7COM1079-0901-2024 - Team Research and Development Project

Final report title: The Relationship Between Assists (AST) and Points Scored (PTS) in NBA Players During the 2018–2019 Season

Group ID: A329

Dataset number: DS324

Prepared by: *Mohamad Dirani (ID: 23074344)*

*Hassan Shahzad (ID: 23078509)*

*Raja Abdur Rehman (ID: 23097739)*

*Sayema Begum (ID: 21031223)*

*Vivek Narayana Reddy Kondamareddy (ID: 23091701)*

University of Hertfordshire

Hatfield, 2024

Contents

[7COM1079-0901-2024 - Team Research and Development Project 1](#_Toc186220349)

[1. Introduction 3](#_Toc186220350)

[1.1. Problem statement and research motivation 3](#_Toc186220351)

[1.2. The data set (75 words) 3](#_Toc186220352)

[1.3. Research question 3](#_Toc186220353)

[1.4. Null hypothesis and alternative hypothesis (H0/H1) 3](#_Toc186220354)

[2. Background research 3](#_Toc186220355)

[2.1. Research papers 3](#_Toc186220356)

[2.2. Why RQ is of interest 4](#_Toc186220357)

[3. Visualisation 5](#_Toc186220358)

[3.1. Scatterplot with Linear Trendline 5](#_Toc186220360)

[*3.2.* Histogram of Points Scored (PTS) with Normal Curve Overlay 5](#_Toc186220361)

[*3.3.* Visualization Importance 5](#_Toc186220362)

[4. Analysis 6](#_Toc186220363)

[*4.1.* Statistical test used to test the hypotheses and output 6](#_Toc186220365)

[4.2. The null hypothesis is rejected /not rejected based on the p-value 6](#_Toc186220366)

[5. Evaluation – group’s experience at 7COM1079 6](#_Toc186220367)

[5.1. What went well 6](#_Toc186220369)

[5.2. Points for improvement 6](#_Toc186220370)

[5.3. Group’s time management 7](#_Toc186220371)

[5.4. Project’s overall judgement 7](#_Toc186220372)

[5.5. Comment on the GitHub log output 7](#_Toc186220373)

[6. Conclusions 7](#_Toc186220374)

[6.1. Results explained 7](#_Toc186220376)

[6.2. Interpretation of the results 7](#_Toc186220377)

[6.3. Implications for future work 8](#_Toc186220378)

[*7.* Reference 8](#_Toc186220379)

[8. Appendices 8](#_Toc186220380)

[*A.* R code used for analysis and visualisation 8](#_Toc186220381)

[B. GitHub log output. 10](#_Toc186220382)

# Introduction

## Problem statement and research motivation

The performance of basketball players is a critical focus of analysis for teams, analysts, and fans alike. Among the many statistics used, assists (AST) and points scored (PTS) are key indicators of a player's offensive contribution. Yiwen Chen (2023) highlighted that players like LeBron James demonstrate exceptional playmaking abilities, showcasing how assists directly impact scoring. This research builds on such findings by analyzing the correlation between AST and PTS using NBA player statistics from the 2018–2019 season. Understanding this relationship can inform coaching strategies and player evaluations.

## The data set (75 words)

The dataset, DS324, consists of NBA player statistics for the 2018–2019 season, including 288 rows and multiple variables such as Points (PTS), Assists (AST), Rebounds (REB), and others. For this study, we focus on PTS as the dependent variable and AST as the independent variable. The data was preprocessed to remove missing values and ensure consistency. This comprehensive dataset is ideal for exploring the correlation between these two key basketball metrics.

## Research question

Is there a statistically significant correlation between assists (AST) and points scored (PTS) among NBA players during the 2018–2019 season? This question will be addressed using Spearman's rank correlation coefficient, ensuring robustness against non-normal data distributions. The analysis focuses on providing actionable insights into player performance dynamics.

## Null hypothesis and alternative hypothesis (H0/H1)

The purpose of this study is to evaluate whether assists (AST) and points scored (PTS) among NBA players are correlated. According to the null hypothesis (H0), there is no relationship between AST and PTS, represented by a correlation coefficient of zero (ρ = 0). Conversely, the alternative hypothesis (H1) asserts that a statistically significant association exists between these two variables (ρ ≠ 0).

