NCAA MEN'S BASKETBALL TEAM SCORE PREDICTION

Light Gradient Boosted Model Prediction; MSBA Capstone Project Spring 2022

Hettinger, Liam

University of Montana | 3-31-2022

Table of Contents

Executive Summary:

During each tournament, Men's NCAA Division 1 basketball fans will fill out a bracket and attempt to pick the winner of each game in the NCAA tournament. In 2022, no one in the nation has ever had a perfect bracket. The American Gaming Association estimates roughly 70 million brackets are completed last year. (Browne)

Statistical models in college sports are popular and necessary for performance analysis. With 358 different universities, most universities will not play each other over the season. Therefore, determining the best team is not as easy as looking at their overall record. Statistical models often provide the most reliable ranking.

The project's goal is to determine if a statistical model is better than intuition at predicting NCAA Division 1 Basketball March Madness scores. Before each round of the 2022 March Madness tournament, I intuitively guessed March Madness team scores using my previous basketball knowledge. The intuitive predictions were then compared to a statistical model to determine accuracy.

The statistical model used was a light gradient boosted model. The features included historical averages of advanced and basic box score statistics for the given team, the given team's opponent, and the given team's previous opponents. The model was trained on NCAA Division 1 regular season and post-season data from 2015 to 2022, excluding the 2022 post-season tournament.

Ultimately, the light gradient boosted model outperformed my intuition. My intuition's mean absolute percentage error was 18.95 percent. The mean absolute error for the light gradient boosted model was 14.9 percent. The model predicted games with over 5 percent greater accuracy.

Introduction:

On March 17th, 2022, I sat in front of my television, awaiting my favorite time of the year. This day marks the start of the 2022 Men's NCAA Division 1 basketball March Madness tournament. March Madness decides the Men's NCAA Division 1 basketball champion, giving the school bragging rights for the following year.

Along with 70 million others worldwide, I filled out a March Madness bracket to predict each March Madness game. March Madness consists of 63 games and 64 teams. Predicting 63 games correctly is a nearly impossible task. Assuming each team has an equal chance of winning, the chances of predicting each game's winner is 0.00000000000000000010842022. (Wilco)

Equation 1:

$$\frac{1}{2^{32} * 2^{16} * 2^{8} * 2^{4} * 2^{2} * 2} = 0.00000000000000000010842022$$

Although I consider myself an avid college basketball fan, my prediction accuracy is likely not much better than random guessing. With over 11,000 games in 2022, it was nearly impossible for anyone to watch every Division 1 regular-season game. With player injuries, the ups and downs of college kids' play, and teams going on a "Cinderella" run, it is hard to understand a team's actual ability. Knowing this, I believed my prediction's accuracy could not be much better than random guessing. This reveals the problem I hope to tackle, predicting NCAA basketball games is nearly impossible. Could I improve my chances of predicting games using a statistical or machine learning model?

Literary Review:

Statistical models are widely used throughout college basketball. Some use them to predict scores and rank teams, such as Evan Miya and Ken Pomeroy. Others strictly are used to predict games, such as those competing in March Machine Learning Mania and Blake Atkinson.

KenPom, a college basketball statistical site created by Ken Pomeroy, is considered by most the leader in statistical college basketball analysis. Kenpom.com uses offensive efficiency, the average number of points a team scores over 100 possessions, defensive efficiency, the average number of points an opponent scores over 100 possessions, and adjusted tempo, the average number of possessions a team has in a game, to predict game outcomes. He also worked with Bill James to create a log5 projection equation that determines a team's likelihood of winning. (Pomeroy)

Most college basketball models are designed to give probabilities that a team will win. Kaggle has popularized this trend with their March Machine Learning Mania competition. In 2022, the three visible codes from the top 5 places all used a boosting model to classify winners and losers for March Madness 2022. Competitors are scored on the log loss to determine who wins one of the $5,000 prizes. (March Machine Learning Mania 2022)

Other models do predict team and player impact directly. For example, Evan Miya, creator of evanmiya.com, a college basketball statistical prediction site, created a Bayesian Performance Rating that determines how effective a team or player is based on Bayesian statistics. Evan Miya's team Bayesian Performance Rating uses a team's true offensive and defensive ratings, four-factor ratings, historical team information, home court advantages, the pace of play, and other predictive statistics to determine a team's performance. The team performance rankings can be used to predict team scores for a given game. (Miya)

Most statistical model features are proprietary. However, Blake Atkinson, a Data Science/Front End Dev that writes on towardsdatascience.com, published an article outlining his creation of a college basketball model. The model used historical data to predict college basketball scores. According to his feature importance method, leave one feature out, home-court advantage had a significant influence on the model. This model is likely intended for regular season predictions since March Madness games are played on neutral courts. (Atkinson)

There has been a variety of statistical models used to identify trends in college basketball. Most can be used for prediction but do not disclose their prediction algorithms completely. The most popular disclosed prediction model type designed for March Madness is a boosted model. XG Boost, Light Gradient Boost, and Cat Boost were the models used by 3 of the top 5 winners in March Machine Learning Mania.

Research Questions:

The model's goal is to predict the Men's NCAA Division 1 basketball team's score using historical averages from the given team, the opponent they are playing against, and the opponents they previously played against. The model will output a predicted score for the given team. After feature selection, the final model must satisfy two parameters:

1. Accuracy
2. Reducing the number of features without sacrificing significant accuracy

Accuracy is essential in prediction. A more accurate model is more successful at predicting a team's score. The end goal is to predict March Madness game accurately. However, reducing features helps generalize the model and avoids overfitting.

An overfitted model predicts residual variation within the training set. Residual variation is random noise within the training set. Although a large sample size reduces the chance of overfitting, ensuring overfitting does not occur is important. Overfitting will cause the model to predict random noise within the data set.

In addition to avoiding overfitting, the model is intended for future use. Fewer features make for a more manageable data collection process. Data collection is automatically done. However, data collection is done through a web scrapping API that needs to be updated after SportReference.com changes its website design.

Feature selection is necessary to prevent overfitting and to ensure more manageable data collection. However, accuracy is still the paramount goal. Therefore, features will only be eliminated if they have less than a .0005 or 0.05 percent effect on mean absolute percentage error.

Once a final model is chosen, this model will be compared to my March Madness bracket based on intuition. Before each round begins, I will intuitively pick the score for each March Madness game. Predictions will be based on previous games watched, analysis from podcasts such as Titus and Tate, articles read, and general knowledge of college basketball.

This will help me answer my research question:

Is a machine learning/ light gradient boosted regression model using historical box scores better than intuition at predicting NCAA Men's Basketball games?

Hypothesis:

Experts commonly believe that data-driven predictions are more effective than intuition-based decision-making. Despite this, I have yet to meet someone that uses a model to fill out their March Madness bracket. Typically, fans will place bets or fill their March Madness bracket out based solely on intuition.

However, intuition is lacking in a variety of ways. It is typically based on less data, places heavier weights on more recent events, and/or is sometimes based on a non-empirical feeling. These biases typically make intuition less accurate than a model. A model is only based on empirical data and weighs all the events the same.

Due to intuition typically being less accurate than a model, I hypothesize that intuitively guessing which teams will win in March Madness will be less accurate than a light gradient boosted model.

Model:

The model I have selected is a light gradient boosted regression model. This model was chosen due to its popularity in the Kaggle's March Machine Learning Mania. The 2022 Kaggle March Machine Learning Mania winner used a gradient boosted model, similar to the light gradient boosted model. The Kaggle's March Machine Learning Mania winning model is the model with the most accurate model based on a log-loss accuracy score.

The light gradient boosted regression model was created by a team at Microsoft in 2016. The model stands out due to its accuracy while also having fast training speeds and low computational power. However, the model can only be used effectively on a large dataset due to down-sampling methods within the algorithm. Down-sampling is a method to reduce the data's complexity.

Due to the complexity of this model, it is necessary to review how this model works. The light gradient boosted regression model uses an ensemble of weak prediction models to make a prediction. These weak prediction models are like a slightly educated guess. To conceptualize this, I like to imagine a food critic scenario. Like linear regression, strong predicters use one prediction model to form the results. This is like a food critic writing a review about his experience at a restaurant. This food critic understands food more than an average person, but the rating is only based on one opinion. An ensemble of weak prediction models is like a thousand average foodies forming a thousand opinions.

These thousand opinions would be aggregated together in a random forest to form a prediction. This method is called bagging. Light gradient boosting models use an ensemble learning method called boosting. Instead of aggregating all weak model results, the results work together to form a prediction. Boosting works like this: the first tree is calculated, and the following trees are used to correct the error in the previous tree's predictions. Each additional tree learns from the previous tree's error.

Since each tree is learning from the previous tree, each tree must be created in order. This can lead to prolonged training times. To solve this issue, Light gradient boosting uses a histogram-based splitting instead of the more common pre-sorting method. Pre-sorting evaluates every possible split in the data and chooses the best possible split for each node. Histogram-based splitting splits continuous samples into discrete bins based on the samples' impact on the model. The bins reduce the number of samples evaluated and, therefore, reduce training time.
Light gradient boosted models evaluate bins using gradient-based one-side sampling. Gradients are used to split node results into bins. The gradient is the error in the sample. A large gradient sample contains more unique information than a small gradient.

Gradient-based one-side sampling works like this. The first tree uses all data and learns how to fit the target variable. The second tree uses gradients from the previous tree prediction. The highest gradients are placed in a bin. The remaining low gradient samples are placed in a bin. The following tree uses the higher gradient bin and a random sample of the lower gradient bin. The smaller gradient randomly sampled values are weighted using the constant (1 – top gradient group) / lower gradient group. Weighting is done to conserve the data's distribution. This group is evaluated and used to train the next tree.

Exclusive feature binding is another down-sampling method used in the light gradient boosted algorithm. Exclusive feature binding is the process of combining mutually exclusive variables into one variable. This light gradient boosted model will not contain any mutually exclusive variables.

Light gradient boosting will continue improving results until its evaluation metric stops improving. The evaluation metric used for this light gradient boosted model is L1, based on absolute loss. The light gradient model will stop after L1 stops improving for ten interactions. L1 is used to prevent overfitting.

Dataset:

All desired data was accessible on Sportsreference.com. The data was acquired using an API called Sportsipy. Sportisipy is a web scraping API created by Robert Clark. The dataset contains 40375 games from 2015 to 2022. This period was selected because the NCAA changed the shot clock length from 35 seconds to 30 seconds before the 2015-16 NCAA Men's basketball season. This change increased the speed of the game. Therefore, the box score data before 2015 was not used in the model.

Table 1 shows all variables that acquired through the Sportsipy API:

Metric	Definition
Assist Percentage	An estimate of the percentage of teammate field goals a player assisted while he was on the floor.
Assists	An estimate of teammate field goals a player assisted while he was on the floor.

