NBA Salary Predictor

As a metric for evaluation player performance

Mark Biernacki

Data Collection and Cleaning

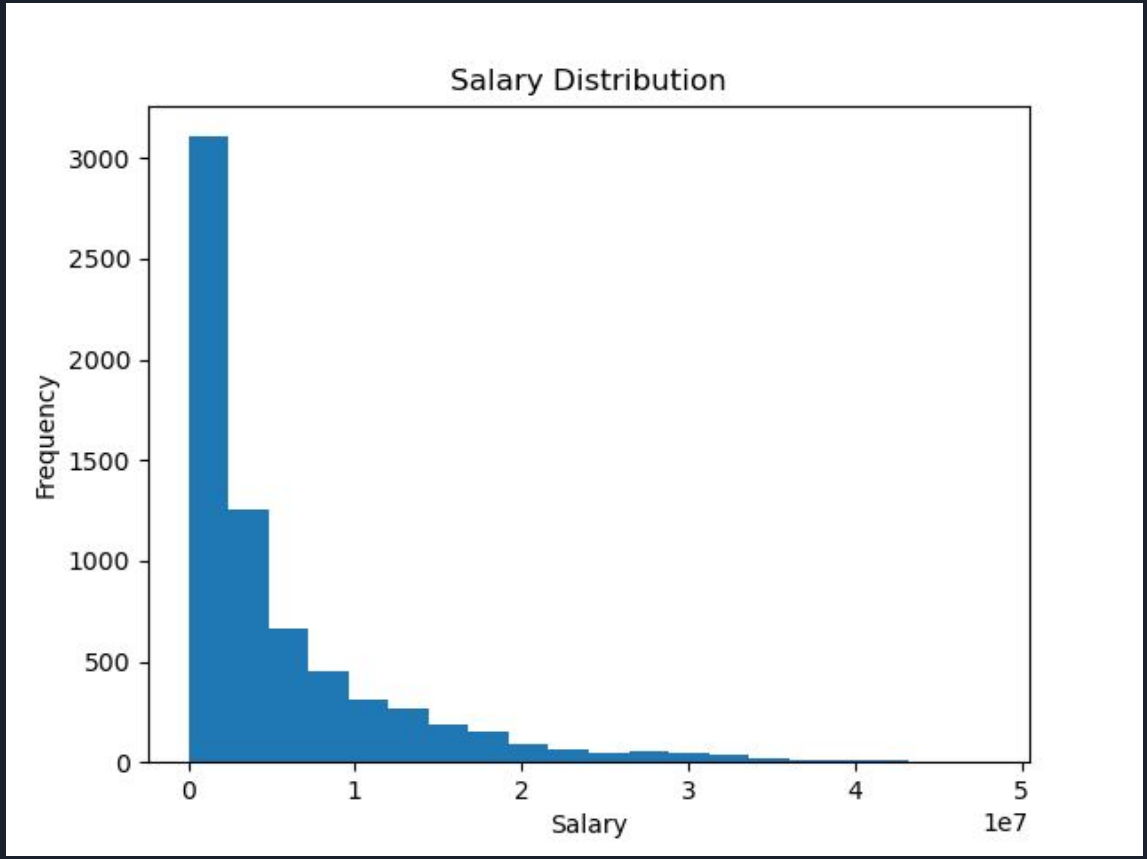
Created 6 functions to scrape data:

- ❑ Per Game Stats, Advanced Stats
- ❑ Salaries
- ❑ Salaries by team
- ❑ Draft Status
- ❑ Salary Cap Maximum

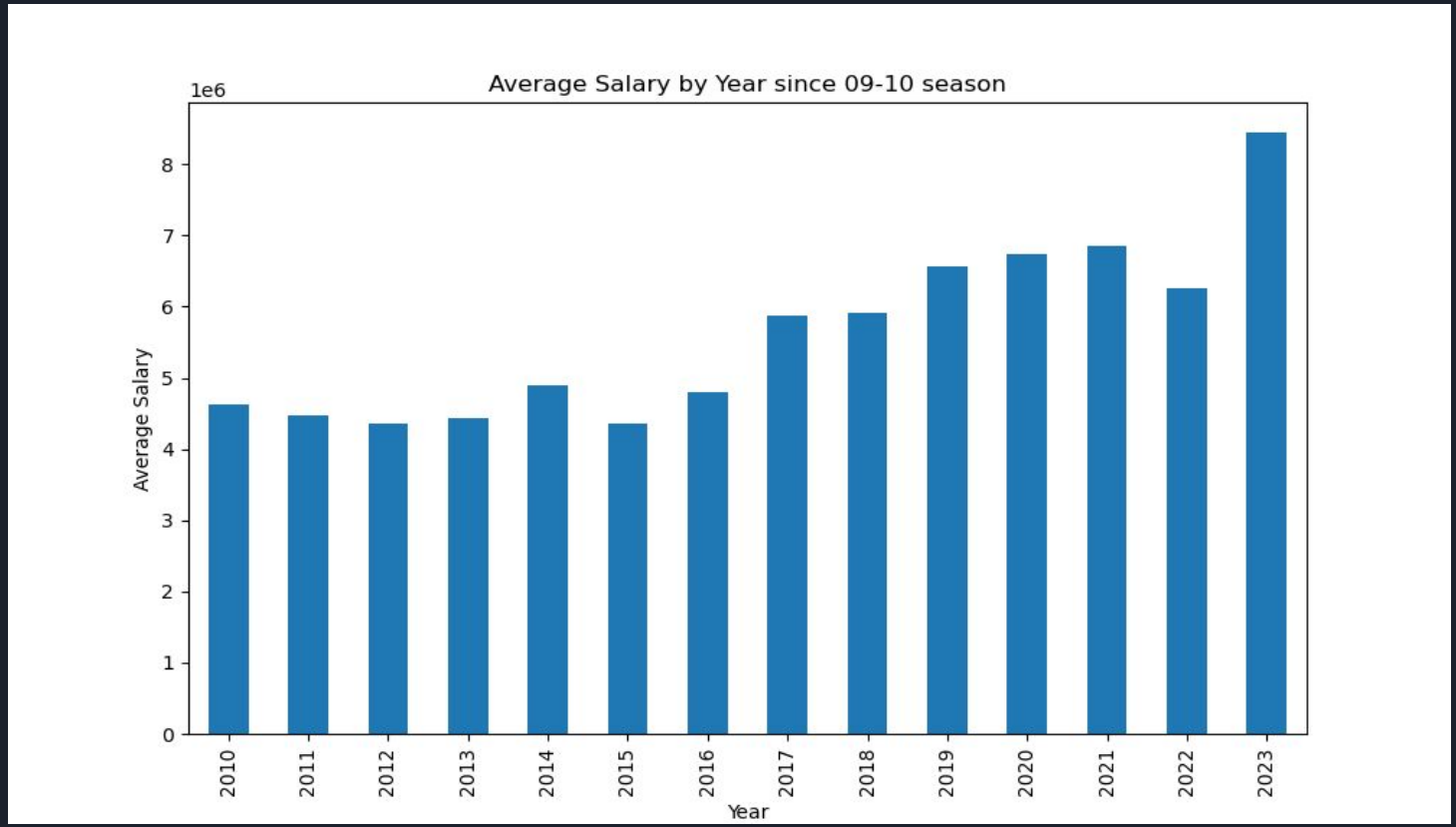
Merged Data!

