IPL Analysis

Jahnvi Patel

2024-03-08

# Loaded required libraries  
library(tidyverse)

library(dplyr)  
  
# Read the dataset  
ipl\_data <- read.csv("IPL Dataset.csv")  
  
# Displayed the first few rows of the dataset  
head(ipl\_data)

## id city date Year Season player\_of\_match  
## 1 7953 Mumbai 5/27/2018 2018 IPL-2018 SR Watson  
## 2 7952 Kolkata 5/25/2018 2018 IPL-2018 Rashid Khan  
## 3 7951 Kolkata 5/23/2018 2018 IPL-2018 AD Russell  
## 4 7950 Mumbai 5/22/2018 2018 IPL-2018 F du Plessis  
## 5 7948 Delhi 5/20/2018 2018 IPL-2018 A Mishra  
## 6 7949 Pune 5/20/2018 2018 IPL-2018 L Ngidi  
## venue team1  
## 1 Wankhede Stadium Sunrisers Hyderabad  
## 2 Eden Gardens Sunrisers Hyderabad  
## 3 Eden Gardens Kolkata Knight Riders  
## 4 Wankhede Stadium Sunrisers Hyderabad  
## 5 Feroz Shah Kotla Delhi Daredevils  
## 6 Maharashtra Cricket Association Stadium Kings XI Punjab  
## team2 toss\_winner toss\_decision result  
## 1 Chennai Super Kings Chennai Super Kings field normal  
## 2 Kolkata Knight Riders Kolkata Knight Riders field normal  
## 3 Rajasthan Royals Rajasthan Royals field normal  
## 4 Chennai Super Kings Chennai Super Kings field normal  
## 5 Mumbai Indians Delhi Daredevils bat normal  
## 6 Chennai Super Kings Chennai Super Kings field normal  
## winner win\_by\_runs win\_by\_wickets umpire1  
## 1 Chennai Super Kings 0 8 Marais Erasmus  
## 2 Sunrisers Hyderabad 14 0 Nitin Menon  
## 3 Kolkata Knight Riders 25 0 Nitin Menon  
## 4 Chennai Super Kings 0 2 Marais Erasmus  
## 5 Delhi Daredevils 11 0 Kumar Dharmasena  
## 6 Chennai Super Kings 0 5 Nitin Menon  
## umpire2  
## 1 S Ravi  
## 2 Kumar Dharmasena  
## 3 Anil Chaudhary  
## 4 C Shamshuddin  
## 5 O Nandan  
## 6 Yeshwant Barde

# Summary statistics  
summary(ipl\_data)

## id city date Year   
## Min. : 1.0 Length:696 Length:696 Min. :2008   
## 1st Qu.: 174.8 Class :character Class :character 1st Qu.:2010   
## Median : 348.5 Mode :character Mode :character Median :2013   
## Mean : 974.1 Mean :2013   
## 3rd Qu.: 522.2 3rd Qu.:2016   
## Max. :7953.0 Max. :2018   
## Season player\_of\_match venue team1   
## Length:696 Length:696 Length:696 Length:696   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## team2 toss\_winner toss\_decision result   
## Length:696 Length:696 Length:696 Length:696   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## winner win\_by\_runs win\_by\_wickets umpire1   
## Length:696 Min. : 0.00 Min. : 0.000 Length:696   
## Class :character 1st Qu.: 0.00 1st Qu.: 0.000 Class :character   
## Mode :character Median : 0.00 Median : 3.000 Mode :character   
## Mean : 13.47 Mean : 3.349   
## 3rd Qu.: 19.00 3rd Qu.: 6.000   
## Max. :146.00 Max. :10.000   
## umpire2   
## Length:696   
## Class :character   
## Mode :character   
##   
##   
##

colnames(ipl\_data)

## [1] "id" "city" "date" "Year"   
## [5] "Season" "player\_of\_match" "venue" "team1"   
## [9] "team2" "toss\_winner" "toss\_decision" "result"   
## [13] "winner" "win\_by\_runs" "win\_by\_wickets" "umpire1"   
## [17] "umpire2"

str(ipl\_data)

