Code ▾

Spring 2024 GE 461 Project 1 - Dodger Promotion Data Analysis

Selin Ataş 21902858 - Ömer Tuğrul 22002723

Show

1. Import the Dataset

Show

Show

```
## The columns in this dataset are:
## month
## day
## attend
## day_of_week
## opponent
## temp
## skies
## day_night
## cap
## shirt
## fireworks
## bobblehead
```

First 5 rows of the dataset

Show

```
##   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
##   fireworks bobblehead
## 1        NO         NO
## 2        NO         NO
## 3        NO         NO
## 4       YES         NO
## 5        NO         NO
```

2. Data Analysis

Summary of the dataset to see the data types and the distribution of the variables.

Show

```
##    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  
## 
## 
## 
```

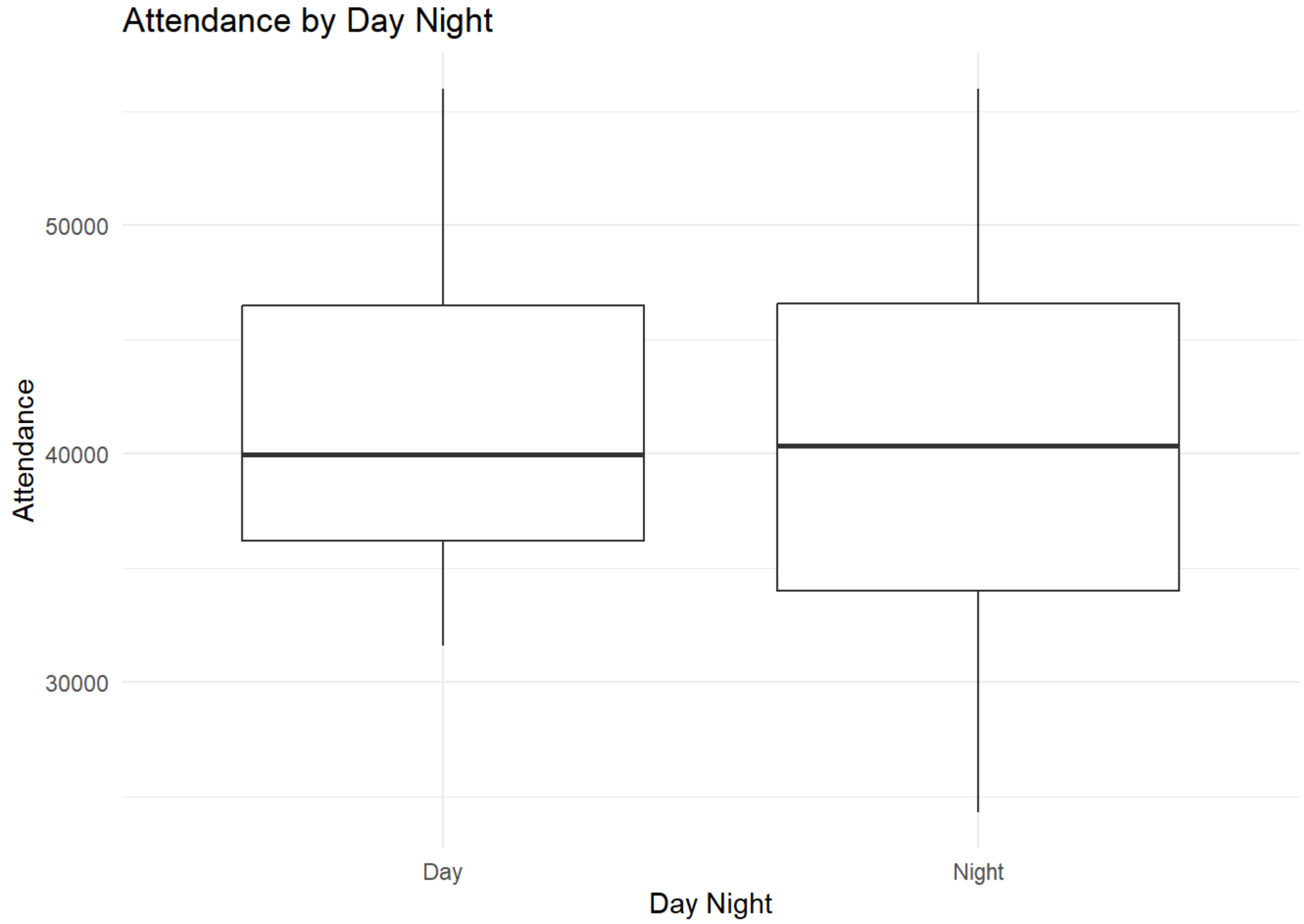
We need to factor our dataset to continue with the analysis.

Show

```
##  month         day            attend          day_of_week     opponent
##  APR:12   Min.   : 1.00   Min.   :24312   Monday   :12   Giants   : 9
##  MAY:18   1st Qu.: 8.00   1st Qu.:34493   Tuesday  :13   Padres   : 9
##  JUN: 9   Median :15.00   Median :40284   Wednesday:12   Rockies  : 9
##  JUL:12   Mean   :16.14   Mean   :41040   Thursday : 5   Snakes   : 9
##  AUG:15   3rd Qu.:25.00   3rd Qu.:46588   Friday   :13   Cardinals: 7
##  SEP:12   Max.   :31.00   Max.   :56000   Saturday :13   Brewers  : 4
##  OCT: 3                                   Sunday   :13   (Other)  :34
##      temp          skies    day_night   cap     shirt    fireworks bobblehea
d
##  Min.   :12.00   Clear :62   Day  :15   NO :79   NO :78   NO :67    NO :70
##  1st Qu.:19.00   Cloudy:19   Night:66   YES: 2   YES: 3   YES:14    YES:11
##  Median :23.00
##  Mean   :22.84
##  3rd Qu.:26.00
##  Max.   :35.00
##
```

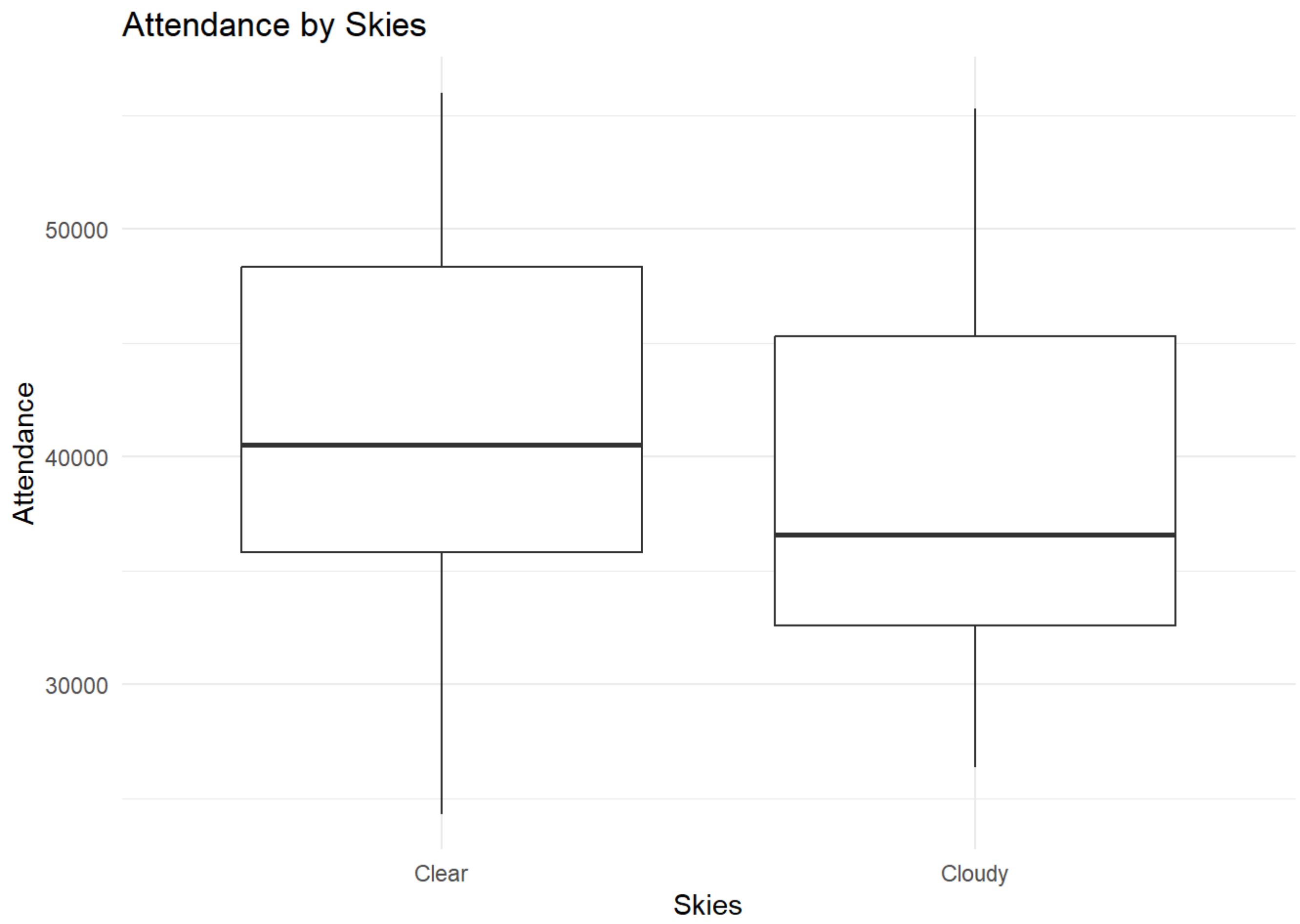
According to this summary, there are 81 observations and 12 variables in the dataset. The most common month is May, the most common sky condition is clear, and the most common day_night condition is night. The average attendance is 41,000. Lets analyze the attendence by month, day_night, skies to see whether the most common values are also the highest attendence values.

Show



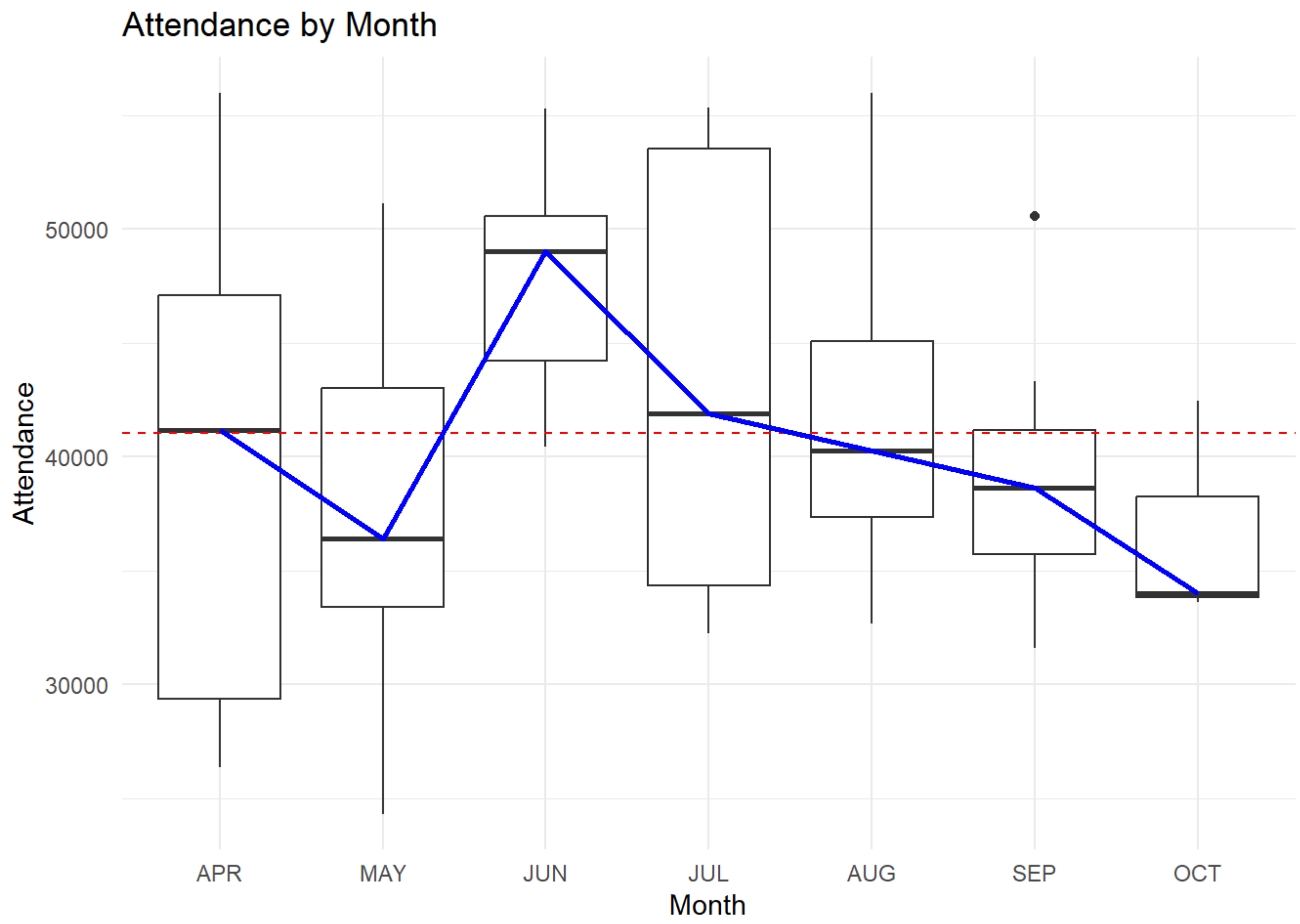
Average attendance by day and by night is almost the same. Distribution of the night games is more spread out than the day games. At this point it is hard to say that the attendance is affected by the day_night. However, we will check this property with the hypothesis testing in further analysis.

Show



Average attendance is higher in the clear days than in the cloudy days. Also, the maximum attendance is higher in the clear days than in the cloudy days, and the minimum attendance is lower in the clear days than in the cloudy days. Clear days may more preferable for the games than the cloudy days.

Show



Show

Average Attendance by Month

month	avg_attend
APR	39592
MAY	37346
JUN	47940
JUL	43884
AUG	42752
SEP	38955
OCT	36704

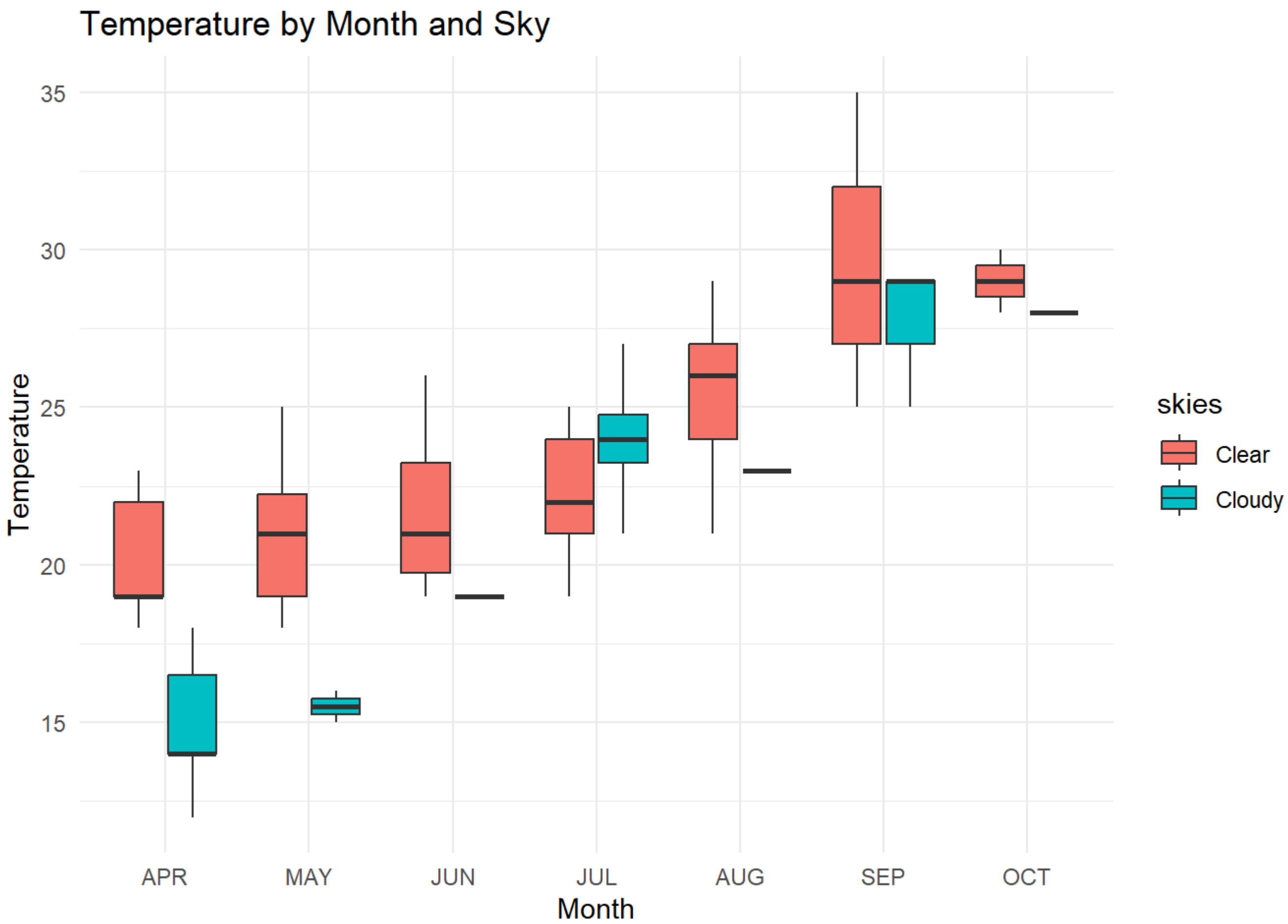
Show

Total Average Attendance

avg_attend
41040

Although the most common month is May, the lowest average attendance is in May as well. It is seen that the peak of the average attendance is in June and it gradually decreases through October. It looks like summer months have higher attendance than the other months. Then, the attendance may be affected by the weather conditions.