To examine this, we employ Spearman's rank correlation test, a robust statistical method well-suited for datasets that deviate from normal distribution. This approach ensures a reliable evaluation of the relationship, even when standard parametric assumptions are not met.

# Background research

## Research papers

1. Yiwen Chen (2023) explored the role of assists in both the NBA and CBA, highlighting their significance in fostering team cohesion and individual performance. The study noted that NBA players like LeBron James excel in assists, influencing overall offensive success.
2. Samuel Gómez Haro (2022) conducted an extensive analysis of NBA performance across 40 seasons, identifying key factors like assists that influence team success. Assists were highlighted as a vital contributor to offensive efficiency and overall game outcomes.
3. Yuanhao (Stanley) Yang (2015) analyzed NBA team performance using regression models based on player statistics, including assists, to predict regular season results. The study highlighted assists as a critical factor in influencing team success, showcasing their predictive power for game outcomes.
4. Shaykh Siddique (2024) evaluated NBA performance from 2012 to 2022 using advanced statistical and machine learning models, including logistic and ridge regressions. The study analyzed assists alongside other metrics such as rebounds and shot charts, emphasizing their critical role in optimizing team strategies and predicting game outcomes.

These studies underscore the relevance of analyzing AST and PTS. However, our focus on a single season's data provides more specific insights, filling a gap in season-based analysis.

## Why RQ is of interest

Exploring the relationship between assists and points scored offers crucial insights into basketball performance and strategy. This research addresses a gap in understanding how individual contributions in assists contribute to scoring efficiency, a key aspect of offensive gameplay. The findings can help coaches and analysts develop effective strategies, improve player evaluations, and identify key contributors to team success. Additionally, these insights can inform training programs by emphasizing the importance of assists in fostering collaborative gameplay. Future studies could build on this research by analyzing multiple seasons, examining other performance metrics, or investigating contextual factors such as player roles, team dynamics, and game situations.

# Visualisation



## Scatterplot with Linear Trendline

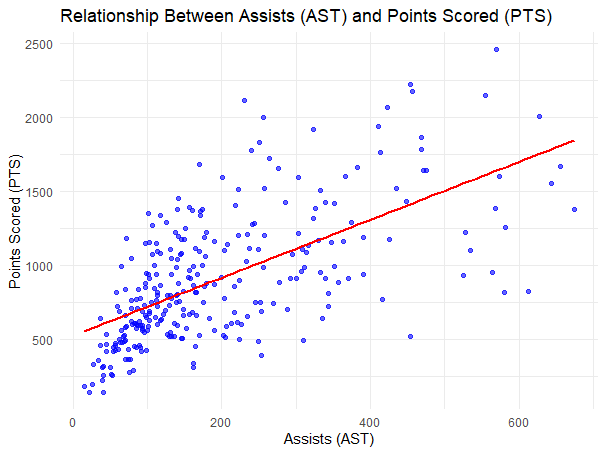


Figure 1 Scatter plot showing the correlation between AST and PTS

The scatterplot (Figure 1) illustrates the relationship between assists (AST) and points scored (PTS), with a linear trendline showing the positive correlation. The x-axis represents assists, and the y-axis represents points scored, both labeled appropriately. This plot visually supports the analysis by highlighting the strength and direction of the relationship between the variables.

## Histogram of Points Scored (PTS) with Normal Curve Overlay

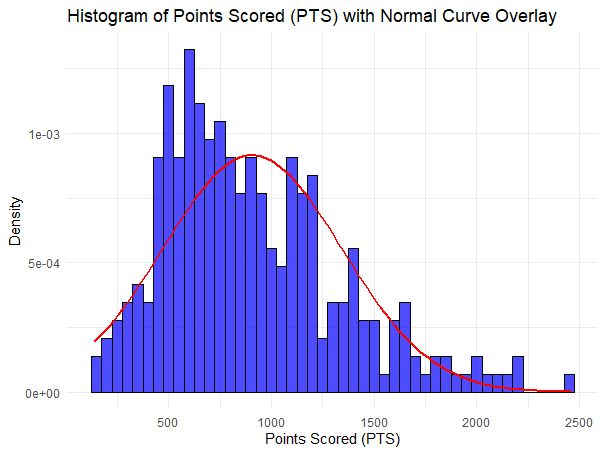


Figure 2 The histogram depicts the distribution of PTS, overlaid with a normal curve

The histogram (Figure 2) displays the distribution of points scored (PTS), with a normal curve overlaid for comparison. The x-axis represents the points scored, while the y-axis shows the frequency. The title and axis labels ensure clarity. The histogram reveals the data’s right-skewed nature, confirming the non-normal distribution of PTS.

