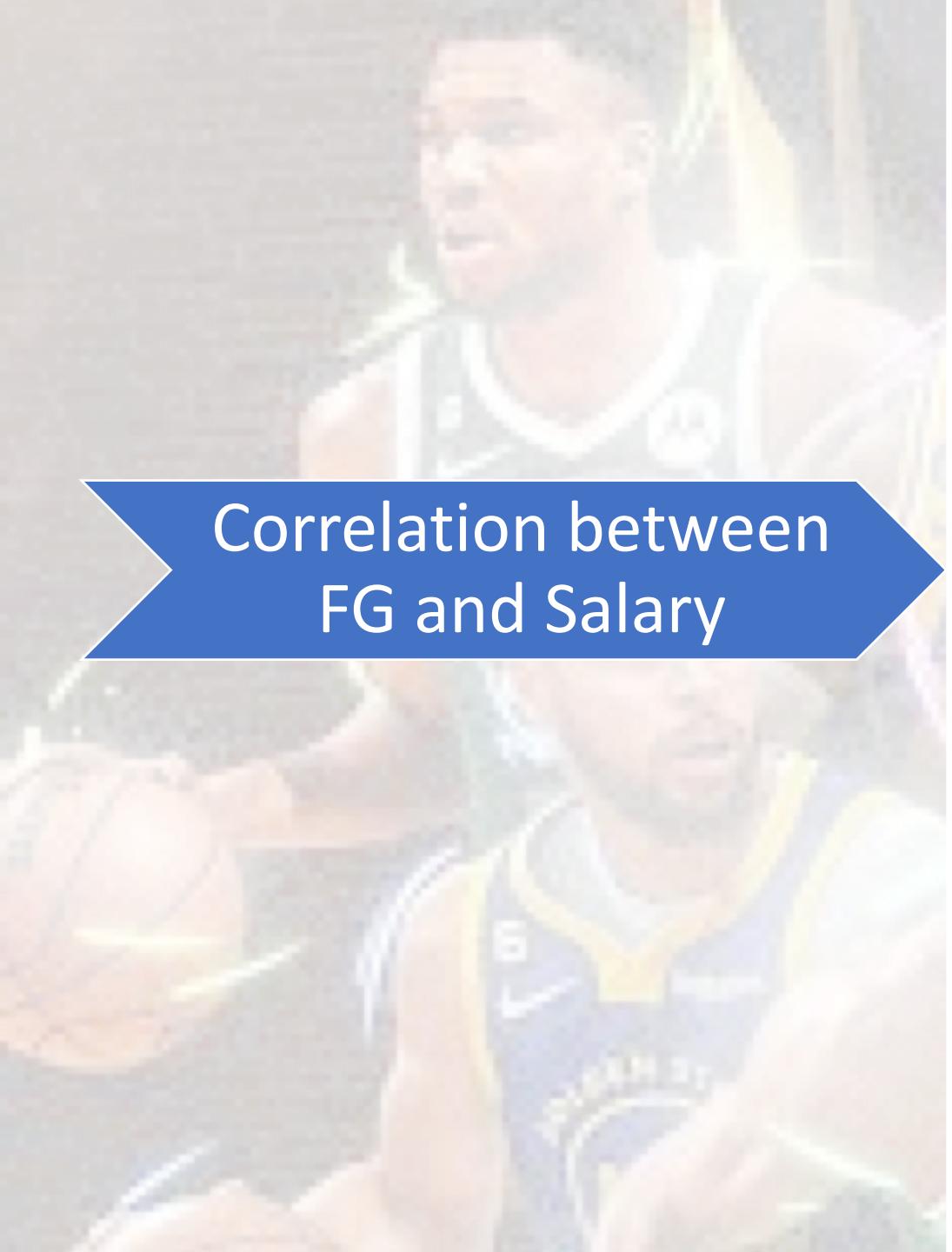
player name	salary	position	age	team	gp	gs	mp	fg	fga	fg%	3p	3pa	3p%	2p	2pa	2p%	efg%	ft	fta	ft%	orb	drb	trb
Stephen Curry	48070014	PG	34	GSW	56	56	34.7	10.0	20.2	0.493	4.9	11.4	0.427	5.1	8.8	0.579	0.614	4.6	5.0	0.915	0.7	5.4	6.1
John Wall	47345760	PG	32	LAC	34	3	22.2	4.1	9.9	0.408	1.0	3.2	0.303	3.1	6.7	0.459	0.457	2.3	3.3	0.681	0.4	2.3	2.7
Russell Westbrook	47080179	PG	34	LAL/LAC	73	24	29.1	5.9	13.6	0.436	1.2	3.9	0.311	4.7	9.7	0.487	0.481	2.8	4.3	0.656	1.2	4.6	5.8
LeBron James	44474988	PF	38	LAL	55	54	35.5	11.1	22.2	0.5	2.2	6.9	0.321	8.9	15.3	0.58	0.549	4.6	5.9	0.768	1.2	7.1	8.3
Kevin Durant	44119845	PF	34	BRK/PHO	47	47	35.6	10.3	18.3	0.56	2.0	4.9	0.404	8.3	13.4	0.617	0.614	6.5	7.1	0.919	0.4	6.3	6.7
Bradley Beal	43279250	SG	29	WAS	50	50	33.5	8.9	17.6	0.506	1.6	4.4	0.365	7.3	13.2	0.552	0.551	3.8	4.6	0.842	0.8	3.1	3.9
Kawhi Leonard	42492492	SF	31	LAC	52	50	33.6	8.6	16.8	0.512	2.0	4.8	0.416	6.6	11.9	0.551	0.572	4.7	5.4	0.871	1.1	5.4	6.5
Paul George	42492492	SF	32	LAC	56	56	34.6	8.2	17.9	0.457	2.8	7.6	0.371	5.4	10.3	0.521	0.536	4.6	5.3	0.871	0.8	5.3	6.1
Giannis Antetokounmpo	42492492	PF	28	MIL	63	63	32.1	11.2	20.3	0.553	0.7	2.7	0.275	10.5	17.6	0.596	0.572	7.9	12.3	0.645	2.2	9.6	11.8
Damian Lillard	42492492	PG	32	POR	58	58	36.3	9.6	20.7	0.463	4.2	11.3	0.371	5.4	9.4	0.574	0.564	8.8	9.6	0.914	0.8	4.0	4.8