Block Percentage	An estimate of the percentage of opponent two-point field goal attempts blocked by the player while he was on the floor.
Blocks	An estimate of opponent two-point field goal attempts blocked by the player while he was on the floor.
Defensive Rating	Points allowed per 100 possessions
Defensive Rebound Percentage	An estimate of the percentage of available defensive rebounds a team grabbed.
Defensive Rebounds	The number of defensive rebounds by a team.
Effective Field Goal Percentage	The formula is (Field Goal + 0.5 * 3-Point) / Field Goal Attempts. This statistic adjusts for the fact that a 3-point field goal is worth one more point than a 2-point field goal.
Field Goal Attempts	Number of baskets scored includes both 2-point field goal attempts and 3-point field goal attempts
Field Goal Percentage	The formula is Field Goals / Field Goal Attempts.
Field Goals	The number of baskets scored excluding free throws.
Free Throw Attempt Rate	the ratio of Free Throw Attempts to Field Goal Attempts
Free Throw Attempts	The number of free throws attempted for a team
Free Throw Percentage	The formula is Free Throws/ Free Throw Attempts
Free Throws	The number of made free throws for a team
Offensive Rating	Offensive points produced per 100 possessions
Offensive Rebound Percentage	An estimate of the percentage of available offensive rebounds a player grabbed while he was on the floor.
Offensive Rebounds	The number of offensive rebounds for a team
Personal Fouls	The number of team personal fouls
Points	The number of points a team scores
Steal Percentage	an estimate of the percentage of opponent possessions that end with a steal by the player while he was on the floor
Steals	The number of steals for a team
Three Point Attempt Rate	Percentage of total field goal attempts from three points.
Three Point Field Goal Attempts	The number of attempted three point shots
Three Point Field Goal Percentage	The formula is Three Point Field Goals/ Three Point Field Goal Attempts
Three Point Field Goals	The number of made three point shots

Total Rebound Percentage	an estimate of the percentage of available rebounds a player grabbed while he was on the floor.
Total Rebounds	The total number of rebounds for a team. The formula is Offensive Rebounds +
True Shooting Percentage	The formula is Points / (2 * True Shooting Attempts). This measure of shooting efficiency that takes into account field goals, 3-point field goals, and free throws.
Turnover Percentage	An estimate of turnovers per 100 plays
Turnovers	The number of turnovers for a team
Two Point Field Goal Attempts	The number of two point shots attempted
Two Point Field Goal Percentage	The formula is Two Point Field Goals/ Two Point Field Goal Attempts
Two Point Field Goals	The number of two point shots made
Pace	an estimate of the number of possessions per 48 minutes by a team.

Sportispy provided box-score statistics for each division 1 men's basketball team. Per game box scores do not show historical trends. The goal of the model is to predict future games. Without historical values, the model cannot predict future games.

The dataset was altered to provide historical trends. A lag value, a rolling average from the last 2-5 games, and a cumulative average for the season with a 2-game minimum for each metric given in Figure 1. Figure 2 represents the suffixes used to indicate its historical trend, a definition of the trend and an example of how it was used for assists:

Table 2 shows the suffixes used to indicate its historical trend:

Suffix	Definition/ Meaning	Example
_cum_avg	Cumulative average for the season with a 2-game minimum	Assists_cum_avg
_RA_1	The lag value	Assists_RA_1
_RA_2	Rolling average over the past two games	Assists_RA_2
_RA_3	Rolling average over the past three games	Assists_RA_3
_RA_4	Rolling average over the past four games	Assists_RA_4
_RA_5	Rolling average over the past five games	Assists_RA_5

Historical trends were only calculated for each given season. This first game a team plays in a season will not have any historical trends. The second game will only have the lag values from the first game. There will be some null independent variables until the season's sixth game. Observations with null values were dropped instead of another preprocessing method to conserve the model's accuracy.

The current historical averages for the predicted team are biased. Each team's schedule varies in difficulty. Some teams play more talented opponents, making it harder for a team to score points.

One team's historical averages could be based on games with terrible defenses. Another could be based on great defenses. Historical averages of previous opponent's defensive metrics were used to account for the difference in defensive ability between teams. This trend accounts for elevated scores due to playing poor defenses in the past. Figure 3 were the historical trends that were calculated:

Table 3 shows previous opponents metrics used in the model:

Previous Opponents Defensive Metrics
block percentage
blocks
defensive rating
defensive rebound percentage
defensive rebounds
personal fouls
steal percentage
steals
turnover percentage
turnovers

To distinguish these variables, a prefix of prev_opponent_ was added and the same suffixes in Table 2 were used to indicate which historical average is represented.

The predicted team opponent's defensive ability for the predicted game must be accounted for. Again, historical trends were used for the predicted team's opponent. The same suffixes from Figure 2 were used to indicate the historical trends. The prefix opponent_ was used to indicate that the metric is about the opponent in the predicted game. Figure 4 shows historical trends used to calculate opponent defense:

Table 4 shows opponents metrics used in the model:

Opponents Defensive Metrics
block percentage
blocks
defensive rating
defensive rebound percentage
defensive rebounds
personal fouls
steal percentage
steals
turnover percentage
turnovers

pace
points

Once the historical trends were calculated, the original pregame box scores were removed from the dataset, excluding the target variable, points. The final dataset had 344 features in total.

Feature Selection:

There are four feature selection methods used to select model features:
- Correlation
- SHAP (Shapley Additive Explanation)
- Feature Gain
- Feature Splits

The goal of feature selection is to improve the model's simplicity and reduce overfitting in the model. The model is intended to be used in the future. If the model can conserve accuracy without a variable, that variable will be removed.

None of these feature selection methods focus on reducing multicollinearity. Light gradient boosted models are mainly immune to multicollinearity. The creator of extreme boosting decision trees (XGBoost), Tianqi Chen, confirms that boosting algorithms will only use one of the highly correlated variables for trees. However, he does say that reducing multicollinearity in any model is beneficial. (Chen)

The first two methods used are built into the light gradient boost python package. These feature importance metrics are called gain and split. Gain is a measure of the loss functions reduction for each variable. Split or weight measures how often those variables were used in each tree.

Both these methods have flaws that should be noted. Using gain as a feature selection method can be flawed when colinear variables are used. The first colinear variable used, such as a previous field goal percentage, will reduce the error in the model. Other multicollinear variables such as the field goal percentage moving average on the previous 2, 3, 4, or 5 games all add less information to the model if they are used in later nodes in the model. The model accuracy will improve significantly for the first variable used from the set. Therefore, the gain for this first multicollinear feature will be higher than proceeding multicollinear variables. Gain will be used as a feature selection method, but its limitations will be acknowledged.

Split is another method that can be helpful as a guide to which features are essential to the model. However, it should not be used alone. Split indicates how many times a variable is used in the model. If a variable is not used in the model, it does not affect the model. However, a variable with a high number of splits does not necessarily mean this variable improves the model's accuracy more than another variable. Splits can be a helpful variable selection method but must be used alongside others.

Split and gain will be used to eliminate unused variables in the model for future feature selection analysis. There is no use in analyzing variables that the model does not use. The entire gain and slip plots can be found in the appendix. One hundred sixty-six variables were not used in any splits and, therefore, had no gain. These variables will be removed immediately from the model since they are not used. Removing these variables does not impact the model's accuracy. It does improve model training time slightly because the model has fewer features.

Figure 5 shows the top 20 performers for gain were. The features in yellow appeared in the top 20 for both split and gain. The Red features were not included in the top 20 for both split and gain. The red features will be further analyzed to insure they have impact in the model. The yellow features likely have an impact on the model.

Table 5 shows the top 20 gain features:

Value	Feature
4296617.67	points_cum_avg
2173834.08	opponent_defensive_rating_cum_avg
2100617.17	opponent_pace_cum_avg
334002.431	points_RA_5
261288.869	opponent_defensive_rebound_percentage_cum_avg
140805.841	opponent_pace_RA_4
118561.298	prev_opponent_defensive_rating_cum_avg
100159.7	opponent_pace_RA_5
79042.869	opponent_block_percentage_cum_avg
74263.4102	opponent_points_cum_avg
69632.3295	turnover_percentage_cum_avg
54646.2799	blocks_cum_avg
54179.55	opponent_pace_RA_2
50538.1406	points_RA_4
40841.8805	opponent_defensive_rating_RA_5
33154.8596	assists_cum_avg
30920.3198	opponent_pace_RA_3
29429.9597	opponent_turnovers_cum_avg
25583.8593	field_goals_RA_5
24295.6798	defensive_rating_cum_avg

Table 6 shows the top 20 split features:

Value	Feature
73	opponent_pace_cum_avg
71	points_cum_avg
66	opponent_defensive_rating_cum_avg

36	opponent_defensive_rebound_percentage_cum_avg
20	opponent_points_cum_avg
17	prev_opponent_defensive_rating_cum_avg
17	blocks_cum_avg
16	turnover_percentage_cum_avg
16	opponent_block_percentage_cum_avg
14	opponent_pace_RA_5
13	personal_fouls_cum_avg
13	opponent_pace_RA_2
12	opponent_pace_RA_4
10	pace_RA_2
9	steals_cum_avg
9	points_RA_5
9	opponent_turnovers_cum_avg
9	opponent_pace_RA_1
9	opponent_defensive_rating_RA_5
9	field_goal_attempts_RA_1

The following feature selection method used is called SHAP. SHAP or Shapley Additive Explanation was used as the third variable selection technique. Lundberg and Lee developed the SHAP algorithm. Many consider it to be the best way to understand variable importance in a machine learning model. The algorithm uses a game theory approach and Shapley values to reproduce outcomes while gaming the features. The higher the SHAP value is, the more influential the feature is to the model.

Table 7 shows the top 20 SHAP features:

Feature	Shap
points_cum_avg	2.44890029
opponent_defensive_rating_cum_avg	1.78611381
opponent_pace_cum_avg	1.62587683
opponent_defensive_rebound_percentage_cum_avg	0.60545073
points_RA_5	0.37135816
prev_opponent_defensive_rating_cum_avg	0.25807374
opponent_block_percentage_cum_avg	0.25203186
opponent_points_cum_avg	0.19240144
opponent_pace_RA_4	0.18910141
turnover_percentage_cum_avg	0.18809822
blocks_cum_avg	0.16939769
defensive_rating_cum_avg	0.14163948
opponent_pace_RA_5	0.14087394

opponent_pace_RA_2	0.14061166
free_throws_cum_avg	0.10566722
personal_fouls_cum_avg	0.10332049
opponent_defensive_rating_RA_5	0.09740856
opponent_steal_percentage_RA_4	0.09703637
pace_RA_1	0.08971892
free_throw_percentage_cum_avg	0.08835068

There is a noticeable drop-off in the usefulness of features. After the first three features, points_cum_avg, opponent_pace_cum_avg, and opponent_defensive_rating_cum_avg, the Shap value falls below one. This means three main features drive this model.

Lastly, correlations between the feature and the target were used to help select variables. Correlation between variables is the classic approach to feature selection. However, using correlation values alone was difficult. Most features had very weak correlations to the target. Due to the small differences in correlation, other feature selection techniques were necessary.

Table 8 shows are the top 20 Pearson correlation coefficients between each feature and the target:

Features	Pearson Correlation Coefficient
points_cum_avg	0.35856018
field_goals_cum_avg	0.32835956
points_RA_5	0.3181103
points_RA_4	0.3025424
opponent_pace_cum_avg	0.29007858
field_goals_RA_5	0.286802
points_RA_3	0.27792767
offensive_rating_cum_avg	0.27647353
field_goals_RA_4	0.27146455
prev_opponent_defensive_rating_cum_avg	0.26758937
opponent_pace_RA_5	0.26459776
opponent_pace_RA_4	0.256777
field_goal_attempts_cum_avg	0.25625111
opponent_defensive_rating_cum_avg	0.25163324
points_RA_2	0.24907028
pace_cum_avg	0.24758206
field_goals_RA_3	0.24721312
opponent_pace_RA_3	0.245226
assists_cum_avg	0.2443845
true_shooting_percentage_cum_avg	0.23979985

Points_cum_avg has the highest correlation with the points scored by a team, which logically makes sense. The cumulative average points a team scores will likely indicate how well that team does. However, the correlation with points for a given game is low. No features had a significant negative correlation to points scored. For example, I would have assumed that turnovers would have a high negative correlation with points scored by a team, but turnover_percentage_cum_avg was negatively correlated with a Pearson's correlation coefficient of -0.159171.

