

Project Part 1

Prithvi Raj, Akhil Rachakonda, Shritan Goki

Link to Github Repository: <https://github.com/whozpj/DataToKnowledgeFinalProject---Prithvi-Raj-Akhil-Rachakonda-Shritan-Goki>

Introduction

The National Basketball Association (NBA) has undergone a dramatic transformation in offensive strategy over the past two decades. The proliferation of three-point shooting, popularized by teams like the Golden State Warriors, has fundamentally changed how the game is played. This shift raises important questions about the evolution of NBA offense and whether these changes have translated to improved scoring efficiency.

This report explores two key research questions:

1. **Three-Point Emphasis:** Has the proportion of total points scored from three-point field goals increased in the NBA from 2019 to 2022?
2. **Scoring Efficiency:** Has overall scoring efficiency, measured as points per field goal attempt, improved among NBA teams from 2019 to 2022?

While broader historical trends (2000-2023) would provide more comprehensive insights, this analysis focuses on recent seasons (2019-2022) for which detailed player-level data is readily available. This four-season window captures the modern NBA era and provides sufficient data to identify meaningful trends. This is an exploratory analysis only, with no statistical inference performed.

Data Summary

Data Source and Collection

The data used in this analysis comes from Basketball Reference (<https://www.basketball-reference.com/>), a comprehensive repository of NBA statistics maintained by Sports Reference LLC. The dataset includes player-level per-game statistics for four NBA seasons: 2018-19 (labeled 2019), 2019-20 (2020), 2020-21 (2021), and 2021-22 (2022).

These data represent a **population** of all NBA players who appeared in games during these seasons. Basketball Reference compiles official statistics recorded by the NBA during each game, including field goals made and attempted, three-pointers made and attempted, points scored, and various other performance metrics.

Data Modifications

After collecting the raw data, several modifications were made:

1. **Season labeling:** Each dataset was tagged with its corresponding season year for tracking trends over time.
2. **Data combination:** All four season datasets were merged into a single dataframe to facilitate year-over-year comparisons.
3. **Missing data removal:** Players with missing values for critical variables (points, field goal attempts, three-pointers) were excluded from analysis, as these represent incomplete records.
4. **Zero-attempt filtering:** Players with zero field goal attempts were removed to prevent division-by-zero errors when calculating efficiency metrics.

These cleaning steps were necessary to ensure accurate calculations and meaningful comparisons. The resulting dataset includes 2851 player-season observations across four years.

Potential Data Issues

Several potential issues should be noted:

- 1. Player-level vs. team-level:** The data are at the player level, but our research questions focus on league-wide trends. Aggregation may obscure team-specific strategies.
- 2. COVID-19 impact:** The 2020 and 2021 seasons were affected by the COVID-19 pandemic, resulting in shortened schedules, altered game conditions, and potential statistical anomalies.
- 3. Sample bias:** Only players who appeared in games are included. Injured players and those on rosters who did not play are excluded, potentially biasing efficiency statistics upward.
- 4. Per-game vs. total statistics:** The data represent per-game averages, which are then summed across players. This approach weights players by games played, which may differ from pure team totals.

Despite these limitations, the data appropriately address the research questions because they capture comprehensive offensive statistics across multiple seasons, enabling calculation of both three-point scoring proportion and overall scoring efficiency at the league level.

