# Exploratory Data Analysis

**Kolby Taylor and Carter Hale**

## 1. Research Question and Dataset Overview

This project aims to predict single-game and overall outcomes for the NCAA Men's Basketball Championship (March Madness) using historical data and machine learning techniques. We will develop both supervised and unsupervised ML models to analyze tournament results.

Our dataset is a compilation of various March Madness sources available on Kaggle. The primary sources include:

- KenPom
- BartTorvik
- Heat Check CBB
- FiveThirtyEight
- ESPN
- College Poll Archive
- Yahoo Sports
- EvanMiya
- TeamRankings

The dataset is in the public domain, as stated by Kaggle:

> "The person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can copy, modify, distribute, and perform the work, even for commercial purposes, all without asking permission."

There are no ethical concerns regarding the use of this data.

The features of the data set are based on each team's regular season performance. These include metrics that describe the team's style of play, their strengths and weaknesses, their strength of resume, and how well they did during the season. Some of these metrics will be described below.

## 2. Data Preprocessing

The dataset required several preprocessing steps before modeling:

- **Handling Missing Values & Duplicates**: Since much of the data was preprocessed, minimal adjustments were needed. We removed a few duplicates.
- **Validation Set**: The dataset does not include 2024 tournament results, so we will use that season as a validation set to assess model performance.
- **Eliminating Metrics**: The dataset includes statistics for home, away, and neutral games. To simplify feature selection, we tested whether using overall metrics alone was sufficient. We trained an XGBoost model to predict the number of tournament rounds a team would win using different feature sets. The overall model outperformed the others, with home and neutral metrics producing identical RMSE and away slightly worse. Based on these results, we chose to use overall metrics only.
- **Eliminating Rankings:** Each metric came with a corresponding feature that ranked it against all of Division 1. We took these out, since we figured it would be a little redundant and that the raw features were better.

- **Merging Data Sources**: We combined all sources into a single dataset containing team-level statistics, shooting splits, and rankings by season. Redundant columns were dropped.
- **Tournament Seeding Adjustment**: To differentiate teams with identical seeds, we redefined tournament seeding from 1-16 (in four regions) to a single ranking from 1-64.
- **Matchup-Level Dataset**: We created a secondary dataset representing each tournament matchup. Team 1 is the higher-seeded team, and Team 2 is the lower-seeded team. We computed feature differences between the two teams (e.g., if "Points Per Game" = 7, Team 1 averaged 7 more points per game than Team 2 during the regular season).

This preprocessing ensures our data is structured for effective modeling and tournament predictions.

**Features**

For most of the features below, there exists another feature that reflects what the team allows from their opponents or what the team does on defense.

| Metric | Description |
|---|---|
| Team | Team name. |
| Conference | The conference the team pertains to. |
| Tempo | Possessions per game. |
| Year | Year of the tournament. |
| Seed | What seed the team was (1-64). |
| Adjusted Tempo | The speed, in possessions per 40 minutes, that the team would expect to have against the average Division 1 team. |
| Offensive/Defensive Efficiency | Points/points allowed per 100 possessions. |
| Adjusted Offensive/Defensive Efficiency | Points/points allowed per 100 possessions, adjusted for their schedule. |
| Adjusted EM | Estimates how much a team would outscore the average Division 1 team over 100 possessions. |
| Adjusted Offense/Defense | Estimates how many points a team would score/allow against the average team over 100 possessions. |
| Power Rating (BARTHAG) | Change of beating the average team. |
| Games | Total games played. |
| Wins | Total wins. |
| Losses | Total losses. |

