GUJARAT TEAM ANALYSIS IN IPL 2022

Jeer shri karthick s

22CSEG12

1st M.sc Data Analytics

ASSUMPTION:

- In this dataset,we can assume that the date column belongs to the team when they perform their matches.
- So we can correlate that date column and then team columns respectively.
- I'm going to correlate the toss winner and toss decision columns which belongs to which team won their toss and which particular team they choose to play, that is either bat or field.
- Next we are going to see the first inning score which is scored by the opponent team which has not won the toss team and their performance according to the toss winner columns.
- And then we are going to analyze that first inning score which is scored by the opponent team and their wickets with respect to second innings score which is scored by the toss winner team and their wickets.
- I can also analyze that match winner column based on their innings score and we can relate to the win by column. This gives the information about which performance makes the team win which wickets or runs.
- Then we can see which player performs well by the player of the match column based on the high score column.
- I can assume that the Gujarat team in the team1 and team2 columns performs well due to their batting capability.

HYPOTHESIS:

- I’m going to analyze the first innings score using histogram with their batting performance and then by using histogram we can analyze the second innings score.
- The comparison of this first innings seems to be higher.
- And then we can analyze that particular team which belongs to team1 column with the comparison of match winners using histogram.
- By this comparison we are going to choose the toss decision which is batting or fielding.
- I can assume that Gujarat won many times due to their batting performance.
- But my assumption was wrong because Gujarat has won many times due to their feilding performance.
- From this we can see which team won the toss and they can choose that particular tram has high chances to win that match.

We can analyze this by using boxplot. and then we are going to analyze the first innings score and second innings score due to their team performance we can see which is highly correlated by using scatterplot.

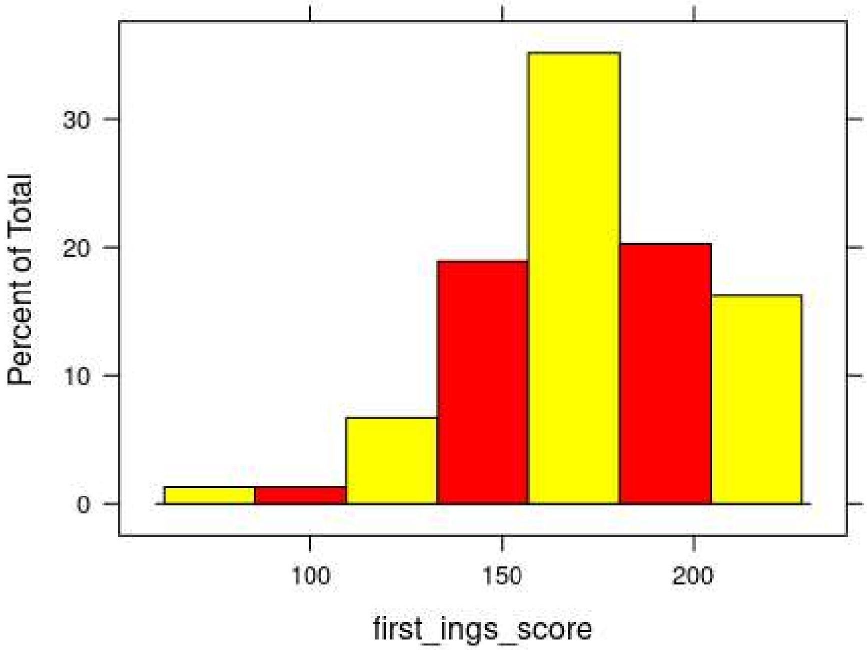
```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(lattice)
library(ggplot2)
library(plyr)

##
-----------------------------------------------------------------------------
-

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first,
then dplyr:
## library(plyr); library(dplyr)

##
-----------------------------------------------------------------------------
-

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

library(readr)

df=read_csv("ipl dataset.csv")

## Rows: 74 Columns: 20

## ── Column specification
────────────────────────────────────────────────────────

## Delimiter: ","
## chr (13): date, venue, team1, team2, stage, toss_winner, toss_decision,
matc...
## dbl  (7): match_id, first_ings_score, first_ings_wkts, second_ings_score,
se...
##
```

```
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

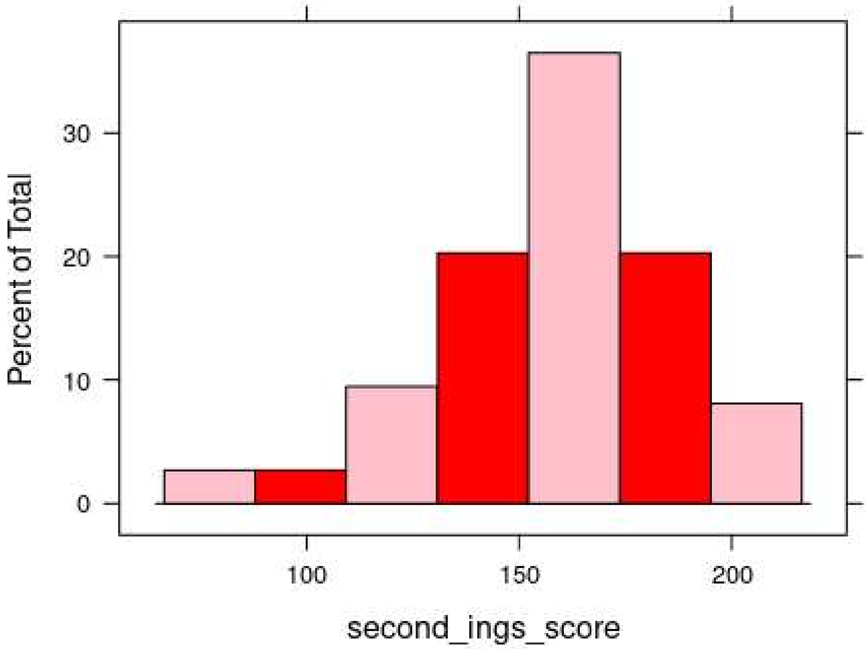
```
summary(df)
```