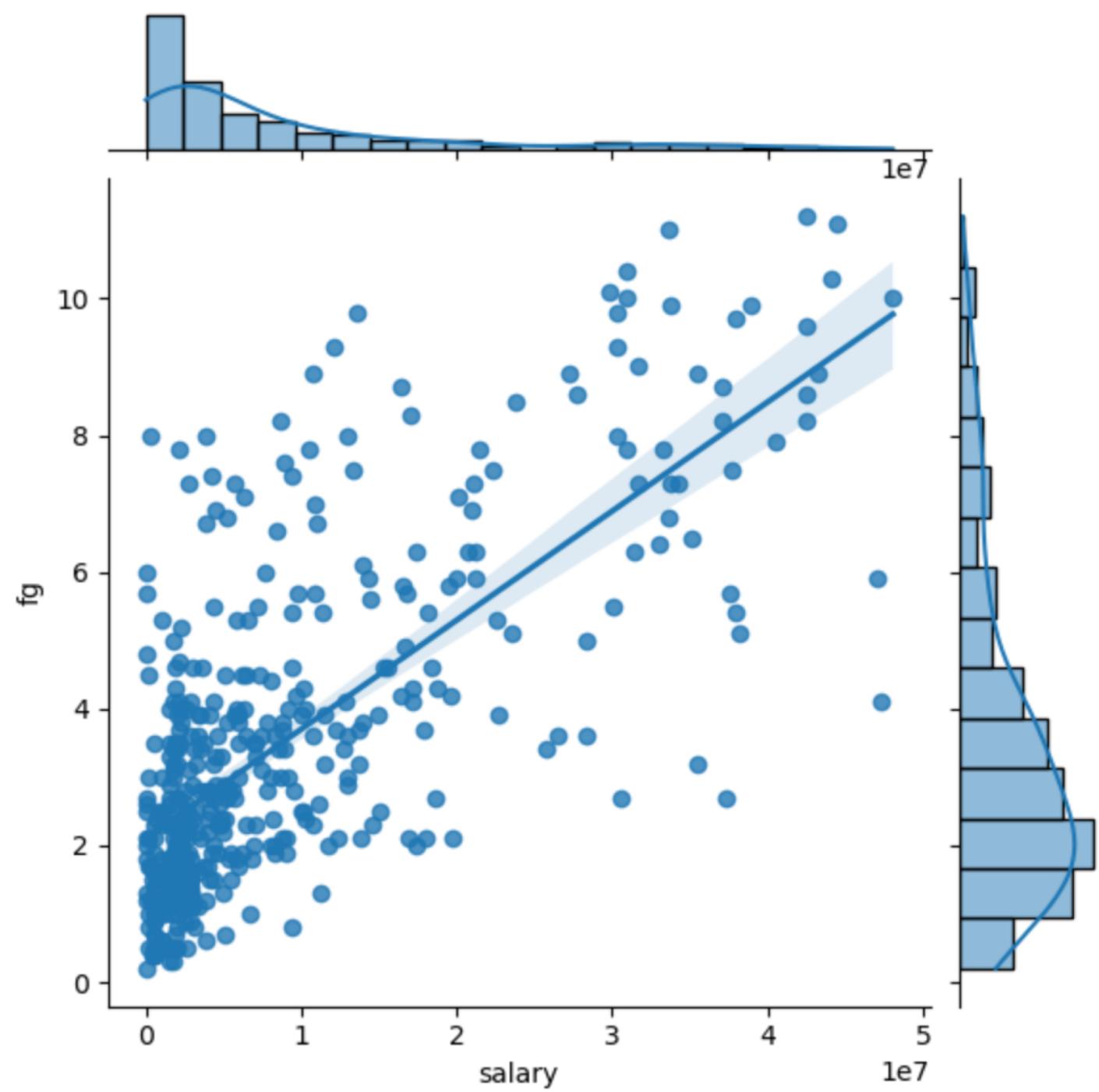
Exploratory Data Analysis



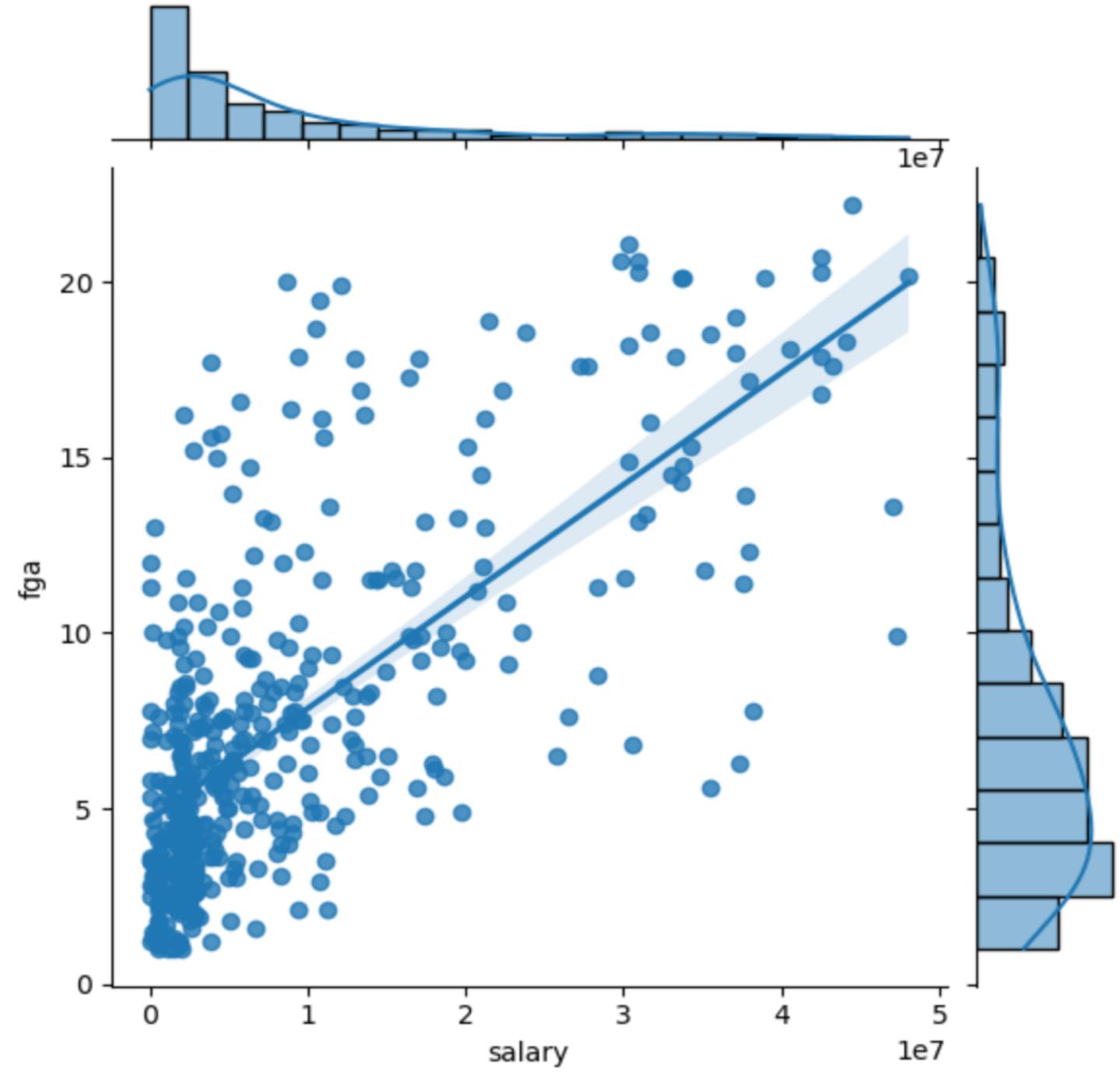




