Name: Muley, Tushar
Case Study

In the analysis we are take data from the Dodger 2012 attendance count and find the lowest number of attendance night and make a recommendation for a promotion that will increase the attendance. I selected to do a day of the week.

My thought was to do analysis of the data first determine what I was looking at. I performance a correlation matrix.

	day	attend	day_of_week	temp	skies	day_night	cap	shirt	fireworks	bobblehead
day	1.000000	0.027093	0.160664	-0.127612	-0.038396	0.039828	-0.202274	-0.030182	0.099528	0.145363
attend	0.027093	1.000000	0.045177	0.098951	-0.150963	0.043544	-0.055002	0.133269	0.002094	0.581895
day_of_week	0.160664	0.045177	1.000000	-0.210856	0.173522	-0.455972	-0.147091	-0.181300	0.423744	-0.003203
temp	-0.127612	0.098951	-0.210856	1.000000	-0.316584	0.272141	0.064521	0.004394	-0.189899	0.049573
skies	-0.038396	-0.150963	0.173522	-0.316584	1.000000	-0.188903	0.099671	-0.108566	-0.021880	-0.049349
day_night	0.039828	0.043544	-0.455972	0.272141	-0.188903	1.000000	0.128951	0.074796	-0.217922	-0.188982
cap	-0.202274	-0.055002	-0.147091	0.064521	0.099671	0.128951	1.000000	-0.031204	-0.072732	-0.063074
shirt	-0.030182	0.133269	-0.181300	0.004394	-0.108566	0.074796	-0.031204	1.000000	-0.089648	-0.077743
fireworks	0.099528	0.002094	0.423744	-0.189899	-0.021880	-0.217922	-0.072732	-0.089648	1.000000	-0.181207
bobblehead	0.145363	0.581895	-0.003203	0.049573	-0.049349	-0.188982	-0.063074	-0.077743	-0.181207	1.000000

Figure 1: Correlation Matric of the Dodger csv file

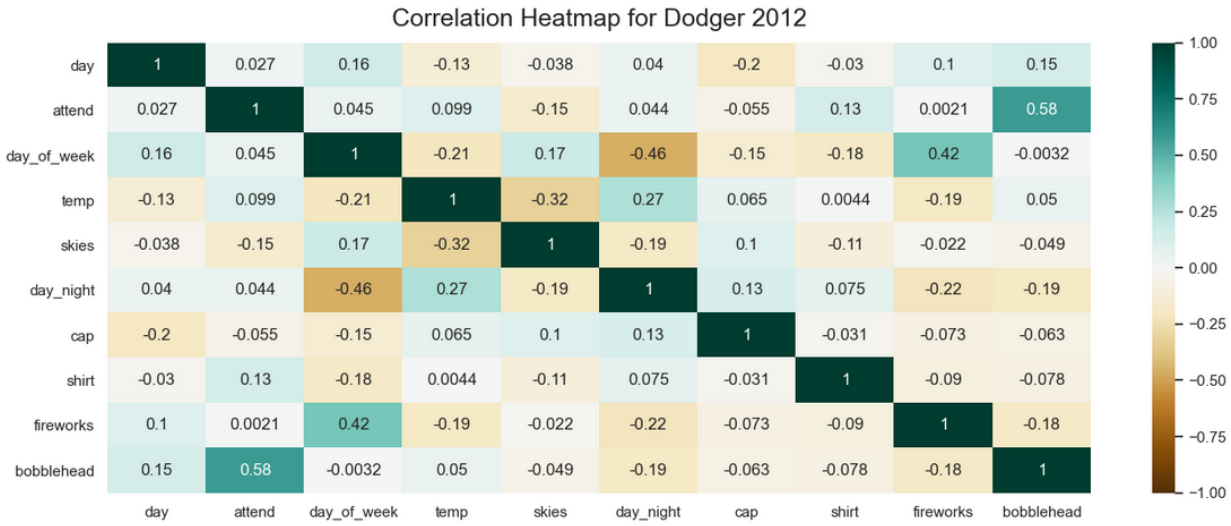

Figure 2: Heatmap of the Correlation Matrix of the Dodger csv file

From the above correlation matrix, I can determine that bobbleheads provide and increase in attendance. Based on the above correlation matrix I plot a boxplot to show the increase that bobblehead promotion provides.

