

NBA Player Role Analysis With K-Means Clustering

By: Marcus Walton



About Me

Description:

I completed this project because I grew up a huge sports fan, but more specifically a huge NBA fan. My passion for analytics stemmed from my interest in analytics and how it impacts the sports field



Project Introduction

Objective:

The objective of this project is to identify distinct player roles in basketball based on key performance metrics, such as points scored, assists, rebounds, and more. By categorizing players into roles, we can better understand their contributions to the game

Why This Analysis Is Important:

1. It aids in talent identification, helping teams find players that fit specific roles
2. It informs team-building strategies by ensuring a balanced roster
3. It supports scouting reports by providing a data-driven understanding of player tendencies



Table of contents

01

Dataset Overview

Examining 2017–2018 NBA season player data

02

Exploratory Analysis

Understand the data and gather some key findings

03


Methodology and Analysis

Walking through the method of my analysis and conducting it

04

Findings and Applications

Showcase my findings and offer future applications where it could be used





01

Dataset Overview



2017-2018 NBA Season Player Data

Source:

The dataset used for this project comes from comprehensive basketball statistics for the 2017-18 NBA regular season. It includes detailed game-by-game performance data for individual players

Dataset Size:

The dataset comprises over 26,000 player-game records, covering performances from hundreds of players across all NBA teams for the season analyzed. This rich dataset provides a robust foundation for clustering and analysis

- 26,109 rows
- 52 columns
- 540 players
- 30 teams

Data Reduction and Transformation

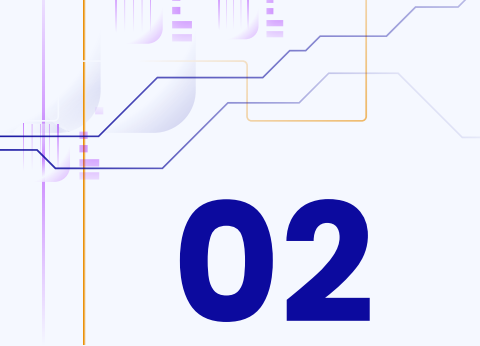
Columns Dropped:

To focus on performance metrics and reduce noise, I dropped columns that were:

- Irrelevant to player performance, such as referee names (offLNm1, offFNm1) and game time (gmTime)
- Redundant, like opponent-related information (oppAbbr, teamConf) since clustering focused on individual players
- Non-actionable for clustering, such as player birthdates (playBDate)


New Dataset Size:

- After preprocessing, the dataset was reduced to 30 columns, containing only key performance metrics such as points, assists, rebounds, shooting percentages, and turnovers



02

Exploratory Data Analysis



Filtering Further

Meaningful Sample Size:

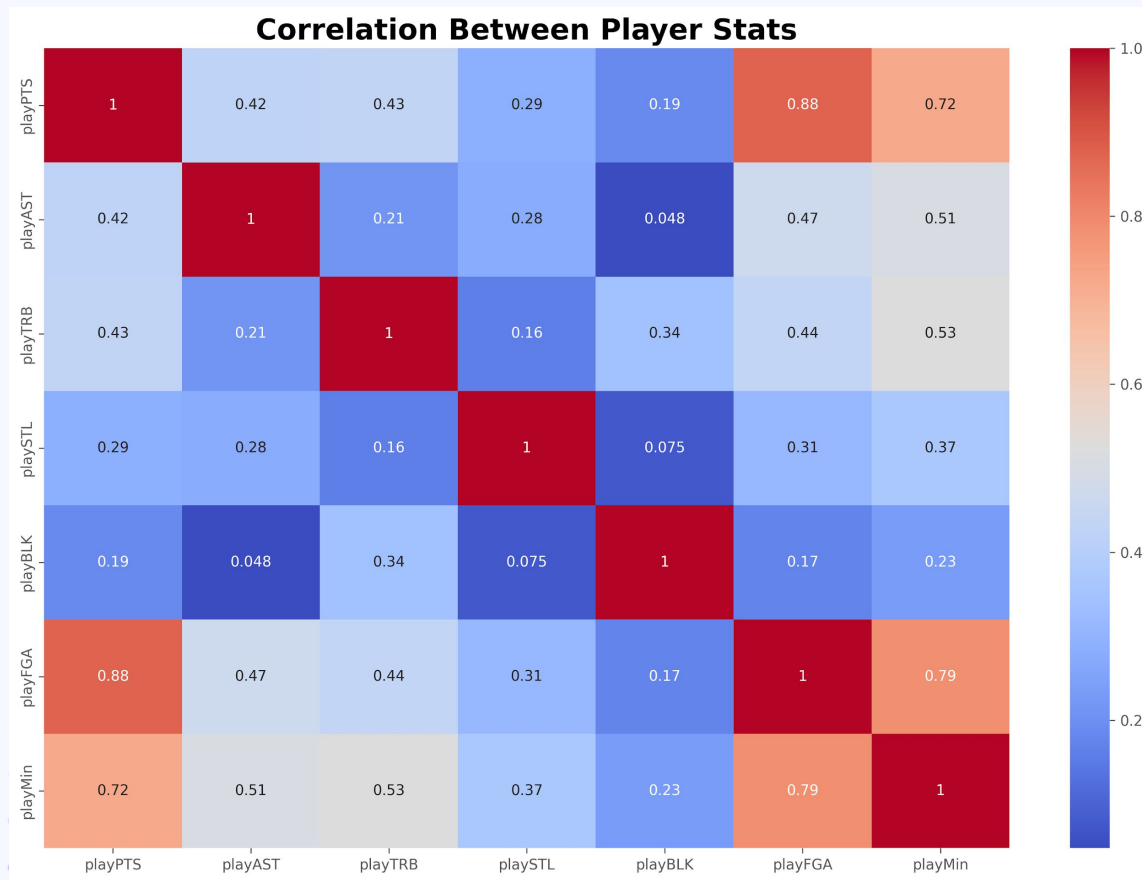
To be able to run a good analysis, it's important to have meaningful useful data that is free of outliers. To combat this I filtered the data even more to get only meaningful player stats