THE FINAL FOURMULAS

Paul Kim, Lisa Lin, Yifan Wang, Jialuo Zhang

TEAM MEMBERS

Lisa Lin

Master of Science
Business Analytics
LinkedIn

Jialuo Zhang

Master of Science
Finance
LinkedIn

Paul Kim

Master of Science
Business Analytics
LinkedIn

Yifan Wang

Master of Science
Business Analytics
LinkedIn

Overview

PURPOSE

To predict the winner of the 2024 NCAA March Madness tournament

DATA

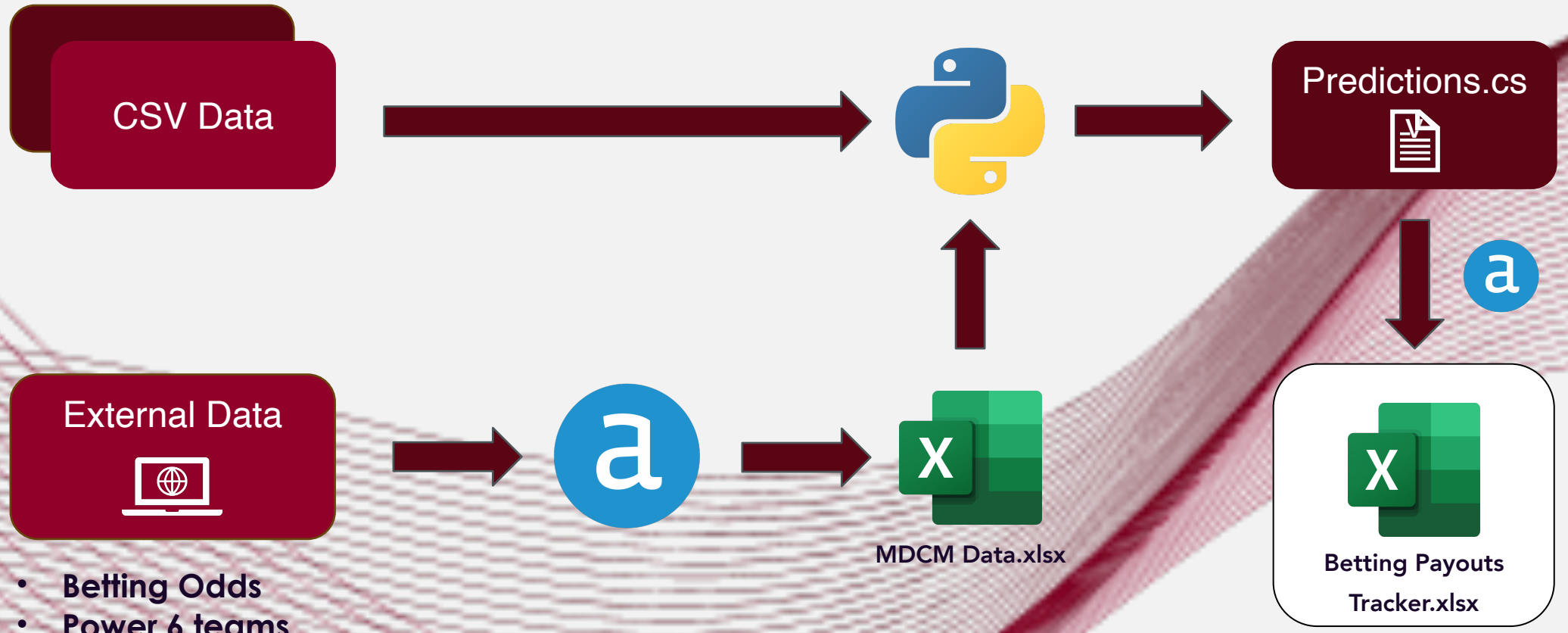
Team statistics for each team from the 2002-2023 NCAA tournaments

SCOPE

- **Tournament teams only** (not all D-1 teams)
- **No player or per-game data**
- Predictions based on **data available at beginning of tournament**

