How To Win the Starting Pitcher Free Agency Market

Michael Scognamiglio
Joe Ramirez
nyc-mnhtn-ds-080320

What Our Project Is About

- This project analyzes strikeout numbers for a starting pitcher in a season.

- The goal of this project is to gain insights into what statistics are most important in determining a starting pitcher's season strikeout numbers.

- Through EDA, hypothesis testing, and model analysis, this project strives to find the 'underrated' or often ignored stats, to provide insight to MLB GMs on who to pursue in the free agency starting pitcher market.

Dataset

- Data Sources: Fangraphs and Baseball Reference via PyBaseball library

- Baseball Reference

  - Uses more traditional and commonly known baseball statistics

  - ERA, BB, Wins, Losses, IP, HR, Strikeouts, etc …

- FanGraphs

  - Uses Sabermetrics and Analytics

  - O-Swing%, F-Strike%, SwStr%, FB Velocity, FIP, Soft%,etc ..

Exploratory Data Analysis / Stat Testing
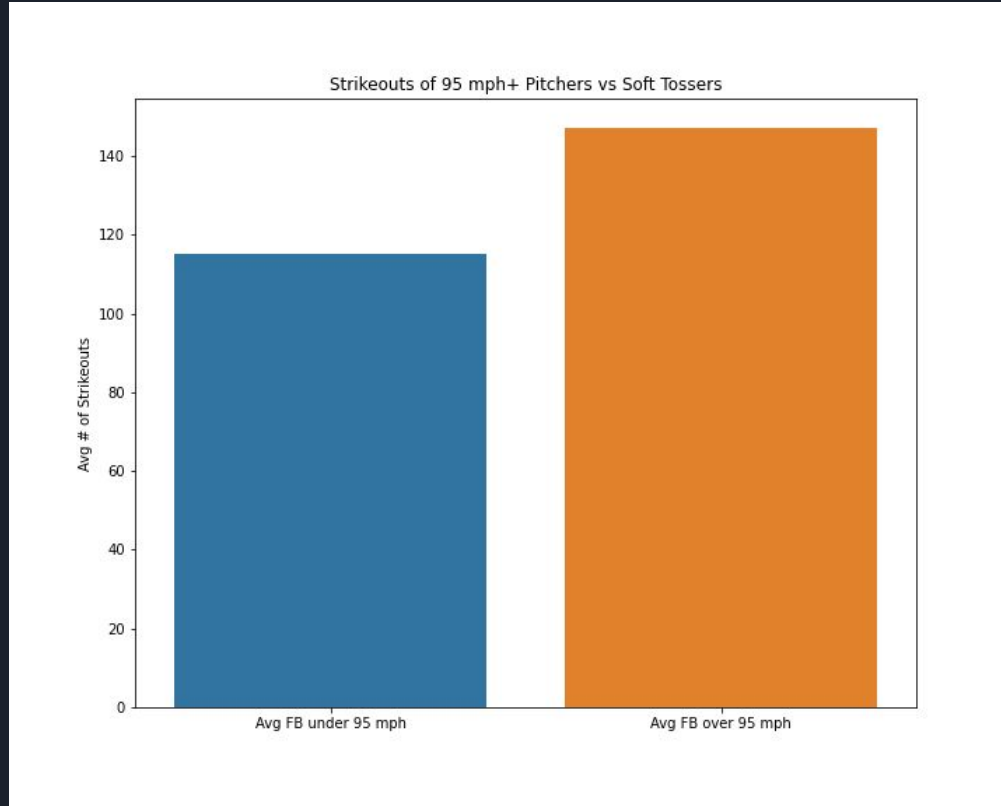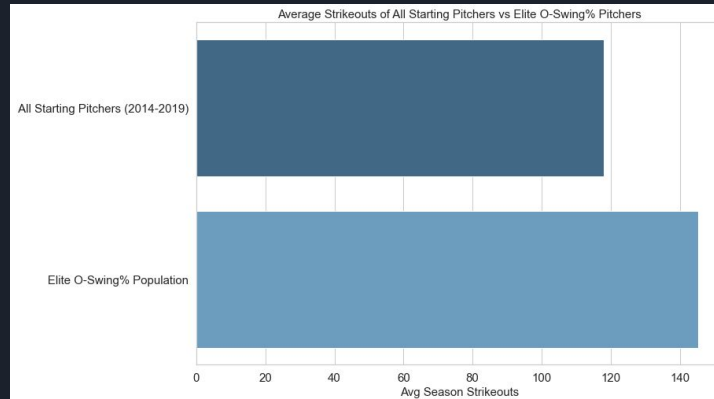

