new analysis

Ajinkya

2/25/2020

```
library(colorspace)
library(RColorBrewer)
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
data=read.csv('cincinnati.csv')
newdata=data


# Alter Date column
class(newdata$Date)
```

```
## [1] "factor"
```

```
newdata$Date1 <- as.Date(newdata$Date,format="%m/%d/%Y")
class(newdata$Date1)
```

```
## [1] "Date"
```

```
# create game differemce column
newdata$gamediff[1] <- 0
for (val in 2:length(newdata$Date1)){
  print(val)
 newdata$gamediff[val] <-  (abs(newdata$Date1[val-1]-(newdata$Date1[val])))

}
```

```
## [1] 2
## [1] 3
## [1] 4
## [1] 5
## [1] 6
## [1] 7
## [1] 8
## [1] 9
## [1] 10
## [1] 11
## [1] 12
## [1] 13
## [1] 14
## [1] 15
## [1] 16
## [1] 17
```

```
newdata$gamediff
```

```
##  [1]  0 13  8 12 22 14 28 14 12  3 13 14  8 13 11  3  8
```

```
#create region

East <- c('Atlanta', 'Philadelphia', 'NY Red Bulls', 'NYCFC', 'D.C. United', 'New England','Columbus')
West <- c('Seattle', 'Portland', 'Los Angeles FC','San Jose', 'LA Galaxy' )
North <- c('Montreal', 'Minnesota', 'Chicago', 'Toronto FC','Vancouver','Toronto FC', 'Montreal')
Central <- c('Sporting KC', 'Real Salt Lake', 'Colorado')
South <- c('Orlando City', 'Houston', 'FC Dallas')


# map opponents to region
for (j in length(newdata$Opponent))
{
  newdata$region <- 0
}

for( i in 1:length(newdata$Opponent))
  {


  for(j in 1:length(East))
  {

    if (newdata$Opponent[i]==East[j])
                            {
                             newdata$region[i] <- 'East'
    }

    if (newdata$Opponent[i]==West[j] & j<=length(West))
                            {
                             newdata$region[i] <- 'West'
    }

    if (newdata$Opponent[i]==North[j] & j<=length(North))
                            {
                             newdata$region[i] <- 'North'
    }

    if (newdata$Opponent[i]==Central[j] & j<=length(Central))
                            {
                             newdata$region[i] <- 'Central'
    }
    if (newdata$Opponent[i]==South[j] & j<= length(South))
                            {
                             newdata$region[i] <- 'South'
    }

    }

  }
```

```
#creating win percentage column
#

w <- 0
j <- 0
newdata$win_per <- 0
for ( i in 1:length(newdata$Result)){

  if(i==0 && newdata$Result[i]=='W')
  {
    j=j+1
    newdata$win_per[i] <- (j/i)*100
  }
  else if (i==0 && newdata$Result != 'W')
    {
      newdata$win_per[i]=0
  }
  else if (i!=0 && newdata$Result[i]=='W'){
    j=j+1
    newdata$win_per[i] <- (j/i)*100
  }
  else if (i!=0 && newdata$Result[i]!='W'){

    newdata$win_per[i] <- (j/i)*100
  }


}
```

```
#finding days between wins
class(newdata$Result)
```

```
## [1] "factor"
```

```
newdata$result1 <- as.character(newdata$Result)
j <- 0
newdata$days_between_wins <- 0
for (i in 1:length(newdata$gamediff)) {

  if( newdata$gamediff[i]==0) {
    newdata$days_between_wins[i] <- 0
  }
  else if (i!=0 && newdata$result1[i]=='W'){

    j<- newdata$gamediff[i-1]
    newdata$days_between_wins [i] <- (newdata$gamediff[i-1])

  }
  else if (i!=0 && newdata$result1[i]!='W') {
      newdata$days_between_wins[i] <- newdata$gamediff[i] + j
      j <- newdata$gamediff[i]+j