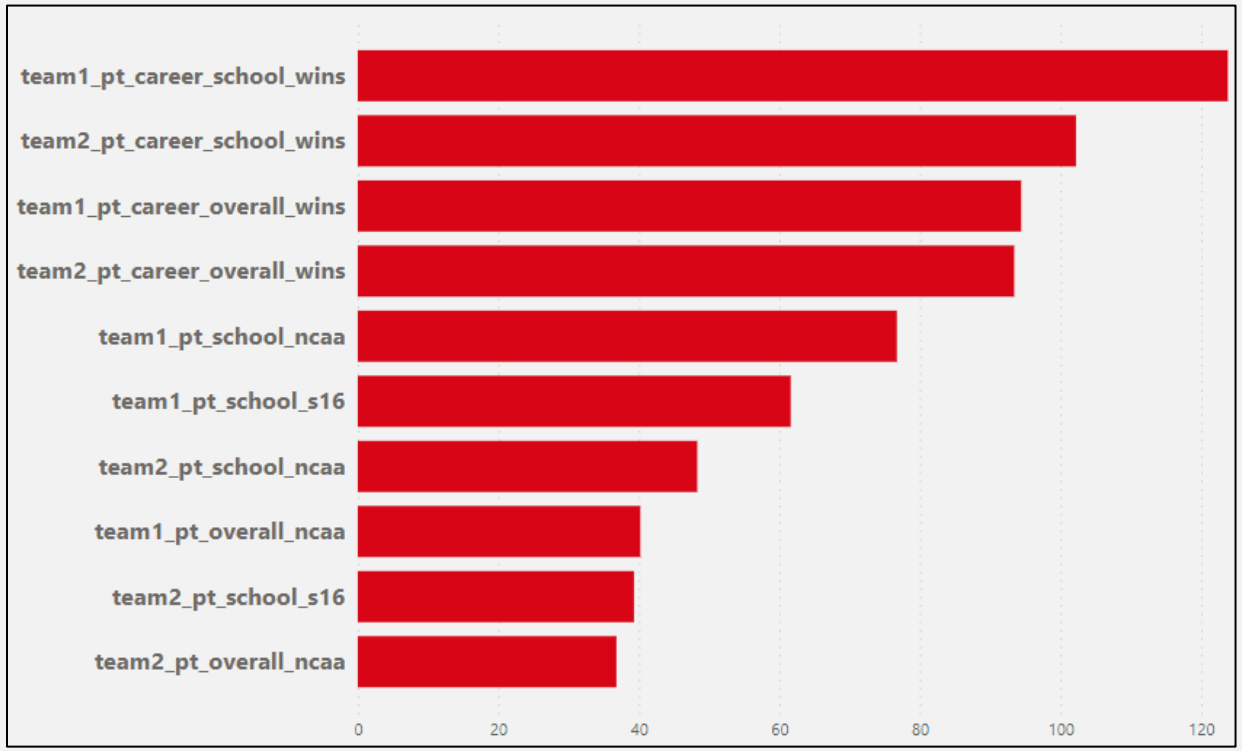
ETL & Modeling Roadmap

Data collection	Data prep, feature selection, & modeling	Analysis & application



- **Betting Odds**
- **Power 6 teams**
- **# All-Americans/team**

Removing Variables - VIF



Variance Inflation Factor (VIF): used to measure the correlation between independent variables

Multicollinearity: strong correlation between multiple variables, causing **redundancy.**

Redundant variables can lead to **overfitting data** & **incorrect predictions**.

Coach-related variables had the **highest VIF**, but they're also correlated with a coach's **age, years of experience, & tenure** at a school

Our solution: **removing most variables** & converting a select few to **ratios** (team winning %, coach winning %, etc.)

Data Preparation

The data was transposed so that each **winner** (team1) and **loser** (team2) is a **unique observation**.

P	Q	CO	CQ	CS	CU	CW	CY
team1_teamname	team2_teamname	team1_adjtempo	team1_adjoe	team1_adjde	team2_adjtempo	team2_adjoe	team2_adjde
North Carolina	Duke	70.1745	113.035	96.9911	67.4185	119.357	95.6444
Kansas	Villanova	69.1275	119.388	93.9009	62.5758	117.921	93.8099
Kansas	North Carolina	69.1275	119.388	93.9009	70.1745	113.035	96.9911
Kansas	Miami FL	69.1275	119.388	93.9009	67.3101	114.757	102.132
North Carolina	St Peter's	70.1745	113.035	96.9911	65.9923	98.7475	94.4745
Villanova	Houston	62.5758	117.921	93.8099	64.0049	117.238	91.368
Duke	Arkansas	67.4185	119.357	95.6444	70.6005	111.046	92.0063
St Peter's	Purdue	65.9923	98.7475	94.4745	65.8191	121.983	99.2547
Kansas	Providence	69.1275	119.388	93.9009	65.2257	111.928	98.0488
North Carolina	UCLA	70.1745	113.035	96.9911	65.5765	116.067	91.2261
Miami FL	Iowa St	67.3101	114.757	102.132	66.0992	104.353	90.4067
Arkansas	Gonzaga	70.6005	111.046	92.0063	72.5152	121.802	88.8352
Villanova	Michigan	62.5758	117.921	93.8099	66.3145	114.142	98.5016

B	C	D	V	W	X	Y
id	**seed**	**year**	**adjtempo**	**adjoe**	**adjde**	**game_id**
1104	2	2002	69.9001	111.4954	93.877	2002-1104
1194	15	2002	71.2446	96.8669	99.9263	2002-1104
1364	14	2002	64.7948	105.4534	97.6704	2002-1112
1112	3	2002	72.8207	117.3877	96.9262	2002-1112
1112	3	2002	72.8207	117.3877	96.9262	2002-1112
1461	11	2002	70.4124	106.037	96.6601	2002-1112
1335	11	2002	66.893	106.9622	95.1915	2002-1143
1143	6	2002	68.0084	108.4214	93.0913	2002-1143
1153	1	2002	68.7152	115.7842	88.0357	2002-1153
1131	16	2002	65.8612	99.5634	101.1011	2002-1153

This **reduced** the number of **variables** by 50% and **doubled** the number of **observations** in our data.

