INFO 348 Final Project Report - NBA Three-Point Shooting Revolution

Sam Warren and Brennan Frost

1- Goals of the Project

If you decide to watch an average NBA game from 2002, and compare it to an average NBA game now, the changes will seem minuscule. The games are still generally high scoring, the playmaking skills of the world's most talented players is still on display, and you still have guards who control the pace of the game and forwards who bruise in the paint and block shots at the rim. However, there has been one significant change that has effectively taken over the league; the three point shot. In 2002, the average number of three pointers attempted in a game was 14.7. In 2024, that number has since skyrocketed to 35.1. Over the course of the past two decades, the three point shot has changed the way teams think about scoring. It has changed the way teams build rosters. With this in mind, we wanted to measure the effect the three point shot has on the game today, and to try and visualize the evolution of scoring in the NBA. Our three main goals for this project were; a) How has scoring evolved throughout the years? b) What things are more or less prevalent? What is the most common form of scoring this year? c) Is there a scoring method that leads to the most wins? Least wins?

2- Dataset Description

The data that we used was pulled from a website called Basketball-Reference (https://www.basketball-reference.com/leagues/NBA_2025.html). Basketball-Reference houses all kinds of individual and team basketball stats for over 50+ years of NBA history. For any given season, Basketball-Reference provides tables that list Per Game, Total, Per 100 Possessions, Advanced, and Shooting stats for all 30 teams, as well as individual stats for specific categories. For our project, we focused on three of these team stats tables;

a) Per Game Stats, which tallied the average per game statistics of basic categories (PTS, FG, 3PT). In total, there were 30 observations (30 NBA teams) across 22 different variables.

b) Advanced Stats, which involves more complex statistics like offensive rating and the four factors (eFG%, TOV%, RB%, FT/FGA). In total, there were 30 observations (30 NBA teams) across 26 different variables.

c) Shooting Stats, which specifically calculated the field goal and field goal attempt percentages by distance. In total, there were 30 observations (30 NBA teams) across 22 different variables. In total, there were 30 observations (30 NBA teams) across 26 different variables.

Dataset Manipulation in the Task Writeups section describes how we wrangled the data so that we combined all three tables and compiled data across 22 seasons.