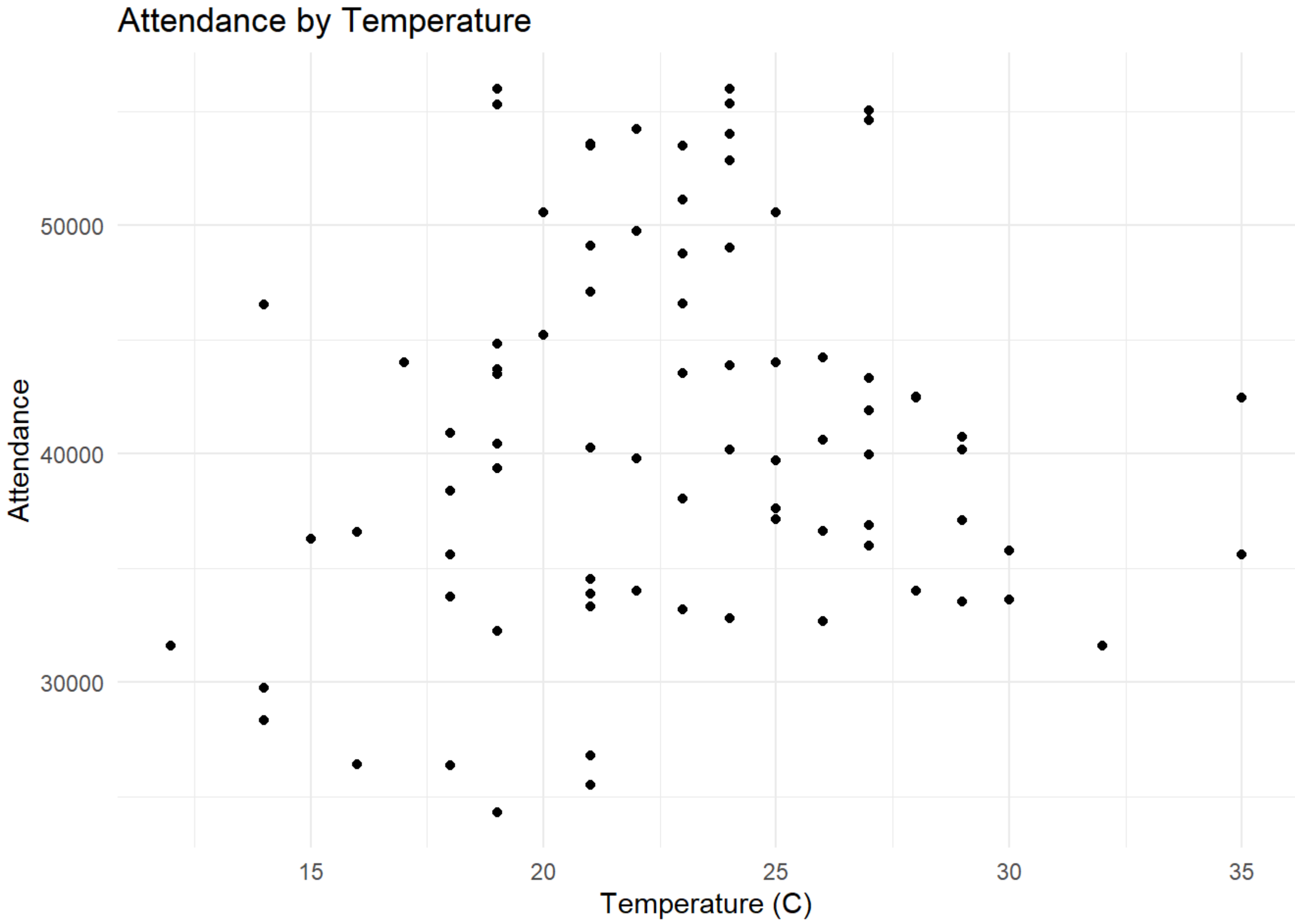
Show



Average temperature is increasing from June to October. From that, we can say that after a certain temperature, the attendance to the games may decrease as the temperature increases. Also, the average temperature is higher in the clear days than in the cloudy days. This may be the reason for the higher attendance in the clear days than in the cloudy days. However, it is hard to make a comment only looking at this graph since the pattern in May and June are very similar but the attendance distribution is highly different.

Then, lets check the relationship between the temperature and the attendence.

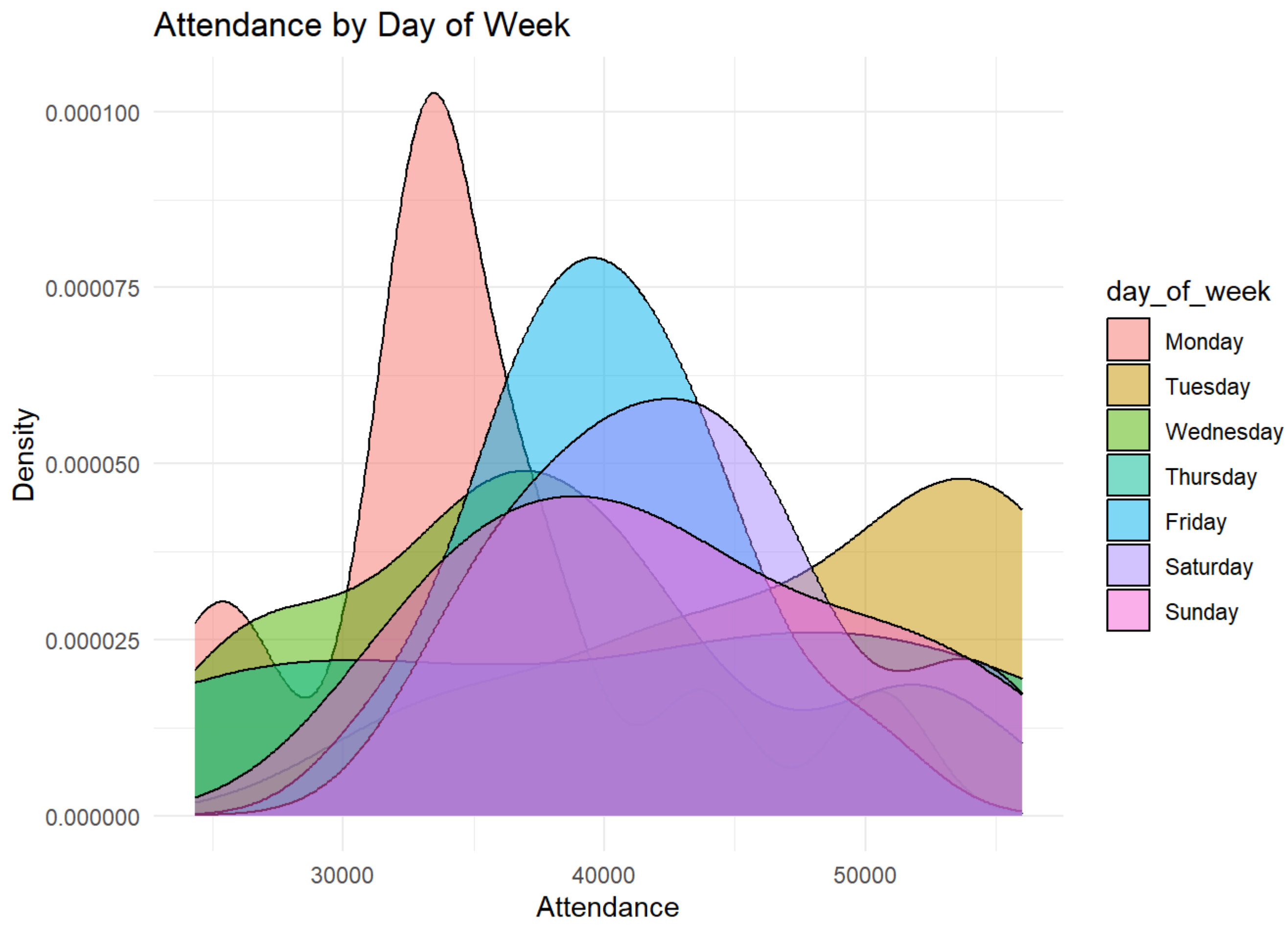
Show



Attandence-temperature plot shows that there is no direct relation between the temperature and the attendence. However, it is seen that the attendence is higher when the temperature is between 20 and 25. It is also lower when the temperature is below 20 and above 25.

After analyzing month, day_night, skies, and temperature, we can say that the attendance is affected by the month and the weather conditions. However, it is hard to predict the attendence by looking at only the month and the weather conditions. Now, we will check is there any relationship between day of week and day of the month with the attendence respectively.

Show



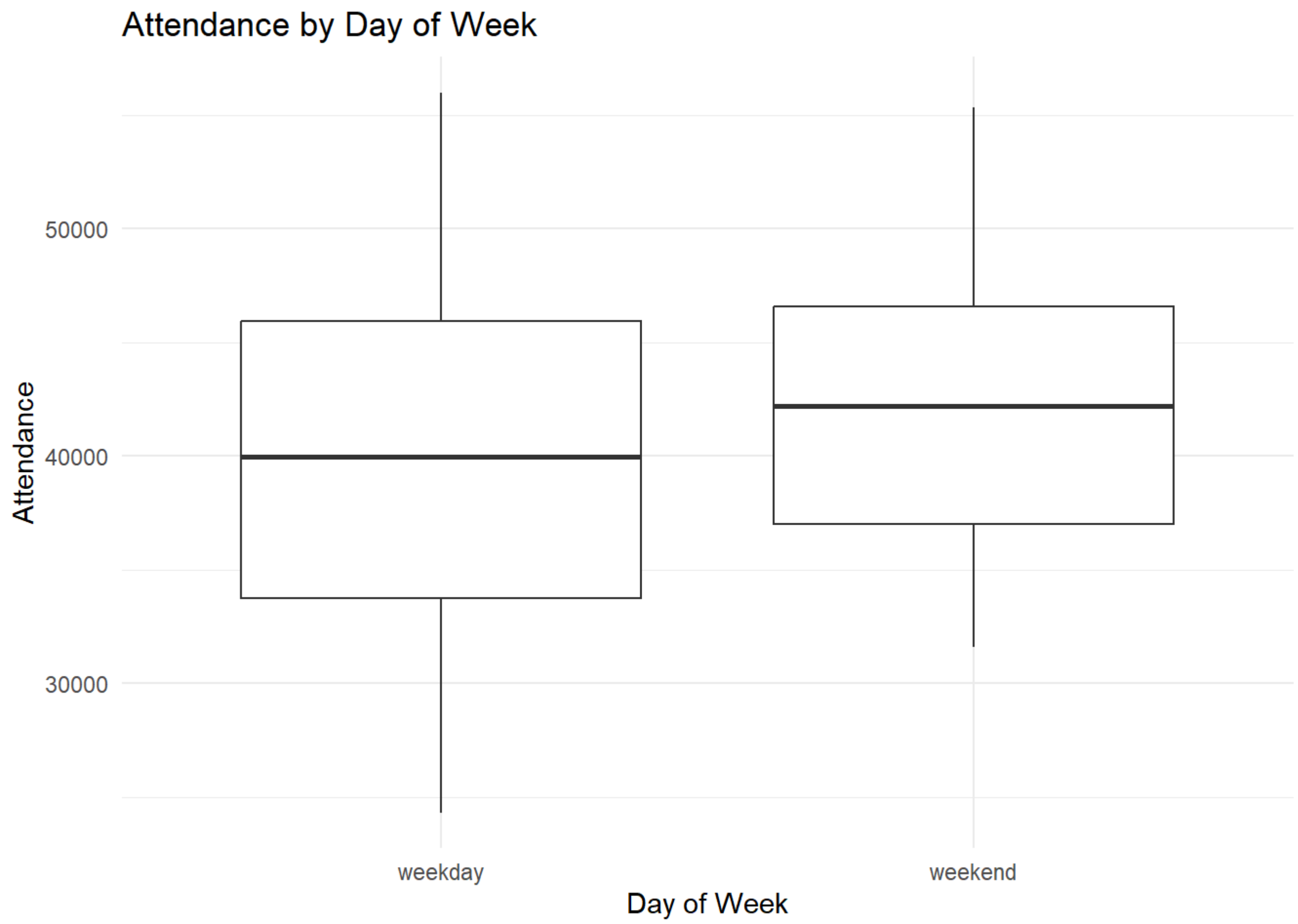
Show

Average Attendance by Day of Week

day_of_week	avg_attend
Monday	34966
Tuesday	47741
Wednesday	37585
Thursday	40407
Friday	40117
Saturday	43073
Sunday	42269

From the density graph, we can say that attendance is generally lower at the beginning of the week, with lowest attendance value of Monday, increasing as the week progresses. However, Tuesday is not following this trend and has the highest average. Nevertheless, there is a increasing trend of attendence towards to weekend. The variance in attendance (how spread out the curves are) also seems to increase as the week progresses, suggesting that at the beginning of the week the attendance may be more predictable.

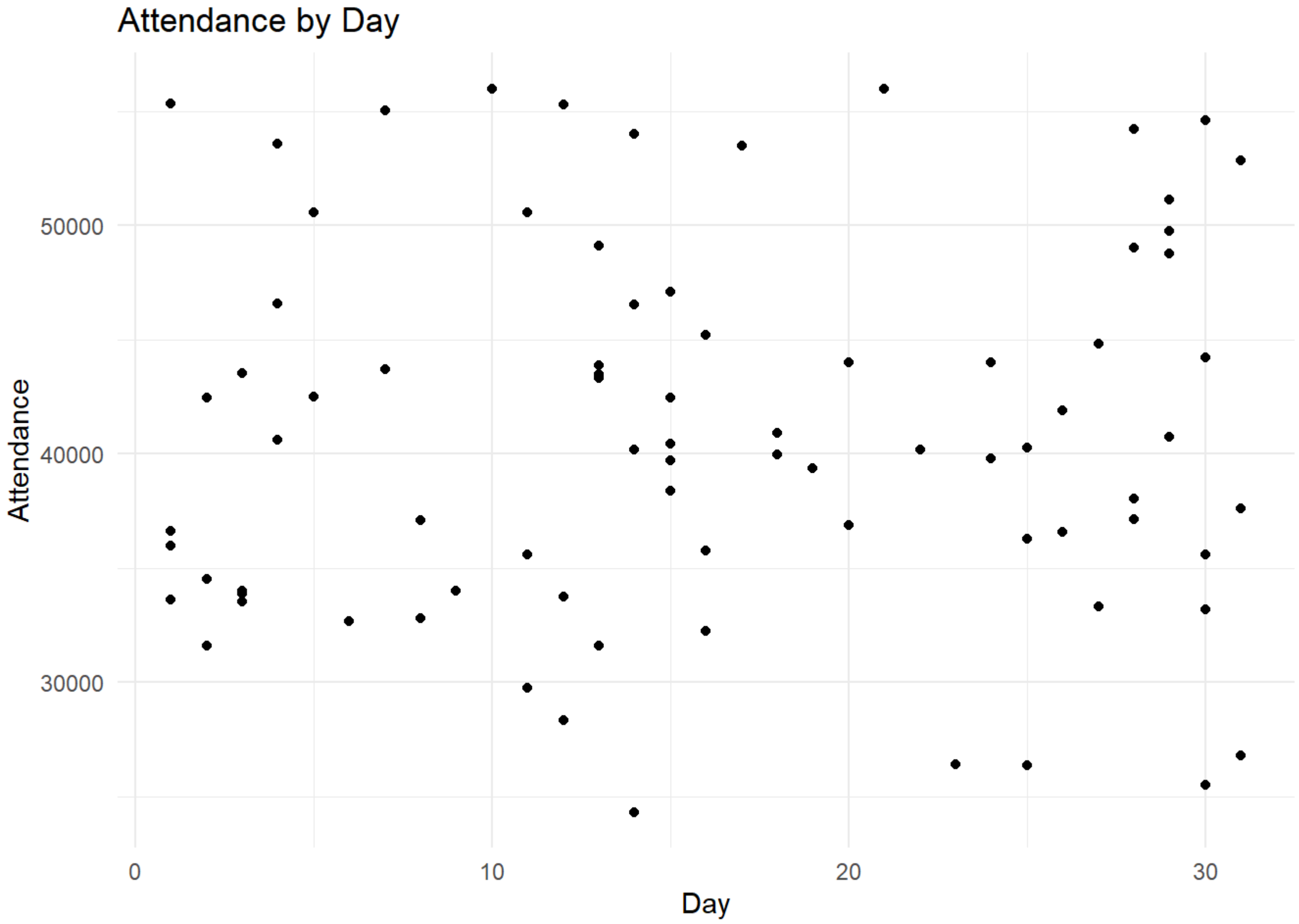
Show



As we discussed above, the average attendance is higher in the weekend than in the weekday. Also, the variance in attendance is higher in the weekday than in the weekend. This suggests that the attendance for the weekend may be more predictable.