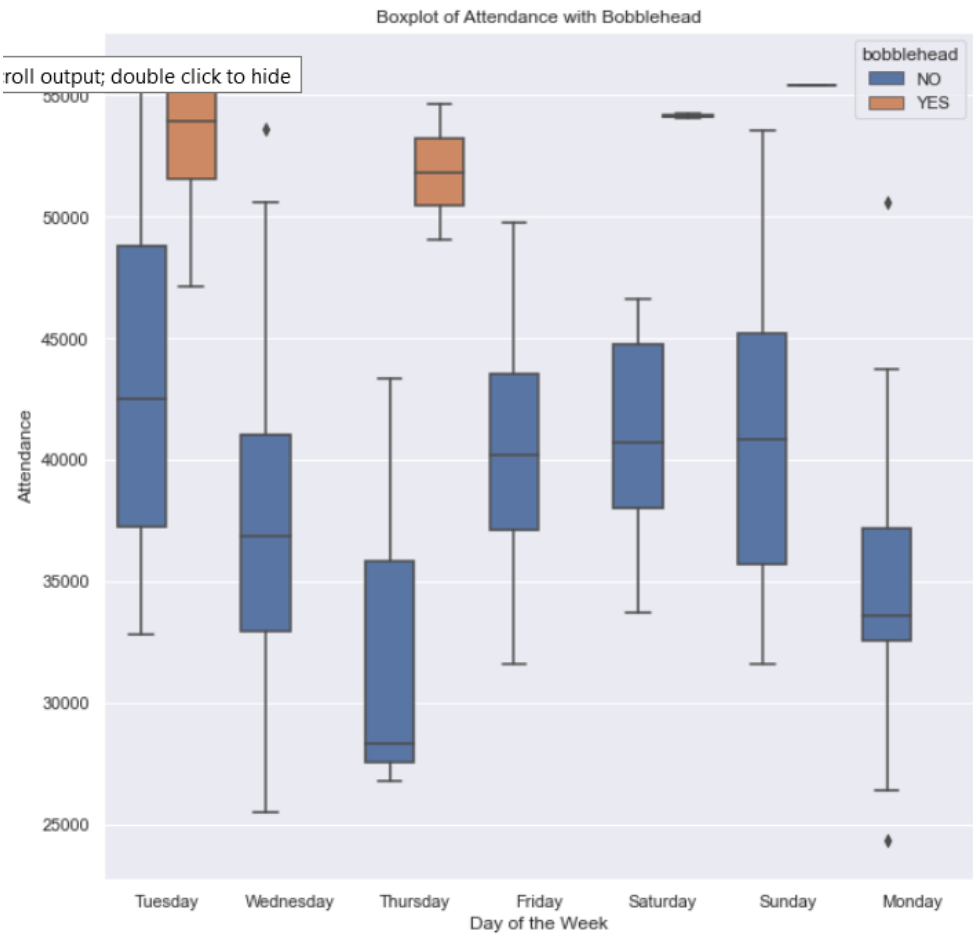

Figure 3: Boxplot of Attendance with Bobblehead

As the boxplot above proves the bump from bobblehead to attendance of games is evident. I also plotted a scatter plot before perform a liner regression. The boxplot also shows that Thursday seem to have the lowest amount of attendance.