3- Task Writeups

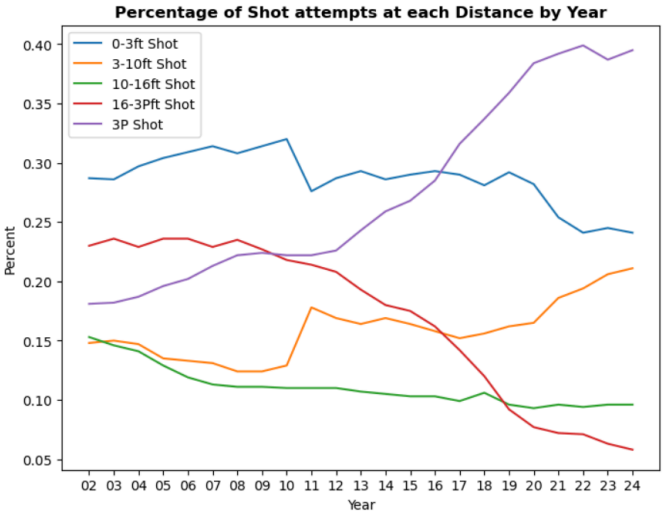
A- Dataset Manipulation

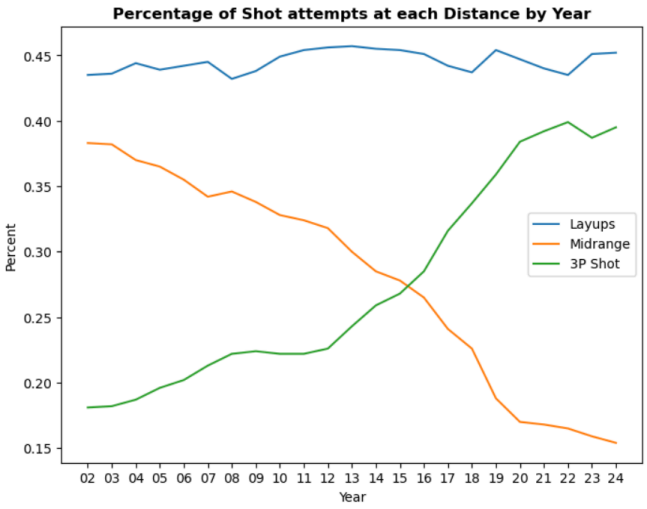
Our code collected and processed NBA statistics spanning multiple seasons by scraping data from Basketball-Reference. We used the *requests* library to fetch HTML content and *BeautifulSoup* library to extract three statistical tables for a season: per-game stats, advanced stats, and shooting stats. These tables were converted into Pandas DataFrames, where irrelevant columns were dropped, column headers were refined, and the data was merged. To uniquely identify teams, we appended the year to team names, and playoff qualification was removed from detecting asterisks in team names.

Our core function, *getData*, iterated through a list of NBA seasons, calling the function for each year. The resulting per-game and shooting DataFrames for each season were stored in lists and then concatenated into two DataFrames: *PG_df* (containing per-game and advanced stats) and *sht_df* (for shooting stats). Historical naming issues ("New Jersey Nets" to "Brooklyn Nets") were resolved to reflect the current franchise names. Finally, we created a final dataset, *shooting_df*, that contained all the columns we would need for analysis.

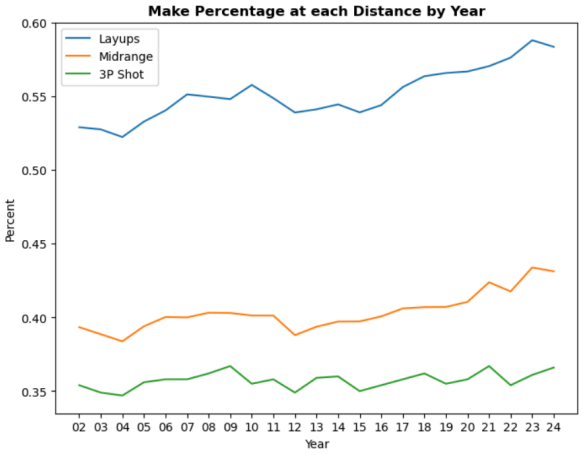
B- Shot Comparisons

Using the data from *shooting_df*, we analyzed and visualized team and league shooting trends in the NBA across different shot distances and time periods. We executed multiple matplotlib and seaborn plots detailing the per-year league average of shot attempts at each distance. We started with plotting five different distance categories but then combined them into three different distance categories that are universally used throughout basketball circles; layups, mid-range, and three pointers. The main takeaway is that the percentage of three-pointers taken by a team (3PAr) has increased significantly (+21.4%), and the percentage of mid range attempts by a team has decreased significantly (-22.9%), while layup attempts have generally stayed consistent over the past two decades (+1.7%). We can see that this trend started occurring around the 2011-12 NBA season. Another interesting takeaway is that while 3PAr has risen significantly since 2011-12, the make