| | |
|---|---|
| Win % | Percentage of games won. |
| Effective Field Goal % | Percentage of shots made, but adjusts for three pointers being more valuable by making them work 1.5 as opposed to 1. |
| Free Throw Rate | The ratio of free throws to their overall shot attempts. |
| Turnover % | How often turnovers occur. |
| Offensive Rebound % | Percentage of all offensive rebounding chances that the team converts. |
| Defensive Rebounds % | Percentage of all defensive rebounding chances that the team converts. |
| Opponent Offensive Rebound % | The ratio of opponent offensive rebounds to the team's defensive rebounds. |
| Opponent Defensive Rebound % | The ratio of opponent defensive rebounds to the team's offensive rebounds. |
| 2PT % | Percentage of two pointers a team makes/allows. |
| 3PT % | Percentage of three pointers a team makes/allows. |
| Block % | Percentage of shots blocked. |
| Asist % | Percentage of made shots that were directly from a teammate's pass. |
| 2PT Ratio | Percentage of shots taken inside the three point line. |
| 3PT Ratio | Percentage of shots taken from three. |
| Average Height | Average height of all players weighted by minutes. |
| Effective Height | Calculate minute weighted height of the power forwards and centers, basically the average height of the tallest players. |
| Experience | Average years played Division 1 college basketball. |
| Talent | Metric based on the recruiting rank of the players. |
| Free throw % | % of free throws the team makes/allows. |
| Points per Possession | Points scored/allowed per possession. |
| Elite Strength of Schedule | % of games an elite team would be projected to lose based on a given schedule. |
| Wins Above Bubble | How many wins a bubble team would have had with a given schedule. |

| | |
|---|---|
| Power Rating | Calculation of the team's strength relative to other teams in the tournament. |
| Path | Metric that measures the difficulty of a team's path in the tournament, based on probability of future matchups. |
| Draw | Compares the path metric to the same seeds in other parts of the bracket. |
| Net RPI | Formula is not publicly available, but it uses adjusted efficiency margin (derived from offense and defensive efficiency), team value index, winning percentage weighted by opponent quality, strength of schedule, and quality of wins and losses. Uses RPI statistic before 2018 and the newer NET RPI after. |
| Resume | Rank based on committee win buckets. |
| B Power | Barttorvik Power Rating rank. |
| Q1 and Q2 wins | Wins against good teams. (Quadrants are defined using Net RPI and whether the game was home, away or neutral. |
| Q3 and Q4 losses. | Losses against bad teams. |
| Plus 500 | Amount of opposing teams with a win % over 50 a team has played. |
| Resume Score | Scores their resume. |

There are also various granular shooting splits for each team.

**Target Variables**
We are primarily concerned with how well each team will do in the tournament. Our target variables include:

- How many rounds the team advanced
- The score of the game
- Whether team 1 or team 2 won the game (1 or 0)
- If it was an upset (1 or 0)

**3. Summary Statistics**

Due to the nature and scale of our problem of interest, we have a lot of variables (92 to be precise). However, as aggregation and partitioning techniques will likely be used throughout our model training process, we decided **not** to remove any features from our data at this point.

As a result, the summary statistics displayed below are in many cases functionally useless and uninterpretable, as they are calculated relative to other teams and are themselves aggregated over the course of the season preceding.

For your viewing pleasure, these summary statistics are included at the end of this report, however they simply took up too much space without adding value to justify their inclusion here.

There are however several key trends that we can pick out from our summary statistics. The first is the prevalance of certain conferences among tournament teams, with a much larger proportion of teams belonging to power conferences such as the Big 12, Big 10, Big East, ACC, and SEC, as opposed to smaller conferences.

We also see an interesting trend in the frequency of individual schools represented in the tournament. While teams from large conferences generally make it to the tournament more often than teams from smaller conferences, some schools such as Yale in the Ivy League or Colgate in the Patriot League have more appearances than their regular season performances would merit. This can likely be attributed to the inclusion of auto-bids in the tournament, where conference champions are guaranteed a spot in the tournament.

An important shift in tournament structure can also be seen in our summary statistics, in that the tournament recently expanded to include more teams. Additionally, many teams have changed conferences over the time frame of our data, so it is important to note that older data will likely be less useful in making predictions than our newer data. Because of this, we will likely have to manually assign higher weight to newer data, at the very least in the creation of our supervised learning model.

