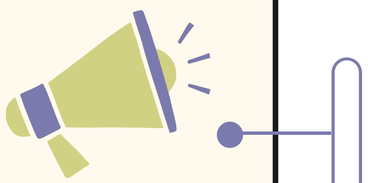
STAT199

Tuesday

5:15 PM

MARCH MADNESS

Morgan Feng, Sophie Mansoor, Sarah Wu



Blue Devils

Basketball!

AGENDA

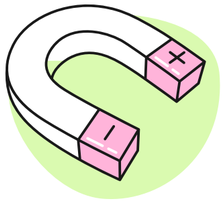
INTRODUCTION

What are we accomplishing?

EDA

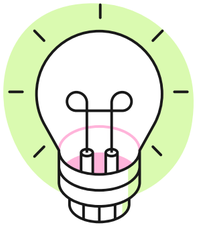
What did we do?

RESULTS



What did we find?

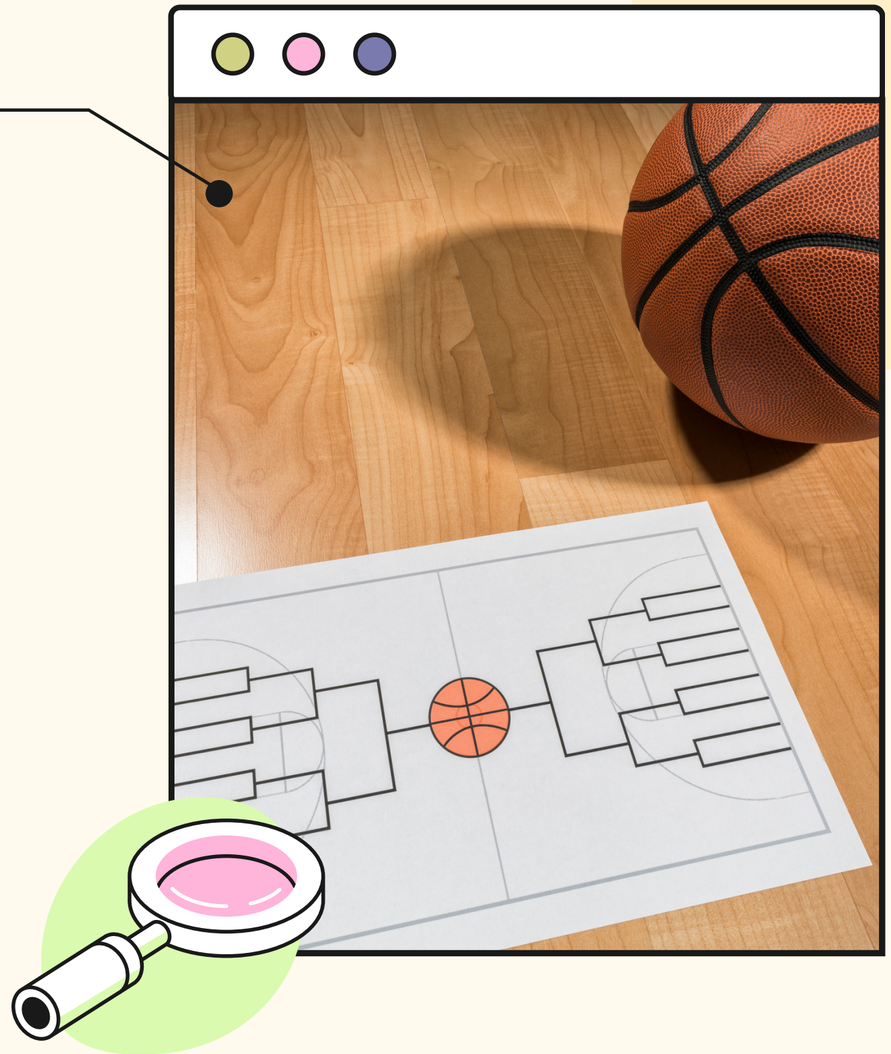
CONCLUSION



Now what?

INTRODUCTION

The March Madness tournament is a highly anticipated event for college basketball fans, who try to fill out the perfect bracket by predicting which teams will make it to each round. This project aims to identify the strongest predictors of success in the tournament to answer our question "What factors are the strongest predictors of success in March Madness?"



