

Week 3 – Assignment 3.2

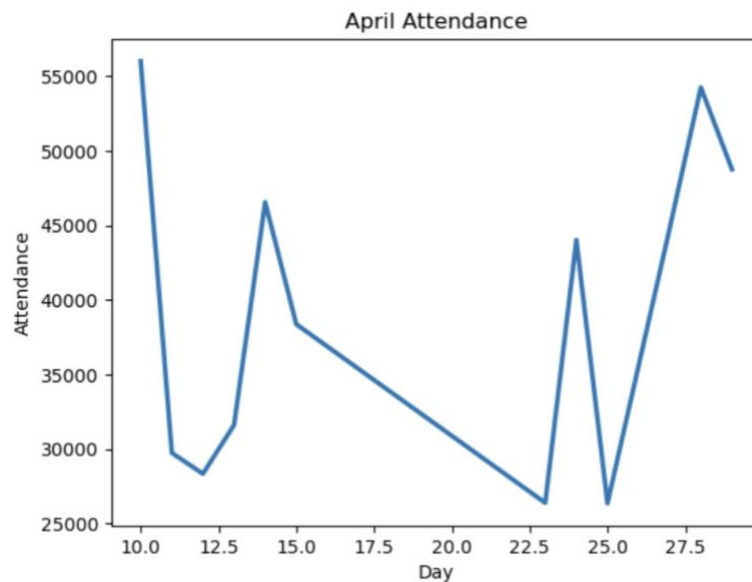
Felipe Rodriguez

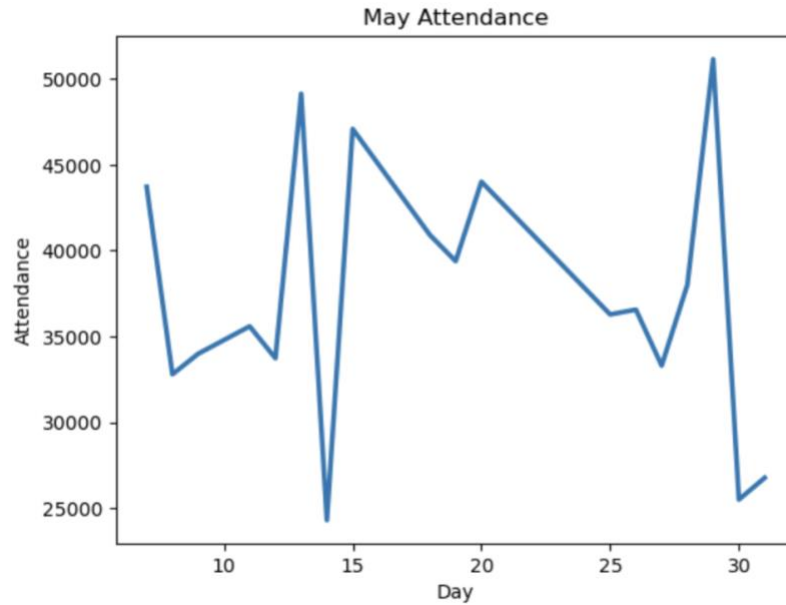
Bellevue University

DSC 630 Predictive Analytics

Professor Andrew Hua

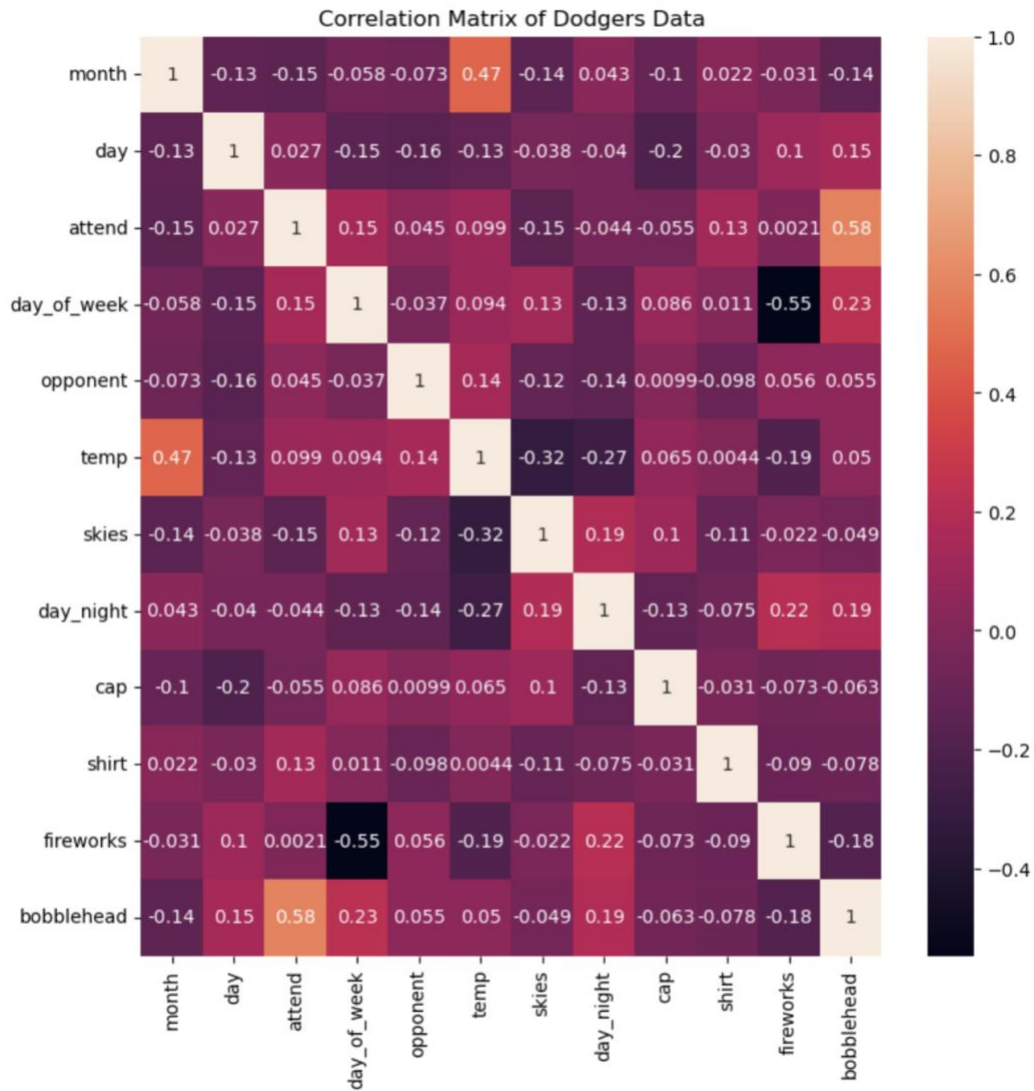
The analysis of this data will include understanding the relationship between attendance and the other variables within the data. Attendance is an important factor for the Los Angeles Dodgers and understanding the driving factors in attendance is crucial to forecast profits for the upcoming seasons. By analyzing the variables that influence attendance, a recommendation can be made to further improve areas that will likely affect the attendance for a baseball game. To gain some insight on the data, the attendance by day for the months of April and May are graphed to show how the attendance changes over time.