```
##     match_id         date              venue              team1          
##  Min.   : 1.00   Length:74          Length:74          Length:74         
##  1st Qu.:19.25   Class :character   Class :character   Class :character  
##  Median :37.50   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :37.50                                                           
##  3rd Qu.:55.75                                                           
##  Max.   :74.00                                                           
##     team2              stage            toss_winner        toss_decision     
##  Length:74          Length:74          Length:74          Length:74         
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##  first_ings_score first_ings_wkts  second_ings_score second_ings_wkts
##  Min.   : 68.0    Min.   : 0.000   Min.   : 72.0     Min.   : 1.000  
##  1st Qu.:154.2    1st Qu.: 5.000   1st Qu.:142.8     1st Qu.: 4.000  
##  Median :169.5    Median : 6.000   Median :160.0     Median : 6.000  
##  Mean   :171.1    Mean   : 6.135   Mean   :158.5     Mean   : 6.176  
##  3rd Qu.:192.8    3rd Qu.: 8.000   3rd Qu.:176.0     3rd Qu.: 8.000  
##  Max.   :222.0    Max.   :10.000   Max.   :211.0     Max.   :10.000  
##  match_winner          won_by             margin      player_of_the_match
##  Length:74          Length:74          Min.   : 2.00   Length:74          
##  Class :character   Class :character   1st Qu.: 5.25   Class :character   
##  Mode  :character   Mode  :character   Median : 8.00   Mode  :character   
##                                        Mean   :16.97                      
##                                        3rd Qu.:18.00                      
##                                        Max.   :91.00                      
##   top_scorer          highscore      best_bowling      
best_bowling_figure
##  Length:74          Min.   : 28.00   Length:74          Length:74         
##  Class :character   1st Qu.: 57.00   Class :character   Class :character  
##  Mode  :character   Median : 68.00   Mode  :character   Mode  :character  
##                     Mean   : 71.72                                        
##                     3rd Qu.: 87.75                                        
##                     Max.   :140.00                                        
```