  }
```

```
  }

newdata$days_between_wins
```

```
##  [1]   0  13  21  33  12  26  54  28  40  43  56  70  78  91 102 105 113
```

```
# categorize the days between wins

# check percentages of days
newdata$days_between_wins
```

```
##  [1]   0  13  21  33  12  26  54  28  40  43  56  70  78  91 102 105 113
```

```
newdata$cat_days_bet_wins <- 0

for (i in 1: length(newdata$days_between_wins)){
  if(newdata$days_between_wins[i]<15) {
    newdata$cat_days_bet_wins[i]='Low'
  }
  else if(newdata$days_between_wins[i]>14 && newdata$days_between_wins[i]<35) {
    newdata$cat_days_bet_wins[i]='Medium'
  }
  else if (newdata$days_between_wins[i]>34) {
    newdata$cat_days_bet_wins[i]='High'
  }

}
newdata$days_between_wins
```

```
##  [1]   0  13  21  33  12  26  54  28  40  43  56  70  78  91 102 105 113
```

```
newdata$cat_days_bet_wins
```

```
##  [1] "Low"    "Low"    "Medium" "Medium" "Low"    "Medium" "High"
##  [8] "Medium" "High"   "High"   "High"   "High"   "High"   "High"
## [15] "High"   "High"   "High"
```

```
#attaching previous GF to next game

newdata$GF1[1] <- 0

for (i in 2:length(newdata$GF)){
  newdata$GF1[i] <- newdata$GF[i-1]

}




#Attaching previous result to next game
class(newdata$Result)
```

```
## [1] "factor"
```

```
# lest assume that crowd will come full in support for the first home game
newdata$result_final[1] <- 'W'
for (i in 2: length(newdata$Result))
{
  print(newdata$Result[i-1])
  newdata$result_final[i] <- as.character(newdata$Result[i-1])
}
```

```
## [1] W
## Levels: D L W
## [1] L
## Levels: D L W
## [1] D
## Levels: D L W
## [1] L
## Levels: D L W
## [1] W
## Levels: D L W
## [1] L
## Levels: D L W
## [1] L
## Levels: D L W
## [1] W
## Levels: D L W
## [1] L
## Levels: D L W
## [1] L
## Levels: D L W
## [1] L
## Levels: D L W
## [1] L
## Levels: D L W
## [1] L
## Levels: D L W
## [1] L
## Levels: D L W
## [1] L
## Levels: D L W
## [1] D
## Levels: D L W
```

```
# attaching previous GA to next game
newdata$GA1[1] <- 0
for (i in 2:length(newdata$GA)){
  newdata$GA1[i] <- newdata$GA[i-1]

}
newdata$GA1
```

```
##  [1] 0 0 2 1 3 1 2 2 2 4 2 2 4 3 5 2 0
```

```r
#creating goal difference
newdata$goaldiff <- newdata$GF1-newdata$GA1

newdata$goaldiff
```

```
##  [1]  0  3 -2  0 -3  1 -2 -2  1 -3 -2 -1 -3 -2 -4 -2  0
```

```r
# categorize goal difference
# positive is FC cin has scored more goals negative FC cin has accepted more goals

for (i in 1: length(newdata$goaldiff)){

  if(as.numeric(newdata$goaldiff[i])<0){
    newdata$goaldiff_cat[i] <- 'Negative'
  } else {
    print(newdata$gamediff[i])
    newdata$goaldiff_cat[i] <- 'Positive'
  }
}
```

```
## [1] 0
## [1] 13
## [1] 12
## [1] 14
## [1] 12
## [1] 8
```

```r
newdata$goaldiff
```

```
##  [1]  0  3 -2  0 -3  1 -2 -2  1 -3 -2 -1 -3 -2 -4 -2  0
```

```r
# lets write this data for Tableau calculations

#write.csv(newdata,'Homedata1.csv')
str(newdata)
```

```
## 'data.frame':    17 obs. of  29 variables:
##  $ ï..Round        : Factor w/ 1 level "Regular Season": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Day             : Factor w/ 5 levels "Fri","Sat","Sun",..: 3 2 3 1 2 2 2 2 4 3 ...
