

MLB Fantasy Points Per Game Predictor



Patrick Cudo



My name is Patrick Cudo and I am a
freelancing data science and recent General
Assembly graduate.

\$4,000,000,000

According to the MLB Hall of Fame, fantasy baseball generates revenues approaching \$4-billion.



Problem Statement

The goal of this project is to build a regression model to predict fantasy points per game for MLB players in the current 2021 season. This model will be evaluated using RMSE and the model should improve upon the baseline by at least 25%. The model's predictions and coefficients will be used to create a lineup building application.



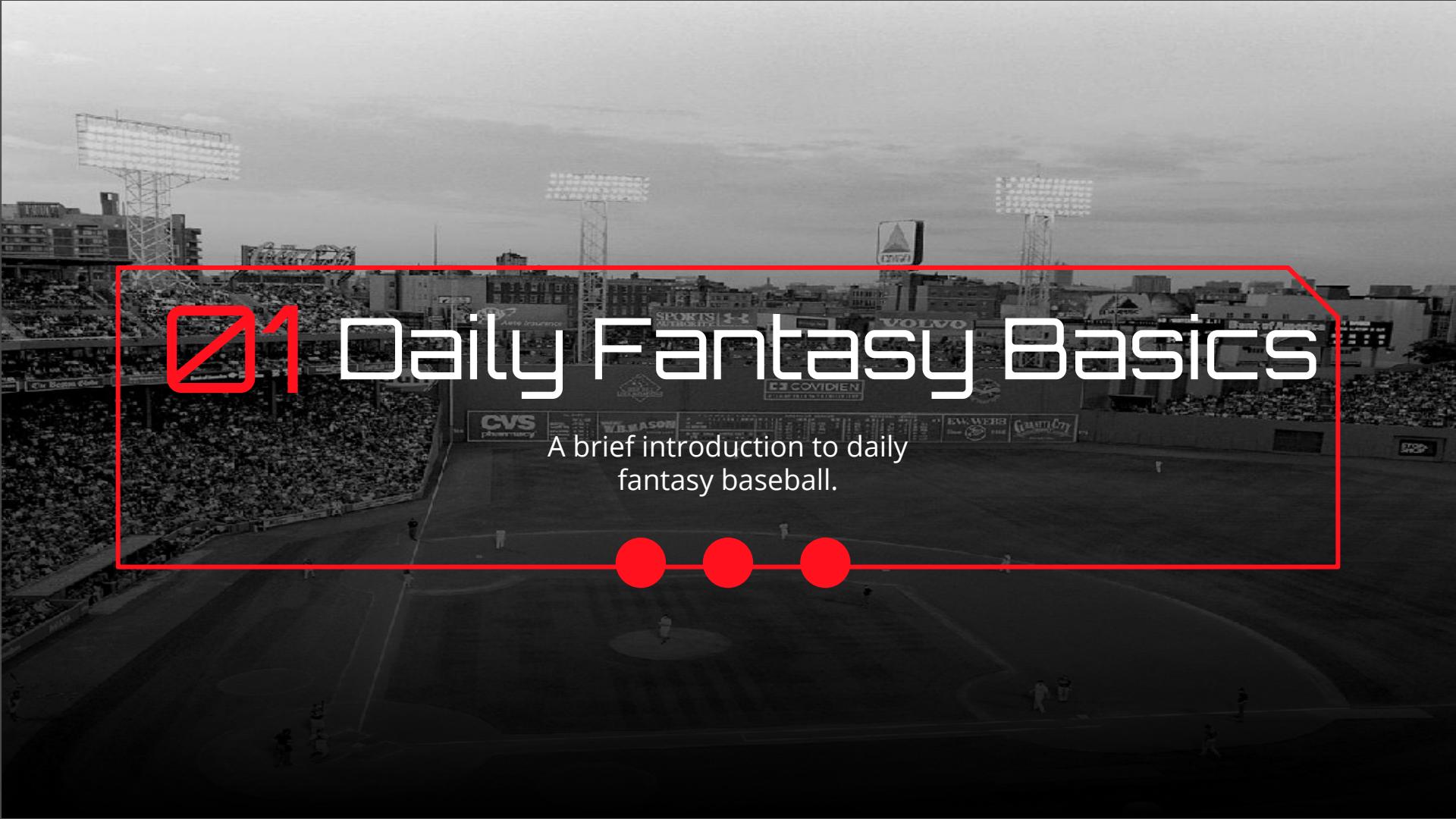
Product Overview

The application will provide a MLB daily fantasy lineup for a user. This product is geared towards companies that provide services to fantasy players. The product can be used in trial periods for potential users or as a product offered to existing users.