Correlation between
FG and Salary

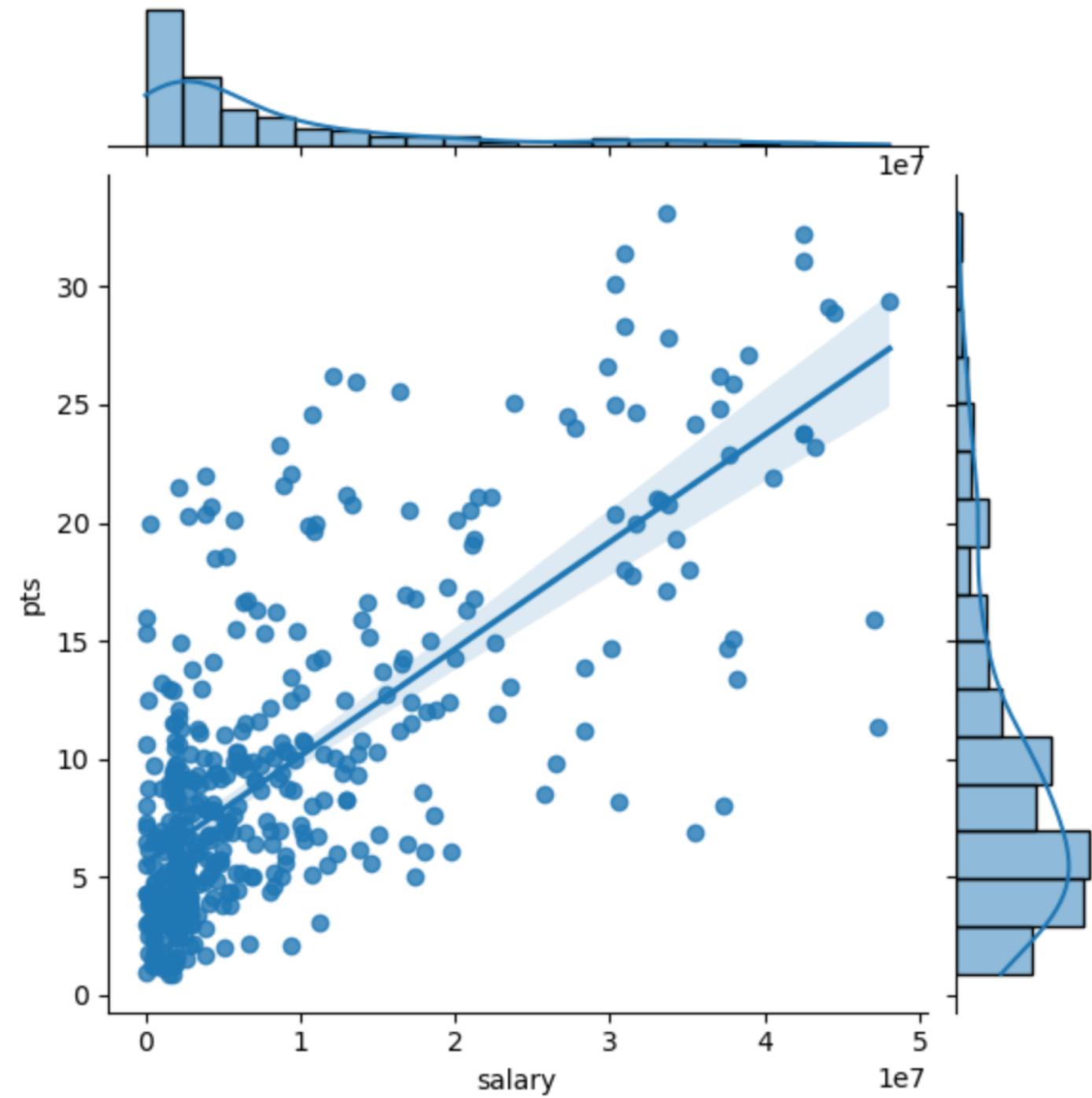


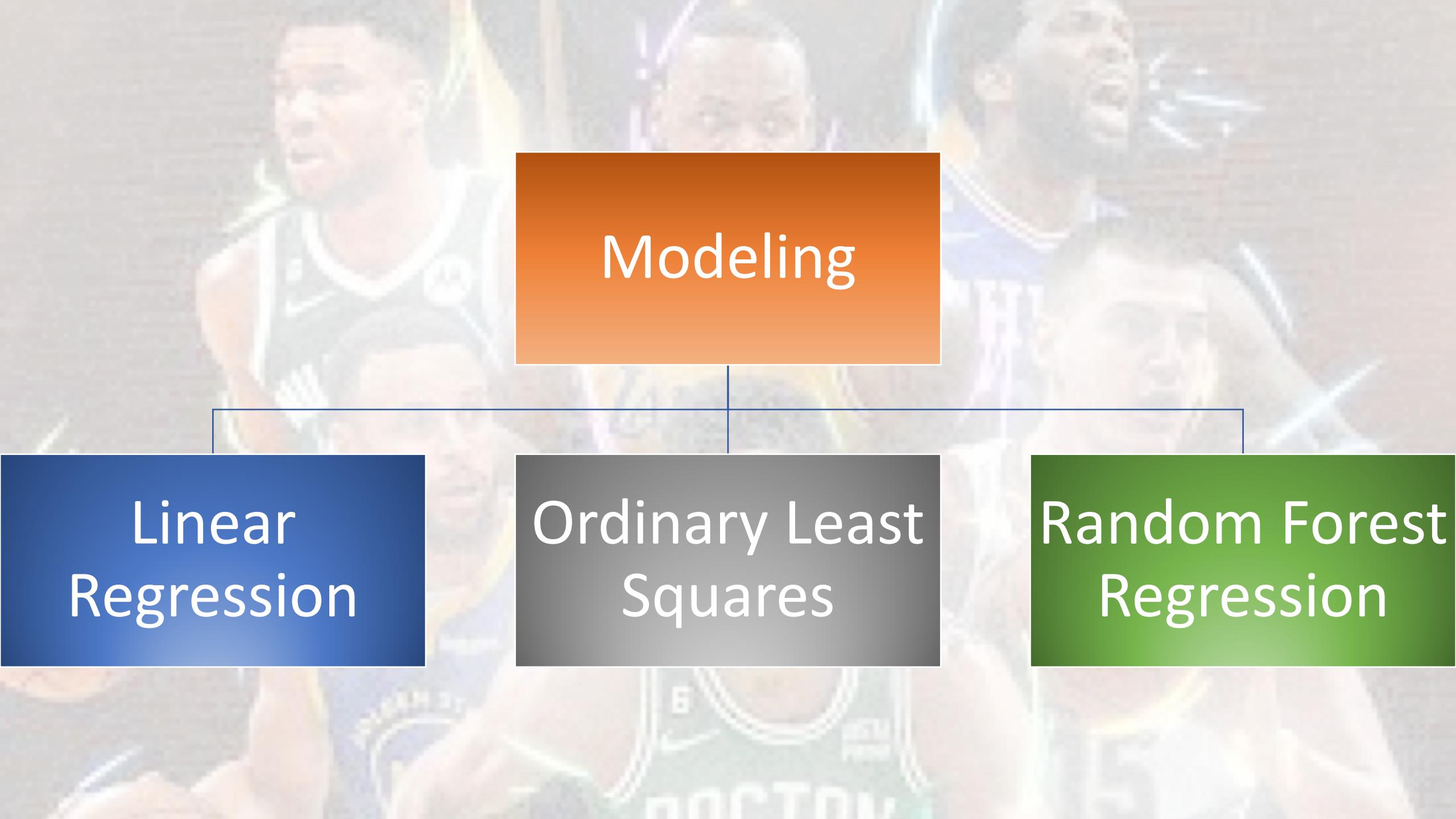