INTRODUCTION

BARTHAG

The team's chance of winning against the average DI team

BARTTORVIK ADJUSTED EFFICIENCY

Bart Torvik's calculation of how efficient a team is offensively and defensively

WINS ABOVE BUBBLE

How many more or less wins the average bubble team would have against the team's schedule

Hypothesis:

Teams with a high BARTHAG value will most be likely to find success in the tournament.

using set.seed(101)

EDA

METHODOLOGY

BOXPLOT

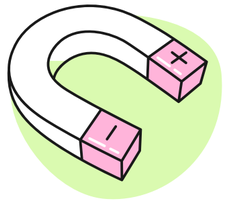
Is there a general trend?

FIT MODELS**

What's the equation?

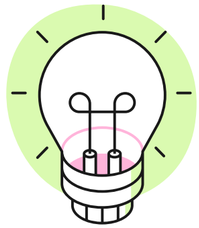
**Start with single variables!

ROC



How does the ROC look?

AUC

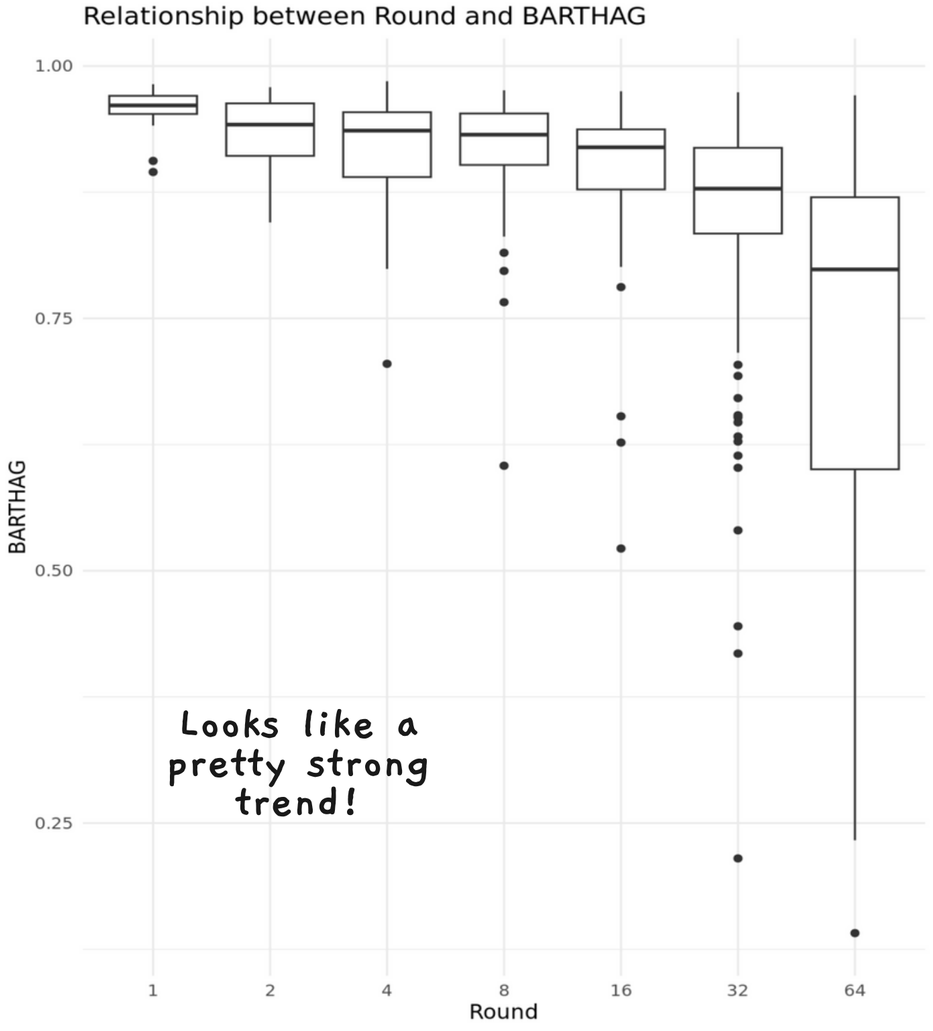


What's the area under the ROC curve?

