NBA Player Performance Analysis

Author: Salman Imtiaz

1. Introduction

This report presents a comprehensive analysis of NBA player performance using the all_seasons dataset. The analysis is divided into three main sections:

Descriptive Analysis: Provides summary statistics and insights into player attributes and performance.

Diagnostic Analysis: Explores relationships between player performance metrics and physical attributes.

Predictive Analysis: Develops a predictive model for estimating points per game using linear regression.

2. Descriptive Analysis

Objective

To summarize the key metrics of NBA players, including average age, height, weight, points, rebounds, and assists. Additionally, this section provides insights into player distribution by team and country.

Descriptive Analysis

```r
# Load necessary libraries
library(ggplot2)
library(dplyr)

# Load the data
data <- read.csv("E:/Baylor Material/STA 5V85 - Capstone/Analysis 1/all_seasons.csv")

# Basic summary of the dataset
summary(data)

# Average age, height, and weight
cat("Average Age: ", mean(data$age), "\n")
cat("Average Height: ", mean(data$player_height), "\n")
cat("Average Weight: ", mean(data$player_weight), "\n")

# Distribution of players by team and country
cat("Top 5 Teams by Player Count:\n")
print(sort(table(data$team_abbreviation), decreasing = TRUE)[1:5])

cat("Top 5 Countries by Player Count:\n")
print(sort(table(data$country), decreasing = TRUE)[1:5])

# Average points, rebounds, and assists per game
cat("Average Points Per Game: ", mean(data$pts), "\n")
cat("Average Rebounds Per Game: ", mean(data$reb), "\n")
cat("Average Assists Per Game: ", mean(data$ast), "\n")
```

Summary

- **Average Age:** 27.0 years
- **Average Height:** 200.6 cm
- **Average Weight:** 100.3 kg
- **Top 5 Teams by Player Count:** CLE, TOR, MIA, DAL, WAS
- **Top 5 Countries by Player Count:** USA, Canada, France, Australia, Spain
- **Average Points Per Game:** 8.2
- **Average Rebounds Per Game:** 3.6
- **Average Assists Per Game:** 1.8

Interpretation

This descriptive analysis provides a basic understanding of the player demographics and performance in the NBA. The average player is 27 years old, with a height of approximately 200 cm and a weight of 100 kg. The USA dominates the player distribution, which is expected given the NBA's origins and popularity in the country.

---

3. Diagnostic Analysis

Objective

To explore the relationships between player performance metrics (points, rebounds, assists) and physical attributes such as age and height.

Diagnostic Analysis

```
# Assuming data is already loaded and 'ggplot2' is available

# Points, Rebounds, and Assists vs. Age
ggplot(data, aes(x=age, y=pts)) + geom_point(alpha=0.3) +
geom_smooth(method="lm", color="red") + ggtitle("Points vs. Age")
ggplot(data, aes(x=age, y=reb)) + geom_point(alpha=0.3) +
geom_smooth(method="lm", color="green") + ggtitle("Rebounds vs. Age")
ggplot(data, aes(x=age, y=ast)) + geom_point(alpha=0.3) +
geom_smooth(method="lm", color="blue") + ggtitle("Assists vs. Age")

# Rebounds vs. Player Height
ggplot(data, aes(x=player_height, y=reb)) + geom_point(alpha=0.3) +
geom_smooth(method="lm", color="orange") + ggtitle("Rebounds vs. Player
Height")
```

Summary

- **Points vs. Age:** The relationship shows that points per game do not significantly change with age.
- **Rebounds vs. Age:** There is a slight positive correlation, indicating rebounds may increase with age to some extent.
- **Assists vs. Age:** A moderate positive correlation is observed, suggesting that assist rates may increase as players age.
- **Rebounds vs. Height:** A positive correlation exists between player height and rebounds, indicating that taller players tend to secure more rebounds.

Interpretation

The diagnostic analysis reveals that while age shows varied relationships with different performance metrics, player height strongly correlates with rebounds. This suggests that certain physical attributes like height play a critical role in a player's ability to perform specific tasks, such as securing rebounds.

---

4. Predictive Analysis

Objective

To develop a linear regression model to predict a player's points per game based on age, height, and weight.

Code

Predictive Analysis

```r
# Load necessary libraries
install.packages("caret")
install.packages("e1071")

library(caret)
library(e1071)

# Assuming data is already loaded

# Prepare the data
set.seed(42)
index <- createDataPartition(data$pts, p=0.8, list=FALSE)
train <- data[index,]
test <- data[-index,]

# Linear regression model
model <- lm(pts ~ age + player_height + player_weight, data=train)

# Model summary
summary(model)

# Predictions
predictions <- predict(model, test)

# Calculate RMSE
rmse <- sqrt(mean((predictions - test$pts)^2))
cat("RMSE: ", rmse, "\n")
```

Summary

- **Model Coefficients:**
    - **Age:** 0.018
    - **Player Height:** -0.624
    - **Player Weight:** 0.336

- **Mean Squared Error (MSE):** 36.49
- **Root Mean Squared Error (RMSE):** 6.04

Interpretation

The linear regression model suggests that while age and weight have a positive influence on points per game, player height negatively influences the points. The model's RMSE of 6.04 indicates a moderate predictive accuracy.

---

5. Conclusion

This analysis provides insights into NBA player performance by examining key metrics, exploring the relationship between performance and physical attributes, and building a predictive model for points per game. The descriptive analysis gave an overview of the player demographics, the diagnostic analysis explored relationships between performance metrics and physical attributes, and the predictive analysis provided a model to estimate points per game.