Now, lets continue with the relationship between the attendence and the day of the month.

Show

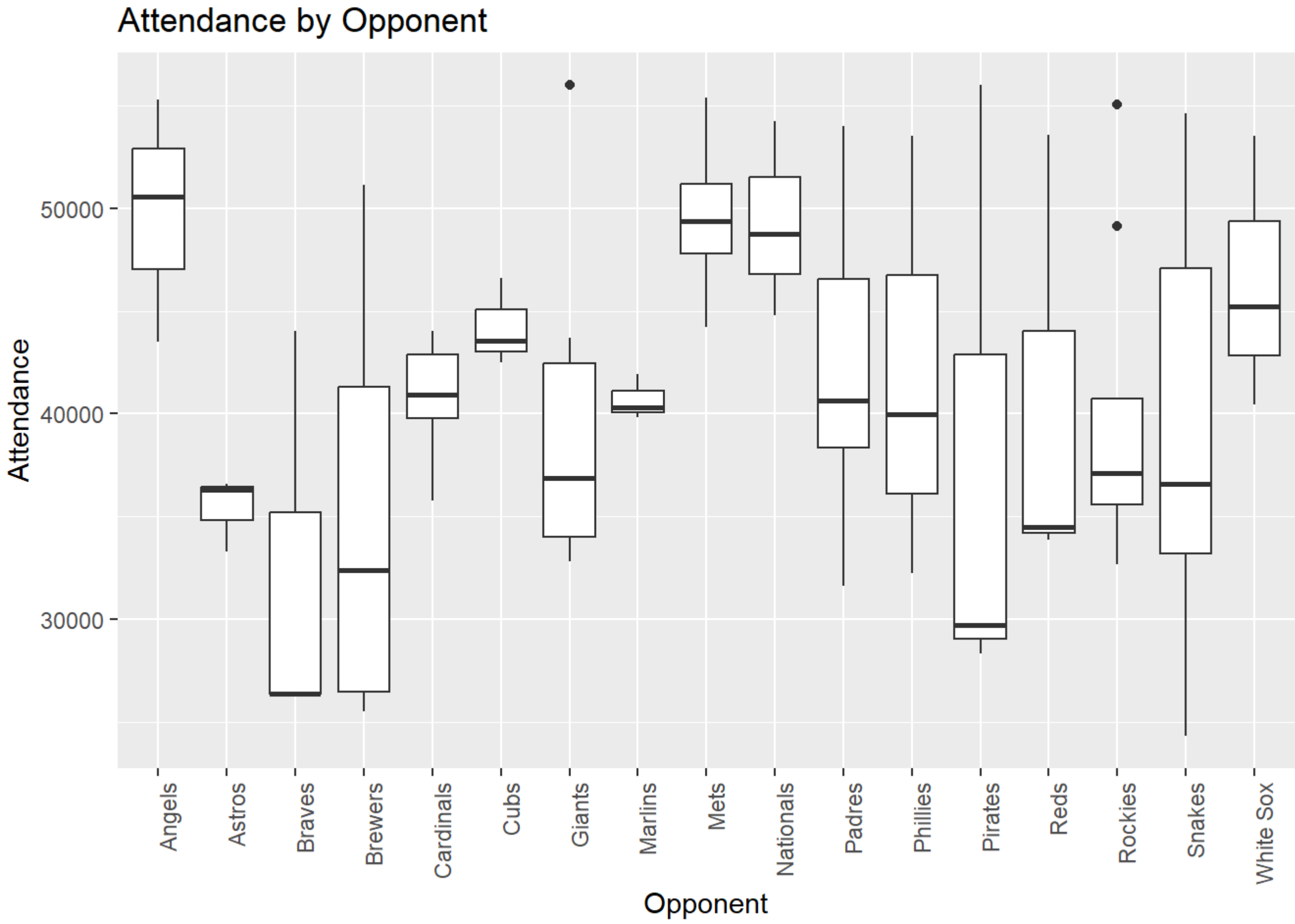


There does not appear to be a clear pattern or trend indicating an increase or decrease in attendance as the month progresses.

Opponent Team Analysis

Now, go on with the relationship between the attendence and the opponent team.

Show

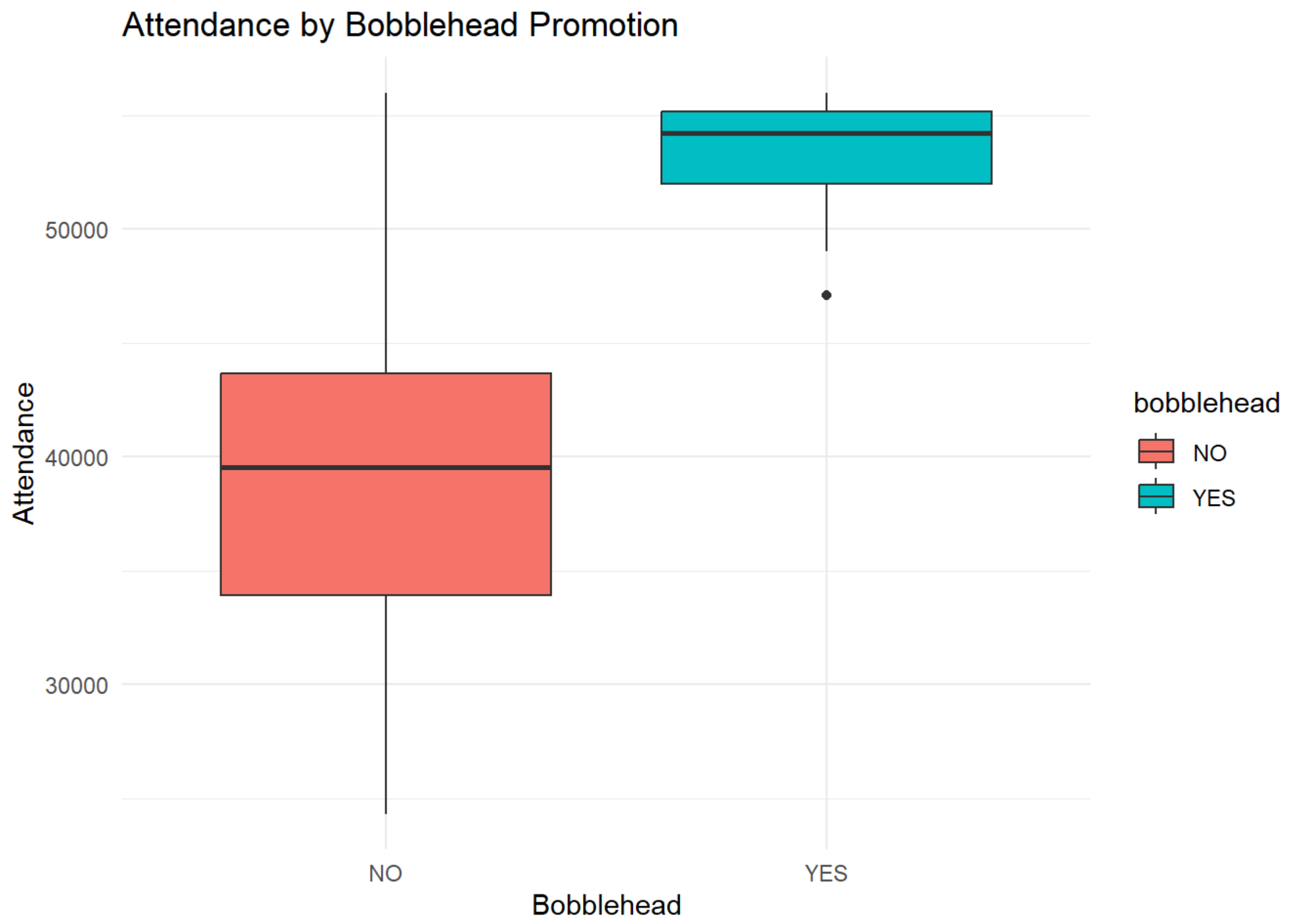


From the boxplot, we can say that for teams like Angels, Mets, and Nationals, which have higher average attendance, and for the teams like Astros, Rockies which have lower average attendance, the variance in attendance is lower. This suggests that the attendance for these teams may be more predictable. However, for some teams there is hgh variance in attendance which makes it hard to predict the attendance by looking at only the opponent team.

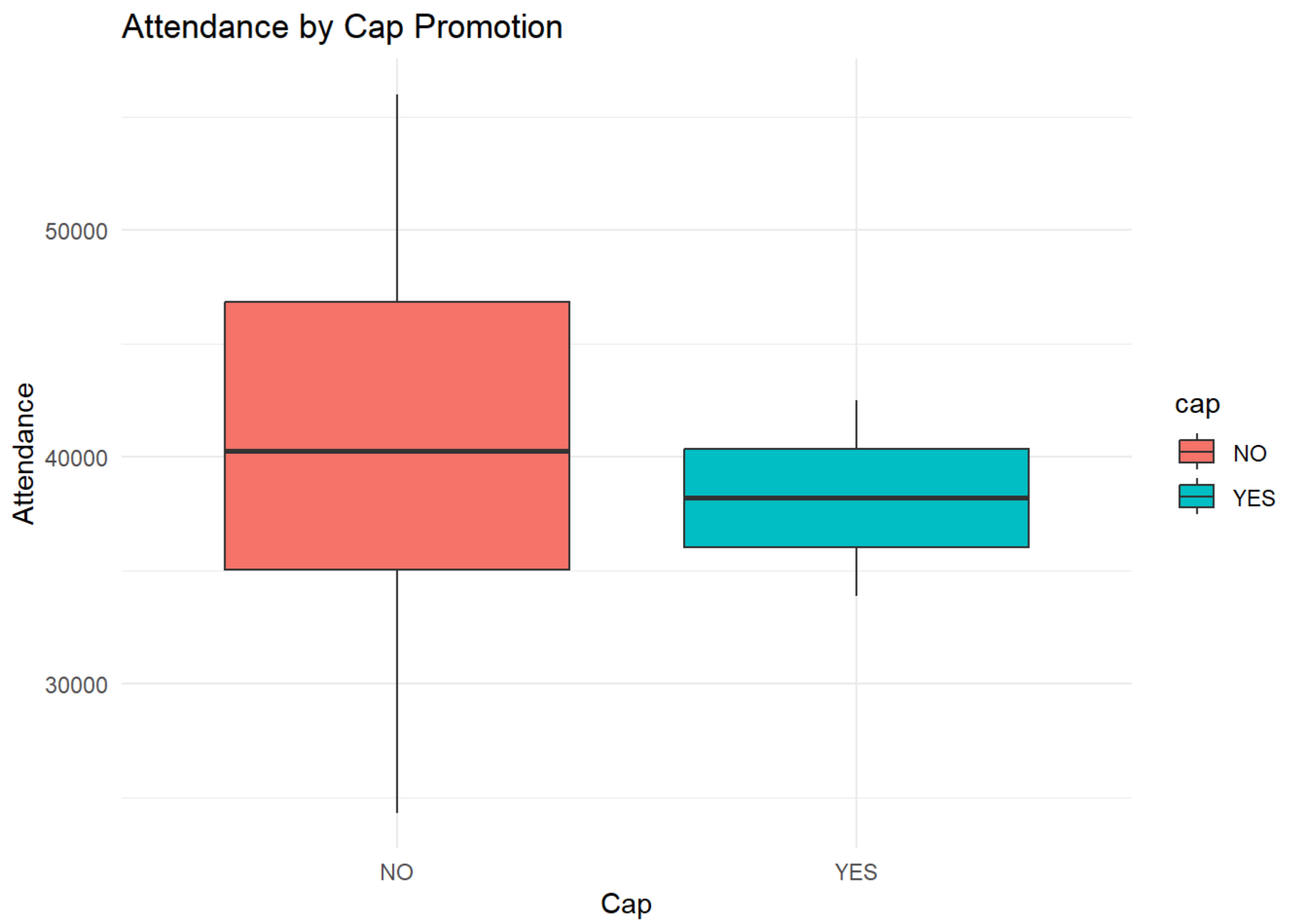
Promotions Analysis

Lets plot the boxplots for the promotions to see if they affect the attendence.

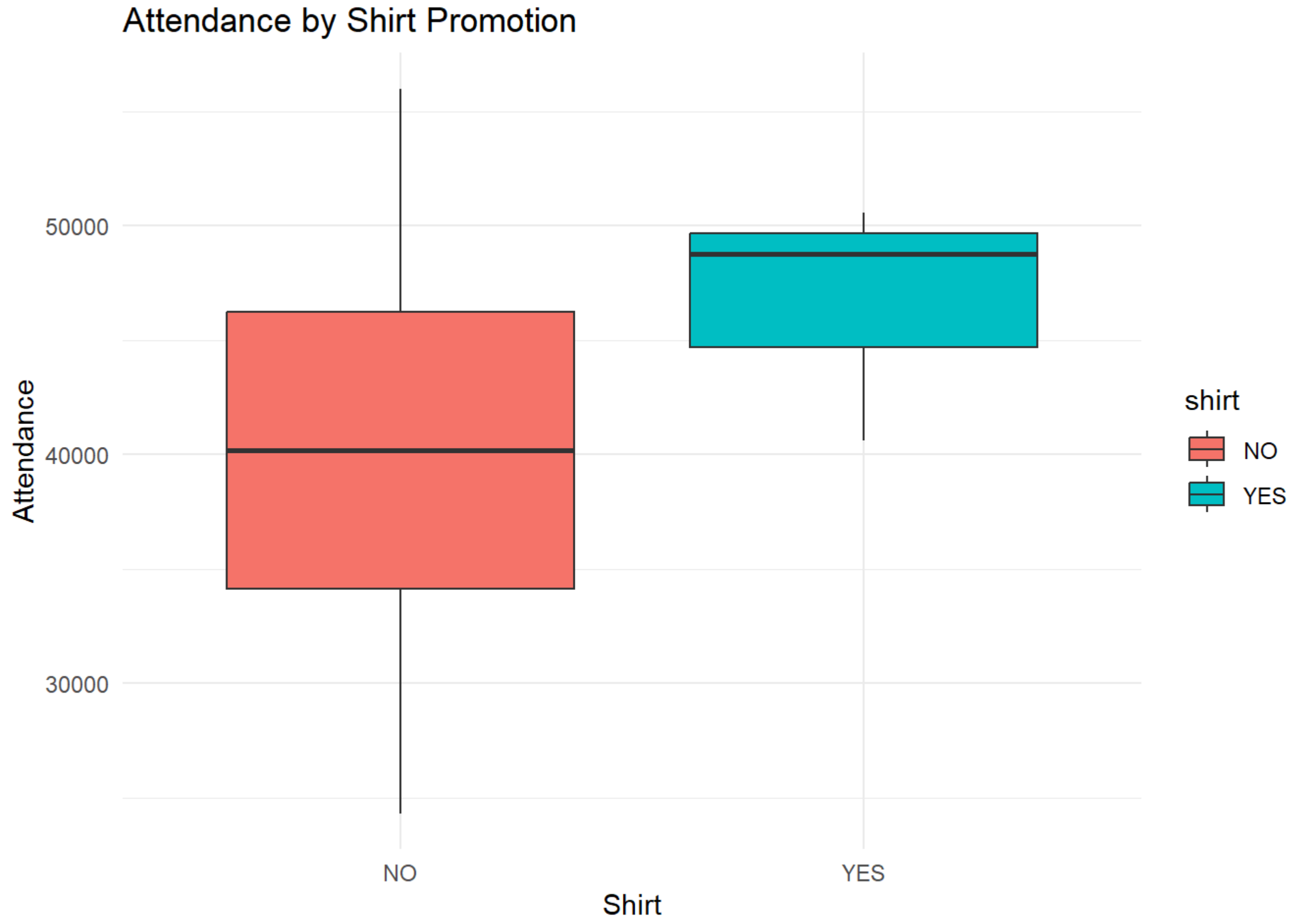
Show


Attendance by Bobblehead Promotion
Attendance
50000
40000
30000
NO
YES
Bobblehead
bobblehead
NO
YES

