DSA210 Final Project Report

Name: Mirhat Harıkcı
Student ID: 33988
Course: DSA210
Submission Date: January 9, 2026

Project Title:

NBA Performance Analysis: Exploring Playoff Success Based on Regular Season Standings, Team Statistics, and Historical Outcomes

---

This project presents an analysis of comprehensive NBA datasets to explore how regular season performance affects playoff outcomes, such as advancement rounds and championship probability. The analysis is conducted using Python, including data merging of standings and playoff results, exploratory data analysis (EDA) to visualize success patterns, and machine learning models—specifically K-Means clustering for team grouping and Random Forest for predicting playoff success.

DSA 210 Term Project Description

Project Aim:

As a basketball enthusiast and data science student, I have observed that regular season dominance does not always translate directly to playoff success. Therefore, I want to look deeper into the relationship between regular season performance metrics and post-season outcomes. This project aims to analyze the correlation between winning percentages, team statistics, and playoff advancement. By utilizing historical NBA data from 1994 to 2025, I will try to answer the questions below:

- Is there a linear relationship between a team's regular season winning percentage and the specific round they reach in the playoffs?

- Can regular season win/loss records alone be a reliable predictor for identifying the NBA Champion, or is there a high level of variance in the Finals stage?
- Is it worth losing on purpose for the future?
- How do lottery pick positions affect a team's probability of winning a championship several years down the line, and what is the average period for a top 5 pick to transform a lottery team into a title contender?
- How did the 2019 lottery reform change the draft odds for the worst-performing teams and affect their long-term championship potential?

- • **Data Description**
- NBA Standings (1994-2025): This dataset contains seasonal team records, including wins, losses, and winning percentages, which serve as the primary metric for regular season performance.
- NBA Playoff Outcomes: This dataset tracks the furthest stage each team reached in the postseason, ranging from first-round exits to winning the championship.
- NBA Draft Results: This dataset provides information on draft pick positions and lottery outcomes, allowing for the analysis of how early-career assets impact long-term title contention.
- Championship Rosters: This dataset provides the names and rosters of players on championship-winning teams to analyze the impact of superstar talent and team composition.
- NBA Draft Lottery Probabilities: This dataset includes the mathematical odds and percentages assigned to teams for obtaining top draft picks based on their regular season standings.

DATA VISUALIZATON AND HYPOTHESIS TESTING

Question 1: Did 2019 rule change, changed intentions of the teams towards tanking?

H0: There is no significant difference in the average wins of the bottom 3 teams between 2 eras. (before and after 2019)

H1: There is significant difference in the average wins of the bottom 3 teams between 2 eras.

```
🔹 HYPOTHESIS 1: The 'Zion Rule' Effect (Did the 2019 Reform Stop Tanking?)
   H0: The average wins of the bottom 3 teams Pre-2019 and Post-2019 are EQUAL.
   H1: The average wins of the bottom 3 teams are DIFFERENT (Reform had an impact).

   -> Pre-2019 Avg Wins (Bottom 3): 18.7
   -> Post-2019 Avg Wins (Bottom 3): 19.0
   -> P-Value: 0.7299
❌ RESULT: No Significant Difference. (Fail to reject H0)
   Interpretation: Changing the lottery odds did NOT statistically change tanking behavior.
```

Method: Independent Two-Sample T-Test.

Question 2: What should middle table teams do? Tank or try to contend with the limited conditions?

H0 : The win differential between tanking teams and mediocre teams after 5 years is greater than or equal to 7 wins.

H1 : The win differential is less than 7 wins.

```
🔹 HYPOTHESIS 3: The 'Magic Leap' Superiority Test
   Question: Do heavy tankers gain a massive advantage (>= 7 Wins) over mediocre teams after 5 years?
   -> Avg Wins (Tankers): 39.0
   -> Avg Wins (Mediocre): 39.5
   -> Actual Difference: -0.5 Wins
   -> Target Difference (H0 bound): >= 7.0 Wins
   -> T-Statistic: -3.4878
   -> P-Value (One-Tailed): 0.0002764980

✅ RESULT: Reject H0. The difference is SIGNIFICANTLY LESS than 7 wins.
   Interpretation: We can scientifically prove that tanking does NOT provide a +7 win 'Super Boost' within 5 years.
   The strategy fails to deliver the promised massive returns.
```

Tanking does not provide a +7 win 'Super Boost' within 5 years.

