

# Analyzing Factors Influencing Batting Strike Rate and the Role of Batting Style in IPL Players

Rohan Dalal

Pratham Patel

2024-12-18

## Introduction

Cricket is a globally popular bat-and-ball sport played between two teams, where the objective is to score runs while preventing the opposing team from doing the same, combining strategy, skill, and athleticism. It is a sport where data analytics plays a crucial role in evaluating player performance and shaping game strategies. Among various performance metrics, the batting strike rate is a standout measure, reflecting a batsman's scoring efficiency as runs scored per 100 balls faced. This metric is particularly vital in shorter formats like T20 cricket, where players often aim for high strike rates to make a decisive impact on the game. However, the strike rate alone does not provide a complete picture of a player's contribution, as it must be analyzed alongside other metrics like batting average and boundary frequency.

In this project, we aim to explore two core research questions:

1. What factors (e.g., runs, boundaries, matches played, balls faced, batting average, and batting style) are most predictive of a batsman's strike rate?
2. Does a batsman's batting style (left-hand bat vs. right-hand bat) influence their batting average and boundary-hitting ability?

These questions aim to fill gaps in understanding how scoring efficiency connects with playing style and consistency. The insights gained could play a pivotal role in shaping team strategies, player selection, and overall performance improvements. The overarching goal of this project is to provide actionable insights that emphasize the importance of strike rate and related metrics in modern cricket.

## Data Provenance

For this project, we used two datasets to explore factors influencing batting strike rate and the role of batting style in IPL players. The datasets were curated from reliable online sources and combined through data wrangling and cleaning processes to produce a final dataset for analysis.

### Primary Dataset

- Source: Kaggle
- Description: The primary dataset contains detailed performance statistics for IPL players from 2008 to 2024, including batting and bowling metrics. It has 604 rows and 16 columns, representing individual player performances across multiple IPL seasons.
- Purpose: This dataset was originally collected to provide comprehensive insights into player performance, including metrics like batting strike rate, boundaries, boundaries percentage, batting average, balls faced, and matches played.
- Cases: Each row represents an IPL player, with columns detailing their batting and bowling statistics.

### Secondary Dataset

- Source: GitHub
- Description: The secondary dataset provides metadata about IPL players who played the 2024 season, including their batting style (left-hand or right-hand), bowling style, and other details like playing role and player profile links. It has 262 rows and 13 columns.
- Purpose: This dataset was collected to offer additional contextual information about players, particularly batting and bowling styles.
- Cases: Each row represents an IPL player, with columns capturing their roles and playing styles.

## Data Wrangling

To address our research questions, we integrated the primary and secondary datasets into a single, cleaned dataset:

### Merging:

- The datasets were merged using the player column from the primary dataset and the Name column from the secondary dataset.

- The merge allowed us to combine batting performance metrics with metadata like batting style.

## **Variable Selection**

- After merging, unnecessary columns such as `imgUrl`, `longName`, and other unrelated attributes were dropped.
- We retained variables directly relevant to our analysis, such as runs, boundaries, balls faced, matches, batting average, batting strike rate, boundaries percentage, and batting style.

## **Filtering and Cleaning**

- Players who faced fewer than 100 balls across matches were excluded to ensure that the analysis focused on consistent performers.
- The final dataset includes only those players who met this threshold, resulting in 92 rows with all required columns.

## **Final Dataset**

- **Description:** The final dataset is a curated selection of the most relevant variables for our analysis, combining performance statistics with metadata like batting styles. It has 92 rows, the 8 selected variables required to answer the research questions, and the player name.
- **Purpose:** This dataset enables us to analyze factors influencing batting strike rate and assess the impact of batting style on performance metrics like boundaries and batting average.
- **Case:** Each row represents an IPL player, with columns capturing their batting statistics and style.

By merging and cleaning the datasets, we ensured that our final data aligns with our research objectives, is reliable, and is focused solely on relevant attributes. This curated dataset provides the foundation for meaningful insights into IPL player performance.

## **FAIR Principles**

The datasets used in this project adhere to the FAIR principles, which ensure data integrity and usability in research:

## **Findable**

- The primary dataset is sourced from Kaggle, a well-known platform for datasets and data challenges, making it easy to locate with clear metadata and descriptions.
- The secondary dataset is publicly available on GitHub, ensuring that both data sources are easily discoverable.

## **Accessible:**

- Both datasets are openly accessible without restrictions, allowing researchers to download and use the data without barriers. Kaggle and GitHub provide reliable hosting platforms with consistent access.

## **Interoperable:**

- The datasets were formatted in .csv files, a universally accepted format that is compatible with most data analysis tools and programming languages like R and Python.
- During data wrangling, we standardized column names and formats to ensure smooth integration, making the data interoperable within our analysis.

## **Reusable:**

- The datasets come with sufficient documentation about their content and purpose, enabling others to use them for similar or extended analyses.
- By cleaning and curating the final dataset, we improved its usability for future research.