```
histogram(~first_ings_score,data=df,col=c("yellow","red"))
```



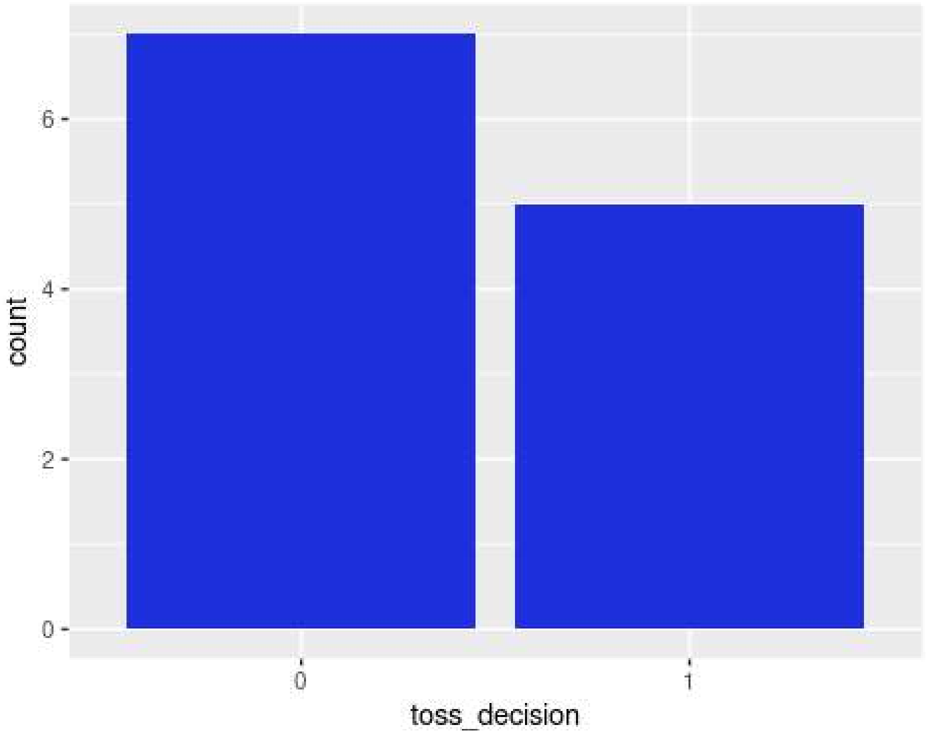
```
histogram(~second_ings_score,data=df,col=c("pink","red"))
```



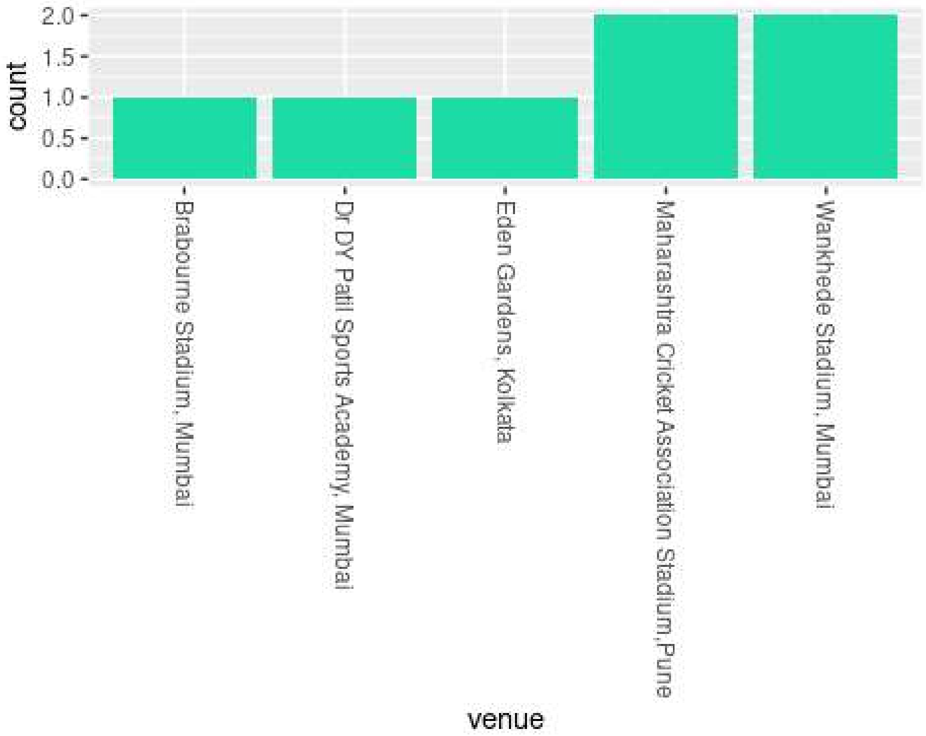
```
count(df,'match_winner')
```