Adjusted Offense vs Adjusted Defense

Definitions

- Offensive Efficiency (OE): **Points scored** per 100 possessions
- Defensive Efficiency (DE): **Points allowed** per 100 possessions
- Adjusted OE (AdjO) & DE (AdjD) are an **estimate** of a team's OE & DE **against an average opponent**

Trends

- **17 of the last 20** champions were **top 10** in **AdjO**
- Since 2002, **every champion** was **Top 25** in **AdjD**

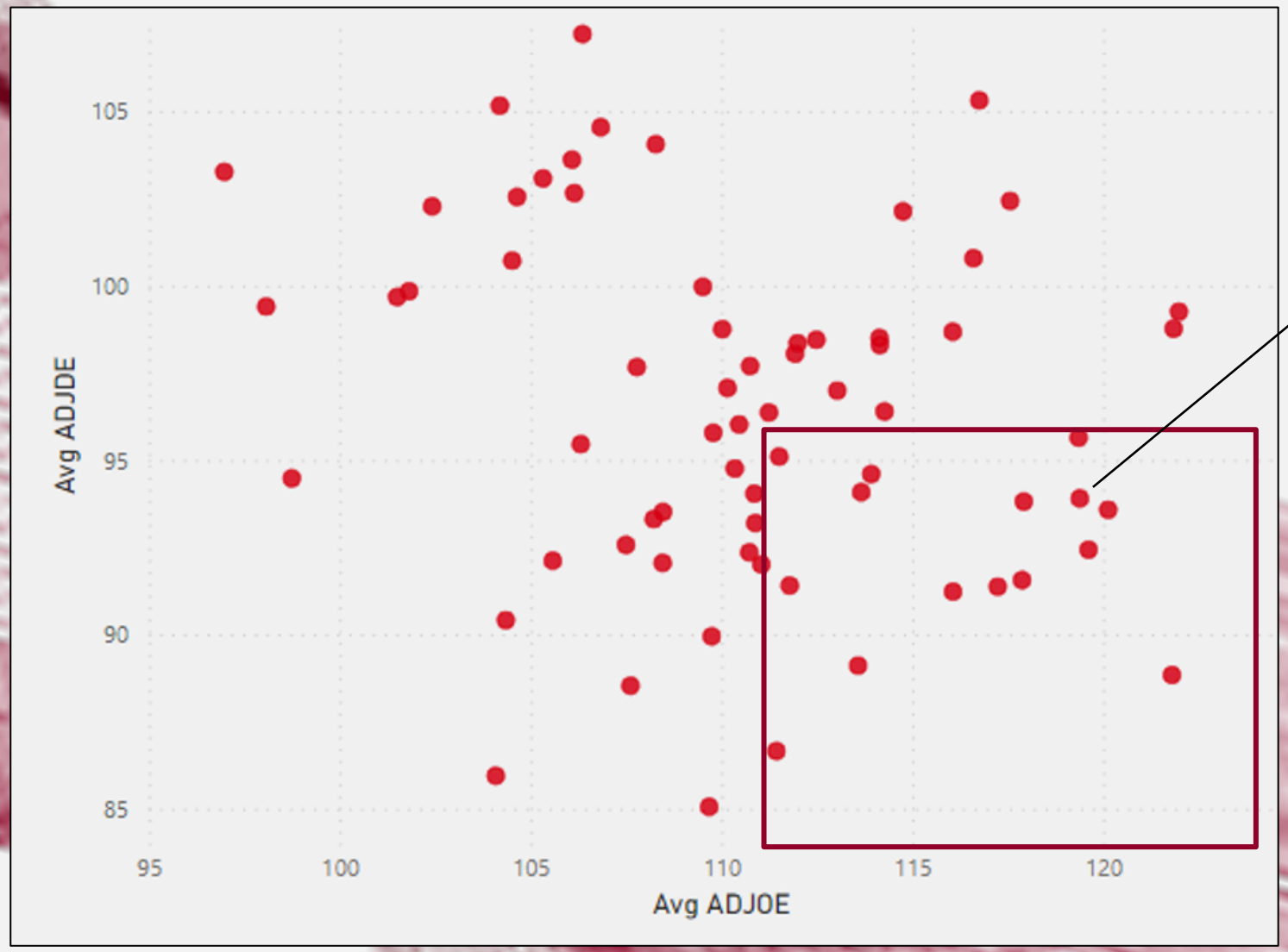
Source: AP News & The Next Round

Conclusion

Look for a
Top 10 Offense
and
Top 25 Defense

The Sweet Spot

2022 Teams Adj. Offense vs Adj. Defense



2022 champion,
Kansas
#6 in AdjO
#24 in AdjD

Top 10 Offense
Top 25 Defense

Higher AdjO is better

Lower AdjD is better

2024 Top Teams by AdjO

Team	Conf	W-L	AdjEM	AdjO		AdjD	
Connecticut 1	BE	31-3	+32.21	126.6	1	94.4	11
Alabama 4	SEC	21-11	+22.48	125.6	2	103.1	112
Illinois 3	B10	26-8	+23.99	125.6	3	101.6	93
Purdue 1	B10	29-4	+29.12	125.0	4	95.9	21
Kentucky 3	SEC	23-9	+20.02	122.7	5	102.7	108
Baylor 3	B12	23-10	+22.30	122.5	6	100.2	64
Duke 4	ACC	24-8	+24.88	121.8	7	97.0	26
Arizona 2	P12	25-8	+26.62	121.1	8	94.5	12
Gonzaga 5	WCC	25-7	+21.56	121.0	9	99.4	46
Auburn 4	SEC	27-7	+28.90	120.6	10	91.7	4

Source: KenPom.com. Ratings are accurate as of 3/18/24.

