

ETC5513-Assignment3

```
library(tidyverse)
```

Warning: package 'ggplot2' was built under R version 4.3.3

Warning: package 'purrr' was built under R version 4.3.3

Warning: package 'lubridate' was built under R version 4.3.3

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.2      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.4
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(kableExtra)
```

Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

group_rows

```
# Import data and process
epl_raw <- read.csv("data/epl_final.csv")

epl_data <- epl_raw %>%
  select(Season, HomeTeam, AwayTeam, FullTimeResult, HomeShots, AwayShots)

# Filter to only include most recent seasons
recent_seasons <- c("2015/16", "2016/17", "2017/18", "2018/19", "2019/20",
                    "2020/21", "2021/22", "2022/23", "2023/24", "2024/25")

epl_recent <- epl_data %>%
  filter(Season %in% recent_seasons)
```

Executive Summary

Introduction

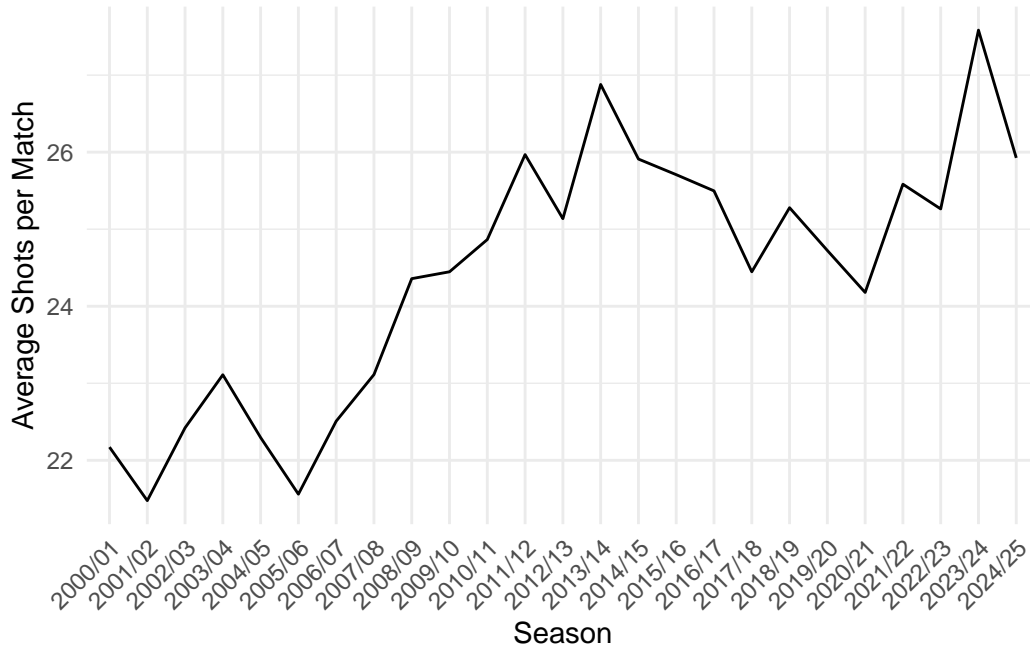
Methodology

Dataset Introduction

```
# Create variable for total number of shots each game
epl_processed_method <- epl_data %>%
  mutate(TotalShots = HomeShots + AwayShots)

# Summarise average shots per match by season
summary_by_season <- epl_processed_method %>%
  group_by(Season) %>%
  summarise(AvgShotsPerMatch = mean(TotalShots, na.rm = TRUE))

# Create line plot of total shots per game
ggplot(summary_by_season, aes(x = Season, y = AvgShotsPerMatch, group = 1)) +
  geom_line() +
  labs(x = "Season",
       y = "Average Shots per Match") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



As illustrated in Figure 1, there has been a noticeable increase in the average number of shots taken per match by English Premier League (EPL) teams over the past two decades. This upward trend suggests a gradual shift in playstyle toward more attack-oriented strategies. This study aims to explore whether this offensive shift is actually linked to better team performance—or whether teams that remain more balanced or conservative may still perform just as well, or even better, amid changing league dynamics.

We used match-level data sourced from Kaggle, covering EPL matches from the 2000/01 season through to 2024/25. From this full dataset, we extracted six variables relevant to offensive output and team performance: `Season`, `HomeTeam`, `AwayTeam`, `FullTimeResult`, `HomeShots`, and `AwayShots`. The processed dataset contained 9,380 observations across 25 seasons.

Data Description and Analytical Approach

```
# Create a description table
variable_description <- tibble::tibble(
  Variable = c("Season", "Team", "TotalMatches", "TotalShots", "TotalWins",
               "AvgShotsPerGame", "WinRatePercent"),

  Description = c(
    "The English Premier League season (e.g., '2015/16')",
    "Name of the football team",
```

```
"Total number of matches played by the team that season",
"Total number of shots taken by the team across the season",
"Total number of wins recorded by the team that season",
"Average number of shots made by the team each game in that season",
"Win rate percentage (TotalWins / TotalMatches * 100)"))
```

```
kable(variable_description)
```

Variable	Description
Season	The English Premier League season (e.g., ‘2015/16’)
Team	Name of the football team
TotalMatches	Total number of matches played by the team that season
TotalShots	Total number of shots taken by the team across the season
TotalWins	Total number of wins recorded by the team that season
AvgShotsPerGame	Average number of shots made by the team each game in that season
WinRatePercent	Win rate percentage (TotalWins / TotalMatches * 100)

For our analysis, we focused on the 10 most recent seasons (2015/16 to 2024/25) to reflect current tactical trends. Using this subset, we constructed a team-season-level dataset, where each observation represents one team’s performance in a single season (see Table 1). Two new variables were created to quantify offensive intensity and performance:

- **Average Shots Per Game:** This serves as a proxy for offensive aggression, indicating a team’s intent to generate scoring opportunities rather than prioritize possession or minimize risk.
- **Win Rate Percent:** This metric considers only wins, rather than combining wins and draws, because wins are the most valuable outcome in league play. They contribute the highest number of points and best reflect a team’s ability to consistently outperform opponents across a season.

To assess the relationship between attacking play and performance, we plotted Win Rate Percent against Average Shots Per Game. Each point in the scatter plot represents one team’s season statistics, allowing us to visually evaluate whether higher offensive output corresponds to improved results.

Results

Discussion, Conclusion & Recommendations

References