## 'data.frame': 696 obs. of 17 variables:  
## $ id : int 7953 7952 7951 7950 7948 7949 7946 7947 7945 7944 ...  
## $ city : chr "Mumbai" "Kolkata" "Kolkata" "Mumbai" ...  
## $ date : chr "5/27/2018" "5/25/2018" "5/23/2018" "5/22/2018" ...  
## $ Year : int 2018 2018 2018 2018 2018 2018 2018 2018 2018 2018 ...  
## $ Season : chr "IPL-2018" "IPL-2018" "IPL-2018" "IPL-2018" ...  
## $ player\_of\_match: chr "SR Watson" "Rashid Khan" "AD Russell" "F du Plessis" ...  
## $ venue : chr "Wankhede Stadium" "Eden Gardens" "Eden Gardens" "Wankhede Stadium" ...  
## $ team1 : chr "Sunrisers Hyderabad" "Sunrisers Hyderabad" "Kolkata Knight Riders" "Sunrisers Hyderabad" ...  
## $ team2 : chr "Chennai Super Kings" "Kolkata Knight Riders" "Rajasthan Royals" "Chennai Super Kings" ...  
## $ toss\_winner : chr "Chennai Super Kings" "Kolkata Knight Riders" "Rajasthan Royals" "Chennai Super Kings" ...  
## $ toss\_decision : chr "field" "field" "field" "field" ...  
## $ result : chr "normal" "normal" "normal" "normal" ...  
## $ winner : chr "Chennai Super Kings" "Sunrisers Hyderabad" "Kolkata Knight Riders" "Chennai Super Kings" ...  
## $ win\_by\_runs : int 0 14 25 0 11 0 30 0 34 14 ...  
## $ win\_by\_wickets : int 8 0 0 2 0 5 0 5 0 0 ...  
## $ umpire1 : chr "Marais Erasmus" "Nitin Menon" "Nitin Menon" "Marais Erasmus" ...  
## $ umpire2 : chr "S Ravi" "Kumar Dharmasena" "Anil Chaudhary" "C Shamshuddin" ...

#Number of Matches Played by Each Team  
team\_matches <- ipl\_data %>%  
 select(team1 = team1, team2 = team2) %>%

# Combine team1 and team2 into a single column using the gather function  
 tidyr::gather(key = "team", value = "team\_name", team1, team2) %>%  
 group\_by(team\_name) %>%  
 summarize(total\_matches = n())  
  
# result  
print(team\_matches)

## # A tibble: 13 × 2  
## team\_name total\_matches  
## <chr> <int>  
## 1 Chennai Super Kings 147  
## 2 Deccan Chargers 75  
## 3 Delhi Daredevils 161  
## 4 Gujarat Lions 30  
## 5 Kings XI Punjab 162  
## 6 Kochi Tuskers Kerala 14  
## 7 Kolkata Knight Riders 164  
## 8 Mumbai Indians 171  
## 9 Pune Warriors 46  
## 10 Rajasthan Royals 133  
## 11 Rising Pune Supergiant 30  
## 12 Royal Challengers Bangalore 166  
## 13 Sunrisers Hyderabad 93

# Count the total wins for each team  
total\_wins <- ipl\_data %>%  
 group\_by(winner) %>%  
 summarize(total\_wins = n())

# Checked the data types of team\_matches  
str(team\_matches)

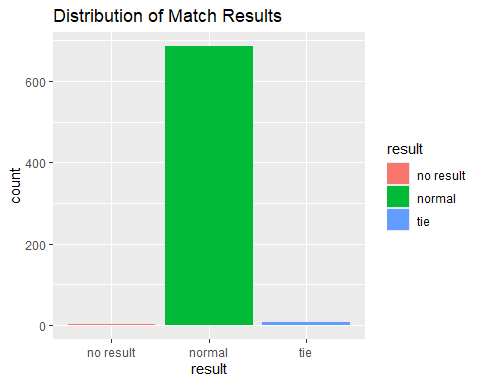
## tibble [13 × 2] (S3: tbl\_df/tbl/data.frame)  
## $ team\_name : chr [1:13] "Chennai Super Kings" "Deccan Chargers" "Delhi Daredevils" "Gujarat Lions" ...  
## $ total\_matches: int [1:13] 147 75 161 30 162 14 164 171 46 133 ...