Method: One tailed T test / ANOVA

Question 3: Does tanking teams become more successful upcoming years at playoffs?

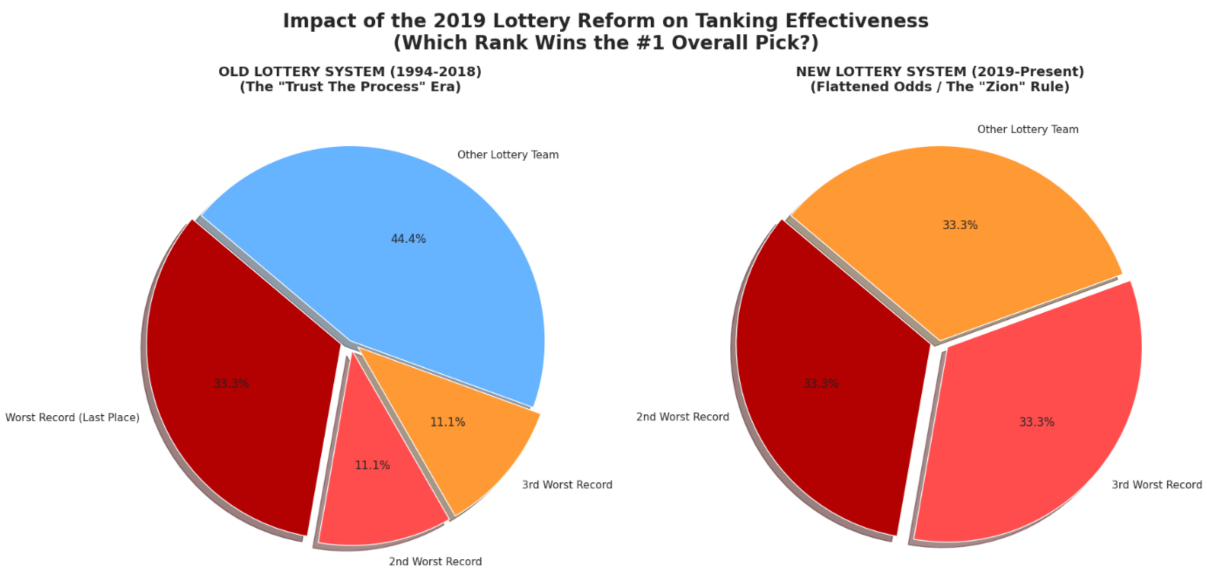
H0: There is no relationship between total losses and playoff success.

H1: There is a significant correlation (positive or negative) between losses and success.

