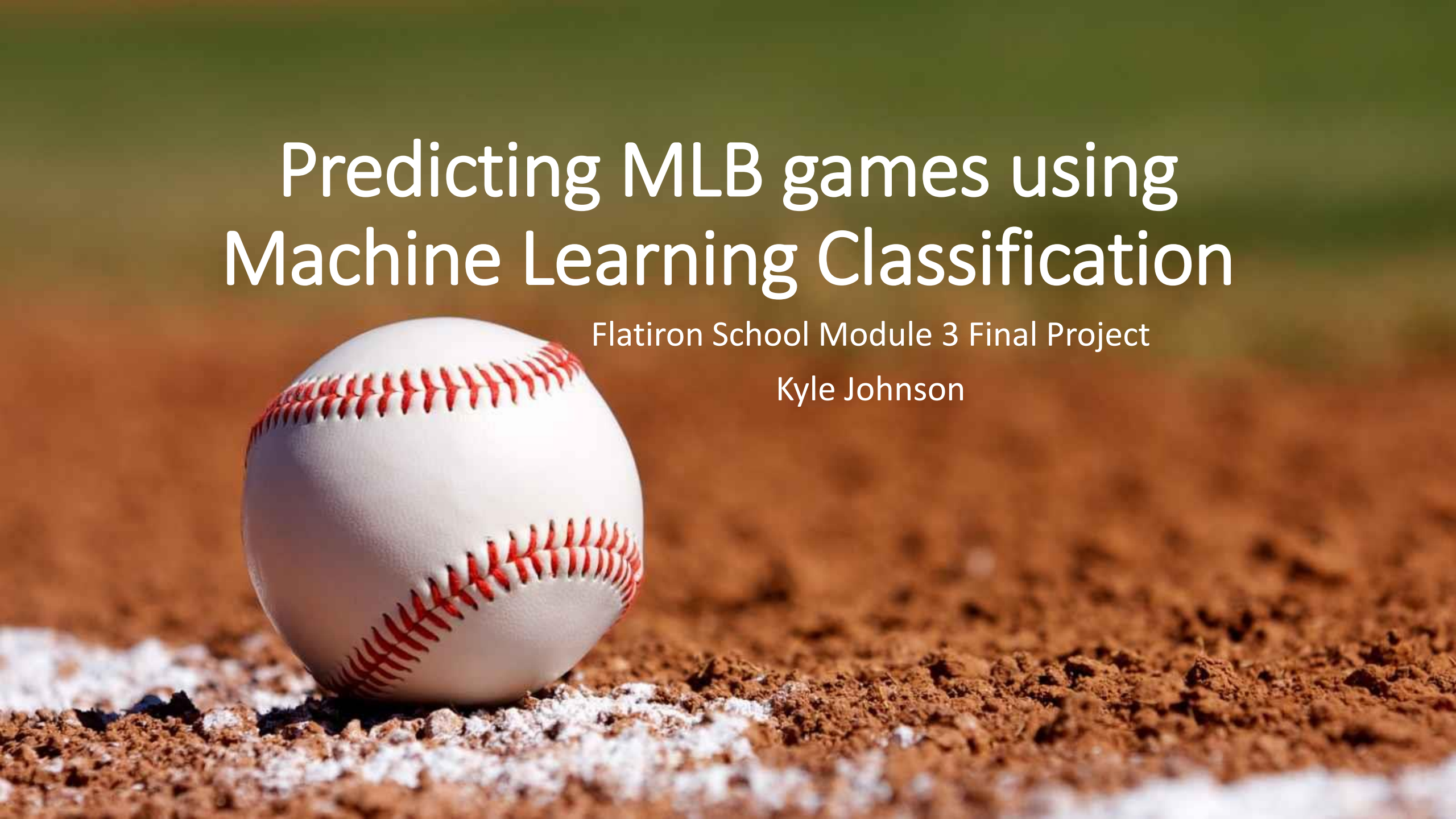
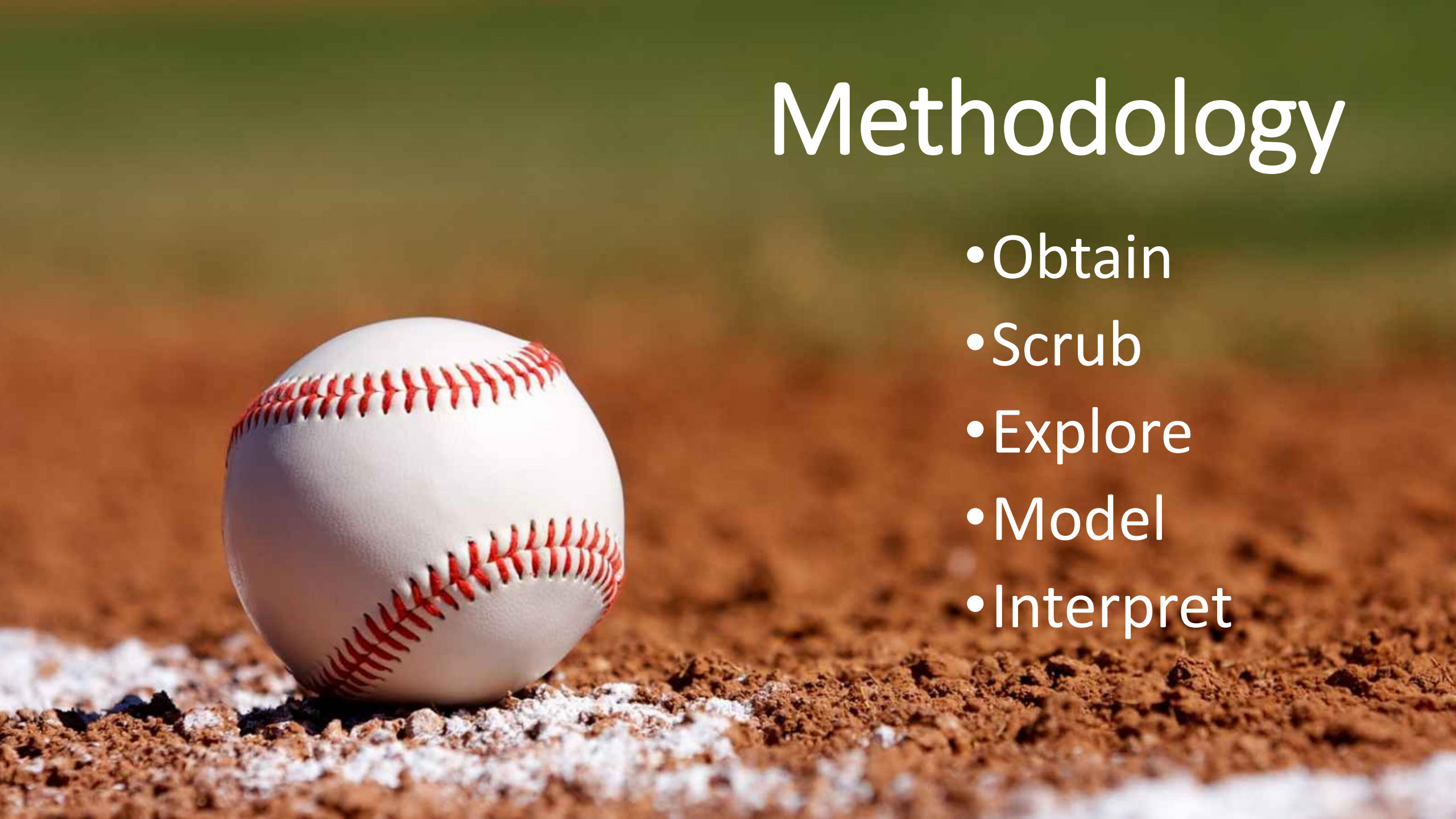
Predicting MLB games using Machine Learning Classification