# Converted total\_matches to numeric  
team\_matches$total\_matches <- as.numeric(team\_matches$total\_matches)

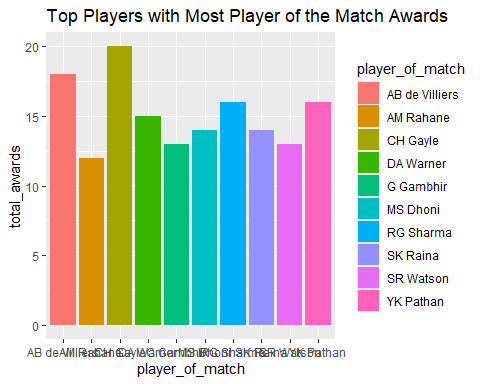
# Calculated win percentage  
team\_summary <- left\_join(total\_wins, team\_matches, by = c("winner" = "team\_name")) %>%  
 mutate(win\_percentage = ifelse(total\_matches > 0, total\_wins / total\_matches \* 100, 0))  
  
# Result  
print(team\_summary)

## # A tibble: 14 × 4  
## winner total\_wins total\_matches win\_percentage  
## <chr> <int> <dbl> <dbl>  
## 1 Chennai Super Kings 90 147 61.2  
## 2 Deccan Chargers 29 75 38.7  
## 3 Delhi Daredevils 67 161 41.6  
## 4 Gujarat Lions 13 30 43.3  
## 5 Kings XI Punjab 76 162 46.9  
## 6 Kochi Tuskers Kerala 6 14 42.9  
## 7 Kolkata Knight Riders 86 164 52.4  
## 8 Mumbai Indians 98 171 57.3  
## 9 No Result 3 NA NA   
## 10 Pune Warriors 12 46 26.1  
## 11 Rajasthan Royals 70 133 52.6  
## 12 Rising Pune Supergiant 15 30 50   
## 13 Royal Challengers Bangalore 79 166 47.6  
## 14 Sunrisers Hyderabad 52 93 55.9

#Distribution of Results  
result\_distribution <- ipl\_data %>%  
 group\_by(result) %>%  
 summarize(count = n())  
  
# Visualized the distribution  
ggplot(result\_distribution, aes(x = result, y = count, fill = result)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Distribution of Match Results")



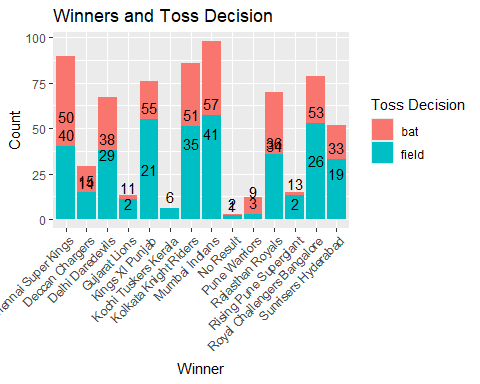
#Top Players with Most Player of the Match Awards  
top\_players <- ipl\_data %>%  
 group\_by(player\_of\_match) %>%  
 summarize(total\_awards = n()) %>%  
 arrange(desc(total\_awards)) %>%  
 head(10)  
  
# Visualized the top players  
ggplot(top\_players, aes(x = player\_of\_match, y = total\_awards, fill = player\_of\_match)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Top Players with Most Player of the Match Awards")



# Counted the occurrences of each combination of winner and toss decision  
win\_toss\_count <- ipl\_data %>%  
 group\_by(winner, toss\_decision) %>%  
 summarize(count = n()) %>%  
 arrange(winner)

## `summarise()` has grouped output by 'winner'. You can override using the  
## `.groups` argument.

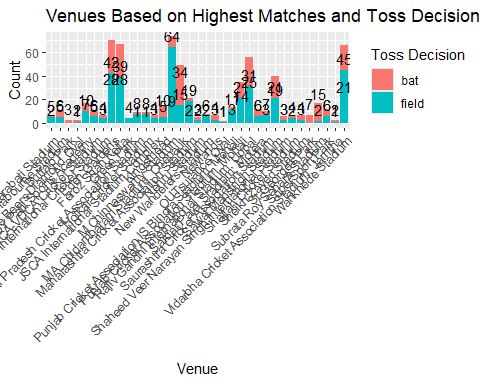
# Created a stacked bar chart  
ggplot(win\_toss\_count, aes(x = winner, y = count, fill = toss\_decision)) +  
 geom\_bar(stat = "identity") +  
 geom\_text(data = win\_toss\_count, aes(label = count), vjust = -0.5) +  
 labs(title = "Winners and Toss Decision",  
 x = "Winner",  
 y = "Count",  
 fill = "Toss Decision") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# Counted the occurrences of each combination of venue and toss decision  
venue\_toss\_count <- ipl\_data %>%  
 group\_by(venue, toss\_decision) %>%  
 summarize(count = n()) %>%  
 arrange(venue)