Agenda

- 01 Daily Fantasy Basics
Provide a brief explanation of daily fantasy baseball.
- 02 Data Collection
Data sources and cleaning techniques.
- 03 Modeling Process
Steps taken in the modeling process.
- 04 Model Evaluation
Review the metrics used to score the model.
- 05 Demo
Demo of application created for specific use case.
- 06 Conclusion and Recommendations
Conclude project and make recommendations.



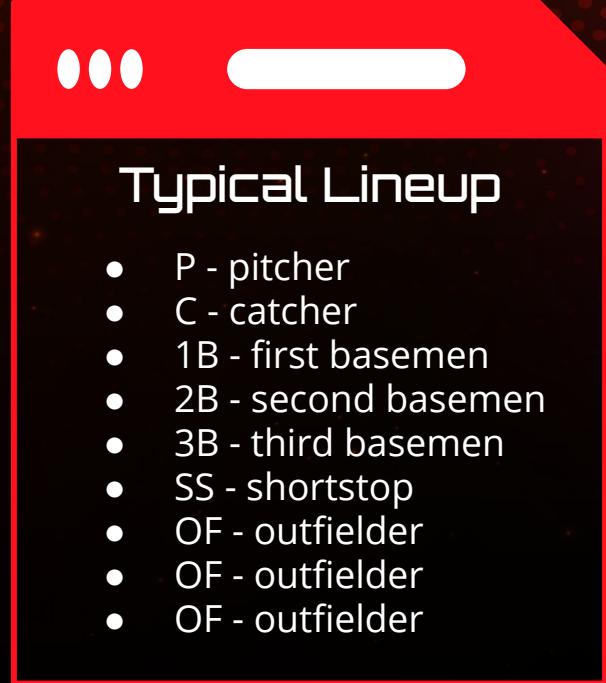
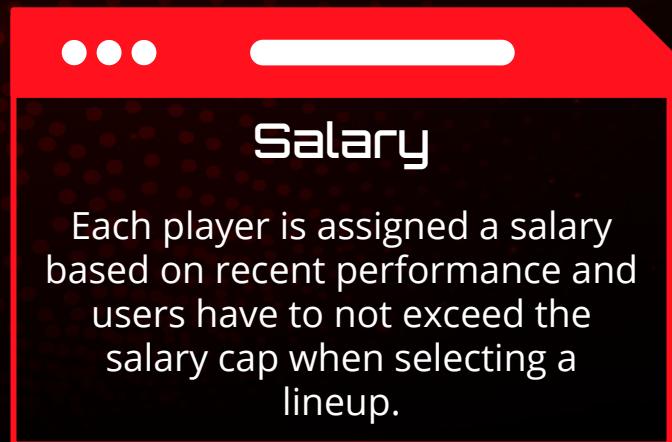
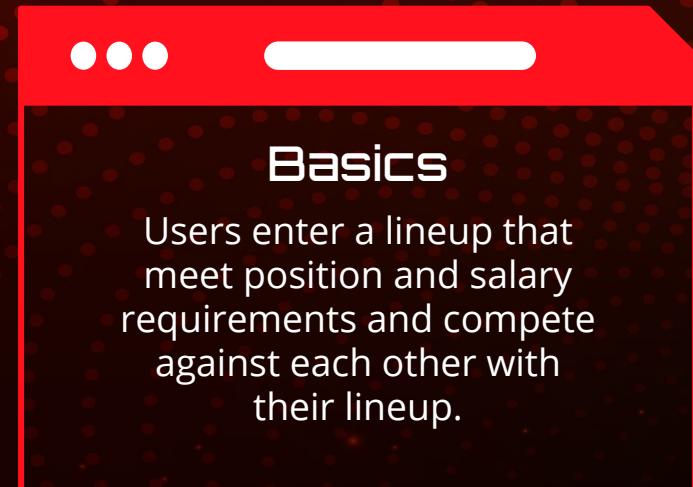
01 Daily Fantasy Basics

A brief introduction to daily fantasy baseball.



