NBA Salary Predictor

As a metric for evaluation player performance

Mark Biernacki

Data Collection

Created 6 functions to scrape data:

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

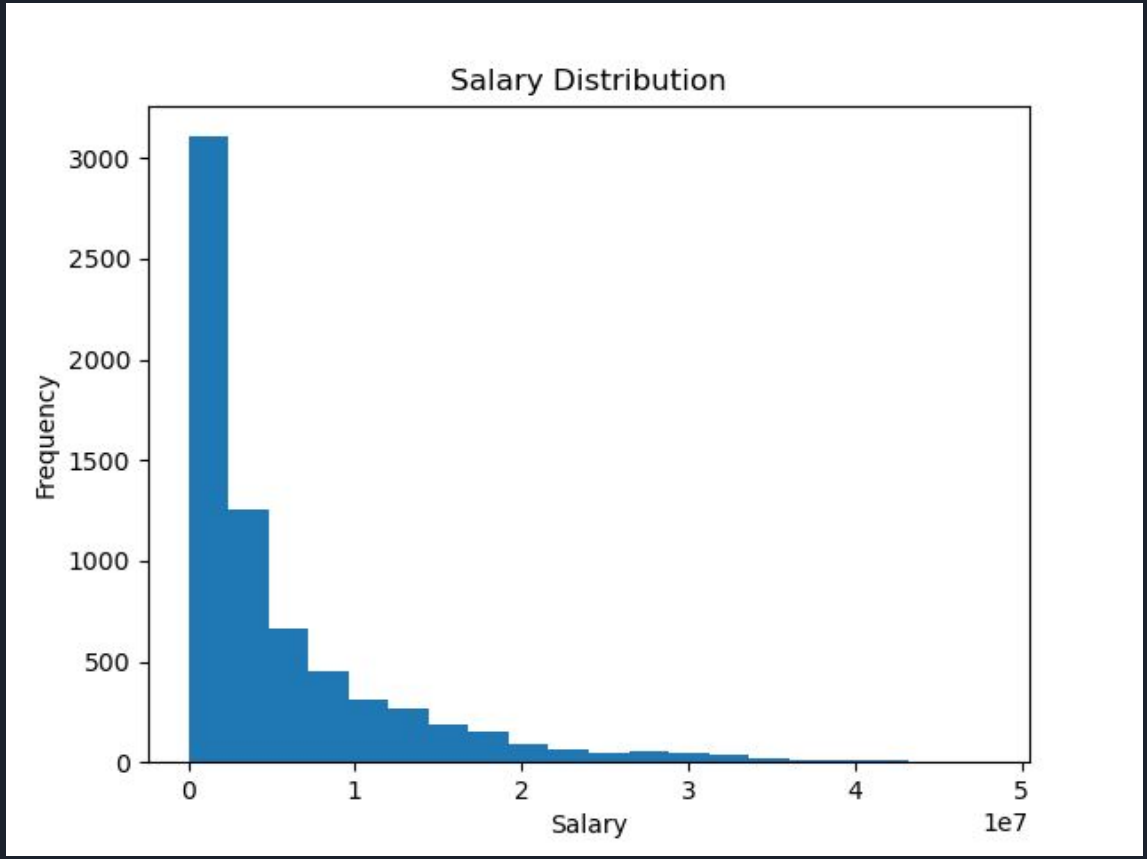
Data Credit:
Basketball-Reference.com
Spotrac.com

Data Cleaning and Feature Engineering

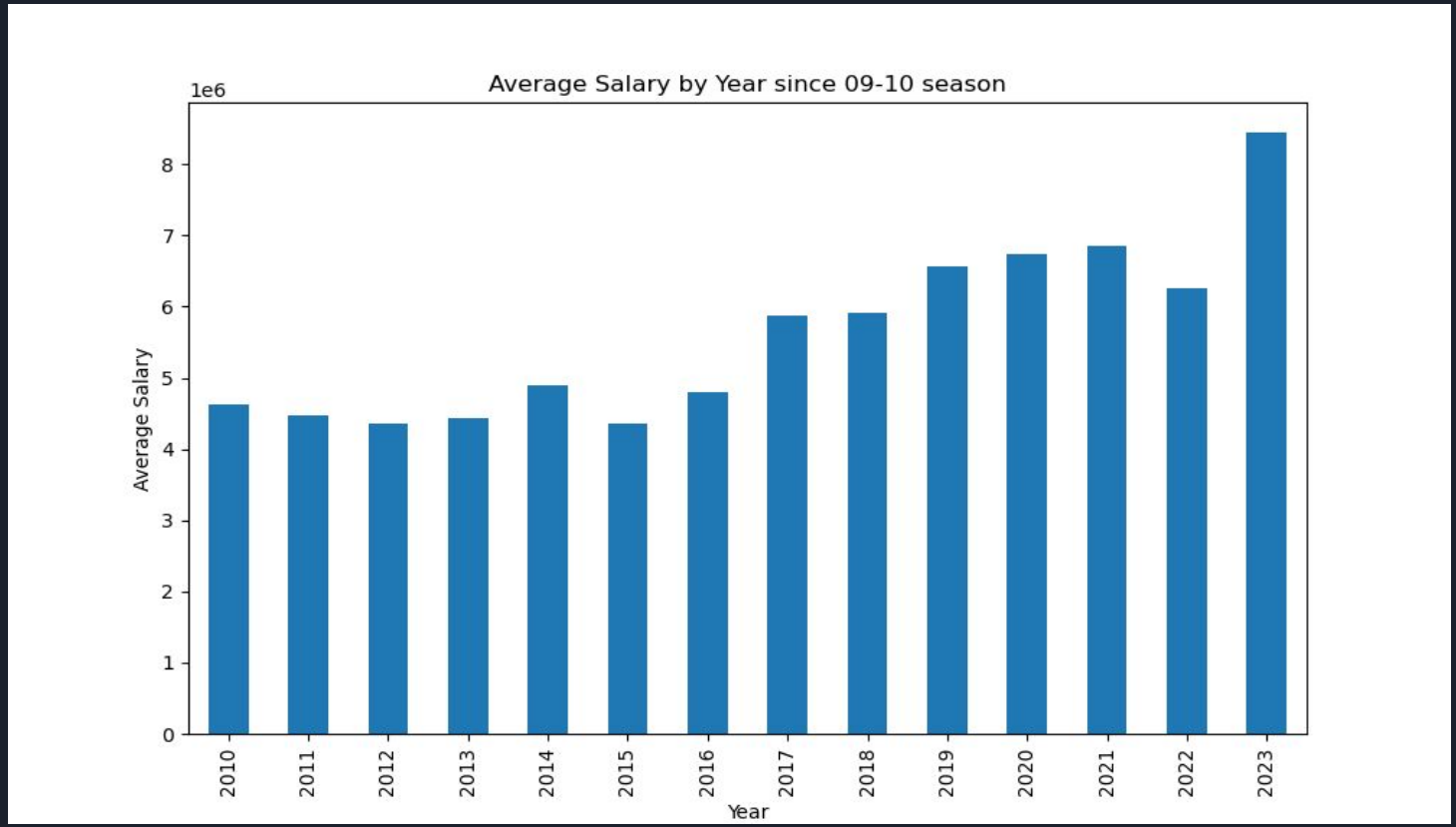
- Merged Dataframes on player link
- Converted columns to correct data types
- Imputed null values with the mean of that position:

Pos	
C	0.212006
PF	0.276355
PG	0.323208
SF	0.321930
SG	0.329178

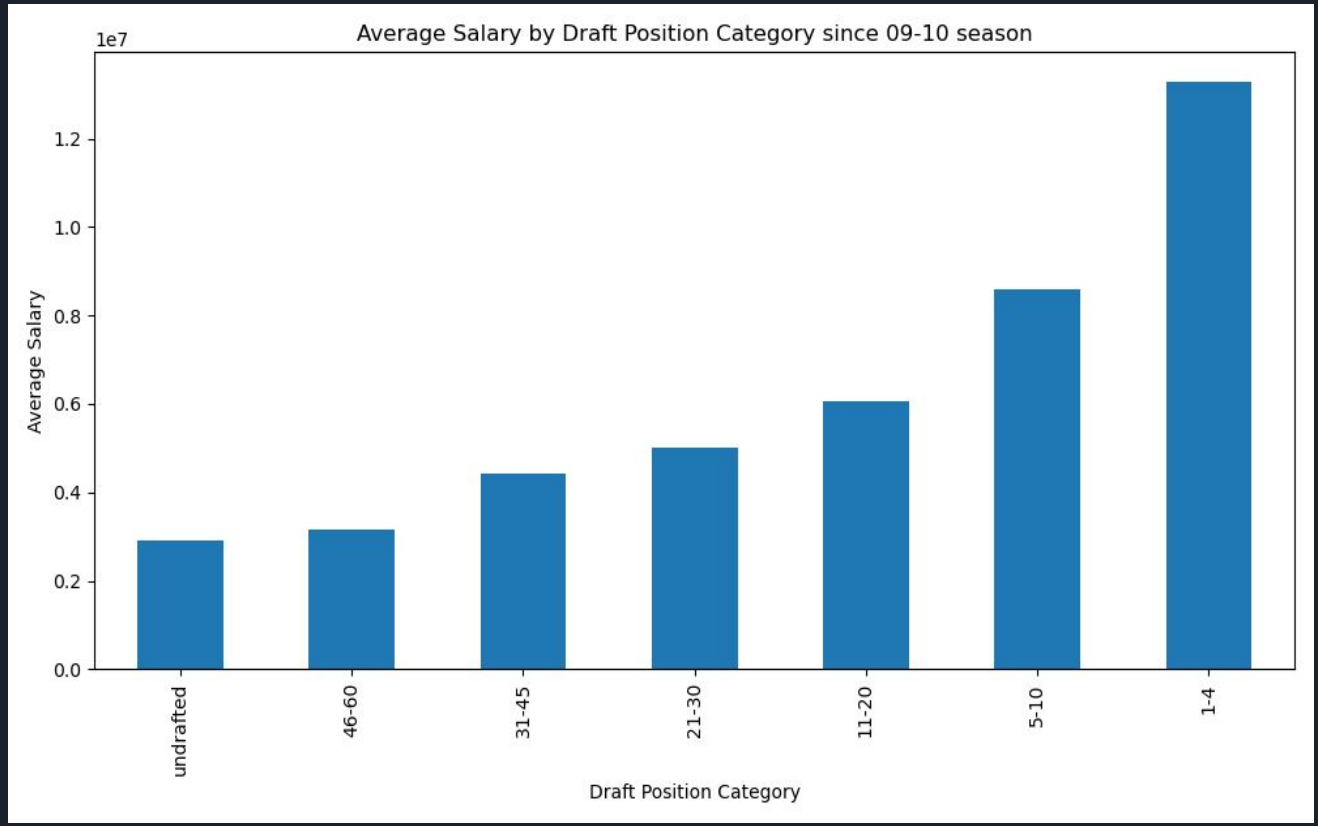
Distribution of Salary



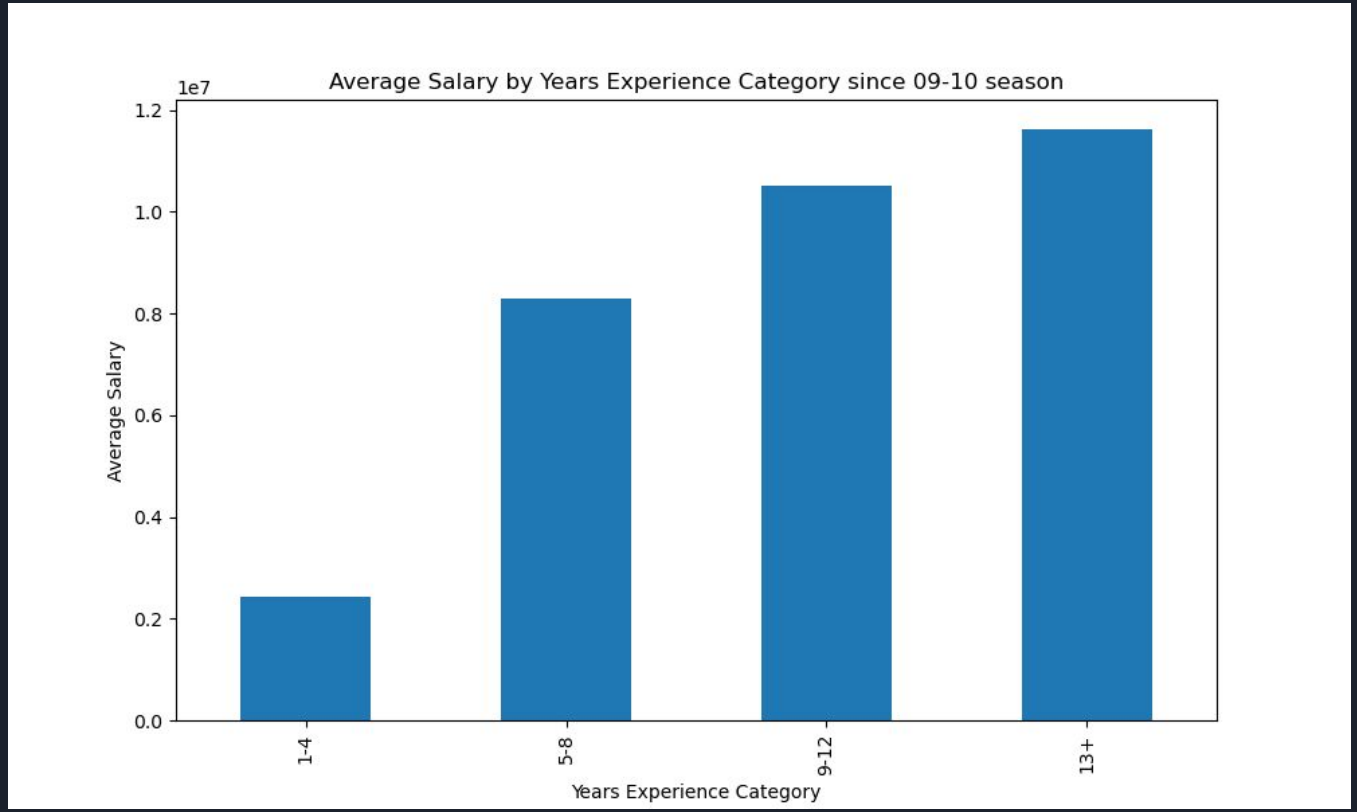