Show


Attendance by Cap Promotion
Attendance
50000
40000
30000
NO
YES
Cap
cap
NO
YES

Show

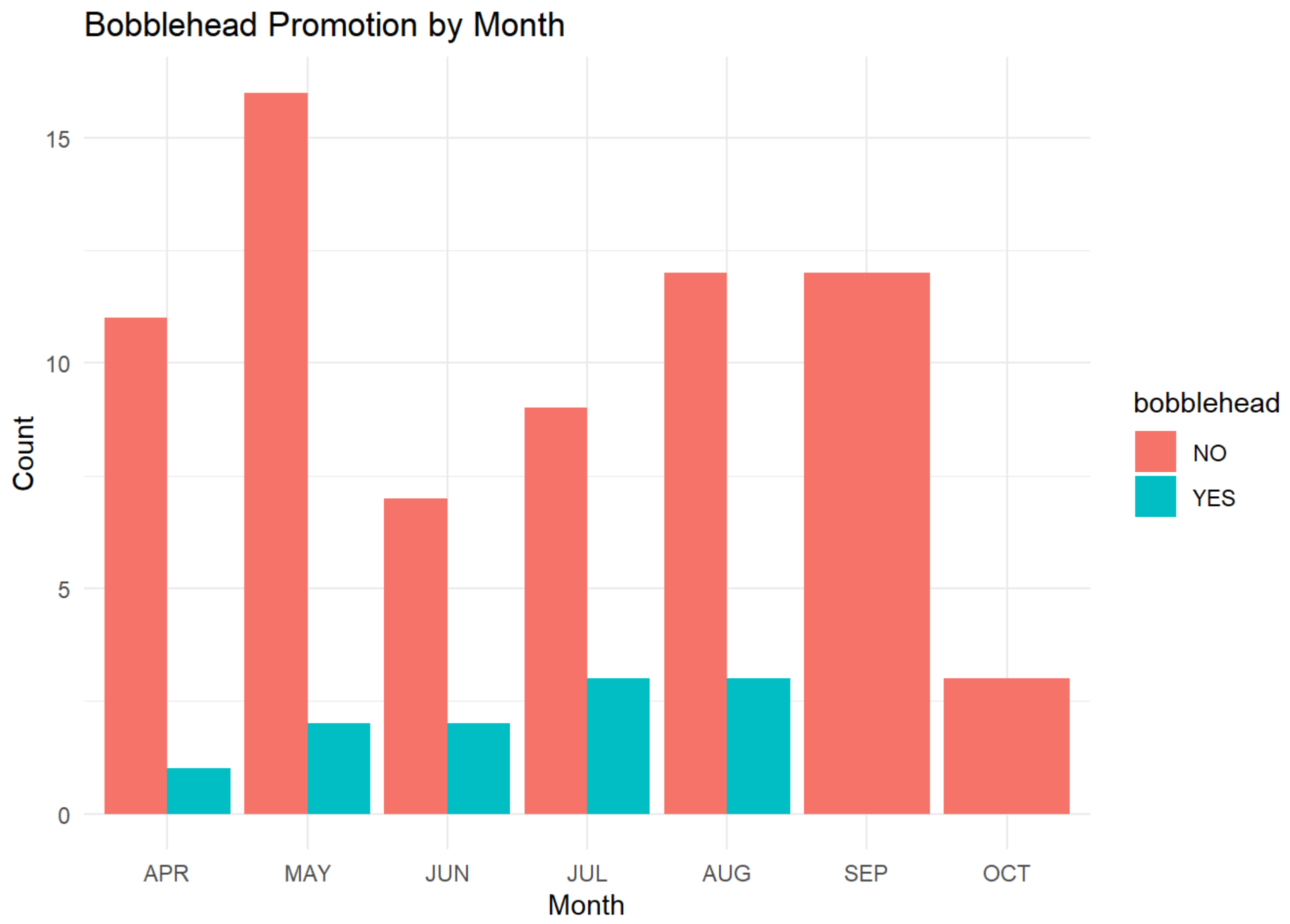


According to the boxplots, the average attendance is higher when there is a bobblehead and shirt promotion. However, the average attendance is lower when there is a cap promotion. Also, we noticed that there is a way more significant difference in the average attendance when there is a bobblehead promotion compared to the shirt promotion.

We plotted the data distribution of the attendence value given different promotion types.

Now we also want to find if these promotion types given in different months or day affect the attendences.

Show



look in more detail with numbers

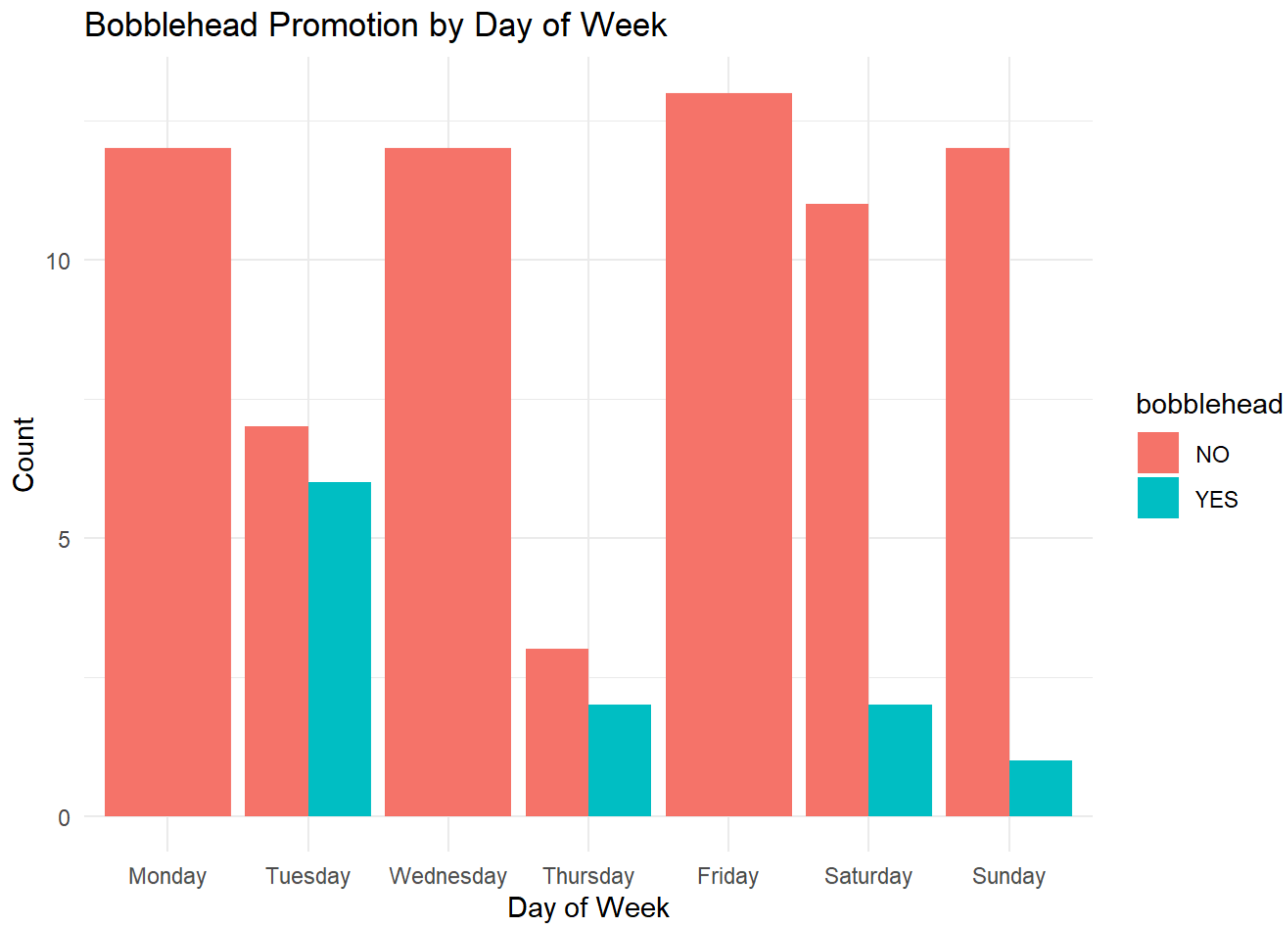
Show

Bobblehead Promotion by Month

month	NO	YES
APR	11	1
MAY	16	2
JUN	7	2
JUL	9	3
AUG	12	3
SEP	12	NA
OCT	3	NA

Now also check with respect to day of week

Show



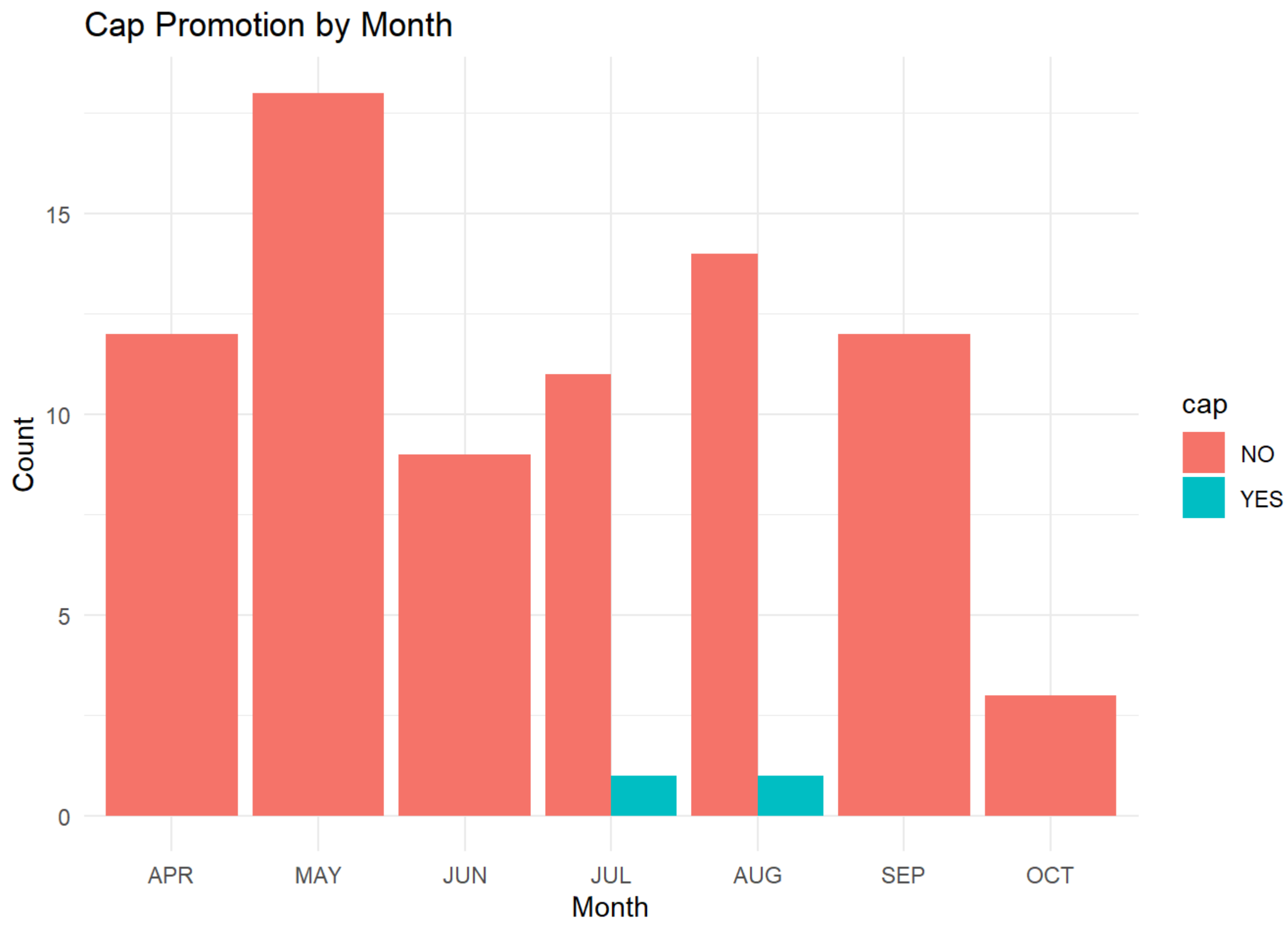
Show

Bobblehead Promotion in each day of week

day_of_week	NO	YES
Monday	12	NA
Tuesday	7	6
Wednesday	12	NA
Thursday	3	2
Friday	13	NA
Saturday	11	2
Sunday	12	1

Now we are going to do same analysis for cap promotion.

