Analyzing NCAA Men's March Madness Team Performance

Max Midlash

Introduction

Each year, the NCAA Division I men's basketball tournament, also known as March Madness, brings together qualifying teams in a single-elimination competition, progressing through rounds until a champion is declared. The tournament format has remained largely consistent, adopting a 64 team bracket in 1985. In 2011, the field of play was expanded to 68 teams, with additional teams competing in a preliminary "play-in" game before the 64 team competition begins.

Despite the relative consistency in the format of this competition, there are constant changes through the sport of basketball. For example, prior to 1986, there was no three-point line in college basketball. The National Collegiate Athletic Association (NCAA) added the three-point line in the 1986-1987 season at 19 ft 9 in, extended it to 20 ft 9 in for the 2008-2009 season, and further increased to 22 ft 1 ¾ in starting in 2019-2020 (National Collegiate Athletic Association, 2019). These rule changes have reshaped team strategies and performance metrics, necessitating analysis to understand their impact on tournament success.

Beyond rule changes, the sport and the style of play of athletes continue to change. Coaches and players are always looking for an edge, and adapting to change is an integral part of this. As teams start to experience success with a certain style of play, others will try to adapt to counter that method. So, at a high level, coaches and athletes may be interested in exploring the style of play of teams across the league and explore the factors that contribute to a winning style of play.

Problem Statement

This project investigates the characteristics of successful March Madness teams by analyzing regular season and tournament data from 2003 to 2025, excluding the 2020 season canceled due to the COVID-19 pandemic. Our goal is to identify patterns in team performance metrics such as three-point shooting, offensive efficiency, and more that distinguish teams that advance to the final 16 teams (also called the Sweet 16) from those that do not. By examining how rule changes and evolving playing styles influence these metrics, this study aims to provide insights for coaching strategies and predictive modeling in college basketball.

Methodology

Data Collection

Data were sourced from the 2025 Kaggle competition "March Machine Learning Mania 2025" (Kaggle, 2025), which provides regular season game statistics, tournament results, and team metadata from 2003 to 2025, excluding the cancelled 2020 season. This dataset was selected for its comprehensive inclusion of performance metrics critical to analyzing March Madness success, such as shooting efficiency and three-point percentages. The data were downloaded to a local repository for analysis.

Data Preprocessing

We loaded the "MRegularSeasonDetailedResults.csv" file, which includes game-level statistics such as field goal attempts, three-point attempts, points scored, rebounds, turnovers and free throws for winning and losing teams. To create a season-level dataset of team performance metrics, we transformed this game level data using Python and pandas.

First, to measure game tempo, we estimated possessions per game for each team using a standard basketball analytics formula incorporating field goal attempts, offensive rebounds, turnovers, and free throw attempts, averaging values across winning and losing games. We then restructured the game-level data to align team performances by matching team IDs across games and combined them into a single dataframe.

Next, we aggregated the data by season and team, calculating mean statistics such as points scored, points allowed, and possessions. From these we derived the following metrics, chosen for their relevance in evaluating offensive and defensive performance:

- Three-point attempt rate = Three-point attempts / Total field goal attempts
- Three-point percentage = Three-point made / Three-point attempts
- Pace = Average possessions per game
- Offensive efficiency = Points scored per 100 possessions
- Defensive efficiency = Points allowed per 100 possessions

The final processed dataset includes columns for 'Season', 'TeamID', and the above listed metrics. Tournament performance was categorized as 'winning' if teams advance to the Sweet 16, or 'losing' if they do not, using tournament results data to identify Sweet 16 teams.

Exploratory Data Analysis

We conducted exploratory data analysis (EDA) to examine the distributions, correlations, and trends of our team performance metrics. First, we computed a correlation matrix for our derived metrics (three-point attempt rate, three-point percentage, pace, offensive efficiency, and defensive efficiency) and visualized it as a heatmap, shown in Figure 1 below. Most correlations were low, with absolute magnitudes below 0.25, indicating the metrics capture mostly distinct aspects of performance. One exception was a positive correlation of 0.63 between three-point percentage and offensive efficiency, which is expected as accurate three-point shooting significantly boosts scoring efficiency.