Data Credit:
Basketball-Reference.com
Spotrac.com

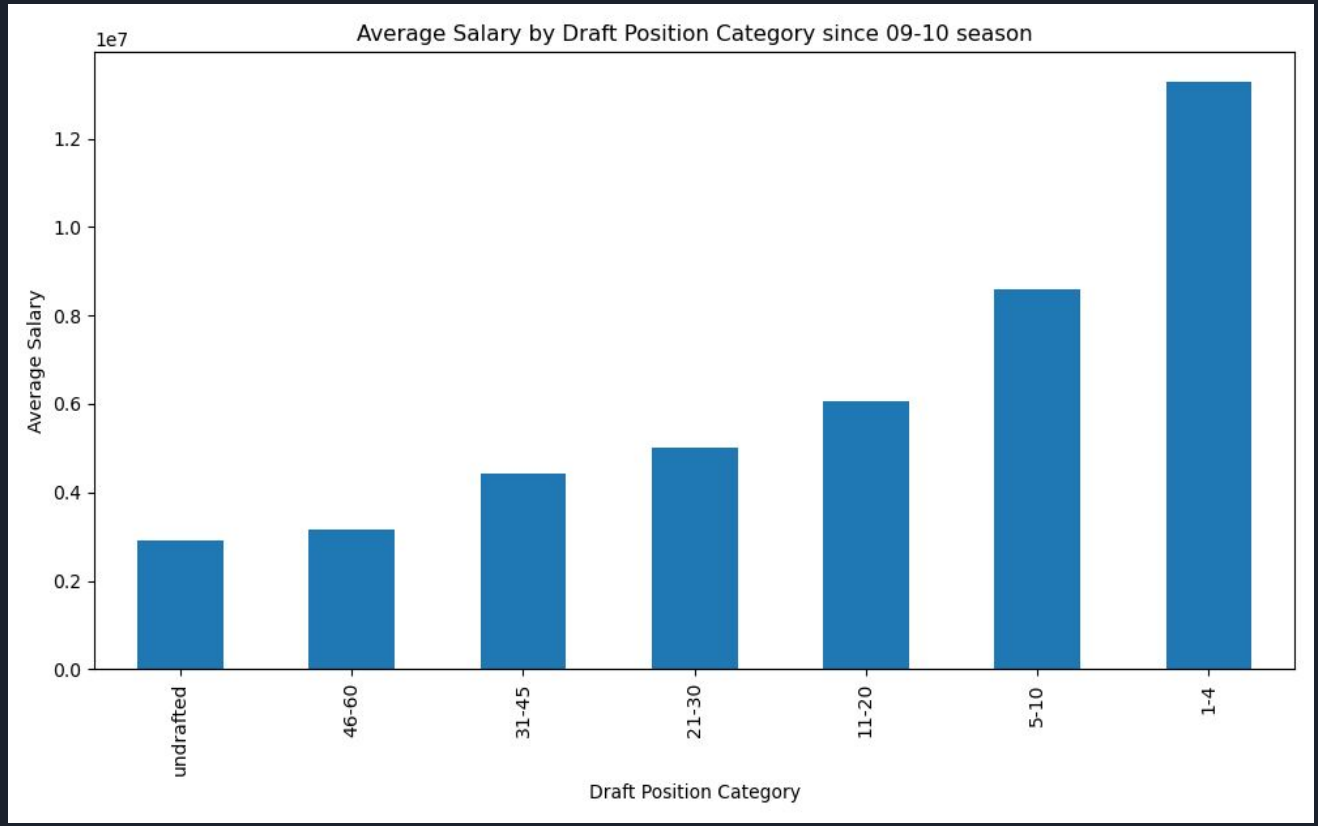
Distribution of Salary Skewed!