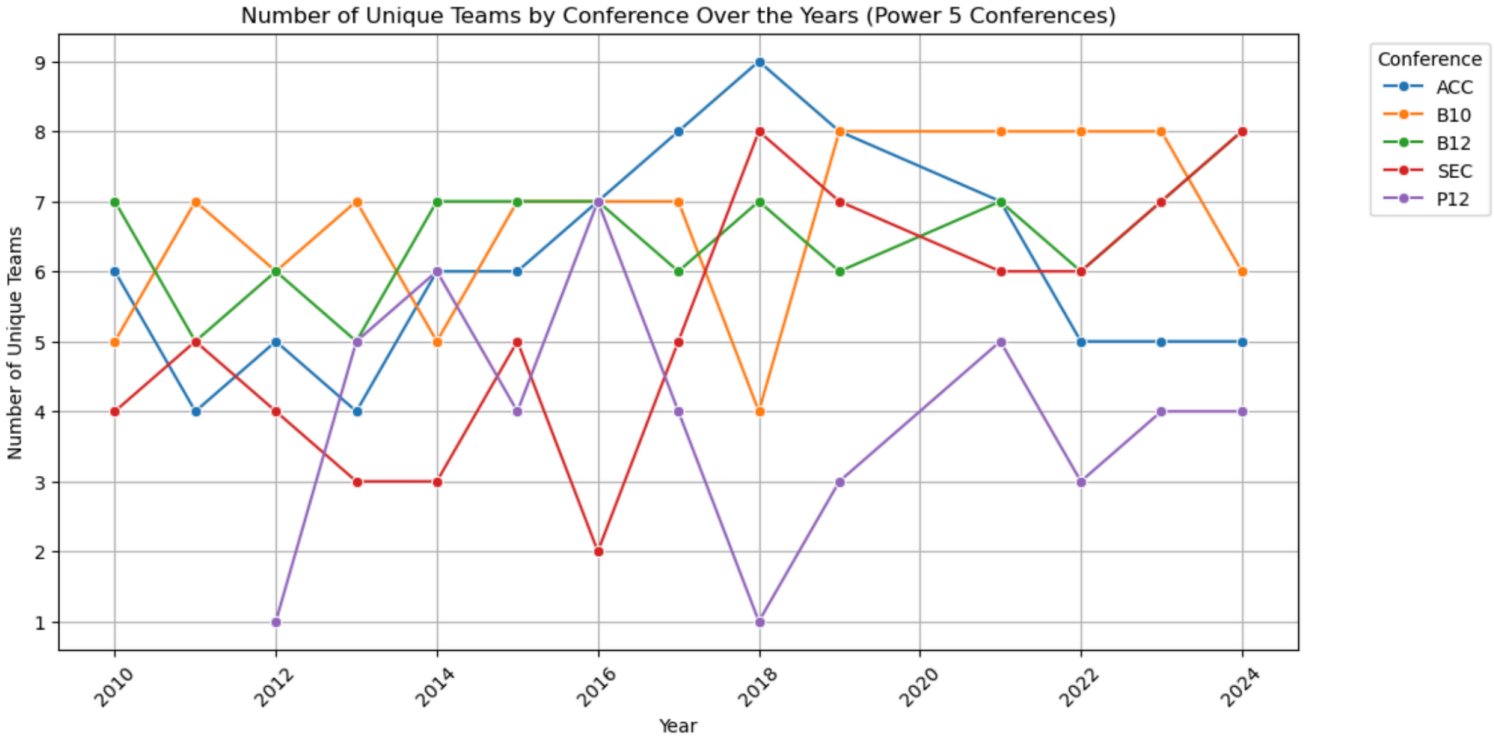
**Correlation Matrix**

As many of our features are directly calculated from each other, or attempt to measure the same things, our correlation matrix is of little value or interpretability as well. As we continue the modeling process, and feature selection occurs, a correlation heatmap would likely reveal invaluable trends in our data, as well as help us to avoid issues such as problematic multicollinearity. Until then, however, we have too many features for a correlation matrix to provide useful insights.

Our mess of a matrix can be found following the tables with feature summary statistics.