Relationship between FGA and Salary





Relationship between PTS and Salary





Modeling

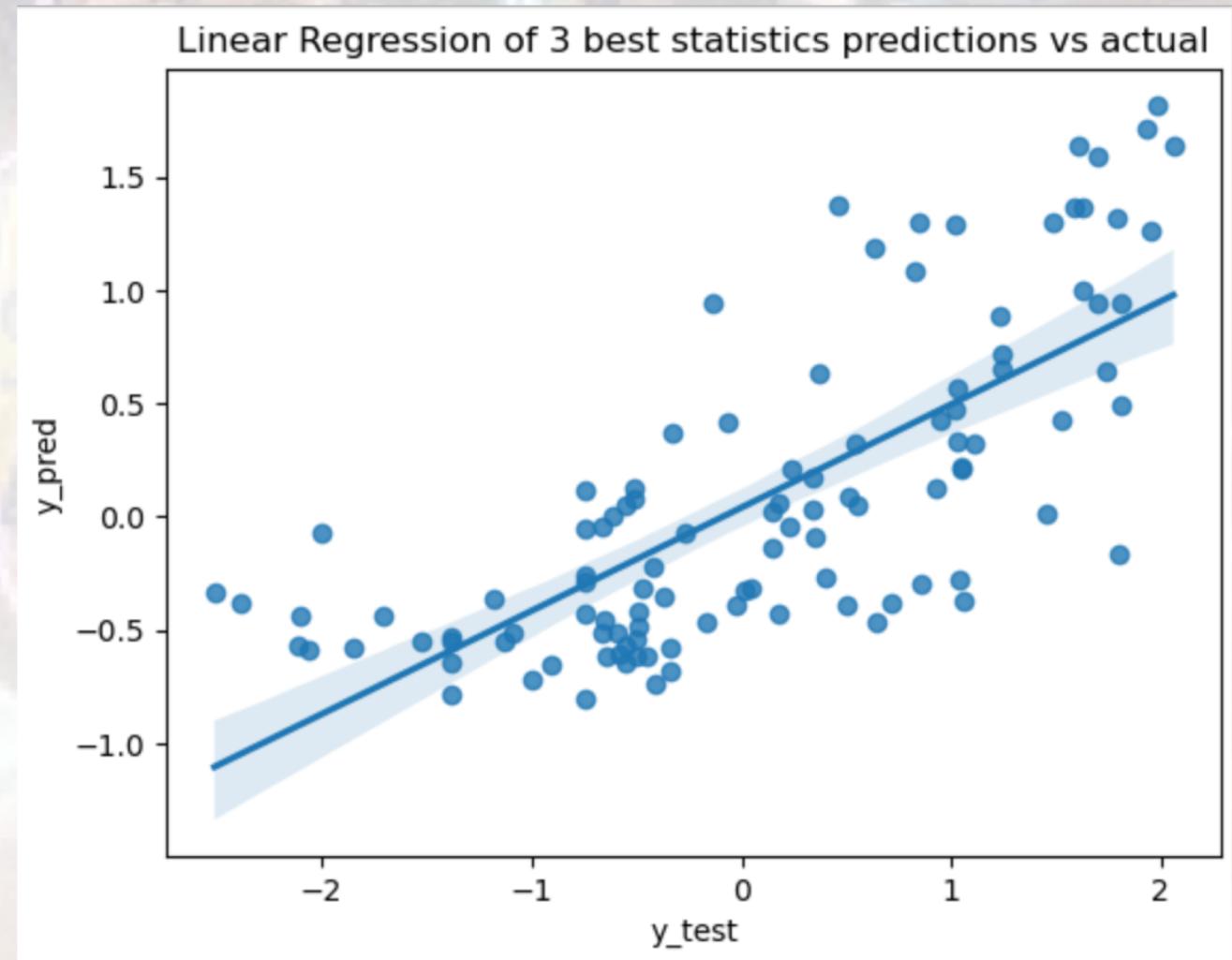
Linear
Regression

Ordinary Least
Squares

Random Forest
