

ADSC1910 Case 4

Neha Malhan(T00757554) , Thiwanka(T00754871)

2024-10-04

Introduction

This report aims to characterize NHL player performance during the 2022-2023 season by focusing on key metrics like positions, games played, goals, assists, shots on goal, and plus/minus rating. These metrics provide valuable insights into offensive and defensive contributions and will help sports writers better understand player performance and team strategies.

Methods

From NHL dataset, goals, assists, shots on goal, plus/minus, and games played were selected because these metrics are central to evaluating a player's offensive and defensive contributions. Goals and shots on goal are used to represent a player's ability to create scoring opportunities, while assists highlight their role in setting up teammates for success and plus/minus is used to reflect their overall impact on team success.

Goals were grouped by position to highlight how different roles contribute to scoring, with a bar plot used for a clear comparison of average goals by position. For other variables like assists, plus/minus, and shots on goal, mean, variance, and confidence intervals were applied to better understand player performance. The mean represents typical contributions, variance highlights differences between players, and confidence intervals show the reliability of performance. These methods provide a clear and simple picture of both scoring and overall player impact.

During the data cleaning process, NA and NaN values were identified, representing missing or unavailable data. To make the report more understandable for the intended audience, these values were replaced with 0, simplifying the presentation while assuming non-participation rather than missing data. This adjustment makes it possible to analyze player metrics more clearly.

Findings

1) Average Goals by Player Position with Consistency Ranges

Table 1: Average Goals by Player Position with Consistency Ranges

Position	Mean Goals	CI Lower	CI Upper
C	12.71	9.19	16.23
C/RW	25.00	0.00	0.00
D	2.88	1.86	3.91
D/RW	6.50	-37.97	50.97
F	2.53	0.44	4.63
G	0.00	0.00	0.00
LW	11.50	7.21	15.79
RW	10.08	6.53	13.63

Table 1 shows how players in different NHL positions perform when it comes to scoring goals. Players in forward positions like centers and left wings tend to score more goals and are reliable contributors to their team's offense. Their performance is steady, meaning they can be counted on to score regularly.

Defensemen, while scoring fewer goals overall, also perform consistently, though their offensive impact is smaller. Positions like defense/right wing, however, show more unpredictable scoring, with some players scoring a lot while others score less often.

As expected, goalies don't score goals, and the data shows no variation here. Some positions, like center/right wing, don't have enough data to make solid conclusions about their goal-scoring abilities. In summary, the data highlights which positions consistently contribute to scoring and which positions have more ups and downs in their performance.