##  $ Date            : Factor w/ 17 levels "3/17/2019","3/30/2019",..: 1 2 4 3 5 6 7 10 8 9 ...
##  $ Time            : Factor w/ 6 levels "13:00 (12:00)",..: 3 5 2 5 1 5 5 5 6 4 ...
##  $ Venue           : Factor w/ 1 level "Home": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Result          : Factor w/ 3 levels "D","L","W": 3 2 1 2 3 2 2 3 2 2 ...
##  $ GF              : int  3 0 1 0 2 0 0 3 1 0 ...
##  $ GA              : int  0 2 1 3 1 2 2 2 4 2 ...
##  $ Opponent        : Factor w/ 17 levels "Atlanta","Chicago",..: 13 12 15 14 7 9 6 5 4 8 ...
##  $ Region          : Factor w/ 5 levels "Central","East",..: 5 2 1 1 3 2 5 4 2 2 ...
##  $ xG              : num  1.5 0.4 1.8 0.9 0.7 1.8 0.6 1.5 1.2 0.9 ...
##  $ xGA             : num  1 2 2.4 2.1 1.2 1.1 1.8 1.9 0.9 1.8 ...
##  $ Attendance      : int  32250 25867 26023 26416 26258 28290 32250 26276 28774 25095 ...
##  $ Captain         : Factor w/ 4 levels "Greg Garza","Justin Hoyte",..: 3 1 3 3 3 3 2 3 3 3 ...
```

```
##  $ Formation        : Factor w/ 4 levels "4-1-4-1","4-2-3-1",..: 4 4 4 4 3 3 2 2 2 2 ...
##  $ Referee          : Factor w/ 13 levels "Allen Chapman",..: 5 12 5 4 6 4 8 3 2 7 ...
##  $ Match.Report     : Factor w/ 1 level "Match Report": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Date1            : Date, format: "2019-03-17" "2019-03-30" ...
##  $ gamediff         : num  0 13 8 12 22 14 28 14 12 3 ...
##  $ region           : chr  "West" "East" "Central" "Central" ...
##  $ win_per          : num  100 50 33.3 25 40 ...
##  $ result1          : chr  "W" "L" "D" "L" ...
##  $ days_between_wins: num  0 13 21 33 12 26 54 28 40 43 ...
##  $ cat_days_bet_wins: chr  "Low" "Low" "Medium" "Medium" ...
##  $ GF1              : num  0 3 0 1 0 2 0 0 3 1 ...
##  $ result_final     : chr  "W" "W" "L" "D" ...
##  $ GA1              : num  0 0 2 1 3 1 2 2 2 4 ...
##  $ goaldiff         : num  0 3 -2 0 -3 1 -2 -2 1 -3 ...
##  $ goaldiff_cat     : chr  "Positive" "Positive" "Negative" "Positive" ...
```

```
newdata$Time <- (substr(newdata$Time, 1, 5))
str(newdata)
```

```
## 'data.frame':    17 obs. of  29 variables:
##  $ ï..Round         : Factor w/ 1 level "Regular Season": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Day              : Factor w/ 5 levels "Fri","Sat","Sun",..: 3 2 3 1 2 2 2 2 4 3 ...
##  $ Date             : Factor w/ 17 levels "3/17/2019","3/30/2019",..: 1 2 4 3 5 6 7 10 8 9 ...
##  $ Time             : chr  "17:00" "19:30" "15:00" "19:30" ...
##  $ Venue            : Factor w/ 1 level "Home": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Result           : Factor w/ 3 levels "D","L","W": 3 2 1 2 3 2 2 3 2 2 ...
