

A REGRESSION ANALYSIS TO DETERMINE PREDICTORS OF NFL PLAYER SALARIES

Jim Peterman, PhD
GitHub: [JimPeterman](#)



INTRODUCTION

Rationale

- One of the most important decisions when running an NFL team is deciding how much to pay the players.

Objective

- Build and compare regression models to determine what features are associated with player salary.

Goal

- Determine what features a team should consider when negotiating player salaries.




METHODS: DATA

- 2020 team and player information scrapped from pro-football-reference.com using Python's BeautifulSoup


Features

- Provided on website:* age, years in NFL, games played, games started, weight, height, position, calculated player “value”, team record
- Created:* general position (offense/defense/special teams), position group (e.g., defensive line), height/weight compared to others within same position, starter (no/sometimes/yes), body mass index (BMI)



Enter Person

PlayersTeamsSeasonsLeadersNFL Scores¹⁴DraftStats



via Sports Logos.net
About logos

2020 Denver Broncos Statistics & Players

« Previous SeasonNext Season »

Record: 5-11-0, 4th in [AFC West Division](#) ([Schedule and Results](#))
Coach: [Vic Fangio](#) (5-11-0)
Points For: 323 (20.2/g) 28th of 32
Points Against: 446 (27.9/g) 25th of 32
Expected W-L: 5.1-10.9
SRS: -6.01 (28th of 32), **SOS:** 1.68
Offensive Coordinator: [Pat Shurmur](#)
Defensive Coordinator: [Ed Donatell](#)

More Team Info ▼

Roster

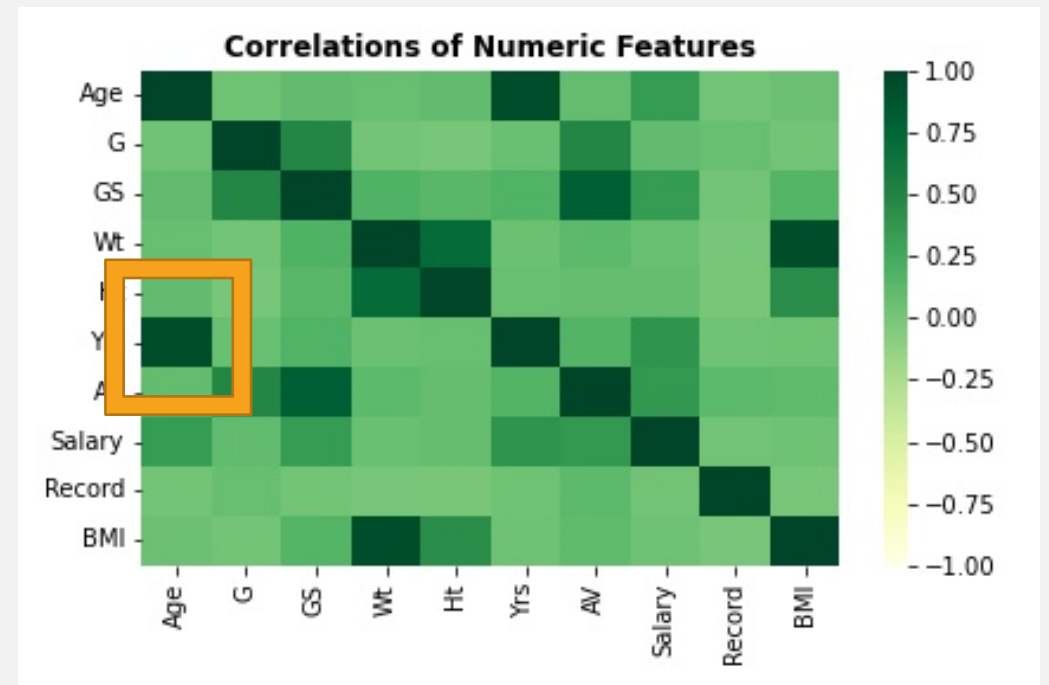
*ProBowl, +1st-tm All-Pro Share & Export ▼ Glossary Toggle Per-Game Stats