Average Salary By Year



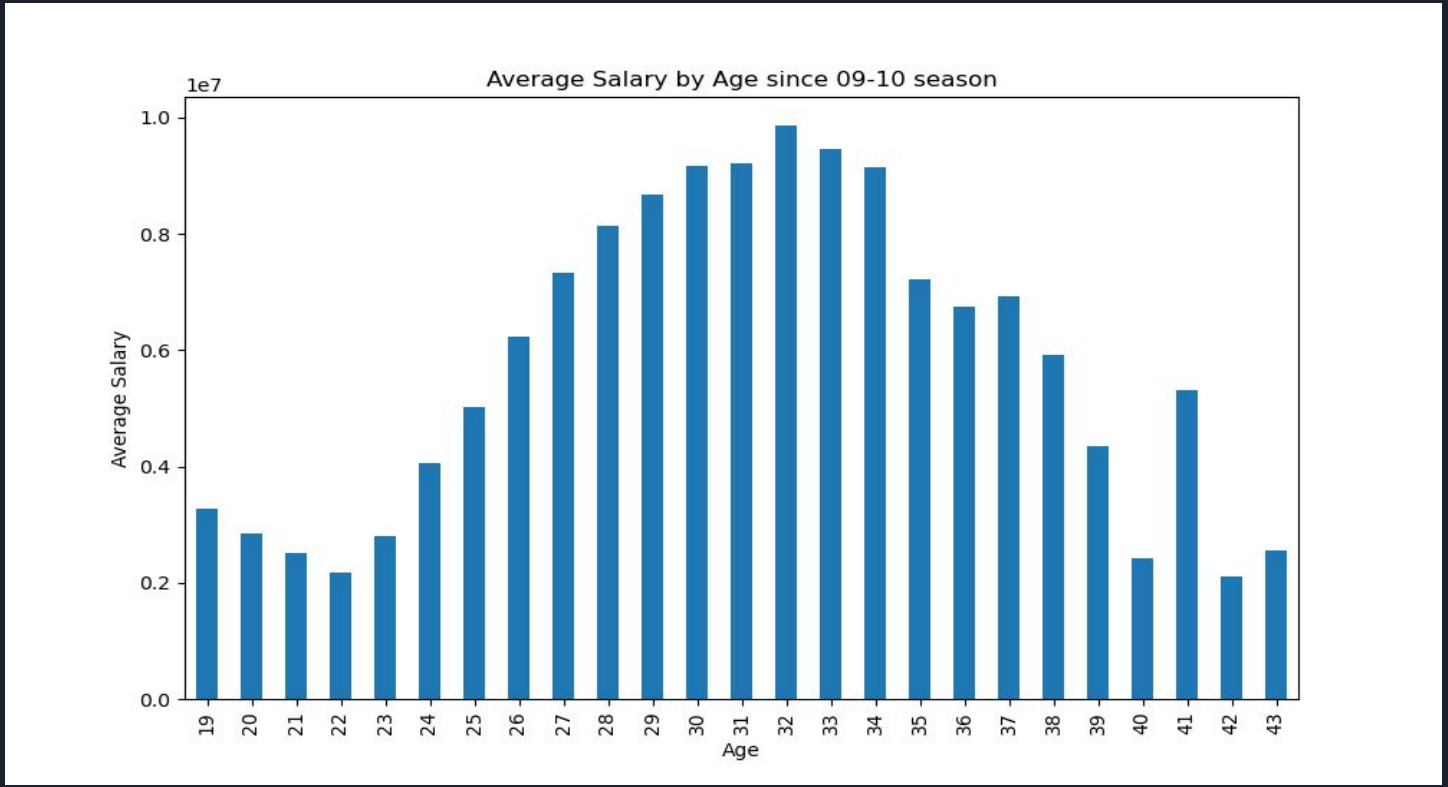
Engineered Features



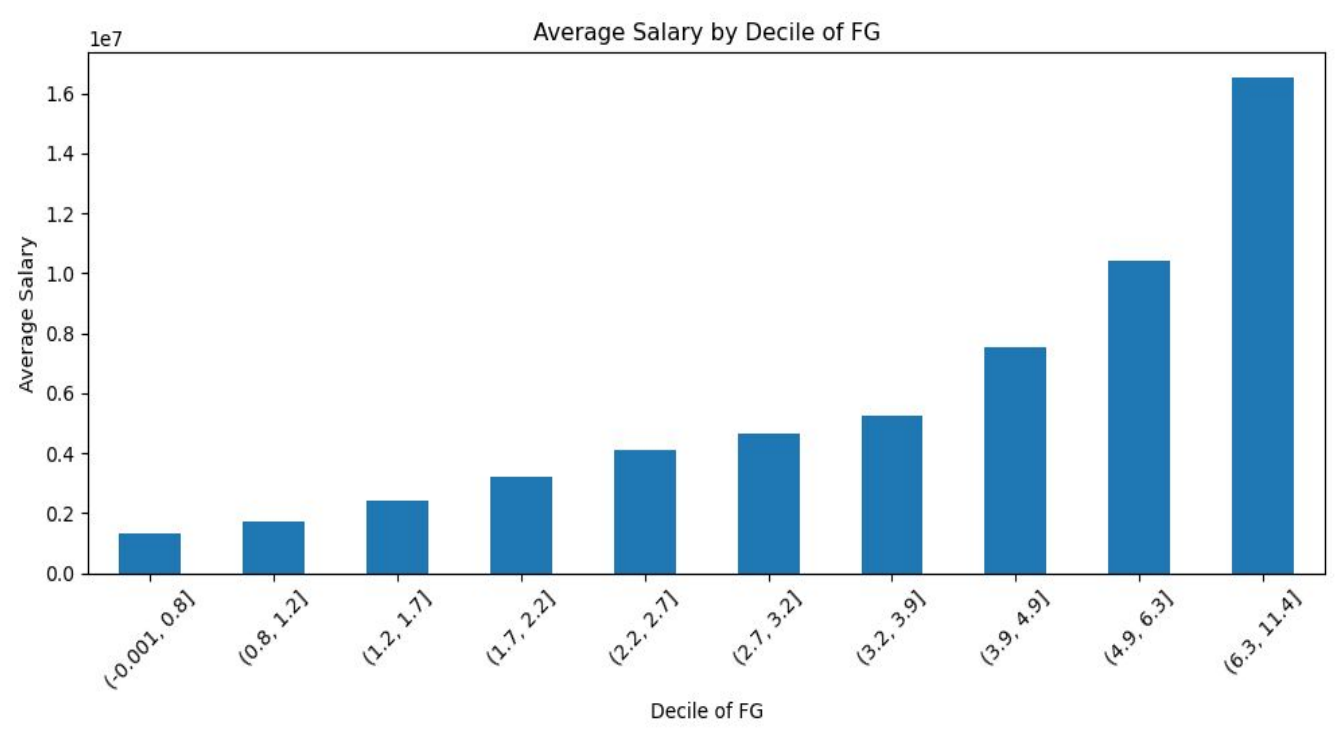
Engineered Features



Average Salary By Age (not linear)