Only **4 teams** are **Top 10 AdjO, Top 25 AdjD**

Other Promising Teams

#7 in AdjO
#26 in AdjD

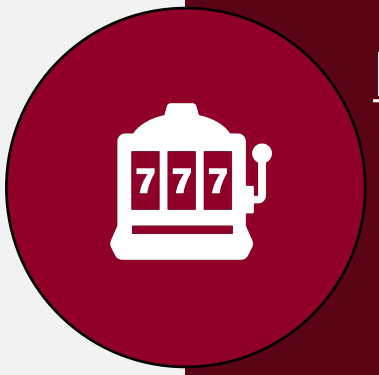
#17 in AdjO
#2 in AdjD

#12 in AdjO
#24 in AdjD

New Data

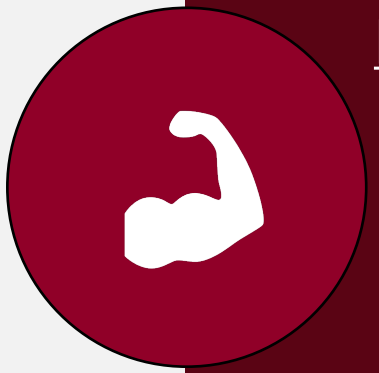
Purpose of attaining new data:

- Capturing **public perception** of a team's expected performance
- Accounting for the **disparity in funding**, **resources**, and **overall success** in college athletics between "big" schools and others
- Examining the impact of **star players** on team success in March



Betting Odds

- **Futures bets** for all **R64 games** from the 2002-2023 NCAA tournaments



Power 6 Teams

- Teams associated with the **6 premier conferences*** in college basketball
- Assigned designation by season to account for **conference realignment**

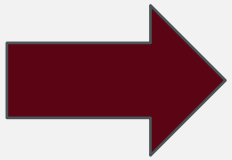
All-Americans

- The number of **All-Americans** on each team for each season

See Appendix for more details

Selecting Features – PCA

Rank	Feature
1	exp_win
2	log5
3	seed_diff
4	seed_tier_1
5	adjoe
6	adjde
7	team_season_win_%
8	Num_AA
9	coach_career_win_%
10	coach_s16_rate
11	Power 6_1.0
12	opp_shooting%
13	oppfg2pct
14	rd1_odds



Rank	Feature
1	log5
2	seed_tier_1
3	adjde
4	team_season_win_%
5	Num_AA
6	coach_career_win_%
7	coach_s16_rate
8	Power 6_1.0
9	opp_shooting%
10	rd1_odds

We used **principal component analysis** (PCA) to determine the **most relevant features** in the **training data**.

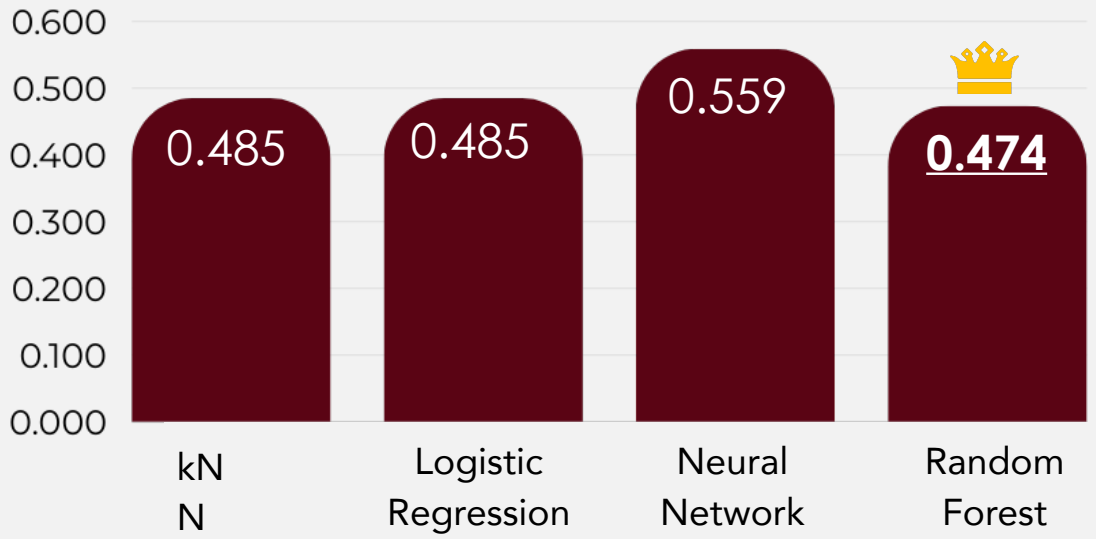
However, multicollinear features were removed to prevent potential overfitting.

Definitions of each feature can be found in the **Appendix**.