Regression

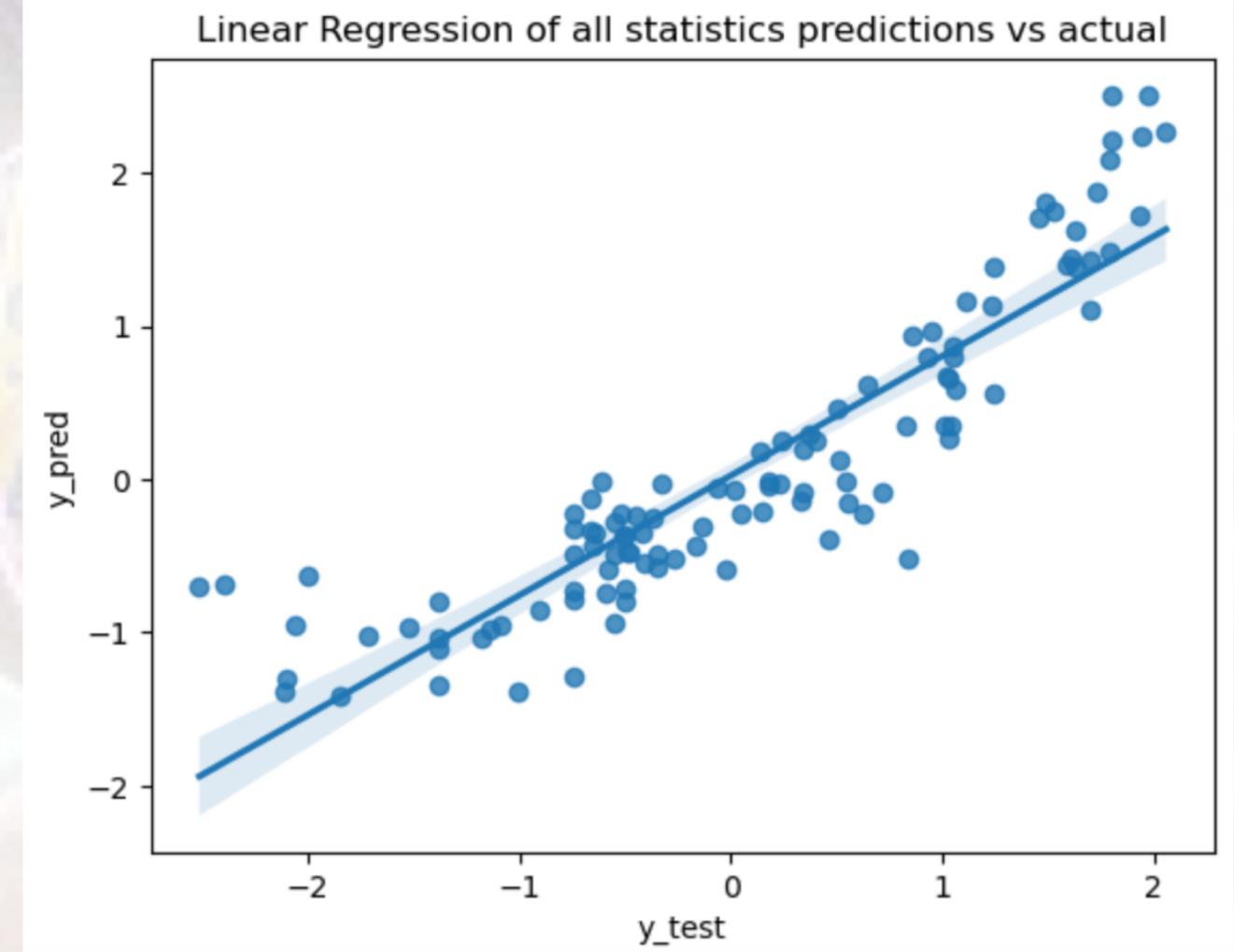
Modeling of our 3 best statistics using linear regression