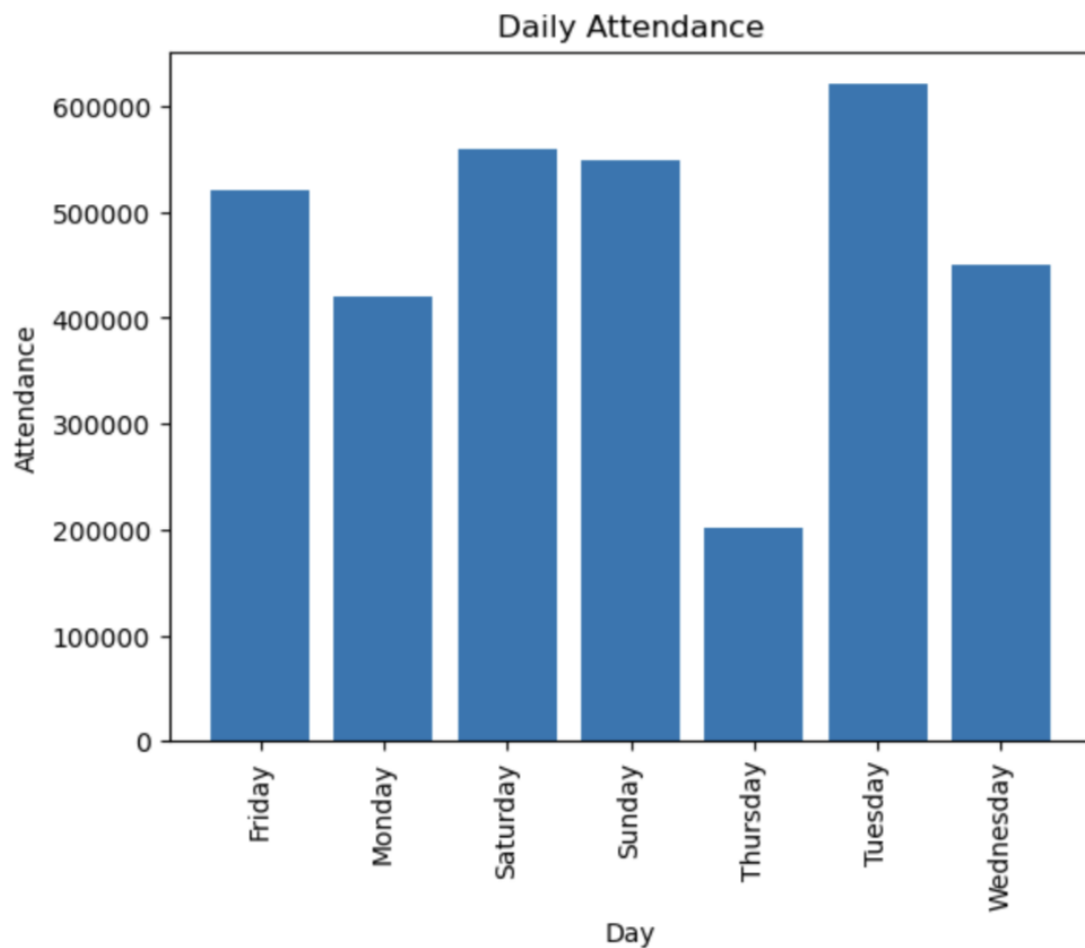
When comparing attendance for both months, there are no similarities between the two. It can be noted that the May attendance is higher than April attendance. This is confirmed by adding attendance for each month. For April, total attendance was 475,103 and for May it was 672,223. One reason May is higher in attendance is that the data contains a full month for May. The data starts from April 10th which omits part of that month.

A great way to view correlation is by creating a correlation matrix; this will display the values that influence each other, and we can focus on attendance to see which variable affects it the most.



The correlation matrix shows us that the variable that positively influences attendance the most is if there was a bobblehead during the game. The variable that negatively affects attendance is skies, however it is not possible to control this variable.

The next factor explored, will be understanding how days of the week play a role in the games. Although day of the week had a slight positive correlation on attendance, further exploration can uncover more details of this variable.



When looking at attendance by day, the weekend days are noticeably the highest, but to our surprise, Tuesday has the most attendance of the games. When looking at the number of games for each day, Friday through Monday had around the same number of days, while Thursday had only 5 games. This explains why Thursdays had the least attendance.

After analyzing the data, we have uncovered three recommendations for attendance. The first recommendation is to include the bobblehead in more games. This was uncovered in the correlation matrix, and it showed to have a positive correlation between attendance and the appearance of the bobblehead. The second recommendation is to redistribute more games to Thursdays. Since this day already has the fewest number of games, spreading the games out to include more Thursdays can create an additional opportunity for fans to attend games. The last recommendation is to continue to promote Tuesday games since they drive the most attendance.