Arbitrary rules were used to choose the final variables in the light gradient boosted model. First, all variables that had no gain or split were removed. Next, all variables with a SHAP value below .1 were be removed. Finally, all features with Pearson's correlation coefficients below .1 were removed leaving 28 features.

Model Evaluation:

The model was evaluated against a separate test set of data. The test set of data is 10 percent of the total data. This test set will have 4,038 game outcomes. The other 90 percent of the data will be used to train the model. The training set is made up of 36337.

The test set was used to evaluate the model's accuracy while training the model. An evaluation set will be used to evaluate the final model's result. The final model's accuracy measurements will be compared to my March Madness intuitive predictions.

Loss Function:

Models will be evaluated against a MAPE, mean absolute percentage error. MAPE is a widely popular accuracy measurement that measures the absolute percent difference between the actual result and the predicted result. The predicted results are the model's predicted score for the test set. The actual result is the actual score in the test set.

Equation 2 shows the mean absolute percentage error equation:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

M is the mean absolute percentage error. At is the actual value. Ft is the forecasted value. N is the number of predictions made.

Model evaluation after feature selection:

The goal of feature selection is to improve training time, improve accuracy and lessen data collection without sacrificing accuracy. The original model with all features had a MAPE score of 12.351910 percent with 342 features.

After removing the features not used in the model, the model had the same MAPE score with just 193 features.

After removing features with a SHAP value below .1 and features not used in the model, the MAPE score was reduced by only .0000179 but only had 16 columns. The MAPE score for the model post SHAP feature selection was 12.350114 compared to the entire model, which was 12.351910. The model's accuracy was slightly improved. Likely, the features that were added in the previous model introduced data noise. Once this noise was eliminated, the model was more accurate on future predictions.

The last feature selection process did hinder the models MAPE score. The last feature selection technique was correlation analysis. Variables from the remaining 16 variables with a Pearson's Correlation coefficient above .1 were kept. This reduced the features from 16 variable to 10 variable but hindered the MAPE score. The MAPE score after these variables was 12.406478 percent compared to 12.350114 after the SHAP feature selection step.

Table 9 shows the MAPE after each feature selection round:

Model	MAPE	Terms	Variable counts
Model 1	0.12351910	All independent variables	342
Model 2	0.12351910	Removed variables not used	193
Model 3	0.12350114	Removed variables with SHAP value less than .1	16
Model 4	0.12406478	Removed variables with Pearson Correlation Coefficients less than .1	10

The final model is Model 3. Model 4's reduction of MAPE was too large of a decrease to use this model as the final model. Model 3 does not lose accuracy and uses 16 terms, which is manageable for data collection. The terms in model three can be found in table 10 below.

Table 10 show the features in model 3:

Model 3 Features

‘points_cum_avg'
'opponent_defensive_rating_cum_avg'
'opponent_pace_cum_avg'
'opponent_defensive_rebound_percentage_cum_avg'
'points_RA_5'
'turnover_percentage_cum_avg',
'opponent_block_percentage_cum_avg',
'opponent_pace_RA_5',
'defensive_rating_cum_avg',
'opponent_pace_RA_4',
'prev_opponent_defensive_rating_cum_avg',
'free_throws_cum_avg',
'opponent_turnovers_cum_avg',
'blocks_cum_avg',
'opponent_pace_RA_1',
'field_goals_RA_5'

Conclusion (Light Gradient Boost Model vs Intuition March Madness):

The light gradient boosted model outperformed my intuition. For each round, excluding the final four round, the light gradient boosted model outperformed my intuitive guess. The most considerable difference in predictions was for the Sweet 16. My intuition was more accurate than the light gradient boosted model for the final four games. However, there are only two games in the final four which is a minimal sample size.

Table 11 shows the MAPE for both the intuitive predictions and the model predictions for March Madness 2022.

	MAPE	
Round	Light GBM Model	Intuitive prediction
Round of 64	16.03%	17.96%
Round of 32	15.06%	18.70%
Sweet 16	9.86%	26.54%
Elite 8	21.37%	35.44%
Final Four	8.07%	7.96%
Championship	4.16%	15.01%
Total	14.9%	18.95%

The light gradient boosted model more accurately predicted March Madness game's average score. The mean for all March Madness games was 68.34. The light gradient boosted model's average prediction was 69.72. My intuition predicted much higher scores, with a mean score of

74.53. This accounts for a large portion of the difference in accuracy between my intuition and the light gradient boosted model. My intuition overestimated the score for many instances while the light gradient boosted model did not.

	Max	Min	Sd	Mean	Median
Light GBM	81.275474	60.880209	3.856769031	69.71629372	69.126
Intuition	101	50	9.679410319	74.53174603	74.5
Actual	97	41	12.15444788	68.34126984	69

My intuition did a better job of predicting the variance within the estimates, though. The standard deviation for the actual March Madness games was 12.15. My intuition had a standard deviation of 9.67, which is much closer to the actual standard deviation of 12.15.

The light gradient boosted model had significantly less variation than my intuition and the actual results. Actual prediction involves noise and random variation. Models are not designed to predict noise in a dataset. Due to the nature of basketball, there is variation that is due to random chance. For example, if someone has a hot hand and makes a higher percentage of field goals, the model cannot account for this random increase in field goal percentages.

The low variance in the light gradient boosted model explains why it did not beat my intuition for the final four. The law of large numbers suggests that the mean for actual scores will approach the true mean as the sample size grows. The final four comprises two games which is a minimal sample size.

Overall, the light gradient boosted model's mean absolute percentage of light gradient models error was 14.9% for march madness games. However, the model had a mean absolute percentage error of 12.35% against a test set of 10% of the entire dataset.

The reason for the difference could be due to smaller sample size. However, it could also be due to the difference between regular-season NCAA Division 1 basketball and NCAA March Madness basketball. NCAA March Madness is different in a variety of ways. Teams typically play away games during March Madness unless they are selected at random to play close to home. In addition, teams do not know whom they are playing next. This gives them less time to prepare. The increased pressure also affects teams. March Madness is televised to millions of people, and it is a loss-and-go-home type tournament. These factors increase pressure for college athletes. Almost always, it is the largest tournament they have ever played in.

Despite the differences between NCAA Division 1 basketball regular season and March Madness, the dataset I used was the best training available. Although there are differences, the teams are playing the same sport, against the same type of opponents, playing by the same rules, and have the same team. Ultimately, these subtle differences are not enough to completely discount this dataset.

In addition, a dataset only consisting of March Madness games from 2015 to 2022 would only consist of 441 observations. This data set is too small to account for the large variation in March Madness games. Including NCAA Division 1 basketball regular season data with past march madness data increases the observation to 67518. This dataset is large enough to account for random variance.

For future predictions, a model is better than intuition. The model is not perfect. However, it is far more accurate than intuitively predicting games. As sports betting continues to grow, I would not suggest using this model to predict games. Although if you are going to bet on college basketball a model will provide more accurate estimates than intuition.

Appendix:

*All code for modeling and cleaning data can be found on https://github.com/Lhett2626/Capstone

Split:

Total list of split values:

73	opponent_pace_cum_avg
71	points_cum_avg
66	opponent_defensive_rating_cum_avg
36	opponent_defensive_rebound_percentage_cum_avg
20	opponent_points_cum_avg
17	prev_opponent_defensive_rating_cum_avg
17	blocks_cum_avg
16	turnover_percentage_cum_avg
16	opponent_block_percentage_cum_avg
14	opponent_pace_RA_5
13	personal_fouls_cum_avg
13	opponent_pace_RA_2
12	opponent_pace_RA_4
10	pace_RA_2
9	steals_cum_avg
9	points_RA_5
9	opponent_turnovers_cum_avg
9	opponent_pace_RA_1
9	opponent_defensive_rating_RA_5
9	field_goal_attempts_RA_1
9	assist_percentage_cum_avg
8	opponent_steal_percentage_RA_4

8	opponent_block_percentage_RA_1
8	defensive_rating_cum_avg
7	pace_cum_avg
7	pace_RA_1
7	opponent_pace_RA_3
7	opponent_block_percentage_RA_4
7	opponent_
7	free_throws_cum_avg
7	free_throw_percentage_cum_avg
7	assists_cum_avg
6	three_point_field_goal_percentage_RA_3
6	defensive_rebounds_cum_avg
5	prev_opponent_block_percentage_RA_2
5	opponent_points_RA_4
5	opponent_block_percentage_RA_5
5	assist_percentage_RA_2
4	two_point_field_goal_percentage_RA_1
4	total_rebounds_RA_5
4	three_point_field_goal_attempts_RA_1
4	points_RA_4
4	pace_RA_4
4	opponent_steal_percentage_RA_5
4	opponent_steal_percentage_RA_3
4	opponent_steal_percentage_RA_2
4	opponent_points_RA_5
4	opponent_defensive_rebounds_cum_avg
4	opponent_defensive_rebounds_RA_4
4	opponent_defensive_rebounds_RA_3
4	opponent_defensive_rebound_percentage_RA_3
4	opponent_blocks_cum_avg
4	effective_field_goal_percentage_RA_4
3	two_point_field_goal_percentage_cum_avg
3	three_point_field_goal_percentage_cum_avg
3	three_point_field_goal_attempts_cum_avg
3	three_point_field_goal_attempts_RA_5
3	three_point_field_goal_attempts_RA_4
3	three_point_attempt_rate_RA_2

3	steal_percentage_RA_4
3	prev_opponent_block_percentage_cum_avg
3	opponent_turnover_percentage_cum_avg
3	opponent_steal_percentage_cum_avg
3	opponent_defensive_rebounds_RA_2
3	offensive_rebound_percentage_RA_1
3	offensive_rating_RA_5
3	free_throw_percentage_RA_3
3	free_throw_attempt_rate_cum_avg
3	field_goals_RA_5
3	defensive_rebounds_RA_4
3	defensive_rebounds_RA_2
3	defensive_rating_RA_4
3	defensive_rating_RA_2
3	block_percentage_RA_1
2	two_point_field_goals_cum_avg
2	two_point_field_goals_RA_4
2	two_point_field_goal_percentage_RA_3
2	two_point_field_goal_attempts_cum_avg
2	turnover_percentage_RA_5
2	turnover_percentage_RA_3
2	true_shooting_percentage_cum_avg
2	true_shooting_percentage_RA_1
2	total_rebounds_RA_4
2	total_rebounds_RA_3
2	three_point_field_goal_percentage_RA_4
2	three_point_field_goal_attempts_RA_3
2	three_point_attempt_rate_RA_5
2	three_point_attempt_rate_RA_3
2	prev_opponent_defensive_rebounds_cum_avg
2	prev_opponent_defensive_rebounds_RA_2
2	personal_fouls_RA_5
2	pace_RA_3
2	opponent_turnover_percentage_RA_3
2	opponent_steals_RA_5
2	opponent_steal_percentage_RA_1
2	opponent_points_RA_3