Daily Fantasy BASICS

Lineups



Daily Fantasy Scoring Basics



Scoring - FanDuel

Pitchers

- ER - earned runs = -3
- IP - innings pitched = 3
- SO - Strikeouts = 3
- W - Wins = 6
- QS - Quality Start = 4

Batters

- 1B - single = 3
- 2B - doubles = 6
- 3B - triple = 9
- BB - walks = 3
- HBP - hit by pitcher = 3
- HR - home runs = 12
- R - runs = 3.2
- RBI - run batted in = 3.5
- SB - stolen bases = 6



02 Data Collection

Sources of data and cleaning
processes.



Data Sources

- Provides data from every player since 1871
- Has advanced statistics, wOBA, ISO and WAR
- Gives user the ability to filter and export data

[Fangraphs.com](https://fangraphs.com)

AND

- Provides fantasy points data from every player since 2010
- Provides points for 5 different scoring platforms
- Gives user the ability to filter and export data

fantasydata.com

Data Combination



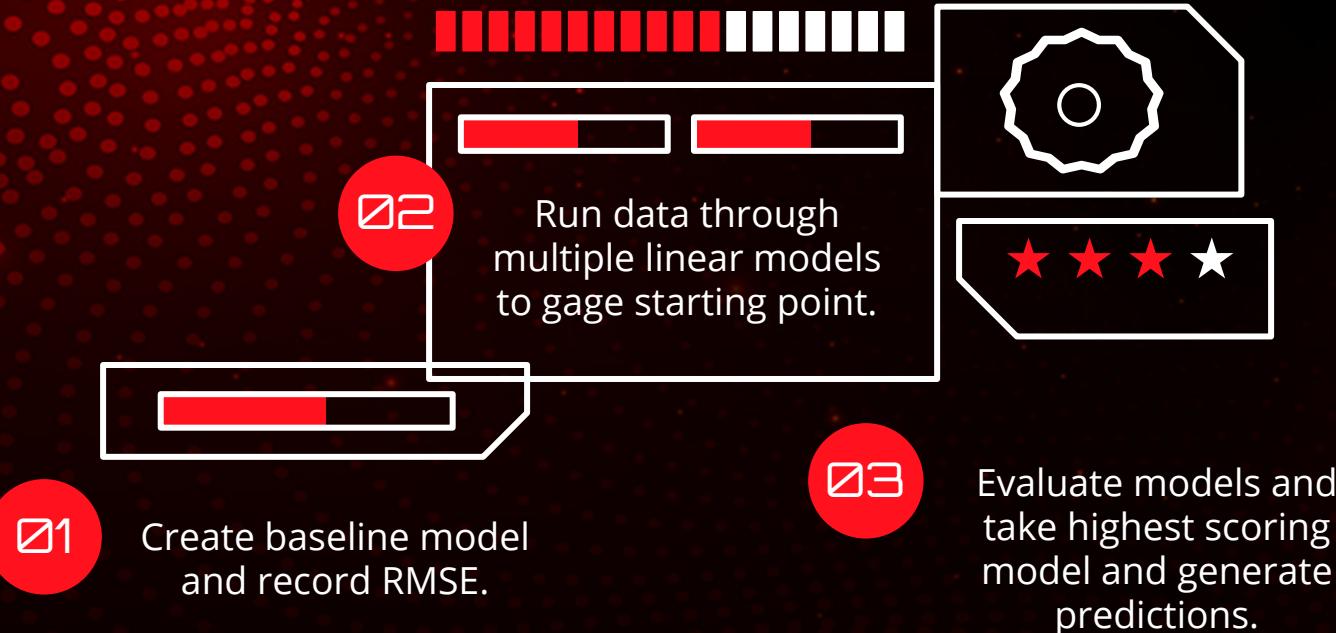


03 Modeling Process

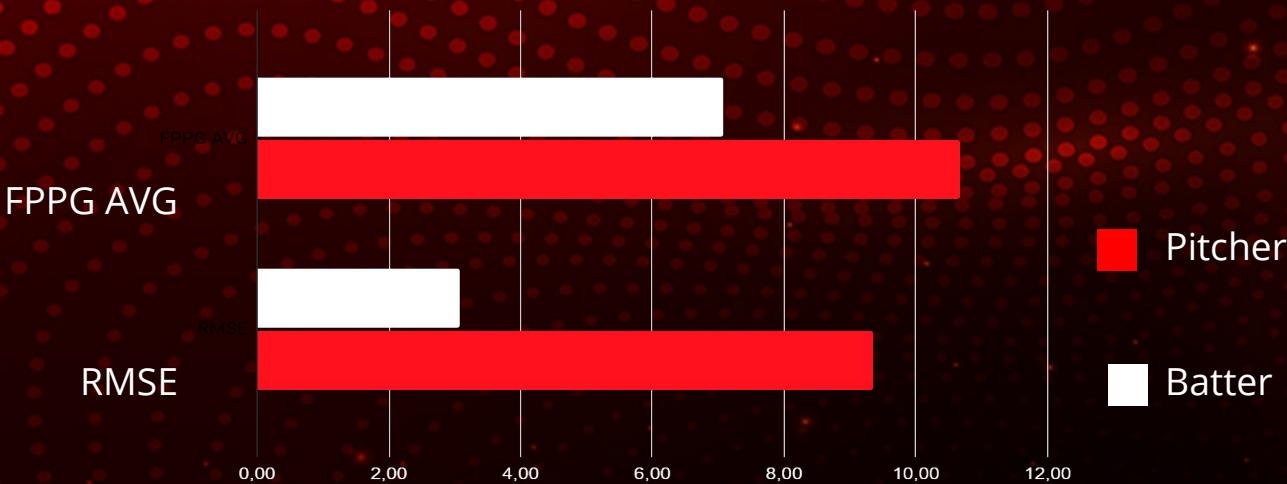
Steps taken in the modeling process to predict FPPG.



Modeling Workflow



Baseline Model



The graph above shows RMSE and FPPG average for each position group.

Pitchers

Pitchers have an FPPG avg of 10.67 and an RMSE of 9.34.

Batters

Batters have and FPPG avg of 7.09 and a RMSE of 3.08.

