Analyzing NBA Performance Trends Through a Database-Driven Visualization Framework

Alejandro Cruz

Deepak Paudal

Krutik Soni

Jonathan Walter

Jordan Staley

Alberta School of Business, College of Social Sciences and Humanities, University of Alberta

Master of Management Analytics

MMA 604 - Database Fundamentals for Business Analytics

Dr. Yonghua Ji

December 4th, 2025

Table of Contents

1. Introduction

The National Basketball Association (NBA) is one of the most prominent professional sports leagues in the world. It consists of 30 teams that compete each year from October to April in an 82-game regular season. The top teams advance to the playoffs, a multi-round elimination tournament that determines the league champion each June ("National Basketball Association," 2025). Beyond being a sports competition, the NBA operates as a major global entertainment enterprise and generates significant revenue through broadcasting rights, ticket sales, sponsorships, merchandise, and international markets ("National Basketball Association," 2025). With a growing global audience and a strong emphasis on data-informed decisions, the league provides an ideal environment for applied analytics.

Analyzing NBA performance is important because franchises rely heavily on analytical insights to guide business decisions that affect both competitive results and financial outcomes. Understanding how players and teams perform over time enables organizations to optimize salary investments, evaluate roster construction strategies, forecast attendance, and assess how on-court performance influences fan engagement and revenue streams. Performance patterns also support decisions on player development, contract negotiations, and long-term team improvement planning. By examining trends in consistency, efficiency, and team success, analysts can uncover relationships that help franchises allocate resources more effectively and identify opportunities for strategic advantage. This type of analysis aligns closely with the broader business questions that guide our project and informs decisions that impact the overall performance and sustainability of NBA organizations.

The objective of this project is to complete a full business analytics workflow that integrates data acquisition, database design, and visualization to understand key factors of the NBA. Our goal is to gather data from multiple NBA sources, process and store it in a relational MySQL database, and analyze performance trends through Tableau. By building a unified and well-structured analytical environment, the project aims to generate insights that support strategic and business decision-making for NBA organizations.

2. Motivation

The motivation behind this project comes from our strong interest in the sports industry, an environment where data has long played a central role in guiding strategic and operational decisions. We find basketball particularly engaging and are motivated by the opportunity to analyze player performance trends and team behaviors across NBA seasons. By exploring how data reflects competitive dynamics and organizational decision making, we aim to gain a deeper understanding of the league and extract meaningful insights. These insights can be valuable for stadium managers, executives responsible for player contracts, and other personnel involved in team planning and performance evaluation. This project also allows us to develop practical analytical skills while generating relevant information that supports informed decision making within NBA organizations.

The central objective of this project is to address the key questions outlined in our project proposal, which guide the analytical direction of our work. These questions are:

1. Which players have shown the most consistent performance across multiple seasons?
2. How do rookie players compare to league veterans in key performance metrics?
3. Can we identify trends in player performance that align with All-Star selections or MVP awards?

4. Are certain positions associated with higher scoring or assists?
5. What is the relationship between team performance and player salary?
6. Which teams have the highest win rates at home versus away games?
7. Is there a significant relationship between a team's improvement or decline in the standings and the change in its average home attendance next season?

3. Database Design

3.1 Data Source:

This project integrates data from multiple external sources, combining API-retrieved information with web-scraped datasets, all of which are publicly available. First, we used the NBA API Python library (Link) (Patel, 2018/2025) to extract a wide range of league data, including active player lists, team metadata, player season game averages, player team history, and major awards such as MVP and All-Star selections. Player statistics were retrieved for every season from 2001 through 2025. Second, we collected historical NBA attendance data by web scraping ESPN's attendance pages (Link) (*2000-2026 NBA Attendance - National Basketball Association - ESPN*, n.d.)for each season from 2000-01 through 2024-25. Data from each season were extracted, standardized, and compiled into a consolidated CSV file. Third, we scraped ESPN's annual NBA salary data (Link) (*NBA Player Salaries - National Basketball Association - ESPN*, n.d.) for the same multi-year timeframe, compiling the results into csv file. In addition to these sources, we also incorporated season standings data using ESPN's public standings API (Link) (Wilson, 2023/2025), which provided team performance metrics across the same multi-year timeframe, compiling the results into a CSV file.

Code:

https://drive.google.com/drive/folders/1paenUyuA7mjzAnzxV3tTKWBp30NBZcju?usp=sharing

Data:

https://drive.google.com/drive/folders/1GpiFDMdTbDLD6FkreLycXLV4MgBkOBjS?usp=sharing