2	opponent_points_RA_2
2	opponent_personal_fouls_cum_avg
2	opponent_personal_fouls_RA_5
2	opponent_personal_fouls_RA_1
2	opponent_defensive_rebounds_RA_5
2	opponent_defensive_rebound_percentage_RA_1
2	opponent_defensive_rating_RA_1
2	opponent_blocks_RA_4
2	opponent_block_percentage_RA_2
2	offensive_rating_cum_avg
2	offensive_rating_RA_1
2	free_throws_RA_3
2	free_throw_percentage_RA_5
2	free_throw_percentage_RA_2
2	free_throw_percentage_RA_1
2	free_throw_attempt_rate_RA_1
2	field_goals_cum_avg
2	field_goals_RA_4
2	field_goal_attempts_cum_avg
2	field_goal_attempts_RA_3
2	defensive_rebounds_RA_5
2	defensive_rating_RA_5
2	defensive_rating_RA_1
2	assists_RA_4
2	assists_RA_1
1	two_point_field_goals_RA_5
1	two_point_field_goals_RA_2
1	two_point_field_goals_RA_1
1	two_point_field_goal_percentage_RA_2
1	two_point_field_goal_attempts_RA_5
1	two_point_field_goal_attempts_RA_3
1	two_point_field_goal_attempts_RA_2
1	turnovers_RA_3
1	turnovers_RA_2
1	turnover_percentage_RA_4
1	true_shooting_percentage_RA_3
1	total_rebounds_RA_2

1	total_rebounds_RA_1
1	total_rebound_percentage_RA_3
1	total_rebound_percentage_RA_2
1	three_point_field_goals_RA_5
1	three_point_field_goal_percentage_RA_2
1	three_point_field_goal_percentage_RA_1
1	three_point_attempt_rate_RA_4
1	three_point_attempt_rate_RA_1
1	steals_RA_2
1	steal_percentage_cum_avg
1	steal_percentage_RA_3
1	prev_opponent_defensive_rebounds_RA_5
1	prev_opponent_blocks_RA_5
1	prev_opponent_blocks_RA_4
1	prev_opponent_blocks_RA_3
1	points_RA_3
1	points_RA_2
1	points_RA_1
1	personal_fouls_RA_3
1	personal_fouls_RA_1
1	pace_RA_5
1	opponent_turnover_percentage_RA_5
1	opponent_turnover_percentage_RA_2
1	opponent_defensive_rebounds_RA_1
1	opponent_defensive_rebound_percentage_RA_2
1	opponent_defensive_rating_RA_4
1	opponent_defensive_rating_RA_2
1	opponent_blocks_RA_2
1	opponent_blocks_RA_1
1	offensive_rebounds_cum_avg
1	offensive_rebounds_RA_2
1	offensive_rebound_percentage_cum_avg
1	offensive_rebound_percentage_RA_4
1	offensive_rebound_percentage_RA_3
1	offensive_rebound_percentage_RA_2
1	free_throw_attempts_RA_5
1	free_throw_attempts_RA_4

1	free_throw_attempts_RA_3
1	free_throw_attempts_RA_2
1	free_throw_attempt_rate_RA_4
1	free_throw_attempt_rate_RA_2
1	field_goals_RA_3
1	field_goals_RA_2
1	field_goals_RA_1
1	field_goal_percentage_cum_avg
1	field_goal_percentage_RA_5
1	field_goal_percentage_RA_3
1	field_goal_attempts_RA_4
1	effective_field_goal_percentage_cum_avg
1	effective_field_goal_percentage_RA_2
1	effective_field_goal_percentage_RA_1
1	defensive_rebounds_RA_1
1	defensive_rebound_percentage_cum_avg
1	defensive_rebound_percentage_RA_4
1	defensive_rebound_percentage_RA_2
1	defensive_rebound_percentage_RA_1
1	assists_RA_5
1	assist_percentage_RA_4
1	assist_percentage_RA_3
1	assist_percentage_RA_1
0	two_point_field_goals_RA_3
0	two_point_field_goal_percentage_RA_5
0	two_point_field_goal_percentage_RA_4
0	two_point_field_goal_attempts_RA_4
0	two_point_field_goal_attempts_RA_1
0	turnovers_cum_avg
0	turnovers_RA_5
0	turnovers_RA_4
0	turnovers_RA_1
0	turnover_percentage_RA_2
0	turnover_percentage_RA_1
0	true_shooting_percentage_RA_5
0	true_shooting_percentage_RA_4
0	true_shooting_percentage_RA_2

0	total_rebounds_cum_avg
0	total_rebound_percentage_cum_avg
0	total_rebound_percentage_RA_5
0	total_rebound_percentage_RA_4
0	total_rebound_percentage_RA_1
0	three_point_field_goals_cum_avg
0	three_point_field_goals_RA_4
0	three_point_field_goals_RA_3
0	three_point_field_goals_RA_2
0	three_point_field_goals_RA_1
0	three_point_field_goal_percentage_RA_5
0	three_point_field_goal_attempts_RA_2
0	three_point_attempt_rate_cum_avg
0	steals_RA_5
0	steals_RA_4
0	steals_RA_3
0	steals_RA_1
0	steal_percentage_RA_5
0	steal_percentage_RA_2
0	steal_percentage_RA_1
0	prev_opponent_turnovers_cum_avg
0	prev_opponent_turnovers_RA_5
0	prev_opponent_turnovers_RA_4
0	prev_opponent_turnovers_RA_3
0	prev_opponent_turnovers_RA_2
0	prev_opponent_turnovers_RA_1
0	prev_opponent_turnover_percentage_cum_avg
0	prev_opponent_turnover_percentage_RA_5
0	prev_opponent_turnover_percentage_RA_4
0	prev_opponent_turnover_percentage_RA_3
0	prev_opponent_turnover_percentage_RA_2
0	prev_opponent_turnover_percentage_RA_1
0	prev_opponent_steals_cum_avg
0	prev_opponent_steals_RA_5
0	prev_opponent_steals_RA_4
0	prev_opponent_steals_RA_3
0	prev_opponent_steals_RA_2

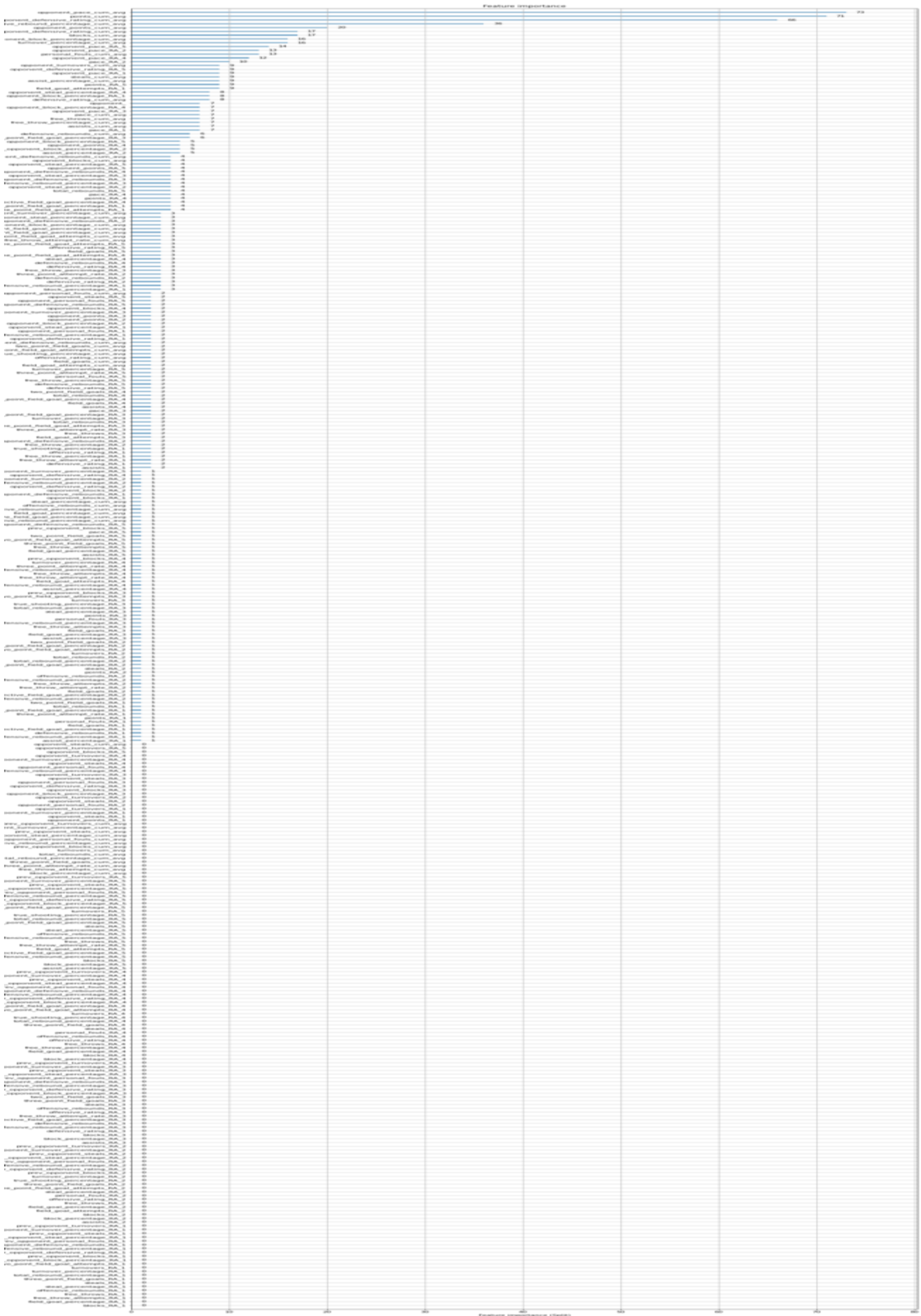
0	prev_opponent_steals_RA_1
0	prev_opponent_steal_percentage_cum_avg
0	prev_opponent_steal_percentage_RA_5
0	prev_opponent_steal_percentage_RA_4
0	prev_opponent_steal_percentage_RA_3
0	prev_opponent_steal_percentage_RA_2
0	prev_opponent_steal_percentage_RA_1
0	prev_opponent_personal_fouls_cum_avg
0	prev_opponent_personal_fouls_RA_5
0	prev_opponent_personal_fouls_RA_4
0	prev_opponent_personal_fouls_RA_3
0	prev_opponent_personal_fouls_RA_2
0	prev_opponent_personal_fouls_RA_1
0	prev_opponent_defensive_rebounds_RA_4
0	prev_opponent_defensive_rebounds_RA_3
0	prev_opponent_defensive_rebounds_RA_1
0	prev_opponent_defensive_rebound_percentage_cum_avg
0	prev_opponent_defensive_rebound_percentage_RA_5
0	prev_opponent_defensive_rebound_percentage_RA_4
0	prev_opponent_defensive_rebound_percentage_RA_3
0	prev_opponent_defensive_rebound_percentage_RA_2
0	prev_opponent_defensive_rebound_percentage_RA_1
0	prev_opponent_defensive_rating_RA_5
0	prev_opponent_defensive_rating_RA_4
0	prev_opponent_defensive_rating_RA_3
0	prev_opponent_defensive_rating_RA_2
0	prev_opponent_defensive_rating_RA_1
0	prev_opponent_blocks_cum_avg
0	prev_opponent_blocks_RA_2
0	prev_opponent_blocks_RA_1
0	prev_opponent_block_percentage_RA_5
0	prev_opponent_block_percentage_RA_4
0	prev_opponent_block_percentage_RA_3
0	prev_opponent_block_percentage_RA_1
0	personal_fouls_RA_4
0	personal_fouls_RA_2
0	opponent_turnovers_RA_5