Show

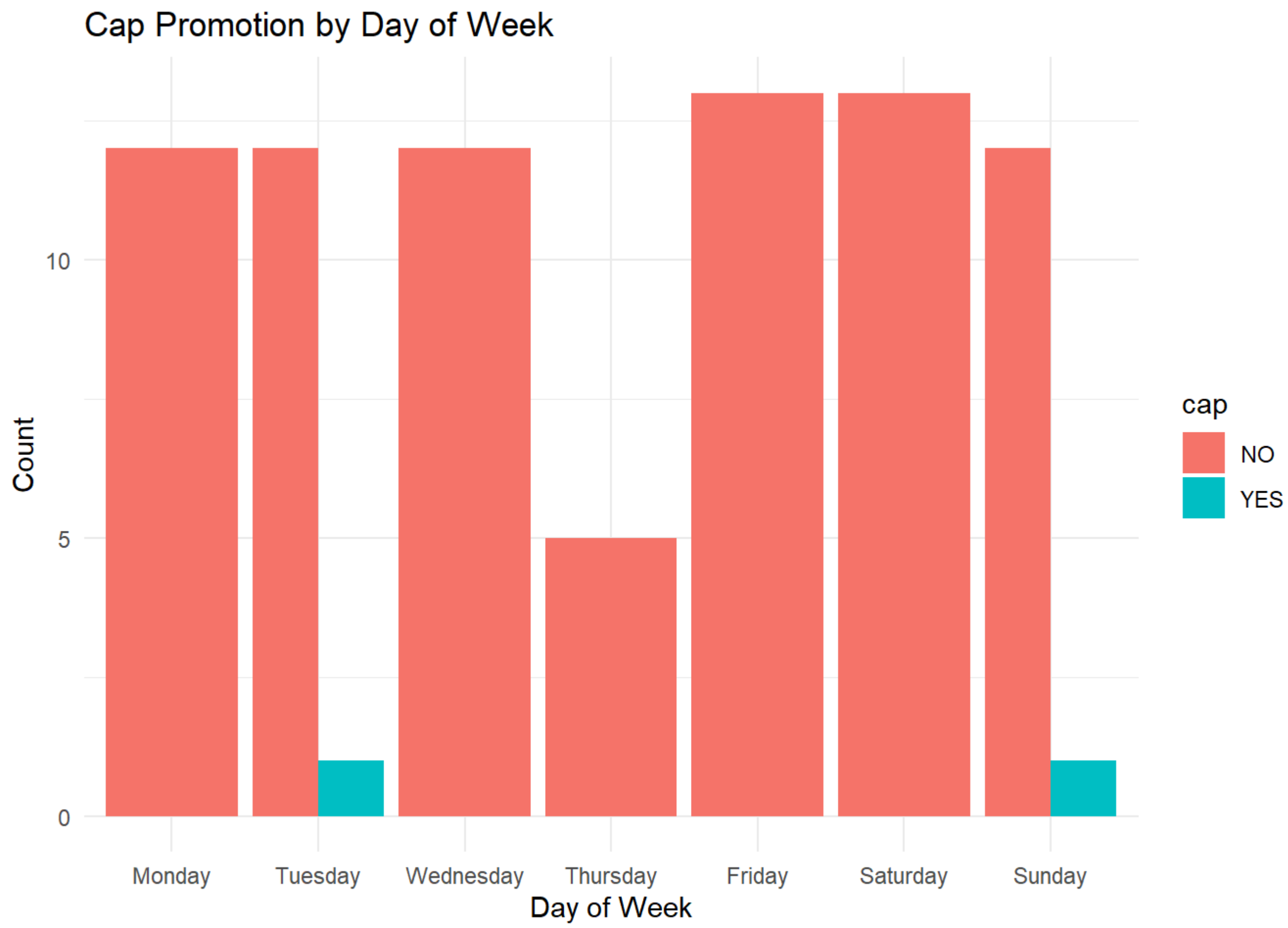


Show

Cap Promotion by Month

month	NO	YES
APR	12	NA
MAY	18	NA
JUN	9	NA
JUL	11	1
AUG	14	1
SEP	12	NA
OCT	3	NA

Show



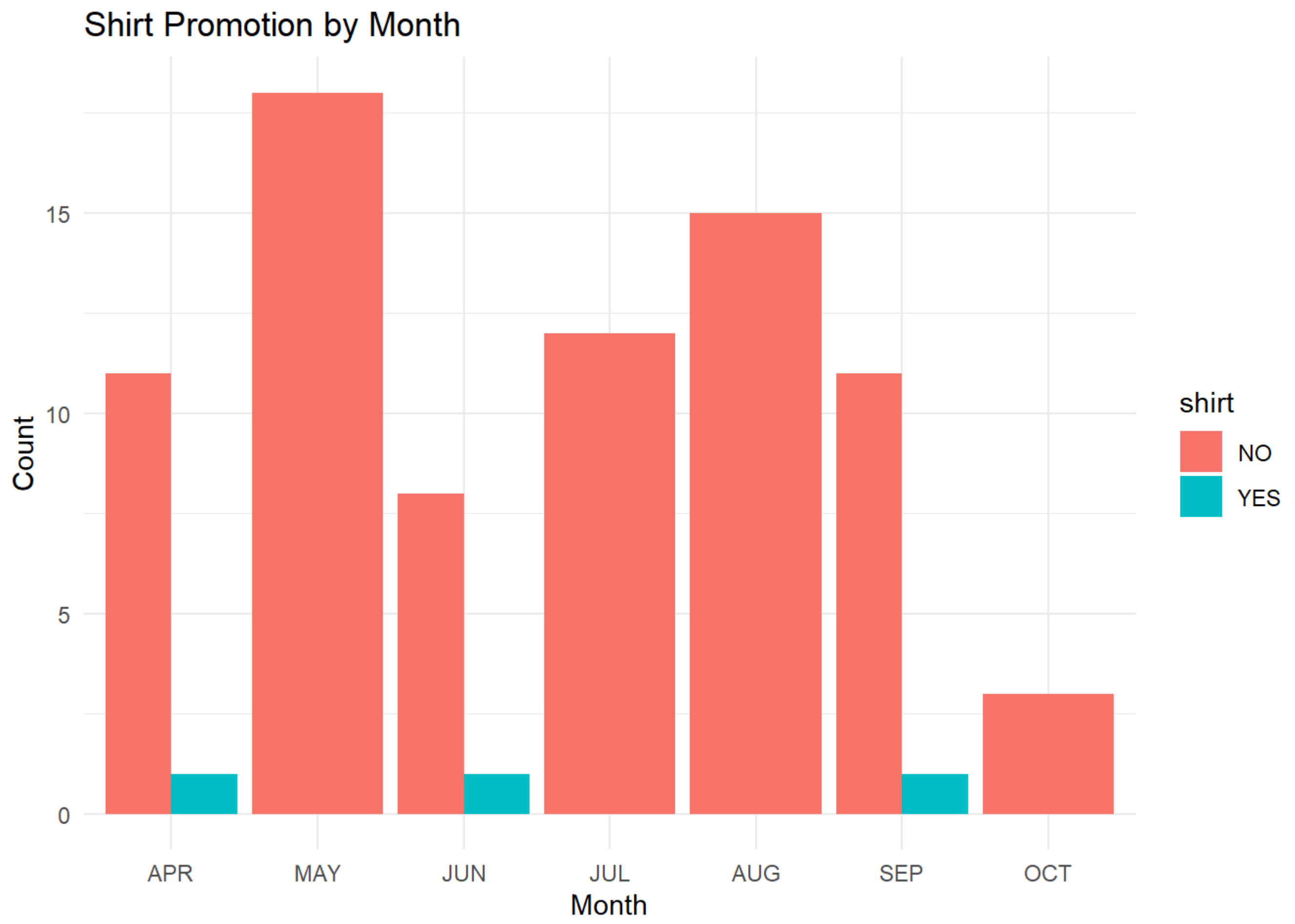
Show

Cap Promotion in each day of week

day_of_week	NO	YES
Monday	12	NA
Tuesday	12	1
Wednesday	12	NA
Thursday	5	NA
Friday	13	NA
Saturday	13	NA
Sunday	12	1

Continue with shirt promotion

Show

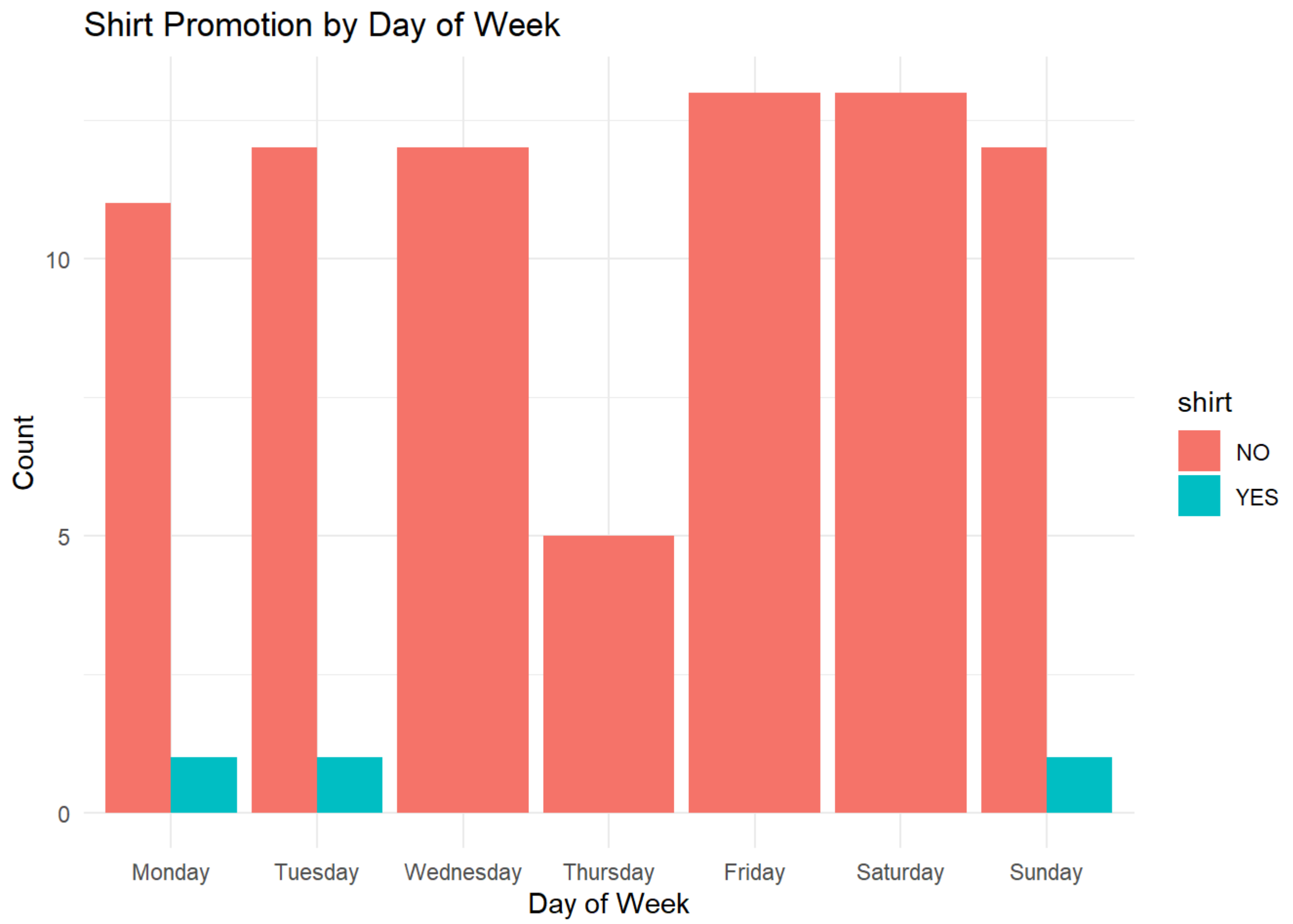


Show

Shirt Promotion by Month

month	NO	YES
APR	11	1
MAY	18	NA
JUN	8	1
JUL	12	NA
AUG	15	NA
SEP	11	1
OCT	3	NA

Show



Show

Shirt Promotion in each day of week

day_of_week	NO	YES
Monday	11	1
Tuesday	12	1
Wednesday	12	NA
Thursday	5	NA
Friday	13	NA
Saturday	13	NA
Sunday	12	1

Combinining the knowledge of other factors (such as month, opponent, etc) with the promotion types, we can make a better analysis.

Booblehead promotion is in April, May, June, July and August. It is also in the Tuesday, Thursday and weekend.

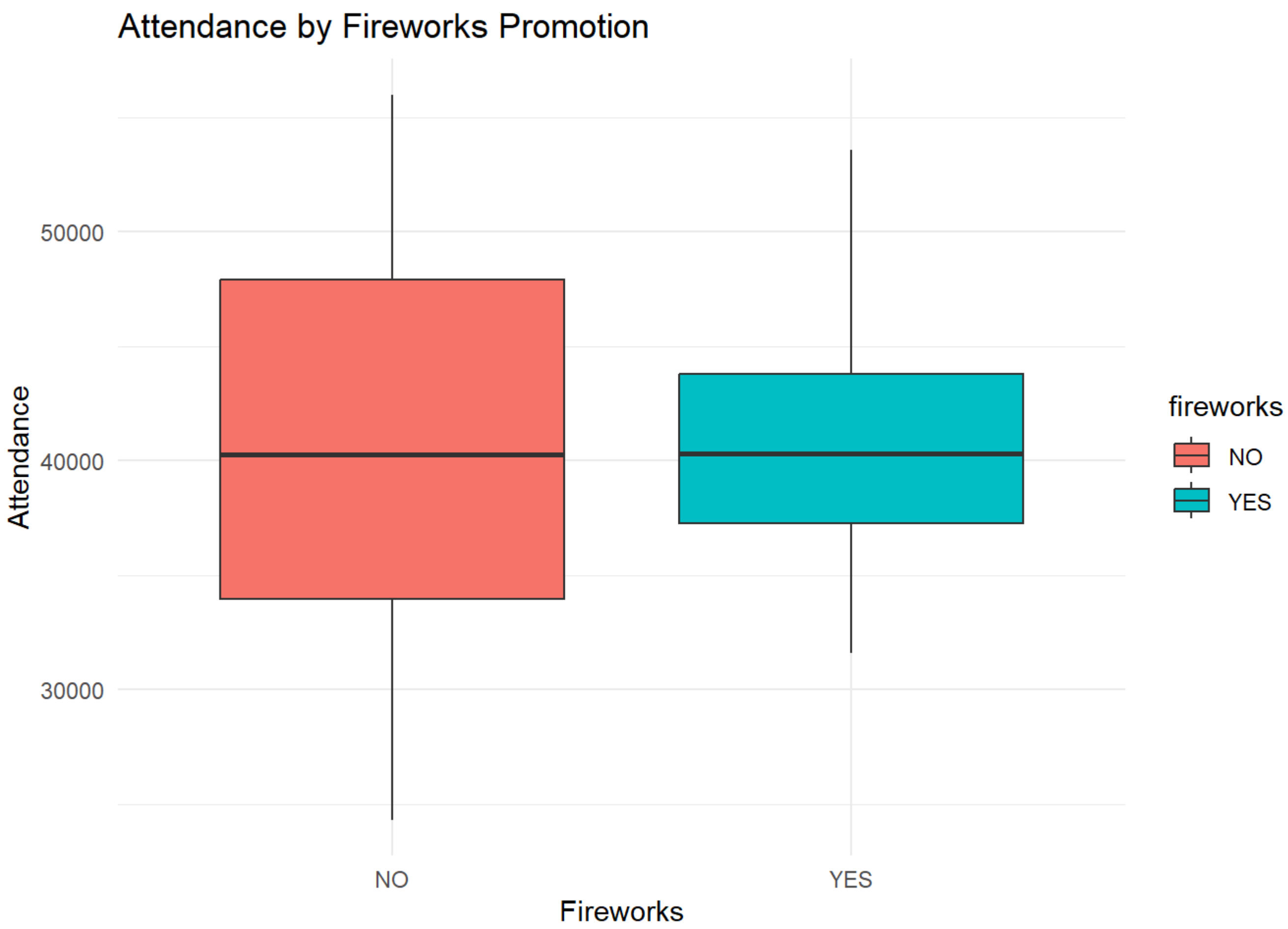
Cap promotion is in July and August. It is also in Tuesday and Sunday.

Shirt promotion is in April, June and September. It is also in Monday, Tuesday and Sunday.

We said above that attendance is higher with booblehead and shirt promotions but it might be also because of the months and days of the week.

Now we will check the relationship between the fireworks promotion and the attendence.

Show



Fireworks promotion does not seem to have a significant effect on the average attendance. The average attendance is almost the same when there is a fireworks promotion and when there is not. However, the variance in attendance is lower when there is a fireworks promotion than when there is not. This suggests that the attendance for the fireworks promotion may be more predictable.

3. Hypotheses

Now we will conduct some hypothesis testing.

H0: There is no relationship between the day_night and the attendence.

Show

```
## 
##  Chi-squared test for given probabilities
## 
## data:  .
## X-squared = 10.337, df = 1, p-value = 0.001304
```

Since the p-value (0.001304) is less than 0.05, we reject the null hypothesis. This suggests that there is evidence to conclude that there is an association between attendance and whether the game is played during the day or at night.

Now check for sky conditions.

H0: There is no relationship between the skies and the attendence.

Show

```
##
##  Chi-squared test for given probabilities
##
## data:  .
## X-squared = 107.19, df = 1, p-value < 0.00000000000000022
```

Since the p-value < 0.05, we reject the null hypothesis. This suggests that there is evidence to conclude that there is an association between attendance and the sky conditions.

We didnt check the relationship between the attendence and fireworks promotion. We will do it now.

H0: There is no relationship between the fireworks and the attendence.

Show

```
##
##  Chi-squared test for given probabilities
##
## data:  .
## X-squared = 0.025411, df = 1, p-value = 0.8733
```

Since the p-value > 0.05 we accept the null hypothesis. This suggests that there is no evidence to conclude that there is an association between attendance and fireworks promotion. Hence, we can say that fireworks promotion does not affect the attendance.

Previously, by looking at the graph we saw some relationship between the attendence and the temperature. Now we will check it. Since we are comparing two numerical variables, we will use correlation test. We can also use scatter plots and regression analysis to see the relationship between the attendence and the temperature.

We first start with the correlation test.

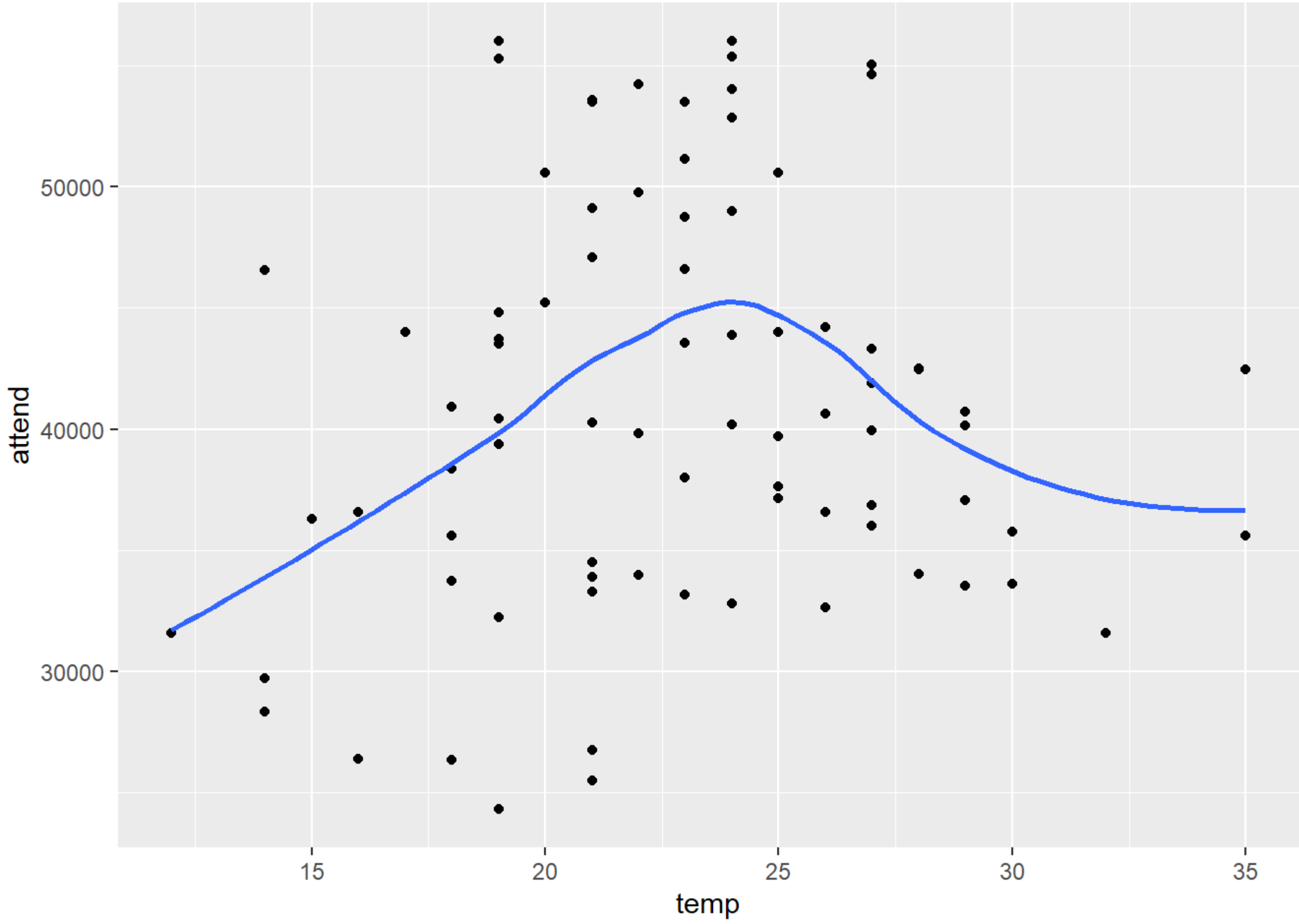
Show

```
##        attend  temp
## attend  1.000 0.097
## temp    0.097 1.000
```

The correlation coefficient is 0.1. We know that it is between (-1,1) and 1 suggest strong positive relationship. Hence, our result suggests that there is a weak positive relationship between the attendance and the temperature.However, it is only looking at the linear relationship. We can also look at the scatter plot and regression analysis to see if there is nonlinear relationship.