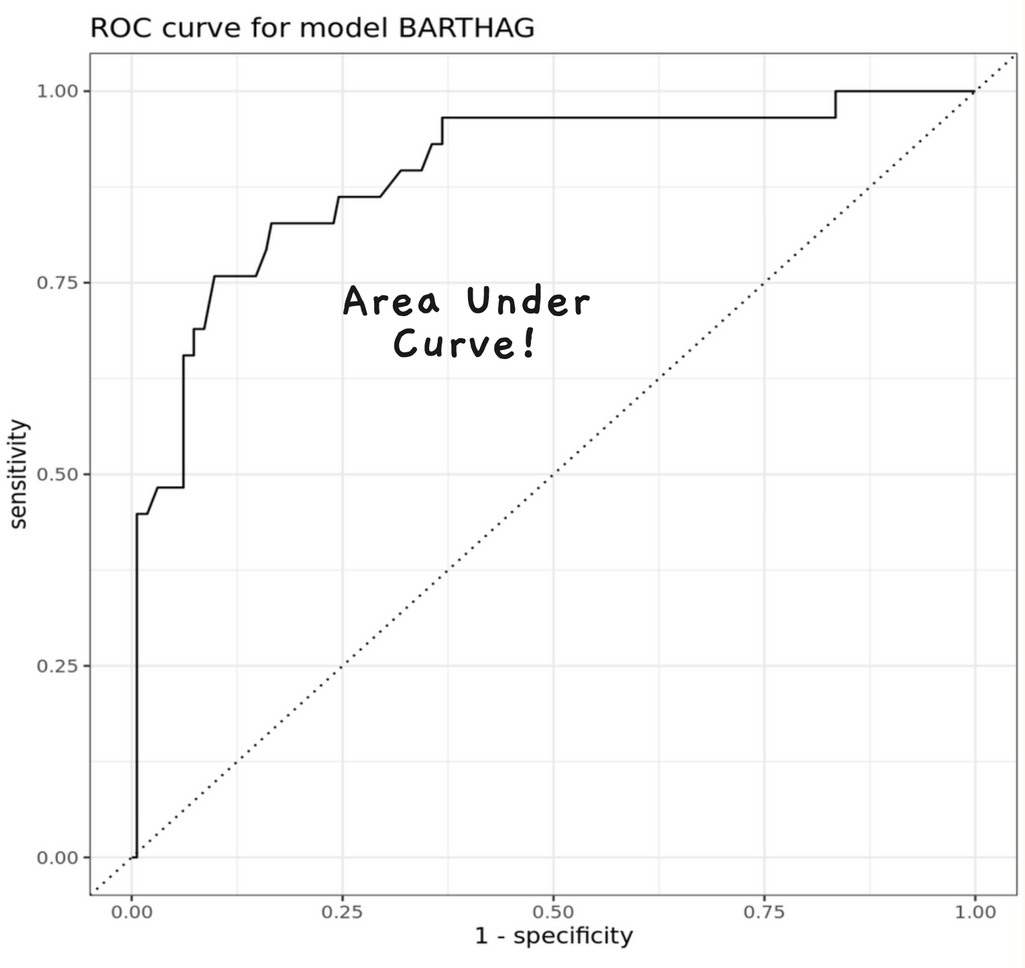
Lastly, MIX AND MATCH variables to find the BEST model!