Flatiron School Module 3 Final Project

Kyle Johnson





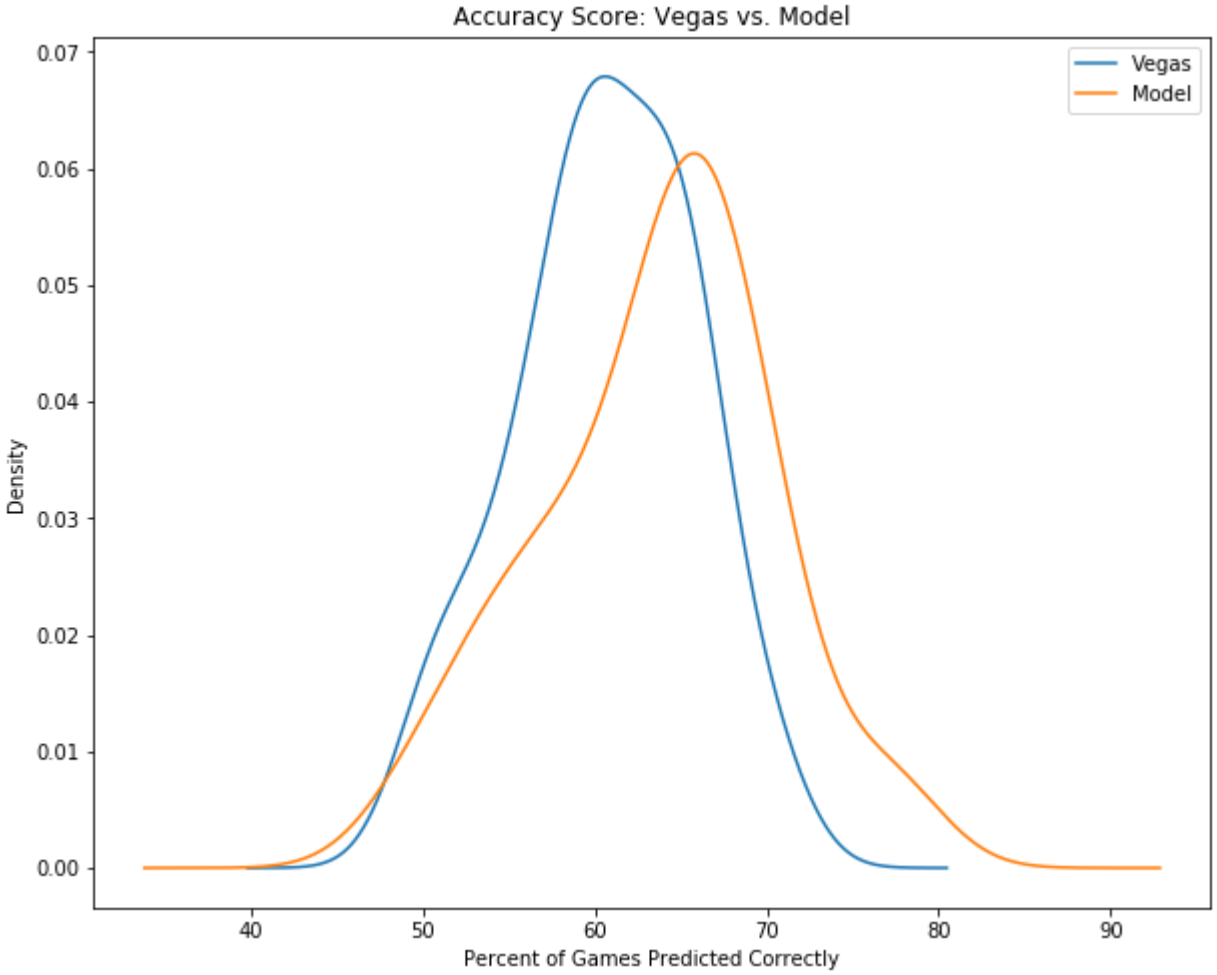
Methodology

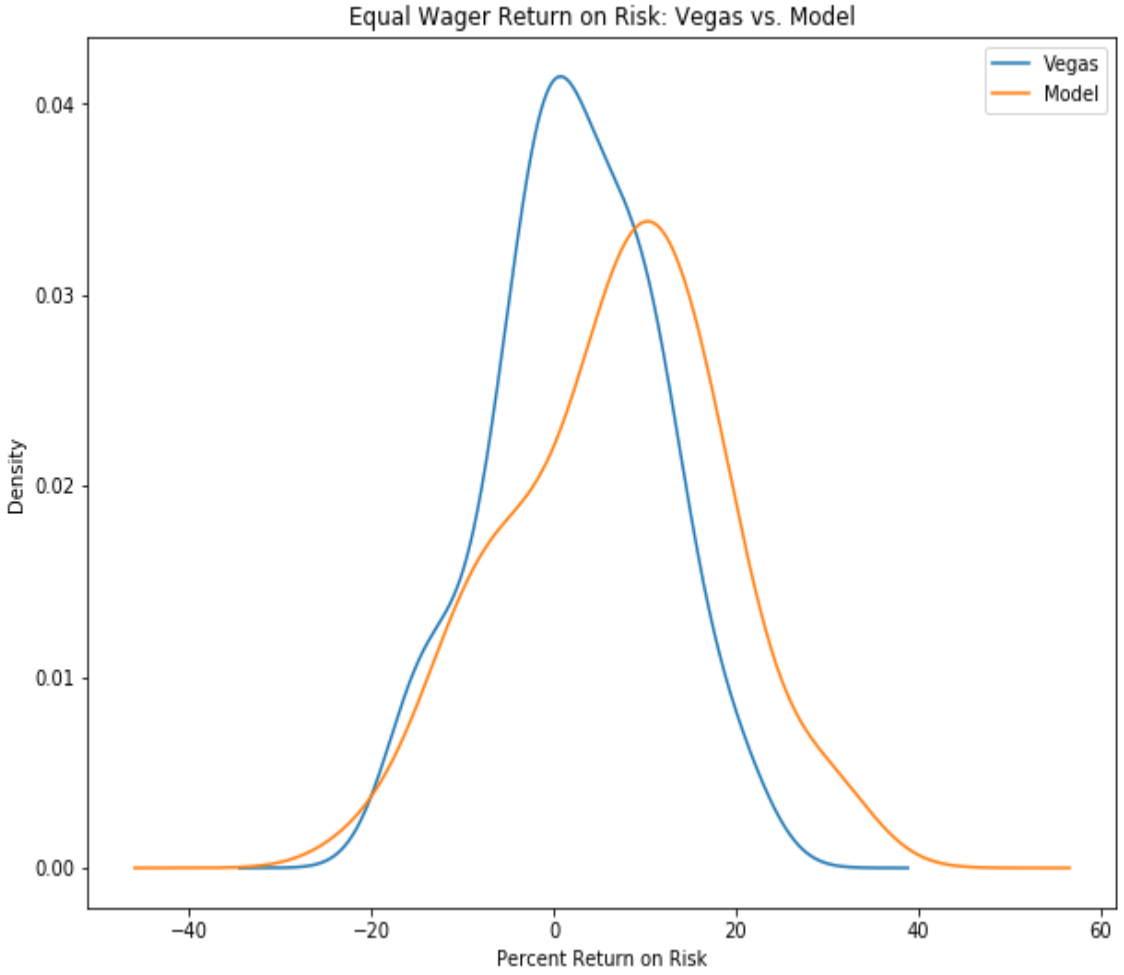
- Obtain
- Scrub
- Explore
- Model
- Interpret

The Data

- Sourced from:
  - baseball-reference.com
  - MLB Advanced Media API
  - Sportsbookreviewsonline.com

