Enkidu Bazua-Kelley (bazuakelley.e@northeastern.edu), Pranav Bollineni (bollineni.p@northeastern.edu), Salvatore Gagliardo (gagliardo.s@northeastern.edu), Sophia Hughes (hughes.so@northeastern.edu), Sebastian Romero (romero.seba@northeastern.edu), Dylan Tu (tu.dy@northeastern.edu)

DS 2001 Project Report
7 April 2025

NBA Data Analysis: Growth of the Three-Point Shot

Problem Statement and Background

Over the last few decades, the NBA has changed in a significant way: teams are shooting more three-pointers than ever before, and this completely reshaped how the game is played. What started as a relatively minor part of the game has now become a central strategy, with many teams building their offense around it. Teams now rely on the three-point shot, but it is still not clear whether this shift has led to better results. Are teams that take more threes consistently more successful? Do they win more games or make the playoffs more often? Looking at teams from different eras—it can be seen not just how much more players shoot from deep now, but how much more important that shot has become. In this analysis, we aim to answer how the rise of the three-point shot has impacted scoring, team success, and the overall strategy of the game.

Introduction and Description of the Data

The focus of our project is to analyze the growth and impact of the three-point shot in the NBA. Over the past few decades, the three-point shot has evolved from a rare scoring option to a cornerstone of today's game, with many organizations building their offenses around it. Our motivation for choosing this topic comes from observing how modern basketball has changed and exploring whether this shift has translated into more successful outcomes for teams, such as more wins, playoff appearances, or championships.

To perform this analysis, we used a dataset sourced from Kaggle, titled "game.csv", which includes over sixty-five thousand rows of game-level data dating back to 1946. From this

dataset, we focused on relevant columns relating to three-point shooting, field goal shooting, and win/loss results for both home and away teams, beginning in 1979 when the three-point shot was officially introduced in the NBA. This dataset allowed us to analyze league-wide trends and team-level performance over time. Additionally, we made two CSV files with statistics published on the NBA website that tracks the average number of three-point shots made by zone of the court per game in the regular season from 1997 to 2025, as well as for the two championship teams during the postseason from 1997 to 2024.

Beyond analytical significance, this project also aims to explain the evolution of the game in a historical sense to fans; with controversy surrounding the increasing reliance on the three-point shot in the modern NBA, we plan to analyze and explain these changes to see if the three-pointer has truly evolved for the good of the game.

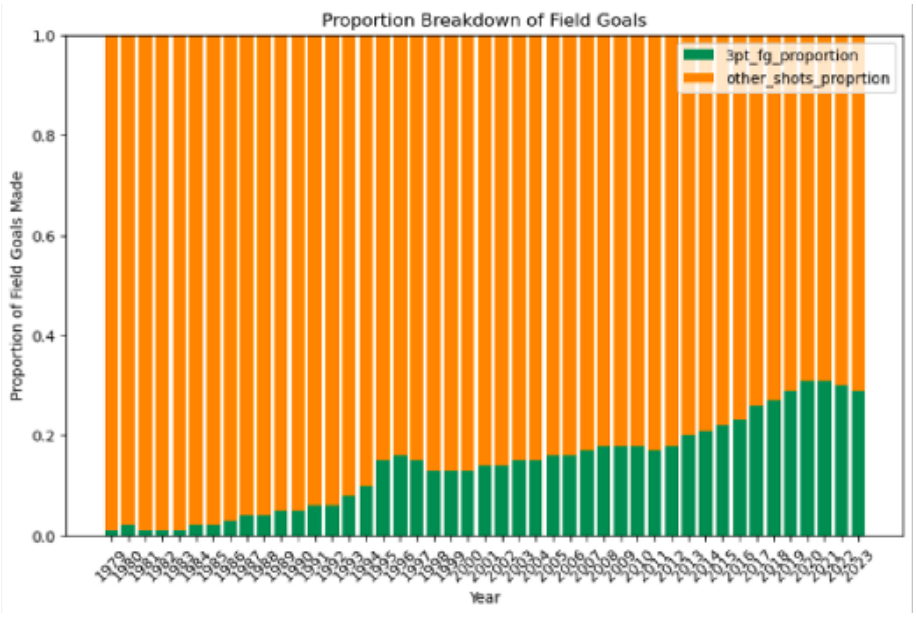
Methods For the Kaggle Dataset

To clean the data and extract the information needed from the "game.csv" file, our programs are designed to only append the value from each row from 1979 that are specific to three-point statistics and winning outcomes. The data was converted to the necessary form, such as floats, allowing for calculations to be performed with these statistics later.