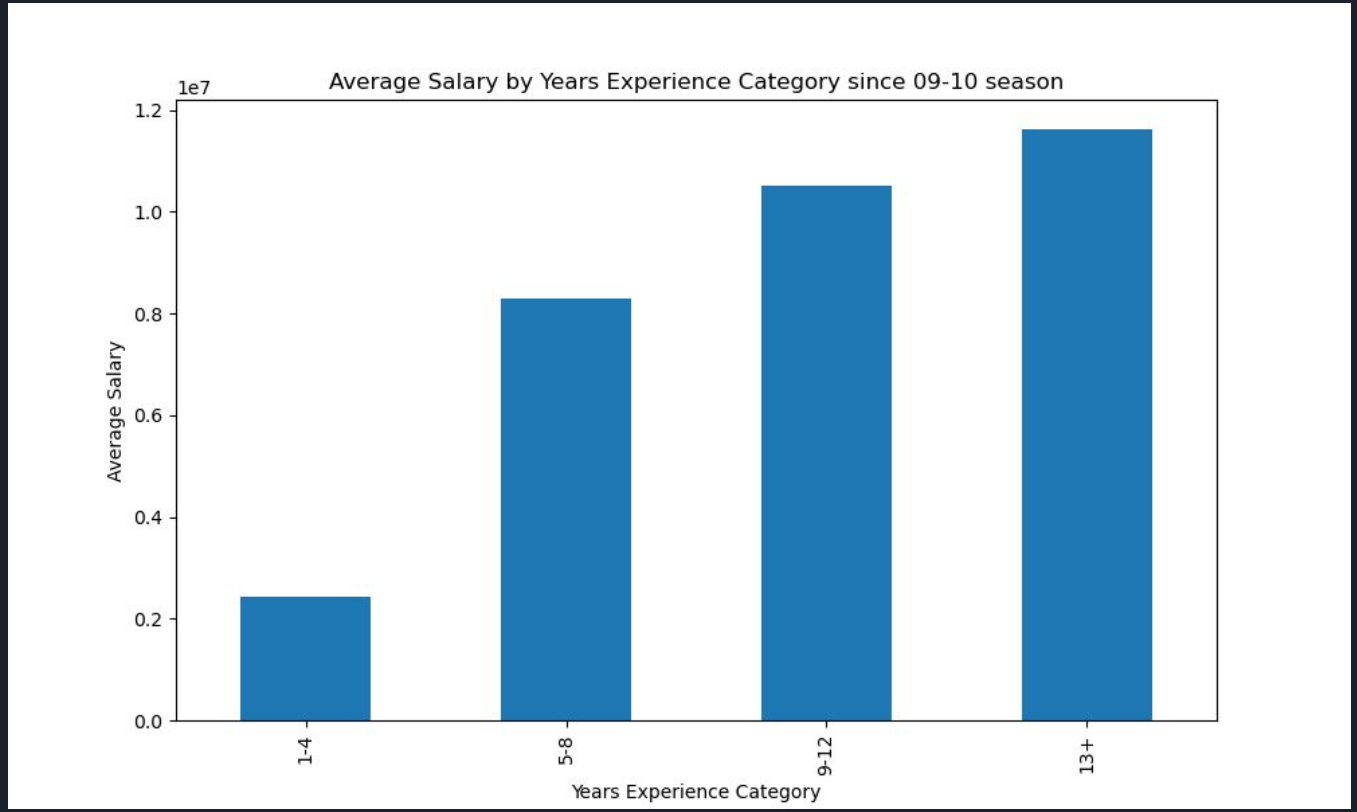
Average Salary By Year



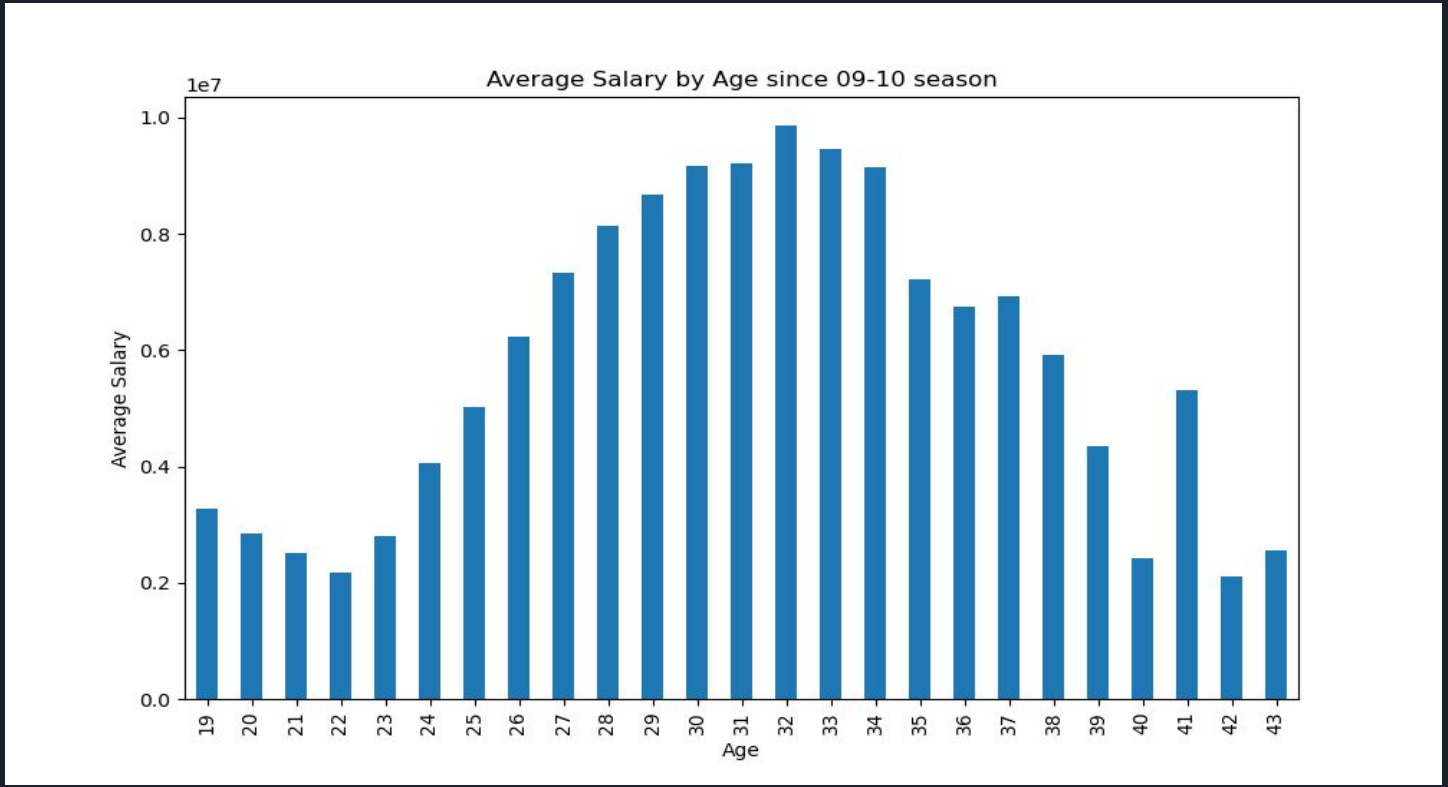
Engineered Features



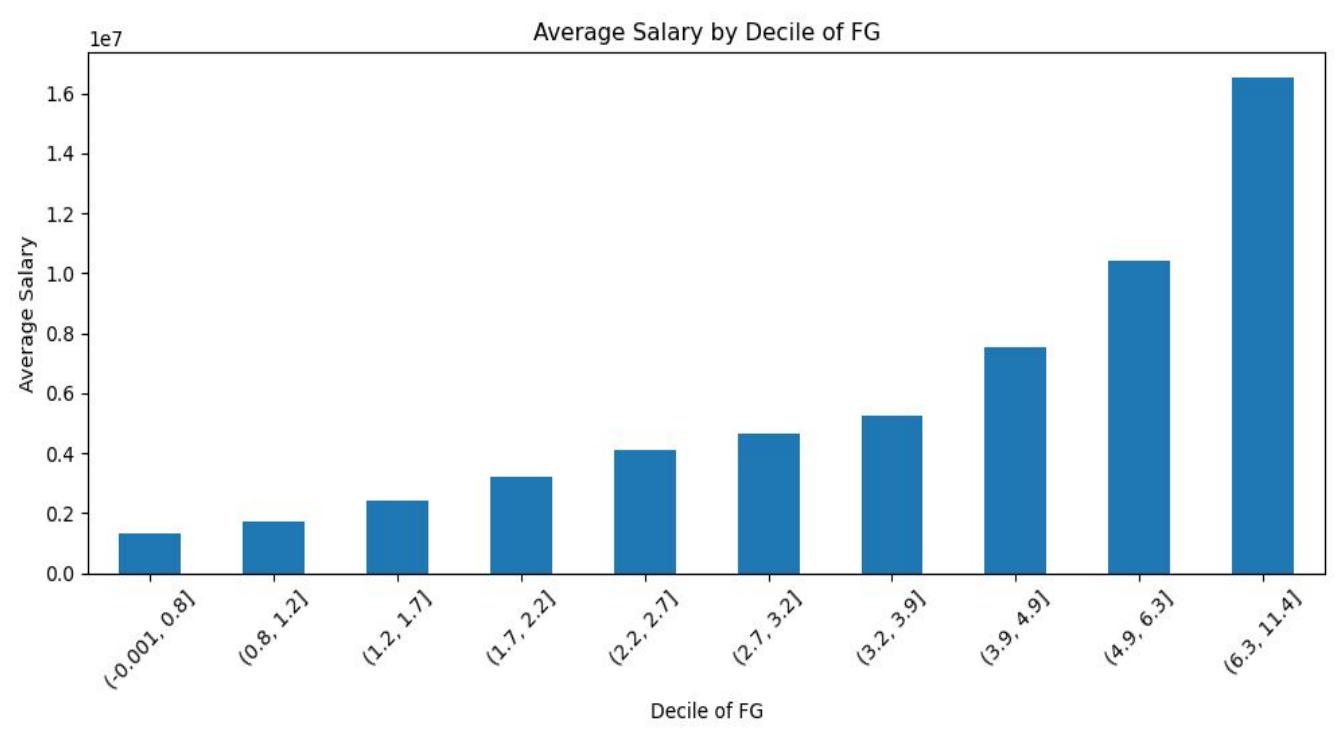
Engineered Features



Average Salary By Age (not linear)