## **CARE Principles**

Although our data doesn't include sensitive or community-specific information, we prioritized ethical standards throughout the project. The datasets we used are publicly available, and our analysis was conducted with a focus on fairness and collective benefit. No private information was accessed, and we ensured that all interpretations remained unbiased and responsible.

## Exploratory Data Analysis (EDA)

To derive meaningful insights from our data and address the research questions, we conducted extensive Exploratory Data Analysis (EDA). This included generating summary tables and creating various visualizations to better understand relationships between key variables and uncover patterns in player performance.

We started by constructing summary tables to provide an overview of key statistics such as runs, boundaries, batting average, strike rate, and more. These tables allowed us to compare metrics across players and batting styles, highlighting trends and differences in performance.

Next, we developed data visualizations to explore deeper relationships and distributions within the dataset:

- Scatter plots: These helped us visualize how variables like runs, boundaries, balls faced, matches, batting average, batting strike rate, and boundaries percentage relate to batting strike rate, with batting styles distinguished by color.
- Box plots: These illustrated the distribution of metrics like batting strike rate, batting average and boundaries across different batting styles.
- Correlation heatmap: This highlighted the strength of relationships between batting strike rate and other performance metrics.
- Histograms: These revealed the overall distribution of batting strike rates across players.

The insights gained from these analyses form the foundation of our results and conclusions. Below, we detail the most critical tables and visualizations, along with the key insights they helped us derive.

### Descriptive Statistics of Batting Metrics

This table (Table 1) provides an overview of the dataset's key batting performance metrics, including runs, boundaries, boundaries\_percent, balls\_faced, matches, batting\_avg, and strike\_rate. The summary statistics (mean, median, minimum, and maximum) offer valuable context about the range and distribution of these variables.

Table 1: Summary Statistics for Key Batting Variables

Variable	Mean	Median	Minimum	Maximum
Runs	1327.45	568.00	96.00	6634.00
Boundaries	170.26	74.50	10.00	838.00
Balls Faced	1029.02	492.50	102.00	5266.00
Matches	72.91	63.00	5.00	221.00
Batting Average	17.98	19.78	1.34	39.34
Boundaries Percentage	15.94	15.67	7.69	26.36

Strike Rate	125.96	126.47	80.77	168.23
-------------	--------	--------	-------	--------

Key Insights:

- The mean batting strike rate is 125.96, with values ranging from 80.77 to 168.23, highlighting variation in scoring efficiency among players.
- On average, players faced 1,029 balls, scored 1,327 runs, and hit 170 boundaries, demonstrating the general scale of performance in the dataset.

This table establishes a foundational understanding of the dataset and help us explore relationships further in the context of our analysis.

## Factors Influencing Batting Strike Rate

### Correlation Between Strike Rate and Other Metrics

This correlation heatmap (Figure 1) illustrates the relationships between `batting_strike_rate` and key batting metrics such as runs, boundaries, boundaries\_percent, matches, balls\_faced, and batting\_avg. The heatmap helps identify the variables that have the strongest influence on batting strike rate

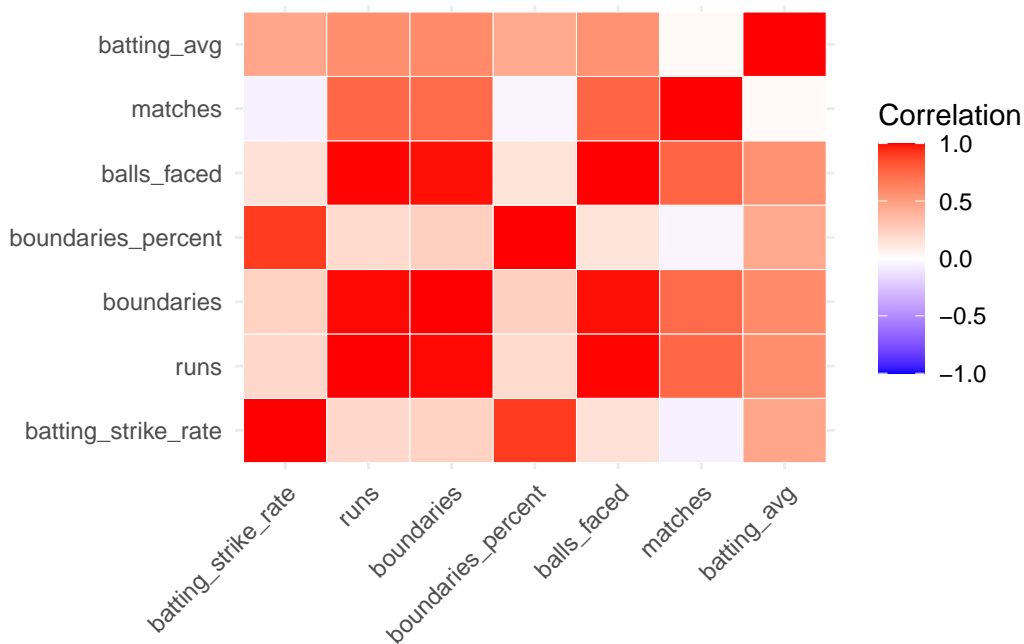


Figure 1: Correlation Heatmap of Key Batting Variables

Key Insights:

- Boundaries\_Percent shows the strongest positive correlation with batting\_strike\_rate, indicating that players who score a higher percentage of their runs through boundaries tend to have higher strike rates. In shorter formats like T20 cricket, where aggressive scoring is crucial, this finding emphasizes the importance of boundary-hitting ability as a primary driver of strike rate. It suggests that players focusing on hitting fours and sixes can significantly impact their teams' scoring pace.
- Batting\_Avg also demonstrates a moderate positive correlation, suggesting that consistent scorers (players with higher batting averages) are more likely to have higher strike rates.
- Other metrics, such as balls\_faced and matches, show minimal or no significant correlation with batting\_strike\_rate. This implies that these variables, while important for contextualizing a player's career or role in a team, are less critical when analyzing scoring efficiency. For example, players who face more balls or play more matches do not necessarily have higher strike rates, highlighting the limited predictive value of these metrics for scoring efficiency.

The heatmap lays the ground for deeper analysis of boundaries\_percent and batting\_avg, which are explored further in the scatter plots.

## Scatter Plots

The following scatter plot (Figure 2) shows the relationship between batting\_strike\_rate and boundaries\_percent, with color-coded points representing batting styles (left-hand bat vs. right-hand bat). A loess curve has been added to identify trends.

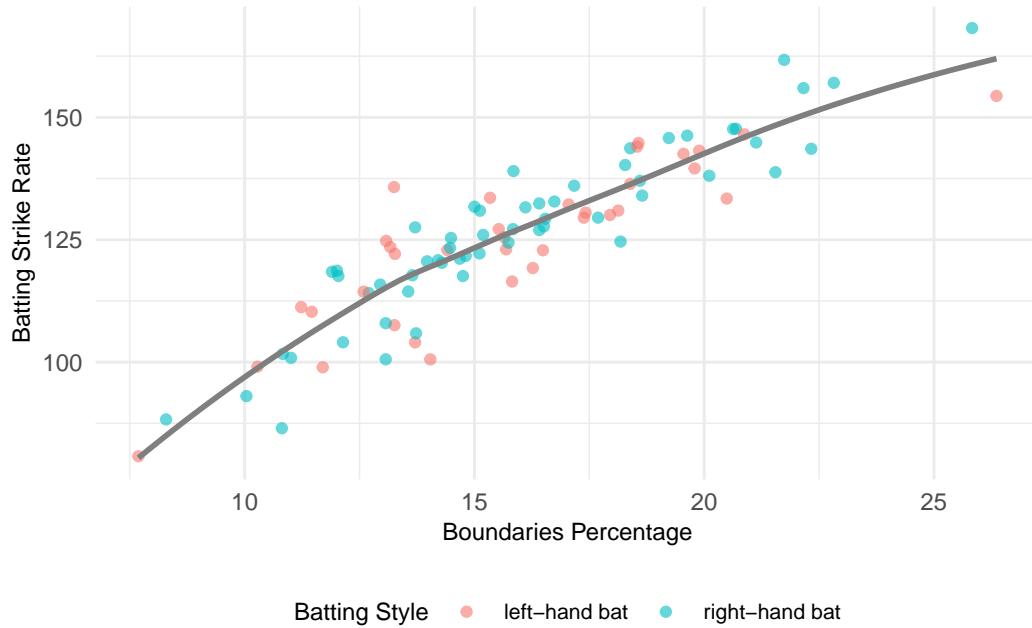


Figure 2: Scatter Plot of Batting Strike Rate vs Boundaries Percentage by Batting Style

#### Key Insights:

- A strong positive relationship is evident between `boundaries_percent` and `batting_strike_rate`. Players with higher percentages of runs from boundaries consistently achieve higher strike rates.
- The relationship holds true across both batting styles, as evidenced by the overlapping trend line. However, the distribution of players shows some variation, with right-handed batsmen slightly more prevalent at higher `boundaries_percent` values.
- This showcases the insight that focusing on aggressive scoring through boundaries is a key factor in driving high strike rates, regardless of batting style, making it an essential strategy for impactful performances in modern cricket.

The scatter plot (Figure 3) examines the relationship between `batting_strike_rate` and `batting_avg`, again differentiating batting styles.