The first function of all the programs reads the data file and returns a dictionary. To ensure the analysis effectively captured the relationship between three-point shooting and team success, the dictionary structure was chosen for its efficiency in organizing data by year, allowing for easy retrieval and trend analysis. When necessary due to program design, a second function was implemented that returned a cleaned and condensed version of the first dictionary to allow for data visualization. To visualize the data, several types of functions were created by utilizing the matplotlib library to render a different type of data visualization. The first function

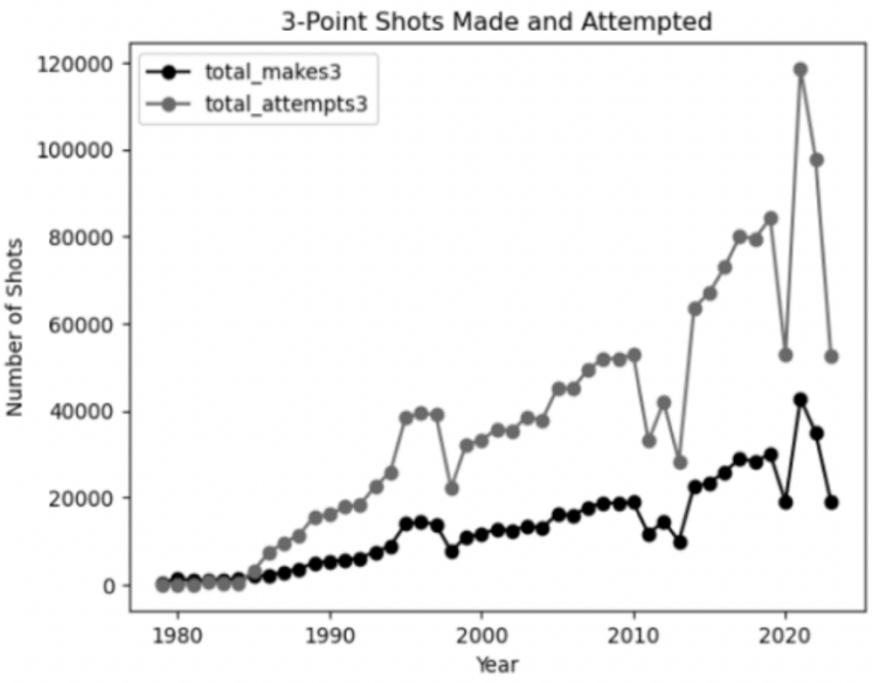
creates a stacked bar chart (Graph 1.1: Stacked Bar-Chart) that demonstrates the proportion of field goals per season from 1979 to 2023 that were three-point shots. Here, it is seen that the use of the three-point shot becomes increasingly higher each year, with more strong increases in 1994 and 2010.

Graph 1.1: Stacked Bar-Chart Proportion of Field Goals



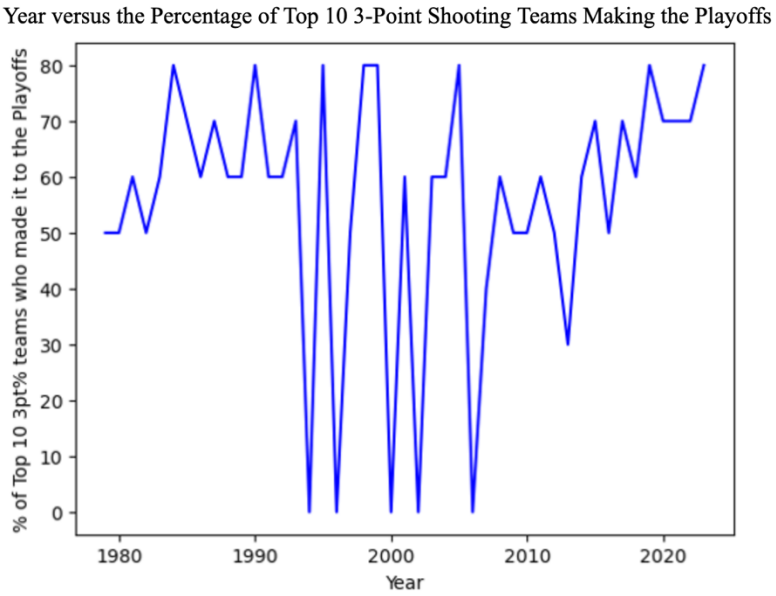
A function was employed to render a line chart visualization, which is used to graph the three-point shots attempted and made from 1979 to 2023 (Graph 1.2: Line-Chart Time Analysis). The data used for the graph came from the returned dictionary after cleaning and condensing the data. As the values in the cleaned dictionary returned a single float for each key in the nested dictionaries, this allowed for easier computations of the proportion of all field goals that were three-pointers for each year. It was found that the line for total attempts closely mirrors the line for total made, suggesting that while more three-point shots are being attempted, the efficiency has remained relatively stable throughout the years.

