

Predicting NBA Salaries

Humza Khan

Agenda

- Motivation
- Data Collection
- Exploratory Data Analysis
- Feature Selection
- Modeling
- Results
- Importance
- Future Improvements

Motivation

- What stats directly affect salary?



Data Collection

- Basketball-Reference
- Tools
 - Beautiful Soup

Contracts Summary Team Payrolls ▾ Player Contracts Glossary

2020-21 NBA Player Contracts

2020-21 Salary Cap: \$109,140,000 ([Salary Cap History](#))

Average Salary: \$7,956,433
Median Salary: \$3,852,027.5

Largest Guarantee: [Damian Lillard, POR](#) (\$227,626,953)

Color Key: [Player Option](#), [Team Option](#)

478 Contracts Share & more ▾ Glossary

 **BASKETBALL**
REFERENCE

Enter Person, Team, Section, etc

Players Teams Seasons Leaders Scores Playoffs Draft Stathead Newsletter


via Sports Logos.net
[About logos](#)

2019-20 NBA Player Stats: Per Game

« 2018-19 Player Stats: Per Game 2020-21 Player Stats: Per Game »

League Champion: [Los Angeles Lakers](#)
Most Valuable Player: [Giannis Antetokounmpo](#) (29.5/13.6/5.6)
Rookie of the Year: [Ja Morant](#) (17.8/3.9/7.3)
PPG Leader: [James Harden](#) (34.3)
RPG Leader: [Andre Drummond](#) (15.2)
APG Leader: [LeBron James](#) (10.2)
WS Leader: [James Harden](#) (13.1)

2019-20 NBA Season 4 Standings Schedule and Results Leaders Coaches Player Stats ▾ Other ▾ 2020 Playoffs Summary