No.	Player	Age	Pos	G	GS	Wt	Ht	College/Univ	BirthDate	Yrs	AV	Drafted (tm/rnd/yr)	Salary
95	McTelvin Agim	23	DL	10	0	309	6-3	Arkansas	9/25/1997	Rook	0	Denver Broncos / 3rd / 95th pick / 2020	\$610,000
76	Calvin Anderson	24	T	16	2	300	6-5	Rice, Texas	3/25/1996	Rook	2		\$675,000
97	Jeremiah Attaochu	27	OLB	13	5	262	6-3	Georgia Tech	1/17/1993	6	4	San Diego Chargers / 2nd / 50th pick / 2014	\$1,000,000
75	Quinn Bailey	25	T	1	0	323	6-6	Arizona St.	10/18/1995	1	0		\$510,000
34	Essang Bassey	22	DB	12	3	190	5-10	Wake Forest	8/12/1998	Rook	3		
41	DeVante Bausby	27	DB	10	3	190	6-2	Pittsburg St.	1/15/1993	4	0		\$388,235
83	Andrew Beck	24	TE	10	0	255	6-3	Texas	5/15/1996	1	0		\$552,942
32	LeVante Bellamy	24	RB	5	0	188	5-9	Western Michigan	11/28/1996	Rook	0		\$502,353
46	Jacob Bobenmoyer	23	LS	16	0	235	6-2	Northern Colorado	5/28/1997	Rook	2		
72	Garett Bolles	28	OT	15	15	300	6-5	Snow (UT), Utah	5/27/1992	3	9	Denver Broncos / 1st / 20th pick / 2017	\$1,969,351
21	A.J. Bouye	29	CB	7	7	191	6-0	Central Florida	8/16/1991	7	0		\$13,000,000
51	Nigel Bradham	31	OLB	1	0	241	6-2	Florida St.	9/4/1989	8	0	Buffalo Bills / 4th / 105th pick / 2012	
19	Fred Brown	27	WR	1	0	195	6-1	Mississippi St.	12/1/1993	2	0		\$585,000
80	Jake Butt	25	TE	5	1	250	6-6	Michigan	7/11/1995	2	0	Denver Broncos / 5th / 145th pick / 2017	\$825,000

METHODS: MODELS

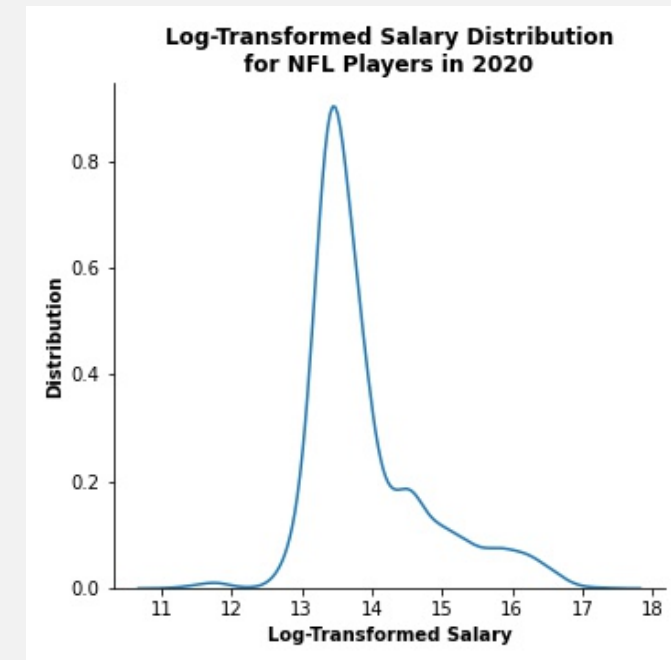
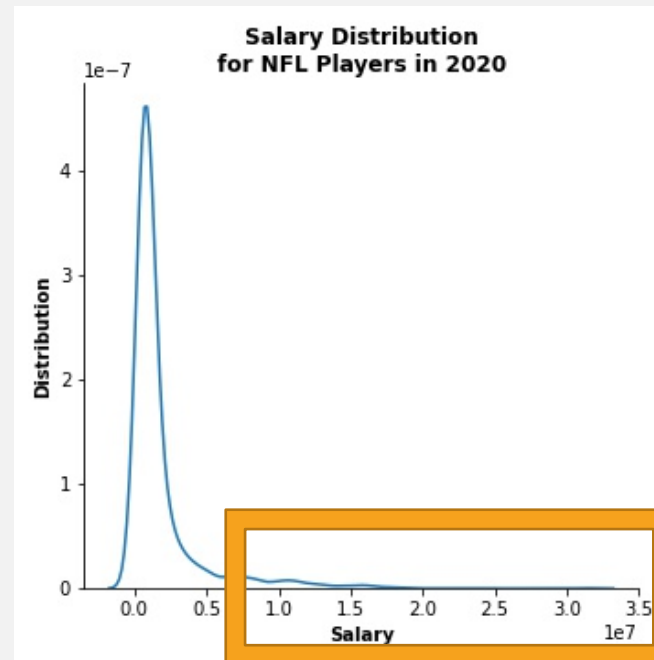
Exploratory Data Analysis