Test Statistic: 5.3

More Stats Testing & EDA

Average Strikeouts for a Starting Pitcher by Best Pitch
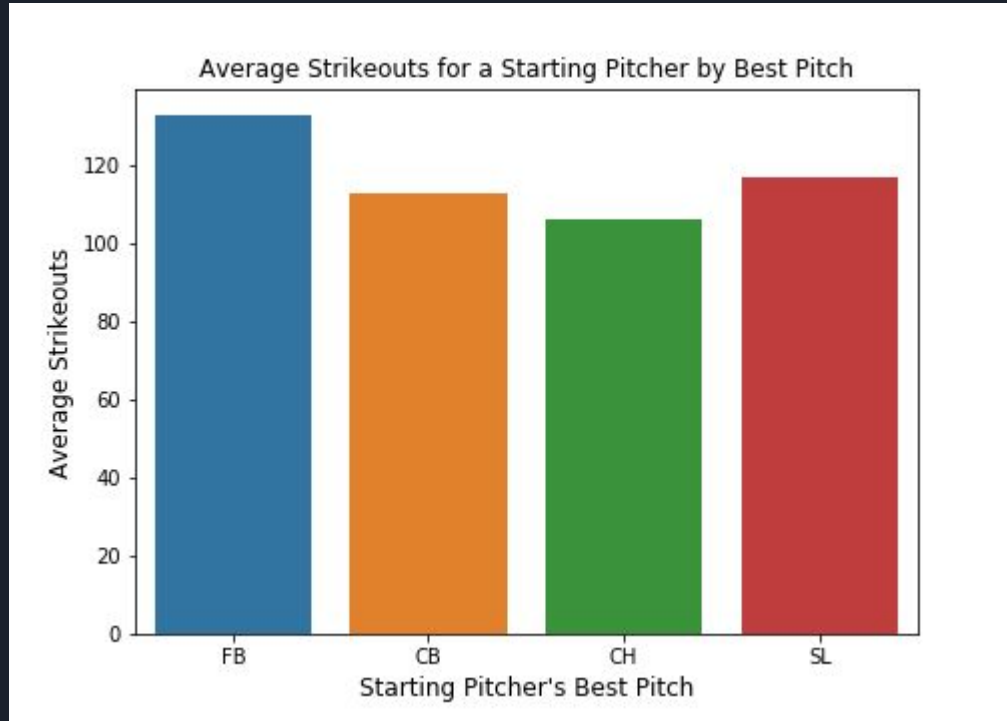
F-Statistic : 12.5

Stat Test/ EDA Summary

- GMs should look very closely at a pitcher's best pitch

- If that best pitch happens to be a fastball, GMs should then look for a pitcher with elite fastball velocity to best capitalize on a starting pitcher's strikeout ability.

- F-Strike% is also vital in determining a starting pitcher's ability to obtain strikeouts controlling for elite fastball velocity and high endurance. (P-Value = .002)

Average Strikeouts of All Starting Pitchers vs Elite O-Swing% Pitchers

All Starting Pitchers (2014-2019)	
Elite O-Swing% Population	

Avg Season Strikeouts

0   20   40   60   80   100   120   140

P-Value=10^-18 (essentially 0)