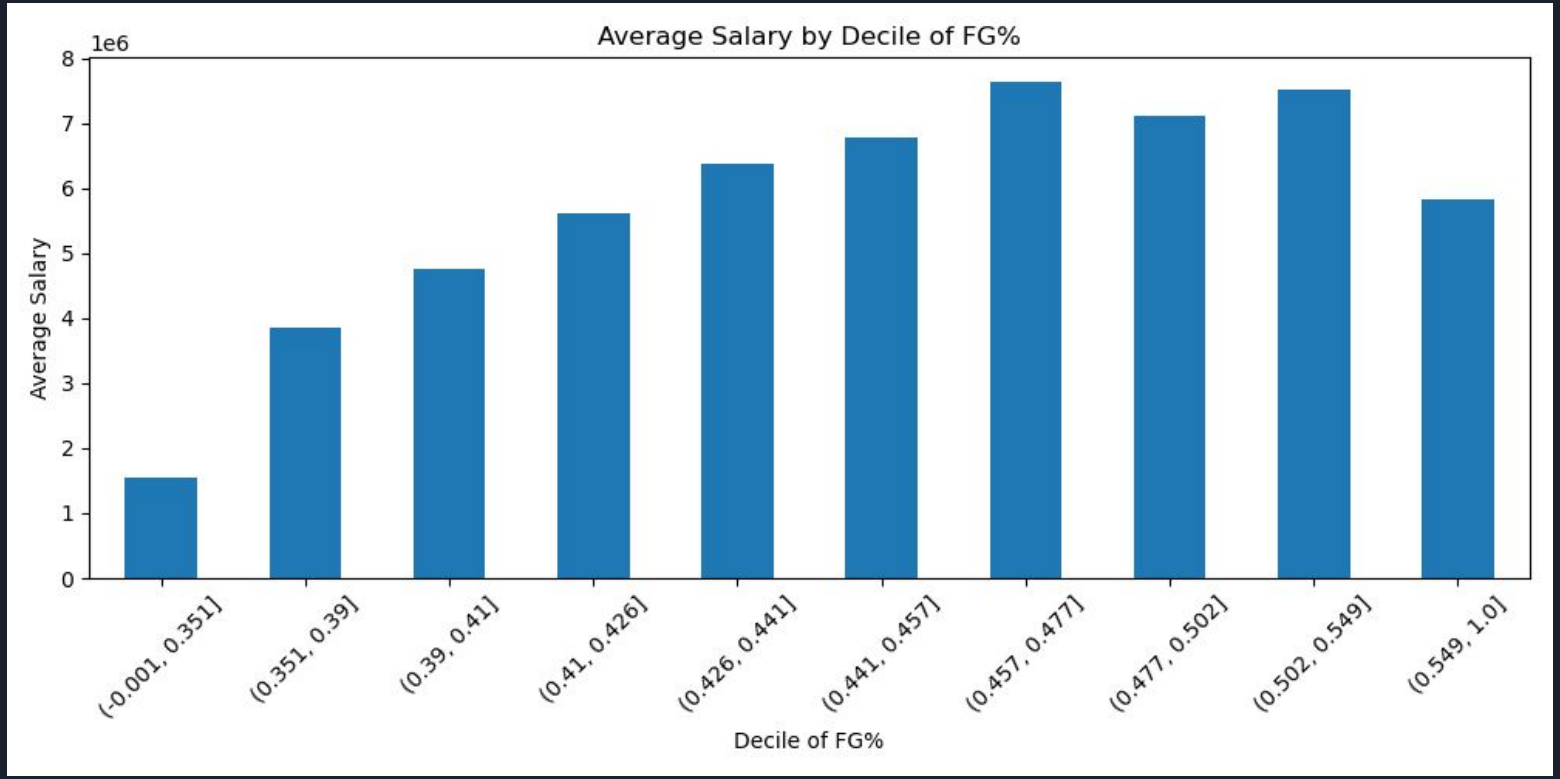


Average Salary by FGM
Grouped by Percentile

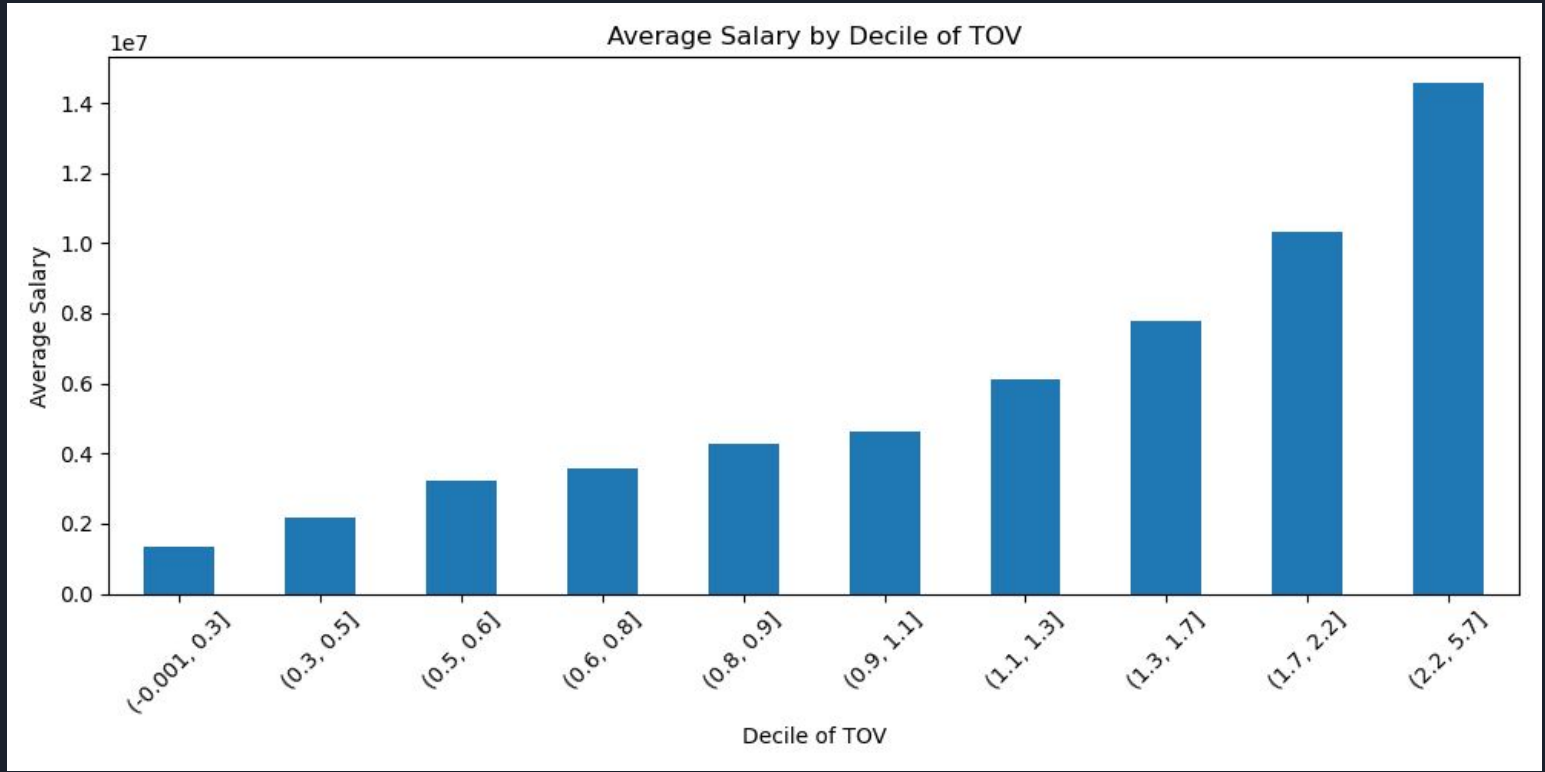


Average Salary by FG%

Notice that last Bar? Not linear

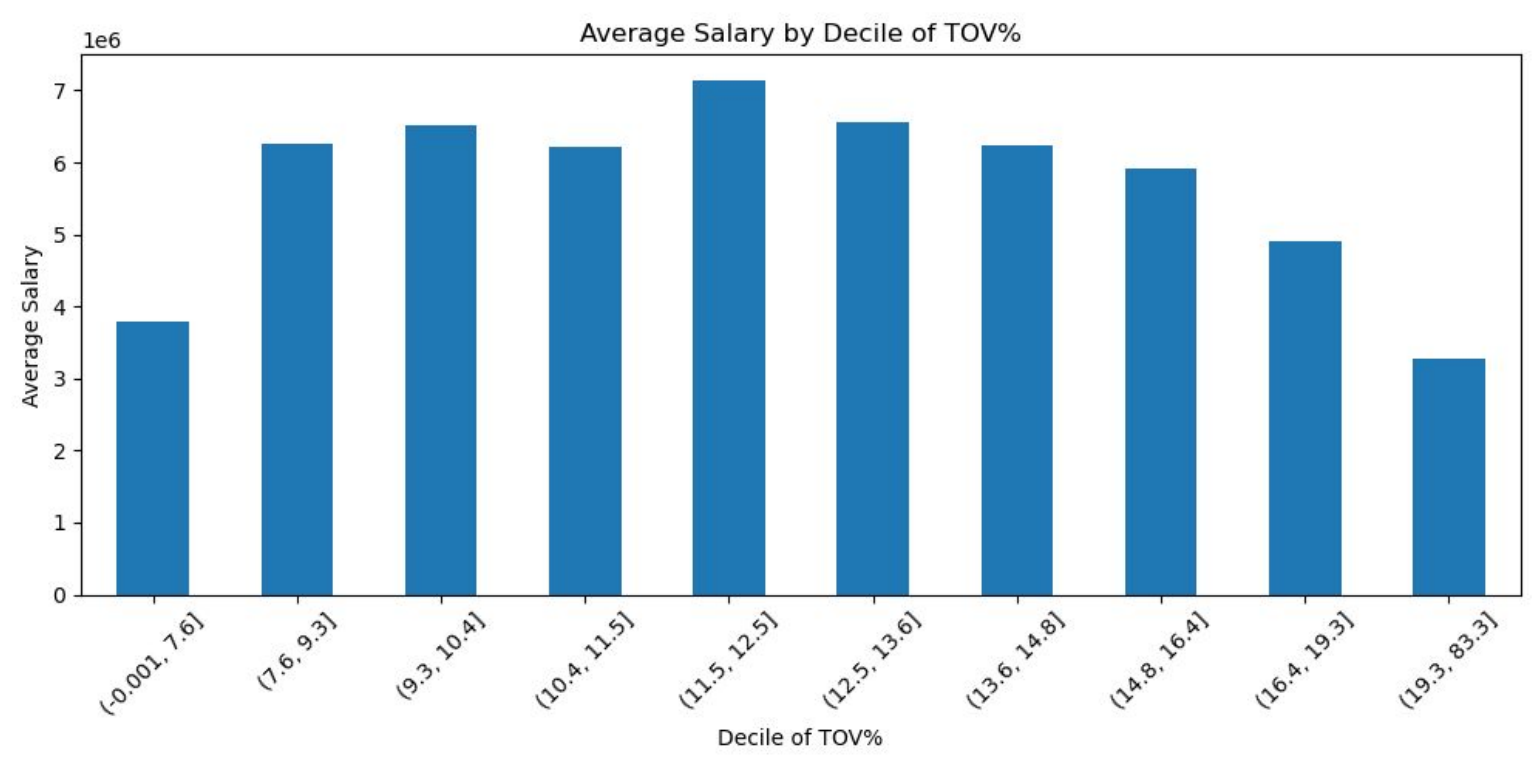


