Assignment_3.2_Vayuvegula_Soma_Shekar

Soma Shekar Vayuvegula

12/18/2022

Importing Data to Dataframe and displaying details of Dataframe

Head of the dataframe

```
##   month day attend day_of_week opponent temp  skies day_night cap shirt
## 1   APR  10  56000     Tuesday  Pirates   67  Clear       Day  NO    NO
## 2   APR  11  29729   Wednesday  Pirates   58 Cloudy     Night  NO    NO
## 3   APR  12  28328    Thursday  Pirates   57 Cloudy     Night  NO    NO
## 4   APR  13  31601      Friday   Padres   54 Cloudy     Night  NO    NO
## 5   APR  14  46549    Saturday   Padres   57 Cloudy     Night  NO    NO
## 6   APR  15  38359      Sunday   Padres   65  Clear       Day  NO    NO
##   fireworks bobblehead
## 1        NO         NO
## 2        NO         NO
## 3        NO         NO
## 4       YES         NO
## 5        NO         NO
## 6        NO         NO
```

Summary of the dataframe

```
##     month               day             attend      day_of_week
##  Length:81          Min.   : 1.00   Min.   :24312   Length:81
##  Class :character   1st Qu.: 8.00   1st Qu.:34493   Class :character
##  Mode  :character   Median :15.00   Median :40284   Mode  :character
##                     Mean   :16.14   Mean   :41040
##                     3rd Qu.:25.00   3rd Qu.:46588
##                     Max.   :31.00   Max.   :56000
##    opponent              temp           skies            day_night
##  Length:81          Min.   :54.00   Length:81          Length:81
##  Class :character   1st Qu.:67.00   Class :character   Class :character
##  Mode  :character   Median :73.00   Mode  :character   Mode  :character
##                     Mean   :73.15
##                     3rd Qu.:79.00
##                     Max.   :95.00
##     cap               shirt             fireworks          bobblehead
##  Length:81          Length:81          Length:81          Length:81
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

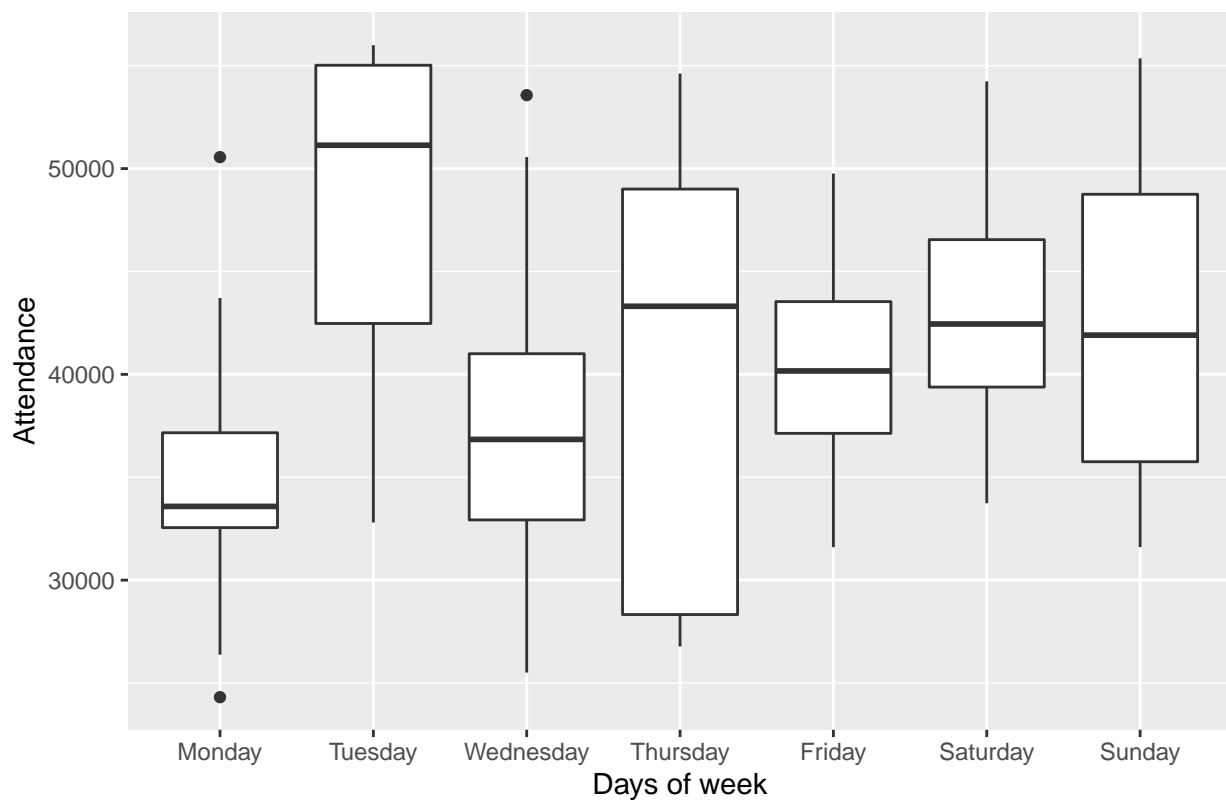Dimension of the dataframe