```
##    match_winner freq
## 1      Banglore    9
## 2       Chennai    4
## 3         Delhi    7
## 4       Gujarat   12
## 5     Hyderabad    6
## 6       Kolkata    6
## 7       Lucknow    9
## 8        Mumbai    4
## 9        Punjab    7
## 10    Rajasthan   10
```

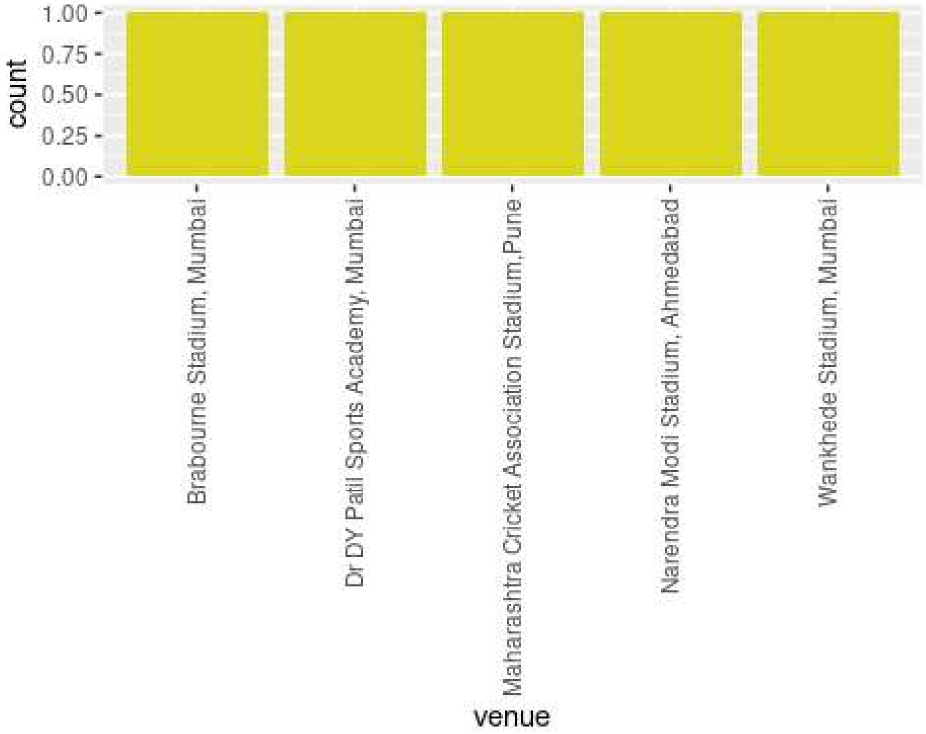
```
df2=subset(df,match_winner=='Gujarat')

df2$toss_decision[df2$toss_decision=='Field']=0
df2$toss_decision[df2$toss_decision=='Bat']=1

df2 %>% ggplot(aes(toss_decision))+ geom_bar(stat="Count",fill='#1d30db')
```



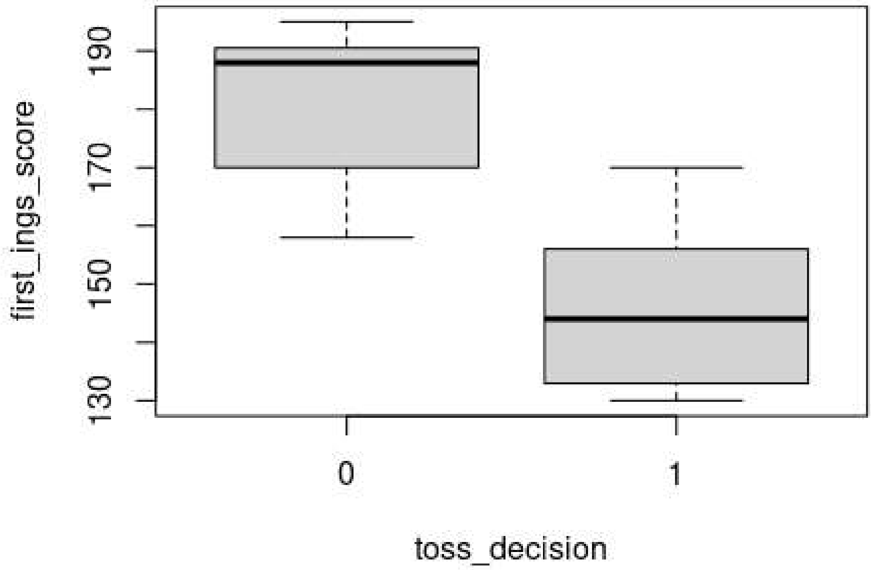
```
df3=subset(df2,toss_decision==0)
df4=subset(df2,toss_decision==1)

