March Madness Mania

Predicting the Mayhem

Eric Choi, Vishesh Patel, Ravikiran Ramjee

March Madness

- College basketball season
  - Regular season
  - Playoffs: March Madness
- Brackets and sports betting

NCAA MARCH MADNESS

Objective

Build a model to help basketball enthusiasts predict March Madness tournament results

The Dataset

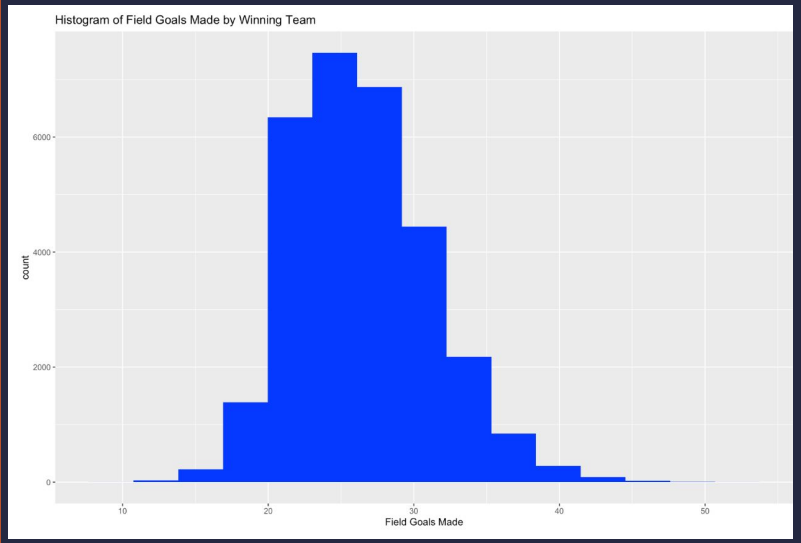
- Consolidated dataset on Kaggle
  - Regular Season Games (2003 - 2022)
  - March Madness Games (2003 - 2021)
  - Tournament Seeds
  - Teams
  - Conferences
  - Cities
  - Coaches
  - Consolidated Ordinal Rankings

Data Wrangling: Challenges

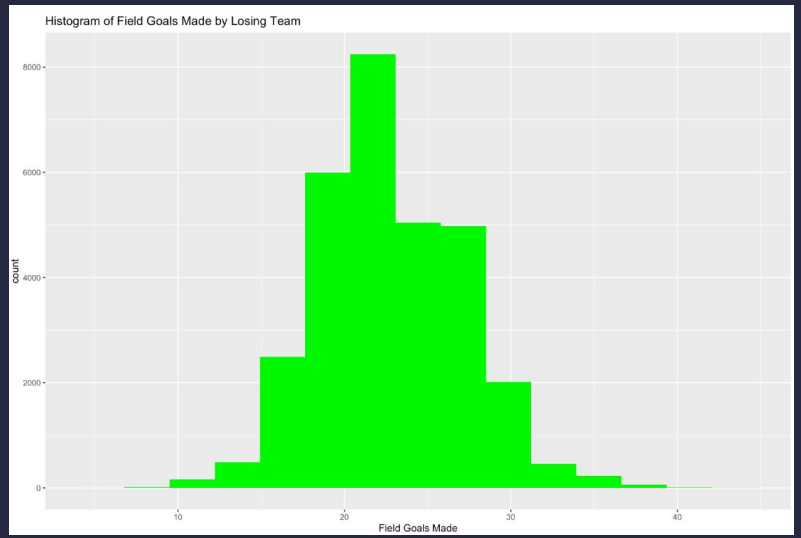
- Merging a multitude of tables together to create 2 uniform datasets for training
- Creating a one-hot encoding for categorical variables
- Randomizing the order of “A” columns and “B” columns
- Averaging the data for prediction