Modeling Selection

| Model Version	Training RMSE	Testing RMSE
Basic Linear Regression	7.2 (strikeouts)	8.6 (strikeouts)
Linear Regression with recursive feature elimination algorithm	7.3 (strikeouts)	8.8 (strikeouts)
Linear regression with K best feature elimination method	8.5 (strikeouts)	10.3 (strikeouts)
Ridge Model	7.5 (strikeouts)	8.8 (strikeouts)

Modeling
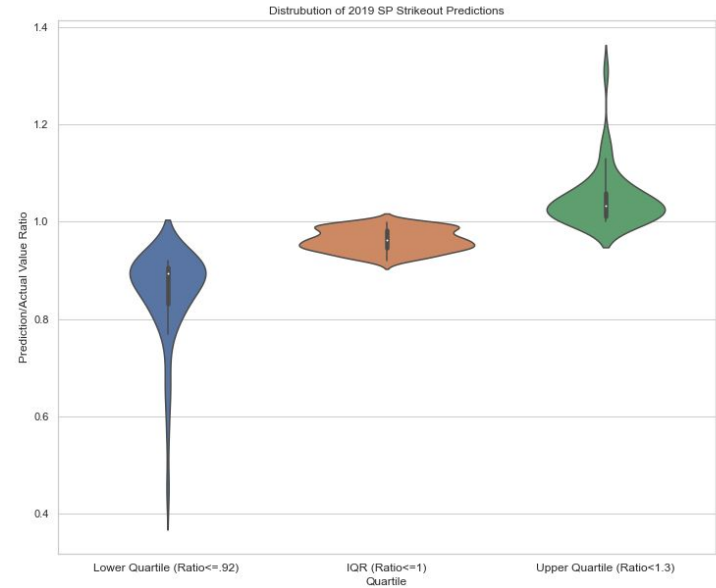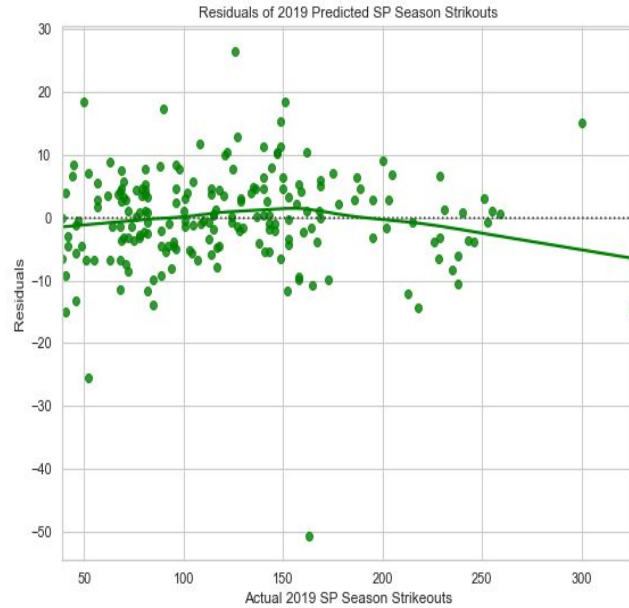
Modeling Real World Takeaways