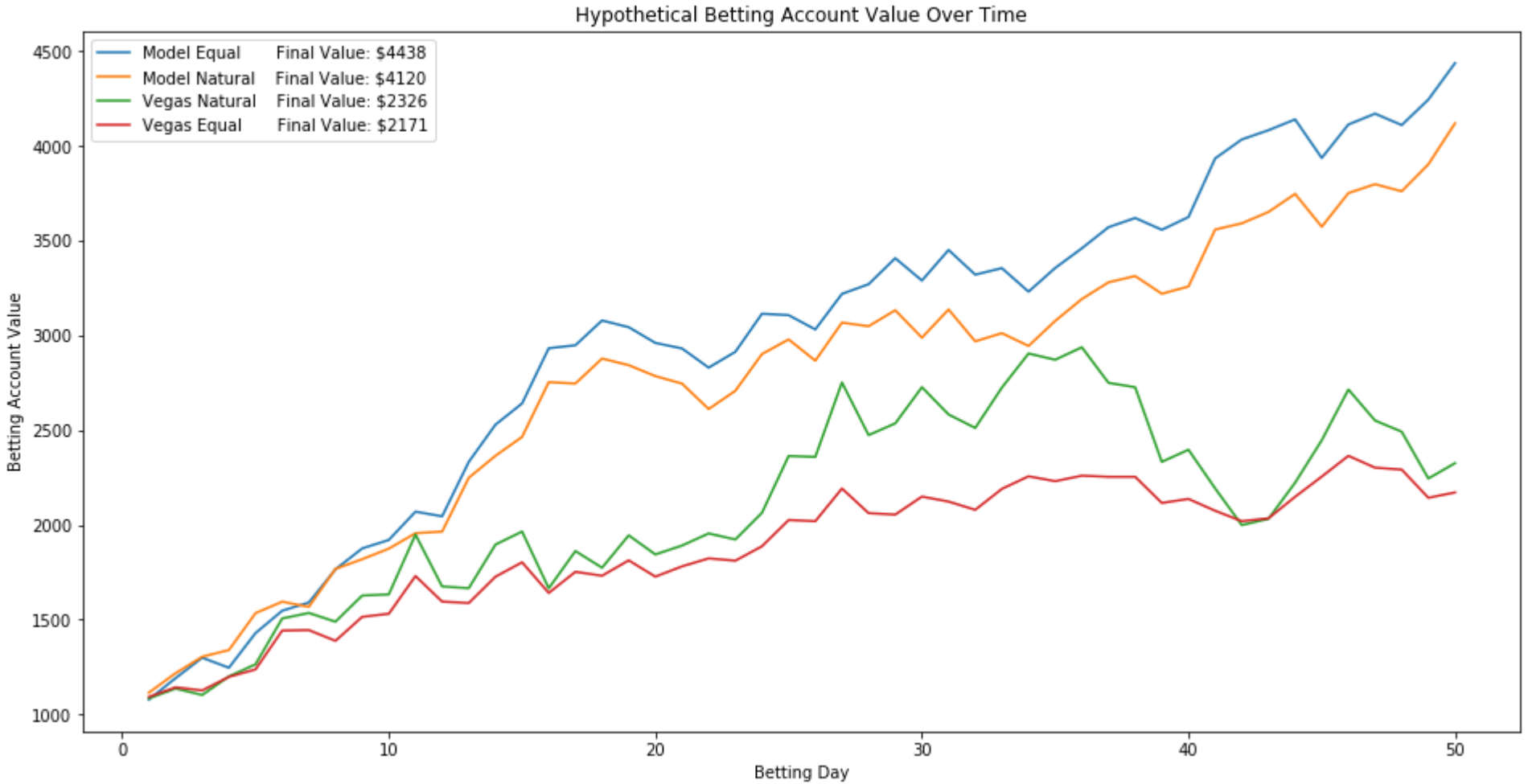
Findings: The model predicts a greater percentage of games correctly and achieves a higher return on risk in a statistically significant manner.