## `summarise()` has grouped output by 'venue'. You can override using the  
## `.groups` argument.

# Identified the toss decision with the maximum count for each venue  
max\_toss\_decision <- venue\_toss\_count %>%  
 group\_by(venue) %>%  
 filter(count == max(count))  
  
# Created a stacked bar chart  
ggplot(venue\_toss\_count, aes(x = venue, y = count, fill = toss\_decision)) +  
 geom\_bar(stat = "identity") +  
 geom\_text(data = venue\_toss\_count, aes(label = count), vjust = -0.5) +  
 labs(title = "Venues Based on Highest Matches and Toss Decision",  
 x = "Venue",  
 y = "Count",  
 fill = "Toss Decision") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# Read the dataset  
ipl\_season\_winners <- read.csv("Winner Data.csv")

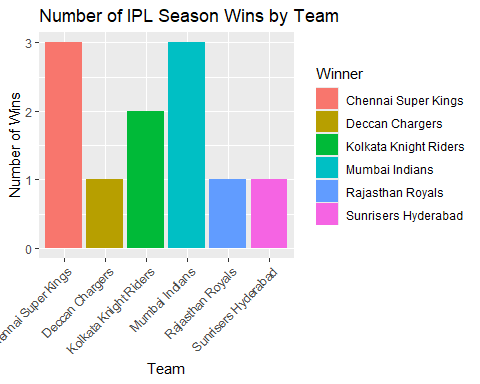
# Displayed the first few rows of the dataset  
head(ipl\_season\_winners)

## Season Winner Runner.Up  
## 1 IPL-2018 Chennai Super Kings Sunrisers Hyderabad  
## 2 IPL-2017 Mumbai Indians Rising Pune Supergiants  
## 3 IPL-2016 Sunrisers Hyderabad Royal Challengers Bangalore  
## 4 IPL-2015 Mumbai Indians Chennai Super Kings  
## 5 IPL-2014 Kolkata Knight Riders Kings XI Punjab  
## 6 IPL-2013 Mumbai Indians Chennai Super Kings  
## Player.of.the.Match Player.of.the.Series  
## 1 Shane Watson Sunil Narine  
## 2 Krunal Pandya Ben Stokes  
## 3 Ben Cutting Virat Kohli  
## 4 Rohit Sharma Andre Russell  
## 5 Manish Pandey Glenn Maxwell  
## 6 Kieron Pollard Shane Watson

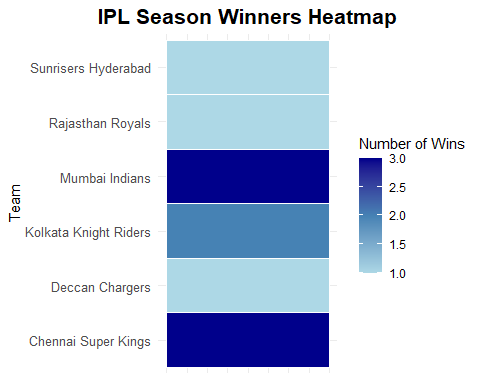
# Summary statistics  
summary(ipl\_season\_winners)

## Season Winner Runner.Up Player.of.the.Match  
## Length:11 Length:11 Length:11 Length:11   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
## Player.of.the.Series  
## Length:11   
## Class :character   
## Mode :character

#Frequency of Season Winners  
# Count the number of times each team won the IPL season  
winner\_counts <- ipl\_season\_winners %>%  
 group\_by(Winner) %>%  
 summarize(win\_count = n())  
  
# Created a bar chart for visualization  
ggplot(winner\_counts, aes(x = Winner, y = win\_count, fill = Winner)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Number of IPL Season Wins by Team",  
 x = "Team",  
 y = "Number of Wins") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