Average Salary by Turnovers (counterintuitive)



Average Salary By Turnover %

(advanced stats are important)

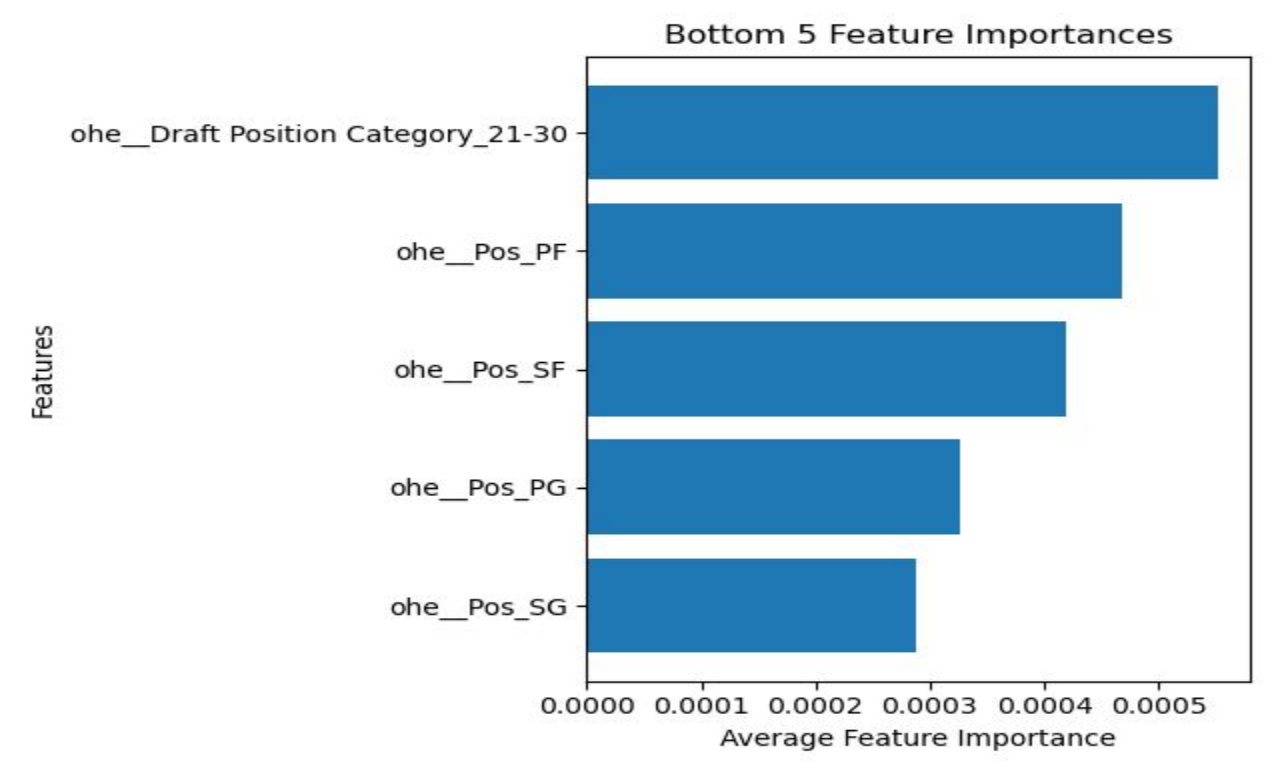


Modeling

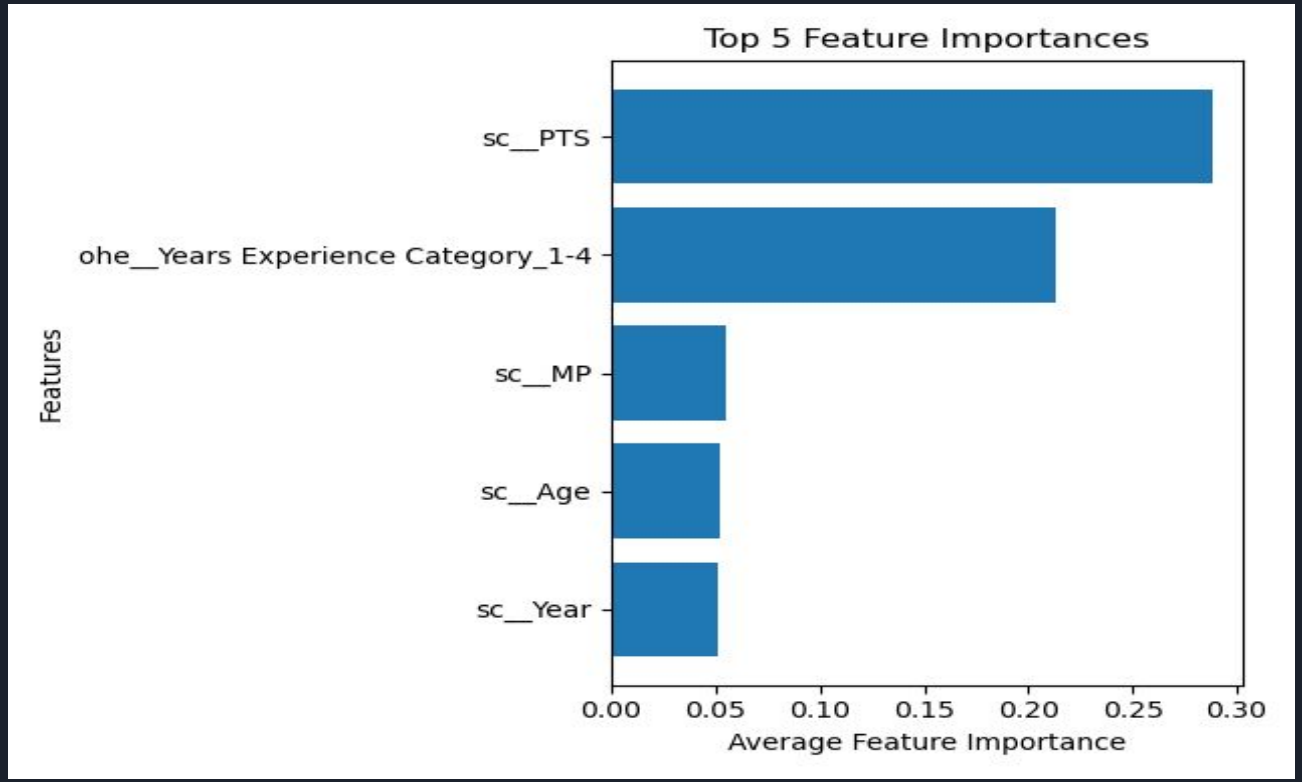
- Because relationships were nonlinear I figured i would need an advanced ensemble learning model.
- Experimented with Random Forest,, Gradient Boosting and stacking algorithms
- The best model was a stacked model using Random Forest and Gradient Boosting algorithms
- R^2 score of .77