Linear Models

Linear Regression



KNeighborsRegressor



Random Forest



Bagging Regressor



Decision Tree Regressor



AdaBoostRegressor





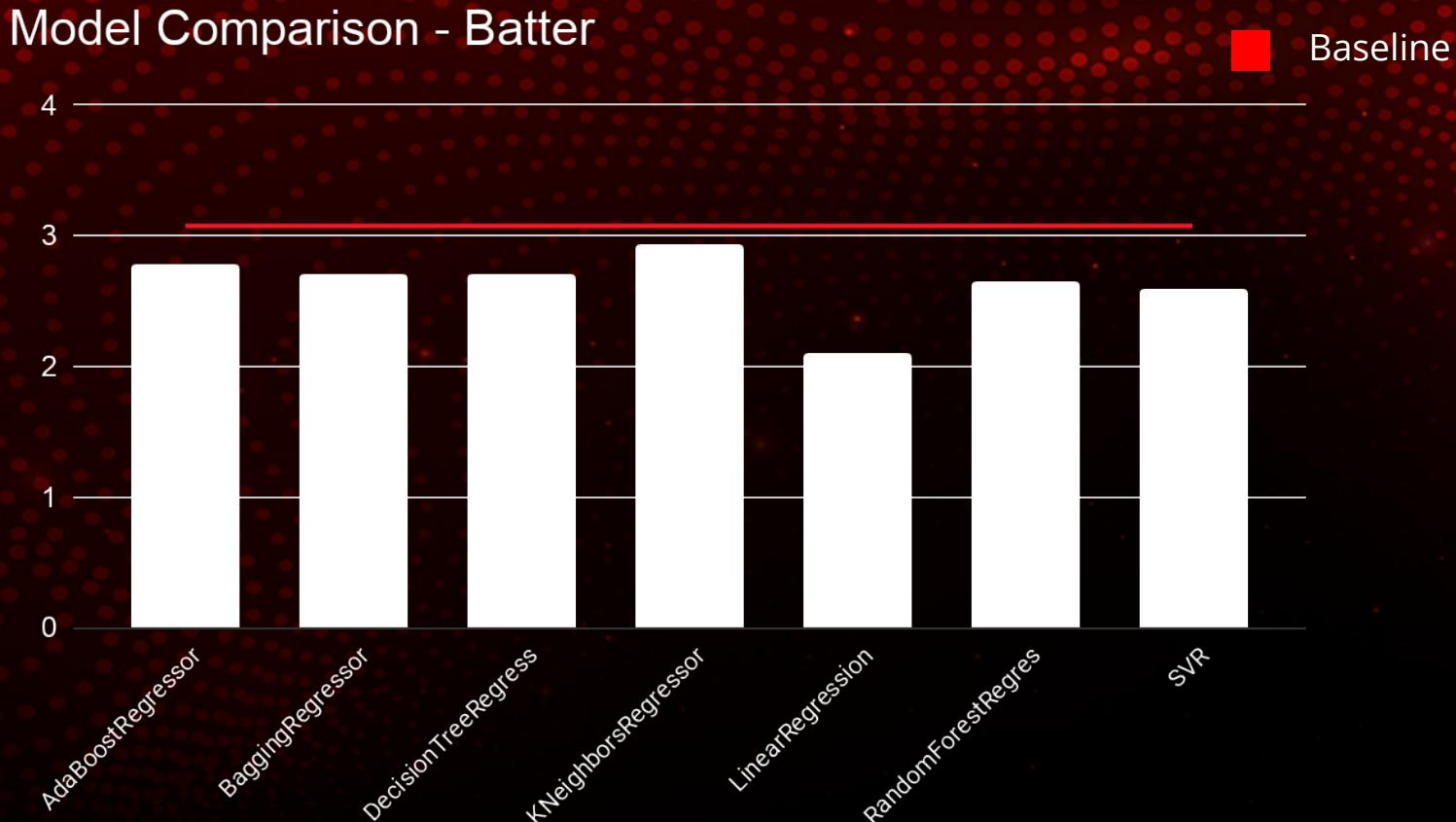
04 Model Evaluation

Evaluate the model's
performance.