Exploratory Data Analysis

Field Goals Made by Both Teams

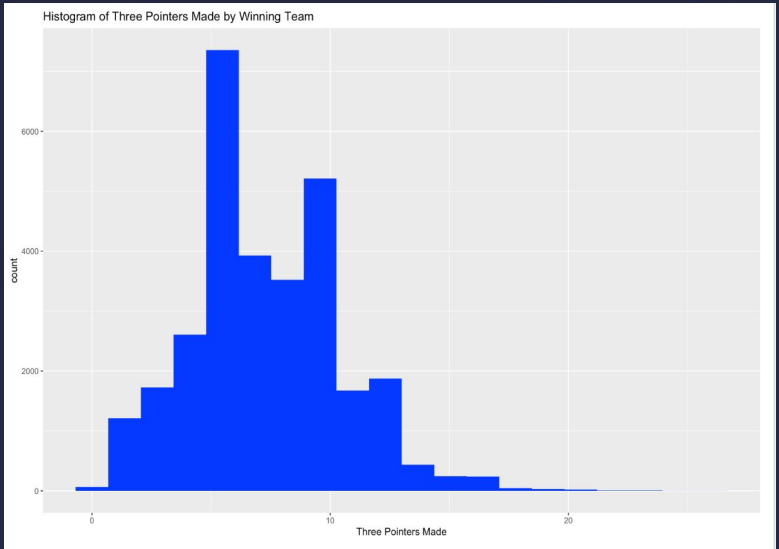


Mean: 26.62
Median: 26.00

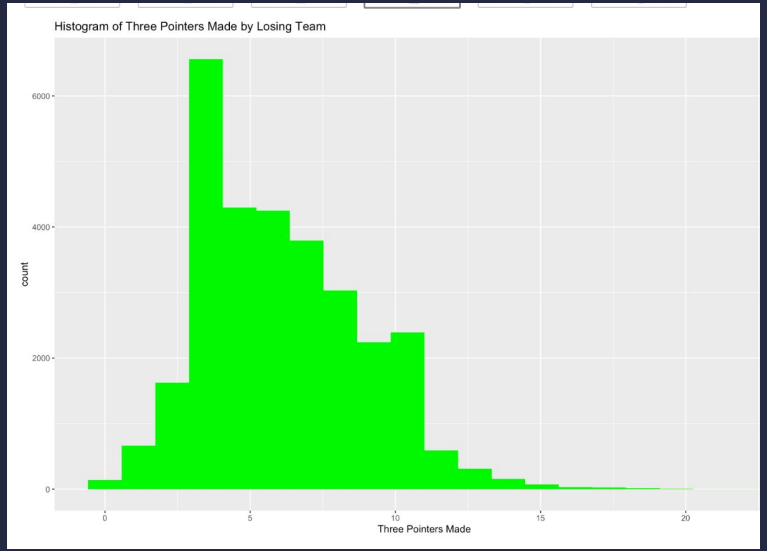


Mean: 23.02
Median: 23.00

Three Pointers Made by Both Teams

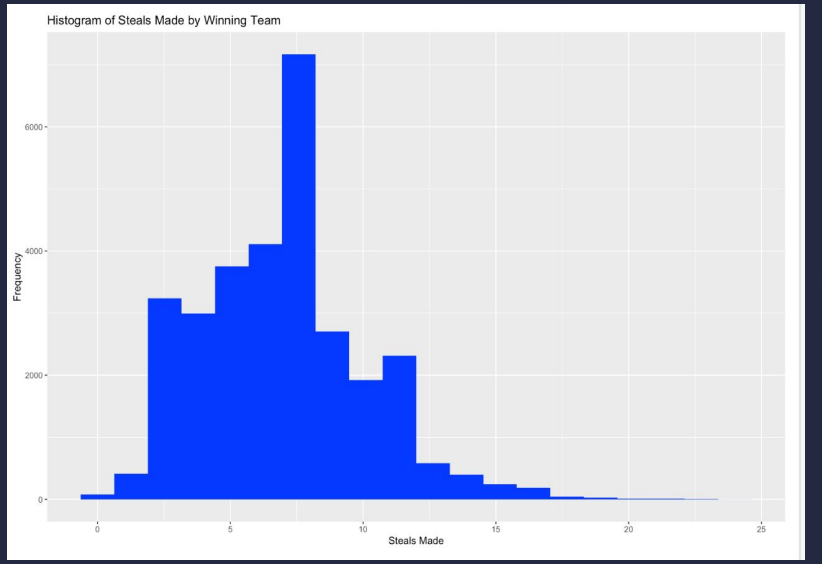


Mean: 6.993
Median: 7.000

