NBA Salary Linear Regression

Kevin Joo - DA Flex Student

Introduction

Objective: Analyze NBA player stat data to predict salary and pinpoint which stats are desirable for a higher salary

Methodology

Data:
-2 years worth of salary data for every player

Tools:
-NumPy and Pandas for data manipulation
-Matplotlib and Seaborn for data visualizations
-BeautifulSoup for web-scraping
-Sklearn for linear modeling

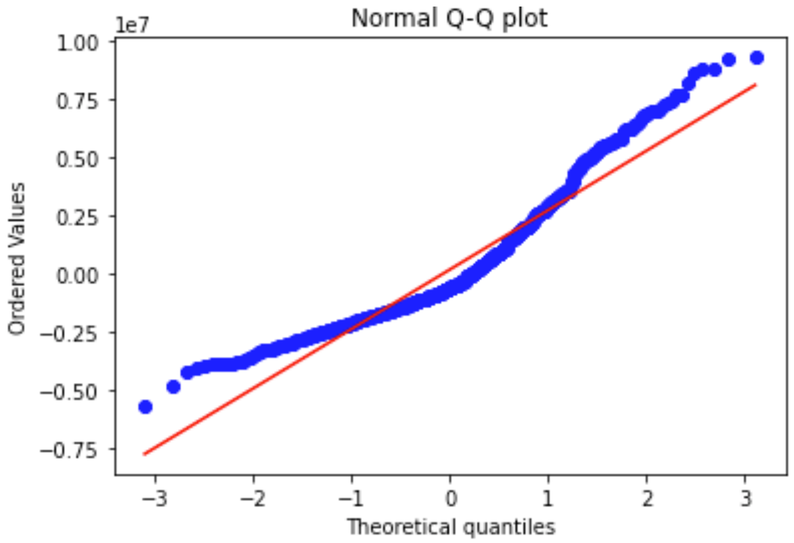
Methodology

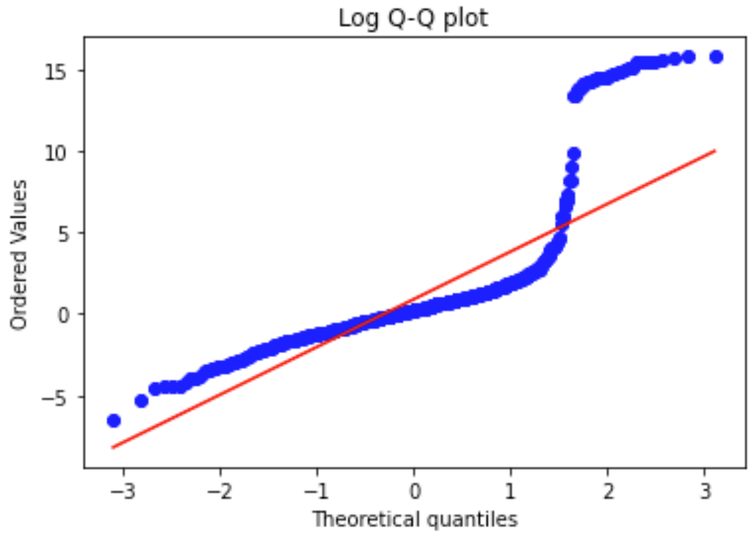
Predictor Values: “3P%”, “2P%”, “FT%”, “ORB”, “DRB”, “AST”, “STL”, “BLK”, “TOV”, “PF”