```
## [1] 81 12
```

Structure of the dataframe

```
## 'data.frame':    81 obs. of  12 variables:
##  $ month      : chr  "APR" "APR" "APR" "APR" ...
##  $ day        : int  10 11 12 13 14 15 23 24 25 27 ...
##  $ attend     : int  56000 29729 28328 31601 46549 38359 26376 44014 26345 44807 ...
##  $ day_of_week: chr  "Tuesday" "Wednesday" "Thursday" "Friday" ...
##  $ opponent   : chr  "Pirates" "Pirates" "Pirates" "Padres" ...
##  $ temp       : int  67 58 57 54 57 65 60 63 64 66 ...
##  $ skies      : chr  "Clear " "Cloudy" "Cloudy" "Cloudy" ...
##  $ day_night  : chr  "Day" "Night" "Night" "Night" ...
##  $ cap        : chr  "NO" "NO" "NO" "NO" ...
##  $ shirt      : chr  "NO" "NO" "NO" "NO" ...
##  $ fireworks  : chr  "NO" "NO" "NO" "YES" ...
##  $ bobblehead : chr  "NO" "NO" "NO" "NO" ...
```
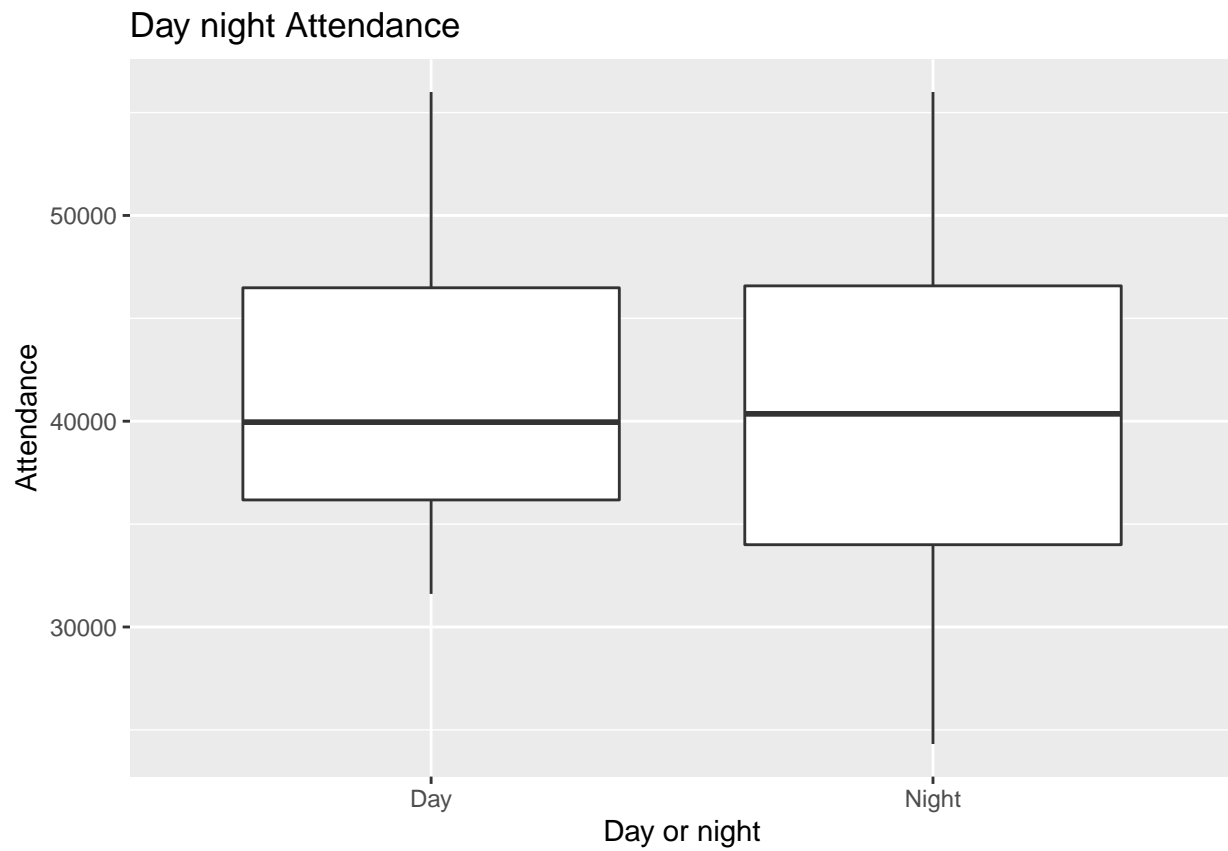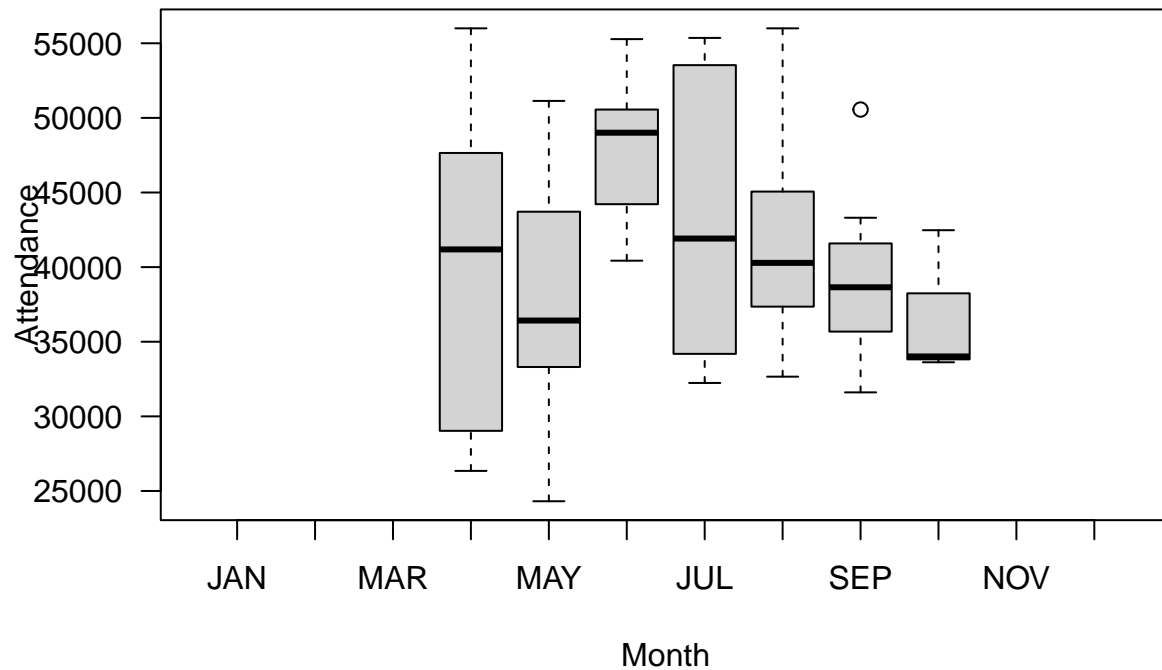
Box plots

Weekdays Attendance

Observations:

As per the above Box plot, Tuedays have the maximum attendance and Mondays have the lowest attendance

Day night Attendance
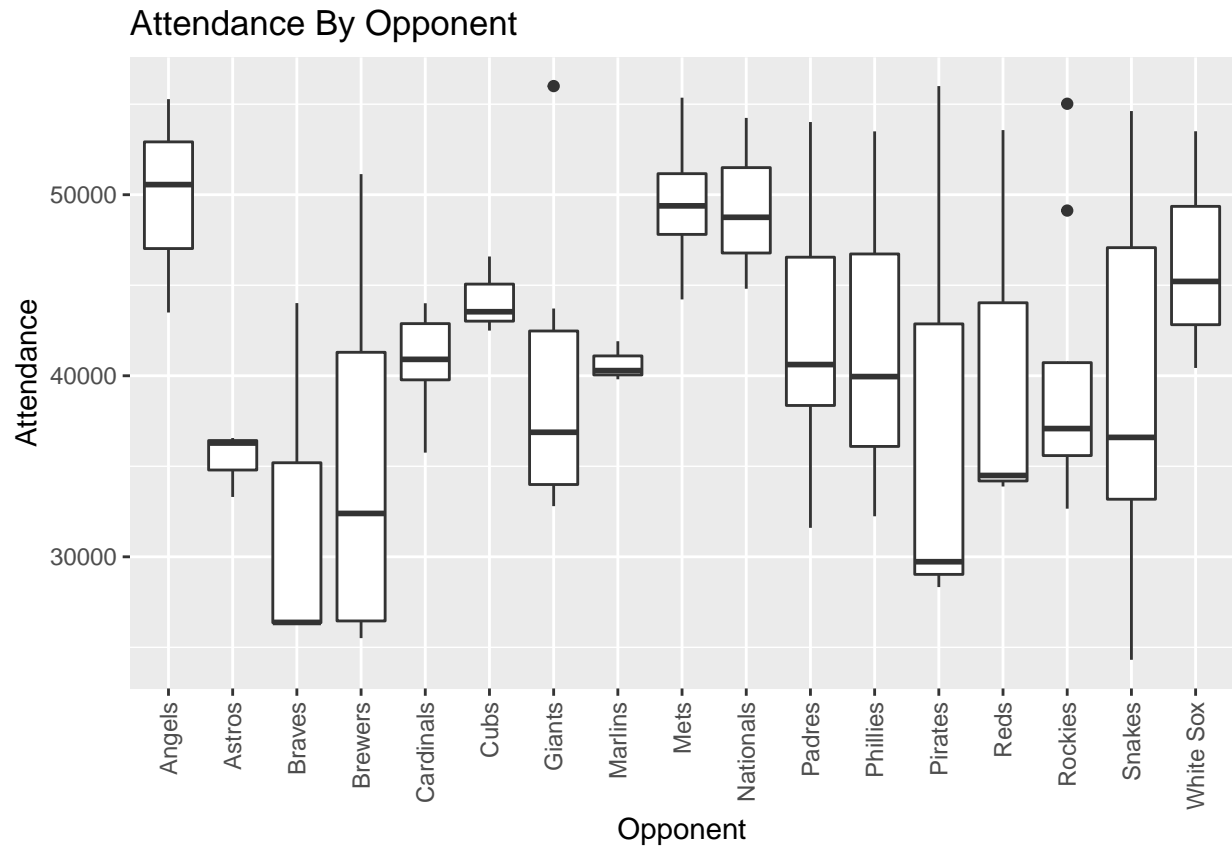

Observations:

As per the above Box plot, there is not big difference in attendance either it is Day or Night.

Attendance per Month

Observations:

As per the above Box plot, Month of June has the maximum attendance and Month of October has the lowest attendance. My assumption is as June and July is the holiday season for students, we are having maximum attendance is June.
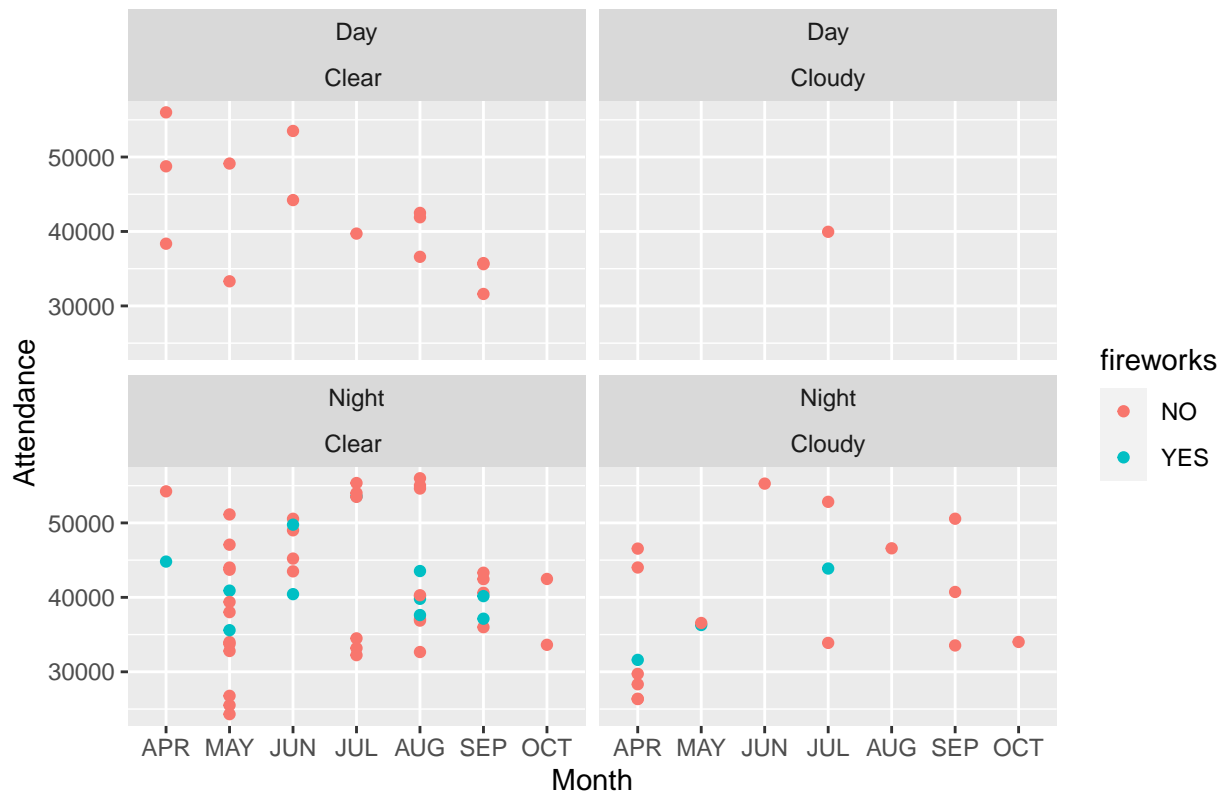
Attendance By Opponent

Observations:

As per the above Box plot,Match against opponents Angels, Mets, Nationals and White Sox has maximum attendance compared to others.

Scatter Plots:

Attendance By Weather

Observations:

As per the above Box plot, Nights with clear skies have the maximum attendance. My assumption is that we are having maximum attendance after office or school hours when the weather permits.

Regression Model