0	opponent_turnovers_RA_4
0	opponent_turnovers_RA_3
0	opponent_turnovers_RA_2
0	opponent_turnovers_RA_1
0	opponent_turnover_percentage_RA_4
0	opponent_turnover_percentage_RA_1
0	opponent_steals_cum_avg
0	opponent_steals_RA_4
0	opponent_steals_RA_3
0	opponent_steals_RA_2
0	opponent_steals_RA_1
0	opponent_points_RA_1
0	opponent_personal_fouls_RA_4
0	opponent_personal_fouls_RA_3
0	opponent_personal_fouls_RA_2
0	opponent_defensive_rebound_percentage_RA_4
0	opponent_defensive_rating_RA_3
0	opponent_blocks_RA_5
0	opponent_blocks_RA_3
0	opponent_block_percentage_RA_3
0	offensive_rebounds_RA_5
0	offensive_rebounds_RA_4
0	offensive_rebounds_RA_3
0	offensive_rebounds_RA_1
0	offensive_rebound_percentage_RA_5
0	offensive_rating_RA_4
0	offensive_rating_RA_3
0	offensive_rating_RA_2
0	free_throws_RA_5
0	free_throws_RA_4
0	free_throws_RA_2
0	free_throws_RA_1
0	free_throw_percentage_RA_4
0	free_throw_attempts_cum_avg
0	free_throw_attempts_RA_1
0	free_throw_attempt_rate_RA_5
0	free_throw_attempt_rate_RA_3

0	field_goal_percentage_RA_4
0	field_goal_percentage_RA_2
0	field_goal_percentage_RA_1
0	field_goal_attempts_RA_5
0	field_goal_attempts_RA_2
0	effective_field_goal_percentage_RA_5
0	effective_field_goal_percentage_RA_3
0	defensive_rebounds_RA_3
0	defensive_rebound_percentage_RA_5
0	defensive_rebound_percentage_RA_3
0	defensive_rating_RA_3
0	blocks_RA_5
0	blocks_RA_4
0	blocks_RA_3
0	blocks_RA_2
0	blocks_RA_1
0	block_percentage_cum_avg
0	block_percentage_RA_5
0	block_percentage_RA_4
0	block_percentage_RA_3
0	block_percentage_RA_2
0	assists_RA_3
0	assists_RA_2
0	assist_percentage_RA_5

Split Plot:

*A more visible plot can be found on https://github.com/Lhett2626/Capstone



Gain:

Total list of gain values:

Value	Feature
4296617.67	points_cum_avg
2173834.08	opponent_defensive_rating_cum_avg
2100617.17	opponent_pace_cum_avg
334002.431	points_RA_5
261288.869	opponent_defensive_rebound_percentage_cum_avg
140805.841	opponent_pace_RA_4
118561.298	prev_opponent_defensive_rating_cum_avg
100159.7	opponent_pace_RA_5
79042.869	opponent_block_percentage_cum_avg
74263.4102	opponent_points_cum_avg
69632.3295	turnover_percentage_cum_avg
54646.2799	blocks_cum_avg
54179.55	opponent_pace_RA_2
50538.1406	points_RA_4
40841.8805	opponent_defensive_rating_RA_5
33154.8596	assists_cum_avg
30920.3198	opponent_pace_RA_3
29429.9597	opponent_turnovers_cum_avg
25583.8593	field_goals_RA_5
24295.6798	defensive_rating_cum_avg
23178.1368	personal_fouls_cum_avg
22836.8092	field_goal_attempts_cum_avg
21985.6	field_goal_attempts_RA_1
21107.4098	pace_cum_avg
20883.53	pace_RA_2
20707.9399	opponent_
20671.6399	pace_RA_1
18585.17	opponent_pace_RA_1
17729.5001	free_throws_cum_avg
17093.2999	assist_percentage_cum_avg
16854.606	opponent_steal_percentage_RA_4
16771.2701	opponent_block_percentage_RA_4
16241.9733	opponent_block_percentage_RA_1
15277.6399	free_throw_percentage_cum_avg

14887.9472	steals_cum_avg
14130.4497	field_goal_attempts_RA_3
12412.21	opponent_steal_percentage_RA_3
10863.58	opponent_block_percentage_RA_5
9955.7002	opponent_steal_percentage_RA_5
9798.06006	defensive_rebounds_cum_avg
9781.75879	three_point_field_goal_percentage_RA_3
9412.37	assist_percentage_RA_2
8791.32983	total_rebounds_RA_5
8673.90002	two_point_field_goal_percentage_cum_avg
8636.6499	prev_opponent_block_percentage_RA_2
8551.93005	opponent_steal_percentage_RA_2
8320.66003	opponent_points_RA_4
8177.04004	defensive_rebounds_RA_5
8149.67993	three_point_field_goal_attempts_RA_1
7918.19006	two_point_field_goal_percentage_RA_1
7823.67004	pace_RA_4
7735.41998	opponent_blocks_cum_avg
7458.44019	effective_field_goal_percentage_RA_4
7446.3999	offensive_rating_RA_5
7317.30981	defensive_rebounds_RA_4
7200.6001	opponent_defensive_rebounds_RA_4
7021.60986	opponent_points_RA_5
6866.84021	opponent_steal_percentage_cum_avg
6743.63013	opponent_defensive_rebound_percentage_RA_3
6664.31995	prev_opponent_block_percentage_cum_avg
6563.41089	opponent_defensive_rebounds_cum_avg
6309.35986	three_point_field_goal_percentage_cum_avg
6303.34998	defensive_rebounds_RA_2
6209.03003	opponent_turnover_percentage_cum_avg
5947.12988	opponent_defensive_rebounds_RA_3
5876.96008	three_point_field_goal_attempts_cum_avg
5865.14001	three_point_field_goal_attempts_RA_4
5787.33008	field_goals_cum_avg
5513.61987	offensive_rating_cum_avg
5427.79004	turnover_percentage_RA_5
5387.44489	three_point_attempt_rate_RA_2
5332.37012	total_rebounds_RA_3
5255.71021	assists_RA_4

5255.20007	three_point_field_goal_attempts_RA_5
5249.92017	two_point_field_goals_cum_avg
5208.91003	defensive_rating_RA_4
5063.47998	three_point_field_goal_attempts_RA_3
4743.68005	prev_opponent_defensive_rebounds_cum_avg
4669.48096	steal_percentage_RA_4
4655.20007	free_throw_percentage_RA_5
4633.10999	free_throw_attempt_rate_cum_avg
4617.3501	opponent_turnover_percentage_RA_3
4602.86597	offensive_rebound_percentage_RA_1
4436.81982	pace_RA_5
4435.72009	defensive_rating_RA_2
4385.82996	free_throw_percentage_RA_3
4325.22009	block_percentage_RA_1
4312.30994	three_point_attempt_rate_RA_3
4259.1001	field_goal_attempts_RA_4
4237.26001	two_point_field_goal_attempts_cum_avg
4123.34998	defensive_rating_RA_5
4040.04004	three_point_attempt_rate_RA_5
4003.58008	points_RA_2
3911.48999	free_throws_RA_3
3833.58984	opponent_steal_percentage_RA_1
3803.17999	opponent_defensive_rebounds_RA_2
3794.12	free_throw_attempt_rate_RA_1
3791.39001	two_point_field_goal_percentage_RA_3
3762.64001	prev_opponent_defensive_rebounds_RA_2
3738.71997	opponent_personal_fouls_RA_5
3667.18005	opponent_block_percentage_RA_2
3627.54993	assists_RA_1
3626.38	opponent_defensive_rebound_percentage_RA_1
3617.69995	field_goals_RA_4
3605.18994	opponent_points_RA_3
3449.15002	three_point_field_goal_percentage_RA_4
3443.59998	true_shooting_percentage_cum_avg
3429.78003	turnover_percentage_RA_3
3332.3999	opponent_defensive_rebounds_RA_5
3331.76001	pace_RA_3
3310.21008	offensive_rating_RA_1
3234.33997	true_shooting_percentage_RA_1

3229.27002	defensive_rating_RA_1
3228.6499	defensive_rebound_percentage_cum_avg
3222.13	free_throw_percentage_RA_2
3188.30005	opponent_blocks_RA_4
3036.92993	opponent_blocks_RA_1
3010.2699	opponent_personal_fouls_cum_avg
2878.36597	two_point_field_goals_RA_4
2860.12988	field_goals_RA_3
2846.56995	opponent_steals_RA_5
2820.09009	opponent_defensive_rating_RA_1
2796.73999	free_throw_percentage_RA_1
2705.92993	three_point_field_goals_RA_5
2661.28003	two_point_field_goals_RA_2
2593.84698	total_rebounds_RA_4
2546.1001	two_point_field_goal_percentage_RA_2
2544.02002	two_point_field_goals_RA_5
2439.95996	opponent_turnover_percentage_RA_2
2410.84009	effective_field_goal_percentage_RA_2
2260.1001	free_throw_attempts_RA_3
2228.88989	offensive_rebounds_RA_2
2198.18994	prev_opponent_blocks_RA_5
2148.52997	personal_fouls_RA_5
2144.07007	turnover_percentage_RA_4
2129.06006	offensive_rebounds_cum_avg
2114.51001	opponent_defensive_rating_RA_4
2081.08008	three_point_field_goal_percentage_RA_1
2057.27301	opponent_personal_fouls_RA_1
2040.09601	opponent_points_RA_2
2017.57996	points_RA_3
1997.71997	opponent_turnover_percentage_RA_5
1960.42004	offensive_rebound_percentage_RA_4
1917.14001	assists_RA_5
1910.03003	opponent_defensive_rebounds_RA_1
1909.27002	personal_fouls_RA_3
1901.73999	two_point_field_goals_RA_1
1891.89001	prev_opponent_blocks_RA_4
1888.55005	steals_RA_2
1867.16003	total_rebound_percentage_RA_2
1866.73999	personal_fouls_RA_1

1860.06006	free_throw_attempts_RA_2
1850.02002	three_point_attempt_rate_RA_1
1846.35999	total_rebounds_RA_1
1825.04004	points_RA_1
1796.80005	steal_percentage_RA_3
1789.85999	three_point_field_goal_percentage_RA_2
1745.45996	offensive_rebound_percentage_RA_3
1744.33997	free_throw_attempts_RA_5
1696.68005	offensive_rebound_percentage_cum_avg
1674.82996	free_throw_attempt_rate_RA_2
1673.81006	two_point_field_goal_attempts_RA_3
1667.47998	defensive_rebounds_RA_1
1666.73999	three_point_attempt_rate_RA_4
1659.18994	free_throw_attempts_RA_4
1656.32996	free_throw_attempt_rate_RA_4
1643.85999	two_point_field_goal_attempts_RA_5
1643.12	field_goal_percentage_RA_3
1600.07996	opponent_blocks_RA_2
1597.89001	prev_opponent_defensive_rebounds_RA_5
1583.58997	assist_percentage_RA_3
1569.29004	assist_percentage_RA_4
1559.43994	turnovers_RA_2
1557.58997	field_goals_RA_2
1547.41003	field_goal_percentage_RA_5
1547.31995	offensive_rebound_percentage_RA_2
1543.63	two_point_field_goal_attempts_RA_2
1542.07996	field_goal_percentage_cum_avg
1541.01001	opponent_defensive_rebound_percentage_RA_2
1489.12	prev_opponent_blocks_RA_3
1459.51001	effective_field_goal_percentage_RA_1
1457.78003	true_shooting_percentage_RA_3
1448.20996	defensive_rebound_percentage_RA_4
1433.10999	total_rebounds_RA_2
1424.93005	field_goals_RA_1
1404.48999	defensive_rebound_percentage_RA_1
1403.67004	turnovers_RA_3
1399.80005	opponent_defensive_rating_RA_2
1336.72998	defensive_rebound_percentage_RA_2
1320.82996	effective_field_goal_percentage_cum_avg