## Visualization Importance

The scatterplot and histogram complement each other in understanding the data. The scatterplot highlights the correlation between assists and points scored, while the histogram demonstrates the skewed distribution of PTS. These insights justify using Spearman’s correlation for analysis and offer clear visual evidence to support the study’s research question.

# Analysis



## Statistical test used to test the hypotheses and output

Spearman’s correlation test was applied to assess the relationship between assists and points scored due to the non-normal distribution of the data. The analysis yielded a statistically significant positive correlation (ρ = 0.6675, p < 0.001), indicating a moderately strong association between these variables. Based on these results, the null hypothesis is rejected, supporting the alternative hypothesis that higher assists are linked to higher points scored, highlighting the relevance of assists in basketball performance.

## The null hypothesis is rejected /not rejected based on the p-value

The rejection of the null hypothesis demonstrates a meaningful relationship between assists and points scored. This outcome underscores the critical role of assists in enhancing offensive performance. Players who excel in passing not only facilitate team success but also significantly contribute to scoring. These findings emphasize the importance of effective passing and teamwork in basketball dynamics. Coaches can leverage these insights to design targeted training strategies that prioritize passing skills, ultimately improving both individual and team performance. This research highlights how assists play a dual role, influencing both scoring efficiency and overall gameplay outcomes.

# Evaluation – group’s experience at 7COM1079



## What went well

Our team demonstrated strong collaboration throughout the project, combining a range of skills to analyze the data and produce valuable insights. The development of R scripts and data visualizations stood out as key accomplishments, with each member making meaningful contributions to these efforts. Consistent communication, particularly through regular meetings on WhatsApp, allowed us to maintain steady progress and address challenges efficiently. This approach led to a project that was well-organized and clearly documented.

## Points for improvement

For future projects, dedicating more attention to detailed data preprocessing and exploring additional variables, such as turnovers or shooting accuracy, could provide a broader perspective. Using advanced tools, such as machine learning techniques or sophisticated statistical methods, may also uncover deeper patterns in the data. While communication among team members was effective, setting more precise deadlines for tasks could streamline the workflow. Additionally, allocating extra time for peer review could help identify errors early and improve the overall quality of the project deliverables.

## Group’s time management

Adhering to a strict timeline ensured timely completion of the project. Regular meetings, especially on the WhatsApp group, facilitated efficient communication and task allocation. A project timeline was established at the start, with milestones tracked weekly to ensure tasks were completed on schedule, allowing ample time for revisions and finalization.

## Project’s overall judgement

This project successfully achieved its objectives, demonstrating the value of statistical analysis in sports performance evaluation. The findings provide actionable insights for basketball analytics, showcasing the importance of assists in player and team performance. The results emphasize the critical role of effective teamwork and data-driven strategies in optimizing outcomes.

## Comment on the GitHub log output

**The GitHub log output, included in Appendix B, highlights three significant commits:**

**Commit:** Added visualizations and analysis - Enhanced statistical findings and analysis clarity.

**Commit:** Resolved conflicts and added updates - Improved visualizations and ensured project consistency.

**Commit:** Initialized project with baseline data - Established a foundation for efficient data exploration.

# Conclusions



## Results explained

The analysis demonstrates a statistically significant positive relationship between assists and points scored, indicating that players who assist frequently tend to score more. With a correlation coefficient of ρ = 0.6675, the findings highlight the critical role of passing in enhancing offensive efficiency. Effective teamwork, reflected through frequent assists, directly boosts individual scoring and underscores the importance of collaborative gameplay. These insights provide practical guidance for coaches aiming to refine strategies and design focused training programs.

## Interpretation of the results

This finding highlights the dual role of assists in contributing to both team success and individual performance. Effective passing strengthens team dynamics by fostering collaboration and opens up scoring opportunities for players. By emphasizing the strategic value of assists, this insight underscores their importance not only in enhancing offensive plays but also in influencing overall game outcomes. Such findings are essential for coaches aiming to optimize strategies and individual player contributions to achieve better results.