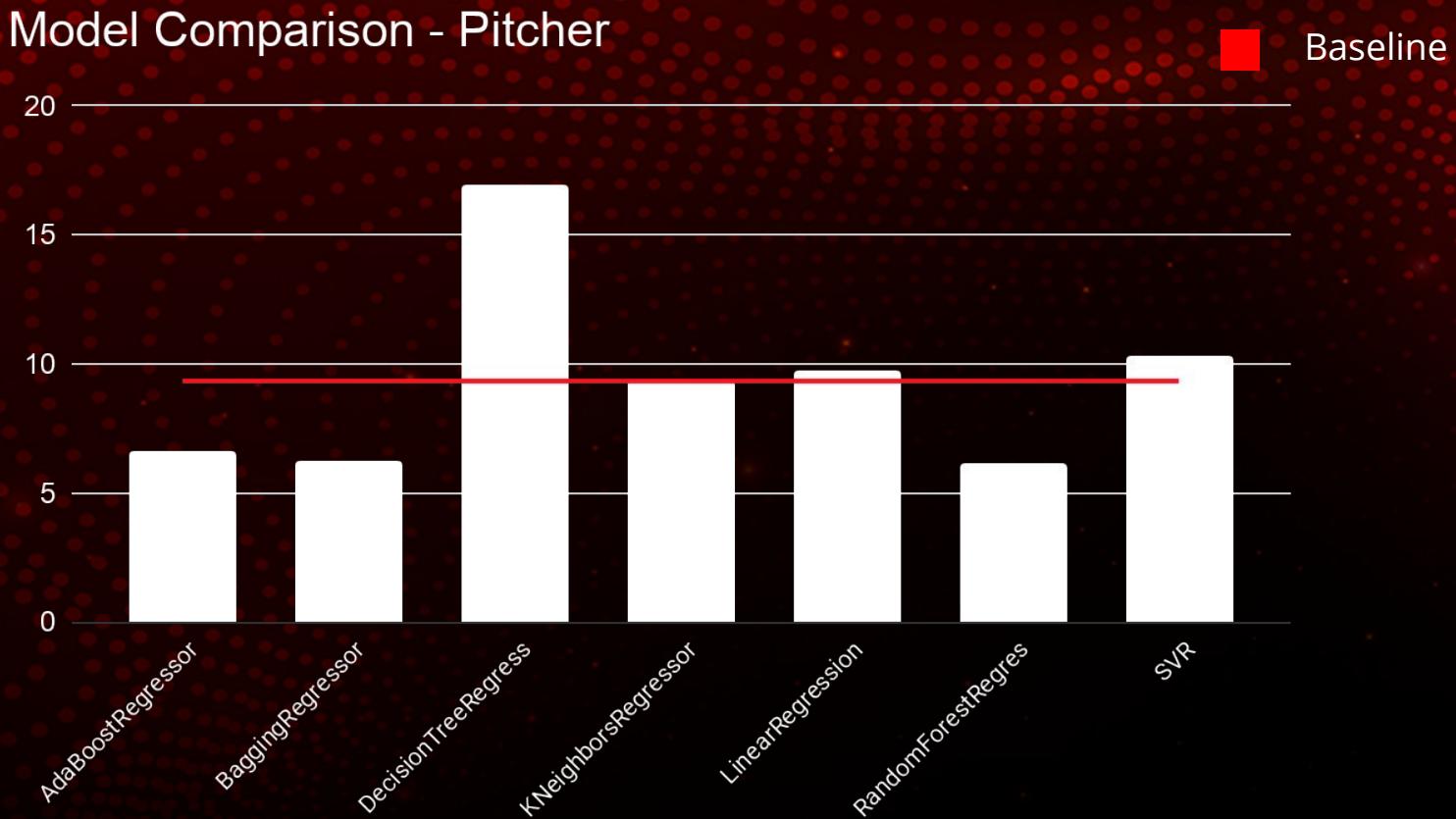


Model Evaluation

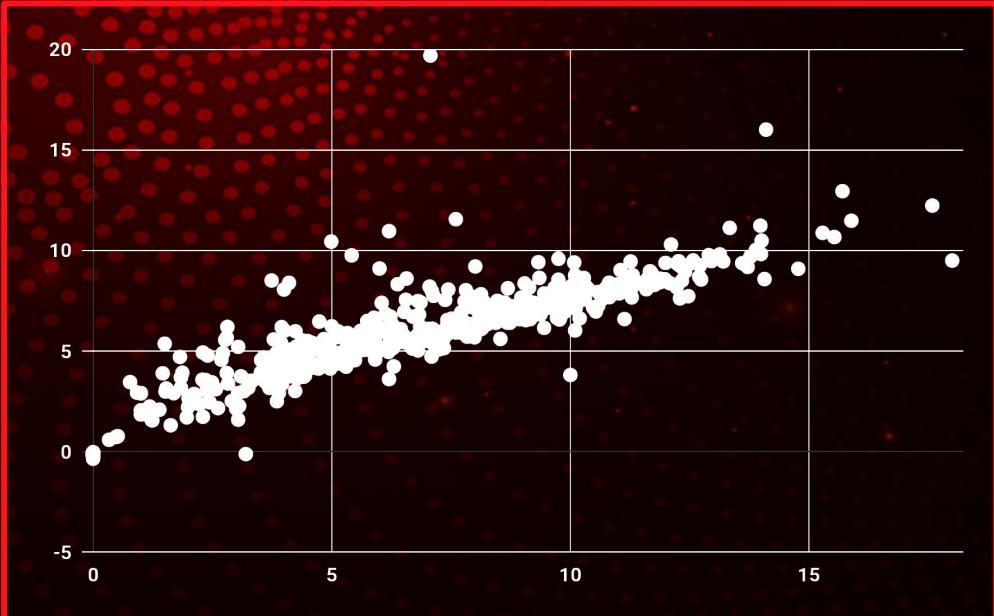
Batter



Model Evaluation Pitcher



Batter Predictions Linear Regression



The scatterplot above shows FPPG vs Projected FPPG for batters.