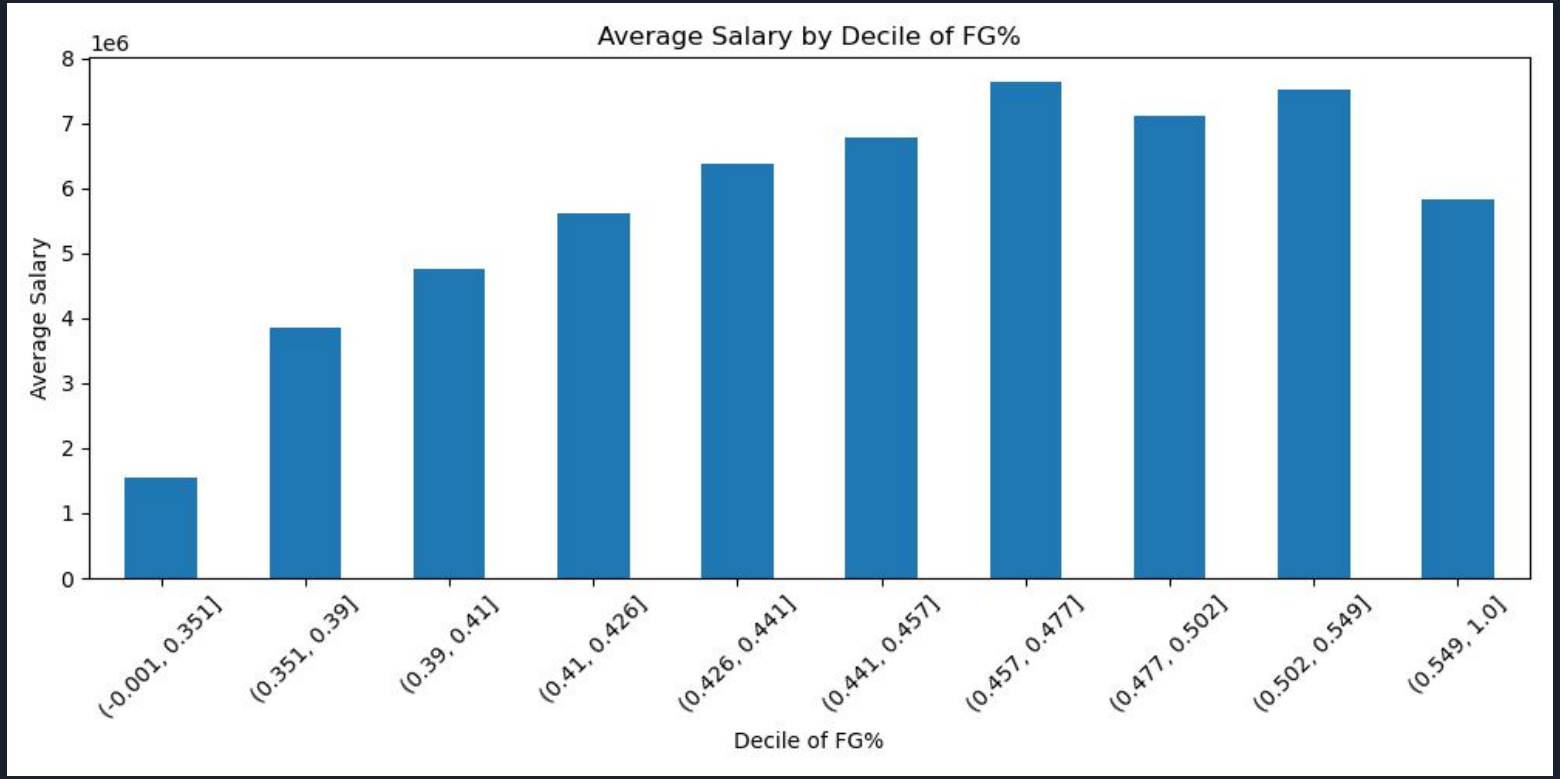


Average Salary by FGM
Grouped by Percentile

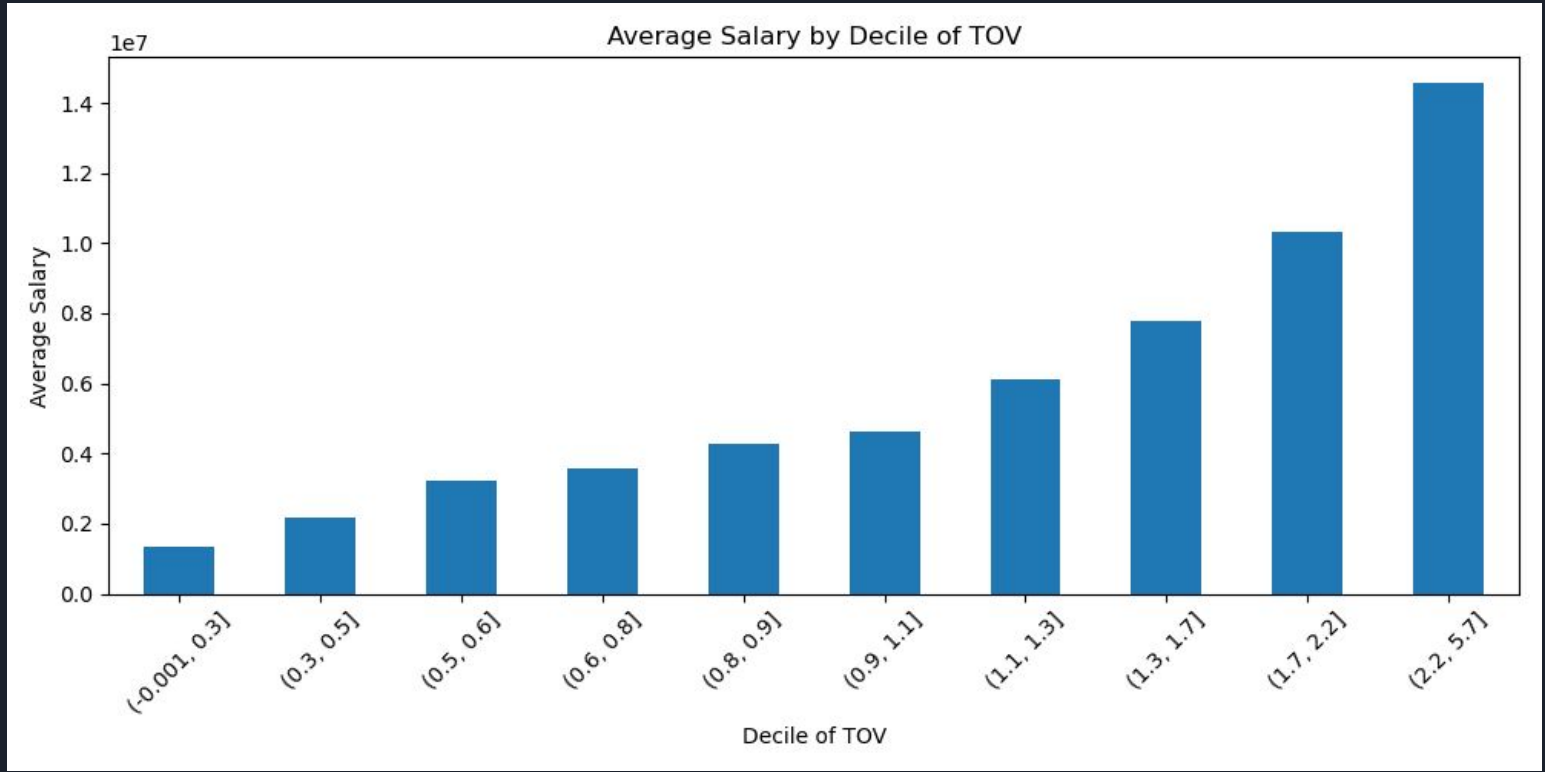


