

# Making March Not So Mad

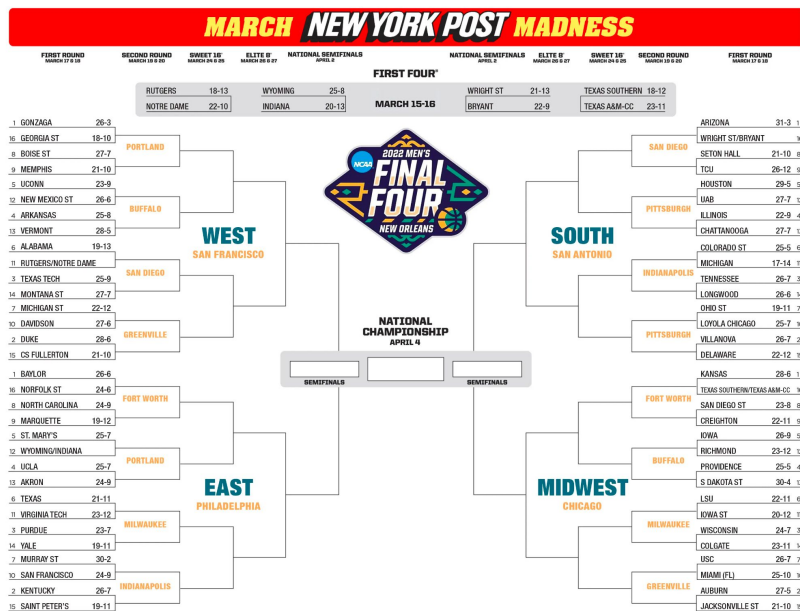
A Beginners Guide into March Madness and Model Selection



**What is March Madness?**

# March Madness Explained

- Postseason College Basketball Tournament
- 68 Teams
- Single Elimination
- Upset-frequent event



# The Freakiness of March

# 12.4



# 1.2



# 1.3



143-1





74

16



FINAL

1



54

SPECTRUM CENTER - CHARLOTTE, N.C.





SOUTH REGION

2ND ROUND

15

LEHIGH

75

NCAA  
MARCH MADNESS

70

DUKE

2

10

GREENSBORO, NC

FINAL

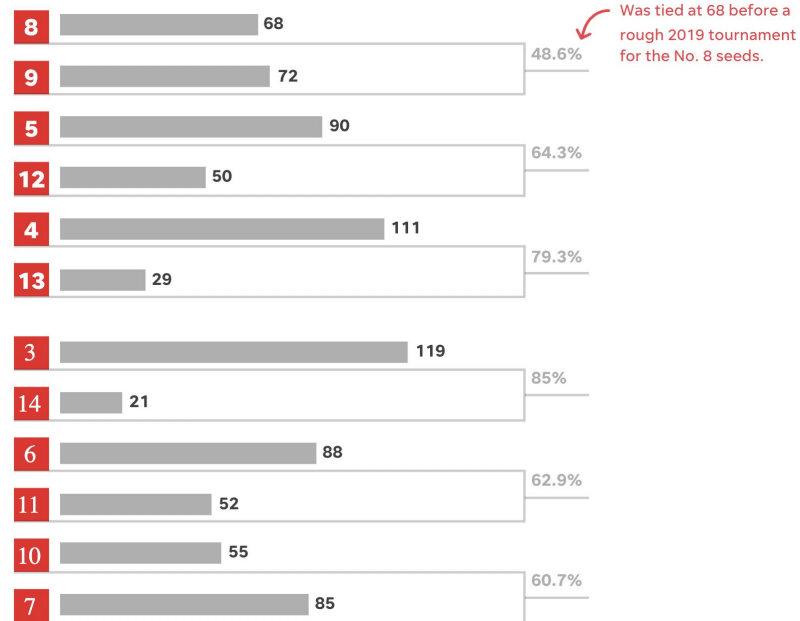
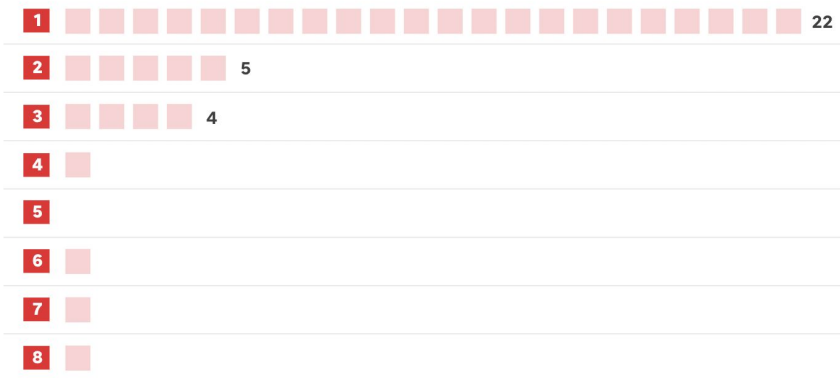
# Bracketology Tips



# Tips

## Seeding of NCAA national championship winner

Seed 1985 to 2019



**Hey TJ We Came for  
Model Building!**





# Making A Model

- Obtain
- Clean
- Test and Train
- Model Selection Methods
  - What are we solving for???
  - Classification vs Regression



# Potential Data Sources

- <https://kenpom.com/>
  - Past Years
- <https://www.kaggle.com/competitions/mens-march-mania-2022/data>
- GitHub

**Code With Me**







# Superior Methods of Model Building

- Gradient Boosting
  - Xgboost Package in R
  - Popular model for most Kaggle competitions
  - Weak Learners – simple decision trees that are constructed based on purity scores (e.g., Gini).
  - Loss Function – a differentiable function you want to minimize. In regression, this could be a mean squared error, and in classification, it could be log loss.
  - Additive Models – additional trees are added where needed, and a functional gradient descent procedure is used to minimize the loss when adding trees.

**Questions??**