

# MLB Contract Tool

---

Group 13:

Will Brannock, Joseph Kaminetz,  
Faizan Khan, Garret Knapp, Nathan Wan



# Project Overview: Research Questions

---

 Using MLB free agent classes from 2003-2015 can we...



**Predict Free Agent's Annual Contract Value (AAV)?**



**Predict Free Agent's Contract Length?**



**What factors impact the predictions the most?**

- Performance statistics, Age, Awards, Etc.



**What characteristics are associated with higher salaries?**

# Data Overview



## MLB Stats API

- Information on Free agent classes from 2003 -Present
- Features:
  - Contract length
  - Player Name
  - Season of free agency



## Lahman Dataset

- Contains comprehensive data from 1871-2023
- Limited Salary data after 2017
- Features:
  - On field statistics (Hits, Home Runs, etc.)
  - Biographical information (age)
  - Award histories (MVP, all star)
  - Salaries



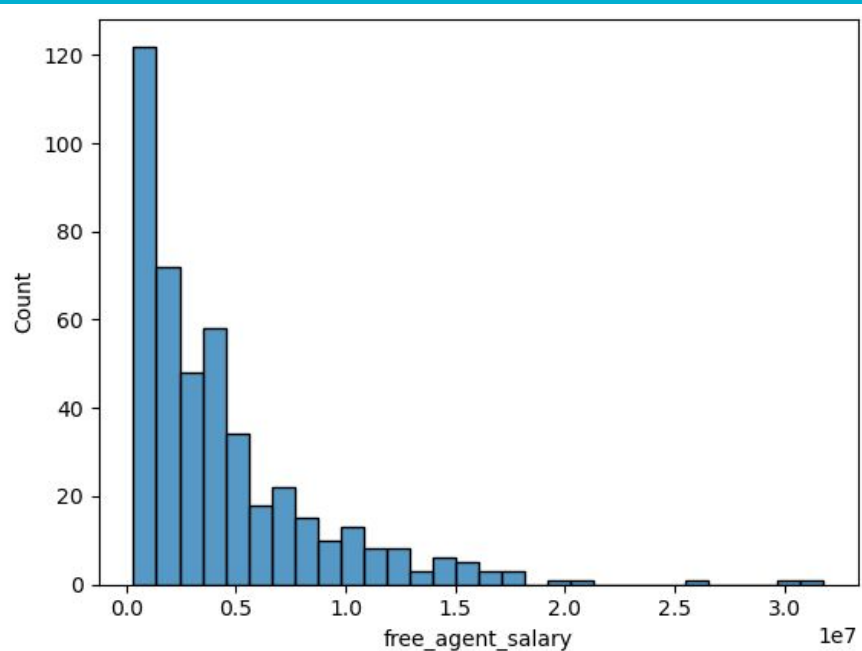
## Cleaned Dataset

Each entry represents a free agent with following features:

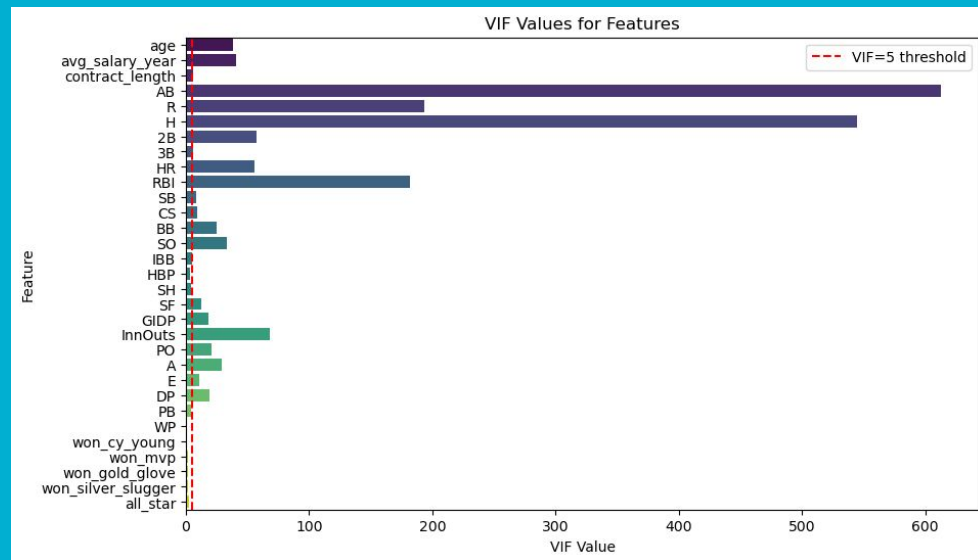
- Averaged/totaled performance statistics
- Binary award categories
- AAV of contract signed
- Contract length
- Age
- Average salary of entire MLB for that season

# EDA

Histogram of Untransformed Response Variable:



Multicollinearity Concerns:



# Dashboard Demo

---

# Conclusion

---

## Model Takeaways

- Salary predictions are good, and have similar errors and  $R^2$
- Batters with better offensive statistics get higher salaries
- Contract prediction able to confidently predict short and long term contracts

## Future Work

- Acquire more data from 2015-present to train and test
- Add ability to factor in team/player options in contracts
- Predict if a player will sign with a different team

Questions?

---