3.2 Data Assembly Process

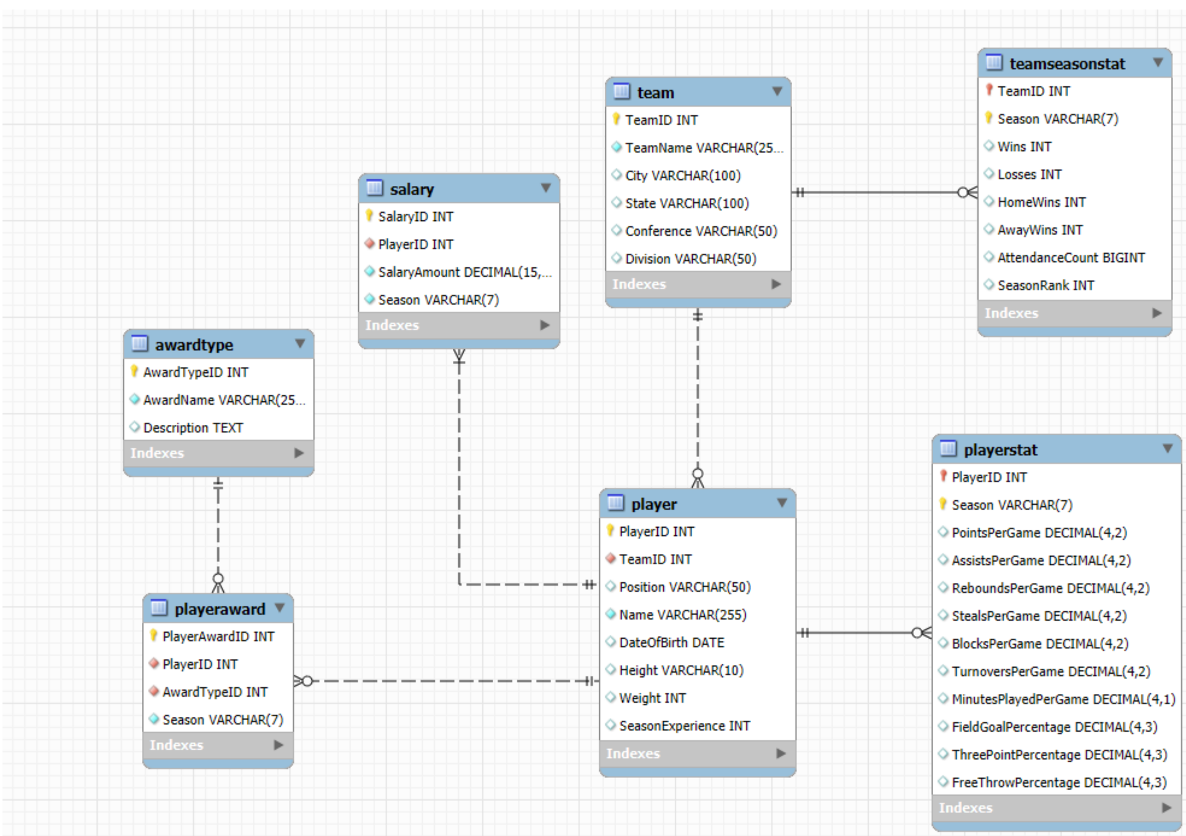
All data in the project was assembled using a Python Extract-Transform-Load (ETL) pipeline, which not only cleaned and standardized each dataset but also created the necessary links to connect the tables in the relational database properly. To connect the team data with all other team-dependent data, we generated a consistent Team Name-to-Team ID mapping after normalizing team names across all sources, including historical names. To connect the player data with its related data, we merged player demographics with the player history file to determine each player's current team, enabling accurate assignment of TeamID as a foreign key. We also created a PlayerName-to-PlayerID mapping so that external datasets such as salaries and awards could correctly reference PlayerID even when names varied across sources. For the award type and player award data, we extracted unique award names, assigned each an AwardTypeID, and then merged these IDs back into the player awards data so every award entry contained both a PlayerID and an AwardTypeID. Player statistics were connected to players by using the PlayerName-to-PlayerID mapping and standardizing seasons so that each statistics row matched the appropriate player and time period. Salary data was linked by mapping scraped ESPN player names to PlayerID using the PlayerName-to-PlayerID mapping and ensuring all season formats aligned with the rest of the database. Finally, the team season data was connected by merging standings and attendance data, normalizing team names, and converting those names to TeamIDs using the previously created mapping. Once the datasets were prepared and properly linked, the ETL pipeline also generated the corresponding SQL INSERT statements and executed them to populate and update all database tables. By performing these mappings, mergers, and automated insert operations, each table was connected through proper foreign key relationships, ensuring that players link to teams, awards link to players and award types, salaries and statistics link to players, and seasonal records link to each team accurately.

Code:

https://drive.google.com/drive/folders/1D7bs-vIWbjBHW_3ZMtk7E-FCLjqYTIxJ?usp=sharing

3.3 Logical Model

Figure 1. Logical Model (ER Diagram) Generated from MySQL Workbench



As shown in figure 1, the diagram represents the structure of the NBA database. At the center of the schema is the player table, which holds one row per NBA player and includes mandatory fields such as PlayerID, Name, and TeamID. The relationship between team and player is modeled as mandatory on the player side but optional on the team side: every player must belong to exactly one team, as we store each player's current active team, but a team can exist in the database even if it currently has no associated players (for example, newly added teams without players). The playerstat table connects to the player table with a mandatory foreign key on the playerstat side, meaning every statistics record must reference a valid PlayerID. However, the relationship is optional from the player's perspective because a player may not have season statistics for certain years for example, if they were injured, inactive, or did not play enough games so some players will have zero or multiple associated stat entries depending on their career activity.

The playeraward table connects players to the awards they have received by storing both a PlayerID and an AwardTypeID for each award entry, and within this table both fields are required because an award record must reference a valid player and a valid award type. However, the relationship is optional from the player's side, since many players will never receive an award, and optional from the award type side, as award categories may exist without having any recipients yet.

Financial information is handled through the salary table, which requires both a PlayerID and a Season for every salary record, ensuring all salary data ties back to a valid player. Because the database focuses on active players, each player is expected to have at least one salary entry for the seasons they appear in, making the relationship mandatory from both the salary side and the player side for all active-season records.

On the team side, the teamseasonstat table stores team performance metrics per season. Each stat record must refer to a valid TeamID, while teams themselves do not need to have season statistics for every year, allowing the model to include new teams.

3.4 Database Access Instructions

Server Information:

- Host: 34.66.156.27
- Port: 3306
- Database: NBAdatabase

Credentials

- Username: readonly_user
- Password: [aB$PROFESSOR!$12]