- Based on our model's performance, we feel that many important insights into our target variable can be found by model analysis.
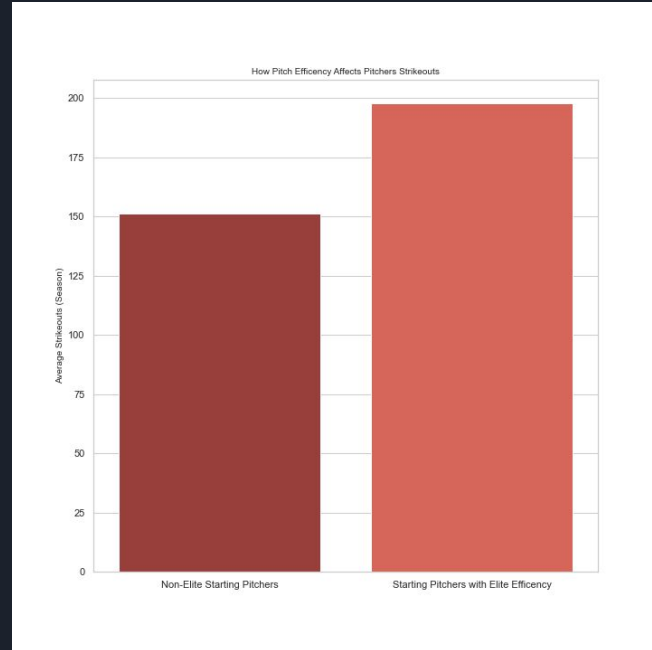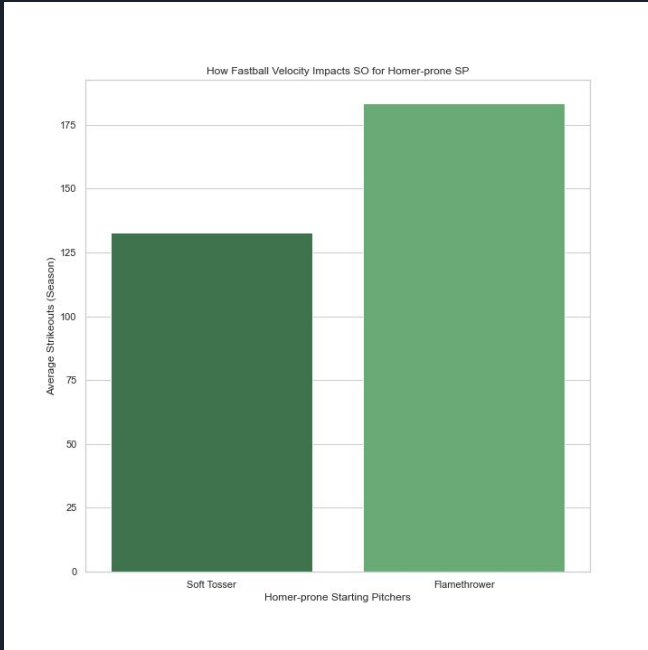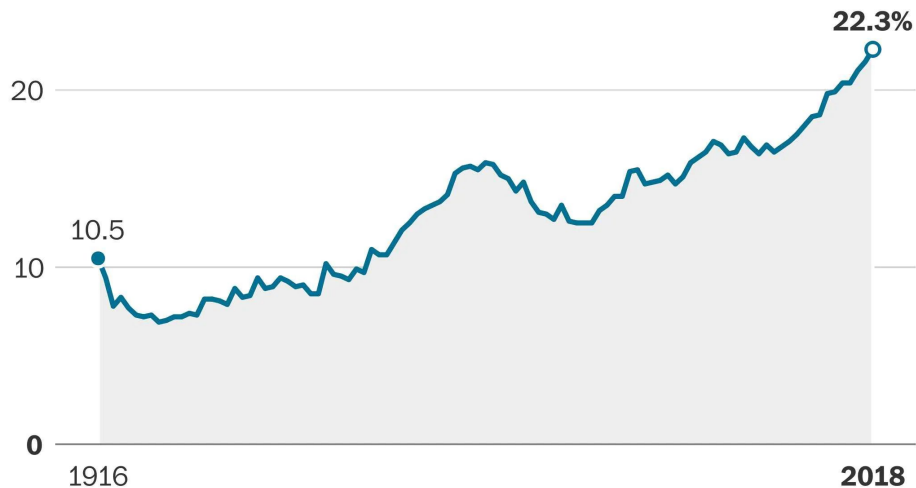
Further Reading / Next Steps

- Create a highly accurate interpretable models for relief pitchers.

  - This model would test to see if insights gained in our current model still hold for relief pitchers.

- Create a General Pitcher Predictive Model for ERA.

  - The goal would be to test whether the insights obtained during this modeling process would help the performance of this new predictive model.

Any Questions?

Strikeouts are on the rise

The strikeout rate in 2018 is higher than at any other time in major league history.

22.3%

20

10.5

10

0

1916   2018

Source: FanGraphs

FANCY STATS