Average Salary by FG%

Notice that last Bar? Not linear

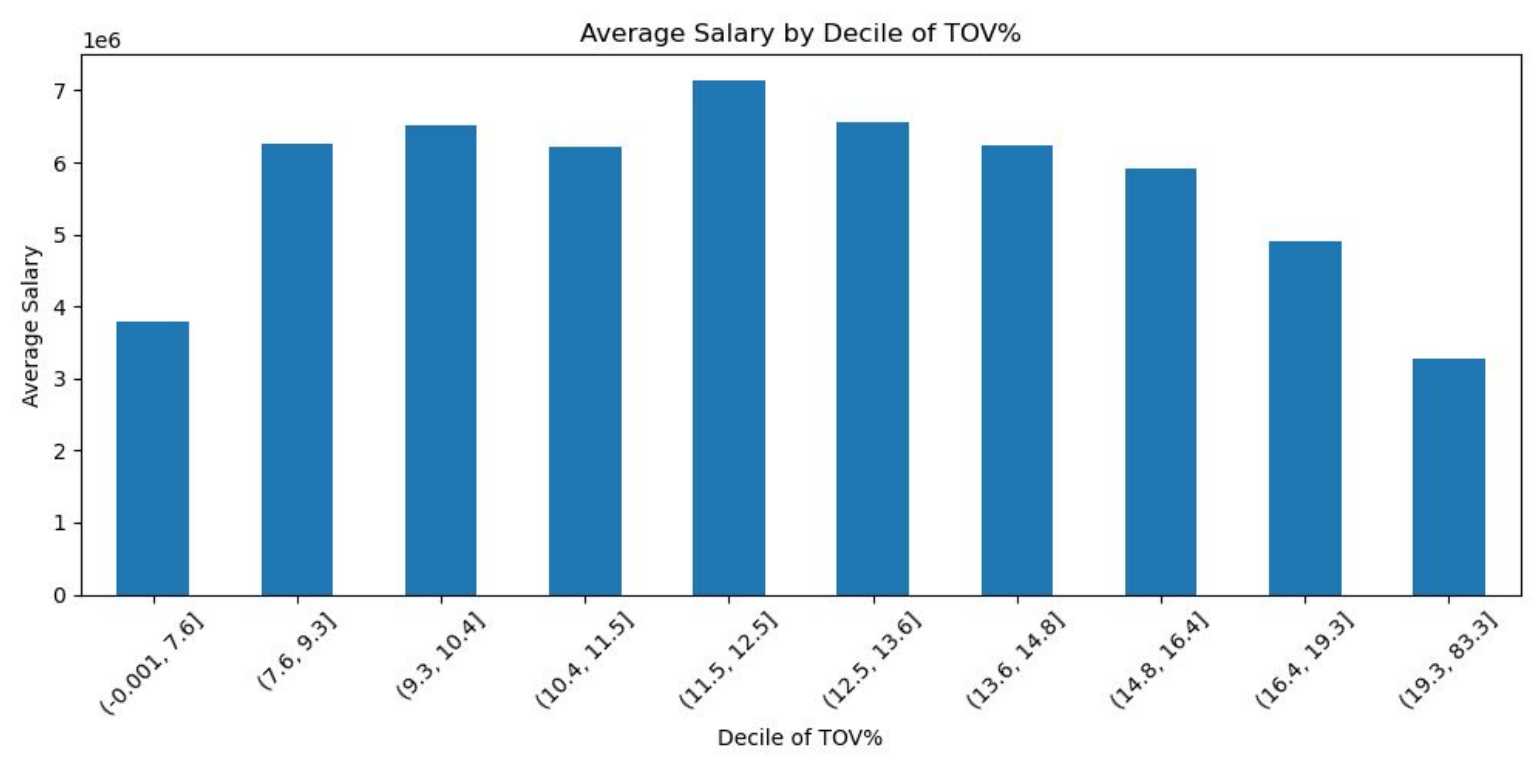


Average Salary by Turnovers (counterintuitive)