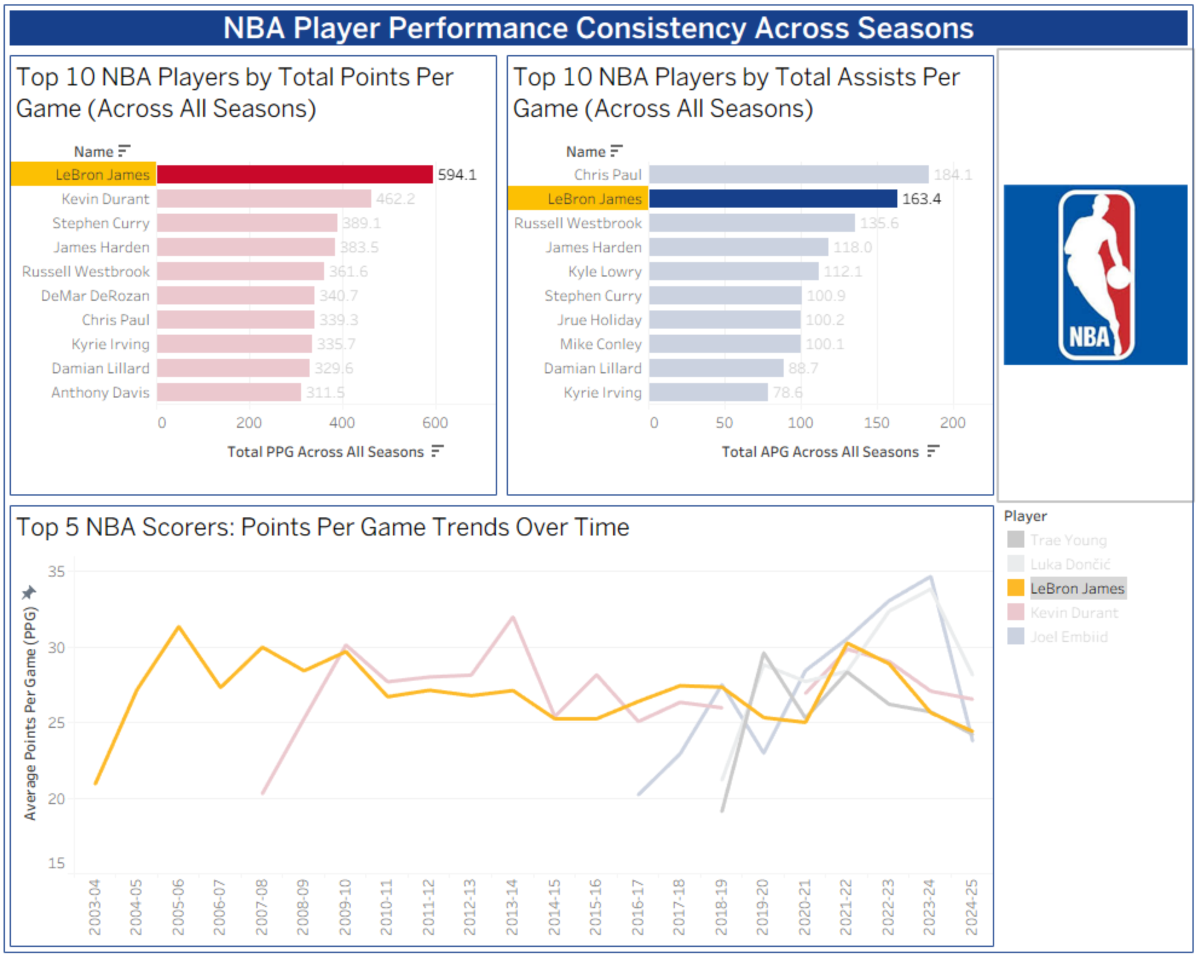
3.5 **Advanced Features in Data Sourcing and Assembly**

Our project incorporates several advanced features that go beyond the material covered in lectures, in the areas of data sourcing and data assembly. First, we developed a fully automated Python-based ETL pipeline that replaces manual data loading by reading all raw CSV files, cleaning and transforming each dataset, dynamically generating SQL INSERT statements. In addition to this, we integrated data sources through both the NBA API Python library and ESPN's public API, which allowed us to retrieve current player lists, player season averages, and award information. The project also employs web scraping using Python library BeautifulSoup and Requests (*Beautifulsoup4*, n.d.; *Requests*, n.d.), enabling us to collect historical attendance figures and player salary information directly from ESPN's public pages. The knowledge and skills required to build these Python programs were developed by two of our team members through their undergraduate studies in computing science as well as their internship experience working as software developers and data systems specialists.

4. Data Visualization

The data visualization section of this project uses Tableau to transform our integrated NBA database into clear and meaningful analytical insights. We present key visuals along with screenshots and brief explanations of how they were created, including the use of calculated fields, filters, and other Tableau features that support deeper analysis. Each visualization is directly linked to the business questions raised earlier, showing how graphical exploration helps reveal patterns in player and team performance across seasons.

Figure 2. Which players have shown the most consistent performance across multiple seasons?



This dashboard highlights player consistency by comparing points and assists across seasons. The visualization shows that LeBron James is the most consistent performer, ranking first in total points per game and second in total assists per game across all seasons. His strong standing is also explained by his long tenure in the league, as he has maintained a high level of performance since entering the NBA in 2003. The line chart further illustrates examples of seasonal performance, such as Joel Embiid averaging 34.69 points per game in the 2023–24 season. LeBron is followed closely by players like Kevin Durant, Stephen Curry, James Harden, and Russell Westbrook, who complete the top five scorers in the trends over time. This dashboard was created using dimensions such as player name and season, along with measures like points per game and assists per game to compare performance across multiple seasons. To focus on the most relevant players, we applied a Top 10 filter by selecting Name and choosing “Top 10 by SUM(Points Per Game)”. We also used dimension and measure filters to isolate specific players and seasons for clearer comparisons. For visual design, we applied a custom light “Hardwood Court” color palette to give the charts a clean and basketball-themed appearance. These elements required exploring Tableau’s filtering and formatting options, helping us create a more polished and focused visualization.

Figure 3. How do rookie players compare to league veterans in key performance metrics?