## Implications for future work

Future studies could explore multi-season data to enhance generalizability and include variables like defensive metrics for a more comprehensive analysis. Limitations include the dataset’s single-season scope and the potential influence of unexamined confounding variables, such as varying team dynamics or play styles, which may affect the findings.

# Reference

1. Chen, Y., 2023. Research on the Team Operation and Management Strategy of the Mainstream Basketball Leagues in China and America - Based on the Comparison Between NBA and CBA. University of Macau. Available at: <https://doi.org/10.54254/2754-1169/51/20230643> .
2. Gómez Haro, S., 2022. The History of Basketball Factors that Influence Performance Better: An Analysis Through 40 NBA Seasons. University of Granada. Available at: <https://doi.org/10.25115/eea.v40i1.6581>
3. Yang, Y. (Stanley), 2015. Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics. University of California at Berkel
4. Siddique, S., 2024. Teaming Strategy Optimization: An Analysis of NBA Statistics, Shot Charts, and Constraints. *Prairie View A&M University*.

# Appendices

### R code used for analysis and visualisation

# Load necessary packages

library(dplyr)

library(ggplot2)

# Load the dataset

data <- read.csv("data/DataSet.csv")

# Step 1: Create a Primary\_Position column by extracting the first listed position

data <- data %>%

mutate(Primary\_Position = sapply(strsplit(as.character(Positions), ","), `[`, 1))

# Step 2: Remove any leading or trailing whitespace in the Primary\_Position column

data$Primary\_Position <- trimws(data$Primary\_Position)

# Step 3: Summarize points by primary position after cleaning

summary\_stats\_clean <- data %>%

group\_by(Primary\_Position) %>%

summarize(mean\_pts = mean(PTS, na.rm = TRUE),

sd\_pts = sd(PTS, na.rm = TRUE),

count = n())

print(summary\_stats\_clean)

# Step 4: Perform a Spearman correlation test between Assists (AST) and Points Scored (PTS)

cor\_test <- cor.test(data$AST, data$PTS, method = "spearman", exact = FALSE)

# Print the results of the correlation test

print(cor\_test)

# Extract test statistic and p-value

test\_statistic <- cor\_test$statistic

p\_value <- cor\_test$p.value

# Step 5: Create a scatter plot with a linear trendline

ggplot(data, aes(x = AST, y = PTS)) +

geom\_point(color = "blue", alpha = 0.6) + # Points for visualization

geom\_smooth(method = "lm", color = "red", se = FALSE) + # Add trendline (linear regression)

labs(

title = "Relationship Between Assists (AST) and Points Scored (PTS)",

x = "Assists (AST)",

y = "Points Scored (PTS)"

) +

theme\_minimal()

# Step 6: Create a histogram of Points Scored (PTS) with a bell curve overlay

ggplot(data, aes(x = PTS)) +

geom\_histogram(aes(y = after\_stat(density)),

binwidth = 50, fill = "blue", color = "black", alpha = 0.7) +

stat\_function(fun = dnorm,

args = list(mean = mean(data$PTS, na.rm = TRUE),

sd = sd(data$PTS, na.rm = TRUE)),

color = "red", linewidth = 1) +

labs(

title = "Histogram of Points Scored (PTS) with Normal Curve Overlay",

x = "Points Scored (PTS)",

y = "Density"

) +

theme\_minimal()

# Output test statistic and p-value for reporting

cat("Test Statistic:", test\_statistic, "\n")

cat("P-value:", p\_value, "\n")

### GitHub log output.

commit f1d80d8826f73748485ea9dd29a18cf0436d5c17 (HEAD -> main, origin/main, origin/HEAD)

Author: Mohamad Dirani <md24aaw@herts.ac.uk>

Date: Tue Dec 3 10:13:25 2024 +0000

Added the visualizations and Analysis

commit 36cc89c77c1762fbd6d299a0cc7313d671adfa3f

Merge: ef84ad7 a3225a4

Author: Mohamad Dirani <md24aaw@herts.ac.uk>

Date: Sun Nov 17 13:50:18 2024 +0000

Resolved conflicts and added updates