- Checked for relationships between features
- Examined distribution of the target



Regression Models

- OLS, polynomial, ridge, and Lasso
- Models trained using cross validation and then tested on holdout data
- *Experimented with dropping outliers when fitting models*
 - *Didn't improve the models*



REGRESSION MODEL RESULTS (ON TEST DATA SAMPLE)

OLS Regression

- R^2 : 0.42
- MAE: \$1,136,849
- RMSE: \$2,839,722

Polynomial Regression

- R^2 : 0.50
- MAE: \$1,068,752
- RMSE: \$2,765,948

Ridge Regression

- R^2 : 0.40
- MAE: \$1,144,876
- RMSE: \$2,855,567

Lasso Regression

- R^2 : 0.39
- MAE: \$1,144,419
- RMSE: \$2,855,079

Predicted values converted back to dollars for MAE and RMSE

REGRESSION MODEL RESULTS (ON TEST DATA SAMPLE)

OLS Regression

- R^2 : 0.42
- MAE: \$1,136,849
- RMSE: \$2,839,722

Polynomial Regression

- R^2 : 0.50
- MAE: \$1,068,752
- RMSE: \$2,765,948

Ridge Regression

- R^2 : 0.40
- MAE: \$1,144,876
- RMSE: \$2,855,567

Lasso Regression

- R^2 : 0.39
- MAE: \$1,144,419
- RMSE: \$2,855,079

REGRESSION MODEL RESULTS (ON TEST DATA SAMPLE)