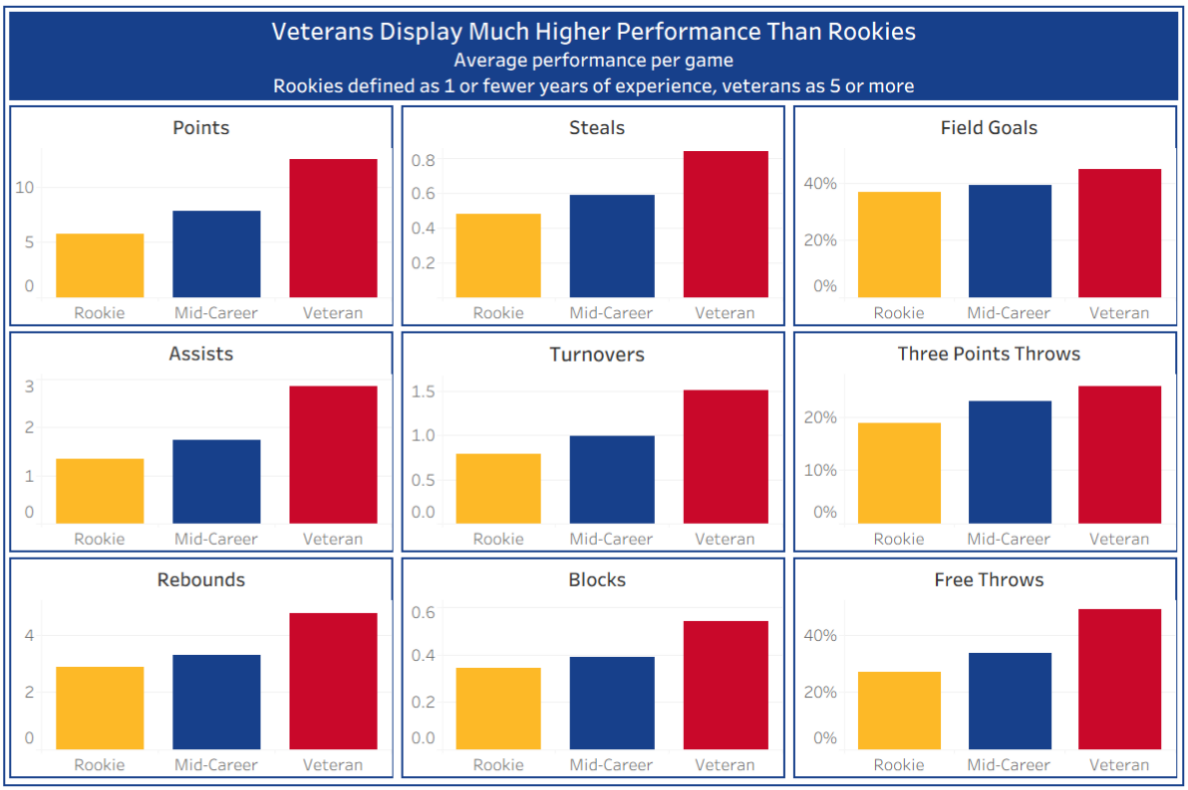


Figure 3 shows many of the performance statistics used in basketball to compare player performance, with players categorized based on their total years or seasons of experience in the NBA. Veterans (players with 5 or more seasons of experience) universally have much higher scores across all performance metrics, up to 2x in some measures when compared to new players.

Players were categorized using a calculated field that returns “Rookie” if a player has 1 or fewer years of experience at the time of data collection, "Veteran" if a player has 5 or more, or "Mid-Career" for players with 2-4 years. Metrics were then averaged across players. Similar proportions of differences between rookies and veterans were found when data was stratified by every player's set of individual career years instead of comparing the average performance of each player across their career. For figures 3 and 4, the *player* table was connected to *playeraward* and *playerstat* by matching their *playerid* columns to connect players to their performance statistics and list of awards received.

Figure 4. Can we identify trends in player performance that align with All-Star selections or MVP awards?