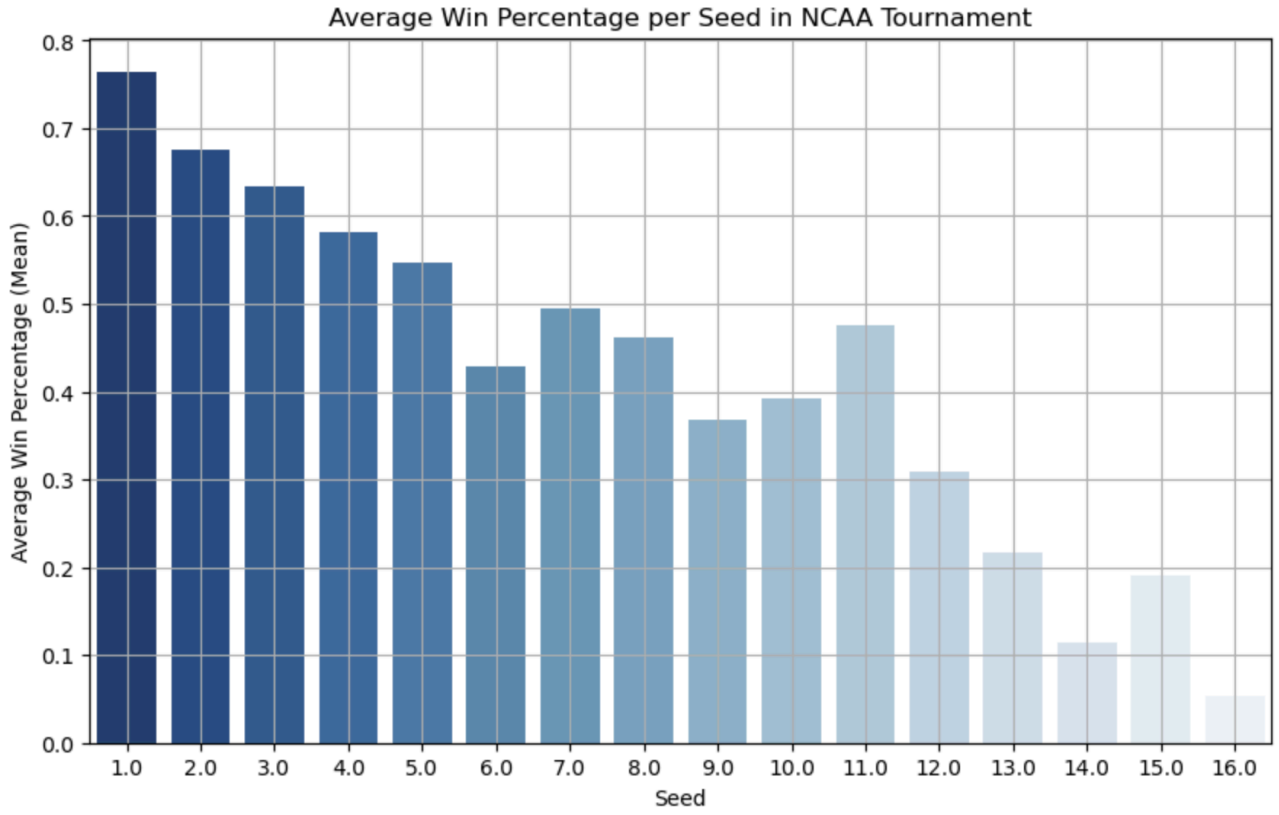
**4. Visual Exploration**

**Plot 1:**




This line plot shows the number of unique teams from five major conferences (ACC, Big Ten, Big 12, Pac-12, SEC) that participated in the NCAA Tournament each year. It tracks the participation of teams from these "Power 5" conferences over time, highlighting any trends or shifts in representation. This plot is relevant for understanding the historical strength and consistency of these conferences in the tournament, which may offer insights into potential advantages for teams from these conferences in predicting future March Madness outcomes.

**Plot 2:**




This bar plot shows the median win percentage for each seed group in the NCAA Tournament, grouping the overall seed(1 - 64) into more commonly used seed groups(1 - 16). It provides insight into the historical performance of teams from different seed groups, with higher seed groups (typically considered "underdogs") shown in lighter colors. Understanding these trends is relevant for predicting March Madness winners, as it highlights how teams from various seed groups have historically fared, helping analysts and fans assess the likelihood of upsets and the overall competitiveness of the tournament.

**5. Challenges and Reflection**

So far, the biggest challenge has been to identify how to best use the wealth of data we have curated to address our problem of interest. While this problem was daunting at first, as we began to break down all of the information we had, we were able to make a plan to incorporate the most useful information from multiple sources. As a result, the process of wrangling our data into a useful form was more intensive than for many previous projects.

We also are currently waiting for the end of regular season games from the 2025 season, as well as the announcement of tournament teams and rankings for this years NCAA Championship Tournament. As this data does not yet exist, we are unable to fully dive into the modeling process. This may actually be benificial however, as this will give us more time to carefullly plan out the features needed, and allow us to analyze older data in order to determine the most important features, as well which features don't matter, and as a result, don't need to be created from forthcoming data.