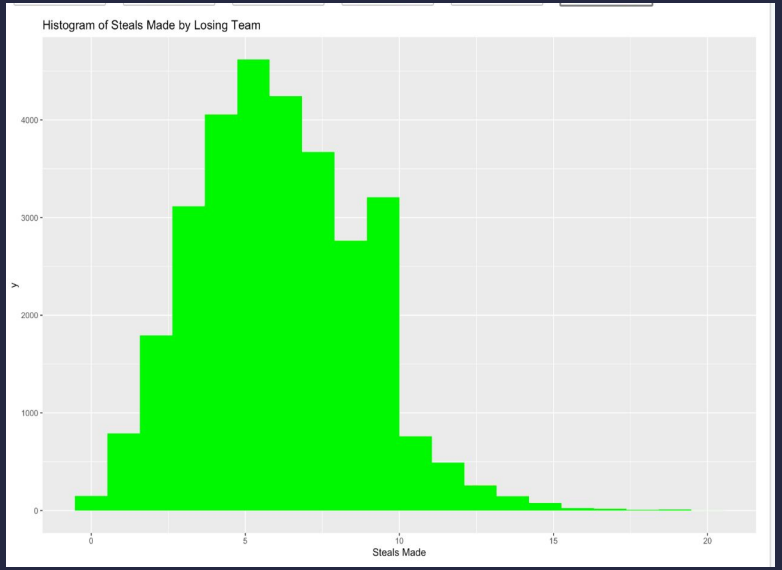


Mean: 6.208
Median: 6.000

Steals Made by Both Teams

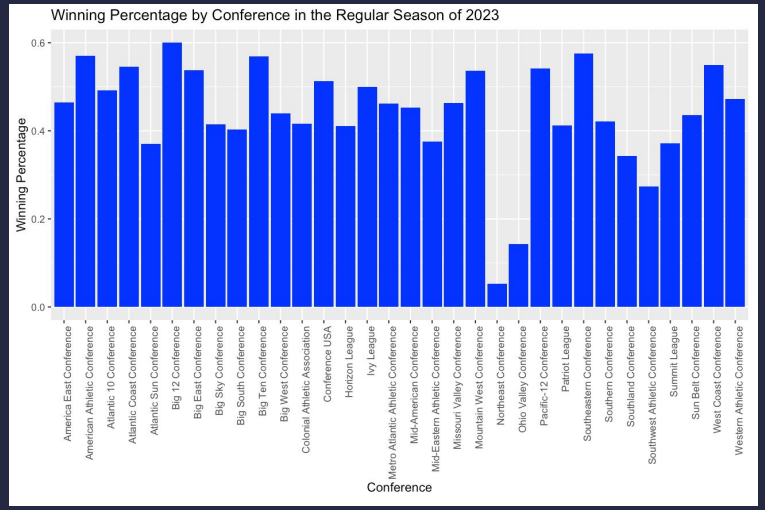


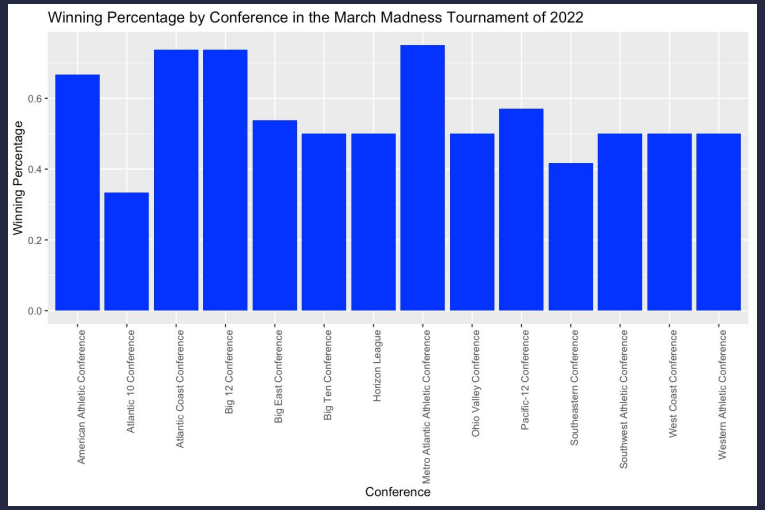
Mean: 6.343
Median: 6.000



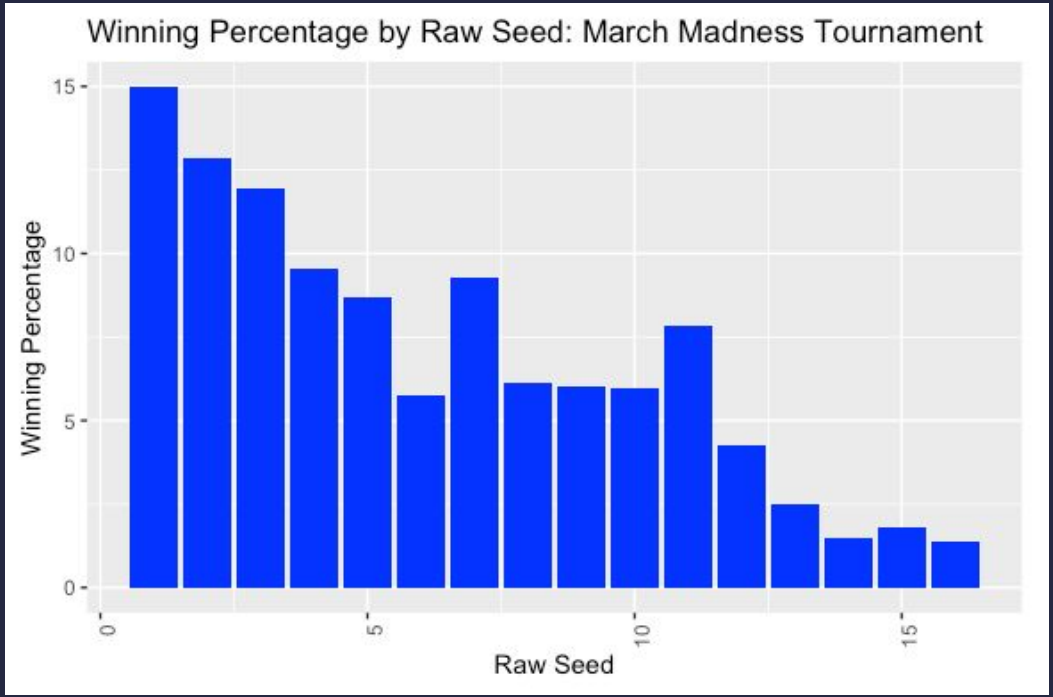
Mean: 5.641
Median: 5.000

Conferences vs. Winning