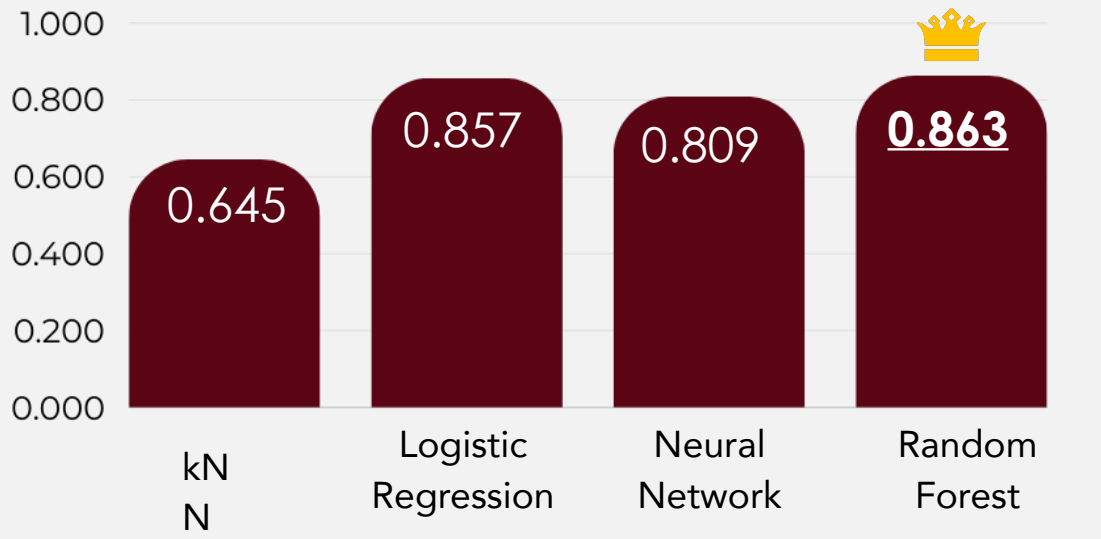
Model Performance

On 2019 test data

Log Loss



AUC



Model	Log Loss	Accuracy	AUC	F-Score	Precision	Recall
kNN	0.485	0.658	0.645	0.667	0.636	0.700
Logistic Regression	0.485	0.754	0.857	0.756	0.750	0.761
Neural Network	0.559	0.724	0.809	0.720	0.720	0.720
Random Forest	**0.474**	**0.776**	**0.863**	**0.762**	**0.814**	0.716

