A close-up photograph of a person's hands typing on a laptop keyboard. The hands are wearing a brown leather jacket sleeve with visible stitching. The background is dark and out of focus.

*Flatiron Capstone
Project: Modeling an
NBA player's Impact
to Winning*





Hello!

My name is Kiarash Ahmadi. I'm an aspiring Data Scientist with Flatiron Academy. I am also pursuing a Master's in Computational Science & Engineering at Georgia Tech. My field of work is in the building energy modeling industry but todays discussion will be about modeling an NBA player's impact to winning





Project Background

- 🏀 Predicting ESPN's RPM (Real Plus-Minus): an advanced NBA stat that measures the impact a player has on the court
- 🏀 Different from traditional plus-minus because it adjusts for the effects of each teammate and opposing player
- 🏀 Predict RPM using traditional statistics and advanced statistics that only use traditional box-score stats



The King of RPM



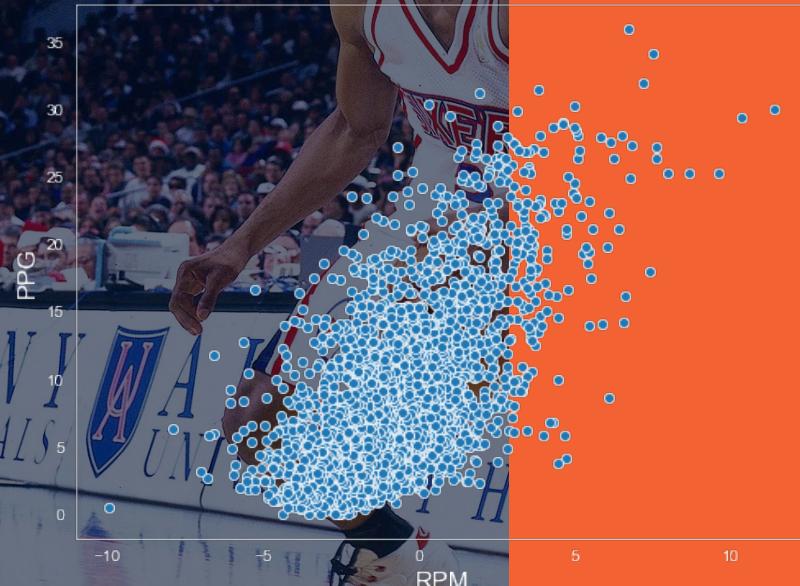
Predicting Player Impact

- How does points-per-game correlate with RPM?
- Hope to address misconception about PPG but also gauge its impact
- What effect do traditional-based advanced statistics have on RPM?
- Advanced stats include Box Plus Minus (BPM) Win Shares (WS) and their dependent statistics:
 - OWS, DWS, WS/48
 - OBPM, DBPM, VORP
- How does the number of 3 pointers made affect RPM?
- Hope to quantify the impact of the 3-point shot via its direct and indirect effects



How does points-per-game correlate with RPM?