Show

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



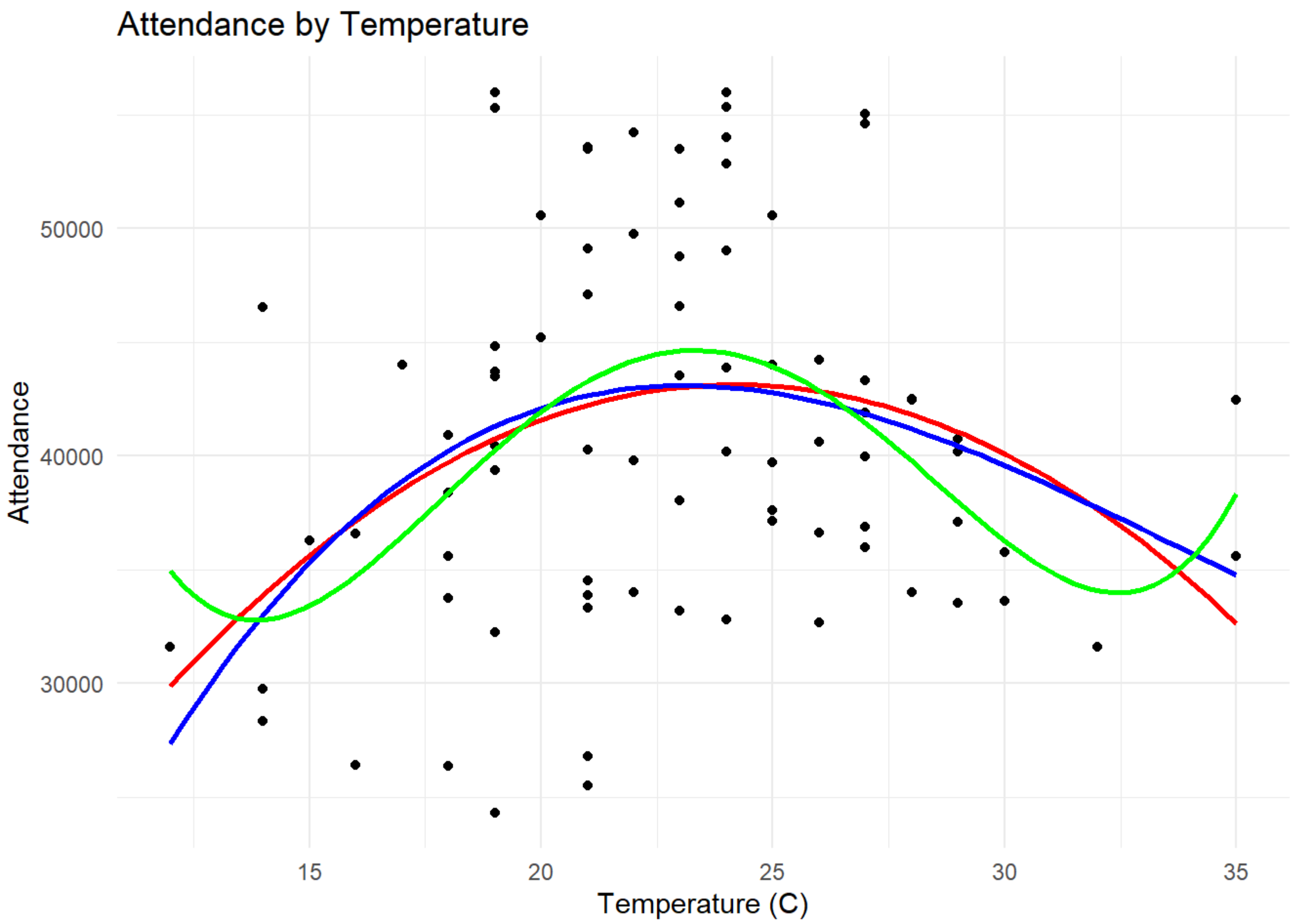
This shows there is a nonlinear relationship between the attendance and the temperature. The attendance is higher when the temperature is between 20 and 25. It is also lower when the temperature is below 20 and above 25.

Now, we will try to fit nonlinear model

Show

```
##
## Call:
## lm(formula = attend ~ poly(temp, 3), data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17138.3  -5642.3     68.4   5660.3  14704.2
##
## Coefficients:
##                 Estimate Std. Error t value             Pr(>|t|)
## (Intercept)      41040.1      880.1  46.629 < 0.0000000000000002 ***
## poly(temp, 3)1    7213.5     7921.2   0.911              0.36532
## poly(temp, 3)2  -24351.1     7921.2  -3.074              0.00292 **
## poly(temp, 3)3    5615.2     7921.2   0.709              0.48054
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7921 on 77 degrees of freedom
## Multiple R-squared:  0.1228, Adjusted R-squared:  0.08865
## F-statistic: 3.594 on 3 and 77 DF,  p-value: 0.01729
```

Show



This model show that there might be a non-linear relationship between temperature and attandence, primarily driven by quadratic term means that second degree polynomial should be the best fit model without overfit.However, the R-squared value is 0.1228 shows that approximately 12.28% of the variation in attendance can be explained by the temperature. This is relatively low suggesting that there are other factors that are influencing attendance.

---

3. Model

Now we will build a model to predict the attendance. First we plan to start by taking all variables into account and then we will remove the variables that are not significant.

Show

```
##
## Call:
## lm(formula = attend ~ ., data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9071.5 -3105.3   -83.3  1398.1 12467.6
##
## Coefficients: (1 not defined because of singularities)
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           49144.43   14120.70   3.480  0.00114 **
## monthMAY               4562.82    6161.08   0.741  0.46288
## monthJUN              -2780.38   12052.35  -0.231  0.81862
## monthJUL               4725.36    6220.19   0.760  0.45150
## monthAUG               8631.58    7717.22   1.118  0.26943
## monthSEP               4296.83    7739.53   0.555  0.58158
## monthOCT               6217.52    9532.41   0.652  0.51763
## day                     134.00     141.29   0.948  0.34811
## day_of_weekTuesday     8670.34    2943.86   2.945  0.00514 **
## day_of_weekWednesday      6.94    2764.44   0.003  0.99801
## day_of_weekThursday     460.46    3897.62   0.118  0.90650
## day_of_weekFriday    -18358.17    9210.84  -1.993  0.05247 .
## day_of_weekSaturday    3674.93    3341.26   1.100  0.27737
## day_of_weekSunday       892.73    4810.77   0.186  0.85364
## opponentAstros       -20678.17   14211.96  -1.455  0.15277
## opponentBraves       -18497.20   13741.48  -1.346  0.18517
## opponentBrewers      -22322.21   15155.70  -1.473  0.14791
## opponentCardinals    -12532.09   13146.23  -0.953  0.34565
## opponentCubs         -10529.08   12640.84  -0.833  0.40938
## opponentGiants       -16880.09   13263.72  -1.273  0.20983
## opponentMarlins      -19147.75   13897.97  -1.378  0.17525
## opponentMets          -4234.12    6444.47  -0.657  0.51459
## opponentNationals     -5850.86   13928.61  -0.420  0.67649
## opponentPadres       -11169.54   11310.93  -0.988  0.32880
## opponentPhillies     -13698.94   12586.97  -1.088  0.28237
## opponentPirates      -12451.05   13301.72  -0.936  0.35436
## opponentReds         -16473.41   11915.48  -1.383  0.17379
## opponentRockies      -16964.63   12946.55  -1.310  0.19687
## opponentSnakes       -19919.33   12719.45  -1.566  0.12450
## opponentWhite Sox      -928.32    5735.06  -0.162  0.87215
## temp                    -43.20     436.72  -0.099  0.92166
## skiesCloudy            -114.01    2385.81  -0.048  0.96210
## day_nightNight        -3662.30    3590.06  -1.020  0.31325
## capYES                -6503.20    5866.45  -1.109  0.27365
## shirtYES                949.27    4594.60   0.207  0.83727
## fireworksYES          20200.13    8352.77   2.418  0.01980 *
## bobbleheadYES          9395.63    3203.05   2.933  0.00531 **
## weeksweekend                NA         NA      NA       NA
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5979 on 44 degrees of freedom
## Multiple R-squared:  0.7145, Adjusted R-squared:  0.4809
## F-statistic: 3.058 on 36 and 44 DF,  p-value: 0.0002485
```

Comments about the base model:

The F-statistic and the p-value (3.058 and 0.0002485) respectively,suggest that the model as a whole is statistically significant, since p<0.05. It means that at least some of the predictors are likely to be useful in explaining the variation in attendance. The adjusted R-squared value is 0.4809, which means approximately the 48.09% of the variation in attendance can be explained by base model. Still half of the variation in attendance is not explained by the model.

From the table, the significant predictors are “monthAPR”, “day_of_weekTuesday”,“fireworksYES” and “bobbleheadYES” with p-value < 0.05, “day_of_weekFriday” is also very close to 0.05.

Also from the analysis above we know that temp is not in linear relationship with the attendence, so we will update model with temp^3.

Show

```
##
## Call:
## lm(formula = attend ~ month + day + day_of_week + opponent +
##     skies + day_night + cap + shirt + fireworks + bobblehead +
##     weeks + poly(temp, 3), data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8645.3 -2500.2   -81.8  1515.8 11192.7
##
## Coefficients: (1 not defined because of singularities)
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            52554.99   14278.93   3.681 0.000658 ***
## monthMAY               -4526.20    7362.45  -0.615 0.542024
## monthJUN               -8420.17   12188.41  -0.691 0.493470
## monthJUL               -4357.60    7462.57  -0.584 0.562393
## monthAUG                 163.40    8658.27   0.019 0.985032
## monthSEP               -3716.01    8468.85  -0.439 0.663066
## monthOCT               -2330.19   10205.96  -0.228 0.820508
## day                       63.38     144.11   0.440 0.662334
## day_of_weekTuesday      6983.76    2972.12   2.350 0.023561 *
## day_of_weekWednesday    -812.67    2716.00  -0.299 0.766250
## day_of_weekThursday      415.37    3796.13   0.109 0.913391
## day_of_weekFriday     -19352.50    8993.64  -2.152 0.037209 *
## day_of_weekSaturday     4101.74    3270.96   1.254 0.216781
## day_of_weekSunday       1916.87    4708.70   0.407 0.686008
## opponentAstros        -14544.84   14386.18  -1.011 0.317792
## opponentBraves        -21482.52   13578.32  -1.582 0.121125
## opponentBrewers       -17828.45   15399.51  -1.158 0.253516
## opponentCardinals      -8558.83   13342.93  -0.641 0.524715
## opponentCubs          -10359.56   12439.17  -0.833 0.409658
## opponentGiants        -13074.49   13400.35  -0.976 0.334808
## opponentMarlins       -16185.26   13808.50  -1.172 0.247754
## opponentMets           -3866.66    6725.78  -0.575 0.568425
## opponentNationals     -11711.15   14221.10  -0.824 0.414872
## opponentPadres         -9675.66   11422.46  -0.847 0.401756
## opponentPhillies       -9221.57   12652.58  -0.729 0.470151
## opponentPirates       -11939.64   13144.15  -0.908 0.368870
## opponentReds          -15519.31   11663.86  -1.331 0.190518
## opponentRockies       -13077.15   13053.22  -1.002 0.322162
## opponentSnakes        -16296.29   12868.78  -1.266 0.212369
## opponentWhite Sox      -1267.46    5671.00  -0.223 0.824232
## skiesCloudy             2046.14    2639.75   0.775 0.442607
## day_nightNight         -2630.62    3529.60  -0.745 0.460239
## capYES                 -4439.99    5803.51  -0.765 0.448518
## shirtYES                1808.07    4524.43   0.400 0.691459
## fireworksYES           22394.06    8209.51   2.728 0.009267 **
## bobbleheadYES          10048.85    3137.02   3.203 0.002593 **
```

```
## weeksweekend                    NA         NA      NA       NA
## poly(temp, 3)1         7546.30   19058.63   0.396 0.694145
## poly(temp, 3)2       -20823.65   10027.23  -2.077 0.043983 *
## poly(temp, 3)3         9905.31    8870.29   1.117 0.270477
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5814 on 42 degrees of freedom
## Multiple R-squared:  0.7422, Adjusted R-squared:  0.509
## F-statistic: 3.182 on 38 and 42 DF,  p-value: 0.0001739
```

Now we see that temp^2 is also significant. Furthermore, our Adjusted R-squared score increased and p-value decreased. This means that the model is better than the previous one.

Now we will remove the variables that are not significant and create the model 3.

Show

```
##
## Call:
## lm(formula = attend ~ day_of_week + month + fireworks + bobblehead +
##     poly(temp, 2), data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10082.1  -3272.2   -218.6   2439.8  13783.0
##
## Coefficients:
##                      Estimate Std. Error t value             Pr(>|t|)
## (Intercept)           36976.3     2917.4  12.674 < 0.0000000000000002 ***
## day_of_weekTuesday     7360.7     2510.2   2.932              0.00466 **
## day_of_weekWednesday    669.8     2395.0   0.280              0.78065
## day_of_weekThursday     474.8     3231.1   0.147              0.88363
## day_of_weekFriday    -12216.2     6668.5  -1.832              0.07162 .
## day_of_weekSaturday    6588.7     2370.3   2.780              0.00713 **
## day_of_weekSunday      6268.3     2436.1   2.573              0.01241 *
## monthMAY              -5002.8     2404.1  -2.081              0.04144 *
## monthJUN               3918.3     2931.5   1.337              0.18607
## monthJUL              -2155.1     3022.5  -0.713              0.47844
## monthAUG              -1394.7     3260.2  -0.428              0.67024
## monthSEP              -2437.4     4039.6  -0.603              0.54839
## monthOCT              -3445.5     4978.5  -0.692              0.49139
## fireworksYES          17343.7     6181.6   2.806              0.00664 **
## bobbleheadYES         10279.8     2300.5   4.468            0.0000328 ***
## poly(temp, 2)1         7623.8    11646.7   0.655              0.51508
## poly(temp, 2)2       -17226.7     6923.6  -2.488              0.01546 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5660 on 64 degrees of freedom
## Multiple R-squared:  0.6278, Adjusted R-squared:  0.5347
## F-statistic: 6.746 on 16 and 64 DF,  p-value: 0.00000001244
```

Again Adjusted R-squared score increased and p-value decreased. This means that the model is better than the previous one.

Include some interactions to the model to see if it is better.

Model 4 includes an interaction between month and a 3rd-degree polynomial of temp, and the main effects of day_of_week:day_night, fireworks, and bobblehead.