- Here we see the relationship between the predicted and actual values.
- We also had a root mean squared error of 0.76. We want to be closer to zero for this metric, so we'll run more models.



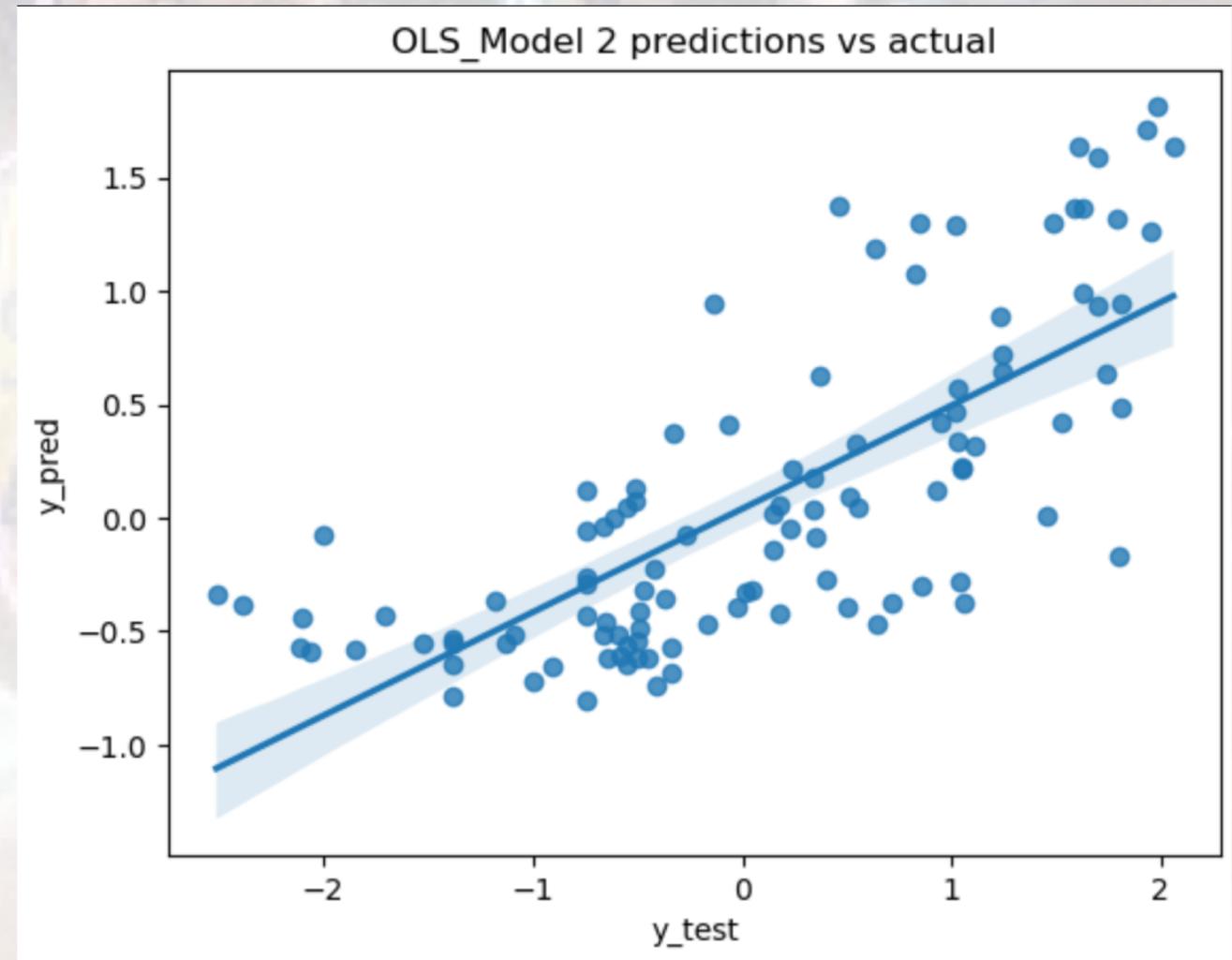
Linear regression of all stats.

