Generating Betting Insight on NBA Player Performance using Machine Learning

Group 42: Sahil Bishnoi, Josh Garretson, Avery Girsky, Oliver Hewett, Hardik Patel, Atticus Rex

Overview

We present an interactive dashboard to inform statistically optimal basketball betting strategy and describe trends in player-level data. We hope the use of interactive visualization around ML prediction should help to distill advanced data science concepts for everyone to receive data-driven advice to assist their intuition.

- **Modeling** To identify profitable bets given historical professional basketball data, our team trained a model outputting a probability distribution of the five main player stats: Points (PTS), Assists (AST), Rebounds (REB), Blocks (BLK), Steals (STL).
- **Visualization** These insights are displayed in a digestible and interactive formats allowing for user inputs to observe probability distributions for model

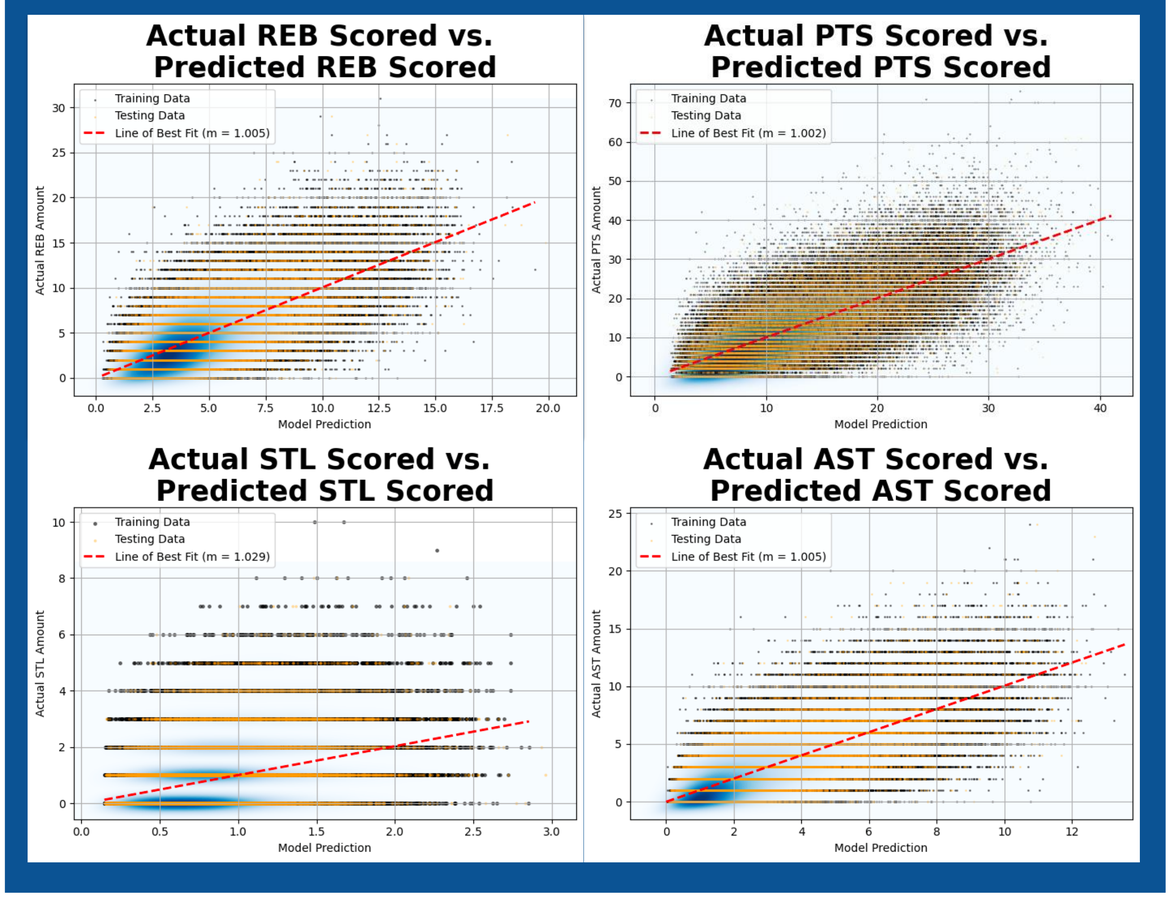
Modeling

- **Data Summary:**
  - 142,320 Observations
  - All Active NBA Players and Regular Season Games since 1995
- **Models Evaluated:**
  - Ridge/Lasso Regression
  - Neural Networks
  - K-Nearest Neighbors Regression
  - AdaBoost and Random Forest Regression

Test MAE By Model, Statistic	PTS	REB	AST	STL	BLK
Ridge Regression					