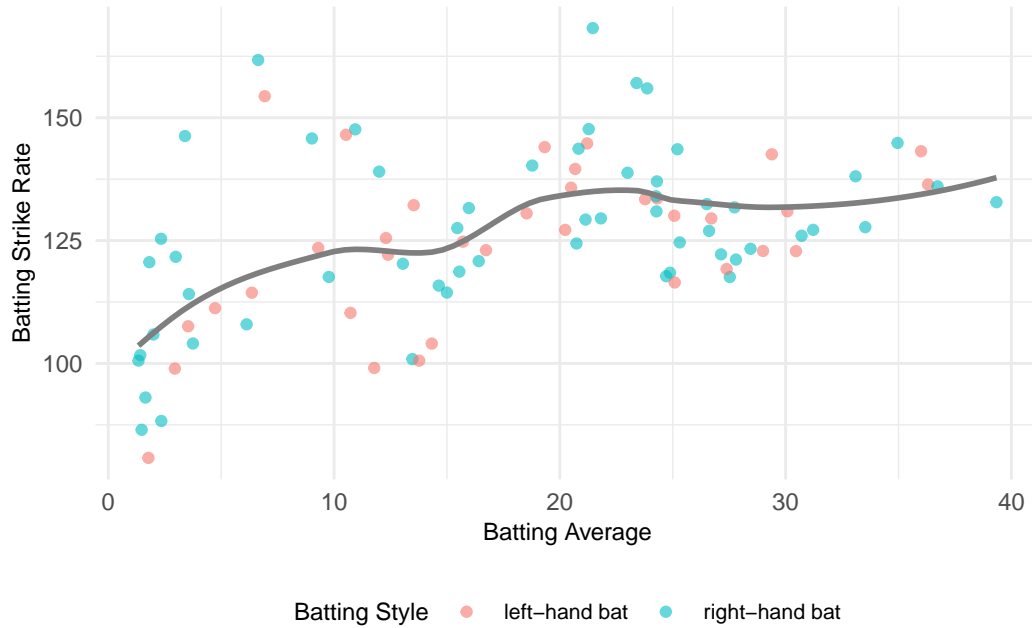


Figure 3: Scatter Plot of Batting Strike Rate vs Batting Average by Batting Style

#### Key Insights:

- A moderate positive relationship is visible between `batting_avg` and `batting_strike_rate`. Players with higher averages tend to maintain higher strike rates, though the trend is less pronounced than with `boundaries_percent`.
- The loess curve reveals a slightly nonlinear relationship, with strike rates stabilizing at higher batting averages.
- This suggests that while batting consistency contributes to scoring efficiency, it is not as impactful as boundary-hitting.

Together, these visualizations highlight the significance of `boundaries_percent` and `batting_avg` in determining a player's batting strike rate, aligning with the findings from the correlation heatmap.

## Influence of Batting Style on Batting Metrics

### Batting Style Comparison

This table (Table 2) compares key batting metrics between left-handed and right-handed batsmen. The metrics include average runs scored, boundaries hit, batting average, and strike rate.

Table 2: Comparison of Batting Metrics by Batting Style

Batting Style	Avg_Runs	Avg_Boundaries	Avg_Batting_Average	Avg_Strike_Rate
Left-Handed Bat	1171.54	153.14	18.04	124.62
Right-Handed Bat	1423.18	180.77	17.94	126.79

Key Insights:

- Right-handed batsmen outperform left-handed batsmen across all metrics, with a higher average of runs (1,423.18 vs. 1,171.54), boundaries (180.77 vs. 153.14), and strike rate (126.79 vs. 124.62). These differences suggest that right-handed batsmen may have an edge in aggressive and impactful scoring.
- Left-handed batsmen show a slightly better batting average (18.04 vs. 17.94) compared to right-handed batsmen, indicating marginally higher consistency in run-scoring. However, the difference is minimal, suggesting that batting average is not significantly influenced by batting style.
- Implications: While the differences in metrics are not drastic, they suggest that batting style does have an influence on certain performance metrics, particularly on the ability to score boundaries and maintain a high strike rate. This supports the notion that batting style could be a factor in performance evaluation, especially in shorter cricket formats.

These findings set the stage for the upcoming visualisations, which delve deeper into the relationship between batting style and boundary-hitting ability, as well as total boundaries scored.

## Box Plots

This box plot (Figure 4) illustrates the distribution of boundaries hit by left-handed and right-handed batsmen. It provides a clear comparison of the performance between the two batting styles.