- You can see how well this model performed with how close to the trend line the data points are.
- We had a RMSE of 0.49 for this model. So way less error than our last model.
- This will ultimately be our best performing model.



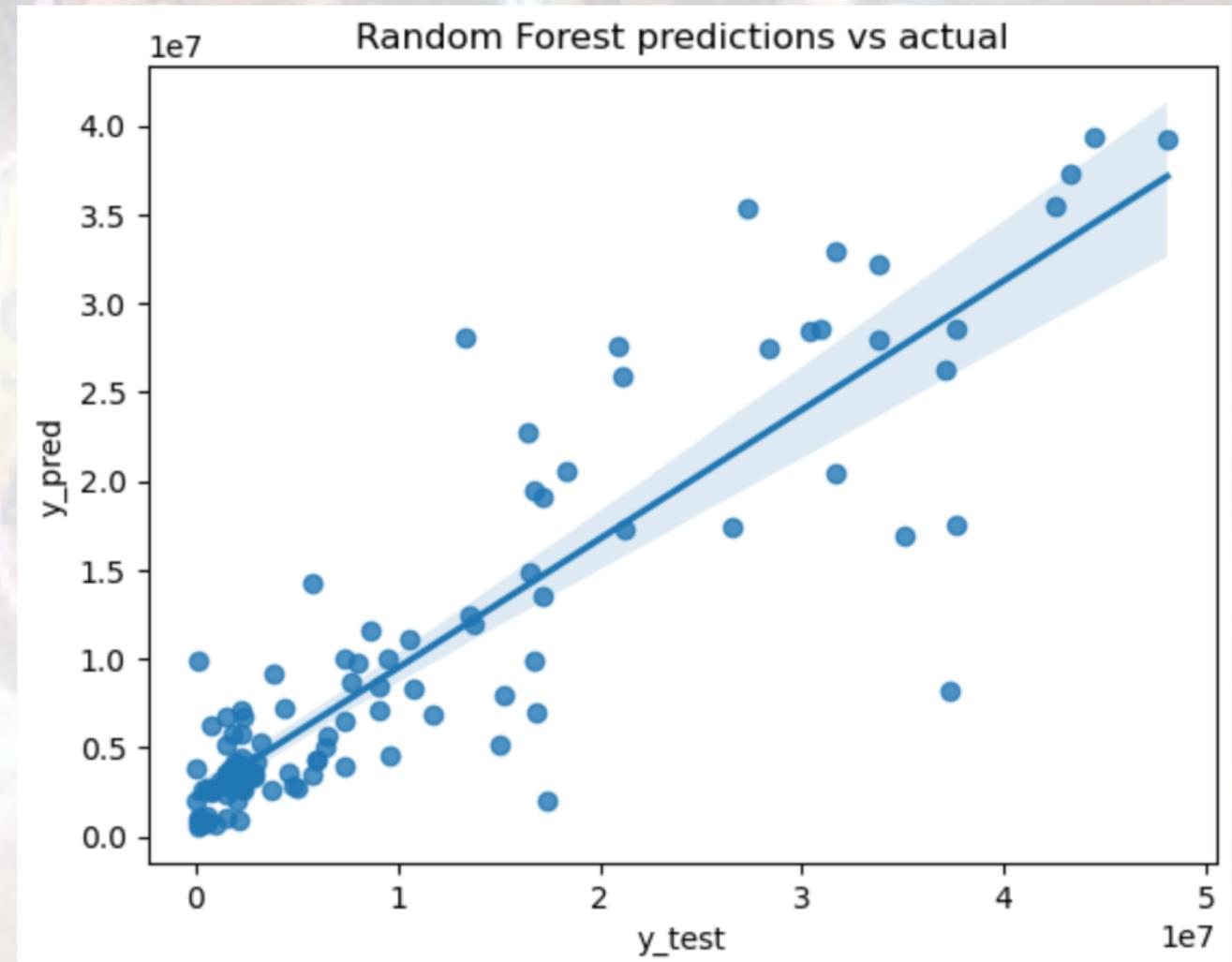
Ordinary Least Squares Model

- This model also performed well but by having an RMSE of 0.51.
- The graph shows the positive correlation between the data.



Random Forest Regression

- This was the worst of our three models that we chose.
- The random forest had an RMSE of 5845684.85. So our model miss calculated the salary by over 5.8 million.
- You can see from the graph how poorly correlated the data is.



Recommendations

1. Using this data set with a linear regression teams will be able to calculate a players salary based off of performance stats within the data set.
2. Teams will also be able to track the progress of each player as they progress on a yearly basis.
3. Lastly this model will help teams calculate each players salary for each year to stay within the salary cap.