Show

```
##
## Call:
## lm(formula = attend ~ fireworks + bobblehead + month:poly(temp,
##     3) + day_of_week:day_night, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8361.3 -2604.2  -379.8  2274.5 11398.8
##
## Coefficients: (5 not defined because of singularities)
##                                    Estimate Std. Error t value     Pr(>|t|)
## (Intercept)                           45016       6697   6.722 0.0000000196
## fireworksYES                          16677       7129   2.339     0.023521
## bobbleheadYES                         11734       3046   3.852     0.000347
## monthAPR:poly(temp, 3)1              -13970      98029  -0.143     0.887277
## monthMAY:poly(temp, 3)1               62803     107019   0.587     0.560061
## monthJUN:poly(temp, 3)1              386063     434216   0.889     0.378382
## monthJUL:poly(temp, 3)1              233431     338800   0.689     0.494144
## monthAUG:poly(temp, 3)1               31137     172198   0.181     0.857272
## monthSEP:poly(temp, 3)1               42632      74422   0.573     0.569423
## monthOCT:poly(temp, 3)1               -4584      62547  -0.073     0.941883
## monthAPR:poly(temp, 3)2              -54443      98459  -0.553     0.582861
## monthMAY:poly(temp, 3)2               55569      97974   0.567     0.573237
## monthJUN:poly(temp, 3)2               -5993     151312  -0.040     0.968572
## monthJUL:poly(temp, 3)2               39359     117059   0.336     0.738164
## monthAUG:poly(temp, 3)2              -16711      64690  -0.258     0.797265
## monthSEP:poly(temp, 3)2              -82943      65874  -1.259     0.214076
## monthOCT:poly(temp, 3)2              -16784     108119  -0.155     0.877286
## monthAPR:poly(temp, 3)3              -26097      46079  -0.566     0.573793
## monthMAY:poly(temp, 3)3               30460      68212   0.447     0.657204
## monthJUN:poly(temp, 3)3              390641     383367   1.019     0.313324
## monthJUL:poly(temp, 3)3              182803     284975   0.641     0.524269
## monthAUG:poly(temp, 3)3               23524     143915   0.163     0.870843
## monthSEP:poly(temp, 3)3               53244      36027   1.478     0.145969
## monthOCT:poly(temp, 3)3                  NA         NA      NA           NA
## day_of_weekMonday:day_nightDay           NA         NA      NA           NA
## day_of_weekTuesday:day_nightDay       11454       8118   1.411     0.164734
## day_of_weekWednesday:day_nightDay     -9016       7086  -1.272     0.209355
## day_of_weekThursday:day_nightDay         NA         NA      NA           NA
## day_of_weekFriday:day_nightDay           NA         NA      NA           NA
## day_of_weekSaturday:day_nightDay       1726      10349   0.167     0.868215
## day_of_weekSunday:day_nightDay        -3058       5346  -0.572     0.569976
## day_of_weekMonday:day_nightNight      -9104       5120  -1.778     0.081723
## day_of_weekTuesday:day_nightNight     -3457       4869  -0.710     0.481165
## day_of_weekWednesday:day_nightNight   -8809       5341  -1.649     0.105625
## day_of_weekThursday:day_nightNight    -8480       5508  -1.540     0.130213
## day_of_weekFriday:day_nightNight     -21856       9256  -2.361     0.022313
## day_of_weekSaturday:day_nightNight    -3520       5225  -0.674     0.503686
```

```
## day_of_weekSunday:day_nightNight           NA          NA      NA            NA
##
## (Intercept)                          ***
## fireworksYES                         *
## bobbleheadYES                        ***
## monthAPR:poly(temp, 3)1
## monthMAY:poly(temp, 3)1
## monthJUN:poly(temp, 3)1
## monthJUL:poly(temp, 3)1
## monthAUG:poly(temp, 3)1
## monthSEP:poly(temp, 3)1
## monthOCT:poly(temp, 3)1
## monthAPR:poly(temp, 3)2
## monthMAY:poly(temp, 3)2
## monthJUN:poly(temp, 3)2
## monthJUL:poly(temp, 3)2
## monthAUG:poly(temp, 3)2
## monthSEP:poly(temp, 3)2
## monthOCT:poly(temp, 3)2
## monthAPR:poly(temp, 3)3
## monthMAY:poly(temp, 3)3
## monthJUN:poly(temp, 3)3
## monthJUL:poly(temp, 3)3
## monthAUG:poly(temp, 3)3
## monthSEP:poly(temp, 3)3
## monthOCT:poly(temp, 3)3
## day_of_weekMonday:day_nightDay
## day_of_weekTuesday:day_nightDay
## day_of_weekWednesday:day_nightDay
## day_of_weekThursday:day_nightDay
## day_of_weekFriday:day_nightDay
## day_of_weekSaturday:day_nightDay
## day_of_weekSunday:day_nightDay
## day_of_weekMonday:day_nightNight     .
## day_of_weekTuesday:day_nightNight
## day_of_weekWednesday:day_nightNight
## day_of_weekThursday:day_nightNight
## day_of_weekFriday:day_nightNight     *
## day_of_weekSaturday:day_nightNight
## day_of_weekSunday:day_nightNight
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5743 on 48 degrees of freedom
## Multiple R-squared:  0.7126, Adjusted R-squared:  0.521
## F-statistic:  3.72 on 32 and 48 DF,  p-value: 0.00002053
```

Model 5 adds interactions between month and a 3rd-degree polynomial of temperature, day of the week and day/night, opponent and day of the week, and fireworks and shirt. Alao it includes the main effects of day of the week, month, fireworks, and bobblehead.

Show

```
##
## Call:
## lm(formula = attend ~ day_of_week + month + fireworks + bobblehead +
##     month:poly(temp, 3) + day_of_week:day_night + day_of_week:opponent +
##     fireworks:shirt, data = d)
##
## Residuals:
##                           1                          2
##     0.0000000000027307921    -0.0000000000000062826
##                           3                          4
##     0.0000000000013668014    -0.0000000000144328257
##                           5                          6
##     0.0000000000284785563    -0.0000000000269651499
##                           7                          8
##    -0.0000000000022681942     0.0000000000003969502
##                           9                         10
##    -0.0000000000000005235     0.0000000000005640046
##                          11                         12
##     0.0000000000047461094     0.0000000000085064155
##                          13                         14
##     0.0000000000124253549    -0.0000000000277813262
##                          15                         16
##     0.0000000000030796341     0.0000000000566380657
##                          17                         18
## -1932.0000000005252331903     0.0000000001364507071
##                          19                         20
##    -0.0000000000756601202    -0.0000000000305544478
##                          21                         22
##  1932.0000000005911715562    -0.0000000001250780470
##                          23                         24
##    -0.0000000000074341969    -0.0000000000049986338
##                          25                         26
##     0.0000000000085039874    -0.0000000000012880022
##                          27                         28
##    -0.0000000000043216295     0.0000000000012778668
##                          29                         30
##    -0.0000000000024564588    -0.0000000000052132342
##                          31                         32
##     0.0000000000702015286     0.0000000000464556100
##                          33                         34
##    -0.0000000000067324900    -0.0000000000206433488
##                          35                         36
##    -0.0000000001092457188    -0.0000000000185314384
##                          37                         38
##    -0.0000000000006836976     0.0000000000406054350
##                          39                         40
##    -0.0000000000016695486    -0.0000000000001519395
##                          41                         42
```

```
##     0.0000000000019088576    -0.0000000000116800197
##                        43                        44
##     0.0000000000000401906     0.0000000000188444314
##                        45                        46
##    -0.0000000000326017156     0.0000000000237985127
##                        47                        48
##    -0.0000000000001517164     0.0000000000073625256
##                        49                        50
##    -0.0000000000026652652     0.0000000000151249525
##                        51                        52
##    -0.0000000000209400009     0.0000000000040998938
##                        53                        54
##    -0.0000000000010002350    -0.0000000000029874455
##                        55                        56
##     0.0000000000006853497    -0.0000000000022594597
##                        57                        58
##     0.0000000000041571607     0.0000000000071552276
##                        59                        60
##    -0.0000000000138057792     0.0000000000258997685
##                        61                        62
##    -0.0000000000049950532    -0.0000000000012844521
##                        63                        64
##    -0.0000000000012821429    -0.0000000000024410383
##                        65                        66
##     0.0000000000000589794    -0.0000000000123689188
##                        67                        68
##     0.00000000001132799232    -0.0000000000031112403
##                        69                        70
##    -0.0000000000044040379     0.0000000000029059230
##                        71                        72
##     0.0000000000027191491     0.0000000000024115256
##                        73                        74
## -1932.0000000000005934452929     0.0000000000001246486070
##                        75                        76
##     0.0000000000074047922    -0.0000000000494189089
##                        77                        78
##  1932.0000000000005138645065    -0.0000000000001207771174
##                        79                        80
##     0.0000000000009033121    -0.0000000000003132397
##                        81
##    -0.0000000000005000137
##
## Coefficients: (77 not defined because of singularities)
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               28871.9    60988.3   0.473    0.719
## day_of_weekTuesday                          729.8    22932.4   0.032    0.980
## day_of_weekWednesday                       6857.9     8974.6   0.764    0.585
## day_of_weekThursday                       29776.0    10929.0   2.724    0.224
```

```
## day_of_weekFriday                       -32768.0   16393.6  -1.999    0.295
## day_of_weekSaturday                     224111.0  248560.6   0.902    0.533
## day_of_weekSunday                         8229.0    7728.0   1.065    0.480
## monthMAY                               -146037.1   69781.1  -2.093    0.284
## monthJUN                               -253378.1  296158.9  -0.856    0.549
## monthJUL                                 64459.0  115396.8   0.559    0.676
## monthAUG                               -220235.8  313659.2  -0.702    0.610
## monthSEP                                -96247.8  101001.7  -0.953    0.515
## monthOCT                                119697.5   88410.7   1.354    0.405
## fireworksYES                             24103.0   10929.0   2.205    0.271
## bobbleheadYES                             7368.0    7728.0   0.953    0.515
## monthAPR:poly(temp, 3)1                -258904.8 1043521.7  -0.248    0.845
## monthMAY:poly(temp, 3)1               -3541784.4 1486308.8  -2.383    0.253
## monthJUN:poly(temp, 3)1               -5582932.3 7511848.3  -0.743    0.593
## monthJUL:poly(temp, 3)1               -3614668.7 3419490.3  -1.057    0.482
## monthAUG:poly(temp, 3)1                4454121.1 4844616.0   0.919    0.527
## monthSEP:poly(temp, 3)1                1178084.0  784872.4   1.501    0.374
## monthOCT:poly(temp, 3)1                -785531.0  321042.3  -2.447    0.247
## monthAPR:poly(temp, 3)2                -157793.7  899309.6  -0.175    0.889
## monthMAY:poly(temp, 3)2               -2727057.9 1105564.0  -2.467    0.245
## monthJUN:poly(temp, 3)2               -4192821.2 5123985.5  -0.818    0.563
## monthJUL:poly(temp, 3)2                  59400.7  387541.8   0.153    0.903
## monthAUG:poly(temp, 3)2               -2360123.5 2939495.5  -0.803    0.569
## monthSEP:poly(temp, 3)2                -839489.8  544648.3  -1.541    0.366
## monthOCT:poly(temp, 3)2                       NA        NA      NA       NA
## monthAPR:poly(temp, 3)3                   2946.7  317070.7   0.009    0.994
## monthMAY:poly(temp, 3)3               -2280015.2  991614.0  -2.299    0.261
## monthJUN:poly(temp, 3)3               -3063377.4 4528587.7  -0.676    0.621
## monthJUL:poly(temp, 3)3               -2913624.5 2679505.9  -1.087    0.473
## monthAUG:poly(temp, 3)3                2887112.1 2418318.9   1.194    0.444
## monthSEP:poly(temp, 3)3                 301855.4  194592.7   1.551    0.365
## monthOCT:poly(temp, 3)3                       NA        NA      NA       NA
## day_of_weekMonday:day_nightNight              NA        NA      NA       NA
## day_of_weekTuesday:day_nightNight        -1915.8   18228.9  -0.105    0.933
## day_of_weekWednesday:day_nightNight     -12460.9    9135.1  -1.364    0.403
## day_of_weekThursday:day_nightNight            NA        NA      NA       NA
## day_of_weekFriday:day_nightNight              NA        NA      NA       NA
## day_of_weekSaturday:day_nightNight     -229460.0  247778.5  -0.926    0.524
## day_of_weekSunday:day_nightNight         32907.3   19454.5   1.692    0.340
## day_of_weekMonday:opponentAstros              NA        NA      NA       NA
## day_of_weekTuesday:opponentAstros             NA        NA      NA       NA
## day_of_weekWednesday:opponentAstros           NA        NA      NA       NA
## day_of_weekThursday:opponentAstros            NA        NA      NA       NA
## day_of_weekFriday:opponentAstros       -170330.8   93282.0  -1.826    0.319
## day_of_weekSaturday:opponentAstros      -79412.8   48537.1  -1.636    0.349
## day_of_weekSunday:opponentAstros        -15818.0    5464.5  -2.895    0.212
## day_of_weekMonday:opponentBraves        -27349.7   17436.1  -1.569    0.361
## day_of_weekTuesday:opponentBraves       -10259.8   12116.4  -0.847    0.553
```

```
## day_of_weekWednesday:opponentBraves       -24023.5   14735.7  -1.630    0.350
## day_of_weekThursday:opponentBraves              NA        NA      NA       NA
## day_of_weekFriday:opponentBraves                NA        NA      NA       NA
## day_of_weekSaturday:opponentBraves              NA        NA      NA       NA
## day_of_weekSunday:opponentBraves                NA        NA      NA       NA
## day_of_weekMonday:opponentBrewers         -16529.5   16540.7  -0.999    0.500
## day_of_weekTuesday:opponentBrewers         -9590.5   10722.0  -0.894    0.535
## day_of_weekWednesday:opponentBrewers       -9783.0    7728.0  -1.266    0.426
## day_of_weekThursday:opponentBrewers       -43898.0   13385.3  -3.280    0.188
## day_of_weekFriday:opponentBrewers               NA        NA      NA       NA
## day_of_weekSaturday:opponentBrewers             NA        NA      NA       NA
## day_of_weekSunday:opponentBrewers               NA        NA      NA       NA
## day_of_weekMonday:opponentCardinals             NA        NA      NA       NA
## day_of_weekTuesday:opponentCardinals            NA        NA      NA       NA
## day_of_weekWednesday:opponentCardinals          NA        NA      NA       NA
## day_of_weekThursday:opponentCardinals     -27808.0   13385.3  -2.078    0.286
## day_of_weekFriday:opponentCardinals        14520.5   11120.9   1.306    0.416
## day_of_weekSaturday:opponentCardinals      20420.0    7728.0   2.642    0.230
## day_of_weekSunday:opponentCardinals        -4162.2    6646.5  -0.626    0.644
## day_of_weekMonday:opponentCubs                  NA        NA      NA       NA
## day_of_weekTuesday:opponentCubs                 NA        NA      NA       NA
## day_of_weekWednesday:opponentCubs               NA        NA      NA       NA
## day_of_weekThursday:opponentCubs                NA        NA      NA       NA
## day_of_weekFriday:opponentCubs             -3176.9   37966.8  -0.084    0.947
## day_of_weekSaturday:opponentCubs           -3441.9   35527.3  -0.097    0.939
## day_of_weekSunday:opponentCubs             15545.1    9580.3   1.623    0.352
## day_of_weekMonday:opponentGiants           19401.0    5464.5   3.550    0.175
## day_of_weekTuesday:opponentGiants          -8088.3    9790.8  -0.826    0.560
## day_of_weekWednesday:opponentGiants       -12130.3    9790.8  -1.239    0.432
## day_of_weekThursday:opponentGiants              NA        NA      NA       NA
## day_of_weekFriday:opponentGiants                NA        NA      NA       NA
## day_of_weekSaturday:opponentGiants              NA        NA      NA       NA
## day_of_weekSunday:opponentGiants                NA        NA      NA       NA
## day_of_weekMonday:opponentMarlins               NA        NA      NA       NA
## day_of_weekTuesday:opponentMarlins              NA        NA      NA       NA
## day_of_weekWednesday:opponentMarlins            NA        NA      NA       NA
## day_of_weekThursday:opponentMarlins             NA        NA      NA       NA
## day_of_weekFriday:opponentMarlins          18982.6   91540.7   0.207    0.870
## day_of_weekSaturday:opponentMarlins        74618.4  173641.5   0.430    0.742
## day_of_weekSunday:opponentMarlins          16201.0    9464.8   1.712    0.337
## day_of_weekMonday:opponentMets                  NA        NA      NA       NA
## day_of_weekTuesday:opponentMets                 NA        NA      NA       NA
## day_of_weekWednesday:opponentMets               NA        NA      NA       NA
## day_of_weekThursday:opponentMets            4171.7   37106.9   0.112    0.929
## day_of_weekFriday:opponentMets                  NA        NA      NA       NA
## day_of_weekSaturday:opponentMets                NA        NA      NA       NA
## day_of_weekSunday:opponentMets            -39795.3   18793.9  -2.117    0.281
## day_of_weekMonday:opponentNationals             NA        NA      NA       NA
```

```
## day_of_weekTuesday:opponentNationals              NA          NA        NA         NA
## day_of_weekWednesday:opponentNationals            NA          NA        NA         NA
## day_of_weekThursday:opponentNationals             NA          NA        NA         NA
## day_of_weekFriday:opponentNationals          -1798.2     16321.4    -0.110      0.930
## day_of_weekSaturday:opponentNationals         6240.1     17275.2     0.361      0.779
## day_of_weekSunday:opponentNationals               NA          NA        NA         NA
## day_of_weekMonday:opponentPadres             -2703.5      6879.3    -0.393      0.762
## day_of_weekTuesday:opponentPadres             1850.2     14253.5     0.130      0.918
## day_of_weekWednesday:opponentPadres          21502.5     12527.1     1.716      0.336
## day_of_weekThursday:opponentPadres                NA          NA        NA         NA
## day_of_weekFriday:opponentPadres              5888.0      7728.0     0.762      0.586
## day_of_weekSaturday:opponentPadres            5345.0     14457.8     0.370      0.775
## day_of_weekSunday:opponentPadres            -25841.5     15689.0    -1.647      0.347
## day_of_weekMonday:opponentPhillies         -105174.7    122863.4    -0.856      0.549
## day_of_weekTuesday:opponentPhillies          19614.0      5464.5     3.589      0.173
## day_of_weekWednesday:opponentPhillies             NA          NA        NA         NA
## day_of_weekThursday:opponentPhillies              NA          NA        NA         NA
## day_of_weekFriday:opponentPhillies                NA          NA        NA         NA
## day_of_weekSaturday:opponentPhillies              NA          NA        NA         NA
## day_of_weekSunday:opponentPhillies                NA          NA        NA         NA
## day_of_weekMonday:opponentPirates                 NA          NA        NA         NA
## day_of_weekTuesday:opponentPirates                NA          NA        NA         NA
## day_of_weekWednesday:opponentPirates        -11221.0     12219.0    -0.918      0.527
## day_of_weekThursday:opponentPirates         -48001.0     13385.3    -3.586      0.173
## day_of_weekFriday:opponentPirates                 NA          NA        NA         NA
## day_of_weekSaturday:opponentPirates               NA          NA        NA         NA
## day_of_weekSunday:opponentPirates                 NA          NA        NA         NA
## day_of_weekMonday:opponentReds                -577.0     13385.3    -0.043      0.973
## day_of_weekTuesday:opponentReds                   NA          NA        NA         NA
## day_of_weekWednesday:opponentReds                 NA          NA        NA         NA
## day_of_weekThursday:opponentReds                  NA          NA        NA         NA
## day_of_weekFriday:opponentReds                    NA          NA        NA         NA
## day_of_weekSaturday:opponentReds                  NA          NA        NA         NA
## day_of_weekSunday:opponentReds                    NA          NA        NA         NA
## day_of_weekMonday:opponentRockies             2920.9      8974.6     0.325      0.800
## day_of_weekTuesday:opponentRockies           31365.0     10929.0     2.870      0.213
## day_of_weekWednesday:opponentRockies              NA          NA        NA         NA
## day_of_weekThursday:opponentRockies               NA          NA        NA         NA
## day_of_weekFriday:opponentRockies            11137.5     11540.2     0.965      0.511
## day_of_weekSaturday:opponentRockies           7897.5      5257.7     1.502      0.374
## day_of_weekSunday:opponentRockies                 NA          NA        NA         NA
## day_of_weekMonday:opponentSnakes                  NA          NA        NA         NA
## day_of_weekTuesday:opponentSnakes                 NA          NA        NA         NA
## day_of_weekWednesday:opponentSnakes               NA          NA        NA         NA
## day_of_weekThursday:opponentSnakes                NA          NA        NA         NA
## day_of_weekFriday:opponentSnakes                  NA          NA        NA         NA
## day_of_weekSaturday:opponentSnakes                NA          NA        NA         NA
## day_of_weekSunday:opponentSnakes                  NA          NA        NA         NA
```

```
## day_of_weekMonday:opponentWhite Sox             NA         NA      NA       NA
## day_of_weekTuesday:opponentWhite Sox            NA         NA      NA       NA
## day_of_weekWednesday:opponentWhite Sox          NA         NA      NA       NA
## day_of_weekThursday:opponentWhite Sox           NA         NA      NA       NA
## day_of_weekFriday:opponentWhite Sox             NA         NA      NA       NA
## day_of_weekSaturday:opponentWhite Sox           NA         NA      NA       NA
## day_of_weekSunday:opponentWhite Sox             NA         NA      NA       NA
## fireworksNO:shirtYES                            NA         NA      NA       NA
## fireworksYES:shirtYES                           NA         NA      NA       NA
##
## Residual standard error: 3864 on 1 degrees of freedom
## Multiple R-squared:  0.9973, Adjusted R-squared:  0.7831
## F-statistic: 4.657 on 79 and 1 DF,  p-value: 0.3556
```