##  $ GF               : int  3 0 1 0 2 0 0 3 1 0 ...
##  $ GA               : int  0 2 1 3 1 2 2 2 4 2 ...
##  $ Opponent         : Factor w/ 17 levels "Atlanta","Chicago",..: 13 12 15 14 7 9 6 5 4 8 ...
##  $ Region           : Factor w/ 5 levels "Central","East",..: 5 2 1 1 3 2 5 4 2 2 ...
##  $ xG               : num  1.5 0.4 1.8 0.9 0.7 1.8 0.6 1.5 1.2 0.9 ...
##  $ xGA              : num  1 2 2.4 2.1 1.2 1.1 1.8 1.9 0.9 1.8 ...
##  $ Attendance       : int  32250 25867 26023 26416 26258 28290 32250 26276 28774 25095 ...
##  $ Captain          : Factor w/ 4 levels "Greg Garza","Justin Hoyte",..: 3 1 3 3 3 3 2 3 3 3 ...
##  $ Formation        : Factor w/ 4 levels "4-1-4-1","4-2-3-1",..: 4 4 4 4 3 3 2 2 2 2 ...
##  $ Referee          : Factor w/ 13 levels "Allen Chapman",..: 5 12 5 4 6 4 8 3 2 7 ...
##  $ Match.Report     : Factor w/ 1 level "Match Report": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Date1            : Date, format: "2019-03-17" "2019-03-30" ...
##  $ gamediff         : num  0 13 8 12 22 14 28 14 12 3 ...
##  $ region           : chr  "West" "East" "Central" "Central" ...
##  $ win_per          : num  100 50 33.3 25 40 ...
##  $ result1          : chr  "W" "L" "D" "L" ...
##  $ days_between_wins: num  0 13 21 33 12 26 54 28 40 43 ...
##  $ cat_days_bet_wins: chr  "Low" "Low" "Medium" "Medium" ...
##  $ GF1              : num  0 3 0 1 0 2 0 0 3 1 ...
##  $ result_final     : chr  "W" "W" "L" "D" ...
##  $ GA1              : num  0 0 2 1 3 1 2 2 2 4 ...
##  $ goaldiff         : num  0 3 -2 0 -3 1 -2 -2 1 -3 ...
##  $ goaldiff_cat     : chr  "Positive" "Positive" "Negative" "Positive" ...
```

```
# 9 is opponent, 10 is region, 15 is formation
data1 <- newdata[,-c(1,3,5,7,8,9,11,10,12,15,16,17,18,22,14,28,26,6,24,29,20)]
str(data1)
```

```
## 'data.frame':    17 obs. of  8 variables:
##  $ Day              : Factor w/ 5 levels "Fri","Sat","Sun",..: 3 2 3 1 2 2 2 2 4 3 ...
##  $ Time             : chr  "17:00" "19:30" "15:00" "19:30" ...
##  $ Attendance       : int  32250 25867 26023 26416 26258 28290 32250 26276 28774 25095 ...
##  $ gamediff         : num  0 13 8 12 22 14 28 14 12 3 ...
##  $ win_per          : num  100 50 33.3 25 40 ...
##  $ days_between_wins: num  0 13 21 33 12 26 54 28 40 43 ...
##  $ GF1              : num  0 3 0 1 0 2 0 0 3 1 ...
##  $ GA1              : num  0 0 2 1 3 1 2 2 2 4 ...
```

```
#create dummies

#install.packages("fastDummies")
data2 <- fastDummies::dummy_cols(data1,remove_first_dummy = TRUE)
str(data2)
```

```
## 'data.frame':    17 obs. of  17 variables:
##  $ Day              : Factor w/ 5 levels "Fri","Sat","Sun",..: 3 2 3 1 2 2 2 2 4 3 ...
##  $ Time             : chr  "17:00" "19:30" "15:00" "19:30" ...
##  $ Attendance       : int  32250 25867 26023 26416 26258 28290 32250 26276 28774 25095 ...
##  $ gamediff         : num  0 13 8 12 22 14 28 14 12 3 ...
##  $ win_per          : num  100 50 33.3 25 40 ...
##  $ days_between_wins: num  0 13 21 33 12 26 54 28 40 43 ...
##  $ GF1              : num  0 3 0 1 0 2 0 0 3 1 ...
##  $ GA1              : num  0 0 2 1 3 1 2 2 2 4 ...
##  $ Day_Sat          : int  0 1 0 0 1 1 1 1 0 0 ...
##  $ Day_Sun          : int  1 0 1 0 0 0 0 0 0 1 ...
##  $ Day_Thu          : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Day_Wed          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Time_19:30       : int  0 1 0 1 0 1 1 1 0 0 ...