# Summary Statistics

## Numeric Variables

| Variable | Sample Size | Mean | Std Dev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|
| SCORE_diff | 818 | 4.74 | 13.73 | -29.00 | -5.00 | 4.00 | 14.00 | 47.00 |
| K TEMPO_diff | 888 | 0.03 | 4.12 | -12.70 | -2.71 | -0.08 | 2.92 | 12.11 |
| KADJ T_diff | 888 | -0.01 | 4.00 | -11.88 | -2.60 | -0.15 | 2.77 | 10.70 |
| K OFF_diff | 888 | 2.18 | 6.94 | -18.02 | -2.13 | 1.82 | 6.80 | 21.90 |
| KADJ O_diff | 888 | 3.79 | 7.86 | -18.12 | -1.97 | 3.10 | 8.74 | 27.47 |
| K DEF_diff | 888 | -1.40 | 6.25 | -25.34 | -5.58 | -1.13 | 2.58 | 16.18 |
| KADJ D_diff | 888 | -3.10 | 6.84 | -24.46 | -7.51 | -2.53 | 1.44 | 22.38 |
| KADJ EM_diff | 888 | 6.89 | 11.45 | -30.48 | -0.86 | 5.58 | 13.76 | 45.95 |
| BADJ EM_diff | 888 | 6.97 | 11.56 | -29.80 | -1.23 | 5.45 | 13.90 | 44.30 |
| BADJ O_diff | 888 | 4.02 | 8.32 | -20.60 | -1.70 | 3.58 | 9.53 | 28.54 |
| BADJ D_diff | 888 | -2.95 | 6.48 | -24.40 | -7.10 | -2.48 | 1.20 | 22.20 |
| BARTHAG_diff | 888 | 0.11 | 0.18 | -0.66 | -0.01 | 0.05 | 0.17 | 0.73 |
| GAMES_diff | 888 | 0.58 | 2.53 | -12.00 | -1.00 | 0.00 | 2.00 | 13.00 |
| W_diff | 888 | 1.87 | 4.92 | -12.00 | -1.00 | 2.00 | 5.00 | 22.00 |
| L_diff | 888 | -1.32 | 4.37 | -19.00 | -4.00 | -1.00 | 2.00 | 11.00 |
| WIN%_diff | 888 | 4.73 | 14.01 | -37.50 | -5.20 | 4.29 | 13.62 | 63.33 |
| EFG%_diff | 888 | 0.69 | 3.90 | -11.30 | -2.00 | 0.60 | 3.40 | 14.30 |
| EFG%D_diff | 888 | -0.68 | 3.27 | -12.90 | -2.80 | -0.70 | 1.40 | 9.90 |
| FTR_diff | 888 | -0.36 | 6.26 | -23.30 | -4.40 | -0.20 | 3.90 | 24.50 |
| FTRD_diff | 888 | -1.55 | 7.86 | -28.90 | -7.00 | -1.70 | 3.70 | 28.00 |
| TOV%_diff | 888 | -0.41 | 2.62 | -10.30 | -2.10 | -0.40 | 1.30 | 8.90 |
| TOV%D_diff | 888 | 0.01 | 3.50 | -11.10 | -2.30 | 0.00 | 2.30 | 10.90 |
| OREB%_diff | 888 | 1.22 | 5.63 | -17.20 | -2.40 | 1.35 | 5.03 | 17.80 |
| DREB%_diff | 888 | 0.25 | 4.15 | -14.50 | -2.52 | 0.20 | 3.20 | 12.20 |
| OP OREB%_diff | 888 | -0.25 | 4.15 | -12.20 | -3.20 | -0.20 | 2.52 | 14.50 |
| OP DREB%_diff | 888 | -1.22 | 5.63 | -17.80 | -5.03 | -1.35 | 2.40 | 17.20 |
| RAW T_diff | 888 | 0.02 | 4.11 | -12.40 | -2.70 | -0.05 | 2.90 | 12.00 |
| 2PT%_diff | 888 | 0.78 | 4.28 | -13.40 | -2.10 | 0.70 | 3.50 | 17.90 |
| 2PT%D_diff | 888 | -0.80 | 3.95 | -12.90 | -3.50 | -0.80 | 1.73 | 11.30 |
| 3PT%_diff | 888 | 0.38 | 3.78 | -10.10 | -2.02 | 0.50 | 2.80 | 14.00 |
| 3PT%D_diff | 888 | -0.29 | 2.99 | -9.70 | -2.30 | -0.25 | 1.80 | 8.50 |
| BLK%_diff | 888 | 0.67 | 4.26 | -12.40 | -2.10 | 0.40 | 3.52 | 14.30 |
| BLKED%_diff | 888 | -0.16 | 2.17 | -7.30 | -1.60 | -0.20 | 1.30 | 7.30 |