EDA

BARTHAG



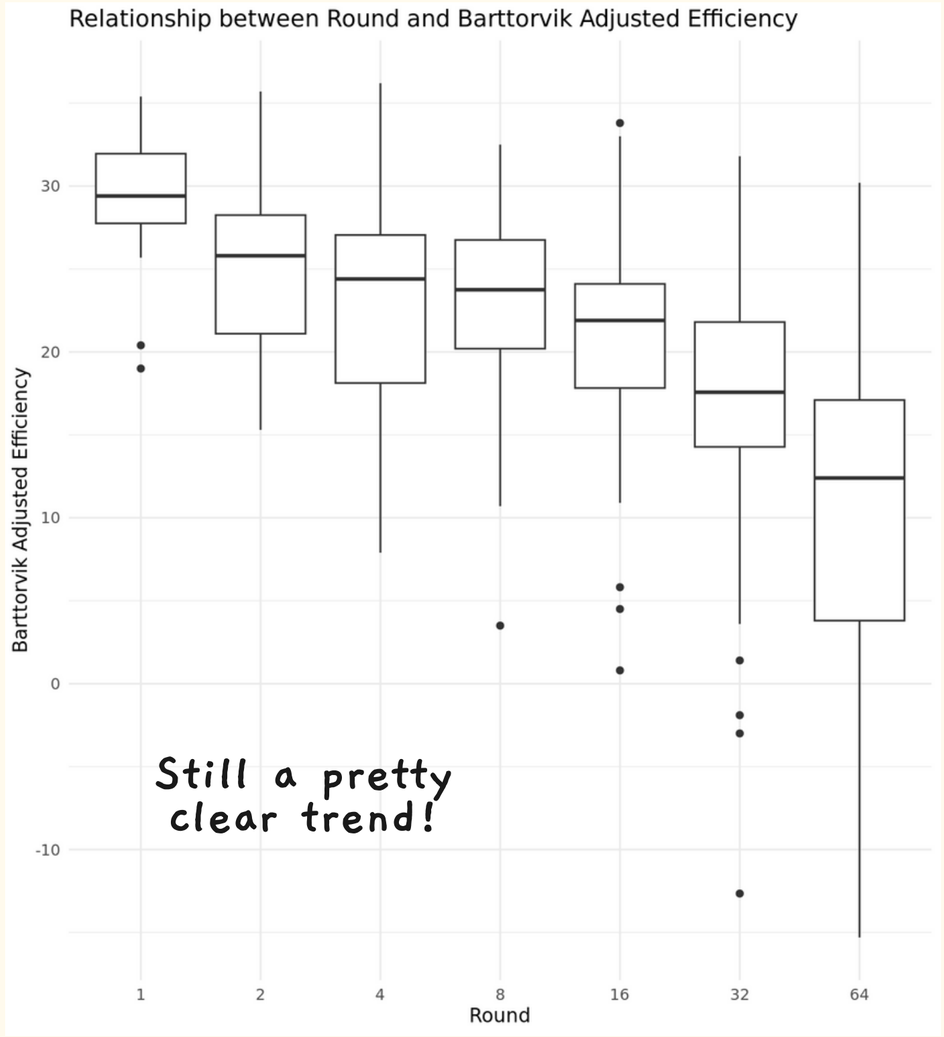
BOXPLOT



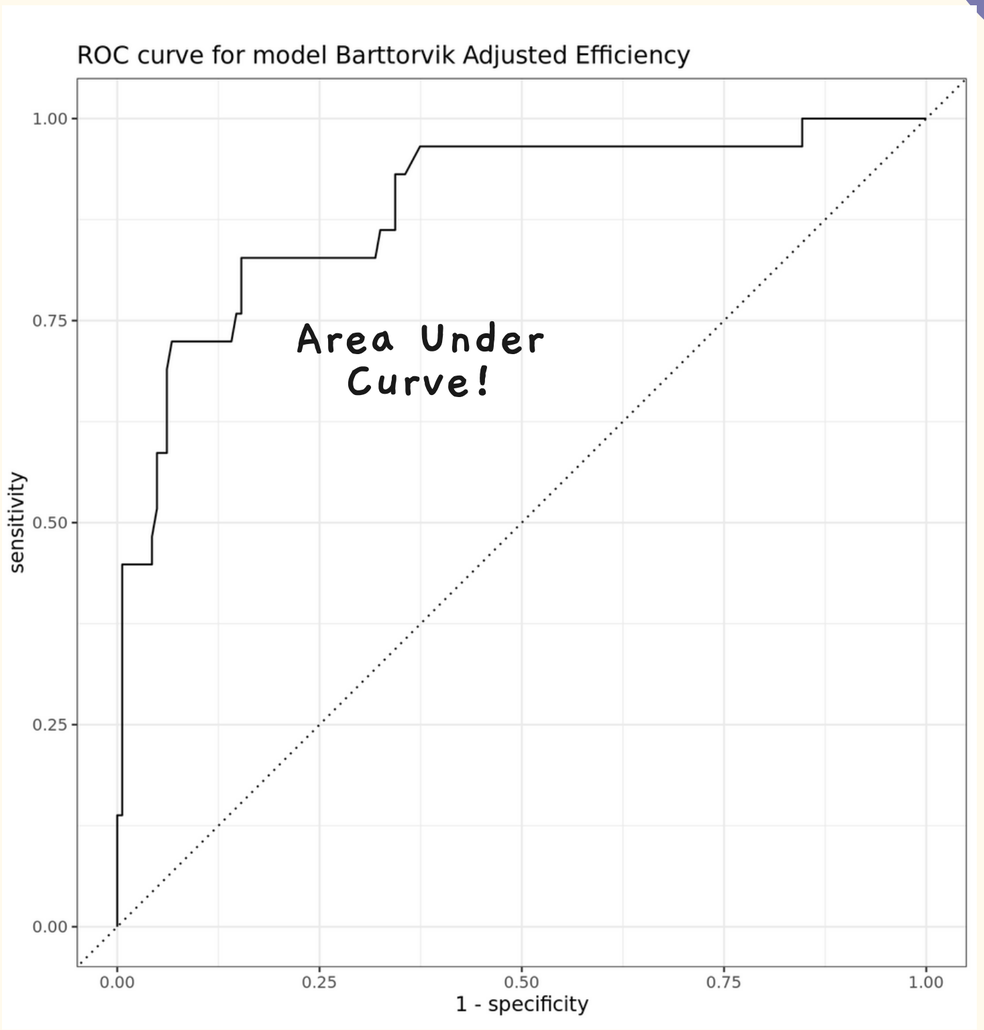
ROC

EDA

BARTTORVIK ADJUSTED EFFICIENCY



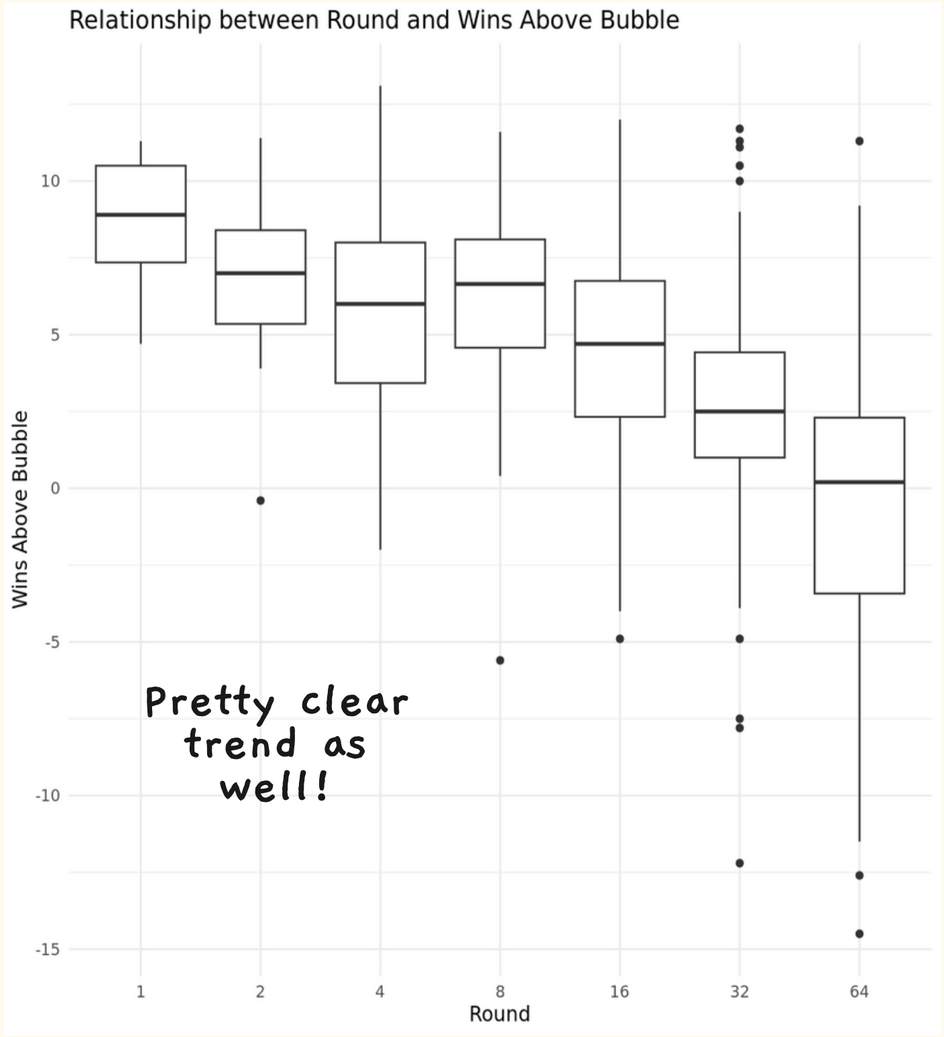
BOXPLOT



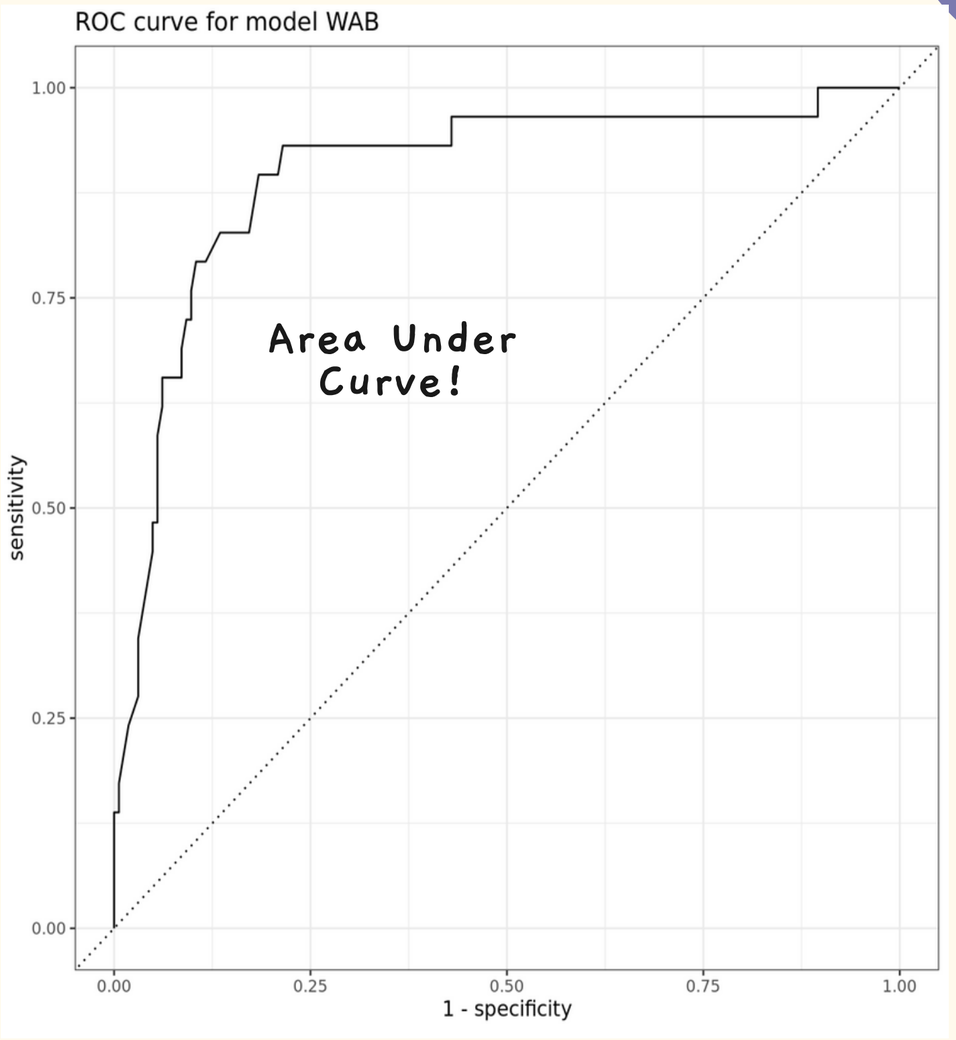
ROC

EDA

WINS ABOVE BUBBLE



BOXPLOT



ROC

RESULTS

SINGLE VARIABLE MODELS

1 BARTHAG MODEL

Equation:

$$log(\hat{p}/(1-\hat{p})) = 20.92 - 21.32 * BARTHAG$$

AUC Value:
0.894

2 BARTTORVIK ADJUSTED EFFICIENCY

Equation:

$$log(\hat{p}/(1-\hat{p})) = 6.09 - 0.208 * b\_adj\_eff$$

AUC Value:
0.891

3 WINS ABOVE BUBBLE

Equation:

$$log(\hat{p}/(1-\hat{p})) = 3.46 - 0.374 * wins\_above\_bubble$$

AUC Value:
0.898

RESULTS

TWO VARIABLE MODELS

1 BARTHAG + WAB

Equation:

$$log(\hat{p}/(1-\hat{p})) = 10.07 - 7.93 * BARTHAG - 0.26 * wins\_above\_bubble$$