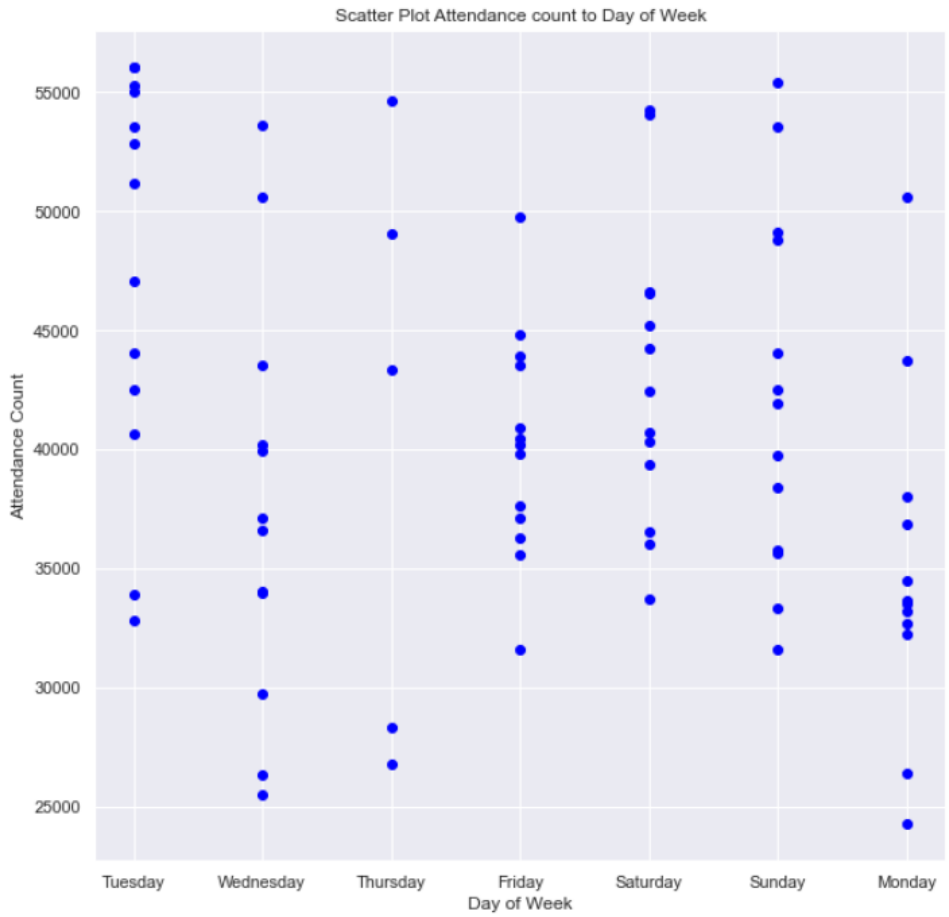

Figure 4: Scatter plot of Attendance and Day of Week

The scatter plot above provides a similar view of the data. You do have days that are lower than Thursdays. I ran a quick count of days.

```
In [27]: df_dodger['day_of_week'].value_counts()

Out[27]: Friday       13
         Tuesday      13
         Sunday       13
         Saturday     13
         Wednesday    12
         Monday       12
         Thursday      5
         Name: day_of_week, dtype: int64
```

Figure 5: Count of day_of_week.

The above show Thursdays are the lowest number of days that a game was held. That might lead to Thursday being lower than all other days.

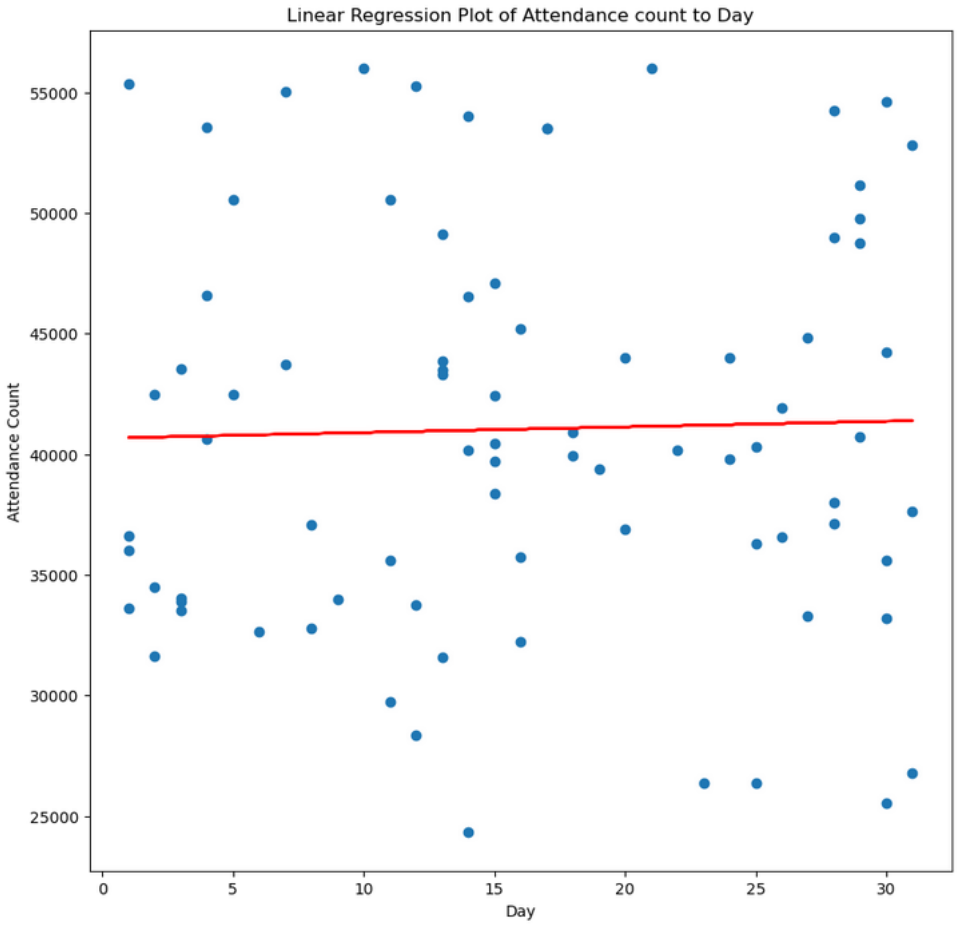

Figure 6: Linear Regression on Attendance and count of Day

After running a linear regression model on Attendance and count of day. Beginning of the month has lower attendance than later in the month. I am making an assumption that Monday and start of months are lower in attendance.

Conclusion:

I am recommending Mondays that are closer to the beginning of the months as good days for bobblehead promotions, which should cause higher bump in attendance. I have ruled out Thursday because there is not enough Thursday compared to all the other days of the week. Based on box plot Monday are lower in attendance compared to all other days which have similar number of games. It also makes sense that earlier. The linear regression shows a increase in attendance the further games are from the start of the month.

I would recommend Monday, away from the start of the month and using bobblehead as the promotion type based on the correlation matrix.