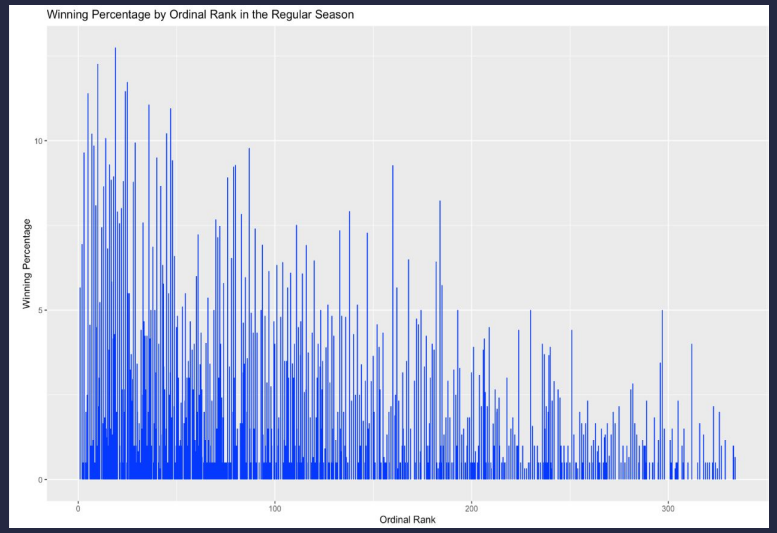


Seed vs. Winning Percentage

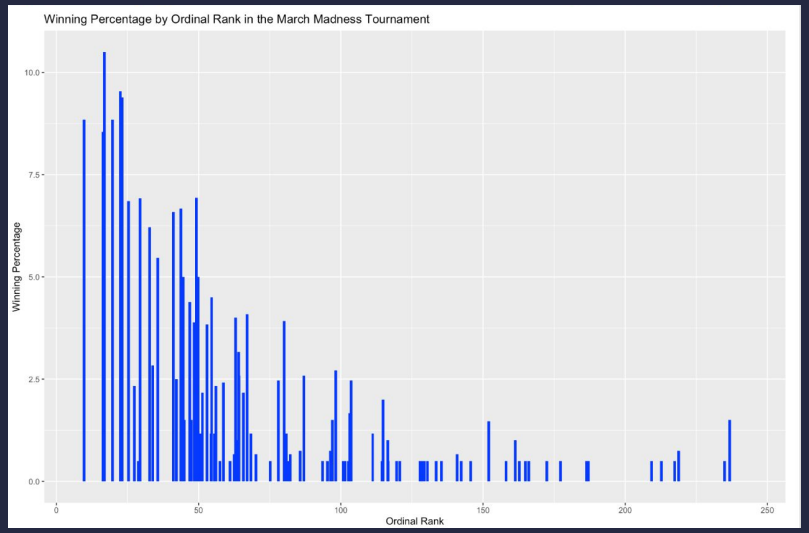


Ordinal Rank vs. Winning Percentage

Regular Season



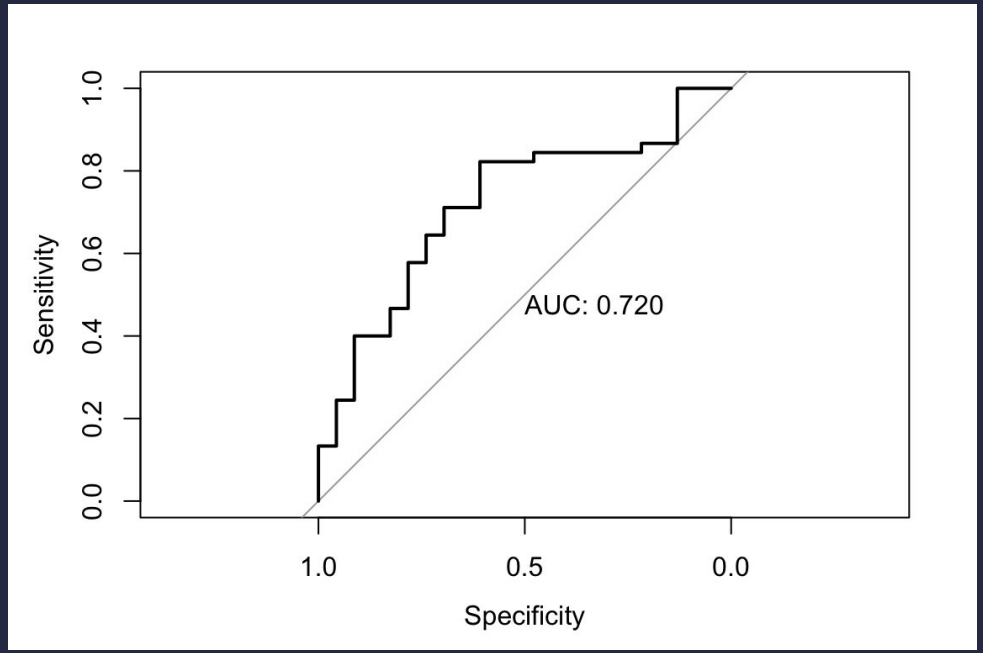
March Madness



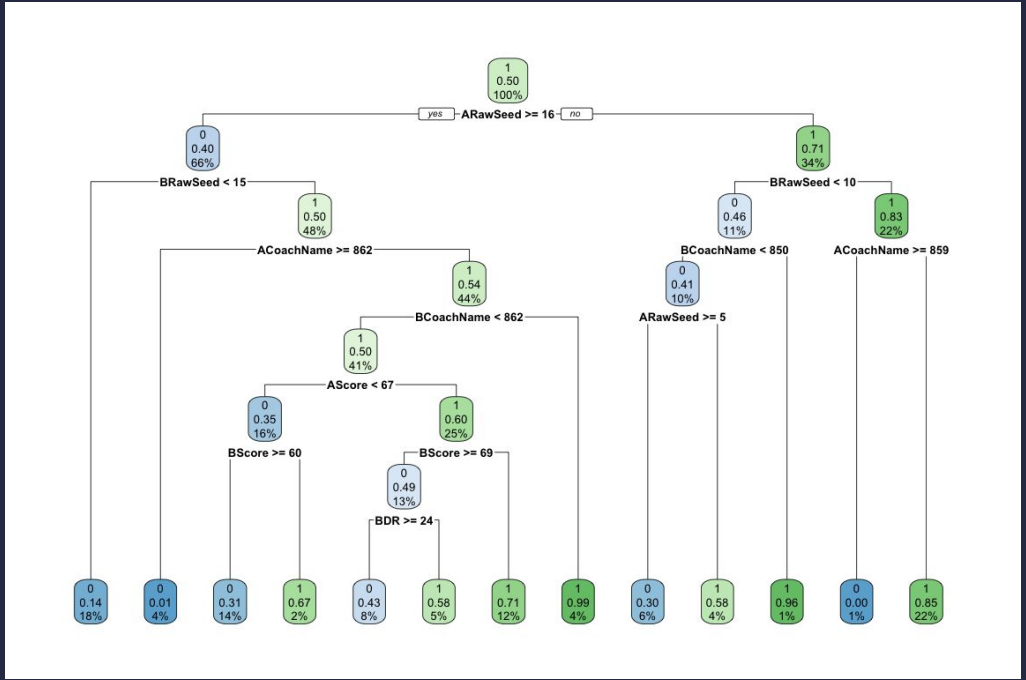
Analysis and Results

Preliminary Models