Average Salary By Turnover %

(advanced stats are important)

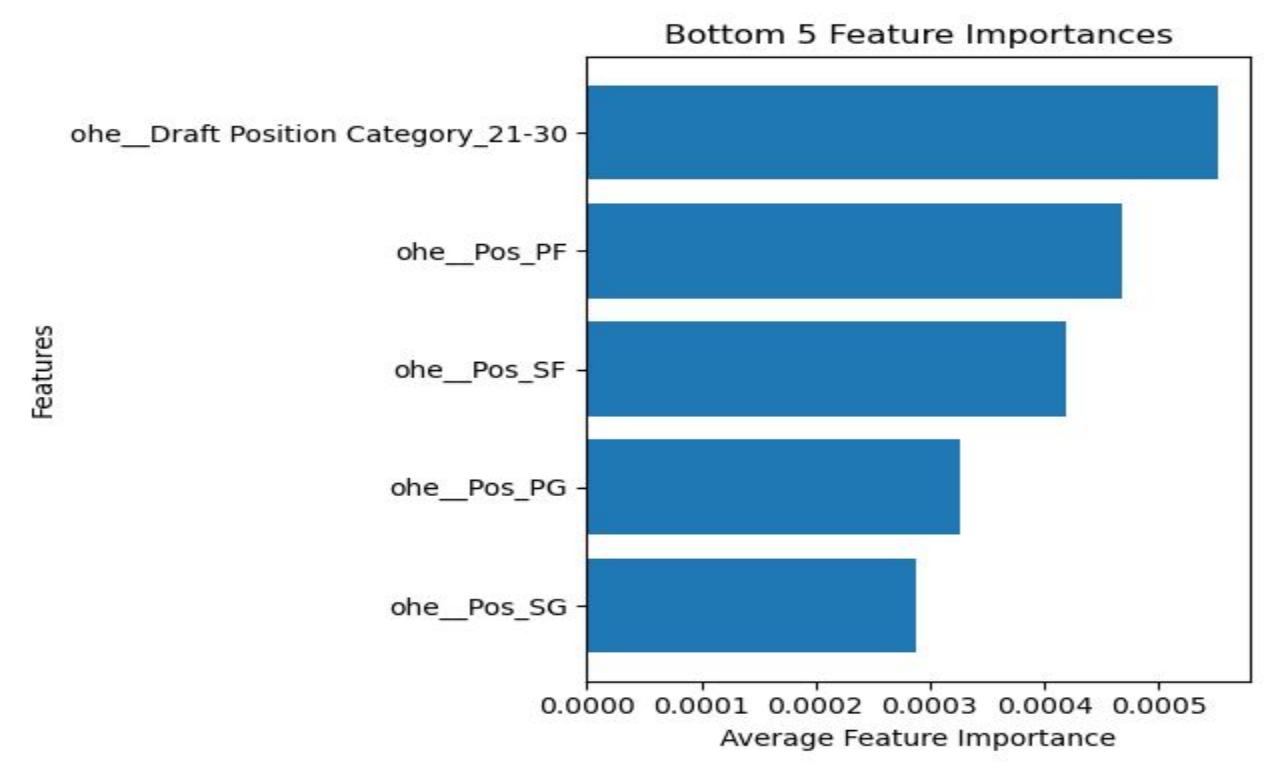


Modeling

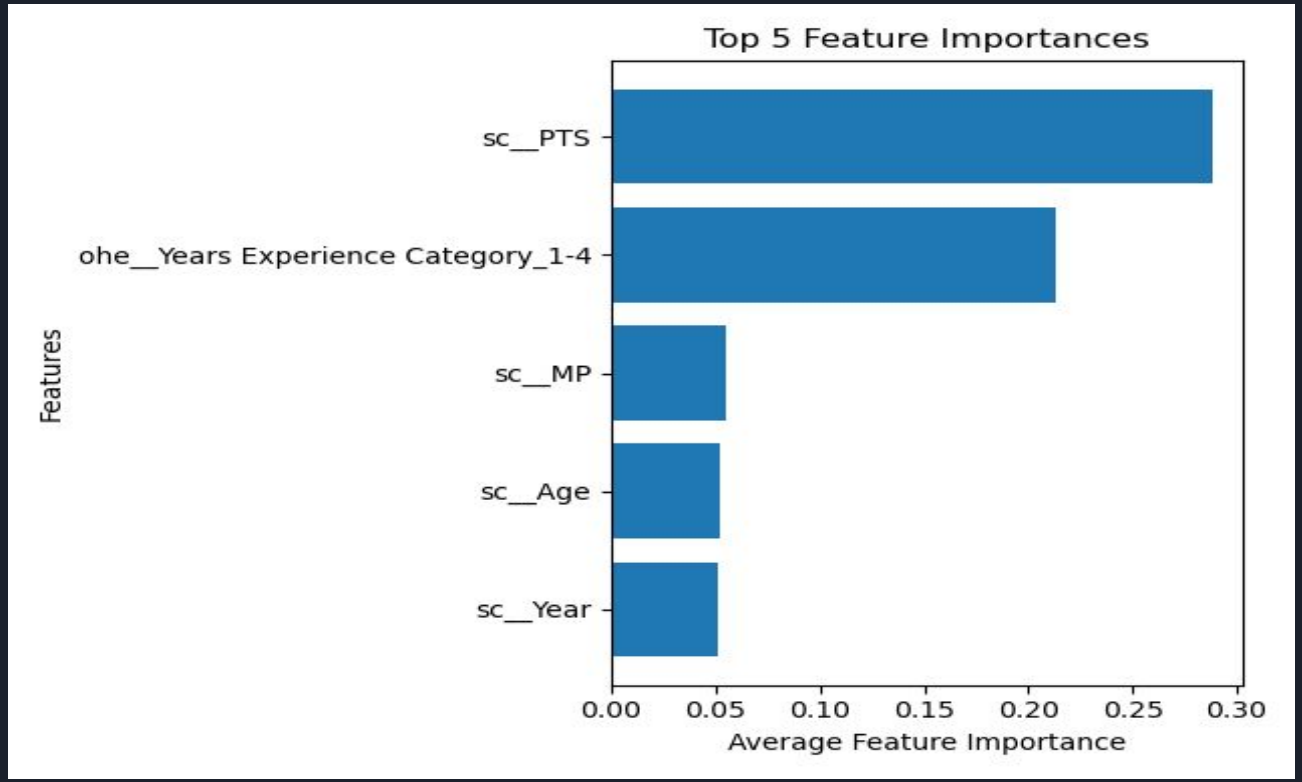
- Because relationships were non-linear I figured i would need an advanced ensemble learning model.
- Experimented with Random Forest, Logistic, Gradient Boosting algorithms
- The best model was a stacked model using Random Forest and Gradient Boosting algorithms
- R^2 score of .77

