

Vis Project sem 1

Anurag Dey Roll no: MDS202210

2022-10-29

Project Name: English Premier League(Analysis of Home Advantage useful for betting purposes)

```
library(ggplot2)
library(tidyverse)
```

```
## — Attaching packages — tidyverse 1.3.2 —
## ✓ tibble 3.1.8      ✓ dplyr 1.0.10
## ✓ tidyr 1.2.1       ✓ stringr 1.4.1
## ✓ readr 2.1.3       ✓ forcats 0.5.2
## ✓ purrr 0.3.5
## — Conflicts — tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
```

```
library(dplyr)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##      combine
```

Introduction

Sports betting is a 500 billion dollar market (Sydney Herald)

Football is played by 250 million players in over 200 countries (most popular sport globally).

The English Premier League is the most popular domestic team in the world.

It contains 20 years of EPL matches dataset. The Analysis consists of several match statistics

generated over the 380 games played

About the dataset

```
setwd("C:/Users/anura/OneDrive/Desktop/visualization/database")
```

```
df=read.csv('2020-2021.csv',header=TRUE)
head(df)
```

```
##      Time      HomeTeam  AwayTeam FTHG FTAG FTR HTHG HTAG HTR   Referee HS AS
## 1 12:30      Fulham    Arsenal    0    3  A    0    1  A C Kavanagh  5 13
## 2 15:00 Crystal Palace Southampton  1    0  H    1    0  H   J Moss    5  9
## 3 17:30    Liverpool      Leeds   4    3  H    3    2  H   M Oliver 22  6
## 4 20:00      West Ham  Newcastle  0    2  A    0    0  D  S Attwell 15 15
## 5 14:00      West Brom  Leicester  0    3  A    0    0  D  A Taylor   7 13
## 6 16:30    Tottenham    Everton   0    1  A    0    0  D M Atkinson  9 15
##      HST AST HF AF HC AC HY AY HR AR
## 1   2   6 12 12  2  3  2  2  0  0
## 2   3   5 14 11  7  3  2  1  0  0
## 3   6   3  9  6  9  0  1  0  0  0
## 4   3   2 13  7  8  7  2  2  0  0
## 5   1   7 12  9  2  5  1  1  0  0
## 6   5   4 15  7  5  3  1  0  0  0
```

variables

- * HomeTeam = Home Team
- * Away team = Away Team
- * FTHG and HG = Full Time Home Team Goals
- * FTAG and AG = Full-Time Away Team Goals
- * FTR and Res = Full-Time Result (H=Home Win, D=Draw, A=Away Win)
- * HTHG = Half Time Home Team Goals
- * HTAG = Half Time Away Team Goals
- * HTR = Half Time Result (H=Home Win, D=Draw, A=Away Win)
- * Referee = Match Referee
- * HS = Home Team Shots
- * AS = Away Team Shots
- * HST = Home Team Shots on Target
- * AST = Away Team Shots on Target
- * HHW = Home Team Hit Woodwork
- * AHW = Away Team Hit Woodwork
- * HC = Home Team Corners

- * AC = Away Team Corners
- * HF = Home Team Fouls Committed
- * AF = Away Team Fouls Committed
- * HFKC = Home Team Free Kicks Conceded
- * AFKC = Away Team Free Kicks Conceded
- * HO = Home Team Offsides
- * AO = Away Team Offsides
- * HY = Home Team Yellow Cards
- * AY = Away Team Yellow Cards
- * HR = Home Team Red Cards
- * AR = Away Team Red Cards

Thought behind the Visual Analysis

Over the different graphs provided throughout the analysis i have tried to boil down the effect

of playing at home and thus help in bringing insights to the current betting system in place.

