# Exploratory Data Analysis (EDA) Report: NBA Dataset

## 1. Dataset Overview

The dataset consists of 12,844 entries and 21 columns containing information about NBA players, including their age, height, weight, college, country, performance statistics, and season-wise data.

## 2. Data Cleaning

Steps taken for data cleaning:  
- Removed an unnecessary 'Unnamed' column.  
- Identified and filled missing values in the 'college' column with the mode.  
- Standardized categorical values (e.g., fixing typos in country and college names).  
- Checked for duplicate rows, none found.  
- Detected and handled outliers using median imputation and percentile capping.

## 3. Exploratory Data Analysis

### 3.1 Univariate Analysis

Key Findings:  
- 83.77% of players are from the USA, while 16.23% are from other countries.  
- The number of players per season has fluctuated, with certain years having more players.  
- The average player age is around 27 years, with heights averaging 200.6 cm.

### 3.2 Bivariate Analysis

Key Findings:  
- Correlation analysis shows that games played (GP) is positively related to points scored (PTS), rebounds (REB), and assists (AST).  
- Players with higher net ratings tend to have played more games.  
- The maximum points per game trend shows peaks in certain seasons, reflecting standout performances.

### 3.3 Multivariate Analysis

Key Findings:  
- Best scoring seasons were led by players like James Harden, Kobe Bryant, and Joel Embiid.  
- Pair plots and heatmaps confirmed significant relationships between player stats.  
- The dataset indicates that player impact is often a combination of scoring, rebounding, and assisting.

## 4. Findings & Conclusion

The NBA dataset provides rich insights into player performance over the years. Data analysis reveals key trends in scoring, efficiency, and player impact. Future analyses could explore advanced predictive modeling for player success.