Lasso Regression	4.813				
Neural Network					
KNN Regression					
AdaBoost					
Random Forest					

- **Final Model: Lasso Regression (α=9e-3)**
  - Chosen for its accuracy, interpretability, and efficiency both in training and evaluation. Notably avoids overfitting.
- **Features include:**
  - Seasonal Running Average of W/L, MIN, FGM, FGA, FG_PCT, FG3M, FG3A, FG3_PCT, FTM, FTA, FT_PCT, OREB, DREB, REB, AST, STL, BLK, TOV, PF, PTS, PLUS_MINUS
  - Whether or not the player is on the Home Team
  - 5-Game Lookback of Stat of Interest
  - Seasonal Running Average of Team Box Score Statistics

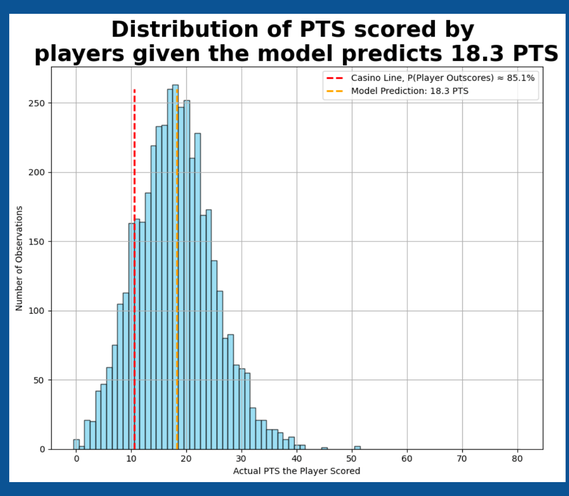
Results

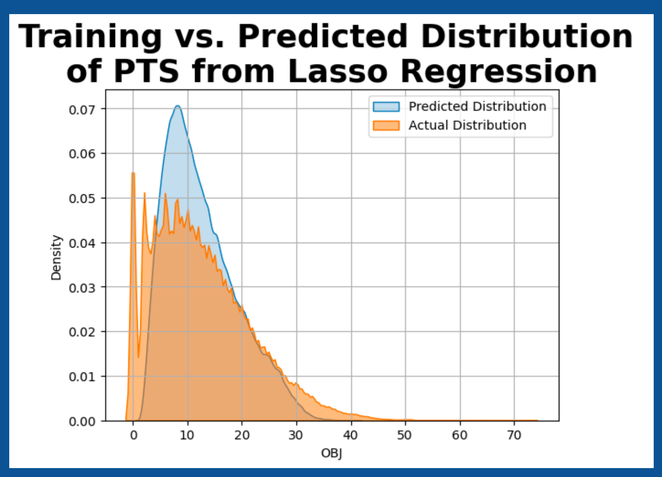


Visualization

Given user-input for a **player** and **statistic**, display two visuals:

- Time-series plot of the player's statistical output over their last ten games combined with the projection and sportsbook line for their upcoming game
- Distribution plot of historical performances amongst all players that had been previously projected the same statistical outcome by our model
  - Vertical lines included for the model projection and the sportsbook line
  - The proportion of the distribution falling above/below the sportsbook line represents the percent "confidence" of the model's recommendation for an over/under bet





Conclusion