**NBA Statistics Analysis Report**

**Introduction**

This report analyzes NBA player performance statistics during the 2023-2024 regular season. The dataset consists of over 400 players and 30 statistical metrics. The analysis aims to identify key trends in player performance across different timeframes.

**Dataset Overview**

The dataset comprises multiple CSV files, each representing player statistics at various points in the season. However, for consistency and relevance, only data from the end of each month has been selected:

* **NBA\_2024\_per\_game(27-12-2023).csv**
* **NBA\_2024\_per\_game(26-01-2024).csv**
* **NBA\_2024\_per\_game(03-01-2024).csv**

**Data Source & Validity**

The dataset was sourced from publicly available NBA statistics repositories such as Basketball Reference and official NBA data sources. The validity of the dataset was ensured through cross-checking with multiple sources and removing inconsistencies.

**Data Cleaning: Handling Multiple Records per Player**

During data preprocessing, it was observed that some players appear multiple times in a given month due to team changes. For example, **Precious Achiuwa** has multiple entries due to switching teams. To ensure accurate and non-redundant representation, only the row where Tm == TOT is retained, as it aggregates performance across all teams played during that period.

The following steps were taken to clean the dataset:

1. Identify players who switched teams within a month.
2. Extract the teams they played for before and after the switch.
3. Retain only the TOT row where available, removing individual team-specific records.
4. Assign Old Team and Current Team columns to maintain team transition information.
5. Concatenate cleaned monthly datasets into a final combined dataframe.
6. Handle missing values by filling them using median values where appropriate to maintain consistency.

**Transferred Players**

A total of **21 players** switched teams during the analyzed period. The following players were identified as transferred:

* Daniel Theis
* Filip Petrušev
* KJ Martin
* Nicolas Batum
* P.J. Tucker
* Robert Covington
* Bruce Brown
* Danilo Gallinari
* Dylan Windler
* Immanuel Quickley
* Jordan Nwora
* Kenneth Lofton Jr.
* Malachi Flynn
* Marvin Bagley III
* Mike Muscala
* OG Anunoby
* Pascal Siakam
* Precious Achiuwa
* RJ Barrett
* Skylar Mays
* Terry Rozier

**Data Dictionary**

| **Column** | **Description** |
| --- | --- |
| Rk | Rank |
| Player | Player's name |
| Pos | Position |
| Age | Player's age |
| Tm | Team |
| G | Games played |
| GS | Games started |
| MP | Minutes played per game |
| FG | Field goals per game |
| FGA | Field goal attempts per game |
| FG% | Field goal percentage (calculated as FG / FGA) |
| 3P | 3-point field goals per game |
| 3PA | 3-point field goal attempts per game |
| 3P% | 3-point field goal percentage (calculated as 3P / 3PA) |
| 2P | 2-point field goals per game |
| 2PA | 2-point field goal attempts per game |
| 2P% | 2-point field goal percentage (calculated as 2P / 2PA) |
| eFG% | Effective field goal percentage (calculated as (FG + 0.5 \* 3P) / FGA) |
| FT | Free throws per game |
| FTA | Free throw attempts per game |
| FT% | Free throw percentage (calculated as FT / FTA) |
| ORB | Offensive rebounds per game |
| DRB | Defensive rebounds per game |
| TRB | Total rebounds per game |
| AST | Assists per game |
| STL | Steals per game |
| BLK | Blocks per game |
| TOV | Turnovers per game |
| PF | Personal fouls per game |
| PTS | Points per game |

**Analysis Plan**

The analysis will be structured as follows:

1. **Data Preprocessing**
   * Handling missing values
   * Standardizing column names
   * Merging selected datasets
   * Recalculating shooting percentages (FG%, 3P%, 2P%, eFG%, FT%)
2. **Exploratory Data Analysis (EDA)**
   * Distribution of key statistics (PTS, AST, REB, FG%)
   * Performance comparison across months
   * Position-based statistical trends
   * Visualizations using histograms, boxplots, and trend lines
3. **Statistical Analysis**
   * Correlation analysis between metrics
   * Regression modeling to predict player performance
   * Evaluating the impact of mid-season transfers on player performance using t-tests or ANOVA
4. **Insights & Findings**
   * Identification of top-performing players
   * Trends in player efficiency and scoring
   * Team-based performance evaluation
   * Impact of mid-season transfers on player performance

**Key Findings & Insights**

* **Scoring Trends**: Analyzing monthly trends reveals fluctuations in scoring efficiency among top players.
* **Top Performers**: Identifying the most consistent players based on advanced metrics.
* **Team Impact**: Evaluating the effect of team changes on individual player performance through statistical significance tests.
* **Regression Insights**: Using linear regression, we analyze which factors contribute the most to player performance variations.

**Player Improvement Trend (Linear Regression)**

* **Brooklyn Nets, Indiana Pacers, and Phoenix Suns** show the most consistent scoring trends, while **New York Knicks** has the lowest average points.
* **Indiana Pacers lead in assists**, followed by **Miami Heat and Cleveland Cavaliers**, with **New York Knicks** ranking the lowest.
* **Boston Celtics, Brooklyn Nets, and Charlotte Hornets** dominate in rebounds, while **Washington Wizards** rank the lowest.

**ANOVA Test Results**

ANOVA tests indicate significant differences in team performance across months:

* **Total Points**: F=1300.87, p-value=1.53e-65 (Reject H0)
* **Total Assists**: F=640.44, p-value=8.93e-53 (Reject H0)
* **Total Rebounds**: F=1351.11, p-value=3.09e-66 (Reject H0)

These results confirm that performance metrics vary significantly across months, highlighting dynamic changes in team effectiveness and player contributions.