Stacked Model

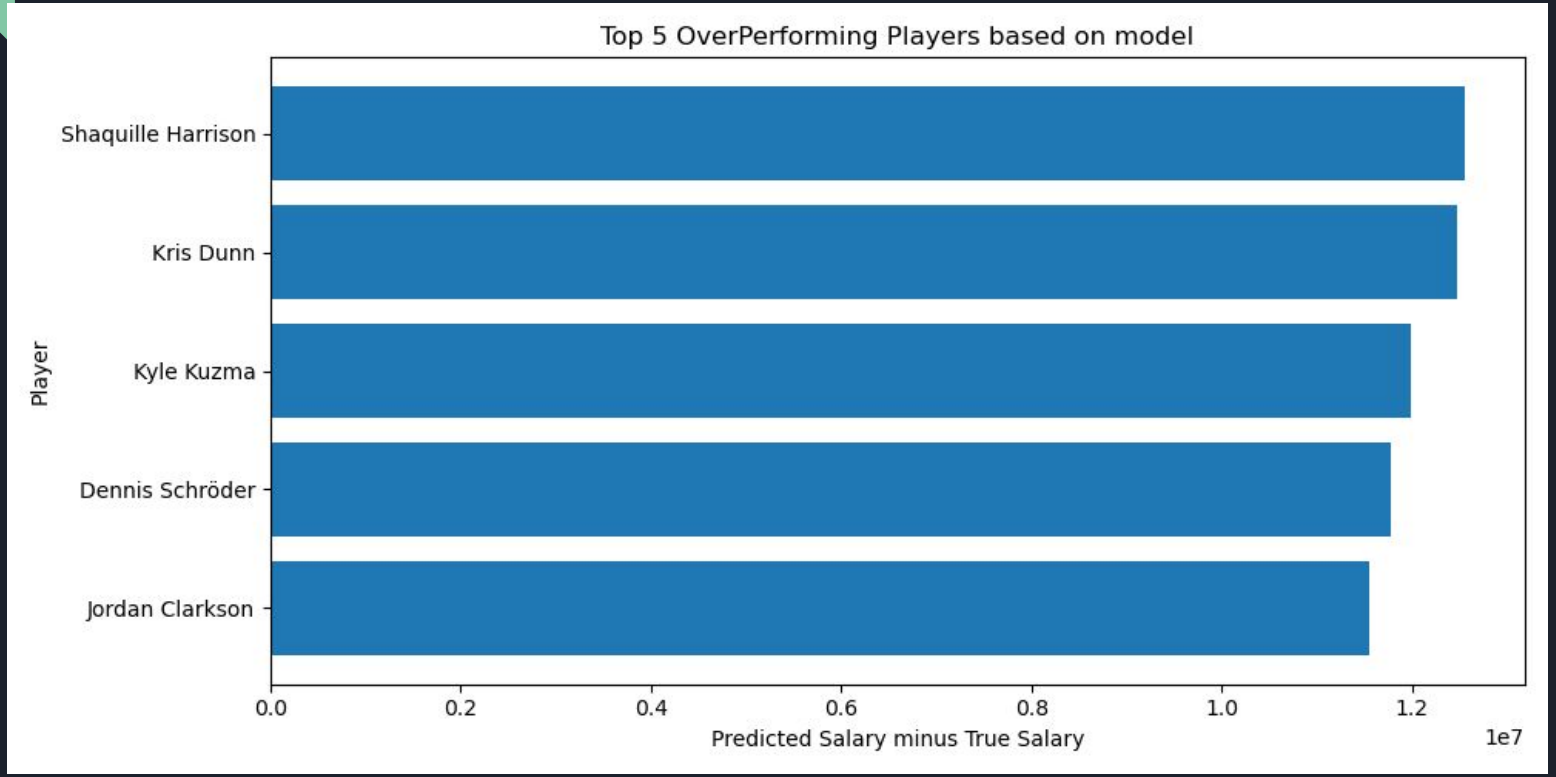
Average importance?