FPPG vs Projected FPPG

2.0962

RMSE

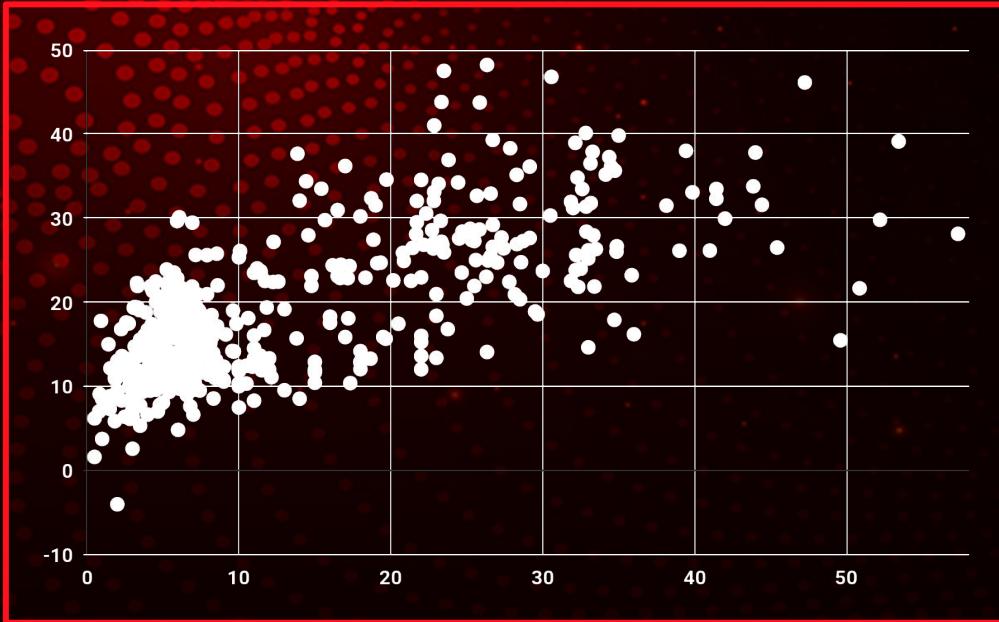
.6601

R2 Score

31.9%

Improvement from baseline.

Pitcher Predictions Random Forest



The scatterplot above shows FPPG vs Projected FPPG for pitchers.

■ FPPG vs Projected FPPG

6.147

RMSE

.6883

R2 Score

34.2%

Improvement from baseline.

Coefficient Values

Feature	Coefficient Value
BB - walks	30.8697
SO - strikeouts	4.359
IP - innings pitched	2.7645

Pitchers

VS

Feature	Coefficient Value
AB - at bats	3.8836
R - runs	1.1378
RBI - runs batted in	1.0872

Batters



05 Demo Application

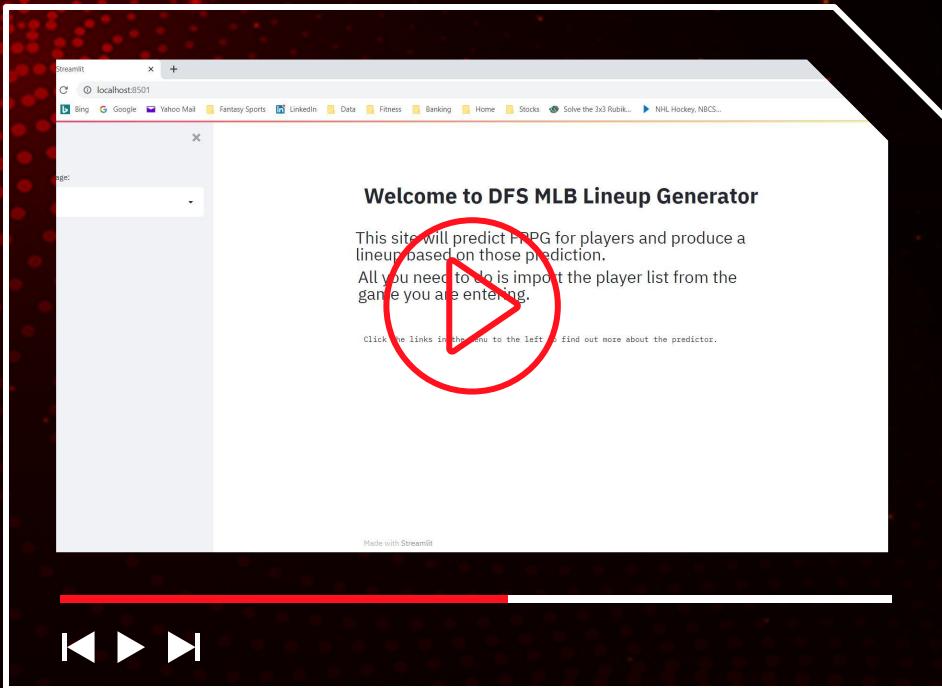
Walkthrough of basic application and comparison.