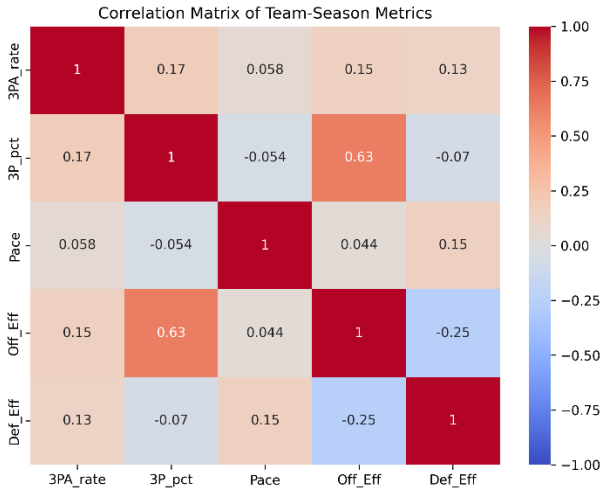

Figure 1 – Correlation Heatmap of team performance metrics.

To investigate trends in team performance over time, we calculated the average metrics across all teams from 2003 to 2025. Figure 2 below shows the average three-point attempt rate over time, and Figure 3 shows average pace over time. These plots allow us to see big picture changes across the dataset. Both plots use season as the x-axis and the corresponding metric on the y-axis. It is worth pointing out that both average three-point attempt rates and pace appear to change greatly around the 2014-2015 season, warranting further investigation.

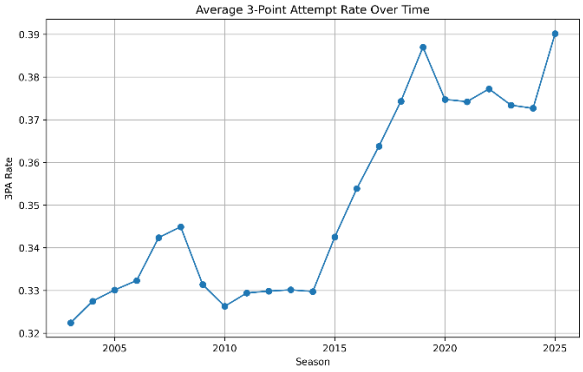

Figure 2 – Average three point attempt rate by season.

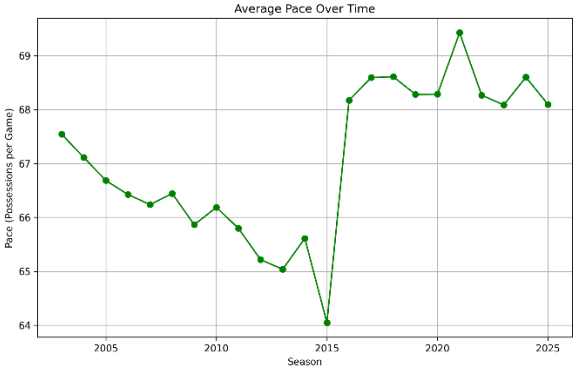

Figure 3 – Average pace by season.

We also examined differences in metrics by tournament success, labeling teams that reached the Sweet 16 as successful (Success = 1) and those that did not as unsuccessful (Success = 0). Figure 4 below contains boxplots comparing the distributions of three-point attempt rate, three-point percentage, pace, offensive efficiency, and defensive efficiency between successful and unsuccessful teams. The boxplots reveal that successful teams have significantly higher three-point percentage and offensive efficiency, and lower defensive efficiency.