1285.04004	assist_percentage_RA_1
951.072021	total_rebound_percentage_RA_3
935.867981	steal_percentage_cum_avg
0	two_point_field_goals_RA_3
0	two_point_field_goal_percentage_RA_5
0	two_point_field_goal_percentage_RA_4
0	two_point_field_goal_attempts_RA_4
0	two_point_field_goal_attempts_RA_1
0	turnovers_cum_avg
0	turnovers_RA_5
0	turnovers_RA_4
0	turnovers_RA_1
0	turnover_percentage_RA_2
0	turnover_percentage_RA_1
0	true_shooting_percentage_RA_5
0	true_shooting_percentage_RA_4
0	true_shooting_percentage_RA_2
0	total_rebounds_cum_avg
0	total_rebound_percentage_cum_avg
0	total_rebound_percentage_RA_5
0	total_rebound_percentage_RA_4
0	total_rebound_percentage_RA_1
0	three_point_field_goals_cum_avg
0	three_point_field_goals_RA_4
0	three_point_field_goals_RA_3
0	three_point_field_goals_RA_2
0	three_point_field_goals_RA_1
0	three_point_field_goal_percentage_RA_5
0	three_point_field_goal_attempts_RA_2
0	three_point_attempt_rate_cum_avg
0	steals_RA_5
0	steals_RA_4
0	steals_RA_3
0	steals_RA_1
0	steal_percentage_RA_5
0	steal_percentage_RA_2
0	steal_percentage_RA_1
0	prev_opponent_turnovers_cum_avg
0	prev_opponent_turnovers_RA_5

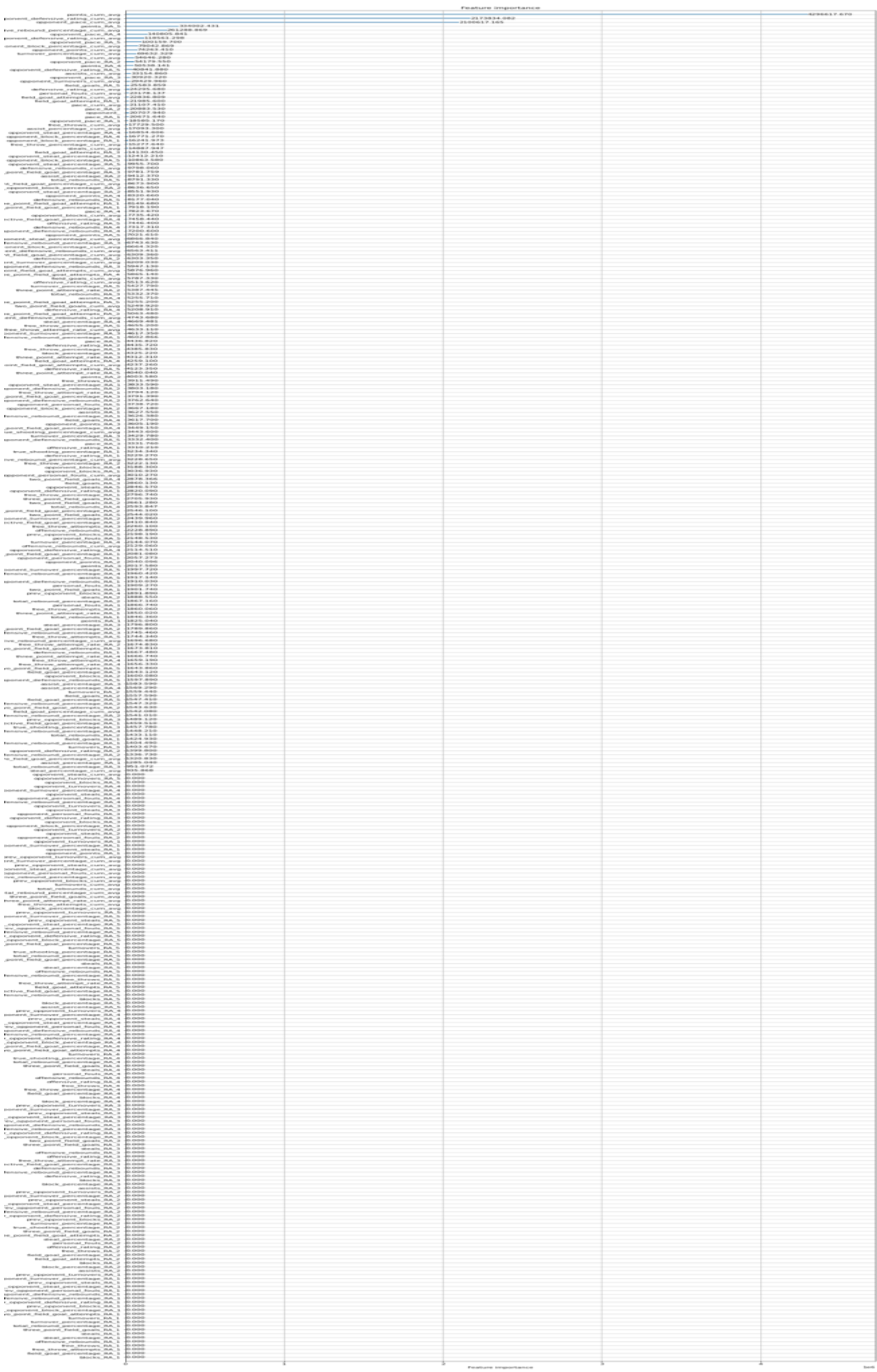
0	prev_opponent_turnovers_RA_4
0	prev_opponent_turnovers_RA_3
0	prev_opponent_turnovers_RA_2
0	prev_opponent_turnovers_RA_1
0	prev_opponent_turnover_percentage_cum_avg
0	prev_opponent_turnover_percentage_RA_5
0	prev_opponent_turnover_percentage_RA_4
0	prev_opponent_turnover_percentage_RA_3
0	prev_opponent_turnover_percentage_RA_2
0	prev_opponent_turnover_percentage_RA_1
0	prev_opponent_steals_cum_avg
0	prev_opponent_steals_RA_5
0	prev_opponent_steals_RA_4
0	prev_opponent_steals_RA_3
0	prev_opponent_steals_RA_2
0	prev_opponent_steals_RA_1
0	prev_opponent_steal_percentage_cum_avg
0	prev_opponent_steal_percentage_RA_5
0	prev_opponent_steal_percentage_RA_4
0	prev_opponent_steal_percentage_RA_3
0	prev_opponent_steal_percentage_RA_2
0	prev_opponent_steal_percentage_RA_1
0	prev_opponent_personal_fouls_cum_avg
0	prev_opponent_personal_fouls_RA_5
0	prev_opponent_personal_fouls_RA_4
0	prev_opponent_personal_fouls_RA_3
0	prev_opponent_personal_fouls_RA_2
0	prev_opponent_personal_fouls_RA_1
0	prev_opponent_defensive_rebounds_RA_4
0	prev_opponent_defensive_rebounds_RA_3
0	prev_opponent_defensive_rebounds_RA_1
0	prev_opponent_defensive_rebound_percentage_cum_avg
0	prev_opponent_defensive_rebound_percentage_RA_5
0	prev_opponent_defensive_rebound_percentage_RA_4
0	prev_opponent_defensive_rebound_percentage_RA_3
0	prev_opponent_defensive_rebound_percentage_RA_2
0	prev_opponent_defensive_rebound_percentage_RA_1
0	prev_opponent_defensive_rating_RA_5
0	prev_opponent_defensive_rating_RA_4

0	prev_opponent_defensive_rating_RA_3
0	prev_opponent_defensive_rating_RA_2
0	prev_opponent_defensive_rating_RA_1
0	prev_opponent_blocks_cum_avg
0	prev_opponent_blocks_RA_2
0	prev_opponent_blocks_RA_1
0	prev_opponent_block_percentage_RA_5
0	prev_opponent_block_percentage_RA_4
0	prev_opponent_block_percentage_RA_3
0	prev_opponent_block_percentage_RA_1
0	personal_fouls_RA_4
0	personal_fouls_RA_2
0	opponent_turnovers_RA_5
0	opponent_turnovers_RA_4
0	opponent_turnovers_RA_3
0	opponent_turnovers_RA_2
0	opponent_turnovers_RA_1
0	opponent_turnover_percentage_RA_4
0	opponent_turnover_percentage_RA_1
0	opponent_steals_cum_avg
0	opponent_steals_RA_4
0	opponent_steals_RA_3
0	opponent_steals_RA_2
0	opponent_steals_RA_1
0	opponent_points_RA_1
0	opponent_personal_fouls_RA_4
0	opponent_personal_fouls_RA_3
0	opponent_personal_fouls_RA_2
0	opponent_defensive_rebound_percentage_RA_4
0	opponent_defensive_rating_RA_3
0	opponent_blocks_RA_5
0	opponent_blocks_RA_3
0	opponent_block_percentage_RA_3
0	offensive_rebounds_RA_5
0	offensive_rebounds_RA_4
0	offensive_rebounds_RA_3
0	offensive_rebounds_RA_1
0	offensive_rebound_percentage_RA_5
0	offensive_rating_RA_4

0	offensive_rating_RA_3
0	offensive_rating_RA_2
0	free_throws_RA_5
0	free_throws_RA_4
0	free_throws_RA_2
0	free_throws_RA_1
0	free_throw_percentage_RA_4
0	free_throw_attempts_cum_avg
0	free_throw_attempts_RA_1
0	free_throw_attempt_rate_RA_5
0	free_throw_attempt_rate_RA_3
0	field_goal_percentage_RA_4
0	field_goal_percentage_RA_2
0	field_goal_percentage_RA_1
0	field_goal_attempts_RA_5
0	field_goal_attempts_RA_2
0	effective_field_goal_percentage_RA_5
0	effective_field_goal_percentage_RA_3
0	defensive_rebounds_RA_3
0	defensive_rebound_percentage_RA_5
0	defensive_rebound_percentage_RA_3
0	defensive_rating_RA_3
0	blocks_RA_5
0	blocks_RA_4
0	blocks_RA_3
0	blocks_RA_2
0	blocks_RA_1
0	block_percentage_cum_avg
0	block_percentage_RA_5
0	block_percentage_RA_4
0	block_percentage_RA_3
0	block_percentage_RA_2
0	assists_RA_3
0	assists_RA_2
0	assist_percentage_RA_5

Gain Plot:

*A more visible plot can be found on https://github.com/Lhett2626/Capstone



SHAP Scores:

Feature	Shap
points_cum_avg	2.44890029
opponent_defensive_rating_cum_avg	1.78611381
opponent_pace_cum_avg	1.62587683
opponent_defensive_rebound_percentage_cum_avg	0.60545073
points_RA_5	0.37135816
prev_opponent_defensive_rating_cum_avg	0.25807374
opponent_block_percentage_cum_avg	0.25203186
opponent_points_cum_avg	0.19240144
opponent_pace_RA_4	0.18910141
turnover_percentage_cum_avg	0.18809822
blocks_cum_avg	0.16939769
defensive_rating_cum_avg	0.14163948
opponent_pace_RA_5	0.14087394
opponent_pace_RA_2	0.14061166
free_throws_cum_avg	0.10566722
personal_fouls_cum_avg	0.10332049
opponent_defensive_rating_RA_5	0.09740856
opponent_steal_percentage_RA_4	0.09703637
pace_RA_1	0.08971892
free_throw_percentage_cum_avg	0.08835068
opponent_turnovers_cum_avg	0.08672503
assists_cum_avg	0.08661249
opponent_block_percentage_RA_1	0.08387353
field_goal_attempts_RA_1	0.08005883
points_RA_4	0.07307025
opponent_defensive_rebounds_RA_4	0.06650444
opponent_points_RA_5	0.06404261
three_point_field_goal_percentage_RA_3	0.06032196
field_goals_RA_5	0.0602412
opponent_	0.05962696
pace_RA_2	0.05733424
assist_percentage_cum_avg	0.05564829
opponent_pace_RA_3	0.05359808
field_goal_attempts_cum_avg	0.04941659
opponent_pace_RA_1	0.04593659
opponent_defensive_rebounds_cum_avg	0.04553704

field_goal_attempts_RA_3	0.04535895
opponent_block_percentage_RA_4	0.04442933
opponent_defensive_rebounds_RA_2	0.04111607
opponent_steal_percentage_RA_3	0.03839301
defensive_rebounds_RA_5	0.03499052
pace_cum_avg	0.03414521
opponent_steal_percentage_cum_avg	0.03318289
opponent_defensive_rebounds_RA_3	0.03257443
opponent_defensive_rebound_percentage_RA_3	0.02876665
personal_fouls_RA_5	0.02846436
opponent_block_percentage_RA_5	0.02769695
steals_cum_avg	0.02725368
three_point_field_goal_attempts_RA_1	0.02612285
prev_opponent_block_percentage_RA_2	0.02577746
opponent_blocks_cum_avg	0.02539304
steal_percentage_RA_4	0.02439662
defensive_rating_RA_5	0.02393733
assist_percentage_RA_2	0.02056246
free_throw_percentage_RA_3	0.0195738
free_throw_attempt_rate_cum_avg	0.01935665
opponent_steal_percentage_RA_5	0.01927589
opponent_steals_RA_5	0.0191343
turnover_percentage_RA_5	0.01883948
effective_field_goal_percentage_RA_4	0.01860059
defensive_rebounds_RA_4	0.01841326
turnover_percentage_RA_3	0.01762258
total_rebounds_RA_5	0.01745049
free_throw_percentage_RA_5	0.01685004
three_point_field_goal_attempts_RA_4	0.01611671
opponent_steal_percentage_RA_1	0.0151704
opponent_points_RA_3	0.01482564
field_goal_attempts_RA_4	0.01478672
offensive_rebound_percentage_RA_1	0.01473631
opponent_steal_percentage_RA_2	0.01456727
free_throw_attempt_rate_RA_4	0.01294372
three_point_attempt_rate_RA_2	0.01225633
three_point_field_goal_percentage_cum_avg	0.01186301
three_point_field_goal_attempts_cum_avg	0.01139038
steal_percentage_cum_avg	0.01104276

pace_RA_5	0.01095497
free_throw_attempt_rate_RA_1	0.0108709
offensive_rating_RA_5	0.01045852
points_RA_2	0.01021396
two_point_field_goal_attempts_cum_avg	0.0098391
three_point_field_goal_attempts_RA_5	0.00976838
three_point_attempt_rate_RA_4	0.00974719
opponent_blocks_RA_1	0.00969302
three_point_attempt_rate_RA_1	0.009452
effective_field_goal_percentage_RA_2	0.00944628
three_point_field_goal_attempts_RA_3	0.00940118
defensive_rebounds_cum_avg	0.00930545
two_point_field_goal_percentage_cum_avg	0.00913053
opponent_defensive_rating_RA_2	0.00897254
opponent_turnover_percentage_RA_2	0.00865713
opponent_points_RA_4	0.00864182
three_point_attempt_rate_RA_3	0.00840821
prev_opponent_defensive_rebounds_RA_2	0.007764
total_rebounds_RA_4	0.0076057
offensive_rebound_percentage_RA_3	0.00730492
assists_RA_4	0.0072361
defensive_rebounds_RA_2	0.00707016
pace_RA_3	0.00670244
true_shooting_percentage_RA_1	0.00668346
total_rebounds_RA_1	0.00664789
field_goals_cum_avg	0.00609966
three_point_field_goal_percentage_RA_4	0.00597305
pace_RA_4	0.0059495
free_throw_percentage_RA_2	0.00593338
opponent_block_percentage_RA_2	0.00592616
defensive_rating_RA_4	0.00587897
offensive_rebound_percentage_RA_4	0.00554619
offensive_rating_cum_avg	0.00547091
opponent_defensive_rating_RA_1	0.00504026
defensive_rating_RA_2	0.00495305
offensive_rebounds_cum_avg	0.00492017
opponent_defensive_rating_RA_4	0.00490037
two_point_field_goal_percentage_RA_1	0.00480057
block_percentage_RA_1	0.00469977

turnover_percentage_RA_4	0.00466328
personal_fouls_RA_3	0.00437133
offensive_rebound_percentage_cum_avg	0.00429547
opponent_personal_fouls_cum_avg	0.00415005
opponent_turnover_percentage_cum_avg	0.00413069
total_rebounds_RA_3	0.00405286
prev_opponent_defensive_rebounds_cum_avg	0.00404476
two_point_field_goals_cum_avg	0.00404126
two_point_field_goals_RA_1	0.00396535
total_rebound_percentage_RA_3	0.00385807
three_point_field_goals_RA_5	0.00358333
steal_percentage_RA_3	0.00338261
free_throw_percentage_RA_1	0.0031634
two_point_field_goal_percentage_RA_2	0.00313368
two_point_field_goals_RA_2	0.00291579
personal_fouls_RA_1	0.00286207
three_point_attempt_rate_RA_5	0.00282665
defensive_rating_RA_1	0.00281091
field_goals_RA_4	0.00280554
assists_RA_1	0.00276333
opponent_defensive_rebound_percentage_RA_2	0.00268417
opponent_personal_fouls_RA_1	0.00262278
field_goal_percentage_RA_3	0.00258467
offensive_rebounds_RA_2	0.00257708
turnovers_RA_2	0.00253423
opponent_personal_fouls_RA_5	0.00248661
true_shooting_percentage_cum_avg	0.00240823
three_point_field_goal_percentage_RA_1	0.00239974
two_point_field_goal_attempts_RA_2	0.00236733
offensive_rating_RA_1	0.00234387
two_point_field_goal_percentage_RA_3	0.0022606
defensive_rebound_percentage_cum_avg	0.00224934
prev_opponent_block_percentage_cum_avg	0.00213262
prev_opponent_blocks_RA_4	0.00211306
opponent_defensive_rebounds_RA_5	0.00203611
opponent_turnover_percentage_RA_3	0.0020227
opponent_defensive_rebound_percentage_RA_1	0.00189958
total_rebound_percentage_RA_2	0.00179012
assist_percentage_RA_4	0.00177819

free_throw_attempts_RA_4	0.00175263
defensive_rebounds_RA_1	0.00169823
prev_opponent_blocks_RA_5	0.00168648
two_point_field_goals_RA_5	0.00167932
two_point_field_goals_RA_4	0.00167774
free_throws_RA_3	0.00167663
free_throw_attempts_RA_2	0.00165548
opponent_blocks_RA_4	0.00149404
field_goal_percentage_cum_avg	0.00145768
free_throw_attempts_RA_3	0.00145718
assist_percentage_RA_3	0.00137055
effective_field_goal_percentage_RA_1	0.00135546
true_shooting_percentage_RA_3	0.00131246
effective_field_goal_percentage_cum_avg	0.00126945
field_goals_RA_3	0.00126652
points_RA_3	0.00119492
opponent_points_RA_2	0.00116354
defensive_rebound_percentage_RA_4	0.00106228
prev_opponent_defensive_rebounds_RA_5	0.00106171
prev_opponent_blocks_RA_3	0.0010612
opponent_blocks_RA_2	0.00105522
field_goals_RA_1	0.00104046
assists_RA_5	0.00099852
points_RA_1	0.00099125
three_point_field_goal_percentage_RA_2	0.00098294
free_throw_attempts_RA_5	0.00094933
steals_RA_2	0.0009406
offensive_rebound_percentage_RA_2	0.0009392
defensive_rebound_percentage_RA_1	0.00092383
defensive_rebound_percentage_RA_2	0.00085298
free_throw_attempt_rate_RA_2	0.00082875
two_point_field_goal_attempts_RA_3	0.00080226
two_point_field_goal_attempts_RA_5	0.00075551
opponent_defensive_rebounds_RA_1	0.00075536
total_rebounds_RA_2	0.00074532
field_goal_percentage_RA_5	0.00068171
opponent_turnover_percentage_RA_5	0.00062518
field_goals_RA_2	0.0005897
turnovers_RA_3	0.00057495

assist_percentage_RA_1	0.00045796

Correlation:

Feature	Shap
points_cum_avg	2.44890029
opponent_defensive_rating_cum_avg	1.78611381
opponent_pace_cum_avg	1.62587683
opponent_defensive_rebound_percentage_cum_avg	0.60545073
points_RA_5	0.37135816
prev_opponent_defensive_rating_cum_avg	0.25807374
opponent_block_percentage_cum_avg	0.25203186
opponent_points_cum_avg	0.19240144
opponent_pace_RA_4	0.18910141
turnover_percentage_cum_avg	0.18809822
blocks_cum_avg	0.16939769
defensive_rating_cum_avg	0.14163948
opponent_pace_RA_5	0.14087394
opponent_pace_RA_2	0.14061166
free_throws_cum_avg	0.10566722
personal_fouls_cum_avg	0.10332049
opponent_defensive_rating_RA_5	0.09740856
opponent_steal_percentage_RA_4	0.09703637
pace_RA_1	0.08971892
free_throw_percentage_cum_avg	0.08835068
opponent_turnovers_cum_avg	0.08672503
assists_cum_avg	0.08661249
opponent_block_percentage_RA_1	0.08387353
field_goal_attempts_RA_1	0.08005883
points_RA_4	0.07307025
opponent_defensive_rebounds_RA_4	0.06650444
opponent_points_RA_5	0.06404261
three_point_field_goal_percentage_RA_3	0.06032196
field_goals_RA_5	0.0602412
opponent_	0.05962696
pace_RA_2	0.05733424
assist_percentage_cum_avg	0.05564829
opponent_pace_RA_3	0.05359808

field_goal_attempts_cum_avg	0.04941659
opponent_pace_RA_1	0.04593659
opponent_defensive_rebounds_cum_avg	0.04553704
field_goal_attempts_RA_3	0.04535895
opponent_block_percentage_RA_4	0.04442933
opponent_defensive_rebounds_RA_2	0.04111607
opponent_steal_percentage_RA_3	0.03839301
defensive_rebounds_RA_5	0.03499052
pace_cum_avg	0.03414521
opponent_steal_percentage_cum_avg	0.03318289
opponent_defensive_rebounds_RA_3	0.03257443
opponent_defensive_rebound_percentage_RA_3	0.02876665
personal_fouls_RA_5	0.02846436
opponent_block_percentage_RA_5	0.02769695
steals_cum_avg	0.02725368
three_point_field_goal_attempts_RA_1	0.02612285
prev_opponent_block_percentage_RA_2	0.02577746
opponent_blocks_cum_avg	0.02539304
steal_percentage_RA_4	0.02439662
defensive_rating_RA_5	0.02393733
assist_percentage_RA_2	0.02056246
free_throw_percentage_RA_3	0.0195738
free_throw_attempt_rate_cum_avg	0.01935665
opponent_steal_percentage_RA_5	0.01927589
opponent_steals_RA_5	0.0191343
turnover_percentage_RA_5	0.01883948
effective_field_goal_percentage_RA_4	0.01860059
defensive_rebounds_RA_4	0.01841326
turnover_percentage_RA_3	0.01762258
total_rebounds_RA_5	0.01745049
free_throw_percentage_RA_5	0.01685004
three_point_field_goal_attempts_RA_4	0.01611671
opponent_steal_percentage_RA_1	0.0151704
opponent_points_RA_3	0.01482564
field_goal_attempts_RA_4	0.01478672
offensive_rebound_percentage_RA_1	0.01473631
opponent_steal_percentage_RA_2	0.01456727
free_throw_attempt_rate_RA_4	0.01294372
three_point_attempt_rate_RA_2	0.01225633