that would contribute to the analysis

What Was Filtered:

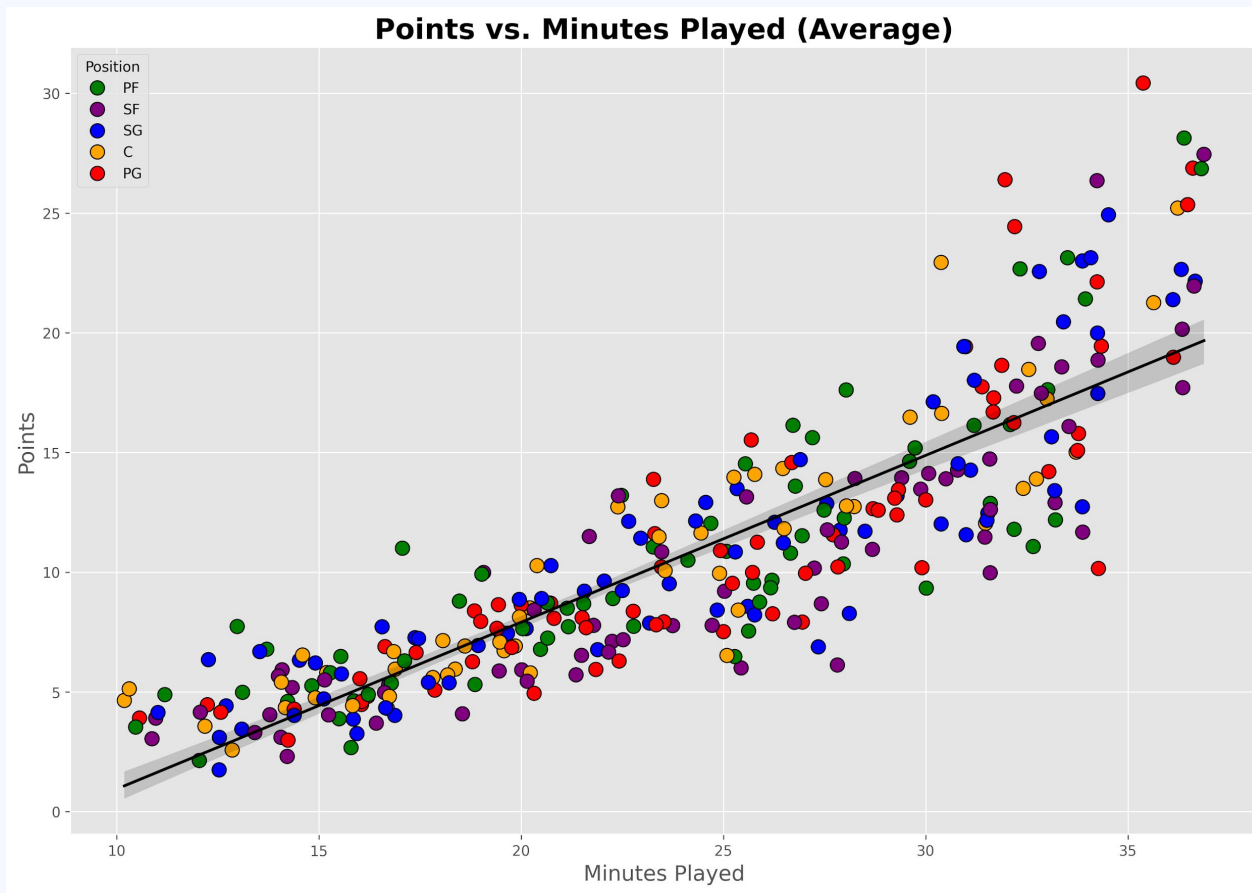
- Averages of all players stats
- Players with at least half of the seasons games played
- Played on average 10 mins or more a night