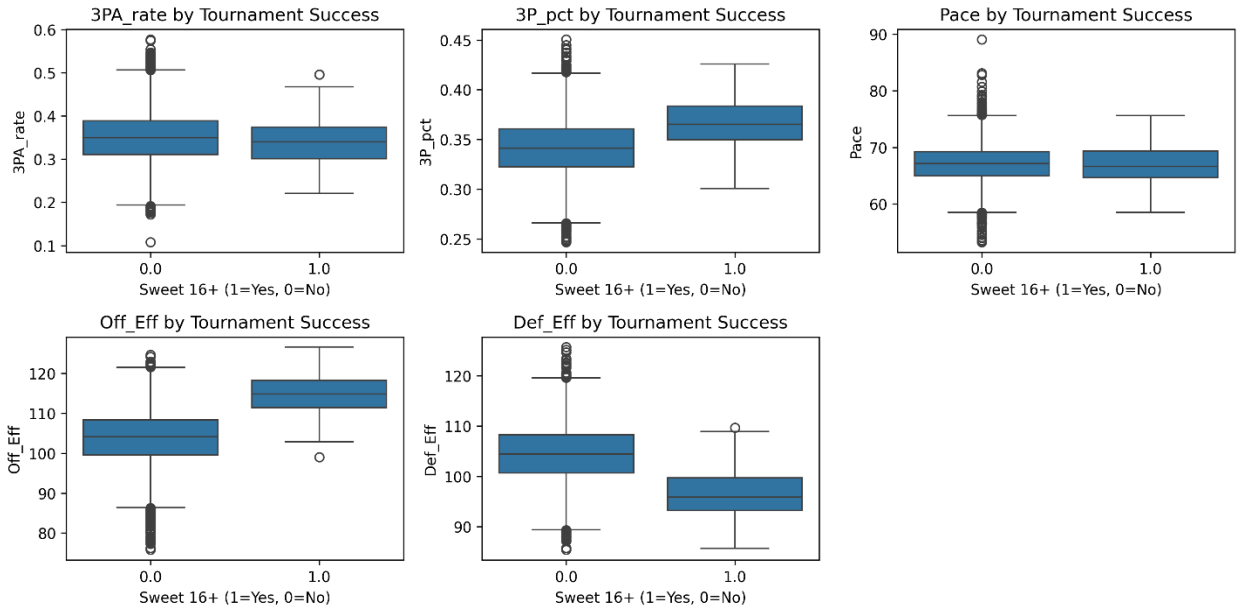

Figure 4 – Boxplots comparing metric distributions for successful vs unsuccessful teams.

Clustering

Building on the EDA's insights into metric trends and independence, we applied K-means clustering to group teams with similar performance profiles based on our standardized metrics. To determine the optimal number of clusters, we used the elbow method. We performed K-means clustering for 1 to 10 clusters using scikit-learn in Python. We plotted the within-cluster sum of squares (WCSS), a measure of cluster compactness, for each number of clusters. Figure 5 below shows this elbow plot, with the number of clusters on the x-axis and WCSS on the y-axis. The plot indicates diminished improvements in WCSS beyond four clusters, suggesting four clusters balance interpretability and generalizability. To ensure robust clusters for analyzing Sweet 16 success while avoiding overfitting, we selected four clusters for our upcoming K-means clustering analysis.

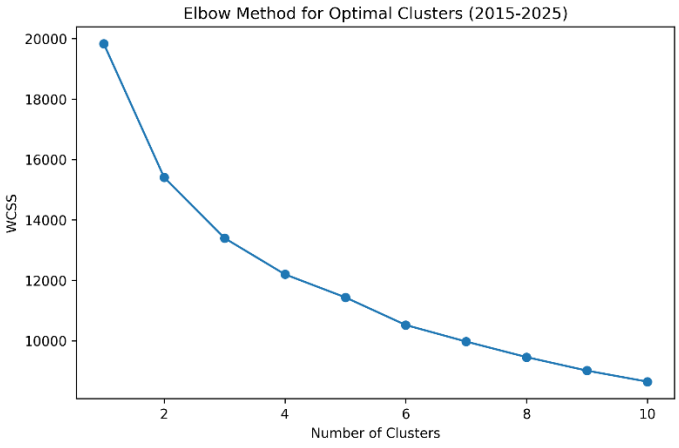

Figure 5 – Elbow plot of within-cluster sum of squares (WCSS) vs. number of clusters.

Results

To characterize team performance profiles, we applied K-means clustering with four clusters to our standardized metrics (three-point attempt rate, three-point percentage, pace, offensive efficiency, and defensive efficiency), grouping teams by similar playing styles. We divided the data into three eras. These eras were 2003-2008, 2009-2014, and 2015-2025. These eras were chosen based on NCAA rule changes, including three-point line extensions in 2008-2009 and 2019-2020, and the shot clock reduction from 35 to 30 seconds in 2015-2016 (Associated Press, 2015). The EDA also found that around the year 2014-2015, there were large changes in three-point attempts and pace (refer back to Figures 2 and 3).