| Variable | Sample Size | Mean | Std Dev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|
| AST%_diff | 888 | 1.15 | 7.26 | -22.10 | -3.50 | 1.40 | 5.90 | 23.90 |
| OP AST%_diff | 888 | -0.07 | 7.31 | -23.10 | -4.90 | -0.15 | 4.43 | 23.30 |
| 2PTR_diff | 888 | 0.28 | 6.78 | -19.80 | -4.30 | 0.25 | 5.10 | 22.00 |
| 3PTR_diff | 888 | -0.28 | 6.78 | -22.00 | -5.10 | -0.25 | 4.30 | 19.80 |
| 2PTRD_diff | 888 | 0.11 | 5.24 | -13.50 | -3.50 | 0.10 | 3.70 | 16.00 |
| 3PTRD_diff | 888 | -0.11 | 5.24 | -16.00 | -3.70 | -0.10 | 3.50 | 13.50 |
| BADJ T_diff | 888 | 0.05 | 4.02 | -11.50 | -2.50 | -0.10 | 2.80 | 10.50 |
| AVG HGT_diff | 888 | 0.35 | 1.23 | -3.55 | -0.44 | 0.40 | 1.12 | 5.19 |
| EFF HGT_diff | 888 | 0.36 | 1.51 | -4.32 | -0.64 | 0.38 | 1.33 | 6.36 |
| EXP_diff | 888 | -0.07 | 0.57 | -1.87 | -0.43 | -0.08 | 0.27 | 1.73 |
| TALENT_diff | 888 | 18.08 | 38.04 | -94.31 | -8.96 | 17.96 | 50.61 | 94.60 |
| FT%_diff | 888 | 0.18 | 4.95 | -17.30 | -3.30 | 0.30 | 3.50 | 16.00 |
| OP FT%_diff | 888 | -0.06 | 3.21 | -9.90 | -2.20 | -0.20 | 2.10 | 9.70 |
| PPPO_diff | 888 | 0.02 | 0.07 | -0.18 | -0.02 | 0.02 | 0.07 | 0.23 |
| PPPD_diff | 888 | -0.01 | 0.06 | -0.25 | -0.06 | -0.01 | 0.03 | 0.16 |
| ELITE SOS_diff | 888 | 5.43 | 11.26 | -21.73 | -2.86 | 4.76 | 14.72 | 31.37 |
| WAB_diff | 888 | 3.68 | 6.18 | -17.00 | -0.40 | 3.10 | 7.20 | 25.70 |
| POWER RATING_diff | 435 | 4.87 | 8.27 | -18.20 | -1.20 | 4.50 | 9.80 | 28.20 |
| POWER_diff | 692 | 9.69 | 15.86 | -29.10 | -2.12 | 9.80 | 20.92 | 49.50 |
| PATH_diff | 692 | -4.53 | 8.06 | -31.60 | -9.12 | -3.80 | 1.30 | 13.70 |
| DRAW_diff | 624 | -0.01 | 1.19 | -3.23 | -0.78 | -0.03 | 0.76 | 4.06 |
| WINS_diff | 692 | 0.59 | 2.04 | -5.00 | -1.00 | 1.00 | 2.00 | 6.00 |
| POOL VALUE_diff | 692 | 17.38 | 40.77 | -112.70 | -4.67 | 14.90 | 46.55 | 120.60 |
| POWER-PATH_diff | 692 | 14.21 | 23.61 | -40.50 | -3.45 | 13.55 | 30.12 | 81.10 |
| NET RPI_diff | 888 | -32.29 | 56.47 | -296.00 | -49.00 | -18.00 | 2.00 | 288.00 |
| RESUME_diff | 888 | -44.50 | 74.15 | -324.00 | -88.25 | -22.50 | 6.00 | 270.00 |
| ELO_diff | 888 | -24.45 | 47.56 | -250.00 | -41.00 | -15.50 | 3.25 | 248.00 |
| B POWER_diff | 888 | -36.99 | 62.86 | -296.00 | -63.08 | -18.00 | 3.30 | 269.00 |
| Q1 W_diff | 888 | 2.58 | 4.93 | -11.00 | -1.00 | 3.00 | 6.00 | 17.00 |
| Q2 W_diff | 888 | 1.54 | 3.80 | -12.00 | -1.00 | 2.00 | 4.00 | 13.00 |
| Q1 PLUS Q2 W_diff | 888 | 4.12 | 7.13 | -15.00 | -1.00 | 4.00 | 10.00 | 22.00 |
| Q3 Q4 L_diff | 888 | -1.64 | 3.15 | -16.00 | -3.00 | -1.00 | 0.00 | 14.00 |
| PLUS 500_diff | 888 | 3.34 | 9.12 | -24.00 | -3.00 | 3.00 | 9.00 | 41.00 |
| R SCORE_diff | 888 | 35.32 | 50.20 | -99.30 | 0.00 | 15.45 | 98.30 | 100.00 |
| DUNKS FG%_diff | 888 | 0.59 | 6.10 | -21.10 | -3.70 | 0.75 | 4.40 | 25.80 |
| DUNKS SHARE_diff | 888 | 1.19 | 4.18 | -13.10 | -1.70 | 1.25 | 4.20 | 15.70 |
| DUNKS FG%D_diff | 888 | -0.70 | 6.97 | -20.80 | -5.30 | -0.70 | 3.62 | 19.50 |
| DUNKS D SHARE_diff | 888 | 0.21 | 2.04 | -5.40 | -1.20 | 0.30 | 1.60 | 7.20 |
| CLOSE TWOS FG%_diff | 888 | 1.80 | 5.63 | -19.10 | -2.10 | 1.70 | 5.30 | 24.10 |