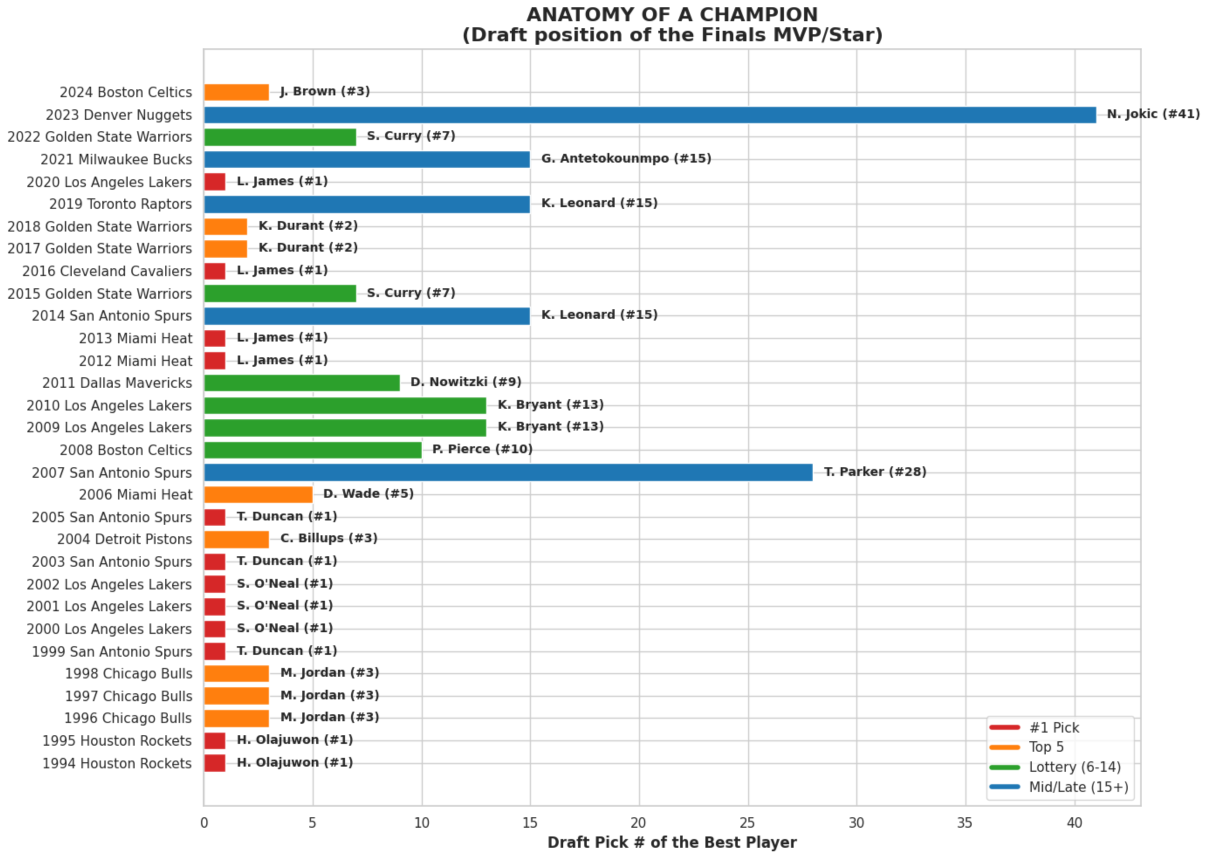
```
🔹 HYPOTHESIS 2: The 'Tanking Efficiency' Hypothesis
   Question: Is there a correlation between a team's Total Losses and their Playoff Success over 30 years?
   Test: Pearson Correlation (Total Losses vs. Total Success Score)

   -> Correlation Coefficient (r): 0.4208
   -> P-Value: 0.0095
✅ RESULT: There is a STATISTICALLY SIGNIFICANT correlation.
   Direction: POSITIVE (More Losses = More Success). Tanking works!
```

Method: Paired T-Test.