Graph 1.2: Line-Chart Time Analysis



Furthermore, a line-chart visualization (Graph 1.3: Line-Chart Top 10%) was generated from the “game.csv” file to visualize the three-point shots of the top ten performing teams for each year from 1979 to 2023. This chart utilized functions that were established to calculate the average three-point percentage per team, identify the top 10 teams, and compute what percentage of them made the playoffs. After reading an article published on the Python website, we utilized the library “datetime” library to aid in our program writing (Python Documentation). Our analysis shows that, on average, fifty-eight percent of the top 10 teams in regular-season three-point shooting make the playoffs, and only five seasons in which none of the top ten teams qualified.

Graph 1.3: Line-Chart Top 10%



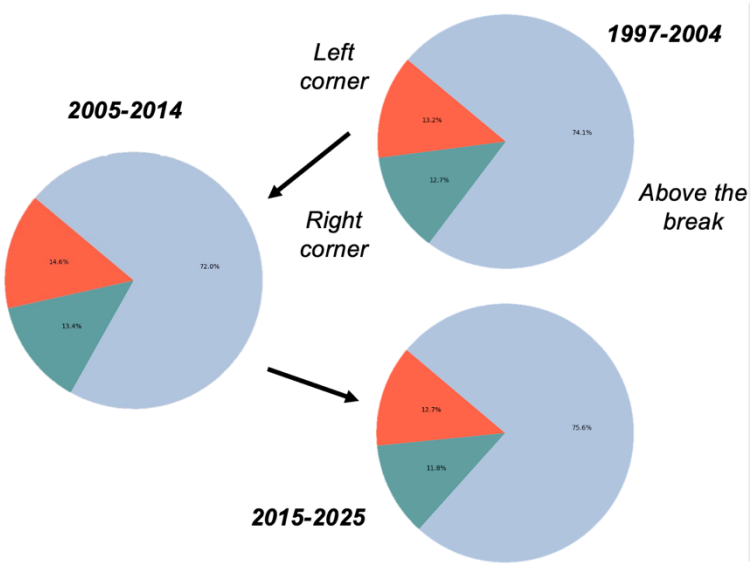
Methods For the NBA Datasets

When looking at three-pointers made per zone with the two datasets made from the NBA statistics, we first standardized all team names in the dataset through Excel to match their current franchise name and city. For example, the Oklahoma City Thunder, up until 2008, were the Seattle SuperSonics, so both teams are attributed as the Oklahoma City Thunder in the dataset.

Similarly to the dataset from Kaggle, our program files for the NBA datasets included a function for reading the CSV file. As there were multiple programs written, some of these programs read the data file using the Pandas library. When Pandas was not utilized, a second function was included in the program that cleaned the data returned from reading the file, to calculate the overall averages for each team for three pointers-made from 1997 to 2025. To understand the evolution of the position of a player taking a three-point chart during the regular season, three functions were implemented to sum the number of shots taken from the left corner, right corner, and above the break. These numbers were then passed through another function that

renders a pie chart visualization. Three pie charts (Graph 2.1: Pie-Charts Shooting Zones) were created to showcase different time periods (1997-2004, 2005-2014, 2015-2025) to analyze whether the playing style has changed. It was found that the split between where shots are taken has not changed much; players have switched to taking more shots behind the arc however they are still taking above seventy percent of their shots above the break. This finding shows the potential for players to improve their shooting and confidence with the corner shot.