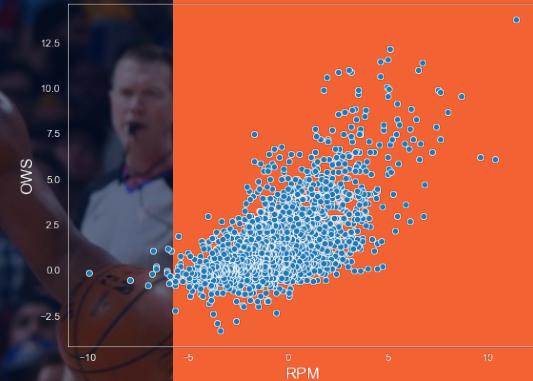
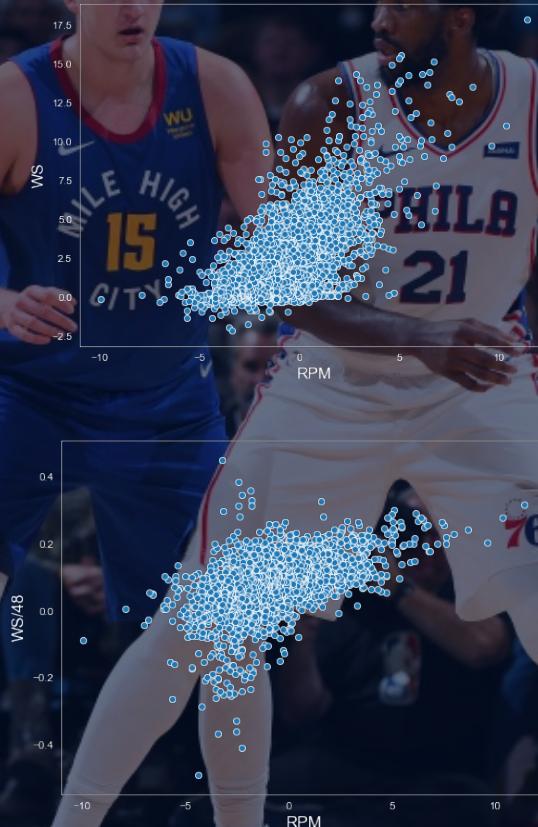
- A strong correlation with points-per-game and RPM is evident
- Pearson Correlation of 0.63
This correlation intuitively makes sense but doesn't necessarily mean that a higher PPG leads to a higher RPM
- Strength of Association based on Pearson Correlation:
 - Small: 0.1 – 0.3
 - Medium: 0.3 – 0.5
 - Large: 0.5 – 1.0





What effect do traditional statistics based advanced statistics have on RPM?

- Win Shares is a player statistic which attempts to divvy up credit for team success to the individuals on the team.
- Pearson Correlations:
 - WS: 0.66
 - WS/48: 0.52
 - DWS: 0.54
 - OWS: 0.64
- Offense is more correlated with RPM than defense





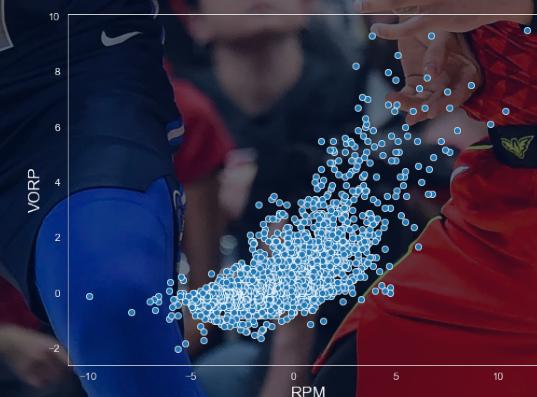
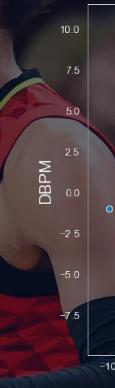
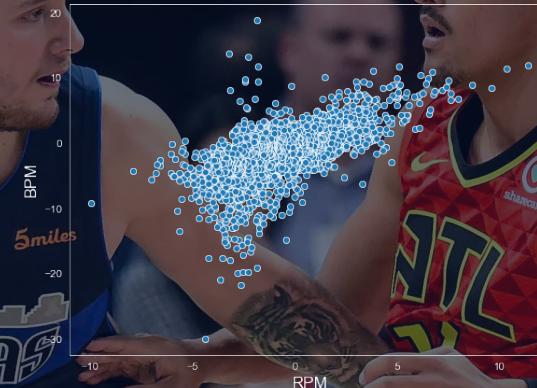
What effect do traditional-based advanced statistics have on RPM?