Application Workflow



Streamlit Application Demo



Now lets predict FPPG and generate a line up.



Conclusions and Recommendations

Conclude presentation with recommendations.



Conclusions & Recommendations

Conclusions

- The final model provided improvement over the baseline RMSE, with a value of .982 batters and 3.014 with pitchers. This made the goal of at least a 25% improvement over the baseline model for each position group.
- With this improvement the predictions and coefficient values are still useful in the context of the model and were used with success in the application.

Recommendations:

- More data, increase the scope of the project data for more details.
 - For example, using SportsData.io's API would more than double the amount of features.
- Provide more feature engineering to improve model performance.
- Split the problem up use a generalized linear model to gather the best coefficient values and a neural network to produce the best predictions.

MLB Hall of Fame

- <https://baseballhall.org/discover/start-of-fantasy-baseball>

PHOTOS

- <https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.pathrise.com%2Fguides%2Fa-review-of-general-assembly-as-a-tech-educational-tool%2F&psig=AOvVaw39Yyao-oPoe-qcs01SqZ1h&ust=1620848349858000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCLC-m-SwwACFQAAAAdAAAABAD>
- <https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.forbes.com%2Fteams%2Fphiladelphia-phillies%2F&psig=AOvVaw3XodH89Cjw1abc0kFxVCox&ust=1620848385426000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCjshfWwwACFQAAAAdAAAABAD>
- <https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.horschgallery.com%2Ffenway-park%2Fblack-white-fenway-park-red-sox&psig=AOvVaw2s4MmGIngMSAVNOFKUNcK6&ust=1620848418579000&source=images&cd=vfe&ved=0CAIQjRxqFwoTClicxoSxwvACFQAAAAdAAAABAE>
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fbaseballhall.org%2Fdiscover-more%2Fstories%2Fbaseball-history%2Fwrigley-field-ivy&psig=AOvVaw3xMAk9l_Fvbag6pWNalE4I&ust=1620848461716000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCNCqipmxwvACFQAAAAdAAAABAF
- <https://www.google.com/url?sa=i&url=https%3A%2F%2Fpixels.com%2Ffeatured%2Fpittsburgh-pirates-pnc-park-bw-x1-david-haskett.html&psig=AOvVaw2VbK-GhGhqRaemJtfbJu5&ust=1620848518378000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCMC67bOxwvACFQAAAAdAAAABAM>
- <https://www.google.com/url?sa=i&url=https%3A%2F%2Ftokens-icons.com%2Fproducts%2Fyankee-stadium-seat-pendant&psig=AOvVaw1jQ2aCTzKsEG4EFciUqCSm&ust=1620848797061000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCKD04biywvACFQAAAAdAAAABAE>
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fphotos.visualjourney.com%2Fblack-and-white%2Fh21D09312&psig=AOvVaw3_26a6ZjOoDwhFj7ESTh5&ust=1620849158643000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCPDsneqzwvACFQAAAAdAAAABAD
- <https://www.google.com/url?sa=i&url=https%3A%2F%2Fmarkoristano.photoshelter.com%2Fimage%2FI0000j4AvhEok48E&psig=AOvVaw3eheGrjISM9JvdE1eLqieS&ust=1620849005686000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCID1Jy0wvACFQAAAAdAAAABAE>

THANK YOU

Do you have any
questions?



CREDITS: This presentation template was created by Slidesgo,
including icons by Flaticon, infographics & images by Freepik

THANK YOU

Do you have any
questions?

CREDITS: This presentation template was created by Slidesgo,
including icons by Flaticon, infographics & images by Freepik