Our Final 4

UConn
Huskies
#1 Seed
EAST

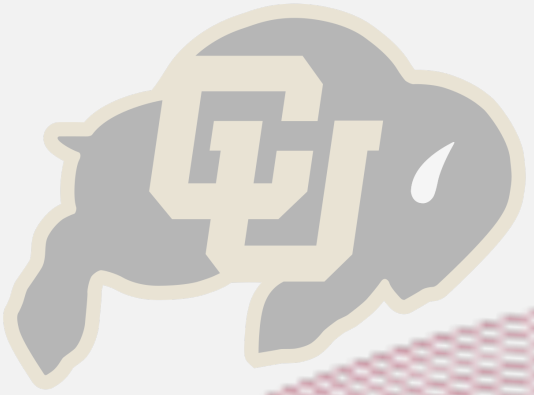
Incumbent

Arizona
Wildcats
#2 Seed
WEST

National Champion

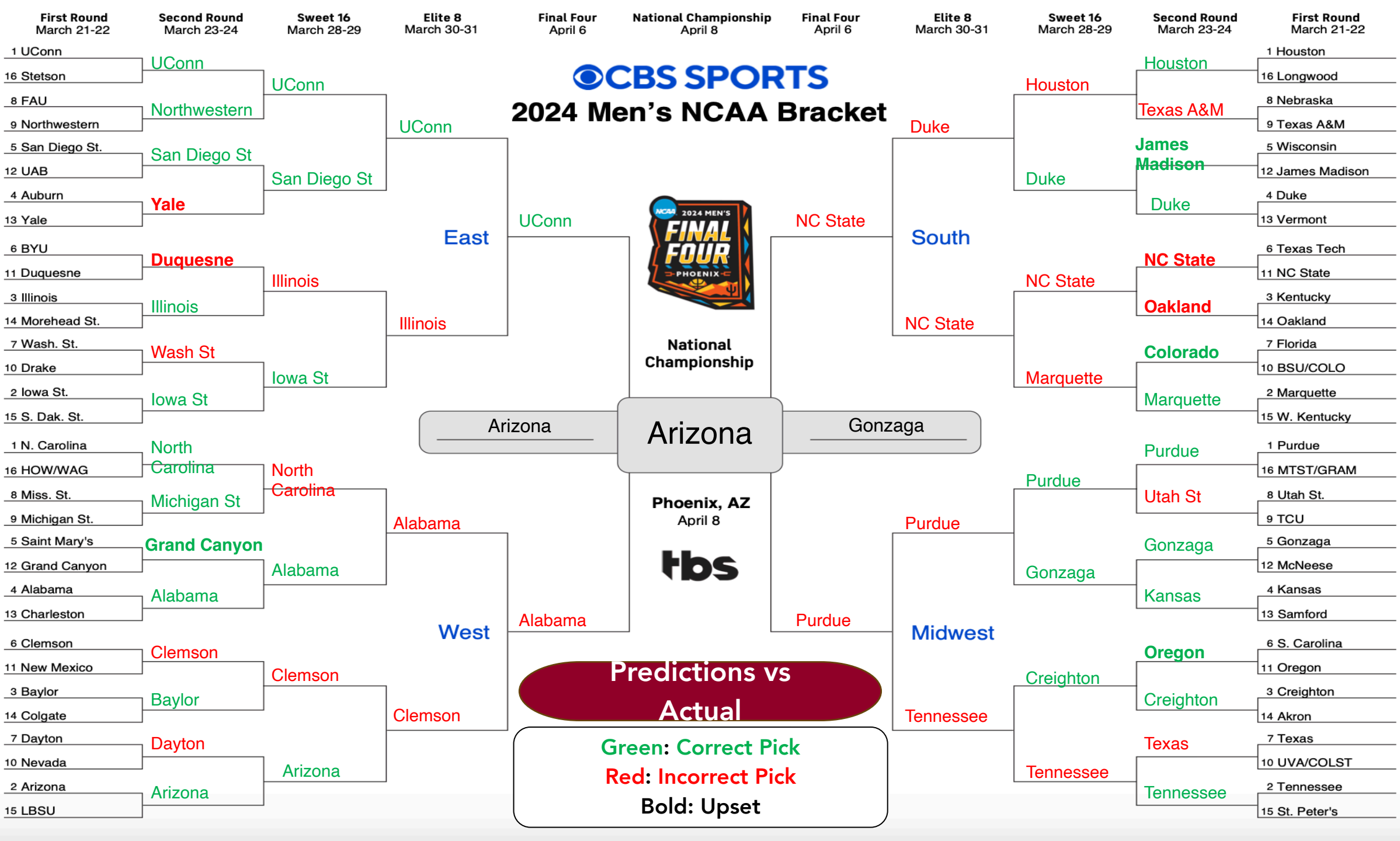
Gonzaga
Bulldogs
#5 Seed
MIDWEST

Runner-Up



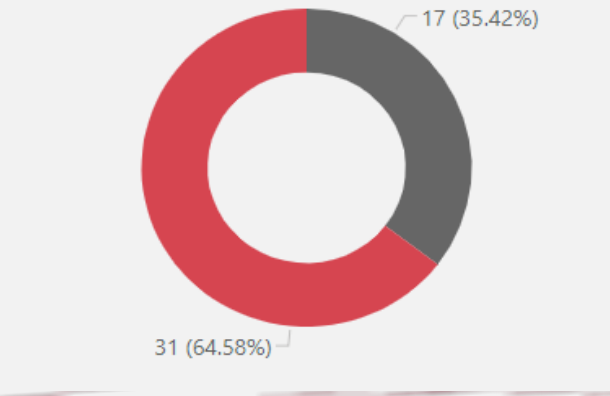
Colorado
Buffaloes
#10 Seed
SOUTH

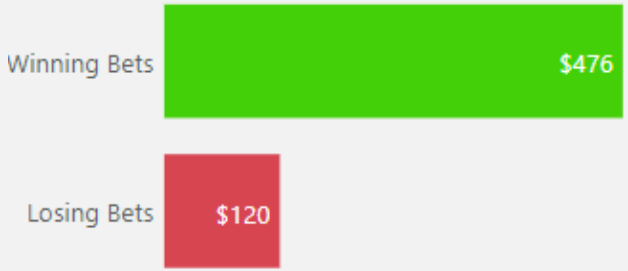
Cinderella


First Round March 21-22
Second Round March 23-24
Sweet 16 March 28-29
Elite 8 March 30-31
Final Four April 6
National Championship April 8
Final Four April 6
Elite 8 March 30-31
Sweet 16 March 28-29
Second Round March 23-24
First Round March 21-22
CBS SPORTS
2024 Men's NCAA Bracket
1 UConn
16 Stetson
8 FAU
9 Northwestern
5 San Diego St.
12 UAB
4 Auburn
13 Yale
6 BYU
11 Duquesne
3 Illinois
14 Morehead St.
7 Wash. St.
10 Drake
2 Iowa St.
15 S. Dak. St.
1 N. Carolina
16 HOW/WAG
8 Miss. St.
9 Michigan St.
5 Saint Mary's
12 Grand Canyon
4 Alabama
13 Charleston
6 Clemson
11 New Mexico
3 Baylor
14 Colgate
7 Dayton
10 Nevada
2 Arizona
15 LBSU
UConn
Northwestern
San Diego St
Yale
Duquesne
Illinois
Wash St
Iowa St
North Carolina
Michigan St
Grand Canyon
Alabama
Clemson
Baylor
Dayton
Arizona
UConn
San Diego St
Illinois
Iowa St
North Carolina
Alabama
Clemson
Arizona
East
West
UConn
Illinois
Alabama
Clemson
2024 MEN'S FINAL FOUR PHOENIX
National Championship
Arizona
Arizona
Gonzaga
Phoenix, AZ
April 8
tbs
NC State
Duke
NC State
Purdue
Tennessee
South
Midwest
Houston
Duke
NC State
Marquette
Purdue
Gonzaga
Creighton
Tennessee
Houston
Texas A&M
James Madison
Duke
NC State
Oakland
Colorado
Marquette
Purdue
Utah St
Gonzaga
Kansas
Oregon
Creighton
Texas
Tennessee
1 Houston
16 Longwood
8 Nebraska
9 Texas A&M
5 Wisconsin
12 James Madison
4 Duke
13 Vermont
6 Texas Tech
11 NC State
3 Kentucky
14 Oakland
7 Florida
10 BSU/COLO
2 Marquette
15 W. Kentucky
1 Purdue
16 MTST/GRAM
8 Utah St.
9 TCU
5 Gonzaga
12 McNeese
4 Kansas
13 Samford
6 S. Carolina
11 Oregon
3 Creighton
14 Akron
7 Texas
10 UVA/COLST
2 Tennessee
15 St. Peter's
Predictions vs Actual
Green: Correct Pick
Red: Incorrect Pick
Bold: Upset