| Variable | Sample Size | Mean | Std Dev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|
| CLOSE TWOS SHARE_diff | 888 | -0.42 | 6.76 | -27.60 | -4.80 | -0.30 | 4.03 | 23.60 |
| CLOSE TWOS FG%D_diff | 888 | 0.01 | 5.84 | -20.00 | -3.73 | -0.05 | 3.90 | 26.70 |
| CLOSE TWOS D SHARE_diff | 888 | -1.32 | 5.63 | -21.50 | -4.90 | -1.35 | 2.23 | 19.00 |
| FARTHER TWOS FG%_diff | 888 | 0.12 | 4.72 | -14.10 | -3.20 | 0.20 | 3.30 | 15.50 |
| FARTHER TWOS SHARE_diff | 888 | 0.72 | 8.10 | -27.60 | -4.70 | 0.60 | 6.00 | 29.40 |
| FARTHER TWOS FG%D_diff | 888 | -0.82 | 4.04 | -14.60 | -3.42 | -0.80 | 2.00 | 10.80 |
| FARTHER TWOS D SHARE_diff | 888 | 1.40 | 6.25 | -21.20 | -3.00 | 1.10 | 5.60 | 23.40 |
| THREES FG%_diff | 888 | 0.41 | 3.68 | -10.00 | -2.00 | 0.50 | 2.73 | 13.50 |
| THREES SHARE_diff | 888 | -0.29 | 6.80 | -23.40 | -5.12 | -0.30 | 4.30 | 20.10 |
| THREES FG%D_diff | 888 | -0.31 | 2.87 | -10.60 | -2.12 | -0.20 | 1.60 | 8.50 |
| THREES D SHARE_diff | 888 | -0.08 | 5.22 | -15.40 | -3.52 | 0.00 | 3.50 | 13.50 |

# Categorical Variables

## YEAR

Sample Size: 888

| Category | Count |
|---|---|
| 2024 | 70 |
| 2023 | 63 |
| 2022 | 63 |
| 2019 | 63 |
| 2018 | 63 |
| 2017 | 63 |
| 2016 | 63 |
| 2015 | 63 |
| 2014 | 63 |
| 2013 | 63 |
| 2012 | 63 |
| 2011 | 63 |
| 2010 | 63 |
| 2021 | 62 |

## ROUND

Sample Size: 888

| Category | Count |
|---|---|
| 64 | 447 |
| 32 | 238 |
| 16 | 112 |
| 8 | 52 |
| 4 | 26 |
| 2 | 13 |