Features

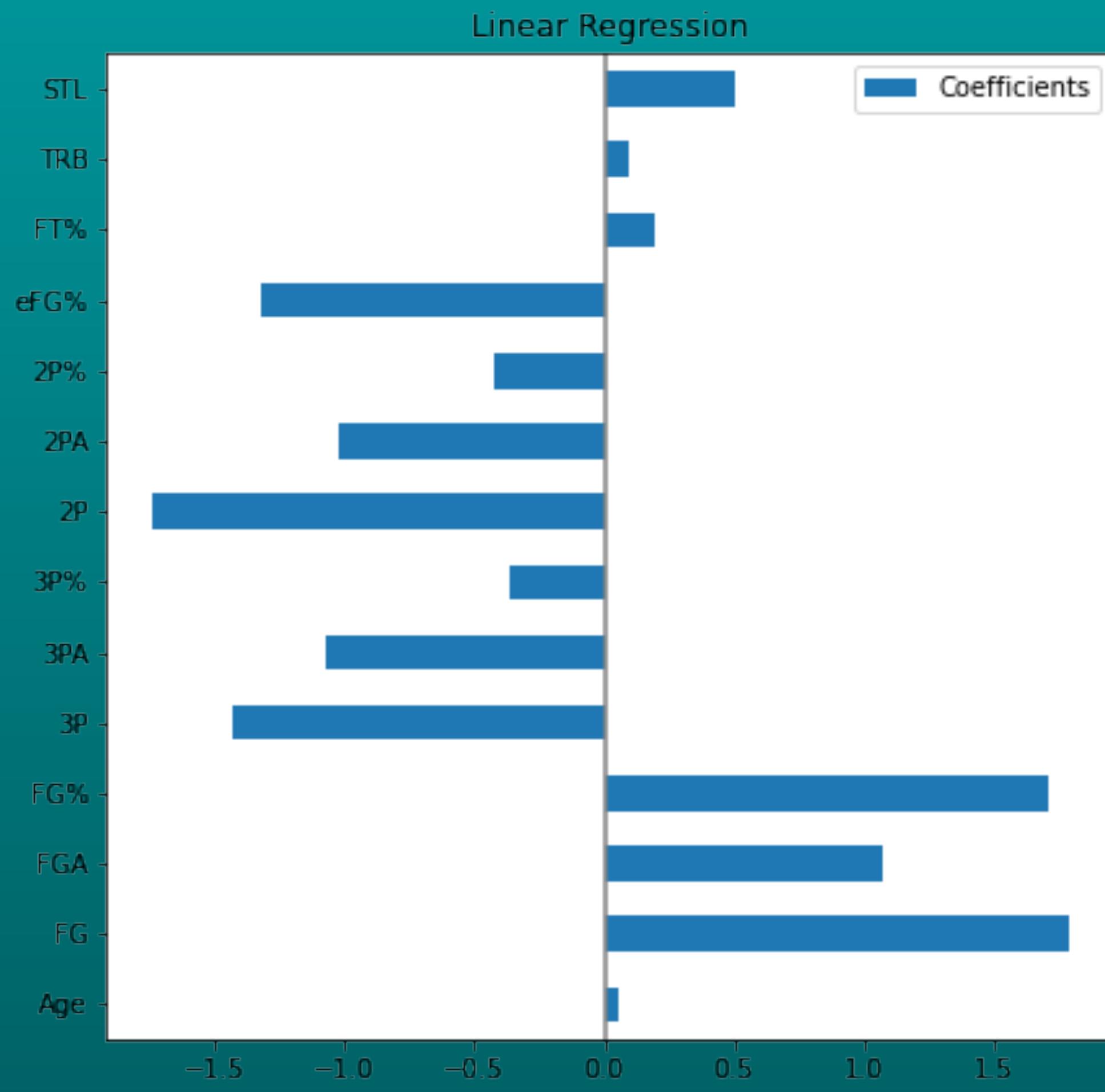
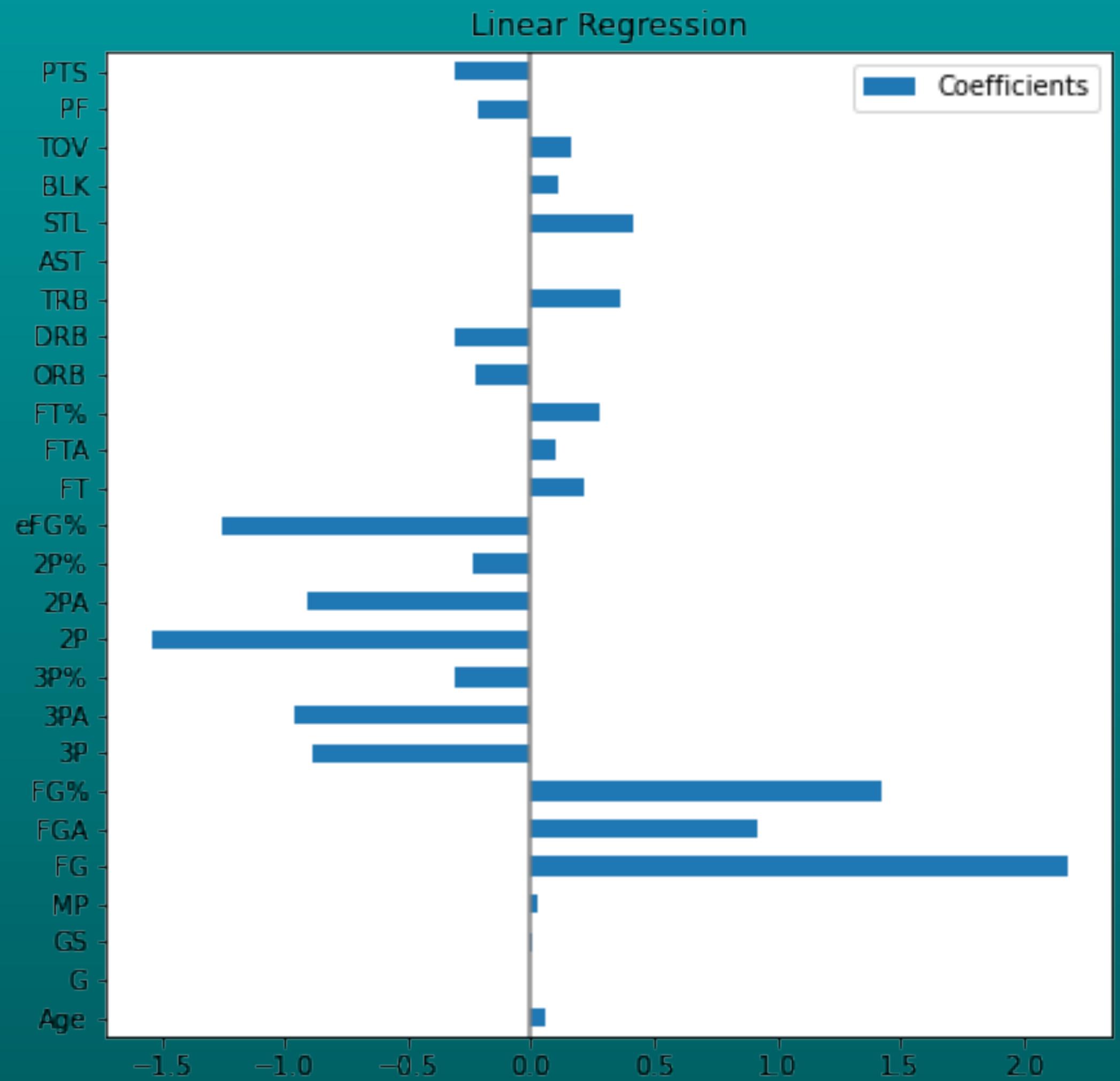
Age	Games Played	Games Started	Minutes Played	Assists per Game
Field Goals per Game	Field Goals Attempted	Field Goal Percentage	3 Pointers per Game	Steals per Game
3 Pointers Attempted per Game	3 Point Percentage	2 Pointers per Game	2 Pointers Attempted per Game	Blocks per Game
2 Point Percentage	Effective Field Goal Percentage	Free Throws per Game	Free Throws Attempted per Game	Turnovers per Game
Free Throw Percentage	Offensive Rebounds per Game	Defensive Rebounds per Game	Total Rebounds per Game	Points per Game