This left 326 players to run the analysis on

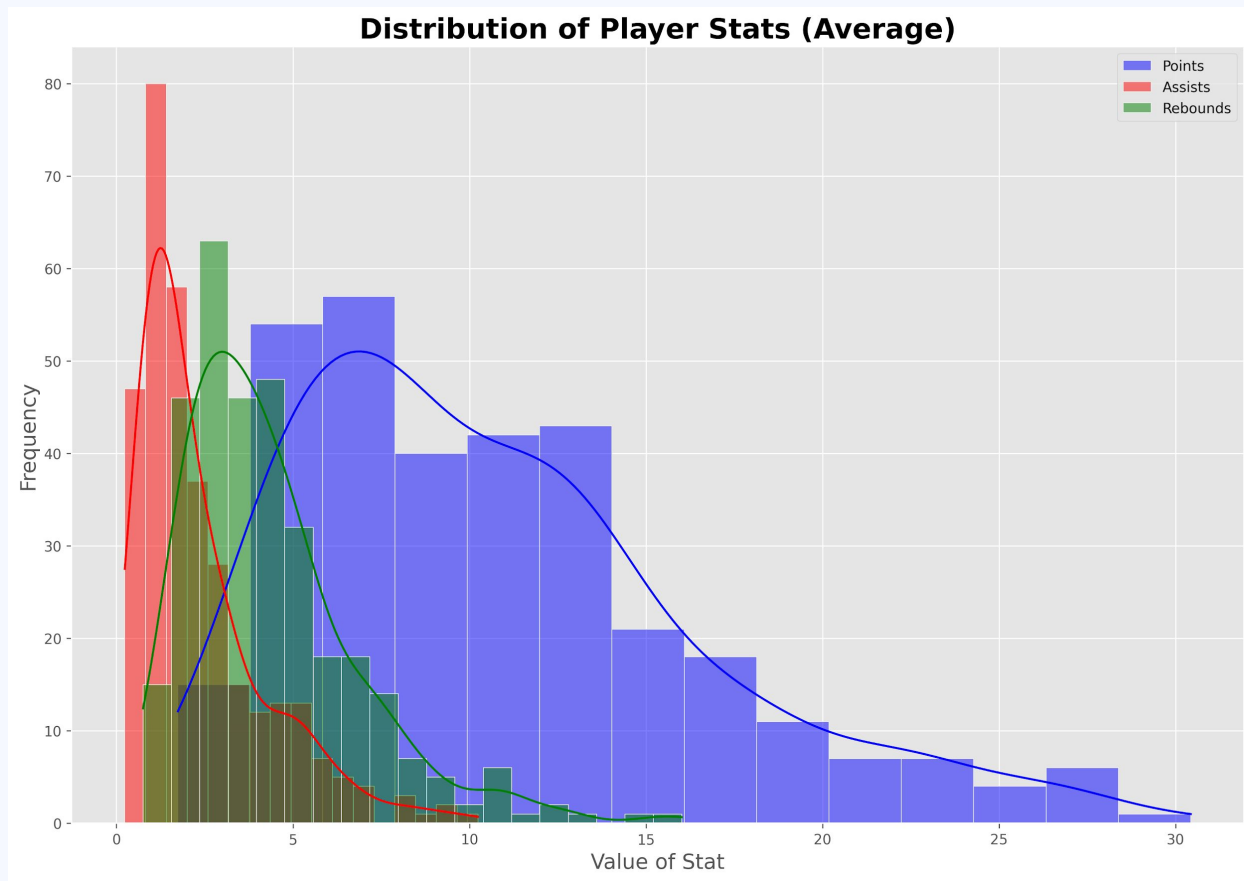
PTS, FGA, Min Only Correlations



More Min Played = More Points

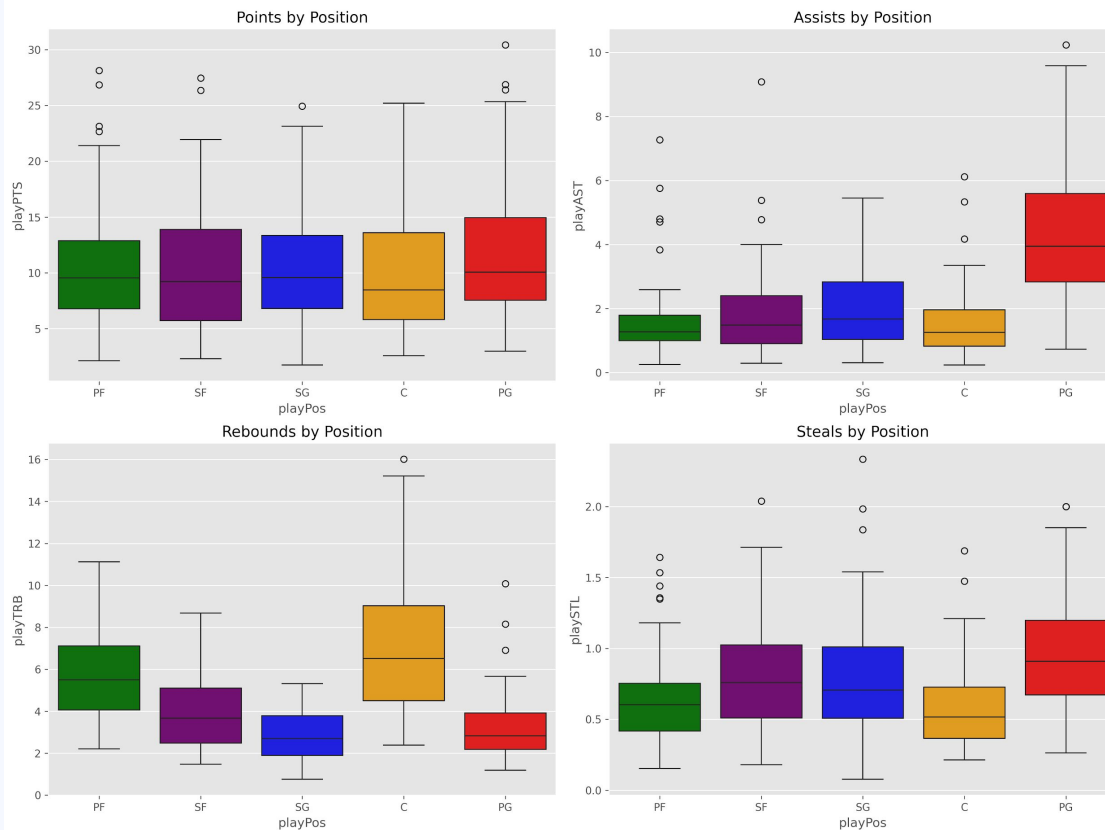


Distribution of Stats



Stats by Position

Player Stats by Position





03

Clustering By Player Role



Methodology - K-Means Clustering

What is K-Means:

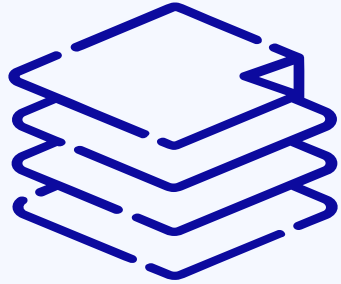
- K-Means is an unsupervised machine learning algorithm used for grouping data into clusters based on similarity
- It minimizes the variance within each cluster, ensuring that players with similar statistical profiles are grouped together
- This method was chosen because it allows for a data-driven approach to categorizing player roles without predefined labels

Features Used:

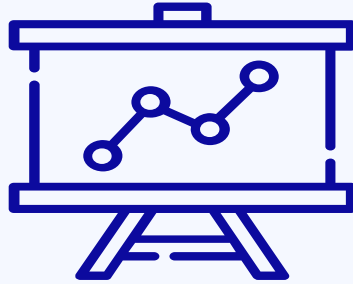
- playPTS, playAST, playTRB, playFGA, playSTL, and playBLK

The goal was to group players based on offensive impact while ensuring balance across roles

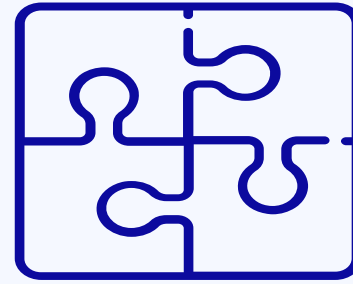
Clustering Process



Data Prep



Normalizing Data



of Clusters



Assign Roles



Player Role Assignment

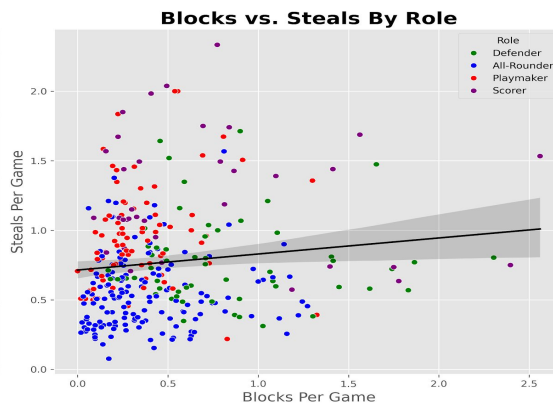
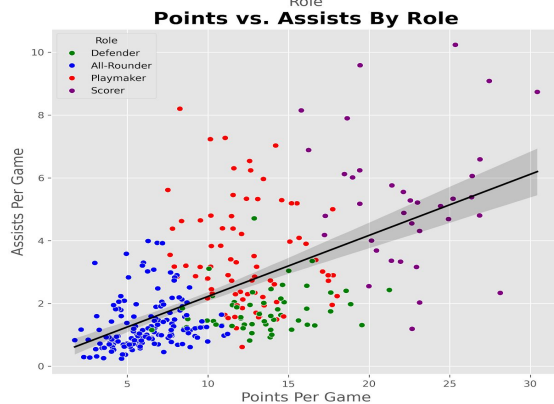
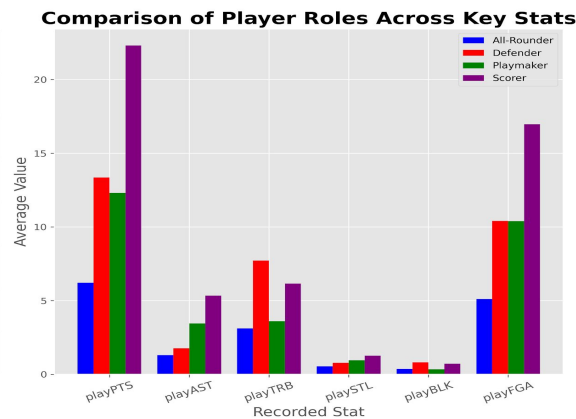
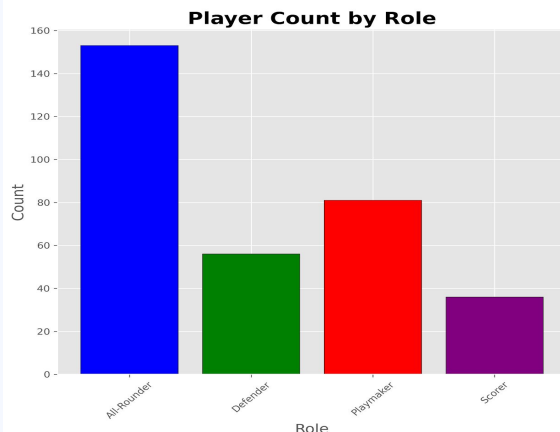
Player Roles:

- Scorer (High points, low assists)
- Playmaker (High assists, moderate points)
- Defender (High steals and blocks)
- All-Rounder (Balanced stats across all categories)



Player Role Dashboard

Basketball Player Role Dashboard



Dashboard Observations

Player Distribution by Role:

- The "Scorer" role has the lowest count, which might indicate that specialized scorers are rarer or less prioritized

Comparison of Player Roles Across Key Stats:

- "Scorers" excel in scoring (playPTS) but have relatively low values in other stats
- "Playmakers" dominate in assists (playAST), as expected, but are not high in other areas
- "Defenders" have high values in defensive metrics like steals (playSTL) and blocks (playBLK)
- "All-Rounders" have a balanced contribution across all metrics but do not lead in any single area
- This plot confirms that roles align well with their statistical strengths

Dashboard Observations

Points vs. Assists by Role:

- "Playmakers" cluster higher in assists with moderate points, highlighting their primary focus on team contribution
- "Scorers" are more spread along the points axis with lower assists, showcasing their specialization in scoring
- "All-Rounders" and "Defenders" are distributed in the lower to middle ranges of both points and assists

Blocks vs. Steals by Role:

- "Defenders" dominate the upper ranges of both blocks and steals, as expected
- "All-Rounders" maintain balanced values but do not excel in these metrics
- "Scorers" and "Playmakers" generally stay in the lower ranges for defensive stats, aligning with their offensive focus

04 Findings And Applications

Example With Boston Celtics



Example With Boston Celtics

Role Distribution:

- **Defensive Strength:** With 5 All-Rounders and 5 Defenders, the team has a strong focus on defense and versatility
- **Limited Offense:** Only 1 Scorer and 2 Playmakers suggest potential challenges in generating consistent scoring and creating offensive plays
- **Team Reliance:** The team likely relies on collective effort rather than individual star power for success

Role	Count
All-Rounder	5
Defender	5
Playmaker	2
Scorer	1

Strategy From Analysis

Roster Suggestions:

- **Add a Primary Scorer:** With only one Scorer, the team may struggle with consistent offensive output. Recruiting a high-efficiency scorer can boost their ability to close games and score under pressure
- **Recruit a Playmaker:** Increasing the number of skilled playmakers will improve ball distribution and create more opportunities for teammates to score
- **Optimize Current Roles:** Develop the offensive skills of current All-Rounders or Defenders to take on secondary scoring or playmaking duties
- **Target Specialists in Free Agency/Draft:** Look for players who excel in scoring or playmaking in specific positions to fill the gaps effectively

Future Applications

Matchup Comparisons:

- Analyze role distributions across teams to identify mismatches and exploit advantages in key areas like scoring, defense, or playmaking

Team Building:

- Guide recruitment and drafting strategies to fill role gaps and create well-rounded rosters

Player Development:

- Identify players who could excel in secondary roles and design targeted training programs

Future Applications

Game Preparation:

- Use role data to develop tailored game plans against specific opponents

League-Wide Insights:

- Compare role distributions across teams to identify league-wide trends and competitive benchmarks



Thanks !

Do you have any questions?

mawalton@marshall.usc.edu
marcuswaltonmsba.com

