stat184 project

Annie Wang, Sanjana Vuppunahalli

2024-12-10

1. Introduction and Research Question

In the National Hockey League, the ability to project a player's future performance is fundamental to effective team management and personal player decisions. Teams face decisions about roster construction, trades, contract negotiations, and game strategies, all of which rely heavily on evaluating a player's historical performance and determine potential. Among various measures, shooting percentage (the ratio of goals scored to shots taken) is a critical measure of a player's scoring efficiency. A higher shooting percentage often indicates a player's ability to capitalize on scoring opportunities, while a lower percentage may suggest inefficiencies or limitations in their ability to convert chances into goals.

Understanding the factors that influence shooting percentage is important because scoring goals is the most direct way to winning games. Shooting efficiency may based on numerous factors, such as the player's position, the quality of their scoring chances, and their ability to perform under pressure. The variability makes it essential to analyze whether certain plater positions, such as center (C), left wings (L), or right wings (R), are associated with differences in shooting percentage. Centers are often tasked with playmaking, faceoffs, and defense play responsibilities, they may take fewer but higher-quality shots, focusing more on setting up teammates for scoring opportunities. Wingers are responsible for offensive plays, they tend to shoot more frequently and from diverse locations on the ice, potentially leading to different scoring outcomes.

Historically, research and analysis in sports statistics have suggested that player position can affect performance metrics. For example, basketball guards tend to have better free-throw percentages than forwards or centers players due to different skillsets and roles. Similarly, in hockey, differences in player responsibilities and positioning could influence scoring efficiency. However, there is limited literature that systematically examines how positional differences impact shooting percentage in professional hockey. This lack of insight motivates this analysis. The dataset used for this story, from Score NHLShots, includes NHL player statistics from 2019-2021 seasons, with information about goals, shots, and calculated shooting percentages. Using this dataset, this analysis can be applied to answer the research question: Does the position of the shooter significantly affect their scoring percentage in NHL games?

This study aims to provide insights for team personnel by identifying whether positional roles influence shooting efficiency. These findings could assist teams in optimizing their rosters, improving in-game strategies, and making data-driven decisions when assessing player potential. By bridging gaps in understanding potential. By bridging gaps in understanding positional differences, this research contributes to the broader field of sports analytics and player evaluation in professional hockey.

2. Exploratory Data Analysis

The data set contains NHL player statistics from the 2019-2021 seasons, focusing on player position (C, L, R) and shooting performance metrics, such as goals, shots, and shooting percentages. There is no missing data in our data set as it had already been removed before uploaded, the final data set contains 347 records. Each record represents an NHL player and includes their position, shooting percentage for 2019 and 2020, and an average shooting percentage calculated for this analysis.