OLS Regression

- R^2 : 0.42
- MAE: \$1,136,849
- RMSE: \$2,839,722

Polynomial Regression

- R^2 : 0.50
- MAE: \$1,068,752
- RMSE: \$2,765,948

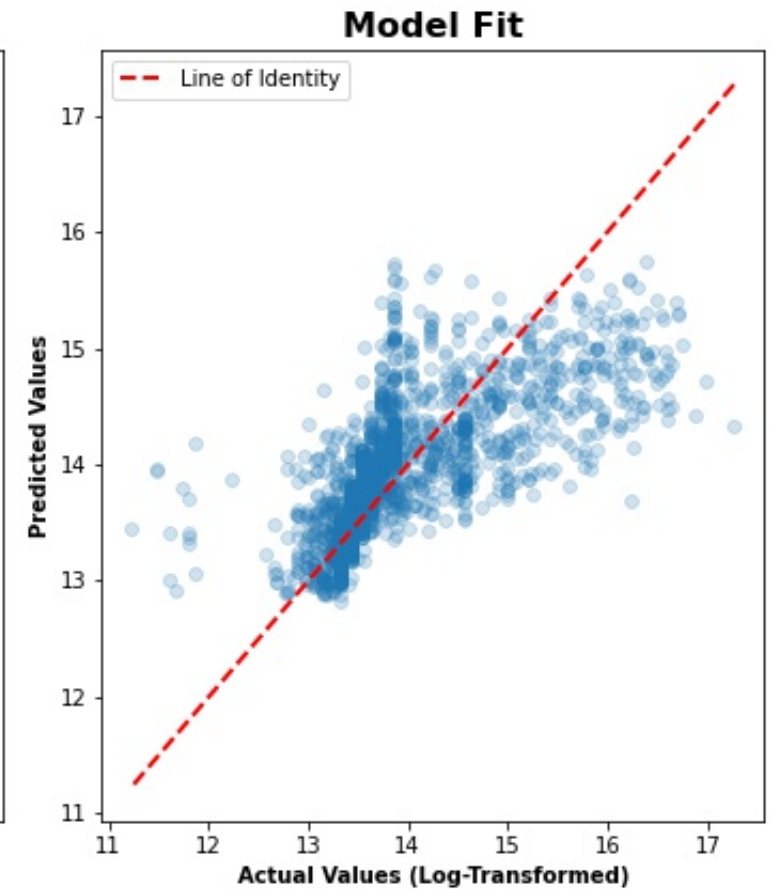
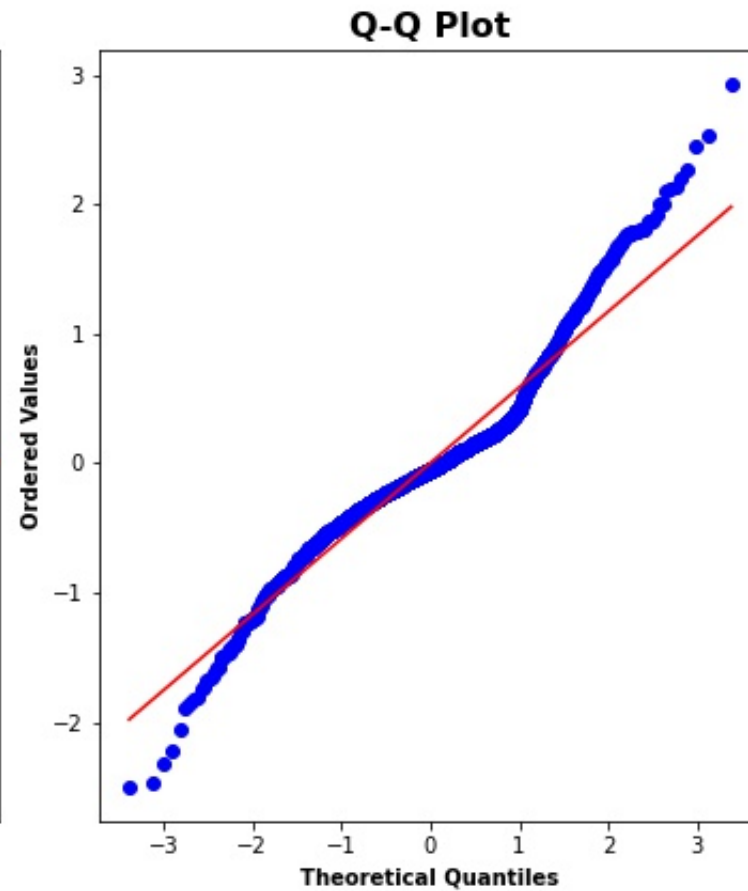
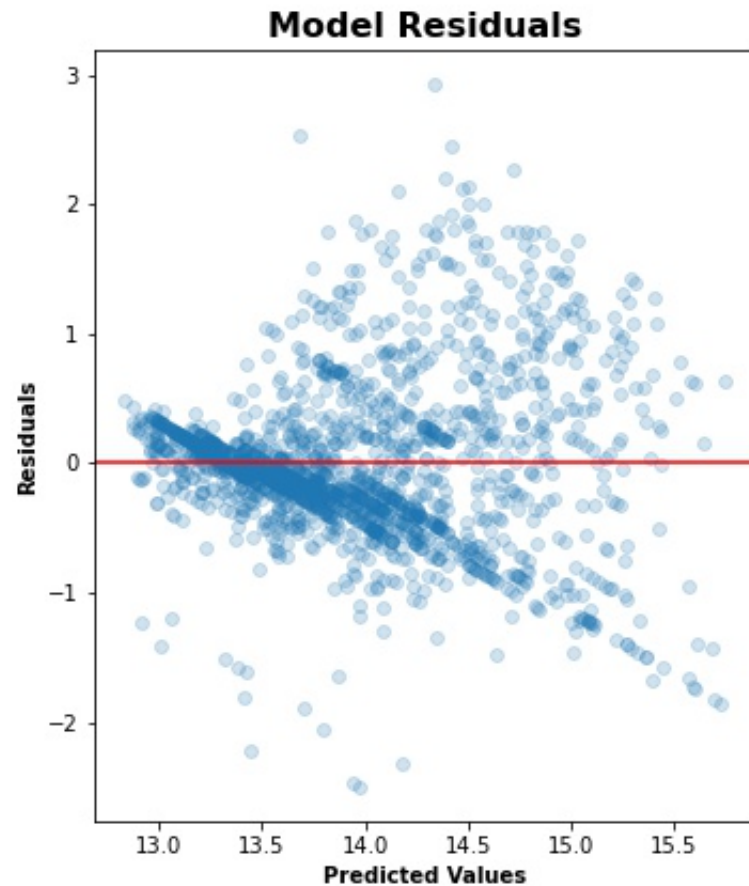
Ridge Regression