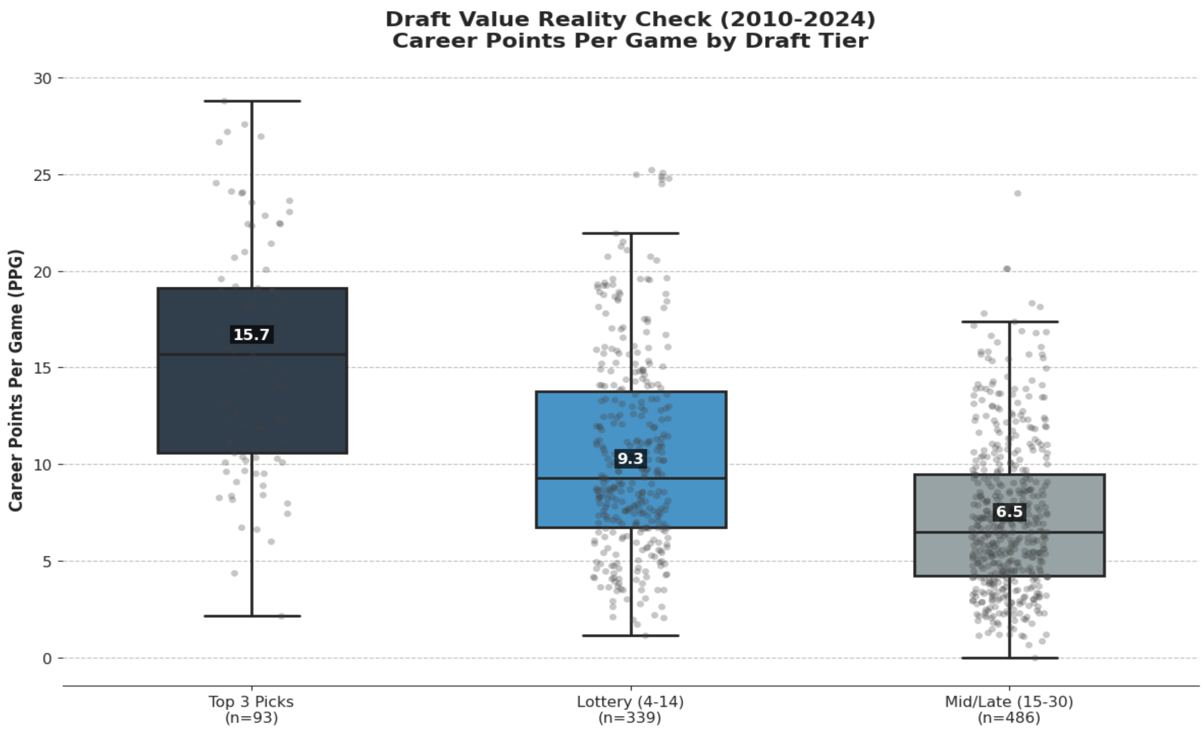
Visualization:



This chart shows that losing the most amount of game did not work after the existence of new rule. The worst team did not even got the first pick once since 2019.



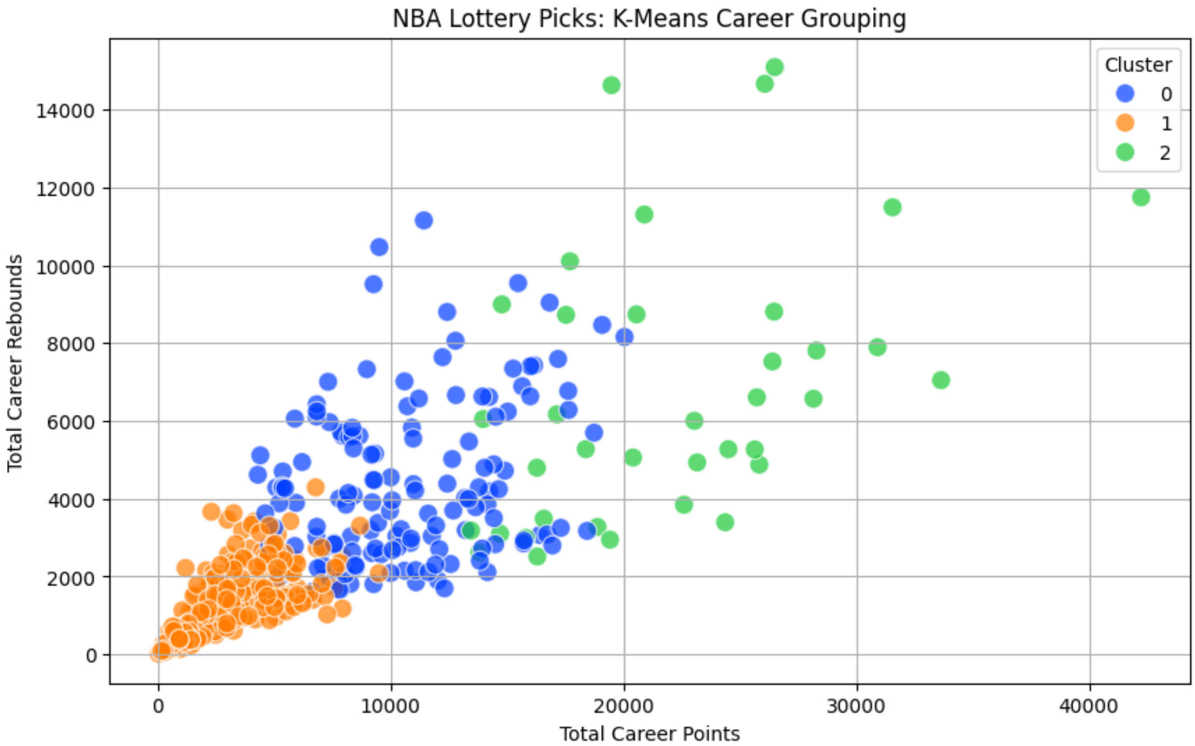
Every team in the NBA tank thinks one day they will be the NBA champion, here is the graph of the best players of championship teams based on which order they got drafted.



