DSC630-T302: Brandon Mather - Week 3 assignment

In this assignment, you will be using data on the Los Angeles Dodgers Major League Baseball (MLB) team. Use this data to make a recommendation to management on how to improve attendance. Tell a story with your analysis and clearly explain the steps you take to arrive at your conclusion. This is an open-ended question, and there is no one right answer. You are welcome to do additional research and/or use domain knowledge to assist your analysis, but clearly state any assumptions you make.

```python
In [2]: import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
```

```python
In [3]: dodgers = pd.read_csv('dodgers-2022.csv')
```
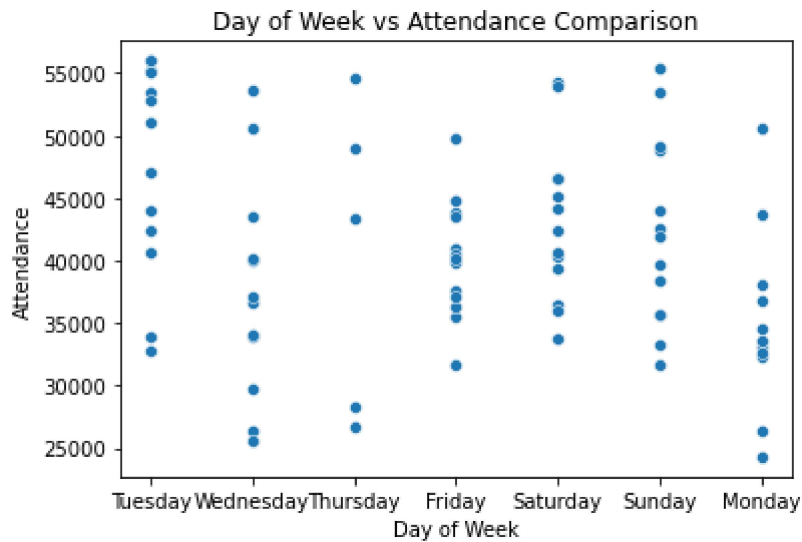
```python
In [4]: dodgers
```

Out[4]:

	month	day	attend	day_of_week	opponent	temp	skies	day_night	cap	shirt	fireworks	bc
0	APR	10	56000	Tuesday	Pirates	67	Clear	Day	NO	NO	NO	
1	APR	11	29729	Wednesday	Pirates	58	Cloudy	Night	NO	NO	NO	
2	APR	12	28328	Thursday	Pirates	57	Cloudy	Night	NO	NO	NO	
3	APR	13	31601	Friday	Padres	54	Cloudy	Night	NO	NO	YES	
4	APR	14	46549	Saturday	Padres	57	Cloudy	Night	NO	NO	NO	
...	...	...	...	...	...	...	...	...	...	...	...	
76	SEP	29	40724	Saturday	Rockies	84	Cloudy	Night	NO	NO	NO	
77	SEP	30	35607	Sunday	Rockies	95	Clear	Day	NO	NO	NO	
78	OCT	1	33624	Monday	Giants	86	Clear	Night	NO	NO	NO	
79	OCT	2	42473	Tuesday	Giants	83	Clear	Night	NO	NO	NO	
80	OCT	3	34014	Wednesday	Giants	82	Cloudy	Night	NO	NO	NO	

81 rows × 12 columns

```python
In [5]: #Seeing if Day of the Week effects attendance using scatter plot
        scatter = sns.scatterplot(data=dodgers, x="day_of_week", y="attend")
        scatter.set(xlabel='Day of Week',
                ylabel='Attendance',
                title='Day of Week vs Attendance Comparison')
```

Out[5]: [Text(0.5, 0, 'Day of Week'),
         Text(0, 0.5, 'Attendance'),
         Text(0.5, 1.0, 'Day of Week vs Attendance Comparison')]

Day of Week vs Attendance Comparison

```
In [6]:  Monday_games = dodgers[dodgers['day_of_week'] == 'Monday']
         Tuesday_games = dodgers[dodgers['day_of_week'] == 'Tuesday']
         Wednesday_games = dodgers[dodgers['day_of_week'] == 'Wednesday']
         Thursday_games = dodgers[dodgers['day_of_week'] == 'Thursday']
         Friday_games = dodgers[dodgers['day_of_week'] == 'Friday']
         Saturday_games = dodgers[dodgers['day_of_week'] == 'Saturday']
         Sunday_games = dodgers[dodgers['day_of_week'] == 'Sunday']
```

```
In [7]:  Monday_games_mean = Monday_games['attend'].mean()
         Tuesday_games_mean = Tuesday_games['attend'].mean()
         Wednesday_games_mean = Wednesday_games['attend'].mean()
         Thursday_games_mean = Thursday_games['attend'].mean()
         Friday_games_mean = Friday_games['attend'].mean()
         Saturday_games_mean = Saturday_games['attend'].mean()
         Sunday_games_mean = Sunday_games['attend'].mean()
```

```
In [17]:  print("Monday Game Average Attendance:", Monday_games_mean)
          print("Tuesday Game Average Attendance:", Tuesday_games_mean)
          print("Wednesday Game Average Attendance:", Wednesday_games_mean)
          print("Thursday Game Average Attendance:", Thursday_games_mean)
          print("Friday Game Average Attendance:", Friday_games_mean)
          print("Saturday Game Average Attendance:", Saturday_games_mean)
          print("Sunday Game Average Attendance:", Sunday_games_mean)
          print("Overall Average Attendance:", attendance_mean)
```

```
Monday Game Average Attendance: 34965.666666666664
Tuesday Game Average Attendance: 47741.230769230766
Wednesday Game Average Attendance: 37585.166666666664
Thursday Game Average Attendance: 40407.4
Friday Game Average Attendance: 40116.92307692308
Saturday Game Average Attendance: 43072.92307692308
Sunday Game Average Attendance: 42268.846153846156
Overall Average Attendance: 41040.07407407407
```