df3 %>% ggplot(aes(venue))+ geom_bar(stat="Count",fill='#1ddba5')+
scale_x_discrete(guide = guide_axis(angle = -90))
```



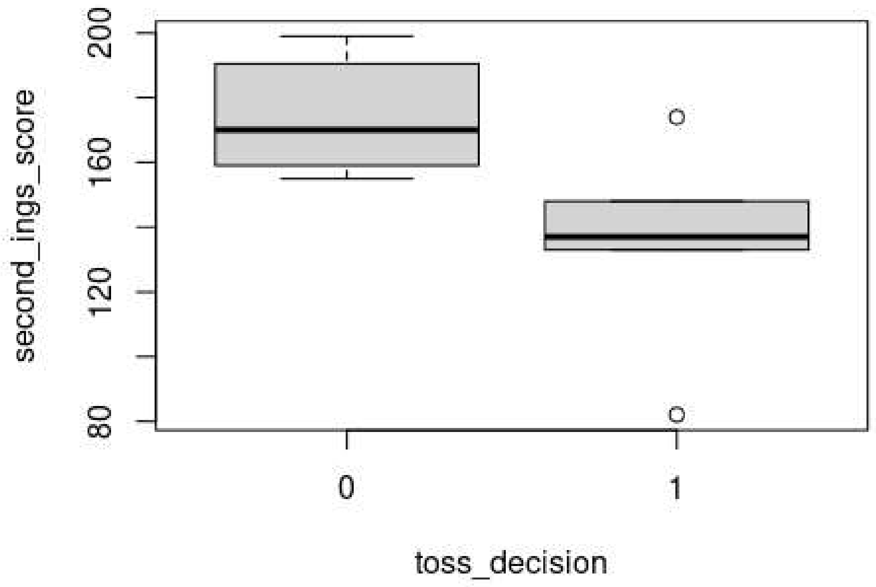
```
df4 %>% ggplot(aes(venue))+
geom_bar(stat="Count",fill='#dbd51d')+scale_x_discrete(guide =
guide_axis(angle = 90))
```



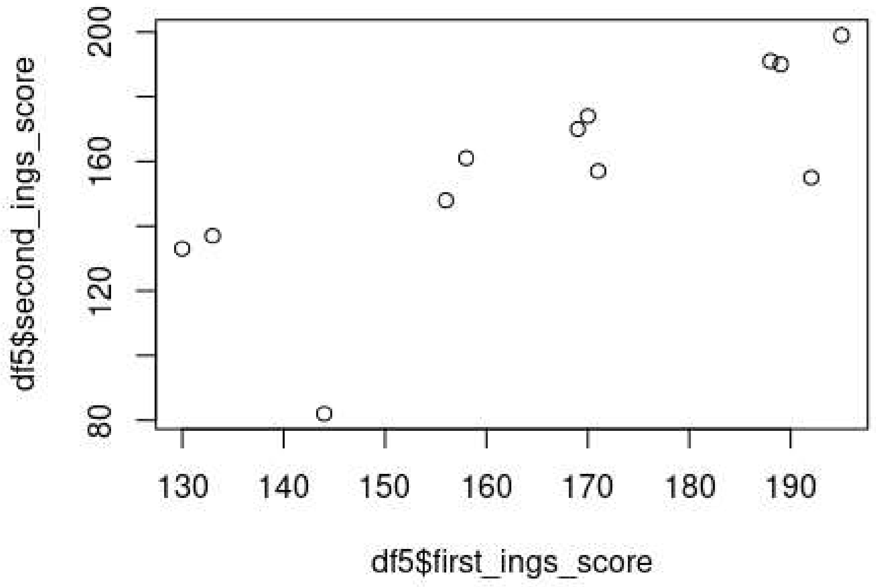
```
df5=rbind(df3,df4)
boxplot(first_ings_score~toss_decision,data=df5,fill='#db401d')
```



```
boxplot(second_ings_score~toss_decision,data=df5)
```



```
plot(x=df5$first_ings_score, y=df5$second_ings_score)
```



INFERENCE:

- BY this comparison of first innings and the second innings by using histogram the first innings score has more than the second innings.
- The first innings ranges from 0 to 15 from 200 above scores but the second innings ranges from 0 to 10 from the above 200 above scores.
- And then comparison of two innings seems that 150 above scores has the highest range by the other innings score.
- In boxplot we can compare the toss decision with the first innings score.
- It seems that field has maximum range value which is 180 and then bat has median range value which is 140 and then comparison of toss decision with second innings using boxplot field has minimum range which is 160 and then bat contains outliers which are extreme values are different from the dataset.
- At last we can analyze the first innings score and then second innings score using scatterplot i shows highly correlated in the range of 170.

INSIGHTS:

- By the comparison of team with toss decision the toss winner and then feilding chosen has high chances of winning that match.
- From this we can compare the Gujarat team in the team1 and team2 column. They can win the highest amount of time because they won the toss and then chose to field.