Performing well at home creates a ripple effect as it also invariably related to the trust that

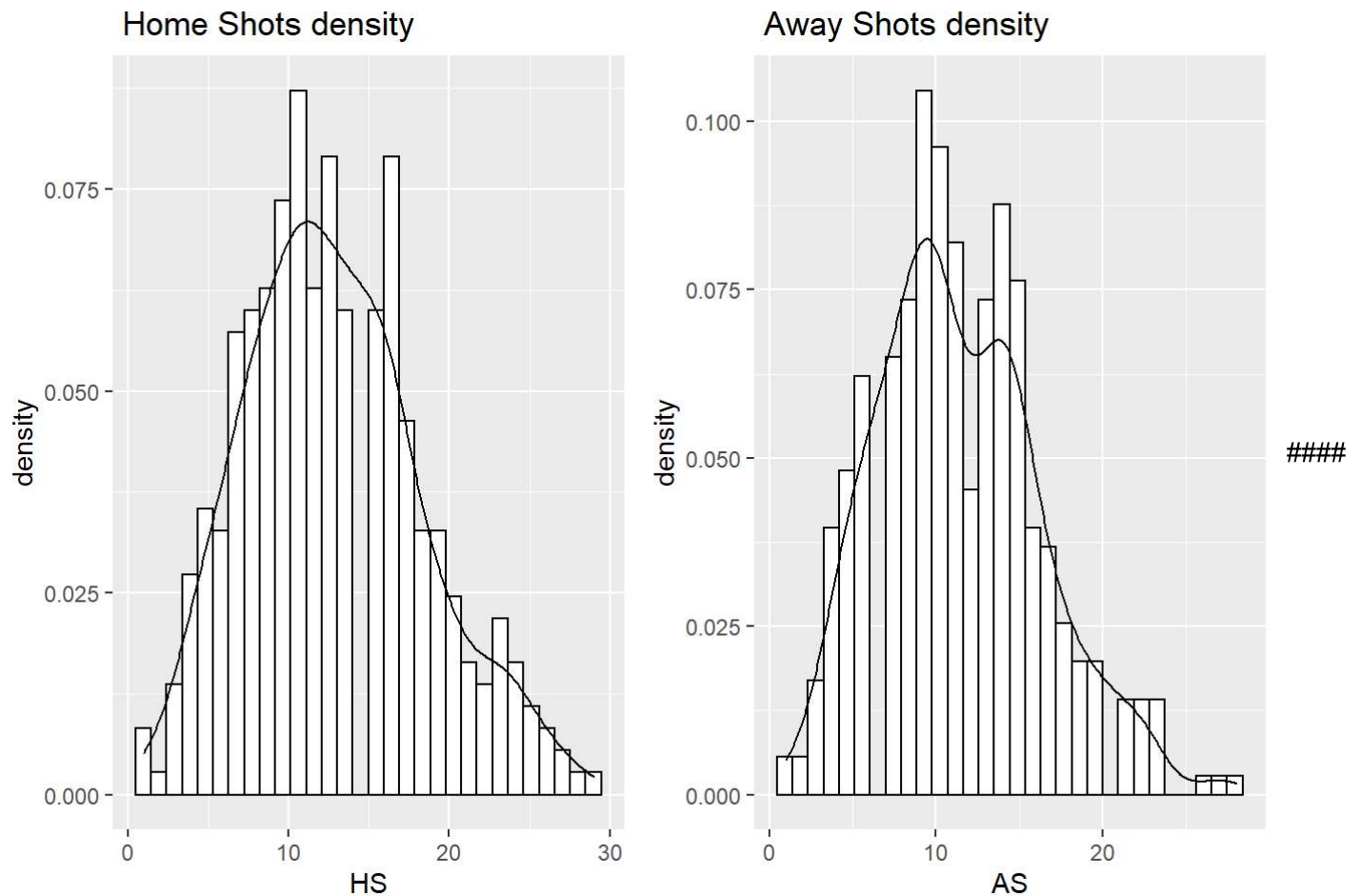
supporters put in and help push the billion dollar betting market. This acts as an incentive for

both parties(Players and Supporters).