Exploratory Data Analysis

- Drop rows with >5 null values
- Impute 0 in missing rows
- Log transformation of salary data



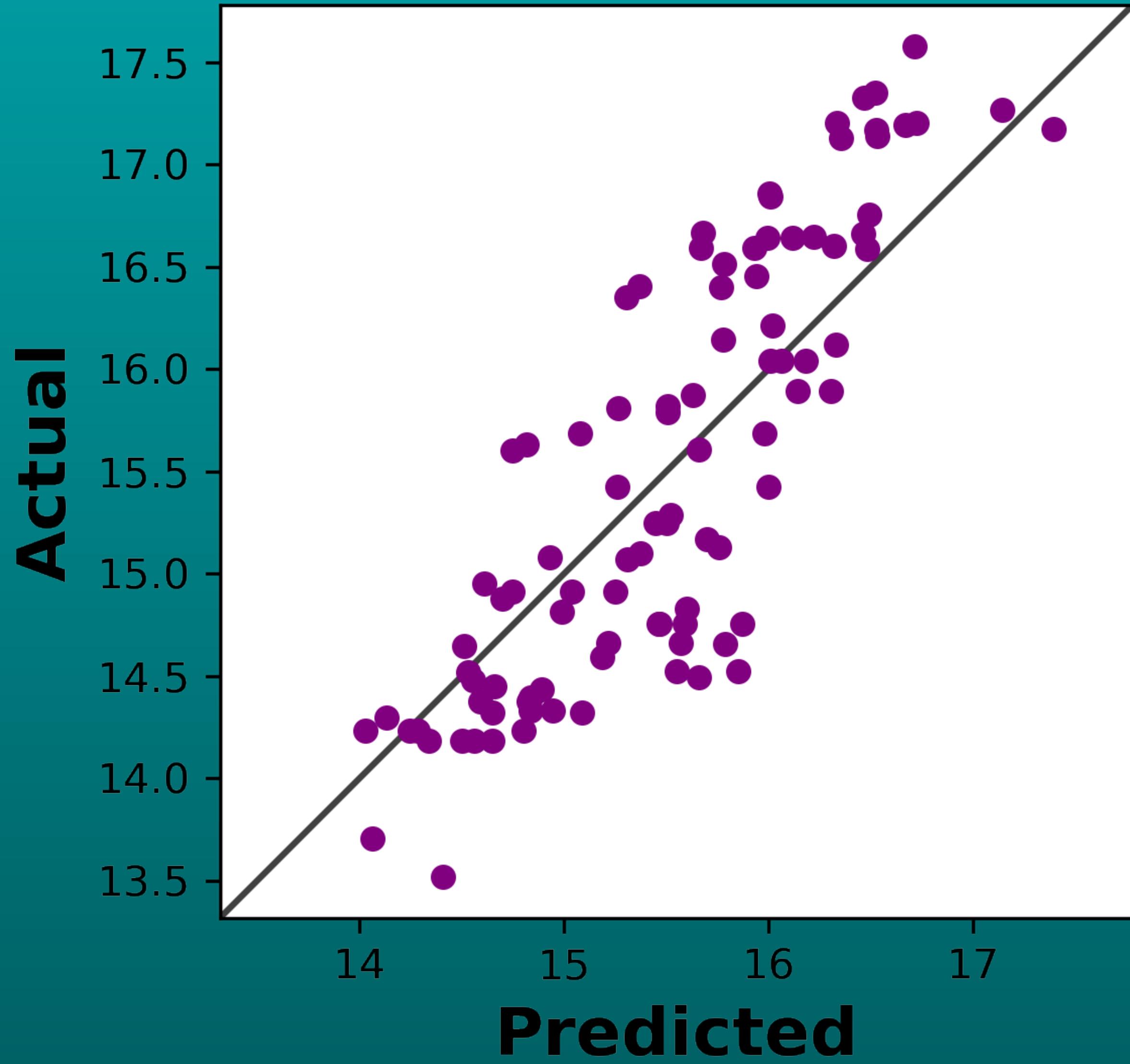
Feature Selection



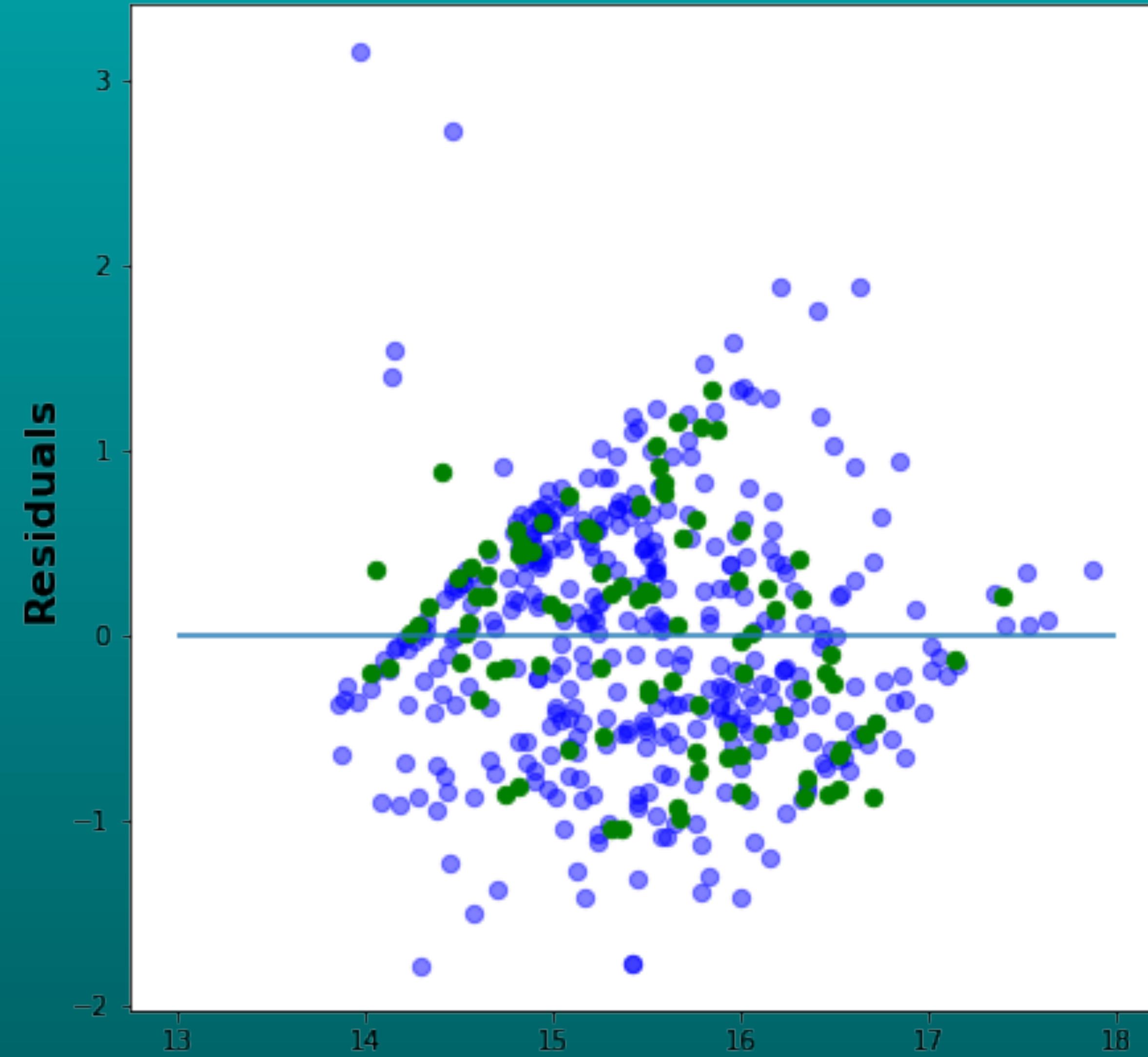
Different Linear Regression Methods

Model	R Squared	MSE
Original	0.693452	0.336988
Lasso	0.685667	0.495522
Ridge	0.698285	0.331675

Model predictions vs. the data points



Residual plot using training (blue) and test (green) data





Importance!



Future Improvements

- Social media impact
- More historial data
- Using different combinations of features