- 🏀 BPM is a basketball box score-based metric that estimates a basketball player's contribution to the team when that player is on the court

- 🏀 Pearson Correlations:

- 🏀 BPM: 0.63
- 🏀 VORP: 0.70
- 🏀 DBPM: 0.22
- 🏀 O BPM: 0.66

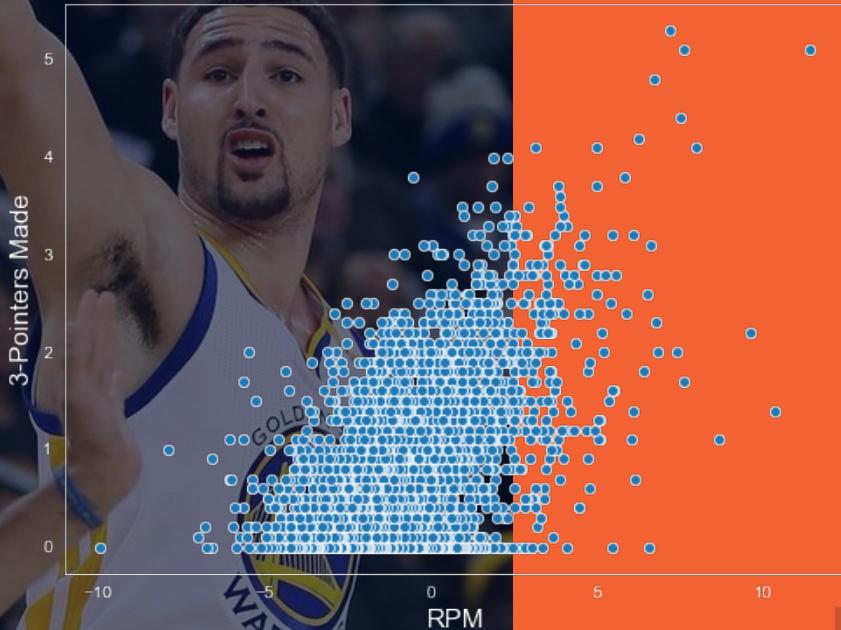
- 🏀 Further indication that offense is more correlated with RPM than defense





How does the number of 3 pointers made affect RPM?

- Done in order to analyze the impact of the 3-point shot on RPM
- Pearson Correlation of 0.54
- Slight correlation between RPM and 3-pointers made but not as much as other metrics we've analyzed



How will we measure model performance?

- 🏀 Mean Squared Error (MSE): Represents the average of the squared difference between the original and predicted values in the data set.
 - 🏀 It measures the variance of the residuals.
- 🏀 Mean Absolute Error (MAE): Represents the average of the absolute difference between the actual and predicted values in the dataset.
 - 🏀 It measures the average of the residuals in the dataset.
- 🏀 Coefficient of determination (R-squared): Represents the proportion of the variance in the dependent variable which is explained by a model



DNY

Best Performing Models

- XGBoost using standard scaled data

- Training Data:

- MSE: 0.791
- MAE: 0.681
- R-Squared: 0.820

- Testing Data

- MSE: 1.640
- MAE: 0.986
- R-Squared: 0.633

- Gradient Boosting using standard scaled data

- Training Data:

- MSE: 1.089
- MAE: 0.814
- R-Squared: 0.752

- Testing Data

- MSE: 1.661
- MAE: 0.993
- R-Squared: 0.627

- Gradient Boosting using non-scaled data

- Training Data:

- MSE: 1.013
- MAE: 0.785
- R-Squared: 0.770

- Testing Data

- MSE: 1.663
- MAE: 0.998
- R-Squared: 0.626



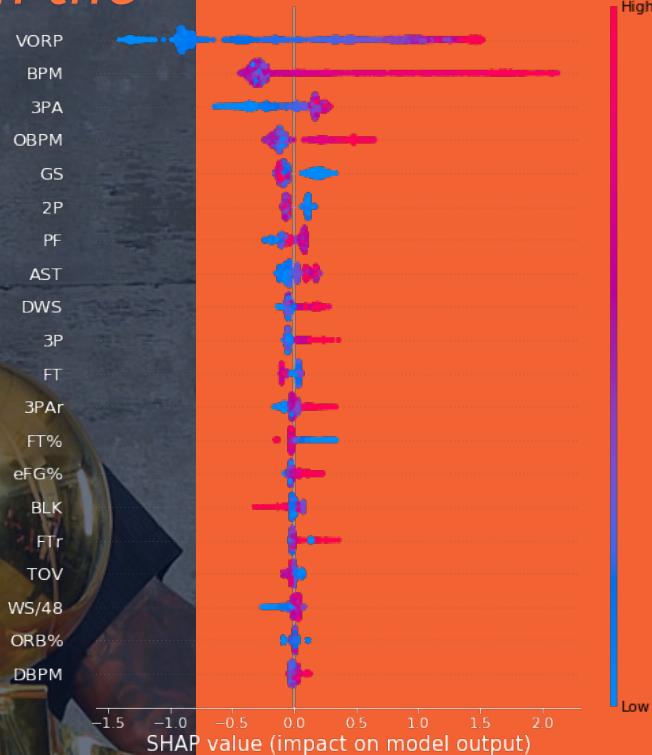
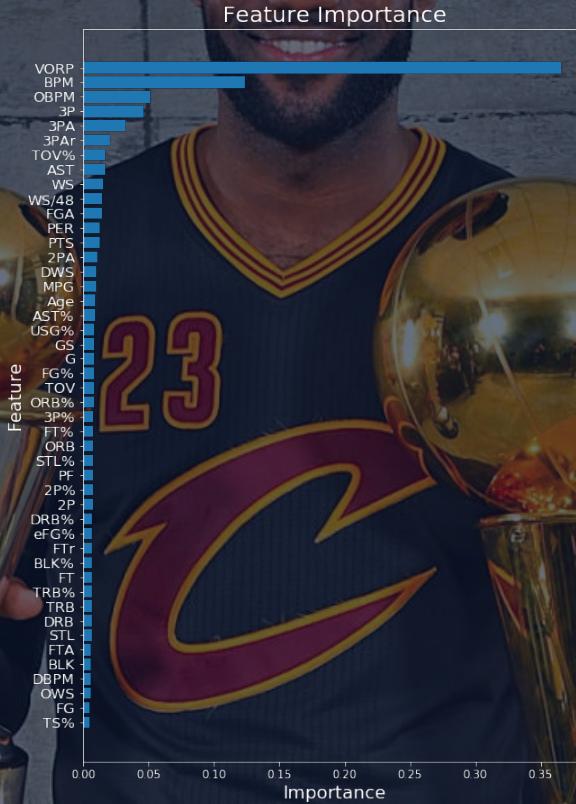
What Features Impact RPM Prediction the Most?

- Advanced statistics, 3-point shooting statistics, and offensive statistics have greatest impact in predicting RPM

- Lower values of 3-pointers attempted lead to a lower prediction whereas higher values lead to a higher RPM prediction

- Higher values of 2 pointers made leads to a lower RPM prediction

- Defensive statistics have less of an impact in predicting RPM





Final Insights



Offense Reigns Supreme

- Often the case where the more influential and important features in predicting RPM were related to offense.
- Helps to confirm eye test of offense having more of an impact to winning than defense



The 3-point Revolution

- Feature importance confirmed the direct and indirect impact of the 3-pointer
- Low volume 3-point shooting is detrimental to winning
- The 3-point shot leads to more gravity and spacing which allows for easier offense



Miscellaneous Insights

- Turnover Percentage is detrimental to the impact of a player on winning
- Points per game is less impactful in predicting RPM than initially hypothesized
- Traditional box-score based advanced statistics such as BPM and VORP have a significant impact in predicting RPM



Future Work

- Using RPM from past years to predict current RPM
- Introducing non-traditional statistics(particularly defensive stats) to see if it improves results
- Focusing on ORPM and the features that affect it the most
- Looking into how different team statistics affect winning percentage over the years as the sport has changed
- More feature selection
- Drop dependent features such as 3PA or 3P
- Try more neural network tuning techniques and architectures



Thanks!

Any questions?

You can find me at:

🏀 mkiah09@gmail.com