##  $ Time_15:00       : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ Time_13:00       : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ Time_20:00       : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Time_18:00       : int  0 0 0 0 0 0 0 0 0 1 ...
```

```
# drop wednesday since only one match has happend on wednesday
data3 <- data2[,-c(1,2,12)]
str(data3)
```

```
## 'data.frame':    17 obs. of  14 variables:
##  $ Attendance       : int  32250 25867 26023 26416 26258 28290 32250 26276 28774 25095 ...
##  $ gamediff         : num  0 13 8 12 22 14 28 14 12 3 ...
##  $ win_per          : num  100 50 33.3 25 40 ...
##  $ days_between_wins: num  0 13 21 33 12 26 54 28 40 43 ...
##  $ GF1              : num  0 3 0 1 0 2 0 0 3 1 ...
##  $ GA1              : num  0 0 2 1 3 1 2 2 2 4 ...
##  $ Day_Sat          : int  0 1 0 0 1 1 1 1 0 0 ...
##  $ Day_Sun          : int  1 0 1 0 0 0 0 0 0 1 ...
##  $ Day_Thu          : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Time_19:30       : int  0 1 0 1 0 1 1 1 0 0 ...
##  $ Time_15:00       : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ Time_13:00       : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ Time_20:00       : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Time_18:00       : int  0 0 0 0 0 0 0 0 0 1 ...
```

```
colnames
```

```
## function (x, do.NULL = TRUE, prefix = "col")
## {
##     if (is.data.frame(x) && do.NULL)
##         return(names(x))
##     dn <- dimnames(x)
##     if (!is.null(dn[[2L]]))
##         dn[[2L]]
##     else {
##         nc <- NCOL(x)
##         if (do.NULL)
##             NULL
##         else if (nc > 0L)
##             paste0(prefix, seq_len(nc))
##         else character()
##     }
## }
## <bytecode: 0x000000001ccecfd8>
## <environment: namespace:base>
```

```
split <- round(nrow(data3) * 0.8)
train.df <- data3[1:split, ]
str(train.df)
```

```
## 'data.frame':    14 obs. of  14 variables:
##  $ Attendance       : int  32250 25867 26023 26416 26258 28290 32250 26276 28774 25095 ...
##  $ gamediff         : num  0 13 8 12 22 14 28 14 12 3 ...
##  $ win_per          : num  100 50 33.3 25 40 ...
##  $ days_between_wins: num  0 13 21 33 12 26 54 28 40 43 ...
##  $ GF1              : num  0 3 0 1 0 2 0 0 3 1 ...
##  $ GA1              : num  0 0 2 1 3 1 2 2 2 4 ...
##  $ Day_Sat          : int  0 1 0 0 1 1 1 1 0 0 ...
##  $ Day_Sun          : int  1 0 1 0 0 0 0 0 0 1 ...
##  $ Day_Thu          : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Time_19:30       : int  0 1 0 1 0 1 1 1 0 0 ...
##  $ Time_15:00       : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ Time_13:00       : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ Time_20:00       : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ Time_18:00       : int  0 0 0 0 0 0 0 0 0 1 ...