Tuesday's, Saturday's, and Sunday's have more higher attendance compared to Wednesday's, Monday's, Thursday's, and Friday's.

```
In [9]:  #Seeing if time of day effects attendence
         night_games = dodgers[dodgers['day_night'] == 'Night']
         day_games = dodgers[dodgers['day_night'] == 'Day']
```

```python
In [10]:   night_games_mean = night_games['attend'].mean()
           day_games_mean = day_games['attend'].mean()
```

```python
In [16]:   print("Night Game Average Attendance:", night_games_mean)
           print("Day Game Average Attendance:", day_games_mean)
           print("Overall Average Attendance:", attendance_mean)
```

```
Night Game Average Attendance: 40868.893939393936
Day Game Average Attendance: 41793.26666666667
Overall Average Attendance: 41040.07407407407
```
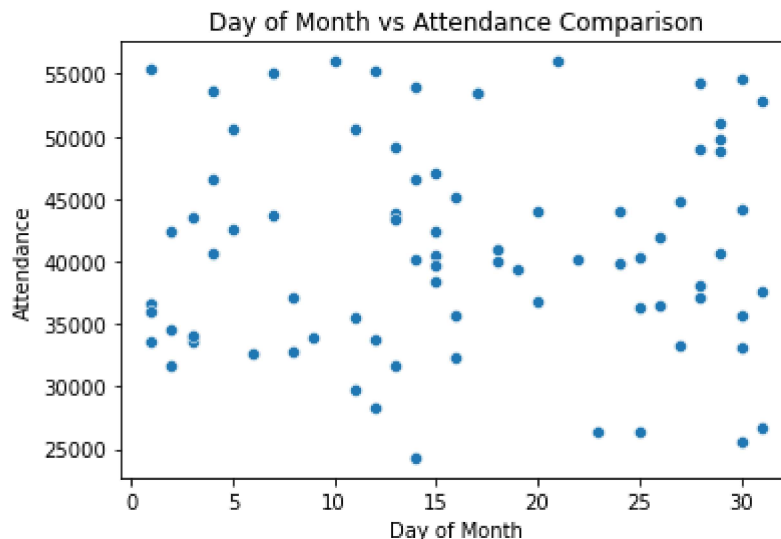
Day Game Average is better then night time, but not a large amount

```python
In [12]:   #Seeing if Day of the Month effects attendance using scatter plot
           scatter2 = sns.scatterplot(data=dodgers, x="day", y="attend")
           scatter2.set(xlabel='Day of Month',
                   ylabel='Attendance',
                   title='Day of Month vs Attendance Comparison')
```

```
Out[12]:   [Text(0.5, 0, 'Day of Month'),
            Text(0, 0.5, 'Attendance'),
            Text(0.5, 1.0, 'Day of Month vs Attendance Comparison')]
```


No meaningful impact on attendace for the day of the month.

```python
In [13]:   #Seeing if having a special event effects attendance
           cap_games = dodgers[dodgers['cap'] == 'YES']
           shirt_games = dodgers[dodgers['shirt'] == 'YES']
           fireworks_games = dodgers[dodgers['fireworks'] == 'YES']
           bobblehead_games = dodgers[dodgers['bobblehead'] == 'YES']
```

```python
In [14]:   cap_games_mean = cap_games['attend'].mean()
           shirt_games_mean = shirt_games['attend'].mean()
           fireworks_games_mean = fireworks_games['attend'].mean()
           bobblehead_games_mean = bobblehead_games['attend'].mean()
           attendance_mean = dodgers['attend'].mean()
```

```python
In [15]:   print("Cap Game Average Attendance:", cap_games_mean)
           print("Shirt Game Average Attendance:", shirt_games_mean)
           print("Fireworks Game Average Attendance:", fireworks_games_mean)
```

```
print("Bobblehead Game Average Attendance:", bobblehead_games_mean)
print("Overall Average Attendance:", attendance_mean)
```

```
Cap Game Average Attendance: 38189.5
Shirt Game Average Attendance: 46643.666666666664
Fireworks Game Average Attendance: 41077.857142857145
Bobblehead Game Average Attendance: 53144.63636363636
Overall Average Attendance: 41040.07407407407
```

In [32]:
```
dodgers.replace(('YES', 'NO'), (1, 0), inplace=True)
```

In [39]:
```
bobblehead_corr = np.corrcoef(dodgers['bobblehead'], dodgers['attend'])[0,1]
print ('Bobblehead correlation to attendance:', bobblehead_corr)
```

```
Bobblehead correlation to attendance: 0.5818949681431969
```

In [42]:
```
shirt_corr = np.corrcoef(dodgers['shirt'], dodgers['attend'])[0,1]
print ('Shirt correlation to attendance:', shirt_corr)
```

```
Shirt correlation to attendance: 0.13326867137616527
```

Giving away a free shirt increases attendance on average by about 5,000, giving away Bobblehad increases attendance average by about 12,000.

Based off my review of the data given for attendance I would say the best way to increase attendance for games would be to hold those games on Tuesday's, Saturday's, and Sunday's, and then have Bobblehead and Shirt giveaways. I looked at possible effects on attendance including day of the week, time of day, day of the month, and special events. There was no considerable effect on attendance looking at time of day or day of the month. The largest increases in average attendance were due to Bobblehead giveaways - 29.4% increase, having a Tuesday game - 16.3%, and Shirt giveaways - 13.6% increase.