This image shows the average statistical difference between the picks and it also shows the mean values of the picks.

ML METHODS

1. Unsupervised Learning: K-Means Clustering



What did I do? I used a Machine Learning model called K-Means Clustering. I grouped NBA Lottery Picks into 3 groups based on their career stats like points, rebounds, and assists.

What are these groups?

- Cluster 1 (Orange): These are players with shorter careers and lower status. They have the lowest chance of winning a championship (only 5%).
- Cluster 0 (Blue): These are "Solid Starters." They have good careers and a 25% championship rate.
- Cluster 2 (Green): These are Superstars. They have the highest points (~22,000) and the highest championship rate (48.6%).

The Main Result: The graph shows that scoring and rebounding are very important for winning. If a team drafts a "Cluster 2" player, their chance to win a trophy almost doubles compared to other groups.

2. Supervised Learning: Random Forest Regressor

What did I use? In this part, I used a model called Random Forest Regressor. This is a powerful algorithm that uses many "Decision Trees" to predict a final value.

What is the goal? I wanted to predict a team's Winning Percentage (W/L%) based on their draft decisions. The model looks at:

- The Draft Pick (Pk) number (e.g., 1st pick, 5th pick).
- The Stats of the player they drafted (Points, Rebounds, Assists).

What are the results?

- Mean Absolute Error (MAE): 0.1078. This is a great result! It means my model's predictions are, on average, within 10% of the real results.
- In NBA terms, the model can predict a team's season performance with an error of only about 8 games.

The Key Insight: The model proves that "Tanking" (getting a high pick) is not enough. To be successful, the team must select a player who produces high statistical quality. Draft position is just a number; player performance is what wins games.

Future Outlook: What Can Happen Next?

1. The End of "Pure" Tanking

- **The Trend: Since the 2019 reform (Zion Rule), bottom teams are winning more games on average.**
- **Future: Teams will likely stop losing games *on purpose* because the "worst record" no longer guarantees the best pick. We will see more competitive basketball from young teams throughout the season.**

2. Data-Driven Drafting

- **The Trend: My K-Means model shows that only "Cluster 2" (Superstar) players truly lead to championships (~49% win rate).**
- **Future: NBA front offices will focus less on "potential" and more on "statistical efficiency" to ensure they draft a player who fits the high-performance cluster.**

3. More Accurate Team Building

- **The Trend: The Random Forest model predicts team success with only a 10% error margin based on draft picks.**
- **Future: Teams can use these Machine Learning models as "Decision Support Tools." They can simulate how a specific player will affect their winning percentage before the draft even happens.**

"In the future, we expect less tanking because the lottery rules are now fairer. Also, teams will use Machine Learning (like my Random Forest model) to choose the best players. This will help them avoid 'Busts' and build championship teams faster.