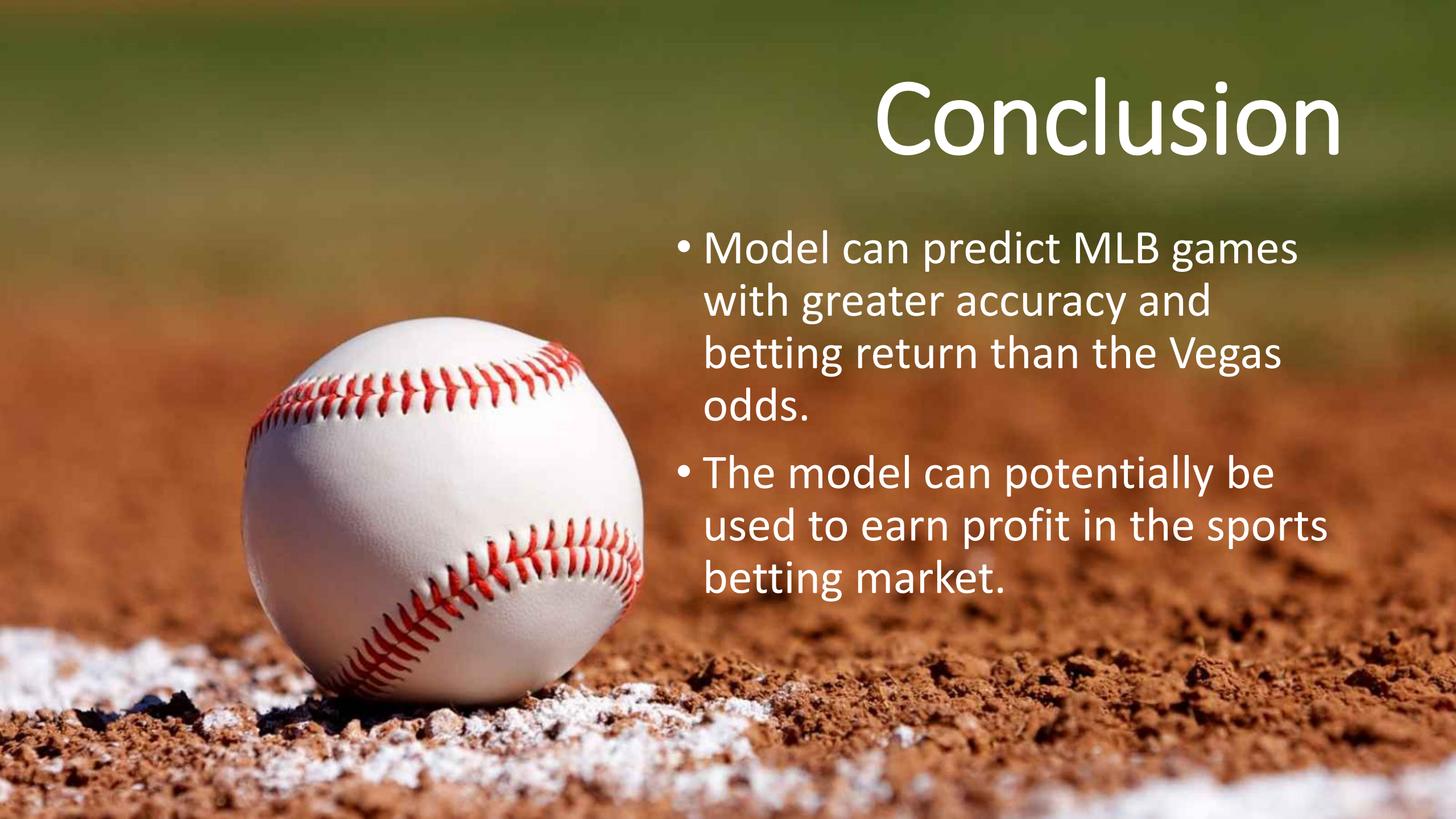


*For details about the models used, please see the appendix on the last slide

Hypothetical Application in the Sports Betting Market:



Conclusion



- Model can predict MLB games with greater accuracy and betting return than the Vegas odds.
- The model can potentially be used to earn profit in the sports betting market.

Future Work

- Use more types of data including the new and highly advanced statistics that have become available with the use of new camera technology
- Use more games from previous seasons
- Create a process to automate and publish the predictions

Appendix

The final model is the combination of the Support Vector Machines and the XG Boost listed in red below.

Model	Accuracy	Natural Return on Risk	Equal Wager Return on Risk
Vegas Odds	61.1%	2.53%	3.11%
Random Forest + Support Vector Machines	60.36%	4.2%	4.76%
Random Forest + XG Boost	61.78%	2.95%	3.35%
Support Vector Machines + XG Boost	63.65%	6.24%	6.88%
All Three	63.77%	5.13%	6.21%