For each era, we clustered teams based on the season-level metrics and computed the cluster centers. These cluster centers represent the average metric values for each cluster. Figures 6, 7, and 8 below display the cluster center values for each era. We then calculated the success rate (proportion of Sweet 16 teams) for each of these clusters. Figures 6-8 below allows us to see which of these clusters are 'winning' styles of play in each era. In the 2003-2008 era, Cluster 0, with a low three-point attempt rate of 0.311, had the highest success rate, suggesting reliance on 'inside scoring' rather than three-point shooting. In the 2009-2014 era, Cluster 1 had the most success, also with a low three point attempt rate. In contrast, in the 2015-2025 era, Cluster 0 had the most success, but this cluster has among the highest three-point attempt rate. This shows how the style of 'winning play' has changed across the dataset.

2003-2008 Cluster Characteristics and Success Rates

	3PA_rate	3P_pct	Pace	Off_Eff	Def_Eff	Success Rate
Cluster 0	0.311	0.365	66.809	109.439	98.03	0.158
Cluster 1	0.307	0.316	66.52	95.926	104.189	0.0
Cluster 2	0.394	0.359	63.515	104.531	104.326	0.005
Cluster 3	0.334	0.352	69.524	104.043	107.099	0.0

Figure 6 – Cluster centers and success rates, 2003-2008.

2009-2014 Cluster Characteristics and Success Rates

	3PA_rate	3P_pct	Pace	Off_Eff	Def_Eff	Success Rate
Cluster 0	0.359	0.346	67.83	103.867	109.284	0.0
Cluster 1	0.295	0.346	66.264	108.227	99.047	0.106
Cluster 2	0.368	0.369	63.318	108.618	102.845	0.065
Cluster 3	0.308	0.313	65.141	96.91	105.419	0.0

Figure 7 – Cluster centers and success rates, 2009-2014.

2015-2025 Cluster Characteristics and Success Rates

	3PA_rate	3P_pct	Pace	Off_Eff	Def_Eff	Success Rate
Cluster 0	0.394	0.373	67.729	113.012	102.125	0.134
Cluster 1	0.352	0.337	66.623	105.141	100.533	0.025
Cluster 2	0.34	0.312	68.014	97.145	107.328	0.0
Cluster 3	0.401	0.345	69.744	104.525	108.952	0.0

Figure 8 - Cluster centers and success rates, 2015-2025.

To quantify the differences between successful (Sweet 16) teams and unsuccessful teams, we conducted independent Welch's t-tests for each metric across all seasons. We used a Bonferroni adjusted significance level of 0.01 for multiple comparisons. Figure 9 below shows the result of this t-test, indicating significant differences in three-point attempt rate, three-point percentage, offensive efficiency, and defensive efficiency, but not pace. This is consistent with our EDA boxplots in Figure 4. These findings highlight that shooting accuracy, scoring, and defense are critical for March Madness success.

T-Test Results for Metrics (Sweet 16+ vs. Others, Bonferroni Adjusted α=0.01)

Metric	T-Statistic	P-Value	Significance
3PA_rate	-3.25	0.0013	significant
3P_pct	16.78	0.0	significant
Pace	0.58	0.5636	not significant
Off_Eff	41.68	0.0	significant
Def_Eff	-31.01	0.0	significant

Figure 9 – T-test results comparing metrics for successful vs. unsuccessful teams.

To visualize the team clusters in the 2015-2025 era, we applied principal component analysis (PCA) to reduce the five metrics to two dimensions. This allows us to see how well the clusters can separate different teams, as well as the distribution of successful teams. Figure 10 below shows a scatterplot of teams in this 2D space, with points colored to indicate cluster membership. A circle on the plot indicates an unsuccessful team, while an X indicates a successful team. To more readily see the successful teams, Figure 11 shows a faceted view of the plot, with unsuccessful teams on the left side, and successful teams on the right. This view allows us to easily see that Cluster 0 and Cluster 1 are "winning" clusters in the 2015-2025 era.

