Machine learning

Sai Gautham Sabhavathu

1/30/2022

```r
#Reference :- https://www.premierleague.com/tables
data1 <- read.csv("C:/Users/gauth/OneDrive/Desktop/Premier League 2021-22 Points Table.csv")
data1
```

```
##    SNO          Squad MP  W  D L GF GA GD Pts   xG  xGA  xGD xGD.90
## 1    1 Manchester City 23 18  3 2 55 14 41  57 51.5 16.0 35.5   1.54
## 2    2       Liverpool 22 14  6 2 58 19 39  48 54.1 23.0 31.1   1.42
## 3    3         Chelsea 24 13  8 3 48 18 30  47 40.8 22.1 18.7   0.78
## 4    4   Manchester Utd 22 11  5 6 36 30  6  38 31.9 32.9 -1.0  -0.05
## 5    5        West Ham 23 11  4 8 41 31 10  37 33.4 29.5  3.8   0.17
## 6    6         Arsenal 21 11  3 7 33 25  8  36 30.9 28.8  2.2   0.10
## 7    7       Tottenham 20 11  3 6 26 24  2  36 31.6 21.4 10.2   0.51
## 8    8          Wolves 21 10  4 7 19 16  3  34 20.9 27.1 -6.2  -0.29
## 9    9        Brighton 22  6 12 4 23 23  0  30 25.7 26.9 -1.3  -0.06
## 10  10   Leicester City 20  7  5 8 34 37 -3  26 31.1 36.4 -5.3  -0.27
##             Leagues
## 1   Champions league
## 2   Champions league
## 3   Champions league
## 4   Champions league
## 5  EUFA Europa League
## 6          No league
## 7          No league
## 8          No league
## 9          Relegation
## 10         Relegation
```

```r
#The selected data is presented above.
#The table represents the points table of the English Premier league of the season.

mean(data1$W)
```

```
## [1] 11.2
```

```r
sd(data1$W)
```

```
## [1] 3.392803
```

```
# The above values represent descriptive statistics
#for a selection of quantitative  variables.

table(data1$Leagues)
```

```
##
##    Champions league EUFA Europa League        No league        Relegation
##                   4                 1                3                 2
```

```
str(data1)
```

```
## 'data.frame':    10 obs. of  15 variables:
##  $ SNO     : int  1 2 3 4 5 6 7 8 9 10
##  $ Squad   : chr  " Manchester City" " Liverpool" " Chelsea" " Manchester Utd" ...
##  $ MP      : int  23 22 24 22 23 21 20 21 22 20
##  $ W       : int  18 14 13 11 11 11 11 10 6 7
##  $ D       : int  3 6 8 5 4 3 3 4 12 5
##  $ L       : int  2 2 3 6 8 7 6 7 4 8
##  $ GF      : int  55 58 48 36 41 33 26 19 23 34
##  $ GA      : int  14 19 18 30 31 25 24 16 23 37
##  $ GD      : int  41 39 30 6 10 8 2 3 0 -3
##  $ Pts     : int  57 48 47 38 37 36 36 34 30 26
##  $ xG      : num  51.5 54.1 40.8 31.9 33.4 30.9 31.6 20.9 25.7 31.1
##  $ xGA     : num  16 23 22.1 32.9 29.5 28.8 21.4 27.1 26.9 36.4
##  $ xGD     : num  35.5 31.1 18.7 -1 3.8 2.2 10.2 -6.2 -1.3 -5.3
##  $ xGD.90  : num  1.54 1.42 0.78 -0.05 0.17 0.1 0.51 -0.29 -0.06 -0.27
##  $ Leagues : chr  "Champions league" "Champions league" "Champions league" "Champions league" ...
```
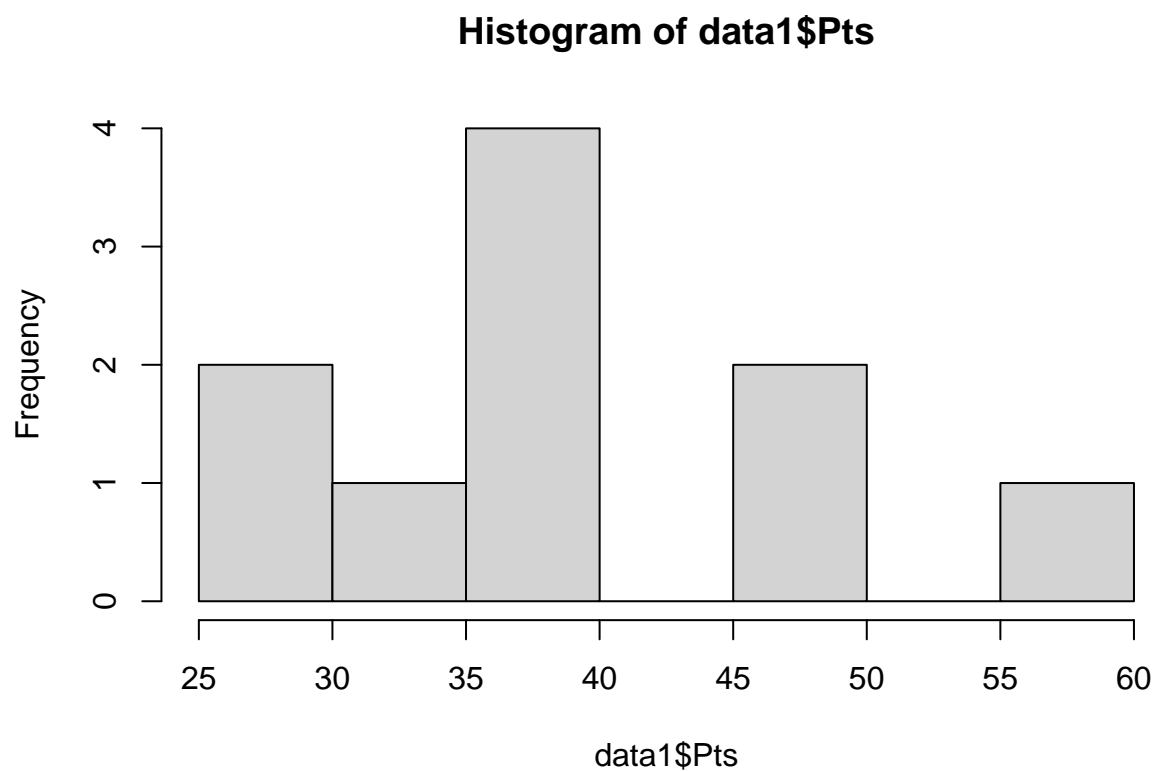
```
# The above values represent categorical
#descriptive analysis of the variables.

data1$W_Trasnformed = (data1$W - mean(data1$W)/sd(data1$W))
data1$W_Trasnformed
```

```
##  [1] 14.698895 10.698895  9.698895  7.698895  7.698895  7.698895  7.698895
##  [8]  6.698895  2.698895  3.698895
```

```
#Transformation of variables has been done above.

hist(data1$Pts)
```