```

```
test.df <- data3[(split+1):nrow(data3), ]

cin.lm <- lm(Attendance ~., data=train.df)
summary(cin.lm)
```

```
##
## Call:
## lm(formula = Attendance ~ ., data = train.df)
##
## Residuals:
##          1          2          3          4          5          6
```

```
## -6.111e-13 -1.469e+02  2.406e-13 -2.674e-13 -1.289e-13  5.394e+02
##          7          8          9         10         11         12
## -2.754e+02  5.210e+02  9.850e-14 -7.909e+02  1.909e-13 -8.838e+02
##         13         14
##  7.909e+02  2.456e+02
##
## Coefficients: (1 not defined because of singularities)
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        31867.32    8155.49   3.907    0.160
## gamediff             291.90     211.39   1.381    0.399
## win_per             -223.33     201.70  -1.107    0.468
## days_between_wins     26.53      73.21   0.362    0.779
## GF1                -1504.09    2038.84  -0.738    0.595
## GA1                -4126.34    3264.20  -1.264    0.426
## Day_Sat             4301.68    4099.71   1.049    0.485
## Day_Sun            22715.65   13780.46   1.648    0.347
## Day_Thu            12551.86   10254.56   1.224    0.436
## `Time_19:30`        1383.99    3094.17   0.447    0.732
## `Time_15:00`      -15755.34   12446.91  -1.266    0.426
## `Time_13:00`        4661.01    7973.34   0.585    0.663
## `Time_20:00`             NA         NA      NA       NA
## `Time_18:00`       -6004.32   14704.27  -0.408    0.753
##
## Residual standard error: 1659 on 1 degrees of freedom
## Multiple R-squared:  0.9639, Adjusted R-squared:  0.5305
## F-statistic: 2.224 on 12 and 1 DF,  p-value: 0.4848
```

```
#subset selection
library(leaps)
search <- regsubsets(Attendance~., data = train.df, nbest = 1, nvmax = dim(train.df)[2],
                     method = "exhaustive")
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
## force.in = force.in, : 1 linear dependencies found
```

```
## Reordering variables and trying again:
```

```
sum <- summary(search)

sum$which
```

```
##    (Intercept) gamediff win_per days_between_wins   GF1   GA1 Day_Sat
## 1         TRUE    FALSE    TRUE             FALSE FALSE FALSE   FALSE
## 2         TRUE    FALSE    TRUE              TRUE FALSE FALSE   FALSE
## 3         TRUE     TRUE    TRUE              TRUE FALSE FALSE   FALSE
## 4         TRUE     TRUE    TRUE              TRUE FALSE FALSE    TRUE
## 5         TRUE     TRUE   FALSE             FALSE FALSE FALSE   FALSE
## 6         TRUE     TRUE   FALSE             FALSE FALSE FALSE   FALSE
## 7         TRUE     TRUE   FALSE              TRUE  TRUE  TRUE   FALSE
## 8         TRUE     TRUE    TRUE             FALSE FALSE  TRUE    TRUE
## 9         TRUE     TRUE    TRUE             FALSE  TRUE  TRUE    TRUE
## 10        TRUE     TRUE    TRUE             FALSE  TRUE  TRUE    TRUE
## 11        TRUE     TRUE    TRUE             FALSE  TRUE  TRUE    TRUE
```

```
## 12        TRUE     TRUE    TRUE              TRUE  TRUE  TRUE    TRUE
##    Day_Sun Day_Thu `Time_19:30` `Time_15:00` `Time_13:00` `Time_20:00`
## 1    FALSE   FALSE        FALSE        FALSE        FALSE        FALSE
## 2    FALSE   FALSE        FALSE        FALSE        FALSE        FALSE
## 3    FALSE   FALSE        FALSE        FALSE        FALSE        FALSE
## 4    FALSE   FALSE        FALSE        FALSE        FALSE        FALSE
## 5     TRUE   FALSE        FALSE         TRUE         TRUE        FALSE
## 6     TRUE    TRUE        FALSE         TRUE         TRUE        FALSE
## 7     TRUE    TRUE        FALSE         TRUE        FALSE        FALSE
## 8     TRUE    TRUE        FALSE         TRUE        FALSE        FALSE
## 9     TRUE    TRUE        FALSE         TRUE        FALSE        FALSE
## 10    TRUE    TRUE        FALSE         TRUE         TRUE        FALSE
## 11    TRUE    TRUE         TRUE         TRUE         TRUE        FALSE
## 12    TRUE    TRUE         TRUE         TRUE         TRUE        FALSE
##    `Time_18:00`
## 1         FALSE
## 2         FALSE
## 3         FALSE
## 4         FALSE
## 5          TRUE
## 6          TRUE
## 7         FALSE
## 8          TRUE
## 9          TRUE
## 10         TRUE
## 11         TRUE
## 12         TRUE
```