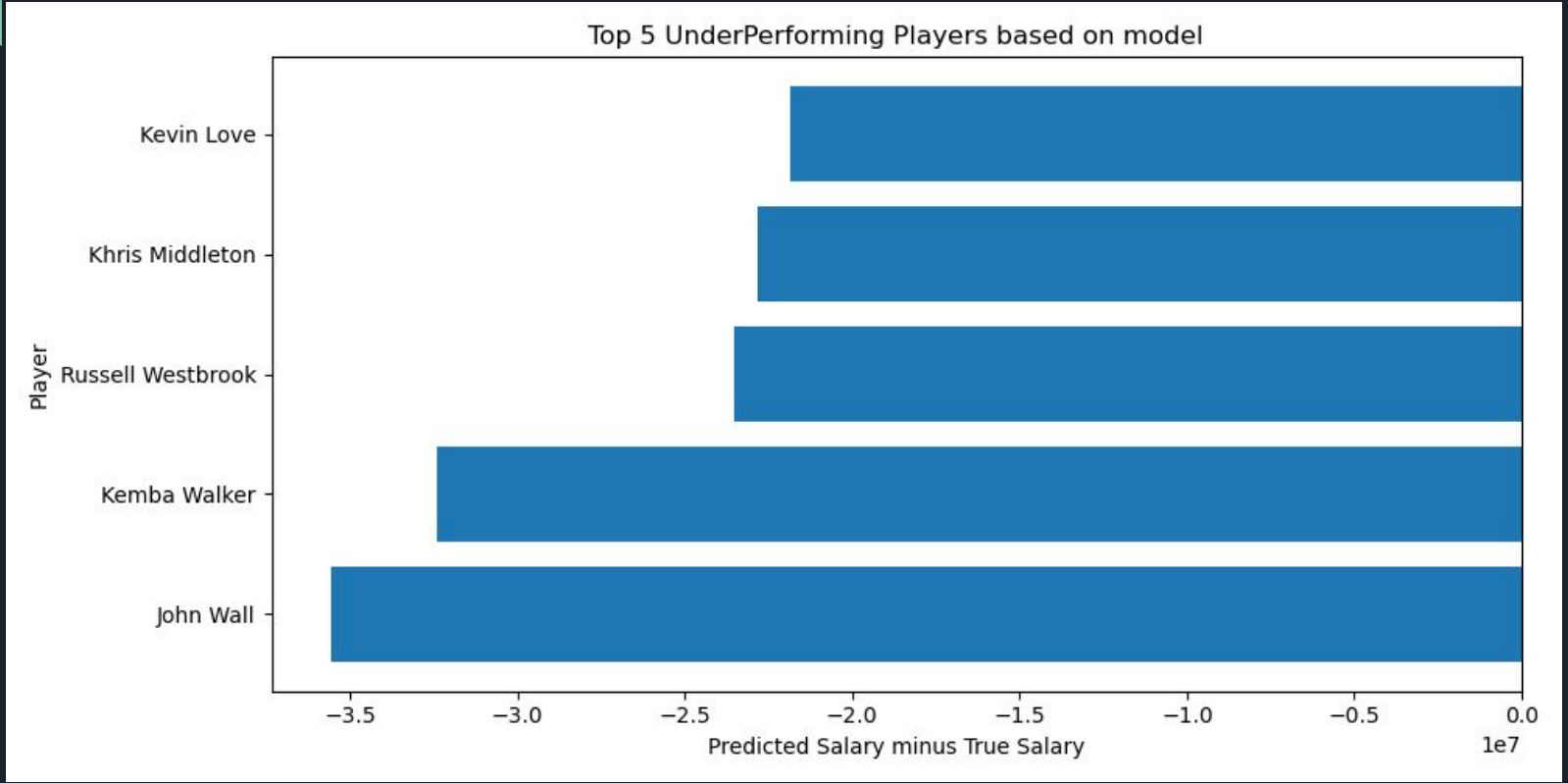
Stacked Model



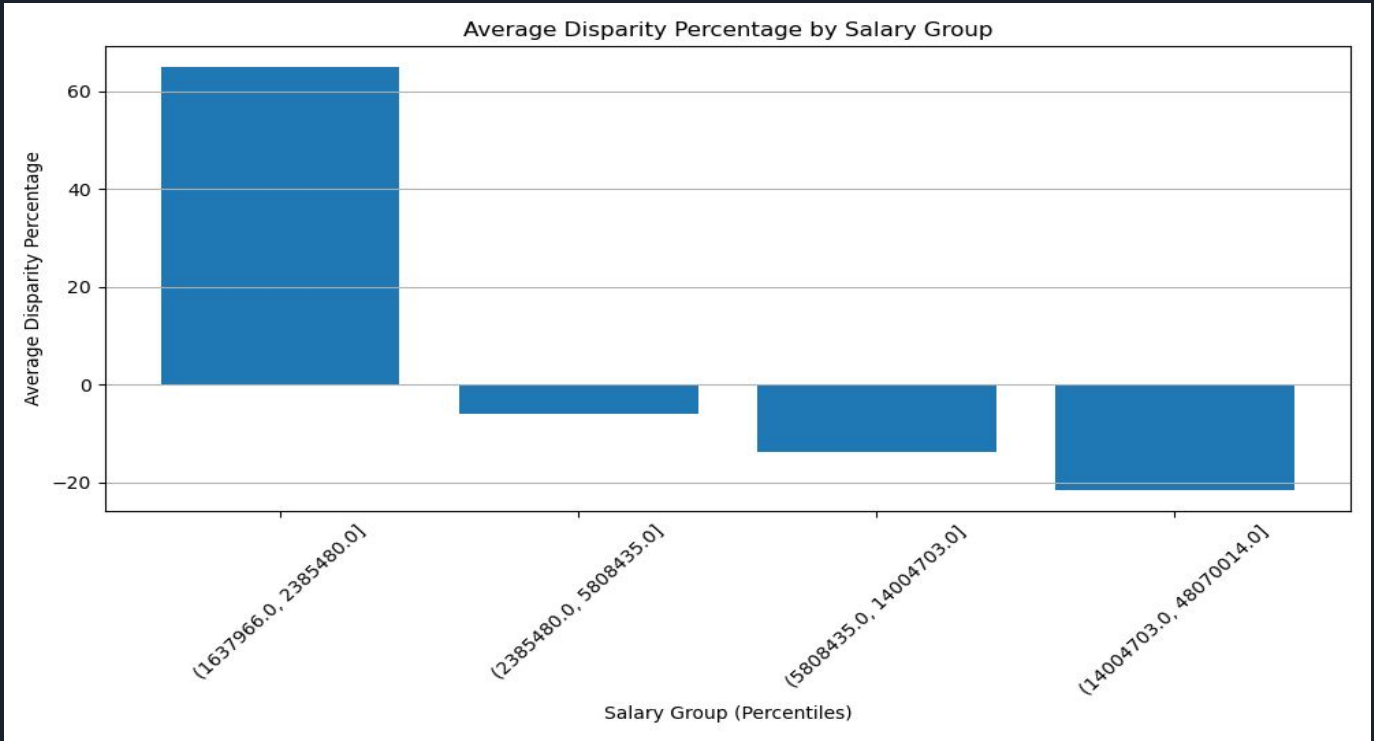
Diamonds in the Rough



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



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



Since the model is based only on stats…

Time to go to the Streamlit App!

Areas 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 fans, and could be valuable as a public resource.

- The model overpredicts lower salaries and underpredicts higher salaries.