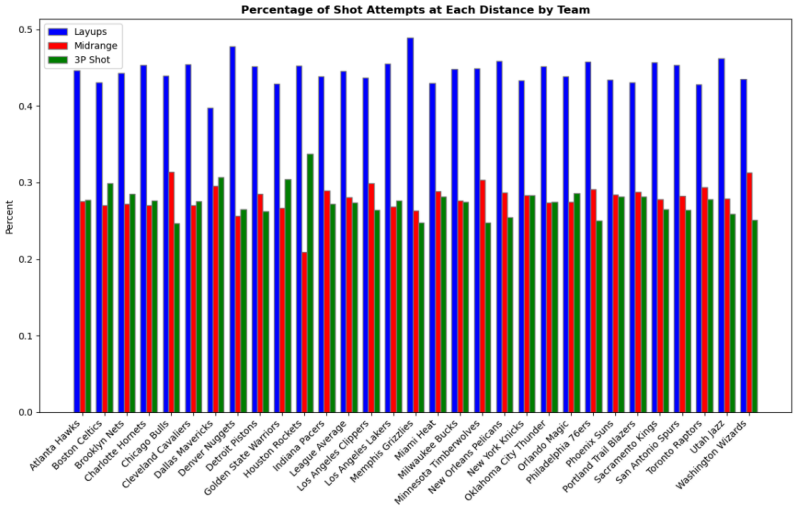


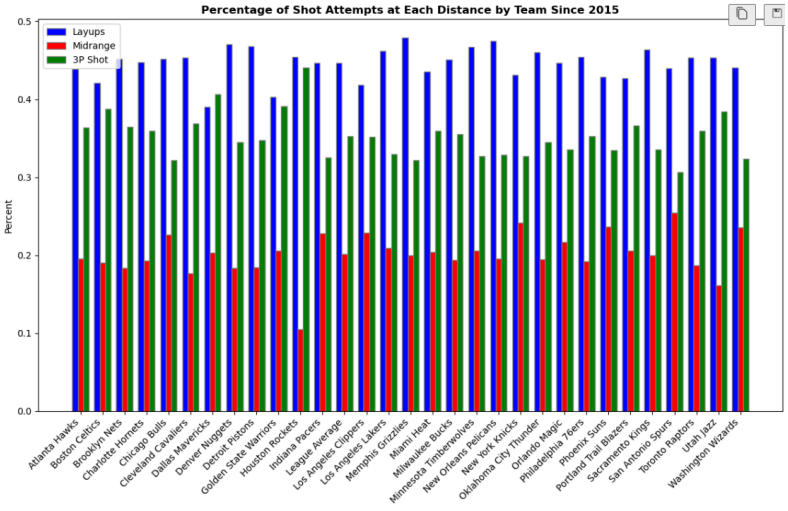
percentage of three-pointers has generally not, while the make percentage of mid-range and layups has risen (albeit by smaller margins).



C- Team-to-Team Comparison

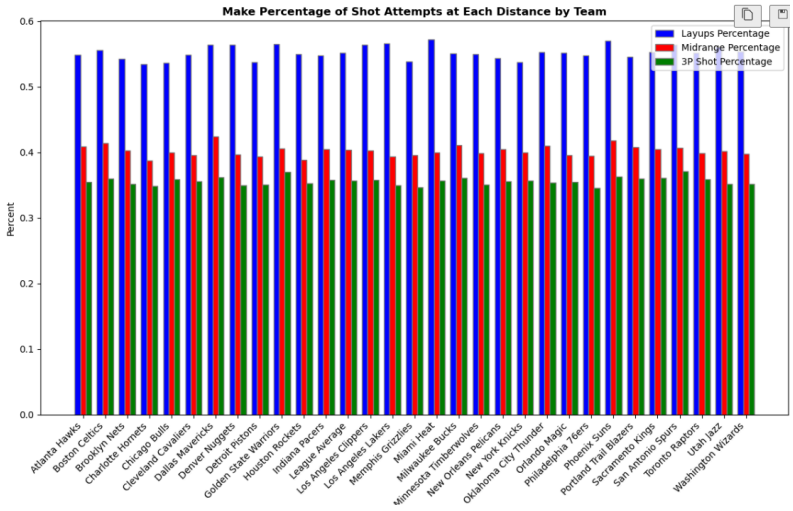
Continuing with the data from *shooting_df*, we analyzed and visualized team-by-team comparisons of shooting trends in the NBA across different shot distances and time periods. We first isolated the numeric columns from the data frame *shooting_df*, and then created a function that standardized team names by stripping out trailing characters. This removed the year from teams changing "Boston Celtics24, to Boston Celtics, making it possible to group teams together over years. We needed to group the data on a team and compute means on all numeric columns for the creation of one dataset comprising all years and another for recent years, starting from 2015. That left us with two team-averaged datasets, which were used to create clustered bar charts again with matplotlib for comparison.