```
sum$adjr2
```

```
##  [1] 0.1047991 0.1792596 0.2655444 0.5324591 0.6320757 0.7379035 0.8130753
##  [8] 0.8466987 0.8419965 0.8063288 0.7344008 0.5304678
```

```
sum$bic
```

```
##  [1]   2.607617   2.812736   2.562369  -2.596618  -4.961059  -8.939745
##  [7] -13.190892 -15.880545 -15.942533 -14.481399 -13.097277 -12.185803
```

```
#LM with BIC=-12.378068 which model 10

model1 <- lm(Attendance~GF1+GA1+Day_Thu+Day_Sat+Day_Sun+`Time_15:00`+`Time_19:30`+`Time_13:00`+win_per+
summary(model1)
```

```
##
## Call:
## lm(formula = Attendance ~ GF1 + GA1 + Day_Thu + Day_Sat + Day_Sun +
##     `Time_15:00` + `Time_19:30` + `Time_13:00` + win_per + days_between_wins,
##     data = train.df)
##
## Residuals:
##          1          2          3          4          5          6
## -7.561e+01  3.596e+02  2.309e-13  2.256e-13 -3.908e-14  2.343e+02
```

```
##          7          8          9         10         11         12
##  1.293e+03 -2.635e+02  4.619e-14 -5.958e+02  6.040e-14 -2.274e+03
##         13         14
##  6.714e+02  6.511e+02
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        39223.07    5416.69   7.241  0.00543 **
## GF1                -3487.93     959.17  -3.636  0.03583 *
## GA1                -7214.58    2036.97  -3.542  0.03831 *
## Day_Thu            23331.93    6342.11   3.679  0.03478 *
## Day_Sat             8292.88    2934.87   2.826  0.06643 .
## Day_Sun            26424.81    7443.38   3.550  0.03809 *
## `Time_15:00`      -15253.46    4356.47  -3.501  0.03944 *
## `Time_19:30`        4395.04    2119.71   2.073  0.12982
## `Time_13:00`       13048.89    4699.10   2.777  0.06917 .
## win_per             -333.22     118.94  -2.802  0.06776 .
## days_between_wins     55.48      36.29   1.529  0.22377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1667 on 3 degrees of freedom
## Multiple R-squared:  0.8907, Adjusted R-squared:  0.5262
## F-statistic: 2.444 on 10 and 3 DF,  p-value: 0.2501
```

```
#LM with Adjr2 which is model 8
model2 <- lm(Attendance~gamediff+win_per+GA1+Day_Sat+Day_Sun+Day_Thu+`Time_15:00`+`Time_18:00`,data=tra
summary(model2)
```

```
##
## Call:
## lm(formula = Attendance ~ gamediff + win_per + GA1 + Day_Sat +
##     Day_Sun + Day_Thu + `Time_15:00` + `Time_18:00`, data = train.df)
##
## Residuals:
##          1          2          3          4          5          6
##  4.263e-13 -4.353e+02 -1.268e-13 -1.659e-13  7.896e+01  3.537e+02
##          7          8          9         10         11         12
## -4.757e+02  1.034e+03  2.320e-13 -1.071e+03  3.575e+02 -3.172e+02
##         13         14
##  1.071e+03 -5.958e+02
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   28046.56    1817.76  15.429 2.08e-05 ***
## gamediff        414.74      65.90   6.293  0.00149 **
## win_per        -187.84      47.27  -3.973  0.01060 *
## GA1           -1911.45     472.48  -4.046  0.00987 **
## Day_Sat        2256.13    1217.70   1.853  0.12311
## Day_Sun       22987.55    3714.16   6.189  0.00161 **
## Day_Thu        5834.79    1549.13   3.767  0.01307 *
## `Time_15:00` -18244.78    3081.15  -5.921  0.00196 **
## `Time_18:00` -12831.38    3025.03  -4.242  0.00816 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 948 on 5 degrees of freedom
## Multiple R-squared:  0.941,  Adjusted R-squared:  0.8467
## F-statistic: 9.975 on 8 and 5 DF,  p-value: 0.01072
```