Assignment 3.2 - Using Data to Improve a Marketing Promotion

Ragunath Gunasekaran

2021-04-04

# 1. Import, Plot, Summarize, and Save Data and summary for every variable, structure and type of data elements

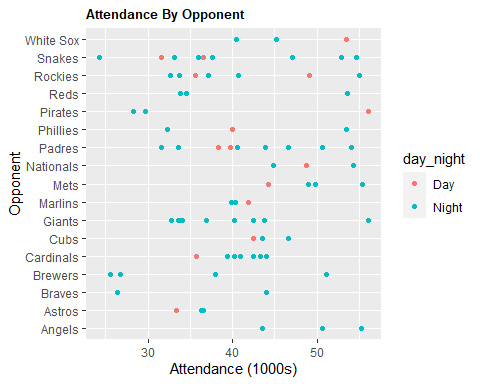
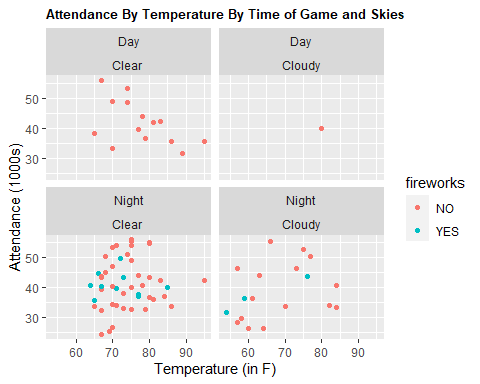
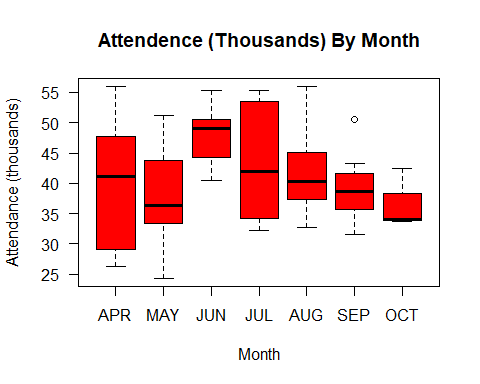
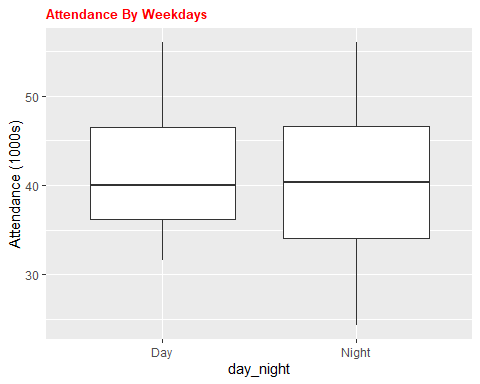
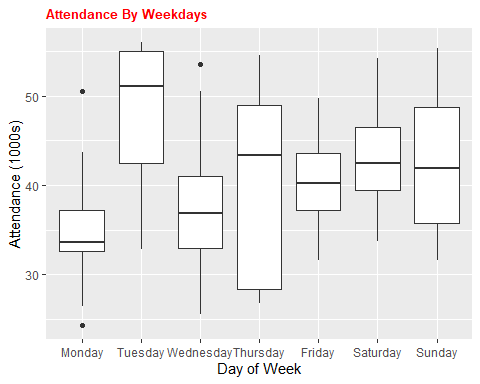
The below is the Structure of datasets which explains the detail of metadata.

## '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" ...

## [1] 81 12

The Dataset contains 81 rows and 12 Variables

## 1. Scatter and Box plots



### 2. Regression Model

##   
## Proportion of Test Set Variance Accounted for: 0.115

##   
## Call:  
## lm(formula = my.model, data = Dodgers\_df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17016.9 -4596.3 -196.4 3019.2 15195.8   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30428.2 4044.9 7.523 1.74e-10 \*\*\*  
## monthMAY -2330.9 2618.9 -0.890 0.37662   
## monthJUN 8839.4 3062.6 2.886 0.00524 \*\*   
## monthJUL 4512.2 2886.0 1.563 0.12266   
## monthAUG 3901.0 2685.8 1.452 0.15104   
## monthSEP -848.8 2848.2 -0.298 0.76660   
## monthOCT -1918.3 4590.3 -0.418 0.67735   
## day\_of\_weekTuesday 13124.6 2781.7 4.718 1.25e-05 \*\*\*  
## day\_of\_weekWednesday 2948.8 2878.8 1.024 0.30936   
## day\_of\_weekThursday 5231.1 3765.1 1.389 0.16932   
## day\_of\_weekFriday 4914.0 2831.3 1.736 0.08723 .   
## day\_of\_weekSaturday 8453.5 2845.9 2.970 0.00413 \*\*   
## day\_of\_weekSunday 10163.8 3822.5 2.659 0.00980 \*\*   
## day\_nightNight 2836.0 3107.9 0.913 0.36478   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6916 on 67 degrees of freedom  
## Multiple R-squared: 0.4182, Adjusted R-squared: 0.3054   
## F-statistic: 3.705 on 13 and 67 DF, p-value: 0.0001937

# Conclusion - Summary of Analysis

Before making a recommendation to the Dodgers Marketing Team on which month and day would be best to run a promotion to increase attendance, several factors were considered and reviewed.

R was utilized for the data preparation and model creation and prediction. The Dodgers data in csv format loaded into dataset.

In reviewing days of the week, Tuesdays were found to have the highest average attendance, followed by the weekend games on Saturdays and Sundays. Mondays averaged the lowest attendance.

With this information and analysis, the data was loaded into R for further review.

Here the multiple linear regression model was created to look at the relationship between month, day of the week, DayNight and attendance for the Dodgers data.

From the data, relationships were found between the month, day of the week head promotions, and attendance for the Dodgers with a p-value of 0.0001937.

once we split the data into testing and training segments, the model was fit with the training set and the test set was used in the prediction.

As part of the analysis in Chart and Regression Model output, we see that there is a positive impact on Tuesday night would be the best to run a marketing promotion to increase attendance - 13125.

# References:

1. <http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visualization>