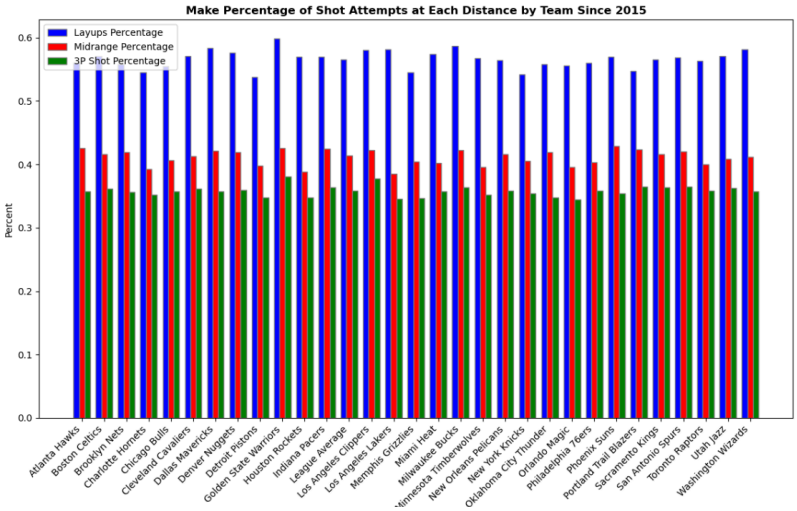




The charts above show the average shot distance per team. With the left showing since 2002, and the right since 2015. As you can see there are some drastic outliers in the dataset. Notably, the Houston Rockets have a very high 3P attempt rate and the lowest midrange attempt rate since 2002, but since 2015 these numbers are even more drastic. With only 10% of their attempted shots being from the midrange, and 44% from the 3-point range. Additionally, the Chicago Bulls and Washington Wizards have the highest midrange attempt rates. Looking at these same teams since 2015, they continue to be in the top 3 of midrange attempts, but it's also

important to note that their 3-point attempt rates are in the bottom 3 of the league. It's pretty clear that as a team's gameplan, they prioritize different shots.



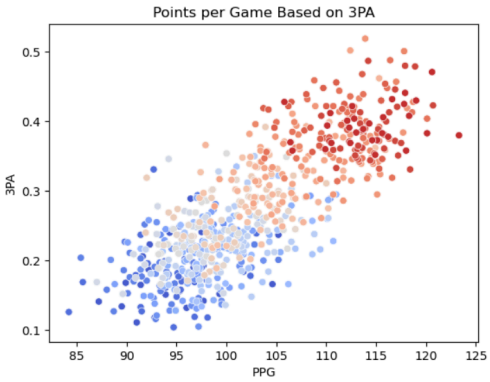


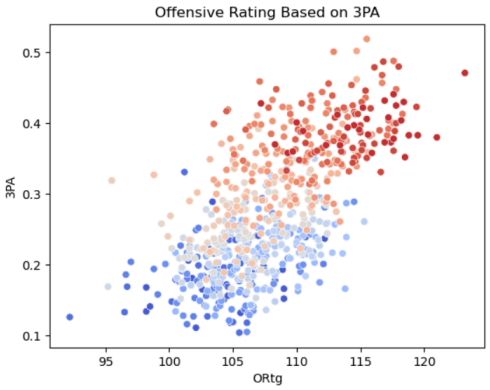
The two charts above show the average shot percentage from the same distances as above. The chart on the left represents all data since 2002, and the right since 2015. While the attempt percentages from these distances are vastly different, the make percentages are relatively consistent between teams. In general, both since 2002 and 2015 the differences between the highest and lowest make percentages from each distance vary 3-4%. This difference is relatively low, whereas differences in attempt percentage vary more than 10%, and get even bigger when comparing only since 2015. Overall, over the last 22 years, there has been relatively little difference between the make percentage of teams, from each shot distance.