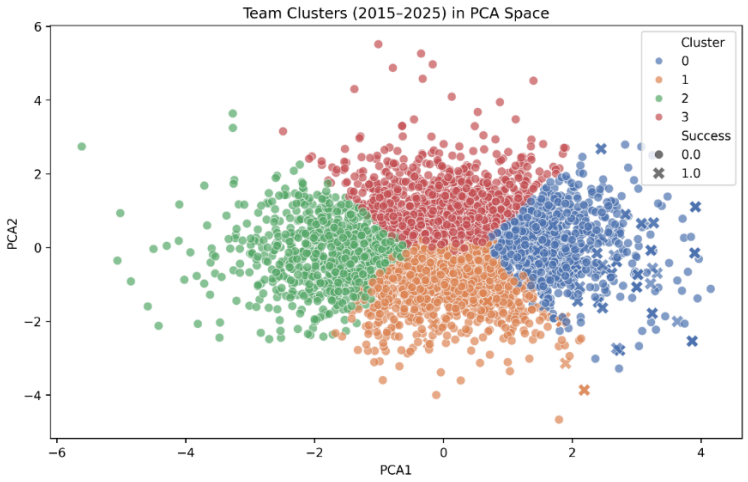

Figure 10 – PCA scatterplot of team clusters, 2015-2025.

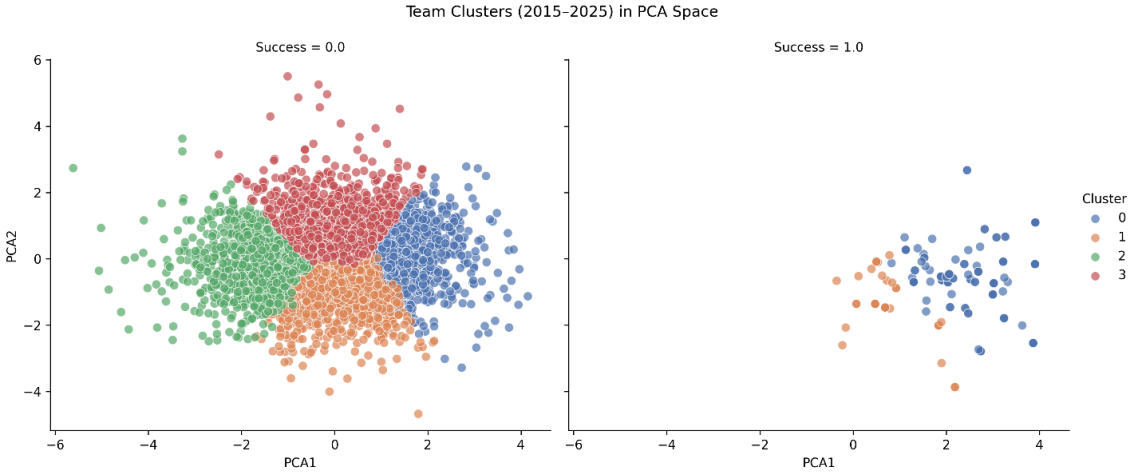

Figure 11 – Faceted PCA scatterplot separating unsuccessful (left) vs successful teams (right), 2015-2025.

Drilling down on the 'winning' clusters from the 2015-2025 era, we identified the top teams in Clusters 0 and 1, which had high success rates in advancing to the Sweet 16. We counted the frequency of each team's appearance in these clusters across seasons, and mapped team IDs to names using the "MTeams.csv" file (Kaggle, 2025). Figure 12 below shows the top five teams by occurrence in Cluster 0 on the left side, and top five teams by occurrence in Cluster 1 on the right side. Figure 13 visualizes these frequencies in a bar plot, with the blue bars for Cluster 0 and orange for Cluster 1. These figures highlight teams consistently playing with performance profiles similar to the successful clusters.

Top 5 Teams in Clusters 0 and 1 (2015–2025)

Cluster 0 Team	Cluster 0 Count	Cluster 1 Team	Cluster 1 Count
Villanova	17	San Diego St	11
Gonzaga	15	Kansas St	10
Duke	13	UC Irvine	10
North Carolina	11	Grand Canyon	10
Toledo	9	Seton Hall	9

Figure 12 – Top five teams by frequency in Clusters 0 and 1, 2015-2025.