Exploratory Analysis

Numerical Summary 1: Season-by-Season Statistics

Table 1: League-wide scoring statistics by season

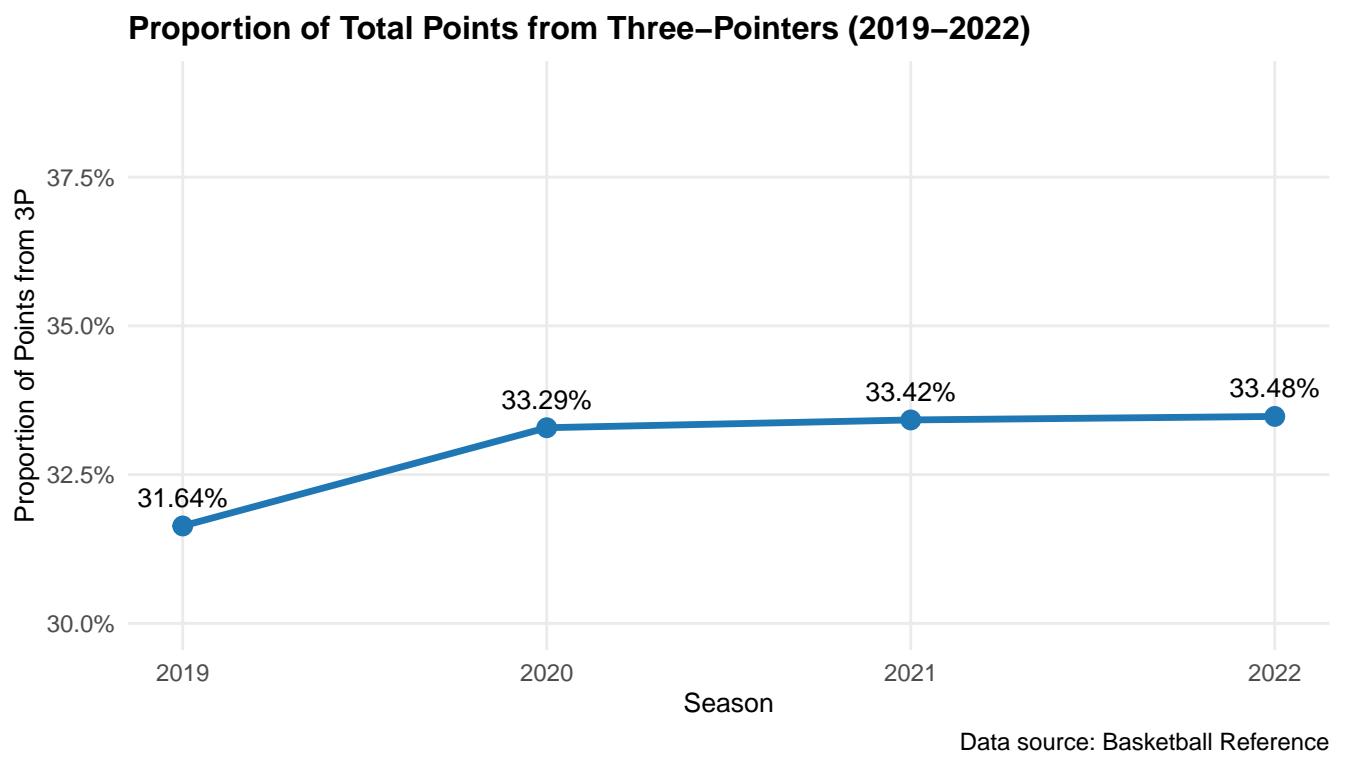
Season	Total Points	3P Points	Total FGA	Prop. from 3P	Pts/FGA
2,019	5,936.1	1,878.0	4,874.6	0.3164	1.218
2,020	5,508.3	1,833.6	4,466.8	0.3329	1.233
2,021	6,074.7	2,030.1	4,896.0	0.3342	1.241
2,022	6,340.0	2,122.5	5,185.7	0.3348	1.223

Numerical Summary 2: Change Over Time

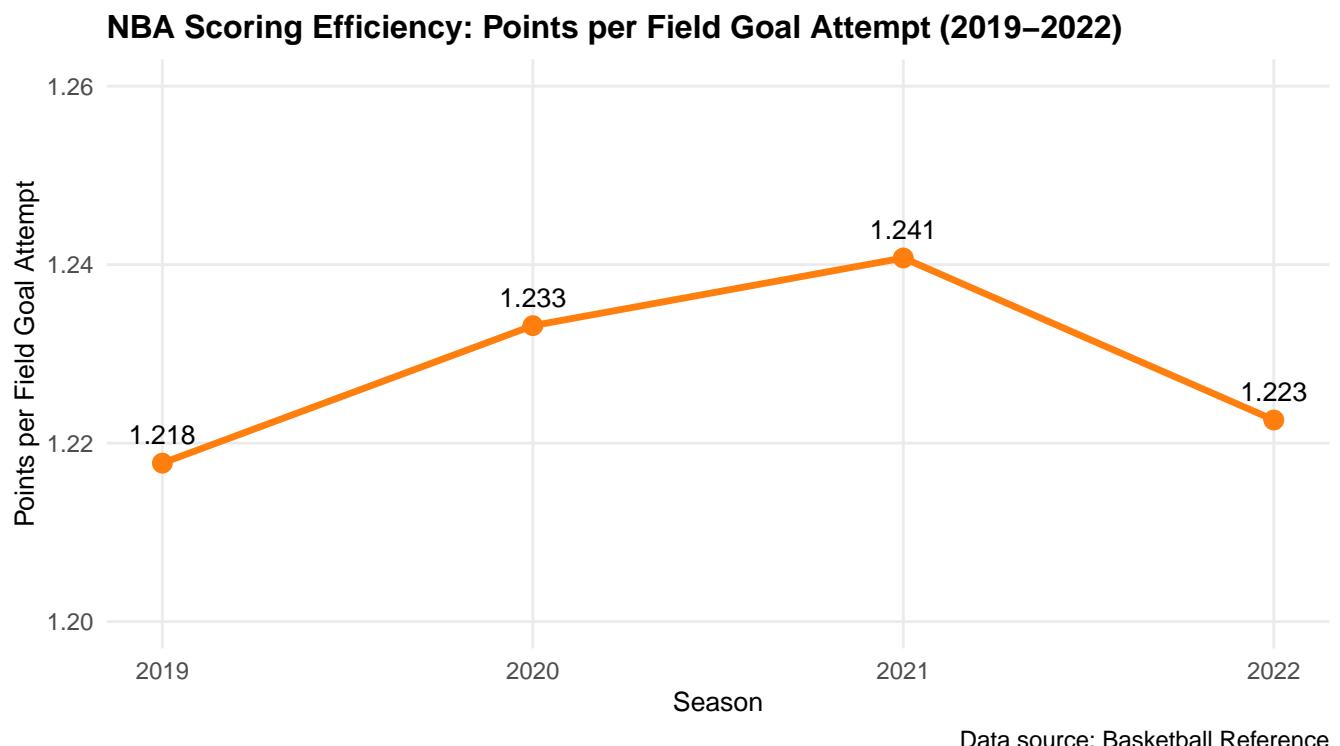
Table 2: Change in key metrics from 2019 to 2022