Analysis

Shots Analysis:

```
plot1=ggplot(df, aes(x = HS)) +  
  geom_histogram(aes(y = ..density..),  
                 colour = 1, fill = "white",bins=30)+  
  ggtitle(" Home Shots density") +  
  geom_density()  
  
plot2=ggplot(df, aes(x = AS)) +  
  geom_histogram(aes(y = ..density..),  
                 colour = 1, fill = "white",bins=30)+  
  geom_density()+  
  ggtitle(" Away Shots density")  
  
grid.arrange(plot1,plot2,ncol=2,nrow=1)
```



A comparison of home team shots against the away team shots. Useful indicator for analysing performance of ##### home teams against away teams. Every Team on an high has produced 10 shots on the goal.

Here i further plan to breakdown the entire HS ans AS into further components:

- * home team shots on target vs away team shots on target
- * Home team hit woodwork vs away team hit woodwork
- * Home Team corners vs Away Team corners

Goal scoring abality:

Goal scored from corners

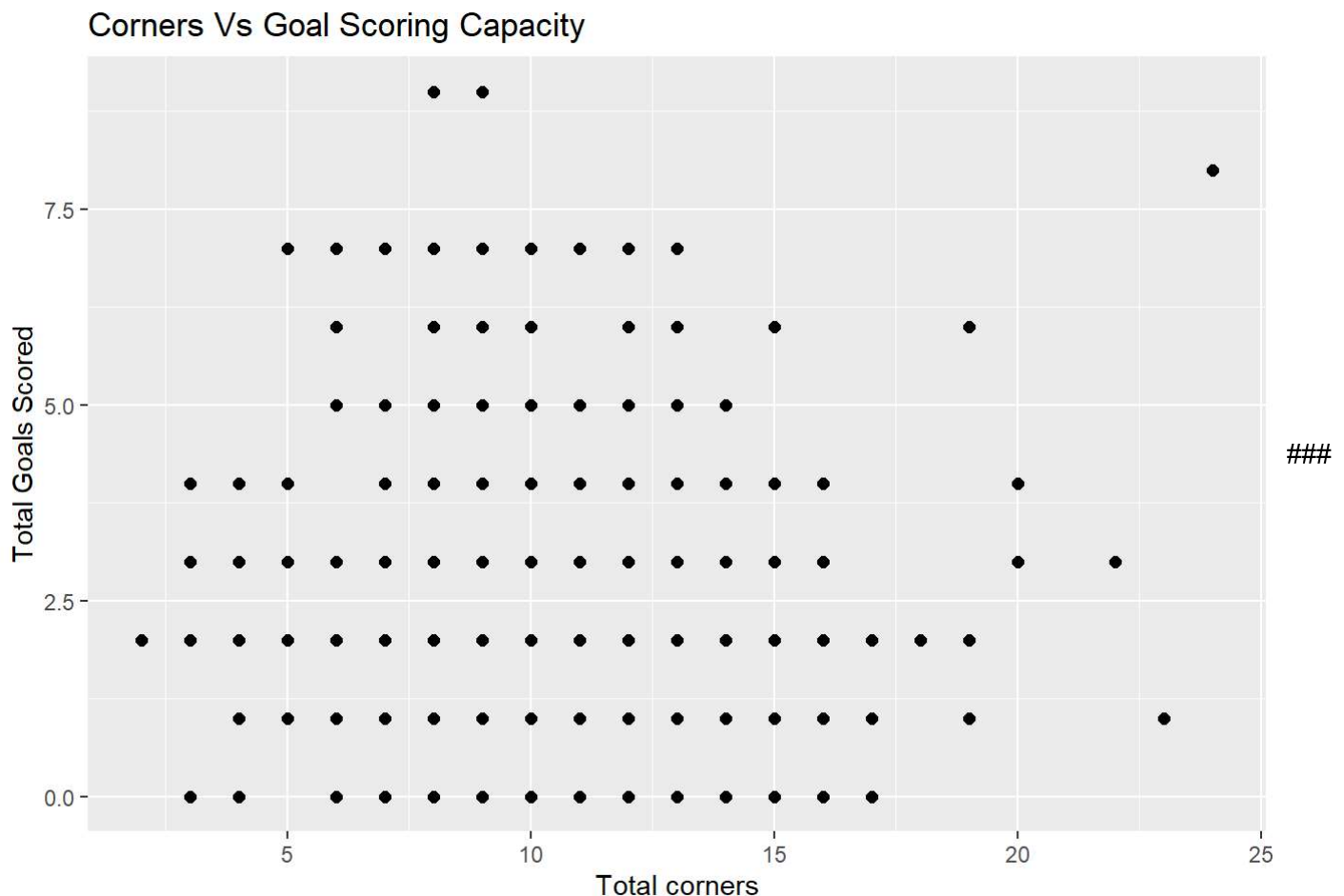
```

for(i in 1:nrow(df)){
  df$tc[i]=df$HC[i]+df$AC[i]
}

for(i in 1:nrow(df)){
  df$tg[i]=df$FTHG[i]+df$FTAG[i]
}

plot3=ggplot(df,aes(x=tc,y=tg,width=0.4))+
  geom_point(size=2)+
  ggtitle("Corners Vs Goal Scoring Capacity")+
  xlab("Total corners")+ylab("Total Goals Scored")
plot3