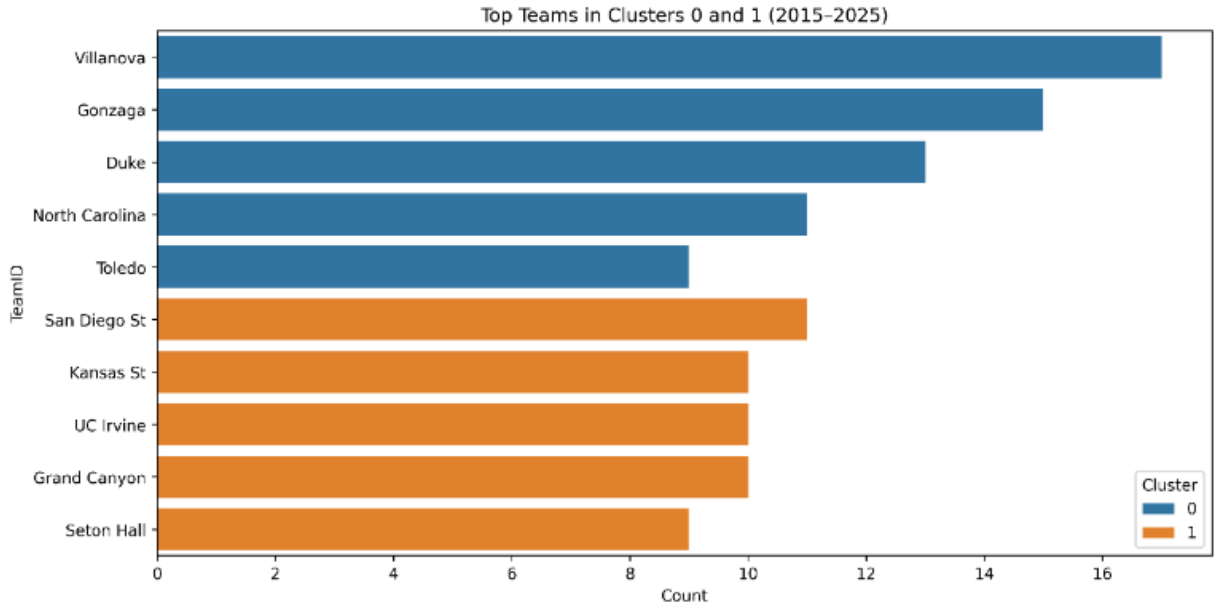

Figure 13 – Bar plot of top team frequencies in Clusters 0 and 1, 2015-2025.

Taking a step back here, we examined each of these teams holistically. Using domain knowledge and outside research, many of these teams are in fact basketball "powerhouses". These teams usually experience successful March Madness tournaments, with the exception of Toledo. Toledo's men's basketball team has not made the tournament at all since 1980 (Yahoo Sports, 2023). This was very surprising to us, so we decided to drill down further. Figure 14 below shows a table summarizing Toledo's lack of success, despite belonging to the 'winning' Cluster 0 nearly every single year in this period. Figure 15 shows a radar plot comparing Toledo's average metrics to the Cluster 0 centers in this 2015-2025 era. The plot scales each metric relative to the maximum values to better show how closely Toledo's playing style aligns with Cluster 0's center. Looking at the values for Toledo's playing style and the 'winning' Cluster 0, the performance metrics are very similar. Toledo's high three-point attempt rate and offensive efficiency closely match Cluster 0's successful profile. These figures suggest a playing style compatible with Sweet 16 success, despite their historical absence from the tournament.

Toledo Team Summary (2015–2025)

Attribute	Value
Team Name	Toledo
TeamID	1405
Success Seasons (Sweet 16+)	None
Cluster 0 Seasons	2015, 2016, 2018, 2019, 2020, 2021, 2022, 2023, 2024

Figure 14 – Table summarizing Toledo's Cluster 0 membership and lack of tournament success, 2015-2025.

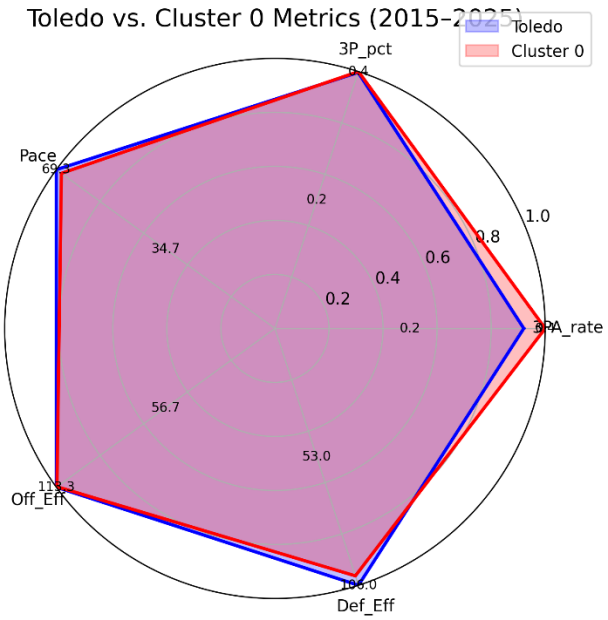

Figure 15 – Radar plot comparing Toledo's average metrics to Cluster 0 centers, 2015-2025.

