DSC 630
Laura Hoffmann
**Assignment 3.2**
<div align="center">

Dodger's Marketing Promotion Report
</div>

Data Exploration and Visuals

      For data exploration I used Python with Jupyter Notebook. After checking the data set for missing values and seeing that there were none, I began exploring the data with simple statistical summaries. There were three numerical columns (one of which was the target variable (attendance) and nine categorical variables contributing to the target variable of attendance. Maximum attendance was 56,000 which occurred twice (Tuesday April 10, and Tuesday August 21) and the average attendance was right around 41,000.

      As far as visuals go, I typically would not have chosen this color scheme, but I thought it would be interesting to go with the Dodger's team colors and keep up with that theme for most of the graphs. First is a simple histogram to show the frequency of attendances throughout the season. The second visual created investigated the relationship of attendance and temperature using a scatterplot. They were not a correlated as I expected, however, I suppose it makes sense because a lot of the time tickets are purchased before the weather is known and would be used regardless. Next I used boxplots to see attendance by month, day of the week, and opponent.
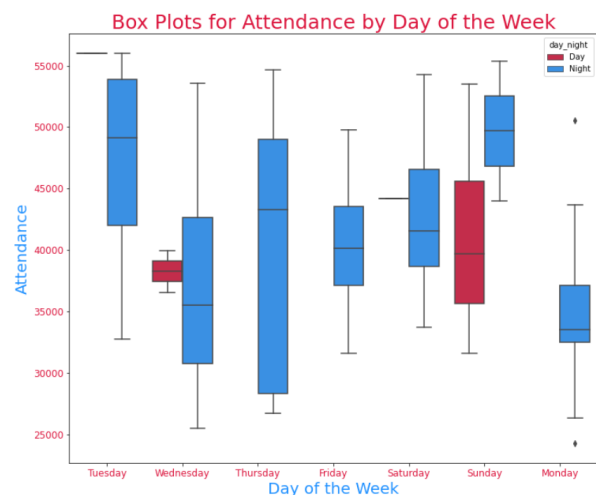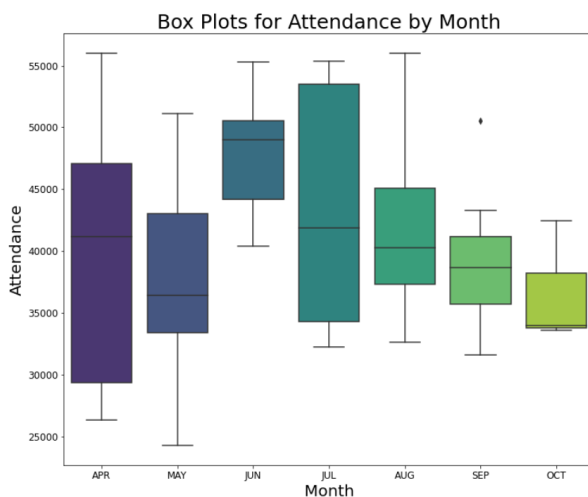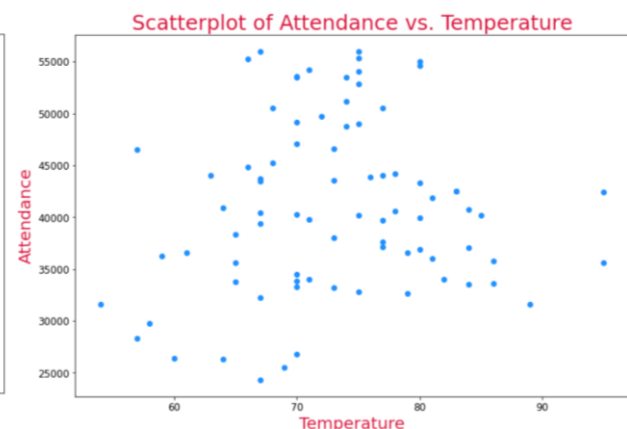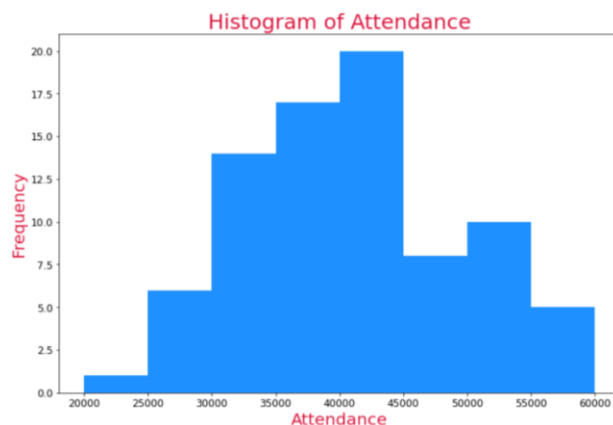
The boxplots were easier to interpret than simply looking at the numbers and visually showed that June and Tuesdays had the highest average attendance in terms of days, because the one day game brought the average up. However, if we incorporated the day_night column we can see Sunday night games had the highest average, and Sunday games during the day dragged that average down. When looking at opponents, we saw that playing the Angels brought the highest average attendance for the Dodgers.