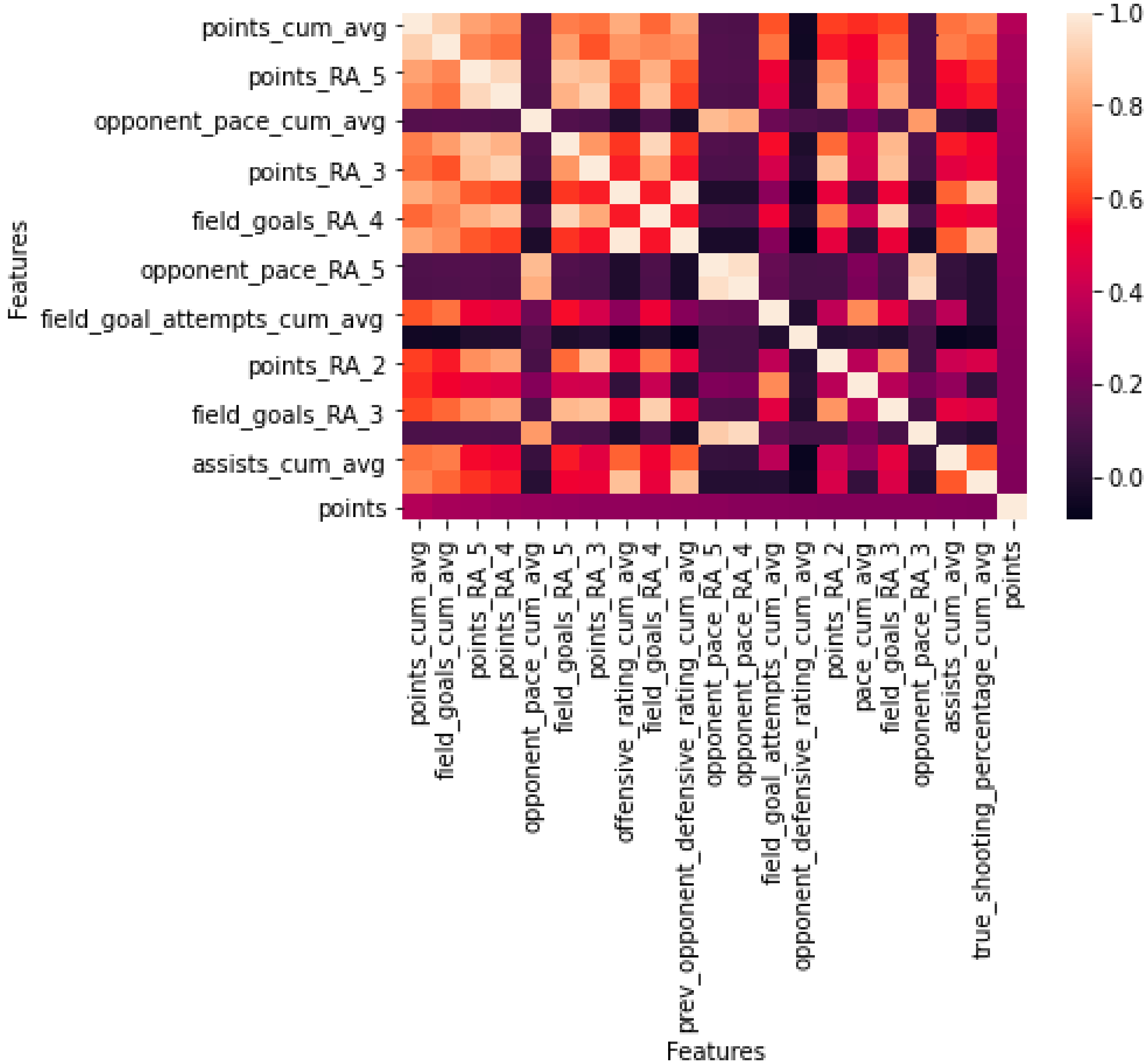
three_point_field_goal_percentage_cum_avg	0.01186301
three_point_field_goal_attempts_cum_avg	0.01139038
steal_percentage_cum_avg	0.01104276
pace_RA_5	0.01095497
free_throw_attempt_rate_RA_1	0.0108709
offensive_rating_RA_5	0.01045852
points_RA_2	0.01021396
two_point_field_goal_attempts_cum_avg	0.0098391
three_point_field_goal_attempts_RA_5	0.00976838
three_point_attempt_rate_RA_4	0.00974719
opponent_blocks_RA_1	0.00969302
three_point_attempt_rate_RA_1	0.009452
effective_field_goal_percentage_RA_2	0.00944628
three_point_field_goal_attempts_RA_3	0.00940118
defensive_rebounds_cum_avg	0.00930545
two_point_field_goal_percentage_cum_avg	0.00913053
opponent_defensive_rating_RA_2	0.00897254
opponent_turnover_percentage_RA_2	0.00865713
opponent_points_RA_4	0.00864182
three_point_attempt_rate_RA_3	0.00840821
prev_opponent_defensive_rebounds_RA_2	0.007764
total_rebounds_RA_4	0.0076057
offensive_rebound_percentage_RA_3	0.00730492
assists_RA_4	0.0072361
defensive_rebounds_RA_2	0.00707016
pace_RA_3	0.00670244
true_shooting_percentage_RA_1	0.00668346
total_rebounds_RA_1	0.00664789
field_goals_cum_avg	0.00609966
three_point_field_goal_percentage_RA_4	0.00597305
pace_RA_4	0.0059495
free_throw_percentage_RA_2	0.00593338
opponent_block_percentage_RA_2	0.00592616
defensive_rating_RA_4	0.00587897
offensive_rebound_percentage_RA_4	0.00554619
offensive_rating_cum_avg	0.00547091
opponent_defensive_rating_RA_1	0.00504026
defensive_rating_RA_2	0.00495305
offensive_rebounds_cum_avg	0.00492017

opponent_defensive_rating_RA_4	0.00490037
two_point_field_goal_percentage_RA_1	0.00480057
block_percentage_RA_1	0.00469977
turnover_percentage_RA_4	0.00466328
personal_fouls_RA_3	0.00437133
offensive_rebound_percentage_cum_avg	0.00429547
opponent_personal_fouls_cum_avg	0.00415005
opponent_turnover_percentage_cum_avg	0.00413069
total_rebounds_RA_3	0.00405286
prev_opponent_defensive_rebounds_cum_avg	0.00404476
two_point_field_goals_cum_avg	0.00404126
two_point_field_goals_RA_1	0.00396535
total_rebound_percentage_RA_3	0.00385807
three_point_field_goals_RA_5	0.00358333
steal_percentage_RA_3	0.00338261
free_throw_percentage_RA_1	0.0031634
two_point_field_goal_percentage_RA_2	0.00313368
two_point_field_goals_RA_2	0.00291579
personal_fouls_RA_1	0.00286207
three_point_attempt_rate_RA_5	0.00282665
defensive_rating_RA_1	0.00281091
field_goals_RA_4	0.00280554
assists_RA_1	0.00276333
opponent_defensive_rebound_percentage_RA_2	0.00268417
opponent_personal_fouls_RA_1	0.00262278
field_goal_percentage_RA_3	0.00258467
offensive_rebounds_RA_2	0.00257708
turnovers_RA_2	0.00253423
opponent_personal_fouls_RA_5	0.00248661
true_shooting_percentage_cum_avg	0.00240823
three_point_field_goal_percentage_RA_1	0.00239974
two_point_field_goal_attempts_RA_2	0.00236733
offensive_rating_RA_1	0.00234387
two_point_field_goal_percentage_RA_3	0.0022606
defensive_rebound_percentage_cum_avg	0.00224934
prev_opponent_block_percentage_cum_avg	0.00213262
prev_opponent_blocks_RA_4	0.00211306
opponent_defensive_rebounds_RA_5	0.00203611
opponent_turnover_percentage_RA_3	0.0020227

opponent_defensive_rebound_percentage_RA_1	0.00189958
total_rebound_percentage_RA_2	0.00179012
assist_percentage_RA_4	0.00177819
free_throw_attempts_RA_4	0.00175263
defensive_rebounds_RA_1	0.00169823
prev_opponent_blocks_RA_5	0.00168648
two_point_field_goals_RA_5	0.00167932
two_point_field_goals_RA_4	0.00167774
free_throws_RA_3	0.00167663
free_throw_attempts_RA_2	0.00165548
opponent_blocks_RA_4	0.00149404
field_goal_percentage_cum_avg	0.00145768
free_throw_attempts_RA_3	0.00145718
assist_percentage_RA_3	0.00137055
effective_field_goal_percentage_RA_1	0.00135546
true_shooting_percentage_RA_3	0.00131246
effective_field_goal_percentage_cum_avg	0.00126945
field_goals_RA_3	0.00126652
points_RA_3	0.00119492
opponent_points_RA_2	0.00116354
defensive_rebound_percentage_RA_4	0.00106228
prev_opponent_defensive_rebounds_RA_5	0.00106171
prev_opponent_blocks_RA_3	0.0010612
opponent_blocks_RA_2	0.00105522
field_goals_RA_1	0.00104046
assists_RA_5	0.00099852
points_RA_1	0.00099125
three_point_field_goal_percentage_RA_2	0.00098294
free_throw_attempts_RA_5	0.00094933
steals_RA_2	0.0009406
offensive_rebound_percentage_RA_2	0.0009392
defensive_rebound_percentage_RA_1	0.00092383
defensive_rebound_percentage_RA_2	0.00085298
free_throw_attempt_rate_RA_2	0.00082875
two_point_field_goal_attempts_RA_3	0.00080226
two_point_field_goal_attempts_RA_5	0.00075551
opponent_defensive_rebounds_RA_1	0.00075536
total_rebounds_RA_2	0.00074532
field_goal_percentage_RA_5	0.00068171

opponent_turnover_percentage_RA_5	0.00062518
field_goals_RA_2	0.0005897
turnovers_RA_3	0.00057495
assist_percentage_RA_1	0.00045796

Correlation Heat Map (Top 20):



Work cited:

Atkinson, Blake. “Guide to Building a College Basketball Machine Learning Model in Python.” *Medium*, Towards Data Science, 12 Dec. 2019, https://towardsdatascience.com/guide-to-building-a-college-basketball-machine-learning-model-in-python-1c70b83acb51.

Browne, Christopher. “Americans to Bet $2 Billion on 70 Million March Madness Brackets This Year, Says New Research.” *American Gaming Association*, 19 Dec. 2018, https://www.americangaming.org/new/americans-to-bet-2-billion-on-70-million-march-madness-brackets-this-year-says-new-research/.

Chen, Tianqi, et al. “Understand Your Dataset with Xgboost.” *Understand Your Dataset with XGBoost - Xgboost 1.6.1 Documentation*, XGBoost, https://xgboost.readthedocs.io/en/stable/R-package/discoverYourData.html.

“March Machine Learning Mania 2022.” *Kaggle*, https://www.kaggle.com/c/mens-march-mania-2022.

Miya, Evan. “Evan Miya CBB Analytics.” *EvanMiya*, https://evanmiya.com/.

Pomeroy, Ken. “2022 Pomeroy College Basketball Ratings.” *Kenpom.com*, https://kenpom.com/.

Wilco, Daniel. “The Absurd Odds of a Perfect NCAA Bracket.” *NCAA.com*, NCAA, 13 Mar. 2022, https://www.ncaa.com/news/basketball-men/bracketiq/2022-03-10/perfect-ncaa-bracket-absurd-odds-march-madness-dream.