title: “ml assignment 1” author: “nemin dholakia” date: “9/10/2021” output: html_document df print:paged

###Q1: Data source Our dataset **Overwise statistics without superover** has been extracted from **KAGGLE** , is a collection of database and datasets used by Machine Learning Community for the analysis of Machine Learning algorithms.

###Q2: Importing Overwise statistics without superover dataset to R Studio

```
setwd("C:/Users/User/Downloads/archive (1)")
OS_WS <- read.csv("Overwise statistics without super over.csv")
head(OS_WS)
```

```
##   match_id inning        batting_team               bowling_team over runs
## 1        1      1 Sunrisers Hyderabad Royal Challengers Bangalore    1    7
## 2        1      1 Sunrisers Hyderabad Royal Challengers Bangalore    2   16
## 3        1      1 Sunrisers Hyderabad Royal Challengers Bangalore    3    6
## 4        1      1 Sunrisers Hyderabad Royal Challengers Bangalore    4    4
## 5        1      1 Sunrisers Hyderabad Royal Challengers Bangalore    5    9
## 6        1      1 Sunrisers Hyderabad Royal Challengers Bangalore    6   17
##   wickets              winner
## 1       0 Sunrisers Hyderabad
## 2       1 Sunrisers Hyderabad
## 3       0 Sunrisers Hyderabad
## 4       0 Sunrisers Hyderabad
## 5       0 Sunrisers Hyderabad
## 6       0 Sunrisers Hyderabad
```

###Q3: Descriptive analysis of OWSW Dataset Inspecting dataset

```
summary(OS_WS)
```

```
##     match_id           inning    batting_team       bowling_team
##  Min.   :    1.0   Min.   :1.0   Length:30240       Length:30240
##  1st Qu.:  189.8   1st Qu.:1.0   Class :character   Class :character
##  Median :  378.5   Median :1.5   Mode  :character   Mode  :character
##  Mean   : 1792.2   Mean   :1.5
##  3rd Qu.:  567.2   3rd Qu.:2.0
##  Max.   :11415.0   Max.   :2.0
##       over            runs           wickets          winner
##  Min.   : 1.00   Min.   : 0.000   Min.   :0.0000   Length:30240
##  1st Qu.: 5.75   1st Qu.: 4.000   1st Qu.:0.0000   Class :character
##  Median :10.50   Median : 7.000   Median :0.0000   Mode  :character
##  Mean   :10.50   Mean   : 7.774   Mean   :0.2917
##  3rd Qu.:15.25   3rd Qu.:11.000   3rd Qu.:1.0000
##  Max.   :20.00   Max.   :37.000   Max.   :4.0000
```

Arithmetic Mean: AM of a set of observation is defined as their sum divided number of observations

```
#Average runs
mean(OS_WS$runs)
```

```
## [1] 7.773942
```

Median: is the middle value of the observation.

```
#Median of wickets
median(sort(OS_WS$wickets))
```

```
## [1] 0
```

Mode : refers to the value which occurs to the maximum frequency.

```
sort(table(OS_WS$winner),decreasing=TRUE)
```

```
##
##              Mumbai Indians         Chennai Super Kings
##                        4360                        4000
##       Kolkata Knight Riders Royal Challengers Bangalore
##                        3680                        3360
##             Kings XI Punjab              Delhi Capitals
##                        3280                        3080
##            Rajasthan Royals         Sunrisers Hyderabad
##                        3000                        2320
##             Deccan Chargers     Rising Pune Supergiants
##                        1160                         600
##               Gujarat Lions               Pune Warriors
##                         520                         480
##        Kochi Tuskers Kerala
##                         240                         160
```

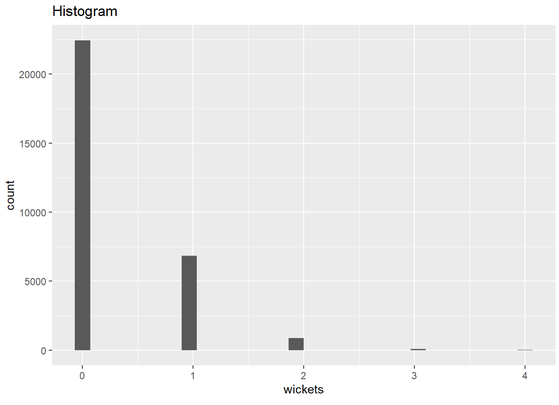
###Q4: Transformation

```
#Applying log transformation on Wicket column
run <- log10(OS_WS$runs)
##displaying first 10 values
run[1:10]
```

```
##  [1] 0.8450980 1.2041200 0.7781513 0.6020600 0.9542425 1.2304489 0.6989700
##  [8] 1.0413927 0.9542425 0.6020600
```

###Q5:Visualisation **Histogram Plot**

```
library(ggplot2)
ggplot(OS_WS,aes(x= wickets))+geom_histogram()+ggtitle('Histogram')
```



Scatter Plot

```
ggplot(OS_WS, aes(x= runs, y=wickets))+geom_point()+ggtitle('Scatter Plot')
```