Following the visual exploration, I used the describe and groupby functions to see summary statistics of attendance by a few of the variables. Throughout this process I reiterated and solidified a few of the statistics discovered by using the boxplots.

To build a correlation heat map I converted all of the categorical columns into integers. We can use this to see that the two most correlated variables were month and temperature, which makes a lot of sense.

Correlation Heatmap
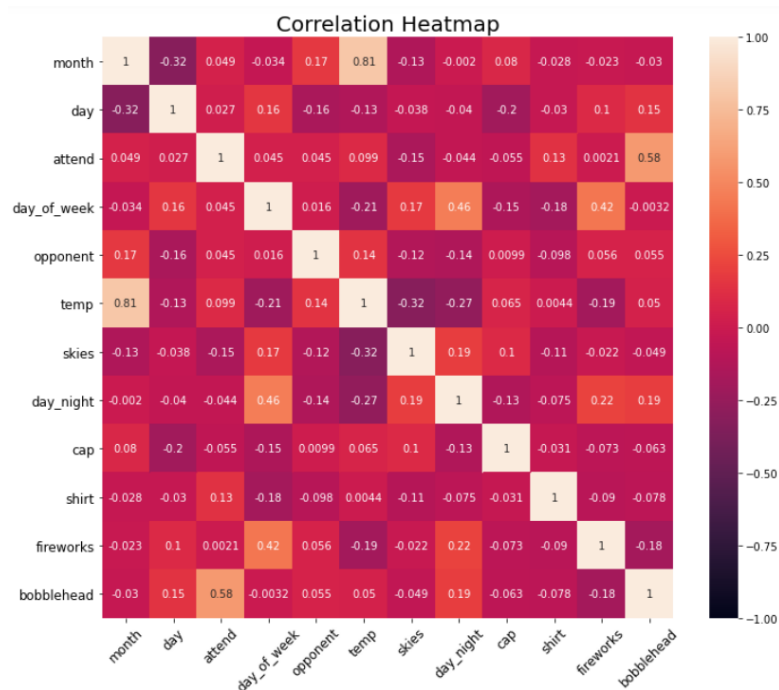
Model Building

For building the model I used both Python and R to develop a better understanding of both when the promotion should take place and what the promotion would be. To build the regression model in Python I divided the data into training and testing groups using a 75% and 25% split. I used the statsmodels package to build a linear regression model in Python using all of the columns as predictors. After building the model in Python, I saw that R-Squared was small for the model so not a lot of the variance could be explained, however, the bobblehead variable had the most statistical significance, looking at the P-value, when considering attendance.

The new CSV file created after converting the categorical columns to numerical values was reopened in R to evaluate a model a little further. Using the predictors of only day, month and bobbleheads the linear model was created and visualized through the graphic below which shows the predicted and actual attendances for each of the games for days of the week by month.