## CONF_1

Sample Size: 888

| Category | Count |
|---|---|
| B12 | 131 |
| B10 | 130 |
| ACC | 128 |
| BE | 118 |
| SEC | 112 |
| P12 | 62 |
| WCC | 42 |
| MWC | 35 |
| A10 | 34 |
| Amer | 27 |
| MVC | 21 |
| P10 | 10 |
| Horz | 7 |
| CUSA | 6 |
| CAA | 6 |
| OVC | 5 |
| Ivy | 3 |
| MAC | 2 |
| SC | 2 |
| BSth | 1 |
| AE | 1 |
| SB | 1 |
| SInd | 1 |
| Sum | 1 |
| NEC | 1 |
| WAC | 1 |

# CONF_2

Sample Size: 888

| Category | Count |
|---|---|
| B10 | 96 |
| ACC | 89 |
| B12 | 80 |
| SEC | 71 |
| BE | 71 |
| P12 | 46 |
| A10 | 33 |
| MWC | 29 |
| Amer | 24 |
| WCC | 23 |
| CUSA | 20 |
| MVC | 20 |
| MAC | 18 |
| Horz | 17 |
| Ivy | 17 |
| CAA | 17 |
| MAAC | 16 |
| BW | 16 |
| Sum | 16 |
| ASun | 15 |
| WAC | 15 |
| Pat | 15 |
| SC | 14 |
| AE | 14 |
| OVC | 14 |
| Slnd | 13 |
| SB | 13 |
| BSth | 13 |
| BSky | 13 |
| MEAC | 11 |
| NEC | 8 |
| SWAC | 8 |
| P10 | 3 |

## Correlation Matrix (Selected Observations)

| | SEED_1 | SEED_2 | SCORE_diff | K TEMPO_diff | \ |
|---|---|---|---|---|---|
| SEED_1 | 1.000000 | -0.315856 | -0.351521 | -0.103653 | |
| SEED_2 | -0.315856 | 1.000000 | 0.421113 | 0.013043 | |
| SCORE_diff | -0.351521 | 0.421113 | 1.000000 | -0.025425 | |
| K TEMPO_diff | -0.103653 | 0.013043 | -0.025425 | 1.000000 | |
| KADJ T_diff | -0.097353 | 0.009324 | -0.029602 | 0.975537 | |
| ... | ... | ... | ... | ... | |
| THREES FG%_diff | -0.133837 | 0.111727 | 0.137223 | -0.016385 | |
| THREES SHARE_diff | 0.017607 | -0.012557 | 0.037950 | -0.086798 | |
| THREES FG%D_diff | 0.154586 | -0.122789 | -0.188784 | -0.027070 | |
| THREES D SHARE_diff | 0.022176 | 0.013140 | -0.046240 | -0.003223 | |
| target | -0.272636 | 0.350063 | 0.790248 | -0.011228 | |

| | KADJ T_diff | K OFF_diff | KADJ O_diff | K DEF_diff | \ |
|---|---|---|---|---|---|
| SEED_1 | -0.097353 | -0.414542 | -0.541173 | 0.320006 | |
| SEED_2 | 0.009324 | 0.370297 | 0.617626 | -0.236152 | |
| SCORE_diff | -0.029602 | 0.326514 | 0.459196 | -0.222068 | |
| K TEMPO_diff | 0.975537 | 0.123477 | 0.120671 | 0.119604 | |
| KADJ T_diff | 1.000000 | 0.134936 | 0.132864 | 0.142452 | |
| ... | ... | ... | ... | ... | |
| THREES FG%_diff | -0.013957 | 0.607516 | 0.499444 | 0.194005 | |
| THREES SHARE_diff | -0.081387 | 0.244320 | 0.163573 | 0.189655 | |
| THREES FG%D_diff | -0.001370 | 0.054617 | 0.049970 | 0.521837 | |
| THREES D SHARE_diff | -0.000746 | -0.144165 | -0.089333 | 0.055800 | |
| target | -0.014754 | 0.243719 | 0.342728 | -0.182959 | |

| | THREES D SHARE_diff | target |
|---|---|---|
| THREES D SHARE_diff | 1.000000 | -0.042285 |
| target | -0.042285 | 1.000000 |

[86 rows x 86 columns]

*Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...*

## Correlation Heatmap (Micro Scale)