- R^2 : 0.40
- MAE: \$1,144,876
- RMSE: \$2,855,567

Lasso Regression

- R^2 : 0.39
- MAE: \$1,144,419
- RMSE: \$2,855,079

MODEL RESIDUALS AND FIT



CONCLUSIONS

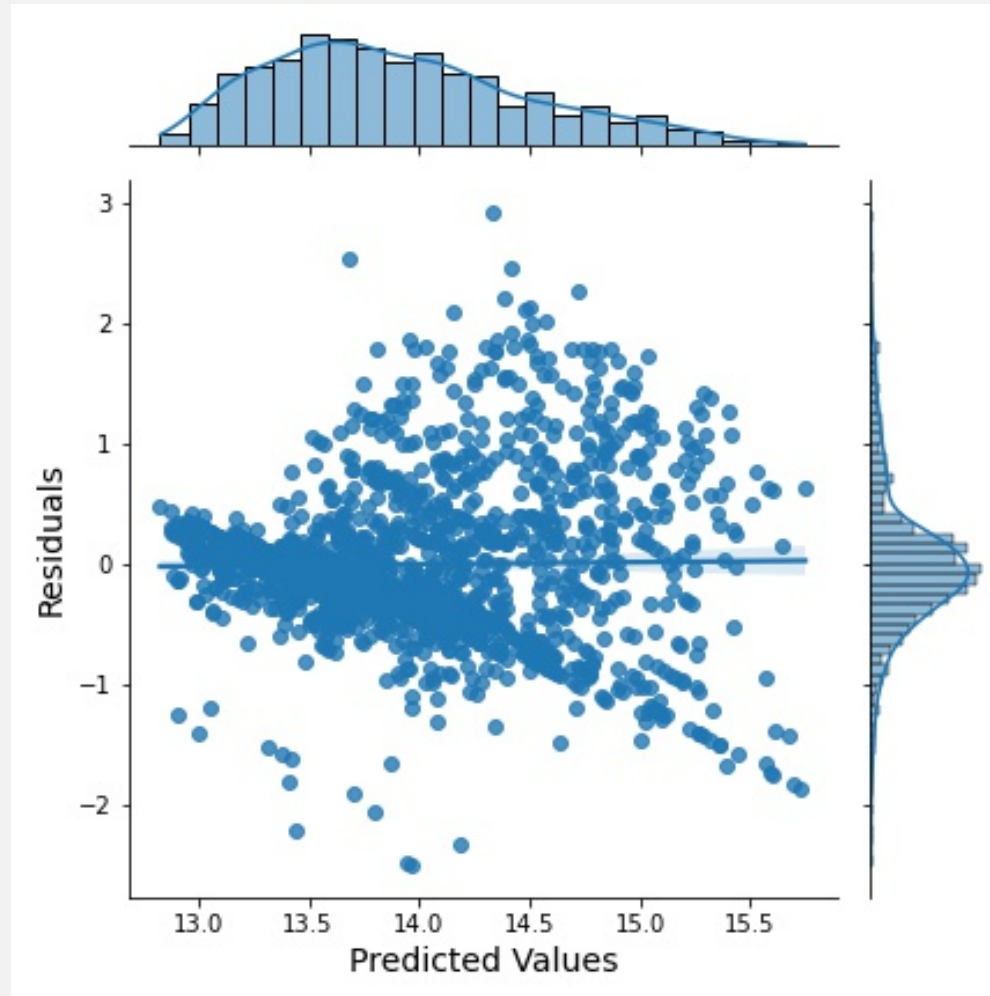
- A polynomial regression was the best performing model for this project.
- The “substantial” mean absolute error suggests other features may improve predictions of player salary
 - Although some features may be hard to quantify (ie, work ethic, quality as a teammate, motivation)
 - Salaries are not always rational



QUESTIONS?



ADDITIONAL RESIDUALS GRAPH



RESULTS FROM DROPPING OUTLIERS TO (POTENTIALLY) IMPROVE PREDICTIVE MODEL

Training Dataset (no outliers)

OLS Regression

- R^2 : 0.40
- MAE: \$862,850
- RMSE: \$1,868,327

Polynomial Regression

- R^2 : 0.46
- MAE: \$816,313
- RMSE: \$1,772,114

Test Dataset (includes outliers)

OLS Regression

- R^2 : 0.39
- MAE: \$1,140,944
 - (~\$4,000 higher than other model)
- RMSE: \$2,899,822

Polynomial Regression

- R^2 : 0.49
- MAE: \$1,071,644
 - (~\$3,000 higher than other model)
- RMSE: \$2,792,746