Attendance Prediction By Months
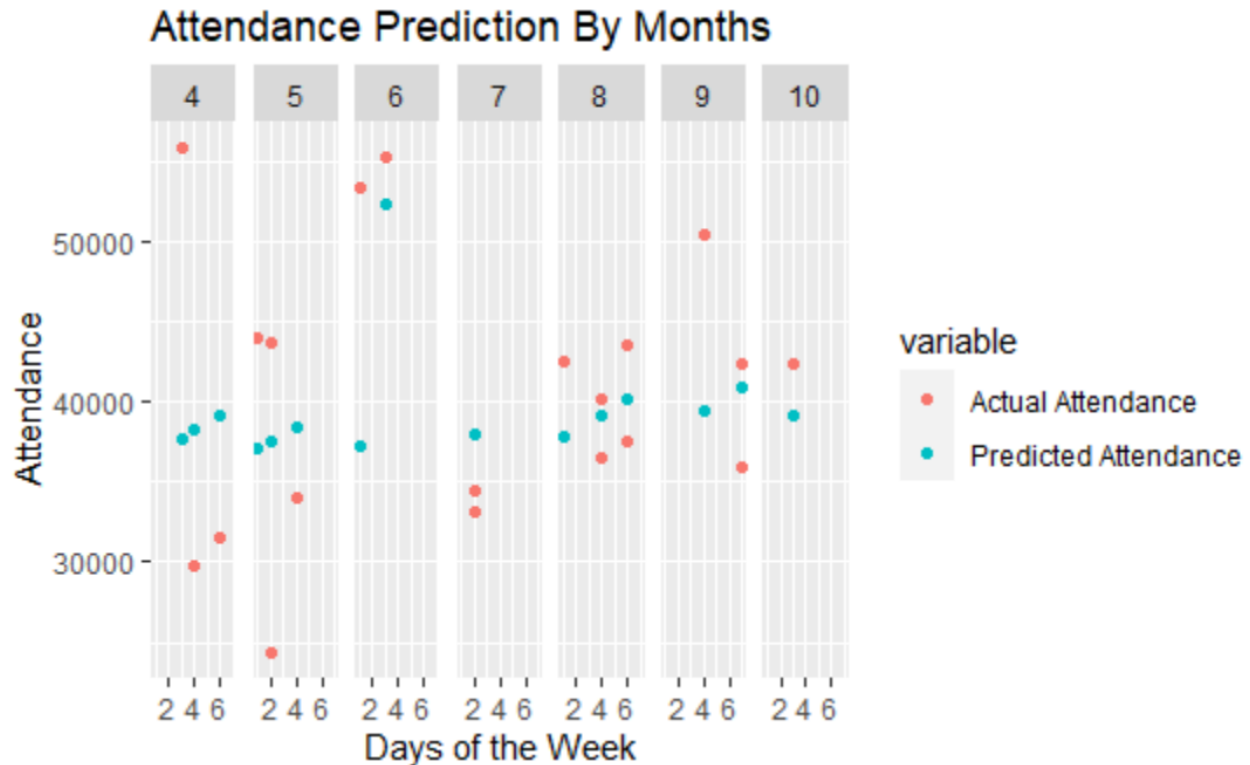
Marketing Promotion Recommendation

After discovering that bobbleheads would be the recommended type of promotion to draw larger crowds to the games, all that was left to do was discover which days or months had games with lower actual attendances than predicted in the models. From the visualized model above, we can see that Monday repeatedly had lower attendances than were predicted, particularly in July. I would recommend that a bobblehead promotion be run on a Monday in July or October for a chance to increase the attendance averages for both Monday (which suffered the lowest average attendance for days of the week) and July or October (which suffered the lowest average attendance for months despite it being the postseason). Promoting a game with bobbleheads in October could draw out a larger crowd for the postseason games, increasing not only the attendance but also the moral of the players for the season in the hopes they perform better and make it further in the season, thus adding games to their season, making room for more attendance and profits.