```



Four Way Analysis: ##### * Teams on the lower left corner receive less corners and invariably score less goals (An educated guess: Teams battling relegation) ##### * Teams on the upper right corner receive less corners but score a lot of goals (An educated guess: Teams with undoubtedly best counterattacking capacity) ##### * Teams on the lower right corner receive a lot of corners but don't score a lot (An educated guess: Mid table teams with inconsistent strikers) ##### * Teams on the upper right corner receive a lot of corners and invariably score a lot of goals (An educated guess: League Leaders/ High performance teams)

Situation analysis:

```

## Win by Home team
d=nrow(df%>%
  group_by(HomeTeam)%>%
  filter(HTR=="H",FTR=="H"))

## dominance by away team
e=nrow(df%>%
  group_by(HomeTeam)%>%
  filter(HTR=="A",FTR=="A"))
## bottled by home team
f=nrow(df%>%
  group_by(HomeTeam)%>%
  filter(HTR=="H",FTR=="A"))
## comeback by home team
g=nrow(df%>%
  group_by(HomeTeam)%>%
  filter(HTR=="A",FTR=="H"))

# draw matches
h=nrow(df%>%
  group_by(HomeTeam)%>%
  filter(FTR=="D"))

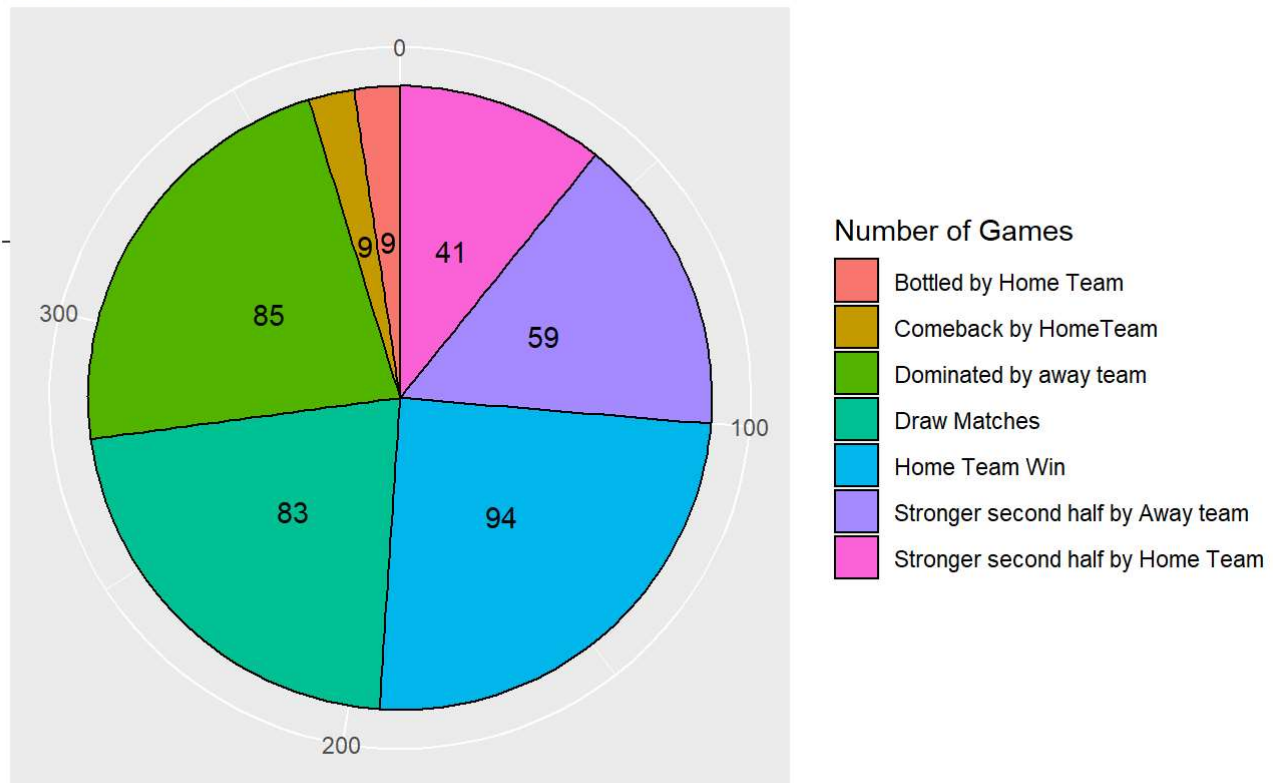
## Stronger second half by Home team
i=nrow(df%>%
  group_by(HomeTeam)%>%
  filter(HTR=="D",FTR=="H"))