2) Player Performance: Average, Consistency, and Reliability

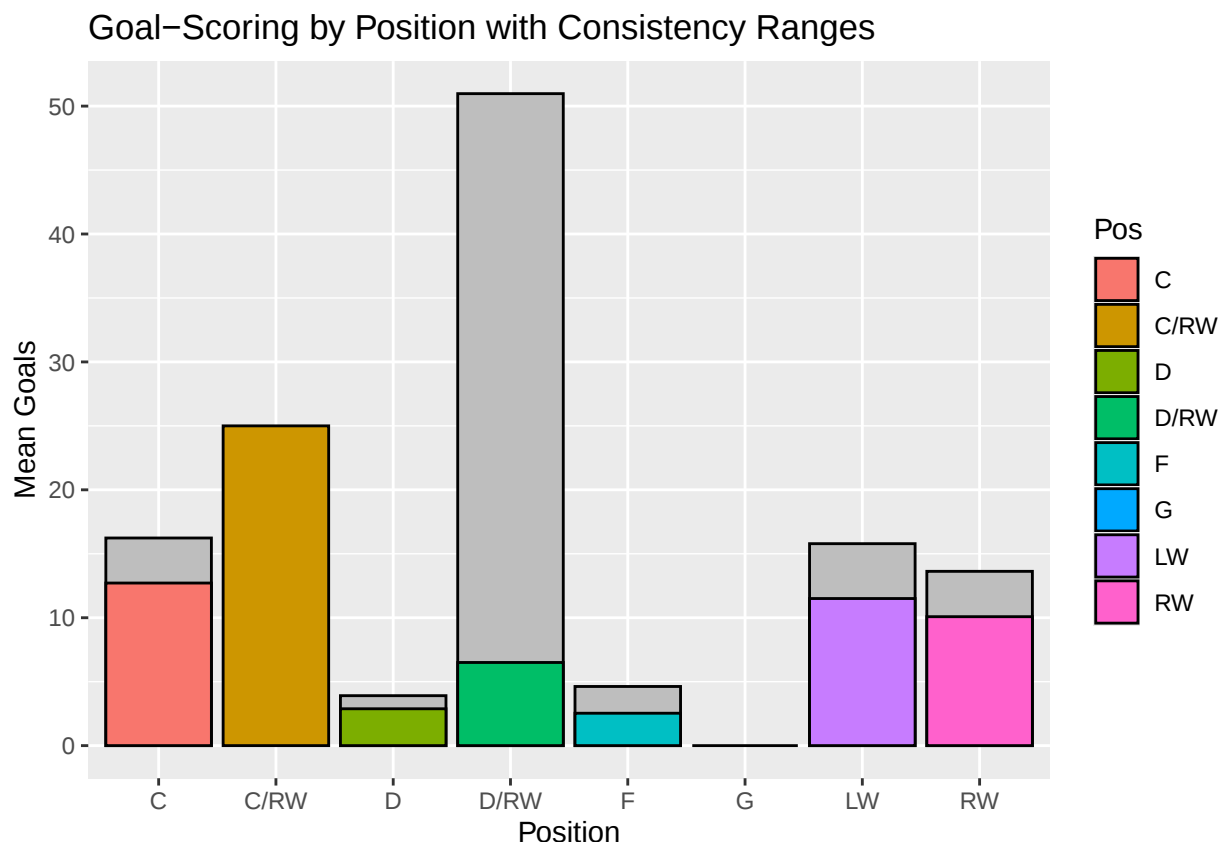
Table 2: Player Performance: Average, Consistency, and Reliability

Variable	Mean	Variance	CI Lower	CI Upper
Games Played	46.57	846.03	42.69	50.45
Assists	12.97	224.17	10.97	14.97
Plus_Minus	0.28	99.31	-1.05	1.61
Shots on Goal	74.90	5725.44	64.80	85.00

Table 2 shows how NHL players performed in different areas. Some players participated regularly throughout the season, while others missed significant time. Assists show that while most players contributed to setting up plays, a few stood out as the main creators for their team's offense. The plus/minus rating reveals a wide range in player impact, with some making a positive difference on the ice, while others struggle to prevent goals. Shots on goal varied widely, with some players taking many more shots than others, creating more scoring chances.

Conclusion

A barplot was created to present the data from Table 1, making it more accessible and relevant to the audience of the sports website.



According to above plot and Table 1, the goal-scoring differs across player positions. Some positions, like C/RW, stand out with strong performances, while D/RW shows more inconsistency in scoring. Positions like Left Wing and Center are more reliable, regularly contributing to the team's offense, whereas Defense and Forward players score less but are steady in their efforts. This table helps show which positions teams can rely on for consistent goal-scoring and where there are more ups and downs.

Table 2 provides a simple look at key player stats. Shots on goal vary the most, meaning some players take a lot of shots and create scoring chances, while others don't shoot as much. Assists and plus/minus are more balanced, showing that most players contribute equally in making plays and defense. This helps teams and sports writers understand which players are active on offense and which ones focus more on supporting roles or defense.

Additionally, the C/RW and D/RW positions had unpredictable performances, with some players scoring inconsistently. This shows that their contributions to scoring can vary. It's crucial to recognize the absence of goal-scoring from goalies, which should be kept in mind when looking at the overall results.

Recommendation

Teams should focus on players who take the most shots and create scoring chances, while also sharing the responsibility of setting up plays among more players to keep the team balanced. Improving defense, especially for players with lower plus/minus ratings, will help the team perform better. Positions like Center and Left Wing, which are more reliable for scoring goals, should be important in the team's game plan, but teams should also work on getting more goals from forwards and defense players. For positions like C/RW and D/RW, which don't score as consistently, teams should change their approach to make better use of these players' abilities.