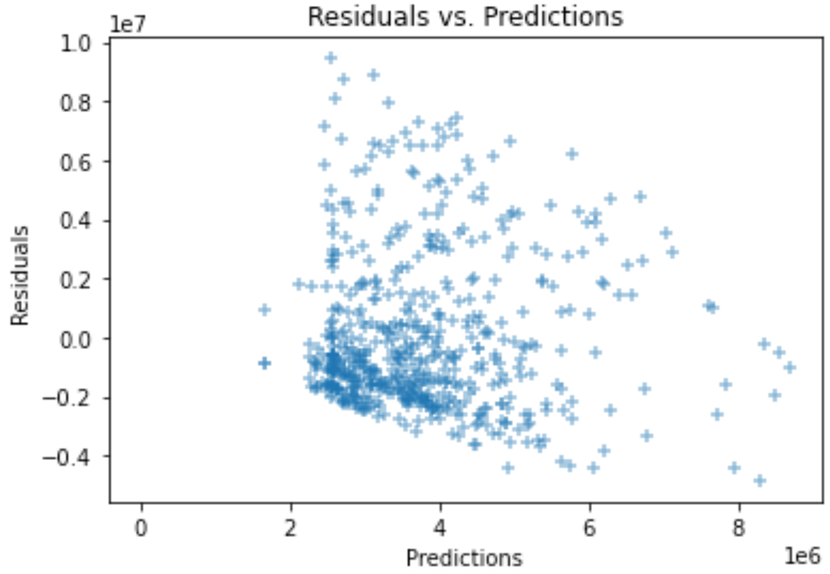
Predicting Value: “Salary”

Results

	variables	vif
0	3P%	8.379234
1	2P%	15.329776
2	FT%	19.967706
3	ORB	8.442780
4	DRB	15.073159
5	AST	12.357243
6	STL	9.077994
7	BLK	5.455680
8	TOV	17.927051
9	PF	15.679048







Results

Simple Linear Regression Model

Split the data: 20/80

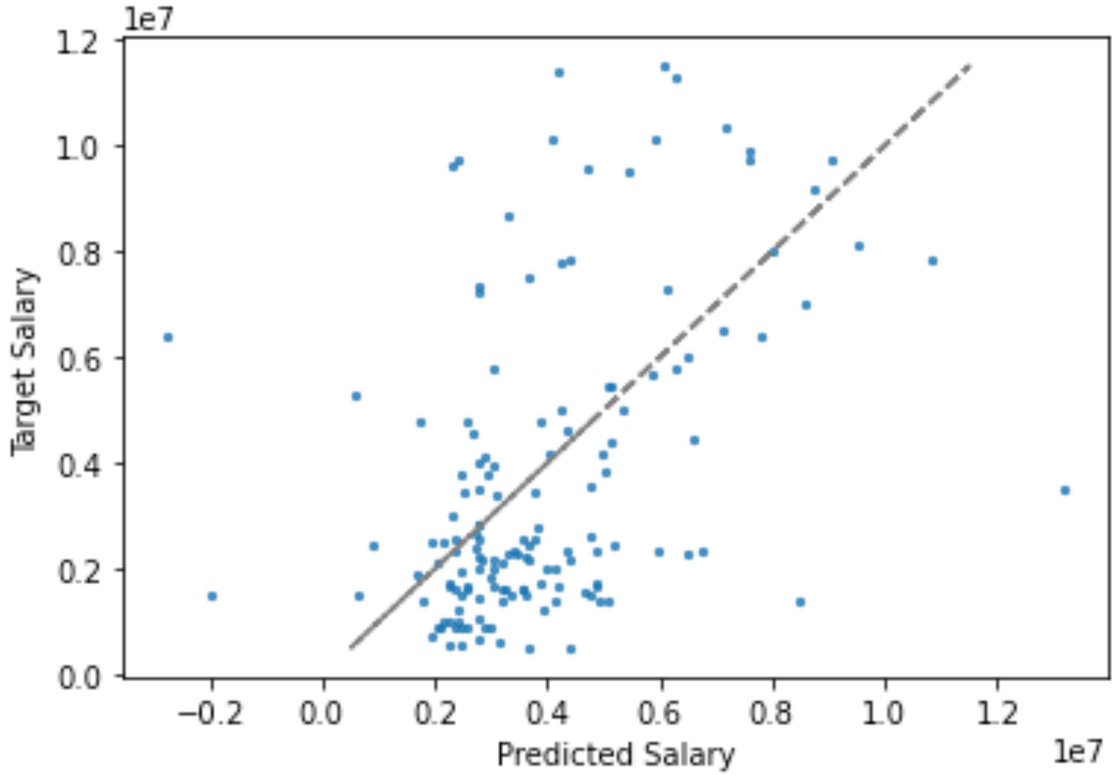
MAE: $1,996,622

R^2: 0.211

Added Polynomial Feature:

MAE: $1,925,832

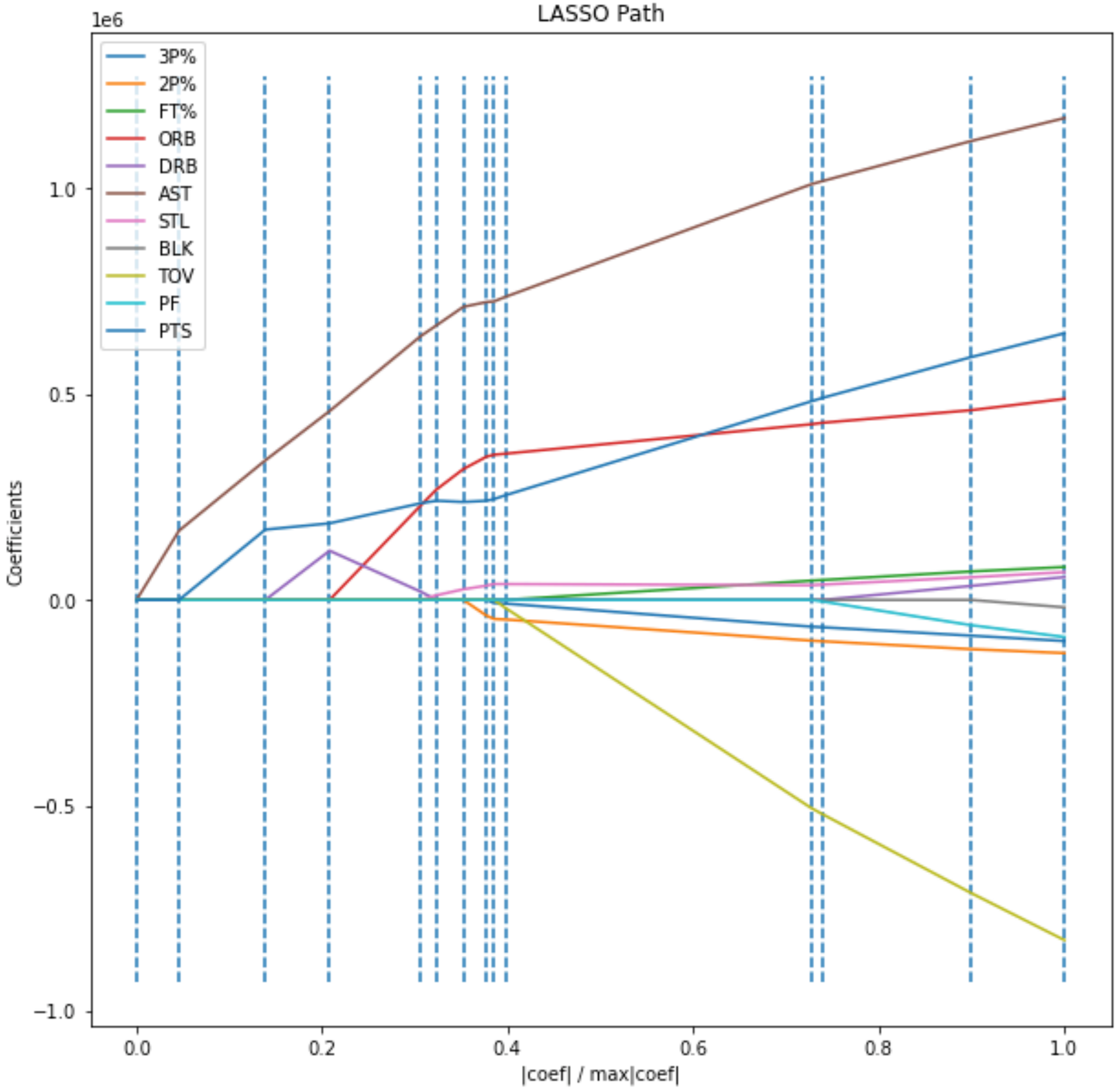
R^2: 0.213



Results

Lasso Model
Split the data: 20/80
R^2 Train: 0.160
R^2 Test: 0.210

Entry Variable:
AST: 1st
3P%: 2nd
DRB: 3nd



Future Work

If I had more time:
- Investigate further into the data to improve upon the R^2
- Lowering the heteroskedasticity
- Incorporating interaction terms to help with the model

End