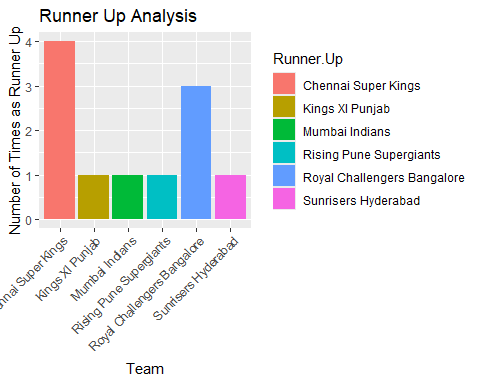
# Counted the number of times each team won the IPL season  
winner\_counts <- ipl\_season\_winners %>%  
 group\_by(Winner) %>%  
 summarize(win\_count = n())  
  
# Created a heatmap  
ggplot(winner\_counts, aes(x = 1, y = Winner, fill = win\_count)) +  
 geom\_tile(color = "white") +  
 scale\_fill\_gradientn(colors = c("lightblue", "steelblue", "darkblue"), na.value = "white") +  
 labs(title = "IPL Season Winners Heatmap",  
 x = NULL,  
 y = "Team",  
 fill = "Number of Wins") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_blank(), # Hide x-axis label  
 axis.title.x = element\_blank(), # Hide x-axis title  
 axis.text.y = element\_text(size = 10), # Adjust y-axis label size  
 plot.title = element\_text(hjust = 0.5, size = 16, face = "bold"))



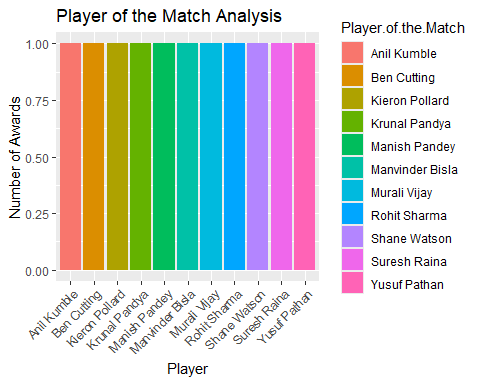
#Bar Chart for Runner Up  
# Check the column names  
colnames(ipl\_season\_winners)

## [1] "Season" "Winner" "Runner.Up"   
## [4] "Player.of.the.Match" "Player.of.the.Series"

# Replaced spaces in column names with underscores  
colnames(ipl\_season\_winners) <- make.names(colnames(ipl\_season\_winners))  
  
# Counted the occurrences of each team as Runner Up  
runner\_up\_counts <- ipl\_season\_winners %>%  
 group\_by(Runner.Up) %>%  
 summarize(count = n())  
  
# Created a bar chart  
ggplot(runner\_up\_counts, aes(x = Runner.Up, y = count, fill = Runner.Up)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Runner Up Analysis",  
 x = "Team",  
 y = "Number of Times as Runner Up") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



#Bar Chart for Player of the Match  
# Replaced spaces in column names with underscores  
colnames(ipl\_season\_winners) <- make.names(colnames(ipl\_season\_winners))  
  
# Counted the occurrences of each player as Player of the Match  
player\_match\_counts <- ipl\_season\_winners %>%  
 group\_by(Player.of.the.Match) %>%  
 summarize(count = n())  
  
# Created a bar chart  
ggplot(player\_match\_counts, aes(x = Player.of.the.Match, y = count, fill = Player.of.the.Match)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Player of the Match Analysis",  
 x = "Player",  
 y = "Number of Awards") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



#Bar Chart for Player of the Series  
# Replaced spaces in column names with underscores  
colnames(ipl\_season\_winners) <- make.names(colnames(ipl\_season\_winners))  
  
# Counted the occurrences of each player as Player of the Series  
player\_series\_counts <- ipl\_season\_winners %>%  
 group\_by(Player.of.the.Series) %>%  
 summarize(count = n())  
  
# Created a bar chart  
ggplot(player\_series\_counts, aes(x = Player.of.the.Series, y = count, fill = Player.of.the.Series)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Player of the Series Analysis",  
 x = "Player",  
 y = "Number of Awards") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))