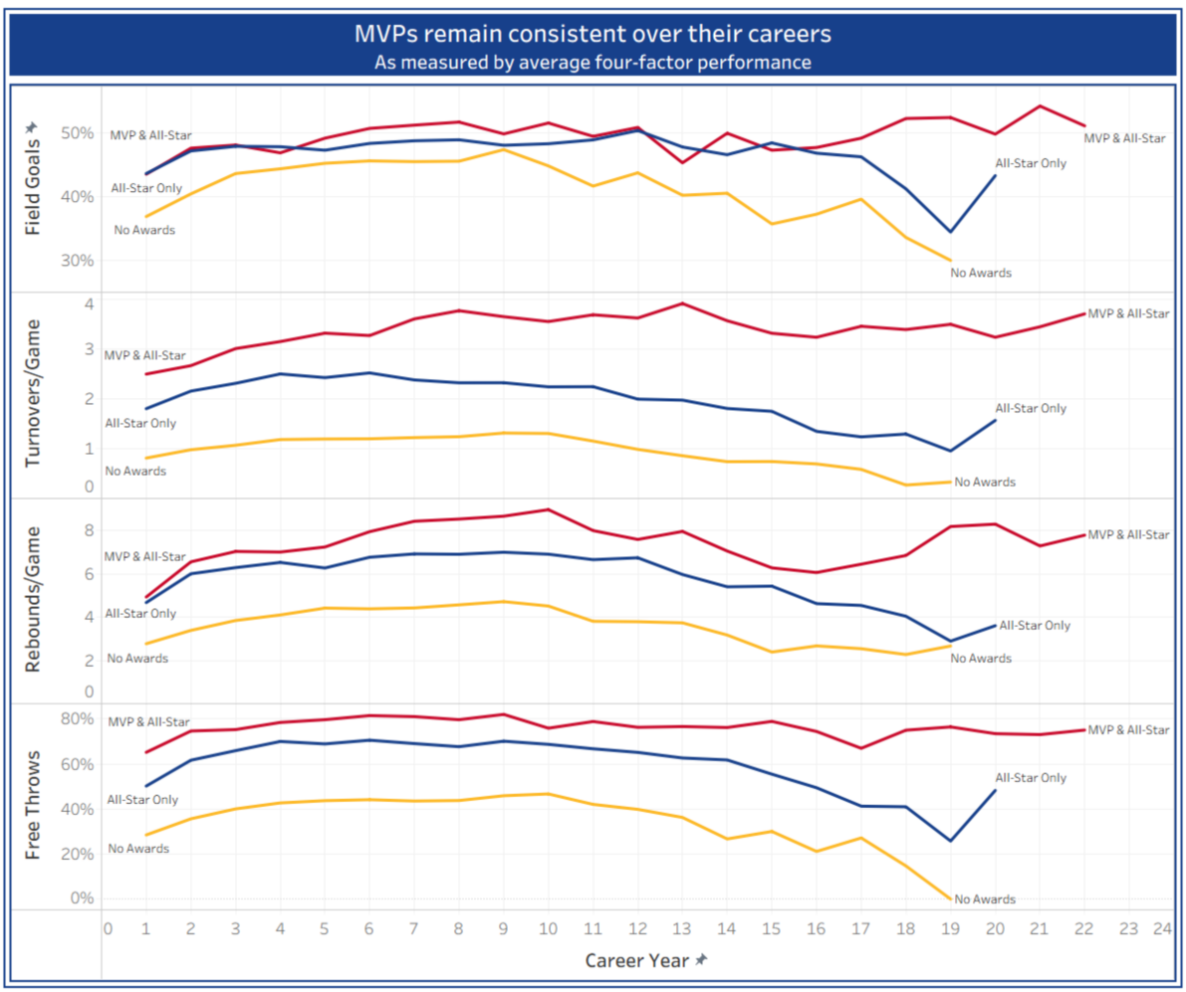
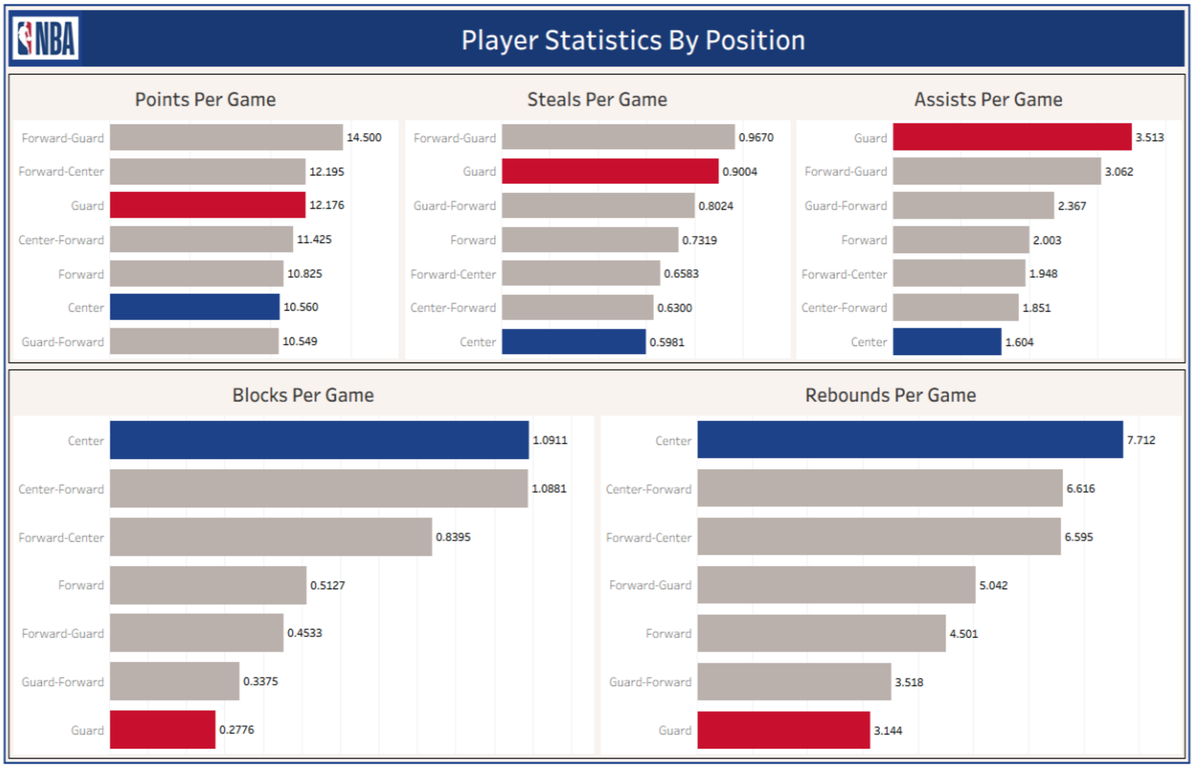


Figure 4 shows average key game performance measures throughout player careers, averaged across players with no awards or who have received MVP or All-Star awards at any point. Players who have received an All-star or MVP award display significantly better performance compared to players who never received an award, and this difference is notable for most metrics from their first season in the NBA. This suggests that the highest performing players can be identified early on in their career. While players who have received no awards or an All-Star award eventually see a decline in their performance, MVP's performance remains consistent. Turnovers and free throw percentages are especially important metrics for distinguishing early on between top picks and average players. The impact of specific performance metrics on a player's likelihood of being a top performer is critical for deciding the length of player contracts.