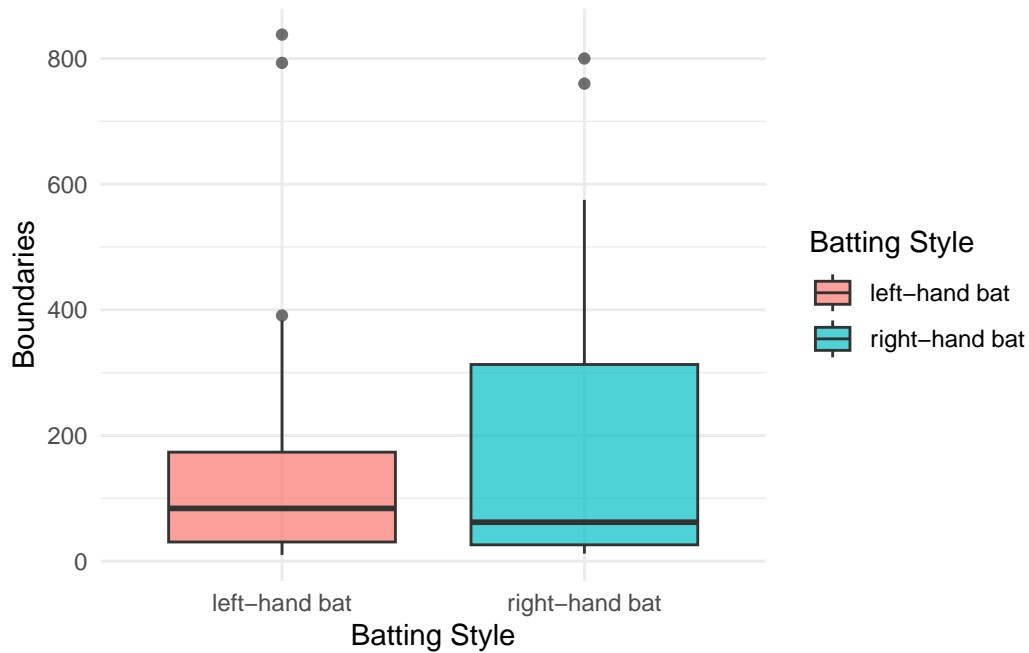


Figure 4: Box Plot of Boundaries by Batting Style

#### Key Insights:

- Building on the findings of Table 2, right-handed batsmen have a significantly wider range in the number of boundaries hit, with a higher median and greater maximum values compared to left-handed batsmen.
- Left-handed batsmen, while consistent, show a smaller range and lower overall performance in boundary hitting.
- This suggests that batting style plays a role in boundary-hitting ability, with right-handed batsmen generally performing better.

This box plot (Figure 5) compares the batting averages of left-handed and right-handed batsmen.

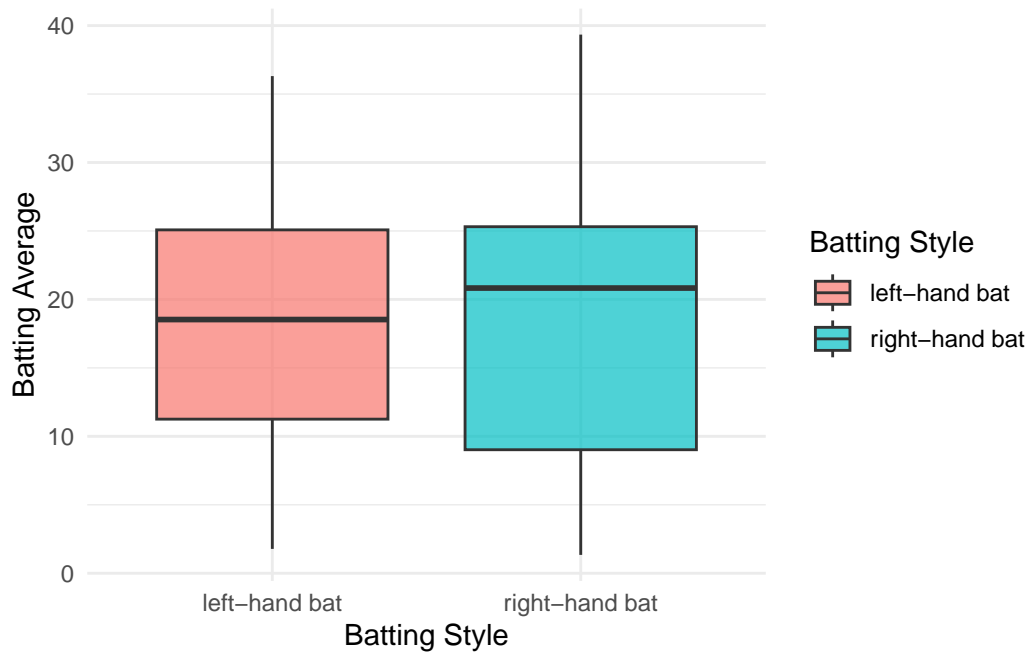


Figure 5: Box Plot of Batting Average by Batting Style

#### Key Insights:

- As seen before, both batting styles show very similar distributions for batting averages, with nearly identical medians and ranges.
- This reinforces the finding from the batting style comparison table that batting style does not significantly influence batting average.

Given the lack of variability in batting average by batting style, this metric appears less relevant to the differences in performance.

#### Stacked Bar Chart

The bar chart (Figure 6) provides a cumulative comparison of the total boundaries hit by left-handed and right-handed batsmen.