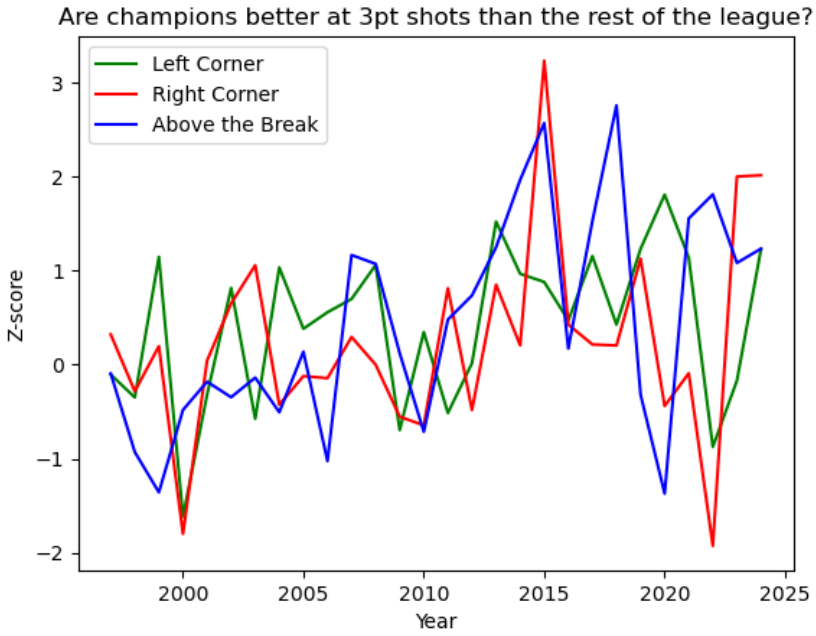
Graph 2.1: Pie-Charts Shooting Zones



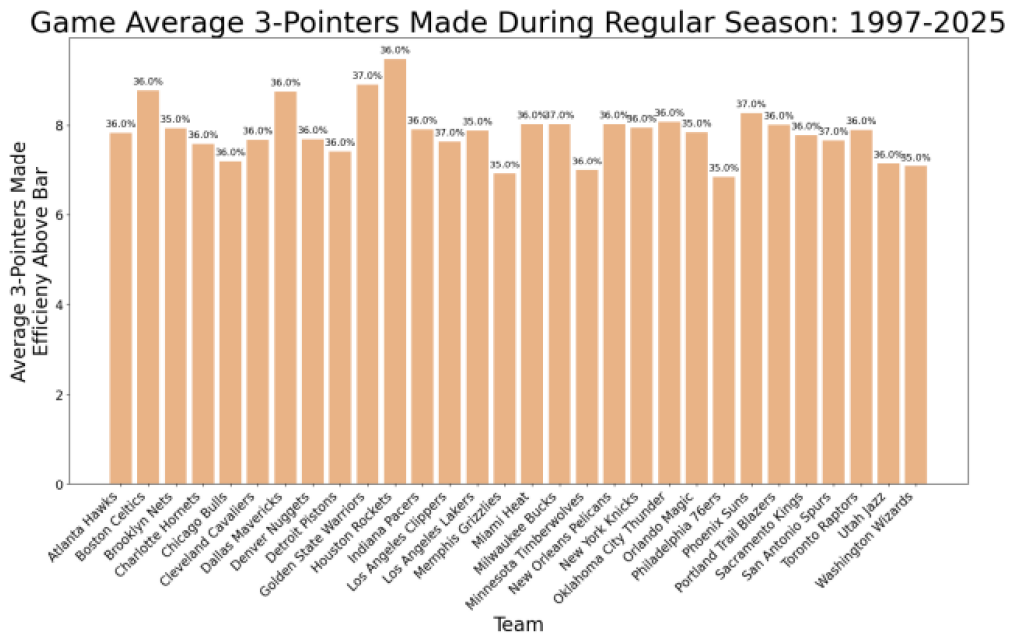
To examine the link between three-point shooting and postseason success, three visualizations were created. The visualization utilized functions to calculate the z-score of regular seasons' three-point percentages (left corner, right corner, and above-the-break) of each championship team from 1997–2024 using Pandas for efficient statistical calculations. Data was restructured into nested dictionaries by year and team, allowing for zone-specific z-score comparisons against league averages. A line plot (Graph 2.2: Z-Scores) was created showing the z-score trends across seasons (green for left corner, red for right corner, blue for above-the-break). It was found that in sixty-eight percent of seasons that occurred from 1997 to 2024,

champions had above-average 3-point performance. This suggests that the number of three pointers is often a contributing factor of team success during the post season, though less dominant in recent years. Additionally, a second visualization takes the returned dictionary from a function that reads the data file and another function that finds the overall average of three-pointer made per game from 1997 to 2025 to make a bar-chart with the shooting efficiencies labeled at the top (Graph 2.3: Regular Season Bar-Chart). After analyzing the top and bottom teams for three-pointers made per game, it was determined that the number of three-pointers made per game is not a determinant on whether a team makes the playoffs. However, the graph suggests a positive relationship between 3-point shooting and team success: embracing high-volume three-point strategies may improve playoff and championship outcomes as the top four teams in average-three-pointers-made-per-game won a combined seven championships whereas the bottom four teams won only one championship combined.