AUC Value:
0.903

2 BARTHAG + BAE

Equation:

$$log(\hat{p}/(1-\hat{p})) = 4.13 + 2.80 * BARTHAG - 0.23 * b\_adj\_eff$$

AUC Value:
0.890

3 WAB + BAE

Equation:

$$log(\hat{p}/(1-\hat{p})) = 5.07 - 0.20 * wins\_above\_bubble - 0.12 * b\_adj\_eff$$

AUC Value:
0.901

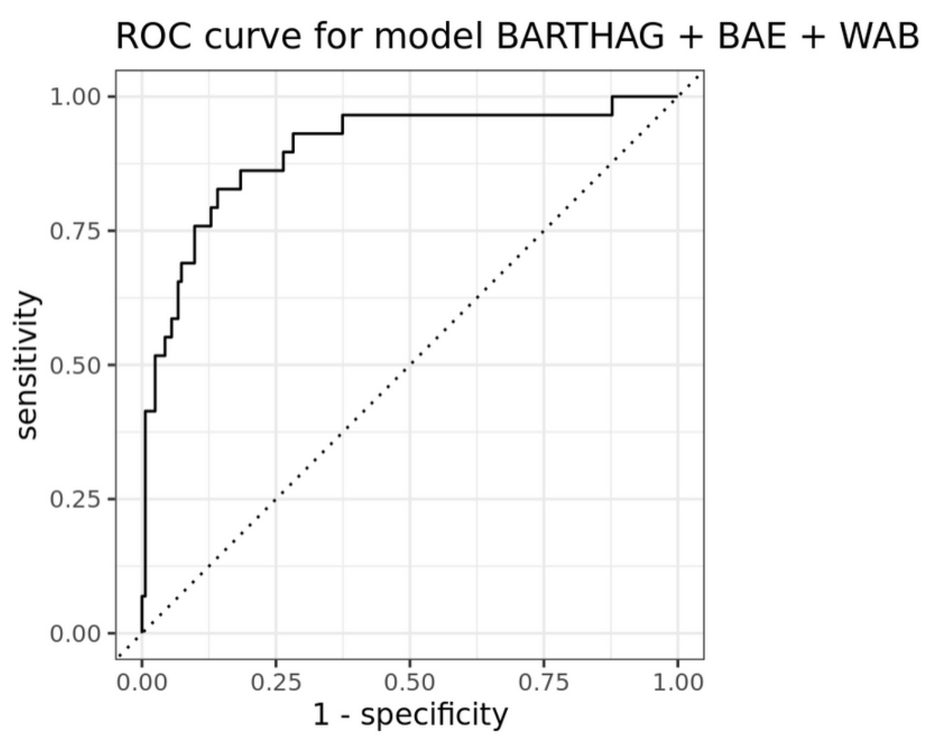
RESULTS

FULL MODEL

Equation:

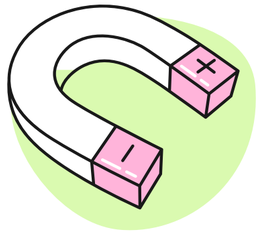
$$log(\hat{p}/(1-\hat{p})) = 2.86 + 3.15 * BARTHAG - 0.20 * wins\_above\_bubble$$
$$-0.14 * b\_adj\_eff$$

AUC:
0.900



COMMENTS

All models, no matter what variables used, performed relatively the same when comparing ROC / AUC



CONCLUSION

Though all models performed roughly the same, we gave our best model the data for the 2023 March Madness tournament and asked it to predict which teams would make the Elite Eight. The model got 2/8 of the teams correctly–Texas and Gonzaga. Our model is far from perfect, which is to be expected given how volatile the NCAA tournament can be.

pred_elite eight <dbl>	.pred_not elite ei... <dbl>	elite_eight <fctr>	team <chr>
0.5768321	0.4231679	not elite eight	Alabama
0.4895942	0.5104058	not elite eight	Kansas
0.4497347	0.5502653	not elite eight	Purdue
0.4368582	0.5631418	not elite eight	Houston
0.4116876	0.5883124	not elite eight	UCLA
0.3586346	0.6413654	elite eight	Texas
0.3219712	0.6780288	not elite eight	Arizona
0.3077924	0.6922076	elite eight	Gonzaga

DIGITAL MARKETING

THANK YOU

We hope you enjoyed!
~ heartemoji <3