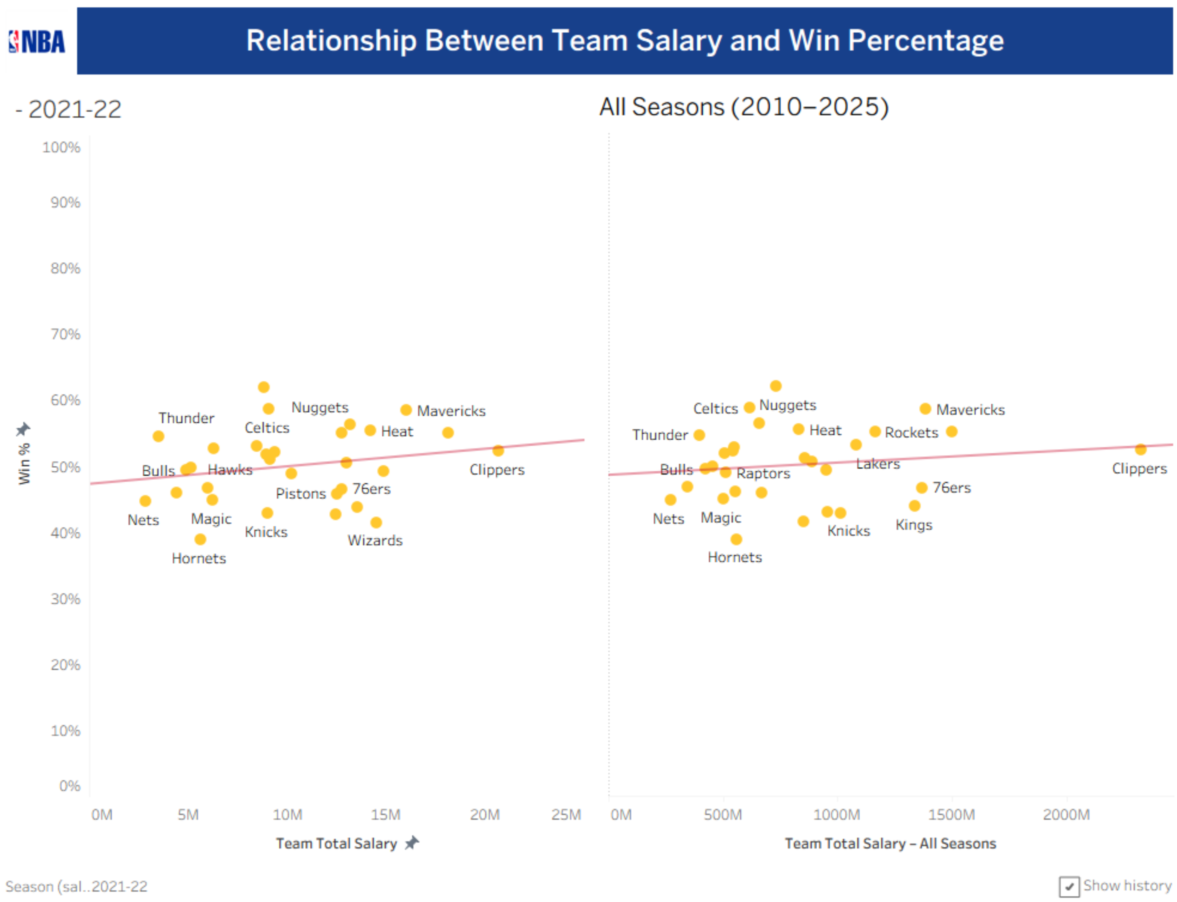
Players were grouped into lifetime award categories through a calculated field checking if a player had ever received a specific award. Career Year was created through a calculated field subtracting a player's current season from the *playerstat* table from the year of their rookie season. For display in Tableau, the *playeraward* and *awardtype* tables were joined across the *awardtypeid* to join a player's award information to the award names and filter by award name.

Figure 5. Are certain positions associated with higher scoring or assists?



This graph highlights clear differences in both offensive and defensive statistics across NBA positions. On the offensive side, the graphs show that guards average about 12.2 points per game and around 3.5 assists per game whereas centers are closer to 10.6 and 1.6. On the defensive side, guards contribute more to the perimeter. Their role is typically chasing opposing guards, while being quick and agile. They contribute more steals than centers at 0.90 per game as compared to centers who average 0.60. Centers however dominate interior defensive statistics such as rebounding. Centers tend to be taller, and therefore, they lead all positions with 7.7 rebounds, and 1.09 blocks per game, while guards sit much lower at only 3.1 rebounds and 0.28 blocks per game. Overall, this graph illustrates that guards are more impactful in offensive creation and perimeter defense, whereas centers provide most of the rim protection and rebounding.

Figure 6. What is the relationship between team performance and player salary?