Stacked Model

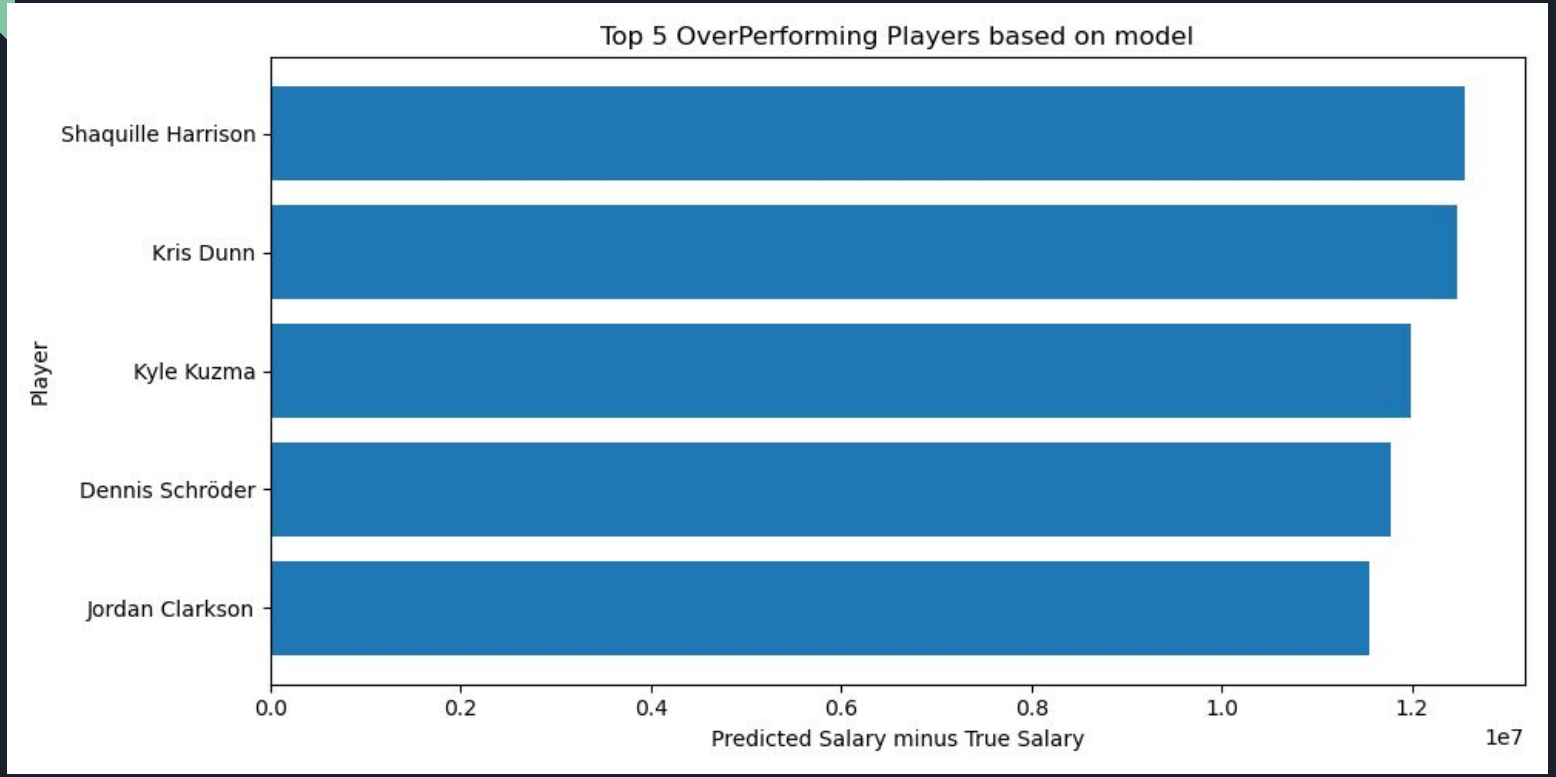
Average importance?