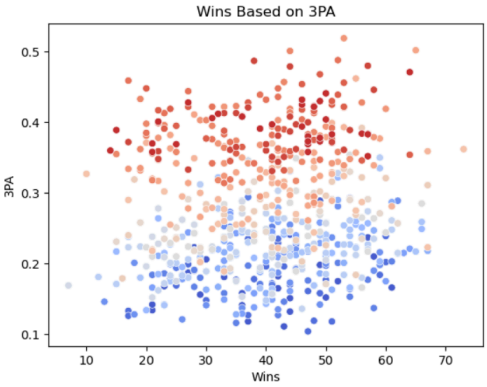
Below there is a table that contains the highest and lowest attempts and percentages of teams. It contains this information for the NBA since 2002, and since 2015.

Statistic	Attempt %	Attempt % since 2015	Make Percentage	Make Percentage since 2015
Highest Layup	0.489	0.479	0.572	0.598
Lowest Layup	0.398	0.390	0.534	0.537
Highest Midrange	0.314	0.254	0.424	0.429
Lowest Midrange	0.209	0.105	0.388	0.385
Highest 3P	0.338	0.441	0.371	0.381
Lowest 3P	0.247	0.307	0.346	0.345

D- Regression Analysis







Based on the shot distance data above, 3P shot attempts have the highest difference over the years. The three scatterplots above represent the correlation between the 3-point attempt rate, and points per game (PPG), offensive rating (ORtg), or Wins. Offensive rating is a statistic that represents the amount of points a player or team will score in 100 possessions. The color of the points on the plots represents years on a scale, with dark blue being 2002, and dark red being 2024. From the three plots, it's clear that PPG and ORtg have a positive correlation with 3PA, that is, as there are a higher percentage of 3-point shots attempted, to all total shots, the higher the ORtg, and PPG for that team.

After creating these plots, we ran regression analysis using each of the shot distances to predict Wins, ORtg, and PPG. The regression analysis for each label was run 200 times, and the average r^2 value was calculated. The predictors were as follows:

Results for target: Wins
Average r^2 Score over 200 runs: 0.042

Results for target: ORtg
Average r^2 Score over 200 runs: 0.433

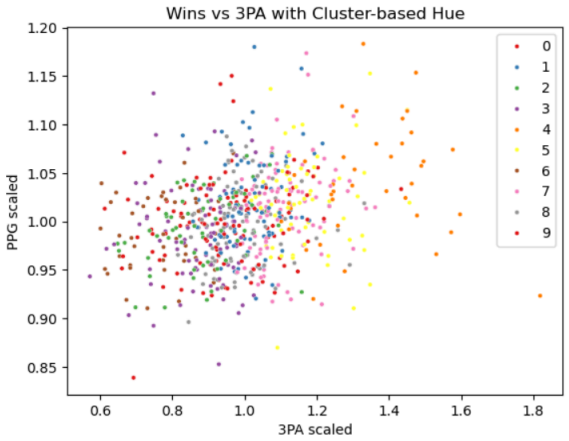
Results for target: PPG
Average r^2 Score over 200 runs: 0.674

Based on these r^2 values and scatterplots it is clear that the rate at which a team shoots 3-pointers has little to no effect on the amount of wins a team has. Additionally, while the scatterplots showed a correlation between 3P attempt rate and PPG and ORtg, the regression analysis shows a different story. With one of the r^2 values being pretty low at 0.433, and the other giving an ok result at 0.674 it suggests that the way a team scores probably doesn't affect scoring. It's important to note that scoring throughout the league has also increased steadily. In general, while there is a correlation between the increase in 3P attempt rate and PPG we cannot

definitively say that the increase in PPG is because of the changes in shot distances. In conclusion, the relation between shot distances and points per game would benefit from further analysis, including comparing PPG independently from defenses or another scaled metric that focuses solely on the offensive output of a team.