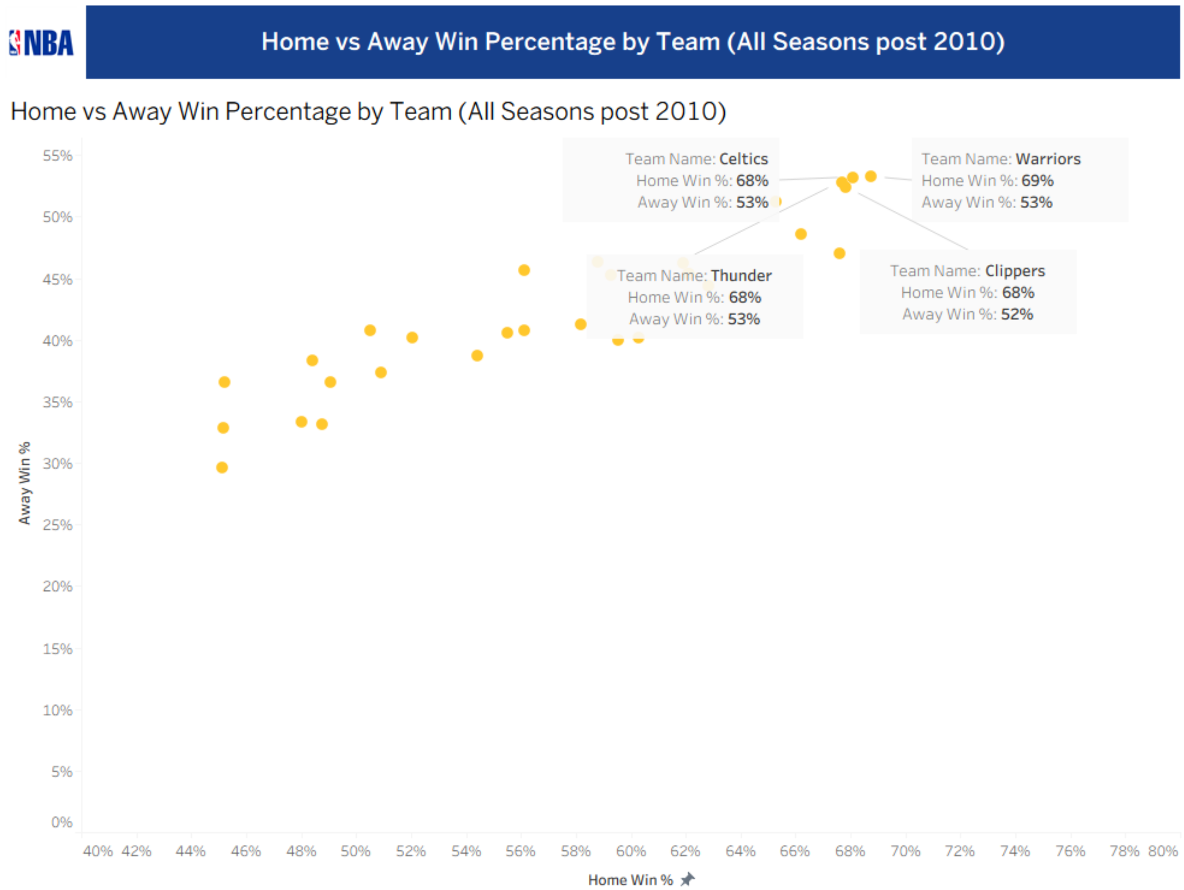


To assess how payroll relates to on-court performance, we plotted team win percentage against total team salary. The left panel illustrates the 2021–22 season, whereas the right panel aggregates all seasons from 2010 to 2025. In both views, the fitted trend line slopes upward, indicating a positive relationship between spending and performance: teams that invest more in player salaries tend, on average, to win a higher share of their games. In 2021-22, for example, higher-payroll teams such as the Mavericks and Nuggets sit above the regression line with win rates in the mid-50s to 60%, whereas lower-spending teams like the Hornets and Magic cluster below 45% wins.

The all-seasons view shows a similar but flatter pattern. Teams such as the Celtics, Nuggets, Heat and Mavericks combine above-average salaries with consistently strong win percentages. In contrast, others, including the Hornets and Nets, exhibit both lower spending and weaker results. The Clippers appear at the extreme right of the all-seasons plot, with the highest total salary but only modestly higher than the league average in win percentage. Overall, the analysis suggests that higher payrolls are associated with better performance. Still, the relatively shallow slope of the trend line indicates that salary is only one of several factors driving team success.

To build this chart in Tableau, we created a simple Win% measure for each team season (wins divided by total games) and plotted it against total team salary, with Season on the Pages shelf to create the year-by-year animation. For the all-seasons panel, we added a small Level of Detail calculation that fixes the math at the team level: Tableau first sums each team's wins and games across 2010 to 2025, then computes an overall win percentage.

Figure 7. Which teams have the highest win rates at home vs. away games?

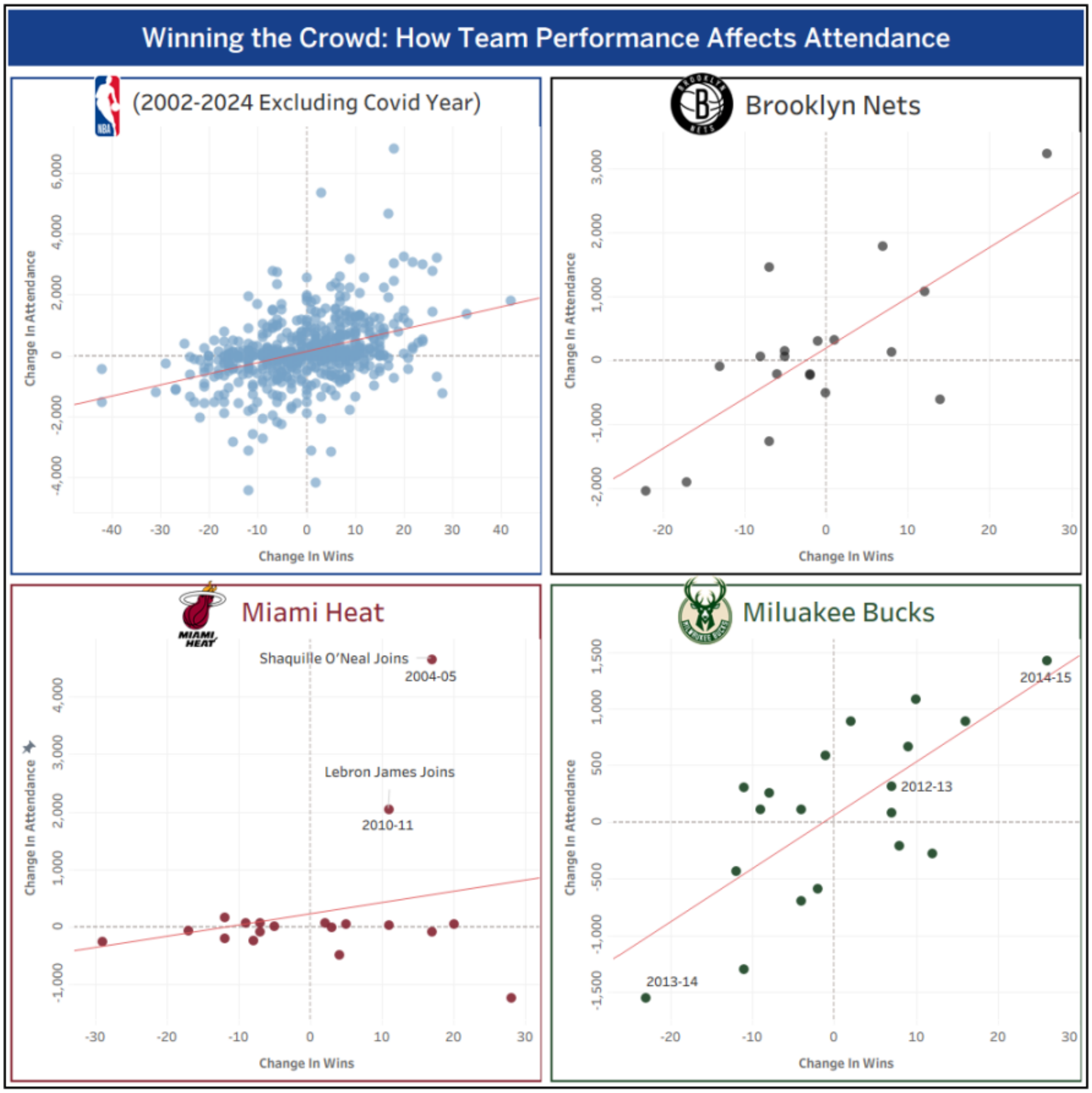


Using all seasons in our database from 2010 onward, we calculated each team's overall home and away win percentages and plotted them on a scatterplot (Home Win % on the x-axis, Away Win % on the y-axis). The teams in the top-right corner of this chart perform well in both environments. In our data, the Warriors, Celtics, Thunder, and Clippers clearly stand out: each win roughly 68-69% of its home games and 52-53% of its away games over the period.