Logistic Regression for Binary Classification

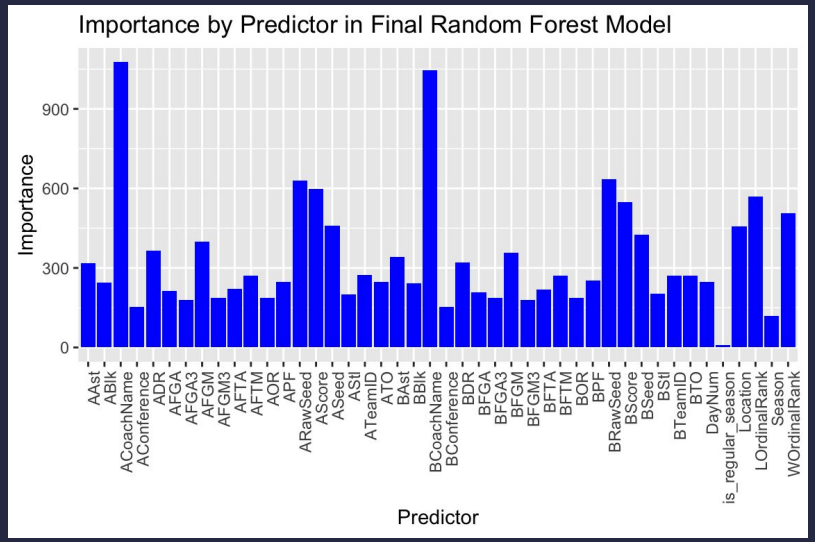


Decision Trees as a Model Free Approach



Final Model and Prediction Accuracy

- Random Forest model with 48 predictors and response of a win or loss for a matchup of two teams
- Prediction Accuracy: 82.4% for 2022 March Madness Tournament



Summary and Future Outlook

- Random Forest model accurately captures results of March Madness tournament for game outcomes of 2022
- Future steps to refine model:
  - Feature engineering to include factors such as injuries and rest time
  - Further hyperparameter tuning

Thank You!