## Stronger second half by away team

j=nrow(df%>%
  group_by(HomeTeam)%>%
  filter(HTR=="D",FTR=="A"))
analysis=data.frame(x=c("Home Team Win","Dominated by away team",
                        "Bottled by Home Team","Comeback by HomeTeam",
                        "Draw Matches","Stronger second half by Home Team",
                        "Stronger second half by Away team"),y=c(d,e,f,g,h,i,j))
ggplot(analysis, aes(x = "", y = y, fill = x)) +
  geom_col(color="black")+
  geom_text(aes(label=y),
            position=position_stack(vjust=0.5))+
  coord_polar(theta = "y")+
  ggtitle("Situational Analysis of 380 games")+
  xlab("")+ylab("")+
  guides(fill=guide_legend(title="Number of Games"))

```

Situational Analysis of 380 games



Team responses to different situations:

- * Bottled by Home Team: Home Team lost from a winning position(9- A low number. Home advantage is definitely playing a huge role)
- * Comeback by Home Team: Homeback comebacks from a losing position(9- A low number. Comebacks are usually tough even with home advantage)
- * Dominated by away team(Teams like Liverpool,manchester city do have outstanding away performances)
- * Draw Matches(Parity)
- * Home Team Win(> Away team wins: Clear reasons helping us understand home ground advantage)