Finally, we used the make percentage of 3-point shots to predict the rate of attempts. Does the amount of shots taken depend on the percentage of successes? After running the regression 200 times the average r^2 Score over 200 runs: was -0.033. The r^2 is negative for this association. This means that using the regression model in the way we did, it predicted worse than random. This is overwhelming evidence that the make percentage of 3-point shots does not affect the rate of attempts.

E- Classification Analysis



To further our analyses, we conducted KMeans clustering with 10 clusters to group teams based on similar patterns in statistics like 3PA and PPG, scaled to the league average. While there is no great way to analyze the specific clusters, we looked over the clusters manually and picked out some of the teams with similar statistics. High three-point teams were clustered together in cluster 4. The 2016-17, 2017-18, and 2018-19 Houston Rockets were clustered with 2015-16 and 2017-18 Golden State Warriors teams, 2022-23 Boston Celtics, and the Utah Jazz teams from 2016-17 through 2020-21. Cluster 5 was composed of teams that had even higher 3-point attempts, with the 2023-24 Boston Celtics and Dallas Mavericks, 2022-23 and 2016-17 Golden State Warriors, 2016-17 Cleveland Cavaliers, and the 2013-14 Miami Heat. Teams that played generally slower than others were clustered together in Cluster 9. The San Antonio Spurs from 2001-02 to 2004-05, the Miami Heat from 2018-19 to 2022-23, the Dallas Mavericks from 2019-20 to 2022-23, and the Milwaukee Bucks from 2019-20 to 2022-23. Finally, low three-point attempting teams like the 2023-24 Los Angeles Lakers the 2023-24 Denver Nuggets, and many early 2000s teams grouped together in Cluster 3. Overall by clustering, we were to associate teams from different years that would have never played against each other.

4- Challenges

There weren't many challenges that we came across in completing this project. However, one challenge we did face happened when we were executing our data wrangling. When looking over the data, there were many cases of the data being dual-indexed. There was a lot of time spent making sure that each data frame was properly formatted the right way so that they could be merged without issue. The process was tedious, but in the end it was beneficial because it

made it possible to compare important stats together. Without formatting the datasets together we wouldn't have been able to compare team statistics like Wins, Offensive rating, and Points per game.

5- Insights

From our analysis, we saw that the NBA has evolved significantly over the past two decades. Even smaller, the NBA has evolved significantly from 2015! The need for the three point shot has only increased, driven by advancements in league-wide available analytics. After conducting a little more research, we read how teams have embraced the mathematical efficiency of three-point shooting to maximize points per possession. A successful possession requires an average of one point, achievable by shooting 33% from three-pointers or 50% from mid-range and layups. That 17% gap in shooting efficiency makes the three-pointer more valuable, as long as teams are taking good three-pointers. Last season and now into this season, the Boston Celtics have taken the league by storm not just with how many three pointers they take, but how a large percentage of those three pointers are classified as "good shots". It played a factor in their championship run last spring and summer, and it will play a factor in the quest to repeat as NBA champions next spring and summer.

6- Future Exploration

There are many avenues for future exploration. One could focus on predicting the upper limit of three-point shots in the NBA. While the current trend shows a sharp increase in the reliance on three-pointers, it raises the question of when this growth will plateau. Another avenue would be to explore how coaching changes of player personnel shifts influence shot attempt distributions. For example, does the hiring of a coach with a reputation for analytics-driven strategy correlate with a higher 3PAr, and potentially wins? One last one could be exploring the defensive impact of the three-point revolution. As three-point attempts continue to rise, how do defensive strategies to counter this trend develop?

7- Participation

Sam Warren: Tasks 1, 2, 3 (worked on final report together)
Brennan Frost: Tasks 4, 5 (worked on final report together)