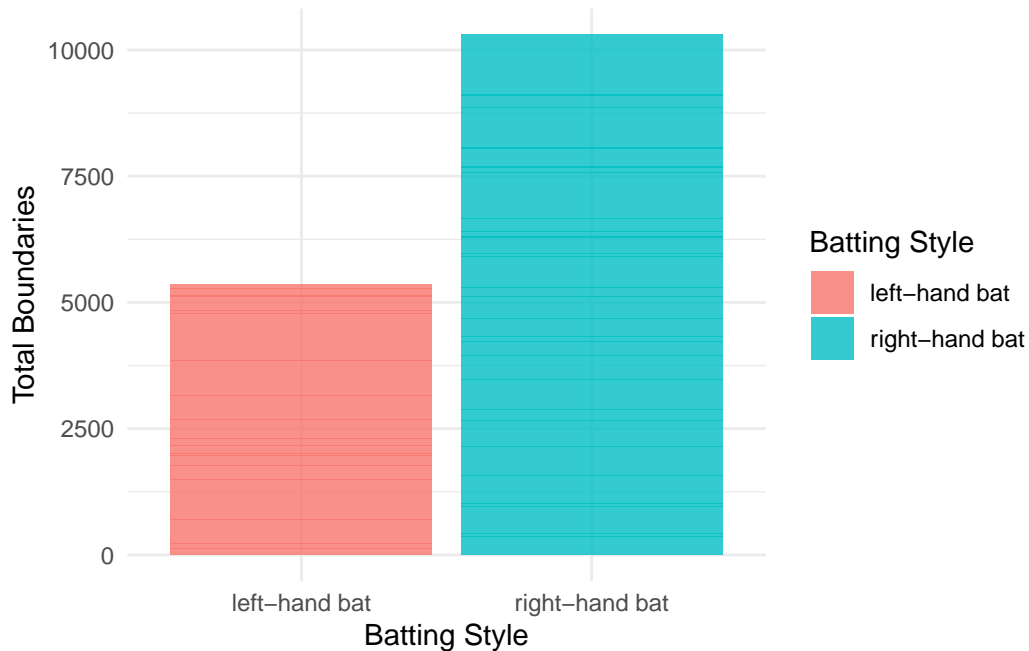


Figure 6: Stacked Bar Chart for Total Boundaries by Batting Style

#### Key Insights:

- Right-handed batsmen have hit a substantially higher number of total boundaries compared to left-handed batsmen.
- This complements the box plot findings and emphasizes the stronger boundary-hitting ability of right-handed batsmen overall.

## Conclusion

This project explored the factors influencing batting strike rate and the impact of batting style on performance metrics such as boundaries and batting average. Our analysis revealed that boundaries percentage has the strongest influence on batting strike rate, followed by batting average, highlighting the importance of aggressive scoring in shorter cricket formats like T20. Batting style significantly impacts boundary-hitting performance, with right-handed batsmen consistently outperforming left-handed batsmen in total and median boundaries hit. However, batting average is largely unaffected by batting style, with both left-handed and right-handed batsmen showing similar distributions and median values. These findings emphasize the nuanced role of batting style in player performance and the critical importance of boundary-hitting ability in driving scoring efficiency. This project demonstrates the value of

data-driven insights in sports analytics and provides a foundation for further exploration into player performance dynamics.

## References/Sources

IPL players statistics. (n.d.). *Kaggle*. Retrieved December 14, 2024, from <https://www.kaggle.com/datasets/mohammadzamakhan/ipl-players-statistics>

IPL-DATASET/csv/2024\_players\_details.csv at main · ritesh-ojha/IPL-DATASET. (n.d.). *GitHub*. Retrieved December 14, 2024, from [https://github.com/ritesh-ojha/IPL-DATASET/blob/main/csv/2024\\_players\\_details.csv](https://github.com/ritesh-ojha/IPL-DATASET/blob/main/csv/2024_players_details.csv)

Strike rate. (2024). In *Wikipedia*. Retrieved December 14, 2024, from [https://en.wikipedia.org/w/index.php?title=Strike\\_rate&oldid=1260997340](https://en.wikipedia.org/w/index.php?title=Strike_rate&oldid=1260997340)