Discussion

Our analysis reveals a clear evolution in college basketball's winning playing styles from 2003 to 2025. Clustering results in Figures 6-8 show that successful teams in the 2003-2008 and 2009-2014 eras relied on lower three-point attempt rates, while the 2015-2025 era favors clusters with high three-point attempt rates and offensive efficiency, reflecting a shift toward perimeter shooting. This aligns with the t-test results from Figure 9, which identified significant differences in three-point attempt rate, three-point percentage, offensive efficiency, and defensive efficiency between Sweet 16 teams and others.

The exploratory data analysis in Figures 2 and 3 identified sharp increases in three-point attempt rate and pace around the 2014-2015 season. This coincides with the NCAA's reduction of the shot clock from 35 to 30 seconds in the 2015-2016 season (Associated Press, 2015). This change likely encouraged faster, more offensive styles of play. This change drove the choice of our eras for our analysis (2003-2008, 2009-2014, 2015-2025). The 2008-2009 season saw the three-point

line extended, so this year was chosen as a natural breakpoint, as well as the 2015-2016 season that saw the shot clock reduced.

To further validate our clusters, PCA visualizations in Figures 10 and 11 confirm that Clusters 0 and 1 in the 2015-2025 era are distinct and contain most successful teams. This validates these clusters of teams represent ‘winning’ styles of play.

The identification of basketball powerhouses in Clusters 0 and 1 in Figures 12 and 13 reinforce the clusters’ relevance, but Toledo’s presence is a notable anomaly. Despite not qualifying for the March Madness tournament since 1980, Toledo consistently aligns with Cluster 0’s high three-point attempt rate and offensive efficiency, shown in Figures 14 and 15. This suggests that Toledo possesses a playing style compatible with tournament success, potentially limited by factors such as the competitive nature of their conference and a lack of tournament bids.

Future Work

Future work could explore Toledo as a case study. By examining what is different about Toledo, we may be able to pinpoint valuable features that were overlooked in this analysis. Toledo’s consistent Cluster 0 membership may warrant deeper investigation into other “soft” metrics, like recruiting, coaching, or culture, which could reveal barriers to tournament success beyond performance metrics.

Conclusion

This study examines the evolution of college basketball winning strategies from 2003 to 2025, identifying a shift toward high three-point attempt rates and offensive efficiency in the 2015-2025 era. Clustering, t-tests, and PCA analyses revealed significant metrics that drive March Madness tournament success, namely three-point attempt rate, three-point percentage, offensive efficiency, and defensive efficiency. The unexpected alignment of Toledo with the successful Cluster 0, despite no March Madness tournament appearances since 1980 underscores it as a potential team to watch. These findings support basketball analytics by guiding strategic decisions and predictive modeling. This analysis contributes to a data-driven understanding of March Madness success, informing future research and team evaluations.

References

Associated Press. (2015, June 8). Shorter shot clock, fewer timeouts among changes coming in 2015-16. https://www.espn.com/mens-college-basketball/story/_/id/13037928/ncaa-announces-rule-changes-2015-16-including-30-second-shot-clock-fewer-outs

Kaggle. (2025). March Machine Learning Mania 2025. https://www.kaggle.com/competitions/march-machine-learning-mania-2025

National Collegiate Athletic Association. (2019, June 5). Men's basketball 3-point line extended to international distance. https://www.ncaa.org/news/2019/6/5/men-s-basketball-3-point-line-extended-to-international-distance.aspx

Yahoo Sports. (2023, March). NCAA basketball betting: MAC leader Toledo looks to break a 42-year March Madness drought. https://sports.yahoo.com/ncaa-basketball-betting-mac-leader-toledo-looks-to-break-a-42-year-march-madness-drought-133311043.html