Overall performance analysis:(team wise)

```

b=df %>%
  group_by(HomeTeam)%>%
  summarise(tgs=sum(FTHG),tgc=sum(FTAG))

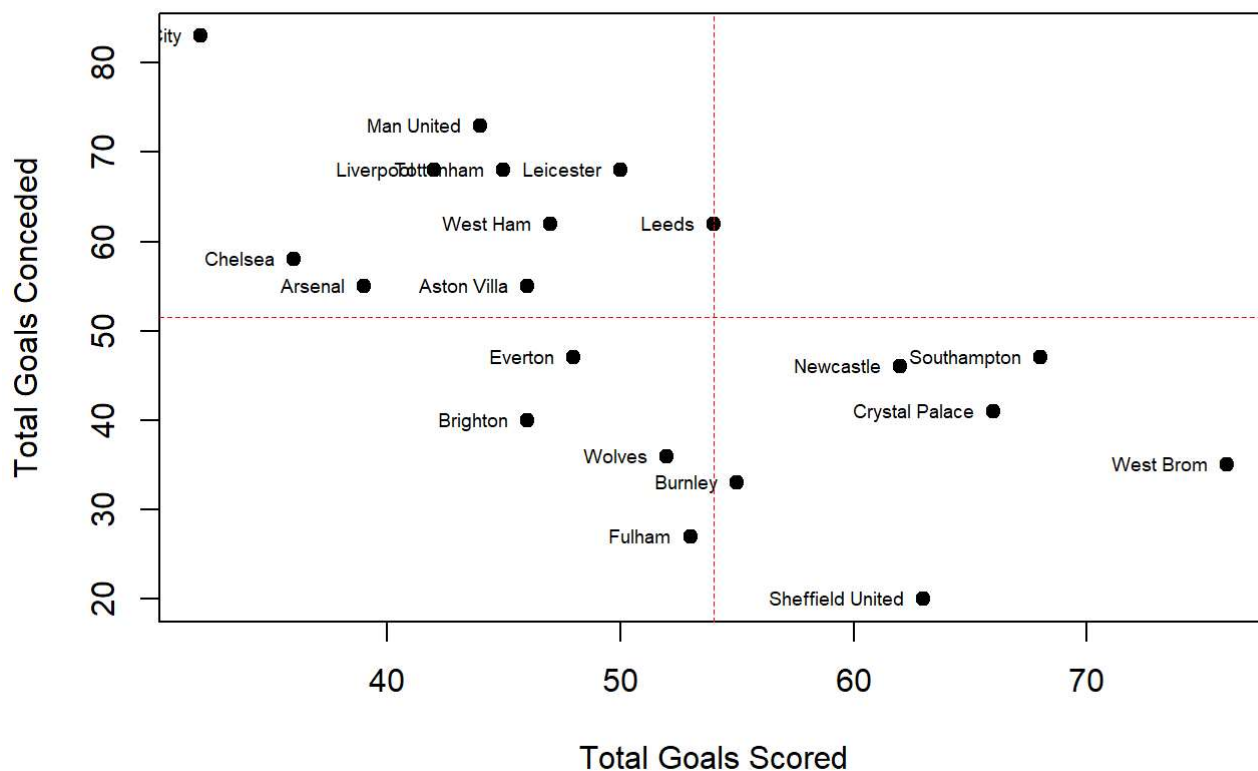
c=df %>%
  group_by(AwayTeam)%>%
  summarise(tgs=sum(FTAG),tgc=sum(FTHG))

Goals_tally=data.frame(Team=c(b$HomeTeam),
                        Total_Goal_Scored=c(b$tgs+c$tgs),
                        Total_goal_Conceded=c(b$tgc+c$tgc))

plot(Goals_tally$Total_goal_Conceded,Goals_tally$Total_Goal_Scored,pch=19,xlab="Total Goals Scored",
     ylab="Total Goals Conceded",main="Team Wise performance analysis")
text(Goals_tally$Total_goal_Conceded,Goals_tally$Total_Goal_Scored,Goals_tally$Team,cex=0.6,pos=
2)
grid(2,2,
     col="red",
     lty="dashed",
     lwd=0.4)

```

Team Wise performance analysis



ManchesterCity (POsitively skewed goal difference-League Winners)

Brighton(Goal Difference is almost 0.A good representation of the mid table teams)

West Brom(Goal Difference is negatively skewed.Relegated from the League)

Conclusion:

Through The various visual analytics i have tried to shown the inmatch, situation wise and overall performance of all the teams participating. This kinds of visualization helps customers chose their betting preferences and maximize profits.However there will always be outliers and ambiguity when it comes to sports.Understanding the nature of the game and analysis of previous games helps us mitigate the risks undertaken over the entire season.