## Code Appendix

```
# Load Required Libraries ----
library(dplyr)
library(knitr)
library(kableExtra)
library(ggplot2)
library(reshape2)

# Load Datasets ----
# Load the primary dataset containing IPL player statistics
primary_data <- read.csv(
  file = "/Users/rohan1809/Desktop/STAT 184/IPL Player Stat.csv"
)

# Load the secondary dataset with additional player details
secondary_data <- read.csv(
  file = "/Users/rohan1809/Desktop/STAT 184/2024_players_details.csv"
)

# Preprocessing: Rename Columns ----
# Rename the `player` column in the primary dataset to match for joining
primary_data <- primary_data %>%
  rename(Name = player)
```

```

# Merge Datasets ----
# Perform a left join to combine datasets based on the `Name` column
merged_data <- primary_data %>%
  left_join(
    secondary_data,
    by = "Name"
  )

# Filter Merged Data ----
# Remove rows with missing values in the `longBattingStyles` column
filtered_merged_data <- merged_data %>%
  filter(!is.na(longBattingStyles))

# View and Save Filtered Data ----
# Check the structure of the filtered data
str(filtered_merged_data)
# View the filtered data in a tabular format
View(filtered_merged_data)

# Save the filtered merged dataset to a CSV file (optional)
# write.csv(
#   filtered_merged_data,
#   file = "Merged_Player_Data.csv",
#   row.names = FALSE
# )

# Data Cleaning and Selection ----
# Select relevant columns, rename for clarity, and filter data
cleaned_data <- filtered_merged_data %>%
  select(
    Name,                # Player's name
    runs,                # Total runs scored
    boundaries,          # Number of boundaries
    boundaries_percent,  # Percentage of boundaries
    balls_faced,         # Total balls faced
    matches,             # Number of matches played
    batting_avg,         # Batting average
    batting_strike_rate, # Strike rate
    longBattingStyles    # Batting style (e.g., left-hand/right-hand)
  ) %>%
  rename(
    batting_style = longBattingStyles # Rename for clarity
  )

```

```

) %>%
  filter(
    balls_faced > 100 # Include only players who faced more than 100 balls
  )

# View and Save Cleaned Data ----
# View the cleaned data
View(cleaned_data)

# Save the cleaned dataset to a new CSV file (optional)
# write.csv(
#   cleaned_data,
#   file = "Cleaned_Merged_Player_Data.csv",
#   row.names = FALSE
# )

# Create Summary Statistics Table ----
# Select relevant columns for summary
batting_data <- cleaned_data %>%
  select(
    runs,
    boundaries,
    balls_faced,
    matches,
    batting_avg,
    boundaries_percent,
    batting_strike_rate
  )

# Calculate summary statistics
summary_table <- batting_data %>%
  summarise(
    Variable = c(
      "Runs",
      "Boundaries",
      "Balls Faced",
      "Matches",
      "Batting Average",
      "Boundaries Percentage",
      "Strike Rate"
    ),
    Mean = c(

```



```

    mean(runs, na.rm = TRUE),
    mean(boundaries, na.rm = TRUE),
    mean(balls_faced, na.rm = TRUE),
    mean(matches, na.rm = TRUE),
    mean(batting_avg, na.rm = TRUE),
    mean(boundaries_percent, na.rm = TRUE),
    mean(batting_strike_rate, na.rm = TRUE)
  ),
  Median = c(
    median(runs, na.rm = TRUE),
    median(boundaries, na.rm = TRUE),
    median(balls_faced, na.rm = TRUE),
    median(matches, na.rm = TRUE),
    median(batting_avg, na.rm = TRUE),
    median(boundaries_percent, na.rm = TRUE),
    median(batting_strike_rate, na.rm = TRUE)
  ),
  Minimum = c(
    min(runs, na.rm = TRUE),
    min(boundaries, na.rm = TRUE),
    min(balls_faced, na.rm = TRUE),
    min(matches, na.rm = TRUE),
    min(batting_avg, na.rm = TRUE),
    min(boundaries_percent, na.rm = TRUE),
    min(batting_strike_rate, na.rm = TRUE)
  ),
  Maximum = c(
    max(runs, na.rm = TRUE),
    max(boundaries, na.rm = TRUE),
    max(balls_faced, na.rm = TRUE),
    max(matches, na.rm = TRUE),
    max(batting_avg, na.rm = TRUE),
    max(boundaries_percent, na.rm = TRUE),
    max(batting_strike_rate, na.rm = TRUE)
  )
)

# Rename columns for clarity
colnames(summary_table) <- c("Variable", "Mean", "Median", "Minimum", "Maximum")

# Display Summary Statistics Table ----
# Output the polished table with improved formatting

```

```

summary_table %>%
  kable(
    caption = "Summary Statistics for Key Batting Variables",
    digits = 2,
    align = c("l", rep("c", 4))
  ) %>%
  kable_styling(
    bootstrap_options = c("striped", "hover", "condensed"),
    full_width = FALSE,
    font_size = 14
  ) %>%
  row_spec(0, bold = TRUE) # Bold header row

# Create Batting Style Comparison Table ----
# Group data by batting style and calculate average metrics
batting_style_comparison <- cleaned_data %>%
  mutate(
    batting_style = case_when(
      batting_style == "left-hand bat" ~ "Left-Handed Bat",
      batting_style == "right-hand bat" ~ "Right-Handed Bat",
      TRUE ~ batting_style
    )
  ) %>%
  group_by(batting_style) %>%
  summarise(
    Avg_Runs = mean(runs, na.rm = TRUE),
    Avg_Boundaries = mean(boundaries, na.rm = TRUE),
    Avg_Boundaries_Percent = mean(boundaries_percent, na.rm = TRUE),
    Avg_Batting_Average = mean(batting_avg, na.rm = TRUE),
    Avg_Strike_Rate = mean(batting_strike_rate, na.rm = TRUE)
  ) %>%
  rename(`Batting Style` = batting_style)

# Display Batting Style Comparison Table ----
# Output the polished table with improved formatting
batting_style_comparison %>%
  kable(
    caption = "Comparison of Batting Metrics by Batting Style",
    digits = 2,
    align = c("l", rep("c", 5))
  ) %>%
  kable_styling(