Metric	2019 Value	2022 Value	Absolute Change	Change %
Proportion from 3-Pointers	0.3164	0.3348	0.0184	5.8191
Points per FGA	1.2178	1.2226	0.0048	0.3967

Graphical Summary 1: Proportion of Points from Three-Pointers



Graphical Summary 2: Scoring Efficiency Over Time



Conclusions

Three-Point Proportion Trends

The data reveal a clear upward trend in the proportion of total points scored from three-point field goals between 2019 and 2022. In the 2019 season, approximately 31.6% of all points came from three-pointers. This proportion increased substantially to 33.3% in 2020, then stabilized at 33.4% in 2021 and 33.5% in 2022, representing an overall absolute increase of 1.9 percentage points and a relative increase of 6.0% from 2019 to 2022.

Statistical interpretation: The trend shows a notable jump between 2019 and 2020 (from 0.316 to 0.333), followed by smaller incremental increases in subsequent years (0.334 in 2021, 0.335 in 2022). This pattern suggests that a significant shift in offensive strategy occurred between 2019-2020, with the new approach becoming the norm in subsequent seasons.

Contextual meaning: This trend confirms that NBA teams are increasingly emphasizing three-point shooting as a central component of their offensive systems. The shift makes mathematical sense: three-pointers are worth 50% more than two-pointers, so even with slightly

lower shooting percentages, they can be more efficient. Teams are recognizing this and adjusting their shot selection accordingly. This evolution reflects broader strategic changes, including increased floor spacing, positionless basketball, and the influence of analytics departments.

Scoring Efficiency Trends

The analysis shows a more complex pattern for overall scoring efficiency (points per field goal attempt). Starting at 1.22 points per FGA in 2019, efficiency increased slightly to 1.23 in 2020 (a 0.82% increase) and peaked at 1.24 in 2021 (a 1.64% increase from 2019). However, efficiency declined back to 1.22 in 2022, returning to the 2019 baseline. The net change from 2019 to 2022 is effectively zero, with no sustained improvement in scoring efficiency over this period.

Statistical interpretation: Unlike the three-point proportion, scoring efficiency does not show a sustained upward trend. The peak efficiency occurred in 2021, but the 2022 season saw efficiency return to 2019 levels. This pattern suggests that any efficiency gains in the middle years were temporary rather than representing a permanent strategic improvement.

Contextual meaning: The temporary improvement in 2021 (peak efficiency of 1.24) followed by decline in 2022 back to 2019 levels suggests that the increased emphasis on three-point shooting has not translated into sustained scoring efficiency gains. While three-pointers should theoretically improve efficiency (worth 50% more than two-pointers), NBA defenses have adapted to defend the perimeter more aggressively. The 2022 return to baseline efficiency indicates that defensive adjustments, rule changes affecting physicality, or officiating patterns may have counteracted the theoretical advantage of increased three-point attempts. This finding reveals an important tension: teams are taking more three-pointers, but the league-wide efficiency remains stable, suggesting a strategic equilibrium between offense and defense.

Summary

This exploratory analysis provides evidence of divergent trends in NBA offense from 2019 to 2022. The proportion of points from three-pointers has increased consistently from 31.6% to 33.5%, reflecting a clear league-wide strategic shift toward perimeter-oriented basketball. However, overall scoring efficiency (points per FGA) has shown no net improvement over this period, returning to 1.22 in 2022 after peaking at 1.24 in 2021.

These findings reveal an important insight: while teams have successfully increased their reliance on three-point shooting, defenses have adapted sufficiently to prevent sustained efficiency gains. The data suggest the NBA may be reaching a strategic equilibrium where offensive and defensive innovations balance each other out.

These findings set the stage for Part 2 of this project, which will employ statistical inference to determine whether the observed changes in three-point proportion represent statistically significant shifts or could reasonably be attributed to chance variation. Understanding these trends has implications for team strategy, player evaluation, coaching decisions, and the future evolution of professional basketball.

References

Basketball Reference. (n.d.). *NBA & ABA League Index*. Sports Reference LLC. Retrieved from <https://www.basketball-reference.com/>

Basketball Reference. (n.d.). *NBA Per Game Stats*. Retrieved from https://www.basketball-reference.com/leagues/NBA_2019_per_game.html

ChatGPT . Used ChatGPT to help grammar check and format written responses as well as debugging plotGG code.

Data files: NBA2019.csv, NBA2020.csv, NBA2021.csv, NBA2022.csv (included in GitHub repository)