Most other teams cluster around the mid-50% range at home and the high-30s to low-40s away, indicating they rely more heavily on home-court advantage. Taken together, this suggests that the Warriors, Celtics, Thunder, and Clippers have been the most consistently strong teams both at home and on the road in the post-2010 era.

To build this view in Tableau, we pulled in HomeWins, AwayWins, Wins, and Losses for each team-season and filtered the data to seasons after 2010. We then created two calculated fields: Home Win %, which divides a team's total home wins by roughly half its total games, and Away Win %, which does the same for away wins. Plotting Home Win % on the x-axis and Away Win % on the y-axis, with one mark per team and a few annotations for the Warriors, Celtics, Thunder, and Clippers, gives us this summary of how consistently teams perform at home versus on the road.

Figure 8. Is there a significant relationship between an NBA team's improvement or decline in the standings and the change in its average home attendance the next season?



This chart illustrates how changes in team performance through wins translate into changes in fan attendance from one season to the next. In the top-left panel, the league wide scatter shows a clear positive relationship. Teams who perform better compared to the previous season also see higher average home attendance while losing causes the opposite. Team specific panels show that the magnitude of this effect is not uniform across all teams. For the Brooklyn Nets and Milwaukee

Bucks, the points cluster tightly around a relatively steep upward trendline. More winning seasons strongly reward the team with larger crowds, and poor seasons do the opposite. The Miami heat panel displays a team that has lower variation in attendance across a wide range of win changes, showing they have a stable fanbase. However, as you would imagine, when superstars such as Shaquille O'Neal or LeBron James join the team, attendance sharply jumps. This insight was created by joining the *teamstat* table on itself, with the join keys being *TeamId* and a calculated field that matches one of the *teamstat* tables' season with the other *teamstat* tables' previous season so that a change in attendance and wins could be calculated. Then the *TeamId* was also joined with the Team table to bring in team names.

5. Summary

The visualizations developed in this project reveal several important patterns in NBA performance that carry meaningful implications for business decision-making. First, the consistency dashboard shows that a small group of elite players (Figure 2), led by LeBron James, Kevin Durant, Stephen Curry, James Harden, and Russell Westbrook, sustain high performance across many seasons. Identifying this level of consistency helps organizations assess long-term player value, contract stability, and investment risk. Second, (Figure 3) shows the comparison of rookies and veterans, demonstrating that experience significantly improves performance across all metrics, supporting decisions related to player retention and mentorship investments. Third, (Figure 4) digs into this rookie veteran gap by comparing core performance statistics across three experience groups: Rookies (one year or less), Mid-Career players (two to four years), and Veterans (five or more years). Across almost every measure, veterans outperform rookies by as much as twice the production, whether we look at career averages or individual seasons. The analysis by player position (Figure 5) highlights that certain roles consistently outperform others in specific skills, offering guidance for scouting, player development, and lineup optimization. Fifth, in Figure 6, the analysis of team salary and win percentage indicates that higher payrolls are only weakly associated with better performance, suggesting that financial efficiency and roster construction strategies may matter more than salary spending alone. Sixth, (Figure 7) comparing home and away win percentages reveals that top-performing franchises maintain strong records across both conditions, which is valuable for stadium operations, revenue forecasting, and planning promotional strategies. The analysis by player position highlights that certain roles consistently outperform others in specific skills, offering guidance for scouting, player development, and lineup optimization. Finally, (Figure 8) the attendance analysis shows that improvements in team performance are positively associated with increases in home attendance, emphasizing the financial importance of competitive success and its influence on ticket sales, fan engagement, and team branding. Together, these findings provide organizations with a data-driven foundation for decisions related to talent management, roster construction, salary planning, and fan engagement strategies.

Overall, a few key findings stand out from our work. At the player level, our results align with research such as the Academy's (Academy, 2018), but our analysis can expand on this: experience and quality really matter. The small core of MVPs, who stay near the top of the league for many years, and the apparent gap between rookies and veterans across almost every metric, both suggest that teams should treat long-run consistency and experience as significant assets when they decide who to keep, how long to sign them for, and how much to invest in developing younger players.

At the team level, we see that money helps, but it is not everything, which contradicts the typical assumption that money provides a better team. Higher payrolls are generally associated with better win percentages, but the relationship is only moderate, and some clubs clearly get more value out of their spending than others. Finally, the attendance analysis shows why these matters for the business side: when teams improve, many see noticeable gains in home attendance the following season, while a few markets remain more stable despite small performance swings. Together, these findings confirm that linking performance, pay, and fan response can provide NBA organizations with a more comprehensive, practical basis for planning contracts, building rosters, and investing in their fan base.

References

2000-2026 NBA Attendance—National Basketball Association—ESPN. (n.d.). Retrieved November 16, 2025, from https://www.espn.com/nba/attendance

Academy, U. S. S. (2018, August 2). NBA Players' Pay and Performance: What Counts? *The Sport Journal*. https://thesportjournal.org/article/nba-players-pay-and-performance-what-counts/

beautifulsoup4: Screen-scraping library (Version 4.14.3). (n.d.). [Python]. Retrieved December 3, 2025, from https://www.crummy.com/software/BeautifulSoup/bs4/

National Basketball Association. (2025). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=National_Basketball_Association&oldid=1324790543

NBA Player Salaries—National Basketball Association—ESPN. (n.d.). ESPN.Com. Retrieved December 3, 2025, from https://www.espn.com/nba/salaries/_/year/

Patel, S. (2025). *Swar/nba_api* [Python]. https://github.com/swar/nba_api (Original work published 2018)

requests: Python HTTP for Humans. (Version 2.32.5). (n.d.). [Python; OS Independent]. Retrieved December 3, 2025, from https://requests.readthedocs.io

Wilson, J. (2025). *Pseudo-r/Public-ESPN-API* [Python]. https://github.com/pseudo-r/Public-ESPN-API (Original work published 2023)