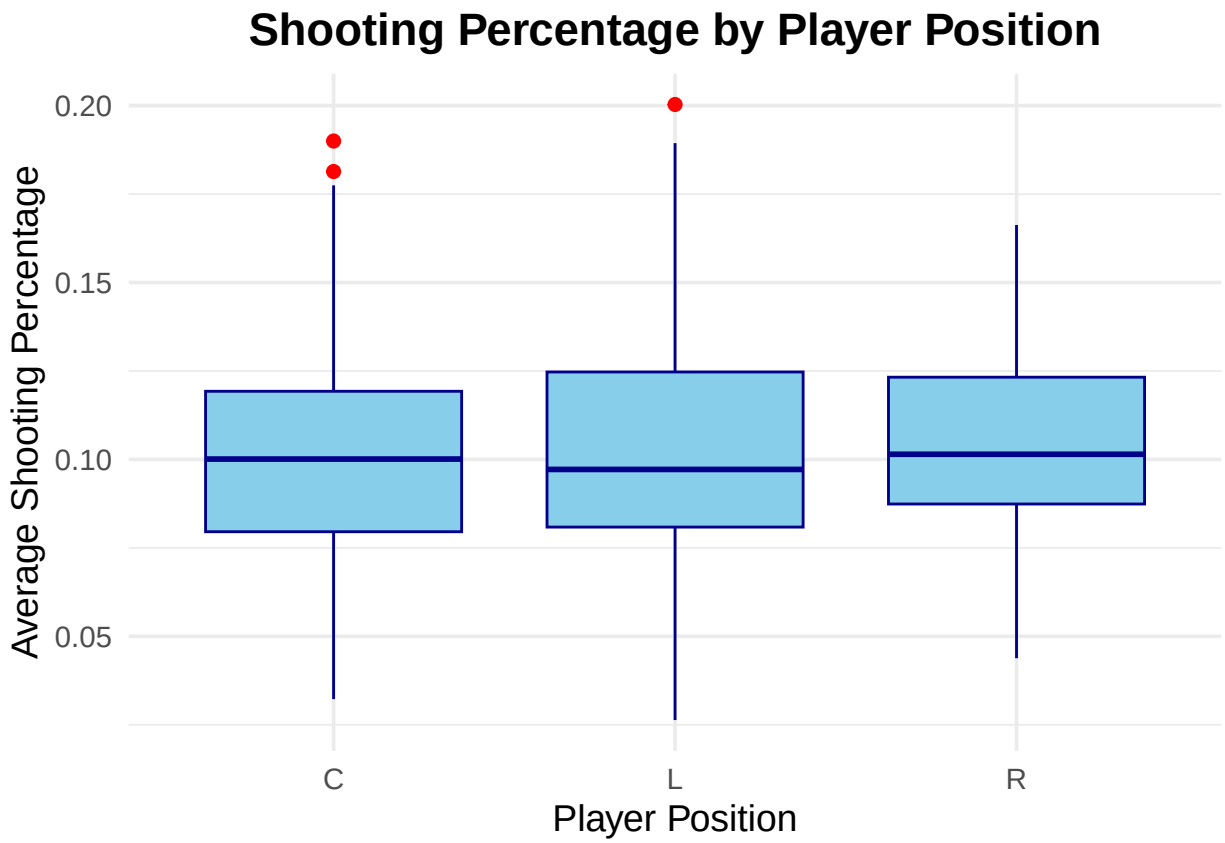
```
##        ID           Player Pos Goals2019 Shots2019     Sh2019 Goals2020
## 1 8479318  Auston Matthews   C        31       199 0.15577889        28
## 2 8478402   Connor McDavid   C        21       143 0.14685315        19
## 3 8478420   Mikko Rantanen   R         9        69 0.13043478        18
## 4 8478493 Joel Eriksson Ek   C         7        92 0.07608696        17
## 5 8478483  Mitchell Marner   R         7        97 0.07216495        17
## 6 8477944      Jakub Vrana   L        22       152 0.14473684        17
##   Shots2020    Sh2020 id
## 1       159 0.1761006  1
## 2       130 0.1461538  2
## 3       107 0.1682243  3
## 4       108 0.1574074  4
## 5       103 0.1650485  5
## 6        76 0.2236842  6
```

```
##   Pos     Sh2019    Sh2020 AverageShootingPercentage
## 1   C 0.15577889 0.1761006                 0.1659398
## 2   C 0.14685315 0.1461538                 0.1465035
## 3   R 0.13043478 0.1682243                 0.1493295
## 4   C 0.07608696 0.1574074                 0.1167472
## 5   R 0.07216495 0.1650485                 0.1186067
## 6   L 0.14473684 0.2236842                 0.1842105
```

Table 1: Summary Statistics of NHL Player Data

Pos	Sh2019	Sh2020	AverageShootingPercentage
Length:347	Min. :0.00000	Min. :0.00000	Min. :0.02633
Class :character	1st Qu.:0.07735	1st Qu.:0.07056	1st Qu.:0.08220
Mode :character	Median :0.10000	Median :0.09804	Median :0.09954
NA	Mean :0.10225	Mean :0.10209	Mean :0.10217
NA	3rd Qu.:0.12500	3rd Qu.:0.13043	3rd Qu.:0.12192
NA	Max. :0.20652	Max. :0.30000	Max. :0.20033

Table 1 provides a summary of the dataset. The majority of the players are distributed across three positions: Center (C), Left Wing (L), and Right Wing (R). Shooting percentages range from approximately 2.6% to 20% across the dataset, with a mean of 10.2%. These statistics suggest variability in scoring efficiency among players, which will be further explored by visualizing positional differences.



The boxplot displays the distribution of average shooting percentages for players grouped by position (C, L, R). This visualization highlights the spread and central tendency of shooting efficiency for each position, as well as any potential outliers. Figure 1 reveals that the distributions of shooting percentages are similar across all three positions, with slight variations in medians and ranges. While outliers are present, particularly among left wingers, the overall trends suggest that positional differences in shooting percentages may not be substantial. This observation will be tested statistically in the next section.

3. Statistical Analysis

The ANOVA test compares the mean shooting percentages across the three positions: Center (C), Left Wing (L), and Right Wing (R). The null hypothesis states that the mean shooting percentages are the same across all positions, while the alternative hypothesis posits at least one positional group has a different mean.