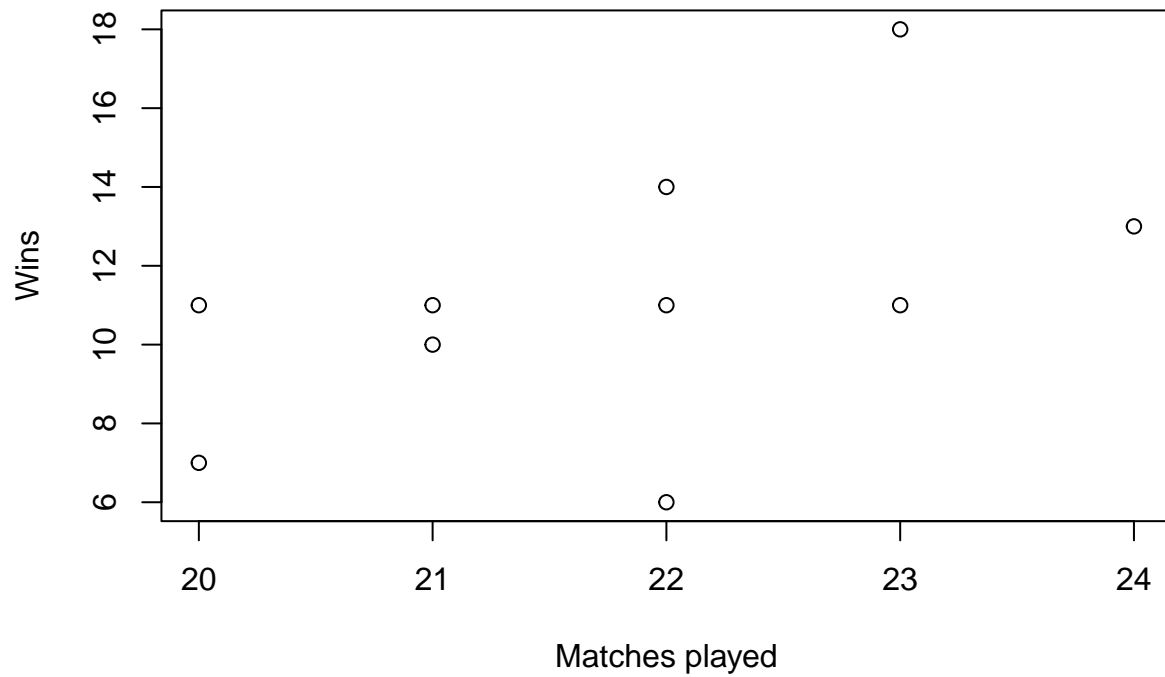## Histogram of data1$Pts


```
#The above graphical representation is a histogram.
#The selected variable is the points.

x <- data1$MP
y <- data1$W
plot(x,y, main = "Matches played and the number of wins", xlab = "Matches played",ylab = "Wins")
```

Matches played and the number of wins

Matches played

#The above graphical representation is a scatterplot.
#The selected variables are Matches played and Wins