

**NBA Data Visualization**

*Milestone 1*

Author: Jake Swartzmiller

Instructor: **Robert Atkinson**

Date of Submission: April 24, 2025

**Table of Contents:**

[Dataset Description](#_bbm8kj16elq8) 3

[Exploratory Focus and User Persona](#_h5lw7u239gt8) 4

[Design Rationale](#_4226xh323d6) 5

## 

**GitHub Repository:**

[**https://github.com/JSwartzmiller/Data\_Visualization\_NBA**](https://github.com/JSwartzmiller/Data_Visualization_NBA)

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## **Dataset Description:**

The dataset that was chosen for this visualization project was a dataset containing detailed statistics of NBA players that have played significant minutes throughout the season. The datasets were retrieved from sources like ESPN and Basketball Reference which are two renowned sources known for their complete coverage of sports statistics. The dataset contains advanced statistics for every single relevant player for each NBA team along with each team's statistics. These metrics are a crucial part in identifying players' performances, identifying trends, and understanding how individual players affect their team’s performance.

Prior to analysis, cleaning and preprocessing steps were performed to ensure that the data’s accuracy along with the data’s usability overall. The first step was to iterate through each team and scrape each player’s game logs for the entire season and store them as a dataframe. Once each team's statistics were retrieved, the entire league's data was combined into one data frame for easy data manipulation. There were many missing values so removing the rows with missing values was a must. When scraping the data, many feature labels stayed, making the full data frame contain more than one line of feature labels throughout the data frame so I had to figure out a way to remove those so the data frame only contained one. Also, a critical step was to make sure that the team names and cities matched throughout the dataset for easy manipulation steps later on.

I chose to work on this dataset because of my interest in sports along with the sheer amount of categorical and continuous variables that sports data offers. This data offers many exploratory angles along with a way to explore trends that might be changing based on the matchups. By selecting this dataset, I hope for users to uncover trends throughout the season.

***Data Resource References:***

* ESPN: <https://www.espn.com/nba/>
* Basketball Reference: <https://www.basketball-reference.com/leagues/NBA_2025.html>

## 

## 

## **Exploratory Focus and User Persona:**

The primary users of the project are sports analysts and sports enthusiasts. These users are people interested in detailed NBA player and team statistics. These are the individuals who rely on data-driven insights to make informative decisions, whether that is because they are trying to gain insight on their favorite team, looking to make moves in a sports fantasy league that they are looking to manage, or make more informative bets using data to back their insights.

Some of the tasks a user might be looking to complete include, discovering who the highest performer in rebounding on a specific team, how has a specific player's points performance been changing throughout the season, or how has the team match up against another future matchup.

**User Persona:**

***Name:*** John Doe

***Role:*** Fantasy Basketball Manager

***Background:*** John is a passionate basketball fan who likes to dive into player statistics and performance metrics to make decisions. He competes in fantasy leagues with his friends where he competes to form the best basketball team based on performance statistics.

***Goals:*** John’s primary goal is to build a lineup for his fantasy team that consistently performs well throughout the season. He aims to make decisions based on data about where he drafts players in the league along with smart decisions throughout the entire season.

***Key Use Case***, when John would use the dashboard:

* Look into the future matchups of players on the current team to see the best game for each player's matchups to lock them into the lineup for that week.
* When offered trades, can look into involved players easily. This can show how the player has been performing throughout the season and how their performance metrics are changing currently.
* Identify trends that might be taking place in the league to try and identify emerging players that could be seen as undervalued by league mates.

## **Design Rationale:**

Describe your choices for the prototype, including your reasoning behind the interactivity features, data selection, and visualization types. Discuss how the design supports the exploratory goals of the project.

For this interactive visualization prototype, I chose to focus on NBA player game logs because they offer a strong combination of **categorical** variables (such as Team and Player) and **continuous** variables (such as Points, Rebounds, Assists, and Minutes Played), making the dataset ideal for **exploratory analysis**.

The visualizations starts by asking the user to select a team, followed by a player from that selected team, and after a statistic that they would like to explore. The layering in filtering supports a **simple exploration** process, helping users focus on players they are interested in without being overwhelmed by the amount of data shown.

To add more exploration features, I incorporated an additional dropdown to allow users to select how many games they want to visualize. This gives the users choice to zoom in on the sample size of data depending on their interests.

Interactivity features I included:

* **Dynamic Bar Chart:** Bar chart that updates based on the user’s selected player and stat they want to explore.
* **Custom Hover Tooltip:** Displays the exact value of the statistic for each game when the user hovers their mouse over a bar in the chart.
* **Baseline Line:** Users can define by inputting a threshold value. The line is dynamically added if the user enters the numbers. When the baseline is chosen, The line appears on the chart and each game is filtered over or under that line by color.

The design emphasizes clarity, customization and easy exploration. By allowing users to customize the statistics they want to see, the dashboard encourages deeper insights, such as performance trends over time or the consistency of a player.

The choice of a bar chart was intentional because it produces a clear, game-by-game comparison and makes the baseline easily shown by sorting the values by color. Overall, the design choices were made to optimize the users options while keeping the dashboard clean and the interactions smooth.