Sweet Moments

Model accuracy up to the Sweet 16





31

Correct predictions in the 1st 2 rounds

4

Correctly-predicted upsets

$356

Total Earnings per [Betting Odds Tracker]

THANK YOU

FORDHAM
UNIVERSITY

Appendix

Final Features – Definitions

log5	The probability of a team winning its matchup by considering each team's AdjO and AdjD
seed_tier_1	A team's grouping based on its seed. Teams with seeds 1-4 are categorized into Tier 1, seeds 5-8 belong to Tier 2, and so on.
adjde	Adjusted Defensive Efficiency (AdjD) - points allowed per 100 possessions
team_season_win_%	Team's regular season winning percentage
Num_AA	Number of All-American players on a team
coach_career_win_%	Coach's career winning percentage
coach_s16_rate	Coach's Sweet 16 appearances / total NCAA tournament appearances
Power 6_1.0	Whether a team is in one of the Power 6 conferences (ACC, BE, B10, B12, P12, SEC)
opp_shooting%	Weighted average of a team's opponents' FT% (17%), 2pt FG% (33%), and 3pt FG% (50%)
rd1_odds	Futures bet for each team just before the beginning of the NCAA tournament

Snubbed – Winning **Isn't** Everything

Why didn't Indiana State make the tournament, even though they **won more games**?

Florida played in a **better conference** & played **better teams.**

- Indiana State Sycamores
- 2024 Record: 28-6
- Conference Champion*: No
- In NCAA Tournament: **No**

- Florida Gators
- 2024 Record: 24-11
- Conference Champion*: No
- In NCAA Tournament: **Yes**

*Conference champions get **automatic bid** into the tournament.

Conference: A group of teams that play against each other.

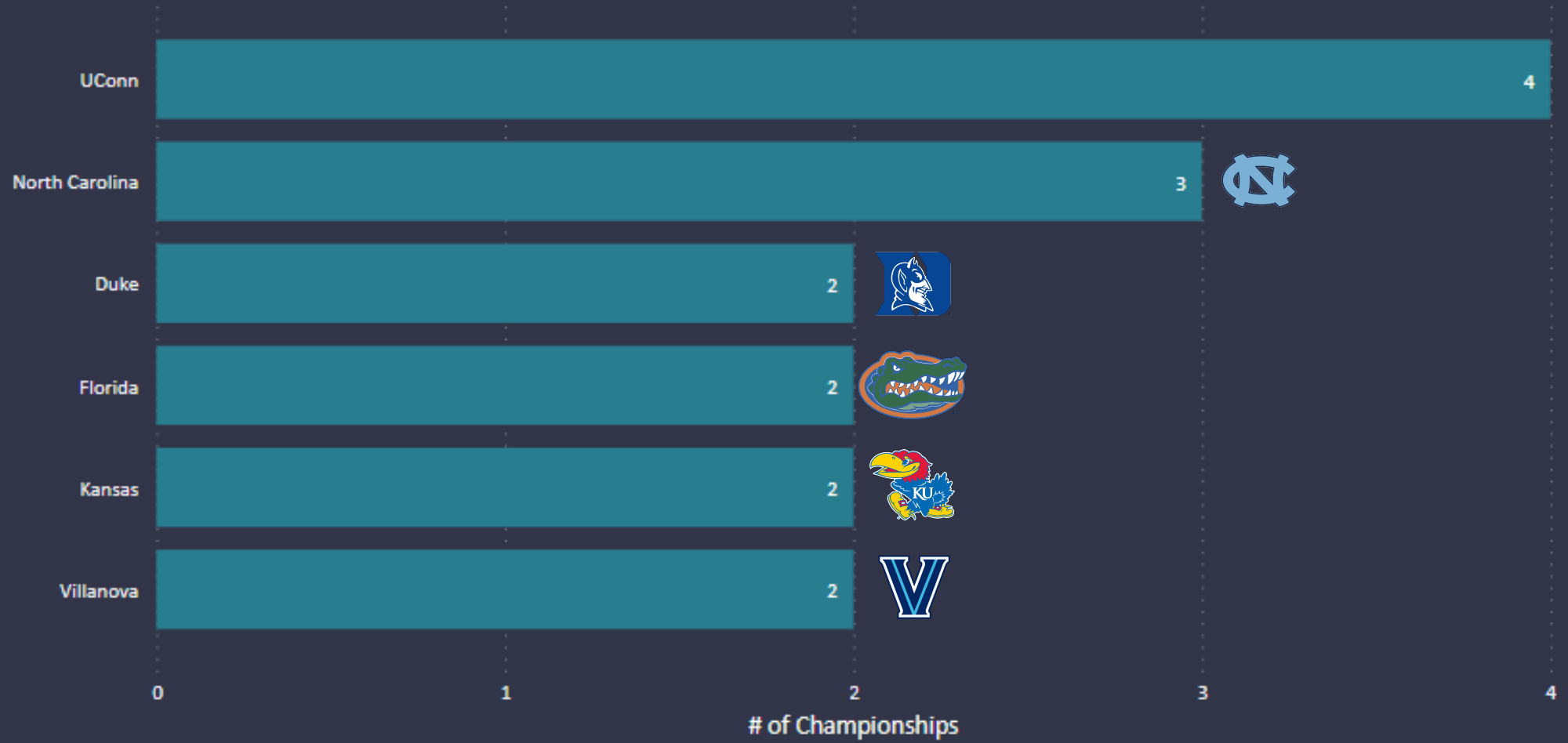
There are **6 conferences** called the **Power 6** that are considered **more competitive** than the others:

ACC, **Big East** (basketball only), **Big10**, **Big12**, **SEC**, **Pac-12**

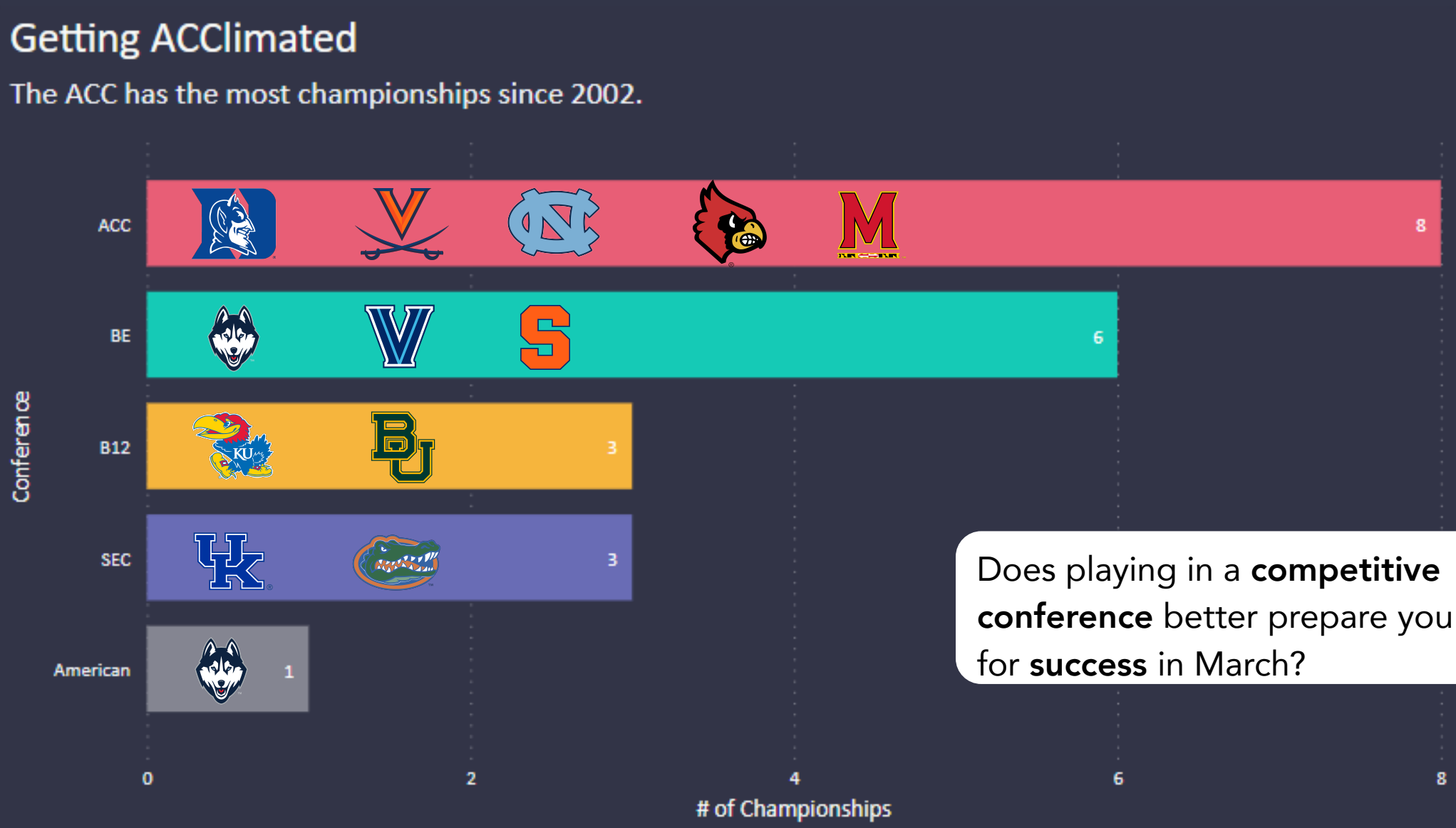
Historical Trends: Best Teams

Who Let the Dogs Out?

The UConn Huskies have the most championships since 2002.



Historical Trends: Best Conferences



Does playing in a **competitive conference** better prepare you for **success** in March?

Canon Events: The Steph Curry Effect

Let Him Cook!

Teams are shooting 3-pointers at a much higher rate since 2016.