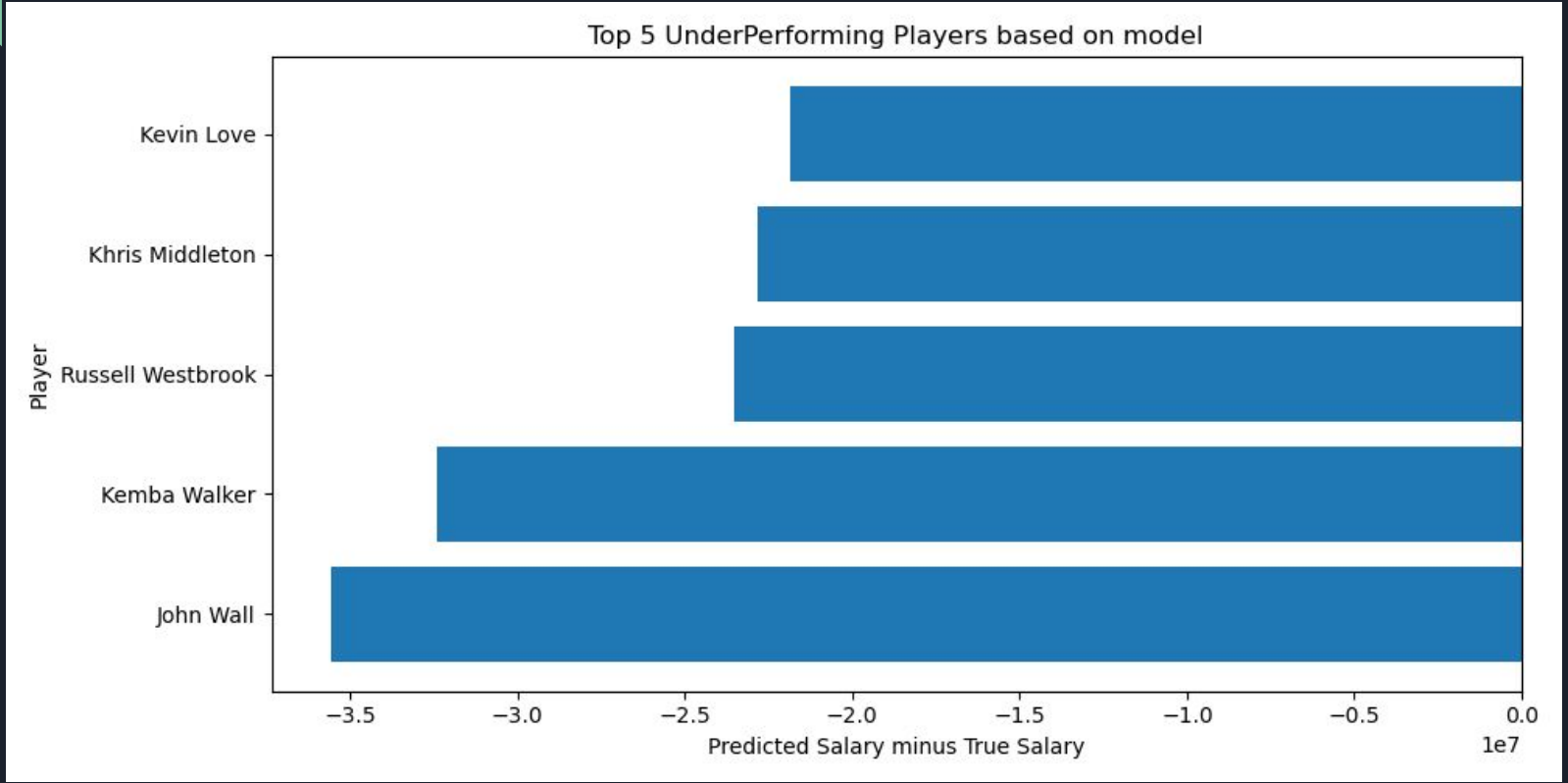
Stacked Model



Diamonds in the Rough



Players who are past their prime
(and still being paid like they're in their prime)



Time to go to the Streamlit App!

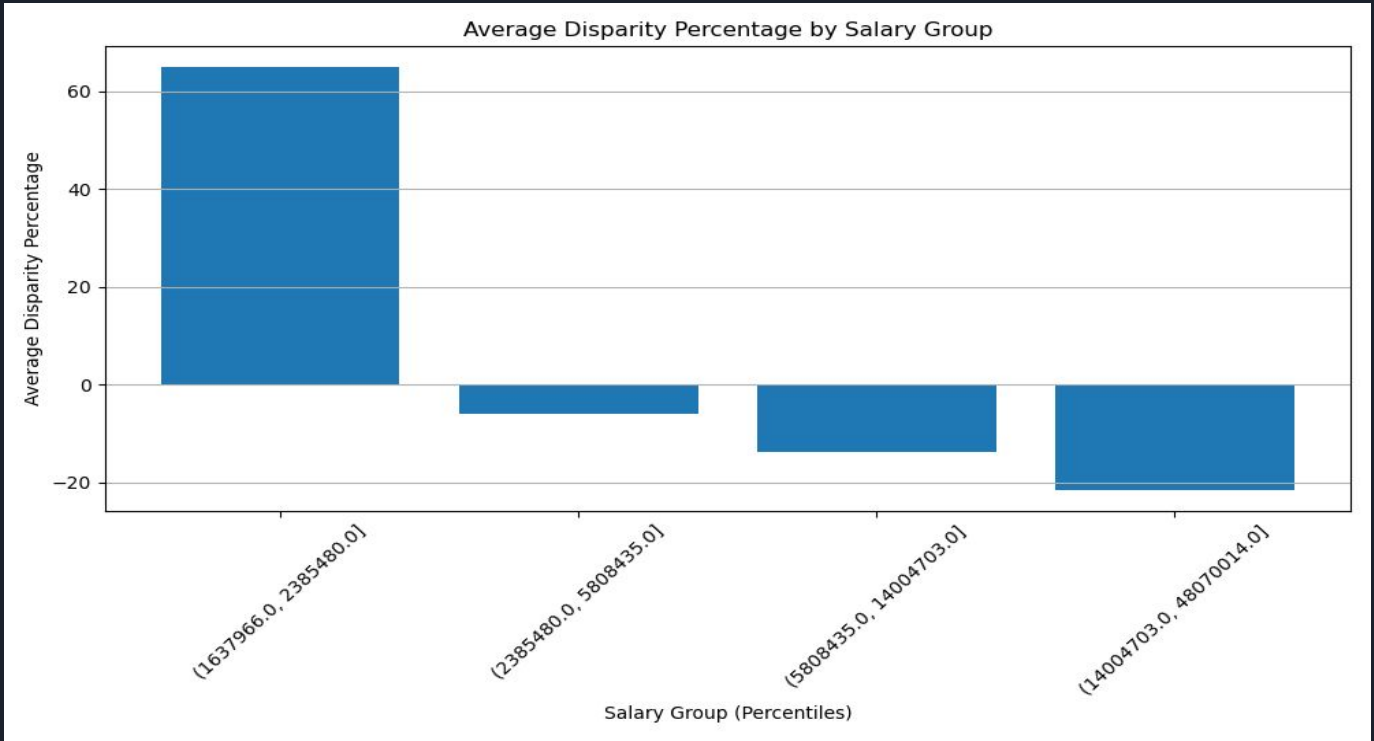
Room for improvement

- More stats could be added, like per 36 minutes stats.
- Could cross validate salary data with another source.
- Experimentation with neural networks
- Incorporating playoff stats

Conclusion

-This could be a valuable tool for NBA player agents and General Managers looking to negotiate. It is also cool for NBA nerds!

The model is biased towards overpredicting lower salaries and underpredicting higher salaries.



Since the model is based only on stats…

This is either a limitation of the model, or a misuse of financial resources across the NBA.