Model 6 includes bobblehead, fireworks, and day_of_week as predictors for attend.

Show

```
##
## Call:
## lm(formula = attend ~ bobblehead + fireworks + day_of_week, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10653.7  -3399.0     50.1   3085.4  15593.3
##
## Coefficients:
##                      Estimate Std. Error t value             Pr(>|t|)
## (Intercept)           34965.7     1816.5  19.249 < 0.0000000000000002 ***
## bobbleheadYES         12618.8     2370.6   5.323           0.00000111 ***
## fireworksYES          17438.0     6572.4   2.653               0.0098 **
## day_of_weekTuesday     6951.5     2746.4   2.531               0.0136 *
## day_of_weekWednesday   1166.3     2626.7   0.444               0.6584
## day_of_weekThursday     394.2     3481.1   0.113               0.9102
## day_of_weekFriday    -12286.7     7038.6  -1.746               0.0851 .
## day_of_weekSaturday    6165.9     2545.3   2.422               0.0179 *
## day_of_weekSunday      6332.5     2525.7   2.507               0.0144 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6293 on 72 degrees of freedom
## Multiple R-squared:  0.4824, Adjusted R-squared:  0.4249
## F-statistic: 8.388 on 8 and 72 DF,  p-value: 0.00000005669
```

7th model predicts attend using bobblehead, a 3rd-degree polynomial of temp, and day_of_week as predictors.

Show

```
##
## Call:
## lm(formula = attend ~ bobblehead + poly(temp, 3) + day_of_week,
##     data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13200.6  -2956.3   -267.3   1951.3  14971.5
##
## Coefficients:
##                       Estimate Std. Error t value             Pr(>|t|)
## (Intercept)              34726       1829  18.985 < 0.0000000000000002 ***
## bobbleheadYES            11244       2439   4.610            0.0000177 ***
## poly(temp, 3)1            4250       6630   0.641              0.52358
## poly(temp, 3)2          -17916       6606  -2.712              0.00841 **
## poly(temp, 3)3            4292       6479   0.663              0.50981
## day_of_weekTuesday        7057       2763   2.554              0.01282 *
## day_of_weekWednesday      2743       2591   1.059              0.29329
## day_of_weekThursday       1606       3542   0.454              0.65156
## day_of_weekFriday         5896       2562   2.302              0.02432 *
## day_of_weekSaturday       7088       2587   2.740              0.00780 **
## day_of_weekSunday         6638       2588   2.565              0.01247 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6326 on 70 degrees of freedom
## Multiple R-squared:  0.4914, Adjusted R-squared:  0.4188
## F-statistic: 6.764 on 10 and 70 DF,  p-value: 0.000000285
```

Lets compares the AIC values of seven different models.

Show

```
##        df      AIC
## model  38 1665.178
## model2 40 1660.898
## model3 18 1646.657
## model4 34 1657.701
## model5 81 1373.950
## model6 10 1657.363
## model7 12 1659.937
```

Now we will compare the BIC values of seven different models.

Show

```
##        df      BIC
## model  38 1756.167
## model2 40 1756.676
## model3 18 1689.757
## model4 34 1739.112
## model5 81 1567.900
## model6 10 1681.307
## model7 12 1688.670
```

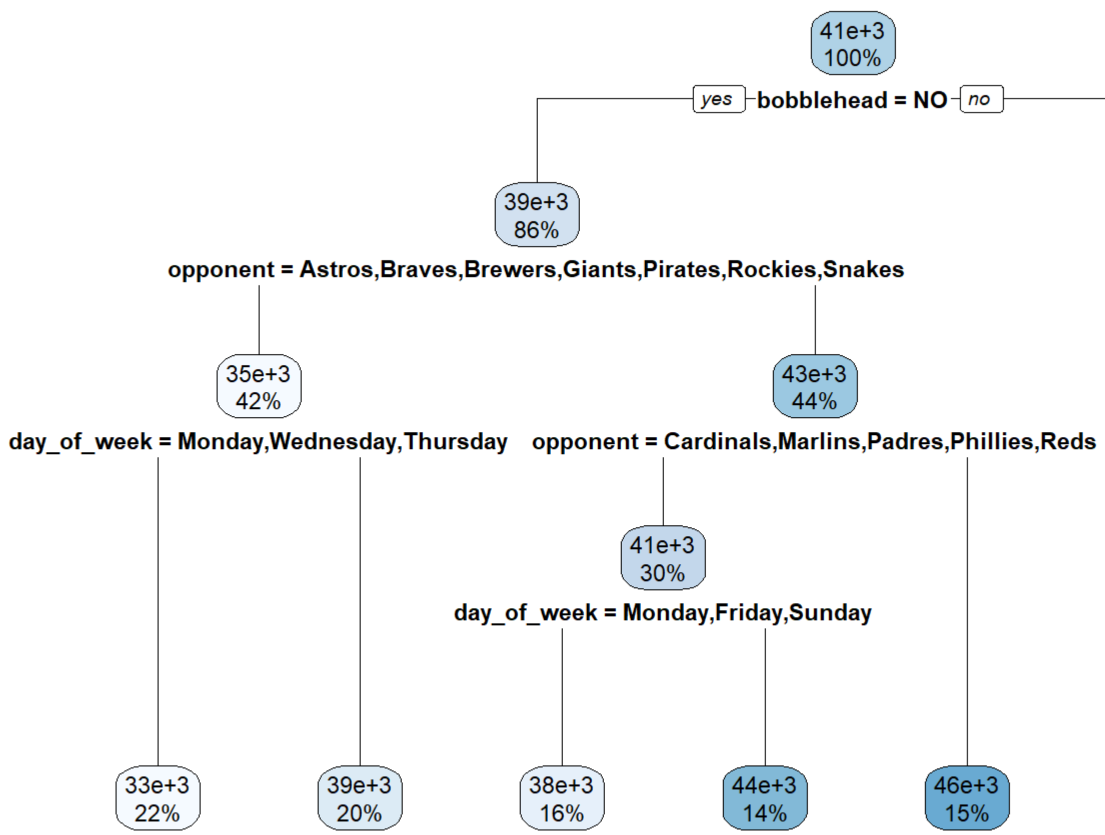
AIC is based on information theory and provides a measure of model quality that balances goodness of fit with the model's complexity. A lower AIC value suggests a better model. Comparing AIC values, model5 has the lowest AIC (1373.950), suggesting it might be the best model among those listed in terms of the trade-off between fit and complexity.

A lower BIC value indicates a better model. BIC also helps to identify the model that is most likely to be the true model among the set of candidates. According to BIC values, model5 has the lowest BIC (1567.900), indicating it is the preferred model among those compared.

The df in the output likely refers to the number of parameters in the model (including the intercept). Model5, while having the most parameters (81 df), still has the lowest AIC and BIC, suggesting that its additional complexity provides a significantly better fit to overcome the penalty for having more parameters.

Decision Tree Model

Show



Show

```
## Call:
## rpart(formula = attend ~ ., data = d)
##   n= 81
##
##           CP nsplit rel error    xerror      xstd
## 1 0.33860175      0 1.0000000 1.0223978 0.1243306
## 2 0.16176050      1 0.6613982 0.6848089 0.1074740
## 3 0.05597094      2 0.4996377 0.6735532 0.1296268
## 4 0.04022084      3 0.4436668 0.7689573 0.1394943
## 5 0.03668867      4 0.4034460 0.7743231 0.1320096
## 6 0.01000000      5 0.3667573 0.7554474 0.1261459
##
## Variable importance
##  bobblehead    opponent day_of_week       month         day       weeks
##          34          23          14          11           6           5
##        temp   fireworks   day_night
##           3           2           1
##
## Node number 1: 81 observations,    complexity param=0.3386018
##   mean=41040.07, MSE=6.799917e+07
##   left son=2 (70 obs) right son=3 (11 obs)
##   Primary splits:
##       bobblehead  splits as  LR, improve=0.3386018, (0 missing)
##       opponent    splits as  RLLLLRLLRRLLLLLLR, improve=0.1729829, (0 missin
g)
##       day_of_week splits as  LRLLLRR, improve=0.1506018, (0 missing)
##       month       splits as  LLRRRLL, improve=0.1349046, (0 missing)
##       temp        < 18.5 to the left,  improve=0.1042029, (0 missing)
##
## Node number 2: 70 observations,    complexity param=0.1617605
##   mean=39137.93, MSE=5.085115e+07
##   left son=4 (34 obs) right son=5 (36 obs)
##   Primary splits:
##       opponent    splits as  RLLLRRLRRRRRLRLLR, improve=0.25030080, (0 missin
g)
##       day_of_week splits as  LRLLRRR, improve=0.13028860, (0 missing)
##       month       splits as  LLRLLLL, improve=0.12627120, (0 missing)
##       temp        < 18.5 to the left,  improve=0.07858422, (0 missing)
##       weeks       splits as  LR, improve=0.03791949, (0 missing)
##   Surrogate splits:
##       month       splits as  LLRRRRL,  agree=0.714, adj=0.412, (0 split)
##       day_of_week splits as  LLLLRRR,  agree=0.657, adj=0.294, (0 split)
##       day         < 19.5 to the right, agree=0.629, adj=0.235, (0 split)
##       temp        < 18.5 to the left,  agree=0.586, adj=0.147, (0 split)
##       weeks       splits as  LR,       agree=0.586, adj=0.147, (0 split)
##
## Node number 3: 11 observations
##   mean=53144.64, MSE=7577867
```

```
##
## Node number 4: 34 observations,    complexity param=0.05597094
##   mean=35466.85, MSE=4.321807e+07
##   left son=8 (18 obs) right son=9 (16 obs)
##   Primary splits:
##       day_of_week splits as  LRLLRRR, improve=0.20980070, (0 missing)
##       opponent    splits as  -RLL--R-----R-RL-, improve=0.13509410, (0 missin
g)
##       day         < 10.5 to the right, improve=0.05456740, (0 missing)
##       temp        < 18.5 to the left,  improve=0.05086383, (0 missing)
##       skies       splits as  RL, improve=0.02912061, (0 missing)
##   Surrogate splits:
##       weeks     splits as  LR, agree=0.765, adj=0.500, (0 split)
##       opponent  splits as  -RLL--L-----L-RL-, agree=0.735, adj=0.438, (0 spli
t)
##       month     splits as  LR-LLRL, agree=0.706, adj=0.375, (0 split)
##       day_night splits as  RL, agree=0.647, adj=0.250, (0 split)
##       fireworks splits as  LR, agree=0.647, adj=0.250, (0 split)
##
## Node number 5: 36 observations,    complexity param=0.04022084
##   mean=42605.06, MSE=3.331112e+07
##   left son=10 (24 obs) right son=11 (12 obs)
##   Primary splits:
##       opponent    splits as  R---LR-LRRLL-L--R, improve=0.18473450, (0 missin
g)
##       month       splits as  LLRLLL-, improve=0.12388540, (0 missing)
##       temp        < 19.5 to the left,  improve=0.08144720, (0 missing)
##       day_of_week splits as  LRRRLRR, improve=0.07953047, (0 missing)
##       day         < 16.5 to the left,  improve=0.07276641, (0 missing)
##   Surrogate splits:
##       month splits as  LLRLLL-,  agree=0.861, adj=0.583, (0 split)
##       day   < 26.5 to the left,  agree=0.778, adj=0.333, (0 split)
##       shirt splits as  LR,       agree=0.694, adj=0.083, (0 split)
##
## Node number 8: 18 observations
##   mean=32627.89, MSE=2.928639e+07
##
## Node number 9: 16 observations
##   mean=38660.69, MSE=3.962345e+07
##
## Node number 10: 24 observations,    complexity param=0.03668867
##   mean=40850.96, MSE=3.407419e+07
##   left son=20 (13 obs) right son=21 (11 obs)
##   Primary splits:
##       day_of_week splits as  LRRRLRL,  improve=0.247106300, (0 missing)
##       day         < 16.5 to the left,  improve=0.038363640, (0 missing)
##       temp        < 26.5 to the right, improve=0.019428240, (0 missing)
##       skies       splits as  RL,       improve=0.008862651, (0 missing)
```

```
##       month       splits as  LR-RLL-,  improve=0.007786163, (0 missing)
##   Surrogate splits:
##       day       < 9    to the right, agree=0.625, adj=0.182, (0 split)
##       opponent  splits as  ----L--L--LR-R---, agree=0.625, adj=0.182, (0 spli
t)
##       fireworks splits as  RL, agree=0.625, adj=0.182, (0 split)
##       month     splits as  LL-LLR-, agree=0.583, adj=0.091, (0 split)
##       temp      < 21.5 to the right, agree=0.583, adj=0.091, (0 split)
##
## Node number 11: 12 observations
##   mean=46113.25, MSE=1.332383e+07
##
## Node number 20: 13 observations
##   mean=38181.77, MSE=1.663378e+07
##
## Node number 21: 11 observations
##   mean=44005.45, MSE=3.631479e+07
```

From the decision tree analysis, it looks like bobblehead has the biggest effect on attandence. The opponent and the day of week is also affecting significantly. By looking at the error at CP as tree grow it become overfit to the data, so maybe keeping tree simpler will can give better results.

Conclusion

Lets compare the best regression model and the decision tree model to see whicj one is better.

In regression analysis, we usually use RMSE and R-squared to compare the models. RMSE is the square root of the average of the squared differences between the predicted and actual values. R-squared is a measure of how well the model explains the variation in the dependent variable. It is a value between 0 and 1, with 1 indicating that the model perfectly explains the variation in the dependent variable.

Show

```
## Zorunlu paket yükleniyor: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

Show

```
## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases
```

```
## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases

## Warning in predict.lm(modelFit, newdata): prediction from rank-deficient fit;
## attr(*, "non-estim") has doubtful cases
```

Show

```
## Warning in nominalTrainWorkflow(x = x, y = y, wts = weights, info = trainInf
o,
## : There were missing values in resampled performance measures.
```

Show

```
## [1] "Regression Model RMSE: 152886.861677824"
```

Show

```
## [1] "Decision Tree RMSE: 6592.33564921623"
```

Show

```
## [1] "Regression Model R-squared: 0.0581554846715227"
```

Show

```
## [1] "Decision Tree R-squared: 0.437942613800504"
```

Show

```
## [1] "Decision Tree model is better based on RMSE and R-squared."
```

Hence, we can say that decision tree model is the optimal model for this task and we can make predictions with this model.

Show

```
##   Actual Predicted
## 6  38359  39137.93
```

The model is able to predict the attendance with a reasonable accuracy.