Graph 2.2: Z-Scores



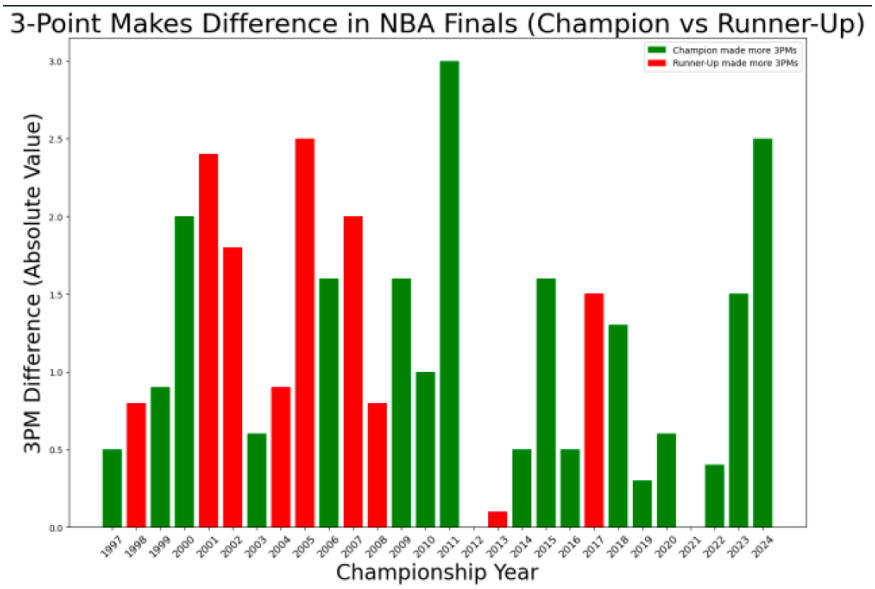
Graph 2.3: Regular Season Bar-Chart



The last visualization utilized the second dataset created from the NBA statistics on postseason three-pointers; this is a bar-chart that shows the difference between the number of three pointers made during the postseason from the championship team and runner-up (Graph 2.5: Difference Bar-Chart). After using a dictionary that a function designated to read the data file returns, a function was created to calculate the difference of three-pointers made by the champion and runner-up teams during the postseason. The cleaned dictionary served as a parameter for a final function that renders the bar-chart, with green representing the team that won the championship made more three-pointers, and red representing the runner-up. From 1997 – 2010, the difference during the postseason is unpredictable with no clear trend on which team will out-perform the other. This possibly shows that the three-point shot was less relied on during these times. From 2011 and onwards, the difference shows a trend where the championship

winning team makes more three-pointers, suggesting that there is a growing importance of three-point shots in postseason and championship success.

Graph 2.5: Difference Bar-Chart



Results, Conclusions and Future Work

We confirmed our hypothesis that three-point shots have increased as a percentage of total shots attempted in the NBA over time. We did observe an inflection in the growth of this ratio in the 2010s, which we believe is also around the time that major sports leagues started emphasizing data analysis as a means of improving and monitoring the health of the sport. Additionally, our analysis indicates that while more three-point shots are being taken (and therefore are made) each game, the scoring efficiency for three-point shots have not changed much since the introduction of the three-point line.

A shortcoming of our project is that we were unable to get data beyond the frequency of a game. We would've liked to see shot-by-shot data, which we could not easily source or access. This would have allowed us to create an interesting visualization of where three-point shots are attempted, and perhaps even explore where specific players are weak or strong at shooting.

Given more time and with the data we have access to, we would have explored team-specific strengths and weaknesses. Since our research question pertains to the league, we did not focus on a specific team. We did differentiate between teams that made it to the playoffs versus those who did not, but avoided analyzing how teams evolve over time. Due to players, coaches and owners changing, we did not think this would be an honest analysis without dimensions of data that is not available.

References

Datetime - Basic Date and Time Types. Python Documentation. (n.d.-a). https://docs.python.org/3/library/datetime.html

GeeksforGeeks. (2025, January 30). *Appending element into a list in python dictionary*. https://www.geeksforgeeks.org/appending-to-list-in-python-dictionary/

Sorting techniques. Python Documentation. (n.d.-b). https://docs.python.org/3/howto/sorting.html

Teams shooting: Stats (Playoffs). NBA. (n.d.-a). https://www.nba.com/stats/teams/shooting?DistanceRange=By%2BZone&SeasonType=Playoffs&Season=2023-24

Teams shooting: Stats (Regular Season). NBA. (n.d.-b). https://www.nba.com/stats/teams/shooting?DistanceRange=By%2BZone

Walsh, W. (2023, July 6). *NBA Database*. Kaggle. https://www.kaggle.com/datasets/wyattowalsh/basketball