References

- 1) Hellingman, S. (2024). ADSC 1000: Sampling and Estimation
- 2) Understanding Ice Hockey
- 3) Information on variables
- 4) groupby(), mutate()

Appendices

We, Neha Malhan(T00757554) and Thiwanka Abeyweera(T00754871), worked together equally on all parts of the case study.

```
knitr::opts_chunk$set(echo = TRUE)

# Importing the data

library(dplyr)

# Load the dataset from the working directory
NHL_data <- read.csv("NHL22.csv")

# Check the structure of the dataset
str(NHL_data)

library(knitr)

# computing the confidence interval for goal when observations are > 2
# ifelse(n() > 1,...) will check the number of observation
# NaN indicates that the CI can't be calculated due to insufficient data points in the group

grouped_data <- NHL_data %>%
  group_by(Pos) %>%
  summarise(
    Mean_Goals = mean(G),
    CI_Lower = ifelse(n() > 1, t.test(G, conf.level = 0.95)$conf.int[1], NaN),
    CI_Upper = ifelse(n() > 1, t.test(G, conf.level = 0.95)$conf.int[2], NaN)
  )

# Data cleaning - replace NA and NaN values with 0
```

```

cleaned_data <- grouped_data %>%
  mutate(
    CI_Lower = ifelse(is.na(CI_Lower) | is.nan(CI_Lower), 0, CI_Lower),
    CI_Upper = ifelse(is.na(CI_Upper) | is.nan(CI_Upper), 0, CI_Upper),
    # Round off all values to two decimal places
    Mean_Goals = sprintf("%.2f", Mean_Goals),
    CI_Lower = sprintf("%.2f", CI_Lower),
    CI_Upper = sprintf("%.2f", CI_Upper)
  )

cleaned_data <- cleaned_data[, c("Pos", "Mean_Goals", "CI_Lower", "CI_Upper")]

# Display the result as a table
kable(cleaned_data,
      col.names = c("Position", "Mean Goals", "CI Lower", "CI Upper"),
      caption = "Average Goals by Player Position with Consistency Ranges")

# Calculate mean, variance, and CI for each variable and rounding off to 2 decimal points

summary_data <- data.frame(
  Variable = c("Games Played", "Assists", "Plus_Minus", "Shots on Goal"),
  Mean = round(c(mean(NHL_data$GP),
                  mean(NHL_data$A),
                  mean(NHL_data$Plus_Minus),
                  mean(NHL_data$SOG)), 2),

  Variance = round(c(var(NHL_data$GP),
                      var(NHL_data$A),
                      var(NHL_data$Plus_Minus),
                      var(NHL_data$SOG)), 2),

  CI_Lower = round(c(t.test(NHL_data$GP, conf.level = 0.95)$conf.int[1],
                          t.test(NHL_data$A, conf.level = 0.95)$conf.int[1],
                          t.test(NHL_data$Plus_Minus, conf.level = 0.95)$conf.int[1],
                          t.test(NHL_data$SOG, conf.level = 0.95)$conf.int[1]), 2),

  CI_Upper = round(c(t.test(NHL_data$GP, conf.level = 0.95)$conf.int[2],
                          t.test(NHL_data$A, conf.level = 0.95)$conf.int[2],
                          t.test(NHL_data$Plus_Minus, conf.level = 0.95)$conf.int[2],
                          t.test(NHL_data$SOG, conf.level = 0.95)$conf.int[2]), 2)
)

kable(summary_data,
      col.names = c("Variable", "Mean", "Variance", "CI Lower", "CI Upper"),
      caption = "Player Performance: Average, Consistency, and Reliability")

# barplot for table 1

library(ggplot2)
ggplot(grouped_data, aes(x = Pos , fill= Pos)) +
  # stat = "identity" to plot data exactly as it is

```

```
geom_bar(aes(y = CI_Upper), stat = "identity",color = "black", fill = "grey") +  
geom_bar(aes(y = Mean_Goals), stat = "identity", color = "black") +  
xlab("Position") +  
ylab("Mean Goals") +  
ggtitle("Goal-Scoring by Position with Consistency Ranges")
```