Table 2: ANOVA Table: Testing Differences in Shooting Percentage by Position

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Pos	2	0.001	0.001	0.55	0.577
Residuals	344	0.364	0.001	NA	NA

Table 2 summarizes the results of the ANOVA test. The F-statistic value is 0.55 with a corresponding p-value of 0.577, larger than the significance level of 0.05. This indicates that there is not engough evidence to reject the null hypothesis. The differences in mean shooting percentages among Centers, Left Wings, and

Right Wings are not statistically significant. This suggests that positional roles do not play a major role in determining shooting efficiency in the NHL.

While the ANOVA test did not find significant differences in shooting percentages based on position, this does not eliminate the possibility that other factors, such as team strategies, individual skill levels, or game contexts, might influence shooting efficiency. Future analyses could incorporate additional variables or interaction effects to uncover more nuanced relationships.

4. Discussion

The results of this analysis indicate that positional roles (Center, Left Wing, Right Wing) in the NHL do not significantly impact a player's shooting efficiency, as evidenced by the ANOVA test with a p-value of 0.577. The exploratory data analysis showed overlapping distributions of shooting percentages across positions, further supporting the conclusion that position alone is not a strong determinant of shooting efficiency. These findings highlight the complexity of player performance in hockey, suggesting that shooting efficiency may be influenced more by individual skill, team dynamics, or game situations than by positional differences.

For NHL teams, these findings suggest that focusing solely on player position when evaluating shooting efficiency may not yield applicable insights. Instead, teams could benefit from a more comprehensive evaluation framework that considers other variables, such as shot quality, game context, or player-specific attributes. This approach could help teams optimize player development, in-game strategies, and roster decisions, ultimately enhancing overall performance.

The data set is openly available on SCORE NHLShots platform, with a clear and persistent identifier. The data set can be accessed by anyone with an internet connection, ensuring transparency and broad usability. The data set is in CSV format and can be processed using common statistical and programming tools. The data set includes sufficient documentation on variable definitions and data collection meothod, which allows researchers to reuse it for additional studies.Therefore, the data aligns with the FAIR principles (Findable, Accessible, Interoperable, and Reusable). The data set benefits the hockey analytics community and it is collected by a trusted academic institution, ensuring ethical data use and proper attribution. The analysis respects the privacy of players by focuing on a aggregated metrics without disclosing sensitive personal data. The data is used in ethical manner to ensure the analysis is transparent. The data set also aligns with key elemetns of the CARE principles (Collective benefit, Authority to control, Responsibility, and Ethics)

Despite the strengths, this study has several limitations. The data set does not account for contextual factors, such as the quality of opposition, game situations, or team strategies, which may influence shooting efficiency. The data set only covers two seasons (2019-2021), which may not capture long-term trends or changes in player performance. Future research could expand this analysis by incorporating additional variables, such as expected goals or shot location data. They can also conduct similar analyses for other leagues or player demographics to generalize findings.

5. Reference

Ventura, Sam. "Please Enable JavaScript in Order to Use Isle." NHL Shooting Percentage, isle.stat.cmu.edu/SCORE/NHLShots/. Accessed 10 Dec. 2024.

6. Code Appendix

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

```r
library(ggplot2)
library(knitr)


# Import data
nhl_data <- read.csv("~/Downloads/NHL Shooter Performance 2019-2021.csv")
head(nhl_data)

# Clean and wrangle data
nhl_cleaned <- nhl_data %>%
  select(Pos, Sh2019, Sh2020) %>%
  mutate(AverageShootingPercentage = (Sh2019 + Sh2020) / 2) %>%
  na.omit()
head(nhl_cleaned)

kable(summary(nhl_cleaned), caption = "Summary Statistics of NHL Player Data")


ggplot(nhl_cleaned, aes(x = Pos, y = AverageShootingPercentage)) +
  geom_boxplot(fill = "skyblue", color = "darkblue", outlier.color = "red", outlier.size = 2) +
  theme_minimal(base_size = 14) +
  labs(title = "Shooting Percentage by Player Position",
       x = "Player Position",
       y = "Average Shooting Percentage") +
  theme(plot.title = element_text(hjust = 0.5, face = "bold"))

anova_result <- aov(AverageShootingPercentage ~ Pos, data = nhl_cleaned)
anova_summary <- summary(anova_result)

# Display ANOVA results as a table
kable(anova_summary[[1]],
      caption = "ANOVA Table: Testing Differences in Shooting Percentage by Position",
      digits = 3,
      align = "c")
```