```

```

    bootstrap_options = c("striped", "hover", "condensed"),
    full_width = FALSE,
    font_size = 14
  ) %>%
  row_spec(0, bold = TRUE) # Bold header row

# Correlation Heatmap ----
# Select relevant columns for correlation analysis
correlation_data <- cleaned_data %>%
  select(
    batting_strike_rate,
    runs,
    boundaries,
    boundaries_percent,
    balls_faced,
    matches,
    batting_avg
  )

# Calculate the correlation matrix
cor_matrix <- cor(
  correlation_data,
  use = "complete.obs"
)

# Melt the correlation matrix for heatmap visualization
cor_melted <- melt(cor_matrix)

# Create and display the correlation heatmap
ggplot(data = cor_melted, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(
    low = "blue",
    high = "red",
    mid = "white",
    midpoint = 0,
    limit = c(-1, 1),
    name = "Correlation"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),

```

```

    axis.title = element_blank()
  ) +
  labs(
    title = "Correlation Heatmap of Key Batting Variables"
  )

# Scatter Plots for Key Metrics ----
# Function to create scatter plots
create_scatter_plot <- function(x_var, x_label) {
  ggplot(data = cleaned_data, aes_string(x = x_var, y = "batting_strike_rate", color = "batt
  geom_point(alpha = 0.6) +
  geom_smooth(method = "loess", color = "gray50", se = FALSE) +
  labs(
    title = paste("Batting Strike Rate vs", x_label, "by Batting Style"),
    x = x_label,
    y = "Batting Strike Rate",
    color = "Batting Style"
  ) +
  theme_minimal() +
  theme(
    legend.position = "bottom",
    plot.title = element_text(size = 16),
    axis.title = element_text(size = 12),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 9)
  )
}

# Generate scatter plots for different metrics
create_scatter_plot("runs", "Runs")
create_scatter_plot("boundaries", "Boundaries")
create_scatter_plot("boundaries_percent", "Boundaries Percentage")
create_scatter_plot("balls_faced", "Balls Faced")
create_scatter_plot("matches", "Matches")
create_scatter_plot("batting_avg", "Batting Average")

# Boxplot for Batting Strike Rate by Batting Style ----
ggplot(data = cleaned_data, aes(x = batting_style, y = batting_strike_rate, fill = batting_s
  geom_boxplot(alpha = 0.7, outlier.color = "red", outlier.shape = 16, outlier.size = 2) +
  labs(
    title = "Batting Strike Rate Distribution by Batting Style",
    x = "Batting Style",

```

```

    y = "Batting Strike Rate",
    fill = "Batting Style"
) +
theme_minimal() +
theme(
  plot.title = element_text(size = 16),
  axis.title = element_text(size = 12),
  legend.position = "none"
)

# Histogram of Batting Strike Rate ----
ggplot(data = cleaned_data, aes(x = batting_strike_rate)) +
  geom_histogram(binwidth = 5, fill = "steelblue", color = "black", alpha = 0.7) +
  labs(
    title = "Distribution of Batting Strike Rate",
    x = "Batting Strike Rate",
    y = "Frequency"
) +
theme_minimal() +
theme(
  plot.title = element_text(size = 16),
  axis.title = element_text(size = 12),
  axis.text = element_text(size = 10)
)

# Boxplot for Batting Average by Batting Style ----
ggplot(data = cleaned_data, aes(x = batting_style, y = batting_avg, fill = batting_style)) +
  geom_boxplot(alpha = 0.7) +
  labs(
    title = "Batting Average by Batting Style",
    x = "Batting Style",
    y = "Batting Average",
    fill = "Batting Style"
) +
theme_minimal()

# Boxplot for Boundaries by Batting Style ----
ggplot(data = cleaned_data, aes(x = batting_style, y = boundaries, fill = batting_style)) +
  geom_boxplot(alpha = 0.7) +
  labs(
    title = "Boundaries by Batting Style",
    x = "Batting Style",

```

```

    y = "Boundaries",
    fill = "Batting Style"
  ) +
  theme_minimal()

# Stacked Bar Chart for Total Boundaries by Batting Style ----
ggplot(data = cleaned_data, aes(x = batting_style, y = boundaries, fill = batting_style)) +
  geom_bar(stat = "identity", alpha = 0.8) +
  labs(
    title = "Total Boundaries by Batting Style",
    x = "Batting Style",
    y = "Total Boundaries",
    fill = "Batting Style"
  ) +
  theme_minimal()

```