```
##
## Call:
## lm(formula = my.model, data = df_dodgers)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10786.5  -3628.1   -516.1   2230.2  14351.0
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          33909.16    2521.81  13.446  < 2e-16 ***
## monthMAY             -2385.62    2291.22  -1.041  0.30152
## monthJUN              7163.23    2732.72   2.621  0.01083 *
## monthJUL              2849.83    2578.60   1.105  0.27303
## monthAUG              2377.92    2402.91   0.990  0.32593
## monthSEP                29.03    2521.25   0.012  0.99085
## monthOCT              -662.67    4046.45  -0.164  0.87041
## day_of_weekTuesday    7911.49    2702.21   2.928  0.00466 **
## day_of_weekWednesday  2460.02    2514.03   0.979  0.33134
```

```
## day_of_weekThursday      775.36      3486.15    0.222  0.82467
## day_of_weekFriday       4883.82      2504.65    1.950  0.05537 .
## day_of_weekSaturday     6372.06      2552.08    2.497  0.01500 *
## day_of_weekSunday       6724.00      2506.72    2.682  0.00920 **
## bobbleheadYES          10714.90      2419.52    4.429 3.59e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6120 on 67 degrees of freedom
## Multiple R-squared:  0.5444, Adjusted R-squared:  0.456
## F-statistic: 6.158 on 13 and 67 DF,  p-value: 2.083e-07

## bobbleheadYES
##         10715
```

Conclusion:

Multiple Regression model is developed to find relationship between month, day of the week, day or night and bobble head promotion. We got relationship woth p-value of 2.083e-07. The lower the p-value, the greater the statistical significance of the observed difference. We can run bobble head promotion to get more attendance of 10715.

Looking at the different charts plotted above, we have more attendance:

```
1. in the month of June
2. During nights with clear skies
3. on Tuesdays
4. and when opponents are Angels, Mets, Nationals and White Sox
```

With below assumptions:

```
1. June is the holiday time for Students
2. During nights indicating people attending after office or school hours,
```

I would suggest the management to run promotions during holiday times with perfect weather and having matches with major opponents during those times (Angels, Mets, Nationals and White Sox) to get maximum attendance.